Information Flow and Pricing Errors: A Unified Approach to Estimation and Testing

Thomas J. George
University of Iowa

Chuan-Yang Hwang
Hong Kong University of Science and Technology

This study examines whether rates of information flow differ between trading and non-trading periods, and whether the variances of pricing errors differ at the open and close of trading. The approach improves on existing methods by allowing for correlation between pricing errors and information flow, and by conducting inferences at the individual security level. The daytime rate of information flow is about seven times the overnight rate, and the variances of pricing errors at the open are not different from those at the close of trading. This evidence differs from existing results based on return variance ratios.

Much of the empirical market microstructure literature predicates its analysis on a model in which price changes have a permanent (information-related) component and a transitory (market-friction or dealer-cost) component. Two questions that have received a great deal of attention relate to the behavior of these components. The first is whether, and to what extent, the rate of information flow into prices differs during trading and nontrading periods. This is equivalent to asking whether the variance of the permanent component per unit time is different during periods when trading occurs than when trading does not occur. The second question is whether "pricing errors" of differing magnitudes are associated with different trading mechanisms or trading preceded by an overnight period of nontrading. This corresponds to asking whether the variance of the transitory component is different depending on which trading mechanism is used or whether trading has been halted for a time. More generally, these questions relate to whether the permanent and transitory components of price changes are heteroskedastic. Although both questions share the context of this model, existing studies typically focus

on only one issue and do not fully account for how correlation between the components affects their results.

This article develops a method to estimate the variances of the temporary and permanent components of returns jointly, in a manner that allows for heteroskedasticity and for correlation between the components. Hypotheses concerning these variances are then tested at the individual security level. Specifically examined is the extent to which information flow differs between daytime and overnight periods, and whether pricing errors are different at the open and close. This complements recent work of Hasbrouck (1991b, 1993) and Madhavan, Richardson, and Roomans (1997), who analyze the informational and frictional components of price changes within the trading day. Madhavan, Richardson, and Roomans base their inferences on a structural model. Like Hasbrouck, we use a time-series approach. Our method is not peculiar to comparisons of day/night and open/close; it could also be applied to estimation and testing of differences in information flow and pricing errors during distinct subperiods of days or weeks within which such differences are thought to exist.

The empirical estimates indicate that the rate of information flow is significantly different during daytime and overnight hours for a majority of stocks in the sample. The median daytime rate is about seven times that of the overnight rate. However, for most of the stocks in our sample, less than 28% of daytime information flow is associated with the flow of orders. This is too weak for the cessation of trading overnight to be the primary explanation for the difference in rates of information flow. Therefore the absence of other sources of information appears to be the primary reason for why value discovery slows overnight.

The evidence concerning pricing errors is that for most stocks, significant differences between the open and close *do not* exist. When differences do exist, pricing errors at the opening tend to be *smaller* than at the close. Of the 100 stocks in the sample, 11 reject the null of equal variances in favor of larger variances at the close, and only 2 reject in favor of larger variances at the open. Most studies that examine this issue using return variance ratios draw the opposite conclusion that the evidence is strong in favor of larger pricing errors at the open than the close. Our estimates also suggest that inventory effects play a larger role at the close, perhaps as a consequence of the overnight trading suspension. Nevertheless, for most of the stocks in our sample, their role is not sufficient to induce pricing errors at the close whose average magnitude is larger than that of pricing errors at the open.

The article is organized as follows. Section 1 reviews the design of existing tests and motivates our approach. Section 2 describes the analytics. Section 3 describes the data and model specification. Section 4 presents empirical results and interpretations. Results of sensitivity tests appear in Section 5. Section 6 concludes.

1. Background and Motivation

This article develops a new method to test hypotheses concerning interday price changes. However, this method is related to methods that others have used to analyze intraday (tick-by-tick) data. In this section the intraday analyses are discussed first to provide context for the approach we advocate for examining interday issues.

For the purpose of illustrating how these methods work, consider this simple structural model of transaction prices

$$p_t = m_t + \kappa_1 v_t + \kappa_2 v_{t-1}$$
$$m_t = m_{t-1} + \kappa_3(v_t - E_{t-1}[v_t|\eta_t]) + \eta_t \qquad (1)$$
$$v_t = \rho v_{t-1} + \gamma \eta_t + u_t.$$

In this model, p_t is the natural log of the transaction price at time t, and v_t is the signed order that is executed at time t. Parameters κ_1 and κ_2 reflect the impact of market frictions on transaction prices: κ_1 reflects the fact that initiators of buy orders pay more (and sell orders receive less) than the security's true value to compensate liquidity providers for the costs of market making; κ_2 reflects the possibility that liquidity providers raise (lower) their quotes when recent trading has increased (decreased) their inventory. The ρ parameter reflects the possibility of persistence in order flow, which could occur if large orders are executed in a sequence of small transactions. Thus m_t is the natural log of what the price would be in the absence of market frictions. In this model, m_t is defined to be its previous value adjusted for the information conveyed by the unexpected component of the current trade, $\kappa_3(v_t - E_{t-1}[v_t|\eta_t])$, and other (nontrade) news, η_t, that arrives between times $t-1$ and t. The γ parameter reflects the impact of news-related portfolio rebalancing on order flow.[1] The structural news and order-flow innovations, η_t and u_t, are assumed to be mutually independent and i.i.d.[2] This model is the basis of more detailed structural specifications in several empirical microstructure studies [e.g., Glosten and Harris (1988), Stoll (1989), Hasbrouck (1991a), Madhavan et al. (1997)]. If $\gamma = 0$, and the κ_j parameters are divided by

---

[1] The expected component of the time t trade, $E_{t-1}[v_t|\eta_t]$, is conditional on η_t because the news that arrives between times $t-1$ and t is known when the time t order arrives.

[2] Under these assumptions m_t is a random walk, and only market frictions generate predictability in price changes. Though typical of empirical microstructure models, this assumption is not necessarily implied even by standard asset pricing models without market frictions or asymmetric information [see Kirby (1998)]. The presumption in microstructure studies is that predictability in prices associated with changes in compensation for bearing systematic risk manifests itself over horizons that are long enough that it would not be detected and misconstrued as market frictions in microstructure studies. In other words, over short time intervals (e.g., a day or less), the predictability in prices caused by market frictions sufficiently dominates that associated with changes in compensation for bearing systematic risk that the latter can be ignored. In this article, we compare predictability of price changes at the open and close. There is no reason to believe that an asset pricing model would generate asymmetric predictability between the open and close. We therefore attribute differences in this predictability to market frictions.

half of the quoted spread, this is a standard model of the components of the bid-ask spread [e.g., George, Kaul, and Nimalendran (1991), and Huang and Stoll (1997)]—i.e, $\kappa_1/(\frac{S}{2})$ is the order-processing component, $\kappa_2/(\frac{S}{2})$ is the inventory component, and $\kappa_3/(\frac{S}{2})$ is the adverse-selection component.

It is often the case in empirical microstructure work that the researcher wishes to test hypotheses concerning variation in security prices that could apply to a class of parametric models. In particular, he or she may simply want to quantify the importance of information and market frictions in explaining variation in security prices. One way to estimate these quantities is first to estimate the parameters of a specific structural model such as Equation (1), then compute $\mathrm{var}\,[\kappa_3(v_t - E_{t-1}[v_t|\eta_t]) + \eta_t]$ and $\mathrm{var}\,[\kappa_1 v_t + \kappa_2 v_{t-1}]$ from the parameter estimates of the structural model. The problem is that if Equation (1) is misspecified, these variance estimates will not be valid.

In two important articles, Hasbrouck (1991b, 1993) demonstrates that it is not necessary to have a detailed structural model in order to estimate the variance in returns attributable to information flow and market frictions. This is useful because it allows estimation of these quantities without the maintained hypotheses, and possible specification error, associated with a particular detailed model. Hasbrouck shows that the information and market friction components of return variance can be estimated provided that the price process can be represented as

$$p_t = m_t + s_t$$
$$m_t = m_{t-1} + \epsilon_t, \tag{2}$$

where m_t is a random walk, and s_t is a covariance stationary stochastic process whose unconditional expectation is zero. Note that Equation (1) fits into this framework with $\epsilon_t = \kappa_3(v_t - E_{t-1}[v_t|\eta_t]) + \eta_t$ and $s_t = \kappa_1 v_t + \kappa_2 v_{t-1}$. With models such as Equation (1) in mind, the variances of ϵ_t and s_t can be interpreted as measures of the impact of information and market frictions on security prices. More generally, changes in security prices that are *permanent* are attributed to information, while changes that are *temporary* are attributed to market frictions and are referred to as "pricing errors."[3]

Beveridge and Nelson (1981) and Watson (1986) show how to identify the variances of ϵ_t and s_t from the moving-average representation of first differences of a univariate series that can be modeled as in Equation (2). Hasbrouck (1991b, 1993) generalizes their results to vector time series (e.g., a vector of price or quote changes and order flow). The intuition is as follows.

---

[3] This term is simply shorthand for "deviation of the current price from its value in a frictionless market," and is not meant to convey that it arises because of irrationality on the part of traders. However, if traders display irrational behavior in a manner that contributes to predictability in price changes, then irrationality will be a determinant of the "pricing error."

The permanent price change, ϵ_t can be written as a revision between times $t-1$ and t of an infinite-horizon forecast of the change in p (see Appendix A),

$$\epsilon_t = \lim_{n\to\infty}\{(E_t[p_{t+n}] - p_{t-1}) - (E_{t-1}[p_{t+n}] - p_{t-1})\}.$$

Intuitively, ϵ_t is the eventual (and hence permanent) impact on p of the conditioning information that arrives between times $t-1$ and t. Likewise, s_t is the component of the price that is expected eventually to disappear (it is temporary). So s_t can be written as the forecast at t of how the price eventually will change from its current level,

$$-s_t = \lim_{n\to\infty}\{E_t[p_{t+n}] - p_t\}.$$

The forecasts that are denoted by these conditional expectations can be made from the univariate moving average representation of first differences of p_t, or from a vector moving average that includes other variables such as order flow. Therefore the *variances* of ϵ_t and s_t can be computed in terms of the parameters and disturbance variance of the moving-average representation.

This approach to identifying the variances of ϵ_t and s_t is due to Beveridge and Nelson. Watson points out that, in general, randomness from sources that are independent of those that cause permanent changes in prices could also affect s_t (modeled by adding an orthogonal random component to the equation for $-s_t$ just above). Though the addition of an orthogonal component would be well suited to capture some microstructure effects (particularly rounding errors due to price discreteness) the variance of s_t is not econometrically identified in this case. However, Watson shows that the variance of s_t is identified if *all* of it comes from the orthogonal component; that is, if s_t is assumed to be *orthogonal* to all polynomial lags of innovations that cause permanent changes in prices. This assumption is not reasonable in a microstructure context. Even in a model as simple as Equation (1), s_t and ϵ_t are correlated because v_t is common to both. Order arrivals convey information *and* cause prices to deviate from true values. Hasbrouck (1993) discusses this identification issue in detail and advocates using the Beveridge–Nelson approach and a *vector* moving average representation that includes both price changes and order flow. He argues that variance contributed to s_t by shocks that are orthogonal to innovations in price changes *and* order flow should be small.

Hasbrouck (1991b) estimates the variance of ϵ_t from NYSE specialists' quote revisions and measures of order flow. He also decomposes that variance into trade-correlated and trade-uncorrelated components to compare the impact of private versus public information on quote revisions. Hasbrouck (1993) estimates the variance of s_t using intraday transaction prices to examine the proportion of variation in trade-to-trade returns that is attributable to market frictions. Taken together, Hasbrouck's (1991b, 1993) analyses provide variance estimates of the components of tick-by-tick changes in security

prices that are less affected by specification error than are estimates based on a specific structural model.

Related issues have been examined using interday data. Two questions that have been studied extensively are whether rates of information flow differ during trading and nontrading periods (e.g., day versus overnight), and whether the impact of market frictions on prices is different at different times in the trading day. These issues have interested researchers because they are important to understanding the extent to which trading itself facilitates the incorporation of information into prices, and whether the use of different trading mechanisms at open and close lead to noticeable differences in price variability related to market frictions.

These questions can be answered without a detailed structural model. They are equivalent to asking whether the variance of ϵ_t differs between daytime and overnight periods, and whether the variance of s_t at the open is different than at the close. In fact, existing studies that compare daytime and overnight price changes base their analyses on a generalization of Equation (2) that accommodates heteroscedasticity in both information and pricing errors:

$$
\begin{aligned}
p_{ct} &= m_{ct} + s_{ct} \\
m_{ct} &= m_{ot} + \epsilon_{ct} \\
p_{ot} &= m_{ot} + s_{ot} \\
m_{ot} &= m_{c,t-1} + \epsilon_{ot},
\end{aligned}
\tag{3}
$$

where the *ot* and *ct* subscripts refer to the open and close on day t [see, e.g., Stoll and Whaley (1990), Jones, Kaul, and Lipson (1994), George and Hwang (1995), Forster and George (1996), Ronen (1997)].[4] Differences in information flow during daytime and overnight periods correspond to differences between var$[\epsilon_{ct}]$ and var$[\epsilon_{ot}]$. Since s_{ot} and s_{ct} have unconditional means of zero, their variances measure the magnitudes of pricing errors at the open and close: var$[s_{ot}] = E[s_{ot}^2]$ and var$[s_{ct}] = E[s_{ct}^2]$.

Although those studies avoid the specification error that is possible with a detailed structural model, their methods for estimation and testing lead to clear conclusions only if pricing errors and information are *uncorrelated*. This assumption is not likely to be satisfied in markets with asymmetric information and costly liquidity provision. In such markets, order arrivals both convey information and impose an order-processing or inventory cost on liquidity providers. This results in correlation between the information and market-friction components of price changes.

The remainder of this section reviews these methods and argues that tests based on a generalization to Beveridge–Nelson that accommodates the heteroscedasticity depicted in Equation (3) would be an improvement. The reason for the improvement is that the Beveridge–Nelson technique delivers tests

---

[4] The random vector of pricing errors $(s_{ct}, s_{ot})'$ is assumed to be covariance stationary with unconditional expectation zero. The change in security value during the day and night $(\epsilon_{ct}, \epsilon_{ot})'$ is assumed to be a zero-mean i.i.d. random vector; with ϵ_{ct} and ϵ_{ot} uncorrelated as they relate to distinct time periods.

with clear interpretations without requiring information and pricing errors to be uncorrelated. Section 3 shows how this generalization can be achieved.

French and Roll (1986) examine whether information flows at different rates during trading and nontrading periods by analyzing ratios of variances of nonoverlapping returns computed from weekday and weekend (or weekday holiday) periods. For a comparison between daytime and overnight periods, their statistic is

$$\frac{\text{var}\,[p_{ct} - p_{ot}]}{\text{var}\,[p_{ot} - p_{ct-1}]} = \frac{\text{var}\,[\epsilon_{ct}] + \text{var}\,[s_{ct} - s_{ot}] + 2\text{cov}\,[\epsilon_{ct}, s_{ct} - s_{ot}]}{\text{var}\,[\epsilon_{ot}] + \text{var}\,[s_{ot} - s_{ct-1}] + 2\text{cov}\,[\epsilon_{ot}, s_{ot} - s_{ct-1}]}. \quad (4)$$

If pricing errors are nonexistent (or nonrandom), terms involving s disappear and this statistic measures differences in information flow into prices between daytime and overnight periods. Jones, Kaul, and Lipson (1994) analyze trading and nontrading days of NASDAQ stocks using variance ratios that are similar to those of French and Roll. They assume that pricing errors and information are independent, and note that the presence of the middle variance term in the numerator and denominator bias these ratios toward one. Note, however, that the nature of the bias is actually indeterminate because the permanent and transitory components of price changes are correlated. Since order arrivals convey information to liquidity providers *and* cause transaction prices to bounce between bid and ask, the covariances in Equation (4) are not likely to be zero.

Amihud and Mendelson (1987, 1991), Stoll and Whaley (1990), and others examine whether the magnitudes of pricing errors are different at the open and close of trading to test whether the existence of an overnight nontrading period or differences in the trading mechanisms used at the open and close affect the transitory component of price changes. They analyze ratios of overlapping 24-hour returns measured open to open and close to close:

$$\frac{\text{var}[p_{ot} - p_{ot-1}]}{\text{var}[p_{ct} - p_{ct-1}]}$$

$$= \frac{\left\{\begin{array}{l}\text{var}[\epsilon_{ot} + \epsilon_{ct-1}] + 2\,\text{var}[s_{ot}] \\ \quad + 2(\text{cov}[s_{ot} - s_{ot-1}, \epsilon_{ot} + \epsilon_{ct-1}] - \text{cov}[s_{ot}, s_{ot-1}])\end{array}\right\}}{\left\{\begin{array}{l}\text{var}[\epsilon_{ot} + \epsilon_{ct}] + 2\,\text{var}[s_{ct}] \\ \quad + 2(\text{cov}[s_{ct} - s_{ct-1}, \epsilon_{ot} + \epsilon_{ct}] - \text{cov}[s_{ct}, s_{ct-1}])\end{array}\right\}}. \quad (5)$$

The leading variance term in the numerator and denominator are equal. If all the covariances are zero, then finding that this ratio is significantly greater than one is evidence that the magnitude of pricing errors at the open is greater than at the close. However, if the covariances are not all zero, the interpretation of this variance ratio is ambiguous.

We are not the first to point out the deficiencies in tests based on Equations (4) and (5). Jones, Kaul, and Lipson (1994) and Smith (1995) were

aware of biases due to pricing errors, and Stoll and Whaley (1990) were aware of biases due to covariances between pricing errors and information. However, highlighting the problems illustrates that an ideal test would *isolate* the variances of information innovations and pricing errors. This is precisely the purpose of the Beveridge–Nelson technique. By extending the Beveridge–Nelson results to handle heteroskedastic information and pricing errors, as depicted in Equation (3), we can dispense with Equations (4) and (5) and use ratios such as $\mathrm{var}[\epsilon_{ct}]/\mathrm{var}[\epsilon_{ot}]$ and $\mathrm{var}[s_{ot}]/\mathrm{var}[s_{ct}]$ as the basis for hypothesis testing. The details of this extension are presented in the next section.

Since our approach generalizes Beveridge–Nelson, variables that are thought a priori to be determinants of pricing errors should be included in the time-series model because exclusion of relevant variables leaves variation in pricing errors undetected. As Hasbrouck (1993) points out, order flow is likely to be the most important variable of this type. Therefore estimates in this article are based on a time-series model that includes both returns and order flow. The presence of order flow also makes it possible to examine the association between order flow and the components of return variance attributable to information and market frictions (see Section 4.3). The strength of this association provides evidence on the importance of trading activity to the flow of information into prices.[5]

Ronen (1997) offers a critique of the existing literature that is different from ours. She points out that most studies in this area draw inferences from a cross-sectional average of Equation (4) or (5), whose standard error is computed from the cross-sectional distribution of estimates. This approach to inference assumes that the estimated ratios are independent draws from an identical distribution. She points out that there is no reason to believe that the sampling distribution of these estimates should be the same across securities, and that the estimates are not likely to be independent because they are obtained from stock price data during a period that is common to each of the securities in the sample. Ronen shows that this approach leads to biased conclusions, and advocates estimation of variance ratios using Hansen's (1982) generalized method of moments (GMM). Inferences can then be conducted without relying on an i.i.d. assumption. Moreover, the ratios can be estimated jointly in a manner that accounts for cross-sectional correlation among the estimates. Using this procedure, she is unable to reject the null that the vector of ratios given by Equation (5) is equal to the vector of ones. Her estimates of Equation (4) are significantly greater than one and similar in magnitude to those of French and Roll, but the *p*-value of her

---

[5] In other related work, Harvey, Ruiz, and Sentana (1992) examine the case in which the variances of ϵ and s in Equation (2) exhibit ARCH. Richardson and Smith (1994), Foster and Viswanathan (1995), and Andersen (1996) estimate models of price changes and volume in which a common source of uncertainty determines the second moments of both series ("mixture" models). Pricing errors associated with market frictions (i.e., s_t) are not explicitly modeled in those studies, however.

test statistic is much lower than theirs. Thus conclusions are different when using test statistics that are not based on the cross-sectional distribution of estimates.

Our objective is to depart from Equations (4) and (5) altogether; but in light of Ronen's criticism, we implement our tests in a manner that is not based on the cross-sectional distribution of our estimates. In particular, we estimate parameters for each individual security using GMM. This gives us the (asymptotic) sampling error structure of the parameters for each security, so inferences can be conducted at the individual security level. For comparison, we also present results in terms of the cross-sectional distribution of estimates, and find them to be biased as Ronen found in her comparisons using Equations (4) and (5). Ideally we would also like to estimate parameters for all securities in our sample jointly. The number of parameters that are estimated for each security in order to identify the variances of the ϵ and s terms in Equation (3) makes joint estimation infeasible, however.

2. Details of the Analysis

This section describes our approach to estimation and hypothesis testing. The stationarity assumptions made in connection with Equation (3) imply that first differences of logarithmic transaction prices—that is, daytime and overnight transaction returns—are covariance stationary; so they have a bivariate infinite moving average representation. Since we want to augment the system with order flow variables, we need to assume that the vector containing returns and order flow has an infinite moving average representation. For estimation, we further assume that this moving average representation is invertible to a finite-order vector autoregression.

To see how a structural model can be rewritten in a way that satisfies these assumptions, consider the structural model of Equation (1). Taking first differences of p_t, substituting for $m_t - m_{t-1}$, substituting for u_t from the equation for order flow, and collecting like terms yields the system

$$r_t \equiv p_t - p_{t-1} = (\kappa_1 + \kappa_3)v_t + (\kappa_2 - \kappa_3\rho - \kappa_1)v_{t-1} - \kappa_2 v_{t-2}$$
$$+ (1 - \gamma\kappa_3)\eta_t \tag{6}$$
$$v_t = \rho v_{t-1} + \gamma \eta_t + u_t.$$

Adding some zeros yields

$$r_t = 0r_{t-1} + 0r_{t-2} + (\kappa_1 + \kappa_3)v_t + (\kappa_2 - \kappa_3\rho - \kappa_1)v_{t-1}$$
$$- \kappa_2 v_{t-2} + (1 - \gamma\kappa_3)\eta_t \tag{7}$$
$$v_t = 0r_{t-1} + 0r_{t-2} + \rho v_{t-1} + 0v_{t-2} + \gamma \eta_t + u_t.$$

Thus the structural model in Equation (1) can be written as a bivariate vector autoregression of order two, albeit with several coefficients of zero. The disturbance to the first equation is $(1 - \gamma\kappa_3)\eta_t$ and the disturbance to the second

is $\gamma\eta_t + u_t$. Note that these disturbances are correlated because order flow depends on news.[6]

The returns variables used in the model we estimate are denoted, $r_{nt} \equiv p_{ot} - p_{ct-1}$ and $r_{dt} \equiv p_{ct} - p_{ot}$—the overnight and daytime transaction returns ending on day t. The order flow variables, denoted by v_{nt} and v_{dt}, are measures of order flow at the opening and during the daytime (excluding the opening transaction) on day t, respectively. The "nt" subscript is used on opening order flow because opening order flow and the opening price (and hence the overnight return) are jointly determined. As noted above, for estimation, the vector of these variables $x_t \equiv (r_{dt}, r_{nt}, v_{dt}, v_{nt})'$, is assumed to have a finite (say, pth order) autoregressive representation:

$$A\mathbf{x}_t = B_1\mathbf{x}_{t-1} + \cdots + B_p\mathbf{x}_{t-p} + \mathbf{u}_t, \qquad \mathbf{u}_t \sim \text{i.i.d. } (0, \Omega). \tag{8}$$

Appendix B shows the standard way in which lags of $\mathbf{x}_t$ can be stacked and the parameter matrices redefined so that a VAR(p) can be written as a VAR(1):

$$A_*\mathbf{x}_{*t} = B_*\mathbf{x}_{*t-1} + \mathbf{u}_{*t}, \qquad \text{where } \mathbf{u}_{*t} \sim \text{i.i.d. } (0, \Omega_*). \tag{9}$$

A_*, B_* and Ω_* are ($4p \times 4p$) matrices, and $\mathbf{x}_{*t}$ and $\mathbf{u}_{*t}$ are ($4p \times 1$). The following result, proved in Appendix A, provides closed-form expressions for the variances of $\epsilon_{ot}, \epsilon_{ct}, s_{ot}$, and s_{ct} in Equation (3) in terms of the autoregressive parameters contained in the matrices A_*, B_*, and Ω_* of Equation (9).

Theorem 1. *With reference to Equation (9), suppose $\mathbf{x}_{*t}$ and $\mathbf{u}_{*t}$ are ($4p \times 1$). Let e_i be the ($4p \times 1$) vector with unity in the ith position and zeros elsewhere, and define the ($4p \times 4$) matrix $H \equiv (e_1\ e_2\ e_3\ e_4)$. If A_*^{-1} exists and the eigenvalues of $A_*^{-1}B_*$ lie inside the unit circle, then the variances of the random-walk innovations in Equation (3) can be written as*

$$\text{var}\,[\epsilon_{ct}] = (e_1 + \Theta'e_2)'\Lambda_{\{1;1\}}(e_1 + \Theta'e_2) + 2(e_1 + e_2)'\Lambda_{\{3;1\}}(e_1 + \Theta'e_2)$$
$$+ (e_1 + e_2)'\Lambda_{\{3;3\}}(e_1 + e_2)$$
$$\text{var}\,[\epsilon_{ot}] = (e_1 + e_2)'\Lambda_{\{24;24\}}(e_1 + e_2);$$

---

[6] An alternative specification that does not include contemporaneous order flow in the returns equation can be obtained by substituting for v_t in the first equation from the second. However, the specification that includes contemporaneous order flow is convenient for the analysis in Section 5.3, so we report results for that specification throughout the article. The conclusions drawn from the results in Tables 2–4 are not sensitive to which specification is used.

the variances of the Beveridge–Nelson stationary components can be written as

$$\text{var}\,[s_{ct}] = (e_1 + e_2)'\big(\mathcal{S} - \Lambda_{\{1234;1234\}}\big)(e_1 + e_2)$$

$$\text{var}\,[s_{ot}] = (e_1 + \Theta' e_2)'\Lambda_{\{24;24\}}(e_1 + \Theta' e_2)$$
$$+ (e_1 + \Theta' e_2)'\big(\mathcal{S} - \Lambda_{\{1234;1234\}}\big)(e_1 + \Theta' e_2);$$

and the variances of innovations in the Beveridge–Nelson stationary components can be written as

$$\text{var}\,[s_{ct} - E_{ot}[s_{ct}]] = (e_1 + e_2)'\Theta\Lambda_{\{13;13\}}\Theta'(e_1 + e_2)$$

$$\text{var}\,[s_{ot} - E_{ct-1}[s_{ot}]] = (e_1 + \Theta' e_2)'\Lambda_{\{24;24\}}(e_1 + \Theta' e_2);$$

where

$$\text{vec}\,[\mathcal{S}] = [I - \Theta \otimes \Theta]^{-1}\text{vec}\,[\Lambda_{\{1234;1234\}}], \qquad \Theta = A_*^{-1}B_*,$$

$$\Lambda_{\{\alpha_1;\alpha_2\}} = \{A_* - B_*\}^{-1}H\big(I_{\{\alpha_1\}}\Omega I_{\{\alpha_2\}}\big)H'\{(A_* - B_*)^{-1}\}',$$

$$\Omega = H'\Omega_* H, \qquad \Omega_* = E[\mathbf{u}_{*t}\mathbf{u}_{*t}'],$$

and $I_{\{\alpha_j\}}$ is the 4×4 matrix with unity in the combination of diagonal positions specified in the vector α_j and zeros elsewhere (e.g., $I_{\{13\}}$ has unity on the first and third diagonals, whereas $I_{\{1234\}}$ is the 4×4 identity matrix).[7]

Though these formulas look daunting, they are straightforward to use. To get estimates of the variances of ϵ_{ot}, ϵ_{ct}, s_{ot}, and s_{ct} in Equation (3), substitute estimates of autoregressive parameters A_*, B_*, and Ω_* into the formulas in Theorem 1. Ratios of these variances can then be used to test the hypotheses described in Section 1. The ratio $\text{var}\,[s_{ot}]/\text{var}\,[s_{ct}]$ indicates the extent to which the magnitude of pricing errors is greater at the open than at the close; and $\text{var}\,[\epsilon_{ct}]/\text{var}\,[\epsilon_{ot}]$ is a measure of the rate of information flow during daytime hours relative to overnight hours. Since pricing errors at a given point in time can relate to return and order flow shocks from the recent or distant past, $\text{var}\,[s_{ct} - E_{ot}[s_{ct}]]/\text{var}\,[s_{ct}]$ and $\text{var}\,[s_{ot} - E_{ct-1}[s_{ot}]]/\text{var}\,[s_{ot}]$ are computed as indicators of the extent to which pricing errors are related to current, rather than past, shocks to returns and order flow.[8] These variances are used in Section 4 to examine the influences of inventory control on return dynamics.

---

[7] If A is an $n \times m$ matrix, then $\text{vec}\,[A]$ denotes the nm column vector whose first n elements are the first column of A, whose second n elements are the second column of A, etc. [see, e.g., Dhrymes (1984, p. 102)].

[8] These cannot be regarded as lagged adjustment to information, since their effect is only temporary. Similar variances in terms of ϵ_{ct} and ϵ_{ot} are not meaningful because information is unforecastable by definition—that is, $E_{ot}[\epsilon_{ct}] = E_{ct-1}[\epsilon_{ot}] = 0$.

An advantage of our approach is that Theorem 1 provides *exact* expressions for the variances of ϵ_{ot}, ϵ_{ct}, s_{ot}, and s_{ct} in terms of the autoregressive parameters, rather than approximating these expressions by truncating the moving average as in Hasbrouck (1991b, 1993). The variances of ϵ_{ot}, ϵ_{ct}, s_{ot}, and s_{ct} can then be estimated as exact functions of estimates of the autoregressive parameters. These formulas are crucial to performing inference at the individual security level because there is no clear way to adjust standard errors of test statistics for the approximation errors introduced by truncation. Numerical partial derivatives can be computed from exact formulas. The asymptotic distributions of test statistics are identified in terms of these partial derivatives and the asymptotic distribution of estimators of the autoregressive parameters. This enables us to draw inferences at the individual security level and avoid the problems that Ronen (1997) shows arise when drawing inferences from a cross-sectional distribution of variance ratios.

In the empirical sections that follow, the autoregressive parameters in A_*, B_*, and Ω_* are estimated using Hansen's (1982) GMM. Estimates of the expressions in Theorem 1 (and their ratios, assuming they are finite at the true parameter values) are nonlinear but smooth functions of GMM estimators. Under standard regularity conditions, their asymptotic distributions are normal, with moments that can be estimated as well-defined functions of the GMM parameter estimates and covariance matrix [see Hamilton (1994, p. 414)].

Finally, it is important to note that even if contemporaneous order flow is included in the returns equations, the disturbances to the return and order flow equations can still be correlated. Therefore the corresponding off-diagonal term of Ω_* should not be constrained to zero. This is apparent in Equation (7). When $\gamma \neq 0$, the disturbances to the equations for r_t and v_t are correlated because news affects order flow.

3. Sample Selection and Model Specification

3.1 Sample

The sample of securities is obtained from the Center for Research on Security Prices (CRSP) files. Stocks that were listed on the NYSE, but not on the London or Tokyo Stock Exchanges, and that traded on at least 800 days over the 1986–1989 period were ranked into quartiles by average daily trading volume in dollars.[9] The top 50 stocks in each quartile were identified. These 50 were sorted in alphabetical order by ticker symbol. The sample contains the first 25 of those stocks for which there were more than 100 days of data on the files of the Institute for the Study of Security Markets (ISSM) for 1986–1989. This latter screen is applied to eliminate when-issued markets.

---

[9] See Barclay, Litzenberger, and Warner (1990) and Forster and George (1995, 1996) for comparisons of internationally cross-listed and non-cross-listed stocks.

The data are drawn from the set of NYSE quotations, transaction prices, and volume files of the ISSM for 1986–1989.[10] For each stock, the time series of daytime and overnight returns is constructed using the prices of the first and last transactions of each day reported in the ISSM files, adjusted for cash distributions, stock dividends, and splits. Order flow at the opening and during the rest of the day is computed using the share volume of individual transactions reported on the ISSM tape (adjusted for splits and stock dividends), and the algorithm suggested by Lee and Ready (1991) to determine whether each order is a purchase or a sale.

Although they are computed in a similar manner, opening and daytime order flow variables have different economic interpretations. If the stock opens with a trade, it is signed using the difference between the opening price and the previous day's closing price (or midquote if the stock closes with the posting of quotes rather than with a trade); if it opens with the posting of quotes, opening order flow is zero. For actively traded stocks, opening transactions involve the crossing of several orders, which are not individually observable. Consequently, for the most active stocks, this variable is almost never a measure of net order flow at the opening, but a measure of total trading at the opening signed by the overnight return. Daytime order flow is computed by cumulating the signed individual transactions that occur throughout the day, excluding the first transaction if the stock opens with a trade. Assuming that the signing algorithm is correct, this variable is a precise measure of net order flow. Since opening and daytime order flow are treated as separate variables, the difference in their construction does not create stationarity problems that would occur if an alternating sequence of these variables were treated as a single order flow series. Nevertheless, they do have different economic interpretations, which might affect the interpretation of the results. This possibility is explored in Section 5.

Descriptive statistics are contained in Table 1. The equity capitalization of stocks in each quartile is approximately four times that of stocks in the next (lowest) quartile of dollar trading volume. This sample is similar to what we would have gotten had securities initially been ranked on equity capitalization rather than dollar volume. Stocks in the highest-volume quartile (quartile 1) tend to have less return volatility than stocks in quartiles 2–4, and greater share volume—measured as the absolute value of the order flow variable—both at the open and during the day. In fact, average absolute order flow during the day decreases by a factor of approximately 2.5 between quartiles 1 and 2, and 3 and 4, and by a factor of two between quartiles 2 and 3. Average absolute order flow at the opening is approximately 8% of its daytime level for stocks in quartiles 2–4, and 14% for stocks in quartile 1.

---

[10] We replicated our tests after excluding October 1987. Those results led to the same conclusions as the results reported in the tables below.

Table 1
Descriptive statistics

	Quartile 1 (most active)		Quartile 2		Quartile 3		Quartile 4 (least active)	
	Mean	Std. dev.	Mean	Std. dev.	Mean	Std. dev.	Mean	Std. dev.
Equity capitalization (in million $)	5,206	4,234	1,222	655	357	384	129	95
Average daytime return	0.0001	0.0004	0.0002	0.0007	0.0002	0.0013	0.0002	0.0008
Std. dev. of daytime return	0.0085	0.0090	0.0099	0.0108	0.0130	0.0182	0.0120	0.0142
Average overnight return	0.0000	0.0005	−0.0003	0.0004	−0.0003	0.0007	−0.0004	0.0006
Std. dev. of overnight return	0.0078	0.0102	0.0081	0.0099	0.0093	0.0140	0.0089	0.0105
Average daytime order flow	74.463	133.877	6.536	44.923	−5.039	22.027	−4.631	14.961
Std. dev. of daytime order flow	1056.255	1483.726	628.073	861.903	293.407	471.734	135.508	335.550
Average opening order flow	10.472	18.515	1.610	2.662	0.469	1.516	0.036	1.194
Std. dev. of opening order flow	133.438	159.535	44.444	51.734	21.175	29.631	12.610	23.429
Average absolute value of daytime order flow	395.866	1427.980	169.151	850.116	86.006	461.092	36.784	333.406
Average absolute value of opening order flow	57.101	144.164	14.809	49.252	5.871	28.680	2.972	23.116

The statistic named in the first column is computed using data from 1986–1989 for each security in the sample. The cross-sectional means and standard deviations are reported by quartile category. Each quartile category contains 25 of the top 50 stocks in each quartile of dollar trading volume for NYSE stocks during the 1986–1989 period. Equity capitalization is computed as of December 31, 1987. Order flow variables are reported in 100-share units.

3.2 Specification

The results reported below are based on a four-variate autoregressive model involving daytime and overnight returns, and daytime and opening order flow. Enough parameters are included to account for dependence of up to two lags on each of the four variables, which accounts for 48 hours in clock time. This specification is relatively parsimonious, yet its lag structure should be sufficient to capture the effects of microstructure frictions whose impact on transaction prices is more complicated than the geometric decay implied by first-order autoregression. The same specification is estimated for all stocks in the sample to avoid data snooping, and facilitate cross-sectional comparisons.[11]

To simplify the description of model specification, it helps to reorder the vector $\mathbf{x}_t$ so that daytime variables appear before overnight variables in the list: $\hat{\mathbf{x}}_t \equiv (r_{dt}, v_{dt}, r_{nt}, v_{nt})'$. The model we estimate is equivalent to

$$\widehat{A}\hat{\mathbf{x}}_t = \widehat{B}_1\hat{\mathbf{x}}_{t-1} + \widehat{B}_2\hat{\mathbf{x}}_{t-2} + \hat{\mathbf{u}}_t, \tag{10}$$

where

$$\widehat{A} = \begin{bmatrix} 1 & a_{13} & a_{12} & a_{14} \\ 0 & 1 & a_{32} & a_{34} \\ 0 & 0 & 1 & a_{24} \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad \widehat{B}_1 = \begin{bmatrix} b_{11} & b_{13} & b_{12} & b_{14} \\ b_{31} & b_{33} & b_{32} & b_{34} \\ b_{21} & b_{23} & b_{22} & b_{24} \\ b_{41} & b_{43} & b_{42} & b_{44} \end{bmatrix}$$

$$\widehat{B}_2 = \begin{bmatrix} c_{11} & c_{13} & 0 & 0 \\ c_{31} & c_{33} & 0 & 0 \\ c_{21} & c_{23} & c_{22} & c_{24} \\ c_{41} & c_{43} & c_{42} & c_{44} \end{bmatrix}$$

and

$$\widehat{\Omega} \equiv E[\hat{\mathbf{u}}_t \hat{\mathbf{u}}_t'] = \begin{bmatrix} \omega_{11} & \omega_{13} & 0 & 0 \\ \omega_{13} & \omega_{33} & 0 & 0 \\ 0 & 0 & \omega_{22} & \omega_{24} \\ 0 & 0 & \omega_{24} & \omega_{44} \end{bmatrix}.$$

The a_{13} and a_{24} parameters incorporate contemporaneous order flow into the equations for daytime and overnight returns, respectively. Note that variables

---

[11] Even if a simpler specification is the true model for some securities, inferences will not be biased because the distributions of test statistics account for the precision with which the parameters are estimated. Parameter estimates that are insignificantly different from zero receive little weight in testing the significance of variance components.

with the same time subscript are ordered in clock time—nt precedes dt. Therefore overnight returns and opening order flow precede daytime returns and order flow. The upper-right (2×2) submatrix of $\widehat{A}$ captures the dependence of daytime variables on variables of the previous overnight. Similarly the upper-right (2×2) submatrix of $\widehat{B}_1$ captures the dependence of daytime variables on returns and order flow from two nights prior. These effects account for two lags of dependence in daytime variables on overnight variables, so the upper-right (2×2) submatrix of $\widehat{B}_2$ contains zeros. The off-diagonal elements of $\widehat{\Omega}$ capture the covariance between disturbances to the equations for contemporaneous returns and order flow. Recall from our discussion of Equation (7) that these covariances can be nonzero even though contemporaneous order flow is included in the returns equations. The zeros in the off-diagonal blocks reflect the assumption that the lag structure implicit in Equation (10) (i.e., two lags of each variable, 48 hours in clock time) is sufficient to capture the cross-serial dependence in the data. This specification involves 40 parameters to be estimated for each security.[12]

Following Gallant, Rossi, and Tauchen (1992), mean return and order flow effects associated with day of the week and turn of the year are accounted for by estimating regressions of each variable on an intercept and indicators for each of Tuesday through Friday, and an indicator for the first 14 days of January. The residuals from these regressions (estimated separately for each security) are then used to estimate the vector autoregressions. GMM estimates of Equation (8) are obtained separately for each of the 100 securities in the sample. Estimation for each security involves 42 orthogonality conditions—32 "least-squares normal equations," and 10 equations for the distinct elements of Ω.[13] Estimates of the variances of $\epsilon_{ot}, \epsilon_{ct}, s_{ot}$, and s_{ct} are obtained by substituting estimates of the autoregressive parameters into the formulas in Theorem 1. Hypothesis tests are based on asymptotic normality of functions of GMM estimators [e.g., see Hamilton (1994, p. 414)].

In principle, parameter estimates should be obtained jointly for all securities in a manner that accounts for correlation of $\mathbf{u}_t$ vectors *across* securities; but the number of parameters involved in such an exercise exceeds what

---

[12] The subscripts on the individual elements of these matrices correspond to their row-column positions in the "nonhat" matrices in Equation (8). It is important to note that the formulas in Theorem 1 are based on the arrangement in Equation (8), and are not valid for the arrangement in Equation (10). Estimates from Equation (10) must be rearranged before using the formulas in Theorem 1 to calculate the variances of return components.

[13] Specifically, the "least-squares normal equations" are of the form

$$E[u_i z_{ik}] = 0,$$

where u_i is the disturbance of equation i in Equation (8) and z_{ik} is predetermined variable k of equation i in Equation (8). All right-hand-side variables are predetermined, with the exception of daytime order flow in the daytime returns equation and opening order flow in the overnight returns equation. The equations for elements of Ω are of the form

$$E[u_i u_j - \omega_{ij}] = 0,$$

where ω_{ij} is set to zero for those elements of Ω that are zero in Equation (8).

can feasibly be estimated. Consequently, hypothesis testing is done at the individual security level, and we report rejection rates in the tables below. That these tests are not necessarily independent should be kept in mind when drawing conclusions.

4. Empirical Results and Interpretations

4.1 Information flow

Table 2 reports the analysis of rates of information flow. The first row of the table reports cross-sectional medians of the difference between the day-to-night ratio of information flow and 0.37, by quartile—$6.5/17.5 = 0.37$ is the ratio of trading to nontrading hours in each 24-hour period. The second row reports the proportion of securities in the quartile category (containing 25 securities total) for which the null hypothesis that the variance ratio equals 0.37 is rejected. The third row reports the cross-sectional average of the z-ratios from the hypothesis tests on individual securities. The results indicate that the null is rejected by a majority of securities in the sample, and by a majority in each quartile. The interpretation of this is that information flow per unit time is greater during business than nonbusiness hours. To understand the economic magnitude of the point estimates, note that $(17.5/6.5) \times \text{var}[\epsilon_{ct}]/\text{var}[\epsilon_{ot}]$ is the proportionality factor by which the rate of daytime information flow exceeds that of overnight information flow. The smallest such median value is $(17.5/6.5) \times (1.8568 + 0.37) = 6.00$ for quartile 4; the values for quartiles 1–3 range between 6.55 and 7.96.[14] These numbers are closer to Jones, Kaul and Lipson's (1994) estimates of 3 to 5 for NASDAQ stocks than to French and Roll's (1986) and Ronen's (1997) estimates of 12 to 13 for NYSE/AMEX stocks.

It is possible that information flow decreases overnight when trading ceases because order flow conveys information about security value (e.g., traders' private information). The flow of non-order-flow news (e.g., public news stories) that is relevant to the values of these securities could also slow substantially when trading in the U.S. ceases. To examine the relative importance of trading, Jones, Kaul, and Lipson and French and Roll compare return volatility during 24-hour periods containing business days on which there is no trading to such periods when trading occurs. The approach in this article allows for a more direct comparison by estimating the proportion of the variance of the permanent component of price changes that is attributable to order flow. These estimates, in Section 4.3 below, indicate that the information revealed through orders is not the dominant effect. For the absence of order flow alone to imply proportionality factors in the 6–8 range, trading

---

[14] To the extent that the first transaction of the day occurs late for stocks in quartile 4, a portion of morning information flow is attributed to the overnight period for these stocks.

Table 2
Information flow during daytime and overnight periods $R = var[\epsilon_{et}]/var[\epsilon_{nt}] - 6.5/17.5$

	All quartiles	Quartile 1 (most active)	Quartile 2	Quartile 3	Quartile 4 (least active)
Cross-sectional median **R**	2.1401	2.0640	2.5848	2.2775	1.8568
Rejection rates for individual tests of H_0: **R** $= 0$	57%/2%	60%/4%	52%/0%	56%/0%	60%/4%
Cross-sectional average z-statistic	2.08	1.61	1.91	2.25	2.55
Cross-sectional average **R**	2.9311	2.3107	2.6883	4.2711	2.4543
Standard error of cross-sectional average	0.4401	0.4409	0.5662	1.5699	0.4581
Average number of time-series observations	895	912	906	879	882
Median number of time-series observations	911	913	911	911	905

The variables ε_{et} and ε_{nt} denote the permanent components of returns during daytime and overnight periods ending on day t, respectively. Parameter estimates from a vector autoregressive model of daytime and overnight returns and daytime and opening order flow, with two lags, are obtained for each security in the sample using GMM. The variances var$[\varepsilon_{et}]$ and var$[\varepsilon_{nt}]$ are computed using the autoregressive parameter estimates and the formulas in Theorem 1, which are then used to construct the **R** statistic. The asymptotic distribution of **R** is normal, and its variance is computed using the estimated asymptotic covariance matrix of GMM estimators and numerical partial derivatives of **R** with respect to the autoregressive parameters. Rejection rates reported in the table are the proportion of significance tests on individual securities within the quartile category that reject the null hypothesis in favor of a positive/negative alternative at the 5% level. The cross-sectional average of the individual test statistics is reported in the third row of the table. The standard error of the cross-sectional average is computed using the distribution of **R** statistics across securities within the quartile category, assuming independence across securities. The estimation period is 1986–1989, and the average and median number of daily observations used to obtain these estimates appear in the last two rows. Each quartile category contains 25 of the top 50 stocks in each quartile of dollar trading volume for NYSE stocks during the 1986–1989 period.

would have to explain between 83% and 88% (i.e., 5/6 to 7/8) of the variability of permanent changes in prices during the day. Estimates in Section 4.3 are almost all less than 28%. This suggests that the overnight reduction in information flow is primarily related to sources of information other than trading activity such as news stories, disclosures by the firms themselves, and prices and trading activity in related securities markets.

4.2 Pricing errors

Table 3 reports the results concerning pricing errors at the open and close. The first row contains cross-sectional medians of the difference between the ratio of variances of pricing errors at the open and close and unity. The second row reports rejection rates for tests of the hypothesis that the variance ratios are equal to unity—that is, that the variances of pricing errors at the open and close are equal—and the third row contains the cross-sectional average z-ratios from these hypothesis tests. Overall, only 13% of the stocks in the sample reject the hypothesis that the variance ratio is unity; most of these rejections (11 of 13) indicate that the *variance of pricing errors at the close is larger than at the open*. These estimates and inferences are very different from those in studies that use ratios of return variances as in Equation (4) [e.g., Amihud and Mendelson (1987, 1991), and Stoll and Whaley (1990), Forster and George (1996)]. In those studies, average and median ratios are greater than one, and large relative to the dispersion of the cross-sectional distribution of ratios. The inference typically drawn in those studies is that pricing errors are larger at the open than at the close.[15]

The difficulty with drawing inferences about variance ratios from the cross-sectional distribution of estimates is apparent in the last two rows in the table. For the sample overall, the cross-sectional average variance ratio indicates that pricing errors at the open are 32.82% larger than those at the close, an estimate that is significant in view of the dispersion of the cross section of estimates. This occurs despite the fact that 87% of the stocks in the sample do not reject the null; and of those that do reject, most of the rejections are because the variance of pricing errors at the close is *greater* than at the open. This indicates that the biases identified by Ronen (1997) for traditional variance ratio computations are also present for the ratios of return component variances estimated here. This highlights the importance of conducting inferences in a manner that avoids computing standard errors from the cross-sectional distribution of estimates.

There are some a priori reasons to believe that pricing errors at the open could be smaller than at the close. First, if liquidity providers are risk averse,

---

[15] An exception is George and Hwang (1995), who analyze returns of stocks traded on the Tokyo Stock Exchange using Ronen's (1997) approach to joint estimation of Equation (4). They document return variance ratios that are not different from, or less than, unity for stocks in the three less-active quartiles of Tokyo trading. Only Tokyo stocks in the most-active quartile have return variance ratios exceeding one.

Table 3
Pricing errors at the open and close $R = \text{var}[s_{ot}]/\text{var}[s_{ct}] - 1$

	All quartiles	Quartile 1 (most active)	Quartile 2	Quartile 3	Quartile 4 (least active)
Cross-sectional median **R**	0.0117	−0.0946	−0.1149	0.2947	−0.0085
Rejection rate for individual tests of H_0: **R** = 0	2%/11%	0%/12%	4%/16%	4%/8%	0%/8%
Cross-sectional average z-statistic	−0.35	−0.58	−0.51	0.19	−0.50
Cross-sectional average **R**	0.3282	0.1767	0.1031	0.7482	0.2847
Standard error of cross-sectional average	0.0962	0.1511	0.1262	0.2610	0.1917

The variables s_{ct} and s_{ot} denote the transitory components of returns ending at the close and open of trading on day t, respectively. Parameter estimates from a vector autoregressive model of daytime and overnight returns and daytime and opening order flow, with two lags, are obtained for each security in the sample using GMM. The variances var$[s_{ct}]$ and var$[s_{ot}]$ are computed using the autoregressive parameter estimates and the formulas in Theorem 1, which are then used to construct the **R** statistic. The asymptotic distribution of **R** is normal, and its variance is computed using the estimated asymptotic covariance matrix of GMM estimators and numerical partial derivatives of **R** with respect to the autoregressive parameters. Rejection rates reported in the table are the proportion of significance tests on individual securities within the quartile category that reject the null hypothesis in favor of a positive/negative alternative at the 5% level. The cross-sectional average of the individual test statistics is reported in the third row of the table. The standard error of the cross-sectional average is computed using the distribution of **R** statistics across securities within the quartile category, assuming independence across securities. The estimation period is 1986–1989, and the average and median number of daily observations used to obtain these estimates appear in the last two rows of Table 2. Each quartile category contains 25 of the top 50 stocks in each quartile of dollar trading volume for NYSE stocks during the 1986–1989 period.

they will demand price concessions to absorb orders that move their inventories out of balance—"static" inventory considerations. The trading day provides opportunities to rebalance inventories that do not exist overnight. Consequently, pricing errors associated with static inventory considerations should be smaller at the open than at the close. We refer to these as "static" to distinguish them from the "dynamic" efforts of liquidity providers to trade out of an inventory imbalance created by previous trades, which are discussed below. Second, the NYSE opening and closing procedures pool orders that are designated for the open or close and execute them at a single price. Madhavan's (1992) model predicts that pooling orders reduces the transitory deviations of transaction prices from true security value because traders' idiosyncratic motivations for trading diversify away. If a greater degree of pooling occurs at the open than at the close, then we would expect pricing errors to be smaller at the open than at the close. Direct evidence on the numbers of orders crossed in the opening and closing transactions for our sample period is not available. However, in each of the more recent years 1991–1999, the number of orders executed at the opening of the NYSE is approximately double the number executed at the close.[16] This suggests that a greater degree of pooling occurs at the opening than at the close.

The importance of the static inventory effect differs depending on how actively the stock is traded. Opportunities to unwind undesirable inventory positions are more abundant for active stocks than inactive stocks. Therefore the price concession charged to absorb order imbalances at the beginning of the day should be smaller for active stocks. This makes pricing errors smaller at the open for active stocks than inactive stocks, predicting that the ratio $\text{var}[s_{ot}]/\text{var}[s_{ct}]$ would be smaller for active stocks. The medians in Table 3 indicate that the variance ratios tend to be smaller for stocks in quartiles 1 and 2 than stocks in quartiles 3 and 4. This pattern across quartiles is consistent with the hypothesis that static inventory effects are important determinants of pricing errors for those stocks whose pricing errors differ between open and close.

To evaluate the importance of inventory effects for the entire sample of stocks, we examine historical dependence in pricing errors. Suppose, for example, that pricing errors occur because transaction prices bounce between the bid and ask of a constant spread that is centered on the security's true value. If purchases and sales are equally likely and serially uncorrelated, the expected transaction price is the security's true value and the expected pricing error is zero. If, instead, order sign is serially correlated, the deviation between the security's true value and future transaction prices can be forecasted. In this case, pricing errors will exhibit historical dependence. If liquidity providers alter their quotations to induce inventory-equilibrating trades—"dynamic" inventory considerations—pricing errors will depend on

---

[16] We are grateful to George Sofianos for providing these data.

the trading history. The duration of such dependence is greater, the longer it takes liquidity providers to resolve the effects of shocks to inventories. Therefore one way to assess the importance of dynamic inventory considerations on pricing errors is to check for significant historical dependence.

Historical dependence can be measured by computing the proportion of pricing error variance that is attributable to an *innovation*; the complementary proportion measures historical dependence. The difference between the realized pricing error at the close and a forecast of it made at the previous open is the innovation in the pricing error at the close: $s_{ct} - E_{ot}[s_{ct}]$. The innovation in the pricing error at the open is defined as the difference between the pricing error realized at the open and a forecast made at the previous close: $s_{ot} - E_{ct-1}[s_{ot}]$. Theorem 1 provides formulas for the variances of these innovations.

Since it is difficult to alter overnight an inventory that is unbalanced at the close, liquidity providers should be more aggressive in resolving inventory imbalances that exist prior to the close and would be carried overnight than those that exist at the opening and would be carried or unwound throughout the day. This implies that less historical dependence should exist in pricing errors at the close than at the open. Whether this difference should be more pronounced for less-active than more-active stocks depends on how much prices are expected to change overnight relative to during the day for these classes of stocks. The results in Table 2 indicate that the variance of permanent changes in prices overnight relative to the daytime is similar across quartiles (perhaps slightly larger for stocks in quartile 4). Consequently inventory control should not lead to large differences across quartiles in the extent to which historical dependence should be greater in pricing errors at the open than at the close.

Table 4 reports one minus the proportions of the pricing error variances at the open and close that are attributable to an innovation. This table shows evidence of statistically significant historical dependence at both the open and the close. Rejection rates are between 48% and 80% for tests of the hypothesis that all of the variation in pricing errors relates to their innovation. In addition, median historical dependence is greater for pricing errors at the open than at the close (63.71% versus 52.70%) for the sample as a whole. Some caution should be exercised in interpreting this difference, however. It is possible that using signed opening volume makes pricing errors at the open more dependent on unexpected overnight returns and opening order flow than they would be if we could observe true net order flow at the open. Nevertheless, these findings are consistent with the hypothesis that dynamic inventory management is a determinant of pricing errors for a majority of stocks in the sample.[17]

---

[17] Price continuity rules could also induce historical dependence in the temporary component of transaction prices. If stocks are not actively traded, closing prices may not reflect information about security value that is

Table 4
Historical dependence in pricing errors at the open and close

	All quartiles	Quartile 1 (most active)	Quartile 2	Quartile 3	Quartile 4 (least active)
Panel A: Historical dependence at the open $R = 1 - \text{var}[s_{ot} - E_{ct-1}[s_{ot}]]/\text{var}[s_{ot}]$					
Cross-sectional median **R**	0.6371	0.6948	0.7500	0.5303	0.6286
Rejection rate for individual tests of H_0: **R** = 0	67%/0%	48%/0%	72%/0%	68%/0%	80%/0%
Cross-sectional average z-statistic	3.81	4.72	4.11	3.30	3.08
Cross-sectional average **R**	0.6173	0.6813	0.6634	0.5326	0.5921
Standard error of cross-sectional average	0.0243	0.0461	0.0488	0.0497	0.0494
Panel B: Historical dependence at the close $R = 1 - \text{var}[s_{ct} - E_{ot}[s_{ct}]]/\text{var}[s_{ct}]$					
Cross-sectional median **R**	0.5270	0.6517	0.4529	0.7081	0.3871
Rejection rate for individual tests of H_0: **R** = 0	68%/0%	64%/0%	68%/0%	64%/0%	76%/0%
Cross-sectional average z-statistic	3.61	3.54	4.64	3.43	2.81
Cross-sectional average **R**	0.5659	0.6407	0.5585	0.5932	0.4711
Standard error of cross-sectional average	0.0275	0.0526	0.0532	0.0624	0.0448

The variables s_{ot} and s_{ct} denote the transitory component of returns ending at the open and close of trading on day t, respectively. Parameter estimates from a vector autoregressive model of daytime and overnight returns and daytime and opening order flow, with two lags, are obtained for each security in the sample using GMM. The variances var$[s_{ot} - E_{ct-1}[s_{ot}]]$, var$[s_{ot}]$, var$[s_{ct} - E_{ot}[s_{ct}]]$, and var$[s_{ct}]$ are computed using the autoregressive parameter estimates and the formulas in Theorem 1, which are then used to construct the **R** statistics. The asymptotic distribution of **R** is normal, and its variance is computed using the estimated asymptotic covariance matrix of GMM estimators and numerical partial derivatives of **R** with respect to the autoregressive parameters. Rejection rates reported in the table are the proportion of significance tests on individual securities within the quartile category that reject the null hypothesis in favor of a positive/negative alternative at the 5% level. The cross-sectional average of the individual test statistics is reported in the third row of the table. The standard error of the cross-sectional average is computed using the distribution of **R** statistics across securities within the quartile category, assuming independence across securities. The estimation period is 1986–1989, and the average and median number of daily observations used to obtain these estimates appear in the last two rows of Table 2. Each quartile contains 25 of the top 50 stocks in each quartile of dollar trading volume for NYSE stocks during the 1986–1989 period.

4.3 Association between trading activity and pricing errors and information flow

This section examines the extent to which variation in permanent and transitory components of returns are associated with innovations in order flow. These associations shed light on some of the hypotheses discussed earlier as possible explanations for the behavior of pricing errors and rates of information flow.

Innovations in order flow are denoted by u_{3t} and u_{4t} in Equations (8)–(10). Our objective is to estimate the portion of the variances of $\epsilon_{ot}, \epsilon_{ct}, s_{ot}$, and s_{ct} in Theorem 1 that is attributable to these innovations. If the disturbances to the returns and order flow equations were uncorrelated, we could achieve this by setting the variances of the returns equation disturbances, u_{1t} and u_{2t}, equal to zero in the equations in Theorem 1. Hasbrouck (1991b), assumes that the time-series innovations to returns and order flow are uncorrelated, and uses this approach to compute the contribution of order flow to the permanent component of intraday quote revisions. This assumption is valid if news affects order flow only indirectly through quote revisions. If news affects order flow directly, as in Equation (1), the disturbances will be correlated.

With correlated disturbances, we can use this idea to approximate the contribution of order flow. To do this we set to zero the variances of the returns equation disturbances and the covariances between them and the order flow equation disturbances. Although this is only an approximation, we can show in the context of Equation (1) that this approximation provides an upper bound on the portion of variance in the permanent and temporary components attributable to innovations in order flow.

To see how this approach produces an upper bound in the context of Equation (1), note that the VAR implied by Equation (1) is given in Equation (6). The disturbance to the returns equation in Equation (6), which we will call $\hat{\eta}$, is equal to a constant times the structural news shock: $\hat{\eta}_t = (1 - \gamma\kappa_3)\eta_t$. The disturbance to the order flow equation, $\hat{u}_t$ is a composite of the structural news and order flow shocks: $\hat{u}_t = \gamma\eta_t + u_t$. Furthermore, the permanent component of the price change is $\epsilon_t = \kappa_3(v_t - E_{t-1}[v_t|\eta_t]) + \eta_t = \kappa_3 u_t + \eta_t$, and the temporary component is $s_t = \kappa_1 v_t + \kappa_2 v_{t-1}$.

---

contained in order flow because of the specialist's smoothing of the price path. For this to affect our estimates, however, the smoothing would have to be slow enough that information in order flow at the opening or earlier is not reflected in the closing price. This seems extreme except, perhaps, for the most inactively traded stocks. Price continuity rules could induce historical dependence in opening prices. Although the specialist is not required to provide enough liquidity so that opening prices are within a certain dollar amount of the previous close, he is generally required to provide liquidity on the side of the market where it is lacking at the opening in case of a large imbalance. This would have the effect of smoothing the price path from the prior close to the opening. It is difficult to know whether this has an important effect on our estimates. If the imbalance at the open is a *surprise*, smoothing over it by providing liquidity will not lead the opening pricing error to be *historically* dependent. However, if the imbalance is forecastable from the prior day's price change and/or order flow, then the specialist's smoothing will result in a pricing error that is historically dependent.

Writing the variance of the permanent component in terms of the VAR disturbances:

$$\epsilon_t = \kappa_3 u_t + \eta_t$$

$$= \kappa_3 \hat{u}_t + \hat{\eta}_t$$

$$\mathrm{var}\,[\epsilon_t] = \kappa_3^2 \mathrm{var}\,[\hat{u}_t] + \mathrm{var}\,[\hat{\eta}_t] + 2\kappa_3\,\mathrm{cov}\,[\hat{\eta}_t, \hat{u}_t]$$

$$= \kappa_3^2\{\sigma_u^2 + \gamma^2\sigma_\eta^2\} + \mathrm{var}\,[\hat{\eta}_t] + +2\kappa_3\mathrm{cov}\,[\hat{\eta}_t, \hat{u}_t].$$

Note that $\mathrm{var}\,[\hat{\eta}_t] = (1 - \gamma\kappa_3)^2\sigma_\eta^2$ and $2\kappa_3\mathrm{cov}\,[\hat{\eta}_t, \hat{u}_t] = 2\kappa_3\gamma(1 - \gamma\kappa_3)\sigma_\eta^2$. Setting these terms to zero yields

$$\mathrm{var}\,[\epsilon_t]|_{\mathrm{var}\,[\hat{\eta}]=0,\mathrm{cov}\,[\hat{\eta},\hat{u}]=0} = \kappa_3^2\{\sigma_u^2 + \gamma^2\sigma_\eta^2\} > \kappa_3^2\sigma_u^2 = \mathrm{var}\,[\epsilon_t]|_{\mathrm{var}\,[\eta]=0}.$$

The left-hand side is the variance of the permanent component computed by setting to zero the variance of the VAR disturbance to the returns equation, and its covariance with the disturbance to the order flow equation. The right-hand side is the variance of the permanent component attributable to the *structural* shock to order flow (i.e., what is left after setting to zero the structural news shock). The inequality indicates that by setting the VAR disturbance variance and covariance to zero, we obtain an upper bound estimate of the true portion of the variance of ϵ attributable to order flow innovations. Note that this is the tightest upper bound we can get with this approach because the sign of the covariance term, $2\kappa_3\gamma(1 - \gamma\kappa_3)\sigma_\eta^2$, can either be positive or negative and larger in absolute value than $\kappa_3^2\gamma^2\sigma_\eta^2$.

Similar logic applies to the temporary component of price changes. In the context of Equation (1), the variance of s_t is

$$\mathrm{var}\,[s_t] = \kappa_1^2\,\mathrm{var}\,[v_t] + \kappa_2^2\,\mathrm{var}\,[v_{t-1}] + 2\kappa_1\kappa_2\,\mathrm{cov}\,[v_t, v_{t-1}],$$

where

$$\mathrm{var}\,[v_t] = \frac{\gamma^2\sigma_\eta^2 + \sigma_u^2}{1 - \rho^2} = \mathrm{var}\,[\hat{u}_t]/(1 - \rho^2)$$

$$\mathrm{cov}\,[v_t, v_{t-1}] = \rho\,\mathrm{var}\,[v_t] = \rho\,\mathrm{var}\,[\hat{u}_t]/(1 - \rho^2).$$

Expressing $\mathrm{var}\,[s_t]$ in terms of the variances and covariances of the VAR disturbances,

$$\mathrm{var}\,[s_t] = [\kappa_1^2/(1 - \rho^2)]\,\mathrm{var}\,[\hat{u}_t] + [\kappa_2^2/(1 - \rho^2)]\,\mathrm{var}\,[\hat{u}_t]$$

$$+ 2\kappa_1\kappa_2\rho\,\mathrm{var}\,[\hat{u}_t]/(1 - \rho^2)$$

$$= \left(\frac{\kappa_1^2 + \kappa_2^2 + 2\kappa_1\kappa_2\rho}{1 - \rho^2}\right)\mathrm{var}\,[\hat{u}_t].$$

Since neither $\mathrm{var}\,[\hat{\eta}_t]$ nor $\mathrm{cov}\,[\hat{\eta}_t, \hat{u}_t]$ enters the expression, setting them equal to zero gives us no tighter a bound on $\mathrm{var}\,[s_t]|_{\sigma_\eta^2=0}$ than $\mathrm{var}\,[s_t]$ itself.

Table 5
Percentage of variance of permanent component of returns attributable to variation in order flow

	All quartiles	Quartile 1	Quartile 2	Quartile 3	Quartile 4
Daytime information flow attributable to daytime order flow:					
90th percentile	28%	35%	26%	19%	24%
75th percentile	19%	28%	15%	13%	19%
Median	9%	18%	7%	7%	11%
25th percentile	5%	9%	1%	1%	6%
10th percentile	1%	6%	0%	0%	1%
Overnight information flow attributable to opening order flow:					
90th percentile	22%	16%	11%	44%	26%
75th percentile	8%	6%	4%	14%	10%
Median	2%	1%	3%	2%	2%
25th percentile	0%	0%	0%	0%	0%
10th percentile	0%	0%	0%	0%	0%

The variance of information flow attributable to order flow is computed by substituting the autoregressive parameter estimates into the formulas for $\text{var}[\varepsilon_{jt}]$ ($j \varepsilon \{o, c\}$) in Theorem 1, and setting the $(1, 1)$, $(2, 2)$, $(1, 3)$ and $(2, 4)$ elements of Ω to zero. The numbers reported in the tables are percentiles of the cross-sectional distribution expressed as a percentage of the total variance of information flow, $\text{var}[\varepsilon_{jt}]$. The estimation period is 1986–1989, and the average and median number of daily observations used to obtain these estimates appear in the last two rows of Table 2. Each quartile category contains 25 of the top 50 stocks in each quartile of dollar trading volume for NYSE stocks during the 1986–1989 period.

However, one could imagine that market frictions that are not modeled in Equation (1) could cause s_t to depend on lagged news shocks. In that case, the expression above would have $\text{var}[\hat{\eta}_t]$ and $\text{cov}[\hat{\eta}_t, \hat{u}_t]$ terms also, and the bound produced by setting them equal to zero would be tighter than the entirety of $\text{var}[s_t]$.

Table 5 reports estimates of the proportion of the permanent component of daytime and overnight returns attributable to unexpected order flow. For the entire sample, the 90th percentile of the distribution is 28% for daytime returns and 22% for overnight returns. The interpretation of these results is that, for most of the sample, less than 28% (22%) of daytime (overnight) information flow is attributable to variation in unexpected order flow. That the daytime proportions are less than 28% for most of the sample is the basis for our earlier conclusion that the reduction in information flow that occurs overnight is not simply a result of the cessation of trading; but that the flow of information from other sources must also decrease during overnight periods.[18]

At first glance it may seem surprising that permanent changes in prices are not more strongly related to trading activity, since private information comes to be reflected in prices through trading. However, Roll's (1984) study of orange juice futures finds that weather explains only about 7% of orange juice futures returns, and that weather and announcements about weather explain

---

[18] That the numbers are larger at the opening could reflect some degree of spurious correlation induced between the overnight return and opening order flow by the fact that opening volume is signed using the overnight return. The first sensitivity test in the next section addresses this possibility further.

Table 6
Percentage of variance of transitory component of returns attributable to variation in order flow

	All quartiles	Quartile 1	Quartile 2	Quartile 3	Quartile 4
Pricing error at close attributable to order flow:					
90th percentile	84%	92%	80%	71%	84%
75th percentile	61%	66%	61%	53%	55%
Median	34%	32%	19%	41%	41%
25th percentile	12%	25%	7%	8%	11%
10th percentile	5%	19%	5%	4%	7%
Pricing error at open attributable to order flow:					
90th percentile	84%	82%	81%	67%	89%
75th percentile	57%	55%	68%	51%	57%
Median	38%	43%	34%	33%	42%
25th percentile	20%	21%	12%	15%	26%
10th percentile	9%	11%	6%	9%	19%

The variance of the transitory component of returns attributable to order flow is computed by substituting the autoregressive parameter estimates into the formulas for $\text{var}[s_{jt}]$ ($j \, \varepsilon \, \{o, c\}$) in Theorem 1, and setting the $(1, 1)$, $(2, 2)$, $(1, 3)$, and $(2, 4)$ elements of Ω to zero. The numbers reported in the tables are percentiles of the cross-sectional distribution expressed as a percentage of the total variance of the transitory component of returns, $\text{var}[s_{jt}]$. The estimation period is 1986–1989, and the average and median number of daily observations used to obtain these estimates appear in the last two rows of Table 2. Each quartile category contains 25 of the top 50 stocks in each quartile of dollar trading volume for NYSE stocks during the 1986–1989 period.

only about 27% of the variability of squared orange juice futures returns. Though his study does not distinguish between permanent and temporary components of returns, it does illustrate how small the contribution to price variability can be from a variable that would intuitively seem to have a great deal of information content.

Table 6 reports the proportions of the variances of pricing errors that are associated with unexpected order flow. These estimates can be interpreted as the extent to which deviations of transaction prices from true value are associated with the unexpected arrival of orders, both currently and in the past. This can be viewed as a measure of the importance of both static and dynamic inventory considerations on pricing errors. For individual firms, the numbers range from 4% to more than 90%, so right tails of the cross-sectional distributions do not provide useful bounds as they did above. Nevertheless, the medians of these distributions indicate that between 34% and 38% of the variation in pricing errors is attributable to order flow for the typical security in our sample. For example, if both the closing and opening prices of the typical security are the result of aggregating orders and crossing them at a single price, these estimates indicate that between 34% and 38% of the variability of the deviation of these prices from the true security value is attributable to current and past surprises in order flow.

5. Sensitivity Tests

This section explores whether the pricing error results in Table 3 that conflict with findings in the existing literature are artifacts of the manner in

which the order flow variables are constructed. Daytime order flow is the sum of signed individual orders, whereas opening order flow is constructed by signing opening volume using the overnight price change. This raises two potential problems. First, some overnight price changes will be related to the disclosure of public information and unrelated to trading at the open. In these instances, signing opening order flow using the overnight price change misallocates to order flow variance that is due to price changes. This will affect the ability of the time-series model to explain price changes. Though we do not know the nature of the effect, it is possible that it could bias estimates in a manner that makes the variance of pricing errors at the close appear larger than those at the open. Specifically, if misallocation of variance lessens the model's ability to forecast daytime price changes *relative* to its ability to forecast overnight price changes, then estimates of pricing error variances will be biased in a way that makes pricing errors at the open small in relation to pricing errors at the close. This is because the pricing error at the open (close) is the component of the opening (closing) price that is forecasted to reverse itself during the day (night) and thereafter.

We cannot observe the degree to which variance might be misallocated to order flow. However, a symptom of this misallocation is that it induces spurious positive correlation between the time series of overnight returns and opening order flow. Suppose that the *true*, but unobservable, correlation between opening order flow and the overnight return is similar across stocks within quartiles. In this case, cross-sectional variation in *estimated* correlations will reflect differing degrees of spurious correlation induced by our sampling procedure. Under this assumption, securities with large estimated correlations will be those with a greater misallocation of variance.

Our first diagnostic examines this possibility. We check whether the ranks of the variance ratios and test statistics in Table 3 are cross-sectionally correlated with the ranks of the estimated correlations between the time series of overnight returns and opening order flow. This will tell us whether the rejections in Table 3 could be associated with misallocated variance. For three of the four quartiles, the rank correlations are not significant at the 10% level. For quartile 4, the rank correlation is significant, but *positive*. This evidence is consistent with the absence of significant misallocation of variance. However, it does not *prove* that such misallocation does not exist. If the assumption made above does not hold, it is possible that variance is misallocated, but the degree of misallocation does not generate cross-sectional differences in correlations between opening order flow and the overnight return because it is overwhelmed by cross-sectional differences in the true correlations.

The second potential problem is that when orders are pooled at the open, opening order flow is a measure of signed volume because the extent to which orders offset each other is not observable. Although both daytime and opening order flow variables are proxies, opening order flow is likely to be a poorer proxy for stocks that frequently open with the pooling of orders.

Having proxies with *unequal* error precisions might bias the comparisons between pricing error variances at the open and close. In particular, precise measurement of daytime order flow may make a greater portion of returns subsequent to the close forecastable conditional on daytime order flow relative to the forecastability of returns subsequent to the open conditional on (a noisier proxy for) opening order flow. If this were true, the variance of pricing errors at the close could erroneously appear to be greater than at the open.

A check for this requires an avoidance of conditioning on order flow variables that are observed with asymmetric levels of precision. Since the true variables are not observable, we cannot estimate the relative precisions of the proxies. However, we can eliminate the asymmetry by reconstructing Table 3 using estimates from a vector autoregression that excludes order flow variables. This also eliminates whatever conditioning information order flow contains for future price changes that is orthogonal to current and past price changes. Nevertheless, the manner in which these results differ from those in Table 3 will provide some indication of how the asymmetry might have affected our conclusions.

Table 7 reports the same statistics as Table 3, computed from estimates of a bivariate vector autoregression of daytime and overnight returns only. There are some striking differences from Table 3. The median ratio for quartile 1 changes sign and is now positive. This is exactly what would be expected if noisy measurement of opening order flow impedes the model's ability to forecast price changes after the opening, because the stocks in quartile 1 are those for which pooling of orders is likely to be most frequent. Despite this, the *overall* sample median ratio is more negative than in Table 3. Moreover, *all* rejections of the null that the pricing error variances are the same at the open and close favor the alternative that pricing error variance is *larger at the close*, even for quartile 1. This suggests that conditioning on order flow variables that are measured with asymmetric precision might have induced noise, but not bias, into the inferences drawn in Table 3. The inferences without order flow indicate uniformly that when pricing error variances differ between the open and close, the variance of the error at the close is greater than at the open.

Beyond their diagnostic value, the Table 7 estimates provide an interesting comparison to extant studies because they condition on the same sample information—opening and closing prices—as the earlier literature's return variance ratio comparisons of the open and close. Table 7 indicates that the findings in Table 3 are not a result of having conditioned on order flow. Instead, it is the combination of estimating variances of return components in a manner that accounts for their correlation and conducting hypothesis tests at the individual security level that overturn the return variance ratio results. These findings show that estimates of pricing error variances at the open and close using our approach are typically not different (or are greater

Table 7
Pricing errors at the open and close $R = \mathrm{var}[s_{ot}]/\mathrm{var}[s_{ct}] - 1$

	All quartiles	Quartile 1 (most active)	Quartile 2	Quartile 3	Quartile 4 (least active)
Cross-sectional median **R**	−0.0092	0.0627	−0.1950	0.5537	−0.2186
Rejection rate for individual t-tests of H_0: **R** $= 0$	0%/15%	0%/20%	0%/16%	0%/8%	0%/16%
Cross-sectional average z-statistic	−0.51	−0.58	−0.88	0.02	−0.62
Cross-sectional average **R**	0.5791	0.3265	0.2085	1.2702	0.5114
Standard error of cross-sectional average	0.1636	0.2012	0.2090	0.4745	0.3410

Estimates based on bi-variate VAR of daytime and overnight returns. The variables s_{ct} and s_{ot} denote the transitory components of returns ending at the close and open of trading on day t, respectively. Parameter estimates from a vector autoregressive model of daytime and overnight returns, with two lags, are obtained for each security in the sample using GMM. The variances $\mathrm{var}[s_{ct}]$ and $\mathrm{var}[s_{ot}]$ are computed using the autoregressive parameter estimates and the formulas in Theorem 1, which are then used to construct the **R** statistic. The asymptotic distribution of **R** is normal, and its variance is computed using the estimated asymptotic covariance matrix of GMM estimators and numerical partial derivatives of **R** with respect to the autoregressive parameters. Rejection rates reported in the table are the proportion of significance tests on individual securities within the quartile category that reject the null hypothesis in favor of a positive/negative alternative at the 5% level. The cross-sectional average of the individual test statistics is reported in the third row of the table. The standard error of the cross-sectional average is computed using the distribution of **R** statistics across securities within the quartile category, assuming independence across securities. The estimation period is 1986–1989, and the average and median number of daily observations used to obtain these estimates appear in the last two rows of Table 2. Each quartile contains 25 of the top 50 stocks in each quartile of dollar trading volume for NYSE stocks during the 1986–1989 period.

at the close) even when conditioned on the same sample information as used in the earlier studies.

6. Conclusion

This article provides methods that improve on return variance ratios to test hypotheses concerning differential rates of information flow and magnitudes of pricing errors (transitory components of price changes). The methods are applied to a comparison of rates of information flow during daytime and overnight periods, and the magnitudes of pricing errors at the open and close of trading.

Estimates of relative rates of information flow indicate that information flow per hour decreases significantly overnight. Median rates of information flow overnight are approximately one-seventh the daytime rate per hour—decreases that are smaller than reported in the existing literature on listed stocks. Estimates of the proportion of variation in information flow that is associated with order flow indicate that this decrease is too large to be explained by the overnight cessation of trading. This evidence suggests that disclosures made by the media during nonbusiness hours, and prices of other securities that trade offshore, are not informative enough for value discovery to continue unimpeded overnight for a majority of the stocks in our sample.[19]

Our estimates of the magnitudes of pricing errors indicate that pricing errors at the open and close are not significantly different for most stocks; but when differences do exist, pricing errors at the close tend to be *greater* than at the open. This is very different from what others have concluded based on estimates of NYSE return variance ratios. The results also reinforce Ronen's (1997) conclusion that caution should be exercised in interpreting tests based on cross-sectional distributions of variance ratios. Those tests ignore correlation between permanent and transitory components of returns, and do not fully account for the effects of estimation error in drawing inferences.

Though these findings challenge views that overnight trading suspensions and opening auction mechanisms are important sources of microstructure-induced pricing errors, our estimates also suggest that inventory effects play a larger role at the close, perhaps as a consequence of the overnight trading suspension. However, for most stocks, their role is not large enough to induce pricing errors at the close whose average magnitude is larger than that of the pricing errors at the open.

Appendix A: Preliminary Results and Proof of Theorem 1

Since we have not, as yet, assumed that returns have an ARMA representation, we need a slightly stronger assumption than covariance stationarity of s_t to prove Facts 1 and 2. In

---

[19] Our sample does not contain stocks that are listed in London or Tokyo.

particular, we assume that the $\{s_{jt}\}_t$ process has an infinite moving average representation (or equivalently that the Wold decomposition is purely indeterministic [see Hamilton (1994, p. 109)]). Under the ARMA assumption made later (Remark 1), s_t does indeed have an infinite moving average representation (Fact 3).

Fact 1. $\quad m_{jt} = \lim_{n\to\infty} E_{jt}[p_{ot+n}] = \lim_{n\to\infty} E_{jt}[p_{ct+n}] \quad for \ j \in \{o, c\}.$

Proof. First, note that for each $j \in \{o, c\}$,

$$E_{jt}[p_{ot+n}] = E_{jt}[m_{ot+n}] + E_{jt}[s_{ot+n}].$$

Since alternate m_{jt} are a random walk,

$$E_{jt}[m_{ot+n}] = m_{jt}.$$

Since s_{ot} has an infinite moving average representation, $E_{jt}[s_{ot+n}]$ and $E_{jt-n}[s_{ot}]$ are equal in distribution for all n—they are identical polynomial lags of zero-mean i.i.d. random variables. Consequently they have the same limiting distribution as $n \to \infty$:

$$\lim_{n\to\infty} E_{jt}[s_{ot+n}] = \lim_{n\to\infty} E_{jt-n}[s_{ot}] = 0.$$

The second term is the unconditional expectation of s_{ot}, which is zero. Therefore both forecasts have distributions that converge to that of a degenerate random variable that is equal to zero with probability one. This implies that

$$\lim_{n\to\infty} E_{jt}[p_{ot+n}] = m_{jt} + \lim_{n\to\infty} E_{jt}[s_{ot+n}]$$
$$= m_{jt}.$$

An identical argument using p_{ct+n} competes the proof. $\quad\quad\quad\quad\quad\quad\quad\quad\quad \|$

Fact 2 (Representation of Components in Terms of Returns). *The random-walk innovations can be written as*

$$\epsilon_{ct} = \lim_{n\to\infty}\{E_{ct}[p_{ct+n} - p_{ot}] - E_{ot}[p_{ct+n} - p_{ot}]\}$$
$$\epsilon_{ot} = \lim_{n\to\infty}\{E_{ot}[p_{ot+n} - p_{ct-1}] - E_{ct-1}[p_{ot+n} - p_{ct-1}]\}.$$

The stationary components can be written as

$$-s_{ct} = \lim_{n\to\infty} E_{ct}[p_{ot+n} - p_{ct}]$$
$$-s_{ot} = \lim_{n\to\infty} E_{ot}[p_{ct+n} - p_{ot}].$$

The surprises in the stationary components can be written as

$$-\{s_{ct} - E_{ot}[s_{ct}]\} = \lim_{n\to\infty}\{E_{ct}[p_{ot+n} - p_{ct}] - E_{ot}[p_{ot+n} - p_{ct}]\}$$
$$-\{s_{ot} - E_{ct-1}[s_{ot}]\} = \lim_{n\to\infty}\{E_{ot}[p_{ct+n} - p_{ot}] - E_{ct-1}[p_{ct+n} - p_{ot}]\}.$$

Proof. From Equation (2) in the text,

$$\epsilon_{ct} = m_{ct} - m_{ot}$$

$$= \lim_{n\to\infty} E_{ct}[p_{ct+n}] - \lim_{n\to\infty} E_{ot}[p_{ct+n}]$$

$$= \lim_{n\to\infty}\{E_{ct}[p_{ct+n} - p_{ot}] - E_{ot}[p_{ct+n} - p_{ot}]\},$$

where the second equality follows from Fact 1. A similar argument provides the formula for ϵ_{ot}. For the stationary components, use Equation (2) again,

$$-s_{ct} = m_{ct} - p_{ct}$$

$$= \lim_{n\to\infty} E_{ct}[p_{ot+n}] - p_{ct}$$

$$= \lim_{n\to\infty}\{E_{ct}[p_{ot+n}] - p_{ct}\},$$

where the second equality follows from Fact 1. A similar argument yields the formula for $-s_{ot}$. For the surprise in s_{ct}, note that

$$-\{s_{ct} - E_{ot}[s_{ct}]\} = -s_{ct} - \{m_{ot} - m_{ot} - E_{ot}[s_{ct}]\}$$

$$= -s_{ct} - \{m_{ot} - (E_{ot}[m_{ct}] + E_{ot}[s_{ct}])\}$$

$$= -s_{ct} - \{m_{ot} - E_{ot}[p_{ct}]\}.$$

We just showed that $-s_{ct} = \lim_{n\to\infty}\{E_{ct}[p_{ot+n}] - p_{ct}\}$; and from Fact 1, we have $m_{ot} = \lim_{n\to\infty} E_{ot}[p_{ct+n}]$. Substituting these into the expression just above yields

$$-\{s_{ct} - E_{ot}[s_{ct}]\} = \lim_{n\to\infty}\{E_{ct}[p_{ot+n}] - p_{ct}\} - \{\lim_{n\to\infty} E_{ot}[p_{ct+n}] - E_{ot}[p_{ct}]\}$$

$$= \lim_{n\to\infty}\{E_{ct}[p_{ot+n} - p_{ct}] - E_{ot}[p_{ct+n} - p_{ct}]\}.$$

A similar argument yields the formula for the surprise in s_{ot}. ∥

Remark 1. Assume that the vector $\mathbf{x}_t \equiv (r_{dt}, r_{nt}, v_{dt}, v_{nt})'$ is covariance stationary, where $r_{dt} = p_{ct} - p_{ot}$ and $r_{nt} = p_{ot} - p_{ct-1}$. It therefore has an infinite moving average representation that can be written as

$$x_{it} = \sum_{k=0}^{\infty} \beta_{i1}(k)u_{1t-k} + \sum_{k=0}^{\infty} \beta_{i2}(k)u_{2t-k} + \sum_{k=0}^{\infty} \beta_{i3}(k)u_{3t-k}$$

$$+ \sum_{k=0}^{\infty} \beta_{i4}(k)u_{4t-k} \quad \text{for } i = 1,\ldots,4 \tag{A.1}$$

where $\mathbf{u}_t \equiv (u_{1t},\ldots,u_{4t})'$ is an i.i.d. random vector of disturbances. The next result provides formulas for the limits in Fact 2 in terms of moving average coefficients and realizations of the disturbance vector $\mathbf{u}_t$. Note that no particular autoregressive structure for $\mathbf{x}_t$ is assumed as yet.

Fact 3 (Representation of Components in Terms of VMA Parameters and Disturbances). *Define $\beta_{ih}^*(j) \equiv \sum_{k=j}^{\infty} \beta_{ih}(j)$. The random-walk innovations can be written as*

$$\epsilon_{ct} = \{\beta_{11}^*(0) + \beta_{21}^*(1)\}u_{1t} + \{\beta_{13}^*(0) + \beta_{23}^*(1)\}u_{3t}$$

$$\epsilon_{ot} = \{\beta_{12}^*(0) + \beta_{22}^*(0)\}u_{2t} + \{\beta_{14}^*(0) + \beta_{24}^*(0)\}u_{4t}.$$

The stationary components can be written as

$$-s_{ct} = \sum_{j=1}^{\infty}\Big\{\big[\beta_{11}^*(j) + \beta_{21}^*(j)\big]u_{1t-j+1} + \big[\beta_{12}^*(j) + \beta_{22}^*(j)\big]u_{2t-j+1}$$

$$+ \big[\beta_{13}^*(j) + \beta_{23}^*(j)\big]u_{3t-j+1} + \big[\beta_{14}^*(j) + \beta_{24}^*(j)\big]u_{4t-j+1}\Big\}$$

$$-s_{ot} = \big\{\beta_{12}^*(0) + \beta_{22}^*(1)\big\}u_{2t} + \big\{\beta_{14}^*(0) + \beta_{24}^*(1)\big\}u_{4t}$$

$$+ \sum_{j=1}^{\infty}\Big\{\big[\beta_{11}^*(j) + \beta_{21}^*(j+1)\big]u_{1t-j} + \big[\beta_{12}^*(j) + \beta_{22}^*(j+1)\big]u_{2t-j}$$

$$+ \big[\beta_{13}^*(j) + \beta_{23}^*(j+1)\big]u_{3t-j} + \big[\beta_{14}^*(j) + \beta_{24}^*(j+1)\big]u_{4t-j}\Big\}.$$

The surprises in the stationary components can be written as

$$-\big\{s_{ct} - E_{ot}[s_{ct}]\big\} = \big\{\beta_{11}^*(1) + \beta_{21}^*(1)\big\}u_{1t} + \big\{\beta_{13}^*(1) + \beta_{23}^*(1)\big\}u_{3t}$$

$$-\big\{s_{ot} - E_{ct-1}[s_{ot}]\big\} = \big\{\beta_{12}^*(0) + \beta_{22}^*(1)\big\}u_{2t} + \big\{\beta_{14}^*(0) + \beta_{24}^*(1)\big\}u_{4t}.$$

Proof.   Available from authors upon request.

Lemma 1 (Component Variances in Terms of VMA Parameters).   *Define* $\sigma_{ih} \equiv \text{cov}[u_{it}, u_{ht}]$. *The variances of the random-walk innovations are*

$$\text{var}[\epsilon_{ct}] = \begin{bmatrix}\beta_{11}^*(0)+\beta_{21}^*(1)\\\beta_{13}^*(0)+\beta_{23}^*(0)\end{bmatrix}'\begin{bmatrix}\sigma_{11} & \sigma_{13}\\\sigma_{13} & \sigma_{33}\end{bmatrix}\begin{bmatrix}\beta_{11}^*(0)+\beta_{21}^*(1)\\\beta_{13}^*(0)+\beta_{23}^*(0)\end{bmatrix}$$

$$\text{var}[\epsilon_{ot}] = \begin{bmatrix}\beta_{12}^*(0)+\beta_{22}^*(0)\\\beta_{14}^*(0)+\beta_{24}^*(0)\end{bmatrix}'\begin{bmatrix}\sigma_{22} & \sigma_{24}\\\sigma_{24} & \sigma_{44}\end{bmatrix}\begin{bmatrix}\beta_{12}^*(0)+\beta_{22}^*(0)\\\beta_{14}^*(0)+\beta_{24}^*(0)\end{bmatrix}.$$

The variances of the stationary components are

$$\text{var}[s_{ct}] = \sum_{j=1}^{\infty}\left\{\begin{bmatrix}\beta_{11}^*(j)+\beta_{21}^*(j)\\\beta_{12}^*(j)+\beta_{22}^*(j)\\\beta_{13}^*(j)+\beta_{23}^*(j)\\\beta_{14}^*(j)+\beta_{24}^*(j)\end{bmatrix}'\text{cov}[\mathbf{u}_t,\mathbf{u}_t]\begin{bmatrix}\beta_{11}^*(j)+\beta_{21}^*(j)\\\beta_{12}^*(j)+\beta_{22}^*(j)\\\beta_{13}^*(j)+\beta_{23}^*(j)\\\beta_{14}^*(j)+\beta_{24}^*(j)\end{bmatrix}\right\}$$

$$\text{var}[s_{ot}] = \begin{bmatrix}\beta_{12}^*(0)+\beta_{22}^*(1)\\\beta_{14}^*(0)+\beta_{24}^*(1)\end{bmatrix}'\begin{bmatrix}\sigma_{22} & \sigma_{24}\\\sigma_{24} & \sigma_{44}\end{bmatrix}\begin{bmatrix}\beta_{12}^*(0)+\beta_{22}^*(1)\\\beta_{14}^*(0)+\beta_{24}^*(1)\end{bmatrix}$$

$$+ \sum_{j=1}^{\infty}\left\{\begin{bmatrix}\beta_{11}^*(j)+\beta_{21}^*(j+1)\\\beta_{12}^*(j)+\beta_{22}^*(j+1)\\\beta_{13}^*(j)+\beta_{23}^*(j+1)\\\beta_{14}^*(j)+\beta_{24}^*(j+1)\end{bmatrix}'\text{cov}[\mathbf{u}_t,\mathbf{u}_t]\begin{bmatrix}\beta_{11}^*(j)+\beta_{21}^*(j+1)\\\beta_{12}^*(j)+\beta_{22}^*(j+1)\\\beta_{13}^*(j)+\beta_{23}^*(j+1)\\\beta_{14}^*(j)+\beta_{24}^*(j+1)\end{bmatrix}\right\}.$$

The variances of the surprises in the stationary components are

$$\text{var}[s_{ct} - E_{ot}[s_{ct}]] = \begin{bmatrix}\beta_{11}^*(1)+\beta_{21}^*(1)\\\beta_{13}^*(1)+\beta_{23}^*(1)\end{bmatrix}'\begin{bmatrix}\sigma_{11} & \sigma_{13}\\\sigma_{13} & \sigma_{33}\end{bmatrix}\begin{bmatrix}\beta_{11}^*(1)+\beta_{21}^*(1)\\\beta_{13}^*(1)+\beta_{23}^*(1)\end{bmatrix}$$

$$\text{var}[s_{ot} - E_{ct-1}[s_{ot}]] = \begin{bmatrix}\beta_{12}^*(0)+\beta_{22}^*(1)\\\beta_{14}^*(0)+\beta_{24}^*(1)\end{bmatrix}'\begin{bmatrix}\sigma_{22} & \sigma_{24}\\\sigma_{24} & \sigma_{44}\end{bmatrix}\begin{bmatrix}\beta_{12}^*(0)+\beta_{22}^*(1)\\\beta_{14}^*(0)+\beta_{24}^*(1)\end{bmatrix}.$$

Proof. These formulas follow directly from computing variances of the expressions in Fact 3. ‖

Remark 2. We now assume that the VMA(∞) representation inverts to a finite VAR. In this case, the moving average coefficients can be solved from the autoregressive parameters, so the variances in Lemma 1 can be expressed in terms of autoregressive parameters. Appendix B shows the standard way in which the parameter matrices of a VAR(p) can be redefined so that the model can be written as a VAR(1); except that $\mathbf{x}_t$ and $\mathbf{u}_t$ become ($4p \times 1$) vectors. It is the first four elements of these vectors whose random-walk and stationary components we care about. Therefore it suffices to work out the details for a VAR(1), where $\mathbf{x}_t$ and $\mathbf{u}_t$ are ($4p \times 1$)

$$A\mathbf{x}_t = B\mathbf{x}_{t-1} + \mathbf{u}_t,$$

then focus our analysis on the first four elements of these vectors. Assuming that A^{-1} exists, the VAR(1) is equivalent to

$$\mathbf{x}_t = \Theta\mathbf{x}_{t-1} + A^{-1}\mathbf{u}_t \qquad \text{where} \qquad \Theta = A^{-1}B,$$

or

$$\mathbf{x}_t = \sum_{k=0}^{\infty} \Theta^k A^{-1}\mathbf{u}_{t-k} \tag{A.9}$$

provided that the eigenvalues of Θ are inside the unit circle. This is the vector form of Equation (A.1). Thus the moving average coefficient, $\beta_{ih}(k)$, is the (i, h)th element of $\Theta^k A^{-1}$. Lemma 2 provides formulas in terms of VAR parameters for the infinite sums of moving average coefficients in Lemma 1. Theorem 1 provides formulas for the variances in Lemma 1 in terms of the VAR parameters.

Lemma 2 (Connection Between VMA and VAR Parameters). *Suppose* $\mathbf{x}_t$ *and* $\mathbf{u}_t$ *are* ($4p \times$ 1). *Let* e_i *be the* ($4p \times 1$) *vector with unity in the ith position and zeros elsewhere, and define the* ($4p \times 4$) *matrix* $H \equiv (e_1 \; e_2 \; e_3 \; e_4)$. *We have*

$$\begin{bmatrix} \beta_{i1}^*(j) \\ \beta_{i2}^*(j) \\ \beta_{i3}^*(j) \\ \beta_{i4}^*(j) \end{bmatrix} = H'\{\Theta^j [I - \Theta]^{-1} A^{-1}\}' e_i,$$

and

$$\sum_{j=1}^{\infty} \left\{ \begin{bmatrix} \beta_{11}^*(j) + \beta_{21}^*(j) \\ \beta_{12}^*(j) + \beta_{22}^*(j) \\ \beta_{13}^*(j) + \beta_{23}^*(j) \\ \beta_{14}^*(j) + \beta_{24}^*(j) \end{bmatrix}' \Omega \begin{bmatrix} \beta_{11}^*(j) + \beta_{21}^*(j) \\ \beta_{12}^*(j) + \beta_{22}^*(j) \\ \beta_{13}^*(j) + \beta_{23}^*(j) \\ \beta_{14}^*(j) + \beta_{24}^*(j) \end{bmatrix} \right\}$$
$$= (e_1 + e_2)'(\mathcal{S} - \Lambda)(e_1 + e_2),$$

1013

and

$$\sum_{j=1}^{\infty} \left\{ \begin{bmatrix} \beta_{11}^*(j) + \beta_{21}^*(j+1) \\ \beta_{12}^*(j) + \beta_{22}^*(j+1) \\ \beta_{13}^*(j) + \beta_{23}^*(j+1) \\ \beta_{14}^*(j) + \beta_{24}^*(j+1) \end{bmatrix}' \Omega \begin{bmatrix} \beta_{11}^*(j) + \beta_{21}^*(j+1) \\ \beta_{12}^*(j) + \beta_{22}^*(j+1) \\ \beta_{13}^*(j) + \beta_{23}^*(j+1) \\ \beta_{14}^*(j) + \beta_{24}^*(j+1) \end{bmatrix} \right\}$$

$$= (e_1 + \Theta' e_2)'(\mathcal{S} - \Lambda)(e_1 + \Theta' e_2),$$

where

$$\text{vec}\,[\mathcal{S}] = \left[I - \Theta \otimes \Theta \right]^{-1} \text{vec}\,[\Lambda],$$

$$\Lambda = \left\{ \left[I - \Theta \right]^{-1} A^{-1} \right\} H \Omega H' \left\{ \left[I - \Theta \right]^{-1} A^{-1} \right\}'$$

$$\Theta = A^{-1} B, \quad \textit{and} \quad \Omega = H' \text{cov}[\mathbf{u}_t, \mathbf{u}_t] H.$$

Proof. By definition of $\beta_{ih}(k)$ as the (i, h)th element of $\Theta^k A^{-1}$,

$$\beta_{ih}^*(j) \equiv \sum_{k=j}^{\infty} \beta_{ih}(k) = \sum_{k=j}^{\infty} e_i' \Theta^k A^{-1} e_h = e_i' \left\{ \sum_{k=0}^{\infty} \Theta^{j+k} \right\} A^{-1} e_h$$

$$= e_i' \Theta^j \left\{ \sum_{k=0}^{\infty} \Theta^k \right\} A^{-1} e_h = e_i' \Theta^j [I - \Theta]^{-1} A^{-1} e_h.$$

This implies that

$$\begin{bmatrix} \beta_{i1}^*(j) \\ \beta_{i2}^*(j) \\ \vdots \\ \beta_{i,4p}^*(j) \end{bmatrix} = \left\{ e_i' \Theta^j \left[I - \Theta \right]^{-1} A^{-1} \right\}'$$

and

$$\begin{bmatrix} \beta_{i1}^*(j) \\ \beta_{i2}^*(j) \\ \beta_{i3}^*(j) \\ \beta_{i4}^*(j) \end{bmatrix} = H' \left\{ e_i' \Theta^j \left[I - \Theta \right]^{-1} A^{-1} \right\}',$$

the first statement in Lemma 2. Consequently

$$\begin{bmatrix} \beta_{11}^*(j) + \beta_{21}^*(j) \\ \beta_{12}^*(j) + \beta_{22}^*(j) \\ \beta_{13}^*(j) + \beta_{23}^*(j) \\ \beta_{14}^*(j) + \beta_{24}^*(j) \end{bmatrix} = H' \left\{ e_1' \Theta^j \left[I - \Theta \right]^{-1} A^{-1} \right\}' + H' \left\{ e_2' \Theta^j \left[I - \Theta \right]^{-1} A^{-1} \right\}'$$

$$= H' \left\{ (e_1 + e_2)' \Theta^j \left[I - \Theta \right]^{-1} A^{-1} \right\}'$$

$$= H' \left\{ \Theta^j \left[I - \Theta \right]^{-1} A^{-1} \right\}' (e_1 + e_2).$$

Define the quadratic form

$$
\begin{aligned}
Q_j &\equiv (e_1 + e_2)' \left\{ \Theta^j \left[I - \Theta \right]^{-1} A^{-1} \right\} H \Omega H' \left\{ \Theta^j \left[I - \Theta \right]^{-1} A^{-1} \right\}' (e_1 + e_2) \\
&= (e_1 + e_2)' \Theta^j \left\{ \left[I - \Theta \right]^{-1} A^{-1} \right\} H \Omega H' \left\{ \left[I - \Theta \right]^{-1} A^{-1} \right\}' (\Theta')^j (e_1 + e_2) \\
&= (e_1 + e_2)' \Theta^j \Lambda (\Theta')^j (e_1 + e_2)
\end{aligned}
$$

where $\Lambda = \{ [I - \Theta]^{-1} A^{-1} \} H \Omega H' \{ [I - \Theta]^{-1} A^{-1} \}'$. Therefore,

$$
\sum_{j=1}^{\infty} Q_j = (e_1 + e_2)' \left\{ \sum_{j=0}^{\infty} \Theta^j \Lambda (\Theta')^j - \Lambda \right\} (e_1 + e_2) = (e_1 + e_2)' (\mathcal{S} - \Lambda)(e_1 + e_2)
$$

where $\mathcal{S} \equiv \sum_{j=0}^{\infty} \Theta^j \Lambda (\Theta')^j$. It remains to show that $\operatorname{vec} [\mathcal{S}] = \left[I - \Theta \otimes \Theta \right]^{-1} \operatorname{vec} [\Lambda]$.
From the definition of $\mathcal{S}$,

$$
\mathcal{S} - \Theta \mathcal{S} \Theta' = \sum_{k=0}^{\infty} \Theta^k \Lambda (\Theta')^k - \sum_{k=0}^{\infty} \Theta^{k+1} \Lambda (\Theta')^{k+1} = \Lambda.
$$

Vectorize both sides and rearrange

$$
\operatorname{vec} [\mathcal{S} - \Theta \mathcal{S} \Theta'] = \operatorname{vec} [\Lambda]
$$

$$
\operatorname{vec} [\mathcal{S}] - \operatorname{vec} [\Theta \mathcal{S} \Theta'] = \operatorname{vec} [\Lambda]
$$

$$
\operatorname{vec} [\mathcal{S}] - (I \otimes \Theta) \operatorname{vec} [\mathcal{S} \Theta'] = \operatorname{vec} [\Lambda]
$$

$$
\operatorname{vec} [\mathcal{S}] - (I \otimes \Theta)(\Theta \otimes I) \operatorname{vec} [\mathcal{S}] = \operatorname{vec} [\Lambda]
$$

using the fact that, for general conformable matrices A and B, $\operatorname{vec} [AB] = (I \otimes A) \operatorname{vec} [B] = (B' \otimes I) \operatorname{vec} [A]$. Since $(A \otimes B)(C \otimes D) = (AC \otimes BD)$ this further simplifies to

$$
\operatorname{vec} [\mathcal{S}] - (\Theta \otimes \Theta) \operatorname{vec} [\mathcal{S}] = \operatorname{vec} [\Lambda]
$$

or

$$
\left[I - \Theta \otimes \Theta \right] \operatorname{vec} [\mathcal{S}] = \operatorname{vec} [\Lambda].
$$

Provided that $\left[I - \Theta \otimes \Theta \right]^{-1}$ exists,

$$
\operatorname{vec} [\mathcal{S}] = \left[I - \Theta \otimes \Theta \right]^{-1} \operatorname{vec} [\Lambda].
$$

Finally, note that

$$
\begin{aligned}
&\begin{bmatrix} \beta_{11}^*(j) + \beta_{21}^*(j+1) \\ \beta_{12}^*(j) + \beta_{22}^*(j+1) \\ \beta_{13}^*(j) + \beta_{23}^*(j+1) \\ \beta_{14}^*(j) + \beta_{24}^*(j+1) \end{bmatrix} \\
&\quad = H' \{ e_1' \Theta^j \left[I - \Theta \right]^{-1} A^{-1} \}' + H' \{ e_2' \Theta^{j+1} \left[I - \Theta \right]^{-1} A^{-1} \}' \\
&\quad = H' \{ (e_1 + \Theta' e_2)' \Theta^j \left[I - \Theta \right]^{-1} A^{-1} \}' \\
&\quad = H' \{ \Theta^j \left[I - \Theta \right]^{-1} A^{-1} \}' (e_1 + \Theta' e_2).
\end{aligned}
$$

Define the quadratic form

$$W_j \equiv (e_1 + \Theta' e_2)' \left\{ \Theta^j \left[I - \Theta \right]^{-1} A^{-1} \right\} H \Omega H' \left\{ \Theta^j \left[I - \Theta \right]^{-1} A^{-1} \right\}' (e_1 + \Theta' e_2).$$

Reasoning identical to that applied to Q_j above yields

$$\sum_{j=1}^{\infty} W_j = (e_1 + \Theta' e_2)'(\mathcal{S} - \Lambda)(e_1 + \Theta' e_2),$$

which completes the proof. ∥

Theorem 1 (Component Variances in Terms of VAR Parameters). *Stated in text.*

Proof. Define $\widehat{I}_{\{ik\}}$ to be the 2×4 matrix with unity in the $(1, i)$th and $(2, k)$th positions and zero elsewhere. From Lemma 2 we have

$$\widehat{I}_{\{24\}} H' \left\{ \Theta^j \left[I - \Theta \right]^{-1} A^{-1} \right\}' e_i = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \beta_{i1}^*(j) \\ \vdots \\ \beta_{i4}^*(j) \end{bmatrix} = \begin{bmatrix} \beta_{i2}^*(j) \\ \beta_{i4}^*(j) \end{bmatrix}$$

and

$$\widehat{I}_{\{24\}} \Omega \widehat{I}_{\{24\}}' = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \sigma_{11} & \cdots & \sigma_{14} \\ \vdots & \ddots & \vdots \\ \sigma_{14} & \cdots & \sigma_{44} \end{bmatrix} \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} \sigma_{22} & \sigma_{24} \\ \sigma_{42} & \sigma_{44} \end{bmatrix}.$$

Combining these formulas and the statements in Lemmas 1 and 2, using the fact that $I_{\{ik\}} = \widehat{I}_{\{ik\}}' \widehat{I}_{\{ik\}}$, implies that

$$\begin{aligned}
\mathrm{var}[\epsilon_{ot}] &= \left[\widehat{I}_{\{24\}} H' \left\{ [I - \Theta]^{-1} A^{-1} \right\}' e_1 + \widehat{I}_{\{24\}} H' \left\{ [I - \Theta]^{-1} A^{-1} \right\}' e_2 \right] \widehat{I}_{\{24\}} \Omega \widehat{I}_{\{24\}}' \\
&\quad \times \left[\widehat{I}_{\{24\}} H' \left\{ [I - \Theta]^{-1} A^{-1} \right\}' e_1 + \widehat{I}_{\{24\}} H' \left\{ [I - \Theta]^{-1} A^{-1} \right\}' e_2 \right] \\
&= \left[e_1' \left\{ [I - \Theta]^{-1} A^{-1} \right\} + e_2' \left\{ [I - \Theta]^{-1} A^{-1} \right\} \right] H \left(I_{\{24\}} \Omega I_{\{24\}} \right) H' \\
&\quad \times \left[\left\{ [I - \Theta]^{-1} A^{-1} \right\}' e_1 + \left\{ [I - \Theta]^{-1} A^{-1} \right\}' e_2 \right] \\
&= (e_1' + e_2') \left\{ [I - \Theta]^{-1} A^{-1} \right\} H \left(I_{\{24\}} \Omega I_{\{24\}} \right) H' \left\{ [I - \Theta]^{-1} A^{-1} \right\}' (e_1 + e_2) \\
&= (e_1 + e_2)' \Lambda_{\{24;24\}} (e_1 + e_2).
\end{aligned}$$

Similar reasoning implies

$$\begin{aligned}
\mathrm{var}[s_{ct} - E_{ot}[s_{ct}]] &= \left[H' \left\{ \Theta[I - \Theta]^{-1} A^{-1} \right\}' (e_1 + e_2) \right]' \\
&\quad \times \left(I_{\{13\}} \Omega I_{\{13\}} \right) \left[H' \left\{ \Theta[I - \Theta]^{-1} A^{-1} \right\}' (e_1 + e_2) \right] \\
&= (e_1 + e_2)' \Theta \Lambda_{\{13;13\}} \Theta'(e_1 + e_2),
\end{aligned}$$

and

$$\begin{aligned}
\mathrm{var}[s_{ot} - E_{ct-1}[s_{ot}]] &= \left[H' \left\{ [I - \Theta]^{-1} A^{-1} \right\}' e_1 \right. \\
&\quad \left. + H' \left\{ \Theta[I - \Theta]^{-1} A^{-1} \right\}' e_2 \right] \left(I_{\{24\}} \Omega I_{\{24\}} \right) \\
&\quad \times \left[H' \left\{ [I - \Theta]^{-1} A^{-1} \right\}' e_1 + H' \left\{ \Theta[I - \Theta]^{-1} A^{-1} \right\}' e_2 \right] \\
&= (e_1 + \Theta' e_2)' \Lambda_{\{24;24\}} (e_1 + \Theta' e_2).
\end{aligned}$$

Noting that Λ in Lemma 2 is the same as $\Lambda_{\{1234;1234\}}$ here, the second expression in Lemma 2 immediately implies

$$\text{var}\,[s_{ct}] = (e_1 + e_2)'\big(\mathcal{S} - \Lambda_{\{1234;1234\}}\big)(e_1 + e_2).$$

The formula for $\text{var}\,[s_{ot} - E_{ct-1}[s_{ot}]]$ just above, and the second expression in Lemma 2, yield

$$\text{var}\,[s_{ot}] = (e_1 + \Theta'e_2)'\Lambda_{\{13;13\}}(e_1 + \Theta'e_2)$$
$$+ (e_1 + \Theta'e_2)'\big(\mathcal{S} - \Lambda_{\{1234;1234\}}\big)(e_1 + \Theta'e_2).$$

The argument for $\text{var}\,[\epsilon_{ct}]$ is slightly more complicated. Define $\widehat{I}_{\{i\cdot\}}$ to be the 2×4 matrix having unity in the $(1, i)$th position and zeros elsewhere; and $\widehat{I}_{\{\cdot k\}}$ to be the 2×4 matrix having unity in the $(2, k)$th position and zeros elsewhere. Note that $\widehat{I}_{\{ik\}} = \widehat{I}_{\{i\cdot\}} + \widehat{I}_{\{\cdot k\}}$. Using these definitions, the first expression in Lemma 2 implies that

$$\begin{bmatrix} \beta_{11}^*(0) + \beta_{21}^*(1) \\ \beta_{13}^*(0) + \beta_{23}^*(0) \end{bmatrix} = \widehat{I}_{\{13\}}H'\big\{[I - \Theta]^{-1}A^{-1}\big\}'e_1$$
$$+ \widehat{I}_{\{1\cdot\}}H'\big\{\Theta[I - \Theta]^{-1}A^{-1}\big\}'e_2 + \widehat{I}_{\{\cdot 3\}}H'\big\{[I - \Theta]^{-1}A^{-1}\big\}'e_2$$
$$= \big(\widehat{I}_{\{1\cdot\}} + \widehat{I}_{\{\cdot 3\}}\big)H'\big\{[I - \Theta]^{-1}A^{-1}\big\}'e_1$$
$$+ \widehat{I}_{\{1\cdot\}}H'\big\{[I - \Theta]^{-1}A^{-1}\big\}'\Theta'e_2 + \widehat{I}_{\{\cdot 3\}}H'\big\{[I - \Theta]^{-1}A^{-1}\big\}'e_2$$
$$= \widehat{I}_{\{1\cdot\}}H'\big\{[I - \Theta]^{-1}A^{-1}\big\}'(e_1 + \Theta'e_2)$$
$$+ \widehat{I}_{\{\cdot 3\}}H'\big\{[I - \Theta]^{-1}A^{-1}\big\}'(e_1 + e_2).$$

Using the first expression in Lemma 1

$$\text{var}\,[\epsilon_{ct}] = \big\{\widehat{I}_{\{1\cdot\}}R'(e_1 + \Theta'e_2) + \widehat{I}_{\{\cdot 3\}}R'(e_1 + e_2)\big\}'\widehat{I}_{\{13\}}\,\Omega\widehat{I}_{\{13\}}'$$
$$\times \big\{\widehat{I}_{\{1\cdot\}}R'(e_1 + \Theta'e_2) + \widehat{I}_{\{\cdot 3\}}R'(e_1 + e_2)\big\}$$

where the substitution $R \equiv H'\big\{[I - \Theta]^{-1}A^{-1}\big\}'$ has been made to simplify notation. Multiplying this expression yields

$$\text{var}\,[\epsilon_{ct}] = \big\{(e_1 + \Theta'e_2)'R\,\widehat{I}_{\{1\cdot\}}'\widehat{I}_{\{13\}}\,\Omega\,\widehat{I}_{\{13\}}' + (e_1 + e_2)'R\,\widehat{I}_{\{\cdot 3\}}'\widehat{I}_{\{13\}}\,\Omega\,\widehat{I}_{\{13\}}'\big\}$$
$$\times \big\{\widehat{I}_{\{1\cdot\}}R'(e_1 + \Theta'e_2) + \widehat{I}_{\{\cdot 3\}}R'(e_1 + e_2)\big\}$$
$$= (e_1 + \Theta'e_2)'R\,\widehat{I}_{\{1\cdot\}}'\widehat{I}_{\{13\}}\,\Omega\,\widehat{I}_{\{13\}}'\widehat{I}_{\{1\cdot\}}R'(e_1 + \Theta'e_2)$$
$$+ (e_1 + e_2)'R\,\widehat{I}_{\{\cdot 3\}}'\widehat{I}_{\{13\}}\,\Omega\,\widehat{I}_{\{13\}}'\widehat{I}_{\{1\cdot\}}R'(e_1 + \Theta'e_2)$$
$$+ (e_1 + \Theta'e_2)'R\,\widehat{I}_{\{1\cdot\}}'\widehat{I}_{\{13\}}\,\Omega\,\widehat{I}_{\{13\}}'\widehat{I}_{\{\cdot 3\}}R'(e_1 + e_2)$$
$$+ (e_1 + e_2)'R\,\widehat{I}_{\{\cdot 3\}}'\widehat{I}_{\{13\}}\,\Omega\,\widehat{I}_{\{13\}}'\widehat{I}_{\{\cdot 3\}}R'(e_1 + e_2).$$

Now, $\widehat{I}_{\{1\cdot\}}'\widehat{I}_{\{13\}} = I_{\{1\}}$ and $\widehat{I}_{\{\cdot3\}}'\widehat{I}_{\{13\}} = I_{\{3\}}$, so this simplifies to

$$
\begin{aligned}
\text{var}[\epsilon_{ct}] &= (e_1 + \Theta'e_2)'R\,I_{\{1\}}\,\Omega\,I_{\{1\}}\,R'(e_1 + \Theta e_2) \\
&\quad + 2(e_1 + e_2)'R\,I_{\{3\}}\,\Omega\,I_{\{1\}}\,R'(e_1 + \Theta'e_2) \\
&\quad + (e_1 + e_2)'R\,I_{\{3\}}\,\Omega\,I_{\{3\}}\,R'(e_1 + e_2) \\
&= (e_1 + \Theta'e_2)'\Lambda_{\{1;1\}}(e_1 + \Theta'e_2) + 2(e_1 + e_2)'\Lambda_{\{3;1\}}(e_1 + \Theta'e_2) \\
&\quad + (e_1 + e_2)'\Lambda_{\{3;3\}}(e_1 + e_2).
\end{aligned}
$$

Finally, the expression for $\Lambda_{\{\alpha_1,\alpha_2\}}$ in the statement of the theorem can be obtained by noting that $[I - \Theta]^{-1}A^{-1} = [I - A^{-1}B]^{-1}A^{-1} = [A - B]^{-1}$. $\qquad\qquad\qquad\|$

Appendix B: Representation of VAR(p) as VAR(1)

Consider a vector autoregression in m variables with p lags. Let $\mathbf{y}_t$ be the $m \times 1$ vector of variables and $\mathbf{u}_t$ be the $m \times 1$ vector of disturbances in the VAR(p):

$$
A\mathbf{y}_t = B_1\mathbf{y}_{t-1} + B_2\mathbf{y}_{t-2} + \cdots + B_p\mathbf{y}_{t-p} + \mathbf{u}_t, \quad \text{where} \quad u_t \sim \text{i.i.d. } (0, \Omega),
$$

A, Bs, and Ω are $m \times m$ matrices. Arrange the variables as follows:

$$
\begin{bmatrix} A & 0 & 0 & \cdots & 0 \\ 0 & I & 0 & \cdots & 0 \\ 0 & 0 & I & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & I \end{bmatrix} \begin{bmatrix} \mathbf{y}_t \\ \mathbf{y}_{t-1} \\ \mathbf{y}_{t-2} \\ \vdots \\ \mathbf{y}_{t-p+1} \end{bmatrix} = \begin{bmatrix} B_1 & B_2 & \cdots & B_{p-1} & B_p \\ I & 0 & \cdots & 0 & 0 \\ 0 & I & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & I & 0 \end{bmatrix} \begin{bmatrix} \mathbf{y}_{t-1} \\ \mathbf{y}_{t-2} \\ \mathbf{y}_{t-3} \\ \vdots \\ \mathbf{y}_{t-p} \end{bmatrix} + \begin{bmatrix} \mathbf{u}_t \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}
$$

Where 0 is the $m \times m$ matrix of zeros and I is the $m \times m$ identity matrix. Rewrite this as

$$
A_*\mathbf{x}_{*t} = B_*\mathbf{x}_{*t-1} + \mathbf{u}_{*t}, \quad \text{where} \quad \mathbf{u}_{*t} \sim \text{i.i.d. } (0, \Omega_*)
$$

and

$$
\Omega_* = \begin{bmatrix} \Omega & 0 & \cdots & 0 & 0 \\ 0 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & 0 \end{bmatrix}.
$$

References

Amihud, Y., and H. Mendelson, 1987, "Trading Mechanisms and Stock Returns: An Empirical Investigation," *Journal of Finance*, 42, 533–553.

Amihud, Y., and H. Mendelson, 1991, "Volatility and Trading: Evidence from the Japanese Stock Market," *Journal of Finance*, 46, 1765–1789.

Andersen, T., 1996, "Return Volatility and Trading Volume: An Information Flow Interpretation of Stochastic Volatility," *Journal of Finance*, 51, 169–204.

Barclay, M., R. Litzenberger, and J. Warner, 1990, "Private Information, Trading Volume and Stock Return Variances," *Review of Financial Studies*, 3, 233–253.

Beveridge, S., and C. Nelson, 1981, "A New Approach to Decomposition of Economic Time Series into Permanent and Transitory Components with Particular Attention to Measurement of the 'Business Cycle'," *Journal of Monetary Economics*, 7, 151–174.

Dhrymes, P., 1984, *Mathematics for Econometrics*, 2nd ed., Springer-Verlag, New York.

Forster, M., and T. George, 1995, "Trading Hours, Information Flow and International Cross-Listing," *International Review of Financial Analysis*, 4, 19–34.

Forster, M., and T. George, 1996, "Pricing Errors at the NYSE Open and Close: Evidence from Internationally Cross-Listed Stocks," *Journal of Financial Intermediation*, 5, 95–126.

Foster, F. D., and S. Viswanathan, 1995, "Can Speculative Trading Explain the Volume-Volatility Relation?" *Journal of Business and Economic Statistics*, 13, 379–396.

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

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

George, T., and C. Hwang, 1995, "Transitory Price Changes and Price Limit Rules: Evidence from the Tokyo Stock Exchange," *Journal of Financial and Quantitative Analysis*, 30, 313–327.

George, T., G. Kaul, and M. Nimalendran, 1991, "Estimation of the Bid-Ask Spread and its Components: A New Approach," *Review of Financial Studies*, 4, 623–656.

Glosten, L., and L. Harris, 1988, "Estimating the Components of the Bid-Ask Spread," *Journal of Financial Economics*, 21, 123–142.

Hamilton, J., 1994, *Time Series Analysis*, Princeton University Press, Princeton, NJ.

Hansen, L., 1982, "Large Sample Properties of Generalized Method of Moment Estimators," *Econometrica*, 50, 1029–1054.

Harvey, A., E. Ruiz, and E. Sentana, 1992, "Unobserved Component Time Series Models with ARCH Disturbances," *Journal of Econometrics*, 52, 129–157.

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

Hasbrouck, J., 1991b, "The Summary Informativeness of Stock Trades: An Econometric Analysis," *Review of Financial Studies*, 4, 571–595.

Hasbrouck, J., 1993, "Assessing the Quality of a Security Market: A New Approach to Transaction-Cost Measurement," *Review of Financial Studies*, 6, 191–212.

Huang, R., and H. Stoll, 1997, "The Components of the Bid-Ask Spread: A General Approach," *Review of Financial Studies*, 10, 995–1034.

Jones, C., G. Kaul, and M. Lipson, 1994, "Information, Trading and Volatility," *Journal of Financial Economics*, 36, 127–154.

Kirby, C., 1998, "The Restrictions on Predictability Implied by Rational Asset Pricing Models," *Review of Financial Studies*, 11, 343–382.

Lee, C., and M. Ready, 1991 "Inferring Trade Direction from Intraday Data," *Journal of Finance*, 46, 733–746.

Madhavan, A., 1992, "Trading Mechanisms in Securities Markets," *Journal of Finance*, 47, 607–641.

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.

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

Roll, R., 1984, "Orange Juice and Weather," *American Economic Review*, 74, 861–880.

Ronen, T., 1997, "Tests and Properties of Variance Ratios in Microstructure Studies," *Journal of Financial and Quantitative Analysis*, 32, 183–204.

Smith, T., 1995, "Econometrics of Financial Models and Market Micro-Structure Effects," *Journal of Financial and Quantitative Analysis*, 29, 519–540.

Stoll, H., 1989, "Inferring the Components of the Bid-Ask Spread: Theory and Empirical Tests," *Journal of Finance*, 44, 115–134.

Stoll, H., and Whaley, R., 1990, "Stock Market Structure and Volatility," *Review of Financial Studies*, 3, 37–71.

Watson, M., 1986, "Univariate Detrending Methods with Stochastic Trends," *Journal of Monetary Economics*, 18, 49–75.