Market quality in the time of algorithmic trading

Nidhi Aggarwal* Susan Thomas

December 19, 2013

Abstract

We contribute to the emerging literature on the impact of algorithmic trading with an analysis of India's National Stock Exchange. This analysis has three strengths: A clean setting with one dominant exchange, a natural experiment where the introduction of co-location was followed by a sharp surge in algorithmic trading, and precise identification of algorithmic orders and trades. The results largely suggest that increased algorithmic intensity has given improvements of market quality.

*The authors are with the Finance Research Group, IGIDR. We are grateful to the National Stock Exchange of India, Ltd. for both research support. We thank Ajay Shah for inputs into the research design and Chirag Anand for technological inputs.

Contents

1 Introduction **3**

2 The impact of algorithmic trading (AT) on markets **4**

3 Research setting **6**
3.1 A clean microstructure 6
3.2 The introduction of the co-location facility (co-lo) 7
3.3 A unique dataset 8

4 Measurement **8**
4.1 Measuring AT intensity 8
4.2 Measuring market quality 9
4.2.1 Liquidity 9
4.2.2 Volatility 11
4.2.3 Market efficiency 11

5 Methodology **12**
5.1 Co-location and AT intensity 12
5.2 Choice of samples 12
5.3 Threats to validity: did other factors cause changes in market quality? 15

6 Data **19**

7 Results **21**
7.1 How did AT impact market quality? 21
7.2 Threats to validity: Did other factors cause the change in market quality? 25

8 Conclusion **29**

A Estimations addressing threats to validity **32**

1 Introduction

In recent years, there has been a surge of interest in the impact of algorithmic trading (AT) upon market quality. In this paper, we examine the impact of AT upon market quality at the National Stock Exchange of India (NSE), which is one of the biggest exchanges of the world, with India as a large emerging market. In addition, this paper improves upon existing measurement of the role of AT through clean identification of the impact of AT upon market quality, owing to a research design with three unique features.

The first concerns fragmentation of the order flow. In countries such as the US, trading takes place at numerous venues, each of which has a different market design. This makes it hard to understand the causal impact of one design feature such as algorithmic trading at any one trading venue. In contrast, in our study of NSE, we have a simple setting where NSE accounts for 80% of equity spot trading and 100% of equity derivatives trading.

The second issue is about disentangling cause and effect where traders voluntarily shift to algorithmic trading. At NSE, there was a sharp date in January 2010 on which the co-location facility was commissioned, after which there was an S-curve adoption of AT. This gives us a dataset with one group of days (prior to the co-location facility) with AT intensity of 15%, and another group of days (after the co-location facility) with AT intensity of 55%.

The third issue is about measurement. Many researchers have been forced to work with crude proxies of algorithmic trading. The data at NSE precisely flags every order, and counterparties on every trade, as being AT or not.

The analysis shows that, on average, market quality has improved in the following ways: lower transactions costs, larger number of shares available for trade, and a reduced imbalance between the number of shares available to buy and sell. It also shows a sharp drop in the volatility of prices and the volatility of transactions costs after the increase in algorithmic trading. However, the depth as measured by the monetary value available to trade *worsens* with higher algorithmic trading, both at the touch (best bid and offer) as well as the value available for trading upto the best 5 prices decreases. These results are similar to the findings of Hendershott *et al.* (2011), who find that increased algorithmic trading activity in the US caused a drop in quoted as well as effective spreads, but also lowered the quoted depth. Other than the depth measures, these results do suggest that, on average, market quality improved with algorithmic trading.

A key threat to validity of the analysis lies in changing macro-economic conditions. The pre-co-location period happens to lie in 2009, which was a time of enhanced macroeconomic uncertainty. This raises the possibility that some or all of the apparent improvements in market quality were merely driven by restoration of normalcy in finance and macroeconomics.

We address these concerns through two strategies. First, financial and macroeconomic uncertainty is controlled for in linear regressions. Second, a matched dataset is constructed containing *days* which have similar aggregate volatility. This is equivalent to controlling for aggregate volatility, without assuming linearity of the relationship between volatility and market quality. This analysis shows that results that are consistent with the earlier analysis in all but the order imbalance which is the difference in the number of shares available to buy and sell. Here, the estimations (after adjusting for macroeconomic volatility) show that there was no significant impact of algorithmic trading on the order imbalance. Thus, we conclude that algorithmic trading improves market quality, other than the depth measures.

The remainder of the paper is organised as follows: Section 3 provides a brief detail on the institutional framework. Section 5 discusses the identification of algorithmic trading activity, measurement of market quality measures and the approach used for analysis in detail. Section 6 describes the data and gives some summary statistics. Section 7 presents the estimation results. Section 8 concludes.

2 The impact of algorithmic trading (AT) on markets

Algorithmic Trading (AT) has been defined as the use of computer algorithms to automatically make trading decisions, submit orders and manage those orders after submission (Hendershott *et al.*, 2011). Many researchers have tried to understand how this shift in the use of technology has changed market quality in terms of liquidity, price efficiency and volatility.

Theoretical models focus on a subset of AT called high frequency trading (HFT), which has a greater focus on low latency of order placement. These models analyse how a rapidly changing trading environment affects liquidity costs and investor welfare. Jovanovic and Menkveld (2010) suggest that to the extent that information is known only to the HFTs, they can reduce welfare. However, in markets where information is continually updated, HFT can improve welfare by posting competitive quotes and reducing informational friction. Biais *et al.* (2013) show that the rapidity of order placement can have a positive impact by providing mutual gains from trade, but also cause negative externalities by increasing adverse selection cost. Cartea and Penalva (2012) build on Grossman and Miller (1988) and show that presence of HFT makes liquidity traders worse off by increasing the price impact of their trade as well as the volatility of the prices. On the other hand, Hoffmann (2012) modifies Foucault (1999) to show how HFT can have a positive impact on market liquidity, conditional on the initial level of market efficiency. Martinez and Rosu (2013) show that HFT does not destabilise the markets, but improves market efficiency by incorporating information into prices quickly.

The dominant strand in the AT literature is empirical measurement of the impact of AT. A major drawback in this literature is poor measurement. Most existing datasets do not have precise flagging of orders or trades that use AT. Zhang (2010) uses a proxy to observe HFT and finds that high frequency trading is negatively related to price formation and also increase volatility. Hasbrouck and Saar (2013) develop a new proxy based on strategic runs in the market, and finds that with increased high frequency trading comes narrower spreads, higher displayed depth and lowered short term volatility. Hendershott *et al.* (2011) treats electronic messages as a proxy for AT, and uses the onset of automated quote dissemination on the New York Stock Exchange as an exogenous event. They find that AT lowers liquidity costs, and improves quote informativeness, particularly for large market capitalization securities.

Another approach to identification is to locate market wide events that are expected to change the level of AT intensity, or exogenous factors indicating the degree of AT in the market. For example, Riordan and Storkenmaier (2012) use the drop in latency at the Deutsche Bourse and find that it is correlated with improved market quality measured by decreased spreads and higher price efficiency.[1] Bohemer *et al.* (2012) use the introduction of co-location facilities at 39 exchanges to locate increases in the level of AT in markets and find that higher AT is correlated with better market liquidity, efficiency, but also higher market volatility. In contrast, Chaboud *et al.* (2009) find no similar relationship between AT and higher volatility on foreign exchange markets.[2]

[1]A few studies such as Viljoen *et al.* (2011), Frino *et al.* (2013) also examine the impact of algorithmic trading on the futures market. They find a similar positive impact of algorithmic trading on liquidity and price efficiency.

[2]Other studies also look at the liquidity provisioning function of algorithmic traders. Hendershott and Riordan (2013) find that algorithmic traders demand liquidity when it is

In the best scenario, researchers access proprietary datasets to examine the role of high frequency traders (HFTs, which is a subset of AT) on price discovery and efficiency. For example, Menkveld (2013), Carrion (2013), Brogaard (2010), Brogaard *et al.* (2012), find that HFTs play a beneficial role in enabling price efficiency and provide liquidity, particularly around times of market stress.

There is a certain contrast between the broadly benign messages of the research literature, and the mistrust that many policy makers and practitioners express about the role of AT. The limitations of the datasets used in the existing literature have generated skepticism about the existing literature. In this paper, we analyse a large exchange with perfect identification of algorithmic orders and trades.

3 Research setting

The research setting in this paper has three strengths compared with the rest of the literature. There is a clean market microstructure setting where most spot trading and all derivatives trading takes place at only one exchange; there is a natural experiment where AT surged after co-location facilities were introduced on this exchange; the underlying data infrastructure precisely flags every order and the counterparties of every trade with a dummy variable signifying AT.

3.1 A clean microstructure

We analyse the impact of AT on market quality of one of three exchanges trading equity in India,[3] the National Stock Exchange (NSE). NSE has the largest share of the domestic market activity, with 80% of the traded volumes on the equity spot market and 100% of the traded volume on equity derivatives (SEBI, 2013). It is also one of the highest ranked equity markets in the world by transaction intensity.[4]

cheap and supply liquidity when it is expensive. Carrion (2013) study the high frequency trading strategies and find that HFTs provide liquidity when it is scarce and consume when it is plentiful.

[3] The other two are the Bombay Stock Exchange and Multi-commodity Stock Exchange.

[4] Source:`http://www.world-exchanges.org/files/statistics/2012%20WFE%20Market%20Highlights.pdf`

The NSE is an electronic limit order book market, where orders are executed on a price-time priority basis. Information about trades, quotes, and quantities are disseminated by the exchange on a real time basis, with traders being able to view the best five bid-ask prices at every given point in time. There are around 1500 securities that trade on the NSE. Current trading hours are from 9:00 am to 3:30 pm (IST), but were from 9:55 am to 3:30 pm before Jan 1, 2010. The market opens with a call auction which runs for 15 minutes, after which trading is done using a continuous order matching system.

All spot trades are cleared with netting by novation at a clearing corporation and settled on a $T + 2$ basis. Netting within the day accounts for roughly 70% of the turnover. Of the trades that are settled, typically around 10-15% have a domestic and foreign institutional investor. Most of the trading can be attributed to retail investors or proprietary trading by securities firms.

3.2 The introduction of the co-location facility (co-lo)

In the international experience, computer technology came into order placement owing to the desire to achieve best execution for customers.[5] This was not a consideration in India, where NSE was and has been the dominant exchange, and there have been no regulations about best execution.

Automated order placement began with a few securities firms establishing technology for equity spot arbitrage between NSE and BSE. The securities regulator issued regulations governing algorithmic trading in April 2008[6] but even after this, the level of AT remained low.[7]

The decisive event was co-location at NSE,w hich began in January 2010.[8] Latency dropped from 10–30 ms to 2–6 ms, and traders who established automated systems in the co-location facility had a significant edge. This gave a surge in AT, as is documented later in this paper.

[5] Some examples of this include *Marketplace rules, 2001* by the Canadian Securities Administration, *Regulation National Market Services* or Regulation NMS, 2005 by the U.S. SEC and *Markets in Financial Instruments Directive* or MiFID, 2007 in Europe.

[6] `http://www.sebi.gov.in/circulars/2008/cir072008.pdf`

[7] `http://www.livemint.com/Opinion/QpU7GHjhTLwClANyUX5T5N/Indian-markets-slowly-warming-up-to-algorithmic-trading.html`

[8] `http://www.nseindia.com/technology/content/tech_intro.htm`

3.3 A unique dataset

The previous literature examining the impact of AT or HFT has generally been limited to observing proxies for AT. For example, Hendershott *et al.* (2011) and Bohemer *et al.* (2012) use electronic message traffic to capture the level of algorithmic trading activity in the market, Hasbrouck and Saar (2013) propose a measure called *RunsInProcess* to identify HFT activity.

The closest that the literature has to a direct measure is where the exchange identifies trading firms as 'engaging primarily in high frequency trading'. Therefore, Brogaard (2010), Brogaard *et al.* (2012) and Carrion (2013) use data from NASDAQ that identifies a subset of HFT on a randomly selected 120 US securities based on the activity of 26 trading firms tagged by NASDAQ as being engaged in HFT. Despite being a very informative dataset, there are concerns about the coverage of the HFT firms in the sample (Brogaard *et al.*, 2012; Bohemer *et al.*, 2012). Another study by Hendershott and Riordan (2013) uses DAX dataset on AT orders for 30 securities on 13 trading days. While this does not suffer from the problem of coverage of AT firms, it suffers from the issue of very small sample that covers only a few securities.

Our analysis uses a dataset of all orders and all trades, timestamped to the millisecond, with an AT flag for every order and for every counterparty of every trade. This data is available from 2009 onwards, which covers the period before the co-location facility also. This is thus a unique dataset within the AT literature.

4 Measurement

As with the rest of the literature, there are two main issues that shape the research design. The first group of questions is about measuring AT intensity and market quality. The second group of questions is about establishing a causal impact of AT intensity upon market quality.

4.1 Measuring AT intensity

We classify a trade as an *AT trade* if either buyer or seller was AT. High-frequency data is used to construct a discrete measure, with a resolution of five minutes, of AT intensity. The AT intensity, $\textsc{at-intensity}_{i,t}$, for any security '$i$' in any given time interval 't', is the share of AT trades within total traded value within the five-minute interval:

$$\text{AT-INTENSITY}_{i,t} = \frac{\text{TTV}_{ATtrade,i,t} \times 100}{\text{TTV}_{i,t}}$$

where $\text{TTV}_{ATtrade,i,t}$ refers to the total traded value of AT trades in time period 't'. $\text{TTV}_{i,t}$ refers to total traded value of all trades in time period 't'.

4.2 Measuring market quality

Liquidity, volatility and market efficiency are measures of market quality, with precise definitions as follows.

4.2.1 Liquidity

Liquiditiy is multi-dimensional in nature, and so we break up measurement into two parts: transactions costs and depth. Transactions costs are higher for markets that are less liquid. Depth is lower for markets that are less liquid. We use the following measures to capture market liquidity:

1. Transactions costs measures:
   a) Quoted Spread (QSPREAD): It is measured as the difference between the best ask and the best bid price at any given point of time. In order to make it comparable across securities, we express it as a percentage of the mid-quote price, which is the average of the best bid and the ask prices. Thus, for a security 'i' at time 't', $\text{QSPREAD}_{i,t}$ is defined as

   $$\text{QSPREAD}_{i,t} = \frac{(\text{P}_{\text{BestAsk}_{i,t}} - \text{P}_{\text{BestBid}_{i,t}}) \times 100}{\text{P}_{\text{MQ}_{i,t}}}$$

   where $\text{P}_{\text{BestAsk}_{i,t}}$ and $\text{P}_{\text{BestBid}_{i,t}}$ are the best ask and bid prices respectively. $\text{P}_{MQ_{i,t}}$ indicates the mid-quote prices.

   b) Impact Cost (IC): The QSPREAD only captures the liquidity available at the best prices. However, the transaction size supported at the best price may not be the typical size at which traders typically execute their trades. Instead, we use the Impact Cost, which measures liquidity for a fixed transaction size Q. For a security 'i', at time 't', $\text{IC}_{Q,i,t}$ is calculated as:

$$\textsc{ic}_{Q_{i,t}} = 100 \times \frac{\text{P}_{Q_{i,t}} - \text{P}_{\text{MQ}_{i,t}}}{P_{\text{MQ}_{i,t}}}$$

where $\text{P}_{\text{Q}_{i,t}}$ is the execution price calculated for a market order of Q and $\text{P}_{\text{MQ}_{i,t}}$ is the mid-quote price. Q is held fixed at a transaction size of USD 500 (Rs 25,000).[9]

Lower values of QSPREAD and IC indicate higher liquidity.

2. Depth (DEPTH) measures:

   a) TOP1DEPTH captures the rupee depth at the best bid and ask prices as:

   $$\textsc{top1depth}_{i,t} = \text{P}_{\text{BestBid},i,t} \times \text{Q}_{\text{BestBid},i,t} + \text{P}_{\text{BestAsk},i,t} \times \text{Q}_{\text{BestAsk},i,t}$$

   where $\text{P}_{\text{BestAsk}_{i,t}}$ and $\text{P}_{\text{BestBid}_{i,t}}$ are the best ask and bid prices respectively of security 'i' at time 't'.

   b) TOP5DEPTH captures the cumulated rupee depth across the best five bid and ask prices as:

   $$\textsc{top5depth}_{i,t} = \Sigma_{k=1}^{5} \text{P}_{\text{BestBid},k,i,t} \times \text{Q}_{\text{BestBid},k,i,t} + \Sigma_{k=1}^{5} \text{P}_{\text{BestAsk},k,i,t} \times \text{Q}_{\text{BestAsk},k,i,t}$$

   where $\text{P}_{\text{BestAsk}_{i,t}}$ and $\text{P}_{\text{BestBid}_{i,t}}$ are the best ask and bid prices of security 'i' at time 't'.

   c) DEPTH measure the total number of shares outstanding at either side of the book available for execution at any point of time. It is expressed as an average, with units of number of shares, and is computed as:

   $$\textsc{depth}_{i,t} = \frac{\text{TSQ}_{i,t} + \text{TBQ}_{i,t}}{2}$$

   where $\text{TSQ}_{i,t}$ denotes the total sell quantity and $\text{TBQ}_{i,t}$ denotes the total buy quantity of security 'i' at time 't'.

   d) Order Imbalance (OIB) is measured as the difference between the buy and sell side depth, and expressed as a percentage of the total depth on average as:

   $$\text{OIB}_{i,t} = \frac{(\text{TSQ}_{i,t} - \text{TBQ}_{i,t}) \times 200}{\text{TBQ}_{i,t} + \text{TSQ}_{i,t}}$$

   A market with lower absolute value of order imbalance is viewed as a high quality market.

[9] This is the average transaction size on the spot market at NSE.

4.2.2 Volatility

There are two elements of volatility observed from market prices: price and liquidity risk.

1. Price risk (RVOL): This is the variance of intra-day returns, where returns are calculated using traded prices at one second frequency as:

$$\text{RVOL}_{i,t} = \sqrt{\frac{\Sigma_{T=1}^{300}(r_{i,T} - \overline{r}_{i,t})^2}{n-1}}$$

   where 't' indexes the five minutes time interval, while 'T' indexes the one second time points. $\overline{r}_{i,T}$ indicates the mean returns within the five minute interval.

2. Liquidity risk (LRISK): We measure liquidity risk as the *variance* of impact cost over a given time interval. It captures the variation in the transactions costs faced when executing a market order of size Q at different times in the trading day.

   LRISK is computed as the standard deviation of IC (computed at one-second frequency) in a five minute interval:

$$\text{LRISK}_{i,t} = \sqrt{\frac{\Sigma_{T=1}^{300}(\text{IC}_{i,T} - \overline{\text{IC}}_{i,t})^2}{n-1}}$$

   where 't' indexes the five minutes time interval, while 'T' indexes the one second time points. $\overline{\text{IC}}_{i,t}$ indicates the mean of IC within the five minute interval.

4.2.3 Market efficiency

While there are several measures to calculate the efficiency of market prices, we use the Variance Ratio (VR) of returns. The variance ratio (Lo and MacKinlay, 1988) is defined as the ratio of $1/k$ times the variance of k-period return to that of one period return, and is calculated as:

$$\text{VR}(k) = \frac{Var[r_t(k)]}{k \cdot Var[r_t]}$$

where r_t is the one period continuously compounded return, $r_t(k) = r_t + r_{t-1} \ldots + r_{t-k}$. k indicates the lag at which the variance ratio (VR)is to be computed.

We compute the VR as the ratio of the variance of returns calculated over ten minutes to the variance of five minute returns. In an efficient market, where prices are expected to approximate a random walk, these values of VR should be not be significantly different from 1.

5 Methodology

We use two approaches to establish a causal relationship between the AT intensity and market quality: (a) An comparative analysis of the average levels of market quality between a period when AT was low to when AT was high. (b) A cross-sectional analysis of the effect of AT on market quality at the level of specific securities between the period of low and high AT.

5.1 Co-location and AT intensity

The introduction of co-location services at the NSE as described in Section 3.2 can be a useful exogenous event which gives us an opportunity to measure the impact of AT intensity upon market quality. Prior to the introduction at co-location services, the level of AT in the market was low.

While we expect a significant increase in the presence of AT trades and orders in the market, we expect that the change would come about through the typical S-curve of adoption of innovations.

Figure 1 shows that before the introduction of the co-lo in January 2010, the AT intensity on the NSE was low. After the introduction of the co-lo (marked by the dashed vertical line in the graph), AT intensity picked up gradually. The period between January 2010 and July 2011 was an adjustment period where participants were adopting the new technology.

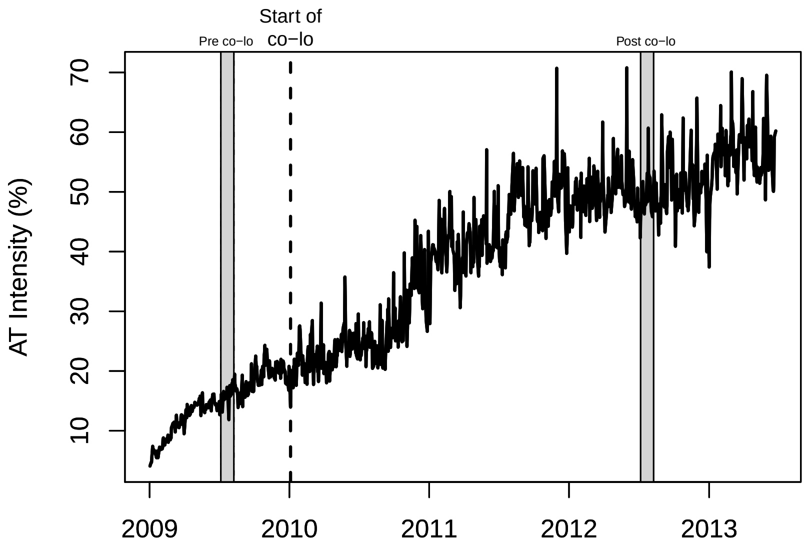
This S-curve of adoption implies that it is not useful to do a sharp study of a few days before and after the introduction of the co-location facility.

5.2 Choice of samples

We use Figure 1 to select two groups of days, one where the data shows a low level of AT-INTENSITY and another where AT-INTENSITY is significantly higher. Before January 2010, the average AT intensity was at around 20%. After January 2010, AT-INTENSITY steadily increased through but stabilized

Figure 1 AT intensity between 2009 and 2013

The graph shows AT intensity for the overall equity spot market at NSE between 2009 and 2013. AT intensity is measured as a fraction of total traded value of AT trades in a day vis-a-vis the total traded value on that day. The dotted line shows the date on which co-lo was introduced by NSE. The shaded region indicates the two periods of study.



at 50% after July 2011. Hence, we choose the following groups of days for analysis:

- The *low-AT* sample: 9 July to 7 August 2009
- The *high-AT* sample: 9 July to 8 August 2012

We examine the average AT-INTENSITY of the overall market[10] in each of the the selected samples in greater detail in Table 1. We also examine the AT-INTENSITY of the top 100 securities by market capitalisation.

We see that the average AT-INTENSITY for the overall market was about 4.33% in the LOW-AT sample. This is significantly higher in the HIGH-AT sample at 16.39%. In the universe of 100 large sized securities, we see that the average AT intensity in the LOW-AT sample was significantly higher at 14.28%, higher by nearly 4× of the overall market. This further rose to 53.97% in the HIGH-AT sample, an increase of 3×.

[10]The 'overall market' consists all the securities traded on the NSE equity spot markets during this period. On average, NSE trades around 1500 securities daily.

Table 1 Summary statistics of AT intensity in the LOW-AT and HIGH-AT periods

The table presents summary statistics of AT-INTENSITY for the overall market and the top 100 securities by market capitalization and liquidity.
AT-INTENSITY is calculated as the percentage of the total traded value of AT trades vis-a-vis the total traded value for a security within a day. It is calculated for each security in the LOW-AT (July 6 to August 8, 2009) and HIGH-AT (July 6 to August 9, 2012) samples separately, and then averaged across all days.

All values in %

	Overall Market		**Top100**	
	LOW-AT	HIGH-AT	LOW-AT	HIGH-AT
Min	0.00	0.00	2.36	22.91
Q1	1.34	5.44	10.15	46.37
Mean	4.10	16.39	14.28	53.97
Median	2.33	12.07	14.34	55.14
Q3	4.00	21.66	19.16	61.83
Max	27.13	77.28	27.13	77.28
SD	4.81	15.18	5.98	11.60

In addition to the comparative analysis of the average market quality, we use a fixed effects regression to adjust for the cross-sectional variation in market quality in relation to the cross-sectional variation in AT intensity. This helps to reduce the endogeneity bias induced as a result of omitted variables. The model used is as follows:

$$\text{M1}: \text{MKT-QUALITY}_{i,t} = \alpha_i + \beta_1 \text{AT-INTENSITY}_{i,t-1} + \beta_2 \text{CO-LO-DUMMY}_t + \epsilon_{i,t}$$

where $i = 1,\ldots,$N indexes firms, $t = 1,\ldots,$T, indexes 5-minute time intervals. α_i captures the firm specific unobserved factors, MKT-QUALITY$_{i,t}$ represents one of the market quality measures (QSPREAD, IC, TOP1DEPTH, TOP5DEPTH, OIB, DEPTH, LRISK, RVOL) for security *'i'* at time *'t'*. CO-LO-DUMMY$_{i,t}$ is a dummy used to capture the differences due to differences in the LOW-AT and HIGH-AT samples. It takes value '1' for the HIGH-AT sample and zero otherwise. That is,

$$\text{CO-LO-DUMMY}_t = \begin{cases} 1 & \text{if 't'} \in \text{Post 2010 period} \\ 0 & \text{otherwise} \end{cases}$$

AT-INTENSITY$_{i,t-1}$ represents the AT intensity in security *'i'* in the previous five minutes. We use the previous five minutes of AT-INTENSITY to address the endogeneity issues that can arise because of the feedback relation between the market quality variables and AT-INTENSITY. The coefficient of interest is β_2 which captures the effect of AT on each market quality variable.

If higher AT-INTENSITY results in better market quality, we expect β_2 to be negative for the market quality variables QSPREAD, IC, LRISK, |OIB|, RVOL and positive for DEPTH, TOP1DEPTH, TOP5DEPTH. Thus, we test the hypothesis (H_0^1):

$$H_0^1 : \beta_2 = 0$$
$$H_A^1 : \beta_2 < 0$$

where MKT-QUALITY $\in$ (QSPREAD, IC, LRISK, |OIB|, RVOL).

We expect that better market quality is associated with higher depth, and set the alternative hypothesis to be:

$$H_A^1 : \beta_2 > 0$$

where MKT-QUALITY $\in$ DEPTH, TOP1DEPTH, TOP5DEPTH.

Since the dataset is a panel with of large number of time-dimensional observations, we report the Driscoll-Kraay standard errors which are robust to heteroskedasticity, autocorrelation and cross-sectional dependence (Driscoll and Kraay, 1998; Hasbrouck and Saar, 2013).

5.3 Threats to validity: did other factors cause changes in market quality?

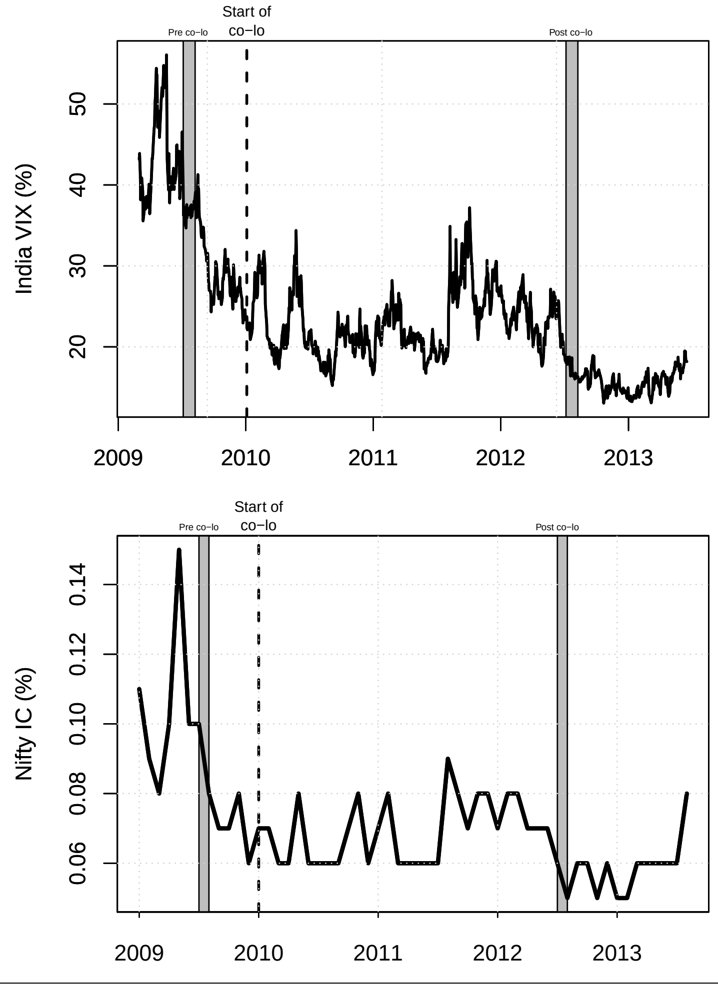
In Section 5.2, we selected samples before and after the introduction of co-lo such that AT intensity was significantly higher level in the HIGH-AT sample. However, as the two periods are three years apart, there is the possibility of many other things having changed. If market volatility is significantly different between the two samples, then the significant changes in market quality might be a consequence of market volatility rather than the change in AT. For example, the LOW-AT sample is observed from the period immediately after the 2008 global financial crisis where market volatility would tend to be systematically higher compared to that in the HIGH-AT period, which is well after the crisis.

A similar argument holds for liquidity measures. The literature on commonality of liquidity across securities shows a significant influence of market

Figure 2 Market volatility and liquidity between 2009 and 2013

The first graph below shows the daily time series of the implied volatility index, India VIX between 2009 and 2013, while the second graph shows the monthly time series of the impact cost of buying and selling Rs.5 million (under USD 80,000) worth of the NSE-50 index.

The dashed line indicates the date on which NSE started co-lo services. The shaded regions indicates the periods of the LOW-AT and the HIGH-AT samples selected for the analysis.



liquidity on the liquidity of all securities (Chordia *et al.*, 2000), and in turn, market liquidity is strongly related to market volatility (Hameed *et al.*, 2010).

We examine the time series of the volatility and liquidity of the market index between January 2009 to August 2013. Market volatility is measured by the daily time series of the Indian implied volatility index, India VIX,[11] while the market liquidity is measured by the monthly time series of the Nifty Impact Cost[12] in the same period. The graphs shows that both market volatility was much higher in 2009 compared to the 2013. The market impact cost was also much higher showing that market liquidity was significantly lower during the selected LOW-AT sample than the HIGH-AT sample.

We address this problem through two strategies.

1. **Including conditioning variables in the models of cross-sectional variation**: We add a control variable that captures market volatility in the specification given by M1. The market volatility is measured using the realised volatilty of the Nifty index.[13] This is used to modify M1 to give M2 as follows:

$$\begin{aligned}\text{MKT-QUALITY}_{i,t} &= \alpha_i + \beta_1 \text{CO-LO-DUMMY}_t + \beta_2 \text{AT-INTENSITY}_{i,t-1} \\ &\quad + \beta_3 \text{NIFTY-VOL}_t + \epsilon_{i,t}\end{aligned}$$

   where $\text{NIFTY-VOL}_{i,t}$ is the variance of five-minute returns on the market index.

   In the aftermath of the global crisis, there were particularly sharp peaks of volatility within the day. In order to control for these, we introduce an intra-day dummy into M3:

$$\begin{aligned}\text{MKT-QUALITY}_{i,t} &= \alpha_i + \beta_1 \text{CO-LO-DUMMY}_t + \beta_2 \text{AT-INTENSITY}_{i,t-1} \\ &\quad + \beta_3 \text{NIFTY-VOL}_t + \beta_4 \text{INTRADAY-DUMMY}_t + \epsilon_{i,t}\end{aligned}$$

   where the intraday-dummy takes value 1 if 't' is the first or the last half hour of the trade, and zero otherwise. The selection of the first half an hour

[11] India VIX is a volatility index based on the Nifty Index Option prices. Nifty is NSE's market index based on 50 securities which constitute about 70% of the free float market captialization of the securities listed on NSE. India VIX uses the Chicago Board Options Exchange (CBOE) computation methodology, with few amendments to suit the Indian markets. See: `http://www.nseindia.com/content/indices/white_paper_IndiaVIX.pdf`

[12] Nifty Impact Cost represents the cost incurred on buying or selling a portfolio of Nifty stocks for a transaction size of Rs. 50 lacs (around USD 83,333). The numbers are disseminated by the NSE on a monthly basis.

[13] Nifty is the market index comprising of the 50 largest firms (that are traded on NSE) in terms of market capitalization and transactions costs that are traded on NSE.

Table 2 Match balance on market volatility for the matched sample

The table presents the match balance statistics between the LOW-AT and HIGH-AT derived after matching. The p-value of the bootstrapped Kolmogorov-Smirnov test results is also reported.

	σ_{market}
Matched LOW-AT	13.05
Matched HIGH-AT	12.66
Kolmogorov-Smirnov test p-value	1

and last half an hour is motivated from Thomas (2010) which finds that the market volatility around these time periods is at its peak.

After the global crisis, many stock prices dropped sharply. To control for this, we use the level of the last traded price (LTP) as a control variable to arrive at M4:

$$\begin{aligned}\text{MKT-QUALITY}_{i,t} &= \alpha_i + \beta_1\text{CO-LO-DUMMY}_t + \beta_2\text{AT-INTENSITY}_{i,t-1} \\ &\quad +\beta_3\text{NIFTY-VOL}_t + \beta_4\text{INTRADAY-DUMMY}_t + \beta_5\text{LTP}_{i,t} + \epsilon_{i,t}\end{aligned}$$

2. **Matched sample**: We use a second approach to account for high market volatility by *matching dates in the sample before the introduction of co-lo that have the same levels of market volatility as in the sample after co-lo*. While matching methods are generally applied at the level of units of observations such as households or firms or countries, they can also be applied to choose time periods that are similar. As an example, the recent paper Moura *et al.* (2013) uses a similar strategy to construct a control and treatment set of *days*.

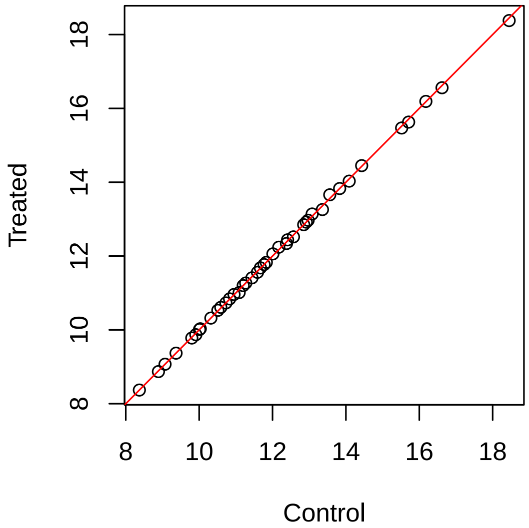
   We find 41 dates in the pre co-lo and post co-lo periods with similar levels of Nifty volatility. This becomes the matched LOW-AT and HIGH-AT samples. Table 2 shows the balance statistics between the matched LOW-AT and HIGH-AT samples on market index (Nifty) volatility.

   The table shows that the mean of the Nifty volatility for the matched HIGH-AT sample is similar to that of matched LOW-AT sample. The Kolmogorov-Smirnov test[14] results confirm this observation. Figure 3 shows the QQ-plot of Nifty volatility of the matched LOW-AT and the HIGH-AT samples. We observe that almost all the points fall on the 45 degree line, indicating a good match between the two periods.

[14] The advantage of the Kolmogorov-Smirnov test as opposed to the standard t-test is that it tests for the significant differences across the entire distribution rather than just the averages

Figure 3 QQ plot of Nifty volatility for the matched samples

The graph shows the QQ-plot of Nifty volatility for the matched LOW-AT and the HIGH-AT samples. Deviations from the 45 degree line indicate a poor match.



We then estimate Models 1-4 using the data from the matched dates in order to arrive at an estimate of the effect of AT on market quality measures, that is robust to the threat of validity posed by systemic changes in the macroeconomy during the sample period.

6 Data

From the analysis in Section 5.2, the LOW-AT sample is selected from July 6 to August 8, 2009, while the HIGH-AT sample is between July 6 to August 9, 2012. This gives us a sample of 23 contiguous trading days in the LOW-AT sample and 25 trading days in the HIGH-AT sample for use in both the comparative analysis of average market quality and market quality across the securities.

In order to address the concerns about the validity of the analysis in Section 5.3, we also identify matched LOW-AT and HIGH-AT samples. The matching is done by selecting dates from the period before the introduction of co-lo (January 1 to December 1, 2009) with the same market volatity as dates in the sample after the introduction of co-lo. We identify 41 days in this matched sample, and we refer to it as *matched sample* in the rest of the paper.

We restrict our analysis to the top 100[15] securities in terms of market capitalisation and liquidity. During the study period, this set of securities accounted for about 65% of the total traded volumes on the NSE. Since liquidity and market capitalization vary over time, the list of the 100 securities varies between the LOW-AT and the HIGH-AT samples. We restrict the analysis for the top 100 securities of 2012. Table 3 gives the descriptive statistics of the sample.

Table 3 Descriptive statistics of the sample

The table presents descriptive statistics for the sample of top 100 stocks used in the analysis. Panel A reports the statistics for the LOW-AT sample, while Panel B shows the statistics for the HIGH-AT sample.

	Market Cap (Rs. Million)	Price (Rs.)	Turnover (Rs. Million)
LOW-AT period			
Mean	325782.06	672.78	1091.25
Median	153426.01	675.15	980.54
SD	470503.74	48.02	460.37
As a % of total	72.44		61.75
HIGH-AT period			
Mean	422742.98	719.49	644.03
Median	232615.13	715.16	543.78
SD	533507.57	24.49	383.39
As a % of total	70.54		68.24

The dataset used in the analysis includes the measures of AT intensity as well as the nine variables of market quality as described in Section 4.2. All these variables are calculated at the frequency of five minutes for all the days in the LOW-AT and HIGH-AT samples, simple and matched. This gives us a total number of 345,282 observations for the one month sample, and 572,094 in the matched sample.

We delete the first fifteen minutes observations belonging to each day in the HIGH-AT sample which comes from the pre-opening call auction session. We also delete the observations belonging to the first ten minutes of trade in the continuous markets in order to reduce the noise caused by high frequency data on the analysis. We are left with 315,115 observations for the one month sample, and 509,376 observations for the matched sample.

[15] Out of these 100 firms, one firm, Coal India Ltd. was not listed in the period before co-location services started on January 2010.

7 Results

7.1 How did AT impact market quality?

We analyse the behavior of market quality variables discussed in Section 4.2 across the two samples – LOW-AT which comes from the one month period from July 6 to Aug 7, 2009 and HIGH-AT which comes from the period from Jul 6 to Aug 9, 2012. Figure 4 shows the average behavior of each market quality variable across the top 100 securities at one-minute frequency in the two samples. Table 4 presents three summary statistics for the the AT-INTENSITY measure as well as for the nine market quality measures. These are the *mean* of the sample, the standard deviation (marked as *SD*) and the *median*. These values are reported separately for the LOW-AT and the HIGH-AT samples.

Figure 4 shows that there is a significant improvement in both measures of transactions costs – QSPREAD and IC – in the HIGH-AT sample. QSPREAD dropped from 5 basis points (bps) to 3bps between these two samples. IC, which measures the cost of a larger transaction at Rs.25,000, dropped from 7 bps to 4bps. As shown in Table 4, both are significant decreases at a 5% level of significance.

Across the two *Rupee depth* measures, there was a decline in the HIGH-AT sample. This was true for both the liquidity available at the touch as well as the best 5 market by price (MBP) limit orders available for the security. The depth at the touch declined by more than 56% while the cumulative depth across the best 5 MBP declined by 28%.

Both the depth measures by number of shares showed significant improvement. The total depth increased, on average, by 25% in the HIGH-AT sample. The average OIB – which is gap between the total depth on the buy side and the sell side – decreased from 17.47% to below 9%, which is a decrease of nearly 50%.

Consistent with several studies of the intra-day impact of AT on market volatility, Figure 4 shows that there was a sharp drop in both the average level of intra-day securities volatility RVOL in the HIGH-AT sample compared to the LOW-AT sample. Table 4 shows that the average RVOL dropped significantly to about 52% in the sample HIGH-AT. Similarly, liquidity risk or LRISK has dropped by more than 50%.

Lastly, we find that the average variance ratio (VR) of 10 minutes compared

Figure 4 Market quality in the LOW-AT vs. the HIGH-AT samples

The graphs plot the average intra-day behavior for eight of the market quality measures between the LOW-AT and the HIGH-AT samples. In each case, the measure is first computed for each security at a one-minute frequency for each day, and then averaged across all securities, all days in each sample.

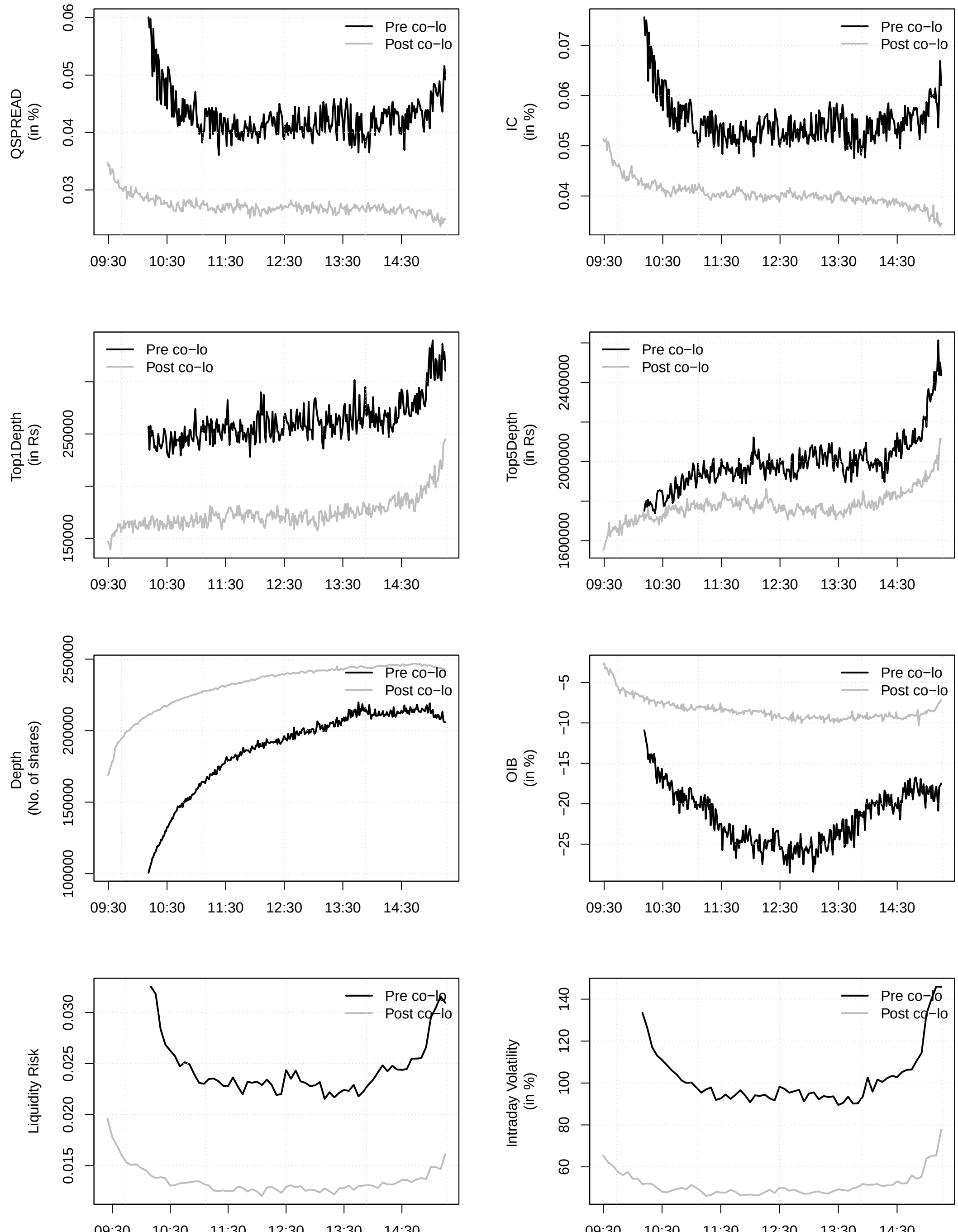


Table 4 Summary statistics of the data

The table presents summary statistics of market quality variable for the one month LOW-AT and HIGH-AT samples.
AT-INTENSITY is measured as the percentage of total traded value in which an AT was present at either one side or both sides of the trade.
QSPREAD is the bid-ask spread as a percentage of mid-quote prices. IC denotes the impact cost computed at a transaction size of Rs. 25,000 (USD 416). TOP1DEPTH shows the Rupee depth at the best bid and ask prices, while TOP5DEPTH shows the cumulated Rupee depth at the top five prices. OIB is the order imbalance measured as the difference between the total outstanding buy side and sell side shares, and expressed as a percentage of average total depth. σ_{ic} is the variance of impact cost. RVOL is the 5-minutes variance of returns of each security. VR is the variance ratio computed as the ratio of ten minutes returns to five minute returns.
Values marked with ** show the values in the HIGH-AT sample which are significantly different from LOW-AT values at 0.05%.

	LOW-AT			HIGH-AT		
	Mean	SD	Median	Mean	SD	Median
AT-INTENSITY (in %)	13.59	7.58	11.04	54.42**	10.69	55.57
Transactions costs						
QSPREAD (in %)	0.05	0.01	0.05	0.03**	0.01	0.03
IC (in %)	0.07	0.02	0.06	0.04**	0.01	0.04
Depth						
TOP1DEPTH (Rs)	619,873	259,074	272,130	267,791**	96,428	176,837
TOP5DEPTH (Rs)	2,850,356	983,148	2,077,488	2,034,486**	579,302	1836,980
DEPTH (No. of shares)	185,447	55,553	194,554	230,949	33,258	239,661
OIB (in %)	-17.14	34.95	-18.66	-8.97**	15.87	-9.26
Risk, annualised (in %)						
RVOL	102.02	13.50	96.83	51.56**	5.71	49.82
LRISK	69.99	15.63	55.25	37.44**	6.18	33.67
Efficiency						
VR (At k=2)	0.96	0.06	0.96	0.94	0.06	0.92

Table 5 Effect of AT on market quality variables

The table presents regression estimation results of model M1 for each of the market quality measures.

- **M1**: $\text{MKT-QUALITY}_{i,t} = \alpha_i + \beta_1 \text{AT-INTENSITY}_{i,t-1} + \beta_2 \text{CO-LO-DUMMY}_t + \epsilon_{i,t}$

where i = 1,...,N indexes firms, t = 1,...,T, indexes 5-minute time intervals. α_i captures the firm specific unobserved factors.
$\text{MKT-QUALITY}_{i,t}$ includes: transactions costs (QSPREAD, IC), depth (TOP1DEPTH, TOP5DEPTH, DEPTH, |OIB|) and market risk (LRISK, RVOL) for security i at t.
The sample includes the days in the LOW-AT sample from Jul 6 to Aug 8, 2009 and the HIGH-AT sample from Jul 6 to Aug 9, 2012. There are 315,115 observations in the data.
The values in parenthesis are Driscoll-Kraay standard errors. Significance levels are marked as: $^+$=p<0.01, **=p<0.05, *=p<0.1. Coefficients and standard errors for AT-INTENSITY are $\times 10^{-2}$.

Panel A: Transactions costs and Rupee depth				
	QSPREAD	IC	TOP1DEPTH	TOP5DEPTH
AT-INTENSITY	-0.01$^+$	-0.01$^+$	-0.09$^+$	-0.17$^+$
	(0.00)	(0.00)	(0.02)	(0.01)
CO-LO-DUMMY	-0.01$^+$	-0.01$^+$	-0.81$^+$	-0.46$^+$
	(0.00)	(0.00)	(0.01)	(0.01)
Obs.	315,115	315,115	315,115	315,115
R^2	0.10	0.07	0.24	0.15
Panel B: Depth and Volatility				
	DEPTH	|OIB|	LRISK	RVOL
AT-INTENSITY	0.10$^+$	4.54$^+$	-0.001**	-5.15$^+$
	(0.01)	(0.49)	(0.000)	(1.12)
CO-LO-DUMMY	0.35$^+$	-30.18$^+$	-0.01$^+$	-46.40$^+$
	(0.01)	(0.96)	(0.00)	(1.77)
R^2	0.18	0.26	0.20	0.26

with 5 minute returns dropped from 0.96 in the LOW-AT sample to 0.94 in the HIGH-AT sample. Although this shows a slight decline in the market efficiency with the increase in AT intensity, the decline is not significant.

Table 5 presents the regression results for the model M1 described in Section 1. This is estimated for all the market quality variables. The coefficient of interest is β_1, which measures the sensitivity of each market quality variable with respect to AT intensity. β_2, which is associated with CO-LO-DUMMY, captures the effect of differences in the time periods of the two samples on market quality variables. This implies that $\hat{\beta}_1$ only captures the effect of AT activity.

These results show a significant negative relation between transactions costs (QSPREAD, IC) and AT-INTENSITY, consistent with the hypothesis that AT improves transactions costs in the market. The results imply that on an average, a 1% increase in AT-INTENSITY reduces QSPREAD and IC by 1 bps.

The total DEPTH (measured as the number of shares) is also positively impacted by AT-INTENSITY. A 1% increase in AT-INTENSITY intensity increases depth by 0.10%. The improvement in transactions costs and total depth is however not matched with an increase in the Rupee depth in the markets. We see that a 1% increase in AT-INTENSITY brings about a 0.09% decline in Rupee depth at best prices, and 0.17% decline at the top five prices, on average. This is contrary to the expectation that AT provides additional liquidity. The coefficient associated with |OIB| also shows an adverse impact of AT on OIB. We see that a 1% increase in AT can bring about 4.54% widening in order imbalance on the limit order books.

Both risk measures – LRISK as well as RVOL – showed a significant decline as a result of AT activity. Higher levels of AT activity are associated with lower levels of liquidity risk.

Overall the results suggest that AT has a positive improvement market quality by way of reduction in transactions costs and reduction in the risks of the security. We also see an improvement in the total depth. But we do not see a corresponding increase in Rupee depth or an improvement in order imbalance. This result is similar to Hendershott *et al.* (2011) who also find a decline in quoted depth after the introduction of co-lo.

7.2 Threats to validity: Did other factors cause the change in market quality?

The results in Section 7.1 show that, around the period of the introduction of co-lo and the subsequent increase of AT intensity, market quality has tended to improve. Transactions costs, securities volatility and liquidity risk have all reduced while total depth has increased as a consequence of AT. On the other hand, the Rupee depth has decreased, and the gap between the buy and the sell side has widened.

However, security liquidity and volatility can be affected by other factors such as macro-economic shocks like the global crisis of 2008, that increase market volatility and liquidity. In that case, could these results be attributed to the increase in AT after co-lo, or were they instead caused reduced market volatility and reduced market price levels (which is the case in the HIGH-AT sample)?

We address such questions using the two approaches described in Section 5.3.

Table 6 Summary statistics of the matched sample

The table presents the summary statistics of the sample where dates were paired from the period before and after the introduction of co-lo that matched on market volatility. QSPREAD is the bid-ask spread as a percentage of mid-quote prices. IC denotes the impact cost computed at a transaction size of Rs. 25,000 (USD 416). TOP1DEPTH shows the Rupee depth at the best bid and ask prices, while TOP5DEPTH shows the cumulated Rupee depth at the top five prices. OIB is the order imbalance measured as the difference between the total outstanding buy side and sell side shares. It is expressed as a percentage of average total depth. σ_{ic} is the variance of impact cost. RVOL is the 5-minutes variance of returns of each security, while Nifty Intraday returns is the 5-minutes variance of Nifty. VR is the variance ratio computed as the ratio of ten minutes returns to five minute returns. Values marked with ** show that the HIGH-AT values are significantly different from LOW-AT values at 0.05%.

	LOW-AT			HIGH-AT		
	Mean	SD	Median	Mean	SD	Median
AT-INTENSITY (in %)	16.81	10.49	14.19	60.76**	10.31	63.13
Transactions costs (in %)						
QSPREAD	0.04	0.02	0.04	0.02	0.10	0.01
IC	0.06	0.02	0.05	0.06	0.09	0.07
Depth						
TOP1DEPTH (Rs.)	648,958	322,657	304,258	248,695**	105,048	155,841
TOP5DEPTH (Rs.)	3,142,270	1,262,099	2,400,423	1,941,818**	611,642	1,752,383
DEPTH (No. of shares)	196,062	66,464	206,005	163,127	55,340	169,969
OIB (in %)	-15.52	32.78	-16.95	-5.85**	24.89	-6.53
Risk, annualised (in %)						
RVOL	94	25	82	73**	19	68
LRISK	61.85	16.25	47.86	42.21**	11.80	38.77
Efficiency						
VR (At k=2)	0.92	0.08	0.94	0.95	0.07	0.96

The first approach includes control variables to capture the macro-economic effects in the fixed effects regression framework. The second is to use the matched sample dataset described in Section 6.

The average behaviour of the AT-INTENSITY and the nine market quality variables for the matched sample are presented in Table 6. These values are similar to the values seen in the unmatched sample. There is a significant increase in the average level of AT-INTENSITY, a decrease in the average level of rupee depth measures, improvement in the OIB and decrease in the average level of intra-day risk measures. However, now the changes in both the measures of transactions costs as well as in total depth by number of shares has become insignificant in the HIGH-AT sample compared to LOW-

AT. This suggests that the some of the improvements in average levels of liquidity seen in Table 4 might have been influenced by the market volatility, rather than the change in average AT-INTENSITY in the market.

Next, we present the estimations of the regression framework with control variables added to capture the macroeconomic factors that can influence market quality. These lead to the three models – M2, M3, M4 – that are described in Section 5.3. The estimates for each of these models for all the nine market quality variables are presented in Appendix A.

Our focus is only on the estimates for the coefficient of AT-INTENSITY, $\hat{\beta}_1$. We ask whether the coefficient value presented in Table 5 has a different magnitude or significance when estimated with controls for macroeconomic factors such as market volatility, or using the matched sample.

Table 7 presents the following four sets of estimates of β_1: (1) **M1** which does not control for macroeconomic factors and other factors (such as the price of the security, intraday effects); (2) **M4**, which controls for macroeconomic factors; (3) **M1'** which is estimated using the matched sample, but without any controls for macroeconomic factors; and (4) **M4'** which is estimated using the matched sample and also includes controls for macroeconomic and other factors. The estimated $\hat{\beta}_1$ is presented for each market quality variable.

The table shows that the results of lower transactions costs – whether for QSPREAD or IC – holds across all the specification. This is also true for the depth that is available at the best price (TOP1DEPTH) and across the best five prices (TOP1DEPTH) in the limit order book. For these four measures of liquidity, these results suggest that AT has improved market quality.

The results also appear to be consistent across all specifications for the two measures of market risk. Intraday price volatility (RVOL) has dropped as a consequence of AT. The magnitude of the drop in RVOL is higher when the estimations use the matched sample, where market volatility is the same for paired dates in the LOW-AT and HIGH-AT samples. This result suggests a stronger result about the impact of AT on intraday volatility compared to what the one-month LOW-AT and HIGH-AT samples suggested. Intraday liquidity risk, LRISK is consistently lower when AT is higher with a negative value of $\hat{\beta}_1$, but the result is weaker than those for intraday price volatility.

In all the above variabels, we infer that AT has improved market quality.

For two of the market quality variables, the results indicate that higher AT leads to poorer market quality. In the case of the difference in the number of shares available on the buy and sell side of the limit order book, OIB,

Table 7 Comparison of the estimated coefficient of AT-INTENSITY on market quality, with and without adjusting for macro-economic effects

The table presents the results of the fixed effects regression, estimated separately using the LOW-AT and HIGH-AT one month and matched samples. The models estimated are:

- **M1**: (estimated using the one month samples, low AT-INTENSITY from July to August 2009 and high AT-INTENSITY from July to August 2012)

  $\text{MKT-QUALITY}_{i,t} = \alpha_i + \beta_1 \text{AT-INTENSITY}_{i,t-1} + \beta_2 \text{CO-LO-DUMMY}_t + \epsilon_{i,t}$

- **M4**: (estimated using the one month sample)

  $\text{MKT-QUALITY}_{i,t} = \alpha_i + \beta_1 \text{AT-INTENSITY}_{i,t-1} + \beta_2 \text{CO-LO-DUMMY}_t + \beta_3 \text{NIFTY-VOL}_t + \beta_4 \text{INTRADAY-DUMMY}_t + \beta_5 \text{LTP}_{i,t} + \epsilon_{i,t}$

- **M1'**: (estimated using the matched sample)

  $\text{MKT-QUALITY}_{i,t} = \alpha_i + \beta_1 \text{AT-INTENSITY}_{i,t-1} + \beta_2 \text{CO-LO-DUMMY}_t + \epsilon_{i,t}$

- **M4'**: (estimated using the matched sample)

  $\text{MKT-QUALITY}_{i,t} = \alpha_i + \beta_1 \text{AT-INTENSITY}_{i,t-1} + \beta_2 \text{CO-LO-DUMMY}_t + \beta_3 \text{NIFTY-VOL}_t + \beta_4 \text{INTRADAY-DUMMY}_t + \beta_5 \text{LTP}_{i,t} + \epsilon_{i,t}$

All coefficients are in terms of 10^{-2}. Significance levels are marked as: $^{+}$=p<0.01, **=p<0.05, *=p<0.1.

	Value of $\hat{\beta}_1$			
	One month sample		Matched sample	
	M1	M4	M1'	M4'
QSPREAD	-0.01^{+}	-0.01^{+}	-0.02^{+}	-0.02^{+}
IC	-0.01^{+}	-0.01^{+}	-0.02^{+}	-0.02^{+}
TOP1DEPTH	-0.09^{+}	-0.10^{+}	-0.08^{+}	-0.10^{+}
TOP5DEPTH	-0.17^{+}	-0.17^{+}	-0.12^{+}	-0.13^{+}
DEPTH	0.10^{+}	0.12^{+}	-0.04^{+}	0.021
|OIB|	4.54^{+}	4.91^{+}	1.45^{+}	2.02^{+}
RVOL	-5.15^{+}	-2.56^{+}	-17.23^{+}	-12.44^{+}
LRISK	-0.001**	-0.00	-0.003^{+}	-0.002^{+}

higher AT leads to a wider gap between the two. These results are consistent across all the models estimated. When the estimated is based on the matched sample, the increase in the gap is smaller.

It is only in the case of DEPTH that the results are inconsistent between the estimation uses the one-month LOW-AT and HIGH-AT samples compared to the matched samples. $\hat{\beta}_1$ when estimations use the matched sample either give a negative value (while the remaining estimates are positive) or an insignificant value. While the results corroborate the findings of Hendershott *et al.* (2011), it needs to be further investigated as to what leads a decline in depth of the markets as a result of high AT-INTENSITY.

Thus, other than for two of the market quality variables, the results show that higher AT leads to improvements in market quality, and that these results are robust even when the estimations adjust for macroeconomic factors.

8 Conclusion

There is a rapidly growing literature on how the presence of algorithmic trading (AT) has changed the liquidity and volatility of markets. This is partly fueled by the regulatory concerns that the use of technology skews the access to markets to a small fraction of the trading community. But, another part of the continuing quest for an answer to this question is founded in the lack of clear identification of whether the orders and trades originate from an AT source or not. While there are some papers that have access to such details from exchanges, the datasets are too small and not comprehensive enough to yield general results. The larger fraction of research is based on proxy measures of AT, which raises questions about the validity of the results.

In this paper, we use access to data from the equity markets of the National Stock Exchange (NSE), where every order is tagged by as AT or non-AT. Unlike in other markets, all equity trading is pooled in two exchanges and the NSE has around 70 percent of the marketshare. Further, the span of this data includes the date when the NSE introduced co-location (co-lo) facilities, which serves to identify a specific date beyond which AT intensity was bound to increase in the market. Therefore, this study helps to address concerns about the lack of generality of previous studies. We use the direct identification of the orders as AT to calculate the AT-INTENSITY in the market any given point in time. We find that the AT-INTENSITY did go up after the introduction of co-lo services, but did not stabilise immediately.

We measure the impact of AT on both the average, or overall market quality, as well as in a cross-sectional analysis. Since the span of the analysis covers a wide period, we control for changes in exogenous factors (such as market volatility) that as likely as higher levels of AT-INTENSITY to cause the changes observed in market quality. We also repeat the comparative analysis on the average and the cross-sectional variation using matched LOW-AT and HIGH-AT samples, where dates are selected that are matched on the level of market volatility.

Both approaches indicate that transactions costs have decreased with higher levels of AT, but that market depth has decreased. The decrease in depth holds for depth measured as total value (in Rupees) available for trade at the touch and at the best five prices available in the limit order book. This result also holds for the overall market depth (in number of shares). These results about lowered costs and worsened depth with higher AT-INTENSITY are similar to those Hendershott *et al.* (2011). The results about intraday price volatility of prices is similar to much of the empirical literature (Hasbrouck and Saar, 2013; Brogaard, 2010). We also test the behaviour of the volatility of transactions costs, and find that liquidity risk has decreased with the rise in AT-INTENSITY. This runs counter to popular arguments that the rise of algorithmic trading has increased liquidity risk in the market. Finally, we analyse the impact of AT on market efficiency as measured by the variance ratio, and find no significant changes in the intra-day behaviour of the variance ratio because of higher AT-INTENSITY.

Thus, the results in the paper mostly validates the findings in the literature about how algorithmic trading affects transactions costs and depth. It adds to the understanding about market volatility by showing that higher AT-INTENSITY significantly improves (has a negative effect) on both intra-day price volatility as well as liquidity risk. Given the clear identification, the comprehensive cover and the span of the data used in the analysis, these results should help address the concerns about the lack of generality of some of the earlier empirical analysis in the literature.

What this work also does is to highlights several new aspects of the impact of AT on the market quality of securities that require further research. One observation that emerges from the analysis is that there is a wide degree of heterogeneity of AT-INTENSITY across securities. In the Indian equity market, where there was a clear event that facilitated AT into the market at the same time, where all trading is pooled into two exchanges with no other dark pools or other avenues of trading available, the question arises as to to why certain securities attract more AT focus than others. Are the selections temporary patterns, driven by the arrival of news, either about the company itself, or the overall market environment? Or are these more structural reasons for these choices, driven by differences in information asymmetry across different companies that are related to differences in corporate disclosure quality? Given the benefits that accrue to the market quality of securities that have a high degree of AT intensity, these are interesting questions for both investors as well as issuers of securities.

A Estimations addressing threats to validity

Table 8 Addressing threats to validity: AT-INTENSITY on transactions costs

The table presents estimation results for transaction costs measures – QSPREAD and IC:

$$\text{M2} : \text{MKT-QUALITY}_{i,t} = \alpha_i + \beta_1\text{AT-INTENSITY}_{i,t-1} + \beta_2\text{CO-LO-DUMMY}_t + \beta_3\text{NIFTY-VOL}_t + \epsilon_{i,t}$$

$$\text{M3} : \text{MKT-QUALITY}_{i,t} = \alpha_i + \beta_1\text{AT-INTENSITY}_{i,t-1} + \beta_2\text{CO-LO-DUMMY}_t + \beta_3\text{NIFTY-VOL}_t + \beta_4\text{INTRADAY-DUMMY}_t + \epsilon_{i,t}$$

$$\text{M4} : \text{MKT-QUALITY}_{i,t} = \alpha_i + \beta_1\text{AT-INTENSITY}_{i,t-1} + \beta_2\text{CO-LO-DUMMY}_t + \beta_3\text{NIFTY-VOL}_t + \beta_4\text{INTRADAY-DUMMY}_t + \beta_5\text{LTP}_{i,t} + \epsilon_{i,t}$$

Panel A presents the regression based control and Panel B show the matched sample estimations. 315,115 observations are used in the Panel A estimations, and 509,376 in the Panel B estimations. Values in parentheses represent Driscoll-Kraay standard errors. Significance levels are marked as: $^+$=p<0.01, **=p<0.05, *=p<0.1. All coefficients and standard errors are $\times 10^{-3}$.

		QSPREAD			IC		
		M2	M3	M4	M2	M3	M4
Panel A	AT-INTENSITY	-0.10$^+$	-0.10$^+$	-0.10$^+$	-0.11$^+$	-0.11$^+$	-0.10$^+$
		(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
	CO-LO-DUMMY	-2.80$^+$	-2.57$^+$	-3.57$^+$	0.08	0.55	-0.74
		(0.70)	(0.72)	(0.71)	(1.02)	(1.04)	(1.02)
	NIFTY-VOL	0.74$^+$	0.76$^+$	0.74$^+$	0.98$^+$	1.02$^+$	0.99$^+$
		(0.07)	(0.07)	(0.07)	(0.10)	(0.10)	(0.10)
	INTRADAY-DUMMY		-1.09**	-1.01**		-2.20$^+$	-2.10$^+$
			(0.46)	(0.46)		(0.69)	(0.68)
	LTP			-0.02$^+$			-0.02$^+$
			(0.00)			(0.00)	
	R^2	0.13	0.13	0.16	0.09	0.09	0.13
Panel B	AT-INTENSITY	-0.19$^+$	-0.19$^+$	-0.17$^+$	-0.22$^+$	-0.22$^+$	-0.19$^+$
		(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
	CO-LO-DUMMY	1.99$^+$	1.97$^+$	1.37$^+$	6.78$^+$	6.73$^+$	5.87$^+$
		(0.43)	(0.43)	(0.40)	(0.65)	(0.65)	(0.61)
	NIFTY-VOL	0.63$^+$	0.64$^+$	0.59$^+$	0.88$^+$	0.92$^+$	0.83$^+$
		(0.04)	(0.04)	(0.03)	(0.05)	(0.06)	(0.05)
	INTRADAY-DUMMY		-0.63	-0.48		-1.66$^+$	-1.44**
			(0.39)	(0.34)		(0.63)	(0.55)
	LTP			-0.01$^+$			-0.02$^+$
				(0.00)			(0.00)
	R^2	0.06	0.06	0.10	0.04	0.04	0.09

Table 9 Addressing threats to validity: AT-INTENSITY on Rupee depth

The table presents estimation results for the two measures of Rupee depth (TOP1DEPTH and TOP5DEPTH) and the following models:

M2 : $\text{MKT-QUALITY}_{i,t} = \alpha_i + \beta_1\text{AT-INTENSITY}_{i,t-1} + \beta_2\text{CO-LO-DUMMY}_t + \beta_3\text{NIFTY-VOL}_t + \epsilon_{i,t}$

M3 : $\text{MKT-QUALITY}_{i,t} = \alpha_i + \beta_1\text{AT-INTENSITY}_{i,t-1} + \beta_2\text{CO-LO-DUMMY}_t + \beta_3\text{NIFTY-VOL}_t + \beta_4\text{INTRADAY-DUMMY}_t + \epsilon_{i,t}$

M4 : $\text{MKT-QUALITY}_{i,t} = \alpha_i + \beta_1\text{AT-INTENSITY}_{i,t-1} + \beta_2\text{CO-LO-DUMMY}_t + \beta_3\text{NIFTY-VOL}_t + \beta_4\text{INTRADAY-DUMMY}_t + \beta_5\text{LTP}_{i,t} + \epsilon_{i,t}$

Panel A presents the regression based control and Panel B show the matched sample estimations. 315,115 observations are used in the Panel A estimations, and 509,376 in the Panel B estimations.
Values in parentheses represent Driscoll-Kraay standard errors. Significance levels are marked as: $^+$=p<0.01, **=p<0.05, *=p<0.1. Coefficients and standard errors for all variables (except CO-LO-DUMMY and INTRADAY-DUMMY) are $\times 10^{-3}$.

		TOP1DEPTH			TOP5DEPTH		
		M2	M3	M4	M2	M3	M4
Panel A	AT-INTENSITY	-0.97$^+$	-0.93$^+$	-0.95$^+$	-1.75$^+$	-1.73$^+$	-1.66$^+$
		(0.16)	(0.16)	(0.16)	(0.13)	(0.13)	(0.13)
	CO-LO-DUMMY	-0.76$^+$	-0.79$^+$	-0.79$^+$	-0.44$^+$	-0.47$^+$	-0.48$^+$
		(0.02)	(0.02)	(0.02)	(0.02)	(0.01)	(0.02)
	NIFTY-VOL	4.66**	0.90	0.95	1.58	-1.26	-1.45
		(1.88)	(1.48)	(1.48)	(1.76)	(1.47)	(1.47)
	INTRADAY-DUMMY		0.18$^+$	0.18$^+$		0.13$^+$	0.13$^+$
			(0.02)	(0.02)		(0.01)	(0.01)
	LTP			0.03$^+$			-0.13$^+$
				(0.01)			(0.01)
	R^2	0.24	0.25	0.25	0.15	0.16	0.159
Panel B	AT-INTENSITY	-0.80$^+$	-0.81$^+$	-1.04$^+$	-1.30$^+$	-1.30$^+$	-1.29$^+$
		(0.21)	(0.20)	(0.19)	(0.17)	(0.17)	(0.16)
	CO-LO-DUMMY	-0.93$^+$	-0.93$^+$	-0.92$^+$	-0.58$^+$	-0.58$^+$	-0.58$^+$
		(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.0)1
	NIFTY-VOL	-0.64	-5.66$^+$	-4.95$^+$	-1.36	-4.58$^+$	-4.62$^+$
		(1.70)	(1.62)	(1.58)	(1.48)	(1.41)	(1.41)
	INTRADAY-DUMMY		0.21$^+$	0.21$^+$		0.13$^+$	0.13$^+$
			(0.02)	(0.02)		(0.01)	(0.01)
	LTP			0.16$^+$			-0.01
				(0.01)			(0.01)
	R^2	0.29	0.30	0.30	0.20	0.20	0.20

Table 10 Addressing threats to validity: AT-INTENSITY on depth

The table presents estimation results for the two depth measures (DEPTH and |OIB|) of models: as:

M2 : $\text{MKT-QUALITY}_{i,t} = \alpha_i + \beta_1\text{AT-INTENSITY}_{i,t-1} + \beta_2\text{CO-LO-DUMMY}_t + \beta_3\text{NIFTY-VOL}_t + \epsilon_{i,t}$

M3 : $\text{MKT-QUALITY}_{i,t} = \alpha_i + \beta_1\text{AT-INTENSITY}_{i,t-1} + \beta_2\text{CO-LO-DUMMY}_t + \beta_3\text{NIFTY-VOL}_t + \beta_4\text{INTRADAY-DUMMY}_t + \epsilon_{i,t}$

M4 : $\text{MKT-QUALITY}_{i,t} = \alpha_i + \beta_1\text{AT-INTENSITY}_{i,t-1} + \beta_2\text{CO-LO-DUMMY}_t + \beta_3\text{NIFTY-VOL}_t + \beta_4\text{INTRADAY-DUMMY}_t + \beta_5\text{LTP}_{i,t} + \epsilon_{i,t}$

Panel A show the regression based control and Panel B show the matched sample estimations. 315,115 observations are used in the Panel A estimations, and 509,376 in the Panel B estimations.
Values in parentheses represent Driscoll-Kraay standard errors. Significance levels are marked as: $^+$=p<0.01, **=p<0.05, *=p<0.1. Coefficients and standard errors for AT-INTENSITY are $\times 10^{-2}$.

		DEPTH			|OIB|		
		M2	M3	M4	M2	M3	M4
Panel A	AT-INTENSITY	0.096$^+$	0.094$^+$	0.121$^+$	4.7307$^+$	4.684$^+$	4.909$^+$
		(0.009)	(0.009)	(0.009)	(0.491)	(0.487)	(0.487)
	CO-LO-DUMMY	0.312$^+$	0.327$^+$	0.298$^+$	-24.769$^+$	-24.231$^+$	-24.473$^+$
		(0.020)	(0.020)	(0.021)	(1.351)	(1.399)	(1.409)
	NIFTY-VOL	-0.004$^+$	-0.002**	-0.003**	0.511$^+$	0.563$^+$	0.557$^+$
		(0.001)	(0.001)	(0.001)	(0.136)	(0.141)	(0.142)
	INTRADAY-DUMMY		-0.071$^+$	-0.069$^+$		-2.537$^+$	-2.520$^+$
			(0.016)	(0.016)		(0.888)	(0.890)
	LTP			-0.001$^+$			-0.004$^+$
				(0.000)			(0.001)
	R^2	0.18	0.18	0.27	0.26	0.26	0.26
Panel B	AT-INTENSITY	-0.045$^+$	-0.045$^+$	0.021	1.685$^+$	1.685$^+$	2.019$^+$
		(0.013)	(0.013)	(0.014)	(0.471)	(0.471)	(0.469)
	CO-LO-DUMMY	-0.189$^+$	-0.192$^+$	-0.212$^+$	-13.595$^+$	-13.594$^+$	-13.697$^+$
		(0.016)	(0.015)	(0.016)	(0.641)	(0.645)	(0.642)
	NIFTY-VOL	-0.006$^+$	-0.003$^+$	-0.005$^+$	0.428$^+$	0.427$^+$	0.417$^+$
		(0.001)	(0.001)	(0.001)	(0.069)	(0.073)	(0.073)
	INTRADAY-DUMMY		-0.110$^+$	-0.105$^+$		0.044	0.072
			(0.015)	(0.016)		(0.558)	(0.559)
	LTP			-0.001$^+$			-0.002$^+$
				(0.001)			(0.030)
	R^2	0.05	0.05	0.15	0.07	0.07	0.07

Table 11 Addressing threats to validity: AT-INTENSITY on risk

The table presents estimation results of two risk based measures (LRISK and RVOL) of models:

M2 : $\text{MKT-QUALITY}_{i,t} = \alpha_i + \beta_1\text{AT-INTENSITY}_{i,t-1} + \beta_2\text{CO-LO-DUMMY}_t + \beta_3\text{NIFTY-VOL}_t + \epsilon_{i,t}$

M3 : $\text{MKT-QUALITY}_{i,t} = \alpha_i + \beta_1\text{AT-INTENSITY}_{i,t-1} + \beta_2\text{CO-LO-DUMMY}_t + \beta_3\text{NIFTY-VOL}_t + \beta_4\text{INTRADAY-DUMMY}_t + \epsilon_{i,t}$

M4 : $\text{MKT-QUALITY}_{i,t} = \alpha_i + \beta_1\text{AT-INTENSITY}_{i,t-1} + \beta_2\text{CO-LO-DUMMY}_t + \beta_3\text{NIFTY-VOL}_t + \beta_4\text{INTRADAY-DUMMY}_t + \beta_5\text{LTP}_{i,t} + \epsilon_{i,t}$

Panel A show the regression based control and Panel B show the matched sample estimations. 315,115 observations are used in the Panel A estimations, and 509,376 in the Panel B estimations.
Values in parentheses represent Driscoll-Kraay standard errors. Significance levels are marked as: $^+$=p<0.01, **=p<0.05, *=p<0.1. Coefficients and standard errors for AT-INTENSITY are $\times 10^{-2}$.

		LRISK			RVOL		
		M2	M3	M4	M2	M3	M4
Panel A	AT-INTENSITY	-0.000*	-0.000*	-0.000	-4.031$^+$	-3.807$^+$	-2.562$^+$
		(0.000)	(0.000)	(0.000)	(0.820)	(0.799)	(0.764)
	CO-LO-DUMMY	-0.003$^+$	-0.004$^+$	-0.004$^+$	-13.679$^+$	-16.338$^+$	-17.683$^+$
		(0.000)	(0.000)	(0.000)	(1.570)	(1.705)	(1.70410)
	NIFTY-VOL	0.001$^+$	0.001$^+$	0.001$^+$	3.091$^+$	2.831$^+$	2.799$^+$
		(0.000)	(0.000)	(0.000)	(0.164)	(0.186)	(0.187)
	INTRADAY-DUMMY		0.001$^+$	0.001$^+$		12.554$^+$	12.653$^+$
			(0.000)	(0.000)		(1.149)	(1.145)
	LTP			-0.000$^+$			-0.022$^+$
				(0.000)			(0.000)
	R^2	0.26	0.26	0.27	0.36	0.37	0.39
Panel B	AT-INTENSITY	-0.002$^+$	-0.002$^+$	-0.002$^+$	-15.581$^+$	-15.624$^+$	-12.440$^+$
		(0.000)	(0.000)	(0.000)	(1.006)	(0.995)	(0.910)
	CO-LO-DUMMY	-0.003$^+$	-0.003$^+$	-0.003$^+$	-3.890$^+$	-3.517$^+$	-4.502$^+$
		(0.000)	(0.000)	(0.000)	(0.949)	(0.893)	(0.844)
	NIFTY-VOL	0.001$^+$	0.001$^+$	0.001$^+$	2.955$^+$	2.643$^+$	2.547$^+$
		(0.000)	(0.000)	(0.000)	(0.126)	(0.116)	(0.112)
	INTRADAY-DUMMY		0.002$^+$	0.002$^+$		13.244$^+$	13.504$^+$
			(0.000)	(0.000)		(0.937)	(0.888)
	LTP			-0.000$^+$			-0.021$^+$
				(0.000)			(0.000)
	R^2	0.09	0.09	0.11	0.13	0.14	0.17

References

Biais B, Foucault T, Moinas S (2013). "Equilibrium Fast Trading." *Working paper*. URL `http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1859265`.

Bohemer E, Fong K, Wu J (2012). "International Evidence on Algorithmic Trading." *Working Paper*. URL `http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2022034`.

Brogaard J (2010). "High Frequency Trading and Its Impact on Market Quality." *Working Paper*. URL `www.futuresindustry.org/ptg/downloads/HFT_Trading.pdf`.

Brogaard J, Hendershott T, Riordan R (2012). "High frequency trading and price discovery." *Working Paper, University of California at Berkeley*.

Carrion A (2013). "Very fast money: High-frequency trading on the {NASDAQ}." *Journal of Financial Markets*, **16**(4), 680 – 711.

Cartea A, Penalva J (2012). "Where is the Value in High Frequency Trading?" *Quarterly Journal of Finance*, **02**(03), 1250014.

Chaboud A, Chiquoine B, Hjalmarsson E, Vega C (2009). "Rise of the machines: algorithmic trading in the foreign exchange market." URL `http://ideas.repec.org/p/fip/fedgif/980.html`.

Chordia T, Roll R, Subrahmanyam A (2000). "Commonality in liquidity." *Journal of Financial Economics*, **56**(1), 3 – 28.

Driscoll JC, Kraay AC (1998). "Consistent Covariance Matrix Estimation With Spatially Dependent Panel Data." *The Review of Economics and Statistics*, **80**(4), 549–560.

Foucault T (1999). "Order flow composition and trading costs in a dynamic limit order market." *Journal of Financial Markets*, **2**(2), 99 – 134.

Frino A, Mollica V, Webb RI (2013). "The impact of co-location of securities exchanges' and traders' computer servers on market liquidity." *Journal of Futures Markets*. ISSN 1096-9934.

Grossman SJ, Miller MH (1988). "Liquidity and Market Structure." *The Journal of Finance*, **43**(3), 617–633.

Hameed A, Kang W, Vishwanathan S (2010). "Stock Market Declines and Liquidity." *The Journal of Finance*, **65**(1), 257–293.

Hasbrouck J, Saar G (2013). "Low-latency trading." *Journal of Financial Markets*, **16**(4), 646 – 679.

Hendershott T, Jones CM, Menkveld AJ (2011). "Does Algorithmic Trading Improve Liquidity?" *The Journal of Finance*, **66**(1), 1–33.

Hendershott T, Riordan R (2013). "Algorithmic trading and the market for liquidity." *The Journal of Financial and Quantitative Analysis, Forthcoming.*

Hoffmann P (2012). "A dynamic limit order market with fast and slow traders." *Technical report*, University Library of Munich, Germany. URL http://ideas.repec.org/p/pra/mprapa/39855.html.

Jovanovic B, Menkveld AJ (2010). "Middlemen in Limit Order Markets." *2010 meeting papers*, Society for Economic Dynamics. URL http://ideas.repec.org/p/red/sed010/955.html.

Lo A, MacKinlay A (1988). "Stock market prices do not follow random walks: evidence from a simple specification test." *Review of Financial Studies*, **1**(1), 41–66.

Martinez VH, Rosu I (2013). "High-frequency traders, news and volatility." *Working paper*. URL http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1859265.

Menkveld AJ (2013). "High frequency trading and the new market makers." *Journal of Financial Markets*, **16**(4), 712 – 740. URL http://www.sciencedirect.com/science/article/pii/S1386418113000281.

Moura M, Pereira F, Attuy G (2013). "Currency Wars in Action: How Foreign Exchange Interventions Work in an Emerging Economy." (304/2013). URL http://www.insper.edu.br/wp-content/uploads/2013/06/2013_wpe304.pdf.

Riordan R, Storkenmaier A (2012). "Latency, liquidity and price discovery." *Journal of Financial Markets*, **15**(4), 416 – 437.

SEBI (2013). "Table 2: The Basic Indicators in Cash Market." SEBI Bulletin, November 2013, www.sebi.gov.in/cms/sebi_data/attachdocs/1385638179478.pdf.

Thomas S (2010). "Call auctions: A solution to some difficulties in Indian finance." *Working Paper*. URL http://www.igidr.ac.in/pdf/publication/WP-2010-006.pdf.

Viljoen T, Zheng H, Westerholm PJ (2011). "Liquidity and Price Discovery of Algorithmic Trading: An Intraday Analysis of the SPI 200 Futures." URL http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1913693.

Zhang F (2010). "High-Frequency Trading, Stock Volatility, and Price Discovery." URL http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1691679.