Commonality and Information in Return/Volatility−Volume Relations

Abstract

This paper develops a multi-asset mixture distribution hypothesis model to investigate commonality in stock returns and trading volume. The model that characterizes how factor structures stem from information effects predicts: first, the factor structure of stock returns is uncorrelated to the factor structure of trading volume although factor impacts on returns and volume share a common dynamic latent information variable. Second, positive volatility-volume linkages exist across assets. The empirical analyses using intraday data on the 28 Dow Jones stocks identify a one-factor structure in both stock returns and trading volume, and the results confirm the model's predictions.

Commonality and Information in Stock Return/Volatility–Volume Relations

1. Introduction

The object 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 model to examine how stock returns and trading volume move together.

Several implications emerge from the model. First, factor structures of both stock returns and trading volume are equilibrium outcomes of information effects. Second, the factor impacts on stock returns and trading volume depends on an underlying latent information variable. Third, the return factor structure is independent of the factor structure of trading volume. Finally, the covariance structure of stock returns is positively related to volume across stocks, and the cross-sectional positive relations result from the latent information variable.

We use half-hour intraday data for Dow Jones stocks over a sample period from April through June, 2007, to fit our common-factor mixture distribution hypothesis model. We apply the method in Lamoureux and Lastrapes (1994) to extract the latent information flow, and the expectation-maximization (EM) algorithm to estimate the model's unobservable factor variables as well as its parameters. We use Akaike's information criterion (AIC) and the Bayesian information criterion (BIC) for model selection. These criteria suggest a one-factor structure in both stock returns and trading volume.

To examine return-volume independence across stocks, we conduct Pearson's correlation test on each return-volume pair for the Dow Jones stocks in our sample. We find that only 13% of return-volume correlations are statistically significant; all the sample stocks show significant return-return and volume-volume correlations across firms, however. We also perform a canonical correlation analysis to examine the positive cross-asset relations between return volatility and volume. The empirical results confirm the implications of our model.

In addition, as the latent information flow plays an important role in price formation in our model, and it is well-known that returns exhibit ARCH (GARCH) dynamics, we remove the information content from returns to see whether return behavior changes. While 13 stocks in our sample show ARCH (GARCH) patterns, and these patterns all disappear after we control for the impact of information on returns.

Our study is inspired by other research in modern portfolio theory and market microstructure theory. Studies of cross-sectional behavior of stock returns have long been separate from work on equity trading volume.[1] The primary study on joint return-volume cross-sectional patterns is Hasbrouck and Seppi (2001), which uses a statistical factor model to examine co-movements in prices and order flows. Their study suggests that the factor structure of valuation and liquidity fundamentals may account for cross-firm commonalities.

More recently, Bernhardt and Taub (2008) present a speculative trading model including a K-fundamental structure in formulating asset value. They provide some analytical characterizations of how the informational structure of the market matters in prices and order flows.

These studies motivate our construction of a theoretical model to unify these lines of research. The model is a multi-asset extension of the mixture distribution hypothesis (MDH) model proposed by Tauchen and Pitts (1983). They derive a joint distribution of price change

---

[1] For studies on stock returns, see, for example, Markowitz (1952), Ross (1978), and Brenner, Pasquariello and Subrahmanyam (2009). For work on trading volume, see, for example, Lo and Wang (2000) and Tookes (2008).

and trading volume for a single security by imposing a variance-component structure on information effects. To incorporate possible interactions among securities, we add K common-factor components to the variance decomposition. These K common-factor components, consistent with the conjectures of Hasbrouck and Seppi (2001), represent various aspects of common information effects on trading activity (such as spillover or market effects.[2])

2. Common-Factor MDH Model

In our multi-asset mixture distribution hypothesis model, J investors choose to trade a portfolio of M assets. Following Tauchen and Pitts (1983), we assume that J is high and fixed within a particular time period.[3] Each time new information, including macroeconomic or liquidity-motivated information, arrives in the market, trader j will take a long or short position in each asset in the portfolio, depending on his or her reservation price (P_j^*) of that asset and the asset's market price (P).[4] We assume that traders will never change the portfolio components when they rebalance their portfolios. This assumption is less restrictive than that of Lo and Wang (2000).

Thus, trader j's equilibrium position, Q_j, of asset m and the aggregate trading volume, V_j, of asset m are given by:

[2] See, for example, Fleming, Kirby and Ostdiek (1998), and Kodres and Pritsker (2002).
[3] A fixed J is for simplicity, and a high J is for normality of trading volume.
[4] It is impossible to distinguish between informed and liquidity trades in our settings.

$$Q_{i,j} = c\left(P_{i,j}^{*} - P_i\right)$$
$$V_i = \frac{c}{2}\sum_{j=1}^{J}\left| Q_{i,j} - Q_{i-1,j} \right|$$

where c is a positive constant, and the subscript i indexes equilibrium phases.[5]

Given the equilibrium condition, $\sum_{j=1}^{J} Q_{i-1,j} = 0$, we obtain the equalities:

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

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

The single-security MDH model in Tauchen and Pitts' (1983) incorporates two components in the change in trader j's reservation price. One component, ϕ_i, is common to all investors trading the stock; the other, ψ_{ij}, is specific to trader j. As we are dealing with a multi-asset market, we extend Tauchen and Pitts' decomposition by introducing a set of common components, $\xi_{ijk}, k = 1, 2, ..., K$. Now, we have:

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

[5] To simplify notations, we omit index *m* in model derivations.

All components in this equation have mutually independent normal distributions with a mean of zero. We further assume that these components are independent across securities and through time. Clearly, the first two components stand for the idiosyncratic effects of the information, and the third component, $\sum_{k=1}^{K}\xi_{ijk}$, represents the *K*-factor structure embedded in the information effects. We are thus able to capture all types of trading activity, whether publicly informed, strategically informed or liquidity trades.

Summing up the price changes and trading volume across all traders, we have:

$$\Delta P_i = (\phi_i + \overline{\psi}_i) + \sum_{k=1}^{K}\overline{\xi}_{ik}$$

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

where $\overline{\psi}_i = \frac{1}{J}\sum_{j=1}^{J}\psi_{ij}, \quad \overline{\xi}_{ik} = \frac{1}{J}\sum_{j=1}^{J}\xi_{ijk}$

Proposition 1: *(a) The price change* ΔP_i *is normally distributed with mean zero and variance* $\sigma_{\Delta P}^2$*. (b) For high J, the volume* V_i *is asymptotically normally distributed with mean* μ *and variance* σ_V^2*. (c) The price change* ΔP_i *and trading volume* V_i *are stochastically independent.*

Proposition 1(a) and 1(b) are evident.[6] Proposition 1(c) follows because the generation of trading volume eliminates the component (ϕ_i) common to all traders of a certain stock; also, components $\overline{\psi}_i$ and $\sum_{k=1}^{K}\overline{\xi}_{ik}$ are independent of their respective deviations from the 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 over a fixed time interval yields the formulas for price changes and volume:

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

where I is the number of the information flow within the fixed time interval. Standardizing notation, we have the equations:

$$\Delta P = \sigma_{10}\sqrt{I}\,z_{10} + \sum_{k=1}^{K}\sigma_{1k}\sqrt{I}\,z_{1k}$$
$$V = \left(\mu_0 I + \sigma_{20}\sqrt{I}\,z_{20}\right) + \sum_{k=1}^{K}\mu_k I + \sum_{k=1}^{K}\sigma_{2k}\sqrt{I}\,z_{2k}$$

where the z_{1k} and z_{2k} $(k = 0, ..., K)$ are all standard normal. All z are mutually independent, and independent of I.

Generalizing these equations for M stocks in matrix form, we have:

---

[6] See Tauchen and Pitts (1983).

$$\Delta P = \varphi G\sqrt{I} + \eta\sqrt{I}$$
$$V = \mu I + \theta F\sqrt{I} + \varepsilon\sqrt{I}$$

where

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

$$\varphi = \begin{bmatrix} \sigma_{111} & \cdots & \sigma_{1k1} & \cdots & \sigma_{1K1} \\ \vdots & \ddots & & & \vdots \\ \sigma_{11m} & & \sigma_{1km} & & \sigma_{1Km} \\ \vdots & & & \ddots & \vdots \\ \sigma_{11M} & \cdots & \sigma_{1kM} & \cdots & \sigma_{1KM} \end{bmatrix}_{M\times K} \quad \boldsymbol{\theta} = \begin{bmatrix} \sigma_{211} & \cdots & \sigma_{2k1} & \cdots & \sigma_{2K1} \\ \vdots & \ddots & & & \vdots \\ \sigma_{21m} & & \sigma_{2km} & & \sigma_{2Km} \\ \vdots & & & \ddots & \vdots \\ \sigma_{21M} & \cdots & \sigma_{2kM} & \cdots & \sigma_{2KM} \end{bmatrix}_{M\times K}$$

$$\boldsymbol{\eta} = \begin{bmatrix} \sigma_{101} z_{101} \\ \vdots \\ \sigma_{10m} z_{10m} \\ \vdots \\ \sigma_{10M} z_{10M} \end{bmatrix}_{M\times 1} \quad \boldsymbol{\varepsilon} = \begin{bmatrix} \sigma_{201} z_{201} \\ \vdots \\ \sigma_{20m} z_{20m} \\ \vdots \\ \sigma_{20M} z_{20M} \end{bmatrix}_{M\times 1}$$

We call the model in matrix form a common-factor MDH model. Vectors of both price changes and of trading volume are governed by three groups of mutually independent variables: idiosyncratic variables ($\boldsymbol{\eta}$ and $\boldsymbol{\varepsilon}$), common factor variables (**G** and **F**) and the mixing variable (***I***). From Proposition 1, we know that **E**(**G**) = **E**(**F**) = 0, **Cov**(**G**) = **Cov**(**F**) = **I**, where **I** is an identity matrix.

Note that the latent information variable enters the factor loadings of both price changes and volume. This gives a dynamic feature to the factor impacts on returns and on volume. Conditional on the mixing variable, the common-factor MDH model reduces to the statistical factor model in Hasbrouck and Seppi (2001). If $K = 0$, our MDH model reduces to the single-security MDH model in Tauchen and Pitts (1983).

Proposition 2: *The cross-sectional interactions among price changes and those among trading volume depend upon the underlying latent information flow.*

Proof:

$$\begin{aligned} \mathrm{Cov}\{\Delta P, \Delta P\} &= E\{\Delta P \Delta P^{T}\} - E\{\Delta P\}[E\{\Delta P\}]^{T} \\ &= E\{(\varphi I^{1/2})\, GG^{T}(\varphi I^{1/2})^{T} + (\eta I^{1/2})\,(\eta I^{1/2})^{T}\} \\ &= E\{\varphi\varphi^{T} I + \eta\eta^{T} I\} \\ &= (\varphi\varphi^{T} + \eta\eta^{T})E\{I\} \neq 0 \end{aligned}$$

as long as $E\{I\} \neq 0$

$$\begin{aligned} \mathrm{Cov}\{V, V\} &= E\{VV^{T}\} - E\{V\}[E\{V\}]^{T} \\ &= E\{(\theta I^{1/2})\, FF^{T}(\theta I^{1/2})^{T} + (\varepsilon I^{1/2})(\varepsilon I^{1/2})^{T}\} \\ &= E\{\theta\theta^{T} I + \varepsilon\varepsilon^{T} I\} \\ &= (\theta\theta^{T} + \varepsilon\varepsilon^{T})E\{I\} \neq 0 \end{aligned}$$

as long as $E\{I\} \neq 0$

Q.E.D.

Proposition 3: *The factor structure of price changes is uncorrelated to the factor structure of trading volume.*

Proof:

$$\begin{aligned} \mathrm{Cov}\{\varphi I^{1/2}G,\ \theta I^{1/2}F\} &= E\{\varphi I^{1/2}G(\theta I^{1/2}F)^{T}\} - E\{\varphi I^{1/2}G\}[E\{\theta I^{1/2}F\}]^{T} \\ &= E\{\varphi\theta I G F^{T}\} = 0 \end{aligned}$$

Q.E.D.

Proposition 4: *The covariance structure of return is positively related to trading volume across firms as long as the mixing variable shows variation.*

Proof:

$$\begin{aligned}\mathbf{Cov}\{\Delta P_i \Delta P_l, V_j\} &= \mathbf{E}\{\Delta P_i \Delta P_l V_j\} - \mathbf{E}\{\Delta P_i \Delta P_l\}\mathbf{E}\{V_j\} \\ &= \left\{\sum_{k=0}^{K} \sigma_{1ki}\sigma_{1kl}\mu_{kj}\right\}\left[\mathbf{E}\{I^2\} - \mathbf{E}^2\{I\}\right] + \mathbf{E}\left\{\sum_{k=1}^{K} \sigma_{1ki}\sigma_{1kl}\sigma_{2kj} z_{1ki} z_{1kj} z_{2kj}\right\}\mathbf{E}\left\{I^{3/2}\right\} \\ &= \left\{\sum_{k=0}^{K} \sigma_{1ki}\sigma_{1kl}\mu_{kj}\right\}\mathbf{Var}\{I\} > 0\end{aligned}$$

as long as $\mathbf{Var}\{I\} \neq 0$.

where subscripts i, j and l index stocks.

Q.E.D.

Proposition 4 indicates that there are positive relations between return volatility and volume not only within a stock but also across stocks.

3. Data Analysis

3.1 Data

The sample includes 28 stocks in the Dow Jones Industrial Average 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 Dow Jones stocks in order to increase the possibility of detecting common factors and to mitigate the nonconcurrent problem.

The transaction data is from NYSE Trade and Quote (TAQ) database. 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 mid-quotes 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, transaction data for the first three minutes of trading are deleted. 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 Dow Jones stock over the entire sample period. General Motors has the highest mean 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).

<Table 1 inserted here>

3.2 Preliminary Analyses

We conduct various correlation analyses to justify the model specifications. Specifically, we need to determine whether the empirical data conform to the implications of propositions two through four.

First, 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. Through canonical correlation analysis, not only can we determine directions of the relation between return volatility and trading volume, but we can also explore whether the relation holds across firms.

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

Table 2 presents the canonical correlation analysis results. Half of the canonical correlations are statistically significant, positive, and less than 1.0. This confirms the cross-asset positive relations between return volatility and volume.

<Table 2 inserted here>

Next, we undertake a simple pair-wise Pearson correlation test to see whether returns and volume are cross-correlated. With 28 stocks in the sample, we have respective 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 reported here).

The canonical redundancy analysis in Table 3 shows that neither the return nor volume canonical variables are good overall predictors of the opposite set of variables. The proportions of variance in returns explained by trading volume are all below 0.02, and vice versa.

<Table 3 inserted here>

3.3 Estimation of the Model

Given the encouraging results in section 3.2, we now proceed to estimate the model. Two main empirical issues arise when using the maximum likelihood estimation (MLE) method. First, the joint distribution of returns and volume is not a multivariate normal distribution unless conditioned on the mixing variable *I*. Second, factors and the mixing variable are unobservable. To address these issues, we follow a natural three-stage estimation strategy by combining the information extraction method in Lamoureux and Lastrapes (1994) and the EM algorithm.[7]

In the first stage, we minimize the conditional moment criterion from Lamoureux and Lastrapes (1994) using the current parameter estimates of the model to extract the information flow. Then we normalize the original return and volume series by eliminating the extracted information flow.

With the normalized return and volume series, our second and third stages are just typical expectation and maximization steps of the EM algorithm for factor analysis. We use Akaike's information criterion (AIC) and the Bayesian information criterion (BIC) to choose the number of factors. Both criteria favor models with low AIC and BIC values.[8] The best model our computed AIC and BIC values indicate has a one-factor structure. All the point estimates of parameters in the one-factor MDH model are statistically significant at the 5% level (not reported here).

3.4 Impact of Information Flow

The underlying information flow plays an important role in cross-firm variations, so we further explore characteristics of the impact using the extracted information flow $\hat{I}_t$. We start the investigation by examining whether the information flow accounts for the serial

---

[7] See, for example, Dempster, Laird a nd Rubin (1977); Louis (1982);Rubin and Thayer (1982); Meng and Rubin (1991); Dyk, Meng and Rubin (1995).
[8] For example, see Tanner (1993).

dependence in returns. Given the extracted information I_t series, $\hat{I}_t$, we adjust the return series for the information flow by:

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

If, as our model implies, the underlying information drives stock returns, the serial dependence of the adjusted return series should disappear. We estimate a GARCH(1, 1) model for raw and adjusted returns of each sample stock, respectively. The GARCH (1, 1) model is as follows:

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

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 4 displays the estimation results (reporting only those where GARCH persistence is evident). The third column reports the estimates of the GARCH (1, 1) model for raw return series. We find that 13 out of the 28 stocks present persistence in variance. The fourth column shows the estimates of the GARCH (1, 1) model for the adjusted return series. After we control for the information effect in raw return series, none of the 28 stocks exhibits persistence in return variance.

<Table 4 inserted here>

Another aspect of the information flow that we can explore is from the prediction of proposition 4, which states that the information flow is the sole determinant of the positive volatility-volume relations. In other words, if we separate the impact of the information flow from return and volume, trading volume will have no explanatory power for return volatility.

We implement a simple regression approach to test this prediction. The regression model takes the following form:

$$Y = \alpha + \beta X + \varepsilon$$

where the dependent variable is return volatility, and the independent variable is trading volume. We run the regression model respectively for the original and the information-adjusted series. The information-adjusted volume is given by $V_{adj} = \frac{V - \hat{\mu}\hat{I}}{\sqrt{\hat{I}}}$.

Table 5 summarizes the results. The results for the original series in the second column reveal a positive and significant volatility-volume relation for 24 stocks, but the majority of transformed volume series appear to have no significant explanatory power for adjusted return volatility (see results in column 3). This differential result supports the prediction of proposition 4. The result also indicates that volume is a rough proxy for information flow.

In practice, the implied volatility index (VIX) is perceived as a benchmark for stock market volatility. Empirical studies have been conducted to test if VIX is a superior informative variable.[9] We perform a similar study by comparing the time series plots of the VIX and the extracted information flow, and find that the time-varying features of both series

---

[9] See, for example, Blair et al. (2001); Degiannakis and Floros (2010).

are quite different. We also find that the correlation between the VIX and estimated information flow is low (0.29) albeit significant.

<Figure 1 inserted here>

4. Summary

We have specified a multivariate mixture distribution hypothesis model of returns and volume. In our model, information affects cross-firm variations in two ways: through impact structure (such as an industry or size or market effect) and through the amount of information within a certain time interval. Using half-hour intraday data for Dow Jones stocks, the model seems to provide a useful framework that captures the important properties of the data. The estimated information flow appears to capture the heteroskedasticity in stock returns, although there is no pattern in the estimated factor loadings.. Therefore, the identity of the effect structure of the information remains unknown.

References

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

Blair, N. J., Poon, S. H. and Taylor, S. J., 2001, “Forecasting S&P 100 volatility: the incremental information content of implied volatilities and high-frequency index returns,” *Journal of Econometrics* 105, 5-26.

Brenner, M., Pasquariello, P. and Subrahmanyam M., 2009, “On the volatility and comovement of U.S. financial markets around macroeconomic news announcements,” *Journal of Financial and Quantitative Analysis* 44, No. 6, 1265-1289.

Degiannakis, S. and Floros, C., 2010, “VIX Index in interday and intraday volatility models,” J*ournal of Money, Investment and Banking*, 13, 21-26.

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.

Dyk, D. A., Meng, X. and Rubin, D. B., 1995, “Maximum likelihood estimation via the ECM algorithm: computing the asymptotic variance,” *Statistica Sinica*, 5, 55-75.

Fleming, J., Kirby, C. and Ostdiek, B., 1998, “Information and volatility linkages in the stock, bond, and money markets,” *Journal of Financial Economics* 49, 111-137.

Hasbrouck, J. and Seppi, D. J., 2001, “Common factors in prices, order flows, and liquidity,” *Journal of Financial Economics* 59, 383-411.

Kodres, L. and Pritsker, M., 2002, "A Rational Expectations Model of Financial Contagion," *Journal of Finance*, 57, 769-799.

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.

Lo, A. and 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.

Markowitz, H., 1952, “Portfolio selection,” *Journal of Finance* 7, 77-91.

Meng, X. and Rubin, D. B., 1991, “Using EM to obtain asymptotic variance-covariance matrices: the SEM algorithm,” *Journal of the American Statistical Association*, 86, No. 416, 899-909.

Ross, S., 1978, “Mutual fund separation in financial theory – the separating distributions,” *Journal of Economic Theory* 17, 254-286.

Rubin, D. B. and Thayer, D. T., 1982, “EM algorithm for ML factor analysis,” *Psychometrika*, 47, No.1, 69-76.

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

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

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

Table 1

Descriptive Statistics of Returns and of Trading Volume

The study sample includes data on 28 Dow Jones stocks from April 1, 2007, to 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	403281	396894
AIG	American International Group Inc.	-0.012	0.195	414760	234542
AXP	American Express Company	0.000	0.290	238669	154571
BA	Boeing Company	-0.010	0.275	187629	106685
C	Citigroup Incorporated	-0.015	0.300	720347	435264
CAT	Caterpillar Incorporated	-0.002	0.346	234245	151744
DD	DuPont	-0.015	0.305	230200	149527
DIS	Walt Disney Company	0.001	0.294	375479	208295
GE	General Electric Company	0.008	0.291	1315054	776955
GM	General Motors Corporation	0.037	0.456	475619	372010
HD	Home Depot	-0.014	0.312	503293	427210
HON	Honeywell International Inc.	0.005	0.315	272203	217163
HPQ	Hewlett-Packard Company	0.019	0.304	505981	256465
IBM	International Business Machines	0.008	0.267	326019	299917
JNJ	Johnson & Johnson	-0.013	0.214	449686	254764
JPM	JPMorgan Chase & Company	-0.027	0.313	529567	313201
KO	Coca-Cola Company	0.000	0.219	360966	200005
MCD	McDonald’s Corporation	-0.003	0.319	305497	177436
MMM	3 M Company	0.013	0.273	403787	242735
MO	Altria Group Incorporated	-0.024	0.239	180852	150207
MRK	Merck & Company, Incorporated	0.008	0.326	457424	337455
PFE	Pfizer Incorporated	-0.015	0.266	1235878	661629
PG	Proctor & Gamble Company	-0.007	0.200	787030	442186
T	AT&T Incorporated	-0.005	0.312	444908	258826
UTX	United Technologies Corporation	0.007	0.255	189547	100528
VZ	Verizon Communications Inc.	0.006	0.286	486536	324010
WMT	Wal-Mart Stores Incorporated	-0.011	0.271	595420	449108

XOM	Exxon Mobil Corporation	0.011	0.296	840262	362339

Table 2

Canonical Correlation Analysis Results for Return Volatility and Trading Volume

The second column reports canonical correlations for each of the 28 stocks in the study. The third column provides canonical correlation test results.

	Canonical Correlations	F Value (P Value)
1	0.819632	3.26 (<.0001)
2	0.722151	2.86 (<.0001)
3	0.691312	2.64 (<.0001)
4	0.677421	2.45 (<.0001)
5	0.638792	2.25 (<.0001)
6	0.604625	2.07 (<.0001)
7	0.549741	1.91 (<.0001)
8	0.529986	1.80 (<.0001)
9	0.474032	1.68 (<.0001)
10	0.467845	1.61 (<.0001)
11	0.440106	1.53 (<.0001)
12	0.43673	1.45 (<.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	0.24181	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)

Table 3

Canonical Redundancy Analysis Results for Return and Trading Volume

The second column reports for each of the 28 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.

	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	0
28	0	0

Table 4

Estimates of GARCH (1,1) for Returns

The GARCH (1,1) model for the sample is:

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

and

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

This table only reports cases in which GARCH persistence is evident. The third column provides results of the original return series. The fourth column provides results of the adjusted return series.

Ticker	Estimate	Before Adjustment of Information	After Adjustment of Information
AIG	γ_1	0.148	–
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	–

Table 5

Regression Results

For each of the 28 Dow Jones stocks in the sample, the regression model is $Y = \alpha + \beta X + \varepsilon$, where the dependent variable is return volatility, and the independent variable is trading volume. The second column presents the beta coefficients for the original series, and the third column reports the beta coefficients for the information-adjusted series.

Ticker	Original Series $\hat{\beta}$ (e-12)	Adjusted Series $\hat{\beta}$ (e-10)
AA	33.4 *	20.0*
AIG	4.55*	2.92
AXP	7.78	6.63
BA	38.6*	18.9
C	7.58*	2.56
CAT	45.2*	14.5
DD	13.4*	4.92
DIS	9.77*	4.23
GE	12.8*	7.31*
GM	30.3*	11.4
HD	11.4*	-3.64
HON	59.5*	-7.48
HPQ	27.0*	1.26
IBM	28.1*	-5.34
JNJ	11.6*	-3.58
JPM	19.8*	-4.91
KO	3.67	13.9
MCD	62.9*	3.08
MMM	-0.85	-6.52
MO	21.1*	-23.1*
MRK	15.8*	10.8
PFE	2.90*	-4.23
PG	3.76*	-2.17
T	4.81	-7.10
UTX	21.0*	8.85
VZ	14.9*	2.30
WMT	13.3*	7.46*
XOM	13.4*	-1.17

*Significant at the 5% level

Figure 1

Daily Time Series Plots of VIX and of Information Flow (thick line = VIX, thin line = square root of the extracted information flow)