Dynamic Volume-Return Relation of Individual Stocks

Guillermo Llorente
Universidad Autonoma de Madrid

Roni Michaely
Cornell University and IDC

Gideon Saar
New York University

Jiang Wang
MIT and National Bureau of Economic Research

We examine the dynamic relation between return and volume of individual stocks. Using a simple model in which investors trade to share risk or speculate on private information, we show that returns generated by risk-sharing trades tend to reverse themselves, while returns generated by speculative trades tend to continue themselves. We test this theoretical prediction by analyzing the relation between daily volume and first-order return autocorrelation for individual stocks listed on the NYSE and AMEX. We find that the cross-sectional variation in the relation between volume and return autocorrelation is related to the extent of informed trading in a manner consistent with the theoretical prediction.

Market participants carefully watch the volume of trade, which presumably conveys valuable information about future price movements. What we can learn from volume depends on why investors trade and how trades with different motives relate to prices. Two reasons are often mentioned for why investors trade: to rebalance their portfolios for risk sharing and to speculate on their private information. These two types of trades, which we call hedging and speculative trades, respectively, result in different return dynamics.

For example, when a subset of investors sell a stock for hedging reasons, the stock's price must decrease to attract other investors to buy. Since the expectation of future stock payoff remains the same, the decrease in the price causes a low return in the current period and a high expected return for the next period. However, when a subset of investors sells a stock for speculative reasons, its price decreases, reflecting the negative private information about

The authors thank I/B/E/S for providing the data on analysts' earnings forecasts, Yiming Qian for her research assistance, and Juan del Hoyo, David Easley, Gustavo Grullon, John Heaton, Soeren Hvidkjaer, Andrew Lo, N. Nimalendran, Maureen O'Hara, Bhaskaran Swaminathan, seminar participants at Cornell University, New York University, and the 1998 Western Finance Association meetings, and especially a referee for helpful comments. Address correspondence to Roni Michaely, 431 Sage Hall, Johnson Graduate School of Management, Cornell University, Ithaca, NY 14853, or e-mail: rm34@cornell.edu.

The Review of Financial Studies Fall 2002 Vol. 15, No. 4, pp. 1005–1047

its future payoff. Since this information is usually only partially impounded into the price, the low return in the current period will be followed by a low return in the next period, when the negative private information is further reflected in the price. This example shows that hedging trades generate negatively autocorrelated returns and speculative trades generate positively autocorrelated returns.

Intensive trading volume can help to identify the periods in which either allocational or informational shocks occur, and thus can provide valuable information to market observers about future price movements of the stock. In periods of high volume, stocks with a high degree of speculative trading tend to exhibit positive return autocorrelation and stocks with a low degree of speculative trading tend to exhibit negative return autocorrelation.

In this article, we construct a simple equilibrium model to derive the return dynamics generated when investors trade both to hedge and to speculate. The model illustrates that the relation of current return, volume, and future returns depends on the relative significance of speculative trade versus hedging trade. If speculative trading in a stock is relatively insignificant, returns accompanied by high volume tend to reverse themselves in the subsequent period. If speculative trading in a stock is significant, conditioned on high volume, returns become less likely to reverse and can even continue in the subsequent period. The difference in the relative importance of speculative trading among different stocks gives rise to the cross-sectional variation in their volume-return dynamics.

We empirically test the predictions of the model by analyzing the daily volume-return dynamics of individual stocks traded on the NYSE and AMEX. The basic structure of the empirical tests is as follows. For each stock in the sample, we use a time-series regression to find the relation between current return and volume and future return. Then, guided by the predictions of the model, we examine how this relation varies across stocks with the extent of speculative trading. We consider several proxies, such as market capitalization and bid-ask spread, for the degree of speculative trading. Consistent with the model, we find significant differences in the dynamics of returns and volume across stocks with different degrees of information asymmetry. Stocks of smaller firms, or stocks with higher bid-ask spreads, show a tendency for return continuation following high-volume days. Stocks of larger firms, or stocks with smaller bid-ask spreads, show almost no continuation and mostly return reversal following high-volume days.

We also test the hypothesis that it is firm-specific private information that affects the cross-sectional variation in the dynamic volume-return relation. We decompose both the volume and return series into systematic and unsystematic components. We find that the relation between information asymmetry and the influence of volume on the autocorrelation of returns persists when we remove the market-wide variations from the analysis.

Many recent articles investigate the relation between return dynamics and trading volume. Several of them focus on aggregate returns and volume [e.g., Duffee (1992), Gallant, Rossi, and Tauchen (1992), LeBaron (1992), and Campbell, Grossman, and Wang (1993) (henceforth CGW)]. These studies find that returns on high-volume days tend to reverse themselves.

A few studies also use returns and volume of individual stocks [e.g., Morse (1980), Conrad, Hameed, and Niden (1992), Antoniewicz (1993), Stickel and Verrecchia (1994)]. In particular, Antoniewicz finds that returns of individual stocks on high-volume days are more sustainable than are returns on low-volume days. Stickel and Verrecchia find that when earnings announcements are accompanied by higher volume, returns are more sustainable in the following days. The results of these two articles, from pooling together individual returns and volume, contrasts with the results from aggregate returns and volume. However, these studies do not provide an explanation that reconciles the two phenomena, nor do they examine the cross-sectional variation in the volume-return relation among the stocks.[1]

Our article provides a model that reconciles the contrasting empirical results on the volume-return relation at the aggregate and the individual levels. The model demonstrates how these results are related to the cross-sectional variation in the volume-return relation. Unlike CGW, we recognize information asymmetry as an important trading motive in addition to risk sharing. We show that for stocks with low information asymmetry [like the market indices and the big firms studied by CGW and LeBaron (1992)], returns following high-volume days exhibit strong reversals, as in CGW. However, for stocks with high information asymmetry, returns following high-volume days exhibit only weaker reversals or even continuations, which is consistent with the findings of Antoniewicz (1993) and Stickel and Verrecchia (1994), who use pooled returns and volume of individual stocks. Our analysis highlights the importance of information asymmetry in understanding the dynamic volume-return relation and demonstrates the more general nature of this relation, which was only partially captured by the aforementioned articles.

On the theoretical side, our model is very close to that of Wang (1994), but less complex. Simplification of the model allows us to obtain sharper predictions on the dependence of a stock's dynamic volume-return relation on the extent of information asymmetry, which we test empirically. Our article is also related to Blume, Easley, and O'Hara (1994), which examines the informational role of volume from the perspective of investors. In their model,

[1] In a recent article, Lee and Swaminathan (2000) show that past volume provides valuable information about future returns over horizons of six months. By assigning stocks to portfolios based on past volume and price changes, they show that using the prior six-months volume in combination with price changes is superior to using past price changes alone in predicting long-term returns. Specifically they show that buying low-volume winners and selling past high-volume losers outperforms a pure price momentum strategy. The nature of their result is similar to that of Antoniewicz (1993), but at a longer horizon.

investors can extract useful information from both volume and prices. In our model, volume provides no additional information to investors. Instead, we focus on how the dynamic relation between volume and returns allows observers (as opposed to participants) of the market to better understand its underlying characteristics, such as the degree of information asymmetry. Of course, in general, volume does provide additional information to investors as well. The impact of the information revealed by volume on price behavior can be complex, depending on the nature of the information. We return to these issues in Section 1.4.

Other related articles include Brown and Jennings (1989) and Grundy and McNichols (1989). In particular, Brown and Jennings examine the characteristics of return serial correlations when investors trade on private information. They show that information asymmetry among investors can affect the serial correlation in returns. They do not consider volume, which is partially exogenous in their model due to the presence of noise traders. Our objective is different from theirs. We focus on the joint behavior of return and volume and the information it provides about the underlying economy. Nonetheless, some of our results are related to theirs and we discuss the relation more specifically in Section 1.3.[2]

On the empirical side, our article is related to Hasbrouck (1988, 1991), who utilizes transactions data to examine the impact of trades on prices and quotes. He uses a linear empirical model to capture how such an impact might be related to the inventory control of specialists and to the private information behind trades. Even though we do not focus on the actual trading process, in many ways our article deals with the same issues, namely, how trades with different motives generate different return dynamics. But our article is different in several ways. First, our analysis focuses on the cross-sectional difference among stocks. Second, while Hasbrouck's analysis is based on heuristic linear specifications, our analysis is based on a specific theoretical model. Third, Hasbrouck's analysis conditions on a trade's direction, relying on a heuristic algorithm to infer its direction from publicly available data, while we use volume that is readily available.

We examine the robustness of our empirical results along several dimensions. First, we explore alternative econometric specifications of our tests and find that they do not change our results. Second, we examine the robustness of our results to data problems, such as bid-ask bounce and variations in the time of the last daily trade, and find that they do not change our conclusions. Third, we replace the daily interval with a measurement period that equates the amount of noise trading across stocks, and show that it does not affect our findings. Fourth, we show that our findings are not sensitive to

[2] It should also be noted that many theoretical articles on market microstructure deal with the impact of private information on asset prices [e.g., Glosten and Milgrom (1985), Kyle (1985), Easley and O'Hara (1987)]. However, the assumption of a competitive market maker in these models makes prices follow a martingale, eliminating richer dynamics.

alternative definitions of trading volume. Fifth, in light of recent articles that identify a larger analyst following with a smaller adverse selection problem, we use analyst following as an additional proxy for information asymmetry. Our results show that stocks that are followed by more analysts exhibit less return continuation following high-volume days.

The article proceeds as follows. Section 1 presents the model and the theoretical predictions that we test. Section 2 describes the empirical methodology. Section 3 provides our empirical results. Section 4 concludes. All proofs are contained in the appendix.

1. The Model

In this section we present a model of the stock market in which investors trade for both allocation and information reasons. We use the model to show how the dynamic relation between return and volume depends on the information asymmetry between investors. Since our goal is to establish this dependence and to illustrate the economic forces behind it, we keep the model as parsimonious as possible. We discuss possible generalizations of the model toward the end of the section.

1.1 Economy

The economy is defined on a discrete time sequence, $t = 0, 1, 2, \ldots$. There are two traded securities, a riskless bond and a stock. The bond is in unlimited supply at a constant, nonnegative interest rate, r. The stock pays a dividend D_{t+1} in period $t+1$, which consists of two additive components:

$$D_{t+1} = F_t + G_t. \tag{1}$$

Shares of the stock are traded in a competitive stock market. Let P_t denote the ex dividend price of the stock at time t.

There are two classes of investors, 1 and 2, with relative population weights of ω and $1-\omega$, respectively. Investors are identical within each class, but are different between the classes in their endowments and information. For convenience, an investor in class i is referred to as investor i, where $i = 1, 2$.

Each investor is initially endowed with $\bar{x}$ shares of the stock. He is also endowed with a flow of income from a nontraded asset. In period t, investor i has $Z_t^{(i)}$ units of the nontraded asset that pays N_{t+1} per unit in the subsequent period.

At time t, all investors observe the current dividend of the stock (D_t), its price (P_t), the current payoff of the nontraded asset (N_t), and their own endowment of the nontraded asset ($Z_t^{(i)}$ for investor i). They also observe F_t, the forecastable part of each stock's dividend next period. In addition, class 1 investors observe G_t. Thus they have private information about future stock payoffs. The information set of investor i at time t is then given by $\mathcal{I}_t^{(1)} = \{D, P, N, F, G, Z^{(1)}\}_{[0,t]}$ and $\mathcal{I}_t^{(2)} = \{D, P, N, F, Z^{(2)}\}_{[0,t]}$, where $\{\cdot\}_{[0,t]}$ denotes the history of a set of variables from time 0 to t. Investor i's information set can potentially contain other variables such as trading volume. However, as we show later (Section 1.2), in the current setting with only two classes of investors, volume provides no additional information to the investors who directly observe their own holdings.

Each investor maximizes expected utility over his wealth next period of the following form:

$$\mathrm{E}\left[-e^{-\lambda W_{t+1}} \middle| \mathcal{I}_t^{(i)}\right], \tag{2}$$

where $\lambda > 0$ is the risk-aversion parameter.

All shocks (i.e., $\{F_t, G_t, N_t, Z_t^{(1)}, Z_t^{(2)}\ \forall t\}$) are assumed to be normally distributed with zero mean and constant variances: σ_F^2 for F_t, σ_G^2 for G_t, σ_N^2 for N_t, and $\sigma_Z^{(i)2}$ for $Z_t^{(i)}$, respectively, where $i = 1, 2$. Furthermore, they are assumed to be mutually independent (contemporaneously and over time), except for D_t and N_t, which are correlated with $\mathrm{E}[D_t N_t] = \sigma_{DN}$. In addition, for convenience in exposition, we set the riskless interest rate at zero and each investor's initial endowment of stock shares at zero ($\bar{x} = 0$) (thus the total supply of the stock is zero). Without loss of generality, we set the investors' risk aversion λ at one.[3]

The model defined above captures two important motives for trading: allocation of risk and speculation on future returns. Each investor holds the stock and the nontraded asset in his portfolio. Since the returns on the two assets are correlated, as his holding of the nontraded asset changes, each investor wants to adjust his stock holdings to maintain an optimal risk profile. This generates allocational trade in the model, which we refer to as hedging trade.[4] In addition, some investors might have private information about future stock payoffs. As new private information arrives, they take speculative positions in the stock in anticipation of high returns. This generates the informational trade in the model that we refer to as speculative trade.

1.2 Equilibrium price and volume

Given a stock price process $\{P_t\}$, the dollar return on one stock share is given by

$$R_t \equiv D_t + P_t - P_{t-1} \quad (t = 1, 2, \ldots). \tag{3}$$

The return consists of two parts, a dividend and a capital gain. Let $\mathrm{E}_t^{(i)}[R_{t+1}]$ denote investor i's conditional expectation of R_{t+1} given his information at

[3] Our model, which uses constant absolute risk aversion preferences and normally distributed shocks, exhibits homotheticity. That is, the implications of the model are invariant to proportional scaling of the variances of all the shocks and the investors' risk aversion. Thus it is convenient to express the results to reflect this invariance. We choose to let $\lambda = 1$ and thank the referee for suggesting this.

[4] Many articles have introduced nontraded assets to generate investors' hedging needs to trade in the market. See, for example, Bhattacharya and Spiegel (1991) and Wang (1994).

t, $\sigma_R^{(i)2}$ its conditional variance, and $X_t^{(i)}$ his stock holding (here, $\sigma_R^{(i)2}$, the conditional variance of stock returns for investor i, has no time subscript because it remains constant over time). We have the following proposition:

Proposition 1. *The economy defined above has an equilibrium in which investor i's stock holding is*

$$X_t^{(i)} = \frac{\mathrm{E}_t^{(i)}[R_{t+1}]}{\sigma_R^{(i)2}} - \frac{\sigma_{DN}}{\sigma_R^{(i)2}} Z_t^{(i)} = \frac{\mathrm{E}_t^{(i)}[D_{t+1}] - P_t}{\sigma_R^{(i)2}} - \frac{\sigma_{DN}}{\sigma_R^{(i)2}} Z_t^{(i)} \quad (i = 1, 2) \quad (4)$$

and the ex dividend stock price is

$$P_t = F_t + \widetilde{P}_t \equiv F_t + a\, G_t - \left(b^{(1)} Z_t^{(1)} + b^{(2)} Z_t^{(2)}\right), \quad (5)$$

where $\mathrm{E}_t^{(1)}[D_{t+1}] = G_t$, $\mathrm{E}_t^{(2)}[D_{t+1}] = \gamma(\widetilde{P}_t - b^{(2)} Z_t^{(2)})$ *and* a, $b^{(1)}$, $b^{(2)}$, $\sigma_R^{(1)2}$, $\sigma_R^{(2)2}$, *and* γ *are constants.*

Each investor's stock holding has two components. The first component is proportional to his risk tolerance and the risk-adjusted, expected stock return given his information. This component reflects the optimal trade-off between the return and risk of the stock. The second component is proportional to the amount of his nontraded asset, and reflects his need to hedge the nontraded risk.

The equilibrium stock price at time t depends on F_t and G_t, and on the amounts of both investors' nontraded asset, $Z_t^{(1)}$ and $Z_t^{(2)}$, respectively. F_t gives the expected dividend next period, based on (nonprice) public information. G_t reflects class 1 investors' private information on the next dividend. $Z_t^{(1)}$ and $Z_t^{(2)}$ give the investors' need to use the stock to hedge their nontraded risk.

An investor changes his stock position when there is a change in his expectation of future stock returns or his exposure to nontraded risk. This generates trading in the market. Given that trading is only between the two classes of investors, the volume of trade, V_t, is given by the change in the total stock holdings of either class. Thus

$$V_t = \omega \left| X_t^{(1)} - X_{t-1}^{(1)} \right| = (1 - \omega) \left| X_t^{(2)} - X_{t-1}^{(2)} \right|. \quad (6)$$

It is worth pointing out that in the current setting, volume is simply proportional to the absolute changes in the stock holdings of each class of investors. For investor i, knowing his own holdings over time allows him to perfectly determine the level of volume in the market. In other words, volume provides no additional information to him given his information on asset payoffs, exposure to nontraded risk, and the stock price. This justifies the omission of volume in the investors' information set in our definition of equilibrium. Such a simplicity is unique to the situation when there are only two classes of investors and is intended to make our analysis tractable. In the more general situation with more than two classes of investors, volume provides information in addition to that in the price. We discuss this issue in more detail in Section 1.4.

1.3 Dynamic relation between return and volume

In our model, returns are generated by three separate sources: public information on future payoffs, investors' hedging trades, and their speculative trades. The returns generated by different sources exhibit different dynamics.

Returns generated by public news on future payoffs are independent over time. As public information about future dividends arrives (i.e., the realization of F_t at t), the stock price changes to fully reflect the public information. As Section 1.2 shows, the price change has no impact on investors' stock demands, despite changes in their wealth. As a result, expected future returns remain unchanged. In other words, public news on future payoffs results in a white noise component in stock returns. In addition, since it does not change investors' demand, it is not associated with abnormal volume.

Returns generated by trading are serially correlated. When investors trade for hedging reasons, the stock price adjusts to attract other investors to take the other side. This price change contains no information about the stock's future payoffs. Thus a price change generated by a hedging trade implies future returns of the opposite sign. For example, when class 1 investors sell the stock to hedge their nontraded risk, the stock price decreases, yielding a negative return for that period. However, the expected payoff in the next period stays the same. Hence the decrease in the price leads to an increase in the expected return in the next period. Thus returns generated by hedging trades tend to reverse themselves.

When investors trade for speculative reasons, the price changes to reflect the informed investors' expectation of the stocks' future payoffs. This expectation is fulfilled later on as private information becomes public. Thus a price change generated by speculative trade implies future returns of the same sign. For example, when class 1 investors sell the stock due to a negative signal on future stock dividends, the stock price decreases, yielding a negative return for the current period. Since the price only partially reflects the private information (in a non-fully revealing equilibrium), the return in the next period is more likely to be negative as the private information becomes public. Thus returns generated by speculative trade tend to continue themselves.

The actual dynamics of returns depend on the relative importance of the three return-generating mechanisms. We are interested in returns generated by trading, with particular attention to the relative amount of hedging trade versus speculative trade and their relative impact on stock prices. By analyzing the serial correlation of returns generated by trading among investors, we could learn about the relative importance of different trading motives. Therefore we would like to separately identify the returns generated by trading from those generated by public news on future payoffs, and examine their dynamics. We use trading volume to facilitate this identification. We observe that in our model, price changes generated by speculative or allocational trading must be accompanied by volume, but those generated by public news about payoffs do not. Thus by conditioning on the current volume return pair, we can (imperfectly) identify trade-generated returns (see CGW for a discussion of this point). Based on those returns, we can further examine how they might predict future returns. When all trades are hedging trades, current returns together with high volume predict strong reversals in future returns (as shown in CGW). When speculative trades are more important, current returns together with high volume predict weaker reversals (or even continuation) in future returns, as suggested in Wang (1994).

We now analyze more formally how the relative importance of hedging trade versus speculative trade might affect return dynamics. For clarity of exposition, we set $Z_t^{(2)} = 0$ for the rest of this section. Thus class 1 investors generate all the trades. We denote $Z_t^{(1)}$ by Z_t. Then we have $P_t = F_t + \widetilde{P}_t$, $\widetilde{P}_t = a(G_t - bZ_t)$ and

$$R_t = G_{t-1} + F_t + (\widetilde{P}_t - \widetilde{P}_{t-1}) \tag{7a}$$

$$V_t = (1-\omega)\alpha|\widetilde{P}_t - \widetilde{P}_{t-1}|, \tag{7b}$$

where $\alpha \equiv \gamma/(\gamma\sigma_G^2 + \sigma_F^2 + \sigma_{\widetilde{P}}^2)$. We can then compute the expectation of future returns conditioned on both current return and volume and the result is given in the following proposition:

Proposition 2. *From Equation* (5),

$$\mathrm{E}[R_{t+1}|\widetilde{V}_t, R_t] = \beta_1 R_t - \beta_2 \widetilde{V}_t \tanh\left(\eta \widetilde{V}_t R_t\right), \tag{8}$$

where $\widetilde{V}_t \equiv V_t/\mathrm{E}[V_t]$ *is volume normalized by its unconditional mean and* β_1, β_2, *and* η *are constants. Moreover,* $\beta_1 \le 0$, $\beta_2 \ge 0$ *and* $\eta \ge 0$.

Equation (8) forecasts future returns using current return and volume. It is obvious that given the current return, the higher the current trading volume implies stronger reversal in the next return.[5] We can further consider an approximation of the forecasting equation when volume and return are small:

$$\mathrm{E}[R_{t+1}|\widetilde{V}_t, R_t] = -\left(\theta_1 + \theta_2 \widetilde{V}_t^2\right)R_t + \text{higher-order terms in } \widetilde{V}_t \text{ and } R_t, \tag{9}$$

where $\theta_1 = -\beta_1 \ge 0$ and $\theta_2 = \beta_2\eta \ge 0$. Equation (9) illustrates the dynamic relation between current return and volume and future return: Volume is

[5] In the simple specification of the model, we set the total supply of the stock ($\bar{x}$) at zero. Thus the unconditional mean of the excess stock return is zero. Since this article focuses on the dynamics of stock returns, this simplification does not affect our results.

related to serial correlation in returns. Even though this result can be stated in the general form of Equation (8), we use the approximate form of Equation (9) in our future analysis for its intuitive appeal. Given the small magnitudes of average daily volume and returns of individual stocks, this approximation is reasonable.

Next, we examine how the dynamic volume-return relation in Equation (9) might depend on the importance of speculative trade in the market, which is driven by information asymmetry. A natural measure of information asymmetry is σ_G^2, the variance of the dividend component on which informed investors have private information. Thus we consider how θ_2 changes with σ_G^2, holding constant the total risk of the stock. The (unconditional) risk of the stock has two components: uncertainty in dividend, given by $\sigma_D^2 \equiv \sigma_F^2 + \sigma_G^2$, and uncertainty in future price, given by $\sigma_{\tilde{P}}^2$ for the uninformed investors.

When $\sigma_G^2 = 0$, there is no information asymmetry and investors trade only to hedge their nontraded risk. When $\sigma_G^2 > 0$, there is information asymmetry and the informed investors trade for hedging *and* speculative reasons. We then have the following proposition:

Proposition 3. *For* $\sigma_G^2 = 0$,

$$\theta_1 = 0 \quad and \quad \theta_2 = \theta_{20} \equiv \frac{\sigma_{\tilde{P}}^2}{\pi \sigma_D^2} = \frac{\omega^2}{\pi} \frac{\sigma_{DN}^2 \sigma_Z^2}{\sigma_D^2}.$$

For $\sigma_G^2 > 0$ *but small and holding* σ_D^2, $\sigma_{\tilde{P}}^2$ *constant,*

$$\theta_1 = \frac{1}{2} \omega \frac{\sigma_G^2}{\sigma_D^2}. \tag{10}$$

and

$$\theta_2 = \theta_{20} \left[1 - \omega \left(\frac{1}{\sigma_D^2} + \frac{3}{2\sigma_{\tilde{P}}^2} \sigma_G^2 \right) \right] + o\left(\sigma_G^2\right). \tag{11}$$

θ_1 *increases with* σ_G^2 *and* θ_2 *decreases with* σ_G^2.

In the absence of information asymmetry ($\sigma_G^2 = 0$), θ_1 equals zero and θ_2 is positive. Thus returns with no volume are uncorrelated with future returns ($\theta_1 = 0$). While returns with volume are more likely to reverse ($\theta_2 > 0$). This is consistent with our intuition that returns generated by public news on payoffs (dividends) give rise to no volume and are serially uncorrelated, while returns generated by investors' hedging trades do give rise to volume and are serially negatively correlated.

In the presence of information asymmetry, holding constant the risk of the stock, θ_1 increases and θ_2 decreases with the degree of information asymmetry, which is measured by σ_G^2. The intuition behind the dependence of θ_1 and θ_2 on information asymmetry is as follows.

From Equation (9), θ_1 gives the return autocovariance conditioned on volume being zero. For $\sigma_G^2 > 0$, a positive θ_1 implies that returns with no volume are more likely to reverse. With no volume, returns come from two sources: public news about future dividends (F_t) and surprises in current dividends ($D_t = F_{t-1} + G_{t-1}$). In the former case, as discussed before, the return comes from the change in price, fully reflecting the new information on future dividends, which contains no information about future returns. In the latter case, however, the return does contain information about future returns. Here is why. Suppose that current return is negative because the current dividend is lower than previously expected, that is, G_{t-1} is lower than $\mathrm{E}_{t-1}^{(2)}[G_{t-1}]$. That implies that the stock was overvalued previously and the informed investors have sold the stock more than what their hedging needs require. In other words, the informed investors' stock demand for their hedging needs in the last period is lower than the previous estimate. The fact that there is no volume in the current period then implies that the informed investors' stock demand for hedging needs has not changed from its previous value. Thus their current demand is low, which implies that the current stock price is low and expected future return is high. Therefore, conditioned on no volume, lower current return implies higher future return. Clearly such a negative serial correlation in no-volume returns arises from the information asymmetry among the investors. As σ_G^2 increases, this correlation becomes more negative, which explains why θ_1 increases with σ_G^2.

The effect of information asymmetry on θ_2 is more straightforward. Conditioned on positive volume, returns are more likely to be generated by trading. In the presence of information asymmetry, some of the trades are motivated by private information, which lead to positively serially correlated returns. Thus they have the effect of decreasing θ_2.

Although the above results are stated only for small σ_G^2 when an analytical proof is available, we also examine its validity numerically when σ_G^2 is large. By computing θ_2 for the complete range of σ_G (between zero and σ_D^2) for a wide range of parameter values (of σ_D^2, σ_Z^2, and σ_{DN}), we find that the dependence of θ_2 on σ_G^2 is always negative.[6]

Propositions 2 and 3 show how current return and volume can predict future returns, and how this predictability depends on the relative significance of hedging trade versus speculative trade. In a cross-sectional context, all else being equal, stocks with higher information asymmetry have lower θ_2 than stocks with lower information asymmetry. It is this dependence of the return-volume dynamics on the degree of information asymmetry in the market, characterized in Equations (9) and (11), that we test empirically.

Although our model also has clear predictions about θ_1, which is consistent with data, our empirical analysis focuses more on θ_2. This bias in focus has several reasons. First, the result on θ_1 can be sensitive to the simplifying

[6] Results of the numerical analysis are available from the authors upon request.

assumptions of our model. For example, when motives for hedging trade (as measured by Z_t) are persistent, the returns can become positively serially correlated (even in the absence of trading). In addition, when private information is long-lived, the behavior of return autocorrelation is more involved, and the impact of information asymmetry on θ_1 becomes more complex [see, e.g., Wang (1993)]. Second, the result on θ_1 is not unique to our model. For example, models with noise trading and asymmetric information, such as Brown and Jennings (1989) or Wang (1993), give the same prediction about θ_1: When there is more information asymmetry, noise trades cause larger price changes, which leads to more significant negative serial correlation in returns.[7] Also, there are other empirical factors, such as bid-ask bouncing and uncertain time intervals in return measurement, that may give rise to similar results about θ_1.[8] In contrast, the result on θ_2 is more robust. For example, Wang (1994) obtains similar (but less sharp) results in the more general setting with persistent motives for hedging trade and private information.[9] Our primary focus on θ_2 is based on these considerations, as well as our objective to understand the joint behavior of return and volume.

1.4 Discussion of the model

Our model is similar to that of Wang (1994) with two simplifying assumptions. First, shocks to the economy are independently (and identically) distributed over time. The independence assumption implies that investors' private information is short-lived: It is only about the next dividend, which is revealed after one period. Thus the less-informed investors do not have to solve the dynamic learning problem (and the corresponding optimization problem for their stock demand), which simplifies their policy.[10] Second, we assume that investors are myopic, which further simplifies the investors' optimization problem.

Our simplification gives sharper results about the dependence of the volume-return relation on information asymmetry, as shown in Proposition 3, while Wang (1994) relies on numerical analysis and provides only examples. However, our model has the restrictive implication that θ_2 remains nonnegative even for high degrees of information asymmetry. As shown in Wang (1994), when private information can be long-lived, θ_2 can become negative as the degree of information asymmetry increases. In our empirical analysis, we allow this possibility.

[7] It should be pointed out that when the "noise trade" are endogenous, as we have here, one obtains different results. See, for example, Hong and Wang (2000) for a more detailed discussion on this issue.

[8] We thank the referee for bringing to our attention the issue of uncertain time intervals in return measurement. We discuss these empirical issues in more detail in Section 3.

[9] When nonspeculative trades are exogenous, the impact of information asymmetry on θ_1 and θ_2 tend to be the same (both negative). As mentioned earlier and shown in the article, when the nonspeculative trades are endogenous, the results are different.

[10] For different cases of long-lived private information in a competitive setting, see, for example, Brown and Jennings (1989), Grundy and McNichols (1989), Wang (1994), and He and Wang (1995).

We assume that investors have constant absolute risk-aversion (CARA) and that assets (the stock and the nontraded asset) have (conditionally) normally distributed payoffs. The combination of these two assumptions allows a closed form solution for the model. However, as a special feature of the CARA preferences, each investor's stock demand is independent of his wealth. Hence there is no income effect in investors' stock demand, and public news on future asset payoffs (and the corresponding price change) does not cause investors to trade in the market. Of course, for more general preferences, investors do rebalance their portfolios in response to public news on future payoffs as their wealth changes, giving rise to another motive for allocational trades. As mentioned earlier, our model introduces a motive for allocational trade by including a set of nontraded assets in investors' portfolios. We note that the detailed motive for allocational trades is not crucial to our main result. This particular choice in the model is for tractability and simplicity.

Despite the simplifying assumptions of the model, it provides a clear illustration of certain volume-return relations, which we believe to be more general than the model itself. As discussed above, relaxing many of these assumptions is possible, but adds little to the main thrust of the article.[11]

Another issue that warrants more discussion is the informational role of volume to investors. As discussed previously, in our model there are only two classes of investors and volume is determined (up to a scaling constant) by changes in the position of each investor. Since investors directly observe their own stock holdings, no additional information is provided to them by volume. This simplification eliminates any informational role of volume from the perspective of investors (as we have shown, however, volume does provide additional information to the observers of the market). Although this situation is consistent with the rational expectations equilibrium of the model, it is quite special nonetheless. In the more general situation when there are many classes of investors, volume does provide additional information. Blume, Easley, and O'Hara (1994) and Bernardo and Judd (1999), for example, consider this situation. The exact impact of the informational role of volume on prices is quite complex to analyze since volume is a nonmonotonic function of changes in investors' stock holdings. For tractability, Blume, Easley, and O'Hara rely on particular specifications of investor behavior and Bernardo and Judd use numerical solutions. The actual outcome depends on the nature of information asymmetry among investors, among other things.

Although the informational role volume can play for investors is an interesting and challenging topic, it is not our focus in this article. Our focus is on how market observers (as opposed to market participants) can use volume, together with prices, to better understand the market, especially to what

[11] There are assumptions in the model that are more substantial, such as those on the preferences and distribution combination and the information structure. Relaxing those would significantly change the model.

extent investors trade for speculation versus for risk sharing and how these trades generate different return dynamics. For this purpose, our simple setting of only two classes of investors is sufficient. When richer heterogeneity among investors is allowed, volume does convey additional information to the investors and the model becomes intractable. But the basic intuition of our analysis is still valid. That is, returns accompanied by volume are more likely to be generated by investors' trading activities as opposed to public news on payoffs. Of course, contributions from offsetting trades among investors with similar information and risk exposure now makes volume a noisier signal about the net speculative or risk-sharing trades of informed investors. The exact impact of volume as a source of information to investors on the dynamic volume-return relation requires further research.

2. Empirical Methodology

2.1 Data and sample description

Our primary sample consists of all common stocks traded on the NYSE and AMEX. From the Center for Research in Security Prices (CRSP), we obtain data on daily return, price, number of shares traded, and shares outstanding. We obtain quotes and bid-ask spreads data from the TAQ database. Our sample period is from January 1, 1993, to December 31, 1998. We choose this six-year sample period for two reasons. First, the TAQ database only starts at the beginning of 1993. Second, the nature of our test requires that stock-specific parameters remain constant over time, which may not be the case over a long period.[12] During the sample period, the CRSP database contained 3538 stocks that were traded on the NYSE and AMEX. To allow a more precise estimation of our time-series regressions, and a more uniform cross-sectional comparison of the parameters, we further require that stocks in the sample trade in at least two-thirds of the days (1000 days out of 1516 possible trading days). This requirement reduces our sample to 2226 stocks.

Table 1 presents the firms' characteristics for the entire sample and for three subgroups according to size. For each firm i, we measure size (AvgCap_i) as the average daily market capitalization (number of shares outstanding multiplied by the daily closing price) over the sample period. The market capitalization of firms in our sample ranges from \$3.61 million to \$147.82 billion. As indicated by columns 2 and 3 of the table, both the average daily number of shares traded (AvgTrd_i) and the average share turnover (AvgTurn_i), which is the number of shares traded relative to shares outstanding, increase with firm size. For example, the daily average turnover is 0.27% for the small size group and is 0.355% for the large size group.

[12] In Section 3.3, we show that increasing the time horizon to ten instead of six years does not affect our results.

Table 1
Summary statistics

Panel A: Characteristics of the entire sample and three size-based subsamples

	AvgCap (in million $)	AvgTrd (in 100s)	AvgTurn (in %)	AvgPrc (in $)	BAsprd (in %)
Entire sample					
Mean	2,587.55	1997	0.321	25.59	2.19
Median	473.52	672	0.258	22.40	1.38
Std. Dev.	8,019.00	3844	0.255	20.49	2.33
Minimum	3.61	2	0.007	0.39	0.25
Maximum	147,817.21	52,735	2.837	330.45	19.51
Observations	2226	2226	2226	2225	2225
Size group: small					
Mean	85.97	288	0.270	10.84	4.11
Median	75.96	164	0.212	8.90	3.12
Std. Dev.	58.18	408	0.213	7.79	2.99
Observations	742	742	742	742	741
Size group: medium					
Mean	525.98	918	0.338	23.93	1.62
Median	473.52	601	0.273	22.91	1.35
Std. Dev.	237.23	994	0.263	11.24	1.13
Observations	742	742	742	742	742
Size group: large					
Mean	7,150.69	4786	0.355	42.01	0.84
Median	3,111.11	2863	0.291	37.20	0.75
Std. Dev.	12,714.35	5597	0.277	24.17	0.48
Observations	742	742	742	741	742

Panel B: Descriptive statistics of the volume (detrended log turnover) series

	Mean	Standard deviation	First autocorrelation	Fifth autocorrelation	Tenth autocorrelation	No. of stocks
Entire sample	0.028	1.068	0.370	0.193	0.143	2226
Size group: low	0.013	1.537	0.313	0.182	0.138	742
Size group: medium	0.038	1.000	0.366	0.196	0.148	742
Size group: high	0.034	0.668	0.430	0.200	0.144	742

The sample includes 2226 common stocks that traded on the NYSE or AMEX between 1993 and 1998 with at least 1000 days with trading. Panel A presents descriptive statistics by size groups. For each firm, AvgCap_i is the average daily market capitalization (number of shares outstanding multiplied by the daily closing price). AvgTrd_i is the average number of shares traded daily. AvgTurn_i is the average daily turnover (number of shares traded divided by the number of shares outstanding), and AvgPrc_i is the average price of a firm's stock over the sample period. BAsprd_i is the average daily opening percentage spread (opening bid-ask spread divided by the opening midquote) over the sample period from the TAQ database. Panel B presents descriptive statistics for the volume series used in the time-series regressions. We define the daily volume series of a stock as the detrended, log-transformed daily turnover series. We report the averages for the entire sample and for three size groups of the mean, standard deviation, and selected autocorrelation coefficients of the volume series of the individual stocks.

The average share price (AvgPrc_i in column 4) exhibits the same pattern [see Lo and Wang (2000) for an extensive analysis of the cross section of stock trading volume].

Using the data from TAQ, we construct a measure of a stock's bid-ask spread (BAsprd_i). In light of the results in Madhavan, Richardson, and Roomans (1997), we use the opening spread to capture the asymmetric information component of the spread. We then define the relative spread for each stock as the average of the daily opening percentage spread (opening bid-ask spread divided by the opening mid-quote) over the sample period. Consistent with many prior studies, the relative spread is high for the small size group (4.11%), and decreases monotonically with firm size (the average for firms in the large size group is only 0.84%).

2.2 Return, volume, and proxies for information asymmetry

We use daily returns and trading volume to analyze the impact of information asymmetry on the dynamic volume-return relation. The main reason to use daily data is to be able to relate our results to those of previous studies [e.g., LeBaron (1992), Antoniewicz (1993), CGW, and Stickel and Verrecchia (1994)].[13] In Section 3.3 we consider an alternative procedure to determine the appropriate time interval for our analysis empirically. The return series we use for the estimation is the daily return of individual stocks from CRSP. We also test the sensitivity of our results to alternative definitions of returns that avoid potential data problems such as variation in the time of the last daily trade and bid-ask bounce. For that purpose we also use quote data from TAQ.

We use daily turnover as a measure of trading volume for individual stocks. We define a stock's daily turnover as the total number of shares traded that day divided by the total number of shares outstanding.[14] Since the daily time series of turnover is nonstationary [see, e.g., Lo and Wang (2000)], we measure turnover in logs and detrend the resulting series. To avoid the problem of zero daily trading volume, we add a small constant (0.00000255) to the turnover before taking logs.[15] We detrend the resulting series by subtracting a 200 trading day moving average:

$$V_t = \text{logturnover}_t - \frac{1}{200}\sum_{s=-200}^{-1} \text{logturnover}_{t+s},$$

where

$$\text{logturnover}_t = \log(\text{turnover}_t + 0.00000255).$$

Panel B of Table 1 provides summary statistics that describe the volume series used in the estimation. While the first-order autocorrelation is higher for larger stocks, the pattern becomes less pronounced in the fifth order and disappears in the tenth order. We test the sensitivity of our results to alternative definitions of volume in Section 3.4.

As in CGW, we note that there is some slippage between the theoretical variables in the model and those in the empirical part. Our model considers

---

[13] CGW also consider the volume-return relation between two-day returns and volume for the market and find it is still present, but weaker than that for daily data. However, Conrad, Hameed, and Niden (1992) and Lee and Swaminathan (2000) examine returns over longer horizons (e.g., six months) and found that in addition to past returns, past volume can be informative about future returns.

[14] Lo and Wang (2000) provide a theoretical justification for using turnover as a measure of trading volume in a detailed study of the turnover of individual stocks.

[15] The value of the constant is chosen to make the distribution of daily trading volume closer to a normal distribution. See Richardson, Sefcik, and Thompson (1986), Ajinkya and Jain (1989), and Cready and Ramanan (1991) for an explanation.

dollar returns per share and normalized turnover, and our empirical analysis considers returns per dollar and detrended log turnover (so that the normalized series are closer to being stationary). The difference between the theoretical and corresponding empirical variables is mainly a matter of normalization. At the daily frequency that we focus on, the relation among these variables should not be very sensitive to the normalizations used.

To test Proposition 3, we also need a measure of information asymmetry for individual stocks that represents the extent of informed trading. Since information asymmetry is not directly observable, we must find a suitable proxy for the empirical investigation. Previous studies use several variables to measure information asymmetry, among which are bid-ask spreads and market capitalization. Some researchers argue that firms with lower bid-ask spreads [e.g., Lee, Mucklow, and Ready (1993)] and larger firms [e.g., Lo and MacKinlay (1990)] have a lower degree of information asymmetry or smaller adverse selection costs. We use both proxies in our empirical analysis for a couple of reasons. First, since there is no agreement on which proxy is the "best," we believe it is prudent not to rely on only one of them. Second, by using more than one proxy, we can examine the sensitivity of our results to various empirical representations of information asymmetry. For this reason, we employ yet another proxy in Section 3.5, the number of analysts who are following a stock, which is linked to the degree of information production in the market. Recent studies [Brennan and Subrahmanyam (1995) and Easley, O'Hara, and Paperman (1998)] find that firms with a larger analyst following have a lower degree of information asymmetry or lower adverse selection costs. Hence we use the number of analysts as a proxy for information asymmetry, and we expect that the more analysts who are following a stock, the less information asymmetry there is about it.

Even with an agreement on the proxies, there is still the question of how to use these proxies in a cross-sectional test. The exact nature of the functional relation between information asymmetry and the proxy is unknown. For example, the last column in panel A of Table 1 indicates that while the average bid-ask spread of the large size group is five times smaller than that of the small size group, the average firm size of the small group is almost 100 times smaller than that of the large group (first column). To work with the two proxies in a unified framework, we adopt an ordinal transformation of the variables. That is, we order the firms in an ascending order according to the proxy and assign a rank of one to the first firm (say, the smallest firm when firm size is the proxy) and a rank of 2226 to the last firm. We then divide the ordinal variable by 2226 so that its range is between zero and one. This monotonic transformation preserves the intuition of the differences between low and high information asymmetry without reading too much into the specific differences in magnitude.[16]

[16] See Johnston (1985) and references therein for a justification of this transformation.

The correlation between ORDCAP and ORDBA (the variables representing the ordinal scales of AvgCap and BAsprd) is −0.876. Although this is a moderately high correlation, it does not suggest that the two proxies represent the exact same phenomenon. We have also repeated our experiments using either the raw variables (AvgCap and BAsprd) or their log transformations. Our results are not materially affected by these alternative representations.

2.3 Experiment design

To test Proposition 3, we estimate the following relation for each individual stock:

$$R_{it+1} = \mathrm{C0}_i + \mathrm{C1}_i \cdot R_{it} + \mathrm{C2}_i \cdot V_{it} R_{it} + \mathrm{error}_{it+1}. \tag{12}$$

While the relation in Equation (9) has squared normalized volume entering the interaction term, we estimate the relation with our definition of normalized volume, the detrended log turnover, without squaring. We do this to allow for comparisons with prior empirical studies (e.g., CGW). We test the relation in Equation (9) with a measure of squared volume in Section 3.4.

In principle, trading contains both hedging and speculative elements. The observed volume-return relation depends on the relative importance of one type of trade relative to the other. We should see statistically significant positive C2 coefficients for stocks that are associated with very significant speculative trade, while for stocks with predominantly hedging trade, the C2 coefficients should be clearly negative. Stocks for which neither speculative nor hedging trade dominates should have C2 coefficients that are insignificantly different from zero. Moreover, the relation between C2 and the significance of speculative trade relative to hedging trade is monotonic.

To examine the relation between the importance of information asymmetry and the C2 coefficients, we use both structured and nonstructured methods. To give a sense of the underlying relation without imposing additional structure, we present a discrete categorization analysis of the results by assigning the stocks into three groups of the information asymmetry proxy. Proposition 3 implies the following relation:

$$\mathrm{C2}_i = f(\mathrm{A}_i), \tag{13}$$

where A_i is a proxy for the degree of information asymmetry of an individual stock. For the bid-ask spread proxy, higher values of A_i are associated with a higher degree of information asymmetry, and we should observe that the mean of $\mathrm{C2}_i$ is more positive for stocks with larger bid-ask spreads. For the market capitalization proxy, higher values of A_i are associated with a lower degree of information asymmetry, and so the mean of $\mathrm{C2}_i$ should be more positive for smaller stocks.

Under the assumption that the relation is linear, we can estimate the cross-sectional relation

$$\mathrm{C2}_i = a + b \cdot \mathrm{A}_i + \mathrm{error}_i. \tag{14}$$

Here we should see $b > 0$ when the information proxy used is the bid-ask spread and $b < 0$ when the information proxy used is market capitalization.

3. Empirical Results

We now present our empirical results in testing the theoretical implications of the model on the dynamic volume-return relation, especially how it is related to the underlying information asymmetry among investors. We first report the test based on the C2 coefficient from the simple ordinary least squares (OLS) regression in Equation (12) and the proxies for information asymmetry. We then examine the robustness of our result with respect to different econometric specifications and empirical adjustments. We further examine the sensitivity of our empirical result with respect to alternative test design and variable choices. Finally, we discuss the empirical result on the C1 coefficient.

3.1 Basic test of the dynamic volume-return relation

Table 2 presents the results from Equation (12) for individual stocks and how the regression coefficients change with the bid-ask spread as the information asymmetry proxy. For each stock in the sample, we estimate the parameters C0, C1, and C2 of Equation (12). In panel A we present summary statistics for these 2226 time-series regressions for each of the three bid-ask spread groups. The table shows that the mean value of C2 decreases monotonically with the stock's bid-ask spread, which is consistent with Proposition 3. Stocks with higher information asymmetry (large bid-ask spreads) are associated with larger and more-positive coefficients (0.035 for the high bid-ask spread group). The mean value becomes negative for stocks in the low bid-ask spread group (−0.003). The nonparametric analysis points in the same direction: only 141 (out of 742) of the stocks in the high group have negative coefficients, compared with 378 in the low group.

Most C2 coefficients of firms with large bid-ask spreads are positive and statistically different from zero, indicating the importance of speculative trading. For many of the stocks with medium spreads, the C2 coefficients are not significantly different from zero, which is consistent with a balance of both speculative and hedging trades. For stocks with small bid-ask spreads, many C2 coefficients are negative and statistically significant, indicating the dominance of hedging trades. The evidence in the table points to a monotone positive relation between C2 and bid-ask spreads.

In panel B we use regression analysis to examine this relation. Equation (14) is estimated using the bid-ask spread as the information asymmetry

Table 2
Bid-ask spread and the influence of volume on the autocorrelation of returns

Panel A: Categorical analysis

	C0 #<0	C1 #<0	C2 #<0	t_{C0} |#|>1.64	t_{C1} |#|>1.64	t_{C2} |#|>1.64	R^2 (%)	BAsprd (in %)
Low	0.000770	0.013323	−0.002814	1.698	0.336	−0.096	0.564	0.719
($n=742$)	30	328	378	409	337	240		
Medium	0.000722	0.010931	0.003168	1.239	0.188	0.232	0.921	1.417
($n=742$)	74	326	357	250	431	257		
High	0.000740	−0.120346	0.035495	0.709	−4.003	2.066	2.754	4.428
($n=742$)	167	609	141	135	548	438		

Panel B: Regression analysis

Dependent variable	a	b	R^2 (%)	Observations
C2	−0.015921	0.055716	10.197	2226
	(−7.863)	(15.891)		

This table shows the relation between information asymmetry and the influence of volume on the autocorrelation of stock returns. The average daily opening percentage spread of a stock over the sample period (BAsprd_i) is used as a proxy for information asymmetry. For each stock the parameter C2_i from the following regression measures the influence of volume on the autocorrelation of stock returns:

$$\text{Return}_{i,t+1} = \text{C0}_i + \text{C1}_i * \text{Return}_t + \text{C2}_i * \text{Volume}_{i,t} * \text{Return}_{i,t} + \text{error}_{i,t+1},$$

where $\text{Volume}_{i,t}$ is the daily detrended log turnover of an individual stock and $\text{Return}_{i,t}$ is the daily return of an individual stock. In panel A, we report the mean value of each parameter for three groups (low, medium, and high) of the information asymmetry proxy (BAsprd), the number of negative parameters and the number of statistically significant (at the 10% level) parameters. In panel B, we provide an analogous analysis using the following cross-sectional regression:

$$\text{C2}_i = \text{a} + \text{b} * \text{ORDBA}_i + \text{ERROR}_i,$$

where ORDBA is a variable representing the ordinal scale of BAsprd. t-statistics appear in parentheses.

proxy, that is, the dependent variable is C2 (the influence of volume on the autocorrelation of returns) and the independent variable is ORDBA (the bid-ask spread rank order). The spread coefficient is positive and highly significant, indicating that stocks with small spreads (i.e., lower information asymmetry) have lower volume-return interaction terms.

In Table 3 we use market capitalization as a proxy for information asymmetry. The results are similar to those in Table 2. Because larger size is associated with lower information asymmetry, the interaction coefficient C2 is the most positive for small firms (high information asymmetry) and decreases as the size of the firm increases. In the low group, 167 stocks have a negative C2 coefficient (mean value 0.030), but there are 354 stocks with a negative C2 coefficient in the high group, with a mean that is very close to zero. The regression results in panel B tell the same story: There is a statistically significant negative relation between our proxy for information asymmetry and the volume-return interaction parameter.

The results in these two tables are consistent with the prediction of Proposition 3. Using two different information proxies, we find that following high volume, stocks that are associated with more informed trading exhibit persistence in their returns and stocks with less informed trading exhibit reversals.

Table 3
Market capitalization and the influence of volume on the autocorrelation of returns

Panel A: Categorical analysis

	C0 $\# < 0$	C1 $\# < 0$	C2 $\# < 0$	t_{C0} $\lvert\#\rvert > 1.64$	t_{C1} $\lvert\#\rvert > 1.64$	t_{C2} $\lvert\#\rvert > 1.64$	R^2 (%)	AvgCap (in million $)
Low	0.000784	−0.104234	0.030277	0.843	−3.538	1.848	2.729	85.97
($n = 742$)	139	557	167	157	550	429		
Medium	0.000672	0.005492	0.004852	1.178	0.028	0.289	0.996	525.98
($n = 742$)	95	348	355	258	433	284		
High	0.000776	0.002650	0.000719	1.625	0.032	0.065	0.514	7150.69
($n = 742$)	37	358	354	379	333	222		

Panel B: Regression analysis

Dependent variable	a	b	R^2 (%)	Observations
C2	−0.034081	−0.044242	6.430	2226
	(16.488)	(−12.362)		

This table shows the relation between information asymmetry and the influence of volume on the autocorrelation of stock returns. The average daily market capitalization of a stock over the sample period (AvgCap_i) is used as a proxy for information asymmetry. For each stock the parameter $C2_i$ from the following regression measures the influence of volume on the autocorrelation of stock returns:

$$\text{Return}_{i,t+1} = \text{C0}_i + \text{C1}_i * \text{Return}_t + \text{C2}_i * \text{Volume}_{i,t} * \text{Return}_i, \text{t} + \text{error}_{i,t+1},$$

where $\text{Volume}_{i,t}$ is the daily detrended log turnover of an individual stock and $\text{Return}_{i,t}$ is the daily return of an individual stock. In panel A, we report the mean value of each parameter for three groups (low, medium, and high) of the information asymmetry proxy (AvgCap), the number of negative parameters and the number of statistically significant (at the 10% level) parameters. In panel B, we provide an analogous analysis using the following cross-sectional regression:

$$\text{C2}_i = \text{a} + \text{b} * \text{ORDCAP}_i + \text{ERROR}_i,$$

where ORDCAP is a variable representing the ordinal scale of AvgCap. t-statistics appear in parentheses.

While we attribute the cross-sectional variation in C2 to different degrees of information asymmetry (or speculative trading), it may also be attributed to other factors such as differences in liquidity across stocks. In particular, for less-liquid stocks, high volume is associated with a higher price impact and a larger subsequent return reversal than for more-liquid stocks. Hence less-liquid stocks should have more negative C2 coefficients. However, natural candidates for stocks with lower liquidity are those stocks with small market capitalizations or large bid-ask spreads. Hence liquidity considerations should cause a larger return reversal following high-volume days for lower market capitalization or large bid-ask spread stocks. This is the opposite of what we find. If a liquidity effect exists, our empirical findings suggest that it is dominated by the information effect.

3.2 Robustness of results

For the results presented in Tables 2 and 3, we estimate both the time-series and the cross-sectional relations using OLS. One possible concern is whether this experiment design is robust to potential econometric problems. One econometric problem could be that the estimated relation in the time-series regression is affected by autocorrelated errors. In this case, a lagged dependent variable among the regressors precludes using OLS for the estimation. To examine how this problem might affect our results, we use a test developed by Breusch (1978) and Godfrey (1978a, b) to identify the most appropriate error structure. For each stock, we test for white noise against the alternative of an autoregressive error structure of orders one through five. We decided to limit the possible orders to five after a lengthy inspection of some stocks. We use a 5% significance level to reject the white noise hypothesis. If the test is significant for any order $p < 5$, but not for higher orders, we test again with the null of AR(p) against an autoregressive structure of orders higher than p, but only up to order five. After identifying the appropriate order, we estimate the relation in Equation (12) using maximum likelihood with the suitable autoregressive structure. We perform this procedure separately for each stock.

We then rerun the cross-sectional regressions with the information asymmetry proxies as the dependent variables and the new C2 coefficients as the independent variable. The results are presented in Table 4, panel A. These results are similar to the OLS findings reported in Tables 2 and 3, and show the same strong relation with the information asymmetry proxies. Therefore our results do not appear to be sensitive to autocorrelation of the error terms in the time-series estimations. To assess the sensitivity of our results to the order identification algorithm, we repeat all estimations, identifying the appropriate error structure only by the white noise test against an AR structure. Panel B of Table 4 presents the results of this specification. Our results appear robust to the exact manner in which the appropriate autoregressive order is identified.

Another possible econometric problem is that if the errors of Equation (12) are correlated across stocks, the $C2_i$ estimates will not be independent. When we estimate Equation (14), the standard error of b is then biased and tests of significance are difficult to interpret. Because a cross-correlation of the errors most likely arises from the sensitivity of the returns to missing common factors, one way to decrease such cross-correlation is to model the factors directly. Following Jorion (1990), we use a market proxy to model the missing common factors for the purpose of decreasing cross-correlation of the error terms. We estimate the following time-series relation for each stock:

$$R_{it+1} = C0_i + C1_i \cdot R_{it} + C2_i \cdot V_{it} R_{it} + C3_i \cdot R_{mt+1} + \text{error}_{it+1}, \qquad (15)$$

where R_{mt+1} is the return on a value-weighted portfolio of all common stocks that are traded on the NYSE or AMEX, and which have valid return and volume information in the CRSP database for that day.

Panel A of Table 5 reports the results of the cross-sectional regressions on the estimates of C2 from Equation (15). The coefficient on the information asymmetry proxy is positive and significant for ORDBA, and negative

Table 4
Alternative econometric specifications: autoregressive structure

Panel A: Autoregressive structure identified using algorithm A

$C2_i = a + b * ORDBA_i + ERROR_i$

A	b	R^2 (%)	Observations
−0.015298	0.060185	10.339	2226
(−7.048)	(16.014)		

$C2_i = a + b * ORDCAP_i + ERROR_i$

A	b	R^2 (%)	Observations
0.038386	−0.047135	6.341	2226
(17.303)	(−12.271)		

Panel B: Autoregressive structure identified using algorithm B

$C2_i = a + b * ORDBA_i + ERROR_i$

A	B	R^2 (%)	Observations
−0.015485	0.060403	10.459	2226
(−7.154)	(16.118)		

$C2_i = a + b * ORDCAP_i + ERROR_i$

A	b	R^2 (%)	Observations
0.038443	−0.047403	6.442	2226
(17.376)	(−12.374)		

The impact of volume on the autocorrelation of stock returns is estimated using the following regression:

$$\text{Return}_{i,t+1} = C0_i + C1_i * \text{Return}_t = C2_i * \text{Volume}_{i,t} * \text{Return}_{i,t} + \text{error}_{i,t+1},$$

where $\text{Volume}_{i,t}$ is the daily detrended log turnover of an individual stock and $\text{Return}_{i,t}$ is the daily return of an individual stock. For each stock, we test for white noise against the alternative of an autoregressive error structure of orders one through five using the Breusch–Godfrey test. We use a 5% significance level to reject the white noise hypothesis. In panel A, we identify the most appropriate autoregressive structure using algorithm A as follows: If the test is significant for any order $p < 5$, but not for higher orders, we test again with the null of AR(p) against an autoregressive structure of orders higher than p, but only up to order five. After identifying the appropriate order, we estimate the above relation using maximum likelihood with the suitable autoregressive structure. In panel B, we identify the autoregressive structure using algorithm B, which uses only the white noise test against an AR structure for each individual stock. Both panels report the results of the cross-sectional regressions:

$$C2_i = a + b * ORDBA_i + ERROR_i$$
$$C2_i = a + b * ORDCAP_i + ERROR_i$$

where ORDBA and ORDCAP are variables representing the ordinal scales of BAsprd and AvgCap, respectively. t-statistics appear in parentheses.

and significant for ORDCAP, consistent with our prior results. Scholes and Williams (1977) show how nonsynchronous trading can introduce econometric problems into the estimation of the market model. We can use their specification as a robustness test on the above formulation. More specifically, we estimate a market model using the Scholes–Williams correction for each of the stocks in our sample. We then take the residuals and use them instead of the regular returns in the time-series regressions described in Equation (12). This eliminates the common market factor from the returns used in the time-series regressions and hence decreases cross-correlations of the error terms when Equation (14) is estimated. Panel B of Table 5 reports the results of the cross-sectional regressions on both information asymmetry

Table 5
Alternative econometric specifications: sensitivity to common factor

Panel A: Using market return in time-series regressions

$C2_i = a + b * ORDBA_i + ERROR_i$

a	b	R^2 (%)	Observations
−0.005528	0.043256	6.693	2226
(−2.795)	(12.631)		

$C2_i = a + b * ORDCAP_i + ERROR_i$

a	b	R^2 (%)	Observations
0.031677	−0.031121	3.465	2226
(15.745)	(−8.934)		

Panel B: Using Scholes–Williams residuals to control for common factor

$C2_i = a + b * ORDBA_i + ERROR_i$

a	b	R^2 (%)	Observations
0.004273	0.035297	4.790	2226
(2.22)	(10.58)		

$C2_i = a + b * ORDCAP_i + ERROR_i$

a	b	R^2 (%)	Observations
0.033704	−0.023539	2.130	2226
(17.25)	(−6.96)		

In panel A, we estimate the parameter $C2_i$ using the following time-series regression:

$$\text{Return}_{i,t+1} = C0_i + C1_i * \text{Return}_{i,t} + C2_i * \text{Volume}_{i,t} * \text{Return}_{i,t} + C3 * \text{MktRet}_{t+1} + \text{error}_{i,t+1},$$

MktRet_{t+1} is the return on a value-weighted portfolio comprised of all common stocks that traded on the NYSE or AMEX that have valid return and volume information in the CRSP database for that day. In panel B, we estimate the market model for each stock using the Scholes and Williams (1977) methodology to account for nonsynchronous trading. The residuals from the market model ($\text{SWret}_{i,t+1}$) are then used instead of returns in the time-series regressions:

$$\text{SWret}_{i,t+1} = C0_i + C1_i * \text{SWret}_{i,t} + C2_i * \text{Volume}_{i,t} * \text{SWret}_{i,t} + \text{error}_{i,t+1}.$$

In both panels, we report the results of the cross-sectional regressions:

$$C2_i = a + b * ORDBA_i + ERROR_i$$
$$C2_i = a + b * ORDCAP_i + ERROR_i,$$

where ORDBA and ORDCAP are variables representing the ordinal scales of BAsprd and AvgCap, respectively. t-statistics appear in parentheses.

proxies. The results are similar in sign and significance to the results using the Jorion (1990) formulation.[17]

Another potential problem has its root in Equation (9) from Proposition 2, where the dynamic volume-return relation is developed using an approximation

[17] Another way to overcome the potential problem of cross-correlations is to estimate both the time-series and cross-sectional relations in a one-step procedure. A particularly suitable procedure for our model is the random coefficient model suggested by Amemiya (1978). The combination of a large sample and a long estimation period makes such estimation computationally challenging. The computation becomes more feasible if one imposes a block diagonal structure on the covariance matrix of the panel data. In other words, each stock is assumed homoscedastic, but contemporaneous correlation across stocks is allowed. In a previous version of the article that analyzed a smaller sample for the period 1983–1992, we used the GLS estimator in Amemiya (1978) to estimate the relation:

$$R_{it+1} = C0_i + C1_i R_{it} + (a + bA_i) V_{it} R_{it} + \epsilon_{it+1}. \tag{16}$$

For both information asymmetry proxies, the b coefficients had the appropriate signs and were highly statistically significant, confirming the results from the two-step procedure. Hence it does not appear as if contemporaneous correlations affect our conclusions in any meaningful way.

that ignores higher-order, nonlinear terms in the product of volume and return. If the product is very large, the approximation may not be good. We could trim observations above a certain bound, but the problem is to choose a sensible bound. So the approach we choose is to use an econometric methodology that identifies observations that are too large relative to a linear structure (i.e., outliers that could be the results of nonlinearities) and eliminate them from the analysis. The methodology that we use is the two-stage least trimmed squares [LTS; see Rousseeuw and Leroy (1987)]. In the first stage, the LTS estimator is applied to the relation. The estimator minimizes the sum of the smallest h residuals, where $h = \frac{2}{3}n$ and n is the number of observations. We use the residuals from the LTS estimation to create weights that identify an observation as an outlier if its residual is too large relative to a measure of the standard errors.[18] In the second stage we use these weights in a weighted least squares (WLS) estimation of the cross-sectional relation. We then report the results that come out of the second-stage estimation.

The two-stage LTS therefore enables us to estimate a cleaner linear relation. It is much less influenced by possible nonlinearities that might produce observations that are too far from the linear approximation, and which could result in biased slope coefficients. Table 6 presents the results of applying the above procedure to our data. In panel A we estimate each time-series regression using the two-stage LTS methodology. Then we estimate Equation (14), the cross-sectional regression, by using OLS. the results are similar to those reported in Tables 2 and 3: the coefficients of the information asymmetry proxies are highly significant and in the right direction. Hence it does not seem as if very large observations are adversely affecting the estimates of the parameter C2.

Because we do not know the functional form of the relation between information asymmetry and the proxy we use, we can also apply the two-stage LTS approach to our cross-sectional estimation. We use the C2 estimates that come out of the OLS time-series estimation to allow for a cleaner comparison with the results in Tables 2 and 3. Panel B of Table 6 presents the result of estimating the cross-sectional regression using a two-stage LTS. The results are similar to those of the OLS estimation.[19]

[18] The LTS estimator is given by

$$\min_{\hat{\theta}} \sum_{i=1}^{h} (r^2)_{i:n},$$

where $\hat{\theta}$ is the vector of estimated parameters and $(r^2)_{1:n} \leq \cdot \leq (r^2)_{n:n}$ are the ordered squared residuals. The weights are then defined as

$$w_i = \begin{cases} 1 & \text{if } |r_i/\hat{\sigma}| \leq 2.5 \\ 0 & \text{if } |r_i/\hat{\sigma}| > 2.5 \end{cases} \quad \text{and} \quad \hat{\sigma} = C_2 \sqrt{\frac{1}{n} \sum_{i=1}^{h} (r^2)_{i:n}},$$

where C_2 is a correction factor.

[19] We also repeated our cross-sectional analysis using nonparametric Spearman correlations in order to measure the association between the C2 coefficients and the information asymmetry proxies without relying on the

Table 6
Alternative econometric specifications: least trimmed squares

Panel A: Two-stage least trimmed squares applied to time-series regressions

$C2_i = a + b * ORDBA_i + ERROR_i$

a	b	R^2 (%)	Observations
−0.011487	0.049199	7.638	2226
(−5.483)	(13.562)		

$C2_i = a + b * ORDCAP_i + ERROR_i$

a	b	R^2 (%)	Observations
0.032918	−0.039572	4.942	2226
(15.487)	(−10.752)		

Panel B: Two-stage least trimmed squares applied to cross-sectional regressions

$C2_i = a + b * ORDBA_i + ERROR_i$

a	B	R^2 (%)	Observations
−0.019349	0.060732	14.328	2226
(−10.40)	(19.00)		

$C2_i = a + b * ORDCAP_i + ERROR_i$

a	b	R^2 (%)	Observations
0.034651	−0.046424	8.586	2226
(18.57)	(−14.25)		

For each stock, we estimate the $C2_i$ parameter from the following regression:

$$\text{Return}_{i,t+1} = C0_i + C1_i * \text{Return}_t + C2_i * \text{Volume}_{i,t} * \text{Return}_{i,t} + \text{error}_{i,t+1},$$

where $\text{Volume}_{i,t}$ is the daily detrended log turnover and $\text{Return}_{i,t}$ is the daily return. We estimate this relation in two stages. In the first stage, we use a least trimmed squares (LTS) estimator to identify outliers (by minimizing the sum of the h smallest squared residuals, where $h = 2/3$ of the observations for each stock). In the second stage, we use a weighted least squares (WLS) estimation in which we make the weights on the outliers identified in the first stage equal to zero. The $C2_i$ parameters for the cross-sectional estimation are taken from the WLS estimation. In panel A, we report the results of the OLS cross-sectional regressions:

$$C2_i = a + b * ORDBA_i + ERROR_i$$
$$C2_i = a + b * ORDCAP_i + ERROR_i,$$

where ORDBA and ORDCAP are variables representing the ordinal scales of BAsprd and AvgCap, respectively. In panel B, we apply the two-stage LTS estimation to the cross-sectional, rather than the time-series, regressions. We take the $C2_i$ coefficients from the OLS time series regressions that were presented in Tables 2 and 3, but we use LTS to estimate the relations with the information asymmetry proxies. We present the results of the second-stage WLS estimation. t-statistics appear in parentheses.

Next we examine the interpretation of our results in light of the bid-ask bounce effect [e.g., Roll (1984)]. Our model shows how volume should interact with the autocorrelation of returns. For stocks with more information asymmetry, greater volume should make the first-order autocorrelation less negative or even positive (hence a positive C2) due to the partial adjustment of prices to information. For stocks with less information asymmetry, greater volume should make the first-order autocorrelation more negative (hence a negative C2) due to the return reversal associated with liquidity shocks. Since

linear specification in Equation (14). The Spearman correlation between C2 and ORDBA is 0.326 (asymptotic standard error 0.02), and the correlation between C2 and ORDCAP is −0.26 (asymptotic standard error 0.02). Hence the Spearman correlations point to the same conclusions as all our other econometric procedures.

bid-ask bounce creates negative autocorrelation, we would expect a less negative or even positive C2 for stocks with large bid-ask spreads if more volume decreases the bid-ask bounce effect. The prediction of our model and the bid-ask bounce effect for this group of stocks would therefore operate in the same direction.

To examine this issue we generate a return series that is free from bid-ask bounce. Using the TAQ database, we generate returns from end-of-day midquotes for all stocks (except Berkshire Hathaway Inc., which is excluded due to its abnormal price range). We note that this return series is less reliable than the CRSP series used in the main analysis. First, there are more days without a valid end-of-day quote in the TAQ database than there are days without a valid return in the CRSP database. Each day without an end-of-day quote results in two days without valid midquote returns. Second, the intraday quote data in TAQ could contain more errors than the heavily used CRSP return series. We reestimate Equation (12) with the midquote return series for the firms in our sample. An indication that the aforementioned problems with respect to the midquote return series might have some effect is that the time-series regressions using TAQ returns produce a few outliers of the C2 coefficient, while the time-series regressions using CRSP returns do not produce any outliers. Hence we estimate the cross-sectional relations using the two-stage LTS procedure described earlier to identify and eliminate the influence of outliers. Panel A of Table 7 presents the cross-sectional regressions of ORDBA and ORDCAP on the interaction parameter C2. The information asymmetry proxies have the appropriate signs and are statistically significant. We note, though, that the proxies explain less of the variation in the C2 coefficients than they do in the cross-sectional regressions reported in Tables 2 and 3.[20]

Another potential data problem that can affect the interpretation of the results is variation in the time of the last daily trade in a stock.[21] Lo and MacKinlay (1990) present a model showing how this problem can lead to negative autocorrelation in the returns of individual stocks. Intuitively, days with higher volume should exhibit less of the problem (since there is a higher likelihood that the last trade is near the closing of the market). Since small stocks and stocks with large bid-ask spreads also tend to be less frequently traded, this data problem may cause a positive C2 for these groups of stocks, which is in the same direction as the prediction of our model.[22] Lo and

[20] The midquote return series does reduce the negative autocorrelation in the returns of small stocks. For example, the first-order autocorrelation for the group of small stocks (742 firms) goes from -0.076 to -0.036, and that for the group of stocks with large bid-ask spreads goes from -0.088 to -0.037.

[21] We thank the referee for drawing our attention to this point.

[22] Note that this data problem cannot explain our findings of negative C2 coefficients for stocks with low information asymmetry, for example, a small bid-ask spread. However, these can be explained within our model since greater volume should make the first-order autocorrelation for these stocks more negative due to the return reversal associated with liquidity shocks.

Table 7
Robustness to data problems: bid-ask bounce and time variation of last trade

Panel A: Using end-of-day midquote returns

$C2_i = a + b * ORDBA_i + ERROR_i$

A	b	R^2 (%)	Observations
−0.011522	0.030290	4.086	2226
(−6.00)	(9.23)		

$C2_i = a + b * ORDCAP_i + ERROR_i$

A	b	R^2 (%)	Observations
0.016266	−0.025829	2.964	2226
(8.69)	(−7.82)		

Panel B: Using 10:00 AM midquote returns

$C2_i = a + b * ORDBA_i + ERROR_i$

A	b	R^2 (%)	Observations
0.014066	0.024287	2.819	2226
(7.44)	(7.56)		

$C2_i = a + b * ORDCAP_i + ERROR_i$

A	b	R^2 (%)	Observations
0.034132	−0.015881	1.208	2226
(18.77)	(−4.91)		

Panel C: Using number of trades to control for freshness of closing price

$C2_i = a + b * ORDBA_i + ERROR_i$

A	b	R^2 (%)	Observations
−0.061201	0.062426	4.900	2226
(−18.17)	(10.70)		

$C2_i = a + b * ORDCAP_i + ERROR_i$

A	b	R^2 (%)	Observations
−0.004547	−0.050832	3.249	2226
(−1.34)	(−8.64)		

To assess the impact of the bid-ask bounce we calculate Return_i, from end-of-day midquotes (panel A) or from 10:00 AM midquotes (panel B) in the following time-series regression:

$$\text{Return}_{i,t+1} = C0_i + C1_i * \text{Return}_{i,t} + C2_i * \text{Volume}_{i,t} * \text{Return}_{i,t} + \text{error}_{i,t+1}$$

Both panels report the results of the cross-sectional regressions:

$$C2_i = a + b * ORDBA_i + ERROR_i$$
$$C2_i = a + b * ORDCAP_i + ERROR_i$$

where ORDBA and ORDCAP are variables representing the ordinal scales of BAsprd and AvgCap, respectively. The cross-sectional regressions are estimated using the two-stage least trimmed squares. In panel C, we run the time-series regressions with an interaction term between return and the number of trades, which serves as a proxy for the freshness of the closing price:

$$\text{Return}_{i,t+1} = C0_i + C1_i * \text{Return}_t + C2_i * \text{Volume}_{i,t} * \text{Return}_{i,t} + C3_i * NT_{i,t} * \text{Return}_{i,t} + \text{error}_{i,t+1},$$

where $NT_{i,t}$ is the number of trades of stock i on day t, and the definition of return used is from CRSP. We then report the results of the OLS cross-sectional regressions:

$$C2_i = a + b * ORDBA_i + ERROR_i$$
$$C2_i = a + b * ORDCAP_i + ERROR_i$$

t-statistics appear in parentheses.

MacKinlay (1990) calibrate their model and examine the implications of nontrading for different return horizons (from short-horizon daily returns to long-horizon annual returns). They end up concluding that with respect to autocorrelation, "the impact of nontrading for individual short-horizon stock returns is negligible" (p. 194). Still it is preferable to empirically examine the robustness of our results to this potential problem.

An alternative specification that alleviates some of the problem is using returns from end-of-day midquotes rather than closing prices. Quotes are binding obligations to trade, and hence presumably incorporate all information available to the market during the time the quote is in effect. By constructing returns from midquotes prevailing at 4:00 PM, one measures return from a commitment to trade at a certain point in time to a commitment to trade at the same point in time the following day, overcoming the problem associated with variation in the time of the last daily trade. As panel A of Table 7 shows, the same pattern in C2 prevails when we use end-of-day midquote returns.

One potential problem with end-of-day midquotes is "stale" limit orders: prices in the market are determined by limit orders that are submitted by investors who do not monitor the market continuously. As new information arrives, without updating the orders by trading or cancellation, quotes may not reflect the "true" price of the security despite being commitments to trade. However, for stocks more prone to the problem of variation in the time of the last daily trade, which tend to be infrequently traded stocks, the impact of stale limit orders are less important. This is because for NYSE stocks, specialists' participation rate is much higher in infrequently traded stocks [Madhavan and Sofianos (1998)], making it much more likely that the quotes for these stocks are set by the specialists and hence are not stale.

Nonetheless, we perform additional tests. One can argue that early morning quotes tend to be fresher than closing quotes. We repeat the experiment, generating returns from the midquotes prevailing at 10:00 AM every day. We reestimate Equation (12) with the 10:00 AM midquote return series for all firms in our sample. Panel B of Table 7 presents the cross-sectional regressions of ORDBA and ORDCAP on the interaction parameter C2. The results are very similar to the end-of-day midquotes used in panel A: The information asymmetry proxies have the appropriate signs and are statistically significant.[23]

Another formulation that can be used to control for the variation in the time of the last trade of the day is to include the number of trades in the

[23] We also repeated the experiment using returns calculated from the first trade of the day. The results are very similar to those of the 10:00 AM midquotes.

time-series estimation as a proxy for the "freshness" of the closing price.[24] More specifically, for each stock we run the following regression:

$$R_{it+1} = \mathrm{C0}_i + \mathrm{C1}_i \cdot R_{it} + \mathrm{C2}_i \cdot V_{it} R_{it} + \mathrm{C3}_i \cdot NT_{it} R_{it} + \mathrm{error}_{it+1}, \qquad (17)$$

where NT_{it} is the number of trades of stock i on day t. The idea behind this specification is that volume and the number of trades are positively correlated, but not perfectly, and hence it may be possible to use volume to partially identify either an information shock or a liquidity shock, while the number of trades is used as a proxy for the "freshness" of the price of the last trade. Panel C of Table 7 presents the cross-sectional regressions of ORDBA and ORDCAP on the interaction parameter C2. The information asymmetry proxies have the appropriate signs and are statistically significant.[25]

We conclude that our results seem robust with respect to adjusting or controlling for various potential econometric problems as well as potential data problems that may affect the analysis.

Beyond the issues discussed above, the design of the experiment requires us to make various decisions that may influence the results. It is therefore useful to examine in more depth a few of these choices. First, how do our choices for the length of the time-series estimation period or the daily intervals for return and volume affect our results? Second, how sensitive are our results to the exact definition of volume that we use? Third, can we relate our findings about information asymmetry to a variable that is more directly associated with information production? Fourth, is firm-specific information asymmetry a driving force behind the dynamic volume-return relation or does the relation disappear when we eliminate market-wide variations? The following sections address these questions.

3.3 Alternative lengths of time intervals

It is possible that the appropriate measurement period differs across stocks. For example, for an infrequently traded stock, the period could be several days, so that more trades are captured within the period. We might choose the appropriate measurement periods to equate the amount of median trading across stocks. The measurement interval is longer for stocks with lower level of hedging trading than it is for stocks with a higher level hedging trading. Thus we can use the average turnover of a stock as a proxy for the normal level of trading. A typical trading intensity measured by the stock's median turnover is less sensitive to informational or allocational volume shocks.[26]

[24] We thank the referee who suggested this specification.

[25] We also tested whether our results disappear when we only use stocks with sufficient trading (on average) to provide us with reasonably fresh prices. For that we used in the cross-sectional regressions only stocks that have on average more than 10 trades per day. This requirement eliminated 352 stocks from the sample. The results were very similar to those reported in Tables 2 and 3 for the entire sample.

[26] Using the cumulative volume as a measure of economic time has been considered by many authors, such as Clark (1973) and Lamoureux and Lastrapes (1990).

Therefore we calculate the median daily turnover for each stock over the sample period (MedTurn) and assign all stocks into three groups according to their median turnover. The average MedTurn for the three groups are 0.0634%, 0.1691%, and 0.384%, respectively. The average MedTurn of the high group is about twice that of the medium group and about five times that of the low group. (The proportions are similar when we use the cross-sectional median rather than the average of each group.) Therefore, given a daily interval for the most active stocks, we choose a two-day interval as the most appropriate for the medium group and a five-day interval for the low-turnover group. To calculate the return and turnover series for the medium and low groups, we compound returns and sum the turnover for the days in the interval.

We then perform a separate time-series analysis for each stock. A stock in the high MedTurn group that is listed for the entire sample period will have 1516 observations in the regressions, while a similar stock in the medium (low) MedTurn group will have 758 (303) observations. Taking the C2 coefficient from each individual stock's time-series regression, we estimate the cross-sectional relation in Equation (14). We present the cross-sectional results in Table 8, panel A. The bid-ask spread coefficient is positive and highly significant, and the size coefficient is negative and highly significant. Our results do not appear to be driven by the choice of the daily interval for the time-series regressions. Thus whether we fix a time interval (a day in our experiment) or a given the amount of trading (as the current test implies) does not affect our findings.

In panel B of Table 8, we use ten instead of six years to estimate the time-series regressions. This experiment allows us to check the sensitivity of our results to the length of the estimation period. We estimate the volume-return interaction parameters (C2) in this panel by using data from 1989 through 1998. The coefficients of the information asymmetry proxies in the cross-sectional analysis have the right signs and are highly statistically significant.[27]

3.4 Alternative definitions of volume

Since there is a slight difference between our detrended volume measure and the theoretical volume measure, Table 9 presents the results using alternative definitions of volume. In panel A we define volume as the daily share turnover of a stock, without taking any transformation or detrending. We reestimate the time-series relation in Equation (12) with this alternative volume definition. The results of the cross-sectional regressions show a statistically significant relation, in the appropriate direction, between C2 and both information asymmetry proxies.

[27] While not reported here, we have also conducted the experiment with a sample of NYSE and AMEX stocks using the period 1983–1992. The results are the same as those presented in the article.

Table 8
Alternative lengths of time intervals

Panel A: Using an alternative to the daily interval for measuring return and volume

$C2_i = a + b * ORDBA_i + ERROR_i$

a	B	R^2 (%)	Observations
−0.015534	0.038606	3.531	2226
(−6.286)	(9.023)		

$C2_i = a + b * ORDCAP_i + ERROR_i$

a	b	R^2 (%)	Observations
0.016285	−0.025003	1.481	2226
(6.521)	(−5.782)		

Panel B: Using ten years of data (1989–1998)

$C2_i = a + b * ORDBA_i + ERROR_i$

a	B	R^2 (%)	Observations
−0.014592	0.057887	13.461	2226
(−8.118)	(18.599)		

$C2_i = a + b * ORDCAP_i + ERROR_i$

a	b	R^2 (%)	Observations
0.037995	−0.047240	8.965	2226
(20.609)	(−14.799)		

Panel A replaces the daily interval for measuring return and volume with an alternative interval that allows for roughly the same level of noise trading across stocks. We use the median daily turnover (MedTurn) of a stock as a measure of the normal level of trading. We group all stocks into three categories (high, medium, and low) according to their MedTurn. We use a daily interval for stocks in the high group, a two-day interval for stocks in the medium group, and a five-day interval for stocks in the low group (so that the average turnover across groups is equal). We calculate the return and turnover series for a stock in the medium and low groups by compounding daily returns and linearly adding turnover for the days in each interval. For each stock, we measure the influence of volume on the autocorrelation of stock returns by the parameter $C2_i$ from the following regression:

$$\text{Return}_{i,t+1} = C0_i + C1_i * \text{Return}_t + C2_i * \text{Volume}_{i,t} * \text{Return}_{i,t} + \text{error}_{i,t+1},$$

where $\text{Volume}_{i,t}$ is the detrended log turnover of an individual stock and $\text{Return}_{i,t}$ is the return of an individual stock. We report the results of the OLS cross-sectional regressions:

$$C2_i = a + b * ORDBA_i + ERROR_i$$
$$C2_i = a + b * ORDCAP_i + ERROR_i,$$

where ORDBA and ORDCAP are variables representing the ordinal scales of BAsprd and AvgCap, respectively. In panel B, we test the robustness of our results to a longer estimation period. We estimate the time-series regressions for the stocks in our sample using 10 years of data (1989–1998). We present the results of the cross-sectional regressions using the same information asymmetry proxies as in panel A. t-statistics appear in parentheses.

In panel B we perform a more direct test of the relation in Equation (9) that comes out of our theoretical model. In the theoretical model, it is the squared volume, rather than a linear term, that affects the subsequent period's returns. We define volume as the logarithm of (1+ daily number of shares traded).[28] Then we estimate the following relation:

$$R_{it+1} = C0_i + C1_i \cdot R_{it} + C2_i \cdot V_{it}^2 R_{it} + \text{error}_{it+1}. \tag{18}$$

[28] To avoid taking the log of zero on days without trading, we add the small constant (1.00) to the daily number of shares traded before making the logarithm transformation.

Table 9
Alternative definitions of volume

Panel A: Volume is measured as turnover
$C2_i = a + b * ORDBA_i + ERROR_i$

a	b	R^2 (%)	Observations
−3.322436	10.685070	2.269	2226
(−3.869)	(7.186)		

$C2_i = a + b * ORDCAP_i + ERROR_i$

a	b	R^2 (%)	Observations
6.408573	−8.768206	1.528	2226
(7.434)	(−5.875)		

Panel B: Volume is measured as $\log(1 + \text{shares traded})^2$
$C2_i = a + b * ORDBA_i + ERROR_i$

a	B	R^2 (%)	Observations
−0.001029	0.003526	19.999	2226
(−11.917)	(23.579)		

$C2_i = a + b * ORDCAP_i + ERROR_i$

a	b	R^2 (%)	Observations
0.002335	−0.003199	16.461	2226
(26.454)	(−20.934)		

In the following regression, $Return_{i,t+1} = C0_i + C1_i * Return_t + C2_i * Volume_{i,t} * Return_{i,t} + error_{i,t+1}$ volume is measured as either turnover—the daily number of shares traded divided by the number of shares outstanding—(panel A), or as the $\log(1 + \text{daily number of shares traded})^2$ (panel B).
Using the results of the time-series regressions, both panels report the results of the cross-sectional regressions:

$$C2_i = a + b * ORDBA_i + ERROR_i$$
$$C2_i = a + b * ORDCAP_i + ERROR_i,$$

where ORDBA and ORDCAP are variables representing the ordinal scales of BAsprd and AvgCap, respectively. t-statistics appear in parentheses.

The resulting cross-sectional analysis shows the same pattern as in Tables 2 and 3. In fact, it appears that the relation is even stronger. Both information asymmetry proxies explain more than 16% of the cross-sectional variation in the parameter C2.

3.5 Analyst following as a proxy for information production

While the information asymmetry proxies we use in the main analysis have received much attention in the literature, several recent articles discuss the relation between the number of analysts who follow a stock and information asymmetry or adverse selection costs. Early articles used the number of analysts as a direct proxy for informed trading, but recent studies by Brennan and Subrahmanyam (1995) and Easley, O'Hara, and Paperman (1998) find that firms that are followed by a larger number of analysts have a lower degree of information asymmetry or lower adverse selection costs.[29] Thus

[29] See Easley, O'Hara, and Paperman (1998) for a discussion of the issue and additional references.

the number of analysts appears to be negatively related to the degree of information asymmetry.

Using the number of analysts as a proxy for the degree of information asymmetry has an intuitive appeal since it directly relates to information production in the market. Nonetheless, this empirical proxy has its share of problems. First, there is still some doubt about the direction and strength of the proxy's relation to the degree of information asymmetry, as work by Brennan and Subrahmanyam (1995) and Easley, O'Hara, and Paperman (1998) suggests. Second, many stocks are not regularly followed by analysts. Third, there is relatively little cross-sectional variation in the number of analysts who follow stocks. Therefore there are reasons to believe that the number of analysts will not exhibit as strong a cross-sectional relation with C2 as will our two main proxies, bid-ask spread and market capitalization.

To construct the analyst-following proxy, we look for the monthly number of analysts who provide I/B/E/S with end-of-fiscal-year earnings forecast for the current year. We define NumEst_i to be the average monthly number of analysts over the sample period (six years). ORDEST_i is the ordinal scale of NumEst_i (constructed like ORDBA and ORDCAP), where two firms that have the same number of analysts receive the same rank. Of the 2226 firms in our sample, 2035 are followed by at least one analyst. Not surprisingly, the majority of firms without analyst coverage are in the small size group. Only 571 of 742 firms in the small size group had forecast records in the I/B/E/S database. The average number of analysts is 2.45 for those small firms that are being followed, compared with 16.51 for firms in the large size group.

Table 10 contains the results of the cross-sectional regression in Equation (14) using either NumEst or ORDEST as the information asymmetry proxy. The coefficient of NumEst is negative and statistically significant and so is the coefficient of ORDEST, though the relation is weaker than the cross-sectional results reported in Tables 2 and 3, where we use our two main proxies. Interpreting this result is straightforward. The more analysts who cover a firm, the better the production of information about the firm's prospects. Investors in a firm with more information production have fewer opportunities to engage in speculative trading, and therefore most trading in these firms' securities is motivated by hedging. Our results are consistent with the evidence in Brennan and Subrahmanyam (1995) and Easley, O'Hara, and Paperman (1998), who find that the number of analysts is negatively related to the degree of information asymmetry.

3.6 Firm-specific information asymmetry

It is possible, and even likely, that both market-wide and firm-specific factors drive the trading and returns of individual stocks. In the model presented in Section 2, trading is only generated by hedging needs and private information that are firm specific. Thus we focus on the different degree of firm-specific

Table 10
Analyst following as an information asymmetry proxy

$C2_i = a + b * NUMEST_i + ERROR_i$

a	b	R^2 (%)	Observations
0.016435	−0.000760	1.395	2035
(9.789)	(−5.362)		

$C2_i = a + b * ORDEST_i + ERROR_i$

a	b	R^2 (%)	Observations
0.022239	−0.026790	2.611	2035
(10.970)	(−7.383)		

We report the results of the cross-sectional regressions:

$$C2_i = a + b * NUMEST_i + ERROR_i$$
$$C2_i = a + b * ORDEST_i + ERROR_i,$$

The $C2_i$ parameter is estimated from the following time-series regression:

$$Return_{i,t+1} = C0_i + C1_i * Return_t + C2_i * Volume_{i,t} * Return_{i,t} + error_{i,t+1},$$

$NUMEST_i$ is the average monthly number of analysts over the sample period (six years). $ORDEST_i$ is a variable representing the ordinal scale of $NUMEST_i$, where two firms that have the same number of analysts receive the same rank. t-statistics appear in parentheses.

private information as the main factor that produces the cross-sectional variation in the dynamic volume-return relation. The model's prediction on the relation between volume and return autocorrelation would therefore most reasonably apply to the firm-specific components of trading and returns.

We can test empirically whether the volume-return relation has a firm-specific component or if it disappears when we eliminate market-wide variations. To do so we use market models to decompose both the volume and return series. Each series is decomposed into a systematic (market) component and a nonsystematic (firm-specific) component. To implement the market models we construct market return and volume series. The market's return for a specific day is defined as the return on a value-weighted portfolio comprised of all common stocks that traded on the NYSE or AMEX, and which have valid return and volume information in the CRSP database for that day. We define the market's turnover in an analogous fashion as the value-weighted average of the turnover of the individual stocks in the portfolio of all NYSE and AMEX common stocks. To maintain compatibility, we detrend the log market turnover series, just as we do with the turnover series of each individual stock. While sensible, this article does not explicitly model this particular volume decomposition. In a recent study, Lo and Wang (2000) present a formal justification for a market model of volume, which we use here.

We reestimate Equation (12) by using residual returns and residual volume from the respective market models. We then examine cross-sectional differences in the resulting C2 coefficients of the volume-return interaction terms.

Table 11
Firm-specific information asymmetry and return-volume relation

$C2_i = a + b * ORDBA_i + ERROR_i$

a	b	R^2 (%)	Observations
0.000740	0.038645	5.409	2226
(0.374)	(11.277)		

$C2_i = a + b * ORDCAP_i + ERROR_i$

a	b	R^2 (%)	Observations
0.033270	−0.026386	2.522	2226
(16.559)	(−7.585)		

The table presents the results of the cross-sectional regressions:

$$C2_i = a + b * ORDBA_i + ERROR_i$$
$$C2_i = a + b * ORDCAP_i + ERROR_i.$$

The $C2_i$ parameter is estimated from the following regression:

$$\text{ResReturn}_{i,t+1} = C0_i + C1_i * \text{ResReturn}_{i,t} + C2_i * \text{ResVolume}_{i,t} * \text{ResReturn}_{i,t} + \text{error}_{i,t+1},$$

where $\text{ResReturn}_{i,t}$ is the daily abnormal return of an individual stock (using the market model), and $\text{ResVolume}_{i,t}$ is the residual volume of an individual stock. ORDBA and ORDCAP are variables representing the ordinal scales of BAsprd and AvgCap. t-statistics appear in parentheses.

Table 11 presents the cross-sectional regressions of the information asymmetry proxies on the C2 coefficients. The coefficient of ORDBA is positive and statistically significant, and the coefficient of ORDCAP is negative and statistically significant. These results suggest that firm-specific information asymmetry is a driving force behind the relation between volume and return autocorrelations.

3.7 Coefficient C1 and return autocorrelation

So far we have focused on the cross-sectional variation in C2 from Equation (12). The main reason is given in Section 1.3. However, as Tables 2 and 3 show, C1 also exhibits clear cross-sectional variation. As mentioned earlier, C1 describes the autocorrelation of returns holding volume at its average level. From Table 2, the mean of C1 is positive but very close to zero for stocks with large bid-ask spreads (though the parameter is negative for almost half of the stocks in this group). From Proposition 3, $C1 = -(\theta_1 + \theta_2 \mathrm{E}[\widetilde{V}^2])$. If $\mathrm{E}[\widetilde{V}^2]$ is small (i.e., volume has small variance), then $C1 \simeq -\theta_1$, which decreases with information asymmetry. In other words, stocks with little information asymmetry are predicted to have C1 that is close to zero. Increasing information asymmetry decreases C1: From Table 2, the parameter is negative for 609 of the 742 stocks in the high bid-ask spread group, with a mean of −0.12, consistent with the prediction of the model.

Table 3 gives basically the same result using market capitalization as the information asymmetry proxy. Panel A shows that about half of the stocks in both the high and medium market capitalization groups have a negative

C1 parameter, while the other half is positive. The mean of both groups is therefore very close to zero. However, 557 of 742 stocks in the low market capitalization, or high information asymmetry, group have negative C1 coefficient, and the mean of the group is significantly negative (−0.10).

The pattern in C1 estimates is consistent with our model, but also with other models [e.g., Brown and Jennings (1989) and Wang (1993)]. The results concerning C1 are also closely related to the short-horizon return autocorrelations literature. Many studies show that short-horizon returns of individual stocks exhibit negative autocorrelation [e.g., French and Roll (1986), Lo and MacKinlay (1988), Conrad, Kaul, and Nimalendran (1991), Jegadeesh and Titman (1995), Canina et al. (1998)]. These autocorrelations are more pronounced in small stocks than in large stocks. French and Roll (1986) and Jegadeesh and Titman (1995) show that the first-order autocorrelation of daily returns is negative for small stocks, increases with the size of the firm, and is positive for large firms. The stocks in our sample exhibit similar return characteristics. The first-order autocorrelation of daily returns is negative for stocks with large bid-ask spreads (−0.088) and small sizes (−0.076). It is positive but very small for large stocks (0.003) and stocks with small bid-ask spreads (0.01). These autocorrelations are similar in sign and relative magnitude to the C1 coefficients from Tables 2 and 3.

Lo and MacKinlay (1988) suggest that these empirical findings are consistent with security returns that reflect three influences: a positively autocorrelated common component, a white noise component, and a negative autocorrelation effect induced by microstructure phenomena such as bid-ask bounce. The pattern in C1 is consistent with this explanation. French and Roll (1986) suggest that the positive autocorrelation arises when the market does not incorporate information as soon as it is released. While prices in our model are partially revealing, the model predicts that stocks with more information asymmetry should have a more negative C1 coefficient, which is what we find empirically. Jegadeesh and Titman (1995) attribute the negative autocorrelation to inventory control by specialists. Since specialists participate more often in the trading of small stocks than in the trading of large stocks [Madhavan and Sofianos (1998)], inventory control is also consistent with the finding of more negative C1 coefficient for small stocks.[30]

Obviously various possible sources can lead to the observed pattern of C1. It would certainly be interesting to be able to further identify the true sources. But it requires additional theoretical and/or empirical input. Given the simple theoretical model we have and the data available to us, we are unable to distinguish these alternative sources. However, this is not the objective of our article. Our objective is to examine the joint behavior of return *and* volume and to link them to the underlying trading motives. Our theoretical

[30] Another explanation suggested in the literature is that return autocorrelation is a result of changes in systematic risk [see, e.g., Conrad and Kaul (1988)].

model leads to sharp and robust predictions on C2 (not C1), which naturally becomes the focus of our empirical analysis.

4. Conclusions

We consider a simple model in which investors trade in the stock market for both hedging and speculation motives. We use the model to investigate the dynamic relation between volume and returns. According to our model, returns generated by hedging-motivated trades reverse themselves, while returns generated by speculation-motivated trades tend to continue themselves. The relative significance of these two types of trades for an individual stock determines whether returns that are accompanied by trading volume exhibit negative or positive autocorrelation.

We test the model's predictions by using daily return and volume data for NYSE and AMEX stocks. We look at how volume affects the first-order autocorrelation of daily stock returns. To proxy for information asymmetry we use bid-ask spreads (larger bid-ask spreads imply a higher degree of information asymmetry) and market capitalization (larger firms are associated with less information asymmetry).

The empirical results support the predictions of the model on the nature of the dynamic volume-return relation. Stocks that are associated with a high degree of informed trading exhibit more return continuation on high-volume days, and stocks that are associated with a low degree of informed trading show more return reversals on high-volume days. Our results are robust to various econometric specifications, potential data problems, alternative definitions of volume, and changes in the lengths of the measurement intervals and the estimation period.

We use analyst following as an additional proxy for the degree of information production about a firm. We find that in the portion of our sample for which we could find data on analyst following, the dynamic volume-return relation shows the same pattern as with the other information asymmetry proxies. We also investigate whether firm-specific information asymmetry is a driving force behind this relation. We use market models to decompose returns and volume, and find that the relation holds even when we use only firm-specific (residual) returns and volume.

The empirical findings support the general notion that volume does tell us something about future price movements. The analysis also suggests that the actual dynamic relation between volume and returns depends on the underlying forces driving trading. Explicitly modeling these driving forces allows us to use volume effectively in making an inference about returns. In particular, by considering both allocational and informational trading, our model gives rise to realistic predictions that seem to encompass the variety that prevails in the market. It is this feature of the model that enables us to reconcile the previous empirical findings of return reversals after high-volume days exhibited by large firms and indices, with the return continuation after high-volume days shown by average firms. The key to generating both results is our ability to use information asymmetry to capture the cross-sectional variation in the dynamic volume-return relation of individual stocks.

Appendix

Proof of Proposition 1. We consider the special case when $Z_t^{(2)} = 0 \; \forall t$. Extending the result to the general case is straightforward. We reexpress the price as

$$P_t = F_t + \widetilde{P}_t,$$

where $\widetilde{P}_t \equiv a(G_t - bZ_t)$, $b \equiv b^{(1)}/a$ and $Z_t \equiv Z_t^{(1)}$. We have

$$E_t^{(i)}[R_{t+1}] = E_t^{(i)}[D_{t+1}] - P_t = E_t^{(i)}[G_t] - \widetilde{P}_t \quad \text{and} \quad \sigma_R^{(i)2} = \sigma_D^{(i)2} + \sigma_F^2 + \sigma_{\widetilde{P}}^2,$$

where $\sigma_D^{(i)2} = E_t^{(i)}[G_t^2]$ and $\sigma_{\widetilde{P}}^2 = a^2(\sigma_G^2 + b^2\sigma_Z^2)$. Note that $E_t^{(1)}[G_t] = G_t$, $E_t^{(2)}[G_t] = \gamma(G_t - bZ_t)$, $\sigma_{DD}^{(1)} = 0$, and $\sigma_D^{(2)2} = \gamma\sigma_G^2$, where $\gamma = (\sigma_G^2 + b^2\sigma_Z^2)^{-1}\sigma_G^2$. Investors' stock demands are

$$X_t^{(1)} = (G_t - \widetilde{P}_t - \sigma_{DN}Z_t)/\sigma_R^{(1)2} \quad \text{and} \quad X_t^{(2)} = [\gamma(G_t - bZ_t) - \widetilde{P}_t]/\sigma_R^{(2)2}.$$

Market clearing requires that $0 = \omega X_t^{(1)} + (1-\omega)X_t^{(2)}$. Substituting in the investors' stock demands, we have

$$0 = \omega(1-a)/\sigma_R^{(1)2} + (1-\omega)(\gamma - a)/\sigma_R^{(2)2}$$
$$0 = \omega(\sigma_{DN} - ab)/\sigma_R^{(1)2} + (1-\omega)(\gamma - a)b/\sigma_R^{(2)2}.$$

We can immediately solve for b: $b = \lambda\sigma_{DN}$. Note that γ and $\sigma_{DD}^{(i)}$ $(i = 1, 2)$ depend only on b, not on a. We have the following equation for a:

$$0 = \omega\sigma_R^{(2)2}(1-a) + (1-\omega)\sigma_R^{(1)2}(\gamma - a).$$

Reorganizing terms, we have

$$0 = f(a) \equiv \left[a^2\left(\sigma_G^2 + b^2\sigma_Z^2\right) + \left(\sigma_F^2 + \omega\gamma\sigma_G^2\right)\right](a - \bar{a}) - \omega\gamma\sigma_G^2(1 - \bar{a}),$$

where $\bar{a} = 1 - (1-\omega)(1-\gamma) \geq 0$. First, note that for $a < \bar{a}, f(\bar{a}) < 0$. Thus $f(a) = 0$ has no real roots less than $\bar{a}$. Second, note that $f(\bar{a}) \leq 0$ and $f(a) \to \infty$ as $a \to \infty$. Thus $f(a) = 0$ has a real root no less than $\bar{a}$. Third, $f'(a) > 0$ for $a > \bar{a}$. We conclude that $f(a) = 0$ has a unique root, which is nonnegative and greater than $\bar{a}$. ∎

Proof of Proposition 2. Suppose that x, y, z are jointly normally distributed with zero means and a covariance matrix of Σ, where

$$\Sigma = \begin{pmatrix} \sigma_{xx} & \sigma_{xy} & \sigma_{xz} \\ \sigma_{xy} & \sigma_{yy} & \sigma_{yz} \\ \sigma_{xz} & \sigma_{yz} & \sigma_{zz} \end{pmatrix} \equiv \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{12}' & \Sigma_{22} \end{pmatrix}$$

$\Sigma_{11} = \sigma_{xx}$, $\Sigma_{12} = (\sigma_{xy}, \sigma_{xz})$, and $\Sigma_{22} = ((\sigma_{yy}, \sigma_{yz}); (\sigma_{yz}, \sigma_{zz}))$. Then, we have

$$\mathrm{E}[x|y, z] = \beta_{xy} y + \beta_{xz} z,$$

where $\beta_{xy} = (\Sigma_{12}\Sigma_{22}^{-1})_1$ and $\beta_{xz} = (\Sigma_{12}\Sigma_{22}^{-1})_2$. Let $f_+ \equiv \exp\{-(\Sigma_{22}^{-1})_{yz} y|z|\}$ and $f_- \equiv \exp\{+(\Sigma_{22}^{-1})_{yz} y|z|\}$. Then

$$\begin{aligned}\mathrm{E}[x|y, |z|] &= \beta_{xy} y + \frac{f_+ - f_-}{f_+ + f_-}\beta_{xz}|z| = \beta_{xy} y - \beta_{xz}|z| \tanh\left[\left(\Sigma_{22}^{-1}\right)_{yz}|z| y\right] \\ &\approx \beta_{xy} y - \beta_{xz}\left(\Sigma_{22}^{-1}\right)_{yz}|z|^2 y.\end{aligned}$$

For our purpose, let $x = R_{t+1}$, $y = R_t$, and $z = \widetilde{P}_t - \widetilde{P}_{t-1}$. Then $\sigma_{xx} = \sigma_{yy} = (1-a)^2\sigma_G^2 + a^2\sigma_{DN}^2\sigma_Z^2 + \sigma_R^2$, $\sigma_{zz} = 2\sigma_{\tilde{P}}^2$, $\sigma_{xy} = a\sigma_G^2 - \sigma_{\tilde{P}}^2$, $\sigma_{xz} = a\sigma_G^2 - \sigma_{\tilde{P}}^2$, and $\sigma_{yz} = 2\sigma_{\tilde{P}}^2 - a\sigma_G^2$, where $\sigma_R^2 = \sigma_F^2 + \sigma_{\tilde{P}}^2$. Then we have

$$\begin{aligned}\mathrm{E}\left[R_{t+1}|R_t, |\widetilde{P}_t - \widetilde{P}_{t-1}|\right] &= \beta_1 R_t - \beta_2|\widetilde{P}_t - \widetilde{P}_{t-1}| \tanh\left(\eta|\widetilde{P}_t - \widetilde{P}_{t-1}| R_t\right) \\ &\approx -\left(\theta_1 + \theta_2'|\widetilde{P}_t - \widetilde{P}_{t-1}|^2\right) R_t,\end{aligned}$$

where

$$\begin{aligned}\beta_1 &= \beta_{xy} = -\frac{a\sigma_G^2(\sigma_{\tilde{P}}^2 - a\sigma_G^2)}{2\sigma_D^2\sigma_{\tilde{P}}^2 - a^2\sigma_G^4} \\ \beta_2 &= \frac{(\sigma_D^2 - a\sigma_G^2)(\sigma_{\tilde{P}}^2 - a\sigma_G^2)}{2\sigma_D^2\sigma_{\tilde{P}}^2 - a^2\sigma_G^4} \\ \eta &= \frac{2\sigma_{\tilde{P}}^2 - a\sigma_G^2}{2\sigma_D^2\sigma_{\tilde{P}}^2 - a^2\sigma_G^4} \\ \theta_1 &= -\beta_1 \\ \theta_2' &= \beta_2\eta\end{aligned}$$

and $\sigma_D^2 = \sigma_F^2 + \sigma_G^2$. Since $\sigma_{\tilde{P}}^2 - a\sigma_G^2 = a(\sigma_G^2 + b^2\sigma_Z^2)(a - \gamma)$ and $a \geq \bar{a} \geq \gamma$, we have $\beta_1 \leq 0$, $\beta_2 \geq 0$ and $\eta \geq 0$. Further note that $|\widetilde{P}_t - \widetilde{P}_{t-1}| = \sqrt{(2/\pi)}\sigma_{\tilde{P}}\widetilde{V}_t$. Letting $\theta_2 = 2\sigma_{\tilde{P}}^2\theta_2'/\pi$, we obtain Equation (9). ■

Proof of Proposition 3. For $\sigma_G^2 = 0$, $a = \omega$, $\theta_2' = 1/(2\sigma_D^2)$, $\theta_1 = 0$, and

$$\theta_2 = \theta_{20} = \frac{\sigma_{\tilde{P}}^2}{\pi\sigma_D^2} \equiv \frac{\omega^2}{\pi}\frac{\sigma_{DN}^2\sigma_Z^2}{\sigma_D^2}.$$

For σ_G^2 small and holding σ_D^2, $\sigma_{\tilde{P}}^2$ constant, we have

$$\theta_1 = \frac{1}{2}a\frac{\sigma_G^2}{\sigma_D^2} + o(\sigma_G^2)$$

and

$$\theta_2 \approx \frac{1}{\pi}\frac{\sigma_D^2}{\sigma_{\tilde{P}}^2}\left[1 - a\left(\frac{\sigma_G^2}{\sigma_D^2} + \frac{3}{2}\frac{\sigma_G^2}{\sigma_{\tilde{P}}^2}\right)\right] + o(\sigma_G^2).$$

Realizing that $a = \omega + o(\sigma_G^2)$, we obtain Equations (10) and (11). ■

References

Ajinkya, B. B., and P. C. Jain, 1989, "The Behavior of Daily Stock Market Trading Volume," *Journal of Accounting and Economics*, 11, 331–359.

Amemiya, T., 1978, "A Note on a Random Coefficients Model," *International Economic Review*, 19, 793–796.

Antoniewicz, R. L., 1993, "Relative Volume and Subsequent Stock Price Movements," working paper, Board of Governors of the Federal Reserve System.

Bernardo, A. E., and K. L. Judd, 1999, "Volume and Price Formation in an Asset Trading Model with Asymmetric Information," working paper, UCLA.

Bhattacharya, U., and M. Spiegel, 1991, "Insiders, Outsiders, and Market Breakdowns," *Review of Financial Studies*, 4, 255–282.

Blume, L., D. Easley, and M. O'Hara, 1994, "Market Statistics and Technical Analysis: The Role of Volume," *Journal of Finance*, 49, 153–181.

Brennan, M., and A. Subrahmanyam, 1995, "Investment Analysis and Price Formation in Securities Markets," *Journal of Financial Economics*, 38, 361–381.

Breusch, T. S., 1978, "Testing for Autocorrelation in Dynamic Linear Models," *Australian Economic Papers*, 17, 334–355.

Brown, D. P., and R. H. Jennings, 1989, "On Technical Analysis," *Review of Financial Studies*, 2, 527–551.

Campbell, J. Y., S. J. Grossman, and J. Wang, 1993, "Trading Volume and Serial Correlation in Stock Returns," *Quarterly Journal of Economics*, 108, 905–939.

Canina, L., R. Michaely, R. Thaler, and K. Womack, 1998, "Caveat Compounder: A Warning About Using the Daily CRSP Equal-Weighted Index to Compute Long-Run Excess Returns," *Journal of Finance*, 53, 403–416.

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

Conrad, J., and G. Kaul, 1988, "Time Variations in Expected Returns," *Journal of Business*, 61, 409–425.

Conrad, J., A. Hameed, and C. M. Niden, 1992, "Volume and Autocovariances in Short-Horizon Individual Security Returns," *Journal of Finance*, 49, 1305–1329.

Conrad, J., G. Kaul, and M. Nimalendran, 1991, "Components of Short-Horizon Individual Security Returns," *Journal of Financial Economics*, 29, 365–384.

Cready, W. M., and R. Ramanan, 1991, "The Power of Tests Employing Log-Transformed Volume in Detecting Abnormal Trading," *Journal of Accounting and Economics*, 14, 203–214.

Duffee, G., 1992, "Trading Volume and Return Reversals," Finance and Economics Discussion Paper Series no. 192, Board of Governors of the Federal Reserve System.

Easley, D., and M. O'Hara, 1987, "Price, Trade Size, and Information in Securities Markets," *Journal of Financial Economics*, 18, 69–90.

Easley, D., M. O'Hara, and J. Paperman, 1998, "Are Financial Analysts Informed Traders?" *Journal of Financial Markets*, 1, 175–201.

French, K. R., and R. Roll, 1986, "Stock Return Variances: The Arrival of Information and the Reaction of Traders," *Journal of Financial Economics*, 17, 5–26.

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

Glosten, L., and P. Milgrom, 1985, "Bid, Ask and Transaction Prices in a Market-Maker Market with Heterogeneously Informed Traders," *Journal of Financial Economics*, 14, 71–100.

Godfrey, L. G., 1978a, "Testing Against General Autoregressive and Moving Average Error Models When the Regressors Include Lagged Dependent Variables," *Econometrica*, 46, 1293–1302.

Godfrey, L. G., 1978b, "Testing Against Higher Order Serial Correlation in Regression Equations When the Regressors Include Lagged Dependent Variables," *Econometrica*, 46, 1303–1310.

Grundy, B., and M. McNichols, 1989, "Trade and Revelation of Information Through Prices and Direct Disclosure," *Review of Financial Studies*, 2, 495–526.

Hasbrouck, J., 1988, "Trades, Quotes, Inventories, and Information," *Journal of Financial Economics*, 22, 229–252.

Hasbrouck, J., 1991, "Measuring the Information Content of Stock Trades," *Journal of Finance*, 46, 179–207.

He, H., and J. Wang, 1995, "Differential Information and Dynamic Behavior of Trading Volume," *Review of Financial Studies*, 8, 919–972.

Hong, H., and J. Wang, 2000, "Trading and Returns Under Periodic Market Closures," *Journal of Finance*, 55, 297–354.

Jegadeesh, N., and S. Titman, 1995, "Short-Horizon Return Reversals and the Bid-Ask Spread," *Journal of Financial Intermediation*, 4, 116–132.

Johnston, J., 1985, *Econometric Methods* (3rd ed.), McGraw-Hill, New York.

Jorion, P., 1990, "The Exchange-Rate Exposure of U.S. Multinationals," *Journal of Business*, 63, 331–345.

Kyle, A. S., 1985, "Continuous Auctions and Insider Trading," *Econometrica*, 53, 1315–1335.

Lamoureux, C., and W. Lastrapes, 1990, "Heteroskedasticity in Stock Return Data: Volume versus GARCH Effects," *Journal of Finance*, 45, 221–229.

LeBaron, B., 1992, "Persistence of the Dow Jones Index on Rising Volume," working paper, University of Wisconsin.

Lee, C. M. C., and B. Swaminathan, 2000, "Price Momentum and Trading Volume," forthcoming in *Journal of Finance*.

Lee, C. M. C., B. Mucklow, and M. J. Ready, 1993, "Spreads, Depths, and the Impact of Earnings Information: An Intraday Analysis," *Review of Financial Studies*, 6, 345–374.

Lo, A. W., and A. C. MacKinlay, 1988, "Stock Market Prices do not Follow Random Walks: Evidence from a Simple Specification Test," *Review of Financial Studies*, 1, 41–66.

Lo, A. W., and A. C. MacKinlay, 1990, "An Econometric Analysis of Nonsynchronous Trading," *Journal of Econometrics*, 45, 181–211.

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

Madhavan, A., and G. Sofianos, 1998, "An Empirical Analysis of NYSE Specialist Trading," *Journal of Financial Economics*, 48, 189–210.

Madhavan, A., M. Richardson, and M. Roomans, 1997, "Why Do Security Prices Change? A Transaction-Level Analysis of NYSE Stocks," *Review of Financial Studies*, 10, 1035–1064.

Morse, D., 1980, "Asymmetric Information in Securities Markets and Trading Volume," *Journal of Financial and Quantitative Analysis*, 15, 1129–1148.

Richardson, G., S. E. Sefcik, and R. Thompson, 1986, "A Test of Dividend Irrelevance Using Volume Reactions to a Change in Dividend Policy," *Journal of Financial Economics*, 17, 313–333.

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

Rousseeuw, P. J., and A. M. Leroy, 1987, *Robust Regression and Outlier Detection*, Wiley, New York.

Scholes, M., and J. Williams, 1977, "Estimating Betas from Nonsynchronous Data," *Journal of Financial Economics*, 5, 309–327.

Stickel, S. E., and R. E. Verrecchia, 1994, "Evidence that Volume Sustains Price Changes," *Financial Analyst Journal*, November–December, 57–67.

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

Wang, J., 1994, "A Model of Competitive Stock Trading Volume," *Journal of Political Economy*, 102, 127–168.