Trading Costs and Returns for US Equities:
The Evidence from Daily Data

Joel Hasbrouck

Department of Finance
Stern School of Business
New York University
44 West 4th St. Suite 9-190
New York, NJ 10012-1126
212.998.0310

jhasbrou@stern.nyu.edu

First draft: August 30, 2002
This draft: October 8, 2004

Preliminary draft
Comments welcome

David Easley, Maureen O'Hara, Soren Hvidkjaer, and Ronnie Sadka generously shared their liquidity estimates. This acknowledgement should not be construed as implying that any of them approve of or endorse the present paper.

For comments on an earlier draft, I am grateful to Yakov Amihud and Lubos Pastor, Bill Schwert and seminar participants at the University of Rochester. All errors are my own responsibility.

The latest version of this paper and a SAS dataset containing the long-run Gibbs sampler estimates are available on my web site at www.stern.nyu.edu/~jhasbrou.

Trading Costs and Returns for US Equities: The Evidence from Daily Data

Abstract

This study examines various measures of trading costs estimated from high-frequency data, the extent to which these measures can be estimated from daily data, and finally the relation between the daily-based proxies and stock returns (where trading cost is viewed as a characteristic). The high-frequency estimates of trading cost achieve partial agreement. Posted spreads and effective costs are highly correlated. Price impact measures and other statistics from dynamic models, however, are only modestly correlated with each other. Among the set of proxies constructed from daily data, a Gibbs estimate of the effective cost stands out, achieving a correlation of 0.89 with effective cost estimated from high-frequency data in a comparison sample. Both the Gibbs estimate of effective cost and the illiquidity ratio covary positively with risk-adjusted returns, but the relationship exhibits marked seasonality.

JEL classification codes: C15, G12, G20

1. Introduction

The notion that agents take into consideration trading costs, and that these costs affect equilibrium expected returns, motivates studies that span market microstructure and asset pricing (recently surveyed in Easley and O'Hara (2002)). Empirical studies that bridge the two fields often encounter difficulties, however, arising from differences in the data samples and frequencies favored by each area. Asset pricing tests generally require daily or monthly samples that are large in cross-section and time span. Microstructure measures, on the other hand, are generally estimated with high-frequency trade and quote data. This limits their availability to the relatively small and recent data samples for which these data exist.[1] In reconciling the conflicting needs of the two approaches, asset pricing considerations appear predominant. This is because precision in estimation of expected returns depends on the length of the data sample, not the sampling frequency (Merton (1980)). This establishes the importance of liquidity measures that can be constructed from data of daily or lower frequency.

The ultimate contribution of the present study is an empirical analysis of the relationship between expected returns and a broad set of liquidity characteristics estimated from daily data. The set includes the Amivest liquidity ratio (Cooper, Groth, and Avera (1985)), the Amihud (2002) illiquidity measure, and the Pastor and Stambaugh (2003) reversal coefficient, and the Gibbs estimate of the effective cost of trading. The last, based on the Roll (1984) model of the spread, is developed and applied to futures data in Hasbrouck (2003). The Gibbs estimate uses a Bayesian perspective and incorporates a non-negativity prior on the spread. (The usual estimate

[1] Microstructure data commonly used in asset pricing studies include: beginning/end of year average spreads for NYSE stocks, 1955-1979 (Stoll and Whalley (1983), based on Fitch data, also used by numerous subsequent studies); average Nasdaq spreads, 1973-1990 (CRSP, see Eleswarapu (1997)); Institute for the Study of Securities Markets (1983-1992); and, TAQ (1993-present).

of the Roll model is infeasible in the frequently encountered case of a positive in-sample return autocovariance.)

The empirical asset pricing framework is that suggested by Brennan, Chordia, and Subrahmanyam (1998) (BCS). Most tests of asset pricing models follow Fama and MacBeth (1973) in forming size- and characteristic-ranked portfolios. When there are few characteristics, this approach may enhance statistical power. The procedure is problematic, however, when potential characteristics are numerous, as results may be sensitive to sort ordering. The BCS procedure avoids this requirement and so allows all of the liquidity measures to be evaluated in an even-handed fashion.

The plausibility of the conclusions from the asset pricing tests, however, depends on the quality of the daily-based liquidity measures. En route to the expected return estimations, therefore, the paper undertakes a detailed analysis of these measures. This is accomplished by comparing daily and commonly used high-frequency measures, drawn from a sample where it is feasible to estimate both. This set is augmented with probability of informed trading (PIN, Easley, Hvidkjaer, and O'Hara (2002)) and permanent/transitory impact coefficients (Sadka (2004)).

The study of asset pricing and liquidity is divided into two streams, according to whether liquidity is viewed primarily as a security characteristic or a risk factor. The present study adopts the characteristic perspective, following Stoll and Whalley (1983), Amihud and Mendelson (1986), Eleswarapu and Reinganum (1993), Chalmers and Kadlec (1998) and Eleswarapu (1997) among others. In this view agents equalize expected returns net of trading costs, and securities with higher trading costs must have higher gross expected returns. Various theoretical models predict that this effect should not be substantial because agents blunt the impact of trading costs by adjusting their portfolios less frequently (Constantinides (1986), Heaton and Lucas (1996), Vayanos (1998)). The extent to which they actually do this, however, remains a puzzle. Actual trading volumes are much higher than the theoretical equilibrium models predict.

The risk-factor perspective stresses stochastic variation in trading costs (Acharya and Pedersen (2002), Pastor and Stambaugh (2003), Sadka (2004)). A security's exposure to aggregate liquidity variation leads to risk that is non-diversifiable and therefore conceivably priced like any other source of risk. The risk-factor and characteristic views are not incompatible. Cross-sectional and deterministic components of variation in liquidity could be priced as characteristics, while stochastic variation over time may give rise to a risk factor. The empirical frameworks are quite different, however. In estimating a liquidity characteristic at a point in time, it is reasonable to take its average over a recent period. Determination of its properties as a risk factor, however, requires a time series sample long enough to accurately characterize the stochastic variation. Risk-factor analysis thus places heavier demands on the data.

The initial conclusions of the paper concern on the set of liquidity measures constructed from high-frequency data. The sample distributions of all estimates exhibit numerous extreme values. Although in some cases these may arise from methodological shortcomings, it is also quite plausible that the extreme values may reflect large cross-sectional variation in actual trading costs. Concordance across the measures also varies. The simple single-trade measures comprising average intraday spread, the closing spread and the effective cost are largely consistent, with pairwise correlations of about ninety-five percent. When the set is expanded to include measures derived from signed trades (the probability of informed trading) and dynamic models of prices and signed trades, however, the correlations deteriorate, generally remaining positive, but only modestly so. Thus, there is no overall concordance, and no single measure that captures all dimensions of liquidity. Stoll (2000) arrives at similar conclusions, using a panel of liquidity measures that partially overlaps the set used in the present study.

The study next examines the correlations between the high-frequency liquidity estimates and the proxies constructed from daily data. The simple single-trade measures are relatively easy to proxy. The Roll estimate of the bid-ask spread (with infeasible estimates set to zero) performs well, achieving a correlation of 0.82 with the average effective cost computed from high-frequency data. The Gibbs estimate, however, performs slight better, attaining a 0.89 correlation.

Thus, effective cost can reasonably be proxied from daily data. Daily proxies for price impact and reversal liquidity measures, however, are less successful.

The expected return estimations provide modest evidence supporting the importance of liquidity as a characteristic. When BCS risk-adjusted returns are regressed against the proxies individually, both the Gibbs estimate of the effective cost and the illiquidity ratio appear by the usual standards of statistical significance to be correlated with BCS risk-adjusted returns. When multiple proxies are included in the specification, however, no single measure emerges as the clear winner. Moreover, the time series of the cross-sectional liquidity coefficients exhibits marked seasonality, with January values being the highest. This is consistent with Eleswarapu and Reinganum (1993).

The paper is organized as follows. The next section summarizes measures of trading cost based on high-frequency trade and quote data. Section 3 describes the proxies constructed from daily data. Reversal measures, however, are sufficiently distinct to warrant a separate discussion in Section 4. Section 5 describes the construction of the high-frequency/daily comparison sample and the estimation details. The properties and interrelations of these measures are discussed in Section 6. The remaining two sections examine the long-term evidence. Section 7 describes the time-series and cross-section properties of the Gibbs estimates of effective cost in the CRSP daily data file. Section 8 presents the asset-pricing specifications. A brief summary concludes the paper in Section 9.

2. Microstructure-based measures of transaction costs and liquidity

This section discusses measures generally motivated by microstructure models and estimated with high-frequency data. The following describes in turn posted and effective spreads, impact measures based on dynamic models of prices and trades. All measures are summarized in Table 1.

a. Spreads: posted and effective

Most computations of transaction costs from the investor's viewpoint can be discussed within the context of the implementation shortfall approach advocated by Perold (1988). This approach focuses on the difference between the actual portfolio return and the return that would have been achieved had all purchases and sales occurred at hypothetical prices that were free of trading costs. The difference is the cost of implementing the portfolio strategy.

For a single executed trade this approach suggests measuring the cost as the difference between the average transaction price and a hypothetical benchmark price taken prior to the initial trade. One common benchmark is the midpoint of the bid and ask prevailing at the time of the order submission. For a small trade executed at the bid or ask, a sensible first estimate of the trading cost is the half-spread. The half-spread associated with the k^{th} quote update is $s_k/2 = (a_k - b_k)/2$, where a_k is the log of the ask and b_k is the log of the bid. (Logs are used here to enhance comparability across a wide range of stock prices.) The present study employs intraday time-weighted averages of half-spreads, denoted $\bar{s}^{Time}/2$. Numerous asset pricing studies, however, rely on closing (end-of-day) spreads (Stoll and Whalley (1983), Amihud and Mendelson (1986), Amihud and Mendelson (1989), Eleswarapu and Reinganum (1993), Kadlec and McConnell (1994), and Eleswarapu (1997), among others). In the present study, average closing half-spreads are denoted $\bar{s}^{Close}/2$

In many markets and for a variety of reasons, market orders often transact at prices better than the posted quotes. This motivates use of the effective cost, defined for the k^{th} trade as

$$c_k = \begin{cases} p_k - m_k, \text{ for a buy order} \\ m_k - p_k, \text{ for a sell order} \end{cases} \quad (1)$$

where p_k is the log trade price and m_k is the log quote midpoint prevailing at the time the order was received. The effective cost is most meaningful for small market orders that can be accommodated in a single trade. The effective cost occupies a prominent role in US securities regulation. Under SEC rule 11ac1-5, market centers must periodically report summary statistics of this measure.

Accurate computation of the effective cost requires knowledge of order characteristics, most importantly the arrival time and direction (buy or sell). Studies of order data are common (e.g., Keim and Madhavan (1995), Harris and Hasbrouck (1996), Chan and Lakonishok (1997) and Conrad, Johnson, and Wahal (2001)), but none of the samples spans a long history. When order data are unavailable, the effective cost is often estimated from transaction and quote data. A trade priced above the midpoint of the bid and ask (prevailing at the time of the trade report or a brief time earlier) is presumed to be a buy order; a trade priced below the midpoint is presumed to be a sale. Effective costs computed in this fashion are extensively used in academic studies. The primary summary measure used in this study is $\bar{c} = \overline{|p_k - m_k|}$, where the average is computed over all trades and is weighted by the dollar value of the trade.

b. Measures based on dynamic models

Many economic models imply joint dynamics for orders and price changes that involve both permanent and temporary effects. The former reflect the information content of the order, with Kyle (1985) and Glosten and Milgrom (1985) exemplifying the two main approaches. Temporary components latter arise from transient liquidity effects, inventory control behavior, price discreteness, etc. In addressing practical trading problems, the principal advantage of the dynamic models over the single-trade approaches discussed above lies in their ability to project execution costs when trades are distributed over time.

The literature contains a large number of approaches. The following specification is representative. The evolution of the log quote midpoint is:

$$m_t = m_{t-1} + \lambda Z_t + u_t \tag{2}$$

Here, t indexes five-minute intervals, and Z_t is the daily cumulative signed logarithm of dollar volume:

$$Z_t = \sum_{k \in N_t} q_k \log\left(V_k^{Dollar}\right) \tag{3}$$

The summation runs over the number of trades in the interval N_t, V_k^{Dollar} is the dollar volume of the k^{th} trade, and q_t is the direction of the trade (+1 if buyer-initiated, –1 if sellerinitiated). The log transformation is motivated by the concavity generally found in these sorts of specifications (e.g., Kempf and Korn (1999)). The study uses two estimates based on eq (2): λ (the impact coefficient) and the coefficient of determination for the regression, i.e., the return variance explained by trades, denoted R_m^2 (see Hasbrouck (1991)).

Sadka (2004) proposes a specification based on Glosten and Harris (1988). Log quote midpoint dynamics are given by:

$$m_k = m_{k-1} + q_k \left[\psi^P + \lambda^P V_k \right] + u_k \tag{4}$$

where k indexes trades, V_k is the share volume of the k^{th} transaction, and q_k is the direction.. The middle term on the right-hand-side is the permanent (informational) impact of the trade, characterized as an affine function of volume. The disturbance u_k reflects public information. The (log) transaction price is

$$p_k = m_k + q_k \left[\psi^T + \lambda^T V_k \right] \tag{5}$$

The last term on the right-hand-side captures non-informational costs paid by the trade initiator, also an affine function of volume. For summary and comparison purposes across firms and time, the present study standardizes the model predictions for a 200-share order. The implied permanent impact of a 200-share order is $\pi_{200}^P = \psi^P + \lambda^P \times 200$; the implied temporary impact is $\pi_{200}^T = \psi^T + \lambda^T \times 200$; and, the total impact is $\pi_{200} = \pi_{200}^P + \pi_{200}^T$.

c. The probability of informed trading (PIN)

The sequential trade models developed in Easley and O'Hara (1987), Easley and O'Hara (1992) and related work feature asymmetric information. Easley, Kiefer, and O'Hara (1997), Easley, Kiefer, O'Hara, and Paperman (1996), and Easley, Hvidkjaer, and O'Hara (2002) discuss empirical implementations. For present purposes, the most important parameters in these models are the arrival rates of informed traders, uninformed traders and information events. These quantities determine a summary measure, the probability of informed trading (*PIN*).

Although the sequential trade models characterize the joint behavior of trades and prices, the *PIN* estimates are based solely on signed trades. Thus, although *PIN* and the price impact measures discussed above are both motivated by asymmetric information, they use market data in fundamentally different ways. The price impact estimates rely mostly on contemporaneous (and near-contemporaneous) comovement of trades and prices. The *PIN* estimates most strongly reflect the incidence of intraday order flows that are buy- or sell-dominated.

3. Transaction cost and liquidity measures based on daily data

Estimation of the measures discussed in the previous section generally requires intraday quote and trade data. I now turn to measures that can be estimated using daily return and volume data.

a. Effective cost estimates based on the Roll model

Roll (1984) suggested a simple model of the spread in an efficient market. Following the notation used earlier, the model may be stated as:

$$\begin{aligned} m_k &= m_{k-1} + u_k \\ p_k &= m_k + c\,q_k \end{aligned}. \tag{6}$$

The time subscript, k, can be thought of as indexing successive trades. m_k is the log quote midpoint prevailing prior to the trade, p_k is the trade price, q_k is the direction indicator (–1 for a sale to the dealer, +1 for a purchase), and c is the effective cost (cf. eq. (1)), which is presumed constant. The model has essentially the same form when it is aggregated over time. In particular, the time subscript can be viewed as indexing days ("t") rather than trades, with q_t being interpreted as the direction variable for the last trade of the day. The estimation approaches described below take this perspective.

The moment estimate of c.

The Roll model is usually estimated by method-of-moments. The model implies

$$\Delta p_t = m_t + c\,q_t - \left(m_{t-1} + c\,q_{t-1}\right) = c\Delta q_t + u_t\,, \tag{7}$$

from which it follows that:

$$\begin{aligned} Var(\Delta p_t) &= \sigma_u^2 + 2c^2 \\ Cov(\Delta p_t, \Delta p_{t-1}) &= -c^2 \end{aligned} \qquad (8)$$

The corresponding sample estimates for the variance and autocovariance imply estimates for σ_u and c that possess all the usual properties of GMM estimators, including consistency and asymptotic normality. Moment estimation for this model is relatively easy to implement and often satisfactory.

A sample moment estimate of c only exists, however, if the first-order sample autocovariance is negative. In samples of daily frequency this is often not the case. In annual samples of daily returns, Roll found positive autocovariances in roughly half the cases. Harris (1990) discusses the incidence of positive autocovariances, and other properties of this estimator. His results show that positive autocovariances are more likely for low values of the spread. Accordingly, one simple remedy to the problem is to assign an a priori value of zero. I define the moment/zero estimate as:

$$c^{MZ} = \begin{cases} \sqrt{-Cov(\Delta p_t, \Delta p_{t-1})}, \text{ if } Cov(\Delta p_t, \Delta p_{t-1}) > 0 \\ 0, \text{ otherwise} \end{cases}$$

The Gibbs-sampler estimate of c, c^{Gibbs}

Hasbrouck (2003) advocates Bayesian estimation using the Gibbs sampler. To complete the Bayesian specification, I assume here that $u_t \overset{d}{\sim} i.i.d.N(0, \sigma_u^2)$ and that the data sample is $\{p_1, p_2, \ldots, p_T\}$. The prior for c is $c \overset{d}{\sim} N^+(0, \sigma_c^{2,prior})$ where the "+" superscript denotes restriction to the positive domain and $\sigma_c^{2,prior} = 1$. Effective costs are generally on the order of ten percent or less, so this is a fairly diffuse prior. The prior for σ_u^2 is inverted gamma distribution. I use $\sigma_u^2 \overset{d}{\sim} IG(\alpha, \beta)$ with $\alpha = \beta = 10^{-12}$, also implying a fairly uninformative prior. In the Bayesian approach, the unknowns comprise both the model parameters $\{c, \sigma_u^2\}$ and the latent data, i.e., the trade direction indicators $q \equiv \{q_1, \ldots, q_T\}$ and the efficient prices $m \equiv \{m_1, \ldots, m_T\}$. The parameter posterior $f(c, \sigma_u | p)$ is not obtained in closed-form, but is instead characterized by a random sample drawn from it. These draws are constructed by iteratively drawing from the full conditional distributions. The Gibbs estimate of c, denoted c^{Gibbs}, is the sample mean of the posterior draws.

The Gibbs estimate of the effective cost offers some advantages over the moment estimate. First, the prior can restrict the estimates to the positive domain. Second, within the framework of the model, the posterior is an exact small sample distribution. A third advantage is particular to the CRSP data and stems from the CRSP convention of reporting the midpoint of the closing bid and ask (flagged as a negative value) in lieu of the transaction price if there are no trades on a particular day.

The implications of the CRSP convention differ for the two estimates. The moment estimate is based on the sample return autocovariance, which is proportional to $\sum_t r_t r_{t-1}$. The summand here is $r_t r_{t-1} = (p_t - p_{t-1})(p_{t-1} - p_{t-2})$. In principle, since the model applies to trade prices, a term should be included only if it encompasses a sequence of three trade prices. For many stocks, however, this would drastically reduce the sample size. To avoid this attenuation, the present study uses all closing prices in the moment estimates irrespective of whether they represent trades or quote midpoints.

The Gibbs estimate, on the other hand, is easily generalized to accommodate quote midpoints. Specifically, if a quote midpoint is reported on day t, we let the trade direction indicator for that day $q_t = 0$. From eqs (6), this implies $p_t = m_t$. Intuitively, this prevents the observation from contributing directly to the estimate of c, but allows one or both of the adjacent prices to contribute (assuming that they have valid transaction prices).[2]

This treatment of the Roll model is almost certainly misspecified in a number of important respects. Actual samples of stock returns contain many more extreme observations than a normal density would likely admit. Trade directions are unlikely to be independent of the

---

[2] Formally, this procedure can be justified by embedding the Roll model in a more general framework in which observation of a quote midpoint or trade price is determined randomly (and independently of the other variables).

efficient price evolution. Realized prices are discrete. The effective cost is unlikely to be constant within a sample. Etc. Hasbrouck (2003) discusses various extensions to deal with some of these features. For computational expediency and programming simplicity, however, the present paper uses the most basic form of the sampler.

Lest misspecification appear to be of major potential importance, it must be emphasized that the Gibbs estimates (like all daily proxies considered here) will to be compared against values constructed independently from high-frequency data. There is accordingly no immediate need to assess the appropriateness of the model assumptions or implementation procedures. If the Gibbs estimates are strongly correlated with the corresponding high-frequency values, these concerns are of secondary importance.

b. The (Amivest) liquidity ratio, L

The effective cost estimates discussed in the prior two section use only daily price data. The remaining measures use volume data as well. This imposes a practical limitation because the interpretation of reported volume may depend on institutional arrangements. Volume in an order-driven market (e.g., NYSE) is not, for example, generally comparable to volume in a quote-driven market (e.g., Nasdaq).

The Amivest liquidity ratio is the average ratio of volume to absolute return:

$$L = \overline{\left(\frac{Vol_d}{|r_d|} \right)} \tag{9}$$

where the average is taken over all days in the sample for which the ratio is defined, i.e., all days with nonzero returns. It is based on the intuition that in a liquid security, a large trading volume may be realized with small change in price. This measure has been used in the studies of Cooper, Groth, and Avera (1985), Amihud, Mendelson, and Lauterbach (1997), and Berkman and Eleswarapu (1998), among others.

c. The illiquidity ratio, I

Amihud (2002) suggests measuring *illiquidity* as:

$$I = \overline{\left(\frac{|r_d|}{Vol_d}\right)} \qquad (10)$$

where r_d is the stock return on day d and Vol_d is the reported dollar volume. The average is computed over all days in the samples for which the ratio is defined, i.e. days with nonzero volume. This measure loosely corresponds to λ in eq. (2), but whereas λ measures the return impact of a cumulative signed order flow, I captures the absolute return impact of a cumulative unsigned volume. This measure is used as a risk factor by Acharya and Pedersen (2002).

4. Reversal measures

Reversal measures of liquidity differ from those considered above. They essentially summarize the association between return and lagged order flow. The intuition is that order flow induces a price adjustment that initially overshoots true value. Reversion occurs with a lag. This might occur, for example, due to inventory adjustments by market makers. Drawing on this intuition, Pastor and Stambaugh (2003) suggest a quote-midpoint return specification of the form (in present notation)

$$r_t = f_t + u_t + \phi\left(X_{t-1} - X_t\right) \qquad (11)$$

where t indexes days, f_t is a market-wide common factor, u_t is a firm-specific component, and X_t is the signed dollar order flow on day t. The reversal coefficient ϕ is presumed to be negative, with a more negative value indicating lower liquidity. (As the dependent variable in (11) is the quote-midpoint return, the bid-ask bounce term included in Pastor and Stambaugh's eq. (2) is omitted.)

Pastor and Stambaugh do not estimate (11) in actual data. Instead, with the aim of obtaining estimates in the absence of signed order flow data, they suggest estimating liquidity by γ in the regression (using present notation):

$$r_t^e = \theta + \phi r_{t-1} + \gamma\, sign\left(r_{t-1}^e\right) V_{t-1}^{Dollar} + \varepsilon_t \qquad (12)$$

where r_t^e is the stock's excess return (over the CRSP value-weighted market return). By comparison with (11), it is apparent that lagged order flow $\left(X_{t-1}\right)$ is being proxied by dollar volume, signed by excess return. Pastor and Stambaugh find that when they simulate data using (11) and estimate using (12), ϕ in eq (11) and γ estimated from eq (12) are very highly correlated.

Pastor and Stambaugh suggest an alternative rationale for return reversals based on Campbell, Grossman, and Wang (1993). The latter model an economy of market makers with fixed risk preferences and investors who have fluctuating risk preferences. An increase in investors' risk-aversion reduces their desired holdings, causing sales to the market makers (negative order flow). To compensate the market makers for bearing more risk, the subsequent expected return must increase. Similarly, a decrease in investors' risk-aversion causes purchases from the market makers (positive order flow) and negative subsequent expected return. Although leading to a similar specification, this mechanism is quite different from the one already suggested. Specifically, r_t in eq (11) is more properly viewed as estimate of an expected return. The expected return reflects risk-aversion and time preference that are usually ignored in the microstructure models underlying specifications such as eq. (4).

5. The comparison sample and estimation

a. Construction of the comparison sample

The comparison sample is a random selection of firm-years for which both TAQ (high frequency) and CRSP (daily) data are available. For each of the ten years 1993-2002, 200 firms were chosen at random. To be eligible for selection in a given year, a firm had to meet the following criteria. According to CRSP:

1. The issue was an ordinary common share, traded at the beginning and end of the year. (Primes, scores, closed end funds and REITs were excluded).
2. There were no changes in listing exchange, cusip identifier or ticker symbol over the year.
3. The number of shares outstanding did not change by more than 20% over the year.

4. The first criterion simply limited the sample to those issues usually considered in asset pricing tests. The second and third criteria are imposed to ensure reasonable homogeneity in a firm's trading characteristics over the year.

It was furthermore required that the firm could be matched to the TAQ database. Of the 2,000 firm-year observations, 794 were NYSE/Amex-listed and 1,206 were Nasdaq-listed.

A common practice in equity market studies is the imposition of a price cutoff. One, five and (occasionally) ten dollars are often set as the minimum acceptable level for the inclusion of an issue. No such restrictions are imposed in the present analysis. This is deliberate, as casual observation suggests that transaction costs are indeed high for low-price stocks. In any investigation of the links between returns and transaction costs, therefore, these firms would normally be considered the most informative.

b. Estimation of liquidity measures and proxies

In general, for a particular firm and year, each TAQ-based measure is estimated monthly. The monthly estimates are averaged to obtain an annual estimate. This grouping is used for computational convenience, since the TAQ files are organized on a monthly basis.

The spread-related measures estimated in this fashion are: the time-weighted quoted half-spread $\bar{s}^{Time}/2$; the daily closing half-spread $\bar{s}^{Close}/2$; and, the trade-weighted effective cost $\bar{c}$.

λ and R_m^2 are estimated from eq (2) applied to five-minute returns and signed order flows. A regression estimation in a typical month comprised roughly 21 days $\times$ 6.5 hours $\times$ 12 intervals/hour = 1,638 observations. The monthly estimates are averaged over the year.

The reversal coefficient for signed volume ϕ is estimated from a modified version of eq. (11):

$$r_t^e = \phi\left(X_{t-1} - X_t\right) + u_t \tag{13}$$

The reversal coefficient for unsigned volume γ is estimated from eq. (12). Both specifications are estimated for each firm using a year's worth of daily data.

PIN estimates were supplied by David Easley, Maureen O'Hara and Soeren Hvidkjaer. Their computation is documented in Easley, Hvidkjaer, and O'Hara (2002), and the sample is NYSE-listed firms. Ronnie Sadka provided estimates of ψ^T, ψ^P, λ^T, and λ^P (Sadka (2004)). This sample comprises NYSE-listed firms with at least 30 trades in the month of estimation. As noted above, these were used as the basis for permanent, temporary and total cost measures representative of a 200-share purchase: π_{200}^P, π_{200}^T, and $\pi_{200} = \pi_{200}^P + \pi_{200}^T$.

The CRSP-based liquidity measures $\left(c^M, c^{MZ}, c^{Gibbs}, I, L, \gamma\right)$ are estimated using a full year of daily return and volume data. In the computation of the moment estimates of effective costs $\left(c^M \text{ and } c^{MZ}\right)$, a day's return is used irrespective of whether one or both of the prices represented a valid transaction or a quote midpoint. For the Gibbs estimate c^{Gibbs} if a quote midpoint is used on a given day, the trade direction indicator for that day is set to zero.

To reduce the sensitivity of the analysis to estimates based on a small number of observations, a CRSP-based estimate was considered 'missing' if there were fewer than fifty valid observations within the year. For *I*, valid observations had nonzero volume; for *L*, nonzero trades; and for c^{MZ} and c^{Gibbs}, observed trade prices (as opposed to quote midpoints). Beyond this, however, no further sample filters are employed. In particular, there is no cutoff on average price or market capitalization.

6. Analysis of the comparison sample.

a. General features

Table 2 reports summary statistics for the total comparison sample and NYSE/Amex and Nasdaq subsamples. The time-weighted quoted half-spread $\bar{s}^{Time}/2$ averages about 1.8 percent, but the median is lower (1.16 percent). Consistent with this, the distribution is skewed to the right and the kurtosis is high. The average and median of the closing half-spread $\bar{s}^{Close}/2$ are near the corresponding figures for $\bar{s}^{Time}/2$ only in the NYSE/Amex sample. For the Nasdaq sample, the median $\bar{s}^{Close}/2$ is about 2.34% while the median $\bar{s}^{Time}/2$ is 1.79%. This suggests that Nasdaq closing spreads, which are used in some asset pricing studies, overstate the spread a representative trade would face within the trading day.

The median effective cost is approximately 0.76%. which is moderately lower than the median $\bar{s}^{Time}/2$, 1.16%. This is consistent with the results of numerous studies and reflects the sundry mechanisms yielding “price improvement”.

The next two measures are derived from the dynamic model of prices and order flows. The average impact coefficient λ is positive. The average R_m^2 implies that roughly 18 percent of the quote change variance is explained by order flow. The average reversal coefficient ϕ is negative in the NYSE/Amex subsample (as predicted by the model), but not in the Nasdaq subsample.

The Easley, Hvidkjaer, and O'Hara (2002) and Sadka (2004) measures are only available for NYSE/Amex issues. In this sample, the probability of informed trading *PIN* is approximately twenty percent. The impact statistics π_{200}^P and π_{200}^T imply that a 200-share order has a 0.1% permanent impact and a 0.26% temporary impact.

The remaining variables are based on daily CRSP data. The moment estimate of the effective cost c^M is feasible for only 1,396 of the 2,000 firms, and averages somewhat higher than the TAQ-based estimate $\bar{c}$. Setting infeasible values to zero helps matters considerably. The mean, median and standard deviation of c^{MZ} are close to the estimates for $\bar{c}$, but the skewness and kurtosis of c^{MZ} are substantially lower. The Gibbs estimate c^{Gibbs} also exhibits sample means, medians and standard deviations that approximate those of $\bar{c}$. The skewness and kurtosis of the Gibbs estimate, however, are also close to the corresponding $\bar{c}$ values. In short, the distributions of $\bar{c}$ and c^{Gibbs} are very similar.

The magnitudes of the liquidity and illiquidity statistics are difficult to interpret. Their distributions exhibit high kurtosis and skewness. The mean and median of the reversal coefficient γ are positive, contrary to the theory. The skewness and kurtosis, however, highlight the sensitivity of this finding to extreme observations.

While the distributions of all estimates exhibit extreme values, this is most pronounced for measures constructed from daily volume and returns (L, I, and γ). Gabaix, Gopikrishnan, Plerou, and Stanley (2003) find that distribution tails for volume and returns follow power laws with exponents 3/2 and 3 respectively, suggesting that volume moments of order greater than 3/2 and return moments of order greater than 3 are not finite. (The normal distribution possesses finite moments of all orders.) The liquidity ratio implicitly uses volume in the numerator, and might therefore be expected to be particularly ill-behaved.

b. Correlations between the TAQ measures

The analysis now turns to the correlations between the high-frequency liquidity measures. These can at best indicate patterns of agreement and disagreement. In cases where the measures purport to be similar economically, low correlation may indicate sensitivity of one or both quantities to statistical specification and estimation. More generally though, in the absence of a presumption of economic similarity, low correlation may simply reflect distinctly different economic dimensions of liquidity.

Table 3 presents the correlations between the high-frequency liquidity measures. To keep the size of the table manageable, only a subset of the variables are used. Correlations involving the omitted measures are occasionally relevant, however, and are discussed in the text. Each entry in the table contains the usual (Pearson) correlation in the top row, the Spearman (rank order) correlation in the middle, and the number of observations at the bottom. Spearman correlations are reported because they may pick up non-linear but nonetheless monotonic associations between variables. Results for the full sample, and NYSE/Amex and Nasdaq subsamples are shown in Panels A, B and C respectively.

In the full sample (Panel A), effective costs $\overline{c}$ and quoted half-spreads $\overline{s}^{Time}/2$ have correlations near unity. This is also true of the correlations between these variables and $\overline{s}^{Close}/2$ (not shown in the table). This is reasonable because $\overline{s}^{Time}/2$ and $\overline{c}$ differ only by price improvement, and variation in price improvement is likely to be small relative to variation in spreads. The price impact coefficient λ and explanatory power of trades R_m^2 are moderately positively correlated with each other (0.458), but are negatively correlated with $\overline{s}^{Time}/2$ and $\overline{c}$. The reversal measure ϕ exhibits no consistent patterns of correlation with the other variables.

The NYSE subsample (panel B) exhibits a similar pattern of correlations. The matrix is somewhat larger: *PIN* and π_{200} are available for many of the NYSE firms. Both *PIN* and R_m^2 are information variables normalized to the unit interval. They are modestly positively correlated, but their patterns of correlations with other variables are not consistent. The total price impact π_{200} is strongly correlated with $\overline{s}^{Time}/2$ and $\overline{c}$, but not with the λ. This may reflect the choice of 200 shares as the reference trade. The total cost for a trade of this size would reflect mostly the transient cost. Correlations in the Nasdaq subsample (Panel C) are similar to those in the full and NYSE/Amex samples.

In summary, there is strong positive correlation between $\overline{s}^{Time}/2$ and $\overline{c}$, and moderate positive correlation between λ and R_m^2, but neither variable from the former group is consistently positively correlated with a variable in the latter group. Thus, spread and impact measures appear to be driven by different underlying factors. From an economic perspective, the spread often presumed to impound impact (information) costs. The absence of a positive relation therefore suggests that variation in spreads is driven mostly by non-impact costs, such as inventory effects. I next turn to the question of whether these variables can be proxied by constructs from daily return and volume data.

c. Proxy relationships

Table 4 presents the correlations between the *TAQ* liquidity measures and the daily liquidity measures. These results bear crucially on the ability of the latter to proxy the former in asset pricing specifications. The layout of the table is similar to that of Table 3. To keep the size of the table manageable, I use a subset of the TAQ liquidity measures, retaining $\overline{c}$, λ, ϕ, PIN , and π_{200}.

Results for the full sample are reported in Panel A of the table. In general, the TAQ measure which can be most readily proxied is the effective cost $\overline{c}$. Both the adjusted moment and Gibbs estimates $\left(c^{MZ} \text{ and } c^{Gibbs}\right)$ are strongly correlated with $\overline{c}$. Between the two proxies, however, the Gibbs estimate appears to have the higher correlations (0.894 vs. 0.818 for the Pearson; 0.809 vs. 0.730 for the Spearman). *L* and *I* also have high Spearman correlations with $\bar{c}$, but their Pearson correlations are substantially lower than those of c^{MZ} and c^{Gibbs}. It therefore appears that $\bar{c}$ be proxied quite well by daily data.

Proxies for price impact and reversal measures, however, are more problematic. The impact coefficient λ is essentially uncorrelated with *L* and *I*. The liquidity and illiquidity measures *L* and *I* have a Pearson correlation near zero (–0.023), but a Spearman correlation of –0.988. The difference probably arises from the extreme observations, and (to a lesser extent) the nonlinearity of the reciprocal relationship. The strong Spearman correlation between *L* and *I* is reflected in the patterns of Spearman correlations with other variables. That is, from a rank-order perspective, *L* and *I* are essentially equivalent.

The reversal coefficient ϕ has a modestly positive Spearman correlation (0.244) with its corresponding proxy γ, but the Pearson correlation is negative. The proxy relationship is much weaker than the simulation analysis of Pastor and Stambaugh would suggest. The apparent inconsistency may be a reflection of the fact that their simulations confined ϕ to the interval (–1,+1), and assumed normal distributions for returns and signed order flows.

The NYSE/Amex subsample (Panel B) exhibits similar correlations for variables already considered, but also reports results for *PIN* and π_{200}. The proxy correlations for *PIN* are not consistent. c^{Gibbs} appears to be the best proxy judging by the Pearson correlation (0.294), but *I* and *L* have larger Spearman correlations. The findings for π_{200} are similar in that c^{MZ} and c^{Gibbs} are the best proxies judging by the Pearson correlation, but *I* and *L* are best judging by the Spearman. Correlations for the Nasdaq subsample (Panel C) are similar to those of the full sample.

In asset pricing applications, market capitalization is a major confounding influence, since it is often cross-sectionally correlated with other firm characteristics. The correlations reported in Table 5 between the liquidity measures and proxies are partial correlations with respect to the logarithm of equity market capitalization. Panel A reports results for the full sample. The partial correlations between $\overline{c}$ and c^{MZ} / c^{Gibbs} remain strongly positive.

Among the measures and proxies considered here, the reversal coefficients ϕ and γ appear to be the most distinctive. ϕ does not correlate strongly with other high-frequency measures, but it also attempts to measure a distinctly different phenomenon (lagged reversal rather than contemporaneous effect). It is also not strongly correlated with its intended proxy variable, γ. This may simply reflect, however, large estimation errors. In this connection, it is worth emphasizing that Pastor and Stambaugh do not use these estimates for individual firms, but form portfolios instead, for which measurement errors should be much lower. Furthermore, Pastor and Stambaugh employ data filters that are likely to reduce the incidence of extreme values for these measures. For example, stocks priced under five dollars are excluded, and estimates are computed only for NYSE stocks.

7. The Gibbs estimates in a broader sample

In view of the strong performance of the Gibbs effective cost estimates in the TAQ/CRSP comparison sample, it is of some interest to investigate the properties of these estimates over the full historical sample (beginning in 1963) for which daily CRSP data are available. To this end, annual estimates of the daily-based trading cost estimates and proxies were computed for all firms in the daily CRSP file. Firms with fewer than fifty days with trades in a given year were excluded for that year.

Nasdaq closing prices are not extensively reported on the CRSP database until the middle of 1982 (with Nasdaq's introduction of the National Market System). Due to relatively small numbers of stocks, however, the Nasdaq estimates developed in this paper are only reported beginning in 1985. The CRSP Nasdaq sample also changed markedly in 1992 with the inclusion of the Nasdaq SmallCap market.

Figure 1 plots the annual average c^{Gibbs} values for exchange and market capitalization subsamples. As in Fama and French (1992), NYSE/Amex breakpoints are also used for Nasdaq sample.

The NYSE/Amex estimates provide a more complete picture of the long-run time-series variation. Although the series appears roughly stationary, there is substantial volatility, with the largest peak occurring around 1975. In 1975, commission levels dropped following the SEC's deregulation. It is possible that liquidity suppliers increased posted and effective spreads to compensate for decreased commission revenue. Another possible explanation is short-run stickiness in absolute dollar spreads. Most market indices dropped over 1974. At the new lower price levels, relative spreads would be higher.

For both NYSE/Amex and Nasdaq firms, time variation in effective costs is concentrated in the lowest-capitalization subsample. This is particularly true for the Nasdaq lowest-capitalization firms, for which average effective cost goes from around one percent in the early 1980's to roughly four percent in the early 1990's. This may in part reflect the changes in composition of the Nasdaq population. Smaller, but still quite noticeable variation in effective cost occurs in the other Nasdaq capitalization quartiles. The NYSE/Amex firms in the lowest capitalization subsample have effective costs that vary approximately between 0.5% and 1.5%. Most notably, however, there has been no dramatic change in effective costs for the higher capitalization quartiles.

8. Liquidity and stock returns

This section examines the relation between expected returns and liquidity, viewed as a characteristic and proxied by one or more of the daily-based measures c^{Gibbs}, I, L, and γ. In studies that focus on a single liquidity proxy, asset pricing tests usually follow the Fama and MacBeth (1973) approach. This requires the formation of portfolios based on size (or beta) and the liquidity proxy. Since the present study aims at an impartial evaluation of a set of proxies, however, approaches that require portfolio construction are undesirable. As an alternative, I follow the approach of Brennan, Chordia, and Subrahmanyam (1998) (BCS).

The BCS procedure may be summarized as follows. The starting point is an approximate factor model in which the return on the *j*th security is given by:

$$R_{jt} = ER_j + \beta_j f_t + e_{jt} , \quad (14)$$

where f_t is a vector of factor realizations at time t, and β_j contains the factor loadings for security j. The APT implies $ER_j - R_F = \beta_j \lambda$, where λ is the vector of factor risk premia, and that realized returns are:

$$R_{jt} - R_{Ft} = \beta_j F_t + e_{jt} \quad (15)$$

where $F_t = \lambda + f_t$. The key question is the extent to which the security characteristics can explain the residual in eq (15). To implement the test, estimates of the factor loadings, denoted $\hat{\beta}_j$, are computed using data prior to time t. The implied risk-adjusted returns are then computed as

$$R^*_{jt} = R_{jt} - \left(R_{Ft} + \hat{\beta}_j F_t \right) \quad (16)$$

Denote by Z_{jt} a vector of predetermined characteristics for security j. In the present application, Z consists mainly of a constant and liquidity proxies. At each t, the risk-adjusted returns are then regressed against this set:

$$R^*_{jt} = d_t Z_{jt} + \tilde{e}_{jt} \quad (17)$$

Under the null hypothesis that the characteristics do not affect returns $E\left[d_t\right] = 0$. Denote by $\hat{d}_t$ the OLS estimate of d_t. BCS suggest two approaches to summarizing the time series of these estimates. The raw overall estimate, denoted $\hat{d}_r$, is simply the average. Alternatively, to alleviate concerns arising from estimation errors in the factor loadings $\hat{\beta}_j$, BCS propose a purged estimator, denoted $\hat{d}_P$. An element of this vector $\hat{d}_{kP}$ is computed as the intercept in a time series regression of $\hat{d}_{kt}$ on the factor realizations F_t.

The full set of characteristics includes the liquidity proxies $\left(c^{Gibbs}, L, I, \gamma \right)$ and other variables suggested by BCS: the log market capitalization, *LMZ*; the lagged return for the stock over the second and third prior month, *RET23*; the return over lagged months four through six, *RET46*; and the return over lagged months seven through twelve, *RET712*. Table 6 reports the descriptive statistics for the variables used in this analysis. Values in the table are time-series averages of statistics computed cross-sectionally for each monthly time period. The correlations are similar those encountered in the comparison sample. The proxy measures are moderately positively correlated with each other and capitalization, except for those involving the liquidity ratio *L* (which has, conceptually, the opposite sign of the others). Patterns across NYSE/Amex and Nasdaq samples are similar.

Table 7 reports the regression estimates of (17). All specifications include and intercept and *LMV*. The first four specifications *(a)-(d)* incorporate one liquidity proxy at a time. In the estimates for the NYSE/Amex sample, the c^{Gibbs} and *I* coefficients have the anticipated sign and significance in the specifications where they are included one at a time. In the Nasdaq sample, this is only the case for c^{Gibbs}. These findings suggest that c^{Gibbs} is the best single proxy. In specification (e), however, which includes all proxies, the picture is less clear. No single proxy is the obvious winner across both samples.

Relative to the others, the c^{Gibbs} measure possesses the virtue of an economically interpretable magnitude. This enables us to address the reasonableness of the coefficient. In the NYSE/Amex sample, the coefficient of c^{Gibbs} is approximately 0.3. This implies that a stock with an average effective cost of one percent would have a monthly expected liquidity premium of 30 basis points (3.6% on an annual basis). This might seem high, but a one percent effective cost is well above average in this sample. Figure 1 suggests that this level is exceeded (on average) only in the lowest market capitalization quartile, and here only a small portion of the time. It should also be noted that the effective cost is generally a fraction of the posted bid-ask spread.

In the Nasdaq sample, the coefficient of c^{Gibbs} is approximately 0.2. Although this point estimate is lower than the NYSE/Amex value, Nasdaq effectives costs are much higher. Figure 1 suggests that an effective cost of two percent would not be extreme in the lowest Nasdaq quartile (and this is using NYSE breakpoints). A two percent effective cost would imply a monthly liquidity premium of forty basis points.

While the overall coefficient magnitudes are reasonable, however, previous research suggests an interaction between seasonality and liquidity coefficients. To assess this, Figure 2 plots the monthly averages for the coefficients of the liquidity proxies. The vertical bar depicts ± two-standard-error bounds. The overall mean and error bounds are shown as horizontal lines. It should be borne in mind that since both monthly and overall averages are computed from the same data the estimates are not independent. Vertical scales for NYSE/Amex and Nasdaq samples differ.

All measures show some evidence of seasonality. There is typically a January elevation, and arguably in some cases a December depression. The pattern is strongest for the c^{Gibbs} coefficients. For this variable, it appears that the liquidity premium is concentrated in January. This is similar to the results obtained by Eleswarapu and Reinganum (1993), with a different sample and liquidity measure. The reasons for the seasonality are unclear. Portfolio rebalancing around the turn of the year could possibly explain the direction of the finding. It the last trade of the year tended to be at the bid, which January trades tended to be at the ask, this would induce an increased return associated with the spread. The magnitude of the January coefficient, however, seems too large. The coefficient implies that the January return is elevated by approximately twice the effective spread. For this to occur, all closing trades in the preceding year would have to have been at the bid price, and all closing trades in the January at the ask.

9. Conclusion

This study moves from analysis of microstructure-based liquidity measures, to evaluation of liquidity proxies computed from daily data, and finally the use of these proxies as characteristics in expected return specifications. There are significant gains in sample size from using daily proxies in lieu of high-frequency measures: the latter are only generally available back to 1983 (the start of ISSM); the former back to 1962 (the start of the CRSP daily file). The study conducts a critical examination of the correlations between the daily proxies and the underlying high-frequency measures, and of the concordance among the latter. These analyses are performed in a comparison sample, consisting of 200 randomly-chosen firms in each of the years 1993-2002, for which both high-frequency and daily data exist.

There exists no single comprehensive measure of liquidity. The microstructure measures constructed here include posted spreads, effective costs, and measures based on dynamic models of prices and signed trades, as well as measures developed by others: the probability of informed trading (*PIN*, Easley, Hvidkjaer, and O'Hara (2002)), and the permanent and transitory cost measures suggested in Sadka (2004). All measures except the ones normalized to the unit interval (PIN and R_m^2) exhibit extreme values. While it is impossible to rule out the possibility that some of these are spurious, it is also likely that trading costs for some companies are truly very high.

Posted spreads (both intraday and closing) and average effective costs are relatively easy to estimate and interpret. They are also highly correlated. The measures derived from dynamic trade and price models, while arguably more comprehensive, are more difficult to estimate and interpret. The correlations within this set suggest modest concordance at best. Reversal measures, which summarize the effect of lagged order flow on future expected returns, appear to be the least correlated with the other measures.

It is quite feasible to estimate effective costs from daily return data. From a year's worth of data, both the Roll estimator (with infeasible estimates set to zero) and the Gibbs estimator perform well. High-frequency dynamic price impact measures of liquidity are more difficult. None of the daily proxies achieves a strong and consistent correlation. Reversal measures are also problematic.

When the daily liquidity proxies are introduced into asset pricing tests modeled on Brennan, Chordia, and Subrahmanyam (1998), both the Gibbs estimate of effective cost and the illiquidity ratio are positively correlated with risk-adjusted returns in the NYSE/Amex sample. In the Nasdaq sample, only the Gibbs estimate is positively correlated. These results provide modest support for the hypothesis that trading cost is a priced characteristic. Moreover, the relation between returns and the liquidity proxies is markedly seasonal for both NYSE/Amex and Nasdaq firms, with the strongest effect arising in January.

There are a number of promising directions for future research. First, since the Gibbs estimate of the effective cost relies solely on the transaction price record, the technique can readily be applied to historical and international settings where only trade prices are available. The present application is to daily data, but there is in principle no reason why the approach would not be useful in weekly or monthly data. Of course, as the frequency drops, drift and diffusion in the efficient price become more pronounced relative to the effective cost, and hence the signal-to-noise ratio is likely to be lower.

A second line of inquiry is refinement of the Gibbs estimation procedure. It seems particularly worthwhile to consider estimation of *c* jointly with β. The estimates of *c* should be improved because the market return is a useful signal in estimating the change in the efficient price $\left(\Delta m_t = u_t\right)$, which is here taken as unconditionally normal. The estimate of β should also be improved, however, because the specification essentially purges the price change of bid-ask bounce in the firm's return.

10. References

Acharya, V. V., Pedersen, L. H., 2002. Asset pricing with liquidity risk. Unpublished working paper. Stern School of Business.

Amihud, Y., Mendelson, H., Lauterbach, B., 1997. Market microstructure and securities values: evidence from the Tel Aviv Exchange. Journal of Financial Economics 45, 365-390.

Amihud, Y., 2002. Illiquidity and stock returns: cross-section and time-series effects . Journal of Financial Markets 5, 31-56.

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

Amihud, Y., Mendelson, H., 1989. The Effects of Beta, Bid-Ask Spread, Residual Risk, and Size on Stock Returns. Journal of Finance 44, 479-486.

Berkman, H., Eleswarapu, V. R., 1998. Short-term traders and liquidity: a test using Bombay Stock Exchange data. Journal of Financial Economics 47, 339-355.

Brennan, M. J., Chordia, T., Subrahmanyam, A., 1998. Alternative factor specifications, security characteristics, and the cross-section of expected stock returns. Journal of Financial Economics 49, 345-373.

Campbell, J., Grossman, S., Wang, J., 1993. Trading volume and serial correlation in stock returns. Quarterly Journal of Economics 108, 905-939.

Chalmers, J. M. R., Kadlec, G. B., 1998. An empirical examination of the amortized spread. Journal of Financial Economics 48, 159-188.

Chan, L. K. C., Lakonishok, J., 1997. Institutional equity trading costs: NYSE versus Nasdaq. Journal of Finance 52, 713-35.

Conrad, J., Johnson, K. M., Wahal, S., 2001. Alternative trading systems. Unpublished working paper. University of North Carolina, Kenan-Flagler Business School.

Constantinides, G. M., 1986. Capital market equilibrium with transaction costs. Journal of

Political Economy 94, 842-862.

Cooper, S. K., Groth, J. C., Avera, W. E., 1985. Liqudity, exchange listing and common stock performance. Journal of Economics and Business 37, 19-33.

Easley, D., Hvidkjaer, S., O'Hara, M., 2002. Is information risk a determinant of asset returns? Journal of Finance 57, 2185-2221.

Easley, D., Kiefer, N. M., O'Hara, M., 1997. One day in the life of a very common stock. Review of Financial Studies 10, 805-835.

Easley, D., Kiefer, N. M., O'Hara, M., Paperman, J., 1996. Liquidity, information and infrequently traded stocks. Journal of Finance 51, 1405-1436.

Easley, D., O'Hara, M., 1987. Price, trade size, and information in securities markets. Journal of Financial Economics 19, 69-90.

Easley, D., O'Hara, M., 1992. Time and the process of security price adjustment. Journal of Finance 47, 576-605.

Easley, D., O'Hara, M., 2002. Microstructure and asset pricing. In: Constantinides, G., Harris, M., and Stulz, R. (Eds.), Handbook of Financial Economics. Elsevier, New York.

Eleswarapu, V. R., 1997. Cost of transacting and expected returns in the Nasdaq market. Journal of Finance 52, 2113-2127.

Eleswarapu, V. R., Reinganum, M. R., 1993. The seasonal behavior of the liquidity premium in asset pricing. Journal of Financial Economics 34, 373-386.

Fama, E. F., French, K. R., 1992. The cross-section of expected stock returns. Journal of Finance 47, 427-465.

Fama, E. F., MacBeth, J. D., 1973. Risk, return and equilibrium. Journal of Political Economy 71, 607-636.

Gabaix, X., Gopikrishnan, P., Plerou, V., Stanley, H. E., 2003. A theory of power law distributions in financial market fluctuations. Nature 423, 267-270.

Glosten, L. R., Harris, L. E., 1988. Estimating the components of the bid/ask spread. Journal of Financial Economics 21, 123-42.

Glosten, L. R., Milgrom, P. R., 1985. Bid, ask, and transaction prices in a specialist market with heterogeneously informed traders. Journal of Financial Economics 14, 71-100.

Harris, L., 1990. Statistical Properties of the Roll Serial Covariance Bid/Ask Spread Estimator. Journal of Finance 45, 579-590.

Harris, L. E., Hasbrouck, J., 1996. Market vs. limit orders: the SuperDOT evidence on order submission strategy. Journal of Financial and Quantitative Analysis 31, 213-31.

Hasbrouck, J., 1991. The summary informativeness of stock trades: An econometric analysis. Review of Financial Studies 4, 571-95.

Hasbrouck, J., 2003. Liquidity in the futures pits: Inferring market dynamics from incomplete data. Journal of Financial and Quantitative Analysis, forthcoming.

Heaton, J., Lucas, D. J., 1996. Evaluating the effects of incomplete markets on risk sharing and asset pricing. Journal of Political Economy 104, 443-487.

Kadlec, G. B., McConnell, J. J., 1994. The effect of market segmentation and illiquidity on asset prices: evidence from exchange listings. Journal of Finance 49, 611-636.

Keim, D. B., and A. Madhavan. 1995. Anatomy of the Trading Process: Empirical Evidence on the Behavior of Institutional Traders. *Journal of Financial Economics; 37(3), March 1995, Pages 371-98.*

Kempf, A., Korn, O., 1999. Market depth and order size. Journal of Financial Markets 2, 29-48.

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

Merton, R., 1980. On estimating the expected rate of return on the market. Journal of Financial Economics 8, 323-362.

Pastor, L., Stambaugh, R. F., 2003. Liquidity risk and expected stock returns. Journal of Political Economy 111, 642-685.

Perold, A., 1988. The implementation shortfall: Paper vs. reality. Journal of Portfolio Management 14, 4-9.

Roll, R., 1984. A simple implicit measure of the effective bid-ask spread in an efficient market. Journal of Finance 39, 1127-1139.

Sadka, R., 2004. Liquidity risk and asset pricing. Unpublished working paper. University of Washington.

Stoll, H. R., 2000. Friction. Journal of Finance 55, 1479-1514.

Stoll, H. R., Whalley, R. H., 1983. Transaction cost and the small firm effect. Journal of Financial Economics 12, 57-79.

Vayanos, D., 1998. Transaction costs and asset prices: a dynamic equilibrium model. Review of Financial Studies 11, 1-58.

Table 1. Summary of trading cost and liquidity measures

Panel A. Measures derived from transaction-level (TAQ) data

$\bar{s}^{Time}/2$	Time-weighted average half-spread: $\bar{s}^{Time}/2 = \overline{(a_k - b_k)}/2$, where a_k and b_k are the log ask and bid prices established at the *k*th quote update. The average is weighted by the time the *k*th quote was in effect.
$\bar{s}^{Close}/2$	Average of daily closing half-spreads: $\bar{s}^{Close}/2 = \overline{(a_d^{Close} - b_d^{Close})}/2$ where a_d^{Close} and b_d^{Close} are the log ask and bid prices prevailing at the close of day *d*, averaged over all the days in sample.
$\bar{c}$	Average effective cost: $\bar{c} = \overline{\lvert p_k - m_k \rvert}$ where p_k is the log trade price of the *k*th trade, m_k is the log prevailing quote midpoint and the average is dollar-value weighted.
λ	Trade price impact coefficient estimated from the specification $m_t = m_{t-1} + \lambda Z_t + u_t$; *t* indexes five-minute intraday intervals; m_t is the log quote midpoint prevailing at the close of the interval; $Z_t = \sum_{k \in N_t} q_k \log\left(V_k^{Dollar}\right)$ where *k* indexes trades occurring in the interval, q_k is the imputed sign of the trade, and V_j^{Dollar} is the dollar volume of the trade
R_m^2	Coefficient of determination in the above regression (annual average of monthly values).
ϕ	Signed reversal coefficient, estimated from $r_t^e = \phi\left(X_{t-1} - X_t\right) + u_t$ where r_t^e is the excess return on day *t* and X_t is the day's signed volume: $X_t = \sum_{k \in N_t} q_k V_k^{Dollar}$ where *k* indexes trades occurring on day *t*, q_k is the imputed sign of the trade, and V_k^{Dollar} is the dollar volume of the trade.
PIN	Probability of informed trading Easley, Hvidkjaer, and O'Hara (2002). Estimates kindly supplied by Soren Hvidkjaer.
π_{200}^P	Permanent proportional price impact associated with a 200-share buy order (Sadka (2004)). Based on estimates kindly supplied by Ronnie Sadka. Annual average of monthly values.
π_{200}^T	Temporary proportional price impact associated with a 200-share buy order (Sadka (2004)). Based on estimates kindly supplied by Ronnie Sadka. Annual average of monthly values.
π_{200}	Total proportional price impact associated with a 200-share buy order: $\pi_{200} = \pi_{200}^P + \pi_{200}^T$

Table 1. Summary of trading cost and liquidity measures (continued)

Panel B. Measures derived from daily (CRSP) data.

c^{Gibbs}	Gibbs estimate of effective cost derived from the Roll model.
c^{M}	Moment estimate of effective cost derived from the Roll model; reported only if this estimate is feasible.
c^{MZ}	Moment estimate of effective cost derived from the Roll model; set to zero if this estimate is infeasible.
L	Liquidity ratio: $L = \overline{Vol_d / \lvert r_d \rvert}$
I	Illiquidity ratio (Amihud (2002)): $I = \overline{\lvert r_d \rvert / Vol_d}$
γ	Reversal coefficient (Pastor and Stambaugh (2003)), estimated from: $r_{t+1}^{e} = \theta + \phi r_t + \gamma \, sign\left(r_t^{e}\right) V_t^{Dollar} + \varepsilon_t$

Table 2. Descriptive statistics for the comparison sample

The comparison sample consists of firms randomly selected those present on both TAQ and CRSP databases. 200 firms are chosen from each of the ten years 1993-2002. Variables are defined as follows (full definitions given in Table 1): $\bar{s}^{Time}/2$, time-weighted relative half-spread; $\bar{s}^{Close}/2$, closing spread; $\bar{c}$, average effective cost; λ, price impact coefficient; R_m^2, R^2 in regression of quote changes on signed order flow; ϕ, reversal coefficient for signed order flow; *PIN*, probability of informed trading; π_{200}^P, permanent impact of a 200-share trade; π_{200}^T, temporary impact of a 200-share trade; π_{200}, total impact of a 200-share trade; c^M, moment estimate of Roll model (with infeasible estimates deleted); c^{MZ}, moment estimate of Roll model (with infeasible estimates set to zero); c^{Gibbs}, Gibbs estimate of Roll model; *L*, liquidity ratio; *I*, illiquidity ratio; γ, reversal coefficient of unsigned volume.

Source	Variable	Sample	N	Mean	Median	Standard Deviation	Skew	Kurt
TAQ	$\bar{s}^{Time}/2$	All	1,998	0.0183	0.0116	0.0209	2.90	13.76
		NYSE/Amex	793	0.0100	0.0051	0.0159	4.44	25.93
		Nasdaq	1,205	0.0238	0.0179	0.0219	2.71	12.91
	$\bar{s}^{Close}/2$	All	1,998	0.0247	0.0146	0.0304	2.85	11.73
		NYSE/Amex	793	0.0101	0.0051	0.0160	4.46	25.84
		Nasdaq	1,205	0.0342	0.0234	0.0337	2.54	9.27
	$\bar{c}$	All	1,998	0.0124	0.0076	0.0155	4.12	29.21
		NYSE/Amex	793	0.0072	0.0037	0.0116	4.95	34.06
		Nasdaq	1,205	0.0158	0.0112	0.0167	4.02	28.25
	λ	All	1,938	1.6259	1.1374	1.8508	5.65	69.62
		NYSE/Amex	780	2.2629	1.6023	2.4407	5.17	50.49
		Nasdaq	1,158	1.1968	0.9395	1.1252	2.31	9.17
	R_m^2	All	1,938	0.1764	0.1666	0.0949	0.78	0.68
		NYSE/Amex	780	0.2046	0.1941	0.0827	0.80	1.43
		Nasdaq	1,158	0.1575	0.1337	0.0979	1.03	0.89
	ϕ	All	1,997	-11.8458	0.2538	620.4580	-16.39	556.18
		NYSE/Amex	793	-47.8823	-0.0292	814.8165	-19.78	448.77
		Nasdaq	1,204	11.8892	1.8287	447.4543	6.40	144.84
	PIN	All	738	0.1961	0.1809	0.0887	1.55	4.08
		NYSE/Amex	738	0.1961	0.1809	0.0887	1.55	4.08
		Nasdaq						
	π_{200}^P	All	521	0.0010	0.0006	0.0011	2.04	4.52
		NYSE/Amex	521	0.0010	0.0006	0.0011	2.04	4.52
		Nasdaq						
	π_{200}^T	All	521	0.0026	0.0014	0.0036	3.78	17.49
		NYSE/Amex	521	0.0026	0.0014	0.0036	3.78	17.49
		Nasdaq						
	π_{200}	All	521	0.0036	0.0022	0.0043	3.39	14.85
		NYSE/Amex	521	0.0036	0.0022	0.0043	3.39	14.85
		Nasdaq						

Table 2. Descriptive statistics for the comparison sample (continued)

The comparison sample consists of firms randomly selected those present on both TAQ and CRSP databases. 200 firms are chosen from each of the ten years 1993-2002. Variables are defined as follows (full definitions given in Table 1): $\bar{s}^{Time}/2$, time-weighted relative half-spread; $\bar{s}^{Close}/2$, closing spread; $\bar{c}$, average effective cost; λ, price impact coefficient; R_m^2, R^2 in regression of quote changes on signed order flow; ϕ, reversal coefficient for signed order flow; *PIN*, probability of informed trading; π_{200}^P, permanent impact of a 200-share trade; π_{200}^T, temporary impact of a 200-share trade; π_{200}, total impact of a 200-share trade; c^M, moment estimate of Roll model (with infeasible estimates deleted); c^{MZ}, moment estimate of Roll model (with infeasible estimates set to zero); c^{Gibbs}, Gibbs estimate of Roll model; *L*, liquidity ratio; *I*, illiquidity ratio; γ, reversal coefficient of unsigned volume.

Source	Variable	Sample	N	Mean	Median	Standard Deviation	Skew	Kurt
CRSP	c^M	All	1,396	0.0178	0.0136	0.0148	1.82	4.40
		NYSE/Amex	430	0.0104	0.0064	0.0120	3.15	11.95
		Nasdaq	966	0.0210	0.0172	0.0147	1.72	4.15
	c^{MZ}	All	1,990	0.0123	0.0083	0.0147	1.83	4.35
		NYSE/Amex	794	0.0056	0.0021	0.0102	3.75	18.29
		Nasdaq	1,196	0.0168	0.0138	0.0155	1.43	2.97
	c^{Gibbs}	All	1,990	0.0121	0.0066	0.0145	3.09	14.32
		NYSE/Amex	794	0.0058	0.0034	0.0083	4.94	31.97
		Nasdaq	1,196	0.0162	0.0103	0.0162	2.72	11.34
	L	All	1,997	0.4876	0.0094	4.7203	24.55	669.01
		NYSE/Amex	793	0.5676	0.0463	1.7057	6.42	53.79
		Nasdaq	1,204	0.4350	0.0042	5.9200	20.59	449.04
	I	All	1,989	0.9668	0.0715	4.2970	11.70	172.57
		NYSE/Amex	793	0.8661	0.0180	5.0608	11.86	166.21
		Nasdaq	1,196	1.0335	0.1468	3.7056	10.43	138.43
	γ	All	1,997	17.7502	0.2078	324.8927	3.05	230.66
		NYSE/Amex	793	9.2409	0.0016	301.9898	2.36	161.35
		Nasdaq	1,204	23.3548	1.1538	339.1403	3.34	256.20
	LMV	All	2,000	11.8435	11.6720	2.0049	0.41	0.14
		NYSE/Amex	794	12.8309	12.9329	2.1335	-0.17	-0.17
		Nasdaq	1,206	11.1934	11.1208	1.6148	0.57	1.58

Table 3. Correlations in high-frequency liquidity measures, comparison sample

The comparison sample consists of firms randomly selected those present on both TAQ and CRSP databases. 200 firms are chosen from each of the ten years 1993-2002. Each entry in the correlation matrix contains the Pearson correlation (top row), Spearman correlation (italics, middle row), and number of observations (bottom row). Variables are defined as follows (full definitions given in Table 1): $\bar{s}^{Time}/2$, time-weighted relative half-spread; $\bar{c}$, average effective cost; λ, price impact coefficient; R_m^2, R^2 in regression of quote changes on signed order flow; ϕ, reversal coefficient for signed order flow; *PIN*, probability of informed trading; π_{200}, total impact of a 200-share trade.

Panel A. Full sample.

	$\bar{s}^{Time}/2$	$\bar{c}$	λ	R_m^2	ϕ
$\bar{s}^{Time}/2$	1.000	0.940	-0.119	-0.370	-0.090
	1.000	*0.975*	*-0.146*	*-0.490*	*0.161*
	1,998	1,998	1,936	1,936	1,997
$\bar{c}$	0.940	1.000	-0.090	-0.322	-0.105
	0.975	*1.000*	*-0.114*	*-0.464*	*0.126*
	1,998	1,998	1,936	1,936	1,997
λ	-0.119	-0.090	1.000	0.458	-0.011
	-0.146	*-0.114*	*1.000*	*0.506*	*-0.135*
	1,936	1,936	1,936	1,936	1,936
R_m^2	-0.370	-0.322	0.458	1.000	0.004
	-0.490	*-0.464*	*0.506*	*1.000*	*-0.163*
	1,936	1,936	1,936	1,936	1,936
ϕ	-0.090	-0.105	-0.011	0.004	1.000
	0.161	*0.126*	*-0.135*	*-0.163*	*1.000*
	1,997	1,997	1,936	1,936	1,997

Table 3. Correlations in high-frequency liquidity measures, comparison sample (continued)

The comparison sample consists of firms randomly selected those present on both TAQ and CRSP databases. 200 firms are chosen from each of the ten years 1993-2002. Each entry in the correlation matrix contains the Pearson correlation (top row), Spearman correlation (italics, middle row), and number of observations (bottom row). Variables are defined as follows (full definitions given in Table 1): $\bar{s}^{Time}/2$, time-weighted relative half-spread; $\bar{c}$, average effective cost; λ, price impact coefficient; R_m^2, R^2 in regression of quote changes on signed order flow; ϕ, reversal coefficient for signed order flow; *PIN*, probability of informed trading; π_{200}, total impact of a 200-share trade.

Panel B. NYSE/Amex

	$\bar{s}^{Time}/2$	$\bar{c}$	λ	R_m^2	ϕ	*PIN*	π_{200}
$\bar{s}^{Time}/2$	1.000	0.957	0.039	-0.020	-0.394	0.389	0.879
	1.000	*0.979*	*0.313*	*-0.017*	*-0.003*	*0.661*	*0.971*
	793	793	780	780	793	738	520
$\bar{c}$	0.957	1.000	0.060	0.015	-0.378	0.397	0.865
	0.979	*1.000*	*0.318*	*-0.004*	*-0.013*	*0.660*	*0.951*
	793	793	780	780	793	738	520
λ	0.039	0.060	1.000	0.670	0.014	0.472	-0.041
	0.313	*0.318*	*1.000*	*0.670*	*-0.051*	*0.498*	*0.110*
	780	780	780	780	780	725	520
R_m^2	-0.020	0.015	0.670	1.000	0.035	0.331	-0.223
	-0.017	*-0.004*	*0.670*	*1.000*	*-0.022*	*0.214*	*-0.220*
	780	780	780	780	780	725	520
ϕ	-0.394	-0.378	0.014	0.035	1.000	-0.041	-0.299
	-0.003	*-0.013*	*-0.051*	*-0.022*	*1.000*	*-0.003*	*0.020*
	793	793	780	780	793	738	520
PIN	0.389	0.397	0.472	0.331	-0.041	1.000	0.309
	0.661	*0.660*	*0.498*	*0.214*	*-0.003*	*1.000*	*0.576*
	738	738	725	725	738	738	492
π_{200}	0.879	0.865	-0.041	-0.223	-0.299	0.309	1.000
	0.971	*0.951*	*0.110*	*-0.220*	*0.020*	*0.576*	*1.000*
	520	520	520	520	520	492	521

Table 3. Correlations in high-frequency liquidity measures, comparison sample (continued)

The comparison sample consists of firms randomly selected those present on both TAQ and CRSP databases. 200 firms are chosen from each of the ten years 1993-2002. Each entry in the correlation matrix contains the Pearson correlation (top row), Spearman correlation (italics, middle row), and number of observations (bottom row). Variables are defined as follows (full definitions given in Table 1): $\bar{s}^{Time}/2$, time-weighted relative half-spread; $\bar{c}$, average effective cost; λ, price impact coefficient; R_m^2, R^2 in regression of quote changes on signed order flow; ϕ, reversal coefficient for signed order flow; *PIN*, probability of informed trading; π_{200}, total impact of a 200-share trade.

Panel C. Nasdaq

	$\bar{s}^{Time}/2$	$\bar{c}$	λ	R_m^2	ϕ
$\bar{s}^{Time}/2$	1.000	0.928	-0.125	-0.463	0.117
	1.000	*0.958*	*-0.201*	*-0.602*	*0.151*
	1,205	1,205	1,156	1,156	1,204
$\bar{c}$	0.928	1.000	-0.108	-0.411	0.071
	0.958	*1.000*	*-0.166*	*-0.574*	*0.099*
	1,205	1,205	1,156	1,156	1,204
λ	-0.125	-0.108	1.000	0.212	-0.048
	-0.201	*-0.166*	*1.000*	*0.415*	*-0.117*
	1,156	1,156	1,156	1,156	1,156
R_m^2	-0.463	-0.411	0.212	1.000	-0.006
	-0.602	*-0.574*	*0.415*	*1.000*	*-0.173*
	1,156	1,156	1,156	1,156	1,156
ϕ	0.117	0.071	-0.048	-0.006	1.000
	0.151	*0.099*	*-0.117*	*-0.173*	*1.000*
	1,204	1,204	1,156	1,156	1,204

Table 4. Correlations between liquidity measures in the comparison sample.

The comparison sample consists of firms randomly selected those present on both TAQ and CRSP databases. 200 firms are chosen from each of the ten years 1993-2002. Each entry in the correlation matrix contains the Pearson correlation (top row), Spearman correlation (italics, middle row), and number of observations (bottom row). Unshaded variables are constructed from high-frequency data; shaded variables are constructed from daily data. Variables are defined as follows (full definitions given in Table 1): $\bar{c}$, average effective cost; λ, price impact coefficient; ϕ, reversal coefficient for signed order flow; *PIN*, probability of informed trading; π_{200} , total impact of a 200-share trade; c^{MZ}, moment estimate of Roll model (with infeasible estimates set to zero); c^{Gibbs}, Gibbs estimate of Roll model; L, liquidity ratio; I, illiquidity ratio; γ, reversal coefficient of unsigned volume.

Panel A. Full sample

	$\bar{c}$	λ	ϕ	c^{MZ}	c^{Gibbs}	L	I	γ
$\bar{c}$	1.000	-0.090	-0.105	0.818	0.894	-0.031	0.619	0.046
	1.000	*-0.114*	*0.126*	*0.730*	*0.809*	*-0.911*	*0.923*	*0.316*
	1,998	1,936	1,997	1,989	1,989	1,997	1,989	1,997
λ	-0.090	1.000	-0.011	-0.190	-0.153	-0.026	-0.033	-0.016
	-0.114	*1.000*	*-0.135*	*-0.273*	*-0.311*	*-0.038*	*0.042*	*-0.169*
	1,936	1,936	1,936	1,936	1,936	1,936	1,936	1,936
ϕ	-0.105	-0.011	1.000	-0.075	-0.080	0.001	-0.297	-0.185
	0.126	*-0.135*	*1.000*	*0.127*	*0.122*	*-0.104*	*0.111*	*0.244*
	1,997	1,936	1,997	1,989	1,989	1,997	1,989	1,997
c^{MZ}	0.818	-0.190	-0.075	1.000	0.876	-0.030	0.508	0.074
	0.730	*-0.273*	*0.127*	*1.000*	*0.861*	*-0.646*	*0.679*	*0.326*
	1,989	1,936	1,989	1,990	1,990	1,989	1,989	1,989
c^{Gibbs}	0.894	-0.153	-0.080	0.876	1.000	-0.025	0.545	0.113
	0.809	*-0.311*	*0.122*	*0.861*	*1.000*	*-0.719*	*0.762*	*0.324*
	1,989	1,936	1,989	1,990	1,990	1,989	1,989	1,989
L	-0.031	-0.026	0.001	-0.030	-0.025	1.000	-0.009	-0.002
	-0.911	*-0.038*	*-0.104*	*-0.646*	*-0.719*	*1.000*	*-0.956*	*-0.296*
	1,997	1,936	1,997	1,989	1,989	1,997	1,989	1,997
I	0.619	-0.033	-0.297	0.508	0.545	-0.009	1.000	0.077
	0.923	*0.042*	*0.111*	*0.679*	*0.762*	*-0.956*	*1.000*	*0.315*
	1,989	1,936	1,989	1,989	1,989	1,989	1,989	1,989
γ	0.046	-0.016	-0.185	0.074	0.113	-0.002	0.077	1.000
	0.316	*-0.169*	*0.244*	*0.326*	*0.324*	*-0.296*	*0.315*	*1.000*
	1,997	1,936	1,997	1,989	1,989	1,997	1,989	1,997

Table 4. Correlations between liquidity measures in the comparison sample. (continued)

The comparison sample consists of firms randomly selected those present on both TAQ and CRSP databases. 200 firms are chosen from each of the ten years 1993-2002. Each entry in the correlation matrix contains the Pearson correlation (top row), Spearman correlation (italics, middle row), and number of observations (bottom row). Unshaded variables are constructed from high-frequency data; shaded variables are constructed from daily data. Variables are defined as follows (full definitions given in Table 1): $\bar{c}$, average effective cost; λ, price impact coefficient; ϕ, reversal coefficient for signed order flow; *PIN*, probability of informed trading; π_{200} , total impact of a 200-share trade; c^{MZ}, moment estimate of Roll model (with infeasible estimates set to zero); c^{Gibbs}, Gibbs estimate of Roll model; *L*, liquidity ratio; *I*, illiquidity ratio; γ, reversal coefficient of unsigned volume.

Panel B. NYSE/Amex

	$\bar{c}$	λ	ϕ	*PIN*	π_{200}	c^{MZ}	c^{Gibbs}	*L*	*I*	γ
$\bar{c}$	1.000	0.060	-0.378	0.397	0.865	0.832	0.880	-0.026	0.797	-0.048
	1.000	*0.318*	*-0.013*	*0.660*	*0.951*	*0.338*	*0.497*	*-0.910*	*0.905*	*0.044*
	793	780	793	738	520	793	793	793	793	793
λ	0.060	1.000	0.014	0.472	-0.041	-0.103	-0.045	-0.035	-0.016	-0.014
	0.318	*1.000*	*-0.051*	*0.498*	*0.110*	*-0.166*	*-0.217*	*-0.478*	*0.506*	*-0.084*
	780	780	780	725	520	780	780	780	780	780
ϕ	-0.378	0.014	1.000	-0.041	-0.299	-0.351	-0.462	0.003	-0.576	-0.366
	-0.013	*-0.051*	*1.000*	*-0.003*	*0.020*	*-0.002*	*0.001*	*0.021*	*-0.025*	*0.174*
	793	780	793	738	520	793	793	793	793	793
PIN	0.397	0.472	-0.041	1.000	0.309	0.180	0.294	-0.045	0.172	0.037
	0.660	*0.498*	*-0.003*	*1.000*	*0.576*	*0.062*	*0.148*	*-0.716*	*0.757*	*-0.004*
	738	725	738	738	492	738	738	738	738	738
π_{200}	0.865	-0.041	-0.299	0.309	1.000	0.701	0.742	-0.173	0.459	0.335
	0.951	*0.110*	*0.020*	*0.576*	*1.000*	*0.224*	*0.318*	*-0.860*	*0.826*	*0.140*
	520	520	520	492	521	521	521	520	520	520
c^{MZ}	0.832	-0.103	-0.351	0.180	0.701	1.000	0.903	-0.022	0.683	0.020
	0.338	*-0.166*	*-0.002*	*0.062*	*0.224*	*1.000*	*0.716*	*-0.293*	*0.281*	*0.084*
	793	780	793	738	521	794	794	793	793	793
c^{Gibbs}	0.880	-0.045	-0.462	0.294	0.742	0.903	1.000	-0.019	0.694	0.118
	0.497	*-0.217*	*0.001*	*0.148*	*0.318*	*0.716*	*1.000*	*-0.411*	*0.416*	*0.046*
	793	780	793	738	521	794	794	793	793	793
L	-0.026	-0.035	0.003	-0.045	-0.173	-0.022	-0.019	1.000	-0.008	-0.001
	-0.910	*-0.478*	*0.021*	*-0.716*	*-0.860*	*-0.293*	*-0.411*	*1.000*	*-0.972*	*-0.034*
	793	780	793	738	520	793	793	793	793	793
I	0.797	-0.016	-0.576	0.172	0.459	0.683	0.694	-0.008	1.000	-0.054
	0.905	*0.506*	*-0.025*	*0.757*	*0.826*	*0.281*	*0.416*	*-0.972*	*1.000*	*0.040*
	793	780	793	738	520	793	793	793	793	793
γ	-0.048	-0.014	-0.366	0.037	0.335	0.020	0.118	-0.001	-0.054	1.000
	0.044	*-0.084*	*0.174*	*-0.004*	*0.140*	*0.084*	*0.046*	*-0.034*	*0.040*	*1.000*
	793	780	793	738	520	793	793	793	793	793

Table 4. Correlations between liquidity measures in the comparison sample. (continued)

The comparison sample consists of firms randomly selected those present on both TAQ and CRSP databases. 200 firms are chosen from each of the ten years 1993-2002. Each entry in the correlation matrix contains the Pearson correlation (top row), Spearman correlation (italics, middle row), and number of observations (bottom row). Unshaded variables are constructed from high-frequency data; shaded variables are constructed from daily data. Variables are defined as follows (full definitions given in Table 1): $\bar{c}$, average effective cost; λ, price impact coefficient; ϕ, reversal coefficient for signed order flow; *PIN*, probability of informed trading; π_{200} , total impact of a 200-share trade; c^{MZ}, moment estimate of Roll model (with infeasible estimates set to zero); c^{Gibbs}, Gibbs estimate of Roll model; *L*, liquidity ratio; *I*, illiquidity ratio; γ, reversal coefficient of unsigned volume.

Panel C. Nasdaq

	$\bar{c}$	λ	ϕ	c^{MZ}	c^{Gibbs}	L	I	γ
$\bar{c}$	1.000	-0.108	0.071	0.794	0.902	-0.070	0.568	0.080
	1.000	*-0.166*	*0.099*	*0.793*	*0.847*	*-0.909*	*0.899*	*0.374*
	1,205	1,156	1,204	1,196	1,196	1,204	1,196	1,204
λ	-0.108	1.000	-0.048	-0.126	-0.098	-0.058	-0.061	0.020
	-0.166	*1.000*	*-0.117*	*-0.204*	*-0.200*	*0.106*	*-0.035*	*-0.157*
	1,156	1,156	1,156	1,156	1,156	1,156	1,156	1,156
ϕ	0.071	-0.048	1.000	0.088	0.094	-0.002	0.137	-0.014
	0.099	*-0.117*	*1.000*	*0.116*	*0.087*	*-0.122*	*0.106*	*0.256*
	1,204	1,156	1,204	1,196	1,196	1,204	1,196	1,204
c^{MZ}	0.794	-0.126	0.088	1.000	0.854	-0.066	0.498	0.091
	0.793	*-0.204*	*0.116*	*1.000*	*0.880*	*-0.760*	*0.767*	*0.379*
	1,196	1,156	1,196	1,196	1,196	1,196	1,196	1,196
c^{Gibbs}	0.902	-0.098	0.094	0.854	1.000	-0.052	0.595	0.115
	0.847	*-0.200*	*0.087*	*0.880*	*1.000*	*-0.840*	*0.840*	*0.382*
	1,196	1,156	1,196	1,196	1,196	1,196	1,196	1,196
L	-0.070	-0.058	-0.002	-0.066	-0.052	1.000	-0.021	-0.005
	-0.909	*0.106*	*-0.122*	*-0.760*	*-0.840*	*1.000*	*-0.944*	*-0.396*
	1,204	1,156	1,204	1,196	1,196	1,204	1,196	1,204
I	0.568	-0.061	0.137	0.498	0.595	-0.021	1.000	0.179
	0.899	*-0.035*	*0.106*	*0.767*	*0.840*	*-0.944*	*1.000*	*0.397*
	1,196	1,156	1,196	1,196	1,196	1,196	1,196	1,196
γ	0.080	0.020	-0.014	0.091	0.115	-0.005	0.179	1.000
	0.374	*-0.157*	*0.256*	*0.379*	*0.382*	*-0.396*	*0.397*	*1.000*
	1,204	1,156	1,204	1,196	1,196	1,204	1,196	1,204

Table 5. Partial correlations (with respect to log equity market value) between high-frequency and daily liquidity measures in the comparison sample.

The comparison sample consists of firms randomly selected those present on both TAQ and CRSP databases. 200 firms are chosen from each of the ten years 1993-2002. Variable definitions are given in Table 1. Each entry in the correlation matrix contains the partial Pearson correlation (top row) and the partial Spearman correlation (italics, bottom row). Unshaded variables are constructed from high-frequency data; shaded variables are constructed from daily data. Variables are defined as follows (full definitions given in Table 1): $\overline{c}$, average effective cost; λ, price impact coefficient; ϕ, reversal coefficient for signed order flow; c^{MZ}, moment estimate of Roll model (with infeasible estimates set to zero); c^{Gibbs}, Gibbs estimate of Roll model; L, liquidity ratio; I, illiquidity ratio; γ, reversal coefficient of unsigned volume.

Panel A. Full Sample

	$\overline{c}$	λ	ϕ	c^{MZ}	c^{Gibbs}	L	I	γ
$\overline{c}$	1.000	-0.174	-0.204	0.716	0.806	0.026	0.566	-0.011
	1.000	*-0.152*	*0.046*	*0.487*	*0.524*	*-0.608*	*0.633*	*0.146*
λ	-0.174	1.000	-0.007	-0.289	-0.237	-0.021	-0.056	-0.022
	-0.152	*1.000*	*-0.130*	*-0.312*	*-0.395*	*-0.170*	*0.198*	*-0.163*
ϕ	-0.204	-0.007	1.000	-0.103	-0.151	-0.005	-0.385	-0.264
	0.046	*-0.130*	*1.000*	*0.066*	*0.059*	*0.010*	*-0.001*	*0.219*
c^{MZ}	0.716	-0.289	-0.103	1.000	0.845	0.024	0.409	0.074
	0.487	*-0.312*	*0.066*	*1.000*	*0.764*	*-0.273*	*0.354*	*0.194*
c^{Gibbs}	0.806	-0.237	-0.151	0.845	1.000	0.027	0.452	0.130
	0.524	*-0.395*	*0.059*	*0.764*	*1.000*	*-0.241*	*0.359*	*0.177*
L	0.026	-0.021	-0.005	0.024	0.027	1.000	0.017	0.004
	-0.608	*-0.170*	*0.010*	*-0.273*	*-0.241*	*1.000*	*-0.771*	*-0.099*
I	0.566	-0.056	-0.385	0.409	0.452	0.017	1.000	0.006
	0.633	*0.198*	*-0.001*	*0.354*	*0.359*	*-0.771*	*1.000*	*0.124*
γ	-0.011	-0.022	-0.264	0.074	0.130	0.004	0.006	1.000
	0.146	*-0.163*	*0.219*	*0.194*	*0.177*	*-0.099*	*0.124*	*1.000*

Table 5. Partial correlations (with respect to log equity market value) between high-frequency and daily liquidity measures in the comparison sample. (continued)

The comparison sample consists of firms randomly selected those present on both TAQ and CRSP databases. 200 firms are chosen from each of the ten years 1993-2002. Variable definitions are given in Table 1. Each entry in the correlation matrix contains the partial Pearson correlation (top row) and the partial Spearman correlation (italics, bottom row). Unshaded variables are constructed from high-frequency data; shaded variables are constructed from daily data. Variables are defined as follows (full definitions given in Table 1): $\bar{c}$, average effective cost; λ, price impact coefficient; ϕ, reversal coefficient for signed order flow; c^{MZ}, moment estimate of Roll model (with infeasible estimates set to zero); c^{Gibbs}, Gibbs estimate of Roll model; L, liquidity ratio; I, illiquidity ratio; γ, reversal coefficient of unsigned volume.

Panel B. NYSE/Amex

	$\bar{c}$	λ	ϕ	c^{MZ}	c^{Gibbs}	L	I	γ
$\bar{c}$	1.000	-0.228	-0.375	0.792	0.833	0.017	0.810	-0.140
	1.000	*-0.141*	*0.026*	*0.215*	*0.277*	*-0.523*	*0.460*	*0.016*
λ	-0.228	1.000	0.074	-0.318	-0.275	-0.016	-0.151	-0.028
	-0.141	*1.000*	*-0.041*	*-0.320*	*-0.475*	*-0.230*	*0.335*	*-0.118*
ϕ	-0.375	0.074	1.000	-0.319	-0.460	-0.007	-0.564	-0.407
	0.026	*-0.041*	*1.000*	*0.008*	*0.013*	*-0.012*	*0.006*	*0.192*
c^{MZ}	0.792	-0.318	-0.319	1.000	0.881	0.006	0.628	-0.005
	0.215	*-0.320*	*0.008*	*1.000*	*0.690*	*-0.118*	*0.080*	*0.076*
c^{Gibbs}	0.833	-0.275	-0.460	0.881	1.000	0.014	0.645	0.095
	0.277	*-0.475*	*0.013*	*0.690*	*1.000*	*-0.033*	*0.031*	*0.034*
L	0.017	-0.016	-0.007	0.006	0.014	1.000	0.013	0.001
	-0.523	*-0.230*	*-0.012*	*-0.118*	*-0.033*	*1.000*	*-0.765*	*0.010*
I	0.810	-0.151	-0.564	0.628	0.645	0.013	1.000	-0.090
	0.460	*0.335*	*0.006*	*0.080*	*0.031*	*-0.765*	*1.000*	*-0.000*
γ	-0.140	-0.028	-0.407	-0.005	0.095	0.001	-0.090	1.000
	0.016	*-0.118*	*0.192*	*0.076*	*0.034*	*0.010*	*-0.000*	*1.000*

Table 5. Partial correlations (with respect to log equity market value) between high-frequency and daily liquidity measures in the comparison sample. (continued)

The comparison sample consists of firms randomly selected those present on both TAQ and CRSP databases. 200 firms are chosen from each of the ten years 1993-2002. Variable definitions are given in Table 1. Each entry in the correlation matrix contains the partial Pearson correlation (top row) and the partial Spearman correlation (italics, bottom row). Unshaded variables are constructed from high-frequency data; shaded variables are constructed from daily data. Variables are defined as follows (full definitions given in Table 1): $\bar{c}$, average effective cost; λ, price impact coefficient; ϕ, reversal coefficient for signed order flow; c^{MZ}, moment estimate of Roll model (with infeasible estimates set to zero); c^{Gibbs}, Gibbs estimate of Roll model; L, liquidity ratio; I, illiquidity ratio; γ, reversal coefficient of unsigned volume.

Panel C. Nasdaq

	$\bar{c}$	λ	ϕ	c^{MZ}	c^{Gibbs}	L	I	γ
$\bar{c}$	1.000	-0.104	-0.123	0.691	0.816	0.108	0.457	0.144
	1.000	*-0.077*	*0.017*	*0.588*	*0.643*	*-0.738*	*0.704*	*0.197*
λ	-0.104	1.000	-0.047	-0.128	-0.091	-0.071	-0.051	0.028
	-0.077	*1.000*	*-0.095*	*-0.140*	*-0.134*	*-0.035*	*0.169*	*-0.110*
ϕ	-0.123	-0.047	1.000	-0.013	-0.062	0.006	0.050	0.504
	0.017	*-0.095*	*1.000*	*0.031*	*0.012*	*-0.053*	*0.014*	*0.217*
c^{MZ}	0.691	-0.128	-0.013	1.000	0.815	0.131	0.390	0.191
	0.588	*-0.140*	*0.031*	*1.000*	*0.788*	*-0.511*	*0.520*	*0.208*
c^{Gibbs}	0.816	-0.091	-0.062	0.815	1.000	0.129	0.495	0.229
	0.643	*-0.134*	*0.012*	*0.788*	*1.000*	*-0.616*	*0.614*	*0.215*
L	0.108	-0.071	0.006	0.131	0.129	1.000	0.064	0.035
	-0.738	*-0.035*	*-0.053*	*-0.511*	*-0.616*	*1.000*	*-0.813*	*-0.232*
I	0.457	-0.051	0.050	0.390	0.495	0.064	1.000	0.228
	0.704	*0.169*	*0.014*	*0.520*	*0.614*	*-0.813*	*1.000*	*0.206*
γ	0.144	0.028	0.504	0.191	0.229	0.035	0.228	1.000
	0.197	*-0.110*	*0.217*	*0.208*	*0.215*	*-0.232*	*0.206*	*1.000*

Table 6. Descriptive statistics for variables used in risk-adjusted return regressions

Values in the table are averages of monthly cross-sectional statistics. The variables are defined as follows: R^* is the risk-adjusted return; c^{Gibbs} is the Gibbs estimate of the effective cost; L is the liquidity ratio; I is the illiquidity ratio; γ is the reversal coefficient. LMV is log of market value of equity; RET23, RET46, and RET712 are the logarithms of the cumulative returns over the second through third, fourth through sixth, and seventh through 12th months prior to the current month, respectively. t-statistics are in parentheses.

			R^*	c^{Gibbs}	L	I	γ	LMV	$RET23$	$RET46$	$RET712$
NYSE/ Amex	Mean		-0.000	0.006	39324	4.487	0.028	11.803	0.009	0.013	0.025
	S.D.		0.121	0.008	1.48E6	20.507	0.501	1.952	0.151	0.180	0.250
	Corr	R^*	1.000	0.011	0.002	0.017	0.005	-0.003	0.009	0.019	0.024
		c^{Gibbs}	0.011	1.000	-0.040	0.705	0.195	-0.515	-0.020	-0.031	-0.084
		L	0.002	-0.040	1.000	-0.022	-0.007	0.179	0.006	0.008	0.018
		I	0.017	0.705	-0.022	1.000	0.239	-0.391	-0.002	-0.005	-0.037
		γ	0.005	0.195	-0.007	0.239	1.000	-0.112	-0.000	0.002	-0.011
		LMV	-0.003	-0.515	0.179	-0.391	-0.112	1.000	0.034	0.057	0.125
		$RET23$	0.009	-0.020	0.006	-0.002	-0.000	0.034	1.000	-0.002	0.052
		$RET46$	0.019	-0.031	0.008	-0.005	0.002	0.057	-0.002	1.000	0.039
		$RET712$	0.024	-0.084	0.018	-0.037	-0.011	0.125	0.052	0.039	1.000
Nasdaq	Mean		0.003	0.018	489.16	8.434	0.023	10.982	-0.009	-0.013	-0.014
	S.D.		0.184	0.019	10281	33.778	0.243	1.567	0.222	0.264	0.363
	Corr	R^*	1.000	0.019	0.002	0.018	0.001	-0.007	0.008	0.018	0.024
		c^{Gibbs}	0.019	1.000	-0.092	0.705	0.156	-0.626	-0.018	-0.034	-0.100
		L	0.002	-0.092	1.000	-0.049	-0.021	0.261	0.007	0.009	0.026
		I	0.018	0.705	-0.049	1.000	0.161	-0.395	-0.007	-0.017	-0.063
		γ	0.001	0.156	-0.021	0.161	1.000	-0.146	-0.011	-0.012	-0.025
		LMV	-0.007	-0.626	0.261	-0.395	-0.146	1.000	0.061	0.102	0.209
		$RET23$	0.008	-0.018	0.007	-0.007	-0.011	0.061	1.000	0.020	0.072
		$RET46$	0.018	-0.034	0.009	-0.017	-0.012	0.102	0.020	1.000	0.057
		$RET712$	0.024	-0.100	0.026	-0.063	-0.025	0.209	0.072	0.057	1.000

Table 7. Risk-adjusted return regression estimates

Fama-MacBeth regression estimates of Eq. (17) using individual security data. Coefficient estimates are time-series averages of cross-sectional OLS regressions. The dependent variable is excess returns risk-adjusted using the FF factors. The variables are defined as follows: c^{Gibbs} is the Gibbs estimate of the effective cost; L is the liquidity ratio; I is the illiquidity ratio; γ is the reversal coefficient. LMV is log of market value of equity; RET23, RET46, and RET712 are the logarithms of the cumulative returns over the second through third, fourth through sixth, and seventh through 12th months prior to the current month, respectively. t-statistics are in parentheses.

		(a)	*(b)*	*(c)*	*(d)*	*(e)*
NYSE/Amex (January 1963 – December 2002, 480 observations)	Intercept	-0.001	0.010	0.002	0.008	-0.000
		(-0.35)	(2.50)	(0.78)	(2.30)	(-0.09)
	c^{Gibbs}	0.328				0.091
		(3.79)				(1.09)
	$L \times 10^{-6}$		-0.064			-0.137
			(-0.42)			(-1.07)
	$I \times 10^{-3}$			0.211		0.197
				(3.13)		(2.75)
	γ				0.004	0.001
					(2.14)	(0.62)
	$LMV \times 10^{-3}$	-0.143	-0.892	-0.312	-0.762	-0.134
		(-0.71)	(-2.91)	(-1.37)	(-2.77)	(-0.60)
	RET23	0.007	0.007	0.007	0.007	0.007
		(2.41)	(2.37)	(2.33)	(2.26)	(2.36)
	RET46	0.013	0.012	0.012	0.012	0.012
		(4.94)	(4.78)	(4.76)	(4.75)	(4.82)
	RET712	0.011	0.010	0.010	0.010	0.010
		(5.46)	(5.05)	(5.16)	(5.12)	(5.38)
	R^2	0.034	0.028	0.033	0.030	0.040

Table 7. Risk-adjusted return regression estimates (continued)

Fama-MacBeth regression estimates of Eq. (17) using individual security data. Coefficient estimates are time-series averages of cross-sectional OLS regressions. The dependent variable is excess returns risk-adjusted using the FF factors. The variables are defined as follows: c^{Gibbs} is the Gibbs estimate of the effective cost; *L* is the liquidity ratio; *I* is the illiquidity ratio; γ is the reversal coefficient. *LMV* is log of market value of equity; RET23, RET46, and RET712 are the logarithms of the cumulative returns over the second through third, fourth through sixth, and seventh through 12th months prior to the current month, respectively. t-statistics are in parentheses.

		(a)	*(b)*	*(c)*	*(d)*	*(e)*
Nasdaq (January 1984 – December 2002)	Intercept	0.003	0.030	0.018	0.029	0.006
		(0.33)	(3.12)	(2.14)	(3.23)	(0.67)
	c^{Gibbs}	0.221				0.110
		(3.48)				(1.56)
	$L \times 10^{-6}$		-0.117			-1.158
			(-0.08)			(-0.87)
	$I \times 10^{-3}$			0.233		0.557
				(0.27)		(0.64)
	γ				-0.005	-0.008
					(-0.30)	(-0.47)
	$LMV \times 10^{-3}$	-0.432	-2.503	-1.513	-2.391	-0.636
		(-0.63)	(-3.22)	(-2.25)	(-3.35)	(-0.87)
	RET23	0.006	0.006	0.006	0.006	0.006
		(1.24)	(1.42)	(1.31)	(1.38)	(1.27)
	RET46	0.010	0.011	0.011	0.011	0.011
		(2.87)	(3.05)	(3.14)	(3.08)	(3.07)
	RET712	0.010	0.010	0.010	0.010	0.010
		(3.91)	(3.81)	(3.89)	(3.83)	(3.86)
	R^2	0.030	0.028	0.030	0.029	0.035

Figure 1. Effective costs 1962-2002

Cross-firm averages of Gibbs estimates of effective trading cost (c^{Gibbs}). NYSE breakpoints used in computing equity capitalization quartiles.

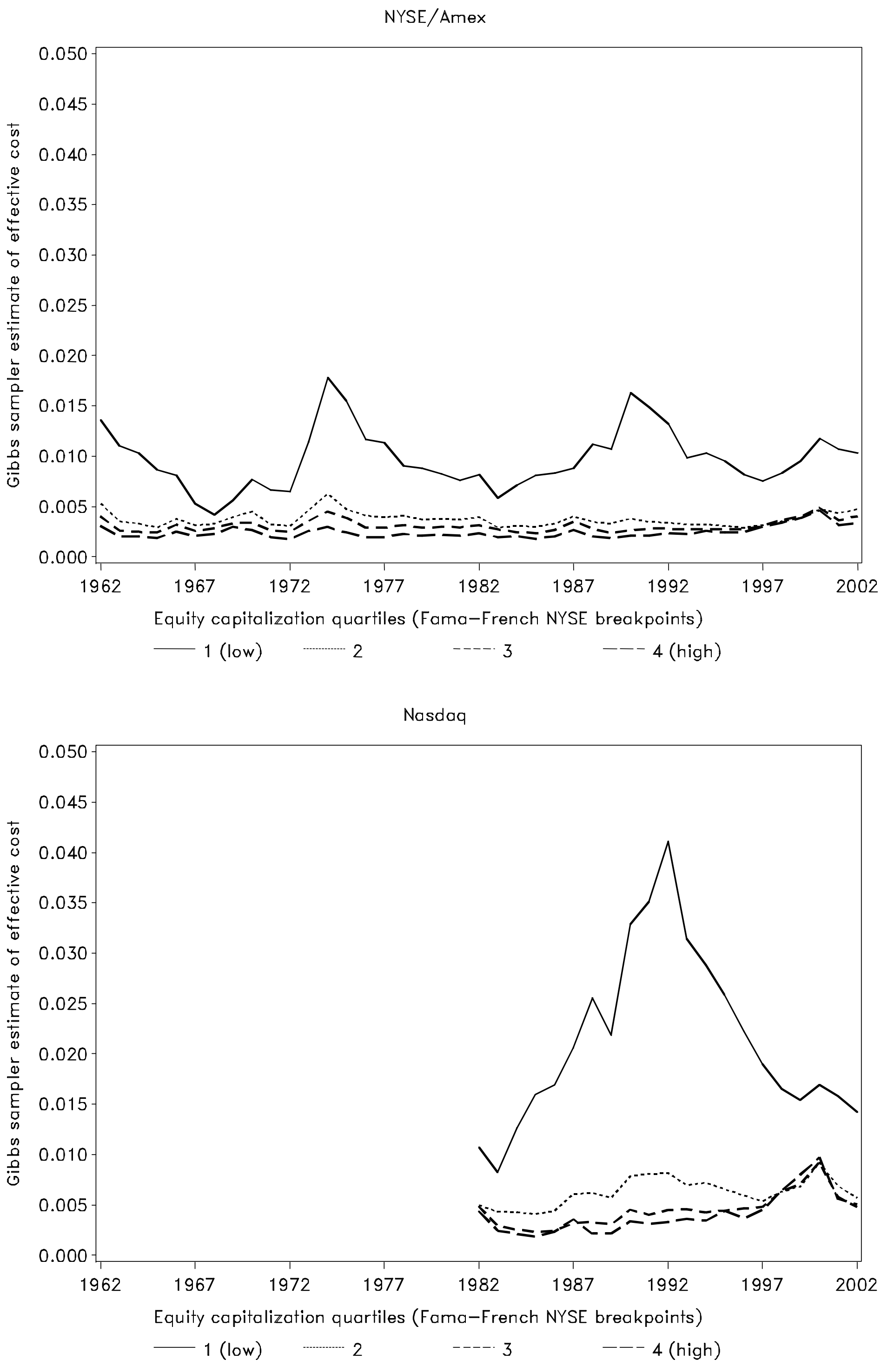

Figure 2. Risk-adjusted return regression estimates, trading cost coefficients by month

Figures are based on Fama-MacBeth regression estimates of Eq. (17) using individual security data. Coeffcient estimates are time-series averages of cross-sectional OLS regressions. The dependent variable is excess returns risk-adjusted using the FF factors. The explanatory variables are the log of market value of equity; logarithms of the cumulative returns over the second through third, fourth through sixth, and seventh through 12th months prior to the current month, and (taken one at a time) a liquidity measure from the set consisting of the Gibbs estimate of the effective cost c^{Gibbs}, the liquidity ratio L, the illiquidity ratio I, and the reversal measure γ. Each graph plots the time series mean of the indicated liquidity coefficient by month. Vertical bars denote two-standard-error bounds. The overall mean coefficient (and error bounds) are indicated by the horizontal solid (and dotted) lines.