Historical Data Dissemination of Capital Market Segment from NSE

From late 1995 onwards, NSE has been India's largest equity market. This has led to a series of requests for historical data, for use in research and analysis amongst market participants, researchers, and policy makers. In response to these requests, NSE has setup a formal framework for historical data dissemination. .

Organisation of Historical Data

Each CD/DVD is identified by the name of the starting directory in yyyymm format. When you mount the CD/DVD for March 1999, the starting directory is 199903.

Underneath this, there are 7 directories:

Bhav copy
: Summary information about each security for each trading day.

Index
: Information about stock market indexes.

Masters
: Database masters, listing out symbols, series, ISINs, etc.

Snapshots
: Snapshots of the limit order book at many timepoints in the month.

Trades
: A database of every single trade that took place.

Circulars
: A database of all circulars issued by NSE or NSCCL. This serves as a formal documentation of the history of market design at NSE.

Bhav copy database

The Bhavcopy directory contains a directory structure organised 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.

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, "N1" for first debenture issue, "W" for warrants, etc. Once a symbol and a series have been specified, a security is uniquely known. A master table of all symbol+series combinations which are traded 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 which 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 database

The Index directory contains databases connected with stock market indexes. Both end-of-day and intra-day information is available. Following indexes are covered: Nifty, CNX Midcap and Defty, these are found in directories called Nifty, CNX Midcap 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 capitalisation of India. 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.

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 capitalisation
   The sum of the market capitalisation 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. Impact cost is provided at Rs.5 million, Rs.10 million and Rs.20 million*
11. Returns on TR index
   Returns on the index, inclusive of dividends.

** Data not available after 2003*

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 database

Each CD/DVD 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, e.g. INE117A01014.
Symbol
   NSE's `symbol', e.g. ABB.

Series
: NSE's `series', e.g. BE

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 database

NSE is a limit order book market, also known to economists as the `open electronic limit order book market (OELOB)', or to practitioners as a market based on `electronic order matching'. Liquidity on the OELOB market is embedded in the limit orders present at any point in time; these limit orders (the right to trade against them, without any obligation) are free options which others can exploit.

Measurement of this liquidity is possible with high accuracy using "order book snapshots": pictures of the complete limit order book at a point in time. This is discrete, in only conveying the picture at a few time points in the day. However, at these time points, a variety of questions about liquidity can be accurately answered. The order book snapshot can yield the bid-ask spread, and it can be used to measure market impact cost for buying or selling any desired transaction (or for testing whether a desired transaction is feasible).

The order book snapshots for 5 March, 1999 are stored in the directory 199903/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/DVD contains 110000.gz, 130000.gz and 140000.gz. These are order book snapshots at 11 AM, 12:00 PM, 1 PM and 2 PM.

These files are databases with 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 a field like 9801050079959 which is a 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 denotes 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 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, (3) Stop loss order. The commonest value is nnn.
11. Book type
    There are two books which 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 meant 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 database

This is a database about every trade that took place. If NSE does 4,00,000 trades in a day, the dataset for that day would have 4,00,000 trades.

Information for the trades of each day is kept in a distinct file. The information for 5 March, 1999 is found in 199903/Trades/1999/Mar/19990305.gz

The files are organised as follows:

1. Trade ID number
   A unique number for each trade, the files on the CD/DVD are sorted by this trade ID.
2. Symbol
   The symbol of the security traded.
3. Series
   The series of the security traded.
4. Timestamp
   The time at which the trade took place, formatted as hh:mm:ss. Many times, many trades are matched within the same second, in which case we see multiple records with the same timestamp. The pattern of seeing a large number of trades for the same stock towards the start or the end of the market is owing to the uniform-price order matching at the end of the call auctions.
5. Price
   The price at this trade.

6. Quantity traded
   The number of shares transacted in this trade.

Database of circulars

Circulars are a formal method of communication between NSE and its member brokerage firms. Each CD/DVD 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.

All the circulars for a month can be accessed by pointing your web browser to the file index.html. For example, the circulars for March, 1999 can be read by loading up 199903/Circulars/1999/Mar/Index.html into your browser.