JOURNAL OF FINANCIAL AND QUANTITATIVE ANALYSIS Vol. 49, No. 1, Feb. 2014, pp. 33–49

doi:10.1017/S0022109014000106

Volume and Volatility in a Common-Factor Mixture of Distributions Model

Xiaojun He and Raja Velu*

Abstract

This paper develops a multi-asset mixture distribution hypothesis model to investigate commonality in stock returns and trading volume. The model makes two main predictions: First, the factor structures of returns and trading volume are independent although they stem from the same valuation fundamentals and jointly depend on a latent information flow; second, cross-sectional positive volatility-volume relations arise solely from the dynamic features of the information flow. Empirical analyses at the market level support these predictions. Furthermore, the results indicate that removing the information flow significantly reduces the return volatility persistence and the extent of the reduction exhibits a size pattern.

I. Introduction

The objective of this study is to provide theory as well as empirical evidence on commonality in stock returns, measured by price changes, and trading volume. To that end, we develop a common-factor mixture distribution hypothesis (MDH) model to explore common cross-firm variations in stock returns and trading volume.

Our study is inspired by other research in modern portfolio theory and market microstructure theory. Although a sizable body of research investigates return-volume relations, studies of the cross-sectional behavior of stock returns have long been separate from work on equity trading volume.[1] Hasbrouck and Seppi (2001)

*He, xihe@yu.edu, Syms School of Business, Yeshiva University, 500 W 185th St, New York, NY 10033; Velu, rpvelu@syr.edu, Whitman School of Management, Syracuse University, 721 University Ave, Syracuse, NY 13244. This paper is partially drawn from He's dissertation. He is grateful to Chihwa Kao, Kasing Man, Chunchi Wu, and especially Philip Griffin, her dissertation committee chair, for their helpful comments. We benefited from the constructive comments of an anonymous referee, Hendrik Bessembinder (the editor), participants at the 2004 American Finance Association (AFA) and 2004 Financial Management Association (FMA), and especially Tom Smith. He is also grateful for the generous financial support provided by the Brethren Operations Management Institute, IBM Research Fund, Entrepreneurship and Emerging Enterprises Research Fund, and Snyder Innovation Management Center of Whitman School of Management at Syracuse University. All errors are our own responsibility.

[1]Gallant, Rossi, and Tauchen (1992), (1993) provide a comprehensive investigation on price-volume relations. For studies on stock returns, see, for example, Markowitz (1952), Ross (1978),

suggest that cross-firm commonalities may arise from the factor structure of asset valuation fundamentals and thus conduct an empirical study to examine joint return-volume cross-sectional patterns.

These studies motivate our construction of a theoretical model to unify these lines of research. The model is a multi-asset version of the standard MDH model. Following Clark (1973) and Epps and Epps (1976), Tauchen and Pitts (1983) derive a dynamic heterogeneous-information equilibrium model of price changes and trading volume for a single asset in the context of a large economy with perfect competition. In their so-called MDH model, price changes and trading volume are equilibrium outcomes of information impact and are jointly related to an unobservable dynamic information flow variable. However, many studies show evidence unfavorable to the standard MDH model (e.g., see Heimstra and Jones (1994), Lamoureux and Lastrapes (1994), Richardson and Smith (1994), and Andersen (1996), among others). We conjecture that the standard MDH model is not wholly successful, probably because it treats securities in multi-asset financial markets as isolated from each other. To investigate possible interactions among securities, we generalize the standard MDH model by separating a cross-asset component arising from the impact of announcements with common effects (e.g., monetary policies, tax regimes, political events) from the variance-component decomposition of Tauchen and Pitts. We call the model with distinct asset interactions a common-factor MDH model.

The model retains all the properties of the standard MDH model for a single asset. In addition, two cross-asset implications emerge from the model. First, the return factor structure results from the expected factor structure of valuation fundamentals, while the volume factor structure stems from the unexpected factor structure of valuation fundamentals. Therefore, the return factor structure is independent of the factor structure of trading volume. Recently, Bernhardt and Taub (2008) report developing a static equilibrium model in line with Admati (1985) and Caballe and Krishnan (1994), in which investors' order flows also have a forecast error structure. Their model predicts that the correlation structures of prices and order flows are driven separately by the correlation structures of asset value fundamentals and liquidity trade, respectively. Because our model does not include liquidity trading, we view their prediction as complementary. Second, the return variance of each stock is positively related not only to its own volume but also to the volume of other stocks. The cross-sectional positive relations result solely from the dynamic features of the latent information flow variable.

For empirical analysis, we use half-hour intraday data for the stocks in the Dow Jones Industrial Average (DJIA) over a sample period from April through June 2007. The results are generally consistent with our predictions. The Pearson correlation test result suggests the existence of factor structures in both returns and trading volume because all the sample stocks show significant return-return and volume-volume correlations across firms. The estimation results indicate that factor structures are indeed present in our sample. However, only 13% of return-volume correlations are statistically significant, implying independence between

and Brenner, Pasquariello, and Subrahmanyam (2009). For work on trading volume, see, for example, Lo and Wang (2000) and Tookes (2008).

the factor structures. The canonical redundancy results further suggest that return and volume canonical variables are not good overall predictors of each other.

The common-factor MDH model, like the standard MDH model, imposes an explicit relation between the latent information flow and stock returns, and it is well-known that returns exhibit autoregressive conditional heteroskedasticity (ARCH) (generalized ARCH (GARCH)) dynamics. To examine these aspects, we employ the conditional moment method in Lamoureux and Lastrapes (1994) to extract the underlying information flow, and then we compare the dynamic characteristics of return series before and after conditioning it on the extracted information flow.[2] We find that 13 stocks in our sample display ARCH (GARCH) patterns, and these patterns all become statistically insignificant after we control for the impact of the information flow. This result is consistent with the model specification.

To investigate our model's market-wide applicability, we also apply the model to daily data for the Standard & Poor's (S&P) Composite 1500 stocks from Jan. 1, 2004, through Dec. 31, 2006. The S&P Composite 1500 is formed by the S&P LargeCap 500, S&P MidCap 400, and S&P SmallCap 600. The daily data are from the Center for Research in Security Prices (CRSP) database. The results are similar to those of the intraday Dow Jones stocks. Moreover, the market-wide analysis on volatility persistence reduction indicates that our model performs better for stocks with large market capitalization and small trading volume.

The role of common cross-equity variation in short-term trade related variables is of interest in financial economics. The factors that influence the prices, order flows, and liquidity are likely to be common among stocks that are exposed to the same risk characteristics. Basic studies that explore the commonality are important both for institutional trading practices such as portfolio rebalancing and, in general, for a better understanding of microstructure theory. Focusing on individual stocks in isolation will not help much in this endeavor. We provide a model-based approach to account for the common variation in the returns and volume of multiple stocks. If return of a stock is found to be related to other stocks' returns and volume, this has implications for pricing of the stock. Empirical results presented in this paper will also aid regulators, exchanges, and other participants in improving market design, especially in episodes of domestic and international financial turmoil. On the other hand, our model is unable to differentiate the informational effect from the impact of uninformed traders' behavior on the cross-equity variation. Our empirical results indicate that there is variation in stock returns that cannot be fully explained by the information flow. This may be explored in future research.

The rest of the paper is organized as follows: Section II describes the multi-asset economy and trading model in more detail. We also analyze the properties of the model in this section. Section III presents data analysis. Section IV provides a brief summary.

[2]The Lamoureux and Lastrapes (1994) extracting method is based upon the property of the standard MDH model that the level of volume and squared returns are influenced by the same latent information flow variable, as well as the property that returns and volume are independent.

II. Common-Factor MDH Model

We develop an M-asset generalization of the single-security MDH model in Tauchen and Pitts (1983). The economy contains J risk-averse speculators, who choose their portfolio positions to maximize their expected constant absolute risk aversion (CARA) utility function. The arrival of a set of information, including asset-specific and common information, causes the movement from one equilibrium phase to the next. At each equilibrium phase i, it is well-known that each trader j uses the information to formulate his or her investment decisions and that his or her optimal demand function is as follows:

$$\mathbf{Q}_{ij} = c\left(\mathbf{P}_{ij}^{*} - \mathbf{P}_{i}\right), \tag{1}$$

where $\mathbf{Q}_{ij} = (Q_{ij1}, Q_{ij2}, \ldots, Q_{ijM})$ is a vector of trader j's demand; c is a constant coefficient of absolute risk aversion; $\mathbf{P}_{ij}^{*} = \left(P_{ij1}^{*}, P_{ij2}^{*}, \ldots, P_{ijM}^{*}\right)$ is a vector of trader j's forecasts, based upon trader j's interpretation of the information, on assets' future values; and $\mathbf{P}_{i} = (P_{i1}, P_{i2}, \ldots, P_{iM})$ is a vector of market prices.

Under the equilibrium condition that markets clear, the market price is simply an average of all investors' forecasts:

$$\mathbf{P}_{i} = \frac{1}{J}\sum_{j=1}^{J}\mathbf{P}_{ij}^{*}. \tag{2}$$

From equations (1) and (2), we can obtain the vectors of price changes and aggregate trading volume between two consecutive equilibrium phases as follows:

$$\begin{aligned} \Delta\mathbf{P}_{i} &= \frac{1}{J}\sum_{j=1}^{J}\Delta\mathbf{P}_{ij}^{*}, \\ \mathbf{V}_{i} &= \frac{c}{2}\sum_{j=1}^{J}\left|\Delta\mathbf{P}_{ij}^{*} - \Delta\mathbf{P}_{i}\right|, \end{aligned} \tag{3}$$

where $\Delta\mathbf{P}_{ij}^{*} = \mathbf{P}_{ij}^{*} - \mathbf{P}_{i-1,j}^{*}$ is a vector of trader j's incremental reservation prices and $\Delta\mathbf{P}_{i} = \mathbf{P}_{i} - \mathbf{P}_{i-1}$ is a vector of the changes in market prices.

Tauchen and Pitts (1983) report conducting a variance-component analysis on model (3) for the case of one asset. They decompose the change in trader j's reservation price into two components: one component common to all investors and the other specific to agent j. Because our economy setting involves more assets, to distinguish the information impact across assets from its impact on a specific asset, we dissect each of their components further into an idiosyncratic component and a common component. We now have a four-component model as follows:

$$\Delta\mathbf{P}_{ij}^{*} = \boldsymbol{\upsilon}_{i}^{I} + \boldsymbol{\Psi}_{ij}^{I} + \boldsymbol{\upsilon}_{i}^{C} + \boldsymbol{\Psi}_{ij}^{C}. \tag{4}$$

The first two terms in model (4) refer to the asset-specific components. The component $\boldsymbol{\upsilon}_{i}^{I}$ is common to all active speculators of a specific stock, and the component $\boldsymbol{\Psi}_{ij}^{I}$ represents heterogeneous beliefs held by investors of that stock.

Such belief heterogeneity may arise from the agents' use of diverse sources to learn about the valuation fundamentals affecting asset values. Similarly, the cross-asset components, $\boldsymbol{\upsilon}_i^C$ and $\boldsymbol{\Psi}_{ij}^C$, represent each trader's common and diverse responses to the cross-asset information impact, respectively. In our multi-asset economy, we can merge the first two terms into one idiosyncratic vector, denoted by $\boldsymbol{\zeta}_{ij} = (\zeta_{ij1}, \zeta_{ij2}, \ldots, \zeta_{ijM})$. Assume that this idiosyncratic vector is independent for all i and j, and it follows a multivariate normal distribution with mean vector $\mathbf{0}$ and a diagonal covariance matrix Σ_ζ. We can also meld the latter two components into one common-factor vector, denoted by $\boldsymbol{\xi}_{ij} = (\xi_{ij1}, \xi_{ij2}, \ldots, \xi_{ijM})$. Assume that it follows a multinormal distribution with mean vector $\mathbf{0}$ and a nondiagonal symmetric positive definite covariance matrix $\Sigma_{\boldsymbol{\xi}}$.[3] We further assume that the common-factor vector $\boldsymbol{\xi}_{ij}$ is independent of the idiosyncratic vector $\boldsymbol{\zeta}_{ij}$ and is mutually independent for all i and j.

Applying model (4) to model (3), we have

$$
\begin{aligned}
\Delta \mathbf{P}_i &= \bar{\boldsymbol{\zeta}}_i + \bar{\boldsymbol{\xi}}_i, \\
\mathbf{V}_i &= \frac{c}{2} \sum_{j=1}^{J} \left| (\boldsymbol{\zeta}_{ij} + \boldsymbol{\xi}_{ij}) - (\bar{\boldsymbol{\zeta}}_i + \bar{\boldsymbol{\xi}}_i) \right|,
\end{aligned} \tag{5}
$$

where $\bar{\boldsymbol{\zeta}}_i = (1/J)\sum_{j=1}^{J} \boldsymbol{\zeta}_{ij}$, and $\bar{\boldsymbol{\xi}}_i = (1/J)\sum_{j=1}^{J} \boldsymbol{\xi}_{ij}$.

The distributional assumptions readily imply that i) the price change $\Delta \mathbf{P}_i$ is an M-variate Gaussian with mean vector $\mathbf{0}$ and a nondiagonal covariance matrix $\Sigma_{\Delta \mathbf{P}_i} = (1/J)(\Sigma_\zeta + \Sigma_\xi)$; ii) the price change $\Delta \mathbf{P}_i$ and trading volume $\mathbf{V}_i$ are stochastically independent because the component $\bar{\zeta}_i + \bar{\xi}_i$ is orthogonal to its deviation from the mean, which is $(\boldsymbol{\zeta}_{ij} + \boldsymbol{\xi}_{ij}) - (\bar{\boldsymbol{\zeta}}_i + \bar{\boldsymbol{\xi}}_i)$; and iii) because the volume $\mathbf{V}_i$ is a folded M-variate Gaussian, for high J, it is asymptotically and normally distributed with a mean vector $\boldsymbol{\mu}$ and a nondiagonal covariance matrix $\boldsymbol{\Sigma}_{\mathbf{V}_i}$:

$$
\begin{aligned}
\boldsymbol{\mu}_i &= \frac{c}{2} \left(\left| \Sigma_\zeta + \Sigma_\xi \right|^{½} \right) \sqrt{\frac{2}{\pi}} \left(\sqrt{\frac{J-1}{J}} \right) J, \\
\boldsymbol{\Sigma}_{\mathbf{V}_i} &= \left(\frac{c}{2} \right)^2 (\Sigma_\zeta + \Sigma_\xi) \left(1 - \frac{2}{\pi} \right) J + o(J),
\end{aligned} \tag{6}
$$

where $|\Sigma_\zeta + \Sigma_\xi|$ is the determinant of $\Sigma_\zeta + \Sigma_\xi$.

When we aggregate the within-day price changes and trading volume and standardize the notation, we have the following:

$$
\begin{aligned}
\Delta \mathbf{P} &= \boldsymbol{\alpha} \mathbf{z}_1 \sqrt{I} + \boldsymbol{\eta} \sqrt{I}, \\
\mathbf{V} &= \boldsymbol{\mu} I + \boldsymbol{\beta} \mathbf{z}_2 \sqrt{I} + \boldsymbol{\varepsilon} \sqrt{I},
\end{aligned} \tag{7}
$$

where I is the number of the within-day information flow; systematic macroeconomic factor variables $\mathbf{z}_1$ and $\mathbf{z}_2$ are standard K-variate Gaussians; firm-specific risk variables $\boldsymbol{\eta}$ and $\boldsymbol{\varepsilon}$ are M-variate normally distributed with mean vectors $\mathbf{0}$ and

[3]For simplicity, we assume that although agents may disagree about how to interpret the information affecting asset values, their forecasts have the same precision expressed by $\Sigma_\zeta + \Sigma_\xi$.

diagonal variance matrices $\boldsymbol{\Sigma}_{\boldsymbol{\eta}}$ and $\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}$; and $\mathbf{z}_1$, $\mathbf{z}_2$, $\boldsymbol{\eta}$, $\boldsymbol{\varepsilon}$ as well as I are all mutually independent.

From equations (6) and (7), the following equalities hold:

$$\boldsymbol{\alpha}\boldsymbol{\alpha}^{\mathrm{T}} = \frac{1}{J}\boldsymbol{\Sigma}_{\boldsymbol{\xi}}, \qquad \boldsymbol{\Sigma}_{\boldsymbol{\eta}} = \frac{1}{J}\Sigma_{\varsigma},$$

and,

$$\boldsymbol{\beta}\boldsymbol{\beta}^{\mathrm{T}} = \left(\frac{c}{2}\right)^2\left(1-\frac{2}{\pi}\right)J\boldsymbol{\Sigma}_{\boldsymbol{\xi}}, \quad \boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}} = \left(\frac{c}{2}\right)^2\left(1-\frac{2}{\pi}\right)J\Sigma_{\varsigma}.$$

We call equation (7) a common-factor MDH model. It is straightforward that both price changes and trading volume depend on three groups of mutually independent variables: idiosyncratic variables ($\boldsymbol{\eta}$ and $\boldsymbol{\varepsilon}$), common-factor variables ($\mathbf{z}_1$ and $\mathbf{z}_2$), and the mixing variable (I). Note that the latent information flow variable is positive and enters the factor loadings of both price changes and volume. This amplifies the factor impacts on returns and volume, and makes them dynamic.

Clearly, in the case of one asset, the common-factor MDH model is identical to the single-asset MDH model in Tauchen and Pitts (1983); thus it retains all the properties of the single-asset MDH model for one asset and extends these properties to the multi-asset economy, as we can see in Propositions 1 and 2.

Proposition 1. The price changes and trading volume are independent within each firm as well as across firms because the factor structure of price changes is uncorrelated with the factor structure of trading volume.

Proof of Proposition 1.

$$\begin{aligned}
\mathrm{Cov}\{\Delta\mathbf{P},\mathbf{V}\} &= \mathrm{Cov}\left\{\boldsymbol{\alpha}I^{1/2}\mathbf{z}_1, \boldsymbol{\beta}I^{1/2}\mathbf{z}_2\right\} + \mathrm{Cov}\left\{\boldsymbol{\eta}I^{1/2}, \boldsymbol{\varepsilon}I^{1/2}\right\} \\
&= \mathrm{E}\left\{\boldsymbol{\alpha}I^{1/2}\mathbf{z}_1(\boldsymbol{\beta}I^{1/2}\mathbf{z}_2)^{\mathrm{T}}\right\} - \mathrm{E}\left\{\boldsymbol{\alpha}I^{1/2}\mathbf{z}_1\right\}\left[\mathrm{E}\left\{\boldsymbol{\beta}I^{1/2}\mathbf{z}_2\right\}\right]^{\mathrm{T}} \\
&= \mathrm{E}\left\{I\boldsymbol{\alpha}\,\mathbf{z}_1\,\mathbf{z}_2^{\mathrm{T}}\,\boldsymbol{\beta}^{\mathrm{T}}\right\} = 0. \quad \square
\end{aligned}$$

Conditional on the mixing variable, the common-factor MDH model reduces to the empirical factor model in Hasbrouck and Seppi (2001). However, Proposition 1 reflects the important difference between our model and theirs. Hasbrouck and Seppi assume that the factor structure of price changes is related to the factor structure of trading volume, which leads to the interdependence between price changes and trading volume. We do not invoke this assumption. Instead, as shown in Proposition 1, the variance-components scheme that we use allows us to derive the relationship between factor structures of price changes and trading volume, and hence the relationship between price changes and trading volume. It is clear that price changes and trading volume are uncorrelated regardless of the presence of common factors.

Proposition 2. The return variance is positively related to trading volume within each firm and across firms as long as the mixing variable shows variation.

Proof of Proposition 2.

$$\begin{aligned}
\mathrm{Cov}\left\{\Delta P_i^2, V_j\right\} &= \mathrm{E}\left\{\left(\Delta P_i^2\right) V_j\right\} - \mathrm{E}\left\{\Delta P_i^2\right\}\mathrm{E}\left\{V_j\right\} \\
&= \left\{\left[\boldsymbol{\eta}_i^2 + \sum_{k=1}^{K}\left(\boldsymbol{\alpha}_{ki}\right)^2\right]\boldsymbol{\mu}_j\right\}\left[\mathrm{E}\left\{I^2\right\} - \mathrm{E}^2\left\{I\right\}\right] \\
&= \left\{\left[\boldsymbol{\eta}_i^2 + \sum_{k=1}^{K}\left(\boldsymbol{\alpha}_{ki}\right)^2\right]\boldsymbol{\mu}_j\right\}\mathrm{Var}\left\{I\right\} \; > \; 0 \text{ as long as } \mathrm{Var}\left\{I\right\} \neq 0.
\end{aligned}$$

Subscripts i and j are index stocks. □

Our principal innovation is the distinction between information cross-asset impact and its asset-specific impact. This enables us to inspect the cross-asset aspects overlooked by the standard MDH model. We state these properties in Propositions 3 and 4.

Proposition 3. The price changes are correlated across firms if and only if common factors exist because the variance-covariance matrix of price changes is nondiagonal as long as factor loadings $\boldsymbol{\alpha} \neq \mathbf{0}$.

Proof of Proposition 3.

$$\begin{aligned}
\boldsymbol{\Sigma}_{\Delta\mathbf{P}} &= \mathrm{E}\left\{\Delta\mathbf{P}\Delta\mathbf{P}^{\mathrm{T}}\right\} - \mathrm{E}\left\{\Delta\mathbf{P}\right\}\mathrm{E}\left\{\Delta\mathbf{P}^{\mathrm{T}}\right\} \\
&= \mathrm{E}\left\{\left(\boldsymbol{\alpha}I^{1/2}\right)\mathbf{z}_1\mathbf{z}_1^{\mathrm{T}}\left(\boldsymbol{\alpha}I^{1/2}\right)^{\mathrm{T}}\right\} + \Sigma_{\boldsymbol{\eta}}\mathrm{E}\{I\} \\
&= \left(\boldsymbol{\alpha}\boldsymbol{\alpha}^{\mathrm{T}} + \boldsymbol{\Sigma}_{\boldsymbol{\eta}}\right)\mathrm{E}\left\{I\right\}. \quad \square
\end{aligned}$$

Proposition 4. The trading volume is correlated across firms if and only if common factors exist because the variance-covariance matrix of trading volume is nondiagonal as long as factor loadings $\boldsymbol{\beta} \neq \mathbf{0}$.

Proof of Proposition 4.

$$\begin{aligned}
\boldsymbol{\Sigma}_{\mathbf{V}} &= \mathrm{E}\left\{\mathbf{V}\mathbf{V}^{\mathrm{T}}\right\} - \mathrm{E}\left\{\mathbf{V}\right\}\mathrm{E}\left\{\mathbf{V}^{\mathrm{T}}\right\} \\
&= \mathrm{E}\left\{\left(\boldsymbol{\beta}I^{1/2}\right)\mathbf{z}_2\mathbf{z}_2^{\mathrm{T}}\left(\boldsymbol{\beta}I^{1/2}\right)^{\mathrm{T}}\right\} + \Sigma_{\boldsymbol{\varepsilon}}\mathrm{E}\{I\} \\
&= \left(\boldsymbol{\beta}\boldsymbol{\beta}^{\mathrm{T}} + \boldsymbol{\Sigma}_{\varepsilon}\right)\mathrm{E}\left\{I\right\}. \quad \square
\end{aligned}$$

The nondiagonal elements in the matrices $\boldsymbol{\alpha}\boldsymbol{\alpha}^{\mathrm{T}}$ and $\boldsymbol{\beta}\boldsymbol{\beta}^{\mathrm{T}}$ give us the cross-firm variations in returns and trading volume, respectively. It is worth mentioning that whereas the cross-firm interactions result from common factors, the cross-sectional positive volatility-volume relations arise solely from the latent information flow variable. We can see that Proposition 2 still holds in the absence of common factors (i.e., $\boldsymbol{\alpha} = \mathbf{0}$ and $\boldsymbol{\beta} = \mathbf{0}$).

III. Data Analysis

In Section III.A, we describe how we form our sample. In Section III.B, we carry out Pearson correlation and canonical correlation analyses to test the model's implications. In Sections III.C and III.D, we explore the role of the information flow in generating volatility persistence. Section III.E provides a robustness check.

A. Data

The sample includes 28 stocks in the DJIA from April 1 through June 30, 2007. Microsoft and Intel do not trade on the New York Stock Exchange (NYSE), so we exclude them from our sample. Similar to Hasbrouck and Seppi (2001), this study focuses on the Dow Jones stocks in order to increase the possibility of detecting common factors and to mitigate the nonconcurrent problem.

The transaction data are from the NYSE Trade and Quote (TAQ) database. Each trading day from 9:30AM to 4:00PM Eastern Standard Time is evenly divided into 13 half-hour intervals. We use midquotes at the beginning and at the end of each interval to compute the log return for each stock at that interval. Volume is in dollars. To avoid potential stale quotes at the market opening, we delete transaction data for the first 3 minutes of trading. We also eliminate observations with zero price changes and observations with overnight price changes and trading volume. This leaves a final sample of 369 half-hour observations for each stock.

Table 1 provides the means and standard deviations of returns and volume for each stock over the entire sample period. General Motors has the highest mean

TABLE 1

Descriptive Statistics of Returns and of Trading Volume

The study sample includes half-hour intraday data on the 28 Dow Jones stocks from April 1 through June 30, 2007. The means and standard deviations of returns and trading volume are calculated over the entire sample period.

		Returns (%)		Volume ($)	
Ticker	Name	Mean	Std. Dev.	Mean	Std. Dev.
AA	Alcoa Incorporated	0.008	0.455	403,281	396,894
AIG	American International Group, Inc.	–0.012	0.195	414,760	234,542
AXP	American Express Company	0.000	0.290	238,669	154,571
BA	Boeing Company	–0.010	0.275	187,629	106,685
C	Citigroup Incorporated	–0.015	0.300	720,347	435,264
CAT	Caterpillar Incorporated	–0.002	0.346	234,245	151,744
DD	DuPont	–0.015	0.305	230,200	149,527
DIS	Walt Disney Company	0.001	0.294	375,479	208,295
GE	General Electric Company	0.008	0.291	1,315,054	776,955
GM	General Motors Corporation	0.037	0.456	475,619	372,010
HD	Home Depot	–0.014	0.312	503,293	427,210
HON	Honeywell International Inc.	0.005	0.315	272,203	217,163
HPQ	Hewlett-Packard Company	0.019	0.304	505,981	256,465
IBM	International Business Machines	0.008	0.267	326,019	299,917
JNJ	Johnson & Johnson	–0.013	0.214	449,686	254,764
JPM	JPMorgan Chase & Company	–0.027	0.313	529,567	313,201
KO	Coca-Cola Company	0.000	0.219	360,966	200,005
MCD	McDonald's Corporation	–0.003	0.319	305,497	177,436
MMM	3M Company	0.013	0.273	403,787	242,735
MO	Altria Group Incorporated	–0.024	0.239	180,852	150,207
MRK	Merck & Company, Incorporated	0.008	0.326	457,424	337,455
PFE	Pfizer Incorporated	–0.015	0.266	1,235,878	661,629
PG	Procter & Gamble Company	–0.007	0.200	787,030	442,186
T	AT&T Incorporated	–0.005	0.312	444,908	258,826
UTX	United Technologies Corporation	0.007	0.255	189,547	100,528
VZ	Verizon Communications Inc.	0.006	0.286	486,536	324,010
WMT	Wal-Mart Stores Incorporated	–0.011	0.271	595,420	449,108
XOM	Exxon Mobil Corporation	0.011	0.296	840,262	362,339

return, 0.037%, while JPMorgan has the lowest mean return, −0.027%. Dollar volume means range from $180,852 (Altria Group Incorporated) to $1,315,054 (General Electric).

B. Preliminary Analyses

Our model indicates that returns and volume are independent, while returns are related across assets and so is volume. We undertake a simple Pearson correlation test to examine these implications. With 28 stocks in the sample, we have 784 pairwise return-return correlations, volume-volume correlations, and return-volume correlations. Only 103 pairs show significant return-volume correlations, but all 784 return-return and volume-volume correlations are statistically significant (not tabulated here). The canonical redundancy analysis results in Table 2 also support the prediction of Proposition 1. The proportions of variance in returns explained by trading volume are all below 0.02, and vice versa, indicating that neither of the return and volume canonical variables are good overall predictors of the opposite set of variables.

TABLE 2

Canonical Redundancy Analysis Results for Return and Trading Volume

The second column of Table 2 reports for each of the 28 Dow Jones stocks in the study the proportion of total variation in returns explained by trading volume. The third column presents the proportion of total variation in volume explained by return.

Ticker	Return Variance Explained by Volume	Volume Variance Explained by Return
1	0.0102	0.0137
2	0.0154	0.0120
3	0.0117	0.0138
4	0.0117	0.0092
5	0.0087	0.0039
6	0.0070	0.0073
7	0.0034	0.0070
8	0.0056	0.0080
9	0.0047	0.0099
10	0.0029	0.0025
11	0.0052	0.0075
12	0.0023	0.0038
13	0.0032	0.0024
14	0.0032	0.0019
15	0.0014	0.0021
16	0.0067	0.0015
17	0.0014	0.0017
18	0.0017	0.0005
19	0.0008	0.0008
20	0.0006	0.0005
21	0.0006	0.0005
22	0.0002	0.0005
23	0.0002	0.0006
24	0.0005	0.0002
25	0.0001	0.0001
26	0.0001	0.0001
27	0.0000	0.0000
28	0.0000	0.0000

Next, we perform a canonical correlation analysis to examine the relation between return volatility and trading volume. Canonical correlation analysis seeks to identify and quantify associations between two sets of variables. The aim is to summarize the associations between two sets of variables via a few carefully chosen pairs of canonical variables (canonical common factors). Through canonical correlation analysis, we can determine whether the positive relation between return volatility and trading volume holds across firms. We can also determine whether common factors underlie the relation.

In our case, for example, if the return volatility of each sample stock is only positively correlated with its own volume but is uncorrelated with other stocks' volume, all the canonical correlations should be 1.0. Moreover, if the 28-dimension volatility-volume relations cannot be concentrated into a few pairs of canonical variables, the positive correlations between return volatility and trading volume may not arise from the common factors.

Table 3 presents the canonical correlation analysis results. Half of the canonical correlations are statistically significant, positive, and less than 1.0. This result is consistent with the prediction of Proposition 2.

TABLE 3

Canonical Correlation Analysis Results for Return Volatility and Trading Volume

The second column of Table 3 reports canonical correlations for each of the Dow Jones 28 stocks in the study. The third and fourth columns provide results for testing the significance of the canonical correlations.

		Test Results	
Order Number	Canonical Correlations	*F*-Value	*p*-Value
1	0.819632	3.26	<0.0001
2	0.722151	2.86	<0.0001
3	0.691312	2.64	<0.0001
4	0.677421	2.45	<0.0001
5	0.638792	2.25	<0.0001
6	0.604625	2.07	<0.0001
7	0.549741	1.91	<0.0001
8	0.529986	1.80	<0.0001
9	0.474032	1.68	<0.0001
10	0.467845	1.61	<0.0001
11	0.440106	1.53	<0.0001
12	0.436730	1.45	<0.0001
13	0.406848	1.34	0.0003
14	0.381699	1.25	0.0086
15	0.334325	1.15	0.0751
16	0.326688	1.10	0.1901
17	0.316317	1.02	0.4230
18	0.279081	0.91	0.7348
19	0.241810	0.83	0.8854
20	0.224843	0.77	0.9321
21	0.214603	0.70	0.9632
22	0.202325	0.59	0.9886
23	0.148921	0.42	0.9992
24	0.106878	0.30	0.9997
25	0.087954	0.22	0.9995
26	0.045166	0.10	0.9996
27	0.026584	0.06	0.9930
28	0.004225	0.01	0.9379

C. Estimation of the Model

Given the encouraging results in Section III.B, we now proceed to estimate the model. Two major empirical issues arise in using the maximum likelihood estimation (MLE) method. First, the joint distribution of returns and volume is not a multivariate normal distribution but mixtures of independent multivariate normals with the mixing variable I. Second, factor and mixing variables are both unobservable. To tackle these issues, we follow a natural iterative three-step estimation strategy by integrating the expectation-maximization (EM) algorithm and the conditional moment method in Lamoureux and Lastrapes (1994).[4]

In the first step (information flow (INF)-step), we apply the Lamoureux and Lastrapes (1994) conditional moment method to find the time-series values of the unobserved information flow variable given $\Delta\mathbf{P}$, $\mathbf{V}$, and $\boldsymbol{\theta}$ ($\boldsymbol{\theta} = [\boldsymbol{\alpha}\ \boldsymbol{\beta}\ \boldsymbol{\mu}\ \boldsymbol{\eta}\ \boldsymbol{\varepsilon}]$). We can obtain the conditional moment criterion for each period t, which is the sum of the standardized squared deviations of squared returns and volume from their conditional means, by the following formula:

$$
\text{(8)} \quad L_t = \sum_{m=1}^{M} \left(\mathbf{e}_{mt} \boldsymbol{\Sigma}_{mt}^{-1} \mathbf{e}_{mt}^{\mathrm{T}}\right) = \sum_{m=1}^{M} \left\{ \frac{\left[\Delta\mathbf{P}_{mt}^2 - \left(\boldsymbol{\eta}_m^2 + \sum_{k=1}^{K} \boldsymbol{\alpha}_{mk}^2\right) I_t\right]}{2\left[\left(\boldsymbol{\eta}_m^2 + \sum_{k=1}^{K} \boldsymbol{\alpha}_{mk}^2\right) I_t\right]^2} - \frac{\left(\mathbf{V}_{mt}^2 - \boldsymbol{\mu}_m I_t\right)^2}{\left(\hat{\varepsilon}_m^2 + \sum_{k=1}^{K} \boldsymbol{\beta}_{mk}^2\right) I_t} \right\},
$$

where

$$
\mathbf{e}_{mt} = \begin{bmatrix} e_{1mt} \\ e_{2mt} \end{bmatrix} = \begin{pmatrix} \Delta\mathbf{P}_{mt}^2 - \mathrm{E}\left(\Delta\mathbf{P}_{mt}^2 \,|I_t, \boldsymbol{\theta}\right) \\ \mathbf{V}_{mt} - \mathrm{E}\left(\mathbf{V}_{mt} \,|I_t, \boldsymbol{\theta}\right) \end{pmatrix}
$$

is the forecast error of return volatility and volume conditional on the current value of the parameter $\boldsymbol{\theta}$ for each stock m at time t, and $\boldsymbol{\Sigma}_{mt} = \mathrm{E}(\mathbf{e}_{mt}\mathbf{e}_{mt}^{\mathrm{T}} \,|I_t, \boldsymbol{\theta})$ is the conditional variance-covariance matrix of $\mathbf{e}_{mt}$. Minimizing this criterion with respect to I_t can give us the value of I_t.

Then we eliminate the effect of the information flow for each period t from the observed return and volume series by the following formulas:

$$
\text{(9)} \quad \Delta\mathbf{P}_t' = \frac{\Delta\mathbf{P}_t}{\sqrt{I_t}},
$$

$$
\text{(10)} \quad \mathbf{V}_t' = \frac{\mathbf{V}_t - \boldsymbol{\mu}_t I_t}{\sqrt{I_t}}.
$$

With the normalized return and volume series, $\Delta\mathbf{P}'$ and $\mathbf{V}'$, model (7) becomes the multivariate normal factor model; hence we can use the standard EM algorithm for factor analysis to estimate the model parameters. Our second and third steps are just the expectation step (E-step) and maximization step (M-step) of the EM algorithm. That is, in the E-step, we calculate the expectation of unobserved factor variables $\mathbf{z}_1$ and $\mathbf{z}_2$ with the current value of the parameter $\boldsymbol{\theta}$, and in the M-step, we obtain the next value of the parameter by maximizing the expected log likelihood given the normalized return and volume data and the computed factor data. Using the new value of the parameter, we compute the next INF-step

[4]See, for example, Dempster, Laird, and Rubin (1977), Louis (1982), Rubin and Thayer (1982), Meng and Rubin (1991), and Van Dyk, Meng, and Rubin (1995).

and E-step and continue until the expected log likelihood difference between two successive iterations is sufficiently small.

We use the Akaike information criterion (AIC) to choose the number of factors. The criterion favors models with low AIC values.[5] Our computed AIC values indicate the existence of factor structures.[6]

D. Information Flow and Volatility Persistence

The underlying information flow plays an important role in generating price changes, so we further explore its characteristics using the extracted information flow, $\hat{I}_t$. Specifically, we examine whether the information flow accounts for the serial dependence in returns. Given the extracted information series, $\hat{I}_t$, we adjust the observed return series for the information flow by equation (9). If, as our model implies, the underlying information drives stock returns, the serial dependence of the adjusted return series should become insignificant. We estimate a GARCH(1, 1) model for raw and adjusted return series for each sample stock, respectively. The GARCH(1, 1) model is as follows:

$$
\begin{aligned}
y_t &= a_0 + a_a y_{t-1} + \varepsilon_t, \\
\varepsilon_t &\sim N(0, h_t), \\
h_t &= \gamma_0 + \gamma_1 \varepsilon_{t-1}^2 + \gamma_2 h_{t-1},
\end{aligned}
\tag{11}
$$

where y_t is the observed stock return at time t.

Table 4 displays the estimation results (reporting only those with significant GARCH persistence). The third column reports the estimates of the GARCH(1, 1) model for the raw return series. We find that 13 out of the 28 stocks present persistence in variance. None of the 13 stocks shows significant persistence in return variance after we control for the information effect. The estimates of the GARCH(1, 1) model for the adjusted return series are all insignificant.

E. Robustness Check

To investigate the model's market-wide applicability, we run a robustness check on the constituents of the S&P Composite 1500 using daily data from Jan. 1, 2004, through Dec. 31, 2006.

The matrix structure inherent in multi-asset settings and the slow convergence rate of the EM algorithm raise technical difficulties that prevent us from estimating the model on a broad cross section of stocks simultaneously. To resolve this technical issue without sacrificing generality, we divide the stocks in each S&P index respectively into six categories according to market capitalization and trading volume. Specifically, in each S&P index, we select 40 stocks with the largest (smallest, closest to category average) market capitalization and

[5]For example, see Tanner (1993).

[6]The point estimate of the parameter, $\hat{\boldsymbol{\theta}}$ ($\hat{\boldsymbol{\theta}} = [\hat{\boldsymbol{\alpha}} \quad \hat{\boldsymbol{\beta}} \quad \hat{\boldsymbol{\mu}} \quad \hat{\boldsymbol{\eta}} \quad \hat{\boldsymbol{\varepsilon}}]$), in the 1-factor MDH model is statistically significant at the 5% level (not reported).

TABLE 4

Estimates of GARCH(1, 1) for Dow Jones Stocks

Table 4 reports estimates for only those Dow Jones stocks that display significant GARCH persistence. The third column provides results of the original return series. None of the estimates is significant when the information effect is controlled.

Ticker	Parameter	Estimate
AIG	γ_1	0.1480
AXP	γ_1	0.1455
	γ_2	0.5978
C	γ_1	0.1592
DD	γ_1	0.1942
GE	γ_1	0.1084
	γ_2	0.7057
HD	γ_1	0.1045
HON	γ_1	0.1891
	γ_2	0.5242
JNJ	γ_2	9.9997×10^{-7}
KO	γ_1	0.2622
MMM	γ_1	0.1822
MRK	γ_1	0.0706
	γ_2	0.8337
T	γ_2	0.9753
WMT	γ_2	0.4553

40 stocks with the largest (smallest, closest to category average) trading volume. Therefore, the sample includes 720 stocks that trade on the NYSE, and each of the 18 categories contains 40 stocks.

The results from the Pearson correlation and canonical analyses are similar to those for the Dow Jones stocks (not reported), which support the predictions of Propositions 1–4. The model estimation results indicate the general existence of factor structures in stock returns and trading volume (see Table 5).

Table 6 reports the before- and after-adjustment GARCH comparison results. The third column displays the number of stocks with significant GARCH persistence in each category before the information adjustment. The fourth column presents the GARCH reduction percentages after the adjustment. Generally, half of the GARCH persistence becomes statistically insignificant. More interestingly, the reduction percentage increases with market capitalization and decreases with trading volume. We conjecture that this is because stocks with small market capitalizations and large trading volumes may have fewer participants and more noise trades, which is inconsistent with our model's normal-approximation requirement for high J as well as no-liquidity-trade specification and may induce more estimation errors.[7]

[7] Our model does not explicitly distinguish how uninformed investors' behavior affects volatility as well. However, many studies suggest that volatility in crises is magnified due to panic selling of uninformed investors, implying an increasing role of uninformed investors in generating volatility (see, e.g., Fleming, Kirby, and Ostdiek (1998), Kyle and Xiong (2001)). We carry out our empirical analysis for the 2008 period. The GARCH persistence reduction percentages in all categories are very low. Therefore, it is desirable to provide a theoretical model where the volatility effects of the information flow and uninformed investors can be separated in empirical work.

https://doi.org/10.1017/S0022109014000106 Published online by Cambridge University Press

TABLE 5

Factor Structures of Returns and Trading Volume for S&P Index Stocks

Table 5 reports some estimation results of the common-factor MDH model for the S&P index stocks. We divide the constituents of each of the three S&P indices into six categories according to market capitalization and trading volume. For estimation convenience, each category contains 40 stocks. The third column provides the number of factors in returns and volume selected by the AIC.

S&P Indices	Categories	Factor Structures of Returns and Trading Volume
S&P 400	Smallest market cap	1
	Average market cap	1
	Largest market cap	1
	Smallest volume	1
	Average volume	1
	Largest volume	1
S&P 500	Smallest market cap	1
	Average market cap	1
	Largest market cap	4
	Smallest volume	4
	Average volume	1
	Largest volume	1
S&P 600	Smallest market cap	1
	Average market cap	1
	Largest market cap	1
	Smallest volume	1
	Average volume	1
	Largest volume	1

TABLE 6

GARCH Comparison Results for S&P Index Stocks

The third column of Table 6 reports the number of stocks in each category whose original return series present GARCH persistence. The fourth column provides the persistence reduction percentages of the information-adjusted return series in each category. The fifth column provides the persistence reduction percentages of the $\widehat{\text{VIX}}$-adjusted return series in each category. $\widehat{\text{VIX}}$ is the predicted VIX from its regression on extracted information, $\hat{I}$.

S&P Indices	Categories	Original Return Series	Information-Adjusted Return Series	$\widehat{\text{VIX}}$-Adjusted Return Series
S&P 400	Smallest market cap	27	70.37%	3.70%
	Average market cap	28	64.29%	7.14%
	Largest market cap	24	50.00%	4.17%
	Smallest volume	28	75.00%	7.14%
	Average volume	30	63.33%	0.00%
	Largest volume	20	60.00%	10.00%
S&P 500	Smallest market cap	29	65.52%	3.45%
	Average market cap	31	77.42%	0.00%
	Largest market cap	31	74.19%	0.00%
	Smallest volume	30	73.33%	0.00%
	Average volume	29	68.97%	3.45%
	Largest volume	30	60.00%	0.00%
S&P 600	Smallest market cap	31	29.03%	3.23%
	Average market cap	30	53.33%	0.00%
	Largest market cap	26	46.15%	3.85%
	Smallest volume	32	62.50%	3.13%
	Average volume	30	46.67%	3.33%
	Largest volume	29	37.93%	6.90%

Because of the notion that the information flow generates return volatility, the implied volatility index (VIX) is perceived as a good proxy for the information flow.[8] However, the VIX is just an average of implied volatility; it is in fact a

[8]Empirical studies have been conducted to test whether the VIX is a superior informative variable (see, e.g., Blair, Poon, and Taylor (2001), Degiannakis and Floros (2010)).

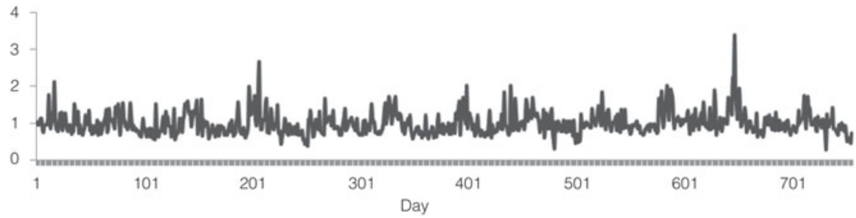
jumble of many influences, and there is not much a priori reason to expect it to rigorously capture informational events. We find that the time-varying features of the VIX and the extracted information flows are quite different by observing their time-series plots (see Figure 1). Graphs A and B present the time-series plots of the estimated daily information flow and intraday information flow, respectively. They appear to be more stationary than the daily VIX series, which shows some trends over time (see Graph C).

FIGURE 1

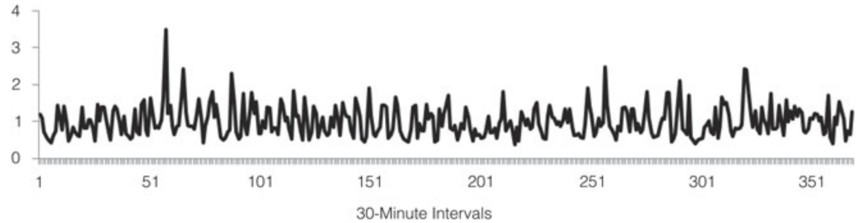
Time-Series Plots of VIX and Estimated Information Flows

Graph A of Figure 1 displays daily information flow from Jan. 1, 2004, through Dec. 31, 2006; Graph B displays half-hour intraday information flow from April 1 through June 30, 2007; and Graph C displays daily VIX from Jan. 1, 2004, through Dec. 31, 2006.

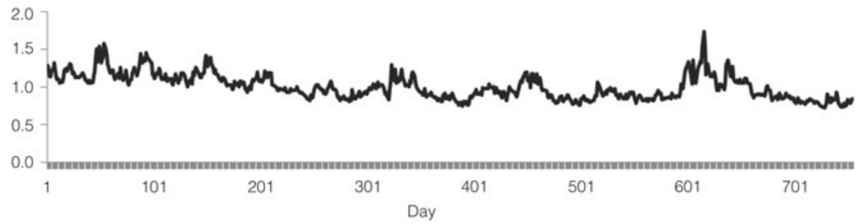
Graph A. Daily Information Flow



Graph B. Intraday Information Flow



Graph C. Daily VIX



To examine directly the relation between the VIX and information flow, we project the extracted information flow $\hat{I}_t$ on the VIX by a simple linear regression model and use the predicted value ($\widehat{\mathrm{VIX}}_t$ in the following equation) as the informational content of the VIX.[9]

[9]The estimated value of b_1, the projection coefficient of $\hat{I}_t$ on the VIX, is 0.00003863 and statistically significant at the 5% level.

$$\widehat{\text{VIX}}_t = \hat{b}_0 + \hat{b}_1 \hat{I}_t. \tag{12}$$

We carry out a similar return adjustment for $\widehat{\text{VIX}}$, $\Delta\mathbf{P}_{\text{VIX}} = \Delta\mathbf{P}/\sqrt{\widehat{\text{VIX}}}$, and run GARCH model (11) on the $\widehat{\text{VIX}}$-adjusted return series for each stock. The fifth column of Table 6 presents the results. Comparing the results for $\widehat{\text{VIX}}$-adjusted returns with those for information-adjusted returns in Table 6, it is clear that the VIX is an inadequate proxy for the information flow.[10]

IV. Summary

We extend a single-asset MDH model in Tauchen and Pitts (1983) to a multiple-asset setting. The model retains all the properties of the standard MDH model. In addition, it explains commonality in stock returns and trading volume. In our model, cross-firm variations result from information in two ways: through its impact structure (e.g., an industry or size or market effect) and through its arrival rate. The model yields a similar econometric specification to Hasbrouck and Seppi's (2001) factor model; however, our model predicts that factor structures of returns and trading volume are independent.

Using intraday data for the Dow Jones stocks and daily data for the S&P index constituents, the model seems to provide a useful framework that captures the stylized facts of the data. The empirical results are generally consistent with our model's implications. Specifically, the estimated information flow appears to capture the heteroskedasticity in stock returns.

Although the estimation result identifies the general existence of factor structures in both returns and volume, we find no patterns in the estimated factor loadings. Therefore, the identity of the impact structure of information remains unknown.

References

Admati, A. R. "A Noisy Rational Expectations Equilibrium for Multi-Asset Securities Markets." *Econometrica*, 53 (1985), 629–658.

Andersen, T. "Return Volatility and Trading Volume: An Information Flow Interpretation of Stochastic Volatility." *Journal of Finance*, 51 (1996), 116–204.

Bernhardt, D., and B. Taub. "Cross-Asset Speculation in Stock Markets." *Journal of Finance*, 63 (2008), 2385–2427.

Blair, B. J.; S. Poon; and S. J. Taylor. "Forecasting S&P 100 Volatility: The Incremental Information Content of Implied Volatilities and High-Frequency Index Returns." *Journal of Econometrics*, 105 (2001), 5–26.

Brenner, M.; P. Pasquariello; and M. Subrahmanyam. "On the Volatility and Comovement of U.S. Financial Markets around Macroeconomic News Announcements." *Journal of Financial and Quantitative Analysis*, 44 (2009), 1265–1289.

Caballe, J., and M. Krishnan. "Imperfect Competition in a Multi-Security Market with Risk Neutrality." *Econometrica*, 62 (1994), 695–704.

Clark, P. K. "A Subordinated Stochastic Process Model with Finite Variance for Speculative Prices." *Econometrica*, 41 (1973), 135–155.

Degiannakis, S., and C. Floros. "VIX Index in Interday and Intraday Volatility Models." *Journal of Money, Investment and Banking*, 13 (2010), 21–26.

[10]We also use a nonlinear regression model to predict the VIX. The GARCH results are slightly better than those for the linearly predicted VIX.

Dempster, A. P.; N. M. Laird; and D. B. Rubin. "Maximum Likelihood from Incomplete Data via the EM Algorithm." *Journal of the Royal Statistical Society*, Ser. B, 39 (1977), 1–38.

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

Fleming, J.; C. Kirby; and B. Ostdiek. "Information and Volatility Linkages in the Stock, Bond, and Money Markets." *Journal of Financial Economics*, 49 (1998), 111–137.

Gallant, A. R.; P. E. Rossi; and G. Tauchen. "Stock Prices and Volume." *Review of Financial Studies*, 5 (1992), 199–242.

Gallant, A. R.; P. E. Rossi; and G. Tauchen. "Nonlinear Dynamic Structures." *Econometrica*, 61 (1993), 871–908.

Hasbrouck, J., and D. J. Seppi. "Common Factors in Prices, Order Flows, and Liquidity." *Journal of Financial Economics*, 59 (2001), 383–411.

Heimstra, C., and J. Jones. "Testing for Linear and Nonlinear Granger Causality in the Stock Price-Volume Relation." *Journal of Finance*, 49 (1994), 1639–1664.

Kyle, A., and W. Xiong. "Contagion as a Wealth Effect." *Journal of Finance*, 56 (2001), 1401–1440.

Lamoureux, C., and W. D. Lastrapes. "Endogenous Trading Volume and Momentum in Stock-Return Volatility." *Journal of Business & Economic Statistics*, 12 (1994), 253–260.

Lo, A. W., and J. Wang. "Trading Volume: Definitions, Data Analysis, and Implications of Portfolio Theory." *Review of Financial Studies*, 13 (2000), 257–300.

Louis, T. A. "Finding Observed Information Using the EM Algorithm." *Journal of the Royal Statistical Society*, Ser. B, 44 (1982), 98–130.

Markowitz, H. "Portfolio Selection." *Journal of Finance*, 7 (1952), 77–91.

Meng, X., and D. B. Rubin. "Using EM to Obtain Asymptotic Variance-Covariance Matrices: The SEM Algorithm." *Journal of the American Statistical Association*, 86 (1991), 899–909.

Richardson, M. P., and T. Smith. "A Direct Test of the Mixture of Distribution Hypothesis: Measuring the Daily Flow of Information." *Journal of Financial and Quantitative Analysis*, 29 (1994), 101–116.

Ross, S. "Mutual Fund Separation in Financial Theory: The Separating Distributions." *Journal of Economic Theory*, 17 (1978), 254–286.

Rubin, D. B., and D. T. Thayer. "EM Algorithm for ML Factor Analysis." *Psychometrika*, 47 (1982), 69–76.

Tanner, M. A. *Tools for Statistical Inference*. New York: Springer-Verlag (1993).

Tauchen, G. E., and M. Pitts. "The Price Variability-Volume Relationship on Speculative Markets." *Econometrica*, 51 (1983), 485–505.

Tookes, H. E. "Information, Trading, and Product Market Interactions: Cross-Sectional Implications of Informed Trading." *Journal of Finance*, 63 (2008), 379–413.

Van Dyk, D. A.; X. Meng; and D. B. Rubin. "Maximum Likelihood Estimation via the ECM Algorithm: Computing the Asymptotic Variance." *Statistica Sinica*, 5 (1995), 55–75.