Commonality, Information and Cross-Sectional Return / Volume Interactions

Abstract

This paper develops a common-factor model to investigate the cross-sectional relation of security returns, return volatility and trading volume. The model generalizes and outperforms the standard MDH model by capturing possible interaction effects among securities. Our model implies: (1) common factor structures stem from information flows; (2) cross-firm variations rely on a latent information variable and common factor structures; (3) common factor structures have no explanatory power in the positive relationship between volatility and volume. We fit the model for intraday data of Dow Jones 30 stocks using EM algorithm. The results support our specifications and show a 3-factor structure.

Commonality, Information and Cross-Sectional Return / Volume Interactions

1. Introduction

On a multi-security market where dynamic co-movements across assets are of interest, it is important to investigate cross-sectional return/trading volume interactions. However, to our knowledge, previous research has not established a unified framework in this area. On the one hand, common factor structures in returns and trading volume are examined isolatedly. While the study of common factors in stock returns based upon portfolio theory has been a classic theme in financial economics, the implication of portfolio theory for the cross-sectional behavior of equity trading volume is a new issue examined by recent study (e.g., Lo and Wang, 2000). On the other hand, theoretical work is separated from empirical work. For example, Caballe and Krishnan (1994) develop a theoretical model of insider trading to show that prices and order flows are governed by a common factor structure. However, there is no empirical evidence to support their model. On the contrary, when Hasbrouck and Seppi (2001) explore the relationship between the common factor structures in returns and trading volume for returns and order flows, they assume an empirical statistical model without any theoretical foundation.

This paper is the first to unify these lines of research. We develop an empirically testable common-factor model from a unified framework of portfolio and market microstructure theories. The purpose of this study is to capture and interpret the commonality in returns and trading volume. Specifically, we examine the cross-sectional relationship of returns, return volatility and trading volume on a multi-security market. Three key implications can be reached from our model: (1) the common factor structures in returns and trading volume stem from information flows; (2) The cross-firm variations for returns and volume rely on both the underlying latent information variable and common factor structures; (3) The positive relation between return volatility and volume, on the other hand, results from the underlying latent information variable only. The common factor structures have no explanatory power in this positive relationship.

Our model is a generalization of the standard Mixture Distribution Hypothesis (MDH) model developed by Tauchen and Pitts (1983). Motivated by the well-documented empirical finding that price volatility is positively related to trading volume, Tauchen and Pitts (1983; hereafter TP) derive a joint distribution of price changes and trading volume of a single stock from economic theory. Their derivation is based upon a variance-component assumption about the within-day revision of traders' reservation prices. The resulting joint distribution is very attractive in that it contains all relevant information underlying both price volatility and trading volume. However, the limitation of the standard MDH model lies in the fact that TP treat securities as isolated from each other. Obviously, this is not realistic for a multi-security market. To incorporate possible interaction effects among securities, we generalize the standard MDH model by introducing K common factor variance components into their model. These K common-factor variance components represent different characteristics of information's effect (such as firm effect, market effect, etc.) on trading.

Moreover, although our model is the same as Haubrouck and Seppi's (2001) empirical model in terms of the form, there are two main differences. First, our model implies that common factor structures in returns and trading volume emerge from underlying latent information. In contrast, since Hasbrouck and Seppi's model (2001) is only an empirical statistical model, it is theoretically incapable to uncover sources of common factor structures in returns and trading volume. Second, the common factors do not play a role in the price volatility-volume relationship in our model, whereas Hasbrouck and Seppi (2001) assume that the positive relationship between price volatility and trading volume extends to their respective common factor structures.

The generalized dynamic model is estimated and tested for half-hour intraday data of Dow Jones 30 stocks. The empirical results confirm the specifications of our model. First, we find that common factors exist in our sample, using two different analytical techniques. The canonical analysis result shows that there are 3-factor structures in returns and trading volume. The estimated results, using EM algorithm, also show that: (i) 9 out of 30 stocks are affected by 3 factors; (ii) 9 out of 30 stocks are affected by 2 factors; (iii) 12 out of 30 stocks are affected by one factors. In other words, the largest number of common factors present in our sample is three. This is consistent with the canonical analysis result. All 30 stocks in our sample are influenced by at least one common factor. This demonstrates that our model is superior to the standard MDH model because the standard MDH model is a special case of our model without the presence of common factors. Second, we test the implication that the cross-sectional return-volume relationship is caused by the latent information variable. As predicted by our model, the test result illustrates that the return-volume relation disappears after the effect of the latent information is evened out. Finally, the result from further test (Lamoureux and Lastrapes 1994) on the source of the persistence in return variance is supportive of general specifications of MDH models.

The paper is organized as follows. In section 2, we describe the model in more detail. Section 3 presents data analysis. Section 4 provides a brief summary.

2. The Common-Factor MDH Model

2.1 The Market

We deal with a market consisting of *M* securities. The market mechanism is the same as TP's[1]. There are *J* traders who choose to trade a specific stock. Each time, traders will take a long or short position in one share of a stock. We make three assumptions: (1) *J* is large and fixed "within a fixed time interval *D*" (hereafter, we use "within-*D*" or "*D*-ly" to stand for it). (2) Traders can trade more than one stock and they are indifferent over stocks. (3) Traders have tendency to hold unchanged portfolios. This last assumption makes assumption (1) possible. Although it seems somewhat contradictory to the "indifference" assumption, the contradiction disappears if we think in the way that traders prefer a certain kind of portfolio just because they are more familiar with it. In other words, traders are indifferent over stocks in terms of the price of stocks, but they have preferences for which stock they choose to trade.

The so-called "unchanged portfolio" is a portfolio with fixed components of securities. That is, people adjust their portfolio holdings by only changing the quantity of each stock in the portfolio. For example, if a trader chooses to trade both stock A and B when she enters the market, she will never trade any other stocks until she exits the market. Lo and Wang (2000) make a similar assumption about portfolio holdings when they derive the factor structure of trading volume. They assume that the proportion of each stock in the portfolio is constant. Comparing to their assumption, we will immediately find that our assumption is weaker because their assumption implies our assumption. That is, the violation of our assumption will lead to the violation of their assumption, but the converse is not necessarily

---

[1] Whereas Tauchen and Pitt (1983) pay their attention only on individual stock in isolation, we focus on the cross-sectional interactions among stocks.

true. For example, if people change the components of their portfolios, let's say, if they exclude a stock from the portfolio or add one to the portfolio, the proportion of this stock will change from a positive value to zero or from zero to a positive value.

We should mention that portfolio rebalance in our model does not arise for variance reduction or budget constraints in traditional CAPM or APT models. It is well known that information plays an important role in modern market microstructure theory. Trading and in turn price changes are induced by information releases. Hence, on a market of imperfect competitions (e.g., asymmetry information, different opinions about information, etc.), we assume that portfolio rebalance arises for strategic purposes. In this way, we can model price and volume dynamics under a unified framework of market microstructure and portfolio theories.

2.2 The Model

This section develops the model for a specific stock in the multi-security market. Suppose that there are I equilibrium phases within-D. The movement from one temporary equilibrium phase to the next is caused by the arrival of new information to the market. We adopt TP's method to formulate prices and trading volume. Suppose that at the time of the (i-1)st equilibrium, trader j's reservation price and the market price are $P^{*}_{i-1,j}$ and P_{i-1}, respectively. Then the desired position $Q_{i-1,j}$ of trader j and the resulting trading volume V_{i-1} are given by the following:

$$Q_{i-1,j} = c\left(P^*_{i-1,j} - P_{i-1}\right) \tag{1}$$

$$V_{i-1} = \frac{c}{2}\sum_{j=1}^{J}\left| Q_{i-1,j} - Q_{i-2,j} \right| \tag{2}$$

$$i = 1,2,\ldots,I;\quad j = 1,2,\ldots,J;\quad k = 1,2,\ldots,K$$

where c>0 is a constant. From (1), we know that if the stock is undervalued, i.e., the trader's reservation price is higher than the market price, the trader will take a long position and $Q_{i-1,j}$ will be positive. On the contrary, if the stock is overvalued, i.e., the trader's reservation price is lower than the market price, the trader will take a short position and this will result in a negative $Q_{i-1,j}$.

According to the equilibrium condition, $\sum_{j=1}^{J} Q_{i-1,j} = 0$, we can derive the following equality from (1):

$$P_{i-1} = \frac{1}{J}\sum_{\mathrm{j=1}}^{\mathrm{J}} P^*_{i\text{-}1,j}$$

which implies that the market price for the stock at the time of the (*i*-1)st equilibrium equals the average of the reservation prices of all traders.

Now suppose that new information arrives at the market. Traders will change their reservation prices accordingly. After the information is completely assimilated into the market, the *i*th equilibrium reaches. Consequently, the market price change at the time of the *i*th equilibrium and the associated trading volume can be respectively written as:

$$\Delta P_i = \frac{1}{J}\sum_{j=1}^{J}\Delta P_{i,j}^{*} \tag{3}$$

$$V_i = \frac{c}{2}\sum_{j=1}^{J}\left|\Delta P_{ij}^{*} - \Delta P_i\right| \tag{4}$$

where $\Delta P_{ij}^{*} = P_{ij}^{*} - P_{i-1,j}^{*}$ is the increment to trader *j*'s reservation price.

TP assume a two-variance-component model for ΔP_{ij}^{*}. One component is common to all traders trading the stock and the other is specific to an individual trader trading the stock. Since we are dealing with a multi-security market, we extend TP's variance-component model by adding a third component $\sum_{k=1}^{K}\xi_{ijk}$. Furthermore, we add proportions to these components, i.e., each component results from only part of the information's effect. Thus, we get

$$\Delta P_{ij}^{*} = \alpha_0\left(\phi_i + \psi_{ij}\right) + \sum_{k=1}^{K}\alpha_k \xi_{ijk} \tag{5}$$

where, the component ϕ_i is common to all traders who choose to trade that stock, while the component ψ_{ij} is specific to trader *j* trading that stock. Both components are normally distributed with mean zero, variances σ_ϕ^2 and σ_ψ^2, respectively. Both are mutually independent and independent across securities and through time. The third component, $\sum_{k=1}^{K}\xi_{ijk}$, is common to at least more than one security in the market, but specific to a certain group of traders. We also assume that the distribution of each component ξ_{ijk} is normal with mean zero and variance $\sigma^2_{\xi_k}$. All ξ_{ijk}'s are mutually independent and independent of other two components, ϕ_i and ψ_{ij}. In the equation, α_i's represent respective proportions for each component, and $\sum_{i=0}^{K} \alpha_i = 1$.

Before going further to the derivation of our model, it is necessary to interpret these three components in terms of information's effect. From previous description, we know that the traders' reservation price changes are caused by the arrival of new information. Therefore, the first two components in equation (5) can be interpreted as the idiosyncratic effect of information on trading of a specific security, and the third component, $\sum_{k=1}^{K} \xi_{ijk}$, can be interpreted as the common effect of information on the trading of securities. Based upon the characteristics of the common effect of information (e.g., firm effect, market effect, etc.), the common variance component can be further classified into *K* categories. These are called *K* common factors and can be treated as the reason why traders change the quantity of their portfolios.

Substitute (5) into (3) and (4), we get

$$\Delta P_i = \alpha_0 \left(\phi_i + \overline{\psi}_i \right) + \sum_{k=1}^{K} \alpha_k \overline{\xi}_{ik}, \quad \overline{\psi}_i = \frac{1}{J} \sum_{j=1}^{J} \psi_{ij}, \quad \overline{\xi}_{ik} = \frac{1}{J} \sum_{j=1}^{J} \xi_{ijk} \tag{6}$$

$$V_i = \frac{\alpha}{2} \sum_{j=1}^{J} \left| \alpha_0 \left(\psi_{ij} - \overline{\psi}_i \right) - \left(\sum_{k=1}^{K} \alpha_k \xi_{ijk} - \sum_{k=1}^{K} \alpha_k \overline{\xi}_{ik} \right) \right| \tag{7}$$

Proposition 2.2.1: (i) The price change ΔP_i is normally distributed with mean zero and variance $\sigma_{\Delta P}^2$. (ii) For large *J*, the volume V_i is asymptotically normally distributed with mean μ and variance σ_V^2. (iii) The price change ΔP_i and trading volume V_i are stochastically independent.

Item (i) and (ii) are evident (see TP's proof). Item (iii) follows because: first, the component (ϕ_i) common to all traders who choose to trade a certain stock has been eliminated in the generation of trading volume; second, the common factor components $\overline{\psi}_i$ and $\sum_{k=1}^{K}\overline{\xi}_{ik}$ are independent of their respective deviations from means $\psi_{ij} - \overline{\psi}_i$ and $\sum_{k=1}^{K}\xi_{ijk}$- $\sum_{k=1}^{K}\overline{\xi}_{ik}$.

Aggregating price changes and trading volume within the time interval *D* yields the following within-*D* price changes and volume:

$$\Delta P = \sum_{i=1}^{I} \Delta P_i$$
$$V = \sum_{i=1}^{I} V_i$$

Simplifying notations, we have the following equations:

$$\begin{aligned} \Delta P &= \alpha_0 \sigma_{10} \sqrt{I} z_{10} + \sum_{k=1}^{K} \alpha_k \sigma_{1k} \sqrt{I} z_{1k} \\ V &= \alpha_0 \left(\mu_0 I + \sigma_{20} \sqrt{I} z_{20} \right) + \sum_{k=1}^{K} \alpha_k \mu_k I + \sum_{k=1}^{K} \alpha_k \sigma_{2k} \sqrt{I} z_{2k} \end{aligned} \tag{8}$$

where $z_{1i}'s$ and $z_{2i}'s$ $(i = 0,...,K)$ are all standard normal. All *z*'s are all mutually independent, and independent of the number of information arrivals within the time interval *D* (denoted as *I*). We refer model (8) as a **Common-factor MDH** model. It can be recognized that when K=0, i.e., there is no common factor present, model (8) is exactly the same as TP's standard MDH model.

2.3 Properties of the Model

Using matrix notation, we generalize model (8) to M securities:

$$\begin{aligned}\mathbf{\Delta P} &= \boldsymbol{\varphi}\mathbf{G} + \boldsymbol{\eta}\\ \mathbf{V} &= \boldsymbol{\mu} I + \boldsymbol{\theta}\mathbf{F} + \boldsymbol{\varepsilon}\end{aligned}$$

where

$$\mathbf{\Delta P} = \begin{bmatrix} \Delta P_1 \\ \vdots \\ \Delta P_M \end{bmatrix}_{M\times 1}, \quad \mathbf{V} = \begin{bmatrix} V_1 \\ \vdots \\ V_M \end{bmatrix}_{M\times 1}, \quad \mathbf{F} = \begin{bmatrix} z_{21} \\ \vdots \\ z_{2K} \end{bmatrix}_{K\times 1}, \quad \mathbf{G} = \begin{bmatrix} z_{11} \\ \vdots \\ z_{1K} \end{bmatrix}_{K\times 1}$$

$$\boldsymbol{\varphi} = \begin{bmatrix} \alpha_{01}\sigma_{111}\sqrt{I_1} & \cdots & \alpha_{K1}\sigma_{1K1}\sqrt{I_1} \\ \vdots & \ddots & \vdots \\ \alpha_{01}\sigma_{11M}\sqrt{I_M} & \cdots & \alpha_{KM}\sigma_{1KM}\sqrt{I_M} \end{bmatrix}_{M\times K}$$

$$\boldsymbol{\mu} = \begin{bmatrix} \alpha_{01}\mu_{01}I_1 + \sum_{k=1}^{K}\alpha_{k1}\mu_{k1}I_1 \\ \vdots \\ \alpha_{0M}\mu_{0M}I_M + \sum_{k=1}^{K}\alpha_{kM}\mu_{kM}I_M \end{bmatrix}_{\mathrm{M}\times 1}, \quad \boldsymbol{\theta} = \begin{bmatrix} \alpha_{01}\sigma_{211}\sqrt{I_1} & \cdots & \alpha_{K1}\sigma_{2K1}\sqrt{I_1} \\ \vdots & \ddots & \vdots \\ \alpha_{0M}\sigma_{21M}\sqrt{I_M} & \cdots & \alpha_{KM}\sigma_{2KM}\sqrt{I_M} \end{bmatrix}_{\mathrm{M}\times\mathrm{K}}$$

$$\boldsymbol{\eta} = \begin{bmatrix} \alpha_{01}\sigma_{101}z_{101}\sqrt{I_1} \\ \vdots \\ \alpha_{0M}\sigma_{10M}z_{10M}\sqrt{I_M} \end{bmatrix}_{M\times 1}, \quad \boldsymbol{\varepsilon} = \begin{bmatrix} \alpha_{01}\sigma_{201}z_{201}\sqrt{I_1} \\ \vdots \\ \alpha_{0M}\sigma_{20M}z_{20M}\sqrt{I_M} \end{bmatrix}_{M\times 1}$$

Both vectors of price changes and of trading volume are governed by three kinds of mutually independent variables: idiosyncratic variables ($\boldsymbol{\eta}$ and $\boldsymbol{\varepsilon}$), common factor variables (**G** and **F**) and mixing variables ($I_{m,}$, m=1,2,...,M). Apparently, conditional on mixing variabls (the number of information arrivals), our model, with **E**(**G**)= **E**(**F**)=0 and **Cov**(**G**)= **Cov**(**F**)=**I** (**I** *here is* identity matrix), is exactly the same as Hasbrouck and Seppi's (2001) empirical model. In addition, notice that we allow the mixing variable for each stock to be different from each other. Consequently, we do not require concurrent trading as Hasbrouck and Seppi (2001) do.

Our model has properties described in the following propositions:

Proposition 2.3.1: The rank of covariance matrix of $\mathbf{\Delta P}$ and $\mathbf{V}$ is equal to the number of common factors K.

Proof: For $\Delta P = (\varphi \mathbf{G} + \boldsymbol{\eta})$, *we have* $\varphi \mathbf{G} = \mathbf{\Delta P} - \boldsymbol{\eta}$. *Treat* $\varphi \mathbf{G} = \mathbf{\Delta P} - \boldsymbol{\eta}$ *as linear systems in* **G**. *If there exist K common factors, i.e.,* **G** *has a unique solution, then* **G** *has full column rank K. Similarly, rank*($\boldsymbol{\theta}$)=*K. According to Matrix Algebra Theorems, since* $(\mathbf{GF'})(\mathbf{GF'}) = \mathbf{GF'FG'} = \mathbf{I}.$, *then* $rank(\mathbf{GF'}) = K$. *Thus, rank*($\varphi\mathbf{GF'}$)=*K, and consequently, rank*($\varphi\mathbf{GF'\theta'}$)=$K$.

Q.E.D.

This is a very useful property. Based upon this property, even though we don't know the joint distribution of ($\Delta P, V$), we can still determine the number of common factors through canonical analysis.

Proposition 2.3.2: *The cross-sectional interactions among price changes and those among trading volume depend upon both underlying latent information and common factor structures.*

Proof:

$$\mathbf{Cov}\{\mathbf{\Delta P}, \mathbf{\Delta P}\} = \mathbf{E}\{\mathbf{\Delta P \Delta P'}\} - \mathbf{E}\{\mathbf{\Delta P}\}[\mathbf{E}\{\mathbf{\Delta P}\}]' = \mathbf{E}\{\boldsymbol{\varphi}\mathbf{GG'}\boldsymbol{\varphi}' + \boldsymbol{\eta\eta}'\} = \mathbf{E}\{\boldsymbol{\varphi\varphi}' + \boldsymbol{\eta\eta}'\}$$

or

$$\mathbf{Var}\{\Delta P_i\} = \mathbf{E}\{I_i\}\left\{\alpha_{0i}^2\sigma_{10i}^2 + \sum_{k=1}^{K}\alpha_{ki}^2\sigma_{1ki}^2\right\}, \quad \mathbf{Cov}\{\Delta P_i, \Delta P_j\} = \mathbf{E}\left\{\sqrt{I_i I_j}\right\}\left\{\sum_{k=1}^{K}\alpha_{ki}\alpha_{kj}\sigma_{1ki}\sigma_{1kj}\right\}$$

$$\mathbf{Cov}\{\mathbf{V}, \mathbf{V}\} = \mathbf{E}\{\mathbf{VV'}\} - \mathbf{E}\{\mathbf{V}\}[\mathbf{E}\{\mathbf{V}\}]' = \mathbf{E}\{\boldsymbol{\theta}\mathbf{FF'}\boldsymbol{\theta}' + \boldsymbol{\varepsilon\varepsilon}'\} = \mathbf{E}\{\boldsymbol{\theta\theta}' + \boldsymbol{\varepsilon\varepsilon}'\}$$

or

$$\mathbf{Var}\{V_i\} = \mathbf{E}\{I_i\}\left\{\alpha_{01}^2\sigma_{2i}^2 + \sum_{k=1}^{K}\alpha_{ki}^2\sigma_{2ki}^2\right\}, \quad \mathbf{Cov}\{V_i, V_j\} = \mathbf{E}\left\{\sqrt{I_i I_j}\right\}\left\{\sum_{k=1}^{K}\alpha_{ki}\alpha_{kj}\sigma_{2ki}\sigma_{2kj}\right\}$$

where subscripts i, j index stocks.

Q.E.D.

Proposition 2.3.3: *Price volatility is positively related to trading volume as long as mixing variables show variation or positive covariation.*

Proof:

$$\begin{aligned}\mathbf{Cov}\{\Delta P_i^2, V_i\} &= \mathbf{E}\{\Delta P_i^2 V_i\} - \mathbf{E}\{\Delta P_i^2\}\mathbf{E}\{(V_i\} \\ &= \left\{\alpha_{0i}^3\sigma_{10i}^2\mu_{0i} + \sum_{k=1}^{K}\alpha_{ki}^3\sigma_{1ki}^2\mu_{ki}\right\}\mathbf{Var}\{I_i\} + \mathbf{E}\left\{\sum_{k=1}^{K}\alpha_{ki}^3\sigma_{1ki}^2\sigma_{2ki}z_{1ki}^2 z_{2ki}\right\}\mathbf{E}\left\{I_i^{3/2}\right\} \\ &= \left\{\alpha_{0i}^3\sigma_{10i}^2\mu_{0i} + \sum_{k=1}^{K}\alpha_{ki}^3\sigma_{1ki}^2\mu_{ki}\right\}\mathbf{Var}\{I_i\} > 0, \text{ as long as } \mathbf{Var}\{I_i\} \neq 0\end{aligned}$$

or

$$\begin{aligned}\mathbf{Cov}\{\Delta P_i^2, V_j\} &= \mathbf{E}\{\Delta P_i^2 V_j\} - \mathbf{E}\{\Delta P_i^2\}\mathbf{E}\{V_j\} \\ &= \left\{\alpha_{0i}^2\alpha_{0j}\sigma_{10i}^2\mu_{0j} + \sum_{k=1}^{K}\alpha_{ki}^2\alpha_{kj}\sigma_{1ki}^2\mu_{kj}\right\}\left[\mathbf{E}\{I_i I_j\} - \mathbf{E}\{I_i\}\mathbf{E}\{I_j\}\right] \\ &\quad + \mathbf{E}\left\{\sum_{k=1}^{K}\alpha_{ki}^2\alpha_{kj}\sigma_{1ki}^2\sigma_{2kj}z_{1ki}^2 z_{2kj}\right\}\mathbf{E}\{I_i I_j^{1/2}\} \\ &= \left\{\alpha_{0i}^3\sigma_{10i}^2\mu_{0i} + \sum_{k=1}^{K}\alpha_{ki}^3\sigma_{1ki}^2\mu_{ki}\right\}\mathbf{Cov}\{I_i, I_j\} > 0, \text{ as long as } \mathbf{Cov}\{I_i, I_j\} > 0.\end{aligned}$$

where subscripts i, j index stocks. ***Q.E.D.***

Clearly, in our model, the relationship between price volatility and trading volume depends only on the underlying latent information. The common factors do not play a role in this relationship. This is different from Hasbrouck and Seppi's (2001) empirical model, where they assume that the positive relationship between price volatility and trading volume extends to their respective common factor structures.

2.4 Estimation of the Model

In general, for each stock, the resulting model (8) is a mixture model. Since price changes and trading volume conditional on I are independent, their joint distribution conditional on I is a bivariate Gaussian mixture that can be written in the following form:

$$f(\Delta p, v \mid I) = \sum_{i=0}^{K}\alpha_i f_i(\Delta p, v \mid I), \qquad \sum_{i=0}^{K}\alpha_i = 1 \tag{9}$$

where

$$f_i(\Delta p, v \mid I) = \frac{1}{2\pi\sigma_{1i}\sigma_{2i}I} e^{-\frac{1}{2}\left\{\frac{(\Delta p)^2}{\sigma_{1i}^2 I} + \frac{(v-\mu_i I)^2}{\sigma_{2i}^2 I}\right\}}, \qquad i = 0, \ldots, K.$$

To estimate the parameters of model (9), we use EM algorithm (Dempster, Laird, and Rubin (1977)). Suppose there are n observations. To apply EM algorithm, the key thing is to introduce a (K+1)-dimensional vector of indictor variable **Z**=(z_1, z_2, ... , z_n). The unobservable random variable z_t can be defined by

$$z_{it} = \begin{cases} 1, & x_t, y_t \in f_i, \\ 0, & x_t, y_t \notin f_i, \end{cases} \quad i = 0, \ldots, K$$

where z_1,..., z_n are independently and identically distributed according to a multinomial distribution consisting of one draw on ($K+1$) categories with probabilities $\alpha_0, \ldots, \alpha_K$, respectively. Then we have the distribution of Z_t as

$$z_1, \ldots z_n \overset{i.i.d.}{\sim} Mult_{K+1}(1, \alpha)$$

In accordance with the bivariate mixture model (9), it is further assumed that (($\Delta p_1, v_1$),..., ($\Delta p_n, v_n$)) given (z_1,..., z_n) respectively are conditionally independent, and ($\Delta p_t, v_t$) given z_n has the following log density

$$\sum_{i=0}^{K} z_{it} \, log(f_i(\Delta p_t, v_t \mid \delta, I_t)), \quad t = 1, \ldots, n$$

where $\delta = (\alpha, \beta, \lambda, \mu)$, and

$$\begin{aligned} \alpha &= (\alpha_0, \ldots, \alpha_K) \\ \beta &= (\beta_0, \ldots, \beta_K) = (\sigma_{10}^2, \ldots, \sigma_{1K}^2) \\ \lambda &= (\lambda_0, \ldots, \lambda_K) = (\sigma_{20}^2, \ldots, \sigma_{2K}^2) \\ \mu &= (\mu_0, \ldots, \mu_K) \end{aligned}$$

Since both Z and I are unidentifiable variables, we treat Z as missing data and I as a parameter, and then we begin the *E-step* by estimating Z.

E step: Suppose $\delta^j = \left(\alpha^j, \beta^j, \lambda^j, \mu^j\right)$ and I^j denote the current guess of parameter δ and I, where $\delta^0 = \left(\alpha^0, \beta^0, \lambda^0, \mu^0\right)$ and I^0 are initial values of δ and I. Then *E-step* requires the calculation of

$$Q\left\{\delta, I \middle| \delta^j, I^j\right\} = \mathbf{E}\left\{L\left(\delta, I \middle| \Delta P_t, V_t, \delta^j, I^j\right)\right\}$$

i.e., the conditional expectation of $log\ p\left(\delta, I \middle| Z, \Delta P, V\right)$ with respect to $p\left(Z \mid \Delta P, V, \delta^j, I^j\right)$. This is equivalent to estimating $\mathbf{E}\left(Z \middle| \Delta P, V, \delta^j, I^j\right)$. With known parameters δ^j and I^j, and the observed data ($\Delta p, v$), we have

$$\mathbf{E}\left(z_{it}\right) = \frac{\alpha_i^j f_i\left(\Delta p_t, v_t \middle| \delta_i^j, I_t^j\right)}{\sum_{i=0}^{K} \alpha_i^j f_i\left(\Delta p_t, v_t \middle| \delta_i^j, I_t^j\right)}$$

M step: There are two steps in the *M-step*. First, the Q function is maximized with respect to δ, i.e.,

$$\left.\frac{\partial Q}{\partial \delta}\right|\left(\delta^j, I^j\right) = 0$$

Then we get the estimated parameter δ^{j+1},

$$\alpha_i^{j+1} = \frac{\sum_{t=1}^{n} \mathbf{E}(z_{ti})}{\sum_{t=1}^{n}\sum_{i=0}^{K} \mathbf{E}(z_{ti})}$$

$$\mu_i^{j+1} = \frac{\sum_{t=1}^{n} \mathbf{E}(z_{it}) v_t}{\sum_{t=1}^{n} \mathbf{E}(z_{it}) I_t^j}$$

$$\beta_i^{j+1} = \frac{\sum_{t=1}^{n} \frac{\mathbf{E}(z_{it})(\Delta p_t)^2}{I_t^j}}{\sum_{t=1}^{n} \mathbf{E}(z_{it})}$$

$$\lambda_i^{j+1} = \frac{\sum_{t=1}^{n} \frac{\mathbf{E}(z_{it})\left(v_t - \mu_i^{j+1} I_t^j\right)^2}{I_t^j}}{\sum_{t=1}^{n} \mathbf{E}(z_{it})}$$

$$i = 0, \ldots, K$$

Second, to extract an estimated I process conditional on the estimated δ^{j+1}, we adopt Lamoureux and Lastrapes's (1994) method to minimize the sum of the standardized squared deviations of squared returns and volume from their means for each fixed time interval. Lamoureux and Lastrapes (1994) conduct a simulation to show that their extraction procedure is effective. It would be more efficient to use the maximum likelihood estimation method (MLE) by incorporating all observations if we assume the distribution of the unobserved information flow. However, there are two reasons that prevent us from doing this. First, "the theory in its pure form puts no restrictions on the intertemporal behavior of the information flow variable" (Andersen 1996, pp. 187-188); second, although empirical work shows the autocorrelation in the information flow variable and hence suggest some time series models (e.g., ARMA, EARMA, SARV or ESARV, etc.) for the dynamic specifications of the information flow variable, "the computational burdens are insurmountable" (TP, pp. 500-501). Nevertheless, "the autocorrelation is probably a very weak threat to the statistical validity of the results" (TP, pp. 501) as long as the sample size is large.

Lamoureux and Lastrapes's (1994) method relies on the fact that the level of squared returns and volume are related by the mixing variable I. Let

$$e_t = \begin{pmatrix} e_1 \\ e_2 \end{pmatrix} = \begin{pmatrix} (\Delta p_t)^2 - \mathbf{E}\left(\Delta p_t \mid I_t, \delta^{j+1}\right) \\ v_t - \mathbf{E}\left(v_t \mid I_t, \delta^{j+1}\right) \end{pmatrix}$$

and

$$\mathbf{E}\left(e_t e_t' \mid I_t, \delta^{j+1}\right) = \Sigma_t = \begin{pmatrix} 2\left(\sum_{i=0}^{K} \alpha_i^2 \sigma_{1i}^2\right)^2 I_t^2 & 0 \\ 0 & \left(\sum_{i=0}^{K} \alpha_i^2 \sigma_{2i}^2\right) I_t \end{pmatrix}$$

Separately, for t=1,...,n, I_t^{j+1} can be obtained by minimizing the following conditional moment criterion:

$$L_t = \hat{e}_t' \hat{\Sigma}_t^{-1} \hat{e}_t = \frac{\left[(\Delta p_t)^2 - \sum_{i=0}^{K} \left(\alpha_i^{j+1}\right)^2 \beta_i^{j+1} I_t\right]^2}{2\left(\sum_{i=0}^{K} \alpha_i^2 \sigma_{1i}^2\right)^2 I_t^2} + \frac{\left[v_t - \sum_{i=0}^{K} \mu_i^{j+1} I_t\right]^2}{\left(\sum_{i=0}^{K} \alpha_i^2 \sigma_{2i}^2\right) I_t}$$

The *E-step* and *M-step* are iterated until $\left| Q\left(\delta^{j+1}, I^{j+1} \middle| \delta^{j}, I^{j}\right) - Q\left(\delta^{j}, I^{j} \middle| \delta^{j}, I^{j}\right)\right|$ is sufficiently small.

Having arrived at the estimated parameter $\hat{\delta}$, one can potentially evaluate

$$-\left.\frac{\partial^2 \log p\left(\delta, I \middle| \Delta P, V\right)}{\partial \delta^2}\right|_{\hat{\delta}}$$

However, this is difficult to evaluate in practice. Meilijson (1989) suggests a numerical differentiation method: perturb $\hat{\delta}$ by adding a small amount $\varepsilon > 0$ to one coordinate and evaluate the score function of the complete data set at the perturbed parameter $\tilde{\delta}$. The *i*th row of the Hessian is approximately equal to:

$$\frac{1}{\varepsilon}\left[S\left(\Delta\tilde{P}, \tilde{V}, I\right)\middle|\tilde{\delta} - S\left(\Delta P, V, I\right)\middle|\hat{\delta}\right]$$

where $S\left(\Delta P, V, I\right)$ is the score function of the complete data set.

Thus far, we only estimate the parameter δ and *I*. There still remains one parameter unestimated: the number of components K+1. This is not easy to estimate by the EM algorithm, because it does not appear in the linear form in Q. To choose the number of groups (K+1), we use the Bayes information criterion (BIC) :

$$\text{BIC} = -2\log(\text{maximized likelihood}) + 4(K+1)\log(n)$$

where n is the number of observations, and there are $4(K+1)$ independent parameters estimated. The criterion favors models with small BIC values.

3. Data Analysis

3.1 Data

The sample includes 30 stocks in Dow Jones Industrial Averages from April 1 to June 30, 1998. Hasbrouck and Seppi (2001) restrict their attention to the 30 Dow Jones stocks for two reasons: (1) to increase the possibility to detect common factors because of indexation; (2) to mitigate the *non-concurrent* problem because the Dow Jones 30 stocks are actively traded stocks. As we mentioned in previous section, non-concurrent trading will not be the problem in our model. Thus, we select the Dow Jones 30 stocks only for the first reason of Hasbrouck and Seppi (2001). The source of transaction data is NYSE Trade and Quote (TAQ). Each trading day from 9:30 a.m. to 4:00 p.m. Eastern Standard Time is evenly divided into 13 half-hour intervals. We use midquote at the beginning and the end of each interval to compute the log return for each stock at that interval. Volume is represented by dollar volume. To avoid the potential problem of stale quotes at the market open, transaction data for the first 3 minutes of trading are excluded. We also delete the observations with zero price changes, and observations with overnight price changes and trading volume. This leaves a final sample of 688 half-hour observations for each individual stock.

Table 1 provides means and standard deviations of returns and volume for each of the 30 Dow Jones stocks over the entire sample period. The means of returns range from –0.056% to 0.0183%. McDonald (MCD) has the largest mean return, which is 0.0183%, whereas Philip Morris Cos. (MO) has the smallest mean return, -0.056%. The means of the dollar volume range from $36,901.6 to $502,051.3. They belong respectively to Philip Morris Cos. (MO) and Goodyear (GT).

<Table 1 inserted here>

3.2 Canonical Analysis

In this section, we use canonical analysis to determine the number of common factors and examine the relationship between price volatility and trading volume. Canonical correlation analysis seeks to identify and quantify the associations between two sets of variables. The main purpose of the technique is to concentrate a high-dimentional relationship between two sets of variables into a few pairs of canonical variables.

Recall that *proposition 2.3.1* states that the number of common factors equals the rank of covariance matrix $\Sigma_{\mathbf{\Delta P,V}}$. In canonical analysis, if the rank($\Sigma_{\mathbf{\Delta P,V}}$)=$p$, then there will be p nonzero canonical correlations. Table 2 reports the canonical correlation analysis results for returns and trading volume. From Table 2, since there are three nonzero canonical correlations (at a 0.05 significance level), we have three common factors in our sample.

<Table 2 inserted here>

As we mentioned earlier, our model is different from that of Hasbrouck and Seppi (2001) in the way that the common factors do not play a role in the relationship between return volatility and trading volume in our model. Also, the *proposition 2.3.3* predicts that return volatility is positively related to trading volume as long as mixing variables show variation or positive covariation. Table 3 presents the canonical correlation test results for return volatility and volume. As shown in Table 3, there are approximate 26 pairs of significant canonical variables, hence, common factors do not appear to have an impact on the relationship between price volatility and trading volume. From Table 3, we can see that all correlation coefficients are positive. This result is consistent with the prediction of *proposition 2.3.3*.

<Table 3 inserted here>

3.3 Estimation Results

Table 4 provides the EM estimation results of model (9) for each of the 30 Dow Jones stocks. The magnitudes of all point estimates appear reasonable. For Eastman Kodak, Hewlett-Packard, International Paper, Coca cola, 3M, Phillip Morris, Sears, United Tech and Exxon, according to the BIC, the best models are those with 3 common factors. The results are shown in Panel A. For American Express, DuPont, General Motors, Goodyear, Johnson & Johnson, JP Morgan, MacDonald, Proctor Gamble and WalMart, the best models are those with 2 common factors. The results are provided in Panel B. The returns and trading volume of the rest stocks are only affected by one common factor (Panel C). As indicated earlier, the standard MDH model is a special case of our model when there is no common factor present. The estimated results, on the other hand, show that all stocks are affected by at least one common factor. Thus, our model outperforms the standard MDH model. In the previous section, we show, through the canonical analysis technique, that the common factor structures in returns and trading volume are 3-factor structures. Therefore, the estimation result is consistent with the canonical analysis result.

<Table 4 inserted here>

3.4 Model Evaluation

In this section, we conduct some tests to examine two properties of our model. According to model (8), stock return series is related to corresponding trading volume by the underlying information variable *I*. Given the extracted I_t series, $\hat{I}_t$, we can construct the adjusted return series as follows:

$$\Delta P_{adj} = \frac{\Delta P}{\sqrt{I}}$$

Thus, the relationship between this new adjusted return series and trading volume will disappear. To check this property, we regress return series on trading volume separately for each of the 30 stocks, before the adjustment of information and after the adjustment of information. We expect that the coefficient estimates between adjusted return series and trading volume will not be significantly different from zero. Table 5 reports the means of the coefficient estimates of trading volume across all stocks in the sample and the standard errors of the mean coefficient estimates. The standard errors are corrected for any cross-sectional correlation in the individual-stock estimators (see the Appendix in Jones et al. (1994)). From Table 5, we can see that the raw return series is positively related to trading volume. The average coefficient estimate is 6.08×10^{-10} and statistically significant (*t*-value=2.17). In contrast, the average coefficient estimate for the regression of the adjusted return series is not statistically different from zero (t-value=-0.26). This result confirms our expectation that the relationship between returns and trading volume will disappear after we control for information in return series.

<Table 5 inserted here>

In this section, following Lamoureux and Lastrapes (1994), we also test the validity of our model in explaining the serial dependence in returns by estimating a GARCH (1,1) model for raw and adjusted return series of each of the 30 stocks, respectively:

$$y_t = a_0 + a_a y_{t-1} + \varepsilon_t$$
$$\varepsilon_t \sim N(0, h_t)$$

and

$$h_t = \gamma_0 + \gamma_1 \varepsilon_{t-1}^2 + \gamma_2 h_{t-1}$$

where y_t is stock returns at time t. Table 6 presents the estimation results (only those cases in which GARCH persistence is evident in the return series are reported). The third column reports the estimates of raw return series. From this column, we can see 24 out of the 30 stocks display the persistence in variance. As our model suggests, the persistence is due mainly to the underlying information variable. Hence, we repeat the estimation of the GARCH (1,1) model for the adjusted return series. The results are provided in the fourth column of Table 6. After we control for information in raw return series, the persistence in variance disappears in most of those stocks that exhibit the persistence before the adjustment.

This result confirms the viability of our model in explaining the source of the persistence in variance.

<Table 6 inserted here>

4. Summary

In this paper, we specify a factor model of returns and volume, based upon market microstructure theory and portfolio theory. We fit our model for the half-hour intraday data of Dow Jones 30 stocks. The empirical results support the specifications of our model. Our model appears to provide a useful framework for capturing the important properties of the data. Especially, our model indicates that common factor structures are due to impacts of information on trading.

It is also important to understand how information flows influence the common factor structures in returns and trading volume. Usually, in factor analysis, we can further identify the source of common factors by observing the pattern of eigenvectors of factors. After looking at the eigenvectors of factors in our sample, we do not see any pattern of those eigenvectors. Therefore, the identification of common factors will leave for future research.

References

Andersen, T.G., 1996, "Return volatility and trading volume: an information flow interpretation of stochastic volatility", Journal of Finance 51, 169-204.

Caballe, J., Krishnan, M., 1994, "Imperfect competition in a multi-security market with risk neutrality", Econometrica 62, 695-704.

Campbell, J., Lo, A. and Mackinglay, C., 1996, "The econometrics of financial markets", Princeton University Press, Princeton, NJ.

Chordia, T., Roll, R., Subrahmanyam, A., 2000, "Commonality in liquidity", Journal of Financial Economics 56, 3-28.

Clark, P.K., 1973, "A subordinate stochastic process model with finite variance for speculative prices", Ecnometrica 41, 135-155.

Dempster, A.P., Laird, N.M. and Rubin, D.B., 1977, "Maximum likelihood from incomplete data via the EM algorithm", Journal of the Royal Statistical Society, Ser. B 39, 1-38.

Epps, T.W., and Epps, M.L., 1976, "The stochastic dependence of security price changes and transaction volumes: implications for the mixture-of-distribution hypothesis", Econometrica 44, 305-321.

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.

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

Hasbrouk, Joel and Seppi, D.J., 2001, "Common factors in prices, order flows, and liquidity", Journal of Financial Economics 59, 383-411.

Heimstra, C., and J.D. Jones, 1994, "Testing for linear and nonlinear granger causality in the stock price-volume relation", *Journal of Finance*, 49, 1639-1664.

Johnson, R.A., and Wichern, D.W., 1998, "Applied Multivariate Statistical Analysis", 4th edition, Prentice Hall.

Jones, C.M., Kaul, G., and Lipson, M.L., 1994, Information, trading, and volatility, *Journal of Financial Economics* 36, 127-154.

Karpoff, J.M., 1987, "The relation between price changes and trading volume: A survey", Journal of Financial and Quantitative Analysis 22, 109-126.

Kyle, A.S., 1985, "Continuous auctions and insider trading", Econometrica 53, 1315-1335.

Lamoureux, C.G., and Lastrapes, W.D., 1994, "Endogenous trading volume and momentum in stock-return volatility", Journal of Business & Economic Statistics 12, 253-260.

Le., N.D., Martin, R.D., and Raftery, A.E., 1996, "Modeling flat stretches, bursts, and outliers in time series using mixture transition distribution models", Journal of the American Statistical Association 91, 1504-1515.

Lo, A., Mamaysky, H. and Wang, J., 2000, "Foundations of technical analysis: computational algorithms, statistical inference, and empirical implementation", Journal of Finance 55, 1705-1770.

Lo, A., Wang, J., 2000, "Trading volume:definitions, data analysis, and implications of portfolio theory", Review of Financial Studies 13, 257-300.

Louis, T.A., 1982, "Finding observed information using the EM algorithm", Journal of the Royal Statistical Society, Ser. B 44, 98-130.

Meilijson, I. , 1989, "A fast improvement to the EM algorithm on its own terms", Journal of the Royal Statistical Society, Ser. B 51, 127-138.

O'Hara, M., 1995, "Market microstructure theory", Blackwell Publishers, Cambridge, MA.

Richardson, M., and T. Smith, 1994, "A direct test of the mixture of distributions hypothesis: measuring the daily flow of information", *Journal of Financial and Quantitative Analysis*, 29, 101-116.

Tanner, M.A., 1993, "Tools for statistical inference", New York: Springer-Verlag.

Tauchen, G. E. and M. Pitts, 1983, "The price variability-volume relationship on speculative markets", Econometrica 51, 485-505.

Roll, R., 1992, "Industrial structure and the comparative behavior of international stock market indices", Journal of Finance 47, 3-41.

Wu, C.F.J., 1983, "On the convergence properties of the EM algorithm", The Annals of Statistics 11, 95-103.

Table 1

Returns and volume

The sample is the 30 Dow Jones stocks from April 1 to June 30, 1998. The means and standard deviations are calculated across the entire sample period.

		Returns (%)		Volume	
Ticker	Name	Mean	Std. Dev.	Mean	Std.Dev.
AA	Alcoa	-0.03361	0.33217	66232.56	71056.96
ALD	Allied Signal	-0.00982	0.50798	90438.08	66779.48
AXP	American Express	-0.00573	0.45492	99180.96	74175.61
BA	Boeing	-0.00665	0.39603	197745.4	129947.5
C	Citigroup	-0.02708	0.50269	294690.6	368115.8
CAT	Caterpillar	-0.02672	0.47116	74050.44	59534.83
CHV	Chevron	-0.00781	0.35952	80350	58059.65
DD	DuPont	0.012513	0.51166	183316.7	153423.5
DIS	Disney	-0.0132	0.3727	114635.9	97870.68
EK	Eastman Kodak	-0.00862	0.41029	80530.52	85130.27
GE	General Electric	-0.01449	0.35931	231183.9	110125.6
GM	General Motors	-0.02618	0.39003	164503.2	110380.3
GT	Goodyear	-0.03168	0.40141	36901.6	46855.73
HWP	Hewlett-Packard	-0.00908	0.56935	253428.1	237430.2
IBM	IBM	-0.00573	0.41158	252827.2	188142.5
IP	International Paper	-0.05164	0.46609	68185.32	66177.55
JNJ	Johnson & Johnson	-0.0171	0.38645	125324.7	73923.52
JPM	JP Morgan	-0.02957	0.40285	53357.12	38624.04
KO	Coca Cola	0.011703	0.34704	173883	92411.77
MCD	McDonalds	0.018305	0.36818	150667.2	125984.6
MMM	3M	-0.01373	0.33819	54731.4	49581.39
MO	Phillip Morris	-0.05604	0.48032	502051.3	453638.3
MRK	Merck	-0.02061	0.37346	162276.5	99498.18
PG	Proctor Gamble	0.011857	0.42987	129214.7	89777.67
S	Sears	0.011988	0.4664	77875.87	57938.14
T	AT&T	-0.02231	0.38216	274291	160592.4
UK	Union Carbide	-0.0305	0.53421	42420.06	44979.81
UTX	United Tech	-0.01027	0.39242	48563.37	42608.85
WMT	WalMart	0.00776	0.42697	166220.9	98751.79
XON	Exxon	0.001859	0.35154	193425.7	101390.8

Table 2

Canonical correlation analysis results for returns and volume

The sample is the 30 Dow Jones stocks from April 1 to June 30, 1998. The second column provides canonical correlations. The third column provides canonical correlation test results. The numbers in the parentheses are p-values (We only provide significant results).

Canonical Correlations	Test Results
0.465	1.31 (<.0001)
0.441	1.21 (<.0001)
0.389	1.10 (0.0258)

Table 3

Canonical correlation analysis results for return volatility and volume

The sample is the 30 Dow Jones stocks from April 1 to June 30, 1998. The second column provides canonical correlations. The third column provides canonical correlation test results. The numbers in the parentheses are p-values.

	Canonical Correlations	Test Results
1	0.649	4.36 (<.0001)
2	0.622	4.16 (<.0001)
3	0.588	3.99 (<.0001)
4	0.570	3.85 (<.0001)
5	0.554	3.71 (<.0001)
6	0.535	3.58 (<.0001)
7	0.528	3.46 (<.0001)
8	0.506	3.32 (<.0001)
9	0.473	3.19 (<.0001)
10	0.464	3.09 (<.0001)
11	0.443	2.98 (<.0001)
12	0.424	2.88 (<.0001)
13	0.416	2.78 (<.0001)
14	0.378	2.66 (<.0001)
15	0.367	2.59 (<.0001)
16	0.351	2.50 (<.0001)
17	0.320	2.42 (<.0001)
18	0.314	2.37 (<.0001)
19	0.299	2.29 (<.0001)
20	0.278	2.20 (<.0001)
21	0.271	2.13 (<.0001)
22	0.248	2.00 (<.0001)
23	0.213	1.87 (<.0001)
24	0.204	1.81 (0.0005)
25	0.175	1.68 (0.0069)
26	0.155	1.60 (0.0304)
27	0.151	1.50 (0.0924)
28	0.108	0.96 (0.4706)
29	0.038	0.24 (0.9143)
30	0.004	0.01 (0.9099)

Table 4

Estimation Results for the Dow Jones 30 Stocks

Table 4 presents the estimation results of the model (9) for each of the 30 Dow Jones stocks from April 1 to June 30, 1998. Returns are log differences of mid-quote changes at half-hour intervals. Returns and dollar volume are used in the estimation. Panel A reports the results for the stocks with 3 common factors; Panel B reports the results for the stocks with 2 common factors; Panel C reports the results for the stocks with one common factor. All the estimates are significant at 5% level (we do not report the standard error).

	α				$\sigma_1^2\,(10^{-4})$			
Panel A	α_0	α_1	α_2	α_3	σ_{10}^2	σ_{11}^2	σ_{12}^2	σ_{13}^2
EK	0.58	0.30	0.02	0.10	0.034	0.036	0.100	0.116
HWP	0.57	0.29	0.10	0.04	0.087	0.087	0.184	0.294
IP	0.17	0.46	0.31	0.06	0.029	0.035	0.075	0.155
KO	0.78	0.09	0.08	0.05	0.050	0.080	0.085	0.131
MMM	0.53	0.20	0.14	0.13	0.025	0.018	0.033	0.081
MO	0.54	0.29	0.11	0.06	0.054	0.054	0.122	0.193
S	0.74	0.12	0.09	0.05	0.045	0.039	0.128	0.159
UTX	0.68	0.17	0.11	0.04	0.033	0.021	0.092	0.123
XON	0.57	0.15	0.20	0.08	0.032	0.040	0.078	0.121

Table 4, **Panel A** (Continued)

	$\mu\,(10^3)$				$\sigma_2^2\,(10^8)$			
Panel A	μ_0	μ_1	μ_2	μ_3	σ_{20}^2	σ_{21}^2	σ_{22}^2	σ_{23}^2
EK	30.42	30.18	23.03	14.34	0.92	1.13	0.03	0.86
HWP	131.73	131.63	80.84	44.22	19.61	19.48	6.59	7.16
IP	21.71	23.06	18.87	8.86	0.05	0.51	0.89	0.49
KO	131.75	86.60	66.52	43.16	18.04	1.77	1.06	0.99
MMM	22.09	21.99	20.49	9.66	0.60	0.08	0.70	0.42
MO	261.79	261.58	157.86	92.58	73.30	73.19	17.69	49.81
S	28.63	27.12	18.21	9.85	0.99	0.02	0.32	0.42
UTX	17.56	17.58	10.38	5.58	0.40	0.04	0.11	0.11
XON	133.38	110.50	80.71	46.28	13.12	0.94	3.34	3.42

Table 4 (Continued)

	α			$\sigma_1^2\,(10^{-4})$			$\mu\,(10^3)$			$\sigma_2^2\,(10^8)$		
Panel B	α_0	α_1	α_2	σ_{10}^2	σ_{11}^2	σ_{12}^2	μ_0	μ_1	μ_2	σ_{20}^2	σ_{21}^2	σ_{22}^2
AXP	0.86	0.12	0.02	0.051	0.134	0.171	50.57	25.44	13.56	3.32	1.17	0.06
DD	0.57	0.28	0.15	0.091	0.092	0.217	111.90	111.00	55.33	14.34	13.97	5.99
GM	0.58	0.24	0.18	0.048	0.048	0.123	106.05	105.91	51.73	11.70	11.69	11.33
GT	0.59	0.30	0.11	0.032	0.033	0.123	12.37	12.35	5.19	0.17	0.19	0.18
JNJ	0.61	0.25	0.14	0.054	0.055	0.136	82.61	82.18	39.60	7.21	7.16	5.07
JPM	0.86	0.09	0.05	0.039	0.101	0.129	25.57	15.08	8.46	0.65	0.23	0.18
MCD	0.64	0.22	0.14	0.036	0.037	0.114	74.75	74.67	33.56	6.26	6.31	6.01
PG	0.70	0.16	0.14	0.064	0.082	0.162	89.15	77.71	42.78	9.75	4.63	5.17
WMT	0.61	0.28	0.11	0.070	0.076	0.186	119.26	113.12	51.45	17.64	16.33	12.16

	α		$\sigma_1^2\,(10^{-4})$		$\mu\,(10^3)$		$\sigma_2^2\,(10^8)$	
Panel C	α_0	α_1	σ_{10}^2	σ_{11}^2	μ_0	μ_1	σ_{20}^2	σ_{21}^2
AA	0.87	0.13	0.047	0.077	33.00	13.60	1.38	0.32
ALD	0.84	0.16	0.103	0.254	59.28	28.40	3.96	2.41
BA	0.86	0.14	0.067	0.139	139.53	68.19	22.75	6.11
C	0.88	0.12	0.077	0.228	121.60	51.40	17.97	6.90
CAT	0.83	0.17	0.082	0.196	43.55	21.99	2.99	0.84
CHV	0.84	0.16	0.062	0.097	58.08	28.28	3.98	1.33
DIS	0.85	0.15	0.069	0.100	80.98	41.26	8.07	1.92
GE	0.85	0.15	0.074	0.127	205.51	108.39	42.72	12.26
IBM	0.89	0.11	0.077	0.125	178.47	96.90	44.23	11.67
MRK	0.90	0.10	0.074	0.119	120.36	59.70	17.50	5.85
T	0.78	0.22	0.080	0.130	237.11	127.93	58.62	23.21
UK	0.90	0.10	0.096	0.209	21.05	8.70	0.64	0.23

Table 5

Estimates of Regressions of Returns on Trading Volume

The sample is the 30 Dow Jones stocks from April 1 to June 30, 1998. The dependent variable is original return series or adjusted return series of information. The independent variable is trading volume. The regressions are separately run for each of the 30 stocks using Ordinary Least Square (OLS) estimation. The standard errors are calculated by the way of Jones et al. (1994). The second column provides results for the original return series. The third column provides results for the adjusted return series.

Estimate	Before Adjustment of Information ($\times 10^{-10}$)	After Adjustment of Information ($\times 10^{-10}$)
Mean Coefficient	6.08	-0.56
Standard Error	2.80	2.17
t-value	2.17	-0.26

Table 6

Estimates of GARCH (1,1) for Returns

The sample is the 30 Dow Jones stocks from April 1 to June 30, 1998. The GARCH (1,1) model is:

$$y_t = a_0 + a_a y_{t-1} + \varepsilon_t$$
$$\varepsilon_t \sim N(0, h_t)$$

and

$$h_t = \gamma_0 + \gamma_1 \varepsilon_{t-1}^2 + \gamma_2 h_{t-1}$$

Only those cases in which GARCH persistence is evident in the return series are reported. The third column provides results for the original return series. The fourth column provides results for the adjusted return series.

Ticker	Estimate	Before Adjustment of Information	After Adjustment of Information
ALD	γ_1	0.0489	0.0867
AXP	γ_1	0.5935	–
BA	γ_1	0.4069	–
C	γ_1	0.6339	0.2293
CAT	γ_1	0.8667	–
	γ_2	0.0294	–
CHV	γ_1	0.4401	–
DIS	γ_1	0.5130	–
EK	γ_1	1.5884	–
GE	γ_1	0.3182	–
GM	γ_1	0.0599	–
GT	γ_1	9.9782×10^{-7}	–
HWP	γ_1	–	0.3671
IP	γ_1	0.1181	–
IBM	γ_1	0.2107	0.1688

Table 6 (Continued)

Ticker	Estimate	Before Adjustment of Information	After Adjustment of Information
JPM	γ_1	0.0633	–
JNJ	γ_2	9.6513×10^{-7}	–
MCD	γ_1	0.3636	–
MMM	γ_1	0.2100	–
MO	γ_1	1.4881	–
MRK	γ_1	0.1440	–
PG	γ_1	0.0781	–
S	γ_1	0.0667	–
T	γ_1	0.4454	–
UTX	γ_1	0.2706	–
WMT	γ_1	0.0767	–