Full-information transaction costs*

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

April 27, 2004

Abstract

In a world with private information and learning on the part of the market participants, transaction costs should be defined as the (positive) differences between transaction prices and full-information prices, i.e., the prices that reflect all information*, private and public,* about the asset of interest.

While current approaches to the estimation of execution costs largely focus on measuring the differences between transaction prices and efficient prices, i.e., the prices that embed all *publicly* available information about the asset, this work provides a simple and robust methodology to identify full-information transaction costs based on high-frequency transaction price data. Our estimator is defined in terms of sample moments and is model-free in nature. Specifically, our measure of transaction costs is robust to unrestricted temporary and permanent market microstructure frictions as induced by operating costs (order-processing and inventory-keeping, among others) and adverse selection.

Using a sample of S&P 100 stocks we provide support for both the operating cost and the asymmetric information theory of transaction cost determination but show that conventional measures of execution costs have the potential to understate considerably the true cost of trade.

JEL classification: G12, G19.

Keywords: Asymmetric information, liquidity, transaction costs.

*Paper written for the conference "Analysis of high-frequency data and market microstructure," Taipei, Taiwan, December 15-16, 2003. We thank David Easley for discussions.

1 Introduction

Measuring the execution costs of stock market transactions and understanding their determinants 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 (in electronic format) execution quality statistics regarding their trades.[1] Not surprisingly, then, a considerable amount of recent work in finance has been devoted to the estimation of execution costs.[2]

Ideally, the execution cost of a trade should be defined as the distance between the corresponding transaction price and the full-information price, i.e., the price that reflects all *information, private and public,* about the asset of interest. Nonetheless, virtually all existing measures of the cost of trade hinge on a different characterization of the execution cost. In effect, the execution cost is typically defined as the (signed[3]) difference between a transaction price and a benchmark price, generally called "the efficient price" (we use the same terminology hereafter), which is intended to reflect all *publicly available information* about the asset.

In a world with heterogeneous (i.e., informed and uninformed) agents, the difference between full-information price and efficient price, might be substantial. The theoretical distinction between the two notions can be easily gauged in the context of a canonical information-based microstructure model (see, for instance, Easly and O'Hara (1987) and Glosten and Milgrom (1985)).[4] In a competitive market with informed and uniformed (liquidity, that is) traders, the market maker's pricing decision is influenced by the fact that the informed agents' trading decisions carry information about the "correct" or full-information value of the asset. A decision to sell to the specialist, for example, might signal that somebody (who is informed) is aware of bad news about the asset. Equivalently, should an uninformed agent be involved in the trade, the trade might mean that there is a need for liquidity on the part on the uninformed agents. Since the specialist cannot clearly infer which is the case, he protects himself by changing his beliefs about the value of the

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

[2]The interested reader is referred to a recent special issue of the Journal of Financial Market (JFM, vol. 6, issue 3, pages 227 - 459) for a thorough discussion of developments on the subject.

[3]Positive for buy orders, negative for sell orders.

[4]For a complete review of this literature, we refer the reader to O'Hara (1995).

asset and setting bid and ask prices that are equal to the conditional expectation of the asset's value given the type of trade that occurs. Hence, the trades provide information. This information is used by the market maker to choose the quoted prices consistently with the expectation of the asset's value given *publicly available information*, while learning about the *full-information* value of the asset. The learning of the specialist leads to quoted prices that eventually converge to full-information values. But, of course, convergence is a limiting notion in information-based sequential trade models.

The present paper aims at measuring transaction costs as defined as the (positive) difference between the transaction price of a trade and the underlying (unobservable) full-information price. We depart from conventional approaches to execution cost estimation in two ways. First, for reasons that were made apparent earlier, our reference price is not the efficient price but the full information price. Second, our interest is in the *absolute* distance between transaction price and full-information price rather than in the *signed* distance between transaction price and efficient price. In other words, we do not simply wish to characterize the gain or loss of market participants placing market sell or buy orders. Instead, we aim at identifying absolute deviations from the correct (or full-information) price. Such deviations represent a cost for certain agents and a gain for others. As an example, market orders are often satisfied through outstanding limit orders. Thus, while current approaches focus on *expected* costs for market orders on either side of the market, we provide measurements of the *standard deviation* of the full-information transaction cost and, under a simple assumption in the spirit of Roll (1984), of the absolute size of the cost. The latter is of importance in understanding the extent of microstructure frictions in transaction prices. The former represents crucial information for risk-averse market participants, such as investors looking to rebalance their portfolios, in that a small expected cost of trade can be associated with arbitrarily large departures of the transaction price from the reference price. Hence, our methodology is founded upon a notion of "efficient market" as one in which trades occur close to the full-information price.

Even when defining the execution cost as the (signed) difference between transaction price and efficient price, as generally done in the literature, the measurement of the cost of trade has posed substantial challenges to the empirical microstructure research. Sensible measures of the effective transaction cost based on high-frequency data, for instance, have the potential to contain substantial measurement error. One such measure is the half difference between the prevailing ask and bid price, i.e., the half-spread. It is widely accepted that the half-spread cannot be regarded as a precise measurement of the cost of trade since transactions often occur within the posted quotes. Conversely, the (signed) difference between a transaction price and the corresponding mid-point bid-ask quote would represent a sensible assessment of the effective transaction cost only if the trades were easily classifiable as being buy or sell (so that transactions can be signed appropriately) and the reference efficient price were easily identifiable. This is typically not the case (Bessembinder (2003), among others).

While straightforward estimates based on the quoted bid-ask spreads and the corresponding transaction prices are bound to provide misleading information due to measurement errors, more complicated measurements relying on structural models of price formation and data sampled at lower (typically daily) frequencies can be subject to misspecifications. In an important contribution, Roll showed that the implicit cost of trade in an efficient market can be written as two times the square root of the (negative) first-order serial covariance between the log price changes taken with a negative sign (Roll (1984)). The Roll's estimator relies on a variety of potentially restrictive assumptions, namely (*i*) the uncorrelatedness of the changes in the underlying efficient log prices (which translates into absence of correlation in the efficient stock returns), (*ii*) the independence between the efficient prices and the execution costs, and (*iii*) the serial uncorrelatedness of the execution costs. Assumption (*i*) is known not to be required in an efficient market (i.e., conditional expected returns can be time-varying) and does not appear to be supported by empirical evidence (see Conrad and Kaul (1988, 1989) and Conrad et al. (1990) for some early evidence). Coherently, George et al. (1991) modify the Roll's estimator to allow for predictability in the underlying efficient returns. Assumption (*ii*) implies that the effective cost of trade is somewhat exogenously set without regard for the efficient price. Glosten and Milgrom (1987), among others, argue that this feature of the specialist's price setting behavior is likely to be incorrect in that the spread is a fundamental device that the market makers use in their interaction with the remaining market participants. As for Assumption (*iii),* its imposition rules out realistic autocorrelations induced by clustering in order flows, limit orders, asymmetries in information, and other market microstructure effects (Garbade and Lieber (1977), for example). In light of this observation, Choi et al. (1988) derive an extension of the Roll's model that allows for positive first-order serial correlation between the underlying transaction types.

Despite its reliance on a stylized market structure and the implicit assumption that (negative) correlations in stock returns are simply induced by transitory order-processing costs, Roll's seminal approach has spurred a considerable amount of work on the measurement of the effective cost of trade. As said, recent work on the subject has partly relaxed Assumption *(i)* through (*iii*) above. Nonetheless, while to the best of our knowledge no existing contribution fully accounts for unrestricted deviations from the Roll's set-up (as summarized by Assumption (*i*) through (*iii*)), the recently proposed estimators generally lose the empirical appeal and simplicity of Roll's original idea.

Building on recent work by Bandi and Russell (2003a,b), we suggest a methodology to consistently estimate the full-information cost of trade, defined here as the (positive) difference between transaction price and unobserved full-information price, using high-frequency transaction price data. The intuition behind our method is as follows. With the exception of relatively infrequent discrete jumps induced by news arrivals to the informed agents, the dynamic behavior of the underlying full-information price process is expected to be rather smooth over time. On the other hand, at the (high) frequencies at which trades arrive intra-daily, the cost of trade necessarily appears as a discrete jump in the underlying full-information log price process. In consequence, the high-frequency continuously-compounded return data are dominated by return components that are induced by microstructure effects since the underlying full-information returns evolve smoothly in time. In this context, we can employ straightforward sample moments of the observed high-frequency return data to learn about moments of the effective cost of trade (and, under relatively innocuous assumptions, the effective cost of trade itself). We do so by using the informational content of return data whose full-information return component is largely swamped by the cost of trade when sampling is conducted at the high frequencies at which transactions occur in practice. To put it simply, we achieve identification by exploiting the econometric implications of an uncontroversial economic assumption: meaningful revisions in the underlying full-information price process occur at a lower frequency than revisions in the efficient price and discrete microstructure frictions (i.e., price discreteness, discrete adjustments to the supply of liquidity via updates in the limit order book and so on). While the latter change with the trading process in that the specialist learns from the trading and adjusts the quotes based on publicly available information, the former (namely, the full-information price) is relatively more stable over time in that it reflects learning about the asset's fundamentals on the part of the informed agents.

Our method is simple and robust. Its simplicity relies on the computation of empirical moments of recorded high-frequency stock returns. In agreement with the approach in Roll (1984), our measure only requires the availability of transaction prices. Its robustness hinges on the observation that, contrary to Roll's estimator, Assumption (*i)* through (*iii*) are not required for the method to provide a consistent measurement of the implicit cost of trade. Specifically, we allow for correlation in the full-information returns. We also allow for correlation between the cost of trade and the underlying full-information price, even though this assumption is somewhat less compelling than when using the efficient price to define transaction costs as in standard approaches. Finally, we account for temporal correlation (possibly of any finite order) between the microstructure frictions. More precisely, we permit the cost of trade to be determined by unrestricted order-processing costs (as in Tinic (1972), among others), inventory-holding costs (Amihud and Mendelson (1980) and Ho and Stoll (1981), *inter alia*), and asymmetric information-related costs (Copeland and Galai (1983) and Glosten and Milgrom (1985), among others).[5]

Based on estimates of the full-information transaction costs for the cross-section of S&P100 stocks, we re-assess the importance of the conventional determinants of the cost of trade as summarized by the traditional taxonomy in the literature: operating costs (order-processing and inventory-carrying, among others) and asymmetric information (see Bagehot (1971), for example). Although we find considerable empirical support for both theories of transaction cost determination, we stress that standard measures of execution costs might understate the true transaction costs faced by the market participants. As expected, the difference between full-information transaction costs and more conventional

[5] Glosten (1987) and Stoll (1989, 2000), *inter alia*, discuss the relative importance of the main theories of effective spread determination.

measures can be imputed to private information.

The paper proceeds as follows. Section 2 briefly reviews the existing literature on the estimation of the cost of trade and provides motivation for the new measure introduced in this work. Section 3 discusses a general model of transaction price determination. In Section 4 we present our nonparametric estimator of the full-information transaction cost. Section 5 provides estimates of execution costs for the cross-section of S&P100 stocks based on high-frequency transaction price data. In Section 6 we study the determinants of the cross-sectional variation in the transaction costs of the S&P 100 stocks. Section 7 concludes. Technical details and proofs are in the Appendix.

2 Which measure of execution costs?

Following Perold (1988), it is generally believed 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). In a world with asymmetric information, a more informative measure of the effective execution cost should aim at capturing the (positive) difference between the trade price of an asset and the corresponding full-information price, namely the price that contains all private and public information about the asset. Such measure is the object of interest in the present paper.

Most available estimates of the cost of trade relying on high-frequency data hinge on the basic logic behind Perold's original intuition. Specifically, there are three measures of the execution cost that have drawn attention in recent years, i.e., the so-called quoted "bid-ask half spread," the "effective half spread," and the "realized half spread." The quoted bid-ask half spread is defined as half the difference between the ask quote and the bid quote. The effective half spread is the (signed) difference between the price at which a trade is executed and the midpoint of the reference bid-ask quotes. As for the realized half spread, this measure is defined as the (signed) difference between the transaction price and the midpoint of a quote in effect some time after the trade.

Some of the limitations of the above measures of the cost of trade have been pointed out in the literature (the interested reader is referred to the special issue of the Journal of Financial Markets on execution costs for updated discussions). The quoted bid-ask half spread, for example, is expected to generally overestimate the true cost of trade in that trades are often executed at prices within the posted quotes. As for the effective and realized spreads, not only they require the trades to be signed as being buyer or seller-initiated but they also require the relevant quotes and transaction prices to be matched.

The first issue (i.e., assigning the trade direction) arises due to the fact that commonly used high-frequency data sets (the TAQ database, for instance) do not contain information about whether a trade is buyer or seller-initiated. Some data sets do provide this information (the TORQ database being an example) but the length of their time series is often insufficient. Naturally, then, a considerable amount of work has been devoted to the construction of algorithms intended to classify trades as being buyer of seller-initiated simply on the basis of transaction prices and quotes (see, for example, Lee and Ready (1991) and Ellis et al. (2000)). Nonetheless, the existing algorithms are somewhat ad-hoc and can missclassify a large number of trades (the Lee and Ready method, for example, is known to categorize incorrectly about 15% of the trades), thereby inducing biases in the final estimates. Bessembinder (2003) and Peterson and Sirri (2003) contain a thorough discussion of the relevant issues.

The second issue (i.e., matching quotes and transaction prices) requires potentially arbitrary judgment calls. Since the trade reports are often delayed, when computing the effective spreads, for example, it seems sensible to compare the trade prices to quotes that occur before the trade report time. The usual allowance is 5 seconds (see Lee and Ready (1991)) but longer lags can of course be entertained. As pointed out by Bessembinder (2003), it would appear appropriate to compare the trade prices to earlier quotes even if there were no delays in the reporting. Such comparison would somehow incorporate the temporal difference between trading decision and implementation of the trade as in Perold's recommendation.

There is another feature of these measures which the extant literature does not emphasize. The implicit assumption underlying all three measures of the cost of trade is that the mid-point bid-ask quote represents a valid estimate of the efficient price. Even though this assumption is coherent with the implications of canonical microstructure models with asymmetric information and is certainly accurate on average, it can not be expected to always hold. For instance, an NYSE specialist might modify the spread (and the corresponding mid-point) to insure transaction prices are recorded inside the quoted spread even in situations where no new public information about the asset is available. In this context, since the above-mentioned estimators are nonlinear in nature, mid-point bid-ask quotes that are only equal to the efficient prices on average introduce biases in the measurement of the effective cost.

The third measure, i.e., the realized half spread, hinges on the additional assumption that the efficient price to be used as a reference is not the prevailing price at the time of the trade but the post-trade value of the stock (proxied by a mid-point quote in effect after the trade). The idea is that the traders possess private information about the security value and the trading costs should be assessed based on the trades' non-informational price impacts (Seppi, 1997). Since the efficient price is bound to converge to the full-information price (the price that contains all private information), the use of a mid-point quote in effect after the trade is intended to capture the difference between transaction price and full-information price *at the time of the trade.* While the logic behind the realized half spread as a measure of transaction costs is coherent with the approach in this paper, this measure is subject to all the limitations that affect the effective spread estimates. Additionally, the choice of the reference (post-trade) mid-point bid-ask value is highly problematic, and always arbitrary, in this case.

Hence, all the measures that were described earlier are at the best noisy and have the potential to be considerably biased. To complicate things, the values that they provide are substantially different. As an example, in the case of NYSE stocks (over the period between July 1 and December 31, 1998 - TAQ database), Bessembinder (2003) finds that the estimated (percentage) cost of trade is about 50 basis point, 25 basis point, and 2 basis point when computed using the midpoint half spread, the effective half spread, and the realized half spread, respectively.

This said, there is a well-known measure (which can be computed using low frequency data) that does not require either the signing of the trades or the matching of quotes and transaction prices, i.e., the Roll effective spread estimator (Roll (1984)). Interestingly, the Roll estimator does not rely on the implausible assumption that the mid-point bid-ask quotes are realistic representations of the unobserved efficient prices. The idea behind the Roll's measure is that the (constant) width of the effective spread can be computed based on the negative first-order autocovariance of recorded stock returns. The estimator is defined as two times the square root of the (negative) first-order autocovariance of the recorded log price changes taken with a minus sign (see Hasbrouck (1999, 2003) for interesting extensions).

Nonetheless, the Roll estimator still hinges on several restrictive assumptions. Let us assume that the recorded log price process is determined based on the unobservable efficient price plus a component capturing microstructure frictions. In the Roll's framework, the efficient returns (i.e, the returns obtained from the efficient prices) are thought to be serially uncorrelated. In addition, the microstructure frictions in returns are (negatively) serially correlated only up to the first order, thereby inducing a similar (negative) first-order autocorrelation in the recorded returns. Finally, the microstructure noise components are modelled as being unocorrelated with the efficient prices.

As discussed in the Introduction and shown in our empirical work below, these assumptions are generally unrealistic in the presence of transaction data. A typical manifestation of the restrictiveness of the Roll's assumptions is the fact that Roll estimator often delivers negative estimates of the effective cost of trade. While recent improvements to the Roll's approach have relaxed some of the assumptions made in Roll's original contribution, no existing extension of the Roll's methodology is, to the best of our knowledge, completely general.

Although in this paper we aim at measuring full-information transaction costs, our method borrows from Roll's methodology without necessitating its original assumptions. Contrary to the approach in Roll (1984), we do not assume that the spread width is constant over time. In agreement with the conventional Roll's estimator, neither do we need to sign the trades nor do the transaction prices need to be matched to the mid-point bid-ask quotes. Finally, the method only uses high-frequency transaction prices, without requiring quote data, and is easy to implement. As shown below, it just entails arithmetic averages of recorded stock returns.

Our approach employes the information potential of transaction price data sampled at high-frequencies. Specifically, as we further explain in the next section, we exploit the different orders of magnitude of the components of recorded high-frequency stock returns, i.e., the unobserved full-information returns and the equally unobserved microstructure frictions in returns, to consistently estimate an important feature of the latter, i.e., the (standard deviation of the) full-information transaction cost. The idea is that at the high frequencies at which trades arrive in practise, the component of the recorded stock returns that is imputable to microstructure frictions (stochastically) dominates the underlying full-information return component. Such dominance hinges on a well-accepted economic idea: the full-information price is revised less frequently than the efficient price and discrete microstructure frictions, thereby generating a full-information return series that is relatively smoother than the recorded stock returns.

We now turn to a description of the model.

3 The price formation mechanism

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 $[0, t]$.

We write the observed price process corresponding to transaction i as

$$\widetilde{p}_i = p_{t_i} \overline{\eta}_i \tag{1}$$

where p_{t_i} is the true price (i.e., the full-information expected value of future cash flows) and $\overline{\eta}_i$ denotes a generically-specified microstructure effect. We now take a log transformation of Eq. (1) and difference it to obtain:

$$\ln(\widetilde{p}_i) - \ln(\widetilde{p}_{i-1}) = \ln(p_{t_i}) - \ln(p_{t_{i-1}}) + \eta_i - \eta_{i-1}, \tag{2}$$

where $\eta = \log(\overline{\eta})$. Naturally, Eq. (2) can be rewritten as

$$\widetilde{r}_i = r_{t_i} + \varepsilon_i, \tag{3}$$

where $\widetilde{r}_i = \ln(\widetilde{p}_i) - \ln(\widetilde{p}_{i-1})$ is the observed continuously-compounded return over the transaction interval (t_{i-1}, t_i), $r_{t_i} = \ln(p_{t_i}) - \ln\left(p_{t_{i-1}}\right)$ is the corresponding full-information continuously-compounded return, and $\varepsilon_i = \log(\overline{\eta}_i) - \log\left(\overline{\eta}_{i-1}\right) = \eta_i - \eta_{i-1}$ denotes the microstructure friction in returns. We now discuss the assumptions that we impose on the (unobserved) full-information log price process.

Assumption 1.

(1) *The full-information log price process* $\ln(p_t)$ *is a continuous semimartingale. Specifically,*

$$\ln(p_t) = A_t + M_t, \tag{4}$$

where A_t *is a continuous finite variation component and* $M_t = \int_0^t \sigma_s dW_s$ *is a local martingale.*

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

In agreement with standard approaches in asset-pricing, we model the underlying log price process as a continuous semimartingale (c.f., Assumption 1(1)). It is known that the semimartingale property of price processes is a necessary condition for the absence of arbitrage opportunities (see Duffie (1990), for example). We also allow for the presence of stochastic volatility. Specifically, provided Assumption 1(2) is satisfied, 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.

As in George et al. (1991), the model permits the expected return of the unobservable true price process to vary through time. This property is supported by a vast empirical evidence (Conrad et al. (1990) for some early findings).

We now describe the assumptions that we impose on the microstructure frictions.

Assumption 2.

(1) *The microstructure effects in prices* η_i *are mean zero and covariance stationary.*

(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) *Furthermore,* $\mathbf{E}(\eta_s \eta_q \eta_{s-u} \eta_{q-p}) < \infty \ \forall s, q$ *and* $\forall u, p = 0, ..., k.$

The microstructure effects are stationary (c.f., Assumption 2(1)). The dependence structure of the spreads is such that all covariances of order smaller than k are 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 (c.f., Assumption 2(2)).

This property is important and, in analogy with Assumption 1, contrasts with Roll's (1984) original specification. In Roll's model the microstructure effects in prices are independently distributed. When combined with the zero correlation of the efficient returns, the independence property of the price frictions η implies a negative first-order autocovariance in the observed return series given by $-\mathbf{E}\left(\eta^2\right)$ as well as higher-order autorrelations that are equal to zero. Even though this characteristic of the Roll's model represents a valid approximation in certain markets, it does rule out temporal dependence in order flows, limit orders, asymmetric information, and other microstructure effects which we wish to accommodate here.

Finally, we allow the microstructure noise to be correlated with the unobservable full-information price process. A standard criticism of the Roll's set-up is that the cost of trade is somewhat exogenously set without regard for the underlying efficient price. In effect, one could argue that the spreads are important tools that the market makers control in their interactions with the remaining market participants (see Glosten and Milgrom (1985)). While this criticism is somewhat less compelling in our case being that the full-information price (rather than the efficient price) is our benchmark here, it is worth pointing out that this standard assumption of the Roll's model can also be relaxed in our framework.

We now turn to the substantive core of the present work, i.e., the estimation of the full-information transaction costs.

4 Measuring transaction costs

The following lemma expresses the square root of the second moment of the microstructure frictions in prices as a function of the cross-moments of the frictions in returns. Our estimator will be a consistent sample analogue of the theoretical standard deviation (σ_η) of the generic spread η as described in the lemma below.

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

$$\sigma_\eta = \sqrt{\mathbf{E}(\eta^2)} = \sqrt{\left(\frac{1+k}{2}\right)\mathbf{E}(\varepsilon^2) + \sum_{s=0}^{k-1}(s+1)\mathbf{E}(\varepsilon\varepsilon_{-k+s})}. \tag{5}$$

Proof. *See Appendix.*

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

$$\sigma_\eta = \sqrt{\mathbf{E}(\varepsilon^2) + \mathbf{E}(\varepsilon\varepsilon_{-1})}. \tag{6}$$

If $k = 2$, i.e., $\mathbf{E}(\eta\eta_{-2}) = \theta_2$,[6] then

$$\sigma_\eta = \sqrt{\frac{3}{2}\mathbf{E}(\varepsilon^2) + 2\mathbf{E}(\varepsilon\varepsilon_{-1}) + \mathbf{E}(\varepsilon\varepsilon_{-2})}. \tag{7}$$

Provided the cross-moments of the microstructure frictions in returns can be consistently estimated using observables, Eq. (5) constitutes an expression that could be readily used to identify the second moment of the full-information transaction cost. Interestingly, the availability of high frequency data offers us a way to do so. The idea goes as follows.

The full-information price is expected to evolve rather smoothly over time. Its evolution is driven by the pace at which informed agents learn and react to new private information. While the arrival of sudden news can induce discrete but infrequent jumps,[7]

[6] Assuming non-zero serial covariances is important. Sometimes transaction types repeat each other, i.e., sales and purchases cluster over brief periods of time. Garbade and Lieber (1977), for example, discuss a situation where a floor broker might split a large (in terms of number of shares) order into smaller orders, thereby making successive sales and purchases (which end up being recorded as different transactions) the result of the same trade. Similarly, limit orders might remain in the market maker's book until there is a change in quotations. When a favorable change occurs, many limit orders might be satisfied at the same time. These transactions are typically recorded separately. Hence, as earlier, they induce several trades on the same side of the market and, consequently, serial correlation in the transaction prices.

[7] Our framework allows us to accommodate these types of shocks. Intuitively, infrequent news arrivals that have an impact on the level of the recorded log price process are, at high frequencies, dominated by the more frequent microstructure effects. In fact, the latter occur any time a transaction takes place. Our consistency arguments (which are discussed below) would then still be valid if we added a bounded jump component of finite intensity to the (compensated) semimartingale process driving the full-information price dynamics. In this case, we would simply have to re-define appropriately the notion of quadratic variation $\left([\log p]_0^{\overline{h}} = \int_0^{\overline{h}} \sigma^2 ds\right)$ in the proof of Theorem 1 and add to it a component $\sum_{0<t\leq\overline{h}} (\Delta \log p_t)^2$,

the underlying dynamics of the true price process is best described by a continuous-time series that slowly adjusts to new private information. At high frequencies, then, the full-information component of the observable return data is stochastically negligible in that its order of magnitude is $O_p\left(\sqrt{\max|t_i - t_{i-1}|}\right)$ (see Assumption 1(1)), where $\max|t_i - t_{i-1}|$ is the maximum duration between quote updates. Of course, the maximum duration is small when using data sampled at the frequencies at which new transactions arrive (see Table 1 below).

Contrary to the underlying full-information returns, the microstructure frictions in the observable return data cannot be thought of as a process whose evolution is continuous in the state space. The observed prices change discretely on the basis of ticks. Furthermore, the very nature of the market-makers' price setting behavior is consistent with the view that quote and price adjustments occur discretely in correspondence with the trading. As pointed out earlier, the dealers react to a variety of factors, such as their inventory levels, the arrival of new orders, and the likelihood of asymmetries in information, among others. The final impact of their reactions is best modelled by discrete adjustments in the recorded transaction prices. Hence, provided we do not sample within transaction arrivals, the variations in the cost of trade should appear as discrete changes in the recorded quotes. It is therefore natural to model variations in the execution costs as being $O_p(1)$, even when the transactions occur very frequently, as typically the case in practise. In light of these observations, when compared to the negligible full-information return component of the recorded quotes, the microstructure frictions in returns are bound to play an important (i.e., dominating) role at high frequencies.

Why are we not identifying deviations from the efficient price, rather than deviations from the full-information price? The answer to this question lies at the heart of our identification procedure and hinges on the different economic meaning of the two price notions. While we can certainly write

$$\widetilde{p}_i = p^e_{t_i}\overline{\xi}_i, \tag{8}$$

where $p^e_{t_i}$ is the unobservable efficient price and $\overline{\xi}_i$ is a microstructure effect different

where the $\Delta \log p_t^{'} s$ would have the interpretation of infrequent (bounded) jumps induced by news arrivals which affect the volatility (over the period $\left[0, \overline{h}\right]$) of the underlying full-information return process as well as the level of the price process itself.

from $\overline{\eta}_i$ in Eq. (1), it is natural to express the efficient price process as a function of the full-information price process, namely

$$p^e_{t_i} = p_{t_i}\overline{\zeta}_i. \tag{9}$$

(Bandi and Russell (2003a,b) explicitly study this case.) Although, as said, the full-information price process p_{t_i} is expected to have a smooth evolution over time (which we model by virtue of Assumption 1 above), the efficient price process changes discretely with the trading mechanism in that the trading process carries information about the full-information value of the asset (see, for example, Hasbrouck (1991) for empirical evidence on the importance of trade price impacts). Hence, let us now plug Eq. (9) into Eq. (8) to obtain

$$\widetilde{p}_i = p_{t_i}\overline{\xi}_i\overline{\zeta}_i = p_{t_i}\overline{\eta}_i,$$

where $\overline{\eta}_i = \overline{\xi}_i\overline{\zeta}_i$. It is now clear that, when taking logs, we can express the log transaction price as the sum of the smooth, full-information, log price process and a "microstructure friction" $\log\overline{\eta}_i$ that accounts for both discrete adjustments in the efficient price and conventional microstructure effects, such as price discreteness..

Finally, we note that 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 suitable to approximate the act of information extraction from a large number of transaction prices sampled at the high-frequencies at which new intra-daily information arrives in practise.

Theorem 1 below lays out the estimator and its asymptotic behavior.

Theorem 1. *Assume Assumptions 1 and 2 are satisfied. Conditional on 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*

$$\widehat{\sigma}_\eta = \sqrt{\left(\frac{k+1}{2}\right)\left(\frac{\sum_{i=1}^{N(\overline{h})}\widetilde{r}_i^2}{N(\overline{h})}\right) + \sum_{s=0}^{k-1}(s+1)\left(\frac{\sum_{i=k-s+1}^{N(\overline{h})}\widetilde{r}_i\widetilde{r}_{i-k+s}}{N(\overline{h})}\right)} \underset{N(\overline{h})\to\infty}{\overset{p}{\to}} \sigma_\eta. \tag{10}$$

Proof. *See Appendix.*

Two observations are in order. First, Eq. (10) gives us an easy method to estimate the second moment of the microstructure friction. Similarly, we could identify higher moments of the cost of trade η. Second, the estimator is defined over a single (generically specified) period $\overline{h}$. In this sense we can readily allow for a time-varying second moment of the transaction cost possibly induced by the convergence dynamics of the transaction prices to the full-information price. Under the assumption that the properties of the $\eta' s$ extend to multiple periods (or when interested in the unconditional expectation of the time-varying second moment of the execution cost), the simple summations over i (which is our index for transactions) in the definition of the estimator in Eq. (10) 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. This is what we do in the following section.

In agreement with more traditional approaches to effective spread estimation (see Roll (1984) and George et al. (1991), for instance), we can easily add some additional structure to the model, assume a constant transaction cost for each trade, and provide simpler implications for the actual cost of trade. This is what Assumption 3 and the Corollary to Theorem 1 accomplish.

Assumption 3. *Assume*

$$\eta_i = sQ_i \qquad \forall i = 1, ..., N(\overline{h}), \tag{11}$$

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.[8]

Should Assumption 3 be satisfied, then $\sigma_\eta = s$. A straightforward corollary to Theorem 1 readily follows.

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

[8] In the Roll's model Q_i can also take on two, equally probable, values, i.e., 1 and -1. $Q_i = 1$ denotes a buyer-initiated trade while $Q_i = -1$ represents a seller-initiated transaction.

as $N(\overline{h}) \to \infty$, *we obtain*

$$\widehat{\sigma}_\eta \underset{N(\overline{h})\to\infty}{\overset{p}{\to}} s, \tag{12}$$

where $\widehat{\sigma}_\eta$ *is defined in Eq. (10) and* s *is the full-information transaction cost.*

Proof. *Immediate given Lemma 1, Theorem 1, and Assumption 3.*

For simplicity, in the sequel we will use the convention of referring to estimates obtained by employing the estimator in Eq. (10) as "the full-information transaction cost" (the *FITC,* hereafter). Of course, it is understood that, when utilizing the unrestricted model in Eq. (1) above, one should really use the wording "the standard deviation of the full-information transaction cost," rather than simply "the full-information transaction cost." In a realistic set-up (such as the one described by the model in Eq. (1)), in fact, the cost of trade ought to be modelled as a process that varies across transactions. Hence, only statements about its statistical properties can be made. On the other hand, if one is willing to impose the restriction that the cost is constant over time (as implied by Assumption 3 above), then referring to the estimator in Eq. (10) as a measurement of the full-information transaction cost accurately characterizes what the estimator achieves.

The next section is about measuring the *FITC*'s for the cross-section of S&P 100 stocks.

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

The data we employ consists of high-frequency transaction prices for the stocks in the S&P 100 index. We use transaction prices, rather than mid-point bid-ask quotes, as typically the case in the realized volatility literature (see Bandi and Russell (2003b), for instance), since our interest here is in the genuine microstructure frictions induced by the trading activity.

The prices were obtained from the TAQ data set for the month of February 2002, a benign month. We restricted our attention to NYSE quote updates. Even though our asymptotic theory permits us to handle infrequent jumps in the underlying full-information price process possibly induced by news arrivals to the informed agents (see the discussion in Section 6), we recognized the potential for finite sample contaminations that might affect our measurements. Consequently, we filtered our data to avoid the use of price changes larger than two times the mean reported spread. While the procedure simply resulted in the elimination of an average of about 1.5 observations per day, the filter had no meaningful impact on our final estimates.

Table 1 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 (computed using the Lee and Ready (1991) algorithm and a standard 5 second time allowance - see Section 2 above for discussions), the annualized standard deviations of the underlying full-information log price processes (calculated using realized volatility and the optimal sampling method discussed in Bandi and Russell (2003a,b)) and the corresponding turn-over measures, i.e., the logarithms of the average number of daily shares transacted to shares outstanding. The Nasdaq stocks are denoted by an asterisk.

In Fig. 1 we present the histograms of the t-ratios of some rapresentative serial correlations for the 100 stocks in our sample, i.e., $\sqrt{n}\widehat{\rho}_{-j}$, with $j = 1, 2, 3, 5, 10, 15$. 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 important 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) testify the necessity of designing an estimation procedure that, contrary to the method in Roll (1984), is robust to deviations from a model of price determination that only allows for a negative first-order autocorrelation in the recorded stock return data. Here, to accommodate non-zero high order autocorrelations, we set k in the $FITC$ estimator in Eq. (10) equal to the maximum order of the statistically significant autocorrelations. We do so for each individual stock.

In agreement with the idea that full-information prices can be away from efficient and transaction prices due to learning on the part of the specialist as well as a variety of market frictions, such as short-sell constraints (Diamond and Verrecchia (1987)), we find that the average $FITC$ is larger than the average effective spread, i.e., 9 bp versus 6 bp.

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. In the sequel we confirm this result.

Since our method relies on the availability of transaction price data sampled at very high frequencies, we evaluate the robustness of our estimates to the use of different frequencies. More precisely, we are interested in assessing the possible biases that are being introduced by virtue of the fact that obvious data limitations prevent us from sampling at the very high frequencies that our limiting results require to achieve consistent measurements. Fig. 4 contains a scatterplot of the estimated percentage *FITC*'s computed using all transactions versus percentage *FITC*'s estimated on the basis of every other transaction. For coherence, after we choose the autocorrelation structure (as represented by the integer k in Eq. (10)) using all transactions, we define the skip sampling estimator by using a maximum number of serial correlations given by the integer part of $(k+1)/2$. If the asymptotic justification for our estimates is satisfied at our sampling frequencies, then the estimates based on all and every other trade should look very similar. This is exactly what we find. We notice that the difference between the two sets of estimates is immaterial in that all measurements fall very close to the 45 degree line. In Fig. 5 we present a scatterplot of the *FITC* estimates using all transactions versus estimates that are obtained by sampling the prices every 30^{th} transaction. As expected, the latter are contaminated by the full-information price component in the recorded prices. Specifically, they appear to be inflated since they capture features of the volatility of the underlying, full-information, price.

To conclude, our asymptotic approximation relying on vanishing durations in the limit is coherent with the features of our data, i.e., it can be validly applied to data with average durations equal to about 14 seconds (see Table 1). Thus, even though the average durations are not literally zero in our case, since doubling the sampling intervals has virtually no impact on our estimates, we expect the finite sample biases in our measurements to be negligible in practise.

We now turn to a study of the cross-sectional determinants of the *FITC*'s.

6 Explaining the cross-section of the *FITC*'s

In this section we study the determinants of the cross-sectional variation in the *FITC*'s. As pointed out earlier, the standard taxonomy in the literature postulates that two are the main economic forces behind the cost of trade: operating costs (order-processing and inventory-keeping, *inter alia*) and adverse-selection (i.e., asymmetric information) costs.

We start with operating costs. The order-processing component of the transaction costs pertains to the service of "predictive immediacy" (Demsetz (1968)) 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 spread to offset positions that are overly long or short with respect to some (desired) inventory target.

Much attention has recently received the adverse-selection component of the effective spreads. The market makers are bound to trade with investors that have superior information. Hence, the asymmetric information component of the cost of trade 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)).

In what follows we will consider proxies for the various effects that are likely to play an important role in explaining the cross-sectional variation of the *FITC*'s.

The "operating cost" channel, for example, implies that virtually any measure that is positively correlated with liquidity and ease of inventory adjustment should have a negative impact on transaction costs. Consider "trading volume per trade," for instance. When faced with high volumes, the dealer knows that imbalances in costly inventories can easily be restored. Hence, higher trading volumes translate into relatively narrower inventory-induced spreads. Consider now the "average number of trades per unit of time" (a day, say). Since the market makers are exposed to a variety of fixed operating costs, the fixed costs ought to be recovered by appropriately setting the spreads. Naturally, then, the higher is the number of trades, the smaller is the fixed cost per transaction and, consequently, the smaller is the necessary spread. Thus, Proposition 1 follows.

Proposition 1. *The "operating cost" theory implies that stocks with higher volumes per trade and a larger number of trades per day should have smaller transaction costs (everything else being the same).*

Volumes also play a role in asymmetric information-based explanations of the cost of trade. Kyle (1985) and Easly and O'Hara (1992), for example, predict increases in trading volumes following arrivals of unevenly distributed information. The specialist interprets the increase in trading activity as an indication of the increased number of informed transactions and protect itself by augmenting the size of the spreads. It is noted that, while volumes matter, the relevant predictions that are made by asymmetric information theories of spread determinations are for *unusually* high trading activity. Thus, the above considerations involve a timing of events that cannot be captured in a cross-sectional context. On the other hand, one can rely on Stoll's intuition (Stoll (1989)) and formulate a statement (contained in Proposition 2 below), whose flavor is consistent with our previous discussion.

Proposition 2. *The "asymmetric information" theory implies that stocks with higher turnover should have larger transaction costs (everything else being the same).*

The volatility of the underlying full-information price plays the same role in both theories. 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, of course, the compensation is 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)). Again, the increased risk ought to be compensated. Thus, one can derive the following proposition.

Proposition 3. *Both the "asymmetric information" theory and the "operating cost" theory imply that stocks with higher underlying volatility should have larger transaction* *costs (everything else being the same).*

We verify the implications of Proposition 1 through 3 by regressing the log of the percentage *FITC*'s on the log of the average dollar volume per trade (*lsize*) and the log of the average number of trades per day (*ltrades*), as well as on the log of the average number of shares transacted to shares outstanding (*lturn*), and the log of the average daily standard deviation of the true price process (*lsdprice*).[9] To control for price discreteness, we add the average price to the regressors, i.e. *price*. As underlined by Stoll (2000), this last variable can also be interpreted as an additional proxy for risk in that low price stocks have a tendency to be riskier. Finally, we account for potential Nasdaq effects by adding a Nasdaq dummy (*nasdaq*).

The results are reported in Table 2. In agreement with the predictions of the operating-cost theory of spread determination *lsize* has a significantly negative impact on the *FITC*'s (i.e., the corresponding elasticity is equal to -0.184 with a t-stat of -2.21). Consistently with the predictions of the asymmetric information theory, the coefficient on *lturn* is significantly positive (i.e., 0.218) with a t-stat of about 4. Similarly, strong and positive is the cross-sectional relation between the *FITC*'s and the volatility of the underlying full-information price (the estimated coefficient is equal to 0.217 with a t-stat of 2.87). As expected, the coefficient on *price* is negative (i.e., -0.0067) and highly statistically significant (t-stat of -4.18).

The signs and/or the significance levels of the estimated coefficient on *ltrades* and *nasdaq* are somewhat surprising at first. We start with *nasdaq*. It is generally believed that the decentralized nature of this venue leaves the dealer more exposed to potential losses coming from trading with the informed agents (Heidle and Huang (2002)). Naturally, the higher risk of informed trading would have to be compensated through larger spreads. The opposite seems to occur here but a simple observation justifies this result. Our sample of Nasdaq stocks (i.e., the Nasdaq stocks in the S&P 100 index) is very small (only 7 companies) and characterized by large cap stocks that trade very frequently and have large (average) volumes. What we are picking up, then, is a liquidity effect, hence

[9]We estimate the daily standard devations by computing the square root of the quadratic variation of the underlying log price process, that is by summing intradaily squared first differences of the log prices (c.f., Andersen et al. (2003) and Barndorff-Nielsen and Shephard (2002)). The optimal number of intradaily observations is chosen to minimize the conditional mean-squared error of the volatility estimator as suggested in Bandi and Russell (2003a,b).

the negative sign. As for *ltrades*, this variable is highly correlated with *lsize* and *nasdaq*. Thus, collinearity is giving rise to an insignificant (positive) coefficient on it. The adjusted R^2 on the regression is equal to 64.98%. Its value barely drops when removing *ltrades* (i.e., 64.94%). Equivalently, the estimated coefficients on the remaining variables do not change by much (see Table 3).

Interestingly, when we add the log of the average half quoted spread (*lspread*) to the regression, the explanatory power of the variables that proxy for liquidity appears to be subsumed by it. On the contrary, *lturns* retains some of its significance (see Table 4). This result is consistent with the idea that one of the components of the variance of the *FITC*'s is the variance of the difference between the efficient price and the full-information price. This component should not be explained perfectly by variables that determine the quoted spreads. In fact, such variables should largely explain the variance of the difference between transaction prices and efficient prices. Instead, the difference between the efficient price and the full-information price should be explained by variables that are correlated with private information. This is what we are finding.

We can look at the same problem from a slightly different angle and regress the difference between the log of the percentage *FITC*'s and the log of the estimated effective spreads on *lturns* (in Table 5). Again, the variance of this difference should be explained by variables that are correlated with private information. In effect, we find that the estimated coefficient on *lturns* is positive (i.e., 0.1834) and highly statistically significant (with a t-stat of 5.354), thereby suggesting that the positive difference between *FITC*'s and conventional measures of the cost of trade can be at least partly imputed to asymmetries in information.

7 Conclusions

In a world with private information and learning on the part of the market participants, transaction costs should be defined as the (positive) differences between transaction prices and full-information prices, i.e., the prices that reflect all information*, private and public,* about the asset of interest. While the current literature largely focuses on measuring the difference between transaction prices and efficient prices, i.e., the prices that embed all *publicly* available information about the asset, this paper proposes a simple and robust methodology to identify what we call "full-information transaction costs."

Our method relies on constructing straightforward sample moments using observable (high-frequency) transaction price data. As such, the method is easy to implement. Furthermore, it is robust to a variety of realistic price formation mechanisms in that it can accommodate predictability in the underlying full-information return process, correlation between the full-information price process and the cost of trade, as well as serial dependence in the cost of trade as implied by conventional operating-cost and adverse-selection theories of price determination.

We confirm the empirical validity of these theories on the basis of estimates of the full-information transaction costs for the cross-section of S&P 100 stocks.

Two final observations are in order. While the present paper focuses on a characterization of aspects of the unconditional distribution of the full-information cost of trade, similar tools can be employed to learn about the conditional properties of it. Such properties are expected to provide much information about the genuine market microstructure dynamics.

Furthermore, since individuals are likely to take into account the effective cost of acquiring and rebalancing their portfolios, expected stock returns should embed transaction 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 Easly and O'Hara (2002)). Nonetheless, the current attempts to characterize the cross-sectional relationship between expected returns and cost of trade either rely on liquidity-based theories of cost determination (Amihud and Mendelson (1986), among others) or they rely on information-based approaches to the same issue (Easly et al. (2002)). Our methodology to measure full-information transaction costs provides a natural set of tools to bridge the two arguments in the study of the cross-sectional dependence between expected stock returns and execution costs. Research on both subjects is being conducted by the authors and will be reported in later work (Bandi and Russell (2004a,b)).

8 Appendix

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

$$\begin{aligned}
\mathbf{E}(\varepsilon\varepsilon_{-j}) &= \mathbf{E}\left(\left(\eta-\eta_{-1}\right)\left(\eta_{-j}-\eta_{-(j+1)}\right)\right) && (13)\\
&= \mathbf{E}(\eta\eta_{-j})-\mathbf{E}(\eta\eta_{-(j+1)})-\mathbf{E}(\eta_{-1}\eta_{-j})+\mathbf{E}(\eta_{-1}\eta_{-(j+1)}) && (14)\\
&= 2\mathbf{E}(\eta\eta_{-j})-\mathbf{E}(\eta_{-1}\eta_{-j})-\mathbf{E}(\eta\eta_{-(j+1)}). && (15)
\end{aligned}$$

Recall, k is the maximum lag for which the covariances in the price contaminations (i.e., $\mathbf{E}(\eta\eta_{-k})$) are different from zero. Hence,

$$\mathbf{E}(\varepsilon\varepsilon_{-j}) = 2\mathbf{E}(\eta\eta_{-j})-\mathbf{E}(\eta_{-1}\eta_{-j}) \quad (16)$$

when $j=k$ and

$$\mathbf{E}(\varepsilon\varepsilon_{-j}) = 2\mathbf{E}(\eta\eta_{-j})-\mathbf{E}(\eta_{-1}\eta_{-j})-\mathbf{E}(\eta\eta_{-(j+1)}) \quad (17)$$

for $1\leq j<k$. We now plug Eq. (16) and Eq. (17) into the right-hand side of the squared version of Eq. (5) and obtain

$$\begin{aligned}
&\left(\frac{1+k}{2}\right)\mathbf{E}(\varepsilon^2)+\sum_{s=0}^{k-1}(s+1)\mathbf{E}(\varepsilon\varepsilon_{-(k-s)})\\
=\ &\left(\frac{1+k}{2}\right)\left[2\mathbf{E}(\eta^2)-2\mathbf{E}(\eta\eta_{-1})\right]\\
&+\sum_{s=1}^{k-1}(s+1)\left[2\mathbf{E}(\eta\eta_{-(k-s)})-\mathbf{E}(\eta_{-1}\eta_{-(k-s)})-\mathbf{E}(\eta\eta_{-((k-s)+1)})\right]\\
&+2\left[\mathbf{E}(\eta\eta_{-k})-\mathbf{E}(\eta_{-1}\eta_{-k})\right] && (18)\\
=\ &(1+k)\left(\mathbf{E}(\eta^2)-\mathbf{E}(\eta\eta_{-1})\right)\\
&+k\left[2\mathbf{E}(\eta\eta_{-1})-\mathbf{E}(\eta^2)-\mathbf{E}(\eta\eta_{-2})\right]\\
&+(k-1)\left[2\mathbf{E}(\eta\eta_{-2})-\mathbf{E}(\eta\eta_{-1})-\mathbf{E}(\eta\eta_{-3})\right]\\
&+(k-2)\left[2\mathbf{E}(\eta\eta_{-3})-\mathbf{E}(\eta\eta_{-2})-\mathbf{E}(\eta\eta_{-4})\right]\\
&+(k-3)\left[2\mathbf{E}(\eta\eta_{-4})-\mathbf{E}(\eta\eta_{-3})-\mathbf{E}(\eta\eta_{-5})\right]\\
&+...\\
&+3[2\mathbf{E}(\eta\eta_{-k+2})-\mathbf{E}(\eta\eta_{-k+3})-\mathbf{E}(\eta\eta_{-k+1})]\\
&+2[2\mathbf{E}(\eta\eta_{-k+1})-\mathbf{E}(\eta\eta_{-k+2})-\mathbf{E}(\eta\eta_{-k})]\\
&+[2\mathbf{E}(\eta\eta_{-k})-\mathbf{E}(\eta\eta_{-k+1})] && (19)
\end{aligned}$$

Finally, we notice that Eq. (19) is equal to $\mathbf{E}(\eta^2)$. This proves the stated result. ■

Proof of Theorem 1. Write

$$\frac{\sum_{i=j+1}^{N(\overline{h})}\widetilde{r}_i\widetilde{r}_{i-j}}{N(\overline{h})} = \underbrace{\frac{\sum_{i=j+1}^{N(\overline{h})} r_{t_i}r_{t_{i-j}}}{N(\overline{h})}}_{\alpha} + \underbrace{\frac{\sum_{i=j+1}^{N(\overline{h})} r_{t_i}\varepsilon_{i-j}}{N(\overline{h})}}_{\beta} + \underbrace{\frac{\sum_{i=j+1}^{N(\overline{h})} \varepsilon_i r_{t_{i-j}}}{N(\overline{h})}}_{\gamma} + \underbrace{\frac{\sum_{i=j+1}^{N(\overline{h})} \varepsilon_i\varepsilon_{i-j}}{N(\overline{h})}}_{\zeta} \quad (20)$$

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

$$\begin{aligned}
\alpha &\leq \frac{\left(\sum_{i=j+1}^{N(\overline{h})} r_{t_i}^2\right)^{1/2}\left(\sum_{i=j+1}^{N(\overline{h})} r_{t_{i-j}}^2\right)^{1/2}}{N(\overline{h})} && (21)\\
&= \frac{\left(\sum_{i=1}^{N(\overline{h})} r_{t_i}^2 - \sum_{i=1}^{j} r_{t_i}^2\right)^{1/2}\left(\sum_{i=1}^{N(\overline{h})} r_{t_i}^2 - \sum_{i=N(\overline{h})-j+1}^{N(\overline{h})} r_{t_i}^2\right)^{1/2}}{N(\overline{h})} && (22)\\
&\leq \frac{\left([\log p]_0^{\overline{h}} + o_p(1) + jO_p\left(\pi_i^{N(\overline{h})}\right)\right)^{1/2}\left([\log p]_0^{\overline{h}} + o_p(1) + jO_p\left(\pi_i^{N(\overline{h})}\right)\right)^{1/2}}{N(\overline{h})} && (23)\\
&\underset{N(\overline{h})\to\infty}{\overset{p}{\to}} 0, && (24)
\end{aligned}$$

where $[\log p]_0^{\overline{h}} = [M]_0^{\overline{h}} = \int_0^{\overline{h}} \sigma_s^2 ds$ is the quadratic variation of the underlying log price process. It is noted that Eq. (21) derives from the Cauchy's inequality while Eq. (23) derives from a standard convergence result in semimartingale process theory (see Revuz and Yor, Proposition 1.18, page 121, 1994, for example). Specifically, under Assumption 1,

$$\sum_{i=1}^{N(\overline{h})}\left(\log p_{t_i} - \log p_{t_{i-1}}\right)^2 \underset{N(\overline{h})\to\infty}{\overset{p}{\to}} [\log p]_0^{\overline{h}} = [M]_0^{\overline{h}} \quad (25)$$

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

$$\begin{aligned}
\mathbf{1}\left\{|\zeta - \mathbf{E}(\varepsilon\varepsilon_{-j})| > \delta\right\} &< \frac{|\zeta - \mathbf{E}(\varepsilon\varepsilon_{-j})|}{\delta} && (26)\\
&< \frac{(\zeta - \mathbf{E}(\varepsilon\varepsilon_{-j}))^2}{\delta^2}, && (27)
\end{aligned}$$

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, taking expectations of both sides of Eq. (27), we obtain

$$\begin{aligned}
P\left\{|\zeta - \mathbf{E}(\varepsilon\varepsilon_{-j})| > \varepsilon\right\} &< \frac{\mathbf{E}(\zeta - \mathbf{E}(\varepsilon\varepsilon_{-j}))^2}{\delta^2} && (28)\\
&= \frac{1}{\delta^2 N^2(\overline{h})}\mathbf{V}\left(\sum_{i=j+1}^{N(\overline{h})}\varepsilon_i\varepsilon_{i-j}\right). && (29)
\end{aligned}$$

But,

$$\begin{aligned}
\mathbf{V}\left(\sum_{i=j+1}^{N(\overline{h})}\varepsilon_i\varepsilon_{i-j}\right) &= \mathbf{E}\left(\sum_{i=j+1}^{N(\overline{h})}\varepsilon_i\varepsilon_{i-j}\right)^2 - \left(\sum_{i=j+1}^{N(\overline{h})}\mathbf{E}(\varepsilon_i\varepsilon_{i-j})\right)^2 && (30)\\
&= \mathbf{E}\left(\sum_{i=j+1}^{N(\overline{h})}\sum_{g=j+1}^{N(\overline{h})}\varepsilon_i\varepsilon_{i-j}\varepsilon_g\varepsilon_{g-j}\right) - \sum_{i=j+1}^{N(\overline{h})}\sum_{g=j+1}^{N(\overline{h})}\mathbf{E}(\varepsilon_i\varepsilon_{i-j})\mathbf{E}(\varepsilon_g\varepsilon_{g-j}) && (31)
\end{aligned}$$

$$= \sum_{i=j+1}^{N(\overline{h})} \sum_{g=j+1}^{N(\overline{h})} \mathbf{E}(\varepsilon_i \varepsilon_{i-j} \varepsilon_g \varepsilon_{g-j}) - \sum_{i=j+1}^{N(\overline{h})} \sum_{g=j+1}^{N(\overline{h})} \mathbf{E}(\varepsilon_i \varepsilon_{i-j}) \mathbf{E}(\varepsilon_g \varepsilon_{g-j}) \quad (32)$$

$$= \sum_{i=j+1}^{N(\overline{h})} \mathbf{E}\left(\varepsilon_i^2 \varepsilon_{i-j}^2\right) + 2 \sum_{g=j+1}^{N(\overline{h})} \sum_{i<g} \mathbf{E}(\varepsilon_i \varepsilon_{i-j} \varepsilon_g \varepsilon_{g-j})$$

$$+ \sum_{i=j+1}^{N(\overline{h})} \left(\mathbf{E}(\varepsilon_i \varepsilon_{i-j})\right)^2 + 2 \sum_{g=j+1}^{N(\overline{h})} \sum_{i<g} \mathbf{E}(\varepsilon_i \varepsilon_{i-j}) \mathbf{E}(\varepsilon_g \varepsilon_{g-j}). \quad (33)$$

The variance term $\mathbf{V}$ grows linearly with $N(\overline{h})$ since the terms in the four summations are bounded by the assumption

$$\mathbf{E}(\eta_s \eta_q \eta_{s-u} \eta_{q-p}) < \infty, \quad (34)$$

$\forall s, q$ and $\forall u, p = 0, ..., k$. Then, the right-hand side of Eq. (29) vanishes as $N(\overline{h}) \to \infty$. This proves convergence in mean-squared (and, by Eq. (29), in probability) of term ζ to $\mathbf{E}\left(\varepsilon \varepsilon_{-j}\right)$. Now we turn to β.

$$\beta \leq \frac{\left(\sum_{i=j+1}^{N(\overline{h})} r_{t_i}^2\right)^{1/2} \left(\sum_{i=j+1}^{N(\overline{h})} \varepsilon_{i-j}^2\right)^{1/2}}{N(\overline{h})} \quad (35)$$

$$= \frac{1}{\sqrt{N(\overline{h})}} \left([\log p]_0^{\overline{h}} + o_p(1) + j O_p\left(\pi_i^{N(\overline{h})}\right)\right)^{1/2} \left(\frac{\sum_{i=j+1}^{N(\overline{h})} \varepsilon_{i-j}^2}{N(\overline{h})}\right)^{1/2} \quad (36)$$

$$= \frac{1}{\sqrt{N(\overline{h})}} O_p(1) \left(\mathbf{E}(\varepsilon^2) + o_p(1)\right) \underset{N(\overline{h}) \to \infty}{\overset{p}{\to}} 0, \quad (37)$$

where, as earlier, Eq. (35) follows from the Cauchy's inequality and the convergence in probability of $\frac{\sum_{i=j+1}^{N(\overline{h})} \varepsilon_{i-j}^2}{N(\overline{h})}$ 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:

$$\gamma \leq \frac{\left(\sum_{i=j+1}^{N(\overline{h})} \varepsilon_i^2\right)^{1/2} \left(\sum_{i=j+1}^{N(\overline{h})} r_{t_{i-j}}^2\right)^{1/2}}{N(\overline{h})} \quad (38)$$

$$= \frac{1}{\sqrt{N(\overline{h})}} O_p(1) \left(\mathbf{E}(\varepsilon^2) + o_p(1)\right) \underset{N(\overline{h}) \to \infty}{\overset{p}{\to}} 0. \quad (39)$$

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 microstructure frictions 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 volatility, microstructure noise, and optimal sampling. *Working paper, GSB - University of Chicago.*

[6] Bandi, F.M., and J. R. Russell (2003b). Separating microstructure noise from volatility. *Working paper, GSB - University of Chicago.*

[7] Bandi, F.M., and J.R. Russell (2004a). Forecasting transaction costs. *Working paper, GSB - University of Chicago.*

[8] Bandi, F.M., and J.R. Russell (2004b). Expected stock returns and full-information transaction costs. *Working paper, GSB - University of Chicago.*

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

[11] 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.

[12] Conrad, J., and G. Kaul (1988). Time-variation in expected returns. *The Journal of Business,* 61, 409-425.

[13] Conrad, J., and G. Kaul (1988). Mean-reversion in short-horizon expected returns. *Review of Financial Studies,* 2, 225-240.

[14] Conrad, J., G. Kaul and M. Nimalendram (1990). Components of short-horizon individual security returns. *Journal of Financial Economics.*

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

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

[17] Diamond, D.W. and R.E. Verrecchia (1987). Constraints on short-selling and asset price adjustments to private information. *Journal of Financial Economics,* 18, 277-311.

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

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

[20] Easley D., and M. O'Hara (1992). Time and the process of security price adjustment. *Journal of Finance,* 47, 577-605.

[21] 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.

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

[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] George, T., K. Gautam, and M. Nimalendram (1991). Estimation of the bid-ask spread and its components: a new approach. *Review of Financial Studies*, 4, 623-656.

[28] Glosten, L.R. (1987). Components of the bid-ask spread and the statistical properties of transaction prices. *Journal of Finance,* 42, 1293-1308.

[29] 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.

[30] Hasbrouck, J. (1991). Measuring the information content of stock trades. *Journal of Finance,* 46, 179-207.

[31] Hasbrouck, J. (1999). Liquidity in the future pits: inferring market dynamics from incomplete data. *Working paper, Stern School of Business, New York University.*

[32] Hasbrouck, J. (2003). Trading costs and returns for U.S. securities: the evidence from daily data. *Working paper, Stern School of Business, New York University.*

[33] 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.

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

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

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

[37] Loeb, T.F. (1983). Trading cost: the critical link between investment information and results. *Financial Analysts Journal,* 39, 39-43.

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

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

[40] 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.

[41] Revuz, D. and M. Yor (1994). *Continuous Martingales and Brownian Motion.* Third Edition, Springer-Verlag, New York.

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

[43] Seppi, D. (1997). Liquidity provision with limit orders and a strategic specialist. *Review of Finacial Studies,* 10, 103-150.

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

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

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

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

Table 1

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 effective spreads (computed using the Lee and Ready (1991) algorithm and a standard 5 second time allowance), the annualized standard deviation of the underlying full-information log price process (calculated using realized volatility and the optimal sampling method discussed in Bandi and Russell (2003a,b)) and the corresponding turnover measures, i.e., the logs of the average number of daily shares transacted to shares outstanding.

Symbol	Duration	Avg Price	*FITC*	$ *FITC*	Eff. Spread	Ann SD	Turnover
AA	11.30	36.10	0.08%	$0.0286	0.05%	30.78%	0.17%
AEP	21.61	41.77	0.05%	$0.0194	0.04%	20.98%	0.22%
AES	15.57	7.79	0.44%	$0.0345	0.27%	168.84%	0.84%
AIG	10.41	72.90	0.06%	$0.0445	0.04%	30.12%	0.09%
ALL	15.37	33.93	0.10%	$0.0346	0.05%	33.80%	0.16%
AMGN*	1.30	57.82	0.04%	$0.0248	0.04%	57.62%	0.04%
AOL	6.29	25.17	0.09%	$0.0223	0.08%	50.30%	0.11%
ATI	64.93	15.64	0.17%	$0.0266	0.09%	34.89%	0.51%
AVP	21.76	49.10	0.06%	$0.0318	0.04%	20.80%	0.33%
AXP	8.72	34.13	0.10%	$0.0344	0.05%	35.50%	0.11%
BA	10.76	43.51	0.07%	$0.0300	0.05%	29.58%	0.17%
BAC	7.18	61.20	0.06%	$0.0344	0.04%	27.06%	0.10%
BAX	17.33	55.32	0.08%	$0.0437	0.04%	22.42%	0.22%
BCC	31.56	34.98	0.11%	$0.0374	0.05%	27.79%	1.07%
BDK	23.97	43.50	0.10%	$0.0445	0.04%	25.98%	0.97%
BHI	14.46	34.53	0.13%	$0.0447	0.06%	37.32%	0.39%
BMY	8.94	45.08	0.08%	$0.0344	0.04%	27.52%	0.09%
BNI	24.76	27.95	0.08%	$0.0231	0.06%	27.11%	0.33%
BUD	20.03	48.69	0.08%	$0.0368	0.04%	16.96%	0.19%
C	5.88	44.43	0.08%	$0.0367	0.05%	34.50%	0.06%
CCU	12.03	47.06	0.07%	$0.0337	0.05%	38.18%	0.24%
CI	20.13	92.46	0.08%	$0.0712	0.05%	33.13%	0.41%
CL	15.19	55.64	0.06%	$0.0321	0.04%	21.68%	0.18%
CPB	25.95	27.02	0.14%	$0.0368	0.07%	26.97%	0.17%
CSC	18.79	47.07	0.11%	$0.0524	0.06%	38.05%	0.46%
CSCO*	0.45	16.69	0.04%	$0.0065	0.10%	33.91%	0.02%
DAL	17.23	32.76	0.13%	$0.0422	0.06%	38.89%	0.60%
DD	10.77	45.10	0.07%	$0.0298	0.04%	27.43%	0.13%
DIS	8.52	23.29	0.07%	$0.0152	0.06%	33.99%	0.11%
DOW	17.08	30.05	0.09%	$0.0279	0.05%	30.45%	0.23%
EK	16.98	29.24	0.07%	$0.0215	0.06%	33.99%	0.52%
EMC	7.83	13.43	0.09%	$0.0117	0.12%	58.67%	0.20%
EP	15.41	37.02	0.06%	$0.0240	0.07%	41.29%	0.27%
ETR	27.00	41.09	0.07%	$0.0289	0.04%	22.19%	0.35%
EXC	22.36	50.01	0.06%	$0.0297	0.04%	25.59%	0.23%
F	12.27	14.69	0.11%	$0.0167	0.07%	28.98%	0.18%
FDX	14.66	54.75	0.06%	$0.0332	0.04%	25.05%	0.25%
G	18.72	33.23	0.08%	$0.0258	0.05%	27.06%	0.14%
GD	17.33	89.14	0.09%	$0.0759	0.05%	23.98%	0.41%
GE	4.68	37.50	0.06%	$0.0215	0.05%	32.52%	0.03%
GM	13.93	51.66	0.06%	$0.0291	0.03%	24.08%	0.35%
GS	8.55	82.28	0.05%	$0.0435	0.05%	31.38%	0.22%
HAL	11.35	15.21	0.17%	$0.0263	0.10%	60.39%	0.51%
HCA	16.42	42.32	0.05%	$0.0219	0.05%	29.83%	0.31%

Symbol	Duration	Avg Price	*FITC*	$ *FITC*	Eff. Spread	Ann SD	Turnover
HD	7.29	50.40	0.05%	$0.0264	0.04%	25.93%	0.07%
HET	29.74	38.34	0.07%	$0.0276	0.07%	31.46%	0.77%
HIG	17.81	65.73	0.06%	$0.0422	0.05%	26.65%	0.23%
HNZ	20.74	40.98	0.09%	$0.0380	0.04%	21.10%	0.22%
HON	12.52	34.34	0.09%	$0.0319	0.06%	53.83%	0.26%
HWP	9.90	20.57	0.07%	$0.0149	0.06%	37.95%	0.13%
IBM	6.51	102.82	0.06%	$0.0583	0.04%	26.65%	0.10%
INTC*	0.52	31.76	0.04%	$0.0116	0.07%	57.94%	0.01%
IP	11.36	42.92	0.06%	$0.0262	0.04%	26.69%	0.25%
JNJ	10.48	57.89	0.05%	$0.0308	0.04%	22.47%	0.08%
JPM	7.64	29.83	0.06%	$0.0177	0.07%	48.35%	0.19%
KO	11.14	46.32	0.06%	$0.0258	0.04%	22.80%	0.07%
LEH	10.98	59.00	0.06%	$0.0378	0.05%	36.61%	0.44%
LTD	21.29	17.60	0.14%	$0.0238	0.06%	37.55%	0.53%
LU	10.84	5.71	0.16%	$0.0093	0.14%	52.87%	0.15%
MAY	24.27	35.55	0.12%	$0.0430	0.05%	28.72%	0.26%
MCD	12.36	26.81	0.05%	$0.0135	0.05%	21.97%	0.12%
MDT	10.77	47.01	0.06%	$0.0293	0.04%	28.24%	0.11%
MEDI*	3.69	40.81	0.08%	$0.0345	0.06%	27.29%	0.17%
MER	7.00	47.63	0.11%	$0.0512	0.06%	38.63%	0.19%
MMM	12.03	114.97	0.05%	$0.0540	0.04%	23.61%	0.21%
MO	10.04	51.28	0.07%	$0.0347	0.04%	15.26%	0.08%
MRK	10.13	59.96	0.06%	$0.0371	0.04%	20.37%	0.08%
MSFT*	0.55	60.16	0.03%	$0.0154	0.05%	66.28%	0.01%
MWD	6.86	49.72	0.07%	$0.0341	0.06%	38.11%	0.13%
NSC	20.71	21.79	0.15%	$0.0321	0.07%	22.58%	0.25%
NSM	13.64	26.60	0.15%	$0.0404	0.09%	47.98%	0.57%
NXTL*	1.10	5.07	0.18%	$0.0093	0.56%	218.34%	0.16%
ONE	13.13	35.58	0.09%	$0.0332	0.05%	32.75%	0.14%
ORCL*	0.77	16.00	0.05%	$0.0075	0.07%	106.93%	0.02%
PEP	9.85	49.71	0.07%	$0.0350	0.04%	20.19%	0.08%
PFE	6.34	41.06	0.06%	$0.0234	0.04%	17.10%	0.03%
PG	9.38	84.00	0.05%	$0.0457	0.03%	50.30%	0.08%
ROK	46.75	18.71	0.11%	$0.0199	0.09%	31.62%	0.48%
RSH	19.33	27.70	0.11%	$0.0308	0.07%	28.59%	0.82%
RTN	19.30	37.89	0.08%	$0.0309	0.05%	51.06%	0.45%
S	14.97	52.76	0.09%	$0.0451	0.04%	27.20%	0.37%
SBC	7.79	36.56	0.08%	$0.0287	0.05%	27.57%	0.05%
SLB	8.79	55.87	0.09%	$0.0478	0.05%	27.70%	0.17%
SLE	21.21	21.36	0.07%	$0.0156	0.06%	14.70%	0.16%
SO	22.73	25.00	0.09%	$0.0227	0.05%	23.35%	0.13%
T	10.92	15.59	0.10%	$0.0155	0.07%	43.01%	0.16%
TOY	26.66	17.63	0.13%	$0.0236	0.07%	45.74%	0.80%
TXN	6.93	30.52	0.11%	$0.0349	0.06%	54.86%	0.14%
TYC	5.16	29.35	0.12%	$0.0360	0.16%	93.69%	0.32%
UIS	32.56	11.66	0.16%	$0.0183	0.09%	12.55%	0.32%
USB	16.97	19.94	0.06%	$0.0119	0.06%	17.54%	0.10%
UTX	11.81	69.40	0.06%	$0.0412	0.04%	48.66%	0.22%
VIAB	11.32	42.44	0.08%	$0.0352	0.05%	25.05%	0.27%
VZ	8.44	45.76	0.07%	$0.0329	0.04%	29.79%	0.05%
WFC	9.13	46.10	0.04%	$0.0176	0.04%	22.86%	0.08%
WMB	11.30	15.97	0.22%	$0.0352	0.12%	69.80%	0.40%

Symbol	Duration	Avg. Price	*FITC*	*$FITC*	Eff. Spread	Ann. SD	Turnover
WMT	8.30	60.02	0.05%	$0.0283	0.04%	24.14%	0.05%
WY	15.77	59.75	0.07%	$0.0435	0.04%	82.04%	0.31%
XOM	7.54	39.41	0.05%	$0.0189	0.04%	19.43%	0.04%
XRX	21.22	10.12	0.15%	$0.0151	0.11%	8.28%	0.35%
Average	**14.22**	**40.70**	**0.09%**	**$0.0307**	**0.06%**	**37.30%**	**0.25%**

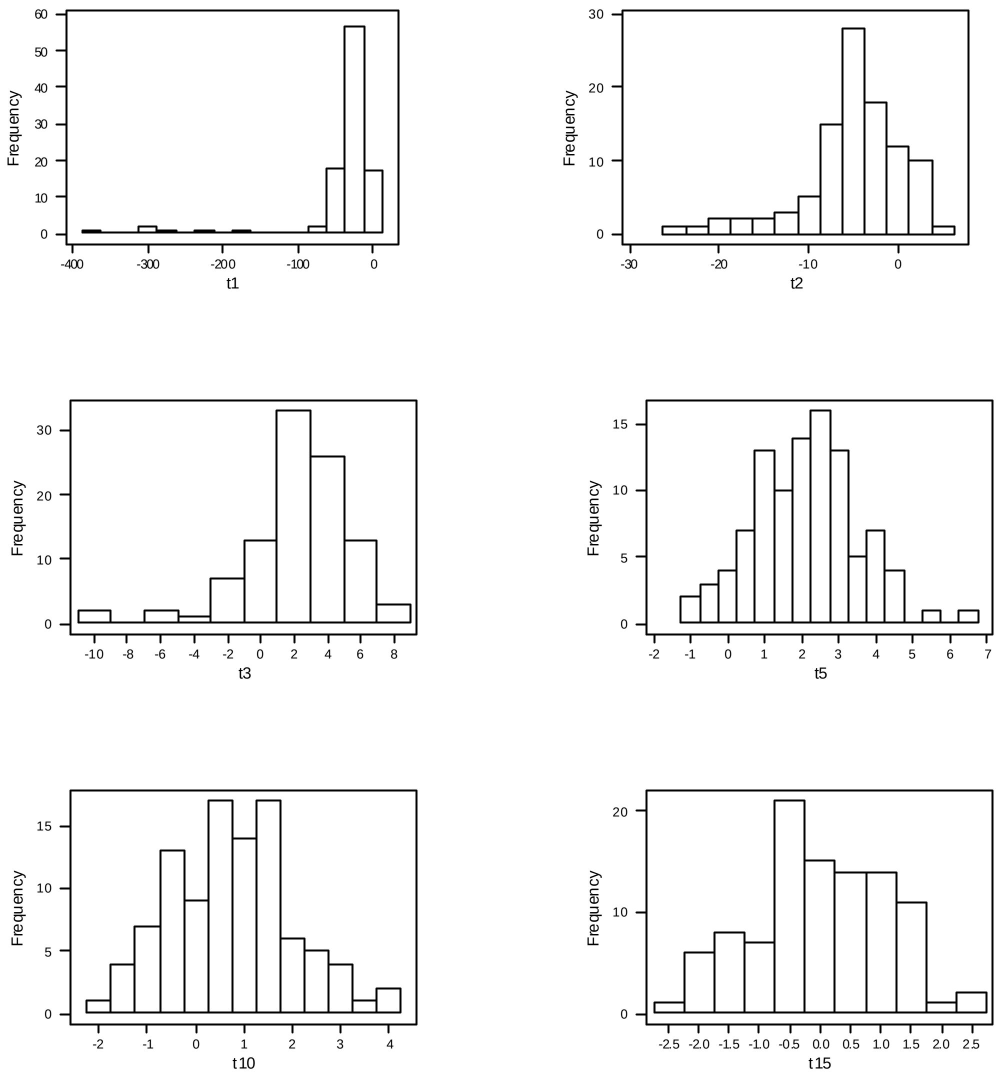

Figure 1. Histograms of the t-ratios of the estimated serial correlations of order 1,2,3,5,10 and 15 of the transaction prices of the S&P 100 stocks.

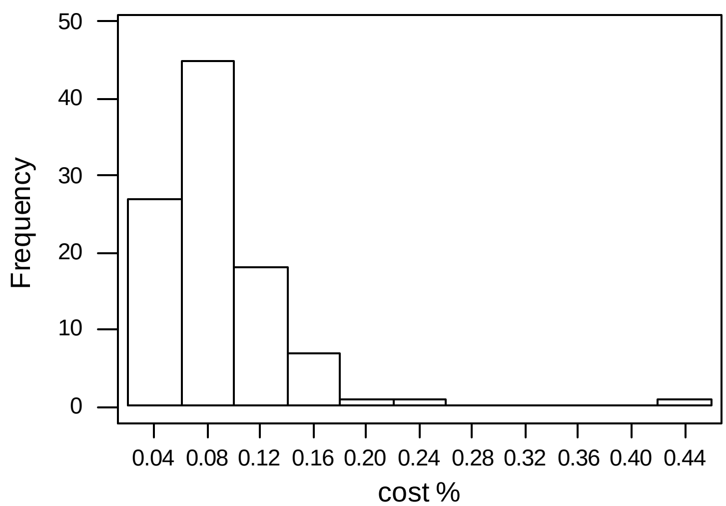

Figure 2. Histogram of the estimated *FITC*'s (in percentage values) of the S&P 100 stocks.

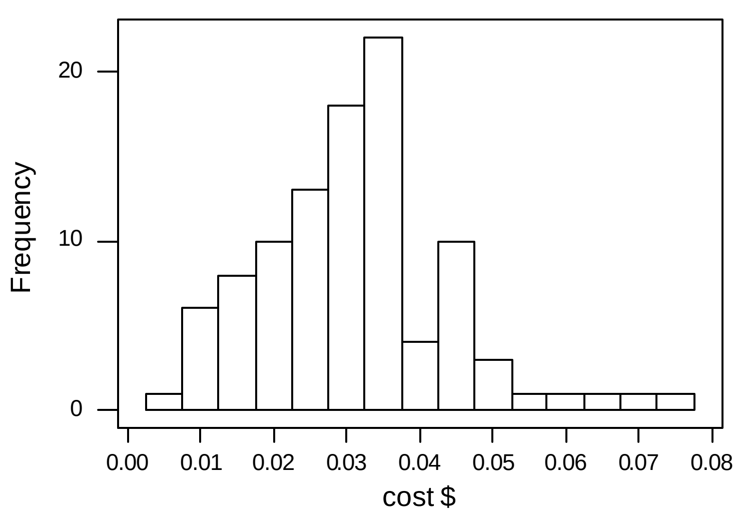

Figure 3. Histogram of the estimated *FITC*'s (in dollar values) of the S&P 100 stocks.

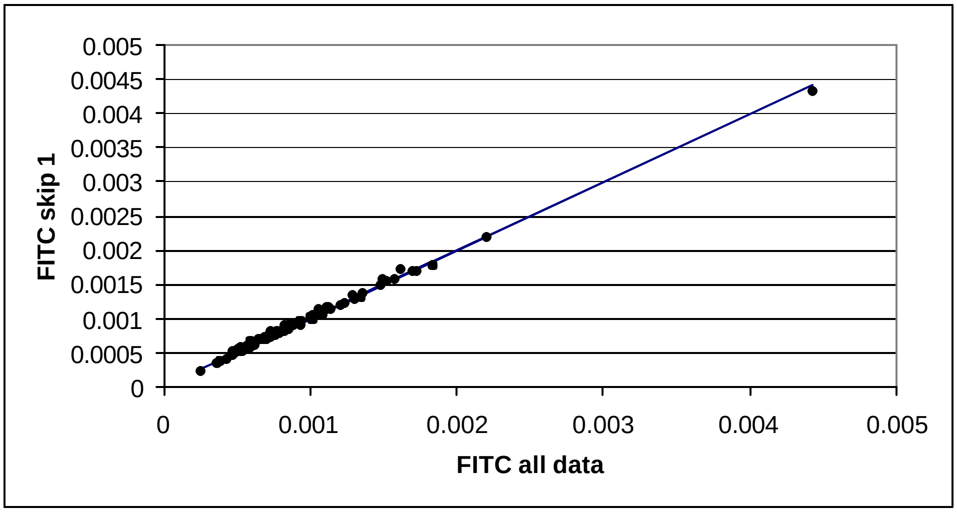

Figure 4. Scatter plot of the estimated *FITC*'s using all transactions versus *FITC*'s that are estimated using every other transaction. The straight line is the 45 degree line.

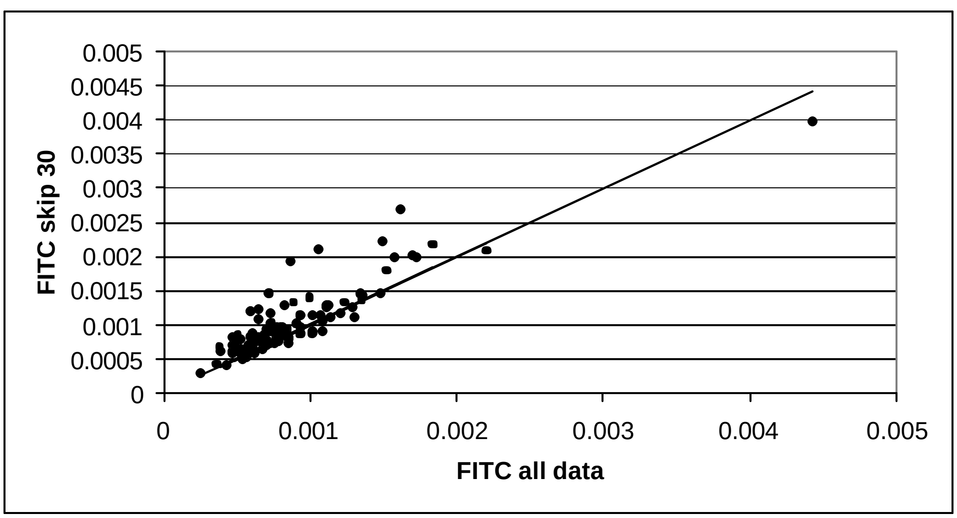

Figure 5. Scatter plot of the estimated *FITC*'s using all transactions versus *FITC*'s that are estimated using every 30th transaction. The straight line is the 45 degree line.

	Coefficient	Std. Errors	T-statistics	P-values
Intercept	-3.4684	0.6968	-4.977	0.000
lturn	0.2186	0.0544	4.017	0.000
lsize	-0.1846	0.0834	-2.212	0.029
lsdprice	0.2177	0.0756	2.878	0.005
ltrades	0.1184	0.1116	1.061	0.291
price	-0.0067	0.0016	-4.182	0.000
nasdaq	-0.7859	0.3123	-2.516	0.0136
	R^2=67.1%	$adjR^2$=64.9%		

Table 2. Outcome of a regression of the log of the percentage *FITC*'s on the log of the average number of daily shares transacted to shares outstanding (lturn), the log of the average dollar volume per trade (lsize), the log of the average daily standard deviation of the true price process (lsdprice), the log of the average number of trades per day (ltrades), the average price (price) and a Nasdaq dummy (nasdaq).

	Coefficient	Std. Errors	T-statistics	P-values
Intercept	-3.2426	0.6640	-4.882	0.000
lturn	0.1812	0.0414	4.369	0.000
lsize	-0.1296	0.0654	-1.980	0.050
lsdprice	0.2599	0.0643	4.041	0.000
price	-0.0071	0.0015	-4.520	0.000
nasdaq	-0.5126	0.1768	-2.898	0.0047
	R^2=66.71%	$adjR^2$=64.94%		

Table 3. Outcome of a regression of the log of the percentage *FITC*'s on the log of the average number of daily shares transacted to shares outstanding (lturn), the log of the average dollar volume per trade (lsize), the log of the average daily standard deviation of the true price process (lsdprice), the average price (price) and a Nasdaq dummy (nasdaq).

	Coefficient	Std. Errors	T-statistics	P-values
Intercept	-1.1169	0.6609	-1.689	0.094
lturn	0.0917	0.0380	2.410	0.017
lsize	-0.040	0.0573	-0.708	0.480
lsdprice	0.1144	0.0594	1.924	0.057
price	0.0005	0.0018	0.282	0.778
nasdaq	-0.1746	0.1597	-1.093	0.277
lspread	0.6528	0.1062	6.143	0.000
	R^2=76.31%	adjR^2=74.79%		

Table 4. Outcome of a regression of the log of the percentage *FITC*'s on the log of the average number of daily shares transacted to shares outstanding (lturn), the log of the average dollar volume per trade (lsize), the log of the average daily standard deviation of the true price process (lsdprice), the average price (price), a Nasdaq dummy (nasdaq), and the log of the average half quoted spread (lspread).

	Coefficient	Std. Errors	T-statistics	P-values
Intercept	1.5166	0.2197	6.901	0.000
lturn	0.1834	0.0342	5.354	0.000
	R^2=22.63%	adjR^2=21.84%		

Table 5. Outcome of a regression of the difference between the log of the percentage *FITC*'s and the log of the estimated effective spreads on the log of the average number of daily shares transacted to shares outstanding (lturn).