# CHAPTER 4

# DATA AND RESEARCH METHODOLOGY

## 4.1 Data and Variable Construction

We use high frequency intraday transactions and order-book snapshot data for equity and options markets separately from NSE for a period of two years from April, 2010 to March, 2012. The transactions data has record of all transactions that took place for the period under study. For the stocks, the trade data comes in a single file with information regarding each and every transaction (with time stamp) happened on that day. NSE collects the snapshots data of the limit order book at four different time periods of the day at 11 A.M., 12 noon, 1 P.M., and 2 P.M. which gives all the information regarding the quotes (with time stamp) placed by various market participants on that particular day. Similarly for the options, the trade & snapshot data is obtained for all the option series except that limit order book snapshot data is collected at five different time periods of the day as provided by NSE at 11 A.M., 12 Noon, 1 P.M., 2 P.M., and 3 P.M. The operating time of stock and options market is synchronized from 9.15 A.M. to 3.30 P.M. for our sample period.[1]

For our sample period, the number of stocks traded on NSE is 1501. The total number of trading days in the sample period is 504.[2] Following prior literature, we apply certain data filters for our equity dataset. Our first filter is to delete all those firms with a price less than Rs. 10 or price greater than Rs. 2500 which results in a sample size of 1484. We apply this filter following

---

[1] The stock market follows a dual auction mechanism of call and continuous auction with call auction at the opening from 9.00-9.15 AM and continuous auction throughout the day from 9.15-3.30.

[2] There are 249 trading days in 2010-11 and 255 trading days in 2011-12.

CRS (2000) to avoid any contaminating effect of tick size. CRS (2000) and Fabre and Frino (2004) argue that stocks with infrequent trades do not provide reliable information. To remove less frequently traded stocks, we delete all those stocks with less than 200 active trading days over our sample period resulting in a sample size of 1404 firms. Using the criteria followed by NSE to identify less liquid stocks, we delete all those stocks with an average daily trading volume less than 10,000 shares and number of trades less than 50 in a quarter which reduces our final sample to 960 firms.[3]

Similar to the equity dataset, following Cao and Wei (2010), we screen the options dataset. The total number of options listed on NSE is 256. We eliminate all those observations having zero trading volume and also to eliminate excessive impact of very less number of trades, we delete options having less than five contracts traded on a given trading day. This screening criterion reduces the sample to 201 options. To safeguard the representativeness of the sample under study, we eliminate those options with very short or very long maturity. To be more precise, we delete all those options with a maturity less than 7 days or more than 365 days. This reduces our sample size to 194. Further, to avoid any pricing related issues caused due to moneyness, we eliminate all those observations with a moneyness lying in the range 0.9 to 1.1[4] which reduces the sample to 191. After the above criteria are satisfied, we filter the sample further including only those options which have at least 300 option observations in a year resulting in a final sample size of 143.

For our analysis of the equity market, we estimate six liquidity proxies; Spread, Pspread, Depth, and Roll (1984) measures are constructed using intraday data. High low spread

---

[3] NSE uses this criterion to separate out less liquid stocks and they are traded under a different window.
[4] Moneyness of an option is the ratio of exercise price of an option to its underlying stock price.

(Spread_HL) estimator of Corwin and Schultz (2012) and Amihud (2002) illiquidity measures are constructed using daily data. The snapshot files containing the limit order book information lists all outstanding orders which are identified as buy or sell at the time when the snapshot is recorded. We estimate the spread and depth measures by extracting the highest ask price (and associated quantity) and the lowest bid price (associated quantity) at each of the snapshot record time on a given trading day.[5] For the options market, we construct four liquidity measures; Spread, Pspread, Depth, and Volume.

To overcome intraday idiosyncrasies and to promote better synchronicity, and also to conveniently manage the data, following CRS (2000) and Brockman and Chung (2002), each liquidity measure for each stock is constructed as an equally weighted average of intraday liquidity measures. Thus for each of the 960 stocks, the sample consists of at most 504 observations, one each for each of the trading days during the sample period. In case of options, we use the volume-weighted average of the intraday liquidity measure over all the option series to arrive at a daily liquidity measure for each option. By doing this, each of the 143 listed options in the sample has at most 504 observations over the sample period.

The construction of the variables for our study is as follows:

a) **Absolute Spread (Spread)**: It is estimated as the difference between highest bid and lowest ask price quoted by the market participants at each snapshot record time. It is one of the high-frequency liquidity measures used in most of the liquidity studies to measure the liquidity of a stock.

$$Spread = P_{Ask} - P_{Bid}$$

Where, $P_{Ask}$ is the lowest asked price quoted and $P_{Bid}$ is the highest bid price quoted in the interval concerned. The absolute quoted spread is in rupee units.

**(b) Percentage Spread (*Pspread*)**: This high frequency liquidity measure is computed as the ratio of absolute quoted spread to the average of asked price and bid price in a given interval.

$$Pspread = \frac{P_{Ask} - P_{Bid}}{P_{Mid}}, \text{where } P_{Mid} \text{ is the average of asked and bid prices.}$$

**(c) Quoted Depth (*DEPTH*)**: This high frequency liquidity measure is signifies a stock's capability to take in the demand for buy and sell orders without much price impact. It is computed as the average quantity of asked shares and the bid shares. $QDEPTH = (Q_{Ask} + Q_{Bid})/2$ . Where, $Q_{Ask}$ is the quantity of asked shares and $Q_{Bid}$ is the quantity of bid shares. It is quantified by number of shares.

**(d) Roll's Spread (*Roll*)**: The Roll's spread is based on the assumption that there would not be any serial correlation in observed price changes when trading costs are zero. It is given by

$$Roll = 2\{-Cov(\Delta P_t, \Delta P_{t-1})\}^{1/2}$$

Where, $P_t$ is the trade price at time t, and $Cov(\Delta P_t, \Delta P_{t-1})$ is the serial covariance between successive price changes.

**(e) High Low Spread Estimator (*Spread_HL*):** This measure is a recent measure proposed by Corwin and Schultz (2012) to estimate the spread from daily low high prices. It is given by

$$Spread\_HL = \frac{2(e^\alpha - 1)}{1 + e^\alpha}$$

Where, $\alpha = (\sqrt{2\beta} - \sqrt{\beta}/(3 - 2\sqrt{2})) - \sqrt{\frac{\gamma}{3-2\sqrt{2}}}$ , $\beta = (Ln\left(\frac{H_t}{L_t}\right))^2 + (Ln\left(\frac{H_{t+1}}{L_{t+1}}\right))^2$ and $\gamma =$

$(Ln\left(\frac{H_{t,t+1}}{L_{t,t+1}}\right))^2$. Here, $H_t$ and $L_t$ are the daily high and low prices.

**(f) Amihud's Liquidity (*Amihud*):** Amihud (2002) compute liquidity measure which captures the daily price impact associated with a stock per one dollar of trading volume and it is defined as follows:

$$Amihud_t = \frac{|Return_t|}{Volume_t}$$

Where, Amihud's liquidity measure on day t is calculated as the ratio of absolute return of a security on day *d* to the total traded volume of that security on that day.

**(g) Market Liquidity (*LIQ_M*):** Market liquidity is calculated as the weighted average of the liquidity of all firms in the sample excluding the firm that is analyzed. We use equally-weighted and value-weighted average in computing this variable.

**(h) Return:** Return of a security measures the gain or loss made by a stock in a particular period of time. It is computed as the ratio of difference in price of a security over two successive time periods to its initial price. $Return_t = (Price_t - Price_{t-1})/Price_{t-1}$. In calculating market return, we use the weighted average (equally-weighted and value-weighted) of returns of all the firms in the sample excluding the firm being analyzed.

**(i) Volatility:** Volatility is a measure which captures the variation in price of a stock over time and is computed here as the squared returns of a stock each day.

For testing the supply and demand-side sources of liquidity commonality we use daily data for a period of 12 years from 2001-2012 to construct quarterly measures of liquidity commonality. As the time period is very long and we don't have the intra-day data for this long period we depend on Amihud's liquidity measure (*LIQ*) to capture liquidity commonality.[6] We use the $R^2$ of regressions of the individual stock liquidity on market liquidity to compute the liquidity commonality measure.[7] First we perform the following filtering regression for each stock *J* based on observations on each day *d* within each month *t*:

$$LIQ_{J,t,d} = \propto_{J,t} LIQ_{J,t,d-1} + \sum_{k=1}^{5} \beta_{J,t,k} Dum_k + \varepsilon_{J,t,d} \qquad (EQ1)$$

Here $Dum_k$ is the weekly dummy for controlling seasonality. We have lagged liquidity measure as an explanatory variable and take the residuals of daily liquidity as our interest lies in examining if the changes in individual liquidity of firms co-move. We use the innovations from *EQ1* to obtain quarterly measures of liquidity commonality for each firm by making use of $R^2$ from the following regressions, using daily observations within a quarter:

$$\varepsilon_{J,t,d} = \alpha_{J,t} + \gamma_{J,t,1}\varepsilon_{M,t,d} + \gamma_{J,t,2}\varepsilon_{M,t,d-1} + \gamma_{J,t,3}\varepsilon_{M,t,d+1} + \vartheta_{J,t,d} \qquad (EQ2)$$

Where $\varepsilon_{M,t,d}$ is the market residuals estimated as a sum total of estimated residuals $\varepsilon_{J,t,d}$ obtained from *EQ1* for all the firms in the sample. We also include one lead and one lag market residuals. This $R_t^2$ measure capturing commonality is not appropriate to use as a dependent variable in the regressions to follow because its value ranges between 0 and 1. So, we apply the logistic transformation of the commonality measure (Morck, Yeung, and Yu, 2000) as shown below:

---

[6] Fong, Holden, and Trzcinka (2010) compare the performance of nine low frequency liquidity measures with four standard high frequency benchmarks and find that Amihud's liquidity measure along with their own new liquidity measure performs the best among all the low frequency liquidity measures.

[7] This $R^2$ measure from regressions has been used by Roll (1988) and Morck, Yeung, and Yu (2000). They perform the regression of stock returns on market return to examine the extent of co-movement of stock prices within a country. Our $R^2$ measure for capturing liquidity commonality computation is similar to their methodology.

$$LiqCom_t = Ln[R_t^2/(1 - R_t^2)]$$

Here, $LiqCom_t$ is the monthly liquidity commonality for all the stocks in the sample. The measure is constructed in a similar fashion for different size-based portfolios.

## 4.2 Research Methodology

To examine market-wide commonality in liquidity for equity market, we follow CRS (2000) and run market model time series regressions given by

$$LIQ_{j,t} = \propto_J + \beta_{1,j} DLIQ_{M,t} + \beta_{2,j} DLIQ_{M,t+1} + \beta_{3,j} DLIQ_{M,t-1} + \delta_{1,j} Return_{M,t} + \delta_{2,j} Return_{M,t+1}$$
$$+ \delta_{3,j} Return_{M,t-1} + \delta_{4,j} Volatility_{j,t} + \varepsilon_{J,t} \qquad (EQ3)$$

Where, $j$ = 1, 2, 3………, 960, $t$ = 1, 2, 3………, 504.

$DLIQ_{j,t} = (LIQ_{j,t} - LIQ_{j,t-1})/LIQ_{j,t-1}$ denotes each of the six liquidity measures used in the study on a given day $t$ for a firm $j$. $DLIQ_{M,t}$ is the concurrent change in the corresponding average market liquidity measure. We also include a lag and lead market liquidity variables in $EQ1$ to capture any nonsynchronous change in liquidity due to thin trading. Cross-sectional means of time series slope coefficients are reported with the t-statistics to test the null hypothesis that there is no market-wide commonality in liquidity for stocks listed on NSE in line with Fama-Macbeth (1973). The concurrent, lag and lead market return along with idiosyncratic firm volatility act as control variables for the model.

The form of the time-series market model regression in the case of options market is given below.

$$DOPLIQ_{j,t} = \propto_J + \beta_{1,j} DLIQ_{j,t} + \beta_{2,J} DOPLIQ_{M,t} + \beta_{3,J} DOPLIQ_{M,t-1} + \beta_{4,J} DOPLIQ_{M,t+1}$$
$$+ \beta_{5,J} DLIQ^{res}{}_{M,t} + \beta_{6,J} DLIQ^{res}{}_{M,t-1} + \delta_{1,J} Return_{j,t} + \delta_{2,J} Volatility_{j,t} + \varepsilon_{j,t}$$

*(EQ4)*

where, $j$ = 1, 2, 3………, 143, $t$ = 1, 2, 3………., 504.

$DOPLIQ_{j,t} = (OPLIQ_{j,t} - OPLIQ_{j,t-1})/OPLIQ_{j,t-1}$, denotes each of the four option market liquidity measures used in the study on a given day $t$ for a firm $j$. Here, $DOPLIQ_{j,t}$ is the percentage change in the option's liquidity measure and $DLIQ_{j,t}$ is the percentage change in the liquidity measure of the stock corresponding to the option. $DOPLIQ_{M,t}$ is the option market's liquidity measure And, $DLIQ^{res}{}_{M,t}$ is the residual from the following regression equation.

$$DLIQ_{M,t} = \alpha_0 + \alpha_1\, DOPLIQ_{j,t} + \varepsilon_t \qquad (EQ5)$$

This is included in *EQ2* to make sure that the coefficients estimated are purely for the options market. The liquidity measure $DLIQ_{j,t}$ which belongs to stock $j$ is included to capture the positive association between liquidities of both the markets due to hedging demand. The underlying firm's return and volatility are included in *EQ2* as additional control variables.

To examine industry-wide liquidity on firm liquidity we use the following regression model:

$$DLIQ_{j,t} = \alpha_j + \beta_{1,j}\, DLIQ_{M,t} + \beta_{2,j} DLIQ_{M,t+1} + \beta_{3,j} DLIQ_{M,t-1} + \beta_{4,j}\, DLIQ_{Ind,t} + \beta_{5,j}\, DLIQ_{Ind,t+1}$$

$$+\, \beta_{6,j}\, DLIQ_{Ind,t-1} + \delta_{1,j}\, Return_{M,t} + \delta_{2,j}\, Return_{M,t+1} + \delta_{3,j} Return_{M,t-1}$$

$$+\, \delta_{4,J}\, Volatility_{j,t} + \varepsilon_{j,t} \qquad (EQ6)$$

where, $j$ = 1, 2, 3………, 960, $t$ = 1, 2, 3………., 504, *Ind* = 1, 2, 3………., 17

For testing the supply and demand-side sources of liquidity commonality we use daily data for a period of 12 years from 2001-2010 to construct quarterly measures of liquidity

commonality. As the time period is very long and we don't have the intra-day data for this long period we depend on Amihud's liquidity measure (*LIQ*) to capture liquidity commonality. We use the $R^2$ of regressions of the individual stock liquidity on market liquidity to compute the liquidity commonality measure.

To test for the impact of asymmetric information on liquidity commonality, we make use of the following regression model:

$$DNTrades_{j,t} = \propto_J + \beta_{1,j} DNTrades_{M,t} + \beta_{2,j} DNTrades_{M,t+1} + \beta_{3,J} DNTrades_{M,t-1}$$
$$+ \beta_{4,j} DNTrades_{I,t} + \beta_{5,j} DNTrades_{I,t+1} + \beta_{6,j} DNTrades_{I,t-1} + \delta_{1,j} Return_{M,t}$$
$$+ Return_{M,t-1} + Return_{M,t+1} + \varepsilon_{J,t} \qquad (EQ7)$$

Where $NumTrades_{J,t}$ measures the transaction frequency which is the overall trades for the firm on a given day. $NTrades_{M,t}$ ($NTrades_{I,t}$) measures the equally-weighted transaction frequency of all the firms in the sample for the market (industry) except firm *j*.

We examine the time-series behavior of supply-side sources of liquidity commonality by using the following model:

$$LiqCom_t = \alpha_t + \beta_1 SInt_t + \beta_2 CPSpread_t + \beta_3 BrokerReturns_t + \beta_4 BankReturns_t$$
$$+ \beta_5 Return_{M,t} + \beta_6 Liq_{M,t} + \beta_7 Volatility_{M,t} + \delta_1 Turnover_{M,t} + \varepsilon_t \qquad (EQ8)$$

*SInt* is the short-term interest rate (%) which is the 91-day treasury-bill rate. *CPSpread* is the commercial paper spread, *BrokerReturns* is the equally-weighted average returns of the brokerage industry. *BankReturns* is the equally-weighted average returns of the banking stocks listed on NSE. The above four variables serve as proxies of supply-side sources of liquidity commonality. Along with the supply-side sources, we include four other market conditions

factors; Market return ($Return_M$), market liquidity ($Liq_M$), market volatility ($Volatility_M$), and market turnover ($Turnover_M$) as additional regressors.

We examine the time-series behavior of demand-side determinants of liquidity commonality by using the following model:

$$LiqCom_t = \alpha_t + \beta_1 NetFII_t + \beta_2 NetMF_t + \beta_3 ExchangeRate_t + \beta_4 LnExports_t + \beta_5 Return_{M,t}$$
$$+ \beta_6 Liq_{M,t} + \beta_7 Volatility_{M,t} + \delta_1 Turnover_{M,t} + \varepsilon_t \qquad (EQ9)$$

*NetFII* is the net FII flow in a month in percentage terms which is calculated as (Net buy/ (buy+sell/2)). *NetMF* is net mutual fund flow in a month calculated to *NetFII*. *ExchangeRate* is the monthly percentage change in exchange rate of Indian rupee vis-à-vis dollar. *LnExports* is the natural logarithm of exports each month. The above four variables serve as proxies of demand-side sources of liquidity commonality. Along with the demand-side sources, we include four other market conditions factors; Market return ($Return_M$), market liquidity ($Liq_M$), market volatility ($Volatility_M$), and market turnover ($Turnover_M$) as additional regressors.