Emerging Topics in Macroeconomics

Richard O. Bailly

Editor

NOVA

EMERGING TOPICS IN MACROECONOMICS

EMERGING TOPICS IN MACROECONOMICS

RICHARD O. BAILLY
EDITOR

Nova Science Publishers, Inc.
New York

LIBRARY OF CONGRESS CATALOGING-IN-PUBLICATION DATA
Emerging topics in macroeconomics / Richard O. Bailly (editor).
p. cm.
ISBN 978-1-61728-395-6 (E-Book)
1. Macroeconomics. I. Bailly, Richard O.
HB172.5.E53 2008
339--dc22
2008030433

Published by Nova Science Publishers, Inc. ✣ New York

CONTENTS

PREFACE

This new book is devoted to new research on macroeconomics which is a branch of economics that deals with the performance, structure, and behavior of a national or regional economy as a whole. Along with microeconomics, macroeconomics is one of the two most general fields in economics. Macroeconomists study aggregated indicators such as GDP, unemployment rates, and price indexes to understand how the whole economy functions. Macroeconomists develop models that explain the relationship between such factors as national income, output, consumption, unemployment, inflation, savings, investment, international trade and international finance. In contrast, microeconomics is primarily focused on the actions of individual agents, such as firms and consumers, and how their behavior determines prices and quantities in specific markets.

While macroeconomics is a broad field of study, there are two areas of research that are emblematic of the discipline: the attempt to understand the causes and consequences of short-run fluctuations in national income (the business cycle), and the attempt to understand the determinants of long-run economic growth (increases in national income).

Macroeconomic models and their forecasts are used by both governments and large corporations to assist in the development and evaluation of economic policy and business strategy.

The combination of a computable general equilibrium model (CGE) and a microsimulation (MS) model is a novel approach designed to build a bridge between macro- and microeconomic analyses. This approach allows for the assessment of the effects of macroeconomic policies on microeconomic issues such as income distribution. However, the behavioural microsimulation models used for this purpose in the literature make no distinction between full-time and part-time work but only for the choice of a sector/occupation. Labour supply modelling has the potential to improve the quality of policy evaluations based on this type of models by allowing for different levels of labour supply responses by individuals.The first short Communication discusses the extension of a CGE-MS model applied to South Africa to allow for labour supply responses. An application shows that although labour supply modelling has the potential to bring additional insights regarding income distribution issues, its contribution is limited in the case of the evaluation of trade liberalisation policies. The impacts are driven by changes in employment levels, while labour supply responses are fairly small and do not significantly affect the results. Nonetheless, this should not be seen as a general case against labour supply modelling in

CGE-MS models since the study of trade liberalisation in South Africa cannot be generalised to other policies or other countries.

The second Short Communication uses a long-run structural VAR approach to study the sources of real exchange rate fluctuations of yen-dollar rates. The authors begin by identifying two types of macroeconomic shock (real and nominal), in order to reveal the sources of movements in real exchange rates. The evidence presented indicates that real shocks play a dominant role in explaining the fluctuations of the real exchange rate. Next, they extend the model to identify three types of macroeconomic shock (real supply, real demand, and nominal), again to reveal the sources of movements in real exchange rates. The evidence from the extended model also indicates that real shocks (real demand and real supply) play a dominant role in explaining the fluctuations of the real exchange rate.

In the third Short Communication, the authors discuss a potentially fruitful but largely neglected area of research in macroeconomics. It has often been argued that catastrophic events have little impact on the aggregate economic performance as these events are typically regional, e.g. the September-11th terrorist attack was mainly concentrated in a district in the city of New York, and hence the aggregate impact in terms of lost output in the US is a negligible portion of its GDP, as argued by Sharpiro (2003) among others. These catastrophic events may also be theoretically insurable in well developed insurance markets. However, the US cannot be generalized, as it is such a large economy with abundant physical and human capital. A question has to be asked whether such a terrorist attack can occur in several financial and economic centers at the same time with similar or even greater scales. A man-made catastrophic event or natural disaster may also occur in a developed or developing country in which almost all human or productive capital stocks are concentrated in a particular region. Conventional arguments that downplay the effects of catastrophic events often fail to take into account other factors such as the level of public confidence, a higher degree of surrounding uncertainty and its associated economic costs and the time duration required to restore productive capital. The conventional arguments also rely on the assumption that the government will take any necessary counter-cyclical measures to avert serious economic downturns, which may not necessarily be the case in many developing countries.

In economic data, skewed and thick tailed frequency curves are the rule rather than the exception. Again, there seems to be an increasing interest in the study of economic variables which can be considered as realizations of a stochastic process. Within this class a major role is played by income distribution.

It is undoubtedly true that the distribution of the command over the goods and services produced by a society is of crucial importance for the student of welfare economics. On a more practical level, producers of consumer goods must study the distribution of income in order to estimate the probable extent of their markets. The size distribution of income also affects the social and economic policy-making when planning the magnitude and the structure of taxation schemes or evaluating the effectiveness of tax reforms. In economic and social statistics, the form of the income distribution is the basis for the measurement of the inequality of incomes and more general social welfare evaluations.

Over the last 100 years, a large number of distributions has been proposed for the modeling of size distribution of personal incomes. The most canonical version of such models was discovered by the mathematical economist and sociologist Vilfredo Pareto in 1897. Pareto formulated his model to explain the distribution of income and wealth, and believed that it was a universal candidate for the mathematical description of given income data for different countries over relatively long time periods. However, more recent studies suggest that it is only the upper end of income and wealth distributions that follows the Pareto's model, with the lower ends following the lognormal form of the Gaussian distribution that is associated with the random walk, originally proposed for the whole of the income distribution by the French economist Robert Gibrat in 1931.

Since the early 1990s, there has been an explosion of work on economic size phenomena in the physics literature, leading to an emerging new field called "econophysics". The list of such phenomena includes, among others, the distribution of returns in financial markets, the distribution of income and wealth, the distribution of economic shocks and growth rate variations, and the distribution of firm sizes and growth rates. A common theme among those who identify themselves as econophysicists is that standard economic theory has been inadequate or insufficient to explain the non-Gaussian properties empirically observed for various of these phenomena, such as excessive skewness and leptokurtotic fat tails. Indeed, many economic phenomena occur according to distributions that obey scaling laws rather than Gaussian normality. Whether symmetric or skewed, the tails are fatter or longer than they would be if Gaussian, and they appear to be linear in figures with the logarithm of a variable plotted against the logarithm of its cumulative probability distribution. This is true of the Pareto's distribution, and soon a variety of efforts have been made by physicists, mathematicians, and economists to model a variety of stochastic economic phenomena using either the Pareto's distribution or one of its relatives or generalizations, such as for instance the Lévy stable distribution.

In the light of these considerations, Chapter 1 offers a reader's guide to "scaling" phenomena in economics with particular focus on the interplay between the dynamics and empirical regularities of income distribution.

Inflation control is the main task of central banks in modern economies. Some attempts have been taken in the recent literature to design systematic feedback rule to regulate inflation at a prescribed level. The Taylor rule isd the simplest instrumental rule to guide monetary policy to control inflation where the instrument (*e*▷*g*▷, a short interest rate) responds to changes in the inflation and the output gaps. The objective of Chapter 2 is to modify the Taylor rule in order to improve its robustness with respect to uncertainties about potential output and unanticipated shocks. To this end, departing from feedback control theory, the Taylor rule is equipped with an adaptive control scheme to reject the adverse effects of shocks and to estimate the deviations of the potential output. It is shown that the proposed adaptation procedure is equivalent to a classical integral feedback controller whose characteristics and implementation issues are well understood in practical control engineering. Singular perturbation methods are used to establish the stability properties of the resulting control system.

Chapter 3 focuses on some methodological perspectives and their applicability to the present from the F. List point of view. The political and economic backgrounds to Friedrich List's ideas are presented in the next chapter; further, the dominated philosophy and methodology of Germany and Britain is presented. Next, I described the main ideas of F. List in the context of German romanticisms, the methodological methods of economics, and socio-economic situation in Germany. Finally, I presented some conclusions and two perspectives of liberalisation of economy in the context of the European Union policy.

After reviewing the basic self-enforcing labour contracts models, Chapter 4 explores some research avenues that can help understand the most important dynamic properties of macroeconomic variables. The idea of adapting the microeconomic theory of Thomas and Worrall [1988] to address the issue of macroeconomic dynamics was first proposed by Calmès [1999, 2003]. The authors thus describe how future research applying the theory of self-enforcing labour contracts will improve the way macroeconomic models can account for the response of the economy to external shocks. First, the introduction, by Calmès [2007], of a state-dependent opportunity for the firm in a model of self-enforcing labour contract represents in itself a novelty which generalizes the properties of this model class [Thomas and Worrall 2007a]. Second, this contribution helps getting rid of the classical result that competition among firms drives profits to zero. Going further in this direction by making capital variable, although desirable, leads to an impasse: the contract set will be no longer compact in this case, and this creates a major difficulty to solve the model. The authors discuss how this issue can be handled with a lottery incorporated in the model to convexify the contract set. For that matter, the authors also discuss the introduction of a third agent and a second relationship. Third, the question of economic growth is examined. More precisely the authors analyse the link between stationarity and set convexity when introducing growth in the model. A stochastic trend may be considered but the difficulty here will come again from the necessity of having a compact set within this class of models. Fourth, the aggregation of heterogeneous individual contracts is another important topic to look at. Fifth, the authors will deal with the problems related to the contracts duration. In other respects, according to Thomas and Worrall [1988, 2007a], risk-sharing is an important motive for contractual solutions. It is therefore relevant to study how the dynamic properties of the model change with the mathematical formulation of the individual utility functions. Utility functions could be assumed nonseparable to better account for the comovements between consumption and hours worked [Calmès 2007].

Knowing the experience of the post-communist countries that faced high inflation rates right after the 1990, there is no wonder that many central banks of those countries embraced the idea of inflation targeting. After a period when most of these countries followed a monetary policy of stabilizing the exchange rate, many of them adopted an inflation targeting strategy. Although the central banks that adopted an inflation targeting strategy want to emphasize the importance of such a strategy in fighting the inflation, this fact is not necessarily true. In most of the Central and Eastern European Countries (CEEC) inflation faced a decreasing trend after the high values recorded right after 1990, no matter their monetary arrangements.

Inflation targeting in the transition economies was a more difficult and provocative task than in the developed countries. So, is this monetary strategy adequate for the transition economies, given the actual economic conditions, and if it is not, which will be the alternative?

Many times, the candidate countries declared that the requests of the Maastricht Treat regarding the inflation and the long-term interest rate are too restrictive for them and that these should not be applied to the CEEC. There are some important studies (the one of the vice-governor of the Hungarian Central Bank, Szapary G., 2000) that prove that, due to the Balassa-Samuelson effect, is rather counter-productive and not likely that the inflation rates in the new EU member states to rapidly converge to the ones of the best performers in the euro-zone.

So, the entry in the Exchange Rate Mechanism (ERM 2) must be preceded by reaching a high level of convergence, namely by a significant progress in the structural reforms area, by a fiscal consolidation process and by a responsible income policy.

If the CEEC enter the ERM 2 before reaching a sufficient advanced stage of convergence or of financial stability, the risk of some speculative attacks increases.

All the countries that acceded or will accede in the future to the European Union want to stay in the ERM 2 the shortest period that is possible (two years), although there are no formal restrictions regarding this period. For avoiding an excessive long participation to the ERM 2 it became important to establish the date of such entrance depending on the stage of reaching the nominal convergence criteria. Another important element is the larger and larger accent that lays on reaching the real convergence criteria before adopting euro.

That is the main issue proposed by Chapter 5: to analyze the complementary of the nominal and real convergence criteria in the CEEC and, especially, in Romania and the convergence level of those countries given the level of the EU-15 countries. In the same time, the paper emphasizes the importance of reaching the real convergence for a smoother entry in the euro-zone. In the conclusions part, the authors proposed some monetary measures, and not only, for achieving the nominal and real convergence in the CEEC and, especially, in Romania.

The economy is a complex system, the dynamics of which can be described by the evolution of the interactions between the actors of the economic process. A measure of the complexity of an economic system must reflect the level of the interdependencies between the component parts of the system. The Input-Output (I-O) tables put into evidence the interconnections between branches as the component parts of the economic system. For the measure of the complexity of the economic system, as it is described by the I-O tables, some different indicators where been proposed, but the most adequate for our approach is the **complexity indicator** which takes in consideration the **network effect** (more connections - more complexity) and the **(inter)dependency effect** (which measures the degree in which the behavior of each component of the system is determined either by the internal connections or by the relations with other parts of the economic system). The multipliers of the I-O tables measures the effects determined by the branches interdependencies when a change of occur in any of the parts of the economic system. In an I-O analysis many different multipliers where been defined, adequate for various analytical instances.

Chapter’s 6 scope is to measure the **complexity** of Romanian economic system determined by the branches interdependencies and to envisage the phenomenon evolution. The **novelty** consist in this approach of the dynamic and complexity of the economic system in a time of great fluctuations of economic indicators (lack of balance and critical points), using concepts and models from physics, linked to the emergence of new properties inside complex systems (differing from system’s synergy).

In recent years, the adoption of flat tax by several countries continues to attract considerable attention, both empirical and theoretical part. Chapter 7 evaluates the main effects of the implementation of flat tax system in Romanian economy. The authors demonstrate, based on the empirical evidence, that their effects in the economy are ambiguous. If accompanying measures are not going to be enforced, the introduction of the flat rate of 16% in Romania will lead to unsustainable budgetary and current account deficits and inflationist pressures.

There are no signals of Laffer-type behavioral responses consolidating budgetary revenue increases from the tax cut. The flat tax favors the workers with big salaries and big and financially solid companies (which, mainly "export" the profit), increasing progressivity and deepening social polarization. It will attack the fragile macroeconomic stability. It is uncertain if it will lead to the increase of the degree of employment, having in view the fact that the contributions to the social insurances have a very high level. The evidence shows that the adoption of flat tax has not resolved the problem of reduction the informal labor market. In the first part of this paper, the authors conclude that the available data do not to make a relevant distinction between the effects of tax reform and other contextual factors.

In the second part, considering that the fiscal policy can play a key role in fostering real convergence, the authors propose measures to increase the long-term sustainability of public finances in Romania, by complying with the Stability and Growth Pact. Romania should have chosen to continue what it was confirmed to be a valid element of the economic evolution towards a European standard (progressive fiscal system). On the other side, to optimize the impact of the flat tax reform, the authors suggest accompanying measures – both fiscal reforms, in particular on the expenditure side, and structural reforms.

In Chapter 8, the authors study an endogenous model of growth cycle in a Keynesian setting. The authors integrate Kaldor (1957)'s model of endogenous growth, that relates the rate of technical progress to the rate of investment by using the 'technical progress function', into Goodwin (1967)'s model of endogenous business cycle that uses a kind of Phillips curve. Our model is an endogenous model in the triple senses. Namely, the 'natural' rate of growth itself as well as 'actual' rate of growth fluctuates endogenously around the endogenously determined equilibrium rate in the course of the endogenous business cycle. Spirit of our model is Keynesian in the sense that full employment of labor is not assumed and the rate of employment as well as the rate of utilization of capital stock is an endogenous variable. This aspect is a notable characteristic of our model which distinguishes it from the usual neoclassical endogenous growth models which assume the full employment of all productive resources including labor. By using such an analytical framework, the authors investigate the impact of the changes of some critical parameter values on the macroeconomic performance theoretically as well as numerically. The authors also refer to the relevance of our model to the analysis of actual macroeconomic problems, in particular, the analysis of the Japanese economy in the 1980s – the 2000s.

Macro price shocks are a concern to the general public and policy makers because these price movements intensify inflationary pressures, increase the uncertainty confronting farmers and agribusiness firms with significant impacts on food security. The current spike in global food prices exceeds all previous records. Food prices increased 2-38 percent per year in Africa during the past three decades, whilst the price of crude oil has more than quadrupled since 2002. The food price increases are driven by several factors, including drought, low food stocks in global market, speculation, declining trends in agricultural production in low-income countries, increasing biofuel production, food demand and income growth in emerging markets. Food accounts for up to 73 percent of household expenditures in Africa. The combination of climate change, rising food and input prices, poor performance of domestic agriculture sector, and high population growth rates is creating significant food shortages in some African countries. Food is imported to bridge the gap between domestic demand and production, thereby, protecting consumption levels. Sub-Saharan Africa currently imports approximately 71.0 percent of its total food supply. Chronic hunger depletes people's ability for increased productivity and increases their susceptibility to diseases. Indeed, whilst high food prices may be good for farmers and food exporters, they could exercabate food insecurity for the poor who may not have access to food at affordable prices. In Chapter 9 the authors use five decades of annual data and a vector error correction model to examine the relationship between macro price shocks and nutrition in Africa. The model identifies and evaluates the strength of the linkages between variables which enables us to trace the dynamic responses in the system at different time horizons in reaction to specific shocks. The empirical results suggest that shocks to food prices, exchange rates and trade policies have significant impacts on nutrition, real incomes and per capita domestic food production. There are strong feedbacks amongst the variables and rapid adjustment towards long-run equilibrium levels after a shock in the system. This information is useful for the formulation of appropriate policies to solve the growing food problems in Africa.

Chapter 10 aims to demonstrate that a green measure of GDP, or what might be better termed as a measure of Hicksian national income, should be included in the formal system of national accounts primarily as an alternative or satellite indicator to GDP. The reason for this is that GDP overstates the national product available for consumption yet is increasingly deployed as a guide to the prudent conduct of national governments (Daly 1989). Since, as the authors aim to show, Hicksian national income constitutes a more appropriate measure of the sustainable national product available for consumption, it can be used as an indispensable tool for both policy-makers and international donor organisations to design the policies required to achieve sustainable development (SD). Although specific policies to achieve SD is not the main aim of this chapter, the results to be later revealed make it possible to: (a) shed some useful light on the nature of Cambodia's recent economic development — a country still recovering from past wars and the devastating impact of the Pol Pot dictatorship; (b) identify what factors have contributed most to the fall or the lack of increase in Cambodia's Hicksian national income; and (c) outline the nature of the policies that are necessary to increase Cambodia's sustainable national product without having to involve the depletion of its income-generating capital.

To achieve its aims, this chapter is organised as follows. Section 2 provides a brief historical overview of Cambodia's social and economic development. In Section 3, a theoretical and empirical overview of green GDP is outlined. In this section, the inadequacies of GDP as a measure of national income are exposed as is the theoretical basis behind the use of green GDP as an alternative measure of sustainable national income. Section 4 presents the methodology employed to calculate the environmental costs and defensive and rehabilitative expenditures associated with the growth of Cambodia's national product. In Section 5, the various costs are deducted from GDP to reveal the Hicksian national income of Cambodia for the period 1988 to 2004. Some general conclusions are outlined in Sections 6 and 7.

Chapter 11 tries to further investigate the factors behind a financial crisis using a large sample of countries in the period 1981 to 1999.

Most methods applied in other researches to analyse financial instability are techniques of statistical nature and the variables employed in these models do not usually satisfy statistical assumptions, what complicates the analysis.

In order to avoid these inconveniences, the authors have approached this problem by applying two algorithms coming from the Machine Learning field, the C4.5 and PART algorithms. By means of theses algorithms the authors will try to analyse the role of a set of macroeconomic and financial variables, both quantitative and qualitative, in explaining banking crises.

Results indicate a high performance of the machine learning techniques, what shows that these methods can be a useful tool to evaluate financial instability and a competitive alternative or complement to existing models related with this kind of problems.

The objective of Chapter 12 is to understand which functions of the financial system are performed well and which ones are not performed well while the credit boom in Bulgaria is unfolding. Bulgaria is interesting because it presents a challenging environment to lenders, similar to that in other transition and developing countries. This allows us to investigate which functions are hampered most by weaknesses in the institutional environment. Institutions have been singled out as some of the most important determinants of financial development (La Porta et. al. 1998) but it is not clear which of the channels through which finance affects growth are blocked most severely when institutions are weak.

Using interviews to address these issues is not the standard practice in the literature and understandably so - interview data are difficult to collect, and the end product is relatively noisy information from only one country at a given point in time. Yet, interviews provide valuable insights that can complement earlier studies that use quantitative data.

SHORT COMMUNICATIONS

In: Emerging Topics in Macroeconomics
Editor: Richard O. Bailly, pp. 3-11
ISBN: 978-1-60692-093-0

LABOUR SUPPLY MODELLING IN SEQUENTIAL COMPUTABLE GENERAL EQUILIBRIUM-MICROSIMULATION MODELS

Nicolas Hérault
Melbourne Institute of Applied Economic and Social Research
The University of Melbourne, Australia

Abstract

The combination of a computable general equilibrium model (CGE) and a microsimulation (MS) model is a novel approach designed to build a bridge between macro- and microeconomic analyses. This approach allows for the assessment of the effects of macroeconomic policies on microeconomic issues such as income distribution. However, the behavioural microsimulation models used for this purpose in the literature make no distinction between full-time and part-time work but only for the choice of a sector/occupation. Labour supply modelling has the potential to improve the quality of policy evaluations based on this type of models by allowing for different levels of labour supply responses by individuals. This chapter discusses the extension of a CGE-MS model applied to South Africa to allow for labour supply responses. An application shows that although labour supply modelling has the potential to bring additional insights regarding income distribution issues, its contribution is limited in the case of the evaluation of trade liberalisation policies. The impacts are driven by changes in employment levels, while labour supply responses are fairly small and do not significantly affect the results. Nonetheless, this should not be seen as a general case against labour supply modelling in CGE-MS models since the study of trade liberalisation in South Africa cannot be generalised to other policies or other countries.

Introduction

The combination of a computable general equilibrium (CGE) model and a microsimulation (MS) model is a novel approach designed to build a bridge between macro- and microeconomic analyses. The approach was first applied by Robilliard *et al.* (2001). The authors combined a macro-oriented CGE model and an MS model in a sequential approach. The top-down model allows for the examination of the microeconomic effects (on poverty and inequality) of macroeconomic policies. In this approach, the CGE model is used to simulate the changes at the macroeconomic level after the policy change. In a second step, the changes estimated by the CGE model are passed on to the MS model. The latter is based on one or more household surveys so that the effects of the policy change can be assessed at the household level.

Following Robilliard *et al.* (2001) numerous CGE-MS models have been developed. Although, the behavioural MS models used for this purpose in the literature generally allow for changes in labour force participation, they make no distinction between full-time and part-time work. This chapter explores the use of labour supply modelling in CGE-MS models. The approach has the potential to improve the quality of policy evaluations by allowing for different levels of labour supply responses by individuals.

First, this chapter discusses some aspects of the flourishing CGE-MS literature. Second, the South African CGE-MS model is briefly presented. In addition, the approach used to allow for labour supply responses is explained. Third, the impact of trade liberalisation on households is assessed using the new model. The results are compared with the results generated without labour supply modelling. Finally, the last section concludes and briefly explores the potential for labour supply modelling in CGE-MS models.

The CGE-MS Literature

CGE models have been widely used in the past 30 years. Many have been applied to assess the impact of macroeconomic policies in developing countries, and more particularly the impact of trade liberalisation policies.[1] However, because representative household groups rather than individual information are used, CGE models are unable to provide detailed analysis of the effects on income distribution.[2] To address this issue Robilliard *et al.* (2001) combine a CGE model and an MS model in a sequential approach. This approach was followed by many since it allows for the effects of macroeconomic policies to be assessed at the household level.[3]

Households are affected by macroeconomic policies through changes in prices and taxes. However, the largest effects are usually driven by the changes occurring in the labour market in terms of earnings and employment. Hence, it is particularly important to account for changes in the labour market when the focus is on income distribution. This is the reason why, in a second step, Robilliard *et al.* (2001) pass the macroeconomic changes on to the MS model. Given any change in the macroeconomic structure of the economy predicted by the CGE model, the MS model predicts how household incomes are affected while accounting for individual heterogeneity.

---

[1] For a literature review of CGE models applied to developing countries see Gunning and Keyser (1997). A set of recent applications can be found in Hertel and Winters (2006).

[2] For a discussion of this issue see Savard (2004).

[3] Alternatively, the microsimulation model can be integrated to the CGE model following the approach introduced by Decaluwé *et al.* (1999) and Cogneau (1999).

Two types of MS models are used in this context. A distinction can be made between behavioural and non-behavioural models. In the latter, individual behaviour is assumed to remain the same before and after the policy change. Behavioural models are generally more complex in that they allow individuals to adjust their behaviour in response to the simulated policy change. A wide variety of behaviours can potentially be modelled depending on individual or household characteristics (see O'Donoghue, 2001). Nevertheless, only a limited range of behavioural responses has been considered in the CGE-MS literature.

In the model developed by Robilliard *et al.* (2001), individual occupational choices are endogenous. Each individual has to choose from three alternatives: being inactive, being a wage-worker, or being self-employed. The model developed by Bussolo and Lay (2003) is similar but there is a fourth option of being both wage-employed and self-employed (in rural areas only). In Bussolo *et al.* (2005) the model simulates transition from agriculture to non-agriculture. The MS model built by Hérault (2006) distinguishes five occupational choices: inactive, unemployed, subsistence agricultural worker, informal worker and formal worker. The recent CGE-MS model by Thierfelder *et al.* (2007) is the most comprehensive as its occupational component contains 16 choices combining formal and informal work, three skill levels and three sectors (agriculture, industry and services). However, it is assumed in all these models that individuals' choices are restricted to working full-time or not working. This is a clear simplification of the labour market as the number of hours worked generally differs substantially across sectors and across demographic groups.

Overview of the Modelling Approach

Extension of the South African Microsimulation Model

The starting point of this project is the CGE-MS model developed by Hérault (2006). Five labour market sectors are considered: inactive, unemployed, subsistence agriculture, informal and formal sectors. Moreover, 'yes/no' choices about labour mobility across sectors are imposed. In this chapter, the model is extended to include different levels of labour supply. The aim it to allow individuals to choose from a range of discrete hours points so changes in labour market choices can be modelled together with the intensity of labour supply.

The MS model developed by Hérault (2006) is extended to nine occupational choices instead of the five initial choices. Three broad activities are now distinguished: (i) inactivity/unemployment (ii) informal work and (iii) formal work. In addition, four hours points are considered in both the formal and the informal sectors. The probabilities of being in each of the nine categories are derived from an estimated implicit utility function based on a mixed logit specification.[4] The utility associated with each category is a linear function of a set of individual characteristics, which include household and individual characteristics such as skill level, age, education, province of residence, racial group or household size. These variables are interacted with a dummy variable indicating the sector (formal or informal) and the number of hours worked. Following Creedy et al. (2002, p. 71), they are also interacted with the predicted net earnings associated with each hours point. This utility function is estimated separately for four demographic groups: single women, partnered

[4] The estimated coefficients are available on request from the author.

women, single men and partnered men. Therefore, the model explicitly accounts for the heterogeneity of labour supply responses across demographic groups. This specification also allows the use of a different set of discrete hours points for each demographic group and each sector. This characteristic of the model seems particularly relevant since the number of hours worked differs substantially across sectors and across demographic groups (see Hérault, 2007).

It is worth mentioning that the estimation of such a labour supply model is problematic in the South African context. The theory behind labour supply modelling assumes that unemployment is voluntary. Hence, unemployed and inactive are grouped together and are both assumed to provide zero hours of work. This is a very strong assumption in the South African case given the very high level of unemployment. In this context, the estimated coefficients lead to inconsistencies with utility maximization for some particular observations.[5]

A regression model is used to predict individual hourly wage rates in the formal and the informal sectors. The regression model, like the selection model, is estimated separately for the four demographic groups. Chow tests confirm that the coefficients are significantly different in the four demographic groups. The results of the regression model are used to assign predicted formal and informal hourly wage rates to all individuals in the sample. However, for those individuals who were formal or informal workers at the time of the survey, the corresponding observed hourly wage rates are used for the sector in which they are observed. Net earnings are predicted by combining hourly wage rates with the number of hours worked. In turn, this variable influences the labour market choices made in the selection model.

Linking CGE and Microsimulation Modelling

The top-down approach used to link the CGE model to the MS model is described in detail in Hérault (2006). It relies on using a CGE and an MS model in a sequential way: first, the CGE model is run, followed by a second step in which the changes in some selected variables are passed on to the MS model.

The first stage consists of running the CGE model to simulate the policy change. The model returns the new macrostructure of the economy after the ‘shock’. In the context of the top-down approach, four sets of variables are of particular interest: prices, direct tax rates, returns from capital and labour, and employment levels. In the second stage, the changes in these variables are passed on to the MS model.

Micro–macro consistency equations, along with the direct transmission of prices and direct tax rates, ensure that changes in prices, direct tax rates, earnings from wages and salaries, returns from capital, and employment levels are transmitted from the CGE to the MS model. Given that the MS model now includes a labour supply component, changes in employment levels can be communicated to the MS model in terms of changes in the total

[5] The estimated utility function was found to increase with the number of hours worked at the observed labour supply point for slightly less than 12 per cent of the individuals included in the model. The addition of two dummy variables (interacted with hours of work) distinguishing workers with high and low hourly wage rates was crucial in reducing this percentage, as well as a dummy variable indicating the presence of an Old Age and War pension recipient in the household.

number of hours worked. The assumption that all new jobs are full-time jobs can be relaxed.

Given any change in the macroeconomic structure of the economy predicted by the CGE model, the MS model predicts how individual agents modify their behaviour and how their incomes are affected, while accounting for individual heterogeneity. Therefore, it provides us with an updated picture of the economy at the microeconomic level taking into account the simulated changes in macroeconomic policies. Top-down models have the advantage of avoiding the use of representative agent assumptions, while accounting for general equilibrium effects.

Results and Analysis

The extended CGE-MS is used to assess the impact of trade liberalisation in South Africa. The aim is to compare the results with previous work by Hérault (2007), in which the same model is used except that it does not include a labour supply module. Trade liberalisation is simulated by the complete removal of import tariffs.

Macro Results from the CGE Model

The CGE results presented in Table 1 reveal a positive but limited impact of trade liberalisation on the South African economy. The lowering of import prices causes a shift towards imported goods and away from domestic production. As domestic and import prices decrease, the real exchange rate depreciates, which promotes exports and contributes to a marginal increase in the current account balance. Regarding the government account, the contraction in nominal government revenue following the decrease in prices is partly offset by the implied decreasing cost of public expenditures and the expansion in gross domestic product. In this context, an increase of 0.8 percentage points in the rate of direct taxes is sufficient to generate an additional revenue of 8 billion rand, which compensates for the loss of import duties. Both government savings and investment are fixed, so that the balance between savings and investment can only be marginally affected by the simulated policy.

Falling prices imply a decrease in the cost of investment and there is a diminution of foreign savings generated by the slightly improving trade balance. The drop in the consumer price index (CPI), resulting from falling import prices, causes nominal earnings of skilled and low-skilled workers to fall because the latter are indexed to the CPI. Therefore, their relative competitiveness improves, which results in a downward pressure on unemployment. Employment goes up by 0.7 and 0.6 per cent for low-skilled and skilled labour respectively. The 0.3 per cent economic growth induced by trade liberalisation calls for more use of the two scarce production factors. As a result, real returns of capital and high-skilled labour rise substantially (by 2.1 and 0.4 per cent respectively).

Table 1. CGE Simulation Results from the Elimination of Import Tariffs

	Base values[a]	Percentage change from base year
Real GDP	R888	0.3
CPI	-	-1.6
Real exchange rate	-	-1.2
Nominal exchange rate	-	0.0
Exports (volume)	R249	2.4
Imports (volume)	R225	1.5
Trade balance[b]	5%	0.3
Private savings	R154	-0.3
Government deficit[b]	-2%	0.0
Investment (volume) [b]	15%	0.0
Factor real returns		
Capital	-	2.1
Low-skilled labour	-	0.0
Skilled labour	-	0.0
High-skilled labour	-	0.4
Factor demand		
Low-skilled labour	3.6	0.7
Skilled labour	2.7	0.6
High-skilled labour	1.1	0.2

Note: (a) Values in billions of rand, percentage of GDP and millions of workers (b) Base values are expressed as a percentage of GDP and changes are expressed in percentage points of GDP.

Microsimulation Results

Table 2 presents the effects of trade liberalisation at the household level for the entire South African population, by racial group and for the two specifications of the MS model. The results generated by the former version of the model, which has no labour supply module, are presented on the left part of the table while the results of the extended MS model, which includes labour supply responses, are presented in the right columns.

Overall, the results from the two models are very similar. Trade liberalisation is pro-poor and has a limited dampening effect on inequality. The expansion of formal employment is the main factor contributing to the observed reduction in poverty, while labour supply responses are fairly small and do not significantly affect the results.

Table 2 shows that percentage changes in the total number of formal workers are almost identical to changes in the total number of hours worked, which indicates that changes in labour supply are very small. Although the new formal workers tend to work longer hours than the existing formal workers, the latter essentially do not change their level of labour supply following the policy change. The explanation is that the impact of trade liberalisation on real net hourly wage rates is too small to trigger substantial labour supply responses.

Table 2. Simulation Results from the two alternative specifications of the MS Model (Percentage Change from the Base Values)

	Base values[(a)]	No labour supply module					Labour supply module				
		All	Blacks	Colou-reds	Asians	Whites	All	Blacks	Colou-reds	Asians	Whites
Inactive/unemployed[(b)]	32,574	-0.1	-0.1	-0.2	-0.2	-0.3	-0.1	0.0	-0.1	-0.3	-0.7
Informal workers	3,357	-0.2	-0.2	-0.4	-0.3	0.6	-0.2	-0.3	-0.3	1.0	1.5
Formal workers	7,307	0.6	0.7	0.6	0.5	0.3	0.6	0.5	0.4	0.5	0.9
Total informal hours	145	-	-	-	-	-	-0.3	-0.4	-0.4	0.5	0.8
Total formal hours	343	-	-	-	-	-	0.6	0.5	0.5	0.5	0.9
Informal earnings[(c)]	13,217	1.4	0.9	1.0	0.7	2.6	0.8	0.6	1.3	1.7	-0.5
Formal earnings[(c)]	52,343	-0.2	-0.2	-0.1	0.0	0.0	-0.1	-0.2	-0.1	-0.1	-0.3
Real income per capita[(d)]	11,098	0.6	0.7	0.6	0.5	0.6	0.6	0.5	0.5	0.5	0.7
Headcount Index[(e)]	29.0	-1.0	-1.0	-0.8	0.0	0.0	-0.9	-0.8	-1.8	0.0	0.0
Gini	0.67	-0.1	-0.2	-0.2	-0.2	0.0	-0.1	-0.1	-0.1	-0.2	0.1

Note: (a) Values in thousands of workers and millions of weekly hours worked; (b) Inactive, unemployed and subsistence agriculture; (c) Average real earnings per worker in Rand per year; (d) Average real disposable income per capita in Rand per year; (e) The poverty line is the international \$2/day poverty line (R2,088/year/capita in 2000 prices).

There is only one remarkable difference between the two sets of results. This difference concerns the labour market mobility of whites.[6] The trends observed in the former model regarding the labour market mobility of whites are now more pronounced. Whites have the highest income and education levels. They seem to attach more importance to the number of hours worked than the other racial groups. In this context, the labour supply extension allows for a better modelling of their behaviour. In the new model, the number of alternatives has been extended beyond the ‘yes/no’ choices about labour mobility across sectors, so that these individuals can now also adjust their labour supply. The new informal workers are found to work fewer hours than the existing informal workers and this is particularly true for high-skilled workers. In turn, this explains why the increase in the average informal earnings per capita is now lower. This phenomenon is particularly visible among whites because the share of high-skilled workers is the highest among this racial group.

The increased labour market mobility of whites also explains the slight decline in poverty reduction. The decrease in poverty is driven by the increase in formal employment among blacks. However, because whites appear to benefit more from the expansion of formal employment under the extended model, the share of the new formal jobs accruing to blacks is now slightly lower.

[6] There are also some changes in informal employment and informal earnings for Asians. However, the size of Asian informal employment is too small for these changes to be considered as being significant. Less than 32,000 Asians were employed in the informal sector in South Africa in 2000. The same argument applies to changes in the headcount index for coloureds because extreme poverty affects only about 10 per cent of this racial group.

Conclusion

This chapter discusses the first attempt in the literature to integrate a labour supply component allowing for discrete hours choices into a sequential computable general equilibrium-microsimulation model. In such top-down models, individuals are generally assumed to be mobile across sectors, and especially in and out of employment. However, it is also assumed that individuals' choices are restricted to working full-time or not working. The use of a labour supply module is a step beyond the 'yes/no' choices across sectors. By allowing for different levels of labour supply, the fact that individuals do not necessarily work full-time, when working, can be taken into account. As a result, the fact that a policy change can affect not only labour market participation but also the level of labour supply is explicitly taken into consideration. That is individuals can adjust their labour market sector choices and also their level of labour supply.

An application of this approach to the case of trade liberalisation in South Africa shows the limited additional insights gained through the use of a labour supply module. The results are largely unaffected by the extension of the model to allow for labour supply responses. The explanation is that the impact of trade liberalisation on hourly wage rates is too small to trigger substantial labour supply responses. However, it may be argued that this is specific to the South African case. Indeed, the total amount of collected import duties represented only 3.6 per cent of the South African government revenue in 2000, which is a much lower share than in many other developing countries. As a result, the removal of tariffs only implies a limited loss of revenue for the South African government. In this context, a small increase in income tax rates is sufficient to compensate for the loss of import duties. Labour supply responses are expected to play a more important role in the context of a policy change involving more substantial changes in income tax rates (or in hourly wage rates).

References

Bussolo, M. & Lay, J. (2003). Globalisation and Poverty Changes in Colombia, OECD Development Centre, *Working Paper* No. **226**.

Bussolo, M., Lay, J. & van der Mensbrugghe, D. (2005). Structural Change and Poverty Reduction in Brazil: *The Impact of the Doha Round, Modeling and Policy Impact Analysis (MPIA) Network*.

Cogneau, D. (1999). *Labour Market, Income Distribution and Poverty in Antananarivo: A General Equilibrium Simulation*. mimeo, DIAL, Paris.

Creedy, J., Duncan, A. S., Harris, M. & Scutella, R. (2002). *Microsimulation Modelling of Taxation and The Labour Market: The Melbourne Institute Tax Transfer Simulator*. Cheltenham: Edward Elgar.

Decaluwé, B., Dumont, J.-C. & Savard, L. (1999). Measuring Poverty and Inequality in a Computable General Equilibrium Model. *Working paper* **99-20**, CREFA, Université Laval, Quebec, Canada.

Gunning, J.W. & Keyzer, M. (1997). Applied General Equilibrium Modeling. In: Behrman, J. and Srinivasan, T. N., editors. *Handbook of Development Economics*, Amsterdam: North-Holland, Vol 3.

Hérault, N. (2006). Building and Linking a Microsimulation Model to a CGE Model for South Africa, *South African Journal of Economics*, Vol. 74:1, pp 34-58, March 2006.

Hérault, N. (2007). Trade Liberalisation, Poverty and Inequality in South Africa: A CGE-Microsimulation analysis. *Economic Record*. **83** (262): 317-328.

Hertel, T.W. & Winters, A.L. (2006). *Poverty and the WTO: Impacts of the Doha Development Agenda*. Palgrave Macmillan and the World Bank.

O'Donoghue, C. (2001). *Introduction to the Special Issue on Dynamic Microsimulation Modelling*. London School of Economics, University of Cambridge, Recife, December 15th.

Robilliard, A.-S., Bourguignon, F. & Robinson, S. (2001). *Crisis and Income Distribution: A Micro-Macro Model for Indonesia*, Paper presented at the OECD Development Center Conference, 9-10 December 2002, Paris, France.

Savard, L. (2004). Poverty and Inequality Analysis within a CGE Framework: a Comparative Analysis of the Representative Agent and Micro-Simulation Approaches, Centre interuniversitaire sur le risque, les politiques économiques et l'emploi, *Working Paper* **04-12.**

Thierfelder, K., Robinson, S., Korman, V., Kearney, M., Go, D.S. & Essama-Ns (2007). Economy-wide and Distributional Impacts of an Oil Price Shock on the South African Economy, *Policy Research Working Paper Series* **4354**, The World Bank.

Reviewed by Associate Professor Guyonne Kalb, Melbourne Institute of Applied Economic and Social Research, University of Melbourne, Australia.

In: Emerging Topics in Macroeconomics
Editor: Richard O. Bailly, pp. 13-27
ISBN: 978-1-60692-093-0

NOMINAL AND REAL EXCHANGE RATE FLUCTUATIONS OF YEN-DOLLAR RATES

Shigeyuki Hamori*
Faculty of Economics, Kobe University
2-1 Rokkodai, Nada-Ku, Kobe 657-8501, Japan

Abstract

This paper uses a long-run structural VAR approach to study the sources of real exchange rate fluctuations of yen-dollar rates. We begin by identifying two types of macroeconomic shock (real and nominal), in order to reveal the sources of movements in real exchange rates. The evidence presented indicates that real shocks play a dominant role in explaining the fluctuations of the real exchange rate. Next, we extend the model to identify three types of macroeconomic shock (real supply, real demand, and nominal), again to reveal the sources of movements in real exchange rates. The evidence from the extended model also indicates that real shocks (real demand and real supply) play a dominant role in explaining the fluctuations of the real exchange rate.

1. Introduction

When modeling exchange rates and studying economic policy, we must measure the relative importance of permanent and temporary shocks on exchange rates. Disequilibrium models of exchange rate determination (e.g., Dornbusch, 1976) mainly attribute variations in real and nominal exchange rates to nominal disturbances of a type likely to have only transitory effects on real exchange rates. Equilibrium models (e.g., Stockman, 1987) rely on permanent, real shocks to explain movements in real and nomina1 rates. Isolating these sources of variation can assist policymakers in determining the extent of excess variability in exchange rates (Lastrapes, 1992). A thorough understanding of the sources underlying real exchange rate fluctuations is also crucial, given that the real exchange rate reflects the performance and

* E-mail address: hamori@econ.kobe-u.ac.jp

competitiveness of an economy. Movements of the real exchange rate may also influence inflation, output, and the balance of payments.

This paper investigates the sources of real exchange rate fluctuations between the US dollar and Japanese yen. By constructing a structural VAR model to analyze the sources of real exchange rate fluctuations, we try to identify the dynamics and forces driving the real exchange rate variations since the beginning of the 1980s. Several studies have attempted to investigate the sources of real exchange rate fluctuations since Blanchard and Quah (1989) published their influential work based on a bivariate structural VAR model for output and unemployment. The paper by Bayoumi and Eichengreen (1992) was among the first to analyze exchange rate variations using the Blanchard and Quah (1989) approach. They distinguish between supply shocks and demand shocks by assuming that the former have permanent effects whereas the latter have only temporary effects. Their empirical results, for the G-7 countries, indicate that the shift from the Bretton Woods system of pegged exchange rates to the post Bretton Woods float can be explained by a modest increase in the cross-country dispersion of supply shocks

Lastrapes (1992) carries out a similar analysis for six industrialized countries from 1973 to 1989. He identifies two types of structural disturbance, nominal shocks and real shocks, with the restriction that the former has no long-run impact on the real exchange rate. His results indicate that real shocks account for the major part of both real and nominal exchange rate fluctuations for all six of the countries analyzed.

Enders and Lee (1997) decompose real and nominal exchange rate movements into the components induced by real and nominal factors over the period from 1973 to 1992. They find that nominal shocks have a minor effect on the real and nominal bilateral exchange rates of the Canadian, Japanese, and German currencies against the US dollar.

Chowdhury (2002) investigates sources of fluctuations in the real and nominal US dollar exchange rates of selected emerging market economies by decomposing the exchange rate series into stochastic components induced by real and nominal factors. His analysis focuses on the dynamic effects and relative importance of real and nominal shocks in explaining the behavior of these exchange rates.

Clarida and Gali (1994) extend the bivariate VAR model to the trivariate VAR model and identify three types of structural disturbance: real aggregate supply shocks (those capable of influencing the levels of all three variables in the long run), real aggregate demand shocks (those with no long-run impact on the real output level), and nominal shocks (those that affect only the price level in the long run). According to their empirical analysis for four industrialized countries (Germany, Japan, the UK, and Canada) over the floating period from 1973 to 1992, nominal disturbances explain a substantial amount of the variance in the real exchange rates against the dollar in two of the countries (Germany and Japan: 41% of the variance of the change in the dollar-deutschmark real exchange rate and 35% of the variance of the change in the dollar-yen real exchange rate), whereas the real exchange rate fluctuations in the other two countries are mainly driven by real demand shocks.

According to a five-dimensional structural VAR analysis on a long-term data set (from 1889 to 1992) for the US and UK by Rogers (1999), monetary shocks typically account for nearly one-half of the forecast error variance of the real exchange rate over short horizons. The evidence indicates that monetary shocks are generally very important for real exchange rate movements. This result contradicts the findings by Lastrapes (1992), Clarida and Gali (1994), and Enders and Lee (1997).

These studies set a benchmark for researchers seeking to explain real exchange rate movements. In this study we begin by constructing a bivariate structural vector autoregressive (SVAR) model, following the example of Enders and Lee (1997), in order to assess the relative importance of two types of shocks: real shocks and nominal shocks. The structural VAR decomposition implies that (i) real shocks are expected to influence real and nominal exchange rates in the long-run; and (ii) nominal shocks are expected to have no long-run impact on real exchange rates. Next, we extend the model to a trivariate SVAR model, in order to assess the relative importance of three types of shocks: aggregate supply shocks, aggregate demand shocks, and nominal demand shocks. The structural VAR decomposition implies that (i) only supply shocks are expected to influence the relative output level in the long run; (ii) both supply and demand shocks are expected to influence the real exchange rate in the long run; and (iii) monetary shocks are expected to have no long-run impact on the relative output levels or the real exchange rate.

The paper is organized as follows. Section 2 explains the data used for empirical analysis. Section 3 shows the empirical techniques and empirical results for the bivariate system. Section 4 shows the empirical techniques and empirical results for trivarite system. Some concluding remarks are summarized in section 5.

2. Data

The data are taken from the *International Financial Statistics* (IFS) of the International Monetary Fund (IFM). The empirical analysis is carried out using monthly observations from January 1980 to April 2006. We use the exchange rate, consumer price indexes of Japan and the United States, and industrial production indexes (seasonally adjusted value) of Japan and the United States. The exchange rate is expressed per US dollar. The log of the nominal exchange rate (e_t), the log of the real exchange rate (r_t), and the log of the relative output (y_t) are used in the empirical analysis. The log-level real exchange rate, r_t, may be expressed as follows:

$$r_t = e_t + p_t^{US} - p_t^{JAPAN}, \tag{1}$$

where p_t^{JAPAN} is the logarithm of the price level in Japan); p_t^{US} is the logarithm of the price level in the United states. The real exchange rate thus measures the relative price of goods in Japan in terms of US goods. $y_t (= y_t^{JAPAN} - y_t^{US})$ is the difference between the real income in Japan and the real income in the United States. Figure 1 shows the movements of real and nominal exchange rates over the period. The value of the Japanese currency appreciated sharply against the US dollar during the 1980's.

This paper uses the augmented Dickey-Fuller (ADF) test (Dickey and Fuller, 1979) to test for the presence of a unit root for each variable in the univariate representations of the log of the nominal exchange rate (e_t), the log of the real exchange rate (r_t), and the log of the relative output (y_t). The null hypothesis of a unit root is not rejected at conventional significance levels for any of the variables. The null hypothesis of a unit root is rejected, however, for the first-differenced for each variable. Thus, e_t, r_t and y_t are each found to be I(1) series.

3. Bivariate System

3.1. Empirical Techniques

The bivariate VAR model developed by Enders and Lee (1997) will serve as a starting point for our discussion. Consider the following infinite-order vector moving average (VMA) representation:

$$\Delta x_t = C(L)\varepsilon_t , \tag{2}$$

where L is a lag operator, Δ is a difference operator, $\Delta x_t = [\Delta r_t, \Delta e_t]'$ is a (2×1) vector of endogenous variables, and $\varepsilon_t = [\varepsilon_{r,t}, \varepsilon_{n,t}]'$ is a (2×1) vector of structural shocks with covariance matrix Σ. The error term can be interpreted as real shocks and nominal shocks. We assume that the structural shocks have no contemporaneous correlation or autocorrelation. This implies that Σ is a diagonal matrix.

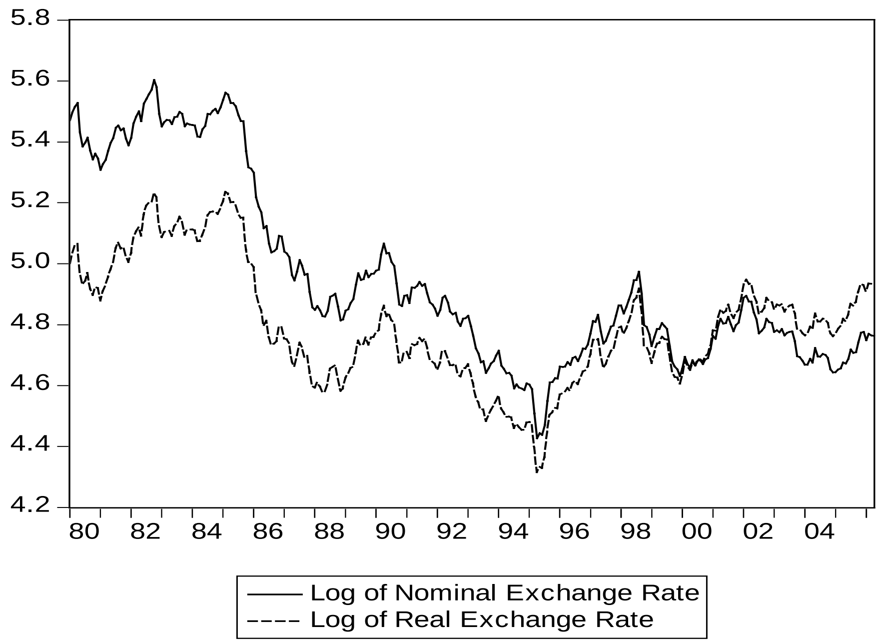


Figure 1. Real and Nominal Exchange Rate Movements.

To implement the econometric methodology, we need to estimate the following finite-order VAR model:

$$[I-\Phi(L)]\Delta x_t = u_t \tag{3}$$

where $\Phi(L)$ is a finite-order matrix polynomial in the lag operator and u_t is a vector of disturbances. If the stationarity condition is satisfied, we can transfer equation (3) to the VMA representation,

$$\Delta x_t = A(L)u_t \tag{4}$$

where $A(L)$ is a lag polynomial.

Equations (2) and (4) imply a linear relationship between ε_t and u_t:

$$u_t = C_0\varepsilon_t \tag{5}$$

C_0 in Equation (4) is a 2×2 matrix that defines the contemporaneous structural relationship that needs to be identified for the vector of structural shocks ε_t in order to be recovered from the estimated disturbance vector u_t. Four parameters are required to convert the residuals from the estimated VAR into the original shocks that drive the behavior of the endogenous variables. Three of the four are given by the elements of $\Sigma = C_0C_0'$, hence we need to add one more identifying restriction. Blanchard and Quah (1989) suggest that economic theory can be used to impose this restriction. By the methodology of Enders and Lee (1997), we thus impose an additional restriction on the long-run multipliers while freely determining the short run dynamics. This restriction is written as follows:

Assumption: Nominal (monetary) shocks have no long-run impact on the real exchange rate;

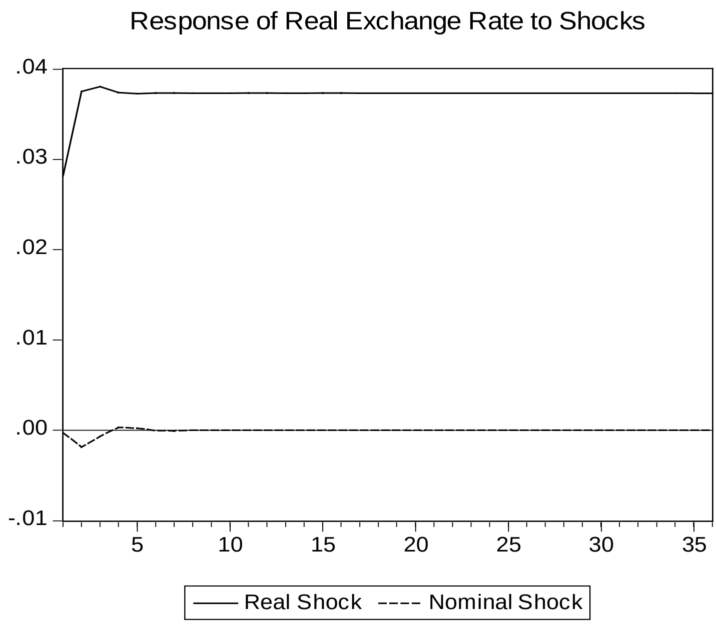
The long-run representation of equation (2) can be written as:

$$\begin{bmatrix}\Delta r_t\\ \Delta e_t\end{bmatrix} = \begin{bmatrix}C_{11}(1) & C_{12}(1)\\ C_{21}(1) & C_{22}(1)\end{bmatrix}\begin{bmatrix}\varepsilon_{r,t}\\ \varepsilon_{n,t}\end{bmatrix} \tag{6}$$

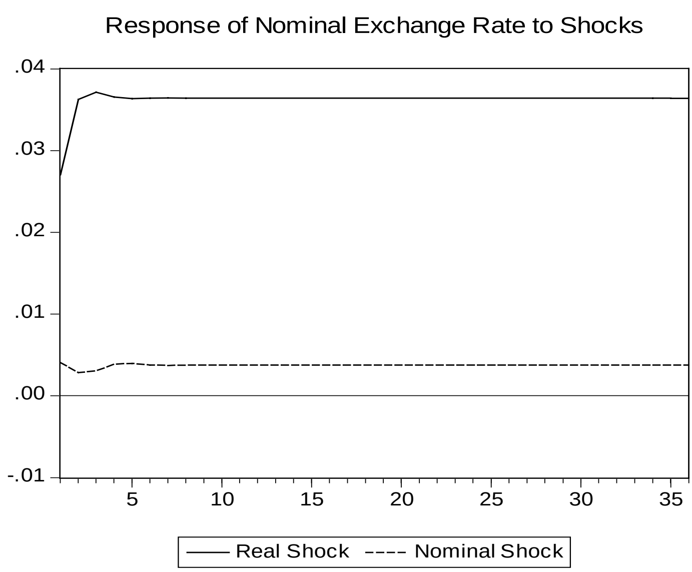
where $C(1) = C_0 + C_1 + C_2 + \cdots$ are long-run multipliers of the structural VAR (long-run effect of ε_t on Δx_t). By the method of Enders and Lee (1997), we stipulate that the long-run multiplier C_{12} is equal to zero, thus making $C(1)$ lower triangular matrix.

3.2. Empirical Results

Using the SBIC (Schwarz Bayesian Information Criterion) to choose the optimal lag length of VAR in empirical analysis, we find that the VAR(2) model is the most appropriate. We present the impulse responses of each of the variables to one standard deviation in each of the fundamental shocks.



(a) Response of Real Exchange Rate



(b) Response of Nominal Exchange Rate

Figure 2. Accumulated Impulse Responses (Bivariate Model).

Figure 2 illustrates the impulse response functions of real and nominal exchange rates to one-standard-deviation structural shocks. Impulse responses are useful for assessing the signs and magnitudes of responses to specific shocks by revealing the dynamic effects of each shock. The results broadly coincide with conventional models of the open economy. Panel (a) shows that real shocks lead to a permanent increase in the real exchange rate. Nominal shocks have smaller effects and leave the long-run real exchange rate unaffected in accordance with the long-run restrictions imposed. Panel (b) demonstrates that a real shock leads a persistent increase in the nominal exchange rate. A nominal shock is associated with a permanent increase in the nominal exchange. Our results show that a real shock induces a jump in the real and nominal rate of nearly the same magnitude. Hence, there is little movement in the price ratio in response to a real shock. We also find that nominal shocks have no long-run effects on real rates. Given that nominal shocks have no effect on relative prices, they can be assumed to affect prices by an equal but opposite amount. These results are consistent with those of Enders and Lee (1997).

Table 1. Forecast Error Variance Decomposition (Bivariate System)

(a) Real Exchange Rate

Horizon (month)	**Real Shock (%)**	**Nominal Shock (%)**
1	99.991	0.009
3	99.539	0.461
6	99.414	0.586
9	99.413	0.587
12	99.413	0.587
24	99.413	0.587
36	99.413	0.587

(b) Nominal Exchange Rate

Horizon (month)	**Real Shock (%)**	**Nominal Shock (%)**
1	97.795	2.205
3	97.849	2.151
6	97.765	2.235
9	97.765	2.235
12	97.765	2.235
24	97.765	2.235
36	97.765	2.235

To shed light on the question of the sources of real exchange rate fluctuations, we then calculate the forecast error variance decompositions. Variance decomposition is a useful technique to evaluate the relative importance of such shocks to the system. Table 1 shows the fraction of the forecast error variance that can be attributed to each shock at different horizons in the model for each variable. Real shocks account for more than 99% of the variance in real exchange rate throughout the estimation horizons. The less then 1% of variance remaining is attributable to nominal shocks. The estimates imply that real shocks explain a substantial amount of the variance of the real exchange rate.

Forecast error variance decompositions for the variations in the nominal exchange rate show that real shocks explain most of the movements in the nominal exchange rate. Specifically, real shocks account for more than 97% of the variance in the nominal exchange rate movement. Nominal shocks, meanwhile, contribute less than 3% to the fluctuations of the nominal exchange rate. To summarize, real shocks explain most of the forecast error variance of the changes in the real and nominal exchange rates.

4. Trivariate System

4.1. Empirical Techniques

Consider the following infinite-order vector moving average (VMA) representation:

$$\Delta x_t = C(L)\varepsilon_t \tag{7}$$

where L is a lag operator, Δ is a difference operator, $\Delta x_t = [\Delta y_t, \Delta r_t, \Delta e_t]'$ is a (3×1) vector of endogenous variables, and $\varepsilon_t = [\varepsilon_{s,t}, \varepsilon_{d,t}, \varepsilon_{n,t}]'$ is a (3×1) vector of structural shocks with covariance matrix Σ. The error term can be interpreted as relative supply shocks, relative real demand shocks, and relative nominal shocks. We assume that the structural shocks have no contemporaneous correlation or autocorrelation. This implies that Σ is a diagonal matrix.

To implement the econometric methodology, we need to estimate the following finite-order VAR model:

$$[I - \Phi(L)]\Delta x_t = u_t \tag{8}$$

where $\Phi(L)$ is a finite-order matrix polynomial in the lag operator and u_t is a vector of disturbances. If the stationarity condition is satisfied, we can transfer equation (8) to the VMA representation,

$$\Delta x_t = A(L)u_t \tag{9}$$

where $A(L)$ is a lag polynomial.

Equations (7) and (9) imply a linear relationship between ε_t and u_t:

$$u_t = C_0\varepsilon_t \tag{10}$$

C_0 in Equation (9) is a 3×3 matrix that defines the contemporaneous structural relationship that needs to be identified for the vector of structural shocks ε_t in order to be recovered from the estimated disturbance vector u_t. Nine parameters are required to convert the residuals from the estimated VAR into original shocks that drive the behavior of the endogenous variables. Blanchard and Quah (1989) suggest that economic theory can be used to impose these restrictions. By the methodology of Clarida and Gali (1994), we thus impose three additional restrictions on the long-run multipliers while freely determining the short-run dynamics. These three restrictions are as follows:

Assumption (i) Nominal (monetary) shocks have no long-run impact on the levels of output;
Assumption (ii) Nominal (monetary) shocks have no long-run impact on the real exchange rate;
Assumption (iii) Real demand shocks have no long-run impact on the levels of output.

The long-run representation of equation (6) can be written as :

$$\begin{bmatrix} \Delta y_t \\ \Delta r_t \\ \Delta e_t \end{bmatrix} = \begin{bmatrix} C_{11}(1) & C_{12}(1) & C_{13}(1) \\ C_{21}(1) & C_{22}(1) & C_{23}(1) \\ C_{31}(1) & C_{32}(1) & C_{33}(1) \end{bmatrix} \begin{bmatrix} \varepsilon_{s,t} \\ \varepsilon_{d,t} \\ \varepsilon_{n,t} \end{bmatrix} \tag{11}$$

where $C(1) = C_0 + C_1 + C_2 + \cdots$ are long-run multipliers of the structural VAR (long-run effect of ε_t on Δx_t). By the method Clarida and Gali (1994), we stipulate that the long-run multipliers C_{12}, C_{13} and C_{23} are equal to zero, thus making the matrix $C(1)$ a lower triangular matrix.

4.2. Empirical Results

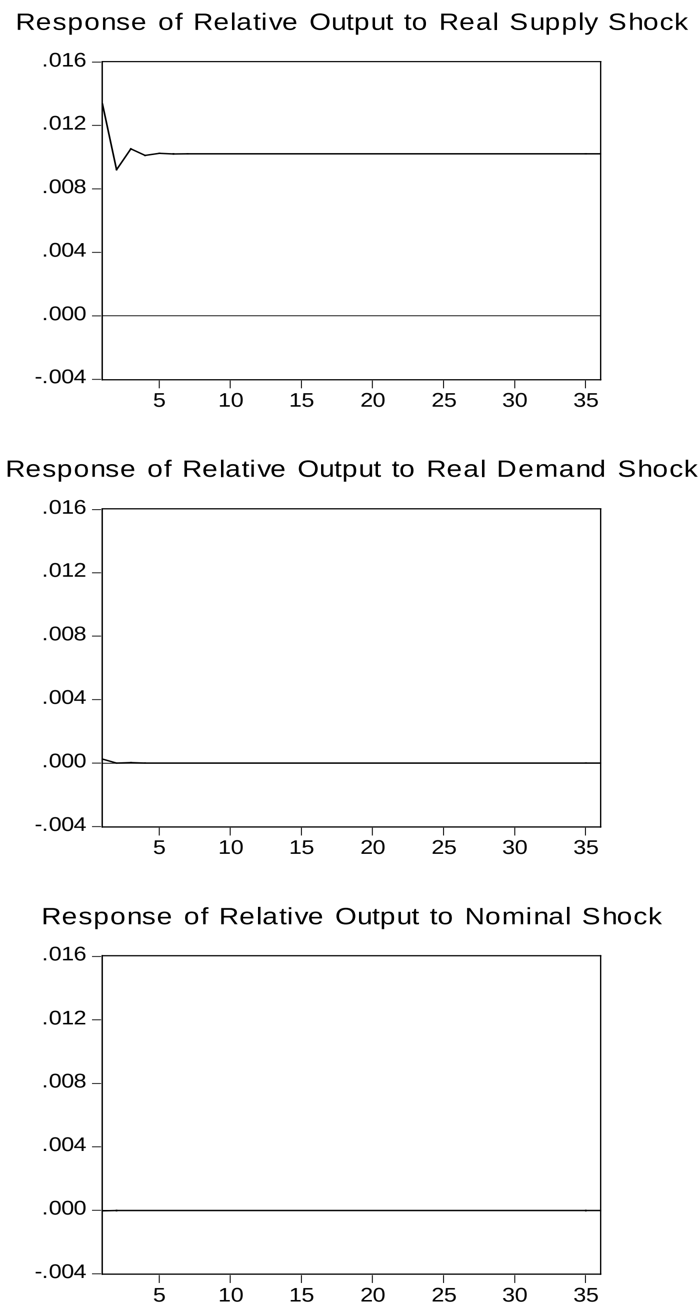
Using the SBIC to choose the optimal lag length of VAR in empirical analysis, we find that the VAR(1) model is the most appropriate. We present the impulse responses of each of the variables to one standard deviation in each of the fundamental shocks.

Figure 3 illustrates the impulse response functions of the explanatory variables to one-standard-deviation structural shocks. Panel (a) shows that supply shocks lead to a permanent increase in relative output. Positive real or nominal demand shocks have smaller effects and leave the long-run relative output level unaffected in accordance with the long-run restrictions imposed.

Panel (b) demonstrates that a positive supply shock leads to a persistent decrease in the real exchange rate. A positive real demand shock is associated with a permanent increase in the real exchange rate, while a nominal shock temporarily increases the real exchange rate and is asymptotic to zero in accordance with the long-run restrictions imposed.

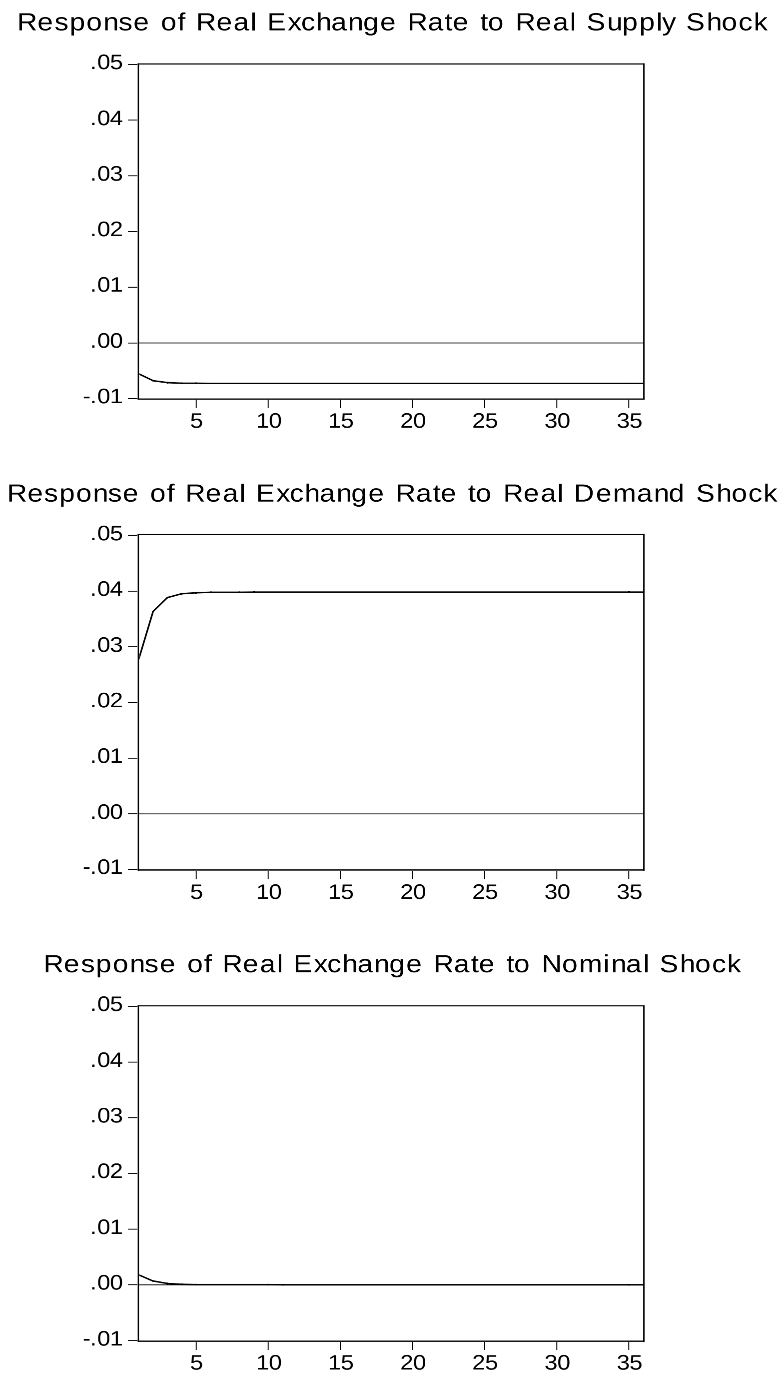
Finally, panel (c) shows that the impulse response of the nominal exchange rate drops immediately and permanently after a supply shock. We can also see that demand shocks have positive permanent effects on nominal exchange rates, while nominal shocks have positive permanent effects on nominal exchange rates.

Next, we calculate the forecast error variance decompositions. Table 2 shows the fraction of the forecast error variance that can be attributed to each shock at different horizons in the model for each variable. Supply shocks account for most of the variance in output growth throughout the estimation horizons and represent the most important factor for variation in the forecast errors of relative output. The rest of the variance is attributable to demand and nominal shocks in more or less similar fractions (less than 1%). The estimates imply that real supply shocks explain a substantial amount of the variance of output growth.



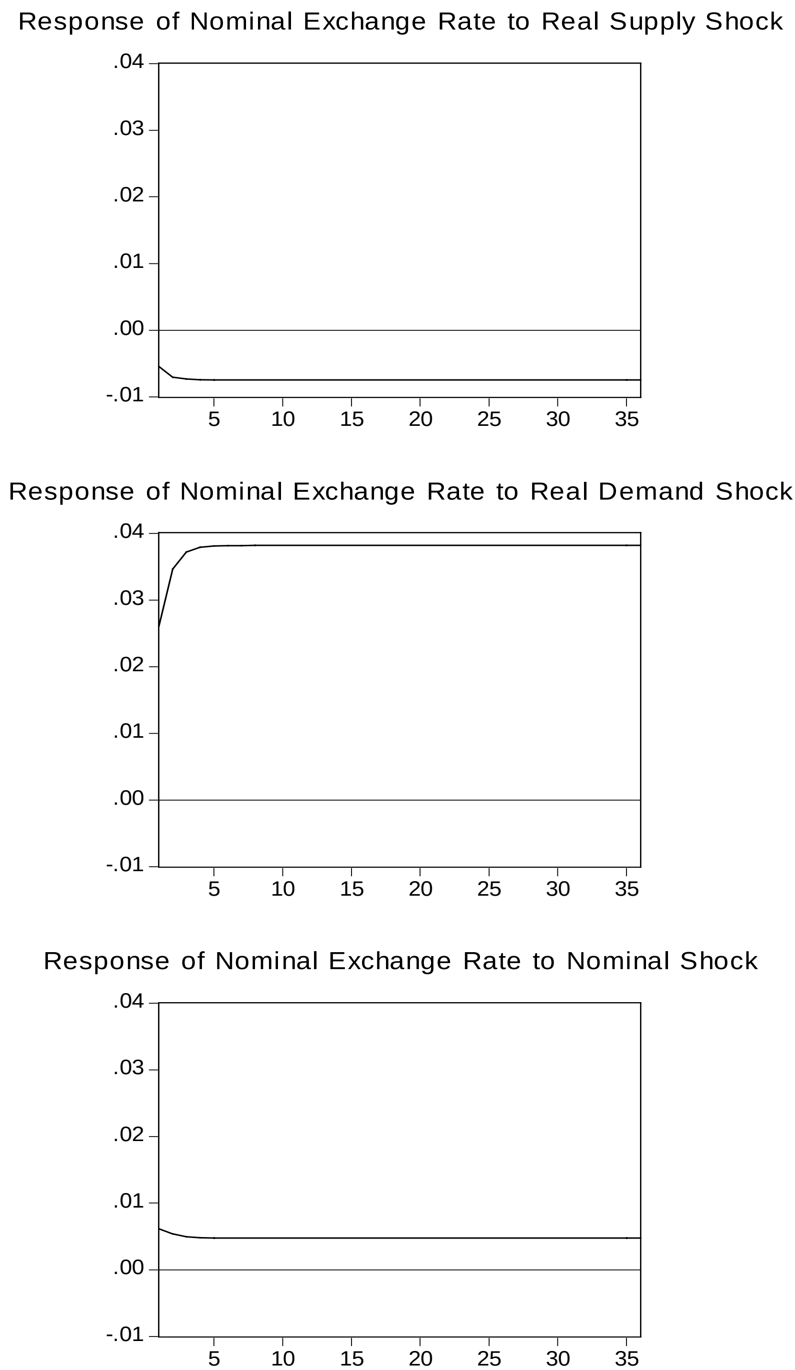
(a) Response of Real Relative Output

Figure 3. Continued on next page.



(b) Response of Real Exchange Rate

Figure 3. Continued on next page.



(c) Response of Nominal Exchange Rate

Figure 3. Accumulated Impulse Responses (Trivariate Model).

Forecast error variance decompositions for the variations in the real exchange rate suggest that real demand shocks explain most of the movements in the real exchange rate. Real demand shocks, the most important factors, account for about 95% of the variance in exchange rate movement. Real supply shocks, meanwhile, appear to play a small role in explaining fluctuations in the real in exchange rate, accounting for roughly 3 to 4% of the forecast error variance. The contribution of nominal shocks to the fluctuations of the real of the real exchange rate amounts to about 0.5%. To summarize, a substantial amount of the forecast error variance of the change in the real exchange rate in the yen-dollar rate is due to real demand shocks.

Table 2. Forecast Error Variance Decomposition (Trivariate System)

(a) Real Relative Output

Horizon (month)	Real Supply Shock (%)	Real Demand Shock (%)	Nominal Shock (%)
1	99.965	0.034	0.000
3	99.936	0.064	0.001
6	99.936	0.064	0.001
9	99.936	0.064	0.001
12	99.936	0.064	0.001
24	99.936	0.064	0.001
36	99.936	0.064	0.001

(b) Real Exchange Rate

Horizon (month)	Real Supply Shock (%)	Real Demand Shock (%)	Nominal Shock (%)
1	3.818	95.770	0.411
3	3.654	95.803	0.543
6	3.652	95.802	0.546
9	3.652	95.802	0.546
12	3.652	95.802	0.546
24	3.652	95.802	0.546
36	3.652	95.802	0.546

(c) Nominal Exchange Rate

Horizon (month)	Real Supply Shock (%)	Real Demand Shock (%)	Nominal Shock (%)
1	3.960	91.009	5.031
3	3.888	91.492	4.620
6	3.887	91.494	4.619
9	3.887	91.494	4.619
12	3.887	91.494	4.619
24	3.887	91.494	4.619
36	3.887	91.494	4.619

Finally, forecast error variance decompositions for nominal exchange rates show that more than 90% of the variation in the changes of nominal exchange rates comes from real demand shock. The rest of the variance is attributable to real supply shocks and nominal shocks in similar fractions (less than 5%).

5. Some Concluding Remarks

This paper examines the sources of real exchange rate fluctuations of yen-dollar rates using a long-run structural VAR approach. We begin by identifying two types of macroeconomic shocks (real and nominal) by the method of Enders and Lee (1997), and then uncovering the sources of movements in real exchange rates by a technique developed by Blanchard and Quah (1989). The evidence presented indicates that real shocks play a dominant role in explaining the fluctuations of the real exchange rate. Real disturbances account for more than 99% of the forecast error variance of the real exchange rate and more than 97% of the nominal exchange rate. These results are consistent with those of Lastrapes (1992) and Enders and Lee (1997).

Next, we extend the model to identify three types of macroeconomic shocks (real supply, real demand and nominal), in order to reveal the sources of movements in real exchange rates. The evidence presented indicates that real shocks (real demand and real supply) play a dominant role in explaining the fluctuations of the real exchange rate. Real disturbances account for more than 99% of the forecast error variance of the real exchange rate and more than 95% of the nominal exchange rate. These results are consistent with those obtained by the bivariate SVAR model.

The dominant influence of real shocks over real exchange rate fluctuations in yen-dollar rates has some implications for policy management and exchange rate modeling. With regard to exchange rate policy, our results imply that the best approach for policy authorities, in their policymaking to improve competitiveness, is to focus on improvements in the real economy, such as in efficiency, technologies, and productivity. This also calls into question monetary policies which seek to promote competitiveness through currency devaluation.

Acknowledgements

The author is grateful to the Japan Society for the Promotion of Science for providing financial support to the 21st Century Center of Excellence Project.

References

Bayoumi, T. and Eichengreen, B., (1992), Shocking aspects of European Monetary Unification, *National Bureau of Economic Research Working Paper* No. **3949**.

Blanchard, O. and Quah, D., (1989), The dynamic effects of aggregate demand and supply disturbances, *American Economic Review,* **79**, 655-673.

Chowdhury, I.S., (2004), Sources of exchange rate fluctuations: empirical evidence from six emerging market countries, *Applied Financial Economics*, **14**, 697–705.

Clarida, R. and Gali, J., (1994), Sources of real exchange rate fluctuations: how important are nominal shocks? *National Bureau of Economic Research Working Paper* No. **4658**.

Dibooglu, S and Kutan, A (2001), Sources of real exchange rate fluctuations in transition economies: the case of Poland and Hungry, *Journal of Comparative Economics*, **29**, 257– 275.

Dickey, D.A. and Fuller, W.A., (1979), Distribution of the estimators for autoregressive time series with a unit root, *Journal of the American Statistical Association*, **74**, 427–431.

Dornbusch, R., (1976), Expectations and exchange rate dynamics, *Journal of Political Economy*, **84**, 1161-1176.

Enders, W. and Lee, B.-S., (1997), Accounting for real and nominal exchange rate movements in the post-Bretton Woods period, *Journal of International Money and Finance*, **16**, 233-254.

Lastrapes, W.D., (1992), Sources of fluctuations in real and nominal exchange rates, *Review of Economics & Statistics*, **74**, 530-9.

Rogers, J.H., (1999), Monetary shocks and real exchange rates, *Journal of International Economics*, **49**, 269-88.

Stockman, A.C., (1987), The equilibrium approach to exchange rates, *Economic Review*, Federal Reserve Bank of Richmond, (Mar./Apr), 12-30.

Wang, T (2004) "China: Sources of Real Exchange Rate Fluctuations," *International Monetary Fund Working Paper* No. **04/18**.

In: Emerging Topics in Macroeconomics
Editor: Richard O. Bailly, pp. 29-34
ISBN: 978-1-60692-093-0

MACROECONOMICS OF CATASTROPHIC EVENT RISKS: RESEARCH COMES OF AGE

***Inwon Jang*[1] *and David Kim*[2]**
[1]Department of Finance, Girard School of Business
Merrimack College, MA, USA
[2]Department of Economics, University of Sydney,
Sydney, NSW, Australia

1. Introduction

Risks exist in various forms and at different levels in an economy. Some risks may be predictable and hence insurable using appropriate market instruments while some others cannot be insured away, which are typically known as systematic risks. Nevertheless, risks affect the economic decisions and outcomes *ex ante*. Insurable risks are well known in standard economic theory, and have been well understood. The effects of unpredictable and uninsurable risks are more difficult to analyze in theory.

In financial economics, risks concerning asset prices and financial transactions can be traded using various financial strategies that have been well developed, such as through trading derivative securities or taking some form of collaterals. Most observable financial and corporate risks are continuously priced in the market as reflected in the prices of bonds, shares, and interest rates. Consequently, a growing volume of financial research has focused on identifying and quantifying how these risks perceived by agents influence asset prices. In particular, understanding the role of default (or credit) risk, also widely known as *event risks*, and how markets process information concerning such risks was the subject of a growing research in finance. In macroeconomics, however, recent developments in research programs have occurred around the role of expectations in the face of aggregate uncertainties, simply known as ‘shocks’. The dynamic stochastic general equilibrium (DSGE) models have been well established as the major methodological workhorse to quantitatively answer a variety of macroeconomic questions. However, given that macroeconomics only deals with aggregate quantities and prices, the risks modeled in conventional macroeconomic research are typically aggregate shocks such as technology (TFP) shocks, output shocks, monetary policy shocks, fiscal shocks, among others. However, a relatively unexplored area in macroeconomic research is to understand the macroeconomic effects of non-economic risks or shocks with catastrophic magnitudes. Examples of non-economic event risks include man-made events such as the September 11th terrorist attack and the Iraq war, or even a potentially worldwide technology breakdown once feared as the Y2K bug, and any other form of major political or social meltdowns. Non-economic, non-human induced event risks would include natural disasters such as the recent earthquake in the Sichuan province in China and tsunamis/hurricanes in Southeast Asia. These non-economic, man-made or natural catastrophic events can occur in both developed and developing countries, and would be very difficult to predict the timing and scale of impacts. The coherent framework of DSGE modeling has not yet been fully explored to examine the macroeconomics of these events. In modern finance, there has been a rapid growth in the developments of complex credit-related financial instruments such as credit default swaps (CDS), collateralized debt obligations (CDO), and so forth, which are marketed to hedge credit/financial risks. Analytical frameworks used in finance have as, a consequence, flourished. Nevertheless, insurance markets hardly exist or at best very inadequate at the aggregate level due to the nature of these risks.

In this essay, we discuss a potentially fruitful but largely neglected area of research in macroeconomics. It has often been argued that catastrophic events have little impact on the aggregate economic performance as these events are typically regional, e.g. the September-11th terrorist attack was mainly concentrated in a district in the city of New York, and hence the aggregate impact in terms of lost output in the US is a negligible portion of its GDP, as argued by Sharpiro (2003) among others. These catastrophic events may also be theoretically insurable in well developed insurance markets. However, the US cannot be generalized, as it is such a large economy with abundant physical and human capital. A question has to be asked whether such a terrorist attack can occur in several financial and economic centers at the same time with similar or even greater scales. A man-made catastrophic event or natural disaster may also occur in a developed or developing country in which almost all human or productive capital stocks are concentrated in a particular region. Conventional arguments that downplay the effects of catastrophic events often fail to take into account other factors such as the level of public confidence, a higher degree of surrounding uncertainty and its associated economic costs and the time duration required to restore productive capital. The conventional arguments also rely on the assumption that the government will take any necessary counter-cyclical measures to avert serious economic downturns, which may not necessarily be the case in many developing countries.

There are several ways of analyzing an economy wide catastrophic events or risks in the literature. Schmitt-Grohe and Uribe (1999) theoretically examine the general equilibrium effects associated with the Y2K (also known as the millennium bug prior to the turn of the 21st century) on the aggregate economy. However, in their theoretical model, the catastrophic event is something that agents could possibly predict, but may never be realized at all, as in the case of the Y2K. Hence, investment is not only affected by the fear of realization of the event but also by agents' attempt to avert the anticipated event. Okuyama (2003) analyzes the effects of natural disasters in the framework of the more traditional Solow-Swan growth model. However, this framework is limited to analyzing how capital is destroyed and then accumulates, and hence cannot explain the full dynamics of other dimensions of the economy. In short, the model framework has severe limitations to be used for a quantitative analysis.

There are some distinctions to be made in economic analysis depending on whether the catastrophic event is terrorism, natural disasters or others such as wars. With terrorism risk, the economy as a whole is exposed to a new type of systematic risk. This increases the costs associated with stepping up the level of security and entails significant opportunity costs, which are similar to social costs, associated with minimizing the outbreak of crimes. While not discussed here, it is natural to presuppose that terrorism risk also increases government spending on military expenditures.

The standard model framework used in modern macroeconomics is the optimal stochastic growth model, which is typically explored in discrete time. The main feature to be added is modeling an exogenous risk term and how it is fed into the aggregate economy. The real business cycle (RBC) models, as a predecessor of the DSGE models, are well known in this regard, as they use an optimal growth framework and exogenous shocks. However, the nature of exogenous shocks modeled is different in macroeconomic models with event risks. In RBC models, shocks are typically to technology, preference, and the terms of trade, etc, with a known probability distribution of the continuous forcing processes to the private agents. The shocks are typically specified as an autoregressive process with independently and identically distributed innovations. The parameters of the forcing processes can then be either calibrated or estimated using appropriate econometric techniques and then simulated. However, macroeconomic models of event risks cannot rely on such assumptions about the exogenous processes. First, shocks are typically one-off, typical of a major catastrophic event, rather than perceived to be continuously occurring with a known probability distribution. The theoretical framework outlined below is based on the recent exploration in bond pricing literature in which event risks are priced in typically corporate bonds using a jump diffusion process. In this literature, default is modeled as an unpredictable jump of a conditional Poisson process as derived by Duffie and Singleton (1999) and Jarrow and Turnbull (1995), and also shown in Driessen (2005).

It is this powerful tool used in credit risk pricing literature that can be combined with a general equilibrium macroeconomics in an optimizing framework. A recent work by Abadie and Gardeazabal (2008) provides an interesting attempt to borrow this tool in an optimizing macroeconomic model, although their focus is on the effects of terrorism on the foreign direct investment (FDI) decisions. The theoretical framework outlined below provides a more generalized account of the theoretical model potentially useful for extensions.

2. Theoretical Framework

Consider the following simple two-country model with production, as analysed by Abadie and Gardeazabal (2008). We assume that capital is perfectly mobile across countries.

Production

In the following, outputs for domestic, Y^D and foreign economy, Y^F are specified as a jump-diffusion process, respectively:

$$dY^D(t) = \alpha\eta(t)K(t)dt + \sigma\eta(t)K(t)dW(t) - l(t)\eta(t)K(t)dN(t-),$$

And, similarly for the foreign output,

$$dY^F(t) = \alpha^*(1-\eta(t))K(t)dt + \sigma^*(1-\eta(t))K(t)dW^*(t) - l^*(t)(1-\eta(t))K(t)dN^*(t-),$$

Note that α, α^*, $\eta(t)$, $\eta^*(t)$ and $l(t)$, $l^*(t)$ are an expected growth rate, a fraction of the capital for production (or net investment), and a loss rate of catastrophic event for domestic and foreign economy, respectively. Note that dW is an increment of Wiener process which is normally distributed with mean of zero and variance of *dt*. In the production process, there is a special term denoted *dN*, which shows an increment of the Poisson process.

A Poisson process is utilized to capture catastrophic events in the model. Thus *dN* = 1 if there is a catastrophic event and *dN* = 0 otherwise. The process N_t can also be represented as $N_t = \sum_{j=0}^{\infty} 1_{\{\tau_j \le t\}}$, where τ_j represent the arrival of the j^{th} catastrophe. When the Poisson process has such a representation, the probability of having *j* events in the time interval $(0,t]$ is given by

$$\text{Prob}(N_t = j) = \frac{(\lambda t)^j}{j!}\exp(-\lambda t)$$

where λ is the intensity (or the instantaneous probability) of arrival of a catastrophe.

Thus, return to capital is represented as the following jump-diffusion process:

$$dR(t) = \frac{dY(t) - \delta\eta(t)K(t)dP(t)}{\eta(t)K(t)} = \alpha dt + \sigma dW(t) - \delta dP(t)$$

Utility

A representative agent maximizes her utility from consumption, *U*(*C*(*t*)). By assumption, the instantaneous utility function *U* is continuously differentiable, strictly increasing, and strictly concave in consumption (i.e. $U_c > 0$ and $U_{cc} < 0$). We also assume that $lim_{c\to 0}U_c = \infty$ and $lim_{c\to\infty}U_c = 0$. Given the production process, the optimization problem can be summarized as follows:

$$\max_{\{C,\eta\}} E\left[\int_0^{\infty} e^{-\rho t}U(C(t))dt\right]$$

subject to

$$dK(t) = (\alpha\eta(t)K(t) + \alpha^*(1-\eta(t))K(t) - C(t))dt + \sigma\eta(t)K(t)dW(t) + \sigma^*(1-\eta(t))K(t)dW^*(t)$$
$$-l(t)\eta(t)K(t)dN(t-) - l^*(t)(1-\eta(t))K(t)dN^*(t-).$$

The solution to this problem can be obtained using the stochastic optimal control theory to derive optimal consumptions, $\{C(t),_{t\geq 0}\}$ and the net investment, $\{\eta(t),_{t\geq 0}\}$.

Let the indirect utility function (or value function) be:

$$I(k,t) = \max_{\{C,\eta\}} E\left[\int_0^\infty e^{-\rho t}U(C(t))dt \mid K(0) = k\right],$$

where the maximum is taken over all feasible consumption plans. Using the property of indirect utility function in an infinite time horizon case, we can simplify the Hamilton-Jacobi-Bellman (HJB) equation as follows:

$$0 = \max_{\{C,\eta\}} \begin{pmatrix} U(C) - \rho I(k) + I_k(\alpha\eta k + \alpha^*(1-\eta)k - c) + \frac{1}{2} I_{kk}(\sigma^2\eta^2 k^2 + (\sigma^*)^2(1-\eta)^2 k^2) \\ +\lambda(I(k - l\eta k) - I(k)) + \lambda^*(I(k - l^*(1-\eta)k) - I(k)) \end{pmatrix}.$$

Solving the maximization problem on the right-hand side of the above equation, it is possible to obtain and evaluate the optimal consumptions and the impact of the catastrophic events on the net domestic investment when the catastrophic events affect domestic and/or foreign capital stocks. As described above, we used a continuous time model where the exogenous process (in this case, catastrophic events) follows a jump process. The standard model one can envisage may have the following features: optimizing agents, production with physical and human capital, a two-country world with capital mobility but with no or limited labour mobility, unpredictable event risk with potential being catastrophic, specified as a jump Poisson process. A closed form solution may or may not exist and one might rely on approximate solutions.

3. Future Work

In this review essay, we outlined a simple illustrative model that ignores other dimensions that could be considered by researchers. Catastrophic events destroy human capital as well as productive capital. Human capital is more costly to replace once destroyed and takes longer to rebuild, and hence having more persistent effects on the economy. Thus, the model can be extended to include human capital, whose fraction is permanently destroyed due to a catastrophic event. In developing countries with scarce human capital, especially, skilled labour, catastrophic events would also adversely affect the long run growth potential. Another possible dimension to model is the regime or structural change in the model triggered by changes in expectations. As catastrophic events might lead to a fundamentally altered perception about risk and hence expectations held by agents, one might consider a model with the feature of regime or structural change. While the above model is a single good model, an interesting extension can be built around a multi-sector model in which inter-sectoral linkages and dynamics of adjustments can be studied. There are various ways in which interesting questions can be analyzed. For example, one might want to explore the dynamic effects of foreign aids in a developing country affected by catastrophic events. This would usefully lead to an area of macro-development economics. Policy questions can also be analyzed. For example, one might want to examine the effects of increasing labour mobility across national borders and how it could contribute to the adjustment of the affected economy. These areas include just some of the obvious but interesting research questions that might be explored in future research. We are of the view that in the current global climate with increased terrorism risks, more frequent natural disasters possibly due to climate changes, among others, exploring the potential of dynamic general equilibrium models with exogenous processes characterized by a jump-diffusion process would open the door for an exciting branch of modern macroeconomics for many years to come.

References

Abadie, A. and Gardeazabal, J. (2008), "Terrorism and the World Economy", *European Economic Review*, **52**, 1-27.

Duffie, D. and Singleton, K. (1999), "Modeling Term Structures of Defaultable Bonds", *Review of Financial Studies*, **12**, 687-720.

Driessen, J. (2005), "Is Default Event Risk Priced in Corporate Bonds?", *Review of Financial Studies*, **18**, 165-195.

Jarrow, R. and Turnbull, S. (1995), "Pricing Derivatives on Financial Securities Subject to Default Risk", *Journal of Finance*, **50**, 53-86.

Okuyama, Y. (2003), "The Economics of Natural Disasters: A Critical Review", Paper presented at the 50th North American Meeting, Regional Science Association International, November, Philadelphia.

Sharpiro, R. (2003), "Al-Qaida and the GDP", The Dismal Science, *Slate*, February 28.

Schmitt-Grohe, S. and Uribe, M. (1999), "Y2K", *Review of Economic Dynamics*, **2,** 850-856.

RESEARCH AND REVIEW STUDIES

In: Emerging Topics in Macroeconomics
Editor: Richard O. Bailly, pp. 37-74 ISBN 978-1-60692-093-0

Chapter 1

SCALING DISTRIBUTIONS IN ECONOMICS: THE SIZE DISTRIBUTION OF INCOMES

Fabio Clementi*
Department of Economics, Polytechnic University of Marche,
Piazzale R. Martelli 8, 60121 Ancona, Italy

Abstract

In economic data, skewed and thick tailed frequency curves are the rule rather than the exception. Again, there seems to be an increasing interest in the study of economic variables which can be considered as realizations of a stochastic process. Within this class a major role is played by income distribution.

It is undoubtedly true that the distribution of the command over the goods and services produced by a society is of crucial importance for the student of welfare economics. On a more practical level, producers of consumer goods must study the distribution of income in order to estimate the probable extent of their markets. The size distribution of income also affects the social and economic policy-making when planning the magnitude and the structure of taxation schemes or evaluating the effectiveness of tax reforms. In economic and social statistics, the form of the income distribution is the basis for the measurement of the inequality of incomes and more general social welfare evaluations.

Over the last 100 years, a large number of distributions has been proposed for the modeling of size distribution of personal incomes. The most canonical version of such models was discovered by the mathematical economist and sociologist Vilfredo Pareto in 1897. Pareto formulated his model to explain the distribution of income and wealth, and believed that it was a universal candidate for the mathematical description of given income data for different countries over relatively long time periods. However, more recent studies suggest that it is only the upper end of income and wealth distributions that follows the Pareto's model, with the lower ends following the lognormal form of the Gaussian distribution that is associated with the random walk, originally proposed for the whole of the income distribution by the French economist Robert Gibrat in 1931.

Since the early 1990s, there has been an explosion of work on economic size phenomena in the physics literature, leading to an emerging new field called "econophysics". The list of such phenomena includes, among others, the distribution of

*E-mail address: fabio.clementi@univpm.it.

returns in financial markets, the distribution of income and wealth, the distribution of economic shocks and growth rate variations, and the distribution of firm sizes and growth rates. A common theme among those who identify themselves as econophysicists is that standard economic theory has been inadequate or insufficient to explain the non-Gaussian properties empirically observed for various of these phenomena, such as excessive skewness and leptokurtotic fat tails. Indeed, many economic phenomena occur according to distributions that obey scaling laws rather than Gaussian normality. Whether symmetric or skewed, the tails are fatter or longer than they would be if Gaussian, and they appear to be linear in figures with the logarithm of a variable plotted against the logarithm of its cumulative probability distribution. This is true of the Pareto's distribution, and soon a variety of efforts have been made by physicists, mathematicians, and economists to model a variety of stochastic economic phenomena using either the Pareto's distribution or one of its relatives or generalizations, such as for instance the Lévy stable distribution.

In the light of these considerations, this chapter offers a reader's guide to "scaling" phenomena in economics with particular focus on the interplay between the dynamics and empirical regularities of income distribution.

1. Introduction: Economic Complexity

Increasingly, in economics what had been considered to be unusual and unacceptable has come to be considered usual and acceptable. Whereas it had been widely believed that economic reality could be reasonably described by sets of pairs of linear supply and demand curves intersecting in single equilibrium points to which markets easily and automatically moved, now it is understood that many markets and situations do not behave so well. The order of the economy appears to emerge from the complex interactions that constitute the evolutionary process of the economy, characterized by nonlinearity, discontinuity, and a variety of other phenomena that are not so easily predicted or understood. This set of phenomena has come to be labeled as *complexity in economics*.

Economic reality reflects the enormous variety and diversity of individuals in their knowledge, attitudes, and behaviors, and what emerges in the aggregate results from their multiform interactions and peculiarities in the enormous range of institutional frameworks and may have little to do with what happens at the individual level. The idea that reality reflects complexity has come from numerous disciplines of thought and has reverberated back and forth across them, from mathematics and computer science to physics and biology to the social sciences. But only in recent years such idea is arising in economics and has increasingly come to have important influence on economic thinking. This change of perspective has stimulated a variety of new approaches to analysis. The acknowledgment that systems are complex has led to a multi-disciplinary literature in which a range of attempts have been made to understand complex systems behavior through the use, for example, of evolutionary computation and artificially intelligent agent models. Possible outcomes and ranges of solutions of emergent phenomena from complex systems, as well as their limits and possibilities, are now inductively discovered by the use of increasingly powerful computer simulations and experimental methods rather than of deductive formal proofs of theorems that sought to derive general solutions broadly applicable.[1]

One reason for economics' preoccupation with these issues is the observation that statistical aggregates are much stabler and more regular than individual behavior. Economists commonly explain that this is due to the Adam Smith's "invisible hand", and that despite the conflicting interests of individuals, the result of the pursuit of their selfish ends is socially satisfactory and some relative order emerges, the market providing the mechanism which links and coordinates all the disparate activities pursued by individuals. As a consequence, when different decision units affect one another only through their effects upon economy-wide variables such as market prices—*i.e.*, when random variations in individual characteristics are largely independent, independent shocks to individual units tend to have an effect on the variability of aggregate quantities that vanishes as the number of units becomes large due to the effects of the *law of large numbers*. But the macroeconomic models which claim to give a simplified picture of the economic reality paradoxically have almost no activity which needs such a coordination. This is because they typically assume that the choices of all the diverse agents in one sector—consumers for example—can be considered as the choices of one *representative* standard utility maximizing individual whose choices coincide with the aggregate choices of the heterogeneous individuals (Kirman, 1992, 1999). This *reductionist* approach, inherited from the methodology of classical mechanics, implies that the optimal aggregate solution can be obtained by means of a simple summation of the equilibrium choices made by each optimizing agent.

The quantum revolution of last century radically changed the perspective in contemporary physics, leading to a widespread rejection of reductionism (Bertuglia & Vaio, 2003). According to the *holistic* approach, the aggregate is different from the sum of its components because of the interaction of sub-units, and the equilibrium of a system does not require any more that every single element is in equilibrium. The type of macroscopic equilibrium that can result from this more complex vision of systemic phenomena has been stud-

[1]As a matter of example, the newly developing field of *Agent-based Computational Economics* (*ACE*) concerns the computational study of economies modeled as evolving systems of autonomous interacting agents. A principal concern of ACE researchers is to understand the apparently spontaneous formation of global regularities in economic processes, such as the unplanned coordination of trading activities in decentralized market economies, and to explain how these global regularities arise from the *bottom-up* through the repeated local interactions of autonomous agents channeled through socio-economic institutions rather than from the top-down imposition of coordination mechanisms, such as market-clearing constraints or assumptions of single representative agents. ACE researchers generally rely on computational laboratories to study the evolution of decentralized market economies under controlled experimental conditions. The ACE modeler starts by constructing an economy with an initial population of agents. These agents include both economic agents (*e.g.*, traders, financial institutions, etc.) and agents representing various other social and environmental phenomena (*e.g.*, government, land, weather, etc.). The ACE modeler specifies the initial state of the economy by specifying the initial attributes of each agent, which might include internalized behavioral norms, internally stored information about itself and other agents, and the like. The economy then evolves over time without further intervention from the modeler. All events that subsequently occur must arise from the historical time-line of agent-agent interactions, and no extraneous coordination devices are permitted. See Tesfatsion (2000, 2001, 2002*a*,*b*,*c*, 2003) and Tesfatsion & Judd (2006) for an illustration of the the key characteristics and goals of the ACE methodology. See also the web site `http://www.econ.iastate.edu/tesfatsi/ace.htm` for extensive resources related to the ACE research ares. A concrete example of ACE modeling strategy applied to economics can be found in Delli Gatti *et al*. (2008).

ied by condensed-matter physicists under the name of *Self-Organized Criticality* (*SOC*).[2] Physicists have noted, in several contexts, the possibility of a *critical state*, in which independent microscopic fluctuations can propagate so as to give rise to some relative order on a macroscopic scale resulting from the balance of actions of a large number of many interacting units. This is a state in which chain reactions initiated by repeated nonlinear local disturbances neither damp out over a short distance (the *subcritical* case) nor propagate explosively so that the system cannot remain in that state (the *supercritical* case). Moreover, the system may spontaneously evolve toward a critical state and return to it even if perturbed by an external shock.[3]

According to the notion of SOC, when mutually interconnections and relationships among units of complex systems are accounted for, *scaling* phenomena characterized by *heavy-tailed* distributions (*e.g.*, power-law distributions) of event size emerge naturally, suggesting that the many detailed features of a system that could in principle affect its sub-parts somehow average so that their individual characteristic scales are not imprinted on the aggregate distribution. Therefore, the occurrence of a power-law may be read as a symptom of self-organizing processes at work. As a consequence, the search for natural laws and the occurrence of scaling phenomena in economics does not necessarily require the adoption of the reductionist paradigm: agents' choice should not necessarily be an equilibrium one, derived from their optimizing behavior, because agents' interaction generates self-organizing solutions giving rise to certain statistical regularities at aggregate level. This turn out to be incompatible with mainstream economics, which, imposing unnatural symmetric conditions, is not able by construction to reproduce the persistent heterogeneity of economic agents captured by skewed non-Gaussian distributions which can not be fully described by mean and variance.

2. High Variability in Complex Systems: Models for Heavy-Tailed Data

A macroeconomic system is a complex system which consists of many individuals interacting with each other. The individuals function in the background provided by the legal and institutional frames, and differ in abilities, education, mentality, historical and cultural background, and so on. They enter the system with different financial and cultural initial conditions. Each of them has his own vision of what is important and of what s/he is willing and able to achieve. Therefore, it is a hard task to formulate a general theory of needs and financial possibilities of a single individual or to create an economic profile of a typical member of such a complicated system. There are too many random factors to be taken into

[2] See *e.g.* Bak (1996) and Bak *et al*. (1987, 1988).

[3] The prototypical example of SOC is a *sandpile*. When the slope of the pile is nowhere too steep, dropping on additional grains of sand at randomly chosen sites has no macroscopic effects, as at most small numbers of grains will shift position in each case. However, randomly dropping on additional sand will eventually result in the slope of the pile increasing to a critical slope, at which point large avalanches can occur in response to the dropping of a single additional grain of sand. A sandpile with a slope that is initially greater than the critical slope also evolves toward it, in this case through an immediate large avalanche that collapses the pile. Thus, while the existence of macroscopic instability without large external shocks depends upon a particular critical slope, the system endogenously evolves toward exactly that state.

account changing in time, sometimes slowly, sometimes faster, sometimes abruptly and in an unpredictable way. Every day some individuals leave the system, some new enter it. It is difficult, if not impossible, to follow individual changes. One can however control their *statistics*, which shape the system on large macroeconomic scale and drive the large scale phenomena observed in the whole macrosystem.

The aim of macroeconomic studies is to extract important factors, understand their mutual relations and describe the development of past events. The ultimate goal is to reach a level of understanding which would also permit to predict the reaction of the system to the change of macroeconomic parameters in the future. Having such a knowledge at hand, macroeconomists would be able to stimulate the optimal evolution by appropriately adjusting the macroeconomic parameters. This level of understanding goes far beyond a formal description and requires modeling and understanding of fundamental principles which are difficult because of the complexity of the problem. Clearly, a model whose main ambition would be to realistically take into account all parameters and factors characterizing the whole network of dependencies in such a complex system would fail to be comprehensive and solvable. One would not be able to learn anything from such a model. It would be even too complicated to properly reflect what it actually intends to describe.

Therefore, one has to find a way of simplifying the underlying complexity to the level that enables a formulation of a treatable model. A danger of a simplification of a complex and nonlinear problem is that by a tiny modification one can loose an important part of the information or introduce some artificial effects.

A common research theme in the study of complex systems is the pursuit of *universal* properties that transcend specific system details. Such universal laws may uncover global regularities which are insensitive to tiny changes of parameters within a given class of them. Such laws also provide a classification of possible universal large scale behaviors that can occur in the system and can be used as a first order approximation in the course of gaining insight into the mechanisms driving the system.

A well known example of the emergence of universal laws is the *Central Limit Theorem* (*CLT*), which can be thought as the cornerstone for understanding collective phenomena.[4] Indeed, the study of natural and social systems involves correlations, nonlinearity, external driving, and so on, which lead to many possible results, but the surviving fact is the emergence, often but not always, of a macroscopic coherent and well-defined universal behavior to understand which the CLT can be used as a starting point. The CLT tells us that the sum of many independent identically distributed (i.i.d.) random numbers polled from a distribution with a finite mean and a finite variance obeys a Gaussian law with the mean and the variance scaling with the number of terms in the sum independently of the particular shape of the distribution. That is: if $\{X_n; n \geq 1\}$ is a sequence of i.i.d. random variables having finite mean μ and finite variance σ^2, and $S_n = X_1 + X_2 + \cdots + X_n$, $n \geq 1$, denote the n^{th} partial sum, then as $n \to +\infty$

$$\tilde{S}_n = \frac{S_n - n\mu}{\sigma\sqrt{n}} \sim N(0,1), \tag{1}$$

where $N(0,1)$ is the standard Gaussian distribution having mean 0 and variance 1. In other words, summing n i.i.d. random variables with probability density functions

[4] See Feller (1971) for basic concepts, notations and extensions related to the Central Limit Theorem.

$p_1(x_1), p_2(x_2), \ldots, p_n(x_n)$, one obtains that the probability density function of the rescaled sum $\tilde{S}_n$, $p_n(\tilde{s}_n)$, is given by the convolution of the n probability density functions of each rescaled random variable, *i.e.*

$$\begin{aligned} p_n(\tilde{s}_n) = p_1(\tilde{x}_1) \otimes p_2(\tilde{x}_2) \otimes \cdots \otimes p_n(\tilde{x}_n) = \\ = \int \mathrm{d}\tilde{x}_1\, p_1(\tilde{x}_1) \int \mathrm{d}\tilde{x}_2\, p_2(\tilde{x}_2) \cdots \\ \cdots \int \mathrm{d}\tilde{x}_{n-1}\, p_{n-1}(\tilde{x}_{n-1})\, p_n(\tilde{s}_n - \tilde{x}_1 - \ldots - \tilde{x}_{n-1}), \end{aligned}$$

and is characterized by a Gaussian probability density with unit variance[5]

$$p_n(\tilde{s}_n) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\tilde{s}_n^2}.$$

Therefore, one can see that the contributions of many random variables lead to a global behavior of the sum which becomes both simple and universal: the original details of the probability density functions of the contributing variables except the mean and the variance are progressively forgotten in the course of enlarging the number of terms in the sum to give rise to the universal Gaussian shape.

The study of limit theorems uses the concept of the *basin of attraction* of a probability distribution, concerning the changes in the functional form of $p_n(s_n)$ that occur when n changes: when n increases, $p_n(s_n)$ changes its functional form, and if the fundamental hypotheses of the CLT are verified—*i.e.*, if both independence and finite variance conditions of the n random variables X_i are satisfied, assumes the Gaussian functional form with mean $n\mu$ and variance $n\sigma^2$ for an asymptotically large value of n.[6] Therefore, the Gaussian probability density function is said to be an *attractor* (or *fixed point*) in the functional space of probability density functions for all the probability density functions that fulfill the requirements of the CLT. The set of these probability density functions constitutes the basin of attraction of the Gaussian probability density function.

The CLT result has encouraged and justified the overwhelmingly widespread adoption of the Gaussian distribution in the most diverse fields. As a matter of example, most of the

[5]This remark is very useful in practice, because it provides a strategy for testing for the existence of Gaussian laws. The idea is to compare different data sets, for instance obtained from different system sizes, and try to collapse the different probability density functions onto a "universal" master curve. This *finite size scaling* technique can be stated more rigorously in the following way: take different samples of different sizes $n_1, n_2, \ldots$; construct the rescaled variables $\tilde{S}_{n_1} = \frac{S_{n_1} - n_1\mu}{\sigma\sqrt{n_1}}, \tilde{S}_{n_2} = \frac{S_{n_2} - n_2\mu}{\sigma\sqrt{n_2}}, \ldots$; check if all the variables exhibit the same probability density function, that is if all the data collapse onto the same Gaussian curve. When such a rescaling holds, one says that the probability density function exhibits scaling properties, that is it remains described by the same law in different measurements. The concept of scaling has been found to be widely shared by equilibrium systems at their critical points, as well as by out-of-equilibrium systems which are self-organizing in states with many different scales.

[6]When the rescaled sum (1) has exactly the same probability density function as the initial variables—*i.e.*, the probability density function of the rescaled sum is independent of n, the Gaussian distribution is said to be *stable*. In other words, a stable distribution is a specific type of distribution encountered in the sum of n i.i.d. random variables having the property that it does not change its functional form for different values of n. Due to the assumption of independence between the initial variables, this allows one to derive the complete knowledge of the probability density function of the sum from the sole knowledge of the probability density function of the individual variables.

statistical models in use are constructed by using the Gaussian as a distribution for the noise terms, where one hypothesizes that the “noise” affecting a certain phenomenon is produced by the joint effect of a set of factors independent of each other. However, at times the empirical evidence does not fit with the theoretical hypothesis above because of empirical distributions of a wide range of phenomena which are not conformable with the Gaussian. Thus in some cases the analysis of estimation residuals of a model contradicts the Gaussian hypothesis because of an excess of *asymmetry* and/or *kurtosis*. This can be caused by model specification errors (nonlinearity, omission of relevant variables, and so on), but also there could be phenomena whose shape of the distribution itself has to be questioned.

The fact that the empirical distribution of the estimation residuals is far from the Gaussian could signal that the individual variances of the components of the disturbance term may not be finite. Indeed, a typical feature of the Gaussian and other commonly-used distribution functions is that their (right) tail decreases exponentially fast, implying that all moments, including exponential moments, exist and are finite. In practice, this property ensures that a random variable X exhibits low variability and thus concentrates tightly around its mean. However, one aspect of many complex systems that has received considerable attention is a tendency toward tremendous variability in event size, such that the observed quantities describing such systems can be reasonably represented by probability distributions that allow *skewness* and *fat* (or *heavy*) *tails*, that is distributions whose (right) tail decays more slowly than any exponential, implying that their regular moments can also be infinite. A well-known example of skewed and heavy-tailed distribution is the Pareto’s (power-law) distribution (Arnold, 1983; Kleiber & Kotz, 2003; Clauset *et al*., 2007)

$$p\left(x \mid k, \alpha\right) = \frac{\alpha k^{\alpha}}{x^{\alpha+1}}, \tag{2}$$

with Complementary Cumulative Distribution Function (CCDF) given by

$$\Pr\left(X > x\right) = 1 - F\left(x\right) = \left(\frac{k}{x}\right)^{\alpha}, \tag{3}$$

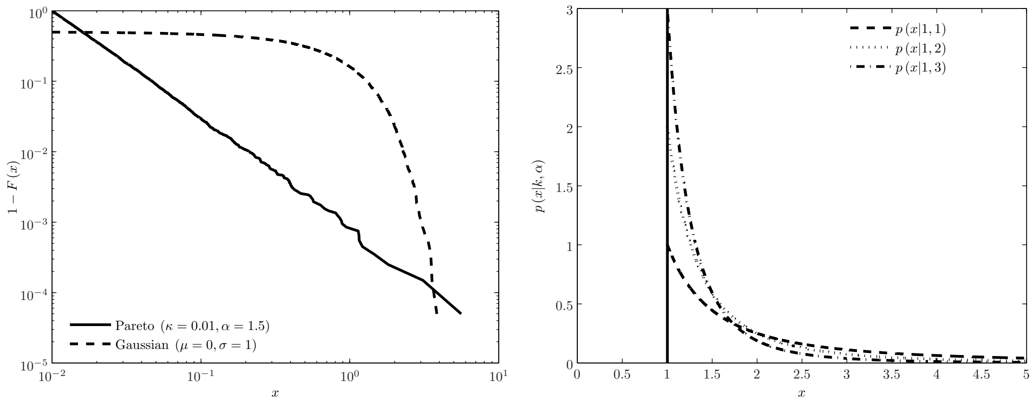
where $k \leq x < \infty$ and $k, \alpha > 0$, k being the (necessarily positive) minimum possible value of X. For $1 < \alpha \leq 2$ X has infinite variance

$$\left\langle x^2 \right\rangle - \left\langle x \right\rangle^2 = \frac{k^2 \alpha}{\left(\alpha - 1\right)^2 \left(\alpha - 2\right)}$$

but finite mean

$$\left\langle x \right\rangle = \frac{k\alpha}{\alpha - 1},$$

while for $0 < \alpha \leq 1$ X has also infinite mean. The Pareto’s probability density function (2) for $\alpha = 1, 2, 3$ and $k = 1$ is shown in Figure 1(a). This relation is usually verified by plotting the logarithm of (3), $\log\left[1 - F\left(x\right)\right]$, as a function of the logarithm of x, $\log x$: one should obtain a curve that for large x becomes a straight line sloping down to the right with an absolute slope equal to α. Figure 1(b) exhibits the CCDFs for two same-sized samples ($n = 10,000$) drawn from a Pareto’s distribution with $k = 0.01$ and $\alpha = 1.5$ and a Gaussian distribution with $\mu = 0$ and $\sigma = 1$ plotted in the double logarithmic scale. Since



(a) PDF (b) CCDF

Figure 1. (a) Pareto's probability density functions for $\alpha = 1, 2, 3$ and $k = 1$. (b) CCDF double logarithmic plot for a Pareto's distribution with $k = 0.01$ and $\alpha = 1.5$ and a Gaussian distribution with $\mu = 0$ and $\sigma = 1$.

the Gaussian distribution is very "short-tailed", one can notice how the slope of the graph becomes asymptotically infinite.[7]

To understand the meaning of (extraordinarily "long-tailed") Paretian laws with infinite population moments, one can use the *sequential moment estimate plot* method (see Mandelbrot, 1963; see also Willinger *et al.*, 2004).This method plots the *running moment*

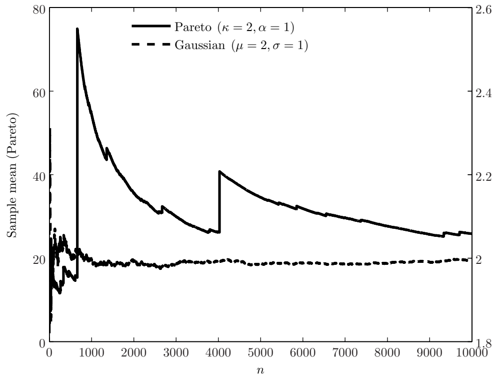
[7]A power-law distribution as the Pareto's distribution is also sometimes called a *scale-free* distribution, because it is the only distribution that is the same whatever scale one looks at it on. To illustrate this fact, one can suppose to have some probability distribution $p(x)$ for a quantity x and to discover or somehow deduce that it satisfies the property that

$$p(kx) = kp(x)$$

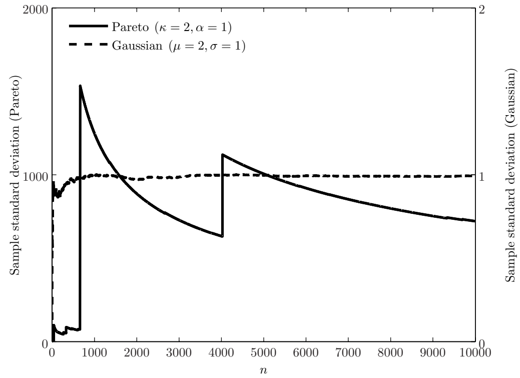
for any k. That is, if one increases the scale or unit by which x is measured by a factor of k, the shape of the distribution $p(x)$ is unchanged, except for an overall multiplicative constant. Thus, for instance, an immediate consequence of the presence of the Pareto's (power-law) tail in the probability distribution of incomes is that the probability that a random person from the richer part of the society is k times richer than another person with income x is independent of x, *i.e.*

$$\frac{p(kx)}{p(x)} = k^{-(1+\alpha)}.$$

Therefore, the Pareto's (power-law) distribution is scale-free, reflecting a certain "self-similarity" of the structure of the richest class. The scale appears in the problem through the parameter k, which provides the lower cutoff above which the power-law part of the distribution sets in. For example, if $k = 10$ and $\alpha = 2$, then the Pareto's law predicts that the number of people ten times richer is roughly one thousand (10^{-3}) times smaller. The suppression factor is very sensitive to α. If the value of α moves toward unity, the suppression factor decreases, and for $k = 10$ it is only 10^{-2}. In other words, in economies with a smaller value of α the tail of the distribution is fatter. This leaves more space for rich individuals. Thus one intuitively expects that for smaller α the economy is more liberal. In a more restrictive economic system, the Pareto's exponent α is larger and hence the richer population is suppressed.



(a) Sequential mean plot

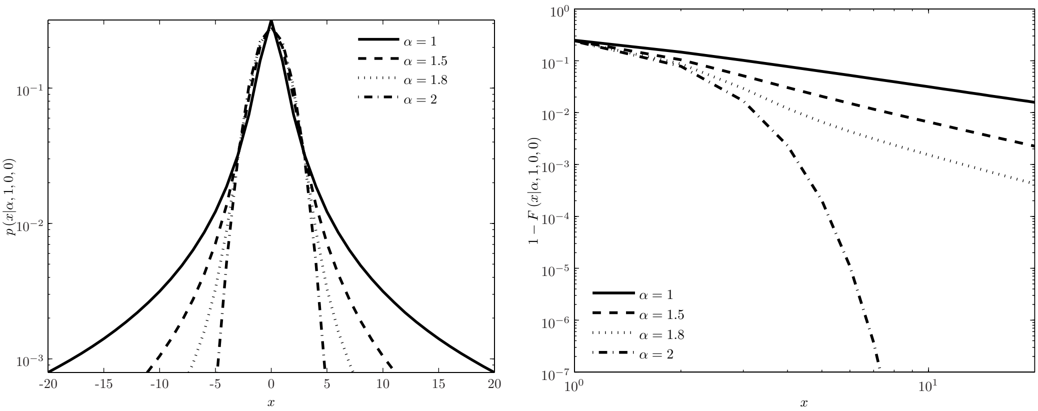


(b) Sequential standard deviation plot

Figure 2. Sequential moment plots corresponding to same-sized ($n = 10,000$) distinct samples of a Paretian random variable with $k = 2$ and $\alpha = 1$, and a Gaussian random variable with $\mu = 2$ and $\sigma = 1$: (a) sequential mean plot; (b) sequential standard deviation plot. The figure gives an idea of how erratic and sample-dependent the moments of Paretian variables can be expected to be.

estimates, that is the value of a moment estimate of the data is plotted as a function of the number of observations used in the estimation of the moment. For example, Figure 2 shows the sequential (a) mean and (b) standard deviation plots corresponding to same-sized ($n = 10,000$) distinct samples of a Paretian random variable with $k = 2$ and $\alpha = 1$ and a Gaussian random variable with $\mu = 2$ and $\sigma = 1$, obtained by simply inverting series of random variables uniformly distributed in the interval $[0, 1]$. As one can easily recognize, while in the Gaussian case the mean and standard deviation estimates exist, are finite, and converge robustly to their theoretical value as the number of observations increases, in the Paretian case the sample mean and standard deviation estimates do not converge as n increases, even if they always exist for any fixed n. This fact confirms that for a Paretian random variable with $\alpha = 1$ the first and second moments do not exist, *i.e* the computer-simulated data set is a sample from an underlying distribution having mean and variance infinite.

Although the Gaussian attractor is the most important and famous attractor in the functional space of probability density functions, the observed lack of convergence of the sequential moment estimate plots for the Paretian random sample in Figure 2(a)–(b) translates directly into inconsistencies of models that assume finite moments, either implicitly or explicitly. Therefore, other existing attractors and limit theorems must be considered when the main conditions under which the classical CLT holds are not satisfied. Lévy (1954) discovered that, in addition to the Gaussian law, there is a rich class of probability distributions allowing both skewness and fat tails and sharing the stability condition. In the mathematical and statistical literature, these probability distribution are called "Lévy α-stable laws". Such a name has been assigned to these distributions because a sum of two or more independent random variables having a Lévy α-stable distribution with index α is again Lévy α-stable with the same index α. This less well-known version of the CLT, the *generalized or*



(a) PDF (b) CCDF

Figure 3. Symmetric ($\beta = 0$) standardized ($\delta = \mu = 0$, $\gamma = 1$) Lévy α-stable probability distribution for $\alpha = 1, 1.5, 1.8, 2$: (a) probability density function semi-logarithmic plot; (b) complementary cumulative distribution function double logarithmic plot. Observe that the Gaussian ($\alpha = 2$) forms a parabola and is the only Lévy α-stable density with exponential tail. This much faster decay is clearly visible in the double logarithmic plot of the complementary cumulative distribution functions, where the power-law tails for $\alpha < 2$ are also clearly visible. Therefore, the smaller the tail index, the stronger is the power decay behavior.

stable-law CLT, can be stated in a more rigorous way as follows. Suppose that $\{X_n; n \geq 1\}$ is a sequence of non-negative, i.i.d. random variables with scaling distribution function F and $1 < \alpha \leq 2$ (implying finite mean but infinite variance). Let $S_n = X_1 + X_2 + \cdots + X_n$, $n \geq 1$, denote the n^{th} partial sum. Then as $n \to +\infty$

$$\tilde{S}_n = \frac{S_n - n\delta}{\sqrt[\alpha]{n}} \sim L_\alpha (1, \beta, 0), \tag{4}$$

where $L_\alpha (1, \beta, 0)$ is the four-parameter Lévy α-stable distribution with the *index of stability* or *characteristic exponent* $\alpha \in (0, 2]$, having location (shift) parameter $\delta = 0$, $\delta \in \mathbb{R}$, and scale parameter $\gamma = 1$, $\gamma > 0$. $\beta \in [-1, 1]$ is the skewness parameter quantifying the asymmetry of the Lévy distribution. As Figure 3(a) shows,[8] the tail exponent α determines the rate at which the tails of the distribution taper off.[9] The generalized CLT (4) applies

[8]The stable probability density and complementary cumulative distribution functions have been computed using the program STABLE provided by John P. Nolan, available as a free download from the following web address: `http://academic2.american.edu/~jpnolan/stable/stable.html`.

[9]When $1 < \alpha \leq 2$, the shift parameter δ is the mean, μ, but for $\alpha \leq 1$ the mean is infinite. Therefore, the sum of n Lévy variables with the same exponent α and distributed according to $p(x|\alpha, \gamma, \beta)$ can be rescaled in a similar way to (4) by the change of variable $\frac{S_n}{\sqrt[\alpha]{n}}$, where the mean $\delta = \mu$ does not appear as it is no longer defined. Lévy variables with $1 < \alpha \leq 2$ are sometimes referred to as "Pareto-Lévy distributions". See Mandelbrot (1960) for a definition and a study of the properties of Pareto-Lévy processes as well as their application within the frame of the theory of income distribution.

under the same restrictions (except for the finiteness of the variance when $\alpha > 2$) of independence and of large n, but an important difference is observed between the Gaussian and stable non-Gaussian attractors: upon n convolutions and as $n \to +\infty$, the Gaussian probability density function "attracts" all the probability density functions decaying as or faster than $\frac{1}{|x|^3}$ at large $|x|$; similarly, upon n convolutions, all probability density functions with exponent $\alpha < 2$ are attracted to the symmetric Lévy law. Therefore, there is an abrupt change in the tail behavior of Lévy α-stable laws at the borderline case with exponent $\alpha = 2$: while for $\alpha < 2$ all the Lévy distributions are heavy-tailed, the case $\alpha = 2$ is special and represents the familiar, not heavy-tailed, Gaussian distribution; *i.e.*, $L_2(1,0,0) = N(0,1)$, and the classical CLT (1) is recovered with the replacement $\alpha = 2$ in (4).

Unfortunately, there are no general closed form of the Lévy distributions, except for a few special cases.[10] However, Lévy laws are fully characterized by the expression of their *characteristic functions*, which are essentially the same as the Fourier transform of the probability density functions. Then, by performing the inverse Fourier transform, one gets

$$p(x|\alpha,\gamma,\beta,\delta) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} \mathrm{d}t\, \varphi(t)\, e^{-itx}, \tag{5}$$

where the characteristic function $\varphi(t)$ is given by[11]

$$\varphi(t) = \begin{cases} \exp\left[it\delta - \gamma\,|t|^{\alpha}\left(1 - i\beta\frac{t}{|t|}\,\tan\frac{\pi\alpha}{2}\right)\right] & \text{if } \alpha \neq 1, \\ \exp\left[it\delta - \gamma\,|t|\left(1 + i\beta\frac{t}{|t|}\frac{2}{\pi}\,\log|t|\right)\right] & \text{if } \alpha = 1, \end{cases}$$

which in the symmetric case ($\beta = 0$) with a zero shift parameter ($\delta = 0$) and $\alpha \neq 1$ reduces

[10]The analytical form of the Lévy α-stable distribution is known only for a few values of α and β:

- $\alpha = \frac{1}{2}$, $\beta = 1$ (Lévy-Smirnov);
- $\alpha = 1$, $\beta = 0$ (Cauchy or Lorentzian);
- $\alpha = 2$ (Gaussian).

Moreover, in the limit as γ or α approach zero, the distribution will approach a Dirac delta function $\delta(x-\mu)$. For a formal treatment of stable distributions and their limit theorems see Nolan (2007).

[11]Note that this is not the sole parameterization in use for stable distributions; it is only the most common. For a discussion of the different possible parameterizations of stable distributions see *e.g.* Nolan (1998).

to the simplest form $\varphi(t) = \exp(-\gamma |t|^{\alpha})$. Thus the function (5) becomes[12]

$$
\begin{aligned}
p(x|\alpha,\gamma) &= \frac{1}{2\pi}\int_{-\infty}^{+\infty} \mathrm{d}t\, e^{-\gamma|t|^{\alpha}} e^{-itx} = \\
&= \frac{1}{2\pi}\int_{-\infty}^{+\infty} \mathrm{d}t\, e^{-\gamma|t|^{\alpha}} [\cos tx - i\sin tx] = \\
&= \frac{1}{2\pi}\int_{-\infty}^{+\infty} \mathrm{d}t\, e^{-\gamma|t|^{\alpha}} \cos tx - \frac{i}{2\pi}\int_{-\infty}^{+\infty} \mathrm{d}t\, e^{-\gamma|t|^{\alpha}} \sin tx = \\
&= \frac{1}{\pi}\int_{0}^{+\infty} \mathrm{d}t\, e^{-\gamma t^{\alpha}} \cos t|x|\,.
\end{aligned} \tag{6}
$$

Integrating by parts, yields

$$
\begin{aligned}
p(|x|\,|\alpha,\gamma) &= \frac{1}{\pi}\left\{\left[e^{-\gamma t^{\alpha}} \sin\frac{t|x|}{|x|}\right]_0^{+\infty} + \frac{\gamma\alpha}{\pi}\int_0^{+\infty} \mathrm{d}t\, e^{-\gamma t^{\alpha}} \sin\frac{t|x|}{|x|} t^{\alpha-1}\right\} = \\
&= \frac{\gamma\alpha}{\pi}\int_0^{+\infty} \mathrm{d}\xi\, e^{-\frac{\gamma\xi^{\alpha}}{|x|^{\alpha}}} \sin\frac{\xi}{|x|}\left(\frac{\xi}{|x|}\right)^{\alpha-1}\frac{1}{|x|} = \\
&= \frac{\gamma\alpha}{\pi|x|^{1+\alpha}}\int_0^{+\infty} \mathrm{d}\xi\, e^{-\frac{\gamma\xi^{\alpha}}{|x|^{\alpha}}} \sin\xi\xi^{\alpha-1},
\end{aligned}
$$

where the two substitutions $\xi = t|x| \Rightarrow \mathrm{d}\xi = \mathrm{d}t|x|$ have been made. Finally, taking the limit of the definite integral as x approaches positive infinity gives

$$
\lim_{x\to+\infty}\int_0^{+\infty} \mathrm{d}\xi\, e^{-\frac{\gamma\xi^{\alpha}}{|x|^{\alpha}}} \sin\xi\xi^{\alpha-1} = \int_0^{+\infty} \mathrm{d}\xi\, \sin\xi\xi^{\alpha-1},
$$

[12] In (6) the Euler's formula

$$
e^{-itx} = \cos -tx + i\sin -tx = \cos tx - i\sin tx,
$$

and the fact that the cosine function is an even symmetry ($\cos x = \cos -x \Rightarrow \cos x + \cos -x = 2\cos -x = 2\cos x$) and the sine function is an odd symmetry ($\sin -x = -\sin x \Rightarrow \sin x + \sin -x = 0$) respectively, have been used.

because the function $e^{-\frac{1}{|x|}} \to 1$ as $x \to +\infty$. Therefore, one gets

$$p\left(|x|\,|\alpha,\gamma\right) \sim \frac{\gamma\alpha}{\pi\,|x|^{1+\alpha}} \int\limits_0^{+\infty} \mathrm{d}\xi\ \sin\xi\xi^{\alpha-1} = \frac{\gamma\alpha\Gamma\left(\alpha\right)\sin\frac{\pi\alpha}{2}}{\pi\,|x|^{1+\alpha}} = \tag{7}$$
$$= \frac{\gamma\Gamma\left(1+\alpha\right)\sin\frac{\pi\alpha}{2}}{\pi\,|x|^{1+\alpha}} =$$
$$= \frac{\gamma C_\alpha}{|x|^{1+\alpha}},$$

where $C_\alpha = \frac{\Gamma(1+\alpha)\sin\frac{\pi\alpha}{2}}{\pi}$.[13] This asymptotic behavior for large values of x is a power-law behavior, a property with deep consequences for the moments of the distribution.[14] Specifically, all Lévy α-stable processes with $\alpha < 2$ have *infinite* variance, and for $\alpha \leq 1$ they also lack a typical size or "scale" around which measurements are centered as in the Gaussian case.

Therefore, distributions with infinite variance belong to the basin of attraction of the Lévy α-stable distributions. Moreover, even if the classical CLT is often cited as the reason for the ubiquity with which Gaussian distributions occur in the study of many natural and social systems, the generalized CLT in which the condition of finite variance is replaced by a much less restricting one concerning a somewhat regular behavior of the tails states that the stable Gaussian distribution having all its moments finite turns out to be, in fact, a very special case (*i.e.*, $\alpha = 2$).[15]

[13] In (7) the Mellin integral transform $\int_0^{+\infty} \mathrm{d}\xi\ \sin\xi\xi^{\alpha-1} = \Gamma\left(\alpha\right)\sin\frac{\pi\alpha}{2}$ has been used, where $\Gamma\left(\alpha\right) = \int_0^\infty \mathrm{d}x\, x^{\alpha-1}e^{-x}$ is the Euler's gamma function that satisfies the property $\Gamma\left(1+\alpha\right) = \alpha\Gamma\left(\alpha\right)$.

[14] The convergence to a power-law tail varies for different α's, and, as can be seen in Figure 3(b), is slower for larger values of the tail index. Moreover, the tails of Lévy α-stable distribution functions exhibit a crossover from a decay with exponent $\alpha \geq 2$ to the true power-law tail with exponent $\alpha < 2$.

[15] The results of the classical and generalized CLTs show that the Gaussian and stable distributions are both invariant under *aggregation*. Mandelbrot (1963) shows that stable distributions are essentially invariant also under transformations such as "maximization" and "mixture". In particular, Gaussian distributions lack this invariance properties. As a matter of example, one can consider the case of n independent random variables denoted $X_1, X_2, \ldots, X_n$, and assume that their distribution functions follows the form of the law of Pareto with the same tail index parameter α but with different scale coefficients $C_i > 0$; *i.e.*

$$\Pr\left(X_i > x\right) = C_i x^{-\alpha},$$

for $1 \leq i \leq n$. By defining the random variables M_k, $1 \leq k \leq n$, to be the k^{th} successive *maxima* given by

$$M_k = \max\left(X_1, X_2, \ldots, X_k\right),$$

and using the fact that $\Pr\left(M_k \leq x\right) = \prod_{i=1}^k \Pr\left(X_i \leq x\right)$, it is straightforward to show that for large x

$$\Pr\left(M_k > x\right) = 1 - \Pr\left(M_k \leq x\right) = 1 - \prod_{i=1}^k \Pr\left(X_i \leq x\right) = 1 - \prod_{i=1}^k \left(1 - C_i x^{-\alpha}\right) = \sum_{i=1}^k C_i x^{-\alpha} = C_{M_k} x^{-\alpha},$$

where $C_{M_k} = \sum_{i=1}^k C_i$. Thus, the k^{th} successive maxima of Pareto's random variables are also Paretian, with the same tail index α but different scale coefficients than the individual X_i's. Moreover, if one considers the *weighted mixture* W_n of the X_i's, and denotes by p_i the probability that $W_n = X_i$, then

$$\Pr\left(W_n > x\right) = \sum_{i=1}^n p_i \Pr\left(X_i > x\right) = \sum_{i=1}^n p_i C_i x^{-\alpha} = C_{W_n} x^{-\alpha},$$

3. Size Distributions in Economics: The Distribution of Incomes

A well known example in the macroeconomic research of universal law governing size distributions of economic variables which can be deduced from the mathematics of large numbers is the issue of the income distribution. The skewness of income distributions is the characteristic which is least easy to disregard, and the relative constancy of form of the distribution in different periods and countries has attracted by far the greatest attention and has been the subject of argument since Pareto (1964, 1965) first stated his general law.

Well aware of the imperfections of statistical data, insufficient reliability of the sources, and lack of veracity of income statements, Pareto boldly analyzed the data for different countries and years using his extensive engineering and mathematical training and succeeded in showing that there is a relation between N_x – the number of individuals with incomes greater or equal to x – and the value of the income x given by the downward sloping line

$$\log N_x = \log C_\alpha - \alpha \log x,$$

or equivalently

$$N_x = \frac{C_\alpha}{x^\alpha},$$

with $C_\alpha = k^\alpha > 0$, $\alpha > 0$, and $x \geq k$, k being the minimum income. Pareto observed that the (absolute) values of the slope α clustered around 1.5; this constancy and the quality of the fit encouraged him to assert a law of income distribution, emphasizing from the start the heavily asymmetric character of his distribution and hence its fundamental difference from a normal curve.[16]

where $C_{W_n} = \sum_{i=1}^{k} p_i C_i$ is the weighted average of the separate scale coefficients C_i. Thus, the distribution of the weighted mixture of Pareto's distributions is also Paretian, with the same tail index α but a different scale coefficient than the individual X_i's. Aggregation, mixture, and maximization are transformations that occur frequently in economic systems, and are inherently part of many measured observations that are collected about them. For example, a number of operations of aggregation occur in the context of firm size when "old" firms merge within a "new" one, as well as aggregates of various kinds of income are better known than each kind taken separately; the very notion of a firm is to some extent indeterminate, and available data often refer to "firms" that actually vary in size between individual establishments and holding companies, so that such a mixture can be represented by random weighting; if one uses historical data such as natural or technological disasters (*i.e.*, droughts, floods, earthquakes, hurricanes, blackouts, nuclear accidents), the fully recorded and commonly available observations reflect a maximizing choice and correspond to the exceptional (*i.e.*, largest, or most catastrophic) events. In turn, these invariance properties suggest that the presence of skewed and heavy-tailed distributions in data obtained from complex systems should be considered the norm rather than the exception, and should not require "special" explanations.

[16]The fact that empirically the values of parameter α remain "stable", if not constant, caused Pareto to conclude that human nature, that is, humankind's varying capabilities, is the main cause of *income inequality*, rather than the organization of the economy and society. Therefore, what presumably determines the distribution in a community in which the production of wealth is the only way to gain an income is the aptitude for work and saving, steadiness and good conduct. This would prevent necessity or desirability of legislative redistribution of income. Several attacks launched on this and others Paretian claims (*e.g.* Shirras, 1935) as well as considerable interest in Pareto's law (Stamp, 1914; MacGregor, 1936; Hayakawa, 1951; Champernowne, 1952) are reviewed in Persky (1992). However, Pareto acknowledged some drift in the data he used toward lower values of the parameter α which he interpreted as a reduction in inequality. Therefore, seeing little reason to expect large changes in α, he encouraged society to focus in raising mean incomes, rather than pursuing short-sighted efforts to change the income distribution. In his view, an increase in the mean would have imply a reduction in α and hence in inequality, the distribution of incomes resting on fundamental realities and deeper necessi-

However, when examining income distribution data, there is no reason to expect that any one mathematical form will be found superior to all others. In many practical cases, the overriding requirement may be that the function should adequately graduate the data, perhaps over a given portion of the range only. In these cases, the choice may be made safely on empirical ground and may depend on which part of the range is relevant. It was originally recognized by Pareto himself that his original function describes only a portion of the reported income distributions, but apparently this point was later underemphasized. The first time the question of whether a Pareto's distribution or a lognormal gives a better fit was fully developed by Aitchison & Brown (1954, 1957), who observe that the lognormal curve is the better approximation in the lower range of incomes, whilst the Pareto's curve is better in the higher range; this is echoed later, among others, by Montroll & Shlesinger (1982, 1983) and Di Matteo *et al*. (2004), who offer a possible theoretical justification for this observation (see *e.g.* footnote 37), and on a more empirical ground by Cowell *et al*. (1998), Souma (2001, 2002) and Clementi & Gallegati (2005*a*,*b*).[17]

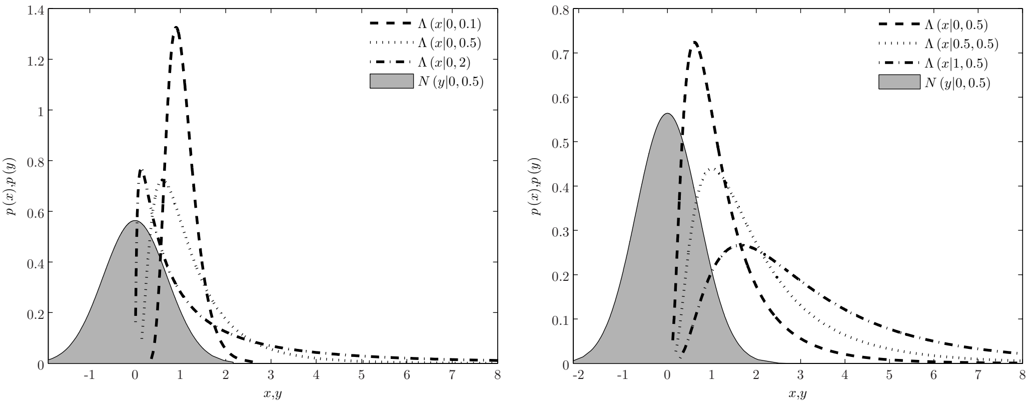
Given that its study consists largely of an interplay of the mathematical properties of the logarithmic function and the statistical properties of the normal distribution, it is not difficult to characterize lognormal distribution mathematically.[18] Indeed, a random variable X is said to be lognormally distributed if $Y = \log X$ is normally distributed with mean μ and variance σ^2. However, despite the familiar properties of the normal distribution have their analogies in the lognormal distribution, certain properties emerge which have no analogues in normal theory. It may be emphasized here that only positive values are possible for the variable X, since the transformation $Y = \log X$ is not defined for $X = 0$. Moreover, the distribution of X is (positively) skewed to the right and has positive kurtosis; these measures of departure from normality increase as σ^2 increases. Figure 4(a)–(b) shows density curves for some selected values of σ^2 and μ compared with the normal distribution $N(y|0, 0.5)$.

Since X and Y are connected by the relation $Y = \log X$, their cumulative distribution

ties of human nature and organization. Pareto's definition of income inequality has led to a continuing debate mostly developed by a number of Italian economists and statisticians, who concentrated on the meaning and significance of the parameter α and suggested alternative indices (see *e.g.* Bresciani-Turroni, 1937; see also Gini, 1921, who demonstrated that in terms of the Lorenz curve lower values of α are associated with increases of inequality). Refer to footnote 20 for further details on the relationship between the Pareto's exponent α, the Lorenz curve and the Gini's coefficient of inequality.

[17]Recently, some researchers report that an *exponential* (Boltzmann-Gibbs) distribution better describes the empirical data in the lower range (see *e.g.* Drăgulescu & Yakovenko, 2001*a*,*b*; see also Willis & Mimkes, 2004, Silva & Yakovenko, 2005, and Nirei & Souma, 2007). A point worth considering here, which may help in explaining the lack of consensus in empirical studies, is that income and wealth can be regarded as *flow* and *stock*, respectively, the former referring to a certain economic quantity in a given period of time and the latter to its accumulation at a point of time. In other words, income depends solely on monies received during an accounting period, whereas wealth depends on both income and spending patterns. Hence, the differences between the empirical studies could be due to differences in whether the measures employed are predominately income measures or wealth measures: measuring the lower end of income usually yields a lognormal distribution, whereas measuring the lower end of wealth yields an exponential. See Wright (2005) for further details on this issue.

[18]See *e.g.* Johnson *et al*. (1994) for a description of the general statistical properties of the lognormal distribution.



(a) PDF for three values of σ^2

(b) PDF for three values of μ

Figure 4. Density curves of selected lognormal distributions compared with a normal distribution. In panel (a), lognormal distributions for $\sigma^2 = 0.1, 0.5, 2$ are compared with the normal distribution $N(y|0, 0.5)$. The plot gives an idea of the flexibility of the lognormal distribution. In panel (b), lognormal distributions for $\mu = 0, 0.5, 1$ are compared with the same normal distribution $N(y|0, 0.5)$. One can note that a change in μ affects the scaling in horizontal and vertical directions, but the essential shape remains the same.

functions are related by

$$\Lambda\left(x|\mu,\sigma^2\right) = \begin{cases} N\left(\log x|\mu,\sigma^2\right) & \text{if } x > 0, \\ 0 & \text{if } x \leq 0. \end{cases}$$

Hence the expression

$$\frac{\mathrm{d}}{\mathrm{d}\,x}\Lambda\left(x|\mu,\sigma^2\right) = p\left(x|\mu,\sigma^2\right) = \frac{1}{x\sigma\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{\log x-\mu}{\sigma}\right)^2}, \tag{8}$$

with $x > 0$, describes the lognormal density curve with the single mode at $x = e^{\mu-\sigma^2}$ and the median at $x = e^{\mu}$. The lognormal distribution possesses moments of any order. From the properties of the moment-generating function of the normal distribution, the r^{th} moment about the origin is given by

$$\langle x^r \rangle = \int_0^{\infty} \mathrm{d}\,\Lambda\left(x|\mu,\sigma^2\right) x^r = \int_{-\infty}^{\infty} \mathrm{d}\,N\left(y|\mu,\sigma^2\right) e^{ry} = e^{r\mu+\frac{1}{2}r^2\sigma^2}. \tag{9}$$

Therefore, the mean (for $r = 1$) is

$$m = \langle x \rangle = e^{\mu+\frac{1}{2}\sigma^2}, \tag{10}$$

and the variance (for $r = 2$) is

$$s^2 = \left\langle x^2 \right\rangle - m^2 = e^{2\mu+2\sigma^2} - e^{2\mu+\sigma^2} = e^{2\mu+\sigma^2}\left(e^{\sigma^2} - 1\right). \tag{11}$$

The relative positions of the mode, median and mean, namely at $x = e^{\mu - \sigma^2}$, $x = e^{\mu}$, and $x = e^{\mu + \frac{1}{2}\sigma^2}$, again emphasize the positive skewness of the distribution.

When considering the statistical description of given income data, the choice of a particular mathematical form should be governed by the economic significance that can be attached to its parameters. For a lognormal distribution of incomes the interpretation of the location parameter μ is straightforward, since it is the logarithm of the median income. It is to be noted that since the arithmetic mean (10) involves both the location and dispersion parameters it is not a pure measure of the level of incomes under the lognormal hypothesis: for this the median is to be preferred. The dispersion parameter σ^2 is of greater interest by virtue of its relation to the concept of concentration of incomes as defined by Lorenz (1905). The so-called *Lorenz diagram* measures the cumulative fraction of population with incomes below x along the horizontal axis and the fraction of the total income this population accounts for along the vertical axis. The points plotted for the various values of x trace out a curve below the 45° line sloping upwards to the right from the origin. In statistical terms the curve describes the relation between the cumulative distribution function, $F(x)$, and the first cumulative moment distribution function, defined by

$$F_1(x) = \frac{\int\limits_0^x \mathrm{d}x'\, x' p\left(x' | \mu, \sigma^2\right)}{\int\limits_0^\infty \mathrm{d}x\, x p\left(x | \mu, \sigma^2\right)}. \tag{12}$$

The measure of income concentration which is naturally suggested by the Lorenz diagram is the *Gini's* (1914) *coefficient* of inequality, which corresponds to the ratio of the area between the Lorenz curve and the 45° line to the area of the triangle under the 45° line.[19] The measure varies from zero, when all persons have the same income (so that the 45° line may be termed the diagonal of "equal distribution"), to unity, when all the available income accrues to one person. The formal definition of the measure is

$$G = \frac{\frac{1}{2} - L}{\frac{1}{2}} = 1 - 2L = 1 - 2\int\limits_0^1 \mathrm{d}F(x)\; F_1(x)\,, \tag{13}$$

where L is the area under the Lorenz curve. For the lognormal hypothesis, the first cumulative moment distribution function (12) can be obtained in the following way. Pulling a

[19] See Cowell (1995) for a discussion of the Gini's coefficient and the measurement of inequality.

factor of x into the exponent of (8) and completing the square yields

$$
\begin{aligned}
xp\left(x|\mu,\sigma^2\right) &= \frac{x}{x\sigma\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{\log x-\mu}{\sigma}\right)^2} = \\
&= \frac{1}{x\sigma\sqrt{2\pi}}e^{-\frac{1}{2\sigma^2}\left(-2\sigma^2\log x\right)}e^{-\frac{1}{2\sigma^2}(\log x-\mu)^2} = \\
&= \frac{1}{x\sigma\sqrt{2\pi}}e^{-\frac{1}{2\sigma^2}\left(\log^2 x-2\mu\log x-2\sigma^2\log x+\mu^2\right)} = \\
&= \frac{1}{x\sigma\sqrt{2\pi}}e^{-\frac{1}{2\sigma^2}\left[\log^2 x-2\left(\mu+\sigma^2\right)\log x+\mu^2\right]} = \\
&= \frac{1}{x\sigma\sqrt{2\pi}}e^{-\frac{1}{2\sigma^2}\left[\log^2 x-2\left(\mu+\sigma^2\right)\log x+\left(\mu+\sigma^2\right)^2-\left(\mu+\sigma^2\right)^2+\mu^2\right]} = \\
&= \frac{1}{x\sigma\sqrt{2\pi}}e^{-\frac{1}{2\sigma^2}\left\{\left[\log x-\left(\mu+\sigma^2\right)\right]^2-\left(\mu+\sigma^2\right)^2+\mu^2\right\}} = \\
&= \frac{1}{x\sigma\sqrt{2\pi}}e^{-\frac{1}{2\sigma^2}\left\{\left[\log x-\left(\mu+\sigma^2\right)\right]^2-2\mu\sigma^2-\sigma^4\right\}} = \\
&= \frac{1}{x\sigma\sqrt{2\pi}}e^{-\frac{1}{2\sigma^2}\left[\log x-\left(\mu+\sigma^2\right)\right]^2}e^{\mu+\frac{\sigma^2}{2}} = \\
&= e^{\mu+\frac{\sigma^2}{2}}p\left(x|\mu+\sigma^2,\sigma^2\right).
\end{aligned}
$$

Therefore, the first cumulative moment distribution function is given by

$$
F_1(x) = \frac{e^{\mu+\frac{\sigma^2}{2}}\int_0^x \mathrm{d}x'\,p\left(x'|\mu+\sigma^2,\sigma^2\right)}{e^{\mu+\frac{\sigma^2}{2}}\int_0^\infty \mathrm{d}x\,p\left(x|\mu+\sigma^2,\sigma^2\right)} = \int_0^x \mathrm{d}x'\,p\left(x'|\mu+\sigma^2,\sigma^2\right) = \Lambda_1\left(x|\mu+\sigma^2,\sigma^2\right),
$$

and the area under the lognormal Lorenz curve is

$$
\begin{aligned}
L &= \int_0^1 \mathrm{d}\Lambda\left(x|\mu,\sigma^2\right)\,\Lambda_1\left(x|\mu+\sigma^2,\sigma^2\right) = \\
&= \int_0^\infty \mathrm{d}x\,p\left(x|\mu,\sigma^2\right)\Lambda_1\left(x|\mu+\sigma^2,\sigma^2\right) = \\
&= \int_0^\infty \mathrm{d}x\left[\int_0^x \mathrm{d}x'\,p\left(x'|\mu+\sigma^2,\sigma^2\right)\right]p\left(x|\mu,\sigma^2\right) = \\
&= \Pr\left(X'\leq X\right) = \\
&= \Pr\left(\frac{X'}{X}\leq 1\right) = \\
&= \Pr\left(\log X'-\log X\leq 0\right),
\end{aligned}
$$

where X' is an independent random variable lognormally distributed as $\Lambda\left(x'|\mu+\sigma^2,\sigma^2\right)$. Because $\log X$ and $\log X'$ are independent normal variates, the variable $Z=\log X'-\log X$

is distributed as

$$N\left(z|\left[\mu+\sigma^2\right]-\mu,\sigma^2+\sigma^2\right)=N\left(z|\sigma^2,2\sigma^2\right).$$

Thus

$$L=\Pr\left(Z\leq 0\right)=N\left(0|\sigma^2,2\sigma^2\right)=N\left(\left[-\sigma^2/\sigma\sqrt{2}\right]|0,1\right)=\Phi\left(-\frac{\sigma}{\sqrt{2}}\right)=1-\Phi\left(\frac{\sigma}{\sqrt{2}}\right),$$

where $\Phi\left(\cdot\right)$ is the standard normal cumulative distribution function. Therefore, the Gini's coefficient of inequality (13) follows from the area under the Lorenz curve as

$$G=1-2\left[1-\Phi\left(\frac{\sigma}{\sqrt{2}}\right)\right]=2\Phi\left(\frac{\sigma}{\sqrt{2}}\right)-1.$$

The Lorenz curves for the lognormal distributions $\Lambda\left(x|0,0.1\right)$, $\Lambda\left(x|0,0.5\right)$ and $\Lambda\left(x|0,2\right)$ with index $G=0.18$, $G=0.38$ and $G=0.68$, respectively, are given in Figure 5(a), showing that the parameter σ^2 may be interpreted as a measure of the concentration of incomes in a sense which is generally acceptable.[20]

It is interesting to notice here that the lognormal probability density function (8) can be mistaken locally for an apparent power-law (Sornette, 2004). This is an important remark for the analysis of data in which the standard procedure is to identify the existence of a

[20] A relationship with Gini's coefficient of inequality can be derived also for the Pareto's distribution (see *e.g.* Cowell, 1995, and Kleiber & Kotz, 2003). The first cumulative moment distribution function (12) written in terms of the cumulative distribution function $F\left(x\right)$ is

$$F_1\left(x\right)=\frac{\int_k^x \mathrm{d}x'\,x'p\left(x'|k,\alpha\right)}{\int_k^\infty \mathrm{d}x\,xp\left(x|k,\alpha\right)}=\frac{\int_0^F \mathrm{d}F'\,x\left(F'\right)}{\int_0^1 \mathrm{d}F'\,x\left(F'\right)},\tag{14}$$

where $x\left(F\right)$ is the inverse of the cumulative distribution function. For the Pareto's distribution

$$F=1-\left(\frac{k}{x}\right)^\alpha\Rightarrow 1-F=\left(\frac{k}{x}\right)^\alpha\Rightarrow x^\alpha=\frac{k^\alpha}{1-F}\Rightarrow x\left(F\right)=\frac{k}{\left(1-F\right)^{\frac{1}{\alpha}}}.$$

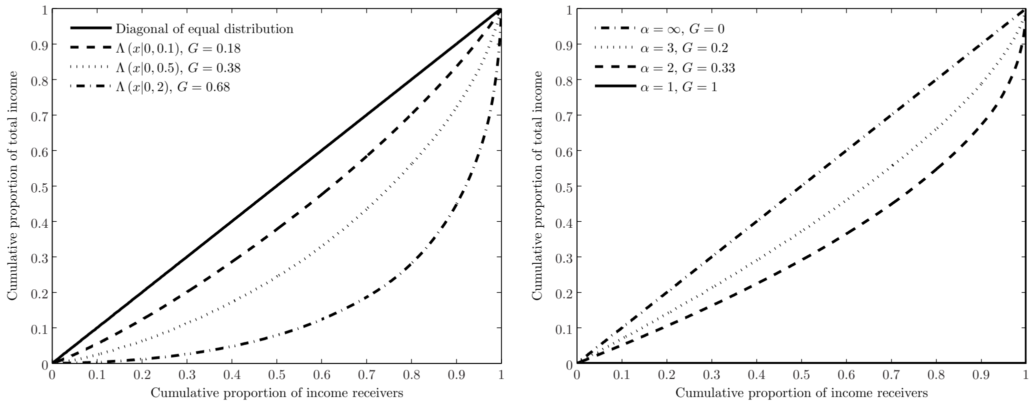
Therefore, the first cumulative moment distribution function (14) is calculated as

$$F_1\left(x\right)=\frac{\int_0^F \mathrm{d}F'\,k\left(1-F\right)^{-\frac{1}{\alpha}}}{\int_0^1 \mathrm{d}F'\,k\left(1-F\right)^{-\frac{1}{\alpha}}}=\frac{\left[-\frac{\left(1-F'\right)^{1-\frac{1}{\alpha}}}{1-\frac{1}{\alpha}}\right]_0^F}{\left[-\frac{\left(1-F'\right)^{1-\frac{1}{\alpha}}}{1-\frac{1}{\alpha}}\right]_0^1}=1-\left(1-F\right)^{1-\frac{1}{\alpha}},$$

where α must be greater than or equal to unity, since the denominator in the expression (14) is just the mean value of x. It follows that the Gini's coefficient (13) for the Pareto's distribution is calculated to be:

$$G=1-2L=1-2\int_0^1 \mathrm{d}F\left(x\right)\left[1-\left(1-F\right)^{1-\frac{1}{\alpha}}\right]=1-2\left\{\left[F\right]_0^1-\left[-\frac{\left(1-F\right)^{2-\frac{1}{\alpha}}}{2-\frac{1}{\alpha}}\right]_0^1\right\}=\frac{1}{2\alpha-1},$$

where L is the area under the Lorenz curve. Examples of the Lorenz curve for a number of Pareto's distributions are shown in Figure 5(b).



(a) Lognormal distribution

(b) Pareto's distribution

Figure 5. (a) The Lorenz diagram for three lognormal distributions of incomes. The curves with index $G = 0.18$, $G = 0.38$ and $G = 0.68$ are derived from $\Lambda(x|0, 0.1)$, $\Lambda(x|0, 0.5)$ and $\Lambda(x|0, 2)$, respectively. (b) Lorenz curves for a number of Pareto's distributions. Note that $\alpha = \infty$ corresponds to perfectly equal distribution ($G = 0$) and $\alpha = 1$ to complete inequality ($G = 1$).

power-law distribution using double logarithmic plots. To see this, one can notice that taking logarithms of both sides of equation (8) yields

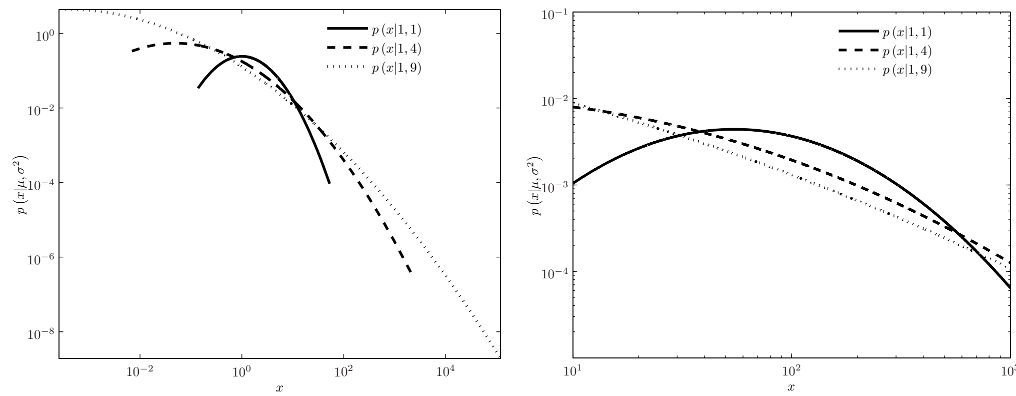
$$\log p\left(x|\mu,\sigma^2\right) = -\log\sigma - \frac{1}{2}\log 2\pi - \frac{1}{2\sigma^2}\log^2 x + \left(\frac{\mu}{\sigma^2} - 1\right)\log x - \frac{\mu^2}{\sigma^2},$$

which is quadratic in $\log x$. However, any quadratic curve looks straight if one views a sufficient small portion of it, so $p\left(x|\mu,\sigma^2\right)$ will look like a power-law distribution when one looks at a small portion of it on a double logarithmic scale. This is clearly seen in Figure 6, where the lognormal probability density function (8) for $\mu = 5$ and $\sigma = 1, 2, 3$ is plotted on a double logarithmic scale. On larger scales, as in panel (a) of the figure, the distribution has some downward curvature which distinguishes the lognormal from a power-law; on the contrary, the magnification of panel (b) shows that the lognormal probability density functions with $\sigma = 2$ and $\sigma = 3$ are close to perfectly linear over about a decade both in abscissa and ordinate. Recalling that a straight line qualifies a power-law in a double logarithmic plot, with some additional noise it would be difficult to distinguish them from pure power-law distributions with constant exponent. In fact, in such a case the effective exponent could be anything, depending on which part of the quadratic the data fall on. To see this, one can notice that

$$e^{-\frac{1}{\sigma^2}(\log x-\mu)^2} = e^{-\frac{1}{\sigma^2}\left[\log\left(\frac{x}{e^\mu}\right)^2\right]} = \left(\frac{x}{e^\mu}\right)^{-\frac{1}{\sigma^2}\log\left(\frac{x}{e^\mu}\right)} = \left(\frac{x}{e^\mu}\right)^{-\alpha(x)}.$$

Using this equality, expression (8) can be rewritten as

$$p\left(x|\mu,\sigma^2\right) = \frac{1}{x\sigma\sqrt{2\pi}}\left(\frac{x}{e^\mu}\right)^{-\alpha(x)} = \frac{1}{x\sigma\sqrt{2\pi}}\frac{e^{\alpha(x)\mu}}{x^{1+\alpha(x)}}, \tag{15}$$



(a) PDF for three values of σ (larger scales) (b) PDF for three values of σ (magnification)

Figure 6. Double logarithmic plot of the lognormal probability density function (8) for $\mu = 5$ and $\sigma = 1, 2, 3$. For the broad range of scales represented in panel (a), a downward curvature is clearly apparent and distinguishes the lognormal from a power-law. The magnification of panel (b) shows that the lognormal probability density functions with $\sigma = 2$ and $\sigma = 3$ are close to perfectly linear over about a decade in abscissa and ordinate.

where the effective x-dependent power exponent $\alpha\,(x)$ is defined by

$$\alpha\,(x) = \frac{1}{\sigma^2} \log\left(\frac{x}{e^{\mu}}\right). \tag{16}$$

The larger the variance σ^2 is, the smaller the exponent (16) is, and therefore the larger is the range of x over which the lognormal distribution (15) will appear to follow a power-law.[21]

This situation arises commonly when considering products of random numbers. Indeed, the lognormal distribution possesses a number of interesting reproductive properties, most of which are immediate consequence of those for the normal distribution. Since the latter has additive reproductive properties, as discussed in §2, it is to be expected from the characteristic property of the logarithmic function $\log X_1 + \log X_2 = \log X_1 X_2$ that the lognormal distribution will have multiplicative reproductive properties. This is in fact the case. Therefore, the next paragraph provides the basis for the discussion of the genesis of the lognormal distribution from the product of random numbers and shows how random multiplication processes can also generate perfect power-law distributions with only slight modifications.

[21]In the limit case, $\alpha \to +\infty$ for $x \to +\infty$, *i.e.* the the distribution (15) behaves as an effective power-law when $x \to +\infty$. However, Huang & Solomon (2000) show that such a case is completely irrelevant for any finite system, because there is simply no variable x with such large values within the real (*e.g.*, financial) systems.

4. Models of Generation of Lognormal and Power-law Distributions

Interestingly enough, income distributions of various types can be obtained as steady-state solutions of stochastic processes. The stochastic theory based on the statistical law of probability, one of the oldest (and still popular) theories of distribution, relies for the skewed shape of income distributions mainly or solely on chance, luck, and random occurrences. The main authorship of this theory is attributed to Gibrat (1931), who viewed income dynamics as a multiplicative random process in which the product of a large number of individual random variables tends to the *lognormal* distribution. The connection between multiplicative processes and the lognormal distribution can be described as follows. Suppose that

$$x_t^i = \left(1 + r_{t-1}^i\right) x_{t-1}^i,$$

where x_t^i is the income of individual i in period t and $\left\{r_t^i\right\}_{t=0,1,\ldots,T-1}$ are the per-period rates of growth in the individual's income. Denoting the individual's per-period growth factors by $\lambda_t^i = \left(1 + r_t^i\right)$, the above expression can be rewritten as

$$x_t^i = \lambda_{t-1}^i x_{t-1}^i, \tag{17}$$

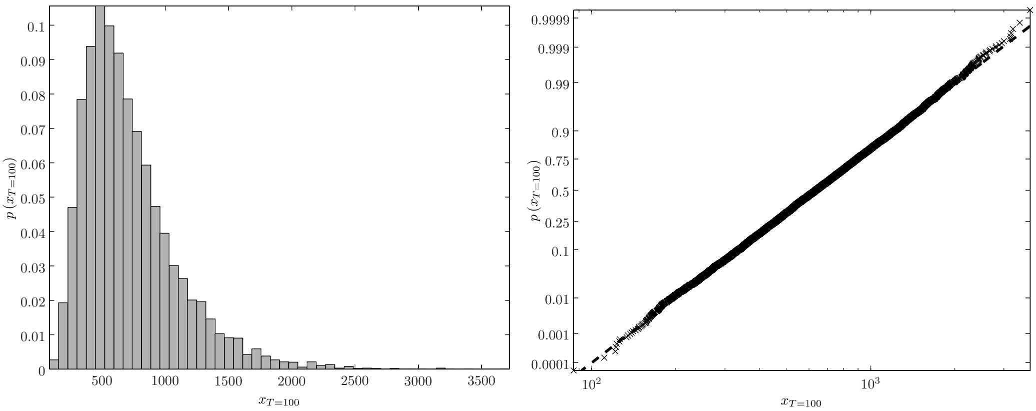
where $\left\{\lambda_t^i\right\}_{t=0,1,\ldots,T-1}$ are independent, identically distributed, continuous, nonnegative random variables. If one studies the above generic multiplicative process (17) with no additional constraints, one gets an ensemble of values x_{t-1} over all possible realizations of the multiplicative factors $\lambda_0, \lambda_1, \lambda_2, \ldots, \lambda_{T-1}$, which is distributed according to the lognormal distribution. Indeed, by iterating (17), the individual's income in period T is

$$x_T^i = x_0^i \lambda_0^i \lambda_1^i \lambda_2^i \cdots \lambda_{T-1}^i = x_0^i \prod_{j=0}^{T-1} \lambda_j^i, \tag{18}$$

where x_0^i represents the starting income of individual i. If the λ_t^i's are all governed by independent lognormal distributions, $\Lambda_i(\lambda)$,[22] then x_T^i is approximately lognormal, since the product of lognormal distributions is again lognormal.[23] However, lognormal distributions

[22]Generally, each individual may have a different distribution of per-period growth factors of his income—hence the subscript i in $\Lambda_i(\lambda)$, even if all the individuals drawn their per-period growth factors from the same distribution $\Lambda(\lambda)$. As recently shown by Levy (2003, 2005), only when *homogeneous* per-period growth factor distributions are assumed—*i.e.*, only when $\Lambda_i(\lambda) = \Lambda(\lambda)$ for all individuals i—the distribution of income converges to the empirically observed distribution, especially at the high-income range. When *heterogeneous* per-period growth factor distributions are assumed—*i.e.*, when $\Lambda_i(\lambda) \neq \Lambda_j(\lambda)$ for all individuals $i \neq j$, the resulting distribution of income is inconsistent with the empirical income distribution. An implication of these findings for the empirically observed Pareto's distribution at the high-income range (mainly influenced by capital investments) is that because individuals in the high-income range seem to have similar investment skills, at the margin it is only *luck* that differentiates between them. These results do not seem to depend on the functional form of the per-period growth factor distribution, $\Lambda(\lambda)$, and generally persist even when investors are finitely lived and their offspring's talents are different than their own.

[23]One example of a random multiplicative process might be wealth generation by investment. If individuals invest money, for instance in the stock market, they will get a percentage return on their investment that varies



(a) Binned distribution plot (b) Lognormal probability plot

Figure 7. The personal income distribution obtained from the simulation of the multiplicative stochastic process (17): (a) distribution plot of the elements into 50 equally spaced bins; (b) lognormal probability plot. λ_t^i is extracted from a Gaussian distribution with $\mu = 1.02$ and $\sigma = 0.05$. The system size is $N = 10,000$, and the measurements have been performed at time $T = 100$. The initial values of individual incomes have been all set to 100.

may be obtained even if the λ_t^is are not themselves lognormal by the additive form of the central limit theorem (see §2). Indeed, from the logarithm of (18)

$$\log x_T^i = \log x_0^i + \sum_{j=0}^{T-1} \log \lambda_j^i,$$

one gets via the CLT that $\sum_{j=0}^{T-1} \log \lambda_j^i$ converges to a normal distribution for sufficiently large t if the $\log \lambda_j^i$'s are independent and identically distributed variables with finite mean and variance; hence, for large times t, x_T^i is well approximated by a lognormal distribution. In the industrial organization and economic geography literature this result is also known as the Gibrat's *Law of Proportionate Effect*, stating that if growth rates of incomes in a fixed population (*i.e.*, abstracting from entry and exit dynamics) are independent of size and uncorrelated, the resulting distribution is lognormal. This is clearly seen in Figure 7(a)–(b), which exhibits the simulated distribution for the multiplicative system (17) by time $T = 100$, with λ_t^i drawn from a Gaussian distribution with $\mu = 1.02$ and $\sigma = 0.05$.[24] The simulation has been run for $N = 10,000$ agents, whose incomes has been initialized to 100. As one can easily recognize, the binned distribution and the logarithmic probability

over time. In other words, in each period of time their investment is multiplied by some factor which fluctuates from one period to the next. If the fluctuations are random and uncorrelated, then after many such periods the value of the investment is the initial value multiplied by the product of a large number of random numbers, and therefore should be distributed according to a lognormal.

[24]The value $\mu = 1.02$ ensures that the experimentally observed individual incomes increase in the overall system at each time step of the simulation, whereas the standard deviation $\sigma = 0.05$ is large enough to enable occasionally values which are smaller than 1.

plots show every sign of being lognormal.[25]

The same result can be also derived using the so-called Master and Fokker-Planck equations (Burda *et al*., 2003; Sornette, 2004). From (17), taking the logarithm of both sides, it is straightforward to obtain

$$y_{t+\tau} = y_t + l_t, \tag{19}$$

where $y_t = \log x_t$ and $l_t = \log \lambda_t$.[26] Expression (19) describes the problem of a random walker at the position y_t at time t making a random step of length l_t (positive or negative) to reach the position $y_{t+\tau}$ at time $t+\tau$.[27] Repeated backwards substitution allows to write the solution of (19) as

$$y_t = y_0 + l_{t-\tau} + l_{t-2\tau} + \ldots + l_\tau + l_0 = \sum_{i=1}^{t/\tau} l_{t-i\tau} = \sum_{i=1}^{N} l_{t-i\tau}, \tag{20}$$

where t is taken as a multiple of the elementary unit time τ, and $y_0 = 0$ is the origin point where the random walker started from at time zero. Expression (20) defines the variable y_t as the sum of $N \equiv \frac{t}{\tau}$ random variables. Since the l_i's are assumed uncorrelated, the average behavior of the random walk is immediately found to be

$$\mu = \langle y_t \rangle = \left\langle \sum_{i=1}^{N} l_i \right\rangle = \sum_{i=1}^{N} \langle l_i \rangle = \sum_{i=1}^{N} \langle l \rangle = N \langle l \rangle = \frac{t}{\tau} \langle l \rangle = vt,$$

where $l_i = l_{t-i\tau}$ and $v = \frac{\langle l \rangle}{\tau}$, while its fluctuations around the average are measured by

$$\sigma^2 = \left\langle y_t^2 \right\rangle - \langle y_t \rangle^2 = \sum_{i=1}^{N} \left[\left\langle l_i^2 \right\rangle - \langle l_i \rangle^2 \right] = \sum_{i=1}^{N} \left[\left\langle l^2 \right\rangle - \langle l \rangle^2 \right] = N \left[\left\langle l^2 \right\rangle - \langle l \rangle^2 \right] = 2Dt,$$

where the two definitions $N \equiv \frac{t}{\tau}$ and $D \equiv \frac{\left\langle l^2 \right\rangle - \langle l \rangle^2}{2\tau}$ have been used in the last computation. The constant D is the *diffusion* coefficient, and relates the variance of the sum of $\frac{t}{\tau}$ random i.i.d. variables, $\sigma^2 = \left\langle y_t^2 \right\rangle - \langle y_t \rangle^2$, to the variance $\left\langle l^2 \right\rangle - \langle l \rangle^2$ of the probability density function of the individual l-variables.

[25]The lognormal probability plot has its abscissa scale graduated logarithmically while, on the scale of ordinates, the distribution function $F(x)$ is plotted as its equivalent normal deviate. The best possible line through the plot points is given by

$$y = \mu + \sigma \Phi^{-1} \left[F(x) \right],$$

where $y = \log x$ and $\Phi^{-1}(\cdot)$ is the quantile function of the standard normal distribution.

[26]For the sake of simplicity, the superscript i has been omitted from expression (19).

[27]Since the individual income per-period growth factors are distributed according to the probability density function $\Lambda(\lambda)$ in (17), the steps l in (19) are distributed according to the density function $\Pi(l) = e^l \Lambda\left(e^l\right)$. This result is immediately found using the *rule of transformation of random variables*

$$p_y(y) = p_x \left[f^{-1}(y) \right] \left| \frac{df^{-1}(y)}{dy} \right|,$$

where $y = f(x)$ and $x = f^{-1}(x)$.

The *Master equation* describes how the probability distribution of the state variable, $p\,(y, t+\tau)$, varies with time. The Master equation corresponding to (19) is

$$p\,(y, t+\tau) = \int_{-\infty}^{+\infty} \mathrm{d}l\,\Pi\,(l)\,p\,(y-l, t)\,. \tag{21}$$

This expression states that, in order to be at y at time $t+\tau$, the walker was at some position $y-l$ at time t with probability $p\,(y-l, t)$ and has then just made the step of the correct length l and direction with probability $\Pi\,(l)$ to reach y at time $t+\tau$.[28] An approximation to (21) can be obtained by performing a Taylor expansion of $p\,(y-l, t)$ around the point $l=0$ up to second order, and of $p\,(y, t+\tau)$ around the point $\tau=0$ up to first order; then one gets

$$p\,(y-l, t) = p\,(y, t) - l\frac{\partial p\,(y, t)}{\partial y} + \frac{1}{2}l^2\frac{\partial^2 p\,(y, t)}{\partial y^2} + \mathcal{O}\,(l^3), \tag{22a}$$

$$p\,(y, t+\tau) = p\,(y, t) + \tau\frac{\partial p\,(y, t)}{\partial t} + \mathcal{O}\,(\tau^2), \tag{22b}$$

where $\mathcal{O}\,(l^3)$ and $\mathcal{O}\,(\tau^2)$ indicate the n^{th} order of the Taylor expansion, *i.e.* where the series has been truncated.[29] Replacing expressions (22a) and (22b) in (21) yields

$$\begin{aligned}
p\,(y, t) + \tau\frac{\partial p\,(y, t)}{\partial t} &= \int_{-\infty}^{+\infty} \mathrm{d}l\,\Pi\,(l)\left[p\,(y, t) - l\frac{\partial p\,(y, t)}{\partial y} + \frac{1}{2}l^2\frac{\partial^2 p\,(y, t)}{\partial y^2}\right] = \\
&= p\,(y, t)\int_{-\infty}^{+\infty} \mathrm{d}l\,\Pi\,(l) - \frac{\partial p\,(y, t)}{\partial y}\int_{-\infty}^{+\infty} \mathrm{d}l\,\Pi\,(l)\,l + \\
&\quad + \frac{1}{2}\frac{\partial^2 p\,(y, t)}{\partial y^2}\int_{-\infty}^{+\infty} \mathrm{d}l\,\Pi\,(l)\,l^2 = \\
&= p\,(y, t) - \langle l\rangle\,\frac{\partial p\,(y, t)}{\partial y} + \frac{\langle l^2\rangle}{2}\frac{\partial^2 p\,(y, t)}{\partial y^2}.
\end{aligned}$$

Therefore[30]

$$\frac{\partial p\,(y, t)}{\partial t} = -v\frac{\partial p\,(y, t)}{\partial y} + D\frac{\partial^2 p\,(y, t)}{\partial y^2}. \tag{23}$$

---

[28]Expression (21) uses the combination rule for independent events, $\Pr\,(A \cap B) = \Pr\,(A)\Pr\,(B)$; *i.e.*: the probability for two independent events A and B to occur is equal to the product of the corresponding probabilities.

[29]This means that all terms of order n or higher have been ignored, and that one should consider only terms of order $n-1$ or lower be of significance in the remainder of the calculation.

[30]Notice that, in the limit, $\frac{\langle l^2\rangle}{2\tau} \to D$. Indeed, by rewriting the coefficient $\frac{\langle l^2\rangle}{2\tau}$ of the second derivative in the r.h.s of (23) under the form $D + \frac{\langle l\rangle^2}{2\tau}$, since $\langle l\rangle = v\tau$ one gets that $\frac{\langle l\rangle^2}{2\tau} = \frac{v^2\tau}{2} \to 0$ for $\tau \to 0$ and $l \to 0$, and therefore $\frac{\langle l^2\rangle}{2\tau} \to D$.

In physics, this linear second-order partial differential equation of parabolic type is called the *Fokker-Planck equation*, and is usually employed to solve for the probability density.[31]

Since from the random walk (19) we get $\mu = vt$ and $\sigma = \sqrt{2Dt}$, one can check that the Gaussian distribution

$$p(y,t) = \frac{1}{\sqrt{4\pi Dt}} e^{-\frac{1}{4Dt}(y-vt)^2} \tag{24}$$

satisfies equation (23) by simple inspection. Indeed, the space derivatives of (24) are

$$\frac{\partial p(y,t)}{\partial y} = -\left(\frac{y-vt}{2Dt}\right) p(y,t), \tag{25a}$$

$$\frac{\partial^2 p(y,t)}{\partial y^2} = \left[\left(\frac{y-vt}{2Dt}\right)^2 - \frac{1}{2Dt}\right] p(y,t), \tag{25b}$$

while the time derivative is

$$\frac{\partial p(y,t)}{\partial t} = \left[v\left(\frac{y-vt}{2Dt}\right) + D\left(\frac{y-vt}{2Dt}\right)^2 - \frac{1}{2t}\right] p(y,t). \tag{26}$$

One can prove that (24) is a solution to the Fokker-Planck equation by substituting expressions (25a) ,(25b), and (26) into (23) and checking, after rearranging terms, that it is satisfied as identity. Therefore, the change of variable $y \to \log x$ in (24) (see footnote 27) gives the expression of the lognormal size distribution of personal income, that is the Gibrat's law

$$p(x,t) = \frac{1}{x\sqrt{4\pi Dt}} e^{-\frac{1}{4Dt}(\log x - vt)^2}. \tag{27}$$

In the discussion of Gibrat's Law of Proportionate Effect one can note that the working of the law is conceived as an ordered sequence of events in time. Indeed, an emphasis placed on the time sequence may lead to certain difficulties which are not easily resolved. A major weakness of the process (17) in fact is that, since time t enters multiplicatively, the standard deviation indicated in (27) increases steadily, as shown in Figure 8, which gives a comparison of the frequency curves obtained for time steps $t = 25, 50, 75$ and $T = 100$ from the same simulation runs of the multiplicative stochastic process (17) presented in Figure 7(a)–(b). This implication may be harmless in a number of cases, but in fields such as the study of the size distribution of incomes if the law operates at all the inequality of incomes measured by the standard deviation must continually increase, which is contrary to the evidence. Kalecki (1945) suggested a method of dealing with this deficiency by abandoning the assumptions of the process (17). He postulated that variations in the inequality

[31] Nearly every system is subjected to complicated external or internal influences that are not fully known and that are often termed "noise" or "fluctuations". The Fokker-Planck equation deals with those fluctuations of systems which stem from many tiny disturbances, each of which changes the variables of the system in an unpredictable way. Because of the fluctuations, one does not generally know the position of the state variable exactly, but instead he has a certain probability to find this variable in a certain region. With the Fokker-Planck equation such a probability density can be determined. In the mathematical literature, it is also called a *forward Kolmogorov equation*. See Risken (1984) for a derivation of the Fokker-Planck equation, the methods of solving it, and some of its applications.

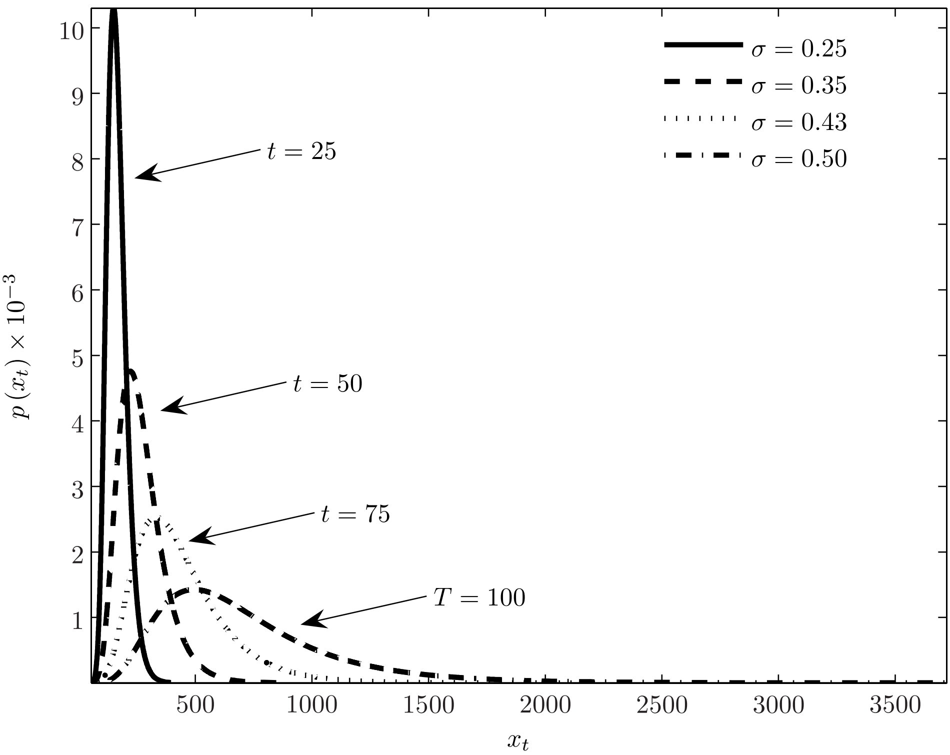


Figure 8. Frequency curves obtained for time steps $t = 25, 50, 75$ and $T = 100$ from the same simulation runs of the multiplicative stochastic process (17) presented in Figure 7(a)–(b). The fact that the standard deviation of $\log x_t$ is an increasing function of time t is clearly visible.

of incomes are to a great extent determined by economic forces, and studied the special case where the inequality of incomes remains constant through time by introducing a linear negative correlation between $\log x_{t-1}$ and $\log \lambda_{t-1}$ that prevents the standard deviation of $\log x_t$ from growing; *i.e.*, by omitting the superscript i to simplify the notation

$$\log \lambda_{t-1} = -\alpha_{t-1} \log x_{t-1} + \eta_{t-1}, \tag{28}$$

where η_{t-1} is independent of $\log x_{t-1}$. Therefore, the new generating equation becomes

$$\left. \begin{aligned} \log x_t &= -\alpha_{t-1} \log x_{t-1} + \eta_{t-1} + \log x_{t-1} = \\ &= (1 - \alpha_{t-1}) \log x_{t-1} + \eta_{t-1} \end{aligned} \right\} \Rightarrow x_t = x_{t-1}^{1-\alpha_{t-1}} e^{\eta_{t-1}}. \tag{29}$$

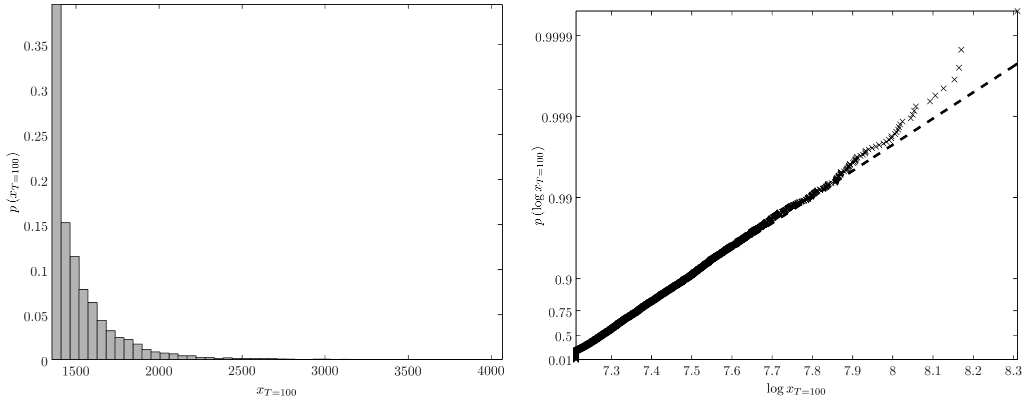
Under fairly general conditions, the generating model (29) returns a final distribution of x_t which is again approximately lognormal. The operation of the negative correlation implied by (28), however, seemed artificial until it received a justification from an unrelated source, namely Friedman (1957) permanent income hypothesis, according to which individuals who have had a high transitory component (included in λ) in the past are more likely to have a lower one now, thus yielding a negative serial correlation.

Lognormal and power-law distributions are intrinsically connected. Indeed, small variations in the underlying model can change the result from one to the other.[32] As a consequence, many variations of the model (17) have been developed in the literature. The main result of this strand of research was to show that even small variations from the pure multiplicative stochastic process lead to a *power-law* distribution. As a matter of example, Champernowne (1953) offered an explanation for the Pareto's power-law distribution of income similar in character to the Law of Proportionate Effect. His model, later on generalized and extended by Simon (1955), views income determination as a Markov process (income for the current period depends only on one's income for the last period and random influence) and depends on the subdivision of incomes into discrete ranges and the specification of a constant matrix of transition probabilities (otherwise, no stationary distribution will emerge from the Markov process) whose typical element (ij) states the probability that an income recipient whose income at time t lies in the i^{th} range will have income in the j^{th} range at time $t+1$. Champernowne showed that if the income intervals defining each class are assumed to form a geometric progression—that is, the limits of class j are higher than those of class $j-1$ by a certain percentage rather than a certain absolute amount of income, and if the transition probabilities $p_t\left(ij\right)$ depend only on t and the differences $j-i$—that is, the prospects for individuals to shift income ranges is independent of their current level of income, the equilibrium distribution tends to that given by the Pareto's law.[33]

The main difference between the multiplicative model (17) and the Champernowne's model is that while in the former income can become arbitrarily close to zero through successive decreases, in the latter model there is a minimum income corresponding to the lowest class below which one can not fall. Generally, for each economic system one can assume the existence of a positive cutoff $x^*_{t-1} > 0$ for the minimal income of each individual: *i.e.*, there is some threshold income which one has to posses to fulfill minimal needs and function in the system. In welfare economies it is provided by the social security system through the economic effects of subsidies, securities, and services. A very general extension to Champernowne's model is contained in Levy & Solomon (1996*a*,*b*), who showed that a power-law distribution can be obtained by adding a reflection condition to the stochastic multiplicative model (17), *i.e.* by assuming that each x^i_{t-1} is bounded from

[32] A rich and long history about generative models leading to either power-law or lognormal distributions, spanning many fields, can be found in work from decades ago. See, for instance, Aitchison & Brown (1957) and Sahota (1978); see also Kleiber & Kotz (2003), Mitzenmacher (2004) and Newman (2005) for pointers to some of the recent and historically relevant scientific literature on the subject.

[33] As previously discussed, the weakness of considering the generation of an equilibrium income distribution as a time process involving transition probabilities is that, in general, the variance of the final distribution increases as the process is continued, apparently in contradiction with the evidence (Aitchison & Brown, 1957). Following Champernowne, but with some variations, a number of others authors, including Rutherford (1955), Wold & Whittle (1957) and Steindl (1965), explained the Pareto's distribution and its stability by incorporating *birth and death* considerations into a Markov model according to which senior persons with growing dispersion of income retire and are replaced by young entrants with predominantly low and relatively uniform incomes. Under this assumption, in these models the overall variance remains constant over time. Moreover, it should also be noted that the Champernowne's and Gibrat's models and some others require long durations of time until the approach to stationarity is obtained. This point has been emphasized by Shorrocks (1975), who criticized previously developed stochastic models for concentrating on equilibrium distributions and proposed a model in which the transition probabilities or parameters of the distribution are allowed to change over time.



(a) Binned distribution plot (b) Pareto's probability plot

Figure 9. The final time step ($T = 100$) personal income distribution of the simulated multiplicative stochastic process (31): (a) distribution plot of the elements into 50 equally spaced bins; (b) Pareto's probability plot. λ_t^i, N and x_0^i are the same of the run presented in Figure 7(a)–(b). The graphs are obtained for fixed $c = 0.9$.

below to a threshold

$$x_{t-1}^* = \frac{c}{N} \sum_{i=1}^{N} x_{t-1}^i = c \langle x_{t-1} \rangle \tag{30}$$

proportional to current average income $\langle x_{t-1} \rangle$. Operationally, each time the generic multiplicative process (17) returns a value x_t^i smaller than x_{t-1}^*, the actual value of x_t^i is restored to

$$x_t^i = x_{t-1}^* = c \langle x_{t-1} \rangle ,$$

where the fraction c is fixed in time.[34] This mechanism can be also viewed as a way of killing off an individual and introducing a new one at the same time, thus incorporating a perfectly balancing "birth and death" process. Then formally the multiplicative model (17) now changes to

$$x_t^i = \begin{cases} \lambda_{t-1}^i x_{t-1}^i & \text{if } x_t^i \geq c \langle x_{t-1} \rangle, \\ c \langle x_{t-1} \rangle & \text{if } x_t^i < c \langle x_{t-1} \rangle. \end{cases} \tag{31}$$

Numerical results probing the properties of this multiplicative stochastic model with reflecting lower bound are shown in Figure 9(a)–(b) for the final time step $T = 100$ and values $c = 0.9$, $N = 10,000$ and $x_0^i = 100$. As in the previous simulation, the random factor λ_t^i is extracted from a Gaussian probability distribution with $\mu = 1.02$ and $\sigma = 0.05$. One can see that the lower cutoff in fact works, and the resulting distribution is no longer lognormal but can be approximated by a power function.[35]

Similarly to the case without the lower bound, one can think of the model (31) as of a random walk, which in the case of a cutoff has the lower barrier $y_t^* = \log x_t^* > 0$.

[34] Clearly, if c is time independent, x^* varies in time.

[35] The coordinates $(\log x_{i,n}; p_{i,n})$ on a Pareto's probability plot, $p_{i,n} = F(\log x_{i,n}) \in [0, 1]$ being the distribution function, follow immediately from the exponential case, since a log-transformed Pareto's random

Therefore, both the cases are described by the same partial differential equation (23), but with a different boundary condition. The stationary solution ($t \to +\infty$) of the Fokker-Planck equation (23) is immediately found to be

$$\frac{\partial p\left(y\right)}{\partial t}=0,$$

i.e.

$$D\frac{\mathrm{d}^2\,p\left(y\right)}{\mathrm{d}\,y^2}-v\frac{\mathrm{d}\,p\left(y\right)}{\mathrm{d}\,y}=Dp''\left(y\right)-vp'\left(y\right)=0. \tag{33}$$

Integrating both sides of (33), one gets

$$Dp'\left(y\right)-vp\left(y\right)=0$$

or

$$p'\left(y\right)+\alpha p\left(y\right)=0,$$

where $\alpha=-\frac{v}{D}$. The solution of this simple type of first order differential equation is

$$p\left(y\right)=Ce^{-\alpha y}, \tag{34}$$

where C is an arbitrary constant. In the presence of the barrier, the most obvious way to deal with expression (34) is to impose the normalization condition

$$\int_{y^*}^{+\infty}\mathrm{d}y\,p\left(y\right)=1,$$

which leads to the determination of the arbitrary constant

$$C=\alpha e^{\alpha y^*}. \tag{35}$$

By substituting expression (35) in (34) and rearranging terms, yields

$$p\left(y\right)=\alpha e^{-\alpha\left(y-y^*\right)},$$

and translating in the initial variable $x=e^y$

$$p\left(x\right)=\frac{\alpha x_*^{\alpha}}{x^{1+\alpha}},$$

which is the power-law distribution.

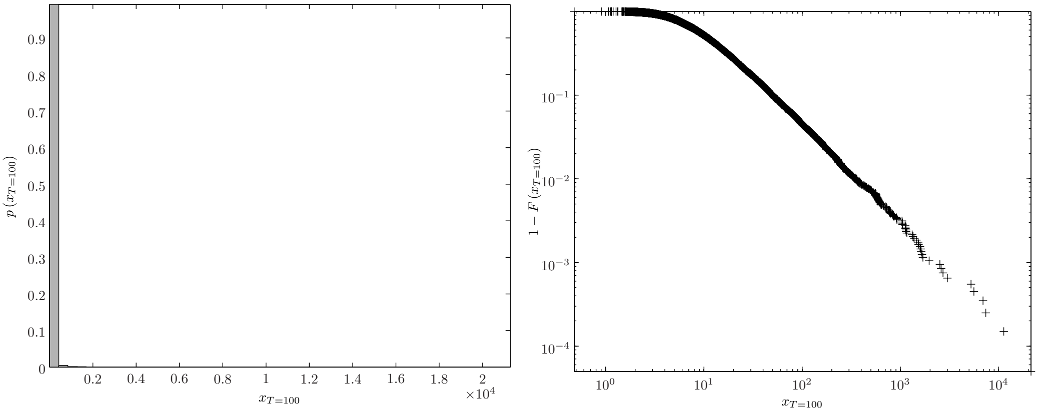
Kesten (1973) considered the following mixture of multiplicative and additive process

$$x_t^i=\lambda_{t-1}^i x_{t-1}^i+b_{t-1}^i, \tag{36}$$

variable is exponentially distributed. The straight line is given by

$$y=\frac{1}{\alpha}\left[-\log\left(1-k^{-\alpha p}\right)\right], \tag{32}$$

where $y=\log x$. The slope of the fitted line (32) can be used as an estimate of the parameter α^{-1}. See also footnote 25.



(a) Binned distribution (b) Cumulative distribution

Figure 10. The final time step ($T = 100$) personal income distribution of the simulated Kesten's model (36) for λ_t^i and b_t^i uniformly sampled in the interval $[0.48, 1.48]$ and in $[0, 1]$, respectively: (a) distribution plot of the elements into 50 equally spaced bins; (b) cumulative distribution (logarithmic scaling on both axes). The values of N and x_0^i are the same of Figure 7(a)–(b) and Figure 9(a)–(b).

$\left\{\lambda_t^i\right\}_{t=0,1,\ldots,T-1}$ and $\left\{b_t^i\right\}_{t=0,1,\ldots,T-1}$ being positive independent random variables. Clearly, for $b_{t-1}^i = 0$ the simple linear stochastic equation (36) recovers the model (17); for $b_{t-1}^i \neq 0$, it generates an ensemble of values x_t for which the power-law tail behavior

$$p\left(x_t\right) \propto x_t^{-(1+\alpha)}$$

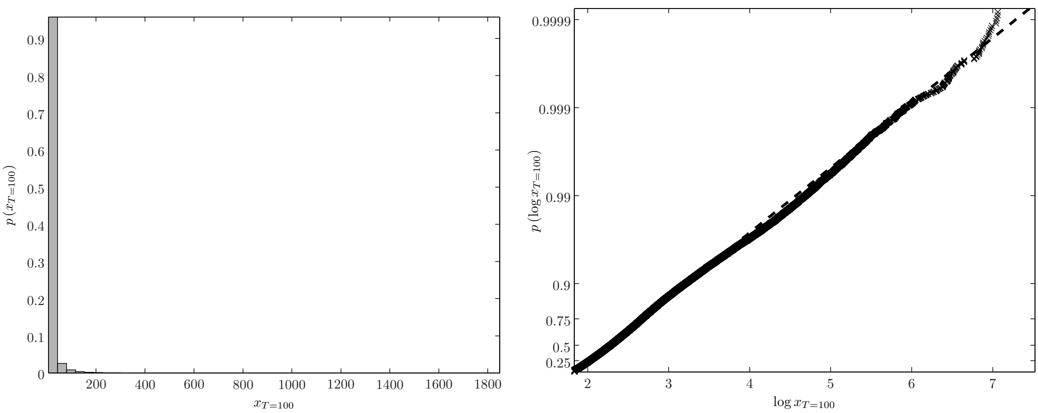
can be observed. The term b_{t-1}^i can be thought of as an effective repulsion from the origin—*i.e.*, a reinjection of the dynamics, and thus acts similarly to the barrier x_{t-1}^* in the previous model (31), making the multiplicative process with the reflective barrier and the Kesten's variable deeply related.[36] The reconstructed final time step ($T = 100$) density of the Kesten's variable (36) for λ_{t-1}^i and b_{t-1}^i uniformly sampled in the interval $[0.48, 1.48]$ and in $[0, 1]$, respectively, and values $N = 10,000$ and $x_0^i = 100$ is shown in Figure 10(a)–(b), where a clear power-law behavior in the tail can be observed from the double logarithmic plot of the cumulative distribution.

Finally, Blank & Solomon (2000) incorporate both entry and exit dynamics by assuming that the individuals disappear from the system if they fall below the threshold (30), and that at each period t

$$\Delta N = N_{t+1} - N_t = K\left(\sum_{i=1}^{N_{t+1}} x_{t+1}^i - \sum_{i=1}^{N_t} x_t^i\right) = K\left(x_{t+1}^{\mathrm{tot}} - x_t^{\mathrm{tot}}\right) \tag{37}$$

new individuals enter the system with the minimal size x_{t-1}^*. This model with a variable number of components whose income sizes evolve according to the multiplicative stochastic

[36] See Sornette (1998) on this subject; see also Sornette & Cont (1997) and Takayasu *et al*. (1997).



(a) Binned distribution plot

(b) Pareto's probability plot

Figure 11. The final time step ($T = 100$) personal income distribution of the simulated Blank and Solomon's model (37) for λ_t^i drawn from a Gaussian distribution with $\mu = 1.02$ and $\sigma = 0.05$ and values $c = 0.4$, $K = 0.1$, $x_0^i = 100$ and $N_0 = 100$: (a) distribution plot of the elements into 50 equally spaced bins; (b) Pareto's probability plot. The final system size is $N = 123,854$.

rule also lead to a power-law distribution, as one can easily recognize from Figure 11(a)–(b), which shows (a) the entire simulated binned distribution and (b) the Pareto's probability plot comparing the distribution of the simulated data to the power-law distribution for $T = 100$.[37]

5. Conclusion

A first goal of research on economic complexity has been the determination of the ways in which complex systems represent an alternative to standard (neoclassical) economic theory. At the same time, economic complexity has from its inception been strongly motivated by a desire to explain substantive empirical phenomena.

One of the main areas of work on the complexity/empirical interface consists of the identification of data patterns that are consistent with some of the features of complex en-

[37]Interestingly, there is another variation of the multiplicative model that yields a power-law behavior, namely the case where the time T itself is a random variable. (As an example, when considering income distribution data different age groups are usually intermixed and one may not know how long each income recipient has lived, so that the number of time steps undergone by each individual may be thought of as a random variable). This effect was firstly noticed by Montroll & Shlesinger (1982, 1983), who showed that a mixture of lognormal distributions based on a geometric distribution would have essentially a lognormal body but a power-law distribution in the tail. Assuming that time T is an exponential random variable and the number of multiplicative steps as being continuous, Reed (2003, 2004) and Reed & Jorgensen (2004) provide in more recent independent works an explanation and an extension for the above result which yields what they call a *double-Pareto distribution*, that is a distribution with one Pareto's tail for small values (below some point) and another Pareto's tail for large values (above the same point). This distribution closely matches the body of a lognormal distribution and the tail of a Pareto's distribution.

vironments (Durlauf, 2005). A major feature of this work has been the effort to identify where power-laws, which represent a particular class of probability distributions, and scaling laws, which describe relationships between variables that appear to be independent of the scale of measurement, occur in various economic data series (Brock, 1999).

The power-law and scaling literature has identified a number of interesting statistical properties of different economic data series, and has made a valuable contribution in identifying a range of "facts" that should help constrain theoretical modeling, such as the presence of skewed and thick tailed densities in data that confirms the existence of a pervasive heterogeneity over individual agents. Scaling laws are, in the context of complex systems, emergent properties, and so their presence would appear to speak to the empirical relevance of complexity. This can be interpreted as evidence of universal properties in economic data.

The attempt to identify the presence in economic data of certain statistical properties that are associated with complex systems has to a substantial extent led by physicists as there is a number of physical systems in which power and scaling laws are present. Indeed, within the physics community there has emerged a subfield known as "econophysics" in which a major research activity is represented by efforts to find power and scaling laws in different socioeconomic data sets. While much of this work focuses specifically on power-laws, it has also considered other probability distributions.

The primary focus of this research has been financial time series, because of the large quantities of data available at high frequencies that permits the evaluation of cross-section distributions over different time horizons.[38] However, a range of other data sets have also been explored. These include income distribution data, for which the general consensus seems to be this: Pareto's (power-law) distribution fits the upper tail of income distributions fairly well but other distributions such as the lognormal do better for the rest of the range. In the course of the survey, this chapter has tried to document this kind of evidence and the progress made in explaining income distribution via stochastic models.

References

Aitchison, J., & Brown, J. A. C. (1954). On criteria for descriptions of income distribution. *Metroeconomica*, **6**, 81–144.

Aitchison, J., & Brown, J. A. C. (1957). *The Lognormal Distribution with Special Reference to its Use in Economics*. New York: Cambridge University Press.

Arnold, B. C. (1983). *Pareto Distributions*. Fairland: International Co-operative Publishing House.

Bak, P. (1996). *How Nature Works: The Science of Self-Organized Criticality*. New York: Copernicus.

Bak, P., Tang, C., & Wiesenfeld, K. (1987). Self-organized criticality: An explanation of $1/f$ noise. *Physical Review Letters*, **59**, 381–384.

[38] This literature is well surveyed in the book by Mantegna & Stanley (1999).

Bak, P., Tang, C., & Wiesenfeld, K. (1988). Self-organized criticality. *Physical Review A*, **38**, 364–374.

Bertuglia, C. S., & Vaio, F. (2003). *Non linearità, caos, complessità. Le dinamiche dei sistemi naturali e sociali*. Torino: Bollati Boringhieri.

Blank, A., & Solomon, S. (2000). Power laws in cities population, financial markets and internet sites (scaling in systems with a variable number of components). *Physica A: Statistical Mechanics and its Applications*, **287**, 279–288.

Bresciani-Turroni, C. (1937). On Pareto's law. *Journal of the Royal Satistical Society*, **100**, 421–432.

Brock, W. A. (1999). Scaling in economics: A reader's guide. *Industrial and Corporate Change*, **8**, 409–446.

Burda, Z., Jurkiewicz, J., & Nowak, M. A. (2003). Is Econophysics a Solid Science? *Acta Physica Polonica B*, **34**, 87–131.

Champernowne, D. G. (1952). The graduation of income distributions. *Econometrica*, **20**, 591–615.

Champernowne, D. G. (1953). A Model of Income Distribution. *The Economic Journal*, **63**, 318–351.

Clauset, A., Shalizi, C. R., & Newman, M. E. J. (2007). *Power-Law Distributions in Empirical Data* (SFI Working Paper No. 07-12-049). Available from `http://www.santafe.edu/research/publications/wpabstract/200712049`

Clementi, F., & Gallegati, M. (2005*a*). Pareto's law of income distribution: Evidence for Germany, the United Kingdom, and the United States. In A. Chatterjee, S. Yarlagadda, & B. K. Chakrabarti (Eds.), *Econophysics of Wealth Distributions* (pp. 3–14). Milan: Springer-Verlag.

Clementi, F., & Gallegati, M. (2005*b*). Power law tails in the Italian personal income distribution. *Physica A: Statistical Mechanics and its Applications*, **350**, 427–438.

Cowell, F. A. (1995). *Measuring Inequality* (2nd ed.). Hemel Hempstead: Harvester Wheatsheaf.

Cowell, F. A., Ferreira, F. H. G., & Litchfield, J. A. (1998). Income distribution in Brazil 1981–1990: Parametric and non-parametric approaches. *Journal of Income Distribution*, **8**, 63–76.

Delli Gatti, D., Gaffeo, E., Gallegati, M., Giulioni, G., & Palestrini, A. (Eds.). (2008). *Emergent Macroeconomics. An Agent-Based Approach to Business Fluctuations*. Milan: Springer-Verlag.

Di Matteo, T., Aste, T., & Hyde, S. T. (2004). Exchanges in complex networks: Income and wealth distributions. In F. Mallamace & H. E. Stanley (Eds.), *The Physics of Complex Systems (New Advances and Perspectives)* (pp. 435–442). Amsterdam: IOS Press.

Drăgulescu, A. A., & Yakovenko, V. M. (2001*a*). Evidence for the exponential distribution of income in the USA. *The European Physical Journal B*, **20**, 585–589.

Drăgulescu, A. A., & Yakovenko, V. M. (2001*b*). Exponential and power-law probability distributions of wealth and income in the United Kingdom and the United States. *Physica A: Statistical Mechanics and its Applications*, **299**, 213–221.

Durlauf, S. N. (2005). Complexity and empirical economics. *The Economic Journal*, **115**, 225–243.

Feller, W. (1971). *An Introduction to Probability Theory and its Applications* (2nd ed., Vol. II). New York: John Wiley & Sons.

Friedman, M. (1957). *A Theory of the Consumption Function*. Princeton: Princeton University Press.

Gibrat, R. (1931). *Les inégalités économiques. Applications: Aux inégalités des richesses, à la concentration des entreprises, aux populations des villes, etc., d'une loi nouvelle: La loi de l'effet proportionel*. Paris: Librairie du Recueil Sirey.

Gini, C. (1914). Sulla misura della concentrazione e della variabilità dei caratteri. *Atti del Reale Istituto Veneto di Scienze, Lettere ed Arti*, **LXXIII**, 1203–1248. (English translation in *Metron*, *73*, 3–38, 2005.)

Gini, C. (1921). Measurement of inequality of incomes. *The Economic Journal*, **31**, 124–126.

Hayakawa, M. (1951). The application of Pareto's law of income to Japanese data. *Econometrica*, **19**, 174–183.

Huang, Z.-F., & Solomon, S. (2000). Finite market size as a source of extreme wealth inequality and market instability. *Physica A: Statistical Mechanics and its Applications*, **294**, 503–513.

Johnson, N. L., Kotz, S., & Balakrishnan, N. (1994). *Continuous Univariate Distributions* (Vol. 2nd). New York: John Wiley & Sons.

Kalecki, M. (1945). On the Gibrat distribution. *Econometrica*, **13**, 161–170.

Kesten, H. (1973). Random difference equations and renewal theory for products of random matrices. *Acta Mathematica*, **131**, 207–248.

Kirman, A. P. (1992). Whom or what does the representative individual represent? *Journal of Economic Perspectives*, **6**, 117–136.

Kirman, A. P. (1999). Interaction and markets. In M. Gallegati & A. P. Kirman (Eds.), *Beyond the Representative Agent.* Cheltenham: Edward Elgar.

Kleiber, C., & Kotz, S. (2003). *Statistical Size Distributions in Economics and Actuarial Sciences*. New York: John Wiley & Sons.

Levy, M. (2003). Are rich people smarter? *Journal of Economic Theory*, **110**, 42–64.

Levy, M. (2005). Is risk-aversion hereditary? *Journal of Mathematical Economics*, **41**, 157–168.

Levy, M., & Solomon, S. (1996*a*). Power laws are logarithmic Boltzmann laws. *International Journal of Modern Physics C*, **7**, 595–601.

Levy, M., & Solomon, S. (1996*b*). Spontaneous scaling emergence in generic stochastic systems. *International Journal of Modern Physics C*, **7**, 745–751.

Lévy, P. (1954). *Théorie de l'addition des variables aléatoires*. Paris: Gauthier-Villars.

Lorenz, M. O. (1905). Methods of measuring the concentration of wealth. *Publications of the American Statistical Association*, **9**, 209–219.

MacGregor, D. H. (1936). Pareto's law. *The Economic Journal*, **46**, 80–87.

Mandelbrot, B. (1960). The Pareto-Lévy law and the distribution of income. *International Economic Review*, **1**, 79–106.

Mandelbrot, B. (1963). New methods in statistical economics. *The Journal of Political Economy*, **71**, 421–440.

Mantegna, R. N., & Stanley, H. E. (1999). *An Introduction to Econophysics: Correlations and Complexity in Finance*. Cambridge: Cambridge University Press.

Mitzenmacher, M. (2004). A brief history of generative models for power law and lognormal distributions. *Internet Mathematics*, **1**, 226–251.

Montroll, E. W., & Shlesinger, M. F. (1982). On $1/f$ Noise and Other Distributions with Long Tails. *Proceedings of the National Academy of Sciences USA*, **79**, 3380–3383.

Montroll, E. W., & Shlesinger, M. F. (1983). Maximum Entropy Formalism, Fractals, Scaling Phenomena, and $1/f$ Noise: A Tale of Tails. *Journal of Statistical Physics*, **32**, 209–230.

Newman, M. E. J. (2005). Power laws, Pareto distributions and Zipf's law. *Contemporary Physics*, **46**, 323–351.

Nirei, M., & Souma, W. (2007). A two factor model of income distribution dynamics. *Review of Income and Wealth*, **53**, 440–459.

Nolan, J. P. (1998). Parameterizations and modes of stable distributions. *Statistics and Probability Letters*, **38**, 187–195.

Nolan, J. P. (2007). *Stable Distributions - Models for Heavy Tailed Data*. Boston: Birkhäuser. (In progress, Chapter 1 online at `http://academic2.american.edu/~jpnolan/stable/stable.html`)

Pareto, V. (1964). Course d'économie politique. In G.-H. Bousquet & G. Busino (Eds.), *Œuvres complètes de Vilfredo Pareto, Tome 1*. Geneva: Libraire Droz.

Pareto, V. (1965). La courbe de la repartition de la richesse. In G. Busino (Ed.), *Œuvres complètes de Vilfredo Pareto, Tome 3: Écrits sur la courbe de la répartition de la richesse* (pp. 1–15). Geneva: Librairie Droz. (English translation in *Rivista di Politica Economica*, *87*, 647–700, 1997.)

Persky, J. (1992). Retrospectives: Pareto's law. *The Journal of Economic Perspectives*, **6**, 181–192.

Reed, W. J. (2003). The pareto law of income—an explanation and an extension. *Physica A: Statistical Mechanics and its Applications*, **319**, 469–486.

Reed, W. J. (2004). On Pareto's law and the determinants of Pareto exponents. *Journal of Income Distribution*, **13**, 7–17.

Reed, W. J., & Jorgensen, M. (2004). The double Pareto-lognormal distribution—A new parametric model for size distributions. *Communications in Statistics - Theory and Methods*, **33**, 1733–1753.

Risken, H. (1984). *The Fokker-Planck Equation: Methods of Solutions and Applications*. Berlin: Springer-Verlag.

Rutherford, R. S. G. (1955). Income distributions: A new model. *Econometrica*, **23**, 277–294.

Sahota, G. S. (1978). Theories of personal income distribution: A survey. *Journal of Economic Literature*, **16**, 1–55.

Shirras, G. F. (1935). The Pareto law and the distribution of income. *The Economic Journal*, **45**, 663–681.

Shorrocks, A. F. (1975). On stochastic models of size distributions. *Review of Economic Studies*, **42**, 631–641.

Silva, A. C., & Yakovenko, V. M. (2005). Temporal evolution of the "thermal" and "superthermal" income classes in the USA during 1983–2001. *Europhysics Letters*, **69**, 304–310.

Simon, H. A. (1955). On a class of skew distribution functions. *Biometrika*, **42**, 425–440.

Sornette, D. (1998). Multiplicative processes and power laws. *Physical Review E*, **57**, 4811–4813.

Sornette, D. (2004). *Critical Phenomena in Natural Sciences. chaos, Fractals, Selforganization and Disorder: Concepts and Tools*. Berlin: Springer-Verlag.

Sornette, D., & Cont, R. (1997). Convergent multiplicative processes repelled from zero: Power laws and truncated power laws. *Journal de Physique I France*, **7**, 431–444.

Souma, W. (2001). Universal structure of the personal income distribution. *Fractals*, **9**, 463–470.

Souma, W. (2002). Physics of personal income. In H. Takayasu (Ed.), *Empirical Science of Financial Fluctuations: The Advent of Econophysics* (pp. 343–352). Tokyo: Springer-Verlag.

Stamp, J. C. (1914). A new illustration of Pareto's law. *Journal of the Royal Statistical Society*, **77**, 200–204.

Steindl, J. (1965). *Random Processes and the Growth of Firms: A Study of the Pareto Law*. New York: Hafner Press.

Takayasu, H., Sato, A.-H., & Takayasu, M. (1997). Stable infinite variance fluctuations in randomly amplified langevin systems. *Physical Review Letters*, **79**, 966–969.

Tesfatsion, L. S. (2000). *Agent-Based Computational Economics: A Brief Guide to the Literature* (Computational Economics No. 0004001). EconWPA. Available from `http://ideas.repec.org/p/wpa/wuwpco/0004001.html`

Tesfatsion, L. S. (2001). Guest editorial: Agent-based modeling of evolutionary economic systems. *IEEE Transactions on Evolutionary Computation*, **5**, 437–441.

Tesfatsion, L. S. (2002*a*). *Agent-Based Computational Economics* (Computational Economics No. 0203001). EconWPA. Available from `http://ideas.repec.org/p/wpa/wuwpco/0203001.html`

Tesfatsion, L. S. (2002*b*). Agent-based computational economics: Growing economies from the bottom up. *Artificial Life*, **8**, 55–82.

Tesfatsion, L. S. (2002*c*). Economic agents and markets as emergent phenomena. *Proceedings of the National Academy of Sciences USA*, **99**, 7191–7192.

Tesfatsion, L. S. (2003). Agent-based computational economics: Modeling economies as complex adaptive systems. *Information Sciences*, **149**, 262–268.

Tesfatsion, L. S., & Judd, K. L. (Eds.). (2006). *Handbook of Computational Economics, Volume 2: Agent-Based Computational Economics*. Amsterdam: North-Holland.

Willinger, W., Alderson, D., Doyle, J. C., & Li, L. (2004). More "normal" than normal: Scaling distributions and complex systems. In R. G. Ingalls, M. D. Rossetti, J. S. Smith, & B. A. Peters (Eds.), *Proceedings of the 2004 Winter Simulation Conference* (pp. 130–141). Piscataway: IEEE Press.

Willis, G., & Mimkes, J. (2004). *Evidence for the Independence of Waged and Unwaged Income, Evidence for Boltzmann Distributions in Waged Income, and the Outlines of a Coherent Theory of Income Distribution* (Microeconomics No. 0408001). EconWPA. Available from `http://ideas.repec.org/p/wpa/wuwpmi/0408001.html`

Wold, H. O. A., & Whittle, P. (1957). A model explaining the Pareto distribution of wealth. *Econometrica*, **25**, 591–595.

Wright, I. (2005). The social architecture of capitalism. *Physica A: Statistical and Theoretical Physics*, **346**, 589–620.

In: Emerging Topics in Macroeconomics
Editor: Richard O. Bailly, pp. 75-96
ISBN 978-1-60692-093-0

Chapter 2

INFLATION FEEDBACK REGULATION UNDER DEMAND SHOCKS AND UNCERTAIN POTENTIAL MACROECONOMIC OUTPUT*

José Alvarez-Ramirez[1] ***and Guillermo Fernández-Anaya***[2†]
[1]Ingeniería de Procesos e Hidráulica,
Universidad Autónoma Metropolitana-Iztapalapa,
Av. San Rafael Atlixco 186,
Col. Vicentina, México, D. F. 09340, MÉXICO
[2]Departamento de Física y Matemáticas,
Universidad Iberoamericana, Prol. Paseo de la Reforma 880,
Lomas de Santa Fe, México, D. F. 01219, MÉXICO

Abstract

Inflation control is the main task of central banks in modern economies. Some attempts have been taken in the recent literature to design systematic feedback rule to regulate inflation at a prescribed level. The Taylor rule isd the simplest instrumental rule to guide monetary policy to control inflation where the instrument ($e.g.$, a short interest rate) responds to changes in the inflation and the output gaps. The objective of this work is to modify the Taylor rule in order to improve its robustness with respect to uncertainties about potential output and unanticipated shocks. To this end, departing from feedback control theory, the Taylor rule is equipped with an adaptive control scheme to reject the adverse effects of shocks and to estimate the deviations of the potential output. It is shown that the proposed adaptation procedure is equivalent to a classical integral feedback controller whose characteristics and implementation issues are well understood in practical control engineering. Singular perturbation methods are used to establish the stability properties of the resulting control system.

*Partial financial support from *Fundacion Simona Lagata* is acknowledged
†E-mail address: guillermo.fernandez@uia.mx

1. Introduction

One of the main goals of modern central banks is the design of successful monetary policies for stabilizing inflation about a prescribed level, with concern for stabilizing output around the intrinsic potential output of the economy. In general, monetary policy establishes the value of the central bank's instrument ($e.g.$, a short interest rate) in response to changes in inflation and output gaps [1]. In order to establish the instrument value, the central bank proceeds along the following steps: i) A desired inflation level, denoted by $\pi_{sp} > 0$, is defined according to certain macroeconomic criteria. Commonly[1], it is suggested that $\pi_{sp} = 2\%$. ii) A small subset of information, including inflation and output gaps, is measured. iii) Based on this information, the central bank adjusts the instrument level in a direction intended to close the inflation and the output gaps. iv) Since the economy is commonly faced with unanticipated supply and demand shocks, the instrument level is reviewed periodically in order to reject the adverse effects of these shocks. This is done by carrying out recursively the steps i) to iii). Roughly, this procedure is a feedback control loop [2-4] where the central bank uses the instrument level as a control input to achieve inflation stabilization about the prescribed ($i.e.$, setpoint) value π_{sp}. That is, by means of a given algorithm ($i.e.$, the feedback control algorithm), the central bank corrects the instrument level in response to variations in inflation and output gaps.

Usually, the procedure that central banks use to decide on the monetary policy is not given in a "mathematical" form. A committee ($e.g.$, a board of governors) judges the available information, including inflation and output gaps and other macroeconomic signals, and, based on prevailing economic theories, the consideration of fundamentals, and expertise, the instrument level is determined. The committee meets periodically in order to correct for deviations in the inflation and output gaps due to, $e.g.$, supply and demand shocks. In this form, it can be argued that the committee has an internal model of the economy, which can be composed of mathematical, algorithmic and handcrafted models, by means of which beneficial and adverse effects of the target instrument level and supply/demand shocks are evaluated. The impressive performance of the USA economy during the past two decades seems to demonstrate the merits and efficacy of prescribing an inflation target π_{sp} and achieving it via a monetary policy based on available macroeconomic information. As mentioned above, this can be located from a methodological viewpoint within the core of what is called as feedback control of dynamical systems [2-4]. As a matter of fact, stabilization and regulation based on *feedback* control methodologies have been the main tool that made possible the practical development of modern technologies, ranging from electronic devices to chemical processes, and moreover, its concepts and methods are of a wide range of applicability, much beyond the engineering framework. In this way, it is not surprising that feedback-based methodologies used by monetary policy committees have been successful in stabilizing macroeconomic activity.

The main criticism some researchers have made to the way central banks proceed in order to determine monetary policy is that decision procedures are not explicit, relying commonly on the expertise (and sometimes bias) of committee participants [5-7]. It has been argued that, if the decision procedures are not explicit and transparent, the monetary policy can have an intrinsic fragility component in the sense that decisions are strongly de-

[1] See, for instance, recent Federal Reserve Board's Monetray Policy Reports to Congress

pendent on the presence or absence of certain personalities ($e.g.$, Alan Greenspan in the Fed). In this way, from a practical viewpoint, it would be desirable to dispose of explicit and systematic procedures for the conduct of monetary policy that is less dependent of discretionary decisions and the particular composition of central bank committees [8]. In principle, an explicit feedback-based procedure can be used as a guideline for conducting monetary policy. In fact, proponents of explicit procedures for monetary policy believe that such a scheme can systematically and compactly summarize much of the relevant information in a manner that potentially provides a convenient benchmark for consideration of policy settings [9]. On the other hand, proponents of the secrecy aspects argue that the latter are necessary in order to avoid that speculators, monopolists, etc., anticipate and neutralize to some extent, the authorities moves.

Early attempts to formulate feedback policies for inflation ($i.e.$, consumer price index) stabilization can be traced back to Ricardo [10], Wicksell [11] and Fisher [12]. Phillips [13] introduced the use of formal models by proposing the classical proportional, integral and derivative stabilization mechanisms. The efforts of Black [14], Fischer [15], McCallum [16], Phelps and Taylor [17] and Benhabib [18] proposed stabilization policies from a rational-expectations perspective [19]. A recent revival of interest in feedback policies is largely attributable to McCallum's work [20-22]. As pointed out by Biederman [23], much of the current debate centers on the proposition that a nation's central bank should adopt an inflation target π_{sp}, and then employ an explicit *policy rule*, as opposed to the use of discretion, to try to achieve the target objective. In an attempt to explain the behavior of the federal funds rate as prescribed by the Fed, Taylor [24] proposed a simple instrument rule, where the instrument rate responds only to the inflation and output gaps according to

$$i_t = \pi_{sp} + f_\pi(\pi_t - \pi_{sp}) + f_y(y_t - y_p) \tag{1}$$

where i_t is the instrument rate in period t, $\pi_t - \pi_{sp}$ is the inflation gap, π_t being the observed rate of inflation, y_t is the output, y_p is the potential output, and $y_t - y_p$ is the output gap. The parameters f_π and f_y are respectively the inflation and output gap gains, which are proposed to be positive constants [24]. The simplicity of the instrument rule (1) (known as Taylor rule) has motivated a large amount of research to support it and to extend its structure within an optimization framework. Commonly, the objective function is formulated as an inter-temporal (quadratic) loss function to be minimized, consisting of the expected sum of discounted current and future losses, $e.g.$,

$$\frac{(1-\delta)}{2}\sum_{\tau=0}^{\infty}\delta^\tau\left[(\pi_t - \pi_{sp})^2 + \lambda(y_t - y_p)^2\right] \tag{2}$$

where $\delta \in (0,1)$ is a discount factor and $\lambda > 0$ is a given weight on output-gap stabilization relative to inflation stabilization. In this form, using either backward- or forward-looking models of transmission mechanism, formal support and extensions to the simple Taylor rule (1) have been obtained [25-28]. In a recent paper, Svensson [1] has provided an excellent review of optimization-based approaches to derive monetary policy rules. However, instrument rules that implement the optimal plan for some given objective are generally very complicated. Because simple rules are more transparent, they are more likely to be understood by private agents, thereby increasing the chance that a committed policy will reap its benefits. Hence, a central bank planning to use explicit instrument rules should, at least during the initial implementation stages, rely on simple rules. As stated above, the Taylor rule (1) is still the starting point to investigate, both from the theoretical and practical perspectives, the applicability and robustness of explicit instrument rules in real economies [1]. In particular, the simple Taylor rule has shown to be robust against model uncertainties and time-varying dynamics [29,30]. Moreover, in some cases a well-tuned Taylor rule can perform better than instrument rules derived from a loss function optimization problem (see, for instance, [30] and [31] and references therein). In this way, studies about the robust stabilization capability of Taylor rules should lead to *improved* and systematic instrument rule designs [1,30].

Some criticisms have been made regarding the practical implementation of the Taylor rule, at least as a guideline for central bank's policy decisions. First, it lacks a mechanism for effectively deal with unanticipated supply/demand shocks. In fact, if the Taylor rule (1) were implemented as it stands, a sustained shock can lead to biased inflation stabilization, so that the inflation target can not achieved. Second, the implementation of (1) requires knowledge of the potential output y_p. Recently, the problems associated with the measurement of potential output have been highlighted in the monetary policy debate in the USA. It has been concluded that estimates of potential output y_p are highly uncertain in real economies affecting seriously the performance of monetary policies [32-35]. These drawbacks impose important limits to the applicability of explicit monetary policies. The present work focuses on the issue of improving the robustness of the Taylor rule (1) to deal with problems of sustained shocks and uncertain potential output. To this end, we resort to existing methodologies in feedback control theory [2-4]. As a first step, the Taylor rule is equipped with a integral/summation feedback compensator to reject the bias effects of non-vanishing shocks. In this way, convergence of the inflation π_t to the target value π_{sp} can be guaranteed. To deal with uncertainties in the potential output y_p, the Taylor rule is endowed with an adaptation rule to learn the deviation of the estimated potential output from the real potential output. The final result is a robust rule retaining the main characteristics of the Taylor rule (1); namely, simplicity and little prior knowledge of the economy dynamics. The stability analysis of the resulting controlled economy is made by means of singular perturbation methodologies.

With respect to existing results in the literature [1,17-18,20-24], our general contribution can be summarized as follows:

- It is shown that simple learning/adaptive schemes can be incorporated into the Taylor rule to correct bias effects induced by unanticipated demand shocks and errors in the (long-term) potential output measurement.

- Along the lines of Onatski and Stock's work [30], it is shown that robust feedback control methodologies can be used to enhance the stabilization properties of proposed monetary policy rules, including both simple rules and rules derived from optimizing a loss function.

In this way, our results suggest that existing results in the feedback control literature can help to understand the dynamics and to design simple robust stabilizing strategies for uncertain economic systems.

2. Problem Statement

Notation. A matrix A is Schur-stable if all its eigenvalues are located in the open unit disc of the complex plane. In this paper, a steady-state equilibrium point x_{eq} will correspond to a fixed point of a dynamical system $x_{t+1} = Ax_t + B$ ($i.e.$, $x_{eq} = (I - A)^{-1}B$ as far the matrix $I - A$ is non-singular). The transpose of a $x \in \mathbb{R}^n$ will be denoted by x^T.

The main contribution of this paper is the modification of the basic structure of the Taylor rule (1) to deal with unanticipated shocks and uncertainties in the potential output. In this form, in order to make clear such modifications and to avoid unnecessary algebraic complexity, in the sequel the analysis will be based on a simple backward-looking model retaining the main stylized features of macroeconomic dynamics. In general, our results can be extended to the case of higher order models after some straightforward modifications.

The model under consideration is similar to that used by Taylor [36], which was motivated from the stylized fact that both inflation and aggregate demand have a lagged reaction to changes in the central bank's instrument, and the lag for inflation is longer than for aggregate demand. In terms of the inflation, π_t, and the output, y_t, dynamics, these stylized facts can be codified in the following equations:

$$\begin{aligned} \pi_{t+1} &= \pi_t + a_1(y_t - y_p) + s \\ y_{t+1} &= y_p + a_2(y_t - y_p) - b(i_t - \pi_t) + d \end{aligned} \tag{3}$$

where i_t is the central bank's instrument (for instance, the federal funds rate in the USA, or a repo rate in several other countries), s is the unanticipated supply shock and d is the unanticipated demand shock. The parameters a_1, a_2 and b are positive constants. The first equation describes a so-called accelerationist Phillips curve, where the change in inflation depends on output with a lag of one time period. The second equation describes an aggregate demand/IS curve, where output depends on lagged output gap and real interest rate[2]. The rationale behind the Taylor rule (1) is implicitly captured in this model: when the output gap is positive and puts an upward pressure on inflation, a central bank can bring inflation back down by increasing the nominal interest rate to push up the real interest rate [37].

Remark 2.1. *Model (3) can be seen as a simplification of more realistic backward looking economy models. More general models involving only inflation and output dynamics can have the following structure:*

$$\begin{aligned} \pi_{t+1} &= \sum_{j=0}^{N_\pi - 1} \alpha_j \pi_{t-j} + a_1(y_t - y_p) + s_t \\ y_{t+1} &= y_p + \sum_{j=0}^{N_y - 1} a_{2,j}(y_{t-j} - y_p) - b\left(\sum_{j=0}^{M_i} \beta_j i_{t-j} - \overline{\pi}_t\right) + d_t \end{aligned} \tag{4}$$

where the α_j's are constants satisfying $\sum_{j=0}^{N_\pi - 1} \alpha_j = 1$ and $\overline{\pi}_t$ is an average inflation at time t, and a_1, b and the $a_{2,j}$'s are positive constants. In this way, inflation depends on

[2]It is noted that, if current inflation is taken as a proxy for expected inflation, $i_t - \pi_t$ is a proxy for the real interest rate.

previous inflation with N_π lags, while output depends on previous output with N_y lags and on expected inflation with M_π leads. Besides, supply and demands shocks in Eq. (4) are of time-varying nature. In this form, model (3) is particular case of model (3) with $N_\pi = 1$, $N_y = 1$, $M_\pi = 1$, $\overline{\pi}_t = \pi_t$ and constant supply and demand shocks ($i.e.$, $s_t = s$ and $d_t = d$). That is, the simplified model (3) considers that the economy is subjected to a non-vanishing supply/demand shock. Hence, the aim of a stabilizing instrument rule should be to reject the adverse effects of such shocks in the inflation dynamics.

Remark 2.2. *Rudebusch and Svensson [26] used the model structure (4) to describe the dynamics of the USA economy. In such case, they used $N_\pi = 4$, $N_y = 2$, $M_i = 4$ and $M_\pi = 1$, with $\beta_j = 1/4$ and*

$$\overline{\pi}_t = (\pi_t + \pi_{t-1} + \pi_{t-2} + \pi_{t-3})/4$$

That is, the expected inflation is an average of past inflation using four time-periods.

Remark 2.3. *Notice that, since we are going to deal with uncertain potential output, the model (3) has been expressed in inflation π and output y coordinates, rather than in inflation π and output gap $y - y_p$ coordinates, as is usual in most inflation control literature. In this form, uncertainties in the potential output can be displayed in an explicit way.*

Let us motivate the Taylor rule re-design problem as follows. Suppose that the monetary policy $i_t = \pi_{sp}$, for all $t \geq 0$, is chosen. It is easy to see that, in the shock-free case ($i.e.$, $s = d = 0$), the steady-state equilibrium point, denoted by π_{eq} and y_{eq}, of the system (3) is given by $y_{eq} = y_p$ and $\pi_{eq} = \pi_{sp}$. If one assumes that the economy is inherently stable in the sense that the matrix

$$M \stackrel{def}{=} \begin{pmatrix} 1 & a_1 \\ b & a_2 \end{pmatrix} \tag{5}$$

is Schur-stable, then the convergence $(\pi_t, y_t) \to (\pi_{sp}, y_p)$ is ensured. In this form, the so-simple assignation $i_t = \pi_{sp}$ leads stabilization of the economic dynamics about the desired inflation value π_{sp}. That is, the output y achieves its potential value y_p, and the inflation becomes equal to the target value π_{sp}. However, this naive approach has the following drawbacks:

i) Although the matrix M is assumed to be Schur-stable, one of its eigenvalues can very close to the boundary of the unit circle. In this way, the convergence $(\pi_t, y_t) \to (\pi_{sp}, y_p)$ can be unacceptably slow in practice.

ii) Real economies are subject to unanticipated shocks, so that in general $s \neq 0$ and/or $d \neq 0$. In this case, although the economy could be sufficiently stable, (π_{sp}, y_p) is no longer a steady-state equilibrium point. In fact, with non-zero shocks the equilibrium point is given by

$$\begin{aligned} \pi_{eq} &= \pi_{sp} + \tfrac{1}{b}\left(\tfrac{a_2-1}{a_1}s - d\right) \\ y_{eq} &= y_p - \tfrac{s}{a_1} \end{aligned} \tag{6}$$

That is, sustained shocks induce an inflation steady-state bias, given by $\left|\frac{1}{b}\left(\frac{a_2-1}{a_1}s-d\right)\right|$.[3]

The first issue can be alleviated by enhancing the dynamical characteristics of the economy by means of a feedback-based reaction function for the interest rate, as done by the Taylor rule. The stability of the controlled economy is governed by the Schur-stability of the matrix

$$M_c \stackrel{def}{=} \begin{pmatrix} 1 & a_1 \\ b(1-f_\pi) & a_2 - bf_y \end{pmatrix} \tag{7}$$

and the corresponding characteristic polynomial is given by

$$P_c(z) = z^2 + [bf_y - a_2 - 1]\, z + [a_2 - bf_y - ba_1\,(1 - f_\pi)] \tag{8}$$

A suitable selection of the instrument rule parameters f_π and f_y can move the eigenvalues of the matrix M_c further inside the unit circle; thus increasing the convergence rate of the controlled economy by virtue of the feedback control action $f_\pi(\pi_t - \pi_{sp}) + f_y(y_t - y_p)$ contained in the Taylor rule[4]. However, the structure of the Taylor rule (1) does not suffice to reject the adverse effects of supply/demand shocks. In fact, the controlled dynamics of the economy (3) under the Taylor rule (1) is given by

$$\begin{aligned} \pi_{t+1} &= \pi_t + a_1(y_t - y_p) + s \\ y_{t+1} &= y_p + (a_2 - bf_y)\,(y_t - y_p) + b(1-f_\pi)(\pi_t - \pi_{sp}) + d \end{aligned} \tag{9}$$

with steady-state equilibrium inflation[5]

$$\pi_{eq} = \pi_{sp} + \frac{1}{b(1-f_\pi)}\left(\frac{a_2 - 1 - bf_y}{a_1}s - d\right) \tag{10}$$

Although the instrument rule parameters f_π and f_y modify the value of the equilibrium inflation, the steady-state bias $\left|\frac{1}{b(1-f_\pi)}\left(\frac{a_2-1-bf_y}{a_1}s-d\right)\right|$ does not vanish unless either $\frac{a_2-bf_y}{a_1}s - d = 0$, or $f_\pi \to \infty$. The first situation is very unlikely while the second situation leads to unstable controlled dynamics since discrete-time systems display bounded stability margins[6]. The bias induced by the Taylor rule (1) is increased if, as it has been

[3]In the existing literature (see, for instance, [25-28]), it is assumed that the shocks s_t and d_t are serially uncorrelated disturbances with zero mean. In such case, although the shocks are non-vanishing, they are in the mean. In this form, the stabilization bias $\left|\frac{1}{b}\left(\frac{a_2-1}{a_1}s-d\right)\right|$ has also a zero mean, looking as a serially uncorrelated noise around the desired inflation π_d. Consequently, considering the shocks as non-zero constants is a worst case analysis where s and d are seen as persistent shocks (Onatski and Stock, 2000).

[4]It is not hard to see that the model (3) is controllable when the instrument rule i_t is seen as a control input. In this form, controllability implies that the instrument rule parameters f_π and f_y can be chosen to assign the roots of $P_c(z)$ at arbitrary positions within the unit open unit circle. In particular, dead-beat control ($i.e.$, finite-time control) conditions are obtained when all the eigenvalues of A_c are located at the origin of the complex plane. From (8), one sees that dead-beat control is obtained with the setting $f_y = (a_2+1)/b$ and $f_\pi = 1 + (a_2 - bf_y)/ba_1$. Dead-beat control is an interesting benchmark since output stabilization is achieved in a single time-period, and inflation stabilization is obtained in two time-periods.

[5]Notice that $y_{eq} = y_p - \frac{s}{a_1}$ regardless the used instrument rule.

[6]In fact, if $|f_\pi| \to \infty$, at least one eigenvalue of the controlled matrix M_c approaches infinity.

discussed by many authors (see, for instance, [32-35]), the potential output y_p is not known exactly and only an estimate of it, denoted by $\overline{y}_p$, is available. In this case, the implemented Taylor rule is

$$i_t = \pi_{sp} + f_\pi(\pi_t - \pi_{sp}) + f_y(y_t - \overline{y}_p) \tag{11}$$

which can be written as $i_t = \pi_{sp} + f_\pi(\pi_t - \pi_{sp}) + f_y(y_t - y_p) - f_y \Delta y_p$. When this expression is used in Eq. (9) and the corresponding steady-state equilibrium point is computed, one obtains that

$$\pi_{eq} = \pi_{sp} + \frac{1}{b(1-f_\pi)}\left(\frac{a_2 - 1 - bf_y}{a_1}s - d + bf_y \Delta y_p\right) \tag{12}$$

where $\Delta y_p = y_p - \overline{y}_p$ is the potential output uncertainty. Interestingly, this gap in the potential output has the effect of a demand shock in the form $-d + bf_y \Delta y_p$, so that the potential output gap Δy_p plays the role of a non-vanishing shock[7].

Despite the above discussed drawbacks of the Taylor rule, it has several advantages. In particular, its simplicity and the fact that its implementation requires little prior information for making it suitable to guide the monetary policies of central banks [38]. From a practical viewpoint it would be desirable to have an instrumental rule similar in structure to the Taylor rule endowed with flexible schemes to deal with non-vanishing shocks and uncertain potential output. More specifically, the problem addressed in this paper can be stated as follows:

"*Endow the Taylor rule (1) with suitable feedback control schemes to reject the steady-state bias induced by non-vanishing shocks and uncertain potential output, while maintaining the stability of the economy dynamics*"

This problem will be addressed under the following assumptions:

- **Assumption 1**. Exact inflation π_t and output y_t measurements are available, for all $t \geq 0$.
- **Assumption 2**. There is a set $\mathcal{F} \subset \mathbb{R}^2$ such that the matrix M_c (see Eq. (7)) is Schur-stable for all $(f_\pi, f_y) \in \mathcal{F}$.
- **Assumption 3**. Supply/demand shocks, s and d, can neither be anticipated nor measured.
- **Assumption 4**. The potential output y_p is not exactly known. Only an estimate of y_p, denoted by $\overline{y}_p$, is available to be used in the instrumental rule.

Remark 2.4. *Assumption 2 is required to guarantee the existence of a stabilizing feedback function of the form $i_t = \pi_{sp} + f_\pi(\pi_t - \pi_{sp}) + f_y(y_t - \overline{y}_p)$. On the other hand, as regarding Assumption 3, if shocks could be anticipated, one could use a standard feedforward scheme to counteract the adverse shock effects. The Taylor rule modifications to be introduced in the following section are aimed at maintaining stability while ensuring convergence $\pi_t \to \pi_{sp}$ in the presence of unanticipated shocks and potential output uncertainties.*

[7] From Eq. (12), one can see that an under-estimation $\Delta y_p > 0$ (resp., over-estimation $\Delta y_p < 0$) of the potential output leads to a negative (resp., positive) steady-state bias in the equilibrium inflation π_{eq}. This result seems to indicate that, in order to not overpassing inflation targets, central bank decision takers or policy makers should under-estimate the capacity of the underlying economy.

3. Main Results

First, note that the steady-state is $y_{eq} = y_p - \frac{s}{a_1}$ regardless the instrument rule. Hence, a non-vanishing supply shock s can be seen as inducing a potential output uncertainty of size $-s/a_1$. Thus, since this problem will be addressed below, in the sequel we will assume without loss of generality that $s = 0$.

3.1. Rejecting Demand Shocks

In a first subsection, the Taylor rule will be modified to deal with demand shocks. In this stage, it will be assumed that the potential output is known with certainty. In the next subsection, uncertain potential outputs will be considered and Taylor rule will be modified accordingly. Under these conditions, consider the following Taylor rule:

$$i_t = \pi_{sp} + f_\pi(\pi_t - \pi_{sp}) + f_y(y_t - y_p) + i_{d,t} \tag{13}$$

where the additional control input $i_{d,t}$ is intended to reject the effects of demand shocks in the steady-state inflation. Eq. (13) can be used in Eq. (3) to obtain the controlled dynamics

$$\begin{aligned} \pi_{t+1} &= \pi_t + a_1(y_t - y_p) \\ y_{t+1} &= y_p + (a_2 - bf_y)(y_t - y_p) + b(1 - f_\pi)(\pi_t - \pi_{sp}) - bi_{d,t} + d \end{aligned} \tag{14}$$

Notice that, if the shock d were anticipated, the simple *feedforward* choice $i_d = d/b$ could completely reject the shock effects. However, since the shock d cannot be anticipated (Assumption 3), the feedforward term $i_d = d/b$ can not be implemented in practice. To solve this problem, as is done in classical control, we resort to feedback strategies. The traditional way to reject the shock disturbance d is by means of an integral/summation feedback scheme; that is,

$$i_{d,t} = g_\pi \sum_{\tau=0}^{t} (\pi_\tau - \pi_{sp}) \tag{15}$$

where g_π is the inflation integral/summation gain. The integral/summation feedback (15) is equivalent to the following difference equation:

$$i_{d,t+1} = i_{d,t} + g_\pi (\pi_t - \pi_{sp}) \tag{16}$$

The *rationale* behind the usage of this integral/summation scheme is the following: If g_π is chosen in a way that the controlled economy composed by eqs. (14) and (16) is stable and convergent in the sense that it converges to a steady-state equilibrium point, say $(\pi_{eq}, y_{eq}, i_{d,eq})$, then Eq. (16) implies that $\pi_{eq} = \pi_{sp}$. That is, the role of the integral/summation feedback is to reject the steady-state inflation bias. This can be seen if one observes that $i_t \to i_{d,eq}$ and $\pi_t \to \pi_{eq} = \pi_{sp}$ imply $i_{d,eq} = d/b$. Hence, the integral/summation *feedback* scheme is able to recover asymptotically the shock rejection effect induced by the *feedforward* scheme $i_d = d/b$.

The stability of the controlled economy is established in the following theorem.

Theorem 1. *Consider the Taylor rule (13) with $f_\pi \neq 1$ and exact knowledge of the potential output y_p. Under assumptions 1 to 3 listed in Section 2, there exists a positive constant*

$g_\pi^{\max}$ *such that, for all* g_π *contained in either* $(-g_\pi^{\max}, 0)$ *if* $1 - f_\pi > 0$ *or in* $(0, +g_\pi^{\max})$ *if* $1 - f_\pi < 0$ *, the controlled economy composed by Eq. (14) and the integral/summation feedback control (16) is stable and the inflation dynamics converges asymptotically to the target level (*$i.e.$*,* $\pi_t \to \pi_{sp}$*).*

Proof. Let $x_{t,1} = \pi_t - \pi_{sp}$ and $x_{t,2} = y_t - y_p$. Then, if $x_t \overset{def}{=} (x_{1,t}, x_{2,t})^T$, the controlled dynamics is given by

$$\begin{aligned} x_{t+1} &= M_c x_t + B i_{d,t} + D \\ i_{d,t+1} &= i_{d,t} + g_\pi C x_t \end{aligned} \tag{17}$$

where the matrix M_c is given by Eq. (7), $B = (0, -b)^T$, $D = (0, d)^T$ and $C = (1, 0)$. Eq. (17) corresponds to a control system of the form (A.1)-(A.2) as considered in the Appendix, with system matrix $A = M_c$, control input $u = i_d$, regulated output $r = x_1 = \pi - \pi_{sp}$, and integral/summation control gain $g = g_\pi$. Consequently, the proof of the Theorem is established by showing that the assumptions of Proposition A.1 (in the appendix) are met. i) Assumption 2 ensures the existence of stabilizing instrument rule parameters f_π and f_y such that the matrix M_c (see Eq. (7)) is Schur-stable. ii) It is straightforward to show that $|I - M_c| = -a_1 b(1 - f_\pi)$. Since $a_1 > 0$, $b > 0$ and $f_\pi \neq 1$ by assumption, one has that $I - M_c$ is a non-singular matrix. iii) Finally, $C(I - M_c)^{-1}B = 1/(1 - f_\pi) \neq 0$ by assumption. Therefore, Proposition A.1 guarantees the existence of a positive constant $g_\pi^{\max}$ as described in the statements of the theorem. □

The following comments are in order:

a) The sign of the integral/summation feedback parameter g_π depends on the sign of $1/(1 - f_\pi)$. In this way, the selected value of the Taylor rule parameter f_π determines the sign of the integral/summation feedback parameter g_π.

b) The integral/summation feedback (15) (or (16)) acts as follows (see Remark A.1 in the Appendix): since g_π is small, $i_{d,t} \simeq 0$ for t small, then the controlled economy behaves approximately as $x_{t+1} = M_c x_t + D$. Since M_c is Schur-stable, x_t reaches a neighborhood of the point $(I - M_c)^{-1} D$ (corresponding to the equilibrium point with $i_{d,t} = 0$) where the demand shock d has an inflation deviation effect ($i.e.$, π_t approaches $\pi_{sp} - d/\left[b(1 - f_\pi)\right]$). As time increases, the integral/summation part $i_{d,t}$ has a more important effect by sliding the trajectory $(x_t, i_{d,t})^T$ along a neighborhood of the equilibrium locus $E = \{(x, i_d) : b(1 - f_\pi)(\pi - \pi_{sp}) - b i_d + d = 0\}$[8] to converge to an equilibrium point in E where $i_{d,eq} = d/b$ and $\pi_{eq} = \pi_{sp}$. In this way, the integral/summation feedback (15) (or 16) can be interpreted as a *learning scheme* [39] which learns the demand shocks and counteracts their adverse effects on the inflation control in a similar form to feedforward schemes.[9]

c) Some researchers have attempted to explain the Federal Reserve monetary policy from the perspective of instrument rules similar to the Taylor rule [26,37,40]. The idea has

[8] Notice that the equilibrium locus E does not depend on y_{eq}, so that E is a horizontal line in the plane i versus y.

[9] That is, the learning scheme given by Eq. (15) eliminates the bias in the inflation feedback control about the target value π_{sp}.

been to fit interest rate movements with a Taylor-type reaction function. A reported feature is that Federal Reserve takes the past values of inflation gap into account, although its response is smaller compared to the response of current inflation [37]. In terms of the modified Taylor rule given by eqs. (13) and (15), such empirical findings mean that $|g_\pi|$ is smaller than $|f_\pi|$. This seems to be in agreement with the result in Theorem 1, which states that the integral/summation gain $|g_\pi|$ should be relatively small.

3.2. Learning Potential Output Deviations

So far, it has been assumed that the potential output y_p is known with certainty. However, in practice only an estimate $\overline{y}_p$ of y_p is known, so that the modified Taylor rule given by eqs. (13) and (15) has to be modified as:

$$\begin{aligned} i_t &= \pi_{sp} + f_\pi(\pi_t - \pi_{sp}) + f_y(y_t - \overline{y}_p) + i_{d,t} \\ i_{d,t} &= i_{d,t-1} + g_\pi\,(\pi_{t-1} - \pi_{sp}) \end{aligned} \tag{18}$$

As discussed before, the deviation Δy_p produces the same steady-state effect of a demand shock in the form $d - bf_y\Delta y_p$. Given the result in Theorem 1, the incorporation of the integral/summation control action into the modified Taylor rule (18) suffices for removing such a demand shock-type effect. From a dynamical viewpoint, one can expect that a small potential output uncertainty would not affect the performance of the Taylor rule (13),(15) where the exact value of the potential output is required. However, large potential output uncertainties can lead to a degraded performance because, as mentioned above, the uncertainty $\Delta y_p = y_p - \overline{y}_p$ induces a demand shock-type effect. Recently, some authors have studied this problem by redesigning monetary policy rules for reducing the effects of uncertainties in potential output [33,35]. In the following we will provide a simple *ad hoc* solution to the problem of errors in measurements of the potential output. Specifically, the Taylor rule given by Eq. (18) will be endowed with a learning scheme to estimate adaptively the uncertainty $\Delta y_p = y_p - \overline{y}_p$. The addition of such a learning scheme into the Taylor rule offers the following advantages:

i) By including information about the potential output uncertainty, the control effort devoted by the central bank's instrument i_t can be reduced and its performance can be enhanced as compared to the Taylor rule (18) where only the estimate $\overline{y}_p$ is considered, and

ii) The exact level of the potential output y_p can be obtained at least asymptotically. This is of particular relevance since the knowledge of the potential output y_p is one of the main indices characterizing the economic behavior.

Let $\overline{\Delta y}_{p,t}$ be an estimate of the potential output uncertainty $\Delta y_{p,t}$. Consider the following modified Taylor rule:

$$\begin{aligned} i_t &= \pi_{sp} + f_\pi(\pi_t - \pi_{sp}) + f_y(y_t - \overline{y}_p - \overline{\Delta y}_{p,t}) + i_{d,t} \\ i_{d,t} &= i_{d,t-1} + g_\pi\,(\pi_{t-1} - \pi_{sp}) \end{aligned} \tag{19}$$

where $\overline{\Delta y}_{p,t}$ is estimated according to the following learning scheme:

$$\overline{\Delta y}_{p,t+1} = \overline{\Delta y}_{p,t} + g_y \left(y_t - \overline{y}_p - \overline{\Delta y}_{p,t} \right) \tag{20}$$

where g_y is a learning parameter to be specified below. The idea behind the adaptation scheme (20) is the following: when the controlled dynamics is stable and converge to an equilibrium point, Eq. (20) implies that $y_{eq} = \overline{y}_p + \overline{\Delta y}_{p,eq} = y_p$, so that $\overline{\Delta y}_{p,eq} = \Delta y_p$. That is, the adaptation scheme (20) would recover asymptotically the uncertainty in the potential output. In this form, the performance of the Taylor rule (19) would converge to the performance of the Taylor rule given by eqs. (13) and (15) where the exact potential output value is known.

The stability of the controlled economy under the instrument rule (19)-(20) is established in the following theorem.

Theorem 2. *Consider the economy (14) under assumptions 1-4 as described in Section 2, and under the action of the Taylor rule (19)-(20) with $f_\pi \neq 1$ and g_π specified as in Theorem 1. Then, there exists a positive number $g_y^{\max}$ such that, for all $g_y \in \left(0, g_y^{\max}\right)$, one has that i) the controlled economy is stable, ii) $\pi_t \to \pi_{sp}$ ($i.e.$, the inflation task is achieved), and iii) $\overline{\Delta y}_{p,t} \to \Delta y_p$ (the exact potential output is recovered).*

Proof. To avoid excessive algebraic manipulations, it can be assumed that $i_{d,t} = 0$. If $e_{p,t} \stackrel{def}{=} \overline{\Delta y}_{p,t} - \Delta y_p$ is the estimation error, the instrument rule can be written as

$$i_t = \pi_{sp} + f_\pi(\pi_t - \pi_{sp}) + f_y(y_t - y_p) + f_y e_{p,t} \tag{21}$$

Introduce the following coordinates: $x_{1,t} = \pi_t - \pi_{sp}$ and $x_{2,t} = y_t - y_p$. Then, the controlled dynamics composed by the system (3) and the Taylor's rule (21) can be described as follows:

$$\begin{aligned} x_{t+1} &= M_c x_t + G_1 e_{p,t} \\ e_{p,t+1} &= -\varepsilon_y G_2 x_t + (1 - \varepsilon_y)\, e_{p,t} \end{aligned} \tag{22}$$

where $G_1 = (0, bf_y)^T$ and $G_2 = (0, 1)$. The matrix M_c is Schur-stable by assumption. Additionally, the matrix $I - M_c$ is non-singular. In fact, $|I - M_c| = -a_1 b\,(1 - f_\pi) \neq 0$ since $f_\pi \neq 1$ by assumption. It can be shown that these properties imply the existence of a positive number $\varepsilon_{y,1}^{\max}$ such that, for all $\varepsilon_y \in \left(0, \varepsilon_{y,1}^{\max}\right)$, the origin is the unique equilibrium point of the system (22). To establish the stability of the system (22), we will consider ε_y as a small parameter, so that the system (22) becomes a singularly perturbed system. Consider the coordinates change $w = x + L(\varepsilon_y) e_p$, where the vector $L(\varepsilon_y) \in \mathbb{R}^2$ will be specified below. Then, one has that

$$w_{t+1} = (M_c - \varepsilon_y L(\varepsilon_y) G_2) w_t + \Phi(L(\varepsilon_y), \varepsilon_y) e_{p,t}$$

where

$$\Phi(L(\varepsilon_y), \varepsilon_y) = G_1 + L(\varepsilon_y) - M_c L(\varepsilon_y) + \varepsilon_y L(\varepsilon_y)(G_2 L(\varepsilon_y) - 1)$$

Let us show that for sufficiently small $\varepsilon_{y,2}^{\max} > 0$, there exists a vector $L(\varepsilon_y) \in \mathbb{R}^2$ such that $\Phi(L(\varepsilon_y), \varepsilon_y) = 0$, for all $\varepsilon_y \in \left(0, \varepsilon_{y,2}^{\max}\right)$. To see this, notice that $\Phi(L(0), 0) = 0$ and the fact that $I - M_c$ is non-singular imply that $L_0 = L(0) = -(I - M_c)^{-1} G_1$. Given that the term $\varepsilon_y L(\varepsilon_y)(G_2 L(\varepsilon_y) - 1)$ is a continuous function of ε_y and is $\mathcal{O}(\varepsilon_y)$, one can ensure the existence of a sufficiently small number $\varepsilon_{y,2}^{\max} > 0$ such that a vector of the form $L(\varepsilon_y) = L_0 + \mathcal{O}(\varepsilon_y) \in \mathbb{R}^2$ is a solution of $\Phi(L(\varepsilon_y), \varepsilon_y) = 0$, for all $\varepsilon_y \in \left(0, \varepsilon_{y,2}^{\max}\right)$, for example, by invoking the continuous implicit function theorem [42]. Hence, in (w, e_p)-coordinates the controlled dynamics (22) becomes

$$\begin{aligned} w_{t+1} &= (M_c - \varepsilon_y L(\varepsilon_y) G_2) w_t \\ e_{p,t+1} &= -\varepsilon_y G_2 w_t + (1 - \varepsilon_y + \varepsilon_y G_2 L(\varepsilon_y))\, e_{p,t} \end{aligned} \tag{23}$$

which is a cascade ($i.e.$, triangular) system. Recall that the matrix M_c is Schur-stable by assumption. Since $\varepsilon_y L(\varepsilon_y) G_2$ is an $\mathcal{O}(\varepsilon_y)$ term, there exists a positive number $\varepsilon_{y,3}^{\max}$ such that the matrix $M_c - \varepsilon_y L(\varepsilon_y) G_2$ is Schur-stable for all $\varepsilon_y \in \left(0, \varepsilon_{y,3}^{\max}\right)$. This implies that $w_t \to (0,0)^T$ for all $\varepsilon_y \in \left(0, \varepsilon_{y,3}^{\max}\right)$. On the other hand, let us show that $\varepsilon_y L(\varepsilon_y) G_2$ is an $\mathcal{O}(\varepsilon_y^2)$ term. In fact, from Eq. (7), one has that $L_0 = -(I - M_c)^{-1} G_1 = \left(\frac{1}{f_\pi - 1} f_y, 0\right)^T$, so that $G_2 L_0 = 0$. Since $L(\varepsilon_y) = L_0 + \mathcal{O}(\varepsilon_y)$, one has that $\varepsilon_y L(\varepsilon_y) G_2$ is of the order of $\mathcal{O}(\varepsilon_y^2)$. Consequently, there is a positive number $\varepsilon_{y,4}^{\max} < 1$ such that $1 - \varepsilon_y + \varepsilon_y G_2 L(\varepsilon_y) < 1$ for all $\varepsilon_y \in \left(0, \varepsilon_{y,4}^{\max}\right)$. Given the cascade structure of the system (23), the above conditions imply that $w_t \to (0,0)^T$ and $e_{p,t} \to 0$, for all $\varepsilon_y \in \left(0, \varepsilon_y^{\max}\right)$, where $\varepsilon_y^{\max} \overset{def}{=} \min\left\{\varepsilon_{y,1}^{\max}, \varepsilon_{y,2}^{\max}, \varepsilon_{y,3}^{\max}, \varepsilon_{y,4}^{\max}\right\}$. Since $L(\varepsilon_y) \neq (0,0)$, one arrives to the conclusion that $x_t \to (0,0)^T$ and $e_{p,t} \to 0$, for all $\varepsilon_y \in \left(0, \varepsilon_y^{\max}\right)$. This implies that i) the closed-loop dynamics is stable, ii) the inflation target is achieved in the sense that $\pi_t \to \pi_{sp}$, and iii) the potential output y_p is recovered asymptotically since $\overline{\Delta y}_{p,t} \to \Delta y_p$. □

Remark 3.1. *It is interesting to note that the adaptation scheme (20) has the structure of an integral/summation feedback acting on the estimated potential gap. In fact,*

$$\overline{\Delta y}_{p,t} = g_y \sum_{\tau=0}^{t} \left(y_\tau - \overline{y}_p - \overline{\Delta y}_{p,\tau}\right)$$

In this way, the signal $\overline{\Delta y}_{p,t}$ is estimated recursively by trying to close the gap $y_t - \overline{y}_p - \overline{\Delta y}_{p,t}$.

Summarizing, the modified Taylor rule endowed with schemes to face demand shocks and uncertainty about potential output (also interpreted as supply shocks) is given by:

$$\begin{aligned} i_t &= \pi_{sp} + f_\pi(\pi_t - \pi_{sp}) + f_y(y_t - \overline{y}_p - \overline{\Delta y}_{p,t}) + i_{d,t} \\ i_{d,t} &= i_{d,t-1} + g_\pi \left(\pi_{t-1} - \pi_{sp}\right) \\ \overline{\Delta y}_{p,t+1} &= \overline{\Delta y}_{p,t} + g_y \left(y_t - \overline{y}_p - \overline{\Delta y}_{p,t}\right) \end{aligned}$$

or as an integral/summation equation as follows:

$$i_t = \pi_{sp} + f_\pi(\pi_t - \pi_{sp}) + f_y \left[y_t - \overline{y}_p - g_y \sum_{\tau=0}^{t} \left(y_\tau - \overline{y}_p - \overline{\Delta y}_{p,\tau}\right) \right] + g_\pi \sum_{\tau=0}^{t} \left(\pi_\tau - \pi_{sp}\right) \tag{24}$$

Now, the Taylor rule has four adjustable parameters; namely, f_π, f_y, g_π, and g_y. The parameters f_π and f_y adjust the response of the central bank's instrument to *current* inflation and output deviations, respectively. On the other hand, the parameters g_π and g_y adjust the response of the instrument to *past* inflation and output deviations, respectively. In this way, the adaptation schemes for $i_{d,t}$ and Δy_p introduce a type of memory effects, by virtue of which biases in the control process are reduced and eventually eliminated from the steady-state conditions. It is interesting to note that such a memory effect is induced by a history-dependence structure similar to instrument rules derived from forward-looking models [43]. In fact, Eq. (24) is similar to that proposed by Svensson [1] to remedy the lack of history-dependence of simple Taylor rules. The main difference is that, while Svensson's proposal is of heuristic nature, Eq. (24) arises from a systematic effort to deal with potential output uncertainty and shocks robustly. The degrees of freedom allowed by the possibility of tuning the four parameters open the way to design complex, rich and flexible policies.

Remark 3.2. *It is interesting to notice that the modified Taylor rule (24) can be written as a standard Taylor rule plus a "driving signal" (denoted by ξ_t) as follows:*

$$i_t = \pi_{sp} + f_\pi(\pi_t - \pi_{sp}) + f_y(y_t - \overline{y}_p) + \xi_t \tag{25.a}$$

where

$$\xi_t = -f_y g_y \sum_{\tau=0}^{t} \left(y_\tau - \overline{y}_p - \overline{\Delta y}_{p,\tau}\right) + g_\pi \sum_{\tau=0}^{t} (\pi_\tau - \pi_{sp}) \tag{25.b}$$

In this form, the modified Taylor rule (25) is similar in structure to policy reaction functions obtained by posing a stochastic optimal linear regulator problem [41] where ξ_t can thought as a policy shock introduced to pick movements in the interest rate that cannot be explained by the model. In our case, the signal ξ_t is used to counteract the effects of modeling errors due to uncertain shocks and potential output. Whether or not the adaptive schemes lumped in the signal ξ_t can explain significant departures from the standard Taylor rule and Federal Reserve interest rate (see, for instance, [40]) is a problem that deserves a detailed study, which should be reported in future work.

Remark 3.3. *Taylor rules with smoothing schemes have been also considered. In fact, smoothing schemes are introduced to avoid excessive interest rate adjustments that can provoke instabilities in the financial sector [43-46]. A typical Taylor rule with first-order ($i.e.$, one lag) smoothing can be expressed as follows:*

$$i_t = \alpha_i i_{t-1} + (1-\alpha_i)\left[\pi_{sp} + f_\pi(\pi_t - \pi_{sp}) + f_y(y_t - y_p)\right] \tag{26}$$

where $\alpha_i \in [0,1]$ is the smoothing parameter. If $\alpha_i = 0$, no smoothing is present and the original Taylor rule given by Eq. (1) is recovered. In the other extreme, if $\alpha_i = 1$ one has perfect smoothing characterized by constant interest rate ($i.e.$, $i_t = \bar{i}$, for all $t \geq 0$). In the interval where $\alpha_i \in (0,1)$, the instrument value i_t is given by the convex combination of previous instrument value i_{t-1} and the suggested correction $\pi_{sp} + f_\pi(\pi_t - \pi_{sp}) + f_y(y_t - y_p)$. In principle, such a convex combination gives smoother interest rate movements with *reduced interest rate volatility*[10]*. The modifications made to the Taylor rule (1) and the corresponding stability analysis can be extended to the case of smoothed Taylor rule. This can be done by noting that, by endowing it with the control input* $i_{d,t}$ *to reject demand shocks (see Eq. (13)), Eq. (26) with estimated potential output* $\overline{y}_p$*, can be written as*

$$i_t = \alpha_i i_{t-1} + \pi_{sp} + f'_\pi(\pi_t - \pi_{sp}) + f'_y(y_t - \overline{y}_p) + i_{d,t} \tag{27}$$

where $\pi'_{sp} = (1-\alpha_i)\pi_{sp}$, $f'_\pi = (1-\alpha_i)f_\pi$ *and* $f'_y = (1-\alpha_i)f_y$*. The resulting closed-loop economy is given by*

$$\begin{aligned} \pi_{t+1} &= \pi_t + a_1(y_t - \overline{y}_p) \\ y_{t+1} &= \overline{y}_p + \left(a_2 - bf'_y\right)(y_t - \overline{y}_p) + b(1-f'_\pi)(\pi_t - \pi_{sp}) - b\alpha_i i_{t-1} - bi_{d,t} + d' \end{aligned} \tag{28}$$

where $d' = d + \alpha_i b\pi_{sp}$*. By introducing the dummy variable* $z_t = i_{t-1}$*, the system (28) can be described as follows:*

$$\begin{aligned} \pi_{t+1} &= \pi_t + a_1(y_t - \overline{y}_p) \\ y_{t+1} &= \overline{y}_p + \left(a_2 - bf'_y\right)(y_t - \overline{y}_p) + b(1-f'_\pi)(\pi_t - \pi_{sp}) - b\alpha_i z_t - bi_{d,t} + d' \\ z_{t+1} &= \alpha_i z_t + \pi'_{sp} + f'_\pi(\pi_t - \pi_{sp}) + f'_y(y_t - y_p) + i_{d,t} \end{aligned} \tag{29}$$

where we have made use of Eq. (27) to express the relationship $z_{t+1} = i_t$*. The effect of the smoothing scheme (27) is to extend the dimensionality of the dynamical system. From Eq. (29), the modifications made to the Taylor rule (1) to reject demand shocks and to learn about uncertain potential output can be made following the same design steps as those used for the non-smoothing case. The main difference relies in the stability analysis where the parameters* f'_π, f'_y *and* α_i *must be chosen in a way that the three-dimensional matrix*

$$M_c = \begin{pmatrix} 1 & a_1 & 0 \\ b(1-f'_\pi) & a_2 - bf'_y & -b\alpha_i \\ f'_\pi & f'_y & \alpha_i \end{pmatrix} \tag{30}$$

be Schur-stable. Notice the important role has the smoothing parameter α_i *in the stability of the closed-loop dynamics. Assumed the existence of a stabilizing instrument rule parameters* $\left(f'_\pi, f'_y, \alpha_i\right)^T \in \mathcal{F}' \subset \mathbb{R}^3$ *(see Assumption 2), theorems 1 and 2 described above can be extended accordingly. The final result is the modified Taylor rule:*

$$i_t = \alpha_i i_{t-1} + (1-\alpha_i) i_t^{ns} \tag{31}$$

where i_t^{ns} *denotes the "non-smoothed" instrument rule given by Eq. (24). In this form, one obtains a modified Taylor rule with smoothing and history-dependent schemes.*

As has been widely discussed, an important feature of the Taylor rule is its simplicity, which makes it attractive to guide monetary policies in practical situations because operators can understand it from fundamental economic concepts. Such simplicity of the Taylor

[10]Clarida *et al.* [45] have suggested that model and parameter uncertainty may induce policymakers to have smoothing incentive. Another reason for a smoothing incentive is to ensure the existence of well-functioning capital markets. In fact, volatile interest rates may result in capital losses, which would be disruptive for the financial sector.

rule relies on the fact that its implementation requires very little prior information; namely, inflation and output measurements. It is a surprising fact that such a simple instrument rule commonly displays better robustness characteristics (across model parameters and structure) than more complex instrument rules derived from optimizing a loss function. We feel that the modified Taylor rule (24) retains to a large extent the simplicity of the original Taylor rule (1) because its adaptation schemes require the same information (inflation and output measurements) and their construction are motivated from basic estimation concepts.

4. Conclusions

In this paper, we have shown that existing methodologies in feedback control theory can help to design and improve instrumental rules for designing monetary policies. In particular, the incorporation of adaptation schemes to deal with uncertain parameters, unanticipated shocks and uncertain potential output leads to improved Taylor-type rules that are still simple in structure, such that they can be adopted by monetary policy practitioners to guide certain central bank's decisions in a more comprehensive way. In this research line, from the feedback control theory viewpoint, some interesting problems deserve to be examined in a future work. Among others, the following questions are worth mentioning: a) The design of tuning guidelines to select the instrument rule parameters. Tuning guidelines based on the optimization of a given performance criteria ($e.g.$, settling time, etc.) can be explored. b) Selection of instrument rule parameters to achieve robust stability in the presence of parametric uncertainties. This problem has a practical importance since the estimated parameters economy models are uncertain. In principle, the robust stabilization problem involves studying the stability of a family of models under a fixed structure controller. c) Smoothing of instrument variations to avoid excessive changes that can induce instabilizing effects in, $e.g.$, financial sector, and d) Extending the economy model to explicitly include financial and private ($i.e.$, productive) sectors. Results along these lines would help monetary policy practitioners to conduct in a more systematic way the stabilization of actual economies faced sometimes to excessive unstabilizing disturbances.

A. Appendix

The proof of Theorem 1 is based in the following proposition, which shows the stability of a control system under integral/summation feedback action. To the best of our knowledge, this result has not been reported previously in the feedback control literature.

Proposition A.1. *Consider the following single-input/single-output control system:*

$$\begin{aligned} x_{t+1} &= Ax_t + Bu_t + D \\ r_t &= Cx_t \end{aligned} \tag{A.1}$$

where $x \in \mathbb{R}^n$ is the system state, $u \in \mathbb{R}$ is the control input, $r \in \mathbb{R}$ is the regulated output, $D \in \mathbb{R}^n$ is a vector of (constant) disturbances, and the matrices A, B and C have suitable dimensions ($i.e.$, $A \in \mathbb{R}^{n\times n}$, $B \in \mathbb{R}^{n\times 1}$ and *$C \in \mathbb{R}^{1\times n}$). Take the following assumptions:*

- *Assumption A.1. The matrix A is Schur-stable.*

- *Assumption A.2. The matrix $I - A$ is non-singular.*
- *Assumption A.3. $C(I - A)^{-1}B \neq 0$.*

Let r_{sp} be a desired (setpoint) value and endow the system (A.1) with the integral/summation feedback control:

$$u_{t+1} = u_t + g(r_t - r_{sp}) \tag{A.2}$$

where g is the integral/summation control gain. Additionally, let

$$sign(g) = -sign(C(I - A)^{-1}B) \tag{A.3}$$

Then, there exists a positive number $g^{\max}$ such that, for all $g \in (0, g^{\max})$ if $sign(C(I - A)^{-1}B) = -1$ or for all $g \in (-g^{\max}, 0)$ if $sign(C(I - A)^{-1}B) = +1$, *the controlled system composed by eqs. (A.1) and (A.2) is stable and $r_t \to r_{sp}$ asymptotically.*

Proof. For $g \neq 0$, Eq. (A.2) implies that, at equilibrium conditions, $r_{eq} = r_{sp}$. Hence, it suffices to show the stability of the controlled system to guarantee the convergence $r_t \to r_{sp}$. For simplicity in notation, let $H_B \overset{def}{=} C(I - A)^{-1}B$ and $H_D \overset{def}{=} C(I - A)^{-1}D$. By virtue of Assumption A.2, the controlled system (A.1)-(A.3) has a unique equilibrium point $(x_{eq}, u_{eq})^T \in \mathbb{R}^{n+1}$. In fact, since $H_B \neq 0$ (Assumption A.3), one has that

$$\begin{aligned} x_{eq} &= F\left[(r_{sp} - H_D)/H_B\right] + (I - A)^{-1}D \\ u_{eq} &= (r_{sp} - H_D)/H_B \end{aligned} \tag{A.4}$$

where $F \overset{def}{=} (I - A)^{-1}B \in \mathbb{R}^n$. Let us shift the origin to the equilibrium. To this end, introduce the new coordinates $w \in \mathbb{R}^n$ and $z \in \mathbb{R}$:

$$\begin{aligned} w &= x - Fu_{eq} - (I - A)^{-1}D \\ z &= u - u_{eq} \end{aligned} \tag{A.5}$$

Using w and z as state variables, and after some tedious but straightforward algebraic manipulations, system (A.1)-(A.2) becomes

$$\begin{aligned} w_{t+1} &= (A + gFC)w_t - gFCFz_t \\ z_{t+1} &= gCw_t + (1 + gCF)z_t \end{aligned} \tag{A.6}$$

Then, eqs. (A.4) and (A.5) imply that the origin is the unique equilibrium point of the system (A.6). For convenience, consider the parametrization $g = \varepsilon\widetilde{g}$, where $\varepsilon > 0$ and $\widetilde{g}$ is a constant with $sign(\widetilde{g}) = sign(g)$. Consider the further coordinate change

$$q = \varepsilon^{-1}w + L(\varepsilon)z \tag{A.7}$$

where $L(\varepsilon) \in \mathbb{R}^{n\times 1}$ is a matrix to be determined later. In this form, Eq. (A.6) is transformed into the following system

$$\begin{aligned} q_{t+1} &= (A + \varepsilon\widetilde{g}FC + \varepsilon^2\widetilde{g}L(\varepsilon)C)q_t + \Psi\left(L(\varepsilon), \varepsilon\right)z_t \\ z_{t+1} &= \varepsilon^2\widetilde{g}Cq_t + (1 + \varepsilon\widetilde{g}CF - \varepsilon^2\widetilde{g}CL(\varepsilon))z_t \end{aligned} \tag{A.8}$$

where

$$\Psi\left(L(\varepsilon),\varepsilon\right) = L(\varepsilon) - AL(\varepsilon) - \widetilde{g}FCF + \varepsilon\widetilde{g}\left(L(\varepsilon)CF - FCL(\varepsilon) - \varepsilon L(\varepsilon)CL(\varepsilon)\right) \quad \text{(A.9)}$$

For sufficiently small $\varepsilon > 0$, there exists a matrix $L(\varepsilon)$ such that $\Psi\left(L(\varepsilon),\varepsilon\right) = 0$. To see this, notice that $\Psi\left(L(0),0\right) = 0$ and the fact that $I - A$ is invertible (Assumption A.2) implies that

$$L_0 \stackrel{def}{=} L(0) = \widetilde{g}(I-A)^{-1}FCF \quad \text{(A.10)}$$

Since the last term in Eq. (A.9) is $\mathcal{O}(\varepsilon)$ (a function $f(\varepsilon)$ is $\mathcal{O}(\varepsilon)$ iff locally $|f(\varepsilon)| \leq M\varepsilon$), continuity of the function $\Psi\left(L(\varepsilon),\varepsilon\right)$ with respect to its arguments implies the existence of a positive constant $\varepsilon_1^{\max}$ such that the matrix $L(\varepsilon) = L_0 + \mathcal{O}(\varepsilon) \in \mathbb{R}^{n\times 1}$ is a solution of $\Psi\left(L(\varepsilon),\varepsilon\right) = 0$, for all $\varepsilon \in (0,\varepsilon_1^{\max})$. As a consequence, system (A.8) is reduced into the following form:

$$\begin{aligned} q_{t+1} &= (A + \varepsilon\widetilde{g}FC + \varepsilon^2\widetilde{g}L(\varepsilon)C)q_t \\ z_{t+1} &= \varepsilon^2\widetilde{g}Cq_t + (1 + \varepsilon\widetilde{g}CF - \varepsilon^2\widetilde{g}CL(\varepsilon))z_t \end{aligned} \quad \text{(A.11)}$$

which is a discrete-time system in cascade form with $(q_{eq}, z_{eq})^T = (0,0)$ as its unique equilibrium point. Since the matrix A is Schur-stable (Assumption A.1), there exists a positive constant $\varepsilon_2^{\max}$ such the perturbed matrix $A+\varepsilon\widetilde{g}FC+\varepsilon^2\widetilde{g}L(\varepsilon)C$ is Schur-stable, for all $\varepsilon \in (0,\varepsilon_2^{\max})$. On the other hand, since $sign(\widetilde{g}FC) = -1$ (see Eq. (A.3)) and since the term $\varepsilon^2\widetilde{g}CL(\varepsilon)$ is $\mathcal{O}(\varepsilon^2)$, there exists a positive constant $\varepsilon_3^{\max}$ such that $|1 + \varepsilon\widetilde{g}CF| < 1$, for all $\varepsilon \in (0,\varepsilon_3^{\max})$. In this way, the cascade system (A.11) is asymptotically stable about the origin for all $\varepsilon \in (0,\varepsilon_4^{\max})$, where $\varepsilon_4^{\max} \stackrel{def}{=} \min\{\varepsilon_2^{\max},\varepsilon_3^{\max}\}$. Given that $q = \varepsilon^{-1}w + L(\varepsilon)z$ is a non-singular coordinates change and $\varepsilon > 0$, $q_t \to 0$ and $z_t \to 0$ imply that $w_t \to 0$ asymptotically, for all $\varepsilon \in (0,\varepsilon^{\max})$, where $\varepsilon^{\max} \stackrel{def}{=} \min\{\varepsilon_1^{\max},\varepsilon_4^{\max}\}$. Finally, from (A.5), $z_t \to 0$ yields $u_t \to u_{eq}$. In this way, $w_t = x_t - Fu_{eq} - (I-A)^{-1}D \to 0$ leads to the following expressions:

$$\begin{aligned} x_t &\to Fu_{eq} + (I-A)^{-1}D \\ &= (I-A)^{-1}(Bu_{eq} - D) \\ &= x_{eq} \text{ (by virtue of Eq. (A.4))} \end{aligned}$$

This concludes the proof. □

Remark 1.1. *By considering the parametrization $g = \varepsilon\widetilde{g}$ taken in the proof of Proposition A.1, it is interesting to note that the system (A.6); namely,*

$$\begin{aligned} w_{t+1} &= (A + \varepsilon\widetilde{g}FC)w_t - \varepsilon\widetilde{g}FCFz_t \\ z_{t+1} &= \varepsilon\widetilde{g}Cw_t + (1 + \varepsilon\widetilde{g}CF)z_t \end{aligned} \quad \text{(A.12)}$$

can be seen as a singularly perturbed system with $\varepsilon > 0$ being the perturbation parameter. The fast subsystem is $w_{t+1} = Aw_t$ and the slow subsystem is $z_{t+1} = (1+\varepsilon\widetilde{g}CF)z_t$. In fact, since ε is small, one has that the singular value $1 + \varepsilon\widetilde{g}CF$ of the slow subsystem is lower but close to the unity, so that z_t converges slowly to zero. The slow and the fast subsystems provide a picture of how dynamics of the full singularly perturbed system (A.12) behaves; namely,

- *For short time-scales ($i.e.$, t small) the dynamics of (A.12) is dominated by the dynamics of the fast subsystem. Since ε is a small parameter, one has that $u_t \simeq 0$, so that the controlled system behaves as $x_{t+1} = Ax_t + D$. Given that A is Schur-stable (Assumption A.3), x_t converges into a neighborhood of the point $x_D \stackrel{def}{=} (I-A)^{-1}D$.*
- *For large time-scales ($i.e.$, t large) the dynamics are dominated by the slow subsystem. Starting from a neighborhood of the point $(x_D^T, 0)$ achieved during the fast transient, the trajectory x_t moves slowly into a neighborhood of the equilibrium manifold $E \stackrel{def}{=} \left\{(x,u)^T : (I-A)x - Bu - D = 0\right\}$ to converge to an equilibrium point $(x_{eq}, u_{eq})^T$ contained in E ($i.e.$, $(I-A)x_{eq} - Bu_{eq} - D = 0$) corresponding to $r_{eq} = Cx_{eq} = r_{sp}$.*

This interpretation of the effect of the integral/summation feedback could be useful for providing a description of the controlled inflation dynamics.

References

[1] L.E.O. Svensson, What is wrong with Taylor rules? Using judgment in monetary policy through targeting rules, *Journal of Economic Literature*, **41** (2003) 426-477.

[2] G.F. Franklin, J.D. Powell, M.L. Workman, *Digital Control of Dynamic Systems*, 2nd. Ed, Addison-Wesley Publishing Company, 1990.

[3] K. Ogata, *Modern Control Engineering*, 2nd Edition, Prentice-Hall, Englewood Cliffs, New Jersey, 1990.

[4] G.F. Franklin, J.D. Powell, A. Emami-Naeini, *Feedback Control of Dynamic Systems*, 2nd Edition, Addison-Wesley, Reading, Massachusetts, 1991.

[5] A. Blinder, *Central Banking in Theory and Practice*, MIT Press, Cambridge Massachusetts, 1998.

[6] J.B. Taylor, *Monetary Policy Rules*, Chicago University Press, Chicago, 1999.

[7] M. Goodfriend, *Inflation targeting in the United States*, NBER Working Paper, 2003.

[8] B.T. McCallum, *The use of policy rules in monetary policy analysis*, Mimeo, Shadow Open Market Committee, 2002.

[9] A. Stuart, Simple monetary policy rules, *Bank of England Quaterly Bulletin*, **36** (1996) 281-287.

[10] D. Ricardo, *The high price of bullion, a proof of the depreciation of booknotes*, In: Sraffa, P., Dobb, M. (Eds.), The works and correspondence of David Ricardo, Vol. III Cambridge University Press, Cambridge, UK, 1951.

[11] K. Wicksell, *Interest and Prices*. Translated, 1936, Kahn, R.F. (Ed.), Macmillan, London. Reprinted, A.M. Kelley, New York, NY, 1965.

[12] I. Fisher, *100% Money*, Alelphi, New York, 1935.

[13] A.W. Phillips, Stabilisation policy in a closed economy, *Economic Journal* **64** (1954) 290-323.

[14] F. Black, Uniqueness of the price level in monetary growth models with rational expectations, *Journal of Economic Theory*, **7** (1974) 53-65.

[15] S. Fischer, Long-term contracts, rational expectations, and the optimal money supply rule. *Journal of Political Economy,* **85** (1977) 191-205.

[16] B.T. McCallum, Price-level stickness and the feasibility of monetary policy stabilization policy with rational expectations, *Journal of Political Economy,* **85** (1977) 627-634.

[17] E.S. Phelps, J.B Taylor, Stabilizing powers of monetary policy under rational expectations, *Journal of Political Economy,* **85** (1977) 163-190.

[18] J. Benhabib, Adaptive monetary policy and rational expectations, *Journal of Economic Theory,* **23** (1980) 261-266.

[19] R.E. Lucas, *Rules, discretion and the role of the economic adviser*, in S. Fischer ed. Rational Expectation and Economy Policy, Chicago University Press-NBER, 1980.

[20] B.T. McCallum, The case for rules in the conduct of monetary policy: A concrete example. Federal Reserve Bank of Richmond *Economic Review* **73** (1987) 10-18.

[21] B.T. McCallum, Robustness properties of a rule for monetary policy, *Carnegie-Rochester Conference Series on Public Policy* **29** (1988) 173-204.

[22] B.T. McCallum, Commentary: How should central banks reduce inflation?- Conceptual issues. In: *Achieving Price Stability (A symposium sponsored by the Federal Reserve Bank of Kansas City),* 1996.

[23] D.K. Biederman, 2002, Stabilizing properties of monetary feedback rules: a representative-agent approach, *Journal of Economic Dynamics & Control*, **26** (2002) 1955-1974.

[24] J.B. Taylor, Discretion versus policy rules in practice. *Carnegie-Rochester Series on Public Policy* **39** (1993) 195-214.

[25] L.E.O. Svensson, Optimal inflation targets, conservative central banks, and linear inflation contracts, *American Economic Review*, **89** (1997) 98-114.

[26] G.D. Rudebusch, L.E.O. Svensson, *Policy rules for inflation targeting*, in Monetary Policy Rules, J.B. Taylor, ed. Chicago: Chicago University Press, (1999) 203-246.

[27] L. Ball, Efficient rules for monetary policy, *International Finance*, **2** (1999) 63-64.

[28] L.E.O. Svensson, M. Woodford, Indicator variables for optimal policy, *Journal of Monetary Policy*, (2002).

[29] A. Levin, V. Wieland, J.C. Williams, Robustness of simple monetary policy rules under model uncertainty, in *Monetary Policy Rules*, J.B. Taylor, ed. Chicago: Chicago University Press, (1999) pp. 203-246.

[30] A. Onatski, J.H. Stock, Robust monetary policy under model uncertainty in a small model of the U.S. economy, *NBER Working Paper* N. **7490**, 2000.

[31] J. Rotemberg, M. Woodford, An optimizztion-based econometric framework for the evaluation of monetary policy. In: Bernanke, B., Rotemberg, J. Eds., *NBER Macroeconomics Annual,* MIT Press, Cambridge, MA, 1997.

[32] A. Orphanides, S. van Norden, *The reliability of output gap estimates in real time*, Board of Governors of the Federal Reserve System, August, 1999.

[33] A. Orphanides, R. Porter, D.R. Reifschneider, R. Tetlow, R., Finan, Errors in the measurement of output gap and the design of monetary policy, *Journal of Economics and Business,* **52** (2000) 117-151.

[34] F. Smets, Output gap uncertainty: Does it matters for the Taylor rule? *Empirical Economics*, **27** (2002) 113-129.

[35] M. Ehrmann, F. Smets, Uncertain potential output: Implications for monetary policy, *Journal of Economic Dynamics & Control,* **26** (2003) 1611-1638.

[36] J.B. Taylor, The inflation/output variability tradeoff revisited, in J.C. Fuhrer, ed., *Goals, Guidelines, and Constraints Facing Monetary Policy Makers,* Federal Reserve Bank of Boston, 1994.

[37] U. Ozlale, Price stability vs. output stability: tales of federal reserve administration, *Journal of Economic Dynamics & Control,* **27** (2003) 1595-1610.

[38] R.L. Hetzel, The Taylor rule: Is it a useful guide to understanding monetary policy? *Federal Reserve bank of Richmond Economic Quarterly*, **86** (2000) 1-33.

[39] Y.Z. Tsypkin, *Adaptation and Learning in Automatic Systems*, Academic Press, New York, 1971.

[40] J.P. Judd, G. Rudebusch, Taylor's rule and the fed: 1970-1997. *FBR San Francisco Economic Review,* **3** (1998) 3-16.

[41] L. Ljungqvist, T. Sargent, *Recursive Macroeconomic Theory.* MIT Press, Cambridge, MA, 2000.

[42] S.G. Krantz, H.R. Parks. *The Implicit Function Theorem*. Birkhauser, Boston (2003).

[43] M. Woodford, Optimal monetary policy inertia, *NBER Working Paper* No. **7261**, 1999.

[44] G.D. Rudebusch, Term structure evidence on interest rate smoothing and monetary policy inertia, *Journal of Monetary Economics*, **49** (1999) 1161-1187.

[45] B. Sack, V. Wieland, Interest-rate smoothing and optimal monetary policy: A review of recent empirical evidence, *Working Paper,* Federal Reserve Board, 1999.

[46] R. Clarida, J. Gali, M. Gertler, The science of monetary policy: a new Keynesian perspective. *Journal of Economic Literature,* (1999) 1661-1707.

In: Emerging Topics in Macroeconomics
Editor: Richard O. Bailly, pp. 97-124
ISBN: 978-1-60692-093-0

Chapter 3

THE PARADIGMS WAR AND THE TWO WAYS OF ECONOMY LIBERALISING BY F. LIST

Dariusz Pieńkowski [*]
University of Physical Education in Poznań, Poland

> History teaches that wherever individuals are engaged in wars, the prosperity of mankind is at its lowest stage, and that it increases in the same proportion in which the concord of mankind increases. In the primitive state of the human race, first unions of families took place, then towns, then confederations of towns, then union of whole countries, finally, unions of several states under one and the same government. If the nature of things has been powerful enough to extend this union (which commenced with the family) over hundreds of millions, we ought to consider that nature to be powerful enough to accomplish the union of all nations. If the human mind were capable of comprehending the advantages of this great union, we ought to venture to deem it capable of understanding the still greater benefits which would result from a union of the whole human race. Many instances indicate this tendency in the spirit of the present times. We need only hint at the progress made in sciences, arts, and discoveries, in industry and social order. It may be already foreseen with certainty, that after a lapse of a few decades the civilised nations of the earth will, by the perfection of the means of conveyance, be united as respects to both material and mental interchange.
>
> ***F. List*** [1]

Introduction

Friedrich List (1789-1846) was a famous German critic of the classical policy which had been presented by A. Smith or J. B. Say. However, he wasn't against the crucial premises of economic liberalism. His point of view resulted from different methodological assumptions made under the philosophy of as well as the political and economic situation in Germany at that time. These premises were developed by the following generations of German

[*] E-mail address: darpie_xl@wp.pl
[1] List, 1909, 100-101

economists who took the social consequences of classical economic policy into consideration. Moreover, the methodology of historical schools has been widely undertaken by institutional schools in the USA. My paper focuses on the work of F. List who described the effective policy of the liberalisation of national economies. This idea, which was presented over 150 years ago, seems to characterise the present policy of the European Union. Thus, F. List presented not only an interesting historical analysis according to the conditions of his time but also described and predicted the effective policy of the liberalisation of national economies in Europe. However, the dominant economics schools of academia are still based on the classical and neoclassical methodologies, despite the fact that they are definitely presented less in the economic policy of the European Union; and his ideas are only treated as an interesting alternative for the economic policies of infant economies.

Moreover, his ideas are taken into consideration as the nationalist perspectives of economic development in contrast to the classical liberalism widely presented by English scholars. He was inspired by A. Smith's theory of economic liberalism, which was influenced by the social liberalism idea; but F. List, similarly to A. Smith, saw many benefits for the economy in economic liberalism and didn't criticize it. His criticisms of classical economics were mostly derived from the different methodology and the subsequent consequences on the economic policy. The methodology was based on a social perspective of economic activity with economic liberalism being treated as a kind of policy, not a universal rule of economic activity. Thus, his idea of protectionism was named contingent protectionism in contrast to mercantilism's protectionism based on the balance of trade surplus; F. List treated it as a path to education in an industrial society (Taylor, 1957, 214-215). This perspective tends to treat English economic liberalism as mercantilism's policy focused on the benefit of one nation, England (i. e. based on the balance of trade surplus).

This article focuses on some methodological perspectives and their applicability to the present from the F. List point of view. The political and economic backgrounds to Friedrich List's ideas are presented in the next chapter; further, the dominated philosophy and methodology of Germany and Britain is presented. Next, I described the main ideas of F. List in the context of German romanticisms, the methodological methods of economics, and socio-economic situation in Germany. Finally, I presented some conclusions and two perspectives of liberalisation of economy in the context of the European Union policy.

Political and Economic Backgrounds to Friedrich List's Ideas

Friedrich List was deeply involved in political activity and his scientific ideas alternated with his political goals[2]. His reformation views focused on the economic processes of Germany's

[2] He loosed lucrative jobs at the University and State administration because of his reformation ideas. He was sentenced to 10 months imprisonment with hard labor accused of acting in interests of foreign state and corruption; and he had to run away to France. When he had come back he had been arrested and when he had finished his sentence he decided to move to the USA. There he became a very famous economist and political activists as well as a successful businessman; he founded the journal *The Readinger Adler* (*Eagle*) and published *The Outlines of American Political Economy* (1827). 5 years later (1930) he came back to German proceeding with his reformation ideas that contributed to economic development of country (as an American consul); he was known as the *Father of German Railways* and strongly advocate of the *Zollverein.* In 1937 he was written his *The Natural System of Political Economy* and *The National System of Political Economy, vol. 1* in 1841; the latter he has never finished because of committed suicide (he was seriously ill and his business in the USA was getting to run wrong).

unification. The idea of trade liberalization widely promulgated by the British Empire and classical economists at that time seems to be the main source of his critiques. Germany had been divided into many small feudal counties and was politically split up. Actually, the Holy Roman Empire established in the late of the year 800 was "a political entity in name, the Empire was never one in substance. . . various Emperors. . . tried to weld the imperial provinces into a coherent unit, but they were never able to succeed, each failure leaving the German lords more independent" (Roussakis, 1968, 13). It had been destroyed by many years of wars in the seventeenth century and the process of industrialization was much slower, despite the fact that the manufacturing sector had been slowly developing under the pressure of changes in England. For example, C. Hayes wrote that during the conflict between the Habsburgs and the Bourbons, the so-called Thirty Years' War between 1618 and 1648 Germany:

> lost at least half of its population and more than two-thirds of its movable property. . . towns and villages were in ashes, and vast districts turned into deserts. Churches and schools were closed by hundreds, and religious and intellectual torpor prevailed. Industry and trade were so completely paralyzed that by 1635 the Hanseatic League was virtually abandoned, because the free commercial cities, formerly so wealthy, could not meet the necessary expenses. Economic expansion and colonial enterprise, together with the consequent upbuilding of a well-to-do middle class, were resigned to Spain, Portugal, Holland, France, or England, without a protest from what had once been a proud burgher class in Germany. This elimination of an influential bourgeoisie was accompanied by a sorry impoverishment and oppression of the peasantry. These native sons of the German soil had fondly hoped for better things from the religious revolution and agrarian insurrections of the sixteenth century; but they were doomed to failure and disappointment. The peasantry were in a worse plight in the eighteenth century in Germany than in any other country of western or central Europe (Hayes, 2004, 321).

As a result, Germany was politically destabilised, economically very weak, and a lot of towns were destroyed. Finally, the Treaty of Westphalia in 1648 definitely stopped the war and the unification processes that being left Germany as a composition of over three hundred sovereign states; the smallest (the estates of the Imperial Knights) consisted of a castle, some land around, and about 400 hundreds despotically ruled inhabitants (Roussakis, 1968, 14-15). Furthermore, in the eighteenth century there were political chaos between separated countries that independently run foreign policies (f. e. coalitions) broken by conflicts and wars between various dynasties:

> the Holy Roman Empire was an unimportant bond of union. Austria, traditionally the chief of the Germanies, was increasingly absorbed in her non-German possessions in Hungary, Italy, and the Netherlands. Prussia, the rising kingdom of the North, comprised a population in which Slavs constituted a large minority. Saxony was linked with Poland; Hanover, with Great Britain. Bavaria was a chronic ally of France. Add to this situation, the political domination of France or Sweden over a number of the petty states of the empire, the selfishness and jealousies of all the German rulers, the looming bitter rivalry between Prussia and Austria, and the sum-total is political chaos, bloodshed, and oppression (Hayes, 2004, 331).

The small German feudal counties were independently governed by methods based on various law systems, economic rules, and trade regulations stated by tariffs: "the good picture of the situation is the example of a merchant who was transporting goods from Hamburg to

Berlin. He had to cover as much as ten state borders or customs clearances. The same merchant on the way from Magdeburg to Dresden (180 km) paid duties 16 times!" (Stankiewicz, 2000, 192). T. Riha reported that "there were no less than 38 kinds of tariffs within the German Confederacy in 1818, (in Prussia alone there were 67 different tariffs) besides different currencies, weights and measures" (Riha, 1985, 60). E. Roussakis claimed that in 1790 there were " some eighteen hundreds customs frontiers in Germany" (Roussakis, 1968, 33). There were even coined their own money.

Commerce and industry developed in the previous centuries had flourished in England and the old craft gild system was broken in the eighteen century while it still maintained its hold in Germany (and most of the European countries). The gilds regulated production systems according to the interests of manufacturers. They effectively used their power to both limits competition and labour costs; and technological improvement was under control of the interests of owners to ensure their market position. In other words, the English economy was much more competitive than the German one strongly regulated by the craft gilds. The gild system completed government policy that used to regulate economic development. Mercantile policy widely regulated industrial production and commerce ensured State interests. France was influenced by the policy of mercantilist J. B. Colbert who postulated full control of all economic activities. System of tariffs effectively restricted the British industrial power during the colonial expansion while, at the same time, the German counties were open for mass exported English goods. English economic expansion weakened the Germany manufacturers, which contributed to widening the gap between Germany and England.

Moreover, unlike in Germany the Great Britain development of industry and commerce resulted in the rapid growth of cities; and the changes of social structure had been attended by it. In the beginning of "the eighteenth century [Great Britain] possessed a foreign commerce estimated at $60,000,000, and that of France was at least two-thirds as great. During the century the volume of commerce was probably more than quadrupled" (Hayes, 2004, 372). London, the biggest European city, had over one million inhabitants in 1800; and it was doubled from 1685 (Hayes, 2004, 369). There were some important trading centres in Germany like Hamburg, Bremen, or Frankfort but generally the leading position held British cities. The growth of cities went along with the growth a new social class i. e. a middle-class consisted of merchants, bankers, wholesalers and so on. They power had been attended by control over business but then it extended to all the world of officialdom by lawyers, judges, intendants, or government secretaries. Thus, they influenced economy as well as social and political life; consequently, reforms had been demanded. This class was mostly developed in England and France while the German society was being left behind with their aristocratic elites, poor peasants and labourers.

At the beginning of the nineteenth century the United Kingdom with a king and a Parliament presented the political and economic interests of England (Wales), Scotland, and Ireland (more or less stable situation); and simultaneously they governed over 30 colonies in the world. The power of the king was mostly shifted to the Parliament consisted of two bodies. The House of Lords represented aristocratic elites but the House of Commons tend to presentation the lower classes. It was a very progressive form of governing that reflected English liberal aspirations despite undemocratic character of it indeed (Hayes, 2004, 402). C. Fyffe wrote that "in the first Assembly of the Revolution it was usual to speak of the English as free men whom the French ought to imitate" (Fyffe, 1896, 41). This tendency was also being presented in the ideas of English philosophers; and philosopher A. Smith reflected the social and economic liberalism in his economics.

However, at the same time (i. e. in the late eighteen century) Prussia was being developed by the absolute power of Frederick II who reformed his country emphasising material prosperity and industrial development. He established a strong army to win new territories (as Silesia or Poland). However, the social system of the society was far the one in the Great Britain:

> the King's Government busied itself with every detail of town and village administration; yet along with this rigorous development of the modern doctrine of the unity and the authority of the State there existed a social order more truly archaic than that of the Middle Ages at their better epochs. The inhabitants of Prussia were divided into the three classes of nobles, burghers, and peasants, each confined to its own stated occupations, and not marrying outside its own order. The soil of the country bore the same distinction; peasant's land could not be owned by a burgher; burgher's land could not be owned by a noble. No occupation was lawful for the noble, who was usually no more than a poor gentleman, but the service of the Crown; the peasant, even where free, might not practise the handicraft of a burgher (Fyffe, 1896, 23).

The rest of the German counties were similarly organised based on the monarchy and the sharp division of society. German national spirit did not exist and was thwarted by the selfishness of the particular German rulers. Prussia first undertook the idea of unity the Holy Roman Empire of the German Nation against France.

At the beginning of nineteenth century Europe was dominated by the tyranny of Napoleon Bonaparte who effectively take-over power in a vast part of Prussia, Austria, and many other German counties (1808); Prussia was deeply degraded and many other German counties accepted protection of Napoleon (the Confederation of Rhine: the kings of Bavaria and Württemberg, the grand-dukes of Baden, Hesse-Darmstadt, Berg, and so on). Whenever Napoleon had got his power over a county the Code Napoleon was implemented: the abolition of feudalism and serfdom, the recognition of equality of all citizens before the law, and so on. Napoleon also simplified the ruled states to make them more manageable. It was a milestone in unification of Germany; moreover, the common enemy rushed the unification.

The movement to free German-speaking counties grew apace during the next 5 years (1813). Prussian monarchy was reformed and the years between 1807-1813 were focused on internal reforms following the French ones made almost 20 years earlier. There were a lot of both social and economic changes as abolishon of serfdom despite peasants were bound to pay rents to landlords; or free trade was established. Finally, the wars with Napoleon caused unification processes in Germany; however, other European countries (as England, France) were already governed by stable and stronger political powers.

The new rules provided by Napoleon divided German society. On the one hand, there were conservatives ready to maintain the tradition or the great power of the Holy Roman Empire; moreover, some of them wanted to unify Austria and Prussia, while some others postulated the Prussian or Austrian rules. They presented nationalistic and romantic movements desired to restore the old Empire. On the other hand, there were liberal postulates emerged to establish a German republic; they were against the interests of the selfish rules of the small counties (Roussakis, 1968, 19). C. Fyffe emphasised that

> it was characteristic of Germany that the demand for free government came not from a group of soldiers, as in Spain, not from merchants and men of business, as in England, but from professors and students, and from journalists, who were but professors in another form. The middle class generally were indifferent: the higher nobility, and the knights who had lost their semi-independence in 1803, sought for the restoration of privileges which were really incompatible with any State-government whatever. The advocacy of constitutional rule and of German unity was left, in default of Prussian initiative, to the ardent spirits of the Universities and the Press, who naturally exhibited in the treatment of political problems more fluency than knowledge, and more zeal than discretion (Fyffe, 1896, 451).

Liberal ideas grew up mostly in critics written by intellectualists as a solution for all the political and economic problems of Germany; thus, they were being mostly presented in newspapers or pamphlets. This power become more influential along with the development of a middle class in Germany as merchants and manufacturers. They interests were more practical and their influence was in line with the power their had got as the crucial actors of economic development. The intellectualists created then theoretical arguments for the middle class interests; and they could undertake more practical activities when the interests of practitioners supported them.

Unification processes started from the Federal Act of 1815 that established the German Confederation of almost forty German states coordinated by the Diet in Frankfurt am Main. It was also used to protect the German order against liberals and nationalists; i. e. the source of the French Revolution. At that time 75% of Germans were farmers mostly settled in the Eastern part, while the Western part was more urban; emancipation of the peasantry was a gradual process and it was not so dramatically conducted like in France. Moreover, there was not a national German market because of the system of tariffs and considerably lower industrialization. Most of the industries were supervised by state and their development was in line with the state political interests. There were no income taxes and customs as well as excise duties were the main source of state revenues. German cameralism had a long tradition in economics, contributing to a broad analysis of economic theories in the context of State activities. E. Reinert wrote that "Whereas early recruitment to the economics profession in England came from the merchants, the ranks of economists were filled from the public sector in Germany. German Cameralists - the equivalent of mercantilists - were generally employed as keepers of the treasury or Schatzkammer - therefore the term cameralists. The German cameralist literature is huge, but largely unknown outside Germany. Like the English mercantilists, the cameralists favoured the protection of manufacturing" (Reinert, 1994, 13).

The wartime resulted in a lack of labour and capital. In the peacetime most of the industries lost their market during the wartime and they were then overtaken by British competitors; but iron industry developed during the war was not adjusted to the market conditions. The German states overtaken by Napoleon developed their markets mostly according to the French interests but in the aftermath of the peacetime even the market was lost as being protected by the French tariff laws. Furthermore, the German transport infrastructure (roads and rivers) was very bad and it technically limited economic development.

There were legally established three economic unions in the 1928 among the German states:

1) In the north, there was the Prusso-Hessian Union - first, it begun from reforms within the Prussia by R. Stein since 1814 (and C. F. Nebenius since 1819), next it was continued with the enclaves surrounded by Prussian territory since 1818, and finally, with two independent sovereign states Prussia and Hesse-Darmstadt. The customs tariffs were being lowered to encourage import, to avoid undue offence from neighbours, and to render smuggling unprofitable.
2) In the south, there was the Bavaria-Württemberg Customs Union firstly organised by F. List with a group of merchants and manufacturers – it begun from meetings, conferences, and petitions to the Diet in 1818 on the abolition of all the internal customs and the adoption of a common external system of tariffs; and the external tariffs policy should be based on the principle of retaliation against foreign nations to let them recognize the principle of European free trade (an ultimate goal of the Union).
3) The Middle German Commercial Union as a response to the attempts of the both Unions to enlarge their customs systems – they declared to further the trade between members, to maintain trade roads, and they were pledged not to join any other union. However, they did not work out a common customs system. This union was welcomed abroad as the Great Britain, the Netherlands, France, or Austria; it was a remedy for the progress toward German economic unification.

The third Union paved the way for rapprochement between the rests, which trade roads, could be partially controlled by the third one. It resulted in alliance between the northern and southern unions in 1829; and it was the next step to creation the Zollverein (i. e. an economic customs union of the German states). Finally, a series of treaties were signed in 1833 following some years of various facilitations between the north and the south, and the particular states of the Middle Union. It induced various offences of the neighbouring countries and the Great Britain, which had taken advantage of the German division. The lost of supremacy of British commerce had been expressed in the Great Britain; and it was being observed soon. Faced with this situation they tried to play off the various German states against one another, they changed they customs tariff policy against the Zollverein, or they tried to pave the way for a commercial agreements with the Zollverein.

The economic unification induced further political and social changes: "from the sleepy land, whose progress was paralyzed by territorial fragmentation, poor communications, and outmoded methods of production, Germany has turned into a bustling nation with new forms of production, new mean of transportation, new social classes, new civic ideas – and all this in unique dimensions" (Roussakis, 1968, 77). E. Roussakis reported numerous examples of it as the emancipation of peasantry completed by 1850 (in the West by 1830), a population increase (i. e. manpower and a larger output for manufactures) by almost 60% between 1816 and 1865, an animal production increase in the same period by 213%, and a vegetable production increase by 62%. The most fruitful period of industrial development between 1850 and 1873 was stimulated by technological changes; in this period Germany widely adapted the improvements of the British Industrial Revolution (Roussakis, 1968, 86), which were widely diffused 50-80 years earlier in the Great Britain.

One of the key sources of the economic development was the improved German transportation infrastructure as roads, railways, and river transportation. A large common home market unleashed the spirit of enterprise suppressed by political boundaries as by the end of the nineteenth century the urban population exceeded the rural one. Nevertheless, it was vigorous supported by states policy and government investment. The German economy "had gradually been transformed from one of the best markets for foreign manufactures to a largely self-sufficient industrial country" (Roussakis, 1968, 89) and a strong competitor of the Great Britain or the United States.

F. List economics coped with the period of the German economic and political transition; moreover, he was a one of the key theoretical and practical designers of it successes and a visionary. Nevertheless, his methodology of economics also reflected the German spirit of that time that was to be in line with the romantic philosophy next undertaken by the German Historical Schools.

Philosophical and Methodological Backgrounds to Economic Liberalism in Germany and Britain

It was conducive to the development of national ideology, postulating the political, cultural, and economic unity of Germany:

> The period of history which is commonly called "modern" has a mental outlook which differs from that of the medieval period in many ways. Of these, two are the most important: the diminishing authority of the Church, and the increasing authority of science. With these two, others are connected. . . There is a tendency for the rich merchants to become absorbed into the aristocracy. From the time of the American and French Revolutions onwards, democracy, in the modern sense, becomes an important political force. Socialism, as opposed to democracy based on private property, first acquires governmental power in 1917. This form of government, however, if it spreads, must obviously bring with it a new form of culture; the culture with which we shall be concerned is in the main "liberal," that is to say, of the kind most naturally associated with commerce. To this there are important exceptions, especially in Germany; Fichte and Hegel, to take two examples, have an outlook which is totally unconnected with commerce. But such exceptions are not typical of their age (Russell, 1945, 491-492).

B. Russell described the history of German philosophy in refer to the socio-political circumstances of that time. He described the specific backgrounds to the German philosophers: "the intellectual predominance of Germany is a new factor, beginning with Kant. Leibniz, though a German, wrote almost always in Latin or French, and was very little influenced by Germany in his philosophy. German idealism after Kant, as well as later German philosophy, was, on the contrary, profoundly influenced by German history; much of what seems strange in German philosophical speculation reflects the state of mind of a vigorous nation deprived, by historical accidents, of its natural share of power" (Russell, 1945, 719-720). Moreover, E. Taylor pointed out that the economic situation there made the classical methodology difficult for German scholars, because of the wider gap between the abstractive models of perfect competition and the virtual market economy of Germany that remained under the influences of state interventionism and political nepotism at the time (Taylor, 1957, 206).

The circumstances that shaped the ideas of the German philosophers similarly shaped the theories of F. List who was born in Swabia; it was a free city municipal self-governed under a democratic constitution placed in a region with its own characteristic, dialect, and traditions.

W. Henderson wrote that “the district has nurtured philosophers such as Hegel and Schelling, and poets as Schiller, Hölderin, and Unland – men who combined a strong local patriotism with an ability to face the challenge of problems of universal significance” (Henderson, 1983, 2). Nevertheless, F. List was a reformer, a revolutionist, and a liberalist for the traditionalist in his state. He postulates in many petitions reforms of civil servants or municipal finances, as well as freedom from serfdom, freedom to express one’s view at meetings or in the press, or liability to pay taxes according to one’s means; his reforms were furnished by practice. Thus, as a civil servant he usually criticised clerks and the procedures used by the clerks whom he worked with or the municipal practices when he reorganised municipal finances; it usually resulted in opposition and aversion the people whom he worked with. Eventually, he was an enemy of the state sentenced to ten months’ imprisonment in a fortress in 1922; a Police Commissioner suppressed his leaflets in 1921 as he claimed that civil servants are linked by a common training, interests, and family ties and “enjoy a high standard of living at the expenses of those whom they were supposed to served” (Henderson, 1983, 55). At the same time he represented nationalistic postulates of Germany from the perspective of the Great Britain policy.

Actually, he was a practitioner focused on solving the problems he met as a citizen, a civil servant, a politician, a businessman, or a journalist. This practical perspective also shaped his life and political activity in the USA and he proceeded with the investigation of economic and political conditions of economic unification in accordance with the economic situation of the USA (1825-30). E. Roussakis reported that in the USA “List also profitted greatly from his observation of the young economy which surrounded him” (Roussakis, 1968, 95); he observed how low-priced British goods flooded this agrarian economy similar to the German one. Thus, F. List’s economic protectionism didn’t only reflect German political nationalism (even if it can be justified by the situation of that time) but it also had roots in the social philosophy that provided for the historical schools other than the methodological perspective of economic analyses. Put it another way, the experience he practised was then generalized; thus, he moved from specific observations to broader generalizations and theories as it is in the inductive methods of reasoning: „he was influenced not by writings of American protectionism but by the experience he had as a newspaper editor, a farmer, and a promotor of mining and railway undertakings. . . his writings on fiscal policy and the promotion of industry were not based entirely upon book knowladge but were the fruits of practical experience in the world of business“ (Henderson, 1983, 155). The experience has been reflected in the economic ideas of F. List and the different method of reasoning (i. e. induction), better postulated in the classical economics by J. S. Mill (Mill, J. S., 1843)[3].

At the time of F. List, German philosophy remained with a romantic outlook which was in contrast to both the classical one and the empiricism based on the materialism of the Enlightenment:

> Enlightened princely despotism was the representative and leader of this great progressive movement; a movement which was destined to annihilate the freedom of the Estates and corporations, to establish freedom of trade and great markets at home, and to combine all the resources of the country, economic as well as financial and military, in face of the foreigner. Those states most quickly became powerful and rich, which carried out this centralising

[3] Nevertheless, as M. Blaug emphasized it, J. S. Mill proposed inductive method a posteriori not as a means of discovering truth, but of verifying it (Blaug, 1992, 58).

> tendency with the greatest energy. Germany remained so far behind the greater Italian states, behind Burgundy, Holland, England, and France, behind even the smaller northern states, because it remained fast bound by mediaeval forms; because, moreover, even its greater territories were too small, too fragmentary, too far from the coast, to pursue this new kind of centralising policy like the western states of Europe (Schmoller, 1897, Appendix I).

Moreover, R. Solomon observed that "The Enlightenment seemed to many Germans to be vulgar, overly concerned with economics, and contemptuous of religion and the 'spiritual' aspects of social life" (Solomon, 1993, 185).

Germany had different economic conditions as well as different philosophical and institutional outlooks. F. List as well as A. Smith mirrored in their methodologies the dominating social philosophies of Germany and England; i. e. the idealism of I. Kant, G. Hegel, or J. G. Fichte and the empiricism of R. Bacon, J. Lock, or D. Hume. Thus, in their economics, both of the economists presented the dominant social philosophies in their nations of the time. Seeing the institutional state at that time from the current point of view, rather the classical schools had perfect political conditions for their ideas of economic liberalism in the Great Britain just as the historical schools in Germany flourished then at the time of G. Schmoller. For example, in the context of the political power of G. von Schmoller, L. Robbins described the German academia as follows: "And Schmoller was an extremely powerful person. At the stage in the history of united Germany, it was difficult to get a Chair, and Chair meant something in Germany in those days – Herr Professor Doktor von So-and-So [von Schmoller]" (Robbins, 1998, 245).

The philosophy of the German idealists is mirrored in the methodological perspective of F. List. The concept of Spirit (Geist) seems to be the key element to define the national perspective of F. List; G. W. F. Hegel an enthusiast of the French Revolution (F. List was also a republican), vastly investigated it in his works. He understood it as "the ultimate unity of the whole of humanity, and of humanity and the world" (Solomon, 1993, 186). The Spirit is the recognition of consciousness other than one's own; thus, the individuality, in his philosophy, is defined by other individuals:

> Self-consciousness exists in itself and for itself when and by the fact that it so exists for another; that is, it exists only in being acknowledged. . . Self-consciousness is faced by another self-consciousness; it has come out of itself. This has a twofold significance: first, it has lost itself, for it finds itself as an other being; secondly, in doing so it has superseded the other, for it does not see the other as an essential being, but in the other sees its own self (Hegel, 1998, 111).

L. Rauch compared the concepts of an individual presented by the English philosophers as J. Lock or T. Hobbes and G. W. F. Hegel:

> Hobbes and Locke regard their protagonist as psychologically complete, endowed with the various personal characteristics shared by the rest of us (although Hobbes and Locke differ diametrically in regard to what those "common" characteristics are). Hegel, on the other hand, regards that picture of the primal human as incomplete: Rousseau introduced the notion of that human as but half-human; Hobbes's and Locke's man is someone we would readily recognize; Hegel sees this creature as incomplete since he will only begin to define himself through his encounter(s) with others. . . Hobbes and Locke see my need for personal security as my reason for entering into social relations; Hegel sees those relations emerging from a recognition of others as beings similar to myself. . . There are dialectical elements in all this as

> well, and these elements, in their very contradictoriness, actually propel the dialectic into creating a social fabric. Thus the fact that a thing is mine (dependent on me) and also it (an object that is independent of me) can be said to raise difficulties which only a society can be called upon to resolve (Rauch, 1993, 271-273).

In G. W. F. Hegel's view is emphasised the social nature of an individual. Consequently, society is much more than a collection of individuals and the Spirit is a product of social life, however, "the Enlightenment as the embodiment of a false because anti-spiritual effort to build a truly universal society" (Solomon, 1993, 208). J. G. Fichte earlier "asserted that the true motive of morality is not the salvation of the individual man but the Progress of humanity" (Bury, 2004, 166). In G. W. F. Hegel's theory an individual as well as a family (two individuals absorbed), or a civil community wherein the family is absorbed in a broader context is an entity; and finally, "the state is the realized ethical idea or ethical spirit. It is the will which manifests itself, makes itself clear and visible, substantiates itself. . . The state finds in ethical custom its direct and unreflected existence, and its indirect and reflected existence in the self-consciousness of the individual and in his knowledge and activity. Self-consciousness in the form of social disposition has its substantive freedom in the state, as the essence, purpose, and product of its activity. . . as universal idea, or kind, or species, it has absolute authority over individual states. This is the spirit which gives itself reality in the process of world-history" (Hegel, 2001, 194-195, 198). Actually, civil society is referred to the individual needs, while the state (universality) goes beyond the particularity (civil society) and subjectivity (an individual). Civil society consist of motivated by private interests individuals that excludes the others' interests, while the state is qualitatively different; it is the dissimilar (dialectic) continuation of the civil society:

> a collective entity wherein individuals fulfill their separate interests by merging them into the interests of the whole. . . The libertarian tradition of Anglo-American thinking sees the individual as the most fundamental political entity; with this in mind, the state is seen as a device for serving the interests of individuals—and therefore the main stress is placed on the limits to be imposed on state power. Hegel, on the other hand, comes from a holistic tradition wherein state power is irreducible to the interests of individuals. The state is a higher entity—even something approaching the mystical—into which the goals of individual persons are to be dissolved (Rauch, 1993, 281).

The state is a syntesis, a higher outcome of the dialectical process between a thesis (the privet interests) and a antythesis (the social ones). For example, for J. Locke the state is a contract between a people born as free individuals:

> men being, as has been said, by nature, all free, equal, and independent, no one can be put out of this estate, and subjected to the political power of another, without his own consent. The only way, whereby any one divests himself of his natural liberty, and puts on the bonds of civil society, is by agreeing with other men to join and unite into a community, for their comfortable, safe, and peaceable living one amongst another, in a secure enjoyment of their properties, and a greater security against any, that are not of it (Locke, 1824, 186).

The contract is conducted to realize the interests of individuals; thus, it is not presented in the terms of the "Idealistic Spirit" and the next, higher stage of the dialectical process of social evolution.

Art, religion, philosophy, and all the creations of social man are a process of the development of the state; they belong to a higher stage of Spirit's self-revelation (Bury, 2004, 165). It should be emphasized that the Spirit embraces both all the material and nonmaterial human advances. The Spirit continually had moved from one nation to another to realize the successive stages of its self-consciousness. It started in China through India, Greece, Rome, and finally its development was completed in the Germanic world; in his philosophy, it is a modern world, the final stage of its development wherein freedom had already been attained (all people are free and the political form is constitutional monarchy). The each nation has its national distinctive Spirit (Volksgeist) on the World Spirit's way in history; it is defined by its culture. G. W. F. Hegel unlike the Enlightenment philosophers (f. e. D. Hume) who argued for the unchanging character of human nature emphasized the evolutionary perspective in the development of the state; thus, human nature should be seen as the process of changes aims at realizing the absolute goal i. e. freedom (as being fully self-conscious, act freely). The rational goals of the Spirit (as freedom) are realized through rational as well as irrational actions; both the types of actions are elements of human history stemmed from human passions, needs, or interests (Rauch, 1993, 255-257).

The investigation of F. List's concept of economics should then proceed with the presentation of premises on the character of social systems. It is involved with a methodological discourse in the context of the explanation. Methodological holism "holds that the laws of more complex cases (e. g. from laws of collective behavior of society) are not deducible in any way from laws of less complex cases (e. g. from laws of individual behavior of society). Its opposite is methodological individualism, the view that such deducibility is possible" (Iannone, 2001, 246). Thus, "the principle of choice for an association of men is interpreted as an extension of the principle of choice for one man" (Rawls, 1999, 21). Consequently, the wealth of a society is a sum of individuals' wealth as in A. Smith's conception or it rather states a new quality formed by individuals as in F. List's economics: "'Labour forms the fund from which every nation derives its wealth, and the increase of wealth depends first on the productive power of labour, namely, on the degree of skill, dexterity, and judgment with which the labour of the nation is generally applied, and secondly, on the proportion between the number of those employed productively and the number of those who are not so employed.' From this we see how clearly Smith in general perceived that the condition of nations is principally dependent on the sum of their productive powers" (List, 1909, 109); F. List was aware the methodological difference and he emphasised the quality determinants furnished by social life which are mirrored in the industry. For F. List a nation, at this stage of socio-economic development, represents "qualitative value added" which is not a simply sum of its individuals; this value is created by specific institutions for each nation or a group of nations because from his historical studies the main factors promoting economic growth were national unity, political freedom, government policy, and the intelligence and industriousness of the people (Henderson, 1983, 167). "Adam Smith has on the whole recognised the nature of these powers so little, that he does not even assign a productive character to the mental labours of those who maintain laws and order, and cultivate and promote instruction, religion, science, and art. His investigations are limited to that human activity which creates material values" (List, 1909, 111). The key concept reflected G. W. F. Hegel's perspective is a productive power[4]. He understood it as political, administrative, and social institutions, natural and human resources, industrial establishments, and public works (Henderson, 1983, 165). The concept reflected the material and nonmaterial values as the concept of the Spirit. It was placed it in the wider social context and all the relationships between economy and society; and the state (or in the future a group of the culturally similar states), in F. List's opinion, is the best level of its organisation at the present stage of social development.

F. List was induced by A. Smith theory of economic liberalism: "not only a very faithful disciple of (Adam) Smith and (J. B) Say, but a very zealous teacher of the infallible doctrine. But the slump in Germany that followed the collapse of Napoleon's Continental System led him to revise his views on fiscal policy. When Napoleon's empire fell, the ports of the Continent were opened and British manufactured goods flooded the German market" (Henderson, 1983, 143). Thus, his liberalism had the same features, but another meaning. This methodology was based on a social perspective of economic activity with economic liberalism being treated as a kind of policy, not a universal rule of economic activity: "and in the last stage, after reaching the highest degree of wealth and power, by gradually reverting to the principle of free trade and of unrestricted competition in the home as well as in foreign markets, that so their agriculturists, manufacturers, and merchants may be preserved from indolence, and stimulated to retain the supremacy which they have acquired" (List, 1909, 94).

Furthermore, his method of reasoning was inductive and concrete while A. Smith, on the contrary, used deductive and abstract methods. The method utilised by F. List "was in fact the staple of the German Historical School, which was a dominant school of economics in continental European countries before the Second World War" (Chang, 2002, 6). J. Schumpeter implemented it to neoclassical economics "provides a bridge from the German historical school to the world of Anglo-Saxon economic theory. In Anglo-Saxon economics Schumpeter stands out as being much more original than he is to someone who has a knowledge of the German historical school of economics" (Reinert, 1994, 7). This was mostly determined by the positions of the two economists. W. Henderson commented his work The Natural System of Political Economy written in 1837 and better-developed four years later as The National System of Political Economy wrote that "although for a short time he had once been a professor and at Tübingen, he was no academic. He was a politician, a journalist, and a businessman" (Henderson, 1983, 165) and once again as I have pictured it above a practitioner; his economics was mostly defined by the problems he discovered in his practice despite it then reflected the specific philosophy of Germany. Driven by practice he postulated policy to improve the condition of the state which he lived, while A. Smith was a philosopher, not without practice, but influenced by the philosophies of J. Lock or his friend D. Hume; the latter created an abstractive model based on the theoretical premises not directly provided by practise. Put it another way, it reflected the British society philosophy (i. e. individualistic ethics) and it was a general theoretical model that is only practical if all the premises are the truth; but F. List only presented when it is possible in the course of socio-economic development.

In F. List's opinion, the political economics of the classical schools presented the political interests of the British Empire as he was going to do so, according to German

[4] Whenever I compare F. List's theory to some German philosophers I do not claim that he was directly inspired by it in the way as A. Smith was directly inspired by the works of J. Locke; it rather reflected similar philosophical notions.

interests, while classical economists claimed that they described the very universal mechanism of economic activity. R. Norgaard (Norgaard, 1994, 62) described the former as contextualism (i. e. phenomena are contingent upon a large number of factors particular to the time and place), while the latter was described as universalism (i. e. diverse, complex phenomena are the results of underlying universal principles, which are few in number and unchanging over time and space). Nevertheless, the critiques of A. Smith's ideas include the acceptance of economic liberalism. However, F. List dealt with it in the context of a policy based on the minimalization of State interventionism, not a lack of any economic policy. Hence, economic liberalism was a style of governing ("a kind of control" or under the control of the State), not a lack of any style of governing:

> It was seen as a self-generating process in which a variety of social forces interacts to produce institutional change. The state played an important facilitating role in the process, however, providing the legal framework that would give free play to the acquisitive motive and protecting property rights in the gains. Implicit in Smith's argument was a hidden assumption about economic institutions: The institutional structure within which acquisitive behavior would lead to progress was either there to begin with or would itself be created by individual action. This was tantamount to assuming either the prior existence of a system of self-adjusting markets or that such an institutional structure would naturally evolve over time as part of the of the progress of opulence. . . The institutional framework within which acquisitiveness could lead to economic progress did not just appear by itself; it had to be created by community action. List, however, applied this idea only to the infrastructure of institutions within which of system of self-adjusting markets functions, not to the system of markets itself. As did Smith, he assumed that to be natural development, if government was able to create the proper environment for individual action (Fusfeld, 1977, 745).

A. Smith created a universal economic law then developed by neoclassical economists in the context of the market efficiency and general market equilibrium, while for F. List economic activities are a process of social development defined by the stages of socio-economic maturity and political circumstances; A. Smith's economics was only applicable at last stage of the evolutionary changes (or between the societies that become at the similar stage[5])

The social perspective was then developed by the following generations of the Historical Schools which contributed to the famous scientific discussion on economic methodology (Methodenstreit) at a time when the historical schools became mainstream economics in German academia. The historical schools' methodology influenced the American scholars who worked in Europe and has proceeded the criticism of the neoclassical economics in the institutional schools developed in the USA and Europe.

F. List can be seen as an opponent of both the British interests and the liberal philosophy mirrored in A. Smith's economics despite that he was also seen as a liberal and an enemy of German interests at his time in Germany; he was too progressive because of his reformation views: "in Germany, List is therefore seen as a free-trader, in the Anglo-Saxon world he is seen as the incarnation of protectionism. Lists argument for free trade in a larger geographical area suggests an understanding of the importance of scale in manufacturing" (Reinert, 1994, 13). The above relativism will be further discussed in the context of both the neo-liberal and

---

Moreover, Hegelian development is more revolutionary, while Listian is rather an evolutionary process.

[5] However, it is important to notice that the industrial development is broadly understood in the context of cultural, religious, scientific, and technological advancements.

the evolutionary standpoints. In the next chapter I present the main ideas of F. List's economics that should be understood in the context of the methodological and historical backgrounds explained above.

Friedrich List's View of National Political Economics

F. List, according to both the situation in Germany drafted above and his methodology, postulated a new outlook on political economy. In his opinion, A. Smith and also J. B. Say omitted a very important social structure developed during the processes of social development. It was created in a long period of historical development and was based on the cultural and social unity of its members:

> From a disorganising particularism and individualism. . . considers private industry only as it would develop itself under a state of free interchange with society (i.e. with the whole human race) were that race not divided into separate national societies. Between each individual and entire humanity, however, stands THE NATION, with its special language and literature, with its peculiar origin and history, with its special manners and customs, laws and institutions, with the claims of all these for existence, independence, perfection, and continuance for the future, and with its separate territory; a society which, united by a thousand ties of mind and of interests, combines itself into one independent whole, which recognises the law of right for and within itself, and in its united character is still opposed to other societies of a similar kind in their national liberty, and consequently can only under the existing conditions of the world maintain self-existence and independence by its own power and resources (List, 1909, 141).

Consequently, the political economists of the classical schools were mistaken. He named their economic policy cosmopolitical policy, which indicated its universal character: "He [A. Smith] entitles his work, 'The Nature and Causes of the Wealth of Nations' (i.e. of all nations of the whole human race). He speaks of the various systems of Political economy in a separate part of his work solely for the purpose of demonstrating their non-efficiency, and of proving that 'political' or national economy must be replaced by 'cosmopolitical or world-wide economy'" (List, 1909, 97-98). However, the interests of individual countries are varied and an economic policy should represent national interests. This resulted from the stage of social development at that time, but A. Smith omitted it, taking into account the same interests of the whole human race. This is involved in the above presented premises on the character of society. Moreover, F. List saw classical liberalism as an expression of the interests of the British Empire.

According to F. List's methodology, he set the economy of individuals against the economy of society. The latter is the subject of political economy (i. e. national) and cosmopolitan economy. However, cosmopolitan economy is applicable to only one global society living in a perpetual state of peace that represents all the national interests. The latter condition is derived from the historical investigations of F. List, which showed that a war contributes to the degradation of an economy, weakening the economic position of the involved countries. A political agreement should ensure the state of economic competition presented in the classical schools. In other words, national interests will identify the interests of the whole human race, which is the next stage of social development[6]:

[6] See also the motto of this paper.

> Thus history shows that restrictions are not so much the inventions of mere speculative minds, as the natural consequences of the diversity of interests, and of the strivings of nations after independence or overpowering ascendancy, and thus of national emulation and wars, and therefore that they cannot be dispensed with until this conflict of national interests shall cease, in other words until all nations can be united under one and the same system of law. Thus the question as to whether, and how, the various nations can be brought into one united federation, and how the decisions of law can be invoked in the place of military force to determine the differences which arise between independent nations, has to be solved concurrently with the question how universal free trade can be established in the place of separate national commercial systems (List, 1909, 92).

and

> If, as the prevailing school requites, we assume a universal union or confederation of all nations as the guarantee for an everlasting peace, the principle of international free trade seems to be perfectly justified (List, 1909, 100).

or

> The popular school has assumed as being actually in existence a state of things which has yet to come into existence. It assumes the existence of a universal union and a state of perpetual peace, and deduces therefrom the great benefits of free trade. In this manner it confounds effects with causes. Among the provinces and states which are already politically united, there exists a state of perpetual peace; from this political union originates their commercial union, and it is in consequence of the perpetual peace thus maintained that the commercial union has become so beneficial to them. All examples which history can show are those in which the political union has led the way, and the commercial union has followed. Not a single instance can be adduced in which the latter has taken the lead, and the former has grown up from it (List, 1909, 102-103).

Referring to some methodological issues presented above, F. List applied the holistic approach, that included the situation and relationships between countries all over the world, while A. Smith extrapolated from British culture and economic conditions to the situation in the rest of the world. Consequently, the latter proposed universal principles of the wealth of nations without any participation of the state (a lack of economic policy was more clearly articulated by J. B. Say): "The natural effort of every individual to better his own condition. . . is so powerful, that it is alone, and without any assistance, not only capable of carrying on the society to wealth and prosperity, but of surmounting a hundred impertinent obstructions with which the folly of human laws too often encumbers its operations" (Smith, 1838, 221). It is derived from the social context of economic development and this social perspective of it is the crucial characteristic of his methodology. The socio-economic context justifies the best economic policy that presently can be effectively made at the national level; thus, economic liberalism has its historical context. Moreover, similarly to A. Smith, he described the stages of economic development: "According to the natural course of things, therefore, the greater part of the capital of every growing society is, first, directed to agriculture, afterwards to manufactures, and last of all to foreign commerce. This order of things is so very natural that in every society that had any territory it has always, I believe, been in some degree observed"

(Smith, 1838, 157)[7]. However, his analyses compromised the classical causes of it, i.e. capital accumulation and individual acquisitiveness depicted by A. Smith as the main causes of economic progress: "In the midst of all the exactions of government, this capital has been silently and gradually accumulated by the private frugality and good conduct of individuals, by their universal, continual, and uninterrupted effort to better their own condition. It is this effort, protected by law and allowed by liberty to exert itself in the manner that is most advantageous, which has maintained the progress of England towards opulence and improvement in almost all former times, and which, it is to be hoped, will do so in all future times" (Smith, 1838, 142). On the contrary, F. List's standpoint is obtained from the social perspective causes of wealth. The critiques are based on the distinction between the sources of the wealth and the meaning of the wealth. The former deals with the possession of wealth in terms of materialistic goals, while the latter is concerned with the possession of the abilities for the production the wealth, i. e. the power of producing it. Thus, wealth can be achieved by conquest, a colonial policy, or a trade policy. However, the wealth of nation can be ensured if the society is able to produce a bigger wealth than it is consuming. The latter refers to the power of producing goods that is not based on the labour or the division of the labour, but on the power of the nation:

> It would be more correct to describe the limbs of men (the head, hands, and feet) as the causes of wealth (we should thus at least approach far nearer to the truth), and the question then presents itself, what is it that induces these heads, arms, and hands to produce, and calls into activity these exertions? What else can it be than the spirit which animates the individuals, the social order which renders their energy fruitful, and the powers of nature which they are in a position to make use of? . . The greatest portion of the consumption of a nation is used for the education of the future generation, for promotion and nourishment of the future national productive powers. The Christian religion, monogamy, abolition of slavery and of vassalage, hereditability of the throne, invention of printing, of the press, of the postal system, of money weights and measures, of the calendar, of watches, of police, 'the introduction of the principle of freehold property', of means of transport, are rich sources of productive power (List, 1909, 110-113).

It also indicates the main goals of social development identifying the wealth of nations with social goals. Thus, the capital accumulation is a mean of social development and the wealth has a social meaning; that is why the only a nation is able to develop the qualities that are not a sum of the qualities of individuals. It also explains the national character of economic policy and the socio-economic states of liberal economic policy (cosmopolitical policy). Moreover, F. List's perspective is focused on the future and when said in the language of the present policy, postulated some ideas of sustainable development. As is done in the Bible, he told a parable to explain it:

> Let us suppose the case of two fathers of families, both being landed proprietors, each of whom saves yearly 1,000 thalers and has five sons. The one puts out his savings at interest, and keeps his sons at common hard work, while the other employs his savings in educating two of his sons as skilful and intelligent landowners, and in enabling the other three to learn a trade after their respective tastes; the former acts according to the theory of values, the latter according to the theory of productive powers. The first at his death may prove much richer

[7] F. List depicted the stages as follows: "As respects their economy, nations have to pass through the following stages of development: original barbarism, pastoral condition, agricultural condition, agricultural-manufacturing condition, and agricultural-manufacturing-commercial condition" (List, 1909, 143).

> than the second in mere exchangeable value, but it is quite otherwise as respects productive powers. The estate of the latter is divided into two parts, and every part will by the aid of improved management yield as much total produce as the whole did before; while the remaining three sons have by their talents obtained abundant means of maintenance. The landed property of the former will be divided into five parts, and every part will be worked in as bad a manner as the whole was heretofore. In the latter family a mass of different mental forces and talents is awakened and cultivated, which will increase from generation to generation, every succeeding generation possessing more power of obtaining material wealth than the preceding one, while in the former family stupidity and poverty must increase with the diminution of the shares in the landed property. So the slaveholder increases by slave-breeding the sum of his values of exchange, but he ruins the productive forces of future generations. All expenditure in the instruction of youth, the promotion of justice, defence of nations, &c. is a consumption of present values for the behoof of the productive powers. The greatest portion of the consumption of a nation is used for the education of the future generation, for promotion and nourishment of the future national productive powers (List, 1909, 112-113).

and

> The prosperity of a nation is not, as Say believes, greater in the proportion in which it has amassed more wealth (i.e. values of exchange), but in the proportion in which it has more developed its powers of production. Although laws and public institutions do not produce immediate values, they nevertheless produce productive powers, and Say is mistaken if he maintains that nations have been enabled to become wealthy under all forms of government, and that by weans of laws no wealth can be created. . . The nation must sacrifice and give up a measure of material property in order to gain culture, skill, and powers of united production; it must sacrifice some present advantages in order to insure to itself future ones (List, 1909, 117).

F. List's liberalism is different from the classical one and also from the neoclassical one. The latter is based on the idea of externalities and the needs for an improvement of the market mechanism in good faith, it being a universal and only possible way of achieving the wealth of nations; this wealth was understood in the same way as Smithian's wealth. The neoclassical economists under the pressure of market failures and the critiques of other schools judged a policy necessarily to include social issues in an economic mechanism, but they have been remaining at the individualistic outlook presented by the classical schools. Thus, the policy was used to achieve A. Smith's abstractive model. As a consequence, the economic policy was not focused on the social issues but on the laissez faire market that should ensure the former. F. List's methodology severs with both of these approaches (of course, the neoclassical one was not known to him). An economic policy was a part of the social policy and economic liberalism was introduced to it only when it was advantageous for a society (a nation); and even then it was under the control of a social policy. This is the idea of the national economy, that could be identified with the cosmopolitan economy (the classical one); economic liberalism as a strategy of social development had to have created the institutional conditions for it to exist. In addition, the classical mechanism seemed to mostly present, in F. List's opinion, the interests of the current generation and the social perspective presented in his work was also focused on the interests of the future generations and a nation has to ensure it.

A Way between Socialism and Liberalism e. i. Social Liberalism

M. Lind recalled F. List's theory as a remedy for the deepening global economic crisis:

> If democratic socialists are still too close to Karl Marx, neoliberals such as Bill Clinton and Tony Blair have made too many concessions to Adam Smith. . . It is important to recall that socialism and classical liberalism are only two of the three rival traditions of political economy that emerged in the half of the 19th century. The three traditions are symbolised by Karl Marx, Adam Smith and Friedrich List. . . This long-neglected figure had more influence on the development of industrial civilisation in Europe and East Asia than either Marx or Smith (Lind, 1998, 25).

He saw the Listian capitalism as a way between two radical proposals, which unites private property, and corporate capitalism with social issues using the theory of historical progress (neglecting, however, the revolutionary character of it); social capitalism seems to exactly characterize the European Union capitalism. Moreover, since socialism has been widely neglected, because of the political and economic crisis in Eastern Europe, the 20th century debate between Marxists and liberals (Smithians) has been shifted in the 21st century onto a debate between two groups of liberals i. e. Smithians and Listians:

> between those who want laissez-fair global capitalism now, and those who envision a social-market version of global capitalism as the remote and negotiated results of generations of differing national and regional paths to development; Smithians believe that global market integration will produce national and regional development; Listians believe that global market integration should follow national and regional development. Smithians want a global market by the year 2000 or 2020 at the latest; Listians are willing to postpone it until 2200 or perhaps 3000 (Lind, 1998, 26).

Although, the European Union mirrors the debate between a more liberal and a more social economic policy, which at last represents the variety of systems among particular nations (and also in the course of time among them), the development of the European Union seems to be much closer to the Listian outlook in the context of both issues: the methodological premises of a political economy (i. e. the political character of economic liberalism based on the holistic outlook and with an historical perspective) and the character of future changes undertaken in Europe in the course of time.

The process of the European Union's creation was primarily focused on the liberalisation of some European economies; however, it proceeds according to F. List's premises. Moreover, it has been mirroring and revealing more social and national outlooks in the course of time which have remained under the political control of decision-makers. The national character of it was foreseen in F. List's works, which explains so many obstacles and constant negotiations between members' Representatives (from the classical point of view, there is no need to worry about national interests when free trade is ensured). As a matter of fact, the policy on infant economies is also widely presented in the process; moreover, the liberalisation of the USA economy was rendered according to the F. List postulates.

F. List's precondition of trade liberalisation, as it was postulated in the classical schools, was continuous peace "which enabled her [England] to dispense with standing armies, while facilitating the early introduction of a consistent customs system" (List, 1909, 43). War was seen as a process of economy and territory devastation which contributed to the weakening of the economic position of a country involved in wars (as in Germany) and simultaneously to the strengthening of the industry of a country not involved in wars[8]. Thus, peaceful relationships are the very preconditions of economic activity and it implicates that economic liberalisation needs to be preceded by effective political agreements. This seems too familiar when the history of the European Union is investigated. World War II contributed to the serious damage of European economies. The first economic steps toward unification were followed by strong political pressure to organise a peaceful state of further development. It has been proceeding to the next stage of social development precisely described by F. List:

> Territorial deficiencies of the nation can be remedied either by means of hereditary succession, as in the case of England and Scotland; or by purchase, as in the case of Florida and Louisiana; or by conquests, as in the case of Great Britain and Ireland. In modern times a fourth means has been adopted, which leads to this object in a manner much more in accordance with justice and with the prosperity of nations than conquest, and which is not so dependent on accidents as hereditary succession, namely, the union of the interests of various States by means of free conventions (List, 1909, 143).

However, the next precondition of the "territory enlargement" and the social development of nations is determined by the next stage of economic development; i. e. an industrial society. Political (i. e. peaceful) strategies are forced by well-developed industrial countries, because of their national interests. War in the context of a strong devastation of a country as well as economic protection is no longer accepted, both no longer being advantageous to the European nations. Thus, European countries will further develop more peaceful strategies for the development national interests. F. List saw the industrial development as the end of conflicts and a precondition of the development of a cosmopolitan society:

> But the more that industry advances, and proportionately extends over the countries of the earth, the smaller will be the possibility of wars. Two nations equally well developed in industry could mutually inflict on one another more injury in one week than they would be able to make good in a whole generation. But hence it follows that the same new forces which have hitherto served particularly for production will not withhold their services from destruction, and will principally favour the side of defence, and especially the European Continental nations. . . The endeavours to settle differences by protocol are clearly already prevailing over those which obtain justice by force of arms (List, 1909, 101).

The latter is more obvious in reference to nuclear technology. The third industrial revolution attributed to the development of global technology (f. e. nuclear technology, IT) definitely seems to finish the warfare strategy of the development of national interests, the potential results of which are rather considered in terms of a loss to the nations all over world. However, the well-developed countries still have to lose much more. It will press them to

[8] However, it seems to important mention that war was also seen by F. List as a state of industrial development acted like a prohibitive tariff system: "If, however, one agricultural nation whose production and consumption are thus diminished by war has already made considerable advances in population, civilisation, and agriculture, manufactures and factories will spring up in it in consequence of the interruption of international commerce by war. War acts on it like a prohibitive tariff system. . . it is called to pass from the condition of a mere agricultural State to the condition of an agricultural-manufacturing State, and in consequence of this transition, to attain to the highest degree of prosperity, civilisation, and power. But if after such a nation has already made considerable progress in the manufacturing career which was opened to it by war, peace is again established, and should both nations then contemplate the resumption of their previously existing commercial intercourse, they will both find that during the war New interests have been formed, which would be destroyed by re-establishing the former commercial interchange" (List, 1909, 147).

choose peaceful strategies for further development and to control the development of the technology they created. Moreover, it can be the means to secure the interests of poorer countries in political negotiations. Therefore, the nuclear issues are widely presented on the political scene nowadays.

H. J. Chang undertook the Listian perspective once again to back his thesis up and to worry contemporary policy-makers (Chang, 2002). The main thesis was an affirmation of the British policy pictured by F. List:

> It is a very common clever device that when anyone has attained the summit of greatness, he kicks away the ladder by which he has climbed up, in order to deprive others of the means of climbing up after him. In this lies the secret of the cosmopolitical doctrine of Adam Smith, and of the cosmopolitical tendencies of his great contemporary William Pitt, and of all his successors in the British Government administrations.
> Any nation which by means of protective duties and restrictions on navigation has raised her manufacturing power and her navigation to such a degree of development that no other nation can sustain free competition with her, can do nothing wiser than to throw away these ladders of her greatness, to preach to other nations the benefits of free trade, and to declare in penitent tones that she has hitherto wandered in the paths of error, and has now for the first time succeeded in discovering the truth (List 1909, 295–96).

In his view the contemporary policy of well-developed countries as well as at the time of F. List the British policy recommended to less-developed countries[9] are the policies that were not used by the well-developed countries when they themselves were in the process of developing (Chang, 2002, 1); in other words, they policies were as nationalistic as we tend to see F. List's economics. Moreover, catch-up strategies the contemporary protagonists of the policies recommended for less-developed countries were almost the opposite of what they usually recommend: restrictive macroeconomic policy, liberalization of international trade and investment, privatization and deregulation (Chang, 2002, 1). Most of the countries as the Great Britain, the USA, Germany, France, or Sweden used the opposite cutch-up strategies despite the way of rendering it was strongly modified by the socio-economic and political circumstances. Moreover, the most liberal present economies were the leaders of protectionism in the time when they were developing countries. Consequently, the USA and the Great Britain are the best examples; and H. J. Chang reported the USA as "the mother country and the bastion of modern protectionism" (Chang, 2002, 61). Using the example vividly presented by F. List, H. J. Chang presented the key policy of Great Britain in the course of time; and before the sixteenth century as a relatively backward economy it imported the Continental technology, exported raw wool and low-value-added wool cloths. The policy of Henry VII (1485-1509) and Elizabeth I (1558-1603) used some measures to improve the woollen sector of economy as an increase export duties on the raw wool or on unfinished cloths; it has been decreased whenever the British economy had no capacity to process all its wool resources. Next, the meaningful measures were introduced to protect Britain against the British colonial economies as the Navigation Act (reinforcing policy that trade should be carried in English vessels) from 1651, the Wool Act from 1699 (that imposed ban on imported wool products from colonies strongly then affected the Irish or the Indian textile sectors), or the series of the Corn Laws passed from 1463. Finally, the crucial protection policy was introduced in 1721 by Robert Walpole (the first British Prime Minister between

[9] J. H. Chang called them "good policies" (Chang, 2002, 1).

1721 and 1742) who, generally speaking, aimed his policy at protecting home raw materials and facilitating export manufactures (as well as some export subsidies, quality regulations of manufactures products, or legal prevention the outflow of its superior technologies and skilled workers). In fact, "with the Industrial Evolution in the second half of the eighteen century, Britain started widening its technological lead over other countries. However, even it continued its policy of industrial promotion until the mid-nineteenth century, by which time its technological supremacy was overwhelming. . . overall liberalization of the British economy that occurred during the mid-nineteenth century, of which trade liberalization was just a part, was a highly controlled affair overseen by the state. . . Britain adopted Free Trade painfully slowly: eighty-four years from the Wealth of Nation to Gladstone's budget" (Chang, 2002, 22 and 24).

Further examples of national policy can be provided from the contemporary history of the European Union. The process of enlarging the European Union after World War II is usually analysed in the context of The Treaty of Paris (1951), that proceeded to an economic liberalisation of the coal and steel trade between France, West Germany, Italy, Belgium, Luxembourg and the Netherlands: the European Coal and Steel Community (ECSC). This Treaty had two main French supporters: the economist J. Monnet and foreign minister R. Schuman[10]. The former is the architect of the peaceful plan of development in Europe. Steel and coal played an important role in the European states and were also important resources in warfare production. The common program for the post-war production and consumption of them intended to control the war industries as well as the post-war industrial development of Europe; it was achieved by trade liberalisation with a short transition period (again according to F. List's postulates on infant economies). The French initiative was derived from the dependency of the French and German coal industry. Economic liberalisation was the tool of the peaceful development of France and Germany, it being based on the "politically neutral" laissez-fair mechanism of allocating the important resources for France and other countries[11]; and without the participation of Great Britain in the process of economic liberalisation this time, although the other participants invited it. And there is an interesting explanation of the British decision in the context of the F. List's nationalism: "more Western European nations would perhaps have responded to the Schuman Declaration, but found themselves precluded from the proposed enterprise for a variety of reasons: Britain because of its national sovereignty consciousness" (Roussakis, 1968, 119). In other words, "political neutrality" in trade was part of the plan for the socio-economic development of Europe, mostly derived from the national economic interests of France and the other participants as the British policy of trade liberalization widely promoted earlier. It seems to reflect the classical view i. e. liberalisation as the natural way of competing for resources in the cosmopolitan economy. However, it rather reflected the political national interests of the participants and the way to secure the conditions for their socio-economic development: "there would be a new order in western Europe, based on the interests its peoples and nations shared together, and it would

[10] W. Churchill in 1946 gave a speech at the University of Zürich calling for a United States of Europe, similar to the United States of America to avoid further conflicts between national European economies.

[11] The J. Monnet's plan was proposed earlier i. e. 1945 but it was taken into consideration in 1949 by R. Schuman when proceeded to problems between Germany and France with control of the Ruhr, the important coal and steel region. He called to take control of German coal and steel areas and redirect it away from German industry into French industry instead; it had to secure the resources for the French industry. The proposition from 1950 is named the Schuman Declaration and was inspired by the J. Monnet's plan; it opposed the Treaty of Versailles following World War I which had imposed severe penalties and reparations on a defeated Germany that were considered humiliating.

be founded upon treaties guaranteeing the rule of law and equality between all countries" (Fontaine, 1996, 3) and "the European Union has been built to achieve political goals, but its dynamism and success spring from its economic foundations – the 'single market' formed by all the EU member states, and the single currency (the Euro) used by 12 of them" (Fontaine, 1996, 5). Once again, not all of them have had an interest in the single currency; among them was Great Britain, the most liberal member of the European Union. Economic liberalism was imposed by the political goals of the further development of particular nations, because economic prosperity is an important part of any national policy. Moreover, contingent protectionism has been forming the policy of the European Union since the beginning, reflecting the political control over liberalisation and the interests of the nations that decided to join the Union. The crucial economic conditions of unification are based on the "industrial (economic) maturity" of the members and are controlled by a huge number of legal rules (acquis communautaire)[12]. Economic liberalisation was a political act of the peaceful development of the industrialised countries; it resulted from the national interests of individual countries.

Finally, in the cosmopolitan economy, the interests of individuals remain institutionalized. Although the cosmopolitan economy promised disperse national interests, individualistic interests do not secure the classical market laissez-fair mechanism, because the individuals with common interests create different groups of interests to secure themselves like nations did so before. And a policy has to control it according to the problems (e. i. market failures) that were common at the national level of the European economy. Primary families, towns, and counties created unions, and then nations. Nowadays, capitalists, workers, carpenters, car producers or petrol producers tend to secure their interests by the creation of different forms of cooperation that are focused on the representation of their interests. Liberalisation has been shifting control from the national level to the international level. However, it does not limit political control over the interests of different groups of people. They are joined by common businesses over the national differences ready to be taken under the control of the market. The neoclassical policy is focused on it in good faith of achieving the market equilibrium that finally reduces the policy to a minimum. On the contrary, the institutionalists postulate social perspective to understand market behaviour and the character of a policy made by state or international institutions. Society is again seen as a system of different values and attitudes that go beyond the individual's behaviour. The final question is rather whether the economic mechanism secures sustainability or social development should tend to create unique qualities that need social contribution (society as whole); the latter strategy deals with a policy as an instrument of planning social development, while the former is focused on solving the current problems. Both need an economic policy, but its character and origin are different.

F. List's theories are named national economics, while A. Smith's ideas a universal one, or, in F. List's words, cosmopolitan. However, 50 years later, the European Union is still based on the national interests of its members. Those same national interests have been

[12] However, the industrial maturity seems to be secondary to political and economic goals of well developed members and the policy is focused on the fast socio-economic improvement of poorer members (or candidates) according to some political and economic goals (for example, the gap between economy of Portugal, Poland, as well as Bulgaria and Romania at the time of joining the European Union was far away from the average of the rest of the members); it (as well as wider extension of the present areas of unification) contributes to longer and variety forms of contingent protectionism.

limiting the processes of unification; Great Britain is a good example of the latter when the monetary union is analysed. Moreover, because an economic liberalisation was created, a new group of interests not always compatible with the socio-ecological interests or even the economic interests of the laissez-fair market (named externalities). Adding to it, the French economic interests in German coal mines can change the peaceful J. Monroe plan into an economic strategy of the European countries that should be secured by peaceful agreements. Europe needed the common resource policy for further development at this stage of an industrial society. Wars were too expensive, but industry was investigating resources for further development; and national polices had common interests in the process of unification. This was the reason for the limited participation of all the western countries (i. e. Great Britain). Both countries knew that further pressure for resources would contribute to costly conflicts and national interests are more and more dependant on international cooperation; the unity between nations was in the interest of all the participants. In other words, the only way to secure further development of the national industry was based on peaceful cooperation:

> The EU countries account for an ever smaller percentage of the world's population. They must therefore continue pulling together if they are to ensure economic growth and be able to compete on the world stage with other major economies. No individual EU country is strong enough to go it alone in world trade. To achieve economies of scale and to find new customers, European businesses need to operate in a bigger market than just their home country. That is why the EU has worked so hard to open up the single European market – removing the old obstacles to trade and cutting away the red tape that entangles economic operators [Fountaine, 2004, 5].

It seems to present the stage of development foreseen by F. List quoted in the motto of this article in which national interests are better represented by the cosmopolitan economy: "If the nature of things has been powerful enough to extend this union (which commenced with the family) over hundreds of millions, we ought to consider that nature to be powerful enough to accomplish the union of all nations. If the human mind were capable of comprehending the advantages of this great union, so ought we to venture to deem it capable of understanding the still greater benefits which would result from a union of the whole human race" [List, 1909, 100-101]; the mottos of F. List's key works also portray these statements: "Et la patrie et l'humanité" ("And the fatherland and humanity").

The vision of F. List's "European Union" seems to correspond to the processes of liberalisation and unification of the European Union. He claimed that a regional agreement is possible as there are similar tradition, religion, or political institutions because the each stage of socio-economic development is defined by the degree of socio-economic structures, political institutions, and power. Thus, the national perspective can be explained by F. List's methodological outlook "his ardent support of the cause of unity was a manifestation of his awareness of the common bonds of language, culture, and history which the fragmented territories had" (Roussakis, 1968, 147) and the same arguments next were used to propose a union of six European countries as "a transitional stage from the condition of national economy to that of world economy. . . universal union" (Roussakis, 1968, 112). On the other hand, the English policy was not cosmopolitan; it was as national as the German one. And it is clear that A. Smith was influenced by the idea of social liberalism; however, there are problems with the clear classification of F. List's works because he was a liberal focus on equity, democracy, and freedom but in his economic theories he postulated contingent protectionism. These differences are furnished by different methodologies. A. Smith as a philosopher and an economist created an abstractive model of economy, while F. List as a politician, a journalist, and an economist created a policy of economic growth focused on solving the problems of his country (or the country he worked as the USA). English policy-makers used A. Smith's model to promote the industrial supremacy of Great Britain; and it reflected the philosophy of the English industrial society. F. List tried to solve the problems of his country; his view was a German politician' perspective and he was focused on the supremacy of his country similarly to the British one: "It is therefore good for England that she should practice resignation betimes, that she should be timely renunciation gain the friendship of European Continental powers, that she should accustom herself betimes to the idea of being only the first among equals" (List, 1909, 340).

Thus, on the one hand he represented national interests and in his works we can find a lot of postulates that reflect the German Spirit invoked by the German philosophers. For example, F. List claimed that the future Continental union of six countries (France, Belgium, Switzerland, Denmark, and Holland) should be led by Germany because of its geographical position, its federal constitution, its religious toleration, its cosmopolitical tendencies, and its civilisation and power (List, 1909, 332). It reflected the German philosophy that can be defined as nationalism – a supremacy of Germany, however, it cannot been understood as chauvinism in the context of the many other economic and political propositions but rather in the context the industrial supremacy of Great Britain that contributed to the socio-economic problems of the Continental countries and the USA; it reflects his postulates as a politician, a journalist, and a German citizen who wish to ensure German interests. To understand his economic theory we have to study his works as J. S. Mill explained A. Smith's views from the perspective of almost 100 years of free trade economy:

> The 'Wealth of Nations' is in many parts obsolete, and in all, imperfect. Political Economy, properly so called, has grown up almost from infancy since the time of Adam Smith: and the philosophy of society, from which practically that eminent thinker never separated his more peculiar theme, though still in a very early stage of its progress, has advanced many steps beyond the point at which he left" (Mill, 1952, v).

To understand his methodology we have to be aware of the situation widely presented in the previous chapters and than separated this political issues from the method of reasoning.

His economic logic is based on the idea of equality in the exchange; thus, the liberal conditions of the market are possible when all the entities included in the common market are enough mature to competition. Otherwise, it is monopoly of a one nation. As A. Smith saw it in the context of equality of all actors he did not include in his model these institutional differences; F. List investigated and postulated a way to achieve the equality. This economic outlook correlates with his liberal social view, as he wanted to create a free, democratic society; this is in the exact way A. Smith's economic liberalism corresponds with his social liberal views and his model of competition based on the similar power of all market agents. However, the equity of market power, in F. List's view, is achievable when the economic and institutional conditions are comparable.

Conclusion

The life of F. List[2] and his ideas seems to perfectly reflect the tragedy of romantic heroes; even so, his view of the political and economic processes in Europe described the present policy of the European Union and should be analysed more extensively to understand its present economic policy based on the economic liberalisation of different countries, nations and cultures in the global world. Thus, the process of the economic liberalisation of the European Union supports the methodological premises undertaken by historical schools. Moreover, Smithian economic liberalism was less present in the national policy of Great Britain at the beginning of the economic liberalization of the European economy; it proves the political character of the most liberal economy of Europe (it is also visible when the monetary union is investigated). F. List suggested the economic interests of Great Britain be obtained from the trade liberalization at his time. He saw Smithian economics as an interesting way of developing industrial societies at the last stage of doing so; thus, it was the part of the socio-economic plan of national development, not a universal way of economic behaviour.

There are some conclusions from the two ways of economy liberalizing as follows:

1. Whenever policy of liberalization is to be planned for countries that perform the different levels of institutional and industrial levels of development the liberalization is more profitable for the better developed ones; moreover, as history teaches, the liberalization jeopardizes the development of the less developed countries.
2. Whenever policy of liberalization is to be designed there are two processes to ponder: institutional and political changes. The latter can be easy changed, while the former needs much more time to adapt to the policy changes (Chang, 2002, 9). The classical perspective omitted the institutional changes in believe that liberal policy is the only arrangement that defined economic development. Moreover, the institutional changes are supposed to be the pre-condition of economic development, while it seems to be that they are rather a product of economic development (Chang, 2002, 11); for example, there is a demand to expect industrial maturity from people who did not experience industrialization.
3. The model of liberalization postulated by F. List is more realistic than the abstractive model of A. Smith because F. List's postulates are furnished by his practice as an economist, a politician, a journalist, and a businessman; in fact, the economy fulfilled premises of A. Smith's model has never existed. It is also mirrored in the divergent methods utilized by the F. List, i. e. induction, and A. Smith, i. e. deduction.
4. There are two perspectives of economic liberalism: historical and classical; the former represents a social approach to economic development (holistic): "Economic unity and political unity. . . are twins; one cannot be born without the other following" (Roussakis, 1968, 102)[13], while the latter is concerned with individualism.
5. Both the ways of liberalizing presented in their works reflected the specific social values of their nations; their utilization by policy-makers was a part of the national policies and was determined by institutional capacities.

[13] A quotation from List, F., 1927-1936. Schriften/Reden/Briefe. Friedrich List Gesellschaft (ed.). Berlin: Reimar Hobbing, pp. 276.

6. The liberalization policy of the European Union is better described by F. List's perspective; the liberalisation rather followed "national maturity" of the members to competition (i. e. institutional and economic maturity).
7. The discourse between both ways of liberalising can be described as discourse between two methodological perspectives furnished by the divergent philosophies that render two opposite perspectives:

   - holism vs. individualism (atomism),
   - abstractivism vs. pragmatism,
   - contextualism vs. universalism (i. e. a sequential stage of socio-economic development vs. universal economic law).

References

[1] Russell, B. (1945). *A History of Western Philosophy*. New York: Simon and Schuster.
[2] Blanning, T. C. W. (Ed.) (2000). *The Oxford History of Modern Europe*. New York: Oxford University Press.
[3] Blaug, M. (1992). *The Methodology of Economics*. New York: Cambridge University Press.
[4] Bury, J. B. (2004). *The Idea of Progress*. Whitefish: Kessinger Publishing Company.
[5] Chang, H. J. (2002). *Kicking Away the Ladder: Development Strategy in Historical Perspective*. London: Athem Press.
[6] Fontaine, P. (2004). *Europe in 12 Lessons*. Luxembourg: Office for Official Publications of the European Communities.
[7] Fusfeld, D. R. (1977). The Development of Economic Institutions. *Journal of Economic Issues*, **4**: 743-784.
[8] Fyffe, C. A. (1896). *A History of Modern Europe*, 1792-1878. New York: H. Holt and Company.
[9] Hayes, C. J. H. (2004). *A Political and Social History of Modern Europe*. Whitefish: Kessinger Publishing Company.
[10] Hegel, G. W. F. (1998) [1806]. *The Phenomenology of Spirit*. Delhi: Motilal Banarsidass Publication.
[11] Hegel, G. W. F. (2001) [1820]. *The Philosophy of Right*. Kitchener: Batoche Books Ltd.
[12] Henderson, W. O. (1983). Friedrich List: *Economist* and *Visionary,* 1789-1846. Frank Cass and Co. Ltd.
[13] Iannone, P. A. (2001). *Dictionary of World Philosophy*. Florence, KY, USA: Routledge.
[14] Lind, M. (1998). *The time is ripe for the Third Man. Forget Marx and Smith.* Friedrich List is the economist for us. New Statesman: 13 November (reprinted from the 5 October issue of "The Nation", New York).
[15] List, F. (1909) [1841*]. The National System of Political Economy*. Translated Sampson S. LLoyd. Longmans, Green and Co. London Edition.
[16] Lock, J. (1824) [1680*]. Two Treatises of Government*. London: C. and J. Rivington.
[17] Mill, J. S. (1974) [1843]. *A System of Logic, Ratiocinative and Inductive*: Vol. VII-VIII. Toronto: University of Toronto Press, Routledge and Kegan Paul.

[18] Mill, J. S. (1852). *Principles of Political Economics*: Vol. I. London: John W. Parker and Son.
[19] Norgaard, R. B. (1994). *Development Betrayed*. Routledge, London-New York.
[20] Preparata, G. G. & Elliott, J. E. (1996). Protecting the infant industry. Cosmopolitan versus nationalist economists. *International Journal of Social Economics*, **2:** 4-34.
[21] Rawls, J. (1999). *A Theory of Justice*. Oxford University Press, New York.
[22] Reinert, E. S. (1994). Competitiveness and its Predecessors - a 500-year Cross-national Perspective. *STEP Raport Series*: **R-03**, Oslo: Stiftelsen - STEP.
[23] Rauch, L. (1993), *Hegel, Spirit, and Politics*. In R. C. Solomon & K. M. Higgins (Eds.), The Age of German Idealism (pp. 254-289). London and New York: Routledge.
[24] Riha, T. (1985). Romanticism and Nationalism in Economics. *International Journal of Social Economics*, **12** (3-5): 52-68.
[25] Robbins, L. (1998). *History of Economic Thought: The LSE Lectures.* Ewing, NJ, USA: Princeton University Press.
[26] Roussakis, E. N. (1968). Friedrich List, the Zollverein, and the Uniting of Europe. Bruges: College of Europe.
[27] Schmoller, G. (1897) [1884]. *The Mercantile System and its Historical Significance*. London: Frank Gass & Go. Ltd.
[28] Smith, A. (1838) [1776]. *An Inquiry into the Nature and Causes of the Wealth of Nations*. Edinburgh: Thomas Nelson.
[29] Solomon, R. C. (1993). Hegel's Phenomenology of Spirit. In R. C. Solomon & K. M. Higgins (Eds.), *The Age of German Idealism* (pp. 181-215). London and New York: Routledge.
[30] Stankiewicz, W. (2000). *Historia Myśli Ekonomicznej*. Warszawa: PWE.
[31] Taylor, E. (1957). *Historia Rozwoju Ekonomiki*: Vol. I. Poznań: PWN.

In: Emerging Topics in Macroeconomics
Editor: Richard O. Bailly, pp. 125-154
ISBN: 978-1-60692-093-0

Chapter 4

THE MACROECONOMIC DYNAMICS OF A LIMITED-COMMITMENT ECONOMY: HOW TO USE SELF-ENFORCING CONTRACTS TO DEAL WITH THE DYNAMICS PUZZLE

Christian Calmès
Département des sciences administratives, Université du Québec (Outaouais)
Laboratory for Research in Statistics and Probability
Raymond Théoret
Département de finance, Université du Québec (Montréal)
Corporate Reporting Chair

Abstract

After reviewing the basic self-enforcing labour contracts models, this chapter explores some research avenues that can help understand the most important dynamic properties of macroeconomic variables. The idea of adapting the microeconomic theory of Thomas and Worrall [1988] to address the issue of macroeconomic dynamics was first proposed by Calmès [1999, 2003]. We thus describe how future research applying the theory of self-enforcing labour contracts will improve the way macroeconomic models can account for the response of the economy to external shocks. First, the introduction, by Calmès [2007], of a state-dependent opportunity for the firm in a model of self-enforcing labour contract represents in itself a novelty which generalizes the properties of this model class [Thomas and Worrall 2007a]. Second, this contribution helps getting rid of the classical result that competition among firms drives profits to zero. Going further in this direction by making capital variable, although desirable, leads to an impasse: the contract set will be no longer compact in this case, and this creates a major difficulty to solve the model. We discuss how this issue can be handled with a lottery incorporated in the model to convexify the contract set. For that matter, we also discuss the introduction of a third agent and a second relationship. Third, the question of economic growth is examined. More precisely we analyse the link between stationarity and set convexity when introducing growth in the model. A stochastic trend may be considered but the difficulty here will come again from the necessity of having a compact set within this class of models. Fourth, the aggregation of heterogeneous individual contracts is another important topic to look at. Fifth, we will deal with the problems related to the contracts duration. In other respects, according to Thomas and Worrall [1988, 2007a], risk-sharing is an important motive for contractual solutions. It is therefore relevant to study how the dynamic properties of the model change with the mathematical formulation of the individual utility functions. Utility functions could be assumed nonseparable to better account for the comovements between consumption and hours worked [Calmès 2007].

Keywords: Internal propagation mechanisms; Real business cycle; Self-enforcing contract; Risk-sharing hypothesis.

JEL: E12, E49, J30, J31, J41.

Introduction

The dynamic contract theory has evolved a long way since the introduction of implicit contracts in Bailey [1974], Gordon [1974] and Azariadis [1975]. Originally, these studies were aiming at explaining the stickiness of real wage, the flexible wage model being unable to account for this phenomenon. Implicit labour contract models, by endogenizing the wage rigidity, provide a more satisfying explanation of the relationship between key labour variables like average productivity, hours worked and real wage. In these models, the fisherian equality between the wage and the marginal product of labour no longer holds and therefore, the microeconomic dynamics of these models is very different from the one associated to the flexible wage models. However, the implicit labour contract models were not well adapted to the study of the macroeconomic dynamics.

Self-enforcing labour contract models may be viewed as a second generation of implicit labour contract models. Contrary to the implicit labour contract models, in a self-enforcing contract the principal and the agent cannot commit. Risk-sharing is the principal motive for such contracts. To be enforceable, the contracts must give to both parties more than what outside opportunities offer.

In a partial equilibrium, microeconomic setting, Thomas and Worrall [1988] were the first to suggest that enforceability considerations could be used to describe the dynamics of labour relationships. Boldrin and Hovarth [1995] then use this theory to study the dynamics of employment, consumption, productivity, hours worked and real wages. However, we must be clear on the definition of dynamics here. This concept has two distinct dimensions. The first dimension is what is generally meant by this concept in applying the theory of self-enforcing contracts to a macroeconomic set up. This dimension is associated to a very restricted form of dynamics, i.e. contemporaneous (instantaneous) dynamics. The article of Boldrin and Hovarth [1995] is a well-known study which is only concerned about contemporaneous macroeconomic dynamics. Incidentally, most models featuring dynamic contracts only consider one-period contracts [Boldrin and Hovarth 1995]. The duration of this kind of "dynamics" is thus only one period.

When adopting this dimension, the researcher is only concerned about replicating very basic stylized facts related to the behaviour of macroeconomic variables, like the contemporaneous correlation between the macroeconomic variables and the autocorrelation functions of these variables. This kind of dynamics is not very far from statics. And even at this rudimentary level of dynamics, a modified RBC model is often unable to replicate important stylised facts.

In other respects, the RBC models are even more challenged. For instance, they generally predict a very high negative correlation between hours worked and average productivity of labour. This correlation is at odds with the empirical evidence: the observed correlation is strongly positive. The negative correlation derived from RBC models is due to the pervasive impact of the income effect which strongly dampens the substitution effect [Calmès 2007]. At the other extreme, models featuring risk-sharing, the principal motive for a self-enforcing relationship, display a weaker income effect because the insurance provided by the contract tends to smooth consumption intertemporally *and* intratemporally so the link between hours worked and average productivity of labour is positive as observed in the data [Calmès 2007]. In this chapter, we will discuss in what respect self-enforcing labour contract models can help explaining central stylised facts related to macroeconomic dynamics.

There is another dimension in which dynamics is quite fundamental and yet often neglected by researchers applying dynamic contract theory. We refer here to the intertemporal dimension of dynamics. For instance, in macroeconomics, the internal propagation mechanisms incorporated in the models must help replicate the actual impulse-response functions of the macroeconomic aggregates, i.e. the propagation of shocks described by these models should tend to match the one observed with the econometric analysis of macroeconomic time series at hand. While progress has been great in this area, especially with models featuring nominal and real rigidities, the RBC models are still quite weak on this respect too. For instance, models often fail to account for the observed permanent and transitory characteristics of the impulse-response function of output [Cogley and Nason 1995]. Furthermore, the amplitude of the cycles generated by RBC models is generally too low and the artificial cycles die too quickly. The simulated volatility of macroeconomic aggregates is also too low in comparison with the facts. These contradictions between the behaviour of macroeconomic time series and their values as simulated with RBC models are quite a challenging puzzle in macroeconomic dynamics theory.

It is in this precise context that the self-enforcing labour contract models reveal promising in helping account for the observed propagation mechanism of shocks and provide an explanation for this dynamics puzzle. Calmès [1999, 2003] was the first to transpose the methodology of Thomas and Worrall [1988] in a macroeconomic setting to study the intertemporal response of an economy to aggregate shocks. Subsequently, as will be presented in this chapter, Calmès [2007] shows how to obtain a rich (intertemporal) dynamics, by its nature quite closer to reality. Calmès [2007] also conjectures that the lack of dynamics generally encountered in RBC models partly results from the embedded weakness of the dampening income effect[1].

The organisation of this chapter is as follows. First, we will present some stylized facts which have been challenging to understand in the field of macroeconomics. Second, we will examine the general structure of a self-enforcing labour contract model. Then, we will study the respective contributions of Thomas and Worrall [1988, 2007a] and Calmès [2007]. We will show how these models help to explain the two dimensions of macroeconomic dynamics, the study of the second dimension in a limited-commitment model being first proposed in

[1] i.e. the strength of the income effect dampens economic fluctuations in standard macroeconomic models like the RBC one.

Calmès [1999, 2003 and 2007]. Finally, we will consider some research avenues in which the limited-commitment framework can improve our understanding of macroeconomic dynamics. The following subjects will be considered: the aggregation of individual, heterogeneous contracts; economic growth and dynamic contracting; the contracts duration; the introduction of variable capital in a limited-commitment model; the endogenization of the external opportunities of the firm and of the worker; the transposition of option theory to self-enforcing labour contract models and the use of nonseparable utility functions to better account for the comovements between consumption and hours worked [Calmès 2007].

Some Important Stylized Facts Related to Macroeconomic Dynamics Challenging Standard Models

We will first examine the stylized facts related to the labour market variables, i.e. real wage, hours worked and the average productivity of labour which, without loss of generality, we equate to the marginal product of labour. We will thus only refer to labour productivity in this section. To get an intuitive understanding of the relations, table 1 illustrates the correlations between these variables in three situations: the observed, actual US data, the ones predicted by a model featuring a full-commitment contract, (i.e. a classical implicit contract in which the principal and the agent commit), and lastly, the correlations derived from the standard flexible wage model, to which we will often refer to as the spot market model.

Table 1. Correlation between labour market variables in three situations

	U.S. data	Contract with full- commit.	Flexible wage model(spot market)
Hours - prod.	strongly positive	1	-1
Hours-wage	mildly negative	-1	-1
wage-prod.	weakly positive	-1	1

Let us consider first the flexible wage model, which in what follows is associated to the spot market. According to table 1, the correlation between real wage and productivity is equal to 1 in this case, since $w = mpl$ in this model, where w is the wage and mpl the marginal product of labour. This correlation is obviously too high in comparison to the observed one which is generally considered to be weakly positive. Furthermore, in this case, the correlation between hours worked and wages is equal to -1. In this simple flexible wage model, the income effect is very strong and neutralizes completely the substitution effect right from the start, i.e. at the shock impact time[2]. This correlation is obviously too low[3] in comparison to

[2] We here assume separable preferences. If this is not the case, the results are not so clear-cut, as discussed later.
[3] In relative value.

the observed correlation. And since $w = mpl$, the correlation between productivity and hours is also -1. This theoretical correlation is also at odds with the actual one.

A contract with full-commitment has no self-enforcing constraints. Full insurance is provided by the firm (i.e. risk sharing is perfect). It is in fact a conventional implicit contract where the agents can commit. In this environment, the income effect is considerably dampened as earnings are fixed across states and time, and consumption completely smoothed. Since earnings are fixed, the correlation between hours and wage is thus -1. The absence of an income effect gives rise to a nearly perfect correlation between hours worked and productivity. However, as earnings do not vary with productivity while hours are nearly perfectly correlated with productivity, there must be a nearly perfect negative correlation between wage and productivity. Compared to the facts, the sign of the correlation between wage and productivity is a non-sense while the other two correlations, although of the good sign, are too high in absolute value. The limited-commitment model being a mix of the two extreme cases reported in table 1, the correlations between the labour market variables delivered by this type of model should be more in line with the facts.

There are other fundamental dynamic aspects of the behaviour of macroeconomic aggregates which must be replicated by the theoretical models. First, the fluctuations generated by most models are often very different from those observed in the data. For instance, consumption and hours are not volatile enough in standard macroeconomic models like the RBC one. Models have also a tendency to understate the volatility of employment and to overstate the volatility of the real wage. Furthermore, there is the question of persistence in the movements of the aggregates. For instance, wage lacks persistence in standard models. Also, there is the related issue of the amplitude and of the propagation of shocks in the models. For example, in flexible wage models, the impact of the substitution effect dies fast because it is neutralized by the income effect, with the consequence that the fluctuations of some aggregates are smaller and short lived. The resulting amplitude of these fluctuations is therefore too low in comparison to the facts. Moreover, there are comovements between macroeconomic aggregates that a model must replicate. In reality, the intertemporal comovement between consumption and hours worked is long lived but standard models do not capture this phenomenon. These discrepancies pose a puzzle we discuss in the following section, in the framework of a self-enforcing labour contract economy, illustrating how its properties, by nature, seem quite relevant to study the issues at hand.

The General Structure of Self-enforcing Labour Contract Models

We first present a very general formulation of relational contract models under limited liability which is an adaptation of Thomas and Worrall [2007b]. Then we transpose this model in continuous time in a simple framework to further assess the properties of such models when they are used to study self-enforcing labour contracts.

General Formulation of Dynamic Relational Contracts under Limited Liability

Let us assume two infinitely lived risk-averse agents who undertake actions[4] to produce a joint output O, of function o such that

$$O = o(a_1, a_2, s)$$

where a_i, i = 1,2, measures the cost-equivalent of the actions and s represents the state of nature, i.e. output is state-dependent. Suppose that the implicit contract linking the two agents cannot be enforced. Let us consider that, if there is disagreement, upon reneging the parties will receive a termination payoff amounting to

$$P_i = p_i(a_1, a_2, s), \ \forall i = 1,2$$

This payoff is the outside opportunity of the agent. It may be linear or nonlinear[5] in its arguments. We want to derive the constrained Pareto efficient self-enforcing contract in the case of infinite horizon.

Let us note the date t state s_t and the history of states up to this date: $s^t = (s_0, s_1, ..., s_t)$. At s^t, which we can call an event, both agents may undertake actions to produce the output. But when s^t is known, the agents can also renege, for instance, when they do not arrive at an agreement on how to share the surplus resulting from their collective action.

The consumption of agent i in event s^t is denoted by $c_i(s^t)$. The agents consume all their output, so

$$\sum_{i=1}^{2} c_i(s^t) = o(a_1, a_2, s_t), \forall s^t$$

The per period or Bernouilli utility of the agents is

$$U_i(s^t) = u_i(c_i(s^t) - a_i(s^t)), \ \forall s^t$$

The agents maximize their discounted expected utility

$$E_0\left[\sum_{t=0}^{\infty} \delta^t U_i(s^t) \mid s_0\right]$$

with δ being the common time discount factor of the agents.

The following Bellman equation characterizes the equilibrium path for agent I

$$V_i(s^t) = U_i(s^t) + E_t\left[\sum_{\tau=t+1}^{\infty} \delta^{\tau-t} U_i(s^\tau) \mid s^t\right], \quad \forall i = 1,2, \ \forall t$$

[4] For example agents providing an effort, with disutility related to their actions.
[5] Like the payoff of an option.

To be self-enforcing, the paths must satisfy the following equation:

$$V_i\left(s^t\right) \geq p_i\left(a_{ij}\left(s^t\right), s_t\right),\ \forall i = 1,2,\ \forall t$$

The equation represents the self-enforcing constraints (SEC)[6] of the program. They mean that the contract must give both parties at least more than what the outside opportunity provides.

It remains to characterize the Pareto-efficient self-enforcing agreements. There are several ways to proceed. The objective of this section is only to present the basics found in the literature related to relational contracts under limited liability, including models of self-enforcing labour contracts as well as many other kinds of contracts[7].

A. One-Sided Self-enforcing Labour Contract Model in Continuous Form

Cahuc and Zylberberg [2001] have developed a one-sided self-enforcing labour contract model in continuous form. It is interesting to summarize this model in order to understand the implications of risk-sharing inherent to this particular form of contracts.

The firm environment includes two agents, a manager and a worker. They are both risk-averse. The preferences of the worker are represented by an utility function given by *u(c,l)*, where *c* is the consumption level of the worker and *l*, his number of hours devoted to leisure. We have $u' > 0$ and $u'' < 0$, which implies risk aversion.

The firm has a production function denoted by $y = f(h, \varepsilon)$, where *h* stands for the hours worked and *ε* is a random variable with density function *g(ε)* accounting for technology shocks. The total hours available to the worker are normalized to 1. The worker consumes all his wage, so $c = w$, where *c* is the consumption level of the worker and *w*, his wage. The profit of the firm, denoted by *π*, is thus:

$$\pi = f(h, \varepsilon) - w$$

Finally assume that the manager is also risk-averse, with an utility function denoted by *v(π)*, $v' > 0$ and $v'' < 0$.

The firm provides an insurance contract to the worker. It is assumed that the firm has access to capital markets to diversify its risk, an access limited to the firm. An insurance contract, denoted by $A = \{w(\varepsilon), h(\varepsilon)\}$, specifies, before the occurrence of a shock *(ε),* the wage *w(ε)* and the hours *h(ε)* provided for any *ε*. This contract is state contingent in the sense that it is contingent on the state that realizes.

The firm chooses a contract which maximizes the manager expected utility and provides the worker an expected gain at least as important as his external opportunity[8]. This external

[6] Also called "participation constraint".

[7] Thomas and Worrall [2007b] note that this theory may be transposed to household production, joint venture and repeated credit relationship, among others.

[8] Note that the firm has no self-enforcing constraint in a one-sided model, for example because it is a pure monopoly with full power and no constraint of any kind.

opportunity is represented by the expected utility level $\overline{U}$. The self-enforcing constraint in this model is thus[9]:

$$EU\left[w(\varepsilon),1-h(\varepsilon)\right]\geq\overline{U}$$

where the external opportunity of the worker is obviously exogenous.

The firm maximizes its expected utility subject to this self-enforcing constraint. The program solves:

$$\underset{A\{h(\varepsilon),W(\varepsilon)\}}{Max}\; Ev\left[\pi(\varepsilon)\right]$$

s.t. the SEC of the worker

$$EU\left[W(\varepsilon),1-h(\varepsilon)\right]\geq\overline{U}$$

The first-order conditions associated to this program may be written as follows. The first condition is related to the worker marginal rate of substitution between consumption and leisure:

$$\frac{U_L\left[w(\varepsilon),1-h(\varepsilon)\right]}{U_C\left[w(\varepsilon),1-h(\varepsilon)\right]}=f_h\left(h(\varepsilon),\varepsilon\right)\quad\forall\varepsilon\in\xi \tag{1}$$

where ξ is the interval of values for ε.

The second condition is linked to inter-state marginal rate of substitution of consumption and profit:

$$\frac{U_C\left[W(\varepsilon),1-h(\varepsilon)\right]}{U_C\left[W(\theta),1-h(\theta)\right]}=\frac{v'\left[\pi(\varepsilon)\right]}{v'\left[\pi(\theta)\right]}\quad\forall(\varepsilon,\theta)\in\xi^2 \tag{2}$$

where θ is a state different from state ε.

For a contract to be efficient, according to equation (1), the marginal rate of substitution between consumption and leisure must be equal to the marginal product of labour. Equation (2) is the well-known Arrow-Borch-Wilson condition associated to optimal risk- sharing. This equation states that risk-sharing is optimal when the marginal rate of substitution of consumption between states ε and θ is equal to the marginal rate of substitution of profit between the same states. In other words, risk sharing is optimal when the marginal rate of substitution of a gain between states ε and θ is the same for the principal and the agent.

In this environment, it can be verified that wage is independent of states ($\frac{dw}{d\varepsilon}=0$) if $U_{cl} = 0$, i.e. when the worker's preferences are assumed separable. The wage becomes perfectly

[9] Note that the binding budget constraint of the worker, i.e. $c = w$, is taken into account in the SEC since $EU(c,l) = EU(w, 1\text{-}h)$.

rigid in this case with no substitution effect nor income effect. And since $w = c$, consumption is completely smoothed across states.

The two models examined so far provide us the basis to understand the working of the models presented in the following sections.

Limited-Commitment Models and the Dynamics Puzzle

Thomas and Worrall Models and Macroeconomic Dynamics

Thomas and Worrall [1988] focus on the rigidity of real wages and its persistence. The most important contribution of the authors is to derive an updating rule for real wage which is in accordance with the properties observed in the data. Interestingly, their updating rule for real wage entails a nonlinear pricing rule which can be linked to option theory. Indeed, according to Rosen [1985, p. 5]: "The opportune substitution of work effort toward more productive states has a value similar to that of an option: that less work is called for in the less favourable states serves to truncate the lower tail of the θ[10] distribution". This connection between option theory and limited- commitment models opens the door to a complete reformulation of these models using options.

In the same vein, Thomas and Worrall [1988] compare the entrepreneur featured in their limited-commitment models to a risk-neutral principal providing insurance to a risk-averse agent facing an i.i.d. income stream. Suffice to view the spot wage as a random income and the contract wage as the net income derived from any transfer issued by the insurer. This comparison to insurance contracting is quite interesting since an insurance contract may be modelled as a put option. Note that in the self-enforcing labour contract models, it is the insurance motive which partially dissociates real wage from marginal productivity. In the real option theory, it is the real options embedded in an investment project which operate the dissociation between real wage and marginal productivity. Hence analogies can be drawn between these two theories. We will come back to this question when discussing the research avenues. More precisely, the outside opportunities of the agents can be modelled as options as will be explained in the last section of this chapter.

In Thomas and Worrall [1988], there is an infinite sequence of dates and a finite sequence of states $s_t \in \{1,2,...,S\}, S \geq 2$, two types of agents, entrepreneurs and workers, and contracts are self-enforcing in the sense that no firm's agent has an incentive to renege. Workers are infinitely lived, risk-averse and have an identical per period state independent utility function $u = u(\omega)$, where ω stands for the wage provided by the contract. $u(.)$ is twice differentiable and strictly concave. All managers are infinitely lived and risk-neutral. Both agents discount the future with a common factor $\alpha \in (0,1)$. In this model, each state is identified solely by the spot market wage $w(s)$, with probability of state s denoted by $p(s)$. Note that there is no production function in this model. Following models will introduce uncertainty directly in the production function.

It is assumed that the entrepreneurs have complete access to capital markets so they can diversify the idiosyncratic risks associated to the operations of the firm they manage. On the

[10] θ being a random variable accounting for technological uncertainty in the model.

other hand, workers consume all their earnings and have no access to capital markets. Being risk-averse, they prefer stable earnings to fluctuating ones.

We keep the same notation as before so the state that occurs at time t is s_t. The history of states up to time t is designated by $h_t = (s_1, s_2, \ldots s_t)$. A contract, δ, is a sequence of functions $(\omega(h_t))_{t=1}^{\infty}$, where $(\omega(h_t))$ is the wage paid if the history is h_t. Note the difference between the spot market wage w and the wage ω offered by the firm. The aim of the dynamic program is to find the set of contracts which is Pareto efficient. A contract is efficient if there exists no other self-enforcing contract offering both parties at least the same expected utility and strictly more to one of the firm's agent. The set of Pareto-efficient contracts maximizes the present value of the firm expected cash flows for a given level of the worker's present value of expected utility. Hence, the entire Pareto frontier can be traced by varying the level of the worker's present value of expected utility.

The dynamic program used to compute the Pareto efficient set of contracts includes two self-enforcing constraints (SEC), one for each representative agent[11]. To be self-enforcing for an agent, a contract must provide him as much payoff as he could get from his outside opportunity. Note that in Thomas and Worrall [1988], this outside opportunity is the spot labour market for both the worker and the entrepreneur, and this opportunity is exogenous. A worker reneging then works on the spot market forever. An entrepreneur can also renege, firing the employees and hiring on the spot market.

Contrary to the full-commitment case where the entrepreneur completely insures the earnings of the worker, in a limited-commitment environment risk-sharing is only partial.

For each contract δ and any history h_t, the net gains accrued to the worker are given by the Bellman equation

$$U(\delta; h_t) = u(\omega(h_t)) - u(w(s_t)) + E \sum_{\tau=t+1}^{\infty} \alpha^{\tau-t} \{u(\omega(h_\tau)) - u(w(s_\tau))\}$$

where SEC for the worker is defined as:

$$U(\delta; h_t) \geq 0$$

Note that if $u(\omega(h_t)) < u(w(s_t))$, then the worker has a short-term incentive to renege because he would be better off with the spot market allocation. This incentive must be compensated by the long-term benefits from compliance to the relationship. More specifically, we can resort to the Holmström backloading principle to prevent reneging in this case. To understand this, we can illustrate the situation in the following Holmström [1983] set up: let us assume for a moment a two period contract. In such a contract, workers pay their insurance premium in advance so that they receive initially a wage below their productivity to be compensated later with wages above productivity. The point is that workers are not attracted by a firm that pays at the marginal product in the first period, as would be the case on the spot market. To avoid this, worker reliability may be rewarded by seniority rules. In

[11] One can argue that there is only one representative agent, the firm, composed of two types, an employee and a manager and described by their strategic interactions.

this case, worker's compensation is thus backloaded, implying a repeated relationship. To summarize, wages are lower in the first period because the firm must offer the worker at least the spot market in the second period.

The net future benefit that the entrepreneur gets from the contract at time t after history h_t is given by the Bellman equation:

$$\Pi(\delta; h_t) = w(s_t) - \omega(h_t) + E \sum_{\tau=t+1}^{\infty} \alpha^{\tau-t} \{w(s_\tau) - \omega(h_\tau)\}$$

where SEC for the entrepreneur is

$$\Pi(\delta; h_t) \geq 0$$

If $w_t(s_t) < \omega(h_t)$, there is an incentive for the entrepreneur to hire employees on the spot market, so, to avoid reneging, the expected wages he offers has to be lower than the expected wages prevailing on the spot market. To summarize, the entrepreneur provides earnings stability to the worker in exchange of lower expected wages. This is the insurance device embedded in self-enforcing models.

At this stage, it is important to note that, if agents could renege and recontract next period without penalty, only trivial self-enforcing contracts would exist. Frictions must thus be introduced in self-enforcing labour contract models in order to avoid this case. For instance, Thomas and Worrall [1988] propose that any reneging agent is observed by everyone else. Reneging must thus harm reputation. Further, it is assumed that once an agent has reneged, he must trade at the spot market wage from then onwards. These strong assumptions are a priori necessary to ensure that, in equilibrium, all contracts are non trivial self-enforcing ones.

To compute the Pareto efficient set of contracts, Thomas and Worrall [1988] define the Pareto frontier recursively, treating the expected utility level as an endogenous state variable[12]. With this method, finding the Pareto frontier amounts to solving recursively a dynamic programming problem. For an history (h_{t-1},s) and a feasible value of the worker's expected utility U_s^t, the Pareto frontier is given by:

$$f_s(U_s^t) = \sup_{\delta \in \Lambda(h_{t-1},s)} \left\{ \Pi(\delta; (h_{t-1}, s)) \,\middle|\, U(\delta; (h_{t-1}, s)) \geq U_s^t \right\}$$

Since an efficient contract cannot be Pareto dominated after any history, it follows that the corresponding dynamic program can be written as[13]

[12] Note that this type of variable could be interpreted as the second jump variable identified by Wen [2005] as necessary to properly account for the dynamic properties of the state space associated to the canonical RBC model.

[13] The binding budget constraint of the worker, i.e. that he consumes all his earnings, is already incorporated in his utility function.

$$f_s\left(U_s^t\right) = \sup_{\omega(s),\left(U_q^{t+1}\right)_{q=1}^{S}} w(s) - \omega(s) + \alpha E f_q\left(U_q^{t+1}\right)$$

s.t.

the two SEC:

$$U_q^{t+1} \geq 0; f_q\left(U_q^{t+1}\right) \geq 0,\ for\ q = 1,2,\ldots,S$$

and the feasibility constraint with the expected utility level of the worker such that

$$u(\omega(s)) - u(w(s)) + \alpha E U_q^{t+1} \geq U_s^t$$

It can be shown that this program is concave. The last constraint ensures that the utility level which the worker gets from the contract is at least equal to his expected utility level at time *t*.

Thomas and Worrall [2007a] introduces a simpler way of finding the Pareto efficient set of contracts, without resorting to a dynamic programming argument but using instead local variational arguments, the conventional method for the Pareto frontier computation. They explain that using local variational arguments instead of dynamic programming avoids the need to establish a number of technical properties of the value function, including the restrictions of twice differentiability. The entire Pareto frontier is thus traced by varying the level of the worker's expected discounted utility.

According to this alternative formulation, the program to retrieve the Pareto efficient frontier becomes

$$\sup_{\omega(h_t)_{t=1}^{T}} \Pi_1(\delta; h_1)$$

s.t.

the two SEC of the worker and the entrepreneur

$$U_1(\delta; h_1) \geq 0$$

$$\Pi_1(\delta; h_1) \geq 0$$

and the fixed utility level of the worker

$$U_1(\delta; h_1) \geq \overline{U_1}$$

The last constraint fixes the level of the worker utility. The term $\overline{U_1}$ measures how much utility the worker gets from the relationship, and as this level is varied across feasible values, all efficient contracts are traced out. The dynamic program gives rise to a simple updating rule for the real wage. For any history (h_{t-1}, s), the wage associated to an efficient contract,

$\omega(h_{t-1},s)$, is included in a closed non-empty interval ($\underline{\omega}_s, \overline{\omega}_s$). If it is equal to $\underline{\omega}_s$, the worker gets no gain from the contract, so $U_s^t = 0$. If, on the other hand, $\omega(h_{t-1}, s) = \overline{\omega}$, the manager gets no gain , so $f_s(\overline{U}_s) = 0$. Contract wages vary over time according to a simple pattern, keeping wages fixed if possible but changing them by the smallest possible amount otherwise. Note also that if the firm's SEC is slack, wages cannot fall. On the other hand, if the worker's SEC is slack, wages cannot raise.

Towards a Macroeconomic Formulation of a Limited-Commitment Economy

Discussing variable hours, Calmès [1999, 2003] describes a simple wage equation much in the spirit of the Thomas and Worrall original formulation, resorting on their methodology to account for the rigidity of real wages in the context of a stochastic marginal productivity of labour fluctuating on a small support. The endogeneization of this rigidity is a desirable property since most macroeconomic models display ad hoc wage rigidity [Rosen 1985]. This equation also shows that the insurance provided by the contract dissociates the behaviour of the real wage from that of the underlying marginal product of labour. It can be read as

$$w = mpl + TR$$

where *TR* stands for the transfer from the entrepreneur to the worker. These transfers may be allowances or penalties depending on the business conditions. They reduce the fluctuations of *w*. For instance, when business conditions are favourable, *mpl* is high and *TR* takes a negative value. The worker must then pay a penalty to the entrepreneur. In such case, $w < mpl$. Conversely, when business conditions deteriorate, *mpl* is low and *TR* is positive. In that scenario, the worker receives a positive transfer from the entrepreneur so $w > mpl$. This insurance device thus stabilizes *w*. Furthermore, *w* varies according to a step function when the increase in productivity is large enough. In this case, real wage behaves according to a nonlinear pricing rule.

In their 1988 paper, Thomas and Worrall focus on the implications of self-enforcing labour contract models on the one period ahead dynamics of wages. The main contribution of their article is to show that the fixed end points of the wage interval imply a rigidity of the response of current wages to past events. In other words, their self-enforcing labour contract model delivers wage persistence, a theoretical prediction later successfully tested by Beaudry and DiNardo [1991]. In their empirical work, Beaudry and DiNardo [1991] found that real wage follows a ratchet-like process, rising when the labour market is tighter but remaining constant otherwise[14]. Furthermore, the current wage is determined by the tightest labour market conditions incurred during the worker's tenure. Although the result of Beaudry and DiNardo tends to favour the relevance of the theory of self-enforcing contracts, Devereux and Hart [2007] argue that this result may be also compatible with a model of job-based wages and intermittent procyclical promotion, a variant of the spot market model obviously at odds with the analysis of Beaudry and DiNardo. However, the authors also acknowledge the fact

[14] This downward rigidity stems from a specific feature of their model. By construction, it obtains from the one-sided limited commitment assumption – the principal can fully commit to the relationship there.

the two views are observationally equivalent - i.e. that the self-enforcing labour market theory cannot be ruled out on that ground: "as such, the cyclical implications of the insurance contract model are observationally equivalent to the implications of the model with job-based wages and intermittent procyclical promotion."[Devereux and Hart 2007, p. 664].

As Calmès [1999], the 2007 paper of Thomas and Worrall focuses on macroeconomic dynamics. Here the model aims at rationalizing some stylized facts regarding macroeconomic dynamics: the weak correlation between real wage and productivity; the high correlation between hours worked and productivity and the negative correlation between hours worked and wages. They also resort to Frisch-type demand functions to study the time path of consumption and hours worked.

Compared to Thomas and Worrall [1988] and as in Calmès [2007], Thomas and Worrall [2007a] introduce variable hours and a stochastic production function. At the start of each period, the entrepreneur and the worker observe the current state s_t. Any agent can quit and take his outside option. Otherwise, agents trade at the agreed terms, in which case the output is realized and the entrepreneur offers the contractual wage payment. The value (discounted utility) of the outside option of the worker and the firm, respectively, is denoted by $\chi_w(s)$ and $\chi_f(s)$ in state s. Contrary to Thomas and Worrall [1988], here the duration of the contract, denoted by T, is finite and random[15]. At $t = T$, after observing the current state s_t, the partnership dissolves and both agents get their outside option. T being a random variable, or a stopping time, the length of the contract will in general depends on the history of shocks.

Let $u_t(h_t)$ denote the continuation utility which the risk-averse worker gets from the contract from t onwards (assuming no termination at time t). The Bellman equation reads

$$U_t(h_t) = u(w_t(h_t), H(h_t)) + E\left[\sum_{\tau=t+1}^{T-1} \beta^{\tau-t} u(w_\tau(h_\tau), H(h_t)) + \beta^{T-t}\chi_w(s_T) \,\middle|\, h_t\right]$$

where $H(.)$ denotes the hours worked, and β the discount factor common to both parties. The other variables are the same as previously defined.

The risk-neutral entrepreneur's continuation profit is given by the following Bellman equation

$$\begin{aligned}\Pi_t(h_t) &= z(s_t)H(h_t) - w_t(h_t)H(h_t)\\ &+E\left[\sum_{\tau=t+1}^{T-1} \beta^{\tau-t}(z(s_\tau)H(h_\tau) - w_\tau(h_\tau)H(h_\tau)) + \beta^{T-t}\chi_f(s_T) \,\middle|\, h_t\right]\end{aligned}$$

It is assumed again that the worker consumes all his earnings so that

$$c(h_t) = w(h_t)H(h_t), \quad \forall t$$

[15] Thomas and Worrall [1988] assumed an infinite horizon to account for the Holmström backloading principle. As a matter of fact, infinite horizon is not necessary nor sufficient to avoid a contract set reduced to the flexible wage trivial singleton [Calmès 2007]. Another way to account for this is to impose some restrictions on the risk-sharing domain. Calmès [2007] proposes imperfect contract competition for that matter. Here, Thomas and Worrall [2007a] formulate an equivalent hypothesis. They assume the existence of some reneging costs C_f and C_w (where f stands for the employer and w the employee).

with *c(.)* the consumption function. We can thus incorporate the consumption function into this Bellman equation and rewrite

$$\Pi_t(h_t) = z(s_t)H(h_t) - c(h_t) + E\left[\sum_{\tau=t+1}^{T-1} \beta^{\tau-t}(z(s_\tau)H(h_\tau) - c_\tau(h_\tau)) + \beta^{T-t}\chi_f(s_T)\middle| h_t\right]$$

The contract is self-enforcing if the following equations hold for all dates *t*, $T-1 \geq t \geq 1$, and for all h_t:

$$U_t(h_t) \geq \chi_w(s_t) - C_w$$

$$\Pi_t(h_t) \geq \chi_f(s_t) - C_f$$

with C_w and C_f being respectively the costs of reneging for the worker and the firm. They may be viewed as mobility costs for these agents. As previously mentioned, these costs indirectly and implicitly account for labour market imperfection, a property required for the self-enforcing contract labour market theory – for its absence would get us back to the flexible wage model view of macroeconomic dynamics [cf Calmès 2005]. These cost equations say that the contract must provide to the worker and the entrepreneur at least what they can get by quitting, net of quitting costs.

The presence of reneging costs is not innocuous. They are required both for the existence of a non trivial contract and for the sustainability of a limited-commitment partial equilibrium. Note that these costs were not required in Thomas and Worrall [1988], since they assumed an infinite contract duration instead. Without these reneging costs, the surplus related to the spot market would be always greater than the one related to the contract, so no non trivial contract would be possible. Actually, these costs represent the down payments made in the first period by the workers, the core of the Holmström backloading principle previously mentioned. They are the equivalent of the imperfect contract competition assumed in Calmès [2007].

The contracts specify the level of consumption c_t and the hours worked H_t. The resulting program, formulated with the variational location method reads as follows

$$\sup_{(c_t(h_t), H_t(h_t))_{t=1}^{T}} \Pi_1(h_1)$$

s.t.

the two SEC

$$U_1(h_1) \geq \chi_w(s_t) - C_w$$

$$\Pi_1(h_1) \geq \chi_f(s_t) - C_f$$

and the participation constraint of the worker

$$U_1(\delta; h_1) \geq \overline{U_1}$$

To find the time path of real wage and hours, Thomas and Worrall [2007a] first show that the condition for an efficient intertemporal allocation of hours, a condition met earlier in the framework of a continuous limited-commitment model (equation 1), holds in their model, so

$$-\frac{u_H(c_t(h_t), H_t(h_t))}{u_c(c_t(h_t), H_t(h_t))} = z(s_t) \qquad (3)$$

The marginal rate of substitution between hours and consumption must equal the marginal product of labour $z(s_t)$[16]. Let define λ the marginal utility of consumption

$$\lambda = u_c(c_t(h_t), H_t(h_t))$$

Equation (3) may thus be rewritten as

$$-u_H(c, H) = \lambda z$$

The solutions to these two last equations are the Frisch-type demand functions $c(\lambda,z)$ and $H(\lambda,z)$. These functions are used to determine the time path of hours and wage. Provided that leisure is a normal good, hours are an increasing function of λ and z. We can resort to the c function to establish that the contractual wage rate is decreasing in λ and z. Incidentally, these results are conditional on the choice of preferences and the degree of separability of the utility function used. For instance, it is easy to show that the wage rate is a decreasing function of λ and z if the assumed utility function is separable. Moreover, if utility is separable, consumption is independent of z for a fixed λ. The time path of hours and wages are thus negatively correlated in the model of Thomas and Worrall, a property consistent with Beaudry and DiNardo [1991, 1995] who found a significant negative relationship between hours and wages.

As noted previously, before Calmès [1999], macroeconomists who used self-enforcing labour contracts were not concerned by the correlation between the labour market variables and the degree of macroeconomic dynamics persistence, and dealt only with the contemporaneous dynamics of the macroeconomic aggregates, not with their intertemporal dynamics [e.g. Boldrin and Horvath 1995]. The first study which discusses the implications of self-enforceability of contracts on the transitory dynamics of labour market variables and other key macroeconomic variables is Calmès [1999] as described in the following section.

[16] See Beaudry and DiNardo [1991]. Note that in Thomas and Worrall [2007a], the production function has a very simple formulation, that is $f = z(s_t)H(s_t)$. The marginal product of labour is thus the derivative of f with respect to $H(s_t)$: $\frac{\partial f(s_t)}{\partial H(s_t)} = z(s_t)$.

Calmès Model [2003, 2007] and Macroeconomic Dynamics

Calmès [1999, 2003, 2007] aims at investigating whether the real wage stickiness implied by risk-sharing can help explain macroeconomic stylized facts that the standard models have difficulty to rationalize. More precisely, he derives the potential effects of risk-sharing on the dynamics of employment, aggregate consumption and aggregate production.

In Calmès [2007], there are two interrelated principles which drive the transitory dynamics of his self-enforcing labour contract model. First, introducing an endogenous real rigidity in the model gives rise to an amplification of the technological shocks. Wage stickiness is an artefact of the risk-sharing hypothesis. Second, he enunciates a conjecture, according to which, for a given self-enforcing labour relationship, economic dynamics partly results from the embedded weakness of the income effect associated to limited commitment. Indeed, following a positive shock, say a positive productivity shock, the substitution effect leads to an increase of aggregate employment and consumption but in standard models, the income effect dampens this dynamics. Calmès [2003, 2007] argues that is not the case with self-enforcing labour contract models because risk-sharing weakens the income effect, allowing for a more potent substitution effect fostering fluctuations. Consequently, we come full circle as the two principles relate to one another. Because real rigidities are endogenized in this self-enforcing labour contract set up, it generates a rich dynamics absent from RBC models. As a matter of fact, in a recent survey on RBC models, Rebelo [2005] noted that Wen [1998] model is one of the few RBC models attempting to match the properties of the spectral density of some important macroeconomic aggregates. To improve the performance of RBC models in terms of macroeconomic dynamics, Wen [1998] resorts to nonseparable preferences and to an employment externality giving way to an aggregate production function with increasing returns to scale. According to Wen [1998, p. 200], nonseparable preferences help propagate shocks while the employment externality helps amplify the propagation mechanism embedded in his model. However, Wen [1998] model too fails to match the empirical spectral density of consumption, a key macroeconomic variable, even if it partly succeeds in matching the spectral densities of output, investment and employment.

To make the propagation mechanism of his model more powerful, Calmès [2007] adds to it two important dimensions: the degree of imperfect contract competition and the initial bargaining power of the worker. We will see that these two dimensions impact greatly on the transitory dynamics of the model.

Calmès [2007] model features three building blocks:

i) The spot market, represented by the flexible wage model where the income effect plays in full;
ii) The full-commitment model, a standard implicit contract model where there is no participation constraint. In this environment, the income effect is completely eliminated with full insurance of earnings provided by the entrepreneur to the worker.
iii) The limited-commitment model itself, where the income effect, although present, is weakened. The macroeconomic dynamics he derives here is a mixture of the two previous ones.

The features of the limited-commitment model are the following. The representative worker is risk-averse while the representative entrepreneur is risk-neutral. Compared to Thomas and Worrall, Calmès introduces variable hours (H). The worker is endowed with a separable utility function taking the general form: $u(c(h_j), H(h_j))$, the notation being the same as earlier. As in Thomas and Worrall, the worker does not save[17] and consumes all his earnings so that $c(h_j) = w(h_j)H(h_j)$. His participation constraint is given by:

$$\forall \tau, \quad E_\tau \sum_{j=\tau}^{\infty} \beta^j u[c(h_j); H(h_j)] \geq E_\tau \sum_{j=\tau}^{\infty} \beta^j u[c^s(\varepsilon_j); H^s(\varepsilon_j)]$$

where $\beta \in (0,1)$ is the intertemporal discount factor assumed to be common to the worker and the entrepreneur, and where the superscripts s denote spot market variables, the outside opportunity being represented by the spot market.

The entrepreneur operates a production function taking the general form $F(\varepsilon_j, H(h_j), k)$. An important innovation in Calmès [2007] is the introduction of capital k in the production function of the firm. This is part of his effort to incorporate the Thomas and Worrall theory to a macroeconomic environment. However, at this stage, k is held constant, for tractability. In his model, uncertainty comes from an exogenous technological shock, denoted ε_t, stemming from production, a state variable that can take N different values each period.

The profit function of the firm, defined for a given history of states (h_j), is:

$$F(\varepsilon_j, k, H(h_j)) - w(h_j)H(h_j) - r(h_j)k, \ \forall j$$

where r is the marginal product of capital. $r(h_j)k$ is thus the opportunity cost of capital for a given h_j. We can resort to this profit function to write the participation constraint of the entrepreneur

$$\forall \tau, \quad \sum_{j=\tau}^{\infty} \beta^j \{F(z_j; k; n(h_j)) - w(h_j)n(h_j) - r(h_j)k\} \geq 0$$

In other words, the self-enforcing contract must offer the entrepreneur at least zero profit, the average outcome associated to the perfect competition occurring in the flexible wage environment, i.e. the spot market economy.

With the introduction of capital in the model, the external opportunity explicitly relates to business conditions. Adding capital in the production function makes the scenario where profit is zero an extreme (limit) case. In general, with the help of capital, the gains of the entrepreneur will be positive and will be conditional on the state of business conditions, especially the level of productivity, because the opportunity cost of capital is a function of

[17] According to Calmès [2007], even if there were savings in the model, hours could be quite volatile. A conjecture

interest rate, which depends on business conditions. This contribution of Calmès [2007] was underlined by Thomas and Worrall [2007a]. They note that, besides having proposed one model where hours worked were assumed variable – an essential generalization if we aim at adapting the theory of self-enforcing contracts to the study of macroeconomic fluctuations – Calmès [2007] has also introduced an external opportunity which really depends on business conditions.

Although a dispensable hypothesis, for simplicity, the duration of a contract is assumed infinite. A history dependent contract

$$\delta_{h_j} = \left\{ w\left(h_j\right); H\left(h_j\right) \right\}_{j=1}^{\infty}$$

specifies a wage and labour input pair for every realization of nature. In the limited-commitment model, this contract is self-enforcing if no party wants to renege on the relationship.

In addition to infinite duration, some frictions are introduced in the models to avoid the degenerate result of a quasi-empty contract set. Indeed, perfect contract competition would drive the allocations toward their first best levels, where no level of risk-sharing is sustainable, and only the flexible wage equilibrium could unfold.

For this matter, Calmès [2007] introduces an exogenous parameter P accounting for imperfect contract competition, $P = 1$ being the maximum value, that is the one associated to the pure contract competition that would only be contemplated in a perfectly flexible wage economy. In this case, the spot market allocation associated to perfect contract competition provides to the worker with the maximum feasible surplus. In other words, here, autarky can be seen as an unreachable first best classical world.

To each contract belonging to the contract set is also associated a unique level of initial bargaining power distribution. This bargaining power is related to the utility level expected by the worker, an endogenous state variable in the Calmès [2007] model. The worker maximum bargaining power is obtained when the entrepreneur's profit is reduced to zero. By varying this initial value the whole contract set obtains.

In Calmès [2007], the dynamic program computing the Pareto efficient contract set associated to the limited-commitment model may be formulated as

$$\Pi\left(\varepsilon_s, U\right) = \sup_{w,H,U_j} F\left(\varepsilon_s, k, H\right) - wH - r\left(\varepsilon_s\right)k + \beta \sum_{j=1}^{N} \Pi\left(\varepsilon_j, U_j\right) p_{sj}$$

s.t.

the self-enforcing constraint of the firm

$$\Pi\left(\varepsilon_j, U_j\right) \geq 0 \quad \forall j = 1, \ldots, N$$

the self-enforcing constraint of the worker

that still needs to be investigated.

$$U_j \geq V^s\left[\varepsilon_j\right] \quad \forall j = 1,...,N$$

the budget constraint of the worker[18]

$$c = wH$$

and the participation constraint related to the expected utility of the worker

$$u(wH, H) + \beta \sum_{j=1}^{N} U_j p_{sj} \geq U$$

In this program, p_{sj} represents the transition probability to be in state ε^j knowing that the previous state was ε^s, and U is the promised utility level chosen in the $\left\{ U_j \right\}_j$ menu given the realization ε^j. The self-enforcing constraint of the worker states that, each period, the entrepreneur must promise at least this level U from the contingent menu $\left\{U_j\right\}_{j=1}^{N}$. This is a major contribution of Thomas and Worrall since it deals directly with the usual problem of the history dependence of the equilibrium allocations[19].

The full-commitment model, i.e. the benchmark model[20], is the limited-commitment model without its two self-enforcing constraints. The dynamic program corresponding to this model is the following:

$$\Pi^*(\varepsilon_s, U) = \sup_{w,H,U} F(\varepsilon_s, k, H) - wH - r(\varepsilon_s)k + \beta \sum_{j=1}^{N} \Pi^*(\varepsilon_j, U_j) p_{sj}$$

s.t.

the binding constraint of the worker

$$c = wH$$

and the participation constraint related to the expected utility of the worker

$$u(wH, H) + \beta \sum_{j=1}^{N} U_j p_{sj} \geq U$$

We must now examine how the degree of contract competition and the bargaining power of the worker are introduced in the dynamic programs.

[18] It binds at equilibrium.
[19] On that matter, see: Ljungqvist and Sargent [2004], chap. 19.
[20] One of the most important contributions of Thomas and Worrall [1988] is to show that the limited commitment allocation can obtain with the first iteration guess set to the full commitment allocation.

Let us begin by the parameter $P \in (0,1)$, which represents the degree of contract competition. In the dynamic program associated to the limited-commitment model, this parameter is used to compute the expected utility level of the outside option V^s. This expected utility represents the external opportunity of the worker he would enjoy if he worked on the spot market. V^s is equal to:

$$V^s = PV^s_{\max}$$

For instance, if imperfect contract competition is associated with a 10% welfare loss, *P* is then equal to 0.9 and $V^s = 0.9V^s_{\max}$.

Importantly, $V^s_{\max}$ is the value function associated to the maximum feasible surplus under autarky. The corresponding spot market economy features a single self-employed agent working in a frictionless environment. For this reason, $V^s_{\max}$ is the first best value only feasible if contracts were perfectly substitutes and wages purely flexible. In other words, this representation of the outside economy is one of perfect competition. In our approach, we rule out the possibility that wages can be fully flexible by assuming $P \neq 100\%$. Because of this imperfect contract competition, there is room for risk-sharing.

$V^s_{\max}$ is equal to

$$V^s_{\max}(\varepsilon_s) = \sup_{\{H^s_{\max}\}} u\left(F\left(\varepsilon_s, k, H^s_{\max}\right) - r(\varepsilon_s)k; H^{\max}_s\right) + \beta \sum_{j=1}^{N} V^s_{\max}(\varepsilon_j) p_{sj} \tag{4}$$

Since, $U = u\left(c(h_j); H(h_j)\right)$, the first argument of the *u(.)* function in equation (4) is what is transferred by the employer to the worker. There is a third (implicit) agent in the Calmès model: the creditor who provides funds to finance the acquisition of capital[21]. Hence the presence of $r(\varepsilon_s)k$ in equation (4). The firm rents the capital to this financial intermediary.

While the degree of contract competition is exogenous in the Calmès [2007] model, the degree of bargaining power of the worker is not: as in Thomas and Worrall [1988], it is endogenous and recorded with the promised utility menu. In other words, the menu function *U* in the dynamic program determines the endogenous bargaining power of the worker. The worker's full bargaining power, U^{max}, is associated to a zero profit for the manager of the firm. It is obtained implicitly by solving the following equation

$$\Pi\left(\varepsilon_j, U^{max}(\varepsilon_j); U_j\right) = 0, \forall j = 1,\dots,N$$

Although the consecutive distribution of bargaining powers is an endogenous state variable of the model, once the relationship starts, Calmès [2007] does not make any

[21] A full representation of this agent is still to be investigated.

particular assumption regarding the initial bargaining power of the worker. However, one of his key contributions is to show that the initial distribution of bargaining power plays a crucial role in explaining the transitory macroeconomic dynamics.

Since Boldrin and Horvath [1995] were the first to consider the macroeconomic implications of self-enforcing contracts, to implement his model, Calmès [2007] calibrates the utility function of the worker and the production function of the entrepreneur using the same kind of functions as these authors[22]. The utility function of the worker is a time-separable Bernouilli utility function of the CES class given by

$$\forall t, \quad u(c_t;H_t)=\frac{1}{1-\gamma}c_t^{1-\gamma}+\frac{\theta}{1-\gamma}(T-H_t)^{1-\gamma}, \ \gamma\in(0,1)$$

where T represents the total of non-sleeping hours and γ captures the risk aversion of the worker. The production function of the entrepreneur is a standard Cobb-Douglas production function, multiplicative in the shock variable ε_t. It has the following expression:

$$\forall t, \quad F(\varepsilon_t;k;H_t)=\varepsilon_t k^{\alpha} H_t^{1-\alpha}$$

As already mentioned, the introduction of capital inside the production function represents a specific novelty of Calmès [2007]: in other well-known limited-commitment models, the production function depends only on hours worked and technological shock. This contribution is crucial as it implies a state dependent outside opportunity. Referring to the Calmès [2007] model, Thomas and Worrall [2007a] note that in much of the existing literature on limited-commitment labour contracts, it is assumed that competition among firms drives profits to zero. This is no longer the case here as firms must cover their capital costs. The marginal product of capital denoted by r is a state-dependent variable in the Calmès model. The outside opportunity of the firm is therefore state dependent and relates to the level of r, which yields profit fluctuations from one state to another.

Before discussing the dynamics of the Calmès [2007] limited-commitment model, it is useful to first compare the dynamics of the two limit cases: the benchmark economy, our flexible wage model, and the full-commitment economy.

The flexible wage model environment (spot market) and the full-commitment economy display very different dynamic properties. Because the income effect has a strong influence in the spot market economy, the model of autarky displays weak dynamics and persistence. As conjectured in Calmès [1999], the income effect prevents the substitution effect to fully play its role, hence dampening the fluctuations (a limitation common to most RBC models). On the spot market, following a technological shock ε_t, consumption and hours worked return to their steady state values very quickly at the end of the first period. Furthermore, the time path of these variables is almost flat. In this respect, the flexible wage model is definitively not a favourable set up to replicate the stylized facts related to macroeconomic dynamics. It severely lacks amplifications.

[22] However, as previously stated, Boldrin and Horvath [1995] consider one period contracts to investigate the contemporaneous dynamics of the macroeconomic aggregates. By contrast, Calmès [2007] is interested by the intertemporal macroeconomic dynamics implied by the Thomas and Worrall theory.

The dynamics of the full-commitment model, a building block of the general model, is substantially different from that of the flexible wage one. Here the risk-sharing is perfect, its domain unbounded, and there is a perfect smoothing of consumption across time and states. With non-separable preferences, a perfect smoothing of marginal utility of consumption will still obtain [Thomas and Worrall 1988]. Since he is perfectly insured, the worker does not have to bother about the income effect in such a case. Indeed, in this case, the income effect is totally absent, eliminated by the perfect risk-sharing. Hence, in the full-commitment model, the propagation mechanism stemming from the substitution effect is fully at work, without impediment, as conjectured in Calmès [2007]. As a matter of fact, when risk-sharing is perfect, initial shocks have an infinitely permanent effect on the economy, and the shocks amplification are much longer than in the classical environment where it is actually absent.

The limited commitment economy is an hybrid situation where the dynamics of the two building blocks formerly described intertwine. In the limited-commitment model, the income effect is not totally removed but it is diminished by the presence of the imperfect risk-sharing. The insurance provided by the entrepreneur smoothes partially consumption across time and states and thus weakens the income effect. In itself, the degree of consumption smoothness is conditional on the level of the initial distribution of bargaining power, on the dynamics of the bargaining power of the worker, and on the degree of imperfect contract competition. As noted previously, these are central contributions of Calmès [2007] and very important to understand how a limited-commitment economy can help explain macroeconomic dynamics. Under limited-commitment, consumption is bounded by the self-enforcing constraint of the worker. The upper bound of the admissible consumption interval corresponds to the consumption associated with the worker's full bargaining power (i.e. U_{MAX}). The lower bound is reached when the entrepreneur has full bargaining power. Following any shocks, even a purely transitory one and before arriving at its steady state level, the economy displays a phase of transitory dynamics, which crucially depends on the initial bargaining power distribution. For example, if the initial bargaining power of the worker is very low relative to that of the employer, the fluctuations of the economy will be larger in this transitory phase[23]. Then incoming shocks have a prolonged impact recorded in *U*. As bargaining power is set higher, with more balanced distribution of its initial value, these fluctuations are smaller because the perfect risk-sharing domain is reached more rapidly.

Another interesting finding of Calmès [2007] relates to the impact of the level of *P*, i.e. the degree of contract competition imperfection, on economic fluctuations. As *P* increases, the shocks have more likely an initial impact on the economy but the amplitude of this impact is smaller[24].

One important contribution of Calmès [2007] is the result that in this type of environment, aggregate dynamics crucially depends on the transitory dynamics observed at the beginning of the contract. This confirms Beaudry and DiNardo [1991]. Calmès [2007] finds that this transitory dynamics is conditioned by the initial bargaining power of the worker and by the degree of imperfection in contract competition. Although not formalized per se, this distribution of the initial bargaining powers between the worker and the

[23] The same is true if the employer has a high initial bargaining power. When the initial bargaining power is evenly distributed, Calmès [2007] explains that the dynamics is similar to the one obtained under full-commitment.

[24] It would be interesting to study to what extent the "great moderation phenomenon" could relate to the institutional innovations of the labour market.

entrepreneur might itself be dependent on the initial business conditions, e.g. the level of capital and its marginal productivity, two new dimensions introduced in this limited-commitment framework.

In summary, Calmès [2007] focuses on the intertemporal macroeconomic dynamics of a limited-commitment economy, a research avenue unexplored in previous studies dealing with self-enforcing contracts. Calmès [2007] findings are very encouraging since they suggest that it is possible, in theory, to generate a rich dynamics using self-enforcing contracts à la Thomas and Worrall, in other words limited-commitment models could help derive a self-enforcing labour market theory of macroeconomic dynamics.

Macroeconomic Dynamics and the Theory of Self-enforcing Labour Contracts: Some Research Avenues

In this section, we explore some avenues for research in the field of self-enforcing labour contract theory which are likely to lead to a better understanding of the dynamics of key macroeconomic variables such as consumption and hours.

The utility function of the worker

According to Thomas and Worrall [1988, 2007a], risk-sharing is an important motive for contractual solutions. Risk-sharing provides insurance to workers who can thus adopt more opportunistic behaviour vis-à-vis business conditions, so that the income effect, which generally dampens the response of the economy to exogenous technological shocks, comes into play later and with diminished strength leaving more room to the substitution effect.

In this respect, one interesting topic to investigate is the role played by preferences in this kind of models, and more precisely, to study how the dynamic properties of the model change with the mathematical formulation of the individual utility function. Indeed, for tractability, preferences of the worker are generally assumed to be separable in consumption and leisure. However, resorting to this kind of preferences in limited-commitment models reduce the comovements between hours and consumption since, in this case, the two variables are perfect substitutes, by construction. The modelled comovements might not be strong enough if a data matching experiment was tried. Hence, it will be useful to consider the more general case of imperfect substitution of consumption and hours and resort to the RBC formulation of preferences[25].

It is worth noting that recently, Jaimovich and Rebelo [2006] have developed a general formulation of nonseparable preferences in consumption (C_t) and hours worked (H_t) given by the following equation:

$$U = E_0 \sum_{t=0}^{\infty} \beta^t \frac{\left(C_t - \psi H_t^{\theta} X_t\right)^{1-\sigma} - 1}{1-\sigma}$$

where

$$X_t = C_t^{\gamma} X_{t-1}^{1-\gamma}$$

[25] As a matter of fact, Calmès [2007] suggested to study the use of nonseparable utility functions.

The presence of X_t implies that, in general, preferences are time nonseparable in consumption and hours worked. But depending on parametrization, the case of separable preferences can also apply here. According to Jaimovich and Rebelo [2006], the two classes of utility functions most widely used in the RBC literature are nested in this function. When $\gamma = 0$, the preferences proposed by Greenwood, Hercowitz and Huffman [1988] obtain and when $\gamma = 1$, the famous King, Plosser and Rebelo[26] [1988] preferences are used.

The utility function proposed by King et al. [1988] becomes separable when expressed in logarithmic form but, in level, the Greenwood et al. [1988] preferences, by construction, completely eliminate the income effect, with the disadvantage, compared to the King et al. [1988] case, that, as a consequence, the economy cannot evolve on a balanced growth path.

In a self-enforcing contract the income effect being weakened by risk-sharing, it might be conjectured that relying on nonseparable preferences (i.e. with $\gamma \rightarrow 0$) could dampen even more this effect, what could, in turn, strengthen the propagation mechanism generated by limited-commitment. On the other hand, there is a possible arbitrage to face because, as discussed in the second part of this chapter, separable preferences might help generate shock propagation by fostering more wage stickiness.

Variable capital and self-enforcing contracts

As Rosen [1985] has quite rightly pointed out, capital allows the aggregate disturbance to be partially diversified through capital accumulation in favourable aggregate conditions and through decumulation in unfavourable circumstances. This classical intertemporal trading reduces the income effect of aggregate shocks on consumption and employment and accentuates pure substitution effects. According to Calmès' [2007] conjecture, this, in turn, works as a propagation mechanism different from the one generated by limited-commitment. For researchers interested by data matching experiments, introducing variable capital in a self-enforcing labour contract model may help replicate stylized facts. In this respect, Calmès [2007] innovates by introducing capital in the production function of his model. But there, capital is held constant for tractability reasons. Actually, making capital variable leads to some complications. The distribution of capital holdings across entrepreneurs will become part of the state. The distribution of workers across entrepreneurs will depend on this distribution of capital holdings and the equilibrium will thus be conditional on the distribution of capital holdings.

Moreover adding these kinds of interactions in a self-enforcing labour contract model entails a contract set which is not necessarily compact, which creates an additional difficulty for solving the model. However, to tackle this issue, a lottery could be used to convexify the contract set as in Ligon, Thomas and Worrall [2000]. For that matter, the third agent (i.e. the financial intermediary) could also be explicitly introduced and a second relationship added to the toy economy, for instance a bank linked to the firm.

The question of the aggregation of contracts

After having considered firms as homogeneous, it is important to consider the more general case of heterogeneous firms. As noted previously, one of the principal results reported in Calmès [2007] is that aggregate dynamics depends crucially on the transitory dynamics observed at the beginning of the contract. There, transitory dynamics is, in turn, conditioned

[26] Their formulation of preferences is standard as it is compatible with a balanced growth path.

by the initial bargaining power of the worker and by the degree of imperfection of contract competition. And the initial distribution of the relative bargaining powers between the worker and the entrepreneur is itself dependent on the hiring conditions. To investigate aggregate dynamics in this context, it will thus be important to build an economy where contracts overlap. To do so, a technique aggregating heterogeneous contracts is again needed. Finally, should this question be investigated, it would be important to get in mind that the contracts analyzed here are dynamic. Hence, to aggregate such contracts, it is important to consider that the external opportunities of the agents should really depend on the states of the economy, especially on other contracts. In this respect, the question also relates to the endogeneization of imperfect contract competition.

The contracts duration

Regarding the duration of a contract, there are two extreme approaches in the limited-commitment models. Some, like Boldrin and Horvath [1995], set the duration of a contract to one period. It is obvious that the dynamic analysis of the macroeconomic aggregates is very limited in such a context. Researchers can at most consider contemporaneous macroeconomic dynamics. However, to avoid the backloading result, most authors in the field of limited-commitment theory set the duration of contracts to infinity. There is a technical reason for this assumption. According to Thomas and Worrall [1988], if workers had only finite life, they would surely renege on any contract which pays below the spot market rate in the final period. The model will then converge to the trivial equilibrium associated to a pure and perfect contract competition. Hence, since the very beginning of a contract, the only sustainable allocation would be the one associated to the regular flexible wage economy.

For this reason, Thomas and Worrall [1988] assume that the entrepreneur will try to accommodate workers in order to maintain a good reputation or to be in a position to attract workers in the future, and the authors rely on infinite duration[27]. However, the assumptions are very strong. For example, they prevent analysing such interesting topics as pension plans and retirement dates. Another way to avoid the backloading result is to directly introduce imperfect contract competition, with variable *P* in Calmès [2007] or by adding reneging costs as done by Thomas and Worrall [2007a]. In this case, it is possible to relax the assumption of infinite duration and consider the case of random termination dates.

Let us notice that, insofar as we incorporate the assumption of a random date for the duration of a contract, we must account for a third kind of heterogeneity. The initial model, where heterogeneity is minimal, because of the presence of only two different representative agents, has to be generalized to deal not only with the distribution of the initial allocation of capital and the initial conditions under which firms begin their activities but also, and this is a third kind of major heterogeneity class, the life expectancy of each firm.

Endogenizing the external opportunities

We have previously exposed how Calmès [2007] endogenizes the external opportunities of the worker and of the manager. As noticed by Thomas and Worrall [2007a], this is the right way to proceed because the transposition of the theory of self-enforcing contracts to macroeconomics must account for the more general case where the external opportunities

[27] The infinite horizon hypothesis is neither sufficient nor necessary to rule out the backloading result. As long as contract competition is perfect, this result obtains because the autarkic Pareto allocation dominates, strictly.

depend on the whole past history of the contracts and on the rational expectations about their future histories.

In Thomas and Worrall [1988] article, which paved the way to all subsequent macroeconomic models featuring self-enforcing labour contracts, the external opportunities were represented by an ad hoc spot market. For the worker, the self-enforcing constraint ensures that he is not better off by taking advantage of an external opportunity, working forever on the spot market if he reneges. After Thomas and Worrall [1988], most of the following studies have tried to consider that the external opportunity of the worker was determined by the discounted expected utility he will get from being employed elsewhere in the economy. But usually, this external opportunity remains partly exogenous or imperfectly endogenous.

There are many ways to endogenize the external opportunities of the agents. One, just mentioned, would consist in relating it to business conditions of a contractual economy. Another is presented next. We consider that the external opportunity is in fact an option and resort to the theory of derivatives to create endogenous external opportunities.

Self-enforcing contract and the option theory

As noted earlier, the outside opportunities of the worker and of the entrepreneur are actually options. The insurance provided by the entrepreneur erases the left tail of the distribution of the worker's payoffs, so this distribution is truncated as it is the case for financial or real options. In the theory of options, insurance is provided by a put. For instance, the outside opportunity of the worker may be considered as a put whose exercise price corresponds to the value of his outside opportunity. This put sets a floor to the gains of the worker, this floor being precisely the value of his outside opportunity. The payoff[28] of such a put is thus

$$payoff = \left(U^s - U^c\right)^+$$

where U^s, the exercise price of the put, corresponds to the value of the outside opportunity of the worker at termination date, say a contract on the spot market, and U^c are the gains that the worker gets from the contract. Importantly U^s can be endogenized by considering that the outside opportunity of the worker is an exchange option. If the worker reneges, he will receive this payoff, that is he will cash his insurance. It could stand for P_i, the payoff received by the worker in the first model presented in this chapter [Thomas and Worrall 2007b], if there were disagreement between the parties.

Let us notice that the computation of the price of an option is perfectly compatible with the methodology of dynamic programming. Assume that $U(S(t),t)$ is the value of an option at time t, $S(t)$ being its underlying asset. We can rely on the following Bellman equation to price this option [Tavella 2002]

$$U = sup\{F,\ \beta E(U')\}$$

[28] The payoff of an option is its value at maturity.

where F stands for the payoff of the option, which is its exercise value, the value of an option being the maximum of its exercise value (stopping value) and its continuation value $\beta E(U')$.

Self-enforcing labour market theory and growth

The question of economic growth can also be investigated in a limited-commitment economy. Most of the researchers in the field of self-enforcing labour contracts theory only consider a stationary economic framework. In such a framework, macroeconomic aggregates record no growth and the equilibrium converges towards a stationary or ergodic state. But it is well-known that most macroeconomic time series have a unit root or a deterministic trend, and in any case grow.

Once again, when the modelled economy is not stationary, we are confronted to the problem related to the non-convexity of the contract set. Should growth be studied in this kind of models, the link between stationarity and set convexity must thus be addressed. Indeed, since contracts depend on the past history of the firm, there exists a different contract for each level of capital and any endogenous state variable of the model. The objective is to arrive at the optimal contract resulting from a multitude of possible allocations.

To add growth to a limited-commitment model, a stochastic or a deterministic trend must be added to all macroeconomic variables depending on their nature. If the trends of the macroeconomic variables incorporated in the model are all deterministic, the model can be made stationary by rescaling these variables, i.e. by dividing them by their deterministic trend. But a macroeconomic variable which has a unit root, or equivalently which has a stochastic trend, cannot be "detrended" by following a procedure similar to the one used for variables having a deterministic trend. These variables must be expressed in first differences to make them stationary. Once transformed a lottery must then again be applied to convexify the contract set solutions.

Conclusion

This chapter exposes how to think about the self-enforcing labour market theory to study macroeconomic dynamic stylized facts, especially transitory macroeconomic dynamics, most previous models in this field being only concerned with wage contract dynamic properties or contemporaneous dynamics.

Since the seminal paper of Thomas and Worrall [1988], who made an allusion to contemporaneous dynamics in their concluding remarks, there have been no comprehensive developments on that matter until the contribution of Calmès [1999, 2003, 2007]. As explained in this chapter, it opens the door to a very promising field of research.

At this stage, these developments are still in their infancy and this avenue remains an emerging topics in macroeconomics. The changes to the existing limited-commitment models which are required to have them match the macroeconomic dynamic stylized facts about transitory dynamics present a challenging task. As mentioned, many of the proposed changes imply non-convex contract sets, which creates an important difficulty for solving these kinds of models. Fortunately, there are ways to deal with it, and with the proposed modifications, the macroeconomic models of limited-commitment economies will feature an internal propagation mechanism of shocks much stronger than the one used in the RBC literature (i.e. capital accumulation) whose dynamics is quite weak, to say the least.

References

Azariadis, C. (1975). Implicit Contracts and Underemployment Equilibria. *Journal of Political Economy,* **83**, 1183-1202.

Bailey, M.N. (1974). Wages and Employment under Uncertain Demand. *Review of Economic Studies*, **41**, 37-50.

Beaudry, P., & DiNardo, J. (1991). The effect of implicit contracts on the movement of wages over the business-cycle – evidence from micro data. *Journal of Political Economy*, **99**, 665-688.

Beaudry, P., & DiNardo, J. (1995). Is the behaviour of hours worked consistent with implicit contract theory?. *Quarterly Journal of Economics*, **110**, 743-768.

Boldrin, M., & Horvath, M. (1995). Labour contracts and business cycles. *Journal of Political Economy*, **103**, 972-1004.

Cahuc, P., & Zylberberg, A. *Le marché du travail*. Bruxelles: De Boeck; 2001.

Calmès, C. (1999). La théorie des contrats de travail autoexécutoires et ses implications macroéconomiques: un survol de la littérature. CIRPÉE, *working paper* **98**.

Calmès, C. (2003). La poignée de main invisible et la persistance des cycles d'affaires. L'Actualité économique, *Revue d'analyse économique*, **79**, 563-583.

Calmès, C. (2007). Self-Enforcing Labour Contracts and Macroeconomic Dynamics. *International Advances in Economic Research*, **13**, 200-213.

Cogley, T., & Nason, J.M. (1995). Output dynamics in real-business-cycle models. *American Economic Review*, **85**, 495-511.

Danthine, J.P. and Donaldson, J.B. (1992). Risk sharing in the business cycle. *European Economic Review*, **36**, 468-475.

Devereux, P.J, and Hart, R.A. (2007). The spot market matters: evidence on implicit contracts from Britain. *Scottish Journal of Political Economy*, **54**, 661-683.

Gordon, D.F. (1974). A Neo-Classical Theory of Keynesian Unemployment. *Economic Inquiry*, **12**, 431-459.

Greenwood, J., Hercowitz, Z., & Huffman, G. (1988). Investment capacity utilization and the real business cycle. *American Economic Review*, **78**, 402-417.

Holmström, B. (1983). Equilibrium long-term labour contracts. *Quarterly Journal of Economics*, **98**, 23-54.

Jaimovich, N., & Rebelo, S. (2006). Can news about the future drive the business cycle?. Standford institute for economic policy research, *discussion paper* no: **07-14**.

King, R., Plosser, C., & Rebelo, S. (1988). Production, growth and business cycles: I. the basic neoclassical model. *Journal of Monetary Economics*, **21**, 195-232.

Ligon, E., Thomas, J.P. & Worrall, T. (2000). Mutual insurance, individual savings, and limited commitment. *Review of Economic Dynamics*, **3**, 216-246.

Ljungqvist, L., & Sargent, T.J.. Recursive Macroeconomic Theory. 2nd edition, Cambridge: The MIT Press; 2004.

Rebelo, S. (2005). Real business cycle models: past, present and future. Northwestern University, *working paper.*

Rosen, S. (1985). Implicit contracts: a survey. *Journal of Economic Literature*, **23**, 1144-1175.

Tavella, D. *Quantitative methods in derivatives pricing: an introduction to computational finance.* New York: John Wiley & Sons; 2002

Thomas, J.P., & Worrall T. (1988). Self-Enforcing Wage Contracts. *Review of Economic Studies*, **55**, 541-554.

Thomas, J.P., & Worrall T. (2007a). Limited Commitment Models of the Labour Market. *Scottish Journal of Political Econonomy,* **54**, 750-773.

Thomas, J.P., & Worrall T. (2007b). *Dynamic relational contracts under limited liability*. Keele University, presentation.

Wen, Y. (1998). Can a real business cycle model pass the Watson test.?. *Journal of Monetary Economics*, **42**, 185-203.

Wen, Y. (2005). Where's the beef? The trivial dynamics of real business cycle models. *working paper* **2005-039A,** Federal Reserve Bank of St. Louis.

In: Emerging Topics in Macroeconomics
Editor: Richard O. Bailly, pp. 155-181
ISBN: 978-1-60692-093-0

Chapter 5

INFLATION TARGETING – THE MONETARY STRATEGY FOR MEETING THE CRITERIA OF REAL AND NOMINAL CONVERGENCE IN CENTRAL AND EASTERN EUROPEAN COUNTRIES (ROMANIA'S PARTICULAR CASE)

Rădulescu Magdalena *
Faculty of Economic Studies, University of Piteşti, ROMANIA

Abstract

Knowing the experience of the post-communist countries that faced high inflation rates right after the 1990, there is no wonder that many central banks of those countries embraced the idea of inflation targeting. After a period when most of these countries followed a monetary policy of stabilizing the exchange rate, many of them adopted an inflation targeting strategy. Although the central banks that adopted an inflation targeting strategy want to emphasize the importance of such a strategy in fighting the inflation, this fact is not necessarily true. In most of the Central and Eastern European Countries (CEEC) inflation faced a decreasing trend after the high values recorded right after 1990, no matter their monetary arrangements.

Inflation targeting in the transition economies was a more difficult and provocative task than in the developed countries. So, is this monetary strategy adequate for the transition economies, given the actual economic conditions, and if it is not, which will be the alternative?

Many times, the candidate countries declared that the requests of the Maastricht Treat regarding the inflation and the long-term interest rate are too restrictive for them and that these should not be applied to the CEEC. There are some important studies (the one of the vice-governor of the Hungarian Central Bank, Szapary G., 2000) that prove that, due to the Balassa-Samuelson effect, is rather counter-productive and not likely that the inflation rates in the new EU member states to rapidly converge to the ones of the best performers in the euro-zone.

So, the entry in the Exchange Rate Mechanism (ERM 2) must be preceded by reaching a high level of convergence, namely by a significant progress in the structural reforms area, by a fiscal consolidation process and by a responsible income policy.

* E-mail address:magdalenaradulescu@yahoo.com

If the CEEC enter the ERM 2 before reaching a sufficient advanced stage of convergence or of financial stability, the risk of some speculative attacks increases.

All the countries that acceded or will accede in the future to the European Union want to stay in the ERM 2 the shortest period that is possible (two years), although there are no formal restrictions regarding this period. For avoiding an excessive long participation to the ERM 2 it became important to establish the date of such entrance depending on the stage of reaching the nominal convergence criteria. Another important element is the larger and larger accent that lays on reaching the real convergence criteria before adopting euro.

That is the main issue proposed by this study: to analyze the complementary of the nominal and real convergence criteria in the CEEC and, especially, in Romania and the convergence level of those countries given the level of the EU-15 countries. In the same time, the paper emphasizes the importance of reaching the real convergence for a smoother entry in the euro-zone. In the conclusions part, we proposed some monetary measures, and not only, for achieving the nominal and real convergence in the CEEC and, especially, in Romania.

Key words: inflation targeting, real and nominal convergence, monetary strategies.

JEL Classification: E31, E32, E52, E58, E61, E63, F31, F36.

1. The Experience of Inflation Targeting in the Transition Countries

Given the experience of ex-communist countries which confronted high rates of inflation right after 1990, is no wonder that many of the central banks in these countries embraced the concept of inflation targeting. After a period when the majority fallowed a monetary policy based on stabilizing the exchange rate, many of them reoriented to inflation targeting. Although central banks that adopted the inflation targeting strategy tend to underline the importance of it in disputing inflation, this thing is not necessarily true. In the majority of Central and eastern European (CEE) countries the inflation had known a descending trend after the high rates registered right after 1990, disregarding the monetary arrangement choused by the central bank. Definitively, the recent global economic environment characterized by low inflation made easier for the CEE countries to control internal inflation. Generally, the implement of fiscal and monetary intern policy sensitive in transition economies, helped by positive evolution of the external inflation, led to this low inflation environment. If these policies will be continued in a responsible way, there is no doubt to believe that inflation won't remain at the low levels registered today even in the absence of inflation targeting policy. As Table no. 1 shows, the Baltic countries and Bulgaria, which had different monetary strategies, also reached low levels of inflation.

Table 1. Annual inflation - average level(%)

Year	Czech Republic	Hungary	Poland	Slovenia	Slovak Republic	Estonia	Latvia	Lithuania	Bulgaria	Romania
1999	2,1	10,0	7,3	6,1	10,4	3,1	2,1	1,5	2,6	45,4
2000	3,9	9,8	10,1	8,9	12,3	3,9	2,6	1,1	10,3	46,3
2001	4,7	9,2	5,5	8,6	7,2	5,6	2,5	1,5	7,4	34,8
2002	1,8	5,5	1,9	7,5	3,5	3,6	2,0	0,4	5,8	22,7
2003	0,1	4,4	0,8	5,7	8,4	1,4	2,9	-1,1	2,4	15,4
2004	2,8	6,8	3,6	3,7	7,5	3,0	6,2	1,2	6,2	12,0

Table 1. Continued

Year	Czech Republic	Hungary	Poland	Slovenia	Slovak Republic	Estonia	Latvia	Lithuania	Bulgaria	Romania
2005	1,8	3,6	2,1	2,5	2,8	4,1	6,9	2,7	5,0	9,1
2006	2,5	3,9	1,0	2,5	4,3	4,4	6,6	3,8	7,3	6,5
2007*	4,9	6,8	4,0	5,0	2,3	3,9	12,6	7,7	9,9	4,8

* partial data
Source: EUROSTAT 2007

Inflation targeting in transition countries was a more complicated and challenging task than in developed countries. As there experience shows, these CEE countries had sometimes missed by a long shot the planned target. Jonas and Mishkin (2003) analyzed the potential difficulties and evaluated the production in these three CEE countries which had an inflation targeting polity: Czech Republic, Poland and Hungary. They have concluded that, even though the progresses in disinflation were substantial, the relatively high incertitude degree in these countries makes the long term prediction of the inflation difficult. This is not a minor impediment. In developed countries, the gear of inflation targeting involves the technical ability of the central bank to:

- prognostic inflation with accuracy (6-8 months ahead)
- know how to adopt their instruments so that they meet the prognostic with the established target.

All the benefits emerging the inflation targeting in terms of measurement and liability depends on these conditions.

Given that the risks of making miscellaneous predictions – and therefore missing the inflation targeting- in transition economies are high, the central bank's liability may be affected by the gear of such a strategy. Adopting an inflexible register in leading monetary policy given the context of not fulfilled proper function criteria, in such an environment, central bank's liability can only be affected.

In a recent study, Stone (2003) shows that, even though series of countries in full developing process practices inflation targeting, in fact they don't meet the projected target. The problems of the policies in these countries concern the engagement that the central bank takes towards only one nominal anchor- the inflation. The same aspect is sooth for transition economies.

Therefore, is the system of inflation targeting adapted to the transition countries, given the present economic conditions, and if not, which is the alternative?

A thoroughgoing study of this three CEE countries experience analyzed by Jonas and Mishkin (2003) could determine some potential challenges for the further monetary policies.

Hungary. Officially, Hungary launched the inflation targeting in the middle of 2001. The first forecasted strap was of +/-1% centered on the targeted ratio of 7%. At the end of 2004, the initially announced targeted strap was of 3.5%+/-1%. As the inflation pressures started in the early 2004, the National Bank revised up the strap to 4%+/-1% at the end of 2005, with a central targeted parity at the end of 2006, 0.5% lower. Is worth mentioned that, even with inflation targeting conditions, the Hungarian Bank simultaneously targeted the maintenance of euro-forint currency inside of an interval of +/-15%, which in fact foreshowed the ERM II system.

Czech Republic. Adapted an inflation targeting system in December 1997 and had an initial strap of 5,5-6,5% and a time horizon of 1 year. At first, The Czech National Bank chose to target core inflation (basic inflation), but changed to CPI in December 2001. Starting with January 2002, Hungary adapted a four prognosis period inside of which the targeted inflation would gradually come down in a strap of 3-5% to 2-4% till December 2005. Since January 2006, they've announced that they will target an inflation of 3% until the integration of this country in Euro area.

Poland. In Poland, the National Bank established a short term inflation target of 8-8,5% in June 1998. All the other targets were settled at the end of each year. At the moment of the implementing announcement of inflation targeting, the National Bank would still keep a flotation strap for exchange rate. This was gradually enlarged in order to allow the free flotation of zloty (PLN) since April 2000. Since the end of 2005 and in 2006, the targeted interval of inflation was of 2,5%+/-1% and the long term monetary policy targets the maintenance of annual inflation as close as possible to 2,5%.

Before 2006, The Monetary Policy Council would announce the public about the monetary policy indicators. This happened with the purpose of limiting the flexibility of monetary policy and therefore, after 2006, the Council announced that will give out instead, throughout public statements, the factors which may affect the inflation hereafter.

Romania. The Romanian National Bank (NBR) announced officially the gear of inflation targeting in August 2005. Anyway, the last three years inflation targets had been published long before this moment. Initially, at the end of 2005, the inflation target was established at 7%. Afterwards, this was altered to a targeted strap of +/-1% centered on 7.5%. The inflation turned out to be of 8,6%. Still in 2006, sustained by the strong appreciation of LEU, the inflation came down to 4,7% lowering inside the strap established by National Bank of Romania (NBR) to 4-6%. For 2007 the NBR's targeted interval for inflation was of 3-5%.

Slovakia. Right after the adherence to EU in 2004, the government and the National Bank of Slovakia announced the strategy of EURO adopting for 2009. Therefore, it was settled a new framework of monetary policy, defined as inflation targeting in ERM II terms. The inflation targets announced for 2006 and 2007 are asymmetric, under 2,5% and 2%. However, these targets could turn out to be ambitions, if the present price of oil is maintained on the world market. In the last trimester of 2005, the inflation grew by its own from 2 to 3,9%. From the inflation evolution in these three analyzed countries we can acknowledge that the central banks had a limited success in reaching the inflation targets. This is not surprising, given the challenges that the transition economies confronts.

2. Adopting Euro by the New EU Members

Many times, the candidate countries made public the fact that, when it comes to inflation criteria and the interest for long term titles, the stipulations of the Maastricht Treaty are much too restrictive and it shouldn't be applied to them. There are important studies, among which we mention the one of the Hungarian National Bank vice governor, Szapary (2000), which proves that, because of the Balassa-Samuelson effect, is highly unlikely and unproductive for the inflations in the new state members to converge rapidly to the levels of the best performers in euro area.

Therefore, the entrance in ERM II should be preceded by the reaching of a high level of convergence, respectively by a consistent progress of the structural reform, by a process of fiscal consolidation and by a responsible policy of the revenues. This phenomenon is demonstrated, in fact, by the Greece's pre-participation experience to EMU, but also by the Hungarian recent experience, which faces economic and monetary difficulties after adopting an exchange rate system ERM II type.

If ERM II is adopted before reaching a sufficiently advance stage of convergence and a financially stable system, the risk of speculative attacks increases substantially.

All countries acceding to EU, plan a period of crossing through ERM II as short as possible (two years), although there are no formal restrictions regarding this period. In order to avoid an excessive participation in ERM II, is important to settle the moment of entry taking into consideration the fulfillment of criteria of nominal convergence. The carry out of the major adjustments regarding the tax system, the external sector and the fixed prices could assure the necessary conditions of a smooth crossing to ERM II.

In order to choose an optimal trajectory for the Romanian economy, when it comes to entering in the EURO area, some aspects should be looked into. A prime aspect regards the liability of each candidate state in elaborating a self strategy. Although is being stipulated that, after the moment of EU adherence, the monetary and the currency strategies of each country becomes a mutual interest, it is as clearly specified the fact that choosing a monetary and currency strategy after the EU adhesion is, first of all, a prerogative of the respective member state.

Another aspect is that it is expected for the authorities of each state not to take unilateral decisions, un-debated with the European Commission and Central European Bank, "Europeanizing" kind. There is a commune agreement that all the countries, including those with Monetary Council (Bulgaria, Estonia and Lithuania), have to participate at least for two years in ERM II before the evaluation of the fulfillment of nominal convergence criteria.

The third significant element is the increasing accent set on the meeting of the real convergence criteria before adopting EURO. Concerning this matter, the European Commission and the CEB advises a temporizing of the process, by one side through adjourning the entrance in ERM II, and by the other side through spending a larger number of years inside this mechanism then it is stipulated, pursuing a better preparation, when it comes to nominal criteria, but also real ones.

Faced with the argument that prolonging the participation period in ERM II increases the risk of exchange rate appreciation beyond the conceded corridor, the European authorities states that the mechanism is flexible enough to allow a resettlement of the central parity (in Greece's case) and, therefore, to the entire corridor, after a consultation of the implicated parties.

In the same time, it should be mentioned the fact that the ERM II mechanism contains, throughout its construction, two confliction objectives: first of all the stability of the exchange rate, and, second of all, the reaching of a lower inflation ratio. Therefore, we can bring to discussion the Hungarian example, in order to observe how the two objectives interact.

Starting with May 2001, the Hungarian authorities adopted unilaterally a similar mechanism to ERM II (with floating ranks of ± 15% on central par of exchange), targeting the EURO adoption for 2006-2007. In parallel, was adopted a inflation targeting strategy, and for reaching the ambitious targets of lowering the inflation, the nominal appreciation of the forint's exchange rate was plenty used, till this was able to allied to the superior edge of the currency strap. The elective cycle was no help: after the elections of April 2002, the budget deficit grew massively, thanks to increasing to 50% the salaries of the budgeting personnel. National Bank had to rise up the level of interest rates (which were already higher than those in region, except for Romanian).

The mix between the high interest rates and the almost fixed exchanged rates gave birth to speculative inflows, in the conditions of the liberalization of the capital account. After the 2004 adherence, the forint's exchange rate continued to appreciate. Anticipating a rank's rearrangement (central parity reinforcement), the speculators made massive portfolio investments in Hungary in only few hours.

In that very moment, the Hungarian Government started being preoccupied by the competitiveness of exports, standing against a nominal appreciation. Confronting with the impossibility of altering the central parity without the government's approval, the Hungarian National Bank, had to defend the rank of exchange rate flotation, which meant sacrificing the inflation objective. So, after two years of standing inside the inflation target (2001-2002), fallowed another two years in which the inflation target was missed (2003-2004), derogating the central bank's liability. In 2002, the data for euro adopting was suspended until 2008, and at the moment a new postponement is being taken into consideration.

In fact, recent studies put into debate even the possibility of adopting euro in a shorter interval than a decade from the Hungary's adherence, not only for this country, but also for Czech Republic and Poland, especially because of problems regarding these countries ability of controlling the budget deficit, because of the aggravation of access conditions for these states to the external financial market.

Is the Hungarian example proving that the prolonged participation to ERM II is risky? The European authorities may object to this interpretation using these arguments:

- The adoption of the mechanism was made unilaterally, without consulting the other involved parts;
- These authorities did not appeal to the available option of rearranging the rank;
- the strategy of direct targeting of the inflation is not compatible with the ERM II mechanism, this being more adequate to targeting the exchange rate;
- being a unilateral arrangement, an explicit or implicit engagement from the central bank in euro area did not exist for interventions supporting forint reaching the edge of the rank. Anyway, we should underline that the intervention in sustaining a participant country to ERM II is not, in fact, self-contained, nor unconditional (if contravenes to the objective of banks handling the inflation, the intervention can be suspended), which increases the risks of prolonged participation in this mechanism.

We most also bring to mind the principles that guides the European Commission and the European Central Bank (ECB) when it comes to the new candidates. The first one is obvious: no country should adopt decisions of monetary of currency policies unilaterally, without consulting the European authorities. The second one – the absence of a unique trajectory for the euro adopting- and the third principle – handling case by case – are extremely relevant for Romania which, because of its particularities, will have a different way to those of the majority of countries in the region which acceded to EU. Finally, the last principle, the one of equal treatment, allows us to make predictions regarding the rules Romania has to fallow, based on the other countries experience. Still, this last principle is debatable and it can be, in certain limits, submitted to derogation, in accordance to each candidate's level of preparation. For example, Italy and Finland were allowed shorter periods than the compulsory years of participation to the ERM II mechanism (of 14, respectively 16 months), but is extremely unlikely and risky (as we've already shown) for the new members to receive such a generosity gesture.

3. The Long Term Romanian Strategy for Adopting Euro

The above presented elements show that some of the criteria of nominal convergence and – especially – the criteria of real convergence can be fulfilled by our country only with sustained efforts on a long period of time. Therefore, the reduction of the proportion of occupied population in agriculture will presume the powerful development of the services and of other alternative occupational fields, which means significant investments, either in the existent cities, or even in rural areas. The EU adherence will undoubtedly contribute to the accelerating of the real convergence, to which will be part private investors and European funds especially intended.

Although it may be admitted that throughout a sustained effort the Maastricht criteria could be fulfilled in a relatively short period (1-2 years), the sustainability of these nominal convergence depends in its final stage on meeting the real convergence. Plus, adopting restrictions, especially with monetary character, with the purpose of meeting the very short terms of the nominal convergence could have a negative impact on maintaining an accelerating rhythm of economic growth, which, in fact, would delay the real convergence. That is way, we consider that adopting programs with clear objectives, but realistic ones, would be able to support the accelerating development of Romanian economy even beyond adherence to one currency.

After the moment of joining ERM II is possible to keep the strategy of direct inflation targeting, but adopting an exchange rate targeting strategy is also a possibility. If we take into consideration the Hungarian experience as being a relevant one, it seems that the second option is preferable, in order to prevent conflicts between objectives. It is possible that until that moment, Romania to increase significantly the degree of economy's opening, so that the transmitting channel for the exchange rate to become uncommonly efficient.

From the Hungary's unpleasant experience inside of the ERM II mechanism we can strip off a fundamental lesion, which can be synthesized as: once announced the due date for euro adherence, the entire society must mobilize in supporting this option, and the decisions of macroeconomic and microeconomic policy should comply. From the Hungarian lesion these ideas stand up:

- The advantages of a low inflation must be explained repeatedly to the political class and to population, otherwise, there support can be absent exactly in the most delicate moments of the anti-inflationary battle (like the episode of increasing the incomes of the budgetary personnel or that of stimulating the exports throughout currency policy);

- The independence of the central bank should be guaranteed not only theoretically, but also in practice, otherwise it risks being captured between confliction objectives;
- reaching both the objectives stresses a powerful support from the government, by impressing a prudent character to fiscal policy and to that concerning the revenues, as well as from the population;
- Another important aspect is that of maintaining under control the credit expansion, because otherwise, the consequences on the current account and inflation can be severe;
- A very prompt communication is necessary between the government and the central bank. For nothing have Hungary and Poland inter-ministerial committees (at finance ministry level and national bank) specialized in supervising the steps to adopt euro, if the respective institutions have different objectives;
- the government should be less concern of the exporters faith; the competitiveness in a unique market is gained not through devaluation, but by innovation, design, deadline meeting shipping, distribution networks, service and guarantees, customers prospecting;

Along with these general aspects, we can also acknowledge instructions with a more technical character:

- The enlarged exchange rank ($\pm$ 15%) or the narrow rank ($\pm$ 2,25%) can be protected only if the fiscal and revenues policies support this demarche;
- is dangerous to leave the currency float close to the edges of the rank, as it can generate speculative attacks;
- In small and opened economies is very difficult the consolidation of disinflation when a deteriorating of the current account takes place, because of importance of exchange rate channel;
- the entire liberalization of the capital account before the inflation and the interests came down to unattractive levels for the inflows of speculative capital carries a incredibly large risk

So all these can be condensate in two recommendations for the monetary policy:

- the period of operating in ERM II shouldn't be longer than two compulsory years, scarcely because of the difficulties of reconciling the currency policy to the inflation objective;
- the ERM II entrance should take place only when the stage of fulfillment for the convergence criteria (nominal and real) will be advance enough; therefore, at the end of the two years the adherence of euro currency will be possible.

The plans of adopting the unique European currency seem to be further away for Romania. After, several years ago, it was promoted a calendar for the years 2010-2012; this was postponed for the period 2012-2014. But, a certain date can't be appointed, especially because the economic analysts say that the inflation will lower fewer than 3% only in 2010.

The disinflation process will get ever slower. After the Romanian economy knew reductions of the inflation rates with proximally 20 percent at the beginning of the 2000, lowering fewer than 3% - which indicates the fulfillment of the criteria regarding the convergence to the euro area – will be established only in 2010.

For Romania the adoption of euro is impossible to be made before 2012. It can actually postpone for a few more years, because, even though the moment of initiation the process of adopting the unique European currency is settled with the Romanian authorities, the effective moment of the adoption is still established by the European Commission. Meaning, the European officials won't approve the entrance of a state in the euro area until they are not convinced that the respective economy can handle some shocks, and the shortest period for testing the economy's resistance is of two years.

One of the main reasons for which the inflation won't slow down to quick is explained by the effect of real price convergence (Balassa-Samuelson effect). That means that, in the next years, at least one segment of the prices will continue growing, so that it reaches the same level of those in other EU countries. Even in the last report concerning inflation, the National Bank's officials specified that, in 2007, some prices will grow, and others will reduce, as a result of Romania accessing to European Union.

Still, the adherence effects on prices will be felt over a longer period of time. It is estimated that a growth, average, by 2,2%-2,3% of the goods and services prices will come only by Romania's integration into EU (Balassa-Samuelson).

Therewith, we must take into consideration the fact that, when the adoption of euro will be firmly settled, the Romanian economy will lose the independence of the monetary policy. Meaning, the decisions of modifying the interest rates will be taken by the Central European Bank, not by the National Bank's officials.

A sudden strategy of adhering to a monetary union presumes that, starting with a certain day, Romania and EU use the same currency, euro, issued by ECB, and the Romanian citizens, are being announced to change LEU in EURO at a certain fixed exchange rate (we suppose a EURO-LEU exchange rate of 1Euro/3Leu; Romanians will purely buy at the established date 1 EURO paying for it 3 LEU).

For Romania, such a strategy would offer an important advantage. First of all because the inflation rate would lower at the level registered in EU, without supplementary costs in terms of unemployment. EURO- currency issued by ECB – is now the legal way of payment in Romania and, therefore, the inflation in Romania equals the inflation in EU. The economic agents acknowledge this fact and, so, they outline accordingly the inflationary anticipations for Romania.

Thus, Romania can immediately benefit by the ECB reputation. Assuming the currency and the EU institutions, Romania can benefit instantly by earnings of prosperity generated by the monetary union. Such a sudden strategy would imply, in Romania's case, the reduction of the inflation ratio before adhering to the monetary union. Else, Romania won't need a monetary reform, which would bring closer the inflation ration in Romania to that in the euro area.

The empiric evidence shows that countries with high inflation which are trying to materialize the convergence criteria inside a transition period continues to register high rates of interest, which shows that the inflationary anticipations tempers extremely slow.

In conditions of fixed exchanged rates, the relatively high rates of interest are a stimulant, as the attract flows of capital on short term. Rational investors, anticipates that countries with high rates of inflation will devalue, at a certain time, their own currency. This short term capital flows tend to increase, in a high inflation country, the money quantity, which complicates the anti-inflationary task of the monetary policy. *Finally, a too long period of transition jeopardizes the odds of reaching the ultimate objective.*

From *the Barro-Gordon pattern* emerges that the countries with small rates of inflation have no significant stimulant in adopting a unique currency. By example, Germany has nothing to gain, as the evenness inflation knows no modification (from this point of view, inflation, respectively unemployment rate) as a result of the fact that it joins a monetary union to which participates countries with high rates of inflation; countries with high rates of inflation will be the ones gaining in this situation.

An analysis on the perspective of *political economy of monetary integration* in Europe brings in front the fact that a country such as Germany is disadvantaged by the entrance in the union of a high rate inflation country. Actually, the interests of such a country could, before any other, be that of reducing the number of countries which adhered to euro and prevent the access of high inflationary countries into the monetary union. Also, by the perspective of *the theory of optimal monetary areas*, a European monetary union, to which will take part all the EU countries, is suboptimal when it comes to well-being.

By the perspective of political economy we can say that The Maastricht Treaty reduces the possibility of the participation of an enlarged number of countries to euro, because of the opposition of the powerful countries with low levels of the inflation ratio.

4. The Complementary of the Nominal and Real Convergence Processes

The process of catching-up in the Central and Eastern Europe is influenced by the quality of management of the two types of processes – real convergence and nominal convergence, the last one indicating new constraining for the economies which intend to adhere to euro area.

The process of nominal convergence was privileged vis-à-vis the real convergence one, because its realization needs a short term horizon.

The progresses in accomplishing the Maastricht criteria influenced the real economic variables, the two processes being in fact complementary. Initially, nominal convergence may generate a reduction of the performances, but the complete fulfillment of the Maastricht criteria is capable of assuring a higher macroeconomic stability, which will create the premises of a superior rate of economic growth. As higher the flexibility of an economy is, as rapid the adoption of a new macroeconomic regime is, which will lead to the intensity of the initial impact.

The analysis of the Maastricht criteria in Romania's case marks out the obedience of fiscal criteria regarding the budget deficit and the public debt, as well as the existence of a significant postponement when it comes to monetary variables – inflation ratio and interest rate. The variable exchange rate will be taken into consideration in the two years of membership to ERM II, when it can't vary more then +/-15%. Thus, the perspective of adhering to this mechanism of exchange rates must involve the assurance of a higher capacity of the real economy to react towards the action of sundries shocks, in order to limit the usage of exchange rate instrument.

In Table no. 2 that we may see below we analyzed three of the criteria of nominal convergence. The one regarding the interest rate to bonds for 10 years wasn't included, as they were issued for the first time in Romania, in 2005. In July 2005, the interest rate for the 10 years Eurobonds was of 3.22% in the three countries with the lower inflation ratio; therefore, the internal interest rate for these titles shouldn't surpass 5.22%.

Table 2. Nominal convergence criteria for Romania (2003-2007)

		2003	2004	2005	2006	2007
Average annual inflation rate (%)	The average of the best three countries with the lowest levels	1,2	0,9	1,0	1,2	1,3
	Reference value (average + 1,5 percentage)	2,7	2,4	2,5	2,7	2,8
	Romania*	14,1	9,7	8,3	7,0	6,5**
Budgetary deficit (max. 3% of GDP)	Romania	2,3	2,5	0,7	1,9	2,5
Public debt (max. 60% of GDP)	Romania	26,8	23,1	20,4	16,2	15,6

* At the end of the period; ** partial data

Source : Eurostat, National Forecasting Commission, Central and Eastern European Report 2006.

The fifth element of nominal convergence, the stability of exchange rate, depends on the fulfillment of the criteria regarding the inflation ratio. A very important conclusion is that a real terms appreciation of the exchange rate balances the inflationary process.

When it comes to exchange currency stability, the national currency must be inside a floating rank tidier then ± 2.25% and a larger rank of ± 15% for two years before meeting the European mechanism of the exchange rates ERM II. In practice, European Commission and ECB can tolerate a rank of (2.25%, 15%).

In 2006, Romania's external financial position seems to be solid comparative to the other advanced CEE countries (see Appendix no.1). The country's rating is medium (and increasing), the level of public debt is relatively low (15%) and there are no difficulties in financing the external debt service.

Still, there is a strong risk regarding the sustainability of the current account deficit on medium and long terms. If, for the period 2001-2005, Romania covered the deficit of current account in a proportion of 70-75% on direct foreign investments, in 2006 and 2007, the rate of covering on FDI will go lower than 40%. This thing is due to reducing FDI as a result of the proximity to finalizing the privatization process. We will have to access funds on the international capital market.

The lack of NBR's intervention on the currency market will have to continue so that the exchange rate to be as flexible as possible. The intervention should be selective only when exists the risk of powerful speculative attacks. The flexibility of the exchange rate will allow NBR – in ceteris paribus conditions – to assure stability.

The exchange rate policy of the central bank shouldn't counteract the Balassa-Samuelson effect, but it should allow an appreciation in real terms of the national currency equivalent to this effect. Because of superior growth of labor productivity in the sector of tradable, this will be able to support the appreciation of the national currency, without losing the external competitiveness.

The sustained growth of the labor productivity, the well-being of the macroeconomic situation and the low efficiency on international environment will lead to the increasing of the volume of capital flows, creating supplementary pressure for real national currency appreciation. In these conditions, in order to maintain the macroeconomic equilibrium, next to the exchange rate policy of the central bank, is necessary a fiscal policy adequate (restrictive) (The NBR's Annual Report, 2004).

A *restrictive fiscal policy* in time of capital inflows helps restraining the aggregate demand and weakening the growth of prices, can limit the real appreciation of the currency float (especially when the largest amount of the governmental expenditures are due to non-tradable), can limit the deficit of current account, can discourage supplementary inflows (by reducing the interest rate) and maybe by growing the degree of savings.

Romania could fulfill entirely the Maastricht criteria beyond 2010, given the fact that the disinflation process in Romania will be slower than the initial estimations. A supplementary problem could appear when it comes to target the budget deficit, because of the supplementary budget expenditures necessary to co-financing the structured funds, participation to the communitarian budget, modernizing the infrastructure, adapting to an economy based on knowledge, as well as protecting the environment. Due to these conditions, the solution of reducing the budget deficit stands on a more efficient collecting of the taxes, as well in diminishing the cvasi-fiscal deficits. Promoting a restrictive fiscal policy in order to increase the budgetary incomes could affect in a negative way the process of real convergence.

The comparison to CEE countries which already adhered to EU (CEE -8) revels a higher convergence in the field of public finances and a more reduced one in monetary plan, Romania registering the higher inflation ratio in the region. The medium estimated budget deficit for these countries for the year 2007 reaches 4.1% from GDP, and the public debt 42.9%. Values beyond the target registers larger countries (Poland, Slovakia, Czech Republic, Hungary), which opted for the postponing of the entry in the euro area, in order to support the real convergence process.

The fulfillment of the nominal convergence criteria are, by one hand, a favorable influence (thanks to the reduction on inflation ratio), and, by the other hand, an unfavorable one (by respecting the public finance criteria) over the process of real convergence. Therefore:

- The reduction of the rates of inflation and interest (in order to respect the Maastricht criteria) determines the growing of the investments and therefore GDP. The empiric evidences suggest that, in terms of a reduced rate of inflation, the economic performances are better then in terms of moderate inflation. Plus, the reducing of the inflation ratio accelerates the process of convergence of the incomes.
- the assessment of respecting the Maastricht criteria (especially when it comes to budget deficit and the public debt) can affect the convergence process of the economies in which the level of investments is low; the existence of sustainable budget deficits (ever higher than 3% from GDP) can contribute to rapidly structural adaptation of these economies to EU demands.

Also, the real convergence process influences nominal variables, favorably and unfavorably:

- The structural reform quickens the GDP/inhabitant convergence which leads to an non-inflationary growth of the salaries; also determines a growth of the incomes, an increasing of the budgetary collecting, therefore to the reduction of the budget deficit and the public debt;
- The differences of productivity between tradable fields and non-tradable fields, given the uniform salary increases in both the fields (Balassa-Samuelson effect) will generate the persistence of a higher level of inflation.

One of the sources often quoted in literature in regards to the relation of dependents between real convergence and nominal one refers to the Balassa-Samuelson effect. According to this, the increasing of the competitiveness in fields exposed to external competition determines the growing of exports and the currency offer which generates the appreciation of the currency, and the tendency of equalizing the salaries between the two sectors (tradable and non-tradable) in economy is responsible for the internal growth. Its manifestation presumes the nominal appreciation of the national currency, along with the increasing of the inflation ratio, given the flexible rates of exchange.

The econometrical estimation of this effect in Romania reflected an impact over the inflation situated between 1.14 percents in 1995 and 1.78 percents in 2004. As average for the period 1995-2004, the inflation caused by the Balassa-Samuelson was of 1.57 percents. The appreciation of the currency floating based on this effect was average 2.03 percents in the same period. The alignment of the prices to utilities to those practiced in EU, given the lack of reorganization in non-tradable field, will generate the persistence of the effect in Romanian economy, with consequences over the intensity of the disinflation process.

Other sources of the correlation between the two forms of the convergence refers to the relation between economies and the national investments, to the evolution and the financing of the deficit of current account, as well as to the implication of liberalizing the capital account and to adopting the strategy of targeting inflation (which regards the reduction of the inflation and production variability).

On regards to the process of economic growth on medium term, the standard answer would be that NBR has to decrease the rate of interest in order to stimulate economic activity; but, this presumes taking into consideration the next adverse effects:

- if the economy passes over a period of economic expansion, as Romania’s situation, in 2004-2005, then the variability of inflation would increase and the liability of the central bank would diminish in reaching the targeted inflation;
-if a weak elasticity of the crediting manifests to the modification of the interest rate in LEI, then the impact on economic growth will be reduced (the transmission channel of the monetary policy throughout the interest rate has a reduced importance just because of currency crediting, stimulated by the nominal appreciation of Romanian currency - leu);
-if the interest rate becomes more reduced than the inflation ratio, then the savings process will be affected; according to Solow pattern, the increasing of its rate assures an economic growth on medium term, because of the large accumulation of capital.

Conclusions of the nominal and real convergence analysis for Romania are as follows:

- The moment of entering in ERM II is provided for 2012 (given the assurance of the necessary period for: fulfilling the criteria of nominal convergence; the realization of significant progresses in the real convergence process)

- The period of ERM II participation will be reduced to the compulsory minimum time of two years
- The entrance into the euro area will take place on the horizon of 2014;
- The fundamental objective of the monetary policy is the diminishing of the inflation, given the conditions of maintaining the process of consistent economic growth;
- For 2007, the target inflation is established to 4% (dec./dec.), with a tolerance rate of one percent in one direction or another;
- For 2008, the central target is of 3.8%(dec./dec.), with the same tolerance rate;
- On medium term, the inflation targets will be settled by taking into consideration the necessity of keeping the process of disinflation on a trajectory which is in accordance with the convergence criteria;
- The reaching of these objectives is fully achieved, by the condition of continuing the appliance of the set of economic policies fallowed for the last years, which, as previously showed, lead to a substantial reduction of the inflation;
- Still, it must be said that, a rhythm of disinflation to high, uncorrelated with the rhythm of adjustment of real economy, would be unsustainable on medium term and would encourage the installment of a policy stop and go type. Thus, the disinflation rhythm must take into account the inflation differential towards euro area, caused by "Balassa-Samuelson effect"
- Finally, we must add that, after the liberalization of the capital account, between the midway objectives of the monetary policy there is a conflict, which undermines the NBR's capacity of applying the disinflation process. So, the increasing of the interest rate with the purpose of decreasing inflation determines the growth of capital inflows, which exerts powerful pressures on the exchange rate.
- The main affront of Romania's entrance into the EU is that of meeting the real convergence;
- Regarding the nominal convergence, the main problem is disinflation consolidation;
- In the particular plan of the monetary policy, the main consequences of the adherence to EMU are:
  1) the alignment of the monetary policy to that in the euro area, in order to avoid shocks produced by the sudden changes of the monetary policy's orientation;
  2) carrying out the efforts of inflation decreasing in a realistic rhythm; is compulsive the continuous enforcing of the "*direct inflation targeting*" for the period preceding the ERM II entrance, but also inside the faze of the participation;
  3) the currency policy must allow the adjustment of the real exchange for the leu; in the onwards period to entering ERM II, is compulsive finding of equilibrium level of the exchange rate, which afterwards will be used as pivot flow;
- Ending the adoption of the *communitarian acquis* in the next two fields:
  a) the banking activity and the surveillance proceedings;
  b) the improvement of the institutional environment, by straitening the central bank's independence and assuring its status's compatibility with the "*European System of Central Banks*"

- the development of banking system and of entire financial system, in order to increase the rationalization of the resources assigning and assuring the monetary policy with the adequate transmission channels
- Updating the payment system and connecting it to the European system T.A.R.G.E.T.
- Finalizing the harmonization of the monetary statistics

5. The Evaluation of the Degree of Real and Nominal Convergence for the CEE Countries with EU-15

When it comes to real convergence process there are no formal criteria or a total agreement concerning the variables which should be taken into consideration; some of those refers to rates of GDP/inhabitant growth and to the productivity of different fields of activity in GDP, to the evolution of the degree of economic integration.

In order to mark out the degree of economic convergence of Romania to EU, relative to other CEE countries, we analyzed the indicators that Deka Bank proposed – Converting Europe Indicator (DCEI) and, respectively, by Deutsche Bank, in 2004, when the first wave of CEE countries acceded to the EU.

The first one refers to four variable categories which dignifies the degree of adjustment of an economy in accordance with the adhering process to EU;

a) *monetary convergence* – evolution of inflation, long term interest rate, nominal exchange rate and the increase of the degree of financial intermediate;
b) *fiscal convergence* – budget deficit, public debt, extern debt;
c) *real convergence*- GDP/inhabitant, percentage of agriculture in GDP, unemployment rate and percentage of commerce to EU;
d) *Institutional convergence* – analysis of transition BERD indicators, as well as the stage of *communitarian acquis* implementation.

The analysis of this indicator in 2004 (according to Table no.3) allows the marking out of the fallowing groups of the states, function of the convergence degree:

- First one, including Estonia, Slovakia and Czech Republic. The second country leads when it comes to real convergence, but Estonia made bigger progresses regarding the nominal convergence (as a result of the successful implementation of the monetary council and fiscal reform) and institutional convergence. Czech Republic registered a lower score, because of the fiscal related problems (significant subsidies from the national budget and the level of public debt);

- Second one, including Poland, Slovakia, Hungary, Lithuania and Leetonia; Slovakia introduced the unique taxing quota of 19%, with positive effects in drawing FDI. Hungary registered some problems concerning the evolution of inflation and budget deficit, which negatively influenced the fulfillment of nominal convergence criteria. In Lithuania’s and Leetonia’s cases, the adoption of the monetary council will allow the faster adhesion to euro area;

- The leader of the third group is Bulgaria, as a result of the progresses concerning the fiscal and monetary convergence. Romania is the last one in this classification, being outmatched by Croatia, because of the decreased performances in assuring the macroeconomic stability. Given this terms, even though the evolution of real GDP after 2000, as well as the decreasing rate of unemployment enhanced the synopsis afferent to real convergence process, Romania did not register a higher degree of convergence in 2004, but even a more reduced one comparing to 2003.

Table 3. Convergence European Index (CEI) 2004 (100 – average level of EU – 15)

	Total convergence	Real Convergence	Institutional Convergence	Monetary Convergence	Fiscal Convergence
Estonia	84	75	85	90	85
Slovenia	83	100	80	80	75
Czech Republic	81	90	80	90	65
Poland	74	60	85	80	75
Slovak Republic	74	60	80	85	75
Hungary	71	85	85	60	60
Lithuania	71	45	80	85	85
Latvia	69	60	85	65	70
Bulgaria	59	25	80	75	80
Croatia	48	30	55	80	40
Romania	41	50	75	10	75

Source: Hanush, H., Balzat, M., "A new era in the dynamic of European integration", 2004.

During 2001-2004, all countries faced progresses in converting to EU process, first of all because of the favorable evolution of the nominal convergence variables. According to this methodology, Romanian economic convergence is higher than year 1995, when the value of DCEI index was of 27. Thus, comparing to Bulgaria, the progresses were smaller, given the fact the value of this index was of 21.

The second indicator of economic convergence – hold forth by Deutsche Bank Research – is build taking into consideration the same criteria as for the DCEI index, but emphasizing the external equilibrium of transition economies;

- *real economy* (GDP/inhabitant, the percentage of agriculture in GDP, rate of unemployment, percentage of private sector in GDP, rate of investments, rate of GDP growth and of productivity);

-*quality of institutions* marked out by ERDB index (legal system, government, banking system, commerce and currency market liberalization);

-*external sector* (balance of current account as percentage in GDP adjusted with FDI inflows, the degree of commercial integration in EU);

- *Monetary and fiscal conditions* (inflation ratio, budget deficit and public debt).

The analysis of evolution of this indicator during 2000-2004 (according to data presented in Table no. 4) shows the enhancement of convergence degree of Romanian economy to that of EU, still this is situating on the last position among the candidate countries. Thus, comparing to DCEI index, the degree of economic convergence is higher, and the postponement towards Bulgaria is a lower one. Therefore, leakage of convergence index values reduced in 2004 vis-à-vis 2003, which makes more difficult the marking out of the different groups of countries by there performances. The most convergent country is Slovakia, and the other adhered to EU countries forms a block almost compact, having an index between 66.3 and 71.3% from EU-15 average. Romania is the only country which enhanced the economic convergence degree every year during this period, thanks to higher rates of economic growth starting with 2001, as well as to disinflation process. But, external sector evolution together with the quality of institution explains the decreased rhythm of convergence growth.

Table 4. Deutche Bank's Convergence Index 1999-2003 (% UE)

	2000	2001	2002	2003	2004
Czech Republic	66	69,9	70,1	73,2	70,6
Estonia	62,7	66,3	70,1	69,4	71,3
Hungary	65,6	70,3	71,8	71,4	69
Latvia	58,7	62,1	64,9	70,6	69,2
Lithuania	45,3	56,9	59,3	67,1	66,3
Poland	60,7	63,5	65,1	65,1	67,4
Slovak Republic	57,4	61,7	64,3	70,3	69,4
Slovenia	57,3	71,3	73,6	82,9	75,5
Bulgaria	53,8	56,5	59	65,8	63,1
Romania	44,2	50,9	53,3	59,3	61,8

Source: Deutsche Bank Research, 2001-2004.

Although this index had been calculated until 2004, still we can estimate its tendency, relative to the variables in its structure. There analysis suggests the growth of convergence degree, because of assertive influence of real variables (GDP and productivity growth, evolution of unemployment rate), presented in the chart below. The deteriorating of balance of current account is no issue, as long as is covered massively by stable flows of capital – FDI, but affects negatively the value of this convergence index. However, results must be read with caution, because the simple comparison of the growing rates of different variables is irrelevant, given the important initial postponements.

The compared analysis of the evolution of real and nominal variables allows the measurement glaring in which economic cycle in Romania is correlated to that of EU (Table no. 5)

Table 5. Comparative trend of the real variables GDP, productivity, unemployment rate

	2002		2003		2004		2005		2006		2007	
	EU-15	**Romania**	**EU-15**	**Romania**	**EU-15**	**Romania**	**EU-25**	**Romania**	**EU-25**	**Romania**	**EU-27**	**Romania**
Real increase of GDP (%)	0,7	5,1	1,0	5,2	2,3	8,3	2,1	4,2	3,0	7,9	2,9	6,0
Labor productivity increase (%)	0,6	7,9	0,6	5,0	1,6	8,4	1,1*	5,6	0,8	4,9	0,9	5,6
Unemployment rate (%)	7,6	8,4	7,9	7,4	8,0	6,3	8,0*	5,9	7,6	5,2	6,9	4,1

* EU - 15

Source: EUROSTAT, National Bank of Romania

Econometrical estimation of the economic convergence degree gives information only concerning orientation of different variables in studied country relative to EU. We shouldn't understand that the degree of divergence presumes an unfavorable evolution of the economy; therefore, Romania registered a real GDP growth each year during the analyzed period, and EU knew a slower growing of it starting with 2000. So, this lack of synchronization of business cycles "enabled" Romania to recover the revenues shortage! Conclusions of the study are more significant regarding the nominal variables, given the fact that this are manifesting a reduced and relatively stable level in EU.

6. Priorities of the Real Convergence

The main objective of adhering to European Union is the real revenues growth. However, catching-up process is more difficult than many imagined and Romania confronts many pressures previously acknowledged in the new member states, which proclaimed there adhesion in 2004. The decision making factors were tented to consider the grown macroeconomic imbalances as being, more likely, inborn part of the convergence process, than the result of weak economic management. Thus, new member countries experience shows that, while transition process can bring to front specific problems, the principles on which the stable economic management lays on can't be ignored. This work analyzes the most commune ten "myths" concerning the convergence process which marked the economic and political thinking in Romania.

Initially, Romania and Bulgaria found themselves behind the other countries adhering no long time ago, which now confront many changing tendencies, as powerful consume and investments and the deepness of external disequilibrium. A few elements were crucial for these evolutions, such as: on-coming of the adhering moment, privatization and structural reforms, development of functional markets as well as macroeconomic stabilization.

Will the EU adhesion lead to rapid revenues enhancement? One of the lessons learned from the previous adhering processes was that the "catching-up" process requires a lot of time and continues long time after adhesion. The transition process doesn't prevent the risk of a stubbed evolution, "stop-and-go" kind. Real convergence is reached throughout sustained macroeconomic policies.

Undoubtedly, the perspectives of EU adhesion quicken the reforms in Central and Eastern European countries being in adhering process. Romania was contemplating the benefits of the process, especially massive investment inflows and rapid economic growth. Still, the experience of the countries which adhered in the previous stages showed that the process of real convergence is slow and must be diligently administrated. Even having a powerful economic growth, Romania would still need many decades in order to reach revenues level in EU. Certainly, many of these pressures are the direct result of disinflation process. However, these pressures will continue for several years and sustained economic growth will depend on the success that the governors achieve in transition administrating.

Disinflation and especially the acceleration of sustained economic growth regarding the acceding of real convergence depends in a very important measure of ensemble coherency of authorities economic programs. Preferable these programs will be followed up during long periods of time, crossing over the horizon of an elective cycle.

First of all, the problem of arrears must be solved and subsidiary, cvasi-fiscal deficits. The arrears mean, in the long run, a form of surviving of inefficient industries which are not and can't be subsidized for ever. The arrears are an inflationary factor of first hand, and the companies which survive on their account contribute to resources dissipation and therefore, to maintaining a suboptimal structure of the economic growth. The fact that arrears stagnates for several years around 40% of GDP shows that the problem's magnitude is still endangering, that the resources expenditure in economy is quantity significant.

Second of all, we should broach and solve the problem of changing the actual pension system (that presumes that the present generation bears pensions of the preceding one, and the next generation will bear the allowance of this generation today), into a system based on funds accumulation, when everybody's pension underlies on their own contribution. The process is, no doubt about it, difficult, long and expensive, but the postponement of adopting some decisions in this matter only makes, in the long run, deepen the gap which will then be covered by public funds.

Once the new pension plan is adopted and if it turns to be accurate and proficiently administrated, it generates important investment funds, which contributes to a better intern accumulation and to development acceleration. Besides, world wide, countries adopting private pension plans, solidly organized and administered (Chile's case), registered spectacular economic growths, based, most of it, on financial resources supplied by this funds. In Romania's case, we shouldn't oversight the failure of very important private investments funds. This negative experience from the past coerces the authorities to maximum attention and to adequate settlement of the new pension plan.

Third of all, our county will have to assign substantial amounts of money for infrastructure projects to bring us closed to EU standards. Road netting, drainage and water supply, urban systems of heating and ecologic matter projects are only few of the directions to which financial and material resources must be nominated which puts together the value of GDP for now and for the next years. Is unrealistic to assume that these resources will be identified entirely inside the public budget, even though there will probably be an important transfer from EU budget with these destinations. Is to be expected that this kind of programs will accelerate economic growth, but at the same time, to put pressure on the public expenditures in a manner which would jeopardize preserving of the fiscal deficit and of public debt store in parameters which seem comfortable at this very moment. In order to realize the equilibrium between the nominal convergence and the real one, the public deficit should diminish uninterrupted, for the sake of unleashing non-inflationary resources given the budget (co)financing of major infrastructure projects.

A suggestion is that of applying a fiscal policy more cathartic, which allows for the close future budget deficits superior to three percent border of GDP stipulated in the Maastricht Treaty. The motivation stands on the necessity of numerous budget expenditures linked to EU and NATO adhesion, on pension system's reform and the expanding of the infrastructure (given the example of some Central and Eastern European countries which took the same path). Although motivated by the real convergence, such an approach involves many risks. Therefore, if for the other countries, with high budget deficits, inflation was brought under control, this was the case for us and one of the main anti-inflationary policies stands in controlling the budget deficit. If it would put aside the small deficits, Romania could be bereaved of one of the few trumps it has and which qualifies her for subsequently adopting euro. Plus, we shouldn't overlook the high cvasi-fiscal deficits: only after their significant and sustainable reduction we can talk about the problem of budget deficit growth.

Settling for 2012 the entrance in ERM II must be considered and threaded as a supplementary opportunity for real convergence and not at all a breathing which would allow the premature relaxation of the fallowed macroeconomic policies.

7. Conclusions

The preconditions for formal inflation targeting impose an operating frame more rigid for the monetary policy and leave less space for central bank's maneuver. Constantly missing the settled inflation targets, the central bank's liability has to suffer sooner or later and this can affect on long term. In the enlightenment of previous mentioned difficulties, it becomes obvious why some central banks adopted a gentle form of inflation targeting regime. An alternative for these banks would be adopting ECB vision, which would not only help them straiten their liability so difficultly earned, but would also allow them the so much wanted flexibility in implementing monetary policy.

Participating to ERM II mechanism is imminent, if changes no happen in Maastricht criteria, which is likely improbable. Therefore, changing the system of inflation targeting to on of exchange rate targeting is unlikely. Central banks in CEE countries invested many resources in trying to implement an inflation targeting and influencing public's inflationary expectations.

An important lesion that we take from transition countries experience with inflation targeting regime is that economic performance will enhance and so the support given to central banks if they avoid over-targeting and sub-targeting of the inflation. Sub-targeting the inflation determined serious economic collapses and weakened this way the crutch for central bank in Poland and Czech Republic as well. Economic performance can also be reached if central banks in these countries won't involve in active manipulating of the exchange rate. This is no current matter in Poland and Czech Republic, but is still an issue in Hungary. A difficult problem in inflation targeting in transition economies is that of the often tempestuous relation between central bank and government. Moreover, having technocrats at the steer of central bank, better then having politicians, may back up the depersonalization of monetary policy management and the growing of support for central bank's independence.

We can conclude that after the EU adhesion, monetary inflation can and will be the main pylon of monetary strategy in these CEE countries till their entrance in euro area. In addition, an important advantage of the inflation targeting regime in transition countries is that central banks in these states already learned how to maneuver monetary policy's instruments for achieving inflation objectives. As these central banks will have an important role in choosing the monetary instruments by ECB once they join EMU, their experience operating inside the inflation targeting system will help them play a more active and positive part in ECB deliberations.

If Romania fallows (with a certain delay) the path chosen by the other advanced transition countries (Poland, Hungary, Czech Republic), it will confront with the massive capital inflows in the next period, which would lead to a sudden and real appreciation of the local currency and for the future. In order to stop the exports decline, taking into consideration social aspects, a positive bump would be for the interest rate to decrease. But independent agents and competitiveness have a counter orientation towards low inflation, as necessary in inflation targeting regime.

The strategy of inflation targeting was introduced in Romania after a long period of getting ready. We think that, even though NBR didn't meet all the preconditions for adopting inflation targeting, as the empiric experience at international level shows, the fulfillment of all preconditions is not by itself an impediment for the success of inflation targeting in Romania. The issues that Romania is confronting with in implementing inflation targeting are, generally, commune issues for transition countries which adopted inflation targeting.

Among the most important challenges in applying the strategy in Romania we identified the problem of persistence of demand excess in economy which creates inflationary pressures and dilemmas consisting with finding the equilibrium between reaching the nominal convergence and assuring in the same time, real convergence in economy. Therefore, the liberalization of capital inflows in the conditions of a high differential between inside interest and outside one will continue to create constrains in using interest rate in monetary policy because of the risk represented by the speculation capital inflows. The persistence of a high level of current account deficit makes the main macroeconomic challenge for the exchange rate, endangering the disinflation process if the extern disequilibrium will deepen continuously and/or if the deficit's financing throughout direct foreign investments will diminish.

Because of existing constrains in using classic monetary policy instruments by NBR, in the last years had be used administrative instruments like limiting credits throughout increasing minimum compulsive reserves or other administrative measures. We find this as being inefficient in limiting the increase of crediting and, therefore, in reducing the excess of demand in economy. As other countries experience showed, administrative measures in limiting credits had no success in adjusting the economic unbalances, especially in decreasing the extern deficit.

Regarding the perspectives of inflation targeting in Romania, after the entrance inside Exchange Rates Mechanism, we consider that a strict form of inflation targeting won't be sufficient as is not assuring the stability of exchange rate. It will be necessary, in our opinion, a more flexible inflation targeting which would allow the targeting of inflation differential predictions towards euro area and a stability of the exchange rate.

Harmonizing the two processes – real convergence and nominal convergence – represents the solution of stimulating the catching-up process inside of an economy. Romania registers a slow rhythm of convergence of monetary variables (inflation rate and interest rate) according to the settled Maastricht criteria; the public finance's situation is favorable when it comes to these criteria, but handles a low degree of economy update. The process of real convergence could be this way stimulated by the lack of a flexible offer. This is the consequence of structural reform in economy, enhance of economic stimulants, as well as attracting outside capital. Such a plead for policies of stimulating the offer is justified by the perspective of adhering to euro area (and therefore to ERM II); the authorities will only dispose of fiscal policy instruments, but their usage will be constrain of Growing and Stability Pact outlines. Another justification is that of increasing the level of integration of the markets, which imposes the growth of competitiveness of home companies in order to assure the sustainability of economic growth process.

For rapid achieving of real convergence between Romania and EU-25 countries, we recommend that Romania would spur labor productivity increase in a higher rhythm of that of euro currency rate appreciation, in order to increase competitiveness, exports and enhancing the situation of balance of trade.

We shouldn't be exclusive concerned by the development of infrastructure in finding the path to a more advantageous integration. If infrastructure development is not empowered by reduction of transaction costs (stable and coherent legislative frame, institutional consolidation, stimulating business environment, liberalization of inflows and outflows on/ of the market, primary infrastructure development, externalities reduction etc.), Romania will become the land of networks of distribution and marketing (disentangling market, only). A redistribution of funds will take place in favor of rich regions (which attracts the majority of conceptualizing activities, with important percentage in production costs).

For the cohesion policy to be efficient funds should be given to regions near the centre (as concentric circles, in wages, towards exterior). Experience shows that isolated areas beneficiary of structural funds only had temporary growths (like Andalusia, Eastern German lands, etc.). Therefore, current patterns must be reformed when it comes to increasing the cohesion of European pattern. Less developed regions should be supported in generating positive technological externalities (to force into the system centripetal strong powers).

A mix of measurements should be applied in order to attract direct foreign investments (for professionals there is no paradox the association of the most reduced timetable cost of labor force between EU countries – members of candidates – and the decreasing of FDI with 40% in first four months of 2005 relative to the same period of the previous year). The error consists in lack vision over adopted measures, which many times were unilateral. Specializing in high rates added value fields is needed, fields which should be interconnected (specializing in many different fields is no indicated), so that the competitiveness of Romanian's economy enhance. Therefore, is worth conceiving a strategy for keeping highly qualified workers throughout: establishment of technological parks, industrial parks; stimulating establishment of educational partnerships – research – business incubators; fiscal stimulation to companies operating in high technology fields.

Is inefficient the predominant usage of exchange rates for increasing the extern commerce efficiency, because it only leads to keeping the economic non-homogenous structure with European Union. Structural measures of growing the exports competitiveness are needed: specializing in fields with high added value, growing quality, reducing prices, publicity, taking part to international expositional fairs. On the other hand, using the instrument of exchange rate may constitute a barrier for getting out the market.

The growth of intra-industrial commerce percentage is necessary and will generate a higher symmetry of socks given the adopting of unique currency; correlation of business cycles will reduce the shortages of commune monetary policy. Plus, development of services sector will contribute to higher structural convergence. Conceiving a strategy of growing the similarity of Romanian exports structure with EU would lead to stimulation of scale economies and productivity growth. If this strategy is combined with exports diversification, positive effects over real convergence would be generated.

Romania has to reduce the percentage of social securities contribution hereby stimulate creation of new jobs. On the other hand, promotion of efficient industrial policy is imposing so that technological externalities are being created. Romanian authorities must stimulate small and medium size companies sector. We find necessary the conception of capitalizing programs for firms throughout the tuning of grace period of three years concerning payment of reinvested profit tax (starting with the first year of firms declaring profit).

Less but not least, fundament energy's strategy which assures a higher independence for Romania represents another desiderates. Decreasing the power of annuity pursuers from energy fields and creating an energy market would be first steps taken into consideration.

Right chosen of central parity when entering Exchange Rates Mechanism and afterwards adoption of euro by Romania is conditioned by knowing the equilibrium exchange rate and establishing the central parity as close to this as possible in order not to jeopardize the process of real and nominal convergence.

Appendix No.1

The share of the current account deficit and the public debt of GDP (%)

Current Account (of GDP)	Czech Republic	Hungary	Poland	Romania	Slovakia
2003	-6,2	-8,7	-2,2	-6,2	-0,9
2004	-5,2	-8,9	-4,3	-7,6	-3,6
2005	-2,1	-6,2	-1,7	-9,1	-8,7
2006	-2,9	-7,3	-2,1	-10,6	-6,4
2007*	-2,0	-6,3	-1,9	-8,8	-3,9
2008**	-2,0	-5,6	-1,5	-8,5	-3,3
Public debt (of GDP)	**Czech Republic**	**Hungary**	**Poland**	**Romania**	**Slovakia**
2003	27,7	59,1	50,1	24,4	42,6
2004	27,8	60,7	48,8	23,7	41,6
2005	28,4	59,9	52,3	20,5	34,5
2006	28,9	59,9	54,4	21,0	38,4
2007*	29,8	59,9	56,7	22,5	37,8
2008**	29,9	59,9	57,1	23,1	37,9

* partial data; ** estimated data

Source: International Monetary Fund's Countries Reports, OECD Country Economic Surveys.

Unemployment rate and the government balance share of GDP (%)

Unemployment rate (%)	Czech Republic	Hungary	Poland	Romania	Slovakia
2003	7,8	5,9	20,0	7,2	17,5
2004	8,3	6,2	19,1	6,2	18,1
2005	8,0	7,3	17,7	5,9	16,2
2006	7,3	7,5	14,3	5,2	13,5
2007*	6,8	7,7	12,6	5,1	12,2
2008**	6,3	7,6	11,3	4,9	11,7

Unemployment rate and the government balance share of GDP (%). Continued

Government balance (of GDP)	Czech Republic	Hungary	Poland	Romania	Slovakia
2003	-6,0	-7,2	-5,9	-2,4	-3,7
2004	-4,0	-5,4	-6,1	-1,2	-4,3
2005	-3,6	-7,8	-4,3	-0,8	-3,1
2006	-3,7	-10,1	-4,1	-1,7	-3,7
2007*	-4,1	-6,9	-3,8	-2,8	-2,7
2008**	-4,3	-6,7	-3,6	-3,0	-2,2

* partial data; ** estimated data

Source: International Monetary Fund's Countries Reports, OECD Country Economic Surveys.

References

[1] Bernanke, Ben S, Laubach, Thomas, Mishkin, Frederic, S. and Posen, Adam S. (1999), "*Inflation Targeting: Lessons from the International Experience"*, Princetown, NJ, Princetown University Press.

[2] Buiter, W., Grafe, C., (2002), „*Anchor, Float or Abandon Ship: Exchange Rate Regimes for Accession Countries*", www.repec.org

[3] Carlstrom, C. and Fuerst, T. S. (2002), "Monetary Policy Rules and Stability: Inflation Targeting versus Price Level Targeting", *Economic Commentary*, Federal Reserve Bank of Cleveland, Feb.

[4] Cerna S., (2006), "*Romania and EU*", presentation made at the Faculty of Economic Sciences, West University, Timisoara.

[5] Ciupagea, C., Atanasiu, I., Câmpeanu, V., Țurlea, G., (2006), „Analysis and evaluation of the development in the economic, social, legislative and institutional area for the new member states of the EU", *Impact Studies* **III**, no. 9, European Institute of Romania, www.ier.ro

[6] Codîrlaşu, A, Chideşciuc, Al., (2004), „Estimating Balassa-Samuelson Effect in Romania (during 1995-2003)", paper presented at the *Conference „Young Economists"*, organized by the National Bank of Romania, September 2004, available on www.bnro.ro.

[7] Daianu, D., Lungu, L. and Vranceanu R.. (2004), "Romania's monetary institutions and policy: meeting the EU challenge", *Pre-accession impact study,* **2**, The European Institute of Romania.

[8] Daianu D., Lungu L., (2007), „Inflation Targeting, between Rhetoric and Reality. The Case of Transition Economies", *The European Journal of Comparative Economics*, vol. 4, nr. 1.

[9] David C., (2006), „Statistic Analysis of the efforts and effects of Romania's accession to the EU", *Doctoral Paper at the Academy of Economic Studies*, Bucharest, Faculty of Cybernetic and Economic Statistics, available on the site www.biblioteca.ase.ro

[10] Dinu, M., Socol, C., Marinaş, M., (2005), "*European Accession Model"*, Economic Publishing House, Bucharest.

[11] Dinu, M., Socol, C., Marinaş, M., (2005), "*Convergence and Cohesion Mechanisms"*, Economic Publishing House, Bucharest.

[12] Dinu, M., Socol, C., Niculescu, A., (2006), "*Coordinating the economic policies into the European Union"*, Economic Publishing House, Bucharest.

[13] Dinu, M., Socol, C., Niculescu, A., (2005), "*Romanian Economy. A vision on the post-communist transition*", Economic Publishing House, Bucharest.

[14] Dinu M., Socol C., (2006), "*Romania in the EU: Convergence potential. A drawn of getting out of the periphery",* General Romanian Economists Association, Bucharest.

[15] Disyatat, P. and G. Galati, (2005), "The Effectiveness of Foreign Exchange Intervention in Emerging Market Countries: Evidence from the Czech Koruna", *BIS Working Papers* 172.

[16] Duenwald C., Gueorguiev N., Schaeter A., (2005), "Too Much of a Good Thing? Credit Booms in Transition Economies: The Cases of Bulgaria, Romania, and Ukraine", *IMF Working Paper* **05/128**, (Washington: International Monetary Fund).

[17] Dumitru I., (2006), „The NBR's monetary policy in the disinflation process from the European integration view", *Doctoral Paper at the Academy of Economic Studies, Bucharest, Faculty of Finance*, available on the site www.biblioteca.ase.ro

[18] Fraga, Arminio, Ilan Goldfajn and André Minella, (2003), "Inflation targeting in emerging market economies", *NBER Working Paper* **10019**.

[19] Gerlach, S., Svensson, LEO, (2003), "Money and inflation in the euro area: a case for monetary indicators?", *Journal of Monetary Economics* **50**, 1649-1672.

[20] Halpern L., Wyplosz C., (2002), "*Catching-Up: The Role of Demand, Supply and Regulated Price Effects on Real Exchange Rates for Four Accession Countries*", ONB Focus on Transition vol. 2.

[21] Hanush, H., Balzat, M., (2004), *"A new era in the dynamic of European integration"*.

[22] Hilbers P., Otker-Robe I., Pazarbasioglu C. Johnsen G., (2005), "Assessing and Managing Rapid Credit Growth and the Role of Supervisory and Prudential Policies", *IMF Working Paper* **05/151.**

[23] IMF WEO, September 2005, Chapter 4, *Does Inflation Targeting Work in Emerging Markets?*

[24] *IMF Country Report*, Czech Republic, No. 04/266, August 2004, www.imf.org

[25] *IMF Country Report*, Hungary, No. 04/145, April 2004, www.imf.org

[26] *IMF Country Report*, Poland, No. 04/217, July 2004, www.imf.org

[27] Iorga E., Salater W., Pitulea G., Cazacu C., Oancea C., (2004), „*Inflation targeting in Czech Republic, Poland and Hungary. Implementation and performances",* Studies and analysis, National Bank of Romania.

[28] Isărescu M., (2007), „*The path to euro",* Conference organized by the Academic College of the Babeş-Bolyai University, March 2007.

[29] Jankovics L., (2003), "Roadmap for the Adoption of the Euro in Hungary: Dangers and Opportunities", paper presented at the *Conference "Institutional and Policy Diversity – Its role in Economic Development*", Debrecen, Hungary.

[30] Jonas J., Mishkin F., (2003), „Inflation Targeting in Transition Countries: Experience and Prospects*", NBER Working Paper Series*, W.P. nr. 9667.

[31] Justice G., Paliu A., (2006), *"Real convergence perspectives",* available on www.fmi.ro

[32] Kemme, David M. and Gavin, William T. (2004), "Measuring the Monetary Transmission Mechanism in Transition Economies*". Paper presented to the 8-th EACES Conference in Belgrade*, September 2004.

[33] Marinaş M.C., (2006), "Correlation analysis between the real and nominal convergence. Romanian Case*", Supplement of the Theoretical and Applied Economics Review*, Bucharest.

[34] Marinaş M.C., (2007), „*Economic Convergence in the EU"*, Doctoral Paper at the Academy of Economic Studies, Bucharest, Faculty of General Economics, available on the site www.biblioteca.ase.ro

[35] Mishkin, Frederic S., Schmidt-Hebbel Klaus (2001), "One decade of inflation targeting in the world: what do we know and what do we need to know?" *NBER Working Paper* **8397**.

[36] Mourmouras I., Arghyrou G., (1999),"Monetary Policy at the European Periphery: Greek Experience and Lessons from Transition Economies", *CERT Discussion Paper*, **99/10.**

[37] Padoa-Schioppa T., (2004), "*The Pre-accession Phase: a resume of the Past Years of Co-operation and Policy Dialogue*", presentation made at the *Seminar on the accession process*, Paris.

[38] Papadia F., (2004), Introduction Speech at the "Monetary and Exchange Rate Policies and the Practical Functioning of the ERM II" Session*, Seminar on the accession process,* Paris.

[39] Socol C., Niculescu A., (2005), "Romanian policy dilemmas regarding the accession to the EU"*, Supplement of the Theoretical and Applied Economics Review*, Bucharest.

[40] Stone Mark, (2003), "Inflation Targeting Lite", *IMF Working Paper* **WP/03/12.**

[41] Svensson Lars, (2000), "*The first year of the Eurosystem: inflation targeting or not*", *NBER Working Paper* **7598**.

[42] Szapary G., (2002), "*Is Maastricht Too Much for the Candidate Countries*?", presentation at the Round Table "Euro 50 Group", Bruxelles.

[43] Truman Edwin M., (2003), "*Inflation Targeting in the World Economy"*, Washington DC: Institute for International Economics.

[44] Ungureanu G., (2006), *"Monetary policy in the transition period. Study for Romania's case",* Doctoral Paper at the Academy of Economic Studies, Bucharest, Faculty of Economics, available on the site www.biblioteca.ase.ro

[45] Wyplosz C., (2002), "*The Path to the Euro for Enlargement Countries*", Briefing Notes to the Committee for Economic Affairs of the European Parliament.

[46] Zanghieri P., (2004), "*Current Account Dynamics in new EU members: Sustainability and Policy Issues*", CEPII, No. 2004.

[47] *** "*Macroeconomic environment of the economic policy",* Chapter 2, Pre/accession Economic Program, National Forecasting Commission, 2005.

[48] *** "*Catching-up, growth and convergence of the new member states"*, European Commission, March 2005

[49] *** "*CEE Countries Report, Nr. 1/2006"*, Bank Austria Creditanstalt.

[50] *** "*Report over the Government's Activity. Forecasting and Analysis Report",* Romanian Academic Society, September 2005.

[51] *** "*Post-accession Romanian Strategy",* Bucharest, 2007.

[52] *** "*EUROSTAT Statistics", 2007.*

[53] *** “*Real Convergence in Candidate Countries-Past Performance and Scenarios in the Pre-Accession Economic Programs*”, European Commission, Directorate General for Economic and Financial Affairs, 2001.

In: Emerging Topics in Macroeconomics
Editor: Richard O. Bailly, pp. 183-196
ISBN: 978-1-60692-093-0

Chapter 6

DYNAMICS OF COMPLEX ECONOMIC SYSTEMS: THE CASE OF ROMANIAN ECONOMY

Cornelia Scutaru[1]
Institute for Economic Forecasting, Romanian Academy, and
Academy of Romanian Scientist, Bucharest

"A competitive economy might be seen as a huge natural computing machine that incessantly mills solutions for a never ending row of quantitative problems with which it is automatically fed. It distributes labor, capital and natural resources among the different production sectors. ... the competitive mechanism produces an adjustment of demand and supply by successive approximations ... Each economic system ... has a complex internal structure. Its results are determined by the reciprocal relationships among its component parts..."

Wassily Leontief[1]

Abstract

The economy is a complex system, the dynamics of which can be described by the evolution of the interactions between the actors of the economic process. A measure of the complexity of an economic system must reflect the level of the interdependencies between the component parts of the system. The Input-Output (I-O) tables put into evidence the interconnections between branches as the component parts of the economic system. For the measure of the complexity of the economic system, as it is described by the I-O tables, some different indicators where been proposed, but the most adequate for our approach is the **complexity indicator** which takes in consideration the **network effect** (more connections - more complexity) and the **(inter)dependency effect** (which measures the degree in which the behavior of each component of the system is determined either by the internal connections or by the relations with other parts of the economic system). The multipliers of the I-O tables measures the effects determined by the branches interdependencies when a change of occur in any of the parts of the economic system. In an I-O analysis many different multipliers where been defined, adequate for various analytical instances.

The paper's scope is to measure the **complexity** of Romanian economic system determined by the branches interdependencies and to envisage the phenomenon evolution. The **novelty** consist in this approach of the dynamic and complexity of the economic system

---

[1] E-mail address: corneliascutaru@yahoo.com

in a time of great fluctuations of economic indicators (lack of balance and critical points), using concepts and models from physics, linked to the emergence of new properties inside complex systems (differing from system's synergy).

Introduction

Economy is a complex system, whose dynamics is very sensitive both to the impact of **random shocks**, with destabilizing consequences, and to the impact of **public policies** (fiscal, monetary, etc.) destined to play a role of economic regulation. One of the random shocks was the sudden and uncontrollable start of the "transition to the market economy": all the countries that passed from the centrally-planned to the market economy went through an interval of economic recession; in the Romanian economy such a phenomenon had two phases: 1989-1992 and 1997-1999, respectively. As regards the impact of the macroeconomic policies, there is a strong interaction of the political instruments used by the public authorities, and such interactions have to be considered in order not to turn into an economic "shock" the effects of a certain political measure. During the pre-election periods, one may notice public policies mixes that have contradictorily impacts: the so-called "populist" policies of increasing wages, with inflationary effects.

The complexity of the economic system rests in the interaction of its parts (participants to the economic process), which leads to the emergence of new features, different from the features of any of the component parts. It may be approached by two ways: at mezzo level, on the basis of the interdependency relationships among the industries in the Input-Output tables (see **Amarral** *et al*., 2007[2]) and at microeconomic level, using the **multi-agent models** (see **Solomon**, 2007[3], **Yaari**, 2006[4], **Stanley** *et al*., 2002[5]).

Stability of the economic systems to shocks results from the combined action of two phenomena: micro/macro aggregation and macro/micro transmission. The aggregation phenomena from the microeconomic level (economic agents) to the macroeconomic one (economic system) lead to the formation of economic clusters[6], whose stability is determined by their inner complexity. We talk about the formation of complex system of production, purchase/distribution, including also centers of research-development and of labor formation/training (see the paper of Yaari, 2006, which reveals clusterization phenomena in Poland over the transition period). As regards the macro/micro transmission, it refers to the macroeconomic policies impacts upon the microeconomic level; over the "transition" period perverse effects of some public policies that have not correctly assessed the interactions among the different instruments used have occurred.

A measure of the **complexity** of an economic system must reflect the level of interdependencies among its component parts. Through their construction, the Input-Output tables show the flows of goods and services among the different sectors and industries of the national economy, thus revealing the connections among the industries/activities considered as parts of the economic system. In order to measure the **complexity** of the system as described by the Input-Output tables, different indicators have been proposed, but for our approach the most adequate is the complexity indicator proposed by **Lopes** *et al*., 2007, which takes into account the **network effect** (more connections – increased complexity) and the **(inter)dependency effect** (which measures the extent to which the behavior of each component of the system is determined either by the internal connections, or by the relationships with other parts of the economic system).

A **network** is a simplified mathematical representation of the functioning of a system. In many disciplines (informatics, economics, biology, and sociology) are used models based on networks; such models are powerful instruments for the evaluation of the dynamics of the Input-Output complex systems (**Moseley,** 2003[7]). **Mauboussin, J.** *et al.* (2000[8]) reveal several concepts specific to the network type of models: **the network effect** – namely the amplification of the phenomena due to the increase in the number of the participants to the network; the **critical mass** – the diffusion process within a network reaches a certain level (known as "critical mass") beyond which the process becomes self-supporting; **the inter(dependency) effect** – the behavior of a part of the system is determined both by the internal relationships and by the inter-relationships with other parts of the system (see **Amaral** *et al*., 2007).

Using the relationships proposed by **Lopes** *et al*. (2007) for the Input-Output tables, we shall analyze several of the complexity indicators for the Romanian economy. The Input-Output tables that are currently available cover the interval 1989-2004 and 105 sectors and industries, using the same methodology. In order to characterize the evolution of the complexity of the economic system, a comparison is drawn between the results obtained for 1989 (the last year of centrally-planned economy) and those obtained for 2004. The main conclusion of the paper points towards a significant increase in the economic system complexity in 2004 as against 1989. The evolution of the sectors and industries groups reveals specific features.

1. Short Presentation of the Economic Developments over the Interval 1989-2004

In 1989[9], the Romanian economy was centrally-planned and exhibited certain specific features that had a crucial impact upon the developments occurred in the period that has followed. With 5.5% economic decline in 1989, the last years of centrally-planned economy were years of deep recession for Romania. At that moment, Romania was a quasi-closed economy, with imports restricted to the minimum and the exports predominantly directed towards the CMEA market. The foreign debt was paid almost entirely, and the domestic consumption was severely restricted. The shares of the sectors in the economy – centrally-directed – did not reflect the potential of available resources, and the economic competitiveness was something of a theoretical notion. The Romanian economy produced on a large scale non-saleable inventories and the tendency to produce "for inventory" was maintained also after 1989. A hidden inflation was present, "masked" by the low consumption supply. There was no unemployment, but the labor productivity level clearly indicated an unemployment hidden through centralized decisions that ensured a compulsory job position for each school and university graduate.

The social and economic system transition led in Romania – as in the other Central and East-European countries – to an initial decline in the overall economic activity. After the first negative impact, against the multiple structural changes aimed at setting a new relationships system, an economic recovery occurred, which in the case of Romania gave up quickly under the pressure of social claims (1997-1999). The transition to the market economy has meant in fact a deep economic, social and institutional crisis. The economic transition crisis continued to influence the economic and social developments towards the market economy, despite all the efforts aiming at macroeconomic stabilization and at creation of a stable and sound economic environment.

The economic opening in the first years of transition has found the Romanian economy structurally and competitively unprepared. Big adaptation problems occurred, both on the domestic and on the international markets. The sudden increase in domestic demand against a low supply led to inflation and increase in consumption imports. The vanishing of the CMEA market forced orientation towards the European market and finding of new markets, from which stemmed the needs to increase products competitiveness and to change the production structures. The former requires investments in new technologies, in research-development, as well as high labor productivity. Two trends specific to the transition crisis revealed at this point: neglect of investments and maximum exploitation of resources in order to maximize profits, and structural reforms accompanied by massive layoffs, which led to increase in unemployment (diminished in the previous years due to economic growth, but also due to the migration of labor to the EU countries).

The centrally-planned countries have started the transition with a labor structure significantly different from that of the European Union countries, the main difference concerning the number of persons employed in agriculture. Overall, the services sector, and especially the trade, recorded the highest labor deficit, around 15% in Romania as compared to the level of the European countries. Romania was one of the countries with the slowest pace of implementing the reforms needed for creating a competitive environment, a fact that generated the inability of economy to create jobs even in the growing sectors. That has led to the preservation of strong distortions regarding labor allocation and contribution of different economic sectors to gross domestic product, as compared to the European Union countries.

As a conclusion, in 1989 the economy was weakly capitalized, with underdeveloped financial and capital markets – an economy that has preserved functioning for a long time the huge industrial enterprises of the centrally-planned era whose inefficiency was well known, but whose privatization raised and still raises large social problems.

Until 2004, the Romanian economy has evolved towards an advanced stage of establishing a functional market economy. The investing climate has improved, the macroeconomic environment stability has increased, the privatization and restructuring processes have continued. Other favoring factor was the preservation of a low budget deficit with all its influences, including the impact upon the inflationary expectations. Despite all the oscillations of transition, one may say that in 2004 the economy capitalization process has continued.

Gradually, prudent budget policies were adopted, which correlated with tempered credit policies through maintaining interest rate at a high level have led to a moderate expansion of domestic consumption. The structural funds coming from the European Union were directed towards development of gross capital formation. The increase in investments in real terms, together with the contribution of foreign capital inflows ensured an increase in the fixed assets, although the depreciation rate was oscillatory. As a result of the increase in the propensity for savings (nevertheless, slowed down due to the interest rates dynamics) and in the foreign direct investments, the financial resources for investments have also increased. The Romanian economy needs to create/support domestic investments and to attract foreign investors, the foreign investors being not only a source of capital, but one of technology as well, the only ones able to promote economic growth. The advantages of Romania for the foreign investors were provided by the labor force features – high education level and low cost, especially in industry and services.

The role of the macroeconomic policies is especially important in the labor force area: the increase in employment is determined not only by the sectors' economic recovery on the basis of stimulating investment policies, but also by other types of measures, such as those regarding retirement, unemployment benefits or professional reorientation and reconversion.

Over the transition period, Romania has approached a gradual system of reforms. Thus, early retirement was preferred over unemployment and orientation of the redundant labor force towards agriculture was encouraged, mainly due to the lack of opportunities in other sectors. Such an approach of reforms has led to the net migration of population from urban to rural areas, a phenomenon unknown to other countries, which has also led to the increase in agricultural employment. The fact that the declining sectors have lost jobs and the other sectors have not created sufficient new jobs has led to massive increases in unemployment, as well as the increase in agricultural employment as a last resort alternative over unemployment. Consequently, one may notice over the transition period an increase and not a decrease in agricultural employment. Such an increase has led to a delay in the convergence of labor employment structure of Romania towards one similar to the European Union countries.

There is a permanent pressure from the part of economic agents and trade unions towards increasing the nominal income; the interest rates dynamics has encouraged the increase in consumption loans, which has led to a significant increase in the domestic demand. In order to meet such a demand, an increase in imports has occurred, which has determined the worsening of the balance of payments.

One may say that despite the impact of transition to the market economy (changes in the sectoral structure of economy, in the relative prices, in technologies, etc.) the long term stability relationships within the economy are preserved, as in a consolidated market system.

An analysis of the structural changes occurred in the Romanian economy was done by Scutaru and Florescu (2003)[10] for the year 1999. The paper analyzed the groups of exporting industries and of industries producing mainly for the domestic markets (intermediate consumption, final consumption and gross fixed capital formation) from the point of view of M. Porter's methodology: general conditions, demand conditions, supporting industries, governmental policies. A first assessment of the disparities within the Romanian transition economy has pointed to the conclusion that the economy has not reacted as a whole: significant structural changes have occurred after 1989, reflected in the different positioning of sectors and industries within the overall economic framework. The exporting industries structure has changed due to the changes on the foreign markets, but also due to the change in the competitiveness ratios among industries. The domestic consumption structure has also changed, and a decline in the industries ensuring gross capital formation and a significant decline in the industries that use the R&D results have been revealed. The main economic policy measures should take into account such changes and should act for improving some of them, while preserving others.

2. Methodological Approach

The problem we attempted to analyze might be written as follows:

- What has happened between 1989 and 2004 to the complexity of the Romanian economy? Has it increased or decreased due to the structural changes occurred over the transition period?
- How it has evolved the degree of autonomy/dependency of groups of industries (sectors) of the Romanian economy? Has it increased or decreased? For all the industries or only for some of them? Which might be the explanation of such a phenomenon?

In order to assess the complexity of the Romanian economy, we have used the Input-Output tables for 105 industries and sectors. We have built the A=(aij) matrix of the input-output technical coefficients, which showed the quantity of output of sector i absorbed by sector j as ratio to the unit of total output j. We have organized the data by blocks of industries with common features (see Appendix 1) and computed the indicators that allowed for assessment of the Romanian economy degree of complexity in the two base years: 1989, as the last year of centrally-planned economy, and 2004, as the last year for which there are available data regarding the Input-Output tables. Different from the cited authors, we have assessed the dependency degree of each block in order to make the comparison between the base years.

We have adapted the methodology developed by Lopes *et al*. (2007) to the above-formulated problem (see Appendix 2) and assessed:

- **The dependency effect**, which allows for assessing how much depends the behavior of a block upon the internal connections and how much upon the relationships with the other parts of the system. It is complementary to the effect of block autonomy as against the other industries: the highest the block autonomy, the less it depends upon which it would receive from the other industries (either it has no exchange relationships with them, or the exchanges are at a very low level). The degrees of dependency and autonomy were computed for the 17 groups of industries identified in the functioning of the Romanian economy.
- **The network effect**, assessed through two indicators:
    - **the number of interconnections** in 1989 and in 2004, respectively, expressed through the number of a_{ij} coefficients different from zero; in order to assess this indicator the intensity of inter-industry exchanges was not taken into account (the value/level/quantity of commodities provided or received by the respective industries, but only the existence of such an exchange – the "zeroes" in the matrix are thus eliminated);
    - the **network effect indicator**, computed on the basis of relationship formulated by **Lopes** *et al*. (2007) as according to the principle: the highest the connections number, the highest the complexity of the system.

3. Results

We shall start with the second problem of the two above-mentioned ones: How it has evolved the degree of autonomy/dependency of groups of industries (sectors) of the Romanian economy? Has it increased or decreased? For all the industries or only for some of them? Which might be the explanation of such a phenomenon?

For this aim, we have delimited 17 groups of industries (see Appendix 1). It is obvious that the industry groups' delimitation may be done differently, according to certain chosen criteria; the criterion that we have followed in the current paper was the specific of the production activity. Considering the importance of the productive inter-exchanges among the industries, expressed through the intermediate consumptions matrix on which basis the A matrix is computed, that has appeared to us as the most rational criterion for grouping the industries. According to the adopted methodology (see Appendix 2) for each group of industries both the level of inter-exchanges (intermediate consumptions) provided/received from inside the respective group, and the level of the inter-exchanges provided to and received from the other industries were taken into account.

Synthetically, the data are presented in Table 1:

Table 1. Autonomy degree of groups of industries

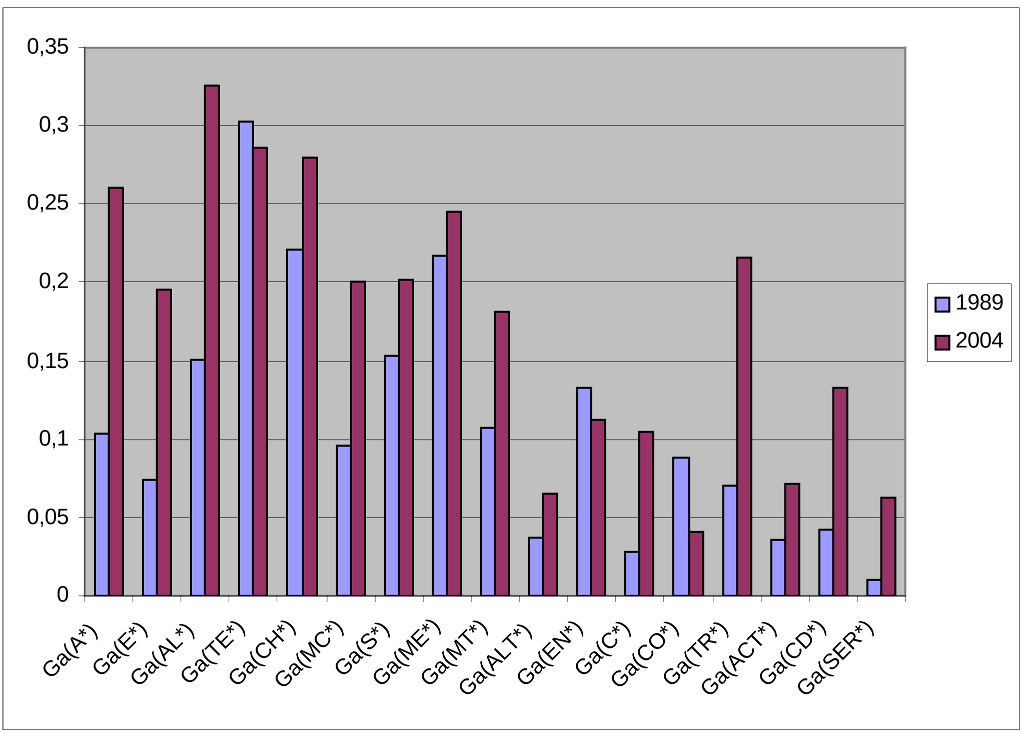
			1989	2004
1	Agricultural industries	Ga(A*)	0.104	0.261
2	Mining and quarrying industries	Ga(E*)	0.075	0.195
3	Food industries	Ga(AL*)	0.150	0.325
4	Textile industries	Ga(TE*)	0.303	0.286
5	Chemical and related industries	Ga(CH*)	0.221	0.280
6	Building materials industries	Ga(MC*)	0.096	0.201
7	Metallurgy	Ga(S*)	0.153	0.202
8	Machinery and equipment	Ga(ME*)	0.217	0.246
9	Transport means	Ga(MT*)	0.107	0.182
10	Other industrial sectors	Ga(ALT*)	0.037	0.065
11	Energy	Ga(EN*)	0.133	0.112
12	Constructions	Ga(C*)	0.028	0.105
13	Trade	Ga(CO*)	0.088	0.041
14	Transports	Ga(TR*)	0.070	0.216
15	Financial activities	Ga(ACT*)	0.036	0.072
16	Research-Development	Ga(CD*)	0.042	0.133
17	Public services	Ga(SER*)	0.011	0.062

As one may see also from the following graph, for most of the industries and sectors an increase in the autonomy degree occurred between 1989 and 2004 and, correspondingly, a decrease in the dependency degree (see Appendix 3). Three groups of industries lost some autonomy: Textile industries (-5.4%), Energy (-15.7%) and Trade (-53.6%). A first explanation is given by the foreign trade structure: the energy and consumption commodities imports have increased significantly, and in what regards the textile industries the OPT system has predominated over the analyzed period, namely production of textile goods based on imported raw materials and processed with domestic labor.

Analyzing by comparison the behavior of groups of industries in what regards the size of the autonomy degree; one may notice that in 1989 a low degree of autonomy (less than 0.100) in the case of the following groups: Public services, Constructions, Financial activities, Other industrial sectors, Research-Development, Transports, Mining and quarrying industries, Trade, Building materials industries. A degree of autonomy ranging between 0.100 and 0.150 was recorded by the groups: Agricultural industries, Transport means, Energy, Food industries, while a degree of autonomy exceeding 0.150 in the case of the groups: Metallurgy, Machinery and equipment, Chemical and related industries, Textile industries.

In 2004, there were less groups with low autonomy degree and more groups with higher autonomy degree: less than 0.100 in the case of the groups Trade, Public services, Other industrial sectors and Financial activities, between 0.100 and 0.150 in that of the groups Constructions, Energy, Research-Development, while the rest of groups recorded higher degrees of autonomy, up to 0.325 (Food industries).

All these mean that the intensity of inter-exchanges within each group has increased as compared to the intensity of the inter-exchanges with the other economic groups. In order to appreciate the evolution of the complexity of the whole system, it is also necessary to assess the network indicators: the number of inter-connections in the two analyzed years and the complexity indicator.



Graph 1. Autonomy degree of groups of industries.

The first formulated problem referred to the complexity degree of the economic system the way it was reflected in the Input-Output tables for 1989 and 2004: What has happened between 1989 and 2004 to the complexity of the Romanian economy? Has it increased or decreased as a consequence of the structural changes occurred during the transition period?"

The system's dependency degree (see Table 2) has decreased to 92.6% in 2004 (together with the corresponding increase in the system's autonomy, as one could have noticed in the analysis of the autonomy degree by groups of industries). In such a case, one may speak about a system where the groups of industries have become more independent over time. An explanation of such a phenomenon is provided also by the increase in the imports for productive consumptions, which has led to the increase in the autonomy towards the domestic production level. Practically, in 1989 the Romanian economy was a quasi-closed system, which imported only what was strictly necessary, while the economic openness and the foreign trade development during the transition period has led not only to an increase in the imports for population consumption, but also in the imports for production consumption, so that one may say that the dependency of industries upon the domestic production has decreased. The phenomenon is much more complex: a foreign trade imbalance has built up, determined, on the one hand, by the increase in the demand for imports, and on the another hand, by a more moderate evolution of exports.

As regards the industries' interdependency degree, it has increased in 2004 by around 40% as compared to 1989. In the technical coefficients matrix an $a_{ij}=0$ signified that either the industry did not receive or it did not provide anything of its output for the consumption of the other industry, namely that there was no dependency of such kind between the two industries. Moreover, two industries are considered as independent one of each other if $a_{ij}=a_{ji}=0$. The decrease in the number of zeroes in the A matrix in 2004 as compared to 1989 reveals an increase in the number of coefficients that differ from zero, namely an increase in the interdependency degree among industries (one of the features of the network effect, as according to the principle: less zeroes – more interconnections, increased interdependency).

Such a thing is confirmed by the increase by near 36% in 2004 as compared to 1989 in the indicator that measures the overall network effect. Since the increase in the network effect was stronger that the decrease in the dependency degree, the economic system is characterized by an **interdependency index** by nearly 26% higher in 2004 as compared to 1989 (Table 2).

Table 2. Complexity indicators of the economic system

	1989	2004	2004/1989
Dependency degree G(A)	0.889965	0.824462	92.6
Number of zeroes in A	5197	3140	60.4
Network effect indicator H(A)	0.524084	0.712454	135.9
Interdependency (complexity) index GA)*H(A)	0.466417	0.587391	125.9

Conclusions:

- Starting in 1989 with a centrally-planned economy and undergoing a transition period towards a functional market economy characterized by many distortions, the Romanian economy system has evolved between 1989 and 2004 towards an increase in complexity;

- Although in 2004 the economic system had a lower dependency degree than in 1989 (increased degree of autonomy) over the analyzed interval the interdependency degree has increased, a phenomenon determined by a stronger network effect;
- Consequently, the increase in the system's complexity stems from the network effect at overall level and not from the dependency effect.

4. New Research Directions

The economy of a country is a complex dynamic system that has both discrete and continuous components, sensitive to the variations of the external and internal parameters. The **complexity** concept involves both structure and order within a system. The dynamic behavior of the complex systems is linked to the modeling of time series and of their informational content. The approach of dynamics and complexity of economic system over a period of large fluctuations in the economic indicators (imbalances and critical points) involves finding and using models and concepts connected with the emergence of the new features (different from the system's synergy).

One of the directions of deepening and detailing the current research is that of finding the distortion points within the evolution of the economic system over the transition period, in order to determine the factors that have influenced its evolution. Considering that the Input-Output tables reflect a network structure, **modeling the dynamics of such a structure and its response to shocks might reveal the intervention capability of the economic policy measures**.

Appendix 1

List of main blocks (groups of industries)

1. **Agricultural industries** comprises the vegetal and animal production industries, fishery, sylviculture, forestry and agricultural services (industries 1-6);
2. **Mining and quarrying industries** comprises all the mining and quarrying industries: coal, petroleum, natural gas, ferrous ores, non-ferrous ores, sand, clay, salt and non-metallic minerals (industries 7-17);
3. **Food industries** comprises food processing industries: meat, fish, dairy, vegetables, fruits, beverages, tobacco (industries 18-27);
4. **Textile industries** comprises textile and leather processing industries, including pulp (industries 28-34);
5. **Chemical and related industries** comprises all the chemical industries, including pesticides and medicines and drugs (industries 35-47);
6. **Building materials industries:** glass, pottery, stones, bricks, cement, etc. (industries 48-54);
7. **Metallurgy:** metallurgical processing, including foundry (industries 55-59);

8. **Machinery and equipment:** metallic constructions, agricultural machinery, tool-machinery, computer hardware, up to electrical and medical appliances (industries 60-71);
9. **Transport means:** for road, ships, for railway, for air flight, including bikes (industries 72-76);
10. **Other industrial sectors** (industries 77-78)
11. **Energy:** production and distribution of electric and thermal power, gas and water (industries 79-82);
12. **Constructions** (industry 83);
13. **Trade,** including hotels, restaurants (industries 84-86);
14. **Transports:** by road, by ship, by railway, by plane, including telecommunications (industries 87-95);
15. **Financial activities,** including real estate (industries 96-97);
16. **Research-development activities,** including informatics and architectural design (industries 98-101);
17. **Public services:** administration, education, health, other public services (industries 102-105).

Appendix 2

Lopes *et al.* **Methodology** (2007)

We shall not present *in extenso* the methodology proposed by **Lopes** *et al*. (2007), which may be found in the cited paper, but we shall attempt a presentation through the means of interpreting upon the technical coefficients matrix (or, equivalently, of the intermediate consumptions) from an Input-Output table.

The economic system is represented through a square matrix, **A**, of order N, with all the values non-negative. A subsystem of order *m* (*m*=1,2,...N-1) of A is a square block $\mathbf{A}^*$ of order *m* that has the main diagonal formed of *m* elements of the main diagonal of A. In an Input-Output matrix, it represents the technical coefficients of a group (block) of industries with similar features (for instance, the agricultural industries, or the services industries – see Appendix 1).

$\mathbf{A}^*$ might be considered a subsystem of **A**. In an Input-Output matrix, such a subsystem comprises the technical coefficients corresponding to the provided technical consumptions (or received, which is equivalent) inside the block; namely the exchanges of commodities among the component industries. Such a subsystem is the more autonomous (or less dependent, which is equivalent) the higher the values of its elements as compared to the intermediate consumptions provided to and received from other industries.

In order to measure how high (or low) is the autonomy of the A* subsystem, we define the **autonomy degree** of A* as ratio of the sum of intermediate consumptions provided/received inside the block to the sum of intermediate consumptions provided and received by the respective block.

$$G_a(A^*) = [A^*]/([A^*]+[A^*]+[A^{***}])$$

where $[A^*]$ measures „sum of intermediate consumptions inside the block”; $[A^{**}]$ measures „sum of intermediate consumptions provided by the respective block to the other industries of the system”, and $[A^{***}]$ measures „sum of intermediate consumptions received by the respective block from the other industries of the system”.

The block’s **dependency degree** is defined as:

$$G_d(A^*) = 1- G_a(A^*)$$

For our analysis, we define the **dependency degree of system A** as being:

$$G^*(A) = (\Sigma_k G_d(A_k^*)) / M$$

where k=1...M, M being the number of blocks taken into account.

By analogy with the above-cited method, we shall define the **network effect indicator** H(A) as follows:

$$H(A) = 1\text{-}h(A)$$

where:

$$h(A) = Z(A) / (2^N\text{-}2)$$

and where Z(A) is the number of zeroes in the A matrix, and 2^N-2 is the maximum possible number of blocks.

Finally, the **complexity index**, which combines the dependency and the network effect, is:

$$I(A) = G(A)*H(A)$$

We have used this measure using the technical coefficients matrix, A = $[a_{ij}]$, for 105 industries, built for the Romanian economy, in years 1989 and 2004.

Appendix 3

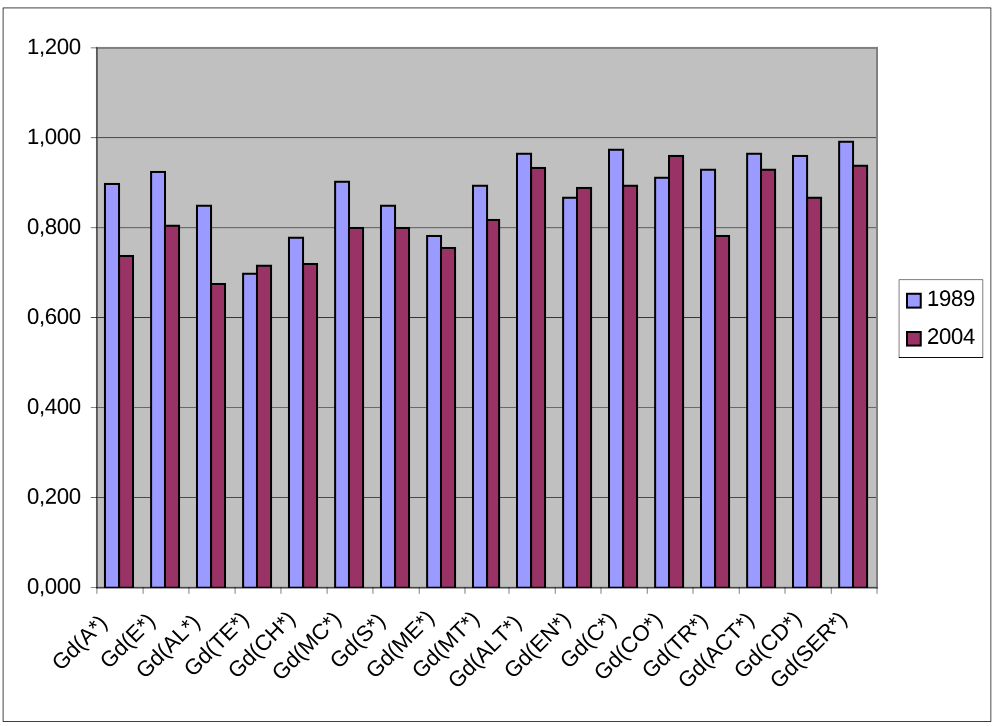
Dependency degree of groups of industries

			1989	**2004**
1	Agricultural industries	Gd(A*)	0.896	0.739
2	Mining and quarrying industries	Gd(E*)	0.925	0.805
3	Food industries	Gd(AL*)	0.850	0.675
4	Textile industries	Gd(TE*)	0.697	0.714

Dependency degree of groups of industries. Continued

			1989	2004
5	Chemical and related industries	Gd(CH*)	0.779	0.720
6	Building materials industries	Gd(MC*)	0.904	0.799
7	Metallurgy	Gd(S*)	0.847	0.798
8	Machinery and equipment	Gd(ME*)	0.783	0.754
9	Transport means	Gd(MT*)	0.893	0.818
10	Other industrial sectors	Gd(ALT*)	0.963	0.935
11	Energy	Gd(EN*)	0.867	0.888
12	Constructions	Gd(C*)	0.972	0.895
13	Trade	Gd(CO*)	0.912	0.959
14	Transports	Gd(TR*)	0.930	0.784
15	Financial activities	Gd(ACT*)	0.964	0.928
16	Research-Development	Gd(CD*)	0.958	0.867
17	Public services	Gd(SER*)	0.989	0.938

Dependency degree of groups of industries



References

[1] Wassily Leontief, "*Input-Output Economics*" New York, Oxford University Press, 1966, p. 33

[2] Amaral, J., Dias and Lopes,J. (2007) „Complexity as Interdependence: A new Measure of Connectedness of Input-Output Systems", *Environment and Planning A,* furthcoming.

[3] Solomon, S., (2007), „Complex Economic Growth Patterns", session at Romanian Academy, may 2007

[4] Yaari, G., (2006), „The AB Model, the GLV, and a Data set from Poland", Presentation of the PhD thesis, under the supervision of Prof, Sorin Solomon, Racah Institute of Physics in the Hebrew University of Jerusalem, Israel; Multi-Agent Systems Division. ISI Torino, Italy

[5] Stanley, H., E., Amaral,L., A., N., Buldyrev, S., V., Gopikrishnan, P., Plerou, V., Salinger, M., A., (2002), „Self-organized complexity in economics and finance" *PNAS*, vol 99, suppl. 1, 2561-2565, febr 2002; http:www.pnas.org/cgi/doi/10.1073/pnas.022582899

[6] In sensul lui Michael Porter (v. *The competitive advantage of Nations* – New York: The Free Press, 1993).

[7] Moseley, N., (2003), "Modeling Economic Time Series Using a Focused Time Lagged FeedForward Neural Network", *Proceedings of Student Research Day, CSIS, Pace University, May 9th, 2003*

[8] Mauboussin, J. Michael, Schay, Alexander, Kawaja G. Stephen, (2000), „Network to Net Worth – Exploring Network Dynamics" Credit Suisse First Boston Corporation http://www.stern.nyu.edu.networks/site.html

[9] Scutaru Cornelia, Florescu Ionel, (2004): Transition to the Market Economy or the Crisis of Capitalism under Formation. The Case of Romania; paper presented at the international workshop „3-ème JUORNEES D'ETUDE DU LAME: Les transformation du capitalisme contemporain: faits et theories. Etat des lieux et perspectives" Univ. Reims – France, 28-30 Mars 2004.

[10] Scutaru Cornelia, Florescu Ionel, (2003): "Global Competitiveness of Sectors - Favoring factors and barriers –" *Proceedings of the Romanian Academy, series C: Humanities and Social Sciences, no. 1-2003, The Publishing House of the Romanian Academy.*

In: Emerging Topics in Macroeconomics
Editor: Richard O. Bailly, pp. 197-218
ISBN: 978-1-60692-093-0

Chapter 7

ADOPTING THE FLAT TAX IN ROMANIA: FROM ORATORY TO EVIDENCE

Cristian Socol, Marinas Marius and Aura Socol
Academy of Economic Studies, Bucharest

Abstract

In recent years, the adoption of flat tax by several countries continues to attract considerable attention, both empirical and theoretical part. This paper evaluates the main effects of the implementation of flat tax system in Romanian economy. We demonstrate, based on the empirical evidence, that their effects in the economy are ambiguous. If accompanying measures are not going to be enforced, the introduction of the flat rate of 16% in Romania will lead to unsustainable budgetary and current account deficits and inflationist pressures.

There are no signals of Laffer-type behavioral responses consolidating budgetary revenue increases from the tax cut. The flat tax favors the workers with big salaries and big and financially solid companies (which, mainly “export” the profit), increasing progressivity and deepening social polarization. It will attack the fragile macroeconomic stability. It is uncertain if it will lead to the increase of the degree of employment, having in view the fact that the contributions to the social insurances have a very high level. The evidence shows that the adoption of flat tax has not resolved the problem of reduction the informal labor market. In the first part of this paper, we conclude that the available data do not to make a relevant distinction between the effects of tax reform and other contextual factors.

In the second part, considering that the fiscal policy can play a key role in fostering real convergence, we propose measures to increase the long-term sustainability of public finances in Romania, by complying with the Stability and Growth Pact. Romania should have chosen to continue what it was confirmed to be a valid element of the economic evolution towards a European standard (progressive fiscal system). On the other side, to optimize the impact of the flat tax reform, we suggest accompanying measures – both fiscal reforms, in particular on the expenditure side, and structural reforms.

Keywords: tax reform, flat tax, public finance sustainability

JEL Classification: E 24, E 62, H 62

1. Introduction

We start our research by comparing the two taxation systems: the progressive one and the one based on the flat tax (Part two). Then, we shall point out that the implementation of the flat tax has been erroneously performed. The moment chosen by the authorities has been an incorrect one, the implementation procedure has been inconsistent and the lack of studies regarding the impact has deprived of consistency the fiscal reform measure, which came into force on January 1st 2005 (Part Three).

The next step in the study is that of pointing out the macro-economic effects generated by introducing the flat tax in Romania (Part Four). Even though there were some other factors with relevant influence on the negative evolution of the macro-economic indicators during the period 2005 – 2008 (after implementing the flat tax), we shall bring arguments referring to the fact that this fiscal reform measure has had as a result the decrease of sustainability of the Romanian public finances. The twin deficits – the budget deficit and the current account deficit – have deepened and the inflation exploded. The proportion of the budgetary revenues in the GDP has very little increased, the signals regarding the existence of a behavior consistent with the Laffer curve, as a result of decrease in taxation, being totally absent. Using the econometrical techniques, we shall demonstrate in our work the fact that the implementation of the flat tax has lowered the sustainability degree of the public finances in Romania, while the structural budget deficit is more and more increased. Regarding the adopting of the flat tax influence on the occupying degree, we may assert that this one is neutral. Moreover, it is uncertain if the introducing of the flat tax has brought to the surface jobs from the underground economy.

The social effects of introducing the flat tax shall be analyzed in Part Five. The adoption of the flat tax has increased the income inequity, thus increasing the social polarity. The GINI coefficient (which measures the income inequities) and the risk-of-poverty rate have increased, thus demonstrating the unequal distribution of the welfare effects induced by the 2005 fiscal relaxation.

In the last Part, we support the necessity to introduce some accompanying measures, which could provide the success of implementing the flat tax in Romania.

2. The Flat Tax – from Politics to Economy

At the beginning of 2005, the new liberal-democrat government was introducing the flat tax of 16% to the natural persons' incomes and to the profit of the companies, thus reforming the fiscal system from Romania from its very bases. After almost three years of implementation, the Romanian Prime Minister was mentioning that the Romanian fiscal system, based on the flat tax, will be undoubtedly maintained during his governing mandate, considering that this taxation level represents an element of stimulating and of attracting investments. He was also considering that the maintaining of the flat tax is necessary for a longer period, and this thing would contribute to attracting investments and to the sustained development of the economy. "The Romanian economy needs a constant, strong increase in order to catch-up the gaps which separate us from the other European countries", the Prime Minister added.

Which are the approaching differences between the social democrat governing up to 2005 and the liberal-democrat governing after 2005? Which of the two fiscal systems is medium-and long-term sustainable, existing a little probability that they generate macro-economic dis-equilibriums? How can the proportion of the revenues paid-up to the budget in the GDP be raised from 29% (in 2005, Romania placing the last in EU27, considering this indicator) to 40-45% of the GDP, the average of the EU27 countries? We shall try to answer these questions in what follows. We shall comparatively analyze the two fiscal systems: the progressive one up to 2004 and the current system, based on the flat tax and initiated on January 1st 2005.

The 2004 Fiscal System: Progressive, Redistributive, Social

In 2004, the fiscal system was a progressive one, with differentiated quotas regarding the tax applied to the personal income. Thus, those who had incomes up to 28 millions ROL were paying 18% income tax, for those who had incomes between 28 millions ROL and 69.6 millions ROL the quota was 23% and the tax could have reached up to 40% on conditions that the income exceeded the amount of 156 millions ROL. The tax applied to the taxable profit of the companies was 25%. The interests were taxed with 1% and the dividends with 5%. The micro-enterprises were paying an income tax of 1.5%.

The International Monetary Fund (2004) was considering the Romanian fiscal system from that period as being a sustainable one, which was not creating macro-economic dis-equilibriums on a medium and long term. The improved savings–investments balance, the rapidly decreasing arrears, the maintaining of the salaries within the limits of the approved budgets and the high collecting rates of 95 – 98% at the main units, all these were scientifically proving the IMF specialists' positive analysis: "In 2004 the budgetary policy will support the objectives of disinflation and of limiting the external current account deficit, at the same time creating conditions for the continuous increase of the investments within the private sector".

What was going to happen to the fiscal system in 2005, in case the social democrat governing was continuing? The Government has already agreed with IMF – in July 2004 – upon a fiscal reform program, which would have come into force on January 1st 2005. The main supports of this program were foreseeing a decrease in the tax on profit from 25% to 19%, a decrease in the revenues tax for the lowest imposing installment from 19% to 14%, partially compensated by freezing the personal deduction at the 2004 level and a decrease in the rate of contributions for the social insurances with 1.25%. In order to compensate a part of the revenues loss, the authorities wanted to simultaneously introduce a land tax, to increase the tax on dividends and on excises, starting with July 1st 2005. The net revenues loss related to the reform would have increased to 0.25% of the GDP compared to the basic reference, due to the effect, during the entire year, of eliminating the decreased rate for imposing the tax on profit for export activities, of the carried forward effect of the excises' increase in July 2004, of some improvements in collecting due to the reform in the fiscal administration reform as well as to a decrease in the tax evasion.

The reform objectives proposed by the social-democrat government were related to the support given to the equal distribution of the gains obtained as a result of the raised economic growth, to the improvement of the business environment and to the strengthening of the Romania's competitive position. Moreover, the fiscal reform wanted to meet the businesspersons' expectances, related to the increase in predictability in the fiscal domain, to the decrease in the administrative costs related to taxes and to the decrease of pressures related to the work taxation.

The 2005-2008 Fiscal System: Stimulating for the Big Businesses and for Consumption

Even though the polls performed among business persons during the last 5 years show that the taxation level is one of the secondary concerns of the companies' management (much less important than the fiscal predictability, for example), the 2005 fiscal reform aimed at encouraging the big businesses, the companies with a high financial force, relying on the carrying over of the huge investments performed by these investors.

Starting with January 1st 2005, the Government has introduced the flat tax of 16% to the natural persons' incomes and to the profit of the companies. In the case of the micro-enterprises, the quota of taxation to the incomes is 2% in 2007, 2.5% in 2008 and 3% in 2009. For the gains obtained from redemption of the titles to the open investment funds there will be applied a differentiate taxation rate depending on the owning period of these titles, as it follows: in case they are owned for a period shorter than 365 days, the 16% general taxation rate will be applied; if they are owned for a period longer than 365 days, a 1% taxation rate will be applied on the net obtained gain. For the securities, for the movables values, bought and redeemed, there will be applied a 16% tax on the net gain obtained from purchasing operations during the fiscal exercise. The dividends, including the amounts received from owning titles to the closed investment funds, obtained by the natural persons, will be taxed with a 16% quota of the value of the due gross dividends.

3. Implementation Strategic Errors, from the Very Beginning

The imitation effect developed strongly in relation to the substantiation of the vision regarding the fiscal reform. After several countries from the Central and Eastern Europe have passed from the progressive system to the flat tax, the Romanian decidents in the economic policies have considered necessary that Romania should considerably lower the taxation level in order to be able to attract a very significant part of the foreign investments performed in the region.

In Romania, at least three strategic errors have been made of substantiation for the implementation of the flat tax.

The first is related to the fact that, compared to the Central and Eastern European countries which applied this taxation reform, in Romania there have not been made cost-benefit analyses related to these measures. The simulations related to a figure or to another regarding the taxation have lacked, and the impact on the budget has not been quantified. There has not been a modelling of the influence on the main macro-economic indicators. The scientifically proving of such a measure has lacked.

The second error has been the settlement of a flat tax to a very low level. If we analyze Table 1, we can notice that almost all the countries which adopted the flat tax have chosen medium levels (between the maximum and the minimum of the taxation quotas previously used in the progressive system). Romania is the only country, which has chosen a 16% flat tax, much lower than the minimum level regarding the tax on profit, and the tax on the natural persons' income. This has happened when Romania had the lowest proportion of the revenues in the GDP, compared to the EU countries and to those who subsequently joined this club.

Table 1. The flat tax in a few Central and Eastern European countries

	Adoption of the flat tax	Personal revenues tax		Profit tax	
		before	after	before	after
Estonia	1994	16-33	26	35	26
Lithuania	1994	18-33	33	29	29
Latvia	1997	10 şi 25	25	25	25
Slovakia	2004	10-38	19	25	19
Romania	2005	18-40	16	25	16

Source: The International Monetary Fund, 2007

The third error of substantiation has been the ignoring of the macro-economic framework within which this measure has been implemented. On conditions that, in 2004, Romania had a positive output gap, in which the GDP increase was effectively higher than the increase of the potential GDP (Table 2), the recommendations were favoring the cautious macro-economic policies. As the governing authorities until 2004 have been succeeding in bringing inflation to a one-figure level, from 40.7% in 2000 to 9.3% in 2004, the economic analysts were considering necessary the strengthening of the macro-economic stability through moderated pro-cyclic or even restrictive policies.

Table 2. Output – gap values in Romania

Year	Output-gap HP (%)
1999	-1.55
2000	-9.13
2001	-3.19
2002	-2.40
2003	-3.92
2004	3.90
2005	-1.67
2006	2.37
2007	2.21

Source: Socol, A. and Chiriacescu, B, 2008

On the basis of the economic inconsistencies created by the erroneous applying of the flat tax and of the negative contribution of agriculture to the GDP (strong drought), the macro-economic volatility appeared again during the period 2005 – 2007 (in Table 2 there can be noticed that, after the expansion output-gap from 2004 of + 3.9% of the GDP, in 2005 there has been a return to the recession output-gap of – 1.67% of the GDP and in 2006 the economy returned to an expansion trend of + 2.37% of the GDP).

4. Macro-economic Effects of Implementing the Flat Tax

When the authorities implemented the flat tax, they have erroneously considered that the decrease in taxation will have as a result a strong increase in the aggregate supply, asserting that the measures of taxation relaxation represent the essence of the economic approach from the point of view of the aggregate supply. The logic was a simple one. The decrease in taxation will lead to an increase of the revenues collected to the budget level and the next step will be a voodoo type economy, as George Bush called it in 2004 (economy within which the families will consume more, the production will grow and the decrease of unemployment will take place instead of the increase of inflation).

The possible expansion of the businesses, the increase of the direct investments, the decrease in the proportion of the underground economy, a sustainable economic growth, more jobs, the increase of the savings and of the investments were listed as representing the ingredients of a successful program. However, the mechanism presented above – though subject to disputes even in the developed countries, which are better structured – has not been carried out in Romania.

In Romania, the macro-economic stability has got increased, the inflation has returned to the increasing trend, the proportion of the budgetary revenues to the GDP has insignificantly got increased, and the twin deficits (the budget deficit and the current account deficit) has got deepened. Moreover, it is not certain whether the measure of introducing the flat tax has brought to the surface jobs from the underground economy and whether it attracted a larger volume of direct foreign investments. To carry on, we shall demonstrate that the effects of introducing the flat tax are mixed, and not at all those predicted by the authorities.

Within the macro-economy, there are considerations about the fact that the main indicator of stability at the macro level is represented by the inflation rate. From this point of view, the expansionist fiscal-budgetary policy promoted by adopting the flat tax was opposite to the postulates in the macro-economic theory. Under the terms of the existence of a positive output gap (as recorded by Romania in 2004), which reflects an excess of aggregate demand, we consider that a shock-type relaxation of taxation was not indicated, but a gradual one. When elaborating the taxation reform in Romania, because of the fact that the power of the effect of multiplying of the taxation rate was not exactly known, an excess aggregate demand has been generated, thus determining inflationary effects on a medium term (2 – 3 years) which could hardly be stopped.

In Romania, the decrease in taxation had effects both on the aggregate demand and on the aggregate supply. However, the effects have been differentiated in terms of proportion, the incidence of the decrease in taxation being much stronger on the aggregate demand ($AD_0 \rightarrow AD_1$) than on the aggregate supply ($AS_0 \rightarrow AS_1$).

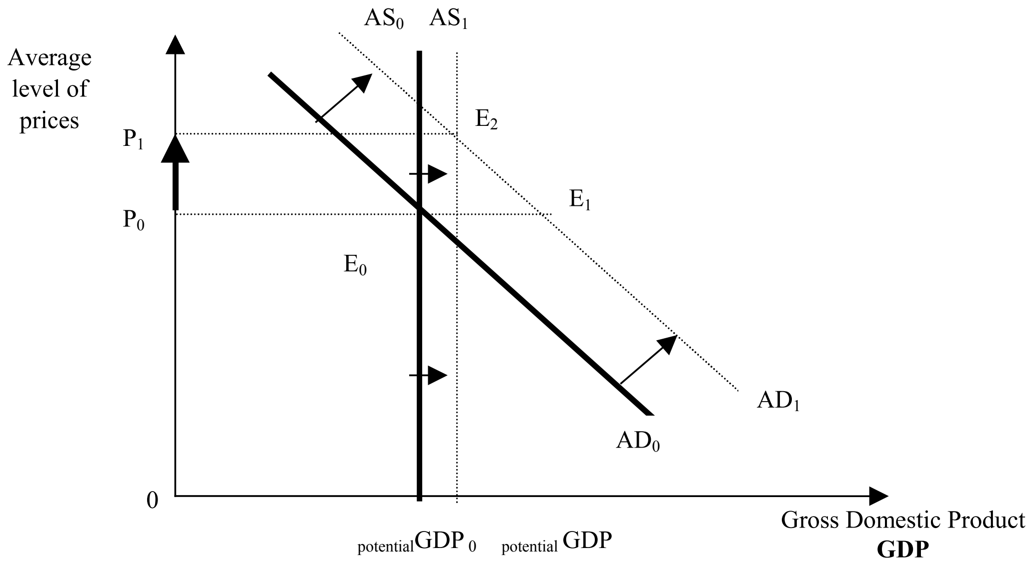
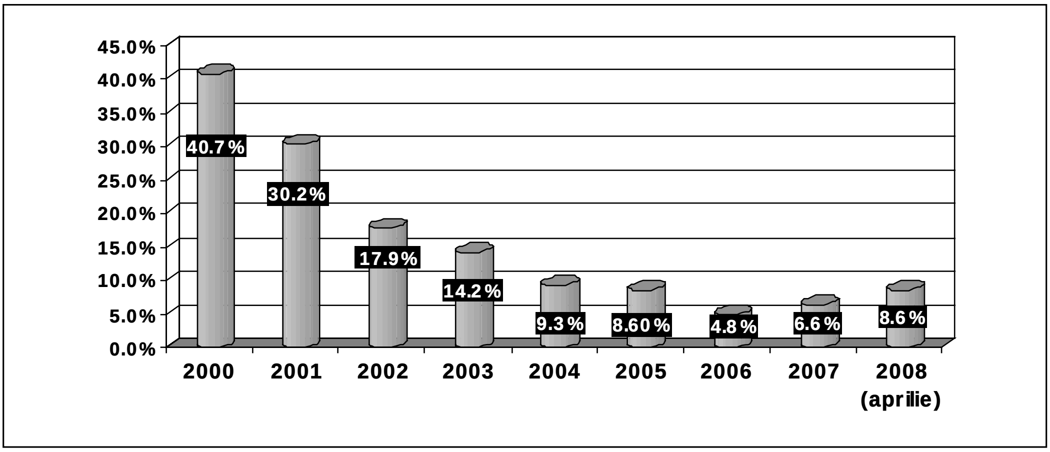


Figure 1. Effects of introducing the flat tax (without accompanying measures).

On a short term, as a result of the decrease in taxation (and also of the loans expansion), the aggregate demand has got increased ($AD_0 \rightarrow AD_1$), so that the economy has moved from E_0 to E_1. The Gross Domestic Product has got increased with a higher proportion during the period 2005 – 2007, leading to an increase in taxation revenues collected to the budget, but with a smaller percentage than that of the decrease in the marginal rate of taxation. In fact, the loss of revenues to the budget from the direct taxes and fees has been exceeded by the increase in revenues collected to the budget from the VAT, based on the explosion of consumption.

On a medium term (and the trend will continue on a long term), the Romanian economy tends towards the point E_2 (as a result of the movement of the aggregate supply $AS_0 \rightarrow AS_1$). As a result of the decrease in taxation the working and investing incentives grow and thus the potential GDP moves from $_{potential}GDP_0$ to $_{potential}GDP_1$. The growth of the $_{potential}GDP$ is much smaller than that of the effective GDP on a short term, obtained by stimulating the aggregate demand. It results that the taxation revenues collected to the state budget will get increased with a much lower rate compared to the rate of decrease in the collections to the budget as a result of applying the flat tax → the budget deficit grows → strong inflationary pressures appear. There also can be noticed that the inflation grows ($P_0 \rightarrow P_1$) from the analysis of Figure 1.

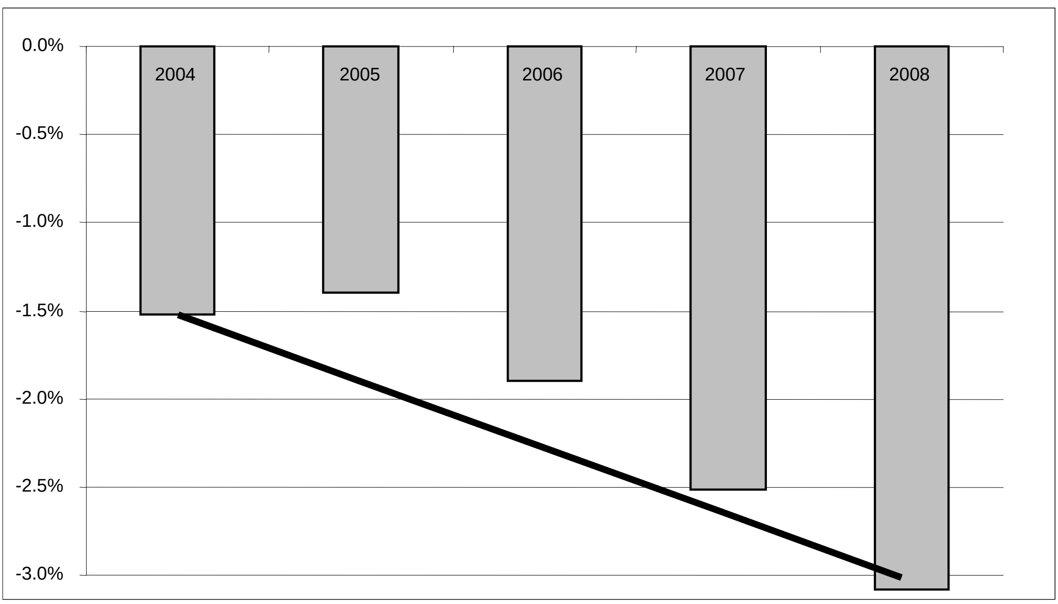
The above-mentioned theoretical model has been materialized in the Romanian economy sooner than the most optimistic economic analysts could have expected. Although the Central Bank of Romania applied the inflation targeting strategy starting with August 2005, it has not succeeded in keeping a tight hand over the inflationary pressures, the strong pro-cyclic characteristic of the taxation policy being appealed to as a cause of this macro-economic stability failure. *The dis-inflationary process has been stopped starting with 2007 and the inflationary pressures are still strong in 2008, the annualized rate of inflation reaching, in April, 8.6% of the level from 2005 (Figure 2).*



Source: The National Bank of Romania, National Institute of Statistics and of Economic Studies, 2008

Figure 2. Inflation rate in Romania during the period 2000-2008 (calculated on the basis of IPC, December/December).

The macro-economic stability, generated by the pronounced consumption growth, by the taxation relaxation determined by the introduction of the flat tax may also be revealed by the deepening of the twin deficits – the budget deficit and the current account deficit. Thus, the proportion of the budget deficit in the GDP will be more than double during the last 3 years from 1.5% in 2004 to 3.2% of the GDP in 2008, according to the spring prediction of the European Commission (Figure 3).

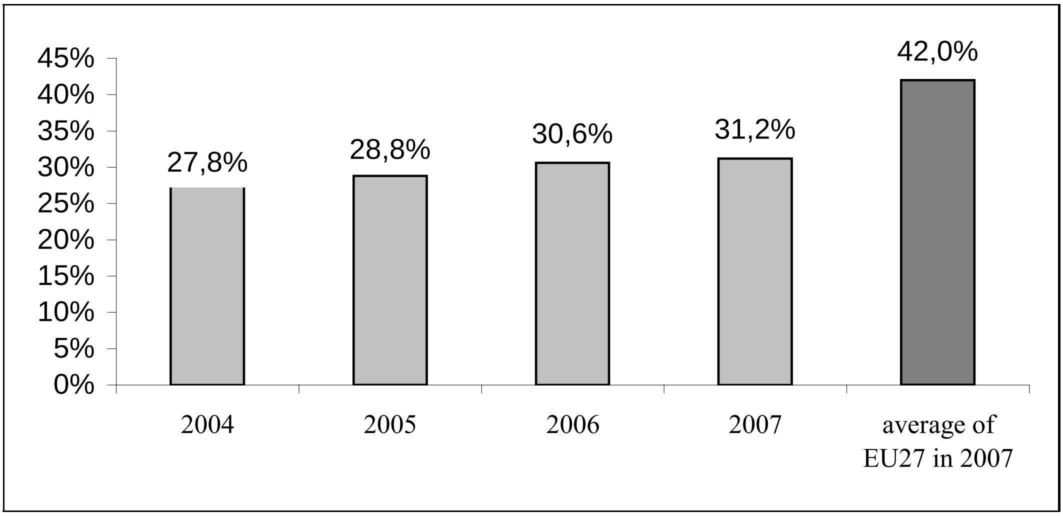


Source: Ministry of Economy and Finances, The European Commission (prognosis for 2008), 2008

Figure 3. The proportion of the budget deficit in the GDP in Romania during the period 2004-2008.

If on a short term (the period during 2005-2006) the growth in the revenues from the VAT to the state budget have got increased as the consumption has got increased, covering, to a great extent, the loss of budget revenues caused by the taxation relaxation, starting with 2007 there can be noticed a saturation of the growth rhythm for the collections from the VAT, thus generating pressures in terms of deepening the public budget deficit.

The introduction of the flat tax has not solved the problem of the increase in revenues to the budget in the GDP. The increase I revenues to the budget during the period 2005-2007 is poor, based, to a great extent, on the increase in the revenues collected from the VAT. These have increased especially on the basis of the increase in consumption, as a result of the loans expansion, of the taxation relaxation and of the remittances.



Source: The Ministry of Finances, The National Bank of Romania, The National Institute of Statistics and of Economic Studies, Eurostat, 2007

Figure 4. The proportion of the budget revenues in the GDP in Romania during the period 2004-2008. Comparison to the average of the EU27 countries.

The introduction of the flat tax in 2005 had as an effect the worsening of the public finances sustainability. During the period 2001-2004, under the terms of existence of the progressive taxation system and of the promoting a few cautious macro-economic policies, the authorities have succeeded in decreasing the structural budget deficit from -2,34% in the GDP in 2001 to -1,17% of the GDP in 2003 and to -2,14% in 2004 (Table 3).

The implementation of the flat tax has and different effects on a short term and on a medium term. On a short term – in 2005 – the structural budget deficit was - 0,78% of the GDP. This improvement – compared to the last year – cumulates, on the one hand, the compensation of the negative effects of the taxation decrease on the budget deficit by increasing the revenues collected from the VAT and, on the other hand, the gapped positive influences which have resulted from the consistent applying of the economic reforms during the period 2001-2004. However, on a medium term, the positive impact of the applying of the flat tax on the public finances sustainability in Romania ceased to manifest. Thus, during the period 2005-2007, the structural budget deficit (the adjusted cyclically budget deficit) has considerably got increased from -0,78% in 2005 to -2,35% in 2006 and to -3,54% in 2007, mainly because of the strongly expansionistic characteristic of the fiscal-budgetary policy. Moreover, during the period 2006-2007, the fiscal-budgetary policy was considerably pro-cyclic, fact which was demonstrated by the existence of a production gap of the positive GDP combined with the expansionist trend of the fiscal-budgetary policy. Anyway, it is obvious that, on conditions that the Romanian economy prepares to enter the Euro zone, the implementation of a restrictive taxation policy is absolutely necessary, and it has to be able to improve both the current budget deficit and the structural one.

Table 3. The production gap and the structural budget deficit in România (1999-2007)

Year	Output-gap HP (%)	Structural budget deficit (%GDP)
1999	-1,55	-1.41
2000	-9,13	-1.40
2001	-3,19	-2.34
2002	-2,40	-1.95
2003	-3,92	-1.17
2004	3,90	-2.14
2005	-1,67	-0.78
2006	2,37	-2.35
2007	2,21	-3.54

Source: Socol, A. and Chiriacescu, B., 2008

On the other hand, we have asked the following question: How did the introduction of the flat tax influence the "health" of the public finances in Romania? With the purpose of quantifying the impact on the budgetary revenues determined by the introduction of the flat tax in 2005, we have introduced a dummy variable in the equation which verifies the long-term relationship between the total revenues and the Gross Domestic Product (GDP). For this reason, we have used the Johansen co-integration procedure (for details related to the presentation of the used methodology, see Annex I). The co-integration coefficients are presented in the Table 4. Based on these coefficients, there have been made estimations related to the long-term equilibrium relationships between the variable.

$$\text{Ln(total revenues)} = \phi_1 \ln(\text{GDP}) + \phi_2 * \text{dummy} + u_t$$

$$\text{Ln(total revenues)} = 1.11 * \ln(\text{GDP}) + (-0.07) * \text{dummy} - 2.42$$

The obtained results (coefficient $\phi_2 = -0.07$) show the fact that the modification brought to the budgetary philosophy introduced in 2005 (by applying the flat tax) had a negative impact on the total budgetary revenues.

A modality of investigating the public finances sustainability, recommended by Bohn (1998, 2005) and used, respectively improved / added by several other authors (Fatas and Mihov, (2003), Gali and Perotti, (2003), de Mello (2005) etc.) is represented by the estimation of a reaction function of the fiscal-budgetary policy (for details related to such a function, see Annex II). Through such a function, we have tested for Romania the reaction of the modification brought to the structural budget balance to the shocks occurred at the public indebtedness level, at the production gap level, at the level of the previous values of the primary budgetary balance and a dummy variable – introduced in order to offer information about the way in which the proportion of the structural balance has been influenced as a result of introducing the flat tax in 2005.

Table 4. Estimation results

Vector Error Correction Estimates
Date: 05/15/08 Time: 10:21
Sample (adjusted): 1999Q2 2007Q3
Included observations: 34 after adjustments
Standard errors in ( ) & t-statistics in [ ]

Cointegrating Eq:	CointEq1		
L_VENITURI_TOTALE(-1)	1.000000		
L_GDP(-1)	-1.118027 (0.07407) [-15.0944]		
DUMMY(-1)	0.071053 (0.03132) [2.26880]		
C	2.420734		
Error Correction:	D(L_VENITURI_TOTALE)	D(L_GDP)	D(DUMMY)
CointEq1	-1.255658 (0.35024) [-3.58512]	0.028309 (0.26916) [0.10517]	0.323692 (2.60161) [0.12442]

Source: authors' calculations

The reaction function of the taxation policy for Romania has been calculated by using one of the models proposed by Golinelli and Momigliano, 2007 (CAPB/PB model, for details see Annex III):

$$\Delta CAPB_{it} = \phi_1 * PB_{it-1} + \phi_2 * DEBT_{it-1} + \phi_3 * GAP_{it(t-1)} + u_{it}$$

where,

$\Delta CAPB_{it}$ = the modification of the primary structural budgetary balance
PB_{it-1} = the primary budgetary balance during the previous period
$DEBT_{it-1}$ = the public debt in t-1
$GAP_{it(t-1)}$ = the output gap in t-1
ϕ_1, ϕ_2, ϕ_3 = coefficients

For the coefficients' estimation we have used the Johansen co-integration procedure. The co-integration coefficients are presented in the Table 5. Based on these coefficients, estimations have been made for the long-term equilibrium relationships between the variables.

$$\Delta CAPB_{it} = \phi_1 * PB_{\ it-1} + \phi_2 * DEBT_{it-1} + \phi_3 * GAP_{\ it(t-1)} + \phi_4 * dummy + u_{it}$$

$$\Delta CAPB_{it} = (-1.18) * PB_{\ it-1} + 0.16 * DEBT_{it-1} + (-0.81) * GAP_{\ it(t-1)} + (-0.086) * dummy + 0.043$$

Table 5. Co-integration vectors for the modification brought to the structural deficit (ΔCAPB)

Vector Error Correction Estimates
Date: 05/15/08 Time: 10:39
Sample (adjusted): 2000Q2 2007Q3
Included observations: 30 after adjustments
Standard errors in ( ) & t-statistics in [ ]

Cointegrating Eq:	CointEq1
MODIF_DEF_STRUCT(-1)	1.000000
DEF_PRIMAR_1(-1)	1.182216
	(0.06982)
	[16.9334]
DAT_PUB_1(-1)	-0.162548
	(0.01313)
	[-12.3775]
OUTPUT_GAP_1(-1)	0.817970
	(0.12370)
	[6.61254]
DUMMY_1(-1)	0.086965
	(0.00506)
	[17.1951]
C	-0.043317

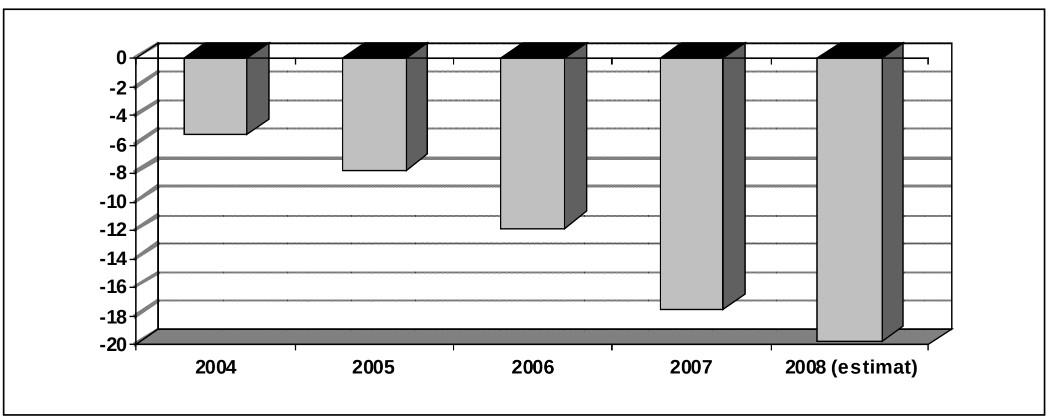
Error Correction:	D(MODIF_ DEF_STRUCT)	D(DEF_PRIMAR_1)	D(DAT_PUB_1)	D(OUTPUT _GAP_1)	D(DUMMY _1)
CointEq1	-3.459910	0.080983	1.214000	-0.078177	-9.458238
	(1.09518)	(0.11253)	(0.59238)	(0.16996)	(4.19220)
	[-3.15921]	[0.71966]	[2.04938]	[-0.45998]	[-2.25615]

Source: authors' calculations

The analysis of the reaction function related to the fiscal-budgetary policy for Romania has lead us to obtaining the following results: the coefficient ϕ_1 = -1,18, is negative, this meaning that *there is not a tendency to balance the budget and this points out a decrease of the public finances sustainability degree.* The coefficient of the public debt proportion in the GDP (φ_2= 0,16) is positive, demonstrating the fact that, *in taking decisions, the fiscal-budgetary authorities have taken into consideration the constraint regarding the public debt, but with a quite low importance* (because the coefficient is low). To a great extent, this fact can be explained for Romania, where the proportion of the public debt in the GDP (average of 12% of the GDP during the period 1999-2007), can be limited without problems to the criterion provided in the Treaty of Maastricht (< 60% of the GDP).

The analysis of the factors which have an influence on the modification brought to the structural budgetary balance demonstrates the fact that the only indicator taken into consideration by the authorities when scientifically proving the decisions is the output-gap. The coefficient (φ_3=-0,81) brings arguments to what we have pointed out earlier in this work, that is the fact that during the analyzed period (1999-2007) *the fiscal-budgetary policy has predominantly been pro-cyclic, thus significantly decreasing the sustainability of the public finances.* The coefficient of the dummy variable is significant from a statistic point of view, this demonstrating that *the introduction of the flat tax since 2005 has had a high impact on the modification to the structural budgetary balance.* Moreover, the influence was a negative one (ϕ_4 = -0.086), thus emphasizing what we have previously demonstrated in this work.

The adoption of the flat tax has also emphasized the external dis-equilibriums. The commercial deficit has got increased four times during the last 3 years, from 5 billions Euro in 2004 to 21.5 billions Euro in 2007, thus causing the strong deterioration of the current account deficit (Figure 5). The deepening of the current account deficit is serious because its financing becomes less and less sustainable, that is based on the direct foreign investments (ISD). Thus, if in 2006 the current account deficit of 10.4% of the GDP was 91% financed from direct foreign investments, in 2007 the degree of covering this deficit on the basis of the direct foreign investments has decreased to 42% (The National Commission on Prognosis, 2008).

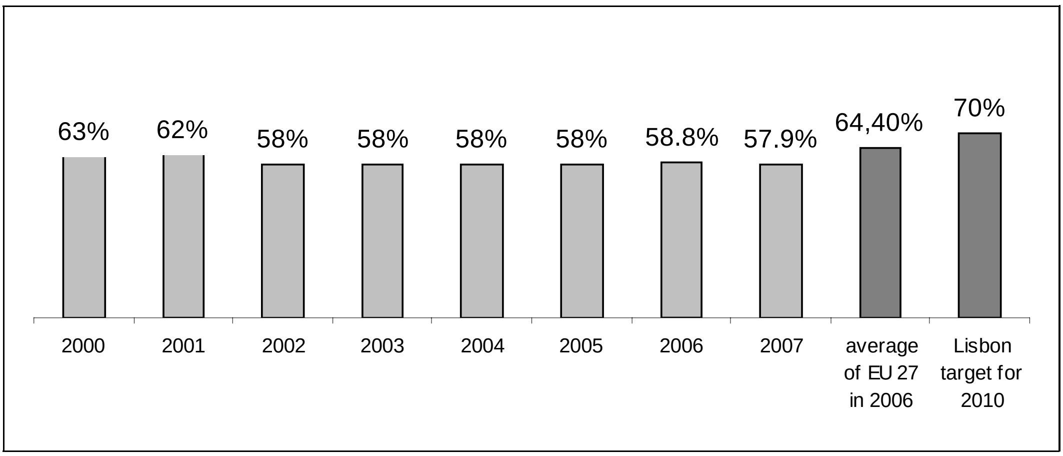


Source: The National Institute of Statistics and of Economic Studies, The National Commission on Prognosis, Bucharest, 2008

Figure 5. Romania's commercial deficit during the period 2004-2008 (export(FOB) – import (FOB), billions euro).

The taxation decrease has left bigger available revenues at the disposal of the individuals with bigger incomes, they strongly increasing the consumption of the imported goods. If we add the increase in the remittances (the amounts of money sent at home by the Romanian people who left to work abroad have raised to 7.16 billions Euro only 2007) and the relaxation of the loans conditions, we can find that the strong increase of the demand for imported goods had solid engines in Romania, which materialized by deteriorations without precedent of the external commercial balance. Moreover, the lack of an active industrial policy made that the Romanian good have a lower and lower external competitiveness, the exports not being able to keep pace with the imports.

The flat tax has not solved the problem regarding the increase in the employment rate. The flat tax has not generated more jobs than the economy would have normally generated and the employment rate has got decreased in 2007, compared to 2004 (Figure 6). The number of the full-time employees has raised with 224.8 thousands persons during the period 2005 – 2006 (out of which: +10.7 thousands persons in agriculture, constructions (+42 thousands persons), commerce (+128.2 thousands persons), transportation, depositing and communications (+39.3 thousands persons), real estates transactions (+46.3 thousands persons) and public administration (+51.1 thousands persons)). It is an insignificant raise, which rather indicates that the Romanian economy has become a commodity market and not a competitive one.



Source: Eurostat, Labour Force Statistics, 2008

Figure 6. Employment rate in Romania (2000-2006) and EU 27.

The introducing of the flat tax has not solved the problem regarding the "black-market for labor". The implementation of the flat tax has not brought to the surface thousands of jobs, as the authorities assert, but the contrary. If in 2005/2004, The National Institute of Statistics (NIS) was estimating a decrease in the number of workers in the underground economy raising to 155,000 persons, in 2006/2005, the NIS estimations showing an increase of 238,000 persons who work in the underground economy, in 2007/2006 a decrease of 47,000 persons, and in 2008/2007 there is an estimation of an increase of 10,000 persons (Table 6).

Table 6. Number of workers in the underground economy (thousands of persons)

	2004	2005	2006	2007	2008
1. Average number of employees according to the AMIGO investigation (investigation made in households)	6036	5921	6167	6173	6278
2. Average number of employees according to the BFM investigation (investigation made in companies)	4469	4559	4667	4770	4865
3. Difference (1-2)	1567	1362	1500	1403	1413
- Employees in the non-civil sector (defense, persons who are in the military service)	250	200	100	50	50
- Employees in the underground economy	1317	1162	1400	1353	1363
4. Evolution of the number of employees in the underground economy	n.a.	-155	+238	-47	+10

Source: The National Institute of Statistics, based on the estimations made by The National Commission on Statistics, 2008

It is obvious that there are solid evidences that the process of creating new jobs in the official economy has been made by consistently decreasing the number of persons employed in the underground economy.

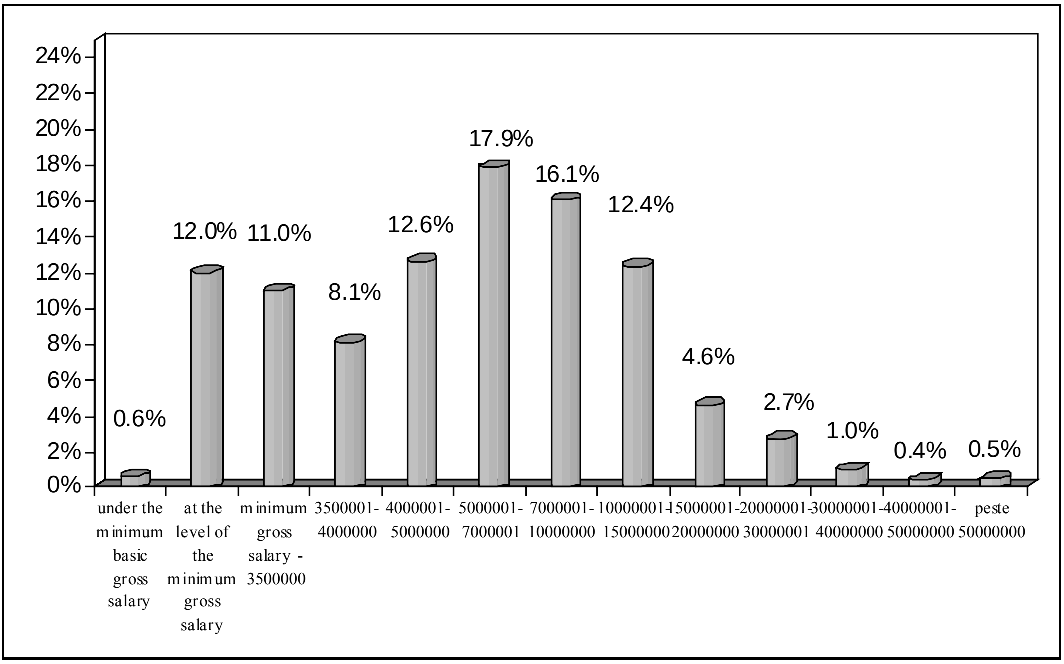
The implementation of the flat tax has not solved the problem regarding the direct foreign investments. In 2005 and 2006, the volume of the direct foreign investments has got increased especially due to the privatization of a few profitable companies (see the case of the Romanian Commercial Bank, the biggest bank in Romania). Moreover, the UNCTAD Report on 2007 shows that the direct foreign investments in Romania have decreased with 21% in 2007 compared to 2006. On the other hand, the analysis of the direct foreign investments' structure in Romania indicate investments in fields with a low value added (consumption, primary processing etc.), the foreign companies taking advantage especially of a cheap labor and which has a low professional training degree.

The flat tax has not solved the problem regarding the compatibility between the monetary policy – the fiscal-budgetary policy. The introduction of the flat tax has lowered the efficiency and the flexibility of the fiscal-budgetary policy regarding the absorption of the symmetrical and asymmetrical shocks which will be faced the economy in the case of adheration to the Euro zone. The implementation of the flat tax has increased the inflationary pressures, rendering the NBR's mission of lowering inflation difficult.

5. Social Effects of the Flat Tax

By introducing the flat tax, the winners were the families with high incomes and the losers were the families with low incomes. Because of the disequilibrium existing in the social structure of the incomes, the burden fell rather on the individuals with low incomes than on those who have high incomes. Even though the authorities asserted that the employees earned amounts in a range between 50,000 ROL (2$) and 10,000,000 ROL (400$), the reality shows that the low pluses in the incomes have immediately melted in the subsequent increases in fees and taxes and in inflation.

The employees with incomes higher than 15 millions ROL (600$) have benefited from the introduction of the flat tax at the end of 2004, and this means 183,062 employees, representing only 4.8% of the overall full-time employees at the end of 2004 (3,813,800 full-time employees) (Figure 7).



Source: The National Institute of Statistics and of Economic Studies, 2005

Figure 7. Distribution of the full-time employees according to the wages earned at the end of 2004.

The introduction of the flat tax has increased the income gaps between the different categories of population, emphasizing on the social polarization. At the end of 2006, the average incomes per household during the last decile D10 were 5.15 times higher than the average incomes per household during the first decile D1 (increasing gap compared to 2005, when the ratio was 4.73:1). The ratio between the incomes which averagely was resting with a person from the households which form the first and the last decile was of 1:8.38 (compared to 1:7.78 in 2005) in December 2006 (Tabel 7).

Table 7. Evolution of the population's incomes in 2006, per deciles

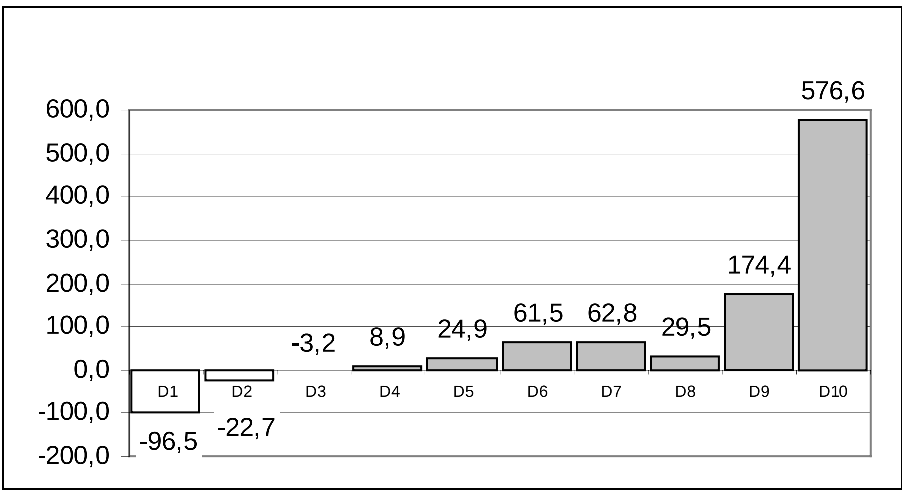
Decile	Average monthly incomes		Ratio compared to decile 1		Overall proportion (%)		
	RON* per a household	RON per a person	Average incomes per a household	Average incomes per a person	Households	Persons	Average incomes per a household
D1	634.7	162.5	1.00	1.00	10	13.33	4.58
D2	765.5	239.8	1.21	1.48	10	10.90	5.52
D3	858.8	288.6	1.35	1.78	10	10.16	6.20

Table 7. Continued

Decile	Average monthly incomes		Ratio compared to decile 1		Overall proportion (%)		
	RON* per a household	RON per a person	Average incomes per a household	Average incomes per a person	Households	Persons	Average incomes per a household
D4	999.,9	337.3	1.58	2.08	10	10.12	7.21
D5	1095.2	382.6	1.73	2.35	10	9.78	7.90
D6	1193.8	435.0	1.88	2.68	10	9.36	8.61
D7	1400.3	499.0	2.21	3.07	10	9.58	10.10
D8	1641.9	594.2	2.59	3.66	10	9.43	11.85
D9	2006.1	748.9	3.16	4.61	10	9.15	14.47
D10	3266.5	1361.6	5.15	8.38	10	8.19	23.56
Households - total	1386.3	473.3	2.18	2.91	100	100.00	100.00

Source: The National Institute of Statistics and Economic Studies, www.insse.ro, 2007
* since 2005, 1 RON = 10.000 ROL

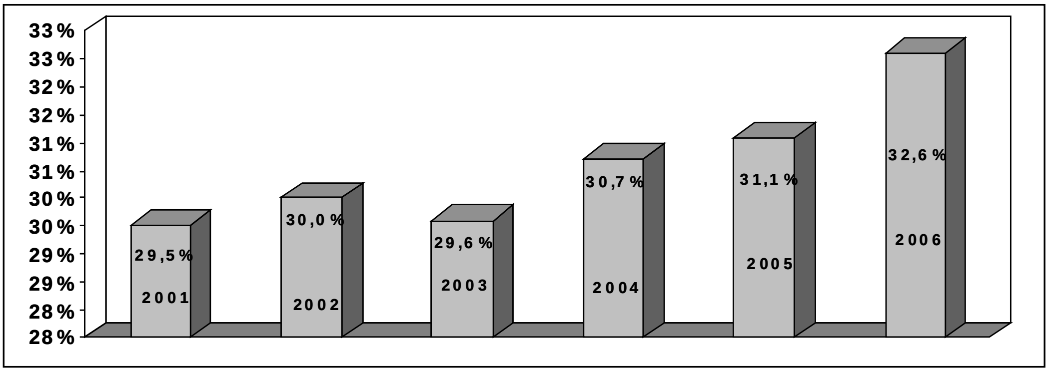
In the Figure 8, there can be noticed that the households from the first 3 deciles (meaning 34.3% of the population) record a deficit of resources.



Source: The National Institute of Statistics and Economic Studies, www.insse.ro, 2007

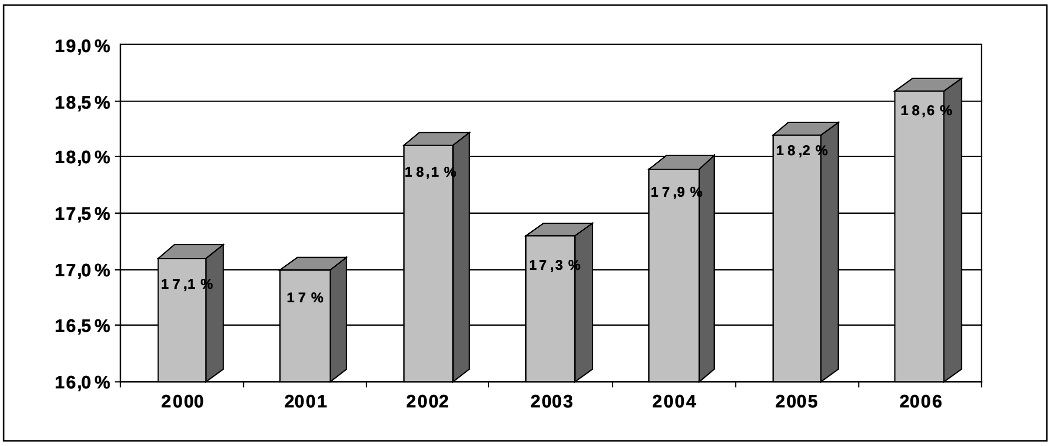
Figure 8. The equilibrium of the balance between incomes and expenses in 2006.

The emphasizing of the social polarization after 2005 may be argued from the analysis of the GINI indicator (which shows the incomes inequity) and of the poverty risk (calculated after the European Union's methodology, at the point of 60% of the incomes median) (Figure 9 and Figure 10). Both the increase in the level of the GINI indicator and the increase of the poverty risk rate in Romania emphasize the negative social effects of introducing the flat tax.



Source: The Report on Social Tendencies, The National Institute of Statistics and Economic Studies, 2007

Figure 9. Evolution of the GINI indicator in Romania during 2001-2006.



Source: The Report on Social Tendencies, The National Institute of Statistics and Economic Studies, 2007

Figure 10. Evolution of the poverty risk rate during 2001-2006 (estimations made on the basis of the available incomes, including the consumption from the personal incomes).

6. Accompanying Measures – Solutions for the Success of the Flat Tax. Conclusion

Analyzing the implementation of the flat tax in Romania, we may say that its introduction represents a "primitive" taxation philosophy, unsustainable, specific to the emergent countries. Moreover, the simulations for applying the flat tax in the Western developed countries lead to unsatisfactory results, being considered an unfeasible measure of taxation reform.

The emphasis of the internal and external economic and social gaps as a result of adopting the flat tax supposes the need to adopt some accompanying measures.

First of all, there are measures leading to the *improvement of the financial system:* the decrease of the public expenses, by decreasing the proportion of the governmental sector (a measure which is difficult to achieve because Romania needs an increase incentive, and the investments in the infrastructure represent the most available instrument, also complying with the modernizing objectives); the enhancement of the forced execution and of the imposing of hard constraints to all the bad payers; the decrease in the proportion of the contribution to the social insurances – a long-term balance between the labor taxation and the capital taxation; the decrease in the para-fiscality; the decrease in the number of fees and taxes (Romania is on the first place in EU in respect with the fees and the taxes)[1]; the rationalizing plan for the budgetary expenses; the limitation of the compensations for the ex-owners (dispossessed by the ex communist regime by nationalization) to a maximum and conferring state certificates instead of shares to the Property Fund; the introduction of the multi-annual budgetary schedule; the implementation of the Law on the Unique Control (in order to avoid the harassing of the business environment, the overlapping of the control organisms etc).

Secondly, it is necessary to *encourage the policies for stimulating the aggregate supply*: a higher legislative stability, institutions capable of exerting the governing; the decrease in the internal lags of the fiscal policy (which are already high); predictability in the taxation reform; a stable, predictable competition environment (operability of the Competition Council); conferring state benefits only in compliance with the related European legislation; stimulation of the internal savings and investments; coherent strategies for attracting the direct foreign investments; complete absorption of the structural and cohesion funds coming from EU (30 billions Euro during the period 2007-2013); the decrease in the social contributions in order to increase the labor occupying degree; the increase in the labor mobility by elaborating a Labor Code without ideological prevalences and which can admit the flexibility on the labor market and which can generate responsibility for the economic players.

Last but not least, we consider that there are necessary measures for *decreasing the market distortion by restraining / stimulating the effects of the negative / positive externalities, by:* the strict settlement of the property rights; the decrease of the transaction costs; the use of taxes and subsidies in order to eliminate the effects of the negative externalities; an efficient process of restructuring and privatization, accompanied by the stimulation in the applying of the active industrial policies; the free entrance on / exit of the market of the economic agents (especially the liberalization regarding the exits of the system).

The accompanying measures matrix of the flat tax (previously proposed) will be capable of leading to a higher increase of the $_{\text{potential}}\text{GDP}$ ($_{\text{potential}}\text{GDP}_0 \rightarrow {}_{\text{potential}}\text{GDP}_1$) compared to the increase of the GDP $_{\text{potențial}}$ in the initial model. The aggregate supply AS will get increased with a higher proportion so that the AS curve could move more towards the right (compared to the initial model) ($AS_0 \rightarrow AS_1$).

Thus, the budget deficits will be lower, and the inflationary pressures will be less overwhelming ($P_0 \rightarrow P_1$). The National Bank of Romania will be able to reach the inflation target for the next years and the current account deficit will not be reach “alarming quotas”. Thus, a vicious circle will be created. We will be able to re-motivate the hope that the introduction of the flat tax may be a success. The taxation reform will not attack any more the fragile macro-economic stability in Romania. The incomes inequity will get diminished and the social polarization will get decreased. Romania will aim at what was being confirmed as a

[1] In the Paying Taxes Report 2008 published by The World Bank, Romania is on the 175th place out of 178 countries taken into account, in respect with the number of fees and taxes which have to be paid to the state. In Romania there are 96 fees and taxes which have to be annually paid by a company, out of which 4 revenues and profit taxes; 60 related to labor and 32 other taxes types.

validated element of the economic evolution towards an European standard (thus, giving up to the myth of the eternal return to the national projects).

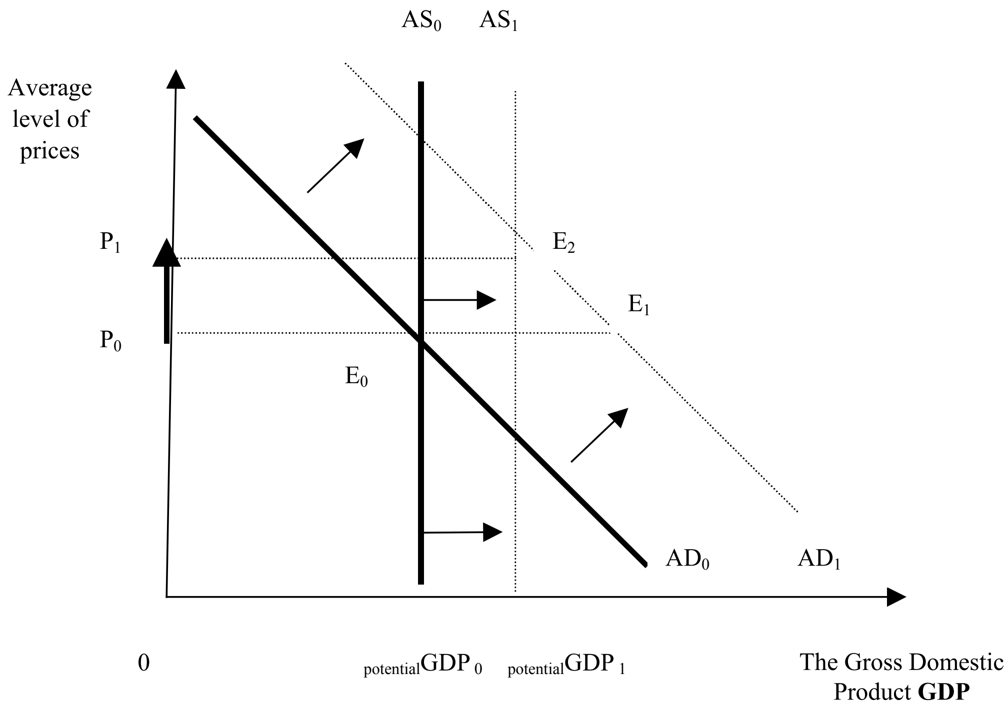


Figure 11. Effects of introducing the flat tax (with accompanying measures).

Annexes

I. The co-integration tests investigate the presence of a mechanism which leads to restoring of the long-term balance between the variables. For this reason, it is necessary to test the order for integration of the variables based on the stationarity tests. The number of the lags used for the stationarity tests have been chosen according to the SC minimization criterion (Schwarz information criterion). The results of the ADF stationarity tests reveal the fact that the above-mentioned series are integrated by order I, and this allows the investigation of the existence of a co-integration relationship between the variables. This means that any long-term relationship between the variables, which are not stationary, supposes that their trends should be correlated. In other words, any long-term balance relationship supposes that the variables cannot evolve independently from one another, and this fact imposes the existence of a stationary combination between the two variables (Enders, 1995). In case two variables x and y are integrated by order *d,* the existence of a co-integration relationship between the two variables supposes the compliance with the following relation:

$$Y_t = \beta * X_t + \varepsilon_t$$

where,

β = the regression parameter

ε_t = the errors term which has to be stationary (integration order 0)

In respect with the number of the lags taken into account within the co-integration, this was determined on the basis of estimating a VAR type model (auto-regressive vector) in which we have introduced the variables used within the analysis. Based on this model, we have chosen the number corresponding to the lags, using the following econometric criteria: Hanan-Quinn Information Criterion (HQ)[2], Akaike Information Criterion (AIC) and Schwarz Information Criterion (SC).

The analyzed period was 1999-2007, and the quarterly data have been supplied by The Ministry of Economy and Finances, Romania.

II. Estimation of a reaction function for the taxation policy[3] is based on the following relation (Gali and Perotti, 2003):

$$PB_{it} = a * PB_{it-1} + b * DEBT_{it} + c * GAP + \text{constant} + \text{error term} \quad (1)$$

where,

PB = primary budget deficit

GAP = output gap

DEBT = public debt

a,b,c = coefficients

The variables (the primary budget deficit, the output gap, respectively the public debt) are expressed as percentage of the GDP. If the coefficient $a > 0$, we interpret that there is a tendency to balance the budget, increasing the public finances sustainability. The coefficient *b* > 0 shows the existence of an active constraint related to the public debt[4]. If $c>0$, then the taxation policy is considered to be anti-cyclic.

III. "The CAPB/PB model" for estimating the taxation policy rule is a model in which the discretionary taxation policy actions are measured by modifying the primary structural budget balance (ΔCAPB). This modification is applied through the initial position of the public finances (measured by the primary budget deficit in the moment t-1 and the public debt in the moment t-1) and by the cyclic conditions (measured through the output gap level). It is considered that the model is stable if the coefficient ϕ_1 is negative, while the public debt coefficient has to be positive. Also, it is considered that the fiscal and the budgetary policy is sustainable if the primary balance's reaction to the shocks of the public debt is instantaneous and not delayed in time. A positive value of the output gap coefficient (ϕ_3) indicates the fact that the taxation policy is anti-cyclic, while a negative value would mean that the taxation policy is pro-cyclic.

[2] Usually, in quarterly time series the Hanan – Quinn (HQ) criterion is the most used

[3] In the economic literature it is aclled „core" reaction function for the taxation policy

[4] Even though this condition does not assure sustainability to the public debt

References

Bohn, H. (1998). The Behavior of U.S. Public Debt and Deficits, *The Quarterly Journal of Economics*, *volume* **113**, pp. 949-963

Bohn, H. (2005). The sustainability of fiscal policy in the United States, *CESifo Working Paper*, No. **1446**, www.cesifo-group.de

Dornbusch, R. & Fischer, S. & Startz, R. (2007). *Macroeconomics*, Bucharest, Economic Publishing House

Enders, W. (1995). *Applied Econometric Time Series*, New York, John Wiley & Sons

Fatás, A., Mihov, I. (2003). On constraining fiscal policy discretion in the EMU, *Oxford Review of Economic Policy,* No. **19**, pp. 112-131

Gali, J. and R. Perotti (2003) Fiscal policy and monetary integration in Europe, *NBER Working Paper*, No. **9773**, www.nber.org

Golinelli, R., Momigliano, S. (2007) The cyclical response of fiscal policies in the euro area. Why do results of empirical research differ so strongly?, *BCE working papers,* www.ecb.int, Frankfurt

Hodrick, F., Prescot, E (1980) Posterior US bussiness cycles. An empirical investigation, *Disscusion Paper*, No. **45**, Carnegie Mellon University

Keen, M., Kim, Y. and Varsano, R. (2006). The “Flat Tax(es)”: Principles and Evidence, *IMF working paper*, Washington, www.imf.org

de Mello, L. (2005) *Estimating a Fiscal Reaction Function: The Case of Debt Sustainability in Brazil*, No **423**, OECD Economics Department, Working Papers, www.oecd.org

Socol, A., Chiriacescu, B. (2008). Estimation of the structural budget deficit for Romania in terms of adheration to the Euro zone, *Theoretical and Applied Economics Review*, www.ectap.ro, No. 3 (520), pp. 13-29

Stiglitz, J.& Walsh, C. (2005). *Economics*, Bucharest, Economic Publishing House,

Reports

World Bank (2008). *Paying taxes* 2008, www.worldbank.org

National Bank of Romania (2008). *Report on inflation* (May), www.bnro.ro

International Monetary Fund (2008). *Statistical data*, www.imf.org

National Institute of Statistics and Economic Studies (2007). *Social Trends*, Bucharest, www.insse.ro

UNCTAD (2008) *World Investment Report*, www.unctad.org

****Stand-By Accord 2004-2006, concluded between Romania and IMF on July 7th 2004, and modified at the beginning of 2005, as a result of Government’s modification, analysis available on www.fmi.ro

Statistical data for the period 1999-2008

National Bank of Romania, www.bnro.ro, 2008

National Institute of Statistics and of Economic Studies Bucharest, www.insse.ro, 2008

National Commission on Prognosis Bucharest, www.cnp.ro, 2008

European Commission (2008), www.eurostat.org

European Commission (2008). EU forecasting 2008-2010, Bruxelles, www.europa.eu.int

Ministry of Economy and Finances Bucharest, www.mefromania.ro, 2008

In: Emerging Topics in Macroeconomics
Editor: Richard O. Bailly, pp. 219-249
ISBN: 978-1-60692-093-0

Chapter 8

A KEYNESIAN MODEL OF ENDOGENOUS GROWTH CYCLE

Toichiro Asada[1,*] and Masahiro Ouchi[2,**]
Faculty of Economics, Chuo University; 742-1 Higashinakano,
Hachioji, Tokyo 192-0393, Japan

Abstract

In this paper, we study an endogenous model of growth cycle in a Keynesian setting. We integrate Kaldor (1957)'s model of endogenous growth, that relates the rate of technical progress to the rate of investment by using the 'technical progress function', into Goodwin (1967)'s model of endogenous business cycle that uses a kind of Phillips curve. Our model is an endogenous model in the triple senses. Namely, the 'natural' rate of growth itself as well as 'actual' rate of growth fluctuates endogenously around the endogenously determined equilibrium rate in the course of the endogenous business cycle. Spirit of our model is Keynesian in the sense that full employment of labor is not assumed and the rate of employment as well as the rate of utilization of capital stock is an endogenous variable. This aspect is a notable characteristic of our model which distinguishes it from the usual neoclassical endogenous growth models which assume the full employment of all productive resources including labor. By using such an analytical framework, we investigate the impact of the changes of some critical parameter values on the macroeconomic performance theoretically as well as numerically. We also refer to the relevance of our model to the analysis of actual macroeconomic problems, in particular, the analysis of the Japanese economy in the 1980s – the 2000s.

Keywords: Keynesian endogenous growth cycle, Kaldorian technical progress function, equilibrium growth rate, dynamic (in)stability, cyclical fluctuation

[*] E-mail address: asada@tamacc.chuo-u.ac.jp
[**] E-mail address: oomiyabi@wta.att.ne.jp

1. Introduction

In the 1980s and the 1990s, the so called ‘endogenous growth theory’ or ‘new growth theory’ flourished. In such an approach, the equilibrium growth rate is determined endogenously by means of the endogenous technological progress through investment on R&D (Research and Development) (cf. Romer 1990) or investment on human capital (cf. Lucas 1988).[1] We can consider that the current fashionable version of such a theory is the development of the so called ‘neoclassical growth theory’, which is characterized by the full employment of labor and the full utilization of capital stock. In fact, all of the above mentioned models are based on the approach of fully optimizing economic agents with rational expectations without disequilibrium in any market including labor market. Such an approach neglects Keynesian situations which is characterized by the involuntary unemployment and underutilization of the productive capital stock. Furthermore, usually such an approach cannot explain the endogenous business cycle without resorting to the exogenous random shocks.[2]

In this paper, we develop an alternative analytical framework of the ‘Keynesian endogenous growth cycle theory’. In fact, the Keynesian tradition of the ‘endogenous growth theory’ is older than the neoclassical tradition of such a theory. We can trace back to Kaldor (1957)(1960) as the source of the ‘Keynesian endogenous growth theory’. Kaldor (1957)(1960) introduced the ‘technical progress function’ to interpret the growth of labor productivity endogenously by means of firms’ investment activities. According to this approach, technical progress is induced by investment expenditure, which means that the rate of technical progress and the ‘natural’ rate of growth are no more independent of the effective demand, because the investment expenditure is an important component of the effective demand. In this approach, the endogenous fluctuations of the investment expenditure induce the endogenous fluctuations of the ‘natural’ rate of growth as well as the ‘actual’ rate of growth, the rate of employment and the rate of capacity utilization of capital stock. Although Kaldor (1957)(1960)’s original formulations assumed the full employment of labor, this somewhat ‘anti-Keynesian’ feature is not the essential aspect of his approach. In fact, Palley (1996) developed a ‘Keynesian-Kaldorian’ endogenous growth theory without full employment. However, Palley (1996)’s model is not complete in the sense that it does not integrate monetary factors and nominal price dynamics into such an analytical framework.

The present paper develops a full fledged Keynesian-Kaldorian model of endogenous growth cycle, which incorporates monetary and nominal price dynamics as well as both of forward-looking and backward-looking inflation expectation dynamics consistently. The analytical framework utilized in this paper is the ‘high-dimensional disequilibrium dynamic Keynesian model’ which was developed by Chiarella and Flaschel (2000), Asada, Chiarella, Flaschel and Franke (2003), Asada (2006) etc., which is an extension of Goodwin (1967)’s model of endogenous business cycle that uses a kind of Phillips curve.[3] While Goodwin (1967)’s original model does not consider monetary and nominal price dynamics, prototype ‘high-dimensional disequilibrium dynamic Keynesian model’ incorporates such factors consistently. Nevertheless, in both models the trend rate of growth (‘natural’ rate of growth)

[1] For extensive exposition of such an analytical framework, see Barro and Sala-i-Martin(1995).

[2] For example, Asada, Semmler and Novak(1998) proved that in Romer(1990)’s model the economic variables converge to the steady state values monotonically without cyclical fluctuations.

[3] ‘High-dimensional’ dynamic model means the dynamic model with many(in fact, more than two) endogenous variables.

is given exogenously from the outside of the system. On the other hand, in our model the 'natural' rate of growth itself as well as 'actual' rate of growth fluctuates endogenously around the endogenously determined equilibrium rate in the course of the endogenous business cycle. In this sense, our model is a true model of endogenous growth cycle. Spirit of our model is Keynesian in the sense that full employment of labor is not assumed and the rate of employment as well as the rate of utilization of capital stock is an endogenous variable.

In section 2, we present an analytical framework of the model. In section 3, we analyze the model mathematically to investigate the impact of the changes of some critical parameter values on the macroeconomic performance such as dynamic stability, instability, and cyclical fluctuations. In section 4, we present some numerical simulations which support the analytical results in section 3. In section 5 which is devoted to concluding remarks, we refer to the relevance of our model to the analysis of actual macroeconomic problems, in particular, the analysis of the Japanese economy in the 1980s – the 2000s. Some complicated mathematical formulae and proofs are relegated to the appendices.

2. Formulation of the Model

In this section, we present a series of equations which constitute the building blocks of our model.

2.1. Dynamics of the Goods Market

We assume the following Keynesian quantity adjustment process of the disequilibrium in the goods market (cf. Keynes1936, Asada1991, 2006, 2008, and Yoshida and Asada2007).[4]

$$\dot{y} = \alpha(c + h + g - y) \; ; \; \alpha > 0, \; y = Y/K, \; c = C/K, \; h = I/K, \; g = G/K \tag{1}$$

where Y = real output (real national income), K = real capital stock, C = real private consumption expenditure, I = real private investment expenditure, G = real government expenditure, α = adjustment speed in the goods market, y = output-capital ratio, c = consumption-capital ratio, h = rate of private investment (private investment-capital ratio), g = government expenditure-capital ratio, and dot over the symbol denotes the derivative with respect to time. As for the production technology, we simply assume the fixed coefficient model. Namely, we assume that the output-capital ratio at the full capacity utilization, which is denoted as y_f, is a positive constant. Nevertheless, in our model the actual output-capital ratio (y) is a variable that reflects the variation of the rate of capital utilization of the capital stock, because in our model the full utilization of the capital stock is not necessarily ensured. More accurately, we have

$$y = uy_f \tag{2}$$

[4] In this formulation, we neglect the international trade for simplicity.

where u is the rate of capital utilization $(0 \leqq u \leqq 1)$. In other words, we can consider that in our model the output-capital ratio is a surrogate variable of the rate of capacity utilization of the capital stock.

As for the determination of c, h, and g, we assume the following quite conventional standard relationships.

$$c = c(y, \pi^e) \ ; \ 0 < c_y = \partial c / \partial y < 1, \ c_\pi = \partial c / \partial \pi^e \geqq 0 \tag{3}$$

$$h = h(y, r - \pi^e) \ ; \ h_y = \partial h / \partial y > 0, \ h_{r-\pi} = \partial h / \partial (r - \pi^e) < 0 \tag{4}$$

$$g = g_0 + \delta(\bar{e} - e) \ ; \ g_0 > 0, \ \delta \geqq 0, \ 0 < \bar{e} < 1 \tag{5}$$

where $\pi^e =$ expected rate of price inflation, $r =$ nominal rate of interest, $r - \pi^e =$ expected real rate of interest, $e =$ rate of employment $= 1 -$ rate of unemployment, and $\bar{e} =$ 'natural' rate of employment.

Eq. (3) is a quite standard Keynesian consumption function, which says that current consumption is the increasing function of the current income and the non-decreasing function of the expected rate of inflation.[5] Eq. (4) is a standard Keynesian investment function, which can be derived from firms' dynamic optimizing behavior and it is also consistent with 'Tobin's q theory'.[6] Eq. (5) formalizes the government's counter-cyclical fiscal stabilization policy.[7]

2.2. Dynamics of the Money Market

We assume that the adjustment speed in the money market is fast enough to establish the equilibrium in the money market instantaneously. This means that in our model the following 'LM equation' is always satisfied.

$$m = l(y, r) \ ; \ m = M/(pK), \ l = L/K, \ l_y = \partial l / \partial y > 0, \ l_r = \partial l / \partial r < 0 \tag{6}$$

---

[5] In general, the current consumption also depends on the expected future income. This is the standard conclusion on the consumption behavior which is derived from consumers' dynamic optimization hypothesis. In our model, however, we simplify the consumption function by eliminating the expected future income assuming implicitly that the expected future income is the increasing function of the current income. In addition, the current consumption will also be the decreasing function of the rate of income tax. But, we need not treat this fact explicitly here, because in our model the rate of income tax is assumed to be a constant parameter.

[6] As for the theoretical interpretation of this statement, see, for example, Yoshikawa(1980), Tobin(1988), and Ouchi(2008a)(2008b). By the way, in this formulation, the 'debt effect' on investment is not considered. As for the theoretical model with such an effect, see Asada(2006).

[7] In this formulation, we neglect the existence of the policy lag for simplicity. As for the detailed analysis of the policy lag in the dynamic setting, see Yoshida and Asada(2007).

where M = nominal money supply, p = price level, L = real money demand, m = real money supply-capital ratio, and l = real money demand-capital ratio. The right hand side of Eq. (6) is a standard Keynesian money demand function.

By differentiating the definitional expression $m = M/(pK)$ with respect to time, we obtain the following expression.

$$\dot{m}/m = \dot{M}/M - \dot{p}/p - \dot{K}/K = \mu - \pi - \dot{K}/K \tag{7}$$

where $\mu = \dot{M}/M$ = rate of increase of money supply, and $\pi = \dot{p}/p$ = actual rate of price inflation. We assume that μ is a positive constant that is determined by the monetary authority (the central bank).

2.3. Dynamics of Wage, Prices and Employment

In this subsection, we consider the dynamics of nominal wage rate, price level and rate of employment. We assume that the dynamic of the labor market is expressed by the following quite standard version of the expectations-augmented wage Phillips curve, which is considered to be the result of wage bargaining between capitalists and workers.

$$\dot{w}/w = \beta(e - \bar{e}) + \dot{a}/a + \pi^e \ ; \ a = Y/N \tag{8}$$

where w = nominal wage rate, N = labor employment, a = average labor productivity, and β is a positive parameter that represents the speed of wage adjustment.

As for the pricing behavior of the firms, we assume the following simple hypothesis of the mark up pricing of the imperfectly competitive firms.

$$p = (1+v)wN/Y = (1+v)w/a \tag{9}$$

where v is a positive constant. We can interpret $(1+v)$ as the mark up over prime cost that reflects the 'degree of monopoly' of the economy.[8] Differentiating Eq. (9) with respect to time, we have

$$\pi = \dot{p}/p = \dot{w}/w - \dot{a}/a. \tag{10}$$

Substituting Eq. (8) into Eq. (10), we obtain the following equation of 'expectations-augmented price Phillips curve'.

$$\pi = \beta(e - \bar{e}) + \pi^e \tag{11}$$

[8] In this formulation, we neglect the prime cost other than labor cost for simplicity.

Next, let us consider the dynamic behavior of the rate of employment (e), which is defined as

$$e = N / N^s = \frac{(Y/K)K}{(Y/N)N^s} = \frac{yK}{aN^s} \tag{12}$$

where N^s = labor supply. Differentiating this equation with respect to time, we have the following relationship.

$$\dot{e}/e = \dot{y}/y + \dot{K}/K - \dot{a}/a - n_s \; ; \; n_s = \dot{N}^s / N^s \tag{13}$$

where n_s is the rate of growth of labor supply, which is assumed to be a positive constant.

2.4. Capital Accumulation and Endogenous Technological Progress

We assume that a part of the government expenditure is the investment expenditure that contributes to the growth of the real capital stock. More accurately, we write the process of capital accumulation as follows.

$$\dot{K} = I + \theta G \; ; \; 0 < \theta < 1 \tag{14}$$

where θ is assumed to be a constant parameter.[9] Substituting equations (4) and (5) into Eq. (14), we obtain the following expression.

$$\dot{K}/K = I/K + \theta G/K = h(y, r - \pi^e) + \theta\{g_0 + \delta(\bar{e} - e)\} \tag{15}$$

Furthermore, we adopt the Kaldorian concept of the endogenous technological progress which is induced by the capital accumulation (cf. Kaldor 1957, 1960, 1978). Kaldor writes as follows.

> "Any sharp or clear-cut distinction between the movement *along* a 'production function' with a given state of knowledge, and a *shift* in the 'production function' caused by a change in the state of knowledge, is arbitrary and artificial. Hence instead of assuming that some given rate of increase in productivity is attributable to technical progress which is superimposed, so to speak, on the growth of productivity attributable to capital accumulation, we shall postulate a single relationship between the growth of capital and the growth of productivity which incorporates the influence of both factors." (Kaldor 1960, p. 265)

Following Kaldor (1959, 1960, 1978), let us express this idea formally by the following Kaldorian 'technical progress function'.

---

[9] In this formulation, we neglect the capital depreciation for simplicity.

$$\dot{a}/a = \varepsilon(\dot{K}/K) + \varepsilon_0 \ ; \ 0 < \varepsilon < 1, \ \ \varepsilon_0 > 0 \tag{16}$$

Substituting Eq. (15) into Eq. (16), we have

$$\dot{a}/a = \varepsilon[h(y, r - \pi^e) + \theta\{g_0 + \delta(\bar{e} - e)\}] + \varepsilon_0 \equiv \varphi(y, r - \pi^e, e; \delta, \theta), \tag{17}$$

which means that in our model the rate of technical progress is an endogenous variable which is influenced by several economic factors.

In this case, we can express the 'natural rate of growth' (n), which is defined as 'the rate of technical progress plus the growth rate of labor supply', as follows.

$$n = \dot{a}/a + n_s = \varphi(y, r - \pi^e, e; \delta, \theta) + n_s \equiv n(y, r - \pi^e, e; \delta, \theta) \tag{18}$$

This equation means that the 'natural rate of growth' itself endogenously fluctuates during the process of the business cycle depending on the fluctuations of the effective demand in our model.[10] This formulation is consistent with the empirical data of some countries such as USA, UK, Japan and Germany (cf. Leon-Ledesma and Thirlwall 2007).

2.5. Inflation Expectation Formation Hypothesis

At this point, to close the model, we must specify how the inflation expectation is formed. The fashionable 'solution' in recent times is to adopt the 'rational expectation' hypothesis, which is equivalent to assume that all economic agents are quite outstanding econometricians. In fact, in the non-stochastic setting, the rational expectation hypothesis is nothing but the perfect foresight hypothesis. In our complicated bounded-rational disequilibrium dynamic model, this type of 'solution' is implausible from the viewpoint of realism. On the other hand, a weaker version of the rational expectation hypothesis is presented in several literatures such as Stein (1971), Asada (1991), and Chiarella and Flaschel (2000). According to their approach, it is not assumed that the economic agents know the current rate of inflation correctly even in the non-stochastic setting, but it is assumed that the economic agents know the long run equilibrium rate of inflation correctly and they use this information to form their inflation expectations.[11] Next, let us consider this 'weak' version of rational expectation hypothesis in the context of our model.

At the long run equilibrium, we have

$$\dot{m} = \dot{e} = \dot{y} = 0 \tag{19}$$

by definition. Substituting Eq. (19) into equations (7) and (13), we have

[10] Note that the private investment expenditure and the government expenditure are two important components of the effective demand.

[11] Asada(1991) called this hypothesis 'quasi rational' expectation hypothesis, and Chiarella and Flaschel(2000) called it 'p-star' expectation hypothesis.

$$\pi = \mu - (\dot{a}/a + n_s) = \mu - n, \tag{20}$$

which means that at the long run equilibrium the rate of price inflation is equal to the difference between the growth rate of nominal money supply and the 'natural rate of growth'.

In our model, the 'natural rate of growth' is an endogenous variable rather than constant. Therefore, the economic agent must know the equilibrium natural rate of growth (n^*) in order to know the equilibrium rate of inflation. The present 'weak' version of the rational expectation hypothesis (or the 'quasi rational' expectation hypothesis) assumes that in fact they know the value of n^*. This will be the case if the government or the central bank has the knowledge on the correct value of n^* and announces the target rate of inflation as the rate $\mu - n^*$ to the public and the economic agents form their inflation expectation in the 'forward looking' manner by using this information in the following way.[12]

$$\dot{\pi}^e = \gamma(\mu - n^* - \pi^e) \ ; \ \gamma > 0 \tag{21}$$

where γ is the speed of the expectation adaptation.

On the other hand, Keynes (1936) asserted that the economic agents will be forced to form their expectations on the shaky ground of *convention* in the situation of true uncertainty. Keynes wrote as follows.

> "In practice we have tacitly agreed, as a rule, to fall back on what is, in truth, a *convention*. The essence of this convention – though it does not, of course, work out quite so simply – lies in assuming that the existing state of affairs will continue indefinitely, except in so far as we have specific reasons to expect a change. This does not mean that we really believe that the existing state of affairs will continue indefinitely. We know from extensive experience that this is most unlikely. ……Nevertheless the above conventional method of calculation will be compatible with a considerable measure of continuity and stability in our affairs, *so long as we can rely on the maintenance of the convention*." (Keynes 1936, p. 152)

The old fashioned adaptive expectation hypothesis rather than various versions of rational expectation hypothesis will capture the essence of such a Keynesian notion. If we apply the adaptive expectation hypothesis to the inflation expectation formation, we have the following formula.

$$\dot{\pi}^e = \gamma(\pi - \pi^e) \ ; \ \gamma > 0 \tag{22}$$

where γ is the speed of expectation adaptation. This equation implies that the economic agents form the inflation expectation based on their experience in the recent past, and they revise their expectation sluggishly when there is the difference between the actual and the expected rates of inflation.

---

[12] Later, we shall see that in fact we have $n^* = \dfrac{\varepsilon_0 + n_s}{1 - \varepsilon} > 0$.

In this paper, we adopt the following mixture of the 'adaptive' (or backward looking) and the 'quasi rational' (or forward looking) expectation hypotheses.

$$\dot{\pi}^e = \gamma[\xi(\mu - n^* - \pi^e) + (1-\xi)(\pi - \pi^e)] \ ; \ \gamma > 0, \ \ 0 < \xi < 1 \tag{23}$$

We can interpret that the parameter ξ is the degree of 'credibility' of the announcement by government or the central bank (cf. Asada 2006, 2008).

2.6. System of Fundamental Dynamical Equations

Now, we are in a position to derive the fundamental dynamical equations in our model. First, we have the following equation substituting equations (3) – (5) into Eq. (1).

$$\dot{y} = \alpha[c(y,\pi^e) + h(y, r - \pi^e) + g_0 + \delta(\bar{e} - e) - y] \tag{24}$$

Second, we have the following equation substituting equations (11) and (15) into Eq. (7).

$$\dot{m} = m[\mu - \beta(e - \bar{e}) - \pi^e - h(y, r - \pi^e) - \theta\{g_0 + \delta(\bar{e} - e)\}] \tag{25}$$

Third, we have the following equation substituting equations (15) and (17) into Eq. (13).

$$\dot{e} = e[\dot{y}/y + (1-\varepsilon)h(y, r - \pi^e) + (1-\varepsilon)\theta\{g_0 + \delta(\bar{e} - e)\} - \varepsilon_0 - n_s] \tag{26}$$

Fourth, we have the following equation substituting Eq. (11) into Eq. (23).

$$\dot{\pi}^e = \gamma[\xi(\mu - n^* - \pi^e) + (1-\xi)\beta(e - \bar{e})] \tag{27}$$

Fifth, we have the following expression solving Eq. (6) with respect to r.

$$r = r(y,m) \ ; \ r_y = -l_y / l_r > 0, \ \ r_m = 1/l_r < 0 \tag{28}$$

Equations (24) – (28) are enough to determine the dynamics of five main endogenous variables y, m, e, π^e, and r.

Substituting Eq. (28) into equations (24) – (26), we obtain the following four dimensional system of nonlinear differential equations, which may be called the system of fundamental dynamical equations in our model.

(i) $\dot{y} = \alpha[c(y,\pi^e) + h(y, r(y,m) - \pi^e) + g_0 + \delta(\bar{e} - e) - y] \equiv F_1(y, m, e, \pi^e; \alpha, \delta)$

(ii) $\dot{m} = m[\mu - \beta(e - \bar{e}) - \pi^e - h(y, r(y,m) - \pi^e) - \theta\{g_0 + \delta(\bar{e} - e)\}]$

$\equiv F_2(y, m, e, \pi^e; \beta, \delta, \theta)$

(iii) $\dot{e} = e[F_1(y,m,e,\pi^e;\alpha,\delta)/y + (1-\varepsilon)h(y,r(y,m)-\pi^e) + (1-\varepsilon)\theta\{g_0 + \delta(\bar{e}-e)\}$
$-\varepsilon_0 - n_s] \equiv F_3(y,m,e,\pi^e;\alpha,\delta,\theta,\varepsilon)$

(iv) $\dot{\pi}^e = \gamma[\xi(\mu - n^* - \pi^e) + (1-\xi)\beta(e-\bar{e})] \equiv F_4(e,\pi^e;\beta,\gamma,\xi)$ (29)

In section 3 – 1, it will be shown that the long run equilibrium solution of this system is consistent only in case of

$$n^* = \frac{\varepsilon_0 + n_s}{1-\varepsilon} > 0. \tag{30}$$

We *assume* that Eq. (30) is in fact satisfied.

3. Mathematical Analysis of the Model

In this section, we investigate the characteristics of the solution of the system (29) mathematically.

3.1. Characteristics of the Long Run Equilibrium Solution

First, let us consider the characteristics of the long run equilibrium solution which satisfies

$$\dot{y} = \dot{m} = \dot{e} = \dot{\pi}^e = 0, \ \ e = \bar{e}. \tag{31}$$

Substituting Eq. (31) into Eq. (29), we obtain the following set of equations.

(i) $c(y,\pi^e) + h(y,r(y,m)-\pi^e) + g_0 - y = 0$

(ii) $\mu - \pi^e - \{h(y,r(y,m)-\pi^e) + \theta g_0\} = 0$

(iii) $(1-\varepsilon)\{h(y,r(y,m)-\pi^e) + \theta g_0\} - \varepsilon_0 - n_s = 0$

(iv) $e = \bar{e}$

(v) $\pi^e = \mu - n^*$ (32)

These equations imply that the long run equilibrium solution is determined *independent* of the parameters α, β, γ, δ, and ξ. We can solve this system of equations as follows. First, let us rewrite Eq. (32) (i) as

$$c(y,\pi^e) + g_0 - y = -h(y,r(y,m)-\pi^e). \tag{33}$$

Substituting this equation into Eq. (32) (ii), we have

$$\mu - \pi^e + c(y, \pi^e) + (1-\theta)g_0 - y = 0. \tag{34}$$

Furthermore, substituting Eq. (32) (ii) into Eq. (32) (iii), we obtain

$$(1-\varepsilon)(\mu - \pi^e) - \varepsilon_0 - n_s = 0. \tag{35}$$

Solving this equation with respect to π^e, we obtain the following explicit solution of the equilibrium (expected and actual) rate of inflation.

$$\pi^e * = \pi * = \mu - \frac{\varepsilon_0 + n_s}{1-\varepsilon} \tag{36}$$

Comparing equations (32) (v) and (36), we can conclude that the long run equilibrium solution is consistent only if Eq. (30) is satisfied. Substituting Eq. (36) into Eq. (34), we have

$$f(y) \equiv \frac{\varepsilon_0 + n_s}{1-\varepsilon} + c(y, \mu - \frac{\varepsilon_0 + n_s}{1-\varepsilon}) + (1-\theta)g_0 - y = 0. \tag{37}$$

This is an equation in which only unknown quantity is y. It is easy to see that the following relationship is satisfied.

$$f'(y) = c_y - 1 < 1,\ \ f(0) > 0,\ \ \lim_{y \to +\infty} f(y) = -\infty \tag{38}$$

Therefore, by continuity, we have the unique equilibrium solution $y^* > 0$, and y^* depends on monetary and fiscal policy parameters μ and g_0. More accurately, we have

$$y^* = y^*(\mu, g_0)\ ;\ y_\mu^* = \partial y^* / \partial \mu \gtreqqless 0,\ \ y_g^* = \partial y^* / \partial g_0 > 0. \tag{39}$$

We *assume* that μ and g_0 are small enough to ensure that $0 < y^* < y_f$.

Finally, we obtain the equilibrium value of m, which is denoted by m^*, substituting equations (36) and (39) into Eq. (33). The mechanism of the determination of m^* is as follows.

At the long run equilibrium point, the real rate of interest (R) and the nominal rate of interest (r) must satisfy the following relationships.

$$h(y^*(\mu, g_0), R) = y^*(\mu, g_0) - c(y^*(\mu, g_0), \mu - \frac{\varepsilon_0 + n_s}{1-\varepsilon}) - g_0 \tag{40}$$

$$r(y^*(\mu, g_0), m) = R + \mu - \frac{\varepsilon_0 + n_s}{1-\varepsilon} \tag{41}$$

Solving Eq. (40) with respect to R, the equilibrium value R^* is uniquely determined. Substituting $R = R^*$ into the right hand side of Eq. (41), we obtain the equilibrium value of nominal rate of interest (r^*), and the equilibrium value of the real money-capital ratio (m^*), that is consistent with Eq. (41), is uniquely determined.[13]

From equations (20) and (36), we can see that the equilibrium value of the natural rate of growth (n^*) becomes as follows.

$$n^* = \frac{\varepsilon_0 + n_s}{1-\varepsilon} > 0 \tag{42}$$

In this case, the rate of technical progress at the long run equilibrium $(\dot{a}/a)^*$ becomes

$$(\dot{a}/a)^* = n^* - n_s = \frac{\varepsilon_0 + \varepsilon n_s}{1-\varepsilon} > 0, \tag{43}$$

which means that the equilibrium rate of technical progress is not independent of the growth rate of labor supply. Next, substituting Eq. (43) into the technical progress function (Eq. (16)) and solving with respect to $\dot{K}/K$, we obtain the following solution of the equilibrium rate of capital accumulation $(\dot{K}/K)^*$.

$$(\dot{K}/K)^* = \frac{\varepsilon_0 + n_s}{1-\varepsilon} = n^* \tag{44}$$

Furthermore, it follows from equations (15), (32) (iv) and (44) that we have the following solution of the equilibrium rate of private investment expenditure (h^*).

$$h^* = \frac{\varepsilon_0 + n_s}{1-\varepsilon} - \theta g_0 = n^* - \theta g_0 \tag{45}$$

Eq. (45) means that the 'crowding out' phenomenon occurs at the long run equilibrium point. Namely, the increase of the government's investment expenditure (θg_0) does not

[13] If $r^* < 0$, the long run equilibrium does not exist because of the nonnegative constraint of the nominal rate of interest. This is the situation of the so called 'liquidity trap'(cf. Asada 2006, 2008). This situation is likely to occur when the central bank selects too small μ compared with ε_0, ε, and n_s. From now on, we assume that μ is large enough to ensure $r^* > 0$.

affect the equilibrium rate of capital accumulation but it induces the corresponding reduction of the equilibrium rate of private investment.

By the way, it is well known that Kaldor (1978) presented the following six 'stylized facts' on the long run tendency of the capitalist countries 'as a starting-point for the construction of theoretical models' (Kaldor 1978, p. 2).

(1) The continued growth in the aggregate volume of production and in the productivity of labor at a steady trend rate.
(2) A continued increase in the amount of capital per worker.
(3) A steady rate of profit on capital.
(4) Steady capital-output ratio over long periods; at least there are no clear long-term trends, either rising or falling.
(5) A steady share of profits (and of wages).
(6) There are appreciable differences in the rate of growth of labor productivity and of total output in different societies.

Now, we can easily prove the following proposition.

Proposition 1.
The long run equilibrium solution of the system (29) is consistent with all of Kaldor's 'stylized facts' (1) – (6).

[Proof.]

(1) Equations (39) and (44) mean that at the equilibrium point $y = Y/K$ is constant and K increases steadily. This implies that Y also increases steadily. On the other hand, Eq. (43) implies that $a = Y/N$ increases steadily.

(2) By definition we have $K/N = (\frac{Y}{N})/(\frac{Y}{K}) = a/y$, while a increases steadily and y is constant from equations (43) and (39). This means that K/N increases steadily.

(3) We can define rate of profit on capital (ρ) as

$$\rho = \frac{pY - wN}{pK} = \frac{Y - \omega N}{K} = y - \frac{\omega}{K/N},$$

where $\omega = w/p$ is the real wage rate. In addition, from Eq. (10) we have $\dot{\omega}/\omega = \dot{w}/w - \dot{p}/p = \dot{a}/a$, which means that the rate of growth of the real wage rate is the same as the rate of technical progress. On the other hand, it follows from the proof of the statement (2) that y is constant and the rate of growth of (K/N) is equal to $\dot{a}/a$ at the long run equilibrium point. In such a situation, ρ must be constant.

(4) It follows from Eq. (39) that $K/Y = 1/y$ is constant at the long run equilibrium point.

(5) The share of profit relative to wages can be expressed as $\frac{pY - wN}{wN} = v =$ constant from the pricing equation (9).

(6) The differences of $(\dot{a}/a)^*$ and $(\dot{Y}/Y)^*$ between different societies can be explained by the differences of the parameter values ε, ε_0 and n_s between different societies. □

3.2. Local Stability/Instability and Existence of Cyclical Fluctuations

Next, let us investigate the out of steady state dynamics of the system. For this purpose, in this subsection, we study the local stability/instability of the equilibrium point and whether cyclical fluctuations around the equilibrium point exist or not. We can write the Jacobian matrix of the system (29), which is evaluated *at the equilibrium point*, as follows.

$$J = \begin{bmatrix} \alpha H_{11} & \alpha H_{12} & -\alpha\delta & \alpha H_{14} \\ mH_{21} & -mH_{12} & m(\theta\delta - \beta) & mH_{24} \\ eH_{31} & eH_{32} & -e\{\alpha / y + (1-\varepsilon)\theta\}\delta & eH_{34} \\ 0 & 0 & \gamma(1-\xi)\beta & -\gamma\xi \end{bmatrix} \tag{46}$$

The detailed expressions of the partial derivatives H_{ij} are contained in **APPENDIX A**. We can write the characteristic equation of this Jacobian matrix as follows.

$$\Delta(\lambda) \equiv |\lambda I - J| = \lambda^4 + b_1\lambda^3 + b_2\lambda^2 + b_3\lambda + b_4 = 0 \tag{47}$$

The coefficients of this characteristic equation are contained in **APPENDIX B**.

It is well known that the equilibrium point of this system is locally asymptotically stable *if and only if* the following set of inequalities is satisfied (cf. Asada, Chiarella, Flaschel and Franke 2003, p. 519).[14]

$$b_j > 0 \ \ (j = 1,2,3,4), \ \ \Phi \equiv b_1 b_2 b_3 - b_1{}^2 b_4 - b_3{}^2 > 0 \tag{48}$$

The following proposition provides us with a set of economically meaningful *sufficient* conditions for local *instability*.

Proposition 2.

Suppose that $\xi \in [0,1]$ is sufficiently close to 0 and $\gamma > 0$ is sufficiently large. Then, the equilibrium point of the system (29) becomes locally unstable.

---

[14] Inequalities (48) are called the ‘Routh-Hurwitz conditions’ for stable roots in case of the four-dimensional system(cf. Asada, Chiarella, Flaschel and Franke 2003, p. 519).

[Proof.]

Suppose that $\xi = 0$. In this case, it follows from Eq. (B2) in **APPENDIX B** that b_2 becomes a linear decreasing function of γ, so that we have $b_2 < 0$ for all sufficiently large values of $\gamma > 0$, which violates one of the local stability conditions (48). By continuity, we have the same qualitative conclusion even if $0 < \xi < 1$ as long as ξ is sufficiently close to zero. □

This proposition means that the high speed of the expectation adaptation is a destabilizing factor if the expectation formation is highly adaptive (or backward-looking). Next, we shall consider some sets of the *sufficient* conditions for local *stability*. For this purpose, we assume as follows.[15]

Assumption 1.

$$0 < \underset{(+)}{h_y} + \underset{(-)}{h_{r-\pi}} \underset{(+)}{r_y} < (1 - \underset{(+)}{c_y}).$$

This assumption means that the 'marginal propensity to invest' that includes the indirect effect through money market $(h_y + h_{r-\pi} r_y)$ is positive but it is less than the 'marginal propensity to save' $(1 - c_y)$. It is easy to see that we have the following set of inequalities under this assumption (cf. **APPENDIX A** and Eq. (B1) in **APPENDIX B**).

$$H_{11} < 0, \ \ H_{21} < 0, \ \ b_1 > 0 \tag{49}$$

Now, we can prove the following important propositions.

Proposition 3.
Under **Assumption 1**, we have the following properties.

(1) Suppose that $\alpha > 0$ and $\gamma > 0$ are sufficiently small. Then, the equilibrium point of the system (29) is locally asymptotically stable irrespective of the value of $\xi \in [0,1]$.

(2) Suppose that $\alpha > 0$ and $\beta > 0$ are sufficiently small, $\theta \in (0,1)$ is sufficiently close to 1, and $\xi \in [0,1]$ is sufficiently close to 1. Then, the equilibrium point of the system (29) is locally asymptotically stable irrespective of the value of $\gamma > 0$.

[Proof.] See **APPENDIX C.**

Proposition 4.

---

[15] It is worth noting that we did not use **Assumption 1** to derive **Proposition 2**.

Under **Assumption 1**, we have the following properties.

(1) Suppose that $\alpha > 0$ is sufficiently small and $\xi \in [0,1]$ is sufficiently close to 0. Then, (i) the equilibrium point of the system (29) is locally asymptotically stable for all sufficiently small values of $\gamma > 0$, (ii) it is locally unstable for all sufficiently large values of $\gamma > 0$, and (iii) cyclical fluctuations occur at some intermediate values of $\gamma > 0$.

(2) Suppose that $\alpha > 0$ and $\beta > 0$ are sufficiently small, $\theta \in (0,1)$ is sufficiently close to 1, and $\gamma > 0$ is sufficiently large. Then, (i) the equilibrium point of the system (29) is locally unstable for all values $\xi \in [0,1]$ which are sufficiently close to 0, (ii) it is locally asymptotically stable for all values $\xi \in [0,1]$ which are sufficiently close to 1, and (iii) cyclical fluctuations occur at some intermediate values of $\xi \in [0,1]$.

[Proof.]

We shall prove only **Proposition 4** (1). The method of the proof of **Proposition 4** (2) is almost the same as that of **Proposition 4** (1). **Proposition 4** (1) (i) directly follows from **Proposition 3** (1). **Proposition 4** (1) (ii) directly follows from **Proposition 2**. The proof of **Proposition 4** (1) (iii) is as follows.

Since the equilibrium point is locally asymptotically stable for all sufficiently small values of $\gamma > 0$ and it is locally unstable for all sufficiently large values of $\gamma > 0$, by continuity, there exists at least one 'bifurcation point' $\gamma_0 > 0$ at which the qualitative characteristics of the system switch. Obviously, at that 'bifurcation point', the real part of at least one root of the characteristic equation (47) must be zero. By the way, the characteristic equation (47) does not have the real root $\lambda = 0$ because of the fact that $\Delta(0) = b_4 > 0$ (cf. equations (47) and (B4)). This means that Eq. (47) has at least one pair of pure imaginary roots. If this equation has only one pair of pure imaginary roots at the point $\gamma = \gamma_0$, that point is the 'Hopf bifurcation point', and the existence of the non-constant closed orbits is ensured at some parameter values $\gamma > 0$ which are sufficiently close to γ_0 because of the 'Hopf bifurcation theorem' (cf. Gandolfo 1996, p. 477 and Asada, Chiarella, Flaschel and Franke 2003, p. 521). If Eq. (47) has two pairs of pure imaginary roots at $\gamma = \gamma_0$, that point is not the Hopf bifurcation point so that the existence of the closed orbits is not necessarily ensured. Even in this case, however, the existence the cyclical fluctuations is ensured at some parameter values $\gamma > 0$ which are sufficiently close to γ_0 because of the existence of the (two pairs of) complex roots. □

These propositions imply that under certain conditions, (1) the low speed of expectation adjustment (γ) and the highly forward-looking expectation formation (high weight of ξ) have stabilizing effects, (2) the high speed of expectation adjustment combined with the highly backward-looking or the adaptive expectation formation has destabilizing effect, and (3) at the intermediate values of the parameters γ and/or ξ, the endogenous cyclical fluctuations occur. In this process, the 'natural' rate of growth itself fluctuates endogenously around the endogenously determined equilibrium value $n^* = \dfrac{\varepsilon_0 + n_s}{1-\varepsilon} > 0$. In this sense, our model is a true 'endogenous growth cycle' model.[16]

4. Numerical Simulations

In this section, we report the results of numerical simulations which support the mathematical analysis in the previous section. We select the unit time period as 0.1 year. and we fix the values of the parameters, functional forms and initial conditions as follows.

(1) Fixed parameters

$\alpha = 0.08,\ \beta = 0.15,\ \bar{e} = 0.95,\ \theta = 0.3,\ \delta = 0.03,\ g_0 = 0.08,$

$\mu = 0.61\%$ per unit time period, $n_s = 0.02\%$ per unit time period, $\varepsilon = 0.8,$

$\varepsilon_0 = 0.1\%$ per unit time period.[17]

(2) Functional forms

Consumption function: $c = 0.7y + 0.3\pi^e$.

Investment function: $h = 0.11\{0.18y^{0.7} - (r - \pi^e)\} + 0.16.$

LM equation: $r = (0.5y - 0.4m)\%$ per unit time period.

Government expenditure function: $g = g_0 + \delta(\bar{e} - e) = 0.08 + 0.03(0.95 - e).$

Technical progress function: $\dot{a}/a = \varepsilon(h + \theta g) + \varepsilon_0 = 0.8(h + 0.3g) + 0.1.$

(3) Initial values

$y(0) \cong 2.232,\ m(0) \cong 2.405,\ e(0) = 0.96,$

$\pi^e(0) = 0.02\%$ per unit time period.

Furthermore, we select two parameters γ and ξ as bifurcation parameters.

Substituting $\varepsilon = 0.8,\ \varepsilon_0 = 0.1,$ and $n_s = 0.02$ into equations (42) and (43), we obtain the equilibrium values of the 'natural rate of growth' and the 'rate of technical progress' as follows respectively.[18]

$$n^* = 0.6\% \text{ per unit time period, } (\dot{a}/a)^* = 0.58\% \text{ per unit time period.} \tag{50}$$

---

[16] Stabilizing effect of the high weight of ξ implies the effectiveness of the 'inflation targeting' by the central bank(cf. Asada 2006, 2008).

[17] These parameter values imply that the annual growth rates of nominal money supply and labor supply are about 6.1%and 0.2%respectively, and the constant part of the annual growth rate of labor productivity is about 1%. These values are not extremely unrealistic as the values of these parameters in modern capitalist countries such as USA, European countries and some Asian countries such as Japan.

[18] This means that the equilibrium annual rates of 'natural rate of growth' and 'rate of technical progress' are about 6% and 5.8% respectively, which are considerably reasonable values for modern capitalist economies.

Figures 1 – 6 summarize the results of our numerical simulations. These figures correspond to six alternative combinations of the parameter values γ and ξ. Constellation of adopted parameter values is summarized in Table 1. In all examples except that of Figure 6, endogenous cyclical fluctuations emerge.[19]

Table 1. Constellation of parameters

γ \ ξ	0.05	0.12	0.3	0.8
0.07		Figure 2		
0.12	Figure 4	Figure 1	Figure 5	Figure 6
0.2		Figure 3		

Figure 1 represents the base line case which produces the limit cycle. In this case, all main variables fluctuate cyclically indefinitely around the equilibrium values. The period of such a cycle is approximately 8 years, which corresponds to the so called 'major cycle' or 'Juglar cycle'.[20] It is worth noting that the growth rate of labor productivity (the rate of technical progress) $\dot{a}/a$ endogenously fluctuates pro-cyclically, which is caused by the fluctuations of investment expenditures.

Comparison of the figures 1, 2, and 3 reveals that the increase of the adjustment speed of inflation expectation (increase of parameter value γ) in case of highly backward looking or adaptive expectation formation (small value of ξ) tends to *destabilize* the system. Obviously, this result is consistent with **Proposition 2**, **Proposition 3** (1) and **Proposition 4** (1) in section 3 – 2. On the other hand, comparison of the figures 1, 4, 5, and 6 reveals that the increase of the weight of the forward-looking inflation expectation formation or the degree of credibility of the inflation targeting by the central bank (the increase of ξ) tends to *stabilize* the system. Needless to say, this result is consistent with **Proposition 3** (2) and **Proposition 4** (2) in section 3 – 2.[21]

[19] In these figures, we have the equilibrium values $y^* \cong 2.197$, $e^* = \bar{e} = 0.95$ and $(\dot{a}/a)^* = 0.58$.

[20] We have $50/6.5 \cong 8$ years since there are about 6.5 cycles during 50 years.

[21] In figures 3 and 4 in which the equilibrium points are unstable, the rate of employment e exceeds its upper bound 1 at the latter stage of the business cycle. We must introduce exogenous ceiling in order to contain the variables in their economically meaningful region in this case, which introduces additional nonlinearity and additional complexity into the system. In this paper, however, we do not go ahead with the detailed analysis of this problem to keep the analysis as simple as possible.

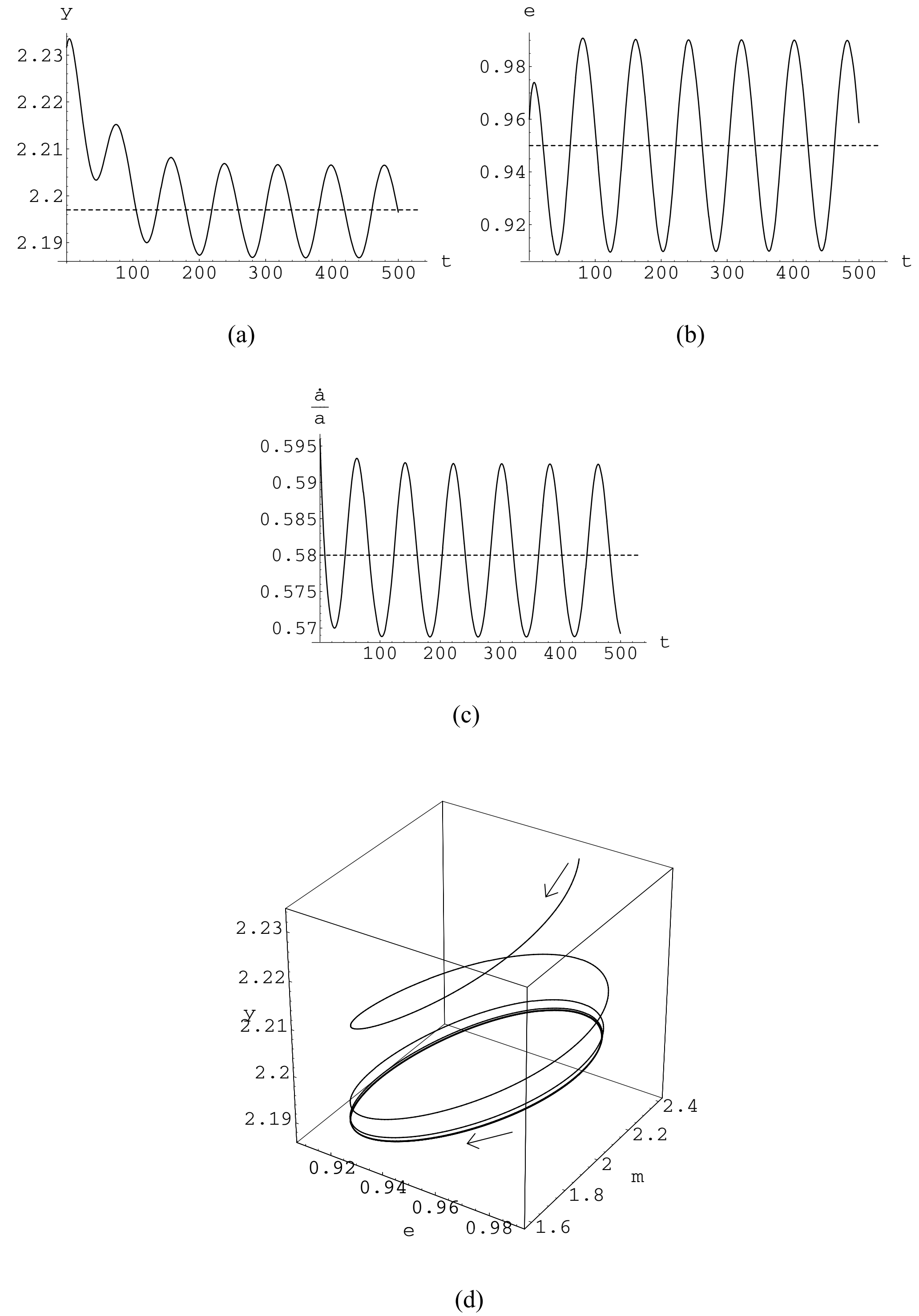
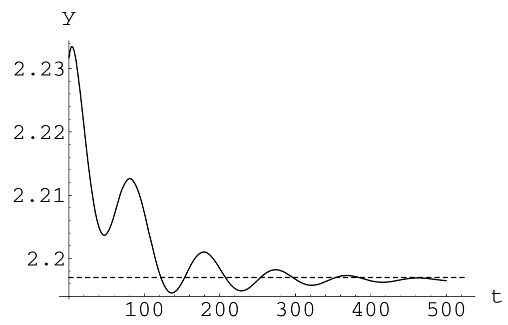
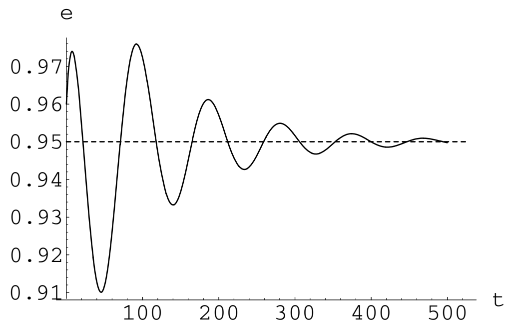


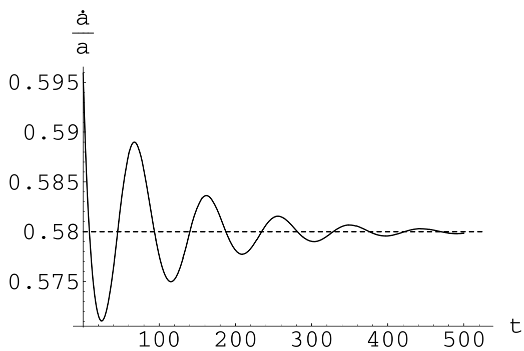
Figure 1. ($\gamma = 0.12, \quad \xi = 0.12$).



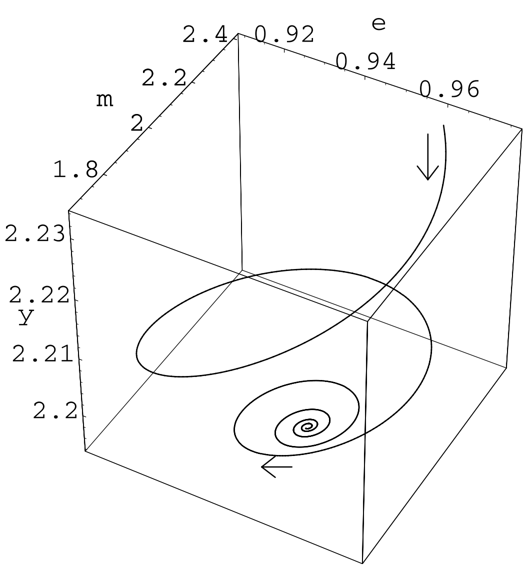
(a)



(b)

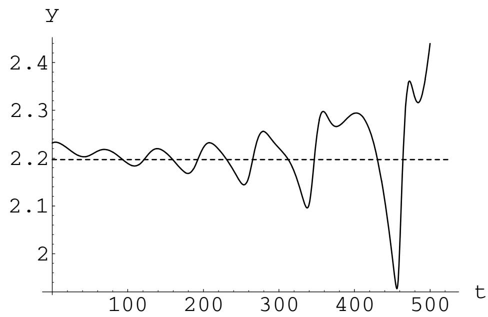


(c)

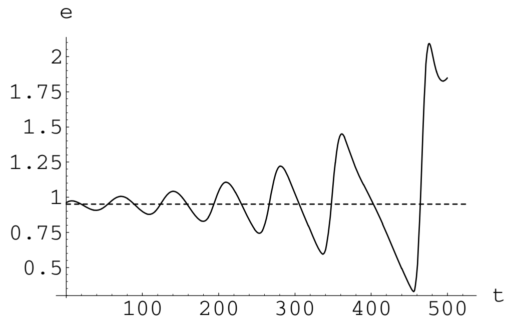


(d)

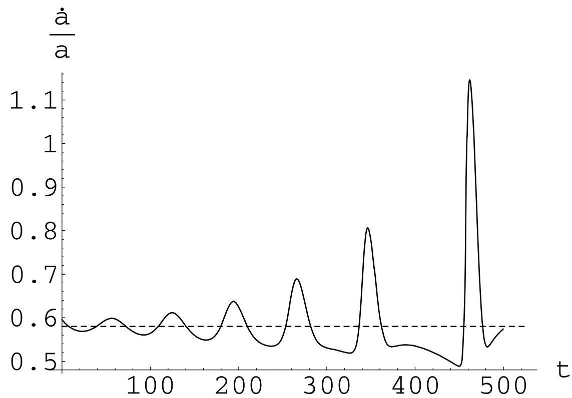
Figure 2. ($\gamma = 0.07, \quad \xi = 0.12$ **).**



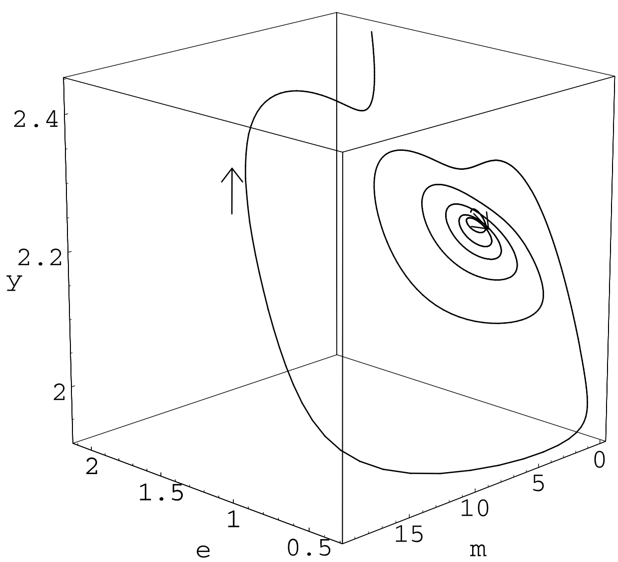
(a)



(b)

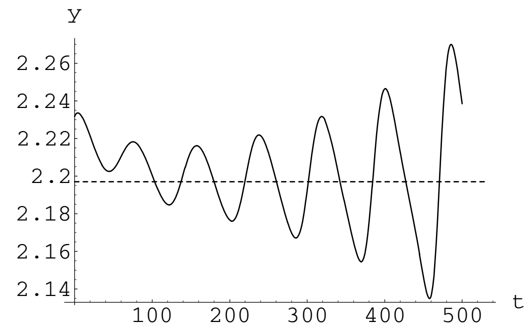


(c)

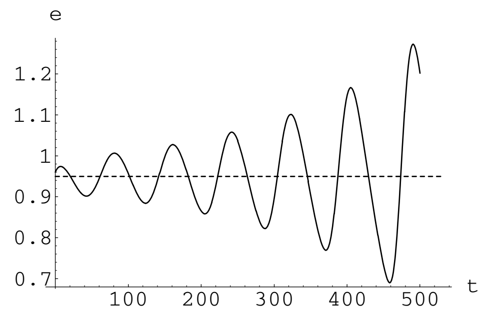


(d)

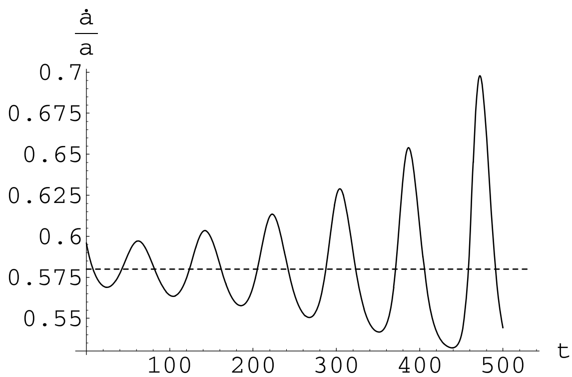
Figure 3. ($\gamma = 0.2, \quad \xi = 0.12$).



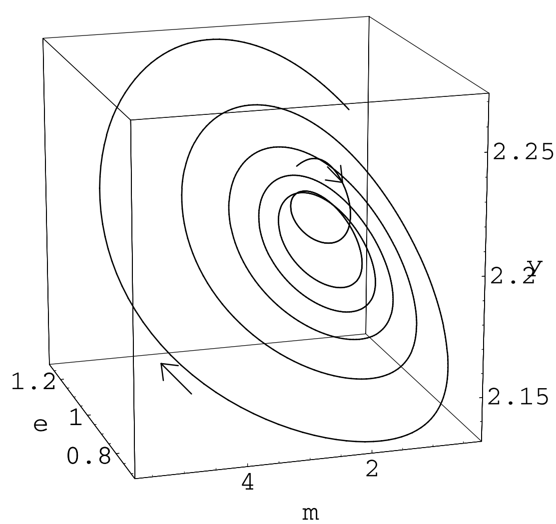
(a)



(b)

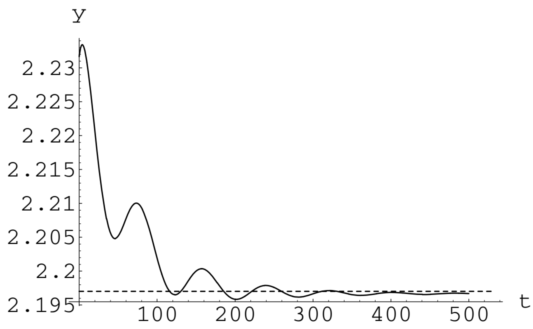


(c)

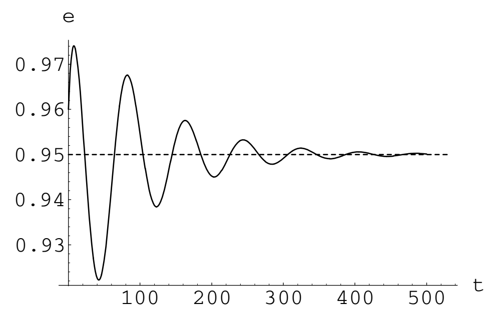


(d)

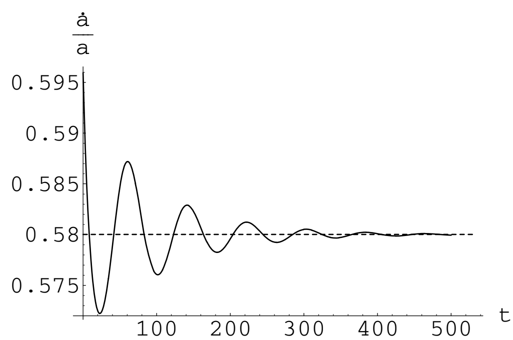
Figure 4. ($\gamma = 0.12, \quad \xi = 0.05$ **).**



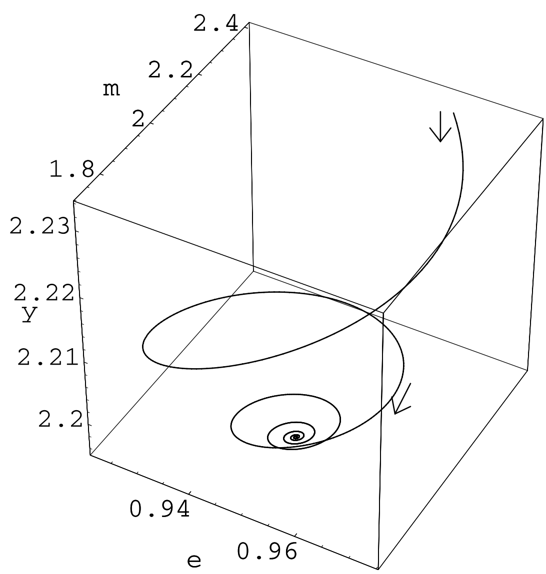
(a)



(b)



(c)



(d)

Figure 5. ($\gamma = 0.12, \quad \xi = 0.3$ **).**

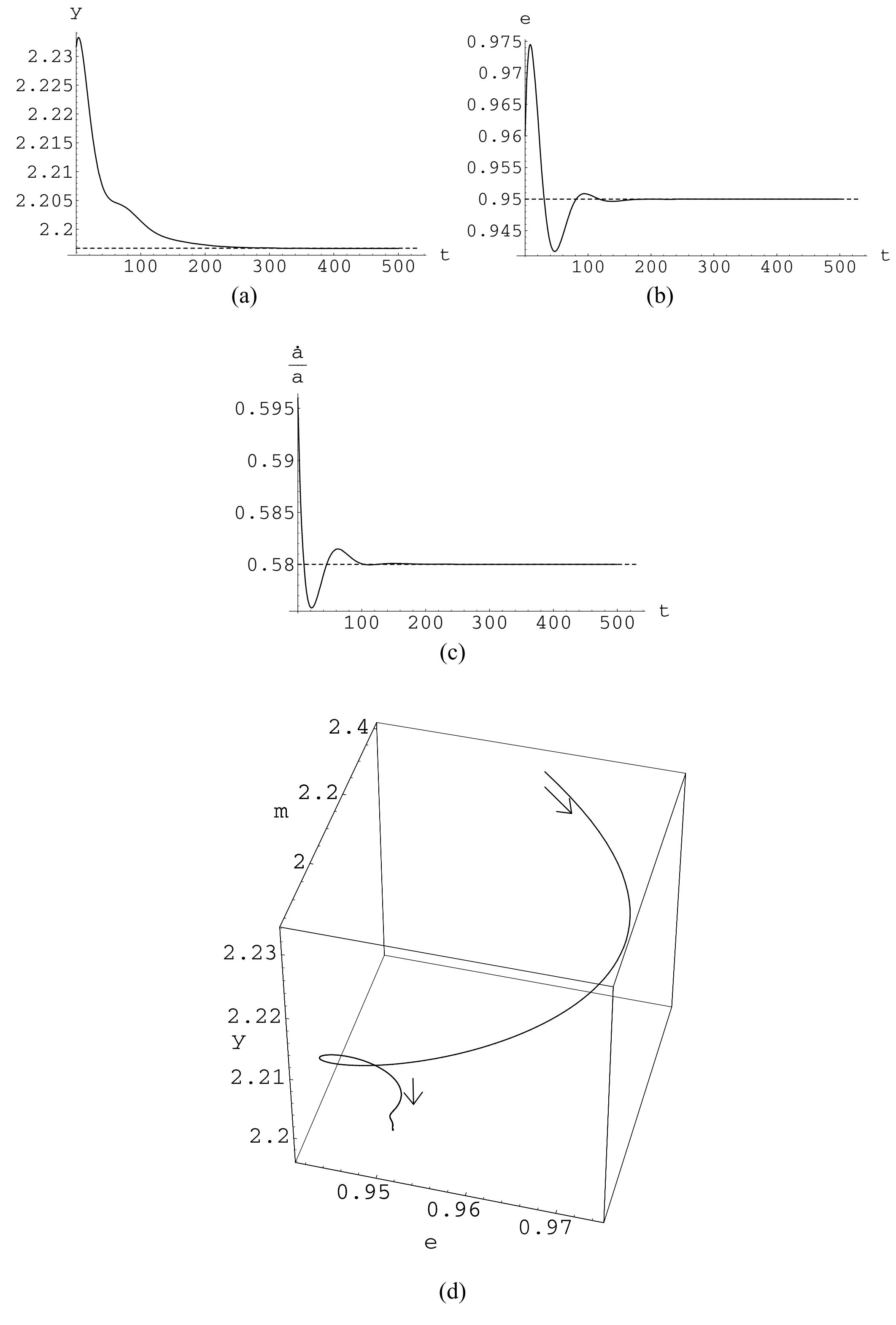


Figure 6. ($\gamma = 0.12, \quad \xi = 0.8$ **).**

5. Concluding Remarks

In this paper, we have developed a Keynesian-Kaldorian model of endogenous growth cycle by means of nonlinear high-dimensional differential equations. In this model, the growth rate of labor productivity (the rate of technical progress) fluctuates endogenously around the endogenously determined equilibrium value during the process of the business cycle, which is induced by endogenous fluctuation of private and public investment expenditures. Since investment expenditures are important components of effective demand, the rate of technical progress is affected by the effective demand in our model. This aspect of our model has relevance to the actual macroeconomic problems, in particular, the analysis of the Japanese economy in the 1880s – the 2000s. Some economists pointed out that the growth rate of labor productivity considerably declined during the serious depression in the 1990s in Japan. Because of this reason, they asserted that the cause of the serious depression of the Japanese economy in the 1990s (so called the 'lost decade') is not the deficiency of the effective demand but the supply side deterioration. Usually such a supply side deterioration is explained by the exogenous 'random shocks'. A typical example of such an assertion is found in Hayashi and Prescott (2002). Our model that has been developed in this paper provides a theoretical foundation to the argument against such an assertion. Our Keynesian-Kaldorian model is consistent with the phenomenon that the growth rate of labor productivity declines during the depression process, but this does not mean that the Keynesian effective demand is irrelevant to the depression, because in our model the productivity deterioration itself is induced by the effective demand decline during the depression.[22]

Acknowledgment

During this research, T. Asada was financially supported by the Japan Society for the Promotion of Science (Grant-in Aid (C) 20530161). Needless to say, however, only the authors are responsible for possible remaining errors.

Appendix A: Partial Derivatives

$$H_{11} = \frac{\partial(F_1/\alpha)}{\partial y} = (\underset{(-)}{c_y} - 1) + \underset{(+)}{h_y} + \underset{(-)}{h_{r-\pi}}\,\underset{(+)}{r_y}. \quad H_{12} = \frac{\partial(F_1/\alpha)}{\partial m} = \underset{(-)}{h_{r-\pi}}\,\underset{(-)}{r_m} > 0.$$

$$H_{14} = \frac{\partial(F_1/\alpha)}{\partial \pi^e} = \underset{(+or\,0)}{c_\pi} - \underset{(-)}{h_{r-\pi}} > 0. \quad H_{21} = \frac{\partial(F_2/m)}{\partial y} = -(\underset{(+)}{h_y} + \underset{(-)}{h_{r-\pi}}\,\underset{(+)}{r_y}).$$

$$H_{24} = \frac{\partial(F_2/m)}{\partial \pi^e} = \underset{(-)}{h_{r-\pi}} - 1 < 0. \quad H_{31} = \frac{\partial(F_3/e)}{\partial y} = \alpha H_{11}/y - (1-\varepsilon)H_{21}.$$

[22] As for more detailed argument on this topic, see Asada(2008).

$$H_{32} = \frac{\partial(F_3/e)}{\partial m} = \{\alpha/y + (1-\varepsilon)\} \underset{(+)}{H_{12}} > 0.$$

$$H_{34} = \frac{\partial(F_3/e)}{\partial \pi^e} = \alpha \underset{(+)}{H_{14}}/y - (1-\varepsilon) \underset{(-)}{H_{24}} > 0.$$

Appendix B: Coefficients of the Characteristic Equation

$$b_1 = -traceJ = -\alpha H_{11} + m \underset{(+)}{H_{12}} + e\{\alpha/y + (1-\varepsilon)\theta\}\delta. + \gamma\xi \equiv b_1(\gamma,\xi) \tag{B1}$$

$b_2 =$ sum of all principal second-order minors of J

$$= \alpha m H_{12}\begin{vmatrix} H_{11} & 1 \\ H_{21} & -1 \end{vmatrix} + \alpha\delta e\begin{vmatrix} H_{11} & -1 \\ H_{31} & -\{\alpha/y + (1-\varepsilon)\} \end{vmatrix} + \alpha\gamma\xi\begin{vmatrix} H_{11} & H_{14} \\ 0 & -1 \end{vmatrix}$$

$$+ me\begin{vmatrix} -H_{12} & \theta\delta - \beta \\ H_{32} & -\{\alpha/y + (1-\varepsilon)\theta\}\delta \end{vmatrix} + m\gamma\xi\begin{vmatrix} -H_{12} & H_{24} \\ 0 & -1 \end{vmatrix}$$

$$+ e\gamma\begin{vmatrix} -\{\alpha/y + (1-\varepsilon)\}\delta & H_{34} \\ (1-\xi)\beta & -\xi \end{vmatrix}$$

$$= -\alpha m \underset{(+)}{H_{12}}(H_{11} + H_{21}) + \alpha\delta e[-H_{11}\{\alpha/y + (1-\varepsilon)\} + H_{31}] - \alpha\gamma\xi H_{11}$$

$$+ me[\underset{(+)}{H_{12}}\{\alpha/y + (1-\varepsilon)\theta\}\delta + (\beta - \theta\delta)\underset{(+)}{H_{32}}] + m\gamma\xi \underset{(+)}{H_{12}}$$

$$+ e\gamma[\xi\{\alpha/y + (1-\varepsilon)\}\delta - (1-\xi)\beta \underset{(+)}{H_{34}}]$$

$$= \alpha\underset{(+)}{(1-c_y)}[m\underset{(+)}{H_{12}} + \delta e(1-\varepsilon)] + me\underset{(+)}{H_{12}}[(1-\theta)\alpha\delta/y + \beta\{\alpha/y + (1-\varepsilon)\}]$$

$$+ \gamma[-\xi\alpha H_{11} + \xi m\underset{(+)}{H_{12}} + e\xi\{\{\alpha/y + (1-\varepsilon)\delta\} - (1-\xi)\beta\underset{(+)}{H_{34}}\}] \equiv b_2(\gamma,\xi) \tag{B2}$$

$b_3 = -($ sum of all principal third-order minors of $J)$

$$= -\gamma me\begin{vmatrix} -H_{12} & \theta\delta - \beta & H_{24} \\ H_{32} & -\{\alpha/y + (1-\varepsilon)\theta\}\delta & H_{34} \\ 0 & (1-\xi)\beta & -\xi \end{vmatrix}$$

$$-\gamma\alpha e\begin{vmatrix} H_{11} & -\delta & H_{14} \\ H_{31} & -\{\alpha/y + (1-\varepsilon)\theta\}\delta & H_{34} \\ 0 & (1-\xi)\beta & -\xi \end{vmatrix} - \gamma\alpha m\begin{vmatrix} H_{11} & H_{12} & H_{14} \\ H_{21} & -H_{12} & H_{24} \\ 0 & 0 & -\xi \end{vmatrix}$$

$$-\alpha me\begin{vmatrix} H_{11} & H_{12} & -\delta \\ H_{21} & -H_{12} & \theta\delta-\beta \\ H_{31} & H_{32} & -\{\alpha/y+(1-\varepsilon)\theta\}\delta \end{vmatrix}$$

$$=\gamma me\{H_{12}\begin{vmatrix} -\{\alpha/y+(1-\varepsilon)\theta\}\delta & H_{34} \\ (1-\xi)\beta & -\xi \end{vmatrix}+H_{32}\begin{vmatrix} \theta\delta-\beta & H_{24} \\ (1-\xi)\beta & -\xi \end{vmatrix}\}$$

$$+\gamma\alpha e\{-H_{11}\begin{vmatrix} -\{\alpha/y+(1-\varepsilon)\theta\}\delta & H_{34} \\ (1-\xi)\beta & -\xi \end{vmatrix}+H_{31}\begin{vmatrix} -\delta & H_{14} \\ (1-\xi)\beta & -\xi \end{vmatrix}\}$$

$$+\gamma\alpha m\xi H_{12}\begin{vmatrix} H_{11} & 1 \\ H_{21} & -1 \end{vmatrix}$$

$$+\alpha me\begin{vmatrix} H_{11} & H_{12} & \delta \\ H_{11}+H_{21} & 0 & \beta+(1-\theta)\delta \\ -(\alpha/y)H_{11}+H_{31} & -(\alpha/y)H_{12}+H_{32} & (1-\varepsilon)\theta\delta \end{vmatrix}$$

$$=\gamma me[\underset{(+)}{H_{12}}\{\{\alpha/y+(1-\varepsilon)\theta\}\delta\xi-(1-\xi)\beta\underset{(+)}{H_{34}}\}+\underset{(+)}{H_{32}}\{(\beta-\theta\delta)\xi-(1-\xi)\beta\underset{(+)}{H_{14}}\}]$$

$$+\gamma\alpha e[-H_{11}\{\{\alpha/y+(1-\varepsilon)\theta\}\delta\varepsilon-(1-\xi)\beta\underset{(+)}{H_{34}}\}+H_{31}\{\delta\xi-(1-\xi)\beta\underset{(+)}{H_{14}}\}]$$

$$+\gamma\alpha m\xi\underset{(+)}{(1-c_y)}\underset{(+)}{H_{12}}+\alpha me(1-\varepsilon)\underset{(+)}{H_{12}}\begin{vmatrix} H_{11} & 1 & \delta \\ -(1-c_y) & 0 & \beta+(1-\theta)\delta \\ -H_{21} & 1 & \theta\delta \end{vmatrix}$$

$$=\gamma me[\{\beta+(1-\theta)\delta\}\{\alpha/y+(1-\varepsilon)\}\xi\underset{(+)}{H_{12}}-\beta(1-\xi)(\underset{(+)}{H_{12}}\underset{(+)}{H_{34}}+\underset{(+)}{H_{14}}\underset{(+)}{H_{32}})]$$

$$+\gamma\alpha e[(1-\varepsilon)\delta\xi(-\theta H_{11}-H_{21})+\beta(1-\xi)(-H_{11}\underset{(+)}{H_{34}}+H_{31}\underset{(+)}{H_{14}})]$$

$$+\gamma\alpha m\xi\underset{(+)}{(1-c_y)}\underset{(+)}{H_{12}}+\alpha me(1-\varepsilon)\underset{(+)}{(1-c_y)}\beta\underset{(+)}{H_{12}}\equiv b_3(\gamma,\xi) \qquad \text{(B3)}$$

$$b_4=\det J$$

$$=\gamma\alpha me\{-(1-\xi)\beta\begin{vmatrix} H_{11} & H_{12} & H_{14} \\ H_{21} & -H_{12} & H_{24} \\ H_{31} & H_{32} & H_{34} \end{vmatrix}-\xi\begin{vmatrix} H_{11} & H_{12} & -\delta \\ H_{21} & -H_{12} & \theta\delta-\beta \\ H_{31} & H_{32} & -\{\alpha/y+(1-\varepsilon)\theta\}\delta \end{vmatrix}\}$$

$$=\gamma\alpha me\{-(1-\xi)\beta\begin{vmatrix} H_{11} & H_{12} & H_{14} \\ H_{11}+H_{21} & 0 & H_{14}+H_{24} \\ -(\alpha/y)H_{11}+H_{31} & -(\alpha/y)H_{12}+H_{32} & -(\alpha/y)H_{14}+H_{34} \end{vmatrix}$$

$$+\xi\begin{vmatrix} H_{11} & H_{12} & \delta \\ H_{11}+H_{21} & 0 & \beta+(1-\theta)\delta \\ -(\alpha/y)H_{11}+H_{31} & -(\alpha/y)H_{12}+H_{32} & (1-\varepsilon)\theta\delta \end{vmatrix}\}$$

$$=\gamma\alpha me(1-\varepsilon)\underset{(+)}{H_{12}}\{(1-\xi)\beta\begin{vmatrix} H_{11} & 1 & H_{14} \\ 1-c_y & 0 & 1-c_\pi \\ -H_{21} & 1 & -H_{24} \end{vmatrix}+\xi\begin{vmatrix} H_{11} & 1 & \delta \\ -(1-c_y) & 0 & \beta+(1-\theta)\delta \\ -H_{21} & 1 & \theta\delta \end{vmatrix}\}$$

$$=\gamma\alpha me(1-\varepsilon)\underset{(+)}{H_{12}}\{(1-\xi)\beta\cdot 0+\xi(1-c_y)\beta\}$$

$$=\gamma\alpha me(1-\varepsilon)\underset{(+)}{(1-c_y)}\beta\xi\underset{(+)}{H_{12}}\equiv b_4(\gamma,\xi)>0 \tag{B4}$$

$$b_1(0,\xi)b_2(0,\xi)-b_3(0,\xi)=[-\alpha H_{11}+m\underset{(+)}{H_{12}}+e\{\alpha/y+(1-\varepsilon)\theta\}\delta][\alpha\underset{(+)}{(1-c_y)}\{m\underset{(+)}{H_{12}}$$
$$+\delta e(1-\varepsilon)\}+me\underset{(+)}{H_{12}}\{(1-\theta)\alpha\delta/y+\beta\{\alpha/y+(1-\varepsilon)\}\}]$$
$$-\alpha me(1-\varepsilon)\underset{(+)}{(1-c_y)}\beta\underset{(+)}{H_{12}} \tag{B5}$$

$$b_1(\gamma,1)b_2(\gamma,1)-b_3(\gamma,1)=[-\alpha H_{11}+m\underset{(+)}{H_{12}}+e\{\alpha/y+(1-\varepsilon)\theta\}\delta+\gamma][\alpha\underset{(+)}{(1-c_y)}\{m\underset{(+)}{H_{12}}$$
$$+\delta e(1-\varepsilon)\}+me\underset{(+)}{H_{12}}\{(1-\theta)\alpha\delta/y+\beta\{\alpha/y+(1-\varepsilon)\}\}+\gamma\{-H_{11}+m\underset{(+)}{H_{12}}$$
$$+\{\alpha/y+(1-\varepsilon)\delta\}\}]-[\gamma me\{\beta+(1-\theta)\delta\}\{\alpha/y+(1-\varepsilon)\}\underset{(+)}{H_{12}}$$
$$+\gamma\alpha e(1-\varepsilon)\delta(-\theta H_{11}-H_{21})+\gamma\alpha m\underset{(+)}{(1-c_y)}\underset{(+)}{H_{12}}+\alpha me(1-\varepsilon)\beta\underset{(+)}{H_{12}}] \tag{B6}$$

Appendix C: Proof of Proposition 3

(1) It follows from **APPENDIX B** that we can write $b_j=b_j(\gamma,\xi)$ $(j=1,2,3,4)$ and it is easy to see that $b_2(0,\xi)>0$, $b_3(0,\xi)>0$. On the other hand, inequalities $b_1>0$ and $b_4>0$ are satisfied for all positive values of γ under **Assumption 1** (cf. equations (B1) and (B4) in **APPENDIX B**). This means that we have

$$b_j>0\quad (j=1,2,3,4) \tag{C1}$$

for all sufficiently small values of $\gamma>0$ by continuity. Next, we can define

$$\Phi(\gamma,\xi) \equiv b_1(\gamma,\xi)b_2(\gamma,\xi)b_3(\gamma,\xi) - b_1(\gamma,\xi)^2 b_4(\gamma,\xi) - b_3(\gamma,\xi)^2 \tag{C2}$$

and we have $b_4(0,\xi) = 0$ (cf. Eq. (B4)). Therefore, we obtain

$$\Phi(0,\xi) = \{b_1(0,\xi)b_2(0,\xi) - b_3(0,\xi)\}b_3(0,\xi),\ \ b_3(0,\xi)>0. \tag{C3}$$

On the other hand, we can see that

$$b_1(0,\xi)b_2(0,\xi) - b_3(0,\xi)>0 \tag{C4}$$

for all sufficiently small values of $\alpha>0$ (cf. Eq. (B5) and inequalities (49)). Inequalities (C3) and (C4) mean that we have

$$\Phi(\gamma,\xi)>0 \tag{C5}$$

for all sufficiently small values of $\alpha>0$ and $\gamma>0$ by continuity. Inequalities (C1) and (C5) are in fact inequalities (48) in the text, which are Routh-Hurwitz conditions for local stability.

(2) It follows from equations (B2), (B3), and inequalities (49) that $b_2(\gamma,1)>0$, $b_3(\gamma,1)>0$. On the other hand, we have $b_1>0$ and $b_4>0$ for all values of $\alpha>0$, $\beta>0$, and $\xi\in[0,1]$. Therefore, we have inequalities (C1) for all values of $\alpha>0$ and $\beta>0$ if ξ is sufficiently close to 1 by continuity. By the way, we have the following relationships (cf. equations (C2), (B3), and (B4)).

$$\lim_{\alpha\to 0}\Phi(\gamma,1) = [\lim_{\alpha\to 0}\{b_1(\gamma,1)b_2(\gamma,1) - b_3(\gamma,1)\}][\lim_{\alpha\to 0} b_3(\gamma,1)] \tag{C6}$$

$$\lim_{\alpha\to 0} b_3(\gamma,1)>0 \ \text{for all}\ \beta>0,\ \gamma>0,\ \theta\in(0,1) \tag{C7}$$

On the other hand, it is easy to see that the following inequality is satisfied for all $\gamma>0$ if $\beta>0$ is sufficiently small and $\theta\in(0,1)$ is sufficiently close to 1 (cf. Eq. (B6)).

$$\lim_{\alpha\to 0}\{b_1(\gamma,1)b_2(\gamma,1) - b_3(\gamma,1)\}>0 \tag{C8}$$

It follows from (C6), (C7) and (C8) that the inequality (C5) is satisfied for all $\gamma>0$ if $\beta>0$ is sufficiently small and $\theta\in(0,1)$ is sufficiently close to 1 even if α is positive and ξ is less than 1, as long as α is sufficiently small and ξ is sufficiently close to 1, by continuity. Inequalities (C1) and (C5) imply Routh-Hurwitz conditions for local stability. □

References

Asada, T. (1991): “On a Mixed Competitive-Monopolistic Macrodynamic Model in a Monetary Economy.” *Journal of Economics* **54**, pp. 33 – 53.

Asada, T. (2006): “Stabilization Policy in a Keynes-Goodwin Model with Debt Accumulation.” *Structural Change and Economic Dynamics* **17**, pp. 466 – 485.

Asada, T. (2008): “Inflation, Deflation and Employment: A Macrodynamic Approach.” in D. T. Bentley and E. P. Nelson (eds.) *Inflation: Roles, Targeting, and Dynamics,* Nova Science Publishers, New York, forthcoming.

Asada, T., C. Chiarella, P. Flaschel and R. Franke (2003): *Open Economy Macrodynamics: An Integrated Disequilibrium Approach.* Springer, Berlin.

Asada, T., W. Semmler and A. J. Novak (1998): “Endogenous Growth and the Balanced Growth Equilibrium.” *Research in Economics* **52**, pp. 189-212.

Barro, R. J. and X. Sala-i-Martin (1995): *Economic Growth*. McGraw Hill, New York.

Chiarella, C. and P. Flaschel (2000): *The Dynamics of Keynesian Monetary Growth.* Cambridge University Press, Cambridge.

Gandolfo, G. (1996): *Economic Dynamics* (Third Edition). Springer, Berlin.

Goodwin, R. M. (1967): “A Growth Cycle.” C. H. Feinstein (ed.) *Socialism, Capitalism and Economic Growth*, Cambridge University Press, Cambridge, pp. 54 – 58.

Hayashi, F. and E. Prescott (2002): “The 1990s in Japan: A Lost Decade.” *Review of Economic Dynamics* **5**, pp. 206-235.

Kaldor, N. (1957): “A Model of Economic Growth.” *Economic Journal* **67,** pp. 591 – 624. (Reprinted in Kaldor 1960, pp. 259 – 300.)

Kaldor, N. (1960): *Essays on Economic Stability and Growth*. Duckworth, London.

Kaldor, N. (1978): *Further Essays on Economic Theory*. Duckworth, London.

Keynes, J. M. (1936): *The General Theory of Employment, Interest and Money.* Macmillan, London.

Leon-Ledesma, M. and A. P. Thirlwall (2007): “Is the Natural Rate of Growth Exogenous?” P. Arestis, M. Badderley and S. L. McCombie (eds.) *Economic Growth: New Directions in Theory and Policy,* Edward-Elger, Cheltenheim UK, pp. 32 – 43.

Lucas, R. E. (1988): “On the Mechanics of Economic Development.” *Journal of Monetary Economics* **22**, pp. 1 – 42.

Ouchi, M. (2008a): *A Theoretical Consideration of the Investment Determination and the Macroeconomic Fluctuations: Firms’ Investment Behaviors under Uncertainty*. Ph. D. Dissertation, Graduate School of Economics, Chuo University, Tokyo. (in Japanese)

Ouchi, M. (2008b): “Investment Behavior of a Monopolistic Firm Under Uncertainty.” *Studies in Regional Science* (The Japan Section of the RSAI), forthcoming.

Palley, T. (1996): “Growth Theory in a Keynesian Mode: Some Keynesian Foundations for New Endogenous Growth Theory.” *Journal of Post Keynesian Economics* **19**, pp. 113 – 135.

Romer, P. (1990): “Endogenous Technological Change.” *Journal of Political Economy* **94**, S71 – S102.

Stein, J. L. (1971): *Money and Capacity Growth*. Columbia University Press, New York.

Tobin, J. (1988): *Money, Credit, and Capital*. McGraw-Hill, New York.

Yoshida, H. and T. Asada (2007): “Dynamic Analysis of Policy Lag in a Keynes-Goodwin Model: Stability, Instability, Cycles and Chaos.” *Journal of Economic Behavior and Organization* **62**, pp. 441 – 469.

Yoshikawa, H. (1980): “On the ‘q’ Theory of Investment.” *American Economic Review* **70**, pp. 739 – 743.

In: Emerging Topics in Macroeconomics
Editor: Richard O. Bailly, pp. 251-274
ISBN: 978-1-60692-093-0

Chapter 9

MACRO PRICE SHOCKS AND NUTRITION IN SUB-SAHARAN AFRICA*

Joseph M. Kargbo**
Ameriprise Financial Services, Inc., RiverSource Life Insurance Company
3883A Steppes Court, Falls Church, Virginia 22041

Abstract

Macro price shocks are a concern to the general public and policy makers because these price movements intensify inflationary pressures, increase the uncertainty confronting farmers and agribusiness firms with significant impacts on food security. The current spike in global food prices exceeds all previous records. Food prices increased 2-38 percent per year in Africa during the past three decades, whilst the price of crude oil has more than quadrupled since 2002. The food price increases are driven by several factors, including drought, low food stocks in global market, speculation, declining trends in agricultural production in low-income countries, increasing biofuel production, food demand and income growth in emerging markets. Food accounts for up to 73 percent of household expenditures in Africa. The combination of climate change, rising food and input prices, poor performance of domestic agriculture sector, and high population growth rates is creating significant food shortages in some African countries. Food is imported to bridge the gap between domestic demand and production, thereby, protecting consumption levels. Sub-Saharan Africa currently imports approximately 71.0 percent of its total food supply. Chronic hunger depletes people's ability for increased productivity and increases their susceptibility to diseases. Indeed, whilst high food prices may be good for farmers and food exporters, they could exercabate food insecurity for the poor who may not have access to food at affordable prices. We use five decades of annual data and a vector error correction model to examine the relationship between macro price shocks and nutrition in Africa. The model identifies and evaluates the strength of the linkages between variables which enables us to trace the dynamic responses in the system at different time horizons in reaction to specific shocks. The empirical results suggest that shocks to food prices, exchange rates and trade policies have significant impacts on nutrition, real incomes and per capita domestic food production. There are strong feedbacks amongst the

* Thanks to Dr. Nomathemba Seme for comments and suggestions on earlier versions of the paper. The views expressed in this paper do not reflect the official position of Ameriprise Financial Services or its affiliates. The author accepts full responsibility for any errors or omissions. May 30, 2008.
** E-mail address: jkargbo@infionline.net. Cell: 757-572-3352. Tel./Fax: 703-820-2995

variables and rapid adjustment towards long-run equilibrium levels after a shock in the system. This information is useful for the formulation of appropriate policies to solve the growing food problems in Africa.

I. Introduction

Macro price (e.g. energy and food prices, exchange rates, wage and interest rate) shocks are a concern to the general public and policy makers because these price movements play a critical role in the welfare of the nation. In particular, the high volatility in oil and food prices tend to intensify inflationary pressures, increases the uncertainty confronting farmers and agribusiness firms with significant impacts on food security. The current spike in global food prices exceeds all previous records. For instance, food prices increased 2-38 percent per year in Africa during the past three decades, whilst the price of crude oil has more than quadrupled since 2002 – exceeding $130.00 per barrel by May 2008. The price increases are driven by drought in major agricultural producing areas, low food stocks in the global market, speculation, stagnant or declining trends in agricultural production in low-income countries, changing diets, increasing food demand and high income growth in emerging markets (e.g. China and India) and other developing countries, and increasing use of food crops, such as corn for biofuel production (World Bank 2008; OECD and FAO 2008).

Food accounted for up to 73 percent of household expenditures in Africa during the past two decades. The combination of climate change, rising food and oil prices, poor performance of domestic agriculture sector, and high population growth rates is creating significant food shortages in some African countries. Food is imported to bridge the gap between domestic demand and production, thereby, protecting consumption levels. Annual food imports in Sub-Saharan Africa increased from $5.43 billion in 1980 to at least $8.35 billion in the early 2000s. Sub-Saharan Africa currently imports approximately 71.0 percent of its total food supply. The cost of food imports increased by 10 to 25 percent over the 2000-2007 period due to the rising food prices. For further details, see for example: Kargbo (2000, 2005, 2007a,b,c; 2008a,b); Faiola (2008a, b); von Braun (2008); and Rosen and Shapouri (2008).

Foreign exchange constraints impede a country’s ability to import food in international markets, and this leads to a deterioration of food insecurity in the country during periods of production shortfalls. Chronic hunger depletes people’s ability for increased productivity and increases their susceptibility to diseases. High food prices may be good for farmers and food exporters, but they could also exercabate food insecurity for the poor who may not have access to food at affordable prices. The behavior of food prices present African policy makers with a dilemma that is not readily solved with traditional monetary and fiscal policies. Despite the implementation of extensive policy reforms in various African countries, the failure of globalization to provide food at affordable prices is worsening the food crisis for the continent’s net food importers. Distortions in the global food market due to agricultural protection policies and subsidies mainly in the United States and the European Union put farmers in developing countries at a disadvantage in the global market. Efforts to settle the thorny issue of agricultural trade liberalization has so far failed in the Doha Round of world trade negotiations (see Kargbo 2006; Faiola 2008a,b).

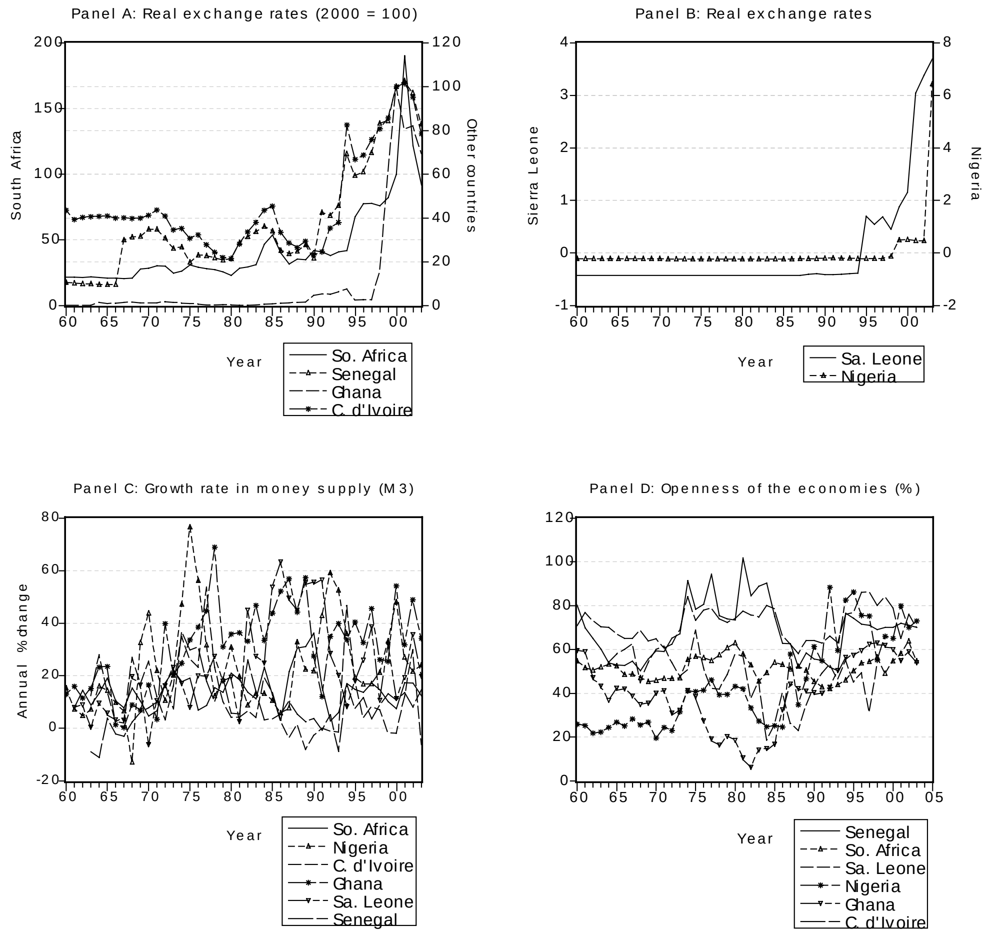
Real shocks in some sectors of the economy, including input costs incurred along the food supply chain raises the prices of food and other commodities paid by consumers in the market. The presence of structural bottlenecks (e.g. foreign exchange constraints, low supply elasticities for agricultural products, and sticky prices and wages in the industrial sector), coupled with differences in prices and output mechanisms across sectors significantly influences the origins and persistence of inflation in developing countries. Recent studies on the impacts of macroeconomic factors on the agricultural sector provide evidence of significant linkages between real supply shocks, food and agricultural prices, money supply, exchange rates and international monetary reserves in various countries (see for example: Kargbo 2000, 2005, 2006, 2007a,b). Exchange rates, interest rates, and the level of money supply are key monetary variables that are determined mainly within domestic or international markets. The manipulation of money supply raises serious concerns about price stability in the agricultural sector, particularly food prices. Macroeconomic variables, including trade policy instruments on imports and exports are determined by domestic policy makers. These variables are viewed as exogenous to the agricultural sector. Moreover, the exchange rate determination process provides a direct link in models that attempt to establish a relationship between the agricultural sector and level of money supply in a particular country.

This study uses data and a vector error correction model (VECM) to examine the relationship between macro price shocks and nutrition in selected African countries during the past five decades. The model identifies and evaluates the strength of the linkages between variables (e.g. daily per capita calorie intake, real food prices, per capita income, domestic food production, real exchange rates and openness of the economy) which enables us to trace the dynamic responses in the system at different time horizons in reaction to specific shocks. This is useful for the formulation of appropriate policies to solve the growing food problems in Africa.

The rest of the analysis in this paper is organized as follows: section two deals with the linkages between food prices and nutrition under policy reforms in Africa. Section 3 presents the model that outlines the relationship between daily per capita calorie intake, food prices and other variables. The empirical results are presented in section 4, and the conclusions are in the final section.

II. Food Prices and Nutrition under Policy Reforms

The implementation of macroeconomic policy reforms has significant impacts on food prices, nutrition, real incomes of farmers, and terms of trade between tradables and non-tradables. The macroeconomic adjustment packages included exchange rates and price reforms that targeted farmers and stabilization of food prices for consumers. In particular, the price reforms included the removal of subsidies on food prices and farm inputs, such as fertilizers and pesticides. Trade liberalization has increased the openness of African economies, and reduction of tariffs and non-tariff barriers to trade. These policies have serious implications for poverty reduction and agricultural growth in Africa (see Kargbo 2005, 2007a,b). In some cases, inconsistent policies and mismanagement have contributed to the poor performance of African agriculture. Figure 1 presents some key macroeconomic variables that have been the focus of policy reforms over the past few decades. The real exchange rate (R_o) is defined herein as: $R_o = eP_f/P_d$; where: e is the bilateral official exchange rate expressed in units of local currency per unit of foreign currency, P_f and P_d are foreign and domestic consumer price indexes, respectively. A rise in the index reflects a depreciation whilst a decline means an appreciation of the domestic currency. The real exchange rates show wide variability during the past five decades.



Source: IMF and World Bank Data.

Notes: The graphs in Panel B are normalized with mean zero and one standard deviation. Openness is calculated as ratio of the sum of imports and exports over GDP in a particular country.

Figure 1. Selected macroeconomic indicators in Sub-Saharan Africa, 1960- 2003.

Food prices and inflation maintain a bilateral relationship as cost-push and demand-pull factors trigger adjustments in various countries. The food and energy prices are increasingly intertwined, and their rising volatility has serious implications for global imbalances, food security and poverty (ECA 2007; Kargbo 2008a,b). In particular, the poor are very much vulnerable to the impacts of rising food and energy prices since these items represent over 70 percent of household budgets in some African countries. Food prices have increased on average 2-38 percent per year in several African countries during the past four decades (see figure 2). Global commodity prices have increased steeply during the past few years, and there seems to be no end in sight. A joint report by the Organization for Economic

Cooperation and Development (OECD) and Food and Agriculture Organization of the United Nations (FAO) published in 2008 predicted that the prices for pork and beef may be 20 percent higher by 2017, compared to 80 percent for vegetable oils, and 60 percent for wheat during this period. So far, wheat prices have increased by 200 percent since 2000, whilst corn, soybean oil, and sugar increased by 54, 71, and 75 percent since 2004. Mauritania, which produces only 30 percent of its domestic food demand has seen the cost of its food staples , such as wheat, rice and cooking oil rose by 67, 117 and 25 percent, respectively during the past year. The FAO estimated that food costs in the most food import-dependent countries will increase by 40 percent this year. The food price increases are driven by drought in key agricultural producing areas of the world, high income growth in emerging markets and other developing countries, and increasing use of food crops, such as corn for biofuel production. The food price increases are putting pressure on governments and have triggered riots and protests in various countries recently.

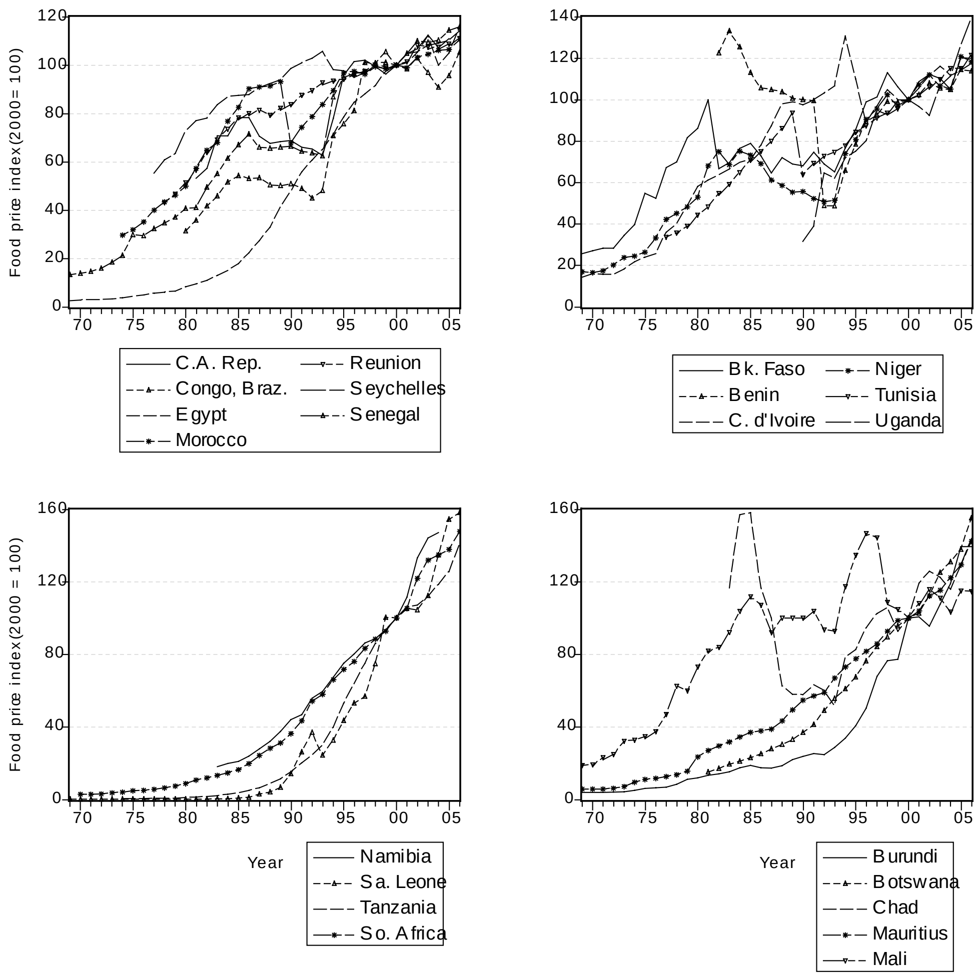


Figure 2. Continued on next page.

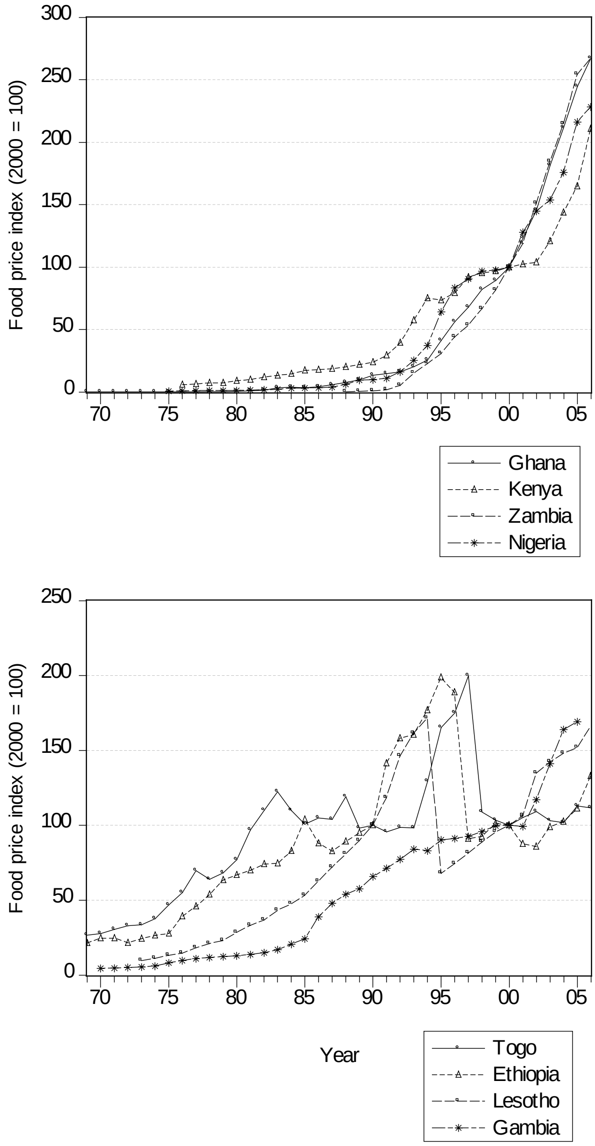
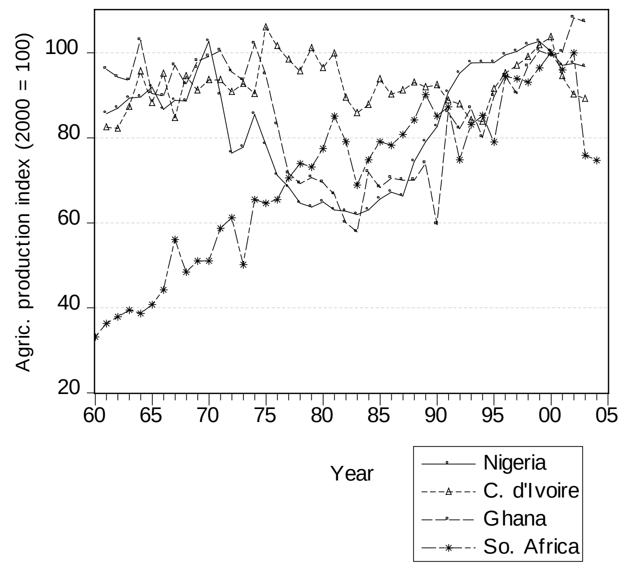


Figure 2. Food prices in Sub-Saharan Africa, 1969 – 2006. (2000 = 100). Source: ILO Data.

Despite the implementation of policy reforms during the past 25 years, per capita domestic agricultural production is either stagnant or declining in a vast majority of African countries during this period (see figure 3). The poor performance of the agricultural sector has serious ramifications for the reduction of rural poverty and welfare of people across Africa. Domestic agricultural production is a major source of food supply, income and livelihood for vast segments of the population in Africa. Food is imported to bridge the gap between domestic demand and production, thereby, protecting consumption levels. The foreign exchange constraints hinder a country's ability to buy food in international markets, which exacerbates food insecurity during periods of domestic production shortfalls. Moreover, food imports compete with domestic development programs, along with external debt and other foreign obligations for foreign exchange. Faiola (2008a) reported that Sub-Saharan Africa imported at least 71 percent of its total food supply in recent years, compared to 68 percent for North Africa, and 37-53 percent for Asia and Pacific region. The rising food crisis has elicited a number of responses from governments around the world, including the implementation of export bans, price controls and subsidies on food items.



Source: FAO Data

Figure 3. Per capita agricultural production in selected African countries, 1960 – 2005.

The rising energy prices combined with the volatility of food and other commodity prices raises concerns for food insecurity, especially for the net food importers in Africa. The price shocks affect people's access to food and consumption patterns. Figure 4 shows the patterns of calorie intake in selected African countries during the past five decades. The average calorie intake ranged from 1,699.52 in Ethiopia to 2,821.01 in South Africa during this periods. The variability in per capita daily calorie intake reveals that consumption in some countries is less than the recommended nutritional requirement of 2,100 calories per person. The share of grains in total calorie intake varies immensely both between and within regions. For example, grains account for 70 percent of the diets in Eritrea and Ethiopia, 54-70 percent in South Africa, compared to 80 percent in Bangladesh (see Missiaen and Shapouri 1994; Rosen and Shapouri 2008). Researchers at the United States Department of Agriculture projected significant food gaps (6-24 percent) that emanate from price shocks in 70 low-income countries during the 2007-2016 period. Rosen and Shapouri defined the food gap as the amount of food needed to raise consumption of all income groups to the recommended nutritional requirement noted above.

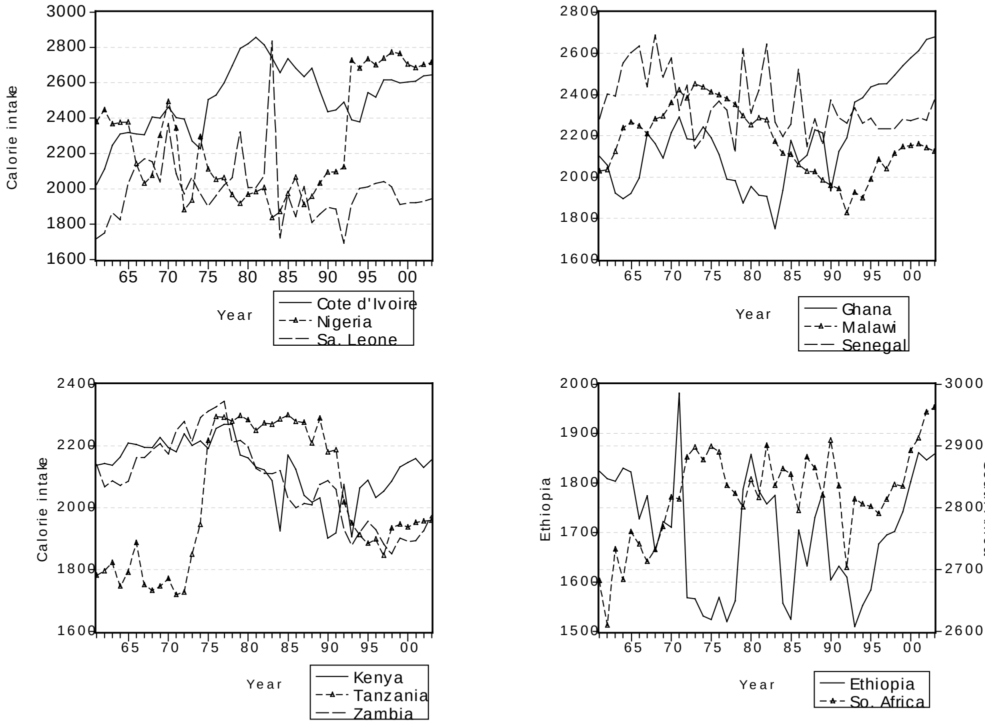


Figure 4. Per capita daily calorie intake in Sub-Saharan Africa, 1961 – 2003. Source: FAO Data.

The World Bank and United Nations have requested $755 million emergency food aid from donor countries to soften the devastating impacts of the food crisis on low-income families around the world. As Robert Zoellick, President of the World Bank noted recently: *"Hunger and malnutrition are the forgotten Millennium Development Goal. It has gotten less attention, but increased food prices and their threat – not only to people but also to political stability – have made it a matter of urgency to draw the attention it needs"*, (World Bank, 2008, p.1)

The World Food Program argues that 22 of the 30 countries experiencing the most severe food insecurity this year are in Africa. Unfortunately, cash-strapped food deficit African countries have difficulties getting the much needed foreign food aid. Food aid has been stagnant or declining at the rate of about 2 percent per year over the 1990-2005 period in 70 low-income countries (Rosen and Shapouri 2008, p.20).

III. Model Specification

Food availability and the ability to acquire it are critical issues in dealing with food security in Africa. Access to food and consumption patterns are affected by both domestic and external factors. Equilibrium food prices are determined by the interaction of producers and consumers in the market under competitive market conditions. We use Johansen's cointegration technique in examining the relationship amongst the various variables (Johansen 1988, 2000). The model is estimated with the method of full information maximum likelihood, and the procedure has the advantage of permitting the joint determination of variables in the system, it takes into account the short-run dynamics of the variables, whilst permitting the system of variables to return to long-run equilibrium. The variables to be tested for any particular country can be written in vector error correction form as:

$$\Delta P_t = \sum_{i=1}^{k-1} \Gamma_i \ \Delta \ P_{t-1} \ + \prod P_{t-k} + D_t + \nu_t \tag{1}$$

where: P_t is an m-dimensional vector of variables in the system, i.e. P_t = ($CALINT_t$, $PCIN_t$, $PFOOD_t$, DOP_t, RER_t, $OPEN_t$)́. D_t refers to the matrix of deterministic variables, such as intercept and time trend, k is the number of lags, and ν_t is the error vector, it is multivariate normal and independent across observations. According to the Granger representation theorem, if the coefficient matrix, Π has reduced rank $r < m$, then there exists m x r matrices α and Φ each with rank such that Π = αΦ′ and Φ′P is stationary (Engle and Granger, 1987). The elements of α are called the adjustment parameters in the vector error correction model, and r is the cointegrating rank. Each column of Φ is the cointegrating vector. We used the *likelihood ratio (trace) test statistic* for testing the hypothesis of at most r cointegrating vectors for each country. An alternative test statistic is the *maximum eigenvalue statistic*. See Johansen (2000) and the *Eviews 3.1 User's Guide* (QMS 1998) for details.

CALINT is the daily per capita calorie intake, $PFOOD_t$ is the real price of food - that is, nominal food price deflated by GDP deflator. $PCIN_t$ (measured as real per capita GDP) is real per capita income. DOP_t is an index of per capita domestic food production, and it is affected by prices, technology, rainfall, political instability and other factors. RER_t is the real bilateral exchange rate. The subscripts: t = 1, 2, ..., n represents annual time periods. $OPEN_t$ is an indicator for trade policy instruments, such as tariffs, quotas and export taxes in African countries. Implementation of import-substitution and other restrictive trade policies contributed to the poor performance of the agriculture sector and high food prices in the region. Consistent data on tariffs and other taxes is lacking. It is difficult to construct a reliable series for trade policy over the estimation period due to pervasiveness of non-tariff barriers and parallel markets for goods and foreign exchange. Thus, OPEN is the ratio of the sum of exports and imports of goods and services over the country's GDP.

An illusion-free individual is assumed to maximize his or her utility subject to a budget constraint, and is said to have demand functions that are homogenous of degree zero in prices and the quantity of financial assets, including money. Money supply and exchange rates enter the model by combining the theory of monetary equilibrium and exchange rate determination. Let MV(i, Y)/P = Y; where: M is the nominal quantity of money demanded, assuming the money market is in equilibrium; V is the velocity of money and it is a function of interest rates (i) and other variables such as real income (Y); and P is the price level. We can rewrite the above equation as: P = MV(i, Y)/Y. Thus, an increase in M is expected to raise the price level in the same direction. Since food accounts for up to 73 percent of household expenditures in Africa, P represents the price of food (Pfood). The strict version of purchasing power parity states that domestic prices are equal to foreign prices (P*) converted at the exchange rate (e) - which is the domestic currency price of foreign exchange. Thus, P = eP*. Combining the above equations, we get: e = [(1/P*)V(i,Y)M]/Y. The equilibrium exchange rate is a function of nominal money, real output and velocity (see Kargbo 2003, 2005, 2007a,b). Since policy reforms have been implemented for nearly two decades in Africa, we argue that domestic prices are becoming fully flexible and are linked to world prices through the exchange rate.

From the above discussion, a six-variable vector error correction model was specified. The annual data for nominal food prices were obtained from various issues of the *Yearbook of Labour Statistics,* published by the International Labour Organization, and also available at http://www.Laborsta.ilo.org. Data for daily calorie intake and indexes of per capita domestic food production were from the FAO database available at http://www.Fao.org. All other data series were from the World Bank Africa data base and various issues of the *International Financial Statistics Yearbook*, published by the International Monetary Fund.

IV. Empirical Results

The unit root tests performed with the Augmented Dickey Fuller (ADF) and Phillips-Perron tests found that all the series are *I(1)* - meaning stationarity was achieved after differencing the series once. All tests were applied to the logarithm of each variable (except M3). Table 1 presents the results of the cointegration tests. The Akaike Information criterion was used in choosing the lag order of the model. The results suggest there is a single cointegrating relationship between the variables in all countries, except Sierra Leone which has 2 cointegrating vectors. The existence of more than a single cointegrating vector implies there are several equilibrium relationships that link the variables in the system, thus, forming an equilibrium sub-space.

Table 2 presents the normalized long-run parameters for variables in each country. There are parameter estimates for each cointegrating vector. Normalizations are on variables with the coefficient 1.000. The implementation of trade policy reforms that open up the countries to international trade would tend to increase daily per capita calorie intake in Ghana, Nigeria and South Africa, but had opposite effects in Cote d'Ivoire. Nutrition (CALINT) increased with upward adjustments in real food prices (PFOOD) in all countries except Sierra Leone. The impacts of domestic production (DOP) and real exchange rates (RER) shocks on CALINT are mixed across the countries studied. This is probably because food production levels were rising from a very low base. An examination of the data shows that these countries have experienced declining trends in per capita agricultural production during the past five decades.

Table 1. Johansen cointegration test and results for nutrition linkages in Sub-Saharan Africa, 1960 – 2003

Country	Number Of lags & DT	Hypothe-sized # of CE(s)	Eigenvalue	Trace (LR) test statistic	5% critical value	1% critical value	Rank(r)
Cote d'Ivoire N = 37	1, CT	r = 0	0.721	119.796*	104.94	114.36	1
		r ≤ 1	0.485	72.523	77.74	85.78	
		r ≤ 2	0.428	47.962	54.64	61.24	
		r ≤ 3	0.396	27.242	34.55	40.29	
		r ≤ 4	0.159	8.617	18.17	23.46	
		r ≤ 5	0.057	2.195	3.74	6.40	
Ghana N = 37	1, C	r = 0	0.674	108.534*	94.15	103.18	1
		r ≤ 1	0.491	67.032	68.21	76.07	
		r ≤ 2	0.444	42.042	47.21	54.46	
		r ≤ 3	0.278	20.291	29.68	35.65	
		r ≤ 4	0.130	8.201	15.41	20.04	
		r ≤ 5	0.079	3.063	3.76	6.65	
Nigeria N = 35	1, C	r = 0	0.761	116.626*	94.15	103.18	1
		r ≤ 1	0.599	63.625	68.21	76.07	
		r ≤ 2	0.362	29.850	47.21	54.46	
		r ≤ 3	0.229	13.194	29.68	35.65	
		r ≤ 4	0.091	3.552	15.41	20.04	
		r ≤ 5	0.0001	0.005	3.76	6.65	
Senegal N= 34	1, CT	r = 0	0.737	109.343**	104.94	114.36	1
		r ≤ 1	0.525	63.864	77.74	85.78	
		r ≤ 2	0.410	38.526	54.64	61.24	
		r ≤ 3	0.245	20.606	34.55	40.29	
		r ≤ 4	0.234	11.057	18.17	23.46	
		r ≤ 5	0.058	2.019	3.74	6.40	
Sierra Leone N = 36	3, CT	r = 0	0.867	193.635*	114.90	124.75	2
		r ≤ 1	0.814	120.890*	87.31	96.58	
		r ≤ 2	0.582	60.344	62.99	70.05	
		r ≤ 3	0.427	28.932	42.44	48.45	
		r ≤ 4	0.183	8.840	25.32	30.45	
		r ≤ 5	0.041	1.541	12.25	16.26	
South Africa N= 40	2, CT	r = 0	0.727	114.022**	104.94	114.36	1
		r ≤ 1	0.514	62.057	77.74	85.78	
		r ≤ 2	0.322	33.170	54.64	61.24	
		r ≤ 3	0.251	17.569	34.55	40.29	
		r ≤ 4	0.099	5.993	18.17	23.46	
		r ≤ 5	0.044	1.804	3.74	6.40	

Notes: DT refers to the type of deterministic trends that are present in the data. For example, a constant (C), or constant and trend (CT) were included in each cointegrating equation (CE).

* Denotes rejection of the hypothesis at 1.0% significance level.

** Denotes rejection of the hypothesis at 5.0% significance level.

The number of lags was determined by minimizing the Akaike information criterion and Schwarz criterion. N = number of observations included in the cointegration test.

Table 2. Vector error correction estimates for nutrition linkages in Sub-Saharan Africa, 1960 – 2003

Country	Variables							
	CALINT	OPEN	PFOOD	RER	PCIN	DOP	TREND	CONSTANT
Cote d'Ivoire	1.000	-0.657	0.658	0.190	-1.267	2.329	0.019	-15.301
(1967 - 2003)a		(2.452)b	(2.488)	(2.195)	(2.890)	(2.490)		
Ghana	1.000	0.024	0.017	-0.019	0.139	-0.391	--	-6.936
(1967 – 2003)		(1.910)	(8.542)	(3.745)	(2.779)	(7.997)		
Nigeria	1.000	0.858	0.266	0.008	-1.553	-1.135	--	4.214
(1967 – 2003)		(1.965)	(2.273)	(0.384)	(2.201)	(3.656)		
Senegal	1.000	-0.181	0.149	-0.009	0.490	0.546	0.020	-14.829
(1970 – 2003)		(1.645)	(1.888)	(0.226)	(1.588)	(2.986)		
Sierra Leone	1.000	0.000	0.015	0.092	-0.186	0.084	0.025	-7.504
(1968- 2003)			(1.091)	(3.555)	(3.204)	(0.313)	(2.695)	
	0.000	1.000	-0.169	-0.112	0.496	-1.365	-0.087	2.824
			(4.842)	(1.702)	(3.376)	(2.022)	(3.776)	
South Africa	1.000	0.064	-0.007	-0.077	-0.290	0.138	0.0004	-5.630
(1964 – 2003)		(2.376)	(0.923)	(7.583)	(8.720)	(4.526)		

Table 2. Continued

Country	Variables							
	CALINT	**OPEN**	**PFOOD**	**RER**	**PCIN**	**DOP**	**TREND**	**CONSTANT**
Adjustment coefficients								
Cote d'Ivoire	-0.122	0.202	-0.785	0.442	-0.037	-0.243	--	--
	(2.666)	(1.542)	(2.059)	(1.600)	(0.575)	(3.456)		
Ghana	-1.232	-0.438	-22.039	2.404	0.251	-1.194	--	--
	(5.024)	(0.244)	(2.410)	(0.683)	(0.465)	(2.204)		
Nigeria	0.044	-0.463	-1.625	-0.495	0.024	0.131	--	--
	(0.713)	(2.203)	(3.352)	(0.815)	(0.287)	(3.033)		
Senegal	-0.140	0.315	0.037	0.721	-0.036	-2.116	--	--
	(1.251)	(1.005)	(0.076)	(1.712)	(0.348)	(5.816)		
Sierra Leone	-1.187	0.851	1.074	-0.466	-0.305	0.204	--	--
	(6.175)	(2.480)	(0.420)	(0.486)	(1.453)	(1.538)		
	-0.289	-0.115	2.746	1.788	-0.152	0.113	--	--
	(4.117)	(0.922)	(2.935)	(5.098)	(1.971)	(2.324)		
South Africa	-0.573	-1.963	-0.250	0.170	-0.456	-3.091	--	--
	(3.883)	(3.044)	(0.160)	(0.094)	(1.332)	(3.566)		

Notes: a.Refers to estimation period.

b. Refers to the absolute t-statistic.

--.Not available.

The coefficients normalized to 1 and others that are not identified do not have standard errors, see QMS (1998).

Table 3. Short-run estimates of nutrition linkages in Sub-Saharan Africa, 1965 - 2003

Independent variables	Country					
	Cote d'Ivoire (1968-2003)a	**Ghana (1967-2003)**	**Nigeria (1967-2003)**	**Senegal (1971-2003)**	**Sierra Leone (1967-2003)**	**South Africa (1965-2003)**
Constant	-0.019 (2.315)	0.045 2.727)	-0.001 (0.070)	0.004 (0.146)	-0.031 (0.779)	0.012 (2.025)
ΔCALINT	0.031(-1)b (0.228)c	0.167(-1) (0.897)	-0.277(-2) (1.876)**	0.041(-1) (0.307)	0.220(-1) (0.846)	-0.072(-2) (0.363)
ΔOPEN	0.009(-1) (0.176)	0.066(-1) (2.258)**	0.051(-1) (1.352)	0.176(-1) (3.186)*	-0.106(-1) (1.111)	-0.027(-1) (0.818)
ΔPFOOD	0.003(-2) (0.292)	0.004(-1) (0.566)	-0.001(-1) (0.072)	-0.061(-2) (2.221)**	0.011(-1) (1.244)	-0.015(-2) (2.670)*
ΔRER	0.014(-1) (0.872)	0.006(-1) (0.464)	-0.023(-2) (2.710)*	-0.040(-1) (1.019)	-0.020(-1) (1.265)	-0.008(-2) (0.502)
ΔPCIN	-0.230(-1) (2.465)**	0.092(-1) (0.781)	-0.122(-1) (0.895)	0.179(-1) (1.873)***	-0.046(-1) (0.712)	0.091(-1) (1.831)***
ΔDOP	0.166(-2) (1.734)***	-0.209(-1) (1.852)***	0.507(-1) (2.112)**	0.185(-1) (8.555)*	0.354(-1) (1.784)***	0.012(-2) (0.638)
MSUP	0.001(-1) (5.086)*	-0.001(-1) (2.515)**	0.0003(-2) (0.378)	-0.001(-1) (0.622)	0.001(-1) (1.003)	-0.001(-1) (1.862)***
POLINST	0.010 (1.393)	--	--	--	0.039 (1.833)***	--
ECM	-0.285(-1) (2.601)**	-1.151(-1) (3.929)*	-0.536(-1) (2.439)**	-1.250(-1) (3.829)*	-1.607(-1) (3.074)*	-0.353(-1) (1.793)***

Table 3. Continued

Independent variables	Country					
	Cote d'Ivoire (1968-2003)a	**Ghana (1967-2003)**	**Nigeria (1967-2003)**	**Senegal (1971-2003)**	**Sierra Leone (1967-2003)**	**South Africa (1965-2003)**
R^2	0.537	0.561	0.450	0.793	0.665	0.453
Adj. R^2	0.377	0.436	0.294	0.712	0.537	0.254
SER	0.026	0.039	0.062	0.042	0.080	0.015
D-W stat	2.024	2.146	2.109	1.392	1.938	2.053
F-statistic	3.355*	4.471*	2.874**	9.802*	5.167*	2.438**
LM: χ^2 (1)	0.267	0.823	0.712	6.282**	0.912	0.536
LB Q-statistic	(12)=17.916	(10)=13.075	(16)=8.081	(16)=11.252	(16)=10.054	(16)=10.854
ARCH: χ^2 (p)	(2)=2.946	(7)=10.482	(1)=0.040	(1)=0.319	(1)=0.452	(1)=0.417
WHET: χ^2 (p)	(17)=17.845	(17)=12.223	(17)=18.234	(17)=21.652	(18)=20.741	(17)=12.776
JB statistic	0.369	1.009	0.802	1.485	27.147*	3.852
RESET	F(2,24)=1.735	F(1,27)=1.756	F(1,27)=0.224	F(1,23)=0.210	F(4,23)=1.896	F(1,29)=0.054
Chow test	--	(1983):1.001	(1985):1.175	(1985):1.260	(1991):0.385	(1994):0.462
No. of observations	36	37	37	33	37	39

Notes: a. Refers to estimation period.
b. The negative number in parentheses refers to number of lags which were determined by minimizing the Akaike information criterion.
c. The number in parentheses refers to the absolute t-value.
* Significant at the 1.0% level;
** Significant at the 5.0% level.
*** Significant at the 10% level.

The regressions for Senegal, Sierra Leone and South Africa were corrected for autocorrelation. LM is the Lagrange multiplier test of residual serial correlation. ARCH and WHET are tests for heteroscedasticity based on Engle (1982) and White (1980), respectively. L-B Q-statistic is used to test whether the residuals are white noise. The number of lags is in parentheses. Jarque-Bera (J-B) statistic tests for normality in the residuals, and it is distributed χ^2 with 2 degrees of freedom. RESET is a test for specification error. The breakpoint Chow test is used to verify whether the coefficient vector is constant over the sub-samples, with the year used as a breakpoint being in parentheses. The degrees of freedom for the other tests are in parentheses adjacent to the distributions. ECM (the error-correction term) is the lagged residual from the cointegration regression, and Δ is the first difference operator.

The adjustment coefficients capture the changes in nutrition, real food prices and other variables required to remove past departures from the equilibrium levels [1]. Table 3 presents the short-run estimates and adjustment coefficients for all the countries covered in this study. Shocks to macro prices trigger various adjustments towards equilibrium levels as the deviation from equilibrium is corrected within a year in Ghana, Sierra Leone and Senegal; compared to 2-4 years for the rest of the countries. No single variable seem to bear the brunt of the adjustment towards long-run equilibrium in all the countries covered in Table 2.

Impulse Response Analysis

The impulse response functions trace the effects of a single standard deviation (S.D.) shock to each of the variables in the system over specific time horizons. To facilitate tracing the patterns of movement, we othogonalized the innovations by a Cholesky decomposition with the variables put in different orders. Sims (1980) argued that there is no unique way of ordering the variables. Thus, the ordering used here is based on the policy environment in each country and expected transmission mechanism. Figure 5 presents the graphs containing the responses of variables to shocks over a 10-year horizon. The own and feedback effects of macro price shocks show very significant and persistent impacts on variables in all six countries. Even though, the empirical evidence is mixed, the graphs show that trade policy and real exchange rate shocks have some impacts on nutrition, food prices and domestic food production (DOP) in these African countries. CALINT responds immediately to international

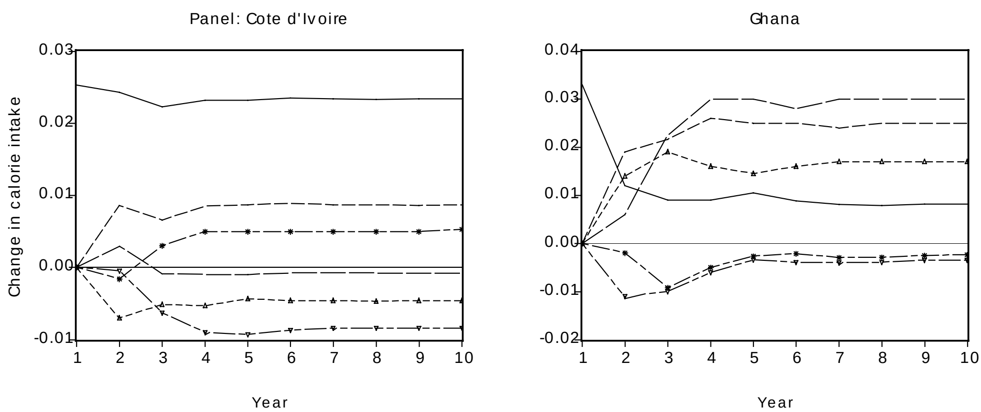


Figure 5. Continued on next page.

[1] 1.The error-correction model is specified as:

$$\Delta \log CALINT_t = \alpha_0 + \beta_1 \Delta \log CALINT_{t-k} + \beta_2 \Delta \log PCIN_{t-k} + \beta_3 \Delta \log RER_{t-k} + \beta_4 \Delta \log OPEN_{t-k} + \beta_5 MSUP_{t-k} + \beta_6 POLINST + \beta_7 \Delta \log DOP_{t-k} + \gamma Ect_{t-1} + \varepsilon_t$$

where: Ect_{t-1} is the lagged residual (error-correction term) from the cointegration regression, ε_t is the stochastic disturbance term, α_0 is a constant, whilst $\beta_1, \beta_2, \ldots, \beta_7$ are parameters representing elasticities, k is the length of lag, and γ is the speed of adjustment. Δ is the first difference operator. POLINST is a binary variable that represents political instability in Africa, with 1 = war time period and 0 otherwise (see Kargbo 1994 for details). MSUP refers to the annual growth rate (percentage change) of broad money supply (M3). The growth rate of M3 is used, instead of levels of the stock of money to reflect its importance from the perspective of macroeconomic stability (see Kargbo 2005).

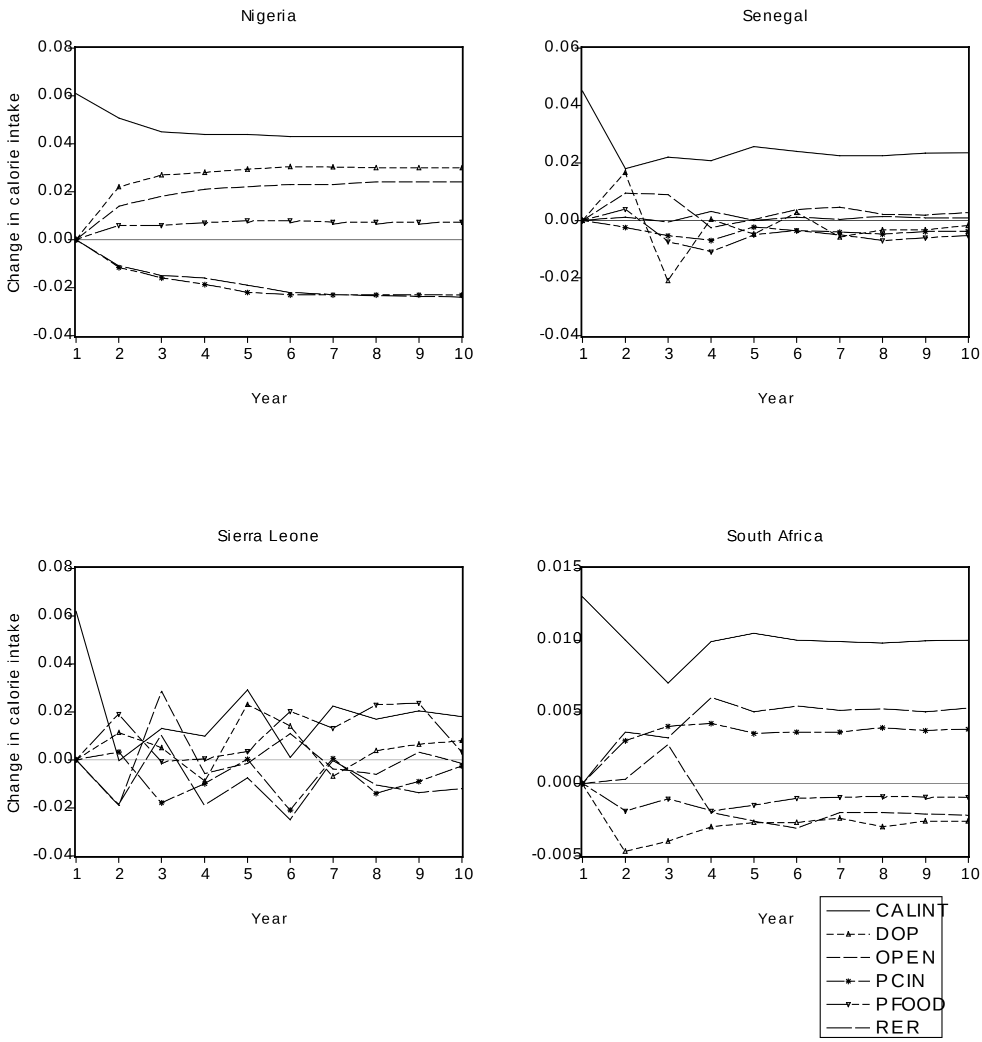


Figure 5. Responses of daily per capita calorie intake to macroeconomic shocks in Africa.

trade shocks (OPEN) in Ghana, Cote d'Ivoire and Nigeria. There is a delayed response (up to 2 years) of CALINT to trade shocks in South Africa, thereafter, nutrition increases to the third year and then becomes negative. Shocks to PFOOD reduces calorie intake in Cote d'Ivoire, Ghana, Senegal and South Africa throughout the forecasting horizon.

Variance Decompositions

The variance decompositions presented in tables 4 through 9 provide us with information about the relative importance of each random innovation to variables in the system. The higher the variance accounted for by cross-variable innovations, the stronger the interactions amongst variables in the system. Calorie intake (CALINT) has more than 50 percent of its variance accounted for by own-innovations for the entire forecasting horizon in all countries, except Ghana and Sierra Leone. The feedback between food prices and the other variables suggest that shocks to real food prices and domestic food production are a major source of macroeconomic instability in Africa. For most countries, shocks to food prices have the most explanatory power for macro variables, such as per capita income, real exchange rates, and trade policy reforms (OPEN); and domestic food production (DOP). The fairly strong feedbacks between real exchange rates and other variables in the system suggest that exchange rate shocks are effective in changing the structure of relative prices in some of these African countries.

Table 4. Decomposition of forecast error variance k-years ahead for nutrition linkages in Cote d'Ivoire

Forecast error in	k-years	S.E.	Shocks to					
			CALINT	OPEN	PFOOD	RER	PCIN	DOP
CALINT	1	0.025	100.000	0.000	0.000	0.000	0.000	0.000
	4	0.052	82.842	6.919	4.491	0.399	1.551	3.797
	7	0.070	78.256	8.389	7.097	0.263	2.704	3.290
	10	0.085	76.841	8.909	7.859	0.208	3.015	3.166
OPEN	1	0.072	6.049	93.950	0.000	0.000	0.000	0.000
	4	0.127	9.441	79.616	7.940	0.993	1.182	0.826
	7	0.165	9.842	75.990	10.741	0.979	1.796	0.651
	10	0.195	10.012	74.633	11.789	0.979	1.993	0.593
PFOOD	1	0.210	8.107	1.596	90.297	0.000	0.000	0.000
	4	0.372	5.506	3.997	76.436	3.580	3.664	6.816
	7	0.450	5.326	3.441	70.068	5.516	7.947	7.701
	10	0.517	5.205	3.179	67.170	6.435	9.823	8.216
RER	1	0.152	0.382	7.414	0.949	91.253	0.000	0.000
	4	0.340	0.102	1.882	5.592	85.063	5.961	1.396
	7	0.458	0.072	1.081	6.545	83.665	7.589	1.041
	10	0.551	0.067	0.779	6.890	83.179	8.159	0.925
PCIN	1	0.035	2.432	4.984	0.002	0.054	92.527	0.000
	4	0.099	7.885	7.602	4.781	0.526	79.106	0.097
	7	0.141	8.001	8.384	5.994	0.588	76.963	0.068
	10	0.173	8.061	8.643	6.339	0.605	76.292	0.059
DOP	1	0.038	8.667	16.723	28.959	0.143	5.957	39.550
	4	0.087	2.170	30.977	47.668	0.889	4.713	13.580
	7	0.119	1.351	31.039	50.079	0.864	4.892	11.774
	10	0.144	1.055	31.136	51.040	0.859	4.874	11.036

Table 5. Decomposition of forecast error variance k-years ahead for nutrition linkages in Ghana

Forecast error in	k-years	S.E.	Shocks to					
			CALINT	OPEN	RER	PFOOD	DOP	PCIN
CALINT	1	0.032	100.000	0.000	0.000	0.000	0.000	0.000
	4	0.074	25.486	27.038	25.500	4.754	15.084	2.137
	7	0.104	15.146	31.018	35.219	2.802	14.553	1.260
	10	0.129	11.197	31.651	39.413	2.094	14.493	0.951
OPEN	1	0.241	3.789	96.210	0.000	0.000	0.000	0.000
	4	0.579	3.286	94.024	2.484	0.019	0.160	0.026
	7	0.816	3.613	91.675	4.571	0.021	0.11	0.017
	10	0.999	3.837	90.618	5.412	0.028	0.093	0.011
RER	1	0.472	0.232	5.567	94.200	0.000	0.000	0.000
	4	1.341	0.978	10.173	83.918	3.103	0.083	1.742
	7	1.930	1.705	9.272	82.822	3.768	0.250	2.182
	10	2.380	1.952	8.774	82.436	4.113	0.303	2.419
PFOOD	1	1.229	18.283	3.413	7.571	70.732	0.000	0.000
	4	2.730	25.001	21.735	6.044	41.998	4.128	1.093
	7	3.762	23.845	26.140	5.390	39.583	3.826	1.214
	10	4.569	23.598	27.940	5.243	38.225	3.802	1.191
DOP	1	0.072	30.413	11.394	0.452	0.231	57.508	0.000
	4	0.150	7.668	32.411	9.201	0.881	49.039	0.799
	7	0.205	4.142	32.025	13.152	1.716	47.763	1.199
	10	0.247	2.849	31.550	14.799	2.064	47.333	1.402
PCIN	1	0.073	13.748	7.806	0.986	24.410	0.242	52.807
	4	0.124	9.807	14.181	0.658	14.722	0.447	60.183
	7	0.160	8.445	14.645	0.878	13.030	0.548	62.453
	10	0.189	7.915	14.707	0.928	12.243	0.6000	63.603

Table 6. Decomposition of forecast error variance k-years ahead for nutrition linkages in Nigeria

Forecast Error in	k-years	S.E.	Shocks to					
			CALINT	RER	OPEN	DOP	PFOOD	PCIN
CALINT	1	0.061	100.000	0.000	0.000	0.000	0.000	0.000
	4	0.122	69.201	4.276	6.923	13.796	0.878	4.925
	7	0.168	57.142	7.255	9.470	16.940	1.062	8.129
	10	0.204	52.123	8.927	10.563	17.913	1.115	9.357
RER	1	0.591	0.051	99.948	0.000	0.000	0.000	0.000
	4	1.890	0.187	92.364	1.301	4.621	0.182	1.342
	7	2.864	0.157	91.679	1.439	5.495	0.217	1.011
	10	3.627	0.147	91.333	1.493	5.878	0.230	0.917
OPEN	1	0.204	0.496	1.743	97.761	0.000	0.000	0.000
	4	0.296	0.624	6.357	87.304	1.239	1.217	3.258
	7	0.376	0.722	13.707	78.538	1.701	1.081	4.249
	10	0.445	0.771	18.142	73.318	2.010	1.005	4.752

Table 6. Continued

Forecast Error in	k-years	S.E.	Shocks to					
			CALINT	RER	OPEN	DOP	PFOOD	PCIN
DOP	1	0.042	40.542	0.008	6.624	52.825	0.000	0.000
	4	0.125	8.397	0.684	20.894	59.238	2.504	8.280
	7	0.187	4.352	0.396	21.641	59.203	2.460	11.946
	10	0.234	3.144	0.379	22.055	58.796	2.447	13.177
PFOOD	1	0.472	0.0001	4.345	9.494	17.987	68.171	0.000
	4	1.047	0.481	39.380	7.804	10.617	27.091	14.625
	7	1.371	0.285	41.495	9.149	6.393	23.742	18.935
	10	1.637	0.201	42.194	9.646	4.582	22.130	21.246
PCIN	1	0.079	15.530	0.651	0.430	1.602	16.122	65.662
	4	0.173	12.609	1.743	1.806	0.713	10.438	72.689
	7	0.228	12.436	3.328	2.356	0.540	10.231	71.106
	10	0.273	12.305	4.135	2.614	0.509	10.093	70.342

Table 7. Decomposition of forecast error variance k-years ahead for nutrition linkages in Senegal

Forecast error in	k-years	S.E.	Shocks to					
			CALINT	PFOOD	OPEN	DOP	RER	PCIN
CALINT	1	0.445	100.000	0.000	0.000	0.000	0.000	0.000
	4	0.066	72.557	4.575	4.127	16.617	0.251	1.871
	7	0.078	77.732	4.202	3.436	12.588	0.197	1.843
	10	0.090	80.105	4.680	2.851	10.083	0.192	2.086
PFOOD	1	0.195	3.653	96.346	0.000	0.000	0.000	0.000
	4	0.373	4.786	88.563	4.921	1.578	0.008	0.142
	7	0.478	4.697	89.392	4.645	1.151	0.017	0.096
	10	0.564	4.564	89.776	4.706	0.864	0.017	0.070
OPEN	1	0.124	4.900	0.713	94.386	0.000	0.000	0.000
	4	0.219	2.708	3.133	91.060	2.797	0.030	0.269
	7	0.287	2.287	3.155	91.957	2.361	0.029	0.209
	10	0.343	2.059	3.308	91.971	2.462	0.026	0.172
DOP	1	0.144	2.771	7.026	3.618	86.583	0.000	0.000
	4	0.206	25.544	10.828	14.988	44.005	0.367	4.267
	7	0.234	32.804	9.554	15.091	37.373	0.304	4.870
	10	0.256	38.265	9.602	14.238	32.110	0.257	5.525
RER	1	0.167	0.739	6.885	14.468	12.211	65.697	0.000
	4	0.299	0.921	4.397	8.752	25.994	59.631	0.304
	7	0.387	0.707	3.883	8.495	26.930	59.753	0.230
	10	0.459	0.643	3.508	8.466	27.728	59.487	0.186
PCIN	1	0.041	0.686	7.916	6.620	0.880	0.391	83.505
	4	0.070	4.868	9.828	3.499	8.187	0.221	73.396
	7	0.088	5.003	10.390	2.499	5.442	0.170	76.494
	10	0.103	5.324	11.283	1.888	4.402	0.155	76.947

Table 8. Decomposition of forecast error variance k-years ahead for nutrition linkages in Sierra Leone

Forecast error in	k-years	S.E.	Shocks to					
			CALINT	OPEN	RER	PFOOD	PCIN	DOP
CALINT	1	0.062	100.000	0.000	0.000	0.000	0.000	0.000
	4	0.085	56.926	16.752	11.466	5.074	6.484	3.296
	7	0.107	48.369	12.025	13.318	8.627	8.454	9.206
	10	0.121	45.394	9.833	13.597	14.338	8.726	8.111
OPEN	1	0.111	13.741	86.258	0.000	0.000	0.000	0.000
	4	0.257	39.418	53.767	1.337	0.988	0.465	4.025
	7	0.297	42.671	44.388	2.467	3.342	3.450	3.680
	10	0.326	44.696	40.857	2.110	2.824	5.467	4.042
RER	1	0.312	0.018	3.508	96.472	0.000	0.000	0.000
	4	0.821	21.529	2.361	22.115	47.389	4.209	2.395
	7	1.265	24.423	20.049	13.858	29.096	10.421	2.151
	10	1.549	19.056	20.996	19.774	19.974	13.040	7.157
PFOOD	1	0.831	1.035	4.540	7.399	87.025	0.000	0.000
	4	1.469	5.891	25.586	4.184	60.065	1.037	3.236
	7	1.822	18.516	24.810	2.938	46.731	4.898	2.107
	10	2.118	19.149	4.116	4.116	42.290	9.875	4.746
PCIN	1	0.068	18.097	3.476	13.343	1.078	64.005	0.000
	4	0.211	12.692	35.101	11.181	1.429	38.431	1.164
	7	0.371	4.162	40.485	13.876	4.927	33.607	2.940
	10	0.468	2.729	38.630	17.735	5.373	32.511	3.020
DOP	1	0.043	0.987	1.416	8.846	11.441	0.573	76.736
	4	0.074	2.900	3.957	5.791	32.305	3.428	51.617
	7	0.100	2.504	3.120	7.278	42.682	2.290	42.123
	10	0.111	2.291	2.608	6.785	45.076	2.435	40.803

Table 9. Decomposition of forecast error variance k-years ahead for nutrition linkages in South Africa

Forecast error in	k-years	S.E.	Shocks to					
			CALINT	PFOOD	RER	OPEN	PCIN	DOP
CALINT	1	0.013	100.000	0.000	0.000	0.000	0.000	0.000
	4	0.024	70.541	1.465	10.334	2.308	7.255	8.095
	7	0.039	68.679	1.210	13.233	3.145	7.545	6.190
	10	0.039	67.473	1.014	14.561	3.091	8.167	5.692
PFOOD	1	0.137	3.829	96.171	0.000	0.000	0.000	0.000
	4	0.295	3.515	89.819	1.233	0.069	5.068	0.292
	7	0.417	2.393	88.452	2.248	0.771	5.533	0.601
	10	0.513	2.186	88.483	2.207	0.619	5.867	0.636
RER	1	0.161	6.054	6.322	87.622	0.000	0.000	0.000
	4	0.292	11.514	4.704	69.991	12.553	0.029	1.206
	7	0.368	11.042	3.916	73.444	10.258	0.041	1.299
	10	0.429	11.616	3.546	73.147	10.189	0.034	1.468

Table 9. Continued

Forecast error in	k-years	S.E.	Shocks to					
			CALINT	PFOOD	RER	OPEN	PCIN	DOP
OPEN	1	0.056	5.583	0.835	2.314	91.266	0.000	0.000
	4	0.109	10.436	3.862	4.943	73.758	1.965	5.034
	7	0.138	9.740	5.562	6.147	71.318	1.786	5.444
	10	0.159	9.558	6.323	6.340	70.255	1.790	5.733
PCIN	1	0.030	6.52E-05	1.860	0.451	38.088	59.601	0.000
	4	0.062	0.363	1.546	4.463	31.208	62.010	0.408
	7	0.081	0.237	2.452	5.077	25.111	66.865	0.255
	10	0.096	0.210	2.954	5.471	22.877	68.306	0.179
DOP	1	0.076	15.642	7.0332	15.034	1.592	1.389	59.311
	4	0.152	34.842	7.782	18.113	3.053	7.901	28.308
	7	0.200	34.637	7.984	20.927	2.099	9.455	24.896
	10	0.238	34.588	8.618	21.148	1.559	10.078	24.009

An examination of the forecast error (S.E) of each variable shows it increases steadily with the length of forecasting horizon, and reaches some upper bound for all of the variables in each country. It confirms the existence of stationary processes in the variables. The source of the forecasting error is the variation in current and future values of the various innovations.

V. Conclusion

Macro price shocks are a concern to the general public and policy makers because these price movements play a critical role in the welfare of the nation. The high volatility in energy and food prices tend to intensify inflationary pressures, increases the uncertainty confronting farmers and agribusiness firms with significant impacts on food security. This has important implications for the farm debt, farm incomes and agricultural productivity. Thus, the stabilization of macro prices is fundamental to the long-term restructuring of African economies. The current spike in global food prices exceeds all previous records. African countries have been experiencing significant increases in food prices over the past few decades. Since food accounts for up to 73 percent of household budgets in some African countries, food price increases have severe political implications because people buy food more frequently than any other item in their household budgets.

This study used data and a vector error correction model to examine the relationship between macro price shocks and nutrition in selected African countries during the past five decades. The model identified and evaluated the strength of the linkages between variables which enabled us to trace the dynamic responses in the system at different time horizons in reaction to specific shocks. The empirical results suggest that shocks to food prices, exchange rates and trade policies have significant impacts on nutrition, real incomes and per capita domestic food production. Moreover, our results show strong feedbacks amongst the variables, and rapid adjustment towards long-run equilibrium levels after a shock in the system. The implementation of monetary and macroeconomic policies has significant impacts on poverty reduction, food security issues and food consumption patterns in Africa. Since declining food production, high food prices, political instability and foreign exchange constraints lower food consumption at the household level, we argue that food security at both the household and national levels is linked directly to improvements in agricultural performance, food policy management, increasing efficiency in the marketing system, and making food available at affordable prices.

Research has shown that inadequate diets inflict tremendous costs to families and the whole country because persistent hunger hinders a nation's productive capacity, thereby, making the people more vulnerable to diseases. Due to the strong linkages between the variables, we suggest that agricultural and macroeconomic policy makers should cooperate more closely in designing economic reform policies in Africa. Neither group of policy makers can ignore the impacts of their policies on agriculture and the macroeconomic environment within which the sector operates. This paper contributes to the policy dialogue going on in Africa, and the empirical results reveal a wide variability amongst the countries. Thus, the policy formulation process in each African country must reflect the economic, social and political environment prevailing in that particular country.

References

Engle, R.F. and Granger, C.W.J. (1987). "Co-integration and error correction: representation, estimation and testing". *Econometrica*, **55** (2), 251-276.

Faiola, A. (2008a). "The new economics of hunger", *The Washington Post*, Sunday April 27, 2008, p. A01.

Faiola, A. (2008b). "Where every meal is a sacrifice", *The Washington Post*, Monday April 28, 2008, pp. A1 & A6.

International Monetary Fund, *International Financial Statistics Yearbook*, IMF, Washington, D.C., various issues.

Johansen, S. (1988). "Statistical analysis of cointegration vectors", *Journal of Economic dynamics and Control*, **12**, 231 -254.

Johansen, S. (2000). "Modelling of cointegration in the vector autoregressive model", *Economic Modelling,* **17**, 359-373.

Kargbo, J.M. (1994). "Analysis of food availability and food policies in the Mano River Union: Implications for political instability and economic growth", Paper presented at the First Joint Annual Meeting of the Sierra Leone Studies and Liberian Studies Associations, held at the College of Charleston, Charleston , South Carolina, April 6-9, 1994.

Kargbo, J.M. (2000). "Impacts of monetary and macroeconomic factors on food prices in eastern and southern Africa", *Applied Economics,* **32**, 1373-1389.

Kargbo, J.M. (2003). "Food prices and long-run purchasing power parity in Africa", *Development Southern Africa Journal,* **20**, # 3, 321 - 336.

Kargbo, J.M. (2005). Impacts of monetary and macroeconomic factors on food prices in west Africa", *Agrekon,* **44** # 2, 205-224.

Kargbo, J.M. (2006) "Exchange rate volatility and agricultural trade under policy reforms in South Africa", *Development Southern Africa,* **23**, # 1, 147- 170.

Kargbo, J.M. (2007a). "Food imports and foreign exchange constraints under macroeconomic adjustment programs in West Africa", *Journal of Policy Modeling*, **29**, 81-85.

Kargbo, J.M. (2007b). "The effects of macroeconomic factors on South African agriculture", *Applied Economics*, **39** # 17, 2211 – 2230.

Kargbo, J.M. (2007c). "Forecasting agricultural exports and imports in South Africa", *Applied Economics*, **39** # 16, 2069 – 2084.

Kargbo, J.M. (2008a). "Energy prices, oil trade flows, and real exchange rates in an integrated global market", in Frank Columbus (Editor), *Trade policy in a globalizing world*, Nova Science Publishers, Inc., forthcoming.

Kargbo, J.M. (2008b). "Energy prices, oil trade flows, and real exchange rates in an integrated global market – Another look at the data", *The International Journal of Energy, Environment and Economics*, **16** # 1, forthcoming.

Missiaen, M. and Shapouri, S. (1994). "Food consumption, nutrition and consumer policies", *Africa and the Middle East: Situation and Outlook Series*. WRS-94-3, USDA/ERS, Washington, DC, pp.45-55.

Organization for Economic Cooperation and Development (OECD) and Food and Agricultural Organization of the United Nations ((FAO), 2008). *OECD-FAO Agricultural Outlook 2008-2017*, Rome, Italy.

Quantitative Micro Software -QMS (1998). *Eviews 3.1 User's Guide*, 2nd Edition. Irvine, California.

Rosen, S. and Shapouri, S. (2008). "Rising food prices intensify food insecurity in developing countries", *Amber Waves*, **6** # 1, 16-21.

Sims, C. (1980). "Macroeconomics and reality", *Econometrica*, **48**, 1-48.

United Nations Economic Commission for Africa ((ECA), 2007). *Eonomic Report on Africa 2007*, Addis Ababa, Ethiopia.

Von Braun, J. (2008). "Rising food prices: what should be done", *IFPRI Policy Brief*, April, 1-2.

World Bank (2008). "High food prices – A harsh new reality", February 29; available at http://www.econ.worldbank.org.

In: Emerging Topics in Macroeconomics
Editor: Richard O. Bailly, pp. 275-301
ISBN: 978-1-60692-093-0

Chapter 10

GREEN NATIONAL ACCOUNTING: THE HICKSIAN NATIONAL INCOME OF CAMBODIA, 1988-2004

Nyda Chhinh [*]
Faculty of Environmental Sciences, Royal University of Phnom Penh
Room 112, RUPP Main Building, Russian Confederation Blvd. Phnom Penh, Cambodia
Philip Lawn [**]
Faculty of Social Sciences, Flinders University
GPO Box 2100, Adelaide, SA, 5001, Australia

1. Introduction

This chapter aims to demonstrate that a green measure of GDP, or what might be better termed as a measure of Hicksian national income, should be included in the formal system of national accounts primarily as an alternative or satellite indicator to GDP. The reason for this is that GDP overstates the national product available for consumption yet is increasingly deployed as a guide to the prudent conduct of national governments (Daly 1989). Since, as we aim to show, Hicksian national income constitutes a more appropriate measure of the sustainable national product available for consumption, it can be used as an indispensable tool for both policy-makers and international donor organisations to design the policies required to achieve sustainable development (SD). Although specific policies to achieve SD is not the main aim of this chapter, the results to be later revealed make it possible to: (a) shed some useful light on the nature of Cambodia's recent economic development — a country still recovering from past wars and the devastating impact of the Pol Pot dictatorship; (b) identify what factors have contributed most to the fall or the lack of increase in Cambodia's Hicksian national income; and (c) outline the nature of the policies that are necessary to increase Cambodia's sustainable national product without having to involve the depletion of its income-generating capital.

[*] E-mail address: nydachhinh32@yahoo.com. Phone : 855 12 955 169
[**] E-mail address: phil.lawn@flinders.edu.au. Phone: 61 8 8201 2838. Fax: 61 8 8201 2644

To achieve its aims, this chapter is organised as follows. Section 2 provides a brief historical overview of Cambodia's social and economic development. In Section 3, a theoretical and empirical overview of green GDP is outlined. In this section, the inadequacies of GDP as a measure of national income are exposed as is the theoretical basis behind the use of green GDP as an alternative measure of sustainable national income. Section 4 presents the methodology employed to calculate the environmental costs and defensive and rehabilitative expenditures associated with the growth of Cambodia's national product. In Section 5, the various costs are deducted from GDP to reveal the Hicksian national income of Cambodia for the period 1988 to 2004. Some general conclusions are outlined in Sections 6 and 7.

2. Cambodia's Economic Development: Historical Background and Recent Developments

2.1. Pre-1953 Independence

The first recognised forms of civilisation in present day Cambodia appeared in the 1st century AD (Tully, 2005). During the 4th and 5th centuries, the Indianised states of Funan and Chenla gradually merged in a region which now constitutes present-day Cambodia and southwestern Vietnam (Chandler, 1983). The process of Indianisation in the region eventually came to a halt following the subjugation of Funan in the 6th century by the rulers of Chenla (Tully, 2005).

In the 9th century, Jayavarman II, a Khmer prince, founded what is conventionally known as the Angkor Empire in an area north of the Tonle Sap (Great Lake) (Chandler, 1983). During the 11th century, Angkor reached its zenith under Suryavarman II, who consolidated Khmer hegemony over the majority of modern Cambodia, Thailand, Laos, and Vietnam. Whilst a succession of battles with neighbouring kingdoms led to a decline in Khmer rule, it remained dominant for the next four centuries until Angkor was politically weakened by Siamese incursions and eventually sacked in the 15th century (Tully, 2005).

It was during the 15th century that the Khmer kingdom sought to regain its former glory through the agency of maritime trade. However, ongoing wars with Thailand and Vietnam continued to result in the loss of Cambodian territory. For the next three centuries, the Khmer kingdom alternated as a vassal state of the Thai and Vietnamese kingdoms with only brief periods of relative independence (Tully, 2005).

In 1863, King Norodom, who had previously been installed by Thailand, appealed to France for protection. The plea proved to be successful with the Thai king in 1867 forced to sign a treaty with France forsaking suzerainty over Cambodia in exchange for control of the Cambodian provinces of Battanbang and Siem Reap (Herz, 1958).[1] Except for the Japanese war-time occupation between 1941 and 1945, Cambodia remained a protectorate of France until 1953, effectively administered in conjunction with the latter's sovereignty over French Indochina (Vietnam).

[1] The two provinces were eventually returned to Cambodia following a border treaty between France and Thailand in 1906.

2.2. Post-1953 Independence to the 1993 Elections

Cambodia finally gained independence from France on November 9, 1953. In the process, it became a constitutional monarchy under King Norodom Sihanouk (Tully, 2005). In 1955, Sihanouk relinquished the throne to his father in order to secure the Prime Ministership of Cambodia. Upon his father's death in 1960, Sihanouk returned as Head of state, this time as the Prince of Cambodia (Osborne, 1973).

With war escalating in neighbouring Vietnam during the 1960s, Sihanouk adopted a neutral stance but was deposed in 1970 by a military coup led by Prime Minister Lon Nol and Prince Sisowath Sirik Matak (Tully, 2005). Safely stationed in Beijing, Sihanouk openly declared his support for the communist Khmer Rouge rebels whom had gradually acquired territory in the remote highland regions of Cambodia.[2] Sihanouk also urged his followers to overthrow the US-supported government of Lon Nol (Tully, 2005). Through his provocative actions, Sihanouk succeeded in triggering a civil war in Cambodia.

Having discovered that the Viet Cong was operating inside the Cambodian border, the USA compounded the deteriorating circumstances in Cambodia by conducting a series of bombing raids on suspected Viet Cong bases and supply routes. Following a brief invasion of Cambodia, the USA expanded its bombing missions to include strikes on the Khmer Rouge (Kiernan, 2002). In 1973, the American bombing ceased. This provided an ideal opportunity for the advancing Khmer Rouge to reach Phnom Penh and seize power, which it accomplished in 1975 (Kiernan, 2002).

Under the dictatorial leadership of Pol Pot (1975-1979), the Khmer Rouge had a devastating impact on Cambodia. Education, religion, and political opposition were almost entirely subverted, while an estimated 1.7 million Cambodians died from execution, starvation, or forced labour (Tully, 2005).[3] Thousands more fled across the border into neighbouring Thailand. Many of the executed were deemed 'enemies of the state' simply by being a civil servant, a person of education or religion, or having had a past association with the previous regime.

Towards the latter stages of Pol Pot's rule, Vietnam invaded Cambodia in an attempt to curtail Khmer Rouge incursions into Vietnam and halt the genocide of Vietnamese Cambodians (Vickery, 1984). Warfare between the two countries continued throughout the 1980s until a peace deal, negotiated in 1989, was finally put into effect in 1991 (Tully, 2005). The peace settlement included a United Nations mandate to enforce a ceasefire as well as mechanisms and institutions, such as the United Nations Transitional Authority in Cambodia (UNTAC), to deal with disarmament and the resettlement of refugees.

2.3. Post-1993 Elections

During 1992, UNTAC arrived in Cambodia to oversee the negotiated ceasefire, disarm combatants, and engage in humanitarian aid and economic reconstruction. UNTAC was also assigned the task of conducting general elections, which were held in May of 1993 (Tully, 2005). Despite threats of Khmer Rouge violence, nearly 90% of enrolled voters exercised

---

[2] The intellectual origins of the Khmer Rouge are detailed in Jackson (1989).

[3] April 17, 1975, the day on which the Khmer Rouge took charge of Cambodia, is often referred to as 'year zero', such was the impact it had on the country and culture (Ponchaud, 1977).

their newly conferred voting rights. Of the 120 National Assembly seats, 58 were won by FUNCINPEC candidates and 51 by the Cambodian Peoples Party (CPP). Owing to the lack of a decisive winner, an interim coalition administration was formed. The National Assembly again recognised Sihanouk as the King of Cambodia while, in order to maintain political harmony, two Prime Ministers were appointed — Prince Norodom Ranariddh (FUNCINPEC) as 'First' Prime Minister and Hun Sen (CPP) as 'Second' Prime Minister (Tully, 2005).

During a period of relative political and economic stability between 1993 and 1997, Cambodia's real GDP rose quite steadily, largely fuelled by the rapid increase in foreign investment and international tourism. However, the growth in real GDP was eventually impeded by two major events. The first was the 1997 Asian financial crisis. The second was the power struggle between the Prime Ministers Ranariddh and Hun Sen which led to a fracture in the relationship between the FUNCINPEC and CPP (Tully, 2005). With the coalition government in turmoil, Ranariddh was forced to flee into temporary exile.

As a consequence of the rising political instability, threats were made by foreign donors to withdraw much needed aid to Cambodia. Pleas were also made for political reconciliation by King Sihanouk. In response, Hun Sen and the CPP agreed to new elections in July 1998. Tarnished by violence and allegations of vote-rigging, the elections resulted in a comfortable victory for the CPP (Tully, 2005). Despite protests emanating from opposition political forces, international agencies proclaimed the poll to be free, fair, and representative of the electorate's preferences.

In 1999, the first full year of relative peace in Cambodia in thirty years, a series of economic reforms were implemented. Real GDP proceeded to grow by 5.0% in both 1999 and 2000. Peace and stability was, however, briefly thrown into chaos in late 2000 by a terrorist attack on government buildings in Phnom Penh involving members of the opposition Sam Rainsy Party. In addition, a coup to oust Prime Minister Hun Sen was attempted by the Cambodian Freedom Fighters (CFF).

Stability was somewhat restored with the strengthening of the CPP's political power following communal elections in February, 2002. Nevertheless, with the CPP winning 1,600 of the 1,621 communes, accusations were again raised regarding voter registration irregularities, vote-buying, intimidation, and media partiality. Violence was also seen to play a part in the unbalanced result with the killing of as many as twenty candidates and activists, mostly from the opposition Sam Rainsy Party (SBS, 2002).

Further instability emerged in Cambodia in 2004 following the shock abdication of King Sihanouk. In order to quell disunity, a special nine-member throne council was hastily formed. The council proceeded to elect Norodom Sihamoni to the position of King of Cambodia on October 14, 2004. In the face of ongoing instability and the fear of renewed political unrest, Cambodia's real GDP managed to increase by 6.3% in 2001 and 5.2% in 2002. In all, Cambodia's real GDP grew at an average annual rate of 6.9% between 1995 and 2004 (World Bank, 2006).

In spite of this decade of rapid growth, the Cambodian economy continues to suffer from internal unrest and the legacy of past wars. Furthermore, the overwhelming majority of the Cambodian population lacks essential education and productive skills. This is particularly so in the poverty-ridden rural areas where basic infrastructure is in short supply.

Due, in part, to the skill shortages just mentioned, the primary industries of fishing, forestry, and agriculture continue to form the backbone of the Cambodian economy. Since many of these industries provide a relatively low financial return, Cambodia possesses one of the highest poverty rates in the Asia-Pacific region (77.7% in 2004).[4]

International concerns over corruption and the possibility of renewed political instability in Cambodia have also kept foreign investment to lower than anticipated levels (Tully, 2005).[5] Moreover, both factors have constantly delayed the receipt and appropriate distribution of foreign aid to needy Cambodians. Recent reports suggest that corruption has led to the illegal transfer of international aid into the private coffers of the Cambodian elite. As a consequence, corruption contributes significantly to the enormous disparity of income within the Cambodian population.

However, the recent economic reforms together with the relative political and economic stability over the past decade (as compared to the pre-1993 circumstances) have no doubt contributed to the increase in Cambodia's average life expectancy and adult literacy rate, as well as the 23% reduction in the number of undernourished citizens as a proportion of the Cambodian population (World Resources Institute, Earthtrends). But with a poverty rate of around 70% (i.e., people living on less than $2US per day) and a population expected to almost double by 2050, Cambodia still has a long way to go to combat widespread impoverishment. What's more, it must approach the poverty issue in the knowledge that its energy consumption and CO_2 emissions are rising steeply and its forest and fish stocks are dramatically in decline. Collectively, these environmental concerns impose a significant constraint on Cambodia's future genuine progress of which the latter will only be sustainably achieved through appropriate and well considered policy measures.

3. Theoretical and Empirical Overview of Green GDP

This section of the chapter details the methodological approach used to calculate the Hicksian national income of Cambodia for the period 1988-2004. The study period thus begins towards the end of the reign of the Khmer Rouge regime and the war with Vietnam. It also includes the decade following the democratic elections in 1993, the relative stability in Cambodia from 1993 to 1997, the Asian financial crisis, Cambodia's own political instability towards the end of 1997, and the high growth period between 1995 and 2004. This Hicksian national income study therefore documents the impact of Cambodia's recent economic development in what can be described as a recovery period in Cambodia's history, yet one frustrated by sporadic instability and a lack of critical infrastructure and civil institutions.

3.1. Theoretical Developments

The theory behind measures of green GDP has developed considerably since the concept first emerged in the early 1970s. The majority of the theoretical developments centre on proposed adjustments to GDP to incorporate the long-term sustainability implications of natural resource depletion and other forms of environmental deterioration. Early on in the development of a green GDP, Weitzman (1976) utilised a dynamic optimisation model to

---

4 The poverty rate revealed here refers to the percentage of the population living on less than $US2 per day.

5 In terms of the Corruption Perceptions Index (Transparency International, www.transparency.org), Cambodia is ranked 151 out of 163 countries.

derive a welfare measure called Net National Product (NNP) — in effect, a variation on the concept of Net Domestic Product (NDP). Weitzman's approach was later translated into a green measure of GDP by Dasgupta and Mäler (2000).

Many other approaches have since been adopted to obtain a measure of green GDP. In the early 1990s, Mäler (1991) suggested that three costs categories should be subtracted from GDP:

- defensive expenditures;
- forms of environmental degradation;
- and the change in the value of stocks of environmental resources.

Asheim (1994) was also prominent in the theoretical developments of green GDP having established indicators of welfare equivalent income, sustainable income, and net social profit. More recently, Cairns (2000) has developed a useful green accounting model to mirror the mismanagement of natural resources in developing countries.

As for the formal national accounting procedures employed by national statistical agencies to calculate, for example, GDP, they are based on the United Nations System of National Accounts (SNA). Introduced more than half a century ago to measure national production levels, the SNA fails to incorporate the impact of environmental damage on national productivity. Stimulated by criticisms levelled at it by the Brundtland Commission in the 1980s (WCED, 1987), the United Nations (UN) revised the SNA in 1993 (UN, 1993). The revised framework is called the System of Integrated Environmental and Economic Accounts (SEEA) and its central aim is to record the stocks and flows of all natural resource assets into the conventional national accounting framework (UN, 1993).

However, the SEEA does not require environmental cost adjustments to measurements of GDP. As such, the SEEA is merely a satellite accounting report to accompany the original SNA framework in readiness for any future estimation of environmentally-adjusted income. To date, environmental cost adjustments to GDP have not been mandated by the United Nations.

3.2. The Green GDP Methodology Used in the Cambodian Study

The methodology used in this chapter for the calculation of Cambodia's true national income is based on the definition of income first posited by Hicks (1946) — namely, that income ought to reflect the maximum quantity of goods a nation can consume without undermining its capacity to consume the quantity of goods over time. One of the essential requirements of the Hicksian concept of income is the need to keep income-generating capital intact. Indeed, any measure of income that counts, as income, the depletion of income-generating capital, effectively overstates 'true' income in the Hicksian sense.

With this in mind, it is worth considering the extent to which GDP constitutes a measure of Hicksian income and, should it not, what adjustments ought to be made to bring GDP into line with this more appropriate definition. GDP is a monetary measure of the goods and services annually produced by domestically *located* factors of production (i.e., by the natural and human-made capital located in a particular country). By natural capital, we mean forests, sub-soil assets, fisheries, water resources, and critical ecosystems. Human-made capital, on the other hand, includes the stock of producer goods (e.g., plant, machinery, and equipment) that are used to produce consumer goods and replacement producer goods.

GDP can be measured in *nominal* or *real* values. If GDP is measured in nominal values, it is measured in terms of the prices at the time of production. On the other hand, if GDP is measured in real values, it is measured in terms of the prices of all goods and services in a particular year — often referred to as the *base* year.

The best way to embark on an assessment of GDP is to first consider whether it would be possible for a nation to consume its entire output and still be in a position to consume at least as much output in the following year and beyond. The answer to this is, of course, no. To begin with, some of a nation's output must be set aside to replace worn out and depreciated human-made capital. Secondly, the generation of a nation's GDP also involves the depletion of natural capital. If a nation failed to replace or maintain these forms of capital, it would be unable, eventually, to maintain its productive capacity at a level necessary to sustain the same consumption stream over time. Clearly, the portion of a nation's output that is required to maintain income-generating capital intact ought not to be used for current consumption purposes. This portion cannot, therefore, be classed as true income.

In addition to this, there are other elements of a nation's annual output that are used for defensive and rehabilitative purposes that, in turn, assist in sustaining output over time (Leipert, 1986). For example, vehicle accident repairs and some medical procedures take place to restore human beings and their productive instruments to something approximating their previous condition. In doing so, the output generated in both instances is not used directly for consumption purposes — it is produced merely to maintain the productivity of human beings, as labour, and the existing stock of human-made capital.

Examples of output produced for defensive purposes include flood mitigation projects and crime prevention measures. Somewhat differently, however, output generated in these circumstances occurs to prevent future economic activity from impacting deleteriously on the existing stock of natural and human-made capital (i.e., to minimise future rehabilitative expenditures).

Not unlike the depreciation of human-made and human-made capital, the value of all output produced for rehabilitative and defensive purposes cannot be directly consumed without undermining the capacity to sustain future output. Nor, then, can it be classed as true income.

In all, it has been suggested that a better measure of national income can be calculated by subtracting from GDP the value of depreciated human-made capital and depleted natural capital as well as all defensive and rehabilitative expenditures. Thus, for the purposes of this study, Cambodia's Hicksian national income has been calculated by adhering to the following equation (Daly, 1996):

$$\text{Hicksian national income} = \text{GDP} - \text{DHK} - \text{DNK} - \text{DRE} \tag{1}$$

where:

GDP = Gross Domestic Product

DHK = depreciation of human-made capital (producer goods)

DNK = depletion of natural capital

DRE = defensive and rehabilitative expenditures.

As mentioned at the beginning of this section, a key aspect of the Hicksian income concept is the need to keep income-generating capital intact. If it is assumed that human-made capital can be substituted for natural capital, it is only necessary to maintain intact a combined stock of human-made and natural capital. This form of sustainability is known as *weak sustainability*. If, however, it is assumed that the two forms of capital are complementary, it is necessary to ensure both forms of capital individually do not decline over time, particularly natural capital. This alternative form of sustainability, which is advocated by ecological economists, is known as *strong sustainability* (Pearce, 1993; Daly, 1996; El Serafy, 1996; Lawn, 2003 and 2007).

For the purposes of this chapter, the strong sustainability concept will be followed. This is based on arguments put forward by Georgescu-Roegen (1979) and Daly (1996) which, in our view, have successfully demonstrated that human-made capital is not a long-term substitute for natural capital. The adoption of the strong sustainability approach has significant implications for the way in which the category of natural resource depletion is calculated prior to its subtraction from GDP. Owing to the limited availability of Cambodian data, many of the valuation methods used in the adoption of a strong sustainability approach to Hicksian national income cannot be employed in a practical sense. Alternative methods have therefore been employed.

3.3. Empirical Overview of Green GDP Studies

In the early 1970s, and in response to the *Limits to Growth* thesis by the Club of Rome (Meadows et al., 1972), Nordhaus and Tobin (1972) adjusted GDP in order to calculate a Measure of Economic Welfare (MEW) for the United States between 1929 and 1965. This work was lauded but also heavily critiqued by Daly and Cobb (1989) who argued that the MEW failed to go far enough in its attempt at calculating a true measure of national income.

Many empirical studies involving environmental adjustments to GDP have since been calculated. They include estimates for Germany (Leipert, 1989), Mexico (Van Tongeren et al., 1993), Ecuador (Kellenberge, 1996), Sweden (Skanberg, 2001), the Netherlands (Gerlagh et al., 2002), and India (Atkinson and Gundimeda, 2005). However, three additional and more comprehensive empirical studies have contributed most to the methodology used in this chapter. They have been performed by Repetto et al. (1989), Young (1990) and Foy (1991).

In the well known Repetto et al. (1989) study on Indonesia, an adjustment to the GDP was undertaken by deducting the costs of three natural capital assets: forests, petroleum, and soil erosion. This study enabled Repetto et al. to conclude that, having subtracting estimates of net natural resource depletion, the conventional measure of Indonesia's GDP overstated the country's Hicksian income. The results of the study showed that the average annual rate of growth in GDP between 1971 and 1984 was 7.1%, while Net Domestic Product — the term given by Repetto et al. for the green GDP estimate — grew by an average annual rate of just 4.0% (Repetto et al., 1989, p. 4). Significantly, the adjustment to Indonesia's GDP was very conservative since the environmental cost deduction was confined to the depletion of three natural resource assets. In addition, there were no calculations and subsequent adjustments for both human-made capital depreciation (HCD) and defensive and rehabilitative expenditures (DRE).

A year later, Young (1990) suggested that environmental and natural resource accounts should be integrated into Australia's national accounting framework in order to provide a more accurate estimate of Australia's national income. The study conducted by Young incorporated a greater range of natural resources than Repetto et al. and included such environmental costs as land degradation, deforestation, and mineral depletion. Young concluded from his study, which revealed a difference in the growth rates of conventional GDP and green GDP, that necessary steps should be taken by the Australian Government to arrive at a more reliable and meaningful national economic indicator than GDP.

In 1991, Foy (1991) adjusted the Gross State Product (GSP) of the US state of Louisiana for the period 1963 to 1986. Foy concluded that, when non-renewable resource depletion was taken into account, the average annual reduction in Louisiana's GSP was 3.3% lower using a depreciation approach, yet as much as 13.8% lower when using the El Serafy (1989) user cost approach. Only two non-renewable resource items — namely, oil and gas — were calculated and incorporated into the study. Again, the adjustment to income, this time at the state level, was very conservative and likely to have underestimated the full cost of environmental damage.

Based on the theoretical and empirical reviews of Hicksian income conducted above, it can be seen that adjustments made to GDP, albeit arbitrary in many cases, constitute a significant step towards: (a) the development of a better measure of green GDP and, therefore, a better indicator of sustainable development, and (b) a more appropriate economic compass for policy-makers to use than GDP. Although, many scholars (e.g., Repetto et al., Young, and Foy) have incorporated only some of the adjustments to GDP suggested by Daly (see equation 1), there is little doubt that GDP is a poor representative of Hicksian national income (likewise GSP at the state or provincial level).

4. Valuation Methods Used to Calculate Cambodia's Hicksian National Income

This study employs a number of valuation approaches to calculate Cambodia's Hicksian national income. The basic doctrine used to compute Hicksian income is modified given the Cambodian data limitations. Moreover, to arrive at the imputed environmental cost attributed to resource depletion, a conservative approach is adopted that is primarily based on the methods and results from other studies, as detailed below.

4.1. Human-Made Capital Depreciation (HCD)

The first of the deduction items in equation (1) is the depreciation of human-made capital. Human-made capital depreciation is often referred to in conventional national accounts as the 'consumption of fixed capital' and equals the amount charged for the use of both private sector and government fixed capital located within a country (e.g., plant, machinery, buildings, etc.). Consumption of fixed capital is invariably estimated by computing the decline in the value of fixed assets due to wear and tear, obsolescence, accidental damage, and aging.

Because the consumption of fixed capital has not been regularly calculated by the Cambodian statistical bureau, this value had to be estimated. For the year 2000, the World Resources Institute (WRI) argued that the Cambodia's consumption of fixed capital was equivalent to 7% of its Gross National Income (GNI) (WRI, 2003). In 2000, Cambodia's GNI was 11,815.0 billion riels[6] (UNSD, 2005). Using equation (2) below, the consumption of

[6] The Cambodia currency is referred to as 'riels'. In 2002, US$1 = 3,850 riels.

fixed capital in 2000 was 827.0 billion riels. Unfortunately, there are no GNI estimates available for the remaining years of the study period. Instead, GDP was used to determine the values for the consumption of fixed capital. GDP can be used because there is a close association between GDP and GNI.[7]

Table 1. Human capital depreciation (HCD) - Cambodia, 1988-2004

Year	GNI at 2000 prices (billion riels) *a*	CFC at 7% of GNI in 2000 (billion riels) *b* (*0.07 × a*)	GDP at 2000 prices (billion riels) *c*	CFC as % of GDP *d* (*b/c*)	CFC (HCD) (billion riels) *e* (*c × d*)
1988	-	-	6.688	0,059	391,45
1989	-	-	6.900	0,059	405,07
1990	-	-	6.981	0,059	409,78
1991	-	-	7.511	0,059	440,89
1992	-	-	8.038	0,059	471,83
1993	-	-	8.496	0,059	498,73
1994	-	-	9.277	0,059	544,58
1995	-	-	9.883	0,059	580,15
1996	-	-	10.411	0,059	611,14
1997	-	-	10.999	0,059	645,66
1998	-	-	11.545	0,059	677,71
1999	-	-	12.994	0,059	762,77
2000	*11.815,0*	*827,1*	*14.089*	*0,059*	*827,05*
2001	-	-	14.863	0,059	872,49
2002	-	-	15.643	0,059	918,27
2003	-	-	16.745	0,059	982,96
2004	-	-	18.032	0,059	1.058,51

Sources: World Bank (1994 and 2006); WRI (2003); UNSD (2005)

At 827.0 billion riels, the consumption of fixed capital constitutes 5.9% of GDP (equation 3). The values of the consumption of fixed capital were therefore assumed to equal 5.9% of GDP in all remaining years of the study period. The value of consumption of fixed capital or human-made capital depreciation is found in Table 1. The negative value of human-made capital depreciation indicates that this item is a *cost* and should be subtracted from GDP when calculating a nation's Hicksian national income.

$$\begin{aligned} HCD_{2000} &= GNI_{2000} \times 0.07 \\ &= 11{,}815 \text{ billion riels} \times 0.07 \\ &= 827.0 \text{ billion riels} \end{aligned} \tag{2}$$

$$\frac{HCD_{2000}}{GDP_{2000}} = \frac{827}{14{,}089} = 0.059 \tag{3}$$

[7] Gross national Income (GNI) measures the total income of all people who are citizens in a given country while GDP measures the value of the total output of all persons living in that country.

4.2. Defensive and Rehabilitative Expenditures (DRE)

Like most studies involving the subtraction of defensive and rehabilitative expenditures (e.g., Daly and Cobb, 1989; Lawn, 2001; Clarke and Lawn, 2005), it was assumed for this study that half of all government expenditure on education and health is required to help maintain the nation's productive capacity due to the ongoing negative impacts of past and present economic activities.

The data available on government health and education expenditure only exists for the period 1996 to 2004. To estimate the data for the missing years (1988-1995), a number of steps were undertaken. Firstly, government health and education expenditure as a percentage of GDP was computed for each year from 1996 to 2004. Secondly, the average percentage value for this period was calculated. It was then assumed that this average (2.1%) was the same percentage value for the missing years. Finally, as explained above, the annual value of defensive and rehabilitative expenditure (DRE) was assumed to be 50% of government expenditure on health and education (Daly and Cobb, 1989; Lawn 2001). That is:

$$\text{DRE} = 0.5 \times \text{Government expenditure on health \& education} \tag{4}$$

The calculations and the annual DRE values for Cambodia are revealed in Table 2 below.

Table 2. Defensive and rehabilitative expenditure (DRE) - Cambodia, 1988-2004

	Govt. health & education exp. (billion riels)	DRE (billion riels)	GDP (billion riels)	DRE as % of GDP
Year	*a*	*b* (*0.5* × *a*)	*c*	*d* (*b/c*)
1988	134,03	67,01	6.668	0,010
1989	138,69	69,34	6.900	0,010
1990	140,30	70,15	6.981	0,010
1991	150,95	75,48	7.511	0,010
1992	161,55	80,77	8.038	0,010
1993	170,76	85,38	8.496	0,010
1994	186,45	93,23	9.277	0,010
1995	198,63	99,32	9.883	0,010
1996	140,46	70,23	10.411	0,007
1997	138,98	69,49	10.999	0,006
1998	130,04	65,02	11.545	0,006
1999	271,35	135,68	12.994	0,010
2000	*344,00*	*172,00*	*14.089*	*0,012*
2001	343,07	171,53	14.863	0,012
2002	444,04	222,02	15.643	0,014
2003	445,93	222,96	16.745	0,013
2004	429,90	214,95	18.032	0,012

Sources: World Bank (1994 and 2006); Cambodian Development Review (2005)

4.3. Natural Capital Depletion (NCD)

The El Serafy (1989) user cost method was employed to calculate the true income and the set-aside components from the net receipts generated from the depletion of Cambodia's natural resource assets. The formula used to calculate the income and set-aside components is:

$$R - X = R\left(\frac{1}{(1+r)^{n+1}}\right) \tag{5}$$

where:

X = income (resource rent);

R = total receipts (the value of harvesting minus the costs incurred from harvesting the resource in excess of the sustainable harvesting rate);

$R - X$ = the user cost or depletion factor that should be set aside as a capital investment to establish a suitable replacement asset. This amount must be deducted from GDP when calculating a nation's Hicksian national income;

r = discount rate or regeneration rate of the replacement asset (alternative renewable resource);

n = number of periods over which the resource would be exhausted if extracted at the current rate.

The user cost formula is normally applied to a non-renewable resource. However, it can be employed to calculate the user cost of a renewable resource if harvesting of the resource exceeds its natural regeneration rate. Excessive harvesting implies that the resource will eventually be exhausted. In these circumstances, harvesting of a renewable resource is no different to the extraction of a non-renewable resource.

Choosing the appropriate discount rate (r) is very critical.[8] Ultimately the discount rate chosen depends on whether one adopts the *weak* or *strong sustainability* objective. If a weak sustainability approach is adopted, the asset established to replace the depleted natural resource can be in the form of human-made capital. The interest rate generated by a typical human-made capital asset is around 7%. However, if a strong sustainability approach is embraced, the replacement asset must be a renewable resource. The interest rate on a renewable resource is effectively its natural regeneration rate. This is typically in the order of 1 to 2%.

The significance of the chosen discount rate can be illustrated by the following. Consider a scenario where the life expectancy of a resource is 60 years (n = *60 years*). At a discount rate of r = *1%* (strong sustainability), 54.5% of the net receipts from resource depletion must be set aside to establish a replacement resource asset (i.e., a cultivated renewable resource). In other words, for every 100 riels received, 45.5 riels would constitute genuine income that could be used for consumption purposes. The remaining

[8] The discount rate deployed when using the El Serafy user cost approach is rationally explained in Lawn (1998, 2001, and 2007).

54.5 riels would constitute the user cost of the resource to be subsequently subtracted when calculating a nation's Hicksian national income. Conversely, at a discount rate of $r = 7\%$ (weak sustainability), just 1.6% of all net receipts would have to be set aside to establish a replacement asset which, under weak sustainability conditions (although incorrectly in our view), can exist in any form. In the weak sustainability case, 98.4 riels would constitute income while only 1.6 riels would need to be set aside for reinvestment. As indicated earlier, this empirical study of Cambodia has employed the strong sustainability approach. A discount rate of 1% was therefore used to calculate the user cost of forest and fish depletion.

Valuing the Cost of Deforestation

The environmental cost of deforestation represents the 'opportunity cost' of forest depletion. It equals the value of the forest that could be obtained over time if the deforestation had not taken place. The larger is the rate of forest clearance, the higher is the opportunity cost of forest depletion. The opportunity cost of forestry operations is negligible if the rate of harvesting does not exceed the regeneration rate of the forest.

To arrive at the imputed environmental cost on deforestation, two variables are required. The first variable is the 'regeneration rate' of Cambodia's forest stocks. This, according to the Department of Fisheries (2001a), is 0.33 m^3 per hectare per year. Given the assumption that the 'volume over bark'[9] of Cambodia's forests is the same as it is for Indonesia's forests — that is, 219 m^3 per hectare (Repetto el al., 1989) — the regeneration rate amounts to:

$$\left(0.33m^3 / 212.9m^3\right) \times 100\% = 0.155\% \tag{6}$$

The second variable is the 'harvesting rate' of Cambodia's forests. According to the Department of Fisheries (2001a), the harvesting rate was 0.57% from 1988 to 1994, but 3% from 1995 and 2004 (World Bank, 2003). Based on the estimated regeneration rate and the 2004 harvesting rate, Cambodia's major forest areas are likely to be fully exhausted in 26 years. Hence, the variables employed to calculate the user cost of deforestation were:

R: Total receipts = Value of the change in forest area

r: Regeneration rate = 0.155 %

n: Number of year to completely deplete the forest (26 years).[10]

The calculations and user cost values of Cambodia's forest depletion are revealed in Table 3.

9 *Volume over bark* is the gross volume in cubic metres per hectare over bark of free bole (from stump or buttresses to crown point of first main branch) of all living trees more than 10 centimetres in diameter at breast height (Repetto et al. 1989, p. 29).

10 The value of n = 31 years is derived on the assumption that the total of harvestable forest areas will be exhausted except for protected areas.

Table 3. User cost of forestry assets - Cambodia, 1988-2004

Year	Annual change in forest area (cubic metres) *a*	Forest value per hectare (riels) *b*	Value of change in forest area (billion riels) *c* ($a \times b$)	User cost (1 - X/R) *d*	User cost (R - X) (billion riels) *e* ($c \times d$)
1988	-10.667.946	6.149	-65,59	0,959	-62,91
1989	-10.623.152	6.149	-65,32	0,959	-62,64
1990	-10.578.547	6.149	-65,04	0,959	-62,38
1991	-10.534.129	6.149	-64,77	0,959	-62,12
1992	-10.489.898	6.149	-64,50	0,959	-61,86
1993	-10.445.852	6.149	-64,23	0,959	-61,60
1994	-10.401.991	6.149	-63,96	0,959	-61,34
1995	-86.358.607	6.149	-530,98	0,959	-509,24
1996	-83.335.490	6.149	-512,39	0,959	-491,41
1997	-80.418.202	6.149	-494,46	0,959	-474,21
1998	-77.603.038	6.149	-477,15	0,959	-457,61
1999	-74.886.423	6.149	-460,44	0,959	-441,59
2000	-72.264.907	6.149	-444,33	0,959	-426,13
2001	-69.735.162	6.149	-428,77	0,959	-411,21
2002	-67.293.974	6.149	-413,76	0,959	-396,82
2003	-64.938.244	6.149	-399,28	0,959	-382,93
2004	-62.664.980	6.149	-385,30	0,959	-369,52

Sources: World Bank (2003); Department of Fisheries (2001b)

Valuing the Cost of Overfishing

The methodology used to calculate the imputed environmental cost or opportunity cost of overfishing is the same used for deforestation. Obviously, the variation in the two calculations relates to the different regeneration rate of fish and as well as the number of years it would take to deplete fish stocks if they were harvested at present rates. The regeneration rate was assumed to be one percent ($r = 1\%$). This regeneration rate is based on evidence showing that, although “reproductive female fish generate large numbers of eggs, in some cases numbering in the millions per female, very few of these eggs grow and survive to reproductive maturity” (Kahn, 2005).

The estimate of the stock changes revealed in Table 4 was based on: (a) the increase in ‘catch per units’ and fish production claimed by Khy (2005), and (b) the annual harvest report by the Department of Fisheries (2001b). Given current fish demands in Cambodia, it was assumed that it would take thirty years (n = *30 years*) to exhaust existing fish stocks at current harvesting rates. Hence, the variables employed to calculate the user cost of fish stock depletion were:

R: total receipts = Value of the change in fish stocks

r: Discount rate = regeneration rate = 1.0% per annum

n: Number of years to completely deplete fish stocks at the current rates of harvesting (assumed to equal 30 years).

Table 4. User cost of fishery assets - Cambodia, 1988-2004

Year	Annual change in fish stocks (tonnes)	Fish value per tonne (billion riels)	Value of change in fish stocks (billion riels)	User cost (1 - X/R)	User cost (R - X) (billion riels)
	a	*b*	*c* ($a \times b$)	*d*	*e* ($c \times d$)
1988	-6.800	0,01114	-75,78	0,795	-60,28
1989	-2.156	0,01114	-24,03	0,795	-19,11
1990	-31.490	0,01114	-350,94	0,795	-279,15
1991	-38.504	0,01114	-429,11	0,795	-341,33
1992	-33.990	0,01114	-378,80	0,795	-301,31
1993	-29.529	0,01114	-329,09	0,795	-261,77
1994	-24.065	0,01114	-268,19	0,795	-213,33
1995	-33.444	0,01114	-372,72	0,795	-296,48
1996	-25.710	0,01114	-286,52	0,795	-227,91
1997	-36.540	0,01114	-407,22	0,795	-323,92
1998	-44.622	0,01114	-497,29	0,795	-395,57
1999	-35.168	0,01114	-391,94	0,795	-311,76
2000	-59.451	0,01114	-662,55	0,795	-527,02
2001	-118.515	0,01114	-1.320,79	0,795	-1.050,61
2002	-349.600	0,01114	-3.896,12	0,795	-3.099,14
2003	-354.338	0,01114	-3.948,92	0,795	-3.141,14
2004	-257.739	0,01114	-2.872,38	0,795	-2.284,81

Sources: World Bank (2006); MAFF (undated)

Valuing the Cost of Soil Erosion

The environmental cost of soil erosion equals the monetary value associated with soil degradation. Given the nature of the Cambodian environment, soil degradation not only reduces agricultural productivity, but also the productivity of its forests.

Unfortunately, there is no reliable study of soil erosion costs in Cambodia. Nonetheless, Khor (undated) has argued that, in the East Asia region, soil degradation comprises 7% of agricultural losses. Moreover, an empirical study by Yang (undated) on the impact of soil erosion on crop yields reveals a 40% decline in agricultural productivity in the northeast region of Cambodia. This study conservatively assumes that the soil erosion cost in Cambodia over the study period equates to 7% of agricultural output. This is relatively small when compared to the estimated cost of soil erosion in Java at 40% of GDP (Magrath et al. in Barbier, 1991). The cost of soil erosion in Cambodia was therefore calculated by adopting the following equation:

$$\text{Soil erosion cost} = \text{Total value of agriculture} \times 0.07 \qquad (7)$$

The total value of Cambodia's agricultural product was derived from the World Bank (1994 and 2006). It is revealed in Table 5 along with the estimated cost of soil erosion in Cambodia.

Table 5. Cost of soil erosion - Cambodia, 1988-2004

Year	Value of agricultural output (billion riels) *a*	Soil erosion cost as 7% of agric. value *b*	Cost of soil erosion (billion riels) *c* ($a \times b$)
1988	1.850,63	-0,07	-129,54
1989	2.044,94	-0,07	-143,15
1990	2.000,53	-0,07	-140,04
1991	2.106,02	-0,07	-147,42
1992	2.144,88	-0,07	-150,14
1993	2.354,00	-0,07	-164,78
1994	2.375,00	-0,07	-166,25
1995	2.605,00	-0,07	-182,35
1996	2.673,00	-0,07	-187,11
1997	2.646,00	-0,07	-185,22
1998	2.846,00	-0,07	-199,22
1999	3.146,00	-0,07	-220,22
2000	3.085,00	-0,07	-215,95
2001	3.178,00	-0,07	-222,46
2002	3.059,00	-0,07	-214,13
2003	3.668,00	-0,07	-256,76
2004	3.612,00	-0,07	-252,84

Sources: World Bank (1994 and 2006); Khor (undated)

Valuing the Cost of Air Pollution

Air pollution constitutes a form of natural capital depletion in the sense that it reflects the erosion of a nation’s waste assimilative capacity. The imputed air pollution cost should thus equal the amount a nation spends to rehabilitate the health of affected citizens, to repair buildings and other structures, plus the cost of reduced agricultural production as a consequence of any reduced assimilative capacity.

No empirical study has been conducted on air pollution costs in Cambodia. The imputed cost of air pollution in Cambodia used for this chapter was therefore based on a case study in China by Deng (2006) using two methods of valuation. For the year 2000, Deng found that the air pollution cost for China was 3.26% of GDP based on a willingness-to-pay approach, and 0.7% of GDP based on a human capital approach.

For this chapter, the latter estimation of the imputed cost of air pollution was adopted — that is, 0.7% of a nation’s GDP. It is conservatively low compared to other studies. For example, Kan and Chen (2004) found that the air pollution cost was 1.03% of GDP in Shanghai, and between 3 to 6% of GDP in India (Lvovshy, 1998). The lower value of 0.7% was chosen because the health cost of air pollution is to some extent already deducted in terms of the item capturing defensive and rehabilitative expenditures. This approach avoids double-counting.

Because the Deng (2006) study is based on China, it was assumed that the per unit GDP cost of air pollution is the same in both Cambodia and China. Thus, in order to impute the air pollution cost in Cambodia, a number of steps were required. Firstly, the cost of air pollution in China for the year 2000 was calculated. In 2000, China’s GDP was US$1,080.7 billion (Deng, 2006). At 0.7% of GDP, the total air pollution cost in China amounts to US$7,565.2 million. Secondly, the unit cost of air pollution per metric tonnes of CO_2 emissions in China was only calculated for the year 2000.[11] In this particular year, China's CO_2 emissions were 3,469,117 metric tonnes (WRI, 2006). By dividing the total pollution cost by total CO_2 emissions, one obtains a pollution cost for China in 2000 equal to US$2,180 per metric tonne of CO_2 emissions.

To calculate the total air pollution cost in Cambodia in 2000 (see Table 6), the unit cost of air pollution in China was converted to Cambodian currency (*riels*) at the prevailing exchange rate in the year 2000 (3,850 riels per US dollar). By performing this calculation, the air pollution cost for Cambodia amounted to 8.415 million riels per metric tonne of CO_2 emissions. This unit price was applied to the remaining years of the study period, as shown in equation (8).

$$\text{Air pollution cost} = \text{total emissions of } CO_2 \times 8.415 \text{ million riels} \quad (8)$$

Table 6. Cost of air pollution - Cambodia, 1988-2004

Year	Total CO_2 emissions (000s of metric tonnes)	Unit price per tonne of CO_2 emissions (riels)	Cost of air pollution (billion riels)
	a	*b*	*c* (-*a* × *b*)
1988	450,70	8.415	-3,79
1989	450,70	8.415	-3,79
1990	450,70	8.415	-3,79
1991	461,70	8.415	-3,89
1992	476,30	8.415	-4,01
1993	476,30	8.415	-4,01
1994	588,50	8.415	-4,95
1995	599,50	8.415	-5,05
1996	700,60	8.415	-5,90
1997	661,00	8.415	-5,56
1998	661,00	8.415	-5,56
1999	524,00	8.415	-4,41
2000	531,30	8.415	-4,47
2001	581,90	8.415	-4,90
2002	596,50	8.415	-5,02
2003	620,67	8.415	-5,22
2004	650,46	8.415	-5,47

Sources: WRI (2006a); Deng (2006)

5. Results of the SNDP Study

5.1. Cambodia's Real GDP and Hicksian National Income

The Hicksian national income results for Cambodia for the period 1988-2004 are summarised in Table 7. As the table reveals, Cambodia's real GDP was 6,668 billion riels in 1988 (column *a*) while its Hicksian income was 5,953 riels (column *e*). By the end of the study period, real GDP had grown to 18,032 billion riels. Although Cambodia's Hicksian income also increased over the study period, it was considerably lower at 13,846 billion riels.

11 The unit price of year 2000 is calculated because year 2000 is the year base of real GDP.

Table 7. Per capita real GDP versus per capita Hicksian income - Cambodia, 1988-2004

Year	Real GDP (billion riels)	HCD (billion riels)	NCD (billion riels)	DRE (billion riels)	Hicksian income (billion riels)	Population (000s)	Per capita real GDP (riels)	Per capita Hicksian income (riels)
	a	*b*	*c*	*d*	*e* (*a* - *b* - *c* - *d*)	*f*	*g* (*a*/*f*)	*h* (*e*/*f*)
1988	6.668	391,45	256,52	67,01	5.953,48	6.463,0	1.031.792,0	921.164,0
1989	6.900	405,07	228,69	69,34	6.197,38	6.978,4	988.839,0	888.083,5
1990	6.981	409,78	485,36	70,15	6.015,41	7.493,8	931.535,0	802.722,6
1991	7.511	440,89	554,76	75,48	6.439,50	8.009,1	937.756,9	804.019,4
1992	8.038	471,83	517,32	80,77	6.967,77	8.524,5	942.890,5	817.379,9
1993	8.496	498,73	492,16	85,38	7.419,73	9.039,9	939.833,8	820.776,3
1994	9.277	544,58	445,87	93,23	8.193,33	9.555,3	970.877,3	857.466,2
1995	9.883	580,15	993,11	99,32	8.210,42	10.070,7	981.366,2	815.282,0
1996	10.411	611,14	912,33	70,23	8.817,30	10.586,0	983.465,6	832.918,0
1997	10.999	645,66	988,91	69,49	9.294,94	11.101,4	990.774,7	837.275,3
1998	11.545	677,71	1.057,96	65,02	9.744,31	11.616,8	993.819,9	838.812,4
1999	12.994	762,77	977,98	135,68	11.117,57	12.132,2	1.071.036,6	916.371,1
2000	14.089	827,05	1.173,57	172,00	11.916,38	12.351,4	1.140.682,1	964.780,8
2001	14.863	872,49	1.689,18	171,53	12.129,80	12.573,6	1.182.081,8	964.705,4
2002	15.643	918,27	3.715,10	222,02	10.787,60	12.803,0	1.221.824,7	842.585,3
2003	16.745	982,96	3.786,05	222,96	11.753,03	13.040,7	1.284.060,0	901.259,5
2004	18.032	1.058,51	2.912,64	214,95	13.845,89	13.287,1	1.357.110,6	1.042.059,0

Note: All values in billions of riels except where indicated

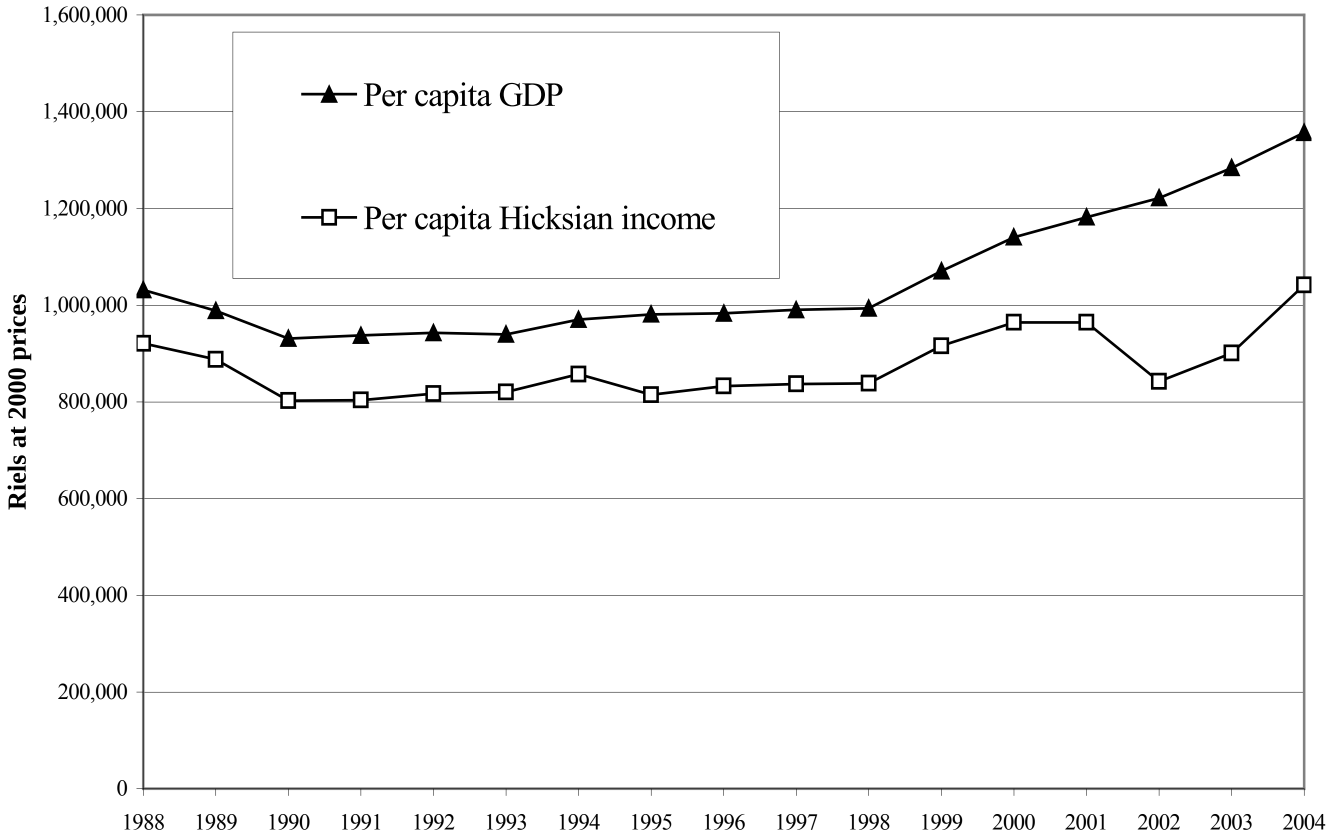


Figure 1. Per capita real GDP versus per capita Hicksian income - Cambodia, 1988-2004.

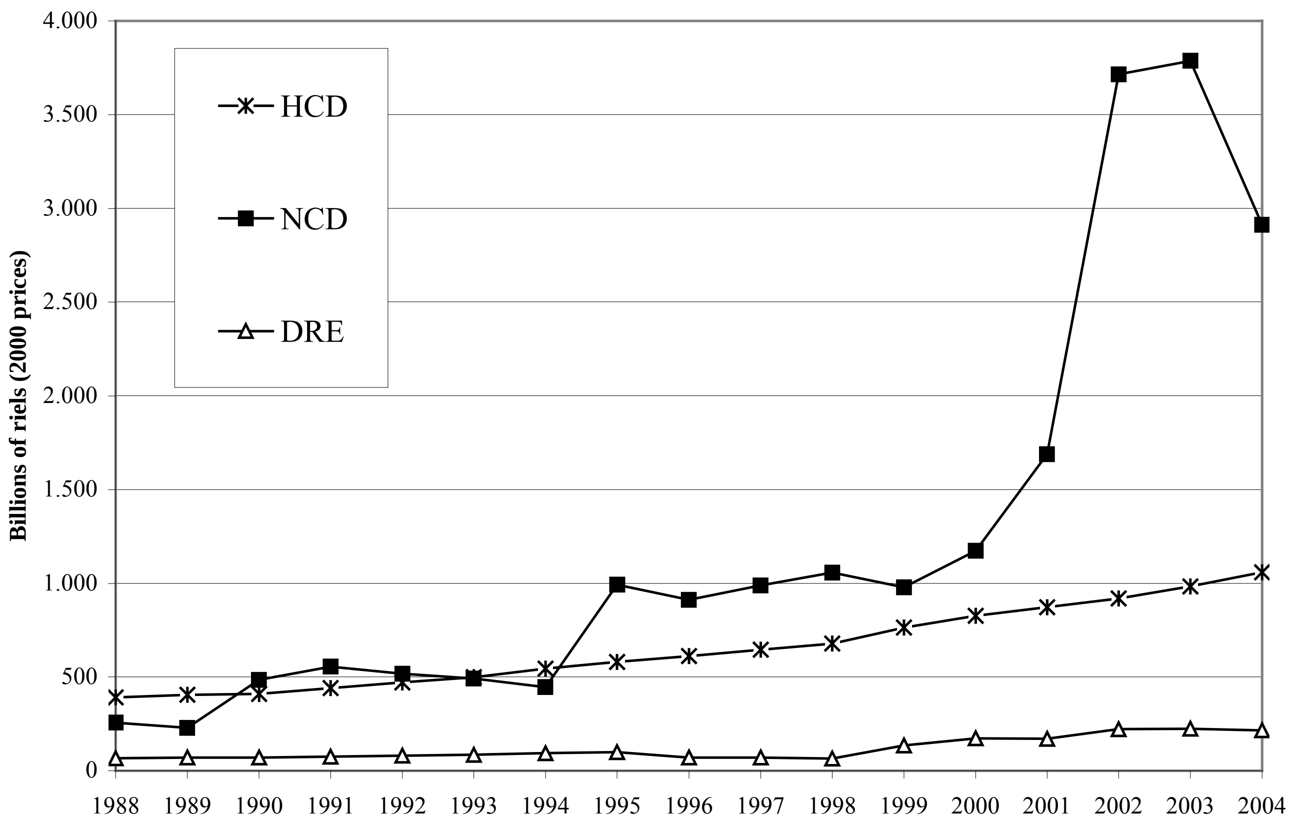


Figure 2. Comparison of NCD, HCD, and DRE - Cambodia, 1988-2004.

Crucially, the disparity between real GDP and Hicksian national income increased from 715 billion riels in 1988 to 4,186 billion riels by 2004. This, in effect, constitutes a 6-fold increase in the gap between Cambodia's real GDP and Hicksian national income over the study period. It is therefore clearly apparent that Cambodia's real GDP substantially overstated Cambodia's sustainable income throughout the study period. Indeed, it was an average of 20.6% higher than Hicksian national income over the entire study period, but a more substantial 30.2% higher in 2004.

Figure 1 provides a comparison between Cambodia's per capita real GDP and per capita Hicksian income (columns *g* and *h* in Table 7). Between 1988 and 1998, there was very little increase in the per capita value of both indicators. In addition, the difference between the two remained steady until 1994 upon which there was a sharp decline in per capita Hicksian income in 1995.

From 1998 to the end of the study period in 2004, Cambodia's per capita real GDP rose significantly. Whilst, in 1999 and 2000, per capita Hicksian income also increased steeply, there was a marked downturn in per capita Hicksian income in 2001 and 2002. Per capita Hicksian income recovered in 2003 and 2004, however, the disparity between Cambodia's per capita real GDP and per capita Hicksian income was more substantial than it was pre-2001. This suggests that the per capita GDP growth spurt between 1988 and 2004 did not genuinely equate to a similar increase in sustainable per capita income and that, relative to the early part of the study period, the much higher per capita level of GDP after 2001 was fuelled by an increased consumption of income-generating capital. Interestingly, this latter phenomenon occurred not long after the implementation of the 1999 economic reforms. This evidence does not imply that each reform had a negative impact and that Cambodia would have fared better towards the end of the study period had they not been introduced. But it does suggest that their full impact should be closely examined to determine which reforms ought to remain and which should perhaps be abandoned, or at least be modified.

5.2. Component Items of Cambodia's Hicksian National Income

To obtain a better understanding of the trend movement of Cambodia's Hicksian income, it is necessary to examine the individual cost items subtracted from GDP. Table 7 (column *c*) and Figure 2 show that the dominant item influencing Cambodia's Hicksian income was the cost of natural capital depletion (NCD). This was followed in importance by the depreciation of human-made capital (HCD) (column *b*) and, finally, DRE (column *d*). In 2004, the cost of NCD was 2,912.6 billion riels, while HCD and DRE were 1,058.5 billion riels and 215.0 billion riels respectively. The cost of NCD was therefore nearly three times greater than HCD and almost 13.6 times bigger than DRE.

There are two other very important aspects about NCD worth noting. Between 1988 and 1994, NCD remained relatively steady. In 1995, however, NCD rose considerably and remained at the much higher level through to 2000. It was in 1995 that Cambodia's per capita Hicksian income dipped. Secondly, NCD increased dramatically in 2001 and 2002. This coincided with the rapid increase in Cambodia's per capita GDP but the sharp decline in per capita Hicksian income.

Of the individual environmental cost items that have been used to calculate the annual value for NCD, the cost of fish and forest depletion made the greatest contribution. Air pollution costs were relatively minor, but rising, while the cost of soil erosion effectively doubled over the study period. Both air pollution and land degradation are environmental issues that should not be overlooked by Cambodian policy makers. However, as we see it, the future management of fish and forest resources is likely to have the most significant implications for Cambodia's short-term and long-term development. For this reason, the following section will concentrate on these two important resource assets.

6. Resource Depletion Issues

6.1. Forestry Issues

Between 1988 and 2004, Cambodia's total forest area was reduced from 11.9 million to 8.4 million hectares. This amounts to a 30% reduction in forests over the study period. For management purposes, Cambodia's forests are classified as either 'protected' forest or forest that exists within management concession zones. The protected areas cover about 18% of the total land area (3.3 million hectares) while the rest is reserved for concession areas (Ministry of Environment, 1998). At current rates of deforestation, the forest area in Cambodia will effectively be reduced to protected areas only by 2030 and, if the situation persists, desertification is likely to occur in less than 5 decades.

Exactly what are the major factors behind the rapid rate of deforestation in Cambodia? In the first instance, there is a strong and increasing demand for fuelwood. Fuelwood is currently the primary source of energy for most Cambodians. As the situation presently stands, Cambodians are permitted to collect fuelwood from any forest area, including protected areas. The fuelwood extracted from forest areas is estimated at 0.3 cubic meters per capita per year (Ministry of Environment, 1998). This consumption rate exceeds the regeneration rate of Cambodia's forests. Thus, commercial logging aside, the local demand for fuelwood is itself contributing to the depletion of Cambodia's forest resources.

The impact of fuelwood demand on Cambodia's forests cannot be underestimated. In 2005, for example, Cambodians were expected to consume up to 19,083 hectares of forest area for fuelwood while forest regeneration was estimated to be 12,577 hectares. This constitutes a forest deficit of 6,506 hectares or a rate of fuelwood demand that exceeds the rate of forest regeneration by 52%.

Secondly, forest stocks have been extensively reduced by commercial logging. Excessive logging has occurred largely as a consequence of the Cambodian Government's desire to generate royalties from commercial logging in concession forest areas for economic rehabilitation and development purposes. On a positive note, the number of concession forest areas has fallen from 33 in 1997 (covering 7 million hectares) to 15 in 2002 (covering 2.5 million hectares). Whilst this has helped to reduce the annual harvest of forests by around 25% since the mid-1990s, it has not been anywhere sufficient to prevent continuing high rates of deforestation.

Clearly, to prevent further deforestation in Cambodia, it will be necessary to deal with the problem of excessive commercial logging, undoubtedly the main offending area. Continued economic development in Cambodia should help alleviate the fuelwood issue since greater access to electricity should follow. Of course, any natural decline in fuelwood demand will depend very much upon how equitably the gains from further rises in real GDP are distributed among the Cambodian population. Hence, a more equitable distribution of income should be a top policy priority for the Cambodian Government.

Returning to the commercial sector, policy makers must ensure that timber is harvested in concession areas in line with sustainable harvesting practices. This means restricting the harvest to a rate that does not exceed the natural regeneration rate of the forests. Access to forests should be contingent upon logging companies meeting such a condition and leases should be revoked if the condition is violated.

To avoid 'hit-and-run' harvesting, logging companies should also be contracted over a lengthy period (i.e., long-term rather than short-term leases). Short-term leases provide a distorted incentive for logging companies to rapidly deplete the resource stock, whereas long-term leases tied to the quantity of timber harvested rather than the area available for harvest guarantee 'resource security'. Long-term leases tied to the quantity of timber harvested also provide two additional benefits. Being quota-based, they allow the relevant government authority to better control the overall quantity of timber harvested by all logging companies. Secondly, the amount paid for a lease equates to a harvest tax which would provide logging companies with an incentive to optimise the harvesting process. It would also encourage the use of quality harvesting techniques that could be provided by expert government advice (Kahn, 2005).

6.2. Fisheries Issues

Based on the methodology employed in this chapter, Cambodian fish stocks have declined over the study period from 8.0 million tonnes to 6.7 million tonnes. The annual harvest has grown rapidly from 86,800 tonnes in 1988 to 325,500 tonnes in 2004. It should be noted that the 2004 harvest was much lower than the peak harvest in 2003 of 425,642 tonnes.[1] The amount of fish harvested from 1988 to 2000 rose quite steadily at an average annual rate of 3.9% but increased dramatically in both 2001 (43.0%) and 2002 (118.2%).

Broadly speaking, there are three management strategies employed in Cambodia in relation to fish resources. The first involves seasonal limits and restrictions on fishing equipment; the second involves the existence of fish sanctuaries in protected areas; and the third involves the establishment of fishing lots (areas designated for commercial fishing only). Fishing lots are granted in such a way as to ensure the access to fish stocks is shared amongst the general community and fishing lot owners. However, while lot owners are permitted to catch fish outside their fishing lots, the general community is prohibited from accessing fishing lots under commercial ownership.

Under current government policy, fishing rights are subject to a public bidding process every two years. In many instances, local communities have difficulty raising the financial capital necessary to participate in the bidding process. Their access to fish stocks is consequently limited, which again raises the equity issue.

Since the fishing lots are a form of commercial property, commercial operators have an incentive to bulk harvest in order to maximise returns from their initial investment. This gives the fishing lot owners the opportunity to engage in excessive harvesting despite the restrictions placed upon them in terms of fishing equipment and closed harvesting seasons. In all, the establishment of fish lots has brought about a massive increase in harvest rates which is clearly unsustainable.

The Cambodia Government, which is the ultimate owner and manager of all fish stocks under the Cambodian Constitution, has recently expressed a strong commitment towards the sustainable management of all fish stocks. It envisages enforcing such a policy by 2010. However, the details of such a policy are yet to be released.

We believe a sustainable management plan will only be successful if it follows a few very basic principles. Firstly, it will need to ensure that access to fish stocks is equitably shared between local communities and commercial operators. This could be achieved by: (a) freely distributing a certain percentage of all fishing rights or permits to local communities via local government agencies; (b) restricting commercial operators to fishing lots; and (c) making all freely distributed permits non-tradeable. Only permits sold through auction would be tradeable. Secondly, fishing rights must be quantity based — that is, act as quotas — in order to restrict catch numbers to sustainable levels. Thirdly, large financial penalties must be imposed on transgressors since only then will permit holders face a disincentive large enough to deter over-fishing.

As an added bonus, the fee paid to acquire fishing rights at auction can serve as a tax to bring the private cost of fishing into line with the full social cost. Because fisheries possess two features typically characteristic of open access resources (see Khan, 2005), there is invariably a disparity between the private cost of fishing and the cost incurred by society as a whole. The fee paid to acquire fishing rights would bridge this gap and ensure an efficient number of operators in the industry. Competition between operators would also induce them to utilise the best available fishing technology, thereby further increasing efficiency. Meanwhile, the government could also use the revenue earned from the auctioning process to restock rivers and lakes or compensate affected parties.

7. Concluding Remarks

Although Cambodia's per capita GDP has increased moderately in recent years, the same cannot be said for Cambodia's per capita Hicksian income. Indeed, this study indicates that per capita GDP was stimulated to a very large extent by the liquidation of Cambodia's natural capital — in particular, its forest and fishery assets. Having said this, Cambodia's per capita Hicksian income was slightly higher in 2004 than it was in 1988 and increased over the last two years of the study period.

Does this mean that Cambodia is on a sustainable economic pathway? No, it simply indicates that Cambodia's per capita income would have risen slightly over the study period had it taken the necessary steps to operate in accordance with the strong sustainability requirement to keep natural capital intact. But it clearly did not do this. It is therefore incumbent upon Cambodian policy makers to limit the rate of resource extraction to rates consistent with their regenerative capacities and, where possible, invest in natural capital as well as human-made capital assets. Although the Cambodian Government has expressed a desire to utilise the nation's natural resources sustainably by 2010, it has yet to indicate how it

[1] It is too early to ascertain whether the decline in the 2004 harvest was due to reduced effort or plummeting fish stocks.

will achieve this. We believe the policies implemented will need to adhere to some basic principles, some of which we have sketched in this chapter.

If there is one positive aspect to emerge from this study, it is the message that a transition to a sustainable economic pathway need not, at least for the time being, result in a radical decline in Cambodia’s per capita income. Cambodian policy makers should therefore seize the opportunity while it exists. After all, there is no telling when it might cease or how costly the transition could eventually become should appropriate action not be immediately taken.

References

Asheim, G. (1994), ‘Net national product as an indicator of sustainability?’, *Scandinavian Journal of Economics*, **96**: 257-265.

Atkinson, G. and Gundimeda, H. (2005), *Accounting for India's forest wealth*, (online), Available: http://www.sciencedirect.com/science/article/B6VDY-4J4B9CJ-1/2/f67d94c6e244d22414a647f416423c3f (4 July 2006).

Barbier, E. (1991), ‘Environmental Degradation of the Third World’, in D. Pearce (ed.), *Blueprint 2: Greening the World Economy*, London: Earthscan.

Cairns, R. D. (2000), ‘Sustainability accounting and green accounting’, *Environment and Development Economics*, **5**(1): 49-54.

Chandler, D. (1983), *A History of Cambodia*, Boulder: Westview Press.

Clarke, M. and Lawn, P. (2005), ‘Measuring Victoria’s Genuine Progress: A Genuine Progress Indicator (GPI) for Victoria’, *Economic Papers*, **24**(4): 368-389.

Daly, H. (1996), *Beyond Growth*, Boston: Beacon Press.

Daly, H. and Cobb, J. (1989), *For the Common Good*, Boston: Beacon Press.

Dasgupta, P. and Mäler, K. (2000), ‘Net national product, wealth, and social well-being’, *Environment and Development Economics*, **5**(1): 69-93.

Deng, X. (2006), ‘Economic costs of motor vehicle emissions in China: A case study’, *Transportation Research Part D: Transport and Environment*, **11**(3): 216-226.

Department of Fisheries (2001a), *Natural Resources Evaluation*, (online), Available: http://www.maff.gov.kh/pdf/NaturalResorcesEvaluation.PDF (15 June 2006).

Department of Fisheries (2001b), *Inland Fisheries Review*, Phnom Penh: Ministry of Agriculture, Forestry and Fisheries.

El Serafy, S. (1989), ‘The proper calculation of income from depletable natural resources’, in Y. Ahmad, S. El Serafy, and E. Lutz, E. (eds), *Environmental Accounting for Sustainable Development*, World Bank, Washington DC, pp. 10-18.

El Serafy, S. (1996), ‘Weak and strong sustainability: natural resources and national accounting - Part I’, *Environmental Taxation and Accounting*, **1**(1): 27-48.

Foy, G. (1991), ‘Accounting for non-renewable natural resources in Louisiana's gross state product’, *Ecological Economics*, **3**(1): 25-41.

Georgescu-Roegen, N. (1979), ‘Comments on the papers by Daly and Stiglitz’, in V. K. Smith (ed.), *Scarcity and Growth Reconsidered*, Baltimore: John Hopkins University Press.

Gerlagh, R., Dellink, R., Hofkes, M. and Verbruggen, H. (2002), ‘A measure of sustainable national income for the Netherlands’, *Ecological Economics*, **41**(1): 157-174.

Herz, M. (1958), *A Short History of Cambodia: From the Days of Angkor to the Present*, London: Atlantic Book Publishing Company.

Hicks, J. (1946), *Value and Capital*, Second Edition, London: Clarendon.

Jackson, K. (1989), 'The intellectual origins of the Khmer Rouge', in K. Jackson (ed.), *Cambodia 1975-1978: Rendezvous with Death*, New Jersey: Princeton University Press.

Kahn, J. (2005), *The Economic Approach to Environmental Natural Resources*, Ohio: Thomson.

Kan, H. and Chen, B. (2004), 'Particulate air pollution in urban areas of Shanghai, China: health-based economic assessment', *Science of the Total Environment*, **322**(1-3): 71-79.

Kellenberge, J. (1996), *Accounting for Natural Resources in Ecuador: Contrasting Methodologies, Conflicting Results*, Washington DC: World Bank.

Khor, M. (undated), *Land Degradation Cause $10 Billion Loss to South Asia Annually*, (online), Available: http://www.twnside.org.sg/title/land-ch.htm (7 August 2006).

Khy, T. (2005), 'Fishery management in Cambodia: A case study in Anlong Rang and Kompong Tralach fishing community', *Economic Institute of Cambodia*, **1**(6): 13-15.

Kiernan, B. (2002), *The Pol Pot Regime*, Second Edition, New Haven: Yale University Press.

Lawn, P. (1998), 'In defence of the strong sustainability approach to national income accounting', *Environmental Taxation and Accounting*, **3**(1): 29-47.

Lawn, P. (2001), *Toward Sustainable Development: An Ecological Economics Approach*, Boca Raton: Lewis Publishers.

Lawn, P. (2003), 'How important is natural capital in terms of sustaining real output? revisiting the natural capital/human-made capital substitutability debate', *International Journal of Global Environmental Issues*, **3**(4): 418-435.

Lawn, P. (2007), *Frontier Issues in Ecological Economics*, Cheltenham, UK: Edward Elgar.

Leipert, C. (1989), 'Social costs of the economic process and national account: The example of defensive expenditure', *Journal of Interdisciplinary Economics*, **3**(2): 27-46.

Lvovshy, K. (1998), *Economic Cost of Air Pollution with Special Reference to India*, (online), Available: http://lnweb18.worldbank.org/SAR/sa.nsf/Attachments/ EconomicCost/$File/ Economic+costs+of+air+pollution+KL.pdf. (30 August 2006).

Mäler, K. (1991), 'National accounts and environmental resources', *Environmental and Resource Economics*, **1**(1): 1-15.

Meadows, D. H., Meadows, D. L., Randers, J., and Behrens, W. III., Eds. (1972), *The Limits to Growth*, New York: Universe Books.

Ministry of Environment (1998), *National Environment Plan 1998-2002*, Phnom Penh, Cambodia.

Nordhaus, W. and Tobin, J. (1972), 'Is economic growth obsolete?', in The National Bureau of Economic Research (ed.), *Economic Growth*, New York: Columbia University Press.

Osborne, M. (1973), *Politics and Power in Cambodia*, Camberwell: Longman.

Pearce, D. (1993), *Blueprint 3: Measuring Sustainable Development*, Earthscan: London.

Ponchaud, F. (1977), *Cambodia: Year Zero*, New York: Holt, Rinehart and Winston.

Repetto, R., Magrath, W., Wells, M., Beer, C. and Rossini, F. (1989), *Wasting Assets: National Resources in the National Income Accounts*, Washington DC: World Resources Institute.

SBS (2002), *World Guide: The Complete Fact File on Every Country*, Tenth Edition, South Yarra: Hardie Grant Books.

Skanberg, K. (2001), *Constructing a Partially Environmentally Adjusted Net Domestic Product for Sweden, 1993 And 1997, Working Paper* **76**, Konjunktur Institute.

Transparency International (various), www.transparency.org

Tully, J. (2005), *A Short History of Cambodia: From Empire to Survival*, Crows Nest: Allen & Unwin.

United Nations (UN) (1993), *Integrated Environmental and Economic Accounting (interim version)*, New York: World Bank.

United Nations Statistics Division (UNSD) (2005), *Cambodia*, (online), Available: http://unstats.un.org/unsd/snaama/resultsCountry.asp?Country=116&Year=0&SLevel=99&Disp=Million (18 August 2005).

Van Tongeren, J., Schweinfest, S., Lutz, E., Gomez Luna, M. and Martin, G. (1993), 'Integrated economic and environmental accounting: a case study for Mexico', in E. Lutz (ed.), *Toward Improved Accounting for the Environment*, Washington DC: World Bank.

Vickery, M. (1984), *Cambodia 1975-1982*, Sydney: Allen & Unwin.

Weitzman, M. (1976), 'On the welfare significance of national product in a dynamic economy', *Quarterly Journal of Economics*, **90**(1): 156-162.

World Bank (1994), *Cambodia: From Reconstruction to Rehabilitation*, Report No. 12667, Phnom Penh, Cambodia.

World Bank (2003), *Cambodia Environment Monitor*, Phnom Penh: World Bank.

World Bank (2006), *2006 World Development Indicators*, Washington, DC: World Bank.

World Commission on Environment and Development (WCED) (1987), *Our Common Future*, Oxford: Oxford University Press.

World Resources Institute (WRI) (2003), *EarthTrends: Economic Indicators, Cambodia*, (online), Available: http://earthtrends.wri.org (3 April 2006).

World Resources Institute (WRI) (2006), *EarthTrends: Climate and Atmosphere, CO2: Total Emissions, Cambodia*, (online), Available: http://earthtrends.wri.org (3 August 2006).

Yang, S. (undated), *Sustainable Agricultural Development Strategies for the Least Developed Countries of the Asian and the Pacific Region, Cambodia*, (online), Available: http://www.unescap.org/rural/doc/sads/cambodia.PDF (12 May 2006).

Young, M. (1990), 'National resource accounting', in M. Common and S. Dovers (eds), *Moving Toward Global Sustainability: Policies and Implication for Australia*, Canberra: Australian National University Press.

In: Emerging Topics in Macroeconomics
Editor: Richard O. Bailly, pp. 303-325
ISBN: 978-1-60692-093-0

Chapter 11

ANALYSIS OF FINANCIAL INSTABILITY BY MEANS OF DECISION TREES AND LISTS

***Z. Díaz*[1], *A. Sanchis*[2] *and M.J. Segovia*[1]**
[1] Department of Financial Economics and Accounting I, Universidad Complutense de Madrid, Spain.
[2] Banco de España. Madrid, Spain.

Abstract

This paper tries to further investigate the factors behind a financial crisis using a large sample of countries in the period 1981 to 1999.

Most methods applied in other researches to analyse financial instability are techniques of statistical nature and the variables employed in these models do not usually satisfy statistical assumptions, what complicates the analysis.

In order to avoid these inconveniences, we have approached this problem by applying two algorithms coming from the Machine Learning field, the C4.5 and PART algorithms. By means of theses algorithms we will try to analyse the role of a set of macroeconomic and financial variables, both quantitative and qualitative, in explaining banking crises.

Results indicate a high performance of the machine learning techniques, what shows that these methods can be a useful tool to evaluate financial instability and a competitive alternative or complement to existing models related with this kind of problems.

Keywords: financial stability, Machine Learning, C4.5, PART.

1. Introduction

The financial system plays a crucial role in economic development as responsible for the allocation of resources over time and among different alternatives of investment by pricing the postposition of consumption (free risk rate) and pricing the risk (risk premium). A correct functioning of financial system allows economies to reach higher levels of real growth as well as more stable macroeconomic conditions. (Stiglitz and Furman, 1998). Therefore, preserving financial stability is one of the main goals for policy makers since the beginning of the monetary systems.

Currently developed countries are embarked in a number of initiatives in the financial field which aim is to reduce the risk of systemic crisis. The globalization of the financial system has increased the risk of contagion between institutions. These developments are forcing financial regulators to revise the regulatory framework. This framework has recently been revised in the banking sector through the so called Basel II project and a similar revision is currently ongoing in the insurance sector (Solvency II).

The recent financial turmoil has further stressed how important it is to guarantee that financial markets are able to correctly price the risk (the US subprime market has triggered the crisis although the repricing has been more general afterwards), how important is globalization in the financial markets (e.g. European banks were exposed to US subprime market in a significant amount) and how close is the relationship between liquidity and solvency and therefore between monetary policy and prudential supervision. In sum, the “originate to distribute” business model that has characterized the last booming period is being questioned. A large number of policy initiatives are at this time under way to draw lessons from this crisis. The lessons cover among others further enhance financial institutions transparency; reassess the liquidity management of the institutions, to review the collateral market, and to improve the exchange of information and cooperation between monetary and prudential authorities.

The especial role that banks play in the financial system and their specificities as money issuers explain why a great number of financial crises had got the banking sector as protagonist, and that is why we focused on banking crises as a financial instability outcome. So the purpose of this paper is to analyze the role of a set of macroeconomic and financial variables suggested by the theory and highlighted by empirical studies in explaining banking crises. Our panel includes all market economies for which data were available in the period 1981-99.

The determination of the future financial stability of a country can be understood as a classification problem: given a set of attributes associated to a country (both macroeconomic and financial variables) we try to assign the country to a predefined class (crisis or financial stability).

Most of the previous empirical studies that deal with this issue are based on classical statistical techniques such as discriminant, logit or probit analysis (Demirgüç-Kant and Detragiache, 1997, 1998, 2000; Domaç and Martínez Peria, 2000; Eichengreen, 1997, 2000; Eichengreen and Arteta, 2000; Eichengreen and Rose, 1998; Frydl, 1999; García Herrero and Del Río, 2003; Glick and Hutchinson, 1999; Gourinchas *et al.*, 1999; Hardy and Pazarbasioglu, 1998; Kaminsky and Reinhart, 1999; Mendis, 1998; Rossi, 1999).

However, although the obtained results have been satisfactory, all these techniques present the drawback that they make some assumptions about the model or the data distribution that are not usually satisfied, and, given the complexity of these techniques, an inexpert user can find difficult to extract conclusions of their results.

In an attempt to overcome the limitations of these methodologies it is recently suggested the use in the economic field of techniques coming from the Artificial Intelligence field, because they are non-parametric, or distribution free, methods. To tackle our problem, the techniques framed in the Machine Learning – the Artificial Intelligence area that develops algorithms which are able "to learn" a model from a set of examples – are of great utility.

For all these reasons the goal of this paper is to show, by means of an empirical study, the efficiency of two Machine Learning techniques, C4.5 and PART algorithms, to detect possible financial crises, and we understand this problem as a classification one with two predefined classes, crisis or financial stability, and using as attributes a set of financial and macroeconomic variables. The scarcity of prior research for prediction of bank failure by means of similar algorithms focused on financial ratios and individual crises (Bonsón Ponte *et al*., 1996; Martín Zamora, 1999; Tam and Kiang, 1992).

The rest of the paper is structured as follows: section 2 introduces the main concepts of the proposed methods. In section 3 we describe the data and input variables. In section 4 we present the obtained results and, finally, section 5 closes the paper with some concluding remarks.

2. C4.5 and PART Algorithms: Main Concepts

2.1. General Considerations

As we have previously mentioned, Machine Learning algorithms are a set of techniques that automatically build models describing the structure at the heart of a set of data, that is, they induce a model or output from a given set of observations or input. Such models have two important applications. First, if they accurately represent the structure underlying the data, they can be used to predict properties of future data points. Second, if they summarize the essential information in human-readable form, people can use them to analyze the domain from which the data originates (Frank, 2000).

These two applications are not mutually exclusive. To be useful for analysis, a model must be an accurate representation of the domain, and that makes it useful for prediction as well. However, the reverse is not necessarily true: some models are designed exclusively for prediction and do not lend themselves naturally to analysis, as it happens in the case of the popular Artificial Neural Networks or the more recent Support Vector Machines. In many applications this “black box” approach is a serious drawback because users cannot determine how a prediction is derived and match this information with their knowledge of the domain.

Two of the most fruitful and widely used approaches in Machine Learning are the Decision Trees and the Classification Rules, that in spite of their differences of formal character keep a narrow relationship that makes they can be considered as different variants of a common methodology. Decision trees and rules are potentially powerful predictors that embody an explicit representation of all the knowledge that has been induced from the dataset. Moreover, compared to other sophisticated models, they can be generated very quickly. Given a decision tree or a set of rules, a user can determine manually how a particular prediction is derived, and which attributes are relevant in the derivation. This makes them extremely useful tools when both prediction and explanation are important.

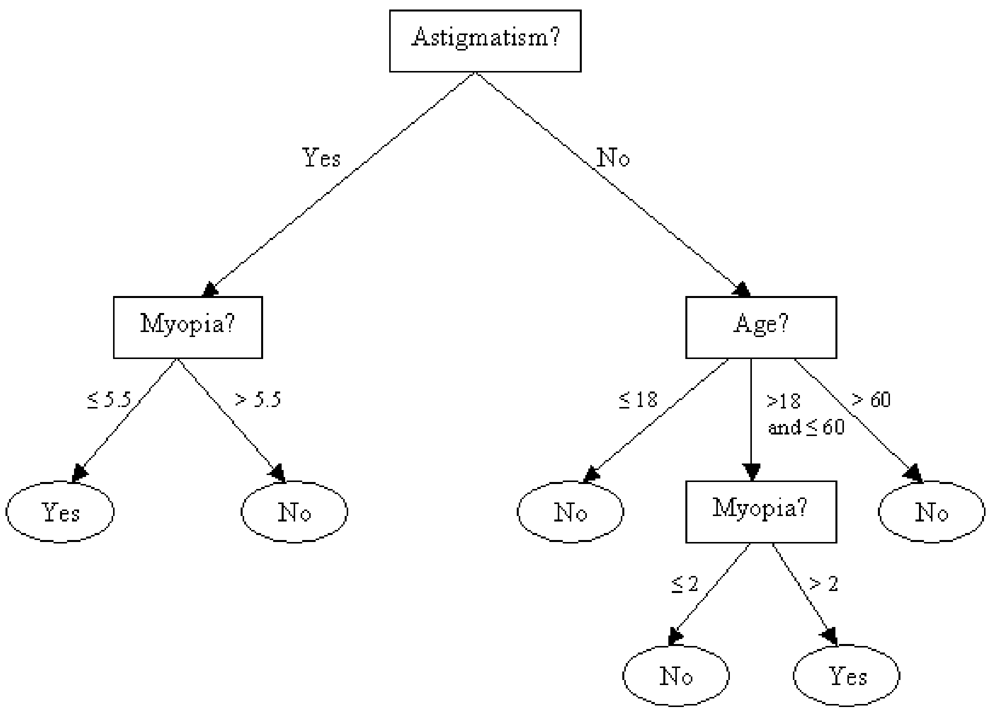
In this paper we will use in fact two of these techniques, the well-known C4.5 decision tree learner, and PART, a learning algorithm for decision lists.

2.2. C4.5 Algorithm

Decision trees are a way of representing the underlying regularity in the data like a set of exhaustive and mutually exclusive conditions which are organized in an arborescent hierarchical structure which is composed by internal and external nodes connected by branches. An internal node contains a test, it is a unit which evaluates a decision function to determine which node will be visited next. In contrast, an external node, which is frequently called leaf or terminal node, doesn't have any son and it is associated with a label or a value which characterizes to the data that are propagated to it.

In general, a decision tree is used in the following way: to derive a prediction, an instance is filtered down the tree, starting from the root node, until it reaches a leaf – in this paper, an instance will be a country described by a set of macroeconomic and financial variables –. At each node one of the instance's attributes is tested, and the instance is propagated to the branch that corresponds to the outcome of the test. The prediction is the class label that is attached to the leaf.

For example, the next figure shows how the recommendation of refractive surgery to myopic patients in a hospital could be determined by means of a decision tree, making the questions of the internal nodes and following the answers until a leaf is reached.



As for the way of generating a tree, standard learning algorithms for decision trees generate a tree structure by splitting the training data into smaller and smaller subsets in a recursive top-down fashion. Starting with all the training data at the root node, at each node they choose a split and divide the training data into subsets accordingly. They proceed recursively by partitioning each of the subsets further. Splitting continues until all subsets are "pure", or until their purity cannot be increased any further. A subset is pure if it contains instances of only one class. The aim is to achieve this using as few splits as possible so that the resulting decision tree is small and the number of instances supporting each subset is large. To this end, various split selection criteria have been designed, and at each node, the learning algorithm selects the split that corresponds to the best value for the splitting criterion.

Some of the most outstanding split selection criteria are the Gini index, which is employed in the CART system – Classification and Regression Trees (Breiman *et al*., 1984) – , and the "information gain" or the "gain ratio", which are used by C4.5. They all provide ways of measuring the purity of a split. C4.5 is the decision tree program most popular and widely used in practice to date (Quinlan, 1993). To carry out the partitions, C4.5 is based on some concepts from Information Theory: The information provided by a message or a random variable *X* is inversely proportional to its probability. In Communication Engineering or in Statistics this quantity is usually measured in *bits* (Krippendorff, 1986; Reza, 1994; Ziemer and Tranter, 2002), obtained through the relation: $\log_2 \frac{1}{p_x}$. The average of this relation for all the possible cases of the random variable *X* is called *entropy* of *X*: $H(X)=\sum_x p(x)\log_2 \frac{1}{p(x)}$. The entropy is a measure of the randomness or uncertainty of *X* or a measure of the average amount of information that is supplied by the knowledge of *X*.

In the same way, we can define the *joint entropy* of two random variables *X* and *Y*: $H(X,Y)=\sum_{x,y} p(x,y)\log_2 \frac{1}{p(x,y)}$, which represents the average amount of information supplied by the knowledge of *X* and *Y*. The *conditional entropy* of *X* given the variable *Y*, $H(X/Y)$, is defined as $H(X/Y)=\sum_{x,y} p(x,y)\log_2 \frac{1}{p(x/y)}$, and this relation is a measure of the uncertainty of *X* when we know the variable *Y*. This is the amount of information necessary to know completely *X* when we know the information provided by *Y*-variable. Naturally, $H(X/Y)\le H(X)$, because if *Y*-variable is known we have more information that can help us to reduce the uncertainty about *X*-variable. This reduction in the uncertainty is called *mutual information* between *X* and *Y*: $I(X\ ;Y)=H(X)-H(X/Y)$, which is the information provided by one of the variables about the other one. It is always verified that $I(X\ ;Y)=I(Y\ ;X)$, therefore the amount of information that each variable provides about the other one is the same.

We can consider that *X* is a random variable that represents the category to which an object belongs. On the other hand, $Y_i,\ i=1,2,\ldots,n$, represents the set of attributes that describe the objects we want to classify.

C4.5 algorithm chooses to make each partition the Y_i-variable that provides the maximum information about *X*-variable, that is, it maximizes the following relation called *gain ratio*: $\frac{I(X\ ;Y_i)}{H(Y_i)}$. This ratio represents the percentage of information provided by Y_i that is useful in order to characterize *X*.

Note that $I(X\ ;Y_i)$ should be large enough to prevent that an attribute could be only chosen because it has a low value for entropy, what would increase the *gain ratio*.

C4.5 works with both continuous and discrete attributes and incorporates several additional features that turn it into a very powerful and flexible technique, such as, for example, its method for handling with missing values. Very briefly, such a method is the following one: once a splitting attribute has been chosen, training cases with unknown values of this attribute cannot be associated with a particular outcome of the test, so a weighting scheme is used to allow recursive application of the decision tree formation procedure on each of the daughter nodes. Instances for which the relevant attribute value is missing are notionally split into pieces, one piece for each branch, in the same proportion as the known instances go down the various branches, so the number of cases that are propagated to the nodes and leaves of the tree could be a fractional value. A similar approach is taken when the decision tree is used to classify a new case. If a decision node is encountered at which the relevant attribute value is unknown, so that the outcome of the test cannot be determined, the system explores all possible outcomes and combines the resulting classifications arithmetically. Since there can now be multiple paths from the root of a tree or subtree to the leaves, a “classification” is a class distribution rather than a single class. When the total class distribution for the case has been established in this way, the class with the highest probability is assigned as the predicted class.

A common problem for most of machine learning techniques is that models they generate can be quite adapted to the training dataset. Consequently, the model developed will be very specific and if we want to classify new objects, the model will not provide good results, especially if the training set has noise. In this last case, the model would be influenced by errors (noise) which would lead to a lack of generalization. This problem is known as *overfitting*.

The usual way of limiting this problem in the context of decision trees and sets of rules consists of deleting some conditions of the tree branches or of the rules. In the case of decision trees, this procedure can be considered as a *pruning* process.

C4.5 incorporates a *post-pruning* method for an original fitted tree. It consists of simplifying the tree by discarding one subtree (or more) and replacing it with a leaf or with its most frequently used branch, provided this replacement lead to a lower predicted error rate. Obviously the probability of error in a node of the tree cannot be determined exactly, and the error rate on the training set from which the tree was built does not provide a suitable estimate. To estimate the error rate, C4.5 works in the following way: suppose that there is a leaf that covers *N* objects and misclassifies *E* of them. This could be considered as a binomial distribution in which the experiment is repeated *N* times obtaining *E* errors. From this issue, the probability of error p_e is estimated, and it will be taken as the aforementioned predicted error rate. So it is necessary to estimate a confidence interval for the error probability of the binomial distribution. The upper limit of this interval will be p_e (note that this is a pessimistic estimate).

Then, in the case of a leaf that covers *N* objects, the number of predicted errors will be $N \times p_e$. Similarly, the number of predicted errors associated with a subtree will be just the sum of the predicted errors of its branches, and the number of predicted errors associated with a branch will be the sum of the predicted errors of its leaves. Therefore, a subtree will be replaced by a leaf or a branch, that is, the subtree will be pruned, when the number of predicted errors for the last ones is lower than that for the subtree.

On the other hand, though the decision trees represent the knowledge in a very simple way, their intelligibility decrease as their size increase. Even though the pruned trees are more compact than the originals, when the problem is very complex the tree is very large and consequently difficult to understand since each node has a specific context established by the outcomes of tests at antecedent nodes. Collections of *if-then* rules are simpler and easier to understand than decision trees and that is why classification rules are a very popular alternative to decision trees.

The decision tree induction algorithms are based on a *divide-and-conquer* approach to the problem of learning from a set of instances: they work top-down (and that is why this approach is often called *Top-Down Induction of Decision Trees* or TDIDT), seeking at each stage an attribute to split on that best separates the classes, and then recursively processing the subproblems that result from the split. This strategy generates a decision tree, which can be converted into a set of rules in a trivial way: one rule is generated for each leaf. The antecedent of the rule includes a condition for every node on the path from the root to that leaf, and the consequent of the rule is the class assigned by the leaf. However, in general rules that are read directly off a decision tree are far more complex than necessary, because the antecedents of individual rules may contain irrelevant conditions which could be deleted without affecting its accuracy. So if it is to produce effective rules, the conversion is not trivial. An important feature of C4.5 is its ability to convert a decision tree into a set of classification rules that is usually about as accurate as a pruned tree, but more easily understandable.

2.3. PART Algorithm

An alternative approach to construct rules is to take each class in turn and seek a way of covering all instances in it, at the same time excluding instances not in the class. This is called a *covering* approach because at each stage you identify a rule that "covers" some of the instances. The partitions of a decision tree take all classes into account, trying to maximize the purity of the split, whereas these rule-generating methods concentrate on one class at a time, disregarding what happens to the other classes. These techniques are based on a *separate-and-conquer* strategy, because they identify a rule that covers instances in the class (and excludes ones not in the class), separate out the covered instances because they are already taken care of by the rule, and continue the process on those that are left.

Covering algorithms operate by adding tests to the rule that is under construction, always striving to create a rule with maximum accuracy. In contrast, divide-and-conquer algorithms operate by adding tests to the tree that is under construction, always striving to maximize the separation between the classes.

The conditions in a decision tree which has been generated by a divide-and-conquer algorithm are exhaustive and mutually exclusive, independently they are represented like a tree or like a set of rules, whereas with rule sets generated by separate-and-conquer algorithms several rules could be applicable to the same instance. Moreover, an example may receive multiple classifications, that is, it may be accepted by rules that apply to different classes. This can be solved by imposing an ordering on the rules (obtaining decision lists) or by weighting and adding the diverse predictions.

Decision lists can be considered like extended IF – THEN rules in the way:

if… then … ; *else*:
  if… then … ; *else*:
    if… then … ; *else*:

.
.
.

The ordered structure of decision lists prevents that rule sets for different classes can "overlap" in instance space, a problem which the inefficiency of some rule induction algorithms is usually attributed to (Berzal Galiano, 2002). With a decision list, during classification the rules are evaluated sequentially and an instance is classified using the first rule that fires. The addition of a default rule (usually the class with the most training examples) at the end serves to ensure that any instance receives a classification. Ordering the rules in this way assures that one and only one rule applies to any given instance.

PART, a rule-learning algorithm based on partial decision trees which was developed by Frank and Witten (1998), represents a hybrid alternative approach to decision list induction because it combines the divide-and-conquer strategy for decision tree learning with the separate-and-conquer one for rule learning. It adopts the separate-and-conquer strategy in that it builds a rule, removes the instances it covers, and continues creating rules recursively for the remaining instances until none are left. However, it differs from the standard approach in the way that each rule is created. In essence, to make a single rule, a pruned decision tree is built for the current set of instances, the leaf with the largest coverage is made into a rule, and the tree is discarded.

The prospect of repeatedly building decision trees only to discard most of them is not as bizarre as it first seems. Using a pruned tree to obtain a rule instead of building it incrementally by adding conjunctions one at a time avoids a tendency to overprune that is a characteristic problem of the basic separate-and-conquer rule learner. Using the separate-and-conquer methodology in conjunction with decision trees adds flexibility and speed. It is indeed wasteful to build a full decision tree just to obtain a single rule, but the process can be accelerated significantly without sacrificing the mentioned advantages in the way implemented in PART: the key idea is to build a partial decision tree instead of a fully explored one. A partial decision tree is an ordinary decision tree that contains branches to undefined subtrees. To generate such a tree, the construction and pruning operations are integrated in order to find a "stable" subtree that can be simplified no further. Once this subtree has been found, tree-building ceases and a single rule is read off.

The process is as follows: first a test is chosen and the instance set is divided into subsets accordingly. The choice is made in the same way as C4.5. Then the subsets are expanded in increasing order of their entropy, starting with the smallest. The reason for this is that the later subsets will most likely not end up being expanded, and a subset with low entropy is more likely to result in a small subtree and therefore produce a more general rule. This proceeds recursively until a subset is expanded into a leaf, and then continues further by backtracking. But as soon as an internal node appears which has all its children expanded into leaves, the algorithm checks whether that node is better replaced by a single leaf. This is just the "subtree replacement" operation of decision tree pruning, and PART makes the decision in exactly the same way as C4.5. If replacement is performed, the algorithm backtracks in the standard way, exploring siblings of the newly replaced node. However, if during backtracking a node is encountered all of whose children are not leaves – and this will happen as soon as a potential subtree replacement is not performed – then the remaining subsets are left unexplored and the corresponding subtrees are left undefined. Due to the recursive structure of the algorithm, this event automatically terminates tree generation (Frank and Witten, 1998).

Once a partial tree has been built, a single rule is extracted from it. Each leaf corresponds to a possible rule, and PART aims at the most general rule by choosing the leaf that covers the greatest number of instances. If the dataset contains missing values, PART deals with them in the same way as C4.5.

According to the experiments performed by the authors of PART, it produces very quickly rule sets as accurate as those generated by C4.5 and more accurate than other fast rule induction methods. However, its main advantage over other schemes is not performance but simplicity, and this is achieved by combining the top-down decision tree induction method with separate-and-conquer rule learning. These are the reasons for we use this algorithm in our paper.

3. Data and Variables Selection

The literature offers several definitions of financial instability - that for our purposes we identity with banking crisis as the most common example of financial instability given the especial role banks play in the financial system - (Bordo, 1986; Caprio and Klingebiel, 1997; Friedman and Schwartz, 1963; Gupta, 1996; Lindgren *et al*., 1996). However, any of these definitions completely solve the problem of how to summarize such description in one single quantitative indicator, or a set of them. Existing indicators, such as those mentioned by Lindgren *et al*. (1996), are not readably available for a large number of countries, or else there is the lack of comparable cross-country data to construct such indicator.

The empirical literature identifies banking crises as events, expressed through a binary variable, constructed with the help of cross-country surveys (Caprio and Klingebiel, 2003; Lindgreen *et al*., 1996). This will be our approach as well.

We will employ a sample of 79 countries in the period 1981-1999 (annual data). The dependent variable can be defined in this way: systemic and non-systemic banking crises dummy equals one during episodes identified as in Caprio and Klingebiel (2003). The independent variables included are dictated by the theory on the determinants of banking crisis.

There are risks common to all companies no matter the business (e.g. bad management). However, banks are exposed to specific risks given the especial nature of their business. The most characteristic risk face by banks is credit risk, which is closely link to cyclical output downturns, terms of trade deteriorations, declines in asset prices such as equity and real estate

(Caprio and Klingebiel, 1997; Gorton, 1988; Kaminsky and Reinhart, 1999; Lindgren *et al.*, 1996). Other typically banking risks are the interest rate risk and the liquidity risk. And depending on the specific circumstances also the currency risk can be a determinant factor in a financial crisis.

As independent variables we employ a set of macroeconomic and financial variables, both qualitative and quantitative. Among the macroeconomic variables we include: the real growth of GDP, the level of real GDP per capita, the inflation rate and the real interest rate to capture the external conditions that countries face. Among the financial variables we include: the credit growth, the banks liquidity, the foreign dependence of the banks and the previous crises as a measure of the degree of confidence of the depositors in the financial stability.

We also want to explore the role of monetary policy. The existing literature on monetary policy has concentrated on issues different than financial stability (mainly price stability but also output stabilization). The impact of the monetary policy design on financial stability is related to the very much debated question of the relation between price stability and financial stability. The economic literature is divided as to whether there are synergies or a trade-off between them (Cukierman, *et al.*, 1992; Mishkin, 1996). If synergies existed between the two objectives it would seem safe to argue that the same monetary policy design which helps achieve price stability also fosters financial stability (Issing, 2003; Padoa-Schioppa, 2002; Schwartz, 1995). However, if there were a trade-off, it would be much harder to establish an a-priori on the impact of price stability on financial stability. Among the variables related with monetary policy we include the monetary policy regime and the degree of independence of the central bank.

This debate has becoming more important in light of the recent turbulences. Although the trigger of the crisis was the US subprime market it was the sudden and dry of the money markets who contribute to the rapid spread of the crisis. US and European monetary policy authorities has confronted the difficult question on how to weight in their reaction functions the two objectives: on one hand the financial stability (and its potential short term effects in the real economy) and on the other the price stability (and its potential long term effects in the real economy).

We have selected the variables having into account the several factors usually highlighted by the literature. We provide a detailed list of variables and sources in the data appendix.

4. Empirical Results

The algorithms have been performed using the data mining package *WEKA* from the University of Waikato (Witten and Frank, 2005). The Weka's implementation of the C4.5 decision tree learner described in section 2 is called J4.8 algorithm (J4.8 actually implements a later and slightly improved version called C4.5 Revision 8, which was the last public version of this algorithm before C5.0, the commercial implementation, was released. We have not used the more recent commercial version because some aspects of its workings have not been described in the open literature).

In order to test the predictive accuracy of the models, we have split the set of original data to form the training set and the holdout sample to validate the obtained models, i.e., the test set (randomly generated). The training set consisted of 421 data from 79 countries in the period 1981-1997 (annual data) described by the variables explained in section 3, and assigned to a decision class (crisis or non-crisis). We have 293 cases for class "non-crisis" and 128 cases for class “crisis”. The test set consisted of 100 data described by the same variables in the period 1997-1999 (36 objects for class “crisis” and 64 objects for class “non-crisis”). Next, the output is shown:

```
=== Classifier model (full training set) ===

J48 pruned tree
------------------

PREVIOUS CRISES = 0
|   DOMESTICCREDITGROWTH <= 3.248933
|   |   Real.interest.rate <= 2.23: non-crisis (9.0/1.0)
|   |   Real.interest.rate > 2.23: crisis (7.0/1.0)
|   DOMESTICCREDITGROWTH > 3.248933: non-crisis (194.0/21.0)
PREVIOUS CRISES = 1
|   FOR_LIAB_REV <= 0.805912
|   |   DOMESTICCREDITGROWTH <= -1.116854: crisis (8.0)
|   |   DOMESTICCREDITGROWTH > -1.116854
|   |   |   GDPPERHEAD <= 3800
|   |   |   |   CBANKINDEP <= 0.45: crisis (8.0)
|   |   |   |   CBANKINDEP > 0.45
|   |   |   |   |   inflation <= 15.94: crisis (8.0/1.0)
|   |   |   |   |   inflation > 15.94: non-crisis (7.0/1.0)
|   |   |   GDPPERHEAD > 3800
|   |   |   |   GDP_GROWTH.ANNUAL... <= 1: crisis (18.0/7.0)
|   |   |   |   GDP_GROWTH.ANNUAL... > 1: non-crisis (48.0/8.0)
|   FOR_LIAB_REV > 0.805912: non-crisis (16.0)
PREVIOUS CRISES = 2
|   X.NETKFLOWS <= 5.919084
|   |   EXCHANGE = CB: non-crisis (12.0/3.0)
|   |   EXCHANGE = FF
|   |   |   X.NETKFLOWS <= 2.898993: crisis (7.0)
|   |   |   X.NETKFLOWS > 2.898993: non-crisis (6.0/2.0)
|   |   EXCHANGE = MF: crisis (17.0/7.0)
|   |   EXCHANGE = P
|   |   |   GDPPERHEAD <= 3167: crisis (6.0)
|   |   |   GDPPERHEAD > 3167: non-crisis (6.0/2.0)
|   X.NETKFLOWS > 5.919084: non-crisis (8.0)
PREVIOUS CRISES = 3: crisis (36.0/9.0)

Number of Leaves  :     18

Size of the tree :      31
```

```
=== Evaluation on training set ===
=== Summary ===

Correctly Classified Instances        358        85.0356 %
Incorrectly Classified Instances       63        14.9644 %
Total Number of Instances             421

=== Detailed Accuracy By Class ===

TP Rate   FP Rate   Precision  Recall  F-Measure   Class
  0.915     0.297      0.876    0.915     0.895     non-crisis
  0.703     0.085      0.783    0.703     0.741     crisis

=== Confusion Matrix ===

   a    b   <-- classified as
 268  25 |   a = non-crisis
  38   90 |   b = crisis


=== Evaluation on test set ===
=== Summary ===

Correctly Classified Instances         80          80 %
Incorrectly Classified Instances       20          20 %
Total Number of Instances             100

=== Detailed Accuracy By Class ===

TP Rate   FP Rate   Precision  Recall  F-Measure   Class
  0.797     0.194      0.879    0.797     0.836    non-crisis
  0.806     0.203      0.69     0.806     0.744     crisis

=== Confusion Matrix ===

  a   b   <-- classified as
 51 13 |  a = non-crisis
  7  29 |  b = crisis
```

At the beginning is a pruned decision tree in textual form which would be read in the following way:

- If previous crises = 0 and domestic credit growth is less than or equal to 3.25 and real interest rate is less than or equal to 2.23, then non-crisis.
- If previous crises = 0 and domestic credit growth is less than or equal to 3.25 and real interest rate is greater than 2.23, then crisis.
- If previous crises = 0 and domestic credit growth is greater than 3.25, then non-crisis.
- If…

and so on.

Every leaf of the tree is followed by a number *n* or *n*/*m*. The value of *n* is the number of cases in the sample that are mapped to this leaf, and *m* (if it appears) is the number of them that are classified incorrectly by the leaf, expressed as a decimal number because of the way the algorithm uses fractional instances to handle missing values.

Beneath the tree structure the number of leaves is printed; then the total number of nodes (*Size of the tree*).

The next part of the output shows the results obtained from the training data. This evaluation is not likely to be a good indicator of future performance. Because the classifier has been learned from the very same training data, any estimate of performance based on that data will be optimistic. Although it is not a reliable predictor of the true error rate on new data, it may still be useful, because it generally represents an upper bound to the model's performance on fresh data. In this case, 358 of 421 training instances – or 85% – are classified correctly. As well as the classification error, the evaluation module also outputs some statistics for each class. TP, FP, TN, and FN are the number of true positives, false positives, true negatives, and false negatives, respectively, and $\text{TP Rate} = \frac{\text{TP}}{\text{TP+FN}}$; $\text{FP Rate} = \frac{\text{FP}}{\text{FP+TN}}$; $\text{Precision} = \frac{\text{TP}}{\text{TP+FP}}$; Recall is the same as TP Rate (different terms are used in different domains), and finally, F-Measure is a weighted average between Precision and Recall:

$$\text{F-Measure} = \text{Precision} \times \frac{\text{TP+FP}}{\text{2TP+FP+FN}} + \text{Recall} \times \frac{\text{TP+FN}}{\text{2TP+FP+FN}} = \frac{\text{2TP}}{\text{2TP+FP+FN}}.$$

From the confusion matrix at the end we can see that 25 instances of class "non-crisis" have been assigned to class "crisis" and 38 of class "crisis" are assigned to class "non-crisis".

To predict the performance of the tree on new data, we need to assess its error rate on a dataset that played no part in the formation of the tree. So the last part of the output gives estimates of the tree's predictive performance that are obtained using the test set of 100 instances. As we can see, 80% of the cases are classified correctly, a quite satisfactory result.

Regarding the economical interpretation of the tree, the most important variable that has a role in the classification, given that it is implied in the classification of all units, is the variable "previous crisis". The variables that influence a country finally suffer a crisis or not are different if the country is not yet in a crisis (previous crisis=0), the country is at the beginning of a crisis (previous crisis=1), it is in a more mature stage of a crisis (previous crisis=2) or is in a really persistent one (previous crisis=3). Confidence is a crucial factor to maintain financial stability, and when a crisis emerges there is a general loss of confidence that is more difficult to recover when more persistent the crisis is. The lack of confidence in turn implies more sacrifice in real terms to come back to a stability scenario.

In fact our tree predicts that countries with a crisis longer than 3 years will be in crisis in the next period. This result of course cannot be read as deterministic but the small numbers of cases we have in our sample for this situation give not enough information to the system to be more precise.

It is interesting to notice that the most important variable for explaining that one country enters in a crisis is the domestic credit growth. There are arguments in the sense that an excessive domestic credit could pose some risks in terms on overheating the economies, but it seems that in general rates of growth above 3% are associated with countries that will not enter in crisis. In general an stable credit growth is associated with growing countries with a developed internal financial market that is a good feature in containing a financial crisis.

For the countries with ratios of domestic credit growth lower than 3%, therefore with an undeveloped financial system, the variable that makes the difference is the real interest rate. Real interest rates higher than 2.25% are associated with a crisis, while lower than 2.25% are compatible with a non crisis situation.

The explanation became more complex for countries that have just started a crisis. Here the second most important variable in explaining a crisis is the foreign liabilities ratio. Although a high foreign liabilities ratio implies a currency risk in case of a depreciation of the national currency given the lower value of the national assets compared with the value of the liabilities, in the short term the possibility to hold foreign deposits within the banking system can help banks, at least temporarily, in their efforts to maintain their deposit base (see García Herrero, 1997), therefore contributing to maintain financial stability.

In the countries with the ratio of foreign liabilities lower than 80% then is again the domestic credit growth the important variable. If this variable is not higher than that threshold then it enters into scene the wealthiest of the country. Richer countries with growing ratios above 1% will go out of the crisis while same countries but growing below this threshold will continue in crisis. An interesting observation is that situation in poorer countries depend on the independence of the central bank. Price stability, and therefore currency stability, is vital in maintaining a climate of confidence. The monetary policy regime is an important factor to maintain price stability. But once a situation of uncontrolled inflation has emerged then the credibility on the monetary policy authorities is an important factor in making easier to recover stability.With an independent central bank in a context of inflation above 15% the country will go out of the crisis. In a context of crisis a dependent central bank has not the credibility to control inflation and therefore to maintain the currency value at a reasonable cost. So the most probable is that deterioration continues. But independence only changes things when inflation is higher than 15%, therefore when the real benefits of price stability are enough to compensate the sacrifice in terms of real income that a restrictive policy implies.

For countries with a more mature crisis the first variable in explaining if a country can go out or not of a crisis is capital net flows. This means that if there is enough external capital, based on the confidence of the external investors on this economy, it plays an important role in overcoming the situation. If external factors are not enough (below 6%) then it is important the exchange regime. The currency regime could be use to recover the confidence in the economy through stabilizing the level of prices. The success of this strategy will depend on the credibility in maintaining the commitments that each regime implies. A currency board regime permits to go out of the crisis, while managed floating doesn't. The first is the regime that constrains the most the internal monetary policy, which will be totally determined by the monetary policy of the benchmark country (usually US). The second regime is the one that is the worst to recover the confidence, it neither implies a commitment in terms of monetary policy nor responds to the market discipline, so is the one who poses more uncertainty in the policy decisions, which are discretionaries, and therefore in the economy. A free floating regime combined with a volume of capital net flows above 3% can lead the economy to overcome the crisis, if the capital net flows are below 3% then crisis will continue. A free floating regime implies that the currency exchange is fixed by the markets and this tool cannot be used to maintain competitiveness. For the pegged regime it depends on the GDP per head, poorer countries cannot support the restriction that this type of commitment implies so this policy is not credible and therefore cannot attract investment and cannot contribute to the end of the crisis. This commitment is not as restrictive for the internal monetary policy as the currency board but also implies a sacrifice. Given that the commitment is not so well defined this regime finds more difficult to be credible, and this difficulty in being credible is more important as higher the sacrifice it implies in real terms (the case of poorer countries). Then it is a combination of a strong commitment with credibility that seems to work.

Next, the decision list obtained with PART algorithm is shown:

```
=== Classifier model (full training set) ===

PART decision list
------------------

PREVIOUS_CRISES = 0 AND
Real.interest.rate <= 7.3: non-crisis (175.0/15.0)

GDPPERHEAD > 1236 AND
FOR_LIAB_REV <= 0.82183 AND
GDP_GROWTH.ANNUAL... > 8: non-crisis (25.0/4.0)

FOR_LIAB_REV > 0.836364: non-crisis (24.0/1.0)

GDPPERHEAD <= 1236: crisis (18.0/1.0)

EXCHANGE = FF AND
X.NETKFLOWS > 2.746635 AND
CBANKINDEP > 0.49: non-crisis (16.0/4.0)

EXCHANGE = FF: crisis (27.0/1.0)

Monetary.Policy = 0 AND
BANKCASH.REV > 0.049501: non-crisis (42.0/10.0)

GDP_GROWTH.ANNUAL... > 1 AND
CBANKINDEP > 0.57: non-crisis (37.0/11.0)

CBANKINDEP > 0.45: crisis (35.0/5.0)

: non-crisis (22.0/10.0)

Number of Rules  :      10
```

Each rule in the decision list consists of one or more conditions that must all be satisfied if the rule is to be applicable, the class predicted by the rule and a number *n* or *n*/*m* which means the same that in the case of the tree leaves. There is also a default class, here "non-crisis", which is used when an object does not match any of the sequential rules.

Performance on the training cases and on the test cases is the same (in global percentage) with this decision list as with the previous tree, as we can see in the following figure:

```
=== Evaluation on training set ===
=== Summary ===

Correctly Classified Instances       359         85.2732 %
Incorrectly Classified Instances      62         14.7268 %
Total Number of Instances            421

=== Detailed Accuracy By Class ===

TP Rate   FP Rate   Precision   Recall  F-Measure   Class
  0.976     0.43      0.839      0.976    0.902     non-crisis
  0.57      0.024     0.913      0.57      0.702     crisis

=== Confusion Matrix ===

   a    b   <-- classified as
 286    7 |   a = non-crisis
  55   73 |   b = crisis



=== Evaluation on test set ===
=== Summary ==

Correctly Classified Instances        80          80      %
Incorrectly Classified Instances      20          20      %
Total Number of Instances            100

=== Detailed Accuracy By Class ===

TP Rate   FP Rate   Precision   Recall  F-Measure   Class
  0.859     0.306     0.833      0.859    0.846     non-crisis
  0.694     0.141     0.735      0.694    0.714     crisis

=== Confusion Matrix ===

  a   b   <-- classified as
 55   9 |   a = non-crisis
 11  25 |   b = crisis
```

Regarding the economical interpretation of the decision list, if we look at the first rule that applies for the 40% of the cases the variables that play a major role in the classification are ¨previous crisis¨ and real interest rates.

The most important variable in explaining the emergence of a crisis in one country is the real interest rate. Real interest rates lower than 7.3% are associated with a situation of non crisis in the next period. Therefore if real interest rates are below this level the probability for a country to enter in a crisis would be quite low.

Here again the real interest rate seems to play a determinant role in the emergence of a crisis in a country. High interest rates are probably the result of a high risk premium as a consequence of a very uncertain economical or political situation.

Following with the decision list, now for countries in which the crisis has already started or countries where there has not been previous crisis but the real interest rate is above 7.3%, the next rule that allows for a clear classification of a group of units is in situations in which the GDP per head is above 1236 $, the ratio of foreign liabilities is below 80% and the ratio of GDP growth is above 8%. For these countries the rule predicts no crisis.

That means that:

1) Countries where no previous crisis, even with real interest rates above 7.3% will not enter in crisis if they are countries with a minimum level of wealth (above 1236), a vigorous growth (above 8%), and a not excessive ratio of foreign liabilities (below 0.8).
2) Countries where are already in a crisis, no matter the level of real interest rates, but that exhibit a minimum level of wealth, a vigorous growth, and a not too high ratio of foreign liabilities will go out of the crisis.

Again it seems that a high level of growth make easier to avoid or to go out of a crisis.

Then we have now countries that do not fulfil any of the previous rules as candidates for the application of the next rule that is quite simple: countries with a foreign liabilities ratio above 0.8 will not experience a crisis in the next period.

That includes:

1) Countries where no previous crisis, high real interest rates and below a minimum level of wealth or/and a not so vigorous growth.
2) Countries where are already in a crisis, no matter the level of real interest rates, below a minimum level of wealth, or/and a vigorous growth.

Recalling that we said when interpreting the tree in the short term the possibility to hold foreign deposits within the banking system can help banks, at least temporarily, in their efforts to maintain their deposit base in terms of instability.

Going ahead for the countries still not classified because do not fulfil any of the previous rules, then the level of wealth of the population appears determinant. For poorer countries crisis will be the outcome.

For these countries it would be very difficult to overcome this situation by using the monetary or exchange rate policies to recover the confidence. The solutions should fall in the real field. Moreover, it is very difficult that any policy that implies a sacrifice was credible because the cost in real terms is so high that is not affordable. These countries have real problems that do not allow for a resolution of the financial crisis. There are some prerequisites for a sound functioning of any financial system that consist of a minimum level of development and stability at a macro level.

For countries that do not fulfil the three first rules but with a minimum level of development the situation will depend on the appropriate policy measures.

For countries that opt for a free floating regime, combined with some minimum independence on monetary policy and with a ratio of capital net flows above 2.7 the result will be no crisis. For countries that opt for free floating but with a less independent monetary policy and/or less capital flows the result will be crisis. When any other currency regime is chosen, then we have that for a monetary policy regime based on inflation rate targets (the less discretional regime) in a situation of enough liquidity in the system then the outcome will be no crisis. In any currency regime other than free floating and excluding the cases belonging to the previous paragraph then for a higher central bank independence and GDP growth above 1% also there will be non crisis in the next period. For the rest a higher independence of the central banks will end in crisis.

Therefore the independence of the central bank is not always a positive contribution to avoid or to overcome a crisis, some minimum conditions should be fulfilled to allow the monetary policy be credible. One of the conditions for that is that the commitment this type of policy implies can be assumed by the population.

5. Conclusions

In this paper we have applied two data analysis methodologies of the field of Machine Learning, C4.5 and PART algorithms, on a sample of countries in the period 1981 to 1999 with the purpose of analyzing the role of a set of variables in explaining banking crises.

The general results we obtained in terms of classification with the selected models are very good, with a 85% of correct classifications. We have checked the robustness of this result by using a test sample and the result we obtained is slightly worse than the previous one but it is still very good (80%).

Besides that our empirical results show that these techniques offer a great predictive accuracy, they are non-parametric, or distribution free, methods, so that they don’t require adopting restrictive assumptions about the characteristics of probability distributions of the variables and errors of the model, and the decision models provided by them (a decision tree or list) are easily understandable and interpretable. Therefore, these representations of the results make easier the economical interpretation than other non-parametric techniques like Neural Networks, Support Vector Machines or Rough Sets. See, for example, Sanchis *et al.* (2006), who trying to get closer to the factors that can cause the financial instability in a country obtained similar results in terms of correct classifications (80% in the test set), but the model they present is more difficult to interpret. On the other hand, our results are quite satisfactory compared with previous analyses based on traditional statistical techniques. For example, Demirgüç-Kant and Detragiache (1997) obtained worse results in terms of in-sample classification accuracy using a multivariate logit model to study the factors associated with the emergence of systemic banking crises, and in general the corrected R-square are well below this percentage (Domaç and Martinez Peria, 2000; Eichengreen and Arteta, 2000; etc.).

Some interesting conclusions emerge from the results in terms of economic policy. The first conclusion is that there are real variables (real interest rates) who determine the probability to enter in a crisis. The second conclusion is that once the crisis is in place then the recovery of the confidence in the system is crucial to overcome the situation. In the begining of a crisis internal policies can be enough to recover the confidence, but when the crisis is more persistent then the government should “borrow” this confidence from a reference country by using the exchange rate regime to anchor their monetary policy and therefore the expectations. The quality of the commitment in terms of monetary policy and the credibility of this commitment are two key factors in determining the success of a policy in going out of a crisis.

Although our sample does not include the recent crisis period the results stressed how important is in containing a crisis to quickly restore the market confidence. Although there were many analysts that had anticipated the possibility of the recent crisis they in general had underestimated the prolonged period of turbulences we are still facing as a consequence of the difficulty in restoring market confidence. One of reasons that explain these difficulties is the lack of transparency that make it difficult for stakeholders to properly discriminate between good and bad institutions, but also the lack of credibility of the financial authorities (also because the important gaps of information these authorities were confronted with).

As shown by the experiment carried out, these machine learning approaches are competitive alternatives to existing prediction models for the same problem and have great potential capacities that undoubtedly make them attractive for application to the field of business classification.

Acknowledgements

This work has been partially supported by *Ministerio de Educación y Ciencia* of Spain through project ref. SEJ2005-06744/ECON.

Data Appendix

Financial Crisis Database

Dependent Variable

Systemic and non-systemic banking crises dummy: equals one during episodes identified as in Caprio and Klingebiel (2003). They present information on 117 systemic banking crises (defined as much or all of bank capital being exhausted) that have occurred since the late 1970s in 93 countries and 51 smaller non-systemic banking crises in 45 countries during that period. The information on crises is cross-checked with that of Domaç and Martínez-Peria (2000) and with IMF staff reports and financial news.

a) Monetary policy variables

* *Monetary policy strategies*: these variables (Exchange rate target, Monetary policy target) are dummies. The exchange rate target takes four values depending on the exchange rate regime: free floating, managed floating, pegged currencies and currency board. The Monetary policy target equals one during periods in which targets were based on monetary aggregates, two when the objective was inflation, three when the two variables are into the objective function and cero in other case, according to the chronology of the Bank of England survey of monetary frameworks, in Mahadeva and Sterne (2000). Since it provides a chronology for the 1990s, we have complemented it with information from other sources for the previous years. Regarding exchange rate arrangements, we use classifications of exchange rate strategies in Reinhart and Rogoff

(2002), Kuttner and Posen (2001), and Berg *et al.,* (2002) for Latin America countries. Data for monetary and inflation targets were complemented with the information taken from Kuttner and Posen (2001) and Carare and Stone (2003). It should be noted that some judgement has gone into the classification of regimes.

* *Central Bank Independence* measures to what extent the central banks are legally independent according to their charters, following the approach of Cukierman *et al*. (1992). This variable goes from 0 (least independent) to 1 (most independent) and is taken from Cukierman *et al*. (1992), for the 1970s and 1980s. For the 1990s, Mahadeva and Sterne (2000) and Cukierman *et al.* (2002). The index of independence is assumed to be constant through every year of each decade.

b) Macroeconomic variables

* *Inflation:* percentage change in the GDP deflator. *Source:* International Monetary Fund, International Financial Statistics, line 99bir.

* *Real Interest Rate*: Nominal interest rate minus inflation in the same period, calculated as the percentage change in the GDP deflator. *Source:* International Monetary Fund, International Financial Statistics. Where available, money market rate (line 60B); otherwise, the commercial bank deposit interest rate (line 60l); otherwise, a rate charged by the Central Bank to domestic banks such as the discount rate (line 60).

**Net Capital Flows to GDP:* Capital Account + Financial Account + Net Errors and Omissions. *Source:* International Monetary Fund, International Financial Statistics, lines (78bcd + 78bjd +78cad).

* *Real GDP per capita in 1995 US dollars:* this variable is expressed in US dollars instead of PPP for reasons of data availability. GDP per capita in PPP was available only for two points in time. *Source:* The World Bank, World Tables; and EBRD, Transition Report, for some transition countries.

* *Real GDP growth* : percentage change in GDP Volume (1995=100). *Source:* International Monetary Fund, International Financial Statistics (line 99bvp) where available; otherwise, The World Bank, World Tables; and EBRD, Transition Report, for some transition countries.

* *World Real GDP growth :* percentage change in GDP Volume (1995=100). *Source:* International Monetary Fund, International Financial Statistics (line 99bvp) where available; otherwise, The World Bank, World Tables; and EBRD, Transition Report, for some transition countries.

c) Financial variables

* *Domestic Credit growth*: percentage change in domestic credit, claims on private sector. *Source:* International Monetary Fund, International Financial Statistics, line 32d.

* *Bank Cash to total assets*: Reserves of Deposit Money Banks divided by total assets of Deposit Money Banks. *Source:* International Monetary Fund, International Financial Statistics, line 20 divided by lines (22a + 22b + 22c +22d +22f).

* *Bank Foreign Liabilities to Foreign Assets*: deposit money banks foreign liabilities to foreign assets. *Source: I*nternational Monetary Fund, International Financial Statistics, lines (26c) divided by lines (21+26c).

* *Previous Crises*: This variable equals zero if the country has not previous crisis; one, if the country has suffered one previous crisis; two, in case of two or three previous crises, and, three, otherwise.

References

Berg, A.; Borensztein, E. and Mauro, P. (2002). An evaluation of monetary regime options for Latin America, *IMF Working Paper* **WP/02/211**.

Berzal Galiano, F. (2002). ART: un método alternativo para la construcción de árboles de decisión. Ph. D. Dissertation, Universidad de Granada, Spain.

Bonsón Ponte, E.; Escobar Rodríguez, T. and Martín Zamora, M.P. (1996). Sistemas de inducción de árboles de decisión: utilidad en el análisis de crisis bancarias. Zaragoza: *Biblioteca Electrónica CiberConta*. (http://ciberconta.unizar.es).

Bordo, M. (1986). Financial crises, banking crises, stock market crashes, and the money supply: some international evidence, 1980-1933, in F. Capie and G. Wood (eds.): *Financial crises and the world banking system*. New York: St. Martin's.

Breiman, L.; Friedman, J.H.; Olshen, R.A. and Stone, C.J. (1984). *Classification and Regression Trees*. Belmont, California: Wadsworth International Group.

Caprio, G. and D. Klingebiel (1997). Bank Insolvencies: Cross-Country Experience, *Policy Research Working Paper* n.º **1620**, The World Bank, Washington DC.

Caprio, G. and Klingebiel, D. (2003). Episodes of Systemic and Borderline Financial Crises, *Dataset mimeo*, The World Bank.

Carare, A and Stone, M. (2003). Inflation Targeting Regimes, *IMF WP*/03/9.

Cukierman, A.; Webb, S.B. and Neyapti, B. (1992). Measuring the Independence of Central Banks and its effect on Policy Outcomes, *The World Bank Economic Review***,** **6**, pp. 353-398.

Cukierman, A.; Miller, G. P. and Neyapti, B. (2002). Central bank reform, liberalization and inflation in transition economies˜ an international perspective, *Journal of Monetary Economics,* **49**, pp. 237-264.

Demirguç-Kunt, A. and Detragiache, E. (1997). The determinants of banking crisis: Evidence from developing and developed countries, *IMF WP* nº. WP/97/106 (September).

Demirguç-Kunt, A. and Detragiache, E. (1998). Financial liberalization and financial fragility, *IMF WP* nº. **WP/98/83** (March).

Demirguç-Kunt, A. and Detragiache, E. (2000). Does deposit insurance increase banking system stability?, *IMF WP* nº. **WP/00/3** (January).

Domaç, I. and Martínez Peria, M. S. (2000). Banking crises and exchange rate regimes: is there a link?, *The World Bank Working Paper* n.º **2489**.

Eichengreen, B. (1997). Exchange rate stability and financial stability. University of California, Berkeley *WP* nº C97-092 (June).

Eichengreen, B. (2000). *When to dollar ice*, unpublished, University of California, Berkeley (January).

Eichengreen, B. and Arteta, C. (2000). *Banking crisis in emerging markets: presumptions and evidence,* Center for International and Development Economics research (University of California, Berkeley).

Eichengreen, B. and Rose, A. (1998). Staying afloat when the wind shifts: external factors and emerging-market banking crisis, *NBER WP* nº **6370** (January).

Frank, E. (2000). Pruning Decision Trees and Lists. Ph. D. Dissertation, University of Waikato, New Zealand.

Frank, E. and Witten, I.H. (1998). Generating Accurate Rule Sets Without Global Optimization, in J. Shavlik (ed.): *Proceedings of the Fifteenth International Conference on Machine Learning*, Madison, Wisconsin. San Francisco: Morgan Kaufmann, pp. 144-151.

Friedman, M. and Schwartz, A. J.(1963). *A Monetary History of the United States*, 1867-1960, Princeton University Press, Princeton, NJ.

Frydl, E. (1999). The length and cost of banking crisis, *IMF Working Paper* nº **WP/99/30** (March).

García Herrero, A. (1997). Monetary impact of a banking crisis and the conduct of monetary policy, *IMF Working Paper* **97/124**.

García Herrero, A. and Del Río, P. (2003). Financial stability and the design of monetary policy, *Documento de trabajo* n. **0315**. Banco de España.

Glick, R. and Hutchinson, M. (1999). Banking and currency crisis: how common are twins?, unpublished, Federal Reserve Bank of San Francisco and UC Santa Cruz (September).

Gorton, G. (1988). Banking panics and business cycles, *Oxford economic papers*, **40**, pp. 751-781.

Gourinchas, P.; Valdes, R. and Landerretsche, O. (1999). Lending booms: some stylized facts, unpublished, Princeton University and Central Bank of Chile (August).

Gupta, P. (1996). Currency crises, banking crises and twin crises: a comprehensive review of the literature, mimeo, International Monetary Fund.

Hardy, D. and Pazarbasioglu, C. (1998). Leading indicators of banking crisis: was Asia different?, *IMF WP* nº **WP/98/91** (June).

Issing, O. (2003). Monetary and financial stability: is there a trade-off?, conference on *Monetary stability, financial stability and the business cycle,* March 28-29, 2003, Bank for International Settlements, Basle.

Kaminsky, G. and Reinhart, C. (1999). The twin crises: the causes of banking and balance-of-payments problems, *American Economic Review,* vol. 89, n.º 3, pp. 473-500.

Krippendorff, K. (1986). *Information theory: structural models for qualitative data.* Beverly Hills: Sage.

Kuttner, K.N. and Posen, A.S. (2001). Beyond Bipolar: A three-Dimensional Assessment of Monetary frameworks, *Oesterreichische Nationalbank Working Paper*, **52**.

Lindgren, C. J.; García, G. and Saal, M.I. (eds.) (1996). Bank soundness and macroeconomic policy, International Monetary Fund.

Mahadeva, L. and Sterne, G. (2000) (eds.). *Monetary Policy Frameworks in a Global Context*, London, Routledge.

Martín Zamora, M.P. (1999). *La solvencia en las cajas rurales provinciales andaluzas (1978-1985)*. Huelva: Universidad de Huelva, Servicio de Publicaciones.

Mendis, C. (1998). External shocks and banking crisis in small open economies: does the exchange rate regime matter?, unpublished, Center for the Study of African Economies, Oxford (October).

Mishkin, F. S. (1996). Understanding financial crises: a developing country's perspective, *NBER Working Paper* n.º **5600**, National Bureau of Economic Research, Cambridge, Mass.

Padoa-Schioppa, T. (2002). Central banks and financial stability: exploring a land in between, paper presented at the Second ECB Central Banking Conference *The transformation of the European financial system,* Frankfurt am Main, October.

Quinlan, J.R. (1993). *C4.5: Programs for Machine Learning*. San Mateo, California: Morgan Kaufmann.

Reinhart , C. M. and Rogoff, K. S. (2002). The modern history of exchange rate arrangements: a reinterpretation, *NBER Working Paper* n.º **8963**, National Bureau of Economic Research, Cambridge, Mass.

Reza, F.M. (1994). *An introduction to Information Theory*. New York: Dover Publications.

Rossi, M. (1999). Financial fragility and economic performance in developing countries: do capital controls, prudential regulation and supervision matter? *IMF Working Paper* **WP/99/66** (May).

Sanchis, A.; Segovia, M.J.; Gil, J.A.; Heras, A. and Vilar, J.L. (2007). Rough Sets and the role of the monetary policy in financial stability (macroeconomic problem) and the prediction of insolvency in insurance sector (microeconomic problem). *European Journal of Operational Research*, **181,** issue 3, pp. 1554-1573.

Schwart, Z.A. (1995). Systemic risk and the macroeconomy, in G. Kaufman (ed.): *Banking financial markets and systemic risk, Research in Financial Services, Private and Public Policy*, vol. 7, JAI Press Inc., Hampton.

Stiglitz, J. and Furman, J. (1998). Evidence and insights from East Asia. *Brookings papers on economic activity* 1998:2.

Tam, K.Y. and Kiang, M.Y. (1992). Managerial applications of neural networks: the case of bank failure predictions, *Management Science*, vol. 38, nº 7, pp. 926-947.

Witten, I.H. and Frank, E. (2005). *Data Mining: Practical Machine Learning Tools and Techniques*, 2nd Edition, San Francisco: Morgan Kaufmann.

Ziemer, R.E. and Tranter, W.H. (2002). *Principles of Communications: Systems, Modulation, and Noise*. New York: John Wiley & Sons.

In: Emerging Topics in Macroeconomics
Editor: Richard O. Bailly, pp. 327-342
ISBN: 978-1-60692-093-0

Chapter 12

VIEWS FROM THE TRENCHES: INTERVIEWING BANK OFFICIALS IN THE MIDST OF A CREDIT BOOM

Neven Valev
Department of Economics, Andrew Young School of Policy Studies, Georgia State University, University Plaza, Atlanta, GA, USA

1. Introduction

Bulgaria experienced a dramatic increase in credit market activity in the last few years. The Bulgarian National Bank (BNB) reports that bank credit to private non-financial enterprises increased by about 360 percent from 1999 to 2005. In some sectors of the economy the credit growth was even faster. For example, the credit to firms in the construction industry increased 14 times during that period!

The credit boom has occurred in an environment that presents a number of challenges to lenders. Bulgaria is a new market economy where most firms have short history and little, if any, experience with credit. Tax evasion is widespread and financial disclosure documents are not very reliable. The legal enforcement of contracts is relatively weak. On the macro level, the 1997 financial crisis is a recent reminder that Bulgaria is susceptible to crises. Yet, credit markets have been buoyant. How are banks able to extend so many credits so rapidly? Are they allocating credit effectively? This paper provides a glimpse into these questions using interviews with 25 banking officials conducted in the fall of 2006.

The interviews are centered on four major theoretical functions of the financial system outlined by Levine (2005). Financial institutions benefit the economy by: 1) producing information about investment opportunities and allocating capital; 2) monitoring borrowers and exerting corporate governance; 3) diversifying cross sectional and liquidity risks; and 4) pooling and aggregating the savings of the population.[1] The interviews reveal that bankers face significant difficulties in terms of risk assessment and exerting control over borrowers related to the ongoing transition process and the relatively low level of development of

[1] Levine (2005) also discusses the role of the financial system in facilitating payments. This was not discussed after the first 2-3 interviews since the interviewees did not see any problems in that area.

Bulgaria. However, the interviews also show that banks are cognizant of these hurdles and adapt to overcome them.

The literature provides ample evidence that financial development enhances economic growth (Levine, 2005). However, this general finding is not without caveats for different regions and for countries at different levels of financial and economic development. For the transition countries, the evidence is mixed. Jaffee and Levonian (2001), Drakos (2002), Fink, Haiss, and Mantler (2005), and Fink, Haiss, and Vukšić (2004) find that financial development has contributed to economic growth whereas Mehl, Vespro, and Winkler (2006), Koivu (2002), and Mehl and Winkler (2003) do not obtain such evidence. Furthermore, the financial crises literature shows that credit booms often end in crises (IMF 2004). Whether rapid credit expansions have a positive outcome (growth) or a negative outcome (crisis) depends on how effectively the financial system performs the functions discussed by Levine (2005). Unfortunately, the finance and growth literature and the financial crises literature shed little light on these theoretical functions. Because of lack of data, the empirical literature uses overall levels of credit to measure financial development whereas, strictly speaking, financial development should be measured by how effectively the financial system performs its theoretical functions.

The objective of this paper is to understand which functions of the financial system are performed well and which ones are not performed well while the credit boom in Bulgaria is unfolding. Bulgaria is interesting because it presents a challenging environment to lenders, similar to that in other transition and developing countries. This allows us to investigate which functions are hampered most by weaknesses in the institutional environment. Institutions have been singled out as some of the most important determinants of financial development (La Porta et. al. 1998) but it is not clear which of the channels through which finance affects growth are blocked most severely when institutions are weak.

Using interviews to address these issues is not the standard practice in the literature and understandably so - interview data are difficult to collect, and the end product is relatively noisy information from only one country at a given point in time. Yet, interviews provide valuable insights that can complement earlier studies that use quantitative data.

2. Credit Boom in Bulgaria

Bulgaria had a typical Soviet-style monobank financial system until 1989. Neither domestic private banks nor foreign banks existed. The State Savings Bank collected most private savings and the central bank was in charge of almost all capital allocation (Miller and Petranov 2001, Koford and Tschoegl 2003). This gradually changed during the transition years but not without bumps along the way. Most of the serious reforms in the financial sectors were implemented after the financial meltdown in 1996-1997. For a number of years Bulgarian banks, both private and public, channeled funds to inefficient but politically important enterprises. Corruption in lending was rampant. When eventually the volume of non-performing loans became unmanageable, the banking system collapsed, the exchange rate devalued, and inflation increased to hyperinflation levels. This, along with a sharp drop in output, erased the savings of the population.

The crisis prompted more serious reforms. A newly elected government introduced a currency board to reduce inflation and privatized the banking system giving favored position to foreign banks (Šević, 2000). At present, almost the entire banking system is privately owned and most banks are foreign owned. The privatization of banks, along with successful macroeconomic stabilization, robust growth, capital inflows, approaching EU entry at that time, and low international interest rates set the stage for the rapid credit expansion.

Table 1 shows that credit to private non-financial enterprises expanded by 361 percent during the period from 1999 to 2005, or nearly 24 percent each year. Credit increased from about 6.2 billion leva in 1999 to about 22.6 billion leva in 2005.[2] The credit expansion varied substantially across economic sectors. Of the main economic sectors in terms of size, credit grew especially rapidly in the construction sector (1,406%), the real estate sector (3,446%), the hotels and restaurants sector (506%), and the trade sector (510%). Credit to manufacturing grew less rapidly (192%). By 2005, the credit to the manufacturing sector was smaller compared to the credit to the service sectors whereas their relative sizes were reversed in 1999 when the manufacturing sector was the largest recipient of credits. Agriculture also lagged behind in terms of credit growth with an increase in credit by about 277 percent during the period.[3]

Table 2 shows that the growth in credit activity in Bulgaria has come primarily from the growth of the number of loans extended rather than the growth in the average size of a loan. The number of loans with size greater than 500,000 leva has grown about 5 times from 701 in 1999 to 3,541 in 2005 whereas the number of loans in the 5,000 leva or less category has grown by almost ten times from 3,394 in 1999 to 30,166 in 2005. Furthermore, the average size of the loans in the latter category has declined from 1,923 leva to 968 leva. Hence, while the banking system is extending more loans in every size category, the tendency is to extend a disproportionately large number of small loans.

Table 3 shows that the maturity of credits has lengthened over time. In 1999, short-term credit with maturity of one year or less was about 74 percent of total credit while long-term credit with maturity of more than five years was only 1.7 percent of total credit. By June 2006, short-term credit had decreased to about 31 percent, while long-term credit had increased to 24 percent of total credit. This is similar to the maturity composition of credit in other Central and Eastern European countries as well as to the maturity of credit in most West European economies (Tasic and Valev 2007).

[2] The lev is fixed to the euro at about 1 euro = 2 leva rate (leva is the plural of lev in Bulgarian).

[3] In what follows I focus on the role of commercial banks and do not discuss stock markets, bond markets, and other non-bank financial institutions as they are only a small portion of the Bulgarian financial system. Furthermore, following the overwhelming majority of papers on finance and growth I focus on business credit to the exclusion of household credit.

Table 1. Commercial Bank Credit (in thousands of BGN)

Sector	1999	2000	2001	2002	2003	2004	06/2005	Total Growth	Yearly Growth
Real estate, renting and business activities	11019	25535	59974	102306	107795	254453	390800	3446.60%	73.95%
Financial intermediation	3428624	4407517	4772322	4366511	3589001	5592031	6338815	84.88%	9.09%
Transport, storage and communication	117706	117860	235602	313236	256886	435969	591591	402.60%	26.40%
Hotels and restaurants	135569	171344	288979	376581	459074	748937	822090	506.40%	33.58%
Trade	710749	833163	1361423	2069236	2974955	4008647	4335534	510.00%	33.93%
Construction	46122	62533	97692	173583	259566	482427	694618	1406.04%	45.72%
Electricity, gas and water supply	105263	37732	67621	70437	301507	355353	368586	250.16%	60.26%
Manufacturing	942392	1275285	1237539	1651510	2125068	2622819	2753627	192.20%	18.86%
Mining and quarrying	57286	55111	56809	76310	85851	107745	117366	104.88%	10.57%
Agriculture, hunting, forestry and fishing	127320	124723	164474	167353	263338	376242	480561	277.44%	18.35%
Education	587	431	1661	1728	7066	11455	8780	1395.74%	110.72%
Public administration	4214	4483	6947	10133	43748	79078	111722	2551.21%	81.75%
Community, social and personal service activities	69503	173552	191457	400916	691674	1120641	1431817	1960.08%	63.71%
Health and social work	1609	3377	18212	35128	65324	98678	152105	9353.39%	123.99%
Total	6270688	7894123	9357511	10784409	13012547	19735855	22647412	361.17%	23.89%

Table 2. The Size Distribution of Loans: Number of Loans and Average Loan Amount (in 1000 BGN) in Each Size Category

	1999	2000	2001	2002	2003	2004	06/2005
up to BGN 5000	3394	3590	3253	4922	14541	24651	30166
	1.9239463	1.672423	2.086382	1.7369141	0.44042504	0.391172378	0.9681047
from BGN 5000 to BGN 20 000	3803	4559	5052	6730	6657	7794	8602
	22.932514	23.35561	24.03891	24.119639	12.0244072	24.19078737	11.909046
from BGN 20 000 to BGN 50 000	2792	3938	4863	5601	5638	7266	8243
	31.00918	32.1191	32.63294	33.845679	24.44388	33.85176796	34.317864
from BGN 50000 to BGN 100 000	1325	1732	1944	2818	3654	5017	5973
	145.53973	137.8408	142.3737	147.11279	74.2788153	149.5305791	74.454405
from BGN 100 000 to BGN 200 000	859	1053	1224	1744	2412	3392	4069
	142.86107	137.7303	142.3301	145.5199	89.9351075	147.6948125	148.74919
from BGN 200 000 to BGN 500 000	637	903	1169	1526	2027	2697	3035
	652.39418	623.9233	628.9853	640.24446	334.964622	637.3726129	317.32429
over BGN 500 000	701	875	1182	1633	2381	3140	3541
	1915.6995	2067.697	2043.574	2158.0916	1008.69491	2277.111095	2185.6064

Table 3. Credit Maturity: Percent short-term (maturity 1 year or less) and percent long-term (maturity longer than 5 years) debt

Sector	1999	2000	2001	2002	2003	2004	2005	06/2006
Real estate, renting and business activities	50.6%	14.5%	44.4%	57.5%	39.7%	38.4%	22.8%	24.5%
	0.0%	5.8%	11.9%	2.8%	10.1%	10.5%	30.5%	26.7%
Financial intermediation	97.5%	96.8%	99.2%	99.0%	96.2%	96.4%	-	-
	0.5%	0.0%	0.1%	0.1%	0.1%	0.2%	-	-
Transport, storage and communication	48.9%	49.1%	34.9%	55.2%	34.6%	33.6%	24.3%	23.1%
	1.0%	0.7%	6.7%	0.4%	1.6%	4.1%	15.7%	30.2%
Hotels and restaurants	31.6%	16.8%	26.7%	29.6%	14.3%	15.9%	6.2%	7.2%
	11.7%	6.2%	1.0%	3.2%	16.2%	26.3%	55.1%	55.1%
Trade	59.2%	56.5%	54.1%	52.5%	46.8%	48.3%	38.6%	36.8%
	0.2%	1.5%	6.9%	3.1%	5.4%	8.8%	19.5%	20.7%
Construction	52.7%	52.1%	65.6%	49.8%	38.1%	38.7%	24.3%	25.9%
	0.0%	0.2%	0.4%	2.9%	5.6%	10.9%	19.1%	13.8%
Electricity, gas and water supply	64.4%	35.1%	31.6%	44.2%	17.5%	31.1%	22.1%	13.5%
	0.0%	1.5%	0.7%	0.7%	51.6%	47.1%	60.6%	71.6%
Manufacturing	54.9%	58.8%	53.1%	54.7%	48.6%	48.8%	34.3%	33.2%
	0.2%	3.5%	3.5%	4.6%	7.8%	10.3%	18.2%	18.2%
Mining and quarrying	51.9%	27.9%	25.8%	67.7%	57.2%	33.6%	52.1%	59.6%
	0.0%	0.0%	7.4%	10.8%	15.3%	0.0%	6.0%	13.6%
Agriculture, hunting, forestry and fishing	44.9%	56.8%	51.3%	51.2%	50.9%	47.4%	27.2%	29.8%
	0.5%	1.6%	8.1%	1.4%	1.8%	3.7%	16.5%	20.8%
Community, social and personal service activities	24.7%	37.9%	23.2%	24.5%	13.4%	16.8%	19.5%	18.5%
	0.7%	4.8%	17.8%	17.5%	11.8%	21.3%	34.9%	33.2%
Health and social work	58.1%	79.9%	34.6%	43.8%	20.7%	34.3%	23.9%	37.6%
	0.0%	0.0%	1.6%	6.8%	20.0%	31.5%	41.2%	31.2%
Education	29.6%	36.9%	41.5%	22.7%	14.5%	69.9%	6.6%	7.5%
	70.2%	0.0%	0.0%	18.7%	6.1%	1.1%	16.6%	21.3%
Public administration	75.4%	89.0%	62.7%	54.5%	13.8%	63.7%	-	-
	0.0%	0.0%	0.0%	0.0%	30.2%	20.5%	-	-
Total	73.7%	74.6%	71.3%	65.7%	52.0%	52.6%	31.0%	30.7%
	1.7%	2.2%	3.4%	3.8%	8.7%	13.7%	23.9%	24.0%

Part of the reason for the rapid growth of credit in Bulgaria is probability the low starting level of credit as percent of GDP. Beck, Demirgüç-Kunt, and Levine (2000) report that credit to the private sector, including firm credit and household credit, in Bulgaria was 10.86 percent of GDP in 1999. For comparison, private credit was 37.41 percent of GDP on average in five Central European transition economies (The Czech Republic, Hungary, Poland, the Slovak Republic, and Slovenia). Their updated data set shows that in 2004, private credit was 30.85 percent of GDP in Bulgaria and a very similar 34.95 percent in those countries (which was actually less than their 1999 levels). However, a low initial level of private credit does not guarantee a subsequent rapid credit growth. For example, consider Romania which is a neighboring country of Bulgaria with a similar level of economic development and the same EU entry date. Private credit in Romania was 8.48 percent of GDP in 1999 and 8.78 percent of GDP in 2004 (although credit appears to have picked up in the years since then).

So far banks appear to be making prudent decisions. The percent “standard loans,” i.e. loans that are not in doubt of repayment, are subject to provisions or are outright uncollectible, were 92.32 percent of total loans at the end of 2005 (BNB, 2005). This is a large change compared to 1996 when standard loans were only 43.67 percent of all loans in the large state-owned banks, 33.41 percent of the loans of smaller privately owned banks, and 0.21 percent (not 21 percent, 0.21 percent) of the credits by foreign banks (BNB, 1996).

3. Interviews with Commercial Bank Officials

To gain insight into the mechanics of the credit process, I interviewed 25 officials from the banking system from August to December 2006. The interviewees included bank officials from a wide variety of banks in the capital Sofia and outside Sofia. Some interviewees were from small banks without a large branch network that specialize in particular financing while other interviewees were employed by the largest banks in Bulgaria. The sample of interviewees also ranged across positions within the banking system. I interviewed credit inspectors for small and micro enterprises at bank branches as well as branch managers, and senior economists. Some interviewees specialized in corporate credits while others had significant experience in the largely dormant but developing credit market for agriculture activities. Many of the bank officials had working experience at other banks since the turnover in the banking system is high and many had experience from several positions, e.g. micro credits, corporate credits, mortgages, research and even board of directors. The interviews took about 1 hour during office hours or after work. The following sections summarize the views of the interviewees. The sections follow the four functions of the banking system discussed earlier.

Producing Information about Possible Investments and Allocating Capital [1]

The bankers explained that having a relatively long (2-3 years or more) history with a bank is an important determinant of whether that bank will extend credit to a borrower. This was true particularly at the branch level from where banks channel financing to micro, small, and medium size firms. If the bank is not familiar with the borrower (and in some banks this is the procedure with all borrowers) the loan application goes through a risk department that checks for links to organized crime and illegal activities, among other things. This is accomplished on the basis of documentation and sometimes site visits. After the borrower is cleared through the risk department, the procedure for extending the credit can be completed.

Credit information sharing

A credit register started operating in Bulgaria only a few years ago. Only then banks could become informed about the total indebtedness of borrowers. Although the credit register is not complete, e.g. it does not include information on leasing credit, it provides valuable information about indebtedness. For example, a bank discovered that one of its borrowers had accumulated 18 credits from 16 different banks when it looked up his indebtedness in the newly opened credit register!

The credit register provides information about the current indebtedness of a borrower but it does not show the credit history of a borrower. The absence of credit history is a major impediment in the credit evaluation process and explains why banks lend primarily to firms that are their long-term clients. This also contributes to the aggressive competition between banks to attract credit inspectors with an established client base from other banks. Attracting bank officials from other banks is a way to gain access to the information held at these institutions.

One interviewee explained that a few large banks with long standing in the economy were reluctant to share their proprietary information about the credit history of borrowers. Sharing that information would have made it easier for firms to shop around for credit from other financial institutions. Therefore, these banks pressed for excluding the credit history from the credit register when it was first designed.

Tax evasion and credit activity

Banks estimate that more than half of the profits of small and medium sized firms are unreported. The overwhelming majority of firms, if not all, keep two sets of books – the official set of books for tax purposes and a second set of “black” books. The official books often show losses or negligible profits whereas the black books show positive and sometimes very healthy profits. Many banks have found ways to incorporate information about the true conditions of a borrower into their decision making process. In fact, some banks specialize in analyzing unofficial information by making site inspections and using the black books. In the cases when the reported official financial situations of a firm cannot substantiate a loan, the loan is justified by preparing a favorable “economic potential” valuation.

[1] If the cost of evaluating projects is high relative to the expected returns accruing to individual investors, investors would prefer assets that have lower returns but are also easier to evaluate. Many high-value projects will not be undertaken leading to slower economic growth. In that environment, financial intermediaries emerge as institutions that specialize in the acquisition, processing, and production of information improving the allocation of savings.

The use of information other than the official books in the evaluation process is not a secret to anyone in the industry including the supervision authorities. Indeed, a bank inspector might raise questions about a credit that appears unjustified by the financials of the borrower. However, the credit will not be reclassified as a problem credit by the supervisor if there is evidence that the borrower is doing well even if the official financial documents do not support this.

A priori, one might suppose that financial development would provide a strong incentive for firms to reveal positive profits. If access to credit is more important relative to the savings from tax evasion, then firms would prefer to reduce the shadow portion of their activities. While there is some validity to this argument, it seems that many banks are also adjusting their practices to the conditions in the real economy. Also, the government apparently is reluctant to fight tax evasion through the financial system at the expense of reducing the flow of external financing to firms. Nonetheless, the discrepancy between official and unofficial information is a problem for banks because unofficial information is costly to verify. Furthermore, not all banks use unofficial information in their evaluation process which limits the range of borrowers who approach them for credit.

Lack of experience on the part of borrowers

Bankers often complained that they receive requests for credit for activities that are far removed from the current business of the applicant. A typical example is someone who owns, say, a restaurant in the capital Sofia and requests funds to build condominiums at one of the resorts in Bulgaria. The applicant has no experience in the construction business and the business plan is weak. Nonetheless, the borrower wants to participate in the construction boom because the profit margins are reportedly high. While this may sound strange in a developed economy where the major players in a sector are established, in a new market economy like the Bulgarian one, many sectors are not yet dominated by established firms. This leaves space for newcomers. Furthermore, the profit margins in a capital poor country like Bulgaria are high. This attracts inexperienced entrepreneurs who may obtain a positive profit from a project even if the execution of the project is less than perfect.

Evaluating the environment of the borrower

Several bankers at the local branches claimed that evaluating the individual borrower is not difficult, especially if the borrower is a long-term client of the bank. However, they complained that obtaining information about the sector or the geographical region in which the borrower operates is a major challenge. These difficulties are due to the lack of comprehensive and timely data on sectoral and regional economic dynamics but also to the rapidly changing economic structure of Bulgaria. The economy still experiences dramatic structural changes as a result of continuing reforms. Some sectors such as construction, tourism, and finance have seen exceptional rates of growth. However this growth has been concentrated mostly in the large cities and the major resorts and has fueled massive internal migration. Forecasting how long these growth rates will last and what is the next region with rapid growth is not an easy task. Recognizing the need for such information, the research department of one of the large banks in Bulgaria has started a project for sectoral and regional analysis. Reportedly, there is significant interest in their analysis by various participants in the financial system.

Who makes the final decisions?

Every level in the banking system faces a limit in terms of the maximum size of a credit it is allowed to extend. The limits are set at the headquarters of the banks and may differ across the various branches of the same bank depending on the location of the bank branch and the experience of the branch officials. For example, in one bank, a new branch starts with 20,000 leva (10,000 euro) limit and the limit can increase up to 500,000 leva. Furthermore, the credit inspectors within the same branch have different limits on the maximum credit size. Any credit with size greater than what can be decided at the branch level is decided in the central office in Sofia and, when the bank is foreign owned, the largest credits are decided in the headquarters abroad. The size limitations were introduced after the 1996 crisis. Before that, branch manages had discretion with large credits which lead to corruption and to misallocation of resources.

Bank officials believe that the size limitations at the branch level do not pose a problem in terms of restricting credit or hampering the assessment process. Similarly, none of the interviewees expressed concern that the big decisions are made abroad. In fact, one official argued that the headquarters of international banks are better equipped to evaluate large projects compared to local credit inspectors since officials at the headquarters can draw on experience from other countries. Overall, the new managerial structures and the incentive schemes introduced by foreign banks were received well by the interviewees.[2] All bank officials were eager to point out that, in terms of management style, banks function much better compared to the period before the crisis. The managerial know-how introduced with the entry of foreign banks seems to have contributed significantly to this positive change.

Monitoring Investments and Exerting Corporate Governance[3]

To understand the challenges involved in the process of monitoring, one must distinguish between the two general types of credits extended by the banking system to non-financial firms – investment credits and turnover/overdraft credits. Investment credits are extended for a particular purpose. They have maturity that reflects the time schedule of implementing the investment project. The disbursement of stages of the investment credits is for particular purposes. For example, if a firm has obtained an investment credit to build a hotel, then a

---

[2] Some examples of new managerial practices include bonuses for the sale of credits adjusted for the quality of credit portfolios; long-term contracts with a higher salary as an effort to retain top performers; salaries based on the performance of the credit department rather than on the performance of the individual member of the department to stimulate team work and internal monitoring; various team building activities and seminars throughout the year; and emphasis on the hiring of young bank officials at some banks. Overall, bankers seem to work in a highly structured environment. The overwhelming majority expressed satisfaction with their jobs although the level of compensation is still relatively low (a top credit officer can be on a 500-600 leva monthly salary) and the working hours are long and often extend beyond the official 9 to 5 hours. The officials feel that there is potential for professional growth in the banking system primarily through frequent mobility between banks. In fact, banks compete to hire strong performers who bring with them a large Rolodex of satisfied clients as clients tend to follow the bank official (assuming that the new bank is of similar reputation).

[3] In its function as a "delegated monitor" (Diamond 1984) the banking system uses economies of scale to reduce the cost of monitoring firms that would otherwise have to be incurred by the individual investors. Furthermore, it reduces the free-rider problem that arises when small individual investors rely on others to monitor the firms. Hence, well developed financial systems improve individuals' ability to monitor and influence the use of capital. Such ability in turn may influence saving behavior and investment decisions, both of which can be beneficial for growth.

stage of financing would be disbursed for the purchase of cement, another for bricks, another for the purchase of furniture, etc. Each disbursement has a particular purpose and is agreed upon at the negotiation of the credit. When the project begins to generate revenues, the proceeds are committed to paying down the interest and principal of the credit. Often the bank has the right to 100% of the generated revenues and, only after the principal is paid off, the borrower can keep the residual stream of revenues.

Obtaining an investment credit is advantageous for the borrowers because they have access to resources for the duration of the implementation of the project. The advantage for the banks is that they have a clear idea what the resources are used for. This clarity makes it easier for banks to perform their function as monitors. The disadvantage of the investment credit is that its extension is cumbersome and costly for the borrower and for the bank. The borrower has to produce a detailed business plan and documentation, and has to provide strong argumentation for the need for financing. The bank has to analyze this information.

In contrast, the overdraft and turnover credits are extended on the basis of evidence for the turnover of the firm and its current financial condition. The credit limit of the turnover credits is set as a percent of the turnover of the firm, e.g. 30% of the turnover. The advantage of this credit for the bank and for the firm is that it is easier to justify and extend. The procedure is shorter and less complicated. Furthermore, the borrower has greater discretion over the use of the borrowed funds. The credit line can be used as intended to meet the liquidity needs of the firm or it can be used for investment purposes. In the latter case the borrower can fund investment projects without a preliminary agreement with the bank. The turnover credits are riskier for the borrower because they are short-term (usually with one year maturity) and have to be used with caution for financing long-term investments. At the same time, however, they give the borrower greater discretion over the use of funds.

The concern of many of the interviewed bank officials is that the banking system is extending too many overdraft and turnover credits, which are then used for investment purposes. With this, the effectiveness of the monitoring of the use of funds is diminished and the risk of imprudent investment decisions by the borrowers increases. It is difficult to determine how widespread this practice is but more than half of the bank officials brought it up as a problem.[4]

The challenges in terms of monitoring discussed here are linked to the information challenges discussed in the previous section. Banks find it easier to extend turnover credits as they are less informationally difficult. This, however, reduces their control over the use of funds as the discretion over the use of funds is shifted to the borrower. Hence, information difficulties breed monitoring difficulties.

Aside from the issue of investment vs. turnover credits, the interviewees stressed that verifying the existence of the collateral is a frequent problem. Despite all the documentation brought to the bank by a borrower, site inspections sometimes reveal that the collateral is not in place. For example, a machine that was used as collateral was sold. Or perhaps, part of the firm was sold or transferred to a relative. A careful inspection can also reveal that the firm has a number of other formal and informal creditors who claim ownership of the collateral.

[4] As far as I am aware, the BNB or other government institutions do not provide a breakdown of commercial credits into investment, turnover, and overdraft credits. Many bank officials also mentioned that household credits, particularly mortgages but also consumer loans (which can reach 10,000 leva or more per person) are also used for investment purposes by some borrowers. Often, they are cheaper in terms of interest rates and easier to obtain in terms of documentation using personal real estate property as collateral.

Hence, the lack of easily verifiable and enforceable property rights is a problem. By law and by contract, borrowers must inform banks of ownership changes but, in practice, this information is not submitted promptly with little or no ensuing penalty to the borrower.

Now, assuming that the bank is aware of the actual use of funds and it has correct and timely information about the financial condition of the firm, what actions does it take when the borrower starts to experience difficulties? First and foremost, all interviewees agreed that banks prefer to work with the borrower to resolve the problem situation and that selling the collateral is done only as a last resort. What usually happens is that officials from the bank start spending quite a bit of time with the firm officials on the premises of the firm. Much information about the business is generated and analyzed. This can take months. If the firm and the bank can reach an agreement, the loan is restructured usually extending its maturity, giving a gratis period or making other allowances. The bank also encourages the borrower to secure funding from other sources. If agreement cannot be reached, then the bank proceeds to selling the collateral.

Selling the collateral is not done by banks but by credit collection agencies that purchase the problem loan from the banks at a discount. Usually, the procedure for selling the credit and selling the collateral takes 6-9 months but it can also take more than a year when the borrower does not cooperate. From the perspective of at least one interviewee, the collections process is difficult, lengthy and costly and this sometimes deters banks from extending credits in more risky situations.

Facilitating the Trading, Diversification, and Management of Risk[5]

Many of the interviewees attributed the 1996 banking crisis partly to the lack of diversification of bank credit portfolios. Many banks had significant exposure to one economic sector and often to one firm. The legacy of the crisis, the Basel capital adequacy standards, and the better portfolio allocation practices all contribute to greater attention paid to cross sectional diversification. However, banks do not seem to apply any particular model, or set of rules or principles to evaluate the effectiveness of their cross-sectional diversification. Overall, the interviews suggested that currently banks are more concerned with growing their market share and less so with overextending themselves in certain sectors. The rapid growth of credit to a wider range of firms in terms of size, location and sectors naturally contribute to the diversification of bank portfolios.

Spreading risk over time requires that the banking system transforms the liquid short-term deposits into less liquid long-term assets. This is beneficial to the economy as banks

[5] Well functioning financial systems provide opportunities to trade, pool, and diversify risk increasing the risk-adjusted return on savings and improving the allocation of resources: Cross sectional risk diversification*:* If individual investors can hold stakes in a pool of projects that includes a mix of low-risk and high-risk projects (e.g. a mutual fund or the credit portfolio of a bank), more high–risk/high-return projects will be funded leading to faster economic growth. Similarly, more innovation activities that are inherently risky will be undertaken leading to faster technological progress. Intertemporal risk sharing*:* A well functioning financial systems can spread the risk of systemic shocks over time. Investing with a long-term perspective, the financial system provides relatively low returns in prosperous times and relatively high returns during recessions. Lowering liquidity risk*:* Liquidity risk arises because of uncertainty associated with converting assets into medium of exchange. In a well functioning financial system individual investors will hold assets that are liquid but the financial system will transform the resources into high-return/less-liquid assets contributing to more investment into long-term/high-yield projects leading to faster economic growth.

finance long-term investment projects that raise productivity. Most interviewees agreed that short-term credits are preferred by the banking industry as banks are wary of a large maturity mismatch on their balance sheets. Often the bankers brought up their experiences during the 1996 banking crisis when demand deposits and short-term time deposits fled the banking system and the banks were left holding long-term illiquid assets. It is understood that one of the functions of the banking system is to transform short-term liabilities into long-term assets but the dislike for this function is strong, particularly at smaller institutions that face more uncertainty on the liabilities side of their balance sheets.

At present the maturity of credits in Bulgaria is similar to that of other transition countries. The maturity of investment credits can reach 5-7 years but some banks already offer 10-year maturity. The prevalent view among the interviewees is that competition between banks will lengthen the maturity of investment credits. If some banks start offering loans with long maturity, other banks will follow suit if the demand for such credits exists. However, some bankers explained that at present the demand for long-term credit is not that strong. Much of the credit demand comes from the construction industry to build hotels and residential buildings in the cities and in Bulgaria's main tourist destinations along the Black Sea coast and the ski resorts in Southwest Bulgaria. Since many buildings were selling off even before they were completed, the average life of a project in the building industry spanned about 3-4 years. Over time, capital will have to be directed to more long-term projects with the closing of short-term profit opportunities. It remains to be seen if investors will be willing to seek funding for long-term project and if banks can provide the resources at reasonable interest rates.

Mobilizing and Pooling Savings[6]

One of the main functions of the financial system is to pool the savings of the individual savers and use these pooled funds to finance large, high-productivity projects. However, according to the statistical data discussed earlier, the credit growth in Bulgaria has originated from a rapid increase in the number of borrowers more so then from an increase in the size of credits extended to the average borrower.

Bankers agreed that most credits in Bulgaria are relatively small in size. There were several reasons brought forth by the interviewees to explain why. First, lending to a greater number of borrowers increases the diversification of the credit portfolios. Second, despite the credit expansion during the last few years and the healthy profits of banks, bank capital has not grown much; this puts limits on exposure to any one firm or to a industry. Third, and most important, the overwhelming majority of firms in Bulgaria are small in size and obtain small credits. Therefore the rapid growth in small credits can be taken as evidence for greater access to finance by small and medium sized firms.

It is important to add that Bulgarian firms do have access to large credits. However, these credits are generated by banks outside Bulgaria. A prime example that was mentioned by

[6] The financial system can overcome the transaction costs of attracting the savings of economic agents into the formal financial system. Furthermore, being able to agglomerate savings, the financial system can have a profound effect on the economy as it can fund projects that need large injections of capital and cannot be completed with staged financing. Such large projects cannot be funded by individual investors but, through economies of scale, they have a strong positive effect on productivity growth and economic growth.

several officials was the credit obtained by Maritza Iztok, a major foreign-owned utility. The credit was for several hundred million euro and was arranged by a consortium of foreign banks. Larger firms, particularly those with foreign participation, can access international capital markets where they obtain larger credits at better terms.[7]

With respect to attracting the savings of the population, bankers almost unanimously claimed that trust in the banking system is strong despite the banking crisis. Depositors can choose between demand deposits, as well as time and savings deposits with various maturity structures and currency denomination in leva, euro and USD. Casual observations from the big cities and even the small towns suggest that banks operate numerous branches making it physically easy for people to go to a bank. Banks also offer innovative products. For example, one bank offered a deposit whose return was linked to an investment in a portfolio of stocks but with a guaranteed return minimum.

The interviews suggested that there are two major deterrents to participating in the financial system. One is that many people in Bulgaria have no savings: incomes are too low. The second reason is tax evasion. Banks are required to submit information about large bank transactions to the central bank. Therefore, savers with undocumented savings prefer alternative assets, mostly real estate. There also appears to be a thriving unofficial lending market which operates with these funds.

4. Conclusions

The banking system in Bulgaria has undergone a profound change in the last several years. Banks have helped relax the financing constraints on firms by offering an array of financial products and competing to improve access and the quality of service. The 1996 crisis has been a catalyst for reforms in the banking sector and for improved vigilance at the supervisory authorities. Credit has increased more than threefold since 1999.

Although their various functions, e.g. to distinguish between good and bad risk, are hampered by institutional and other factors, the functions are performed reasonably well as banks adapt to the environment. In the absence of shared credit history and with a large part of the economy underground, banks rely on relationship lending based mostly on “soft” information accumulated by doing business with a firm over time. Berger and Udell (2002) point out that relationship lending requires a shift in discretion over the lending decision down to the credit officers who are in direct contact with the borrower. However, this shift in decision making power also requires an organizational structure for effective monitoring of the lending decisions by upper management. Indeed, the conversations with the interviewees suggest that the banks strike this balance fairly well.

One relatively weak area that emerged from the interviews is the monitoring of borrowers. It seems that banks’ efforts are devoted primarily to expanding their credit portfolios and growing their market share with less attention paid to active subsequent monitoring and control. Much of the time of credit inspectors is devoted to new credits.

[7] This suggests that, ideally, the discussions of credit activity in Bulgaria (and in other countries) should incorporate international as well as domestic credit. In the case of Bulgaria, which is a net international borrower, domestic credit is an underestimate of the overall amount of credit to the private sector. However it is not clear by how much as data on international borrowing are not available. Because of similar data limitations in other countries, the finance and growth literature has focused on the effect of domestic credit on growth.

Bonuses for performance are often based on the rate of expanding a credit portfolio. Furthermore, weak contract enforcement and poorly defined and protected property rights over the collateral make it difficult to influence the behavior of borrowers.

The analysis leaves no doubt that a well functioning financial system is instrumental for economic growth. Much of the immediate boost to growth can probably be attributed to capital accumulation – the purchase of plant and equipment that are essential in a capital poor economy like the Bulgarian. Over time, the efficient allocation of credit and monitoring of borrowers will be important for improving corporate governance and productivity. To date, however, the most important task has been to satisfy the large dormant demand for financing. Talking to businesspeople, one hardly hears that lack of external financing is a major impediment to doing business in Bulgaria whereas in the late 1990's this was a major issue.

The analysis also highlights the importance of sound internal practices in the financial system. It seems that much of the improvements in the system are due to new managerial procedures and culture. In this respect, the value added of international banks and the strong supervision by the central bank are apparent.

References

Beck, Thorsten, Asli Demirgüç-Kunt, and Ross Levine. 2000. A New Database on Financial Development and Structure. *World Bank Economic Review* **14**, no. 3: 597-605.

Berger, Allen and Gregory Udell (2002). "Small Business Credit Availability and Relationship Lending: The Importance of Bank Organisational Structure," *The Economic Journal* **112**.

Diamond, D. (1984). "Financial Intermediation and Delegated Monitoring," *Review of Economic Studies* **51**, 393-414.

Drakos, K. (2002), "Imperfect competition in banking and macroeconomic performance: Evidence from the transition economies," *Bank of Finland working paper*.

Fink, G., P. Haiss, and H. C. Mantler (2005), "The Finance-Growth Nexus: Market Economies vs. Transition Countries," Working Paper, EUROPAINSTITUT, Vienna.

Fink, G., P. Haiss, and G. Vukšić (2004), "Changing Importance of Financial Sectors for Growth from Transition to Cohesion and European Integration" Working Paper, EUROPAINSTITUT, Vienna.

Jaffee, D. and M. Levonian (2001), "The structure of banking systems in developed and transition economies," *European Financial Management*, **7**(2): 161–81.

Koford, K. and A. E. Tschoegl (2003), "Foreign banks in Bulgaria, 1875-2002," *William Davidson Working Paper* Number **537**.

Koivu, T. (2002) "Do efficient banking sectors accelerate economicgrowth in transition countries?" *Bank of Finland Institute for Economies in Transition - Discussion Papers* No. **14**.

La Porta, Rafael, Florencio Lopez-de-Silanes, Andrei Shleifer, and Robert W. Vishny (1998). "Law and Finance." *Journal of Political Economy* **106**, 1113-1155

Levine, R. (2005), "Finance and Growth: Theory and Evidence," in Philippe Aghion and Steven Durlauf, eds. Handbook on Economic Growth, The Netherlands: Elsevier Science.

Mehl, A, C. Vespro, and A. Winkler, (2006), "The finance-growth nexus and financial sector environment: new evidence from Southeast Europe," *ECB-CFS Research Network working paper*.

Mehl, A. and A. Winkler (2003), "The finance-growth nexus and financial sector environment: New evidence from Southeast Europe," paper presented at The Ninth Dubrovnik Economic Conference on "Banking and the Financial Sector in Transition and Emerging Market Economies," organized by the Croatian National Bank, June, http://www.hnb.hr/dub-konf/9-konferencija-radovi/mehl-winkler.pdf

Miller, J. and S. Petranov (2001), "The Financial System in the Bulgarian Economy," *Bulgarian National Bank Discussion Papers* no. **19/2001**.

Šević, Ž. (2000), "Banking Reform in South East European Transitional Economies: An Overview," *MOCT-MOST*, **10** (3-4): 271-283.

Valev, Neven and Nikola Tasic (2007). "Finance and Growth in Bulgaria – the Effects of Credit Maturity and Credit Size across Economic Sectors," Georgia State University working paper.

INDEX

A

B

C

D

E

F

G

H

I

J

K

L

M

N

S

U

V

W

Y