Heterogeneous VAR Dynamics of Limit Order Book Depth, Trade Imbalance, and Volatility on the NYSE

Eldar Nigmatullin

Konstantin Tyurin

Hao Yin*

October 4, 2007

*Corresponding author. We thank Reuters for providing the data used in this paper. We thank Mark Hooker for helpful comments. We are responsible for any errors. Eldar Nigmatullin is from the Advanced Research Center, State Street Global Advisors, Email: eldar_nigmatullin@ssga.com, address: Advanced Research Center, State Street Global Advisors, One Lincoln Street, Boston, MA 02111. Konstantin Tyurin is from the Department of Economics, Indiana University, Email: ktyurin@indiana.edu, address: Department of Economics, Indiana University, 100 S. Woodlawn, Wylie Hall 105, Bloomington, IN 47405. Hao Yin is from the Department of Economics, Indiana University, Email: haoyin@indiana.edu, address: Department of Economics, Indiana University, 100 S. Woodlawn, Wylie Hall 105, Bloomington, IN 47405. The research reported here was completed when Hao Yin was at the Advanced Research Center, State Street Global Advisors. The opinions expressed in this paper do not necessarily reflect those of the employees, members, or directors of State Street Global Advisors.

Heterogeneous VAR Dynamics of Limit Order Book Depth, Trade Imbalance, and Volatility on the NYSE

Abstract

This paper uses a recent sample of data on DJIA stocks traded on the NYSE to study relationships between the state of the limit order book and high-frequency returns, volatility, and other market variables in the framework of heterogeneous vector autoregressions. It is shown that most of the information contained in the limit order book can be captured by four interpretable common factors related to the book's "shape". We demonstrate that these common factors have significant interactions with price volatility and instantaneous returns, even after controlling for total trade volume and signed order flow. However, the estimated coefficients across stocks exhibit considerable heterogeneity, which should be taken into account for proper inference. The impulse-response analysis reveals significant feedback between the shape of the limit order book and transaction volume, bid-ask spread, return, and short-run volatility.

JEL Classification Codes: G1, C33, C51

Keywords: limit order book, common factors, principal components, heterogeneous panel vector autoregression

1. Introduction

Historically, analysis of the relationship between the state of a limit order book and prevailing market conditions primarily focused on one way effects. Recently there appeared several studies that emphasized the two-way dynamic interactions (e.g., Hall and Hautsch, 2007). In this paper we contribute to this growing literature by analyzing the relevance of limit order book (LOB) information for explaining the price fluctuations and order flow dynamics at intraday (minute-by-minute) frequencies. We distinguish between two pieces of limit order book information: the depth level, associated with overall liquidity supplied by limit order traders across the broad spectrum of prices on the limit order book, and the LOB rotation, associated with liquidity provided by impatient limit order traders in a close vicinity of the inside spread. We report that the limit order book helps explain the upcoming demand for liquidity and thus the price movement in near future, although different features of the book play different roles. At the same time, our results suggest that limit order traders monitor the market situation closely and determine their liquidity supply and aggressiveness accordingly.

More specifically, we find that the overall liquidity supply in the limit order book has at least as much predictive power to the price volatility as the overall liquidity demand (measured by trading volume) in the reduced form vector autoregressive model (VAR). In addition, the imbalance in liquidity supply between the bid and ask sides of the book provides at least as much explanatory power to the instantaneous price returns as the imbalance in the liquidity demand (measured by signed volume, or the imbalance between transaction volumes initiated by buyers and sellers). This conclusion is hardly surprising if we recall the fundamental roles of both supply and demand in equilibrium price determination, but the present paper demonstrates this property more systematically in a limit order market environment.

Furthermore, we apply a structural VAR model of liquidity provision and consumption to study responses of the limit order book and market variables to unobservable shocks. We show that generally both the response of the limit order book to market shocks and the response of market conditions to the limit order book shocks

behave as predicted by the classical microstructure models. In this structural VAR framework, the limit order book information still matters for price and volatility dynamics in the short run, although the magnitude of this effect is found to be relatively mild.

The growing literature on limit order markets can be broadly categorized into two partially overlapping strands. One strand focuses on the order flow component, aiming to infer limit order traders' submission and cancellation strategies from disaggregated and partially disaggregated order level data. The recent studies of order submission and cancellation strategies include, among others, Ellul *et al.* (2007), Bessembinder *et al.* (2006), Chakrabarty *et al.* (2007) and Hall and Hautsch (2006). The second strand focuses on the informational role of a limit order book, and the interplay of limit order book information with other market characteristics. Among the central issues addressed by the second strand are whether the limit order book beyond the best quotes provides incremental information at all, how different parts of the limit order book differ in their information content, how the properties of the limit order book interact with market conditions. One of the earliest studies of this sort, by Biais *et al.* (1995), concludes that the limit order book information can facilitate prediction of order flow in the Paris Bourse. Cao *et al.* (2004) examine liquidity provided at different price levels of the limit order book at the Australian Stock Exchange and argue that parts of the book beyond the best quotes are also informative, showing, in particular, that the imbalance between the levels of liquidity provided on bid and ask sides of the book helps explain the future price movements. Kavajecz and Odders-White (2004) demonstrate that some of the technical analysis rules used by practitioners to predict price trends can stem from the incremental information transmitted by the shape of a limit order book. In addition to price returns, some studies found a close relationship between price volatility and shape of the limit order book. For example, both Ahn *et al.* (2001) and Coppejans *et al.* (2001) find a robust inverse relationship between the limit order book depth and the high-frequency price volatility. In addition, studies conducted on daily data (for instance, Nas and Skjeltorp, 2006) provide evidence that the slope of a limit order book is negatively related to the price volatility measured at daily frequency.

Our paper extends the existing literature in several ways. First, it integrates several key variables including order flow, spread, rate of return, and volatility into a unified model, and investigates the dynamic properties of their "two-way" interactions with the limit order book. In our paper, a large number of market variables and limit order book characteristics are studied simultaneously, which is important if we want to isolate the information share of the limit order book from other covariates and control potential endogeneity among the variables involved. Chan (2005), which is closely related to our paper methodologically, examines the price movement impact on the state of limit order book at the Hong Kong Stock Exchange. Our paper takes a different perspective by focusing on joint dynamics between endogenous market variables and the limit order book shape.

Second, most of the existing empirical work on the information content of a limit order book emphasizes certain characteristics of the book. In our paper, we take an agnostic approach, attempting to capture parsimoniously the most informative features of the book without making the choice of these features up front. More specifically, we rely on principal component analysis to extract the most important common factors across a broad spectrum of limit order book price levels. In that respect, our approach is similar to the one undertaken by Beltran *et al*. (2006), which employs principal component analysis to uncover common factors driving the price impact dynamics implied by the limit order book. The present paper focuses on common factors based on the limit order book depth and studies interaction of those factors with market conditions.

Furthermore, although the availability of order level data to researchers has increased dramatically in the last decade, most of the empirical papers studying the limit order book use data that is more than five years old from now. For example, Ronaldo (2004) uses the data of 1997 from the Swiss Stock Exchange; Beltran *et al*. (2006) use data of 2004 from the German Stock Exchange, Pascual and Veredas (2006a and 2006b) use data of 2000 from the Spanish Stock Exchange, Hall and Hautsch (2007) use the data of 2002 from the Australian Stock Exchange, and Large (2007) use the data of 2002 from the London Stock Exchange. As of the U.S. equity markets, Bacidore et al. (2003) use data of 2000 from NYSE, Bae *et al*. (2003) and Harris and Panchapagesan (2005) use the TORQ data of 1990-1991, and Ellul et al. (2007) use data of 2001 from NYSE. Relying on the 2006

DJIA stock data from NYSE, the present paper analyzes the dynamics in the prevailing limit order market. It would be helpful to compare our results with the previous findings, considering the fast-changing environment of market microstructure, such as the decimalization and initiation of hybrid market on NYSE.

The remainder of this paper is organized as follows. Section 2 formulates empirical hypotheses based on the theoretical microstructure literature. Section 3 briefly reviews the econometric methods used throughout the paper. Section 4 describes the data and the definitions of variables. The main results are discussed in section 5. Section 6 concludes and highlights the directions for future research.

2. Empirical Hypotheses

At any moment of time, the electronic screen of a prototypical limit order book displays outstanding buy and sell orders up to a certain number of ticks away from the current best bid and ask prices. It is common to summarize the limit order data by reporting the cumulative depth at each limit price level on the bid and ask sides of the book, and visualize cumulative depth by plotting the limit order book as shown in Figure 1. Such snapshots evolving in real time can be interpreted as instantaneous "ersatz" demand and supply schedules for the traded financial instrument.

It is easy to see from Figure 1 that at least four ingredients are necessary to "anchor" a limit order book snapshot on the price-quantity diagram. The first ingredient is the mid-quote, which pins down the relative position of the snapshot across the price axis; the shifts of the mid-quote level characterize the evolution of an equilibrium price level for the financial instrument over time. The second ingredient is the inside spread, which determines how far apart from each other are the best prices of outstanding buy and sell limit orders. The third characteristic is the level of cumulative depth schedules derived from the outstanding buy and sell limit orders, which can be used to characterize the average amounts of liquidity provided by the limit order book within a specified range of bid and ask prices. Finally, since the demand and supply schedules tend to "curve" around pivotal points near the origin of those curves, we might want to introduce some "rotation" measure to characterize the extent of "tilting" experienced by the bid and ask

sides of the book. The level and rotation ingredients may provide a parsimonious summary of the information contained in the shape of a limit order book. We will apply principal component analysis to the cumulative depth on the bid and ask sides of the book to formally extract the "level" and "rotation" factors that would explain most of cross-sectional variation of the limit order book shape.

Among the four characteristics that could be used to summarize the limit order book information at any time, the empirical microstructure literature has focused primarily on the dynamics of the mid-quote and the inside spread (see, for example, Roll, 1984, and McInish and Wood, 1992). The early work on depth focused primarily on the depth provided at the best bid and ask price levels, since this is the only data made available by major exchanges (NYSE TAQ database in the equity market, EBS and Reuters limit order books in the FX market, etc.) The work on the role played by limit order book depth at the deeper price levels has peaked in popularity only recently, as richer limit order book data sets become widely available to researchers. How does the shape of the limit order book evolve over time? How do the key characteristics of LOB shape interact with other determinants of the limit order book and transactions? Do the limit order book depth level and rotation dynamics provide incremental information about the future stock price dynamics? These are just a few of the questions that have been asked in the recent limit order book literature and will be re-examined in the present paper.

The importance of limit order book information has been emphasized by various theoretical models, which posit testable predictions about the origin and character of interactions between the shape of the limit order book, the limit and market order submission strategies, and the prevailing market conditions.

2.1. Depth Average:

As a measure of overall level of the outstanding liquidity provided by limit order traders, the depth is expected to shrink when the potential risks of posting a limit order increase. One such risk faced by a limit order trader is the "picking-off" risk, which goes up at times of more severe informational asymmetries among the traders. Almost all adverse selection models (Kyle, 1985, Glosten and Milgrom, 1985, Handa, Schwartz, and Tiwari, 2003, among many others) predict the wider inside spreads at the times of

adverse selection. In addition to wider spreads, liquidity suppliers may also cancel the existing limit orders and refrain from submitting new limit orders. In other words, the following hypotheses can be posited about the relationship between the depth average and the inside bid-ask spread:

Depth and Spread*: An increase (decrease) of the limit order book depth leads to a narrower (wider) inside spread in the short run.*

$$Depth_{t-1} \uparrow \Rightarrow Spread_t \downarrow$$

Spread and Depth*: A wider (tighter) inside spread leads to a smaller (larger) depth level in the short run.*

$$Spread_{t-1} \uparrow \Rightarrow Depth_t \downarrow$$

In the information-based model of Easley and O'Hara (1987), the presence of asymmetric information is reflected in an unexpectedly large transaction volume, driven primarily by informed traders, which may discourage the limit order activity by liquidity traders. Accordingly, we can formulate the following hypothesis:

Trade Volume and Depth*: An increase (decrease) of trade volume decreases (increases) the depth of the limit order book in the short run.*

$$TVolume_{(t-1;t]} \uparrow \Rightarrow Depth_t \downarrow$$

While trade volume can have a negative impact on the depth levels of the limit order book, the sign will switch if we approach this problem from the opposite direction. A higher level of liquidity supplied by limit order traders may attract more trades to the market. This leads to the following hypothesis:

Depth and Trade Volume*: An increase (decrease) of the limit order book depth tends to increase (decrease) the trade volume in the short run.*

$$Depth_{t-1} \uparrow \Rightarrow TVolume_{(t-1;t]} \uparrow$$

Copeland and Galai (1983) argue that traders submitting bid (ask) limit orders *de facto* offer free put (call) options with strike prices equal to the limit prices. As the short-term volatility of the future stock price goes up, so does the intrinsic value of these free options, which have the higher probability to be in-the-money in a volatile market, since limit order traders have a higher chance of being "picked off" by informed traders when the uncertainty about the future stock price gets higher. And the higher "picking-off" risk will make liquidity suppliers refrain from submitting limit orders. This predicted

relationship between price volatility and limit order book depth is summarized by the following hypothesis:

Volatility and Depth*: An increase (decrease) of future price volatility decreases (increases) the depth of the limit order book in the short run.*

$$Volatility_{(t-1;t]} \uparrow \Rightarrow Depth_t \downarrow$$

On the other hand, greater limit order book depth may reflect liquidity supply that would absorb the liquidity demand from the incoming market orders. As a result, such market orders would be less likely to "walk the book", and we would expect smaller levels of price volatility, as summarized in the following hypothesis:

Depth and Volatility*: An increase (decrease) of limit order book depth reduces (increases) the future price volatility in the short run.*

$$Depth_{t-1} \uparrow \Rightarrow Volatility_{(t-1;t]} \downarrow$$

2.2. Depth Imbalance:

A limit order book which is unbalanced with respect to its depth on the bid and ask sides can originate from a liquidity supply shock experienced by either of its sides, such as, for instance, unexpectedly abundant or scarce limit order submissions. Since traders tend to condition their order submission strategies, which cause future market movement, on the current state of the limit order book, we can bridge the depth imbalance of the book to subsequent shift in the market price. As an example, in the open limit order book market modeled by Parlour (1998), the sequentially arriving traders choose between submitting limit order or market orders depending on the endogenous execution probability given the current book status and the predicted future order flow. One implication of this model is the negative correlation between consecutive limit order submissions. If a buyer observes a "thick" book on the bid side, she will increase her chance of execution by submitting market buy order instead of submitting limit buy order. The market order will consume depth at the best ask price and move up the price (measured by mid-quote or the transaction price). On the other hand, Biais *et al*. (1995), Chakravarty and Holden (1995), and Yeo (2006) propose models suggesting a "follow-on" strategy executed by limit order traders who tend to undercut each other either because of information asymmetry or the pure competition for execution. All of these

models point to a same prediction of the dynamics between depth imbalance of the book and the price movement, formulated as follows:

Depth Imbalance and Return*: An increase (decrease) of the limit order book depth imbalance (measured by the difference between bid and ask levels of depth) increases (decreases) the instantaneous price return in the next moment.*

$$BidDepth_{t-1} - AskDepth_{t-1} \uparrow \Rightarrow Return_{(t-1;t]} \uparrow$$

Although all of the above theoretical models predict the same pattern between depth imbalance and instantaneous return, the implications are different with regard to the sign of serial correlation between limit orders. In Parlour (1998), a large "crowd" of sequential traders hit either side of the book with market orders. At the same time, Chakravarty and Holden (1995) and Yeo (2006) argue that the informed sequential traders may be submitting the limit rather than market orders, leading to improvements in bid and ask limit prices. These two classes of models have different implications to the relationship between depth imbalance and signed volume. Although the "follow-on" models do not make clear prediction about the trading volume, the sequential market order traders observing an unusually large depth on the bid side would be submitting market buy orders in Parlour (1998) model, leading to an imbalance between the buyer- and seller-initiated trade volumes, at least in the short run. Thus we have the following hypothesis to test:

Depth Imbalance and Signed Volume*: An increase (decrease) of the depth imbalance of the limit order book implies an increase (decrease) of the signed volume in the short run.*

$$BidDepth_{t-1} - AskDepth_{t-1} \uparrow \Rightarrow BuyVolume_{(t-1;t]} - SellVolume_{(t-1;t]} \uparrow$$

2.3. Rotation Average:

While the limit order book depth is traditionally used to characterize the level of liquidity provided by limit order traders, the limit order book rotation, or "tilting", can be used for inference about the aggressiveness of limit order traders on the bid and ask sides of the market. By aggressiveness, we mean the limit order price is close to the best quote as in the scheme proposed by Biais et al (1995), and an alternative term could be "urgency". *Other things being equal*, we would expect more limit orders clustering close

to the best quotes when we observe the limit order book curve rotating around the average depth as illustrated by the arrows in Figure 1. In the extreme case where all limit orders locate at the best quote, the limit order book shape would be completely flat.[1] Furthermore, Handa and Schwartz (1996) demonstrate that a limit order market is basically driven by two types of traders, patient and urgent. Within their framework, the depth average would measure the overall size of both patient and urgent traders and the rotation average the relative proportion of urgent traders.

Limit orders are subject to two main risks: the risk of non-execution and the risk of being "picked-off". Higher aggressiveness enhances the execution probability of the limit order book, but at the cost of higher probability of being "picked-off". So following the logic of Kyle (1985), Easley and O'Hara (1992) and Foucault (1999), we would expect the rotation average interacts in the same way as the depth average with volume, spread and volatility. To conserve space, we do not list the hypotheses here.

It would be interesting to see the effect of overall depth on overall aggressiveness of the market. According to the "crowding-out" hypothesis, thick book induces the traders on the same side to be more aggressive to improve the execution opportunity. Thus we have the following hypothesis:

Depth average and Rotation average: *An increase (decrease) of the depth average of the limit order book implies an increase (decrease) of the rotation average in the next moment.*

$$Depth_{t-1} \uparrow \Rightarrow Rotation_t \uparrow$$

2.4. Rotation Difference:

Rotation difference is a proxy of the imbalance of aggressiveness of traders on the two sides of the market, which we would expect to associate closely with the market sentiment. According to the "information effect", one of the hypotheses can be formulated as following:

[1] However, we should notice that this rotation factor can only measure the relative aggressiveness of "at-the-quote" or "behind-the-quote" limit orders. If traders submit market orders or "inside-the-quote" limit orders, the position of mid quote and the spread will be different. This is why we emphasize "other things being equal".

Return and Rotation difference: *A higher (lower) price return increases (decreases) the rotation difference of the limit order book in the next moment.*

$$Return_{(t-1;t]} \uparrow \Rightarrow BidRotation_t - AskRotation_t \uparrow$$

Rotation difference and Return*: An increase (decrease) of the rotation difference of the limit order book increases (decreases) the instantaneous price return in the next moment.*

$$BidRotation_{t-1} - AskRotation_{t-1} \uparrow \Rightarrow Return_{(t-1;t]} \uparrow$$

The effect of asymmetry of depths on the rotation difference of limit order book can be predicted with the "crowding-out" hypothesis as well. A shock to the depth difference caused by unexpectedly thick or thin book on one side will induce the traders on the same side to act more aggressively in order to execute their orders, thus resulting in a positive response of the rotation difference, as stated in the following hypothesis:

Depth difference and Rotation difference*: An increase (decrease) of depth difference of the limit order book increases (decreases) the rotation difference of the limit order book in the next moment.*

$$BidDepth_{t-1} - AskDepth_{t-1} \uparrow \Rightarrow BidRotation_t - AskRotation_t \uparrow$$

3. Econometric Methods

In this section we briefly discuss the econometric methods used in this paper. We start with the brief description of principal component analysis (PCA) in the context of its applications to dimension reduction of LOB data. Then we provide a summary of the vector autoregressive (VAR) model and introduce the concepts related to structural interpretation of its coefficients. Finally, we provide a brief description of robust test statistics that will be used to evaluate significance and consistency of stock-specific structural VAR analysis results across stocks.

3.1. Principal component analysis

The underlying assumption of principal component analysis (PCA) is that the limit order book depth for stock k across the spectrum of limit price levels can be summarized by a few fundamental factors, called principal components, obtained as linear combinations of columns in the $N\times1$ depth vector d_{kt}.[2] The goal of PCA is to extract and interpret such fundamental factors.

We assume, without loss of generality, that the depth vector d_{kt} has been logged and diurnally adjusted, so that $E[d_{kt}] = 0$. The output of linear dimension reduction technique can be viewed as a system of $q \leq N$ components $f_{kt} = B'_k d_{kt}$, where B_k is a $N\times q$ loading matrix. Let $\Omega_k = Var(d_{kt})$ and $\Omega_k = C_k'\Lambda_k C_k$ be the eigenvalue decomposition of the variance-covariance matrix $\Omega_k = Var(d_{kt})$, where $\Lambda_k = \text{diag}(\lambda_{k1},\lambda_{k2},\ldots,\lambda_{kN})$ with $\lambda_{k1} \geq \lambda_{k2} \geq \ldots \geq \lambda_{kN} > 0$, and C_k is an orthogonal $N\times N$ matrix. Then the elements of vector $f_{kt} = C_k'd_{kt}$ are called the *principal components* of stock k data d_{kt}. Since $Var(f_{kt}) = \Lambda_k$, the principal components are uncorrelated. The matrix C'_{kq} that consists of the first q columns of C_k is the loading matrix of the first q principal components is defined uniquely, up to the columns sign reversal and exchange of columns with identical eigenvalues. Denote by Λ_{kq} the matrix of q largest eigenvalues, corresponding to the first q principal components.

As discussed above, we decompose the limit order book cumulative depth into the first four principal components, which have straightforward interpretation as level and rotation factors of cumulative depth schedules on the ask and bid sides of the book.

3.2. Vector autoregression

Vector autoregression (VAR) is one of the most popular econometric techniques widely used in applied macroeconomics and empirical finance. The underlying assumption in VAR analysis is that the evolution of the endogenous variables of interest z_{kt} can be represented by a set of simultaneous linear equations, with feedback from the own lags of z_{kt} and with the error terms u_{kt} in the right-hand side of those equations. The assumption usually made in applied VAR analysis is that all VAR equations include lags up to a maximum order p_k, which is subsequently estimated using information criteria.

[2] In the subsequent analysis, N=20, since we consider the depth at the ten best limit price levels on the ask side and ten best limit price levels on the bid side of the limit order book.

The assumption of finite order VAR is rarely literally true, so it can be viewed as a convenient approximation of a VAR system of equations with infinite number of lags.

The joint dynamics of the endogenous variables for each of the *K* stocks are modeled by the reduced form VAR

$$z_{kt} = \varphi_{kt} + \Phi_{k1} z_{k,t-1} + ... + \Phi_{kp} z_{k,t-p} + u_{kt},\ k=1,\ldots,K, \qquad (1)$$

where φ_{kt} is the vector of deterministic terms that may include diurnal factors characterizing the deterministic intraday variation in the components of vector z_{kt}, while the error terms u_{kt} are assumed to have a non-diagonal variance-covariance matrix $E[u_{kt}u'_{kt}] = \Sigma_k$, $E[u_{kt}u'_{k't}] = 0$ for $k \neq k'$. We assume, for the time being, that the error terms are serially uncorrelated and mutually uncorrelated across stocks, although the latter assumption can be subsequently relaxed in the extensions of this work.

Another choice that must be made is with regard to the number of lags in the vector autoregression. There is a tradeoff between parsimony of the model and degree of the time series dependence that can be adequately captured by the model. We use Akaike information criterion (AIC) which seems to provide a good compromise keeping the number of lags in check but allowing rich dynamic structure. Note that the number of lags is chosen separately for each stock, and the selected number of lags across 28 stocks varies from 8 to 19[*].

3.3. Impulse responses

The aim of impulse response analysis is to identify structural shocks that are mutually uncorrelated and have a diagonal variance-covariance matrix $E[v_{kt}v'_{kt}] = D_k$, which is allowed to be stock-specific. To determine impulse responses of endogenous variables z_{kt} to the structural shocks, one needs to find a matrix A_k that satisfies the relationship $u_{kt} = A_k v_{kt}$, which obviously implies $\Sigma_k = A_k D_k A'_k$. If each A_k is a full-rank $n \times n$ matrix, then we can assume, without loss of generality, that all its main diagonal elements are unity, and define as an impulse vector any column a_k of such matrix A_k.

To compute impulse responses of endogenous variables, we stack equation (1) as

$$Z_k = \widetilde{Z}_k \Phi_k + U_k, \qquad (2)$$

[*] Certain calculations derived in part from data provided by Reuters.

where $Z_k = [z_{k1},...,z_{kT}]'$, $\tilde{z}_{kt} = [z'_{k,t-1},...,z'_{k,t-p}]'$, $\tilde{Z}_k = [\tilde{z}_{k1},...,\tilde{z}_{kT}]'$, $\Phi = [\Phi_1,...,\Phi_p]'$, and $U_k = [u_{k1},...,u_{kT}]'$. For any $n \times 1$ impulse vector a_k, define $\tilde{a}_k = [a'_k, 0_{1,(p-1)n}]'$ and

$$\Gamma_k = \begin{bmatrix} \Phi'_k \\ I_{(p-1)n} \quad 0_{(p-1)n,n} \end{bmatrix}.$$

Then the impulse response $r_{i,ak}(h)$ of endogenous variable i=1,…,n for stock k to an impulse a_k at horizon h is computed as

$$r_{i,a_k}(h) = (\Gamma_k^h \tilde{a}_k)_i \quad \text{for } h = 0,1,...,H. \tag{3}$$

3.4. Identification of structural shocks

Structural shocks in VAR are commonly associated with exogenous orthogonal "surprise" disturbances hitting endogenous variables. In impulse response analysis, the goal is to examine the effect of a typical (usually normalized) "surprise" (or "shock") on the time path of variables in the model. Shocks to observable variables can be calculated directly using generalized impulse response functions (GIRFs). However, if we wish to know where the shocks are coming from, we may want to decompose the shocks to observables into unobserved components, called structural shocks, which are usually assumed to be orthogonal. To disentangle the shocks that might affect the variables of stock k, we make a set of just-identifying (and therefore non-testable) assumptions, corresponding to a particular causal ordering of the endogenous variables in a Cholesky decomposition of matrix Σ_k. For example, a popular identification scheme used to identify the "information" shock assumes that price variables most commonly associated with economic profits reflect contemporaneously all information that may be relevant to their determination. From this perspective, the price variables are allowed to contemporaneously react to all structural shocks. However, this identification scheme would imply that the price shocks do not have contemporaneous effects on other variables, which is unlikely, given that prices themselves are the product of evolution of the limit order book and other endogenous variables. In the absence of structural model of the limit order book, "surprises" in the price levels and volatility may be associated with simultaneous changes in trading volume and shape of the limit order book.

An alternative approach would be to use long-term economic relations such as the long-term neutrality of endogenous variables to specific shocks, or linear combinations of shocks. One such relationship imposes the zero restriction on the long-term effect of price shocks on endogenous variables other than price level. Such shocks can be associated with stock-specific private information effects. The second possible relationship can be obtained by imposing a zero restriction on the long-run effect of signed volume.

In the absence of a specific structural model of limit order market, we postulate that the price and volatility shocks do not contemporaneously affect the buy and sell trading volume and the limit order book depth factors on bid and ask sides of the book. Moreover, we postulate that depth shocks do not affect contemporaneously the buy and sell trading volumes, and that the price range shock does not affect contemporaneously the bid and ask price shocks. One possible intuition behind this structure is that market order submitters are more exogenous with regard to the depth of the limit order book, as shown empirically by Pascual and Veredas (2006a).

4. Data

The data is extracted from the NYSE OpenBook™, Trades and Quotes data feed provided by Reuters. The period of interest covers 42 trading days, from June 1, 2006 to July 31, 2006. An advantage of this data set is that it consists of the price and number of standing shares at each price level up to 10 steps from the best bid and ask quotes. In other words, the data has the information of the snapshots of limit order book, sampled at a pre-defined time interval[3]. Unlike most of the existing research on limit order books, we do not have to reconstruct the limit order book based on the order level data, which tremendously reduces the chance of making mistakes during this step. Since the focus of this study is on the most liquid stocks in the U.S. equity market, our sample is confined to the Dow Jones Industrial Average (DJIA) component stocks. The list of these stocks

[3] The sampling interval in our data is mostly 5 seconds and sometimes longer if there was no change to any of the price levels in that period. Although the updating of the book during the time period under consideration did not actually occur in real time, it was fine enough considering the one-minute interval data frequency that we use in the present study. At the same time, Pascual and Veredas (2000a) have shown that the limit order book is more informative at 1-minute than 5-minute time interval. Currently NYSE has begun to offer real-time OpenBook™ data to its subscribers.

traded on NYSE as of June 1st, 2006 includes 28 tickers. All data are filtered as explained below:

1) All snapshots corresponding to "locked" or "crossed" market conditions where the best bid is greater or equal to the best ask are removed.
2) All snapshots lasting less than one second are removed, to reduce the high-frequency noise contained in the book.
3) All snapshots, trades, and quotes with the timestamps arriving before 9:45 AM EST or after 3:45 PM EST are removed, since we only focus on market dynamics during the normal trading hours and therefore would like to eliminate the intervals corresponding to opening and closing "abnormal" market conditions.

After the data cleaning, we sample from the remaining snapshots at one-minute frequency. Table I presents summary statistics of the limit order book and trading activities for our sample. Other than the variation in price level across the stocks, all of them have considerably narrow spread, indicating a quite liquid market which is not surprising for the DJIA stocks. "Level 1Ask (Bid) depth" is the number of shares offered at the best ask (bid) price, and "Level 1-10 Ask (Bid) depths" is the cumulative number of shares at the ten best ask (bid) prices. It can be seen from the table that although the spread is relative homogeneous across stocks, the biggest value is more than three times larger than the smallest value for level 1 depths and twice larger for level 1-10 depths. This demonstrates the heterogeneity of limit order book depth across these even most liquidity stocks.

To study the interactive dynamics between limit order book information and market variables, we extract the principal components using the cumulative depth levels of the limit order book at ten best bid and ask prices. In addition, we generate the following five variables:

1) *Trade volume*: Each trade in our data is categorized as either buyer-initiated or seller-initiated, using the Lee and Ready (1991) algorithm, with adjustment introduced in Bessembinder (2003) to accommodate the more active modern trading environment. The sum of buy and sell volumes $BuyVolume_t$ and $SellVolume_t$, measured as the number of shares, is calculated over the one-minute interval preceding time t. The variable *volume* that will be used as one of the VAR

endogenous variables is defined as $log(BuyVolume_t)+log(SellVolume_t)$. Since trade volume and price volatility are strongly correlated (see, for example, Jones et al (1994), Huang and Masulis (2003)), we control for the impact of volume to test whether the information on the limit order book can provide additional explanatory power on price volatility.

2) *Signed volume*: Similar to the definition of *volume*, the signed volume is defined as the difference of logged buy and sell volumes $log(BuyVolume_t)–log(SellVolume_t)$ over the last minute. Inclusion of this variable is motivated by the relationship between the order imbalance and returns established in the literature (Chordia *et al*., 2002, Chordia *et al*., 2004, and Chordia *et al*., 2006). Therefore, we need to control for this relationship to study the incremental contribution of information into the limit order book.
3) *Inside spread*: At any point in time, inside spread is defined as the difference between logged best ask and bid prices $log(Ask_t)–log(Bid_t)$. Thus we consider the relative spread measured as a percentage of current price.[4]
4) *One-minute return*: We calculate the average of logged ask and bid prices at the beginning and end of each one-minute interval, and construct the one-minute mid-quote return from this time series by taking first difference.
5) *Price variation (volatility)*: We use the bid-ask mid-quote to obtain the ratio of maximum mid-quote to minimum mid-quote over the last one-minute interval. After adjusting for the intraday seasonality (see below), we use logarithm of this ratio as a measure of intraday volatility in our model. To make the three price variables comparable to the limit order book depths, we multiply the spread, return and volatility by 10000 to scale them up.

Intraday seasonality is one well-documented phenomenon of high-frequency financial data. Figure 2 shows the plots of buy_vol, sell_vol, spread=(askt/bidt), one-minute price variation, and cumulative bid depths at 1st, 5th, and 10th price levels as a function of time-of-day across the DJIA stocks. Each of the above variables are first averaged within each of the 24 fifteen-minute bins coveting the regular intraday trading hours (9:45 am

[4] Our definition of *inside spread* is closely related to the definition of relative inside spread traditionally used in the literature. Indeed, if $Ask_t/Bid_t=1+x$, then we can easily see that $log(Ask_t)–log(Bid_t)=log(1+x)\sim x–x^2/2$, whereas $(Ask_t–Bid_t)/((Ask_t+Bid_t)/2)=2x/(2+x)\sim x–x^2/2$, if x is close to zero.

EST to 15:45 pm EST) for each stock and each trading day, and then the medians of those averages are calculated across all days for each bin and for each stock. The box-whisker plot is generated across stocks for each fifteen-minute bin. The intraday seasonality is obvious for all the variables. Specifically, the "U-shape" pattern is evident and consistent across the stocks on the plots of buyer- and seller-initiated trade volumes, as documented in Biais et al (1995) and Barclay and Hendershott (2003). The spread tends to be wide at the beginning of trading day and tightens toward the day end, which is consistent with the argument that information accumulates overnight and informed trading is more severe in the morning. Also as shown in Barclay and Hendershott (2003), price variation demonstrates a "U-shape" pattern, which is expected considering its close relationship to the trading activities. Cumulative depths of the limit order book display very similar intraday patterns at all price levels, with small depth levels at the early trading session gradually picking up to the end of the day. This is in line with the result found by Anand et al (2005) that traders tend to use more limit orders as approaching the end of the day.

To filter out this intraday seasonality, we diurnally adjust all endogenous variables except for instantaneous return using the procedure described in Engle (2000). We first fit all endogenous variables by the time-of-day indicators using a linear spline function with knots at every half-hour from 10:00 am to 15:30 pm, and then express the diurnally adjusted values of each endogenous variable as a ratio of its actual value to its linear spline-fitted value.

After having diurnally adjusted the depths of the limit order book at 20 price levels (ten levels for each side of the book), we log them to use as the input in the principal component analysis. To determine the relative explanatory power of principal components, we generate the scree plot of their eigenvalues for the first 10 factors across the 28 DJIA stocks in Figure 3, where on the vertical axis are the logarithms of eigenvalues and on the horizontal axis are the factor numbers. There is clearly a large break between the first and the second pair of factors, and between the second and the third pair of factors, which isolates the first and second pairs of factors as much more important factors from others. In addition, the "short" boxes imply that the incremental explanatory power of the first principal components is consistent across the stocks.

To give economic interpretation of the principal components, we present the loadings of the first four principal components in Table II. First observation from the table is that the overall variation of the loading coefficients across stocks is rather small, which is in line with the "short" boxes we see on the scree plot. More interestingly, each of the first four principal components is related to an important character of the limit order book shape. For example, factor 1 always has positive and similar coefficients over all the bid and ask depths and thus can be treated as the average of bid and ask depths, or "depth average". Meanwhile factor 2 always has positive coefficients for all the bid depths and negative coefficients for all the ask depths and thus can be treated as the difference of bid and ask depths, or "depth difference". While principal components $F1$ and $F2$ capture, respectively, the overall depth level and depth imbalance on bid and ask sides, principal components $F3$ and $F4$ are related, respectively, to rotation of the cumulative depth schedule on the bid and ask sides. Specifically, $F3$ has significant loading coefficients on the ask side and the sign switches from positive to negative when we move away from the best ask price. Therefore, $F3$ measures the "tilting", or rotation, of cumulative depth on the ask side of the book. Similarly, $F4$ measures the rotation of the bid side of the book. Together these four factors explain around 85%[*] of the total variance, indicating they are quite successful in capturing all the information contained in the depths of limit order book.

To be consistent with the meanings of factor 1 and factor 2, we do not use factor 3 or factor 4 directly in our model. Instead, we define two new variables: one is factor 3 plus factor 4; the other is factor 4 minus factor 3. Then the former is the average of bid and ask rotations and the later is the difference between them.

Thus we have extracted four factors from the depths of the limit order book, named "depth average ($F1$)", "depth difference ($F2$)", "rotation average ($F3$)", and "rotation difference ($F4$)". The "depth average" and "rotation average" can be labeled as "variational factors" and are likely to interact closely with the level variables of the market, such as volume, spread, and volatility. On the other hand, the "depth difference" and "rotation difference" are related to the asymmetry between the bid and ask sides of the book, labeled as "directional factors" and we expect them to interact closely with the

[*] Certain calculations derived in part from data provided by Reuters.

difference variables such as instantaneous return. Figure 4 gives a graphic illustration of these four factors by a prototypical limit order book, where the depths are set to be the median in our sample.

5. Empirical Results

Based on their definitions, our variables can be grouped into two blocks: one contains all "variational" variables such as volume, depth average, rotation average, spread, and volatility; the other contains all the "directional" variables such as signed volume, depth differences, rotation differences, and return. Panels A and B of Figure 5 present cross-correlograms of variables within these two blocks. As can be seen from the plots of autocorrelograms on the diagonal of each panel, all variables with exception of mid-quote return are mildly persistent, with level average ($F1$), level imbalance ($F2$), and trade volume (*TVol*) displaying a substantial degree of persistence, and signed volume (*SVol*) is only mildly persistent. The limit order book depth-related variables $F1$, $F2$, $F5$, and $F6$ are strongly serially correlated, implying that the overall liquidity level supplied by the limit order book adjusts rather sluggishly, and this slow adjustment pattern holds consistently across all DJIA stocks. In contrast, the instantaneous return is significantly short-memory, confirming that the stock price movement is close to a random-walk and virtually unpredictable based on the recent mid-quote data.

The cross-correlation functions within each block of variables provide additional insights into the patterns of interactions between the dependent variables across the spectrum of leads and lags. Note, for instance, that the higher trade volume (*TVol*) leads to rather persistent increases in the magnitude of both level ($F1$) and rotation ($F5$) factors, and to a temporary (indeed, rather short-lived) drop in the level of bid-ask spread. Curiously, we do not observe a consistent reverse lead-lag pattern from the liquidity supply factor ($F1$, $F5$) to the liquidity demand, represented by trade volume. As expected, the levels of bid-ask spread and short-term price variation (*Range*) are positively correlated at all leads and lags; however, the similar symmetric positive cross-correlation pattern can be detected only between trade volume and price variation, but not between trade volume and bid-ask spread. As expected, the increased levels of liquidity supply

lead to persistently lower levels of both bid-ask spread and price variation in the future. Conversely, the greater levels of spread and price variation are associated with persistently lower levels of liquidity supply in the future.

On the other hand, we find that the signed volume (*SVol*) is positively correlated with contemporaneous rate of return (*Return*) but this mutual correlation pattern does not persist over time, with the current minute positive return leading to the next minute positive signed order flow, and the reverse relationship between current minute positive order flow not leading to higher than usual returns in the future. Similarly, the current positive return leads to greater imbalance between buyer- and seller-initiated liquidity supply factors (*F*2 and *F*6) within the next minute. Conversely, the rotation discrepancy factor (*F*6) has a mild predictive power for the next minute rate of mid-quote return, while the level discrepancy factor (*F*2) has a mild but persistent negative relation to future mid-quote returns.

5.1. Testing Granger causality

The four principal components (*depth average*, *depth difference*, *rotation average*, *rotation difference*) extracted from the vector of cumulative depths at all tick levels are included as components of the 9×1 vector of endogenous variables in the reduced form VAR, along with the following five variables defined above: (1) *Volume*, (2) *Signed volume*, (3) *Spread*, (4) *Return*, and (5) *Volatility*. We apply the standard *F*-tests of Granger causality for each of the DJIA stocks, collecting K=28 test statistics for every pair of nine endogenous variables. If we maintain the standard assumptions underlying the heterogeneous VAR (Pesaran and Smith, 1995) that the VAR coefficients Φ_k vary across individual stocks k =1,…,K according to a simple random-coefficient specification $\Phi_k = \Phi+\eta_k$, where η_k are random matrices distributed identically and independently of u_{kt} and z_{kt} with zero mean and a constant covariance matrix, then we could have based our inference on the mean group estimator of Φ since it was found to be consistent as $T\rightarrow \infty$ and $K\rightarrow \infty$ (Pesaran *et al*., 1996). However, since those assumptions (in particular, $K\rightarrow\infty$) may be debatable, we prefer to report the medians, 10th percentiles, and 90th percentiles of the sample of stock-specific *F*-statistics, and make inference about Granger causality across stocks based on the 10th percentiles of those samples of *F*-statistics.

The Granger causality tests for endogenous variables in the reduced form VAR are summarized in Table III. To make these results comparable across stocks, we report the tests assuming the same number of lags in VAR for all stocks in our sample. We show the results for several such common number of lags choices p = 5, 10, and 15. For each pair of endogenous variables, we obtain the values of F-test statistics for 28 stocks, rank those statistics by their size, and report the median, 10^{th}, and 90^{th} percentiles in the table. The values of F-tests for which the 10^{th}-percentile is significant at 1% level are highlighted by stars (***).

Based on the results of this analysis, all nine endogenous variables, with a notable exception of mid-quote return, have considerable time series dependence, whereas the one-minute mid-quote return is only weakly correlated with lagged values of other endogenous variables. We also find that transaction volume, for instance, is Granger-caused by the lagged values of bid-ask spread and the level and rotation factors of the book, as expected. The signed volume, in contrast, can not be explained by lagged values of most of the other variables, except the level difference in the book. At the same time, the level and rotation factors of the book are Granger-caused by bid-ask spread. We also observe strong interactions between the level and rotation factors of the book and between the level and rotation discrepancies across the bid and ask sides of the book.

One of the goals of this paper is to determine if the limit order book information helps to predict the future dynamics of bid and ask prices. The results can be illustrated by the last three equations of the reduced form VAR. The quoted spread, which can be viewed as a measure of roundtrip transaction cost incurred by small liquidity traders, is Granger-caused by the level factor of the book. The one-minute rate of return is partly explained by the past signed trading volume. However, the dependence is rather weak and fades away very quickly. On the other hand, Table III shows that the difference between the limit order book levels on bid and ask sides can also be viewed as a predictive factor for one-minute return. While Cao *et al.* (2004) report a similar finding for the predictive power of the depth differential, the signed volume is not included in their model and thus is not controlled for. It has also been documented extensively that the trade volume is important for prediction future price volatility. However, as shown in Table III, the average limit order book depth level has at least as much predictive power for mid-quote

price volatility as the transaction volume. This result is also consistent with the findings of Ahn *et al*. (2001) and Coppejans *et al*. (2001).

5.2. Structural VAR analysis

We use the structural VAR framework to identify the contemporaneous links between endogenous variables. As described in Section 3 of this paper, we will identify matrix A_k by selecting a particular identification scheme $\Sigma_k = A_k D_k A_k'$ based on a strict triangular restriction underlying the Cholesky factorization of matrix Σ_k. The recursive ordering of endogenous variables used in this section is (*Volume*, *Signed volume*, *Depth average*, *Depth difference*, *Rotation average*, *Rotation difference*, *Spread*, *Return*, *Volatility*)[5]. The intuition behind this ordering is that trading volume is the most exogenous variable, which affects the state of limit order book and the market variables. This is partly motivated by the Easley and O'Hara (1992) model and consistent with the finding by Pascual and Veredas (2006b) that market orders are relatively exogenous to the shape of limit order book. After estimating the reduced form VAR with the number of lags selected by Akaike information criterion for each stock, we obtain the estimates of matrices A_k and $B_k = A_k^{-1}$, and summarize their coefficients reporting their medians, 10^{th}, and 90^{th} percentiles across stocks k=1,…,K. The results for matrices A_k and B_k are presented in Tables IV and V, respectively.

A preliminary check of the results reveals that shocks associated with the overall level of liquidity supplied by the limit order book tend to interact strongly with transaction volume, spread, and volatility, while the shocks associated with imbalance of liquidity supplied by the limit order book tend to interact with the signed volume and rate of return. We call the first group of variables "variational" and the second group of variables "directional" variables.

The interpretation of results within those two groups can be performed within the context of the specific identification scheme used in this paper. For instance, we can see from Table V that a shock to the trade volume, which can be interpreted as a liquidity

[5] An alternative ordering of variables we also tried is (*Depth average*, *Depth difference*, *Rotation average*, *Rotation difference*, *Volume*, *Signed volume*, *Spread*, *Return*, *Volatility*), which gives the same results. To conserve space, we do not report the results here, but they are available from the authors upon request.

demand shock within a time period (t–1; t] affects positively the volatility and spread within the same time period, while its effect on the factors associated with the limit order book liquidity is significantly negative. The shocks to both "variational" LOB factors ($F1$ and $F5$), which are assumed to be orthogonal to the liquidity demand shock, can be associated with the liquidity supplied by limit order traders. Both of these shocks are positively related with the bid-ask spread but negatively related to volatility, so those shocks (especially the positive shock to the rotation average factor) can be viewed as price stabilizing shocks. Finally, a shock to the bid-ask spread, assumed to be orthogonal to both liquidity demand and supply shocks, is positively related to the mid-quote price volatility.

The interpretation of contemporaneous results for the four "directional" variables in the bottom part of Tables IV and V can be performed similarly. The signed volume shock, associated with the imbalance of liquidity demanded by aggressive buyers and sellers, not only is strongly contemporaneously associated with positive returns, but also positively associated with both "directional" LOB factors ($F2$ and $F6$), suggesting that there is a positive contemporaneous relationship between the imbalance of liquidity supply (by limit order traders) and the imbalance in liquidity demand (from market order traders). The shocks to both "directional" LOB factors that are orthogonal to the directional liquidity demand are positively correlated but have distinct contemporaneous associations with the mid-quote returns. While the shock to the total depth level imbalance affects the mid-quote return negatively, the signs of associations of shock to the LOB rotation difference vary across stocks in a broad range from –0.41 (10^{th} percentile) to +0.76 (90^{th} percentile)[*]. This finding suggests that the role played by "urgent" liquidity suppliers cannot be summarized uniformly across all stocks, but rather should be viewed specifically within the context of a particular market and time period.

Based on the estimated matrices A_k and B_k, we then obtain impulse response functions (IRFs) of the nine endogenous variables to nine structural shocks. To conserve space, Figure 6 presents only responses of variables to shocks within the group of "variational" endogenous variables and within the group of "directional" endogenous variables, since interaction between those two groups is found to be insignificant across the DJIA stocks.

[*] Certain calculations derived in part from data provided by Reuters.

Panel *A* shows the IRFs of the "variational" variables to each of the shocks. A positive shock to the general level of liquidity supplied on both sides of the limit order book leads to an increase in transaction volume within the next several minutes. Similar to the result in Biais et al (1995), this finding firms the "depth and volume" hypothesis, which is consistent with the theoretical prediction of Parlour (1998) that exogenous shocks to the liquidity provided by the limit order book encourages arrival of market orders: A similar pattern is observed for a symmetric rotation shock experienced by both sides of the limit order book, implying that the more active supply of liquidity near the best quotes enhances the future trade volume. The negative IRF of trade volume to the positive shock in spread implies that the wider transaction cost in the market discourages trading activity, at least in the short run, which is actually the possible intention of the "market maker" in the Kyle (1985) and Glosten and Milgrom (1985) models. Interestingly, there is no clear evidence that volatility shock (as long as it is orthogonal to other shocks) discourages trading activity in the market as predicted in Foucault (1999).

The second row in Panel *A* of Figure 6 presents the responses of depth level average of the limit order book to other shocks. As a measure of overall level of supplied liquidity, the average level responds to a larger number of shocks. Although the impact of general liquidity demand shock to the level average is inconclusive, the negative response of the average LOB depth level to the rotation shock indicates that an increased liquidity supply at the levels close to current bid and ask quotes ultimately leads to reduction of the overall liquidity supply on the market. While the exact reason is not clear, one possible explanation could be that the more severe competition induced by the higher level of aggressiveness induces the existing limit orders to be cancelled and re-submitted as market orders. At the same time, a larger roundtrip transaction cost leads to the lower level of liquidity supplied by limit order traders, confirming our "spread and depth" hypothesis and the finding by Lee, Mucklow, and Ready (1993) that traders tend to adjust both spread and depth when facing severe adverse selection issue. Finally, the volatility shock reduces the level of depth supplied by limit order traders, which is consistent with our "volatility and depth" hypothesis. Coppejans *et al*. (2001) and Goldstein and Kavajecz (2004) reported the similar finding, although it is in contrast to Ahn et al (2001).

The responses of the limit order book rotations behave as we expected too. Other than that the shock to trading volume does not have a clear effect on the rotation average, we can see the traders become less aggressive when observing higher volatility or wider spread. This "rotation" effect of the spread on the limit order book is also documented by Chan (2005). Meanwhile, the "depth average and rotation average" hypothesis is confirmed pronouncedly, showing that thicker book forces the traders to act more aggressively, as also found by Ranaldo (2004).

As predicted by Easley and O'Hara (1992), the bid-ask spread widens as more market orders arrive, possibly in response to the greater information asymmetry across the traders. The greater aggressiveness of traders, associated with a positive shock to liquidity provided near the top of the limit order book, induces some limit order traders jump the queue, thereby reducing the bid-ask spread. We also find mild evidence in favor of our "depth and spread" hypothesis that depth and spread dynamically complement each other. Meanwhile, the bid-ask spread tends to increase in response to the price volatility shock, although the effect of this shock lasts only several minutes. This is in line with the results shown by Copeland and Galai (1983), but in contrast to the finding in Bae et al (2003) that lagged price volatility shrinks the spread.

The last row in Panel A is the IRFs of volatility. There have been several recent empirical studies of the effect of limit order book on the volatility. While Ahn *et al*. (2001) and Coppejans *et al*. (2001) argue that the average liquidity supply affects volatility; Nas and Skjeltorp (2006) relate volatility to the homogeneity of the traders' expectations. As can be seen from the principal component analysis of this paper, the two measures capture in fact different aspects of shape of the limit order book. It is shown clearly that both a greater liquidity supply near the top of the book (at the expense of liquidity supplied deeper within the book) and a shock to the average liquidity have negative effects on the volatility, while the later is more persistent. This can be viewed as evidence in support of our "depth and volatility" hypothesis. Furthermore, we also observe the well-known volume-volatility association and the spread-volatility association.

Now we turn to Panel *B* of Figure 6 for the responses of endogenous variables to the "directional" shocks. First of all it shows that shocks to liquidity supply decrease, instead

of increasing, the signed trade volume, meaning that the liquidity shock on one side actually attracts more market orders from the other side. This is in contrast to the "depth imbalance and signed volume" hypothesis implied by Parlour (1998), but in line with the evidence found by Hall and Hautsch (2007) that trades tend to take liquidity when the liquidity supply is high. One possible explanation for this conflict is that in Parlour (1998)'s model, traders can only choose either limit order or market order. In reality, however, traders have many more dimensions of aggressiveness to explore; for example they may indeed submit more aggressive limit orders, as stated in the "depth difference and rotation difference" hypothesis.

Note that the depth level difference has negative responses to the rotation difference, similar to the one between the level average and rotation average as observed above.

Next row is the responses of LOB rotation imbalance to the other three "directional" shocks. Notice that the unlike the negative response of depth difference to rotation difference, a shock to depth difference increases the rotation difference, supporting the "depth difference and rotation difference" hypothesis. Meanwhile, higher return in the previous one or two minutes induces more rotation difference too, which is in line with the "return and rotation difference" hypothesis.

Finally, the responses of return tell a more nuanced story. First, the rate of return is unaffected by its own past shocks, which is consistent with weak market efficiency. Second, a positive shock to liquidity demand induces a positive response of the rate of return at the next minute. Additionally, the imbalances of depth level and rotation factors in the limit order book have positive impact on the instantaneous return, although the effect completely dissipates within the next few minutes. This is in line with the "depth imbalance and return" hypothesis, which is also consistent with the theoretical prediction of Bias et al (1995), Chakravarty and Holden (1995) and Parlour (1998).

To investigate how each of the shocks contributes to variation of endogenous variables over time, we present the forecast error variance decomposition (FEVD) at several different horizons in Table VI. However, as a word of warning, we have to realize that this approach hinges on the structure of identifying restrictions imposed on matrix A in the structural VAR. So, unavoidably, the results are subject to misspecification of the structural shocks.

By and large, the FEVDs reveal a similar picture as we have seen from IRFs in Figure 6. There are at least two additional observations. First, the variance decomposition of endogenous variables is relatively stable as a function of horizon. Consistent with the IRF results, the variances of VAR forecasts for signed trade volume and liquidity supply near the top of the limit order book are among the most stable, while the variance of VAR forecasts for the LOB depth level factors is growing with the horizon considerably, and subject to contributions from a more diverse set of shocks, as shown in Chan (2005). Second, the contribution of liquidity supply shocks associated with the changes of shape of the limit order book as a whole have roughly as much impact on the endogenous variables as other shocks. For instance, the four shocks associated with the shape of the limit order book together contribute as much to the variance of return as liquidity demand imbalance shocks. Although they contribute much less to the price volatility than symmetric liquidity demand shocks, their fraction is at the same level as spread. As said before, the FEVD may be sensitive to the specification of structural shocks. However, alternative recursive identification schemes that we tried lead to similar results, indicating importance of the limit order book in explaining market dynamics.

Pascual and Veredas (2006a, 2006b) demonstrate that the time resolution also matters in explain the behavior of price volatility. As a robustness check, we also computed the volatility at a 5-minute frequency, which is defined as the price range within the previous 5 minutes. The impulse responses between the 5-minute price volatility and other variational variables are presented in Figure 7. As shown in the graphs, the main results we saw before still remain. One thing worth pointing out is, now the higher price volatility induces larger trading volume, implying that although liquidity demand is not responsive to immediate volatility shock, it is in a longer horizon.

6. Conclusion

This paper attempts to shed light on the question, “what information surprises to the limit order book changes convey about the future evolution of trading volume, return and volatility of the same stock, and across the DJIA stocks?”. This goal is achieved by studying the impulse responses of VAR endogenous variables the four limit order book

depth components. We find that the information of limit order book captured by these four components is important in predicting the price volatility and instantaneous return in the future. The impulse responses reveal that the state of limit order book is affected by the shocks to market variables such as spread and volatility. At the same time, price volatility is responsive to the overall level of liquidity supply in the limit order book, even after controlling for incoming trading volume and spread. Price return can be explained by shocks to the imbalance in liquidity supply in the book, even after controlling for the order imbalance, although the effect is only short-run and mild. In Table VII, we present the summary of empirical findings by recent literature and show how our results fit in.

One intriguing extension of this work would be to incorporate common factors in the VAR error components, augmenting the stock-specific VAR regressions with contemporaneous and lagged cross-sectional averages across stocks, as described in Pesaran (2006). If successful, this approach has a potential to characterize one or several factors driving common liquidity across stocks. Another important extension would be to introduce nonlinearities into the model by augmenting it with GARCH and regime-switching structures, which would produce a more realistic picture of market dynamics. Those extensions will be pursued in our future research.

Disclaimer

This material is for your private information. The views expressed are the views of Eldar Nigmatullin, Konstantin Tyurin, and Hao Yin only through the period ended October 4, 2007 and are subject to change based on market and other conditions. The opinions expressed may differ from those with different investment philosophies. The information we provide does not constitute investment advice and it should not be relied on as such. It should not be considered a solicitation to buy or an offer to sell a security. It does not take into account any investor's particular investment objectives, strategies, tax status or investment horizon. We encourage you to consult your tax or financial advisor. All material has been obtained from sources believed to be reliable, but its accuracy is not guaranteed. There is no representation or warranty as to the current accuracy of, nor

liability for, decisions based on such information. Past performance is no guarantee of future results.

References

Ahn, Hee-Joon, Kee-Hong Bae, and Kalok Chan, 2001, Limit orders, depth and volatility: evidence from the stock exchange of Hong Kong, *Journal of Finance*, 56, 767-788.

Anand, A., S. Chakravarty, and T. Martell, 2005, Empirical evidence on the evolution of liquidity: choice of market versus limit orders by informed and uninformed investors, *Journal of Financial Markets* 8, 265-287.

Bacidore, J., R. Battalio, and R. Jennings, 2003, Order Submission Strategies, Liquidity Supply, and Trading in Pennies on the New York Stock Exchange, *Journal of Financial Markets* 6,337-62.

Bae, K-H., H. Jang, and K. Park, 2003, Traders' Choice between Limit and Market Orders: Evidence from NYSE Stocks, *Journal of Financial Markets* 6, 517-38.

Barclay, M. and T. Hendershott, 2003, Price Discovery and Trading After Hours, *Review of Financial Studies* 16, 1041-1073.

Beltran-Lopez, Helena M, Giot, Pierre and Grammig, Joachim, 2006, Commonalities in the Order Book, CORE Discussion Paper

Bessembinder, Hendrik, 2003, Issues in assessing trade execution costs, *Journal of Financial Markets* 6, 233-257

Bessembinder, Hendrik, Panayides, Marios and Venkataraman, Kumar, 2006, In Search of Liquidity: An Analysis of Order Submission Strategies in Automated Markets, Working Paper, University of Utah

Biais, B., Hillion, P. and C. Spatt, 1995, An empirical analysis of the limit order book and the order flow in the Paris bourse, *Journal of Finance* 50, 1655-1689.

Cao, C, Hansch, O, and Wang, X. 2004, The informational content of an open limit order book. EFA 2004 Working Paper, Penn State University and Southern Illinois University

Chakravarty, S., and C. Holden, 1995, An Integrated Model of Market and Limit Orders, *Journal of Financial Intermediation* 4, 213-41.

Chan, Yue-cheong, 2005, Price Movement Effects on the State of the Electronic Limit-Order Book, *Financial Review* 40, 195–221

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

Chordia, T., and A. Subrahmanyam, 2004, Order imbalance and individual stock returns: Theory and evidence, *Journal of Financial Economics* 72, 485-518.

Chordia, Tarun, Roll, Richard W. and Subrahmanyam, Avanidhar, 2006, Liquidity and market efficiency, Available at SSRN: http://ssrn.com/abstract=794264

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

Coppejans, Mark, Ian Domowitz, and Ananth Madhavan, 2001, Liquidity in an automated auction, Working Paper, Duke University.

Easley, D. and M. O'Hara, 1992, Time and the process of security price adjustment, *Journal of Finance* 47, 577-605

Ellul, Andrew, Holden, Craig W., Jain, Pankaj K. and Jennings, Robert H., 2007, Order dynamics: recent evidence from the NYSE, EFA 2003 Annual Conference Paper No. 836

Engle,R.F., 2000. The economics of ultra-high-frequency data. *Econometrica* 68, 1-23

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

Glosten, L. R. and Milgrom, P. R., 1985, Bid, Ask, and Transaction Prices in a Specialist Market With Heterogeneously Informed Traders, *Journal of Financial Economics* 14, 71-100

Goldstein, M., and K. Kavajecz, 2004, Trading strategies during circuit breakers and extreme market movements, *Journal of Financial Market*s 7, 301–333

Hall, A. D. and Hautsch N., 2007, Modelling the buy and sell intensity in a limit order book market, *Journal of Financial Markets*, In Press

Handa, P. and Schwartz, R.A., 1996, Limit order trading. *Journal of Finance* 51**,** 1835–1861

Handa, Puneet, Robert A. Schwartz, and Ashish Tiwari, 2003, Quote setting and price formation in an order driven market, *Journal of Financial Markets*, 6, 461-489

Harris L. E, and Panchapagesan, V., 2005, The information content of the limit order book: evidence from NYSE specialist trading decisions, *Journal of Financial Markets* 8, 25-67.

Huang, R. and Masulis, R., 2003, Trading Activity and Stock Price Volatility: Evidence from the London Stock Exchange, *Journal of Empirical Finance* 10, 249-269

Jones, C., Kaul, G., Lipson, M., 1994. Transactions, volume, and volatility. *Review of Financial Studies* 7, 631–651.

Kavajecz, Kenneth A. and Elizabeth Odders-White, 2004, Technical analysis and liquidity provision, *Review of Financial Studies* 17: 1043-1071

Large, Jeremy, 2007, Measuring the resiliency of an electronic limit order book, *Journal of Financial Markets* 10, 1-25

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

Lee, C., B. Mucklow, and M. Ready, 1993, Spreads, depths and the impact of earnings information: An intraday analysis, *Reviews of Financial Studies* 6, 345-374.

McInish, Thomas H. and Wood, Robert A., 1992, An Analysis of Intraday Patterns in Bid/Ask Spreads for NYSE Stocks**,** *Journal of Finance* 47, 753-764

Næs, Randi and Johannes A. Skjeltorp, 2006, Order book characteristics and the volume–volatility relation: Empirical evidence from a limit order market, *Journal of Financial Markets* 9, 408–432.

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

Pascual, Roberto and Veredas, David, 2006a, What Pieces of Limit Order Book Information Matter in Explaining the Behavior of Aggressive and Patient Traders?, Working paper.

Pascual, Roberto and Veredas, David, 2006b, Does the open limit order book matter in explaining long run volatility?, Working paper

Pesaran, M. Hashem, 2006, Estimation and Inference in Large Heterogeneous Panels with a Multifactor Error Structure, *Econometrica* 74, 967-1012.

Pesaran, M. Hashem, and Ronald P. Smith, 1995, Estimating Long-Run Relationships from Dynamic Heterogeneous Panels, *Journal of Econometrics* 68, 79-113.

Pesaran, M. Hashem, Ronald P. Smith, and K. Im, 1996, Dynamic linear models for heterogeneous panels in Matyas L. and P. Sevestre (editors), *The Econometrics of Panel Data*, Kluwer Academic Publishers, London.

Ranaldo, A., 2004, Order Aggressiveness in Limit Order Book Markets, *Journal of Financial Markets* 7, 53-74.

Roll, R., 1984, A Simple Implicit Measure of the Effective Bid-Ask Spread in an Efficient Market, *Journal of Finance* 39, 1127-1139.

Yeo, Wee Yong, 2006, Persistence in limit order flow, available at SSRN: http://ssrn.com/abstract=923022

Table I
Summary Statistics of Trading Activity and Book Depth

This table gives daily average descriptive statistics of the trading activity and book depth on the 28 DJIA stocks traded on the NYSE between 9:45 am and 3:45 pm EST from 06/01/2006 to 07/31/2006. The symbols are sorted in descending order of the common stock market value as of May 31st, 2006. "Mid-Quote" is the average of the best bid and ask prices (in dollars). "Relative Spread" is the ratio of distance between the best bid and ask prices to the prevailing mid-quote (in 1/100ths percent). "Level 1 Ask (Bid) depth" is the number of shares offered at the best ask (bid) price as a fraction of daily volume (in percent). "Level 1–10 Ask (Bid) depth" is the accumulated number of shares offered at the ten best ask (bid) prices as a fraction of daily volume (in percent). "Daily Volume" is the average of traded volume (in million shares).

Symbol	**Mid-Quote** (dollars)	**Relative Spread** (1/100ths of a percent)	**Level 1 Ask Depth** (percent of daily volume)	**Level 1 Bid Depth** (percent of daily volume)	**Level 1–10 Ask Depth** (percent of daily volume)	**Level 1–10 Bid Depth** (percent of daily volume)	**Daily Volume** (million shares)
XOM	61.78	2.25	0.055	0.009	8.49	9.05	8.375
GE	33.33	3.30	0.132	0.028	9.15	10.64	8.290
C	48.45	2.65	0.075	0.012	9.15	13.24	5.754
WMT	46.42	3.04	0.046	0.013	13.41	13.00	5.255
JNJ	61.02	2.49	0.066	0.011	13.32	11.50	3.770
PG	55.68	2.89	0.055	0.010	12.13	12.47	3.809
PFE	23.63	4.61	0.105	0.070	11.19	8.66	9.901
AIG	59.59	2.97	0.051	0.009	12.85	13.63	3.235
MO	74.67	2.49	0.051	0.009	11.29	10.71	3.244
JPM	42.24	3.22	0.067	0.018	10.51	12.53	4.942
IBM	77.03	2.46	0.042	0.013	10.81	9.86	3.538
KO	43.38	3.29	0.077	0.026	15.17	12.99	2.956
T	27.57	4.05	0.197	0.080	7.54	8.65	7.425
VZ	32.45	3.95	0.117	0.039	10.78	11.98	4.069
HPQ	31.81	3.88	0.103	0.048	9.03	11.12	6.449
HD	35.71	3.82	0.049	0.056	14.42	13.03	5.687
MRK	36.06	3.60	0.107	0.040	9.11	10.80	4.162
AXP	52.54	3.18	0.078	0.022	12.03	15.04	2.249
BA	81.33	3.37	0.044	0.013	11.27	12.85	2.228
UTX	61.52	3.89	0.048	0.015	12.03	13.31	2.288
MMM	76.94	3.45	0.047	0.015	10.49	11.83	2.163
DIS	29.52	4.18	0.070	0.060	16.74	13.99	5.231
CAT	70.81	3.37	0.043	0.016	10.58	14.59	2.551
MCD	33.62	4.23	0.098	0.050	13.60	15.26	2.588
DD	40.85	3.70	0.086	0.043	11.88	14.02	2.182
HON	38.75	4.84	0.060	0.037	15.64	16.13	2.265
AA	30.79	5.33	0.065	0.053	13.64	18.28	3.087
GM	27.82	5.82	0.080	0.155	8.26	12.52	4.959
Average	47.69	1.61	0.075	0.035	11.59	12.56	4.380
Std.Dev.	17.48	0.45	0.034	0.031	2.34	2.23	2.136

Certain calculations derived in part from data provided by Reuters.

Table II
Loading Patterns of Principal Components of Limit Order Book Depth

This table presents the loading patterns of the first four factors (principal components) extracted from the cumulative depth of NYSE OpenBook for 28 DJIA stocks traded on NYSE from 06/01/2006 to 07/31/2006. The percentiles are based on the ranked patterns of loadings across 28 stocks. The "cumulative proportion" is the fraction of total variance of the limit order book cumulative depth explained jointly by the first 1, 2, 3, and 4 principal components.

	Factor1			**Factor2**			**Factor3**			**Factor4**		
	Median	10th Percentile	90th Percentile	**Median**	10th Percentile	90th Percentile	**Median**	10th Percentile	90th Percentile	**Median**	10th Percentile	90th Percentile
asize1	**0.42**	[0.37;	0.48]	**–0.32**	[–0.36;	–0.27]	**0.56**	[0.45;	0.65]	**0.00**	[–0.44;	0.37]
asize2	**0.58**	[0.52;	0.64]	**–0.45**	[–0.49;	–0.41]	**0.47**	[0.39;	0.60]	**–0.01**	[–0.40;	0.34]
asize3	**0.65**	[0.59;	0.71]	**–0.53**	[–0.57;	–0.49]	**0.32**	[0.26;	0.43]	**–0.02**	[–0.26;	0.23]
asize4	**0.70**	[0.64;	0.75]	**–0.57**	[–0.61;	–0.54]	**0.15**	[0.11;	0.21]	**–0.01**	[–0.14;	0.12]
asize5	**0.72**	[0.67;	0.77]	**–0.58**	[–0.65;	–0.54]	**0.01**	[–0.01;	0.04]	**–0.01**	[–0.04;	0.02]
asize6	**0.73**	[0.67;	0.77]	**–0.59**	[–0.67;	–0.55]	**–0.10**	[–0.13;	–0.08]	**–0.01**	[–0.08;	0.06]
asize7	**0.72**	[0.66;	0.76]	**–0.60**	[–0.67;	–0.55]	**–0.20**	[–0.23;	–0.15]	**–0.02**	[–0.15;	0.16]
asize8	**0.70**	[0.62;	0.74]	**–0.59**	[–0.66;	–0.53]	**–0.27**	[–0.32;	–0.21]	**–0.01**	[–0.20;	0.19]
asize9	**0.67**	[0.57;	0.72]	**–0.58**	[–0.64;	–0.51]	**–0.31**	[–0.37;	–0.24]	**–0.02**	[–0.24;	0.23]
asize10	**0.64**	[0.52;	0.70]	**–0.54**	[–0.62;	–0.46]	**–0.33**	[–0.39;	–0.25]	**–0.03**	[–0.25;	0.25]
bsize1	**0.40**	[0.36;	0.45]	**0.30**	[0.26;	0.33]	**0.04**	[–0.39;	0.47]	**0.57**	[0.46;	0.67]
bsize2	**0.55**	[0.50;	0.60]	**0.45**	[0.40;	0.48]	**0.03**	[–0.32;	0.40]	**0.51**	[0.40;	0.58]
bsize3	**0.62**	[0.58;	0.70]	**0.54**	[0.49;	0.57]	**0.03**	[–0.22;	0.27]	**0.36**	[0.25;	0.41]
bsize4	**0.67**	[0.63;	0.74]	**0.58**	[0.53;	0.62]	**0.01**	[–0.10;	0.13]	**0.17**	[0.09;	0.20]
bsize5	**0.70**	[0.65;	0.76]	**0.60**	[0.55;	0.65]	**0.00**	[–0.02;	0.02]	**0.01**	[–0.02;	0.04]
bsize6	**0.71**	[0.65;	0.76]	**0.61**	[0.57;	0.68]	**–0.01**	[–0.09;	0.07]	**–0.11**	[–0.13;	–0.08]
bsize7	**0.70**	[0.64;	0.75]	**0.61**	[0.57;	0.68]	**–0.02**	[–0.16;	0.13]	**–0.20**	[–0.24;	–0.15]
bsize8	**0.68**	[0.61;	0.74]	**0.60**	[0.56;	0.67]	**–0.03**	[–0.23;	0.18]	**–0.27**	[–0.31;	–0.19]
bsize9	**0.66**	[0.59;	0.73]	**0.59**	[0.53;	0.65]	**–0.04**	[–0.26;	0.21]	**–0.32**	[–0.36;	–0.21]
bsize10	**0.63**	[0.57;	0.71]	**0.57**	[0.49;	0.62]	**–0.04**	[–0.27;	0.20]	**–0.32**	[–0.37;	–0.22]
Cumulative Proportion	**0.41**	**[0.38;**	**0.47]**	**0.72**	**[0.69;**	**0.76]**	**0.79**	**[0.77;**	**0.82]**	**0.85**	**[0.83;**	**0.89]**

Certain calculations derived in part from data provided by Reuters.

Table III
Tests of Granger-causality in the heterogeneous VAR model

This table presents the tests of Granger-causality based on the vector autoregressions estimated separately for 28 DJIA stocks traded on NYSE from 06/01/2006 to 07/31/2006. All variables are expressed in logarithms and diurnally adjusted. The percentiles are calculated for the samples of Granger-causality test statistics ranked across 28 stocks. The null hypothesis is that the lagged values of the variable in the second column jointly do not improve VAR prediction of the variable in the first column. Certain calculations derived in part from data provided by Reuters.

		Lags=5			Lags=10			Lags=15		
Equation	**Variable**	**Median**	**10th Percentile**	**90th Percentile**	**Median**	**10th Percentile**	**90th Percentile**	**Median**	**10th Percentile**	**90th Percentile**
Volume	*Volume*	361.83***	[217.44;	545.04]	182.78***	[117.11;	282.43]	121.49***	[79.18;	191.08]
	Signed volume	1.47	[0.39;	3.80]	1.22	[0.53;	2.22]	1.15	[0.72;	2.04]
	Level average	12.51***	[2.13;	42.40]	7.34	[1.96;	20.88]	5.53	[1.70;	13.81]
	Level difference	2.03	[0.38;	3.25]	1.40	[0.54;	1.99]	1.18	[0.64;	1.64]
	Rotation Average	16.09***	[6.52;	34.96]	7.85***	[3.63;	14.25]	5.17***	[2.70;	8.78]
	Rotation Difference	5.02	[1.48;	15.58]	3.04	[1.20;	5.88]	2.44	[0.89;	4.09]
	Spread	17.32***	[7.47;	36.41]	9.20***	[3.78;	16.70]	6.32***	[3.09;	11.00]
	Return	1.09	[0.35;	2.90]	0.96	[0.59;	1.94]	0.95	[0.63;	1.66]
	Volatility	2.88	[1.63;	8.79]	1.89	[1.12;	4.21]	1.51	[1.08;	2.88]
Signed volume	*Volume*	1.46	[0.44;	2.98]	1.17	[0.65;	1.94]	1.09	[0.68;	1.58]
	Signed volume	19.11***	[12.07;	34.23]	11.30***	[6.98;	20.35]	8.40***	[5.71;	13.88]
	Level average	1.23	[0.33;	2.80]	1.07	[0.49;	2.01]	0.93	[0.50;	1.76]
	Level difference	5.74***	[3.46;	10.92]	3.40	[2.06;	5.47]	2.43	[1.48;	3.73]
	Rotation Average	1.41	[0.53;	3.71]	1.29	[0.58;	2.17]	1.27	[0.62;	2.03]
	Rotation Difference	2.17	[0.60;	3.84]	1.76	[0.70;	2.42]	1.33	[0.73;	2.16]
	Spread	1.02	[0.51;	1.86]	0.93	[0.51;	1.52]	0.97	[0.61;	1.31]
	Return	1.13	[0.34;	2.48]	1.09	[0.60;	1.87]	1.07	[0.67;	1.64]
	Volatility	1.08	[0.41;	2.03]	1.01	[0.52;	1.98]	0.99	[0.57;	1.57]
Level average	*Volume*	3.84	[1.32;	6.46]	1.95	[1.08;	3.96]	1.53	[1.01;	3.11]
	Signed volume	0.77	[0.48;	1.57]	0.89	[0.59;	1.39]	0.91	[0.60;	1.30]
	Level average	1827.61***	[966.16;	3793.46]	896.63***	[477.30;	1740.90]	595.66***	[316.30;	1111.35]
	Level difference	1.93	[0.71;	5.67]	1.71	[1.00;	3.25]	1.37	[0.94;	2.58]
	Rotation Average	70.92***	[36.29;	115.21]	33.85***	[17.09;	56.51]	21.87***	[11.49;	37.38]
	Rotation Difference	13.85***	[2.84;	45.75]	7.18***	[2.28;	23.90]	5.40***	[2.08;	16.46]
	Spread	25.36***	[15.04;	47.58]	12.29***	[6.96;	23.69]	8.93***	[4.66;	15.96]
	Return	1.38	[0.58;	3.81]	1.35	[0.73;	2.57]	1.25	[0.91;	2.05]
	Volatility	6.02	[2.52;	10.58]	3.44	[1.64;	5.41]	2.41	[1.29;	3.67]
Level difference	*Volume*	1.06	[0.46;	2.31]	0.98	[0.49;	2.45]	0.93	[0.59;	1.90]
	Signed volume	1.73	[1.03;	3.68]	1.32	[0.72;	2.55]	1.35	[0.84;	2.08]
	Level average	1.86	[0.38;	4.57]	1.62	[0.47;	3.63]	1.34	[0.70;	3.13]
	Level difference	1273.48***	[667.42;	2323.92]	624.75***	[342.65;	1145.87]	409.55***	[228.51;	755.39]
	Rotation Average	24.22	[1.36;	68.18]	12.13	[0.96;	30.79]	8.49	[0.93;	20.74]
	Rotation Difference	97.66***	[42.97;	168.73]	44.38***	[19.99;	79.79]	29.39***	[14.13;	53.18]
	Spread	0.66	[0.25;	2.08]	1.07	[0.37;	1.92]	1.02	[0.42;	1.59]
	Return	4.32	[1.55;	17.26]	3.35	[1.36;	10.59]	2.56	[1.23;	7.47]
	Volatility	1.11	[0.47;	2.63]	1.14	[0.44;	2.15]	1.14	[0.57;	1.94]

Rotation Average	*Volume*	3.35	[1.28;	9.19]	1.98	[0.86;	3.87]	1.51	[0.85;	2.84]
	Signed volume	1.01	[0.52;	1.79]	1.10	[0.48;	1.75]	0.96	[0.59;	1.42]
	Level average	3.82	[1.16;	14.11]	2.37	[1.27;	7.21]	1.80	[1.06;	4.74]
	Level difference	2.02	[1.05;	4.87]	1.57	[1.02;	2.69]	1.29	[1.04;	1.87]
	Rotation Average	193.67***	[95.09;	473.06]	98.34***	[51.60;	236.50]	67.95***	[35.09;	155.83]
	Rotation Difference	2.82	[0.59;	7.23]	1.64	[0.82;	4.10]	1.33	[0.84;	3.19]
	Spread	17.23***	[5.70;	29.49]	8.99***	[3.44;	12.97]	5.78***	[2.49;	8.65]
	Return	1.46	[0.36;	2.96]	1.33	[0.54;	2.75]	1.27	[0.70;	2.06]
	Volatility	1.67	[0.73;	5.07]	1.42	[0.89;	2.61]	1.37	[0.79;	2.10]
Rotation Difference	*Volume*	1.91	[0.87;	4.89]	1.33	[0.96;	2.25]	1.16	[0.80;	1.92]
	Signed volume	1.28	[0.47;	1.97]	1.19	[0.57;	1.95]	1.14	[0.64;	1.86]
	Level average	2.74	[0.62;	6.96]	1.74	[0.78;	3.10]	1.51	[0.81;	2.58]
	Level difference	1.95	[0.69;	5.61]	1.38	[0.57;	3.15]	1.30	[0.69;	2.18]
	Rotation Average	2.89	[0.85;	5.09]	1.69	[0.74;	2.98]	1.46	[0.92;	2.10]
	Rotation Difference	128.59***	[81.52;	294.52]	67.89***	[44.50;	152.83]	46.27***	[30.39;	99.18]
	Spread	4.58	[0.65;	11.16]	2.47	[0.76;	5.71]	1.90	[0.67;	4.23]
	Return	3.13	[1.04;	6.76]	2.42	[1.03;	3.91]	1.94	[0.99;	3.10]
	Volatility	0.98	[0.28;	3.15]	1.28	[0.54;	1.89]	1.21	[0.68;	1.74]
Spread	*Volume*	1.71	[0.54;	3.29]	1.36	[0.48;	2.21]	1.32	[0.62;	2.07]
	Signed volume	1.28	[0.35;	2.30]	1.06	[0.47;	1.69]	1.01	[0.52;	1.52]
	Level average	9.42***	[3.45;	14.06]	4.42***	[2.24;	6.58]	3.10	[1.73;	4.31]
	Level difference	1.29	[0.59;	2.79]	1.20	[0.69;	2.23]	1.21	[0.66;	1.96]
	Rotation Average	3.77	[1.84;	7.59]	2.31	[1.02;	4.33]	1.86	[0.92;	2.98]
	Rotation Difference	1.91	[1.05;	4.64]	1.58	[0.94;	2.74]	1.52	[0.81;	2.07]
	Spread	39.03***	[13.43;	71.23]	22.41***	[8.06;	41.71]	15.50***	[5.42;	29.21]
	Return	1.52	[0.31;	3.04]	1.23	[0.69;	2.32]	1.28	[0.79;	1.92]
	Volatility	13.28***	[2.89;	24.97]	6.42	[1.83;	11.61]	4.46	[1.64;	7.66]
Return	*Volume*	1.15	[0.44;	1.92]	0.87	[0.45;	1.36]	0.91	[0.63;	1.58]
	Signed volume	7.40***	[4.26;	13.13]	3.85***	[2.61;	6.86]	3.30	[1.94;	5.21]
	Level average	1.30	[0.53;	3.10]	1.12	[0.80;	2.15]	1.33	[0.80;	2.14]
	Level difference	21.60***	[7.21;	30.02]	10.99***	[4.20;	16.04]	7.90***	[3.41;	11.09]
	Rotation Average	2.50	[1.02;	4.54]	1.81	[1.11;	3.58]	1.56	[0.88;	3.34]
	Rotation Difference	6.66	[2.00;	22.18]	3.87	[1.54;	11.10]	2.79	[1.27;	7.52]
	Spread	1.12	[0.42;	3.55]	1.37	[0.75;	2.63]	1.44	[0.86;	1.98]
	Return	6.08***	[2.64;	20.28]	4.14	[1.66;	10.92]	3.25	[1.50;	7.83]
	Volatility	1.53	[0.73;	3.55]	1.83	[1.10;	2.74]	1.51	[1.01;	2.66]
Volatility	*Volume*	16.69***	[8.10;	26.54]	7.83***	[4.11;	11.96]	5.00***	[2.98;	7.75]
	Signed volume	1.03	[0.40;	2.11]	1.06	[0.55;	1.78]	1.11	[0.55;	1.58]
	Level average	30.44***	[14.85;	50.04]	12.89***	[5.52;	19.55]	8.10***	[3.34;	12.34]
	Level difference	1.03	[0.21;	3.71]	1.33	[0.62;	2.44]	1.22	[0.68;	2.23]
	Rotation Average	2.50	[0.69;	6.05]	1.64	[0.97;	3.84]	1.55	[0.85;	2.78]
	Rotation Difference	1.56	[0.40;	4.03]	1.59	[0.52;	2.31]	1.20	[0.50;	1.94]
	Spread	12.96***	[4.18;	14.66]	6.12***	[2.69;	7.28]	4.36***	[2.17;	5.58]
	Return	1.86	[0.83;	4.35]	1.50	[0.80;	3.27]	1.55	[0.99;	2.65]
	Volatility	98.90***	[60.17;	141.90]	52.03***	[34.16;	75.32]	35.77***	[22.96;	50.92]

*** indicates significance at 1% level for all the 10^{th}, 50^{th} and 90^{th} percentiles. Certain calculations derived in part from data provided by Reuters.

Table IV
Estimates of Matrix *A* in the Structural Vector Autoregressions

This table gives the estimated matrix A in the structural VAR models for 28 DJIA stocks traded on NYSE from 06/01/2006 to 07/31/2006. The model is given by $Au_t = D\varepsilon_t$, where u_t is the vector of reduced form VAR residuals, ε_t is the vector of standardized structural residuals, and D is a diagonal matrix of standard deviations for structural residuals. The percentiles are based on the samples of the corresponding element estimates across 28 stocks.

Equation ↓	Volume	Signed volume	Level average	Level difference	Rotation Average	Rotation Difference	Spread	Return	Volatility
Volume	1	0	0	0	0	0	0	0	0
Level average	0.17 [–0.01; 0.29]	0.02 [–0.06; 0.09]	1	0	0	0	0	0	0
Rotation Average	0.02 [–0.01; 0.04]	0.00 [–0.03; 0.03]	–0.28 [–0.34; –0.23]	–0.03 [–0.22; 0.23]	1	0	0	0	0
Spread	–0.05 [–0.10; –0.03]	0.00 [–0.02; 0.01]	–0.06 [–0.10; –0.01]	0.00 [–0.03; 0.01]	–0.05 [–0.11; 0.00]	–0.01 [–0.06; 0.05]	1	0	0
Volatility	–0.53 [–0.82; –0.44]	0.01 [–0.03; 0.06]	0.01 [–0.02; 0.10]	0.00 [–0.02; 0.02]	0.27 [0.17; 0.35]	–0.01 [–0.21; 0.21]	–0.35 [–0.43; –0.29]	0.01 [–0.01; 0.03]	1
Signed volume	0.01 [–0.04; 0.03]	1	0	0	0	0	0	0	0
Level difference	–0.01 [–0.04; 0.04]	–0.06 [–0.19; 0.00]	0.01 [–0.03; 0.06]	1	0	0	0	0	0
Rotation Difference	0.00 [–0.02; 0.02]	–0.02 [–0.04; 0.01]	0.03 [–0.19; 0.24]	–0.31 [–0.42; –0.22]	–0.01 [–0.15; 0.11]	1	0	0	0
Return	–0.03 [–0.21; 0.07]	–0.70 [–1.13; –0.49]	0.04 [–0.09; 0.16]	0.33 [0.03; 0.78]	–0.03 [–0.39; 0.42]	–0.47 [–0.76; 0.41]	–0.03 [–0.10; 0.11]	1	0

*The number above the bracket is the median. The numbers in the bracket are the 10^{th} percentile and the 90^{th} percentile.

Certain calculations derived in part from data provided by Reuters.

Table V
Estimates of Matrix $B=A^{-1}$ in the Structural Vector Autoregressions

This table gives the estimated matrix $B=A^{-1}$ in the structural VAR models for 28 DJIA stocks traded on NYSE from 06/01/2006 to 07/31/2006. The model is given by $Au_t = D\varepsilon_t$, where u_t is the vector of reduced form VAR residuals, ε_t is the vector of standardized structural residuals, and D is a diagonal matrix of standard deviations for structural residuals. The percentiles are based on the samples of the corresponding element estimates across 28 stocks.

Shock to →	Volume	Signed Volume	Depth Level Average	Depth Level Difference	Rotation Average	Rotation Difference	Spread	Return	Volatility
Volume	1	0	0	0	0	0	0	0	0
Level Average	–0.17 [–0.29; 0.01]	–0.02 [–0.08; 0.06]	1	0	0	0	0	0	0
Rotation Average	–0.07 [–0.13; –0.01]	–0.01 – [0.06; 0.04]	0.28 [0.22; 0.34]	0.03 [–0.23; 0.22]	1	0	0	0	0
Spread	0.05 [0.02; 0.09]	0.00 [–0.01; 0.02]	0.07 [0.04; 0.13]	0.00 [–0.01; 0.03]	0.05 [–0.01; 0.11]	0.01 [–0.05; 0.06]	1	0	0
Volatility	0.57 [0.50; 0.84]	–0.02 [–0.06; 0.02]	–0.08 [–0.19; –0.05]	0.01 [–0.03; 0.03]	–0.25 [–0.30; –0.15]	0.02 [–0.20; 0.21]	0.35 [0.29; 0.43]	–0.01 [–0.03; 0.01]	1
Signed volume	–0.01 [–0.03; 0.04]	1	0	0	0	0	0	0	0
Level Difference	0.01 [–0.04; 0.05]	0.06 [0.00; 0.19]	–0.01 [–0.06; 0.03]	1	0	0	0	0	0
Rotation Difference	0.00 [–0.05; 0.04]	0.05 [0.01; 0.10]	–0.03 [–0.20; 0.19]	0.33 [0.23; 0.43]	0.01 [–0.11; 0.15]	1	0	0	0
Return	0.02 [–0.07;0.18]	0.73 [0.49; 1.16]	–0.03 [–0.16; 0.04]	–0.18 [–0.39; –0.03]	0.03 [–0.41; 0.54]	0.47 [–0.41; 0.76]	0.03 [–0.10; 0.10]	1	0

*The number above the bracket is the median. The numbers in the bracket are the 90th percentile and the 10th percentile.

Certain calculations derived in part from data provided by Reuters.

Table VI
Forecast Error Variance Decompositions for Endogenous Variables

This table shows the forecast error variance decompositions (FEVD) obtained for the endogenous variables in the structural vector autoregressions estimated for 28 DJIA stocks traded on the NYSE from 06/01/2006 to 07/31/2006. Entry (i,j) of each panel shows the part of the variance of the ith forecast variable (in the row) attributed to the jth component of structural shock (associated with the variable in the column). Certain calculations derived in part from data provided by Reuters.

	Volume	Signed Volume	Level Average	Level Difference	Rotation Average	Rotation Difference	Spread	Return	Volatility
				Panel A: Forecast horizon = 1					
Volume	2.50	0	0	0	0	0	0	0	0
Signed Volume	0.00	2.13	0	0	0	0	0	0	0
Level Average	0.07	0.00	12.20	0	0	0	0	0	0
Level Difference	0.00	0.01	0.00	8.06	0	0	0	0	0
Rotation Average	0.01	0.00	0.98	0.01	1.62	0	0	0	0
Rotation Difference	0.00	0.00	0.01	0.86	0.00	1.71	0	0	0
Spread	0.01	0.00	0.05	0.00	0.00	0.00	2.18	0	0
Return	0.00	1.13	0.01	0.27	0.00	0.37	0.00	27.04	0
Volatility	0.81	0.00	0.08	0.00	0.10	0.00	0.27	0.00	10.33
				Panel B: Forecast horizon = 5					
Volume	2.65	0.00	0.06	0.00	0.01	0.00	0.02	0.00	0.00
Signed Volume	0.00	2.14	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Level Average	0.15	0.00	16.44	0.00	0.54	0.00	0.21	0.00	0.04
Level Difference	0.00	0.01	0.00	9.55	0.00	0.34	0.00	0.01	0.00
Rotation Average	0.02	0.00	1.10	0.01	1.71	0.00	0.02	0.00	0.00
Rotation Difference	0.00	0.01	0.01	0.90	0.00	1.78	0.00	0.00	0.00
Spread	0.01	0.00	0.06	0.00	0.00	0.00	2.20	0.00	0.01
Return	0.00	1.22	0.01	0.88	0.00	0.41	0.00	27.07	0.00
Volatility	1.00	0.00	0.16	0.00	0.11	0.00	0.34	0.00	10.62
				Panel C: Forecast horizon = 15					
Volume	2.81	0.00	0.07	0.00	0.01	0.00	0.02	0.00	0.00
Signed Volume	0.00	2.14	0.00	0.01	0.00	0.00	0.00	0.00	0.00
Level Average	0.21	0.00	19.00	0.00	1.16	0.01	0.37	0.00	0.16
Level Difference	0.00	0.02	0.01	10.29	0.00	0.63	0.00	0.05	0.00
Rotation Average	0.02	0.00	1.16	0.01	1.75	0.00	0.04	0.00	0.00
Rotation Difference	0.00	0.01	0.01	0.91	0.00	1.81	0.00	0.00	0.00
Spread	0.01	0.00	0.07	0.00	0.00	0.00	2.22	0.00	0.02
Return	0.00	1.22	0.01	0.88	0.00	0.41	0.00	27.07	0.00
Volatility	1.19	0.00	0.21	0.00	0.12	0.00	0.38	0.00	10.83

Table VII
Comparison of Empirical Results to Recent Literature

This table summarizes the findings by the recent literature on the interaction between the state of limit order book and market situation and the compares to the results in the present paper. Up arrow means a positive shock or response and down arrow means a negative response. Check means the result is confirmed by the present paper.

Shock in Previous Moment	Response in Current Moment	Literature Reference	Present Paper
Volume ↑	Depth ↓	Lee, Mucklow and Ready (1993)	
	Spread ↑	Lee, Mucklow and Ready (1993)	√
	Volatility ↑	Jones et al (1994), Huang and Masulis (2003), Xu et al. (2006)	√
Depth ↑	Volume ↑	Biais et al (1995)	√
	Spread ↓	Copeland and Galai (1983), Lee, Mucklow and Ready (1993) Ellul et al (2007)	√
	Volatility ↓	Coppejans et al (2001), Ahn et al (2001)	√
	Aggressiveness ↑	Ranaldo (2004)	√
Aggressiveness ↑	Volume ↑		√
	Depth ↓		√
	Spread ↓		√
Spread ↑	Volume ↓		√
	Depth ↓	Lee, Mucklow and Ready (1993)	√
	Aggressiveness ↑	Ranaldo (2004)	√
	Aggressiveness ↓	Foucault (1999), Handa et al (2003), Chan (2005), Pascual and Veredas (2006a)	√
Volatility ↑	Depth ↓	Coppejans et al (2001) Goldstein and Kavajecz (2004)	√
	Depth ↑	Ahn et al (2001) Chan 2005	
	Spread ↑	Copeland and Galai (1983)	√
	Spread ↓	Bae et al (2003)	
	Aggressiveness ↑	Ranaldo (2004)	
	Aggressiveness ↓		√
Order Imbalance ↑	Return ↑		√
	Return ↓	Chordia and Subrahmanyam (2004) find lagged order imbalance is NEGATIVELY related to return when controlling for contemporaneous imbalance, in DIALY basis.	
Depth Imbalance ↑	Order Imbalance ↓	Hall and Hautsch (2007) "take liquidity when supply is high"	√
	Return ↑	Cao et al (2003) Harris and Panchapagesan (2005)	√
	Aggressiveness Imbalance ↑	Cao et al (2003) Ranaldo (2004), Pascual and Veredas (2006a)	√
Return ↑	Aggressiveness Imbalance ↑	Chan (2005)	√
Aggressiveness Imbalance ↑	Depth Imbalance ↓		√

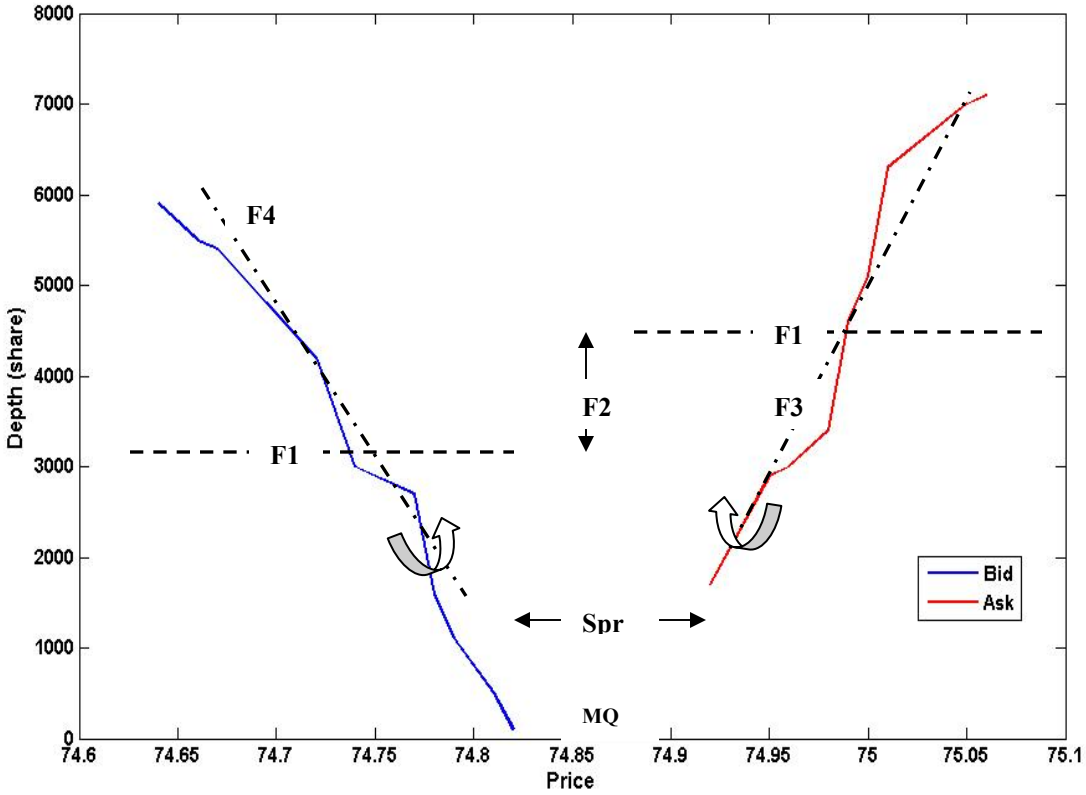


Figure 1. Screenshot of the limit order book for IBM at 10:30:05 AM on July 13, 2006. "*MQ*" is the mid quote, anchoring the position of this snapshot. "*Spr*" is the spread, anchoring the relative distance between the bid and ask cumulative depth curves. "*F*1" ("*F*2") is the average depth level (difference of depth levels) provided by limit order traders on the bid and ask sides of the book, anchoring the height level of the cumulative depth curves. "*F*3" and "*F*4" are the rotation factors of bid and ask curves, respectively, anchoring the extent to which the cumulative depth curves rotate around pivot points in the middle.

Certain calculations derived in part from data provided by Reuters.

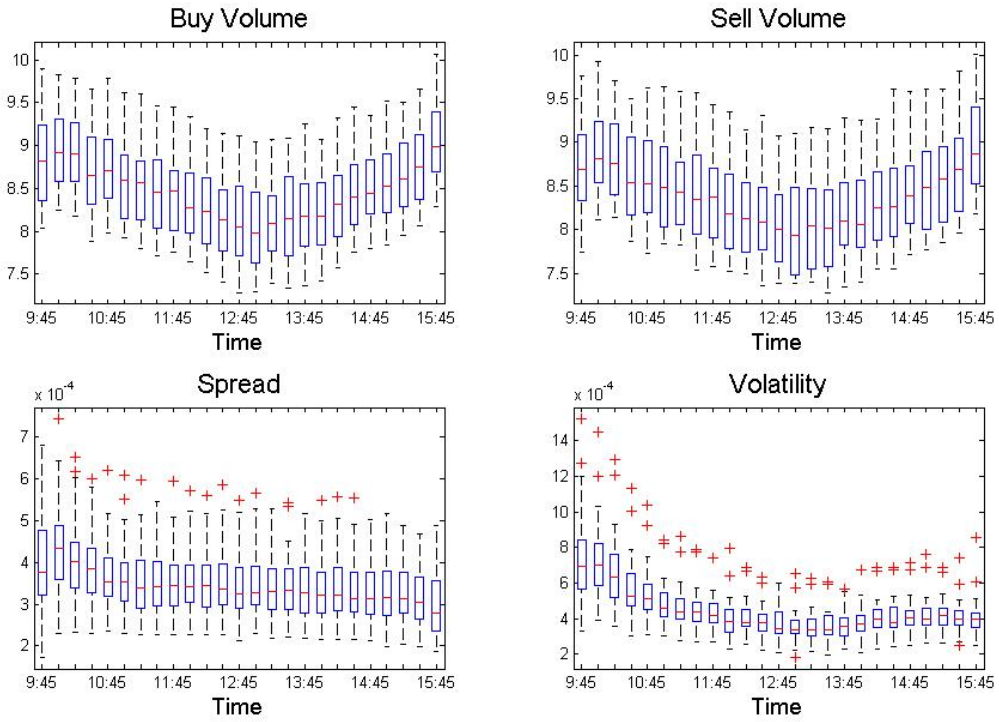


Figure 2. *Panel A*: Box-and-whisker plots of intraday (diurnal) patterns of market variables for 28 DJIA stocks traded on the NYSE. "Buy (Sell) volume" is the total number of shares for buyer-initiated (seller-initiated) trades during the last minute. "Spread" is the ratio of best ask price to best bid price at the end of each minute. "Volatility" is the ratio of the maximum and minimum mid-quotes during the last minute. Each of these variables is averaged within a 15-minute bin starting from 9:45 am to 3:45 pm EST for each stock and each day. The medians are calculated across 42 trading days for each time-of-the-day bin and each stock. All values are expressed in logarithmic terms.

Certain calculations derived in part from data provided by Reuters.

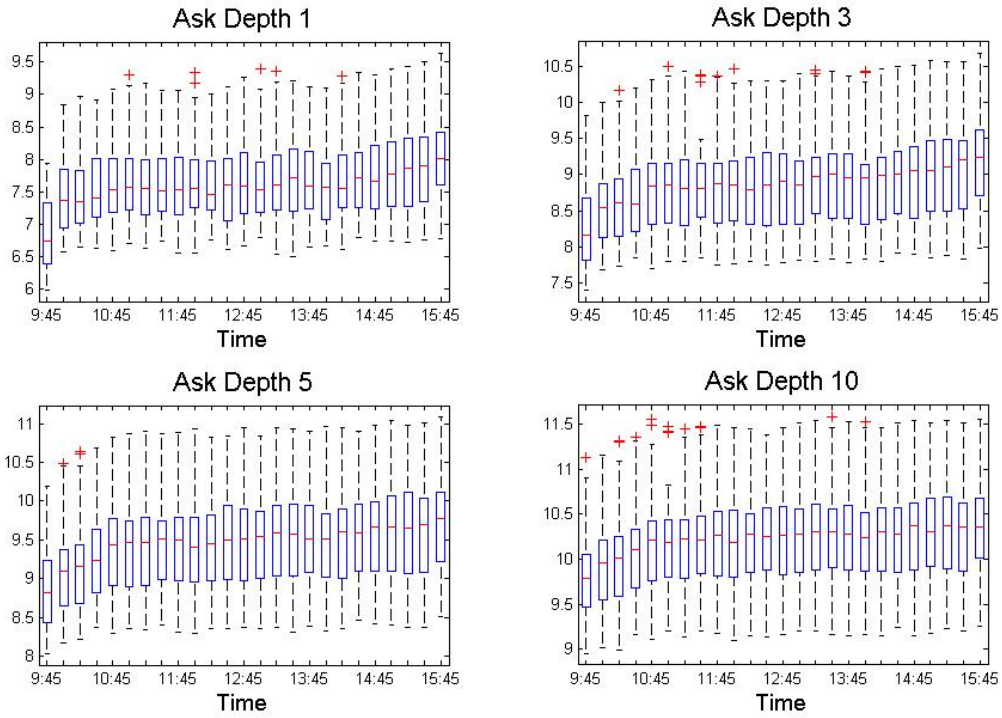


Figure 2. *Panel B*: Box-and-whisker plots of the intraday pattern of depth (accumulated up to the first best price, the third best price, the fifth best price, and the tenth best price) on the ask side of the limit order book for 28 DJIA stocks traded on the NYSE. The cumulative depths are first averaged within a 15-minute bin starting from 9:45 am to 3:45 pm EST for each stock and each day. The median values are calculated across 42 trading days for each time-of-the-day bin and each stock. All values of depth are expressed as logarithms of the number of shares.

Certain calculations derived in part from data provided by Reuters.

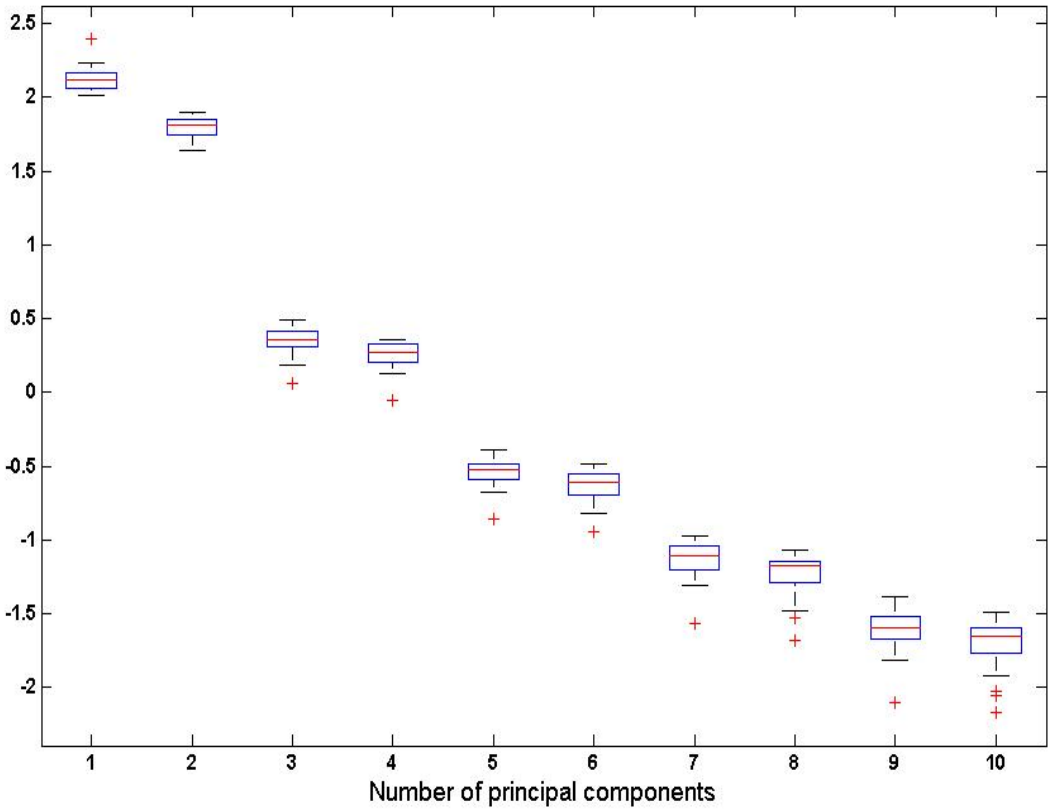


Figure 3: Box-and-whisker scree plot of the principal components obtained for the cumulative limit order book depths for 28 DJIA stocks traded on the NYSE. The vertical axis shows the logarithm of eigenvalues associated with the first ten principal components extracted for each stock. Only the first ten eigenvalues are shown.

Certain calculations derived in part from data provided by Reuters.

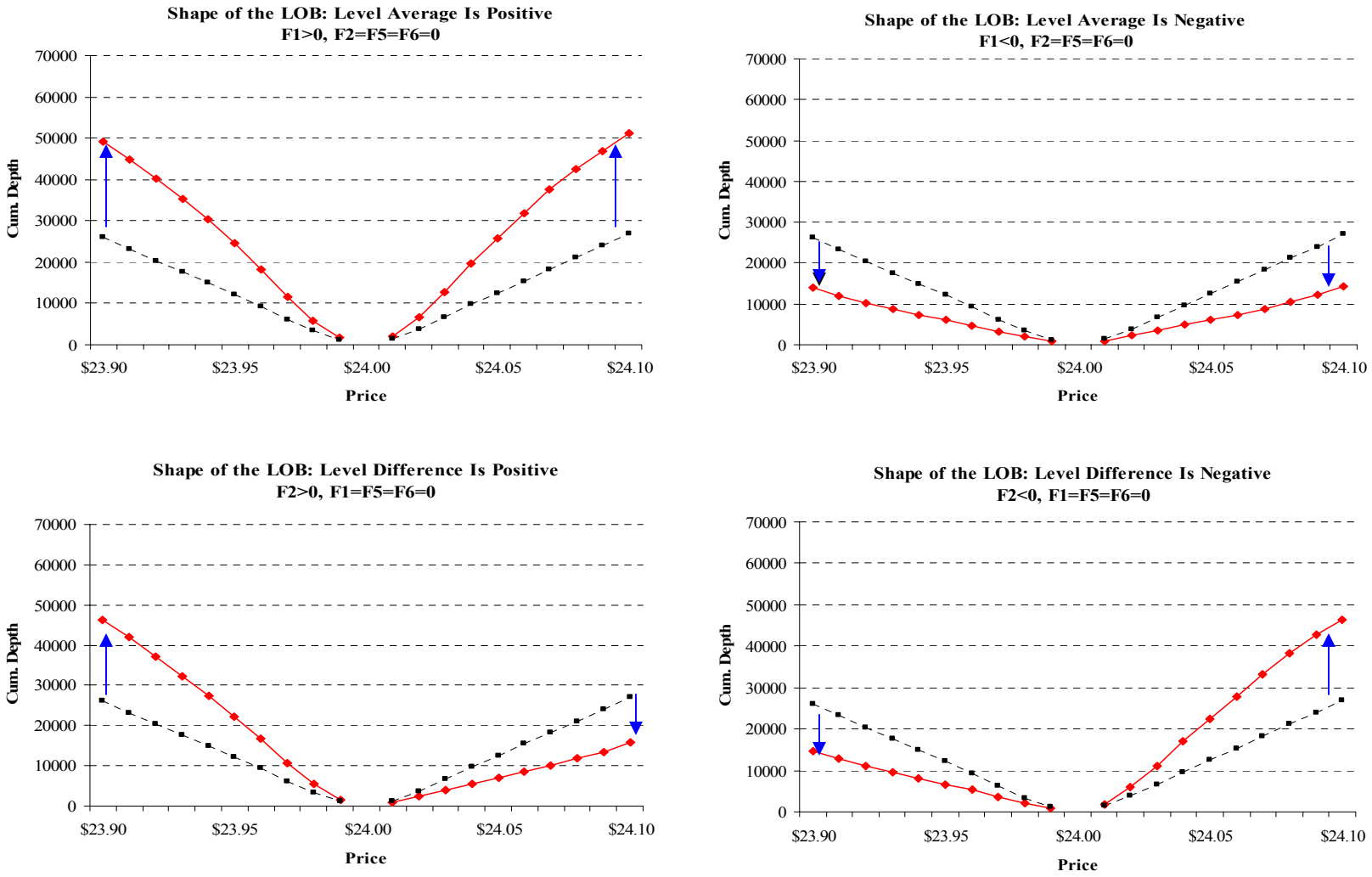


Figure 4. Illustration of the principal components extracted from the cumulative limit order book depth levels on the bid and ask sides of a prototypical limit order book. Certain calculations derived in part from data provided by Reuters.

Panel A: Depth Level Factors *F*1 and *F*2.

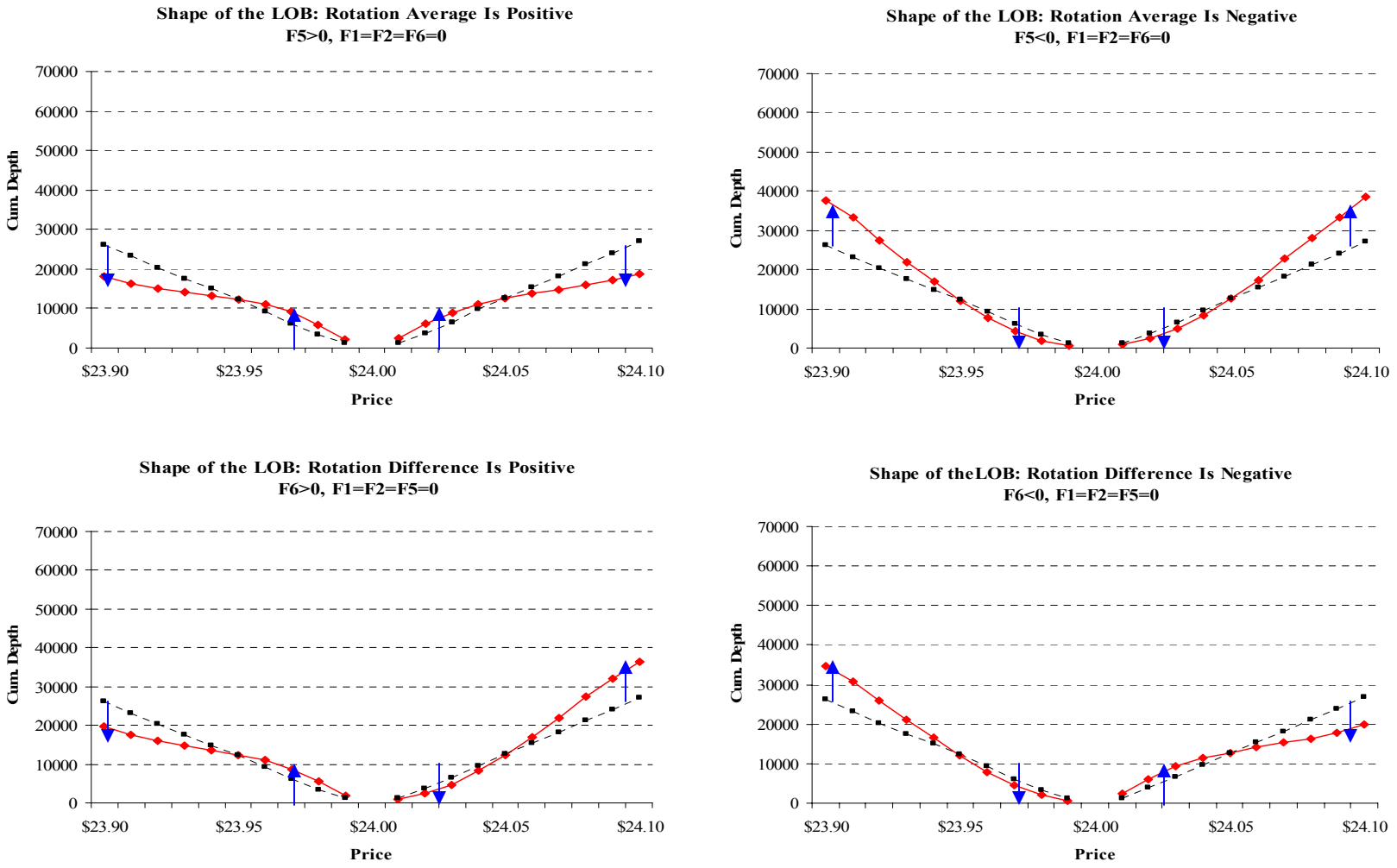


Figure 4. Illustration of the principal components extracted from the cumulative limit order book depth levels on the bid and ask sides of a prototypical limit order book (*continued*). Certain calculations derived in part from data provided by Reuters.

Panel B: Rotation Factors *F*5 and *F*6.

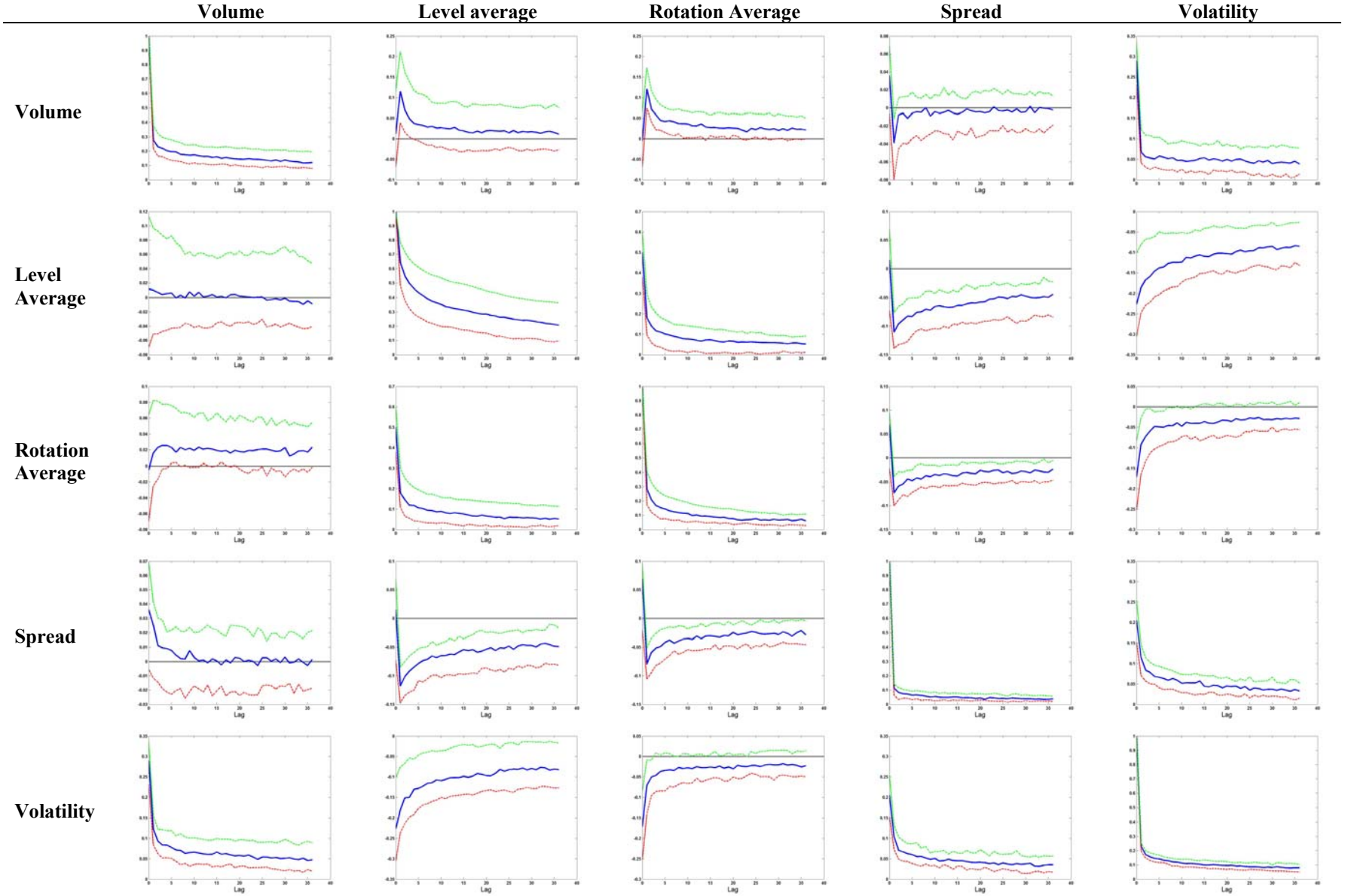


Figure 5. *Panel A*: Autocorrelograms and cross-correlograms of logged diurnally adjusted volume, depth level average, rotation average, spread, and volatility. The upper, middle, and lower lines are, respectively, the 90th, 50th, and 10th percentiles based on the sample of auto- and cross-correlograms for 28 stocks. The cross-correlograms between X_{it} and X_{jt} at horizon h are defined as Corr(X_{it}, $X_{j,\,t+h}$), where $X_{i\,t}$ in the column is the input variable and $X_{j,t+h}$ in the row is the response variable.

Certain calculations derived in part from data provided by Reuters.

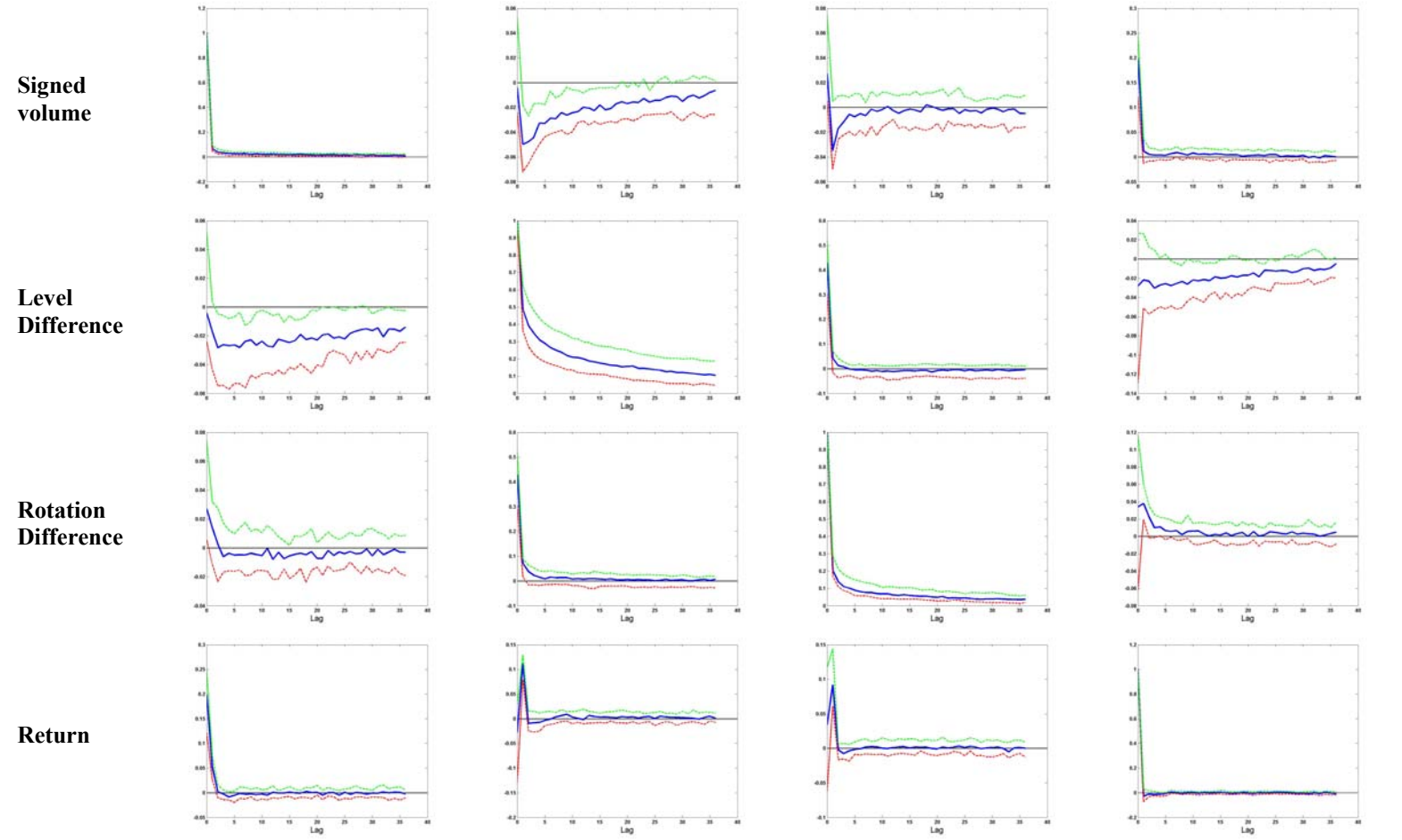


Figure 5. *Panel B*: Autocorrelograms and cross-correlograms of logged diurnally adjusted signed volume, depth level difference, rotation difference, and return. The upper, middle, and lower lines are, respectively, the 90th, 50th, and 10th percentiles based on the sample of auto- and cross-correlograms for 28 stocks. The cross-correlograms between X_{it} and X_{jt} at horizon h are defined as Corr(X_{it}, $X_{j,\,t+h}$), where $X_{i\,t}$ in the column is the input variable and $X_{j,t+h}$ in the row is the response variable.

Certain calculations derived in part from data provided by Reuters.

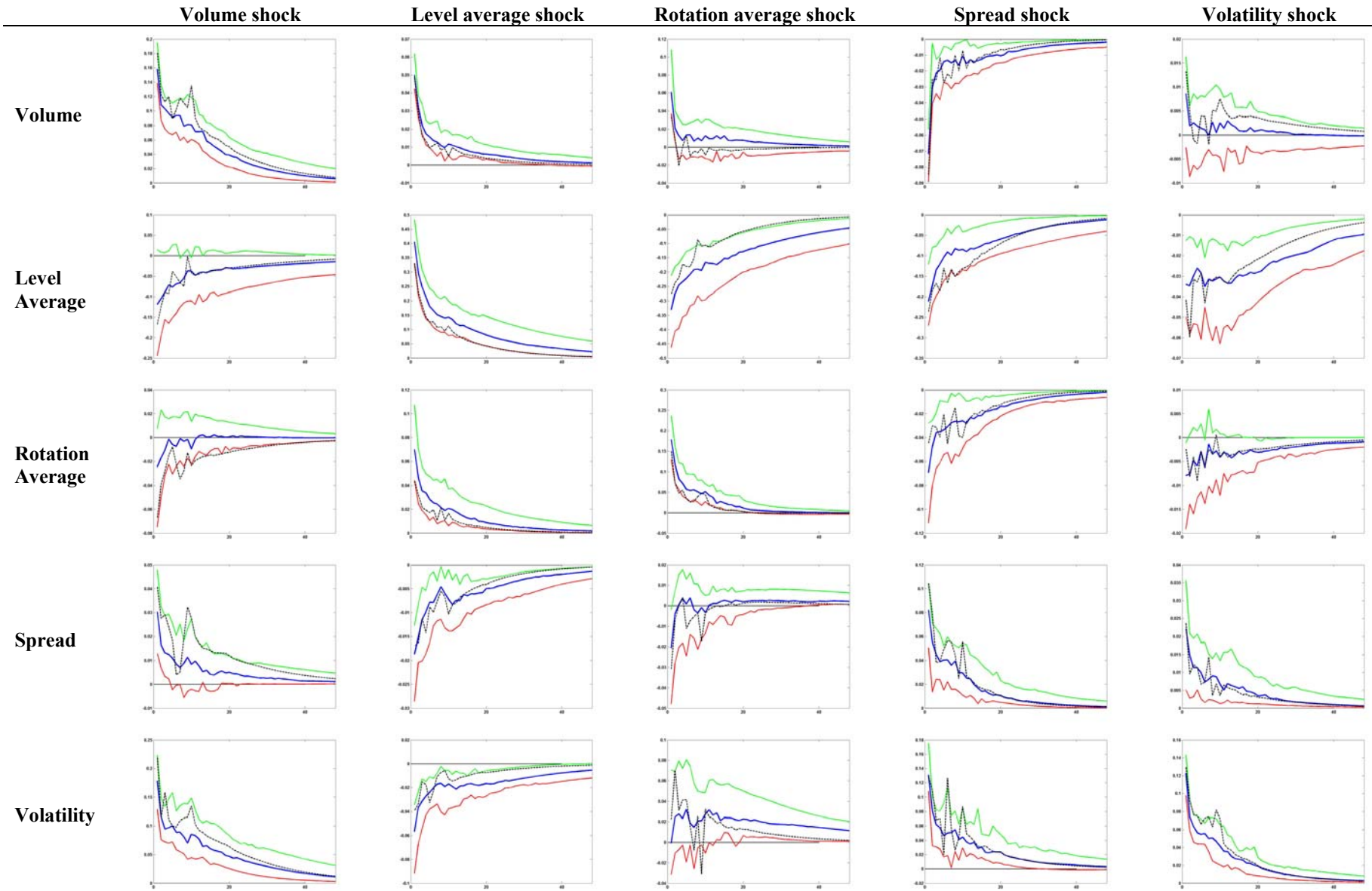


Figure 6. *Panel A*: Impulse response functions of "variational" endogenous variables to the structural VAR shocks estimated for each of 28 DJIA stocks traded on the NYSE from 06/01/2006 to 07/31/2006. The upper, middle and lower lines are, respectively, the 90th, 50th, and 10th percentile based on the sample of estimated response functions for 28 stocks. The dashed line is an example of the estimated impulse response functions for IBM as a representative stock. The impulse variables associated with shocks are listed in the columns. The response variables are shown in the rows.

Certain calculations derived in part from data provided by Reuters.

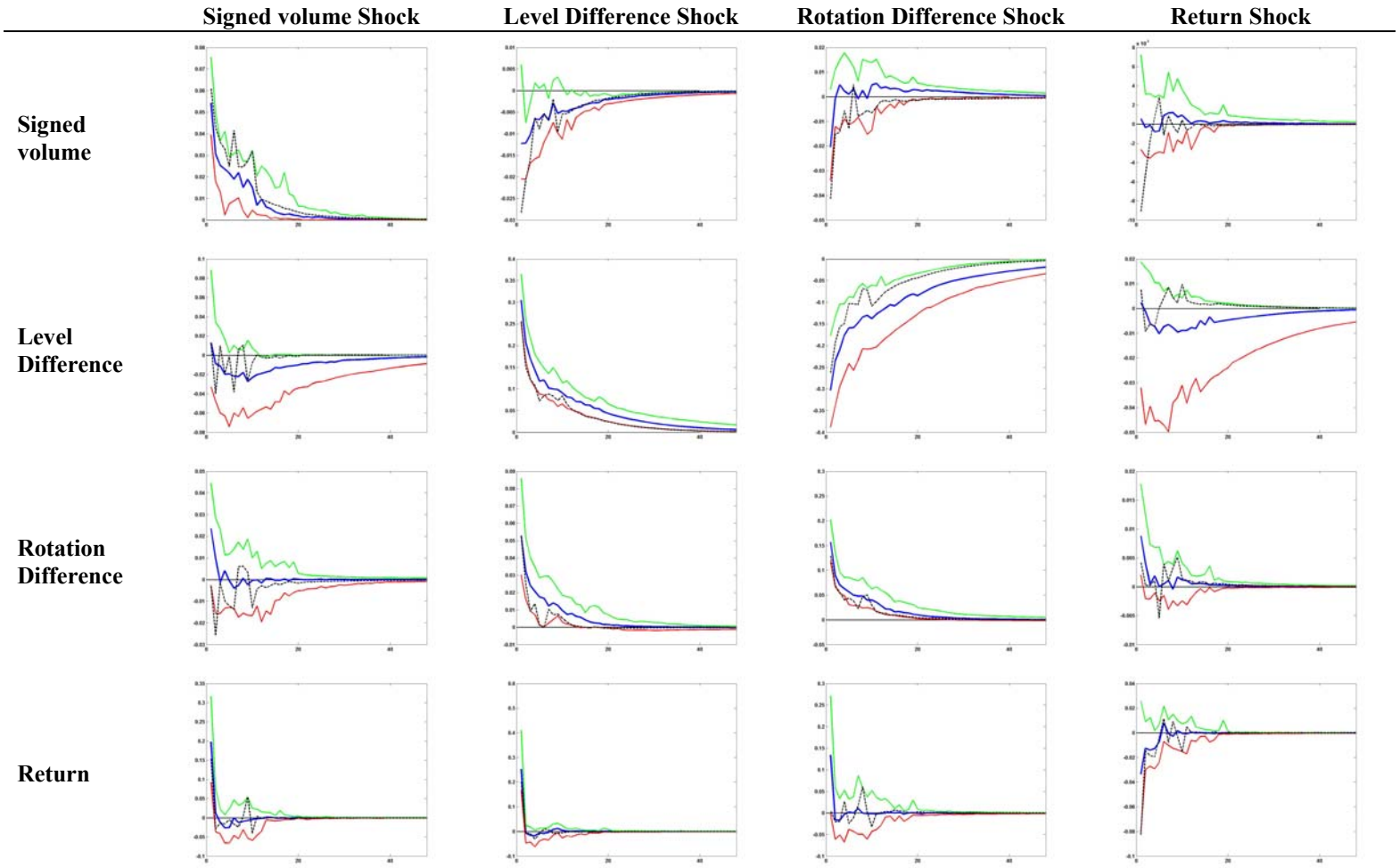


Figure 6. Panel *B*: Impulse response functions of "directional" endogenous variables to the structural VAR shocks estimated for each of 28 DJIA stocks traded on the NYSE from 06/01/2006 to 07/31/2006. The upper, middle and lower lines are, respectively, the 90th, 50th, and 10th percentile based on the sample of estimated response functions for 28 stocks. The dashed line is an example of the estimated impulse response functions for IBM as a representative stock. The impulse variables associated with shocks are listed in the columns. The response variables are shown in the rows.

Certain calculations derived in part from data provided by Reuters.

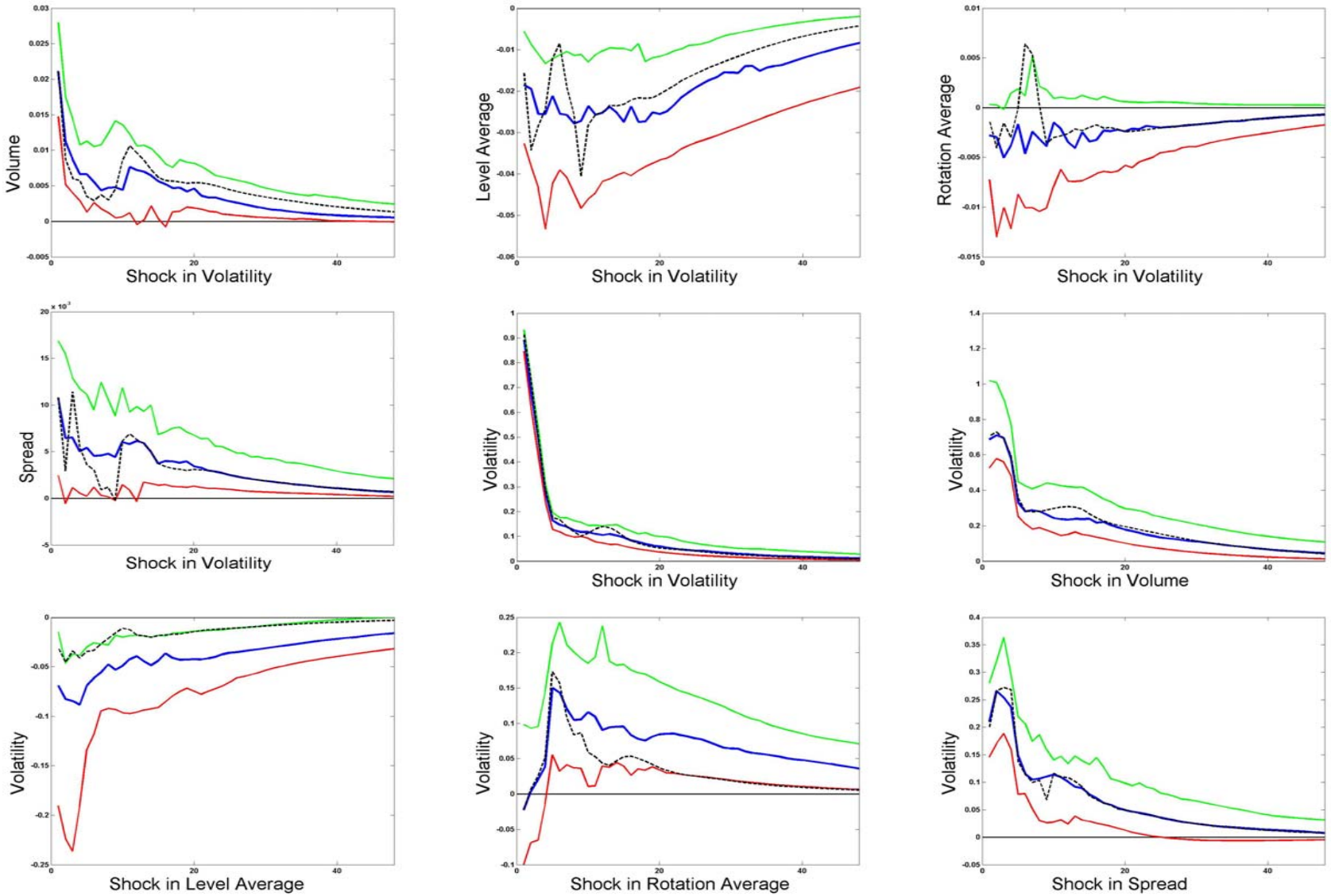


Figure 7. Impulse response functions of variables to the structural VAR shocks between "variational" endogenous variables and 5-minute volatility estimated for each of 28 DJIA stocks traded on the NYSE from 06/01/2006 to 07/31/2006. The volatility is calculated as the range of mid quote in the previous 5 minutes. The upper, middle and lower lines are, respectively, the 90th, 50th, and 10th percentile based on the sample of estimated response functions for 28 stocks. The dashed line is an example of the estimated impulse response functions for IBM as a representative stock.

Certain calculations derived in part from data provided by Reuters.