A Model of Instantaneous Price Impact
and Implied True Price∗

Christopher Ting† and K. C. Tsui

School of Business, Singapore Management University, Singapore 259756

ABSTRACT

We show that the S-shaped hyperbolic tangent function of signed volume is appropriate to model the price impact of a trade. This model enables an implied true price to be obtained without relying on the quotes. We compare this implied true price with the quotes’ midpoint. For the 1,748 common stocks traded on the NYSE in 1997, we find that the implied true price is superior than the midpoint in proxying for the unobservable true price. The difference between these two proxies is statistically and economically significant. In addition, there is evidence that the price impact function changes over time and relates to the information environment of the company. In light of our trade-based approach to estimating transaction costs, the quote-based effective spread is found to have an upward bias. In sum, this article reveals that both the midpoint and the effective spread are not as reliable as most researchers have implicitly predicated.

∗We would like to thank Liuren Wu for comments on the initial draft. We are also indebted to Joel Hasbrouck for his insights and suggestions for improvement. A partial funding from the Wharton-SMU Research Center is gratefully acknowledged.

†Corresponding author. Email: christopherting@smu.edu.sg.
The true price is an important quantity in financial economics. Since it is not directly observable, the midpoint of the prevailing bid and ask quotes is often used as its proxy. Among others, Hasbrouck (1991), Huang and Stoll (1997), Engle and Russell (1998), Engle (2000) as well as Dufour and Engle (2000) employ the midpoint to study the dynamics of price impact and trade duration. As many empirical studies and the estimation of effective spread rely on the midpoint, it is important to ascertain how well this quantity proxies for the true price.

This paper attempts to estimate the true price and transaction costs without using the prevailing quotes. The motivation for using only trade records is threefold. First, it is important to examine the fidelity of the quote-based midpoint in representing the unobservable true price. This examination is necessary because at any given time, so long as one of the two quotes is stale, the midpoint and the transaction cost estimation will be unreliable. Second, if a trade-based proxy is found and proven to be superior, then the midpoint has little additional information about the true price that is not already in the better proxy. Finally, there are important cases such as the Chicago Mercantile Exchange, Eurex and Euronext where only the historical trades but not the quotes are recorded. In these instances, a purely trade-based approach is the only way forward.

We model the instantaneous price impact of a trade by expressing the transaction price in terms of the true price and a price impact function. Our mathematical and empirical analyses show that the instantaneous price impact that is contemporaneous with the true price is well captured by the hyperbolic tangent function of signed volume. This functional form is consistent with the economic theory of Bhattacharya and Spiegel (1991) and Spiegel and Subrahmanyam (2000). Serving both as a trade sign indicator as well as a differentiator between large and small trades, the hyperbolic tangent function is special in that the price impact is bounded for all trade sizes.

The sample comprises 1,748 stocks traded on the NYSE in 1997. We pay due attention to the bid-ask bounce and price discreteness by removing the first-order serial correlation in the trade-to-trade price change prior to estimation. Once the parameters of the model are estimated, the implied true price is obtained by discounting the instantaneous price impact from the transaction price. A relative comparison reveals that the implied true price is closer to the unobservable true
price than the midpoint. Moreover, the difference between these two proxies is found to be both statistically and economically significant. It appears that the midpoint is not as good a proxy as one might have hitherto assumed.

This article also studies the shape of the price impact curve over time. Our formulation has a scaling parameter that determines the shape of the curve. With increasing value of this parameter, the shape of the hyperbolic tangent function ranges from effectively a straight line to an S-shaped curve and to the Heaviside function. When the scaling parameter is small, the price impact of a large trade is much larger than a small trade. The market is deemed to be illiquid because a block trade cannot be executed without creating a larger price impact. This lack of liquidity may be attributable to the information asymmetry among traders. We find that the shape of the price impact function varies over time, though we do not find evidence that suggests a close association with announcements of material information.

However, there is a correspondence between the information environment and the scaling parameter. Along with the volatility of price change, we use the number of shareholders and the number of analysts following a company as explanatory variables. These variables are motivated by Spiegel and Subrahmanyam (2000). Our cross-sectional regression results are in line with their prediction. For stocks that are followed by many analysts and are widely held, asymmetric information is likely to be less of a problem.

To demonstrate the usefulness of our approach, we define the difference between the transaction price and the implied true price as a measure of transaction costs. Consistent with existing literature, we find that large market capitalization stocks traded on the NYSE have smaller transaction costs. In addition, the costs are 38.9% lower on average after the tick size reduction in 1997, which is anticipated by Harris (1994).

This paper is organized as follows. The next section presents the model as well as insights gained from comparing our formulation with related econometric models. The strong assumption of buyer- and seller-initiated trades being equally likely is relaxed in Section II. In Section III, we describe our sample and the data used for estimation. Section IV documents the estimation results and Section V compares our trade-based statistics with the corresponding quote-based
numbers. We conclude our paper in Section VI.

I. Model of Instantaneous Price Impact and Econometric Implications

In this section, we present a model of instantaneous price impact that allows the implied true price to be obtained from transaction data only. The implied true price is an alternative to the midpoint of the prevailing quotes in estimating the real true price. With the implied true price, we define the friction spread as a measure of transaction costs that complements the quote-based effective spread.

Assuming that the transaction price has a role in price discovery, and that buyer- and seller-initiated trades are equally likely, this paper argues that the S-shaped hyperbolic tangent function is a germane model to capture the instantaneous price impact of a trade. Also presented is an econometric specification that allows this paper’s approach be compared with related models. The insights gained from constructing a model of symmetric price impact function are applied in the next section when the strong assumption of symmetrically distributed trade sign is relaxed.

A. Transaction Price and True Price

The key components of a trade are the time of transaction $t$, the transaction price, the volume and the sign indicating whether the trade is a purchase or sale. In an exchange where there are designated market makers, the signed volume $x$ is positive when investors buy from them at the ask price and negative when investors sell to them at the bid price. In a fully automated exchange, investors who use market orders for immediate transactions are said to be the trade initiators. Here, one defines analogously a signed volume $x$ as positive for a buy market order and negative for a sell market order.

As highlighted by O’Hara (1995), transaction price depends on trade size. Thus, we consider the transaction price $P(t, x)$ as a stochastic process with respect to time $t$ and signed volume $x$. 

4
At any given time $t$, two different signed volumes $x_a$ and $x_b$ will have different transaction prices. Namely,

$$P(t, x_a) < P(t, x_b) \quad \text{whenever} \quad x_a < x_b.$$  \hfill (1)

Equation (1) describes the circumstances in which investors sell at a price lower than the price at which they buy. Investors also face a lower (higher) transaction price if they sell (buy) a larger amount of shares. These considerations motivate a parametric formulation for the transaction price as

$$P(t, x) = S(t) \exp(\varphi(x)),$$  \hfill (2)

where $\varphi(x)$ is an increasing function of signed volume $x$. Intuitively, the transaction price $P(t, x)$ is decomposed into two parts, the unobservable true price denoted by $S(t)$ and a price impact function $\varphi(x)$ that captures the friction in the trading process.

Expressed in logarithmic levels, we obtain

$$\ln P(t, x) = \ln S(t) + \varphi(x),$$  \hfill (3)

which appears in Hasbrouck (1993) where $\varphi(x)$ corresponds to his pricing error. The price impact function $\varphi(x)$ must be an odd function with $\varphi(-x) = -\varphi(x)$ so that a sell (buy) order is transacted at a price lower (higher) than the true price. More importantly, $\varphi(x)$ must be bounded when $x \to \pm\infty$. Otherwise, the transaction price will deviate too significantly from the true price, which is at odds with financial economics and market practice.

In our framework, the true value $S(t)$ of the asset evolves according to some stochastic process even in the absence of trades. This is because when the market is closed or when trading halts occur, information production, gathering and processing activities are still ongoing. All these activities are likely to result in a true price $S(t)$ that is different from that of the last trade. Since the true price $S(t)$ is not directly dependent on $x$ and other intertemporal trade and quote variables, our framework captures only the transitory price impact of a trade.
B. Functional Forms and Parameters of Instantaneous Price Impact

The true price cannot be measured precisely because in order to observe it at a particular instance \( t \), one has to trade an amount \( |x| \), which perturbs the true price \( S(t) \) to \( P(t, x) \). But, if the price impact function is ascertained, we obtain

\[
S(t) = P(t, x) \exp\left(-\varphi(x)\right).
\]  
(4)

The true price thus inferred from the observed transaction price and signed volume is called the implied true price.

B.1. Symmetric Functional Form

What is the functional form for \( \varphi(x) \) that would fit the data? Along this direction, Hausman, Lo, and MacKinlay (1992) as well as Hasbrouck (1993) empirically find the price impact to be concave with respect to the volume \( |x| \). From a mathematical standpoint, however, when the defining parameter is given and fixed, equation (3) makes it evident that any concave function of the form \( \text{sign}(x) |x|^k \) with exponent \( k \) drives the transaction price close to zero if a very large sale occurs. This family of polynomial functions is not bounded for large \( |x| \). To be consistent with the observation that the transaction price \( P(t, x) \) is strictly positive even if an investor sells a large number of shares, one has to consider a function that is asymptotically bounded with respect to the trade size.

**Proposition 1.** Let \( \alpha \) be a non-negative parameter, and \( \beta \) a strictly positive scaling parameter. If each transaction with signed volume \( x \) and price \( P(t, x) \) contributes to the discovery of the true price \( S(t) \), and if buyer- and seller-initiated trades are equally likely to occur, then a suitable functional form for the instantaneous price impact is \( \alpha \tanh(\beta x)/2 \).

**Proof.** See Appendix A.

It is interesting that two assumptions commonly employed in the finance literature lead to the S-shaped hyperbolic tangent function. This function is strictly increasing and asymptotically
bounded. The assumption that the transaction price \( P(t, x) \) plays a role in discovering the unobservable true price \( S(t) \) is one of the core tenets of financial economics. The other premise of treating the trade sign as a fair coin is a common practice in the market microstructure literature. A noteworthy aspect of this proposition is that it is independent of the exchange design such as whether there are designated market makers.

Therefore, we consider the hyperbolic tangent\(^1\) as the price impact function with \( \alpha \) and \( \beta \) as the parameters:

\[
\varphi(x) = \frac{\alpha}{2} \tanh(\beta x).
\]

The volume \( |x| \) measured in lots\(^2\) is unique up to a scaling parameter \( \beta \). This scaling parameter allows us to examine how the transaction price \( P(t, x) = S(t) \exp\left(\frac{\alpha \tanh(\beta x)}{2}\right) \) approaches \( S(t)e^{\pm\alpha/2} \) when \( |x| \) becomes large.

An economic justification for equation (5) is found in Bhattacharya and Spiegel (1991). Their theory suggests that feasible aggregate demand curves are generally S-shaped in the classical Walrasian setting in which noisy rational expectations equilibria are obtained. Spiegel and Subrahmanyam (2000) confirm that pricing functions are continuous, differentiable and nonlinear. Their theoretical results and numerical simulations with various probability density functions for the variance of private information suggest an S-shaped pricing function. A key feature of their function is that it is bounded even as the volume \( |x| \) becomes asymptotically large. Their supply curve corresponds to our \( P(t, x) \), which is convex for \( x < 0 \) and concave for \( x > 0 \) with the inflection at the origin.

**B.2. Economic Interpretation of Scaling Parameter \( \beta \)**

To illustrate the effects of different values of \( \beta \), we simulate a family of \( P(t, x) \) and plot the curves in Figure 1. The true price is assumed to be \( S(t) = 10 \) dollars at a given time. The price impact parameter \( \alpha \) is fixed at a large value of 10%. Five different values for \( \beta \) are used in

---

\(^1\)In a different setting, Wang (1994) obtains an asset pricing theorem that the expected future excess return conditional on current volume and dividend change is related to the hyperbolic tangent function of these two quantities.

\(^2\)Throughout the paper, we use the board lot as the unit of trade size. This is in line with the market practice that the board lot is the unit used in trading rather than the number of shares.
the illustration. If $\beta$ is large, say $1 \times 10^6$, then $\tanh(\beta x)$ is economically no difference from the functional form $H(x) - H(-x)$ where $H(x)$ is the Heaviside function: $H(x) = 1$ if $x > 0$ and zero otherwise. In this case of large $\beta$, the transaction price is independent of the trade size. Only the trade sign matters. Even if $\beta = 1$, the difference in the price impacts from large and small trades is marginal and economically insignificant.

On the other hand, when $\beta = 0.5$, the transaction prices for one lot and five lots differ by about 27 cents in this illustration. The market is not ready to entertain larger trades without incurring a larger price impact. In this situation, the security is deemed to be less liquid. The other case of $\beta = 0.1$ transforms the nonlinear functional form effectively to a linear function of $x$, i.e. $P(t, x) \approx S(t)(1 + \alpha \beta x/2)$. This is because by Taylor’s expansion, $\tanh(\beta x) \approx \beta x$ when $\beta|x|$ is small, and $\exp(\alpha \beta x/2) \approx 1 + \alpha \beta x/2$. Nonetheless, for very large trades, the price impact is still bounded even for small values of $\beta$.

Therefore, one could broadly consider different values of $\beta$ as different functional forms. The pair of parameters, $\{\alpha, \beta\}$, defines the two asymptotic values of $P(t, x)$ and governs the manner in which these values are approached as $|x|$ becomes large.

In Spiegel and Subrahmanyam (2000), the Heaviside function corresponds to the linear equilibrium under the assumption that all traders know the variance of private information is at its maximum. This variance is a random variable in their setting. Conditional on the market-wide knowledge of the variance of private information and that an equilibrium with trade exists, a larger variance implies that insiders themselves disagree in their valuations of the private information. As a result, the privately informed traders have less of an advantage over the uninformed. In this instance where adverse selection is not an issue, the price impact is less dependent on the trade size. Conversely, when even the variance of private information is uncertain, the degree of asymmetry is larger. It follows that the price impact is larger for larger trade size. In light of their insights, $\beta$ should be negatively related to the extent of information asymmetry.

Before discussing the economic interpretation of $\alpha$, two other remarks are in order. Firstly, since the proposition of this paper is obtained irrespective of whether market makers are des-
ignated, the model should be applicable to both fully automated exchanges as well as markets where specialists or dealers are officially appointed intermediaries. Secondly, the price impact is instantaneous and contemporaneous with the transaction. There is a one-to-one correspondence between the trade and the transitory price impact at the moment of transaction. In contrast, other papers consider the permanent price impact over an arbitrarily chosen unit of time or number of trades. However, such price impact is dependent on the time interval chosen. Our model of instantaneous price impact does not have this problem.

B.3. Economic Interpretation of Price Impact Parameter $\alpha$

An important feature of $\alpha \tanh(\beta x)/2$ is that both the price impact parameter $\alpha$ and the scaling parameter $\beta$ are positive. This is required by the upward sloping nature of our postulate in equation (1). Also, the $\alpha$ estimate should be of the right magnitude to ensure that the transaction price is not too different from the true price. The reason for this requirement is as follows. When the trade size $|x|$ is large, and as $\tanh(\beta x)$ asymptotes to $\pm 1$, the following inequalities provide the bounds:

$$S(t) e^{-\alpha / 2} < P(t, x) < S(t) e^{\alpha / 2}. \quad (6)$$

If the price impact parameter $\alpha$ is small, then the bounds are tight, implying that investors are able to trade at a price near the true price $S(t)$. The smaller $\alpha$ is, the better is the market in price discovery.

To provide further economic interpretation of the price impact parameter $\alpha$, consider the Taylor series representation of $e^{\alpha \tanh(\beta x)/2}$:

$$\exp\left(\frac{\alpha}{2} \tanh(\beta x)\right) = 1 + \frac{\alpha}{2} \tanh(\beta x) + \frac{\alpha^2}{8} \tanh^2(\beta x) + \cdots. \quad (7)$$

It is noteworthy that when the trade size $|x|$ is large,

$$\tanh(\beta x) \approx \text{sign}(x). \quad (8)$$
It follows that if \( \alpha \) is small and the volume \( |x| \) is large, \( e^{\alpha \tanh(\beta x)/2} \) is well approximated by

\[
\exp\left(\frac{\alpha}{2} \tanh(\beta x)\right) \approx 1 + \text{sign}(x) \frac{\alpha}{2},
\]

(9)
as the marginal price impact declines at the exponential rate. Thus we obtain from equation (2),

\[
P(t, x) \approx S(t) + \text{sign}(x) \times S(t) \frac{\alpha}{2}.
\]

(10)

In the market microstructure literature, the transaction price is often written as the sum of the true price and the transaction cost per unit,

\[
P(t, x) = S(t) + \text{sign}(x) \times \text{Transaction cost}.
\]

(11)

For example, with the premise expressed in equation (11), Roll (1984) shows that the autocovariance of returns computed using transaction prices is negative as a result of the transaction cost. The bias in Roll’s estimate is investigated by Harris (1990b), who suspects the discrete nature of the price grid is the most likely culprit. Hasbrouck (1993) uses equation (11) to study the pricing error and finds the transaction costs to be 0.26% of the NYSE stock prices in 1989. In light of these papers, and comparing equation (10) with equation (11), it is clear that \( S(t)\alpha/2 \) corresponds to the transaction cost. Consequently, the price impact parameter \( \alpha \) is interpreted as the percentage round-trip transaction cost for large trades.

### C. Trade-Based Transaction Costs

The importance and implications of transaction costs can be discerned from a diverse range of research papers and reports. For instance, Stoll and Whaley (1983) find that expected returns are related to transaction costs. In asset pricing, Bensaid, Lesne, Pagès, and Scheinkman (1992) show that transaction costs make the market incomplete. Admati and Pfleiderer (1991) argue that fund managers are concerned with transaction costs and the price impacts of their trades. Keim and Madhavan (1997) find transaction costs vary with investment style, order submission strategy as well as exchanges. The SEC uses the quote-based effective spread as one of the
measures to compare transaction costs across market centers (see SEC’s report (2001)).

If the implied true price \( S(t) \) can be inferred from the data, one defines the transaction price as

\[
F(t, x) = 2|P(t, x) - S(t)|. \tag{12}
\]

For ease of exposition, we refer to \( F(t, x) \) as the friction spread. It is analogous to the effective spread. Instead of the quotes’ midpoint, the true price is used as a reference.

When \( \alpha \) and \( \beta \) are estimated, one obtains an implied true price:

\[
S(t) = P(t, x) \exp\left(-\frac{\alpha}{2} \tanh(\beta x)\right). \tag{13}
\]

As \( x \) and \( P(t, x) \) are in principle observable, the implied true price \( S(t) \) is computed readily using this formula. From equations (12) and (13), the friction spread is then obtained as follows:

\[
F(t, x) = 2P(t, x) \left|1 - \exp\left(-\frac{\alpha}{2} \tanh(\beta x)\right)\right|. \tag{14}
\]

Equation (14) offers a new avenue to measure trading friction, which is different but complementary to existing quote-based measures such as the quoted and effective spreads. It is noteworthy that the friction spread is contemporaneous\(^3\) in that the true price is updated whenever a transaction takes place.

As a summary, we see from equation (14) that \( \alpha \) is approximately the friction spread in percent for large trades. Substituting in equations (7) and (8), equation (14) becomes

\[
\frac{F(t, x)}{P(t, x)} \approx \alpha \left| \tanh(\beta x) \right| \approx \alpha. \tag{15}
\]

The definition of the friction spread is therefore consistent with the economical interpretation of \( \alpha \) as percentage transaction cost in the previous subsection.

---

\( ^3 \)In contrast, there is usually a time lag between a quote update and the actual transaction.
D. Comparison with other Approaches

In this subsection, we first discuss the econometric specification. This is to facilitate the comparison with related models. Subsequently in Section II, the specification is generalized with the relaxation of a strong assumption underlying the arguments for the symmetrical price impact function.

To estimate the price impact parameter \( \alpha \), we follow the standard practice in the literature by assuming that the true price \( S(t) \) follows a geometric Brownian motion. Let the drift rate of the true price in logarithmic levels be \( \mu \) and the diffusion coefficient be \( \sigma \). The geometric Brownian motion is stochastically driven by the Wiener process \( W(t) \) with \( W(0) = 0 \) as follows:

\[
\ln S(t) = \ln S(0) + \mu t + \sigma W(t). \tag{16}
\]

Suppose \( N \) trades have been observed for a given security. We let \( t_i \) be the intra-day time and denote the signed volume at time \( t_i \) as \( x_i \), and the transaction price as \( P(t_i, x_i) \), for \( i = 1, 2, \ldots, N \). Although \( S(0) \) is unobservable, this term can be eliminated by considering two consecutive intra-day transactions. The discrete-time version of equation (3) is \( \ln P(t_i, x_i) = \ln S(t_i) + \varphi(x_i) \) and that of equation (16) is \( \ln S(t_i) = \ln S(0) + \mu t_i + \sigma W(t_i) \). The logarithmic transaction price change is then given by

\[
\ln \left( \frac{P(t_i, x_i)}{P(t_{i-1}, x_{i-1})} \right) = \varphi(x_i) - \varphi(x_{i-1}) + \mu (t_i - t_{i-1}) + \sigma \epsilon_{t_i, t_{i-1}}. \tag{17}
\]

The Brownian increment \( \epsilon_{t_i, t_{i-1}} \equiv W(t_i) - W(t_{i-1}) = \sqrt{t_i - t_{i-1}} \psi_i \), where \( \psi_i \) is a normally distributed random variable with mean 0 and variance 1.

For notational simplicity, we write the square root of the trade duration as

\[
\tau_i \equiv \sqrt{t_i - t_{i-1}} \tag{18}
\]
and define the trade-to-trade price change as

\[ r_i \equiv \frac{\ln P(t_i, x_i) - \ln P(t_{i-1}, x_{i-1})}{\tau_i}. \]  

(19)

It follows that a specific model for \( r_i \) is

\[ r_i = \frac{\alpha}{2} \left( \frac{\tanh(\beta x_i) - \tanh(\beta x_{i-1})}{\tau_i} \right) + \mu \tau_i + \sigma \psi_i. \]  

(20)

Equation (20) is obtained after setting \( \varphi(x_i) = \tanh(\beta x_i) \) and dividing both sides of equation (17) by \( \tau_i \). The motivation for the division is to disengage the stochastic fluctuation in equation (17) from the temporal clustering of trade arrivals. We see that the trade-to-trade price change \( r_i \) is dependent on the observable \( \tau_i \) and the change in price impact, which is \( \frac{\tanh(\beta x_i) - \tanh(\beta x_{i-1})}{\tau_i} \).

At this juncture, a few remarks on the linkages between our specification with previous research papers are in order. Glosten and Harris (1988) study a transaction price change process. In our notations, the main model they estimate is

\[ r_i \tau_i = \frac{c}{2} (\text{sign}(x_i) - \text{sign}(x_{i-1})) + a x_i + f_i, \]  

(21)

where \( c \) is the transitory spread component associated with non-information related costs such as order processing costs. The last term \( f_i \) is the residual. In their framework, the parameter \( a \) is the adverse selection coefficient. Huang and Stoll (1997) develop a similar price change process. In our notations, equation (5) in their paper is written as

\[ r_i \tau_i = \frac{c}{2} (\text{sign}(x_i) - \text{sign}(x_{i-1})) + \lambda \text{sign}(x_{i-1}) + g_i. \]  

(22)

The parameter \( \lambda \) in this model of Huang and Stoll (1997) is the sum of adverse selection and inventory holding components. The residual is denoted by \( g_i \). An econometric model similar to equation (21) is also developed and estimated by Madhavan, Richardson, and Roomans (1997).

Interestingly, the first terms in equations (21) and (22) coincide with our \( \frac{\alpha}{2} \left( \frac{\tanh(\beta x_i) - \tanh(\beta x_{i-1})}{\tau_i} \right) \) when the trade size \( |x| \) is large. Viewed in light of these three models, we draw an
analogy by considering the price impact parameter $\alpha$ as the transitory component of the transaction costs. In other words, our $\alpha$ partakes in the same semantics of transitory price impact with their $c$. This analogy is consistent with the analysis presented earlier in Subsection B.3. Moreover, when $\beta$ becomes infinite,

$$\lim_{\beta \to \infty} \tanh(\beta x) = \text{sign}(x). \quad (23)$$

With the functional form a priori assumed to be $\text{sign}(x)$, and as far as the transitory price impact is concerned, these papers are special cases in our framework.

II. Asymmetry in Price Impact and Estimation Methodology

Thus far, like many other authors, we have employed the assumption that buyer- and seller-initiated trades are equally likely. This is a strong assumption hardly fulfilled in the market. In this section, we relax the symmetric assumption and derive a more generic price impact function that accounts for the asymmetric distribution of trade signs.

A. Asymmetric Functional Form

The asymmetry is characterized by a parameter $\kappa$ such that $(1 + \kappa)/2$ is the probability of a purchase and $(1 - \kappa)/2$ a sale. To obtain a functional form that is consistent with this probability measure, we employ the following guiding principles, which are based upon the economics of financial markets, as well as the insights gained from Section I.

- Both the transaction price $P(t, x)$ and the true price $S(t)$ are bounded.
- Although it is no longer zero due to asymmetry, $E[\ln (P(t, x)/S(t))]$ should still be small relative to the bid-ask spread in percent.
- The asymmetric functional form must satisfy a condition of consistency described in Proposition 2 below.
PROPOSITION 2. Let \( \tilde{\alpha}_b \) and \( \tilde{\alpha}_s \) be two generically different non-negative parameters. In addition, let \( \tilde{\beta} \) be a strictly positive scaling parameter. Then, a functional form that is congruent with the asymmetric probability measure is

\[
\varphi(x) = \begin{cases} 
\tilde{\alpha}_b \tanh(\tilde{\beta}x) & \text{if } x > 0; \\
\tilde{\alpha}_s \tanh(\tilde{\beta}x) & \text{if } x < 0,
\end{cases}
\]  

(24)

provided the following consistency condition is fulfilled:

\[
\kappa = \frac{2\eta - \tilde{\alpha}_b J_b - \tilde{\alpha}_s J_s}{\tilde{\alpha}_b J_b - \tilde{\alpha}_s J_s}.
\]  

(25)

Here, \( \eta \) is the expected value of \( \ln(P(t,x)/S(t)) \). The two quantities \( J_\pm \) are, respectively, the average of \( \tanh(\tilde{\beta}x) \) conditional on \( x > 0 \) and that of \( \tanh(\tilde{\beta}x) \) conditional on \( x < 0 \).

PROOF. See Appendix B for a proof of equation (24) and the derivation of condition (25).

Equation (24) is a generalization of the symmetric functional form. When \( \kappa = 0 \), it reduces to the special case enunciated in Proposition 1. The asymmetrical S-shaped curves are now parameterized by three parameters, \( \tilde{\alpha}_b, \tilde{\alpha}_s \) and \( \tilde{\beta} \). The price impact function is still S-shaped. But, it is no longer symmetric as in Figure 1 because the two asymptotic values of \( S(t) \exp(\tilde{\alpha}_b) \) for buy and \( S(t) \exp(-\tilde{\alpha}_s) \) for sell are different. Nevertheless, a common\(^4\) scaling parameter \( \tilde{\beta} \) governs the manner in which these two values are attained as the trade size \( |x| \) increases. Moreover, the origin is still the inflection point of the curve, and the function is still continuous and differentiable for all \( x \).

To gain some insight on condition (25), we first consider the special case where there is no asymmetry. In this symmetric case, \( \kappa = 0, \tilde{\alpha}_b = \tilde{\alpha}_s \) and \( J_s = -J_b \). Substituting these relations into equation (25), we obtain \( \eta = 0 \), which is consistent with the notion that both prices of buyer- and seller-initiated trades are symmetrically distributed with respect to the true price. On the

\(^4\)One may imagine that a more general version is to have different scaling parameters for buy and sell. However, two different scaling parameters result in a discontinuity at the origin and the annihilation of its property as the inflection point. The requirement for a common \( \tilde{\beta} \) in our derivation of the functional form in Appendix B is consistent with the theoretical paper by Spiegel and Subrahmanyam (2000). They prove that if an equilibrium with trade exists, the price function is necessarily continuous with at least one inflection point.
other hand, when $\kappa$ is non-zero, the price discovery is asymmetric, as one of the two transaction prices is closer to the true price than the other. The bounds corresponding to equation (6) are

$$S(t) e^{-\tilde{\alpha}_s} < P(t, x) < S(t) e^{\tilde{\alpha}_b}.$$ 

(26)

Equivalently

$$-\tilde{\alpha}_s < \ln \left( \frac{P(t, x)}{S(t)} \right) < \tilde{\alpha}_b.$$ 

(27)

Therefore, when $\tilde{\alpha}_b \neq \tilde{\alpha}_s$, the buyer-initiated price is not equi-distant from the true price as the seller-initiated price. As a result, the transaction cost in percent for a buyer-initiated trade is different from a seller-initiated transaction.

**B. Econometrics of Implied True Price and Friction Spread**

In view of the asymmetry $\kappa$, the implied true price $S(t)$ in equation (13) and the friction spread $F(t, x)$ in equation (14) become, respectively,

\[
\tilde{S}(t) = P(t, x) \exp \left( -\tilde{\alpha}_b \tanh(\tilde{\beta} x)D_{x>0} - \tilde{\alpha}_s \tanh(\tilde{\beta} x)D_{x<0} \right),
\]

(28)

\[
\tilde{F}(t, x) = 2P(t, x) \left| 1 - \exp \left( -\tilde{\alpha}_b \tanh(\tilde{\beta} x)D_{x>0} - \tilde{\alpha}_s \tanh(\tilde{\beta} x)D_{x<0} \right) \right|,
\]

(29)

where $D_{x>0}$ is the dummy variable that equals one if $x > 0$ is one and zero otherwise. The other dummy variable $D_{x<0}$ is defined analogously.

The implied true price $\tilde{S}(t)$ and the friction spread $\tilde{F}(t, x)$ are the main quantities we estimate in this paper. These two quantities are obtainable from the transaction price $P(t, x)$ and the signed volume $x$ when the three parameters, $\tilde{\alpha}_b, \tilde{\alpha}_s$ and $\tilde{\beta}$ are available.

To estimate the parameters from tick-by-tick data, we need to pay due attention to the microstructure effects. It is well known that high-frequency price change $r_i$ as defined in equation (19) is negatively correlated due to the bid-ask bounce (Roll (1984), Harris (1990b)) and

16
price discreteness (Harris (1990a)). To control for these microstructure effects, we consider the AR(1)-adjusted price change as the residual $y_i$ of the following regression:

$$r_i = \rho_0 + \rho_1 r_{i-1} + y_i.$$  \hspace{1cm} (30)

Here, $\rho_0$ is the intercept and $\rho_1$ the autocorrelation coefficient at one lag. With $y_i$ as the dependent variable, and in view of the asymmetry, the multivariate regression model we estimate empirically is

$$y_i = \tilde{y}_0 + \left( \tilde{\alpha}_b \tanh(\tilde{\beta}x_i)D_{x_i>0} + \tilde{\alpha}_s \tanh(\tilde{\beta}x_i)D_{x_i<0} \\ - \tilde{\alpha}_b \tanh(\tilde{\beta}x_{i-1})D_{x_{i-1}>0} - \tilde{\alpha}_s \tanh(\tilde{\beta}x_{i-1})D_{x_{i-1}<0} \right) / \tau_i \\ + \tilde{\mu} \tau_i + \tilde{e}_i ,$$ \hspace{1cm} (31)

where $\tilde{y}_0$ is the intercept and $\tilde{e}_i$ the error term.

This specification is a generalized version of equation (20). To support this claim, suppose the asymmetry is absent. In other words, $\kappa = 0$ and there is only one parameter, $\alpha / 2$. The price impact is symmetric with $\tilde{\alpha}_b = \alpha / 2 = \tilde{\alpha}_s$ and $\tilde{\beta}$ is replaced by $\beta$. In this special case, the four terms in the parentheses become

$$\frac{\alpha}{2} \tanh(\beta x_i) (D_{x_i>0} + D_{x_i<0}) - \frac{\alpha}{2} \tanh(\beta x_{i-1}) (D_{x_{i-1}>0} + D_{x_{i-1}<0}).$$  \hspace{1cm} (32)

Since buyer- and seller-initiated trades are mutually exclusive, which is echoed by $D_{x>0} + D_{x<0} = 1$, the above expression reduces to

$$\frac{\alpha}{2} \left( \tanh(\beta x_i) - \tanh(\beta x_{i-1}) \right).$$ \hspace{1cm} (33)

The associated specification is then

$$y_i = y_0 + \frac{\alpha}{2} \left( \frac{\tanh(\beta x_i) - \tanh(\beta x_{i-1})}{\tau_i} \right) + \mu \tau_i + e_i ,$$ \hspace{1cm} (34)

which is essentially equation (20), a special case discussed in Section I.D.
C. Asymmetric and Symmetric Parameter Estimates

What is the relation between the symmetric $\alpha$ parameter and $\tilde{\alpha}_b + \tilde{\alpha}_s$? We note that buyer-initiated purchases and seller-initiated sales are mutually exclusive events. If the price impact is constrained to be modeled by only one parameter $\alpha$ as in the symmetric case, then when the constraint is removed, $\alpha$ will be split into $\tilde{\alpha}_b$ and $\tilde{\alpha}_s$ to account for these two mutually exclusive events. Worded differently, when the transaction cost is broken down into two separate accounts for buy and sell, the sum of the two should more or less tally with the total value in the account before the breakdown. Otherwise, it implies that the two events somehow interact at any given time $t$ to reinforce or decimate each other. Clearly, this is at odds with buy and sell being mutually exclusive.

**Proposition 3.** When the magnitude of the asymmetry $|\kappa|$ is small, the symmetric price impact parameter $\alpha$ is approximately the sum of two asymmetric price impact parameters. More specifically, the relations between the parameters in equations (34) and (31) are as follows:

\begin{align}
  y_0 & \approx \tilde{y}_0; \quad \mu \approx \tilde{\mu}; \\
  \beta & \approx \tilde{\beta}; \quad \alpha \approx \tilde{\alpha}_b + \tilde{\alpha}_s.
\end{align}

**Proof.** See Appendix C.

Proposition 3 is important in justifying that the symmetric assumption is not that bad after all. If this proposition were not obtained, earlier papers such as Roll (1984), Glosten and Harris (1988), Hasbrouck (1993), Madhavan, Richardson, and Roomans (1997), Huang and Stoll (1997) and many others that assume the price impact is symmetric would be wrong. It is noteworthy that their observations are pooled over a long period of a year or so. Such pooling is likely to result in a small asymmetry between buy and sell. Namely, $|\kappa| \approx 0$ and Proposition 3 is applicable, which implies that these earlier papers are not off the mark in their estimations of transitory transaction costs despite assuming that a purchase has the same price impact as a sale.
D. Comparison of Quotes’ Midpoint and Implied True Price

The importance of midpoint is widely recognized in the literature. Among many others, Hasbrouck (1991) and Dufour and Engle (2000) employ the midpoint to study the dynamics of quote revisions and trades. It is therefore of interest to compare the midpoint with the implied true price for every trade, which is readily computed from equations (28). The difference between these two prices may provide valuable insight into the linkage between quote revisions and implied true price movements. The comparison is particularly interesting because the implied true price uses only trade data whereas the midpoint is based purely on quotes.

To compare the quotes’ midpoint and the implied true price in terms of their relative proximity to the real true price, we use the following method. Suppose \( z(t_i) \) is the real true price in logarithmic levels. The index \( i \) is the intra-day transaction number. If the candidate true price in logarithmic levels is denoted by \( c(t_i) \), then it is reasonable to write

\[
  c(t_i) = z(t_i) + \zeta(t_i),
\]

where \( \zeta(t_i) \) is the deviation from the real true price.

Owing to the discrete price grid and trade size, the second-order moment of the price change estimated with the midpoint or the implied true price is larger than that estimated with the real true price. Thus, we consider the trade-to-trade price change:

\[
  u_i = c(t_i) - c(t_{i-1}) = v_i + w_i,
\]

where the change processes for the real true price and deviation are given by \( v_i = z(t_i) - z(t_{i-1}) \) and \( w_i = \zeta(t_i) - \zeta(t_{i-1}) \), respectively. The second-order moment is then obtained as follows:

\[
  E[u_i^2] = E[v_i^2] + E[w_i^2] + 2E[v_i w_i].
\]

We assume that the real true price change \( v_i \) is independent of the change in deviation, \( w_i \). No further assumption is made concerning the process \( w_i \). Now, \( E[v_i w_i] = E[v_i]E[w_i] = 0 \), as
\[ E[v_i] = 0, \] which is attributed to the real true price being a martingale. If the order of magnitude of the deviation is small, then its second-order moment \( E[w_i^2] \) will also be small. Since we are only performing a relative comparison of the implied true price and the quotes’ midpoint, \( E[v_i^2] \) is irrelevant. Therefore, between the two proxies, whichever that has a smaller second-order moment is deemed to be relatively closer to the real true price \( z(t_i) \).

To ascertain whether there is a significant difference between these two proxies, we consider the spread \( L_i \) between the implied true price \( \tilde{S}_i \) and the midpoint \( M_i \) associated with the \( i \)-th trade:

\[ L_i = 2 |\tilde{S}_i - M_i|. \quad (40) \]

To gauge the magnitude of \( L_i \), we use the prevailing bid-ask spread \( Q_i \) as a yardstick and consider the daily average ratio from \( N \) pairs of \( L_i \) and \( Q_i \) values as follows:

\[ R_d \equiv \frac{1}{N} \sum_{i=1}^{N} \frac{L_i}{Q_i}. \quad (41) \]

The ratio \( R_d \) is the daily average difference as a percentage of the bid-ask spread. Obviously, a small \( R_d \) indicates that the implied true price and the midpoint are not different.

When \( R_d \) is larger than 50%, it means that the implied true price is closer to the transaction price. To understand this, suppose the two price impact parameters are vanishingly small. Then the model suggests that

\[ P(t, x) \approx \tilde{S}(t)(1 + \tilde{\alpha}_b \tanh(\tilde{\beta} x)D_{x>0} + \tilde{\alpha}_s \tanh(\tilde{\beta} x)D_{x<0}) \approx \tilde{S}(t). \quad (42) \]

In other words, the implied true price \( \tilde{S}_i \) approximately equals the transaction price when \( \tilde{\alpha}_b \) and \( \tilde{\alpha}_s \) are close to zero. If transactions take place at the bid or the ask, \( L_i \) will be approximately equal to the bid-ask spread \( Q_i \), which leads to a large \( R_d \). Conversely, \( R_d \) smaller than 50% indicates that the implied true price is closer to the midpoint. In other words, both proxies for the real true price are more or less the same.
III. Data

Our sample consists of companies listed on the NYSE. The sample period is the year 1997. Like Peterson and Sirri (2003), we choose 1997 for two reasons. Firstly, on June 24, 1997, NYSE's minimum tick size was reduced from 1/8 to 1/16 of a dollar. Among others, Chordia, Roll, and Subrahmanyam (2001) report that the quoted and effective spreads are lower after the tick size reduction. If our model is correctly specified and the estimation is reliable, then the price impact parameters should exhibit a similar behavior. Secondly, on October 27, the DJIA fell 544 points or -7.2 percent and rebounded 337 points the following day. These large and unanticipated moves in the market provide a good test bed to examine the robustness of our approach.

Motivated by these considerations, we extract all the NYSE listed companies from the CRSP database. After removing two securities with zero market capitalization and four with negative stock prices, we have 1,781 left. For companies that have more than one class of shares, we remove those securities that are not Class A. We then use the CUSIP numbers and the ticker symbols to query the TAQ database. The final number of stocks that have matching descriptions in the CRSP and TAQ databases is 1,753. Like Chordia, Roll, and Subrahmanyam (2001), we remove a company whose stock price is greater than $999 to avoid the influence of unduly high-priced stocks. However, we do not discard stocks that are below a dollar. We also drop four stocks that do not have 50 observations. Thus, our sample comprises 1,748 common stocks traded primarily on the NYSE.

The trade records of these stocks are obtained from the TAQ for all 253 trading days in 1997. The descriptive statistics for our NYSE sample are given in Table I. Throughout the sample period, the cross-sectional average number of trades on the NYSE alone is 25,816. This translates to approximately 102 trades per trading day for each stock. Therefore, there is sufficient number of observations to estimate the parameters on a daily basis.

We use the tick rule to infer whether a transaction is buyer- or seller-initiated. This signing rule requires only the trade records. The tick rule signs an up-tick\(^5\) trade as buyer-initiated and

\(^5\)An up tick is defined as an upward move of the current transaction price from the previous transaction. Conversely, a down tick occurs when the current transaction price is less than the previous transaction price. A zero tick corresponds to no change in transaction prices.
a down-tick transaction as seller-initiated. Zero-tick trades are classified according to the last up and last down ticks. For trades that cannot be signed, they are excluded from the sample. These trades typically occur at the beginning of the regular hours and re-opening after the trading halts.

The motivation for using the tick rule instead of the Lee and Ready (1991) algorithm is to demonstrate that our approach is still applicable when quotes are not available. Moreover, Ellis, Michaely, and O'Hara (2000) and Finucane (2000) show that using only the tick rule results in a marginal loss of classification accuracy compared to the Lee and Ready (1991) algorithm.

For further analysis, I/B/E/S database is used to compute the standardized unexpected earnings, and to derive the number of analysts following a stock in 1997. Moreover, we tap into the Compustat to obtain the information on the number of shareholders (Annual Data Code Number 100) and whether the stock is a component of the S&P 500 (Annual Data Code Number 276).

### IV. Estimation and Analysis

In our empirical study, three estimation periods are used: Daily and over two subsample periods delineated by the change in the minimum tick size from 1/8 to 1/16 of a dollar. Daily frequency enables the day-to-day variation in the estimates to be studied as time series. The problem of infrequent trading is ameliorated to a large extent when observations are pooled. So long as there are 50 trades or more in the estimation period, a regression is performed\(^6\). Since there are 23,400 seconds per trading day, the trade duration \(t_i - t_{i-1}\) is normalized by this number. The unit we use for the signed volume \(x\) is the number of lots.

**A. Estimation and Consistency Tests**

Section II reveals three important relations that the parameter estimates must satisfy. They are re-produced in the following order to underscore their importance:

\(^6\)Beyond demonstrating that it is possible to pool the observations over a few days or months to perform estimations for illiquid stocks, no insight is gained in addition to what is already gained from the findings based on daily estimations. These subperiod estimation results are not reported in detail to curtail the length of this paper.
A. \[ \kappa = \frac{2\eta - \tilde{\alpha}_bJ_b - \tilde{\alpha}_sJ_s}{\tilde{\alpha}_bJ_b - \tilde{\alpha}_sJ_s}; \]

B. \[ |\eta| = \left| E \left[ \ln \left( \frac{P(t,x)}{S(t)} \right) \right] \right| \leq \tilde{\alpha}_b + \tilde{\alpha}_s; \]

C. \[ \tilde{y}_0 \approx y_0; \quad \tilde{\mu} \approx \mu; \quad \tilde{\beta} \approx \beta; \quad \tilde{\alpha}_b + \tilde{\alpha}_s \approx \alpha. \]

To examine whether Condition A is fulfilled, we need to estimate \( \kappa \), which measures the deviation from \( 1/2 \). Suppose \( n_b \) out of a total of \( N \) trades are buyer-initiated. The parameter \( \kappa \) on the left side of the condition is estimated by

\[ \kappa = \frac{2n_b}{N} - 1. \quad (43) \]

The quantity \( J_b \) is estimated as the average of \( \tanh(\tilde{\beta}x_i) \) with \( x_i \) being positive, and \( J_s \) as the average of \( \tanh(\tilde{\beta}x_i) \) with \( x_i \) being negative:

\[ J_b = \frac{1}{n_b} \sum_{i=1}^{N} \tanh (\tilde{\beta}x_i) D_{x_i>0} \quad (44) \]
\[ J_s = \frac{1}{n_s} \sum_{i=1}^{N} \tanh (\tilde{\beta}x_i) D_{x_i<0}, \quad (45) \]

where \( n_s = N - n_b \) is the number of seller-initiated trades.

A few words concerning the statistical significance of the price impact parameters are in order. Positive but insignificant \( \tilde{\alpha}_b \) and \( \tilde{\alpha}_s \) estimates are associated with observations that have little price variation. To understand this, suppose there is only one transaction price. Then, the left side of equation (31) is zero for all \( i \) and thus all the parameter estimates must be zero. This is consistent with the fact that if the investors trade at only one specific price, there is no price impact as the transaction price itself is the true price. In this circumstance, the only way to satisfy \( P(t,x) = S(t) \) with non-zero \( x \) is to have \( \tilde{\alpha}_b = \tilde{\alpha}_s = 0 \). Thus, on days when the number of transaction prices is only one or two, the estimates for the price impact parameters are usually not significant.
For this reason, Condition A is much more stringent than the tests for the statistical significance of the model parameters. The quantity computed based on the parameters’ estimates on the right side of Condition A must equal \( \kappa \), which is a parameter exogenous to our model\(^7\). Estimated non-parametrically and independent of our model, Condition A provides a crucial litmus test. If indeed Condition A holds, then there is strong evidence that the model is well specified and the estimation results are reliable.

To estimate the parameters of interest, we use equation (30) to first obtain the AR(1)-adjusted price change\(^8\) \( y_i \). Nonlinear least squares (NLS) regression is then performed according to the specification in equation (31). In total, there are seven parameters in equations (30) and (31). However, only three of them, \( \tilde{\alpha}_b, \tilde{\alpha}_s \) and \( \tilde{\beta} \), are crucial from the standpoint of obtaining the implied true price as well as estimating the friction spread.

A total of 190,119 NLS regressions are performed for each NYSE sample stock and each day in the sample period. Table II provides a breakdown of the statistics by decile and tick size regime. An NLS regression result is said to be valid if Condition A is satisfied (up to the double precision of the software). Remarkably, we find that 188,945 regressions or 99.38% of the total are valid. This high percentage bears testimony to the validity of the model and of the estimation. In addition, in over 85% of these valid regressions, \( \tilde{y}_0 \) and \( \bar{\mu} \) are not statistically significant whereas \( \tilde{\alpha}_b + \tilde{\alpha}_s \) and \( \beta \) are significant at the 5% level.

Condition B can be verified by obtaining the implied true price with equation (28) and the three parameter estimates: \( \tilde{\alpha}_b, \tilde{\alpha}_s \) and \( \tilde{\beta} \). As for \( \eta \), it is estimated as the average,

\[
\frac{1}{N} \sum_{i=1}^{N} \left( \ln P(t_i, x_i) - \ln \tilde{S}(t_i) \right),
\]

where \( N \) is the number of transactions. Remarkably, we find that in all the valid regressions, Condition B is satisfied.

---

\(^7\)The estimator of \( \kappa \) as in equation (43) produces a population estimate for each stock and each day. To understand this, consider a company that has \( N \) historical transactions over a period. Obviously, these \( N \) transactions are the only observations for the company and for that period. In other words, given these \( N \) observations, regardless of whether \( N \) is large or small, equation (43) is the only estimator.

\(^8\)We have also performed the same analysis without the AR(1)-adjustment. As anticipated, the sum of the two estimates for the price impact parameters is larger and the regression residual is negatively serial correlated. As no additional insight is gained from this exercise, we do not report the results.
Finally, to check if Condition C is satisfied, we need the estimates for $y_0$, $\mu$ and $\alpha$. These parameter estimates are obtained from the regressions based on the symmetric specification, equation (34). For each stock that has at least 20 valid sets of daily parameter estimates in the first subperiod (120 trading days) or in the second subperiod (133 trading days), we conduct a test comparing the mean of daily $y_0$ estimates and the mean of daily $\tilde{y}_0$ estimates. The two-sided tests do not reject the null hypothesis of $y_0 - \tilde{y}_0 = 0$. Remarkably, the minimum $p$ values for these tests are 44.3% and 38.7% for the first and second subperiods, respectively. Similarly, two-population mean tests are conducted to compare $\mu$ with $\tilde{\mu}$, $\beta$ with $\tilde{\beta}$, as well as $\alpha$ with $\tilde{\alpha}_b + \tilde{\alpha}_s$. The statistics also do not reject the null hypotheses.

As a robustness check, for each stock, we pool the observations over the first subperiod when the tick size is 1/8 of a dollar. We find that both Conditions A and B pertaining to Proposition 2 are satisfied for all of the 1,748 stocks in our sample. Tests of Condition C also do not reject the null hypotheses.

B. Asymmetry in the Price Impact of Buy and Sell

Naturally, it is of interest to ascertain which of the two price impact parameter estimates is larger. Of the 1,748 sample stocks, 1,561 have at least one day with valid estimation results. Among these 1,561 stocks, 924 have more days with $\tilde{\alpha}_b$ estimates larger than $\tilde{\alpha}_s$ estimates, 91 have equal number, and the remaining 546 stocks have less number of days with larger $\tilde{\alpha}_b$ estimates. Therefore, 59.19% of these stocks have asymmetrically larger price impact in buyer-initiated trades for more than 50% of the trading days in 1997. The $\chi^2$ statistic for $\tilde{\alpha}_b > \tilde{\alpha}_s$ is 97.2, which is significant.

In magnitude, however, the difference $\tilde{\alpha}_b - \tilde{\alpha}_s$ is small relative to $\tilde{\alpha}_b + \tilde{\alpha}_s$. We consider a subsample of 1,131 stocks that have at least 20 days with valid estimation results. The average value of the difference $\tilde{\alpha}_b - \tilde{\alpha}_s$ over that of the sum $\tilde{\alpha}_b + \tilde{\alpha}_s$ is a ratio $\gamma$ of small percent. For 738 stocks that have more days with $\tilde{\alpha}_b > \tilde{\alpha}_s$, the average value of $\gamma$ is only 1.85% and the standard deviation is 8.89%. Weighted by the percentage of days with larger $\tilde{\alpha}_b$, $\gamma$ is 2.02% with a standard deviation of 8.89%. In this subsample, 738 have more days with $\tilde{\alpha}_b > \tilde{\alpha}_s$, 28 have equal number, and the remaining 365 stocks have less number of days with asymmetrically larger $\tilde{\alpha}_b$ estimates. The $\chi^2$ statistic for $\tilde{\alpha}_b > \tilde{\alpha}_s$ is 123.7.

9In this subsample, 738 have more days with $\tilde{\alpha}_b > \tilde{\alpha}_s$, 28 have equal number, and the remaining 365 stocks have less number of days with asymmetrically larger $\tilde{\alpha}_b$ estimates. The $\chi^2$ statistic for $\tilde{\alpha}_b > \tilde{\alpha}_s$ is 123.7.
deviation of 2.59%.

Therefore, while the non-parametric $\chi^2$ statistics are supportive of an asymmetrical larger price impact of a purchase, the difference between the two transitory price impacts is not economically significant. In a way, this outcome is not surprising. From Condition A, since $J_b \approx 1$ and $J_s \approx -1$, we have

$$\kappa \approx \frac{2\eta}{\alpha_b + \alpha_s} - \frac{\tilde{\alpha}_b - \tilde{\alpha}_s}{\alpha_b + \alpha_s}.$$  \hspace{1cm} (47)

In Table II, the subsample averages of $\kappa$ and $\eta$ are small positive numbers, especially for larger market capitalization stocks that have more days with 20 valid estimation results. If $\alpha_b$ were to be much larger than $\alpha_s$ frequently, then $\gamma$, the average of the second term in equation (47), would be large and $\kappa$ would be negative rather than positive.

Overall, for all these 1,131 stocks, $\gamma$ is -0.477% on average with the standard deviation being 9.71%. Only 1% of these stocks have $|\gamma|$ larger than 44.32%. In sum, despite the fact that $\tilde{\alpha}_b > \tilde{\alpha}_s$ occurs more regularly\(^{10}\), the transaction cost in percent of a buyer-initiated trade is not economically much different from that of a seller-initiated trade.

C. Shape of the Price Impact Function and Information Environment

As discussed in Section I.B.2, the scaling parameter $\tilde{\beta}$ is negatively related to the information asymmetry. Does the scaling parameter remain constant for a given security over time? A time-varying $\tilde{\beta}$ corresponds to a changing price impact function in a broader sense. We find that $\tilde{\beta}$ estimate varies from day to day. In Figure 2, we show two examples. One is a relatively liquid stock (Ticker Symbol: GE) and the other (Ticker Symbol: BNI) is less liquid. It is evident that $\tilde{\beta}$ does not remain constant. To reveal the nature of the fluctuation more clearly, the $y$ axis is plotted as $\ln(\tilde{\beta})$ instead. In this way, $\tilde{\beta}$ estimates that are smaller than one show up as negative values. As anticipated, less liquid stocks have more days with negative $\ln(\tilde{\beta})$ values, which suggest that the price impact is more dependent on trade size.

\(^{10}\)A concise yet comprehensive review of this temporary asymmetry between buy and sell is found in footnote 1 of Saar (2001).
ments? To answer this question, we obtain the dates of announcements from the I/B/E/S database and compute the standardized unexpected earnings. We apply the standard event study methodology (see MacKinlay (1997) for a review). This methodology is able to pick up the price change, especially the jump between the opening price and the closing price of previous trading day when there is an earnings surprise. However, we do not find statistically significant evidence that the \( \tilde{\beta} \) estimate is much smaller around the announcement date\(^{11}\).

The finding that \( \tilde{\beta} \) is unaffected by information events may be attributable to the inappropriateness of the event study methodology employed to probe the change in \( \tilde{\beta} \), or to the possibility that the \( \tilde{\beta} \) estimates from nonlinear least squares are poor and grossly inaccurate. To ascertain which is the case, we perform a cross-sectional study to examine the relation between the scaling parameter \( \tilde{\beta} \) and the information environment of a firm. Intuitively, the \( \tilde{\beta} \) parameter should be positively related to the number of shareholders. This is because the more shareholders there are, the larger is the heterogeneity in valuation and fewer trades are motivated by information asymmetry. Also, if a firm has more analysts monitoring it, the \( \tilde{\beta} \) parameter should be larger because analysts’ reports should in general reduce the information asymmetry between insiders and outsiders. In the same vein, a stock that is a component of the benchmark S&P 500 Index should have larger \( \tilde{\beta} \) estimates. Finally, it is of interest to ascertain whether the variance of implied true price change is related to the scaling parameter. When trading is volatile, one may interpret the volatility of the implied true price change as a result of larger heterogeneity in traders’ belief and smaller information asymmetry. Therefore, the variance of the implied true price change should also be positively related to \( \tilde{\beta} \).

To examine the validity of these intuitive arguments, we consider the following cross-sectional regression:

\[
\tilde{\beta}_j = b_0 + b_1 \ln(\text{Shareholders}_j) + b_2 \text{Analysts}_j + b_3 \text{Variance}_j + b_4 \text{S&P 500}_j + \epsilon_j. \tag{48}
\]

In this regression, \( \tilde{\beta}_j \) is the median of the daily \( \tilde{\beta} \) estimates for stock \( j \). The number of shareholders for firm \( j \) is \( \text{Shareholders}_j \) and the number of analysts who forecast earnings of firm \( j \) is denoted as \( \text{Analysts}_j \). The variable \( \text{Variance}_j \) is the average of the daily estimates for the

\(^{11}\)From the standpoint of event study, this finding is qualitatively consistent with Saar and Yu (2002).
variance of the implied true price change. S&P 500, is the dummy variable to indicate whether firm j was a constituent of the S&P 500 Index in 1997. Since the S&P 500 constituent stocks are widely held and followed by analysts, this dummy variable serves as a control in the regression. As usual, $\epsilon_j$ is the residual. The coefficients of the four variables are postulated to be positive and statistically significant.

In addition, we regress the standard deviation of the daily $\tilde{\beta}$ estimates on these variables. This standard deviation indicates the volatility of the shape of the price impact curve. It therefore measures the corresponding amount of uncertainty in the information asymmetry level. Intuitively, Shareholders, Analysts, and S&P 500 should be negatively related to this type of uncertainty because the liquidity-motivated trades by many shareholders as well as the coverage by more analysts are likely to reduce the amount of information shocks and thereby make the $\tilde{\beta}$ fairly stable. On the other hand, Variance, the intrinsic volatility in the implied true price, should be positively correlated with the uncertainty in the shape of the price impact curve. This is because both are proxies for uncertainty in the price change.

To perform these two regressions, we take a sample of firms that have at least 20 daily $\tilde{\beta}$ estimates so that the standard deviation of these $\tilde{\beta}$ estimates is obtainable. Statistics for Shareholders, and S&P 500 are extracted from the Compustat. Using the I/B/E/S, we derive Analysts by counting the number of analysts who had made at least one earnings forecast for firm j in 1997. Owing to the availability of these data, the sample size is reduced to 548.

In Table III, we report the results of these two cross-sectional regressions. Based on the Newey-West (1987) $t$-statistics in two-sided tests, we find that all the coefficients are significant at the 5% level when the dependent variable is the level of $\tilde{\beta}$ estimate. Therefore, there is reason to believe that asymmetric information is less of a problem for stocks that are followed by many analysts and are widely held. This result is consistent with the prediction by Spiegel and Subrahmanyam (2000). We also find that Variance, and S&P 500 are significant at the 5% level.

Statistics for the second regression with the standard deviation of the daily $\tilde{\beta}$ estimates as the dependent variable are also broadly consistent with the intuitive arguments. For this regression,
Variance, is the main determinant. In addition, the coefficient of Analysts, is negative with a p value of 5.34%. The coefficient of Shareholders, is also negative. Despite being statistically insignificant, we find that firms with more shareholders tend to reduce the volatility in the scaling parameter.

In summary, with regard to the earlier question concerning the reliability of \( \tilde{\beta} \) estimates, we have evidence to suggest that the possibility of \( \tilde{\beta} \) estimates being grossly unreliable is less likely. If these estimates were wrongly obtained, then one would not find any statistical significance in these cross-sectional regressions. It is not by chance that the coefficients in equation (48) are significant with the correct sign. Furthermore, the analysis indicates that the \( \tilde{\beta} \) estimate does reflect the information environment of a publicly listed company traded on the NYSE. When \( \tilde{\beta} \) is small (i.e. the price impact of block trades is significantly larger than small trades), and when there is more intertemporal variation in \( \tilde{\beta} \) (i.e. the shape of price impact curve fluctuates more dramatically), traders are likely to have asymmetric information among them.

V. Comparison of Trade-Based and Quote-Based Estimations

The approach of this paper is based solely on transaction data. In contrast, the midpoint and the effective spread are estimated with prevailing quotes. This section documents a comparison of the trade-based and quote-based statistics. Intriguingly, we find that the midpoint is not as good a proxy as one might have hitherto assumed. Since the transaction cost is computed with reference to the true price, the accuracy of the transaction cost measure is determined by the proximity of the proxy to the real true price. As the implied true price is better in this regard, the friction spread will be relatively more accurate than the effective spread.

A. Implied True Price versus Quotes’ Midpoint

Given that the implied true price \( \tilde{S}(t) \) is reliably estimated with the methodology described in the previous section, a natural question concerning its proximity to the real true price arises. Put it the other way round, in light of the implied true price, what is the fidelity of the midpoint in reflecting the real true price?
In Section I.D, we show that the second-order moment of price change can be used to compare the relative proximity of the proxy to the unobservable true price. Therefore, for the 188,945 valid NLS regressions of Section IV.A, we estimate the second-order moments of the price change in the prevailing midpoint and the change in the implied true price. Pairwise comparison results reveal that in 89.19% of these estimations, the implied true price is a better proxy than the midpoint. To the best of our knowledge, this is the first documentation that there is a proxy in the implied true price that is better than the midpoint.

To ascertain whether the difference between the implied true price and the midpoint is economically significant, we compute the ratio $R_d$ using equation (41), which is the daily average difference as a percentage of the bid-ask spread. The percentage of cases for which the implied true price is a better proxy than the midpoint is 94.34% before the tick size reduction and 85.10% after the reduction. Despite this decline, the ratio $R_d$ increases from 52.29% to 58.62%. The summary statistics by decile and tick size are documented in Table IV.

These statistics suggest that the difference is economically significant. Over the entire sample period, $R_d$ is 55.66% with a standard deviation of 18.68%. This average difference between the implied true price and the midpoint is about three standard deviations from zero. As an illustration, if the average quoted spread is 15 dollars per lot, then the difference between these two proxies is 4.175 dollars per lot on average.

Given these results, we conclude that the real true price is more likely to be closer to the implied true price than to the midpoint. More importantly, the difference between the implied true price and the midpoint is economically significant. A plausible reason for this outcome is that bid and ask quotes are mostly determined by limit orders and market makers’ inventories, which can have temporary imbalances that are unrelated to the asset’s value. Consequently, the midpoint may deviate from the true price. Moreover, so long as one of the two quotes is stale, there will be a bias in the midpoint.

**B. Friction Spread and Quote-Based Spreads**

This section documents the comparison of the friction spread with the quoted and effective spreads. The 1,748 stocks are grouped into deciles with the largest market capitalization stocks
in the first decile. The friction spread is calculated using the implied true price and equation (29), whereas the two quote-based measures use the prevailing midpoint as the benchmark to estimate the spreads.

We compute the volume-weighted averages of these spreads on a daily basis. For each quantity, we obtain a time series of 253 cross-sectional averages. The summary statistics for these measures of transaction costs are given in Table V. Panel A lists the statistics before the tick size reduction and Panel B displays the corresponding statistics after the reduction. As expected, large market capitalization stocks have lower transaction costs in percent. The volume-weighted mean values for all the three measures in percent increase monotonously as the market capitalization decreases from the first decile to the last decile.

We find that the friction spread is more compatible with the effective spread than the quoted spread, as anticipated. Before the tick size reduction, the volume-weighted average friction spreads range from 12.01 dollars per lot to 13.45 dollars per lot. This range overlaps with the range of 12.35 dollars to 14.52 dollars per lot for the volume-weighted average effective spreads. However, after the tick size was reduced to 1/16 of a dollar, the difference between the friction and effective spreads is significantly larger. As tabulated in Panel B, average friction spreads range from 7.16 to 9.57 dollars per lot while the average effective spreads are from 9.87 to 12.64 dollars per lot. The respective declines from pre- to post-reduction for these three measures are statistically significant at the 1% level. These declines concur with the prediction of Harris (1994) that the spread will be narrower when the minimum tick size is smaller.

To provide a visual presentation of these observations, Panel A of Figure 3 plots the time series of the cross-sectional volume-weighted averages for the largest market capitalization stocks. It is evident that the difference between the friction spread and the effective spread is much smaller than that between the friction spread and the quoted spread. Before the tick size reduction, the friction spread is quite similar to the effective spread. After the reduction, however, the

---

12 As highlighted by Peterson and Sirri (2003), we are mindful of the upward biases when using the TAQ data to estimate the effective spread. Thus, the actual effective spread may be lower than the values that we have computed. Owing to the limitation of the TAQ database and the signing algorithm, it is not possible to estimate the magnitude of the bias directly.

13 Another important observation is that the mean friction spreads in percent are compatible with the respective sums of the two average price impact parameter estimates in Table II. This close correspondence reinforces our economic interpretation of the price impact discussed in Section I.B.3, and provides further support for Proposition 3.
friction spread is smaller than the effective spread by quite a significant amount. It seems that the friction spread responses to the reduction in minimum tick size more sensitively. On average, the decline in friction spreads is about 40.4% for these stocks, which shows up as a precipitous drop on June 24.

It is also apparent in Panel A that the spreads are more constant before the reduction than after. As anticipated, there are sharp rises on black Monday. In fact, the peaks in the respective spreads are observed on the following day. One possible explanation for this phenomenon is that the NYSE specialists of large market capitalization stocks became more risk averse when the market crashed. From microstructure theory, the spread will widen when market makers attempt to protect themselves against unfavorable price movements.

In Panel B, the time series of the cross-sectional volume-weighted averages for the small market capitalization (ninth decile) stocks in our sample are plotted. Again, the friction spread becomes more different from the effective spread after the tick size was halved. Intriguingly, however, there is no sharp rise in response to the black Monday. It is also noteworthy that the friction spread is relatively less volatile than the two quote-based measures.

In summary, the evidence suggests that even the effective spread tends to be larger than the friction spread, especially after the tick size was reduced by half. Since the implied true price is a better proxy than the midpoint, the friction spread \( \tilde{F}(t, x) \) defined in equation (29) ought to be a better measure of implicit transaction costs than the effective spread.

### VI. Summary and Conclusions

Given trading friction, investors buy at a price higher than the true price and vice versa when they sell. In addition, transaction prices are bounded and play a significant role in discovering the unobservable true price. These insights motivate the formulation of the transaction price as a product of the true price and a function of signed volume. In particular, the hyperbolic tangent function is shown to be both theoretically and empirically well suited to capture the instantaneous price impact of a trade.
Our model allows the true price to be inferred from the transaction price and the signed volume alone. Having this implied true price as an alternative proxy to the quotes’ midpoint is useful in research that contributes to a better understanding of the high-frequency dynamics of trades and quotes.

The implied true price is found to be closer to the real true price than the midpoint. This result is obtained from the analysis of 1,748 firms traded on the NYSE in 1997. Moreover, the difference between the implied true price and the midpoint is both statistically and economically significant. This finding has implication for researchers who rely on the midpoint as a proxy for the true price.

In our formulation, a scaling parameter determines the shape of the price impact function. It is shown to be interpretable as an indicator of information asymmetry. A small scaling parameter implies that the disparity in the price impacts of large and small trades is large. Cross-sectional analysis demonstrates that the scaling parameter is positively related to the number of shareholders, number of analysts who provide earnings forecasts as well as the variance of intra-day return computed with the implied true price. Overall, after controlling for whether the stock is a component of the S&P 500 Index, we find that the scaling parameter correlates with these information environment variables of a firm.

We apply the model to study the costs of trading on the NYSE. Our trade-based measure of transaction costs is tested with 1,748 liquid and illiquid stocks traded on the NYSE, over two sub-periods delineated by the reduction of the minimum tick size in 1997. As anticipated, large market capitalization stocks have lower transaction costs. On average, transaction costs decline by 38.9% after the minimum tick size is halved. More important, using our measure, the quote-based effective spread is diagnosed to have an upward bias.

With these findings, this paper has extended the literature on price impacts and transaction costs. We highlight that the time-varying nature of the price impact function is a source of uncertainty additional to the market risk. An understanding of this uncertainty is relevant to asset pricing and optimal liquidation of portfolios.
Appendix A: Proof of Proposition 1

In this Appendix, we demonstrate that \( \tanh(\beta x) \) is a germane function to model the instantaneous price impact of a trade with signed volume \( x \) as a random variable. Although the transaction price \( P(t, x) \) differs from the true price \( S(t) \), our mathematical analysis is motivated by a common assumption in the market microstructure literature that the ratio \( P(t, x)/S(t) \) fluctuates around unity. Equivalently, \( \ln \left( \frac{P(t, x)}{S(t)} \right) \), which is the price impact in percent, is expected to vanish: \( E \left[ \ln \left( \frac{P(t, x)}{S(t)} \right) \right] = 0 \).

**Proof.** By definition, the signed volume \( x \) is given by \( x = I|x| \), where \( I \) is the trade sign and \( |x| \) the volume. The trade sign is a Bernoulli random variable that equals one or minus one for buyer- or seller-initiated trade, respectively. To construct a representation for \( P(t, x)/S(t) \) in terms of the signed volume \( x \), we invoke a common assumption in the market microstructure literature: Buyer- and seller-initiated trades are equally likely. The probability measure \( \Pr \) associated with the signed volume \( x \) is

\[
\Pr(x > 0) = \frac{1}{2} = \Pr(x < 0).
\]

(A.1)

Let the scaling parameter \( \beta \) be a constant and consider the expected value of \( e^{\beta I|x|} \) under the probability measure \( \Pr \). It follows that

\[
E \left[ e^{\beta I|x|} \right] = \frac{1}{2} \left( e^{\beta x} + e^{-\beta x} \right) = \cosh(\beta x).
\]

(A.2)

Since \( \cosh(\beta x) \neq 0 \) for all \( x \), we can divide both sides of the equation by \( \cosh(\beta x) \) to obtain

\[
E \left[ \frac{e^{\beta I|x|}}{\cosh(\beta x)} \right] = E \left[ \frac{e^{\beta x}}{\cosh(\beta x)} \right] = 1,
\]

(A.3)

which implies that

\[
E \left[ \frac{e^{\beta x}}{\cosh(\beta x)} - 1 \right] = E \left[ \tanh(\beta x) \right] = 0.
\]

(A.4)

If the transaction price \( P(t, x) \) contributes to the discovery of \( S(t) \), then \( \ln \left( \frac{P(t, x)}{S(t)} \right) \) is
equally likely to be positive and negative. Therefore, under the measure \( \Pr \), this assumption is expressed as

\[
E \left[ \ln \left( \frac{P(t, x)}{S(t)} \right) \right] = 0. \tag{A.5}
\]

Consequently, up to a positive constant \( \alpha/2 \), we obtain from equations (A.4) and (A.5) that \( \tanh(\beta x) \) is a version, or a representation of \( \ln \left( \frac{P(t, x)}{S(t)} \right) \). Namely,

\[
\ln \left( \frac{P(t, x)}{S(t)} \right) = \frac{\alpha}{2} \tanh(\beta x) \quad \text{almost surely}, \tag{A.6}
\]

since

\[
E \left[ \ln \left( \frac{P(t, x)}{S(t)} \right) \right] = 0 = \frac{\alpha}{2} E \left[ \tanh(\beta x) \right]. \tag{A.7}
\]

**Appendix B: Proof of Proposition 2**

In this Appendix, we relax the strong assumption that buyer- and seller-initiated trades are equally likely. The outcome is that the price impact of a purchase is different from that of a sale.

**PROOF.** Let \( \kappa \) be a real number in the open interval \((-1, 1)\). The asymmetric probability measure that corresponds to different frequencies of buy and sell is parameterized by \( \kappa \) as follows:

\[
\Pr(x > 0) = \frac{1 + \kappa}{2}; \quad \Pr(x < 0) = \frac{1 - \kappa}{2}. \tag{B.1}
\]

Since the trade sign is asymmetrically distributed, the expected value of the deviation from the true price in percent is no longer zero:

\[
\eta = E \left[ \ln \left( \frac{P(t, x)}{S(t)} \right) \right] \neq 0. \tag{B.2}
\]

Following the same method in Appendix A but using the measure defined by equation (B.1), we arrive at

\[
\ln \left( \frac{P(t, x)}{S(t)} \right) = \frac{\tilde{\alpha}}{2} \left( \frac{\tanh(\tilde{\beta} x)}{1 + \kappa \tanh(\tilde{\beta} x)} \right) + \eta, \quad \text{almost surely.} \tag{B.3}
\]
To obtain this representation of $\ln \left( \frac{P(t, x)}{S(t)} \right)$, we compute the expected value of $e^{\tilde{\beta} x}$ under the asymmetric probability measure:

$$E \left[ e^{\tilde{\beta} x} \right] = \frac{1 + \kappa}{2} e^{\tilde{\beta} x} + \frac{1 - \kappa}{2} e^{-\tilde{\beta} x} = \cosh(\tilde{\beta} x) + \kappa \sinh(\tilde{\beta} x).$$  \hspace{1cm} (B.4)

The expected value of a strictly positive quantity $e^{\tilde{\beta} x}$ must be positive. Therefore, the right side of this equation is non-zero for all $x$ and we have

$$E \left[ \frac{e^{\tilde{\beta} x}}{\cosh(\tilde{\beta} x) + \kappa \sinh(\tilde{\beta} x)} - 1 \right] = 0. \hspace{1cm} (B.5)$$

Consequently,

$$E \left[ \frac{\sinh(\tilde{\beta} x) - \kappa \sinh(\tilde{\beta} x)}{\cosh(\tilde{\beta} x) + \kappa \sinh(\tilde{\beta} x)} \right] = (1 - \kappa)E \left[ \frac{\tanh(\tilde{\beta} x)}{1 + \kappa \tanh(\tilde{\beta} x)} \right] = 0. \hspace{1cm} (B.6)$$

To satisfy this equation for all possible $\kappa$, it must be that $E \left[ \frac{\tanh(\tilde{\beta} x)}{1 + \kappa \tanh(\tilde{\beta} x)} \right] = 0$. Since

$$E \left[ \ln \left( \frac{P(t, x)}{S(t)} \right) - \eta \right] = 0, \hspace{1cm} (B.7)$$

equation (B.3) is thus a representation of $\ln \left( \frac{P(t, x)}{S(t)} \right)$.

From equation (B.3), one sees that as the trade size becomes asymptotically large,

$$-\frac{\tilde{\alpha}}{2} \left( \frac{1}{1 - \kappa} \right) + \eta < \ln \left( \frac{P(t, x)}{S(t)} \right) = \frac{\tilde{\alpha}}{2} \left( \frac{1}{1 + \kappa} \right) + \eta. \hspace{1cm} (B.8)$$

Therefore, two different price impact parameters implicit in this representation are

$$\tilde{\alpha}_b = \frac{1}{2} \left( \frac{\tilde{\alpha}}{1 + \kappa} \right) + \eta; \hspace{1cm} (B.9)$$

$$\tilde{\alpha}_s = \frac{1}{2} \left( \frac{\tilde{\alpha}}{1 - \kappa} \right) - \eta. \hspace{1cm} (B.10)$$

As expected, when $\kappa$ and $\eta$ are both zero, the asymmetric formalism reduces to the symmetric
case with $\tilde{\alpha} \to \alpha$. These two equations appear to be singular when $\kappa \to 1$ or $\kappa \to -1$. But, it is not to be. Since $\kappa = 1$ means that all the trades are buyer-initiated, there is not a single seller-initiated trade and therefore no price impact from a sale. Because no trade means no impact, we have $\tilde{\alpha}_s = 0$. To satisfy equation (B.10) when $\kappa = 1$ and $\tilde{\alpha}_s = 0$, the only possibility is $\tilde{\alpha} = 0$ and $\eta = 0$. In turn, equation (B.9) suggests that $\tilde{\alpha}_b = 0$. Taken together, the conclusion is that at any given time, the transaction price of the buyer-initiated trade is the only price in the market because there is no seller-initiated trade at all. As there is only one price, that price must be the true price at the time of trade. Similar arguments apply when $\kappa = -1$. As $\tilde{\alpha}_b$ and $\tilde{\alpha}_s$ are continuous with respect to $\kappa$, both the price impact parameters will be infinitesimally small when $\kappa$ approaches $\pm 1$.

Knowing that the two asymptotic bounds are given by equation (B.8), which is re-written as

$$-\tilde{\alpha}_s < \ln \left( \frac{P(t,x)}{S(t)} \right) < \tilde{\alpha}_b,$$  \hspace{1cm} (B.11)

we construct a functional form that is equivalent to equation (B.3). In view of these two bounds, we postulate

$$\ln \left( \frac{P(t,x)}{S(t)} \right) = \begin{cases} \tilde{\alpha}_b \tanh(\beta x) & \text{if } x > 0, \\ \tilde{\alpha}_s \tanh(\beta x) & \text{if } x < 0. \end{cases}$$ \hspace{1cm} (B.12)

and derive a condition that is consistent with the original equation (B.3).

Suppose $D_{x>0}$ is the dummy variable (indicator function) associated with a purchase and $D_{x<0}$ with a sale. Then, this equation can be written as

$$\ln \left( \frac{P(t,x)}{S(t)} \right) = \tilde{\alpha}_b \tanh(\beta x) D_{x>0} + \tilde{\alpha}_s \tanh(\beta x) D_{x<0} \hspace{1cm} (B.13)$$

Under the measure $Pr$ expressed in equation (B.1), we have

---

14When $\kappa \neq \pm 1$, it is possible that by chance $2\eta$ is approximately $\tilde{\alpha}/(1 - \kappa)$. From equation (B.10), we have $\tilde{\alpha}_s \approx 0$. In this instance, equation (B.9) suggests that the price impact of buy $\tilde{\alpha}_b$ is approximately $\tilde{\alpha}/(1 - \kappa^2)$, which is non-zero in general. Conversely, if $2\eta \approx -\tilde{\alpha}/(1 + \kappa)$, then from equation (B.10), $\tilde{\alpha}_b \approx 0$ but $\tilde{\alpha}_s \approx \tilde{\alpha}/(1 - \kappa^2) \neq 0$. Therefore, when $\kappa \neq \pm 1$, it is possible to have one of the two parameters to be close to zero. But, if $\kappa = \pm 1$, then both $\tilde{\alpha}_b$ and $\tilde{\alpha}_s$ will have to vanish.
\[ E \left[ \ln \left( \frac{P(t, x)}{S(t)} \right) \right] = E[\tilde{\alpha}_b \tanh(\tilde{\beta}x)D_{x>0} + \tilde{\alpha}_s \tanh(\tilde{\beta}x)D_{x<0}] \]
\[ = \tilde{\alpha}_b E[\tanh(\tilde{\beta}x)|x>0] \Pr(x>0) + \tilde{\alpha}_s E[\tanh(\tilde{\beta}x)|x<0] \Pr(x<0) \]
\[ = \frac{\tilde{\alpha}_b(1 + \kappa)}{2} E[\tanh(\tilde{\beta}x)|x>0] + \frac{\tilde{\alpha}_s(1 - \kappa)}{2} E[\tanh(\tilde{\beta}x)|x<0]. \quad (B.14) \]

With \[ E \left[ \ln \left( \frac{P(t, x)}{S(t)} \right) \right] = \eta \] and denoting \[ E[\tanh(\tilde{\beta}x)|x>0] \] by \( J_b \) and \[ E[\tanh(\tilde{\beta}x)|x<0] \] by \( J_s \), equation (B.14) becomes
\[ \eta = \frac{\tilde{\alpha}_b(1 + \kappa)}{2} J_b + \frac{\tilde{\alpha}_s(1 - \kappa)}{2} J_s. \quad (B.15) \]

Note that the symmetric case discussed in Appendix A corresponds to \( \kappa = 0 \), \( J_s = -J_b \) and \( \tilde{\alpha}_b = \tilde{\alpha}_s = \alpha / 2 \). Substituting these quantities into equation (B.15), it is obvious that \( \eta = 0 \), which is consistent with equation (A.5) as required.

Rearranging the terms in equation (B.15), the consistency condition is obtained as follows:
\[ \kappa = \frac{2\eta - \tilde{\alpha}_b J_b - \tilde{\alpha}_s J_s}{\tilde{\alpha}_b J_b - \tilde{\alpha}_s J_s}. \quad (B.16) \]

This is a condition that the parameters and equation (B.12) must satisfy. Since \( J_b > 0 \) and \( J_s < 0 \), the denominator is always positive and the ratio is non-singular.

**Appendix C: Proof of Proposition 3**

In this appendix, we provide a more analytical presentation that relates the asymmetric framework with its symmetric counterpart. The implication of \( \tilde{\alpha}_b + \tilde{\alpha}_s \approx \alpha \) is that despite the strong assumption required in the derivation of symmetric price impact function, the associated price impact parameter \( \alpha \) is still a a good estimate of the round-trip transaction cost in percent when the asymmetry \( \kappa \) is small in magnitude.

PROOF. The key step of the proof lies in working out the expected values of both sides of equa-
tions (31) and (34). First, we note that the mean of the dependent variable \( y_i \) in these two equations is a statistic by itself, irrespective of the econometrical specification. Given \( N \) observations, it can be consistently estimated as

\[
E[y_i] = \frac{1}{N} \sum_{j=1}^{N} y_j.
\]

(C.1)

Therefore, we can equate the expected values for the right side of equations (31) and (34), which spells out the econometrical specifications for the symmetric and asymmetric scenarios, respectively.

By construction, \( E[\epsilon] = 0 \) and \( E[\tilde{\epsilon}] = 0 \). Since \( E[y_i] \) is given by equation (C.1) regardless of the econometrical specification, it follows that

\[
y_0 + \mu E[\tau_i] + \alpha 2 E \left[ \frac{\tanh(\beta x_i) - \tanh(\beta x_{i-1})}{\tau_i} \right] = \tilde{y}_0 + \tilde{\mu} E[\tau_i] \\
+ \tilde{\alpha}_b E \left[ \frac{\tanh(\tilde{\beta} x_i) D_{x_i,>0} - \tanh(\tilde{\beta} x_{i-1}) D_{x_{i-1},>0}}{\tau_i} \right] \\
+ \tilde{\alpha}_s E \left[ \frac{\tanh(\tilde{\beta} x_i) D_{x_i,<0} - \tanh(\tilde{\beta} x_{i-1}) D_{x_{i-1},<0}}{\tau_i} \right].
\]

(C.2)

There are three regressors (“variables”) on the left and four on the right. But, there is only one equation. Despite having more variables than equations, the following three relations constitute an approximate solution of equation (C.2):

\[
y_0 \approx \tilde{y}_0; \quad \mu \approx \tilde{\mu}; \quad \frac{\alpha}{2} K \approx 1 + \frac{k}{2} \tilde{\alpha}_b K_b + \frac{1}{2} \tilde{\alpha}_s K_s,
\]

(C.3)

(C.4)

(C.5)

where

\[
K \equiv E \left[ \frac{\tanh(\beta x_i) - \tanh(\beta x_{i-1})}{\tau_i} \right].
\]

(C.6)
under the symmetric probability measure and
\[
K_b \equiv E \left[ \frac{\tanh(\tilde{\beta}x_i)}{\tau_i} \mid x_i > 0 \right] - E \left[ \frac{\tanh(\tilde{\beta}x_{i-1})}{\tau_i} \mid x_{i-1} > 0 \right]; \quad (C.7)
\]
\[
K_s \equiv E \left[ \frac{\tanh(\tilde{\beta}x_i)}{\tau_i} \mid x_i < 0 \right] - E \left[ \frac{\tanh(\tilde{\beta}x_{i-1})}{\tau_i} \mid x_{i-1} < 0 \right] \quad (C.8)
\]
under the asymmetric probability measure.

In contrast to equation (B.16), the relations expressed in equations (C.3) to (C.5) are approximate because the variables or degrees of freedom may not be exactly orthogonal. After a re-arrangement, equation (C.5) is
\[
\alpha K \approx \tilde{\alpha}_b K_b + \tilde{\alpha}_s K_s + \kappa (\tilde{\alpha}_b K_b - \tilde{\alpha}_s K_s). \quad (C.9)
\]
A crucial property of the hyperbolic tangent function is that \( \tanh(\tilde{\beta}x) \) and \( \tanh(\beta x) \) are almost identical when the scaling parameters \( \beta \) and \( \tilde{\beta} \) are not small, or when the trade size \(|x|\) is very large. Therefore, if the magnitude of asymmetry \(|\kappa|\) is small, we have
\[
\beta \approx \tilde{\beta} \quad (C.10)
\]
and
\[
K_b \approx K_s \approx K. \quad (C.11)
\]
It follows that
\[
\alpha \approx \tilde{\alpha}_b + \tilde{\alpha}_s. \quad (C.12)
\]
References


Our NYSE sample consists of 1,748 common stocks for the year 1997. In that year, there were 253 trading days, of which 120 days were traded with the minimum tick size of 1/8 of a dollar. The remaining 133 days were traded with 1/16 of a dollar. We use CRSP's December 1996 data to obtain the market capitalizations, shares outstanding and prices of the 1,748 common stocks traded primarily on the NYSE. Statistics for trading activities are computed with the TAQ database. The consolidated number of trades and volume traded are annual aggregated transactions that took place on the NYSE and other exchanges in 1997. Like other researchers, we do not use trades executed at other exchanges in our study. The mean, standard deviation (Std), minimum (Min), 1st percentile (1st per), median, 99th percentile (99th per) and maximum (Max) are computed cross-sectionally.

<table>
<thead>
<tr>
<th>Stock</th>
<th>NYSE</th>
<th>Consolidated</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Market Capitalization (in Million)</td>
<td>Shares Outstanding (in Million)</td>
</tr>
<tr>
<td>Mean</td>
<td>3,488.4</td>
<td>134.9</td>
</tr>
<tr>
<td>Std</td>
<td>9,787.3</td>
<td>400.7</td>
</tr>
<tr>
<td>Min</td>
<td>2.9</td>
<td>0.8</td>
</tr>
<tr>
<td>1st per</td>
<td>15.3</td>
<td>3.5</td>
</tr>
<tr>
<td>Median</td>
<td>791.0</td>
<td>39.1</td>
</tr>
<tr>
<td>99th per</td>
<td>46,232.6</td>
<td>1,818.6</td>
</tr>
<tr>
<td>Max</td>
<td>162,789.9</td>
<td>9,878.5</td>
</tr>
</tbody>
</table>
Table II
Summary Statistics for the Estimation Results

This table reports the NLS regression results. Only the transaction data, namely, the transaction price $P(t_i, x_i)$, time of transaction $t_i$ and signed volume $x_i$ are required in the estimation. With $\rho_0$ and $\rho_1$ being the parameters, the dependent variable $y_i$ is first obtained from the residual in the AR(1) process:

$$r_i = \rho_0 + \rho_1 r_{i-1} + y_i.$$  

The price change $r_i$ is the difference $\ln(P(t_i, x_i)) - \ln(P(t_{i-1}, x_{i-1}))$ divided by the square root of the trade duration, which is denoted by $\tau_i$. The motivation for this AR(1) adjustment is to mitigate the microstructure effects of price discreteness and bid-ask bounce. The AR(1)-adjusted price change $y_i$ is specified as

$$y_i = \tilde{y}_0 + (\tilde{\alpha}_b \tanh(\tilde{\beta} x_i) D_{x_i > 0} + \tilde{\alpha}_s \tanh(\tilde{\beta} x_i) D_{x_i < 0} - \tilde{\alpha}_b \tanh(\tilde{\beta} x_{i-1}) D_{x_{i-1} > 0} - \tilde{\alpha}_s \tanh(\tilde{\beta} x_{i-1}) D_{x_{i-1} < 0}) / \tau_i + \tilde{\mu} \tau_i + \tilde{e}_i,$$

where $\tilde{y}_0$ is the intercept and $\tilde{e}_i$ the error term. The dummy variables of the trade sign are $D_{x_i > 0}$ and $D_{x_i < 0}$. Although there are a total of seven parameters in this specification, only $\tilde{\alpha}_b$, $\tilde{\alpha}_s$ and $\tilde{\beta}$ are of interest, as their estimates allow the implied true price to be inferred from the model.

In this table, “Total” refers to the total number of regressions performed for each decile of stocks sorted on market capitalization. The row designated by “Valid” contains the numbers of valid estimation results that pass the test of Condition A. We report the average number of observations in the regression, $N$, along with the standard deviation (Std). Also tabulated are the summary statistics for the asymmetry measure $\kappa$ and the expected value of $\ln P(t, x) / \tilde{S}(t)$, where $\tilde{S}(t)$ is the implied true price. We employ the root mean square error (RMSE) of the regression and the adjusted $R^2$ denoted by $\widehat{R}^2$ to measure the goodness of fit.
Panel A: Subsample period from January 2, 1997 to June 23, 1997

<table>
<thead>
<tr>
<th>Decile</th>
<th>Largest</th>
<th>2nd</th>
<th>3rd</th>
<th>4th</th>
<th>5th</th>
<th>6th</th>
<th>7th</th>
<th>8th</th>
<th>9th</th>
<th>10th</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>20,709</td>
<td>19,643</td>
<td>16,261</td>
<td>11,139</td>
<td>6,145</td>
<td>5,145</td>
<td>2,238</td>
<td>1,568</td>
<td>746</td>
<td>240</td>
</tr>
<tr>
<td>Valid</td>
<td>20,708</td>
<td>19,626</td>
<td>16,229</td>
<td>11,114</td>
<td>6,142</td>
<td>5,122</td>
<td>2,222</td>
<td>1,552</td>
<td>743</td>
<td>239</td>
</tr>
<tr>
<td>% Valid</td>
<td>100.00</td>
<td>99.91</td>
<td>99.80</td>
<td>99.78</td>
<td>99.55</td>
<td>99.29</td>
<td>98.98</td>
<td>99.60</td>
<td>99.58</td>
<td></td>
</tr>
</tbody>
</table>

| Mean $N$ | 385.2 | 170.0 | 135.3 | 99.7 | 91.9 | 100.9 | 90.1 | 83.1 | 86.4 | 88.3 |
| Std $N$  | 279.6 | 146.5 | 118.5 | 65.7 | 77.2 | 88.5 | 60.3 | 44.8 | 45.9 | 61.7 |
| Mean $\kappa$ in % | 5.48 | 3.13 | 2.57 | 2.75 | 5.59 | 6.84 | 5.93 | 8.15 | 9.44 | 16.46 |
| Std $\kappa$ in % | 21.44 | 27.08 | 29.48 | 31.66 | 33.08 | 34.73 | 33.74 | 32.50 | 33.36 |
| Mean $\eta$ in % | 0.004 | 0.003 | 0.005 | 0.003 | 0.012 | 0.015 | 0.016 | 0.053 | 0.043 | 0.130 |
| Std $\eta$ in % | 0.050 | 0.070 | 0.120 | 0.123 | 0.146 | 0.185 | 0.249 | 0.309 | 0.360 | 0.917 |
| Mean $\alpha_b$ in % | 0.116 | 0.153 | 0.200 | 0.251 | 0.267 | 0.340 | 0.387 | 0.485 | 0.566 | 1.247 |
| Std $\alpha_b$ in % | 0.367 | 0.377 | 0.221 | 0.972 | 0.317 | 0.755 | 0.344 | 0.472 | 0.505 | 1.301 |
| Mean $\alpha_s$ in % | 0.115 | 0.151 | 0.201 | 0.236 | 0.263 | 0.344 | 0.397 | 0.469 | 0.591 | 1.405 |
| Std $\alpha_s$ in % | 0.129 | 0.088 | 0.652 | 0.164 | 0.193 | 0.660 | 0.349 | 0.377 | 0.604 | 1.436 |
| Mean $\beta$ | 4.21 | 4.08 | 4.27 | 4.15 | 3.87 | 4.15 | 3.90 | 4.16 | 4.03 | 5.19 |
| Std $\beta$ | 3.02 | 3.15 | 8.56 | 5.84 | 3.84 | 4.01 | 3.59 | 4.36 | 3.84 | 9.65 |

| Mean $F$ Statistic | 1254 | 577 | 540 | 426 | 341 | 382 | 376 | 301 | 254 | 326 |
| Std $F$ Statistic | 1292 | 844 | 891 | 950 | 778 | 766 | 1057 | 822 | 393 | 678 |
| Mean RMSE in % | 2.22 | 2.36 | 2.81 | 3.12 | 3.77 | 4.72 | 5.33 | 6.51 | 8.97 | 18.00 |
| Std RMSE in % | 0.92 | 1.19 | 2.16 | 1.78 | 2.63 | 2.87 | 3.49 | 4.34 | 6.76 | 17.03 |
| Mean $\widehat{R}^2$ in % | 68.6 | 62.9 | 63.4 | 60.8 | 56.6 | 57.2 | 56.1 | 54.5 | 56.9 | 57.8 |
| Std $\widehat{R}^2$ in % | 15.8 | 20.7 | 22.7 | 24.4 | 25.6 | 25.4 | 25.8 | 26.0 | 23.7 | 23.4 |
Panel B: Subsample period from June 24, 1997 to December 31, 1997

<table>
<thead>
<tr>
<th>Decile</th>
<th>Largest</th>
<th>2nd</th>
<th>3rd</th>
<th>4th</th>
<th>5th</th>
<th>6th</th>
<th>7th</th>
<th>8th</th>
<th>9th</th>
<th>10th</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>22,144</td>
<td>21,660</td>
<td>19,498</td>
<td>14,967</td>
<td>9,625</td>
<td>9,148</td>
<td>3,869</td>
<td>3,154</td>
<td>1,537</td>
<td>662</td>
</tr>
<tr>
<td>Valid</td>
<td>22,142</td>
<td>21,636</td>
<td>19,463</td>
<td>14,909</td>
<td>9,534</td>
<td>9,066</td>
<td>3,821</td>
<td>3,132</td>
<td>1,521</td>
<td>658</td>
</tr>
<tr>
<td>% Valid</td>
<td>99.99</td>
<td>99.89</td>
<td>99.82</td>
<td>99.61</td>
<td>99.05</td>
<td>99.10</td>
<td>98.76</td>
<td>99.30</td>
<td>98.96</td>
<td>99.40</td>
</tr>
</tbody>
</table>

| Mean $N$ | 476.6 | 212.5 | 157.1 | 117.7 | 99.3 | 104.1 | 90.5 | 87.5 | 83.4 | 103.3 |
| Std $N$ | 333.5 | 156.5 | 131.6 | 85.7 | 77.5 | 78.7 | 56.5 | 59.5 | 41.8 | 72.5 |
| Mean $\kappa$ in % | 6.37 | 5.42 | 3.66 | 4.39 | 5.84 | 5.35 | 6.82 | 5.32 | 6.11 | 11.94 |
| Std $\kappa$ in % | 15.88 | 20.14 | 23.68 | 26.25 | 28.06 | 27.97 | 30.25 | 30.84 | 30.64 | 28.86 |
| Mean $\eta$ in % | 0.003 | 0.003 | 0.001 | 0.004 | 0.006 | 0.006 | 0.009 | 0.013 | 0.008 | 0.073 |
| Std $\eta$ in % | 0.027 | 0.041 | 0.094 | 0.069 | 0.084 | 0.111 | 0.138 | 0.195 | 0.263 | 0.554 |

| Mean $\alpha_b$ in % | 0.062 | 0.091 | 0.119 | 0.142 | 0.186 | 0.193 | 0.252 | 0.294 | 0.409 | 0.691 |
| Std $\alpha_b$ in % | 0.035 | 0.543 | 0.265 | 0.297 | 0.922 | 0.221 | 1.221 | 0.358 | 1.779 | 1.043 |
| Mean $\alpha_s$ in % | 0.063 | 0.093 | 0.122 | 0.147 | 0.182 | 0.195 | 0.271 | 0.306 | 0.372 | 0.706 |
| Std $\alpha_s$ in % | 0.046 | 0.490 | 0.450 | 0.666 | 0.676 | 0.197 | 1.261 | 0.594 | 0.727 | 1.039 |

| Mean $\hat{\beta}$ | 3.12 | 3.32 | 3.37 | 3.25 | 3.24 | 3.14 | 3.08 | 3.20 | 2.91 | 3.64 |
| Std $\hat{\beta}$ | 2.70 | 3.37 | 3.16 | 4.10 | 13.17 | 4.57 | 7.08 | 6.94 | 3.67 | 5.70 |

| Mean $F$ Statistic | 744 | 307 | 256 | 210 | 175 | 171 | 156 | 172 | 120 | 152 |
| Std $F$ Statistic | 783 | 429 | 386 | 563 | 455 | 394 | 317 | 508 | 258 | 244 |
| Mean RMSE in % | 2.04 | 2.33 | 2.63 | 2.90 | 3.48 | 4.05 | 4.61 | 5.46 | 7.22 | 13.41 |
| Std RMSE in % | 1.02 | 1.11 | 2.19 | 1.68 | 2.03 | 2.55 | 3.19 | 3.97 | 5.62 | 17.07 |
| Mean $\hat{R}^2$ in % | 53.7 | 48.0 | 49.2 | 47.3 | 43.7 | 44.6 | 43.1 | 45.2 | 42.2 | 46.8 |
| Std $\hat{R}^2$ in % | 15.6 | 18.9 | 21.1 | 22.9 | 23.5 | 23.3 | 24.2 | 24.2 | 22.7 | 21.3 |
Table III
Cross-Sectional Regressions of $\tilde{\beta}$ Estimates on Information Environment Variables

This table reports the regression statistics for the following cross-sectional specification:

$$Y_j = b_0 + b_1 \ln(\text{Shareholders}_j) + b_2 \text{Analysts}_j + b_3 \text{Variance}_j + b_4 \text{S&P 500}_j + \epsilon_j.$$ 

The subsample comprises 548 firms traded on the NYSE in 1997. Two regressions are performed. For the first regression, the dependent variable $Y_j$ is the median of the daily scaling parameter ($\tilde{\beta}$) estimates of firm $j$. For the second regression, $Y_j$ is the standard deviation of these daily estimates. The explanatory variables are Shareholders$_j$, the number of shareholders of firm $j$; Analysts$_j$, the number of analysts following the firm; Variance$_j$, the average of the daily variances of the implied true price change; and S&P 500$_j$, the dummy variable indicating whether the firm is a component of the S&P 500. The residual is denoted by $\epsilon_j$, and $b_0$ to $b_4$ are the coefficients (Coeff) of these explanatory variables. To indicate the statistical significance, Newey-West $t$ statistics ($t$ stat) and the corresponding $p$ values are tabulated.

<table>
<thead>
<tr>
<th></th>
<th>Median $\tilde{\beta}$</th>
<th></th>
<th>Standard Deviation of $\tilde{\beta}$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coeff</td>
<td>$t$ stat</td>
<td>$p$ value</td>
</tr>
<tr>
<td>Intercept</td>
<td>1.3375</td>
<td>10.50</td>
<td>0.0000</td>
</tr>
<tr>
<td>ln (Shareholders)</td>
<td>0.1003</td>
<td>3.53</td>
<td>0.0004</td>
</tr>
<tr>
<td>Analysts</td>
<td>0.0145</td>
<td>3.48</td>
<td>0.0005</td>
</tr>
<tr>
<td>Variance</td>
<td>0.0463</td>
<td>2.50</td>
<td>0.0128</td>
</tr>
<tr>
<td>S&amp;P 500</td>
<td>0.3220</td>
<td>2.92</td>
<td>0.0036</td>
</tr>
<tr>
<td>Adjusted $R^2$</td>
<td>12.84%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Durbin-Watson Stat.</td>
<td>2.009</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table IV

Summary Statistics for the Difference between the Implied True Price and Midpoint

This table provides a comparison between the implied true price and the midpoint by decile and by tick size regime. The decile is formed on market capitalization. Given the parameter estimates $\tilde{\alpha}_b$, $\tilde{\alpha}_s$ and $\tilde{\beta}$, the implied true price $\tilde{S}_i$ at time $t_i$ where the $i$-th trade of the day has occurred is computed as

$$\tilde{S}_i = P(t_i, x_i) \exp \left( -\tilde{\alpha}_b \tanh(\tilde{\beta} x_i) D_{x_i>0} - \tilde{\alpha}_s \tanh(\tilde{\beta} x_i) D_{x_i<0} \right).$$

In this equation, $x_i$ is the signed volume of the trade transacted at the price $P(t_i, x_i)$. The two dummy variables are denoted by $D_{x_i>0}$ and $D_{x_i<0}$, respectively. For every trade at time $t_i$, the midpoint of the prevailing bid and ask quotes $M_i$ is compared with $\tilde{S}_i$ to ascertain which is closer to the real true price. The method used in the comparison is based on the second-order moments of price change.

The spread between these two prices is computed as $L_i = 2|\tilde{S}_i - M_i|$. To gauge the magnitude of $L_i$, we use the prevailing bid-ask spread $Q_i$ and compute a daily average ratio from $N$ pairs of $L_i$ and $Q_i$ values:

$$R_d \equiv \frac{1}{N} \sum_{i=1}^{N} \frac{L_i}{Q_i}.$$

In each panel, the column designated by “Valid” contains the numbers of valid estimation results that pass the consistency tests. The mean, standard deviation (Std), minimum (Min), median (Med) and maximum (Max) statistics for $R_d$ in percent are tabulated.
Panel A: Minimum Tick Size of 12.5 Cents (1/8 of a Dollar)

<table>
<thead>
<tr>
<th>Decile</th>
<th>Valid</th>
<th>Number</th>
<th>%</th>
<th>Mean</th>
<th>Std</th>
<th>Min</th>
<th>Med</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Largest</td>
<td>20,707</td>
<td>19,853</td>
<td>95.88</td>
<td>49.32</td>
<td>17.43</td>
<td>5.98</td>
<td>45.50</td>
<td>157.5</td>
</tr>
<tr>
<td>2</td>
<td>19,618</td>
<td>18,821</td>
<td>95.94</td>
<td>51.09</td>
<td>18.70</td>
<td>3.84</td>
<td>47.86</td>
<td>236.9</td>
</tr>
<tr>
<td>3</td>
<td>16,203</td>
<td>15,285</td>
<td>94.33</td>
<td>52.45</td>
<td>20.56</td>
<td>1.85</td>
<td>48.98</td>
<td>206.4</td>
</tr>
<tr>
<td>4</td>
<td>11,075</td>
<td>10,269</td>
<td>92.72</td>
<td>53.44</td>
<td>20.82</td>
<td>2.54</td>
<td>50.28</td>
<td>185.8</td>
</tr>
<tr>
<td>5</td>
<td>6,097</td>
<td>5,603</td>
<td>91.90</td>
<td>56.61</td>
<td>21.42</td>
<td>2.50</td>
<td>52.93</td>
<td>172.4</td>
</tr>
<tr>
<td>6</td>
<td>5,100</td>
<td>4,670</td>
<td>91.57</td>
<td>56.71</td>
<td>22.30</td>
<td>5.29</td>
<td>53.47</td>
<td>201.4</td>
</tr>
<tr>
<td>7</td>
<td>2,208</td>
<td>1,989</td>
<td>90.08</td>
<td>56.66</td>
<td>21.66</td>
<td>7.22</td>
<td>53.69</td>
<td>152.7</td>
</tr>
<tr>
<td>8</td>
<td>1,540</td>
<td>1,409</td>
<td>91.49</td>
<td>58.33</td>
<td>22.08</td>
<td>10.54</td>
<td>55.24</td>
<td>165.8</td>
</tr>
<tr>
<td>9</td>
<td>739</td>
<td>684</td>
<td>92.56</td>
<td>57.69</td>
<td>22.66</td>
<td>6.11</td>
<td>53.09</td>
<td>201.1</td>
</tr>
<tr>
<td>Smallest</td>
<td>239</td>
<td>218</td>
<td>91.21</td>
<td>58.97</td>
<td>23.54</td>
<td>15.60</td>
<td>55.71</td>
<td>179.9</td>
</tr>
<tr>
<td>Overall</td>
<td>83,526</td>
<td>78,801</td>
<td>94.34</td>
<td>52.29</td>
<td>19.61</td>
<td>1.85</td>
<td>48.84</td>
<td>236.9</td>
</tr>
</tbody>
</table>

Panel B: Minimum Tick Size of 6.25 Cents (1/16 of a Dollar)

<table>
<thead>
<tr>
<th>Decile</th>
<th>Valid</th>
<th>Number</th>
<th>%</th>
<th>Mean</th>
<th>Std</th>
<th>Min</th>
<th>Med</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Largest</td>
<td>22,140</td>
<td>18,547</td>
<td>83.77</td>
<td>53.38</td>
<td>14.65</td>
<td>8.30</td>
<td>50.69</td>
<td>151.7</td>
</tr>
<tr>
<td>2</td>
<td>21,621</td>
<td>19,033</td>
<td>88.03</td>
<td>56.86</td>
<td>16.29</td>
<td>9.88</td>
<td>54.12</td>
<td>167.1</td>
</tr>
<tr>
<td>3</td>
<td>19,421</td>
<td>16,550</td>
<td>85.22</td>
<td>57.91</td>
<td>18.00</td>
<td>6.32</td>
<td>55.22</td>
<td>211.2</td>
</tr>
<tr>
<td>4</td>
<td>14,827</td>
<td>12,662</td>
<td>85.40</td>
<td>59.69</td>
<td>19.35</td>
<td>2.89</td>
<td>56.90</td>
<td>242.6</td>
</tr>
<tr>
<td>5</td>
<td>9,417</td>
<td>7,768</td>
<td>82.49</td>
<td>64.67</td>
<td>21.09</td>
<td>6.10</td>
<td>61.59</td>
<td>285.1</td>
</tr>
<tr>
<td>6</td>
<td>8,974</td>
<td>7,643</td>
<td>85.17</td>
<td>64.15</td>
<td>20.59</td>
<td>6.89</td>
<td>61.46</td>
<td>302.6</td>
</tr>
<tr>
<td>7</td>
<td>3,764</td>
<td>3,123</td>
<td>82.97</td>
<td>63.75</td>
<td>21.41</td>
<td>6.49</td>
<td>61.54</td>
<td>178.9</td>
</tr>
<tr>
<td>8</td>
<td>3,092</td>
<td>2,578</td>
<td>83.38</td>
<td>63.64</td>
<td>20.98</td>
<td>7.19</td>
<td>60.66</td>
<td>245.8</td>
</tr>
<tr>
<td>9</td>
<td>1,507</td>
<td>1,242</td>
<td>82.42</td>
<td>65.43</td>
<td>19.46</td>
<td>18.31</td>
<td>62.89</td>
<td>185.6</td>
</tr>
<tr>
<td>Smallest</td>
<td>656</td>
<td>566</td>
<td>86.28</td>
<td>62.67</td>
<td>19.77</td>
<td>10.28</td>
<td>60.31</td>
<td>214.4</td>
</tr>
<tr>
<td>Overall</td>
<td>105,419</td>
<td>89,712</td>
<td>85.10</td>
<td>58.62</td>
<td>17.86</td>
<td>2.89</td>
<td>55.89</td>
<td>302.6</td>
</tr>
</tbody>
</table>
This table provides the statistics for transaction costs of stocks traded on the NYSE in 1997. The 1,748 sample stocks are sorted and grouped into deciles according to their market capitalizations as at December 31, 1996. The first decile comprises 175 largest market capitalization stocks.

The friction spread \( \tilde{F}(t, x) \) is computed from the trade record given the parameter estimates \( \tilde{\alpha}_b, \tilde{\alpha}_s \) and \( \tilde{\beta} \) as follows:

\[
\tilde{F}(t, x) = 2P(t, x) \left| 1 - \exp \left( -\tilde{\alpha}_b \tanh(\tilde{\beta}x)D_{x>0} - \tilde{\alpha}_s \tanh(\tilde{\beta}x)D_{x<0} \right) \right|.
\]

The transaction price is \( P(t, x) \) at time \( t \) and the signed volume is \( x \). Two dummy variables that indicate buyer- and seller-initiated trades are denoted by \( D_{x>0} \) and \( D_{x<0} \), respectively. For each NYSE stock, the daily volume-weighted friction spread is computed. To compare our trade-based friction spread with the quote-based measures, we compute the daily volume-weighted effective and quoted spreads.

For each decile, we compute the cross-sectional average for each trading day. We tabulate the following statistics in dollars per lot and percent before (Panel A) and after (Panel B) the tick size reduction: mean, standard deviation (Std), minimum (Min) and maximum (Max).
### Panel A: Before Tick Size Reduction (120 Trading Days)

<table>
<thead>
<tr>
<th></th>
<th>1st</th>
<th>2nd</th>
<th>3rd</th>
<th>4th</th>
<th>5th</th>
<th>6th</th>
<th>7th</th>
<th>8th</th>
<th>9th</th>
<th>10th</th>
</tr>
</thead>
<tbody>
<tr>
<td>Friction Mean ($)</td>
<td>12.01</td>
<td>12.41</td>
<td>12.46</td>
<td>12.57</td>
<td>12.98</td>
<td>13.06</td>
<td>13.00</td>
<td>13.45</td>
<td>12.97</td>
<td>12.11</td>
</tr>
<tr>
<td>Spread Std ($)</td>
<td>0.12</td>
<td>0.17</td>
<td>0.27</td>
<td>0.32</td>
<td>0.46</td>
<td>0.58</td>
<td>1.03</td>
<td>1.04</td>
<td>1.44</td>
<td>2.33</td>
</tr>
<tr>
<td>Mean (%)</td>
<td>0.23</td>
<td>0.30</td>
<td>0.39</td>
<td>0.47</td>
<td>0.52</td>
<td>0.66</td>
<td>0.77</td>
<td>0.96</td>
<td>1.11</td>
<td>2.50</td>
</tr>
<tr>
<td>Std (%)</td>
<td>0.01</td>
<td>0.01</td>
<td>0.02</td>
<td>0.02</td>
<td>0.04</td>
<td>0.06</td>
<td>0.13</td>
<td>0.24</td>
<td>0.39</td>
<td>1.43</td>
</tr>
<tr>
<td>Spread Std ($)</td>
<td>0.29</td>
<td>0.26</td>
<td>0.35</td>
<td>0.44</td>
<td>0.56</td>
<td>0.59</td>
<td>1.04</td>
<td>1.44</td>
<td>1.70</td>
<td>2.62</td>
</tr>
<tr>
<td>Mean (%)</td>
<td>0.23</td>
<td>0.31</td>
<td>0.41</td>
<td>0.49</td>
<td>0.55</td>
<td>0.69</td>
<td>0.80</td>
<td>1.01</td>
<td>1.18</td>
<td>2.77</td>
</tr>
<tr>
<td>Std (%)</td>
<td>0.01</td>
<td>0.01</td>
<td>0.02</td>
<td>0.02</td>
<td>0.05</td>
<td>0.05</td>
<td>0.12</td>
<td>0.23</td>
<td>0.37</td>
<td>1.59</td>
</tr>
<tr>
<td>Quoted Mean ($)</td>
<td>17.40</td>
<td>17.90</td>
<td>17.53</td>
<td>17.89</td>
<td>18.82</td>
<td>18.21</td>
<td>18.47</td>
<td>18.80</td>
<td>18.62</td>
<td>17.28</td>
</tr>
<tr>
<td>Spread Std ($)</td>
<td>0.37</td>
<td>0.33</td>
<td>0.37</td>
<td>0.44</td>
<td>0.67</td>
<td>0.72</td>
<td>1.29</td>
<td>1.73</td>
<td>2.02</td>
<td>4.27</td>
</tr>
<tr>
<td>Mean (%)</td>
<td>0.32</td>
<td>0.42</td>
<td>0.54</td>
<td>0.66</td>
<td>0.74</td>
<td>0.89</td>
<td>1.05</td>
<td>1.26</td>
<td>1.55</td>
<td>3.34</td>
</tr>
<tr>
<td>Std (%)</td>
<td>0.01</td>
<td>0.01</td>
<td>0.03</td>
<td>0.03</td>
<td>0.06</td>
<td>0.07</td>
<td>0.14</td>
<td>0.26</td>
<td>0.46</td>
<td>1.72</td>
</tr>
</tbody>
</table>

### Panel B: After Tick Size Reduction (133 Trading Days)

<table>
<thead>
<tr>
<th></th>
<th>1st</th>
<th>2nd</th>
<th>3rd</th>
<th>4th</th>
<th>5th</th>
<th>6th</th>
<th>7th</th>
<th>8th</th>
<th>9th</th>
<th>10th</th>
</tr>
</thead>
<tbody>
<tr>
<td>Friction Mean ($)</td>
<td>7.16</td>
<td>7.70</td>
<td>7.75</td>
<td>8.00</td>
<td>8.98</td>
<td>8.69</td>
<td>9.07</td>
<td>9.17</td>
<td>9.57</td>
<td>9.07</td>
</tr>
<tr>
<td>Spread Std ($)</td>
<td>0.38</td>
<td>0.33</td>
<td>0.38</td>
<td>0.47</td>
<td>0.78</td>
<td>0.68</td>
<td>1.19</td>
<td>1.22</td>
<td>1.42</td>
<td>2.25</td>
</tr>
<tr>
<td>Mean (%)</td>
<td>0.13</td>
<td>0.17</td>
<td>0.23</td>
<td>0.27</td>
<td>0.31</td>
<td>0.37</td>
<td>0.45</td>
<td>0.54</td>
<td>0.64</td>
<td>1.27</td>
</tr>
<tr>
<td>Std (%)</td>
<td>0.01</td>
<td>0.01</td>
<td>0.03</td>
<td>0.02</td>
<td>0.03</td>
<td>0.06</td>
<td>0.11</td>
<td>0.13</td>
<td>0.22</td>
<td>0.88</td>
</tr>
<tr>
<td>Effective Mean ($)</td>
<td>9.87</td>
<td>10.71</td>
<td>10.48</td>
<td>10.66</td>
<td>11.84</td>
<td>11.50</td>
<td>12.64</td>
<td>12.27</td>
<td>12.82</td>
<td>11.89</td>
</tr>
<tr>
<td>Spread Std ($)</td>
<td>1.00</td>
<td>0.86</td>
<td>0.80</td>
<td>0.75</td>
<td>2.04</td>
<td>1.16</td>
<td>1.69</td>
<td>1.82</td>
<td>1.83</td>
<td>2.47</td>
</tr>
<tr>
<td>Mean (%)</td>
<td>0.17</td>
<td>0.23</td>
<td>0.30</td>
<td>0.35</td>
<td>0.40</td>
<td>0.47</td>
<td>0.57</td>
<td>0.69</td>
<td>0.83</td>
<td>1.65</td>
</tr>
<tr>
<td>Std (%)</td>
<td>0.02</td>
<td>0.02</td>
<td>0.04</td>
<td>0.03</td>
<td>0.04</td>
<td>0.07</td>
<td>0.09</td>
<td>0.15</td>
<td>0.26</td>
<td>1.14</td>
</tr>
<tr>
<td>Quoted Mean ($)</td>
<td>13.59</td>
<td>14.28</td>
<td>13.79</td>
<td>14.10</td>
<td>15.48</td>
<td>14.81</td>
<td>16.17</td>
<td>15.60</td>
<td>16.59</td>
<td>15.79</td>
</tr>
<tr>
<td>Spread Std ($)</td>
<td>1.18</td>
<td>0.98</td>
<td>0.94</td>
<td>0.97</td>
<td>2.10</td>
<td>1.06</td>
<td>1.62</td>
<td>1.96</td>
<td>2.27</td>
<td>3.09</td>
</tr>
<tr>
<td>Mean (%)</td>
<td>0.23</td>
<td>0.31</td>
<td>0.39</td>
<td>0.46</td>
<td>0.52</td>
<td>0.60</td>
<td>0.72</td>
<td>0.85</td>
<td>1.07</td>
<td>2.10</td>
</tr>
<tr>
<td>Std (%)</td>
<td>0.02</td>
<td>0.03</td>
<td>0.05</td>
<td>0.04</td>
<td>0.05</td>
<td>0.08</td>
<td>0.11</td>
<td>0.17</td>
<td>0.31</td>
<td>1.45</td>
</tr>
</tbody>
</table>
This graph illustrates the transaction price $P(t, x)$ as a function of signed volume $x$ at a given time $t$. Our model of transaction price is

$$P(t, x) = S(t) \exp \left( \frac{\alpha}{2} \tanh(\beta x) \right).$$

In this illustration, the true price $S(t)$ is fixed at ten dollars. The price impact parameter $\alpha$ is set at a large value of 10%. The scaling parameter $\beta$ determines the rate at which the asymptotic values of $S(t) \exp(\pm \alpha/2)$ are attained for large trades. Broadly speaking, one could regard different values of $\beta$ as different functional forms or shapes, although the price impact function is still $\tanh(\beta x)$. 
The sample period is January 2, 1997 to December 31, 1997. As two examples of our sample of 1,748 stocks, the time series of the $\tilde{\beta}$ estimates for General Electric (GE) and Burlington Northern Santa Fe Company (BNI) are plotted. Different $\tilde{\beta}$ corresponds to different functional form or S-shaped curve. On most of the 253 trading days, the $\tilde{\beta}$ estimates are larger than one. Equivalently, $\ln(\tilde{\beta})$ is positive. GE is more liquid than BNI. All the $\tilde{\beta}$ estimates of GE are larger than one. The price impact of trading GE’s stock is less dependent on the trade size than BNI’s stock, which has several days with negative $\ln(\tilde{\beta})$. 
Figure 3. Time Series of Transaction Costs.

Transaction costs of stocks traded on the NYSE in 1997 are plotted. The friction spread \( \tilde{F}(t, x) \) is computed from the trade record given the parameter estimates \( \tilde{\alpha}_b, \tilde{\alpha}_s \) and \( \tilde{\beta} \) as follows:

\[
\tilde{F}(t, x) = 2P(t, x) \left| 1 - \exp \left( -\tilde{\alpha}_b \tanh(\tilde{\beta}x)D_{x>0} - \tilde{\alpha}_s \tanh(\tilde{\beta}x)D_{x<0} \right) \right|.
\]

The transaction price is \( P(t, x) \) at time \( t \) and the signed volume is \( x \). Two dummy variables that indicate buyer- and seller-initiated trades are denoted by \( D_{x>0} \) and \( D_{x<0} \), respectively. For each NYSE stock, the daily volume-weighted friction spread is computed. To compare our trade-based friction spread with the quote-based measures, we compute the daily volume-weighted effective and quoted spreads. Panel A shows the cross-sectional average transaction costs of 175 largest market capitalization stocks in the first decile. Panel B displays the cross-sectional average transaction costs of 175 small market capitalization stocks in the ninth decile.