The information content of high frequency trading volume

The information content of high frequency trading volume

KHALADDIN RZAYEV
University of Edinburgh, United Kingdom

GBENGA IBIKUNLE*
University of Edinburgh, United Kingdom
European Capital Markets Cooperative Research Centre, Pescara, Italy
Centre for Responsible Banking & Finance, University of St Andrews, United Kingdom

Abstract

We propose a state space modelling approach for decomposing high frequency trading volume into liquidity-driven and information-driven components. Based on a set of high frequency S&P 500 stocks data, we show that informed trading increases pricing efficiency by reducing volatility, illiquidity and toxicity/adverse selection during periods of non-aggressive trading. We also find that our estimated informed trading component of volume is a statistically significant predictor for one-second stock returns, but is not a significant predictor for one-minute stock returns; this disparity is explained by high frequency trading activity, which leads to the elimination of pricing inefficiencies at high frequencies.

JEL Classification: G12; G14; G15

Keywords: trading volume; expected component; unexpected component; volatility; liquidity; market toxicity, unobserved components time series models; high-frequency data.

* Corresponding author. Contact information: University of Edinburgh Business School, 29 Buccleuch Place, Edinburgh EH8 9JS, United Kingdom; e-mail: Gbenga.Ibikunle@ed.ac.uk; phone: +441316515186.

1. Introduction

Market participants' trades are driven by either information or the search for liquidity (see Admati & Pfleiderer 1988). Liquidity traders do not trade on the basis of any specific information; their trading strategies are therefore not directly related to future payoffs. The trading strategies of informed traders, on the other hand, are based on private information and are directly related to future payoffs. The activities of these two fundamental types of traders have been extensively analysed in seminal papers in the larger financial markets literature and more so in the market microstructure literature (see as examples Glosten & Milgrom 1985; Kyle 1985; Collin-Dufresne & Fos 2016). For example, Kyle (1985) predicts that the volatility of asset prices partially reflects inside information (informed trading) and is independent of liquidity-driven trading effects, while Glosten and Milgrom (1985) predict that the breadth of the bid-ask spread is primarily driven by informed trading, which incorporates adverse selection costs into the spread. In both models, it is assumed that traders execute their trading strategies by using marker orders; thus, all traders trade aggressively in both models. More recently however, Collin-Dufresne and Fos (2016) extend Kyle's (1985) model and show that the relationship between stock price volatility and informed trading depends on the aggressiveness of traders. Furthermore, in contrast to Glosten and Milgrom's (1985) model, Collin-Dufresne and Fos (2016) predict that informed trading may be negatively correlated with adverse selection if informed traders execute their strategies using limit orders. Using a comprehensive sample of trades from Schedule 13D filings by activist investors, Collin-Dufresne and Fos (2015) show that informed traders with long-lived information tend to use limit orders, which leads to a negative correlation between adverse selection and informed trading (see also Kaniel & Liu 2006).

This paper builds on the above predictions and findings. We first develop a general state space-based methodology for decomposing trading volume into liquidity-driven and information-driven components. Specifically, we demonstrate that observable percentage change in trading volume is a sum of two unobserved series: a nonstationary series (expected component) and a stationary series (unexpected component). We argue that the expected component of trading volume is mainly driven by liquidity traders, whereas the unexpected component is primarily driven by informed traders. Our argument is anchored on two reasons. Firstly, the expected component in the state space model is a nonstationary series and follows a random walk. Consistent with the literature, it is reasonable to argue that liquidity traders trade randomly (i.e. the reference to noise trading in the market microstructure literature), and thus we model the trading volume of liquidity traders as a random walk (see as examples Kyle 1985; Admati & Pfleiderer 1988). Secondly, in state space models, changes in the expected component affect the observable variable permanently, while changes in the unexpected component have a transitory impact on the observable variable, in this case, trading volume (see Hendershott & Menkveld 2014).

Secondly, using the estimated expected and unexpected components of trading volume, we examine the role of liquidity and informed traders on market quality metrics, such as volatility, liquidity and toxicity. This part of our analysis serves two purposes, by being a joint test of the empirical relevance of our state space model and the impact of different traders on market quality. The relevance of our state space approach is examined by relating our empirical findings to model predictions in the existing relevant theoretical market microstructure literature.

Thirdly, we examine the predictive power of the estimated information-driven/unexpected component of trading volume on short-horizon returns. This analysis furthers our aim of demonstrating the relevance of the state space approach to decomposing trading volume into informed and liquidity components. It is also a direct test of the efficiency of the price discovery process (see Chordia *et al.* 2005, 2008). Similar to order imbalance metrics employed in Chordia *et al.* (2008), the unexpected component also signals private information, and we expect it to be a predictor of short-horizon returns.

Our results are generally consistent with our expectations. Based on our state space-estimated information and liquidity-driven components of trading volume, we find that stock price volatility is independent of liquidity trading, but impacted by information-motivated trading (see Glosten & Milgrom 1985; Kyle 1985). We also find that information-motivated trading volume improves pricing efficiency by reducing price volatility and market toxicity, and improving liquidity; the results are robust to alternative estimation frequencies, and volatility and liquidity proxies. This finding is in line with the theoretical model developed by Collin-Dufresne and Fos (2016), which predicts that the price volatility-informed trading relationship is influenced by two effects. On the one hand, informed trading reveals information, and this decreases uncertainty in financial markets, which reduces price volatility. On the other hand, the aggressive behaviour of informed traders could increase volatility. Thus, the net impact of informed trading on stock price volatility depends on which effect dominates. Thus, our finding in relation to volatility is linked to the period of relative calm in S&P 500 stocks, which we examine. Furthermore, Menkveld (2013) shows that aggressive trading is not profitable during normal trading periods, i.e. trading periods are considered normal if there is no excessive aggressiveness, such as a flash crash. This implies that informed traders do not tend to use aggressive orders during periods of relative calm in financial markets; thus, their activities could lead to a reduction of volatility in the markets, as predicted by Collin-Dufresne and Fos (2016). The results are also consistent with the findings of Avramov *et al.* (2006) and Collin-Dufresne and Fos (2015), who find that price volatility and adverse selection are negatively correlated with informed trading. The negative relationships of informed trading with order flow toxicity and illiquidity are linked to informed traders' use of limit orders rather than (aggressive) market orders. In a large part of the market microstructure literature, it is generally assumed that informed traders use only market orders, and therefore it is expected that informed traders increase aggressiveness and widen the bid-ask spread, a proxy for illiquidity and, by extension one of its components, adverse selection (or its high frequency equivalent, market toxicity). However, Kaniel and Liu (2006), modifying Glosten and Milgrom's (1985) model, demonstrate that if there is a high probability that the information to be exploited is long-lived, then informed traders tend to submit limit orders. The prediction of Kaniel and Liu's (2006) model is empirically confirmed by Collin-Dufresne and Fos (2015), who find that informed traders with long-lived information tend to use limit orders, which leads to a reduction in adverse selection.[1]

Finally, we find that information-motivated trading is a significant predictor of one-second stock returns. It implies that although financial markets are efficient in the long-term, there are short-term inefficiencies in markets because investors need time to absorb new information (see Chordia *et al.* 2008). However, we find that the horizon for short-term stock returns predictability has decreased substantially since the five-minute window reported by Chordia *et al.* (2008). We find that the predictability of short-horizon returns only holds on a per second basis, and no longer at the minutes-long threshold reported in earlier studies. This is linked to the impact of high frequency trading activity.

A few streams of the literature are related to this study. There are those studies delineating traders into liquidity-driven and information-driven traders (see as an example Avramov *et al.* 2006), and another, extensive, stream examining the role of the different types of traders on price volatility and liquidity (see as examples Daigler & Wiley 1999; Avramov *et al.* 2006; Van Ness *et al.* 2016). This current paper differs from these studies in several aspects. Firstly, the approach we present is fundamentally different to those employed in the existing studies. Secondly, we generally examine the role of informed trading activity in the evolution of specific market quality metrics, including a new proxy, market toxicity – a role

[1] The rational expectation model developed by Wang (1993) also predicts a negative correlation between informed trading and stock price volatility, but via a different mechanism. Furthermore, Admati and Pfleiderer (1988) also argue in favour of a negative relationship between adverse selection and informed trading.

not well documented in the literature. Finally, we present new evidence on the speed of price adjustment in financial markets.

2. Theory and the previous literature

In this paper, we decompose trading volume into liquidity and information-driven components, and thereafter test the empirical relevance of our model and the role of liquidity and informed traders in the price discovery process. Our empirical analysis is based on the predictions of widely accepted theories as proposed in existing studies. Thus, this paper is related to the stream of literature investigating the impact of asymmetric information on asset prices' volatility and liquidity. Kyle (1985) presents one of the first and best-established models deriving equilibrium security prices when traders possess asymmetric information. The model assumes three types of traders in a market: a market maker, a noise trader that trades randomly, and an informed trader, and also provides a framework for determining the price impact of trading volume. The model shows that stock price volatility partially reflects inside information, which is independent of noise trading volatility. Furthermore, the model predicts that informed traders trade more actively when there is a higher level of noise trading volume in the markets, because the higher uninformed trading volume provides a "camouflage" for informed order flow. Glosten and Milgrom (1985) model the bid-ask spread and propose a then new explanation on why it arises in financial markets. The model predicts that adverse selection implies that the market maker makes losses whenever trading with insiders, and hence she is forced to impose different charges on buy and sell volumes in order to compensate for her potential losses. In other words, the model predicts that the bid-ask spread depends on informed trading activity and the independence of liquidity traders. Moreover, the model predicts that the higher the variance of prices, the greater the impact of insiders/informed traders on the bid-ask spread. Consistent with Glosten and Milgrom (1985), Easley and O'Hara (1987) also suggest that stock illiquidity should increase in the presence of informed traders, as information asymmetry increases adverse selection, which widens the spread.

In both Kyle's (1985) and Glosten and Milgrom's (1985) models, the liquidity traders trade randomly. By contrast, Admati and Pfleiderer (1988) argue that this is a strong assumption and it might be more reasonable to assume that at least some liquidity traders can select the timing of their transactions. Consistent with the literature, this model predicts that the information-motivated trades increase as liquidity driven trading volumes rise, and the variance of price changes is independent of the variance of liquidity traders. However, surprisingly, the theoretical framework predicts that adverse selection decreases the number of informed traders. Admati and Pfleiderer (1988) argue that informed traders in possession of the same set of information will compete, and that this competition reduces adverse selection and increases benefits to liquidity traders.

As already noted, generally, theoretical models examining information asymmetry in the price discovery process assume that informed traders execute their trading strategies by using market orders, i.e. they are aggressive traders (see as examples Glosten & Milgrom 1985; Kyle 1985). Popular models such as the probability of informed trading (PIN) model, developed by Easley *et al.* (1996) and Easley *et al.* (1997), also make this assumption. In contrast to these models, Kaniel and Liu (2006) argues that the assumption is unnecessarily strong. By extending the Glosten and Milgrom's (1985) model, the authors show that informed traders with long lived information strategically tend to use limit orders instead of market orders (see also Sun & Ibikunle 2016). Collin-Dufresne and Fos (2016) also extend Kyle's (1985) model of insider trading and show that the impact of informed trading on the price discovery process is two-fold and could be explained by two mechanisms. Firstly, informed traders reveal information, which decreases the level of price uncertainty in the market; thus, stock price volatility is negatively correlated with informed trading. Secondly, informed traders could trade aggressively, and this aggressive behaviour increases stock price volatility in financial markets; hence, stock price volatility is positively correlated with informed traders. Therefore, the relationship between market quality characteristics, such as price volatility, and informed traders depends on which effect dominates the other. The majority of market microstructure models predict positive correlations between informed trading and stock price volatility because they assume that informed traders will aim to quickly take advantage of private information by seeking to execute market orders based on such information. However, Menkveld (2013) and Rzayev and Ibikunle (2017) show that aggressive trading is not profitable for informed traders if there is no widespread aggression in the market. This implies that during calmer periods, we would expect to see a negative relationship between informed trading volume and stock price volatility (see also Kaniel & Liu 2006; Collin-Dufresne & Fos 2015). The negative informed trading-price volatility relationship is also predicted by rational expectations models (see as examples Hellwig 1980; Wang 1993).

While the relationship between informed trading volume and price volatility is nuanced, a positive relationship between aggregate trading volume, i.e. containing informed and uninformed volume, and stock price volatility, is widely documented (see as an example the studies summarized in Karpoff 1987). Generally, the impact of trading volume on stock price volatility is explained by some related theories. We mainly focus on two well-known and widely accepted theories: information theories and dispersion of beliefs theories. Information theories, such as a mixture of distributions models and sequential arrival of information models, suggest that both volatility and volume are determined by information arrivals (see Copeland 1976; Epps & Epps 1976; Copeland 1977). The dispersion of beliefs theory, modelled by Harris and Raviv (1993) and Shalen (1993), argues that both unusual volume and volatility are associated with the differences in traders' beliefs. To put it simply, the dispersion of beliefs model/theory incorporates the role of different types of traders into the relationship between trading volume and stock price volatility.

In most existing studies, trading activity is measured by total trading volume. However, as already noted, the dispersion of beliefs models argue that this relationship depends on the differences in traders' beliefs, and thus linking volatility to total trading volume conceals some important information (see also Chordia *et al.* 2002). Therefore, some studies decompose trading volume into its components and then examine the role of different trading components on market quality characteristics, such as stock price volatility and market liquidity (see as examples Bessembinder & Seguin 1993; Daigler & Wiley 1999; Avramov *et al.* 2006). Avramov *et al.* (2006) partition trades into two components: herding (non-informed) and contrarian (informed) trades. Consistent with the rational expectation models, Avramov *et al.* (2006) find that herding trades increase stock price volatility, however contrarian trades reduce it. Collin-Dufresne and Fos (2015) directly examine the role of informed traders in the pricing process by using a comprehensive sample of trades from Schedule 13D filings by activist investors, and conclude that when informed traders can select when (they could strategically trade when noise trading is high) and how (they might strategically select to use limit orders) to trade, their trading activity decreases adverse selection in financial markets.

We extend this study to examine the effects of informed trading on market toxicity, and then relate it to Van Ness *et al.* (2016). Van Ness *et al.* (2016) investigate the role of high frequency traders (HFTs) in order flow toxicity by employing the Easley *et al.* (2011, 2012) volume-synchronized probability of informed trading (VPIN) metric as a measure of order flow toxicity. Their study finds a negative correlation between HFT activity and order flow toxicity. It indicates that, as HFT increases, average order flow toxicity decreases. Furthermore, the authors observe a negative correlation between trading volume and order flow toxicity; specifically, as volume increases, average market toxicity decreases.

Finally, our approach for decomposing trading volume into informed and uninformed components is based on state space modelling; therefore, our paper is also related to yet another stream of the market microstructure literature, which employs state space models. Generally, the existing body of literature on market microstructure uses state space modelling only for decomposing price into two components (see as examples Menkveld *et al.* 2007; Brogaard *et al.* 2014; Hendershott & Menkveld 2014) rather than volume. Menkveld *et al.* (2007) use the approach to analyse around-the-clock price discovery for cross-listed stocks in the Amsterdam exchange and NYSE. Their study finds that NYSE plays a minor role in the price discovery process for Dutch stocks. Similar to Menkveld *et al.* (2007), Brogaard *et al.* (2014) use a state space model in order to analyse the price discovery process in the US market. More precisely, they examine the role of high frequency trading (HFT) in the price discovery process. The study reports a positive role for HFT in the price discovery process. Durbin and Koopman (2012) provide a more detailed discussion on the advantages of state space models.

3. Data and Methodology

3.1 Data

The data employed consists of ultra-high frequency tick-by-tick data for the most active 100 S&P 500 stocks sourced from the Thomson Reuters Tick History (TRTH) database. Appendix B lists the stocks that are examined. The data spans October 2016 – September 2017. In the data, each message is recorded with a time stamp to the nearest millisecond. The following variables are included in the dataset: Reuters Identification Code (RIC), date, timestamp, price, volume, bid price, ask price, bid volume, and ask volume. We then follow Chordia *et al.* (2001) and Ibikunle (2015) in applying a standard set of exclusion criteria to the data, with the aim of excluding inexplicable values that may arise due to erroneous data entries.

Table 1 presents the summary statistics of trading activities for the final sample of stocks.

INSERT TABLE 1 ABOUT HERE

In order to classify trades as buyer- or seller-initiated, we apply the Lee and Ready (1991) algorithm.[2] Going by the number of transactions and nominal and dollar-denominated trading volume, the sell side appears marginally more active than the buy side over the sample period. This view is further underscored by the average trade sizes for both buys and sells. The sellers also appear more aggressive, based on the average sizes of their trades.

3.2 Main Variables

A key aim of this study is to examine the role of informed and liquidity traders in the evolution of price volatility, liquidity and market toxicity. This inevitably translates into a joint test of the empirical relevance of the state space model we employ, as well as the impact of the different types of traders on several market quality metrics. Specifically, we build a set of predictive regressions to test the impact of expected and unexpected components of traded volume on price volatility, liquidity and market toxicity. Thus, our volatility, liquidity and market toxicity measures are the main variables of interest.

Consistent with the literature, we use absolute price changes to measure stock price volatility. For robustness, we also use the standard deviation of stock returns (see as examples Karpoff 1987; Lamoureux & Lastrapes 1990) as a proxy for stock price volatility. Absolute price change is defined as the absolute value of the differences between prices at time t and t_{-1}, and we use one-second intervals for computing the absolute price changes. To compute the standard deviation of stock returns, firstly we employ the midpoint of the bid and ask quotes corresponding to every transaction.[3] For robustness, we also compute the standard deviation of stock returns by computing the returns from the execution price for each transaction rather than the midpoint of the prevailing quotes.

[2] Chakrabarty *et al.* (2015) compare the different trades classification methods and conclude that Lee and Ready's (1991) is the most accurate method.
[3] Chordia *et al.* (2008) and Avramov *et al.* (2006) employ midpoint returns to reduce bid-ask bounce.

For robustness, we employ three spread measures as proxies for liquidity; the spread metrics are effective spread, quoted spread, and relative spread. The relative and quoted spread measures are computed using the best bid and ask prices for each interval, *t*, which corresponds to one second.[4] The relative bid-ask spread is obtained by dividing the difference between ask and bid prices by the midpoint of both prices, while the quoted spread is simply the difference between the ask and bid prices. The effective spread is twice the absolute value of the difference between the last transaction price in an interval, *t*, which corresponds to one second, and the midpoint of the prevailing bid and ask prices.

We use the order imbalance (*OIB#*) metric proposed by Chordia *et al.* (2008) as a proxy for the level of order toxicity in the market. This is because existing order toxicity measures, such as the volume synchronised probability of informed trading (VPIN – see Easley *et al.* 2012), essentially capture the essence of order imbalance in the market and thus are highly correlated with *OIB#*. *OIB#* is computed as the absolute value of the number of buyer-initiated trades minus the number of seller-initiated trades divided by the total number of trades during the interval, *t*. In this case, *t* equals one minute rather than one second. We employ the one-minute interval to compute market toxicity, because it is challenging to obtain enough trading volume for the lower volume stocks to compute unbiased order imbalance metrics within a one-second interval.[5]

Apart from the main variables discussed above, there are a few other variables that are critical to our analysis. In our state space model, trading volume change is an observable variable, which is decomposed into two unobservable variables – the expected/uninformed/liquidity and unexpected/informed components. Thus, the unexpected and expected components should be mechanically correlated with trading activity and volume. This implies that we need to include at least one proxy for trading volume and activity in our

[4] For robustness, we also employ the last bid and ask quotes for each interval.
[5] For robustness and consistency, we also employ the one-minute estimation interval for price volatility and bid-ask spread models; the results are presented in Appendix A.

secondary models to control for volume. To this end, we employ the percentage change in trading volume as the first and main control for trading volume, since the state space-estimated components are driven changes in trading volume (see also Chordia *et al.* 2002).[6] Our second proxy is the absolute value of buyer-less seller-initiated trades, which should adequately proxy trading activity because of Chordia *et al.* (2002)'s argument that the metric should strongly affect prices and liquidity (see also Collin-Dufresne & Fos 2015).[7] Table 2 presents summary statistics associated with our variables.[8]

INSERT TABLE 2 ABOUT HERE

Table 2 presents the descriptive statistics for measures of liquidity, volatility, toxicity and return used in this study. The average effective, relative and quoted spreads are about 0.009, 0.0004 and 0.018, respectively. Average returns are weakly negative from October 2016 to September 2017. The mean and median for the absolute price change are about 0.0092 and 0.009 respectively. The average percentage changes in trading volume is positive at 28.22; hence, trading volume increases during our sample period. The average market toxicity metric (order imbalance developed by Chordia *et al.* 2008) is high at 0.54067, since it is computed over one-minute intervals.

3.3 State Space Model

Transactions in financial markets are motivated either by liquidity or information (see Admati & Pfleiderer 1988). As predicted by the theoretical models of Kyle (1985) and Glosten and Milgrom (1985), liquidity and informed order flows have different impacts on price changes and the bid-ask spread (see also Wang 1993; Collin-Dufresne & Fos 2016). Avramov *et al.* (2006) empirically measure the relative impact of informed and liquidity traders on

[6] For robustness, we also use the natural logarithm of trading volume as a proxy for trading activities and obtain completely consistent results.
[7] The correlation between these two proxies is very low.
[8] Descriptive statistics for expected and unexpected components are provided in Section 4.

financial instruments and document the different impacts of these traders (see also Collin-Dufresne & Fos 2015). In this paper, we aim to disentangle liquidity and informed trading volume and examine their relative impacts on price volatility, liquidity and market toxicity, using the state space approach. State space models are a natural tool for modelling an observed variable as the sum of two unobserved variables (see Hendershott & Menkveld 2014). Thus, our approach involves showing observable, high-frequency percentage changes in trading volume series as the sum of an unobservable nonstationary series (the expected component) and stationary series (the unexpected component). We argue that the expected component is primarily driven by liquidity trades and the unexpected component is mainly driven by information-motivated trades.

The expected component is mainly driven by liquidity traders, for the following reasons. Firstly, consistent with the literature, liquidity-motivated traders trade randomly (see as examples Glosten & Milgrom 1985; Kyle 1985). In state space representation, the expected component is modelled as a random walk, and hence it is reasonable to argue that liquidity traders drive the expected component, since if the random walk holds, all available information would have been incorporated into stock prices. Secondly, market makers are considered as liquidity-motivated traders since they are responsible for the provision of liquidity in financial markets. Large, institutional traders, whose trades are typically motivated by liquidity requirements, are usually designated as market makers, with obligations to provide liquidity when there are liquidity constraints. Thus, some liquidity traders should be a permanent feature in the market. Furthermore, Menkveld (2013) shows that the profit of market makers comes from the bid-ask spread. Therefore, they need to trade consistently to obtain and increase their profits; it again indicates that some liquidity traders are permanent players in financial markets and suggests that any change in the structure of designed market makers will have a permanent impact on trading volume. According to the structure of the state space model, only the changes in the expected component affect the observable variable permanently, and therefore we can again argue that this component is driven by liquidity traders.

Similarly, information-motivated traders drive the unexpected component of trading volume, for the two reasons. Firstly, the information arrival process is an ‘unexpected’ process, and hence simple intuition suggests that information-motivated traders should be modelled as an unexpected component. Secondly, according to Chordia *et al.* (2002), private information should impact liquidity temporarily in financial markets.[9] Thus, any changes in the information-driven component of trading volume, while effecting a durable impact on price, should affect trading volume temporarily, and thus in state space models, the unexpected component has a transitory impact on the observable trading volume variable (see Hendershott & Menkveld 2014).

We model percentage changes in trading volume as a sum of a non-stationary expected (liquidity-driven) component and a stationary unexpected (information-driven) component. In its simplest form, the structure of the state space model for the percentage changes in trading volume can be expressed as:

$$v_{it} = m_{it} + s_{it} \tag{1}$$

and

$$m_{it} = m_{it-1} + u_{it} \tag{2}$$

where

$$v_{it} = \frac{TVolume_{it} - TVolume_{it-1}}{TVolume_{it-1}} \tag{3}$$

$TVolume_{it}$ is a trading volume of stock *i* at time *t*, $TVolume_{it-1}$ is a trading volume of stock *i* at time *t-1*, m_{it} is a non-stationary expected component of stock *i* at time *t*, s_{it} is a stationary

[9] Although information is a permanent component of stock prices (see Menkveld *et al.* 2007), it has a temporary impact on trading volume. The reason is that, according to the Efficient Market Hypothesis (EMH), any new information is simultaneously absorbed by traders and hence, it can only cause transitory (short-term) changes in trading volume (see Fama 1970).

unexpected component of stock *i* at time *t* and u_{it} is an idiosyncratic disturbance error. s_{it} and u_{it} are assumed to be mutually uncorrelated and normally distributed. Time, *t*, equals one-second in the main estimations; however, we also employ one-minute interval analysis for robustness.[10] The structure of the model shows that only changes on u_{it} affect the changes in trading volume permanently; s_{it} is temporary because it affects trading volume changes only at a particular time. By using maximum likelihood (likelihood is constructed using the Kalman filter), we can easily estimate $\sigma_{it}^{2_u}$ and $\sigma_{it}^{2_s}$. According to the structure of our state space model, any changes in the expected component of trading volume are sourced by changes in one fraction of the market, which is populated by liquidity traders; any changes in the other fraction of market, which is controlled by informed traders, should reflect the changes in the unexpected components. It implies that our estimations ($\sigma_{it}^{2_u}$ and $\sigma_{it}^{2_s}$) can be used as proxies for the two fractions of the market's trading volume, i.e. $\sigma_{it}^{2_u}$ is a proxy for liquidity-motivated traders and $\sigma_{it}^{2_s}$ is a proxy for information-motivated traders. To jointly test the empirical relevance of the state space model and the role of informed and liquidity traders in functionality and the efficiency of financial markets, we employ multivariate regressions as motivated in the next section.

3.4 A joint test of the empirical relevance of the state space model and the impact of different types of trading volume on price volatility, liquidity and market toxicity.

As already noted, we employ a state space approach to decompose trading volume into expected and unexpected components, and argue that the expected component is mainly driven by liquidity-motivated traders and the unexpected component is primarily driven by

[10] The results of the one-minute estimation results are presented in Appendix A and are qualitatively similar to the one-second interval estimations.

information-motivated traders. In order to jointly test the empirical relevance of the state space model and the role of liquidity and informed traders on the functioning and efficiency of financial markets, we employ predictive multivariate regressions.

Kyle (1985) develops a theoretical model deriving equilibrium security prices when traders' information sets are asymmetric. The model predicts that price volatility depends only on the informed trading volume and is independent of liquidity-based trading volume. In an associated work, Collin-Dufresne and Fos (2016) extend and generalize Kyle's (1985) model and show that the informed trading-induced price volatility depends on the aggressiveness of informed traders.

Thus, motivated by the predictions of the above-mentioned theoretical models, we jointly test the empirical relevance of the state space model and the roles of informed and liquidity traders in inducing price volatility by estimating the following regression:

$$\left|\Delta p_{it}\right| = \alpha + \beta_1 Espread_{it-1} + \beta_2 CTV_{it-1} + \beta_3 BSI_{it-1} + \beta_4 \sigma_{it-1}^{2_S} + \beta_5 \sigma_{it-1}^{2_U} + \varepsilon_{i,t} \tag{4}$$

where $\left|\Delta p_{i,t}\right|$ is the absolute value of price changes for stock *i* at time *t*, $Espread_{i,t-1}$ is the effective spread, measured as twice the absolute value of the difference between the last transaction price at time *t-1* minus the prevailing bid-ask spread at the transaction time, for stock *i* at time *t-1*, $CTV_{i,t-1}$ is the percentage changes in trading volume for stock *i* at time *t-1*, $BSI_{i,t-1}$ is the absolute difference between buyer- and seller-initiated traders for stock *i* at time *t-1*. $\sigma_{it-1}^{2_S}$ is the proxy for informed trading volume for stock *i* at time *t-1* and $\sigma_{it-1}^{2_U}$ is the proxy for liquidity trading volume for stock *i* at time *t-1*; both variables are obtained by maximum likelihood and from the state space estimation described in Section 3.3. The model is estimated at one-second intervals. Consistent with literature, we use absolute price changes to measure price volatility and employ effective spread for controlling liquidity. As mentioned, we use percentage changes in trading volume as the observable variable in the state space model. It implies that our proxies for informed and liquidity traders are mechanically correlated with percentage changes in trading volume. Chordia *et al.* (2002) argue that prices and liquidity in financial markets are strongly affected by the difference between buyer- and seller-initiated trades. Therefore, we use the absolute difference between buyer- and seller-initiated trades as the additional proxy to control for the effect of trading volume, in addition to percentage change in trading volume. $\sigma^{2_S}_{it-1}$ and $\sigma^{2_U}_{it-1}$ are the most important variables in the regression. If indeed our state space model correctly decomposes trading volume into liquidity and informed traders, we expect to see an insignificant relationship between $\sigma^{2_U}_{it-1}$ and price volatility after controlling for volume and liquidity, as Kyle (1985) argues that price volatility is not affected by liquidity traders. $\sigma^{2_S}_{it-1}$ on the other hand should be negatively and significantly correlated with price volatility, due to the absence of excessive aggressiveness in our sample period (see Collin-Dufresne & Fos 2016). As noted, we employ the absolute price changes as the dependent variable in the main regression, and for robustness, we also use the standard deviation of stock returns to measure price volatility (see as an example Lamoureux & Lastrapes 1990). Consistent with literature, we include the lagged value of the standard deviation of stock returns as an additional explanatory variable.

$$\sigma^{p}_{it} = \alpha + \beta_1\sigma^{p}_{it-1} + \beta_2 Espread_{it-1} + \beta_3 CTV_{it-1} + \beta_4 BSI_{it-1} + \beta_5\sigma^{2_S}_{it-1} + \beta_6\sigma^{2_U}_{it-1} + \varepsilon_{it} \tag{5}$$

Glosten and Milgrom's (1985) model is based on the idea that the extent of the adverse selection problem facing specialists when they trade with informed traders is one of the factors that the bid-ask spread is influenced by. The model predicts that the bid-ask spread is positively correlated with informed traders, however it is independent of the liquidity traders. This model is based on the assumption that traders adopt their trading strategies by using market orders, i.e. they trade aggressively. However, Kaniel and Liu (2006) modify the Glosten and Milgrom (1985) model and show that the informed traders with long-lived information tend to use limit orders rather than market orders (see also Menkveld 2013). It implies that by submitting limit orders, informed traders might improve liquidity. In addition, the theoretical model presented by Admati and Pfleiderer (1988) shows that informed traders who observe the same signal will compete against each other in exploiting the information signal, and this may lead to the market maker facing a smaller adverse selection problem. When faced with reduced adverse selection, market markets will respond with tighter spreads. Hence, motivated by the predictions of above-mentioned theoretical models, we jointly test the empirical relevance of the state space model and the role of informed and liquidity traders in liquidity by using the following regression:

$$Spread_{it} = \alpha + \beta_1 \sigma^{p}_{it-1} + \beta_2 CTV_{it-1} + \beta_3 BSI_{it-1} + \beta_4 \sigma^{2_S}_{it-1} + \beta_5 \sigma^{2_U}_{it-1} + \varepsilon_{it} \quad (6)$$

where $Spread_{i,t}$ corresponds to one of relative spread, quoted spread and effective bid-ask spread. Quoted spread is the difference between the last ask price minus the last bid price at time *t*, while the relative spread is the quoted spread divided by the last mid-point at time *t*. σ^{p}_{it-1} is the standard deviation of stock returns, $CTV_{i,t-1}$ is the percentage changes in trading volume, $BSI_{i,t-1}$ is the absolute difference between buyer- and seller-initiated traders, $\sigma^{2_S}_{it-1}$ is the proxy for informed traders, and $\sigma^{2_U}_{it-1}$ is the proxy for liquidity traders. We employ one-second frequency for the regression; *t* indexes the one-second interval. Additional explanatory variables, $CTV_{i,t-1}$, $BSI_{i,t-1}$ and σ^{p}_{it-1}, are included to control for trading volume and volatility. $\sigma^{2_S}_{it-1}$ and $\sigma^{2_U}_{it-1}$ are the key variables in our model. If indeed our state space model correctly decomposes trading volume into liquidity and informed traders, we expect to see no significant relationship between $\sigma^{2_U}_{it-1}$ and the various bid-ask spread metrics we use as dependent variables after controlling for volume, since Glosten and Milgrom (1985) argue that the bid-ask spread is not affected by the liquidity traders. By contrast, $\sigma^{2_S}_{it-1}$ should be significantly and negatively related with the bid-ask spread variables, because informed trading induces adverse selection, which is the major determinant of how wide the market maker spread is. The negative relationship between σ^{2s}_{it-1} and the spread is expected also because there is no evidence of excessive aggressiveness in our sample period (see Menkveld 2013; Collin-Dufresne & Fos 2015, 2016).

Finally, we investigate the role the informed trader plays in the creation of a toxic trading environment in the market. This is because the relationship between informed trading and market toxicity is a flipside question of the impact of informed traders on the functionality and efficiency of financial markets. In other words, questions about the role of informed traders in the inducement of market efficiency and the impact of informed traders on market toxicity are natural extensions of each other and one may not be fully explored without the other. Thus, we employ the following model to examine the relationship between market toxicity and informed trading:

$$MT_{it} = \alpha + \beta_1 Espread_{it-1} + \beta_2 CTV_{it-1} + \beta_3 BSI_{it-1} + \beta_4 \sigma^{2s}_{it-1} + \beta_5 \sigma^{2u}_{it-1} + \varepsilon_{it} \tag{7}$$

where MT_{it} is the proxy for market toxicity, $Espread_{i,t-1}$ is the effective spread, $CTV_{i,t-1}$ is the percentage changes in trading volume, $BSI_{i,t-1}$ is the absolute difference between buyer- and seller-initiated traders, σ^{2s}_{it-1} is the proxy for informed traders, and σ^{2u}_{it-1} is the proxy for liquidity traders, as computed from the state space model. We use the nominal order imbalance (OIB#) developed by Chordia *et al.* (2008), which captures buying and selling pressure, as proxy for order flow toxicity; the Lee and Ready (1991) algorithm is used to classify trading volume into buys and sells. Thus, MT_{it} is calculated as the absolute value of the difference between the numbers of buy and sell trades, divided by the total number of trades:

$$MT = \frac{|\#\,Buy\ trades - \#\,Sell\ trades|}{\#\,Buy\ trades + \#\,Sell\ trades} \tag{8}$$

Apart from these, we again employ some additional explanatory variables ($Espread_{i,t-1}$ and $CTV_{i,t-1}$ ($BSI_{i,t-1}$)) to control for trading volume and liquidity. In a departure from the other models already presented, we estimate this model only at the one-minute frequency. This is because it is difficult to obtain enough trading volume to compute MT_{it} within the one second period in an unbiased manner. According to Collin-Dufresne and Fos (2016) and Kaniel and Liu (2006), informed traders strategically choose to trade more when noise trading volume is high, and execute their trading strategies by submitting limit orders (passive orders) (see also Menkveld 2013), which leads to a negative relationship between informed trading volume and market toxicity during normal trading sessions (see also Admati & Pfleiderer 1988). Thus, we expect to see a negative correlation between σ^{2s}_{it-1} and market toxicity (see also Collin-Dufresne & Fos 2015).

INSERT TABLE 3 ABOUT HERE

Table 3 presents a correlation matrix with all the variables featured in the above-presented models. The low correlation coefficient estimates among the variables (except for the liquidity proxies, which is expected) suggest that we do not have multicollinearity issues with the regression models.

3.5 The predictability of short-horizon returns from unexpected (information-driven) components of trading volume

According to Fama (1970), (developed) financial markets are largely informationally efficient over a daily horizon. Chordia *et al.* (2008) argue that although markets are quite efficient over a long-horizon, there are inefficiencies in markets at shorter horizons because traders need time to act on new information. Motivated by this, Chordia *et al.* (2008) examine the predictability of short-term returns from past order imbalance and document that, indeed, markets are inefficient over short periods. The study employs order imbalance as an explanatory variable because it is argued that order imbalance signals private information, due to its capturing of buying and selling pressure. In their model, Chordia *et al.* (2008) show that short horizon returns predictability is smaller when markets are more liquid. We contend that the elimination of short horizon predictability is driven by the information-driven component of the order flow rather than increased order flow as a whole. Thus, we expect our estimated information-driven component of trading volume to be negatively correlated with short-horizon returns. This is because informed trading eliminates arbitrage opportunities, and by so doing engenders a market where short-horizon returns are minimal. In addition to eliminating short horizon return predictability, informed trading decreases price volatility as long as there is no case of excessive aggressiveness in financial markets. Therefore, the risk premium demanded by the traders should decrease with the volume of information-motivated traders in the market (see Wang 1993). This regression also serves as a further test of the empirical relevance of the state space modelling approach to estimating liquidity and informed trading components of trading volume. The estimated model is as follows:

$$R_{it} = \alpha + \beta_1 Espread_{it-1} + \beta_2 CTV_{it-1} + \beta_3 BSI_{it-1} + \beta_4 \sigma^{2s}_{it-1} + \varepsilon_{it} \tag{9}$$

where R_{it} is a midpoint return, $Espread_{i,t-1}$ is the effective spread, $CTV_{i,t-1}$ is the percentage changes in trading volume, $BSI_{i,t-1}$ is the absolute difference between buyer- and seller-initiated traders, and σ^{2s}_{it-1} is the proxy for informed traders. t indexes the one-second interval. $Espread_{i,t-1}$ and $CTV_{i,t-1}$ ($BSI_{i,t-1}$) are included for control of liquidity and trading volume, respectively. All variables are computed over a one-second frequency. σ^{2s}_{it-1} is the most important variable in this regression; we expect to see a significant and negative relationship between informed traders and future short-horizon return.

While we estimate the above regression over one-second intervals, it could be insightful to also do so over a lower frequency, such as the one-minute interval. The reason for this is that the trading volume in our sample appears to be mainly driven by HFTs, given the sample period and market we focus on (see Brogaard *et al.* 2014). Thus, if HFTs constitute the bulk of the informed trading volume, the predictability of return should be greatly diminished over a one-minute interval, since a one-minute interval cannot be considered a short-horizon for an HFT-driven market. Thus, we estimate the following regression at a one-minute frequency; the only difference to Equation (9) is the addition of MT_{it}, which we can only validly compute at a minimum frequency of one-minute:

$$R_{it} = \alpha + \beta_1 Espread_{it-1} + \beta_2 CTV_{it-1} + \beta_3 BSI_{it-1} + \beta_4 \sigma_{it-1}^{2s} + \beta_5 MT_{it-1} + \varepsilon_{it} \quad (10)$$

We expect that both MT_{it} and σ_{it-1}^{2s} should be insignificant at the one-minute interval because of the superfast trading systems of HFTs trading in S&P 500 stocks.

4. Results and discussion

4.1 State Space Estimates

Before presenting the results of the joint test of the empirical relevance of the state space model and the role of liquidity and informed traders on the functioning and efficiency of financial markets, we report the estimates of the general state space model as presented in Equations (1) – (3).

INSERT TABLE 4 ABOUT HERE

Table 4 presents the standard deviation estimates of the expected (liquidity-driven) and unexpected (information-driven) components of trading volume as decomposed using the state space model. As expected, the standard deviation of the unexpected component is higher than the standard deviation of the expected component. The estimates for the unexpected component's standard deviation in each quartile is higher than the corresponding estimates for the expected component. There are at still two reasons for this distribution in the estimates. Firstly, consistent with the structure of our state space approach, informed trades are more informative than the liquidity trades. Secondly, some liquidity traders (market makers) should trade consistently as they are under obligations to provide liquidity in the markets. By contrast, informed traders are not obligated to provide liquidity in the markets, and hence they are likely to trade only if they have an informational advantage over other traders. It implies a higher variance for informed traders and our results are consistent with this expectation.

Informed traders strategically trade more actively when trading volume and liquidity trading is high, as a higher trading volume provides a greater "camouflage" for informed trades. The estimates presented in Table 4 are consistent with this widely-held view in the market microstructure literature. For clarity, we divide our sample into quartiles according to their level of trading activity/activeness. The stocks in Quartile 1 are the least active ones, whereas Quartile 4 contains the most active stocks. The average daily trading volume estimates for Quartile 4 is 13.76 million, whereas the average daily trading volume estimates for Quartile 1 is 1.59 million for the least active stocks; thus, the trading volume estimates in Quartile 4 is about nine times higher than the trading volume for the typical stock in Quartile 1. Correspondingly, the average daily standard deviation of liquidity-motivated trades in Quartile 4 is about 9 times higher than the average daily standard deviation of liquidity traders in Quartile 1. This suggests that informed traders should be more active in Quartile 4; the estimates in the penultimate row of Table 4 are completely in line with this expectation; the average daily standard deviation of the unexpected component in Quartile 4 (14) is about 7 times larger than that of the unexpected component at 1.99 in Quartile 1.

4.2 Joint tests: the empirical relevance of the state space model and the impact of informed and liquidity trading volume on price volatility, liquidity and market toxicity.

In order to jointly test the relevance of the state space estimates obtained above and to investigate the impact of liquidity and informed trading volume components on several market quality proxies, we estimate the predictive regressions shown in Equations (4) – (7); the regression estimates for the first market quality proxy (volatility – Equations 4 and 5) are now presented in Table 5.

INSERT TABLE 5 ABOUT HERE

The estimates show that the lagged unexpected (information-driven) component of trading volume is a significant predictor of one-second absolute price changes. In contrast, the liquidity/expected component is not a significant predictor of one-second absolute price changes once we control for volume and liquidity. This is unsurprising since the latter component is liquidity driven and it is ‘expected’ in the sense that the trading activity generating it is based on information already incorporated into the price of the traded financial instruments. The results hold for both measures of price volatility that we employ, i.e. absolute price changes (presented in Panel A) and standard deviation of stock returns (Panel B), although the unexpected component coefficient is larger in Panel A. The negative coefficient indicates that increases in information-motivated trades reduces price volatility in financial markets. This result is consistent with the result of the empirical study of Avramov *et al.* (2006), who find that stock price volatility is negatively correlated with informed traders. The significant unexpected component and the insignificant expected component estimates imply a validation of the empirical relevance of our state space approach to decomposing trading volume into informed and liquidity-drive components. As predicted by Kyle's (1985) model, the informed trading volume captured by our state space approach is significantly related to price volatility, however, liquidity trading component is not. The estimated coefficients for all the other explanatory variables are consistent with the existing literature; trading volume and the effective spread are both positively and significantly correlated with price volatility (see Epps & Epps 1976; Glosten & Milgrom 1985). The explanatory power of the regression is small, however, with the R^2 being only about 0.33% for absolute price changes and 0.88% for standard deviation of stock returns. This is unsurprising and is due to our employment of a one-second frequency for the models' estimations (see Chordia *et al.* 2008).

The above-outlined results are consistent with the model presented by Collin-Dufresne and Fos (2016). The relationship between informed trading and price volatility is subject to two impacts. Firstly, informed traders’ activity in the market leads to the revelation of information and this new information reduces price uncertainty in financial markets. The reduction in price uncertainty in turn spurs a reduction in price volatility. Secondly, informed traders may trade aggressively in a liquidity-constrained environment and thereby increase aggressiveness in financial markets and this may increase price volatility. Thus, the relationship between informed traders and price volatility depends on the aggressiveness of informed traders. The relationship will be positive if informed traders use aggressive orders (market orders) and create excessive aggressiveness in the market. Interestingly, in related papers, Menkveld (2013) and Rzayev and Ibikunle (2017) show that aggressive orders are not profitable during normal trading periods, i.e. if there is no extreme volatility in financial markets, then the use of market orders offers no trading advantage to informed traders. The implication here is that informed traders seldom submit aggressive orders during normal trading days. Hence, as we do not observe any instance of excessive aggressiveness in our sample for the period we focus on, the negative impact of informed trading on stock price volatility reported in Table 5 is what we would expect to find (see also Wang 1993).

INSERT TABLE 6 ABOUT HERE

We now turn to the relationship between liquidity and the decomposed trading volume components, and estimate Equation (6) for this purpose. In Table 6 we present the model's estimates, and Panels A, B and C show the results with relative, quoted and effective spread measures as respective proxies for liquidity. The estimates show that, consistent with the predictions of Glosten and Milgrom's (1985) model predictions, the lagged unexpected component is a significant predictor of liquidity. The estimates for the lagged unexpected component of trading volume are negative and statistically significant at the 0.01 level irrespective of which liquidity proxy we employ. By contrast, the expected component is not significantly related with bid-ask spread after controlling for volume. The results in all of Table 6's panels indicate that the state space model we employ in this study appropriately decomposes trading volume into liquidity- and information-driven components. Consistent with the results in Table 5, our results show that the information-driven component is negatively (positively) correlated with the bid-ask spread (liquidity). Negative coefficients indicate that informed traders are more likely to consume liquidity in financial markets rather than provide it; in this case, they are liquidity consumers. The results are consistent with the findings of Collin-Dufresne and Fos (2015). The coefficients of all control variables are in line with the consistent literature. Similar to the price volatility model, R^2 values in Panels A, B and C are very small at only 0.51%, 0.23% and 0.15% respectively, because of the estimation frequency we use, i.e. the one-second frequency.

INSERT TABLE 7 ABOUT HERE

Finally, in this section, we examine the predictive regression estimates based on an investigation of the impact of liquidity and informed traders on market toxicity (as shown in Equation 7). Table 7 presents the estimated coefficients. Consistent with the results in Tables 5 and 6, the lagged unexpected component of trading volume is negatively and significantly correlated with market toxicity, however the expected component is not, after we control for volume and liquidity. The negative correlation suggests that information-motivated trading volume reduces order flow toxicity in financial markets even after controlling for the overall impact of trading volume and liquidity. At least two mechanisms could explain this observed effect. Firstly, theoretical models like Glosten and Milgrom (1985) assume that informed traders use aggressive orders (market orders) to execute their trading strategies, and hence they increase the bid-ask spread and induce adverse selection risk/market toxicity. However, Kaniel and Liu (2006) modify Glosten and Milgrom's (1985) model and show that informed traders with long-loved information tend to use limit orders rather than market orders during normal trading periods (see also Menkveld 2013). The prediction of Kaniel and Liu's (2006) model is empirically confirmed by Collin-Dufresne and Fos (2015). Thus, informed traders might use limit orders, which contributes to a reduction of the bid-ask spread by removing uncertainty in instruments' prices, as long as the trading period is not unnecessarily aggressive. In addition, the theoretical model presented by Admati and Pfleiderer (1988) shows that informed traders who observe the same signal will compete against each other in exploiting the information signal, and this may lead to the market maker facing a smaller adverse selection problem. When faced with reduced adverse selection, market makers will respond with tighter spreads.

Although all other control variables are significant in model, the explanatory power of the regression is small with the R^2 being only about 0.07%, again owing to the short horizon over which we estimate Equation (7).

4.3 The predictability of short-horizon return using the information-driven component of trading volume

In the previous section, we show that the unexpected (information-driven) component obtained from our state space model is significantly correlated with future price volatility, liquidity and market toxicity. The expected (liquidity-driven) component, on the other hand, is not significantly correlated with future volatility, liquidity and toxicity after controlling for volume, and in the case of volatility and toxicity, after controlling for liquidity. These results are consistent with the predictions of the Kyle (1985) and Glosten and Milgrom (1985) models and offer a strong support to our argument that the state space approach correctly decomposes trading volume into liquidity- and information-motivated trades. In other words, these results show that the unexpected component, as estimated, signals private information. Chordia *et al.* (2002) and Chordia *et al.* (2008) argue that short-horizon return can be predicted by order imbalance, as order imbalance can signal private information, and they show this empirically by estimating a series of short-horizon predictive regressions. Thus, if indeed the unexpected component signals private information, then stock returns should be predictable by the unexpected component as well. We employ one-second and one-minute frequency to empirically test the predictive power of the lagged unexpected component of trading volume for one-second and one-minute price returns respectively. As explained, we believe that although stock returns might be predictable within the one-second horizon in a market dominated by HFTs, such predictability dissipates over a longer horizon, such as one-minute, due to the ability of HFTs to eliminate arbitrage opportunities at much lower frequencies.

We first estimate the predictive regression Equation (9) at one-second intervals.

INSERT TABLE 8 ABOUT HERE

Table 8 presents the estimated coefficients for Equation (9). All of the coefficients, including the unexpected component variable, are statistically significant at the 0.01 level. Consistent with estimates from the previous section, the unexpected component estimate is also negative, suggesting that increasing levels of informed trading volumes eliminates returns/arbitrage. Thus, the unexpected component of trading volume as obtained using the state space model approach signals private information similar to the order imbalance metrics developed by Chordia *et al.* (2008). The adjusted R^2 is 0.02%, due to the frequency of the estimated model – one-second intervals.

INSERT TABLE 9 ABOUT HERE

We next estimate a similar regression model (Equation 10) over a longer time horizon of one-minute. As expected, the unexpected component is not significant after controlling for volume and liquidity, and the adjusted R^2 coefficient is 0.06% for this model. The lack of statistical significance for the unexpected component in the one-minute frequency regression model is due to the prevalence of HFT activity in the data we use, and the ability of HFTs to eliminate arbitrage opportunities at much lower frequencies. We also include the order imbalance metric used by Chordia *et al.* (2008) in the regression model and, in contrast to Chordia *et al.* (2008)'s results, the measure is not significant here. This shows that while one-second stock return is predictable from lagged metrics that signal private information, one-minute stock returns are not predictable in financial markets dominated by HFTs.

A key finding here is that although the lag of the unexpected component predicts one-second stock returns, one-minute stock returns are not predictable using either the unexpected component or order imbalance (as computed by Chordia *et al.* 2008). Thus, the latter part of the findings are not consistent with the results presented by Chordia *et al.* (2008) as Chordia *et al.* (2008) show that even five-minute stock returns can be predicted from past order imbalance. The inconsistency here is linked to the data period employed by both studies. While Chordia *et al.* (2008) employ a dataset covering 1993 to 2002, when HFTs are not the main drivers of trading in financial markets, we employ a much more recent dataset from 2016 to 2017. For example, based on an analysis of similar data, which is older than ours by a few years, Brogaard *et al.* (2014) show that at least fifty percent of New York's trading volume is driven by HFTs. It implies that the speed of price adjustment through the incorporation of new information has become much lower. Specifically, HFTs do not need a full minute to absorb and act on new information. Furthermore, Brogaard *et al.* (2014) show that HFTs are more active in large stocks. As our sample consists of the most active and largest stocks in U.S. financial markets, we expect that HFTs are the dominant traders in our sample period. Thus, the definition of short-horizon has shifted since the period investigated by Chordia *et al.* (2008); the one or five-minute (as in the case of Chordia *et al.* 2008) horizons cannot be considered as short-horizons for the purpose of predicting short-horizon returns. The negative relationship between the unexpected component and the one-second short-horizon return documented above is due to a decrease in the risk premium demanded by the traders when informed trading reduces volatility in the absence of excessive aggressiveness in the market.

5. Conclusion

In this paper, we develop a state space model for decomposing trading volume into liquidity-driven (expected) and information-driven (unexpected) components. We find evidence of the empirical relevance of our approach to estimating liquidity and information-driven components of trading volume. This paper is based on two central arguments related to the specification of the state space approach we use. Firstly, we argue that the expected component we obtain from the state space model is mainly driven by liquidity-seeking order flow, and secondly, that the unexpected component as motivated is primarily driven by information-motivated order flow. In addition to providing a robust set of arguments to back up our claims, we further develop a set of multivariate regression models to formally test these arguments. We find that the unexpected component obtained from the state space model is significantly correlated to lead volatility, liquidity and toxicity, even after controlling for volume (and in the case of volatility and toxicity, we also control for liquidity), whereas the expected component is not significantly related with them once volume and liquidity are controlled for. These results are consistent with the theoretical models presented in Kyle (1985) and Glosten and Milgrom (1985); the consistency therefore implies that the expected and unexpected components can be viewed as encapsulating the liquidity- and information-motivated trades in our sample, respectively. The findings can also be linked to informed traders not trading by using market (aggressive) orders during normal trading periods, when there are no upheavals or extreme liquidity constraints in the market, as predicted by Kaniel and Liu (2006) and Menkveld (2013).

Furthermore, we demonstrate that, similar to the order imbalance metrics developed by Chordia *et al.* (2008), the unexpected component we compute is also a significant predictor of short-horizon returns. This again shows that the unexpected component signals private information, which is due to its capturing information-motivated trading volume. The estimated and statistically significant negative relationship between the lag unexpected component of trading volume (informed trading) and one-second short-horizon return is linked to a reduction in the risk premium demanded by the traders, given that increased informed trading is linked with a reduction in price volatility during normal trading period, i.e. in the absence of excessive aggressiveness in trading. However, in contrast to Chordia *et al.* (2008), we find that one-minute return cannot be predicted using either the unexpected component metric or order imbalance, as employed by Chordia *et al.* (2008) for a five-minute return. This implies that in today's high frequency trading environment, arbitrage opportunities are eliminated at a much faster rate than in the early 2000s period examined by the latter study.

References

Admati, A.R., Pfleiderer, P., 1988. A Theory of Intraday Patterns: Volume and Price Variability. The Review of Financial Studies 1, 3-40

Avramov, D., Chordia, T., Goyal, A., 2006. The Impact of Trades on Daily Volatility. The Review of Financial Studies 19, 1241-1277

Bessembinder, H., Seguin, P.J., 1993. Price Volatility, Trading Volume, and Market Depth: Evidence from Futures Markets. Journal of Financial and Quantitative Analysis 28, 21-39

Brogaard, J., Hendershott, T., Riordan, R., 2014. High-Frequency Trading and Price Discovery. The Review of Financial Studies 27, 2267-2306

Chakrabarty, B., Moulton, P., Shkilko, A., 2015. Evaluating Trade Classification Algorithms: Bulk Volume Classification versus the Tick Rule and the Lee-Ready Algorithm. Journal of Financial Markets 25, 52-79

Chordia, T., Roll, R., Subrahmanyam, A., 2001. Market Liquidity and Trading Activity. Journal of Finance 56, 501-530

Chordia, T., Roll, R., Subrahmanyam, A., 2002. Order imbalance, liquidity, and market returns. Journal of Financial Economics 65, 111-130

Chordia, T., Roll, R., Subrahmanyam, A., 2005. Evidence on the speed of convergence to market efficiency. Journal of Financial Economics 76, 271-292

Chordia, T., Roll, R., Subrahmanyam, A., 2008. Liquidity and market efficieny. Journal of Financial Economics 87, 249-268

Collin-Dufresne, P., Fos, V., 2015. Do Prices Reveal the Presence of Informed Trading? The Journal of Finance 70, 1555-1582

Collin-Dufresne, P., Fos, V., 2016. Insider Trading, Stochastic Liquidity, and Equilibrium prices. Econometrica 84, 1441-1475

Copeland, T.E., 1976. A Model of Asset Trading Under the Assumption of Sequential Information Arrival. The Journal of Finance 31, 1149-1168

Copeland, T.E., 1977. A Probability Model of Asset Trading. Journal of Financial and Quantitative Analysis 12, 563-578

Daigler, R.T., Wiley, M.K., 1999. The Impact of Trader Type on the Futures Volatility-Volume Relation. The Journal of Finance 54, 2297-2316

Durbin, J., Koopman, S., 2012. Time Series Analysis by State Space Models. Oxford University Press, Oxford, UK.

Easley, D., De Prado, M., O'Hara, M., 2011. The microstructure of the "flash crash": flow toxicity, liquidity crashes, and the probability of informed trading. Journal of Portfolio Management 37, 118-129

Easley, D., De Prado, M., O'Hara, M., 2012. Flow Toxicity and Liquidity in a High-frequency World. The Review of Financial Studies 25, 1457-1493

Easley, D., Kiefer, N., O'Hara, M., Paperman, J., 1996. Liquidity, Information, and Infrequently Traded Stocks. The Journal of Finance 51, 1405-1436

Easley, D., Kiefer, N.M., O'Hara, M., 1997. One Day in the Life of a Very Common Stock. The Review of Financial Studies 10, 805-835

Easley, D., O'Hara, M., 1987. Price, trade size, and information in securities markets. Journal of Financial Economics 19, 69-90

Epps, T.W., Epps, M.L., 1976. The Stochastic Dependence of Security Price Changes and Transaction Volumes: Implications for the Mixture-of-Distributions Hypothesis. Econometrica 44, 305-321

Fama, E., 1970. Efficient capital markets: a review of theory and empirical work. The Journal of Finance 25, 383-417

Glosten, L., Milgrom, P., 1985. Bid, ask, and transaction prices in a specialist market with heterogeneously informed agents. Journal of Financial Economics 14, 71-100

Harris, M., Raviv, A., 1993. Differences of opinion make a horse race. The Review of Financial Studies 6, 473-506

Hellwig, M.F., 1980. On the Aggregation of Information in Competitive Markets. Journal of Economic Theory 22, 477-498

Hendershott, T., Menkveld, A.J., 2014. Price pressures. Journal of Financial Economics 114, 405-423

Ibikunle, G., 2015. Opening and closing price efficiency: Do financial markets need the call auction? Journal of International Financial Markets, Institutions & Money 34, 208-227

Kaniel, R., Liu, H., 2006. So What Orders Do Informed Traders Use? The Journal of Business 79, 1867-1913

Karpoff, J.M., 1987. The Relation Between Price Changes and Trading Volume: A Survey. Journal of Financial and Quantitative Analysis 22, 109-126

Kyle, A.S., 1985. Continuous Auctions and Insider Trading. Econometrica 53, 1315-1335

Lamoureux, C.G., Lastrapes, W.D., 1990. Heteroskedasticity in Stock Return Data: Volume versus GARCH effects. The Journal of Finance 45, 221-229

Lee, C., Ready, M., 1991. Inferring Trade Direction from Intraday Data. The Journal of Finance 46, 733-746

Menkveld, A.J., 2013. High frequency trading and the new market makers. Journal of Financial Markets 16, 712-741

Menkveld, A.J., Koopman, S.J., Lucas, A., 2007. Modeling Around-the-Clock Price Discovery for Cross-Listed Stocks Using State Space Methods. Journal of Business & Economic Statistics 25, 213-225

Rzayev, K., Ibikunle, G., 2017. Order aggressiveness and flash crashes.

Shalen, C.T., 1993. Volume, volatility, and the dispersion of beliefs. The Review of Financial Studies 6, 405-434

Sun, Y., Ibikunle, G., 2016. Informed trading and the price impact of block trades: A high frequency trading analysis. International Review of Financial Analysis

Van Ness, B.F., Van Ness, R.A., Yildiz, S., 2016. The role of HFTs in order flow toxicity and stock price variance, and predicting changes in HFTs' liquidity provisions. Journal of Economics and Finance 41, 739-762

Wang, J., 1993. A Model of Intertemporal Asset Prices Under Asymmetric Information. The Review of Economic Studies 60, 249-282

Table 1. Summary of trading activities

The table presents trading summary statistics for the most active 100 S&P 500 stocks from September 1, 2016 through to October 30, 2017. The Lee and Ready (1991) algorithm is used to classify trades as buyer- and seller-initiated.

Buyer-initiated (000,000s)	Seller-initiated (000,000s)	Total trades (000,000s)
106.89	109.48	216.37

Buyer-initiated (00,000,000s)	Seller-initiated (00,000,000s)	Total trading volume (00,000,000s)
347.71	375.61	723.32

Buyer-initiated	Seller-initiated	Average trade sizes
325.30	343.09	334.30

Buyer-initiated ($'0,000,000,000)	Seller-initiated ($'0,000,000,000)	Total USD volume ($'0,000,000,000)
156.70	171.66	328.36

Table 2. Summary statistics for variables

The table presents the descriptive statistics for variables of interest. *Espread* is the effective spread, computed as twice the absolute value of the difference between the last execution price for each interval and the midpoint of the prevailing bid and ask prices. *Rspread* is the relative spread, and is obtained by dividing the difference between the best ask and bid prices for each interval by the midpoint of both prices. *Qspread* is the quoted spread, and is simply the difference between the best ask and bid prices for each interval. *CTV* is the percentage change in trading volume, *BSI* is the absolute difference between buyer- and seller-initiated traders, $|\Delta p|$ is absolute value of price change, *R* is the one-second midpoint return, σ^p is the standard deviation of mid-price returns, and *MT* is the proxy for market toxicity, calculated as the absolute value of the difference between the numbers of buy and sell trades. One-second frequency is used for all variables, except *MT*. *MT* is computed by using one-minute frequency. The sample contains the most active 100 S&P 500 stocks traded between September 1, 2016 through to October 30, 2017 on NYSE and NASDAQ.

Variables	Mean	Median	Standard Deviation
Espread	0.00906	0.01000	0.04625
Rspread	0.00039	0.00028	0.00090
Qspread	0.01863	0.01000	0.05640
CTV	28.218	1.00	1026
BSI	1584.05	424.00	35771
$\lvert\Delta p\rvert$	0.00918	0.00900	0.06707
R	$-0.412\text{x}10^{-6}$	0.00	0.00139
σ^p	$0.92\text{x}10^{-4}$	$0.59\text{x}10^{-4}$	0.00091
MT	0.54067	0.50375	0.34194

Table 3. Correlation matrix for variables

The table plots the correlation matrix of the variables employed in this study's models. *Espread* is the effective spread, computed as twice the absolute value of the difference between the last execution price for each interval and the midpoint of the prevailing bid and ask prices. *Rspread* is the relative spread, and is obtained by dividing the difference between the best ask and bid prices for each interval by the midpoint of both prices. *Qspread* is the quoted spread, and is simply the difference between the best ask and bid prices for each interval. *CTV* is the percentage change in trading volume, *BSI* is the absolute difference between buyer- and seller-initiated traders, $|\Delta p|$ is absolute value of price change, σ^p is the standard deviation of mid-price returns, and σ^{2}_{S} and σ^{2}_{U} are the state space model-estimated proxies for informed and liquidity trading volumes respectively. The sample contains the most active 100 S&P 500 stocks traded between September 1, 2016 through to October 30, 2017 on NYSE and NASDAQ.

	Qspread	*Rspread*	*Espread*	*CTV*	*BSI*	σ^{2}_{S}	σ^{2}_{U}	$\lvert\Delta p\rvert$	σ^{p}
Qspread	1								
Rspread	0.80004	1							
Espread	0.91242	0.72760	1						
CTV	0.00059	0.00499	0.00126	1					
BSI	0.00031	0.00914	0.00392	0.12078	1				
σ^{2}_{S}	0.00010	0.00023	0.00050	0.21591	0.09836	1			
σ^{2}_{U}	0.00007	0.00002	0.00004	0.00000	0.00001	0.00000	1		
$\lvert\Delta p\rvert$	0.08367	0.05412	0.06475	0.00134	0.00698	0.00002	0.00008	1	
σ^{p}	0.13243	0.17431	0.12296	0.00140	0.00637	0.00004	0.00001	0.45337	1

Table 4. State Space Estimates

The table contains trading volume statistics and average daily standard deviation estimates of unexpected (information-driven) and expected (liquidity-driven) components of trading volume for the most active 100 S&P 500 stocks trading between September 1, 2016 through to October 30, 2017. Stocks are divided into quartiles according to their level of trading activity. Quartile 1 contains the least active companies, while Quartile 4 contains the most active stocks. The estimates are based on the following state space model for decomposing percentage change in trading volume:

$$v_{it} = m_{it} + s_{it}; \; m_{it} = m_{it-1} + u_{it}$$

where $v_{it} = \frac{TVolume_{it} - TVolume_{it-1}}{TVolume_{it-1}}$, $TVolume_{it}$ corresponds to trading volume of stock i at time t and $TVolume_{it-1}$ is trading volume of stock i at time t_{-1}, m_{it} is a non-stationary expected component of stock i at time t, s_{it} is a stationary unexpected component for stock i at time t and u_{it} is an idiosyncratic disturbance error. *TV* in the table below is the average daily trading volume, while $\sigma^s{}_{it}$ and $\sigma^u{}_{it}$ are the standard deviation estimates of the unexpected and expected components of trading volume respectively, estimated by maximum likelihood (constructed using the Kalman filter).

	Quartiles			
Variable	1	2	3	4
Trading volume ('000,000)	1.59	2	3.02	13.76
$\sigma^s{}_{it}$	1.99	3.34	5.25	14.00
$\sigma^u{}_{it}$	0.92	1.92	3.18	8.61

Table 5. Predictive power of lagged expected and unexpected components of trading volume on market volatility

The predictive power of one-second expected and unexpected components of trading volume is estimated using the following models:

$$\left|\Delta p_{it}\right| = \alpha + \beta_1 Espread_{it-1} + \beta_2 CTV_{it-1} + \beta_3 BSI_{it-1} + \beta_4 \sigma^{2_S}_{it-1} + \beta_5 \sigma^{2_U}_{it-1} + \varepsilon_{i,t}$$

$$\sigma^{p}_{it} = \alpha + \beta_1 \sigma^{p}_{it-1} + \beta_2 Espread_{it-1} + \beta_3 CTV_{it-1} + \beta_4 BSI_{it-1} + \beta_5 \sigma^{2_S}_{it-1} + \beta_6 \sigma^{2_U}_{it-1} + \varepsilon_{it}$$

where $|\Delta p|$ is the absolute value of price change, *Espread* is the effective spread, computed as twice the absolute value of the difference between the last execution price for each interval and the midpoint of the prevailing bid and ask prices. σ^{p}_{it-1} is the standard deviation of stock returns, *CTV* is the percentage changes in trading volume, *BSI* is the absolute difference between buyer- and seller-initiated traders, $\sigma^{2_S}_{it-1}$ and $\sigma^{2_U}_{it-1}$ are the state space model-based proxies (estimated using Kalman filter constructed maximum likelihood) for informed and uninformed trading. The sample contains the most active 100 S&P 500 stocks traded between September 1, 2016 through to October 30, 2017 on NYSE and NASDAQ. ***, ** and * correspond to statistical significance at the 0.01, 0.05 and 0.10 levels, respectively.

Panel A

Dependent Variable: $\left|\Delta p_{it}\right|$		
	Coefficient	*t*-Statistics
Intercept	0.844x10^{-2}***	707.48
$Espread_{i,t-1}$	0.724x10^{-1}***	287.01
$CTV_{i,t-1}$	0.123x10^{-6}***	18.95
$BSI_{i,t-1}$	0.543x10^{-7}***	157.79
$\sigma^{2_S}_{it-1}$	-0.422x10^{-13}***	-5.37
$\sigma^{2_U}_{it-1}$	0.195x10^{-44}	0.90
Adjusted R^2	0.0033	

Panel B

Dependent Variable: σ^{p}_{it}		
	Coefficient	*t*-Statistics
Intercept	0.754x10^{-4}***	464.92
σ^{p}_{it-1}	0.382x10^{-2}***	50.93
$Espread_{i,t-1}$	0.158x10^{-2}***	460.13
$CTV_{i,t-1}$	0.147x10^{-8}***	16.59
$BSI_{i,t-1}$	0.125x10^{-8}***	268.05
$\sigma^{2_S}_{it-1}$	-0.193x10^{-14}***	-18.07
$\sigma^{2_U}_{it-1}$	0.587x10^{-48}	0.02
Adjusted R^2	0.0088	

Table 6. Predictive power of lagged expected and unexpected components of trading volume on market liquidity

The predictive power of one-second expected and unexpected components of trading is estimated using the following model:

$$Spread_{it} = \alpha + \beta_1 \sigma^{p}_{it-1} + \beta_2 CTV_{it-1} + \beta_3 BSI_{it-1} + \beta_4 \sigma^{2_S}_{it-1} + \beta_5 \sigma^{2_U}_{it-1} + \varepsilon_{it}$$

where *Spread* corresponds to one of effective, quoted and relative spreads respectively. Effective spread is computed as twice the absolute value of the difference between the last execution price for each interval and the midpoint of the prevailing bid and ask prices. Relative spread is obtained by dividing the difference between the best ask and bid prices for each interval by the midpoint of both prices. Quoted spread is simply the difference between the best ask and bid prices for each interval. σ^{p}_{it-1} is the standard deviation of stock returns, *CTV* is the percentage change in trading volume, *BSI* is the absolute difference between buyer- and seller-initiated transactions, and $\sigma^{2_S}_{it-1}$ and $\sigma^{2_U}_{it-1}$ are the state space model-based proxies (estimated using Kalman filter constructed maximum likelihood) for informed and uninformed trading. The sample contains the most active 100 S&P 500 stocks traded between September 1, 2016 through to October 30, 2017 on NYSE and NASDAQ. ***, ** and * correspond to statistical significance at the 0.01, 0.05 and 0.10 levels, respectively.

Panel A

Dependent Variable: $RSpread_{i,t}$		
	Coefficient	*t*-Statistics
Intercept	0.390×10^{-3}***	2475.95
σ^{p}_{it-1}	0.194×10^{-1}***	262.16
$CTV_{i,t-1}$	0.661×10^{-9}***	7.56
$BSI_{i,t-1}$	0.144×10^{-8}***	311.17
$\sigma^{2_S}_{it-1}$	-0.152×10^{-14}***	-14.42
$\sigma^{2_U}_{it-1}$	-0.832×10^{-47}	-0.28
Adjusted R^2	0.0051	

Panel B

Dependent Variable: $QSpread_{i,t}$		
	Coefficient	*t*-Statistics
Intercept	0.185×10^{-1}***	1872.84
σ^{p}_{it-1}	0.936***	201.72
$CTV_{i,t-1}$	0.961×10^{-8}*	1.75
$BSI_{i,t-1}$	0.527×10^{-7}***	181.90
$\sigma^{2_S}_{it-1}$	-0.360×10^{-13}***	-5.45
$\sigma^{2_U}_{it-1}$	0.478×10^{-46}	0.03
Adjusted R^2	0.0023	

Panel C

Dependent Variable: $ESpread_{i,t}$

	Coefficient	*t*-Statistics
Intercept	$0.896\text{x}10^{-2***}$	1105.90
σ_{it-1}^{p}	0.626^{***}	164.50
$CTV_{i,t-1}$	$-0.305\text{x}10^{-8}$	-0.68
$BSI_{i,t-1}$	$0.349\text{x}10^{-7***}$	146.80
$\sigma_{it-1}^{2_S}$	$-0.354\text{x}10^{-13***}$	-6.51
$\sigma_{it-1}^{2_U}$	$-0.293\text{x}10^{-45}$	-0.19
Adjusted R^2	0.0015	

Table 7. Predictive power of lagged expected and unexpected components of trading volume on market toxicity

The predictive power of one-minute expected and unexpected components of trading volume is estimated using the following model:

$$MT_{it} = \alpha + \beta_1 Espread_{it-1} + \beta_2 CTV_{it-1} + \beta_3 BSI_{it-1} + \beta_4 \sigma^{2_S}_{it-1} + \beta_5 \sigma^{2_U}_{it-1} + \varepsilon_{it}$$

where *MT* is a proxy for market toxicity, which is computed as the absolute value of the difference between the numbers of buy and sell trades over a one-minute interval, divided by the total number of trades for that interval. *Espread* is the effective spread, computed as twice the absolute value of the difference between the last execution price for each one-minute interval and the midpoint of the prevailing bid and ask prices. *CTV* is the percentage change in trading volume, *BSI* is the absolute difference between buyer- and seller-initiated transactions, and $\sigma^{2_S}_{it-1}$ and $\sigma^{2_U}_{it-1}$ are the state space model-based proxies (estimated using Kalman filter constructed maximum likelihood) for informed and uninformed trading. The sample contains the most active 100 S&P 500 stocks traded between September 1, 2016 through to October 30, 2017 on NYSE and NASDAQ. ***, ** and * correspond to statistical significance at the 0.01, 0.05 and 0.10 levels, respectively.

Dependent Variable: MT_{it}		
	Coefficient	*t*-Statistics
Intercept	0.539***	4664.09
$Espread_{i,t-1}$	0.769x10^{-1}***	57.41
$CTV_{i,t-1}$	0.137x10^{-5}***	9.62
$BSI_{i,t-1}$	0.877x10^{-7}***	52.32
$\sigma^{2_S}_{it-1}$	-0.530x10^{-13}***	-3.82
$\sigma^{2_U}_{it-1}$	0.113x10^{-43}	1.02
Adjusted R^2	0.0007	

Table 8. Predictive power of lagged unexpected component of trading volume on one-second stock returns

The predictive power of one-second expected and unexpected components of trading volume is estimated using the following model:

$$R_{it} = \alpha + \beta_1 Espread_{it-1} + \beta_2 CTV_{it-1} + \beta_3 BSI_{it-1} + \beta_4 \sigma^{2s}_{it-1} + \varepsilon_{it}$$

where *R* is the midpoint one-second return, *Espread* is the relative spread, computed as twice the absolute value of the difference between the last execution price for each one-second interval and the midpoint of the prevailing bid and ask prices. *CTV* is the percentage changes in trading volume, *BSI* is the absolute difference between buyer- and seller-initiated traders, and σ^{2s}_{it-1} is the state space model-based proxy (estimated using Kalman filter constructed maximum likelihood) for informed trading. The sample contains the most active 100 S&P 500 stocks traded between September 1, 2016 through to October 30, 2017 on NYSE and NASDAQ. ***, ** and * correspond to statistical significance at the 0.01, 0.05 and 0.10 levels, respectively.

Dependent Variable: R_{it}		
	Coefficient	*t*-Statistics
Intercept	$-0.444\text{x}10^{-5***}$	-17.96
$Espread_{i,t-1}$	$0.390\text{x}10^{-3***}$	74.48
$CTV_{i,t-1}$	$0.129\text{x}10^{-8***}$	9.61
$BSI_{i,t-1}$	$0.316\text{x}10^{-9***}$	44.37
σ^{2s}_{it-1}	$-0.581\text{x}10^{-15***}$	-3.57
Adjusted R^2	0.0002	

Table 9. Predictive regressions of one-minute returns on lagged unexpected component

The predictive power of one-minute expected and unexpected components of trading volume is estimated using the following model:

$$R_{it} = \alpha + \beta_1 Espread_{it-1} + \beta_2 CTV_{it-1} + \beta_3 BSI_{it-1} + \beta_4 \sigma^{2s}_{it-1} + \beta_5 MT_{it-1} + \varepsilon_{it}$$

where R is the midpoint one-minute return, *Espread* is the relative spread, computed as twice the absolute value of the difference between the last execution price for each one-minute interval and the midpoint of the prevailing bid and ask prices. *CTV* is the percentage changes in trading volume, *BSI* is the absolute difference between buyer- and seller-initiated traders. *MT* is a proxy for market toxicity, which is computed as the absolute value of the difference between the numbers of buy and sell trades over a one-minute interval, divided by the total number of trades for that interval, and σ^{2s}_{it-1} is the state space model-based proxy (estimated using Kalman filter constructed maximum likelihood) for informed trading. The sample contains the most active 100 S&P 500 stocks traded between September 1, 2016 through to October 30, 2017 on NYSE and NASDAQ. ***, ** and * correspond to statistical significance at the 0.01, 0.05 and 0.10 levels, respectively.

Dependent Variable: R_{it}		
	Coefficient	t-Statistics
Intercept	-0.707x10^{-5***}	-3.02
$Espread_{i,t-1}$	0.867x10^{-3***}	59.39
$CTV_{i,t-1}$	0.677x10^{-8***}	4.36
$BSI_{i,t-1}$	0.781x10^{-9***}	42.70
σ^{2s}_{it-1}	0.223x10^{-15}	1.48
MT_{it}	-0.184x10^{-5}	-0.50
Adjusted R^2	0.0006	

Appendix A. Predictive power of lagged expected and unexpected components of trading volume on market volatility and liquidity II

The predictive power of one-minute expected and unexpected components of trading volume is estimated using the following models:

$$|\Delta p_{it}| = \alpha + \beta_1 Espread_{it-1} + \beta_2 CTV_{it-1} + \beta_3 BSI_{it-1} + \beta_4 \sigma^{2_S}_{it-1} + \beta_5 \sigma^{2_U}_{it-1} + \varepsilon_{i,t}$$

$$Rspread_{it} = \alpha + \beta_1 \sigma^{p}_{it-1} + \beta_2 CTV_{it-1} + \beta_3 BSI_{it-1} + \beta_4 \sigma^{2_S}_{it-1} + \beta_5 \sigma^{2_U}_{it-1} + \varepsilon_{it}$$

where $|\Delta p|$ is the absolute value of price change, *Espread* is the effective spread, computed as twice the absolute value of the difference between the last execution price for each interval and the midpoint of the prevailing bid and ask prices, *Rspread* is the relative spread and is obtained by dividing the difference between the best ask and bid prices for each interval by the midpoint of both prices σ^{p}_{it-1} is the standard deviation of stock returns, *CTV* is the percentage changes in trading volume, *BSI* is the absolute difference between buyer- and seller-initiated traders, and $\sigma^{2_S}_{it-1}$ and $\sigma^{2_U}_{it-1}$ are the state space model-based proxies (estimated using Kalman filter constructed maximum likelihood) for informed and uninformed trading. The sample contains the most active 100 S&P 500 stocks traded between September 1, 2016 through to October 30, 2017 on NYSE and NASDAQ. ***, ** and * correspond to statistical significance at the 0.01, 0.05 and 0.10 levels, respectively.

Panel A

Dependent Variable: $\lvert\Delta p_{i,t}\rvert$		
	Coefficient	*t*-Statistics
$Intercept$	0.200x10^{-1***}	388.04
$Espread_{i,t-1}$	0.460x10^{-1***}	76.75
$CTV_{i,t-1}$	0.489x10^{-6***}	7.69
$BSI_{i,t-1}$	0.775x10^{-7***}	103.69
$\sigma^{2_S}_{it-1}$	-0.132x10^{-13**}	-2.13
$\sigma^{2_U}_{it-1}$	-0.653x10^{-45}	-0.13
Adjusted R^2	0.0019	

Panel B

Dependent Variable: $RSpread_{i,t}$		
	Coefficient	*t*-Statistics
$Intercept$	0.440x10^{-3***}	824.36
σ^{p}_{it-1}	0.203x10^{-3***}	58.80
$CTV_{i,t-1}$	0.251x10^{-8***}	3.82
$BSI_{i,t-1}$	0.134x10^{-8***}	174.63
$\sigma^{2_S}_{it-1}$	-0.522x10^{-15***}	-8.16
$\sigma^{2_U}_{it-1}$	0.277x10^{-47}	0.05
Adjusted R^2	0.0038	

APPENDIX B. List of the sample stocks

ISIN CODE	RIC	SECURITY NAME
US02376R1023	AAL.OQ	American Airlines Group Inc.
US0378331005	AAPL.OQ	Apple Inc.
US00287Y1091	ABBV.N	AbbVie Inc.
US0028241000	ABT.N	Abbott Laboratories
US00130H1059	AES.N	AES Corp.
US0268747849	AIG.N	American International Group Inc.
US0382221051	AMAT.OQ	Applied Materials Inc.
US0079031078	AMD.OQ	Advanced Micro Devices Inc.
US0325111070	APC.N	Anadarko Petroleum Corp.
US00507V1098	ATVI.OQ	Activision Blizzard Inc.
US0605051046	BAC.N	Bank of America Corp.
US1101221083	BMY.N	Bristol-Myers Squibb Co.
US1011371077	BSX.N	Boston Scientific Corp.
US1729674242	C.N	Citigroup Inc.
US1746101054	CFG.N	Citizens Financial Group Inc.
US1651671075	CHK.N	Chesapeake Energy Corp.
US20030N1019	CMCSA.OQ	Comcast Corp.
US1270971039	COG.N	Cabot Oil & Gas Corp.
US20825C1045	COP.N	ConocoPhillips
US2220702037	COTY.N	Coty Inc.
US17275R1023	CSCO.OQ	Cisco Systems Inc.
US1264081035	CSX.OQ	CSX Corp.
US1567001060	CTL.N	CenturyLink Inc.
US1266501006	CVS.N	CVS Health Corp.
US1667641005	CVX.N	Chevron Corp.
US2473617023	DAL.N	Delta Air Lines Inc.
US2546871060	DIS.N	Walt Disney Co.
US2786421030	EBAY.OQ	eBay Inc.
US2944291051	EFX.N	Equifax Inc.
US3453708600	F.N	Ford Motor Co.
US30303M1027	FB.OQ	Facebook Inc.
US35671D8570	FCX.N	Freeport-McMoRan Inc.
US3167731005	FITB.OQ	Fifth Third Bancorp
US90130A2006	FOXA.OQ	Twenty-First Century Fox Inc.
US3696041033	GE.N	General Electric Co.
US3700231034	GGP.N	General Growth Properties Inc.
US3755581036	GILD.OQ	Gilead Sciences Inc.
US2193501051	GLW.N	Corning Inc.
US37045V1008	GM.N	General Motors Co.

US3647601083	GPS.N	Gap Inc.
US4062161017	HAL.N	Halliburton Co.
US4461501045	HBAN.OQ	Huntington Bancshares Inc.
US4103451021	HBI.N	HanesBrands Inc.
US42824C1099	HPE.N	Hewlett Packard Enterprise Co.
US4282361033	HPQ.N	Hewlett-Packard Co.
US44107P1049	HST.N	Host Hotels & Resorts Inc.
US4581401001	INTC.OQ	Intel Corp.
US4783661071	JCI.N	Johnson Controls
US4781601046	JNJ.N	Johnson & Johnson International Plc.
US46625H1005	JPM.N	JPMorgan Chase & Co.
US4932671088	KEY.N	KeyCorp
US49456B1017	KMI.N	Kinder Morgan Inc.
US1912161007	KO.N	The Coca Cola Co.
US5010441013	KR.N	Kroger Co.
US5486611073	LOW.N	Lowe's Companies Inc.
US8447411088	LUV.N	Southwest Airlines Co.
US55616P1049	M.N	Macy's Inc.
US5770811025	MAT.OQ	Mattel Inc.
US6092071058	MDLZ.OQ	Mondelez International Inc.
IE00BTN1Y115	MDT.N	Medtronic Plc.
US59156R1086	MET.N	MetLife Inc.
US5529531015	MGM.N	MGM Resorts International
US02209S1033	MO.N	Altria Group Inc.
US58933Y1055	MRK.N	Merck & Co Inc.
US5658491064	MRO.N	Marathon Oil Corp.
US6174464486	MS.N	Morgan Stanley
US5949181045	MSFT.OQ	Microsoft Corp.
US5951121038	MU.OQ	Micron Technology Inc.
NL0011031208	MYL.OQ	Mylan NV Inc.
US6516391066	NEM.N	Newmont Mining Corp.
US64110L1061	NFLX.OQ	Netflix Inc.
US6541061031	NKE.N	Nike Inc.
US6293775085	NRG.N	NRG Energy Inc.
US67066G1040	NVDA.OQ	NVIDIA Corp.
US68389X1054	ORCL.N	Oracle Corp.
US7170811035	PFE.N	Pfizer Inc.
US7427181091	PG.N	Procter & Gamble Co.
US70450Y1038	PYPL.OQ	PayPal Holdings Inc.
US7475251036	QCOM.OQ	Qualcomm Inc.
US7591EP1005	RF.N	Regions Financial Corp.

US75281A1097	RRC.N	Range Resources Corp.
US8552441094	SBUX.OQ	Starbucks Corp.
US8085131055	SCHW.N	Charles Schwab Corp.
AN8068571086	SLB.N	Schlumberger NV
IE00B58JVZ52	STX.OQ	Seagate Technology Plc.
US87165B1035	SYF.N	Synchrony Financial
US8715031089	SYMC.OQ	Symantec Corp.
US00206R1023	T.N	AT&T Inc.
US87612E1064	TGT.N	Target Corp.
US8825081040	TXN.OQ	Texas Instruments Inc.
US9043111072	UAA.N	Under Armour Inc.
US9029733048	USB.N	U.S. Bancorp
US92826C8394	V.N	Visa Inc.
US92553P2011	VIAB.OQ	Viacom Inc.
US92343V1044	VZ.N	Verizon Communications Inc.
US9314271084	WBA.OQ	Walgreens Boots Alliance Inc.
US9497461015	WFC.N	Wells Fargo & Co.
US9694571004	WMB.N	Williams Companies Inc.
US9311421039	WMT.N	Wal-Mart Stores Inc.
US30231G1022	XOM.N	Exxon Mobil Corp.