Available online at www.sciencedirect.com
ScienceDirect
Journal of Empirical Finance 15 (2008) 1 – 16
Journal of EMPIRICAL FINANCE
www.elsevier.com/locate/jempfin

A functional approach to the price impact of stock trades and the implied true price ☆

Roger D. Huang [a,*], Christopher Ting [b,1]

[a] Department of Finance, Mendoza College of Business, University of Notre Dame, Notre Dame, Indiana 46556, USA
[b] Lee Kong Chian School of Business, Singapore Management University, 50 Stamford Road, 178899 Singapore

Received 6 February 2006; received in revised form 27 November 2006; accepted 10 May 2007
Available online 18 May 2007

Abstract

We propose a functional approach to estimate the instantaneous price impact of a trade and to infer an implied true price. Our model expresses price impact as an S-shaped function of signed volume. It has two parameters, one for price impact and one for liquidity depth. The latter measures the differential impact of small and large trades. The price impact is instantaneous, that is, it occurs at the instant of trade execution. Our specification also permits the price impact of buys and sells to be asymmetric. We compute an implied true price from our model, and we find that it is closer than the quote midpoint to the unobservable true price. Our empirical analysis also shows that the effective spread, which is computed using the midpoint, has an upward bias, and that the implicit transaction costs may be lower than previous estimates.

JEL classification: G10; C51
Keywords: Price impact; Implied true price; Quote midpoint; Friction spread; Effective spread

1. Introduction

Transaction prices deviate from the true value of stocks because of trading frictions. The price impact of stock trades is a measure of trading friction; it is also an implicit cost for traders in the stock market. A theoretical justification for using price impact as a measure of implicit trading costs is provided by models such as Kyle (1985), Admati and Pfleiderer (1988), Holden and Subrahmanyam (1992), and Foster and Viswanathan (1993). Some studies have provided further insights on price impact. For example, Brennan and Subrahmanyam (1996) find that price impact is associated with cross-sectional variation in required rates of return on equity. Keim and Madhavan (1997) report that for certain type of trades, price impact is larger than explicit costs. Further, in view of recent market policy decisions that have had an impact on implicit trading costs, such as decimalization, an empirical analysis of these costs has become even more important.

☆ We are grateful to two anonymous referees, Joel Hasbrouck, K.C. Tsui, Liuren Wu, and seminar participants at Singapore Management University for their comments.
* Corresponding author. Tel.: +1 574 631 6370.
E-mail addresses: roger.huang.31@nd.edu (R.D. Huang), christopherting@smu.edu.sg (C. Ting).
[1] Tel.: +65 6828 0364.

0927-5398/$ - see front matter
doi:10.1016/j.jempfin.2007.05.001

This paper offers a new functional approach to modeling price impact. We present evidence that our model provides an improved measure of price impact, and our empirical analysis suggests that implicit trading costs may be lower than previously thought. We use our model to infer the unobservable true price. This true price is the "unperturbed price," the price that would have prevailed had the trade not occurred. The observed trade price is different from the true price, and its deviation from the true price is the instantaneous price impact. In our model, price impact is instantaneous because it occurs at the instant a trade is executed, and the true price fully reflects all public information at that instant.

Under the assumption that trades contribute to price discovery, we show that it is suitable to model instantaneous price impact as an S-shaped hyperbolic tangent function of signed volume with two parameters. The first is the price impact parameter, which is a measure of overall deviation of the trade price from the true price. The second parameter corresponds to the notion of market depth enunciated by Black (1971) and Kyle (1985). When this liquidity depth parameter is small, the market has little depth and price impact increases rapidly as trade size increases. A liquid market is characterized by a small price impact parameter and a large liquidity depth parameter.

The functional approach allows us to obtain a parametric estimate of the unobservable true price, which we refer to as the implied true price. It provides an alternative to the quote midpoint as an estimate of the true price. In empirical market microstructure analyses, the quote midpoint is often assumed to be an unbiased estimator of the unobservable true price. However, there is reason to believe that the midpoint may not accurately reflect the true price. In a recent paper, Goettler et al. (2005) construct a dynamic limit-order market with strategic traders. Both their theoretical model and their simulation results suggest that, in the presence of a trade, the midpoint is not a good estimate of the true price. Moreover, Amihud and Mendelson (1980), Ho and Stoll (1981), and others have shown that the midpoint may deviate from the true price due to market makers' inventory costs. Market makers may adjust their spreads to keep their inventories from deviating too much from an optimal level, and these adjustments may cause the midpoint to differ from the true price. The midpoint may also be inaccurate because the quotes are stale or because there are recording errors in the trade and quote data (Peterson and Sirri, 2003). By contrast, our implied true price is based entirely on transaction prices, and it is updated at the instant of each trade.

Our most important empirical result is that the implied true price is closer to the unobservable true price than the midpoint is, and the difference between the implied true price and the midpoint is both statistically and economically significant. One may infer from this result that estimates of implicit trading costs are more accurate if the implied true price, rather than the quote midpoint, is used as the reference price. We refer to the spread computed using the implied true price as friction spread, and we find that it is smaller than the effective spread, which is computed using the quote midpoint. Thus, it appears that the effective spread has an upward bias, and that implicit transaction costs may be lower than previous estimates.

The rest of the paper is organized as follows. The next section presents the basic model. The strong assumption that buyer- and seller-initiated trades are equally likely is relaxed in Section 3. In Section 4, we describe our sample and the data used in the estimations. Section 5 documents the estimation results. In Section 6, we compare the deviations of our implied true price and the quote midpoint from the true price, and in addition, we compare the friction and effective spreads. In Section 7, we summarize and conclude our paper.

2. The basic model of price impact and implied true price: symmetric price impact for buys and sells

In the basic model, we make the assumptions that trade price has a role in price discovery, and that buyer- and seller-initiated trades are equally likely. We propose that an S-shaped hyperbolic tangent function has characteristics that make it well suited for capturing the instantaneous price impact of a trade.

2.1. The relation between transaction price and true price

The key characteristics of a trade are transaction time, share volume, trade sign, and transaction price. In a market where liquidity is provided by designated market makers, the trade sign is positive when investors buy from them at the offer price and negative when investors sell to them at the bid price. On a fully automated exchange, where traders use market orders to ensure an immediate execution, the trade sign is positive for a market buy order and negative for a market sell order.

If there were no trading frictions, the true price $S(t)$ would also be the transaction price. But there are trading frictions, and trade price deviates from $S(t)$. This deviation is the instantaneous price impact of a trade, and we propose that it is a function of signed volume. Thus, the transaction price $P(t, x)$ has two components, the true price and the price impact $\varphi(x)$:

$$\ln P(t,x) = \ln S(t) + \varphi(x). \tag{1}$$

This structure is commonly used for modeling transaction prices in the market microstructure literature. For example, in Hasbrouck (1993), $\varphi(x)$ corresponds to the notion of pricing error. We express Eq. (1) as

$$P(t,x) = S(t)\exp(\varphi(x)). \tag{2}$$

The functional form of $\varphi(x)$ must be consistent with market realities and with financial economics. First, we note that $S(t)$ reflects all public information, and it is not influenced by $\varphi(x)$. Even in the absence of trade, that is, $x=0$, the true price continues to change in response to events. We express this aspect of Eq. (2) as the no-trade condition:

$$\varphi(0) = 0. \tag{3}$$

Thus, when $x=0$, the true price $S(t)$ is the "unperturbed price" $P(t, 0)$, the price that would have prevailed had the trade not occurred. Second, transaction prices for buyer-initiated (seller-initiated) trades should be higher (lower) than the true price. If the price impact function is an odd function, that is, if $\varphi(-x)=-\varphi(x)$, then this condition will be met. Moreover, the odd function has the property that $\varphi(0)=0$, which is consistent with the no-trade condition. Third, $\varphi(x)$ should be bounded. This insures that when trade size is large, $P(t, x)$ does not deviate too much from the true price under normal market conditions.

Since the unobservable true price is a component of the transaction price, we can back out an estimate of the true price denoted by $S^*(t, x)$. If we have identified the price impact function, then we obtain

$$S^*(t,x) = P(t,x)\exp(-\varphi(x)). \tag{4}$$

We refer to $S^*(t, x)$ as the implied true price.

2.2. Symmetric functional forms and the parameters of instantaneous price impact

Both empirical and theoretical studies suggest that an S-shaped pricing function is appropriate. In empirical studies, Hausman et al. (1992), Madhavan and Smidt (1991), and Hasbrouck (1993) find that the price impact is concave with respect to trade size (the absolute value of signed volume $|x|$); that is, it is concave for positive x and convex for negative x.

As for theory, Bhattacharya and Spiegel (1991) find that within the classical Walrasian setting, noisy rational expectations equilibria generally have feasible aggregate demand curves that are S-shaped. Spiegel and Subrahmanyam (2000) confirm that pricing functions are continuous and non-linear. Their theoretical results and numerical simulations both suggest an S-shaped pricing function. A key feature of their function is that it is bounded as trade size becomes very large.

In Proposition 1, we offer a suitable S-shaped price impact function.

Proposition 1. *Let α be a non-negative parameter, and let β be a strictly positive parameter. If it is the case that each trade with signed volume x and transaction price contributes to discovery of the true price S(t); that buyer- and seller-initiated trades are equally likely to occur; and that the condition $\varphi(0)=0$ is met, then $\alpha \tanh(\beta x)/2$ is a suitable functional form for modeling the instantaneous price impact.*

Proof. See Appendix A.

The first assumption, that the transaction price $P(t, x)$ plays a role in discovery of the unobservable true price, is a basic tenet of financial economics. The second assumption, that the trade sign is a fair game, is also common in the market microstructure literature. And the third assumption, that in the absence of trade, the true price reflects all public information, is common in many research studies as well. We emphasize that the proposition is independent of market design. The functional form is applicable to pure limit-order markets, exchanges with designated market makers, and hybrid markets such as the NYSE and NASDAQ.

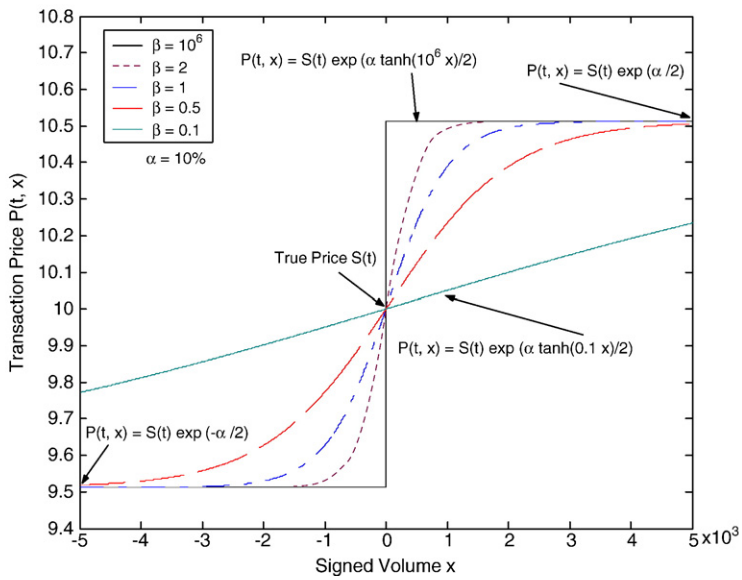

Fig. 1. A family of S-shaped curves. This graph illustrates transaction price $P(t, x)$ as a function of signed volume x at intraday time t. Transaction price is $P(t,x)=S(t)\exp\left(\frac{\alpha}{2}\tanh(\beta x)\right)$. In this illustration, the true price $S(t)$ is fixed at ten dollars. The price impact parameter α is set at 10%. The liquidity depth parameter β determines the rate at which the transaction price approaches the asymptotes, $S(t)\exp\left(\pm\frac{\alpha}{2}\right)$, as trade size increases.

In view of Proposition 1, we express the price impact function $\varphi(x)$ as a hyperbolic tangent function characterized by parameters α and β:

$$\varphi(x)=\frac{\alpha}{2}\tanh(\beta x). \tag{5}$$

The hyperbolic tangent function is S-shaped, strictly increasing and it has the unique property that when trade size is large, it asymptotes to ± 1. Therefore, the bounds of transaction price are

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

which satisfy the economic requirement that transaction price be bounded. We note that our model of hyperbolic tangent function of signed volume satisfies the no-trade condition.

If the parameter α is small, then the bounds are tight, implying that investors are able to trade at a price near the true price $S(t)$. We refer to α as the price impact parameter that quantifies the spread width. The parameter β is the liquidity depth parameter. It is a measure of the rate at which price impact increases as trade size increases. If β is large, then the price impact of large trades is only a little larger than that of small trades; that is, the market is deep. Fig. 1 presents illustrations of price impact curves with different values of β, holding α constant. As β increases, the slopes of the price impact functions increase substantially in the area around $x=0$. Although α does not vary in our illustration, if α were larger (smaller), then the two asymptotes would be further apart (closer), which would indicate that the spread for the largest trades is wider (narrower).

2.3. The implied true price and the friction spread

Recall that the model is able to yield an estimate of the true price, which we call the implied true price. Having identified that $\varphi(x)$ is the hyperbolic tangent function, the implied true price is given by

$$S^{*}(t,x)=P(t,x)\exp\left(-\frac{\alpha}{2}\tanh(\beta x)\right). \tag{7}$$

Since both trade size and the transaction price are observable, we can back out an estimate of the true price once we have estimated α and β.

As discussed earlier, the price impact is a half-spread, and therefore, it is natural to define a spread using the implied true price as the reference price:

$$F(t,x)\equiv 2|P(t,x)-S^*(t,x)|. \tag{8}$$

We refer to $F(t, x)$ as the friction spread. Substituting Eq. (7) into Eq. (8), the friction spread is

$$F(t,x)=2P(t,x)|1-\exp\left(-\frac{\alpha}{2}\tanh(\beta x)\right)|. \tag{9}$$

The friction spread is a new measure of implicit trading costs. This measure is contemporaneous with a transaction because the implied true price $S^*(t, x)$ is updated instantaneously when a trade occurs at time t. In contrast, the effective spread uses the quote midpoint as the reference price, and there is a time lag between the most recent quote update and a transaction.

2.4. Econometric specification

To estimate the model, we assume that the unobservable true price $S(t)$ follows a geometric Brownian motion. Let the drift rate of the log of true price be μ, and let the diffusion coefficient be σ. The geometric Brownian motion is driven by the Weiner process $W(t)$ with $W(0)=0$:

$$\ln S(t)=\ln S(0)+\mu t+\sigma W(t). \tag{10}$$

Suppose a stock has N trades for a particular trading day. With each trade indexed by i, t_i is the intraday time at which the ith trade occurs; x_i is the signed volume for the ith trade; and $P(t_i, x_i)$ is the transaction price. From Eq. (1), we have

$$\ln P(t_i,x_i)=\ln S(t_i)+\varphi(x_i), \tag{11}$$

and from Eq. (10), we have

$$\ln S(t_i)=\ln S(0)+\mu t_i+\sigma W(t_i). \tag{12}$$

Although $S(0)$ is unobservable, this term can be eliminated by using two consecutive prices. The log of change in trade price is thus given by

$$\ln\left(\frac{P(t_i,x_i)}{P(t_{i-1},x_{i-1})}\right)=\varphi(x_i)-\varphi(x_{i-1})+\mu(t_i-t_{i-1})+\sigma\varepsilon_{t_i,t_{i-1}}. \tag{13}$$

In this equation, the Brownian increment is $\varepsilon_{t_i,ti-1}\equiv W(t_i)-W(t_{i-1})$, and $\varepsilon_{t_i,t_{i-1}}=\sqrt{t_i-t_{i-1}}\psi_i$, where ψ_i is a normally distributed random variable with mean=0 and variance=1.

For simplicity, we write the square root of the time interval between two consecutive trades as

$$\tau_i\equiv\sqrt{t_i-t_{i-1}}, \tag{14}$$

and we define the trade-to-trade price change as

$$r_i\equiv\frac{\ln P(t_i,x_i)-\ln P(t_{i-1},x_{i-1})}{\tau_i}. \tag{15}$$

In view of Proposition 1, we substitute $\varphi(x_i)=\alpha\tanh(\beta x_i)/2$ in Eq. (13) and then divide both sides of the equation by τ_i. It follows that

$$r_i=\frac{\alpha}{2}\left(\frac{\tanh(\beta x_i)-\tanh(\beta x_{i-1})}{\tau_i}\right)+\mu\tau_i+\sigma\psi_i. \tag{16}$$

The innovation of this specification for r_i is $\sigma\psi_i$, which has mean 0 and variance σ^2.

We note that when the liquidity depth parameter increases without bound,

$$\lim_{\beta\to\infty} \tanh(\beta x) = \operatorname{sign}(x), \tag{17}$$

where sign(x) is defined as +1 if $x>0$ and −1 if $x<0$. After substituting Eq. (17) in the numerator of the first term in the right hand side of Eq. (16), the numerator simplifies to $\operatorname{sign}(x_i)-\operatorname{sign}(x_{i-1})$. This special case is reminiscent of regressions of price change on signed unit trades that appear in spread decomposition models such as Glosten and Harris (1988), Huang and Stoll (1997), and Madhavan et al. (1997). The focus of these papers is on identifying the spread components, rather than on measuring the instantaneous price impact, which is the main thrust of this paper.

3. Asymmetric price impact and methodology to compare two proxies for the true price

In Proposition 1, we assume that not only are buyer- and seller-initiated trades equally likely, but also that the distributions of trade size for buys and sells are identical. In this section, we relax these strong assumptions and develop a more generic price impact function that reflects the asymmetry in the distributions.

3.1. The model when buyer-initiated and seller-initiated trades are not equally likely

Suppose a stock has N trades over a trading day. For each stock-day, the number of buys N_b is most probably different from the number of sells N_s. These two numbers can be summarized with a single statistic κ that ranges from −1 to 1 by expressing N_b/N as $(1+\kappa)/2$, and N_s/N as $(1-\kappa)/2$, since $N=N_b+N_s$. Thus, $(1+\kappa)/2$ is the probability of a buyer-initiated trade and $(1-\kappa)/2$ is that of a seller-initiated trade. We note that the basic model developed in Section 2 assumes that $\kappa=0$.

Even when κ is non-zero, the true price $S(t)$ is still the "unperturbed price," that is, the no-trade condition (Eq. (3)) must still hold; the transaction price $P(t, x)$ of a buyer-initiated trade is higher than that of a seller-initiated trade; and the price impact is bounded. In Proposition 2 below, we present an asymmetric price impact function of signed volume that is consistent with these characteristics of market realities and with financial economics. Further, the parameters of the asymmetric price impact function are shown to satisfy an additional condition that is consistent with κ and the average value of $\ln(P(t, x)/S(t))$ for each stock-day.

Proposition 2. *Let $\tilde{\alpha}_b$ and $\tilde{\alpha}_s$ be two different non-negative parameters, and let $\tilde{\beta}$ be a strictly positive parameter. Then, a functional form that permits different probabilities for buys and sells is*

$$\varphi(x) = \begin{cases} \tilde{\alpha}_b \tanh(\tilde{\beta}x) & \text{if } x>0 \\ \tilde{\alpha}_s \tanh(\tilde{\beta}x) & \text{if } x<0, \end{cases} \tag{18}$$

provided that the following consistency condition is satisfied:

$$\kappa = \frac{2\eta - \tilde{\alpha}_b J_b - \tilde{\alpha}_s J_s}{\tilde{\alpha}_b J_b - \tilde{\alpha}_s J_s}, \tag{19}$$

where η is the mean of $\ln(P(t, x)/S(t))$, J_b and J_s are the respective average values of $\tanh(\tilde{\beta}x)$ when $x>0$ and when $x<0$.

Proof. See Appendix B.

To see that Proposition 2 is a generalization of Proposition 1, consider the special case of Proposition 2 in which there is no asymmetry. In this case, $\kappa=0$, $\tilde{\alpha}_b=\tilde{\alpha}_s$, and $J_s=-J_b$. Substituting these values into Eq. (19), we obtain $\eta=0$, which is an assumption in Proposition 1. When κ is non-zero, price discovery is asymmetric because the bounds corresponding to Eq. (6) become asymmetric:

$$S(t)e^{-\tilde{\alpha}_s} < P(t,x) < S(t)e^{\tilde{\alpha}_b}. \tag{20}$$

In spite of this asymmetry, a common liquidity depth parameter $\tilde{\beta}$ determines the rate at which transaction prices approach the two bounds in Eq. (20) as trade size increases.[2]

[2] An even more general version would have different liquidity depth parameters for buy and sell trades. However, it is not possible to have two depth parameters in our derivation of the functional form in Appendix A. More importantly, the consistency condition Eq. (19) is not met in our empirical analyses when two different depth parameters are specified.

We note that κ and η are exogenous to our model. These two statistics are dependent on the distributions of the trade sign and of the signed volume only, and not on any price impact model. In our empirical analysis, the estimates for the three parameters, $\tilde{\alpha}_b$, $\tilde{\alpha}_s$, and $\tilde{\beta}$, must satisfy the consistency condition (Eq. (19)) for the estimation to be considered valid. Thus, this consistency condition provides a powerful test to examine the validity of each estimation result.

3.2. Econometrics of the implied true price and friction spread

Given the asymmetry characterized by κ, the implied true price $S^*(t, x)$ (Eq. (7)) and the friction spread $F(t, x)$ (Eq. (9)) become, respectively,

$$\tilde{S}^*(t,x) = P(t,x)\exp(-\tilde{\alpha}_b \tanh(\tilde{\beta}x)D_{x>0} - \tilde{\alpha}_s \tanh(\tilde{\beta}x)D_{x<0}), \tag{21}$$

and

$$\tilde{F}(t,x) = 2P(t,x)|1 - \exp(-\tilde{\alpha}_b \tanh(\tilde{\beta}x)D_{x>0} - \tilde{\alpha}_s \tanh(\tilde{\beta}x)D_{x<0})|, \tag{22}$$

where $D_{x>0}$ ($D_{x<0}$) is a dummy variable that equals one if $x>0$ ($x<0$) and is zero otherwise.

In estimating the parameters from transaction data, we need to take into account widely recognized microstructure phenomena. It is well known that high-frequency price changes are negatively auto-correlated due both to the bid-ask bounce (Roll, 1984; Harris, 1990b) and to price discreteness (Harris, 1990a). To control for these effects, we first estimate the AR(1)-adjusted price change, which is the residual y_i in the following regression:

$$r_i = \rho_0 + \rho_1 r_{i-1} + y_i, \tag{23}$$

where ρ_0 is the intercept and ρ_1 is the lag one auto-correlation coefficient. Then we estimate the following multivariate regression model:

$$y_i = \tilde{y}_0 + (\tilde{\alpha}_b \tanh(\tilde{\beta}x_i)D_{x_i>0} + \tilde{\alpha}_s \tanh(\tilde{\beta}x_i)D_{x_i<0} - \tilde{\alpha}_b \tanh(\tilde{\beta}x_{i-1})D_{x_{i-1}>0} - \tilde{\alpha}_s \tanh(\tilde{\beta}x_{i-1})D_{x_{i-1}<0})/\tau_i + \tilde{\mu}\tau_i + \tilde{e}_i, \tag{24}$$

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

Eq. (24) is a generalized version of Eq. (16). If there is no asymmetry, then $\kappa=0$ and price impact is symmetric: $\tilde{\alpha}_b=\alpha/2=\tilde{\alpha}_s$ and $\tilde{\beta}=\beta$. In this special case, the four terms in the parentheses simplify to $\frac{\alpha}{2}\tanh(\beta x_i)(D_{x_i>0} + D_{x_i<0}) - \frac{\alpha}{2}\tanh(\beta x_{i-1})(D_{x_{i-1}>0} + D_{x_{i-1}<0})$. Since buyer- and seller-initiated trades are mutually exclusive, $D_{x>0}+D_{x<0}=1$, and this expression reduces to $\frac{\alpha}{2}(\tanh(\beta x_i) - \tanh(\beta x_{i-1}))$. The specification for price change becomes

$$y_i = y_0 + \frac{\alpha}{2}\left(\frac{\tanh(\beta x_i) - \tanh(\beta x_{i-1})}{\tau_i}\right) + \mu\tau_i + e_i, \tag{25}$$

which is essentially the same as Eq. (16), the special case discussed in Section 2.4.

3.3. Parameter estimates for symmetric and asymmetric models

The following proposition summarizes the relations between corresponding parameters of our basic model and the asymmetric model.

Proposition 3. *When the magnitude of the asymmetry $|\kappa|$ is small, the relations between the corresponding parameters in Eqs.*(24) *and (25) are as follows*:

$$y_0 \approx \tilde{y}_0; \ \ \mu \approx \tilde{\mu}; \tag{26}$$

$$\beta \approx \tilde{\beta}; \ \ \alpha \approx \tilde{\alpha}_b + \tilde{\alpha}_s. \tag{27}$$

Proof. The mean of the dependent variable, y_i, is the same statistic irrespective of its econometric specification. Thus, the expected values for both sides of Eqs. (24) and (25) are approximately the same.

Proposition 3 is important for justifying the assumption of symmetry that is commonly made in the literature. If the requirements of this proposition are not met in market data, the assumption of symmetric price impact would cause previous estimates to be unreliable. In reality, when observations are pooled over a long sample period, there is likely to be a small asymmetry between the distributions of buyer- and seller-initiated trades. Since this asymmetry is small, that is, $|\kappa|\approx 0$ and $\eta\approx 0$, Proposition 3 applies. Consequently, we are assured that the results reported in previous research are not substantially distorted by the assumption that the price impact of buys and sells are symmetric.

3.4. Methodology for comparing the implied true price and the quote midpoint

The quote midpoint has been used extensively as a proxy for the unobservable true price. Our model yields an implied true price (Eq. (21)) as an alternative proxy for the true price. Therefore, it is of interest to compare the two proxies. The comparison is particularly interesting because estimation of the implied true price uses only trade data whereas the midpoint is calculated from quotes.

First, we begin by defining $c(t_i)$ as the estimate of the log of true price $z(t_i)\equiv \ln S(t_i)$ at the time of the ith transaction of the trading day,

$$c(t_i) = z(t_i) + \zeta(t_i), \tag{28}$$

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

Next, consider the trade-to-trade price change

$$u_i = c(t_i) - c(t_{i-1}) = \upsilon_i + \omega_i, \tag{29}$$

where $\upsilon_i = z(t_i) - z(t_{i-1})$ is the change in the true price and $\omega_i = \zeta(t_i) - \zeta(t_{i-1})$ is the change in the unobservable deviation. The second moment of Eq. (29) is

$$E[u_i^2] = E[\upsilon_i^2] + E[\omega_i^2] + 2E[\upsilon_i\omega_i]. \tag{30}$$

We assume that the change in true price is independent of the change in deviation. Then, $E[\upsilon_i\omega_i]=E[\upsilon_i]E[\omega_i]$. Since the true price is a martingale, $E[\upsilon_i]=0$, it follows that $E[\upsilon_i\omega_i]=0$, and Eq. (30) becomes $E[u_i^2]=E[\upsilon_i^2]+E[\omega_i^2]$.

We know that the true price is common to both the implied true price and the midpoint, and that only their deviations differ. Therefore, we use the following observation to compare the deviations of the two proxies from the true price. If a stochastic process X_i is statistically smaller than another process Y_i, that is, $X_i<Y_i$, then the second moments of the differences will have the same inequality: $E[(X_i-X_{i-1})^2]<E[(Y_i-Y_{i-1})^2]$. Thus, if the magnitude of the deviation $\zeta(t_i)$ is comparatively smaller, then the second moment of the change in deviation $E[\omega_i^2]$ is also smaller in comparison.

Table 1
Descriptive statistics

	Stock attributes			NYSE		Consolidated	
	Market capitalization (millions)	Shares issued (millions)	Price ($)	Number of trades	Volume traded (1000s)	Number of Trades	Volume Traded (1000s)
Mean	3488.4	134.9	49.99	25 816	61 610	41 984	72 478
Std	9787.3	400.7	81.71	40 467	114 579	85 609	138 604
Min	2.9	0.8	0.28	59	42	87	42
1st per	15.3	3.5	1.33	552	690	688	854
Median	791.0	39.1	26.21	11 782	21 191	15 712	25 681
99th per	46 232.6	1818.6	102.17	224 558	574 370	460 610	699 993
Max	162 789.9	9878.5	335.13	391 769	1 231 586	1 055 752	1 602 050

Our sample consists of 1748 common stocks listed on the NYSE. The sample period is 1997. In that year, there were 253 trading days, of which 120 fall within the 1/8 tick-size regime, and 133 fall within the 1/16 tick-size regime. We use December 1996 data from CRSP for market capitalization and shares outstanding. The trade prices are obtained from CRSP. Statistics for number of trades and trade volume are computed from TAQ data. The consolidated number of trades and volume traded are annual aggregates of all transactions that took place on the NYSE and other exchanges. Following standard research practice, we omit trades executed on other exchanges. The mean, standard deviation (Std), minimum (Min), first percentile (1st per), median, 99th percentile (99th per), and maximum (Max) are cross-sectional statistics.

Specifically, we compare the two second moments: $\sigma_S^2 = E[(\ln \tilde{S}_i^* - \ln \tilde{S}_{i-1}^*)^2]$ and $\sigma_M^2 = E[(\ln M_i - \ln M_{i-1})^2]$, where $\tilde{S}_i^*$ is our implied true price and M_i is the quote midpoint. If σ_S^2 is smaller (larger) than σ_M^2, then the implied true price is relatively closer to (farther from) the true price than the midpoint is.

Suppose the implied true price $\tilde{S}_i^*$ is a better proxy. But, even if that was true, the difference between it and the quote midpoint M_i might not be economically meaningful. To examine whether the difference is economically significant, we first define a dollar measure of the difference,

$$L_i \equiv 2|\tilde{S}_i^* - M_i|. \tag{31}$$

To gauge the economic magnitude of L_i, we then divide it by the prevailing bid-ask spread Q_i, and compute the average of the resulting ratio for a stock that has N trades on day d.

$$R_d \equiv \frac{1}{N} \sum_{i=1}^{N} \frac{L_i}{Q_i}. \tag{32}$$

If we find a large R_d, we will conclude that the difference between these two proxies is economically significant.

4. Data

Our sample consists of 1748 companies listed on the NYSE. The sample period is 1997, which includes two events that provide tests of our model's reliability. First, on June 24, 1997, the minimum tick size was reduced from 1/8 to 1/16 of a dollar. Earlier research has shown that quoted and effective spreads decline when tick size is reduced. If our model is correctly specified and if the estimates are reliable, then the price impact parameters should decline. Second, October 27, 1997 is referred to as the Black Monday of 1997. The Dow Jones Industrial Average Index fell 544 points (7.2%) on that day, and rebounded 337 points on the following day. These large market-wide price swings provide a good setting in which to examine the robustness of our results.

We obtain trade data from the TAQ database. The descriptive statistics in Table 1 show that our sample includes not only large market capitalization firms but also penny stocks, which tend to be illiquid. Throughout the sample period, the 253 trading days in 1997, the average number of trades per day on the NYSE is 25 816, which is approximately 102 trades per stock-day. Thus, we have an adequate number of observations to estimate the parameters for each stock-day.

We employ the tick test to infer whether a transaction is buyer- or seller-initiated. The tick test classifies an up-tick trade as buyer-initiated and a down-tick trade as seller-initiated.[3] Zero-tick trades are signed using the up- or down-tick of the most recent previous trade. Trades that cannot be signed by a previous trade on the same day are excluded from the sample. These unclassified zero-tick trades typically occur at the start of regular trading hours and at re-openings after trading halts. They account for only 1.61% of the 45.1 million trade records in our sample. Previous research has shown that tick-test rules are remarkably accurate for NYSE stocks; see Lee and Ready (1991), Ellis et al. (2000) and Finucane (2000). The reason for using the tick test is that it requires only trade records and more importantly, we want to compare our trade-based friction spread with the quote-based effective spread in estimating the implicit trading costs. By not using the quotes to sign the trades, we are assured that our approach is purely based on trades and that our friction spread is not directly influenced by quotes.

5. Estimation and empirical results

As discussed in Section 3.2, our estimations require two steps. First, we compute the AR(1)-adjusted price change y_i (Eq. (23)).[4] Then we use non-linear least squares regression to estimate the parameters in Eq. (24). The 1/8 and 1/16 tick-size sub-periods are analyzed separately. Although there are seven parameters in Eqs. (23) and (24), only three, $\tilde{\alpha}_b$,

[3] An up (down) tick occurs when the current transaction price is larger (less) than the previous transaction price. A zero tick occurs when there is no price change.

[4] We have also performed the analysis without the AR(1)-adjustment. As anticipated, the sum of the two estimates for the price impact parameters is marginally larger than the AR(1) adjusted sum, and the regression residual is characterized by negative serial correlation. As no additional insight is gained from this exercise, we do not report the results.

Table 2
Estimates of the price impact parameters and the average price impact

Decile	Largest	2nd	3rd	4th	5th	6th	7th	8th	9th	10th
Panel A: January 2, 1997 to June 23, 1997 (120 trading days) sample period (before tick-size reduction)										
Mean buy $\tilde{\alpha}_b$ parameters in %	0.116	0.153	0.200	0.251	0.267	0.340	0.387	0.485	0.566	1.247
Std in %	0.367	0.377	0.221	0.972	0.317	0.755	0.344	0.472	0.505	1.301
Mean sell $\tilde{\alpha}_s$ parameters in %	0.115	0.151	0.201	0.236	0.263	0.344	0.397	0.469	0.591	1.405
Std in %	0.129	0.088	0.652	0.164	0.193	0.660	0.349	0.377	0.604	1.436
Mean liquidity depth parameter $\tilde{\beta}$	4.21	4.08	4.27	4.15	3.87	4.15	3.90	4.16	4.03	5.19
Std	3.02	3.15	8.56	5.84	3.84	4.01	3.59	4.36	3.84	9.65
Mean buy price impact in %	0.113	0.150	0.197	0.235	0.258	0.321	0.375	0.470	0.554	1.192
Std in %	0.064	0.083	0.179	0.144	0.175	0.219	0.311	0.419	0.505	1.294
Mean sell price impact in %	0.114	0.150	0.195	0.233	0.258	0.325	0.386	0.451	0.565	1.338
Std in %	0.065	0.083	0.153	0.143	0.177	0.222	0.319	0.358	0.455	1.450
Panel B: June 24, 1997 to December 31, 1997 (133 trading days) sample period (after tick-size reduction)										
Mean buy $\tilde{\alpha}_b$ parameters in %	0.062	0.091	0.119	0.142	0.186	0.193	0.252	0.294	0.409	0.691
Std in %	0.035	0.543	0.265	0.297	0.922	0.221	1.221	0.358	1.779	1.043
Mean sell $\tilde{\alpha}_s$ parameters in %	0.063	0.093	0.122	0.147	0.182	0.195	0.271	0.306	0.372	0.706
Std in %	0.046	0.490	0.450	0.666	0.676	0.197	1.261	0.594	0.727	1.039
Mean liquidity depth parameter $\tilde{\beta}$	3.12	3.32	3.37	3.25	3.24	3.14	3.08	3.20	2.91	3.64
Std	2.70	3.37	3.16	4.10	13.17	4.57	7.08	6.94	3.67	5.70
Mean buy price impact in %	0.062	0.085	0.113	0.132	0.149	0.176	0.212	0.265	0.320	0.632
Std in %	0.035	0.052	0.180	0.092	0.109	0.154	0.192	0.281	0.379	1.030
Mean sell price impact in %	0.063	0.086	0.113	0.132	0.152	0.178	0.219	0.264	0.318	0.609
Std in %	0.036	0.051	0.188	0.093	0.115	0.153	0.209	0.318	0.387	0.935

This table reports the statistics for the three parameter estimates of the asymmetric model (Eq. (24)), and for the price impact of an average size trade calculated with those estimates. We present the mean and standard deviation (Std) for our sample of 1748 NYSE common stocks, which are grouped into market capitalization deciles.

$\tilde{\alpha}_s$ and $\tilde{\beta}$, are needed for estimating the implied true price and the friction spread. To be included in the estimations, a stock-day must have at least 50 trades.

In total, we perform 190 119 regressions. More than 99.5% of the regressions produce estimated values for $\tilde{\alpha}_b$, $\tilde{\alpha}_s$, and $\tilde{\beta}$ that satisfy the consistency condition specified in Proposition 2 (Eq. (19)). The average adjusted R^2 value for the entire sample is 54.7%, with a standard deviation of 20.8%. The average F values are in the hundreds, and thus the parameter estimates are jointly non-zero. We have also performed an extensive analysis to verify that our estimation results are robust to serial correlation in signed volume (in round lots).[5] All these results indicate that the hyperbolic tangent function is well suited for capturing the instantaneous price impact.

Table 2 presents the estimation results by market capitalization decile and by tick-size regime. The mean price impact parameters, $\tilde{\alpha}_b$ and $\tilde{\alpha}_s$, decrease monotonically with market capitalization. This pattern is observed in both tick-size periods (Panels A and B). The average estimates of $\tilde{\beta}$ are quite stable across firm size deciles. However, the distribution of $\tilde{\beta}$ estimates is highly skewed, with a median of approximately 1.01, compared to the smallest mean of 2.91 (9th decile, 1/16 tick-size period).[6] As one would expect, all parameters decline after the tick-size reduction. The decline is especially large for $\tilde{\alpha}_b$ and $\tilde{\alpha}_s$. The mean liquidity depth parameter $\tilde{\beta}$ also declines from greater than 4.0 during the 1/8 tick-size period (Panel A) to approximately 3.2 during the 1/16 tick-size period (Panel B). These findings, that the price impact of stock trades declines as firm size increases and when tick size is reduced, are consistent with results reported in previous studies. See Goldstein and Kavajecz (2000), Chung and Chuwonganant (2004), Mackinnon and Nemiroff (2004), and Ronen and Weaver (2001), among many others.

We find that the price impact of buys is greater than that for sells. Of the 1748 sample stocks, 1561 have at least one day with valid estimation results that satisfy the consistency condition (Eq. (19)). Among these 1561 stocks, 924 have more days on which $\tilde{\alpha}_b > \tilde{\alpha}_s$, 91 have an approximately equal number of days, and the remaining 546 stocks have fewer

[5] The results of these consistency and robustness analyses are available from the authors.

[6] In the post-tick-size reduction sub-period (Panel B), the standard deviation of $\tilde{\beta}$ for the fifth decile is large due to two very large stock-day estimates. When these two outliers are excluded, the standard deviation declines to 3.56 (based on 9352 stock-days).

days. Therefore, for 59.19% of the stocks with valid estimates, buyer-initiated trades have a larger price impact on more than 50% of the trading days in 1997. The χ^2 statistic for $\tilde{\alpha}_b > \tilde{\alpha}_s$ is 97.2, which is statistically significant. However, the difference $\tilde{\alpha}_b - \tilde{\alpha}_s$ is small relative to $\tilde{\alpha}_b + \tilde{\alpha}_s$. In sum, although the instantaneous price impact of buyer-initiated trades is larger than that of seller-initiated trades, the difference between them is not economically significant.

In the lower section of each panel in Table 2, we show an average price impact for each firm size decile. This average is computed using the parameter estimates in the upper section of the panel and the average trade size for that decile. In the 1/8 tick-size period, the average trade size is approximately 20 to 25 round lots. The average price impact for the largest firm size decile is approximately 0.11% of the stock price. The average price impact increases monotonically to more than 1.27% for the smallest firm size decile. In the 1/16 tick-size period, the average trade sizes, 17 to 22 round lots, are smaller than those in the pre-reduction period. The price impact values are approximately half as large as in the pre-reduction period, but they rise as firm size declines just as they do in the pre-reduction period.

6. Comparison of the two proxies for true price and implications for transaction costs

In this section, we compare the accuracy of our two proxies for true price, the implied true price $\tilde{S}_i^*$ and the quote midpoint M_i. In principle, the implicit transaction cost should be defined with reference to the true price. Since the true price is unobservable, the accuracy of any cost measure is determined by the accuracy of the proxy used as the reference price. Our results show that the implied true price is closer to the true price than the midpoint. Therefore, we conclude that the friction spread that uses the implied true price is a more accurate estimate of implicit transaction costs than the effective spread. Further, since the friction spread is smaller, it appears that the effective spread is an upwardly biased estimator.

Table 3
Comparison of the implied true price and the midpoint as proxies for the true price

Firm size decile	Valid	$\tilde{S}_i^*$ Closer ($\sigma_s^2 < \sigma_M^2$)		R_d in %				
		Number	%	Mean	Std	Min	Median	Max
Panel A: Before tick-size reduction (120 trading days)								
Largest	20,707	19,943	96.31	49.32	17.43	5.98	45.50	157.5
2nd	19,622	18,852	96.08	51.09	18.70	3.84	47.86	236.9
3rd	16,218	15,307	94.38	52.45	20.56	1.85	48.98	206.4
4th	11,088	10,286	92.77	53.44	20.82	2.54	50.28	185.8
5th	6103	5614	91.99	56.61	21.42	2.50	52.93	172.4
6th	5111	4686	91.68	56.71	22.30	5.29	53.47	201.4
7th	2217	1996	90.03	56.66	21.66	7.22	53.69	152.7
8th	1547	1411	91.21	58.33	22.08	10.54	55.24	165.8
9th	742	685	92.32	57.69	22.66	6.11	53.09	201.1
Smallest	239	218	91.21	58.97	23.54	15.60	55.71	179.9
Overall	83,594	78,998	94.50	52.29	19.61	1.85	48.84	236.9
Panel B: After tick-size reduction (133 trading days)								
Largest	22,141	18,961	85.64	53.38	14.65	8.30	50.69	151.7
2nd	21,644	19,168	88.56	56.86	16.29	9.88	54.12	167.1
3rd	19,451	16,649	85.59	57.91	18.00	6.32	55.22	211.2
4th	14,868	12,722	85.57	59.69	19.35	2.89	56.90	242.6
5th	9481	7824	82.52	64.67	21.09	6.10	61.59	285.1
6th	9030	7694	85.20	64.15	20.59	6.89	61.46	302.6
7th	3799	3155	83.05	63.75	21.41	6.49	61.54	178.9
8th	3108	2594	83.46	63.64	20.98	7.19	60.66	245.8
9th	1521	1253	82.38	65.43	19.46	18.31	62.89	185.6
Smallest	656	566	86.28	62.67	19.77	10.28	60.31	214.4
Overall	105,699	90,586	85.70	58.60	17.85	2.89	55.87	302.6

This table presents the statistics by firm size decile for the comparison of our implied true price $\tilde{S}_i^*$ with the midpoint M_i in estimating the true price. The second column shows the number of stock-day estimations that pass the model's consistency test (Valid). The third and forth columns show the numbers and percentages of estimations for which $\tilde{S}_i^*$ is closer than M_i to the true price ($\tilde{S}_i^*$ Closer). The remaining columns present the statistics for R_d (Eq. (32)), which measures the economic significance of the difference between the two proxies as a percentage of the bid-ask spread.

6.1. Comparison of the accuracy of the implied true price and the quote midpoint

Recall that in Section 3.4, we show that the second moment of the log of price change is a measure of the proxy's distance from the true price, which is a measure of the estimator's accuracy. To perform the comparative analysis, we select all stock-days that have valid parameter estimates, and we compute the second moments σ_S^2 and σ_M^2.

Table 3 presents the comparison statistics by firm size decile and tick-size regime. The results show that the implied true price is closer to the true price than the midpoint ($\sigma_S^2<\sigma_M^2$) for all firm size deciles in both sub-periods of 1997. In the pre-tick-size reduction period, the implied true price is closer in 94.50% of the 83 594 comparisons and in the post-tick-change period, it is closer in 85.70% of the 105 699 comparisons. For the full sample, we find that σ_S^2 is smaller than σ_M^2 in 89.59% of comparisons. Therefore, we conclude that the implied true price is closer to the true price and thus more accurate than the midpoint. To the best of our knowledge, this is the first time that empirical tests have shown an alternative proxy for the true price to be better than the midpoint. This finding is consistent with the simulation results reported by Goettler et al. (2005), who find that the midpoint is not a good proxy for the true price around the time a trade occurs.

We want to know whether the difference between these two proxies is economically significant. Accordingly, we compute the ratio R_d, which is the daily average difference between the two proxies as a percentage of the bid-ask spread (Eq. (32)). The summary statistics for R_d, which are reported in Table 3, suggest that the difference is economically significant. Over the entire sample period, R_d is 55.66% of the quoted spread, with a standard deviation of 18.68%; that is, the average difference between the implied true price and the midpoint is about three standard deviations from zero. To illustrate the dollar magnitude of the average difference, let us assume that the average quoted spread is \$15 per round lot. In this case, the average difference between the two estimates is \$4.17 per round lot, which is estimated by $\$15\times0.5566/2$.

Table 4
Comparison of volume-weighted friction, effective, and quoted spreads

		Largest	2nd	3rd	4th	5th	6th	7th	8th	9th	Smallest
Panel A: Before tick-size reduction (120 trading days)											
Friction spread	Mean (\$)	12.01	12.41	12.46	12.57	12.98	13.06	13.00	13.45	12.97	12.11
	Std (\$)	0.12	0.17	0.27	0.32	0.46	0.58	1.03	1.04	1.44	2.33
	Mean (%)	0.23	0.30	0.39	0.47	0.52	0.66	0.77	0.96	1.11	2.50
	Std (%)	0.01	0.01	0.02	0.02	0.04	0.06	0.13	0.24	0.39	1.43
Effective spread	Mean (\$)	12.35	12.93	13.04	13.19	13.76	13.80	13.74	14.52	14.02	13.59
	Std (\$)	0.29	0.26	0.35	0.44	0.56	0.59	1.04	1.44	1.70	2.62
	Mean (%)	0.23	0.31	0.41	0.49	0.55	0.69	0.80	1.01	1.18	2.77
	Std (%)	0.01	0.01	0.02	0.02	0.05	0.05	0.12	0.23	0.37	1.59
Quoted spread	Mean (\$)	17.40	17.90	17.53	17.89	18.82	18.21	18.47	18.80	18.62	17.28
	Std (\$)	0.37	0.33	0.37	0.44	0.67	0.72	1.29	1.73	2.02	4.27
	Mean (%)	0.32	0.42	0.54	0.66	0.74	0.89	1.05	1.26	1.55	3.34
	Std (%)	0.01	0.01	0.03	0.03	0.06	0.07	0.14	0.26	0.46	1.72
Panel B: After tick-size reduction (133 trading days)											
Friction spread	Mean (\$)	7.16	7.70	7.75	8.00	8.98	8.69	9.07	9.17	9.57	9.07
	Std (\$)	0.38	0.33	0.38	0.47	0.78	0.68	1.19	1.22	1.42	2.25
	Mean (%)	0.13	0.17	0.23	0.27	0.31	0.37	0.45	0.54	0.64	1.27
	Std (%)	0.01	0.01	0.03	0.02	0.03	0.06	0.11	0.13	0.22	0.88
Effective spread	Mean (\$)	9.87	10.71	10.48	10.66	11.84	11.50	12.64	12.27	12.82	11.89
	Std (\$)	1.00	0.86	0.80	0.75	2.04	1.16	1.69	1.82	1.83	2.47
	Mean (%)	0.17	0.23	0.30	0.35	0.40	0.47	0.57	0.69	0.83	1.65
	Std (%)	0.02	0.02	0.04	0.03	0.04	0.07	0.09	0.15	0.26	1.14
Quoted Spread	Mean (\$)	13.59	14.28	13.79	14.10	15.48	14.81	16.17	15.60	16.59	15.79
	Std (\$)	1.18	0.98	0.94	0.97	2.10	1.06	1.62	1.96	2.27	3.09
	Mean (%)	0.23	0.31	0.39	0.46	0.52	0.60	0.72	0.85	1.07	2.10
	Std (%)	0.02	0.03	0.05	0.04	0.05	0.08	0.11	0.17	0.31	1.45

The sample consists of 1748 stocks listed on the NYSE in 1997. The values in the table are cross-sectional statistics and are presented by firm size decile. The friction spread is computed using the implied true price, and the effective and quoted spreads are computed using the midpoint. The mean and standard deviation (Std) are reported in dollars per round lot (\$) and in percentages (%).

Our finding that the implied true price is a better estimate of true price than the midpoint is based on propositions in Sections 2 and 3 that the hyperbolic tangent function is appropriate for capturing the instantaneous price impact. For purposes of comparison, we examine an alternative S-shaped function, α sign $(x)|x|^b$, which also satisfies the no-trade condition. This unbounded function fails to fit the data well. The estimate of the parameter b is close to zero and is statistically insignificant, which suggests that price impact is independent of trade size. Furthermore, when we use α sign $(x)|x|^b$ as the price impact function, the inferred true price is farther from the true price than the midpoint. This example illustrates why a functional form must be bounded in order to estimate the true price reliably.

There are several reasons why our implied true price is a better proxy for the true price than the midpoint. As discussed in the introduction, inventory costs and stale quotes may cause the midpoint to deviate from the true price. A more intuitive reason is that our implied true price $\tilde{S}^*(t, x)$ is an estimate of the true price $S(t)$ at the instant a trade occurs. By contrast, the information content of the quote midpoint may lag changes in the true price. For example, when the true price is rising, more trades will occur at the current ask price, the next ask price and so on. In this case, the quote midpoint underestimates the true price. This intuition is consistent with the results reported by Goettler et al. (2005), who find that the midpoint is not a good estimate of true price at the time a trade occurs.

6.2. Comparison of transaction cost proxies: the friction spread and quote-based spreads

This subsection documents our comparison of transaction costs computed using the two proxies for true price: the friction spread, which is computed using the implied true price, and the quoted and effective spreads, which are computed using the prevailing quote midpoint.

The first step in our comparison is to compute daily volume-weighted averages of the three spread measures. Using these averages, we create three time series of 253 daily cross-sectional averages for each firm size decile. Summary statistics for the three measures of implicit transaction costs are given in Table 4. As before, Panel A (Panel B) presents statistics for the pre-(post) tick-size reduction. Consistent with previously reported results, the statistics indicate that large market capitalization stocks have lower percentage transaction costs. The volume-weighted mean values for all three measures of transaction costs increase monotonically as market capitalization decreases.

In the pre-tick-size reduction period, the average friction spreads range from \$12.01 to \$13.45 per round lot. This range overlaps with the range of \$12.35 to \$14.52 for the effective spreads. However, in the post-tick-size reduction period, the difference between the friction and effective spreads is significantly larger; the average friction spreads

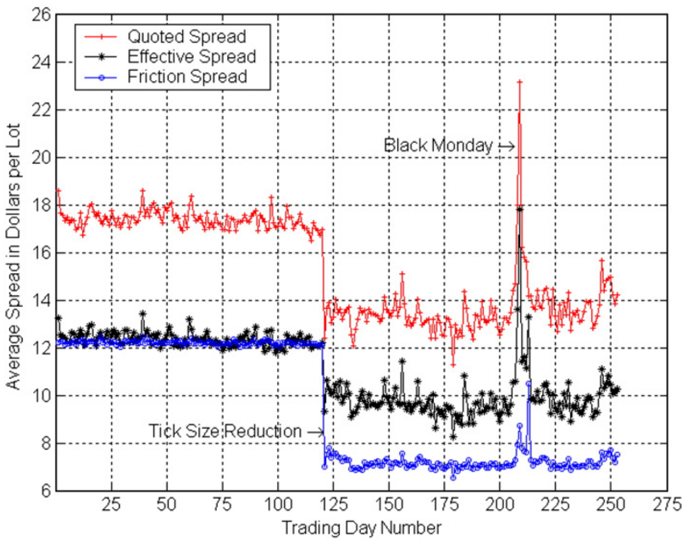

Fig. 2. Time series of volume-weighted spreads. The figure plots the average friction, effective, and quoted spreads for the 175 largest market capitalization firms (the first decile) in our sample of NYSE listed firms in 1997. During 1997, 120 trading days fall within the 1/8 tick-size regime, and 133 days fall within the 1/16 tick-size regime. The friction spread $\tilde{F}(t, x)$ is computed from trade records, using estimates of the parameters $\tilde{\alpha}_b$, $\tilde{\alpha}_s$ and $\tilde{\beta}$ as follows: $\tilde{F}(t,x) = 2P(t,x)|1 - \exp(-\tilde{\alpha}_b \tanh(\tilde{\beta}x)D_{x>0} - \tilde{\alpha}_s \tanh(\tilde{\beta}x)D_{x<0})|$, where, $P(t, x)$ is the transaction price at intraday time t with signed volume, and $D_{x>0}$ ($D_{x<0}$) is a dummy variable that equals one for buyer- (seller-) initiated trades.

range from \$7.16 to \$9.57 per round lot, whereas the average effective spreads range from \$9.87 to \$12.64 per round lot. The decline in each of the three spread measures from pre- to post-tick-size reduction is statistically significant at the 1% level. The price impact parameters after tick-size reduction are almost half the size of the pre-reduction parameters, and consequently the friction spread is smaller. This decline is clear in Fig. 2, which shows a time series plot of the daily cross-sectional volume-weighted averages for the 175 largest market capitalization stocks.

In sum, since the implied true price is a better proxy for the true price than the midpoint, the friction spread is a better measure of implicit transaction costs than the effective spread. Compared to the friction spread, the effective spread appears to have an upward bias, especially after the tick-size reduction. This result is consistent with findings reported by Peterson and Sirri (2003). They report that effective spreads overestimate the true execution costs and attribute the error to problems in inferring trade direction and in using inappropriate benchmark quotes.

7. Summary and conclusions

We show that the S-shaped hyperbolic tangent function of signed volume is well suited for capturing the instantaneous price impact of a trade. In addition to the price impact parameter, our model has another parameter that reflects the liquidity depth, which is the rate at which price impact increases as trade size increases. A liquid market has a small price impact parameter and a large liquidity depth parameter.

In the more general form of our model, the price impact of buyer- and seller-initiated trades can be different. The general model is estimated for 1748 stocks traded on the NYSE in 1997, which has two sub-periods defined by the reduction of the tick size. Our estimates of the parameters satisfy the model's consistency condition, and they indicate that the price impact of buyer-initiated trades is statistically greater than that of seller-initiated trades, although their difference is not economically significant.

With our model, the true price may be inferred from the transaction price and signed volume. We find that our implied true price is closer to the unobservable true price than the quote midpoint, and that the difference between these two estimators of the true price is statistically and economically significant. These results show that our model provides an alternative proxy for the true price that is more accurate than the midpoint. When we use our implied true price to compute a trade-based spread that we refer to as the friction spread, we find that it is smaller than the effective spread, which is computed with the less accurate midpoint. This finding is important because the effective spread is commonly used for estimating the implicit transaction cost, and it appears to be upwardly biased. Our results indicate that implicit trading costs may be lower than previous estimates that are based on the effective spread.

Appendix A. Proof of Proposition 1

Signed volume x is defined as $x=I|x|$, where $I\equiv\text{sign}(x)$ is the trade sign and $|x|$ is the volume. To construct a representation for $P(t, x)/S(t)$ in terms of signed volume, we assume that buyer- and seller-initiated trades are equally likely:

$$Pr(x>0) = \frac{1}{2} = Pr(x<0). \tag{A.1}$$

Under the probability measure Pr, and given that β is a constant, the expected value of $e^{\beta I|x|}$ is

$$E_0\left[e^{\beta I|x|}\right] = \frac{1}{2}\left(e^{\beta x} + e^{-\beta x}\right) = \cosh(\beta x). \tag{A.2}$$

Since $\cosh(\beta x)\neq 0$ for all x, we can divide both sides of Eq. (A.2) by $\cosh(\beta x)$:

$$E_0\left[\frac{e^{\beta I|x|}}{\cosh(\beta x)}\right] = E_0\left[\frac{e^{\beta x}}{\cosh(\beta x)}\right] = 1. \tag{A.3}$$

Thus, $e^{\beta x}/\cosh(\beta x)$ fluctuates symmetrically around one for all x, which implies that

$$E_0\left[\frac{e^{\beta x}}{\cosh(\beta x)}-1\right]=E_0\left[\frac{\sinh(\beta x)}{\cosh(\beta x)}\right]=E_0[\tanh(\beta x)]=0. \tag{A.4}$$

If the transaction price $P(t, x)$ contributes to the discovery of true price $S(t)$, then $\ln(P(t, x)/S(t))$ is equally likely to be positive or negative:

$$E_0\left[\ln\left(\frac{P(t,x)}{S(t)}\right)\right]=0. \tag{A.5}$$

Consequently, we obtain from Eqs. (A.4) and (A.5) that up to a non-negative parameter $\alpha/2$, $\tanh(\beta x)$ is a representation or a version of $\ln(P(t, x)/S(t))=\varphi(x)$ by Kolmogorov's fundamental theorem:

$$\varphi(x)=\frac{\alpha}{2}\tanh(\beta x)\quad\text{almost surely,} \tag{A.6}$$

since

$$E_0\left[\ln\left(\frac{P(t,x)}{S(t)}\right)\right]=0=\frac{\alpha}{2}E_0[\tanh(\beta x)]. \tag{A.7}$$

It is easy to verify that Eq. (A.6) satisfies the no-trade condition, $\varphi(0)=0$.

Appendix B. Proof of Proposition 2

When buys and sells are not equally likely, we use a number κ to characterize the asymmetry:

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

We note that κ depends on the frequencies of buyer- and seller-initiated trades, but not on trade size $|x|$. For emphasis, $E_x[\cdot]\equiv E[\cdot]$ denotes the average under the probability measure of signed volume x, which encompasses the measure Pr in Eq. (B.1). If there are more (few) buyer-initiated trades, then more values of $P(t, x)/S(t)$ are positive (negative). Under the probability measure of signed volume, we have

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

Since η is non-zero, the price impact of buyer- and that of seller-initiated trades are not equal. Nevertheless, both price impacts are bounded under normal market conditions:

$$-\tilde{\alpha}_s<\ln\left(\frac{P(t,x)}{S(t)}\right)<\tilde{\alpha}_b. \tag{B.3}$$

The formulation of price impact in Eq. (18) has a functional form that is asymmetric, strictly increasing, continuous and bounded as in Eq. (B.3), which can be expressed as,

$$\ln\left(\frac{P(t,x)}{S(t)}\right)=\tilde{\alpha}_b\tanh(\tilde{\beta}x)D_{x>0}+\tilde{\alpha}_s\tanh(\tilde{\beta}x)D_{x<0},\quad\text{almost surely.} \tag{B.4}$$

This functional form is a generalization of Eq. (5). Like Eq. (5), it satisfies the no-trade condition, $\varphi(0)=0$.

To derive the consistency condition, we compute

$$\begin{aligned} E_x\left[\ln\left(\frac{P(t,x)}{S(t)}\right)\right] &= E_x[\,\tilde{\alpha}_b\tanh(\tilde{\beta}x)D_{x>0},+\tilde{\alpha}_s\tanh(\tilde{\beta}x)D_{x<0}] \\ &= \tilde{\alpha}_b E_x[\tanh(\tilde{\beta}x)|x>0]Pr(x>0)+\tilde{\alpha}_s E_x[\tanh(\tilde{\beta}x)|x<0]Pr(x<0) \\ &= \frac{\tilde{\alpha}_b(1+\kappa)}{2}E_x[\tanh(\tilde{\beta}x)|x>0]+\frac{\tilde{\alpha}_s(1-\kappa)}{2}E_x[\tanh(\tilde{\beta}x)|x<0]. \end{aligned} \tag{B.5}$$

Let $J_b=E_x[\tanh(\tilde{\beta}x)|x>0]$ and let $J_s=E_x[\tanh(\tilde{\beta}x)|x<0]$. Then we have

$$\eta=\frac{\tilde{\alpha}_b(1+\kappa)}{2}J_b+\frac{\tilde{\alpha}_s(1-\kappa)}{2}J_s. \tag{B.6}$$

Rearranging the terms in Eq. (B.6), we obtain the consistency condition, Eq. (19).

References

Admati, Anat R., Pfleiderer, Paul, 1988. A theory of intraday patterns: volume and price variability. Review of Financial Studies 1, 3–40.

Amihud, Yakov, Mendelson, Haim, 1980. Dealership market: market-making with inventory. Journal of Financial Economics 8, 31–53.

Bhattacharya, Utpal, Spiegel, Matthew, 1991. Insiders, outsiders, and market breakdowns. Review of Financial Studies 4, 255–282.

Black, Fischer, 1971. Toward a fully automated exchange: part I. Financial Analysts Journal 29–35, 44 (27 July/Aug).

Brennan, Michael J., Subrahmanyam, Avanidhar, 1996. Market microstructure and asset pricing: on the compensation for illiquidity in stock returns. Journal of Financial Economics 41, 441–464.

Chung, Kee H., Chuwonganant, Chairat, 2004. Tick size, order handling rules and trading costs. Financial Management 33, 47–62.

Ellis, Katrina, Michaely, Roni, O'Hara, Maureen, 2000. The accuracy of trade classification rules: evidence from Nasdaq. Journal of Financial and Quantitative Analysis 35, 529–551.

Finucane, Thomas J., 2000. A direct test of methods for inferring trade direction from intraday data. Journal of Financial and Quantitative Analysis 35, 553–576.

Foster, Douglas F., Viswanathan, S., 1993. The effect of public information and competition on trading volume and price volatility. Review of Financial Studies 6, 23–56.

Glosten, Lawrence R., Harris, Lawrence E., 1988. Estimating the components of the bid/ask spread. Journal of Financial Economics 21, 123–142.

Goettler, Ronald, Parlour, Christine A., Rajan, Uday, 2005. Equilibrium in a dynamic limit order market. Journal of Finance 60, 2149–2192.

Goldstein, Michael A., Kavajecz, Kenneth A., 2000. Eighths, sixteenths, and market depth: changes in tick size and liquidity provision on the NYSE. Journal of Financial Economics 56, 125–149.

Harris, Lawrence, 1990a. Estimation of stock price variances and serial covariances from discrete observations. Journal of Financial and Quantitative Analysis 25, 291–306.

Harris, Lawrence, 1990b. Statistical properties of the Roll serial covariance bid-ask spread estimator. Journal of Finance 45, 579–590.

Hasbrouck, Joel, 1993. Assessing the quality of a security market: a new approach to transaction cost measurement. Review of Financial Studies 6, 192–212.

Hausman, Jerry A., Lo, Andrew W., Mackinlay, Craig, 1992. An ordered probit analysis of transaction stock prices. Journal of Financial Economics 31, 319–379.

Holden, W. Craig, Subrahmanyam, Avanidhar, 1992. Long-lived private information and imperfect competition. Journal of Finance 47, 247–270.

Ho, Thomas, Stoll, Hans R., 1981. Optimal dealer pricing under transactions and return uncertainty. Journal of Financial Economics 9, 47–73.

Huang, Roger D., Stoll, Hans R., 1997. The components of the bid-ask spread: a general approach. Review of Financial Studies 10, 995–1034.

Keim, Donald B., Madhavan, Ananth, 1997. Transactions costs and investment style: an inter-exchange analysis of institutional equity trades. Journal of Financial Economics 46, 265–292.

Kyle, Albert S., 1985. Continuous auctions and insider trading. Econometrica 53, 1315–1335.

Lee, Charles M.C., Ready, Mark J., 1991. Inferring trade direction from intraday data. Journal of Finance 46, 733–746.

MacKinnon, Greg, Nemiroff, Howard, 2004. Tick size and the returns to providing liquidity. International Review of Economics & Finance 13, 57–73.

Madhavan, Ananth, Smidt, Seymour, 1991. A Bayesian model of intraday specialist pricing. Journal of Financial Economics 30, 99–134.

Madhavan, Ananth, Richardson, Matthew, Roomans, Mark, 1997. Why do security prices change? A transaction-level analysis of NYSE stocks. Review of Financial Studies 10, 1035–1064.

Peterson, Mark, Sirri, Erik, 2003. Evaluation of the biases in execution cost estimation using trade and quote data. Journal of Financial Markets 6, 259–280.

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

Ronen, Tavy, Weaver, Daniel G., 2001. 'Teenies' anyone? Journal of Financial Markets 4, 231–260.

Spiegel, Matthew, Subrahmanyam, Avanidhar, 2000. Asymmetric information and news disclosure rules. Journal of Financial Intermediation 9, 363–403.