Measuring transaction costs in the absence of time stamps

Filip Zikes*

January 30, 2016

Preliminary, not for circulation.

Abstract

This paper develops consistent estimators of transaction costs in the absence reliable transaction time-stamps and information about who initiates trades. In addition to extending some current measures, I propose new measures of effective spread and investigate their statistics properties in asymptotic theory and simulations. I then allow the effective spread to smoothly vary over time and propose consistent kernel-based estimator of the time-varying spread. Finally, I study the relative performance of all estimators in a Monte Carlo experiment and present an empirical application to small-cap stocks using the TAQ data for the period 2005-2014. The theoretical, simulation and empirical results presented in the paper may help guide future empirical work on measuring transactions costs using data that suffer from the absence of time stamps and order direction information.

JEL Classification: C14, C15, G20
Keywords: transaction costs, effective spread, simulated method of moments, time-varying estimation

*Federal Reserve Board, Office of Financial Stability Policy and Research, 1801 K Street NW, Washington, D.C., 20006, United States. Phone: +1-202-475-6617. Email: filip.zikes@frb.gov. I am grateful to Margaret Yellen for excellent research assistance. All results reported in the paper were generated using programs written in the Ox language of Doornik (2007); the programs are available on request. The views expressed in this paper are the sole responsibility of the author and should not be interpreted as representing the views of the Federal Reserve Board or any other person associated with the Federal Reserve System.

1 Introduction

This paper develops consistent estimators of transaction costs in the absence reliable transaction time-stamps and information about who initiates trades. Several interesting over-the-counter (OTC) transactional datasets that have recently become available to academic and central-bank researchers suffer from these limitations.[1] While these transactions reports often contain rich information, such as the identity of counterparties, the lack of time stamps and trade direction entails a significant loss of infomation (in the statistical sense) and makes measuring transaction costs much more challenging. Yet, understanding liquidity and the cost of trading in these large financial markets, and how they vary over time, has probably never been of greater interest than in the recent turbulent years.

The literature on measuring transaction costs is fairly large (see Harris, 2015, for a review), but only a few papers deal with the specific data limitations described above. While all of the existing measures have sound micro- or statistical foundations, their sampling properties are not well-understood, and no comparison, in a unified theoretical framework, exists that would guide empirical work. Indeed, none of the existing estimators of the effective spread are consistent as the number of transactions increases, unlike, for example, the well-known measure of Roll (1984). Inspired by the ideas underlying the measures of Corwin and Schulz (2012) and Benos and Zikes (2015), I propose consistent estimators of the effective spread and study their sampling properties.

More specifically, I propose three measures. The first one is an extension of Benos and Zikes (2015) and it is available in closed form. This allows me to study its finite-sample properties analytically and compare it to the Roll's (1984) measure, explicitly quantifying the loss of information due to the missing time stamps. The other two measures combine the ideas of Corwin and Schulz (2012) and Benos and Zikes (2015) yielding more accurate estimators, albeit not in closed form. Since I make weaker assumptions on who initiates the trades than Corwin and Schulz (2012) do, I have to

[1] Examples include Chen et al. (2011), Benos et al. (2013), Biswas et al. (2014), Du et al. (2015) and Siriwardane (2015) who study the credit default swap market using data from the Depository Trust and Clearing Corporation (DTCC), Chen et al. (2012) who describe the interest rate swap market using data from the same source, Benos et al. (2015) who look at the interest rate swap market using proprietary transactions data from the London Clearing House, Benos and Zikes (2014) who investigate the UK sovereign bond market using audit-trail data obtained from the UK's Financial Conduct Authority (FCA), and FCA (2015) who examine all UK-listed bonds.

resort to the simulated method of moments (SMM). Despite having to approximate one moment by simulation, the estimator turns out to be computationally very cheap and easy to implement in practice. An important advantage of the SMM framework is that it allows me to test the validity of my model assumptions using the well-known test for overidentifying restrictions.

I then extend the effective spread measures in two important directions. First, I show some of my measures are robust to stationary stochastic volatility of the efficient price. This is clearly a desirable feature given the well-known evidence of volatility clustering in financial returns. Second, I allow the effective spread itself to vary (smoothly) over time and employ the recent advances in time-varying estimation to recover the sample path of the effective spread by kernel methods, originally proposed by Giraitis, Kapetanios and Yates (2014) in a different context. This allows me to uncover smooth changes in transaction costs occuring with the evolving structure of finacial markets over long periods of time.

Thoughout the paper, I run Monte Carlo simulations to corroborate the theoretical findings and to study how the various estimators perform in small samples. I also provide an empirical application to small-cap stocks listed on the New York Stock Exchange using data from the Trade and Quote database (TAQ). The results of these exercises may help guide future empirical work as they shed light on the relative performance of the various measures of effective spread in a controlled environment as well as in real data.

The rest of the paper is organized as follows. Section 2 reviews the existing measures of effective spread that are applicable in the absence of time stamps and trade direction. In Section 3, I propose the three new measures of effective spread, study their properties in theory and Monte Carlo simulations, and compare their performance. In Section 4, I allow for time-varying volatility and show how to consistently estimate time-varying effective spread. Section 6 presents an empirical application to small-cap stocks using TAQ data and Section 7 concludes. Technical results are collected in the Appendix.

2 Measures of effective spread

The effective spread is defined as two times the difference between the actual transaction price (P) and the prevailing mid-quote or some proxy for the true value of the

asset (efficient price) (M) at the time of the transaction. It can be expressed in absolute terms, i.e. $2|P-M|$, or in relative terms, i.e. $2|P-M|/M$ or $2|\log(P)-\log(M)|$. Like the bid-offer spread, the effective spread measures round-trip transaction costs, but it is based on actual transaction price rather than on quoted prices. The effective spread can be also seen as a measure of price impact of a trade, and since price impact and transaction costs tend to vary inversely with liquidity, it can be viewed as a measure of liquidity (Foucault, Pagano and Roell, 2013). Many different estimators of the effective spread exists depending on what information is available to the econometrician; see Harris (2015, Section 3.1) for an overview. In this section, I review those measures that can be used when time-stamps and trade direction are not observable. The consistent estimators of the effective spread that I propose in the next section build on, and extend the ideas undelying these measures.

Jankowitsch et al. (2011) propose a micro-founded absolute effective spread measure based on the dispersion of transaction prices from some benchmark price, typically a mid-quote. Their dispersion metric is given by

$$d_t = \sqrt{\frac{1}{\sum_{i=1}^{n} V_{i,t}} \sum_{i=1}^{n} (P_{i,t} - M_{0,t})^2 V_{i,t}}, \tag{1}$$

where $P_{i,t}$ and $V_{i,t}$ denote the price and volume associated with transaction i on day t and $M_{0,t}$ is some benchmark price or mid-quote on day t. The key assumption underlying this measure is that the benchmark price does not vary over the course of the day. If it does, d_t will pick up the volatility of the benchmark price, and it will be an upward biased measure of the effective spread.

Corwin and Schultz (2012) suggest a measure of relative effective spread based on daily high and low prices. Their key insight is that if the true price of the security follows a diffusion process "the sum of the price ranges over two consecutive single days reflects two days volatility and twice the spread, while the price range over one 2-day period reflects 2 days volatility and one spread". This gives two equations in two unknowns which can be solved to obtain an estimator of the effective spread. The main advantage of the Corwin and Schultz (2012) measure is that only intraday high and low prices are required, and these are available for many assets over long sample periods, unlike detailed transactions data.

Formally, let H_t^o and H_t^o denote the observed high and low prices on day t and

let $H^o_{t,t+1}$ and $L^o_{t,t+1}$ denote the high and low prices over two consecutive days t and $t+1$. Define

$$\beta_t = \sum_{j=0}^{1} \left[\log \left(\frac{H^o_{t+j}}{L^o_{t+j}} \right) \right]^2, \quad \gamma_t = \left[\log \left(\frac{H^o_{t,t+1}}{L^o_{t,t+1}} \right) \right]^2 \tag{2}$$

Then under the assumptions that the observed high and low prices are related to the high and low efficient prices by $H^o_t = H^A_t(1 + S/2)$ and $L^o_t = L^A_t(1 - S/2)$, where S is the proportional effective spread, Corwin and Schultz (2012) show that S can be estimated by

$$S_t = \frac{2(e^{\alpha_t} - 1)}{1 + e^{\alpha_t}}, \quad \alpha_t = \frac{\sqrt{2\beta_t} - \sqrt{\beta_t}}{3 - 2\sqrt{2}} - \sqrt{\frac{\gamma_t}{3 - 2\sqrt{2}}} \tag{3}$$

Unlike the measure of Jankowitsch et al. (2011), the range-based estimator allows the efficient price to vary intraday and corrects for the contribution of the intraday variation to the daily and two-day ranges. The measure S_t rests on two important assumptions though, namely that the daily high (low) prices are buyer (seller) - initiated and that the overnight returns are zero. Corwin and Schultz (2012) suggest some adjustments for the latter, but the former assumption is maintained. While this is a valid assumption for equity data as shown in their paper, this may not always be the case for infrequently OTC-traded assets. In the next section, I exploit the idea of using the range for estimating the effective spread, but I dispense with both of the above mentioned assumptions.

Inspired by Jankowitsch et al. (2011), Benos and Zikes (2015) also rely on the dispersion of transaction prices around some benchmark price, but they recognize that the dispersion metric is affected by the intraday volatility of the benchmark price in a non-trivial way. They suggest to use two dispersion metrics,

$$\hat{d}^2_t = \frac{1}{n} \sum_{i=1}^{n} (p_{i,t} - m_{0,t})^2, \quad \tilde{d}^2_t = \frac{1}{n-1} \sum_{i=1}^{n} (p_{i,t} - \bar{p}_t)^2, \tag{4}$$

where $p_{i,t} = \log P_{i,t}$ and $m_{0,t} = \log M_{0,t}$, and show that under the assumptions of the Roll (1984), which we spell out explicitly in the next section, the two metrics satisfy

$$\mathrm{E}(\hat{d}^2_t) = \frac{s^2}{4} + \frac{\sigma^2}{2} \left(\frac{n+1}{n} \right), \tag{5}$$

$$\mathrm{E}(\tilde{d}_t^2) = \frac{s^2}{4} + \frac{\sigma^2}{6}\left(\frac{n+1}{n}\right). \tag{6}$$

Solving for s^2, censoring at zero and taking the square root yields the relative effective spread measure:

$$ES_t = \sqrt{\max\{2(3\tilde{d}_t^2 - \hat{d}_t^2), 0\}}. \tag{7}$$

The framework of Benos and Zikes (2012) is similar to Corwin and Schulz (2012) in that the efficient price follows random walk, although the former do not require any assumptions on the overnight return or the distribution of the random walk innovations. On the other hand, Corwin and Schultz (2012) only require daily high and low prices, which are typically available for much longer periods of time than the more granular transactions data used by Jankowitsch et al. (2011) and Benos and Zikes (2015).

3 Consistent estimation of effective spread

Having introduced the main ideas used in the literature to measure the effective spread, I now exploit these ideas to construct consistent estimators of s. Achieving consistency essentially requires averaging the Corwin and Schulz (2012) and the Benos and Zikes (2015) estimators over an increasing number of days. I first study the latter, as it is available in closed form and permits analytical results, and then borrow the idea of the former and use the range together with the metrics of Benos and Zikes (2015) to construct new measuring via the simulated method of moments.

3.1 Framework and assumptions

Suppose we have a sample of T days and divide each day into n subintervals of equal length. I assume that a transaction arrives at the beginning of each of these subintervals and that the associated logarithmic transactions prices, $p_{i,t}$, $i = 1, ..., n$, $t = 1, ..., T$, are related to the logarithmic efficient price, $m_{i,t}$, by

$$p_{i,t} = m_{i,t} + \frac{s}{2} q_{i,t}, \quad i = 1, ..., n, \tag{8}$$

where s is the proportional effective spread and $q_{i,t}$ is a binary variable indicating whether the i-th transaction on day t is buyer-initiated $(+1)$ or seller-initiated (-1).

I initially assume that the efficient price is observable at the end of the day, i.e. at the end of the last subinterval n. I later relax this assumption and propose estimators that do not require observing m at all. Following Roll (1984) and Benos and Zikes (2015), I assume that the logarithmic efficient price m follows random walk with independent and identically distributed increments:

$$m_{i+1,t} = m_{i,t} + \epsilon_{i+1,t}, \quad i = 0, ..., n-1, \tag{9}$$

where $\mathrm{E}(\epsilon_{i,t}) = 0$ and $\mathrm{E}(\epsilon_{i,t}^2) = \sigma^2/n$. Thus, the daily integrated variance of the efficient price equals σ^2 for any n and t. This assumption will be relaxed in Section 4. Finally, I assume that $q_{i,t}$ is uncorrelated with $m_{j,s}$ for all i, j, t, s and that $q_{i,t}$ is serially uncorrelated with $\mathrm{E}(q_{i,t}) = \frac{1}{2}$, that is, there is the same number of buyer and seller initiated trades on average. I make no assumptions on the overnight return of the efficient price, $m_{0,t} - m_{n,t-1}$.

3.2 Baseline estimator

My baseline estimator is based on the idea of Benos and Zikes (2015). Define $\hat{s}_t^2 = 2(3\tilde{d}_t^2 - \hat{d}_t^2)$, where $\hat{d}_t^2$ and $\tilde{d}_t^2$ are given in (4). Provided that the fourth moment of $\epsilon_{1,1}$ exists, I show in the Appendix that the variance of $\hat{s}_t^2$ is given by

$$\mathrm{Var}(\hat{s}_t^2) = \frac{9s^4}{2n(n-1)} + \frac{2s^2\sigma^2(2n^2+3n+1)}{n^2(n-1)} + \frac{2(2n\sigma^4+\sigma^4+2\kappa)(2n^3+7n^2+7n+2)}{15n^3(n-1)} \tag{10}$$

where $\kappa = \mathrm{E}(\epsilon_{1,1}^4) - 3\sigma^4$ is the excess kurtosis of $\epsilon_{1,1}$. Equation (10) implies that while $\hat{s}^2$ is an unbiased estimator of s^2, it is not consistent as the number of intraday transactions increases since $\mathrm{Var}(\hat{s}_t^2) = \frac{8}{15}\sigma^4 + O(n^{-1})$ as $n \to \infty$. This is because we are averaging random walks in levels (prices) as opposed to first differences (returns), which cannot be constructed due to missing time stamps.

To derive a consistent estimator of s based on $\hat{s}_t^2$, we need to average $\hat{s}_t^2$ over an increasing number of days before truncating as zero and taking the square root as in (7). The resulting estimator, which I denote by $ES_T^{(1)}$, is thus given by:

$$ES_T^{(1)} = \sqrt{\max\left\{\frac{1}{T}\sum_{t=1}^{T} 2(3\tilde{d}_t^2 - \hat{d}_t^2), 0\right\}}. \tag{11}$$

Given the non-linear nature of the estimator, $\mathrm{E}(ES_T^{(1)})$ and $\mathrm{Var}(ES_T^{(1)})$ are not available in closed form. I employ a Taylor series expansion of $ES_T^{(1)}$ around s, $s > 0$, together with (10) to establish the leading terms (as $T \to \infty$). The leading term of the bias reads

$$\lim_{T\to\infty} \mathrm{E}[T(ES_T^{(1)} - s)] = -\frac{1}{8s^3}\mathrm{Var}(\hat{s}^2) = -\frac{1}{15}\frac{\sigma^4}{s^3} - \left(\frac{1}{3}\frac{\sigma^4}{s^3} + \frac{1}{15}\frac{\kappa}{s^3} + \frac{1}{2}\frac{\sigma^2}{s}\right)\frac{1}{n} + O\left(\frac{1}{n^2}\right). \tag{12}$$

implying that $ES_T^{(1)}$ tends to underestimate the true effective spread. The limiting variance reads

$$\lim_{T\to\infty} \mathrm{Var}[\sqrt{T}(ES_T^{(1)} - s)] = \frac{1}{4s^2}\mathrm{Var}(\hat{s}^2) = \frac{2}{15}\frac{\sigma^4}{s^2} + \left(\frac{2}{3}\frac{\sigma^4}{s^2} + \frac{2}{15}\frac{\kappa}{s^2} + \sigma^2\right)\frac{1}{n} + O\left(\frac{1}{n^2}\right). \tag{13}$$

As expected, the absolute bias and variance increase with the noise-to-signal ratio σ/s, i.e. it is more difficult to estimate the effective spread when it is small relative to the volatility of the efficient price. The absolute bias and variance of $ES_T^{(1)}$ also increase with excess kurtosis, but this only matters when the number of transactions is small; the contribution of κ vanishes as $n \to \infty$. The second term in the expansion also shows that for sufficiently large n the absolute bias and variance of $ES_T^{(1)}$ decrease with the number of transactions as the coefficients on the n^{-1} terms in (12) and (13) are always positive.

It follows from the assumptions stated in Section 3.1 and standard limit theorems that as $T \to \infty$, $ES_T^{(1)} \to_p s$ and if $s > 0$ and $\kappa < \infty$, $\sqrt{T}(ES_T^{(1)} - s) \to_d \mathrm{N}(0, \omega^2)$, where $\omega^2 = \frac{1}{4s^2}\mathrm{Var}(\hat{s}^2)$. The limiting variance of $ES_T^{(1)}$ has a particularly simple form if we consider the asymptotics where both $T, n \to \infty$, which may be appropriate in situations where the number of daily transactions is large. Then from (13) we obtain $\sqrt{T}(ES_T^{(1)} - s) \to_d \mathrm{N}(0, \frac{2\sigma^4}{15s^2})$. Feasible inference can be obtained by replacing the unknown s and σ^2 in the limiting variance by their sample counterparts, $ES_T^{(1)}$ and $\hat{\sigma}_T^2$, respectively, where

$$\hat{\sigma}_T^2 = \max\left\{\frac{1}{T}\sum_{t=1}^{T} 3(\tilde{d}_t^2 - \hat{d}_t^2), 0\right\}. \tag{14}$$

is a consistent estimator of σ^2.

To assess how accurately the asymptotic properties of $ES_T^{(1)}$ approximate their

finite-sample counterparts, I run a Monte Carlo experiment. I set the daily integrated volatility of the efficient price (σ) to 0.35%, which is approximately equal to the daily volatility of the 10Y Treasury futures price, and let the efficient price innovations follow normal distribution. I vary the true effective spread (s) between 5 and 50 basis points, the number of daily trasactions (n) between 5 and 250 and the number of days (T) in the sample between 25 and 250. I run 1 million Monte Carlo replications.

Table 1 reports the results. I find that the bias of the estimator can be either positive or negative in small samples depending on the true effective spread. But as predicted by theory (see equation (12)), the bias does become negative for sufficiently large T before eventually converging to zero as $T \to \infty$. The asymptotic variance approximates the finite-sample variance fairly well for relatively large values of s, but in small samples the approximation can substantially overestimate or underestimate the true variance when s is small relative to σ. Finally, the incidence of zero estimates increases as the true s decreases relative to σ, but it does converge to zero as $T \to 0$ as predicted by theory.

3.3 Comparison with Roll's measure

Should time stamps be available, one would typically use Roll's (1984) estimator, which is equal to minus 4 times the sample first-order autocovariance of intraday returns:

$$\hat{\gamma}_t^2 = -\frac{4}{n-2}\sum_{i=3}^{n}(p_{i,t} - p_{i-1,t})(p_{i-1,t} - p_{i-2,t}), \tag{15}$$

It is easy to show that in my set-up the variance of this estimator reads

$$\mathrm{Var}(\hat{\gamma}_t^2) = \left(\frac{\sigma^4}{n^2} + \frac{s^2\sigma^2}{n} + \frac{3s^4}{16} + \frac{2s^4(n-3)}{16(n-2)}\right)\frac{1}{n-2}. \tag{16}$$

Clearly, $\mathrm{Var}(\hat{\gamma}_t^2) = \frac{5s^4}{16n} + O(n^{-2})$, and $\hat{\gamma}_t^2$ is a consistent estimator of γ^2 as $n \to \infty$. When the number of transactions is large, Roll's estimator will clearly dominate the baseline estimator $ES_T^{(1)}$. But for small n, the relative variance of these estimators depends on the parameters s, σ and κ. For illustration, we plot in Figure 1 the standard deviations of $\hat{s}_t^2$ and $\hat{\gamma}_t^2$ as a function of n for $\sigma = 35$ bps, $\kappa = 0$ and $s = 50$ (right panel) or $s = 10$ bps (left panel). We find that when the spread s is relatively large, there exist a fairly wide range of ns for which $\hat{s}_t^2$ actually dominates Roll's $\hat{\gamma}_t^2$

in terms of standard deviation.

3.4 Range-based estimators

The benchmark estimator is simple to compute, but it requires observing the benchmark price m. When these prices or mid-quotes are not available, $\hat{d}_t$ cannot be calculated and we need an alternative moment condition to use together with (6). Inspired by Corwin and Schulz (2012), I use the daily range:

$$\hat{r}_t^2 = (\max_j p_{j,t} - \min_j p_{j,t})^2. \tag{17}$$

The range-based estimator has a long tradition in financial econometrics, dating back to Parkinson (1980), and has been widely used for estimating volatility from intraday data (Christensen and Podolskij, 2007, and the refereneces therein). It is clear that in expectation $\hat{r}_t^2$ depends on both s and σ, as do $\tilde{d}_t^2$ and $\tilde{d}_t^2$, but a complication arises in that $\mathrm{E}(\hat{r}_t^2)$ is not available in closed form and so a simple method-of-moments estimation does not apply. Under additional assumptions on the distribution of the innovations of the efficient price, ϵ, the expectation can be nontheless approximated by simulation and the simulated method of moments (SMM) employed to consistently estimate s.

I proceed as follows. Let $\boldsymbol{\theta} = (s, \sigma^2)'$ and let $\boldsymbol{p}_s^* = (p_{1s}^*, p_{2s}^*, ..., p_{ns}^*)'$ denote a random draw from model (8) given $\boldsymbol{\theta}$. Taking S independent draws, I approximate the expectation of $\hat{r}^2$ by

$$m_S(\boldsymbol{\theta}, n) = \frac{1}{S}\sum_{s=1}^{S}(\max_j p_{js}^* - \min_j p_{js}^*)^2. \tag{18}$$

The SMM estimator is then obtained by

$$\hat{\boldsymbol{\theta}}_T = \underset{\theta \in \mathbb{R}_{++}}{\arg\min}\, \boldsymbol{g}_T' \boldsymbol{g}_T \tag{19}$$

where $\boldsymbol{g}_T = \frac{1}{T}\sum_{t=1}^{T} \boldsymbol{g}_t$, $\boldsymbol{g}_t = (g_{1t}, g_{2t})'$, $g_{1t}(\boldsymbol{\theta}, n) = \tilde{d}_t^2 - \mathrm{E}(\tilde{d}_t^2)$ and $g_{2t}(\boldsymbol{\theta}, n) = \hat{r}_t^2 - m_S(\boldsymbol{\theta}, n)$. The objective function $\boldsymbol{g}_T' \boldsymbol{g}_T$ must be minimized numerically under the restrictions that s and σ are non-negative. The range-based estimator of s, which I denote by $ES_T^{(2)}$, is then given by $ES_T^{(2)} = \hat{\theta}_{1,T}$. It follows that if $T/S \to 0$ as

$T \to \infty$, $ES_T^{(2)} \xrightarrow{p} s$ and the SMM estimator is asymptotically equivalent to GMM (Gourieroux and Monfort, 1996, Chapter 2).

In Table 2, I report the results of a Monte Carlo simulation of $ES_T^{(2)}$ where I employ the same simulation design as in Section 3.2. I find that the estimator exhibits a bias that can be either positive or negative depending on n, T and s. The standard deviation and the incidence of zero estimates of s declines with T as expected. To compare the performance of $ES_T^{(2)}$ with the baseline estimator $ES_T^{(1)}$, I report in Table 5 their root mean-square error (RMSE). Interestingly, I find that the two estimators can perform quite differently. $ES_T^{(1)}$ does well when s is large and T is small; take for example the case $s = 50$, $n = 100$ and $T = 25$, where the RMSE of $ES_T^{(1)}$ is around three times smaller than that of $ES_T^{(2)}$. On the other hand, $ES_T^{(2)}$ works relatively well when s is small and T is large; as an example, consider the case $s = 5$, $n = 100$ and $T = 250$, where the RMSE of $ES_T^{(1)}$ is almost four times larger than that of $ES_T^{(2)}$.

My final estimator follows naturally from the previous two. If the benchmark prices are observable, it is clearly desirable to use all three moment conditions at the same time. This not only improves efficiency, it also provides a way of testing the validity of the model assumptions. Formally, define $g_{3t}(\boldsymbol{\theta}, n) = \hat{d}_t^2 - \mathrm{E}(\hat{d}_t^2)$ and $\boldsymbol{g}_t = (g_{1t}, g_{2t}, g_{3t})'$, where g_{1t} and g_{2t} are given above. The the overidentified SMM estimator of $\boldsymbol{\theta}$ is given by

$$\tilde{\boldsymbol{\theta}}_T = \underset{\theta \in \mathbb{R}_{++}}{\arg\min}\, \boldsymbol{g}_T' \boldsymbol{W}_T \boldsymbol{g}_T \tag{20}$$

for some positive definite matrix $\boldsymbol{W}_T$. I follow the standard two-stage approach, whereby I first use $\boldsymbol{W}_T = \boldsymbol{I}$ to obtain a preliminary estimate $\hat{\boldsymbol{\theta}}_T$ and then use the optimal $\widehat{\boldsymbol{W}_T}$ (sample variance of $\boldsymbol{g}_t$ evaluated at $\hat{\boldsymbol{\theta}}_T$) in the second stage to obtain $\tilde{\boldsymbol{\theta}}_T$. My third estimator of s is then given by $ES_T^{(3)} = \tilde{\theta}_{1,T}$. Again, if $T/S \to 0$ as $T \to \infty$, $ES_T^{(3)} \xrightarrow{p} s$. Additionally, if the assumptions of the model are correct, the statistic

$$J_T = T \tilde{\boldsymbol{g}}_T' \widetilde{\boldsymbol{W}}_T \tilde{\boldsymbol{g}}_T, \tag{21}$$

where the tilde denotes quantities evaluated at $\tilde{\boldsymbol{\theta}}_T$, satisfies $J_T \xrightarrow{d} \chi^2(1)$ as $T \to \infty$. This is, of course, the standard test for an overidentifying restriction.

I again run a Monte Carlo simulation to investigate the behavior of $ES_T^{(3)}$ in finite samples. The results are reported in Table 3. I find that the estimator exhibits

small bias in small samples, generally smaller standard deviation than $ES_T^{(3)}$ and lower incidence of zero estimates of s than either $ES_T^{(1)}$ or $ES_T^{(2)}$. To comparte the precision of the three estimators, I add to Table 5 the RMSE of $ES_T^{(3)}$. The overidentified estimator is typically the most precise of the three, except when T is small. This is not surprising given the differences in the performance of $ES_T^{(1)}$ and $ES_T^{(2)}$: using all three moment conditions simultaneously, together with the optimal weighting matrix, efficiently combines the two estimators. Thus, the overidentified estimator $ES_T^{(3)}$ should be used in practice whenver possible (i.e. when $m_{0,t}$ is observable). Finally, in Table 4, I report the simulated size of the J_T test in (21) for the 10%, 5% and 1% nominal size. I find some size distortions for small values of T, but when $T \geq 100$, the simulated size is generally quite close to its nominal counterpart.

4 Extentions

4.1 Time-varying volatility

I first provide conditions under which the baseline estimator $ES_T^{(1)}$ is robust to time-varying volatility of the efficient price innovations ϵ in (9). It turns out that these conditions are fairly straightforward. By examining the condition expectation of $\hat{s}_T^2$ given the sample path of volatility of the efficient price, $\mathrm{E}(\hat{s}_T^2|\{\sigma\})$, which I provide in the appedix in closed form, I find that if $\mathrm{E}(\sigma_{i,t}^2) = const$, the Law of Iterated Expectations implies that $\mathrm{E}(\hat{s}_T^2) = s^2$. Thus, if the variance is stochastic with constant and finite mean, $\hat{s}_T^2$ remains an unbiased estimator of s^2. If, however, the variance is non-stationary or deterministic (and time-varying), this is no longer true in general, and the estimator $\hat{s}_T^2$ becomes biased and inconsistent for s^2. The bias is a function of the entire path of σ, but it cannot be evaluated analytically in the absence of time-stamps.

Now to establish consistency of $ES_T^{(1)}$ under stochastic volatility, we also need to show that the variance of $\hat{s}_T^2$ converges to zero as $T \to \infty$. Sufficient conditions for this, which I state here without proof, are more involved and require the variance process to be fourth-order stationary and weakly dependent. These conditions are satisfied by a wide range of popular GARCH and stochastic volatility models.

To corroborate the conjecture of consistency of $ES_T^{(1)}$ under stochastic volatility, I run a Monte Carlo experiment similar to the one in the previous section. The variance

process is a standard Gaussian exponential volatility model, where $\sigma_{i,t}^2 = \exp(\alpha + z_{i,t})$ and $z_{i,t}$ is a stationary Gaussian AR(1) process with autoregressive parameter ρ. I set $\rho = 0.75$ and choose α such that $\mathrm{E}(\sigma_{i,t}^2) = 1225$bps for comparison with the constant-variance simulation (where I set $\sigma^2 = 1225$bps). The simulation results, reported in Table 6, strongly echo those for the baseline estimator (Table 1). The variance of the estimator declines with T and the bias eventually approaches zero.

4.2 Time-varying effective spread

My second extension considers time-varying effective spread. The lack of transaction time stamps does not allow me to estimate time-varying intraday spreads, but I can allow the effective spread to vary over days. For any given day t, I thus assume that the effective spread within the day equals s_t. I follow the recent advances by Giraitis et al. (2014) and Giraitis et al. (forthcoming) and adopt a nonparametric approach whereby the law of motion of the paramters is left unspecified, up to a class of processes, and the parameters are estimated by local averaging. The processes I consider for s_t are stochastic and/or deterministic processes satisfying the smoothness condition:

$$\sup_{j:|j-t|<h} ||s_t - s_j||^2 = O_p(h/t) \tag{22}$$

as $t \to \infty$, $h \to \infty$, $h = o(t)$. Example of processes that belong into this class include

$$s_t = t^{-1}x_t, \quad s_t = \frac{x_t}{\max_{j \le T} |x_j|}, \quad s_t = g\left(\frac{t}{n}\right), \tag{23}$$

where $\{x\}$ is a unit root process with stationary increments and g is a smooth deterministic function on the unit interval. These processes are bounded in probability and are smoother than random walks. Figure 2 provides for illustration some sample trajectories of these processes.

To estimate s_t, I take some weights $w_{j,t} = \tilde{w}_{j,t} / \sum_j \tilde{w}_{j,t}$ where $w_{j,t} = K((j-t)/H)$ for some kernel function K and bandwidth paramter H. I then define the kernel estimator of the effective spread at time t by

$$ES_{t,T} = \sqrt{\max\left\{2\sum_{j=1}^{T} w_{j,t}(3\tilde{d}_j^2 - \hat{d}_j^2), 0\right\}}. \tag{24}$$

The kernel can be uniform, leading to simple rolling estimation with window of size H, or with unbounded support, such as the Gaussian kernel. Precised conditions on the kernel function are stated in the appendix. The key to achieving consistency of the estimator is the choice of H relative to T as the following proposition shows.

Proposition 1 *For any* $t = [\tau T], 0 < \tau < 1$, *if* $H \to \infty$, $(H \log^{1/2} H)/T \to 0$ *as* $T \to \infty$, $ES_{t,T} \xrightarrow{p} s_t$.

The proposition shows that the bandwith parameter has to grow with T, but not as fast as T. When choosing H one faces the familiar trade-off between bias and variance: smaller (larger) H produces less (more) biased by more (less) volatile estimates. There is currently no data-driven method for choosing H, but taking $H = \sqrt{H}$ seems to work well in existing applications (Giraitis et al., 2014, forthcoming) and also in the simulation reported below.

I run a Monte Carlo simulation with the following design. The true effective spread follows $s_t = s + \frac{3}{5}\frac{x_t}{\max_{j \leq T} |x_j|}$, where x_t is a Gaussian white noise process with unitary volatility and $s \in \{5, 10, 20, 50\}$ is as in Section 3. With this parametrization, the effective spread fluctuates between $\frac{2s}{5}$ and $\frac{8s}{5}$ and has a mean equal to s; for example, for $s = 50$bps, s_t varies between 20 and 80bps. The efficient price innovations are normally distributed with volatility of 35 bps as before. I use the Gaussian kernel and set the bandwith parameter according to $H = \sqrt{T}$ following Giraitis et al. (2014, forthcoming).

For illustration, I first plot in Figure 2 three randomly selected sample realizations of $\{s_t\}$ together with the kernel estimate $\{ES_{t,T}\}$ for T ranging between 1,000 and 5,000. The Figure shows that the kernel estimator works well. I then report in Table 7 the average difference between the true and estimated spreads and the associated root mean-square error (RMSE). Consistent with Proposition 1 the bias and RMSE decline as $T \to \infty$. The rate of convergence may appear slow, but recall that with $H = \sqrt{T}$ the effective sample size equals $\sqrt{T}$ rather than T. The kernel estimator therefore requires a lot of data and should always be applied to the full sample of observations.

5 Observable trade direction

Some databases contain information about trade direction, i.e. whether a trade is buyer or seller initiated. A prominent example is the TRACE database of US corporate bonds. Observing q in model (8) makes the estimation of transaction costs significantly easier, despite the absence of time stamps. Schultz (2001) proposes to use linear regression to estimate the effective spread in this case. In a simple form, the regression reads

$$p_{i,t} - m_{0,t} = \beta_t q_{i,t} + u^0_{i,t}. \tag{25}$$

Given model (8), the regression innovation u^0_i is given by $u^0_{i,t} = m_{i,t} - m_{0,t} + u_{i,t}$, and $u_{i,t}$ is a measurement error unrelated to m or q. It is easy to show that the OLS estimator of β_t in a regression of $(p_{i,t} - m_{0,t})$ on $q_{i,t}$, $i = 1, ..., n$, satisfies under my assumptions $\mathrm{E}(\hat{\beta}_t) = \frac{s}{2}$ and $\mathrm{Var}(\hat{\beta}_t) = \frac{\sigma^2}{2n} + \frac{\sigma_u^2}{n}$. Replacing $m_{i,t}$ by $m_{0,t}$ therefore increases the variance of the OLS estimator but it does not render it biased or inconsistent as it did for $\hat{d}_t^2$ in Section 3. The OLS estimator will converge in probability to $2s$ as $n \to \infty$. Note also that the OLS estimator above is equal to the difference between the average purchase price and the average sales price. The latter estimator was proposed by Hong and Warga (2000) and it was subsequently used by a number of papers measuring transaction costs in the US bond markets.

In a recent paper, Biswas, Nikolova and Stahel (2015) apply the regression-based method in a situation where $q_{i,t}$ is not directly observable. They suggest to approximate $q_{i,t}$ by $q^0_{i,t} = 1\{p_{i,t} > m_{0,t}\} - 1\{p_{i,t} \leq m_{0,t}\}$ and run the regression

$$p_{i,t} - m_{0,t} = \beta_t q^0_{i,t} + u^0_{i,t}. \tag{26}$$

Similarly to the Jankowitsch et al. (2011) metric, replacing $m_{i,t}$ by $m_{0,t}$ in the definition of $q^0_{i,t}$ renders the OLS estimator of β_t in (26) biased and inconsistent for $2s$. Given the non-linearity of the OLS estimator in m and q, the bias is a complicated function of the paramteres and distributional assumptions of the model in equations (8) and (9) and I do not present it here. Instead, I explore the behavior of the estimator in a Monte Carlo simulation with the same design as in Section 3. The results are reported in Table 8. I find that the OLS estimator in regression (26) is inconsistent and significantly upward biased. The bias is particularly severe for small values of s, but this is not surprising since the main driver of the bias is the volatility of the

efficient price, just like in the case of the $\hat{d}_t^2$ and $\tilde{d}_t^2$ metrics, see equations (5) and (6), respectively.

6 Empirical illustration

Having explored the behavior of the various estimators of s in theory and simulations, I now turn to an empirical illustration. The purpose of this exercise is to compare the performance of the various measures with the "first best", that is with the effective spread one would calculate if m, q and time-stamps were observable. Thus, I employ time-stamped trade and quote data for selected NYSE-listed stocks from the Trade and Quote (TAQ) database maintained by WRDS. This approach to assessing the performance of my measures is similar to Goyenko, Holden and Trzcinka (2009) and Corwin and Schultz (2012), who compare liquidity measures constructed from low-frequency data with their high frequency data-based counterparts.

In progress...

References

Ait-Sahalia, Y. and J. Jacod, 2014, *High-Frequency Financial Econometrics*, Princeton University Press.

Benos, E., Wetherilt, A. and F. Zikes, 2013, The structure and dynamics of the UK credit default swap market, Financial Stability Paper No. 25, Bank of England.

Benos, E., Payne, R. and M. Vasios, 2015, Centralized Trading, Transparency and Interest Rate Swap Market Liquidity: Evidence from the Implementation of the Dodd-Frank Act, Bank of England.

Benos, E. and F. Zikes, 2015, Liquidity and dealer activity in the UK gilt market during the financial crisis, Bank of England.

Biswas, G., Nikolova, S. and C. W. Stahel, 2014, The transaction costs of trading corporate credit, Securities and Exchange Commission.

Chen, K., Fleming, M., Jackson, J., Li, A. and A. Sarkar, 2011, An Analysis of CDS Transactions: Implications for Public Reporting, Staff Report No. 517, Federal Reserve Bank of New York.

Chen, K., Fleming, M., Jackson, J., Li, A. and A. Sarkar, 2012, An Analysis of OTC Interest Rate Derivatives Transactions: Implications for Public Reporting, Staff Report No. 557, Federal Reserve Bank of New York.

Christensen, K. and M. Podolskij, 2007, Realized range-based estimation of integrated variance, *Journal of Econometrics*, **141**, 323-349.

Corwin, S. and P. Schultz, 2012, A Simple Way to Estimate Bid-Ask Spreads from Daily High and Low Prices, *Journal of Finance*, **67**(2), 719–759.

Financial Conduct Authority, 2015, Transparency in the UK Bond Markets: An Overview, Occasional Paper No. 6.

Doornik, J.A., 2007, *Object-Oriented Matrix Programming Using Ox*, 3rd ed. London: Timberlake Consultants Press and Oxford: www.doornik.com.

Du, W., Gadgil, S., Gordy, M.B. and C. Vega, 2015, Counterparty risk and counterparty choice in the credit default swap market, Federal Reserve Board.

Foucault, T., Pagano, M. and A. Roell, 2013, *Market Liquidity: Theory, Evidence and Policy*, Oxford University Press.

Giraitis, L., Kapetanios, G. and T. Yates, 2013, Inference on stochastic time-varying coefficient models, *Journal of Econometrics*, **179**, 46–65.

Giraitis, L., Kapetanios, G., Wetherilt, A. and F. Zikes, forthcoming, Estimating the dynamics and persistence of financial networks, with an application to the sterling money market, *Journal of Applied Econometrics*.

Gourieroux, C. and A. Monfort, 1996, *Simulation-Based Econometric Methods*, Oxford University Press.

Goyenko, R.Y., Holden, C.W. and C.A. Trzcinka, 2009, Do liquidity measures measure liquidity?, *Journal of Financial Economics*, **92**, 153–181.

Harris, L., 2015, Transaction costs, trade throughs, and riskless principle trading in corporate bond markets, USC Marshall School of Business.

Harris, L. E. and M. S. Piwowar, 2006, Secondary trading costs in the municipal bond market, *Journal of Finance*, **61**(3), 1361–1397.

Hong, G. and A. Warga, 2000, An empirical study of bond market transactions, *Financial Analyst Journal*, **56**, 32-46.

Jankowitsch, R., Nashikkar, A. and M.G. Subrahmanyam, 2011, Price Dispersion in OTC Markets: A New Measure of Liquidity, *Journal of Banking and Finance*, **35**, 343–357.

Parkinson, M., 1980, The extreme value method for estimating the variance of the rate of return, *Journal of Business*, **53**, 61-65.

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

Schultz, P., 2001, Corporate bond trading costs: A peak behind the curtain, *Journal of Finance*, **56**(2), 677–698.

Siriwardane, E.N., 2015, Concentrated Capital Losses and the Pricing of Corporate Credit Risk, Working Paper 16-007, Harvard Business School.

A Figures and Tables

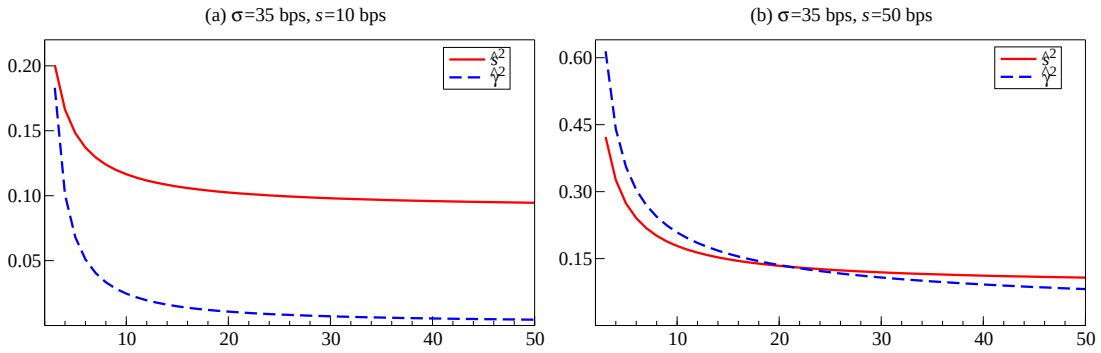

Figure 1: Exact standard deviation of $\hat{s}^2$ and $\hat{\gamma}^2$ as a function of n. The parameter values are $\sigma = 35$ bps, $\kappa = 0$ and $s = 10$ bps (left panel) and $s = 50$ bps (right panel).

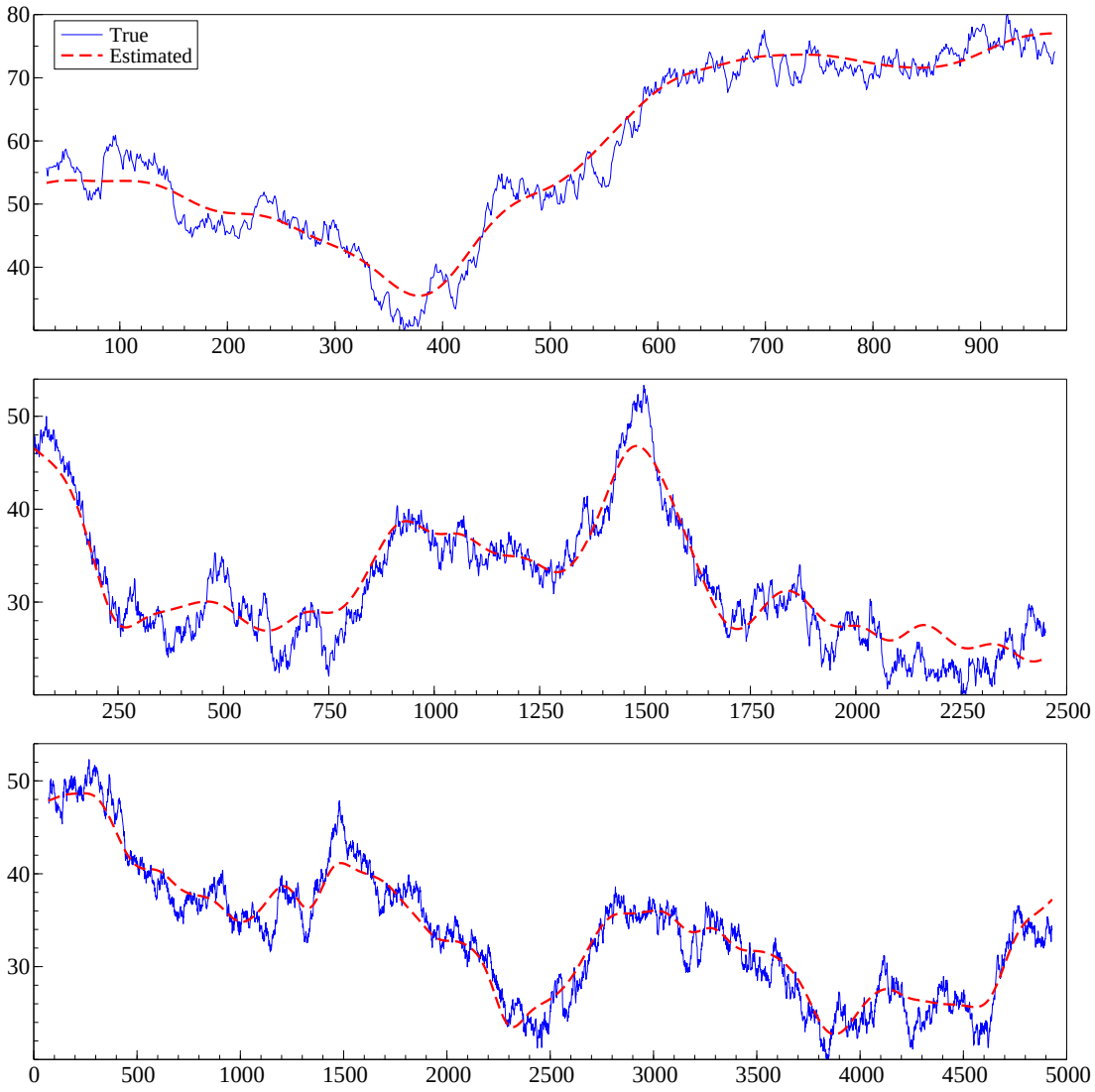

Figure 2: Sample realization of s_t and $ES_{t,T}$ for $T = 1,000$ (top panel), $T = 2,500$ (middle panel) and $T = 5,000$ (bottom panel). The true effective spread follows $s_t = s + \frac{3}{5}\frac{x_t}{\max_{j\leq T}|x_j|}$, where x_t is a Gaussian white noise process with unitary volatility and $s = 50$bps. The efficient price innovations are normally distributed with volatility of 35 bps. Gaussian kernel is used for estimation and the bandwith parameter is set according to $H = \sqrt{T}$.

	T					T				
n	25	50	100	250	1000	25	50	100	250	1000
	$s = 50$ bps					$s = 20$ bps				
5	49.67	49.85	49.92	49.97	49.99	18.38	18.92	19.46	19.82	19.96
	[1.033]	[1.014]	[1.008]	[1.002]	[1.002]	[1.101]	[1.173]	[1.139]	[1.043]	[1.008]
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(11.29)	(4.65)	(0.94)	(0.01)	(0.00)
10	49.87	49.93	49.97	49.99	50.00	18.83	19.36	19.71	19.90	19.98
	[1.017]	[1.008]	[1.004]	[1.000]	[1.001]	[1.170]	[1.164]	[1.082]	[1.025]	[1.005]
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(6.02)	(1.56)	(0.13)	(0.00)	(0.00)
25	49.94	49.97	49.99	49.99	50.00	19.14	19.58	19.81	19.93	19.98
	[1.009]	[1.005]	[1.003]	[1.002]	[0.999]	[1.184]	[1.130]	[1.053]	[1.019]	[1.005]
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(3.36)	(0.56)	(0.02)	(0.00)	(0.00)
50	49.96	49.98	49.99	49.99	50.00	19.23	19.64	19.84	19.94	19.99
	[1.009]	[1.004]	[1.003]	[1.002]	[1.000]	[1.184]	[1.114]	[1.046]	[1.015]	[1.003]
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(2.63)	(0.35)	(0.01)	(0.00)	(0.00)
100	49.96	49.98	49.99	50.00	50.00	19.29	19.67	19.85	19.95	19.99
	[1.008]	[1.004]	[1.003]	[1.000]	[1.000]	[1.183]	[1.108]	[1.043]	[1.014]	[1.004]
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(2.30)	(0.28)	(0.01)	(0.00)	(0.00)
250	49.97	49.98	49.99	50.00	50.00	19.31	19.69	19.86	19.95	19.99
	[1.008]	[1.003]	[1.001]	[1.001]	[1.000]	[1.181]	[1.102]	[1.039]	[1.014]	[1.003]
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(2.11)	(0.23)	(0.00)	(0.00)	(0.00)
	$s = 10$ bps					$s = 5$ bps				
5	9.938	9.303	9.018	9.077	9.644	7.757	6.727	5.929	5.154	4.541
	[0.592]	[0.720]	[0.867]	[1.064]	[1.153]	[0.284]	[0.344]	[0.416]	[0.535]	[0.784]
	(35.21)	(30.81)	(24.51)	(14.28)	(1.72)	(44.83)	(43.93)	(42.22)	(38.67)	(28.99)
10	9.534	9.118	9.016	9.245	9.790	7.104	6.194	5.513	4.893	4.485
	[0.674]	[0.818]	[0.971]	[1.141]	[1.105]	[0.320]	[0.388]	[0.470]	[0.607]	[0.882]
	(31.77)	(26.36)	(19.20)	(8.93)	(0.39)	(43.62)	(42.55)	(40.44)	(35.90)	(24.28)
25	9.334	9.066	9.066	9.387	9.854	6.717	5.900	5.295	4.763	4.481
	[0.732]	[0.883]	[1.036]	[1.169]	[1.073]	[0.345]	[0.419]	[0.509]	[0.658]	[0.947]
	(29.27)	(23.16)	(15.65)	(5.92)	(0.10)	(42.89)	(41.51)	(39.07)	(33.88)	(21.10)
50	9.296	9.040	9.092	9.438	9.873	6.602	5.802	5.222	4.718	4.477
	[0.753]	[0.907]	[1.058]	[1.172]	[1.063]	[0.355]	[0.431]	[0.523]	[0.675]	[0.969]
	(28.21)	(22.07)	(14.44)	(4.98)	(0.06)	(42.61)	(41.24)	(38.58)	(33.17)	(20.01)
100	9.258	9.044	9.102	9.456	9.882	6.548	5.775	5.193	4.694	4.478
	[0.765]	[0.919]	[1.069]	[1.174]	[1.057]	[0.360]	[0.437]	[0.530]	[0.685]	[0.980]
	(27.83)	(21.51)	(13.88)	(4.55)	(0.04)	(42.39)	(40.88)	(38.27)	(32.89)	(19.47)
250	9.255	9.039	9.123	9.473	9.883	6.501	5.739	5.159	4.690	4.479
	[0.771]	[0.926]	[1.073]	[1.175]	[1.056]	[0.362]	[0.440]	[0.535]	[0.691]	[0.986]
	(27.48)	(21.13)	(13.39)	(4.28)	(0.03)	(42.38)	(40.83)	(38.26)	(32.54)	(19.10)

Table 1: Monte Carlo simulation results for the baseline effective spread estimator, $ES_T^{(1)}$. For each T (number of days in the simulated sample) and n (number of transactions on each day in the simulated sample), the table reports the average simulated $ES_T^{(1)}$, the ratio of the simulated standard deviation of $ES_T^{(1)}$ and its asymptotic standard deviation (ω^2) in brackets and the percentage of zero $ES_T^{(1)}$ in the simulation in parentheses. The efficient price innovations are normally distributed and the daily integrated volatility of the efficient price is set to 0.35% or 35 bps. The results are based on 1 million Monte Carlo replications.

	T				T			
n	25	50	100	250	25	50	100	250
	$s = 50$ bps				$s = 20$ bps			
10	49.67	49.89	49.95	49.99	21.33	21.24	21.18	21.11
	[3.200]	[2.127]	[1.499]	[0.941]	[5.530]	[4.625]	[3.766]	[2.834]
	(0.01)	(0.00)	(0.00)	(0.00)	(1.92)	(0.90)	(0.18)	(0.00)
25	49.65	49.94	49.91	50.07	20.58	20.41	19.95	20.54
	[4.072]	[1.889]	[1.247]	[0.564]	[4.067]	[3.346]	[2.665]	[2.288]
	(0.50)	(0.08)	(0.03)	(0.00)	(0.09)	(0.00)	(0.00)	(0.00)
50	48.79	49.65	49.92	49.98	19.67	20.02	19.70	19.49
	[7.542]	[4.562]	[2.113]	[0.436]	[3.973]	[3.404]	[2.781]	[2.192]
	(2.26)	(0.80)	(0.16)	(0.00)	(0.09)	(0.00)	(0.00)	(0.00)
100	48.81	49.62	49.90	50.00	19.20	19.15	19.06	19.35
	[7.456]	[4.153]	[1.752]	[0.355]	[4.121]	[3.519]	[2.984]	[2.437]
	(2.23)	(0.67)	(0.11)	(0.00)	(0.00)	(0.04)	(0.00)	(0.00)
	$s = 10$ bps				$s = 5$ bps			
10	9.33	9.45	9.66	9.90	6.66	5.96	5.40	4.78
	[6.983]	[6.324]	[5.498]	[3.976]	[6.541]	[5.862]	[5.082]	[3.872]
	(27.21)	(20.47)	(12.79)	(3.91)	(40.73)	(38.77)	(35.44)	(28.51)
25	9.92	10.04	10.62	9.55	5.41	4.86	5.02	3.90
	[4.942]	[4.016]	[3.036]	[1.929]	[4.724]	[3.884]	[3.084]	[2.267]
	(8.88)	(3.28)	(0.41)	(0.01)	(31.10)	(26.13)	(15.73)	(14.76)
50	10.43	10.11	10.46	10.43	5.50	4.64	4.84	5.00
	[4.019]	[3.329]	[2.630]	[1.883]	[4.054]	[3.128]	[2.297]	[1.425]
	(2.64)	(0.69)	(0.02)	(0.00)	(20.38)	(17.59)	(7.60)	(1.25)
100	10.48	10.76	10.83	10.52	5.55	5.29	5.16	5.00
	[3.639]	[3.169]	[2.710]	[2.185]	[3.652]	[2.751]	[1.906]	[1.089]
	(0.97)	(0.08)	(0.00)	(0.00)	(13.78)	(7.22)	(2.30)	(0.14)

Table 2: Monte Carlo simulation results for the exactly-identified range-based effective spread estimator, $ES_T^{(2)}$. For each T (number of days in the simulated sample) and n (number of transactions on each day in the simulated sample), the table reports the average simulated $ES_T^{(2)}$, the standard deviation of $ES_T^{(2)}$ and the percentage of the realizations of $ES_T^{(2)}$ below 0.1bps in the simulation in parentheses. The efficient price innovations are normally distributed and the daily integrated volatility of the efficient price is set to 0.35% or 35 bps. The results are based on 10,000 Monte Carlo replications.

	T				T			
n	25	50	100	250	25	50	100	250
	$s = 50$ bps				$s = 20$ bps			
10	50.03	50.04	50.02	50.02	19.33	19.62	19.81	19.91
	[2.916]	[2.002]	[1.425]	[0.896]	[6.146]	[4.426]	[3.118]	[1.976]
	(0.00)	(0.00)	(0.00)	(0.00)	(2.16)	(0.37)	(0.01)	(0.00)
25	49.97	50.01	49.95	50.06	19.73	19.76	19.78	19.89
	[1.750]	[1.220]	[0.874]	[0.548]	[4.445]	[3.410]	[2.460]	[1.810]
	(0.00)	(0.00)	(0.00)	(0.00)	(0.06)	(0.01)	(0.00)	(0.00)
50	49.95	50.04	50.01	49.99	19.76	19.71	19.74	19.85
	[1.347]	[0.936]	[0.665]	[0.425]	[4.015]	[3.280]	[2.405]	[1.518]
	(0.00)	(0.00)	(0.00)	(0.00)	(0.02)	(0.00)	(0.00)	(0.00)
100	49.91	49.95	49.95	50.00	19.66	19.58	19.61	19.87
	[1.101]	[0.795]	[0.557]	[0.345]	[3.896]	[3.094]	[2.310]	[1.435]
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
	$s = 10$ bps				$s = 5$ bps			
10	10.58	10.10	9.96	9.96	7.17	6.01	5.30	4.81
	[6.673]	[5.370]	[4.049]	[2.467]	[6.419]	[5.214]	[4.192]	[3.185]
	(15.76)	(10.14)	(4.76)	(0.56)	(31.26)	(30.05)	(26.77)	(19.52)
25	11.00	10.61	10.62	9.70	6.24	5.30	5.20	4.21
	[5.104]	[3.811]	[2.556]	[1.509]	[5.155]	[3.883]	[2.823]	[2.042]
	(4.47)	(1.42)	(0.06)	(0.00)	(20.78)	(20.04)	(11.49)	(9.61)
50	11.55	10.71	10.57	10.28	6.34	5.11	5.05	5.08
	[4.432]	[3.466]	[2.425]	[1.363]	[4.618]	[3.311]	[2.128]	[1.259]
	(1.22)	(0.19)	(0.00)	(0.00)	(14.71)	(12.87)	(4.92)	(0.76)
100	11.62	11.20	10.75	10.26	6.48	*	5.28	5.05
	[4.305]	[3.309]	[2.393]	[1.472]	[4.464]	*	[1.804]	[0.994]
	(0.34)	(0.03)	(0.00)	(0.00)	(9.68)	*	(1.51)	(0.05)

Table 3: Monte Carlo simulation results for the over-identified range-based effective spread estimator, $ES_T^{(3)}$. For each T (number of days in the simulated sample) and n (number of transactions on each day in the simulated sample), the table reports the average simulated $ES_T^{(3)}$, the standard deviation of $ES_T^{(3)}$ and the percentage of the realizations of $ES_T^{(3)}$ below 0.1bps in the simulation in parentheses. The efficient price innovations are normally distributed and the daily integrated volatility of the efficient price is set to 0.35% or 35 bps. The results are based on 10,000 Monte Carlo replications. (* simulation in progress)

		T				T			
n	α	25	50	100	250	25	50	100	250
		$s = 50$ bps				$s = 20$ bps			
10	0.10	0.123	0.112	0.098	0.102	0.107	0.103	0.104	0.106
	0.05	0.067	0.060	0.052	0.052	0.051	0.051	0.050	0.053
	0.01	0.017	0.014	0.011	0.010	0.007	0.009	0.010	0.010
25	0.10	0.131	0.113	0.110	0.102	0.122	0.108	0.119	0.090
	0.05	0.076	0.063	0.059	0.051	0.064	0.059	0.066	0.043
	0.01	0.020	0.018	0.016	0.011	0.009	0.012	0.015	0.006
50	0.10	0.129	0.113	0.110	0.103	0.150	0.109	0.108	0.109
	0.05	0.077	0.062	0.059	0.052	0.086	0.061	0.058	0.059
	0.01	0.023	0.018	0.014	0.012	0.017	0.017	0.017	0.016
100	0.10	0.128	0.118	0.108	0.106	0.160	0.142	0.126	0.106
	0.05	0.076	0.065	0.058	0.053	0.095	0.085	0.073	0.055
	0.01	0.022	0.021	0.015	0.012	0.023	0.024	0.023	0.013
		$s = 10$ bps				$s = 5$ bps			
10	0.10	0.102	0.096	0.101	0.103	0.115	0.118	0.119	0.113
	0.05	0.045	0.044	0.046	0.050	0.047	0.053	0.057	0.053
	0.01	0.006	0.007	0.008	0.009	0.003	0.006	0.008	0.009
25	0.10	0.085	0.084	0.091	0.108	0.085	0.101	0.103	0.110
	0.05	0.037	0.036	0.041	0.052	0.029	0.039	0.049	0.054
	0.01	0.005	0.004	0.007	0.010	0.002	0.004	0.008	0.011
50	0.10	0.105	0.080	0.087	0.100	0.077	0.093	0.101	0.105
	0.05	0.051	0.037	0.041	0.050	0.029	0.038	0.051	0.050
	0.01	0.009	0.006	0.005	0.009	0.002	0.004	0.009	0.009
100	0.10	0.119	0.106	0.101	0.091	0.072	*	0.098	0.099
	0.05	0.066	0.056	0.051	0.044	0.027	*	0.048	0.048
	0.01	0.014	0.012	0.011	0.008	0.002	*	0.008	0.009

Table 4: Monte Carlo simulation results for the $\chi^2(1)$ test for an overidentifying restriction given in (21). For each T (number of days in the simulated sample) and n (number of transactions on each day in the simulated sample), the table reports the frequency with which the J_T statistics exceeds $(1-\alpha)$-percentile of the $\chi^2(1)$ distribution in the simulation. The efficient price innovations are normally distributed and the daily integrated volatility of the efficient price is set to 0.35% or 35 bps. The results are based on 10,000 Monte Carlo replications. (* simulation in progress)

		T				T			
n		25	50	100	250	25	50	100	250
		$s = 50$ bps				$s = 20$ bps			
10	$ES_T^{(1)}$	3.631	2.536	1.780	1.125	7.339	5.119	3.336	1.998
	$ES_T^{(2)}$	3.217	2.131	1.500	0.941	**5.687**	4.789	3.947	3.043
	$ES_T^{(3)}$	**2.916**	**2.002**	**1.425**	**0.896**	6.183	**4.443**	**3.124**	**1.978**
25	$ES_T^{(1)}$	2.538	1.765	1.249	0.802	6.212	4.129	2.724	1.687
	$ES_T^{(2)}$	4.086	1.890	1.251	0.569	**4.109**	**3.371**	2.665	2.350
	$ES_T^{(3)}$	**1.750**	**1.220**	**0.875**	**0.552**	4.453	3.418	**2.469**	**1.813**
50	$ES_T^{(1)}$	2.160	1.509	1.068	0.679	5.732	3.736	2.511	1.544
	$ES_T^{(2)}$	7.639	4.575	2.114	0.436	**3.987**	3.404	2.797	2.252
	$ES_T^{(3)}$	**1.348**	**0.937**	**0.665**	**0.425**	4.022	**3.293**	**2.419**	**1.525**
100	$ES_T^{(1)}$	1.994	1.402	0.983	0.622	5.588	3.672	2.425	1.486
	$ES_T^{(2)}$	7.550	4.170	1.755	0.355	4.197	3.620	3.128	2.522
	$ES_T^{(3)}$	**1.105**	**0.797**	**0.559**	**0.345**	**3.911**	**3.122**	**2.343**	**1.441**
		$s = 10$ bps				$s = 5$ bps			
10	$ES_T^{(1)}$	7.867	6.744	5.681	4.254	7.612	6.391	5.377	4.397
	$ES_T^{(2)}$	7.015	6.349	5.508	3.977	**6.749**	5.940	5.098	3.878
	$ES_T^{(3)}$	**6.698**	**5.371**	**4.049**	**2.467**	6.776	**5.310**	**4.202**	**3.191**
25	$ES_T^{(1)}$	7.353	6.269	5.279	3.790	6.961	5.898	5.039	4.141
	$ES_T^{(2)}$	**4.943**	4.016	3.098	1.981	**4.742**	**3.886**	3.084	2.520
	$ES_T^{(3)}$	5.201	**3.859**	**2.631**	**1.537**	5.302	3.894	**2.830**	**2.188**
50	$ES_T^{(1)}$	7.137	6.108	5.101	3.587	6.885	5.770	4.938	4.046
	$ES_T^{(2)}$	**4.041**	**3.331**	2.670	1.930	**4.085**	**3.149**	2.303	1.425
	$ES_T^{(3)}$	4.696	3.539	**2.490**	**1.393**	4.809	3.312	**2.129**	**1.262**
100	$ES_T^{(1)}$	7.065	6.095	5.023	3.421	6.789	5.747	4.861	3.990
	$ES_T^{(2)}$	**3.671**	**3.259**	2.833	2.246	**3.693**	2.765	1.913	1.089
	$ES_T^{(3)}$	4.599	3.522	**2.507**	**1.495**	4.703	*	**1.826**	**0.995**

Table 5: Root mean-squared error for the three estimators of effective spread. The efficient price innovations are normally distributed and the daily integrated volatility of the efficient price is set to 0.35% or 35 bps. The results are based on 10,000 Monte Carlo replications. (* simulation in progress)

	T					T				
n	25	50	100	250	1000	25	50	100	250	1000
	$s = 50$ bps					$s = 20$ bps				
5	49.61	49.81	49.91	49.96	49.99	18.42	18.71	19.20	19.70	19.93
	[6.248]	[4.278]	[2.988]	[1.879]	[0.935]	[9.895]	[7.926]	[5.833]	[3.445]	[1.600]
	(0.04)	(0.00)	(0.00)	(0.00)	(0.00)	(13.96)	(7.80)	(2.79)	(0.19)	(0.00)
10	49.83	49.92	49.96	49.98	49.99	18.72	19.14	19.56	19.84	19.96
	[4.044]	[2.805]	[1.971]	[1.241]	[0.620]	[8.164]	[6.165]	[4.240]	[2.462]	[1.187]
	(0.01)	(0.00)	(0.00)	(0.00)	(0.00)	(8.73)	(3.62)	(0.77)	(0.01)	(0.00)
25	49.91	49.95	49.97	49.99	49.99	18.98	19.42	19.73	19.90	19.97
	[2.80]	[1.96]	[1.38]	[0.87]	[0.43]	[6.86]	[4.88]	[3.22]	[1.91]	[0.93]
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(5.33)	(1.53)	(0.15)	(0.00)	(0.00)
50	49.94	49.97	49.98	49.99	49.99	19.13	19.55	19.80	19.92	19.98
	[2.356]	[1.653]	[1.163]	[0.734]	[0.367]	[6.248]	[4.296]	[2.819]	[1.705]	[0.841]
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(3.91)	(0.83)	(0.05)	(0.00)	(0.00)
100	49.95	49.97	49.98	49.99	49.99	19.22	19.61	19.83	19.93	19.98
	[2.094]	[1.473]	[1.039]	[0.656]	[0.328]	[5.821]	[3.925]	[2.587]	[1.579]	[0.781]
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(3.01)	(0.50)	(0.01)	(0.00)	(0.00)
250	49.96	49.98	49.99	49.99	49.99	19.28	19.67	19.85	19.94	19.98
	[1.928]	[1.355]	[0.956]	[0.604]	[0.302]	[5.516]	[3.660]	[2.430]	[1.494]	[0.740]
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(2.39)	(0.30)	(0.01)	(0.00)	(0.00)
	$s = 10$ bps					$s = 5$ bps				
5	10.50	9.725	9.241	9.035	9.453	8.475	7.376	6.470	5.541	4.693
	[9.433]	[8.195]	[7.064]	[5.670]	[3.446]	[8.921]	[7.681]	[6.589]	[5.392]	[3.968]
	(35.61)	(32.55)	(27.83)	(19.18)	(4.693)	(44.00)	(43.73)	(42.68)	(40.18)	(32.70)
10	9.971	9.389	9.095	9.135	9.674	7.692	6.715	5.921	5.155	4.553
	[8.355]	[7.242]	[6.198]	[4.818]	[2.592]	[7.870]	[6.778]	[5.832]	[4.780]	[3.514]
	(32.46)	(28.45)	(22.76)	(13.16)	(1.510)	(42.87)	(42.26)	(40.98)	(37.69)	(28.17)
25	9.616	9.187	9.059	9.272	9.798	7.129	6.240	5.549	4.914	4.497
	[7.655]	[6.602]	[5.576]	[4.162]	[2.018]	[7.195]	[6.200]	[5.332]	[4.376]	[3.188]
	(30.14)	(25.23)	(18.67)	(8.801)	(0.394)	(42.38)	(41.53)	(39.67)	(35.54)	(24.18)
50	9.477	9.130	9.078	9.359	9.839	6.883	6.043	5.392	4.817	4.484
	[7.356]	[6.318]	[5.277]	[3.826]	[1.789]	[6.912]	[5.956]	[5.117]	[4.191]	[3.034]
	(28.94)	(23.58)	(16.47)	(6.75)	(0.172)	(42.20)	(41.11)	(39.03)	(34.31)	(22.13)
100	9.372	9.085	9.092	9.406	9.864	6.705	5.895	5.289	4.744	4.482
	[7.171]	[6.125]	[5.082]	[3.609]	[1.649]	[6.739]	[5.795]	[4.978]	[4.075]	[2.935]
	(28.24)	(22.37)	(15.11)	(5.564)	(0.081)	(42.25)	(40.95)	(38.54)	(33.56)	(20.76)
250	9.294	9.062	9.107	9.443	9.877	6.573	5.803	5.213	4.701	4.482
	[7.040]	[5.996]	[4.937]	[3.451]	[1.558]	[6.614]	[5.686]	[4.885]	[3.991]	[2.860]
	(27.74)	(21.55)	(14.01)	(4.757)	(0.049)	(42.31)	(40.79)	(38.28)	(32.99)	(19.67)

Table 6: Monte Carlo simulation results for the baseline effective spread estimator, $ES_T^{(1)}$, under stochastic volatility of the efficient price. For each T (number of days in the simulated sample) and n (number of transactions on each day in the simulated sample), the table reports the average simulated $ES_T^{(1)}$, the simulated standard deviation of $ES_T^{(1)}$ in brackets and the percentage of zero $ES_T^{(1)}$ in the simulation in parentheses. The efficient price innovations are normally distributed and the daily integrated volatility of the efficient price is set to 0.35% or 35 bps. The results are based on 1 million Monte Carlo replications.

	T				T			
n	500	1000	2500	5000	500	1000	2500	5000
	$E(s_t) = 50$ bps				$E(s_t) = 20$ bps			
10	0.177	0.120	0.075	0.062	-0.385	-0.303	-0.222	-0.154
	[3.807]	[3.198]	[2.523]	[2.125]	[4.518]	[3.796]	[3.017]	[2.526]
25	0.197	0.138	0.088	0.066	-0.281	-0.221	-0.155	-0.099
	[3.501]	[2.930]	[2.320]	[1.929]	[3.863]	[3.252]	[2.560]	[2.144]
50	0.204	0.145	0.094	0.065	-0.234	-0.192	-0.128	-0.097
	[3.418]	[2.869]	[2.286]	[1.904]	[3.650]	[3.088]	[2.426]	[2.034]
100	0.198	0.142	0.092	0.070	-0.238	-0.185	-0.125	-0.082
	[3.373]	[2.827]	[2.242]	[1.887]	[3.555]	[3.007]	[2.356]	[1.961]
	$E(s_t) = 10$ bps				$E(s_t) = 5$ bps			
10	-0.599	-0.632	-0.619	-0.549	0.897	0.563	0.215	0.028
	[5.978]	[5.363]	[4.605]	[4.069]	[5.857]	[5.370]	[4.796]	[4.434]
25	-0.609	-0.612	-0.561	-0.459	0.636	0.357	0.052	-0.059
	[5.438]	[4.865]	[4.122]	[3.615]	[5.441]	[5.006]	[4.470]	[4.134]
50	-0.586	-0.587	-0.520	-0.457	0.588	0.302	0.031	-0.122
	[5.265]	[4.698]	[3.964]	[3.476]	[5.317]	[4.892]	[4.370]	[4.040]
100	-0.607	-0.576	-0.513	-0.438	0.521	0.271	0.000	-0.134
	[5.176]	[4.599]	[3.882]	[3.400]	[5.245]	[4.828]	[4.318]	[3.989]

Table 7: Monte Carlo simulation results for the time-varying effective spread estimator, $ES_{t,T}$. For each T (number of days in the simulated sample) and n (number of transactions on each day in the simulated sample), the table reports the average deviation of the estimated spread from the true value and the root mean-squared error in brackets. The efficient price innovations are normally distributed and the daily integrated volatility of the efficient price is set to 0.35% or 35 bps. The results are based on 10,000 Monte Carlo replications.

	T					T				
n	25	50	100	250	1000	25	50	100	250	1000
	$s = 50$ bps					$s = 20$ bps				
5	60.43	60.47	60.45	60.47	60.47	45.09	45.12	45.09	45.13	45.11
	[4.486]	[3.211]	[2.249]	[1.438]	[0.712]	[4.793]	[3.423]	[2.398]	[1.538]	[0.763]
	(11.11)	(10.84)	(10.63)	(10.54)	(10.48)	(25.25)	(25.21)	(25.13)	(25.15)	(25.12)
10	59.42	59.37	59.41	59.40	59.42	43.21	43.16	43.21	43.20	43.22
	[3.610]	[2.538]	[1.831]	[1.159]	[0.580]	[4.342]	[3.050]	[2.185]	[1.377]	[0.693]
	(9.83)	(9.58)	(9.52)	(9.45)	(9.43)	(23.30)	(23.21)	(23.24)	(23.21)	(23.22)
25	58.88	58.83	58.82	58.82	58.81	42.15	42.07	42.07	42.06	42.04
	[3.087]	[2.188]	[1.526]	[0.970]	[0.482]	[4.094]	[2.900]	[2.020]	[1.279]	[0.636]
	(9.15)	(8.97)	(8.89)	(8.85)	(8.81)	(22.21)	(22.10)	(22.08)	(22.06)	(22.04)
50	58.60	58.59	58.61	58.62	58.61	41.66	41.65	41.64	41.67	41.66
	[2.839]	[2.013]	[1.423]	[0.903]	[0.454]	[3.932]	[2.807]	[1.982]	[1.263]	[0.630]
	(8.80)	(8.69)	(8.66)	(8.64)	(8.62)	(21.68)	(21.67)	(21.66)	(21.67)	(21.66)
100	58.49	58.54	58.53	58.53	58.51	41.43	41.51	41.49	41.49	41.46
	[2.726]	[1.928]	[1.371]	[0.871]	[0.430]	[3.898]	[2.760]	[1.961]	[1.246]	[0.612]
	(8.66)	(8.63)	(8.57)	(8.55)	(8.51)	(21.45)	(21.52)	(21.50)	(21.49)	(21.46)
250	58.48	58.46	58.46	58.46	58.46	41.37	41.36	41.37	41.36	41.36
	[2.688]	[1.863]	[1.335]	[0.845]	[0.425]	[3.913]	[2.720]	[1.955]	[1.227]	[0.616]
	(8.64)	(8.53)	(8.50)	(8.48)	(8.47)	(21.40)	(21.37)	(21.37)	(21.36)	(21.36)
	$s = 10$ bps					$s = 5$ bps				
5	42.67	42.70	42.67	42.72	42.70	42.06	42.08	42.05	42.10	42.08
	[4.898]	[3.490]	[2.448]	[1.569]	[0.780]	[4.931]	[3.509]	[2.463]	[1.576]	[0.784]
	(32.75)	(32.75)	(32.69)	(32.73)	(32.70)	(37.09)	(37.12)	(37.06)	(37.11)	(37.08)
10	40.58	40.54	40.59	40.57	40.59	39.90	39.87	39.92	39.90	39.92
	[4.527]	[3.180]	[2.275]	[1.430]	[0.722]	[4.578]	[3.217]	[2.298]	[1.444]	[0.730]
	(30.61)	(30.56)	(30.60)	(30.57)	(30.59)	(34.90)	(34.87)	(34.92)	(34.90)	(34.92)
25	39.34	39.25	39.26	39.24	39.23	38.62	38.52	38.53	38.51	38.50
	[4.325]	[3.061]	[2.133]	[1.351]	[0.672]	[4.388]	[3.105]	[2.164]	[1.372]	[0.682]
	(29.36)	(29.26)	(29.26)	(29.24)	(29.23)	(33.61)	(33.51)	(33.52)	(33.51)	(33.50)
50	38.77	38.76	38.75	38.78	38.77	38.02	38.01	38.00	38.02	38.01
	[4.176]	[2.985]	[2.108]	[1.343]	[0.669]	[4.244]	[3.034]	[2.143]	[1.366]	[0.680]
	(28.76)	(28.76)	(28.75)	(28.78)	(28.77)	(32.98)	(32.99)	(32.99)	(33.02)	(33.01)
100	38.50	38.58	38.56	38.56	38.53	37.72	37.80	37.78	37.78	37.75
	[4.157]	[2.945]	[2.092]	[1.328]	[0.653]	[4.229]	[2.997]	[2.129]	[1.351]	[0.665]
	(28.48)	(28.57)	(28.56)	(28.56)	(28.53)	(32.68)	(32.79)	(32.78)	(32.78)	(32.75)
250	38.42	38.41	38.41	38.40	38.41	37.63	37.62	37.62	37.61	37.62
	[4.180]	[2.909]	[2.091]	[1.310]	[0.658]	[4.254]	[2.962]	[2.129]	[1.333]	[0.669]
	(28.41)	(28.40)	(28.41)	(28.40)	(28.41)	(32.60)	(32.60)	(32.62)	(32.61)	(32.62)

Table 8: Monte Carlo simulation results for the regression-based effective spread estimator $2\hat{\beta}$ from regression (26). For each T (number of days in the simulated sample) and n (number of transactions on each day in the simulated sample), the table reports the average simulated $2\hat{\beta}$, the variance of $2\hat{\beta}$ in brackets and the root mean-squared error in parentheses. The efficient price innovations are normally distributed and the daily integrated volatility of the efficient price is set to 0.35% or 35 bps. The results are based on 1 million Monte Carlo replications.

B Technical appendix

B.1 Preliminaries

Dropping the subscript t to simplify notation, we have

$$\begin{aligned}\hat{s}^2 - s^2 &= 2(3\tilde{d} - \hat{d}) - s^2, \\ &= \frac{3s^2}{2}\left[\left(\frac{1}{n-1}\sum_{i=1}^{n}(q_i - \bar{q})^2\right) - 1\right] \\ &\quad + 2s\left[\frac{3}{n-1}\sum_{i=1}^{n}(m_i - \bar{m})(q_i - \bar{q}) - \frac{1}{n}\sum_{i=1}^{n}(m_i - m_0)q_i\right] \\ &\quad + 2\left[\frac{3}{n-1}\sum_{i=1}^{n}(m_i - \bar{m})^2 - \frac{1}{n}\sum_{i=1}^{n}(m_i - m_0)^2\right] \\ &=: A_n + B_n + C_n. \end{aligned} \tag{27}$$

B.2 Derivation of $\mathrm{Var}(\hat{s}^2)$

By construction, we have $\mathrm{E}(A_n) = \mathrm{E}(B_n) = \mathrm{E}(C_n) = 0$, and it is easy to show that $\mathrm{E}(A_nB_n) = \mathrm{E}(A_nC_n) = \mathrm{E}(B_nC_n) = 0$, since $\mathrm{E}(m_iq_j) = 0$ for all i and j. Thus, $\mathrm{Var}(\hat{s}^2) = \mathrm{E}(A_n^2) + \mathrm{E}(B_n^2) + \mathrm{E}(C_n^2)$. It is clear from the equation above that $\hat{s}^2$ does not depend on m_0 so we will set it to zero to simplify notation.

Starting with $\mathrm{E}(A_n^2)$, write

$$A_n^2 = \frac{9s^4}{4}\left[\frac{1}{(n-1)^2}\sum_{i=1}^{n}\sum_{j=1}^{n}(q_i - \bar{q})^2(q_j - \bar{q})^2 - \frac{1}{n-1}\sum_{i=1}^{n}(q_i - \bar{q})^2 + 1\right]. \tag{28}$$

Since $\mathrm{E}(q_iq_j) = 0$ if $i \neq j$ and $q_i^2 \equiv 1$, we have

$$\mathrm{E}\left(\sum_{i=1}^{n}\sum_{j=1}^{n}(q_i - \bar{q})^2(q_j - \bar{q})^2\right) = n^2 - 2\mathrm{E}\left(\sum_{i=1}^{n}\sum_{j=1}^{n}q_iq_j\right) + \frac{1}{n^2}\mathrm{E}\left(\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\sum_{l=1}^{n}q_iq_jq_kq_l\right), \tag{29}$$

$$= n^2 - 2n + 3 - \frac{2}{n}. \tag{30}$$

This, together with $\mathrm{E}\left(\sum_{i=1}^{n}(q_i - \bar{q})^2\right) = n - 1$, gives after some algebra

$$\mathrm{E}(A_n^2) = \frac{9s^4}{2n(n-1)}. \tag{31}$$

Turning to $\mathrm{E}(B_n^2)$, write

$$B_n^2 = 4s^2\left[\frac{9}{(n-1)^2} - \frac{6}{n(n-1)} + \frac{1}{n^2}\right]\sum_{i=1}^{n}\sum_{j=1}^{n} m_i m_j q_i q_j, \tag{32}$$

$$+ 4s^2\left[\frac{6}{n^2(n-1)} - \frac{18}{n(n-1)^2}\right]\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n} m_i m_j q_j q_k, \tag{33}$$

$$+ \frac{36s^2}{n^2(n-1)^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\sum_{l=1}^{n} m_i m_j q_k q_l. \tag{34}$$

Since $\mathrm{E}(\epsilon_i\epsilon_j) = 0$ if $i \neq j$,

$$\mathrm{E}\left(\sum_{i=1}^{n}\sum_{j=1}^{n} m_i m_j q_i q_j\right) = \sum_{i=1}^{n}\sum_{j=1}^{n}\mathrm{E}(m_i m_j)\mathrm{E}(q_i q_j) = \sum_{i=1}^{n}\mathrm{E}(m_i^2)q_i^2 = \sum_{i=1}^{n}\sum_{p=1}^{i}\mathrm{E}(\epsilon_p^2) = \frac{\sigma^2}{2}(n+1). \tag{35}$$

Similarly, $\mathrm{E}\left(\sum_i\sum_j\sum_k m_i m_j q_j q_k\right) = \sum_i\sum_j \mathrm{E}(m_i m_j)$ and $\mathrm{E}\left(\sum_i\sum_j\sum_k\sum_l m_i m_j q_k q_l\right) = n\sum_i\sum_j \mathrm{E}(m_i m_j)$. Thus, it remains to derive $\sum_i\sum_j \mathrm{E}(m_i m_j)$. The case $i = j$ follows from above so we focus on the case when $i \neq j$.

$$\sum_{\substack{i=1\\ i\neq j}}^{n}\sum_{j=1}^{n}\mathrm{E}(m_i m_j) = 2\sum_{i=1}^{n}\sum_{j=i+1}^{n}\mathrm{E}(m_i m_j), \tag{36}$$

$$= 2\sum_{i=1}^{n}\sum_{j=i+1}^{n}\mathrm{E}\left(\sum_{p=1}^{i}\left(\sum_{r=1}^{i} + \sum_{r=i+1}^{j}\right)\epsilon_p\epsilon_r\right), \tag{37}$$

$$= 2\sum_{i=1}^{n}\sum_{j=i+1}^{n}\sum_{p=1}^{i}\mathrm{E}(\epsilon_p^2), \tag{38}$$

$$= \frac{2\sigma^2}{n}\sum_{i=1}^{n}\sum_{j=i+1}^{n} i, \tag{39}$$

$$= \sigma^2 n(n+1) - \frac{1}{3}\sigma^2(n+1)(2n+1). \tag{40}$$

Plugging into (10) and simplifying gives

$$\mathrm{E}(B_n^2) = \frac{2s^2\sigma^2(2n^2+3n+1)}{n^2(n-1)}. \tag{41}$$

Finally, we derive $\mathrm{E}(C_n^2)$. Write

$$C_n^2 = 4\left[\frac{9}{(n-1)^2} - \frac{6}{n(n-1)} + \frac{1}{n^2}\right]\sum_{i=1}^{n}\sum_{j=1}^{n} m_i^2 m_j^2, \tag{42}$$

$$+ 4\left[\frac{6}{n^2(n-1)} - \frac{18}{n(n-1)^2}\right]\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n} m_i m_j m_k^2, \tag{43}$$

$$+ \frac{36}{n^2(n-1)^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\sum_{l=1}^{n} m_i m_j m_k m_l. \tag{44}$$

We focus on the last term since the other two terms follow from the derivation of this term. First observe that

$$\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\sum_{l=1}^{n} m_i m_j m_k m_l = \sum_{i=1}^{n} m_i^2 + 3\sum_{\substack{i=1\\ i\neq j}}^{n}\sum_{j=1}^{n} m_i^2 m_j^2 + 4\sum_{\substack{i=1\\ i\neq j}}^{n}\sum_{j=1}^{n} m_i^3 m_j, \tag{45}$$

$$+ 12\sum_{\substack{i=1\\ i<j<k}}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n} m_i^2 m_j m_k + 12\sum_{\substack{i=1\\ i<j<k}}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n} m_i m_j^2 m_k \tag{46}$$

$$+ 12\sum_{\substack{i=1\\ i<j<k}}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n} m_i m_j m_k^2 + 24\sum_{\substack{i=1\\ i<j<k<l}}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\sum_{l=1}^{n} m_i m_j m_k m_l \tag{47}$$

To save space, we only derive here the expectation of the last term; the other terms follow using the same approach.

$$\mathrm{E}\left(\sum_{\substack{i=1\\ i<j<k<l}}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\sum_{l=1}^{n} m_i m_j m_k m_l\right) \tag{48}$$

$$= \mathrm{E}\left(\sum_{i=1}^{n}\sum_{j=i+1}^{n}\sum_{k=j+1}^{n}\sum_{l=k+1}^{n}\sum_{p=1}^{i}\sum_{r=1}^{j}\sum_{s=1}^{k}\sum_{t=1}^{l}\epsilon_p\epsilon_r\epsilon_s\epsilon_t\right), \tag{49}$$

$$= \mathrm{E}\left(\sum_{i=1}^{n}\sum_{j=i+1}^{n}\sum_{k=j+1}^{n}\sum_{l=k+1}^{n}\sum_{p=1}^{i}\left(\sum_{r=1}^{i}+\sum_{r=i+1}^{j}\right)\left(\sum_{s=1}^{i}+\sum_{s=i+1}^{j}+\sum_{s=j+1}^{k}\right)\right. \tag{50}$$

$$\left.\times\left(\sum_{t=1}^{i}+\sum_{t=i+1}^{j}+\sum_{t=j+1}^{k}+\sum_{t=k+1}^{l}\right)\epsilon_p\epsilon_r\epsilon_s\epsilon_t\right), \tag{51}$$

$$= \sum_{i=1}^{n}\sum_{j=i+1}^{n}\sum_{k=j+1}^{n}\sum_{l=k+1}^{n}\left[\mathrm{E}\left(\sum_{p=1}^{i}\sum_{r=1}^{i}\sum_{s=1}^{i}\sum_{t=1}^{i}\epsilon_p\epsilon_r\epsilon_s\epsilon_t\right) + 3\mathrm{E}\left(\sum_{p=1}^{i}\sum_{r=1}^{i}\sum_{s=i+1}^{j}\sum_{t=i+1}^{j}\epsilon_p\epsilon_r\epsilon_s\epsilon_t\right)\right. \tag{52}$$

$$+\mathrm{E}\left(\sum_{p=1}^{i}\sum_{r=1}^{i}\sum_{s=j+1}^{k}\sum_{t=j+1}^{k}\epsilon_p\epsilon_r\epsilon_s\epsilon_t\right)\Bigg], \tag{53}$$

$$= \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=i+1}^{n}\sum_{k=j+1}^{n}\sum_{l=k+1}^{n} 3\sigma^4 i^2 + \kappa i + 3\sigma^4 i(j-i) + \sigma^4 i(k-j), \tag{54}$$

$$= \frac{\sigma^4}{3}n^4 + \left(\sigma^4 + \frac{\kappa}{5}\right)n^3 + \left(\frac{13\sigma^4}{12} + \frac{\kappa}{2}\right)n^2 + \left(\frac{\sigma^4}{2} + \frac{\kappa}{3}\right)n + \frac{\sigma^4}{12} - \frac{\kappa}{30}\frac{1}{n}, \tag{55}$$

where $\kappa = \mathrm{E}(\epsilon_1^4) - 3\sigma^4$. Above, we use the fact that

$$\mathrm{E}\left(\sum_{p=1}^{i}\sum_{r=1}^{i}\sum_{s=1}^{i}\sum_{t=1}^{i}\epsilon_p\epsilon_r\epsilon_s\epsilon_t\right) = \sum_{p=1}^{i}\sum_{r=1}^{i}\sum_{s=1}^{i}\sum_{t=1}^{i}\mathrm{E}(\epsilon_p\epsilon_r\epsilon_s\epsilon_t), \tag{56}$$

$$= 3\sum_{\substack{p=1\\p\neq r}}^{i}\sum_{r=1}^{i}\mathrm{E}(\epsilon_p^2\epsilon_r^2) + \sum_{p=1}^{i}\mathrm{E}(e_p^4), \tag{57}$$

$$= \frac{1}{n^2}(3\sigma^4 i^2 + \kappa i), \tag{58}$$

and

$$\mathrm{E}\left(\sum_{p=1}^{i}\sum_{r=1}^{i}\sum_{s=i+1}^{j}\sum_{t=i+1}^{j}\epsilon_p\epsilon_r\epsilon_s\epsilon_t\right) = \mathrm{E}\left(\sum_{p=1}^{i}\sum_{r=1}^{i}\epsilon_p\epsilon_r\right)\mathrm{E}\left(\sum_{s=i+1}^{j}\sum_{t=i+1}^{j}\epsilon_s\epsilon_t\right), \tag{59}$$

$$= \left(\sum_{p=1}^{i}\mathrm{E}(\epsilon_p^2)\right)\left(\sum_{r=i+1}^{j}\mathrm{E}(\epsilon_r^2)\right), \tag{60}$$

$$= \frac{1}{n^2}\sigma^4 i(j-i). \tag{61}$$

The expectation of the other terms in C_n^2 can be obtained analogously. We obtain

$$\mathrm{E}(C_n^2) = \frac{2(2\sigma^4 n + \sigma^4 + 2\kappa)(2n^3 + 7n^2 + 7n + 2)}{15n^3(n-1)}. \tag{62}$$

The variance of $\hat{s}^2$ then follows.

B.3 Time-varying volatility

Clearly, $\mathrm{E}(A_n) = \mathrm{E}(B_n) = 0$ even under time-varying volatility, since q and m are independent. Now for C_n, we have after some algebra

$$\mathrm{E}(C_n|\{\sigma_i\}_{i=1}^n) = \left[\frac{1}{n} + \frac{1}{n-1} - \frac{1}{n(n-1)}\right] \sum_{i=1}^{n}\sum_{p=1}^{i} \mathrm{E}(\epsilon_p^2|\sigma_p^2) - \frac{1}{n(n-1)} \sum_{i=1}^{n}\sum_{j=i+1}^{n}\sum_{p=1}^{j} \mathrm{E}(\epsilon_p^2|\sigma_p^2) \tag{63}$$

Now since $\epsilon_i = \frac{\sigma_i}{\sqrt{n}} z_i$, where z_i is *iid* with zero mean and unit variance, $\mathrm{E}[\mathrm{E}(\epsilon_p^2|\sigma_p^2)] = \frac{1}{n}\mathrm{E}(\sigma_i^2)$, and if $\mathrm{E}(\sigma_i^2) = const$, by the LIE, $\mathrm{E}[\mathrm{E}(C_n|\{\sigma_i\}_{i=1}^n)] = 0$. Thus, $\mathrm{E}(\hat{s}^2) = s^2$.

B.4 Time-varying effective spread

We first state the conditions the kernel K has to satisfy and then prove Proposition 1. $K(x) \geq 0$, $x \in \mathbb{R}$, is a continuous bounded function with a bounded first derivative such that $\int K(x)dx = 1$. $K(x) = O(e^{-cx^2})$, $\exists c > 0$, $|K'(x)| = O(|x|^{-2})$, $x \to \infty$. Several popular kernel functions satisfy these conditions, e.g. the Gaussian and quartic kernels.

Proof of Proposition 1. *To be typeset.*