Common Factors in Prices, Order Flows and Liquidity

Joel Hasbrouck

Stern School of Business
New York University
New York, NY 10012-1126
jhasbrou@stern.nyu.edu
212/998-0310

and

Duane J. Seppi*

Graduate School of Industrial Administration
Carnegie Mellon University
Pittsburgh, PA 15213-3890
ds64@andrew.cmu.edu
412/268-2298

This draft: April 4, 2000

Comments welcome

* We thank Ian Domowitz, Roger Edelen, Dennis Epple, Bill Schwert (the editor), Myron Slovin, an anonymous referee and seminar participants at HEC and the 2000 AFA meetings for helpful comments. The Institute for Quantitative Research in Finance provided generous financial support. All errors are our own responsibility.

Common Factors in Prices, Order Flows and Liquidity

Abstract

How important are cross-stock common factors in the price discovery/liquidity provision process in equity markets? We investigate two aspects of this question for the thirty Dow stocks. First, using principal components and canonical correlation analyses we find that both returns and order flows are characterized by common factors. Commonality in the order flows explains roughly two-thirds of the commonality in returns. Second, we examine variation and common covariation in various liquidity proxies and market depth (trade impact) coefficients. Liquidity proxies such as the bid-ask spread and bid-ask quote sizes exhibit time variation which helps explain time variation in trade impacts. The common factors in these liquidity proxies are, however, relatively small.

1. Introduction

An open issue in the microstructure of equity markets is the role of common cross-firm variation in short-horizon returns, order flows and liquidity. Since order flows are generally held to contain informed components, does common covariation in stocks' orders account for the covariance structure of short-term returns? Furthermore, is liquidity driven by strong common factors? The equity market breaks of 1987 and 1989, as well as the debt market crisis of 1998, for example, are widely perceived as systematic breakdowns in liquidity.

These issues are important for both microstructure theory and for institutional trading practice. Subrahmanyam (1991), Chowdhry and Nanda (1991), Kumar and Seppi (1994), Caballe and Krishnan (1994) have all extended the work of Kyle (1984, 1985) to multiasset markets by adding investors who are informed about macroeconomic factors and/or who have portfolio-wide liquidity shocks (e.g., portfolio substitution). In such environments intermarket price discovery and order flow dynamics are obviously more subtle than when private information and/or trading noise is purely idiosyncratic.

Until recently, however, little direct empirical research has been conducted on the magnitudes of cross-stock interactions at the microstructure level. Given the sheer size of the Fitch, ISSM, TAQ and TORQ databases, it is perhaps not surprising that previous work has tended to focus on individual stocks in isolation from each other. This focus on stocks in isolation has, however, left us ignorant of even the most basic facts about *cross-sectional* interactions between stocks.

This paper answers these questions in two ways. First, we use principal components analysis to show that common factors exist in the order flows and returns of the 30 stocks in the Dow Jones Industrial Average (DJIA). In addition, canonical correlation analysis documents that the common factor in returns is highly correlated with the common factor in order flows. Second, we find some evidence of a common factor in quote-based proxies for liquidity, and to a lesser degree, in inferred price impact coefficients, after controlling for previously documented time-of-day seasonalities.

We choose the thirty Dow stocks as our sample because the rapid pace of trading there allows us to construct high-frequency trading measures which approximate the idea of contemporaneous (i.e. simultaneous) order flow across stocks (as in Subrahmanyam (1991) and op. cit.) as well as giving us frequently updated prices. In particular, we aggregate trading for each stock over fifteen minute intervals and measure price changes using the quote mid-points at the beginning and end of each interval.

Our investigation of cross-stock interactions builds on a foundation of prior work on the price, volume, and liquidity properties of individual stocks viewed in isolation (surveyed in O'Hara (1995) and Hasbrouck (1996a)). The study of common factors in stock returns is a classic theme in financial economics. Although the standard asset pricing models (e.g., CAPM, APT) do not assign a significant role to trading per se, Lo and Wang (2000) show that, under certain assumptions, portfolio rebalancing and liquidation imply a factor structure for trading volume.

The approach in our paper, in contrast to Lo and Wang (2000), is more descriptive and statistical, and assigns no distinctive role to factor portfolios. This last difference may be justified by noting that modern microstructure theory ascribes a prominent informational role to trading. Hence, cross-firm commonalities in order flows may be influenced by the differential liquidity of individual stocks as well as by the factor structure of valuation fundamentals.

Another line of research on cross-sectional price/order flow interactions is the work on index arbitrage and the cash/futures basis.[1] In this case, however, strong interactions across markets are expected a priori, since, after all, it is exactly the same

[1] See MacKinlay and Ramaswamy (1998), Chan, Chan, and Karolyi (1991), Chan (1993), Hasbrouck (1996b), Harris, Sofianos, and Shapiro (1994), and Miller, Muthuswamy, and Whaley (1994).

portfolio which is traded in different locations. In contrast, the Dow stocks in our study, while closely related, are far from perfect substitutes.

Several forms of time-varying liquidity have been previously documented. First, Wood, McInish, and Ord (1985), Jain and Joh (1988) and Foster and Viswanathan (1990) study deterministic (e.g., time-of-day) components. Second, variation has been studied around earnings reports (Lee, Mucklow, and Ready (1993)), dividends (Koski (1996)), stock splits (Desai, Nimalendran, and Venkataraman (1998)), take-over announcements (Foster and Viswanathan (1995b)) and other identifiable events. In contrast, we are interested in stochastic variation in liquidity and, in particular, in possible co-variation due to common components rather than largely idiosyncratic firm-specific events.

An important paper which does look at stochastic liquidity is Foster and Viswanathan (1995a). They use simulated method of moments to estimate a repeated one-period Kyle model with time-varying parameters. Caballe and Krishnan (1994) extend their approach by adding a second stock to look at interaction effects. In contrast, ours is a less structural approach distinguishing between common and idiosyncratic factors for a broader cross-section of stocks.

Chordia, Roll, and Subrahmanyam (2000) also explore cross-sectional interactions in liquidity measures using quote data. Like Lo and Wang (2000), they too assign a special role to the market portfolio. In contrast, our study characterizes relationships involving returns and order flows as well as liquidity. However, while we study a cross-section of just the thirty (actively traded) Dow firms, Chordia et al. use a cross-section of roughly one thousand stocks. Huberman and Halka (1999) estimate time series models for quotes and depths for market capitalization-weighted portfolios. They find evidence of commonality in liquidity in that the estimated model residuals are correlated across portfolios.

This paper is organized as follows. Section 2 establishes the economic framework for the study and outlines the initial analyses. Section 3 describes the data. The joint statistical properties of returns and signed order flows are analyzed in Section 4; those of absolute returns and unsigned order flows, in Section 5. Section 6 explores variation and covariation in liquidity proxies derived from quote data. Section 7 attempts to relate variation in these proxies to the price impacts of trades. A brief summary concludes the paper in Section 8.

2. Economic Framework

The starting point for our analysis is the hypothesis that a set of common and idiosyncratic variables underlies both stock returns and order flows. These variables represent, on the one hand, information about economic fundamentals and, on the other, non-informational demands for liquidity/immediacy. In particular, we assume that the cross-section of orders for a set of *n* stocks can be represented statistically by the linear factor model:

$$x_t = \theta F_t + \varepsilon_t \tag{1}$$

where x_t is a column *n*-vector of the order flows arriving at time *t*; F_t is a column vector of order flow common factors; θ is a conformable matrix of factor loadings; and ε_t is a vector of idiosyncratic disturbances where $E\,F_t\varepsilon_t' = 0$ and $E\varepsilon_{it}\varepsilon_{jt} = 0$ for $i \neq j$. With the normalizations $EF_t = 0$ and $Cov(F_t) = I$ these imply that $Cov(x_t) = \theta\,\theta' + \Omega_\varepsilon$, where Ω_ε is diagonal, i.e., the covariance structure of the order flows is explained by the common factors and their loadings. The model is parsimonious if the number of factors (rows of F_t) is smaller than the number of firms (*n*). A similar factor model describes returns r_t in terms of a vector of factors G_t:

$$r_t = \varphi G_t + \eta_t \tag{2}$$

From an economic perspective, the order flow model (1) allows for both liquidity- and information-motivated components in both common factors and idiosyncratic disturbances. For example, dynamic hedging strategies, tax- and calendar-related effects and naïve momentum trading could all plausibly lead to correlated liquidity trading, whereas private forecasts about macroeconomic variables would cause correlated informed trading. Similarly, commonalities in innovations to discount rates and expectations of future cash flows will give rise to common return components in (2).

In contrast to the statistical representations in (1) and (2), microstructure models establish causal links by decomposing returns into a portion due to a stock's own order flow and a non-trade component, e.g., for the *i*th return, $r_{it} = \lambda_i x_{it} + u_{it}$ where λ_i is a liquidity or price-impact coefficient. In this paper, we specifically consider a multivariate extension of the standard univariate price/order flow relationship:[2]

$$r_t = \Lambda\, x_t + u_t \tag{3}$$

where Λ is an $n \times n$ matrix of price impact coefficients and the u_t are disturbances with $E\, x_t u_t' = 0$. (Λ is initially assumed constant; it is generalized to a time-varying stochastic process in Sections 6 and 7.) The first term on the right in (3) captures the portion of returns due to conditioning on the cross-section of order flows. The residual reflects public non-trade information. The model, in particular, allows for the possibility that quote-setters observe and learn from the order flows of other stocks, as well their own.

It is clear from (3) that a factor structure for the x_t can induce a factor structure for r_t. In addition, the non-trade disturbance may have its own factor structure:

$$u_t = \xi\, H_t + \omega_t \tag{4}$$

---

[2] The standard univariate model is a special case of (3) where Λ is diagonal. Note also that, even if Λ is diagonal, common factors in the "own" order flows can pass through into returns as described below.

Common factors here could arise from announcements with common effects (e.g., monetary policy).

Substituting equations (1) and (4) into (3) yields $r_t = \Lambda\left(\theta F_t + \varepsilon_t\right) + \left(\xi H_t + \omega_t\right)$. The correspondence between this and the return factor model (2) then implies

$$\varphi G_t = \Lambda\theta F_t + \xi H_t \tag{5}$$

This says that the factor structure for returns may have a microstructure foundation in either or both order flows (e.g., if market makers condition on realized orders x_t) as well as from the non-trade residuals u_t. However, the factor structure of orders does not *automatically* pass through into returns. For example, if a particular order flow factor (component of F_t) is observable and uninformative, market makers could, if they chose, purge that factor by choosing Λ so that the kth column of $\Lambda\theta$ is zero.

We ask a variety of questions of this model. Are cross-stock commonalities in returns and order flows significant? Are the return and order flow factors correlated? How much of the commonality in returns can be attributed to order flow commonalities versus common factors in the non-trade residuals u_t? Do the common factors in order flows represent information or cross-stock liquidity shocks? Do other stock's order flows have additional explanatory power for pricing beyond a stock's own order flow?

3. Data

The data for this study are from the NYSE's TAQ database, which contains all trades and quotes for stocks listed on the NYSE, the AMEX and NASDAQ's National Market System. Our sample is limited to the thirty Dow stocks. This selection is motivated by 1) our intention to include firms for which common factors in liquidity trading (e.g., because of indexation) and information are plausible a priori and 2) the fact that we need actively traded stocks to construct approximately *concurrent* order flows at high frequencies. The sample covers the 252 trading days in 1994. Table 1 gives summary statistics for market activity in the sample.

We establish a common time-frame for the data series using fifteen-minute intervals covering 9:30 to 9:45, 9:45 to 10:00, . . . 15:45 to 16:00 for a total of 26 intervals per trading session on the NYSE. Hereafter, the time subscript *t* indexes these intervals. A fifteen-minute time resolution represents a compromise between, on the one hand, needing to look at correlations in contemporaneous order flows across stocks (e.g., at a one-second resolution few trades are contemporaneous) and, on the other, seeking to minimize return/order flow simultaneity problems. In particular, at shorter horizons there is less time for feed-back effects from prices into subsequent order submissions due to portfolio insurance and other positive feed-back strategies. In addition, specification (3) ignores transitory mid-point dynamics. Using intervals shorter than fifteen minutes would exacerbate these omitted dynamics.

We calculate the log quote midpoint return as

$$r_{i,t} = \log\left(m_{i,t} / m_{i,t-1}\right) \tag{6}$$

where $m_{i,t}$ is the midpoint of the NYSE bid and offer quotes for firm *i* prevailing at the end of interval *t*.

Unsigned order flow measures are derived from the consolidated trade data. Denote the number of trades for firm *i* in interval *t* by n_{it}. For the *j*th trade, $j = 1, \ldots, n_{it}$, in interval *t*, let p_{ij} and v_{ij} be the price per share and share volume. The total share volume in the interval is $\sum_{j=1}^{n_{it}} v_{ij}$, and the total dollar volume is $\sum_{j=1}^{n_{it}} p_{ij} v_{ij}$.

Studies of short-term price-trade dynamics suggest that the trade impact is concave in size.[3] We therefore also examine the cumulative square-root of the dollar volume ("SRD volume") $\sum_{j=1}^{n_{it}} \sqrt{p_{ij} v_{ij}}$. We also explore size effects by constructing order flow measures based on small (≤2,000 shares), medium (2,001-10,000 shares) and large (>10,000 shares) trades.

Signed order flow measures, $sign\left(v_{ij}\right)v_{ij}$, corresponding to the above are derived by letting the imputed direction of a trade be the sign of the difference $p_{ij} - m_{i,j-1}$. Thus, a trade at the ask price is positive; a trade at the bid is negative. The individual signed trades are then summed over period *t* to obtain cumulative signed number, share volume, dollar volume and square-root dollar volumes. Trades occurring at the quote midpoint are dropped from the sum (effectively assigned a sign of zero).

To differentiate stochastic sources of common time-variation from deterministic sources, the series are standardized to remove the time-of-day effects documented in Wood, McInish, and Ord (1985) etc. For a representative variable "*z*", let $z_{i,d,k}$ denote the observation for firm *i* for fifteen-minute subperiod *k* on day *d*. The standardized value is $z^*_{i,d,k} = \left(z_{i,d,k} - \mu_{i,k}\right)/\sigma_{i,k}$ where $\mu_{i,k}$ and $\sigma_{i,k}$ are the mean and standard deviation for firm *i* and subperiod *k*, estimated across days.

4. Returns and signed volume measures

In this section we estimate the factor and regression models for order flows and returns discussed in Section 2. To investigate the factor structures we rely on the standard techniques of principal component and canonical correlation analysis. These

---

[3] See Madhavan and Smidt (1991) and Hasbrouck (1991). A concave price-order flow relation is predicted by models of reputation, stealth trading and counterparty search in Seppi (1990), Barclay and Warner (1993) and Keim and Madhavan (1996).

procedures model the factors as linear compounds (i.e., functions) of order flows and returns, e.g., ax_t and br_t where a and b are fixed coefficient vectors.

Briefly, a principal components analysis constructs factors to maximize explanatory power within a set of related variables. For example, given an order flow factor ax_t, we can use n regressions (linear projections) of the form

$$x_{jt} = \gamma_j \left(ax_t\right) + e_{jt} \text{ for } j = 1,\ldots n \tag{7}$$

to decompose the total order flow variance as follows:

$$\sum_{j=1}^{n} Var\left(x_{it}\right) = \left(\sum_{j=1}^{n} \gamma_j^2\right) Var\left(a\,x_t\right) + \sum_{j=1}^{n} Var\left(e_{jt}\right) \tag{8}$$

where the first term on the right is the total variation attributable to (i.e., explained by) the factor. If the vector a is chosen (up to a normalization) to maximize the explanatory power in (8), then ax_t is called the first principal component of x_t. Computationally, the first term on the right, $\left(\sum_{j=1}^{n} \gamma_j^2\right) Var\left(a\,x_t\right)$, is the first eigenvalue of $Cov\left(x_t\right)$, and a is the first eigenvector. Up to n principal components may be constructed successively, subject to the restriction that the ith component is uncorrelated with the i-1 components already extracted.

Principal component analyses are generally sensitive to the units in which the underlying variables are measured. It is customary, therefore, to standardize variables to unit variances, or equivalently to extract the eigenvalues and eigenvectors from the correlation matrix. This is not a major concern here because the elements of x_t and r_t are all order flows or all returns, and so are in consistent units. Our motivation for standardizing the variables is to remove deterministic time-of-day effects, as described in the previous section. Standardization moreover facilitates certain comparative analyses. For these reasons, the present analyses are based on standardized variables, except where noted.

The need to summarize the explanatory power of a panel of regressions (such as equation (7)) arises frequently in this paper. For the sake of consistency, we use the metric underlying the principal components analysis, the proportion of explained total variation of the standardized variables. In most cases, this summary R^2 was virtually identical to the average of the individual regression R^2s, which is a possible alternative measure.

Whereas principal components maximize explanatory power within one set of variables, factors constructed using canonical correlation analysis maximize the power of the factors to explain the covariance between two sets of variables. Specifically, given the order flow and return vectors, the *first canonical variates* are the pair of linear compounds ax_t and br_t that maximize $Corr\left(a\,x_t, b\,r_t\right)$. Succeeding pairs of canonical variates are determined subject to the restriction that they be uncorrelated with the canonical variates previously constructed.[4]

Since the models given in equations (1)-(4) are factor models it would seem logical to estimate them directly using factor analysis. We chose to use principal component and canonical correlation analyses because factor models are generally

---

[4] One might incorrectly conjecture that principal components and canonical variates must necessarily be virtually identical. By way of counterexample, consider a panel sample of two variables $\left(x_{it} \text{ and } y_{it}\right)$ for $i = 1, ..., n$ firms. Suppose that the underlying factor structure reflects three independent standard normal variables, F_t^x, F_t^y and z. All but the last pair of x_{it} and y_{it} depend on the first two factors: $x_{it} = F_t^x$ and $y_{it} = F_t^y$ for $i = 1, ..., n-1$. The last pair are driven by z: $x_{nt} = y_{nt} = z$. A principal components analysis will determine that $(n\text{-}1)/n$ of the total variation in the x's can be attributed to a single principal component, and similarly for the y's. These (first) x and y components are independent. In contrast, the first canonical variates are x_{nt} and y_{nt}, which are perfectly correlated but independent of the principal components.

estimated via maximum likelihood, which necessitates distributional assumptions (usually normality) that are implausible in the present application.

a. Principal components

Table 2 first reports means and standard deviations for the various unstandardized series (pooled across the 30 stocks) to indicate the scale and variability of the raw data. For reasons noted above, however, the principal component analysis is based on the covariance matrix of the standardized variables.

Since the variance of a standardized variate is unity, the total variation on the left hand side of equation (8) is simply the size of the variable set, *n*. Hence if the variables were perfectly positively correlated, the first eigenvalue of this covariance matrix would be *n*, i.e., a single factor would suffice to explain all of the variation. If instead the *n* variates were uncorrelated, the covariance matrix would have a single eigenvalue (unity) with a multiplicity of *n*.

Commonalities in returns, of course, are long-established results. We report them here as a reference point for assessing the other commonalities in our sample. The first eigenvalue of (standardized) returns in Table 2 is 6.32. This implies that $6.32/30 = 21\%$ of the total variation in fifteen-minute returns can be explained by a single common factor. The second and third eigenvalues are close to one, however, indicating that additional common factors are of negligible importance.

The first eigenvalues for the signed volume measures also suggest commonality. This is most evident for small trades ($\leq 2{,}000$ shares) and medium trades (2,001-10,000 shares), and less so for large block trades (>10,000 shares).[5] The weakness of the

---

[5] Although we do not identify the sources of the order flow common factors, Edelen and Warner (1999) find that mutual fund flows are highly correlated with returns at a daily

common factor in large-sized trades is not particularly surprising. Even if two block trade orders for different stocks arrived simultaneously, it is unlikely, given the mechanics of how blocks are shopped in the upstairs market, that counterparties would be located and the trades executed within the same fifteen minute interval.

Although the standard sampling results for principal components presuppose multivariate normality, they provide at least some guidance in assessing statistical significance. In particular, if the data were multivariate normal, then the eigenvalues of the sample covariance matrix would have a known asymptotic distribution (Morrison (1976)): with n observations on $z \overset{d}{\sim} N(\mu, \Sigma)$, $\sqrt{n}(l_i - \psi_i) \overset{asy.dist.}{\sim} N(0, \; 2\psi_i^2)$ where ψ_i and l_i are the population and sample values of the ith eigenvalue. Here we are working with an estimated correlation matrix and normality is doubtful. Nevertheless, direct application of the asymptotic approximation for returns with $n \approx 6{,}000$ yields an estimated standard error for the first eigenvalue of $\sqrt{2(6.32)^2/6{,}000} = 0.12$. The analogous standard error for the first eigenvalue of signed square-root dollar volume is 0.074. Even allowing for substantial understatement due to violation of the distributional assumptions, these calculations are highly suggestive of statistical significance for the first eigenvalues of both returns and order flows.

As an aside, one might wonder if the common factor in orders is simply measuring program trading. The first order flow principal component explains $2.36/30 \approx 7.8\%$ of the total variation. The NYSE reports that in 1994, 11.6% of total volume was attributable to program trades (New York Stock Exchange (1996)). Despite the closeness of the two numbers, however, the measurements are conceptually quite different. Since volume is the absolute value of signed order flow, volume proportions reflect ratios of standard deviations (not variances). Thus, index arbitrage and other

---

level. To the extent that they constitute broadly diversified portfolios, the commonality that we find in order flows may well derive, at least in part, from mutual fund flows.

reported program trading does not fully account for all of the observed commonality in signed order flow from principal component analysis.[6]

b. Canonical correlations

Given the presence of common factors F_t in order flows and G_t in returns, the next question is whether they are statistically correlated with each other. We temporarily defer the causal question of whether the common factor in orders is "informative" in a microstructure sense until the next section. For brevity we limit our analysis to the signed square-root dollar (SSRD) volume. We chose this variable because, among all of the signed volume measures, it is generally the most highly correlated with returns at the individual firm level.

Table 3 reports the canonical correlation analyses. Panel A reports the correlations between successively constructed canonical variates (factors). The first entry says that it is possible to construct linear compounds of standardized returns and order flows for which the correlation is maximized at 0.829. By way of contrast, if each r_{it} were perfectly correlated with its own order flow x_{it}, but returns and order flows were uncorrelated across firms, this correlation would be $\sqrt{1/n} = 0.183$ (ignoring sampling error). Thus the commonalities in returns and order flows are statistically interrelated.

---

[6] Suppose that the signed order flow for the ith stock is $x_{it} = x_t^P + x_{it}^N$ where x_t^P is a zero-mean program order flow (common across stocks) with $Var\left(x_t^P\right) = 0.078$ and x_{it}^N is a zero-mean nonprogram order flow, i.i.d. across stocks with $Var\left(x_{it}^N\right) = 0.922$. The proportion of program trading in total volume, however, is $\left(n\left|x_t^P\right|\right)\big/\left(\sum\left|x_{it}\right|\right)$.
For a standard normal variate z, $E\left|z\right| = \sqrt{2/\pi}$, so $E\left|x_t^P\right| = \sqrt{0.078\times 2/\pi}$ and $E\left|x_{it}\right| = \sqrt{2/\pi}$. The expected proportion of program trading therefore converges to $\sqrt{0.078} \approx 28\%$, which is much larger than 11.6%.

A canonical variate can be described by the proportion of variation it explains within its own variable set (like a principal component), but also by the proportion of variation explained in the other variable set. These values, comprising a canonical redundancy analysis, are reported in Panel B of Table 3. The first canonical variate for returns, for example, explains 20.6% of the return variation and 9.1% of the signed order flow variation. For comparison purposes, the table also reports variation explained by the principal components. In the case of returns, this is 21.0%, which suggests that the first return canonical variate and the first principal component are functionally equivalent. The corresponding analysis of signed order flows is similar.

The statistical similarity of the order flow factors constructed by principal components and those obtained from canonical correlation has a natural economic interpretation. It suggests that investors with macroeconomic information and those motivated by systematic liquidity shocks both trade roughly the same basket of stocks. In other words, it is not possible to identify particular waves of orders arriving in the market as purely liquidity-motivated simply by looking at the amounts (i.e., loadings) of the different stocks traded. This would be the case, for example, if both types of investors trade indices.

c. Residual commonality

From equation (5) it is clear that commonality in returns can arise either from order flows or the return residuals. How much covariability remains in returns after accounting for the part associated with the common factor in order flow? Let P_t^x denote the first principal component of order flow and consider the following regressions:

$$r_{it} = \lambda_i^{common} P_t^x + u_{it} \text{ for } i = 1, \ldots, n \qquad (9)$$

After removing the order flow factor, we perform a principal components analysis on the residuals from (9) to assess the strength of the remaining commonality. Denote by P_t^u the first principal component of the residuals. If the u_{it} are projected onto this component, $u_{it} = \gamma_i P_t^u + \varepsilon_{it}$, we can decompose the total return variation as follows:

$$\sum Var\left(r_{it}\right) = \left(\sum\left[\lambda_i^{common}\right]^2\right) Var\left(P_t^x\right) + \left(\sum \gamma_j^2\right) Var\left(P_t^u\right) + \sum Var\left(\varepsilon_{it}\right) \tag{10}$$

where the first term on the right is due to order flow commonality and the second is due to residual commonality. The two components are orthogonal by construction. This mode of analysis is termed partial principal components since it involves a principal components analysis on the r_{it} after controlling for P_t^x.

Panel A of Table 4 reports the results of this analysis. P_t^x explains 14.6% of the total variation, and P_t^u explains another 7.1% for a cumulative total of 21.7%. To put these numbers in perspective, the first principal component of returns explains 21.0% of the total variation. Thus, roughly two-thirds of the common variation in returns may be accounted for by the order flow commonality.

This attribution suggests that both order flow and residual commonality contribute to return commonality. The existence of two prominent (and uncorrelated) sources of commonality might seem to suggest that returns have at least two prominent factors. The principal components analysis of returns, however, displays only one dominant component (cf. Panel B of Table 3). These two results can be reconciled by considering equation (5), which expresses the return commonality $\left(\varphi G_t\right)$ in terms of commonalities in order flows $\left(\Lambda\theta F_t\right)$ and residuals $\left(\xi H_t\right)$, and considering the conditions under which the latter two components will be indistinguishable in a principal components analysis.

One possibility is that F_t and H_t are perfectly correlated. If $H_t = cF_t$, where c is a coefficient of proportionality, then $\varphi G_t = \Lambda\theta F_t + \xi H_t = \left(\Lambda\theta + \xi c\right) F_t$. By construction, however, this is impossible. The principal component of the regression residuals in equation (9) is a linear combination of the u_{it}, which are each orthogonal to the order flow principal component.

A second possibility is that the proportionality occurs in the loadings, i.e., that $\xi = c\Lambda\theta$. In this case, $\varphi G_t = \Lambda\theta F_t + \xi H_t = \Lambda\theta\left(F_t + cH_t\right)$, where the term in parentheses will be indistinguishable from a single factor. Examination of the estimated parameters shows that the elements of $\Lambda\theta$ and ξ are in fact moderately positively correlated, so this appears to be the most probable explanation.

d. The explanatory power of common and firm-specific order flow

Having documented a statistical relation between the dominant common factors of returns and order flows, we next turn to a more detailed modeling of this relationship. The return model in equation (3) is a flexible structure that allows the return on the *i*th stock to depend on the full set of order flows. Yet within that set, economic logic suggests an importance ordering. A stock's own order flow x_{it} should be a primary determinant, followed by the common factor of order flow, and then by the other idiosyncratic order flows. This view suggests a linear model such as:

$$r_{it} = \lambda_i^{own} x_{it} + \lambda_i^{common} P_{xt} + \sum_{\substack{j=1 \\ j \neq i}}^{n} \lambda_{i,j}^{other} x_{jt} + u_{it} \tag{11}$$

where P_t^x is the first principal component of order flows, taken as a proxy for the order flow common factor.

One likely source of misspecification (or at least misattribution of effects) arises from the connections between the common component of order flow and the futures market. Two important mechanisms here are index futures arbitrage and positive-feedback trading on index futures returns.

The index arbitrage story says that if macroeconomic news is first incorporated in the futures price, this induces index arbitrage trades in the cash market which then causes adjustments in the individual stock prices (see Kumar and Seppi (1994) and references in footnote 1). The positive feedback story says that trend-chasers, portfolio insurers, and other positive-feedback investors respond to a positive futures return by buying stocks, even after any inter-market arbitrage is gone. To investigate these possibilities, we expand the explanatory variable set to include the contemporaneous futures return over the fifteen-minute interval.

Since the explanatory variables are multicollinear, we compute their incremental explanatory power by adding the variables sequentially. This is tantamount to orthogonalizing the later variables with respect to the variables earlier in the ordering. Since different economic stories suggest different causal ordering, we perform these calculations for two alternative orderings.

Table 4 presents these results. The variable ordering in Panel B corresponds to that given in equation (11), plus the futures return. The explanatory power of own-order flow is a substantial 24.9%, the incremental explanatory power of the order flow principal component P_t^x is lower but still appreciable at 4.3%. The vector of order flows in all other stocks adds little (0.5%). The futures return adds 2.6%. The ordering in Panel C moves the futures return earlier in the specification. Once own order flow and the futures return are included, P_t^x and other order flows contribute little explanatory power (a total of only 0.9%).

In interpreting these results, the question is not whether the futures price is a source of common factor pricing information. Clearly it is. Rather we are asking whether the common factor in orders is a significant independent source of price covariability taking into account the role of the futures market. Comparing the bounds of 0.5 and 4.3 percent on the incremental explanatory power of P_t^x (i.e., given the own order flow) with the bounds of 2.6 and 6.5 percent for $r_t^{futures}$ suggests that, while the order flow common factor may indeed play a role in the price discovery process for common factor (macroeconomic) pricing information, that role is likely dominated by the futures market. The multicollinearity of P_t^x and $r_t^{futures}$ is consistent with both index arbitrage and positive-feedback trading.

As an aside, if the lower bound of 0.5 percent is closer to the truth – that is, if the common factor in orders plays no causal role in pricing and its apparent explanatory power in Panel B is simply "spurious correlation" due to positive feed-back trading – this would have important implications for institutional trading practices. For example, mutual funds and pensions could follow multi-stock portfolio rebalancing strategies which would simply aggregate the solutions from separate optimization problems for each individual stock (see Bertsimas and Lo (1998)) without worrying about cross-stock interactions of orders.

e. Intraday patterns

Our intent in standardizing the data is to isolate stochastic commonalities, the focus of our paper, from the well-known time-of-day effects. Intraday patterns in the commonalities themselves are, however, of interest. To investigate such patterns, we performed a principal components analysis of returns, signed SRD volume, and the log quote slope (described in Section 6) for each fifteen-minute period within the day. The eigenvalues are plotted in Figure 1.

The noise in the plots is largely due to sampling error. The eigenvalues for all three series display a peak around 11:00AM. This peak is attributable to a single day's activity, Thursday, March 31, 1994. Media commentary on the following day note contemporaneous volatility in the stock and bond markets, but do not attribute this to any obvious new information. [7] Nevertheless, the concurrent bond market activity is

---

[7] The Wall Street Journal (April 1, 1994) reports, "STOCKS RECOVERED after a rollercoaster ride that included a sharp morning drop … Long-term bond prices rose in a wild day of trading."

consistent with a commonality related to interest rates. Aside from this irregularity, return commonality is slightly elevated at the beginning and end of trading (a "U" shape). The order flow commonality exhibits an upward trend throughout the day. This might arise from program trading activity. Madhavan and Cushing (2000) find that portfolio return variances are concentrated near the end of the day, which they attribute to institutional trading interest. Orders to unwind futures-linked stock positions on expiration days would generally appear as market-on-open orders, which are not included here. Finally, the commonality in liquidity (discussed further in Section 6) is relatively low and constant over the day.

5. Absolute returns and trading volume

In this section we redo parts of the above analysis for the unsigned (absolute value) counterparts of the signed variables. There are two reasons for this. First, although equations (1) through (4), the motivation for our analysis, involve signed order flows; as a practical matter, the signing of trades by reference to the prevailing quote is subject to error. For example, roughly one third of the trades are typically priced at the quote midpoint and, therefore, cannot be signed. Secondly, our analysis of the unsigned variables is a logical extension of the "volume/volatility" literature for individual securities and indexes. Trading volume and absolute price change are generally found to be positively correlated (see Karpoff (1987) and Gallant, Rossi, and Tauchen (1992)). The present analysis can be viewed as a multifirm extension, asking whether the price/volume correlation extends to *common factors* in prices and volumes.

Table 5 reports statistics for the absolute returns and unsigned volume measures (corresponding to the signed variables in Table 2). The first eigenvalue of absolute returns is smaller than that of signed returns (3.64 vs. 6.32). In contrast, the first eigenvalues for the unsigned volume measures are generally larger than the corresponding values for signed volumes. This may, in part, be a consequence of errors in signing trades.

Table 6 summarizes the canonical correlations for the absolute measures. The general pattern is similar to Table 3. The first canonical variates are moderately highly correlated (0.726) and are also closely related to their corresponding principal components. Thus, the comovements in absolute intraday price changes also have a strong microstructure foundation in absolute SRD volume.

6. Time-varying liquidity measures and their common factors

Leaving orders and returns for now, we turn next to investigate time-variability and cross-sectional commonality in market liquidity. This section looks at aggregate liquidity – including both permanent and transitory price impacts – as measured by posted bid and ask quotes and depths. In the following section we focus specifically on the permanent (or informational) component of liquidity.

Traders can estimate their trading costs ex ante (i.e., prior to submission of an order) on the basis of the displayed quotes and sizes of the quotes. Suppressing the firm subscript *i* for the sake of notational economy, let B_k and A_k denote the per share bid and ask for quote record *k*, and let N_k^B and N_k^A denote the respective number of shares posted at these quotes. Thus, a prospective purchaser knows that, if hers is the first market buy order to arrive, she can buy at least N_k^A shares at the ask price A_k. In our analysis, we employ the following measures of quoted liquidity.

$$\text{Spread}_k = A_k - B_k$$

$$\text{Log Spread}_k = \log\left(A_k / B_k\right)$$

$$\text{Log Size}_k = \log\left(N_k^A\right) + \log\left(N_k^B\right)$$

$$\text{Quote Slope}_k = \left(A_k - B_k\right) \Big/ \left(\log\left(N_k^A\right) + \log\left(N_k^B\right)\right)$$

$$\text{Log Quote Slope}_k = \log\left(A_k / B_k\right) \Big/ \left(\log\left(N_k^A\right) + \log\left(N_k^B\right)\right)$$

The first three of these are standard. The last two, which combine both price and quantity information, may be viewed as summary measures of the quoted liquidity supply curve. As depicted in Figure 2, the quote slope is the slope of the dotted line connecting the bid and ask price/quantity pairs. If more quantity is added at either the bid or ask, or if either quote is moved closer to the other, the line flattens and the market is more liquid. As drawn here, the line joining the quote/quantity pairs for any particular observation need not intersect the vertical axis at the quote midpoint. The log quote slope is defined in a similar fashion, except that log prices are used on the vertical axis.

To align these liquidity measures with the transaction-based data, we examine averages over the fifteen-minute time intervals. The averages are time-weighted, according to the number of seconds that the quote prevailed.

Liquidity can also be measured by the effective spread. The transaction price, p_j, on trade j may be expressed as the quote midpoint, m_j, prevailing immediately prior to trade j plus a disturbance:

$$p_j = m_j + s_j \tag{12}$$

Intuitively, $\left|s_j\right|$ is the effective half-spread, an approximate measure of the trading cost to the active side of the transaction. For each fifteen-minute interval, we compute the volume-weighted average of the $\left|s_j\right|$. The effective spread is smaller than the quoted half-spread for a trade that receives price improvement, and (rarely) larger for a block trade that is crossed outside of the posted quotes. This quantity is easy to compute, but some care must be taken with its economic interpretation. Being the difference between the transaction price and the *pre-trade* quote midpoint, s_j impounds the information inferred from the trade. The purely transient component of the transaction price is the difference between the transaction price and the quote midpoint immediately *post-trade*, i.e., $p_j - \left(m_j + \lambda_j x_j\right) = s_j - \lambda_j x_j$.

Table 7 presents descriptive statistics on the liquidity proxies and the eigenvalues of their standardized covariance matrices. Because spreads and related quantities are likely to be strongly affected by the relative tick size, stocks that split during the year were dropped from the liquidity analysis. The sample here is the remaining Dow stocks (n=24) that did not split in 1994. Therefore, the eigenvalues in Table 7 are constructed from “standardized” covariance matrices of dimension twenty-four.

Among the liquidity measures, our log quote slope measure exhibits the most commonality, with a first eigenvalue of 3.07. Under a normality assumption this is statistically significant, and implies that $3.07/24 \approx 13\%$ of the total variation can be attributed to the first common factor. Commonalities in the other quoted depth and spread measures appear weaker. Furthermore, we find little evidence of economically significant common liquidity factors in the effective spreads calculated from realized transactions. One interpretation of this finding is that liquidity providers at the NYSE appear to offset the weak commonality in quoted liquidity. Taken together, our results suggest that the strong common liquidity shocks suggested by the brief (but intense) market crises (e.g., 1987, 1989) do not exist in “normal” trading regimes.

Chordia, Roll, and Subrahmanyam (2000) also find only weak commonality in liquidity. Their methodology, however, differs from ours in several respects. First, our intervals of observation are shorter (fifteen-minutes vs. daily). Secondly, our liquidity variables are in levels form, whereas Chordia et al. work with *changes*. Generally, variables are differenced when one suspects that they may contain unit-root (i.e., random-walk) components. Spreads and other liquidity measures are usually not so characterized. Overdifferencing (i.e., differencing series that are already stationary) induces autocorrelation in computed residuals. For these reasons, we believe that analysis of levels is more economically meaningful and statistically appropriate. Thirdly, the multivariate techniques used here do not impose any a priori restrictions on the common factor, while Chordia et al. use a market-capitalization weighted average, analogous to the market return, as their factor. Unlike returns in the CAPM, however, there is no theory motivating a capitalization-weighted liquidity factor. A fourth difference is that we also look, in the next section, specifically at the permanent component of aggregate liquidity.

7. Time-varying price impacts

Bid-ask spreads can be decomposed into permanent (informational) and transitory (immediacy-related) components (Glosten (1987)). In this section we allow the price impact of orders on quote midpoints, a standard measure of the informational component of liquidity, to time-vary stochastically so that $r_{it} = \lambda_{it} x_{it} + u_{it}$.

Since λ_{it} is not directly observable, inferences about this parameter are necessarily indirect. We try two simple approaches to circumvent this problem. First, we use the quote-based liquidity measures from Section 6 (and their common factors) as instruments for the λ_{it}. Secondly, we use daily dummy variables as instruments for a common daily liquidity shock to the λ_{it}. Both approaches let us examine time- and common-variation within a linear multivariate statistical framework.

To manage the complexity of the analysis, we employ a basic specification in which the only order variable is the firm's own:

$$r_{it} = \lambda_{it} x_{it} + u_{it} . \tag{13}$$

The time-varying liquidity parameter here is in turn parameterized as a linear function:

$$\lambda_{it} = \alpha_i + \delta_i\, d_t + \gamma_i\, LQS_{it} + h_i\, P_t^{LQS} + \varepsilon_{it} \tag{14}$$

where d_t is a vector of time-of-day dummy variables; LQS_{it} is the standardized loq-quote slope (a liquidity proxy); and P_t^{LQS} is the first principal component (across firms) of the LQS_{it} (a proxy for the common factor of liquidity). For the sake of brevity, just the log quote slope results are reported.[8] Results using other quoted liquidity proxies were similar. When this expression for λ_{it} is substituted in equation (13), we have:

$$r_{it} = \alpha_i\, x_{it} + \delta_i\, d_t x_{it} + \gamma_i\, LQS_{it} x_{it} + h_i\, P_t^{LQS} x_{it} + \left(\varepsilon_{it} x_{it} + u_{it}\right) \tag{15}$$

This is a linear specification in which returns are regressed against order flow x_{it} and order flow/liquidity interaction terms $d_i x_{it}$, $LQS_{it}\, x_{it}$ and $P_t^{LQS} x_{it}$. In light of the concave relation between orders and price changes described in Section 3, we use signed square-root dollar volume for our order flow variable x_{it}.

We are primarily interested in the explanatory power of the liquidity/signed order flow term, $LQS_{it} x_{it}$ and the liquidity common factor $P_t^{LQS} x_{it}$. Time-of-day dummy variables are included because, even though r_{it} and x_{it} are standardized to remove time-of-day effects, this need not pick up time-of-day variation in λ_{it}.

To estimate equation (15) via least squares, the composite residual $\left(\varepsilon_{it} x_{it} + u_{it}\right)$ must be uncorrelated with the explanatory variables. Since x_{it} appears in both the residual and the explanatory variable set, a sufficient condition for least squares consistency is that the ε_{it} are zero-mean and independent of x_{it}, $d_t\, x_{it}$, $LQS_{it} x_{it}$ and $P_t^{LQS} x_{it}$.[9]

---

[8] Intuitively, the log quote slope reflects the permanent price impact, λ_{it}, plus the transitory price impact excluding any price improvement from the floor. Thus, the question we are asking is whether time-variation in quoted liquidity *LQS* is informative about (i.e., related to) time-variation in the permanent/informational component of effective liquidity.

[9] The dummy variable coefficients δ_i in equation (15) are not identified without further restrictions (e.g., that one of them is zero). Present purposes require, however, identifying only the explanatory power associated with a set of variables (for which coefficient identification is not necessary).

Table 8 summarizes the explanatory power of the specification. We use two alternative orderings of the variables that effectively attribute the *joint* explanatory power of the common-factor liquidity term $P_t^{LQS} x_{it}$ and the own-firm liquidity term $LQS_{it} x_{it}$ to one variable or the other. In Panel A, the ordering is x_{it}, $d_t x_{it}$, $P_t^{LQS} x_{it}$ and $LQS_{it} x_{it}$, which attributes any joint power to the common factor term $P_t^{LQS} x_{it}$. In Panel B, the ordering is x_{it}, $d_t x_{it}$, $LQS_{it} x_{it}$ and $P_t^{LQS} x_{it}$, which attributes the joint power to $LQS_{it} x_{it}$. The sum of the two terms, 4.7%, is unaffected by their ordering. The decomposition of the sum, of course, differs between the two panels. The principal component term explains at most 1.0% of the variance, leaving 3.7% explained by the own-firm term (Panel A). When the principal component term is added last, however, its incremental explanatory power is a mere 0.1%. Even taking the higher figure, however, the common covariation in the permanent (informational) component of liquidity appears to be dominated by firm-specific variation.[10]

We extend this analysis in several respects. First, we estimate specifications parallel to those reported in Table 8 where the total signed order flow is replaced by the three separate small/medium/large signed order flow variables described in Section 3. Our intent is to allow for the possibility that the liquidity proxy coefficients are not constant across different order size categories. The full results are not presented for the sake of brevity, but may be summarized as follows. When the size-based order flow measures are used in lieu of the total, the explanatory power of the specification increases somewhat, from 0.292 (in Table 8) to 0.353. However, virtually all of this increase

---

[10] The composite specification does not constrain λ_{it} to be nonnegative. This does not appear to be a significant problem, however. As a check, we computed the fitted values $E\left[\lambda_{it}\right] = \alpha_i + \delta_i\, d_t + \gamma_i\, LQS_{it} + h_i\, P_t^{LQS}$ from equation (14). These were positive for over 99% of the observations.

results from the order flow variables per se, not from the components that impound the liquidity proxies.

The composite specification (15) also admits stochastic variation in λ_{it} via the ε_{it} term. However, since this variation is confounded with other sources of residual variation, it is not econometrically identified. To test for this possibility we use a panel regression to estimate a variant of equation (15) based on a liquidity parameterization:

$$\lambda_{it} = \alpha_i + \delta_i d_t + \gamma_i LQS_{it} + \eta D_t + \varepsilon_{it} \tag{16}$$

where D_t is a vector of date dummy variables (i.e., one for each day). The associated coefficient vector η is the same for all firms. The coefficients η_1 (for day 1), η_2 (for day 2), . . . can be interpreted as estimates of the daily realizations of a random (daily) common liquidity factor. In other words, on a given day *k*, liquidity is partially driven by a realized factor η_k common to all firms.

The new composite specification for returns (corresponding to equation (15)) is:

$$r_{it} = \alpha_i x_{it} + \delta_i d_t x_{it} + \gamma_i LQS_{it} x_{it} + \eta D_t x_{it} + \left(\varepsilon_{it} x_{it} + u_{it}\right) \tag{17}$$

Because the η coefficient vector is common across all firms, this specification is estimated jointly as a panel regression for all twenty-four firms. The results (not reported for brevity) are similar to the regressions in Table 8. The incremental explanatory power of the date dummy terms is sensitive to the ordering of the variables, but is always dominated by the own-firm liquidity term. (The incremental R^2 of the date dummies is at most 0.002, while that of $LQS_{it}x_{it}$ is at least 0.044.) As before, time variation in the informational component of liquidity seems to be largely firm-specific.

Why is it important whether liquidity is correlated or idiosyncratic? Amihud and Mendelson (1986) and Brennan and Subrahmanyam (1996) have argued that predictable differences in liquidity lead to cross-sectional differences in expected returns. A natural extension of this argument is that if liquidity is random and covaries across stocks, then a stock's sensitivity to systematic liquidity randomness could potentially play the role of a priced risk factor. If our results – namely, that randomness in effective liquidity (and its permanent price impact component) is idiosyncratic – are correct, however, then random liquidity risk can be diversified away at the portfolio level and, hence, is unlikely to be a priced source of risk. Thus, any liquidity-linked differences in expected returns are most likely due to predictable differences in the level of liquidity, rather than to variability in liquidity per se.

8. Conclusions

Taking as our starting point a linear microstructure specification in which returns are driven by signed order flows and public news, this paper assesses the extent and role of cross-firm common factors in returns, order flows and market liquidity. We implement the analysis for the thirty Dow stocks in 1994 using time-aggregated trade and quote date over fifteen-minute intervals.

We find that common factors exist in both signed and absolute order flow. These explain part, but not all, of the common variation in signed and absolute returns. This conclusion does not depend on whether the common factors were constructed using principal components or canonical correlations. However, multicollinearity of the S&P 500 futures return with the common factor in order flows complicates the economic interpretation of this result. At a fifteen-minute sampling frequency we cannot distinguish futures-induced positive feedback trading from index arbitrage.

Our findings are less supportive of economically significant common factors in liquidity. After removing time-of-day effects, the strength of any common factors in spreads and related liquidity measures, as judged by the first principal components, is modest. This is confirmed by cross-sectional regressions in which price impact coefficients are projected on various explanatory variables. Own-firm variables dominate the principal component (common factor) and daily liquidity shock estimates. Thus, the systematic liquidity fluctuations visible during market crises such as 1987 and 1998 do not appear to characterize normal trading.

While we do find common factors in the microstructure of equity markets, it is perplexing that these variables do not, on balance, play a more prominent and unambiguous role given the importance traders attach to proximity to other traders. Clearly further study is needed. For example, "event studies" tracing through the impact at a transactional level of macroeconomic news (as opposed to idiosyncratic corporate events) on order flows, individual stock prices and S&P 500 futures prices might be helpful in distinguishing between the various causal explanations for return covariability. In terms of liquidity, it would be interesting to see whether liquidity is more strongly correlated for stocks with lower market capitalizations (i.e., than the Dow stocks used here) or, perhaps, for stocks traded by the same specialist firm.

9. References

Amihud, Y., Mendelson, H., 1986. Asset pricing and the bid-ask spread. Journal of Financial Economics 17, 223-249.

Barclay, M. J., Warner, J. B., 1993. Stealth trading and volatility: Which trades move prices? Journal of Financial Economics 34, 281-305.

Bertsimas, D., Lo, A., 1998. Optimal control of execution costs. Journal of Financial Markets 1, 1-50.

Brennan, M. J., Subrahmanyam, A., 1996. Market microstructure and asset pricing: On the compensation for illiquidity in stock returns. Journal of Financial Economics 41, 441-464.

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

Chan, K., Chan, K. C., Karolyi, G. A., 1991. Intraday volatility in the stock index and stock index futures market. Review of Financial Studies 4, 657-684.

Chan, K., 1993. Imperfect information and cross-autocorrelation among stock prices. Journal of Finance 48, 1211-30.

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

Chowdhry, B., Nanda, V., 1991. Multimarket trading and market liquidity. Review of Financial Studies 4, 483-511.

Desai, A., Nimalendran, M., Venkataraman, S., 1998. Changes in trading activity following stock splits and their effect on volatility and the adverse information component of the bid-ask spread. Journal of Financial Research 21, 159-184.

Edelen, R. M., Warner, J. B., 1999. Why are mutual fund flow and market returns related? Evidence from high-frequency data. Unpublished working paper. Wharton School.

Foster, F. D., Viswanathan, S., 1990. A theory of the interday variations in volume, variance, and trading costs in securities markets. Review of Financial Studies 3, 593-624.

Foster, F. D., Viswanathan, S., 1995a. Can speculative trading explain the volume-volatility relation? Journal of Business and Economic Statistics 13, 379-396.

Foster, F. D., Viswanathan, S., 1995b. Trading costs for target firms around takeovers. In: Hirschey, M. and Marr, M. W. (Eds.), Advances in Financial Economics, JAI Press, Greenwich, Connecticut, 37-57.

Gallant, A. R., Rossi, P. E., Tauchen, G., 1992. Stock prices and volume. Review of Financial Studies 5, 199-242.

Glosten, L. R., 1987. Components of the bid-ask spread and the statistical properties of transaction prices. Journal of Finance 42, 1293-1307.

Harris, L. E., Sofianos, G., Shapiro, J. E., 1994. Program trading and intraday volatility. Review of Financial Studies 7, 653-85.

Hasbrouck, J., 1991. Measuring the information content of stock trades. Journal of Finance 46, 179-207.

Hasbrouck, J., 1996a. Modeling microstructure time series. In: Maddala, G. S. and Rao, C. R. (Eds.), Handbook of Statistics 14: Statistical Methods in Finance, Elsevier

North Holland, Amsterdam, 647-692.

Hasbrouck, J., 1996b. Order characteristics and stock price evolution: An application to program trading. Journal of Financial Economics 41, 129-49.

Huberman, G., Halka, D., 1999. Systematic liquidity. Unpublished working paper. Columbia Business School.

Jain, P. C., Joh, G. H., 1988. The dependence between hourly prices and trading volume. Journal of Financial and Quantitative Analysis 23, 269-283.

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

Keim, D., Madhavan, A., 1996. The upstairs market for block trades: Analysis and measurement of price effects. Review of Financial Studies 9, 1-36.

Koski, J. L., 1996. A microstructure analysis of ex-dividend stock price behavior before and after the 1984 and 1986 tax reform acts. Journal of Business 69, 313-338.

Kumar, P., Seppi, D. J., 1994. Information and index arbitrage. Journal of Business 67, 481-509.

Kyle, A. S., 1984. Market structure, information, futures markets, and price formation. In: Storey, G. G., Schmitz, A., and Sarris, A. H. (Eds.), International Agricultural Trade: Advanced Readings in Price Formation, Market Structure, and Price Instability, Westview Press, Boulder, Colorado and London.

Kyle, A. S., 1985. Continuous auctions and insider trading. Econometrica 53, 1315-1336.

Lee, C. M., Mucklow, B., Ready, M. J., 1993. Spreads, depths and the impact of earnings information: An intraday analysis. Review of Financial Studies 6, 345-374.

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

MacKinlay, A. C., Ramaswamy, K., 1998. Index-futures arbitrage and the behavior of stock index futures prices. Review of Financial Studies 1, 137-158.

Madhavan, A., Cushing, D., 2000. Stock returns and trading at the close. Journal of Financial Markets , forthcoming.

Madhavan, A., Smidt, S., 1991. A Bayesian model of intraday specialist pricing. Journal of Financial Economics 30, 99-134.

Miller, M. H., Muthuswamy, J., Whaley, R. E., 1994. Mean reversion of Standard & Poor's 500 index basis changes: arbitrage-induced or statistical illusion? Journal of Finance 49, 479-513.

Morrison, D. F. 1976. Multivariate Statistical Methods, Second Edition. McGraw-Hill, New York.

New York Stock Exchange. 1996. Fact Book. New York Stock Exchange, New York.

O'Hara, M. 1995. Market Microstructure Theory. Blackwell Publishers, Cambridge, USA.

Seppi, D. J., 1990. Equilibrium block trading and asymmetric information. Journal of Finance 45, 73-94.

Subrahmanyam, A., 1991. A theory of trading in stock index futures. Review of Financial Studies 4, 17-51.

Wood, R. A., McInish, T. H., Ord, J. K., 1985. An investigation of transaction data for NYSE stocks. Journal of Finance 25, 723-739.

Table 1. Descriptive Statistics

The sample is the thirty Dow stocks for all trading days in 1994. The daily return is computed as the first difference of the log end-of-day quote midpoint.

Symbol	Name	Mean Price ($/Share)	Mean Bid-Ask Spread ($/Share)	Average Daily Trades	Std. Dev. of Daily Return x100 (percent)
	Cross-stock mean:	51	0.16	608	1.5
AA	Alcoa	78	0.21	200	1.4
ALD	Allied Signal	44	0.18	191	1.5
AXP	American Express	29	0.15	501	1.6
BA	Boeing	45	0.16	554	1.2
BS	Bethlehem Steel	20	0.15	211	2.4
CAT	Caterpillar	90	0.21	333	1.6
CHV	Chevron	63	0.17	408	1.1
DD	DuPont	57	0.15	447	1.2
DIS	Disney	43	0.15	780	1.4
EK	Eastman Kodak	47	0.15	519	1.8
GE	Gen'l Electric	68	0.16	1,238	1.1
GM	Gen'l Motors	51	0.16	1,226	1.8
GT	Goodyear	38	0.19	260	1.5
IBM	IBM	63	0.16	1,251	1.7
IP	Int'l Paper	72	0.19	232	1.3
JPM	JP Morgan	63	0.17	362	1.1
KO	Coca Cola	45	0.14	902	1.1
MCD	McDonalds	43	0.15	737	1.3
MMM	MMM	67	0.18	363	1.1
MO	Phillip Morris	56	0.15	1,200	1.4
MRK	Merck	33	0.14	1,707	1.4
PG	Proctor Gamble	58	0.16	503	1.3
S	Sears	48	0.16	425	1.6
T	ATT	53	0.14	1,152	1.0
TX	Texaco	63	0.16	342	0.9
UK	Union Carbide	28	0.15	382	2.0
UTX	United Tech	64	0.19	159	1.3
WX	Westinghouse	13	0.13	548	1.7
XON	Exxon	61	0.14	698	1.0
Z	Woolworth	18	0.15	400	2.4

Table 2. Returns and Signed Volume Measures

The sample is the thirty Dow stocks for all trading days in 1994. Trades are signed by reference to the quote midpoint prevailing at the time of the trade. The signed volume measures are cumulated over fifteen-minute intervals. Small trades are 2,000 shares or less; medium trades are 2,001-10,000 shares; large trades are >10,000 shares. The means and standard deviations are for the raw data, pooled across firms and time.

Variable		Mean (not standardized by time-of-day)	Std. Dev. (not standardized by time-of-day)	Eigenvalues		
				First	Second	Third
Return		-0.001	0.266	6.32	1.04	1.00
Signed trades	Total	0.635	8.784	3.36	1.44	1.27
	Small	0.521	7.831	2.75	1.61	1.36
	Medium	0.071	1.975	3.60	1.14	1.07
	Large	0.043	0.771	1.67	1.09	1.08
Signed share volume (100-shr lots)	Total	16.259	334.253	2.36	1.08	1.07
	Small	0.542	42.607	3.92	1.15	1.11
	Medium	5.556	103.557	3.03	1.12	1.09
	Large	10.161	290.502	1.48	1.11	1.10
Signed dollar volume ($10,000)	Total	9.123	153.406	2.38	1.09	1.07
	Small	0.057	21.103	3.90	1.11	1.11
	Medium	3.356	52.483	2.99	1.12	1.08
	Large	5.711	130.690	1.48	1.11	1.10
Signed square root of dollar volume $\left(\times 10^{-2}\right)$	Total	1.206	19.798	4.06	1.08	1.05
	Small	0.229	10.517	3.51	1.32	1.20
	Medium	0.489	9.545	3.36	1.13	1.08
	Large	0.488	8.278	1.64	1.10	1.08

Table 3. Canonical and Principal Factors of Returns and Signed Square-root Dollar Volume

The sample is the thirty Dow stocks for all trading days in 1994. Panel A reports the correlations between the return and order flow (signed SRD volume) canonical variates. Panel B reports a canonical redundancy analysis: the proportion and cumulative proportion of total variation in returns and order flow that is explained by the canonical variates and (for comparison purposes) the principal components.

Panel A: Canonical correlations

First	0.829
Second	0.578
Third	0.556

Panel B: Canonical redundancy analysis

Total variation in returns explained by:

Return principal components			Return canonical variates			Signed square-root dollar volume canonical variates		
	Prop.	Cum.		Prop.	Cum.		Prop.	Cum.
First	0.210	0.210	First	0.206	0.206	First	0.142	0.142
Second	0.034	0.244	Second	0.028	0.234	Second	0.009	0.151
Third	0.033	0.277	Third	0.029	0.262	Third	0.009	0.160

Total variation in signed square-root dollar volumes explained by:

Signed square-root dollar volume principal components			Signed square-root dollar volume canonical variates			Return canonical variates		
	Prop.	Cum.		Prop.	Cum.		Prop.	Cum.
First	0.135	0.135	First	0.132	0.132	First	0.091	0.091
Second	0.036	0.171	Second	0.031	0.163	Second	0.010	0.101
Third	0.035	0.205	Third	0.031	0.193	Third	0.009	0.111

Table 4. Explained return variation.

The sample is the thirty Dow stocks for all trading days in 1994. The proportional and cumulative measures of incremental explanatory power are analogous to R^2s. They are derived from a set of regression specifications in which the dependent variables are the standardized fifteen-minute returns r_{it} for $i = 1,\ldots,30$ firms. Entries are of the form $\left(\sum_{i=1}^{30}\text{Explained variance for stock } i\right)\Big/\left(\sum_{i=1}^{30}Var\left(r_{it}\right)\right)$. Regression variables are defined as follows: x_{it} is a stock's own order flow; x_{jt} (for j=1,…,30; $j\neq i$) are the order flows for other stocks; P_t^x is the first principal component of order flows; P_t^u is the first principal component of the regression residuals; $r_t^{futures}$ is the return on the near-maturity S&P 500 futures contract.

	Variable Ordering	Proportional	Cumulative
Panel A:	Common order flow, P_t^x	0.146	0.146
	Residual common factor, P_t^u	0.071	0.217
Panel B:	Own order flow, x_{it}	0.249	0.249
	Common order flow, P_t^x	0.043	0.292
	Other order flow, x_{jt}, for $j = 1,\ldots,30$; $j \neq i$	0.005	0.296
	$r_t^{futures}$	0.026	0.322
Panel C:	Own order flow, x_{it}	0.249	0.249
	$r_t^{futures}$	0.065	0.313
	Common order flow, P_t^x	0.005	0.318
	Other order flow, x_{jt}, for $j \neq i$	0.004	0.322
Panel D:	$r_t^{futures}$	0.134	0.134
	Own order flow, x_{it}	0.179	0.313
	Common order flow, P_t^x	0.005	0.318
	Other order flow, x_{jt}, for $j \neq i$	0.004	0.322

Table 5. Absolute Returns and Unsigned Volume

The sample is the thirty Dow stocks for all trading days in 1994. Data series are aggregated over 15-minute intraday intervals. The means and standard deviations are for the raw series pooled across firms and time. Eigenvalues are based on series standardized to remove time-of-day effects. The return is the log quote-midpoint return (using only NYSE quotes); trades, share volume, dollar volume and (cumulative) square root dollar volume are computed for all trades. Small trades are 2,000 shares or less; medium trades are 2,001-10,000 shares; large trades are >10,000 shares.

Variable		Mean (not standardized by time-of-day)	Std. Dev. (not standardized by time-of-day)	Eigenvalues		
				First	Second	Third
|Return|		0.161	0.211	3.64	1.10	1.09
Number of Trades	Total	22.581	14.713	5.00	1.78	1.44
	Small	19.344	12.672	4.61	1.82	1.47
	Medium	2.626	2.737	4.43	1.21	1.19
	Large	0.610	0.968	1.78	1.29	1.14
Share Volume (100-shr lots)	Total	388.823	501.121	3.41	1.29	1.17
	Small	87.706	64.724	4.97	1.39	1.25
	Medium	127.992	143.251	3.81	1.20	1.18
	Large	173.124	421.857	2.21	1.22	1.16
Dollar Volume ($10,000)	Total	188.941	233.844	3.60	1.23	1.15
	Small	43.528	32.304	5.55	1.51	1.14
	Medium	65.417	73.138	3.94	1.27	1.18
	Large	79.996	194.405	2.23	1.18	1.14
Square root of dollar volume $\left(\times 100^{-2}\right)$	Total	44.164	32.036	5.53	1.36	1.25
	Small	25.169	16.783	5.35	1.44	1.21
	Medium	12.564	13.311	4.27	1.22	1.15
	Large	6.431	10.535	2.16	1.24	1.13

Table 6. Canonical and Principal Factors of Absolute Returns and Unsigned Square-root Dollar Volume

The sample is the thirty Dow stocks for all trading days in 1994. Data series are aggregated over 15-minute intraday intervals and then standardized to remove time of day effects.

Panel A: Canonical correlations

First	0.726
Second	0.525
Third	0.502

Panel B: Canonical redundancy analysis
Total variation in absolute returns explained by:

Absolute return principal components			Absolute return canonical variates			Square-root dollar volume canonical variates		
	Prop.	Cum.		Prop.	Cum.		Prop.	Cum.
First	0.120	0.120	First	0.116	0.116	First	0.061	0.061
Second	0.037	0.157	Second	0.036	0.152	Second	0.010	0.071
Third	0.036	0.193	Third	0.032	0.184	Third	0.008	0.079

Total variation in square-root dollar volume explained by:

Square-root dollar volume principal components			Square-root dollar volume canonical variates			Absolute return canonical variates		
	Prop.	Cum.		Prop.	Cum.		Prop.	Cum.
First	0.171	0.171	First	0.159	0.159	First	0.084	0.084
Second	0.046	0.217	Second	0.034	0.193	Second	0.009	0.093
Third	0.042	0.259	Third	0.032	0.225	Third	0.008	0.101

Table 7. Liquidity Measures

The sample is the twenty-four Dow stocks that did not split during 1994 for all trading days in 1994. The bid and ask for quote record k are denoted by B_k and A_k (dollars per share); N_k^B and N_k^A are the numbers of shares (in 100-share round lots) posted at the bid and ask. The spread is $A_k - B_k$; the log spread is $\log(A_k/B_k)$; the log size is $\log(N_k^A)+\log(N_k^B)$; the quote slope is $(A_k - B_k)\big/\left(\log(N_k^A)+\log(N_k^B)\right)$; the log quote slope is $\log(A_k/B_k)\big/\left(\log(N_k^A)+\log(N_k^B)\right)$. For all of the preceding, we employ time-weighted averages over the fifteen-minute intervals. For trade j at price p_j, the effective spread is defined as $\left|p_j - m_j\right|$ where m_j is the prevailing quote midpoint. These are averaged by trade size over the interval. Small trades are 2,000 shares or less; medium trades are 2,001-10,000 shares; large trades are >10,000 shares. The means and standard deviations are for the raw data pooled across firms and time.

Variable		Mean (not standardized by time-of-day)	Std. Dev. (not standardized by time-of-day)	Eigenvalues		
				First	Second	Third
Spread		0.15746	0.03830	2.00	1.10	1.07
Log Spread		0.00394	0.00224	2.84	1.67	1.32
Log Size		9.36776	1.91521	2.20	1.40	1.36
Quote Slope		0.01811	0.00844	2.58	1.27	1.15
Log Quote Slope		0.00043	0.00024	3.07	1.55	1.16
Effective Spread	Total	0.05932	0.01967	1.13	1.12	1.10
	Small	0.05726	0.01657	1.14	1.13	1.09
	Medium	0.05828	0.02416	1.17	1.14	1.11
	Large	0.06077	0.03129	1.38	1.32	1.29

Table 8. Price Impact Regressions

The sample is the twenty-four Dow stocks that did not split during 1994 for all trading days in 1994. For each stock the estimated specification is $r_{it} = \alpha_i x_{it} + \delta_i d_t x_{it} + \gamma_i LQS_{it} x_{it} + h_i P_t^{LQS} x_{it} + e_{it}$ where r_{it} is the standardized log quote midpoint return for firm i, x_{it} is the standardized order flow (signed square root dollar volume), d_t is a vector of time-of-day dummies; LQS_{it} is the (standardized) log quote slope (a liquidity proxy); and P_t^{LQS} is the first principal component of the loq quote slope (across firms).

	Variable Ordering	Proportion	Cumulative
Panel A:	X_{it}	0.240	0.240
	$d_t x_{it}$	0.005	0.245
	$P_t^{LQS} x_{it}$	0.010	0.255
	$LQS_{it} x_{it}$	0.037	0.292
Panel B:	x_{it}	0.240	0.240
	$d_t x_{it}$	0.005	0.245
	$LQS_{it} x_{it}$	0.046	0.291
	$P_t^{LQS} x_{it}$	0.001	0.292

Figure 1. Intraday Commonality

The sample is the thirty Dow stocks in 1994, over fifteen-minute intraday intervals. Return is the log quote-midpoint return; order flow is signed square-root dollar volume; liquidity is the log quote slope. Figure depicts the first eigenvalue (an indication of the strength of the first principal component) for each of the three series by time of day.

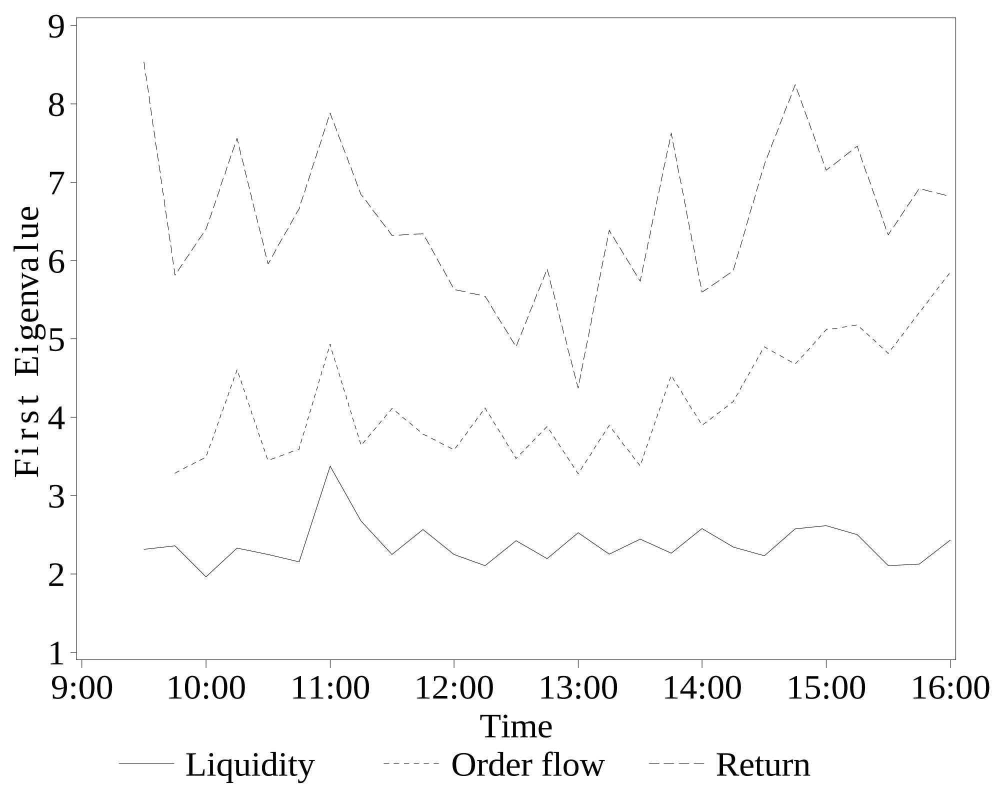

Figure 2. The Quote Slope

For quote record k, the bid and ask prices are B_k and A_k. N_k^B and N_k^A are the numbers of shares sought at the bid and available at the ask. The quote slope is the slope of the heavy dotted line in the figure.