

Serial correlation in high-frequency data and the link with liquidity

Susan Thomas

Tirthankar Patnaik

December 12, 2002

Abstract

This paper tests for market efficiency at high-frequencies of the Indian equity markets. We do this by testing the behaviour of serial correlation in firm stock prices using the Variance Ratio test on high frequency returns data. We find that at a frequency interval of five minutes, all the stocks show a pattern of mean-reversion. However, different stocks revert at different rates. We find that there is a correlation between the time the stock takes to revert to the mean and the liquidity of the stock on the market. Here, liquidity is measured both in terms of impact cost as well as trading intensity.

Keywords: Variance-Ratios, High Frequency Data, Market Liquidity

Contents

1	Introduction	3
2	Issues	5
2.1	Choice of frequency for the price data	6
2.2	Concatenation of prices across different days	6
2.3	Intra-day heteroskedasticity in returns	7
2.4	Measuring intra-day liquidity	7
3	Econometric strategies	8
3.1	The Variance Ratio methodology	8

3.2	Variance Ratios with HF data	10
3.3	Issues of inference	10
4	Data description	11
5	Results	13
5.1	Serial correlation in the Market Index	13
5.2	Serial correlations in individual stocks	16
6	Conclusion	21

1 Introduction

The earliest tests of the “Efficient Market Hypothesis” (EMH) have focussed on serial correlation in financial market data. The presence of significant serial correlation indicates that prices could be forecasted. This, in turn, implies that there might be opportunities for rational agents to earn abnormal profits if the forecasts were predictable after accounting for transactions costs. Under the null of the EMH though, serial correlations ought to be negligible. Most of the empirical literature on market efficiency documents that serial correlations in daily returns data are very small, which supports the hypothesis of market efficiency.

However, most of these tests are vulnerable to problems of low power. There are papers that point out that tests of serial correlation are vulnerable to low power (Summers, 1986) given that:

- There is strong heteroskedasticity in the data.
- There are changes in the characteristics of the data caused by changes in the economic environment, changes in market microstructure, the presence of economic cycles, etc.
- The changes in the DGP can also impact upon the character of heteroskedasticity as well.

Therefore, the number of observations that are typically available at the monthly, weekly or even daily levels usually lead to situations where the tests have weak power.

In the recent past, intra-day financial markets data has become available for analysis. This *high frequency data* (HF data) constitutes price and volumes data that can be observed at intervals that are as small as a second. The analysis of serial correlations with such high frequency data is particularly interesting for several reasons.

One reason is that economic agents are likely to require time in order to react to opportunities for abnormal profits that appear in the market during the trading day. While the time required for agents to react may not manifest themselves in returns observed at a horizon of a day, we may observe agents taking time to react as patterns of forecastability in intra-day high-frequency data. For example, one class of models explaining the behaviour of intra-day trade data are based on the presence of asymmetric information between informed traders and uninformed traders. When a large order hits the market, there will be temporary uncertainty about whether this is a speculative order placed by an informed trader, or a liquidity-motivated order placed by an uninformed trader. This phenomenon could generate short-horizon mean reversion in stock prices.

Another reason why HF data could prove beneficial for market efficiency studies is the sheer abundance of the data. This could yield highly powerful statistical tests of efficiency, that would be sensitive enough to reject subtle deviations from the null. However, the use of HF data also introduces problems such as those of asynchronous data, especially in cross-sectional studies. Since

different stocks trade with different intensity, using HF data is constrained to not suffer too much of a missing data problem across all the stocks in the sample.

In this paper, we study the behaviour of serial correlation of HF stock price returns from the National Stock Exchange of India (*NSE*). We use variance ratio (*VR*) tests which was first applied to financial data in Nelson and Plosser (1982). We study both the returns of the S & P CNX Nifty (*Nifty*) market index of the NSE, and to a set of 100 stocks that trade on the NSE. We choose those stocks which are the most liquid stocks in India and which, therefore, have the least probability of missing data even at a very small intervals.

The literature leads us to expect *positive deviations from the mean* in the index returns (Atchison et al., 1987) and *negative serial correlations* in stocks (Roll, 1984). The positive correlation in index returns is attributed to the asynchronous trading of the constituent index stocks: information shocks would first impact on the prices of stocks that are more liquid (and therefore, more actively traded) and impact on price of less active stocks with a lag. When the index returns are studied at very short time intervals, this effect ought to be even more severe as compared to the correlations in daily data with the positive serial correlations perhaps continuously growing for a period of time before returning to the mean (Low and Muthuswamy, 1996). The negative serial correlations in stock returns is attributed to the “bid-ask bounce”: here, the probability of a trade executing at the bid price being followed by a trade executing at the ask price is higher than a trade at the bid followed by another trade at the bid.

However, we find that there is no significant evidence of serial correlations in the index returns, even at an interval of five minutes. In fact, the *VR* at a lag of one is non-zero and *negative*. One implication is that there is no asynchronous trading within the constituent stocks of the index at a five-minute interval. Either we need to examine index returns at higher frequencies, in order to find evidence of asynchronous trading in the index or find an alternative factor to counter the positive correlations expected in a portfolio’s returns.

Our results on the *VR* behaviour of individual stocks is more consistent with the literature. All the 100 stocks show significant negative deviations from the mean at a lag of five minutes. This means that stock returns at five minute intervals do have temporal dependence. However, there is a strong heterogeneity in the behaviour of the serial correlation across the 100 stocks. While the shortest time to mean reversion is ten minutes, the longest is across several days!

While Roll (1984) shows how the liquidity measure of the bid-ask bounce can lead to negative serial correlations in price, Hasbrouck (1991) shows that the smaller the bid-ask spread of this stock, the smaller is the impact of a trade of a given size, which would lead to a smaller observed correlation in returns. Therefore, an illiquid stock with the same depth but larger spread would suffer a larger serial correlation at the same lag. Thus, liquidity could have an impact not only on the *sign* of the serial correlation but also its magnitude. This, in turn, could affect the rate at which the *VR* reverts back to the mean.

The hypothesis is that the cross-sectional differences in serial correlation across stocks is driven by the cross-sectional differences in their liquidity. We base our finding of heterogeneity in mean reversion as a test of this hypothesis.

The NSE disseminates the full limit order book information available for all listed stocks, observed at four times during the day. We construct the spread and the *impact cost* of a trade size of Rs.10,000 at each of these times for each of the stocks. We use the average impact cost as a measure of intra-day liquidity of the stocks. We then construct deciles of the stocks by their liquidity, and examine the average VRs observed for each decile.

We find that the top decile by liquidity (ie, the most liquid stocks) have the smallest deviation from mean (in magnitude at the first lag). This decile also has the shortest time to mean reversion in VRs. The bottom decile by liquidity have the largest deviation from mean at the first lag as well as the longest time to mean reversion. The pattern of increasing time to mean reversion is consistently observed as correlated with decreasing liquidity as measured by the impact cost. Thus, we find that there is a strong correlation between mean reversion in returns of stocks and their liquidity.

The paper is organised as follows: Section 2 presents the issues in analysing intra-data of returns as well as liquidity. In Section 3 focusses on developments in the VR methodology since Nelson and Plosser (1982), and the method of inference we employ in this paper. We describe the dataset in Section 4. Section 5 presents the results, and we conclude our findings in Section 6.

2 Issues

High frequency finance is a relatively new field (Goodhart and O'Hara (1997), Dacorogna et al. (2001)). The first HF data that became available was the time series of every single traded price on the New York Stock Exchange. The first studies of HF data were based on HF data from foreign exchange markets made available by *Olsen and Associates*. Studies based on these datasets were the first to document time series patterns of intra-day returns and volatility. Wood et al. (1985), McNish (1993), Harris (1986), Lockwood and Linn (1990) were some of the first studies of the NYSE data, while Goodhart and Figliuoli (1991) and Guillaume et al. (1994) are some of the first papers on the foreign exchange markets. Since then, there have been several papers that have worked on further characterising the behaviour of intra-day prices, returns and volumes of financial market data (Goodhart and O'Hara (1997), Baillie and Bollerslev (1990), Gavridis (1998), Dunis (1996), ap Gwilym et al. (1999)).

These papers raised several issues for consideration about peculiarities of HF data that are not present in the frequencies that have been traditionally analysed such as daily or weekly data. The two issues that we had to deal with at the outset of our analysis of the Indian HF data were:

1. The issue of the frequency at which to analyse the data.

2. How to concatenate returns across days.
3. The problem of higher volumes/volatility at the start and at the end of every trading day.

2.1 Choice of frequency for the price data

Data observed from financial markets in real-time have the inherent problem of being irregular in traded frequency (Granger and Ding, 1994). Part of the problem arises because of the quality of published data - trading systems at most exchanges can execute trades at finer intervals than the frequency at which exchanges record and publish data. For example, the trading system at the NSE can do trades at intervals of $1/64^{th}$ of a second. However, the data is published with timestamps at the smallest interval of a second. There could be multiple trades within the same timestamp.

On the other hand, some stocks do trade with the intensity of several trades within a second, whereas other may do just one. Thus, the notion of the “last traded price” (LTP) for two different stocks might actually mean prices at two different real times, which leads to comparing asynchronous or *irregular* data.¹

However, when we study the cross-sectional behaviour of (say) serial correlation of stock prices, there is a need to synchronise the price data that is being studied. There are several methods of synchronising irregular data that is used in the literature. One approach is to model the time series of every stock directly in trade-time, rather than in clock or calendar time (Marinelli et al., 2001). Another approach is of Dacorogna et al. (2001), who follow a time-deformation technique called ν -time. Here, they model the time series as a subordinated process of another time series.

We follow the approach of Andersen and Bollerslev (1997) who impose a discrete-time grid on the data. In this approach, the key parameter of choice is the width of the grid. If the width of the interval is too high, then information about the temporal pattern in the returns may be lost. On the other hand, a thin interval may lead to high incidence of “non-trading” which is associated with spurious serial-correlation in the returns. Therefore, the grid interval must be chosen to minimize the informational loss while avoiding the problem of spurious autocorrelation.

2.2 Concatenation of prices across different days

When intraday data is concatenated across days, the first return on any day is not really an intraday return from the previous time period, but an *overnight* return. This is asynchronous with the inter-

¹For example, stock A at the NSE trades four times within the second, with trades at the 1st, 2nd, 3rd and 15th 64^{th} interval of a second. Stock B trades once with a trade at the 63rd interval. Then the LTP recorded for A is at 15/ 64^{th} of a second and B at 63/ 64^{th} of a second which is asynchronous in real time.

vals implied in other returns data. Returns over differing periods can lead to temporal aggregation problems in the data, and result in spurious autocorrelation.

This is analogous to the practice of ignoring weekends when concatenating daily data. However, the same cannot be done for the intraday data. In the case of daily data, dropping the weekend means that every Monday is in reality a three-day return rather than a one-day return. However, the interval between the last return of a day and the first return on the next day will typically mean a much larger number of intervals in the intervening period.² Overnight returns have quite different properties as compared to intraday returns (Harris, 1986).

(Andersen and Bollerslev, 1997) solve this problem by dropping the first observation of every trading day .

2.3 Intra-day heteroskedasticity in returns

Another issue that arises when analysing the serial correlations in the data are the U-shaped patterns in volumes and volatility. It is observed that returns in the beginning and the end of the trading day tend to be different when compared with data from the mid-period of the trading day. This manifests itself in a U-shaped pattern of volatility. These patterns have been documented for markets over the world (Wood et al. (1985); Stoll and Whaley (1990); Lockwood and Linn (1990); McNish and Wood (1990b,a, 1991, 1992); Andersen and Bollerslev (1997)).

If there are strong intraday seasonalities in HF data, then this could cause problems with our inference of the serial correlations of returns. Andersen et al. (2001) regress the data using it's *Fourier Flexible Form* and analyse the behaviour of the residuals, rather than the raw data itself.

2.4 Measuring intra-day liquidity

Kyle (1985) characterises three aspects of liquidity: the *spread*, the *depth* of the limit order book (LOB) and the *resiliency* with which it reverts to its original level of liquidity. Papers from the market microstructure literature have analysed the link between measures of liquidity and price changes. These start from Roll (1984) who established a link between the sign of the serial correlation to the bid-ask spread to Hasbrouck (1991) who established the depth of the LOB to the magnitude of the serial correlation and ? who attempts to develop models of asymmetry of information to the resiliency of the LOB.

²For example, if we are using returns on a five minute interval, and the market closes at 4pm and opens at 10am, the interval between the last return on one day and the opening return of the next day would mean a gap of 216 intervals in between!

Typically, theory models change in liquidity measures as arising out of trades by informed or uninformed trades and the impact these trades on price changes. While there is no clear consensus from empirical tests on the precise role of information and the path through which it can impact upon prices through liquidity changes, there is consensus on the positive link between the effect of liquidity upon serial correlation in stock returns (Hasbrouck (1991), Dufour and Engle (2000)).

The traditional metric used for real-time data on liquidity is the bid-ask spread that is available for many markets along with the traded price. Our dataset does not have real-time data on bid-ask spreads. Instead, we have access to two measures that capture liquidity: *trading intensity* and *impact cost*.

- We define trading intensity as the number of trades in a given time interval of the trading day. It fluctuates in real-time and can either be measured either in value or in number of shares.
- Impact cost is the estimated cost of transacting a fixed value in either buying or selling a stock. It is measured as the actual price paid (or received) with respect to the “fair price” of the stock, which is measured by $bid + ask/2$. The impact cost is calculated as

$$\text{actual price}/0.5 * (\text{bid} + \text{ask})$$

This measure is like the bid-ask spread but is a standardised measure which makes it directly comparable across different stocks trading at different price levels. The impact cost is a function of the size of the trade. It will depend upon the depth of the LOB and is a measure of the available liquidity of the stock. The impact cost is measured in basis points.

Both these liquidity measures are visible for an open electronic LOB. While the trading intensity captures the liquidity that has taken place thus far, the impact cost is a predictive measure of liquidity. We will use both of these measures to test for the correlation between liquidity and serial correlations in HF data.

3 Econometric strategies

3.1 The Variance Ratio methodology

VRs have been often used to test for serial correlations in stock market prices (Lo and MacKinlay (1988), Poterba and Summers (1988)). The VR statistic measures the serial correlation over q period, as given by:

$$VR(q) = \frac{Var[r_t(q)]}{Var[r_t]}, \quad \text{where } r_t(q) = \sum_{k=1}^q r_t \quad (1)$$

$$= 1 + \sum_{k=1}^{q-1} \left(1 - \frac{k}{q}\right) \rho_k \quad (2)$$

Here $r_t(q)$ is the q -period return, and ρ_k is the k -period autocorrelation. For a random walk, $\rho_k \equiv 0 \quad \forall k$. Hence the null of market efficiency is defined as,

$$VR(q) \equiv 1$$

which is what the value of the VR ought to be for a random walk.

The VR test has more power than the other tests for randomness, such as the Ljung-Box and the Dickey-Fuller test (Lo and MacKinlay, 1989). It does not require the normality assumption, and is quite robust. Lo and MacKinlay (1988) derived the sampling distribution of the VR test statistic under the null of a simple homoscedastic Random Walk (RW), and a more general uncorrelated but possible heteroscedastic RW.³ The empirical literature on testing serial correlations in financial data have largely used the heteroskedasticity consistent Lo and MacKinlay (1988) form.

There have been several papers extending and improving the tests of the VR: Chow and Denning (1993) extended the original form to jointly test multiple VRs, Cecchetti and sang Lam (1994) proposed a MonteCarlo method for exact inference when using a joint test of multiple VRs for small samples, Wright (2000) proposed exact tests of VRs based on the ranks and signs of time series, Pan et al. (1997) use a bootstrap scheme to obtain an empirical distribution of the VR at each q .

One of the divides in the literature is on calculating VRs using overlapping data vs. non-overlapping data. In the former, the aggregation of data over q periods is done using overlapping windows of length q , whereas in the latter the data is aggregated over windows of data that do not overlap. If the time series is short, then overlapping windows re-use old data to give a larger number of points available to calculate the VR at q . However, the efficiency of this form of the VR estimator is lower.

Richardson and Stock (1989) develop the theoretical distribution of both the overlapping and the non-overlapping VR statistic: they show that the limiting distribution in the case of the overlapping statistic is a chi-squared distribution that is robust to heteroskedasticity and non-normality. The distribution of the non-overlapping statistic is a functional of a Brownian Motion and can only be estimated using Monte Carlo. Richardson and Smith (1991) find that overlapping VRs have 22% higher standard errors compared with non-overlapping VRs.

³For more details on the various types of the RW that can be tested, see Campbell et al. (1997, Page. 28).

3.2 Variance Ratios with HF data

One of the earliest papers on the use of HF data in VR studies is Low and Muthuswamy (1996). They tested the serial correlations in quotes from three foreign exchange markets: USD/JPY, DEM/JPY, and USD/DEM for the period October 1992 to September 1993. They calculate variance ratios for these FX returns where the holding period (aggregation) ranges from 5 minutes till 3525 minutes (corresponding to 750 intervals).

They find that the VR do not immediately mean-revert - that negative correlation in the returns becomes stronger as the holding period increases. Variance Ratio (VR)s are found to grow faster in the very short-run *i.e.*, less than 200 minutes. This leads them to suggest that “serial dependencies are stronger in the long-run”.

The use of the VR in the HF data literature has been relatively limited so far. But the existing work highlight one interesting new facet of the issue of using over-lapping versus non-over-lapping data in VR studies when applied to HF data:

- The abundance of HF data and the lower efficiency of overlapping VR would imply that we should use non-overlapping VRs to test for serial correlation in HF data.
- Andersen et al. (2001) study the shift in volatility patterns in HF data. They use non-overlapping returns to analyse the behaviour of the VR in this data. They show that the standard VR tests could be seriously misleading when applied on intraday data, due to the inherent intraday periodicity present. Thus, they suggest applying a Fourier Flexible Form (FFF) Gallant (1981) when analyzing such data.

3.3 Issues of inference

Lo and MacKinlay (1988) developed the heteroskedasticity-consistent overlapping VR statistic. If the sample of HF prices has T observations and we need to calculate the VR at an aggregation of q , then:

$$\begin{aligned}\hat{\mu} &= \frac{1}{nq} \sum_{k=1}^{nq} (p_k - p_{k-1}) \\ \sigma_a^2 &= \frac{1}{nq-1} \sum_{k=1}^{nq} (p_k - p_{k-1} - \hat{\mu})^2 \\ \sigma_b^2(q) &= \frac{1}{m} \sum_{k=q}^{nq} (p_{qk} - p_{qk-q} - q\hat{\mu})^2\end{aligned}$$

$$\begin{aligned}
\sigma_c^2(q) &= \frac{1}{m} \sum_{k=q}^{nq} (p_k - p_{k-q} - q\hat{\mu})^2 \\
m &= q(nq - q + 1) \left(1 - \frac{q}{nq}\right) \\
\tilde{V}R(q) &= \frac{\sigma_b^2(q)}{\sigma_a^2} \\
\overline{V}R(q) &= \frac{\sigma_c^2(q)}{\sigma_a^2}
\end{aligned}$$

Here, $\tilde{V}R(q)$ is the non-overlapping and $\overline{V}R(q)$ is the overlapping variance-ratio statistic. Under the normality assumption for returns with heteroscedasticity in the error terms, the asymptotic distribution of $\overline{V}R(q)$ is:

$$\begin{aligned}
\hat{\delta}_k &= \frac{nq \sum_{j=k+1}^{nq} (p_j - p_{j-1} - \hat{\mu})^2 (p_{j-k} - p_{k-1} - \hat{\mu})^2}{\left[\sum_{j=1}^{nq} (p_j - p_{j-1} - \hat{\mu})^2 \right]^2} \\
\hat{\theta}(q) &= 4 \sum_{k=1}^{q-1} \left(1 - \frac{k}{q}\right)^2 \hat{\delta}_k \\
\sqrt{nq}(\overline{V}R(q) - 1) &\sim N(0, \theta).
\end{aligned}$$

4 Data description

We analyse serial correlations of both the Nifty as well as for the constituent stocks of the index. The data used in this study comprises all trades in the Capital Market (NSE CM) segment of the National Stock Exchange in the period Mar 1999-Feb 2001.⁴

The trading system on the NSE is a continuous open electronic limit-order book market, with order-matching done on price-time priority. Trading starts every day at 0955 in the morning, and continues without break till 1530. The NSE is one of the most heavily traded stock exchanges in the world where more than 490,000 trades take place daily, on average.

The selection of our period of study has some market microstructure issues to consider. Unlike exchanges all over the world, the NSE had in place “price limits” on trading of all stocks. Though it initially started without any price limits, NSE shifted to a $\pm 10\%$ band on all stocks in November 1995. On 11 September 1996, all stocks were classified into three different categories, based on their historical volatility, with wider bands for the more volatile stocks. On 10 Sep 1997, stocks

⁴The source of this data is a monthly CD that is disseminated by the NSE called the “NSE Release A CD”.

were again classified into three categories, this time based on liquidity, with wider bands for the more liquid stocks. A series of changes in the price-limits of stocks have followed since then. In March 2001, the exchange changed over to a rolling settlement system. Further, there have been a number of changes in the trading-times of the exchange over the years. For the period under study, the exchange traded in the interval 0955:1530.

We select March 1999 to February 2001 as our period of study because this period has the least microstructural changes to the price discovery on the exchange. A total of 1384 stocks traded in this period. The data set has about 253,717,939 records in 514 days. These are “time and sales” data, (as defined in MacGregor (1999)).

We deal with the issues raised in Section 2 in the following manner:

Selection of data frequency We have chosen a 300 second interval as the base frequency for this grid. Our choice is based on various tests and diagnostic measures (Patnaik and Shah, 2002). The 300s discretization gives about 67 data points per day. On average, we observe that at a 300 second interval, there is an incidence of 10% missing observations.

In addition, the data were filtered of outliers, and anomalous observations. An outlier observation results when a trade is recorded for a stock, outside its price band. Anomalous observations could be caused when the position of the decimal point is misplaced while recording the data. Errors range from incorrect recording, to human errors. A number of papers have dealt with the issues of cleaning up and filtering of intraday data (Dacorogna et al., 2001). The specifics of data-filtering and cleaning for HF data from the NSE can be found in Patnaik and Shah (2002).

Concatenation of data across different days We follow the solution adopted by Andersen and Bollerslev (1997) of ignoring the first return in the day, and *then* concatenating the data.

Trading intensity NSE has one of the highest trading intensities in the world. In this period studied here, the mean trading intensity was about 24.56 trades per second.⁵ For the study, we chose the hundred most traded stocks in the period. On an average, each of these stocks traded about 4211 times a day, about one trade in every 5 seconds. These 100 stocks comprise about 83.32% of all the trades recorded in this period at the NSE.

Calculation of intra-day impact costs The impact cost for a stock is calculated using order-book snapshots of the market which are recorded four times a day by the National Stock Exchange (NSE), at 1100, 1200, 1300, and 1400 respectively. The complete Market By Price (MBP)⁶

⁵253717939 trades over 514 days, with 20100 trading seconds in a day, gives 24.56.

⁶This is the set of all the limit orders that are available in the market at that time, both on the buy and on the sell side. Since the limit order shows both the price and the quantity, it is easy to calculate the impact cost for any stock given a certain size of the trade. Given the MBP we can calculate the graph of the impact cost at all trade sizes for a given stock. This graph is called the liquidity supply schedule and is always empirically different for the buy and the sell side of the LOB.

is available at these times.

For our study, we calculate the buy/sell impact cost for an order of Rs. 10,000 for each of the 1382 stocks that traded in the sampling period. The calculation was done for the LOB data for each of the four time points for every traded day in this period. We then selected the 100 most liquid stocks on the basis of mean buy and sell impact cost.⁷

We show the top and bottom ten stocks by liquidity by both the impact cost and the trading intensity measures in Table 1. We can see that the most liquid stock is RELIANCE, with an impact cost of 7 basis points for a transaction size of Rs.10,000, and the least liquid stock is CORPBANK with a median impact cost of 26 basis points for the same transaction size.

There appears also some amount of difference in the ordering by the two liquidity measures. However, at level of quartiles of stocks, the differences are not significant.

5 Results

We estimate the overlapping VRs for the NSE-50 index as well as all the 100 stocks in our sample. We depict the results as a set of two graphs for each of the returns:

1. The first graph shows the VRs themselves starting from an aggregation of two (which is the serial correlation for returns at ten minute intervals) and continues upto an aggregation of 350 lags (which is the serial correlation of returns at one trading week or five days).

The null that we test is that if the returns are truly random, then the VRs should be not significantly different from a value of one. The graph shows two sets of confidence intervals, the inner intervals are for the 95% band and the outer ones are for the 99% bands.

2. The second graph shows the non-overlapping heteroskedasticity consistent VR statistic. Once again, the statistic is plotted for aggregation levels from two to 350. The null implies that the statistic should have a value of 0. Both the 95% and the 99% confidence intervals are drawn around the statistics.

The graphs for the NSE-50 index and the stocks are shown below.

5.1 Serial correlation in the Market Index

We find that the index shows no pattern of significant serial correlations even at the five-minute interval as can be seen in Figure 1.

⁷The list of all the 100 stocks is in the appendix.

Table 1 Top and bottom ten stocks by the two liquidity measures

The liquidity measures have been applied for the period of March 1999-February 2001 for data from the NSE. The table shows the top most liquid stocks at the NSE, and the bottom least liquid stocks, amongst the top 100 stocks by liquidity.

The trading intensity is the total number of trades in the entire period.

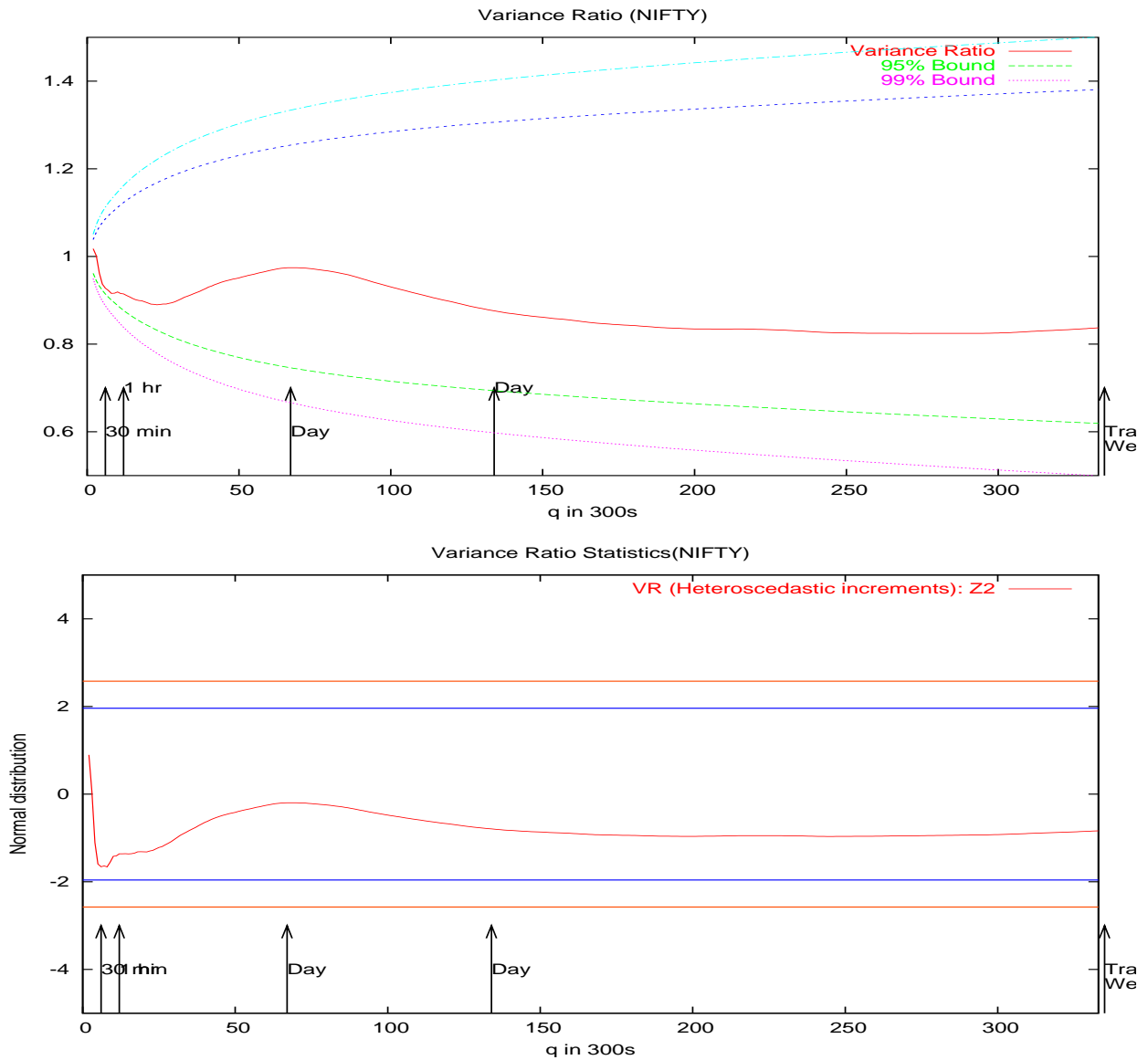
The impact cost figures are calculated in percent and is based on a order size of Rs.10,000.

Stock	Trading Intensity	Stock	Mean Impact cost
SATYAMCOMP	17971195	RELIANCE	0.07
ZEETELE	14746798	INFOSYSTCH	0.08
HIMACHLFUT	12609773	HINDLEVER	0.09
SILVERLINE	10987497	ITC	0.09
PENTSFWARE	10606971	SBIN	0.09
GLOBALTELE	10415715	ACC	0.11
SQRDSFWARE	9982348	L&T	0.11
RELIANCE	9204846	RANBAXY	0.11
ROLTA	7125411	SATYAMCOMP	0.11
INFOSYSTCH	7024350	MTNL	0.12
BOMDYEING	311934	ORCHIDCHEM	0.25
GOLDTECHNO	311070	ORIENTINFO	0.25
BEL	307093	PHILIPS	0.25
STERLITIND	303246	POLARIS	0.25
ORCHIDCHEM	302334	PUNJABTRAC	0.25
POLARIS	299354	SMITHKLPHA	0.25
ORIENTBANK	295265	TATAPOWER	0.25
LML	294189	WOCKPHARMA	0.25
BANKBARODA	293264	AUROPHARMA	0.26
DABUR	290017	CORPBANK	0.26

Figure 1 VRs and Heteroskedasticity-consistent standardized VR statistics for Nifty.

The first graph is that of variance ratios for Nifty over one week, for the period March 1999 to February 2001. Nifty index values are discretised at 300s intervals.

The second graph is that of heteroskedasticity-consistent standardized variance ratios statistics for Nifty for the same period.



This is contrary to what is documented in the literature where stock market indexes show “positive” correlations even in studies based on daily data. In this case, the very first VR is slightly above one, but the next three values are negative.

This is an interesting paradox to the typical rationale that is given to explain the behaviour of intra-day index returns, which is the asynchronous trading of the constituent stocks. If the NSE-50 index shows negative correlations at the five-minute intervals, that could mean that there is so much negative correlation that it beats the positive aspect we expect from asynchronous trading.

The other interesting aspect is that the Z-stats are significant if we ignore heteroskedasticity but are insignificant when we take heteroskedasticity into account. We infer therefore that the serial correlation patterns in the VR values are being driven largely by the intra-day U-shaped pattern of volatility. However, there does not appear to be any serial correlations net of these intra-day patterns.

5.2 Serial correlations in individual stocks

We examine the graphs of the VRs for individual stocks. Here, we find that the results are more consistent with the literature. All the individual stocks VRs show a negative correlation in returns. Some stocks (such as RELIANCE and ZEETELE) have VRs that are negative and significant at aggregations of two or three (which are returns at ten or fifteen minutes), but not significant after that. Others have VRs that are significantly lower than one even beyond aggregations of 350 (which are returns at a trading week!).

If we organise the graphs in order of stocks with decreasing liquidity, we find that there appears to be a relation between the time taken for the VRs to mean-revert and the liquidity of the stock. This is independent of which liquidity metric we use as we can see in the graphs below:

In order to better understand the link between the liquidity of the stock and significant deviations in the VR values, we organise the stocks into deciles by liquidity. The summary statistics of the liquidity characteristics of the deciles are in Table 2. We note that the highest heterogeneity in the liquidity characteristics are in the top decile. By the time we reach the bottom deciles, the liquidity characteristics are much more homogenous.

If liquidity is a factor that affects the serial correlation of the stock, then the behaviour of the average VRs for the first decile stocks should be different compared with the second or the last decile. Since high liquidity causes smaller values of correlations in stock returns, we would expect that the mean reversion pattern of the first decile stocks (most liquid) should show the most rapid reversion to the mean. The last decile should show the least rapid reversion to the mean. Since we are looking at the average of the VRs for the 10 stocks in a decile, we should expect that the VRs should show negative deviations in all cases.

Figure 2 VR and Heteroskedasticity-consistent standardized VR statistics for the top two stocks by trading intensity.

These are the variance ratios and heteroskedasticity consistent standardised VR statistics for the top two stocks by trading intensity: Satyam Computers and Zeetele for the period from March 1999 to February 2001.

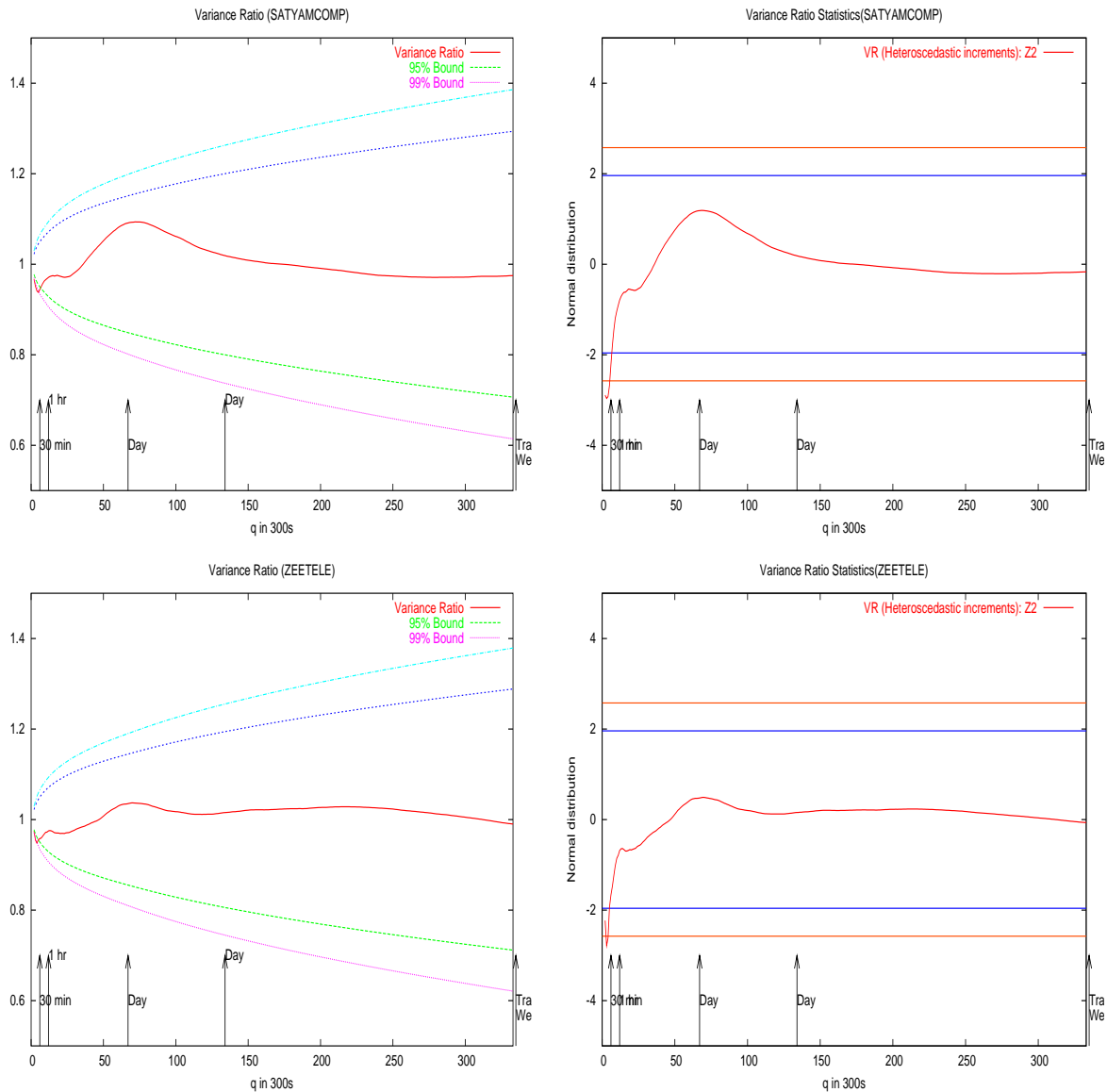


Figure 3 VR and Heteroskedasticity-consistent standardized VR statistics for the bottom two stocks by trading intensity.

These are the variance ratios and heteroskedasticity consistent standardised VR statistics for the bottom two stocks by trading intensity: Bank of Baroda and Dabur Industries for the period from March 1999 to February 2001.

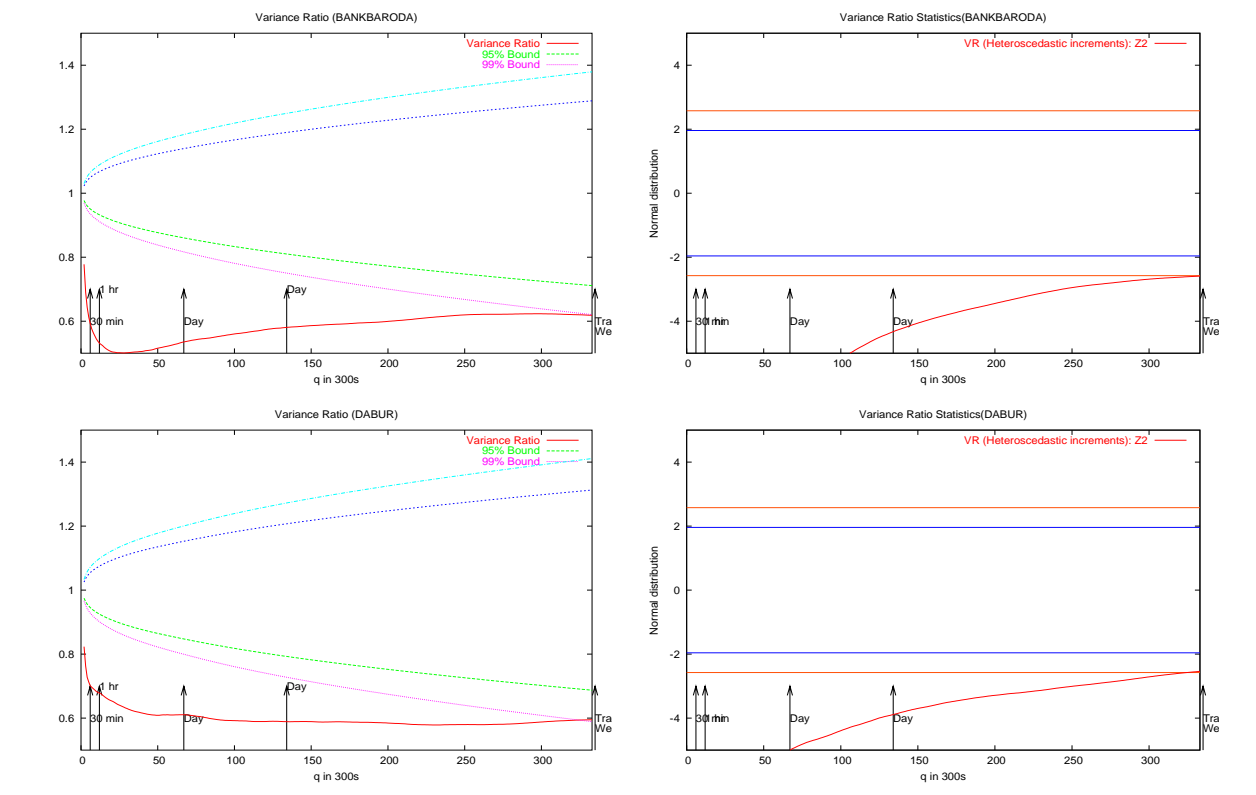


Figure 4 Variance Ratios and Heteroskedasticity-consistent standardized VR statistics for the top two stocks by impact cost.

These are the variance ratios and heteroskedasticity consistent standardised VR statistics for the top two stocks by impact cost: Reliance Industries and Infosys Technologies for the period from March 1999 to February 2001.

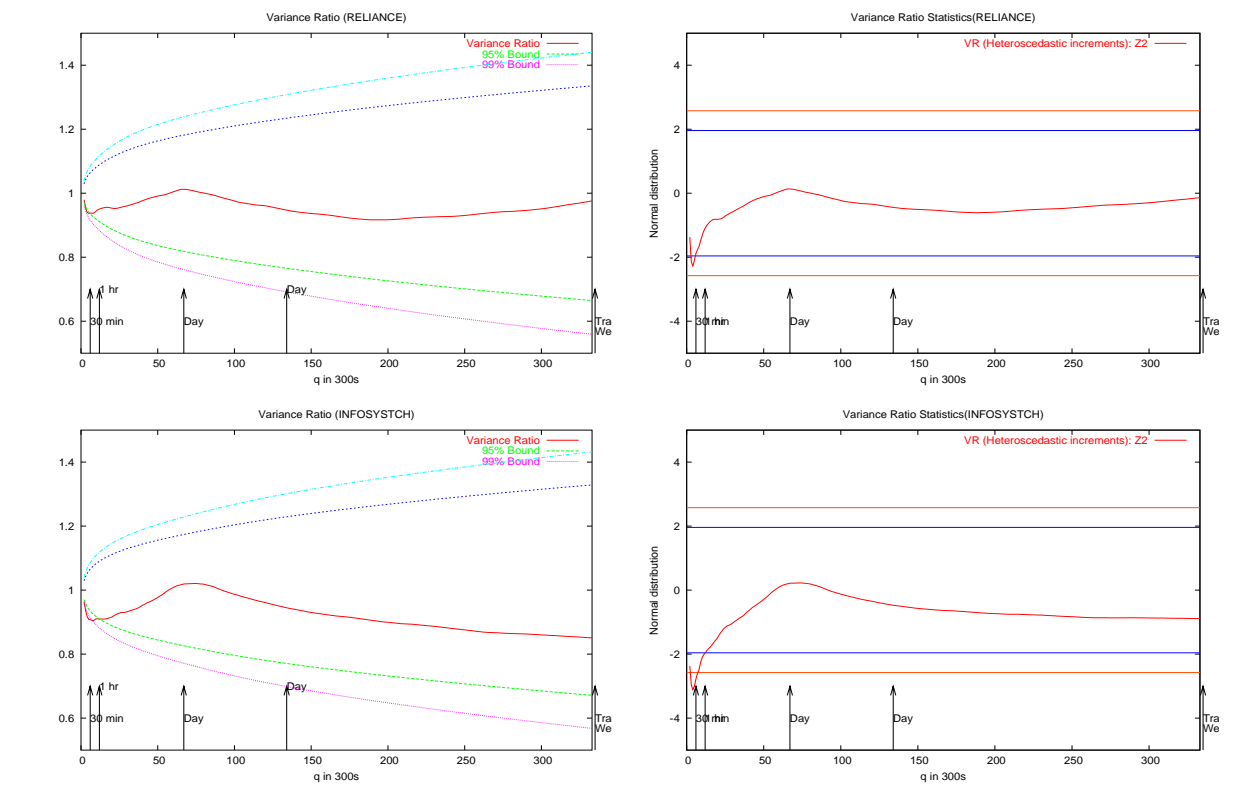


Figure 5 Variance Ratios and Heteroskedasticity-consistent standardized VR statistics for the bottom two stocks by impact cost.

These are the variance ratios and heteroskedasticity consistent standardised VR statistics for the top two stocks by impact cost: Auro Pharmaceuticals and Corporation Bank for the period from March 1999 to February 2001.

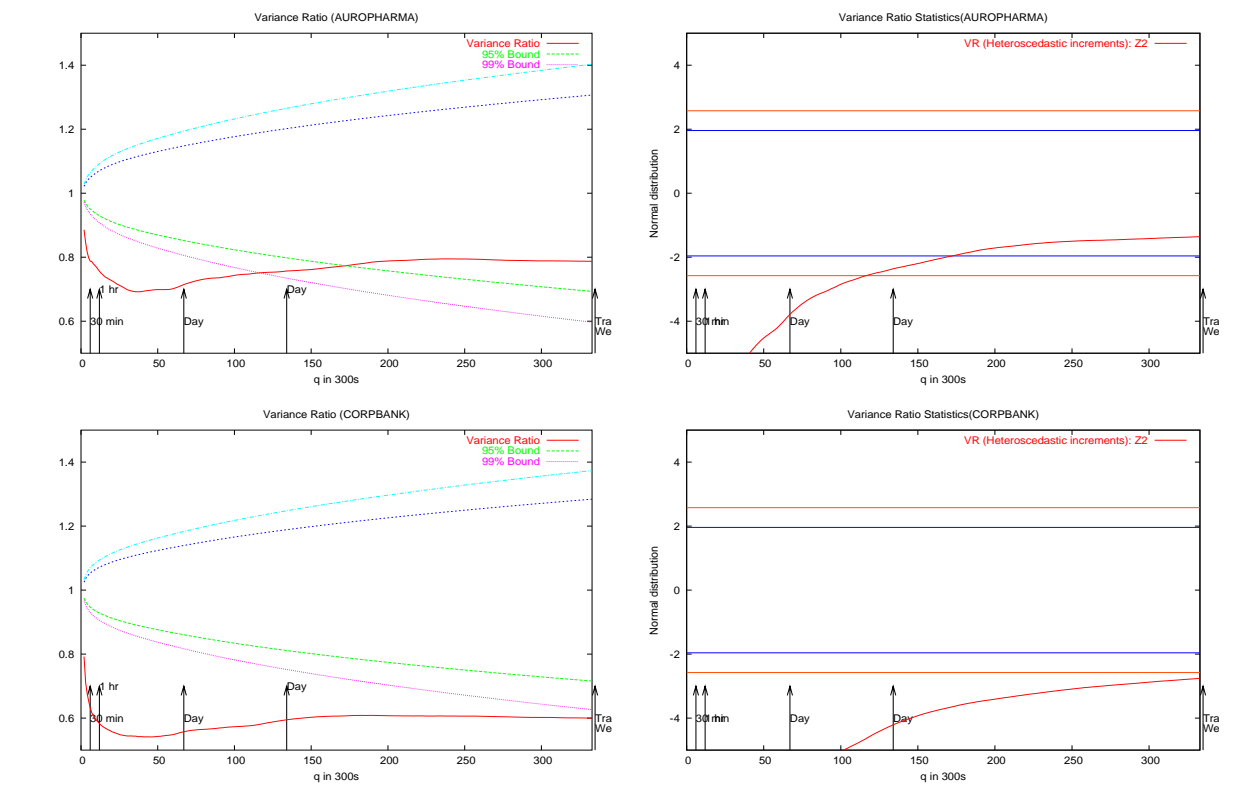


Table 2 Liquidity characteristics for stock deciles.

Deciles of the 100 stocks are selected on the basis of impact costs for an order of Rs.10,000. The data used here is using the four snapshots of the LOB every trading day in the period from Mar 1999-Feb 2001.

Decile	Mean Trades	Mean IC	Minimum	Maximum
1	12323.29	0.098	0.07	0.12
2	11878.61	0.131	0.12	0.14
3	6615.22	0.144	0.14	0.15
4	4587.33	0.158	0.15	0.17
5	1783.87	0.174	0.17	0.18
6	2115.46	0.191	0.18	0.21
7	1146.91	0.219	0.21	0.23
8	594.96	0.231	0.23	0.24
9	904.73	0.245	0.24	0.25
10	666.27	0.252	0.25	0.26

The behaviour of the VR averaged for each decile are presented in Figure 6 below. We see that there is a fall off in the time to mean reversion as we move from the average VR of the top decile to the bottom decile by liquidity. Furthermore, the fall off is consistent: the top decile has the shortest time to mean reversion, the next decile has a longer time, and so on, till the last decile with the longest time to mean reversion.

Therefore, liquidity can be used as a factor understand the serial correlations in stock returns. We see that at the level of individual stocks, the impact of liquidity differences is a larger deviation from the mean in the VR tests, and a longer time to revert to the mean.

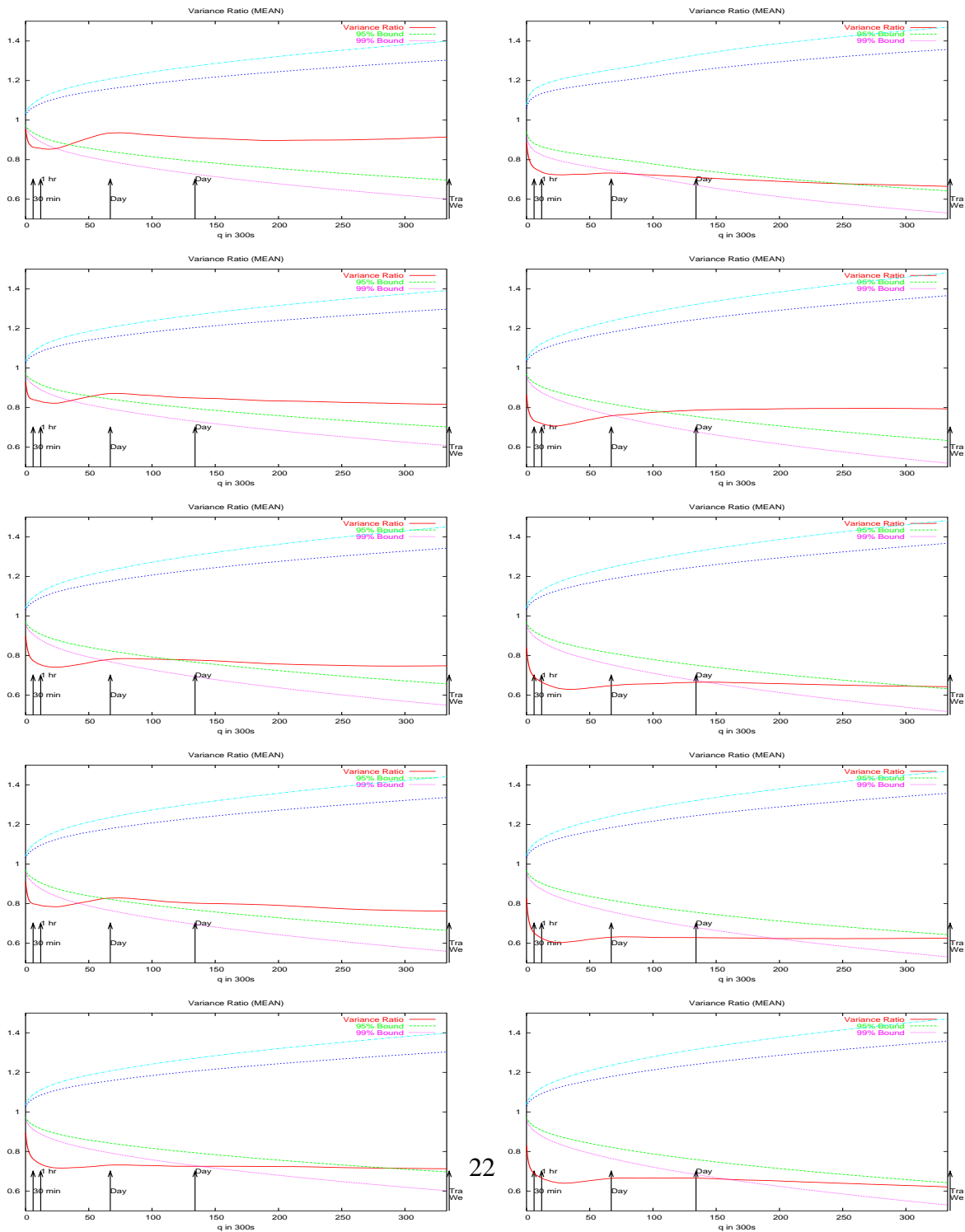
6 Conclusion

High frequency data in finance has been gaining the spotlight in the last decade of empirical work in finance. On the one hand, the abundance of this data helps to eliminate the problems of weak power of tests of market efficiency. On the other hand, the data has to be treated with caution since, very high frequency observations bring with it new factors that introduce noise for analysis.

In this paper, we have used high–frequency price data in a variance ratio framework to test patterns of serial correlation in returns from the Indian stock markets. We apply these tests to both the index and to individual stocks. We find that our results on high frequency index returns are unlike the results from other markets in the world with no significant correlations even at ten minute intervals. However, individual stocks show negative correlations at intervals of thirty minutes which is in

Figure 6 Variance Ratios for liquidity deciles of stocks.

These are the variance ratios for the deciles of stocks from the set of 100 stocks. The deciles were created by calculating the average impact cost of each stock for a trade size of Rs.10,000 over all the trading days from March 1999 to February 2001. The set of all stocks were sorted by increasing impact cost and deciles were created from this sorted list.



accordance with the literature from other markets.

One aspect of the individual stock correlations is the high degree of heterogeneity of mean reversion that we see across the set of 100 stocks. We analyse the heterogeneity in serial correlations in terms of heterogeneity of liquidity and find there is substantial evidence of a link between the serial correlation and the liquidity in stock returns. We find that the average variance ratios across a decile of stocks with the best liquidity reverts to mean at a much more rapid rate (thirty minutes) as compared with that observed for the decile of the least liquid stocks.

Thus, there appears to be a close link from liquidity of a stock and the efficiency of its market price.

One possible extension to this paper may be to ask the question of whether we can design an arbitrage strategy that can profit from this liquidity factor in market efficiency. One possible design might be to use data upto a period and identify sets from highly traded stocks that differ in liquidity in terms of their impact cost. Once these stocks have been identified and sorted, would an arbitrage strategy played out over a period of two weeks, of creating a portfolio of long the less liquid stocks and short the more liquid stocks result in arbitrage profits? Since we are going long and short a set of stocks, we are likely to be hedged against losses from market index movements. What remains would only be the lagged adjustment of prices in the less liquid stocks with respect to the more liquid ones.

Alternatively, such a strategy can be applied to identify “pairs of stocks” and an arbitrage strategy put into place over historical data to see if there is arbitrage profits to be made from this.

References

- Andersen, T., Bollerslev, T., and Das, A. (2001). Variance-ratio statistics and high-frequency data: Testing for changes in intraday volatility patterns. *Journal of Finance*, LVI(1):305–327.
- Andersen, T. G. and Bollerslev, T. (1997). Intraday seasonality and volatility persistence in financial markets. *Journal of Empirical Finance*, 4:115–58.
- ap Gwilym, O., Buckle, M., and Thomas, S. (1999). The intra-day behaviour of key market variables for liffe derivatives. chapter 6, pages 151–189.
- Atchison, M. D., Butler, K. C., and Simonds, R. R. (1987). Nonsynchronous security trading and market index autocorrelation. *Journal of Finance*, 42(1):111–119.
- Baillie, R. and Bollerslev, T. (1990). Intraday and intermarket volatility in foreign exchange markets. *Review of Economic Studies*, 58:565–585.
- Campbell, J. Y., Lo, A. W., and MacKinlay, A. C. (1997). *The Econometrics of Financial Markets*. Princeton University Press.
- Cecchetti, S. G. and sang Lam, P. (1994). Variance-ratio tests: Small-sample properties with an application to international output data. *Journal of Business Economics & Statistics*, 12(2):177–186.
- Chow, K. V. and Denning, K. C. (1993). A simple multiple variance ratio test. *Journal of Econometrics*, 58:385–401.
- Dacorogna, M. M., Gencay, R., Muller, U. A., Olsen, R. B., and Pictet, O. V. (2001). *An Introduction to High-Frequency Finance*. Academic Press, 525 B Street, Suite 1900, San Diego California 92101-4495.
- Dufour, A. and Engle, R. (2000). Time and price impact of stock trades. *Journal of Finance*, *Forthcoming*.
- Dunis, C., editor (1996). *Forecasting Financial Markets: Exchange Rates, Interest Rates and Asset Management*. Wiley Series in Financial Economics and Quantitative Analysis. John Wiley & Sons, Baffins Lane, Chichester, West Sussex PO19 1UD, England.
- Gallant, A. R. (1981). On the bias in the flexible functional forms and an essentially unbiased form: The fourier flexible form. *Journal of Econometrics*, 15:211–245.
- Gavridis, M. (1998). Modelling with high frequency data: A growing interest for financial economists and fund managers. chapter 1, pages 3–22.
- Goodhart, C. A. E. and Figliuoli, L. (1991). Every minute counts in financial markets. *Journal of International Money and Finance*, 10:23–52.

- Goodhart, C. A. E. and O'Hara, M. (1997). High frequency data in financial markets: Issues and applications. *Journal of Empirical Finance*, 4:73–114.
- Granger, C. W. J. and Ding, Z. (1994). Stylized facts on the temporal and distributional properties of daily data from speculative markets. Technical report, University of California, San Diego and Frank Russell Company, Tacoma, Washington.
- Guillaume, D. M., Dacorogna, M. M., Davutyan, R. R., Muller, U. A., Olsen, R. B., and Pictet, O. V. (1994). From the bird's eye to the microscope: A survey of new stylized facts of the intra-daily foreign exchange markets. Technical report, Olsen & Associates Research Group.
- Harris, L. (1986). A transactions data study of weekly and intradaily patterns in stock returns. *Journal of Financial Economics*, 16:99–117.
- Hasbrouck, J. A. (1991). Measuring the information content of stock trades. *Journal of Finance*, 46:179–207.
- Kyle, A. S. (1985). Continuous auctions and insider trading. *Econometrica*, 53:1315–1335.
- Lo, A. W. and MacKinlay, A. C. (1988). Stock market prices do not follow random walks: Evidence from a simple specification test. *Review of Financial Studies*, 1(1):44–66.
- Lo, A. W. and MacKinlay, A. C. (1989). The size and the power of the variance ratio test in finite samples: a monte carlo investigation. *Journal of Econometrics*, 40:203–238.
- Lockwood, L. J. and Linn, S. C. (1990). An examination of stock market return volatility during overnight and intraday periods, 1964-1989. *Journal of Finance*, 45:591–601.
- Low, A. and Muthuswamy, J. (1996). Information flows in high frequency exchange rates. In Dunis (1996), chapter 1, pages 3–32.
- MacGregor, P. (1999). The sources, preparation and use of high frequency data in the derivative markets. chapter 11, pages 305–311.
- Marinelli, C., Rachev, S., and Roll, R. (2001). Subordinated exchange-rate models: Evidence for heavy-tailed distributions and long-range dependence. *Mathematical and Computer Modelling*, 34(9-11):1–63.
- McNish, T. A. (1993). A geographical model for the daily and weekly seasonal volatility in the foreign exchange market. *Journal of International Money and Finance*, 12(4):413–438.
- McNish, T. H. and Wood, R. A. (1990a). An analysis of transactions data for the toronto stock exchange. *Journal of Business & Finance*, 14:458–491.
- McNish, T. H. and Wood, R. A. (1990b). A transaction data analysis of the variability of common stock returns during 1980-1984. *Journal of Business & Finance*, 14:99–112.

- McNish, T. H. and Wood, R. A. (1991). Hourly returns, volume, trade size, and number of stocks. *Journal of Financial Research*, pages 303–315.
- McNish, T. H. and Wood, R. A. (1992). An analysis of intraday patterns in bid/ask spreads for nyse stocks. *Journal of Finance*, 47:753–764.
- Nelson, C. R. and Plosser, C. I. (1982). Trends and random walks in macroeconomic time series. *Journal of Monetary Economics*, 10:139–162.
- Pan, M.-S., Chan, K. C., and Fok, R. C. W. (1997). Do currency futures follow random walks? *Journal of Empirical Finance*, 4:1–15.
- Patnaik, T. C. and Shah, A. (2002). An empirical characterization of the national stock exchange, using high-frequency data. Technical report, IGIDR, Mumbai.
- Poterba, J. and Summers, L. (1988). Mean Reversion in Stock Returns: Evidence and Implications. *Journal of Financial Economics*, 22:27–60.
- Richardson, M. and Smith, T. (1991). Tests of financial models in the presence of overlapping observations. *Review of Financial Studies*, 4(2):227–254.
- Richardson, M. and Stock, J. H. (1989). Drawing inferences from statistics based on multiyear asset returns. *Journal of Financial Economics*, 25:323–348.
- Roll, R. (1984). A simple measure of the implicit bid-ask spread in an efficient market. *Journal of Finance*, 39:1127–1139.
- Stoll, H. R. and Whaley, R. E. (1990). Stock market structure and volatility. *Review of Financial Studies*, 3:37–71.
- Summers, L. H. (1986). Does the Stock Market Rationally Reflect Fundamental Values? *Journal of Finance*, XLI(3):591–602.
- Wood, R. A., McNish, T. H., and Ord, J. K. (1985). An investigation of transaction data for NYSE stocks. *Journal of Finance*, 40:723–741.
- Wright, J. H. (2000). Alternative variance-ratio tests using ranks and signs. *Journal of Business Economics & Statistics*, 18(1):1–9.