Full-information transaction costs*

Federico M. Bandi and Jeffrey R. Russell
GSB, University of Chicago

First draft: December 2003
This version: January 2006

Abstract

In a world with private information and learning on the part of the market participants, the (positive) difference between the *observed* transaction price of an asset and the corresponding *unobserved* full-information price (the price that reflects private and public information about the asset) represents an ideal measure of market quality. We call this difference "full-information transaction cost." We propose a simple and robust methodology to measure full-information transaction costs. Its simplicity is due to reliance on sample moments of *observed* high-frequency transaction price data. Its robustness hinges on the fact that the deviations of the observed transaction prices from the unobserved full-information prices can be imputed to fairly unrestricted operating (order-processing and inventory keeping) costs, adverse-selection costs, and learning in the marketplace. We estimate full-information transaction costs for a sample of S&P100 stocks and find that, while related to existing measures of transaction costs, they differ in ways predicted by their theoretical construct. Specifically, we show that our approach captures a large (asymmetric information - induced) component of market quality which is missing from existing measures, such as effective spreads and bid-ask spreads.

JEL classification: G12, G19.

Keywords: Asymmetric information, liquidity, market quality.

---

*We are especially grateful to John Cochrane, Robert Engle, Joel Hasbrouck, Jennifer Huang, and Ingrid Werner for their many useful comments. We also thank David Easley, Darren Roulstone, and the seminar participants at Queen Mary and Westfield College, Cass Business School, LSE (Department of Accounting and Finance), the conferences "Analysis of high-frequency data and market microstructure" (Taipei, Taiwan, December 15-16, 2003), "Econometric forecasting and high-frequency data analysis" (Singapore, May 7-8, 2004), the NBER/NSF Time Series Conference (Dallas, TX, September 19-20, 2004), the NYU conference in honor of the nobel prize winner Robert Engle (NYC, September 30, 2004), the China International Conference in Finance (Kunming, China, July 5-7, 2005), and the 2006 Winter Meetings of the Econometric Society (Boston, January 6-8), for discussions.

> *"...All estimates of value are noisy, so we can never know how far away price is from value. However, we might define an efficient market as one in which price is within a factor of 2 of value, i.e., the price is more than half of value and less than twice of value. The factor 2 is arbitrary, of course. Intuitively, though, it seems reasonable to me, in the light of sources of uncertainty about value and the strength of the forces tending to cause price to return to value. ... . Because value is not observable, it is possible for events that have no information content to affect price. For example, the addition of a stock to the S&P 500 index will cause some investors to buy it. Their buying will force the price up for a time. Information trading will force it back, but only gradually..."*
>
> *Fisher Black - "Noise" - 1986 Presidential Address to the American Finance Association.*

1 Introduction

Traditional measures of financial transaction costs, such as bid-ask spreads and effective spreads, quantify average deviations of transaction prices from a notional efficient price generally defined as the conditional expectation of future discounted cash flows given all publicly available information. An alternative literature on price discovery introduces the notion of a full-information price. The full-information price is the discounted liquidation value of the asset at a terminal date. More generally, it can be defined as the conditional expectation of future discounted cash flows given all private and public information. When only a subset of the agents are informed, the full-information price can differ from the efficient price. While formal definitions of market quality are elusive, arguably average deviations of transaction prices from full-information prices form a more complete measure of market quality than average deviations of transaction prices from efficient prices. We name these deviations "full-information transaction costs" or $FITC$s. Using a novel identification approach, we show how $FITC$s can be estimated consistently under general assumptions.

Our proposed measure of market quality is very much in the spirit of the Commodity Exchange Act of 1974 which defines a market as meeting the public interest if it satisfies three requirements: reliable price discovery, broad based price dissemination, and effective hedging against price risk. While the definition entertained in the 1974 Act is not quantitative in nature, it can serve as a guide as to what characteristics a market quality measure should posses. We capture both reliable price discovery and effective hedging against price risk attributes. The former requires that prices adjust to new information in a timely manner. The latter requires that markets provide insurance to liquidity traders by facilitating trades at, or near, efficient prices. Our method accounts for both departures of transaction prices from efficient prices, as in much existing work on transaction cost evaluation, and deviations of efficient prices from full-information prices. Specifically, we evaluate the positive distance between observed transaction prices and full-information prices. In the context of our measure, a high-quality market is one in which price discovery occurs quickly so that the efficient price is generally near the full-information price and transaction prices occur near their efficient values.

Our identification approach solely requires transaction price data. The intuition is simple. Consider a trading day. Write an observed transaction price as $\widetilde{p} = p + \eta$, where p denotes the full-information price and η denotes the corresponding *FITC*. If the full-information price is constant over the day (see, e.g., Admati and Pfleiderer (1988), Kyle (1985), Easley, Kiefer and O'Hara (1997)), then $\Delta\widetilde{p} = \Delta\eta$. Sample moments of the observed return data $\Delta\widetilde{p}$ can be employed to learn about moments of $\Delta\eta$ and, more importantly for our purposes, moments of the η's (given their assumed parametric structure). This procedure can be generalized to allow for empirically richer, ever evolving full-information price dynamics over the day. In this more general case, $\Delta\widetilde{p} = \Delta p + \Delta\eta$. If changes in the observed transaction prices are dominated by changes in the η's, then trivially $\Delta\widetilde{p} \approx \Delta\eta$. Specifically, provided non-negligible revisions to the full-information prices occur less frequently than revisions to the observed transaction prices, as implied by classical market microstructure theory, sample moments of the observed return data will still identify the moments of the η's at *high sampling frequencies.*

This paper focuses on the standard deviation of the η's, σ_η. We put the previous intuition to work to propose a measure of σ_η which is very easy to compute in that it is only based on sample moments of the observed intra-daily returns. The measure allows for a rich dependence structure in the deviations η's as well as for dependence between the full-information price and the deviations η's. We show consistency of our proposed measure for σ_η in an asymptotic framework explicitly designed to capture availability of high-frequency transaction prices. Using a suitable extension of Hasbrouck and Ho's model (Hasbrouck and Ho (1987)), we also show analytically that our measure performs very satisfactorily for durations commonly encountered in practise.

The closest approach to our approach is Hasbrouck (1993). Hasbrouck (1993) was, to the best of our knowledge, the first to focus on the standard deviation of the difference between transaction prices and a notional equilibrium price (defined as an efficient price) as a measure of market quality, as we do in this paper. Similarly to our framework, Hasbrouck also allowed for dependence between his notional equilibrium price and the deviations η's as well as for dependence in the η's. His important early work differs from ours along two main dimensions. First, identification in his framework hinges on the separation of permanent versus transitory price components as in the empirical macro literature (see, e.g., Stock and Watson (1988) for a review). We rely on a novel identification method designed to exploit the different stochastic orders of Δp and $\Delta \eta$. Second, his results provide a lower bound for σ_η. We estimate σ_η consistently and show finite sample validity of our asymptotic approximations using an empirically relevant price formation mechanism.

Our empirical work provides measurements of the (standard deviation of the) $FITC$s for a cross-section of S&P100 stocks. Consistent with our economic structure, we find the $FITC$s are more correlated with measures of private information like the PIN, turnover, and the number of analysts following the stock than traditional transaction cost measures such as effective spreads and bid-ask spreads. We also find that the asymmetric information or "learning" component in the estimated $FITC$s is substantial. Importantly, we show that the distance between efficient prices and full-information prices can be as large as the distance between transaction prices and efficient prices or larger. Therefore, measures of market quality which abstract from the difference between full-information prices and efficient prices have the potential to omit an important component of market quality.

The paper proceeds as follows. Section 2 discusses the price formation mechanism. In Section 3 we discuss estimation of σ_η. The section lays out the asymptotic and finite sample properties of our proposed measure. Section 4 provides σ_η estimates for a cross-section of S&P100 stocks and studies the determinants of the cross-sectional variation of the $FITC$s. Section 5 provides measurements of the distances between efficient prices and full-information prices for our sample of S&P100 stocks. Section 6 concludes. Technical details and proofs are in the Appendix.

2 The economics of high-frequency price formation

Our methodology assumes the existence of two equilibrium prices, namely the efficient price and the full-information price. Much of the literature analyzing agents' trading decisions under uncertainty relies on the notion of a terminal payoff of the security (see, for example, Kyle (1985) and Easley and O'Hara (1987)). In this context the distinction between these two equilibrium prices is easily understood. The terminal payoff is observed by the informed agents and is referred to as the full-information price. More generally, the full-information price can be regarded as the conditional expectation of future discounted cash flows given the private information of the informed agents. In a rational expectation setting, the uninformed agents, or noise traders, learn about the full-information price by observing order flow. Hence, market participants update their beliefs about the asset's value given publicly available information (order flow) and formulate equilibrium or "efficient" prices. The specific trading mechanism determines the difference between these two prices since it determines how trades affect prices, which, in turn, affect traders' order strategies. Kyle (1985) and Easley and O'Hara (1987), among others, provide explicit representations for the dynamics of the efficient price (relative to a full-information price) for batch auction and specialist markets, respectively. The learning of the market participants leads to efficient prices that eventually converge to full-information values. As O'Hara emphasizes in her discussion of trading and asymmetric information, "the eventual convergence of beliefs and thus of prices to full-information levels follows from standard Bayesian learning results" (O'Hara, 1995, page 64). Hence, the efficient price will differ from the full-information price in general.

The observed price, in general, will be neither the full-information price nor the efficient price. Deviations of the observed prices from the efficient values are induced by market frictions. The standard taxonomy in the literature postulates that two are the main economic forces behind market frictions: operating (i.e., order-processing and inventory-keeping) costs and adverse-selection costs. The order-processing component of frictions largely pertains to the service of "predictive immediacy" (Demsetz (1968)) or liquidity provision for which the market makers need to be compensated in equilibrium. Smidt (1971) suggests that the market makers are not just providers of liquidity but actively modify the spreads based on variation in their inventory levels (see, also, Garman (1976)). The idea is that the market makers wish not to be excessively exposed on just one side of the market and therefore adjust the spreads to offset positions that are overly long or short with respect to some desired inventory target. Much attention has recently been placed on the asymmetric information component of frictions. The market makers are bound to trade with investors that have superior information. Hence, the asymmetric information component of frictions is the profit that the dealers extract from the uninformed traders to obtain compensation for the expected losses to the informed traders (see Copeland and Galai (1983) and Glosten and Milgrom (1985)).

We now formalize these ideas. We consider a certain time period $\overline{h}$ (a trading day, for instance). Let t_i denote the arrival time of the i^{th} transaction. The counting function $N(t)$, which is defined over $t \in [0, \overline{h}]$, denotes the number of transactions occurred over the period $[0, t]$. The following prices are logarithmic prices. We write the *unobserved* efficient price $p^e_{t_i}$ as a function of the *unobserved* full-information price p_{t_i}, namely

$$p^e_{t_i} = p_{t_i} + \eta_i^{asy}, \tag{1}$$

where η_i^{asy} is a purely information-based component capturing differences between the efficient price and the full-information price. We write the *observed* price $\widetilde{p}_i$ corresponding to transaction i as

$$\widetilde{p}_i = p^e_{t_i} + \eta_i^{fri}, \tag{2}$$

where $p^e_{t_i}$ denotes the efficient price in Eq. (1) and η_i^{fri} denotes standard market frictions. Combining Eq. (1) and Eq. (2), we obtain that the observed transaction price can be expressed as

$$\widetilde{p}_i = p_{t_i} + \eta_i, \tag{3}$$

where p_{t_i} is the full-information price and $\eta_i = \eta_i^{asy} + \eta_i^{fri}$. We call the term η_i, i.e., the combined effect of frictions and departures of the efficient price from the full-information price, "market effect."

The dynamic properties of the full-information price p_{t_i} and the market effects η_i

are crucial for our measurement methodology. Economic theory sheds light on these properties. The full-information set, by definition, contains all information used by the market participants in their decisions to transact. Transactions, therefore, cannot carry any new information relative to the full-information set. Hence, the full-information price p_{t_i} is unaffected by the order flow. The dynamics of the full-information price, instead, are driven by the arrival of new information to the private information set. The efficient price, on the other hand, is updated as a function of order flow (i.e., each time a trade occurs). A cornerstone of market microstructure theory is that the uninformed agents learn about existing private information from observed order flow (see O'Hara (1995)). Since each trade carries information, meaningful revisions to the efficient price will be made regardless of the time interval between trade arrivals. Thus, the efficient price process (and equivalently η_i^{asy}) is naturally thought of as a process with non-negligible revisions associated with each transaction arrival time, no matter how close in time the transactions occur. Finally, the presence of separate prices for buyers and sellers and price discreteness, alone, suggest that the changes in the market friction components from trade to trade are discrete in nature and therefore so is η_i^{fri}.

In sum, our identification hinges on the basic premise that trades, and correspondingly updates to transaction prices, occur more frequently than updates to the full-information set and, therefore, the full-information price. In fact, most theoretical microstructure models assume that a single update to the full-information price occurs over a given trading period (i.e., Kyle (1985) and Easley and O'Hara (1987)). While our formal assumptions about the full-information price dynamics are more general than the assumptions used in these models, the important common theme is that, over short time horizons, the full-information price is relatively constant when compared to the market effects. The exact meaning of the words "relatively constant" is made clear below.

Assumption 1 (The full-information price.)

(1) *The full-information logarithmic price process p_t is a discontinuous semimartingale. Specifically,*

$$p_t = A_t + M_t + K_t, \tag{4}$$

where A_t *is a continuous finite variation component,* $M_t = \int_0^t \sigma_s dW_s$ *is a local martingale driven by the Brownian motion* W_t*, and* $K_t = \int_0^t \left(J_s dZ_s - \mu_j \lambda_s ds \right)$ *is an independent, compensated, jump process with* Z_t *denoting a counting process with finite intensity* λ_t *and* J_t *denoting a random jump size with mean* μ_j *and variance* σ_j^2*.*

(2) *The spot volatility process* σ_t *is càdlàg and bounded away from zero.*

The full-information price is observed only by the informed agents, but not by the rest of the market. Our assumed dynamic specification for this price is general. The models of Kyle (1985) and Easley and O'Hara (1987) assume that the liquidation value is fixed and observed by the informed agents thereby implying a constant full-information price. If $A_t = c$, $M_t = 0$, and $\lambda_t = 0$, then our Assumption 1 exactly replicates the static structure of the theoretical models. More generally, one can think of the full-information price as being random and interpreted as the conditional expectation of the discounted liquidation value given the private information set. If $A_t = c$, $M_t = 0$, and $\lambda_t > 0$, then the full-information price becomes a stochastic jump process. Alternatively, a continuously-evolving full-information price is obtained when we relax the assumption that $M_t = 0$ and set $\lambda_t = 0$. In all of these cases (and in the most general specification of Assumption 1) the expected full-information price variation becomes small over short time horizons.

Asset-pricing models require the semimartingale property of the price process from Assumption 1(2) as a necessary condition for the absence of arbitrage opportunities (Duffie (1990), for example). We also allow for the presence of stochastic volatility. Specifically, under Assumption 1(2), the volatility process can display long-memory properties, diurnal effects, and nonstationary dynamics. In addition, the innovations in returns can be correlated with the innovations in volatility. Hence, our specification can feature leverage effects.

We now formally state assumptions on the market effects η which are comprised of the efficient price deviations from the full-information price (η^{asy} in Eq. (1)) and the market frictions term η^{fri}. The assumptions reflect the non-negligible update features of this price component as discussed above.

Assumption 2 (The market effects.)

(1) *The market effects* η_i*'s are mean zero and covariance stationary with standard deviation* σ_η.

(2) *Their covariance structure is such that* $\mathbf{E}(\eta\eta_{-j}) = \theta_j \neq 0$ *for* $j = 1, ..., k < \infty$ *and* $\mathbf{E}(\eta\eta_{-j}) = 0$ *for* $j > k$.

(3) $\sum_{s=0}^{\infty} |w_s| < \infty$ *with* $w_s = \mathbf{E}\left[\left(\varepsilon_s\varepsilon_{s-j} - \mathbf{E}\left(\varepsilon\varepsilon_{-j}\right)\right)\left(\varepsilon\varepsilon_{-j} - \mathbf{E}\left(\varepsilon\varepsilon_{-j}\right)\right)\right]$, *where* $\varepsilon = \eta - \eta_{-1}$ *for* $j = 0, 1, ..., k < \infty$

The market effects are mean zero and stationary in transaction time (Assumption 2(1)).[1] The average market effect is zero so that transaction prices are equal to full-information prices on average but at any point in time there may be positive or negative deviations. It is natural to consider deviations of the efficient price from the full-information price η^{asy} as unconditionally mean zero in a rational expectation setting with learning on the part of the uninformed agents. If in addition the market frictions component η^{fri} is mean zero, also a natural assumption, then so is the entire market effect component, as assumed.

The dependence structure of the market effects is such that all covariances of order smaller than k can be different from zero while the covariances of order higher than k are equal to zero. The value of k and the signs of the covariances for values that are smaller than k is left unrestricted (Assumption 2(2)). Assumption 2 permits the deviations of the transaction prices from the efficient prices η^{fri} to be determined by virtually unrestricted order-processing costs (Tinic (1972), among others), inventory-holding costs (Amihud and Mendelson (1980) and Ho and Stoll (1981), *inter alia*), and adverse-selection costs (Copeland and Galai (1983) and Glosten and Milgrom (1985), among others). In effect, we accommodate temporal dependence in order flows, limit orders, and asymmetric information. As an example, transaction types sometimes repeat each other, i.e., sales and purchases cluster over brief periods of time. It is well known (see, for example, Garbade and Lieber (1977)) that a floor broker might split a large order into smaller orders, thereby inducing successive recorded sales and purchases. Similarly, limit orders might

---

[1] Stationarity in transaction time is a classical assumption in this literature (see, e.g., Hasbrouck, 1993).

remain in the market maker's book until there is a change in quotation. When a favorable change occurs, many limit orders might be satisfied at the same time. These transactions are typically recorded separately. As before, they induce several trades on the same side of the market and, consequently, serial correlation in the transaction prices. In addition Assumption 2 allows the deviations of the efficient price from the full-information price η^{asy} to be correlated with the unobservable full-information price process as required by asymmetric information and learning on the part of the market participants. Generally, transaction cost studies must assume that the market effects are uncorrelated with the equilibrium prices. It is worth noting that one of the nice features of our new identification approach is that it does not require this restrictive assumption. We now turn to our estimation procedure.

3 Measuring full-information transaction costs

Eq. (3) can be written in terms of returns as

$$\widetilde{r}_i = r_{t_i} + \varepsilon_i, \tag{5}$$

where $\widetilde{r}_i = \widetilde{p}_i - \widetilde{p}_{i-1}$ is the observed continuously-compounded return over the transaction interval (t_{i-1}, t_i), $r_{t_i} = p_{t_i} - p_{t_{i-1}}$ is the corresponding full-information continuously-compounded return, and $\varepsilon_i = \eta_i - \eta_{i-1}$ denotes market effects in the observed return process.

Lemma 1 expresses the square root of the second moment of the market effects in the price process σ_η as a function of the cross moments of the market effects in returns. Our estimator will be a consistent sample analogue of σ_η.

Lemma 1. *Write $\varepsilon = \eta - \eta_{-1}$. Then, under Assumptions 2(1) and 2(2),*

$$\sigma_\eta = \sqrt{\mathbf{E}(\eta^2)} = \sqrt{\left(\frac{1+k}{2}\right)\mathbf{E}(\varepsilon^2) + \sum_{s=0}^{k-1}(s+1)\mathbf{E}(\varepsilon\varepsilon_{-k+s})}. \tag{6}$$

Proof. *See Appendix.*

For clarity, we illustrate two subcases of the general result in Lemma 1. Assume $k = 1$, i.e., $\mathbf{E}(\eta\eta_{-1}) = \theta_1$. Hence,

$$\sigma_\eta = \sqrt{\mathbf{E}(\varepsilon^2) + \mathbf{E}(\varepsilon\varepsilon_{-1})}. \tag{7}$$

If $k = 2$, i.e., $\mathbf{E}(\eta\eta_{-2}) = \theta_2$, then

$$\sigma_\eta = \sqrt{\frac{3}{2}\mathbf{E}(\varepsilon^2) + 2\mathbf{E}(\varepsilon\varepsilon_{-1}) + \mathbf{E}(\varepsilon\varepsilon_{-2})}. \tag{8}$$

Provided the relevant moments of the market effects in returns can be consistently estimated using observables, Eq. (6) constitutes an expression that can be readily used to identify the second moment of the market effects in the price process. The availability of high-frequency price data offers us a unique way to do so.

More intuition about our identification procedure seems warranted. The market effects in the observed returns ε are $O_p(1)$, i.e., a stochastic process with a variance which does not vanish with increasing sampling frequencies. The full-information component of the observed returns is of order $O_p\left(\sqrt{\max_{1\leq i\leq N(\overline{h})} |t_i - t_{i-1}|}\right)$, where $\max_{1\leq i\leq N(\overline{h})} |t_i - t_{i-1}|$ is the maximum duration between price updates. In other words, the variance of this component of the observed returns does vanish as the sampling frequency increases. In light of these observations, variation in the market effects in the return process ε dominates variation in the full-information return component r at high sampling frequencies. This is simply a mathematical representation of the economic intuition developed in Section 2. Specifically, non-negligible revisions to the full-information price occur less frequently than non-negligible revisions to the market effects. Hence, sample moments of the *observed* high-frequency return data contain a negligible full-information return component and can be used to identify moments of the *unobserved* market effects.

Theorem 1 formalizes this idea. Our asymptotic design, which hinges on an increasing number of transactions (i.e., $N(\overline{h}) \to \infty$) over a fixed interval of time ($\overline{h}$), is meant to represent availability of a very large number of transactions over the time interval, not an endogenous increase in the number of trades possibly due to the state of the market. In the context of our asymptotic approximation, it is natural to regard the transaction arrival times as being deterministic.

Theorem 1. *Assume Assumptions 1 and 2 are satisfied. Given a sequence of trade arrival times such that max* $\left\{|t_{i+1} - t_i|, i = 1, ..., N(\overline{h})\right\} \to 0$ *as* $N(\overline{h}) \to \infty$, *we obtain*

$$\widehat{\sigma}_\eta = \sqrt{\left(\frac{k+1}{2}\right)\left(\frac{\sum_{i=1}^{N(\overline{h})}\widetilde{r}_i^2}{N(\overline{h})}\right) + \sum_{s=0}^{k-1}(s+1)\left(\frac{\sum_{i=k-s+1}^{N(\overline{h})}\widetilde{r}_i\widetilde{r}_{i-k+s}}{N(\overline{h})-k+s}\right)} \underset{N(\overline{h})\to\infty}{\overset{p}{\rightarrow}} \sigma_\eta. \quad (9)$$

Proof. *See Appendix.*

In what follows, we will use the convention of referring to estimates obtained by employing the estimator in Eq. (9) as *FITC*s. The estimator is local in nature and defined over a single, generically-specified, period $\overline{h}$. In this sense we can readily allow for a time-varying second moment of the market effects possibly induced by the convergence dynamics of the transaction price to the full-information price. Under the assumption that the properties of the η's extend to multiple periods (or when interested in the unconditional expectation of the time-varying second moment of the market effects), the simple summations over i (which is our index for transactions) in the definition of the estimator in Eq. (9) can be replaced by double summations over j, say, where j denotes the j^{th} period in the sample and, again, over i, where i denotes the i^{th} transaction during the generic j^{th} period. We use this procedure in what follows.

Some recent work has advocated the use of HAC-type estimators applied to high-frequency price data in order to estimate the quadratic variation of asset prices over a fixed time period such as a day (see Hansen and Lunde (2006) and Oomen (2005)) when market frictions play a role. These estimators focus on the volatility of the efficient price. On the contrary, our estimator focuses on estimation of the high-frequency variance associated with the market effects. While our estimator may appear similar to a HAC-type estimator, it is not in that the weights are derived specifically for estimating the variance of the market effects in the observed price process under our assumed correlation structure for the market effects.

3.1 Measuring the positive difference between transaction prices and full-information prices

The *FITC* measure is a standard deviation of departures of transaction prices from full-information prices. Alternatively, one might want to quantify the expected (positive) deviation of transaction prices from full-information levels. In this subsection we show that this expectation can be estimated from the *FITC*s under further assumptions. Specifically, we consider two alternative assumptions below.

Assumption 3. *Assume η is normal.*

Assumption 4. *Assume*

$$\eta_i = sQ_i \qquad \forall i = 1, ..., N(\overline{h}), \tag{10}$$

where Q_i is a random variable representing the direction (i.e., higher or lower) of the transaction price with respect to the full-information price and s is the full-information transaction cost. Specifically, assume Q_i can take on only two values, -1 and 1, with equal probabilities.

Under Assumption 3, $\mathbf{E}\left(|\widetilde{p} - p|\right) = 0.7979\sigma_\eta$. If Assumption 4 is satisfied, then $\mathbf{E}\left(|\widetilde{p} - p|\right) = \sigma_\eta = s$.

Corollary to Theorem 1. *i) Assume Assumptions 1, 2, and 3 are satisfied. Given a sequence of trade arrival times such that* $\max\left\{|t_{i+1} - t_i|, i = 1, ..., N(\overline{h})\right\} \to 0$ *as* $N(\overline{h}) \to \infty$, *we obtain*

$$0.7979\widehat{\sigma}_\eta \underset{N(\overline{h})\to\infty}{\overset{p}{\to}} \mathbf{E}\left(|\widetilde{p} - p|\right), \tag{11}$$

where $\widehat{\sigma}_\eta$ is defined in Eq. (9).

ii) Assume Assumptions 1, 2, and 4 are satisfied. Given a sequence of trade arrival times such that $\max\left\{|t_{i+1} - t_i|, i = 1, ..., N(\overline{h})\right\} \to 0$ *as* $N(\overline{h}) \to \infty$, *we obtain*

$$\widehat{\sigma}_\eta \underset{N(\overline{h})\to\infty}{\overset{p}{\to}} s = \mathbf{E}\left(|\widetilde{p} - p|\right), \tag{12}$$

where $\widehat{\sigma}_\eta$ is defined in Eq. (9).

Proof. *Immediate given Lemma 1, Theorem 1, and Assumptions 3 and 4.*

Assumption 4 is in the spirit of Roll's fundamental approach to effective transaction cost estimation (Roll (1984)). Choi et al. (1988) and Hasbrouck (1999, 2003), among others, provide interesting extensions of Roll's method. In Roll's model, r_{t_i} denotes the efficient return rather than the full-information return, $Q_i = 1$ corresponds to a buyer-initiated trade, and $Q_i = -1$ denotes a seller-initiated trade. Under *(i)* uncorrelatedness of the efficient return process, *(ii)* uncorrelatedness between the efficient return process and the order flows, and *(iii)* uncorrelatedness of the order flows, Roll shows that $Cov(\widetilde{r}, \widetilde{r}_{-1}) = -2s^2$. Hence, a consistent estimate of s is given by $2\sqrt{\widehat{\mathbf{E}}(\widetilde{r}, \widetilde{r}_{-1})}$. While our approach uses recorded asset returns to measure unobserved transaction costs as in Roll's approach, our definition of transaction costs is different from Roll's in that Roll's benchmark price is the efficient price. Furthermore, Assumptions *(i)* through (*iii*) were shown to be unnecessary in our framework.

3.2 The finite sample properties of the *FITC*s

Estimation of the *FITC*s requires the availability of high-frequency transaction price data. When the arrival times are not very frequent, there could be residual contaminations induced by the dynamics of the underlying full-information price process. Specifically, if the full-information price is unpredictable, i.e., $A_t = 0$, and changes in the market effects do not predict the full-information returns, i.e., $\mathbf{E}(r\varepsilon_{-s}) = 0 \ \forall s \geq 1$, we can write

$$\mathbf{E}\left(\widehat{\sigma}^2_\eta\right) = \left(\frac{k+1}{2}\right)\mathbf{E}(\varepsilon^2) + \underbrace{\left(\frac{k+1}{2}\right)\mathbf{E}\left(\frac{\sum_{i=1}^{N(\overline{h})} r_i^2}{N(\overline{h})}\right)}_{\alpha} + \sum_{s=0}^{k-1}(s+1)\mathbf{E}(\varepsilon\varepsilon_{-k+s}) + \tag{13}$$

$$\underbrace{+(k+1)\,\mathbf{E}\left(\frac{\sum_{i=1}^{N(\overline{h})} r_i\varepsilon_i}{N(\overline{h})}\right) + \sum_{s=0}^{k-1}(s+1)\mathbf{E}\left(\frac{\sum_{i=k-s+1}^{N(\overline{h})} \varepsilon_i r_{i-k+s}}{N(\overline{h}) - k + s}\right)}_{\beta}. \tag{14}$$

The finite sample bias induced by the full-information price dynamics is given by $\alpha + \beta$. The *FITC* estimates are consistent since $\alpha + \beta \to 0$ as $N(\overline{h}) \to \infty$. The relevant moments in the bias cannot be nonparametrically identified from the data. Any inference on the size of the finite sample bias must therefore be examined from the perspective of a given parametric model. In this spirit, we consider a generalization of the model proposed by Hasbrouck and Ho (1987) (see, also Beja and Goldman (1980) and Goldman and Beja (1979)). In the context of this model we can asses the potential bias associated with our estimator. Write

$$\widetilde{p}_t = p_t^e + \eta_t^{fri} \tag{15}$$

$$p_t^e - p_{t-1}^e = \theta(p_t - p_{t-1}^e) + u_t \tag{16}$$

$$p_t = p_{t-1} + \xi_t, \tag{17}$$

with ξ_t and u_t i.i.d. mean zero, and η_t^{fri} stationary, mean zero. Assume ξ_t, η_t^{fri} and u_t are uncorrelated.[2] Under this model, it is easy to show that

$$\alpha + \beta \approx \left(\frac{k+1}{2}\right)\sigma_\xi^2 - (k+1)(1-\theta)\sigma_\xi^2 + \sum_{s=0}^{k-1}(s+1)(1-\theta)^{k-s}\theta\sigma_\xi^2, \tag{18}$$

where σ_ξ^2 is the variance of the innovations in the full-information price.[3] For our sample of stocks, we estimate σ_ξ^2 using daily realized variance divided by the average number of transactions over a day. Realized variance is optimally sampled on the basis of the procedure proposed by Bandi and Russell (2003a,b). Given an estimate of σ_ξ^2, the size of the bias depends on the parameter θ controlling the speed of convergence of the efficient price to the full-information price. In our sample, for a typical stock, we find that values of θ between 0.1 and 0.9 are associated with a difference between bias-corrected $FITC$ and raw $FITC$ which is, as a percentage of raw $FITC$, between about 1% and about 30% in absolute value.

This potential finite sample bias highlights the difference between the nonparametric approach adopted in our work and an alternative parametric approach. Our consistent estimates promise to be roughly unbiased if the number of trades within a day is large, but are potentially biased for less frequently-traded assets. Alternatively, parametric models will likely yield inconsistent estimates in both finite samples and asymptotically unless they are correctly specified. While one could bias-correct the nonparametric estimates using bias estimates constructed from a realistic parametric model, such as the one above,

---

[2] It should be noted that in Hasbrouck and Ho (1987) p_t and p_t^e denote the efficient price and the mid-point of the bid-ask quotes, respectively.

[3] The model implies that

$$\eta^{asy} = -\sum_{j=0}^{\infty}(1-\theta)^{j+1}\xi_{t-j} + \sum_{j=0}^{\infty}(1-\theta)^j u_{t-j}.$$

Hence, $\eta^{asy} = MA(k=\infty)$. Empirically (see Section 4) we find a maximum $k = 15$. Thus, we regard the model as giving us a useful assessment of the types of biases that we face for our empirically-motivated choice of k.

here we prefer to remain completely nonparametric. We view the bias calculations above as being simply suggestive as to the size of the bias. Alternative models would obviously provide alternative results.

4 The *FITCs* of the S&P 100 stocks

The data consist of one month of high-frequency transaction prices for the stocks in the S&P 100 index. The prices were obtained from the TAQ data set for the month of February 2002. Our sample contains 93 NYSE stocks and 7 NASDAQ stocks. Transactions from the primary exchanges only are used. The data are filtered to remove any zeros.

The *FITC* estimates require a choice for k, the number of non-zero autocorrelations. In Fig. 1 we present the histograms of the t-ratios of several autocorrelations for the 100 stocks in our sample, i.e., $\sqrt{n}\widehat{\rho}_j$, with $j = 1, 2, 3, 5, 10$, and 15. The autocorrelation structure in the high-frequency transaction prices is significantly negative at lag one and quite negative at lag two. It is generally positive at lags higher than two but largely statistically insignificant at lags around 15 and higher. These features of the data, which are likely to be induced by bid-ask bounce effects at small lags and clustering in order-flows at higher lags, demonstrate the need to consider estimation procedures that are robust to deviations from a model of price determination that only allows for a negative first-order autocorrelation in the recorded stock return data (as in Roll's approach, for example). To accommodate non-zero high order autocorrelations, we set k in the *FITC* estimator in Eq. (9) equal to 15 for all stocks.

We begin by comparing the *FITC*s to a conventional measure of market quality which is meant to quantify deviations of transaction prices from efficient prices, namely *effective spread*. This measure is a natural benchmark to use since, like our measure, it uses trade-by-trade data.

Ignoring private information, Perold (1988) suggested that an ideal measure of the execution cost of a trade should be based on the comparison between the trade price for an investor's order and the efficient price prevailing at the time of the trading decision. Although individual investors can plausibly construct this measure, researchers and regulators do not have enough information to do so (see Bessembinder (2003) for a discussion). Virtually all available estimates of the cost of trade utilizing high-frequency data hinge on the basic logic behind Perold's original suggestions. The effective spread is defined as the (weighted) average of

$$Q_t(\widetilde{p}_t - m_t), \tag{19}$$

where Q_t is an indicator equal to 1 (-1) for buyer- (seller-)initiated trades, $\widetilde{p}_t$ is the logarithm of the transaction price and m_t is the midpoint of the bid and ask quotes. The latter is used as a proxy for the unobserved efficient price prevailing at the time of the trading decision.

The limitations of this measure have been pointed out in the literature. First, the effective spread measure requires the trades to be signed as buyer- or seller-initiated. Commonly used high-frequency data sets, such as the TAQ dataset used in this paper for instance, do not contain information on trade origin. Lee and Ready (1991) and Ellis et al. (2000), among others, propose algorithms intended to classify trades as buyer- of seller-initiated simply on the basis of transaction prices and quotes. While these algorithms perform reasonably well, they have the potential to missclassify a large number of trades, thereby inducing biases in the final estimates. Bessembinder (2003) and Peterson and Sirri (2003) contain a thorough discussion of these issues. Second, effective spreads require the relevant quotes and transaction prices to be matched. Since the trade reports are often delayed, it is difficult to accurately match trade prices to quotes when computing effective spreads. However, it seems sensible to compare the trade prices to quotes that occur before the trade report time. In our work we compute effective spreads by using the conventional Lee and Ready (1991) algorithm and a standard 5-second time allowance.

Table I contains summary statistics for the stocks in our sample. Specifically, we report the average durations, the average prices, the $FITC$s as a percentage of the average prices, the $FITC$s in dollars, the effective spreads as a percentage of the average prices, and the effective spreads in dollars. An asterisk is placed after NASDAQ stocks.

In Figures 2 and 3 we report the histogram of the estimated $FITC$s as a percentage of the corresponding average prices as well as in dollar values. The cross-sectional distribution of the $FITC$s is considerably more left-skewed when reported in percentage values than in absolute values, thereby suggesting that, on average, stocks with higher percentage $FITC$s tend to have lower average prices. Figure 4 presents a plot of the

$FITC$s and the effective spreads, both expressed as a percentage of the corresponding average price. Since the $FITC$s and the effective spreads should be capturing a common market friction component we expect the two measures to be positively correlated and, in fact, their correlation is 0.9. Under Assumption 4 the magnitudes of the $FITC$s and the effective spreads can be meaningfully compared. We should expect the $FITC$s to be larger than the effective spreads and, in fact, they are always larger than the effective spreads by a factor of anywhere between about 5% and 140%. Finally, any sensible measure of transaction cost should be highly correlated with the quoted spreads. We find that the correlation between the $FITC$s and the quoted spreads is 0.94 and stronger than the correlation between the effective spreads and the quoted spreads (0.88).

4.1 The cross-sectional determinants of the $FITC$s, effective spreads, and quoted spreads

Traditional market microstructure literature suggests two main economic forces behind the determination of the quoted spreads: operating (order-processing and inventory-keeping) costs and adverse-selection costs. Liquidity and asymmetric information proxies are known to explain the cross-sectional variation of the effective spreads and quoted spreads. Similarly, if the $FITC$s contain a component that can be imputed to standard frictions as well as a component that can be imputed to learning on the part of the market participants, the same proxies should explain the cross-sectional variation of the $FITC$s. However, due to the additional component capturing the difference between the efficient price and the full information price, we expect the $FITC$s to be more correlated with the asymmetric information proxies than other measures are.

By performing a standard cross-sectional regression in the empirical microstructure literature on friction determination (see Stoll (2000), for example), we show that (i) the $FITC$s are highly correlated with traditional measures of liquidity and private information and (ii) the $FITC$s are more correlated with the asymmetric information proxies than other transaction cost measures.

We regress the logarithm of the percentage $FITC$s ($lfitc$) on the logarithm of the average dollar volume per trade ($lsize$), the logarithm of the average number of shares transacted to shares outstanding ($lturn$), the logarithm of the average daily standard deviation of the true price process (*lsdprice*),[4] the logarithm of the average price (*lprice*), and an NYSE dummy (*nyse).* Table II, column 1, contains the results. The same regressions with logarithmic half-quoted spreads and logarithmic effective spreads as regressors are in Table II, columns 2 and 3.

The variable *lsize* proxies for liquidity and ease of inventory adjustment. The operating cost channel implies that higher *lsize* should translate into smaller spreads. When faced with high dollar volume, the market maker knows that imbalances in risky inventories can easily be restored. Similarly, the market maker is exposed to a variety of fixed operating costs which he recovers by setting transaction costs appropriately. The higher *lsize*, the smaller the fixed cost per transacted share and, consequently, the smaller the necessary transaction cost. The variable *lturn* proxies for the extent of informed trading. The asymmetric information channel implies that higher *lturn* should determine larger spreads. As Stoll (1989) points out, without informed trading, stocks would be traded in proportion to their shares outstanding. Trading rates in excess of this proportion should be associated with informed trading. The variable *lsdprice* proxies for both asymmetric information and ease of inventory adjustment. In both cases, higher *lsdprice* should lead to larger spreads. Higher uncertainty about the fundamental value of the asset increases the risk of transacting with traders with superior information. The increased risk needs to be compensated and the compensation should be proportional to the degree of asymmetry in the market. Equivalently, higher uncertainty about the underlying stock's value implies higher potential for adverse price moves and hence higher inventory risk, mostly in the presence of severe imbalances to be offset (Garber and Silber (1979) and Ho and Stoll (1981)). The variable *lprice* is included to control for price discreteness. As suggested by Stoll (2000), this variable can also be interpreted as an additional proxy for risk in that low price stocks have a tendency to be riskier. Finally, the *nyse* dummy allows us to account for potential exchange effects.

The adjusted R^2 of the $FITC$ regression is 0.95. The variable *lsize* has a significantly negative impact on the $FITC$s with an elasticity of -0.158 and a t-stat of -5.50 supporting the predictions of the operating-cost theory of friction determination. The variable *lturn* is

---

[4] As in the previous section, we estimate the daily standard deviations by using daily realized volatilities. The optimal number of intradaily observations is chosen to minimize the conditional mean-squared error of the realized volatility estimator as proposed in Bandi and Russell (2003a,b).

significantly positive with an elasticity of 0.165 and a t-stat of about 9.16 in agreement with the predictions of the asymmetric information theory of friction determination. Similarly, strong and positive is the cross-sectional relation between the *FITC*s and the volatility of the underlying full-information price. The corresponding coefficient is equal to 0.57 with a t-stat of 12.89. As expected, the coefficient on *lprice* is negative (-0.136) and highly statistically significant with a t-stat of -4.34. The (statistically significant) positive sign of the estimated coefficient on *nyse* (0.887) is somewhat surprising at first. It is widely believed that the decentralized nature of NASDAQ leaves the dealer more exposed to potential losses coming from trading with the informed agents (Heidle and Huang (2002)). The higher risk of informed trading would have to be compensated through larger transaction costs. The opposite result emerges from our sample but a simple observation justifies this outcome. Our sample of NASDAQ stocks is very small (only 7 companies) and characterized by large cap stocks that trade very frequently and have large average volumes. The exchange dummy is likely to pick up an unaccounted for liquidity effect, hence the negative sign. This outcome is not specific to our *FITC* measure. When regressing the quoted and effective spreads on the same controls we also find a significantly positive parameter estimate (Table II, columns 2 and 3). It therefore seems likely that the NASDAQ stocks in our sample of S&P 100 stocks are not representative of the universe of NASDAQ stocks. A thorough analysis of the efficiency properties of NASDAQ stocks is of interest for future research but is beyond the scope of the present paper.

Interestingly, the main asymmetric information proxy in the regression, *lturn*, appears to be considerably more correlated with the *FITC*s than with the quoted and effective spreads. In the effective spread regression the corresponding coefficient is equal to 0.058 with a t-stat of 3. In the half-quoted spread regression the corresponding coefficient is equal to 0.08 with a t-stat of 3.67.

We provide two robustness checks. We run the same regression but replace *lturn* with two alternative measures of asymmetric information that have been widely used in the recent literature, namely the logarithm of the probability of informed trading or *PIN* (*lpin*) (see, for example, Easley et al. (1996)) and the logarithm of the number of analysts following the stock (*lanalysts*). We start with the former (see Table III). Our sample of *PIN* estimates covers 70 NYSE stocks out of the 100 stocks in our original sample. Specifically, we use annual *PIN* measures pertaining to the year 2001.[5] We expect more informed trading to take place in the presence of larger deviations between the full-information prices and the efficient prices. Also, higher informed trading should imply higher adverse selection costs for the market maker. Both effects should lead to larger values of the two components of the *FITC*s. The estimated *lpin* coefficient in the *FITC* regression is equal to 0.187 with a t-stat of 2.32. The corresponding values for the half-quoted spreads and the effective spreads are 0.166 and 0.118 with t-stats of 2.07 and 1.66, respectively. Hence, *lpin* is insignificant in the effective spread regression and barely significant in the quoted spread regression.

We now turn to the number of analysts (Table IV). We use the (logarithm of the) number of analysts following the stocks in our sample over the quarter that includes February 2002 and the number of analysts following the stocks over the entire 2002 year.[6] It is known that *lanalysts* is negatively correlated with *lpin* (Easley et al. (1998)). We confirm this result in our sample (the correlation is $-.2$). We expect a larger number of analysts to induce faster distribution and incorporation of information, resulting in lower risk for the market maker (Brennan and Subrahmanyam (1995)), and hence smaller deviations between transaction prices and efficient prices, as well as smaller deviations of the efficient prices from the full-information prices. A higher *lanalyst* value, therefore, should be associated with smaller *FITC*s. The estimated coefficient in the *FITC* regression is equal to -0.114 with a t-stat of -2.68. The corresponding values for quoted and effective spreads are -0.026 and -0.045 with t-stats of -0.64 and -1.25, respectively. Hence, *lanalysts* is insignificant in both the effective spread regression and the quoted spread regression.

5 How big is the asymmetric information component in the *FITCs?*

This section provides a test of the importance of asymmetric information and quantifies the asymmetric information component in the estimated *FITC*s.

We start with the former. The price formation mechanism in Section 2 implies that

---

[5] We thank Soeren Hvidkjaer for making the *PIN* measures available to us.

[6] The number of analysts is obtained from the Institutional Brokers Estimation System (I/B/E/S) database.

the market effects η are induced by standard market frictions and a pure asymmetric information component, i.e., $\eta = \eta^{asy} + \eta^{fri}$. We recall that η^{fri} denotes the difference between the transaction price and the efficient price whereas η^{asy} denotes the difference between the efficient price and the full-information price. Hence, our model implies that

$$\sigma_{\eta}^2 = \sigma_{\eta^{asy}}^2 + \sigma_{\eta^{fri}}^2 + 2\sigma_{\eta^{asy}\eta^{fri}}, \tag{20}$$

where $\sigma_{\eta^{asy}\eta^{fri}}$ is the covariance between η^{asy} and η^{fri}. Similarly, we can write

$$\sigma_{\eta}^2 - \sigma_{\eta^{fri}}^2 = \sigma_{\eta^{asy}}^2 + 2\sigma_{\eta^{asy}\eta^{fri}}. \tag{21}$$

Proposition. *Since the FITCs contain an asymmetric information component purely driven by the difference between the efficient price and the full-information price, the cross-sectional variation of the difference between the squared FITCs and classic market friction variances* ($\sigma_{\eta}^2 - \sigma_{\eta^{fri}}^2$) *should be explained by the extent of the assets' private information component or variables which are correlated with private information.*

Interestingly, this difference can be estimated from the data. The first term, σ_{η}^2, is the square of the *FITC* measure. The second term is the variance of the difference between the transaction price and the efficient price. Under a standard assumption in the literature, we use the midpoint of the bid and ask prices as a proxy for the unobserved efficient price. We test the prediction in the proposition by regressing the logarithm of $FITC^2 - \widehat{\sigma}_{\eta^{fri}}^2$ on *lturn* (Table V). The variable *lturn* is expected to be positively related to $\sigma_{\eta^{asy}}^2$. The relation between *lturn* and $\sigma_{\eta^{asy}\eta^{fri}}$ is not obvious. The estimated coefficient on *lturn* is positive (1.01) and very statistically significant with a t-stat of 10.72.

As in the previous section, we provide two robustness checks, namely we replace *lturn* with *lpin* and *lanalysts*. Since we expect more informed trading to take place in the presence of larger deviations between the full-information prices and the efficient prices, a higher *lpin* value should be associated with a larger $\sigma_{\eta^{asy}}^2$. The effect on $\sigma_{\eta^{asy}\eta^{fri}}$ is less clear. When we regress the logarithm of $FITC^2 - \widehat{\sigma}_{\eta^{fri}}^2$ on *lpin* we find a positive estimate of 1.33 with a t-stat of 2.6. Since we expect a larger number of analysts to induce faster distribution and incorporation of information, a higher *lanalyst* value should be associated with a smaller $\sigma_{\eta^{asy}}^2$. As earlier in the case of *lturns* and *lpin*, the effect on $\sigma_{\eta^{asy}\eta^{fri}}$ is not obvious. When we regress the logarithm of $FITC^2 - \widehat{\sigma}^2_{\eta^{fri}}$ on *lanalysts* we find a negative coefficient of -1.29 with a t-stat of -4.31.

In sum, these results confirm the presence of a pure private information component in the estimated *FITC*s. It is now interesting to quantify the magnitude of this component. Write

$$\mathbf{E}\left|\widetilde{p}-p\right| \leq \mathbf{E}\left|\widetilde{p}-p^e\right| + \mathbf{E}\left|p^e-p\right| \tag{22}$$

where, as earlier, $\widetilde{p}$, p, and p^e denote the transaction price, the full-information price and the efficient price, respectively. We can now provide a lower bound for the expected difference between the *unobserved* efficient price and the *unobserved* full-information price, namely

$$\mathbf{E}\left|\widetilde{p}-p\right| - \mathbf{E}\left|\widetilde{p}-p^e\right| \leq \mathbf{E}\left|p^e-p\right|. \tag{23}$$

Under a log-normality assumption for the market effect η, the term $\mathbf{E}\left|\widetilde{p}-p\right|$ can be estimated as in Section 3, Corollary to Theorem 1. Using the midpoint of the bid and the ask price as a proxy for the efficient price as earlier, the term $\mathbf{E}\left|\widetilde{p}-p^e\right|$ can be estimated based on transaction prices and midpoints. Hence, the difference $\mathbf{E}\left|\widetilde{p}-p\right| - \mathbf{E}\left|\widetilde{p}-p^e\right|$ can be easily evaluated using *FITC*s. A formal test of the hypothesis that the mean bound is zero is overwhelmingly rejected by our data with a t-stat of 15.

In Figure 5 we plot the histogram of the estimated lower bounds as a percentage of the *FITC* estimates. The mean value is about 35%, the maximum value is about 50%. Hence, the asymmetric information component in the *FITC*s is substantial. In Figure 6 we plot the histogram of the estimated lower bounds as a percentage of the difference between transaction prices and efficient prices. We notice that the mean value is about 80% but values as large as 160% are possible.

Hence, substantial deviations of the efficient price from the full-information price can occur. These departures can be as large as the departures of the transaction prices from the efficient prices. Measures of market quality that do not account for these deviations, like the effective spreads and the half-quoted spreads, have the potential to overstate the extent of market quality substantially.

6 Conclusions

In a world with private information and learning on the part of the market participants, the (positive) differences between *observed* transaction prices and *unobserved* full-information prices, i.e., the prices that reflect all public and private information about the assets, constitute ideal measures of market quality. We call these differences "full-information transaction costs." While the current literature on market quality focuses on measuring the differences between transaction prices and efficient prices, i.e., the prices that embed all publicly available information about the assets, this paper proposes a methodology to study full-information transaction costs.

Our method relies on sample moments of high-frequency transaction return data. As such, it is easy to implement. Furthermore, the method is robust to a variety of realistic price formation mechanisms in that it can accommodate *(i)* predictability in the underlying full-information return process, *(ii)* correlation between the full-information price process and the market effects (i.e., the combined effects of standard frictions and deviations of the efficient price from the full-information price), (*iii)* nonlinearities in the full-information price and market effects, as well as (*iv*) serial dependence in the market effects. If market quality is believed to be affected by realistic operating and adverse-selection costs, as implied by accepted theories of market friction determination and learning, these properties are important.

Using estimates of (second moments of) full-information transaction costs for a sample of S&P 100 stocks, we provide further support for the existing theories of market friction determination and show the importance of learning on the part of the market participants. Importantly, we stress that the deviations of the efficient prices from their full-information levels, as determined by the existence of private information in the marketplace, can be as large as the departures of the transaction prices from the efficient prices, as induced by standard frictions.

Since individuals are likely to take into account the effective cost of acquiring and rebalancing their portfolios, expected stock returns should embed effective execution costs in equilibrium. This observation has given rise to a convergence between market microstructure work on price determination and asset pricing in recent years. The interested reader is referred to the recent survey of Easley and O'Hara (2002). However, the current attempts to characterize the cross-sectional relation between expected returns and execution costs either rely on liquidity-based theories of transaction cost determination (Amihud and Mendelson (1986), among others) or they rely on information-based approaches to the same issue (Easley et al. (2002)). Our methodology to measure full-information transaction costs may be viewed as providing a natural framework to bridge the two arguments in the study of the cross-sectional dependence between expected stock returns and execution costs. Research on this subject is being conducted by the authors and will be reported in later work.

7 Appendix

Proof of Lemma 1. The price formation mechanism in Section 2 implies that the generic serial covariance of order j of the market effects in the observed returns can be expressed as

$$\begin{aligned}\mathbf{E}(\varepsilon\varepsilon_{-j}) &= \mathbf{E}\Big(\left(\eta-\eta_{-1}\right)\left(\eta_{-j}-\eta_{-(j+1)}\right)\Big) && (24)\\ &= \mathbf{E}(\eta\eta_{-j})-\mathbf{E}(\eta\eta_{-(j+1)})-\mathbf{E}(\eta_{-1}\eta_{-j})+\mathbf{E}(\eta_{-1}\eta_{-(j+1)}) && (25)\\ &= 2\mathbf{E}(\eta\eta_{-j})-\mathbf{E}(\eta_{-1}\eta_{-j})-\mathbf{E}(\eta\eta_{-(j+1)}). && (26)\end{aligned}$$

Recall, k is the maximum lag for which the serial covariances of the market effects are different from zero. Hence,

$$\mathbf{E}(\varepsilon\varepsilon_{-j}) = 2\mathbf{E}(\eta\eta_{-j})-\mathbf{E}(\eta_{-1}\eta_{-j}) \tag{27}$$

when $j = k$ and

$$\mathbf{E}(\varepsilon\varepsilon_{-j}) = 2\mathbf{E}(\eta\eta_{-j})-\mathbf{E}(\eta_{-1}\eta_{-j})-\mathbf{E}(\eta\eta_{-(j+1)}) \tag{28}$$

for $1 \leq j < k$. We now plug Eq. (27) and Eq. (28) into the right-hand side of the squared version of Eq. (6) and obtain

$$\begin{aligned}& \left(\frac{1+k}{2}\right)\mathbf{E}(\varepsilon^2)+\sum_{s=0}^{k-1}(s+1)\mathbf{E}(\varepsilon\varepsilon_{-(k-s)})\\ =\ & \left(\frac{1+k}{2}\right)\left[2\mathbf{E}(\eta^2)-2\mathbf{E}(\eta\eta_{-1})\right]\\ & +\sum_{s=1}^{k-1}(s+1)\left[2\mathbf{E}(\eta\eta_{-(k-s)})-\mathbf{E}(\eta_{-1}\eta_{-(k-s)})-\mathbf{E}(\eta\eta_{-((k-s)+1)})\right]\\ & +2\left[\mathbf{E}(\eta\eta_{-k})-\mathbf{E}(\eta_{-1}\eta_{-k})\right] && (29)\\ =\ & (1+k)\left(\mathbf{E}(\eta^2)-\mathbf{E}(\eta\eta_{-1})\right)\\ & +k\left[2\mathbf{E}(\eta\eta_{-1})-\mathbf{E}(\eta^2)-\mathbf{E}(\eta\eta_{-2})\right]\\ & +(k-1)\left[2\mathbf{E}(\eta\eta_{-2})-\mathbf{E}(\eta\eta_{-1})-\mathbf{E}(\eta\eta_{-3})\right]\\ & +(k-2)\left[2\mathbf{E}(\eta\eta_{-3})-\mathbf{E}(\eta\eta_{-2})-\mathbf{E}(\eta\eta_{-4})\right]\\ & +(k-3)\left[2\mathbf{E}(\eta\eta_{-4})-\mathbf{E}(\eta\eta_{-3})-\mathbf{E}(\eta\eta_{-5})\right]\\ & +\ldots\\ & +3[2\mathbf{E}(\eta\eta_{-k+2})-\mathbf{E}(\eta\eta_{-k+3})-\mathbf{E}(\eta\eta_{-k+1})]\\ & +2[2\mathbf{E}(\eta\eta_{-k+1})-\mathbf{E}(\eta\eta_{-k+2})-\mathbf{E}(\eta\eta_{-k})]\\ & +[2\mathbf{E}(\eta\eta_{-k})-\mathbf{E}(\eta\eta_{-k+1})] && (30)\end{aligned}$$

Finally, we notice that Eq. (30) is equal to $\mathbf{E}(\eta^2)$. This proves the stated result. ■

Proof of Theorem 1. Given a fixed j such that $1 \le j \le N(\overline{h})-1$, write

$$\frac{\sum_{i=j+1}^{N(\overline{h})}\widetilde{r}_i\widetilde{r}_{i-j}}{N(\overline{h})-j} = \underbrace{\frac{\sum_{i=j+1}^{N(\overline{h})} r_{t_i}r_{t_{i-j}}}{N(\overline{h})-j}}_{\alpha} + \underbrace{\frac{\sum_{i=j+1}^{N(\overline{h})} r_{t_i}\varepsilon_{i-j}}{N(\overline{h})-j}}_{\beta} + \underbrace{\frac{\sum_{i=j+1}^{N(\overline{h})} \varepsilon_i r_{t_{i-j}}}{N(\overline{h})-j}}_{\gamma} + \underbrace{\frac{\sum_{i=j+1}^{N(\overline{h})} \varepsilon_i\varepsilon_{i-j}}{N(\overline{h})-j}}_{\zeta}. \tag{31}$$

We start with term α. Define $\pi^{N(\overline{h})}$ as $\max\left\{|t_{i+1}-t_i|, i=1,...,N(\overline{h})\right\}$. Then,

$$\begin{align}
\alpha &\le \frac{\left(\sum_{i=j+1}^{N(\overline{h})} r_{t_i}^2\right)^{1/2}\left(\sum_{i=j+1}^{N(\overline{h})} r_{t_{i-j}}^2\right)^{1/2}}{N(\overline{h})-j} \tag{32}\\
&= \frac{\left(\sum_{i=1}^{N(\overline{h})} r_{t_i}^2 - \sum_{i=1}^{j} r_{t_i}^2\right)^{1/2}\left(\sum_{i=1}^{N(\overline{h})} r_{t_i}^2 - \sum_{i=N(\overline{h})-j+1}^{N(\overline{h})} r_{t_i}^2\right)^{1/2}}{N(\overline{h})-j} \tag{33}\\
&\le \frac{\left([p]_0^{\overline{h}} + o_p(1) + jO_p\left(\pi^{N(\overline{h})}\right)\right)^{1/2}\left([p]_0^{\overline{h}} + o_p(1) + jO_p\left(\pi^{N(\overline{h})}\right)\right)^{1/2}}{N(\overline{h})-j} \tag{34}\\
&\underset{N(\overline{h})\to\infty}{\overset{p}{\rightarrow}} 0, \tag{35}
\end{align}$$

where $[p]_0^{\overline{h}} = [M]_0^{\overline{h}} + \sum_{0<s\le\overline{h}}(\Delta p_s)^2 = \int_0^{\overline{h}}\sigma_s^2 ds + \sum_{0<s\le\overline{h}}(\Delta p_s)^2$ is the quadratic variation of the underlying logarithmic price process. It is noted that Eq. (32) derives from the Cauchy's inequality while Eq. (34) derives from a standard convergence result in semimartingale process theory (see Protter, Theorem 22, page 59, 1995, for example). Specifically, under Assumptions 1(1)(2),

$$\sum_{i=1}^{N(\overline{h})}\left(p_{t_i}-p_{t_{i-1}}\right)^2 \underset{N(\overline{h})\to\infty}{\overset{p}{\rightarrow}} [p]_0^{\overline{h}} = [M]_0^{\overline{h}} + \sum_{0<s\le\overline{h}}(\Delta p_s)^2 \tag{36}$$

if $lim_{N(\overline{h})\to\infty}\pi^{N(\overline{h})} = 0$. Now consider the term ζ and write,

$$\begin{align}
\mathbf{1}\left\{|\zeta - \mathbf{E}(\varepsilon\varepsilon_{-j})| > \delta\right\} &\le \frac{|\zeta - \mathbf{E}(\varepsilon\varepsilon_{-j})|}{\delta} \tag{37}\\
&\le \frac{\left(\zeta - \mathbf{E}(\varepsilon\varepsilon_{-j})\right)^2}{\delta^2}, \tag{38}
\end{align}$$

where $\mathbf{1}\{A\}$ is the indicator function of the generic set A and the first line follows from Markov's inequality for any positive and arbitrarily small δ. By the monotonicity property of the expectation operator, taking expectations of both sides of Eq. (38), we obtain

$$P\left\{|\zeta - \mathbf{E}(\varepsilon\varepsilon_{-j})| > \delta\right\} \le \frac{\mathbf{E}\left(\zeta - \mathbf{E}(\varepsilon\varepsilon_{-j})\right)^2}{\delta^2}. \tag{39}$$

By Assumption 2(1), we note that

$$\begin{align}
&\mathbf{E}\left(\zeta - \mathbf{E}(\varepsilon\varepsilon_{-j})\right)^2 \nonumber\\
=\; & \frac{1}{\left(N(\overline{h})-j\right)^2}\mathbf{E}\left(\sum_{i=j+1}^{N(\overline{h})}\left(\varepsilon_i\varepsilon_{i-j} - \mathbf{E}(\varepsilon\varepsilon_{-j})\right)\right)^2 \tag{40}\\
=\; & \frac{1}{\left(N(\overline{h})-j\right)^2}\left(\left(N(\overline{h})-j\right)\lambda_0 + 2\left(N(\overline{h})-j-1\right)\lambda_1 + ... + 2\lambda_{N(\overline{h})-j}\right), \tag{41}\\
\le\; & \frac{1}{\left(N(\overline{h})-j\right)}\left(|\lambda_0| + 2|\lambda_1| + ... + 2\left|\lambda_{N(\overline{h})-j}\right|\right), \tag{42}
\end{align}$$

where

$$\lambda_s = \mathbf{E}\left[\left(\varepsilon_s\varepsilon_{s-j} - \mathbf{E}\left(\varepsilon\varepsilon_{-j}\right)\right)\left(\varepsilon\varepsilon_{-j} - \mathbf{E}\left(\varepsilon\varepsilon_{-j}\right)\right)\right] \tag{43}$$

with $0 \leq s \leq N(\overline{h}) - j$. Hence,

$$\begin{aligned} P\left\{\left|\zeta - \mathbf{E}\left(\varepsilon\varepsilon_{-j}\right)\right| > \varepsilon\right\} &\leq \frac{1}{\delta^2(N(\overline{h}) - j)}\left(2\left|\lambda_0\right| + 2\left|\lambda_1\right| + \ldots + 2\left|\lambda_{N(\overline{h})-j}\right| - \left|\lambda_0\right|\right) \\ &\underset{N(\overline{h})\to\infty}{\rightarrow} 0 \end{aligned}$$

since $\sum_{s=0}^{\infty}|\lambda_s| < \infty$ given Assumption 2(3). This proves convergence in probability of the term ζ to $\mathbf{E}\left(\varepsilon\varepsilon_{-j}\right)$. Now we turn to β.

$$\begin{aligned} \beta &\leq \frac{\left(\sum_{i=j+1}^{N(\overline{h})} r_{t_i}^2\right)^{1/2}\left(\sum_{i=j+1}^{N(\overline{h})} \varepsilon_{i-j}^2\right)^{1/2}}{N(\overline{h}) - j} && (44) \\ &= \frac{1}{\sqrt{N(\overline{h}) - j}}\left([p]_0^{\overline{h}} + o_p(1) + jO_p\left(\pi^{N(\overline{h})}\right)\right)^{1/2}\left(\frac{\sum_{i=j+1}^{N(\overline{h})} \varepsilon_{i-j}^2}{N(\overline{h}) - j}\right)^{1/2} && (45) \\ &= \frac{1}{\sqrt{N(\overline{h}) - j}}O_p(1)\left(\mathbf{E}(\varepsilon^2) + o_p(1)\right) \underset{N(\overline{h})\to\infty}{\overset{p}{\rightarrow}} 0, && (46) \end{aligned}$$

where Eq. (44) follows from Cauchy's inequality and the convergence in probability of $\frac{\sum_{i=j+1}^{N(\overline{h})}\varepsilon_{i-j}^2}{N(\overline{h})-j}$ to $\mathbf{E}(\varepsilon^2)$ derives from an argument that is similar to the argument used for term ζ. The quantity γ can be examined in the same fashion. Write

$$\begin{aligned} \gamma &\leq \frac{\left(\sum_{i=j+1}^{N(\overline{h})} \varepsilon_i^2\right)^{1/2}\left(\sum_{i=j+1}^{N(\overline{h})} r_{t_{i-j}}^2\right)^{1/2}}{N(\overline{h})} && (47) \\ &= \frac{1}{\sqrt{N(\overline{h})}}O_p(1)\left(\mathbf{E}(\varepsilon^2) + o_p(1)\right) \underset{N(\overline{h})\to\infty}{\overset{p}{\rightarrow}} 0. && (48) \end{aligned}$$

Finally, the statement in Theorem 1 readily derives from Slutsky's theorem given the continuity of σ_η as a function of the cross-moments of the market effects in returns.■

References

[1] Amihud, Y., and H. Mendelson (1980). Dealership markets: market making with inventory. *Journal of Financial Economics,* 8, 31-53.

[2] Amihud, Y., and H. Mendelson (1986). Asset pricing and the bid-ask spread. *Journal of Financial Economics,* 17, 223-249.

[3] Andersen, T.G., T. Bollerslev, F.X. Diebold, and P. Labys (2003). Modeling and forecasting realized volatility. *Econometrica,* 71, 579-625.

[4] Bandi, F.M., and J.R. Russell (2003a). Realized variance, microstructure noise, and optimal sampling. *Working paper.*

[5] Bandi, F.M., and J.R. Russell (2003b). Separating microstructure noise from volatility. *Journal of Financial Economics,* forthcoming.

[6] Barndorff-Nielsen, O.E., and N. Shephard (2002). Econometric analysis of realized volatility and its use in estimating stochastic volatility models. *Journal of the Royal Statistical Society, Series B, 64, 253-280.*

[7] Beja, A., and B. Goldman (1980). On the dynamics of behavior of prices in disequilibrium. *Journal of Finance,* 35, 235-248.

[8] Bessembinder, H. (2003). Issues in assessing trade execution costs. *Journal of Financial Markets,* 6, 233-257.

[9] Biais, B., P. Hillion, and C. Spatt (1995). An empirical analysis of the limit order book and the order flow in the Paris Bourse. *Journal of Finance,* 50, 1655-1689.

[10] Biais, B., P. Hillion, and C. Spatt (1999). Price discovery and learning during the pre-opening period in the Paris Bourse. *Journal of Political Economy,* 107, 1218-1248.

[11] Black, F. (1986). Noise. *Journal of Finance,* 41, 529-543.

[12] Brennan, M.J., and A. Subrahmanyam (1995). Investment analysis and price formation in securities markets. *Journal of Financial Economics,* 38, 361-381.

[13] Choi, J.Y., D. Salandro and K. Shastri (1988). On the estimation of the bid-ask spreads: theory and evidence. *Journal of Financial and Quantitative analysis,* 23, 219-230.

[14] Copeland, T., and D. Galai (1983). Information effects in the bid-ask spread. *Journal of Finance,* 38, 1457-1469.

[15] Demsetz, H. (1968). The cost of transacting. *Quarterly Journal of Economics,* 82, 33-53.

[16] Duffie, D. (1990). *Dynamic asset pricing.* Princeton Press.

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

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

[19] Easley D., S. Hvidkjaer, and M. O'Hara (2002). Is information risk a determinant of asset returns? *Journal of Finance,* 57, 2185-2221.

[20] Easley D., N. Kiefer, and M. O'Hara (1996). Cream-skimming or profit sharing? the curious role of purchased order flow. *Journal of Finance*, 51, 811-833.

[21] Easley D., M. O'Hara, and J. Paperman (1998). Financial Analysts and information-based trade. *Journal of Financial Markets*, 1, 175-201.

[22] Easley D., N. Kiefer, M. O'Hara, and J. Paperman (1996). Liquidity, information and infrequently traded stocks. *Journal of Finance*, 1, 1405-1436.

[23] Ellis, K., R. Michaely, and M. O'Hara (2000). The accuracy of trade classification rules: evidence from Nasdaq. *Journal of Financial and Quantitative Analysis,* 35, 529-552.

[24] Garbade, K., and Z. Lieber (1977). On the independence of transactions on the NYSE. *Journal of Banking and Finance,* 1, 151-172.

[25] Garber, K.D., and W.L. Silber (1979). Structural organization of secondary markets: clearing frequency, dealer activity, and liquidity risk. *Journal of Finance,* 34, 577-593.

[26] Garman, M. (1976). Market microstructure. *Journal of Financial Economics,* 3, 257-275.

[27] Germain, L., N. Meddahi, and E. Renault (1996). The speed of learning in financial markets. *Working paper.*

[28] Glosten, L.R., and P. Milgrom (1985). Bid, ask and transaction prices in specialist market with heterogenously informed traders. *Journal of Financial Economics,* 14, 71-100.

[29] Goldman, B., and A. Beja (1979). Market prices versus equilibrium prices: return's variance, serial correlation, and the role of the specialist. *Journal of Finance,* 34, 595-607.

[30] Hansen, P.R. and A. Lunde (2006). Realized variance and market microstructure noise. *Journal of Business and Economic Statistics,* forthcoming.

[31] Hasbrouck, J. (1993). Assessing the quality of a security market: a new approach to transaction cost measurement. *Review of Financial Studies,* 6, 191-212.

[32] Hasbrouck, J. (1999). Liquidity in the future pits: inferring market dynamics from incomplete data. *Working paper.*

[33] Hasbrouck, J. (2003). Trading costs and returns for U.S. securities: the evidence from daily data. *Working paper.*

[34] Hasbrouck, J. and T. Ho (1987). Order arrival, quote behavior, and the return-generating process. *Journal of Finance,* 4, 1035-1048.

[35] Heidle, H, and R. Huang (2002). Information-based trading in dealer and auction markets: an analysis of exchange listings. *Journal of Financial and Quantitative Analysis* 37, 391 - 424.

[36] Ho, T., and H.R. Stoll (1981). Optimal dealer pricing under transactions and return uncertainty. *Journal of Financial Economics,* 9, 47-73.

[37] Kyle, A. (1985). Continuous auctions and insider trading. *Econometrica,* 53, 1315-1335.

[38] Lee, C. and M. Ready (1991). Inferring trade direction from intraday data. *Journal of Finance,* 46, 733-746.

[39] O'Hara, M. (1995). *Market Microstructure Theory.* Blackwell Publishers Ltd.

[40] Oomen, R.C.A. (2005). Properties of bias corrected realized variance under alternative sampling schemes. *Journal of Financial Econometrics,* 3, 555-577.

[41] Perold, A (1988). The implementation shortfall. *Journal of Portfolio Management,* 14, 4-9.

[42] Peterson, M. and E. Sirri (2002). Evaluation of the biases in execution cost estimates using trade and quote data. *Journal of Financial Markets,* 6, 259-280.

[43] Protter, P. (1995). *Stochastic Integration and Differential Equations.* Springer, Berlin.

[44] Roll, R. (1984). A simple measure of the effective bid-ask spread in an efficient market. *Journal of Finance,* 39, 1127-1139.

[45] Smidt, S. (1971). Which road to an efficient stock market: free competition or regulated monopoly? *Financial Analysts Journal,* 27, 64-69.

[46] Stock, J. and M. Watson (1988). Variable trends in economic time series. *Journal of Economic Perspectives* 2, 147-174.

[47] Stoll, H.R. (1989). Inferring the components of the bid-ask spread: theory and empirical tests. *Journal of Finance,* 44, 115-134.

[48] Stoll, H.R. (2000). Friction. *Journal of Finance,* 55, 1479-1514.

[49] Tinic, S. (1972). The economics of liquidity services. *Quarterly Journal of Economics,* 86, 79-93.

[50] Vives, X. (1995). The speed of information revelation in a financial market mechanism. *Journal of Economic Theory,* 67, 178-204.

Table I

Descriptive statistics for the S&P 100 stocks in our sample.

The table contains the average durations, the average prices, the estimated full-information transaction costs (as a percentage of the average prices), the estimated full-information transaction costs (in dollar values), the percentage effective spreads (computed using the Lee and Ready (1991) algorithm and a standard 5 second time allowance), the dollar effective spreads.

Symbol	Duration	Avg Price	FITC (%)	FITC ($)	ESpd. (%)	ESpd. ($)
AA	11.35	$36.09	0.12%	$0.044	0.055%	$0.0199
AEP	21.64	$41.77	0.11%	$0.048	0.043%	$0.0182
AES	15.79	$7.75	0.64%	$0.050	0.274%	$0.0212
AIG	10.63	$72.89	0.10%	$0.073	0.043%	$0.0313
ALL	15.42	$33.92	0.14%	$0.046	0.054%	$0.0184
AMGN[*]	1.32	$57.82	0.05%	$0.027	0.030%	$0.0172
AOL	6.45	$25.17	0.15%	$0.037	0.079%	$0.0199
ATI	64.87	$15.64	0.29%	$0.046	0.089%	$0.0140
AVP	21.81	$49.10	0.11%	$0.054	0.039%	$0.0194
AXP	8.80	$34.13	0.12%	$0.040	0.054%	$0.0184
BA	10.88	$43.52	0.11%	$0.048	0.050%	$0.0219
BAC	7.25	$61.20	0.09%	$0.053	0.042%	$0.0256
BAX	17.49	$55.33	0.11%	$0.059	0.043%	$0.0239
BCC	31.58	$34.98	0.17%	$0.060	0.058%	$0.0202
BDK	23.95	$43.50	0.14%	$0.059	0.043%	$0.0189
BHI	14.56	$34.53	0.16%	$0.056	0.061%	$0.0211
BMY	9.02	$45.08	0.10%	$0.044	0.044%	$0.0197
BNI	24.83	$27.95	0.16%	$0.043	0.063%	$0.0175
BUD	20.15	$48.69	0.10%	$0.048	0.042%	$0.0207
C	6.00	$44.43	0.10%	$0.046	0.053%	$0.0234
CCU	12.11	$47.06	0.14%	$0.067	0.055%	$0.0259
CI	20.13	$92.46	0.14%	$0.134	0.051%	$0.0471
CL	15.20	$55.64	0.09%	$0.052	0.037%	$0.0207
CPB	25.98	$27.02	0.17%	$0.046	0.072%	$0.0195
CSC	18.84	$47.07	0.18%	$0.084	0.063%	$0.0296
CSCO[*]	0.45	$16.69	0.05%	$0.009	0.051%	$0.0086
DAL	17.24	$32.76	0.17%	$0.055	0.060%	$0.0197
DD	10.83	$45.08	0.10%	$0.046	0.043%	$0.0194
DIS	8.59	$23.29	0.12%	$0.029	0.057%	$0.0133
DOW	17.33	$30.05	0.14%	$0.042	0.057%	$0.0171
EK	17.05	$29.24	0.15%	$0.044	0.058%	$0.0170
EMC	7.91	$13.42	0.20%	$0.026	0.121%	$0.0163
EP	15.48	$37.01	0.18%	$0.065	0.076%	$0.0283
ETR	26.96	$41.08	0.13%	$0.052	0.043%	$0.0177
EXC	22.35	$50.00	0.13%	$0.067	0.046%	$0.0231
F	12.29	$14.69	0.13%	$0.019	0.067%	$0.0099
FDX	14.68	$54.75	0.11%	$0.060	0.044%	$0.0239
G	18.77	$33.23	0.13%	$0.043	0.050%	$0.0165
GD	17.40	$89.13	0.11%	$0.096	0.047%	$0.0416
GE	4.73	$37.49	0.08%	$0.029	0.044%	$0.0167
GM	14.07	$51.65	0.09%	$0.047	0.031%	$0.0163
GS	8.73	$82.27	0.10%	$0.085	0.047%	$0.0386
HAL	11.38	$15.21	0.21%	$0.033	0.101%	$0.0153
HCA	16.62	$42.31	0.13%	$0.057	0.051%	$0.0215
HD	7.38	$50.40	0.08%	$0.042	0.038%	$0.0194
HET	29.73	$38.34	0.19%	$0.071	0.073%	$0.0281
HIG	17.86	$65.73	0.13%	$0.082	0.047%	$0.0312
HNZ	20.73	$40.98	0.11%	$0.043	0.040%	$0.0165
HON	12.73	$34.34	0.21%	$0.071	0.067%	$0.0228
HWP	9.97	$20.57	0.14%	$0.028	0.062%	$0.0128

Symbol	Duration	Avg Price	FITC (%)	FITC ($)	ESpd. (%)	Espd. ($)
IBM	6.58	$102.81	0.08%	$0.082	0.038%	$0.0393
INTC*	0.52	$31.74	0.04%	$0.013	0.037%	$0.0116
IP	11.46	$42.92	0.10%	$0.044	0.044%	$0.0187
JNJ	10.54	$57.89	0.08%	$0.047	0.041%	$0.0237
JPM	7.91	$29.83	0.16%	$0.049	0.076%	$0.0226
KO	11.28	$46.32	0.09%	$0.043	0.039%	$0.0181
LEH	11.12	$58.99	0.14%	$0.082	0.050%	$0.0297
LTD	21.50	$17.60	0.18%	$0.032	0.067%	$0.0118
LU	10.86	$5.71	0.22%	$0.013	0.140%	$0.0080
MAY	24.32	$35.55	0.16%	$0.056	0.051%	$0.0180
MCD	12.42	$26.81	0.11%	$0.028	0.051%	$0.0137
MDT	10.84	$47.01	0.10%	$0.046	0.042%	$0.0200
MEDI*	3.73	$40.80	0.11%	$0.043	0.059%	$0.0241
MER	7.14	$47.63	0.12%	$0.059	0.061%	$0.0288
MMM	12.16	$114.96	0.10%	$0.116	0.039%	$0.0444
MO	10.07	$51.28	0.07%	$0.037	0.038%	$0.0196
MRK	10.23	$59.96	0.08%	$0.048	0.036%	$0.0217
MSFT*	0.55	$60.16	0.03%	$0.018	0.026%	$0.0155
MWD	6.97	$49.72	0.11%	$0.056	0.054%	$0.0269
NSC	20.77	$21.79	0.18%	$0.040	0.077%	$0.0167
NSM	13.73	$26.60	0.21%	$0.055	0.093%	$0.0247
NXTL*	1.09	$5.07	0.25%	$0.013	0.187%	$0.0095
ONE	13.28	$35.58	0.13%	$0.047	0.050%	$0.0179
ORCL*	0.77	$16.00	0.06%	$0.009	0.053%	$0.0085
PEP	9.88	$49.71	0.08%	$0.041	0.037%	$0.0184
PFE	6.38	$41.06	0.07%	$0.030	0.037%	$0.0152
PG	9.49	$83.99	0.08%	$0.063	0.031%	$0.0261
ROK	46.56	$18.71	0.30%	$0.056	0.098%	$0.0184
RSH	19.39	$27.70	0.21%	$0.057	0.077%	$0.0214
RTN	19.43	$37.90	0.13%	$0.050	0.050%	$0.0188
S	14.98	$52.76	0.10%	$0.055	0.045%	$0.0235
SBC	7.87	$36.56	0.10%	$0.035	0.047%	$0.0171
SLB	8.82	$55.86	0.11%	$0.060	0.049%	$0.0273
SLE	21.26	$21.36	0.13%	$0.028	0.060%	$0.0127
SO	22.79	$25.00	0.12%	$0.030	0.056%	$0.0141
T	11.04	$15.59	0.15%	$0.024	0.071%	$0.0111
TOY	26.90	$17.63	0.23%	$0.041	0.078%	$0.0138
TXN	7.12	$30.52	0.14%	$0.044	0.065%	$0.0198
TYC	5.30	$29.34	0.23%	$0.068	0.152%	$0.0445
UIS	32.61	$11.65	0.23%	$0.027	0.097%	$0.0113
USB	17.03	$19.93	0.15%	$0.030	0.066%	$0.0131
UTX	11.98	$69.41	0.09%	$0.063	0.039%	$0.0269
VIAB	11.56	$42.42	0.15%	$0.063	0.054%	$0.0230
VZ	8.54	$45.76	0.09%	$0.041	0.042%	$0.0193
WFC	9.17	$46.10	0.08%	$0.037	0.037%	$0.0171
WMB	11.44	$15.97	0.28%	$0.045	0.125%	$0.0200
WMT	8.37	$60.02	0.08%	$0.047	0.038%	$0.0227
WY	15.86	$59.75	0.12%	$0.070	0.045%	$0.0267
XOM	7.56	$39.41	0.07%	$0.029	0.040%	$0.0157
XRX	21.27	$10.12	0.24%	$0.024	0.107%	$0.0108
Average	14.30	$40.69	0.14%	$0.048	0.061%	$0.0207

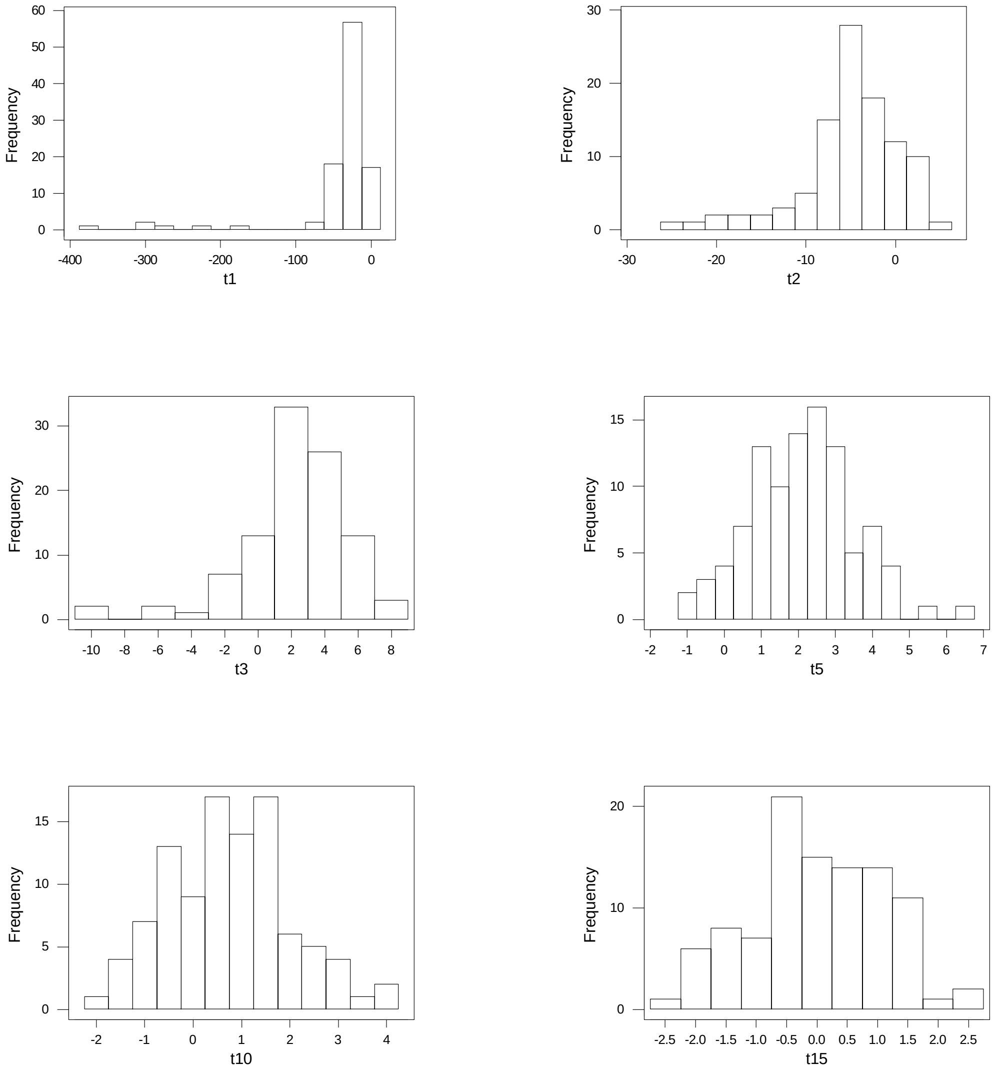

Figure 1. Histograms of the t-ratios of the estimated serial correlations of order 1,2,3,5,10 and 15 of the transaction prices of the S&P 100 stocks.

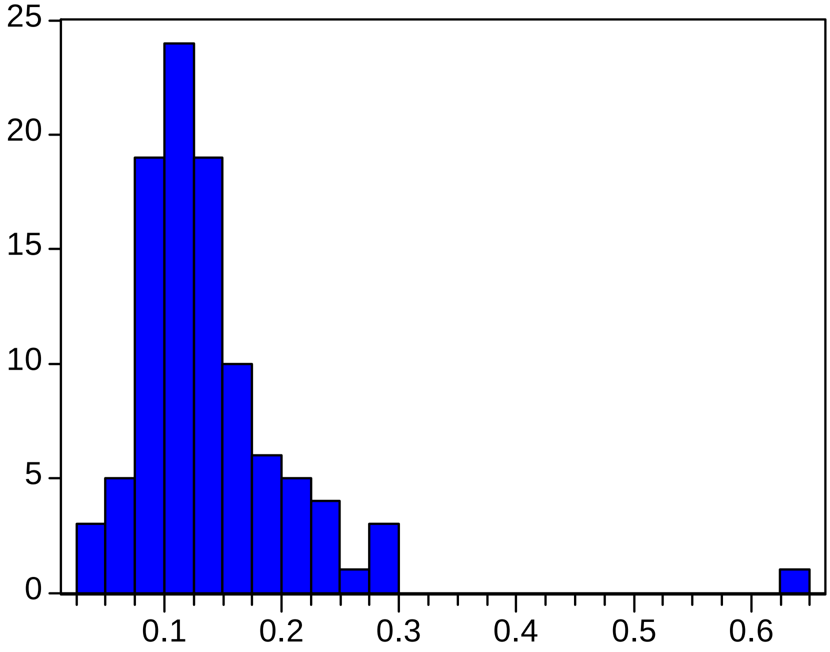

Figure 2. Histogram of the estimated *FITC*'s of the S&P 100 stocks (in percentage values).

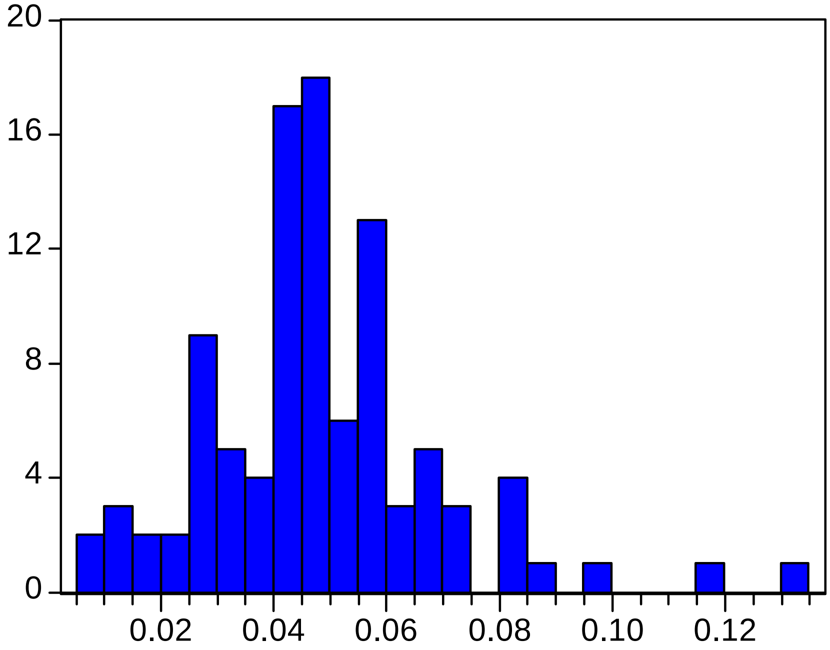

Figure 3. Histogram of the estimated *FITC*'s of the S&P 100 stocks (in dollar values).

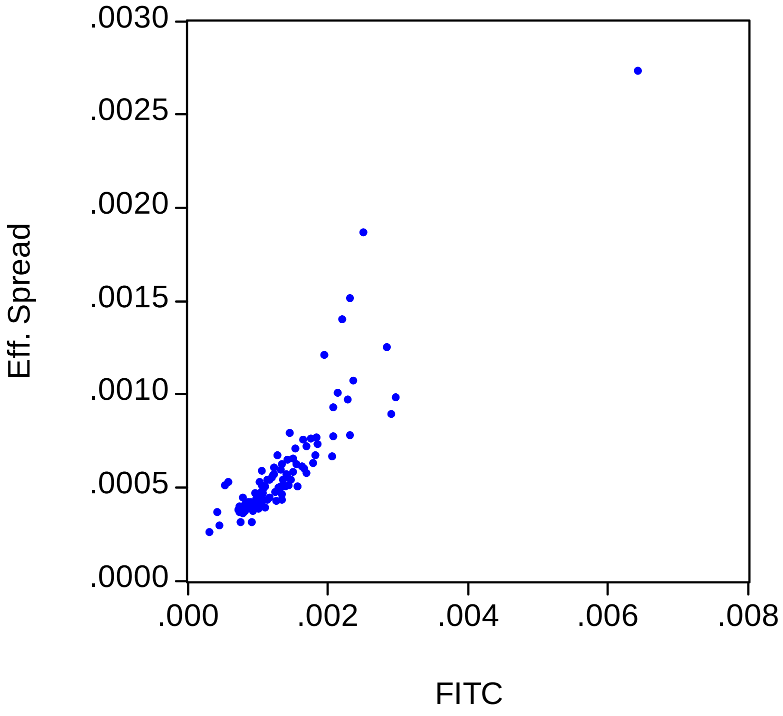

Figure 4. Plot of the estimated effective spreads and the *FITC*’s of the S&P 100 stocks.

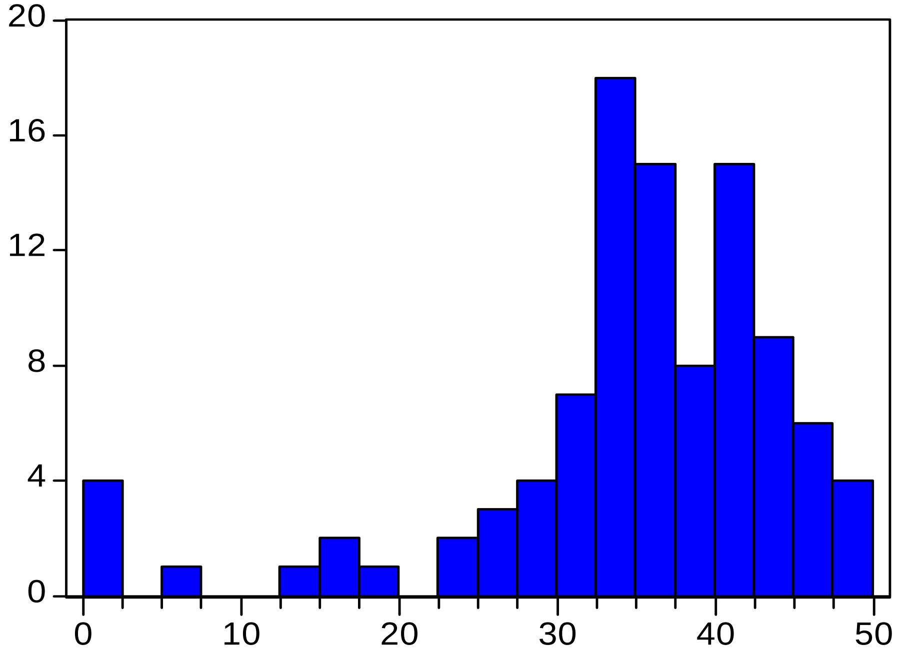

Figure 5. Histogram of the estimated lower bound of the asymmetric information component expressed as a percentage of the *FITC*'s.

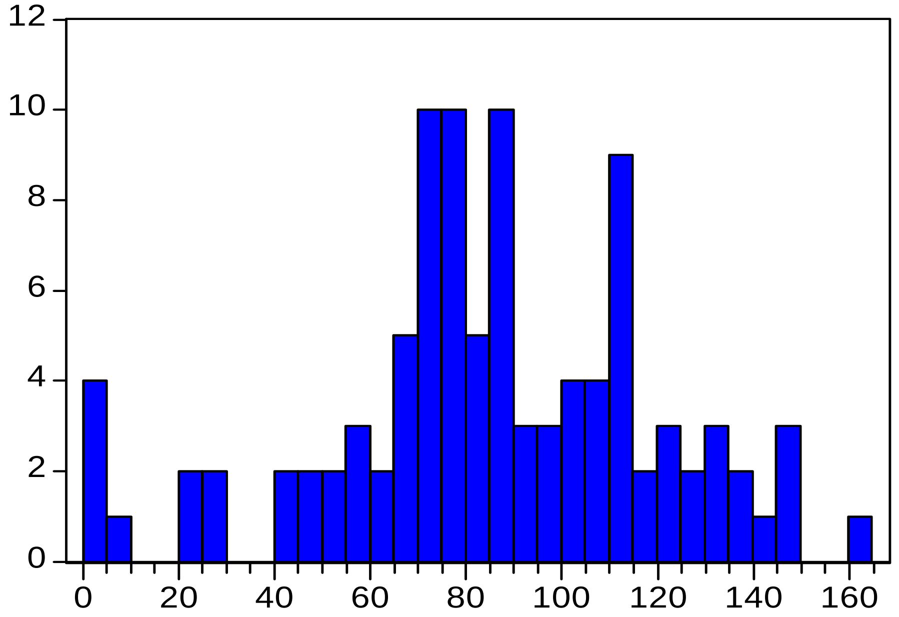

Figure 6. Histogram of the estimated lower bound of the asymmetric information component expressed as a percentage of the effective spreads.

	Log FITC	*Log Half-Spread*	*Log Eff. Spread*
Intercept	-2.79	-3.62	-3.93
	(-10.01)**	(-10.80)**	(-13.14)**
Log Turnover	**0.165**	**0.079**	**0.058**
(lturn)	**(9.16)****	**(3.67)****	**(3.00)****
Log Size	-0.158	-0.145	-0.045
(lsize)	(-5.50)**	(-4.21)**	(-1.47)
Log SD Price	0.570	0.419	0.618
(lsdprice)	(12.89)**	(7.88)**	(13.04)**
Log Price	-0.136	-0.380	-0.273
(lprice)	(-4.34)**	(-10.14)**	(-8.16)**
NYSE Dummy	0.887	0.727	0.439
(nyse)	(12.03)**	(8.20)**	(5.56)**
	adjR^2*=94.9%*	*adj*R^2*=92.9%*	*adj*R^2*=93.0%*

Table II. Outcome of regressions of the logarithm of the *FITC*'s, the logarithm of the half-spreads, and the logarithm of the effective spreads on the logarithm of the average number of daily shares transacted to shares outstanding, the logarithm of the average dollar volume per trade, the logarithm of the average daily standard deviation of the true price process, the logarithm of the average price and an NYSE dummy. ** denotes significance at the 5% level.

	Log FITC	*Log Half-Spread*	*Log Eff. Spread*
Intercept	-0.175	-1.95	-2.79
	-0.524	(-5.89)**	(-9.48)**
Log PIN	**0.187**	**0.166**	**0.118**
(lpin)	**(2.32)***	**(2.07)***	**(1.66)****
Log Size	-0.255	-0.192	-0.099
(lsize)	(-8.22)**	(-6.25)**	(-3.64)
Log SD Price	0.732	0.524	0.646
(lsdprice)	(14.15)**	(10.24)**	(14.24)**
Log Price	-0.092	-0.313	-0.222
(lprice)	(-2.11)**	(-7.25)**	(-5.79)**
	adjR^2*=90.9%*	*adj*R^2*=91.3%*	*adj*R^2*=91.6%*

Table III. Outcome of regressions of the logarithm of the *FITC*'s, the logarithm of the half-spreads, and the logarithm of the effective spreads on the logarithm of the *PIN* measures, the logarithm of the average dollar volume per trade, the logarithm of the average daily standard deviation of the true price process, the logarithm of the average price and an NYSE dummy. ** denotes significance at the 5% level.

	Log FITC	*Log Half-Spread*	*Log Eff. Spread*
Intercept	-1.19	-2.87	-3.36
	(-4.03)**	(-10.01)**	(-13.55)**
Log # Analysts	**-0.114**	**-0.0265**	**-0.0449**
(lanalysts)	**(-2.68)****	**(-.644)**	**(-1.25)**
Log Size	-0.299	-0.222	-0.0935
(lsize)	(-9.55)**	(-7.33)**	(-3.56)
Log SD Price	0.786	0.527	0.693
(lsdprice)	(16.15)**	(11.18)**	(16.98)**
Log Price	-0.027	-0.324	-0.236
(lprice)	(-0.74)	(-8.84)**	(-7.44)**
NYSE Dummy	1.32	0.960	0.590
(nyse)	(19.13)**	(14.32)**	(10.14)**
	adjR^2=95%	*adjR^2=92.9%*	*adjR^2=92.4%*

Table IV. Outcome of regressions of the logarithm of the *FITC*'s, the logarithm of the half-spreads, and the logarithm of the effective spreads on the logarithm of the number of analysts following the stock for the sample month, the logarithm of the average dollar volume per trade, the logarithm of the average daily standard deviation of the true price process, the logarithm of the average price and an NYSE dummy. ** denotes significance at the 5% level.

	Estimates	*Estimates*	*Estimates*
Intercept	-12.56	-10.68	-1103
	(-95.23)**	(-8.77)**	(-16.42)**
Log Turnover	0.89		
	(12.33)**		
Log PIN		1.347	
		(2.66)**	
Log # Analysts			-1.08
			(-4.46)**
	R^2=62.3%	*R^2=9.4%*	*R^2=17.8%*

Table V. Outcomes of regressions of the logarithm of the difference between the squared *FITC*'s and the average squared distance between the transaction prices and the quote midpoints on the logarithm of the average number of daily shares transacted to shares outstanding, the logarithm of the *PIN* measures, and the logarithm of the number of analysts following the stocks.