A Simple Estimation of Bid-Ask Spreads from Daily Close, High, and Low Prices

Farshid Abdi and Angelo Ranaldo*

This version: July 2016

ABSTRACT

To estimate the bid-ask spread, we propose a new method that resembles the Roll measure (1984) but has some key advantages: it is fully independent of bid-ask bounces and benefits from a wider information set, namely, close, high, and low prices, which are readily available. Assessed against other low-frequency estimates, our estimator generally provides the highest cross-sectional and average time-series correlations with the TAQ effective spread benchmark. Moreover, it delivers the most accurate estimates for less liquid stocks. Finally, our estimator improves the measurement of systematic liquidity risk and commonality in liquidity for individual stocks and sorted portfolios.

* University of St. Gallen. Address: Swiss Institute of Banking and Finance, University of St. Gallen, Rosenbergstrasse 52, CH-9000 St. Gallen, Switzerland. E-mail addresses: farshid.abdi@unisg.ch (F. Abdi), angelo.ranaldo@unisg.ch (A. Ranaldo). We thank Yakov Amihud, Allaudeen Hameed, Robert Korajczyk, Asani Sarkar, Avanidhar Subrahmanyam, Paul Söderlind, and Jan Wrampelmeyer, as well as the participants of the 2013 CFE conference in London, and 2016 SFI research days in Gerzensee, for comments and suggestions. We acknowledge financial support from the Swiss National Science Foundation (SNSF grants 159418 and 154445).

Market liquidity is critical to market efficiency and financial stability. Inaccurate estimates of market liquidity can create misinformation about actual transaction costs and market malfunctioning, thus leading to an inefficient allocation of resources and a misperception of risks. As a result, the measurement of transaction costs has become a major topic in the financial literature. Whether and how one can precisely measure transaction costs if quote data are not available are key questions.

This paper provides a new method to accurately estimate the bid-ask spread based on readily available daily close, high, and low prices. Akin to the seminal *model* proposed by Roll (1984), the rationale of our estimator is the departure of the security price from its efficient value because of transaction costs. However, our estimator improves the Roll *measure* in two important respects: First, our method exploits a wider information set, namely, close, high, and low prices, which are readily available, rather than only close prices like in the Roll measure. Second, our estimator is completely independent of trade direction dynamics, unlike in the Roll measure, which relies on bid-ask bounces expected to occur only for 25% of the days. By virtue of its closed-form solution and straightforward computation, our method delivers very accurate estimates of effective spreads, both numerically and empirically. Compared with other daily estimates of transaction costs, our estimator generally provides the highest cross-sectional and average time-series correlation with the effective spread based on Trade and Quotes (TAQ) data, which serve as the benchmark measure. Our estimator can be applied for a number of research purposes and to a variety of markets and assets because it is derived under very general conditions and is easy to compute.

Our estimation of the effective spread shares the theoretical framework with the Roll (1984) model, in which the efficient price of an asset follows a geometric Brownian motion. Within this framework, we follow three innovative steps to derive our simple estimator. First, we build a simple proxy for the efficient price using the mid-range, which we define as the mean of the daily high and low log-prices. The mid-range of every day represents (at least) one point in the continuous path of the efficient log-price process as half-spreads included in the high and low prices cancel out in the mid-range calculation.

Moreover, the mean of two consecutive daily mid-ranges represents a natural proxy for the midpoint or efficient price at the time of the closure. In fact, the continuous efficient price path of day t ($t+1$) hits the mid-range before (after) the closing time on day t. Second, we calculate the squared distance between the close log-price and the midpoint proxy at the time of the closure. We show that this squared distance is composed of the efficient-price variance and the squared effective spread at the closing time. As the third step, we derive an efficient-price variance estimator as a function of mid-ranges. The efficient-price variance is then removed from the squared distance between the close price and midpoint proxy (obtained in the previous step). The outcome is a simple measure for the proportional spread, $Spread = 2\sqrt{E[(c_t - \eta_t)(c_t - \eta_{t+1})]}$, in which c is the daily close log-price and η is the daily mid-range, that is, the average of daily high and low log-prices. This simple closed-form solution resembles the Roll's autocovariance measure. However, instead of the covariance of consecutive close-to-close price returns like in the *Roll* measure, our estimator relies on the covariance of close-to-mid-range returns around the same close price.

Accurate measures of transaction costs also can be obtained from quote data, such as intraday or end-of-the-day bid and ask quotes. However, quote data at any frequency entail some difficulties and limitations. The use of high-frequency quote data involves drawbacks, including very limited access to recent data and to some securities, a restricted and delayed use, and the need for time-consuming data handling and filtering techniques.[1] The *approximation* of transaction costs using bid and ask quotes snapped at daily or lower frequencies (e.g., Stoll and Whaley (1983)) is still relatively accurate (Chung and Zhang (2014)). However, it shares many drawbacks with the use of intraday quotes, such as limited access to bid and ask quotes that, in turn, are affected by intraday patterns at the point of time when recorded.[2] The previous literature overcomes mentioned issues by employing price data to *estimate* the

[1] The key advantages of using daily data, including large computational time savings, are comprehensively discussed by Holden, Jacobsen, and Subrahmanyam (2014).

[2] Rather than approximating and estimating transaction costs, an alternative approach to measuring illiquidity is to use proxies for the price impact, in particular the Amihud (2002) measure.

effective spread. Starting with the Roll (1984) measure (hereafter *Roll*), a number of models have been proposed. Hasbrouck (2004, 2009) proposes a Gibbs sampler Bayesian estimation of the Roll model (hereafter *Gibbs*). Lesmond, Ogden, and Trzcinka (1999) introduce an estimator based on zero returns (*LOT*). Compared with *Roll*, estimating the *LOT* measure is computationally intensive since it relies on optimizing the maximum likelihood function for every single month to get the monthly estimates. Following the same line of reasoning, Fong, Holden, and Trzcinka (2014) develop a new estimator (*FHT*) that simplifies existing *LOT* measures. Holden (2009), jointly with Goyenko, Holden, and Trzcinka (2009), introduces the Effective Tick measure based on the concept of price clustering (*EffTick*). And, more recently, Corwin and Schultz (2012) put forward an original estimation method using high and low prices rather than close prices (*HL*).

This paper contributes to the literature by providing a new estimation method of transaction costs *jointly* based on close, high, and low prices. The rationale of our model is to bridge the two above-mentioned estimation methodologies, that is, the long-established approach based on close prices originated from Roll (1984) and the most recent one relying on high and low prices (Corwin and Schultz (2012)). In doing so, our model has four main advantages over the previous estimation methods. First, the joint utilization of the daily high, low, and close prices allows our model to benefit from the richest readily available information set of price data.[3] Second, unlike Roll (1984), our measure is independent of trade direction dynamics and, consequently, it does not rely on bid-ask bounces to capture the effective spreads. Third, unlike Corwin and Schultz's (2012) *HL* estimator, our model neither needs to violate Jensen's inequality in order to construct the closed-form estimator nor does it need *ad hoc* adjustments for nontrading periods, such as weekends, holidays, and overnight closings. Finally, our estimates are only marginally sensitive to the number of trades per day, whereas the Corwin and Schultz (2012) method generates substantially lower estimates of effective cost when the daily number of trades are lower, that is, when stocks (and markets) are less liquid.

[3] Unlike close, high, and low prices, the availability of open prices is subject to additional limitations. For example, open prices are missing in the CRSP data between July 1962 and June 1992.

We empirically test our method by using daily CRSP data to estimate bid-ask spreads and compare the monthly estimates to TAQ data, which serves as the benchmark to compute the effective spread. As recommended by Holden and Jacobsen (2014), we use Daily (Millisecond) TAQ data to enhance the precision of our analysis. Thus, the availability of the Daily TAQ data naturally defines our main sample period, which spans from 2003 to 2014, that is, 135 months. Then, we assess the performance of our method by comparing bid-ask spread estimates with the Monthly TAQ data between January 1993 and September 2003 thus extending our analysis to 22 years of TAQ data, that is, from the beginning of 1993 to the end of 2014. As emphasized in the literature, for example, by Goyenko, Holden, and Trzcinka (2009), the decision criteria for selecting the best estimator depends on the particular application of the estimates. To cover the widest range of possible applications, we use three different criteria to gauge the quality of the estimators: cross-sectional correlation, time-series correlation, and prediction errors. To ensure a comprehensive assessment, we consider the average correlations for all the available stocks, as well as for subsamples, based on a variety of criteria, including shorter time periods, primary exchanges (NYSE, AMEX, and NASDAQ), market capitalization, and the magnitude of bid-ask spreads.

Three clear results emerge from our study. First, our estimator provides the highest cross-sectional correlation with the intraday effective spread compared with other bid-ask spread estimators, namely, the *HL*, *Roll*, *Gibbs*, *EffTick*, and *FHT* measures. On a monthly basis, the average cross-sectional correlation of our estimates with the Daily TAQ effective spreads is 0.74, whereas the other estimators range from 0.38 to 0.65. The analysis of Monthly TAQ data from 1993 to 2003 delivers consistent results, that is, our estimates have the highest average cross-sectional correlation of 0.86, whereas those of other estimators range from 0.606 to 0.833. These results are consistent whether correlations are taken for estimates in levels or in changes, across subperiods, and across different markets. When breaking down the cross-section of stocks into quintiles based on companies' size and effective spread size, we find that our estimator provides the highest cross-sectional correlations for small to medium market capitalizations and for a medium to large effective spread size. This can be seen as a suitable characteristic because accurate estimates of transaction costs are particularly needed for less liquid securities. Second, our estimator also delivers the highest average time-series correlations with the effective spread benchmark. Compared with other estimators, it provides the highest average correlations over the entire sample period, across two out of three market venues (AMEX and NASDAQ), for small to medium market capitalizations, and for a medium to large effective spread size. Finally, our estimates generally exhibit the lowest prediction errors in terms of root-mean-squared errors (RMSEs) when compared with the TAQ benchmark. Overall, the empirical evidence indicates that our straightforward method generally offers the most accurate estimation of the intraday effective spread in three respects: cross-sectional correlation, time-series correlation, and RMSEs.

A natural question is whether our estimator outperforms *combinations* of other estimators. We answer this question in two ways. First, we compare our estimates with the average of other estimates. Compared with the (overidentified) combinations of all other estimates, our estimates provide the highest average cross-sectional and time-series correlations with the TAQ effective spread benchmark, as well as the lowest average prediction errors. Second, we measure partial correlations between our estimates and the TAQ benchmark, while controlling for *HL*, *Roll, Gibbs, EffTick*, and *FHT* estimates. We find that the average partial cross-sectional and partial time-series correlations for our estimates are significantly positive for the entire sample, for every primary exchange, and for every effective-spread quintile. This confirms the additional explanatory power of our estimates. Average partial correlations are especially higher for quintiles with a medium to large effective spread size; that is, our estimator is even more effective in predicting the effective spread of less liquid stocks. We provide a numerical and empirical analysis showing that this quality is due to the marginal sensitivity of our estimates to the number of trades per day, whereas Corwin and Schultz's (2012) method produces substantially smaller estimates of transaction costs for less-frequently traded stocks.

An accurate measurement of transaction costs is important for at least two applications: First, to analyze how and to what extent transaction costs erode asset returns (e.g., Amihud and Mendelson

(1986)). To illustrate the potential application in this respect, we compute estimates of bid-ask spreads for NYSE and AMEX stocks for the period from 1926 through 2014. Then we discuss the reliability of our estimator in describing the developments of transaction costs over long time spans and for large cap and small cap stocks. Second, investors demand a premium for liquidity risk, that is, the chance that liquidity disappears when it is needed to trade. To comprehend this issue, it is necessary to obtain accurate estimates of transaction costs for individual stocks, stock portfolios, and the whole market. Through the lens of the liquidity-adjusted capital asset pricing model (LCAPM), proposed by Acharya and Pedersen (2005), we analyze which model provides accurate estimates of systematic liquidity risk for individual stocks and sorted portfolios, that is, estimates close to those based on the TAQ effective spreads. We show that our model precisely captures all the different components of systematic liquidity risk, in particular the component originated by comovements of liquidity of individual stocks and that of the whole market, that is, commonality in liquidity, as well as positive covariations between stock returns and liquidity. Overall, our model provides more accurate estimates for (liquidity) systematic risk than do the *Roll* and *HL* estimators, and it can be used to analyze commonality in liquidity and return-liquidity covariations. Our estimator has many potential applications in areas other than asset pricing, including corporate finance, risk management, and other important research areas that need an accurate measure of trading costs over long periods.

I. The Estimator

In this section, we explain our model in theory and provide details for best use in practice. At the end of the section, we numerically examine our estimator under different simulation scenarios.

A. Model

Our model relies on assumptions similar to those made in the Roll (1984) model. We assume that the efficient price follows a geometric Brownian motion (GBM) and the observed price at each time point can be either buyer initiated or seller initiated. To keep the notation concise, we directly implement the model on *log-price*, and the superscript *e* refers to efficient prices. Equation (1) shows how the observed market price and efficient price at the closing time are related. The random variable c_t represents the observable close log-price, and the random variable c_t^e represents the efficient log-price at the closing time. The random variable q_t is the trade direction indicator, and s is the relative spread, which we aim to estimate. In line with Roll (1984), we assume that trade directions are independent of the efficient price.

Based on the assumption of a continuous efficient price path, the daily high price (h_t) will always be buyer initiated ($q_t = 1$), and the daily low price (l_t) will always be seller initiated ($q_t = -1$). Equations (2) and (3) represent these points.

$$c_t = c_t^e + q_t \frac{s}{2} \quad , \; q_t = \pm 1 \, , \tag{1}$$

$$h_t = h_t^e + \frac{s}{2} \, , \tag{2}$$

$$l_t = l_t^e - \frac{s}{2} \, . \tag{3}$$

For the moment, we assume that the efficient-price movement during nontrading periods (e.g., overnight closings) is zero, but we will show later, both in numerical simulations and analytically, that our results are robust to the relaxation of this assumption (see the proof in Appendix B). We start with defining mid-range and then derive our estimator using the mid-range.

Definition 1. We define the mid-range as the average of daily high and low log-prices:

$$\eta_t \equiv (l_t + h_t)/2. \tag{4}$$

One can replace the efficient high and low log-prices with the observed values since the spreads cancel out.

Proposition 1. Assuming that the efficient price follows a continuous path (in our case a GBM):

(i) The mid-range of observed prices coincides with mid-range of efficient price:

$$\eta_t = (l_t^e + h_t^e)/2. \tag{5}$$

(ii) η_t represents at least one point in the efficient-price process. In other words, the efficient price hits η_t at least once during the day.

(iii) A straightforward and unbiased proxy for the end-of-the-day midquote of day t is the average of mid-ranges of the same day and the next day, since the end of the day midquote of day t occurs between the time at which η_t and $\eta_{t+1}$are hit. As shown in equation (6), this proxy is unbiased:

$$E[c_t^e - (\eta_t + \eta_{t+1})/2] = 0. \tag{6}$$

Proposition 2. The squared distance between close log-price of day t and the proposed mid-point proxy includes two components: bid-ask spread component and efficient price variance component. Equation (7) shows this relation:

$$E[(c_t - (\eta_t + \eta_{t+1})/2)^2] = s^2/4 + (1/2 - k_1/8)\sigma_e^2\ , \ \ k_1 \equiv 4\ln(2). \tag{7}$$

Garman and Klass (1980) use the value of $k_1$for the purpose of estimating volatility using the daily price range. Proofs for Propositions 2 and 3 are available in Appendix A. The effective spread, by definition, is the distance between the price and the contemporaneous midquote. We interpret equation (7) as a characterization of the standard definition of the effective spread, that is, when the unobservable midpoint is proxied by the average mid-ranges. We argue that the average of the consecutive mid-ranges of days t and $t+1$ is a natural proxy for the midquote or the efficient price at the closing time of day t since the mid-range of day t occurs before the closing time and the mid-range of the next day occurs after it. As expressed in equation (7), the squared distance between the close price and the proxy for the midquote contains two components: the squared effective spread and the transitory variance. The squared effective spread term represents the squared distance between the observed close price and the midquote at the time of market closure. The transitory variance term represents the squared distance between the midquote at the closure time and its approximation, that is, the average of two consecutive mid-ranges. Figure 1 provides a graphical illustration of the two components of the dispersion measure introduced in equation (7) in the framework of the Roll (1984) model. The figure illustrates that the distance between the close price and the average of the two consecutive mid-ranges reflects two quantities, namely, the effective spread and the intraday efficient-price variation (σ_e^2). As the next step, we propose a way to compute a measure of intraday volatility, which we will remove from the dispersion between the close price and the midquote proxy.

Proposition 3. The variance of changes in mid-ranges is a linear function of efficient price variance. Equation (8) provides the accurate relation:

$$E[(\eta_{t+1} - \eta_t)^2] = (2 - k_1/2)\sigma_e^2 \ , \ \ k_1 \equiv 4\ln(2). \tag{8}$$

Since the mid-ranges are both independent of the spread, their difference only reflects the volatility of the efficient-price path. We also perform several numerical simulations to assess the quality of the estimate of the efficient price volatility in Proposition 3. We find two main results: First, the estimated efficient price volatility implied by our model closely follows the "true" efficient price volatility. Second, our volatility estimate is less sensitive to the trading frequency. In other words, it is still accurate and less biased than the high-low volatility estimates, even for a very low frequency of trades. This is a favorable property of our volatility estimates compared to the use of price range, which, as shown in Garman and Klass (1980), is considerably biased if the trades are observed less frequently. Figure 2 illustrates the explained simulation results. Proposition 3 provides us with a way to remove the efficient price variance part introduced in Proposition 2.

Theorem 1. The squared effective spread can be estimated as shown in equation (9):

$$s^2 = 4E[(c_t - (\eta_t + \eta_{t+1})/2)^2] - E[(\eta_{t+1} - \eta_t)^2] = 4E[(c_t - \eta_t)(c_t - \eta_{t+1})]. \tag{9}$$

Proof of Theorem 1: Multiplying both sides of equation (7) by four, subtracting equation (8), and simplifying the outcome expression leads to equation (9).

Interestingly, the estimator derived in Theorem 1 resembles the *Roll* autocovariance measure, in which $c_{t+1}(c_{t-1})$ is replaced with $\eta_{t+1}(\eta_t)$. However, this simple and intuitive formulation leads to some important improvements. Hereafter, we compare our estimator to the *Roll* and *HL* measures.

As illustrated in equation (10), the expected value of every two-day observation of our estimator gives the squared effective cost to trade independent of trade direction patterns. This independence provides two key advantages. First, contrary to the *Roll* measure, there is no need for any restrictive assumptions on the dynamics of the trade direction, for example, on the serial independence of trades and equal likelihood of buyer-initiated and seller-initiated trades. Second, there is no need for observing the bid-ask bounces because each two-day observation of days t and $t+1$ only relies on c_t. In contrast, the *Roll* autocovariance estimates involve three-day closing prices ($c_{t-1}, c_t,$ and c_{t+1}) that reveal the transaction costs only in case of reversals of trade direction at time t. As explained in Equation (11), these reversals across three days (that is, buy, sell, and buy or sell, buy, and sell) are expected to occur only one out of four times. This means that, on average, 15 observations in a 20-day month are completely uninformative about the bid-ask spread.[4]

$$E[4(c_t-\eta_t)(c_t-\eta_{t+1})|q_t] = q_t^2 s^2 = s^2, \tag{10}$$

$$E[-4(c_{t+1}-c_t)(c_t-c_{t-1})|(q_{t+1},q_t,q_{t-1})] = -s^2(q_{t+1}-q_t)(q_t-q_{t-1}) =$$

$$\begin{cases} 4s^2, (q_{t+1}-q_t)(q_t-q_{t-1}) = -4, & [Prob. = 0.25] \\ 0 \quad , (q_{t+1}-q_t)(q_t-q_{t-1}) = 0, & [Prob. = 0.75] \end{cases}. \tag{11}$$

Compared to the *HL* estimates' performance (Corwin and Shultz (2012)), our model should perform better since it uses a wider information set, that is, the close, high, and low prices, whereas the *HL* estimator only relies on high and low prices. Moreover, as discussed in Proposition 3, since our estimates

[4] Choi, Salandro, and Shastri (1988) extend the Roll (1984) model by allowing for serial correlation in trade direction. With realistic conditional probabilities of trade continuation, this implies that trade reversals occur even more infrequently, that is, fewer than one out of ten times.

do not depend on the range, they are less sensitive to the number of trades per day. This feature is crucial to estimate accurately the transaction costs for less liquid stocks. We will return to this issue in Sections III and IV, when we empirically analyze various estimators of the effective spread.

One practical way to get an estimate of s^2 is to replace the term on the right-hand-side in equation (9) with monthly averages. However, the estimates of s^2 can become negative. In the next section, we explain how to deal with this issue.

B. Use in practice: How to deal with negative estimates

We aim to use the model to estimate effective spreads for every month-stock. The first approach that comes to mind is to replace expectations with monthly averages in equation (9), and if the result is negative, report it as zero. We call this version the *monthly corrected* version.

An alternative method is the following: First, we calculate estimates of squared spreads over two-day periods. If the two-day estimates are negative, then we set them to zero. Second, we take their square roots. Finally, we average them over a month. The three steps are summarized in equation (12):

$$\hat{s}_t = \sqrt{\max\{4(c_t - \eta_t)(c_t - \eta_{t+1})\,,0\}}\ , \qquad \hat{s}_{month} = \frac{1}{N_{month}} \sum_{t \in month} \hat{s}_t. \tag{12}$$

The estimate of the two-day spread and the monthly spread are represented by $\hat{s}_t$ and $\hat{s}_{month}$. We call this version the *two-day corrected* version, as we replace negative two-day estimates with zero. This is similar to the correction method applied by Corwin and Schultz (2012). Although the two-day correction approach increases the bias because of setting more negative values to zero compared with the monthly corrected version, it provides better results in terms of higher correlation with the high-frequency benchmark, as documented by Corwin and Schultz (2012).

The better performance of the two-day corrected version can be explained by some restrictive assumptions in the Roll (1984) model, which our estimator also relies on, in particular the constant spread and volatility. First, the monthly corrected estimate hinges on $E(s^2)$, which consists of the squared mean, plus the variance of bid-ask spreads. This is larger than the squared mean when the spread is not constant. With the use of a two-day period for the spread estimation, we isolate a single incident of a close-price transaction, and, therefore, no assumption on the distribution of the spread over consecutive days is needed. Second, the two-day time window is more prone to capturing transient price patterns, such as heteroskedasticity and volatility clustering.

In the next part, by numerical simulations under different settings, we show that the two-day corrected version of our estimates is more precise, in terms of having a smaller estimation variance. We also show that when the trading cost is not constant, the two-day corrected version can be less biased than the monthly corrected version.

C. Simulation results

In this subsection, we perform several numerical simulations under different settings. For ease of comparison, we set the setting of simulations, including the less-than-ideal conditions, similar to the settings proposed by Corwin and Schultz (2012). We compare two versions of our measure with the *HL* and *Roll* estimates, that is, the monthly corrected and the two-day corrected versions.[5]

Table I shows the results of the simulation exercise in five different scenarios. We label our estimator *CHL* because it is based on the close, high, and low prices. Panel A shows the results for the near-ideal settings. For each relative spread under analysis, we perform 10,000 time simulations for 21-day months of the price process. Each day consists of 390 minutes in which trades are observable. We simply draw from $M_t = M_{t-1}e^{z\sigma/\sqrt{390}}$, $P_t = M_t e^{q_t s/2}$, $z \sim N(0,1)$, where M_t and P_t represent the efficient price and observed transaction price at time t, respectively. We set the daily standard deviation of efficient-price returns, σ, to be 3%. Panel B has the same setting as panel A, but a lower trade frequency. Under this setting, each per-minute trade has a of 10% chance of being observed. This "imperfection"

[5] Shane Corwin has kindly provided the SAS codes for the HL estimator on his personal Web site. The code produces several versions of spread estimates. We consider two of them in our simulations. The first version, named MSPREAD_0, is calculated by setting two-day negative estimates to zero and then taking the monthly average. The second version, named XSPREAD_0, is calculated by directly setting the negative monthly averaged estimates to zero. Although the second version produces less-biased results in some simulation cases, Corwin and Schultz (2012) advocate the former method, which is better associated with the TAQ benchmark.

represents an interesting case as Garman and Klass (1980) introduce the bias in the range-based volatility estimation when the full price path is not observable. The setting in panel C is the same as in panel A, except that the spreads are no longer constant. By considering various spread sizes ($a\%$ spreads), the spreads for each day are randomly drawn from a uniform distribution with the range of $(0\,,2a)$. The setting in panel D adds an "overnight" price change corresponding to a half standard deviation of daily price returns to panel A. Overnight characterization actually represents more general nontrading periods, such as weekends, holidays, and overnight closings. Finally, panel E, all at once, encompasses the previous imperfections analyzed in panels B, C, and D.

Four main findings arise from the simulation analysis in Table I. First, in almost all of the cases, the standard deviations of two-day corrected spread estimates from the *HL* and our model are lower than those of the monthly corrected estimates. This finding supports the idea of tolerating a relatively small bias added by the two-day correction method in order to benefit from lower estimation variances. This lower estimation variances might also explain why the two-day corrected estimates are more strongly correlated with the TAQ benchmark than the monthly corrected ones. Second, the introduction of random spreads does not change the two-day corrected *CHL* estimates as much as it changes the monthly corrected estimates. This evidence also motivates the use of the two-day corrected version, rather than the monthly corrected one. Third, the standard deviations of the two-day corrected version of the *HL* and *CHL* estimates are very similar and about one-fourth of those of the *Roll* estimates, suggesting higher precision in the *HL* and *CHL* estimates. Finally, considering all imperfections together in the last panel of Table I, for the cases of larger spreads (from 3% to 8%) *CHL* two-day corrected estimates show the lowest bias and mean-squared errors, computed as the sum of estimation variances and the squared biases. This finding suggests that our model provides more accurate estimates than do the *HL* and *Roll* models for the less liquid securities, for which transaction costs and liquidity issues are of much more concern.

II. Other Spread Estimators that Use Daily Data

Of the different estimators developed by researchers, none jointly utilize close, high, and low prices data. In the rest of this section, we concisely describe the most common methods for bid-ask spread estimation, which we empirically analyze in the next sections.

A. Estimation based on return autocovariance and its improvements

Roll (1984) assumes that the efficient price follows a random walk and the observed transaction prices recorded at the end of day differ a half spread from the efficient price. The observed price can be a half spread higher or lower than the efficient price depending on whether the trade is buyer or seller initiated. Assuming an equal chance of observing a buyer- or seller-initiated price, the proportional spread can be estimated with the following formula:

$$s = 2\sqrt{-E[\Delta p_t \Delta p_{t-1}]}, \tag{13}$$

where s represents percentage spread and p is the logarithm of the close price.[6] To return a nonnegative spread, the first-order autocovariance of the price changes in equation (13) must be negative. However, Roll (1984) finds positive estimated autocovariances for several stocks, even over a one-year sample period. Hasbrouck (2004, 2009) uses a Gibbs sampler Bayesian estimator to overcome this issue. Using annual estimates, Hasbrouck (2009) shows that the spreads originated from the *Gibbs* method have higher correlations with the high-frequency benchmark. Following Corwin and Schultz (2012), among others, we perform our empirical analysis on a monthly basis.[7] This time granularity facilitates the time-series and cross-sectional comparisons with the previous literature and the high-frequency benchmark.

[6] As discussed in Roll (1984), the spread estimator can be computed with price changes or percent returns to estimate the spreads or proportional spreads, respectively. We use log-price and log-returns to estimates the proportional spreads. Percent and log returns deliver similar results. Furthermore, the use of log-prices eases the comparison with other models studied in the literature.

[7] Joel Hasbrouck has kindly provided the SAS codes for the Gibbs sampler estimator on his personal Web page. We modify the codes by altering the estimation windows from stock-years into stock-months. We only consider stock-months in which there are at least 12 days with trades. As he already noted in the Web page, the monthly estimator is less accurate than the annual version because of the overweigh role of the prior density in the outputs.

B. Estimation based on zero returns

Lesmond, Ogden, and Trzcinka (1999) introduce an effective-spread estimator based on the idea that the stock price return follows a market model. It is larger than the spread and zero otherwise. They use a maximum likelihood approach for their estimations. Goyenko, Holden, and Trzcinka (2009) propose two alternative methods to define the three feasible regions for the maximum likelihood estimation. The first alternative, named *LOT Mixed*, separates the different regions of observations according to the stock return and market return. In the second alternative, named *LOT Y-Split*, the separation of the three regions is solely based on the stock return. Goyenko, Holden, and Trzcinka (2009) document that these two measures can produce somewhat different results.

More recently, Fong, Holden, and Trzcinka (2014) develop another estimator, named *FHT*, which simplifies the existing *LOT Mixed* measure. Assuming that nonzero returns are observable only if they are larger than the spread, they calculate the proportional spread as follows:

$$s = 2\sigma N^{-1}\left(\frac{1+\text{Zeros}}{2}\right), \qquad \text{Zeros} = \frac{\text{ZRD}}{\text{TD}+\text{NTD}}, \tag{14}$$

where ZRD refers to the number of zero return days, TD refers to the number of trading days, and NTD refers to the number of no-trade days in a given stock-month. They assess the quality of their measure in estimating liquidity of the global equity market and find that it is one of the most accurate measures.[8]

C. Estimation based on price clustering

Holden (2009), jointly with Goyenko, Holden, and Trzcinka (2009), develops a proxy for the effective spread based on observable price clustering. Larger spreads are associated with larger effective tick sizes. Assuming that trade prices are clustered to minimize the negotiation costs among dealers, they consider the four possible S_j spreads of \$1/8, \$1/4, \$1/2, and \$1 to be randomly drawn with

[8] Fong, Holden, and Trzcinka (2014) consider three measures as the best bid-ask spread proxies, namely, their *FHT* estimator, the *HL* estimator, and the approximation using end-of-the-day bid and ask quotes as explained by Chung and Zhang (2014). We consider the *HL* and their *FHT* estimator in this paper, but we skip Chung and Zhang's (2014) measure because it directly uses quote data.

probability γ_j. They assume that price clustering is completely determined by the spread size. Then they calculate the empirical probabilities of trades on prices corresponding the jth spread as

$$F_j = \frac{N_j}{\sum_{j=1}^{J} N_j}.$$

The unconstrained probability of the effective spread is

$$U_j = \begin{cases} 2F_j, & j = 1 \\ 2F_j - F_{j-1}, & j = 2, \ldots, J-1. \\ F_j - F_{j-1}, & j = J \end{cases} \quad (15)$$

To avoid negative probabilities, and to make sure that the probabilities sum to one, they construct a constrained probability of the *j*th spread. Going from smallest to largest spreads, they compute the constrained probabilities as follows:

$$\hat{\gamma}_j = \begin{cases} \text{Min}\left[\text{Max}\{U_j, 0\}, 1\right], & j = 1 \\ \text{Min}\left[\text{Max}\{U_j, 0\}, 1 - \sum_{k=1}^{j-1} \hat{\gamma}_k\right], & j = 2, \ldots, J \end{cases} . \quad (16)$$

The reliance on the constrained probabilities makes it possible to calculate the expected spreads for the time interval i.

$$\text{Effective Tick}_i = \frac{\sum_{j=1}^{J} \hat{\gamma}_j s_j}{\bar{P}_i}. \quad (17)$$

We compute this measure by considering records of days for which a positive price and volume are recorded.

D. Estimation using high and low prices

Corwin and Schultz (2012) develop an estimator based on daily high and low prices. They argue that high (low) prices are almost always buyer (seller) initiated. Therefore, the daily price range reflects both the stock's volatility and its bid-ask spread. They build their model on the comparison of one- and two-day price ranges. The latter should twice reflect the variance of the former, but they should have the same bid-ask spread. This reasoning results in a nonlinear system of two equations with two unknowns that does not have a closed-form solution. The authors provide a simplified solution by neglecting Jensen's inequality and under the assumption that the expected efficient-price range squared has to be equal to the square of the expected efficient-price range.

By using the above simplification, they calculate the two-day spread as

$$s = \frac{2(e^{\alpha}-1)}{1+e^{\alpha}}, \qquad \alpha = \frac{\sqrt{2\beta}-\sqrt{\beta}}{3-2\sqrt{2}} - \sqrt{\frac{\gamma}{3-2\sqrt{2}}}, \tag{18}$$

where β represents a two-day sum of the daily log-price squared range and γ represents one observation of the two-day log-price squared range. The monthly estimates are calculated as the average of the above two-day estimates, where the negative two-day estimates are set to zero.

Their method needs an additional adjustment for nontrading periods, such as weekends, holidays, and overnight closings. If the close price on day t is higher (lower) than the high (low) price for day $t+1$, they adjust day $t+1$ high and low prices by shifting both of them together such that the high (low) price of day $t+1$ and close price of day t match.

III. A Comparison of Spread Estimates from Daily Data Using the TAQ Benchmark

We now turn to the analysis and comparison of the main estimation methods of transaction costs used in the literature. Using CRSP daily data, we estimate the effective spreads for common stocks listed in the main three stock markets in the United States, namely, NYSE, AMEX, and NASDAQ. In addition to our estimator, labeled *CHL*, we estimate the spreads originating from the following estimators: *Roll* (Roll (1984)), *Gibbs* (Hasbrouck (2009)), *EffTick* (Holden (2009); Goyenko, Holden, and Trzcinka (2009)), *HL* (Corwin and Schultz (2012)), and *FHT* (Fong, Holden, and Trzcinka (2014)). In the following analysis, we use the two-day corrected version for our estimator and for the HL measure, as recommended by Corwin and Schultz (2012).

To calculate our measure, we do the following: (1) we keep the previous daily high, low, and close prices on those days when a stock does not trade; (2) we use the two-day corrected version; that is, we set negative two-day estimates of squared spreads to zero and then take the square roots and average over the month; and (3) we discard estimates for months in which there are less than 12 trading days.[9,10]

To calculate the *HL* estimates, we exactly follow Corwin and Schultz (2012). More specifically, (1) we keep the previous daily high, low, and close prices on those days when a stock does not trade; (2) we perform the ad hoc overnight adjustment as explained in their paper; (3) we use the two-day corrected version; that is, we set negative two-day estimates to zero; and (4) we discard stock-months with less than 12 two-day estimates. We then calculate the other measures and merge all the estimations. We finally discard stock-months in which (1) any of the estimates produce a missing value, (2) a stock split or enormous distribution occurred, (3) a change of the primary exchange occurred, or (4) a stock has a time-series of less than six monthly estimates.[11, 12]

We construct the main high-frequency benchmark for our analysis by calculating the effective spread from Daily TAQ data. Equation (19) defines the proportional effective spread at time t. As recommended by Holden and Jacobsen (2014), we use Daily TAQ data, with milliseconds time stamps, instead of the Monthly TAQ data. In fact, the authors show that in fast, competitive markets of today, the Daily TAQ granularity is more precise, whereas the usage of Monthly TAQ data might lead to incorrect statistical inferences.

$$ES_t = \frac{2|P_t - M_t|}{M_t} \quad , \qquad M_t = \frac{B_t + A_t}{2}. \tag{19}$$

The time span of the data set, covering 135 months of Daily TAQ data, starts in October 2003 and ends in December 2014. To calculate the effective spread from Daily TAQ data, we closely apply the

[9] A trading day is defined as one with a positive closing price, high price, and low price and positive volume. Inclusion or exclusion of the volume criterion does not change any conclusions. It is also possible and accurate to replace missing η_{t+1} values, for the two-day estimates in which no trade occurs on day $t+1$, with readily available mid-quotes. However, to have a fair comparison with other estimates, we refrain from using mid-quotes.

[10] As we merge the estimates in the next step, this filter will be applied to other estimates as well. Therefore, all the estimates will have similar quality in terms of the selected sample.

[11] We discard stock-months in which the cumulative price adjustment factor (cfacpr) changes more than 20%.

[12] For example, the *Gibbs* estimator's code returns errors for the few stock-months in which the price remains constant for most of the days in the month, because the initial trade directions used in the simulations are calculated as sign of daily returns.

procedure explained by Holden and Jacobsen (2014).[13] More precisely, we first clean up the National Best Bid and Offer (NBBO) data set by removing any best bid (ask) in which the bid-ask spread is above five dollars and the bid (ask) is more than 2.5 dollars above (below) the previous midpoint. We also remove any quotes from the consolidated quotes (CQ) file if the spread is more than five dollars. Second, we merge the CQ and NBBO (cleaned) data to construct a complete official NBBO data set. Third, we match trades with constructed official NBBO quotes one millisecond before them. In addition to the above-mentioned filters, we discard all trades outside the market opening hours and with proportional effective spreads above 40%. We compute the dollar-weighted average for intraday proportional effective spreads to obtain the average daily spreads. Then we take the average of daily spreads to construct the monthly benchmark.

The final step in the data preparation is to link the CRSP and Daily TAQ using CUSIPs in the TAQ monthly master files.[14] This matching strategy allows us to cover 98% of stock-months estimates from the CRSP. We provide the summary statistics for the estimates in Table II. As we compare the pooled data in Table II, both the mean and standard deviation convey valuable information about the explanatory power of the estimators. The mean provides a simple measure for the level or size of the estimated transaction costs, and the standard deviation gives information about the time-series and cross-sectional dispersion of spread estimates around the mean. We also include overall correlations of estimates with the effective spreads benchmark, confirming the choice of two-day corrected estimates over monthly corrected estimates for our measure, as well as for the *HL* measure. Calculating *EffTick* estimates, we observe some stock-months in which none of the prices are divisible by the base-eight denomination increments. We report the estimates for these stock-months as zeros. For the purpose of comprehensiveness, we consider a second variant of the *EffTick* measure using the tick sizes of 1¢, 5¢, 10¢, 25¢, 50¢, or $1.00 as our sample time span lies after the decimalization of stock markets. In Table II,

[13] To calculate the effective spreads using Daily TAQ data, we use the same SAS codes kindly provided by Craig Holden on his Web site. We add additional criteria to keep the trades/quotes records with no symbol suffixes.

[14] We use the information in monthly master, files rather than that in the daily master files, because only the monthly master files are available for the entire sample period.

this second variant is labeled - *EffTick – Alt. Incr*. Clearly, this variant underestimates both the mean and the variations of the effective spread. Therefore, in the next sections we perform our analysis using the original base-eight denomination.

In addition to the effective spread and its estimates, we include Amihud (2002) illiquidity measure as the last row of the Table II. We multiply Amihud illiquidity proxy by one million to obtain easier-to-read values. The Amihud measure is a price impact proxy, rather than a transaction cost measure. This explains why (1) its mean and standard deviation are not directly comparable with the other spread estimates and (2) the correlation between the Amihud measure and the effective spread benchmark is generally lower than that for the spread estimators.

As a decomposition of the standard deviations reported in Table II, we also compute the cross-sectional standard deviation of the estimates on a monthly basis to assess how well the estimators' dispersion follows that of the TAQ benchmark across time. Figure 3 shows the results for some estimators. It is clearly evident that the cross-sectional dispersions from our estimator most closely track that of the benchmark.

We now turn to identifying which criteria should be used to assess the measurement performance of the effective spread estimators. As stressed by Goyenko, Holden, and Trzcinka (2009), the choice of the best estimator, depending on the specific application, should be based on different criteria. For the sake of completeness, our analysis encompasses the three main criteria used in the literature: cross-sectional correlation, time-series correlation, and prediction errors. This set of criteria should support a complete assessment and cover a wide range of applications. In the following part, we analyze the measurement performance of a set of estimators applying the three above-mentioned criteria.

A. Cross-sectional correlations

Cross-sectional correlation is an important criterion in assessing the estimation performance. This criterion is relevant in many applications, especially when the analysis is conducted for levels of transaction costs and for cross-sectional comparisons.

For each month, we calculate the correlation of the estimates with TAQ effective spreads that serve as the benchmark. Panel A in Figure 4 shows the development across time for the cross-sectional correlations of different estimators with the spread benchmark. It is clearly discernable that our estimator always provides the highest cross-sectional correlation. The results in panel A of Table III confirm that the estimates from our model (labeled *CHL*) have the highest cross-sectional correlation over the entire sample and across all subperiods. We apply the approach proposed by Goyenko, Holden, and Trzcinka (2009) to perform the statistical inferences to assess whether the average correlations are significantly different. More specifically, to compare the average correlation of the two estimators, we compute the pairwise difference of their cross-sectional correlations with the benchmark at each month. We then test if the average value for this time series is significantly different from zero, while adjusting for the autocorrelations using Newey-West (1987) standard errors with four lags. The findings in Table III indicate that the higher correlation coefficients of our estimator are statistically significant compared with any other measures.

We substantiate the previous analysis by examining the cross-sectional correlations in first differences, that is, taking the changes in monthly (estimated) spreads. Panel B of Figure 4 shows their developments across time, and panel B of Table III indicates whether these correlations are statistically significant. As expected, correlations based on changes in spreads are lower than those based on spread levels. However, the result is qualitatively the same. As for the correlation in levels, the average correlation in first differences of our estimator with the benchmark is the highest and significantly different from the other estimates.

Next, we perform a subsampling analysis of the cross-sectional correlation for levels of effective spreads across three dimensions: market venues, market capitalization, and effective spread size. First, we identify the three primary exchanges in which the stocks are listed using the CRSP exchange codes, that is, NYSE, AMEX, and NASDAQ. Second, we examine whether our results from the cross-sectional analysis depend on firm size. To do this, we decompose the entire sample into five quintiles by the firm's market capitalization value for each individual stock at the last observed period. Third, we consider whether our findings are sensitive to the magnitude of transaction costs. As before, we form five quintiles according to the average effective spread size over the entire sample period. The results of these three subsampling analyses are reported in panels C, D, and E of Table III, respectively.

Three main findings arise: First, our estimator provides the best results for stocks listed on the AMEX and NASDAQ. Though, when only NYSE stocks are considered, it shares the highest cross-sectional correlation with the *HL* and *EffTick* measures. Second, our estimator significantly outperforms the other measures across all market capitalizations, except for the fifth quintile (Quintile 5), which includes the largest capitalization. Third, our estimator performs significantly better than the other estimators for stocks traded with medium and large transaction costs (from Quintiles 3 to 5 sorted by smallest to largest effective spreads). In sum, our estimator provides the overall highest cross-sectional correlations with the effective spread benchmark. Its estimates are particularly accurate for stocks with lower liquidity, proxied by small-medium market capitalizations and effective spreads of medium and large magnitude. [15]

B. Time-series correlation

As the second criterion, we analyze stock-by-stock time-series correlations between the different spread estimates and the TAQ effective spread. We first calculate the time-series correlation between bid-ask spread estimates and the effective spread benchmark for each individual stock and each individual estimator. Then we compute the average of these time-series correlations across all sample stocks for each individual estimator. To compare the average correlations originating from different estimates, we use paired *t*-test.

[15] To make sure that the outperformance of our estimates are not potentially driven by the overweight of illiquid stocks in the linear correlation, we also calculate spearman's rank correlations for the cross-section of stocks. On a monthly basis, the average cross-sectional spearman's correlation of our estimates with the TAQ effective spreads is 0.68, whereas the other estimators range from 0.28 to 0.62.

Table IV shows the main results. Similar to Table III, panel A (B) of Table IV shows the average time-series correlations for the levels (changes) of effective spreads, while stocks are sorted by exchanges, market capitalization, and spread size in panels C, D, and E, respectively. Our model provides the highest average time-series correlation for monthly spreads of the overall sample and for two out of three subperiods. The only exception is the 2012–2014 subperiod, in which time-series correlations drop for all estimators and the *HL* measure performs slightly better than our model. For changes of spreads, the *HL* method generates the highest time-series correlation. Our estimator provides the second-highest time-series correlations, except for the 2008–2011 subperiod, which, in statistical terms, is not significantly lower than the *HL* one. The remaining parts of Table IV suggest that (1) our estimators outperform the others for stocks listed on the AMEX and NASDAQ, whereas the *HL* has the highest time-series correlation for NYSE stocks; (2) our measure (the *HL* measure) performs best for small- and medium-sized (large-sized) firms; and (3) our measure (the *HL* measure) performs best when stocks are traded with large (small) effective spreads.[16] The time-series correlation analysis confirms the previous findings that our estimator generally provides the most accurate estimates of effective costs, especially for less liquid stocks.

C. Prediction errors

A straightforward way to assess the quality of an estimator is to observe how far it is from its benchmark value. We measure this by root-mean-squared errors (RMSEs) and mean-absolute errors (MAEs) of monthly estimates with respect to the TAQ effective benchmark. In line with Goyenko, Holden, and Trzcinka (2009), we calculate prediction errors every month and then average them during that time.

As shown in panel A of Table V, our estimator (labeled *CHL*) provides the lowest RMSEs compared with other estimators across the entire time series. The difference between average RMSEs of our

[16] As an additional test, which we report in the Internet Appendix, we construct equally weighted portfolios of stocks and then compare the correlation of the estimated portfolios' spread to that of the high-frequency benchmark. The market-wide portfolio shows a time-series correlation of 0.965 with the TAQ benchmark.

estimates and the other estimates is also significant, using Newey-West (1987) standard errors with four lags to test whether the time-series of pairwise difference of RMSEs is statistically different from zero.

As our sample includes several stock-months without any observed zero returns, *FHT* estimates are zero for about 60% of stock-months (see Table II). Therefore, one shall interpret the reported prediction errors of *FHT* estimates considering this limit; doing so translates into lower prediction errors at the cost of lower association with the benchmark, in terms of time-series and cross-sectional correlations (see Tables III and VI). Using MAEs, as in panel B of Table V, we observe even lower prediction errors for *FHT* estimates. Interestingly, even in this setting, our estimates provide the lowest MAEs for AMEX and NASDAQ stocks.

All in all, our estimates generally show the highest average time-series and cross-sectional correlations, as well as the lowest RMSEs, with respect to the Daily TAQ benchmark. As showed in Table VI, these results are fully confirmed when we repeat the analysis for the period of January 1993 to September 2003 using the Monthly TAQ effective spreads.[17]

IV. Further Analysis of *CHL* Performance

Since our *CHL* estimates jointly use close, high, and low prices, which are also partially used in other estimators discussed in the paper, it is worth comparing the quality of our estimates with *combinations* of estimates originated by the other models. The main purpose of this analysis is to make certain that the *CHL* estimates contain additional information compared to that provided by other models. To do this, we proceed in two steps. First, we compare the performance of *CHL* with a set of overidentified models, which are averages of spread estimates. Second, and more rigorously, we regress TAQ effective spreads on a set of estimates, including *CHL*. Using this setting, we calculate the partial correlation between *CHL* and TAQ effective spreads, controlling for the explanatory power of the other estimates. The main result of these analyses is that the *CHL* estimates do provide additional explanatory power to explain the

[17] See the Internet Appendix for more details on the construction of Monthly TAQ benchmark and additional analysis.

variations in effective spreads, even after controlling for its common sources of variation with other estimates. In the final part of this section, we try to understand why.

A. Comparison with naïve combinations of models

Using the estimates from Section III, we compare the performance of the *CHL* estimates with combinations of other estimates. As is common in the literature (e.g. Chordia, Roll, and Subrahmanyam 2000), we compute averages across estimates, as we expect doing so will reduce the noise due to estimation errors in different estimates.[18] As all the estimates proxy the size of the effective spread, we are entitled to compute a simple average. As comparison criteria, we use cross-sectional correlation, time-series correlation, and RMSEs with respect to the TAQ benchmark.

Table VII shows the main results. As observable in panel A of Table VII, the average of the *Roll* and *HL* estimates (fourth column) and of any combinations of the other estimates (fifth and sixth columns) provide lower average cross-sectional correlations with the TAQ benchmark than the *CHL* estimates, even when the *CHL* estimates are included (last two columns). This result holds true when we consider changes in bid-ask spread estimates, subsamples of the stocks and primary markets. Furthermore, the results are fully consistent when we analyze average time-series correlations with the TAQ benchmark (panel B of Table VII) and prediction errors (panel C of Table VII).

All in all, the results of this subsection show that our method provides more accurate estimates of effective costs than do (overidentified) combinations of estimates obtained from other methods. This speaks in favor of the joint utilization of the close, high, and low prices in building estimates of transaction costs, like in our model, rather than using more limited information sets, like in other models. We combine the results. In the next subsections, we shed light on the additional explanatory power of the *CHL* estimates.

[18] We also consider weighted average measures. To do so, we perform a principal component (PC) analysis and use the coefficients in first PC as the weights of the estimates in the weighted average. The results using this approach are consistent and do not outperform *CHL*. For instance, the pooled correlation for the first PC of *HL*, *Roll*, *Gibbs*, *EffTick*, and *FHT* with the TAQ effective spread is 0.51, compared to 0.62 for the simple average and 0.75 for *CHL* alone.

B. Partial correlations

In a more rigorous attempt to analyze the explanatory power of *CHL*, we examine the ability of *CHL* to predict the effective spread benchmark, while the predictive power of other estimates is already taken into account. We do this by using the partitioned regression framework of equation (20):

$$ES_{i,t} = \alpha + \boldsymbol{\beta}\,\boldsymbol{EST}_{i,t} + \gamma CHL_{i,t} + \varepsilon_{i,t} \ , \tag{20}$$

where $ES_{i,t}$ represents the TAQ effective spread for stock i in month t, and $\boldsymbol{EST}_{i,t}$ is a vector of other estimates including *HL* and *Roll*. By using this setting, we are interested in calculating the partial correlation between $CHL_{i,t}$ and $ES_{i,t}$, while controlling for the other estimates. By using the Frisch-Waugh-Lovell theorem, we first regress $ES_{i,t}$ and $CHL_{i,t}$ on $\boldsymbol{EST_{i,t}}$ and then calculate the correlation between the orthogonal complements yielded from of above regressions, that is, $\epsilon_{ES_{i,t}|\boldsymbol{EST}_{i,t}}$, and $\epsilon_{CHL_{i,t}|\boldsymbol{EST}_{i,t}}$. To calculate time-series (cross-sectional) partial correlations, we perform the above regressions in the time-series (cross-sectional) dimension for every stock (month) and average the calculated partial correlations across stocks (months).

Table VIII shows the average partial correlations calculated while controlling for different set of estimates in the following order: we control for *HL* (third column titled "CHL|HL"), we add *Roll* (fourth column titled "CHL|HL, Roll"), and we move forward by adding other estimators to the set of controls (adding *Gibbs* in the fifth column, *EffTick* in the sixth column, and *FHT* in the last column). Panel A of Table VIII shows the average partial cross-sectional correlations testing whether they are different from zero by using Newey-West (1987) standard errors with four lags in the time-series of monthly-estimated cross-sectional correlations. All average cross-sectional correlations are significantly different from zero and positive, indicating that *CHL* has some additional explanatory power, not already included in any overidentified models, in predicting the effective spread. For instance, the average partial cross-sectional correlation of *CHL* and TAQ effective spreads after controlling for *HL*, *Roll*, *Gibbs*, *EffTick*, and *FHT* is 0.426 for the entire sample and 0.166, 0.401, and 0.444 for NYSE, AMEX, and NASDAQ stocks, respectively. Another interesting result is that the additional explanatory ability of *CHL* is larger for less liquid stocks as indicated by the increasing partial correlations from quintiles 1 to 5 in rows 8 to 12. All these findings remain consistent when average partial time-series correlations are considered (panel B of Table VIII).

C. Illiquidity and accuracy of estimates

The previous analysis suggests that the additional explanatory power of *CHL* is larger for less liquid stocks. To confirm that this additional explanatory ability is related to illiquidity rather than to volatility, we double sort the stocks by these two properties. First, we construct illiquidity tertiles by sorting the stocks by average effective spreads across the entire sample. Then we construct volatility tertiles within every illiquidity tertile by sorting stocks according to their daily price volatility across the entire sample. We then calculate average partial cross-sectional and time-series correlations with the TAQ effective spread benchmark for the nine groups controlling for the explanatory power of *HL* and *Roll*. Panel A (B) of Figure 5 shows the average partial cross-sectional (time-series) correlations for the nine groups. It delivers two main messages: First, correlations are considerably higher for the illiquid tertiles corroborating the previous findings. Second, there is no discernable pattern in terms of volatility within the three illiquidity tertiles, suggesting that illiquidity rather than volatility explains the additional explanatory power of *CHL*.

We substantiate this idea by performing a numerical simulation about the sensitivity of the *CHL* effective spread estimates to the number of trades per day. We repeat this analysis for the *HL* estimator, which is the nearest competitor, as documented in Section III. For this purpose, we simulate 10,000 months of 21 days, in which the stocks trade certain number of times per day. For simplicity, we assume that the intraday trades are equally distant from each other. We simply draw from $M_t = M_{t-1}e^{z\sigma/\sqrt{n}}$, $P_t = M_t e^{q_t s/2}$, $z \sim N(0,1)$, where M_t and P_t represent the efficient price and observed transaction price at time t, respectively. We set the daily standard deviation of efficient-price returns, σ, to be 3% and the bid-ask spread s to be constant at 1%. We perform the simulations for different number of trades per day, n, ranging from 3 to 390, and calculate the *CHL* and *HL* estimates in every attempt.[19]

Figure 6 shows two main findings: First, the *CHL* estimates are only marginally sensitive to the number of observed trades per day and this is essentially due to the two-day correction approach. On the other hand, the *HL* estimates are much more sensitive to the number of trades per day. In comparative terms, from 3 to 390 trades per day, the *HL* estimates range from 70 to 175 bps (instead of the 100-bps "true" proportional spread), whereas the C*HL* estimates remain in a narrow range from 129 to 132 bps. The steepness of the *HL* curve in Figure 6 illustrates the high sensitivity of the *HL* estimates to the number of trades per day. Perhaps a more important concern is the direction of this sensitivity, which entails that the *HL* estimates indicate a narrower spread when fewer transactions take place, contrary to common wisdom that the occurrence of fewer trades indicates more illiquid stocks or markets.[20]

To have a better sense of the actual number of trades per day, we look into the TAQ consolidated dataset and count how many regular trades occurred between 9:30 a.m. to 4:00 p.m. EST. We refer to the landmark of 100 trades represented by the dotted line in Figure 6. We find two revealing values. First, 25.6% of stock-days in our sample include less than 100 trades; second, 77.5% of stocks experienced at least one day with fewer than 100 trades. These numbers suggest that the *HL* estimates' sensitivity to the daily number of trades can be a broader issue that goes way beyond a limited number of illiquid stocks.

To provide empirical support to the numerical analysis above, we group stocks into five quintiles sorting them by their average number of daily trades during the sample period. Panel A of Table IX shows the correlation coefficients between different estimates and the TAQ effective spread benchmark

---

[19] When we run simulations for a large number of trades, we obtain estimates close to those for 390 trades.

[20] In the Internet Appendix, we provide a similar figure based on actual estimates. We sort the stock-month estimates into decile groups based on monthly averages of number of trades per day. We then calculate the bias of *CHL* (*HL*) estimates as the average difference between the estimates and the TAQ effective spread benchmark. This additional analysis confirms that the *HL* negative bias is substantially larger for less frequently traded stock-months.

for each quintile and for the entire sample (already seen in Table II). As expected, the *CHL* estimates have the highest correlation with the TAQ benchmark for the entire sample, as well as for the first three quintiles representing stocks less frequently traded.

By virtue of the low sensitivity of the *CHL* estimates to the number of trades per day, we expect that the *CHL* estimates generally explain market illiquidity better than the *HL* estimates do. To confirm this, we compute the correlations of spread estimates with the Amihud illiquidity measure, as a broader measure for market illiquidity. As observable in panel B of Table IX, the *CHL* estimates show the highest correlation coefficients for the entire sample and for the first four quintiles, that is, for all stocks, except for the ones with the highest number of trades.

In sum, the results in this section provide evidence that the *CHL* estimates have explanatory ability that goes beyond other estimators and are even more informative for less liquid stocks. Numerical simulations, followed by an empirical analysis, suggest that the *CHL* (*HL*) estimates are marginally (substantially) sensitive to the number of trades per day.

V. Example Applications

Well-performing estimators of transaction costs can be applied in a variety of research areas. To illustrate their potential uses, we propose two simple applications. The first example is a description of the historical spread estimates for stocks listed on NYSE (AMEX) from 1926 (1962) to 2014. In the second example, the spread estimates are applied to measure systematic risks originating from liquidity issues.

A. Estimating historical spreads for U.S. stocks

By using the close, high, and low price data from CRSP and the methodology explained above, we calculate the estimates of the bid-ask spreads based on our model. Specifically, we use the price values from previous days for the days with missing price values and construct the two-day corrected version of our estimates. We finally discard stock-months with fewer than 12 trading days.

Figure 7 shows the time development of the estimated spreads computed for three equally weighted portfolios: the smallest and largest market capitalization deciles, as well as the entire stocks sample. The spreads originated from our model display relatively stable variation over time. Reassuringly, this also applies to the smallest market capitalization decile. In contrast, Corwin and Schultz (2012) document that the spread estimates generated by their model display considerable variation over time, and these are extraordinarily high during the Great Depression, in which the market-wide average estimates of the effective spreads are as high as 20% for NYSE stocks and 50% for small cap stocks. Instead, panel A (B) of Figure 7 shows that our estimates for the NYSE (AMEX) stocks evolve pretty steadily across every decade, remaining within an economically reasonable range; that is, the market-wide estimated effective spread does not exceed 4% (6%) for NYSE (AMEX) stocks. Moreover, the average estimated effective spread for the small cap stocks listed on the NYSE (AMEX) does not exceed 12% (19%) during the whole sample.

The results in this subsection suggest that our estimator can be used in various research areas across many types of markets and assets, including the less liquid ones. This is especially true for researchers interested in the ability of an estimator to capture the temporal evolution of spreads over long time spans that predate intraday data or international markets without intraday data.

B. Estimating systematic liquidity risk for individual stocks

The results presented in Section III show that the spread estimates from our model closely follow the effective spread benchmark, suggesting that our estimator can be adopted for gauging transaction costs and liquidity. Another crucial application of spread estimates is liquidity *risk*. As liquidity risk is not diversifiable, its accurate measurement is crucial for at least two purposes: first, to identify and gauge systematic risk stemming from illiquidity issues, and, second, to perform effective asset and risk management. The collapse of Long-Term Capital Management L.P. (LTCM) and more recent experiences during the last financial crises are vivid examples of an incorrect consideration of liquidity risk.

Acharya and Pedersen (2005) propose a liquidity-adjusted capital asset pricing model (LCAPM) in which expected returns in time *t* for stock *i* (r_t^i) net of its transaction costs (s_t^i) are explained by the risk-free interest rate (r_f) and expected market returns (r_t^M) net of market transaction costs (s_t^M). Then the systemic risk of return net of trading costs is decomposed into four components:

$$E\left(r_{t+1}^i - s_{t+1}^i\right) = r_f + E\left(r_{t+1}^M - s_{t+1}^M - r_f\right)\frac{cov(r_{t+1}^i - s_{t+1}^i, r_{t+1}^M - s_{t+1}^M)}{var(r_{t+1}^M - s_{t+1}^M)}, \quad (21)$$

$$E\left(r_{t+1}^i\right) = r_f + E\left(s_{t+1}^i\right) + \lambda_t(\beta_1 + \beta_2 - \beta_3 - \beta_4), \quad (22)$$

$$\beta_1 = \frac{cov(r_{t+1}^i, r_{t+1}^M)}{var(r_{t+1}^M - s_{t+1}^M)},$$

$$\beta_2 = \frac{cov(s_{t+1}^i, s_{t+1}^M)}{var(r_{t+1}^M - s_{t+1}^M)},$$

$$\beta_3 = \frac{cov(r_{t+1}^i, s_{t+1}^M)}{var(r_{t+1}^M - s_{t+1}^M)},$$

$$\beta_4 = \frac{cov(s_{t+1}^i, r_{t+1}^M)}{var(r_{t+1}^M - s_{t+1}^M)}.$$

While β_1 represents the standard market beta, β_2, β_3, and β_4 capture important aspects of systematic risk due to liquidity issues. β_2 measures the commonality of liquidity with the market-wide liquidity and is expected to be positive (Chordia, Roll, and Subrahmanyam (2000)). Higher β_2 translates into less liquid stocks in times of market illiquidity. Huberman and Halka (2001) and Hasbrouck and Seppi (2001) document the presence of a systematic, time-varying component of liquidity that co-moves with individual stock's liquidity. Kamara, Lou, and Sadka (2008) show important implications of the cross-sectional variation of commonality in liquidity, including the decline over time of diversification benefits against aggregate liquidity shocks by holding large-cap stocks. β_3 is typically negative as market liquidity tends to dry up when stock prices decline. Pastor and Stambaugh (2003) show that investors demand a premium for the sensitivity of stock returns to aggregate liquidity shocks. Watanabe and Watanabe

(2008) document that aggregate liquidity is priced and the liquidity risk premium is twice as high as the value premium in high-beta states. β_4 is also expected to be negative as the liquidity of individual stocks tend to decrease in downturn markets. Hameed, Kang, and Viswanathan (2010) provide empirical evidence of significant increases of bid-ask spreads when the stock market experiences large negative returns.

The above-mentioned literature points to the importance of an accurate measurement of different dimensions of liquidity risk and its variation in the cross section of stocks. By using the effective spread estimates of Section III, we calculate the four systematic risk components of equation (22) for each stock in our sample based on the Daily TAQ effective spreads, as well as the *Roll* estimates, the *HL* estimates, and our estimates. In addition to the filtrations explained in Section III, we discard stocks with fewer than 30 months of data and the stock-months in which the monthly CRSP return is missing. Following Asparouhova, Bessembinder, and Kalcheva (2010, 2013), we use a gross return-weighted portfolio of all the stocks to construct the market return and market liquidity to avoid biases calculating portfolio returns. To assess the quality of the estimates for systematic risk, we compare them to those based on the TAQ effective spreads. In other words, we gauge how close the cross-sectional variation of liquidity risks generated by the *Roll*, the *HL*, and our model are to those obtained from Daily TAQ data. To be consistent with the previous tables, we report the results in the form of correlation coefficients. In fact, the square of the correlation coefficients reported in Table X are R^2s of the regression framework shown in equation (23), which also shows the proportion of cross-sectional variation in the systematic liquidity risks that can be explained by the estimates.

$$\beta_{k,i}^{Estimated} = \delta_k + \gamma_k \beta_{k,i}^{ES} + \varepsilon_{k,i} \ , k = 1,..,4 \, , i = 1,..,N. \qquad (23)$$

$\beta_{k,i}^{Estimated}$ represents the LCAPM systematic risk component *k* for stock *i,* computed using the *Roll* estimates, the *HL* estimates, or our estimator, while $\beta_{k,i}^{ES}$ represents the same factor computed with TAQ effective spreads. We then compare the correlation between the liquidity risk estimates from our model and from the TAQ data to that between the other estimates and TAQ data. Table X shows the main results of this analysis for β_2, β_3, and β_4. Because the correlations for $\beta_1$are close to one as a result of the secondary importance of transaction costs to compute standard betas, they are not tabulated. Thus, we concentrate our analysis on β_2, β_3, and β_4 because they are more influenced by transaction costs. Overall, the results for all stocks, shown in panel A of Table X, clearly indicate the superiority of our estimator to capture the cross-sectional estimation of systematic liquidity risks. The correlation between *CHL* estimates and the benchmark spreads for β_2, β_3, and β_4 are pretty high, that is, 0.829, 0.937, and 0.761, respectively. The same correlation coefficients for the *HL* and the *Roll* estimators are lower, especially for β_2 and β_4, which are around 0.1 lower for the *HL* and around 0.2 lower for the *Roll*. In many of the cases, these differences are statistically significant using a two-tailed Fisher's z-test with a 5% significance level. The same picture generally holds when we perform the subsampling analysis across the 2003–2007, 2008–2011, and 2012–2014 subperiods.

Following Acharya and Pedersen (2005), we analyze liquidity innovations generated from an AR(2) model. The analysis of liquidity in innovations, rather than by levels, helps us control for the persistence in the transaction cost process, thereby capturing the unexpected component of transaction costs. The results in panel B of Table X confirm the high accuracy of *CHL* estimates to gauge systematic liquidity risks using spread innovations. The correlation coefficients between the estimates of β_2, β_3, and β_4 from our model and the TAQ spreads are 0.511, 0.860, and 0.535, while the same correlations for *HL* estimates are 0.439, 0.879, and 0.258, and those for the *Roll* estimates are 0.152, 0.588, and 0.251, respectively. The subsampling analysis across shorter periods delivers consistent results, confirming that *CHL* estimates provide systematic risk estimates that behave more similarly to the TAQ benchmark, no matter if transaction costs are in levels or innovations.

As in Sections III and IV, we reiterate the subsampling analysis across primary exchange, market capitalization, and effective spread size (panels C, D, and E).[21] Overall, our estimator outperforms the

[21] To facilitate comparisons, we use the same quintile groups like in Section III. However, here we remove a few more stocks that have fewer than 30 months of data.

other measures when stocks are grouped by market venues, market capitalization, and spread size. The only exceptions are β_4 for the NYSE and the largest capitalization quintile, for which the correlation between *HL* estimator and TAQ benchmark is higher, but the difference is not statistically significant. When stocks are subsampled from smallest to largest transaction costs, our estimator (the *HL* estimator) performs better across less (more) liquid stocks.

C. Estimating systematic liquidity risk for portfolios of stocks

A common approach in the empirical asset-pricing literature is to use portfolios rather than individual stocks. The final question we address is whether transaction cost estimates can be reliably used in applying this approach. As in Acharya and Pedersen (2005), we construct 25 portfolios sorted by illiquidity level to fit the LCAPM model. We create 25 portfolios sorted by the TAQ effective spread benchmark and the bid-ask spread estimates from the *CHL*, *HL*, and *Roll* models. Then we measure how close the LCAPM systematic risk betas of the estimated portfolio spreads follow the betas of the TAQ benchmark. More precisely, we apply the following procedure: First, we sort stocks using the TAQ effective spreads of previous month and construct 25 gross-return-weighted portfolios using gross returns of the previous month as weights. Second, we calculate betas using the TAQ effective spread benchmark. Third, we repeat the same steps of sorting stocks, constructing portfolios, and calculating betas for the bid-ask spread estimates of *CHL*, *HL*, and *Roll*. Finally, we apply the previous framework explained in equation (23) to assess the overall quality of the estimates.

Table XI reports the main results. As the estimates are also used in the portfolio construction, β_1 of the estimated portfolios might differ from the TAQ benchmark ones. Therefore, Table XI includes all four betas. The row labeled "levels" reports betas calculated with effective spreads in levels. The overall result is that for every β, *CHL* portfolio betas more accurately follow the TAQ ones. Compared with the single-stock analysis, the correlations for β_1 are no longer close to one, ranging from 0.238 for *Roll*, 0.621 for *HL*, and to 0.693 for *CHL*. While the correlations for β_2 are pretty high for the three estimation models, the differences are quite large for β_3 and β_4. For β_3, the highest correlation is 0.352 for the *CHL* estimates, while the correlation between β_3 for *HL* estimates and that of that of the TAQ benchmark is negative (-0.473). For β_4, correlations span from a relatively low 0.155 for *Roll* to 0.901 for *CHL* estimates. In general, these results point to the importance of using accurate estimates in sorting stocks and call for some caution in the selection of estimation models to construct portfolios.

We repeat our analysis of betas for liquidity shocks of an AR(2) model fitted to portfolio effective spreads. As discernable in the row labeled "shocks" of Table XI, the results of liquidity shocks are in line with those of liquidity levels. The *CHL* estimates provide very high correlations (β_2, β_3, and β_4, are 0.966, 0.952, and 0.957, respectively), which are the highest across the estimation models, except for β_2. Now, the *HL* correlation for β_3 is closer to the *CHL* correlation.

In sum, this simple asset-pricing analysis shows that our estimator provides accurate estimations of LCAPM systematic risks, commonality in liquidity, and covariation between stock returns and liquidity using sorted portfolios and individual stocks.

VI. Conclusion

Building on the seminal model proposed by Roll (1984), we derive a new way to estimate bid-ask spreads. Compared with the *Roll* measure, our model has two important benefits: First, it takes advantage of a richer information set of daily close, high, and low prices, whereas the *Roll* measure solely relies on the close prices. Thereby, our model improves estimation accuracy. From the high and low prices, we can compute the mid-range, that is, the mean of the daily high and low log-prices, that proxies the efficient price. Second, our estimator is fully independent of order-flow dynamics, and therefore it does not rely on bid-ask bounces, as the original *Roll* measure does. Our method of estimating effective spreads is straightforward, easy to compute, and has an intuitive closed-form solution that resembles the

Roll measure. While the *Roll* measure relies on the covariance of consecutive close-to-close price returns, our estimator relies on the covariance of close-to-mid-range returns around the same close price.

We numerically and empirically test our method by using Daily Trade and Quotes (Millisecond TAQ) data. The simulation analysis shows that considering all imperfections together (that is, infrequent trading, inconstant spreads, and nontrading periods), our model provides more accurate estimates than those from the high-low estimator proposed by Corwin and Schultz (2012) and the *Roll* model for less liquid securities, for which transaction costs and liquidity issues are of much more concern. In the empirical analysis, the effective spread computed with Daily TAQ data serves as the benchmark for our comparative considerations. Assessed against other daily estimates, our estimator generally provides the highest cross-sectional and average time-series correlation with the TAQ effective spread benchmark, as well as the smallest prediction errors.

We also document the additional explanatory ability of our estimates that systematically goes beyond that of other estimates. This additional predictive ability is especially larger for less liquid stocks. The numerical and empirical analyses suggest that our estimates are stable and much less sensitive to the number of trades per day, whereas the Corwin and Schultz (2012) high-low estimates produce substantially smaller spread estimates for lower number of trades per day , that is, for more illiquid stocks. The ability of our estimator to provide much more accurate spread estimates for less liquid stocks is a suitable characteristic because accurate estimates of transaction costs are particularly needed for less liquid securities and markets.

To illustrate some potential applications, we reconstruct the historical development of our spread estimates for stocks listed on NYSE (AMEX) from 1926 (1962) through 2014. These patterns display relatively stable variation over time and remain within an economically meaningful range, even for small-cap stocks. Then we estimate the components of systematic liquidity risk like in the liquidity-adjusted capital asset pricing model (LCAPM), which was postulated by Acharya and Pedersen (2005). The overall result is that our estimator provides very accurate estimations of the systematic liquidity risks for individual stocks and sorted portfolios, in the sense that systematic risk betas based on our estimates are close to those of the TAQ benchmark and that our model generally outperforms other models in estimating systematic risk originating from commonality in liquidity and covariation between stock returns and illiquidity.

Our estimator has many potential applications for future research. It should be useful for researchers who work in asset pricing, corporate finance, risk management, and other important research areas and need a simple but accurate measure of trading costs over long periods. Our model could be suitably applied to many securities, including emerging markets or over-the-counter markets, and not only U.S. stocks, for which data are of limited quality or availability.

References

Acharya, Viral V., and Lasse H. Pedersen, 2005, Asset pricing with liquidity risk, Journal of Financial Economics 77, 375–410.

Amihud, Yakov, 2002, Illiquidity and stock returns: Cross section and time-series effects, Journal of Financial Markets 5, 31–56.

Amihud, Yakov, and Haim Mendelson, 1986, Asset pricing and the bid-ask spread, Journal of Financial Economics 17, 223–249.

Asparouhova, Elena N., Hendrik Bessembinder, and Ivalina Kalcheva, 2010, Liquidity biases in asset pricing tests, Journal of Financial Economics 96, 215–237.

Asparouhova, Elena N., Hendrik Bessembinder, and Ivalina Kalcheva, 2013, Noisy prices and inference regarding returns, Journal of Finance 68, 665–714.

Choi, J.Y., Dan Salandro, and Kuldeep Shastri, 1988, On the estimation of bid-ask spread: Theory and evidence, Journal of Financial and Quantitative Analysis 23, 219–229.

Chordia, Tarun, Richard Roll, and Avanidhar Subrahmanyam, 2000, Commonality in liquidity. Journal of Financial Economics 56, 3–28.

Chung, Kee H., and Hao Zhang, 2014, A simple approximation of intraday spreads using daily data, Journal of Financial Markets 17, 94–120.

Corwin, Shane A., and Paul Schultz, 2012, A simple way to estimate bid-ask spreads from daily high and low prices, Journal of Finance 67, 719–759.

Fong, Kingsley Y. L., Craig W. Holden, and Charles Trzcinka, 2014, What are the best liquidity proxies for global research?, Working Paper, University of New South Wales, Indiana University.

Garman, Mark B., and Michael J. Klass, 1980, On the estimation of security price volatilities from historical data, Journal of Business 53, 67–78.

Goyenko, Ruslan Y., Craig W. Holden, and Charles A. Trzcinka, 2009, Do liquidity measures measure liquidity?, Journal of Financial Economics 92, 153–181.

Hameed, Allaudeen, Wenjin. Kang, and S. Viswanathan, 2010, Stock market declines and liquidity, Journal of Finance 65, 257–93.

Hasbrouck, Joel, 2004, Liquidity in the futures pits: Inferring market dynamics from incomplete data, Journal of Financial and Quantitative Analysis 39, 305–326.

Hasbrouck, Joel, 2009, Trading costs and returns for US equities: the evidence from daily data, Journal of Finance 64, 1445–1477.

Hasbrouck, Joel, and Duane J. Seppi, 2001, Common factors in prices, order flows, and liquidity, Journal of Financial Economics 59, 383–411.

Holden, Craig W., 2009, New low-frequency liquidity measures, Journal of Financial Markets 12, 778-813.

Holden, Craig W., and Stacey Jacobsen, 2014, Liquidity measurement problems in fast, competitive markets: expensive and cheap solutions, Journal of Finance 69, 1747–1785.

Holden, Craig W., Stacey Jacobsen, and Avanidhar Subrahmanyam, 2014, The empirical analysis of liquidity, Foundations and Trends in Finance 8, 263–365.

Huberman, Gur, and Dominika Halka, 2001, Systematic liquidity. Journal of Financial Research 24, 161–178.

Kamara, Avraham, Xiaoxia Lou, and Ronnie Sadka, 2008, The divergence of liquidity commonality in the cross-section of stocks, Journal of Financial Economics 89, 444–466.

Lesmond, David A., Joseph P. Ogden, and Charles A. Trzcinka, 1999, A new estimate of transaction costs, Review of Financial Studies 12, 1113–1141.

Newey, Whitney K., and Kenneth D. West, 1987, A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix, Econometrica 55, 703–708.

Pastor, Lubos, and Robert F. Stambaugh, 2003, Liquidity risk and expected stock returns, Journal of Political Economy 111, 642–685.

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

Stoll, Hans, and Robert E. Whaley, 1983, Transaction costs and the small firm effects, Journal of Financial Economics 12, 57–79.

Watanabe, Akiko, and Masahiro Watanabe, 2008, Time-varying liquidity risk and the cross section of stock returns, Review of Financial Studies 21, 2449–2486.

Appendix A. Proof of Propositions 2 and 3

We first derive two propositions A1 and A2 that we need for the proofs.

Proposition A1. Under the model assumptions, equation (A1) holds:

$$E[(h_t^e - c_t^e)(c_t^e - l_t^e)] = (2\log(2) - 1)\sigma_e^2. \quad \text{(A1)}$$

Proof of proposition A1: to prove Proposition A1, we use the two following equations from Garman and Klass (1980):

$$E[(h_t^e - l_t^e)^2] = 4\log(2)\,\sigma_e^2, \quad \text{(A2)}$$

$$E[(h_t^e - c_t^e)^2] = E[(l_t^e - c_t^e)^2] = \sigma_e^2. \quad \text{(A3)}$$

Plugging (A2) and (A3) into (A1) leads to the proof

$$E[(h_t^e - c_t^e)(c_t^e - l_t^e)] = \frac{1}{2}E[(h_t^e - c_t^e + c_t^e - l_t^e)^2 - (h_t^e - c_t^e)^2 - (c_t^e - l_t^e)^2] = (2\log(2) - 1)\,\sigma_e^2. \quad \text{(A4)}$$

Proposition A2. Under the model assumptions, equation (A5) holds:

$$E[(c_t^e - h_t^e)^2] = E[(c_t^e - h_{t+1}^e)^2]. \quad \text{(A5)}$$

Proposition A2 is the result of the symmetry of Brownian motion in forward-looking and backward-looking expressions. More specifically, the distance between the efficient close price of day t and the efficient high (low) price of the same day is equal to the distance between the efficient open price and the efficient high (low) price of the next day:

$$E[(c_t^e - h_t^e)^2] = E[(o_{t+1}^e - h_{t+1}^e)^2]. \quad \text{(A6)}$$

Proof of Proposition A2: We assume no overnight price movements, so the efficient close price of day t and the efficient open price of day $t + 1$ are identical, and, therefore, replacing o_{t+1}^e with c_t^e leads to the proof.

Proof of Proposition 2

Now we use the two propositions for the proof of Proposition 2 of the paper. The stepwise proof is as follows:

$$E[(c_t - (\eta_t + \eta_{t+1})/2)^2] = E[(c_t^e + q_t s/2 - \eta_t/2 \ - \eta_{t+1}/2)^2] \quad \text{(A7)}$$

$$= E[q_t^2 s^2/4 + 1/4(c_t^e - \eta_t)^2 + 1/4(c_t^e - \eta_{t+1})^2 + 1/2(c_t^e - \eta_{t+1})(c_t^e - \eta_t) + qs/4\,(c_t^e - \eta_{t+1}) + qs/4\,(c_t^e - \eta_t)]\ , \quad \text{(A8)}$$

$$= s^2/4 + 1/2E[(c_t^e - h_t^e/2 - l_t^e/2)^2]\ , \quad \text{(A9)}$$

$$= s^2/4 + (1/2)E[(1/4)(c_t^e - h_t^e)^2 + (1/4)(c_t^e - l_t^e)^2 + (1/2)(c_t^e - l_t^e)(c_t^e - h_t^e)]\ , \quad \text{(A10)}$$

$$= s^2/4 + \sigma_e^2/4 - (\log(2)/2 - 1/4)\sigma_e^2 = s^2/4 + (1/2 - \log(2)/2)\ \sigma_e^2\ . \quad \text{(A11)}$$

Equation (A7) is the result of the definition of the Roll (1984) model. Equation (A9) is the result of Proposition A2, and, finally, we derive equation (A11) using Proposition A1.

Proof of Proposition 3

The proof for Proposition 3 of the paper is similar to that of Proposition 2:

$$E[(\eta_{t+1} - \eta_t)^2] = E[(\eta_{t+1} - c_t^e + c_t^e - \eta_t)^2], \quad \text{(A12)}$$

$$= 2E[(c_t^e - \eta_t)^2] = 2E[(c_t^e - h_t^e/2 - l_t^e/2)^2], \quad \text{(A13)}$$

$$= 2E[(1/4)(c_t^e - h_t^e)^2 + (1/4)(c_t^e - l_t^e)^2 + (1/2)(c_t^e - l_t^e)(c_t^e - h_t^e)]\ , \quad \text{(A14)}$$

$$= (2 - 2\log(2))\sigma_e^2. \quad \text{(A15)}$$

Equation (A13) is the result of Proposition A2, and, finally, we derive equation (A15) using Proposition A1.

Appendix B. Proof of robustness to nontrading periods

To prove the robustness of our estimator to nontrading periods, we repeat the logical steps followed in the paper by including the nontrading period in the efficient price variance. We then show that this term cancels out when we derive the outcome expression.

Definition B1. The nontrading period (e.g., overnight) efficient-price variance is defined as follows:

$$\sigma^2_{Nontrading} = E[(o^e_{t+1} - c^e_t)^2]. \tag{B1}$$

Proposition B1. If we consider a price movement during nontrading periods with the variance of $\sigma^2_{Nontrading}$, equation (B2) holds:

$$E[(c_t - (\eta_t + \eta_{t+1})/2)^2] = s^2/4 + (1/2 - \log(2)/2)\sigma^2_e + 1/4\ \sigma^2_{Nontrading}. \tag{B2}$$

Proof of Proposition B1: The proof is similar to the proof of Proposition 2, which is explained in Appendix A. The only difference arises because the distance between efficient close price of day t and the efficient high (low) price of day $t+1$ is higher than the distance between efficient close price of day t and the efficient high (low) price at the same day. Therefore, equation (A5) no longer holds, and, instead, equation (B3) shows the link between the two quantities. Using equation (B3) and following the steps of the proof in in Appendix A leads to the proof of Proposition B1.

$$E[(c^e_t - h^e_{t+1})^2] = E[(c^e_t - o^e_{t+1} + o^e_{t+1} - h^e_t)^2] = \sigma^2_{Nontrading} + E[(o^e_{t+1} - h^e_t)^2] = \sigma^2_{Nontrading} + E[(c^e_t - h^e_t)^2]. \tag{B3}$$

Proposition B2. If we consider a price movement during nontrading periods (e.g., overnight) with the variance of $\sigma^2_{Nontrading}$, equation (B4) holds:

$$E[(\eta_{t+1} - \eta_t)^2] = (2 - 2\log(2))\sigma^2_e + \sigma^2_{Nontrading}. \tag{B4}$$

Proof of Proposition B2: The proof is very similar to the proof of Proposition B1.

Proof of robustness to nontrading periods

When calculating s^2 using the two equations of proposition B1 and B2, the nontrading variance terms cancel out, and the result is identical to equation (9):

$$s^2 = 4\,E[(c_t - \eta_t)(c_t - \eta_{t+1})]. \tag{B5}$$

Therefore, the spread estimates are independent of price movements during nontrading periods.

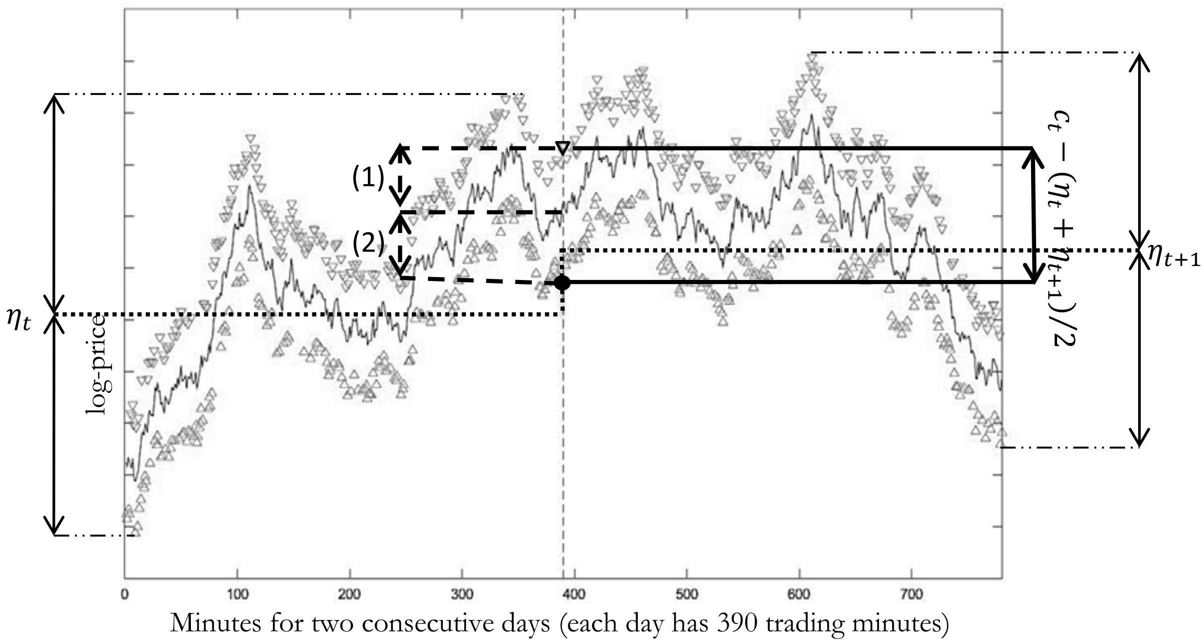


Figure 1. The Schematic Decomposition of the Distance between Closing Price and Average Mid-ranges

The log-price process is simulated with minute increments for the duration of two days of working hours. Each working day consists of 390 minutes. The figure provides a simple illustration that the distance between c_t and $(\eta_t + \eta_{t+1})/2$ can be decomposed into two compnents: (1) the distance between close price and efficient close price, that is, the effective spread, and (2) the distance between efficient close price and the midquote proxy.

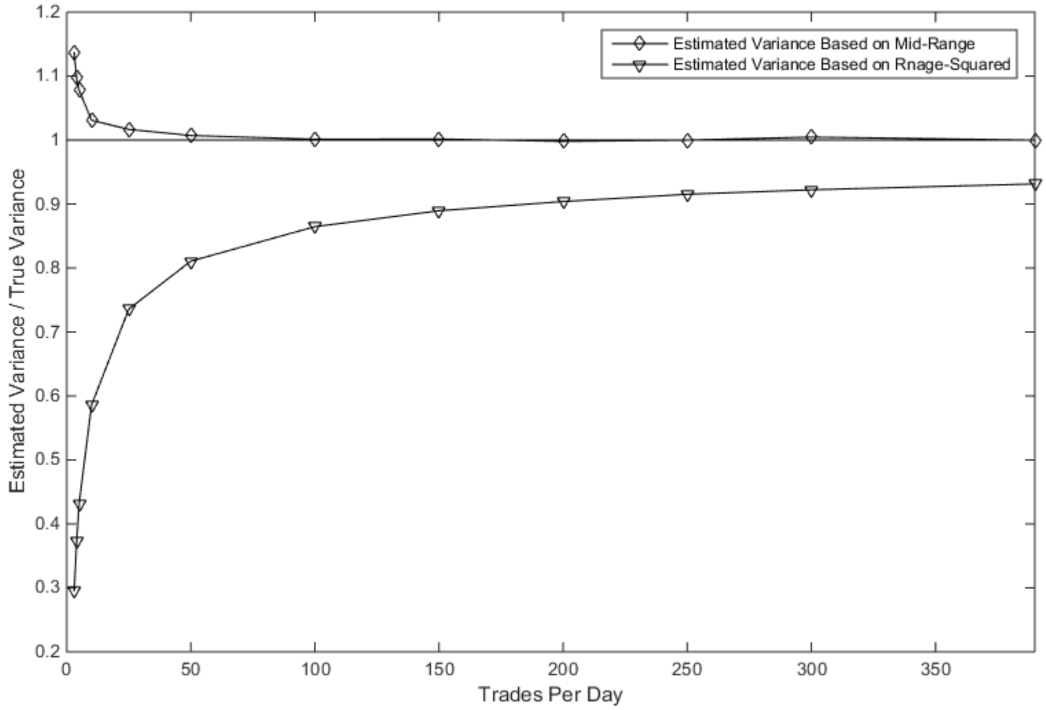


Figure 2. Estimated Variance from a Simulated Discrete Random Walk

For every case with a certain number of trades per day, ranging from 3 to 390, simulations of 200,000 days with pre-assigned daily volatility of 3% are performed. The variance based on the mid-range is calculated as $\sigma_\eta^2 = 1/(2 - 2\log(2))\, E[(\eta_t - \eta_{t-1})^2]$, and the range-based variance is calculated as $\sigma_{h-l}^2 = 1/(4\log(2)) E[(h_t - l_t)^2]$. Expected values are estimated by using the means of a sample of 200,000 simulations. The estimation outputs are divided by the pre-assigned variance of 0.03^2 in order to be comparable with 1.

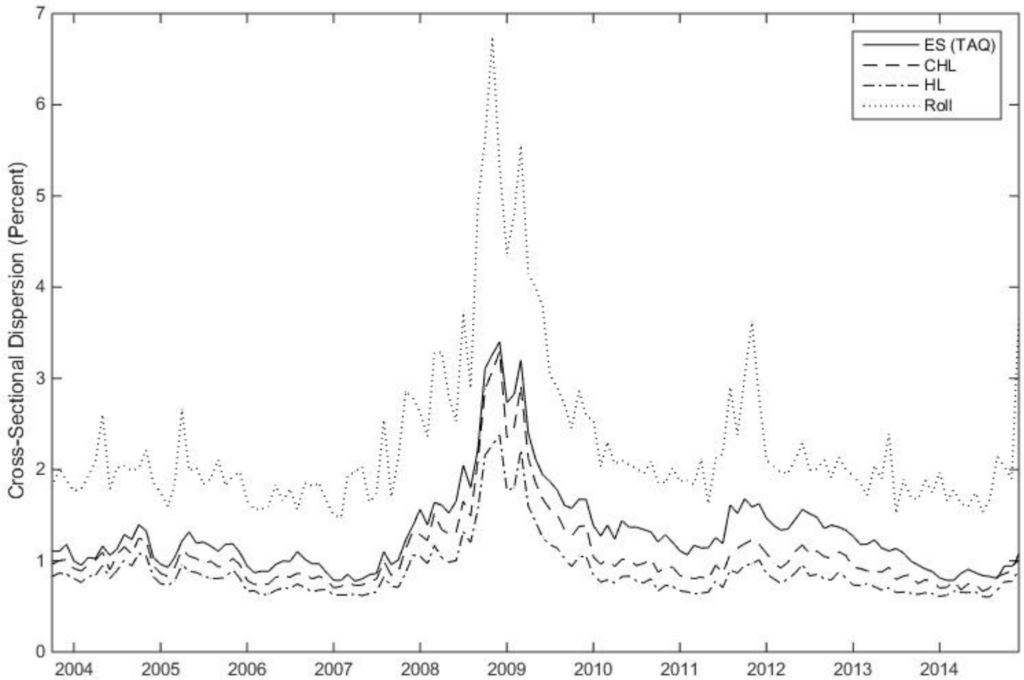
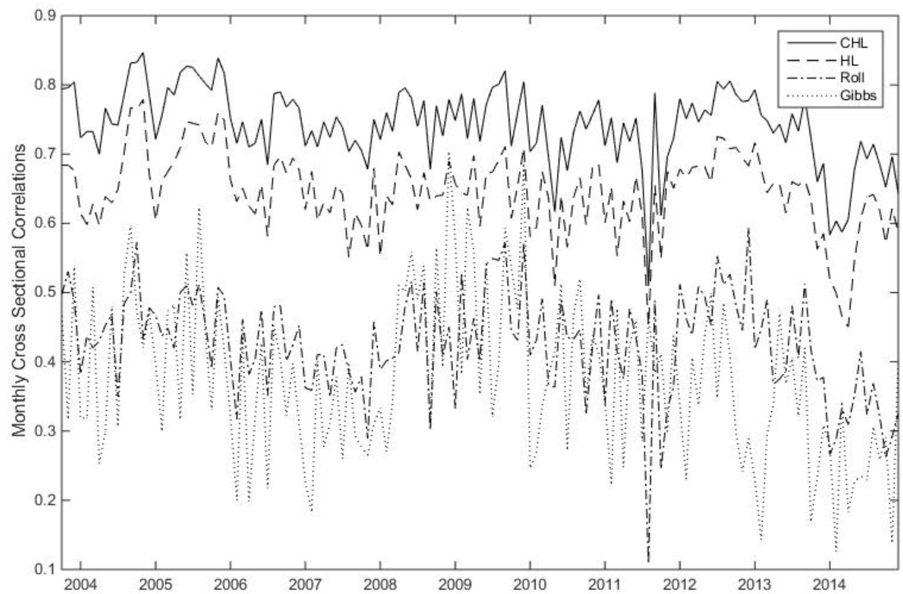


Figure 3. Cross-sectional Dispersion of Monthly Spread Estimates

This figure shows the standard deviations of spread estimates across stocks for each month from October 2003 to December 2014. In addition to the effective spread based on the Daily TAQ data, the labels refer to our estimator (CHL) and the estimators proposed by Corwin and Schultz (HL; 2012) and Roll (Roll; 1984).

Panel A. Correlations Based on Percentage Spreads



Panel B. Correlations Based on First Difference in Percentage Spreads

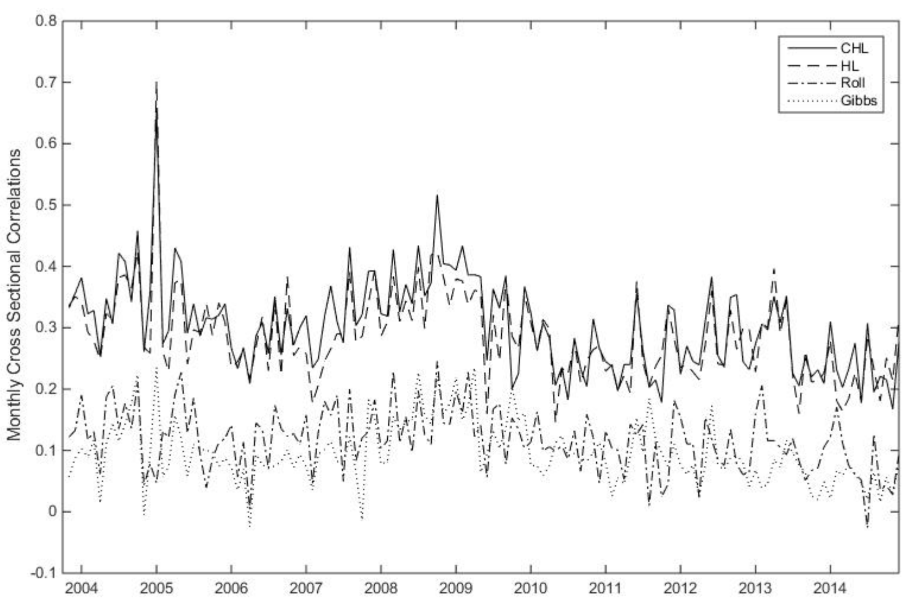
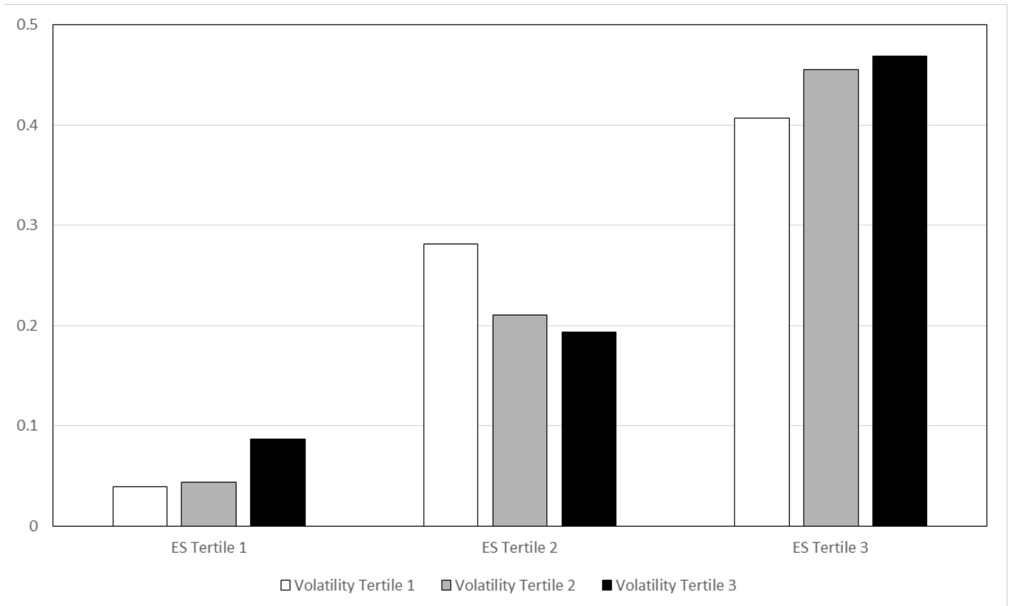


Figure 4. Cross-sectional Correlation of Monthly Spread Estimates

This figure shows the cross-sectional correlation between model-implied spread estimates and effective spreads from the Daily TAQ data for each month from October 2003 to December 2014. The labels refer to our model (CHL) and the models proposed by Corwin and Schultz (HL; 2012), Roll (Roll; 1984), and Hasbrouck (Gibbs; 2009). Panel A (B) displays correlations for levels (changes).

Panel A. Average Partial Cross-Sectional Correlations



Panel B. Average Partial Time-Series Correlations

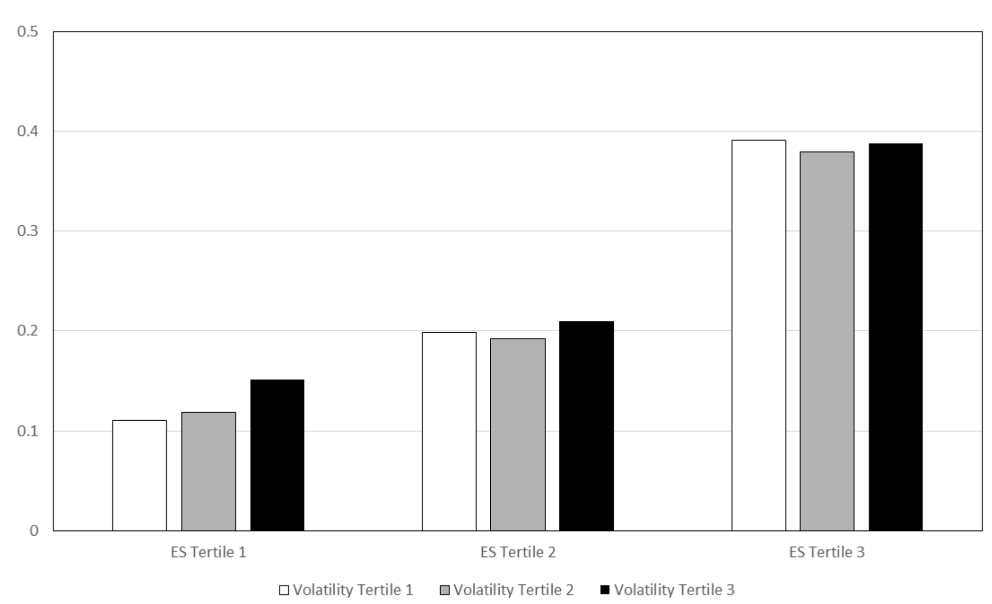


Figure 5. Average Partial Correlations after Controlling for HL and Roll
We split the stocks sample into three illiquidity tertiles by sorting them with their average effective spread during the sample period. Then we break down each illiquidity tertile into three volatility tertiles using the daily volatility of the stocks during the sample period. The partial correlations are the correlations between the residuals of regressing TAQ effective spreads and our estimates (CHL) on Corwin and Schultz's (HL; 2012) and Roll's (Roll; 1984) estimates.

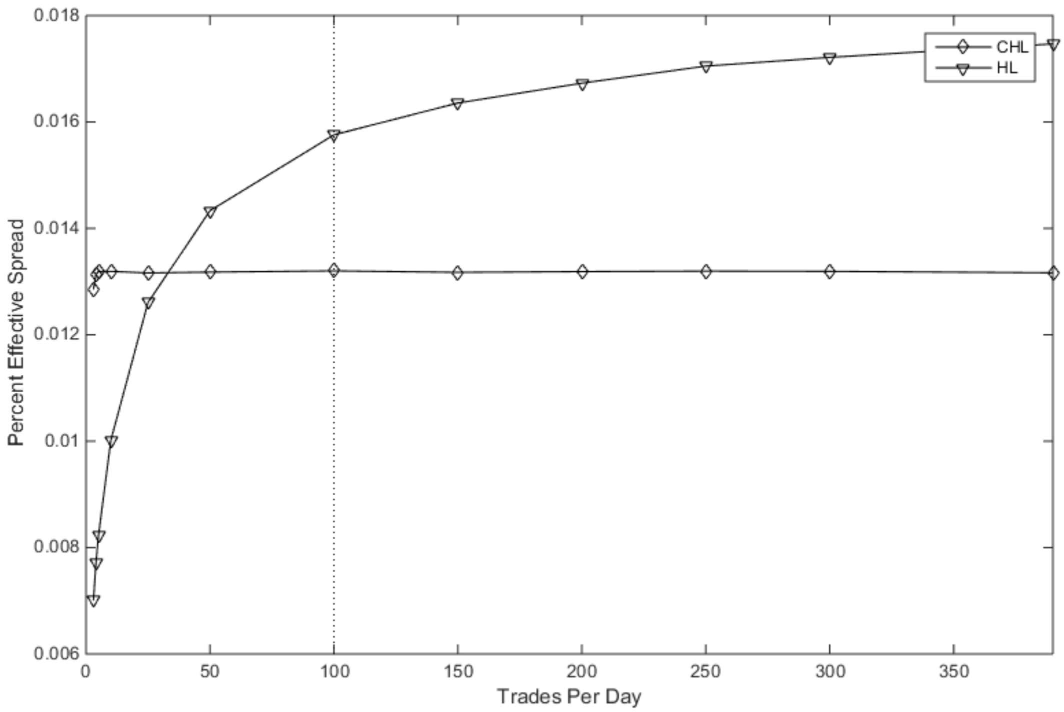
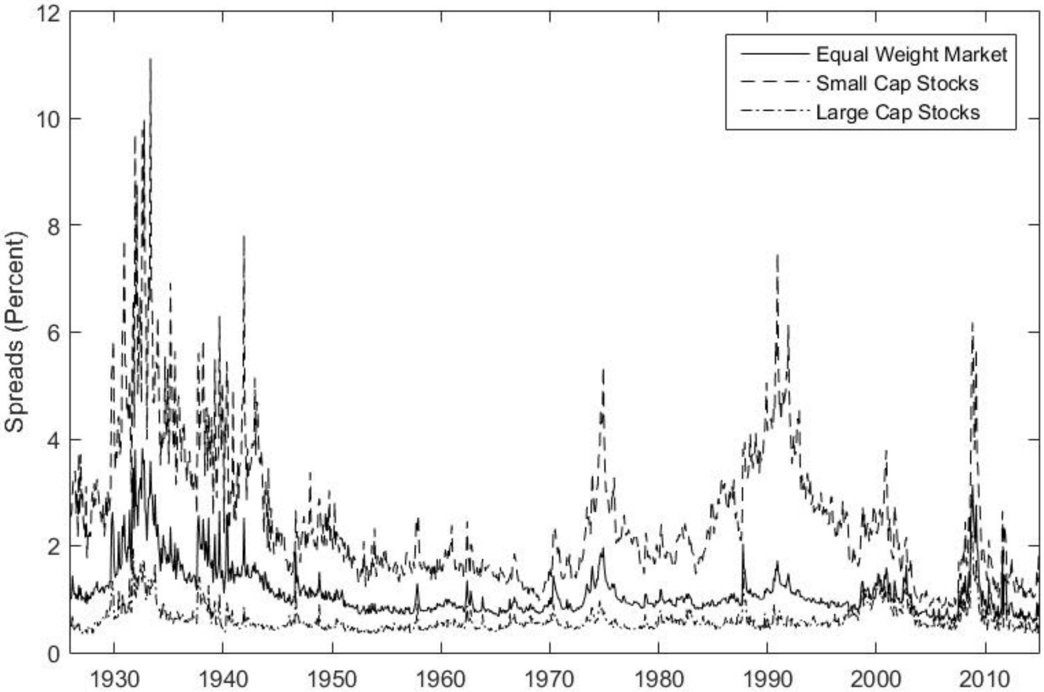


Figure 6. Sensitivity of Estimates to the Number of Daily Trades

The horizontal axis refers to the equally spaced number of trades per day, ranging from 3 to 390, in the simulations of 10,000 months with 21 days. For each case, we simulate the trajectory of observed prices at the times of trades following a geometric Brownian motion, with a daily volatility of 3%, and constant relative spread of 1%. The labels in the legend refer to the estimators from our model (CHL) and Corwin and Schultz model (HL; 2012).

Panel A. NYSE Listed Stocks



Panel B. AMEX Stocks

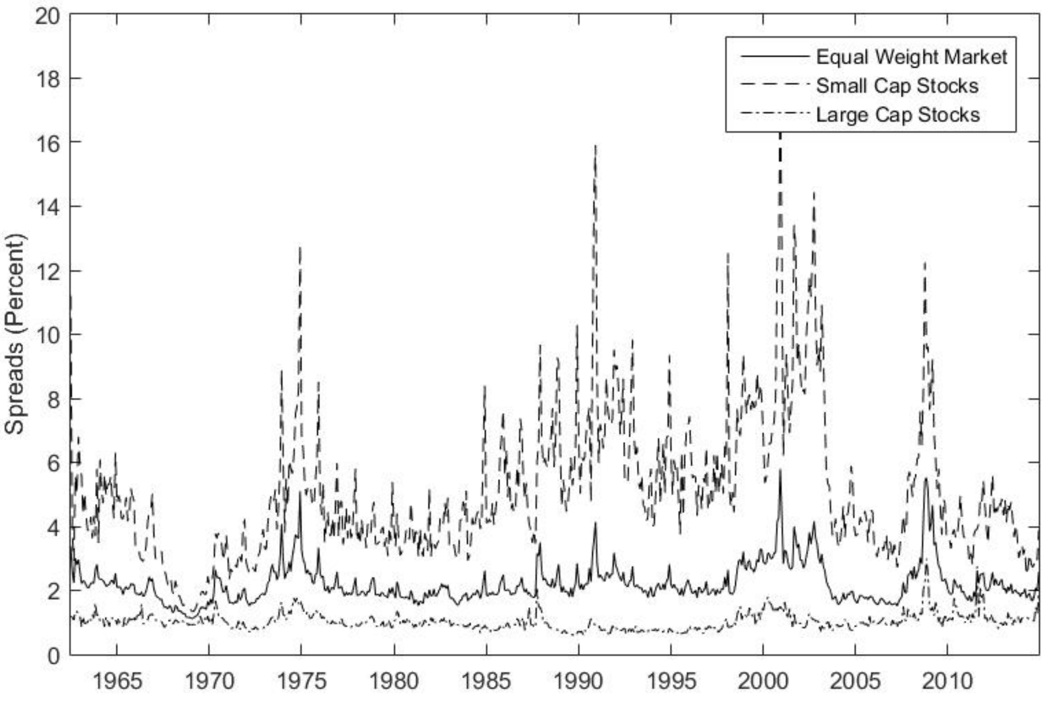


Figure 7. Time-Series Evolution of Estimated Spread, Calculated as Equally Weighted Portfolios of Stocks
This figure shows the monthly historical developments of spread estimates from our model. Small cap and large cap portfolios are represented by the first and last decile of stocks sorted by market capitalization at the end of each month. Panel A (B) shows the estimates for stocks listed on the NYSE (AMEX) between 1926 (1962) and 2014**.**

Table I. Estimated Bid-Ask Spreads from Simulations

Each simulation consists of 10,000 21-day months of stock prices, and each day consists of 390 minutes. For each minute, the trajectory of a geometric Brownian motion with daily volatility of 3% and a constant relative spread with the values mentioned in the table is simulated. The labels in the first row refer to the estimators from the following models: ours (CHL), Corwin and Schultz's (HL; 2012), and Roll's (Roll; 1984). Two-day and monthly refer to the two-day corrected and monthly corrected versions, in which two-day or monthly negative estimates are set to zero. We run the simulations in five separate scenarios. Panel A shows the results in the near-ideal situation. Panel B shows the results when trades in each minute are observable with only a 10% chance. Panel C shows the results when the spreads of each day are uniformly distributed between zero and twice the nominal average value. Panel D shows the results when there are "overnight" price movements and the standard deviation of the overnight price change is 50% of the standard deviation of the daily price change. The overnight adjustment procedure for *HL* estimates is exactly as in Corwin and Schultz (2012). Panel E encompasses the "imperfections" in scenarios B, C, and D at the same time.

		CHL		HL		Roll
		Two-Day	Monthly	Two-Day	Monthly	
Panel A: Near-Ideal Conditions						
0.5% Spread	Mean	1.2%	0.7%	1.4%	0.6%	1.2%
	σ	0.4%	0.8%	0.3%	0.5%	1.4%
1.0% Spread	Mean	1.3%	0.9%	1.7%	1.0%	1.3%
	σ	0.4%	0.8%	0.4%	0.6%	1.4%
3.0% Spread	Mean	2.4%	2.9%	3.2%	2.9%	2.6%
	σ	0.5%	0.7%	0.5%	0.6%	1.8%
5.0% Spread	Mean	4.3%	5.0%	5.0%	4.9%	4.6%
	σ	0.6%	0.6%	0.6%	0.6%	2.2%
8.0% Spread	Mean	7.6%	8.0%	7.8%	7.8%	7.6%
	σ	0.6%	0.5%	0.6%	0.6%	2.6%
Panel B: Each Trade Is Visible with a Chance of 10%						
0.5% Spread	Mean	1.2%	0.7%	1.0%	0.2%	1.2%
	σ	0.4%	0.8%	0.3%	0.3%	1.4%
1.0% Spread	Mean	1.3%	1.0%	1.3%	0.4%	1.3%
	σ	0.4%	0.8%	0.3%	0.5%	1.4%
3.0% Spread	Mean	2.4%	2.9%	2.6%	2.1%	2.6%
	σ	0.5%	0.7%	0.5%	0.6%	1.8%
5.0% Spread	Mean	4.3%	4.9%	4.2%	4.0%	4.6%
	σ	0.6%	0.6%	0.6%	0.6%	2.1%
8.0% Spread	Mean	7.5%	7.9%	7.0%	7.0%	7.6%
	σ	0.6%	0.5%	0.6%	0.6%	2.6%
Panel C: Random Spreads						
0.5% Spread	Mean	1.2%	0.7%	1.4%	0.6%	1.2%
	σ	0.4%	0.8%	0.3%	0.5%	1.3%
1.0% Spread	Mean	1.4%	1.0%	1.7%	1.0%	1.4%
	σ	0.4%	0.9%	0.4%	0.6%	1.5%
3.0% Spread	Mean	2.7%	3.4%	3.1%	2.8%	3.0%
	σ	0.6%	0.8%	0.5%	0.7%	2.0%
5.0% Spread	Mean	4.7%	5.7%	4.7%	4.5%	5.3%
	σ	0.8%	0.8%	0.7%	0.8%	2.3%
8.0% Spread	Mean	7.7%	9.2%	7.0%	7.0%	8.8%
	σ	1.1%	1.0%	1.1%	1.1%	3.0%

(continued)

Table I. Continued

		CHL		HL		Roll
		Two-Day	Monthly	Two-Day	Monthly	
Panel D: Overnight Price Movement with Half of the Daily Standard Deviation						
0.5% Spread	Mean	1.4%	0.8%	1.4%	0.4%	1.3%
	σ	0.4%	0.9%	0.3%	0.4%	1.5%
1.0% Spread	Mean	1.5%	1.0%	1.6%	0.6%	1.4%
	σ	0.4%	1.0%	0.4%	0.6%	1.6%
3.0% Spread	Mean	2.5%	2.9%	3.0%	2.4%	2.6%
	σ	0.6%	0.9%	0.5%	0.7%	2.0%
5.0% Spread	Mean	4.2%	4.9%	4.6%	4.3%	4.5%
	σ	0.7%	0.7%	0.6%	0.7%	2.3%
8.0% Spread	Mean	7.4%	8.0%	7.4%	7.3%	7.6%
	σ	0.7%	0.6%	0.7%	0.7%	2.8%
Panel E: All Imperfections Together						
0.5% Spread	Mean	1.4%	0.8%	1.1%	0.2%	1.3%
	σ	0.4%	0.9%	0.3%	0.3%	1.5%
1.0% Spread	Mean	1.5%	1.1%	1.3%	0.3%	1.5%
	σ	0.4%	1.0%	0.3%	0.4%	1.6%
3.0% Spread	Mean	2.8%	3.3%	2.5%	1.8%	3.0%
	σ	0.6%	0.9%	0.5%	0.8%	2.1%
5.0% Spread	Mean	4.6%	5.7%	4.0%	3.6%	5.3%
	σ	0.8%	0.9%	0.7%	0.9%	2.5%
8.0% Spread	Mean	7.6%	9.2%	6.3%	6.1%	8.7%
	σ	1.2%	1.1%	1.1%	1.2%	3.1%

Table II. Summary Statistics for Different Estimators

This table provides the main summary statistics for the pooled sample of the main estimators considered in this paper. The column labeled N refers to the number of stock-months of estimates in the sample. The column labeled $\rho\ (\,.\,, ES_{i,t})$ refers to the correlation of different estimates with the TAQ effective spread benchmark. The row labels refer to the TAQ effective spread benchmark (Effective Spread), our estimator (CHL), and the estimators proposed by Corwin and Schultz (HL; 2012), Roll (Roll; 1984), Hasbrouck (Gibbs; 2009), Holden (EffTick; 2009), and Fong, Holden, and Trzcinka (FHT; 2014). For calculating the CHL estimates, we replace the missing high, low, and close price with the previous days' values. We then discard monthly estimates for the months with fewer than 12 trading days (that is, days with positive high, low, and close price, as well as positive volume). The *HL* estimates are exactly calculated as in Corwin and Schultz (2012); that is, (1) missing daily high and low prices are replaced with those of previous days, (2) overnight adjustments are applied, and (3) monthly estimates with fewer than 12 two-day estimates are discarded. We merge the results of different estimators and discard stock-months in which any of the estimates are missing. We compute two versions of the *HL* (*CHL*) estimator, that is, the two-day corrected and monthly corrected versions labeled two-day and monthly. In the two-day corrected version for *HL* (*CHL*), we set each negative two-day spread (squared spread) to zero, and then the spreads (square roots of estimated squared spreads) are averaged within a month. The monthly corrected *HL* estimates are calculated by averaging all the two-day spreads within the month and then setting negative monthly averages to zero. The monthly *CHL* estimates are calculated by inserting monthly averages in equation (9), setting negative estimates of squared spreads to zero and, finally, taking the square roots. The *Roll* estimates are calculated by setting positive monthly autocovariance estimates to zero. The zeros reported for *EffTick* estimates reflect the months in which none of the prices are divisible by the base-eight denomination increments. We consider a second variant of *EffTick* measure (*EffTick – Alt. Incr.*) by using the tick sizes of 1¢, 5¢, 10¢, 25¢, 50¢, or \$1.00 as our sample time span lies after the decimalization of stock markets. For the sake of completeness, we include the Amihud price impact measure (*Amihud ILLIQ*, 2002), which captures the price impact. For the sake of convenience, the *Amihud ILLIQ* measure is multiplied by 10^6 in reporting the results.

	N	Mean	Median	Standard Deviation	$\rho\ (\,.\,, ES_{i,t})$	$\% \leq 0$
Effective Spread	536,275	0.84%	0.28%	1.43%	1.00	0.00%
CHL - Two-Day	536,275	1.40%	1.03%	1.32%	0.75	0.00%
CHL - Monthly	536,275	0.63%	0.38%	0.94%	0.68	33.69%
HL - Two-Day	536,275	1.22%	0.94%	1.04%	0.66	0.00%
HL - Monthly	536,275	0.58%	0.32%	0.88%	0.63	24.33%
Roll	536,275	1.52%	0.72%	2.59%	0.46	42.88%
Gibbs	536,275	2.14%	1.48%	2.95%	0.41	0.00%
EffTick	536,275	2.06%	0.67%	4.82%	0.42	26.94%
EffTick – Alt. Incr.	536,275	0.25%	0.07%	0.72%	0.52	0.00%
FHT	536,275	0.26%	0.00%	0.69%	0.44	61.14%
Amihud ILLIQ	536,275	0.847	0.0087	8.816	0.45	0.00%

Table III. Average Cross-Sectional Correlations with the TAQ Benchmark

This table shows the average cross-sectional correlations between estimates of transaction costs and the TAQ benchmark for each month. The monthly correlations are averaged over the specified sample periods. The labels in the first row refer to our estimator (CHL) and the estimators proposed by Corwin and Schultz (HL; 2012), Roll (Roll; 1984), Hasbrouck (Gibbs; 2009), Holden (EffTick; 2009), and Fong, Holden, and Trzcinka (FHT; 2014). N is the average number of stocks per month. An asterisk indicates numbers not significantly different from the estimator with the highest correlation marked in bold in every row. We test our hypotheses on the time series of pairwise difference in correlations for two estimators and assess whether the mean is significantly different from zero. We adjust for any potential time-series autocorrelation by using Newey-West (1987) standard errors with four lags autocorrelation. The size quintiles are sorted by increasing market capitalization at the last observed period for each individual stock. The spread quintiles are sorted by increasing average effective spreads during the whole sample period.

	N	CHL	HL	Roll	Gibbs	EffTick	FHT
Panel A: Average Cross-sectional Correlations with Effective Spreads for Monthly Estimates							
Full Period	3,972.4	**0.742**	0.646	0.427	0.376	0.414	0.492
2003–2007	4,381.1	**0.762**	0.664	0.435	0.374	0.458	0.570
2008–2011	3,870.9	**0.736**	0.635	0.428	0.430	0.391	0.471
2012–2014	3,528.9	**0.722**	0.633	0.414	0.307	0.381	0.407
Panel B: Average Cross-sectional Correlations with Changes in Effective Spreads for Monthly Estimates							
Full Period	3,920.8	**0.303**	0.287	0.117	0.096	0.026	0.054
2003–2007	4,266.7	**0.328**	0.306	0.128	0.093	0.029	0.075
2008–2011	3,765.9	**0.304**	0.292	0.121	0.120	0.028	0.052
2012–2014	3,410.5	**0.248**	0.241*	0.089	0.060	0.019	0.017
Panel C: Analysis across Different Markets							
NYSE	1,337.7	0.496*	0.481*	0.212	0.232	**0.511**	0.423
AMEX	306.0	**0.743**	0.655	0.458	0.547	0.324	0.516
NASDAQ	2,328.7	**0.712**	0.590	0.406	0.344	0.371	0.416
Panel D: Analysis across Market Capitalization							
Size Quintile 1	587.2	**0.695**	0.570	0.383	0.384	0.243	0.386
Size Quintile 2	682.1	**0.557**	0.378	0.330	0.272	0.192	0.249
Size Quintile 3	789.5	**0.455**	0.341	0.201	0.158	0.263	0.273
Size Quintile 4	867.9	**0.457**	0.439*	0.166	0.166	0.383	0.331
Size Quintile 5	1046	0.426	**0.477**	0.147	0.175	0.475*	0.356
Panel E: Analysis across Effective Spread Size							
ES Quintile 1	1,031.5	0.381	**0.450**	0.132	0.170	0.353	0.153
ES Quintile 2	842.6	0.375	**0.419**	0.143	0.156	0.342	0.217
ES Quintile 3	764.4	**0.403**	0.389	0.150	0.160	0.339	0.290
ES Quintile 4	725.5	**0.494**	0.384	0.245	0.204	0.256	0.279
ES Quintile 5	608.4	**0.700**	0.588	0.420	0.482	0.229	0.363

Table IV. Average Time-Series Correlations for Spread Estimates of Individual Stocks Compared to the TAQ Benchmark
The labels in the first row refer to our estimator (CHL) and the estimators proposed by Corwin and Schultz (HL; 2012), Roll (Roll; 1984), Hasbrouck (Gibbs; 2009), Holden (EffTick; 2009), and Fong, Holden, and Trzcinka (FHT; 2014). N is the number of stocks in the subsamples with, at least, six months of estimates. The averages are computed across stocks. An asterisk indicates numbers not significantly different from the estimator with the highest correlation marked in bold in every row. We use a paired t-test for the statistical inferences. The size quintiles are sorted by increasing market capitalization at the last observed period for each individual stock. The spread quintiles are sorted by increasing average effective spreads during the whole sample period.

	N	CHL	HL	Roll	Gibbs	EffTick	FHT
Panel A: Average Time-Series Correlations with Effective Spreads: Monthly Estimates							
Full Period	6,961	**0.523**	0.513	0.246	0.339	0.311	0.182
2003–2007	5,652	**0.393**	0.377	0.140	0.247	0.252	0.124
2008–2011	4,783	**0.611**	0.604	0.317	0.435	0.267	0.150
2012–2014	4,144	0.287	**0.295**	0.096	0.166	0.127	0.060
Panel B: Average Time-Series Correlations with Changes in Effective Spreads: Monthly Estimates							
Full Period	6,882	0.292*	**0.295**	0.116	0.170	0.052	0.023
2003–2007	5,574	0.256*	**0.258**	0.096	0.167	0.040	0.012
2008–2011	4,727	0.340	**0.351**	0.146	0.211	0.064	0.034
2012–2014	4,074	0.187*	**0.192**	0.067	0.100	0.017	-0.003
Panel C: Analysis across Different Markets							
NYSE	2,100	0.430	**0.447**	0.186	0.281	0.301	0.127
AMEX	815	**0.557**	0.508	0.259	0.417	0.292	0.268
NASDAQ	4,411	**0.548**	0.534	0.263	0.345	0.309	0.186
Panel D: Analysis across Market Capitalization							
Size Quintile 1	1,392	**0.702**	0.654	0.386	0.526	0.334	0.340
Size Quintile 2	1,392	**0.578**	0.528	0.294	0.405	0.314	0.191
Size Quintile 3	1,392	**0.482**	0.476*	0.215	0.289	0.307	0.156
Size Quintile 4	1,392	0.436	**0.460**	0.164	0.230	0.305	0.134
Size Quintile 5	1,393	0.416	**0.448**	0.173	0.248	0.297	0.089
Panel E: Analysis across Effective Spread Size							
ES Quintile 1	1,392	0.412	**0.442**	0.170	0.246	0.264	0.063
ES Quintile 2	1,392	0.429	**0.470**	0.165	0.233	0.310	0.122
ES Quintile 3	1,392	0.467	**0.478**	0.194	0.277	0.338	0.203
ES Quintile 4	1,392	**0.591**	0.537	0.285	0.372	0.335	0.248
ES Quintile 5	1,393	**0.715**	0.640	0.418	0.570	0.310	0.274

Table V. Prediction Errors

We measure the accuracy of different monthly estimates by computing their root-mean-squared errors (RMSEs), as well as mean absolute errors (MAEs) with respect to the TAQ benchmark**.** Prediction errors are calculated every month and then averaged through the months in the sample. N is the average number of stocks per month. The labels refer to our estimator (CHL) and the estimators proposed by Corwin and Schultz (HL; 2012), Roll (Roll; 1984), Hasbrouck (Gibbs; 2009), Holden (EffTick; 2009), and Fong, Holden, and Trzcinka (FHT; 2014). An asterisk indicates numbers not significantly different from the estimator with the lowest average prediction error marked in bold in every row. We test our hypotheses on the time series of pairwise difference in prediction errors for two estimators and assess whether the mean is significantly different from zero. We adjust for any potential time-series autocorrelation by using Newey-West (1987) standard errors with four lags autocorrelation

	N	CHL	HL	Roll	Gibbs	EffTick	FHT
Panel A: RMSEs, Breakdown for different periods, and across different markets							
Full Period	3,972.4	**0.0105**	0.0107	0.0223	0.0285	0.0440	0.0131
2003–2007	4,381.1	**0.0084**	0.0086	0.0182	0.0252	0.0369	0.0101
2008–2011	3,870.9	**0.0141**	0.0141*	0.0292	0.0325	0.0551	0.0175
2012–2014	3,528.9	**0.0087**	0.0092	0.0187	0.0280	0.0394	0.0117
NYSE	1,337.7	0.0089	0.0077	0.0163	0.0237	0.0165	**0.0030**
AMEX	306.0	**0.0115**	0.0125	0.0286	0.0237	0.0981	0.0192
NASDAQ	2,328.7	**0.0112**	0.0119	0.0239	0.0308	0.0436	0.0155
Panel B: MAEs, Breakdown for different periods, and across different markets							
Full Period	3,972.4	0.0083	0.0083	0.0134	0.0145	0.0182	**0.0068**
2003–2007	4,381.1	0.0067	0.0068	0.0112	0.0130	0.0155	**0.0055**
2008–2011	3,870.9	0.0113	0.0110	0.0180	0.0177	0.0234	**0.0087**
2012–2014	3,528.9	0.0065	0.0068	0.0105	0.0125	0.0153	**0.0059**
NYSE	1,337.7	0.0076	0.0068	0.0099	0.0131	0.0074	**0.0016**
AMEX	306.0	**0.0082**	0.0092	0.0189	0.0146	0.0506	0.0136
NASDAQ	2,328.7	**0.0087**	0.0091	0.0148	0.0154	0.0204	0.0089*

Table VI. Comparison with the Monthly TAQ (MTAQ) Benchmark, January 1993-September 2003

This table compares different estimates with the MTAQ benchmark between January 1993 and September 2003. The labels in the first row refer to our estimator (CHL) and the estimators proposed by Corwin and Schultz (HL; 2012), Roll (Roll; 1984), Hasbrouck (Gibbs; 2009), Holden (EffTick; 2009), and Fong, Holden, and Trzcinka (FHT; 2014). The spread quintiles are sorted by increasing average effective spreads during the whole sample period. In Panel A, N refers to the average number of stocks per month. Cross-sectional correlations are calculated per month and averaged across the sample. We test our hypotheses on the time series of pairwise difference in correlations for two estimators and assess whether the mean is significantly different from zero. In Panel B, N refers to the number of stocks in the subsamples with at least 6 months of estimates. Time-series correlations are calculated for each individual stock and then averaged across assets. We use a paired t-test for the statistical inferences. In Panel C, N refers to the average number of stocks per month. RMSEs are calculated for every month and then averaged through time. We test our hypotheses on the time series of pairwise difference in prediction errors for two estimators and assess whether the mean is significantly different from zero. We adjust for any potential time-series autocorrelation by using Newey-West (1987) standard errors with four lags autocorrelation. An asterisk indicates numbers not significantly different from the highest correlation marked in bold in every row of Panel A and B, and from the estimator with the lowest average prediction error marked in bold in every row in Panel C.

	N	CHL	HL	Roll	Gibbs	EffTick	FHT
Panel A: Average Cross-Sectional Correlations with the TAQ Benchmark							
All Stocks, Levels	5057.2	**0.860**	0.833	0.606	0.715	0.637	0.645
All Stocks, Changes	4973.6	**0.470**	0.459	0.206	0.267	0.191	0.152
NYSE	1616.5	**0.814**	0.810*	0.461	0.636	0.811*	0.757
AMEX	363.0	**0.927**	0.915	0.653	0.849	0.788	0.740
NASDAQ	3077.7	**0.817**	0.774	0.571	0.660	0.604	0.585
ES Quintile 1	1385.3	0.428	0.440	0.153	0.296	**0.658**	0.425
ES Quintile 2	1125.6	**0.579**	0.575*	0.275	0.339	0.526*	0.426
ES Quintile 3	967.7	**0.668**	0.627	0.359	0.431	0.482	0.413
ES Quintile 4	923.9	**0.750**	0.708	0.450	0.567	0.491	0.459
ES Quintile 5	654.7	**0.799**	0.776	0.537	0.698	0.481	0.511
Panel B: Average Time-Series Correlations for Spread Estimates of Individual Stocks							
All Stocks, Levels	10788	**0.586**	0.580	0.281	0.445	0.464	0.403
All Stocks, Changes	10683	**0.401**	0.399*	0.155	0.284	0.220	0.116
NYSE	2759	0.343	0.353	0.096	0.285	**0.459**	0.409
AMEX	1084	**0.644**	0.623	0.286	0.507	0.601	0.355
NASDAQ	7744	**0.645**	0.632	0.326	0.482	0.450	0.383
ES Quintile 1	2157	0.200	0.209	0.015	0.169	0.333	**0.402**
ES Quintile 2	2157	0.526	**0.552**	0.189	0.325	0.458	0.400
ES Quintile 3	2157	**0.686**	0.674	0.325	0.483	0.483	0.376
ES Quintile 4	2157	**0.743**	0.716	0.406	0.574	0.524	0.402
ES Quintile 5	2160	**0.773**	0.747	0.469	0.674	0.521	0.434
Panel C: Root-Mean-Squared Errors w.r.t TAQ Benchmark							
All Stocks	5057.2	**0.0143**	0.0152	0.0309	0.0251	0.0466	0.0225
NYSE	1616.5	0.0065	**0.0059**	0.0169	0.0161	0.0186	0.0075
AMEX	363.0	**0.0125**	0.0162	0.0326	0.0189	0.0535	0.0246
NASDAQ	3077.7	**0.0223**	0.0234	0.0408	0.0341	0.0591	0.0326

Table VII. Comparison with Combinations of Models

We combine the bid-ask spread estimates of different models by computing simple averages and then compare the results with our estimates. The labels in the first row refer to our estimator (CHL) and the estimators proposed by Corwin and Schultz (HL; 2012), Roll (Roll; 1984), Hasbrouck (Gibbs; 2009), Holden (EffTick; 2009), and Fong, Holden, and Trzcinka (FHT; 2014). "The Rest" refers to the three latter mentioned estimators. In panels A and C, N refers to the average number of stocks per month. In panel B, N refers to the number of stocks in the subsamples with at least 6 months of estimates. The spread quintiles are sorted by increasing average effective spreads during the whole sample period. Cross-sectional correlations are calculated per month and averaged across the sample. Time-series correlations are calculated for each individual stock and then averaged across assets. RMSEs are calculated for every month and then averaged through time.

	N	CHL	HL & Roll	HL, Roll, & the Rest	CHL, HL, & Roll	All
Panel A: Average Cross-Sectional Correlations with the TAQ Benchmark						
All Stocks, Levels	3,972.4	0.742	0.547	0.585	0.631	0.621
All Stocks, Changes	598.1	0.303	0.171	0.133	0.216	0.159
NYSE	1,337.7	0.496	0.314	0.477	0.380	0.494
AMEX	306.0	0.743	0.584	0.549	0.660	0.588
NASDAQ	2,328.7	0.712	0.511	0.537	0.596	0.575
ES Quintile 1	1,031.5	0.381	0.232	0.296	0.284	0.319
ES Quintile 2	842.6	0.375	0.234	0.322	0.284	0.341
ES Quintile 3	764.4	0.403	0.236	0.359	0.293	0.377
ES Quintile 4	725.5	0.494	0.323	0.356	0.389	0.382
ES Quintile 5	608.4	0.700	0.531	0.476	0.610	0.522
Panel B: Average Time-Series Correlations for Spread Estimates of Individual Stocks						
All Stocks, Levels	6,961	0.523	0.353	0.418	0.420	0.448
All Stocks, Changes	6882	0.292	0.173	0.182	0.215	0.207
NYSE	2,100	0.430	0.277	0.357	0.330	0.380
AMEX	815	0.557	0.369	0.437	0.447	0.471
NASDAQ	4,411	0.548	0.374	0.432	0.444	0.466
ES Quintile 1	1,392	0.412	0.260	0.312	0.312	0.338
ES Quintile 2	1,392	0.429	0.268	0.329	0.324	0.355
ES Quintile 3	1,392	0.467	0.298	0.396	0.359	0.420
ES Quintile 4	1,392	0.591	0.398	0.477	0.479	0.512
ES Quintile 5	1,393	0.715	0.541	0.574	0.625	0.616
Panel C: Root-Mean-Squared Errors w.r.t TAQ Benchmark						
All Stocks, Levels	3,972.4	0.0105	0.0141	0.0147	0.0123	0.0135
All Stock, Changes	3,920.8	0.0083	0.0150	0.0147	0.0118	0.0129
NYSE	1,337.7	0.0089	0.0109	0.0098	0.0099	0.0094
AMEX	306.0	0.0115	0.0170	0.0232	0.0141	0.0202
NASDAQ	2,328.7	0.0112	0.0152	0.0156	0.0131	0.0144
ES Quintile 1	1,031.5	0.0077	0.0089	0.0075	0.0083	0.0074
ES Quintile 2	842.6	0.0098	0.0120	0.0108	0.0109	0.0104
ES Quintile 3	764.4	0.0106	0.0141	0.0139	0.0124	0.0130
ES Quintile 4	725.5	0.0103	0.0148	0.0174	0.0125	0.0157
ES Quintile 5	608.4	0.0146	0.0208	0.0232	0.0176	0.0208

Table VIII. Partial Correlations

We calculate the partial correlation between the TAQ effective spread and our estimates (CHL) removing the effects explained by other estimates, that is, $\rho\left(\epsilon_{ES_{i,t}|\boldsymbol{EST}_{i,t}}, \epsilon_{CHL_{i,t}|\boldsymbol{EST}_{i,t}}\right)$. $\boldsymbol{EST}_{i,t}$ includes a constant, Corwin and Schultz (HL; 2012), Roll (Roll; 1984), Hasbrouck (Gibbs; 2009), Holden (EffTick; 2009), and Fong, Holden, and Trzcinka (FHT; 2014) estimates. The spread quintiles are sorted by increasing average effective spreads during the whole sample period. In panel A, N refers to the average number of stocks per month, and, in panel B, N refers to the number of stocks in the subsamples with at least 24 months of estimates. Panel A shows the average partial cross-sectional correlations. The bold numbers are significantly different from zero using a 5% two-tailed confidence interval. The statistical test for the average of cross-sectional correlations is based on Newey-West (1987) standard errors with four lags autocorrelation. Panel B shows the average partial time-series correlations, as the average of partial time-series correlations for individual stocks. The bold numbers are significantly different from zero using a t-test for the average of time-series correlations. To avoid overfitting in calculating the partial time-series correlations, we discard the stocks with fewer than 24 months of estimates.

	N	CHL|HL	CHL|HL, Roll	CHL|HL, Roll, Gibbs	CHL|HL, Roll, Gibbs, EffTick	CHL|HL, Roll, Gibbs, EffTick, FHT
Panel A: Average Partial Cross-Sectional Correlations with the TAQ Benchmark						
All Stocks, Levels	3,972.4	**0.482**	**0.460**	**0.453**	**0.442**	**0.433**
All Stocks, Changes	3,920.8	**0.162**	**0.157**	**0.152**	**0.152**	**0.151**
NYSE	1,337.7	**0.194**	**0.209**	**0.198**	**0.173**	**0.167**
AMEX	306.0	**0.476**	**0.440**	**0.415**	**0.411**	**0.408**
NASDAQ	2,328.7	**0.498**	**0.471**	**0.465**	**0.457**	**0.450**
ES Quintile 1	1,031.5	**0.049**	**0.072**	**0.061**	**0.053**	**0.051**
ES Quintile 2	842.6	**0.084**	**0.100**	**0.094**	**0.079**	**0.075**
ES Quintile 3	764.4	**0.164**	**0.181**	**0.179**	**0.167**	**0.161**
ES Quintile 4	725.5	**0.338**	**0.324**	**0.329**	**0.324**	**0.318**
ES Quintile 5	608.4	**0.473**	**0.429**	**0.400**	**0.399**	**0.399**
Panel B: Average Partial Time-Series Correlations for Spread Estimates of Individual Stocks Compared to the TAQ Benchmark						
All Stocks, Levels	5,748	**0.222**	**0.231**	**0.200**	**0.186**	**0.183**
All Stocks, Changes	5,683	**0.134**	**0.139**	**0.120**	**0.119**	**0.119**
NYSE	1,818	**0.125**	**0.152**	**0.122**	**0.113**	**0.113**
AMEX	575	**0.323**	**0.310**	**0.262**	**0.250**	**0.247**
NASDAQ	3,558	**0.252**	**0.256**	**0.226**	**0.211**	**0.206**
ES Quintile 1	1,280	**0.090**	**0.122**	**0.099**	**0.093**	**0.093**
ES Quintile 2	1,185	**0.098**	**0.134**	**0.114**	**0.104**	**0.102**
ES Quintile 3	1,108	**0.171**	**0.192**	**0.173**	**0.155**	**0.153**
ES Quintile 4	1,125	**0.337**	**0.331**	**0.299**	**0.275**	**0.266**
ES Quintile 5	1,050	**0.454**	**0.410**	**0.342**	**0.331**	**0.325**

Table IX. Correlations for Quintiles Based on Average Number of Trades

We group the stocks into five quintiles sorting them by their average number of trades per day during the sample period. The daily number of trades is counted using TAQ consolidated trades data for trades that occur between 9:30 and 16:00 and have a positive price and volume. The first four quintiles are constructed of 1,392 stocks, and the fifth is constructed of 1,393 stocks. The labels in the first row refer to our estimator (CHL) and the estimators proposed by Corwin and Schultz (HL; 2012), Roll (Roll; 1984), Hasbrouck (Gibbs; 2009), Holden (EffTick; 2009), and Fong, Holden, and Trzcinka (FHT; 2014). N refers to the number of stock-months of estimates for the entire sample, as well as for each quintile. Panel A shows the correlation coefficients between different monthly estimates and the TAQ effective spread benchmark. Panel B shows the correlation coefficients between different monthly estimates and the Amihud illiquidity measure (Amihud 2002). The bold numbers are the highest correlations coefficients in each row. An asterisk indicates numbers not significantly different from the estimator with the highest correlation, using Fisher's z-test to compare the correlation coefficients.

	N	CHL	HL	Roll	Gibbs	EffTick	FHT
Panel A: Correlation of monthly estimates with the TAQ effective spread benchmark							
Full Sample	536,275	**0.747**	0.663	0.456	0.406	0.424	0.445
ANTD Quintile 1	72,873	**0.823**	0.767	0.579	0.707	0.378	0.380
ANTD Quintile 2	96,875	**0.789**	0.725	0.461	0.465	0.412	0.463
ANTD Quintile 3	100,951	**0.707**	0.681	0.362	0.330	0.440	0.460
ANTD Quintile 4	120,935	0.619	**0.634**	0.290	0.268	0.493	0.445
ANTD Quintile 5	144,641	0.529	**0.561**	0.241	0.242	0.547	0.474
Panel B: Correlation of monthly estimates with the Amihud illiquidity measure							
Full Sample	536,275	**0.409**	0.330	0.264	0.221	0.189	0.199
ANTD Quintile 1	72,873	**0.521**	0.452	0.368	0.442	0.230	0.252
ANTD Quintile 2	96,875	**0.411**	0.331	0.275	0.243	0.199	0.176
ANTD Quintile 3	100,951	**0.342**	0.285	0.209	0.154	0.196	0.199
ANTD Quintile 4	120,935	**0.175**	0.140	0.087	0.067	0.122	0.093
ANTD Quintile 5	144,641	0.179	0.158	0.088	0.070	**0.218**	0.178

Table X. Estimating Systematic Liquidity Risks from the LCAPM Model for Individual Stocks

We calculate the components of systematic risk implied by the LCAPM model (Acharya and Pedersen (2005)) by using the TAQ effective spreads, *Roll* model estimates (Roll; 1984), the *HL* estimates (Corwin and Schultz; 2012), and the estimates from our model (labeled *CHL*). N refers to the number of stocks. The table reports the cross-sectional correlation of betas based on *Roll*, *HL*, and *CHL* estimates ($\beta_i^{Estimates}$), with betas based on the TAQ effective spreads (β_i^{ES}). We discard stocks with fewer than 30 months of effective spread estimates. Betas are calculated for the spreads in levels and the residuals of AR(2) regressions in panels A and B, respectively. Panels C, D, and E show the results from subsampling analyses across exchanges (NYSE, AMEX, and NASDAQ), market capitalization, and spread size. In panel D, the size quintiles are sorted by increasing market capitalization at the last observed period for each individual stock. In panel E, the spread quintiles are sorted by increasing average effective spreads during the whole sample period. An asterisk indicates values not significantly different from that with the higher correlation marked in bold for every set of values. The statistical inferences are performed using Fisher's z-test.

		$\rho(\beta_2^{ES}, \beta_2^{Estimates})$			$\rho(\beta_3^{ES}, \beta_3^{Estimates})$			$\rho(\beta_4^{ES}, \beta_4^{Estimates})$		
	N	CHL	HL	Roll	CHL	HL	Roll	CHL	HL	Roll
Panel A: Cross-section of Estimated Systematic Risks: All Stocks										
Full Period	5396	**0.829**	0.742	0.650	**0.937**	0.930	0.832	**0.761**	0.672	0.484
2003-2007	4119	**0.501**	0.457	0.171	**0.671**	0.659*	0.407	**0.532**	0.439	0.214
2008-2011	3575	**0.736**	0.604	0.353	**0.971**	0.970*	0.896	**0.557**	0.427	0.296
2012-2014	3068	**0.615**	0.319	0.125	**0.591**	0.386	0.223	**0.407**	0.325	0.102
Panel B: Cross-section of Estimated Systematic Risks Considering Liquidity Shocks of AR(2) Model										
Full Period	5,308	**0.511**	0.439	0.152	0.860	**0.879**	0.588	**0.535**	0.258	0.251
2003–2007	4,012	**0.287**	0.161	-0.077	**0.795**	0.776	0.554	**0.369**	0.148	0.066
2008–2011	3,565	**0.394**	0.267	-0.030	0.892	**0.914**	0.627	**0.520**	0.108	0.143
2012–2014	3,057	0.087*	0.084*	**0.088**	0.676	**0.736**	0.540	**0.206**	0.089	0.053
Panel C: Analysis across Different Markets										
NYSE	1,724	**0.696**	0.679*	0.434	**0.937**	0.934*	0.837	0.678*	**0.683**	0.257
AMEX	510	**0.882**	0.825	0.686	**0.923**	0.916*	0.819	**0.780**	0.701	0.536
NASDAQ	3,316	**0.854**	0.780	0.696	**0.938**	0.930	0.830	**0.786**	0.692	0.493
Panel D: Analysis across Market Capitalization										
Size Quintile 1	987	**0.865**	0.784	0.745	**0.949**	0.941*	0.849	**0.806**	0.724	0.575
Size Quintile 2	990	**0.741**	0.600	0.560	**0.926**	0.915*	0.787	**0.626**	0.513	0.373
Size Quintile 3	1,053	**0.591**	0.458	0.296	**0.931**	0.930*	0.848	**0.520**	0.417	0.235
Size Quintile 4	1,128	**0.558**	0.502*	0.297	**0.917**	0.903*	0.790	**0.590**	0.575*	0.209
Size Quintile 5	1,238	**0.478**	0.465*	0.290	**0.932**	0.930*	0.835	0.456*	**0.491**	0.119
Panel E: Analysis across Effective Spread Size										
ES Quintile 1	1,237	0.556*	**0.558**	0.339	0.908*	**0.908**	0.778	0.538*	**0.574**	0.210
ES Quintile 2	1,119	0.709*	**0.721**	0.360	**0.940**	0.934*	0.838	0.623	**0.672**	0.141
ES Quintile 3	1,035	**0.719**	0.707*	0.375	**0.939**	0.929*	0.841	**0.675**	0.645*	0.258
ES Quintile 4	1,055	**0.786**	0.727	0.508	**0.928**	0.921*	0.811	**0.738**	0.699*	0.387
ES Quintile 5	950	**0.881**	0.819	0.776	**0.944**	0.939*	0.845	**0.826**	0.754	0.585

Table XI. Estimating Systematic Liquidity Risks from the LCAPM Model for 25 Portfolios of Stocks Sorted by Illiquidity Level

We calculate the components of systematic risk implied by the LCAPM model (Acharya and Pedersen (2005)) for 25 portfolios sorted by illiquidity levels, constructed using the TAQ effective spreads, Roll model estimates (Roll; 1984), the HL estimates (Corwin and Schultz; 2012), and the estimates from our model (labeled CHL). The table reports the cross-sectional correlation of portfolio betas based on Roll's, HL's, and CHL's estimates ($\beta_i^{Estimates}$), with portfolio betas based on the TAQ effective spreads (β_i^{ES}). Portfolio spreads are in levels or as shocks, defined by residuals of AR(2) regressions. An asterisk indicates values not significantly different from that with the higher correlation marked in bold for every set of values. The statistical inferences are performed using Fisher's z-test.

Cross-Sectional Correlations of Betas with those Calculated using TAQ Effective Spreads *(continued)*

	$\rho(\beta_1^{ES},\beta_1^{Estimates})$			$\rho(\beta_2^{ES},\beta_2^{Estimates})$			$\rho(\beta_3^{ES},\beta_3^{Estimates})$		
	CHL	HL	Roll	CHL	HL	Roll	CHL	HL	Roll
Levels	**0.693**	0.621*	0.238	**0.978**	0.976*	0.933	**0.352**	-0.479	0.132
Shocks	**0.693**	0.621*	0.238	0.966*	**0.970**	0.159	**0.952**	0.940*	0.668

Table XI. Continued

Cross-Sectional Correlations of Betas with those Calculated using TAQ Effective Spreads

	$\rho(\beta_4^{ES},\beta_4^{Estimates})$			$\rho(\beta_{Net}^{ES},\beta_{Net}^{Estimates})$		
	CHL	HL	Roll	CHL	HL	Roll
Levels	**0.901**	0.848*	0.155	**0.754**	0.698*	0.311
Shocks	**0.957**	0.930*	0.509	**0.744**	0.682*	0.296