Finance and Economics Discussion Series
Divisions of Research & Statistics and Monetary Affairs
Federal Reserve Board, Washington, D.C.

When do low-frequency measures really measure transaction costs?

Mohammad R. Jahan-Parvar and Filip Zikes

2019-051

Please cite this paper as:
Jahan-Parvar, Mohammad R., and Filip Zikes (2019). "When do low-frequency measures really measure transaction costs?," Finance and Economics Discussion Series 2019-051. Washington: Board of Governors of the Federal Reserve System, https://doi.org/10.17016/FEDS.2019.051.

When do low-frequency measures really measure transaction costs?

Mohammad R. Jahan-Parvar Filip Zikes*

June 28, 2019

Abstract

We compare popular measures of transaction costs based on daily data with their high-frequency data-based counterparts. We find that for U.S. equities and major foreign exchange rates, (i) the measures based on daily data are highly upward biased and imprecise; (ii) the bias is a function of volatility; and (iii) it is primarily volatility that drives the dynamics of these liquidity proxies both in the cross section as well as over time. We corroborate our results in carefully designed simulations and show that such distortions arise when the true transaction costs are small relative to volatility. Many financial assets exhibit this property, not only in the last two decades, but also in the previous century. We document that using low-frequency measures as liquidity proxies in standard asset pricing tests may produce sizable biases and spurious inferences about the pricing of aggregate volatility or liquidity risk.

*Jahan-Parvar, mohammad.jahan-parvar@frb.gov, Division of International Finance, Federal Reserve Board; Zikes, filip.zikes@frb.gov, Division of Financial Stability, Federal Reserve Board. We thank Karim Abadir, Alain Chaboud, Anusha Chari, Sergio Correia, Dobrislav Dobrev, Michael Gordy, Luca Guerrieri, Peter Reinhard Hansen, Erik Hjalmarsson, Aytek Malkhozov, Valery Polkovnichenko, David Rappoport, Marius Rodrigues, Pawel Szerszen, Clara Vega, Lingling Zheng, Molin Zhong, and seminar participants at the University of North Carolina at Chapel Hill, Federal Reserve Board, Market Microstructure: Confronting Many Viewpoints #5, and Econometric Conference on Big Data and AI in Honor of Ronald Gallant for useful comments and discussions. We are grateful to Mark Morley and Charles Horgan for excellent research assistance and to William McClennan for helping us run our programs on a computer cluster. The views expressed in this paper are those of the authors and should not be interpreted as representing the views of the Federal Reserve Board or any other person associated with the Federal Reserve System.

1 Introduction

Transaction costs are a key input in many financial decisions. The literature on measuring transaction costs and market liquidity more generally, has evolved into two distinct strands. With the increasing availability of high-frequency data since the 1990s, one strand emphasizes the importance of using transactions data to calculate trading costs. Chordia, Sarkar, and Subrahmanyam (2005) who study stock and bond market liquidity and Mancini, Ranaldo, and Wrampelmeyer (2013) who examine foreign exchange market liquidity are examples of papers that use high-frequency measures. The other strand argues that the availability of long histories of daily data—dating back to early 1900s—and the significant cost associated with acquiring and processing high-frequency data makes it desirable to estimate transaction costs from low-frequency, daily data. Roll (1984), Hasbrouck (2009), Corwin and Schultz (2012), and Abdi and Ranaldo (2017) propose low-frequency measures that have become popular in the literature. As the latter are essentially statistical proxies for the former, we would expect them to deliver similar results on average.

In this paper, we find that, in general, the low-frequency measures of transaction costs are not good estimators for the high-frequency-based trading costs. We show that for U.S. equities and major foreign exchange rates, (i) several widely used measures suffer from a large bias that cannot be diminished by using more data; (ii) more importantly, the bias is a function of volatility, and hence it induces a positive correlation between the low-frequency measures and volatility above and beyond what is implied by the correlation between volatility and the true transaction costs; and (iii) the volatility-induced bias has important implications for empirical asset pricing. Put simply, volatility explains the wedge between the low-frequency measures and their high-frequency counterparts.

Specifically, for the S&P 1500 stocks between 2003 and 2017, we show that the average effective spread computed from the Trade and Quote (TAQ) data is around 12 basis points, while the monthly low-frequency measures–measures calculated from one month of daily data–yield spreads between 26 and 168 basis points on average. The five major foreign exchange rates trade in the Electronic Broking Services (EBS) market with an effective spread of less than one basis point on average between 2008 and 2015, but the monthly low-frequency measures deliver estimates between 13 and 48 basis points on average. When we regress the monthly low-frequency measures on the true spread and realized volatility, we find that the coefficient on realized volatility is highly statistically significant. Moreover, relative to a simple regression of the low-frequency measures on the true spread alone, adding realized volatility significantly increases the R^2, often more than doubling it. For the foreign exchange rates, adding realized volatility to the regression even renders the coefficient on the true spread statistically insignificant for some low-frequency measures.

How does the volatility-induced bias arise? When recovering the transaction costs from daily data, one is attempting to estimate a small, positive-valued object from noisy data. This requires methods that guarantee non-negativity of the transaction costs estimates, such as censoring, truncation, or an appropriate prior distribution if one chooses a Bayesian approach. It is well-known from the study of limited dependent variable models that truncating or censoring a random variable at zero will increases the mean and the mean becomes a function of volatility; see Greene (2017). We show by carefully designed simulations that the same problem arises in the construction of the low-frequency measures. The magnitude of the problem depends on the relative size of the true transaction costs and the volatility of the daily asset price. The lower (higher) the transaction costs relative to volatility, the higher (lower) the bias and the stronger (weaker) the dependence on volatility are.

Since many financial markets experienced a significant decline in transaction costs, both in absolute terms and relative to the volatility of asset returns, over the past two decades, the issues we uncover have become more acute. But they have also implications for the most liquid financial assets in the previous century. Jones (2002) documents that the proportional effective spread was only around 60 basis points on average for the Dow Jones Industrial Average 30 stocks between 1928 and 2000. Thus, there has always been a nontrivial subset of U.S. stocks for which the issues we document mattered. Bessembinder (1994) reports that the average bid-ask spread in the interdealer foreign exchange market was between 5 and 8 basis points on average for major currencies between 1979 and 1992. Although such values are much larger than in the current century, our results indicate that they are still small enough for the volatility to have a sizable effect.

We use two well-known asset pricing models to illustrate how using transaction costs measures that suffer from the presence of a non-negligible, volatility-driven bias may produce misleading inference and spurious results. First, we document the consequences of using low-frequency liquidity measures as liquidity proxies in the liquidity-adjusted CAPM of Acharya and Pedersen (2005). Our simulation results show that both the liquidity betas and the market price of risk suffer from sizable biases; in empirically relevant calibrations, popular low-frequency measures yield negative market price of risk estimates even though the true value in the simulation is positive. Second, we study the pricing of aggregate volatility and liquidity risks in an unconditional asset pricing model motivated by Ang, Hodrick, Xing, and Zhang (2006). We demonstrate through simulations that using low-frequency measures as a proxy for aggregate liquidity may yield spurious pricing of liquidity risk due to the correlation of these low-frequency measures with volatility. The estimated price of liquidity risk can be as low as -10% per annum or as high as 2.5% per annum on average, depending on which low-frequency measure is used, even though the true price of risk in the simulation is zero.

Accurate measures of transaction are essential in many other areas of empirical finance. Our findings about the upward bias of the low-frequency measures raise questions about their suitability for optimal portfolio choice problems with transaction costs, where the so-called no-trade region is a function of the level of transaction costs (Constantinides, 1986), and for evaluation of trading strategies and asset pricing anomalies, where small changes in transaction costs have a large impact on performance (Novy-Marx and Velikov, 2015, Chen and Velikov, 2018, and Patton and Weller, 2018). The dependence of the low-frequency measures on volatility limits their use in studies of the commonality in liquidity, where liquidity proxies are used to measure co-movements in liquidity across assets and markets (Chordia, Roll, and Subrahmanyam, 2000, Hasbrouck and Seppi, 2001, Korajczyk and Sadka, 2008, Karolyi, Lee, and Van Dijk, 2012, Mancini, Ranaldo, and Wrampelmeyer, 2013). In all of these applications, using imperfect proxies for transaction costs can lead to suboptimal financial decisions.

The literature on low-frequency measures of liquidity is fairly large. A number of influential studies introduce measures that build upon the seminal work of Roll (1984) and his proposed measure of transaction costs. Among these studies, we examine the measures introduced by Hasbrouck (2009), Corwin and Schultz (2012), and Abdi and Ranaldo (2017).

Alternative measures include those based on the frequency of zero returns (Lesmond, Ogden, and Trzcinka, 1999), effective tick, which is based on the concept of price clustering (Holden, 2009 and Goyenko, Holden, and Trzcinka, 2009), or the simplified version of the Lesmond, Ogden, and Trzcinka (1999) measure introduced by Fong, Holden, and Trzcinka (2017). We do not study these measures here for two reasons. First, Corwin and Schultz (2012) and Abdi and Ranaldo (2017) show that their proposed measures outperform these alternatives, and second, these measures are less suitable for our empirical exercise. For example, we virtually never observe a zero daily return in our data.

Additionally, previous literature has examined the performance of a variety of low-frequency measures against high-frequency benchmarks. Hasbrouck (2009), Goyenko, Holden, and Trzcinka (2009), Corwin and Schultz (2012), Abdi and Ranaldo (2017) provide evidence from the U.S. stock market; Fong, Holden, and Trzcinka (2017) extend the analysis to global equity markets; Karnaukh, Ranaldo, and Söderlind (2015) study the foreign exchange market; Marshall, Nguyen, and Visaltanachoti (2011) study the commodity market; and Schestag, Schuster, and Uhrig-Homburg (2016) examine the U.S. corporate bond market. All of these studies conclude that the low-frequency measures are good proxies of true transaction cost, as they closely correlate with their high-frequency benchmarks. So why do we reach a different conclusion? We argue that while these measures are correlated with the true spreads as the literature finds, this correlation is to a large extent driven by volatility. Since the realized volatility and the true spread are positively correlated as predicted by market microstructure theory, omitting the realized volatility from the regression inflates the magnitude and statistical significance of the correlation between the low-frequency measures and the true spread. To the best of our knowledge, this issue has not been recognized by the existing literature. [1]

We do not claim that low-frequency measures of transaction costs are not generally useful. They are certainly useful for assets with relatively higher transaction costs and lower volatility, such as corporate bonds. However, we argue that caution is needed when applying these measures–whether in today's electronic markets or to data from the previous

[1] Our foreign exchange results differ from those of Karnaukh, Ranaldo, and Söderlind (2015). We discuss the sources of these differences in Section 4.2.2.

century–unless one is reasonably confident that the transaction costs are large relative to the daily volatility. It is worth mentioning that Hasbrouck (2009) already noted that his measure requires a significant amount of data and should not be expected to perform well for highly liquid assets. Contrary to his recommendation, however, the common practice in the literature is to construct the measures from only one month worth of data and apply them to the entire cross section of U.S. stocks and foreign exchange rates. In this study, we show that for less liquid assets, the consistent version of Corwin and Schultz (2012) performs better than other low-frequency liquidity measures, provided that one uses a fairly long window of data (at least one year) for constructing the measure. Otherwise, high-frequency measures are preferable.

Our analysis bears considerable similarity to several important debates in finance and economics literature. First, recovering a small positive object from noisy data is notoriously difficult. Examples include estimating the expected stock return from low-frequency data (Merton, 1980), recovering a small predictable component of consumption growth from consumption data (Bansal and Yaron, 2004), or expected dividend growth (Lettau and Ludvigson, 2005). Second, the necessity of using high-frequency data for measuring transaction costs resemble the discussion in the early 2000s about the use of parametric, low-frequency data based measures of volatility against high-frequency nonparametric volatility, the so-called realized measures (Andersen, Bollerslev, Diebold, and Labys, 2003, Andersen, Bollerslev, and Diebold, 2010, Bollerslev, Hood, Huss, and Pedersen, 2018). Spiegel and Zhang (2013) critically assess the convex relation between flow and past performance of mutual funds. They show that several studies that report a convex flow response function suffer from misspecification in their empirical models. Finally, our exercise is also similar in spirit to the paper by Farre-Mensa and Ljungqvist (2016), who study the performance of various measures of financial constraints. They find that instead of measuring financial constraints, these measures in reality capture firms' growth and financing policies; our results show that popular low-frequency measures of transaction costs are largely driven by volatility.

The rest of the paper is organized as follows. In Section 2, we describe the four low-frequency measures of transaction costs, which have become popular in the literature. In

Section 3, we study the properties of these measures in simulations and document their bias and dependence on volatility. Section 4 presents our empirical results for U.S. stocks and major foreign exchange rates, and in Section 5 we discuss the implications of our results for empirical finance. Section 6 concludes.

2 Low-frequency effective spread measures

The four low-frequency measures of transaction cost that we discuss in this paper are all derived within the well-known model of Roll (1984). In this section, we first present the model and then briefly describe the measures. For more details on the model, and its many generalizations, see Foucault, Pagano, and Röell (2013).

2.1 Notation and framework

We assume there are T days with n transactions each. For every $i = 1, ..., n$ and $t = 0, ..., T-1$, the logarithmic (log) transaction price p and the log efficient price m follow

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

$$m_{t+i/n} = m_{t+(i-1)/n} + \epsilon_{t+i/n}, \tag{2}$$

where ϵ, the efficient price innovation, is independently and identically distributed (*iid*) with variance σ^2/n; q, the buy/sell indicator, is *iid* and $q_{t+i/n} = \pm 1$ with equal probability; ϵ and q are independent at all leads and lags; and $m_0 = 0$ without loss of generality. The sequence of daily closing log prices is thus given by $\{p_1, p_2, ..., p_T\}$, the sequence of daily high log prices by $\{h_1, h_2, ..., h_T\}$, where $h_t = \max_{1 \le i \le n} p_{t-1+i/n}$, and the sequence of daily low log prices by $\{l_1, ..., l_T\}$, where $l_t = \min_{1 \le i \le n} p_{t-1+i/n}$. The sequence of daily returns is given by $\{r_2, r_3, ..., r_T\}$, where $r_t = p_t - p_{t-1}$. It follows from the assumptions above that the daily efficient price innovations, $m_t - m_{t-1}$, are *iid* with zero mean and variance σ^2. The parameter of interest is s, the proportional effective spread.

2.2 Roll (1984)

Roll (1984) exploits the negative serial correlation in returns, induced by the bid-ask bounce. In our framework, the first-order serial covariance is $\mathrm{Cov}(r_t, r_{t-1}) = -\frac{s^2}{4}$. For a sample of T days, the Roll measure is obtained by replacing the true covariance with its sample counterpart:

$$R = 2\sqrt{\max\left\{-\frac{1}{T-2}\sum_{t=3}^{T} r_t r_{t-1}, 0\right\}}, \tag{3}$$

where the censoring at zero ensures nonnegative estimates, as the sample first-order serial covariance is not generally guaranteed to be negative.

2.3 Hasbrouck (2009)

Hasbrouck (2009) proposes to estimate the Roll model by Bayesian methods. He develops a Gibbs sampler to approximate the posterior distribution of s, given a sample of T daily closing transaction prices. The prior on s is half-normal, $s \sim \mathrm{N}_+(0, \sigma_s^2)$, and the prior on σ is inverse Gamma, $\sigma \sim \mathrm{IG}(\alpha, \beta)$. Following Hasbrouck (2009), we set $\alpha = \beta = 1\mathrm{e}^{-12}$ and $\sigma_s^2 = 0.05^2$, but we also consider a tighter prior for s, namely $\sigma_s^2 = 0.01^2$. We denote the resulting estimators by H_L and H_T, respectively.

2.4 Corwin and Schultz (2012)

Corwin and Schultz, henceforth CS, exploit the information contained in daily high and low prices. They observe that "the sum of the price ranges over 2 consecutive single days reflects 2 days volatility and twice the spread, while the price range over one 2-day period reflects 2 days volatility and one spread". This yields two equations and two unknowns (s and σ), which is solved analytically. The two-day CS estimator of the spread is given by

$$S_t = \frac{2(e^{\alpha_t} - 1)}{1 + e^{\alpha_t}}, \tag{4}$$

$$\alpha_t = \frac{\sqrt{2\beta_t} - \sqrt{\beta_t}}{3 - 2\sqrt{2}} - \sqrt{\frac{\gamma_t}{3 - 2\sqrt{2}}}, \tag{5}$$

$$\beta_t = (h_{t+1} - l_{t+1})^2 + (h_t - l_t)^2, \tag{6}$$

$$\gamma_t = (\max\{h_{t+1}, h_t\} - \min\{l_{t+1}, l_t\})^2. \tag{7}$$

For a sample of T days, the Corwin and Schultz (2012) measures are obtained by averaging the 2-day estimates:

$$CS_M = \max\left\{\frac{1}{T-1}\sum_{t=1}^{T-1} S_t, 0\right\}, \tag{8}$$

$$CS_D = \frac{1}{T-1}\sum_{t=1}^{T-1} \max\{S_t, 0\}, \tag{9}$$

$$CS_P = \frac{1}{\sum_{t=1}^{T-1}\mathbb{I}_{\{S_t\geq 0\}}}\sum_{t=1}^{T-1} S_t \mathbb{I}_{\{S_t\geq 0\}}, \tag{10}$$

where $\mathbb{I}_{S_t\geq 0}$ is an indicator function that takes the value of one if $S_t \geq 0$. The first measure, CS_M, is the censored version of CS and thus guaranteed to be positive definite. The second measure, CS_D, is constructed by censoring the 2-day spreads at zero before averaging. This version is recommended by Corwin and Schultz for U.S. equities and is one of the most widely used measures in the literature. Finally, the last measure, CS_P, is obtained by only averaging over the non-negative 2-day spreads. Karnaukh, Ranaldo, and Söderlind (2015) use this measure for spot foreign exchange rates, because it is more closely correlated with the true spread than CS_D.

The CS measures are derived under the assumption of continuous trading, that is, the overnight return between two consecutive trading days is assumed to be zero. When this is not the case, such as in most equity markets, CS propose a simple adjustment that mitigates the effect of non-zero overnight returns (Corwin and Schultz, 2012, p. 726). We employ this correction in our empirical applications.

2.5 Abdi and Ranaldo (2017)

Abdi and Ranaldo (2017), henceforth AR, combine the ideas of Roll and CS. They define the daily mid-range by $\eta_t = (h_t + l_t)/2$, the average of the daily high and low log prices, and observe that the statistic

$$\delta_t = 4(p_t - \eta_t)(p_t - \eta_{t+1}), \tag{11}$$

satisfies $\mathrm{E}(\delta_t) = s^2$, Thus, using a sample of T days, they propose to estimate s by

$$AR_M = \sqrt{\max\left\{\frac{1}{T-1}\sum_{t=1}^{T-1}\delta_t, 0\right\}}, \tag{12}$$

$$AR_D = \frac{1}{T-1}\sum_{t=1}^{T-1}\sqrt{\max\{\delta_t, 0\}}, \tag{13}$$

$$AR_P = \frac{1}{\sum_{t=1}^{T-1}\mathbb{I}_{\{\delta_t \geq 0\}}}\sum_{t=1}^{T-1}\sqrt{\delta_t}\mathbb{I}_{\{\delta_t \geq 0\}}, \tag{14}$$

where, similar to CS, AR employ censoring or truncation to achieve nonnegative estimates. Unlike the CS measures, the AR measures are robust to the presence of overnight returns and so no adjustment is required when applying these measures to non-24-hour markets.

3 Statistical properties and the role of volatility

Having described the low-frequency measures, we now document their statistical properties. First, we show by simulation that the low-frequency measures are significantly upward biased when the true spread is small. Second, we study the source of the bias and document its dependence on the volatility of the efficient price. Finally, we study the implications of this dependence for the dynamics of the low-frequency measures when both volatility and spreads vary over time.

3.1 Bias and precision

For comparison with CS and AR, we adopt their "near-ideal" simulation design, except that we consider smaller values of the effective spread, ranging from 5 to 300 basis points, as these values are more consistent with the transaction costs observed in our data and sample period. Specifically, we assume there are 390 transactions per day, the daily volatility of the efficient price innovation equals 3%, and the sample size is either 21 trading days (approximately one month) or 251 trading days (approximately one year). All simulations are based on 10,000 replications.

Table 1 reports the simulation results. We find that all measures are significantly upward biased when the spread is small. For example, when the true spread equals 10 bps, the bias ranges from 35 bps for the CS_M measures to 236 bps for the AR_P measure, or 250% and 2,260% of the true spread, respectively. But for some measures, the bias remains substantial, even for moderate values of s. When the spread is large, the bias is reasonable, and this is known from previous studies as well. The measures based on censoring and truncation before averaging (the "D" and "P" measures) exhibit a much higher bias than measures obtained by censoring after averaging (the "M" measures). Hasbrouck's measures are also highly biased when the spread is small or moderate, and the bias is highly dependent on the prior.

In terms of precision measured by the root mean square error (RMSE), CS_M generally performs the best, but this comes at a cost: a large fraction of the estimates is equal to zero. For example, for $s = 10$ bps, the measure equals zero 38% of the time. The AR_M produces a higher RMSE and many more zeros, even for moderate values of s. The "D" and "P" measures perform better in terms of standard deviation, but the bias dominates and produces inferior RMSE to those of the "M" measures. When the sample size increases from $T = 21$ to $T = 251$, the bias and RMSE of the consistent estimators—the "M" estimators and the Hasbrouck measures—improve significantly, as expected, while those of the "D" and "P" measures do not.

3.2 Source of bias

What drives the bias when the true effective spread is small? Recall that to ensure non-negativity, the Roll, CS, and AR measures are all based on some form of censoring or truncation either before or after averaging. These operations induce a positive bias and dependence on the volatility of the efficient price σ as we now demonstrate.

To develop some intuition, consider a simple case where x_t is *iid* normal with mean μ, $\mu > 0$, and variance ω^2. Of course, the summands ($r_t r_{t-1}$, S_t, δ_t) in the various measures are neither normal nor *iid*, but the normal *iid* case provides informative predictions for the actual behavior of the effective spread measures. It is well-known that in the case of *iid* normal x_t,

$$\mathbb{E}\left(\max\left\{\frac{1}{T}\sum_{t=1}^{T} x_t, 0\right\}\right) = \Phi\left(\frac{\sqrt{T}\mu}{\omega}\right)\mu + \phi\left(\frac{\sqrt{T}\mu}{\omega}\right)\frac{\omega}{\sqrt{T}}, \tag{15}$$

$$\mathbb{E}\left(\frac{1}{T}\sum_{t=1}^{T}\max\{x_t,0\}\right)=\Phi\left(\frac{\mu}{\omega}\right)\mu+\phi\left(\frac{\mu}{\omega}\right)\omega, \tag{16}$$

$$\mathbb{E}\left(\frac{1}{\sum_{t=1}^{T}\mathbb{I}_{\{x_t\geq 0\}}}\sum_{t=1}^{T}x_t\mathbb{I}_{\{x_t\geq 0\}}\right)=\left[\Phi\left(\frac{\mu}{\omega}\right)\right]^{-1}\left[\Phi\left(\frac{\mu}{\omega}\right)\mu+\phi\left(\frac{\mu}{\omega}\right)\omega,\right]. \tag{17}$$

First, these results show that all three forms of imposing non-negativity introduce a bias that depends on both s and ω. Second, only censoring *after* averaging can deliver consistent estimates of the effective spread, equation (15). Recall that an estimator $\hat{\theta}_T$ is consistent for the true parameter θ_0 when for $T\to\infty$, $\hat{\theta}_T\xrightarrow{p}\theta_0$. Censoring or truncation *before* averaging produces a bias that does not vanish as the sample size increases, and hence measures introduced in equations (16) and (17) are not consistent. Finally, truncation delivers a higher bias and greater dependence on volatility than censoring. Henceforth, we refer to "M" measures (CS_M, AR_M, and Roll measures) as "consistent" measures, and "D" and "P" measures (CS_D, CS_P, AR_D, and AR_P) as "inconsistent".

Analogous analytical results for the four low-frequency measures we consider are not available in closed form, so we approximate the expectations by simulation over a grid for s and σ, where both parameters vary between 10 and 500 basis points. The simulation design is otherwise identical to that in the previous subsection. The results are summarized in Figures 1, 2, and 3. In each of these figures, the left panel plots the expectations as a function of volatility for a given spread, while the right panel plots the expectations as a function of the spread for a given level of volatility.

Starting with the Roll measure shown in the top panel of Figure 1, we find that when the spread is large and/or the volatility is small, the Roll measure is fairly insensitive to volatility. However, for small values of s, the dependence on volatility becomes apparent. For example, when the spread equals 50 bps, the measure varies almost linearly with volatility even when volatility is as low as 1%. At the same time, at this level of volatility, the sensitivity to changes in the spread is significantly less than one (right panel); increasing volatility further makes this sensitivity even smaller.

Turning to the CS measures shown in Figure 2, censoring after averaging (CS_M) produces a positive bias that increases with σ, especially for small s, but this bias disappears as the sample size increases (not shown). Censoring before averaging (CS_D) delivers a more pronounced dependence on volatility, regardless of the level of the spread. Truncation before censoring (CS_P) makes the dependence on volatility even stronger. Moreover, when the spread is small, both measures are essentially a linear function of volatility, even for low σ. Also, as shown in the right panel, the sensitivity to s declines as volatility increases. Increasing the sample size does not resolve these issues (not shown).

Finally, the AR measures are shown in Figure 3. We notice two differences relative to CS. First, the dependence on volatility is less pronounced for large values of s. This comes at a cost, however, and that is a lower sensitivity to s when s is small and volatility is high. Similar to CS, when the spread is small or when the volatility is high, the dependence of the AR measures on volatility is essentially linear.

In the case of the Hasbrouck measures, non-negativity is achieved by choosing an appropriate prior. It is difficult to make predictions about how exactly the choice of the prior affects the bias and dependence on volatility, if any. But a simulation can shed some light on this issue. In Figure 1, we plot the expected values of the Hasbrouck measures approximated by simulation as we did for the other measures. We consider two different priors for s, one where we set $\sigma_s^2 = 0.05^2$ as in Hasbrouck's original paper (middle panel), and one where we consider a tighter prior by setting $\sigma_s^2 = 0.01^2$ (bottom panel). In the former case, we find a behavior that is qualitatively similar to that of the Roll measure, although the dependence on volatility is somewhat stronger. In the latter case, the dependence on volatility varies with the size of the spread; for small values of s, the dependence is positive, while for large values of s it is negative. As volatility increases, the posterior increasingly resembles the prior and, regardless of the true spread, the measure produces very similar estimates. As a result, the sensitivity of the measure to changes in the spread becomes very small (right panel).

3.3 Correlation with true spread and volatility

The key finding from the simulations so far is that when the effective spread is small, the variation in the low-frequency measures can be driven by volatility. In practice, both spreads and volatility vary over time and in the cross section, as predicted by microstructure theory, and they tend to be contemporaneously correlated (see Foucault, Pagano, and Röell, 2013, for a textbook treatment).

In the final set of simulations, we therefore explore the implications of time-varying and correlated spread and volatility for the dynamics of the effective spread measures. Specifically, we study the correlation of the spread estimates with the true spreads and volatility by calculating simple pairwise correlations. More importantly, though, we run OLS regressions of the spread estimates on the true spread and volatility:

$$\hat{s}_t = \alpha + \beta_s s_t + \beta_\sigma \sigma_t + u_t. \tag{18}$$

If the low-frequency measures ($\hat{s}_t$) truly capture the effective spread, these regressions would exhibit a spread coefficient β_s close to unity, a volatility coefficient β_σ close to zero, and a high R^2. If, on the other hand, volatility plays an important role in the dynamics of these measures, β_σ will be different from zero, and the R^2 in the bivariate regression in equation (18) will be substantially higher than that in a univariate regression of $\hat{s}_t$ on s_t alone.

Since we cannot calculate the population coefficients in equation (18) in closed form, we approximate them by simulation. As before, we simulate samples of size $T = 21$ (one month) or $T = 252$ (one year) days with $n = 390$ intraday returns each. All simulations are based on 10,000 replications. We keep the spread and volatility fixed within each sample (month or year). However, we allow the spread and volatility to vary randomly across samples (months or years). Both the spread and volatility are assumed to be independently log-normally distributed, either uncorrelated, or contemporaneously correlated with $\rho = 0.5$.[2] We set the mean daily volatility of the efficient price equal to 2% and the volatility of volatility equal to 2%. A 2% average daily volatility is approximately what we observe for the S&P 1500 stocks during our sampling period. We consider three scenarios for the effective spread. First, we consider an effective spread with a mean of 10bps and spread volatility of 10 bps, i.e., the implied ratio of expected spread to expected daily volatility is 5%, which is close to what we observe for the large- and mid-cap stocks in our sample. Second, we consider an effective spread with a mean spread of 50 bps and spread volatility of 50bps, so that the signal-to-noise ratio equals 25%. This value is large for modern electronic

[2] We also run simulations with $\rho = 0.75$. These results are available in Appendix A.

markets, but it is close to what the DJIA 30 stocks experienced in the previous century (see Jones, 2002). Finally, we consider an effective spread of 3% with volatility of 3%, implying a signal-to-noise ratio of 150%. We run this last simulation mainly to show that when the signal-to-noise ratio is high, the consistent low-frequency measures perform as dictated by statistical theory.

Table 2 reports the simulation results for the case of no correlation between the spread and volatility. In the left part of Panel A, we show the results for the small, tranquil spread and $T = 21$ observations (days) used to calculate the low-frequency measures. We find that the correlation with the true spread is very small for all measures, while the correlation with volatility is high, especially for the D and P measures, where the correlation exceeds 0.9. The slope coefficient in a univariate regression on the true spread is well below one and the intercept is well above zero, consistent with the large bias documented in Table 1. The bivariate regressions on the true spread and volatility show that it is almost exclusively volatility that drives the dynamics of the low-frequency measures: the slope coefficient for volatility is well above zero and the R^2 increases from essentially zero to values that range from 25% for the consistent, M, measures to over 90% for the P measures. In the right part of Panel A, we repeat the simulation with $T = 251$ observations. Except for CS_M, we find that the performance does not really improve, and in some cases it actually deteriorates: for the inconsistent D and P measures, the correlation with volatility is now near perfect. Increasing the number of observations reduces the variance of these measures and the bias, which is a function of volatility and does not diminish asymptotically, dominates.

Panel B of Table 2 reports the simulation results for the medium, moderately volatile spread. There are clear improvements in terms of the correlations with the true spread relative to Panel A, but all measures still exhibit significant correlation with volatility. This is also evident in the bivariate regressions, where, with the exception of CS_M, the volatility coefficient is far from zero and the R^2 is much higher than in the univariate regression on the true spread alone. For example, the most widely used CS_D measure exhibits an $R^2 = 0.17$ in the univariate regression and an $R^2 = 0.89$ in the bivariate regression when the number of observations equals 21 days, suggesting that even with a fairly large signal-to-noise ratio of 25%, the dynamics of the measures is mainly determined by volatility and not the true spread.

Finally, in Panel C we report the results for the high, volatile spread. We find that with such a high noise ratio of 150%, the measures generally work as expected, exhibiting a high correlation with the true spread and a slope coefficient close to unity in a univariate regression on the true spread alone (except H_T). The consistent measures CS_M and AR_M work particularly well, even when $T = 21$. The inconsistent measures, D and P, while much less plagued by volatility than in Panels A and B, still exhibit a nontrivial dependence on volatility as evidenced by a correlation between 0.10 and 0.36, respectively, when $T = 21$, and this correlation barely changes as T increases to 251.

While the simulation results with uncorrelated spread and volatility clearly illustrate the impact of volatility on the low-frequency measures, in practice, the spread and volatility tend to be positively correlated. To be able to better interpret our empirical findings, we therefore repeat the simulations setting the correlation between the spread and volatility equal to 0.5 but leaving the simulation design otherwise unchanged. The results are reported in Table 3. We first note how the contemporaneous correlation between the spread and volatility masks the problem with dependence on volatility when one judges performance by correlation with the true spread. Panel A shows that even if the true spread is very small, the correlation between the low-frequency measures and the true spread tends to be quite high. However, when one runs a univariate regression on the true spread, the issue becomes clearly apparent and manifests itself in a slope coefficient that is much larger than one. Moreover, when volatility is added to the regression, the R^2 increases significantly. For example, with $T = 21$, the CS_D measure exhibits an increase in R^2 from 0.26 to 0.88, while the CS_P's R^2 increases from 0.27 to 0.94; the AR_D and AR_P measures exhibit a similar behavior. Increasing T to 251 does not yield any improvement for either set of metrics. This problem does not go away even when we increase signal-to-noise ratio to 25%, as shown in Panel B. Including the volatility in the regression yields a small increase in R^2 for the inconsistent D and P measures only with a large signal-to-noise ratio of 150%, reported in Panel C.

In summary, the simulation results confirm the fact that the inconsistent low-frequency measures are contaminated by volatility, and it is primarily volatility that drives the dynamics of these measures for parametrizations that are empirically relevant in electronic markets. To end on a positive note, we find that the CS_M measure, while being quite noisy when the true spread is small or when the number of observations is small, does deliver almost unbiased estimates for moderate (and larger) spreads and when the number of observations is large. Unfortunately, the measure is almost never applied in practice, as the authors themselves advise against it on the grounds that it exhibits low correlation with the true spread. Our simulations show that while the D and P measures, which are recommended by both CS and AR, are correlated with the true spread, they are correlated with the true spread for the wrong reason—they load on volatility.

4 Estimating transaction costs in equity and foreign exchange markets

4.1 U.S. equities

4.1.1 Data

Our U.S. equity sample consists of the S&P 1500 constituent stocks during the period between October 2003 and December 2017. We focus on the post-decimalization period and the S&P 1500 stocks for several reasons. First, the S&P 1500 index comprises over 90% of U.S. stock market capitalization, and thus our results apply to the vast majority of economically significant U.S. stocks. Second, very small stocks do not necessarily trade often enough (sometimes not even once a day), which necessitates imposing onerous conditions when implementing the low-frequency methods. We do not claim that these stocks do not matter. However, if we insist on incorporating them in our analysis, more robustness checks and alternative treatments for issues related to infrequent trading need to be carefully explored, which we believe is best left for a separate paper. Third, estimating realized volatility, which is a key variable in our empirical analysis, is only possible for sufficiently liquid assets.

The data on the historical S&P 1500 and other index membership come from COMPUSTAT. For each month in our sample, we select all stocks that were included in the S&P 1500 index between the 5th and 25th of the month and obtain daily high, low, close, bid, and ask prices for these stocks from CRSP, matching by stock CUSIP. We use the daily CRSP price data to calculate the low-frequency measures at the stock-month and stock-year level.

Our high-frequency effective spread benchmarks and various realized volatility measures are calculated from the Daily TAQ data accessed via Wharton Research Data Services (http://wrds-web.wharton.upenn.edu/wrds/). We use the Holden and Jacobsen (2014) SAS code, kindly shared by the authors on their web site, to first construct the national best bid and offer (NBBO) data and then calculate daily dollar-volume-weighted percent effective spreads for all stock in our sample. In addition to the filters employed by Holden and Jacobsen (2014), we also remove transactions where the transaction price differs from the prevailing NBBO quote by more than 10%—i.e. the implied effective spread is larger than 20%—before calculating the daily effective spreads. This helps remove obvious outliers, but it does not materially affect our results, as the effective spreads of our stocks are two orders of magnitude smaller on average. Effective spread estimates at the stock-month or stock-year level are then obtained by simply averaging the daily estimates for a given stock and month or year, respectively.

Novel in our paper is the use of volatility in explaining the behavior of the low-frequency measures. We construct high-frequency-based proxies for the daily volatility of the efficient price by calculating 5-minute realized volatilities from the NBBO mid-quotes. Liu, Patton, and Sheppard (2015) and Bollerslev, Hood, Huss, and Pedersen (2018) show that the 5-minute realized volatility tends to perform very well across different asset classes. To ensure that our results are not plagued by market microstructure noise (see Bandi and Russell, 2006, for an in-depth treatment), we also consider the 30-minute realized volatility. Before calculating the daily realized volatilities, we employ some additional data cleaning methods in the spirit of Barndorff-Nielsen, Hansen, Lunde, and Shephard (2009) and Bollerslev, Hood, Huss, and Pedersen (2018). Specifically, we remove (1) quotes where the bid or ask price is missing or equal to zero, or the bid-ask spread is less than or equal to zero; (2) quotes with bid-ask spreads larger than 50 times the median bid-ask spread for a given stock and day; and (3) quotes where the absolute difference between the mid-quote and the median daily mid-quote exceeds 10 times the average value of this difference on a given day. These filters are designed to remove obvious outliers, such as spurious jumps of the bid and ask prices to \$10,000, for example. To calculate monthly realized volatility, we sum the daily realized volatilities within a given month and add the sum of squared overnight returns, that is, returns between 4:00pm and 9:30am the following trading day.

Finally, we merge the stock-month and stock-year variables created from CRSP and TAQ by the stock CUSIP. In order for a stock-month to be included in our final sample, the CRSP and TAQ have to have data for the same number of trading days for a given stock and month. Following Abdi and Ranaldo (2017), we also eliminate stock-months with stock splits or unusally large distributions ($> 20\%$). Finally, we drop May 2010 from our sample, as a well-known flash crash occurred on the 6th of this month and was associated with unusual price and liquidity dynamics in the U.S. equity markets (Kirilenko, Kyle, Samadi, and Tuzun, 2017). This leaves us with 241,236 stock-months. In case of stock-years, a stock-year is included in our sample, as long as the CRSP and TAQ contain data for a given stock for at least 231 (11 months) trading days and the number of days with data differs by no more than 21 (1 month) between CRSP and TAQ. We also eliminate stock-years with stock splits and unusually large distributions ($> 20\%$), and in case of 2010, eliminate May from all stock-years. Thus, the results for 2010 are only based on 11 months of data. This leaves us with 18,239 stock-years.

Table 4 reports daily summary statistics for our final sample. The daily average number of trades in our sample period is 8,606 and the average daily trading volume equals \$71 million. The average quoted spread equals 16.3 basis points, while the average effective spread is 12.4 basis points. The average daily realized volatility stands at 2.17% and the signal-to-noise ratio–the ratio of effective spread to realized volatility–equals a mere 6% on average. The table also reports analogous statistics for the four sub-indices that comprise the S&P 1500 index. The largest stocks (DJIA 30) trade with an effective spread of 4.3 basis points on average and daily realized volatility of 1.42%, and a signal-to-noise ratio of 3.4%. The S&P 500 index constituents have an effective spread of 6.1 basis points, daily realized volatility of 1.83% and a signal-to-noise ratio of 3.7% on average. For the mid caps (S&P 400), these figures stand at 9.1 basis points, 2.06%, and 5%, while for the small caps (S&P 600), they are 20 basis points, 2.55%, and 8.5%, respectively. Thus, even the mid and small caps trade with a fairly small effective spread and a signal-to-noise ratio well below

10%.

4.1.2 Results

We show the empirical results for U.S. stocks in three stages. First, we present the results for the whole sample, i.e., the S&P 1500 stocks. Since there is significant variation in the effective spread and the signal-to-noise ratio in the cross section (Table 4), we then repeat the exercise for the largest stocks, i.e., the DJIA 30 stocks, and the small caps, i.e., S&P 600 stocks.

Full-sample results: Monthly

Starting with the full sample, Panel A of Table 6 summarizes descriptive statistics for the TAQ effective spread and the low-frequency measures pooled over all stock-months in the sample. Clearly, all low-frequency measures exhibit a sizable upward bias relative to the TAQ spread. While the true spread equal 12.4 bps on average, the low-frequency measures average between 26.4 bps for CS_M and 168 bps for H_L. As expected, for the CS and AR measures, the average bias increases when one censors or truncates before averaging the daily estimates within a month; the mean CS_D and AR_D are 76.2 bps and 82.2 bps, respectively, while the mean CS_P and AR_P are 123 bps and 157 bps, respectively. The median and the RMSE exhibit a similar pattern.

In Panel B, we report the cross-sectional average of time-series correlations between the low-frequency measures and the TAQ spread and realized volatility. We find that the consistent measures exhibit low time-series correlation with the TAQ spread, ranging between 0.19 and 0.28, while the inconsistent measures exhibit roughly twice as high correlation with the TAQ spread. At the same time, the inconsistent measures tend to be correlated with realized volatility, and the correlation coefficient range from 0.62 to 0.74, roughly double the correlation of the consistent measures with the realized volatility. In Panel C, we present analogous results for the cross-sectional correlations averaged over time. These correlations exhibit a similar pattern.

In Table 7 we report results of regressions of the low-frequency measures on the TAQ spread and realized volatility. In Panel A, we run pooled OLS regressions and use the Driscoll and Kraay (1998) standard errors, which are robust to heteroscedasticity, serial correlation, and cross-sectional dependence for inference. We find that our estimation results are qualitatively consistent with our simulation results reported in Table 3. First, the realized volatility is highly statistically significant for all measures; second, including RV significantly reduces the estimated coefficient for the TAQ spread and increases the R^2. The increase in the R^2 is particularly pronounced for the inconsistent measures, where it at a minimum doubles the value of the R^2. For example, for the CS_D measure the R^2 goes from 0.26 to 0.66, while for the AR_D measure, it increases from 0.22 to 0.64.

In the next two panels, we examine the variation separately in the time-series and cross-section. Starting with the time-series variation, in Panel B, we report the "within" estimation, that is, pooled OLS on data that were de-meaned at the stock level, which is equivalent to running OLS with stock fixed effects. We again use the Driscoll and Kraay (1998) standard errors for inference. In Panel C, we run cross-sectional regressions for each month in the sample and report the mean cross-sectional parameter estimates together with the associated Newey and West (1987) Student's t-statistics. The two sets of results are qualitatively consistent with the pooled results in Panel A. One notable difference between the time-series and cross-sectional estimations is that the increase in the R^2 is more pronounced for the former.

The results reported so far have been calculate using the 5-minute realized volatility. To check for robustness to this sampling frequency, in Table 24 in Appendix B we rerun our regression using the 30-minute realized volatility. Sampling at lower frequency should alleviate any concerns about the effect of market microstructure noise on the volatility estimates, see Bandi and Russell (2006). As is seen from the table, our results are virtually unchanged by replacing the 5-minute RV by the 30-minute RV and the point estimates, as well as the associated t-statistics, are remarkably close.

Full-sample results: Annual

We now turn to annual results, where all low-frequency measures are calculated using one year's worth of daily data. Table 8 report the annual descriptive statistics. Several differences relative to the monthly results are immediately apparent. First, the bias and RMSE have significantly declined for the consistent measures: the mean H_L declined from 168 to 64.2 bps, the mean CS_M from 26.4 to 18.3 bps, and the AR_M from 58.3 to 46.0 bps. In contrast, the annual inconsistent measures exhibit almost the same bias as the monthly ones. The correlation with the true spread also increased for the consistent measures, especially the time-series correlations (Panel C). For example, the annual CS_M measure exhibits an average time-series correlation with the TAQ spread of 0.64 relative to the monthly 0.37. Similar increases are observed for the other consistent measures as well, but not for the inconsistent ones. These findings are perfectly in line with our simulation results, the asymptotic theory, and show that only the consistent measures can deliver increasingly precise estimates as the sample size increases.

In Table 9 we run the regressions of low-frequency measures on the TAQ spread and realized volatility at the annual frequency. We again find a very different change relative to the monthly results for the consistent and inconsistent measures. The former now exhibit substantially smaller coefficients on realized volatility, and in some cases the realized volatility is only marginally statistically significant, especially in the cross-sectional regressions (Panel C); take, for example, the case of CS_M and AR_M, where the coefficient on RV is essentially zero. In contrast, the inconsistent ("D" and "P") measures exhibit no such changes and, if anything, the R^2 in the annual regressions with RV increase relative to the monthly regressions. The dependence on volatility of the inconsistent measures cannot be simply undone by increasing the sample size.

Results by size

Tables 10–17 report analogous monthly and annual results—descriptive statistics and regression results—for very large stocks (DJIA 30) and small caps (S&P 600). Starting with the monthly results, we find that the realized volatility plays a more important role for the very large stocks: the low-frequency measures tend to be more correlated with RV for these stocks, and adding RV to the regression on the TAQ spread increases the R^2 more (see Tables 10 and 11). But even for the small stocks, the realized volatility is a significant factor. All low-frequency measures exhibit a non-negligible correlation with realized volatility, and the realized volatility is highly significant in both time-series and cross-sectional regressions (see Tables 14 and 15). In Tables 25 and 26 reported in Appendix B we again check for robustness of our results to the choice of the sampling frequency at which we calculate realized volatility. The table shows that using the 30-minute RV in place of the 5-minute

RV leaves our results intact.

Turning to the annual results reported in Tables 12 and 13 for the DJIA 30 stocks and 16 and 17 for the S&P 600 stocks, we find that the bias of the consistent measures improves more for the small caps than for the very large caps. For example, the average CS_M measure drops from 36 bps to 27 bps, which is much closer to the TAQ spread of 20 bps, while the same measure declines from 16 bps to 11 bps for the very large caps, which have a TAQ spread of mere 4 bps on average. So for the very large caps, the bias is almost 156% of the TAQ spread even at the annual frequency. The bias associated with the inconsistent measures does not, of course, change as the sample size increases.

Second, the correlation of the consistent measures with the TAQ spreads increases more for the small caps than for the large caps when moving to the annual frequency. Both of these results are consistent with our simulation results and the intuition developed in equations (15)-(17) that increasing the sample size reduces the bias and increases the precision of the consistent measures more when the signal-to-noise ratio is high. However, the regression results show that even for the small caps and the annual frequency (Table 17), all consistent measures still exhibit statistically significant dependence on volatility both in the cross section and time series. For the inconsistent measures, this dependence actually becomes stronger at the annual frequency, as can be seen from the magnitude of the coefficients on RV, their statistical significance, and the incremental R^2 in the bivariate regressions. This is not surprising; the variance of the measures is reduced in larger samples, but the bias, which is a function of volatility, is not, and hence the measures now have a less noisy relationship with volatility.

In summary, our empirical results for the very large and small caps are perfectly in line with our simulation evidence. The consistent low-frequency measures perform better for less liquid assets, i.e., assets with a higher ratio of transaction costs to volatility, and this performance improves more when increasing the sample size. The inconsistent low-frequency measures, on the other hand, exhibit no such improvement. Even for the small caps and at the annual frequency, however, all measures suffer from the dependence on volatility both in the cross section as well as in the time series.

4.2 Foreign exchange rates

4.2.1 Data

Our foreign exchange sample consists of five major currency pairs for which we have intraday data: EUR/USD, EUR/CHF, EUR/JPY, USD/CHF, USD/JPY. The source of our intraday data is EBS, one of the largest electronic interdealer platforms for trading spot FX. Although other currency pairs are traded on this platform, it is well known that EBS is the primary source of price discovery for the euro, Japanese yen, and Swiss franc, while Reuters is the primary venue for other currencies (Hasbrouck and Levich, 2018). Thus, we exclude these currencies from our analysis because the transaction costs estimates obtained from the EBS data might not be representative of the true costs of trading these currencies. The EBS data contain transactions time stamped to the millisecond and classified as either buyer or seller initiated, and top-of-the-book quotes sampled at regular 100ms intervals. Unlike for the U.S. stocks, we cannot therefore align quotes with trades exactly, and use instead the most recent—rather than the prevailing—mid-quote when calculating dollar-volume-weighted effective spreads, as in Karnaukh, Ranaldo, and Söderlind (2015).[3] We calculate daily realized volatility from 5-minute mid-quotes; Chaboud, Chiquoine, Hjalmarsson, and Loretan (2010) show that the 5-minute sampling frequency is sufficiently low to avoid contamination by microstructure noise. Our sample period is from January 2008 to December 2015. We drop holidays, and in case of the exchange rates involving the Swiss Franc, we also drop the months when the exchange rate floor was introduced and abandoned by the Swiss National Bank, as these were months of unusual market developments.

Table 5 reports summary statistics for the foreign exchange data. The EUR/USD and USD/JPY are the most frequently traded currency pairs, with around 39,000 and 21,000 trades per day on average, and \$52 billion and \$28 billion in average daily volume, respec-

[3] An additional source of measurement error stems from the fact that trading on EBS takes place in three different locations—London, New York, and Tokyo—and the associated latencies may give rise to inaccuracies in trade time stamps. Addressing this issue directly requires information on the geographical location of trades, but such data are currently not available. As a result, the effective spread may be negative for some trades. In Appendix C, we consider two alternative ways of constructing the dollar-volume-weighted effective spreads for robustness and show that our results remain unchanged.

tively. The other currency pairs are considerably less frequently traded. The average quoted spread is very small for all currencies, as is the effective spread. The most liquid currency pair is the EUR/USD with an average quoted spreads of 1.07 bps and an effective spread of 0.50 bps; USD/JPY exhibits slightly higher transaction costs, 1.45 bps and 0.68 bps, respectively; the remaining currency pairs trade with an average quoted spread between 2.6 and 2.8 bps and an average effective spread between 0.69 and 0.83 bps. The average daily realized volatility ranges from 45 bps for EUR/CHF and 85 bps for EUR/JPY. These values are almost two orders of magnitude larger that the average effective spreads, producing signal-to-noise ratios between 1 and 2%. In this sense, the spot foreign exchange rates are even more liquid than the DJIA 30 stocks discussed in the previous section.

Following Karnaukh, Ranaldo, and Söderlind (2015), we obtain daily high, low, and closing prices for the five exchange rates from Thompson Reuters. These prices are based on quotes rather than transactions, and to the best of our knowledge, daily transaction prices are not publicly available. This means that we can only calculate the Corwin and Schultz (2012) effective spread measures from these data, since other measures explicitly require daily closing transaction prices; the effective spread is otherwise not identified. We calculate the Roll, Hasbrouck, and Abdi and Ranaldo measures from the daily high, low, and closing prices extracted from our intraday EBS data. We acknowledge that this is not feasible for researchers who do not have access to the proprietary EBS data, but it is nonetheless instructive to study how these measures would perform if the daily transaction-based data were publicly available.

4.2.2 Results

Table 18 summarizes our foreign exchange results. Similar to the analysis of the U.S. equities, we calculate statistics and run regressions for FX-month data pooled across the five currency pairs in our sample. We do not perform the analysis at the annual frequency because we only have five exchange rates and 8 years in our sample, i.e., only 40 exchange rate-years, which is too small a sample for any meaningful statistical analysis. Starting with the summary statistics reported in Panel A, we find that all low-frequency measures are severely upward biased. The true effective spread is 0.68 bps on average, while the average

Roll measure equals 25.9 bps, the Hasbrouck measures vary between 44.8 and 46.3 bps depending on the prior, the Corwin and Schultz average measures range from 13.3 to 44.7 bps, and the Abdi and Ranaldo from 13.7 to 47.6 bps. The standard deviations of the low-frequency measures are also two orders of magnitude higher than that of the true spread. These results are in stark contrast to those of Karnaukh, Ranaldo, and Söderlind (2015), who report a much smaller bias for their currencies and sample period (2007-2012). When replicating their results, we find that they report the summary statistics in percent rather than basis points, as claimed. For example, Karnaukh, Ranaldo, and Söderlind (2015) report a mean CS_P for EUR/USD of 0.476 basis points,[4] while our replication shows that it is actually 0.476%, or 47.6 basis points. Given that the effective spreads from EBS data is reported to be 0.584 basis points on average in Karnaukh, Ranaldo, and Söderlind (2015), the bias is actually very large, and very similar to what we obtain for EUR/USD during our sample period (2008-2015).

Panel B of Table 18 reports regression results from regressing the low-frequency measures on the true effective spread. We find that the intercept is statistically indistinguishable from zero, but the slope coefficient is far from unity, as would be required of a well-performing measure. The large slope coefficients reflect, of course, the large bias the low-frequency measures suffer from. The coefficient of determination is generally not very large, reaching a maximum of 0.32 for the CS_D measure. In Panel C of Table 18, we add realized volatility to the regression. Echoing the results for the DJIA 30 stocks, three main results emerge from these regressions: (i) realized volatility is highly statistically significant in all regressions except for CS_M; (ii) compared to the univariate regressions in Panel B, the R^2 more than doubles for all measures except for consistent M measures (those based on censoring after averaging); and (iii) for the Hasbrouck measures and the "P" measures, adding realized volatility renders the spread coefficient (β_{ES}) statistically insignificant.

In summary, volatility is an important driver of the Hasbrouck and the "D" and "P" measures in the foreign exchange market, explaining why these measures exhibit a higher

[4]Page 3080, Table 2, "Corwin-Schultz high-low estimate/2 (LF), bps". Since we focus on the full spread in our paper, while Karnaukh, Ranaldo, and Söderlind (2015) report half-spread, multiplying their 0.238 by two produces 0.476.

correlation with the true spread documented in Panel A and in previous studies. Volatility affects the "M" measures by a much smaller degree. As a result, these measures exhibit a much smaller correlation with the true spread.

5 Implications for empirical asset pricing

Thus far, we have documented the statistical properties of the low-frequency measures. In this section, we highlight the implications of these properties for empirical asset pricing. We provide two case studies. In the first case, we investigate how the imprecision of the low-frequency measures affects the estimates of betas and market price of risk in the liquidity-adjusted CAPM of Acharya and Pedersen (2005). In the second case, we follow Ang, Hodrick, Xing, and Zhang (2006) and add a liquidity factor to their unconditional factor model with market and aggregate volatility risks. Our goal is to examine if and to what extent the dependence of the low-frequency measures on volatility affects the estimation of liquidity and volatility betas and the associated prices of risk if one employs low-frequency measures to construct an aggregate liquidity factor.

5.1 Liquidity-adjusted CAPM

We assume that stock and market returns follow:

$$r_t^i - s_t^i = \mu_i + \beta_i(r_t^m - s_t^m) + \epsilon_t^i, \tag{19}$$

$$r_t^m - s_t^m = \mu_m + \sigma_t^m \epsilon_t^m, \tag{20}$$

where r_t^i and r_t^m are the gross stock and market returns, and s_t^i and s_t^m are the stock and market transaction costs. Without loss of generality, we assume that the risk-free rate is zero. We assume that stock idiosyncratic volatility is constant, but this assumption can be easily relaxed and a factor structure assumed following Herskovic, Kelly, Lustig, and Van Nieuwerburgh (2018). Acharya and Pedersen (2005) decompose the beta in equation

(19) into four separate betas, $\beta_i = (\beta_i^{(1)} + \beta_i^{(2)} - \beta_i^{(3)} - \beta_i^{(4)})$, where

$$\beta_i^{(1)} = \frac{\mathrm{Cov}(r_t^i, r_t^m)}{\mathrm{Var}(r_t^m - s_t^m)}, \beta_i^{(2)} = \frac{\mathrm{Cov}(s_t^i, s_t^m)}{\mathrm{Var}(r_t^m - s_t^m)}, \beta_i^{(3)} = \frac{\mathrm{Cov}(r_t^i, s_t^m)}{\mathrm{Var}(r_t^m - s_t^m)}, \beta_i^{(4)} = \frac{\mathrm{Cov}(s_t^i, r_t^i)}{\mathrm{Var}(r_t^m - s_t^m)}. \tag{21}$$

Thus, the expected return is:

$$\mathrm{E}(r_t^i) = \mathrm{E}(s_t^i) + \lambda_m(\beta_i^{(1)} + \beta_i^{(2)} - \beta_i^{(3)} - \beta_i^{(4)}). \tag{22}$$

The asset-pricing model in equations (19) and (20) is written in terms of monthly returns. However, to calculate the various low-frequency effective spread measures, we need to simulate intraday returns. To do so, we continue with our simulation design of Section 3 and assume that the transaction costs and market volatility are constant within each month and divide the month into $n_T = T \times n$ periods each, where T is the number of days in the month and n is the number of intraday periods as before. Consistent with equations (19) and (20), we assume that gross returns at the intraday level follow:

$$r_{tj}^i - s_t^i/n_T = \beta_i(r_{tj}^m - s_t^m/n_T) + (\sigma^i/\sqrt{n_T})\epsilon_{tj}^i, \tag{23}$$

$$r_{tj}^m - s_t^m/n_T = \mu_m/n_T + (\sigma_t^m/\sqrt{n_T})\epsilon_{tj}^m, \tag{24}$$

where $j = 1, ..., n_T$ and ϵ_{tj}^i and ϵ_{tj}^m are independent standard normal random sequences. Aggregating to the monthly frequency by summing up the intraday returns within the month reproduces equations (19) and (20). The intraday efficient prices and transaction prices are then generated as in the Roll (1984) model:

$$m_{t-j/n_T}^i = m_0 + \sum_{l=1}^{j} r_{tj}^i, \tag{25}$$

$$p_{t-j/n_T}^i = m_{t-j/n_T}^i + \frac{s_t^i}{2} q_{t-j/n_T}^i, \tag{26}$$

where q_{t-j/n_T}^i is an independent sequences of binary trade indicators. Once we generate the intraday transaction prices for a given month, we then compute the high, low, and closing prices for each day in the month, followed by the low-frequency spread measures and realized volatilities for each stock. Armed with the monthly time-series of effective spread estimates and stock and market returns, we first estimate the time-series regression in equation (19) separately for each stock. In the second step, we run a cross-sectional regression of the stock net returns on the estimated betas to recover the market price of risk. We do this exercise separately for each low-frequency measure. That is, we substitute s_t^i by a low-frequency measures and substitute s_t^m by the cross-sectional average of the measure.

We calibrate the simulations as follows. In each replication, we generate 360 months (or 30 years) worth of intraday data for 26 stocks. As in the previous simulations, we assume that each month has $T = 21$ trading days with $n = 390$ transactions each. We set the gross expected return of the market (μ_m) equal to 5% per annum and let the stocks' betas to be equidistantly spaced between 0.5 and 1.5. The market transaction costs and volatility are *iid* and jointly log-normally distributed with correlation of 0.5.[5] The mean and volatility of the market spread are set to 10 basis points each, and the mean and volatility of daily market volatility are set to 1% each; the ratio of expected spread to expected daily volatility is thus 10%. We posit that individual stocks' transaction costs follow a factor model:

$$s_t^i = \beta_i s_t^m u_t^i, \quad i = 1, ..., 26, \tag{27}$$

where u_t^i is an independent log-normal innovation with unit mean and standard deviation equal to 0.5. Since the average beta equals one, the cross sectional mean of the stocks' spread equals the mean of the market spread, i.e., 10 bps, but the high beta stocks have a higher effective spread and the low beta stock have a lower effective spread. Finally, to calibrate the idiosyncratic volatility, σ^i, we observe that conditional on σ_t^m, s_t^m, and s_t^i, the variance of stock returns in equation (19) is $(\sigma_t^i)^2 = \beta_i^2(\sigma_t^m)^2 + (\sigma^i)^2$. We set $(\sigma^i)^2 = 0.25\beta_i^2\mathrm{E}((\sigma_t^m)^2)$ so that, on average, the idiosyncratic variance accounts for 20% of the total stock variation. With this choice, the stock beta also governs the total stock volatility, with higher beta stocks having higher total volatility. Moreover, the ratio of the expected stock spread to the expected stock volatility is constant across stocks. We run 10,000 replications.

We report the simulation results in Table 19. To save space, we only report three sets of

[5] We also run simulations with $\rho = 0.75$ and these simulations are reported in Appendix A.

estimated betas and the associated decompositions—small (0.5), medium (1.0), and large (1.5). We report the true values of the four Acharya and Pedersen (2005) liquidity betas in the first column. The second column reports the (infeasible) estimation results when we use the true spread (s) to estimate the liquidity-adjusted CAPM. Clearly, using the true spread yields estimates which are, on average, very close to the true values of the betas and the market price of risk, λ_m. The remaining columns report the results for the low-frequency measures. While, on average, the estimated β_1 is very close to the true value, this is not the case for the remaining betas. For β_2, the bias is uniformly positive and very large, ranging from 5 times to 54 times the actual value for CS_M and AR_P, respectively. For β_3 and β_4, the H_L measure yields a large positive bias, while the bias associated with the other measures is either positive or negative depending on the form of censoring or truncation. The market price of risk estimates, reported in Panel D, exhibit sizable negative biases. For the consistent measures CS_M and AR_M, the bias is around -1.1% and -2.3%. The bias gets worse when the censoring is done before averaging: -5.6% for CS_D and -5.1% for AR_D. Still, moving from censoring to truncation before averaging yields even more biased estimates of market price of risk. On average, the price of risk is estimated at -5% by CS_P and -6.2% by AR_P. The H_L measures produces a similarly biased price of risk estimates, about -4.6%. In summary, these simple simulations show that using low-frequency liquidity measures to estimate the liquidity-adjusted CAPM could easily lead to misleading inferences about both the size and sign of the market price of risk and the liquidity betas.

5.2 Asset pricing with aggregate liquidity and volatility risks

As mentioned earlier, we use a variation of Ang, Hodrick, Xing, and Zhang (2006) model to demonstrate the potentially spurious inference resulting from using low-frequency measures of liquidity in settings similar to theirs. We assume that stock and market returns evolve according to:

$$r_t^i - \mu_i = \beta_i^m (r_t^m - \mu_m) + \beta_i^l (l_t - \bar{l}) + \beta_i^\sigma (\sigma_t^m - \bar{\sigma}_m) + \epsilon_t^i, \tag{28}$$

$$r_t^m - \mu_m = \sigma_t^m \epsilon_t^m, \tag{29}$$

where μ_i is stock i's expected return, $\bar{l} = \mathbb{E}(l_t)$ is the average liquidity, and $\bar{\sigma}_m = \mathbb{E}(\sigma_t^m)$ is the average market volatility. In the absence of arbitrage, the expected return is given by

$$\mathbb{E}(r_t^i) = \beta_i^m \lambda_m + \beta_i^l \lambda_l + \beta_i^\sigma \lambda_\sigma, \tag{30}$$

where λ_m, λ_l, and λ_σ are the market, liquidity, and volatility prices of risk, respectively.

Consistent with equations (28) and (32), we assume that at the intraday level, stock and market returns follow:

$$r_{tj}^i = \mu_i/n_T + \beta_i^m (r_{tj}^m - \mu_m/n_T) + \beta_i^l (l_t - \bar{l})/n_T + \beta_i^\sigma (\sigma_t^m - \bar{\sigma}^m)/n_T + (\sigma^i/\sqrt{n_T})\epsilon_{tj}^i, \tag{31}$$

$$r_{tj}^m = \mu_m/n_T + (\sigma_t^m/\sqrt{n_T})\epsilon_{tj}^m, \tag{32}$$

$j = 1, ..., n_T$, ϵ_{tj}^i and ϵ_{tj}^m are independent standard normal random sequences. Aggregating to the monthly frequency by summing up the intraday returns within the month reproduces equations (28) and (32).

We set the liquidity factor in equation (32) according to $l_t = 10\sqrt{T} s_t^m$, that is, the liquidity factor equals the common factor driving the stock effective spreads in equation (27), adjusted by a factor of $10 \times \sqrt{T}$ in order to make the liquidity and volatility factors of the same order of magnitude. Recall that σ_t^m is the monthly stock volatility and the expected daily volatility equals 10 times the average effective spread. s_t^m and σ_t^m are jointly log-normal with the same parameters as in the previous subsection, s_t^i follows the factor model in equation (27), and σ^i is also set as in the previous subsection. The intraday efficient prices and transaction prices evolve according to equations (25) and (26), with r_{tj}^i and r_{tj}^m given in by equations (31) and (32), respectively.

We consider five different market betas, ranging from 0.5 to 1.5, five different liquidity betas, ranging from -2 to 2, and five different volatility betas, ranging from -2 to 2. Thus, we generate data for $5 \times 5 \times 5 = 125$ stocks, one for each combination of the three betas. We consider three scenarios for the risk prices. In the first scenario all three risks are priced with $\lambda_m = 5\%$, $\lambda_s = \lambda_\sigma = -1\%$ per annum. In the second scenario, only the market and volatility risk are priced, i.e. $\lambda_m = 5\%$, $\lambda_\sigma = -1\%$, and $\lambda_s = 0$. In the third scenario, only the market and liquidity risks are prices, i.e. $\lambda_m = 5\%$, $\lambda_s = -1\%$, and $\lambda_\sigma = 0$.

In each replication, we generate 360 months of data with 21 days in each month and 390 returns for each day for each of the 125 stocks. Once the data are generated, we calculate the low-frequency measures for each stock and month. For each low-frequency measure, we then run the usual two-stage asset pricing exercise, where we use the measure's cross-sectional average, multiplied by $10 \times \sqrt{T}$, as the aggregate liquidity factor in equation (28). To proxy for the aggregate volatility factor, we use either the monthly realized volatility based on daily closing values of the market factor or the monthly range-based volatility estimates based on the daily high and low values of the market factor (Parkinson, 1980). Thus, we only use daily data to estimating the market volatility. In the first stage, we estimate the time-series regression in equation (28) for each stock, while in the second stage, we estimate the cross-sectional regression in equation (30) using the first-stage beta estimates in order to estimate the prices of risk. We repeat this 10,000 times and report the average prices of risk obtained in the simulation.

Table 20 summarizes the results. In Panel A, we use the realized volatility as a proxy for aggregate volatility, while in Panel B, we use the range-based volatility. The first column reports the true values of the price of risk parameter values. In the second column, we show that using the true effective spread in the estimation yields prices of risk that are on average very close to the true values, regardless of the volatility proxy used.

Turning to the results for the low-frequency measures, we find that using these measures produces large biases in estimated prices of risk, which can lead to spurious inferences about the sign and the magnitude of the risk prices. First, when only the volatility risk is priced, i.e., the price of liquidity risk is zero, all measures produce a non-zero price of liquidity risk on average, and the price of risk is typically large and negative. For example, the CS_D and AR_D measures yield a price of liquidity risk of -2.22% and -6.86%, respectively, when daily realized volatility is used, and 2.56% and -7.20% when the daily range is used. Second, when only liquidity risk is priced, that is when the price of volatility risk is zero, most measures produce significantly biased liquidity risk prices and often non-trivial volatility risk prices on average. Also, the sign of the estimated liquidity risk premium varies across the different measures. For example, the AR measures yield on average positive prices of liquidity risk ranging from 0.81% to 5.93% even though the true value is -1%, while the

CS measures produce uniformly negative prices of liquidity risk ranging from -4.18% to -11.1%. Finally, when both volatility and liquidity risks are priced in the cross section, all measures produce sizable negatively biased estimates of liquidity risk premia, while the volatility prices of risks are on average quite close to the true value.

6 Conclusion

Many financial decisions crucially depend on accurate estimates of transaction costs. The availability of long histories and the relative ease of handling daily data motivate researchers and practitioners to employ low-frequency transaction cost measures. However, as we show in this study, a number of well-known low-frequency measures are biased and inconsistent. This bias is significant, positive, stems from the construction of these measures, is a function of volatility, and hence it induces a positive correlation between the low-frequency measures and volatility above and beyond what is implied by the correlation between volatility and the true transaction costs. The relative size of this bias is particularly large for liquid assets such as S&P 1500 stocks or heavily traded foreign currencies.

Through careful simulation, we document the properties and problems of several popular low-frequency measures when they are used in analyzing highly liquid assets, where the size of transaction costs relative to volatility is small. We then document the problems that arise, including biased or outright spurious results, when one uses these measures in asset pricing applications. Often, proponents of these measures point to their applicability for historical data. However, at least two studies (Jones, 2002 and Bessembinder, 1994) report results that point to trading costs that are much lower than those implied by low-frequency measures for the largest, most liquid U.S. equities and exchange rates as early as 1920s and early 1980s, respectively. We do not claim that low-frequency measures are not applicable in general. They are useful for illiquid assets, where trading costs are relatively large in comparison to volatility. Examples may include high-yield corporate debt or infrequently traded equities.

Awareness of potential pitfalls of using the low-frequency measures of transaction costs matters in practice. Similar to the estimation of mean of daily equity returns, recovering liquidity measures—which are small, positive-valued quantities—from noisy transaction data is challenging. For highly liquid assets, using high-frequency transactions data is probably the only reasonable way for estimation of accurate liquidity measures; see Chordia, Sarkar, and Subrahmanyam (2005) and Mancini, Ranaldo, and Wrampelmeyer (2013). If one must use a low-frequency measure, then as we show in this study, for less liquid assets the consistent version of Corwin and Schultz (2012) performs better than other competing measures, provided that one uses a fairly long window of data (at least one year) for constructing the measure.

References

ABDI, F., AND A. RANALDO (2017): "A Simple Estimation of Bid-Ask Spreads from Daily Close, High, and Low Prices," *Review of Financial Studies*, 30(12), 4437–4480.

ACHARYA, V. V., AND L. H. PEDERSEN (2005): "Asset pricing with liquidity risk," *Journal of Financial Economics*, 77(2), 375 – 410.

ANDERSEN, T. G., T. BOLLERSLEV, AND F. X. DIEBOLD (2010): *Parametric and Nonparametric Volatility Measurement*chap. 2, pp. 67–128, Handbook of Financial Econometrics. Elsevier Science B.V., Amsterdam.

ANDERSEN, T. G., T. BOLLERSLEV, F. X. DIEBOLD, AND P. LABYS (2003): "Modelling and Forecasting Realized Volatility," *Econometrica*, 71, 579–625.

ANG, A., R. J. HODRICK, Y. XING, AND X. ZHANG (2006): "The cross-section of volatility and expected returns," *The Journal of Finance*, 61(1), 259–299.

BANDI, F. M., AND J. R. RUSSELL (2006): "Separating microstructure noise from volatility," *Journal of Financial Economics*, 79(3), 655 – 692.

BANSAL, R., AND A. YARON (2004): "Risks for the long run: A potential resolution of asset pricing puzzles," *Journal of Finance*, 59, 14811509.

BARNDORFF-NIELSEN, O., P. R. HANSEN, A. LUNDE, AND N. SHEPHARD (2009): "Realized kernels in practice: trades and quotes," *Econometrics Journal*, 12(3), 1–32.

BESSEMBINDER, H. (1994): "Bid-ask spreads in the interbank foreign exchange markets," *Journal of Financial Economics*, 35(3), 317–348.

BOLLERSLEV, T., B. HOOD, J. HUSS, AND L. H. PEDERSEN (2018): "Risk Everywhere: Modeling and Managing Volatility," *Review of Financial Studies*, 31(7), 2729–2773.

CHABOUD, A. P., B. CHIQUOINE, E. HJALMARSSON, AND M. LORETAN (2010): "Frequency of observation and the estimation of integrated volatility in deep and liquid financial markets," *Journal of Empirical Finance*, 17, 212–240.

CHEN, A. Y., AND M. VELIKOV (2018): "Accounting for the Anomaly Zoo: A Trading Cost Perspective," *Available at SSRN: https://papers.ssrn.com/abstract=3073681.*

CHORDIA, T., R. ROLL, AND A. SUBRAHMANYAM (2000): "Commonality in liquidity," *Journal of Financial Economics*, 56(1), 3–28.

CHORDIA, T., A. SARKAR, AND A. SUBRAHMANYAM (2005): "An Empirical Analysis of Stock and Bond Market Liquidity," *Review of Financial Studies*, 18(1), 85–129.

CONSTANTINIDES, G. (1986): "Capital market equilibrium with transaction costs," *Journal of Political Economy*, 94, 842–862.

CORWIN, S. A., AND P. SCHULTZ (2012): "A Simple Way to Estimate Bid-Ask Spreads from Daily High and Low Prices," *Journal of Finance*, 67(2), 719–760.

Driscoll, J. C., and A. C. Kraay (1998): "Consistent covariance matrix estimation with spatially dependent panel data," *Review of Economics and Statistics*, 80(4), 549–560.

Farre-Mensa, J., and A. Ljungqvist (2016): "Do Measures of Financial Constraints Measure Financial Constraints?," *Review of Financial Studies*, 29(2), 271–308.

Fong, K. Y., C. W. Holden, and C. A. Trzcinka (2017): "What are the best liquidity proxies for global research?," *Review of Finance*, 21(4), 1355–1401.

Foucault, T., M. Pagano, and A. Röell (2013): *Market liquidity: theory, evidence, and policy.* Oxford University Press.

Goyenko, R. Y., C. W. Holden, and C. A. Trzcinka (2009): "Do liquidity measures measure liquidity?," *Journal of Financial Economics*, 92(2), 153 – 181.

Greene, W. H. (2017): *Econometric Analysis.* Pearson, New York, N.Y., 8 edn.

Hasbrouck, J. (2009): "Trading Costs and Returns for U.S. Equities: Estimating Effective Costs from Daily Data," *Journal of Finance*, 64(3), 1445–1477.

Hasbrouck, J., and R. M. Levich (2018): "FX Market Metrics: New Findings Based on CLS Bank Settlement Data," *Working Paper, Stern School of Business, New York University.*

Hasbrouck, J., and D. J. Seppi (2001): "Common factors in prices, order flows, and liquidity," *Journal of Financial Economics*, 59(3), 383–411.

Herskovic, B., B. T. Kelly, H. N. Lustig, and S. Van Nieuwerburgh (2018): "Firm Volatility in Granular Networks," *Chicago Booth Research Paper No. 12-56; Fama-Miller Working Paper.*

Holden, C. W. (2009): "New low-frequency spread measures," *Journal of Financial Markets*, 12(4), 778 – 813.

Holden, C. W., and S. Jacobsen (2014): "Liquidity Measurement Problems in Fast, Competitive Markets: Expensive and Cheap Solutions," *Journal of Finance*, 69(4), 1747–1785.

Jones, C. M. (2002): "A Century of Stock Market Liquidity and Trading Costs," *Working Paper, Graduate School of Business, Columbia University.*

Karnaukh, N., A. Ranaldo, and P. Söderlind (2015): "Understanding FX Liquidity," *Review of Financial Studies*, 28(11), 3073–3108.

Karolyi, G. A., K.-H. Lee, and M. A. Van Dijk (2012): "Understanding commonality in liquidity around the world," *Journal of Financial Economics*, 105(1), 82–112.

Kirilenko, A., A. S. Kyle, M. Samadi, and T. Tuzun (2017): "The flash crash: High-frequency trading in an electronic market," *The Journal of Finance*, 72(3), 967–998.

Korajczyk, R. A., and R. Sadka (2008): "Pricing the commonality across alternative measures of liquidity," *Journal of Financial Economics*, 87(1), 45–72.

LESMOND, D. A., J. P. OGDEN, AND C. A. TRZCINKA (1999): "A New Estimate of Transaction Costs," *the Review of Financial Studies*, 12(5), 1113–1141.

LETTAU, M., AND S. C. LUDVIGSON (2005): "Expected returns and expected dividend growth," *Journal of Financial Economics*, 76(3), 583 – 626.

LIU, L. Y., A. J. PATTON, AND K. SHEPPARD (2015): "Does anything beat 5-minute RV? A comparison of realized measures across multiple asset classes," *Journal of Econometrics*, 187(1), 293–311.

MANCINI, L., A. RANALDO, AND J. WRAMPELMEYER (2013): "Liquidity in the Foreign Exchange Market: Measurement, Commonality, and Risk Premiums," *Journal of Finance*, 68(5), 1805–1841.

MARSHALL, B. R., N. H. NGUYEN, AND N. VISALTANACHOTI (2011): "Commodity liquidity measurement and transaction costs," *The Review of Financial Studies*, 25(2), 599–638.

MERTON, R. C. (1980): "On estimating the expected return on the market: An exploratory investigation," *Journal of Financial Economics*, 8(4), 323 – 361.

NEWEY, W. K., AND K. D. WEST (1987): "A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica*, 55(3), 703–708.

NOVY-MARX, R., AND M. VELIKOV (2015): "A taxonomy of anomalies and their trading costs," *The Review of Financial Studies*, 29(1), 104–147.

PARKINSON, M. (1980): "The Extreme Value Method for Estimating the Variance of the Rate of Return," *Journal of Business*, 53(1), 61–65.

PATTON, A., AND B. M. WELLER (2018): "What You See is Not What You Get: The Costs of Trading Market Anomalies," *Working Paper, Duke University.*

ROLL, R. (1984): "A Simple Implicit Measure of the Effective Bid-Ask Spread in an Efficient Market," *Journal of Finance*, 39(4), 1127–1139.

SCHESTAG, R., P. SCHUSTER, AND M. UHRIG-HOMBURG (2016): "Measuring Liquidity in Bond Markets," *Review of Financial Studies*, 29(5), 1170–1219.

SPIEGEL, M., AND H. ZHANG (2013): "Mutual fund risk and market share-adjusted fund flows," *Journal of Financial Economics*, 108(2), 506 – 528.

Table 1: Monte Carlo simulation of the effective spread estimators. The simulation design follows the "Near-Ideal" conditions of Corwin and Schultz (2012), who set $n = 390$ and $\sigma = 3\%$. We consider sample sizes of $T = 21$ days (approximately one trading month) and $T = 251$ (approximately one trading year). For each low-frequency measure, "Mean" is the average estimate in the simulation, "Std" is the corresponding standard deviation of the estimates, "RMSE" the root mean square error, and "≤ 0" the proportion of non-positive estimates in the simulation. The results are based on 10,000 replications.

		$T = 21$									$T = 251$								
s		R_M	H_L	H_T	CS_M	CS_D	CS_P	AR_M	AR_D	AR_P	R_M	H_L	H_T	CS_M	CS_D	CS_P	AR_M	AR_D	AR_P
0.05	Mean	1.15	2.03	1.47	**0.34**	1.21	2.06	0.64	1.18	2.36	0.63	0.99	0.91	**0.18**	1.21	2.06	0.35	1.18	2.36
	Std	1.36	0.64	0.41	0.40	**0.31**	0.37	0.75	0.35	0.47	0.72	0.35	0.31	0.15	**0.09**	0.10	0.40	0.10	0.13
	RMSE	1.75	2.08	1.47	**0.50**	1.20	2.05	0.95	1.18	2.36	0.93	1.00	0.91	0.19	1.17	2.01	0.50	1.14	2.31
	≤ 0	0.50	0.00	0.00	0.38	0.00	0.00	0.50	0.00	0.00	0.49	0.00	0.00	0.18	0.00	0.00	0.50	0.00	0.00
0.1	Mean	1.14	2.03	1.47	**0.35**	1.22	2.07	0.64	1.18	2.36	0.62	0.99	0.91	**0.19**	1.22	2.07	0.35	1.18	2.36
	Std	1.36	0.64	0.41	0.41	**0.31**	0.37	0.75	0.35	0.48	0.73	0.35	0.31	0.15	**0.09**	0.10	0.40	0.10	0.13
	RMSE	1.71	2.04	1.43	**0.48**	1.16	2.00	0.93	1.14	2.31	0.89	0.96	0.87	**0.18**	1.13	1.98	0.47	1.09	2.27
	≤ 0	0.51	0.00	0.00	0.38	0.00	0.00	0.50	0.00	0.00	0.50	0.00	0.00	0.15	0.00	0.00	0.49	0.00	0.00
0.2	Mean	1.15	2.04	1.47	**0.39**	1.27	2.12	0.66	1.19	2.37	0.62	1.00	0.91	**0.25**	1.26	2.12	0.36	1.19	2.37
	Std	1.36	0.64	0.41	0.43	**0.31**	0.38	0.75	0.35	0.48	0.73	0.35	0.31	0.16	**0.09**	0.10	0.41	0.10	0.13
	RMSE	1.66	1.95	1.33	**0.47**	1.11	1.95	0.88	1.05	2.22	0.84	0.87	0.78	**0.17**	1.07	1.92	0.44	0.99	2.17
	≤ 0	0.49	0.00	0.00	0.33	0.00	0.00	0.49	0.00	0.00	0.50	0.00	0.00	0.08	0.00	0.00	0.48	0.00	0.00
0.5	Mean	1.20	2.06	1.48	**0.59**	1.44	2.29	0.72	1.22	2.40	0.69	1.03	0.94	**0.52**	1.43	2.28	0.48	1.22	2.41
	Std	1.39	0.65	0.41	0.50	**0.34**	0.38	0.78	0.36	0.49	0.75	0.36	0.32	0.17	**0.09**	0.11	0.44	0.10	0.13
	RMSE	1.55	1.69	1.06	**0.51**	0.99	1.83	0.81	0.80	1.97	0.77	0.65	0.55	**0.17**	0.94	1.79	0.44	0.72	1.91
	≤ 0	0.49	0.00	0.00	0.19	0.00	0.00	0.46	0.00	0.00	0.46	0.00	0.00	0.00	0.00	0.00	0.37	0.00	0.00
1.0	Mean	1.33	2.14	1.52	**1.00**	1.74	2.59	0.93	1.31	2.53	0.93	1.17	1.06	**0.99**	1.74	2.58	0.91	1.32	2.53
	Std	1.46	0.67	0.42	0.59	**0.38**	0.40	0.86	0.38	0.50	0.81	0.42	0.37	0.18	**0.11**	0.11	0.43	0.11	0.14
	RMSE	1.50	1.33	0.67	0.59	0.83	1.64	0.86	**0.49**	1.61	0.82	0.45	0.38	**0.18**	0.75	1.58	0.44	0.34	1.53
	≤ 0	0.46	0.00	0.00	0.06	0.00	0.00	0.37	0.00	0.00	0.34	0.00	0.00	0.00	0.00	0.00	0.10	0.00	0.00
3.0	Mean	2.61	2.94	1.89	**2.92**	3.22	3.84	2.90	2.40	3.54	**2.93**	2.76	2.52	2.92	3.22	3.84	3.00	2.41	3.54
	Std	1.90	0.89	0.54	0.63	0.50	**0.47**	0.75	0.53	0.55	0.63	0.61	0.63	0.18	0.14	**0.13**	0.19	0.15	0.16
	RMSE	1.94	0.90	1.23	0.64	**0.55**	0.96	0.75	0.79	0.77	0.64	0.65	0.80	**0.19**	0.26	0.85	0.19	0.61	0.56
	≤ 0	0.23	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Table 2: Correlations of low-frequency measures with the true spread (s_t) and volatility (σ_t) and regression of the low-frequency measures on the true spread and/or volatility. The average volatility equals 2.0% and volatility of volatility equals 2.0%. Both volatility and spread are *iid* lognormally distributed and mutually independent. All simulations are based on 10,000 replications.

		$T=21$									$T=251$								
		R_M	H_L	H_T	CS_M	CS_D	CS_P	AR_M	AR_D	AR_P	R_M	H_L	H_T	CS_M	CS_D	CS_P	AR_M	AR_D	AR_P
A. Small, tranquil spread: $\mathbb{E}(s)=0.1$, $\mathrm{Std}(s)=0.1$																			
Corr. with s_t		0.02	0.01	0.03	0.17	0.06	0.03	0.05	0.02	0.01	0.05	0.05	0.07	0.42	0.08	0.05	0.14	0.04	0.02
Corr. with σ_t		0.46	0.90	0.67	0.47	0.94	0.97	0.50	0.92	0.96	0.48	0.89	0.79	0.53	0.99	1.00	0.50	0.99	1.00
Reg. on s_t	α	0.73	1.31	0.92	0.19	0.79	1.36	0.39	0.77	1.56	0.39	0.64	0.55	0.08	0.77	1.33	0.19	0.76	1.54
	β_s	0.21	0.07	0.13	0.71	0.48	0.37	0.40	0.14	0.09	0.42	0.35	0.29	0.83	0.63	0.65	0.63	0.30	0.34
	R^2	0.00	0.00	0.00	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.18	0.01	0.00	0.02	0.00	0.00
Reg. on s_t, σ_t	α	0.07	0.15	0.56	-0.02	-0.02	-0.02	-0.03	-0.01	-0.04	0.01	-0.01	0.21	-0.02	-0.01	-0.01	-0.03	-0.02	-0.02
	β_s	0.32	0.17	0.16	0.73	0.57	0.52	0.45	0.22	0.26	0.43	0.43	0.25	0.83	0.57	0.55	0.61	0.24	0.23
	β_σ	0.33	0.59	0.18	0.11	0.40	0.68	0.21	0.39	0.80	0.19	0.32	0.17	0.05	0.40	0.68	0.11	0.40	0.79
	R^2	0.21	0.80	0.44	0.25	0.88	0.94	0.25	0.85	0.92	0.23	0.80	0.62	0.45	0.99	0.99	0.27	0.99	0.99
B. Medium, moderately volatile spread: $\mathbb{E}(s)=0.5$, $\mathrm{Std}(s)=0.5$																			
Corr. with s_t		0.23	0.21	0.38	0.71	0.41	0.23	0.47	0.36	0.16	0.47	0.48	0.65	0.95	0.46	0.26	0.76	0.40	0.19
Corr. with σ_t		0.42	0.86	0.54	0.24	0.84	0.94	0.37	0.83	0.94	0.38	0.73	0.50	0.11	0.88	0.96	0.21	0.89	0.98
Reg. on s_t	α	0.59	1.18	0.83	0.13	0.73	1.34	0.26	0.64	1.48	0.25	0.49	0.40	0.04	0.70	1.27	0.08	0.61	1.42
	β_s	0.65	0.51	0.40	0.91	0.76	0.67	0.84	0.64	0.54	0.80	0.73	0.70	0.96	0.78	0.72	0.94	0.64	0.58
	R^2	0.05	0.04	0.14	0.51	0.17	0.05	0.22	0.13	0.03	0.22	0.23	0.43	0.90	0.22	0.07	0.58	0.16	0.04
Reg. on s_t, σ_t	α	-0.02	0.06	0.52	-0.03	-0.05	-0.05	-0.08	-0.11	-0.10	-0.07	-0.08	0.13	-0.01	-0.04	-0.03	-0.05	-0.09	-0.08
	β_s	0.66	0.53	0.40	0.92	0.78	0.71	0.85	0.66	0.59	0.81	0.72	0.69	0.96	0.78	0.71	0.94	0.63	0.57
	β_σ	0.30	0.56	0.16	0.08	0.38	0.67	0.16	0.36	0.77	0.16	0.28	0.14	0.03	0.37	0.66	0.06	0.36	0.76
	R^2	0.23	0.78	0.44	0.57	0.89	0.94	0.36	0.83	0.91	0.36	0.76	0.67	0.91	0.98	0.99	0.62	0.94	0.99
C. Large, volatile spread: $\mathbb{E}(s)=3.0$, $\mathrm{Std}(s)=3.0$																			
Corr. with s_t		0.82	0.90	0.58	0.98	0.97	0.92	0.97	0.97	0.90	0.97	0.98	0.92	1.00	0.98	0.92	0.99	0.98	0.90
Corr. with σ_t		0.09	0.20	-0.16	0.02	0.18	0.36	0.03	0.10	0.36	0.02	0.07	-0.08	0.00	0.19	0.37	0.00	0.10	0.38
Reg. on s_t	α	0.12	0.62	1.27	0.00	0.38	0.88	0.01	0.04	0.80	0.01	0.08	0.29	-0.02	0.39	0.88	-0.01	0.04	0.79
	β_s	0.94	0.86	0.34	0.99	0.95	0.92	1.00	0.97	0.91	0.99	0.98	0.85	0.99	0.95	0.91	1.00	0.97	0.91
	R^2	0.68	0.82	0.33	0.97	0.94	0.84	0.94	0.95	0.82	0.94	0.97	0.84	1.00	0.95	0.84	0.99	0.96	0.82
Reg. on s_t, σ_t	α	-0.18	-0.03	1.56	-0.06	-0.17	-0.24	-0.09	-0.26	-0.33	-0.09	-0.10	0.53	-0.03	-0.17	-0.24	-0.03	-0.29	-0.35
	β_s	0.94	0.86	0.34	0.99	0.95	0.92	1.00	0.97	0.91	0.99	0.98	0.85	0.99	0.95	0.92	1.00	0.97	0.91
	β_σ	0.15	0.33	-0.15	0.03	0.28	0.56	0.05	0.15	0.57	0.05	0.09	-0.12	0.01	0.28	0.56	0.01	0.16	0.58
	R^2	0.69	0.86	0.35	0.97	0.98	0.97	0.94	0.96	0.95	0.94	0.97	0.85	1.00	0.99	0.98	0.99	0.97	0.97

Table 3: Correlations of low-frequency measures with the true spread (s_t) and volatility (σ_t) and regression of the low-frequency measures on the true spread and/or volatility. The average volatility equals 2.0% and volatility of volatility equals 2.0%. Both volatility and spread are *iid* lognormally distributed and contemporaneously correlated with $\rho = 0.5$. All simulations are based on 10,000 replications.

		$T=21$									$T=251$								
		R_M	H_L	H_T	CS_M	CS_D	CS_P	AR_M	AR_D	AR_P	R_M	H_L	H_T	CS_M	CS_D	CS_P	AR_M	AR_D	AR_P
A. Small, tranquil spread: $\mathbb{E}(s) = 0.1$, Std$(s) = 0.1$																			
Corr. with s_t		0.27	0.49	0.41	0.32	0.51	0.52	0.32	0.48	0.50	0.31	0.46	0.46	0.58	0.55	0.53	0.33	0.52	0.51
Corr. with σ_t		0.49	0.88	0.65	0.51	0.94	0.97	0.54	0.91	0.96	0.58	0.89	0.77	0.65	0.99	1.00	0.55	0.99	1.00
Reg. on s_t	α	0.38	0.69	0.70	0.11	0.40	0.67	0.18	0.38	0.77	0.16	0.34	0.37	0.04	0.38	0.66	0.09	0.38	0.76
	β_s	3.98	6.32	2.26	1.49	4.37	7.32	2.62	4.11	8.05	2.60	3.35	2.00	1.18	4.41	7.24	1.50	4.02	7.92
	R^2	0.08	0.24	0.17	0.10	0.26	0.27	0.10	0.23	0.25	0.10	0.21	0.21	0.34	0.30	0.29	0.11	0.27	0.26
Reg. on s_t, σ_t	α	0.01	0.18	0.56	0.00	0.00	-0.02	-0.03	-0.01	0.00	-0.05	0.03	0.22	-0.01	-0.01	-0.01	-0.03	-0.01	0.00
	β_s	0.53	0.45	0.56	0.40	0.49	0.66	0.52	0.33	0.50	0.14	0.07	0.39	0.69	0.45	0.47	0.32	0.11	0.11
	β_σ	0.36	0.55	0.16	0.11	0.39	0.68	0.21	0.38	0.77	0.23	0.32	0.16	0.05	0.40	0.68	0.12	0.39	0.79
	R^2	0.24	0.78	0.43	0.27	0.88	0.94	0.29	0.83	0.92	0.34	0.79	0.60	0.51	0.99	1.00	0.31	0.99	0.99
B. Medium, moderately volatile spread: $\mathbb{E}(s) = 0.5$, Std$(s) = 0.5$																			
Corr. with s_t		0.38	0.55	0.49	0.72	0.70	0.63	0.51	0.60	0.56	0.51	0.65	0.67	0.95	0.75	0.65	0.78	0.66	0.59
Corr. with σ_t		0.51	0.87	0.62	0.50	0.90	0.95	0.49	0.90	0.96	0.50	0.84	0.67	0.53	0.94	0.98	0.47	0.96	0.99
Reg. on s_t	α	0.28	0.66	0.70	0.06	0.37	0.66	0.14	0.34	0.74	0.14	0.28	0.34	0.02	0.38	0.69	0.02	0.35	0.77
	β_s	1.23	1.44	0.54	1.00	1.40	1.91	0.95	1.07	1.83	0.93	1.02	0.67	0.99	1.37	1.85	0.97	1.04	1.78
	R^2	0.14	0.30	0.24	0.51	0.49	0.39	0.26	0.37	0.31	0.26	0.42	0.45	0.91	0.56	0.42	0.61	0.44	0.34
Reg. on s_t, σ_t	α	0.00	0.19	0.58	-0.01	0.00	0.01	-0.01	-0.01	-0.01	-0.01	0.02	0.24	-0.01	-0.01	0.00	-0.02	-0.02	-0.01
	β_s	0.52	0.33	0.25	0.87	0.67	0.64	0.66	0.38	0.37	0.64	0.47	0.45	0.95	0.67	0.61	0.90	0.38	0.37
	β_σ	0.32	0.52	0.13	0.07	0.37	0.65	0.15	0.35	0.75	0.14	0.26	0.11	0.02	0.37	0.66	0.03	0.35	0.75
	R^2	0.28	0.77	0.42	0.54	0.90	0.94	0.34	0.85	0.93	0.35	0.77	0.59	0.91	0.99	1.00	0.62	0.97	0.99
C. Large, volatile spread: $\mathbb{E}(s) = 3.0$, Std$(s) = 3.0$																			
Corr. with s_t		0.84	0.91	0.47	0.98	0.98	0.94	0.97	0.98	0.94	0.97	0.98	0.83	1.00	0.99	0.96	1.00	0.99	0.96
Corr. with σ_t		0.46	0.53	0.05	0.49	0.59	0.71	0.48	0.48	0.69	0.49	0.47	0.22	0.48	0.58	0.70	0.48	0.47	0.68
Reg. on s_t	α	-0.05	0.43	1.24	0.00	0.18	0.45	-0.04	-0.07	0.36	0.00	0.00	0.58	-0.01	0.17	0.44	-0.01	-0.08	0.35
	β_s	0.95	0.82	0.21	0.98	0.98	1.00	1.00	0.94	0.98	0.99	0.97	0.68	0.98	0.99	1.00	1.00	0.94	0.98
	R^2	0.70	0.82	0.22	0.96	0.95	0.89	0.94	0.95	0.88	0.95	0.97	0.69	1.00	0.98	0.92	0.99	0.98	0.92
Reg. on s_t, σ_t	α	-0.10	0.35	1.40	-0.02	-0.03	-0.03	-0.05	-0.07	-0.08	-0.01	0.06	0.90	-0.01	-0.03	-0.05	-0.01	-0.05	-0.07
	β_s	0.94	0.79	0.27	0.98	0.91	0.84	1.00	0.94	0.83	0.99	0.99	0.80	0.98	0.92	0.85	1.00	0.95	0.84
	β_σ	0.05	0.09	-0.17	0.02	0.21	0.49	0.00	0.00	0.45	0.00	-0.06	-0.33	-0.01	0.20	0.48	-0.01	-0.03	0.42
	R^2	0.70	0.83	0.27	0.96	0.97	0.97	0.94	0.95	0.95	0.95	0.97	0.75	1.00	0.99	0.99	0.99	0.98	0.98

Table 4: Daily summary statistics for the S&P 1500 data. The table reports averages over stock-days in the sample. The quoted spread, effective spread, and realized volatility were winsorized at the 99.9% level prior to calculating the means. The daily realized volatility is based on 5-minute intraday mid-quote returns and close-open mid-quote returns. The sample period is from October 2003 to December 2017.

	DJIA30 (very large)	S&P500 (large)	S&P400 (mid)	S&P600 (small)	S&P1500 (all)
# Trades	51,070	16,691	5,555	2,132	8,606
Volume ($mn)	609.8	150.3	30.12	8.632	71.03
Quoted spread (bps)	2.920	5.617	11.01	29.17	16.35
Effective spread (bps)	4.292	6.069	9.065	19.98	12.36
Realized vol (bps)	142.3	183.5	206.2	254.6	217.1
Signal-to-noise	0.034	0.037	0.050	0.085	0.059

Table 5: Daily summary statistics for the foreign exchange rates. The table reports averages over stock-days in the sample. The daily realized volatility is based on 5-minute mid-quotes. The sample period is January 2008 to December 2015.

	EUR/USD	EUR/CHF	EUR/JPY	USD/JPY	USD/CHF
# Trades	39,390	3,635	5,529	21,033	6,241
Volume ($mn)	51,837	4,668	6,584	28,118	7,547
Quoted spread (bps)	1.071	2.579	2.819	1.452	2.722
Effective spread (bps)	0.505	0.721	0.687	0.680	0.826
Realized vol (bps)	66.31	45.24	84.58	66.68	73.55
Signal-to-noise	0.008	0.023	0.009	0.011	0.012

Table 6: Monthly descriptive statistics for S&P 1500 stocks. Panel A reports descriptive statistics calculated across all stock-months (pooled). Panel B reports cross-sectional averages, and the associated standard deviations in parentheses, of stock-specific time-series standard deviation and correlation with the TAQ effective spread (ES) and 5-minute realized volatility RV. Panel C reports the time-series averages, and the associated standard devations in parentheses, of cross-sectional standard deviation and correlation with the TAQ effective spread and realized volatility. The realized volatility is based on 5-minute midquotes. All variables were winsorized at the 99.9% level prior to calculating the descriptive statistics. The sample period is October 2003 to December 2017 and the sample size is 241,236 stock-months.

	ES	RV	R_M	H_L	H_T	CS_M	CS_D	CS_P	AR_M	AR_D	AR_P
A. Pooled											
Mean	12.4	217	90.0	168	107	26.4	76.2	123	58.3	82.2	157
10pct	3.68	105	0.00	56.7	54.1	0.00	33.6	57.6	0.00	33.1	70.2
25pct	5.39	134	0.00	79.0	73.0	0.00	45.0	74.6	0.00	46.1	93.1
Med	8.69	182	35.5	117	101	15.4	63.1	102	34.4	66.6	129
75pct	14.9	257	138	181	135	38.8	90.9	144	89.4	98.4	186
90pct	24.6	364	245	289	166	67.9	131	207	152	146	269
RMSE			153	261	104	35.6	78.2	132	89.8	88.7	176
B. Cross-sectional averages of time-series moments											
Std	5.69	105	121	198	40.1	30.7	38.1	58.2	71.5	46.2	80.3
	(6.18)	(55.1)	(51.1)	(139)	(9.49)	(14.3)	(21.0)	(33.2)	(31.5)	(24.8)	(44.5)
$\rho(ES,\cdot)$		0.43	0.19	0.19	0.28	0.20	0.45	0.47	0.19	0.41	0.44
		(0.28)	(0.22)	(0.27)	(0.23)	(0.23)	(0.27)	(0.27)	(0.23)	(0.26)	(0.28)
$\rho(RV,\cdot)$	0.43		0.32	0.60	0.51	0.16	0.62	0.70	0.29	0.65	0.74
	(0.28)		(0.26)	(0.20)	(0.19)	(0.26)	(0.26)	(0.25)	(0.26)	(0.23)	(0.21)
C. Time-series averages of cross-sectional moments											
Std	11.7	99.3	111	199	39.8	30.8	36.3	53.7	67.5	41.8	72.2
	(3.28)	(31.4)	(46.3)	(46.8)	(4.60)	(11.8)	(14.9)	(22.9)	(28.1)	(19.1)	(32.4)
$\rho(ES,\cdot)$		0.48	0.20	0.16	0.30	0.37	0.54	0.50	0.29	0.47	0.46
		(0.10)	(0.08)	(0.08)	(0.10)	(0.09)	(0.07)	(0.07)	(0.10)	(0.10)	(0.09)
$\rho(RV,\cdot)$	0.48		0.30	0.55	0.52	0.24	0.67	0.74	0.33	0.68	0.76
	(0.10)		(0.08)	(0.05)	(0.13)	(0.08)	(0.07)	(0.07)	(0.08)	(0.06)	(0.06)

Table 7: Regressions of low-frequency measures on TAQ effective spread and 5-minute realized volatility for S&P 1500 stocks at the monthly frequency. Panel A reports pooled OLS regression estimates. Panel B reports within estimates, that is, pooled OLS regressions on data de-meaned at the stock level. In both panels, the Driscoll-Kraay robust t-statistics are reported in parentheses. Panel C reports the pooled mean-time estimates, that is, the time-series averages of month-by-month cross-sectional regression estimates. The t-statistics, reported in parentheses, are calculated using the Newey-West estimator of the variance of the cross-sectional estimates. The sample period is October 2003 to December 2017 and the sample size is 241,236 stock-months.

	R_M		H_L		H_T		CS_M		CS_D		CS_P		AR_M		AR_D		AR_P	
Panel A. Pooled estimation																		
const	58.5	-5.78	129	-14.6	92.9	63.3	12.9	4.77	50.3	8.56	85.1	12.0	31.6	-0.60	54.4	3.75	109	7.38
	(16.1)	(-0.88)	(33.8)	(-0.97)	(28.7)	(13.3)	(16.2)	(3.69)	(25.6)	(2.96)	(25.2)	(2.35)	(17.9)	(-0.14)	(25.3)	(0.97)	(25.7)	(1.04)
β_{ES}	2.54	0.36	3.10	-1.79	1.14	0.14	1.09	0.81	2.08	0.66	3.00	0.52	2.15	1.05	2.24	0.52	3.90	0.46
	(6.76)	(3.65)	(8.78)	(-5.62)	(30.0)	(2.05)	(14.7)	(17.9)	(8.54)	(19.6)	(7.29)	(8.69)	(7.82)	(7.58)	(7.06)	(11.4)	(6.50)	(7.22)
β_{RV}		0.42		0.94		0.19		0.05		0.27		0.48		0.21		0.33		0.66
		(9.73)		(9.41)		(9.83)		(5.38)		(13.6)		(14.0)		(7.80)		(14.2)		(15.7)
R^2	0.06	0.19	0.03	0.30	0.10	0.34	0.15	0.18	0.26	0.66	0.23	0.73	0.11	0.20	0.22	0.64	0.21	0.75
B. Within estimation																		
β_{ES}	3.34	0.44	3.30	-3.30	1.07	-0.12	1.09	0.73	2.55	0.83	3.93	0.94	2.50	1.05	2.90	0.77	5.22	1.02
	(4.61)	(3.53)	(4.06)	(-8.57)	(13.8)	(-1.22)	(7.91)	(14.5)	(5.36)	(15.9)	(4.86)	(12.1)	(5.24)	(5.52)	(4.91)	(8.39)	(4.58)	(7.08)
β_{RV}		0.44		1.01		0.18		0.06		0.26		0.46		0.22		0.33		0.64
		(9.43)		(8.77)		(11.4)		(4.78)		(11.5)		(11.9)		(7.28)		(13.4)		(14.5)
R^2	0.04	0.17	0.02	0.28	0.04	0.24	0.07	0.10	0.20	0.58	0.20	0.68	0.07	0.16	0.18	0.58	0.19	0.70
C. Pooled mean-time estimation																		
β_{ES}	1.87	0.62	2.52	-2.85	1.10	0.25	0.99	0.89	1.67	0.84	2.30	0.82	1.70	1.05	1.68	0.65	2.86	0.73
	(17.2)	(6.63)	(24.9)	(-12.1)	(19.1)	(7.34)	(20.4)	(18.3)	(22.7)	(20.6)	(22.3)	(16.6)	(15.7)	(10.8)	(18.4)	(12.2)	(19.0)	(9.88)
β_{RV}		0.30		1.33		0.22		0.03		0.20		0.35		0.16		0.25		0.50
		(27.9)		(23.6)		(20.8)		(8.75)		(30.4)		(30.3)		(22.8)		(33.5)		(35.7)
R^2	0.04	0.11	0.03	0.33	0.10	0.29	0.14	0.15	0.29	0.52	0.26	0.58	0.09	0.14	0.23	0.50	0.22	0.60

Table 8: Annual descriptive statistics for S&P 1500 stocks. Panel A reports descriptive statistics calculated across all stock-years (pooled). Panel B reports cross-sectional averages, and the associated standard deviations in parentheses, of stock-specific time-series standard deviation and correlation with the TAQ effective spread (ES) and 5-minute realized volatility RV. Panel C reports the time-series averages, and the associated standard devations in parentheses, of cross-sectional standard deviation and correlation with the TAQ effective spread and realized volatility. The realized volatility is based on 5-minute midquotes. All variables were winsorized at the 99.9% level prior to calculating the descriptive statistics. The sample period is October 2003 to December 2017 and the sample size is 18,239 stock-years.

	ES	RV	R_M	H_L	H_T	CS_M	CS_D	CS_P	AR_M	AR_D	AR_P
A. Pooled											
Mean	12.1	231	71.8	64.2	60.0	18.3	75.9	122	46.0	82.2	157
10pct	3.83	122	0.00	29.2	28.8	0.00	40.6	66.2	0.00	43.9	84.7
25pct	5.51	154	0.00	37.6	37.0	4.45	49.9	80.9	0.00	54.2	104
Med	8.72	203	47.3	51.8	50.3	13.3	66.1	106	34.2	70.8	136
75pct	14.9	274	110	75.7	71.8	25.3	89.5	143	68.8	97.1	185
90pct	23.9	372	187	114	104	41.5	125	198	112	135	256
RMSE			107	65.2	57.8	16.7	71.8	123	62.1	79.5	163
B. Cross-sectional averages of time-series moments											
Std	3.33	64.6	63.4	25.9	21.8	11.1	21.4	34.1	36.4	24.3	45.0
	(4.41)	(52.2)	(43.6)	(21.0)	(16.5)	(8.89)	(17.2)	(27.2)	(27.1)	(19.6)	(36.2)
$\rho(ES,\cdot)$		0.39	0.16	0.34	0.33	0.27	0.46	0.44	0.20	0.42	0.42
		(0.49)	(0.48)	(0.48)	(0.48)	(0.50)	(0.47)	(0.48)	(0.48)	(0.48)	(0.48)
$\rho(RV,\cdot)$	0.39		0.30	0.56	0.54	0.21	0.74	0.77	0.23	0.75	0.77
	(0.49)		(0.52)	(0.49)	(0.49)	(0.54)	(0.43)	(0.41)	(0.53)	(0.42)	(0.41)
C. Time-series averages of cross-sectional moments											
Std	10.5	84.6	73.1	29.5	25.5	17.4	27.7	42.7	47.5	29.9	55.4
	(1.96)	(22.8)	(31.4)	(14.4)	(10.4)	(6.24)	(8.72)	(13.9)	(20.9)	(11.2)	(20.0)
$\rho(ES,\cdot)$		0.54	0.28	0.48	0.47	0.64	0.66	0.58	0.47	0.62	0.56
		(0.08)	(0.08)	(0.08)	(0.09)	(0.06)	(0.07)	(0.07)	(0.07)	(0.10)	(0.10)
$\rho(RV,\cdot)$	0.54		0.24	0.61	0.57	0.35	0.86	0.89	0.31	0.87	0.90
	(0.08)		(0.08)	(0.09)	(0.11)	(0.08)	(0.03)	(0.04)	(0.09)	(0.03)	(0.03)

Table 9: Regressions of low-frequency measures on TAQ effective spread and 5-minute realized volatility for S&P 1500 stocks at the annual frequency. Panel A reports pooled OLS regression estimates. Panel B reports within estimates, that is, pooled OLS regressions on data de-meaned at the stock level. In both panels, the Driscoll-Kraay robust t-statistics are reported in parentheses. Panel C reports the pooled mean-time estimates, that is, the time-series averages of year-by-year cross-sectional regression estimates. The t-statistics, reported in parentheses, are calculated using the Newey-West estimator of the variance of the cross-sectional estimates. The sample period is October 2003 to December 2017 and the sample size is 18,239 stock-years.

	R_M		H_L		H_T		CS_M		CS_D		CS_P		AR_M		AR_D		AR_P	
Panel A. Pooled estimation																		
const	42.8	-12.7	44.5	0.62	44.1	10.2	4.74	-0.11	51.3	3.63	87.4	6.25	16.1	-7.24	56.4	2.29	113	5.73
	(4.66)	(-2.37)	(7.93)	(0.22)	(8.07)	(6.01)	(5.35)	(-0.14)	(12.0)	(1.76)	(11.8)	(1.85)	(7.94)	(-1.66)	(12.4)	(0.82)	(12.3)	(1.16)
β_{ES}	2.39	0.69	1.63	0.28	1.31	0.28	1.12	0.97	2.04	0.58	2.84	0.35	2.47	1.76	2.13	0.47	3.63	0.34
	(6.87)	(3.60)	(6.80)	(2.97)	(10.7)	(2.55)	(10.9)	(14.0)	(7.02)	(15.6)	(6.50)	(8.71)	(6.18)	(6.88)	(5.74)	(4.51)	(5.65)	(2.98)
β_{RV}		0.33		0.26		0.20		0.03		0.28		0.48		0.14		0.32		0.64
		(6.33)		(10.00)		(12.8)		(4.29)		(20.2)		(20.5)		(5.77)		(20.7)		(22.1)
R^2	0.08	0.20	0.17	0.53	0.16	0.48	0.38	0.40	0.34	0.87	0.26	0.88	0.23	0.28	0.30	0.85	0.25	0.88
B. Within estimation																		
β_{ES}	3.48	0.28	2.42	0.33	1.80	0.17	1.15	0.83	2.71	0.67	4.03	0.61	3.05	1.71	2.97	0.63	5.25	0.70
	(3.64)	(1.91)	(4.08)	(4.99)	(4.77)	(1.56)	(7.59)	(10.6)	(4.32)	(9.70)	(3.99)	(7.09)	(4.93)	(5.47)	(4.01)	(4.46)	(3.78)	(3.37)
β_{RV}		0.45		0.29		0.23		0.04		0.28		0.48		0.19		0.33		0.63
		(10.1)		(11.9)		(12.1)		(5.06)		(18.5)		(19.8)		(6.82)		(23.3)		(24.3)
R^2	0.05	0.23	0.13	0.52	0.11	0.45	0.18	0.24	0.25	0.83	0.22	0.86	0.12	0.22	0.23	0.82	0.21	0.86
C. Pooled mean-time estimation																		
β_{ES}	1.89	1.41	1.29	0.58	1.10	0.53	1.04	1.05	1.73	0.70	2.36	0.55	2.13	1.97	1.75	0.61	2.95	0.56
	(11.7)	(6.05)	(12.0)	(5.13)	(12.8)	(5.24)	(13.5)	(11.6)	(13.8)	(10.1)	(13.9)	(7.31)	(8.28)	(7.49)	(9.68)	(5.21)	(10.5)	(3.84)
β_{RV}		0.11		0.16		0.13		0.00		0.23		0.41		0.04		0.26		0.54
		(3.84)		(11.6)		(16.1)		(0.06)		(31.9)		(29.4)		(2.06)		(23.3)		(25.7)
R^2	0.08	0.10	0.24	0.42	0.23	0.38	0.41	0.41	0.44	0.80	0.35	0.81	0.22	0.24	0.39	0.79	0.33	0.82

Table 10: Monthly descriptive statistics for DJIA 30 stocks. Panel A reports descriptive statistics calculated across all stock-months (pooled). Panel B reports cross-sectional averages, and the associated standard deviations in parentheses, of stock-specific time-series standard deviation and correlation with the TAQ effective spread (ES) and 5-minute realized volatility RV. Panel C reports the time-series averages, and the associated standard devations in parentheses, of cross-sectional standard deviation and correlation with the TAQ effective spread and realized volatility. The realized volatility is based on 5-minute intraday mid-quotes. The sample period is October 2013 to December 2017 and the sample size is 4,791 stock-months.

	ES	RV	R_M	H_L	H_T	CS_M	CS_D	CS_P	AR_M	AR_D	AR_P
A. Pooled											
Mean	4.30	142	56.9	104	81.0	15.8	49.3	80.8	37.4	54.8	105
10pct	1.91	80.6	0.00	41.3	40.1	0.00	25.4	43.7	0.00	25.1	52.7
25pct	2.40	94.5	0.00	55.0	52.6	0.00	32.0	53.1	0.00	33.2	67.2
Med	3.67	118	11.4	76.1	71.0	10.5	41.5	67.5	25.3	45.1	87.3
75pct	5.48	155	88.5	110	96.6	24.2	55.2	89.1	59.1	62.9	116
90pct	7.37	221	153	174	136	39.2	76.8	124	91.3	89.5	166
RMSE			104	148	87.7	22.4	54.6	92.7	59.1	63.7	124
B. Cross-sectional averages of time-series moments											
Std	1.88	73.4	81.2	91.2	38.2	18.9	26.7	42.3	46.3	32.7	57.8
	(0.97)	(52.8)	(32.1)	(49.5)	(9.89)	(4.93)	(14.3)	(27.8)	(13.8)	(18.7)	(37.4)
$\rho(ES,\cdot)$		0.40	0.14	0.26	0.26	0.14	0.36	0.38	0.14	0.35	0.37
		(0.20)	(0.21)	(0.22)	(0.19)	(0.24)	(0.26)	(0.25)	(0.24)	(0.20)	(0.24)
$\rho(RV,\cdot)$	0.40		0.38	0.65	0.59	0.16	0.70	0.80	0.26	0.73	0.83
	(0.20)		(0.34)	(0.20)	(0.18)	(0.27)	(0.23)	(0.17)	(0.33)	(0.21)	(0.15)
C. Time-series averages of cross-sectional moments											
Std	1.57	45.1	60.6	72.7	29.7	15.2	16.2	25.1	36.8	20.9	36.8
	(0.78)	(39.3)	(37.5)	(63.0)	(14.7)	(7.78)	(11.9)	(21.4)	(20.3)	(15.8)	(29.9)
$\rho(ES,\cdot)$		0.40	0.07	0.23	0.20	0.07	0.32	0.36	0.08	0.28	0.33
		(0.20)	(0.23)	(0.23)	(0.20)	(0.22)	(0.23)	(0.21)	(0.23)	(0.22)	(0.21)
$\rho(RV,\cdot)$	0.40		0.16	0.64	0.53	-0.01	0.51	0.65	0.16	0.59	0.71
	(0.20)		(0.28)	(0.20)	(0.22)	(0.26)	(0.22)	(0.19)	(0.28)	(0.20)	(0.16)

Table 11: Regressions of low-frequency measures on TAQ effective spread and 5-minute realized volatility for DJIA 30 stocks. Panel A reports pooled OLS regression estimates. Panel B reports within estimates, that is, pooled OLS regressions on data de-meaned at the stock level. In both panels, the Driscoll-Kraay robust t-statistics are reported in parentheses. Panel C reports the pooled mean-time estimates, that is, the time-series averages of month-by-month cross-sectional regression estimates. The t-statistics, reported in parentheses, are calculated using the Newey-West estimator of the variance of the cross-sectional estimates. The sample period is October 2003 to December 2017 and the sample size is 4,791 stock-months.

	R_M		H_L		H_T		CS_M		CS_D		CS_P		AR_M		AR_D		AR_P	
Panel A. Pooled estimation																		
const	25.4	-9.10	54.7	-0.69	62.0	42.5	8.21	5.71	24.5	4.84	39.5	3.32	20.7	8.55	26.7	1.33	51.1	0.68
	(3.17)	(-1.52)	(4.92)	(-0.10)	(16.3)	(7.55)	(6.22)	(4.90)	(4.72)	(3.55)	(4.03)	(1.53)	(6.14)	(2.83)	(4.22)	(1.05)	(3.88)	(0.32)
β_{ES}	7.33	-0.61	11.4	-1.35	4.44	-0.03	1.76	1.19	5.77	1.24	9.60	1.27	3.87	1.07	6.54	0.70	12.5	0.90
	(2.36)	(-0.56)	(2.90)	(-1.63)	(3.42)	(-0.07)	(4.82)	(5.22)	(3.06)	(4.38)	(2.78)	(3.41)	(3.39)	(2.11)	(2.88)	(2.47)	(2.73)	(2.13)
β_{RV}		0.48		0.78		0.27		0.04		0.28		0.51		0.17		0.35		0.71
		(10.2)		(13.2)		(8.01)		(5.98)		(20.4)		(22.9)		(9.24)		(28.6)		(35.1)
R^2	0.04	0.24	0.07	0.41	0.07	0.34	0.05	0.07	0.20	0.72	0.20	0.83	0.04	0.12	0.17	0.72	0.19	0.84
B. Within estimation																		
β_{ES}	8.32	0.06	11.6	-1.57	4.71	0.24	1.94	1.27	6.08	1.46	9.99	1.55	4.40	1.40	6.84	0.91	12.7	1.03
	(2.17)	(0.05)	(2.54)	(-1.78)	(2.93)	(0.46)	(3.94)	(4.58)	(2.77)	(4.36)	(2.54)	(3.48)	(3.04)	(2.63)	(2.64)	(3.13)	(2.49)	(2.35)
β_{RV}		0.49		0.79		0.27		0.04		0.28		0.50		0.18		0.35		0.70
		(9.84)		(13.8)		(9.16)		(6.42)		(18.9)		(21.2)		(9.30)		(26.8)		(31.6)
R^2	0.04	0.23	0.05	0.39	0.06	0.31	0.04	0.07	0.18	0.69	0.18	0.81	0.04	0.12	0.15	0.69	0.16	0.81
C. Pooled mean-time estimation																		
β_{ES}	2.93	-0.94	10.6	-3.32	3.59	-0.73	0.73	0.88	3.45	1.14	6.10	1.58	2.46	0.76	3.98	0.42	8.38	0.95
	(3.62)	(-0.96)	(7.14)	(-3.46)	(7.55)	(-1.59)	(4.22)	(4.86)	(9.97)	(6.13)	(8.48)	(6.32)	(5.27)	(1.54)	(8.27)	(1.62)	(8.33)	(2.27)
β_{RV}		0.29		1.23		0.45		-0.01		0.19		0.37		0.15		0.29		0.61
		(8.49)		(13.3)		(15.8)		(-1.69)		(18.0)		(20.7)		(8.19)		(25.8)		(29.1)
R^2	0.06	0.15	0.11	0.48	0.08	0.35	0.05	0.11	0.15	0.36	0.17	0.49	0.06	0.14	0.13	0.42	0.15	0.56

Table 12: Annual descriptive statistics for DJIA 30 stocks. Panel A reports descriptive statistics calculated across all stock-years (pooled). Panel B reports cross-sectional averages, and the associated standard deviations in parentheses, of stock-specific time-series standard deviation and correlation with the TAQ effective spread (ES) and 5-minute realized volatility RV. Panel C reports the time-series averages, and the associated standard devations in parentheses, of cross-sectional standard deviation and correlation with the TAQ effective spread and realized volatility. The realized volatility is based on 5-minute midquotes. All variables were winsorized at the 99.9% level prior to calculating the descriptive statistics. The sample period is October 2003 to December 2017 and the sample size is 388 stock-years.

	ES	RV	R_M	H_L	H_T	CS_M	CS_D	CS_P	AR_M	AR_D	AR_P
A. Pooled											
Mean	4.32	157	45.0	43.2	41.5	11.0	49.8	81.5	32.0	55.7	107
10pct	1.97	94.3	0.00	20.6	20.5	0.00	31.9	51.2	0.00	34.5	65.7
25pct	2.43	110	0.00	26.3	25.6	3.91	36.1	59.0	0.00	40.0	76.5
Med	3.90	130	23.2	34.4	33.9	9.65	43.0	69.5	30.2	48.0	91.5
75pct	5.46	170	65.4	48.0	46.6	16.2	53.8	88.7	45.4	59.6	115
90pct	7.00	239	122	80.8	78.0	22.8	73.0	125	66.8	84.8	159
RMSE			74.5	48.0	44.5	10.9	50.4	87.1	42.7	58.1	117
B. Cross-sectional averages of time-series moments											
Std	1.60	70.6	47.2	21.9	19.3	6.68	17.1	30.1	25.7	20.5	40.7
	(1.59)	(80.4)	(27.9)	(14.1)	(11.0)	(3.19)	(13.1)	(26.8)	(16.7)	(17.9)	(40.4)
$\rho(ES, \cdot)$		0.40	0.11	0.31	0.31	0.33	0.47	0.43	0.22	0.42	0.40
		(0.33)	(0.31)	(0.37)	(0.38)	(0.47)	(0.33)	(0.35)	(0.41)	(0.35)	(0.36)
$\rho(RV, \cdot)$	0.40		0.47	0.74	0.73	0.18	0.84	0.85	0.22	0.85	0.85
	(0.33)		(0.43)	(0.28)	(0.28)	(0.51)	(0.31)	(0.32)	(0.51)	(0.30)	(0.32)
C. Time-series averages of cross-sectional moments											
Std	1.66	49.6	42.0	16.2	14.6	6.34	10.9	20.1	25.6	14.5	29.7
	(1.07)	(50.7)	(21.9)	(9.82)	(7.28)	(2.86)	(8.11)	(17.3)	(14.4)	(11.6)	(26.6)
$\rho(ES, \cdot)$		0.49	0.06	0.25	0.23	0.09	0.55	0.53	0.19	0.49	0.50
		(0.18)	(0.22)	(0.22)	(0.20)	(0.19)	(0.12)	(0.15)	(0.22)	(0.16)	(0.17)
$\rho(RV, \cdot)$	0.49		0.11	0.45	0.43	-0.15	0.79	0.84	0.07	0.80	0.85
	(0.18)		(0.31)	(0.18)	(0.18)	(0.22)	(0.09)	(0.09)	(0.22)	(0.11)	(0.09)

Table 13: Regressions of low-frequency measures on TAQ effective spread and 5-minute realized volatility for DJIA 30 stocks at the annual frequency. Panel A reports pooled OLS regression estimates. Panel B reports within estimates, that is, pooled OLS regressions on data de-meaned at the stock level. In both panels, the Driscoll-Kraay robust t-statistics are reported in parentheses. Panel C reports the pooled mean-time estimates, that is, the time-series averages of year-by-year cross-sectional regression estimates. The t-statistics, reported in parentheses, are calculated using the Newey-West estimator of the variance of the cross-sectional estimates. The sample period is October 2003 to December 2017 and the sample size is 388 stock-years.

	R_M		H_L		H_T		CS_M		CS_D		CS_P		AR_M		AR_D		AR_P	
Panel A. Pooled estimation																		
const	29.5	10.5	22.9	8.91	25.1	13.6	6.69	6.92	26.3	14.4	39.3	16.7	14.3	9.21	28.6	13.2	50.7	18.2
	(2.33)	(0.62)	(5.53)	(1.80)	(6.51)	(2.56)	(1.84)	(2.12)	(8.17)	(4.00)	(5.59)	(4.50)	(2.21)	(1.41)	(6.56)	(4.25)	(4.75)	(5.35)
β_{ES}	3.58	-3.18	4.68	-0.30	3.79	-0.29	1.00	1.09	5.42	1.17	9.76	1.74	4.09	2.28	6.27	0.77	13.0	1.48
	(1.70)	(-1.31)	(2.43)	(-0.44)	(2.51)	(-0.45)	(2.05)	(1.81)	(3.26)	(2.77)	(2.73)	(2.88)	(3.56)	(1.74)	(2.95)	(1.69)	(2.68)	(2.14)
β_{RV}		0.31		0.23		0.19		-0.00		0.19		0.36		0.08		0.25		0.52
		(22.8)		(15.7)		(14.8)		(-0.47)		(23.3)		(16.8)		(7.07)		(18.9)		(15.9)
R^2	0.02	0.18	0.17	0.56	0.15	0.50	0.08	0.08	0.36	0.81	0.36	0.86	0.10	0.14	0.32	0.83	0.34	0.87
B. Within estimation																		
β_{ES}	5.05	-2.83	5.33	0.07	4.37	0.02	1.22	1.11	5.74	1.33	10.3	2.13	4.80	2.42	6.59	0.94	13.3	1.70
	(1.29)	(-1.26)	(2.27)	(0.09)	(2.23)	(0.03)	(2.78)	(1.90)	(3.09)	(2.57)	(2.82)	(2.69)	(3.89)	(2.23)	(2.89)	(2.08)	(2.85)	(2.88)
β_{RV}		0.36		0.24		0.20		0.01		0.20		0.37		0.11		0.26		0.53
		(12.9)		(23.4)		(25.0)		(0.46)		(26.9)		(28.9)		(6.81)		(27.8)		(24.6)
R^2	0.03	0.23	0.16	0.58	0.14	0.52	0.09	0.09	0.31	0.81	0.32	0.86	0.10	0.17	0.28	0.83	0.29	0.88
C. Pooled mean-time estimation																		
β_{ES}	1.20	-0.43	2.45	0.25	2.12	0.25	0.51	0.97	3.64	1.35	6.25	1.91	3.26	3.70	4.19	1.01	8.70	1.96
	(0.65)	(-0.25)	(2.97)	(0.29)	(2.92)	(0.30)	(1.91)	(3.78)	(11.3)	(7.36)	(6.81)	(4.67)	(3.42)	(3.21)	(6.54)	(1.93)	(5.57)	(1.90)
β_{RV}		0.16		0.17		0.16		-0.04		0.18		0.34		-0.03		0.25		0.52
		(1.88)		(9.75)		(8.87)		(-3.96)		(14.8)		(13.4)		(-0.93)		(12.4)		(13.1)
R^2	0.05	0.14	0.11	0.28	0.09	0.26	0.04	0.13	0.32	0.69	0.30	0.75	0.09	0.13	0.26	0.69	0.28	0.77

Table 14: Monthly descriptive statistics for S&P 600 stocks. Panel A reports descriptive statistics calculated across all stock-months (pooled). Panel B reports cross-sectional averages, and the associated standard deviations in parentheses, of stock-specific time-series standard deviation and correlation with the TAQ effective spread (ES) and 5-minute realized volatility RV. Panel C reports the time-series averages, and the associated standard devations in parentheses, of cross-sectional standard deviation and correlation with the TAQ effective spread and realized volatility. The realized volatility is based on 5-minute intraday mid-quotes. The sample period is October 2003 to December 2017 and the sample size is 95,923 stock-months.

	ES	RV	R_M	H_L	H_T	CS_M	CS_D	CS_P	AR_M	AR_D	AR_P
A. Pooled											
Mean	20.1	255	108	197	119	35.5	91.3	143	73.7	96.9	182
10pct	8.22	135	0.00	69.0	64.4	0.00	42.9	72.0	0.00	40.6	86.3
25pct	10.9	167	0.00	95.1	85.1	0.00	56.8	92.2	0.00	56.3	113
Med	15.6	218	53.8	139	114	24.0	77.7	122	50.5	80.3	153
75pct	23.4	298	169	212	146	52.7	109	168	114	116	215
90pct	35.7	414	289	330	175	86.8	153	235	185	169	305
RMSE			173	298	108	42.8	86.8	146	105	97.5	195
B. Cross-sectional averages of time-series moments											
Std	8.05	110	133	219	40.2	35.7	40.5	59.9	80.6	49.2	83.5
	(8.12)	(59.5)	(57.8)	(161)	(10.6)	(16.9)	(23.3)	(35.8)	(36.8)	(27.7)	(48.7)
$\rho(ES,\cdot)$		0.39	0.17	0.16	0.25	0.20	0.42	0.44	0.19	0.38	0.40
		(0.30)	(0.24)	(0.28)	(0.25)	(0.26)	(0.30)	(0.30)	(0.26)	(0.29)	(0.30)
$\rho(RV,\cdot)$	0.39		0.28	0.57	0.45	0.15	0.56	0.65	0.28	0.60	0.69
	(0.30)		(0.28)	(0.22)	(0.21)	(0.28)	(0.28)	(0.26)	(0.28)	(0.25)	(0.23)
C. Time-series averages of cross-sectional moments											
Std	14.9	108	128	226	40.4	37.2	40.3	58.1	79.0	46.9	79.0
	(5.12)	(33.0)	(48.7)	(61.9)	(4.52)	(13.9)	(16.3)	(24.2)	(31.8)	(21.0)	(34.8)
$\rho(ES,\cdot)$		0.42	0.16	0.11	0.23	0.33	0.46	0.43	0.27	0.42	0.40
		(0.14)	(0.09)	(0.10)	(0.10)	(0.10)	(0.11)	(0.11)	(0.11)	(0.12)	(0.12)
$\rho(RV,\cdot)$	0.42		0.27	0.52	0.44	0.20	0.61	0.69	0.30	0.64	0.72
	(0.14)		(0.10)	(0.07)	(0.15)	(0.10)	(0.09)	(0.09)	(0.10)	(0.08)	(0.07)

Table 15: Regressions of low-frequency measures on TAQ effective spread and 5-minute realized volatility for S&P 600 stocks at the monthly frequency. Panel A reports pooled OLS regression estimates. Panel B reports within estimates, that is, pooled OLS regressions on data de-meaned at the stock level. In both panels, the Driscoll-Kraay robust t-statistics are reported in parentheses. Panel C reports the pooled mean-time estimates, that is, the time-series averages of month-by-month cross-sectional regression estimates. The t-statistics, reported in parentheses, are calculated using the Newey-West estimator of the variance of the cross-sectional estimates. The sample period is October 2003 to December 2017 and the sample size is 95,923 stock-months.

	R_M		H_L		H_T		CS_M		CS_D		CS_P		AR_M		AR_D		AR_P	
Panel A. Pooled estimation																		
const	64.0	-3.88	154	-11.8	104	76.6	16.0	5.72	56.8	11.2	94.4	16.4	33.2	-5.76	57.8	2.75	116	7.31
	(16.5)	(-0.43)	(41.7)	(-0.69)	(34.3)	(16.0)	(16.8)	(3.69)	(26.2)	(2.98)	(25.9)	(2.48)	(10.4)	(-0.91)	(19.9)	(0.50)	(21.4)	(0.72)
β_{ES}	2.19	0.53	2.13	-1.92	0.73	0.06	0.97	0.72	1.72	0.60	2.42	0.52	2.02	1.06	1.95	0.60	3.28	0.62
	(5.75)	(4.92)	(5.62)	(-6.88)	(23.7)	(1.03)	(12.9)	(20.9)	(7.48)	(20.7)	(6.33)	(11.3)	(6.85)	(7.73)	(6.23)	(8.97)	(5.62)	(6.60)
β_{RV}		0.40		0.97		0.16		0.06		0.27		0.46		0.23		0.32		0.64
		(8.84)		(9.80)		(10.8)		(6.50)		(13.2)		(13.1)		(7.92)		(13.3)		(14.1)
R^2	0.06	0.16	0.02	0.27	0.07	0.25	0.14	0.17	0.26	0.61	0.22	0.68	0.12	0.21	0.23	0.60	0.22	0.70
B. Within estimation																		
β_{ES}	2.73	0.68	2.41	-2.76	0.71	-0.03	1.00	0.70	2.04	0.80	3.04	0.92	2.27	1.10	2.36	0.82	4.12	1.11
	(5.10)	(4.36)	(4.29)	(-8.24)	(15.5)	(-0.34)	(9.70)	(14.6)	(6.37)	(15.1)	(5.64)	(12.2)	(5.81)	(6.36)	(5.55)	(8.48)	(5.09)	(7.71)
β_{RV}		0.42		1.06		0.15		0.06		0.25		0.43		0.24		0.32		0.62
		(8.22)		(8.88)		(11.3)		(6.08)		(12.0)		(11.9)		(7.55)		(13.3)		(13.8)
R^2	0.05	0.14	0.01	0.26	0.04	0.19	0.08	0.11	0.23	0.55	0.23	0.64	0.09	0.17	0.21	0.55	0.22	0.66
C. Pooled mean-time estimation																		
β_{ES}	1.44	0.55	1.41	-2.48	0.70	0.21	0.85	0.77	1.28	0.68	1.70	0.63	1.47	0.98	1.35	0.60	2.18	0.64
	(13.7)	(6.45)	(13.1)	(-11.0)	(16.4)	(7.69)	(17.2)	(16.5)	(17.8)	(14.9)	(17.0)	(11.1)	(13.9)	(10.6)	(15.2)	(10.2)	(15.0)	(7.39)
β_{RV}		0.28		1.32		0.17		0.03		0.19		0.34		0.16		0.24		0.49
		(23.3)		(20.8)		(17.2)		(8.20)		(25.3)		(26.3)		(18.9)		(29.6)		(31.1)
R^2	0.04	0.09	0.02	0.30	0.06	0.22	0.12	0.13	0.23	0.44	0.19	0.51	0.09	0.13	0.19	0.45	0.18	0.54

Table 16: Annual descriptive statistics for S&P 600 stocks. Panel A reports descriptive statistics calculated across all stock-years (pooled). Panel B reports cross-sectional averages, and the associated standard deviations in parentheses, of stock-specific time-series standard deviation and correlation with the TAQ effective spread (ES) and 5-minute realized volatility RV. Panel C reports the time-series averages, and the associated standard devations in parentheses, of cross-sectional standard deviation and correlation with the TAQ effective spread and realized volatility. The realized volatility is based on 5-minute midquotes. All variables were winsorized at the 99.9% level prior to calculating the descriptive statistics. The sample period is October 2003 to December 2017 and the sample size is 7,122 stock-years.

	ES	RV	R_M	H_L	H_T	CS_M	CS_D	CS_P	AR_M	AR_D	AR_P
A. Pooled											
Mean	19.7	267	88.8	75.3	69.7	26.9	90.9	142	63.5	96.7	182
10pct	8.81	159	0.00	36.1	35.5	0.81	53.5	85.1	0.00	55.5	106
25pct	11.4	192	0.00	46.2	45.0	10.6	64.1	102	0.00	66.8	128
Med	16.0	242	68.5	61.8	59.4	22.2	80.0	126	52.9	84.4	161
75pct	23.2	312	139	90.1	83.6	36.7	105	164	93.8	113	211
90pct	34.4	407	218	130	117	56.1	142	221	142	152	283
RMSE			121	69.5	60.5	20.7	79.5	135	75.5	86.8	179
B. Cross-sectional averages of time-series moments											
Std	4.60	59.1	64.7	25.2	20.9	12.7	20.9	32.1	39.6	23.5	42.3
	(5.91)	(53.3)	(48.6)	(22.3)	(17.4)	(10.8)	(19.4)	(29.4)	(31.8)	(21.6)	(38.4)
$\rho(ES, \cdot)$		0.31	0.15	0.30	0.29	0.30	0.39	0.36	0.23	0.35	0.33
		(0.56)	(0.53)	(0.54)	(0.54)	(0.54)	(0.54)	(0.55)	(0.53)	(0.55)	(0.56)
$\rho(RV, \cdot)$	0.31		0.24	0.46	0.44	0.22	0.67	0.70	0.23	0.67	0.70
	(0.56)		(0.56)	(0.55)	(0.55)	(0.58)	(0.50)	(0.47)	(0.57)	(0.48)	(0.46)
C. Time-series averages of cross-sectional moments											
Std	12.8	84.1	83.7	32.2	27.6	20.7	28.9	43.5	55.3	31.6	56.5
	(3.10)	(23.3)	(32.4)	(14.5)	(10.2)	(7.59)	(9.98)	(15.1)	(23.5)	(12.5)	(21.3)
$\rho(ES, \cdot)$		0.47	0.24	0.41	0.40	0.59	0.58	0.50	0.46	0.56	0.49
		(0.14)	(0.07)	(0.08)	(0.07)	(0.04)	(0.12)	(0.12)	(0.10)	(0.15)	(0.14)
$\rho(RV, \cdot)$	0.47		0.19	0.51	0.46	0.31	0.84	0.87	0.30	0.84	0.88
	(0.14)		(0.09)	(0.11)	(0.12)	(0.08)	(0.04)	(0.04)	(0.11)	(0.04)	(0.04)

Table 17: Regressions of low-frequency measures on TAQ effective spread and 5-minute realized volatility for S&P 600 stocks at the annual frequency. Panel A reports pooled OLS regression estimates. Panel B reports within estimates, that is, pooled OLS regressions on data de-meaned at the stock level. In both panels, the Driscoll-Kraay robust t-statistics are reported in parentheses. Panel C reports the pooled mean-time estimates, that is, the time-series averages of year-by-year cross-sectional regression estimates. The t-statistics, reported in parentheses, are calculated using the Newey-West estimator of the variance of the cross-sectional estimates. The sample period is October 2003 to December 2017 and the sample size is 7,122 stock-years.

	R_M		H_L		H_T		CS_M		CS_D		CS_P		AR_M		AR_D		AR_P	
Panel A. Pooled estimation																		
const	44.8	-13.0	47.3	1.23	48.1	14.4	6.30	-3.03	57.7	0.79	97.8	4.08	14.6	-21.0	60.3	-2.86	123	0.39
	(7.70)	(-1.67)	(11.3)	(0.30)	(10.5)	(4.94)	(6.58)	(-1.49)	(21.2)	(0.23)	(20.4)	(0.81)	(4.42)	(-2.36)	(22.6)	(-0.65)	(21.7)	(0.05)
β_{ES}	2.24	1.01	1.42	0.44	1.10	0.38	1.05	0.85	1.69	0.48	2.26	0.27	2.49	1.73	1.85	0.51	2.99	0.38
	(4.98)	(6.74)	(5.70)	(6.12)	(8.73)	(4.97)	(9.59)	(15.8)	(5.57)	(11.6)	(5.00)	(4.79)	(5.33)	(6.50)	(4.75)	(4.19)	(4.50)	(2.53)
β_{RV}		0.31		0.24		0.18		0.05		0.30		0.50		0.19		0.34		0.65
		(8.10)		(10.4)		(13.6)		(4.59)		(18.0)		(19.3)		(6.64)		(20.7)		(22.1)
R^2	0.09	0.18	0.18	0.46	0.16	0.39	0.35	0.39	0.31	0.85	0.24	0.87	0.25	0.33	0.30	0.83	0.24	0.87
B. Within estimation																		
β_{ES}	2.99	0.81	1.93	0.48	1.39	0.31	1.10	0.73	2.17	0.58	3.09	0.52	2.91	1.70	2.40	0.63	4.04	0.67
	(4.29)	(3.93)	(4.68)	(5.38)	(5.73)	(3.32)	(8.32)	(10.1)	(4.89)	(10.6)	(4.47)	(6.80)	(5.13)	(5.72)	(4.41)	(5.19)	(4.13)	(4.28)
β_{RV}		0.42		0.28		0.21		0.07		0.31		0.50		0.23		0.34		0.65
		(14.0)		(14.1)		(13.8)		(5.69)		(20.1)		(22.8)		(9.05)		(30.8)		(30.0)
R^2	0.07	0.21	0.16	0.48	0.12	0.39	0.22	0.31	0.29	0.85	0.25	0.87	0.17	0.28	0.28	0.83	0.25	0.86
C. Pooled mean-time estimation																		
β_{ES}	1.62	1.33	1.03	0.59	0.86	0.55	0.95	0.92	1.31	0.54	1.68	0.36	2.01	1.83	1.38	0.56	2.17	0.46
	(8.09)	(5.67)	(10.3)	(6.24)	(12.4)	(6.49)	(11.6)	(11.1)	(10.0)	(8.27)	(9.54)	(4.52)	(7.07)	(6.94)	(7.57)	(4.80)	(7.71)	(2.98)
β_{RV}		0.09		0.14		0.10		0.01		0.25		0.42		0.06		0.27		0.55
		(2.72)		(10.5)		(10.2)		(2.04)		(28.2)		(25.3)		(2.91)		(23.7)		(25.3)
R^2	0.06	0.08	0.18	0.32	0.16	0.28	0.35	0.35	0.35	0.76	0.26	0.78	0.22	0.23	0.34	0.75	0.26	0.79

Table 18: Empirical results for foreign exchange rates. The table reports pooled descriptive statistics for the monthly low-frequency measures and pooled least squares regressions of the low-frequency measures on the true spread (ES) and 5-minute realized volatility (RV) at the monthly frequency. Driscoll-Kraay standard errors, which are robust to heteroskedasticity, serial correlation, and cross-sectional dependence are reported in parentheses. The sample period is January 2008 to December 2015 and the sample size is 480 exchange rate-months.

	ES	RV	R_M	H_L	H_T	CS_M	CS_D	CS_P	AR_M	AR_D	AR_P
A. Summary statistics											
Mean	0.68	70.6	25.9	46.3	44.8	13.3	29.0	44.7	13.7	24.2	47.6
10pct	0.43	34.2	0.00	15.5	15.0	0.00	14.1	20.0	0.00	7.74	17.7
25pct	0.53	52.7	0.00	27.8	27.1	4.12	19.1	30.1	0.00	14.4	29.6
Med	0.64	68.1	6.61	40.6	39.6	12.5	26.8	41.8	1.19	22.4	44.4
75pct	0.81	83.6	44.6	56.4	54.6	18.2	36.3	55.5	23.3	31.0	60.3
90pct	1.10	123	95.2	97.7	88.5	34.2	53.6	80.3	49.6	48.3	93.0
Std	0.22	32.7	34.0	32.0	31.0	11.5	13.3	21.5	19.5	14.7	26.7
RMSE			42.3	55.7	53.8	17.0	31.2	48.9	23.4	27.7	53.9
$\rho(ES,\cdot)$		0.56	0.29	0.39	0.38	0.34	0.56	0.52	0.31	0.51	0.51
$\rho(RV,\cdot)$	0.56		0.37	0.71	0.65	0.16	0.80	0.89	0.31	0.82	0.91
B. Regression on ES											
const			-4.80	7.69	8.92	1.21	5.88	9.96	-4.61	1.05	5.85
			(-0.74)	(0.87)	(1.11)	(0.86)	(1.59)	(1.36)	(-1.35)	(0.22)	(0.57)
β_{ES}			45.2	56.8	52.7	17.8	34.0	51.1	27.0	34.0	61.4
			(4.50)	(4.06)	(4.16)	(9.57)	(6.02)	(4.45)	(4.91)	(4.36)	(3.64)
R^2			0.09	0.15	0.14	0.12	0.32	0.27	0.09	0.26	0.26
C. Regression on ES and RV											
const			-9.43	-2.44	0.09	1.43	1.70	1.56	-6.37	-4.00	-4.96
			(-1.89)	(-0.92)	(0.03)	(1.02)	(1.11)	(0.84)	(-2.09)	(-2.90)	(-2.61)
β_{ES}			19.1	-0.32	2.99	19.1	10.5	3.76	17.1	5.53	0.48
			(2.06)	(-0.06)	(0.53)	(7.71)	(5.15)	(1.31)	(3.91)	(2.37)	(0.20)
β_{RV}			0.32	0.69	0.60	-0.02	0.29	0.57	0.12	0.35	0.74
			(5.23)	(16.4)	(12.0)	(-0.66)	(8.14)	(13.7)	(2.56)	(13.0)	(26.2)
R^2			0.15	0.50	0.42	0.12	0.66	0.80	0.12	0.67	0.82

Table 19: Simulation results for the liquidity-adjusted CAPM of Acharya and Pedersen (2005). Panels A–C report the mean beta estimates obtained in the simulation using different measures of stock and market transaction costs. The column labeled "True" reports the true betas, the column labeled "s" reports beta estimates when the true effective spread is used in estimation, and the remaining columns show mean beta estimates when the low-frequency measures are used for estimation. Panel D reports the mean annualized price of market risk (λ_m). The simulations are based on 10,000 replications.

	True	s	R_M	H_L	CS_M	CS_D	CS_P	AR_M	AR_D	AR_P
A. Small β										
β_1	0.50	0.50	0.49	0.51	0.50	0.50	0.50	0.50	0.50	0.50
$\beta_2 \times 10^4$	1.19	1.29	57.0	53.4	6.37	19.8	50.6	17.7	18.2	64.8
$\beta_3 \times 10^4$	1.19	1.39	-3.38	52.8	-1.53	2.65	8.17	-1.72	2.81	10.3
$\beta_4 \times 10^4$	1.19	1.31	-2.95	32.0	-1.45	2.62	8.10	-1.58	2.90	10.2
B. Medium β										
β_1	1.00	1.02	1.01	1.03	1.02	1.02	1.01	1.02	1.02	1.01
$\beta_2 \times 10^4$	2.38	2.66	117.3	105.3	13.1	40.4	103.5	36.4	37.3	132.2
$\beta_3 \times 10^4$	2.38	2.91	-7.59	108.0	-3.26	5.25	16.58	-3.84	5.60	20.8
$\beta_4 \times 10^4$	2.38	2.90	-7.89	106.3	-3.01	5.24	16.67	-3.68	5.65	21.2
C. Large β										
β_1	1.50	1.50	1.48	1.52	1.50	1.50	1.49	1.49	1.50	1.49
$\beta_2 \times 10^4$	3.57	3.96	173.8	146.1	19.5	59.4	152.2	54.0	54.9	194.7
$\beta_3 \times 10^4$	3.57	4.23	-10.4	160.2	-4.65	8.09	25.0	-5.24	8.74	31.5
$\beta_4 \times 10^4$	3.57	4.56	-9.98	188.3	-4.14	8.01	24.3	-4.64	8.62	31.5
D. Price of risk										
λ_m (% p.a.)	5.00	4.97	0.11	-4.58	3.91	-0.58	-5.01	2.69	-0.05	-6.21

Table 20: Simulation results for the three-factor model with aggregate liquidity and volatility risks. The table reports the mean price of market risk (λ_m), aggregate liquidity risk (λ_s), and aggregate volatility risk (λ_σ) obtained in the simulation. The prices of risk are expressed in % per annum. The column labeled "True" reports the true prices of risk, the column labeled "s" reports the estimated prices of risk when the true effective spread is used in estimation, and the remaining columns report the mean estimated prices of risk when the loq-frequency measures are used in estimation. Panel A report results based on daily realized volatility as a proxy for the true volatility, while panel B reports results based on daily range-based volatility. All results are based on 10,000 replications.

	True	s	R_M	H_T	CS_M	CS_D	CS_P	AR_M	AR_D	AR_P
A. Daily realized volatility										
Liquidity and volatility risks priced										
λ_m	5.00	5.02	5.01	5.00	5.01	5.01	5.01	5.01	5.01	5.01
λ_s	-1.00	-0.99	3.59	1.23	-7.12	-7.85	-8.50	-3.93	-1.13	-4.24
λ_σ	-1.00	-0.93	-1.00	-0.97	-0.62	-0.34	-0.34	-1.01	-1.06	-1.16
Volatility risk priced										
λ_m	5.00	5.02	5.01	5.01	5.01	5.01	5.01	5.01	5.01	5.01
λ_s	0.00	0.02	-23.1	-3.54	2.54	-2.22	-4.60	-9.74	-6.86	-1.53
λ_σ	-1.00	-0.96	-0.92	-0.85	-1.14	-1.03	-1.23	-0.85	-0.82	-1.09
Liquidity risk priced										
λ_m	5.00	5.02	5.01	5.00	5.01	5.01	5.01	5.01	5.01	5.01
λ_s	-1.00	-0.99	20.4	9.42	-11.1	-7.12	-4.18	0.81	5.29	1.36
λ_σ	0.00	0.10	-0.03	-0.02	0.65	1.01	1.11	-0.09	-0.18	-0.17
B. Daily realized range										
Liquidity and volatility risks priced										
λ_m	5.00	5.02	5.01	5.00	5.01	5.01	5.01	5.01	5.01	5.01
λ_s	-1.00	-0.99	3.65	0.79	-8.50	-7.11	-8.29	-3.76	-0.73	-3.46
λ_σ	-1.00	-0.91	-0.97	-0.92	-0.34	-0.85	-0.83	-0.97	-0.95	-0.95
Volatility risk priced										
λ_m	5.00	5.02	5.01	5.01	5.01	5.01	5.01	5.01	5.01	5.01
λ_s	0.00	0.02	-23.0	-4.71	-4.60	2.56	-1.61	-9.74	-7.20	-0.13
λ_σ	-1.00	-0.91	-0.90	-0.84	-1.23	-0.94	-0.89	-0.84	-0.91	-0.78
Liquidity risk priced										
λ_m	5.00	5.02	5.01	5.00	5.01	5.01	5.01	5.01	5.01	5.01
λ_s	-1.00	-0.99	20.4	9.18	-4.18	-11.1	-8.11	0.99	5.93	1.46
λ_σ	0.00	0.07	-0.04	0.01	1.11	0.16	0.17	-0.07	0.00	-0.07

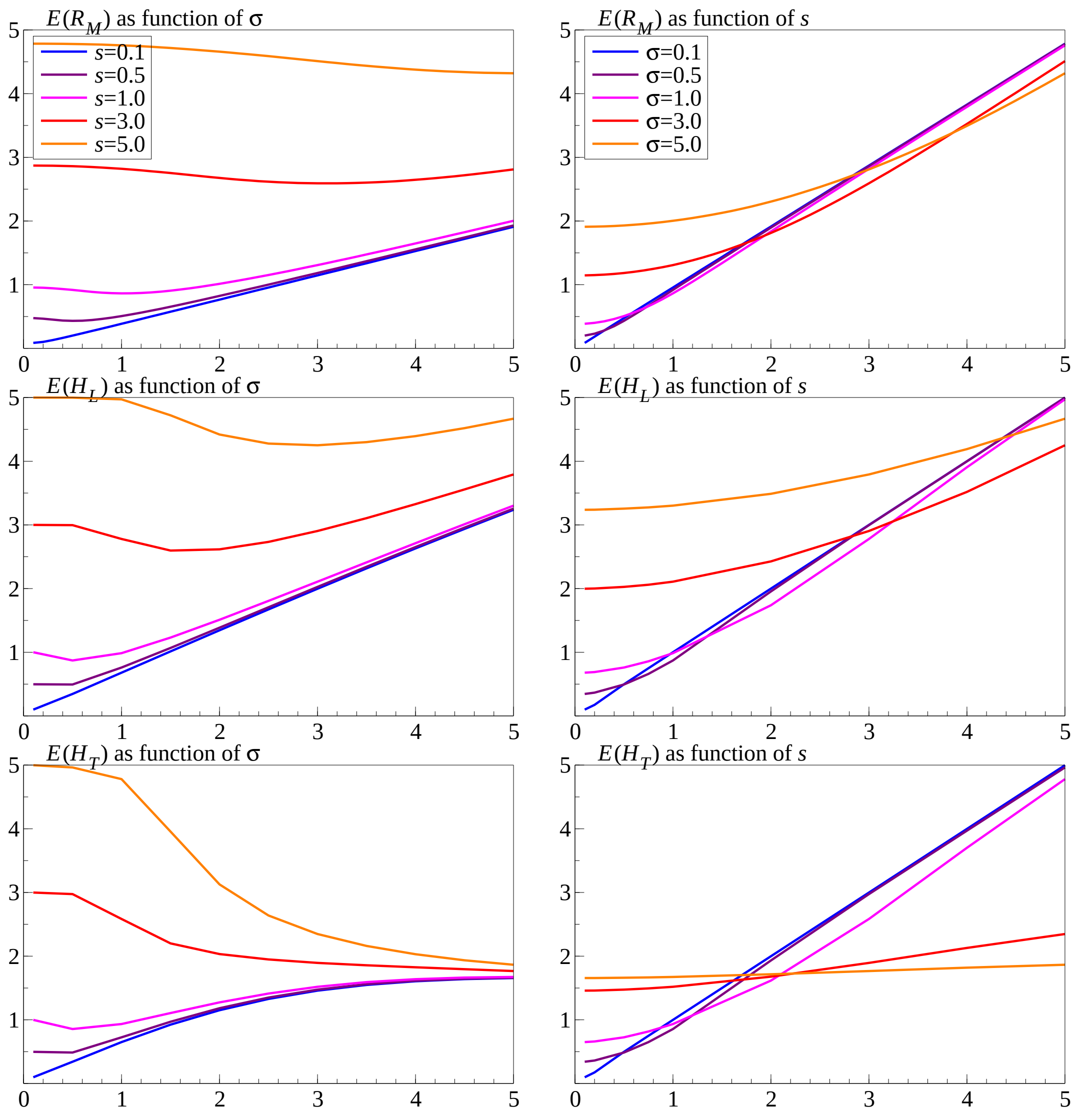

Figure 1: Expectations of the low-frequency measures. The figure shows $\mathbb{E}(R_M)$, $\mathbb{E}(H_T)$, and $\mathbb{E}(H_L)$ as a function of volatility (left panel) and true spread (right panel) when $T = 21$. The expectations are approximated by simulation with 10,000 replications.

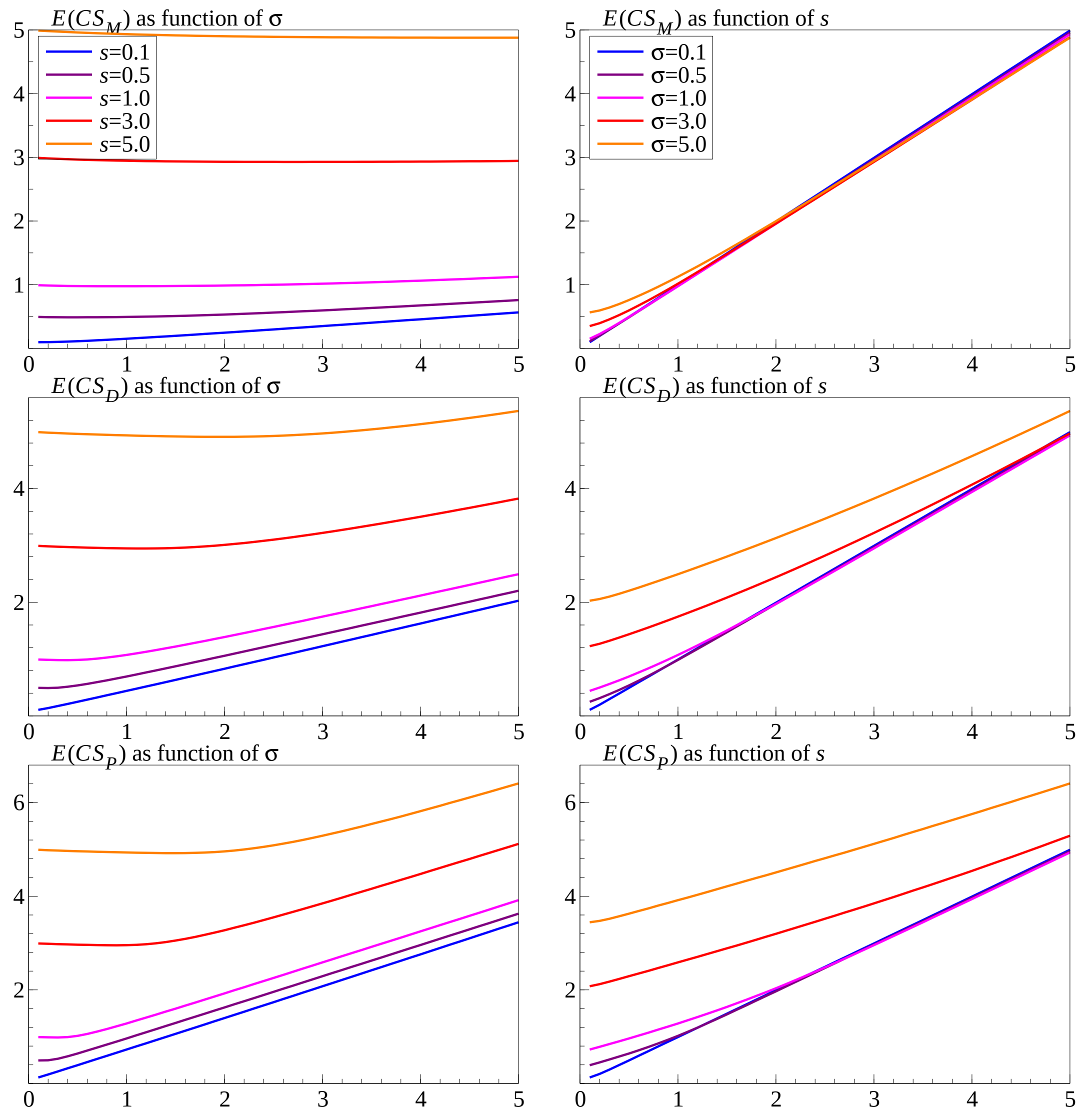

Figure 2: Expectations of the low-frequency measures. The figure shows $\mathbb{E}(CS_M)$, $\mathbb{E}(CS_D)$, and $\mathbb{E}(CS_P)$ as a function of volatility (left panel) and spread (right panel) when $T = 21$. The expectations are approximated by simulation with 10,000 replications.

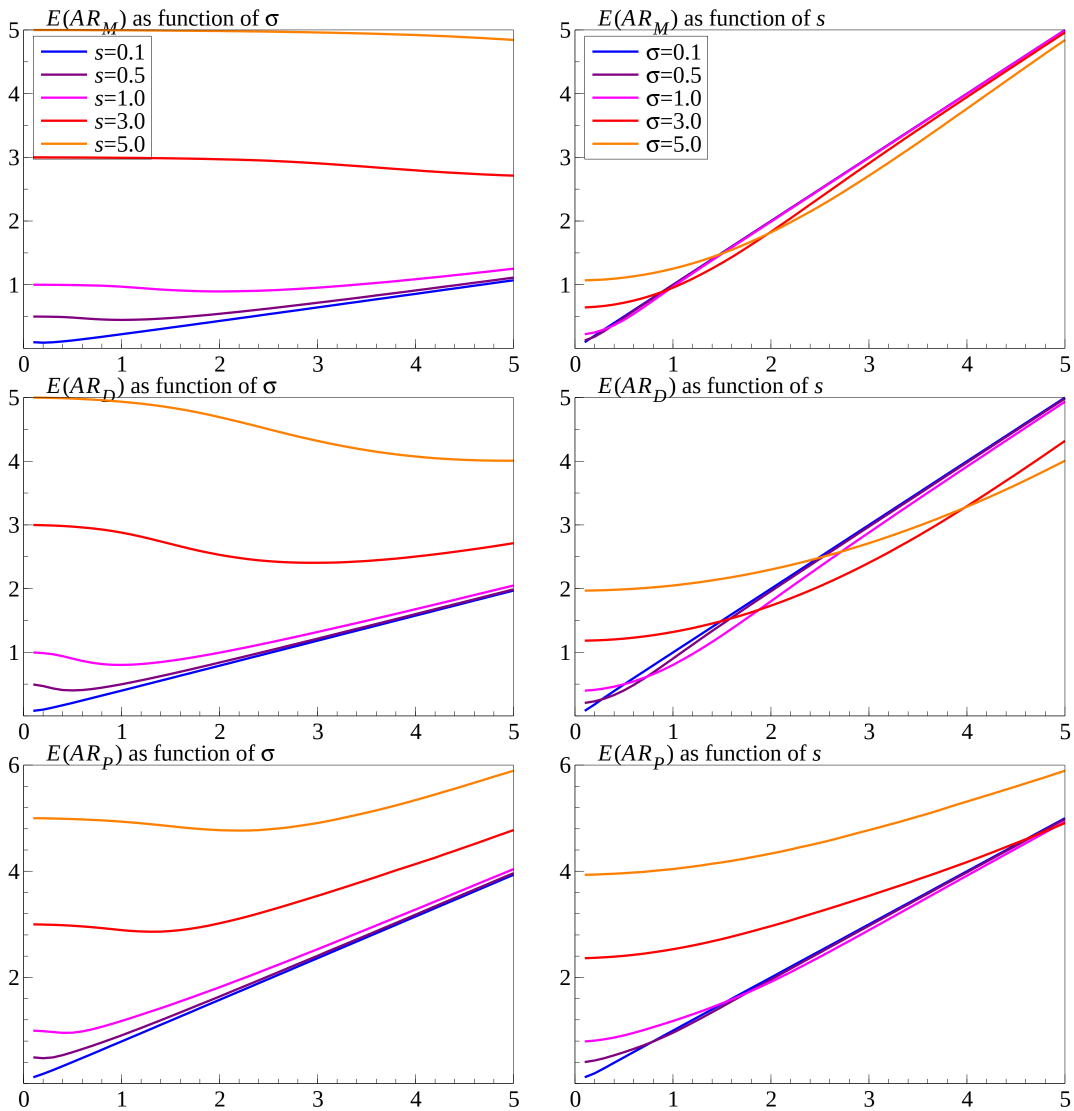

Figure 3: Expectations of the low-frequency measures. The figure shows $\mathbb{E}(AR_M)$, $\mathbb{E}(AR_D)$, and $\mathbb{E}(AR_P)$ as a function of volatility (left panel) and spread (right panel) when $T = 21$.The expectations are approximated by simulation with 10,000 replications.

A Simulation results with $\rho = 0.75$

In this section, we report simulations results with the correlation between the effective spread and volatility set equal to 0.75 rather than 0.5 as in our baseline results reported in Tables 3, 19, and 20. The results are reported in Tables 21, 22, and 23.

Table 21: Correlations of low-frequency measures with the true spread (s_t) and volatility (σ_t) and regression of the low-frequency measures on the true spread and/or volatility. The average volatility equals 2.0% and volatility of volatility equals 2.0%. Both volatility and spread are *iid* lognormally distributed and contemporaneously correlated with $\rho = 0.75$. All simulations are based on 10,000 replications.

		$T=21$									$T=251$								
		R_M	H_L	H_T	CS_M	CS_D	CS_P	AR_M	AR_D	AR_P	R_M	H_L	H_T	CS_M	CS_D	CS_P	AR_M	AR_D	AR_P
A. Small, tranquil spread: E(s) = 0.1, Std(s) = 0.1																			
Corr. with s_t		0.40	0.68	0.55	0.42	0.74	0.75	0.46	0.72	0.74	0.43	0.67	0.62	0.65	0.77	0.77	0.44	0.75	0.76
Corr. with σ_t		0.51	0.89	0.65	0.52	0.94	0.97	0.57	0.92	0.96	0.60	0.89	0.77	0.69	0.99	1.00	0.55	0.99	1.00
Reg. on s_t	α	0.19	0.45	0.63	0.06	0.19	0.32	0.06	0.17	0.36	0.05	0.19	0.30	0.02	0.19	0.33	0.04	0.19	0.38
	β_s	5.99	8.67	2.98	1.93	6.37	10.8	3.84	6.26	12.2	3.66	4.84	2.73	1.34	6.34	10.6	2.01	5.97	11.8
	R^2	0.16	0.46	0.30	0.18	0.54	0.57	0.21	0.52	0.55	0.18	0.45	0.39	0.42	0.60	0.59	0.19	0.57	0.57
Reg. on s_t, σ_t	α	-0.02	0.18	0.56	0.00	0.00	-0.02	-0.05	-0.03	-0.03	-0.05	0.04	0.22	-0.01	0.00	-0.01	-0.02	-0.01	0.00
	β_s	0.30	0.74	0.79	0.30	0.51	0.69	0.59	0.55	0.74	-0.38	0.04	0.41	0.61	0.41	0.42	0.21	0.08	0.10
	β_σ	0.39	0.53	0.15	0.11	0.39	0.68	0.22	0.38	0.77	0.26	0.31	0.15	0.05	0.40	0.68	0.12	0.39	0.79
	R^2	0.26	0.79	0.44	0.27	0.89	0.94	0.32	0.84	0.92	0.36	0.79	0.60	0.51	0.99	1.00	0.31	0.99	0.99
B. Medium, moderately volatile spread: E(s) = 0.5, Std(s) = 0.5																			
Corr. with s_t		0.45	0.71	0.57	0.73	0.84	0.82	0.56	0.77	0.76	0.54	0.75	0.69	0.95	0.88	0.83	0.78	0.82	0.79
Corr. with σ_t		0.52	0.88	0.64	0.62	0.93	0.96	0.54	0.92	0.96	0.54	0.87	0.73	0.75	0.97	0.99	0.61	0.99	0.99
Reg. on s_t	α	0.12	0.45	0.65	0.02	0.18	0.32	0.05	0.17	0.37	0.09	0.17	0.31	0.00	0.20	0.36	-0.01	0.19	0.40
	β_s	1.51	1.85	0.63	1.05	1.75	2.58	1.06	1.36	2.55	0.97	1.17	0.66	1.00	1.72	2.51	0.97	1.33	2.49
	R^2	0.21	0.51	0.33	0.53	0.70	0.66	0.32	0.59	0.58	0.30	0.56	0.48	0.90	0.77	0.70	0.60	0.67	0.63
Reg. on s_t, σ_t	α	-0.01	0.19	0.58	-0.01	0.00	0.00	-0.01	0.00	-0.01	0.02	0.04	0.26	-0.01	0.00	0.01	-0.02	-0.01	0.00
	β_s	0.51	0.33	0.24	0.88	0.67	0.67	0.67	0.31	0.32	0.58	0.34	0.32	0.95	0.65	0.60	0.92	0.28	0.31
	β_σ	0.32	0.51	0.13	0.06	0.37	0.64	0.13	0.36	0.75	0.13	0.27	0.11	0.02	0.37	0.66	0.02	0.36	0.75
	R^2	0.28	0.79	0.43	0.54	0.91	0.95	0.35	0.86	0.93	0.34	0.78	0.58	0.91	0.99	1.00	0.61	0.98	0.99
C. Large, volatile spread: E(s) = 3.0, Std(s) = 3.0																			
Corr. with s_t		0.82	0.90	0.42	0.98	0.98	0.97	0.97	0.98	0.97	0.98	0.98	0.77	1.00	1.00	0.98	1.00	0.99	0.98
Corr. with σ_t		0.63	0.68	0.18	0.73	0.78	0.85	0.71	0.68	0.83	0.74	0.70	0.39	0.73	0.78	0.85	0.73	0.69	0.84
Reg. on s_t	α	-0.06	0.44	1.27	0.00	0.09	0.24	-0.04	-0.09	0.16	0.01	0.02	0.86	-0.01	0.08	0.23	-0.01	-0.08	0.17
	β_s	0.94	0.76	0.14	0.98	0.99	1.04	1.00	0.91	1.00	0.98	0.96	0.55	0.98	1.00	1.04	1.00	0.91	1.00
	R^2	0.68	0.81	0.17	0.96	0.96	0.93	0.95	0.95	0.93	0.95	0.96	0.59	1.00	0.99	0.96	1.00	0.99	0.96
Reg. on s_t, σ_t	α	-0.06	0.42	1.33	0.00	0.00	0.01	-0.03	-0.03	-0.03	0.02	0.09	1.08	0.00	-0.01	-0.01	-0.01	-0.01	-0.03
	β_s	0.93	0.74	0.21	0.97	0.90	0.81	1.01	0.97	0.81	0.99	1.03	0.79	0.98	0.91	0.82	1.00	0.98	0.81
	β_σ	0.01	0.03	-0.13	0.01	0.18	0.47	-0.03	-0.12	0.39	0.00	-0.14	-0.46	-0.01	0.17	0.47	0.00	-0.13	0.38
	R^2	0.68	0.81	0.21	0.96	0.97	0.97	0.95	0.96	0.96	0.95	0.97	0.67	1.00	1.00	0.99	1.00	0.99	0.99

Table 22: Simulation results for the liquidity-adjusted CAPM of Acharya and Pedersen (2005). Panels A–C report the mean beta estimates obtained in the simulation using different measures of stock and market transaction costs. The column labeled "True" reports the true betas, the column labeled "s" reports beta estimates when the true effective spread is used in estimation, and the remaining columns show mean beta estimates when the low-frequency measures are used for estimation. Panel D reports the mean annualized price of market risk (λ_m). The simulations are based on 10,000 replications.

	True	s	R_M	H_L	CS_M	CS_D	CS_P	AR_M	AR_D	AR_P
A. Small β										
β_1	0.50	0.50	0.49	0.51	0.50	0.50	0.50	0.50	0.50	0.50
$\beta_2 \times 10^4$	1.19	1.26	57.4	53.6	6.61	20.6	52.2	17.8	18.4	65.3
$\beta_3 \times 10^4$	1.19	1.31	-2.06	53.5	-1.38	3.10	8.90	-1.24	3.15	11.1
$\beta_4 \times 10^4$	1.19	1.34	-2.53	32.9	-1.38	3.13	8.85	-1.47	3.04	11.2
B. Medium β										
β_1	1.00	1.02	1.01	1.03	1.02	1.02	1.01	1.02	1.02	1.02
$\beta_2 \times 10^4$	2.38	2.60	118.1	105.5	13.6	42.16	106.5	36.6	37.6	133.4
$\beta_3 \times 10^4$	2.38	2.75	-4.73	109.1	-2.89	6.19	17.98	-2.64	6.33	22.4
$\beta_4 \times 10^4$	2.38	2.91	-4.48	108.4	-2.58	6.70	18.9	-2.44	6.60	23.1
C. Large β										
β_1	1.50	1.50	1.48	1.52	1.50	1.50	1.49	1.49	1.50	1.49
$\beta_2 \times 10^4$	3.57	3.87	175.0	146.2	20.2	62.1	156.7	54.3	55.4	196.2
$\beta_3 \times 10^4$	3.57	4.17	-8.14	161.4	-4.50	9.25	27.0	-4.56	9.27	33.3
$\beta_4 \times 10^4$	3.57	3.68	-5.78	189.5	-4.28	9.34	26.9	-3.25	9.71	33.1
D. Price of risk										
λ_m (% p.a.)	5.00	5.02	0.17	-4.53	4.02	-0.50	-4.93	2.77	0.00	-6.17

Table 23: Simulation results for the three-factor model with aggregate liquidity and volatility risks. The table reports the mean price of market risk (λ_m), aggregate liquidity risk (λ_s), and aggregate volatility risk (λ_σ) obtained in the simulation. The prices of risk are expressed in % per annum. The column labeled "True" reports the true prices of risk, the column labeled "s" reports the estimated prices of risk when the true effective spread is used in estimation, and the remaining columns report the mean estimated prices of risk when the loq-frequency measures are used in estimation. Panel A report results based on daily realized volatility as a proxy for the true volatility, while panel B reports results based on daily range-based volatility. All results are based on 10,000 replications.

	True	s	R_M	H_T	CS_M	CS_D	CS_P	AR_M	AR_D	AR_P
A. Daily realized volatility										
Liquidity and volatility risks priced										
λ_m	5.00	4.97	4.96	4.96	4.96	4.96	4.96	4.96	4.96	4.96
λ_s	-1.00	-1.04	-6.57	-5.62	-4.18	-5.51	-7.28	-5.11	-3.13	-4.45
λ_σ	-1.00	-1.00	-1.09	-1.06	-0.88	-0.80	-0.87	-1.07	-1.09	-1.18
Volatility risk priced										
λ_m	5.00	4.97	4.96	4.96	4.96	4.96	4.96	4.96	4.96	4.96
λ_s	0.00	-0.03	-25.7	-11.3	4.59	-0.75	-3.38	-9.94	-7.41	-4.16
λ_σ	-1.00	-1.08	-0.93	-0.89	-1.28	-1.29	-1.62	-0.85	-0.80	-0.92
Liquidity risk priced										
λ_m	5.00	4.97	4.96	4.96	4.96	4.96	4.96	4.96	4.96	4.96
λ_s	-1.00	-1.04	11.4	4.65	-9.80	-5.46	-3.89	0.40	3.60	2.33
λ_σ	0.00	0.07	-0.20	-0.18	0.43	0.58	0.72	-0.22	-0.29	-0.36
B. Daily realized range										
Liquidity and volatility risks priced										
λ_m	5.00	4.97	4.96	4.96	4.96	4.96	4.96	4.96	4.96	4.96
λ_s	-1.00	-1.04	-5.34	-1.75	-7.28	-3.83	-5.47	-3.57	-2.11	-3.20
λ_σ	-1.00	-1.01	-1.07	-1.04	-0.87	-1.00	-1.00	-1.06	-1.03	-0.97
Volatility risk priced										
λ_m	5.00	4.97	4.96	4.96	4.96	4.96	4.96	4.96	4.96	4.96
λ_s	0.00	-0.03	-28.8	-15.5	-3.38	4.28	-4.46	-11.5	-9.99	-13.1
λ_σ	-1.00	-0.98	-0.90	-1.00	-1.62	-0.93	-0.77	-0.85	-0.97	-0.96
Liquidity risk priced										
λ_m	5.00	4.97	4.96	4.96	4.96	4.96	4.96	4.96	4.96	4.96
λ_s	-1.00	-1.04	16.2	14.7	-3.89	-9.12	-2.69	4.39	7.50	11.5
λ_σ	0.00	-0.04	-0.21	-0.06	0.72	-0.05	-0.18	-0.22	-0.08	0.01

B Equity results with 30-minute realized volatility

In this section, we re-run our monthly regressions of low-frequency measures on the TAQ effective spreads and realized volatility, calculating RV from 30-minute intraday returns rather than 5-minute returns used for our baseline results reported in Tables 7, 11, and 15. As can be clearly seen from the following tables, the results with 30-minute RV are remarkably similar to those with 5-minute RV for all stocks in our sample (Table 24) as well as for the very large caps (Table 25) and small caps (Table 26).

Table 24: Regressions of low-frequency measures on TAQ effective spread and 30-minute realized volatility for S&P 1500 stocks at the monthly frequency. Panel A reports pooled OLS regression estimates. Panel B reports within estimates, that is, pooled OLS regressions on data de-meaned at the stock level. In both panels, the Driscoll-Kraay robust t-statistics are reported in parentheses. Panel C reports the pooled mean-time estimates, that is, the time-series averages of month-by-month cross-sectional regression estimates. The t-statistics, reported in parentheses, are calculated using the Newey-West estimator of the variance of the cross-sectional estimates. The sample period is October 2003 to December 2017 and the sample size is 241,236 stock-months.

	R_M		H_L		H_T		CS_M		CS_D		CS_P		AR_M		AR_D		AR_P	
Panel A. Pooled estimation																		
const	58.5	-3.02	129	-11.0	92.9	65.0	12.9	5.65	50.3	11.5	85.1	16.8	31.6	0.62	54.4	6.45	109	12.9
	(16.1)	(-0.50)	(33.8)	(-0.72)	(28.7)	(13.7)	(16.2)	(4.73)	(25.6)	(4.22)	(25.2)	(3.48)	(17.9)	(0.15)	(25.3)	(1.79)	(25.7)	(1.96)
β_{ES}	2.54	0.46	3.10	-1.65	1.14	0.20	1.09	0.84	2.08	0.76	3.00	0.69	2.15	1.10	2.24	0.62	3.90	0.66
	(6.76)	(4.73)	(8.78)	(-6.14)	(30.0)	(2.76)	(14.7)	(19.1)	(8.54)	(19.9)	(7.29)	(9.98)	(7.82)	(8.52)	(7.06)	(15.1)	(6.50)	(11.0)
β_{RV}		0.45		1.02		0.20		0.05		0.28		0.50		0.22		0.35		0.69
		(9.82)		(9.34)		(9.80)		(5.05)		(12.6)		(13.0)		(8.02)		(13.7)		(14.9)
R^2	0.06	0.19	0.03	0.32	0.10	0.33	0.15	0.17	0.26	0.65	0.23	0.72	0.11	0.21	0.22	0.65	0.21	0.75
B. Within estimation																		
β_{ES}	3.34	0.54	3.30	-3.22	1.07	-0.07	1.09	0.77	2.55	0.94	3.93	1.12	2.50	1.09	2.90	0.86	5.22	1.21
	(4.61)	(4.50)	(4.06)	(-10.2)	(13.8)	(-0.62)	(7.91)	(15.6)	(5.36)	(18.7)	(4.86)	(14.8)	(5.24)	(6.25)	(4.91)	(10.9)	(4.58)	(9.96)
β_{RV}		0.47		1.09		0.19		0.05		0.27		0.47		0.24		0.34		0.67
		(9.52)		(8.91)		(11.6)		(4.45)		(10.4)		(10.8)		(7.54)		(12.7)		(13.4)
R^2	0.04	0.17	0.02	0.30	0.04	0.24	0.07	0.09	0.20	0.57	0.20	0.66	0.07	0.16	0.18	0.58	0.19	0.69
C. Pooled mean-time estimation																		
β_{ES}	1.87	0.69	2.52	-2.52	1.10	0.33	0.99	0.91	1.67	0.93	2.30	0.98	1.70	1.09	1.68	0.74	2.86	0.92
	(17.2)	(7.83)	(24.9)	(-11.8)	(19.1)	(10.2)	(20.4)	(18.6)	(22.7)	(23.5)	(22.3)	(20.8)	(15.7)	(11.7)	(18.4)	(15.1)	(19.0)	(14.3)
β_{RV}		0.32		1.42		0.22		0.02		0.20		0.35		0.17		0.25		0.52
		(28.4)		(24.0)		(20.9)		(7.64)		(28.3)		(27.9)		(23.4)		(31.7)		(33.4)
R^2	0.04	0.11	0.03	0.35	0.10	0.29	0.14	0.15	0.29	0.50	0.26	0.56	0.09	0.14	0.23	0.50	0.22	0.59

Table 25: Regressions of low-frequency measures on TAQ effective spread and 30-minute realized volatility for DJIA 30 stocks at the monthly frequency. Panel A reports pooled OLS regression estimates. Panel B reports within estimates, that is, pooled OLS regressions on data de-meaned at the stock level. In both panels, the Driscoll-Kraay robust t-statistics are reported in parentheses. Panel C reports the pooled mean-time estimates, that is, the time-series averages of month-by-month cross-sectional regression estimates. The t-statistics, reported in parentheses, are calculated using the Newey-West estimator of the variance of the cross-sectional estimates. The sample period is October 2003 to December 2017 and the sample size is 4,791 stock-months.

	R_M		H_L		H_T		CS_M		CS_D		CS_P		AR_M		AR_D		AR_P	
Panel A. Pooled estimation																		
const	25.4	-7.54	54.7	1.01	62.0	43.4	8.21	5.96	24.5	6.08	39.5	5.51	20.7	9.06	26.7	2.68	51.1	3.38
	(3.17)	(-1.26)	(4.92)	(0.13)	(16.3)	(7.83)	(6.22)	(5.08)	(4.72)	(4.03)	(4.03)	(2.31)	(6.14)	(3.04)	(4.22)	(2.01)	(3.88)	(1.49)
β_{ES}	7.33	-0.60	11.4	-1.53	4.44	-0.02	1.76	1.22	5.77	1.33	9.60	1.41	3.87	1.06	6.54	0.75	12.5	1.02
	(2.36)	(-0.56)	(2.90)	(-1.87)	(3.42)	(-0.05)	(4.82)	(5.35)	(3.06)	(4.53)	(2.78)	(3.57)	(3.39)	(2.13)	(2.88)	(2.67)	(2.73)	(2.45)
β_{RV}		0.52		0.84		0.29		0.04		0.29		0.53		0.18		0.38		0.75
		(10.1)		(12.9)		(8.10)		(5.52)		(17.7)		(20.4)		(9.78)		(26.2)		(32.2)
R^2	0.04	0.24	0.07	0.43	0.07	0.35	0.05	0.07	0.20	0.71	0.20	0.82	0.04	0.12	0.17	0.72	0.19	0.84
B. Within estimation																		
β_{ES}	8.32	0.09	11.6	-1.73	4.71	0.27	1.94	1.31	6.08	1.57	9.99	1.74	4.40	1.41	6.84	0.99	12.7	1.18
	(2.17)	(0.08)	(2.54)	(-1.96)	(2.93)	(0.50)	(3.94)	(4.66)	(2.77)	(4.48)	(2.54)	(3.62)	(3.04)	(2.70)	(2.64)	(3.35)	(2.49)	(2.72)
β_{RV}		0.53		0.85		0.28		0.04		0.29		0.53		0.19		0.38		0.74
		(9.61)		(13.6)		(9.27)		(5.90)		(16.4)		(18.8)		(9.84)		(24.2)		(28.4)
R^2	0.04	0.23	0.05	0.41	0.06	0.32	0.04	0.07	0.18	0.68	0.18	0.79	0.04	0.12	0.15	0.69	0.16	0.81
C. Pooled mean-time estimation																		
β_{ES}	2.93	-0.83	10.6	-2.95	3.59	-0.52	0.73	0.91	3.45	1.27	6.10	1.82	2.46	0.80	3.98	0.57	8.38	1.27
	(3.62)	(-0.87)	(7.14)	(-3.21)	(7.55)	(-1.17)	(4.22)	(5.03)	(9.97)	(7.05)	(8.48)	(7.55)	(5.27)	(1.64)	(8.27)	(2.25)	(8.33)	(3.21)
β_{RV}		0.30		1.27		0.47		-0.02		0.19		0.37		0.16		0.30		0.62
		(8.85)		(13.8)		(15.1)		(-1.95)		(17.0)		(19.1)		(8.47)		(24.7)		(27.7)
R^2	0.06	0.15	0.11	0.49	0.08	0.35	0.05	0.11	0.15	0.35	0.17	0.48	0.06	0.14	0.13	0.41	0.15	0.55

Table 26: Regressions of low-frequency measures on TAQ effective spread and 30-minute realized volatility for S&P 600 stocks at the monthly frequency. Panel A reports pooled OLS regression estimates. Panel B reports within estimates, that is, pooled OLS regressions on data de-meaned at the stock level. In both panels, the Driscoll-Kraay robust t-statistics are reported in parentheses. Panel C reports the pooled mean-time estimates, that is, the time-series averages of month-by-month cross-sectional regression estimates. The t-statistics, reported in parentheses, are calculated using the Newey-West estimator of the variance of the cross-sectional estimates. The sample period is October 2003 to December 2017 and the sample size is 95,923 stock-months.

	R_M		H_L		H_T		CS_M		CS_D		CS_P		AR_M		AR_D		AR_P	
Panel A. Pooled estimation																		
const	64.0	-1.49	154	-6.86	104	78.1	16.0	6.85	56.8	14.6	94.4	21.9	33.2	-4.35	57.8	5.74	116	13.6
	(16.5)	(-0.18)	(41.7)	(-0.40)	(34.3)	(16.3)	(16.8)	(4.76)	(26.2)	(4.21)	(25.9)	(3.58)	(10.4)	(-0.74)	(19.9)	(1.14)	(21.4)	(1.47)
β_{ES}	2.19	0.58	2.13	-1.81	0.73	0.10	0.97	0.74	1.72	0.68	2.42	0.65	2.02	1.09	1.95	0.67	3.28	0.77
	(5.75)	(6.29)	(5.62)	(-7.68)	(23.7)	(1.54)	(12.9)	(21.4)	(7.48)	(23.1)	(6.33)	(14.4)	(6.85)	(8.49)	(6.23)	(11.3)	(5.62)	(9.53)
β_{RV}		0.43		1.05		0.17		0.06		0.28		0.47		0.25		0.34		0.67
		(9.06)		(9.76)		(10.6)		(6.16)		(12.6)		(12.5)		(8.26)		(13.3)		(13.9)
R^2	0.06	0.17	0.02	0.29	0.07	0.26	0.14	0.16	0.26	0.60	0.22	0.68	0.12	0.22	0.23	0.61	0.22	0.71
B. Within estimation																		
β_{ES}	2.73	0.73	2.41	-2.67	0.71	0.01	1.00	0.73	2.04	0.89	3.04	1.06	2.27	1.14	2.36	0.89	4.12	1.26
	(5.10)	(5.31)	(4.29)	(-9.28)	(15.5)	(0.08)	(9.70)	(15.5)	(6.37)	(16.9)	(5.64)	(14.4)	(5.81)	(7.10)	(5.55)	(10.2)	(5.09)	(9.76)
β_{RV}		0.45		1.13		0.16		0.06		0.26		0.44		0.25		0.33		0.64
		(8.48)		(9.17)		(11.6)		(5.54)		(10.9)		(10.9)		(7.90)		(12.7)		(13.1)
R^2	0.05	0.15	0.01	0.28	0.04	0.19	0.08	0.11	0.23	0.54	0.23	0.63	0.09	0.18	0.21	0.56	0.22	0.67
C. Pooled mean-time estimation																		
β_{ES}	1.44	0.59	1.41	-2.23	0.70	0.25	0.85	0.79	1.28	0.75	1.70	0.74	1.47	1.01	1.35	0.66	2.18	0.77
	(13.7)	(7.68)	(13.1)	(-10.5)	(16.4)	(9.89)	(17.2)	(16.5)	(17.8)	(17.0)	(17.0)	(14.1)	(13.9)	(11.3)	(15.2)	(12.1)	(15.0)	(10.1)
β_{RV}		0.31		1.41		0.18		0.02		0.19		0.34		0.17		0.25		0.51
		(24.1)		(21.5)		(17.6)		(7.49)		(24.0)		(24.7)		(19.8)		(28.6)		(29.7)
R^2	0.04	0.10	0.02	0.32	0.06	0.22	0.12	0.13	0.23	0.43	0.19	0.49	0.09	0.14	0.19	0.45	0.18	0.54

C FX results with alternative high-frequency effective spread proxies

As mentioned in Section 4.2.1, we do not observe continuous quotes for our foreign exchange rates but only 100–millisecond snapshots. Thus, transaction prices cannot be directly compared to prevailing mid-quotes but only to the most recently observable mid-quotes. Additionally, the latencies associated with trading in three distant geographical locations—London, New York, and Tokyo—can give rise to measurement error in trade time stamps. As a result, the volume-weighted effective spread calculated for day t according to

$$ES_t = \frac{1}{\sum_i v_{i,t}} \sum_i v_{i,t} q_{i,t} (p_{i,t} - m_{i,t})/m_{i,t} \tag{33}$$

can potentially over– or underestimate the true spread depending on how the mid-quote moves between the last snaphot and the time of the transaction; in principle, $q_{i,t}(p_{i,t} - m_{i,t})$ can even turn negative. For robustness, we therefore consider two alternative ways of calculating the effective spread from the EBS data, where we impose the restriction that the effective spread cannot be negative for any transaction:

$$\widehat{ES}_t = \frac{1}{\sum_i v_{i,t}} \sum_i v_{i,t} \max\{q_{i,t}(p_{i,t} - m_{i,t}), 0\}/m_{i,t}, \tag{34}$$

$$\widetilde{ES}_t = \frac{1}{\sum_i v_{i,t}} \sum_i v_{i,t} |p_{i,t} - m_{i,t}|/m_{i,t}. \tag{35}$$

Note that the latter definition is also used to calculate the TAQ effectives spreads for U.S. equities. We then repeat the analysis with these spreads in place of ES_t. In Table 27 we report some summary statistics for the alternative EBS spreads, and in Table 28 we report the regression results analogous to those reported in Table 18.

Table 27: Descriptive statistics for alternative measures of effective spread for FX rates. The mean, median, standard deviation, and percentiles are expressed in basis points. The sample period is 2008–2015.

	Mean	10pct	25pct	Med	75pct	90pct	Std	$\rho(\cdot, RV)$	$\rho(\cdot, \widehat{ES})$	$\rho(\cdot, \widetilde{ES})$
ES	0.68	0.43	0.53	0.64	0.81	1.10	0.22	0.56	0.97	0.91
$\widehat{ES}$	0.86	0.52	0.66	0.82	1.03	1.37	0.28	0.64		0.99
$\widetilde{ES}$	1.04	0.63	0.79	0.98	1.24	1.63	0.34	0.67		

Table 28: Empirical results for foreign exchange rates with alternative EBS effective spread proxies. The table reports pooled least squares regressions of the low-frequency measures on the EBS effective spread ($\widehat{ES}$ or $\widetilde{ES}$) and 5-minute realized volatility (RV) at the monthly frequency. Driscoll-Kraay standard errors, which are robust to heteroskedasticity, serial correlation, and cross-sectional dependence are reported in parentheses. The sample period is January 2008 to December 2015 and the sample size is 479 exchange rate-months.

	R_M	H_L	H_T	CS_M	CS_D	CS_P	AR_M	AR_D	AR_P
A.1. Regression on $\widehat{ES}$									
const	-4.95	1.30	4.18	1.31	3.13	4.33	-4.33	-1.62	-0.19
	(-0.79)	(0.16)	(0.56)	(0.92)	(1.01)	(0.67)	(-1.35)	(-0.39)	(-0.02)
β_{ES}	36.0	53.1	47.4	14.0	30.2	47.0	21.0	30.1	55.7
	(4.66)	(4.94)	(5.08)	(8.51)	(8.11)	(5.82)	(5.00)	(5.50)	(4.46)
R^2	0.09	0.19	0.18	0.11	0.39	0.36	0.09	0.32	0.33
A.2. Regression on $\widehat{ES}$ and RV									
const	-6.81	-2.52	0.70	1.50	1.58	1.12	-5.01	-3.57	-4.40
	(-1.36)	(-0.88)	(0.21)	(1.11)	(1.07)	(0.62)	(-1.77)	(-2.63)	(-2.33)
β_{ES}	11.1	1.98	0.81	16.4	9.35	4.19	12.0	3.90	-0.75
	(1.38)	(0.34)	(0.15)	(8.15)	(5.65)	(1.69)	(3.27)	(2.08)	(-0.41)
β_{RV}	0.33	0.68	0.62	-0.03	0.27	0.57	0.12	0.35	0.75
	(4.98)	(12.9)	(10.9)	(-1.39)	(7.73)	(12.9)	(2.37)	(12.5)	(26.7)
R^2	0.14	0.44	0.43	0.12	0.66	0.80	0.11	0.67	0.82
B.1. Regression on $\widetilde{ES}$									
const	-2.94	1.09	3.95	2.20	3.17	3.53	-2.92	-1.54	-0.74
	(-0.49)	(0.15)	(0.59)	(1.59)	(1.22)	(0.63)	(-0.98)	(-0.43)	(-0.09)
β_{ES}	27.8	44.2	39.5	10.7	24.9	39.7	16.1	24.8	46.6
	(4.54)	(5.45)	(5.63)	(7.67)	(9.68	(6.84)	(4.77)	(6.25)	(4.98)
R^2	0.08	0.20	0.19	0.10	0.41	0.40	0.08	0.34	0.36
B.2. Regression on $\widetilde{ES}$ and RV									
const	-4.49	-1.94	1.19	2.35	1.96	1.04	-3.50	-3.09	-4.07
	(-0.94)	(-0.68)	(0.37)	(1.82)	(1.31)	(0.58)	(-1.31)	(-2.25)	(-2.15)
β_{ES}	5.53	0.67	-0.21	12.9	7.57	3.87	7.81	2.55	-1.29
	(0.83)	(0.13)	(-0.05)	(7.57)	(5.66)	(1.91)	(2.54)	(1.66)	(-0.87)
β_{RV}	0.35	0.68	0.62	-0.03	0.27	0.56	0.13	0.35	0.75
	(5.07)	(12.4)	(10.6)	(-1.46)	(7.65)	(12.7)	(2.47)	(12.4)	(27.3)
R^2	0.14	0.44	0.43	0.11	0.66	0.80	0.11	0.67	0.82