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

August 30, 2002
This draft: March 14, 2003

Preliminary draft
Comments welcome

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 approaches for estimating effective costs and price impacts using data of daily frequency. The daily-based estimates are evaluated by comparison with corresponding estimates based on high-frequency TAQ data. The analysis suggests that the best daily-based estimate of effective cost is the Gibbs sampler estimate of the Roll model (suggested in Hasbrouck (1999)). The correlation between this estimate and the TAQ value is about 0.90 for individual securities and about 0.98 for portfolios. Daily-based proxies for price impact costs, however, are more problematic. Among the proxies considered here, the illiquidity measure (Amihud (2000)) appears to be the best: its correlation with the TAQ-based price impact measure is 0.47 for individual stocks and 0.90 for portfolios.

The study then extends the Gibbs effective cost estimate to the full sample covered by the daily CRSP database (beginning in 1962). These estimates exhibit considerable cross-sectional variation, consistent with the presumption that trading costs vary widely across firms, but only modest time-series variation. In specifications using Fama-French factors, the Gibbs effective cost estimates are found to be positive determinants of expected returns.

1. Introduction

The notion that individuals must take into account the costs of acquiring, divesting and rebalancing their portfolios, and that these costs affect equilibrium expected returns, is driving a convergence of market microstructure and asset pricing (see the recent survey of Easley and O'Hara (2002)). Asset pricing tests generally require large cross-sectional and long time samples in order to reliably estimate differences in expected returns on risky assets. Measures of trading cost common in empirical microstructure work, on the other hand, are generally based on high-frequency trade and quote data, and are so limited to the samples for which these data exist.[1] Trading cost measures based on data of daily or lower frequency are therefore highly desirable. In this context, the present paper seeks to examine various daily-based trading cost measures utilized in other studies, explore their relationships to high-frequency cost measures, and to discuss a promising new daily-based liquidity measure.

The analysis starts with existing high-frequency measures of trading cost. Roughly speaking, these fall into two categories: spread-related and price-impact measures. Neither of these is a comprehensive measure of trading cost. Spread-related measures reflect the cost faced by a trader contemplating a single order of small size. Price impact functions are more indicative of the costs associated with dynamic strategies in which an order is broken up and fed to the market over time.

The paper then turns to the various proxies that may be constructed from daily return and (in some cases) volume data. As with the high-frequency measures, these tend to resemble either spread-related or price impact proxies. Spread proxies include the standard moment-based estimates of the Roll model and a new measure, the Gibbs sampler estimate of the Roll model. Price impact proxies include the liquidity ratio (used

---

[1] Representative studies using high-frequency data in asset pricing specifications include Brennan and Subrahmanyam (1996) and Easley, Hvidkjaer, and O'Hara (1999).

by Cooper, Groth, and Avera (1985) and others), the illiquidity ratio (Amihud (2000)) and the reversal (gamma) measure of Pastor and Stambaugh (2002).

Both high-frequency and daily-based measures are constructed for a comparison sample comprising roughly 1,800 firm-years in 1993 to 2001 (for which we possess both daily CRSP and high-frequency TAQ data). The correlations between the CRSP-based and TAQ-based measures are used as a guide to the validity of the former as proxies. By this measure, the Gibbs estimate is an excellent proxy for effective cost, both for individual stocks and portfolios. Among the price impact proxies, the illiquidity ratio is modestly correlated with the TAQ measure for individual stocks, and more strongly for portfolios.

In view of the strong performance of the Gibbs effective cost estimate in the comparison analysis, these estimates are constructed for the CRSP daily database. These estimates exhibit substantial cross-sectional variation. Time-series variation, however, is large only for subsamples that have low market capitalization.

The paper also presents a preliminary analysis of the relation between the effective cost estimates and returns, to assess whether effective cost is associated with a liquidity premium. In specifications that include the three Fama-French factors, excess returns are found to be positively related to effective costs.

The paper is organized as follows. The next section discusses empirical measurement of trading costs from a microstructure perspective, and develops the two classes of measures (spread and price impact). The following two sections discuss proxies that may be constructed from daily data. Section 3 analyzes spread proxies; Section 4, price impact proxies. Section 5 describes the data samples. The properties of the cost estimates in the TAQ/CRSP comparison sample are discussed in Sections 6 (TAQ) and 7 (CRSP). Section 8 describes the correlations between the CRSP estimates and the TAQ values they are supposed to proxy. Effective cost estimates for the full daily CRSP sample are described in Section 9, and their relation to returns is analyzed in Section 10. A brief summary concludes the paper in section 11.

2. Measures of trading cost based on detailed (high-frequency) data

Most computations of transaction costs can be discussed within the context of the implementation shortfall approach advocated by Perold (1988). For an executed trade (or, more generally, a sequence of trades), 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. The most straightforward calculations are for institutional traders, for whom the time of the trading decision and the exact sequence of trades are usually well-documented (Keim and Madhavan (1995), Keim and Madhavan (1996), Chan and Lakonishok (1997), Conrad, Johnson, and Wahal (2001)). More commonly, data limitations or the need for prospective (rather than retrospective) measures introduce complications. In such situations, it is useful to group the measures as "spread-related" and "impact-related", and the following discussion is organized on these lines.

a. Spreads: posted and effective

A trader who demands to trade a small amount of the security immediately must be prepared to pay the market's prevailing ask price (if buying) or receive the market's prevailing bid (if buying). The difference between the two is the posted market spread. Taking the midpoint as a benchmark, the half- spread is a sensible first estimate of the trading cost. For comparability across firms and time, this paper's definition uses logs. The half-spread is:

$$s_t = \frac{1}{2}(a_t - b_t)$$

where a_t is the log of the ask price prevailing at time t and b_t is the log of the bid. For a particular firm over some time period, a useful summary measure of the half-spread is the time-weighted average of s_t. Posted bid-ask spreads have been used in asset pricing studies by Stoll and Whalley (1983), Amihud and Mendelson (1986), Amihud and Mendelson (1989), Eleswarapu and Reinganum (1993), Kadlec, McConnell, and Purdue U (1994), and Eleswarapu (1997), among others.

In many markets, however, for a variety of reasons, actual trade prices are often better than the posted quotes (due to “price improvement”). Accordingly, the effective cost is defined as

$$c_t = \begin{cases} p_t - m_t, \text{ for a buy order} \\ m_t - p_t, \text{ for a sell order} \end{cases} \tag{1}$$

where p_t is the actual log price of the t^{th} trade and m_t 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. For a particular firm over a given time interval, a useful summary measure of c_t is the dollar-volume-weighted average.

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, disaggregated by order size and security characteristics.

Accurate computation of the effective cost requires knowledge of order characteristics, most importantly the arrival time and direction (buy or sell). Since order data are not widely available, the effective cost is commonly 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.

b. Price impact measures

Incoming orders give rise to both temporary and permanent effects on the security price. From an economic perspective, temporary components may arise from transient liquidity effects, inventory control behavior, price discreteness, etc. Permanent effects are generally attributed to the information content of the order. It is difficult to differentiate permanent and transient effects empirically: what appears to be permanent over a window of five minutes may be transitory over a day. Nevertheless, for simplicity, the following discussion will assume that the entire impact is permanent.

To illustrate the importance of price impact for trading costs, suppose that the evolution of the quote midpoint is given by:

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

where x_t is the signed order flow, λ is a liquidity parameter and u_t is a zero-mean disturbance reflecting newly arriving non-trade-related public information. Suppose that the effective cost on each trade is c, and consider a buy order that is broken into two trades of n_1 and n_2 shares. Relative to the initially prevailing quote midpoint, the expected total cost of the order is

$$n_1\left(p_1 - m_1\right) + n_2\left(p_2 - m_1\right) = n_1\left(p_1 - m_1\right) + n_2\left(p_2 - m_2 + m_2 - m_1\right) = c\left(n_1 + n_2\right) + \lambda n_1$$

That is, in addition to the effective cost on the order, the total cost reflects the price impact of the first trade. Extension to more than two trades is straightforward.

Although some theoretical models imply a relation close to eq. (2), market features such as discreteness, inventory control, serial correlation in order flow, etc., militate in favor of more general specifications. These specifications are often estimated at the transaction level, and often involve the joint dynamics of order flow and other variables, as well as prices.

To facilitate estimation over a large sample of stocks, estimations in the present paper are based on returns and signed order flows aggregated over fifteen-minute intervals. The empirical evidence is mixed on the exact specification of the order variables. Accordingly, four variants of eq. (2) were considered (for each stock), using singly and jointly the following order flow variables. Let $v_{i,t}$ represent the signed dollar volume of the ith trade in fifteen-minute interval t, signed in the usual fashion (by comparing the trade price to the prevailing midpoint). Then V_t is the cumulative signed dollar volume, $V_t = \sum_i v_{i,t}$, where the summation is over all trades in interval t; N_t is the signed number of trades, $N_t = \sum_i Sign\left(v_{i,t}\right)$ where $Sign\left(x\right) = +1 \text{ if } x > 0 \ \ -1 \text{ if } x > 0$, and 0 if x=0; and, S_{it} is the signed square-root dollar volume, $S_t = \sum_i Sign\left(v_{i,t}\right)\sqrt{\left|v_{it}\right|}$. With these definitions, the models are:

$$\begin{aligned}&\text{Model I: } r_t = \lambda^{I}\, N_t + u_t\\&\text{Model II: } r_t = \lambda^{II}\, S_t + u_t\\&\text{Model III: } r_t = \lambda^{III}\, V_t + u_t\\&\text{Model IV: } r_t = \lambda_1^{IV}\, N_t + \lambda_2^{IV}\, S_t + \lambda_3^{IV}\, V_t + u_t\end{aligned} \tag{3}$$

Price-impact related measures of trading cost have been used in asset pricing specifications by Brennan and Subrahmanyam (1996). A related measure is the *PIN* statistic used in Easley, Hvidkjaer, and O'Hara (1999). *PIN* is a measure of information asymmetry, but it is not a measure of price impact. It is based solely on signed order flows, and most directly reflects the strength of one-sided runs in the order flow.

3. Estimates of the effective cost constructed from daily data

The effective cost and price impact measures described above may be estimated from transactions level data. This section turns to consider of estimates constructed from data at a daily or longer frequency.

a. The Roll model

Roll (1984) suggested a simple model of the spread in an efficient market. A variant of this model is as follows. Let the logarithm of the efficient price, m_t, evolve as:

$$m_t = m_{t-1} + u_t\,. \tag{4}$$

where $Eu_t = 0$ and $Eu_t u_s = 0$ for $t \neq s$. The term "efficient price" is used here in the sense common to the sequential trade models, i.e., the expected terminal value of the security conditional on all public information (including the trade history). The u_t reflect new public information. The (log) bid and ask prices are given as

$$\begin{aligned}b_t &= m_t - c\\a_t &= m_t + c\end{aligned} \tag{5}$$

where c is the nonnegative half-spread, presumed to reflect the quote-setter's cost of market-making.

The direction of the incoming order is given by the Bernoulli random variable $q_t \in \{-1,+1\}$, where -1 indicates an order to sell (to the quote-setter) and $+1$ indicates an order to buy (from the quote-setter). Buys and sells are assumed equally probable. In the standard implementation, q_t is assumed independent of $\Delta m_t = u_t$, i.e., that the direction of the trade is independent of the efficient price movement. Depending on q_t, the (log) transaction price is either at the bid or the ask:

$$p_t = \begin{cases} b_t \text{ if } q_t = -1 \\ a_t \text{ if } q_t = +1 \end{cases} \tag{6}$$

The cost parameter is c. Inference is based on a time series sample of trade prices $p = \{p_1, p_2, \ldots, p_T\}$. Since the prices are those at which transactions actually occur, c is in principle the effective cost, rather than half the posted spread.

b. Moment estimates of c (c^M and c^{MZ})

Roll proposed estimation 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:

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

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. The estimate obtained in this fashion will be denoted c^M. 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 autocovariance is negative. In finite samples, particularly of the sort that arise with daily data, however, this is often not the case. When estimating the spread using annual samples of daily return data, 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 expedient to the problem of infeasible moment estimates is to simply assign an a priori value of zero. This gives rise to a moment/zero estimate: $c^{MZ} = c^M$, if c^M is defined, and zero otherwise.

c. The Gibbs-sampler estimate, $c^{Gibbsibbs}$

Hasbrouck (1999) advocates Bayesian estimation using the Gibbs sampler. To complete the Bayesian specification, I assume here that $u_t \overset{d}{\sim} i.i.d.N\left(0,\sigma_u^2\right)$. In this approach, the unknowns comprise both the model parameters $\left\{c,\sigma_u^2\right\}$ and the latent data, i.e., the trade direction indicators $q \equiv \left\{q_1,\ldots,q_T\right\}$ and the efficient prices $m \equiv \left\{m_1,\ldots,m_T\right\}$. The parameter posterior $f\left(c,\sigma_u \middle| p\right)$ 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.

To start this process, suppose that q and σ_u are given, and consider the construction of $f\left(c \middle| \sigma_u, q, p\right)$. Given q, eq (7) is a simple linear regression with c as the coefficient. With a normal prior for c, this is a standard Bayesian regression model (see, for example, Kim and Nelson (2000)). The prior used here is actually a modification, specifically, $c \overset{d}{\sim} N^+\left(0,\sigma_c^{2,prior}\right)$ where the "+" superscript denotes restriction to the positive domain. In the implementation, I set $\sigma_c^{2,prior} = 1$. As posted spreads are usually much lower than 50%, this implies a prior that is relatively flat over the region of interest.[2] The posterior for c is also normal, and a random draw is made from this.

Next, given q, p and the newly-drawn value of c, we may compute the residuals in eq (7). A convenient prior for σ_u^2 is the inverted gamma distribution. I use $\sigma_u^2 \overset{d}{\sim} IG\left(\alpha,\beta\right)$ with $\alpha = \beta = 10^{-12}$, implying a fairly uninformative prior. The posterior is also inverted gamma, and a draw of σ_u^2 is made from this distribution. The next step is to make random draws of m and q, conditional on c, σ_u^2, and p. The details of this are

---

[2] Choice of $\sigma_c^{2,prior}$ is important in one respect. The Gibbs estimate of c is essentially formed by estimating equation (7) as a simple linear regression conditional on simulated values of q_t. It is possible that in some draws, all of the values of q_t are either +1 or –1. In this case, all of the Δq_t are zero, the regression is uninformative, and the posterior distribution for c is identical to the prior. An extremely large draw of c can draw the Gibbs sampler into a region where mixing is poor.

discussed in Hasbrouck (1999). This completes one cycle ("sweep") of the Gibbs sampler.[3] The appendix to the present paper provides an illustration.

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 efficient price evolution. Realized prices are discrete. The effective cost is unlikely to be constant within a sample. Etc. Hasbrouck 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 are 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.

4. Measures of price impact constructed from daily data.

The price impact parameters, the λs in Models I-IV, are coefficients of signed order flow variables. These are generally not available. The closest thing reported by most markets on a daily basis is the trade volume, the total number of shares that changed hands. A number of price impact measures based on (unsigned) volume have been proposed by other researchers. I examine here three representative proxies, and propose a third.

---

[3] For each stock, I ran 1,000 sweeps of the sampler, discarding the first 200 as a burn-in period. The mean of the c draws in the remaining 800 sweeps was taken as the summary estimate of c.

a. The (Amivest) liquidity ratio, L

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.

b. The illiquidity ratio, I

Amihud (2000) suggests measuring *illiquidity* as:

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

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. In terms of units (return per dollar volume), this measure roughly corresponds to the price impact coefficient λ^{III} in Model III, eq. (3). The variables are substantially different, however, as the λ^{III} relates signed volume to signed return, whereas I relates absolute return and cumulative (unsigned) volume. This measure is also used by Acharya and Pedersen (2002).

c. The reversal measure, γ

Pastor and Stambaugh (2002) suggest measuring liquidity by γ in the regression

$$r_{d+1} = \theta + \phi r_d + \gamma\, sign\left(r_d^e\right) Vol_d + \varepsilon_d \tag{11}$$

where $r_{d,m}^e$ is the stock's excess return (over the CRSP value-weighted market return) on day d in month m, and $Vol_{d,m}$ is the dollar volume. The liquidity measure is the coefficient of lagged signed volume. Intuitively, it measures the subsequent day's "correction" to an order flow shock. In principle, it should be negative, with more negative values implying lower liquidity.

5. Data for the comparison analysis

The analysis draws on TAQ data for the high-frequency measures, and on CRSP for daily data. There are two samples. The TAQ comparison sample is a random sample of firms that are present and could be matched on TAQ and CRSP databases. Estimation of the high-frequency specifications was performed on a sample drawn from the NYSE's TAQ database, from 1993 through 2001. The sample was constructed as follows. For a given year, a stock was eligible if it was a common stock, was present on first and last TAQ master file for the year, had the NYSE, Amex or Nasdaq as the primary listing exchange, and didn't change primary exchange, ticker symbol or cusip over the year. (Constancy of primary exchange, ticker symbol and cusip facilitated the subsequent matching to the CRSP data.) All eligible stocks for a year were randomly permuted, and the first 200 were selected. Each of the nine years was sampled separately, resulting in a total sample of 1,800 firm years. Firms that could not be matched subsequently to CRSP were deleted. Most of the results discussed below are based on annual estimations. Alternative analyses employed monthly and quarterly estimations.

6. Cost estimates based on TAQ data

Table 1 reports counts for the TAQ comparison sample by year and listing exchange. In all years, Nasdaq firms are most numerous.

a. Spread-related measures

Table 2 reports summary statistics for the spread-related measures estimated from the TAQ data. The mean log effective cost is 0.014 (roughly 1.4%), while the mean posted log half-spread is about a third higher (0.020). This is the usual result, consistent with widespread betterment of the posted quotes. The relatively high skewness and kurtosis, however, reflect a distribution that has an extreme upper tail. That is, a few stocks have very high posted and effective costs.

b. Price impact measures

Table 2 also reports summary statistics for R^2s for each of the four return/signed order flow models discussed in Section 2.b. Among models I, II and III, which each rely on a single signed order flow proxy, Model II (which uses the square-root order flow, S_t) is, by a small margin, the best (judging by mean R^2). Interestingly, Model III (the signed dollar volume) is the worst. Finally, the incremental improvement in fit (relative to model II) achieved by putting in all three variables (model IV) is small. Accordingly in the interests of proceeding on to the next stage of the analysis with a single price impact coefficient, Model II is preferable to the others.

It is worth emphasizing, however, that the distribution of the price impact coefficient λ^{II}, is sharply skewed and highly leptokurtotic. While some of the high values might arise from estimation error, it is also acknowledged as a practical matter that impact costs in thinly traded issues are extremely high. There is no obvious reason to exclude these observations from the analysis. The character of the distributions does argue, however, in favor of robust statistical analysis.

7. Daily return-based (CRSP) cost estimates in the TAQ comparison sample

Table 3 presents summary statistics for the cost estimates in the TAQ comparison sample that are based solely on CRSP daily data. That is, although the firms and time periods were selected to match the TAQ sample, the estimates discussed in this section do not employ the TAQ data.

a. Spread-related cost measures.

Summary statistics for the Gibbs-sampler estimate of the effective cost, c^{Gibbs}, are presented in the first row of Table 3. The similarity of the distributional parameters to those for the effective cost estimated from the transaction data (Table 2) is striking. Means, standard deviations, and skewness and kurtosis parameters are very close.

The moment estimate of effective cost, c^{M}, does not fare as well. In over one-third of the cases it is undefined (due to positive return autocovariances). The alternative estimate that sets otherwise undefined values to zero, c^{MZ}, offers some improvement. While the mean and standard deviation of c^{MZ} are similar to those for the TAQ-based effective cost, the skewness and kurtosis are somewhat lower.

b. Impact-related cost measures

The bottom part of Table 3 presents summary statistics for the price impact proxies. Unlike the effective cost estimates discussed above, however, we would not expect the distributions for these proxies to closely resemble that of the price impact parameter (λ^{II}) described in Table 2. The proxies in Table 3 nevertheless exhibit a kurtosis that is even higher than that of λ^{II}.

8. Correlations and proxy relationships in the TAQ comparison sample

Table 4 presents correlation matrices for the principal TAQ and CRSP measures used in the study. The variables are grouped as to type of measure (effective cost or price impact) and source of data (TAQ or CRSP). The table presents both Pearson and Spearman (rank order) correlations. Furthermore, in view of prominent role of market capitalization as an explanatory variable in both microstructure and asset pricing analyses, the table presents both and partial (with respect to log market capitalization) correlations.

The upper left-hand section of each matrix summarizes correlations between effective cost estimates. Across all correlation measures, both c^{Gibbs} and c^{MZ} are positively correlated with the c^{TAQ}, but the Gibbs correlation is stronger. The Pearson correlation between the c^{TAQ} and c^{Gibbs} is 0.901, while that for the moment/zero estimate is 0.825. This pattern also appears in the Spearman correlations (Panel B), Pearson partial correlations (Panel C), and Spearman partial correlations (Panel D). The relationship between c^{TAQ} and c^{Gibbs} is illustrated visually in Figure 1, which presents a scatter plot and best-fit regression lines. The relationship is visually strong both in the overall sample (Panel A) and a sample restricted to low values of TAQ effective cost (Panel B). The fit is, however, obviously looser for firms with high effective costs, suggesting that measurement error is higher for these firms.

In summary, within the class of effective cost estimates based on daily data, the Gibbs estimate consistently dominates. It is always feasible (in contrast with the moment estimate). Furthermore, however measured (simple or partial, Pearson or Spearman correlation), it has the strongest relationship to the target value.

Correlations between the impact measures appear in the lower right-hand corner of each correlation matrix. A high value of λ^{II} suggests illiquidity. In principle, therefore, the correlation should be negative for the liquidity ratio *L,* and positive for the illiquidity ratio *I* and the reversal measure γ. The results suggest proxy relationships that are weaker and more variable than those found for the effective cost estimates. Judging by the simple Pearson correlations (Panel A), only *I* is strongly correlated in the expected direction. Given the distributional extremities noted above, however, the Spearman correlation may well be more meaningful. Here, all proxies are correlated in the expected direction, with the liquidity ratio *L* being highest, followed closely by *I*, and then γ. The partial correlations, which measure the residual relationship after controlling for log market capitalization suggest a similar story. The Spearman partial correlations, however, are substantially reduced relative to the corresponding Spearman simple correlations in Panel A.

We now turn to the correlations between price impact and effective cost measures. First note that the Pearson correlation between TAQ-based measures c^{TAQ} and λ^{II} is moderately positive (0.515). The strength of this relation is not uniform, however, across all types of correlation. The partial Spearman correlation is only 0.090. In principle, some positive correlation would be expected. Price impacts arise from asymmetric information considerations, which would presumably be impounded into posted and effective spreads and costs. Spreads should also be driven, however, by inventory and clearing costs, which would not necessarily be reflected in the price impact coefficient. Thus it is not too surprising that these two measures appear to be capturing different things.

In considering the correlations between c^{TAQ} and the CRSP-based price impact proxies, and between λ^{II} and the CRSP-based effective cost estimates, it is worth noting that both the liquidity L and illiquidity I measures are (in the Spearman full and partial correlations) strongly correlated with the TAQ estimate of the effective cost c^{TAQ}. This may reflect the fact that both L and I use volume information and c^{TAQ} is a volume-weighted measure. The corresponding Pearson full and partial correlations are weaker.

The analysis to this point has involved correlations between estimates constructed at the firm level. In many asset pricing applications, however, these estimates are averaged over portfolios. To the extent that estimation errors are uncorrelated, these averages should have lower measurement errors. To assess the improvement offered by forming portfolios, correlation analyses parallel to the ones discussed above were performed for grouped data. The grouping was by year, and within each year by c^{TAQ} or λ^{II}. In the c^{TAQ} analysis, for example, ten portfolios were formed for each year by ranking on c^{TAQ}.

Table 5 reports the correlations between the various measures for portfolios grouped by c^{TAQ}. (For the sake of brevity, Spearman correlations are not reported.) The results for the effective cost proxies are striking. The correlation between c^{TAQ} and the Gibbs estimate c^{Gibbs} is 0.987, while that for c^{MZ} is only slightly lower (Panel A). The partial correlations (net of log market capitalization) are also high, although c^{TAQ} is now more clearly preferable to c^{MZ} (Panel B). Table 6 presents correlations in portfolios grouped by the *TAQ* impact measure λ^{II}. All of the proxy relations are somewhat strengthened, but as with the ungrouped estimates, the illiquidity ratio I is apparently the best proxy. Its full correlation with λ^{II} is 0.899, and its partial correlation is 0.837. This is markedly better than either L or γ.

The results of this section may be summarized as follows. Most importantly, for both individual stocks and portfolios, the CRSP-based Gibbs estimate of effective cost is an excellent proxy for the corresponding TAQ-based estimates. Among the price impact proxies, the illiquidity ratio appears to offer the most consistent relationship to the TAQ-based λ^{II}. The CRSP-based measures of price impact are weaker proxies than the effective cost estimates.Whereas Corr(c^{TAQ}, c^{Gibbs}) is 0.901 for individual stocks (and 0.987 for portfolios), however, Corr(I, λ^{II}) is 0.473 for individual stocks (and 0.899 for portfolios).

9. The Gibbs estimates in a broader sample

Given 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 few valid observations in a given year were excluded.

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 2 depicts the average estimates of effective cost for the NYSE/Amex and Nasdaq samples. The estimates for Nasdaq are substantially higher than those of the NYSE/Amex sample. This is not surprising given the differences in market structure and listed companies.

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.

The graphs in Figure 3 plot average effective costs within subsamples constructed as quintiles on equity market capitalization. These quintiles were formed by collapsing the CRSP market value deciles. From these graphs, it becomes apparent that most of the variation is occurring in relatively low-capitalization stocks. This is particularly true of the NYSE/Amex sample, for which the variation in the effective costs for the third and higher market capitalization quintiles is essentially minor. It should be emphasized that neither sample was subject to any minimum price filters.

It is useful to compare this figure with Jones (2001) historical series for the posted bid-ask spread on the Dow stocks (his Figure 1). Compared with the Dow posted spreads (Jones), the effective spread estimates in the largest market capitalization quintile appear to be more stable over time.

10. Effective costs and stock returns

This section describes the relations between the Gibbs estimates of the effective cost and returns over the period covered by the daily CRSP database, 1962-2001. Because CRSP coverage of Nasdaq is more extensive in the later portion of this sample, separate analyses are performed for NYSE/Amex and Nasdaq issues. The analysis proceeds by constructing portfolios sorted on market capitalization and effective cost, and analyzing the average monthly returns in a standard multifactor framework (Fama and French (1992)).

More specifically, the portfolios are formed by independent sorts on end-of-year market capitalization and the effective cost estimates formed over the year. The grouping is by quintiles. Market capitalization quintiles are formed by collapsing the CRSP market value deciles. These portfolios are then used as groupings for excess returns over the subsequent year. The excess return on a stock in a given month is the total return less the one-month T-bill return.[4]

---

[4] The risk-free return and Fama-French factors used in this study are from the U.S. Research Returns data on Ken French's web site.

Table 7 reports results for the NYSE/Amex sample. From Panel A, the highest average excess returns are found in highest effective cost quintile. This is consistent with the hypothesis of a liquidity premium. Within the lower effective cost quintiles, however, the average excess return is not monotonically increasing in effective cost. Panel B reports the average effective costs for the portfolios. It is important to note here that the values for the highest effective cost quintile are markedly higher than the others. The increase in going from the fourth to the highest quintiles is several times larger that the difference between the fourth and lowest quintiles. This suggests that there is relatively little cost variation apart from the highest quintile. This may contribute to the absence of a relationship between costs and average expected returns in the lower effective cost quintiles.

The results for the Nasdaq sample, reported in Table 8 are slightly stronger. The highest average excess returns are found in the highest effective cost quintile. For all except the second market capitalization quintile, the second-highest average excess returns are found in the fourth effective cost quintile. As in the NYSE/Amex sample, the largest variation in average effective costs is found in the higher quintiles.

Although these results provide some evidence for a liquidity premium, there is a strong possibility that effective cost is acting as a proxy for some priced risk factor. To investigate this, two types of return specifications were estimated: a one-factor market model and a three-factor Fama-French model. The one-factor market model is estimated to provide a point of comparison for Amihud and Mendelson (1986) and Amihud (2000) and other studies.

The one-factor specification is:

$$r_{i,j,t} = a_{i,j} + \beta_{i,j}\, r_{m,t} + e_{i,j,i} \tag{12}$$

where i and j index portfolios: i indexes market capitalization quintiles, j indexes effective cost quintiles. $r_{i,j,t}$ and $r_{m,t}$ are excess returns on the portfolio and the Fama-French market factor. The model is estimated in a GMM framework: the reported t-statistics are based on an error covariance structure that allows for heteroskedasticity and cross-sectional correlation in the $e_{i,j,t}$.

Table 9 reports the estimates of the model intercepts (the $a_{i,j}$). The estimates for the NYSE/Amex sample (Panel A) display a cross-sectional pattern similar to that found for the average excess returns (cf. Table 7). As before, the highest values are found in the highest effective cost quintiles, but there is no evident pattern in the lower effective cost groups. Estimates for the Nasdaq sample (Panel B) are also similar to the corresponding averages (cf. Table 8).

The three-factor Fama-French specification is:

$$r_{i,j,t} = a_{i,j} + \beta_{i,j}\, r_{m,t} + s_{i,j}\, SMB_t + h_{i,j}\, HML_t + e_{i,j,i} \qquad (13)$$

where $r_{i,j,t}$ is the average portfolio excess return in month t, i and j index market capitalization and effective cost quintiles. $r_{m,t}$, SMB_t and HML_t are respectively the Fama-French excess market return, size and book-to-market factors.

Panel A of Table 10 reports estimates of the intercepts for the NYSE/Amex sample. The results are less conclusive than the corresponding single-factor estimates. In the lowest market capitalization quintile, consistent with a positive liquidity premium, the highest intercept is found in the highest effective cost portfolio. This is not the case, however, for the other market capitalization quintiles. The results for the Nasdaq sample, however, reported in Panel B, remain essentially unaltered from those for the single-factor model.

The discussion to this point has been based on the intercepts in multifactor specifications. Parametric models offer another perspective. The Fama-French factor model in eq (13) is modified to include a term that is quadratic in the effective cost:

$$r_{i,j,t} = \left(a_0 + a_1 c_{i,j}^{Gibbs} + a_2 \left[c_{i,j}^{Gibbs}\right]^2\right) + \beta_{i,j}\, r_{m,t} + s_{i,j}\, SMB_t + h_{i,j}\, HML_t + e_{i,j,i} \qquad (14)$$

where $c_{i,j}^{Gibbs}$ is the mean Gibbs estimate of effective cost in portfolio i, j. Table 11 reports estimates of a_0, a_1, and a_2. Figure 1 depicts the relation between effective costs and excess returns implied by the estimates.

For the NYSE/Amex analysis, the implied function is essentially flat over the region that encompasses the preponderance of stocks in this sample. The highest mean value of the effective cost in the portfolios is approximately 0.02 (cf. Panel B of Table 7).

Beyond this point, however, the relation is strongly positive. The Nasdaq sample, in contrast displays a consistently positive relationship throughout the range. It is noteworthy that in both samples, the curvature is convex, rather than concave (as suggested by the model of Amihud and Mendelson (1986)).

With respect to the existence and direction of the liquidity premium, these results are broadly consistent with the results of earlier studies based on alternative measures. As the present study is based on effective costs, the most directly comparable earlier studies at those based on posted bid-ask spreads: Ho and Stoll (1981), Amihud and Mendelson (1980), Amihud and Mendelson (1989), Eleswarapu and Reinganum (1993), Kadlec, McConnell, and Purdue U (1994), and Eleswarapu (1997). Most of these studies find a positive liquidity premium, with stronger results for Nasdaq than NYSE/Amex. To the extent that the effective cost is partially proxying for asymmetric information and/or price impact, the present results can be viewed as consistent with the studies of Brennan and Subrahmanyam (1996), Easley, Hvidkjaer, and O'Hara (1999), Chordia, Subrahmanyam, and Anshuman (2001), and Amihud (2002).

11. Conclusion

Motivated by the need for trading cost measures in samples where we don't possess detailed trading data, this paper addresses the problem of inferring trading costs from daily data. The first step of the analysis is to construct a set of trading cost measures from daily CRSP price and volume data, and then to compare these proxies to measures constructed from TAQ trade and quote data.

Two common TAQ-based trading cost measures are the effective cost (the difference between the trade price and the prevailing quote midpoint) and the price impact coefficient (the permanent impact on the price for a trade of a given size). To measure effective cost in daily data, this study examines two estimates of the bid ask spread based on the Roll (1984) model: the conventional moment estimate (a transformation of the first-order return autocovariance) and a Gibbs sampler estimate. In this context, the Gibbs estimate is the clear winner. Its correlation with the TAQ-based estimate of effective cost is 0.90 in individual stocks and 0.98 in portfolios. Furthermore, unlike the moment estimate of the effective cost, the Gibbs estimate is always defined and positive in small samples.

Price impact measures, however, are more difficult to proxy. The present paper examines the relationship between price impact coefficients estimated 15-minute return/signed order flow specifications for the TAQ data and three proxies estimated from daily return/volume data. These proxies are the liquidity ratio, the illiquidity ratio and the reversal measure. Among these, the illiquidity ratio appears to have the strongest correlation with the transaction-level estimated impact coefficient. The sample distributions of all estimates, however, exhibit an extreme tail. This suggests that when these estimates are used as proxy variables in subsequent analyses, robust statistical methods should be considered.

The strong performance of the Gibbs effective cost estimates in the TAQ comparisons supports reliance on these estimates outside of the TAQ sample period. The second part of this paper considers Gibbs effective cost estimates computed over the full range of the daily CRSP file (beginning in 1962) and their relation to returns. The estimates suggest that average effective cost has varied substantially over the past forty years, but that this variation is largely driven by low-capitalization issues. Effective trading cost for the highest market value quintile has remained relatively stable over the period.

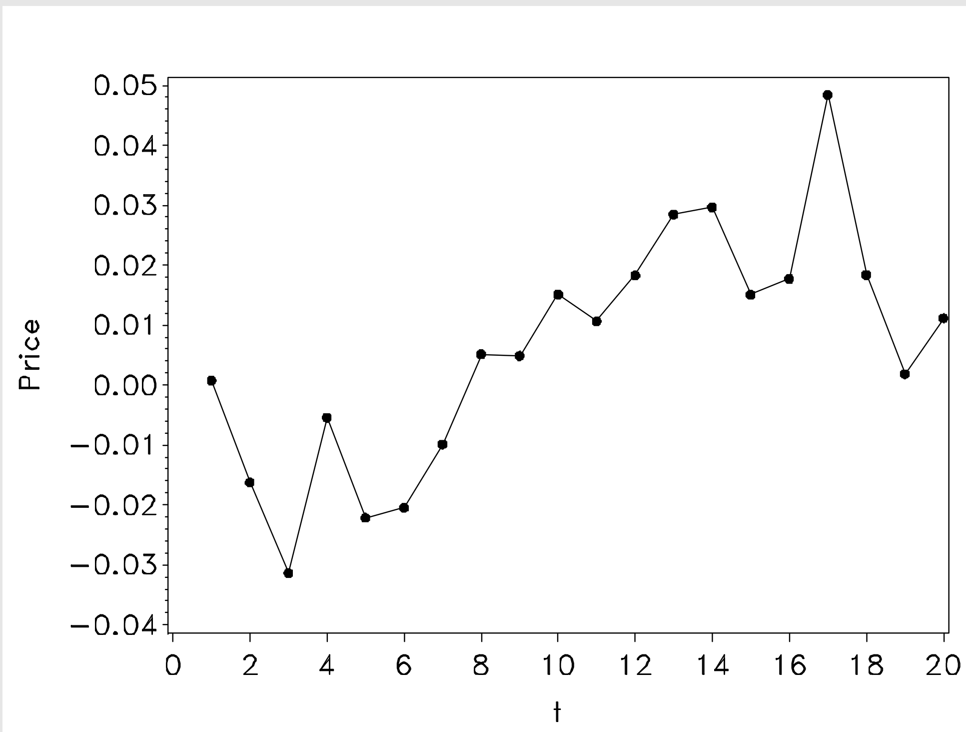
Portfolios are formed by grouping on effective cost and market capitalization. The pattern of average excess returns on these portfolios is suggestive of a trading cost ("liquidity") premium. Portfolios with high average effective costs exhibit relatively high average excess returns. The same pattern arises in the intercepts of one-factor market-model specifications. When excess returns are estimated in a three-factor Fama-French model, the pattern in the intercepts is less conclusive. The Nasdaq sample, (which exhibits the largest cross-sectional variation in effective cost), still displays evidence of a liquidity premium, but the NYSE/Amex sample (which covers the longest time period) does not. In parametric specifications, however, where the dependence of excess returns on effective cost is specified as a quadratic function, both samples evince economically and statistically significant evidence of a liquidity premium. This is broadly consistent with the results of earlier studies.

The analysis suggests 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.

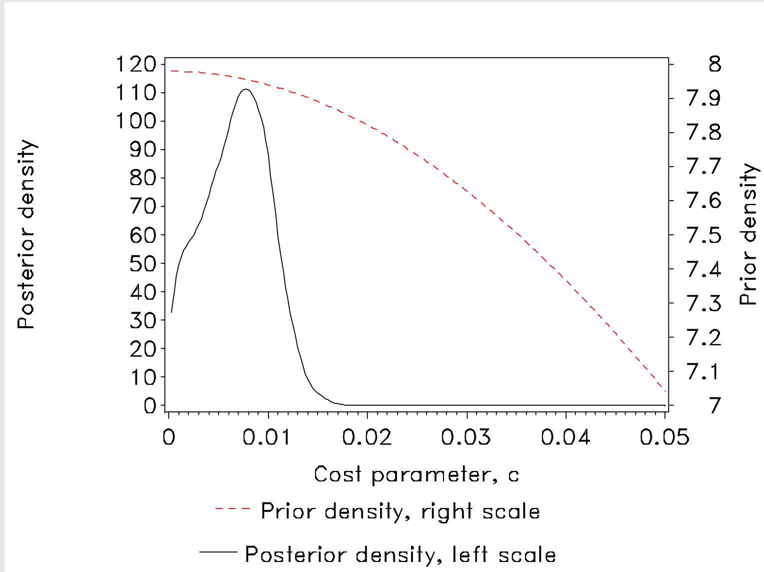
12. Appendix: An illustration of Gibbs estimates of the Roll Model

This section discusses the analysis of a simple simulated price record using the Gibbs sampler. The model is described in section 3.a and the Gibbs estimator is described in section 3.c. The parameter values used for the simulation are c=0.01 and $\sigma_u = 0.01$. Since the model is stated in log terms, these values imply a standard deviation and half-spread of approximately one percent. Starting at an initial valueof zero, twenty price observations were simulated. The price path exhibits both nonstationarity from the random-walk component of the price, and also reversals from bid-ask bounce (Figure 1).

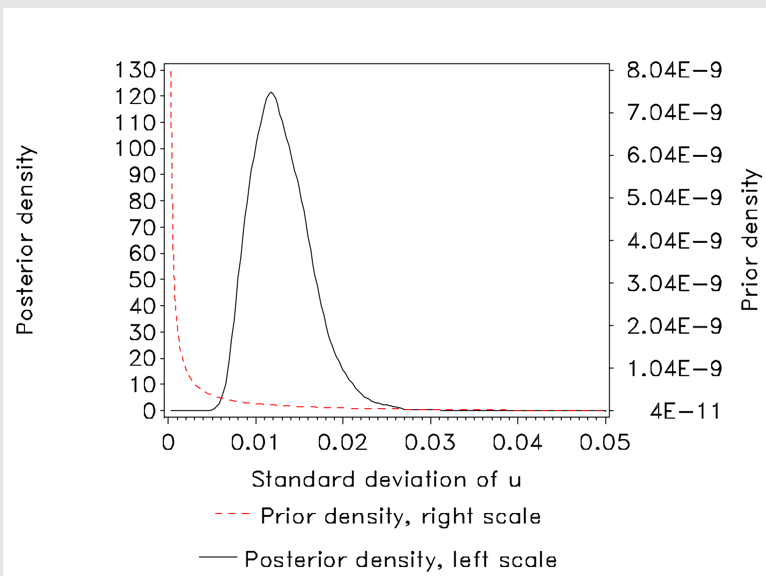


Appendix Figure 1. The simulated (log) price path.

Prior and (smoothed, simulated) posterior distributions are presented for c in Figure 2, and for σ_u in Figure 3. The prior for c used in this appendix is $N^+\left(0, \sigma_c^{2,prior} = 0.01\right)$.



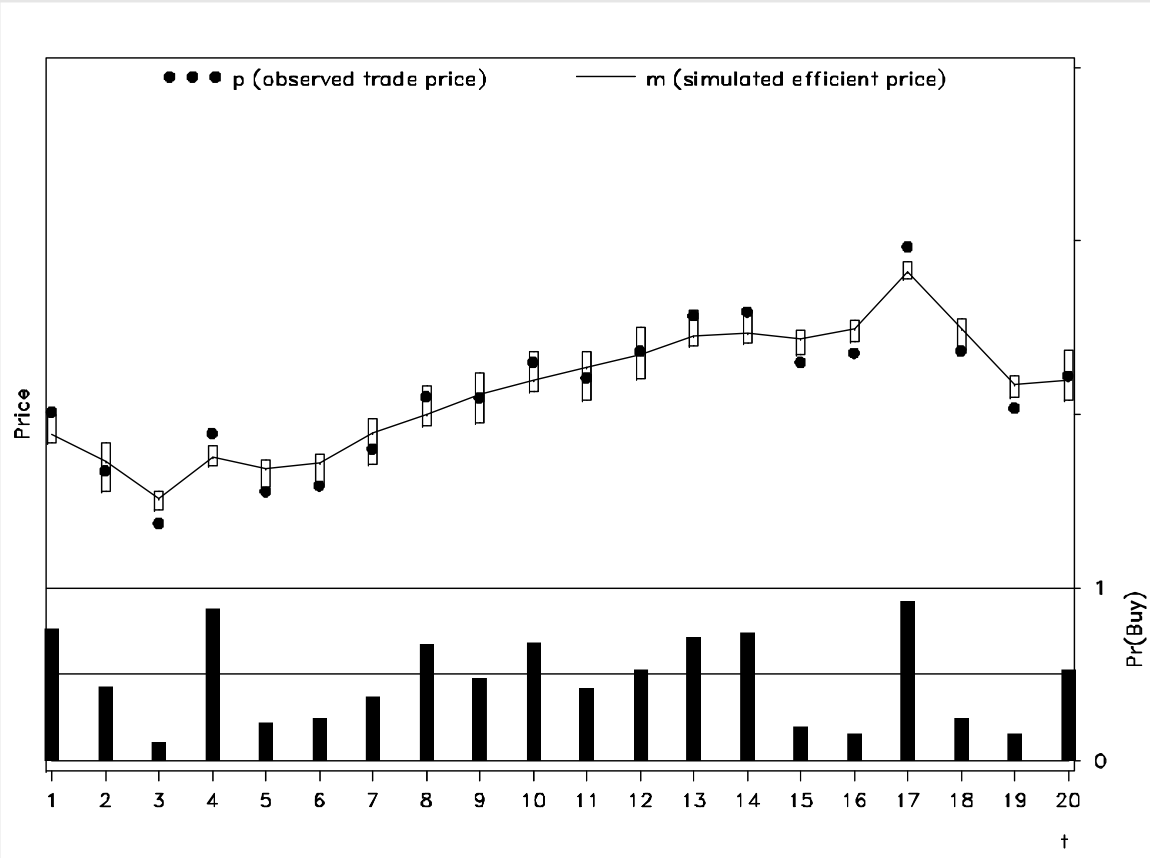
Appendix Figure 2. Prior and posterior distributions for the cost parameter.



Appendix Figure 3. Prior and posterior distributions for σ_u.

In both figures the dotted lines depict the prior distributions. The solid lines describe the posteriors. The latter are constructed as the kernel smoothed distributions of the Gibbs draws. Note that the scales of the priors and posteriors are different. The posteriors are concentrated in regions where the priors are relatively flat. Essentially, the posteriors are data dominated.

In addition to parameter posteriors, the Gibbs procedure also produces posteriors for the latent data in the model – in this case the implicit efficient prices m_t and the trade direction indicators q_t. Although these are not analyzed in the main body of the paper, they provide useful confirmation for the reasonableness of the procedure. Figure 4 describes the distributions of the *m* and *q*.



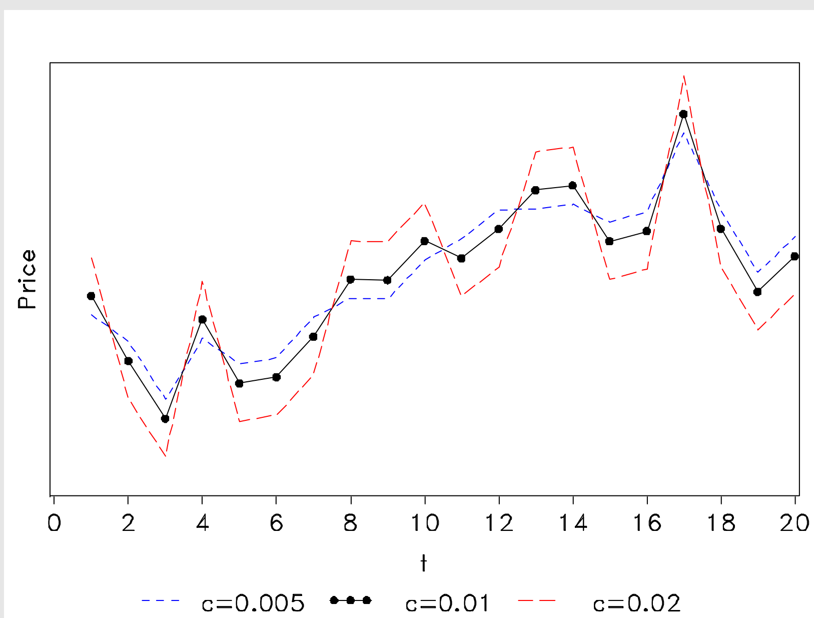
Appendix Figure 4. Gibbs estimates of latent data.

The figure presents two stacked graphs aligned by time. In the top section, the observed prices are plotted as dots. At each time, the posterior distribution of the efficient price is indicated by the box plots. The limits of the box represent the twenty-fifth and seventy-fifth percentiles of the distribution. A line joins the medians.

Visually, the posteriors for the efficient prices resemble a smoothed version of the observed prices. This is reasonable, because the efficient prices are in principle purged of bid-ask bounce. Note too that the posteriors are not uninformly tight. When the observed prices exhibit a well-defined reversal (at times 3, 4, and 17, for example), the posteriors are more concentrated than when the price path is smoother (in the middle range of the sample).

The bottom section in the figure graphs the posterior probability that the trade was a "buy". A value near one (cf. times 4 and 17) indicates a relatively high certainty that the trade was a buy. A value near zero (time 3, for example) suggests a relatively high certainty that the trade is a sell. Certainty is highest when there is a clear reversal, as one would expect. In the middle range of the sample, the posterior probabilities are approximately fifty percent.

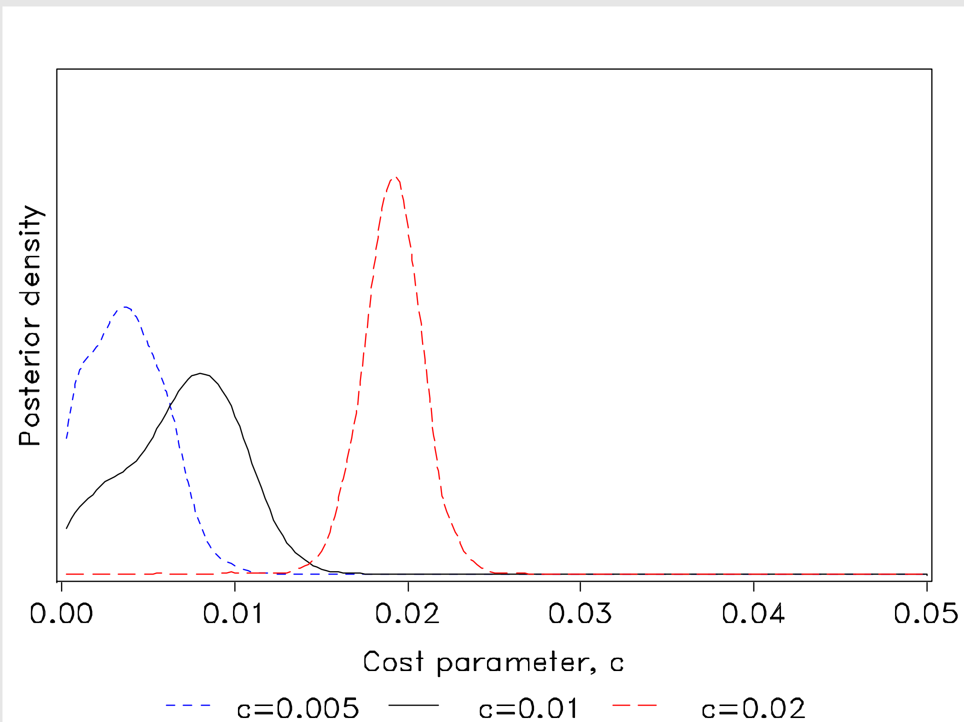
It is also useful to consider how inference changes when the relative values of c and σ_u change. Figure 5 presents three versions of the original simulated price path. Each uses a different value of c, while keeping constant the latent efficient price and trade direction series. The central line marked by black dots is the base case (c=0.01); the dashed lines follow from lower or higher values of c. Changing the value of c has the effect of exaggerating or attenuating the bid-ask bounce.



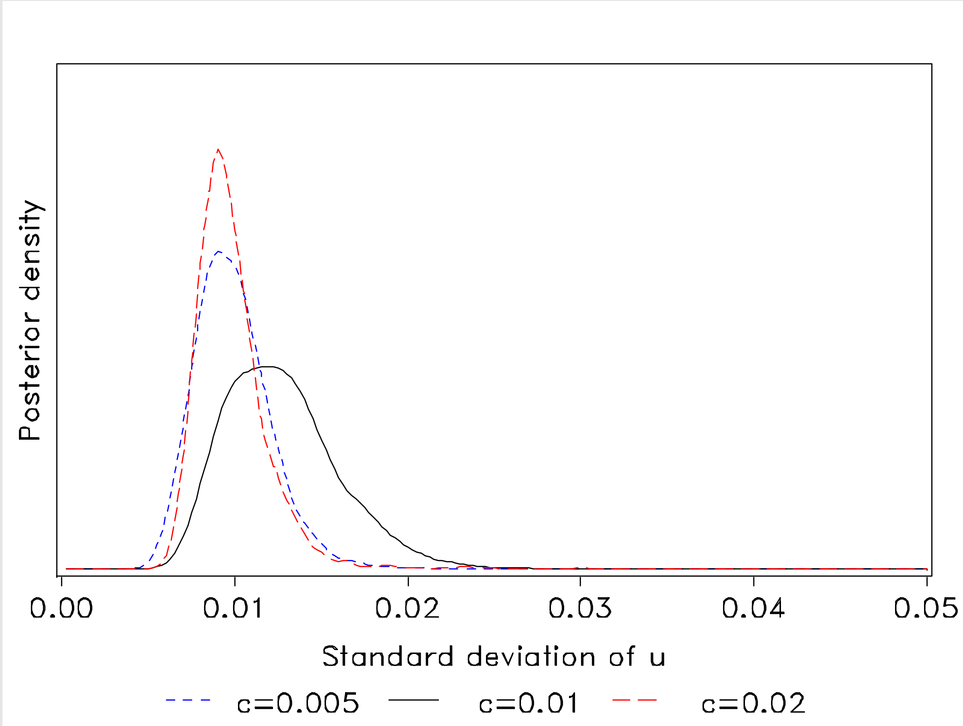
Appendix Figure 5. Alternative simulated price paths.

Figure 6 depicts the parameter posteriors for c. The sharpest (most well-defined) posterior is obtained for the highest value of c. This is the case where bid-ask bounce is most well-defined, and it is easiest (both visually and in the estimation) to judge trade direction. For the lowest value of c, it is difficult to separate out the bid-ask bounce and random-walk components. This translates into a relatively broad posterior that runs up against the nonnegativity constraint for the parameter (implied by the prior).

Figure 7 depicts the parameter posteriors for σ_u. It is noteworthy that these posteriors are relatively sharp for both the higher and lower values of c. For the higher value of c, the well-defined bid-ask bounce noted above also provides good identification of the efficient price. In the case of low c, the bid-ask bounce is not well-defined, but its magnitude is sufficiently low that the observed price changes are dominated by the efficient price changes. It is in the intermediate case, when the bid-ask bounce and efficient price change components are of comparable magnitudes that resolution is most difficult.



Appendix Figure 6. Alternative parameter posteriors for *c*.



Appendix Figure 7. Alternative parameter posteriors for σ_u.

13. 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., 2000. Illiquidity and stock returns: Cross-section and time-series effects. Unpublished working paper. Stern School, NYU.

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

Amihud, Y., Mendelson, H., 1980. Dealership markets: Market making with inventory. Journal of Financial Economics 8, 31-53.

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

———. 1989. The Effects of Beta, Bid-Ask Spread, Residual Risk, and Size on Stock Returns. *Journal-of-Finance; 44(2), June 1989, Pages 479-86.*

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

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

Chordia, T., Subrahmanyam, A., Anshuman, V. R., 2001. Trading activity and expected stock returns. Journal of Financial Economics 59, 3-32.

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

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., 1999. Is information risk a determinant of asset returns? Unpublished working paper. Johnson School, Cornell University.

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

Eleswarapu, V.-R. A. U. A. 1997. Cost of Transacting and Expected Returns in the Nasdaq Market. *Journal-of-Finance; 52(5), December 1997, Pages 2113-27*.

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

Harris, L. E., 1990. Statistical properties of the roll serial covariance bid/ask spread estimator. Journal of Finance 45, 579-90.

Hasbrouck, J., 1999. Liquidity in the futures pits: Inferring market dynamics from incomplete data. Unpublished working paper. Stern School of Business, New York University, www.stern.nyu.edu\~jhasbrou.

Ho, T. S. Y., Stoll, H. R., 1981. Optimal dealer pricing under transactions and return uncertainty. Journal of Financial Economics 9, 47-73.

Jones, C. M., 2001. A century of stock market liquidity and trading costs. Unpublished working paper. Columbia University Graduate School of Business.

Kadlec, G.-B., J.-J. A. V. P. I. &. S. U. McConnell, and Purdue U. 1994. The Effect of Market Segmentation and Illiquidity on Asset Prices: Evidence from Exchange Listings. *Journal-of-Finance; 49(2), June 1994, Pages 611-36*.

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

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.

Kim, C.-J., Nelson, C. R., 2000. State-space models with regime switching. MIT Press, Cambridge, Massachusetts.

Pastor, L., Stambaugh, R. F., 2002. Liquidity risk and expected stock returns. Unpublished working paper. University of Chicago.

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.

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

Table 1. Summary of the TAQ comparison sample

From the TAQ database, 200 firms were randomly drawn for each of the years 1993-2001 (1,800) firms. Only those firms that could be matched to the CRSP database were retained. The table reports numbers of firms by year and listing exchange.

	All	Exchange		
		Amex	NYSE	Nasdaq
	N	N	N	N
All	1,665	166	516	983
Year				
1993	189	23	54	112
1994	188	23	56	109
1995	190	16	66	108
1996	188	16	49	123
1997	187	17	60	110
1998	190	14	62	114
1999	180	21	62	97
2000	181	12	53	116
2001	172	24	54	94

Table 2. Trading cost measures based on transactions data (TAQ comparison sample)

The TAQ comparison sample consists of 1,800 firm-years randomly drawn from the TAQ database (200 in each year, 1993 to 2001), restricted to those that could be matched to the CRSP database. For a given stock, the average effective cost is the average absolute difference between the log trade price and the prevailing log quote midpoint, over all trades in the year, weighted by dollar volume of the trade. The half log spread is the time-weighted average of $\log(ask/bid)/2$ using all primary market quotes for the year. Models I-IV refer to linear specifications, estimated separately for each firm, of fifteen-minute returns and fifteen-minute aggregate signed volume:

Model I: $r_t = \lambda^I N_t + u_t$
Model II: $r_t = \lambda^{II} S_t + u_t$
Model III: $r_t = \lambda^{III} V_t + u_t$
Model IV: $r_t = \lambda_1^{IV} N_t + \lambda_2^{IV} S_t + \lambda_3^{IV} V_t + u_t$

where N_t is the signed number of trades in fifteen-minute interval t; V_t is the signed dollar volume; and, S_t is the cumulative signed square-root dollar volume.

Variable	N	Mean	Std. Dev.	Skewness	Kurtosis
Effective cost	1,665	0.014	0.017	3.418	17.751
Half log spread	1,665	0.020	0.024	2.704	9.856
R^2 for Model I	1,664	0.127	0.103	1.141	1.532
R^2 for Model II	1,664	0.123	0.092	0.886	1.069
R^2 for Model III	1,664	0.050	0.055	1.982	6.108
R^2 for Model IV	1,664	0.159	0.108	0.847	0.819
λ^{II}	1,664	0.00003	0.00006	6.898	65.612

Table 3. Trading cost measures based on daily CRSP data (TAQ comparison sample)

Estimates for each firm are based on CRSP daily returns for the year. The number of firms is less than 1,800 due to matching failures between CRSP and TAQ. c^{Gibbs} is the Gibbs-sampler estimate of the effective cost; c^{M} is the moment estimate of the effective cost; and c^{MZ} is equal to c^{M} (when defined) and zero otherwise. L is the liquidity ratio $L = \overline{\left(Vol_d / |r_d|\right)}$ where Vol_d is the dollar volume on day d, and r_d is the return on day d, and the average is taken over all days in the year. I is the illiquidity ratio $I = \overline{\left(|r_d| / Vol_d\right)}$. γ is the reversal liquidity measure estimated from the regression $r_{d+1} = \theta + \phi r_d + \gamma \, sign\left(r_d^e\right) Vol_d + \varepsilon_d$, where r_d^e is the excess return on day d.

	Variable	N	Mean	Std. Dev.	Skewness	Kurtosis
Spread-related cost proxies	c^{Gibbs}	1,668	0.014	0.019	3.427	16.073
	c^{M}	1,201	0.019	0.017	2.153	7.035
	c^{MZ}	1,668	0.014	0.017	2.232	7.476
Impact-related cost proxies	L	1,668	794	5251	17.237	388
	I	1,668	6.286	29.923	13.866	272
	γ	1,668	0.0051	0.318	-6.536	148

Table 4. Correlations in the TAQ/CRSP comparison sample

The TAQ/CRSP comparison sample comprises roughly 200 firms per year, for the years 1993-2001, randomly chosen from the TAQ database, that could be subsequently matched to CRSP (a total of 1,664 firms). c^{TAQ} is the effective cost estimated from transaction-level TAQ data; c^{Gibbs} and c^{MZ} are estimates of the effective cost based on daily CRSP data: c^{Gibbs} is the Gibbs-sampler estimate of the effective cost; c^{MZ} is the moment estimate of the effective cost (when defined) and zero otherwise. λ^{II} is a signed-trade price impact measure estimated from TAQ data using the specification: $r_t = \lambda^{II} S_t + u_t$, where r_t is the return, S_t is the cumulative signed square-root dollar volume, and t indexes fifteen-minute intervals. L, I, and γ are impact proxies based on daily CRSP data: L is the liquidity ratio $L = \overline{\left(Vol_d / |r_d|\right)}$ where Vol_d is the dollar volume on day d, and r_d is the return on day d, and the average is taken over all days in the year. I is the illiquidity ratio $I = \overline{\left(|r_d| / Vol_d\right)}$. γ is the reversal liquidity measure estimated from the regression $r_{d+1} = \theta + \phi r_d + \gamma\, sign\left(r_d^e\right) Vol_d + \varepsilon_d$, where r_d^e is the excess return on day d.

Panel A. Correlations (Pearson, full)

	Effective cost measures			Impact measures			
	TAQ	CRSP		TAQ	CRSP		
	c^{TAQ}	c^{Gibbs}	c^{MZ}	λ^{II}	L	I	γ
c^{TAQ}	1.000	0.901	0.825	0.515	-0.112	0.641	-0.051
c^{Gibbs}	0.901	1.000	0.880	0.397	-0.071	0.657	-0.028
c^{MZ}	0.825	0.880	1.000	0.391	-0.084	0.562	0.078
λ^{II}	0.515	0.397	0.391	1.000	-0.060	0.473	-0.058
L	-0.112	-0.071	-0.084	-0.060	1.000	-0.031	-0.004
I	0.641	0.657	0.562	0.473	-0.031	1.000	0.178
γ	-0.051	-0.028	0.078	-0.058	-0.004	0.178	1.000

Panel B. Correlations (Spearman, full)

	Effective cost measures			Impact measures			
	TAQ	CRSP		TAQ	CRSP		
	c^{TAQ}	c^{Gibbs}	c^{MZ}	λ^{II}	L	I	γ
c^{TAQ}	1.000	0.851	0.756	0.658	-0.924	0.934	0.312
c^{Gibbs}	0.851	1.000	0.867	0.461	-0.741	0.782	0.293
c^{MZ}	0.756	0.867	1.000	0.405	-0.671	0.706	0.300
λ^{II}	0.658	0.461	0.405	1.000	-0.763	0.737	0.213
L	-0.924	-0.741	-0.671	-0.763	1.000	-0.968	-0.287
I	0.934	0.782	0.706	0.737	-0.968	1.000	0.297
γ	0.312	0.293	0.300	0.213	-0.287	0.297	1.000

Table 4. Correlations in the TAQ/CRSP comparison sample (continued)

Panel C. Correlations (Pearson, partial with respect to log market capitalization)

	Effective cost measures			Impact measures			
	TAQ	CRSP		TAQ	CRSP		
	c^{TAQ}	c^{Gibbs}	c^{MZ}	λ^{II}	L	I	γ
c^{TAQ}	1.000	0.850	0.725	0.365	0.137	0.610	-0.093
c^{Gibbs}	0.850	1.000	0.821	0.225	0.151	0.620	-0.055
c^{MZ}	0.725	0.821	1.000	0.211	0.145	0.498	0.072
λ^{II}	0.365	0.225	0.211	1.000	0.088	0.404	-0.080
L	0.137	0.151	0.145	0.088	1.000	0.077	0.009
I	0.610	0.620	0.498	0.404	0.077	1.000	0.176
γ	-0.093	-0.055	0.072	-0.080	0.009	0.176	1.000

Panel D. Correlations (Spearman, partial with respect to log market capitalization)

	Effective cost measures			Impact measures			
	TAQ	CRSP		TAQ	CRSP		
	c^{TAQ}	c^{Gibbs}	c^{MZ}	λ^{II}	L	I	γ
c^{TAQ}	1.000	0.662	0.544	0.090	-0.639	0.682	0.110
c^{Gibbs}	0.662	1.000	0.768	-0.108	-0.310	0.444	0.119
c^{MZ}	0.544	0.768	1.000	-0.086	-0.307	0.411	0.150
λ^{II}	0.090	-0.108	-0.086	1.000	-0.387	0.292	-0.001
L	-0.639	-0.310	-0.307	-0.387	1.000	-0.819	-0.044
I	0.682	0.444	0.411	0.292	-0.819	1.000	0.062
γ	0.110	0.119	0.150	-0.001	-0.044	0.062	1.000

Table 5. Correlations in the TAQ/CRSP comparison sample with grouping by effective cost.

The TAQ/CRSP comparison sample comprises roughly 200 firms per year, for the years 1993-2001, randomly chosen from the TAQ database, that could be subsequently matched to CRSP (a total of 1,664 firms). c^{TAQ} is the effective cost estimated from transaction-level TAQ data; c^{Gibbs} and c^{MZ} are estimates of the effective cost based on daily CRSP data: c^{Gibbs} is the Gibbs-sampler estimate of the effective cost; c^{MZ} is the moment estimate of the effective cost (when defined) and zero otherwise. λ^{II} is a signed-trade price impact measure estimated from TAQ data. L, I, and γ are impact proxies based on daily CRSP data: L is the liquidity ratio $L = \overline{\left(Vol_d / |r_d|\right)}$ where Vol_d is the dollar volume on day d, and r_d is the return on day d. I is the illiquidity ratio $I = \overline{\left(|r_d| / Vol_d\right)}$. γ is the reversal liquidity measure. Within each year, ten groups were formed by ranking on c^{TAQ}. Reported correlations are between group means (90 observations).

Panel A. Correlations (Pearson, full)

	Effective cost measures			Impact measures			
	TAQ	CRSP		TAQ	CRSP		
	c^{TAQ}	c^{Gibbs}	c^{MZ}	λ^{II}	L	I	γ
c^{TAQ}	1.000	0.987	0.970	0.753	-0.287	0.877	-0.001
c^{Gibbs}	0.987	1.000	0.962	0.727	-0.220	0.897	-0.001
c^{MZ}	0.970	0.962	1.000	0.759	-0.258	0.804	0.091
λ^{II}	0.753	0.727	0.759	1.000	-0.210	0.754	-0.300
L	-0.287	-0.220	-0.258	-0.210	1.000	-0.134	-0.042
I	0.877	0.897	0.804	0.754	-0.134	1.000	-0.180
γ	-0.001	-0.001	0.091	-0.300	-0.042	-0.180	1.000

Panel B. Correlations (Pearson, partial with respect to log market capitalization)

	Effective cost measures			Impact measures			
	TAQ	CRSP		TAQ	CRSP		
	c^{TAQ}	c^{Gibbs}	c^{MZ}	λ^{II}	L	I	γ
c^{TAQ}	1.000	0.981	0.923	0.537	0.415	0.884	-0.115
c^{Gibbs}	0.981	1.000	0.931	0.509	0.419	0.880	-0.095
c^{MZ}	0.923	0.931	1.000	0.547	0.531	0.762	0.030
λ^{II}	0.537	0.509	0.547	1.000	0.306	0.644	-0.462
L	0.415	0.419	0.531	0.306	1.000	0.286	0.019
I	0.884	0.880	0.762	0.644	0.286	1.000	-0.266
γ	-0.115	-0.095	0.030	-0.462	0.019	-0.266	1.000

Table 6. Correlations in the TAQ/CRSP comparison sample with grouping by trade impact coefficient, λ^{II}.

The TAQ/CRSP comparison sample comprises roughly 200 firms per year, for the years 1993-2001, randomly chosen from the TAQ database, that could be subsequently matched to CRSP (a total of 1,664 firms). c^{TAQ} is the effective cost estimated from transaction-level TAQ data; c^{Gibbs} and c^{MZ} are estimates of the effective cost based on daily CRSP data: c^{Gibbs} is the Gibbs-sampler estimate of the effective cost; c^{MZ} is the moment estimate of the effective cost (when defined) and zero otherwise. λ^{II} is a signed-trade price impact measure estimated from TAQ data. L, I, and γ are impact proxies based on daily CRSP data: L is the liquidity ratio $L = \overline{\left(Vol_d / |r_d|\right)}$ where Vol_d is the dollar volume on day d, and r_d is the return on day d. I is the illiquidity ratio $I = \overline{\left(|r_d| / Vol_d\right)}$. γ is the reversal liquidity measure. Within each year, ten groups were formed by ranking on λ^{II}. Reported correlations are between group means (90 observations).

Panel A. Correlations (Pearson, full)

	Effective cost measures			Impact measures			
	TAQ	CRSP		TAQ	CRSP		
	c^{TAQ}	c^{Gibbs}	c^{MZ}	λ^{II}	L	I	γ
c^{TAQ}	1.000	0.977	0.957	0.792	-0.310	0.855	0.245
c^{Gibbs}	0.977	1.000	0.950	0.785	-0.229	0.851	0.234
c^{MZ}	0.957	0.950	1.000	0.784	-0.257	0.789	0.257
λ^{II}	0.792	0.785	0.784	1.000	-0.171	0.899	0.128
L	-0.310	-0.229	-0.257	-0.171	1.000	-0.142	-0.049
I	0.855	0.851	0.789	0.899	-0.142	1.000	0.302
γ	0.245	0.234	0.257	0.128	-0.049	0.302	1.000

Panel B. Correlations (Pearson, partial with respect to log market capitalization)

	Effective cost measures			Impact measures			
	TAQ	CRSP		TAQ	CRSP		
	c^{TAQ}	c^{Gibbs}	c^{MZ}	λ^{II}	L	I	γ
c^{TAQ}	1.000	0.953	0.865	0.596	0.466	0.799	0.138
c^{Gibbs}	0.953	1.000	0.876	0.588	0.476	0.760	0.123
c^{MZ}	0.865	0.876	1.000	0.578	0.564	0.648	0.158
λ^{II}	0.596	0.588	0.578	1.000	0.384	0.837	-0.007
L	0.466	0.476	0.564	0.384	1.000	0.365	0.101
I	0.799	0.760	0.648	0.837	0.365	1.000	0.228
γ	0.138	0.123	0.158	-0.007	0.101	0.228	1.000

Table 7. Summary statistics for the NYSE/Amex portfolios, 1963-2001

Average monthly excess returns for NYSE/Amex portfolios formed by independent ranking on the market capitalization at the end of the prior year and the Gibbs estimate of the effective cost formed over the prior year. Market capitalization quintiles were constructed by collapsing Crsp market capitalization deciles.

Panel A. Average excess returns

		Effective Cost (c^{Gibbs}) Quintiles				
		Low	2	3	4	High
	Low	0.0067	0.0069	0.0067	0.0067	0.0132
Market	2	0.0068	0.0075	0.0077	0.0067	0.0110
Capitalization	3	0.0072	0.0072	0.0068	0.0071	0.0112
Quintiles	4	0.0060	0.0077	0.0073	0.0067	0.0105
	High	0.0055	0.0080	0.0059	0.0045	0.0094

Panel B. Average Effective Cost

		Effective Cost (c^{Gibbs}) Quintiles				
		Low	2	3	4	High
	Low	0.0014	0.0023	0.0033	0.0053	0.0194
Market	2	0.0014	0.0023	0.0033	0.0053	0.0159
Capitalization	3	0.0015	0.0023	0.0033	0.0052	0.0157
Quintiles	4	0.0015	0.0023	0.0033	0.0052	0.0154
	High	0.0016	0.0023	0.0033	0.0052	0.0163

Panel C. Average Market Capitalization

		Effective Cost (c^{Gibbs}) Quintiles				
		Low	2	3	4	High
	Low	759,190	595,538	345,181	167,253	39,301
Market	2	1,560,007	917,546	729,977	401,997	257,749
Capitalization	3	1,845,771	1,658,789	801,740	542,726	107,720
Quintiles	4	2,601,945	1,996,642	1,305,900	915,959	249,015
	High	3,248,395	2,821,790	1,995,721	993,622	467,564

Panel D. Counts

		Effective Cost (c^{Gibbs}) Quintiles				
		Low	2	3	4	High
	Low	79	57	55	59	71
Market	2	119	80	64	59	53
Capitalization	3	108	95	81	69	60
Quintiles	4	80	102	98	84	69
	High	34	81	109	115	91

Table 8. Summary statistics for the Nasdaq portfolios, 1985-2001

Average monthly excess returns for Nasdq portfolios formed by independent ranking on the market capitalization at the end of the prior year and the Gibbs estimate of the effective cost formed over the prior year. Market capitalization quintiles were constructed by collapsing Crsp market capitalization deciles.

Panel A. Average excess returns

		Effective Cost (c^{Gibbs}) Quintiles				
		Low	2	3	4	High
	Low	0.0066	0.0048	0.0035	0.0060	0.0217
Market	2	0.0067	0.0036	0.0081	0.0066	0.0096
Capitalization	3	0.0073	0.0054	0.0060	0.0083	0.0138
Quintiles	4	0.0083	0.0031	0.0046	0.0101	0.0178
	High	0.0051	0.0005	0.0027	0.0065	0.0197

Panel B. Average Effective Cost

		Effective Cost (c^{Gibbs}) Quintiles				
		Low	2	3	4	High
	Low	0.0035	0.0070	0.0123	0.0210	0.0511
Market	2	0.0032	0.0068	0.0116	0.0208	0.0469
Capitalization	3	0.0031	0.0066	0.0116	0.0199	0.0447
Quintiles	4	0.0032	0.0065	0.0116	0.0199	0.0448
	High	0.0035	0.0063	0.0113	0.0196	0.0482

Panel C. Average Market Capitalization

		Effective Cost (c^{Gibbs}) Quintiles				
		Low	2	3	4	High
	Low	212,978	135,114	78,360	57,744	20,711
Market	2	406,857	168,146	108,608	50,293	23,074
Capitalization	3	587,282	214,919	131,928	75,140	35,741
Quintiles	4	811,351	303,948	222,518	84,883	31,685
	High	1,476,968	469,413	440,918	289,626	95,571

Panel D. Counts

		Effective Cost (c^{Gibbs}) Quintiles				
		Low	2	3	4	High
	Low	31	45	88	139	199
Market	2	86	93	126	166	153
Capitalization	3	140	120	132	131	124
Quintiles	4	180	155	133	112	102
	High	214	213	142	86	66

Table 9. Market-model estimates

Estimates of the intercepts $a_{i,j}$ in the monthly excess return regression

$$r_{i,j,t} = a_{i,j} + \beta_{i,j} r_{m,t} + e_{i,j,i}$$

where $r_{i,j,t}$ is the average portfolio excess return in month t, i and j index market capitalization and effective cost (c^{Gibbs}) quintiles, and $r_{m,t}$ is the excess return on the Fama-French market factor. Reported values are GMM estimates where the error covariance matrix allows for heteroskedasticity and cross-sectional dependence. NYSE/Amex estimations span 1963-2001; Nasdaq estimations are from 1985-2001.

Panel A. NYSE/Amex

		Effective Cost (c^{Gibbs}) Quintiles				
		Low	2	3	4	High
Market Capitalization Quintiles	Low	0.0034	0.0033	0.0028	0.0026	0.0081
		(3.06)	(2.65)	(2.06)	(1.58)	(2.76)
	2	0.0029	0.0032	0.0030	0.0016	0.0053
		(3.11)	(2.82)	(2.28)	(0.95)	(1.96)
	3	0.0027	0.0022	0.0015	0.0014	0.0053
		(2.98)	(1.90)	(1.06)	(0.78)	(1.86)
	4	0.0010	0.0022	0.0013	0.0005	0.0041
		(1.00)	(1.90)	(0.92)	(0.29)	(1.39)
	High	-0.0003	0.0016	-0.0009	-0.0027	0.0026
		(-0.24)	(1.17)	(-0.58)	(-1.40)	(0.79)

Panel B. Nasdaq

		Effective Cost (c^{Gibbs}) Quintiles				
		Low	2	3	4	High
Market Capitalization Quintiles	Low	-0.0010	-0.0024	-0.0009	0.0018	0.0098
		(-0.25)	(-0.74)	(-0.29)	(0.69)	(2.64)
	2	-0.0005	-0.0023	0.0020	0.0036	0.0063
		(-0.22)	(-1.06)	(0.76)	(1.35)	(1.82)
	3	0.0001	-0.0019	-0.0010	0.0026	0.0093
		(0.02)	(-0.72)	(-0.31)	(0.72)	(2.19)
	4	0.0003	-0.0053	-0.0042	0.0033	0.0124
		(0.13)	(-1.57)	(-1.02)	(0.69)	(2.28)
	High	-0.0040	-0.0093	-0.0062	-0.0021	0.0140
		(-1.18)	(-2.13)	(-1.23)	(-0.34)	(1.89)

Table 10. Regressions of returns on Fama-French factors

Estimates of the intercepts $a_{i,j}$ in the regression

$$r_{i,j,t} = a_{i,j} + \beta_{i,j}\, r_{m,t} + s_{i,j}\, SMB_t + h_{i,j}\, HML_t + e_{i,j,i}$$

where $r_{i,j,t}$ is the average portfolio excess return in month t, i and j index market capitalization and effective cost (c^{Gibbs}) quintiles. $r_{m,t}$, SMB_t and HML_t are respectively the Fama-French excess market return, size and book-to-market factors. Reported values are GMM estimates where the error covariance matrix allows for heteroskedasticity and cross-sectional dependence. NYSE/Amex estimations span 1963-2001; Nasdaq estimations are from 1985-2001.

Panel A. NYSE/Amex

		Effective Cost (c^{Gibbs}) Quintiles				
		Low	2	3	4	High
Market Capitalization Quintiles	Low	0.0003	-0.0005	-0.0009	-0.0015	0.0024
		(0.31)	(-0.47)	(-0.78)	(-1.34)	(1.22)
	2	0.0003	-0.0002	-0.0009	-0.0027	-0.0005
		(0.44)	(-0.25)	(-0.99)	(-2.51)	(-0.26)
	3	0.0004	-0.0012	-0.0026	-0.0034	-0.0004
		(0.49)	(-1.41)	(-2.74)	(-2.86)	(-0.22)
	4	-0.0012	-0.0007	-0.0024	-0.0040	-0.0017
		(-1.37)	(-0.73)	(-2.27)	(-3.08)	(-0.84)
	High	-0.0023	-0.0010	-0.0039	-0.0064	-0.0033
		(-1.73)	(-0.83)	(-3.19)	(-4.57)	(-1.38)

Panel B Nasdaq

		Effective Cost (c^{Gibbs}) Quintiles				
		Low	2	3	4	High
Market Capitalization Quintiles	Low	-0.0039	-0.0022	-0.0010	0.0018	0.0095
		(-1.18)	(-0.78)	(-0.37)	(0.90)	(3.14)
	2	-0.0027	-0.0027	0.0016	0.0033	0.0054
		(-1.67)	(-1.65)	(0.91)	(1.72)	(1.96)
	3	-0.0013	-0.0017	-0.0002	0.0039	0.0104
		(-0.98)	(-0.97)	(-0.08)	(1.36)	(3.14)
	4	0.0008	-0.0034	-0.0010	0.0064	0.0153
		(0.46)	(-1.63)	(-0.32)	(1.70)	(3.34)
	High	-0.0002	-0.0038	-0.0004	0.0037	0.0180
		(-0.09)	(-1.30)	(-0.11)	(0.81)	(2.80)

Table 11. Factor return models with quadratic effective cost.

The specification is:

$$r_{i,j,t} = \left(a_0 + a_1 c_{i,j} + a_2 c_{i,j}^2\right) + \beta_{i,j}\, r_{m,t} + s_{i,j}\, SMB_t + h_{i,j}\, HML_t + e_{i,j,i}$$

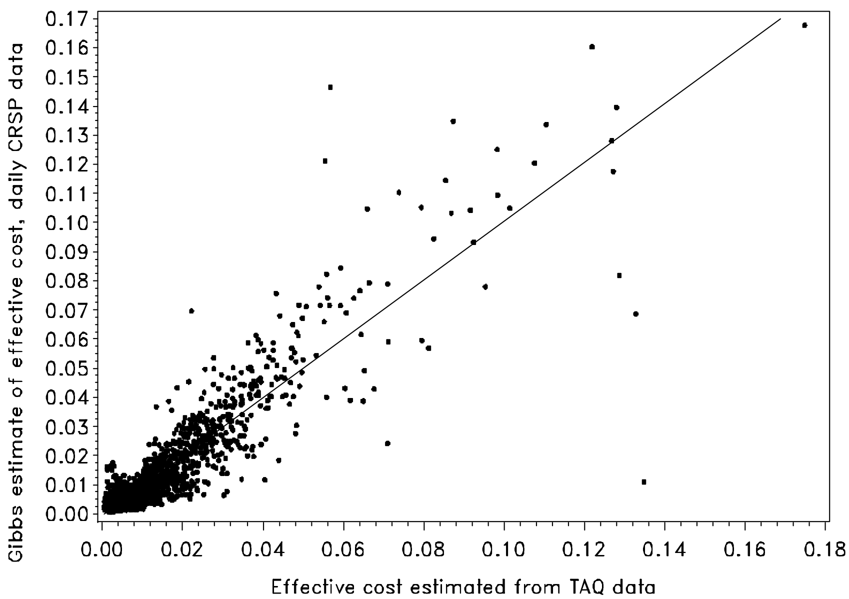
where $r_{i,j,t}$ is the average portfolio excess return in month t, i and j index market capitalization and effective cost (c^{Gibbs}) quintiles. $r_{m,t}$, SMB_t and HML_t are respectively the Fama-French excess market return, size and book-to-market factors; $c_{i,j}$ is the mean effective cost in portfolio (i, j). Reported values are GMM estimates where the error covariance matrix allows for heteroskedasticity and cross-sectional dependence. NYSE/Amex estimations span 1963-2001; Nasdaq estimations are from 1985-2001.

	a_0	a_1	a_2
NYSE/Amex	–0.001	–0.316	20.260
	(–6.54)	(–6.16)	(12.10)
Nasdaq	–0.002	0.180	1.493
	(–3.65)	(4.98)	(3.35)

Figure 1. TAQ vs. Gibbs (CRSP) estimates of effective cost (TAQ comparison sample)

The TAQ comparison sample comprises approximately 1,800 firm-years (200 firms randomly drawn from each year, 1993-2001). Only firm-years that could be matched to CRSP data were retained. The figure depicts for each firm-year the average effective cost estimated from the TAQ data vs. the Gibbs estimate based on daily CRSP returns (c^{Gibbs}).

Panel A: Full TAQ comparison sample



Panel B: Detail (TAQ effective cost estimates < 0.04)

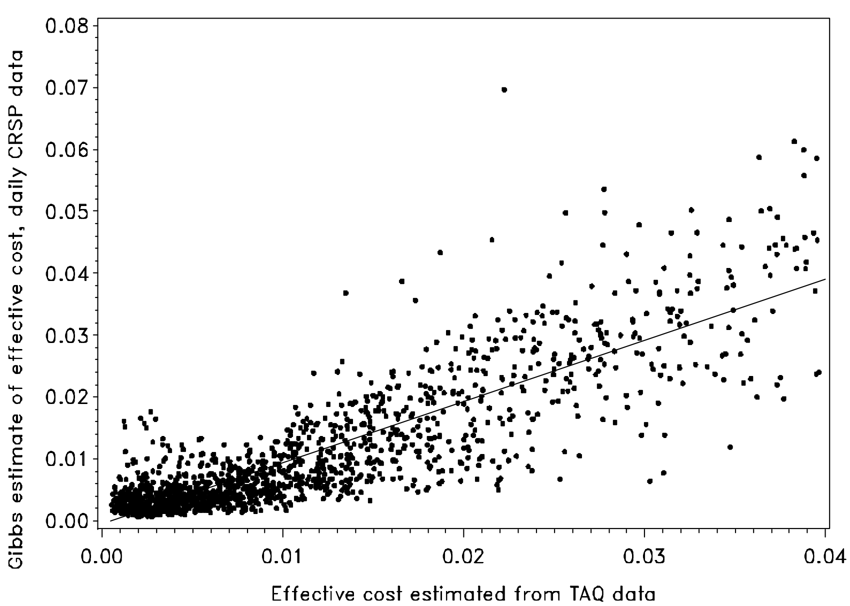


Figure 2. Gibbs estimates of effective cost by listing exchange

The sample is all ordinary common equity issues on the CRSP daily database.

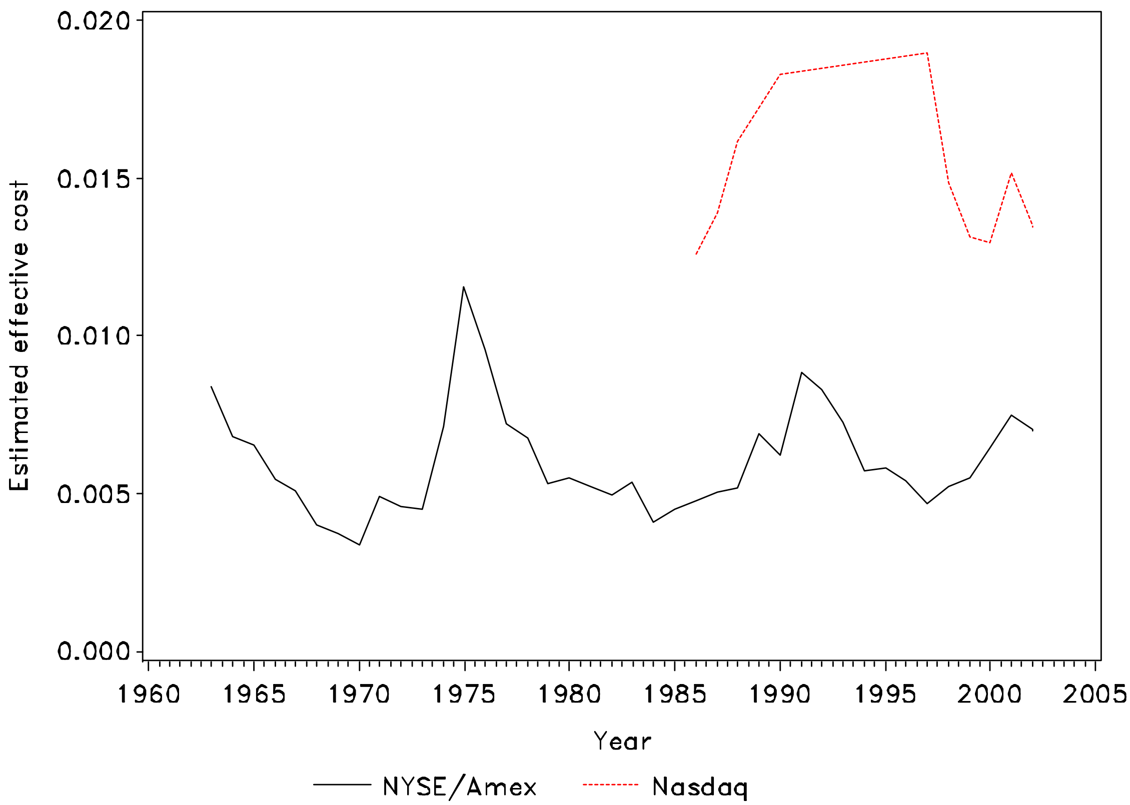
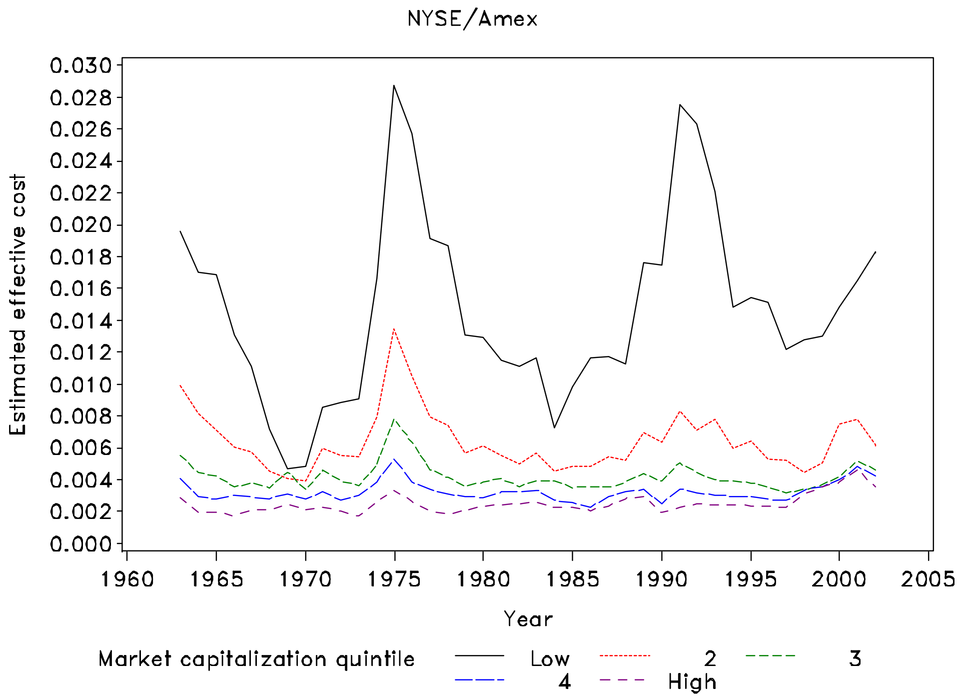


Figure 3. Gibbs estimates of effective cost by market capitalization quintile.

The sample is all ordinary common equity issues on the CRSP daily database.



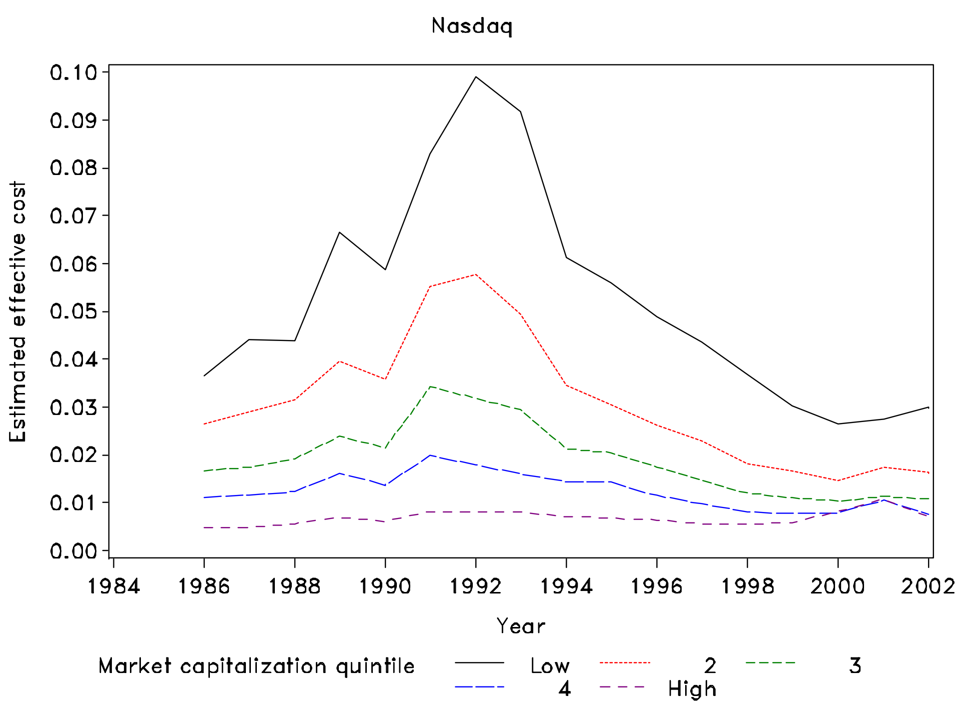


Figure 4. Relationship between effective cost and excess return implied by parametric specification.

The plots are based on the factor return model:

$$r_{i,j,t} = \left(a_0 + a_1 c_{i,j} + a_2 c_{i,j}^2\right) + \beta_{i,j}\, r_{m,t} + s_{i,j}\, SMB_t + h_{i,j}\, HML_t + e_{i,j,i}$$

where $r_{i,j,t}$ is the average portfolio excess return in month t, i and j index market capitalization and effective cost (c^{Gibbs}) quintiles. $r_{m,t}$, SMB_t and HML_t are respectively the Fama-French excess market return, size and book-to-market factors; $c_{i,j}$ is the mean effective cost in portfolio (i, j). The figures plot the estimated functions $a_1 c + a_2 c^2$, with two-standard-error bounds. The figures are based on GMM estimates where the error covariance matrix allows for heteroskedasticity and cross-sectional dependence. NYSE/Amex estimations span 1963-2001; Nasdaq estimations are from 1985-2001.

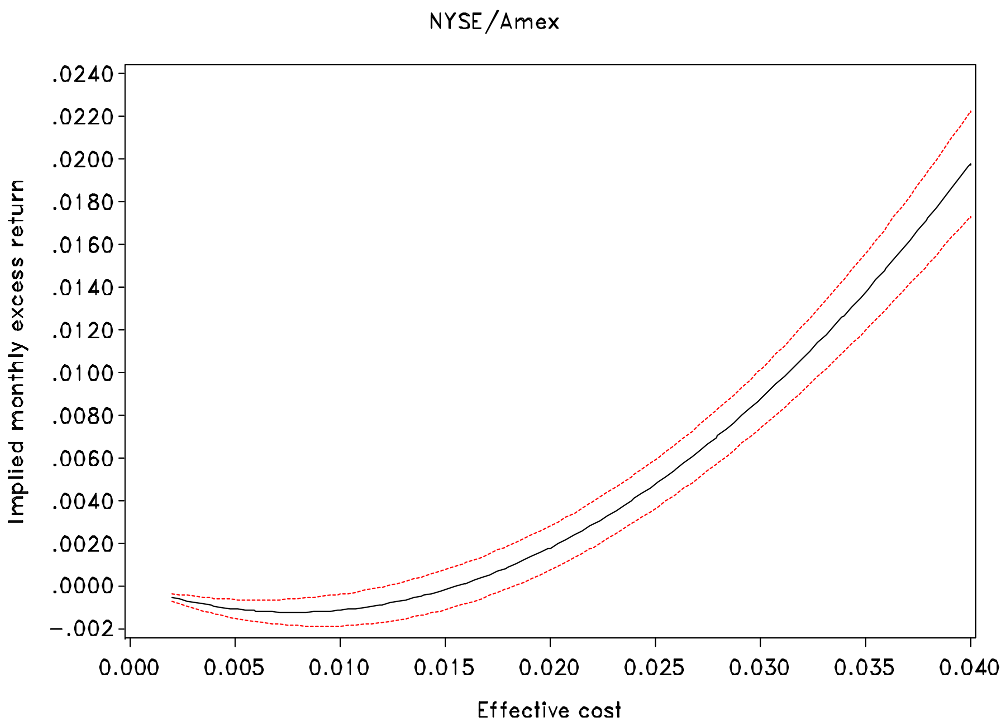


Figure 4. Relationship between effective cost and excess return implied by parametric specification. (Continued)