

Full-information transaction costs*

Federico M. Bandi and Jeffrey R. Russell
Graduate School of Business, The University of Chicago

First draft: December 2003
This version: December 2004

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 efficiency. We call this difference “Full-information transaction cost.” We propose a simple and robust methodology to evaluate 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.

JEL classification: G12, G19.

Keywords: Asymmetric information, liquidity, market efficiency.

*We are especially grateful to John Cochrane, Robert Engle, Joel Hasbrouck, and Ingrid Werner for their many useful comments. We also thank David Easley, Darren Roulstone, and the seminar participants at 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), and the NYU conference in honor of the nobel prize winner Robert Engle (NYC, September 30, 2004) 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

A fundamental question in economics is whether or not financial markets are efficient (Fama (1970, 1991)). If not, how inefficient are they and what are the determinants of the existing inefficiencies? Measuring stock market efficiency is of crucial importance to a variety of market participants, such as individual investors and portfolio managers, as well as regulators. In November 2000, the Security and Exchange Commission issued Rule 11 Ac. 1-5 requiring market venues to widely distribute execution quality statistics regarding their trades in electronic format.¹ From a regulator’s perspective, learning about market efficiency and its determinants can lead to more effective market designs. From an investor’s perspective, less efficient markets translate into higher rebalancing costs, higher risk, and, potentially, higher required and expected asset returns.² This paper proposes a new measure of market efficiency.

In a world with heterogeneous (informed and uninformed) agents, two equilibrium prices can be defined: the price that would prevail in absence of market frictions given publicly available information (the public information set) and the price that would prevail in absence of market frictions given both public and private information (the full-information set). The theoretical market microstructure literature has termed the former the “efficient price” and the latter the “full-information price.” Sequential and batch trade models posit that uninformed agents learn about private information from order flow (Kyle (1985) and Easley and O’Hara (1987), among others). Hence, in a rational ex-

¹The corresponding SEC regulation can be accessed at <http://www.sec.gov/rules/final/34-43590.htm>.

²When interpreting market frictions as being mainly determined by liquidity effects, as typically the case in the asset pricing literature, this statement relates to a substantial recent work devoted to assessing whether expected stock returns include a liquidity premium (Amihud and Mendelson (1986), Brennan and Subrahmanyam (1996), Brennan, Chordia, and Subrahmanyam (1998), Datar, Naik, and Radcliffe (1998), Fiori (2000), Hasbrouck (2003), and Pastor and Stambaugh (2003), among others)

pectation setting with learning on the part of the market participants, the efficient price converges to the full-information price. The eventual convergence of the efficient price to the full-information price is implied by Walras' tâtonnement process, namely the process by which the order flow leads to the full-information equilibrium price (Walras (1889)). Biais et al. (1995, 1999) offer empirical evidence of learning in the marketplace. Biais et al. (1999) quantify the speed of learning empirically. Vives (1995) and Germain et al. (1998) provide theoretical discussions of this issue.

In an asymmetric information setting, there are two sources of market inefficiency. First, transaction prices differ from the efficient prices (the equilibrium prices given public information) but tend to cluster around them. Deviations of transaction prices from efficient prices are driven by *market frictions*,³ such as price discreteness. These deviations are the focus of virtually all existing transaction cost measures.⁴ Second, efficient prices can differ from full-information prices (the equilibrium prices given public and private information).

This paper proposes a novel measure of market efficiency⁵ which accounts for *both* departures of transaction prices from efficient prices, as in much existing work on execution cost evaluation, *and* deviations of efficient prices from full-information prices. Specifically, we measure the positive distance between observed transaction prices and full-information prices. Our measure is termed "full-information transaction cost" or *FITC*. In the jargon of general economic theory, we characterize the extent of deviations of asset prices from strong-form market efficiency while more conventional measures using the efficient price as the benchmark price can be interpreted as measuring deviations of asset prices from semi-strong form market efficiency. In our framework, a nearly strong-form efficient market can thus be characterized as one in which transaction prices occur close to the full-information prices or, equivalently, have small *FITC's*. Fama (1991) writes: "Since there are surely positive information and trading costs, the extreme version of the market

³As in Stoll's 2000 presidential address to the Americal Finance Association we use the word "frictions" to define real as well as information-induced deviations of transaction prices from efficient prices. We expand on the determinants of frictions in Sections 2 and 4.

⁴Among others, *half-quoted spreads, effective spreads, realized half spreads, traded spreads, Roll's effective bid-ask measure*. The interested reader is referred to a recent special issue of the Journal of Financial Market (vol. 6, issue 3, pages 227 - 459) for a thorough discussion of developments on the subject.

⁵We follow Black (1986) in calling both the difference between the transaction price and the efficient price and the difference between the efficient price and the full-information price "sources" of market inefficiency.

efficiency hypothesis is surely false. Its advantage, however, is that it is a clean benchmark that allows me to sidestep the messy problem of deciding what are reasonable information and trading costs.” The objective of the present work is to quantify these costs through *FITC*'s.

Since the difference between the transaction price and the efficient price is generally estimated to evaluate “market quality” (see, for example, Hasbrouck (1993)) and the *FITC*'s account for this difference, as well as for the difference between efficient price and full-information price, one can also interpret the *FITC*'s as measures of market quality.

We view our approach as providing a bridge between the above mentioned literatures, namely the literature that studies the extent of deviations of transaction prices from “equilibrium prices” on the one hand and the literature devoted to assessing the rate at which market participants learn about the full-information price on the other hand. Differently from our approach, the former takes the efficient price, rather than the full-information price, as the equilibrium price of interest while the latter is typically not concerned with providing measurements of the distance between transaction prices and their full-information levels. This is the first paper, to our knowledge, which provides estimates of the magnitudes of both components.

Our identification procedure uses high-frequency data and relies on the different orders of magnitude of the components of the observed return data at high-frequency. The full-information price is expected to evolve rather “smoothly,” although with unpredictable variations given full-information, over time. Differently from the full-information price, the efficient price, as well as the market frictions, have the potential to be adjusted in response to the arrival of each transaction. Classic asymmetric information theory predicts that the uninformed agents learn about private information from order flow (O'Hara (1995)). Hence, meaningful revisions in the efficient price can arise no matter how close in time the transactions occur. Similarly, the random arrival of buyers and sellers combined with the discreteness inherent in financial prices imply analogous discrete increments to the microstructure noise component. In this setting, the observed high-frequency continuously-compounded return data are dominated by return components that are induced by the efficient price and market frictions since the underlying full-information returns evolve *relatively* more smoothly in the state space. Hence, our estimation procedure utilizes

sample moments of the *observed* high-frequency return data to learn about moments of the *unobserved FITC's*.

Our method is simple and robust. It is simple because we only require the computation of empirical moments of available high-frequency stock returns. It is robust because weak assumptions are sufficient for the method to provide consistent measurements of market efficiency through *FITC's*. The traditional taxonomy in the literature imputes deviations of the transaction price from the efficient price to operating (order-processing and inventory-carrying) costs and asymmetric information (see Bagehot (1971)). The deviations of the efficient price from the full-information price are due to learning. Our assumptions permit the deviations of the transaction prices from the full-information price 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), as well as learning in the marketplace.

Our empirical work provides measurements of the *FITC's* for a cross-section of S&P100 stocks. We find that the asymmetric information or “learning” component in the estimated *FITC's* is substantial. Importantly, we show that the magnitude of the expected distance between efficient prices and full-information prices can be as large as the distance between transaction prices and efficient prices or larger. Therefore, traditional measures of market quality which abstract from the difference between full-information prices and efficient prices have the potential to severely overstate the extent of actual market efficiency.

The paper proceeds as follows. Section 2 discusses the price formation mechanism. In Section 3 we present our nonparametric measure of market efficiency, the *FITC*. Section 4 provides *FITC* 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 *unobserved* efficient prices and *unobserved* 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

We start by assuming absence of market frictions. Deviations of the efficient price from the full-information price are, thus, the only source of inefficiency. In a competitive market with informed and uninformed (liquidity) traders, the informed agents' trading decisions carry information about the full-information value of the asset. Generally, it is not possible to fully distinguish between informed and uninformed traders. A decision to sell, for example, might signal that a trader is informed and aware of bad news about the asset. Alternatively, a decision to sell might simply be the result of a liquidity need on the part of an uninformed agent. Market participants cannot clearly infer which is the case. However, the trades provide information. This information is used by the market participants to update their beliefs about the asset's value given publicly available information and formulate equilibrium or "efficient" prices. 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, even in the absence of market frictions, the efficient price differs from the full-information price in general.

In practise, market frictions do exist (see Stoll's 2000 presidential address to the American Finance Association for recent discussions). The actual transaction prices differ from the efficient prices (the equilibrium prices given public information) but cluster around them. 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 \bar{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 $t \in [0, \bar{h}]$, denotes the number of transactions occurred over the period $[0, t]$. The following prices are logarithmic prices. We write the *unobserved* efficient price $p_{t_i}^e$ as a function of the *unobserved* full-information price p_{t_i} , namely

$$p_{t_i}^e = p_{t_i} + \eta_i^{asy}, \quad (1)$$

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 \tilde{p}_i corresponding to transaction i as

$$\tilde{p}_i = p_{t_i}^e + \eta_i^{fri}, \quad (2)$$

where $p_{t_i}^e$ 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

$$\tilde{p}_i = p_{t_i} + \eta_i, \quad (3)$$

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.”

Economic theory sheds light on the properties of the relevant components of the model, namely p_{t_i} and η_i . The full-information set, by definition, contains all information used by the market participants in their decisions to transact. Hence, the full-information price p_{t_i} is unaffected by the present order flow. With the exception of infrequent and unexpected news arrivals to the informed agents, the dynamic behavior of the full-information price is

best modelled as being rather “smooth.” We capture these features of the full-information price process by representing it as a discontinuous semi-martingale process with infrequent jumps.

Assumption 1 (The full-information price.)

- (1) *The full-information logarithmic price process p_t is a discontinuous semi-martingale. 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, and $K_t = \int_0^t (J_s dZ_s - \mu_j \lambda_s ds)$ 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.*

In asset-pricing models, the semimartingale property of the price process from Assumption 1(2) is a necessary condition for the absence of arbitrage opportunities (Duffie (1990), for example). In information-based models of price determination, the full-information price is a martingale, i.e., $A_t = 0$. This case is a sub-case of our more general set-up. We also allow for the presence of stochastic volatility. Specifically, under Assumption 1(2), the volatility process can display long-memory properties, diurnal effects, jumps, 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 turn to the market effects η . 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. Hence, the efficient price process 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. This is what the presence of the term η^{asy} in Eq. (1) accomplishes.

The term η^{fri} reflects conventional microstructure frictions. The presence of separate prices for buyers and sellers and price discreteness, alone, suggest that the changes in the frictions from trade to trade are discrete in nature.

Assumption 2 (The market effects.)

- (1) *The microstructure 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 $j = 1, \dots, k < \infty$ and $\mathbf{E}(\eta\eta_{-j}) = 0$ for $j > k$.*
- (3) $\sum_{s=0}^{\infty} |\lambda_s| < \infty$ with $\lambda_s = \mathbf{E}[(\varepsilon_s \varepsilon_{s-j} - \mathbf{E}(\varepsilon \varepsilon_{-j})) (\varepsilon \varepsilon_{-j} - \mathbf{E}(\varepsilon \varepsilon_{-j}))]$ where $\varepsilon = \eta - \eta_{-1}$ for $j = 0, 1, \dots, k < \infty$

The market effects are stationary (Assumption 2(1)). Their dependence structure 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)). This property permits us to 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 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. Finally, we allow the microstructure effects to be correlated with the unobservable full-information price process as required by asymmetric information and learning on the part of the market participants.

Our interest will be in measuring the standard deviation of η , i.e., σ_η , a natural measure of distance between the actual transaction prices and the prices that would

prevail given full-information. Hasbrouck (1993) was the first, to our knowledge, to focus on the standard deviation as a measure of market quality as we do in this paper. His work differs from ours, however, in that his reference price is the efficient price, not the full-information price. His results provide a lower bound for his object of interest, namely σ_η^{fri} . We now turn to our identification procedure.

3 Measuring full-information transaction costs

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

$$\tilde{r}_i = r_{t_i} + \varepsilon_i, \quad (5)$$

where $\tilde{r}_i = \tilde{p}_i - \tilde{p}_{i-1}$ 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),*

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

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,

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

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

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

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.

The intuition behind our identification procedure is as follows. The market effects in the observed returns (ε) are $O_p(1)$ (see Assumption (2)). At high frequencies, the full-information component of the observed returns is of order $O_p\left(\sqrt{\max_{1 \leq i \leq N(\bar{h})} |t_i - t_{i-1}|}\right)$ (see Assumption (1)), where $\max_{1 \leq i \leq N(\bar{h})} |t_i - t_{i-1}|$ is the maximum duration between price updates. Of course, the maximum duration is small when using data sampled at the frequencies at which transactions arrive in practise (see Table I). In light of these observations, the market effects in the return process (ε) dominate the full-information return component (r) at high frequencies. Hence, we can use sample moments of the *observed* high-frequency return data to identify moments of the *unobserved* market effects (ε) by employing the informational content of high-frequency return data.

Finally, we note that our asymptotic design, which hinges on an increasing number of transactions (i.e., $N(\bar{h}) \rightarrow \infty$) over a fixed interval of time (\bar{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 below lays out the estimator and its asymptotic behavior.

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

$$\hat{\sigma}_\eta = \sqrt{\left(\frac{k+1}{2}\right) \left(\frac{\sum_{i=1}^{N(\bar{h})} \tilde{r}_i^2}{N(\bar{h})}\right) + \sum_{s=0}^{k-1} (s+1) \left(\frac{\sum_{i=k-s+1}^{N(\bar{h})} \tilde{r}_i \tilde{r}_{i-k+s}}{N(\bar{h}) - k + s}\right)} \xrightarrow{p}_{N(\bar{h}) \rightarrow \infty} \sigma_\eta. \quad (9)$$

Proof. *See Appendix.*

In what follows, we will use the convention of referring to estimates obtained by employing the estimator in Eq. (9) (and Eq. (14) below) as *FITC's*. The estimator

is local in nature and defined over a single, generically-specified, period \bar{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.

Recently, a growing literature has attempted to accommodate market microstructure noise effects on realized volatility estimates. In the realized volatility literature, sums of squared continuously-compounded returns over a period are used to estimate the quadratic variation of the underlying efficient price (see the review paper by Andersen et al. (2003) for discussions). The approaches that allow for general dependence in the microstructure noise and are closest to our approach are Bandi and Russell (2003a), Hansen and Lunde (2003) and Oomen (2004). While both our current work and this realized volatility literature require the modeling and estimation of features of the noise components, the objects of econometric interest are different. With the exception of some aspects of the analysis in Bandi and Russell (2003a), the identification techniques are also completely different. Hansen and Lunde (2003) and, subsequently, Oomen (2004) (but in the context of a pure jump process for the underlying efficient price) rely on a clever cancellation technique using a HAC type estimator to purge realized volatility estimates of their market microstructure noise-induced bias component. The bias there is given by the sum of the variance of the noise in return component and the covariance between the noise in return and the underlying efficient price. Clearly, our measure requires estimation of the variance of the noise (or, more precisely in our model, combined market effect) component contained in the price process only. The techniques of Hansen and Lunde (2003) and Oomen (2004) are not able to separately identify this variance under our assumptions. Here, we rely on limiting arguments and show consistency of our estimator. Our identification procedure hinges on the different orders of the full-information returns and market effects. Our estimator is not a HAC type estimator in that the weights are derived specifically

for estimating the variance of the market effects in the 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, consider Assumptions 3 and 4 below.

Assumption 3. *Assume η is normal.*

Assumption 4. *Assume*

$$\eta_i = sQ_i \quad \forall i = 1, \dots, N(\bar{h}), \quad (10)$$

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

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

$$0.7979\hat{\sigma}_\eta \xrightarrow[N(\bar{h}) \rightarrow \infty]{P} \mathbf{E}(|\tilde{p} - p|), \quad (11)$$

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

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

$$\widehat{\sigma}_\eta \xrightarrow[N(\bar{h}) \rightarrow \infty]{p} s = \mathbf{E}(|\tilde{p} - p|), \quad (12)$$

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 $Cov(\tilde{r}, \tilde{r}_{-1}) = -2s^2$. Hence, a consistent estimate of s is given by $2\sqrt{\widehat{\mathbf{E}}(\tilde{r}, \tilde{r}_{-1})}$. 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 A finite sample bias-correction

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, under the assumption that $A_t = 0$, which implies, as in conventional market microstructure theory, unpredictability of the full-information price process, we can write

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

The finite sample contaminations are α and β . While one can characterize α by using a standardized (by $\frac{1}{N(\bar{h})}$) estimate of the variance of the full-information price process over

\bar{h} , the identification of β is much more allusive.

We propose bias-correcting $\hat{\sigma}_\eta$ by subtracting a model-free estimate of α obtained by using realized variance as in Andersen et al. (2003) and Barndorff-Nielsen and Shephard (2002). We estimate the h -period variance by summing intraperiod squared continuously-compounded returns. The optimal number of intraperiod observations is chosen to minimize the conditional mean-squared error of the variance estimator in the presence of market microstructure effects as suggested by Bandi and Russell (2003a,b). The resulting estimator is:

$$\bar{\sigma}_\eta = \sqrt{\left(\frac{k+1}{2}\right) \left(\frac{\sum_{i=1}^{N(\bar{h})} \tilde{r}_i^2}{N(\bar{h})} - \hat{\alpha}\right) + \sum_{s=0}^{k-1} (s+1) \left(\frac{\sum_{i=k-s+1}^{N(\bar{h})} \tilde{r}_i \tilde{r}_{i-k+s}}{N(\bar{h}) - k + s}\right)} \quad (14)$$

where

$$\hat{\alpha} = \frac{1}{N(\bar{h})} \left(\sum_{j=1}^{\overline{M}(\bar{h})} \tilde{r}_j^2\right), \quad (15)$$

and the optimal number $\overline{M}(\bar{h})$ of equispaced intraperiod returns \tilde{r}_j is chosen on the basis of the procedure proposed by Bandi and Russell (2003a,b). We use $\bar{\sigma}_\eta$ in what follows.

The estimated biases $\hat{\alpha}$ are very small for the S&P 100 stocks in our sample. Specifically, the average and maximum bias are 7.5% and 10.5% of the corresponding *FITC*'s, respectively.

Contrary to virtually all existing approaches to transaction cost estimation, our approach does not require independence between the full-information price process and the market effects to provide consistent estimates of market quality measures, namely for either $\hat{\sigma}_\eta$ or $\bar{\sigma}_\eta$ to consistently estimate σ_η in our framework. However, when independence or uncorrelatedness between the full-information price process and η is not satisfied, our estimates might contain a finite sample bias component, β , which cannot be characterized. Even if β is non zero, this bias will be small for stocks with high trading rates. The stocks in our sample are traded at very high frequencies and we therefore expect this bias, just like the bias represented by α , to be small. Our asymptotic approximation is of course bound to improve as stocks get traded more and more frequently in the future.

4 The *FITC*'s of the S&P 100 stocks

The data analyzed 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. (14) 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

$$Q_t(\tilde{p}_t - m_t), \tag{16}$$

where Q_t is an indicator equal to 1 (-1) for buyer (seller) initiated trades, \tilde{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 (the TAQ database that we use in the paper, for instance) do not contain information about whether a trade is buyer or seller-initiated. Lee and Ready (1991) and Ellis et al. (2000), among others, propose algorithms intended to classify trades as buyer or seller-initiated simply on the basis of transaction prices and quotes. While these algorithms perform reasonably well, they have the potential to misclassify 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, the 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 the 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 the 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 Figs. 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 distributions of the *FITC's* are 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.

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*),⁶ 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

⁶As 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).

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 amount 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 -0.153 and a t-stat of -5.29 supporting the predictions of the operating-cost theory of friction determination. The variable *lturn* is significantly positive with an elasticity of 0.16 and a t-stat of about 8.85 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 0.55 with a t-stat of 12.44 . As expected, the coefficient on *lprice* is negative (-0.15) and highly statistically significant with a t-stat of -4.79 . The (statistically significant) positive sign of the estimated coefficient on *nyse* (0.86) is somewhat surprising at first. It is widely believed that the decentralized nature of NASDAQ leaves the dealer more ex-

posed 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 2001.⁷ 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 0.184 with a t-stat of 2.3. 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

⁷We thank Soeren Hvidkjaer for making the *PIN* measures available to us.

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.⁸ 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 -0.11 with a t-stat of -2.66 . 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 $FITC$'s?

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 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

$$\sigma_{\eta}^2 = \sigma_{\eta^{asy}}^2 + \sigma_{\eta^{fri}}^2 + 2\sigma_{\eta^{asy}\eta^{fri}}, \quad (17)$$

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

⁸The number of analysts is obtained from the Institutional Brokers Estimation System (I/B/E/S) database.

$$\sigma_{\eta}^2 - \sigma_{\eta^{fri}}^2 = \sigma_{\eta^{asy}}^2 + 2\sigma_{\eta^{asy}\eta^{fri}}. \quad (18)$$

Proposition. *If the FITC's contain a pure asymmetric information component, then the cross-sectional variation of the difference $\sigma_{\eta}^2 - \sigma_{\eta^{fri}}^2$ should be explained by variables that 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 $FITC^2 - \hat{\sigma}_{\eta^{fri}}^2$ 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 $FITC^2 - \hat{\sigma}_{\eta^{fri}}^2$ 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 $FITC^2 - \hat{\sigma}_{\eta^{fri}}^2$ 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

$$\mathbf{E} |\tilde{p} - p| \leq \mathbf{E} |\tilde{p} - p^e| + \mathbf{E} |p^e - p| \quad (19)$$

where, as earlier, \tilde{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

$$\mathbf{E} |\tilde{p} - p| - \mathbf{E} |\tilde{p} - p^e| \leq \mathbf{E} |p^e - p|. \quad (20)$$

Under a log-normality assumption for the market effect η , the term $\mathbf{E} |\tilde{p} - p|$ 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 $\mathbf{E} |\tilde{p} - p^e|$ can be estimated based on transaction prices and midpoints. Hence, the difference $\mathbf{E} |\tilde{p} - p| - \mathbf{E} |\tilde{p} - p^e|$ 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 4 we plot the histogram of the estimated lower bounds as a percentage of the *FITC* estimates. The mean value is about 30%, the maximum value is about 50%. Hence, the asymmetric information component in the *FITC's* is substantial. In Figure 5 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 70% but values as large as 140% 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 (or market efficiency) 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 efficiency. 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, the method 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. We argue that it is important to account for all of these assumptions if market efficiency is believed to be affected by realistic operating and adverse-selection costs as implied by accepted theories of market friction determination as well as by learning. To our knowledge, no existing approach to measuring market quality allows for *(i)* through *(iv)*.

Using estimates of 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 market place, can be as large as the departures of the transaction prices from the efficient prices, as induced by standard frictions.

Much is left for future work. While the present paper focuses on unconditional, cross-sectional, measures of market efficiency, our tools can be employed to learn about the conditional properties of the full-information transaction costs. Such properties are expected to provide valuable information about the genuine market dynamics. Furthermore, 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 relationship 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 provides a natural framework to bridge the two arguments in the study of the cross-sectional dependence between expected stock returns and our more general notion of transactions costs. Research on both subjects is being conducted by the authors and will be reported in later work (Bandi and Russell (2004a,b)).

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

$$\mathbf{E}(\varepsilon\varepsilon_{-j}) = \mathbf{E}\left((\eta - \eta_{-1})(\eta_{-j} - \eta_{-(j+1)})\right) \quad (21)$$

$$= \mathbf{E}(\eta\eta_{-j}) - \mathbf{E}(\eta\eta_{-(j+1)}) - \mathbf{E}(\eta_{-1}\eta_{-j}) + \mathbf{E}(\eta_{-1}\eta_{-(j+1)}) \quad (22)$$

$$= 2\mathbf{E}(\eta\eta_{-j}) - \mathbf{E}(\eta_{-1}\eta_{-j}) - \mathbf{E}(\eta\eta_{-(j+1)}). \quad (23)$$

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

$$\mathbf{E}(\varepsilon\varepsilon_{-j}) = 2\mathbf{E}(\eta\eta_{-j}) - \mathbf{E}(\eta_{-1}\eta_{-j}) \quad (24)$$

when $j = k$ and

$$\mathbf{E}(\varepsilon\varepsilon_{-j}) = 2\mathbf{E}(\eta\eta_{-j}) - \mathbf{E}(\eta_{-1}\eta_{-j}) - \mathbf{E}(\eta\eta_{-(j+1)}) \quad (25)$$

for $1 \leq j < k$. We now plug Eq. (24) and Eq. (25) into the right-hand side of the squared version of Eq. (6) and obtain

$$\begin{aligned} & \left(\frac{1+k}{2}\right) \mathbf{E}(\varepsilon^2) + \sum_{s=0}^{k-1} (s+1) \mathbf{E}(\varepsilon\varepsilon_{-(k-s)}) \\ = & \left(\frac{1+k}{2}\right) [2\mathbf{E}(\eta^2) - 2\mathbf{E}(\eta\eta_{-1})] \\ & + \sum_{s=1}^{k-1} (s+1) [2\mathbf{E}(\eta\eta_{-(k-s)}) - \mathbf{E}(\eta_{-1}\eta_{-(k-s)}) - \mathbf{E}(\eta\eta_{-((k-s)+1)})] \\ & + 2 [\mathbf{E}(\eta\eta_{-k}) - \mathbf{E}(\eta_{-1}\eta_{-k})] \quad (26) \\ = & (1+k) (\mathbf{E}(\eta^2) - \mathbf{E}(\eta\eta_{-1})) \\ & + k [2\mathbf{E}(\eta\eta_{-1}) - \mathbf{E}(\eta^2) - \mathbf{E}(\eta\eta_{-2})] \\ & + (k-1) [2\mathbf{E}(\eta\eta_{-2}) - \mathbf{E}(\eta\eta_{-1}) - \mathbf{E}(\eta\eta_{-3})] \\ & + (k-2) [2\mathbf{E}(\eta\eta_{-3}) - \mathbf{E}(\eta\eta_{-2}) - \mathbf{E}(\eta\eta_{-4})] \\ & + (k-3) [2\mathbf{E}(\eta\eta_{-4}) - \mathbf{E}(\eta\eta_{-3}) - \mathbf{E}(\eta\eta_{-5})] \\ & + \dots \\ & + 3[2\mathbf{E}(\eta\eta_{-k+2}) - \mathbf{E}(\eta\eta_{-k+3}) - \mathbf{E}(\eta\eta_{-k+1})] \\ & + 2[2\mathbf{E}(\eta\eta_{-k+1}) - \mathbf{E}(\eta\eta_{-k+2}) - \mathbf{E}(\eta\eta_{-k})] \\ & + [2\mathbf{E}(\eta\eta_{-k}) - \mathbf{E}(\eta\eta_{-k+1})] \quad (27) \end{aligned}$$

Finally, we notice that Eq. (27) is equal to $\mathbf{E}(\eta^2)$. This proves the stated result. ■

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

$$\frac{\sum_{i=j+1}^{N(\bar{h})} \tilde{r}_i \tilde{r}_{i-j}}{N(\bar{h}) - j} = \underbrace{\frac{\sum_{i=j+1}^{N(\bar{h})} r_{t_i} r_{t_{i-j}}}{N(\bar{h}) - j}}_{\alpha} + \underbrace{\frac{\sum_{i=j+1}^{N(\bar{h})} r_{t_i} \varepsilon_{i-j}}{N(\bar{h}) - j}}_{\beta} + \underbrace{\frac{\sum_{i=j+1}^{N(\bar{h})} \varepsilon_i r_{t_{i-j}}}{N(\bar{h}) - j}}_{\gamma} + \underbrace{\frac{\sum_{i=j+1}^{N(\bar{h})} \varepsilon_i \varepsilon_{i-j}}{N(\bar{h}) - j}}_{\zeta}. \quad (28)$$

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

$$\alpha \leq \frac{\left(\sum_{i=j+1}^{N(\bar{h})} r_{t_i}^2\right)^{1/2} \left(\sum_{i=j+1}^{N(\bar{h})} r_{t_{i-j}}^2\right)^{1/2}}{N(\bar{h}) - j} \quad (29)$$

$$= \frac{\left(\sum_{i=1}^{N(\bar{h})} r_{t_i}^2 - \sum_{i=1}^j r_{t_i}^2\right)^{1/2} \left(\sum_{i=1}^{N(\bar{h})} r_{t_i}^2 - \sum_{i=N(\bar{h})-j+1}^{N(\bar{h})} r_{t_i}^2\right)^{1/2}}{N(\bar{h}) - j} \quad (30)$$

$$\leq \frac{\left([p]_0^{\bar{h}} + o_p(1) + jO_p\left(\pi^{N(\bar{h})}\right)\right)^{1/2} \left([p]_0^{\bar{h}} + o_p(1) + jO_p\left(\pi^{N(\bar{h})}\right)\right)^{1/2}}{N(\bar{h}) - j} \quad (31)$$

$$\stackrel{P}{\underset{N(\bar{h}) \rightarrow \infty}{\rightarrow}} 0, \quad (32)$$

where $[p]_0^{\bar{h}} = [M]_0^{\bar{h}} + \sum_{0 < s \leq \bar{h}} (\Delta p_s)^2 = \int_0^{\bar{h}} \sigma_s^2 ds + \sum_{0 < s \leq \bar{h}} (\Delta p_s)^2$ is the quadratic variation of the underlying logarithmic price process. It is noted that Eq. (29) derives from the Cauchy's inequality while Eq. (31) 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),

$$\sum_{i=1}^{N(\bar{h})} (p_{t_i} - p_{t_{i-1}})^2 \stackrel{P}{\underset{N(\bar{h}) \rightarrow \infty}{\rightarrow}} [p]_0^{\bar{h}} = [M]_0^{\bar{h}} + \sum_{0 < s \leq \bar{h}} (\Delta p_s)^2 \quad (33)$$

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

$$\mathbf{1}\{|\zeta - \mathbf{E}(\varepsilon \varepsilon_{-j})| > \delta\} \leq \frac{|\zeta - \mathbf{E}(\varepsilon \varepsilon_{-j})|}{\delta} \quad (34)$$

$$\leq \frac{(\zeta - \mathbf{E}(\varepsilon \varepsilon_{-j}))^2}{\delta^2}, \quad (35)$$

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. (35), we obtain

$$P\{|\zeta - \mathbf{E}(\varepsilon \varepsilon_{-j})| > \delta\} \leq \frac{\mathbf{E}(\zeta - \mathbf{E}(\varepsilon \varepsilon_{-j}))^2}{\delta^2}. \quad (36)$$

By Assumption 2(1), we note that

$$\begin{aligned} & \mathbf{E}(\zeta - \mathbf{E}(\varepsilon \varepsilon_{-j}))^2 \\ &= \frac{1}{(N(\bar{h}) - j)^2} \mathbf{E} \left(\sum_{i=j+1}^{N(\bar{h})} (\varepsilon_i \varepsilon_{i-j} - \mathbf{E}(\varepsilon \varepsilon_{-j})) \right)^2 \end{aligned} \quad (37)$$

$$= \frac{1}{(N(\bar{h}) - j)^2} \left((N(\bar{h}) - j) \lambda_0 + 2(N(\bar{h}) - j - 1) \lambda_1 + \dots + 2\lambda_{N(\bar{h})-j} \right), \quad (38)$$

$$\leq \frac{1}{(N(\bar{h}) - j)} \left(|\lambda_0| + 2|\lambda_1| + \dots + 2|\lambda_{N(\bar{h})-j}| \right), \quad (39)$$

where

$$\lambda_s = \mathbf{E}[(\varepsilon_s \varepsilon_{s-j} - \mathbf{E}(\varepsilon \varepsilon_{-j}))(\varepsilon \varepsilon_{-j} - \mathbf{E}(\varepsilon \varepsilon_{-j}))] \quad (40)$$

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

$$P\{|\zeta - \mathbf{E}(\varepsilon \varepsilon_{-j})| > \varepsilon\} \leq \frac{1}{\delta^2(N(\bar{h}) - j)} \left(2|\lambda_0| + 2|\lambda_1| + \dots + 2|\lambda_{N(\bar{h})-j}| - |\lambda_0| \right) \xrightarrow[N(\bar{h}) \rightarrow \infty]{} 0$$

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

$$\beta \leq \frac{\left(\sum_{i=j+1}^{N(\bar{h})} r_{t_i}^2 \right)^{1/2} \left(\sum_{i=j+1}^{N(\bar{h})} \varepsilon_{i-j}^2 \right)^{1/2}}{N(\bar{h}) - j} \quad (41)$$

$$= \frac{1}{\sqrt{N(\bar{h}) - j}} \left([p]_0^{\bar{h}} + o_p(1) + j O_p(\pi^{N(\bar{h})}) \right)^{1/2} \left(\frac{\sum_{i=j+1}^{N(\bar{h})} \varepsilon_{i-j}^2}{N(\bar{h}) - j} \right)^{1/2} \quad (42)$$

$$= \frac{1}{\sqrt{N(\bar{h}) - j}} O_p(1) (\mathbf{E}(\varepsilon^2) + o_p(1)) \xrightarrow[N(\bar{h}) \rightarrow \infty]{p} 0, \quad (43)$$

where Eq. (41) follows from Cauchy's inequality and the convergence in probability of $\frac{\sum_{i=j+1}^{N(\bar{h})} \varepsilon_{i-j}^2}{N(\bar{h}) - j}$ 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

$$\gamma \leq \frac{\left(\sum_{i=j+1}^{N(\bar{h})} \varepsilon_i^2 \right)^{1/2} \left(\sum_{i=j+1}^{N(\bar{h})} r_{t_{i-j}}^2 \right)^{1/2}}{N(\bar{h})} \quad (44)$$

$$= \frac{1}{\sqrt{N(\bar{h})}} O_p(1) (\mathbf{E}(\varepsilon^2) + o_p(1)) \xrightarrow[N(\bar{h}) \rightarrow \infty]{p} 0. \quad (45)$$

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] Bagehot, W. (1971). The only game in town. *Financial Analysts Journal*, 27, 12-14.
- [5] Bandi, F.M., and J.R. Russell (2003a). Realized variance, microstructure noise, and optimal sampling. *Working paper*.
- [6] Bandi, F.M., and J.R. Russell (2003b). Separating microstructure noise from volatility. *Working paper*.
- [7] Bandi, F.M., and J.R. Russell (2004a). Forecasting transaction costs. *Working paper*.
- [8] Bandi, F.M., and J.R. Russell (2004b). Expected stock returns and full-information transaction costs. *Working paper*.
- [9] 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.
- [10] Brennan, M.J., T. Chordia and A. Subrahmanyam (1998). Alternative factor specifications, security characteristics, and the cross-section of expected stock returns. *Journal of Financial Economics*, 49, 345-373.
- [11] Brennan, M.J., and A. Subrahmanyam (1995). Investment analysis and price formation in securities markets. *Journal of Financial Economics*, 38, 361-381.
- [12] Brennan, M.J., and A. Subrahmanyam (1996). Market microstructure and asset pricing: on the compensation for illiquidity in stock returns. *Journal of Financial Economics*, 41, 441-464.

- [13] Bessembinder, H. (2003). Issues in assessing trade execution costs. *Journal of Financial Markets*, 6, 233-257.
- [14] 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.
- [15] 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.
- [16] Black, F. (1986). Noise. *Journal of Finance*, 41, 529-543.
- [17] 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.
- [18] Copeland, T., and D. Galai (1983). Information effects in the bid-ask spread. *Journal of Finance*, 38, 1457-1469.
- [19] Datar, V.T., N.Y. Naik and R. Radcliffe (1998). Liquidity and stock returns: an alternative test. *Journal of Financial Markets*, 1, 203-219.
- [20] Demsetz, H. (1968). The cost of transacting. *Quarterly Journal of Economics*, 82, 33-53.
- [21] Duffie, D. (1990). *Dynamic asset pricing*. Princeton Press.
- [22] Easley D., and M. O'Hara (1987). Price, trade size, and information in securities markets. *Journal of Financial Economics*, 19, 69-90.
- [23] Easley D., and M. O'Hara (1992). Time and the process of security price adjustment. *Journal of Finance*, 47, 577-605.
- [24] 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.
- [25] Easley D., S. Hvidkjaer, and M. O'Hara (2002). Is information risk a determinant of asset returns? *Journal of Finance*, 57, 2185-2221.

- [26] 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.
- [27] Easley D., M. O'Hara and J. Paperman (1998). Financial Analysts and information-based trade. *Journal of Financial Markets*, 1, 175-201.
- [28] Easley D., N. Kiefer, M. O'Hara and J. Paperman (1996). Liquidity, information and infrequently traded stocks. *Journal of Finance*, 1, 1405-1436.
- [29] 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.
- [30] Fama, E. (1970). Efficient capital markets: a review of theory and empirical work. *Journal of Finance*, 25, 383-417.
- [31] Fama, E. (1991). Efficient capital markets: II. *Journal of Finance*, 25, 1575-1617.
- [32] Fiori, F. (2000). Liquidity premia in equity markets: an investigation into the characteristics of liquidity and trading activity. *Working paper*.
- [33] Garbade, K., and Z. Lieber (1977). On the independence of transactions on the NYSE. *Journal of Banking and Finance*, 1, 151-172.
- [34] 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.
- [35] Garman, M. (1976). Market microstructure. *Journal of Financial Economics*, 3, 257-275.
- [36] Germain, L., N. Meddahi and E. Renault (1996). The speed of learning in financial markets. *Working paper*.
- [37] Glosten, L.R. (1987). Components of the bid-ask spread and the statistical properties of transaction prices. *Journal of Finance*, 42, 1293-1308.
- [38] 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.

- [39] Hansen, P.R. and A. Lunde (2003). An unbiased measure of realized variance. *Working paper*.
- [40] Hasbrouck, J. (1993). Assessing the quality of a security market: a new approach to transaction cost measurement. *Review of Financial Studies*, 6, 191-212.
- [41] Hasbrouck, J. (1999). Liquidity in the future pits: inferring market dynamics from incomplete data. *Working paper*.
- [42] Hasbrouck, J. (2003). Trading costs and returns for U.S. securities: the evidence from daily data. *Working paper*.
- [43] 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.
- [44] Ho, T., and H.R. Stoll (1981). Optimal dealer pricing under transactions and return uncertainty. *Journal of Financial Economics*, 9, 47-73.
- [45] Kyle, A. (1985). Continuous auctions and insider trading. *Econometrica*, 53, 1315-1335.
- [46] Lee, C. and M. Ready (1991). Inferring trade direction from intraday data. *Journal of Finance*, 46, 733-746.
- [47] Loeb, T.F. (1983). Trading cost: the critical link between investment information and results. *Financial Analysts Journal*, 39, 39-43.
- [48] O'Hara, M. (1995). *Market Microstructure Theory*. Blackwell Publishers Ltd.
- [49] Oomen, R.C.A. (2004). Properties of bias corrected realized variance in calendar time and business time. *Working paper*.
- [50] Pastor, L. and R.F. Stambaugh (2003). Liquidity risk and expected stock returns. *Journal of Political Economy*, 111, 642-685.
- [51] Perold, A (1988). The implementation shortfall. *Journal of Portfolio Management*, 14, 4-9.

- [52] 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.
- [53] Protter, P. (1995). *Stochastic Integration and Differential Equations*. Springer, Berlin.
- [54] Roll, R. (1984). A simple measure of the effective bid-ask spread in an efficient market. *Journal of Finance*, 39, 1127-1139.
- [55] Roulstone, D.T. (2003). Analysts following and market liquidity. *Contemporary Accounting Research*, 20, 551-578.
- [56] Smidt, S. (1971). Which road to an efficient stock market: free competition or regulated monopoly? *Financial Analysts Journal*, 27, 64-69.
- [57] Stoll, H.R. (1989). Inferring the components of the bid-ask spread: theory and empirical tests. *Journal of Finance*, 44, 115-134.
- [58] Stoll, H.R. (2000). Friction. *Journal of Finance*, 55, 1479-1514.
- [59] Tinic, S. (1972). The economics of liquidity services. *Quarterly Journal of Economics*, 86, 79-93.
- [60] Vives, X. (1995). The speed of information revelation in a financial market mechanism. *Journal of Economic Theory*, 67, 178-204.
- [61] Walras, L. (1889). *Eléments d'économie politique pure, ou théorie de la richesse sociale*. Second Edition, Lausanne, Rouge.

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 (in percentage values), 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	0.112%	\$0.0405	0.055%	\$0.0199
AEP	21.64	\$41.77	0.105%	\$0.0438	0.043%	\$0.0182
AES	15.79	\$7.75	0.582%	\$0.0451	0.274%	\$0.0212
AIG	10.63	\$72.89	0.092%	\$0.0667	0.043%	\$0.0313
ALL	15.42	\$33.92	0.126%	\$0.0426	0.054%	\$0.0184
AMGN [*]	1.32	\$57.82	0.043%	\$0.0249	0.030%	\$0.0172
AOL	6.45	\$25.17	0.136%	\$0.0343	0.079%	\$0.0199
ATI	64.87	\$15.64	0.267%	\$0.0417	0.089%	\$0.0140
AVP	21.81	\$49.10	0.102%	\$0.0500	0.039%	\$0.0194
AXP	8.80	\$34.13	0.109%	\$0.0374	0.054%	\$0.0184
BA	10.88	\$43.52	0.103%	\$0.0450	0.050%	\$0.0219
BAC	7.25	\$61.20	0.081%	\$0.0496	0.042%	\$0.0256
BAX	17.49	\$55.33	0.099%	\$0.0547	0.043%	\$0.0239
BCC	31.58	\$34.98	0.157%	\$0.0549	0.058%	\$0.0202
BDK	23.95	\$43.50	0.123%	\$0.0536	0.043%	\$0.0189
BHI	14.56	\$34.53	0.149%	\$0.0514	0.061%	\$0.0211
BMJ	9.02	\$45.08	0.092%	\$0.0413	0.044%	\$0.0197
BNI	24.83	\$27.95	0.145%	\$0.0404	0.063%	\$0.0175
BUD	20.15	\$48.69	0.092%	\$0.0448	0.042%	\$0.0207
C	6.00	\$44.43	0.096%	\$0.0427	0.053%	\$0.0234
CCU	12.11	\$47.06	0.131%	\$0.0618	0.055%	\$0.0259
CI	20.13	\$92.46	0.131%	\$0.1210	0.051%	\$0.0471
CL	15.20	\$55.64	0.086%	\$0.0480	0.037%	\$0.0207
CPB	25.98	\$27.02	0.158%	\$0.0427	0.072%	\$0.0195
CSC	18.84	\$47.07	0.163%	\$0.0766	0.063%	\$0.0296
CSCO [*]	0.45	\$16.69	0.052%	\$0.0086	0.051%	\$0.0086
DAL	17.24	\$32.76	0.154%	\$0.0504	0.060%	\$0.0197
DD	10.83	\$45.08	0.095%	\$0.0426	0.043%	\$0.0194
DIS	8.59	\$23.29	0.116%	\$0.0270	0.057%	\$0.0133
DOW	17.33	\$30.05	0.131%	\$0.0393	0.057%	\$0.0171
EK	17.05	\$29.24	0.139%	\$0.0406	0.058%	\$0.0170
EMC	7.91	\$13.42	0.181%	\$0.0243	0.121%	\$0.0163
EP	15.48	\$37.01	0.162%	\$0.0600	0.076%	\$0.0283
ETR	26.96	\$41.08	0.116%	\$0.0478	0.043%	\$0.0177
EXC	22.35	\$50.00	0.123%	\$0.0615	0.046%	\$0.0231
F	12.29	\$14.69	0.121%	\$0.0178	0.067%	\$0.0099
FDX	14.68	\$54.75	0.100%	\$0.0548	0.044%	\$0.0239
G	18.77	\$33.23	0.121%	\$0.0402	0.050%	\$0.0165
GD	17.40	\$89.13	0.100%	\$0.0889	0.047%	\$0.0416
GE	4.73	\$37.49	0.072%	\$0.0269	0.044%	\$0.0167
GM	14.07	\$51.65	0.085%	\$0.0437	0.031%	\$0.0163
GS	8.73	\$82.27	0.096%	\$0.0787	0.047%	\$0.0386
HAL	11.38	\$15.21	0.199%	\$0.0302	0.101%	\$0.0153
HCA	16.62	\$42.31	0.123%	\$0.0522	0.051%	\$0.0215
HD	7.38	\$50.40	0.078%	\$0.0394	0.038%	\$0.0194
HET	29.73	\$38.34	0.169%	\$0.0649	0.073%	\$0.0281
HIG	17.86	\$65.73	0.116%	\$0.0760	0.047%	\$0.0312
HNZ	20.73	\$40.98	0.098%	\$0.0401	0.040%	\$0.0165
HON	12.73	\$34.34	0.193%	\$0.0664	0.067%	\$0.0228
HWP	9.97	\$20.57	0.126%	\$0.0259	0.062%	\$0.0128
IBM	6.58	\$102.81	0.074%	\$0.0765	0.038%	\$0.0393
INTC [*]	0.52	\$31.74	0.040%	\$0.0126	0.037%	\$0.0116
IP	11.46	\$42.92	0.094%	\$0.0405	0.044%	\$0.0187
JNJ	10.54	\$57.89	0.076%	\$0.0438	0.041%	\$0.0237

Symbol	Duration	Avg Price	FITC (%)	FITC (\$)	ESpd. (%)	Espd. (\$)
JPM	7.91	\$29.83	0.153%	\$0.0456	0.076%	\$0.0226
KO	11.28	\$46.32	0.086%	\$0.0400	0.039%	\$0.0181
LEH	11.12	\$58.99	0.128%	\$0.0757	0.050%	\$0.0297
LTD	21.50	\$17.60	0.168%	\$0.0295	0.067%	\$0.0118
LU	10.86	\$5.71	0.210%	\$0.0120	0.140%	\$0.0080
MAY	24.32	\$35.55	0.146%	\$0.0518	0.051%	\$0.0180
MCD	12.42	\$26.81	0.100%	\$0.0267	0.051%	\$0.0137
MDT	10.84	\$47.01	0.091%	\$0.0426	0.042%	\$0.0200
MEDI	3.73	\$40.80	0.098%	\$0.0399	0.059%	\$0.0241
MER	7.14	\$47.63	0.115%	\$0.0550	0.061%	\$0.0288
MMM	12.16	\$114.96	0.095%	\$0.1088	0.039%	\$0.0444
MO	10.07	\$51.28	0.069%	\$0.0353	0.038%	\$0.0196
MRK	10.23	\$59.96	0.074%	\$0.0445	0.036%	\$0.0217
MSFT	0.55	\$60.16	0.029%	\$0.0174	0.026%	\$0.0155
MWD	6.97	\$49.72	0.106%	\$0.0525	0.054%	\$0.0269
NSC	20.77	\$21.79	0.171%	\$0.0373	0.077%	\$0.0167
NSM	13.73	\$26.60	0.194%	\$0.0515	0.093%	\$0.0247
NXTL	1.09	\$5.07	0.240%	\$0.0122	0.187%	\$0.0095
ONE	13.28	\$35.58	0.123%	\$0.0439	0.050%	\$0.0179
ORCL	0.77	\$16.00	0.055%	\$0.0088	0.053%	\$0.0085
PEP	9.88	\$49.71	0.076%	\$0.0380	0.037%	\$0.0184
PFE	6.38	\$41.06	0.069%	\$0.0285	0.037%	\$0.0152
PG	9.49	\$83.99	0.070%	\$0.0586	0.031%	\$0.0261
ROK	46.56	\$18.71	0.274%	\$0.0512	0.098%	\$0.0184
RSH	19.39	\$27.70	0.192%	\$0.0532	0.077%	\$0.0214
RTN	19.43	\$37.90	0.121%	\$0.0460	0.050%	\$0.0188
S	14.98	\$52.76	0.096%	\$0.0505	0.045%	\$0.0235
SBC	7.87	\$36.56	0.090%	\$0.0327	0.047%	\$0.0171
SLB	8.82	\$55.86	0.099%	\$0.0551	0.049%	\$0.0273
SLE	21.26	\$21.36	0.126%	\$0.0270	0.060%	\$0.0127
SO	22.79	\$25.00	0.116%	\$0.0290	0.056%	\$0.0141
T	11.04	\$15.59	0.142%	\$0.0221	0.071%	\$0.0111
TOY	26.90	\$17.63	0.212%	\$0.0373	0.078%	\$0.0138
TXN	7.12	\$30.52	0.133%	\$0.0406	0.065%	\$0.0198
TYC	5.30	\$29.34	0.213%	\$0.0626	0.152%	\$0.0445
UIS	32.61	\$11.65	0.214%	\$0.0250	0.097%	\$0.0113
USB	17.03	\$19.93	0.142%	\$0.0283	0.066%	\$0.0131
UTX	11.98	\$69.41	0.084%	\$0.0584	0.039%	\$0.0269
VIAB	11.56	\$42.42	0.137%	\$0.0580	0.054%	\$0.0230
VZ	8.54	\$45.76	0.083%	\$0.0381	0.042%	\$0.0193
WFC	9.17	\$46.10	0.075%	\$0.0346	0.037%	\$0.0171
WMB	11.44	\$15.97	0.263%	\$0.0420	0.125%	\$0.0200
WMT	8.37	\$60.02	0.073%	\$0.0436	0.038%	\$0.0227
WY	15.86	\$59.75	0.107%	\$0.0640	0.045%	\$0.0267
XOM	7.56	\$39.41	0.071%	\$0.0280	0.040%	\$0.0157
XRX	21.27	\$10.12	0.222%	\$0.0224	0.107%	\$0.0108
Average	14.30	\$40.69	0.128%	\$0.0448	0.061%	\$0.0207

	Dependent Variable		
	<i>Log FITC</i>	<i>Log Half-Spread</i>	<i>Log Eff. Spread</i>
<i>Intercept</i>	-2.92 (-10.38)*	-3.62 (-10.80)*	-3.93 (-13.14)*
<i>Log Turnover</i> <i>(lturn)</i>	0.161 (8.85)*	0.079 (3.67)*	0.058 (3.00)*
<i>Log Size</i> <i>(lsize)</i>	-0.153 (-5.29)*	-0.145 (-4.21)*	-0.045 (-1.47)
<i>Log SD Price</i> <i>(lsdprice)</i>	0.554 (12.44)*	0.419 (7.88)*	0.618 (13.04)*
<i>Log Price</i> <i>(lprice)</i>	-0.15 (-4.79)*	-0.380 (-10.14)*	-0.273 (-8.16)*
<i>NYSE Dummy</i> <i>(nyse)</i>	0.86 (11.68)*	0.727 (8.20)*	0.439 (5.56)*
	<i>adjR</i> ² =94.9%	<i>adjR</i> ² =92.9%	<i>adjR</i> ² =93.0%

Table II. Outcome of regressions of the logarithm of the *FITC*'s, logarithm of the Half-Spread, and the logarithm of the Effective Spread 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.

	Dependent Variable		
	<i>Log FITC</i>	<i>Log Half-Spread</i>	<i>Log Eff. Spread</i>
<i>Intercept</i>	-0.379 -1.14	-1.95 (-5.89)*	-2.79 (-9.48)*
<i>Log PIN</i> <i>(lpin)</i>	0.184 (2.31)*	0.166 (2.07)*	0.118 (1.66)*
<i>Log Size</i> <i>(lsize)</i>	-0.247 (-8.07)*	-0.192 (-6.25)*	-0.099 (-3.64)
<i>Log SD Price</i> <i>(lsdprice)</i>	0.709 (13.89)*	0.524 (10.24)*	0.646 (14.24)*
<i>Log Price</i> <i>(lprice)</i>	-0.107 (-2.48)*	-0.313 (-7.25)*	-0.222 (-5.79)*
	<i>adjR</i> ² =91.0%	<i>adjR</i> ² =91.3%	<i>adjR</i> ² =91.6%

Table III. Outcome of regressions of the logarithm of the *FITC*'s, logarithm of the Half-Spread, and the logarithm of the Effective Spread 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.

	Dependent Variable		
	<i>Log FITC</i>	<i>Log Half-Spread</i>	<i>Log Eff. Spread</i>
<i>Intercept</i>	-1.36 (-4.63)*	-2.87 (-10.01)*	-3.36 (-13.55)*
<i>Log # Analysts</i> <i>(lanalysts)</i>	-0.112 (-2.66)*	-0.0265 (-.644)	-0.0449 (-1.25)
<i>Log Size</i> <i>(lsize)</i>	-0.290 (-9.34)*	-0.222 (-7.33)*	-0.0935 (-3.56)
<i>Log SD Price</i> <i>(lsdprice)</i>	0.764 (15.82)*	0.527 (11.18)*	0.693 (16.98)*
<i>Log Price</i> <i>(lprice)</i>	-0.046 (-1.22)	-0.324 (-8.84)*	-0.236 (-7.44)*
<i>NYSE Dummy</i> <i>(nyse)</i>	1.29 (18.81)*	0.960 (14.32)*	0.590 (10.14)*
	<i>adjR</i> ² =94.9%	<i>adjR</i> ² =92.9%	<i>adjR</i> ² =92.4%

Table IV. Outcome Outcome of regressions of the logarithm of the *FITC*'s, logarithm of the Half-Spread, and the logarithm of the Effective Spread 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
<i>Intercept</i>	-12.70 (-73.57)	-11.02 (-8.93)*	-10.82 (-13.06)*
<i>Log Turnover</i> <i>(lturn)</i>	1.01 (10.72)*		
<i>Log PIN</i> <i>(lpin)</i>		1.336 (2.60)*	
<i>Log # Analysts</i> <i>(lanalysts)</i>			-1.29 (-4.31)*
	<i>R</i> ² =55.8%	<i>R</i> ² =9%	<i>R</i> ² =17.0%

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.

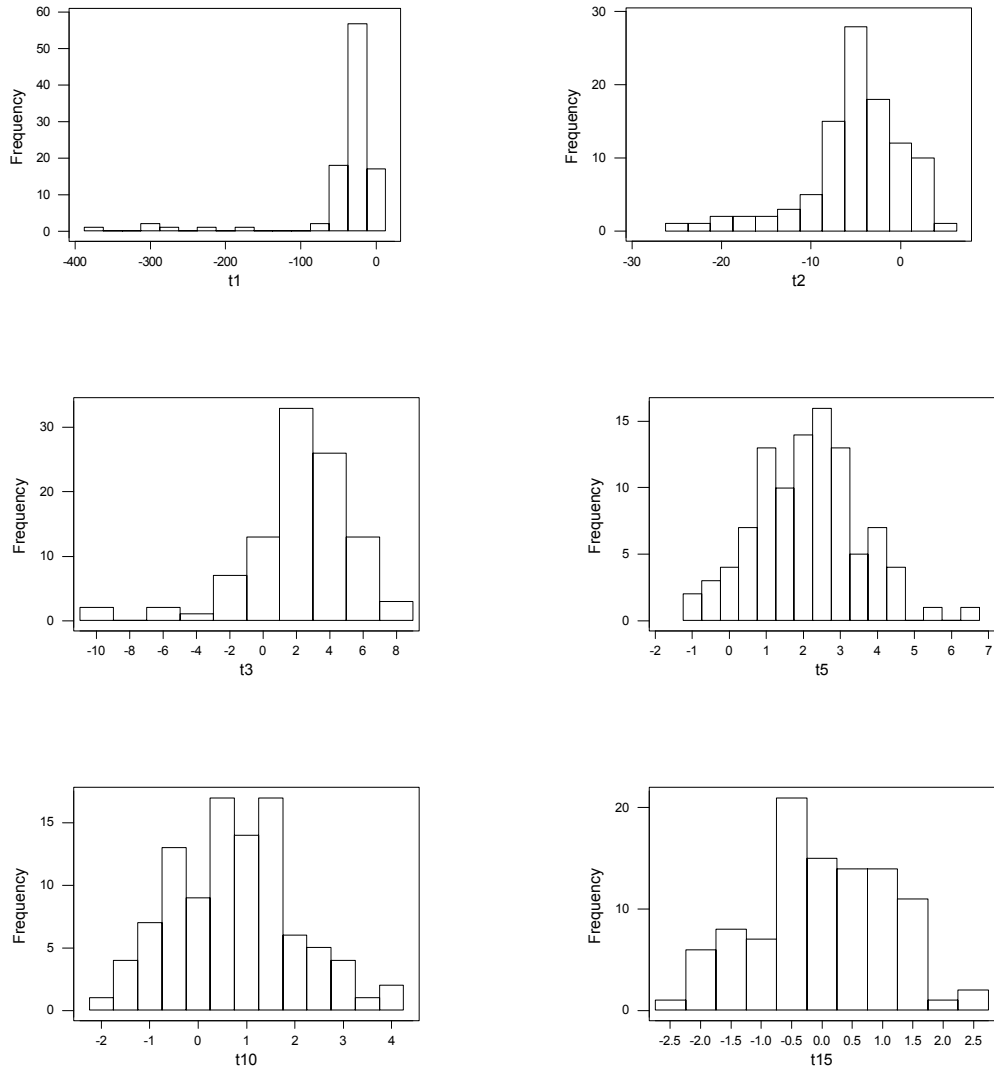


Figure 1. Histograms of the t-ratios of the estimated serial correlations of order 1,2,3,5,10 and 15 of the transaction prices of the S&P 100 stocks.

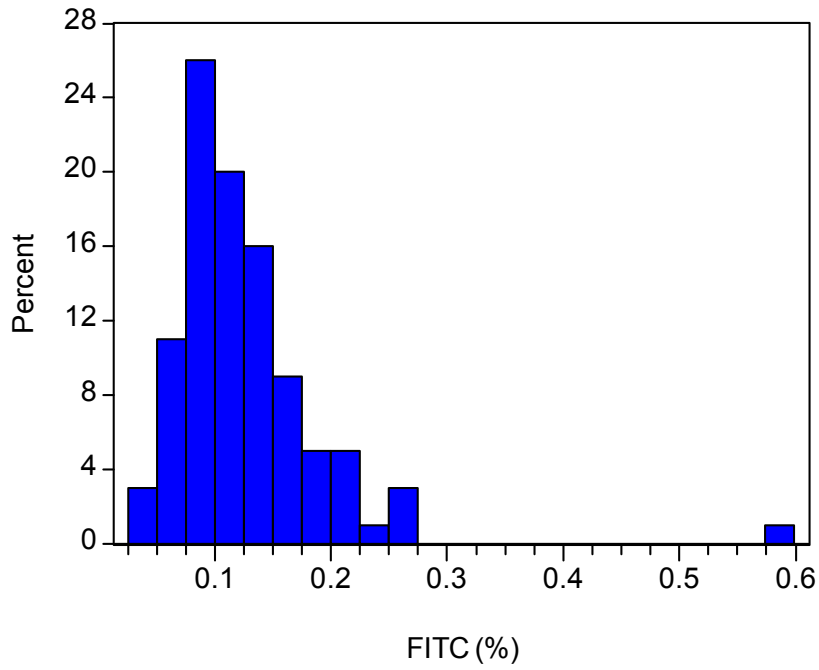


Figure 2. Histogram of the estimated *FITC*'s (in percentage values) of the S&P 100 stocks.

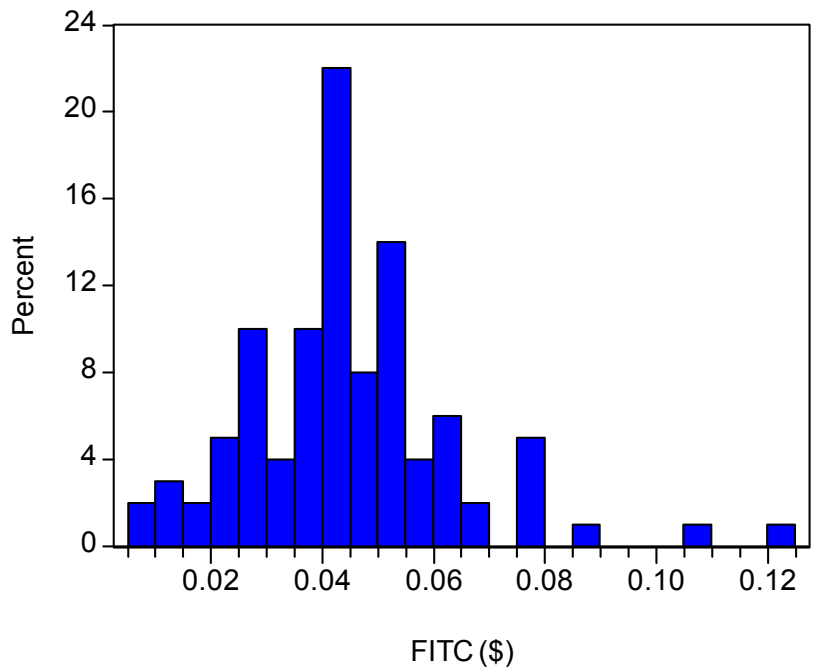


Figure 3. Histogram of the estimated *FITC*'s (in dollar values) of the S&P 100 stocks.

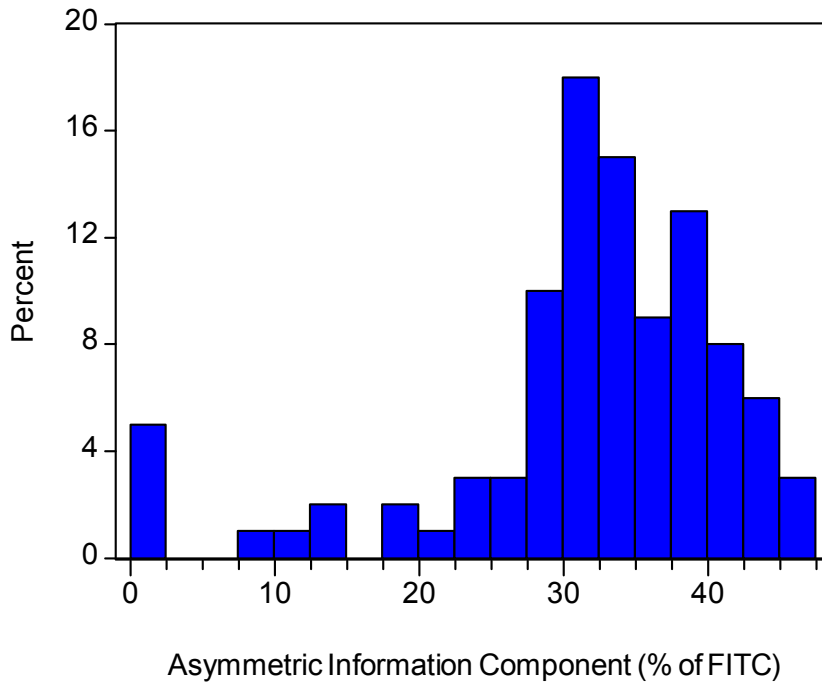


Figure 4, Estimated lower bound of the asymmetric information component expressed as a percentage of the FITC.

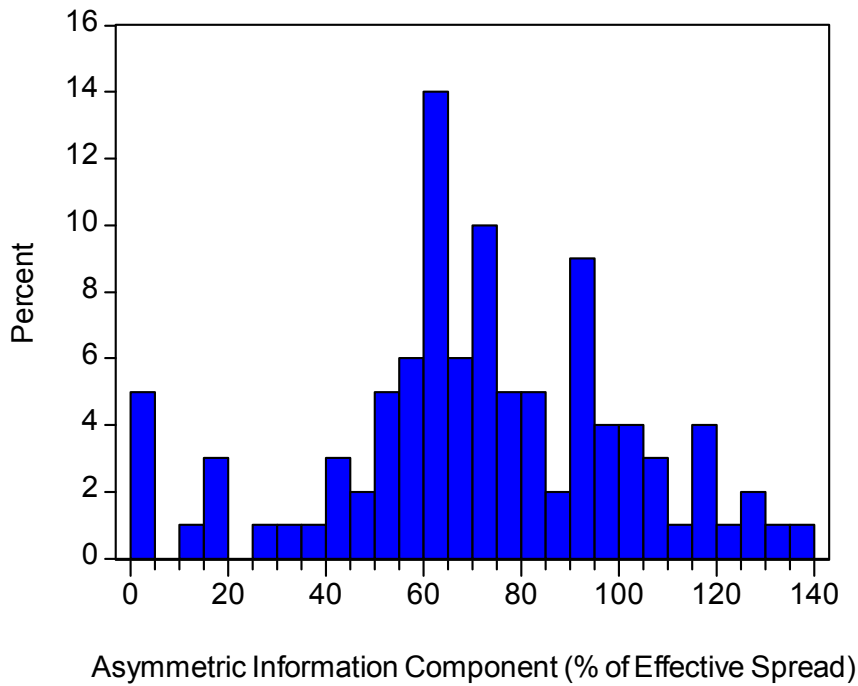


Figure 5, Estimated lower bound of the asymmetric information component expressed as a percentage of the effective spread.