

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

December 31, 1998

Comments welcome

^{*} We thank Dennis Epple 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 half 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 relatively small, however.

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.

To date, 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 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 should closely 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.

This paper is organized as follows. Section 2 establishes the motivation for the study and reviews relevant earlier work. A simple microstructure specification that serves as the basis for the empirical analysis is discussed in Section 3. Section 4 describes the data. The joint statistical properties of returns and signed order flows are analyzed in Section 5; those of absolute returns and unsigned order flows, in Section 6. Section 7 explores variation and covariation in liquidity proxies derived from quote data. Section 8 attempts to relate variation in these proxies to the price impacts of trades. A brief summary concludes the paper in Section 9.

2. Economic Framework

a. Model

We start with a simple and conventional linear model of transaction-time market dynamics for the i th firm:

$$r_{i,\tau} = m_{i,\tau} - m_{i,\tau-1} = \lambda_{i,\tau} x_{i,\tau} + u_{i,\tau} \quad (1)$$

where $m_{i,\tau}$ is the quote midpoint subsequent to trade τ , with associated first difference $r_{i,\tau}$ and $x_{i,\tau}$ is the signed volume of trade τ . Equation (1) partitions the quote midpoint change into a component, $u_{i,\tau}$, due to public information unrelated to trade τ and a component, $\lambda_{i,\tau} x_{i,\tau}$, due to the private information inferred from trade τ . The impact coefficient, $\lambda_{i,\tau}$,

is potentially time-varying. These components are permanent, in the sense of being permanently impounded into the stock price.

If we adopt the timing convention that public information in $u_{i,\tau}$ arrives after trade τ , then the quote midpoint prevailing prior to trade τ is $m_{i,\tau-1}$. The transaction price, $p_{i,\tau}$, of trade τ is given by the prevailing midpoint plus a disturbance:

$$p_{i,\tau} = m_{i,\tau-1} + s_{i,\tau} \quad (2)$$

Intuitively, $|s_{i,\tau}|$ is the effective half-spread, an approximate measure of the trading cost to the active side of the transaction. This quantity is easy to compute, but some care must be taken with its economic interpretation. It is the difference between the transaction price and the *pre-trade* quote midpoint, and so impounds the information inferred from the trade. The difference between the transaction price and the quote midpoint immediately *post-trade* is $p_{i,\tau} - (m_{i,\tau-1} + \lambda_{i,\tau} x_{i,\tau}) = s_{i,\tau} - \lambda_{i,\tau} x_{i,\tau}$. It is this last expression that constitutes the transient component of the transaction price.

We ask several broad empirical questions about this model. First, in equation (1) do common factors exist in stocks' order flows, $x_{i,\tau}$, due either to information or liquidity shocks? And if so, to what extent can commonality in order flows explain any commonality in returns, $r_{i,\tau}$? Second, does market depth (as measured by $\lambda_{i,\tau}$) vary stochastically over time. And if so, is this time variation systematic across stocks? Similarly, we also ask whether there are commonalities in the effective half-spreads $|s_{i,\tau}|$ in equation (2).

In a portfolio context, comovements in stock price changes, $r_{i,\tau}$, are generally held to arise from common factors in future cash flows and discount rates (or in revisions of expectations of these quantities). Most microstructure models distinguish between trade (signed order flow) and non-trade components of returns. Order flows, $x_{i,\tau}$, are usually viewed as comprising liquidity- and information-motivated components. Either of these might reasonably be hypothesized to derive (partly) from influences common across firms.

Dynamic hedging strategies, tax and calendar-related effects and simple momentum trading could all plausibly lead to correlated liquidity trading. Private information about common, macroeconomic variables would cause correlated informed trading. Correlation in the non-trade permanent components of returns, $u_{i,\tau}$, could arise from announcements with common effects (e.g., monetary policy).

Informed order flow is not separately observed by uninformed market participants, including the quote setters. The expected intensity of informed order flow is, however, reflected in the impact coefficients $\lambda_{i,\tau}$. We hypothesize that time variation and common factors may exist in these coefficients. Fluctuations in the supply of liquidity may, similarly, lead to time variation in $|s_{i,\tau}|$.

b. Previous Literature

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 (1997) show that certain assumptions about portfolio rebalancing and liquidation imply a factor structure for trading volume.

The approach in our paper, in contrast to Lo and Wang (1997), is more descriptive and statistical, and assigns no distinctive role to factor portfolios. This may be justified by noting that modern microstructure theory ascribes a prominent informational role to trading. Cross-firm commonalities in informationally-motivated order flows need not necessarily parallel the factor structure in returns.

Chordia, Roll, and Subrahmanyam (1998) explore cross-sectional interactions in liquidity measures using quote data. They assign a particular 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 Dow firms, Chordia et al use a cross-section of roughly one thousand stocks.

Another line of research on cross-sectional 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 portfolio which is traded in different locations. In contrast, the Dow stocks in our study, while closely related (i.e., given their common price factors), are far from perfect substitutes. In some sense, we are trying to quantify how close is “closely related” in terms of common factors in prices, volume and liquidity.

Two forms of time-varying liquidity have been previously documented. First, Wood, McInish, and Ord (1985), Jain and Joh (1988) and Foster and Viswanathan (1990) (among others) 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 Venkataramaan (1998)), take-over announcements (Foster and Viswanathan (1994)) and other identifiable events. We, in contrast, are interested in *stochastic* (rather than predictable) 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 (1995) who 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.

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).

3. Methodology

The primary purpose of this paper is to characterize the properties of the model in equations (1) and (2) across firms. To achieve this, we use panel specifications across firms and across time periods. In particular, we aggregate trades within fifteen minute intervals to get a time-aggregated version of equation (1):

$$r_{i,t} = m_{i,t} - m_{i,t-1} = \lambda_{i,t} x_{i,t} + u_{i,t} \approx \sum_{\tau} \lambda_{i,\tau} x_{i,\tau} + u_{i,t} \quad (3)$$

where $m_{i,t}$ is the quote midpoint prevailing at the close of interval t , $x_{i,t}$ is the sum $\sum_{\tau} x_{i,\tau}$ of the trades in interval t , and the market depth parameter $\lambda_{i,\tau}$ for each trade τ is assumed to be a constant, $\lambda_{i,t}$, *within* each interval.

We employ principal component and canonical correlation techniques to investigate commonalities in returns and order flow. These approaches assume latent random components. Principal components analysis is used to determine the linear compounds with maximal power in explaining the total variance of a set of variables (e.g., returns or order flows) across firms. Canonical correlation analysis determines the linear compounds with maximal power in explaining the covariances between order flows and returns.

Unlike order flows and returns, the market impact parameters are linear projection (regression) coefficients that are not directly observable. Time variation and common factors in the $\lambda_{i,t}$ are characterized here using regressions in which the depth coefficients are projected onto a set of predetermined variables.

To obtain a panel series for the effective half-spreads, we construct within-period averages, $\bar{s}_{i,t}$, of the $|s_{i,\tau}|$ in equation (2) for each trade within the fifteen minute interval. The panel series are then studied using principal components.

4. Data

a. Overview

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.

It is necessary to establish a common time-frame for the data series. We use 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. A 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 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. An interval significantly shorter than fifteen minutes would exacerbate these omitted dynamics.

b. Constructed data series

Three groups of variables are constructed from the TAQ data: returns, order flows and liquidity measures.

Returns. We use the log quote midpoint return, defined as

$$r_{i,t} = \log(m_{i,t}/m_{i,t-1}) \quad (4)$$

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

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

Because studies of short-term price-trade dynamics suggest that the trade impact is concave in size (see Madhavan and Smidt (1991) and Hasbrouck (1991)), we also examine the cumulative square-root of the dollar volume (“SRD volume”) $\sum_{\tau=1}^{n_{i,t}} \sqrt{p_{i,\tau} v_{i,\tau}}$. We also explore size effects by constructing order flow measures based on small ($\leq 2,000$ shares), medium (2,001-10,000 shares) and large ($> 10,000$ shares) trades.

Signed order flow measures, $\text{sign}(v_{i,\tau})v_{i,\tau}$, corresponding to the above are derived by letting the imputed sign (i.e., direction) of a trade be the sign of the difference $p_{i,\tau} - m_{i,t-1}$. 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).

Liquidity measures. A specification such as equation (3) with observations aggregated over fifteen-minute intervals is a convenient way to estimate the “permanent” component of market depth. But the specification will not be informative about the transient component in the individual transaction prices. One of the measures we use to study this is the effective half-spread $|s_{i,\tau}|$ averaged (weighted by dollar volume of trade) across

trades in fifteen-minute intervals to get the (previously described) $\bar{s}_{i,t}$. This reflects (in a general sense) the cost paid by the trade initiator.

The effective spread is an ex post measure of liquidity. A trader may estimate liquidity ex ante (i.e. prior to submission of an order) on the basis on the displayed quotes and the sizes of the quotes. Suppressing the firm subscript i for the sake of notational economy, let B_τ and A_τ denote the per share bid and ask at trade time τ , and let N_τ^B and N_τ^A denote the posted size of these quotes. Thus, a prospective purchaser knows that, if hers is the first market buy order to arrive, she can buy at least N_τ^A shares at the ask price A_τ . We employ the following measures:

$$\text{Spread}_\tau = A_\tau - B_\tau$$

$$\text{Log Spread}_\tau = \log(A_\tau/B_\tau)$$

$$\text{Log Size}_\tau = \log(N_\tau^A) + \log(N_\tau^B)$$

$$\text{Quote Slope}_\tau = (A_\tau - B_\tau) / (\log(N_\tau^A) + \log(N_\tau^B))$$

$$\text{Log Quote Slope}_\tau = \log(A_\tau/B_\tau) / (\log(N_\tau^A) + \log(N_\tau^B))$$

The first three of these are standard. The last two combine both price and quantity information. Intuitively, they may be viewed as summary measures of the liquidity supply curve. As depicted in Figure 1, 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 will flatten. As drawn in the figure, the line joining the quote/quantity pairs for any particular observation need not pass through the “origin” $(0, m_\tau)$. The log quote slope is defined in a similar fashion, except that log prices are used on the vertical axis.

Standardizations. To differentiate stochastic sources of common time-variation from deterministic sources, series are standardized to remove time-of-day effects. 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 $z_{i,d,k}^* = (z_{i,d,k} - \mu_{i,k}) / \sigma_{i,k}$ where $\mu_{i,k}$ and $\sigma_{i,k}$ are the mean and standard deviation for firm i and subperiod k , estimated across days.

5. Returns and signed volume measures

The central relationship in market microstructure is that between returns and order flows (i.e., signed as buyer- or seller-initiated) as in equation (1) or (3). This section describes the commonalities in returns and order flows, considered both separately and jointly.

a. Principal components

Table 2 summarizes features of returns and signed measures of trading activity. The reported means and standard deviations are computed from the unstandardized series to indicate the scale and variability of the raw data. For each standardized variable, we estimate the correlation matrix $[\rho_{i,j}]$ where $\rho_{i,j}$ is the correlation $\text{Corr}(z_{i,t}^*, z_{j,t}^*)$, i.e., the correlation between standardized variates for firms i and j . Due to the standardization, this correlation matrix is also the covariance matrix.

For a given variable, the importance of common factors across firms is measured in a principal components framework. The first principal component of the standardized random vector $Z_t^* = (z_{1,t}^*, z_{2,t}^*, \dots, z_{30,t}^*)$ is the linear combination aZ_t^* for a coefficient vector a that maximizes $\text{Var}(aZ_t^*)$ subject to normalization restrictions. This variance is the largest eigenvalue of the correlation matrix. The second principal component is the variance-maximizing linear combination, subject to the requirement that it is uncorrelated with the first principal component, and so on. Since the eigenvalues sum to the trace of the correlation matrix (which equals the dimension of Z_t^* , or thirty here), the extraction of the eigenvalues partitions the total standardized variance into uncorrelated components of decreasing importance. For example, if the standardized variables were perfectly positively correlated (i.e., identical), the largest eigenvalue would be thirty, and the

remaining eigenvalues would be zero. If all standardized variables were uncorrelated, each of the thirty eigenvalues would be one.

The first eigenvalue of (standardized) returns in Table 2 is 6.32.² This implies that $6.32/30 = 21\%$ of the total variation in 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 of the signed volume measures also suggest commonality. This is most evident for trades in the small ($\leq 2,000$ shares) and medium (2,001-10,000 shares) size classes, and less so for the large block trades ($>10,000$ shares).

b. Canonical correlations

With two sets of variables, such as returns and signed order flow for all firms, common factors may be constructed using canonical correlation analysis. Denoting the two sets of variables as Z_t^* and Y_t^* , the first canonical components are the linear combinations aZ_t^* and bY_t^* such that the correlation $\text{Corr}(aZ_t^*, bY_t^*)$ is maximized.

From among the various trading volume measures, we restrict our analysis to signed square-root dollar (SRD) volume, for brevity. We chose this variable simply because, among all of the signed volume measures, it is the one that is generally most highly correlated with returns at the individual firm level. Table 3 reports correlations of

² If the data are multivariate normal, then eigenvalues of the sample covariance matrix have a known asymptotic distribution (Morrison (1976)): with n observations on $z \sim N(\mu, \Sigma)$, $\sqrt{n}(l_i - \psi_i) \stackrel{asy.dist.}{\sim} N(0, 2\psi_i^2)$ where ψ_i and l_i are the population and sample values of the i th eigenvalue. In this application we are working with an estimated correlation matrix and normality is doubtful. Nevertheless, direct application of the asymptotic approximation yields an estimated standard error for the first eigenvalue of $\sqrt{2(6.32)^2/6,000} = 0.12$.

the canonical and principal components. For each set of variables (i.e., returns and signed square-root dollar volume), “ C ” denotes a canonical factor and “ P ” denotes a principal component.

The correlation structure of the factors is revealing in several respects. Note first that the first principal component (P_1) of returns is highly correlated (0.82) with the first principal component for signed SRD volume. Recall that the first principal component maximizes explained variance only within a single set of variables (i.e., just returns or just signed volume). This does not imply anything about the correlation between principal components constructed for different variable sets. Maximizing the correlation between linear compounds is, however, the objective of a canonical correlation procedure. In a sense, we may take the 0.82 as a starting point. The canonical correlation analysis maximizes this correlation at 0.83. The canonical factors are also highly correlated with the principal components. For returns, the correlation between the first canonical factor and first principal component is 0.99; for signed SRD volume, 0.98. This suggests that the commonality within returns or signed SRD volume is essentially identical to the commonality between these variables.

The second canonical factors are also highly correlated (0.58), suggesting that the first canonical factors do not capture all of the commonalities among the two sets of variables. The second canonical factors are not strongly correlated with the second principal components (as the first), however.

The degree of residual commonality may also be illustrated by estimating the following (panel) regression

$$r_{i,t}^* = \alpha_i + \beta_i C_{1,t}^{SSRDV} + e_{i,t}, \text{ for } i = 1, \dots, 30 \quad (5)$$

where $C_{1,t}^{SSRDV}$ is the first canonical factor for the signed square-root dollar volume series. The residual commonality may be quantified by performing a principal component analysis on the residual covariance matrix. This is formally known as a partial principal

components analysis. From Table 2, the first eigenvalue of the standardized return covariance matrix is 6.32. The first eigenvalue of the residual covariance matrix from regression (5) is 2.13, implying that roughly two-thirds of the first return component is explained by the first canonical SSRDV factor. This suggests that the factor structure of short-horizon returns has a strong microstructure foundation in the factor structure of order flows.

6. Absolute returns and trading volume

Although the motivating specification (1) for our analysis involves signed order flows and returns, it is also worthwhile to consider their unsigned (absolute value) counterparts. There are two reasons for this. First, signing a trade by reference to the prevailing quote midpoint is a procedure subject to uncertainties and errors. For example, roughly one third of the trades are typically priced at the quote midpoint and, therefore, cannot be signed. Secondly, our analysis is a logical extension of the “price/volume” literature for individual securities and indexes. Trading volume (variously measured) and absolute price change are generally found to be positively correlated (see Karpoff (1987) and Gallant, Rossi, and Tauchen (1992), among others). The present analysis can be viewed as a multifirm extension. In particular, it asks whether the price/volume correlation extends to *common factors* in prices and volumes.

Table 4 reports statistics for the absolute returns and unsigned volume measures (corresponding to the signed variables in Table 2). Most importantly, the first eigenvalue of absolute returns is smaller than that of signed returns (3.64 vs. 6.32), suggesting a weaker common factor. This finding does not carry through, however, to the volume measures, for which the first eigenvalues are generally larger than the corresponding values for signed volumes. This may, in part, be a consequence of errors in signing trades.

Table 5 summarizes the canonical correlations for the absolute measures (and corresponds to the signed measures given in Table 3). The general pattern is similar to

that of Table 3. The first canonical factors are moderately highly correlated (0.73) and are also correlated with their corresponding principal components. Thus, the comovements in absolute intraday price changes also have a strong microstructure foundation in absolute SRD volume.

7. Liquidity measures

Table 6 presents descriptive statistics on the liquidity proxies and the eigenvalues of their correlation 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 twenty-four Dow firms that did not have stock splits in 1994.

The eigenvalues in Table 6 are, therefore, constructed from correlation matrices of dimension twenty-four. Even allowing for this, they are generally smaller than the corresponding magnitudes for returns or order flows (cf. Tables 2 and 4). This suggests that common variation in liquidity, at least over fifteen minute aggregation intervals is not large and, thus, that the determinants of variation in liquidity are largely (though not exclusively) firm-specific.

8. Price impacts

This section attempts to assess the variation and common covariation in the market depth parameters, $\lambda_{i,t}$, in equation (3). Up to this point we have characterized common co-variation in returns, order flows, etc. using general multivariate techniques. This approach does not work for the depth parameters because they are not directly observed. We therefore proceed by positing specific determinants.

Natural candidates for these determinants are the liquidity proxies and summary measures (e.g., spreads, depths, etc.) from the last section. These may be presumed to impound (in part) ex ante the expected cost of asymmetric information, which is, in turn,

linked by economic theory to the price impact of a trade (the depth coefficient). With this in mind we posit a linear specification for the liquidity parameter in equation (3):

$$\lambda_{i,t} = f(d_t, L_{i,t}^*, P_t^L) + \varepsilon_{i,t} \quad (6)$$

where f denotes a linear function, d_t is a vector of time-of-day dummy variables, $L_{i,t}^*$ is one of the liquidity proxies from the last section (standardized by firm and time-of-day), P_t^L is the first principal component (across firms) of the chosen standardized liquidity proxy and $\varepsilon_{i,t}$ is the residual. P_t^L is included as a proxy for the common liquidity factor. When substituted into equation (3), we have:

$$\begin{aligned} r_{i,t}^* &= \lambda_{i,t} x_{i,t}^* + u_{i,t} = [f(d_t, L_{i,t}^*, P_t^L) + \varepsilon_{i,t}] x_{i,t}^* + u_{i,t} \\ &= \alpha_i x_{i,t}^* + \delta_i d_t x_{i,t}^* + \gamma_i L_{i,t}^* x_{i,t}^* + h_i P_t^L x_{i,t}^* + (\varepsilon_{i,t} x_{i,t}^* + u_{i,t}) \end{aligned} \quad (7)$$

where the asterisks denote variables standardized by time-of-day. This is a linear specification in which returns are regressed against order flow, $x_{i,t}^*$, and order flow/liquidity interaction terms $d_t x_{i,t}^*$, $L_{i,t}^* x_{i,t}^*$ and $P_t^L x_{i,t}^*$.

We are primarily interested in the explanatory power of the liquidity proxy/signed order flow term, $L_{i,t}^* x_{i,t}^*$ and the associated common factor proxy $P_t^L x_{i,t}^*$. Time-of-day dummy variables are included because, even though $r_{i,t}^*$ and $x_{i,t}^*$ are standardized to remove time-of-day effects, this need not pick up time-of-day variation in $\lambda_{i,t}$.

To estimate equation (7) via least squares, the composite residual $(\varepsilon_{i,t} x_{i,t}^* + u_{i,t})$ must be uncorrelated with the explanatory variables. Since $x_{i,t}^*$ appears in both the residual and the explanatory variable set, there is an obvious potential for dependence. A sufficient condition for least squares consistency is that that the $\varepsilon_{i,t}$ are zero-mean and independent of $x_{i,t}^*$, $d_t x_{i,t}^*$, $L_{i,t}^* x_{i,t}^*$ and $P_t^L x_{i,t}^*$.³

³ The dummy variable coefficients δ_i in equation (7) are not identified without further restrictions (e.g., that one of them is zero). Present purposes require, however, identifying

We estimated specification (7) separately for each firm. Table 7 reports the cross-firm means of incremental R^2 contributions. We use two alternative orderings of the variables that effectively attribute the *joint* explanatory power of the common-factor liquidity term $P_t^L x_{i,t}^*$ and the own-firm liquidity term $L_{i,t}^* x_{i,t}^*$ to one variable or the other. In Panel A, the ordering is $x_{i,t}^*$, $d_t x_{i,t}^*$, $P_t^L x_{i,t}^*$ and $L_{i,t}^* x_{i,t}^*$, which attributes any joint power to the common factor term $P_t^L x_{i,t}^*$. In Panel B, the ordering is $x_{i,t}^*$, $d_t x_{i,t}^*$, $L_{i,t}^* x_{i,t}^*$ and $P_t^L x_{i,t}^*$, which attributes the joint power to $L_{i,t}^* x_{i,t}^*$. The sum of the two terms is unaffected by the ordering (up to rounding error), and is on the order of 3–5%. The decomposition of the sum, of course, differs between the two panels. For the log quote slope liquidity proxy, the principal component term explains at most 0.9% 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, it appears that common covariation in liquidity is dominated by firm-specific variation. Thus, the strong common liquidity factors suggested by the brief (but intense) periods of market crisis (e.g., 1987, 1989, etc.) do not appear to exist in “normal” trading regimes.

The composite specification (7) admits stochastic variation in $\lambda_{i,t}$ (via the $\varepsilon_{i,t}$ term). But this variation is confounded with other sources of residual variation, and so is not econometrically identified. To test for this possibility we use a panel regression to estimate a variant of equation (6):

$$\lambda_{i,t} = \alpha_i + \delta_i d_t + \gamma_i L_{i,t}^* + \eta D_t + \varepsilon_{i,t} \quad (8)$$

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

only the explanatory power associated with a set of variables (for which coefficient identification is not necessary).

2), . . . can be interpreted here as estimates of the daily realizations of a random (daily) liquidity factor. In other words, on a given day k , liquidity is partially driven by a realized factor η_k common across all firms.

The new composite specification for returns (corresponding to equation (7)) is:

$$\begin{aligned} r_{i,t}^* &= \lambda_{i,t} x_{i,t}^* + u_{i,t} = (\alpha_i + \delta_i d_t + \gamma_i L_{i,t}^* + \eta D_t + \varepsilon_{i,t}) x_{i,t}^* + u_{i,t} \\ &= \alpha_i x_{i,t}^* + \delta_i d_t x_{i,t}^* + \gamma_i L_{i,t}^* x_{i,t}^* + \eta D_t x_{i,t}^* + (\varepsilon_{i,t} x_{i,t}^* + u_{i,t}) \end{aligned} \quad (9)$$

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 7. 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. As before, time variation in effective liquidity seems to be largely firm-specific.

9. Conclusions

Taking as our starting point a linear microstructure specification in which returns are driven by signed order flow, 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 at short-term fifteen-minute intervals using time-aggregated trade and quote data.

We find that common factors exist in both absolute and signed order flow. These explain part, but not all, of the common variation in absolute and signed returns. This conclusion does not depend on whether the common factors were constructed using principal components (given cross-sectional variance within a single set of variables, e.g., returns) or canonical correlations (given cross-sectional covariance between two sets of variables, e.g., returns and signed order flow).

Our assessment of liquidity is less supportive of economically significant common factors. After standardizing to remove time-of-day effects, the strength of any common

factors in spreads and related liquidity measures, as judged by the first principal components of the correlation matrices, is modest. This is confirmed by cross-sectional regressions in which price impact coefficients are projected on various explanatory variables. Own-firm effects dominate the principal component (“common factor”) effects and daily liquidity factor estimates. Thus, the systematic liquidity effects visible during market crises such as 1987 and 1989 do not appear to characterize normal trading.

10. References

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

Chan, K., K.C. Chan, and G.A. Karolyi, 1991. Intraday volatility in the stock index and stock index futures market. *Review of Financial Studies* 4, 657-84.

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

Chordia, T., R. Roll, and A. Subrahmanyam, 1998, Commonality in liquidity, Working Paper, Anderson Graduate School of Management, UCLA.

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

Desai, A., M. Nimalendran, and S. Venkataramaan, 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*.

Foster, F.D., and S. Viswanathan, 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., and S. Viswanathan, 1994. Trading costs for target firms around takeovers. *Advances in Financial Economics* 1.
- Foster, F.D., and S. Viswanathan, 1995. Can speculative trading explain the volume-volatility relation? *Journal of Business and Economic Statistics* 13, 379-96.
- Gallant, A.R., P.E. Rossi, and G. Tauchen, 1992. Stock prices and volume. *Review of Financial Studies* 5, 199-242.
- Harris, L. E., G. Sofianos, and J. E. Shapiro. 1994. Program trading and intraday volatility. *Review of Financial Studies* 7, no. 4: 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 G. S. Maddala and C. R. Rao, eds.: *Handbook of Statistics 14: Statistical Methods in Finance*, (Amsterdam: Elsevier North Holland).
- Hasbrouck, J., 1996b. Order Characteristics and Stock Price Evolution: An Application to Program Trading. *Journal-of-Financial-Economics* 41, 129-49.
- Jain, P.C., and G.H. Joh, 1988. The dependence between hourly prices and trading volume. *Journal of Financial and Quantitative Analysis* 23, 269-83.
- Karpoff, J.M., 1987. The relation between price changes and trading volume: a survey. *Journal of Financial and Quantitative Analysis* 22, 109-26.
- 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, 313-38.
- Kumar, P., and D.J. Seppi, 1994. Information and index arbitrage. *Journal of Business* 67,

- 481-509.
- Kyle, A. S., 1984, Market structure, information, futures markets, and price formation, in Gary G. Storey, Andrew Schmitz, and Alexander H. Sarris, eds.: *International Agricultural Trade: Advanced Readings in Price Formation, Market Structure, and Price Instability*, (Boulder, Colorado and London: Westview Press).
- Kyle, A.S., 1985. Continuous auctions and insider trading. *Econometrica* 53, 1315-36.
- Lee, C.M., B. Mucklow, and M.J. Ready, 1993. Spreads, depths and the impact of earnings information: An intraday analysis. *Review of Financial Studies* 6, 345-74.
- Lo, A.W. and J. Wang, 1997, Trading volume: Definitions, data analysis, and implications of portfolio theory, Working Paper, Laboratory for Financial Engineering.
- MacKinlay, A.C., and K. Ramaswamy, 1998. Index-futures arbitrage and the behavior of stock index futures prices. *Review of Financial Studies* 1, 137-58.
- Madhavan, A., and S. Smidt. 1991. A Bayesian model of intraday specialist pricing. *Journal of Financial Economics*; 30(1), November 1991, Pages 99-134.
- Miller, M. H., J. Muthuswamy, and R. E. Whaley. 1994. Mean reversion of Standard & Poor's 500 index basis changes: arbitrage-induced or statistical illusion? *Journal of Finance* 49, no. 2: 479-513.
- Morrison, Donald F. 1976. *Multivariate Statistical Methods, Second Edition*. New York: McGraw-Hill.
- O'Hara, Maureen. 1995. *Market Microstructure Theory*. Cambridge, USA: Blackwell Publishers.
- Subrahmanyam, A. 1991. A theory of trading in stock index futures. *Review of Financial Studies* 4, no. 1: 17-51.

Wood, R.A., T.H. McInish, and J.K. Ord, 1985. An investigation of transaction data for NYSE stocks. *Journal of Finance* 25, 723-39.

Table 1. Descriptive Statistics

The sample is the thirty Dow firms 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-firm 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. Signed Returns and Signed Volume Measures

The sample is the thirty Dow firms for all trading days in 1994. The means and standard deviations are for the raw series across firms and time. Eigenvalues are based on series standardized to remove time-of-day effects. 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.

Variable		Mean (not standardized by time-of-day)	Std. Dev. (not standardized by time-of-day)	Eigenvalues		
				First	Second	Third
Return		-0.001	0.005	6.32	1.04	1.00
Trades	Total	0.635	3.344	3.36	1.44	1.27
	Small	0.521	3.356	2.75	1.61	1.36
	Medium	0.071	0.105	3.60	1.14	1.07
	Large	0.043	0.054	1.67	1.09	1.08
Share volume (100-shr lots)	Total	16.259	17.810	2.36	1.08	1.07
	Small	0.542	9.135	3.92	1.15	1.11
	Medium	5.556	5.446	3.03	1.12	1.09
	Large	10.161	16.595	1.48	1.11	1.10
Dollar volume (\$10,000)	Total	9.123	10.728	2.38	1.09	1.07
	Small	0.057	4.400	3.90	1.11	1.11
	Medium	3.356	3.572	2.99	1.12	1.08
	Large	5.711	9.357	1.48	1.11	1.10
Square root of dollar volume ($\times 10^{-2}$)	Total	1.206	3.208	4.06	1.08	1.05
	Small	0.229	3.350	3.51	1.32	1.20
	Medium	0.489	0.566	3.36	1.13	1.08
	Large	0.488	0.668	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 firms for all trading days in 1994. C_1 and C_2 are the first and second canonical factors; P_1 and P_2 are the first and second principal components.

		(Signed) Returns				Signed square-root dollar volume			
		C_1	C_2	P_1	P_2	C_1	C_2	P_1	P_2
(Signed) Returns	C_1	1.00							
	C_2	0.00	1.00						
	P_1	0.99	0.04	1.00					
	P_2	0.04	0.08	0.00	1.00				
Signed square-root dollar volume	C_1	0.83	0.00	0.82	0.03	1.00			
	C_2	0.00	0.58	0.02	0.05	0.00	1.00		
	P_1	0.82	0.03	0.82	0.04	0.98	0.05	1.00	
	P_2	-0.04	0.10	-0.05	0.24	-0.05	0.17	0.00	1.00

Table 4. Absolute Returns and Unsigned Volume

The sample is the thirty Dow firms 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 across firms and time. Eigenvalues are based on series standardized to remove time-of-day effects. Letting $z_{i,t}^*$ denote the standardized value of a variable for firm i at time t , we construct the correlation matrix $[\rho_{i,j}]$ where $\rho_{i,j} = \text{Corr}(z_{i,t}^*, z_{j,t}^*)$. For each variable, the table reports the largest eigenvalues of this correlation matrix. 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	0.032	3.64	1.10	1.09
Number of Trades	Total	22.581	14.879	5.00	1.78	1.44
	Small	19.344	13.114	4.61	1.82	1.47
	Medium	2.626	1.559	4.43	1.21	1.19
	Large	0.610	0.398	1.78	1.29	1.14
Share Volume (100-shr lots)	Total	388.823	241.130	3.41	1.29	1.17
	Small	87.706	53.096	4.97	1.39	1.25
	Medium	127.992	78.103	3.81	1.20	1.18
	Large	173.124	119.032	2.21	1.22	1.16
Dollar Volume (\$10,000)	Total	188.941	131.576	3.60	1.23	1.15
	Small	43.528	29.335	5.55	1.51	1.14
	Medium	65.417	46.371	3.94	1.27	1.18
	Large	79.996	60.174	2.23	1.18	1.14
Square root of dollar volume ($\times 100^{-2}$)	Total	44.164	28.341	5.53	1.36	1.25
	Small	25.169	16.440	5.35	1.44	1.21
	Medium	12.564	8.168	4.27	1.22	1.15
	Large	6.431	4.496	2.16	1.24	1.13

Table 5. Canonical and Principal Factors of Absolute Returns and Unsigned Square-root Dollar Volume

The sample is the thirty Dow firms for all trading days in 1994. Data series are aggregated over 15-minute intraday intervals and then standardized to remove time of day effects. C_1 and C_2 are the first and second canonical factors; P_1 and P_2 are the first and second principal components.

		Absolute returns				Square-root dollar volume			
		C_1	C_2	P_1	P_2	C_1	C_2	P_1	P_2
Absolute returns	C_1	1.00							
	C_2	0.00	1.00						
	P_1	0.97	-0.10	1.00					
	P_2	0.17	0.63	0.00	1.00				
Square-root dollar volume	C_1	0.73	0.00	0.70	0.12	1.00			
	C_2	0.00	0.53	-0.05	0.33	0.00	1.00		
	P_1	0.69	-0.04	0.69	0.07	0.95	-0.07	1.00	
	P_2	-0.10	-0.20	-0.07	-0.08	-0.13	-0.38	0.00	1.00

Table 6. Liquidity Measures

The sample is the twenty-four Dow firms that did not split during 1994 for all trading days in 1994. The bid and ask for trade τ are denoted by B_τ and A_τ (dollars per share); N_τ^B and N_τ^A are the numbers of shares (in 100-share round lots) posted at the bid and ask. The spread is $A_\tau - B_\tau$; the log spread is $\log(A_\tau/B_\tau)$; the log size is $\log(N_\tau^A) + \log(N_\tau^B)$; the quote slope is $(A_\tau - B_\tau)/(\log(N_\tau^A) + \log(N_\tau^B))$; the log quote slope is $\log(A_\tau/B_\tau)/(\log(N_\tau^A) + \log(N_\tau^B))$. For all of the preceding, we employ time-weighted averages over the fifteen-minute intervals. For a given trade at price p , the effective spread is defined as $|p - (A + B)/2|$. These are averaged by trade size over the interval. The means and standard deviations are for the raw series across firms and time. Eigenvalues are based on series standardized to remove time-of-day effects. Small trades are 2,000 shares or less; medium trades are 2,001-10,000 shares; large trades are >10,000 shares.

Variable		Mean	Std. Dev.	Eigenvalues		
		(not standardized by time-of-day)	(not standardized by time-of-day)	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
	Total	0.05932	0.01967	1.13	1.12	1.10
Effective Spread	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 7. Price Impact Regressions

For each firm the estimated specification is $r_{i,t}^* = \alpha_i x_{i,t}^* + \delta_i d_t x_{i,t}^* + \gamma_i L_{i,t}^* x_{i,t}^* + h_i P_t^L x_{i,t}^* + e_{i,t}$ where $r_{i,t}^*$ is the standardized log quote midpoint return for firm i , $x_{i,t}^*$ is the standardized order flow (signed square root dollar volume), d_t is a vector of time-of-day dummies and $L_{i,t}^*$ is one of the following (standardized) liquidity proxies. The bid and ask for trade τ are denoted by B_τ and A_τ (dollars per share); N_τ^B and N_τ^A are the numbers of shares (in 100-share round lots) posted at the bid and ask. The spread is $A_\tau - B_\tau$; the log spread is $\log(A_\tau/B_\tau)$; the log size is $\log(N_\tau^A) + \log(N_\tau^B)$; the quote slope is $(A_\tau - B_\tau)/(\log(N_\tau^A) + \log(N_\tau^B))$; the log quote slope is $\log(A_\tau/B_\tau)/(\log(N_\tau^A) + \log(N_\tau^B))$. For all of the preceding, we employ time-weighted averages over the fifteen-minute intervals. P_t^L is the first principal component (constructed across all firms) of the liquidity proxy. The incremental explanatory power is computed assuming that the variables are added to the specification in the indicated column order (left to right). A separate regression is estimated for each firm. Numbers in the table are means across firms.

Panel A. Ordering of variables has $L_{i,t}^* x_{i,t}^*$ last.

Liquidity Proxy, L_t	Incremental contribution to R^2 from:			
	$x_{i,t}^*$	$d_t x_{i,t}^*$	$P_t^L x_{i,t}^*$	$L_{i,t}^* x_{i,t}^*$
Spread	0.240	0.005	0.006	0.022
Log Spread	0.240	0.005	0.007	0.023
Log Size	0.240	0.005	0.007	0.024
Quote Slope	0.240	0.005	0.009	0.037
Log Quote Slope	0.240	0.005	0.010	0.037

Panel B. Ordering of variables has $P_t^L x_{i,t}^*$ last.

Liquidity Proxy, L_t	Incremental contribution to R^2 from:			
	$x_{i,t}^*$	$d_t x_{i,t}^*$	$L_{i,t}^* x_{i,t}^*$	$P_t^L x_{i,t}^*$
Spread	0.239	0.005	0.027	0.003
Log Spread	0.240	0.005	0.026	0.002
Log Size	0.240	0.005	0.029	0.001
Quote Slope	0.240	0.005	0.045	0.001
Log Quote Slope	0.240	0.005	0.046	0.001

Figure 1. The Quote Slope

The bid and ask quotes prevailing at time τ are denoted by B_τ and A_τ . N_τ^B and N_τ^A are the numbers of shares sought at the bid and available at the ask. The quote midpoint is $m_\tau = (A_\tau + B_\tau)/2$. The quote slope is the slope of the dotted line in the figure.

