Final Paper No. 128

Understanding the Microstructure of Indian Markets

ABSTRACT

In this study, we examine the functioning of the financial markets in India at the microstructure level. Specifically, for all stocks listed on the NSE we estimate the bid-ask spread and relate it to firm specific characteristics. The median spread is 0.45 rupee in the limit order book (LOB) and 0.72 rupee according to the Roll estimator. This difference is consistent with a U-shaped pattern of intra-day spreads. The LOB spread represents 0.77% of the median price in the limit order book. Rupee spreads increase with price and return variance and decline with market capitalization, trading volume and the number of trades

JEL Classification Code: G14

Keywords: NSE, bid-ask spreads, limit order snapshots

1. Introduction

Market design is increasingly becoming a focal issue in the market microstructure literature. The primary reason for this interest is the fact that market design influences market conditions, which in turn impact the price formation process. In this study we propose to empirically document the market conditions-liquidity and informational environment that characterize how security markets in India function at the microstructure level. Our study will be based on the stocks listed on the National Stock Exchange (henceforth, NSE) of India which is a completely order driven electronic marketplace. The metric of our primary focus will be the bid-ask spread for a stock.

Bid-ask spread (the difference between the lowest ask price and highest bid price for a stock) represents the (variable) cost that an investor pays for a (round trip) trade. Thus, the magnitude of the spread is an important decision variable that an investor considers in choosing a trading venue, as well as the stocks to buy/sell. In classical economic theory, price formation is envisioned as an equilibrium process the outcome of which is an appropriate clearing price at which the demand from buyers equals the supply from sellers. Earlier models assumed a converging process where market frictions are non-existent or irrelevant.

Beginning with the work of Demsetz (1968), the existence of frictions in markets was recognized. There are costs of negotiation, both in terms of time and resources. There are search and bargaining costs. Incorporating these sources of friction in markets, two sets of equilibria were proposed: one for buyers, and one for sellers. Hence two sets of prices, the bid or buy price and the ask or offer price. In an exchange where one can get the bid and ask quotes separately and (price-quantity) orders arrive intermittently, the bid-ask spread depends upon a number of factors, including the probability of information based trading, the volume of trade, the pricing grid (tick size) and share price.

Researchers have, for a long time now, been interested in factors affecting the spread, viz., inventory cost, order processing cost, and adverse selection cost, and what role they play in

influencing the demand for and supply of liquidity and on listing decisions. For example, other things being equal, if the order processing costs for a stock are higher in exchange A than on stock exchange B, then a firm would have an incentive to list in exchange B. Likewise, if an exchange mandates a higher tick size than a competing exchange for the same stock, it de-facto sets a floor on the stock's spread. Thus, the trading platform and supervisory regime that an exchange offers impacts the bid ask spread for a stock. A lower bid ask spread implies higher liquidity for a firm's stock. Studies have shown that firms with higher quality disclosures have lower effective bid-ask spreads and lower adverse selection spread components (Welker, 1995, Healy, Hutton, and Palepu, 1999). Ownership concentration of a firm impacts the spread of its stock (Attig, Gadhoum and Lang, 2003). Concentrated or pyramid shareholding has been shown to widen a stock's bid ask spread.

Although it is hardly necessary to emphasize that the bid ask spread is arguably the most important metric that measures a stock's or market's liquidity, empirical evidence on spread has been established mostly in the context of the developed capital markets. To date, there has been little scientific work that presents a rigorous characterization of the market conditions, bid ask spread and price formation processes for the Indian stock markets, although India ranks second (after the United States) in terms of the number of stocks listed on its exchanges.

The Indian economy has seen phenomenal growth rates in the past decade, much of it fuelled by the large scale outsourcing of business processes that it has attracted since the mid 1990s. Indian GNP at 1980-81 prices rose from Rs. 7878.1 billion in 1997 to Rs. 11151.6 billion in 2002, which constitutes a 41% increase at constant prices. The growing economy has contributed to a strong and growing stock market. Between July 1997 and February 2005, the BSE Sensex Index went from 4305.76 to 6713.34 and the NSE Nifty index rose from 1221.5 to 2103.25. This represents a nearly 56% growth rate in the Sensex index and a 72.2% growth rate in the Nifty index. Figure 1 shows the GDP growth rate between 1950 and 2004, and figure 2 shows the performance of the major stock market indexes between 1997 and 2004. The rapid growth beginning in the late 1990s is clear from the graphs.

This fast growth, coupled with the prominence of job outsourcing to India as a recent election issue, has helped generate current interest in the performance of the Indian economy in general, and the financial sector in particular. A number of studies have focused on firm and industry-level issues – response of business groups to primary market deregulation (Khanna and Palepu, 1998), efficiency improvements following privatization of government owned enterprises (Gupta, 2004), financial liberalization and banking sector efficiency (Ataullah and Hang, 2004), contribution of the information technology sector to India's economic development (Singh, 2003), etc.

Using impact cost as a proxy for liquidity in the Indian markets, Kumar (2003) finds that more than 50% of the variation in liquidity can be attributed to inventory cost and adverse selection risk. Dalvi et. al (2004) add to Kumar's definition of liquidity by including the *immediacy* of supply and examine changes in liquidity following a move from fixed account period settlement to a rolling settlement regime. They find that the competition between the two premier exchanges of India intensify as measured by immediacy of order fulfillment. Gupta, Metia and Trivedi (2004) document some impact of options expiration on NSE trading volume on the expiration days. While the recent availability of NSE trading data has led to some interest in exploring the liquidity characteristics in this limit order market, at the microstructure level, there has not been adequate characterization of the Indian stock markets – the efficiency of trading mechanisms, the regulatory processes and price formation – issues that are crucial in the examination of financial markets.

In this study, we examine the functioning of the financial markets in India at the microstructure level. Specifically, for all stocks listed on the NSE we estimate the bid-ask spread and relate it to firm specific characteristics. By the nature of the data available from the NSE, continuous information of the limit order book is not available, so we compute estimated spread from observed price series using the covariance method developed in Roll (1984). This estimate is then correlated to various firm specific characteristics of the sample securities, using a modeling approach that closely follows the seminal work of Stoll (2002).

We find that average (rupee) spread for all stocks listed on the NSE is 2.17, which is about 3.2% of the average price. This is much larger than the average percentage spreads observed for NYSE and NASDAQ stocks. Comparing this to the tick size of Rs. 0.05 (same across all stocks as per NSE regulations), the spread to tick ratio is 43.4, which is large by international standards. By contrast, median spread is only Rs. 0.72. We find that a significant part of this unusually high spread can be attributed to trading volume. Since (rupee) trading volume is extremely skewed (mean = Rs. 62,545,287; median = Rs. 96,802), we examine volume portfolios and show that stocks with higher trading volume have lower spreads, even after controlling for share price.

While the estimated spreads are based on trade prices, we also collect information on limit order book snapshots to compute direct measures of bid ask spread. The median spread in the LOB is Rs. 0.45, which is 0.77% of the median LOB price. Since the LOB snapshots are not collected for the opening and closing sessions of the market, lower LOB spreads, as compared to estimated spreads, is consistent with spreads being higher during the opening and closing sessions. Coupled with the drop in spread from 11 a.m. to 2 p.m. that we document in our LOB analysis, we conclude that the familiar U-shaped pattern of intra-day spreads is evident in the Indian equity markets.

In our judgment, this work can benefit four different groups of market participants: investors, firms, regulators, and stock exchange management.

(a) Liquidity and transactions cost directly affect the investors' net returns from their equity portfolios. Transaction costs also influence the portfolio selection strategy of investors if they can pick stocks with low spread among comparable alternatives. Although direct calculation of spread is not possible given the data limitations, our study provides the first empirical estimation of rupee spreads, and hence of (variable) transaction costs, for stocks listed on the NSE.

(b) Firms' listing choice can possibly take these costs into account. Although, listing fee of a particular exchange is the direct cost paid by the firm, a much more important indirect issue is the liquidity in firm's stocks. Improved liquidity can potentially reduce the cost of

equity for listed firms (Amihud and Mendelson (1986)). And as has been explained before, the bid-ask spread is arguably the most important metric that captures liquidity.

(c) Results of this study will also help achieve the policy goals of lawmakers who want to increase the efficiency of securities markets.

(d) Stock exchanges can become more competitive[1] and attract more investors for trading, and more firms for listing their stocks, once they are aware of the dynamics of trading costs.

The rest of this paper is organized as follows. Section 2 presents a literature review on bid ask spread and its importance as a metric that captures (albeit not completely) the variable cost of trading borne by an investor. Section 3 briefly describes the informational and regulatory environment of the Indian equity markets, with special emphasis on the NSE trading platform. We then describe the dataset used, sample selection criteria and present summary statistics of interest. Section 4 outlines the methodology for estimating spread using the serial covariance approach and presents the results of our regression models that relate estimated spread to firm level characteristics. Since most of the variables of interest are extremely (right) skewed, we also conduct portfolio regressions and present our results. Section 5 analyses the LOB snapshots and provides some indicative results on the intra-day variations in the metrics of interest. Section 6 concludes with a review of the rationale for why we consider this an important study in the context of Indian stock markets.

2. Literature review on the bid ask spread

As a wedge between the price that a buyer pays and the price that a seller receives for a stock is the bid ask spread, which has interested researchers in finance for a long time. Arguably, it is the

[1] Previous literature has documented various instances where trading turnover is found to be very sensitive to trading costs and market structure. See for example Pagano and Stiel (1996) who document that in 1989, French order handling rules made block trades unattractive and majority of block trades in French stocks were executed anonymously on the London SEAQ-International exchange.

single most important metric for measuring friction in asset markets. As Stoll (2000) defines it, friction in financial markets measures the difficulty with which an asset is traded. It can either be measured in time – how long it takes to trade – or in price – the price concession it takes to trade immediately. Bid ask spread related to the latter measure. Buyer initiated trades are usually made at the ask price and seller initiated trades at the bid price. Thus the spread between the bid price and the ask price is a measure of friction.

A number of variables related to a firm's trading characteristics have been found to be empirically robustly related to spread. A firm's trading volume, the number of trades and its market value of equity are generally found to be negatively related to spread. That is because, increased volume and number of trades makes it more likely to locate a counterparty to trade and reduces the necessity of holding large and expensive inventory for very long. Larger capitalization firms have more information and reduces uncertainty, hence a lower spread. A stock's return variance is positively related to spread, mostly for information reasons. Return variance measures the likelihood of adverse price change of a stock, and the higher the variance, the higher is the friction to trading and hence the higher the spread. Price is a proxy for risk – the higher the stock price, the lower is risk and hence spread.

Apart from trading characteristics, research has shown that the informational environment of a firm also affects spread. Earnings announcements impact the bid ask spread by increasing the adverse selection component of spread (Krinsly and Lee, 1996). Conrad and Niden (1992) find weak evidence of spread changes around corporate information events. Barclay and Smith (1988), and Conrad, Mandelker, Niden, Rosenfeld and Shastri (1991) find a significant change in spreads during stock repurchases and corporate acquisition announcements.

Trading mechanisms and market design also impact the spread. Harris (1995) relates spread to tick size and provides evidence on the relationship between tick size increase and spread increase. Others have shown that spreads in specialist markets (like the NYSE) are smaller than in dealer markets (such as the NASDAQ).

In this study, we examine the relationship of the bid-ask spread to a firm's trading characteristics for an emerging market such as India. We also use the limit order book information on order arrival to infer about intra day variation in spread. Our study is based on the stocks listed on the NSE, which is a fully automated limit order market.

3. Informational environment at the NSE, data and sample characteristics

The NSE came into existence at a time when securities markets in India were characterized by little transparency, limited participation of retail investors and dominated by entrenched brokerage firms. It was incorporated in 1992 and recognized as a stock exchange in 1993, and by October 1995, it became the biggest exchange in the country.

3.1. The NSE trading platform

The NSE is a completely automated limit order market, whose main contribution to Indian financial markets has been to incorporate forward-looking technology in the trading process and enabling equal and fair access to all investors through its electronic online trading platform. The exchange played a pivotal role in implementing market reforms in India and establishing industry benchmarks in the adoption of best practices. One of the key ways it protects the interests of the investors is by ensuring the separation of ownership, management and trading rights. It is owned by a set of leading financial institutions and managed by professionals who do not, directly or indirectly, trade on the exchange. Another innovation it introduced, which was particularly useful after India's largest financial scandal involving powerful and entrenched brokers, was to eliminate settlement risks by setting up a settlement guarantee fund and reducing the settlement cycle.[2] Additionally, it has also created an investors protection fund that compensates investors for financial damages resulting due to default of brokers.

[2] Similar issued are highlighted in Khwaja and Mian's (forthcoming) JFE article about Pakistani brokers' use of pumping-and-dumping schemes to manipulate stock prices and defraud uninformed investors.

Currently, the NSE operates three different segments – the Wholesale Debt Market (WDM), the Capital Market (CM) and the Futures and Options Segment (F&O). The WDM provides a platform for trading a wide range of debt securities. In March 2004 (our sample period), this segment reported a net traded value of Rs. 98,806.35 crores, with average trade size of Rs. 7.05 crores. The CM offers a screen based trading system for equities/preference shares, debentures, ETFs as well as retail government securities. The trading platform, known as NEAT (National Exchange for Automated Trading) operates on a price/time priority basis, in much the same way as ECNs operate in the US. For our sample period, the CM segment had a market capitalization of Rs. 1,120,976 crores with average daily turnover of Rs. 4,767 crores. The F&O segment provides trading in derivative instruments including index futures, index options, stock futures, stock options, interest rate futures and other credit derivatives. In March 2004, the total number of contracts traded in this segment was 7,006,620, representing an annual daily turnover of Rs. 11,840 crores.

To promote informationally efficient financial markets and keep trading costs to a minimum, the NSE has relied heavily on state-of-the-art technology. NSE is the first exchange in the world to use satellite communication technology for trading. Its client server based trading platform NEAT operates on 2,888 VSATs in 365 cities spread all over the country. The system boasts an uptime in excess of 99.7% with data integrity record of one single error in 10 million bits. For all trades entered into the NEAT system, there is a uniform response time of less than 1.5 seconds. It has recently launched the NIBIS (NSE's Internet Based Information System) for online real-time dissemination of trading information over the Internet, very much like the 'Book Viewer' provided on INET ATS's web site in the US. In all, it has established a remarkable record as an automated exchange and changed the very structure of and practices in securities markets in India.

3.2. Data files, sample selection and summary statistics

How does this trading platform facilitate efficient price formation? That is the main question we answer in this study, using the sample of all stocks that traded on the NSE in March 2004. The

NSE provides trade and quotes data, much like the TAQ data disseminated by the NYSE. For each month, there is one (or more, depending on trading activity) CD that is named in the year-month format.[3] Under the starting directory, there are seven sub-directories, or files. The 'Masters' file is the database master file, it contains information on all securities that traded in that month, along with identifying information like the series class (equity/debt/etc), the International Security Identification Number (ISIN) which is the NSE equivalent of CUSIP number, and appropriate flags to show listing status. The 'Bhav-copy' file provides summary information about each security for each trading day. The 'Index' file provides tick-by-tick information on index movements for three broad stock indexes – the Nifty, Nifty Junior and Defty.[4] The 'Trades' file contains real time information on all trades that take place in each security for that month. Unlike the TAQ data, the NSE provides quotes data at various points in time, and not continuously throughout the day. The 'Snapshots' file contains these quotes, which are snapshots of the limit order book at given hours of the day.[5] Finally, each CD also contains a 'Circulars' file that lists all notices and circulars served by the NSE for any of the listed stocks. Appendix I provides a detailed description of the data structure.

For our sample month, the number of stocks that traded on the NSE exceeded one thousand. However, after eliminating stocks with missing data our universe of stocks was reduced to 893. Our summary measures presented in table 1 and distributional characteristics of these measure presented in table 2 are calculated for these 800 plus stocks. Since trading activity on a typical day can be quite different across firms, we consider several variables that reflect these cross sectional variations. All variables are standardized to daily basis. Since there are over 800 sample firms' trading data for 22

[3] For example, in Release A version of the data, when you mount the CD for March 1999, the starting directory is 199903.

[4] Nifty is the main stock market index in India; it is composed of the top 50 highly liquid stocks in India which make up roughly half of the market capitalization of India. Nifty Junior is the second tier of 50 less liquid stocks. Nifty Junior accounts for around 10% of the market capitalization of India. Nifty and Nifty Junior are always disjoint sets: there is never any common index member. Defty is the same as Nifty, expressed in dollars.

[5] For example, the order book snapshots for 5 March, 1999 are stored in the directory 19903/Snapshots/19990305. The files that are found inside this have names of the form hhmmss.gz, to convey the time at which the snapshot was taken. For example, for 5 March, 1999 the CD for Release A contains 110000.gz, 130000.gz and 140000.gz. These are order book snapshots at 11 AM, 1 PM and 2 PM.

trading days, the summary measures involve some degree of aggregation. We first calculate the average measure for each security for each day and then average these daily measures over the entire month for each stock. Thus, we obtain one summary measure per security for the month and have over 800 such measures for the sample. The mean, median, minimum, maximum and the various percentile values are calculated over these averages.

Share price ranges from a minimum of Rs. 0.29 to a maximum of Rs. 14,061.32, with the average value of Rs. 179.11. In table 2 we find that the median share price is Rs. 21.00. Share price distribution shows that the mean lies between the 75th and the 90th percentile. Thus, a large majority of the stocks that trade on the NSE are relatively low priced stocks. Average daily number of trades ranges from a minimum of 1 trade to a maximum of 86,741.10 trades, with the mean (median) value being 1922.87 (39.97) trades. Looking at the distribution of this variable, we find that the mean lies between the 75th and 90th percentile, and is close to the latter. In terms of market capitalization also, the sample shows a similar (right) skewness. While the lowest market value of equity (MVE) is Rs. 0.19 crores, the highest is Rs. 792,722.75 crores, with the mean (median) value being Rs. 17,310.78 crores (Rs. 2,193.08 crores). It is clear from the range and the mean (median) that the sample is highly skewed. In the absence of symmetric distribution of the variables of interest, it is conceivable that the full-sample results may not generalize to sub samples. More seriously, violation of the normality assumptions may render our OLS estimates inefficient. Hence, for our ensuing regressions analyses, we also consider sub samples of the universe to draw conclusions.

To further explore the ramifications of this skewness, we partition the sample into four price quartiles, Q1 being the lowest priced stocks and Q4 the highest. Table 3 shows the same variables, now measured for each of the price quartiles. We find that the average share price for the lowest 25% of the stocks is only Rs. 13.79 as compared to Rs. 622.63 for the highest 25% of the sample securities. The low priced stocks also have much lower trading volume and market capitalization. The highest 25% stocks have an average market value of equity more than ten times that of the lowest 25%. This

severe skewness causes some abnormal results in our spread estimation and regressions, as we shall document in the subsequent sections.

4. Estimating Spread

In a market that is informationally efficient, there will be no serial dependence in observed price changes if trading costs are zero. However, the zero costs scenario is a theoretical abstraction. Once we introduce the possibility that investors incur variable costs of trading, for holding inventory, processing orders and search costs, it becomes clear that even in an efficient market, there will be a positive bid-ask spread. As a result of the spread, trades take place at either the bid price or the ask price. Because trade prices 'bounce' between the bid and the ask prices, Roll (1984) showed that this induces a negative serial dependence in observed prices. This is true even if the conditional distribution of transaction types is assumed to be independent. Roll then shows that this negative serial dependence implies a bid ask spread that can be estimated as

$$S = 2\sqrt{-COV(\Delta P_t, \Delta P_{t-1})} \qquad (1)$$

where P_t is the transaction price at time t, and COV (ΔP_t, ΔP_{t-1}) is the covariance between two successive price changes. We use this measure here since the estimation requires only trade prices, and not actual bid and ask price quotes. We calculate spreads for 776 stocks on all 21 trading days in March giving us 15,161 stock-days. On each day trade-to-trade returns for all transactions are utilized for estimating the spread. Then we drop 5,945 observations (stock-days) which had positive correlation. This enables us to keep a given stock in the sample as long as it has at least one negative correlation day. Roll's estimates are obtainable for majority of the stocks (584) on 10 or more trading days. There are only 28 stocks which had 3 or fewer trading days with negative correlation.

Table 1 shows that the average estimated (Roll) rupee spread is 2.17 with a very high standard deviation (10.55). In table 2, we find that the bottom one percent of all stocks have a spread close to

zero. The median rupee spread is 0.72, which is much smaller than the mean. This is not surprising, given the extreme right skewness we documented in the distribution of stock price, market value of equity and trading volume. A more detailed picture of the distribution of spread is available in table 3. We find that the average (raw) spread is under one rupee for the bottom two price quartiles. The average spread for the third price quartile is Rs. 1.56, which is less than the average spread for the entire sample. This highlights that an overwhelming majority of the sample lies to the left of the mean. Percent spread, calculated as (Roll) rupee spread divided by share price, also shows the same trend. While the mean percent spread is 0.03, the median is much smaller (0.01). In table 3 we see that the average percent spread for the bottom 75% of the stocks (sorted on price) is less than the average percent spread for the entire sample.

The cross sectional relation of spreads to a firm's trading characteristics was first characterized by Demsetz (1968) as the price for immediacy. Later, Stoll (2000) used a similar model to show the robustness of the bid-ask spread as a measure of trading friction in capitalist markets. However, one has to remember that most of the empirical work that relates spreads to various aspects of a firm's trading characteristics were derived from mature financial markets, predominantly the US. There has been little work to examine whether the same empirical regularities hold for an emerging market such as India. In this study, we use the model developed by Stoll (2000) to examine the variables that affect the bid-ask spread in Indian equity markets. Specifically, we estimate the following models:

$$Y = \alpha + \beta_1 Ln(RupeeVolume) + \beta_2 \sigma_r^2 + \beta_3 Ln(MVE) + \beta_4 Ln(Share\,\mathrm{Pr}ice) + \beta_5 Ln(NumTrades) + \varepsilon$$

Y= Roll Percent Spread; Roll Rupee Spread (2)

$$Y = \alpha + \beta_1 (RupeeVolume) + \beta_2 \sigma_r + \beta_3 (MVE) + \beta_4 (MedianShare\,\mathrm{Pr}ice) + \beta_5 (NumTrades) + \varepsilon$$

Y= Roll Percent Spread; Roll Rupee Spread (3)

Here the dependent variable is (a) percent spread, which is the estimated Roll spread divided by stock price, and (b) the raw Roll spread measured in rupee. The percent spread specification is

used to benchmark our results with Stoll's model and provide some comparison between observed empirical regularities in the developed (US) markets and those from an emerging market. Standardizing spread using share price is a common practice, and it has the added benefit of improving the adjusted-R^2.[6] However, as Bollen, Smith and Whaley (2003) point out, there are certain limitations to using the percentage spread specification for the dependent variable. When spreads do not vary much but price dispersion is considerable, the percentage spread specification provides coefficient estimates that reflect the response of price volatility to the explanatory variables, and not the relationship between spread and said variables. In our summary statistics discussions, it is clear that the range of share prices for our sample stocks is considerable. While the minimum share price is Rs 0.26, the maximum is Rs. 14,061.32. To provide some comparison, in March 2004, the minimum (maximum) share price for NYSE listed stocks was \$0.5 (\$3,111) and for NASDAQ listed issues, the minimum (maximum) was \$0.18 (\$181.5). Clearly stock price dispersion is much higher in the NSE. To address this issue, we also use the absolute (or raw) spread as a dependent variable in our regression specifications. Our methodology, therefore, serves to not only relate our results to those established in the literature, but also to the peculiarities of the Indian market.

Volume is measured as share volume in rupee, MVE stand for market value of equity. The regression results anticipate spread increase with greater price volatility of stock, with higher stock price, and with lower volume. Tables 4 and 5 present the estimates of equation (2). Panel A shows the full sample results and Panel B shows the results for the four price and four volume portfolios. In Table 4 panel A, we find that apart from the expected negative relationship between the number of trades and percent spread (β_5 = -0.0413, *t*-stat = -3.04), all other coefficients are insignificant. This is not unexpected. One of the peculiarities of the Indian stock markets is the fact that although the number of listed stocks is large (second only to the US), regular trading takes place only in a very

[6] Tinic and West (1974) estimate a model of absolute spread with price per share, log of trading volume, price volatility (measured by high-low price range divided by share price), trading continuity and the number of markets in which that security traded, and they report an adjusted-R^2 of 0.499. When they estimate the same model using percent spread as the dependent variable (and dropping the explanatory price variable), the adjusted-R^2 increases to 0.804.

small subset of all listed firms. Most stocks trade infrequently. To account for this, in panel B, we show results for subgroups of our sample, sorted on (a) volume, and (b) price.

Directionally, we find the expected relationship in most of the sub groups. Our estimates of β_1 are negative in three of the four volume portfolios, showing the negative relationship between trading volume and proportional spread. Volatility is positively related to proportional spread in three of the four volume portfolios, although the magnitudes are not statistically significant. Unlike the US markets, where the adjusted-R^2 for such a model would typically lie between 0.65 and 0.75,[7] we find our adjusted-R^2 to be much lower, 0.09 for full sample and between 0.12 and 0.46 for the sub group results. This is attributable to the thin trading phenomenon that we mentioned earlier.

Table 5 shows the relationship of rupee spread to various measures of a firm's trading characteristics. Here the dependent variable is rupee spread so we expect a positive coefficient on (logarithm of) share price. β_4 is positive and significant in the whole sample as well as in three of the four volume portfolios. We do not expect to see this relationship to be significant in the price-sorted portfolios. Number of trades has a robust and significant negative relationship to the dependent variable, which implies that as trading frequency increases, there is a reduction in bid-ask spread.

Tables 6 and 7 present results on the same model, but now without logarithmic transform of the independent variables. The results are consistent with what we find in the previous two tables. Volume is negatively related to both raw and proportional spread and raw spread increases as share price increases. The estimates for the price and volume portfolios are as expected and statistically significant, the full sample results are often rendered weak due to the inclusion of a large number of stocks with very thin trading.

5. Snapshot data analysis

[7] In Stoll's (2000) table 1, for the NYSE/AMEX stocks, the adjusted R^2 lies between 0.77 and 0.79, and for the Nasdaq stocks it is between 0.64 and 0.67.

Our estimates of spread in the previous section were obtained from the trade prices for the sample stocks. To obtain actual (not estimated) spreads, one need the bid and ask quotes which can be obtained from the limit order book (LOB). The NSE, at this time, does not distribute continuous LOB data, hence calculating actual bid ask spreads from the LOB is not directly possible. However, the NSE does collect snapshots of the LOB at four different times of the trading day – 11 a.m., 12 noon, 1 p.m. and 2 p.m. The LOB snapshot lists all outstanding orders, identified as buy and sell orders, that are in the book at the time the snapshot is recorded. There are stop-loss orders, as well as regular limit orders, and for our study, we have ignored all the stop-loss orders. Each order entry also identifies the time at which it was entered into the LOB, the associated price and quantity (depth).

Table 8 shows some summary statistics from analysis of the LOB. Average ask price is Rs. 183.66 and average bid price is Rs. 181.83. The mean quote mid-point is Rs. 182.75, which is a little higher than average trade price of Rs. 179.11 that we obtain from the bhavcopy (trade) files. Average raw spread calculated from the LOB snapshots is Rs. 1.82, which is lower than the estimated Roll spread of Rs. 2.17. Our explanation of this difference is tied to the intra-day variation in spread, which we discuss in Table 9.

The last two rows of Table 8 show the *age* of the outstanding orders in the LOB at the time of each snapshot. The age of an order is the difference in the snapshot date-time and the date-time when the order was entered into the LOB. We note here that some of the snapshot files are recorded with a lag, depending on the time it takes to record all the outstanding orders for all stocks into one snapshot file. This delay may result in some stocks showing a negative age. We do not eliminate these negative age orders, since doing so would eliminate the freshest orders. This is procedurally identical to what Dalvi et. al. (2005) document in their study of liquidity on the NSE. The minimum age of orders (both bid and ask) is –10815. These are orders that entered the LOB after the snapshot time and are the freshest orders. We find that up to the 10^{th} percentile of all orders, on both the bid and ask side, have a negative age, which implies that these came into the book after the snapshot time.

In Table 9 we present some indicative results on intra-day variations on some of the variables we have analyzed so far. As explained earlier, we have four time stamped LOB snapshots per day, so analyzing continuous movements in variables is beyond the scope of the data. However, the snapshots are collectedly at hourly intervals for four hours in a five-and-a-half hour trading day, and thus provide a good sense of intra-day variations, if not a complete picture.

Both bid and ask prices drop uniformly from 11 a.m. to 2 p.m. This is true of the averages, as well as the entire distribution. Both bid and ask depth increase in the morning hours and after 1 p.m. there is a fall. Both raw and proportional spreads decrease uniformly between 11 a.m. and 2 p.m. In our estimate of Roll spread from the trade files, we found average spread to be higher than the spread calculated from the LOB. One explanation that reconciles these two pieces of evidence is that spreads during the opening (9:55 a.m.) and closing (3:30 p.m.) sessions are higher than the rest of the day. This would also be in agreement with the intra-day pattern in spread that has been documented for other mature financial markets. The age of orders uniformly increase as the trading day progresses. This, again, is expected. Since outstanding orders are (mostly) cancelled at the end of each trading day, the amount of time that an order can sit in the LOB, after being entered from 9:55 a.m. onwards, increases as the day progresses.

Since the snapshot data allows direct calculation of the bid-ask spread for those stocks that trade show reasonable activity in the limit order book, we expect the mean spreads calculated from these snapshots to be free of the thin-trading problem that renders the estimated (Roll) spread regression coefficients weak. So in this section, we re-estimate the same models but now using the snapshot raw and proportional spreads as the dependent variables. The equations are the same as specified in (2) and (3) above.

Tables 10 and 11 present the estimates for equation (2) using snapshot raw spread and snapshot percent spread respectively as the dependent variable. In Table 10 we find the expected negative relationship between percent spread and share price, showing that as a proportion of price, spread decreases as share price increases. Higher trading volume and larger number of trades are both

associated with lower proportional spread, as expected. All the coefficients are also statistically significant.

In Table 10, we report the estimated coefficients using raw spread as the dependent variable. As expected, we find a positive and significant relation (β_4 = 2.0106, *t*-stat = 33.86) between share price and raw spread. In both tables, the coefficient of return variance is positive and significant, which is also a result that was documented, amongst others, by Stoll (2000).

Tables 12 and 13 estimate equation (3), using snapshot percent spread and raw spread respectively as the dependent variable. Again, the results are consistent with our previous findings. We note here that using the snapshot spreads not only improves the coefficient estimates, but also increases the fit of the model to the data. For three of the four models, the adjusted-R^2 increases when the Roll spreads are replaced with comparable snapshot spreads.

6. Review and conclusions

In this paper we examine, for the first time, the bid ask spread for stocks trading on the NSE, India. This allows, for the first time, to compare the frictions to trading in an emerging market like India to the developed western securities markets. We find that average (rupee) spread for all stocks listed on the NSE is 2.17, which is about 3.2% of the average price. This is much larger than the average percentage spreads observed for NYSE and NASDAQ stocks. Comparing this to the tick size of Rs. 0.05 (same across all stocks as per NSE regulations), the spread to tick ratio is 43.4, which is also large by international standards. Variables that affect the bid-ask spread, viz. trading volume, market capitalization and share price all show extremely (right) skewed distributions. Any empirical methodology that attempts to characterize patterns in the data needs to address this issue.

While the estimated spreads are based on trade prices, we also collect information on limit order book snapshots to compute direct measures of bid ask spread. The median spread in the LOB is Rs. 0.45, which is 0.77% of the median LOB price. Since the LOB snapshots are not collected for the opening and closing sessions of the market, lower LOB spreads, as compared to estimated spreads,

is consistent with spreads being higher during the opening and closing sessions. Coupled with the drop in spread from 11 a.m. to 2 p.m. that we document in our LOB analysis, we conclude that the familiar U-shaped pattern of intra-day spreads is evident in the Indian equity markets.

Bid ask spread is one of the most important determinants of market liquidity, which influences a trader's choice of trading venue and stocks to buy/sell. By providing a scientific analysis of the spread, we propose to highlight the nature of this decision variable from an investor's perspective. This work has very broad implications and can benefit four different groups of market participants: investors, firms, regulators, and stock exchange management, in maximizing their goals of highest returns, lowest cost of capital, fair markets, and market shares, respectively. From an academic viewpoint, our work examines the microstructure of the Indian financial markets for the first time.

References

Ahn, H-J., J. Cai, Y. Hamao, and R. Ho, 2004. The components of the Bid-Ask spread in Limit Order Market: Evidence from the Tokyo Stock Exchange. Journal of Empirical Finance, forthcoming.

Amihud, Y. and H. Mendelson, 1986, Asset pricing and the bid-ask spread, Journal of Financial Economics, 17, 223-249.

Biais, B., P. Hillion, and C. Spatt, 1995, An Empirical Analysis of the Limit-order Book and the Order Flow in the Paris Bourse, Journal of Finance 50, 1655-89.

Clarke, J. & Shastri, K., 2000. On information asymmetry metrics. Working Paper, Georgia Institute of Technology.

Copeland, T., Galai, D., 1983. Information effects on the bid-ask spread. Journal of Finance 38, 1457–1469.

Dalvi, M., M. Krishnan, G. C. Nath, 2004. Liquidity in an Emerging Market, Working Paper, CUNY

De Winne, R., D'Hondt, C., 2003. Rebuilding the limit order book on Euronext or how to improve market liquidity assesment. Working Paper, Facultés Universitaires Catholiques de Mons.

Declerck, F., 2002. Trading costs on a limit order book market: Evidence from the Paris bourse. Banque et Marché 57, 31–45.

Demsetz, H., 1968. The cost of transacting, Quarterly Journal of Economics 82, 33-53.

Easley, D., Kiefer, N., O'Hara, M., Paperman, J., 1996. Liquidity, information, and infrequently traded stocks. Journal of Finance 51, 1405-1436.

Easley, D., Kiefer, N. M., O'Hara, M., Paperman, J., 1996. Liquidity, information, and infrequently traded stocks. Journal of Finance 51, 1405–1436.

Grossman, S., M. Miller, 1988. Liquidity and Market Structure. Journal of Finance, 43, 617-633.

Ho, T., Stoll, H., 1981. Optimal dealer pricing under transactions and return uncertainty. Journal of Financial Economics 9, 47–73.

Huang, R., Stoll, H., 1997. The components of the bid-ask spread: A general approach. Review of Financial Studies 10, 995–1034.

Krinsky and Lee, JF 1996, Earnings Announcement and the components of the bid ask spread, No 4, 1523 – 1535.

Kumar, S., 2003. Determinants of liquidity in a Open Electronic Limit Order Book Market, Working Paper, Indira Gandhi Institute of Research and Development.

Lee, C., Ready, M., 1991. Inferring trade direction from intraday data. Journal of Finance 46, 733–746.

Madhavan, A., Richardson, M., Roomans, M., 1997. Why do security prices change? A transaction-level analysis of NYSE stocks. Review of Financial Studies 10, 1035–1064.

Menyah, K., Paudyal, K., 2000. The components of bid-ask spreads on the London Stock Exchange. Journal of Banking and Finance 24, 1767–1785.

Pagano, M., and B. Steil, 1996, Equity trading I: The evolution of European trading systems, The European Equity Markets: The State of the Union and an Agenda for the Millennium, A Report of European Capital Market Institute, 1-58.

Parlour, C., 1998, Price Dynamics in Limit Order Markets, Review of Financial Studies 11, 789-816.

Roll, R., 1984. A simple implicit measure of the effective bid-ask spread in an efficient market. Journal of Finance 39, 1127–1139.

Saporta, V., Trebeschi, G., Vila, A., 1999. Price formation and transparency on the London Stock Exchange. Working Paper, Bank of England.

Silva, A., Chavez, G., 2002. Components of execution costs: Evidence of asymmetric information at the Mexican Stock Exchange. Journal of International Financial Markets, Institutions & Money 12, 253–278.

Stoll, H. R., 2000. Friction, Journal of Finance 55, 1479-1514.

Van Ness, B., Van Ness, R., Warr, R., 2001. How well do adverse selection components measure adverse selection? Financial Management 30, 77–98.

Vandelanoite, S., 2002. Takeover announcements and the components of the bid-ask spread. Working Paper, University of Paris.

Wang, J., 1999. Asymmetric information and the bid-ask spread: an empirical comparison between automated order execution and open outcry auction. Journal of International Financial Markets, Institutions & Money 9, 115–128.

Table 1: Summary Statistics of the variables computed from the trade files

The analysis is based on a sample of all stocks that traded on the National Stock Exchange of India in March 1, 2004 to March 31, 2004. All variables are first calculated as the average for each security for each day and we calculate a daily average for each trading day of the sample month. Roll spread is estimated from trade prices using Richard Roll's covariance estimator where spread = $S = 2\sqrt{-COV(\Delta P_t, \Delta P_{t-1})}$. Percent spread is Roll's spread divided by share price. Rupee volume is the trading volume multiplied by the (INR) price per share traded. Return variance is calculated over the past one year. Market value of equity is calculated as the share price multiplied by the number of shares outstanding. Number of trades refers to the number of trades for each stock each day.

Variable	N	Minimum	Mean	Standard deviation	Maximum
Roll Spread	768	0	2.17	10.55	196.28
Percent Spread	768	0	0.0320	0.0974	2.08
Number of Trades	776	1.00	1922.87	7195.26	86741.10
Share Price (Median)	776	0.29	179.11	608.18	14061.32
Market Value of Equity	884	0.19	17310.78	59228.61	792722.75
Return Variance	893	0	0.0016	0.0020	0.0139
Rupee Trading Volume	776	3	62545287	323367004	3667265383

Table 2: Distribution of the variables

The analysis is based on a sample of all stocks that traded on the National Stock Exchange of India in March 1, 2004 to March 31, 2004. All variables are first calculated as the average for each security for each day and we calculate a daily average for each trading day of the sample month. Roll spread is estimated from trade prices using Richard Roll's covariance estimator where spread = $S = 2\sqrt{-COV(\Delta P_t, \Delta P_{t-1})}$. Percent spread is Roll's spread divided by share price. Rupee volume is the trading volume multiplied by the (INR) price per share traded. Return variance is calculated over the past one year. Market value of equity is calculated as the share price multiplied by the number of shares outstanding. Number of trades refers to the number of trades for each stock each day.

Variable	Percentile								
	1st	5th	10th	25th	50th	75th	90th	95th	99th
Roll Spread	0	0.0161	0.0354	0.1415	0.7244	5.1414	13.0000	16.0000	23.3064
Percent Spread	0	0.0003	0.0006	0.0026	0.0150	0.1254	12.0000	15.0000	18.0000
Number of Trades	1.00	3.00	7.00	20.10	39.97	288.00	2126.95	6529.00	34125.00
Share Price (Median)	0.2182	1.0000	2.2289	8.8289	21.0000	80.6000	269.7750	480.2417	1069.7000
Log Share Price	-1.52	0.00	0.80	2.17	3.04	4.38	5.59	6.17	6.97
Mkt Val. of Equity (MVE)	9.06	37.00	127.22	571.77	2193.08	8498.22	36039.30	103338.50	348863.13
Log MVE	2.20	3.61	4.85	6.34	7.69	9.04	10.49	11.54	12.76
Return Variance	0	0	.0003	0.0007	0.0012	0.0018	0.0027	0.0038	0.0091
Rupee Trading Volume	27	575	2197	9261	96802	2351673	29285831	125012885	1535421169

Table 3: Summary statistics by price portfolios

The analysis is based on a sample of all stocks that traded on the National Stock Exchange of India between March 1, 2004 and March 31, 2004. All variables are first calculated as the average for each security for each day and we calculate a daily average for each trading day of the sample month. Stocks are sorted on share price and grouped into four portfolios, with Price Quartile 1 being the lowest priced and Price Quartile 4 being the highest priced. Roll spread is estimated from trade prices using Richard Roll's covariance estimator where spread = $S = 2\sqrt{-COV(\Delta P_t, \Delta P_{t-1})}$. Percent spread is Roll's spread divided by share price. Rupee volume is the trading volume multiplied by the rupee price per share traded. Return variance is calculated over the past one year. Market value of equity is calculated as the share price multiplied by the number of shares outstanding. Number of trades refers to the number of trades for each stock each day.

Variables		Price Quartile 1	Price Quartile 2	Price Quartile 3	Price Quartile 4
Roll Spread	Min	0.0153	0.0000	0.0408	0.0442
	Mean	0.5736	0.6541	1.5633	4.4682
	Max	6.9527	7.6183	12.6716	110.3849
	Standard Dev.	0.8964	1.1853	2.1050	14.6173
Percent Spread	Min	0.0010	0.0000	0.0002	0.0001
	Mean	0.0206	0.0125	0.0211	0.0583
	Max	0.1746	0.1302	0.2360	2.0836
	Standard Dev.	0.0261	0.0194	0.0360	0.2394
Number of Trades	Min	5.29	2.22	10.38	9.70
	Mean	316.65	779.21	2955.80	5577.98
	Max	3590.95	10630.05	47510.71	86741.10
	Standard Dev.	646.39	1607.74	7346.70	13240.70
Share Price	Min	1.0452	29.8988	83.4774	245.2976
	Mean	13.7914	52.7097	152.7879	622.6263
	Max	28.7440	83.2357	244.7655	5094.9464
	Standard Dev.	7.9625	15.8593	44.2355	697.1846
MVE	Min	18.22	127.22	37.00	330.41
(Rs. Crores)	Mean	4488.94	4903.55	16402.23	54061.51
	Max	147359.94	96146.56	383016.38	792722.75
	Standard Dev.	17304.82	13455.08	52077.82	106000.85
Variance	Min	0.0000	0.0000	0.0000	0.0000
	Mean	0.0021	0.0015	0.0011	0.0008
	Max	0.0100	0.0043	0.0032	0.0021
	Standard Dev.	0.0016	0.0007	0.0006	0.0004
Trading Volume	Min	6139.00	11069.08	68977.70	167821.73
(Number of trades)	Mean	1977675.84	8389771.87	64233068.16	243215321.97
	Max	27512569.86	119208114.77	1535421169.37	3658459601.72
	Standard Dev.	4738093.01	20464797.12	195975845.29	661547685.53

Table 4: Regression Model with Roll's Estimated Percent Spread as the Dependent Variable

The table reports the results of the equation below:

$$Y = \alpha + \beta_1 Ln(RupeeVolume) + \beta_2 \sigma_r^2 + \beta_3 Ln(MVE) + \beta_4 Ln(Share\,\mathrm{Pr}ice) + \beta_5 Ln(NumTrades) + \varepsilon$$

All variables are first calculated as the average for each security for each day and we calculate a daily average for each trading day of the sample month, March 2004. Volume is the trading volume multiplied by the rupee price per share traded. σ_r^2 is the variance of returns, calculated over the past one year. Share Price is the median share price, MVE is the market value of equity, and is calculated as the share price multiplied by the number of shares outstanding. NumTrades is the number of trades for each stock each day. Ln represents the natural logarithmic transform of the variables. In Panel A, we present the full sample results, and in Panel B, we present the results for Volume and Price portfolios. All stock are ranked on the basis of their rupee trading volume (share price), the bottom 25% forms the first quarter (Q1), the 25% to 50% range forms Q2, 50% to 75% Q3 and finally 75% to 100% forms the top quarter Q4. t-statistics are reported in parentheses, and * (**) denotes significance at the 95% (99%) level of confidence.

Panel A

	α	β_1	β_2	β_3	β_4	β_5
Coefficient Estimate	-0.1499	0.0257	5.0027	-0.0034	0.0105	-0.0413
t-stat	(-1.56)	(1.89)	(0.77)	(-0.62)	(1.23)	(-3.04**)
p-value	0.11	0.06	0.44	0.53	0.21	0.00
	Adj-R^2 = 0.0898			F= 8.95**		

Panel B

Estimated	Volume Portfolios				Price Portfolios			
Coefficients	Q_1	Q_2	Q_3	Q_4	Q_1	Q_2	Q_3	Q_4
α	0.1942	-0.2622	-0.0974	-0.8867	0.1489	0.0544	-0.0715	-1.4235
	(2.55*)	(-0.32)	(-0.79)	(-3.82**)	(3.93**)	(1.51)	(-0.97)	(-3.07**)
β_1	-0.0175	-0.0011	0.0149	-0.1462	-0.0087	-0.0033	0.0090	-0.1185
	(-1.82)	(-0.01)	(1.30)	(-5.17**)	(-1.58)	(-0.71)	(1.32)	(-2.54*)
β_2	4.1450	14.0664	-1.2505	15.4865	-2.2911	2.4105	-4.5865	-11.2645
	(1.08)	(0.73)	(-0.17)	(0.65)	(-1.73)	(1.04)	(-0.73)	(-0.19)
β_3	-0.0012	-0.0408	-0.0009	-0.0182	0.0004	0.0036	0.0012	-0.0320
	(-0.40)	(-1.68)	(-0.20)	(-1.60)	(0.32)	(1.60)	(0.33)	(-1.37)
β_4	0.0203	0.1168	0.0031	-0.0486	-0.0058	0.0011	0.0087	0.1125
	(3.29**)	(3.38**)	(0.58)	(-3.47**)	(-1.42)	(0.18)	(0.74)	(2.86**)
β_5	-0.00348	0.0213	-0.0205	-0.1620	-0.0003	-0.0054	-0.0163	-0.1370
	(-0.32)	(0.49)	(-2.14*)	(-5.76**)	(-0.05)	(-1.19)	(-2.48*)	(-2.96**)
Adj-R^2	0.2534	0.1605	0.1235	0.2363	0.4687	0.3013	0.1892	0.2169
F	7.69**	4.75**	3.76**	7.37**	18.46**	9.54**	5.62**	6.59**

Table 5: Regression Model with Roll's Rupee Spread as the Dependent Variable

The table reports the results of the equation below:

$$Y = \alpha + \beta_1 Ln(RupeeVolume) + \beta_2 \sigma_r^2 + \beta_3 Ln(MVE) + \beta_4 Ln(Share\,\mathrm{Pr}ice) + \beta_5 Ln(NumTrades) + \varepsilon$$

All variables are first calculated as the average for each security for each day and we calculate a daily average for each trading day of the sample month, March 2004. Volume is the trading volume multiplied by the (INR) price per share traded. σ_r^2 is the variance of returns, calculated over the past one year. Share Price is the median share price, MVE is the market value of equity, and is calculated as the share price multiplied by the number of shares outstanding. NumTrades is the number of trades for each stock each day. Ln represents the natural logarithmic transform of the variables. In Panel A, we present the full sample results, and in Panel B, we present the results for Volume and Price portfolios. All stock are ranked on the basis of their rupee trading volume (share price), the bottom 25% forms the first quarter (Q1), the 25% to 50% range forms Q2, 50% to 75% Q3 and finally 75% to 100% forms the top quarter Q4. t-statistics are reported in parentheses, and * (**) denotes significance at the 95% (99%) level of confidence.

Panel A

	α	β_1	β_2	β_3	β_4	β_5
Coefficient Estimate	-5.3648	0.8283	356.77	-0.3676	1.8043	-1.9117
t-stat	(-0.95)	(1.04)	(0.93)	(-1.13)	(3.60**)	(-2.38*)
p-value	0.34	0.29	0.35	0.25	0.00	0.01
	Adj-R^2 = 0.1600			F= 16.35**		

Panel B

Estimated	Volume Portfolios				Price Portfolios			
Coefficients	Q_1	Q_2	Q_3	Q_4	Q_1	Q_2	Q_3	Q_4
α	15.4861	-12.6164	-0.0891	-34.7658	0.2049	-1.1567	-5.0259	-70.2322
	(1.17)	(-0.29)	(-0.04)	(-3.95**)	(0.13)	(-0.51)	(-1.44)	(-2.74**)
β_1	-2.6527	0.3651	0.2093	5.6571	0.1470	0.5085	0.8262	3.7542
	(-1.51)	(0.09)	(0.89)	(5.47**)	(0.64)	(1.74)	(2.55*)	(1.46)
β_2	735.7436	893.3372	25.9568	648.9146	6.9810	275.0518	-501.5474	-1449.1673
	(1.14)	(0.90)	(0.17)	(0.75)	(0.13)	(1.88)	(-1.69)	(-0.44)
β_3	-0.6915	-1.9090	0.0136	-0.6626	0.0786	0.2128	-0.2061	-2.5096
	(-1.21)	(-1.51)	(0.14)	(1.62)	(1.37)	(1.51)	(-1.20)	(-1.94)
β_4	5.0753	5.6821	0.4194	-1.6289	0.0817	-0.5709	0.7108	11.4767
	(4.41**)	(3.17**)	(3.75**)	(-3.16**)	(0.48)	(-1.44)	(1.27)	(5.28**)
β_5	1.7036	-0.2867	-0.7268	-6.3971	-0.4920	-0.9757	-1.2934	-5.2253
	(0.86)	(-0.12)	(-3.68**)	(-6.33**)	(-1.87)	(-3.43**)	(-4.14**)	(-2.05*)
Adj-R^2	0.2378	0.1961	0.5845	0.3001	0.2182	0.2496	0.4442	0.3596
F	7.18**	5.83**	28.85**	9.49**	6.53**	7.59**	16.83**	12.34**

Table 6: Estimated Regression Model with Roll's Percent Spread as the Dependent Variable

$$Y = \alpha + \beta_1(RupeeVolume) + \beta_2\sigma_r + \beta_3(MVE) + \beta_4(MedianSharePrice) + \beta_5(NumTrades) + \varepsilon$$

All variables are first calculated as the average for each security for each day and we calculate a daily average for each trading day of the sample month, March 2004. Volume is the trading volume multiplied by the (INR) price per share traded. σ_r is the standard deviation of returns, calculated over the past one year. Share Price is the median share price, MVE is the market value of equity, and is calculated as the share price multiplied by the number of shares outstanding. NumTrades is the number of trades for each stock each day. In Panel A, we present the full sample results, and in Panel B, we present the results for Volume and Price portfolios. All stock are ranked on the basis of their rupee trading volume (share price), the bottom 25% forms the first quarter (Q1), the 25% to 50% range forms Q2, 50% to 75% Q3 and finally 75% to 100% forms the top quarter Q4. t-statistics are reported in parentheses, and * (**) denotes significance at the 95% (99%) level of confidence.

	α	β_1	β_2	β_3	β_4	β_5
Coefficient Estimate	0.0025	-0.0000	-0.0517	-0.0001	0.0002	0.0000
t-stat	(0.16)	(-2.20*)	(-0.13)	(-2.16*)	(1.45)	(1.40)
p-value	0.87	0.04	0.89	0.03	0.13	0.16
	Adj-R^2 = 0.2500			F= 27.86**		

Panel B

Estimated	Volume Portfolios				Price Portfolios			
Coefficients	Q_1	Q_2	Q_3	Q_4	Q_1	Q_2	Q_3	Q_4
α	0.0638	-0.0138	0.0273	0.0754	0.0454	0.0084	0.0283	0.0631
	(5.01**)	(-0.84)	(1.41)	(1.46)	(5.42**)	(0.80)	(1.70)	(0.71)
β_1	-0.0000†	-0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	(-1.53)	(-4.24**)	(0.35)	(0.48)	(1.66)	(1.40)	(1.11)	(-1.57)
β_2	-0.1314	0.1205	-0.4338	-1.4667	-0.1490	0.0848	-0.5960	-3.9645
	(-0.53)	(0.38)	(-0.91)	(-1.03)	(-1.11)	(0.44)	(-1.69)	(-1.38)
β_3	-0.0000	-0.0000	-0.0000	-0.0000	0.0000	0.0000	0.0000	0.0000
	(-0.70)	(-0.91)	(-0.65)	(-0.40)	(-0.03)	(0.85)	(1.30)	(-1.10)
β_4	0.0001	-0.0005	0.0001	-0.0000	-0.0010	0.0001	0.0001	0.0002
	(4.48**)	(-47.23)	(2.60**)	(-0.38)	(-3.07**)	(0.68)	(1.14)	(6.02**)
β_5	-0.0003	0.0003	-0.0001	-0.0000	0.0000	0.0000	0.0000	0.0000
	(-1.96**)	(4.52)	(-0.86)	(-0.61)	(-2.24*)	(-1.88)	(-1.88)	(1.54)
Adj-R^2	0.2741	0.9597	0.1019	0.0001	0.1492	0.0354	0.0815	0.2666
F	8.48**	473.72**	3.25**	0.79	4.47**	1.73	2.76*	8.34**

†: Numbers in this table that are reported as zero up to the fourth decimal place are not exactly equal to zero, but may have numerical integers in the fifth or higher decimal place.

Table 7: Estimated Regression Model with Roll's Rupee Spread as the Dependent Variable

$$Y = \alpha + \beta_1(RupeeVolume) + \beta_2\sigma_r + \beta_3(MVE) + \beta_4(MedianSharePrice) + \beta_5(NumTrades) + \varepsilon$$

All variables are first calculated as the average for each security for each day and we calculate a daily average for each trading day of the sample month, March 2004. Volume is the trading volume multiplied by the (INR) price per share traded. σ_r is the standard deviation of returns, calculated over the past one year. Share Price is the median share price, MVE is the market value of equity, and is calculated as the share price multiplied by the number of shares outstanding. NumTrades is the number of trades for each stock each day. In Panel A, we present the full sample results, and in Panel B, we present the results for Volume and Price portfolios. All stock are ranked on the basis of their rupee trading volume (share price), the bottom 25% forms the first quarter (Q1), the 25% to 50% range forms Q2, 50% to 75% Q3 and finally 75% to 100% forms the top quarter Q4. t-statistics are reported in parentheses, and * (**) denotes significance at the 95% (99%) level of confidence.

Panel A

	α	β_1	β_2	β_3	β_4	β_5
Coefficient Estimate	0.0052	-0.0000	-8.8898	-0.0001	0.0122	0.0002
t-stat	(0.01)	(-2.77**)	(-0.39)	(-2.90**)	(15.63**)	(2.05*)
p-value	0.99	0.00	0.69	0.00	0.00	0.04
	Adj-R^2 = 0.3370			F= 51.88**		

Panel B

Estimated	Volume Portfolios				Price Portfolios			
Coefficients	Q_1	Q_2	Q_3	Q_4	Q_1	Q_2	Q_3	Q_4
α	0.8985	0.1235	0.9819	3.5787	0.6139	0.1260	2.7550	3.3051
	(1.72*)	(0.15)	(2.85*)	(1.91*)	(2.00*)	(0.20)	(2.97**)	(0.67)
β_1	-0.0000†	-0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	(-4.59**)	(-3.63**)	(0.17)	(0.70)	(0.74)	(1.41)	(1.24)	(-1.99*)
β_2	-1.0514	7.2325	-12.0326	-69.7894	-0.6975	15.3367	-59.6586	-235.8605
	(-0.11)	(0.42)	(-1.42)	(-1.35)	(-0.14)	(1.30)	(-3.04**)	(-1.48)
β_3	-0.0003	-0.0000	-0.0000	-0.0000	0.0000	0.0000	0.0000	-0.0000
	(-0.32)	(-0.18)	(-1.58)	(-0.75)	(0.05)	(0.93)	(0.00)	(-1.33)
β_4	0.0357	0.0245	0.0038	0.0001	0.0086	0.0028	0.0060	0.0144
	(49.34**)	(47.91**)	(11.16**)	(0.15)	(0.73)	(0.36)	(1.30)	(7.91**)
β_5	0.0123	0.0081	-0.0006	-0.0000	-0.0008	-0.0008	-0.0002	0.0007
	(1.93*)	(2.52**)	(-2.01*)	(-0.83)	(-1.26)	(-1.85)	(-1.96*)	(1.93)
Adj-R^2	0.9643	0.9618	0.6935	0.0101	0.0262	0.0292	0.1411	0.3934
F	535.26**	498.91**	45.81**	0.79	1.53	1.60	4.25**	14.10**

†: Numbers in this table that are reported as zero up to the fourth decimal place are not exactly equal to zero, but may have numerical integers in the fifth or higher decimal place.

Table 8: Analysis of Snapshot Data

This table reports the distribution of variables calculated from snapshots of the limit order book (LOB) for all stocks that traded on the NSE between March 1, 2004 and March 31, 2004. The snapshots are recorded at hourly interval, beginning at 11 a.m. and ending at 2 p.m. Ask (Bid) price is the lowest (highest) sell (buy) price amongst all outstanding orders in the LOB at each snapshot time. Depth is the quoted quantity associated with the ask and bid prices. Spread is the rupee spread, calculated as the difference between ask and bid prices for each stock-snapshot combination. Percentage spread is rupee spread divided by quote mid-point. For each order we define its age as the difference between the snapshot date-time and the order date-time. It note that the snapshot files are sometimes recorded with a lag due to the time taken to record all of the outstanding orders across all stocks into one file. Thus, a 1 p.m. snapshot may actually contain some orders that entered the system after 1 p.m. By our definition then, calculated age of such an order will be negative. We retain them in the sample as they represent the freshest orders. This accounts for the minimum age for both bid and ask orders being negative.

N=59997	Mean	Stdev	Min	Percentile							Max
				5th	10th	25th	50th	75th	90th	95th	
Ask Price	183.66	616.61	0.30	3.30	6.45	18.45	58.95	185.00	439.95	623.40	15870.00
Bid Price	181.83	611.60	0.00	2.95	6.15	18.05	57.95	183.00	436.00	621.25	15815.00
Ask Depth	602.90	3001.46	1.00	2.00	10.00	50.00	100.00	427.00	1000.00	2000.00	415433
Bid Depth	561.58	1813.61	1.00	1.00	9.00	50.00	100.00	500.00	1000.00	2000.00	100000
Spread	1.82	8.50	0.00	0.05	0.05	0.15	0.45	1.30	3.40	5.50	293.00
Percent Spread	0.02	0.05	0.00	0.00	0.00	0.00	0.01	0.02	0.06	0.11	2.00
Ask Age	2582.47	3640.98	-10815.00	-398.00	-275.00	67.00	1299.00	3879.00	7862.00	10539.00	14675.00
Bid Age	2158.09	3445.46	-10815.00	-424.00	-320.00	-24.00	956.00	3309.00	7016.00	9879.00	14669.01

Table 9: Analysis of time of day variations

This table reports the intra-day variations in the variables calculated from snapshots of the limit order book (LOB) for all stocks that traded on the NSE between March 1, 2004 and March 31, 2004. The snapshots are recorded at hourly interval, beginning at 11 a.m. and ending at 2 p.m. Ask (Bid) price is the lowest (highest) sell (buy) price amongst all outstanding orders in the LOB at each snapshot time. Depth is the quoted quantity associated with the ask and bid prices. Spread is the rupee spread, calculated as the difference between ask and bid prices for each stock-snapshot combination. Percentage spread is rupee spread divided by quote mid-point. For each order we define its age as the difference between the snapshot date-time and the order date-time.

N=15009	Time	Mean	Min	Percentile							Max
				5th	10th	25th	50th	75th	90th	95th	
Ask Price	11 a.m.	184.99	0.30	3.35	6.60	18.70	59.45	185.45	440.55	624.00	15870
	12 noon	184.24	0.30	3.30	6.40	18.45	58.95	185.40	439.75	623.00	15825
	1 p.m.	183.09	0.30	3.30	6.35	18.20	58.18	183.95	435.88	621.00	15750
	2 p.m.	182.32	0.30	3.25	6.10	17.85	57.20	181.95	433.80	622.00	15460
Bid Price	11 a.m.	182.86	0.25	3.00	6.30	18.20	58.15	183.60	435.60	622.00	15815
	12 noon	182.42	0.00	2.95	6.20	18.05	58.00	183.85	435.85	621.00	15710
	1 p.m.	181.39	0.00	2.95	6.00	17.80	57.28	182.00	432.78	619.55	15610
	2 p.m.	180.67	0.00	2.90	5.90	17.50	56.50	180.50	427.55	619.35	15410
Ask Depth	11 a.m.	567.98	1.00	2.00	10.00	50.00	100.00	400.00	1000	2000	52000
	12 noon	603.92	1.00	3.00	10.00	50.00	100.00	450.00	1000	2000	136674
	1 p.m.	631.34	1.00	2.00	10.00	50.00	100.00	429.00	1000	2000	415433
	2 p.m.	604.36	1.00	2.00	10.00	50.00	100.00	401.00	1000	2000	96000
Bid Depth	11 a.m.	520.88	1.00	1.00	6.00	50.00	100.00	500.00	1000	2000	50000
	12 noon	573.52	1.00	1.00	10.00	50.00	100.00	500.00	1000	2000	50000
	1 p.m.	593.63	1.00	1.00	9.00	50.00	100.00	500.00	1000	2003	100000
	2 p.m.	558.16	1.00	1.00	10.00	50.00	100.00	500.00	1000	2000	50000
Spread	11 a.m.	2.12	0.00	0.05	0.10	0.20	0.50	1.50	4.00	6.90	293
	12 noon	1.82	0.30	0.05	0.05	0.15	0.45	1.30	3.45	5.70	249
	1 p.m.	1.70	0.00	0.05	0.05	0.15	0.40	1.20	3.00	5.00	249
	2 p.m.	1.64	0.00	0.05	0.05	0.15	0.40	1.20	2.95	4.90	255
% Spread	11 a.m.	0.03	0.00	0.00	0.00	0.01	0.01	0.03	0.08	0.12	0.40
	12 noon	0.02	0.00	0.00	0.00	0.00	0.01	0.02	0.07	0.11	0.20
	1 p.m.	0.02	0.00	0.00	0.00	0.00	0.01	0.02	0.06	0.10	0.20
	2 p.m.	0.02	0.00	0.00	0.00	0.00	0.01	0.02	0.05	0.09	0.20
Ask Age	11 a.m.	677	-10815	-528	-410	-166	475	2018	3273	3651	3870
	12 noon	2663	-475	-58	63	928	1935	4303	6474	7073	7470
	1 p.m.	3161	-598	-342	-257	23	1233	6173	9545	10461	11075
	2 p.m.	3837	-585	-413	-301	48	1576	6492	12515	13833	14675
Bid Age	11 a.m.	424	-10815	-562	-427	-226	254	1600	2992	3463	3866
	12 noon	2284	-476	-101	-3	758	1437	3626	6041	6832	7467
	1 p.m.	2695	-598	-357	-280	-44	865	5103	8943	10190	11066
	2 p.m.	3235	-585	-430	-339	-58	1069	4739	11502	13441	14669

Table 10: Regression Model with Snapshot Mean Percent Spread as the Dependent Variable

The table reports the results of the equation below:

$$Y = \alpha + \beta_1 Ln(RupeeVolume) + \beta_2 \sigma_r^2 + \beta_3 Ln(MVE) + \beta_4 Ln(Share\,\mathrm{Pr}ice) + \beta_5 Ln(NumTrades) + \varepsilon$$

All variables are first calculated as the average for each security for each day and we calculate a daily average for each trading day of the sample month, March 2004. Volume is the trading volume multiplied by the (INR) price per share traded. σ_r^2 is the variance of returns, calculated over the past one year. Share Price is the median share price, MVE is the market value of equity, and is calculated as the share price multiplied by the number of shares outstanding. NumTrades is the number of trades for each stock each day. Ln represents the natural logarithmic transform of the variables. t-statistics are reported in parentheses, and * (**) denotes significance at the 95% (99%) level of confidence.

	α	β_1	β_2	β_3	β_4	β_5
Coefficient Estimate	0.0634	-0.0027	2.2644	0.0023	-0.0026	-0.0035
t-stat	(24.01**)	(-7.26**)	(12.68**)	(15.08**)	(-11.16**)	(-9.47**)
p-value	0.00	0.00	0.00	0.00	0.00	0.00
	Adj-R^2 = 0.1801			F= 1451.94**		

Table 11: Regression Model with Snapshot Mean Raw Spread as the Dependent Variable

The table reports the results of the equation below:

$$Y = \alpha + \beta_1 Ln(RupeeVolume) + \beta_2 \sigma_r^2 + \beta_3 Ln(MVE) + \beta_4 Ln(Share\,\mathrm{Pr}ice) + \beta_5 Ln(NumTrades) + \varepsilon$$

All variables are first calculated as the average for each security for each day and we calculate a daily average for each trading day of the sample month, March 2004. Volume is the trading volume multiplied by the (INR) price per share traded. σ_r^2 is the variance of returns, calculated over the past one year. Share Price is the median share price, MVE is the market value of equity, and is calculated as the share price multiplied by the number of shares outstanding. NumTrades is the number of trades for each stock each day. Ln represents the natural logarithmic transform of the variables. t-statistics are reported in parentheses, and * (**) denotes significance at the 95% (99%) level of confidence.

	α	β_1	β_2	β_3	β_4	β_5
Coefficient Estimate	-2.8449	0.2983	381.9695	-0.2134	2.0106	-1.3521
t-stat	(-4.23**)	(3.17**)	(8.39**)	(-5.55**)	(33.86**)	(-14.21**)
p-value	0.00	0.00	0.00	0.00	0.00	0.00
	Adj-R^2 = 0.1401			F= 1075.54**		

Table 12: Estimated Regression Model with Snapshot Mean Percent Spread as the Dependent Variable

$$Y = \alpha + \beta_1(RupeeVolume) + \beta_2\sigma_r + \beta_3(MVE) + \beta_4(MedianShare\,\mathrm{Pr}ice) + \beta_5(NumTrades) + \varepsilon$$

All variables are first calculated as the average for each security for each day and we calculate a daily average for each trading day of the sample month, March 2004. Volume is the trading volume multiplied by the (INR) price per share traded. σ_r is the standard deviation of returns, calculated over the past one year. Share Price is the median share price, MVE is the market value of equity, and is calculated as the share price multiplied by the number of shares outstanding. NumTrades is the number of trades for each stock each day. t-statistics are reported in parentheses, and * (**) denotes significance at the 95% (99%) level of confidence.

	α	β_1	β_2	β_3	β_4	β_5
Coefficient Estimate	0.0112	0.0001	0.1708	0.0000	-0.0001	-0.0001
t-stat	(20.97**)	(17.69**)	(12.77**)	(-6.68**)	(-13.33**)	(-22.38**)
p-value	0.00	0.00	0.00	0.00	0.00	0.00
	Adj-R^2 = 0.3026			F= 2023.56**		

Table 13: Estimated Regression Model with Snapshot Mean Raw Spread as the Dependent Variable (*Y*) (Model 4)

$$Y = \alpha + \beta_1(RupeeVolume) + \beta_2\sigma_r + \beta_3(MVE) + \beta_4(MedianShare\,\mathrm{Pr}ice) + \beta_5(NumTrades) + \varepsilon$$

All variables are first calculated as the average for each security for each day and we calculate a daily average for each trading day of the sample month, March 2004. Volume is the trading volume multiplied by the (INR) price per share traded. σ_r is the standard deviation of returns, calculated over the past one year. Share Price is the median share price, MVE is the market value of equity, and is calculated as the share price multiplied by the number of shares outstanding. NumTrades is the number of trades for each stock each day. In Panel A, we present the full sample results, and in Panel B, we present the results for Volume and Price portfolios. t-statistics are reported in parentheses, and * (**) denotes significance at the 95% (99%) level of confidence.

Panel A

	α	β_1	β_2	β_3	β_4	β_5
Coefficient Estimate	-0.3194	-0.0001	-3.9490	-0.0001	0.1223	0.0002
t-stat	(-2.91**)	(-24.53**)	(-1.44)	(-22.74**)	(128.44**)	(18.84**)
p-value	0.00	0.00	0.14	0.00	0.00	0.00
	Adj-R^2 = 0.3442			F= 3467.34**		

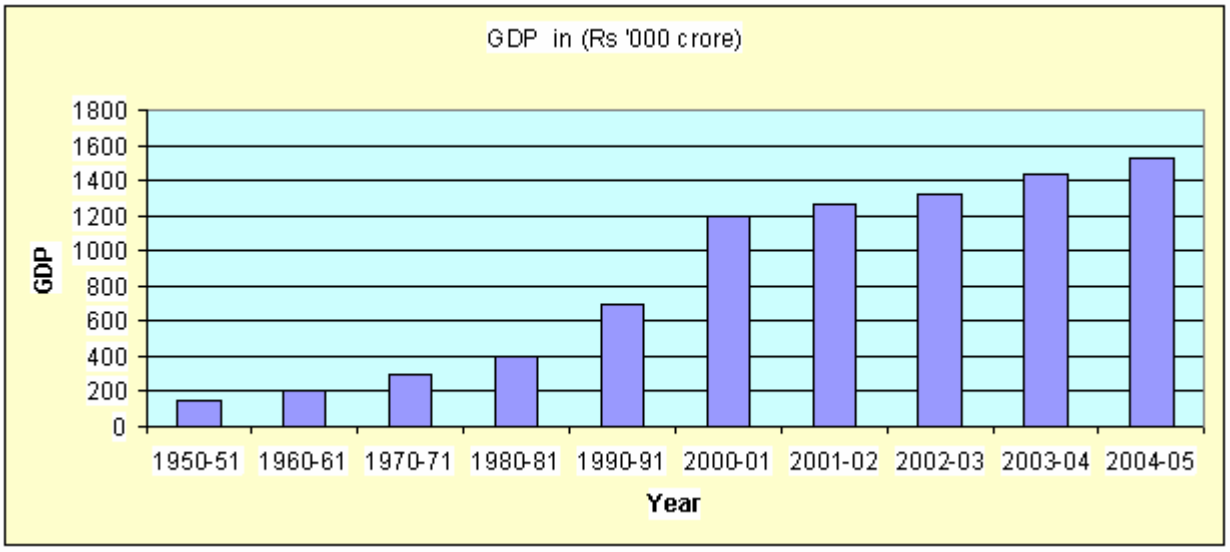

Figure 1: India's GDP growth

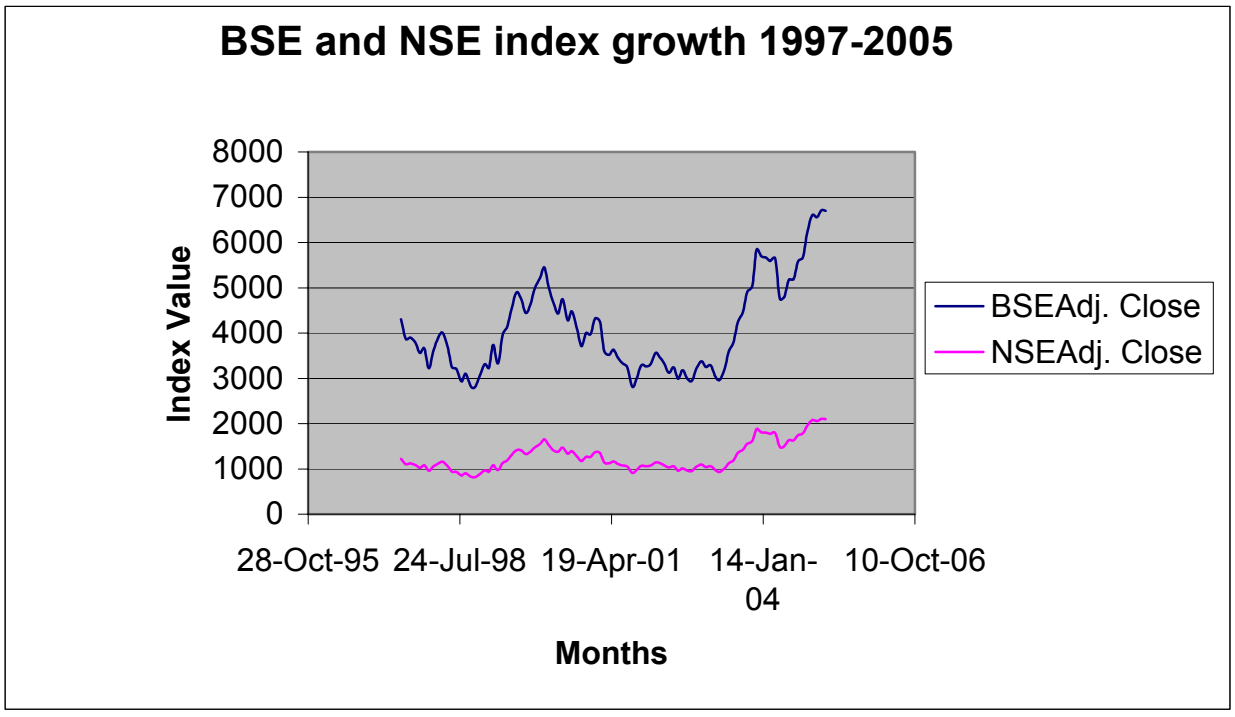

Figure 2: Growth rate of the two major stock market indexes

Appendix I: The NSE data files

Data organization

One (or more) CD is produced for each month of the operations of NSE, containing 7 files that are organized as sub-directories. Hence, in the first level directory (which has the CD name in the yyyymm format), the user will find the following sub-directories: Bhavcopy, Circulars, Index, Masters, Snapshots, Software and Trades. Each of contain data files. All data files are plain ascii format gunzipped files.

Bhav copy file

The Bhavcopy contains a directory structure organized as 1999/Mar/date.gz where the date is represented as yyyymmdd. Thus, the file for 5th March, 1999 would be located at Bhavcopy/1999/Mar/19990305.gz. This gives us one file for each trading day. The lines in this file have 11 fields per line (each line is one observation), delimited by the pipe "|" character. These fields are:
1. Symbol - Symbol for each company given by NSE; e.g. "SBIN" for State Bank of India or "INFOSYSTCH" for Infosys. A master table of all symbols is found in the Masters database.

2. Series - Series Symbol for each security given by NSE; e.g. "EQ" for common stock. A symbol and a series uniquely identify a security. A master table of all symbol plus series combinations is in the Masters database.

3. Open price - The opening price of the day. On some days, when the pre-opening call auction has been used, and if the security trades in the call auction, the Open price is the (single price) from the call auction. Otherwise, the Open price is the price at the first trade of the day.

4. High price - The highest traded price of the day.

5. Low price - The lowest traded price of the day.

6. Closing price - This is the official closing price reported by NSE.

7. LTP - The last traded price of the day. In general, this need not be equal to the official closing price because the official closing price is calculated using a variety of rules (e.g. averaging of trades over the last 30 minutes), etc.

8. Traded quantity - The number of shares traded in the day.

9. Value of shares traded - The rupee value of all shares traded in the day. The volume weighted average (VWA) price is field 9 divided by field 8.

10. Number of trades. - Number of trades that took place in the normal market (i.e. excluding trades in the auction market. The average trade size is field 8 divided by field 10 (in number of shares) or field 9 divided by field 10 (in rupees).

11. Corporate action flags - Ex date indicators, e.g. XD for a dividend, etc.

Index file

The Index directory contains databases connected with stock market indexes. Both end-of-day (summary file) and intra-day information (intra-day file) is available. Three indexes are covered: Nifty, Nifty Junior and Defty, these are found in directories called Nifty, Junior and Defty. Nifty is the main stock market index in India; it is composed of the top 50 highly liquid stocks in India, which make up roughly half of the market capitalization of India. Nifty Junior is the second tier of 50 less liquid stocks. Nifty Junior accounts for around 10% of the market

capitalization of India. Nifty and Nifty Junior are always disjoint sets: there is never any common index member. Defty is the same as Nifty, expressed in dollars. The end-of-day data for the month of March '99 for Nifty is found in the file Index/Nifty/1999/Mar/summary.gz. Intra-day data for Nifty for the 5th of March, 1999 is found in the file Index/Nifty/1999/Mar/19990305.gz. Files for Junior and Defty are similarly organised.

The summary file

The summary file has lines with 14 fields per line (day).
1. Name of index - This identifies the index, e.g. Nifty.

2. Date - The date, formatted as yyyymmdd.

3. Open - The opening level of the index.

4. High - The highest level of the index in the day.

5. Low - The lowest level of the index in the day.

6. Close - The official closing price of the index: this is the reference rate that is used for measuring the expiration value of index futures or index options.

7. Number of shares traded - The sum of the number of shares traded of each of the components of the index.

8. Value of shares traded - The sum of value of the shares traded of each of the components of the index.

9. Market capitalization - The sum of the market capitalization of all the components of the index.

10. Impact cost at a program trade of Rs.2.5 million - Average impact cost (measured in percent) faced when doing program trades on this index for a transaction size of Rs.2.5 million. Typically, the average is taken over the values seen in three snapshots of the limit order book on the day.

11. Impact cost at Rs.5 million - The same, at Rs.5 million.

12. Impact cost at Rs.10 million - The same, at Rs.10 million.

13. Impact cost at Rs.20 million - The same, at Rs.20 million.

14. Returns on TR index - Returns on the index, inclusive of dividends.

The intra-day files

The intra-day files show a fresh calculation of the market index every time a trade takes place for an index component. Most of the time, more than one trade takes place in a given second, so multiple records are found for the same second. Hence, we often see days where there are more than 100,000 observations for Nifty. The records shown are in correct time-sorted order, even though it appears that they all have the same timestamp.
1. Index name - This is a string identifying the index, e.g. "Nifty".

2. Timestamp - This is formatted as hh:mm:ss

3. Index level - The level of the index, rounded off to 10 decimal places.

Masters file

Each CD shows a snapshot of the Masters table as of the end of the preceding month. Hence, the Master applicable for March 1999 is found in the file 199903/Masters/1999/Mar/19990228.gz.
The 5 fields in this database are:

ISIN - The International Security Identification Number (ISIN), if it has been allocated

Symbol - NSE's `symbol', e.g. ABB.

Series - NSE's `series', e.g. AE

Name - A descriptive string about the product.

Deleted - A flag which takes the value `N' if the security has not been deleted.

Order book snapshots file

NSE is an open electronic limit order book market (OELOB). The order book snapshots are taken at given times each day and those times are identified in the filename. For example, for 5 March, 1999 snapshots data are stored in the directory 19903/Snapshots/19990305. The files that are found inside this have names of the form hhmmss.gz, to convey the time at which the snapshot was taken, like 110000.gz, 130000.gz and 140000.gz. These are order book snapshots at 11 AM, 1 PM and 2 PM. These files contain one record per line, and each record pertains to one limit order. The files are sorted by price. They have 14 fields per record:

1. Order ID number - This is unique code given to every limit order on NSE.

2. Symbol - The symbol for the security.

3. Series - The series for the security. Every security is uniquely defined once its symbol and series have been specified. A symbol of GLAXO and a series of EQ denote common stock of Glaxo.

4. Quantity - The size of the limit order.

5. Price - The limit price on the limit order.

6. Timestamp - The time at which the order was placed (or last modified). This is formatted as hh:mm:ss.

7. Buy/sell - This is B for buy limit orders and S for sell limit orders.

8. Day flags - This is a set of four boolean (yes/no) flags: (1) Day order, (2) Good till date, (3) Cancel, (4) Immediate/Cancel. The commonest value found is ynnn.

9. Quantity flags - This is a set of three quantity flags: (1) Minimum fill, (2) All or none, (3) Disclosed quantity. The commonest value found is nnn.

10. Price flags - This is a set of three price flags: (1) At the open (ATO) price, (2) Market price, and (3) Stop loss order. The commonest value is nnn.

11. Book type - There are two books that can be used: RL and SL. RL is the most common.

12. Minimum fill quantity - If the order specifies a minimum fill, then this field shows the minimum fill quantity specified.

13. Quantity disclosed - If the order discloses a smaller quantity as compared with the true order size, then this field shows the smaller quantity that is to be disclosed.

14. Date for GTD - GTD orders need to specify a date until which the order is good: that date is specified here.

Trades file

This database records every trade that took place. Each day's trades are recorded in a file, organized as follows:

1. Trade ID number - A number for each trade, starting at 1 for each day and going higher. Sorting by this field gives a sequence of prices, which are sorted in time order.

2. Symbol - The symbol of the security traded.

3. Series - The series of the security traded.

4.Timestamp – The time at which a trade took place, formatted as hh:mm:ss. Often multiple trades are matched within the same second, in which case we see multiple records with the same timestamp.

5. Price - The trade price.

6. Quantity traded - The number of shares transacted in a trade.

Database of circulars

Circulars are a formal method of communication between NSE and its member brokerage firms. Each CD contains a comprehensive set of circulars issued in the month. Every development on the market in terms of market design is documented in these circulars.

*****