Review of Finance, 2017, 1355–1401
doi: 10.1093/rof/rfx003
Advance Access Publication Date: 13 March 2017

OXFORD

What Are the Best Liquidity Proxies for Global Research?*

Kingsley Y. L. Fong[1], Craig W. Holden[2], and Charles A. Trzcinka[2]

[1]University of New South Wales Business School and [2]Indiana University

Abstract

Liquidity plays an important role in global research. We identify high-quality liquidity proxies based on low-frequency (daily) data, which provide 1,000× to 10,000× computational savings compared to computing high-frequency (intraday) liquidity measures. We find that: (i) Closing Percent Quoted Spread is the best monthly percent-cost proxy when available, (ii) Amihud, Closing Percent Quoted Spread Impact, LOT Mixed Impact, High–Low Impact, and FHT Impact are tied as the best monthly cost-per-dollar-volume proxy, (iii) the daily version of Closing Percent Quoted Spread is the best daily percent-cost proxy, and (iv) the daily version of Amihud is the best daily cost-per-dollar-volume proxy.

JEL classification: C15, G12, G20

Keywords: Liquidity, Transaction costs, Effective spread, Price impact

Received March 16, 2015; accepted January 1, 2017 by Editor Burton Hollifield.

1. Introduction

Rapidly expanding global research analyzes the connection between monthly market liquidity and global asset pricing,[1] global corporate finance,[2] and global market microstructure.[3]

* We thank seminar participants at the China International Conference in Finance, Hong Kong University, Hong Kong University of Science and Technology, Indiana University, Michigan State University, University of New South Wales, University of Sydney Microstructure Meeting, University of Technology, Sydney, and University of Wisconsin, Milwaukee, Burton Hollifield (the editor), and an anonymous referee. We are solely responsible for any errors.

1 See Stahel (2005), Liang and Wei (2006), Bekaert, Harvey, and Lundblad (2007), Chan, Jain, and Xia (2008), Griffin, Kelly, and Nardari (2010), Hearn, Piesse, and Strange (2010), Griffin, Hirschey, and Kelly (2011), Lee (2011), Asness, Moskowitz, and Pedersen (2013), and Bekaert *et al.* (2014).

2 See Bailey, Karolyi, and Salva (2006), LaFond, Lang, and Skaife (2007), Lang, Lins, and Maffett (2012), and Hearn (2014).

3 See Jain (2005), Levine and Schmukler (2006), Henkel, Jain, and Lundblad (2008), Henkel (2008), DeNicolo and Ivaschenko (2009), and Clark (2011).

Further global research analyzes daily liquidity,[4] including: (i) the pricing of daily liquidity risk, (ii) the impact of firm announcements and regulatory changes on daily liquidity, (iii) the interaction between daily market liquidity and daily funding liquidity, (iv) the determinants of daily liquidity, and (v) the commonality of daily liquidity across countries. All of this literature faces great difficulty in trying to compute standard measures of liquidity for a global sample of stocks using intraday trade and quote data, because global intraday data (i) is relatively expensive and (ii) is very large and growing exponentially over time. As an example of the later point, the primary intraday sample used in this article has 8.0 billion trades and 17.7 billion quotes, and is growing at a 32.8% compound annual growth rate.[5] This exponential growth rate of intraday data has kept pace with the exponential growth rate of computer power.[6] Thus, it will continue to be very difficult to compute liquidity for a global sample based on intraday data for the foreseeable future.

A recent literature proposes the use of low-frequency (monthly and daily) liquidity proxies that can be calculated from daily data. These liquidity proxies offer the (globally untested) potential benefit of well-capturing intraday-based liquidity benchmarks and an enormous savings in computational time compared to using intraday data. New liquidity proxies continue to be developed. Corwin and Schultz (2012) develop the High–Low percent-cost proxy and find that it performs better in US data than any other proxy that they test. Chung and Zhang (2014) develop the Closing Percent Quoted Spread percent-cost proxy and find that it generally, but not always,[7] performs better in US data than any other proxy that they test. Neither paper tests these two proxies against each other. We develop a new percent-cost proxy, FHT, which simplifies the existing LOT Mixed measure. It is easy to implement yet retains the core elements of LOT Mixed. Our goal in this article is to identify the best liquidity proxies for global research.

To get a feel for the approximate magnitude of computation savings, we estimate the ratio of high-frequency data points to low-frequency data points.[8] In our sample, this ratio starts at 42× in 1996, rises to 962× by 2007, and reaches 12,616× in 2014. Undoubtedly,

4 Bhattacharya *et al.* (2000), Attig, Gadhoum, and Lang (2003), Gomez-Puig (2006), Gersl and Komarkov (2009), Erten and Okay (2012), Karolyi, Lee, and Van Dijk (2012), Beber and Pagano (2013), and Lee, Tseng, and Yang (2014).

5 To determine the compound annual growth rate, we select the twenty most active stocks on the thirty-seven exchanges for which we have data in 1996 and compare to the twenty most active stocks on the same exchanges in 2007. The quantity of trades and quotes is 22.7 times larger in 2007 than 1996, which translates into a 32.8% annual growth rate.

6 Hennessy and Patterson (2012) report a 31.0% compound annual growth rate of computer power. Specifically, they report that CPU performance based on the SPECint benchmark for the fastest personal computer available each year grew at a 52% annual growth rate from 1986 to 2002 and then slowed to a 20% rate post-2002.

7 In sharp contrast to the rest of their results, they find that for NYSE/AMEX stocks from 1993 to 1996 the Closing Percent Quoted Spread has a −0.5073 time-series correlation with intraday effective spread. This result demonstrates there is no strictly mechanical reason why Closing Percent Quoted Spread must be highly correlated with intraday effective spread.

8 Specifically, we estimate: ratio = (high-frequency data points) / (low-frequency data points) = (high-frequency quotes + trades) / (2*Number of stock-days). We multiple the denominator by two, because liquidity proxies use up to two data points per stock-day. Depending on the particular liquidity proxy being computed, the two data points might be price and volume, high and low price, or closing bid and ask.

the computation savings will continue to grow larger in the years ahead as intraday data continues to grow exponentially versus a linear growth rate in daily data.

Given the enormous computational savings and the potential benefit, low-frequency liquidity proxies have been widely adopted by researchers, including all of the global studies mentioned in the introductory paragraph. Considering that "market liquidity" is a multidimensional concept, there are two major categories of low-frequency liquidity proxies. First are "percent-cost" liquidity proxies, which represent the transaction cost required to execute a small trade. Second are "cost-per-dollar-volume" liquidity proxies, which represent marginal transaction costs per US dollar of volume. They are useful for assessing the marginal cost of trading an additional dollar amount of a large trade.[9] Of the twenty papers using monthly proxies mentioned above, thirteen use percent-cost proxies and thirteen use cost-per-dollar-volume proxies.

Our research design is to compare liquidity proxies to accurate liquidity benchmarks computed using more than a decade of global intraday data. Our primary sample contains 8.0 billion trades and 17.7 billion quotes representing 24,240 firms on 42 exchanges around the world from January 1996 to December 2007. Our secondary sample contains 1.8 billion trades and 14.7 billion quotes representing thirty firms per exchange listed on forty-two exchanges around the world from January 2008 to December 2014. Specifically, we evaluate ten monthly percent-cost proxies relative to four monthly percent-cost benchmarks: percent effective spread, percent quoted spread, percent realized spread, and percent price impact. These benchmarks are standard measures of liquidity from the microstructure literature. We examine thirteen monthly cost-per-dollar-volume proxies relative to a monthly cost-per-dollar-volume benchmark: the slope of the price function, which is often called "lambda" by reference to the same concept in Kyle (1985).

We also test liquidity proxies at the daily frequency. Most liquidity proxies require a reasonable sample size in order to yield reliable results and thus cannot be meaningfully updated at the daily frequency in a responsive[10] manner. However, we are able to examine two daily percent-cost proxies relative to the daily version of the same four percent-cost benchmarks and four daily cost-per-dollar-volume proxies relative to daily lambda.

At both monthly and daily frequencies, we test the proxies using four performance dimensions: (i) higher average cross-sectional correlation with the benchmarks, (ii) higher portfolio time-series correlation with the benchmarks, (iii) higher individual stock time-series correlation with the benchmarks, and (iv) lower average root mean squared error (RMSE) relative to the benchmarks. We find that: (i) Closing Percent Quoted Spread is the best monthly percent-cost proxy when available, (ii) Amihud, Closing Percent Quoted

9 The two categories are apples and oranges, because they are measured in different units and are on different numerical scales. Percent-cost proxies and benchmarks are unit-less measures (i.e., they are measured in percent). Cost-per-dollar-volume proxies and benchmarks are denominated in percent-cost per dollar-of-volume (i.e., %/$). In our sample, all of the percent-cost benchmarks are relatively similar in magnitude and all of them are $10\times$ to $10{,}000\times$ larger than the corresponding cost-per-dollar-volume benchmark. Thus, we strictly compare proxies in one category against benchmarks in the same category.

10 For instance, proxies such as Roll and LOT require at least five daily returns observations to yield a model estimate; Effective Tick requires a distribution of closing prices. Implementing these proxies at daily frequency can only be achieved by using a many-days-long window of data and rolling it forward one day at a time with the bulk of the past data driving the current-day estimate.

Spread Impact, LOT Mixed Impact, High–Low Impact, and FHT Impact are tied as the best monthly cost-per-dollar-volume proxy, (iii) the daily version of Closing Percent Quoted Spread is the best daily percent-cost proxy, and (iv) the daily version of Amihud is the best daily cost-per-dollar-volume proxy. If Closing Percent Quoted Spread is not sufficiently available for a given research purpose, we find that the High–Low and FHT proxies are the next best monthly percent-cost proxies.

Specifically, we find that for both the monthly and daily frequencies Closing Percent Quoted Spread has the highest correlations with percent effective spread, percent quoted spread, percent realized spread, and percent price impact. It provides significant performance gains over the monthly proxies that global research has used to date (Zeros, LOT Mixed, etc.). For example, the global average cross-sectional correlation between monthly Zeros and monthly percent effective spread is 0.406. The corresponding correlation for Closing Percent Quoted Spread is 0.799. At both frequencies, Closing Percent Quoted Spread also does the best job of capturing the level of percent effective spread and percent quoted spread. At both frequencies, High–Low does the best job of capturing the level of percent realized spread and percent price impact. These are the first findings at the daily frequency that liquidity proxies can perform well, which both validates existing research (see footnote 4) and lays the foundation for further daily liquidity studies.

We find that five best monthly cost-per-dollar-volume proxies (Amihud, Closing Percent Quoted Spread Impact, LOT Mixed Impact, High–Low Impact, and FHT Impact) are highly correlated with monthly lambda, but do not capture its level. We find that the daily version of Amihud is highly correlated with daily lambda, but does not capture its level.[11]

High-frequency liquidity measures are used as our benchmarks throughout the article, but we recognize some qualifications in doing this. In particular, the high-frequency cost-per-dollar-volume measure lambda may be noisy. Further, it is possible that some low-frequency measures may pick up relevant facets of liquidity that the high-frequency measures miss. Therefore, our identification of the "best liquidity proxies" should be interpreted with caution.

We extend previous liquidity proxy research such as Lesmond (2005) and Goyenko, Holden, and Trzcinka (2009) by including new proxies that have not been tested against one another (High–Low, Closing Percent Quoted Spread, and FHT), by including the daily liquidity proxy that has never been examined, by including new markets, and by introducing a new proxy evaluation criteria: stock-level time-series correlation that addresses liquidity proxy performance requirements of stock-level studies. We also contribute to the literature by examining the characteristics of a relatively new global intraday equity dataset: Thomson Reuters Tick History (TRTH). We examine how well our TRTH sample matches with Datastream (i.e., matching security identifiers and matching prices) and find that we can match 84.7% of Datastream stock-years from 1996 to 2007. We also compare TRTH's intraday data to Bloomberg's intraday data. For a random sample of fifty stocks per exchange in December 2011, we found the difference between Bloomberg and TRTH

11 For US research at both frequencies covering 1993 to present, the Supplementary Appendix provides a bonus result is that Closing Percent Quoted Spread is the best percent-cost proxy. For US research at the monthly frequency, High–Low is the best percent-cost proxy available before 1993 and, indeed, it goes all the way back to 1926.

percent effective spreads to be 0.07% and the correlation between Bloomberg and TRTH percent effective spreads to be 99.19%. We report the median ratio of the sum of intraday share volume reported by TRTH divided by the share volume reported by Datastream per stock per day. We find that 91% of the exchange-year ratios are exactly 100% and 97% of the exchange-year ratios are in the range (95%, 102%). Combining all of this evidence, we conclude that TRTH is a high-quality, reliable dataset for global research.

The article is organized as follows. Section 2 explains the high-frequency benchmarks. Section 3 introduces a new low-frequency proxy. Section 4 describes the data and our analysis of the TRTH dataset. Section 5 presents monthly percent-cost results. Section 6 presents monthly cost-per-dollar-volume results. Section 7 presents monthly robustness checks by time period and developed versus emerging countries. Section 8 presents daily percent-cost results. Section 9 present daily cost-per-dollar-volume results. Section 10 concludes. The Appendix summarizes the formulas for the low-frequency proxies from the existing literature. An Online Supplementaty Appendix provides additional robustness checks.

2. High-Frequency Benchmarks

The liquidity benchmarks that we study include percent-cost benchmarks, which measure small-trade transaction costs as a percentage of the price, and a cost-per-dollar-volume benchmark, which captures the marginal transaction costs per US dollar of volume. We analyze four high-frequency percent-cost benchmarks and one high-frequency cost-per-dollar-volume benchmark.

2.1 Percent-Cost Benchmarks

Our first percent-cost benchmark is percent effective spread. For a given stock, the percent effective spread on the k-th trade is defined as

$$\text{Percent Effective Spread}_k = 2D_k(\ln(P_k) - \ln(M_k)), \tag{1}$$

where D_k is an indicator variable that equals $+1$ if the k-th trade is a buy and -1 if the k-th trade is a sell, P_k is the price of the k-th trade, and M_k is the midpoint of the consolidated BBO prevailing immediately prior to the time of the k-th trade (i.e., 1 second prior or 1 millisecond prior depending on the unit of time used by each exchange's time-stamp). Aggregating over period (day or month) i, a stock's Percent Effective Spread$_i$ is the volume-weighted[12] average of Percent Effective Spread$_k$ computed over all trades in period i.

Our second percent-cost benchmark is percent quoted spread. For a given time interval s, the percent quoted spread is defined as

$$\text{Percent Quoted Spread}_s = (\text{Ask}_s - \text{Bid}_s)/((\text{Ask}_s + \text{Bid}_s)/2) \tag{2}$$

where Ask$_s$ is the best ask quote and Bid$_s$ is the best bid quote in that time interval. Over period i, the stock's Percent Quoted Spread$_i$ is the time-weighted average of Percent Quoted Spread$_s$ computed over all time intervals in the period.

12 We compute the volume weights in local currency, but we would get the identical weights if we converted both the numerator and denominator to US dollars. We follow the same approach when aggregating Percent Realized Spread and Percent Price Impact (both defined below).

Our third percent-cost benchmark is the percent realized spread, which is the temporary component of the spread (Huang and Stoll, 1996). For a given stock, the percent realized spread on the k-th trade is

$$\text{Percent Realized Spread}_k = 2D_k(\ln(P_k) - \ln(M_{k+5})), \tag{3}$$

where $M_{(k+5)}$ is the midpoint 5 min after the k-th trade and D_k is the buy–sell indicator variable as defined above. We follow the Lee and Ready (1991) method, which specifies that a trade is a buy when $P_k > M_k$, is a sell when $P_k < M_k$, and the tick test is used when $P_k = M_k$. The tick test specifies that a trade is a buy (sell) if the most recent prior trade at a different price was at a lower (higher) price than P_k. Aggregating over period i, a stock's Percent Realized Spread$_i$ is the volume-weighted average of Percent Realized Spread$_k$ computed over all trades in period i.

Our fourth percent-cost benchmark is percent price impact, which is the permanent component of the spread (Huang and Stoll, 1996). For a given stock, the percent price impact on the k-th trade is

$$\text{Percent Price Impact}_k = 2D_k(\ln(M_{k+5}) - \ln(M_k)). \tag{4}$$

For a given stock aggregated over a period i, the Percent Price Impact$_i$ is the volume-weighted average of Percent Price Impact$_k$ computed over all trades in period i.

2.2 Cost-per-Dollar-Volume Benchmarks

Our cost-per-dollar-volume benchmark is λ, which is the slope of the price function. We follow Goyenko, Holden, and Trzcinka (2009) and Hasbrouck (2009), and calculate λ as the slope coefficient of

$$r_n = \lambda \cdot S_n + u_n, \tag{5}$$

where for the n-th 5 min period, r_n is the stock return, $S_n = \sum_k \text{Sign}(v_{kn})\sqrt{|v_{kn}|}$ is the signed square root of US dollar volume, v_{kn} is the signed US dollar volume of the k-th trade in the n-th 5 min period, and u_n is the error term.

3. Low-Frequency Proxies

We analyze ten monthly percent-cost proxies and thirteen monthly cost-per-dollar-volume proxies computed from low-frequency (daily) data. Most of the low-frequency proxies require a number of daily observations to compute because they require statistics such as daily return variance and the proportion of zero return days or they require regression or maximum likelihood estimation. Hence, the liquidity proxy literature has been focusing on evaluating monthly and annual proxies as well as applying filters such as a minimum of ten nonzero return days in a month. While most proxies cannot be computed on a daily basis other than using an incremental approach of moving a large window 1 day at a time, a few liquidity proxies can be meaningfully updated and we evaluate them as daily proxies.

We begin with a description of the monthly proxies that we use in the study. We follow with a description of the daily proxies and how we handle situations where the value of a proxy cannot be computed.

3.1 Monthly Percent-Cost Proxies

Nine of the ten percent-cost proxies that we analyze are from the prior literature: "Roll"from Roll (1984); "LOT Mixed" and "Zeros" from Lesmond, Ogden, and Trzcinka

(1999); "LOT Y-Split" and "Zeros2" from Goyenko, Holden, and Trzcinka (2009); "Effective Tick" from Goyenko, Holden, and Trzcinka (2009) and Holden (2009); "Extended Roll" from Holden (2009); "High–Low" from Corwin and Schultz (2012); and "Closing Percent Quoted Spread" from Chung and Zhang (2014).[13] We introduce a new percent-cost proxy, FHT, which is a simplification of the LOT Mixed model. We start by describing the setup of the LOT Mixed model.

3.1.a *The setup of the LOT Mixed Model*

Lesmond, Ogden, and Trzcinka (1999) develop a percent-cost proxy based on the idea that transaction costs cause a distortion in observed stock returns. The LOT Mixed model assumes that the unobserved "true return" of a stock j on day t is given by

$$R^*_{jt} = \beta_j R_{mt} + \varepsilon_{jt}, \tag{6}$$

where β_j is the sensitivity of stock j to the market return R_{mt} on day t and ε_{jt} is a public information shock on day t. They assume that ε_{jt} is normally distributed with mean zero and variance σ_j^2. Let $\alpha_{1j} \leq 0$ be the percent transaction cost of selling stock j and $\alpha_{2j} \geq 0$ be the percent transaction cost of buying stock j. Then the observed return R_{jt} on a stock j is given by

$$\begin{aligned} R_{jt} &= R^*_{jt} - \alpha_{1j} && \text{when } R^*_{jt} < \alpha_{1j} \\ R_{jt} &= 0 && \text{when } \alpha_{1j} < R^*_{jt} < \alpha_{2j} \\ R_{jt} &= R^*_{jt} - \alpha_{2j} && \text{when } \alpha_{2j} < R^*_{jt}. \end{aligned} \tag{7}$$

The LOT Mixed liquidity measure is simply the difference between the percent buying cost and the percent selling cost:

$$\text{LOT Mixed} = \alpha_{j2} - \alpha_{j1}, \tag{8}$$

where the model's parameters are estimated by maximizing a likelihood function. Goyenko, Holden, and Trzcinka (2009) developed a new version of the measure, which they called LOT Y-Split, by maximizing the same likelihood function over different spatial regions (see the Appendix for details).

Both LOT measures contain two core elements: the proportion of zero returns (from the middle region of Equation (7)) and return volatility. This combination of core elements enables both LOT measures to outperform either Zeros or return volatility separately as shown by Goyenko, Holden, and Trzcinka (2009). However, the complexity and nonanalytic character of the LOT measures open the door to our new liquidity proxy.

3.1.b *FHT*

We create a new percent-cost proxy, FHT, by simplifying the LOT model. First, we assume that transaction costs are symmetric. Let $\alpha_{j2} = S/2$ be the percent transaction cost of buying a stock and $\alpha_{j1} = -S/2$ be the percent transaction cost of selling the same stock, where S is

13 We analyze neither the Gibbs measure from Hasbrouck (2004) nor the Holden measure from Goyenko, Holden, and Trzcinka (2009) and Holden (2009), because both measures are very numerically intensive. Given our large sample, they would be infeasible.

the round-trip, percent transaction cost. Substituting this assumption into Equation (7) and suppressing the subscripts, the observed return R on an individual stock is given by

$$\begin{aligned} R &= R^* + S/2 \quad \text{when } R^* < -S/2 \\ R &= 0 \quad \text{when } -S/2 < R^* < S/2 \\ R &= R^* - S/2 \quad \text{when } S/2 < R^*. \end{aligned} \tag{9}$$

Second, we focus on the return distribution of an individual stock and provide no role for the market portfolio. Specifically, the unobserved "true return" R^* of an individual stock on a single day is assumed to be normally distributed with mean zero and variance σ^2. Thus, the theoretical probability of a zero return is the probability of being in the middle region, which is given by

$$N\left(\frac{S}{2\sigma}\right) - N\left(\frac{-S}{2\sigma}\right). \tag{10}$$

The empirically observed frequency of a zero return is given by the Zeros proxy:

$$z \equiv \text{Zeros} = \frac{\text{ZRD}}{\text{TD}+\text{NTD}}, \tag{11}$$

where ZRD = the number of zero returns days, TD = number of trading days, and NTD = number of no-trade days in a given stock-month. Equating the theoretical probability of a zero return to the empirically observed frequency of a zero return, we obtain

$$N\left(\frac{S}{2\sigma}\right) - N\left(\frac{-S}{2\sigma}\right) = z \tag{12}$$

By the symmetry of the cumulative normal distribution, Equation (12) can be rewritten as

$$N\left(\frac{S}{2\sigma}\right) - \left[1 - N\left(\frac{S}{2\sigma}\right)\right] = z \tag{13}$$

Solving for S, we obtain

$$\text{FHT} \equiv S = 2\sigma N^{-1}\left(\frac{1+z}{2}\right), \tag{14}$$

where $N^{-1}()$ is the inverse function of the cumulative normal distribution. The FHT measure is an analytic measure that can be computed 1,000 times faster than LOT, with a single line of SAS code,[14] and using only return data. For example, Marshall, Nguyen, and Visaltanachoti (2012) are able to compute FHT in commodity markets using only commodity prices. Researchers have already used the FHT measure in recent studies, including Bundgaard and Ahm (2012), Marshall, Nguyen, and Visaltanachoti (2012, 2013), Edmans, Fang, and Zur (2013), Karnaukh, Renaldo, and Soderlind (2015), and Schestag, Schuster, and Uhrig-Homburg (2016).

The intuition of the FHT measure follows from the simple idea that a zero return is the result of the true return being in-between the upper bound given by the transaction cost for buying and the lower bound given by the transaction cost for selling. Holding the volatility

14 The one-line SAS code to compute FHT is: Sigma=Std(NonZeroReturns); Zeros=ZeroReturnDays/TotalDays; FHT = 2*Sigma*Probit((1+Zeros)/2).

of the true return distribution constant, a greater proportion of zero returns implies wider bounds and thus a wider spread. Holding the proportion of zero returns constant, a higher volatility of the true return distribution implies that the transaction cost bounds and bid–ask spread must be larger in order to achieve the same proportion of zero returns. In summary, the percent spread is an increasing function of both the proportion of zero returns and the volatility of the return distribution.

3.2 Monthly Cost-per-Dollar-Volume Proxies

Twelve of the thirteen monthly cost-per-dollar-volume proxies that we study are from the prior literature: ‘‘Amihud’’from Amihud (2002), ‘‘Pastor and Stambaugh’’ from Pastor and Stambaugh (2003), “Amivest” and the Extended Amihud class of proxies from Goyenko, Holden, and Trzcinka (2009). We test ten versions of the Extended Amihud class of proxies by dividing ten different percent-cost proxies by the average US dollar value of daily volume. Nine of these Extended Amihud proxies are from the prior literature: Roll Impact, Extended Roll Impact, Effective Tick Impact, LOT Mixed Impact, LOT Y-Split Impact, Zeros Impact, Zeros2 Impact, High–Low Impact, and Closing Percent Quoted Spread Impact. The tenth version, FHT Impact, is based on dividing our new percent-cost proxy FHT (defined in Section 3.1.b) by the average US dollar value of daily volume. The Appendix summarizes the formulas for the low-frequency proxies from the existing literature.

A key step in making the cost-per-dollar-volume proxies comparable across countries is converting the local currency value of volume into a common currency unit (i.e., US dollars). Thus, the local currency value of volume is converted to US at the average daily exchange rate over the month for monthly measures.

3.3 Daily Liquidity Proxies

We examine the daily version of two percent-cost proxies: High–Low and Closing Percent Quoted Spread. We examine the daily version of four cost-per-dollar-volume proxies: Amihud, Amivest, Closing Percent Quoted Spread Impact, and High–Low Impact. The local currency value of volume is converted to US dollars at the daily exchange rate.

3.4 Trading Activity Filters and Default Value

The computation of most liquidity requires multiple daily observations, hence we impose two trading activity filters in order to have reliable and consistent proxy estimates. We require that a stock have at least five positive-volume days and eleven nonzero return days in the month.

We set up the data such that there is a numerical value for all monthly liquidity proxies. The only exception that we allow to this policy is for Closing Percent Quoted Spread, which is available for 85.6% of stock-years in the primary sample (1996–2007) and for 95.2% of stock-years in the secondary sample (2008–14). We set Closing Percent Quoted Spread variable to missing when quote data is not available in a particular stock-month or on a particular stock-day. We carefully detail its availability by exchange and over time in Table III. The daily sample is based on the stock-days contained within the set of stock-months that have at least five positive-volume days and eleven nonzero return days. For the daily liquidity proxy analysis, we also require that both Closing Percent Quoted Spread and High–Low be nonmissing.

To elaborate on how we make sure that a numerical value is always available for a monthly liquidity proxy given the trading activity filters and quote data availability, we note that seven of the ten percent-cost proxies can always be computed. The three problematic cases are Roll, Extended Roll, and High–Low. For the Roll proxy, the serial correlation of price changes is supposed to be negative. But if the measured serial correlation is positive, which would imply an imaginary value for Roll, we assign Roll to be the default value of zero. This is reasonable approximation since a positive sample serial correlation is most likely to occur when the true, population value of the serial correlation is very small, which corresponds to a highly liquid stock. The Extended Roll proxy has the same problem and same fix as the Roll proxy. For the High–Low proxy, the 2 day volatility is supposed to be twice the 1 day volatility. But if the measured 2 day volatility is much larger than double the 1 day volatility, then the High–Low estimate for that 2 day period will be negative. If this happens for any 2 day period, Corwin and Schultz recommend adjusted the estimate to a default value of zero. Empirically, the High–Low measure for a month is nearly always positive, even if the default value of zero is used for some of the 2 day periods within the month. So the monthly High–Low proxy is nearly always fine. However, the daily High–Low proxy may yield a negative spread, in which case we set the daily High–Low proxy equal to the default value of zero.

Similarly, ten of the thirteen monthly cost-per-dollar-volume proxies can always be computed given the trading activity filters and quote data availability. The three problematic cases are Roll Impact, Extended Roll Impact, and High–Low Impact, where numerator of these measures inherits the same problems as the Roll, Extended Roll, and High–Low proxies. The fixes in these three cases are the same as discussed above.

A subtle point regarding the Amihud measure, which is the average of the ratio of absolute return on day t divided by dollar volume on day t, is that the average is computed over positive volume days only. Since our monthly sample selection filter requires at least five positive volume days, the monthly Amihud measure can always be computed. However, at the daily frequency, if a given day has zero volume, then daily High–Low, daily Closing Percent Quoted Spread, and daily Amihud cannot be computed and we treat this observation as missing. Similarly, daily Amivest has absolute return in the denominator. If the absolute return is zero, then we treat this observation as missing.

4. Data

4.1 Thomson Reuters Tick History

We obtain US intraday trades and quotes data from the New York Stock Exchange Trade and Quote (TAQ) database and other data such as returns and market capitalization from the Center for Research in Security Prices (CRSP) and Compustat. We obtain intraday trades and quotes data of international markets from the TRTH database, and other international data such as returns, market capitalization, securities level information, and daily exchange rates from Datastream. Datastream adds high and low prices and bid and ask prices for a small number of countries beginning in 1987. These variables only become available for a sizable global sample starting in 1994. TAQ, CRSP, Compustat, and Datastream are widely used databases, but the TRTH database is relatively new. Hence, we explain the TRTH database in detail and test how well its data matches Datastream.

The TRTH database is supplied by the Securities Industry Research Centre of Asia-Pacific (SIRCA). TRTH contains historical Reuters data feeds beginning January 1996 on over 5 million instruments from various exchanges. We obtain equity trades and quotes that are time-stamped to whatever time unit an exchange uses and by Reuters to the millisecond.

The TRTH equity database is a survivor-bias-free database that covers both active and inactive stocks. It organizes data by the Reuters Instrument Code (RIC). A “RIC table” includes information such as asset class (e.g., equity), market, currency denomination, the first and the last data date, and the International Securities Identification Number (ISIN) where applicable.[15] A company may have a number of RICs that represent different classes of common shares, preference shares, depository receipts, cross-listings, and securities in special trading status such as deferred settlement after stock split. In order to create a representative sample of RICs of each stock market and to avoid multiple counting, we focus on one common stock per company, traded in the home country and in the local currency. TRTH, however, has limited historical coverage of some of these screening variables so we construct our sample by collecting the securities screening variables from Datastream and identify the matching RICs for the list of screened Datastream securities identifiers.

Datastream identifies each stock by its DSCODE, which is a unique identifier to a security-trading venue combination. Each DSCODE is associated with a comprehensive list of DSCODE information, including, critically, stock split information. We retain only the DSCODEs with an ISIN, in the local market, traded in the local currency and identified as “major security” and “primary quote.” These screening criteria lead to one DSCODE per domestic company per ISIN.

While the TRTH database covers all historically traded symbols on an exchange and their associated intraday data, matching RICs to other databases is not a trivial task. Our experience with the RIC table of the standard TRTH database indicates that comprehensive coverage of ISIN starts from June 2008. Hence, many stocks that became inactive prior to June 2008 often do not have ISIN information from the RIC table. Our data period starts from January 1996, so we need additional data and alternative methods to match RICs and DSCODEs. To this end, we obtain from SIRCA a RIC-DSCODE listing that SIRCA created upon our request from two sources of information. The first source of information is a RIC-DSCODE match list from another commercially available Thomson Reuter database. The second source of information is SIRCA’s RIC-DSCODE matches based on their historical ISIN and SEDOL records. This panel data of RIC attributes data allow us to identify periods in which a RIC is referring to the same attributes, for example, ISIN and company name. We validate each RIC-DSCODE match by checking two variables. First, we check that there are at least 12 month end prices with positive monthly volume from the RIC firm in TRTH and from the DSCODE firm in Datastream. Second, we verify that these TRTH prices and corresponding Datastream prices match within a 10% range at least 90% of the time when stated in the original currency.[16] Since RIC may refer to different stocks over time, we use DSCODE as our primary security identifier after merging the data.

15 The RIC for equity has the structure of company code (often, but not always, the same as the local ticker) plus a security class modifier called the brokerage character and the exchange code. The brokerage character varies by market and we obtain the brokerage character information from TRTH’s date sensitive market and securities reference system “Speedguide.”

16 Specifically, we validate the match by comparing the Datastream price history to the TRTH price history after adjusting for currency reporting differences. TRTH prices are historical prices in the original currency. Datastream unadjusted prices are historical prices in the current currency unit, for

The TRTH data have qualifiers in many markets that contain market specific codes denoting whether a trade is the first trade of the day, an auction trade, and an irregular trade (such as an off-market trade or a trade related to exercising an option). In computing intraday bid–ask spreads, effective spreads, intraday returns, and related measures, we exclude these irregular trades and quotes.

Trading hours differ across exchanges and over time. We determine each exchange's historical trading hour regime by examining for sharp increases and decreases in exchange-level aggregated trade frequency at 5 min intervals in the time series. We cross-check the trading hour regimes based on aggregated trade frequency against the trading hour regimes listed in Reuter's Speedguide and the *Handbook of World Stock, Derivative and Commodity Exchanges*. The liquidity benchmarks that we compute are based on data during trading hours only.

4.2 Our Samples

Our primary sample covers forty-two exchanges in thirty-eight countries. We analyze the leading exchange by volume in thirty-six countries, plus three exchanges in China (the Hong Kong Stock Exchange, Shanghai Stock Exchange, and Shenzhen Stock Exchange), and three exchanges in the US (the New York Stock Exchange, American Stock Exchange, and NASDAQ). Given the large number of stocks and large amount of data in the US market, we select a random sample of 400 firms out of the universe of all eligible US firms in 1996, replace any firms that are ineligible in 1997 with randomly drawn firms out of the universe of all eligible US firms in 1997, and so on rolling forward to 2007. Following the methodology of Hasbrouck (2009), a stock must meet five criteria to be eligible: (i) it has to be a common stock, (ii) it has to be present on the first and last TAQ master file for the year, (iii) it has to have the NYSE, AMEX, or NASDAQ as the primary listing exchange, (iv) it does not change primary exchange, ticker symbol, or cusip over the year, and (v) has to be listed in CRSP. We use the sample of Goyenko, Holden, and Trzcinka (2009) for the years 1996–2005 and extend the sample through 2007. This had the additional advantage of facilitating comparison to the Goyenko, Holden, and Trzcinka (2009) results. Most of the analysis in Tables I–IX is based on the primary sample.

Our secondary sample, which extends forward to cover the same forty-two exchanges from 2008 to 2014, is constructed in a similar manner. We select a random sample of thirty stocks per exchange—stratified by size tercile—out of all eligible firms on a given exchange in 2008. Our stratification is to randomly select ten large stocks, ten medium stocks, and ten small stocks from each exchange. We replace any firms that are ineligible in 2009 with a randomly drawn firm from the same exchange and the same size tercile, and repeat this process rolling forward year by year to 2014. Six of our nine tables include some analysis from the secondary sample as well.

We impose several filters in order to have reliable and consistent proxy estimates. First, we require that a stock have at least five positive-volume days and eleven nonzero return days in the month. The daily sample is based on the stock-days contained within the set of stock-months that have at least five positive-volume days and eleven nonzero return days. Second, for Datastream we follow the recommendation of Ince and Porter (2006) to

example, French stocks prior to 1999 were traded in French franc but reported in euro in Datastream. We convert Datastream prices to the original trading currency. Some differences are not avoidable due to noise. For instance, the bid–ask spread can be over 20% for illiquid stocks, and that Datastream's algorithm to sample end of day price is not stated for each market and over time.

Table I. TRTH match with Datastream and comparison with Bloomberg

Match with Datastream is the percentage of Datastream stock-years in 1996–2007 where we could match TRTH and Datastream records (i.e., matching security identifiers RIC and ISIN and verifying that the month-end prices are within 10% at least 90% of the time). The median TRTH to Datastream volume is the median daily ratio of the sum of intraday share volume reported by TRTH divided by share volume reported by Datastream. The TRTH comparison with Bloomberg is the difference in the TRTH and Bloomberg percent effective spreads and the correlation of the TRTH and Bloomberg percent effective spreads based on a random sample of ten stocks per exchange in December 2011.

		Match with Datastream (1996–2007)				Comparison with Bloomberg			
Country	Exchange	No. of Dstrm Stk-Yrs	Match with Dstrm Stk-Yrs	Match (%)	Med TRTH/ Dstrm Vol (%)	Bloom Eff Spr (%)	TRTH Eff Spr (%)	Dif in Eff Spr (%)	Corr Bloom and TRTH Eff Spd (%)
Argentina	Bue. Ar.	794	679	85.5	100	1.53	1.36	0.17	97.85
Australia	Australian	14,072	11,855	84.2	100	0.44	0.51	−0.07	99.69
Austria	Vienna	999	785	78.6	100	0.32	0.33	−0.01	98.98
Belgium	Brussels	1,480	1,361	92.0	100	0.08	0.09	−0.01	98.38
Brazil	Sao Paulo	910	740	81.3	100	0.72	1.06	−0.34	99.99
Canada	Toronto	12,466	7,254	58.2	100	1.17	0.43	0.74	95.30
Chile	Santiago	1,993	905	45.4	100	1.15	0.95	0.20	99.75
China	Hong K.	8,986	7,945	88.4	100	0.21	0.23	−0.02	99.96
China	Shanghai	7,263	7,042	97.0	99	0.18	0.20	−0.02	99.54
China	Shenzhen	5,437	5,287	97.2	105	0.20	0.20	0.00	99.99
Denmark	Copenhag.	2,208	1,912	86.6	100	0.65	0.63	0.02	99.97
France	Paris	9,662	7,527	77.9	100	0.24	0.26	−0.02	99.99
Finland	Helsinki	1,411	1,313	93.1	100	1.35	1.31	0.04	99.95
Germany	Frankfurt	1,996	1,546	77.5	100	6.51	3.83	2.68	99.14
Greece	Athens	3,174	2,940	92.6	100	3.35	3.26	0.09	99.99
India	Bombay	12,811	10,929	85.3	100	1.45	1.52	−0.07	97.52
Indonesia	Jakarta	3,360	3,325	99.0	100	2.55	2.73	−0.18	99.99
Ireland	Irish	422	345	81.8	100	2.16	2.34	−0.18	95.41
Israel	Tel Aviv	4,957	3,996	80.6	100	2.07	1.79	0.28	99.91
Italy	Milan	2,872	2,735	95.2	100	0.17	0.16	0.01	99.40
Japan	Tokyo	25,834	23,220	89.9	100	0.24	0.25	−0.01	99.98
Malaysia	Kuala Lum.	8,490	8,076	95.1	100	4.40	4.71	−0.31	100.00
Mexico	Mexican	1,303	1,093	83.9	100	0.50	0.53	−0.03	99.96
Nether.	Amsterdam	1,885	1,353	71.8	100	0.10	0.11	−0.01	99.99
New Zea.	New Zea.	923	720	78.0	100	1.60	1.64	−0.04	96.74
Norway	Oslo	2,215	2,059	93.0	100	0.39	0.39	0.00	99.73
Philip.	Philippines	2,289	2,141	93.5	100	–	–	–	–
Poland	Warsaw	992	837	84.4	100	1.54	1.39	0.15	99.15
Portugal	Lisbon	883	158	17.9	100	0.36	0.38	−0.02	99.78
Singapore	Singapore	4,528	4,281	94.5	100	3.06	3.23	−0.17	99.92
S. Africa	Johannes.	4,894	4,403	90.0	100	1.10	1.13	−0.03	99.79
S. Korea	Korea	7,738	7,097	91.7	100	0.22	0.24	−0.02	98.68
Spain	Madrid	1,498	1,406	93.9	100	0.49	0.48	0.01	99.99

(continued)

Table I. Continued

Country	Exchange	Match with Datastream (1996–2007) No. of Dstrm Stk-Yrs	Match with Dstrm Stk-Yrs	Match (%)	Med TRTH/ Dstrm Vol (%)	Comparison with Bloomberg Bloom Eff Spr (%)	TRTH Eff Spr (%)	Dif in Eff Spr (%)	Corr Bloom and TRTH Eff Spd (%)
Sweden	Stockholm	3,768	3,164	84.0	100	0.34	0.36	−0.02	99.89
Switzer.	SWX Swiss	2,872	2,366	82.4	108	0.71	0.73	−0.02	99.86
Taiwan	Taiwan	6,986	6,156	88.1	100	0.32	0.34	−0.02	99.30
Thailand	Thailand	4,536	4,273	94.2	100	0.94	0.95	−0.01	100.00
Turkey	Istanbul	3,020	2,958	97.9	100	0.74	0.74	0.00	99.98
UK	London	18,650	13,382	71.8	100	0.38	0.43	−0.05	100.00
Global Median		200,577	169,564	84.5	100.3	1.16	1.08	0.07	99.30

remove any stock-month with extreme return reversal. Third, we winsorize our data for each liquidity variable by replacing values above the 99th percentile with the 99th percentile value and replacing values below the 1st percentile with the 1st percentile value. Finally, there are two additional subtle filters. By necessity, the data in both monthly and daily frequencies is conditional on the benchmark of the relevant analysis being available. Given that there are only two percent-cost proxies at daily frequency we further require that both proxies be available in daily proxy evaluation.

Our final primary, high-frequency sample has 8.0 billion trades and 17.7 billion quotes. We compute the corresponding benchmarks and proxies for 24,240 firms in 1,491,930 stock-months. Our final secondary, high-frequency sample has 1.8 billion trades and 14.7 billion quotes. We compute the corresponding benchmarks and proxies for 3,087 firms in 84,789 stock-months.

Table I examines how well our TRTH sample matches with Datastream. Each row represents a different exchange. For example, looking at the first row, the country is Argentina and the exchange Bue. Ar., which is short for the Buenos Aires Stock Exchange. The third column lists the number of Datastream stock-years in the sample period 1996–2007. The fourth column lists the number of stock-years where we could match TRTH and Datastream records (i.e., matching security identifiers RIC and ISIN and verifying that the month-end prices are within 10% at least 90% of the time). For the global sample, our percent matched was 84.7%.

We also compare TRTH's intraday data to Bloomberg's intraday data. Since Bloomberg only retains historical data for a few months to a few years, we checked a random sample of ten stocks per exchange in December 2011. For the global sample, we find that the Bloomberg percent effective spread is 1.16% and the TRTH percent effective spread is 1.08% yielding a difference of 0.07%. We also find that the correlation between the Bloomberg percent effective spread and the TRTH percent effective spread is 99.19%. This high correlation implies that correlations between liquidity proxies and TRTH percent effective spread would be nearly identical to correlations between liquidity proxies and Bloomberg percent effective spread.[17]

17 As an additional data integrity test, we checked the trades in our database against the Nordic Security Depository, which is the central clearing agency for all trading in Finland. It includes the

Table II. TRTH and Datastream trading volume comparison

This table reports the median ratio of TRTH to Datastream volume, which is the median value of the daily ratio of the sum of intraday share volume reported by TRTH divided by share volume reported by Datastream.

Country	Exchange	1996 (%)	1997 (%)	1998 (%)	1999 (%)	2000 (%)	2001 (%)	2002 (%)	2003 (%)	2004 (%)	2005 (%)	2006 (%)	2007 (%)	2008–09 (%)	2010–14 (%)	All years (%)
Argentina	Bue. Ar.			100	101	101	103	102	100	100	100	100	100	100	100	100
Australia	Australian	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
Austria	Vienna				50	50	56	100	100	100	100	100	100	100	100	100
Belgium	Brussels				100	100	100	100	100	100	100	100	100	100	100	100
Brazil	Sao Paulo			100	100	93	100	100	100	100	65	100	100	100	100	100
Canada	Toronto	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
Chile	Santiago							100	100	100	100	100	100	100	100	100
China	Hong Kong	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
China	Shanghai	99	99	100	100	100	100	100	100	100	100	100	97	100	100	100
China	Shenzhen	97	95	97	98	97	98	99	100	100	100	100	100	100	100	99
Denmark	Copenhag.	100	100	100	100	100	100	100	100	100	100	100	101	100	100	100
France	Paris				95	96	97	97	100	100	100	100	100	100	100	100
Finland	Helsinki				100	100	100	100	100	100	100	100	100	100	100	100
Germany	Frankfurt				100	100	100	100	100	100	100	100	100	100	100	100
Greece	Athens						100	100	100	100	100	100	100	100	100	100
India	Bombay	100	100	100	100	100	100	100	74	51	61	100	100	100	100	100
Indonesia	Jakarta	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100

(continued)

Table II. Continued

Country	Exchange	1996 (%)	1997 (%)	1998 (%)	1999 (%)	2000 (%)	2001 (%)	2002 (%)	2003 (%)	2004 (%)	2005 (%)	2006 (%)	2007 (%)	2008–09 (%)	2010–14 (%)	All years (%)
Ireland	Irish					100	100	100	100	100	100	100	100	100	100	100
Israel	Tel Aviv			100	100	100	100	100	100	100	100	100	100	100	100	100
Italy	Milan				77	71	75	73	100	102	102	100	100	100	100	100
Japan	Tokyo	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
Malaysia	Kuala Lum.	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
Mexico	Mexican			100	100	100	100	100	100	100	100	100	100	100	100	100
Nether.	Amsterdam				100	100	99	97	100	100	100	101	100	100	100	100
New Zea.	New Zea.	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
Norway	Oslo	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
Philip.	Philippines	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
Poland	Warsaw					100	100	100	100	100	100	100	100	100	100	100
Portugal	Lisbon										100	100	100	100	100	100
Singapore	Singapore	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
S. Africa	Johannes.	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
S. Korea	Korea								100	100	100	100	100	100	100	100
Spain	Madrid				100	100	100	100	100	100	100	100	100	100	100	100
Sweden	Stockholm	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
Switzer.	SWX Swiss	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
Taiwan	Taiwan	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
Thailand	Thailand	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
Turkey	Istanbul										100	100	100	100	100	100
UK	London	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
Global median		100	100	100	100	100	100	100	100	100	100	100	100	100	100	100

Table III. Availability of closing bid and ask prices in Datastream

The percentage of stocks in Datastream that have daily bid and ask prices by exchange-year for the primary sample and by time period for the secondary sample. To be considered, we require that a stock have more than ten nonzero return days.

Country	Exchange	1996 (%)	1997 (%)	1998 (%)	1999 (%)	2000 (%)	2001 (%)	2002 (%)	2003 (%)	2004 (%)	2005 (%)	2006 (%)	2007 (%)	2008–09 (%)	2010–14 (%)
Argentina	Bue. Ar.			90	89	91	89	95	99	94	97	99	99	96	96
Australia	Australian	70	73	74	74	72	73	76	90	92	93	94	96	100	95
Austria	Vienna				83	83	85	85	83	83	90	92	95	81	89
Belgium	Brussels				93	93	92	92	94	93	93	89	96	98	94
Brazil	Sao Paulo			64	81	83	82	81	86	89	93	95	95	100	97
Canada	Toronto	49	49	50	52	52	53	54	56	61	61	69	75	100	97
Chile	Santiago							81	83	82	82	88	90	99	92
China	Hong Kong	87	87	88	89	89	89	90	90	90	89	88	85	100	97
China	Shanghai	100	100	100	100	96	95	96	97	97	98	96	96	97	97
China	Shenzhen	95	95	94	94	95	96	96	99	99	99	99	100	97	96
Denmark	Copenhag.	77	79	78	79	83	83	83	92	94	96	96	97	98	96
France	Paris				75	73	74	74	74	76	91	93	96	100	97
Finland	Helsinki				91	92	94	96	96	99	100	100	100	100	96
Germany	Frankfurt				67	69	72	75	77	81	84	85	90	98	97
Greece	Athens						93	96	98	98	98	99	99	57	89
India	Bombay	82	81	79	82	82	74	81	83	88	91	92	94	62	64

(continued)

Table III. Continued

Country	Exchange	1996 (%)	1997 (%)	1998 (%)	1999 (%)	2000 (%)	2001 (%)	2002 (%)	2003 (%)	2004 (%)	2005 (%)	2006 (%)	2007 (%)	2008–09 (%)	2010–14 (%)
Indonesia	Jakarta	98	99	99	99	99	99	99	99	98	98	99	99	99	95
Ireland	Irish					76	80	81	82	82	83	83	88	99	96
Israel	Tel Aviv	10	14	74	77	79	80	82	83	85	92	95	99	96	96
Italy	Milan				94	95	95	96	97	98	99	97	98	100	96
Japan	Tokyo	79	79	80	81	88	89	91	92	94	96	97	98	100	97
Malaysia	Kuala Lum.	84	85	85	85	96	97	98	99	99	99	99	99	100	96
Mexico	Mexican	19	20	77	83	84	87	94	93	97	98	95	97	98	91
Nether.	Amsterdam				28	86	89	89	91	91	99	99	100	100	96
New Zea.	New Zea.	64	65	65	67	61	66	68	76	83	88	91	93	100	95
Norway	Oslo	73	79	90	90	89	89	96	96	97	98	97	100	100	97
Philip.	Philippines	87	89	90	92	94	92	95	94	95	93	97	97	100	96
Poland	Warsaw					97	98	98	99	99	99	99	100	99	91
Portugal	Lisbon										79	80	92	100	95
Singapore	Singapore	90	94	93	92	91	92	93	94	95	96	97	97	100	96
S. Africa	Johannes.	80	83	83	87	89	89	93	93	93	95	96	97	99	96
S. Korea	Korea	1	81	85	86	87	89	92	96	98	98	99	99	99	97
Spain	Madrid				92	93	92	95	95	95	97	98	98	100	95
Sweden	Stockholm	88	77	75	78	81	82	84	85	86	87	89	91	100	97
Switzer.	SWX Swiss	66	74	77	80	80	82	74	88	86	88	88	88	100	96
Taiwan	Taiwan	92	83	80	75	80	85	88	91	94	94	93	95	91	97
Thailand	Thailand	93	92	92	93	94	95	96	95	95	94	95	95	100	97
Turkey	Istanbul										100	99	99	98	96
UK	London	66	60	62	63	59	72	75	76	78	76	77	80	80	89
Global average		71.7	75.6	81.0	81.5	84.3	85.6	87.2	89.5	90.6	92.3	93.2	94.9	95.9	94.4

Table IV. Mean of the monthly percent-cost benchmarks and proxies

The percent-cost benchmarks (percent effective spread, percent quoted spread, percent realized spread, and percent price impact) are calculated from every trade and corresponding BBO quote in the SIRCA TRTH database for a sample stock-month. The primary sample spans forty-two exchanges around the world from 1996 to 2007. It consists of all stock-months with at least five positive-volume days and eleven nonzero return days. The secondary sample spans forty-two exchanges (thirty stocks per exchange) with the same activity filters from 2008 to 2014. All variables are winsorized at 1st and 99th percentile in each exchange-year.

		Percent-cost benchmarks				Percent-cost proxies														
Country	Exchange	Eff Spr (%)	Quo Spr (%)	Real Spr (%)	Price Imp (%)	Roll	Ext Roll	Eff Tick	LOT Mix	LOT Y	FHT	Zeros	Zeros2	High–Low	Clos Quo Spr (%)	Stocks	Stock-Mon	Quotes (Mil.)	Trades (Mil.)	Start Date
Argentina	Bue. Ar.	0.017	0.025	0.009	0.008	0.002	0.010	0.021	0.035	0.020	0.014	0.203	0.146	0.009	0.017	75	3,965	7	5	7/98
Australia	Australian	0.024	0.029	0.012	0.012	0.009	0.025	0.036	0.064	0.034	0.025	0.238	0.191	0.018	0.031	1,922	84,495	344	155	1/96
Austria	Vienna	0.016	0.016	0.009	0.007	0.004	0.011	0.004	0.021	0.009	0.006	0.117	0.084	0.009	0.015	106	4,459	19	7	1/99
Belgium	Brussels	0.012	0.013	0.006	0.006	0.004	0.012	0.003	0.022	0.009	0.007	0.131	0.104	0.009	0.013	151	8,721	71	29	1/99
Brazil	Sao Paulo	0.023	0.035	0.009	0.014	0.005	0.016	0.008	0.034	0.016	0.011	0.124	0.082	0.013	0.028	139	2,893	29	10	6/98
Canada	Toronto	0.025	0.029	0.011	0.014	0.012	0.026	0.023	0.050	0.022	0.016	0.161	0.124	0.017	0.020	1,235	73,105	927	237	1/96
Chile	Santiago	0.017	0.021	0.011	0.005	0.001	0.006	0.003	0.025	0.014	0.009	0.195	0.154	0.006	0.027	101	3,256	4	2	6/02
China	Hong K.	0.022	0.029	0.013	0.009	0.006	0.021	0.022	0.055	0.029	0.020	0.211	0.158	0.017	0.027	925	66,075	267	140	1/96
China	Shanghai	0.004	0.003	0.000	0.004	0.001	0.008	0.003	0.015	0.005	0.003	0.045	0.031	0.010	0.002	812	77,271	1,109	2,098	1/96
China	Shenzhen	0.004	0.003	0.000	0.004	0.001	0.008	0.003	0.014	0.004	0.003	0.044	0.029	0.010	0.003	677	57,488	502	466	1/96
Denmark	Copenhag.	0.015	0.022	0.010	0.006	0.005	0.013	0.011	0.032	0.016	0.011	0.209	0.149	0.010	0.020	253	12,157	37	17	1/96
France	Paris	0.016	0.018	0.009	0.007	0.004	0.012	0.004	0.025	0.010	0.007	0.120	0.091	0.011	0.017	915	41,098	671	376	1/99

(continued)

Table IV. Continued

Country	Exchange	Percent-cost benchmarks: Eff Spr (%)	Quo Spr (%)	Real Spr (%)	Price Imp (%)	Percent-cost proxies: Roll	Ext Roll	Eff Tick	LOT Mix	LOT Y	FHT	Zeros	Zeros2	High–Low	Clos Quo Spr (%)	Stocks	Stock-Mon	Quotes (Mil.)	Trades (Mil.)	Start Date
Finland	Helsinki	0.013	0.017	0.011	0.003	0.006	0.016	0.013	0.032	0.016	0.011	0.168	0.123	0.010	0.018	161	10,209	120	41	1/99
Germany	Frankfurt	0.020	0.022	0.005	0.015	0.005	0.017	0.010	0.034	0.014	0.009	0.110	0.080	0.010	0.025	592	29,411	443	144	1/99
Greece	Athens	0.019	0.019	0.009	0.010	0.004	0.012	0.011	0.034	0.014	0.010	0.138	0.127	0.013	0.019	356	23,603	68	60	1/01
India	Bombay	0.052	0.060	0.029	0.023	0.015	0.036	0.013	0.070	0.032	0.022	0.123	0.047	0.030	0.073	1,667	82,624	114	61	4/96
Indonesia	Jakarta	0.026	0.041	0.012	0.014	0.010	0.028	0.031	0.081	0.050	0.035	0.290	0.236	0.023	0.038	380	14,278	52	37	1/96
Ireland	Irish	0.017	0.023	0.008	0.010	0.007	0.017	0.018	0.059	0.044	0.014	0.182	0.135	0.012	0.022	55	2,708	9	1	2/00
Israel	Tel Aviv	0.030	0.042	0.009	0.021	0.003	0.012	0.002	0.036	0.020	0.011	0.150	0.048	0.014	0.026	580	22,686	56	30	12/98
Italy	Milan	0.007	0.008	0.004	0.004	0.007	0.015	0.005	0.022	0.008	0.007	0.087	0.075	0.009	0.009	371	23,976	228	243	1/99
Japan	Tokyo	0.009	0.012	0.003	0.006	0.004	0.014	0.010	0.026	0.010	0.007	0.114	0.096	0.008	0.011	2,803	2.7E+05	1,882	754	1/96
Malaysia	Kuala Lum.	0.017	0.028	0.005	0.012	0.007	0.020	0.020	0.043	0.023	0.016	0.223	0.182	0.013	0.025	960	79,565	189	90	1/96
Mexico	Mexican	0.015	0.029	0.005	0.010	0.002	0.010	0.006	0.028	0.014	0.009	0.122	0.078	0.008	0.026	116	5,042	33	12	5/98
Nether.	Amsterdam	0.013	0.013	0.005	0.009	0.004	0.012	0.007	0.026	0.011	0.008	0.115	0.095	0.010	0.014	190	11,620	251	125	1/99
New Zea.	New Zea.	0.017	0.015	0.012	0.005	0.003	0.009	0.022	0.030	0.017	0.012	0.260	0.229	0.007	0.015	99	5,052	4	2	1/96
Norway	Oslo	0.019	0.024	0.013	0.006	0.005	0.017	0.018	0.061	0.041	0.014	0.188	0.140	0.011	0.021	331	15,111	82	34	1/96
Philip.	Philippines	0.024	0.038	0.013	0.011	0.007	0.023	0.032	0.070	0.042	0.029	0.286	0.232	0.017	0.033	218	9,540	17	10	1/96
Poland	Warsaw	0.027	0.034	0.007	0.020	0.007	0.019	0.011	0.063	0.047	0.013	0.152	0.110	0.014	0.028	222	6,819	15	12	11/00
Portugal	Lisbon	0.009	0.008	0.003	0.006	0.003	0.008	0.009	0.019	0.009	0.006	0.173	0.159	0.007	0.008	44	987	11	6	7/05
Singapore	Singapore	0.016	0.025	0.007	0.009	0.007	0.020	0.030	0.051	0.028	0.020	0.250	0.213	0.015	0.024	644	32,313	141	47	1/96

(continued)

Table IV. Continued

		Percent-cost benchmarks				Percent-cost proxies														
Country	Exchange	Eff Spr (%)	Quo Spr (%)	Real Spr (%)	Price Imp (%)	Roll	Ext Roll	Eff Tick	LOT Mix	LOT Y	FHT	Zeros	Zeros2	High–Low	Clos Quo Spr (%)	Stocks	Stock-Mon	Quotes (Mil.)	Trades (Mil.)	Start Date
S. Africa	Johannes.	0.026	0.032	0.011	0.015	0.006	0.019	0.026	0.056	0.030	0.021	0.224	0.177	0.014	0.032	658	28,049	64	34	3/96
S. Korea	Korea	0.015	0.012	0.005	0.010	0.004	0.016	0.004	0.029	0.008	0.007	0.079	0.073	0.014	0.010	750	76,246	817	807	10/97
Spain	Madrid	0.007	0.007	0.002	0.004	0.004	0.009	0.004	0.018	0.006	0.004	0.092	0.088	0.008	0.007	171	11,043	319	129	1/99
Sweden	Stockholm	0.016	0.021	0.010	0.006	0.006	0.017	0.015	0.037	0.017	0.012	0.168	0.134	0.011	0.021	526	30,020	187	92	1/96
Switzer.	SWX Swiss	0.015	0.015	0.006	0.009	0.005	0.014	0.004	0.028	0.012	0.009	0.173	0.119	0.010	0.016	311	21,154	65	33	8/96
Taiwan	Taiwan	0.007	0.007	0.002	0.004	0.002	0.009	0.007	0.022	0.007	0.006	0.110	0.107	0.009	0.006	752	68,365	451	251	1/96
Thailand	Thailand	0.016	0.024	0.009	0.007	0.006	0.019	0.014	0.046	0.025	0.018	0.225	0.183	0.014	0.014	561	33,960	192	71	1/96
Turkey	Istanbul	0.009	0.008	0.001	0.007	0.003	0.011	0.011	0.029	0.012	0.010	0.162	0.160	0.011	0.008	313	10,141	11	66	1/05
UK	London	0.016	0.024	0.014	0.003	0.001	0.008	0.002	0.036	0.020	0.013	0.192	0.178	0.007	0.029	2,187	77,972	882	413	1/96
US	NYSE	0.006	0.005	0.007	0.000	0.011	0.000	0.000	0.017	0.004	0.003	0.050	0.050	0.007	0.012	199	13,052	2,895	293	1/96
US	AMEX	0.034	0.031	0.030	0.004	0.022	0.001	0.005	0.049	0.020	0.015	0.147	0.118	0.018	0.050	74	3,868	58	4	1/96
US	NASDAQ	0.024	0.027	0.025	−0.001	0.027	0.001	0.001	0.045	0.015	0.011	0.100	0.084	0.020	0.026	638	37,015	4,103	555	1/96
Global 1996–2007		0.018	0.022	0.009	0.009	0.006	0.014	0.012	0.038	0.019	0.012	0.158	0.124	0.012	0.021	24,240	1.5E+06	17,744	7,998	
Global 2008–2009		0.019	0.024	0.009	0.010	0.006	0.019	0.013	0.037	0.018	0.012	0.131	0.098	0.015	0.024	1,380	24,708	2,757	467	
Global 2010–2014		0.015	0.017	0.006	0.008	0.006	0.015	0.010	0.028	0.013	0.009	0.127	0.104	0.012	0.017	1,707	60,081	11,897	1,344	

As a further data integrity check, Table II reports the median ratio of the sum of intraday share volume reported by TRTH divided by the share volume reported by Datastream per stock per day. We find that 91% of the exchange-year ratios in the primary sample (1996–2007) are exactly 100%. We find that 97% of these exchange-year ratios are in the range (95%, 102%).[18] The exchanges with the most prolonged deviation from this range are Milan (4 years), Vienna (3 years), and Bombay (3 years). All exchanges have median ratios of 100% in the secondary sample (2008–14). With full acknowledgment of the early deviations, we note that the vast majority of exchange-year volume ratios are close to or exactly equal to 100%.

Combining all the evidence above, we conclude that the TRTH intraday equity dataset is a high-quality, reliable dataset for global research. Our evidence does not imply anything about any other TRTH data (e.g., futures, options, commodities, foreign exchange, fixed income, etc.).

Table III describes the availability of closing bid and ask prices in Datastream, which is the information that is required to compute the Closing Percent Quoted Spread proxy. Each value represents the fraction of stocks in an exchange-year with more than 10 nonzero return days in the year where we observe closing bid and ask prices. We find that global average availability of closing bid and ask data in Datastream rises from 71.7% in 1996 to 94.9% in 2007 and stays steady at 95.9% in 2008–09 and 94.4% in 2010–14. Five exchanges have less than 70% availability in 1996 and the number declines to zero in 2007. Seventeen exchanges have less than 90% availability in 1996 and the number declines to five in 2007 and to four in 2010–14 (Austria, Greece, India, and the UK) India fell from 94% coverage in 2007 to 64% coverage in 2014. However, for the most part, the data inputs required to compute the Closing Percent Quoted Spread are widely available in Datastream.

4.3 Descriptive Statistics

Table IV provides the equally weighted mean of the monthly percent-cost benchmarks and proxies. Each row represents a different exchange. The last three rows are the global average of all forty-two exchanges for 1996–2007, 2008–09, and 2010–14, respectively. Of particular importance, the 1996–2007 global average of the Closing Percent Quoted Spread proxy is 0.021 (the last column of the proxies) that is relatively close to the 1996–2007 global average of the (intraday) percent quoted spread benchmark of 0.022 (second column of the benchmarks) and to the global average of the percent effective spread benchmark of 0.018 (first column of the benchmarks). The same is true for the global averages in 2008–09, and 2010–14.

Table V provides the equally weighted median of the monthly cost-per-dollar-volume benchmarks and proxies. Of particular importance, the 1996–2007 global median for each of the cost-per-dollar-volume proxies is an order of magnitude larger than the global

complete, official trading records of all trading in securities listed on the Helsinki Stock Exchange. The random checks we performed showed the trades agree so that if a trade of 200 shares at 10kr shows in the TRTH database, we will see a purchase of 200 shares at 10kr and a corresponding sale of 200 shares in the Depository data. We performed the random checks across all 12 years of our data and we believe that for this market the TRTH database exactly replicates trades reported in the central clearing agency.

18 There are several reasons why TRTH and Datastream may differ. First, the basis of volume quotation on TRTH can change from rounding to the nearest 1000 or 100, although it is mostly in one share. When there is rounding, there is rounding down errors. Some of the larger differences may be due to the fact that Datastream includes after hours trades, whereas our TRTH sample does not.

Table V. Median of the monthly cost-per-dollar-volume benchmarks and proxies

The cost-per-dollar-volume benchmark (slope of the price function lambda) is calculated from every trade and corresponding BBO quote in the SIRCA TRTH database for a sample stock-month. All cost-per-dollar-volume proxies are calculated from daily stock price and volume data for a sample stock-month. The primary sample spans forty-two exchanges around the world from 1996 to 2007. It consists of all stock-months with at least five positive-volume days and five non-zero return days. The secondary sample spans forty-two exchanges (thirty stocks per exchange) with the same activity filters from 2008 to 2014. The medians of all cost-per-dollar-volume benchmarks and proxies are multiplied by 1,000 except for the median of Amivest that is divided by 1,000,000. All variables are winsorized at 1st and 99th percentile in each exchange-year.

		Cost/$ Vol Bench	Cost-per-dollar-volume proxies													
Country	Exchange	Lambda	Roll Imp	Ext Roll Imp	Eff Tick Imp	LOT Mix Imp	LOT Y Imp	FHT Imp	Zeros Imp	Zeros2 Imp	High–Low Imp	Clos Quo Sprd Imp (%)	Amihud	Pastor and Stam-baugh	Amivest	Stock -Mon
Argentina	Bue. Ar.	0.075	0.145	0.601	1.414	2.385	1.564	1.017	14.834	7.969	0.597	1.430	2.125	0.000	0.214	3,965
Australia	Australian	0.064	0.723	1.822	2.560	4.474	2.804	1.908	12.960	8.065	1.372	2.277	7.340	0.000	0.576	84,495
Austria	Vienna	0.016	0.281	0.614	0.489	1.467	0.931	0.586	6.934	3.535	0.439	1.781	1.444	0.000	0.415	4,459
Belgium	Brussels	0.019	0.175	0.440	0.150	0.970	0.541	0.367	5.378	3.218	0.305	0.616	0.876	0.000	0.543	8,721
Brazil	Sao Paulo	0.201	0.274	0.852	0.672	2.135	1.315	0.863	8.131	3.251	0.634	1.980	5.681	0.000	17.472	2,893
Canada	Toronto	0.092	1.935	2.780	2.631	5.063	2.770	2.031	11.735	6.448	1.969	0.686	8.601	0.000	0.741	73,105
Chile	Santiago	0.058	0.021	0.117	0.068	0.541	0.377	0.218	3.832	2.567	0.102	0.228	0.825	0.000	172383	3,256
China	Hong K.	0.074	0.429	1.062	1.114	2.556	1.647	1.070	8.466	4.058	0.882	1.774	5.497	0.000	11.976	66,075
China	Shanghai	0.029	0.002	0.012	0.006	0.026	0.008	0.005	0.092	0.075	0.015	0.006	0.038	0.000	23.293	77,271

(continued)

Table V. Continued

		Cost/$ Vol Bench	Cost-per-dollar-volume proxies													
Country	Exchange	Lambda	Roll Imp	Ext Roll Imp	Eff Tick Imp	LOT Mix Imp	LOT Y Imp	FHT Imp	Zeros Imp	Zeros2 Imp	High–Low Imp	Clos Quo Sprd Imp (%)	Amihud	Pastor and Stam-baugh	Amivest	Stock -Mon
China	Shenzhen	0.034	0.002	0.012	0.006	0.018	0.005	0.005	0.089	0.072	0.014	0.006	0.037	0.000	20.692	57,488
Denmark	Copenhag.	0.047	0.184	0.373	0.280	0.824	0.489	0.334	4.544	2.240	0.264	0.602	0.953	0.000	13.783	12,157
France	Paris	0.029	0.325	0.735	0.366	1.715	0.919	0.635	8.533	4.600	0.822	1.601	1.785	0.000	1.419	41,098
Finland	Helsinki	0.003	0.310	0.664	0.627	1.359	0.827	0.546	6.664	3.493	0.419	0.814	2.088	0.000	0.619	10,209
Germany	Frankfurt	0.059	0.606	1.462	1.572	3.223	1.623	1.124	11.844	6.757	0.915	2.732	4.030	0.000	0.032	29,411
Greece	Athens	0.058	0.360	0.943	0.952	2.324	1.137	0.831	9.838	7.575	0.914	1.331	2.882	0.000	0.069	23,603
India	Bombay	5.814	129.3	272.1	137.0	592.3	348.9	230.0	936.4	216.2	220.8	504.3	874.8	−0.002	4.083	82,624
Indonesia	Jakarta	1.260	2.753	6.625	7.702	15.856	11.306	7.104	39.316	18.304	5.767	12.069	74.396	0.000	4.E+06	14,278
Ireland	Irish	0.027	0.252	0.427	0.379	1.042	0.799	0.377	2.720	1.343	0.309	0.531	2.358	0.000	0.801	2,708
Israel	Tel Aviv	0.208	0.472	1.427	0.230	4.343	2.995	1.564	17.067	3.616	1.420	2.430	6.128	0.000	1.497	22,686
Italy	Milan	0.011	0.113	0.211	0.112	0.379	0.206	0.146	2.239	1.278	0.140	0.249	0.439	0.000	1.639	23,976
Japan	Tokyo	0.278	0.127	0.317	0.281	0.670	0.364	0.249	3.113	1.951	0.169	0.327	0.529	0.000	7.E+03	2.7E+05
Malaysia	Kuala Lum.	0.197	1.412	2.999	2.884	6.411	4.007	2.628	28.306	16.350	1.907	5.221	10.092	0.000	0.301	79,565
Mexico	Mexico	0.085	0.108	0.333	0.488	0.912	0.652	0.415	4.042	1.501	0.313	0.949	7.732	0.000	75.505	5,042
Nether.	Amsterdam	0.032	0.228	0.509	0.570	1.199	0.684	0.459	4.392	2.743	0.387	0.788	1.608	0.000	3.195	11,620
New Zea.	New Zea.	0.040	0.169	0.463	0.987	1.461	0.947	0.634	10.087	7.312	0.367	0.788	1.640	0.000	0.354	5,052

(continued)

Table V. Continued

Country	Exchange	Cost/$ Vol Bench	Cost-per-dollar-volume proxies												
		Lambda	Roll Imp	Ext Roll Imp	Eff Tick Imp	LOT Mix Imp	LOT Y Imp	FHT Imp	Zeros Imp	Zeros2 Imp	High–Low Imp	Clos Quo Sprd Imp (%)	Amihud	Pastor and Stam-baugh	Amivest	Stock -Mon
Norway	Oslo	0.040	0.182	0.515	0.526	1.221	0.834	0.458	3.570	1.686	0.300	0.716	1.481	0.000	18.847	15,111
Philip.	Philippines	0.399	1.039	2.803	3.507	7.388	5.022	3.209	22.930	12.428	2.003	4.602	16.592	0.000	41.517	9,540
Poland	Warsaw	0.347	8.561	13.384	10.008	29.442	21.717	10.940	79.250	37.145	7.777	30.786	61.429	0.001	0.296	6,819
Portugal	Lisbon	0.013	0.223	0.284	0.232	0.545	0.290	0.203	4.986	3.514	0.204	0.407	0.863	0.000	0.953	987
Singapore	Singapore	0.033	0.734	1.360	1.984	2.927	1.909	1.287	11.535	7.086	0.989	1.735	4.833	0.000	0.274	32,313
S. Africa	Johannes.	0.135	1.468	3.074	3.648	7.917	5.295	3.341	15.943	9.836	2.328	4.696	28.413	0.000	14.938	28,049
S. Korea	Korea	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.001	0.000	0.000	0.001	0.000	3.E+02	76,246
Spain	Madrid	0.009	0.034	0.063	0.050	0.142	0.069	0.048	1.036	0.938	0.053	0.079	0.124	0.000	3.283	11,043
Sweden	Stockholm	0.058	0.486	1.314	1.104	2.721	1.630	1.089	8.224	3.901	0.770	2.448	2.864	0.000	30.383	30,020
Switzer.	SWX Swiss	0.028	0.133	0.289	0.112	0.577	0.321	0.220	2.775	1.423	0.211	0.383	0.664	0.000	1.388	21,154
Taiwan	Taiwan	0.003	0.001	0.004	0.004	0.012	0.005	0.004	0.064	0.046	0.004	0.012	0.022	0.000	13.400	68,365
Thailand	Thailand	0.820	1.794	4.007	2.362	9.373	5.921	3.908	29.641	13.043	3.349	2.494	35.324	0.000	72.238	33,960
Turkey	Istanbul	0.022	0.015	0.057	0.065	0.144	0.068	0.051	0.889	0.877	0.057	0.048	0.124	0.000	0.378	10,141
UK	London	0.036	0.000	0.001	0.001	0.006	0.004	0.003	0.029	0.024	0.001	0.006	0.005	0.000	146.156	77,972
US	NYSE	0.008	0.021	0.001	0.000	0.047	0.018	0.013	0.188	0.182	0.014	0.045	0.053	0.000	6.228	13,052
US	AMEX	0.194	2.437	0.075	1.138	4.768	2.375	1.726	9.512	6.008	1.978	6.650	9.808	0.047	0.036	3,868
US	NASDAQ	0.079	1.048	0.018	0.069	1.672	0.705	0.515	3.500	2.096	0.739	1.376	3.186	−0.005	0.992	37,015
Global 1996–2007		0.052	0.263	0.512	0.508	1.464	0.877	0.566	6.799	3.503	0.429	0.881	2.106	0.000	2.417	1.5E+06
Global 2008–2009		0.033	0.349	0.918	0.793	1.898	1.152	0.789	7.339	3.163	0.680	1.307	3.620	0.000	2.494	24,708
Global 2010–2014		0.043	0.306	0.698	0.501	1.518	0.952	0.624	5.857	3.283	0.502	0.920	2.784	0.000	2.906	60,081

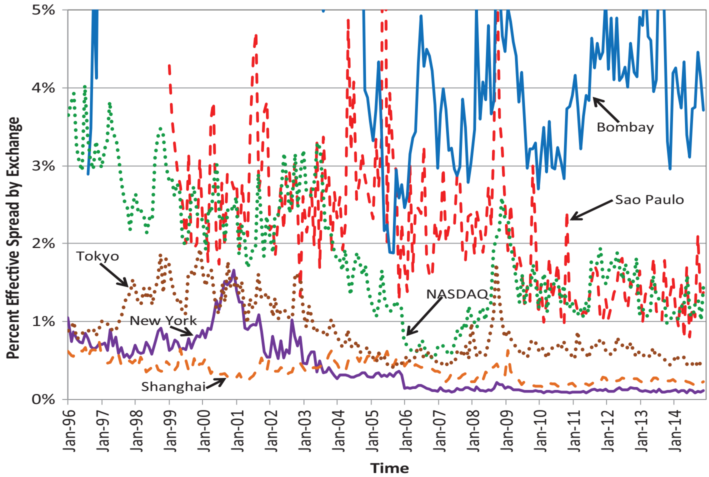

Figure 1. Percent effective spread by exchange over time.

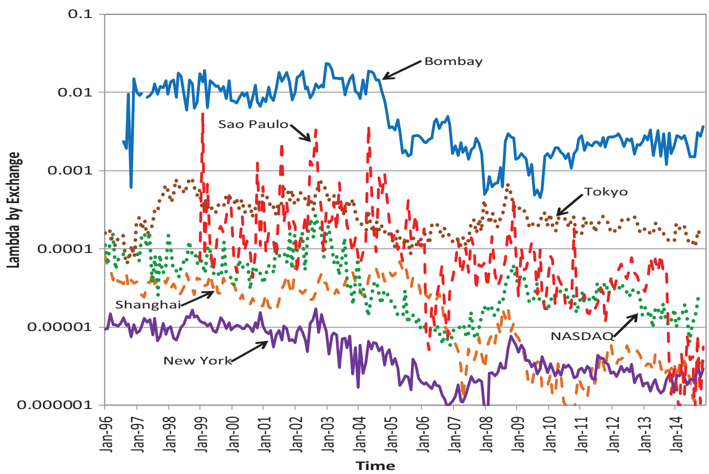

Figure 2. Lambda by exchange over time.

median of lambda at $0.052{*}10^{-3}$. The closest proxy is Roll Impact at 0.263×10^{-3}, which off by a factor of 5×. None of the cost-per-dollar-volume proxies are on the same scale as lambda. The same is true for the global medians in 2008–09, and 2010–14.

Figures 1 and 2 allow us to look at patterns in the data over the combined sample period from 1996 to 2014. Figure 1 presents the equally weighted mean of the monthly percent effective spread for six exchanges around the world from 1996 to 2014. In general, percent effective spreads have declined over time, but the pattern and timing is idiosyncratic to each exchange. Bombay hovered around 7% for a long time and then declined to around 4% during 2005. Sao Paulo fluctuated around 3% for a much of the sample and then declined to around 1.5% in 2009. NASDAQ declined by a one-third in 1997 and declined further from 2004 to 2007. Tokyo increased in 1997 and declined gradually from 2002 to 2005.

New York increased in 2000, declined sharply in 2001, and declined gradually since then. Perhaps the most surprising is Shanghai and Shenzhen (data not shown), which both have among the lowest percent effective spreads in the world over the entire sample period.

Figure 2 presents the median of the monthly lambda for six exchanges around the world from 1996 to 2014. The *y*-axis is on a log scale because the values of lambda by exchange differ by many orders of magnitude. Again the pattern and timing of lambda is idiosyncratic to each exchange. Bombay declined sharply in 2004 and 2005. Sao Paulo declined sharply in 2006 and in 2013. Both NASDAQ and New York declined gradually from 2003 to 2007. Tokyo increased in 1998 and declined from 2003 to 2007. Shanghai and Shenzhen (data not shown) both have among the lowest lambda in the world over the sample period and both declined in 2006 and 2009.

Finally, while the analysis in this article is based on data winsorized at the extreme 1% and using default value for three proxies when they cannot be computed, we conduct an assessment of the extent of outliers and assignment of default or missing values in the data. Supplementary Table 10 presents the global pooled percentile distribution of unwinsorized monthly benchmarks and proxies, and the frequency of missing value. The table shows that Extended Roll, LOT Mixed, and LOT Y-Split have maximum values that exceed the range of the benchmark Percent Effective Spread. Further, the default value is used for Roll, Extended Roll, and High–Low in 87, 35, and 3% of stock-months, respectively. Closing Percent Quoted Spread is not available for 25% of stock-months. For the cost-per-dollar-volume distribution, Roll Impact, Extended Roll Impact, High–Low Impact, and Closing Percent Quoted Spread Impact have similar percent default or missing value as their percent-cost counterparts.[19] Supplementary Table 11 shows the corresponding distribution at daily frequency. Both of the daily percent-cost proxies yield have a minimum value of zero and a maximum value that is less than the benchmark percent effective spread maximum value. High–Low and Closing Percent Quoted Spread are always available due to data filters but High–Low is based on the default value in 13% of stock-days. If we do not condition on the joint availability of these two daily proxies, High–Low would have 17% default value and Closing Percent Quoted Spread would have 26% missing value. Both High–Low Impact and Closing Percent Quoted Spread Impact have a maximum value less than the benchmark maximum. Both Amihud and Amivest have maximum values that are orders of magnitude larger than the benchmark maximum. Amivest has missing value in 54% of stock-days due to zero return. High–Low Impact is set to the default value of zero in 15% of stock-days. Closing Percent Quoted Spread Impact and Amihud are always available given the data filters.

5. Monthly Percent-Cost Results

Table VI provides a global overview. Panels A–D report the global performance of ten monthly percent-cost proxies compared to four monthly percent-cost benchmarks (percent effective spread, percent quoted spread, percent realized spread, and percent price impact). The four panels report four performance dimensions: average cross-sectional correlations,

19 The percent missing for the cost-per-dollar-volume proxies are slightly different than their percent-cost counterparts, because in 17% of stock-months the benchmark lambda has an insufficient number of trades to be computed. Unreported results confirm that the conclusion of this article is qualitatively unaffected by the treatment of missing value with respect to whether a default value is assigned and whether to include the observation in performing comparison.

Table VI. Global performance of liquidity proxies compared to liquidity benchmarks (1996–2007)

The percent-cost benchmarks (percent effective spread, percent quoted spread, percent realized spread, and percent price impact) and a cost-per-dollar-volume benchmark (slope of the price function lambda) are calculated from every trade and corresponding BBO quote in the SIRCA TRTH database for a sample stock-month. All percent-cost proxies and cost-per-dollar-volume proxies are calculated from daily stock price data for a sample stock-month. The primary sample spans forty-two exchanges around the world from 1996 to 2007. It consists of all stock-months with at least five positive-volume days and eleven nonzero return days. An open box means the highest correlation or the lowest average RMSE in the row. Shaded boxes mean correlations that are statistically indistinguishable from the highest correlation or average RMSEs that are statistically indistinguishable from the lowest average RMSE in the row at the 5% level. Bold-faced numbers are statistically different from zero or proxies have predictive power that is significant at the 5% level.

	Roll	Ext Roll	Eff Tick	LOT Mix	LOT Y	FHT	Zeros	Zeros2	High–Low	Closing Quo Sprd (%)
Panel A: Average cross-sectional correlation of monthly percent-cost proxies compared to percent-cost benchmarks										
Effective Spread (%)	**0.224**	**0.362**	**0.349**	**0.534**	**0.526**	**0.587**	**0.406**	**0.210**	**0.564**	**0.799**
Quoted Spread (%)	**0.233**	**0.381**	**0.363**	**0.573**	**0.576**	**0.642**	**0.458**	**0.209**	**0.594**	**0.915**
Realized Spread (%)	**0.189**	**0.253**	**0.254**	**0.362**	**0.371**	**0.407**	**0.317**	**0.176**	**0.384**	**0.589**
Price Impact (%)	**0.126**	**0.265**	**0.251**	**0.399**	**0.374**	**0.427**	**0.261**	**0.129**	**0.423**	**0.567**
Panel B: Portfolio time-series correlation of monthly detrended percent-cost proxies compared to detrended percent-cost benchmarks										
Effective Spread (%)	0.059	**0.454**	**0.404**	**0.426**	**0.451**	**0.531**	0.066	−0.081	**0.616**	**0.764**
Quoted Spread (%)	0.039	**0.451**	**0.430**	**0.463**	**0.529**	**0.607**	**0.142**	−0.063	**0.657**	**0.870**
Realized Spread (%)	0.063	**0.308**	**0.271**	**0.268**	**0.302**	**0.352**	0.097	0.002	**0.378**	**0.526**
Price Impact (%)	−0.021	**0.340**	**0.296**	**0.351**	**0.356**	**0.418**	0.002	−0.122	**0.495**	**0.572**
Panel C: individual stock time-series correlation of monthly detrended percent-cost proxies compared to detrended percent-cost benchmarks										
Effective Spread (%)	0.040	0.114	0.080	**0.144**	0.134	**0.157**	0.031	−0.028	**0.223**	**0.429**
Quoted Spread (%)	0.049	**0.140**	0.120	**0.186**	**0.183**	**0.217**	0.060	−0.046	**0.306**	**0.677**
Realized Spread (%)	0.047	0.058	0.049	0.045	0.060	0.070	0.051	0.015	0.093	**0.220**
Price Impact (%)	−0.008	0.060	0.040	0.117	0.088	0.105	−0.022	−0.047	**0.143**	**0.226**
Panel D: Average RMSE of monthly percent-cost proxies compared to percent-cost benchmarks										
Effective Spread (%)	0.0259	0.0228	0.0280	0.0402	0.0266	**0.0172**	0.1850	0.1461	**0.0164**	**0.0152**
Quoted Spread (%)	0.0307	0.0257	0.0309	0.0371	**0.0263**	**0.0201**	0.1800	0.1439	**0.0214**	**0.0096**
Realized Spread (%)	0.0216	0.0240	0.0271	0.0478	0.0302	0.0178	0.1935	0.1523	0.0157	0.0226
Price Impact (%)	0.0207	0.0205	0.0243	0.0479	0.0292	0.0158	0.1942	0.1526	0.0132	0.0227

(continued)

Table VI. Continued

	Roll Imp	Ext Roll Imp	Eff Tick Imp	LOT Mix Imp	LOT Y Imp	FHT Imp	Zeros Imp	Zeros2 Imp	High–Low Imp	Clos Quo Sprd Imp (%)	Amihud	Pastor and Stambaugh	Amivest
Panel E: Average cross-sectional correlation of monthly cost-per-dollar-volume proxies compared to lambda													
Lambda	**0.249**	**0.450**	**0.416**	**0.539**	**0.494**	**0.524**	**0.494**	**0.455**	**0.554**	**0.565**	**0.515**	**0.041**	**−0.208**
Panel F: Portfolio time-series correlation of monthly detrended cost-per-dollar-volume proxies compared to the detrended lambda													
Lambda	0.132	**0.387**	**0.347**	**0.426**	**0.398**	**0.416**	**0.296**	**0.275**	**0.410**	**0.427**	**0.418**	0.082	**−0.210**
Panel G: Individual stock time-series correlation of monthly detrended cost-per-dollar-volume proxies compared to the detrended lambda													
Lambda	0.049	**0.159**	**0.165**	**0.227**	**0.185**	**0.208**	**0.156**	0.126	**0.295**	**0.315**	**0.234**	0.013	−0.121
Panel H: Average RMSE of cost-per-dollar-volume proxies compared to lambda/median of lambda													
Ratio	381.6	474.7	326.0	882.8	547.6	347.6	1419.3	564.4	323.7	696.8	1642.2	14.6	4.E+06

portfolio time-series correlations, individual stock time-series correlation, and average RMSEs.

Panel A reports the average cross-sectional correlation for each monthly percent-cost proxy compared to the four monthly percent-cost benchmarks. The average cross-sectional correlations are computed in the spirit of Fama and MacBeth (1973) by: (i) calculating for each month the cross-sectional correlation across all firms and then (ii) calculating the average correlation value over all months. The convention that we will use throughout the rest of the article is to place an open box around the highest correlation in the row and a shaded box around any correlations that are statistically indistinguishable from the highest correlation in the row at the 5% level.[20] The idea is to identify the best proxy relative to a particular benchmark and the full "leadership group" that is statistical indistinguishable from the best proxy. For example, in the first row the proxy Closing Percent Quoted Spread has the highest average cross-sectional correlation with percent effective spread at 0.799 (open box) and there are no shaded boxes—so all of the rest of the correlations in the first row are significantly lower than 0.799. Bold-faced correlations are statistically different from zero at the 5% level.[21] All correlations in this panel are statistically different from zero.

Closing Percent Quoted Spread dominates all of the row comparisons for the four percent-cost benchmarks in Panel A. Closing Percent Quoted Spread has the highest correlation (open box) on all four rows and the Closing Percent Quoted Spread correlation is statistically higher than the correlation of any other proxy on all four rows. FHT has the second best correlations in all four rows and High–Low has the third best correlations in all four rows. This is evidence that Closing Percent Quoted Spread, FHT, and High–Low are the top three percent-cost proxies. This article is the first to test any of these top three proxies against the others.

Closing Percent Quoted Spread is the winner by a wide margin. It provides enormous performance gains over the proxies that global research has used to date (Zeros, LOT Mixed, etc.). For instance, results of Panel A imply that its mean cross-sectional correlation is 2.0 times the correlation of Zeros and 1.5 times the correlation of LOT Mixed. Interestingly, Closing Percent Quoted Spread has relatively higher correlations with percent effective spread (0.799) and percent quoted spread (0.915) and relatively lower correlations with percent realized spread (0.589) and percent price impact (0.567).

Figure 3 plots the global average of the cross-sectional correlations of six percent-cost proxies with percent effective spread over the combined sample period from 1996 to 2014. The global average of the cross-sectional correlation for Closing Percent Quoted Spread stays primarily in the range 0.70–0.90 over the entire sample period. It is typically 0.15–

20 In all tables with cross-sectional correlations, we test if the correlations are different between proxies on the same row by t-tests on the time series of correlations in the spirit of Fama–MacBeth. Specifically, we calculate the cross-sectional correlation of each proxy for each month and then regress the correlations of one proxy on the correlations of another proxy. We assume that the time series of correlations of each proxy is i.i.d. over time, and test if the regression intercept is zero and the slope is one. Standard errors are adjusted for autocorrelation with a Newey–West correction using four lags.

21 In all tables with correlations, we test if the correlations are statistically different from zero and highlight the correlations that are significant in boldface. For an estimated correlation σ, Swinscow (1997, chapter 11) gives the appropriate test statistic as $t = \sigma\sqrt{(D-2)/(1-\sigma^2)}$, where D is the sample size.

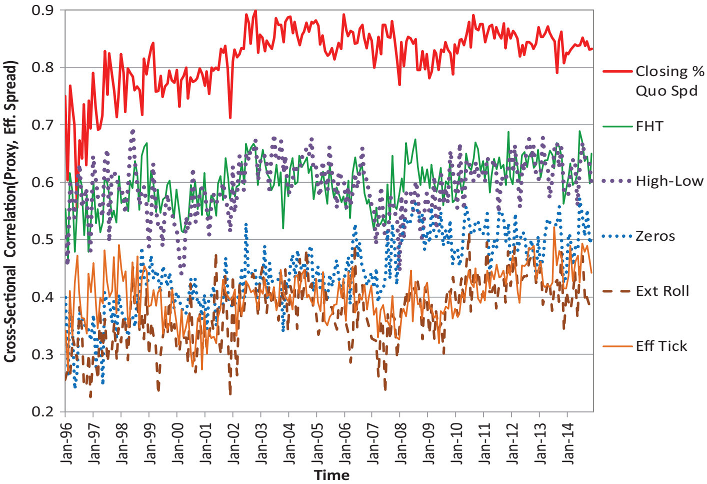

Figure 3. Global average of cross-sectional correlations (proxy, percent effective spread) over time.

0.20 above FHT and High–Low. It is typically 0.30–0.40 above Zeros. In other words, the large increase in performance occurs continuously throughout the sample period. Figure 3 also shows that the correlations of High–Low and FHT are the clear second choice and especially dominant over the measures commonly used in global research (e.g., Zeros, Roll).

Table VI Panel B is based on equally weighted portfolios across all stocks for month *i*. That is, we compute a portfolio percent-cost proxy (benchmark) in month *i* by taking the average of that percent-cost proxy (benchmark) over all stocks in month *i*. We remove the trend of each proxy and each benchmark by taking the first difference. Then, Panel B reports the time-series correlation between each detrended (i.e., first difference of) portfolio percent-cost proxy and the detrended portfolio percent-cost benchmarks. Closing Percent Quoted Spread dominates all of the row comparisons for the four percent-cost benchmarks. Closing Percent Quoted Spread has the highest correlation (open box) on all four rows and the Closing Percent Quoted Spread correlation is statistically higher than the correlation of any other proxy on three rows and higher than all but High–Low on the fourth row. We test whether time-series correlations are statistically different from each other using a Fisher's *Z*-test. As in Panel A, Closing Percent Quoted Spread, FHT, and High–Low are the top three percent-cost proxies on all four rows. Once again, Closing Percent Quoted Spread provides enormous performance gains over Zeros, LOT Mixed, etc. Again we find that Closing Percent Quoted Spread has relatively higher correlations with percent effective spread (0.764) and percent quoted spread (0.870) and relatively lower correlations with percent realized spread (0.526) and percent price impact (0.572). In unreported results, we get the same qualitative results when using the time-series with the trends left in.

Panel C reports the individual stock time-series correlations between the detrended individual stock percent-cost proxy and the detrended individual stock percent-cost benchmark. Closing Percent Quoted Spread has the highest correlation (open box) on all four rows and the Closing Percent Quoted Spread correlation is statistically higher than the correlation of any other proxy on all four rows. Again, Closing Percent Quoted Spread provides enormous performance gains over the other proxies, many of which are not significantly different from zero. It has relatively higher correlations with percent effective spread and percent quoted spread and relatively lower correlations with percent realized spread and percent price impact. For realized spread it is the only measure significantly different from zero.

Panel D reports the average RMSE between each percent-cost proxy and percent-cost benchmarks based on individual firms. The average RMSE tells us whether a particular proxy does a good job of capturing the level of a benchmark, not just whether it is correlated with the benchmark. The RMSE is calculated every month for a given exchange and then averaged over all sample months. In this case, an open box identifies the lowest average RMSE in the row and a shaded box indicates RMSEs that are statistically indistinguishable from the lowest average RMSE in the row. We test whether RMSEs are statistically different from each other using a paired *t*-test. Boldfaced RMSE indicates that the ability of the proxy to predict the benchmark is statistically greater than zero at the 5% level.[22]

Closing Percent Quoted Spread has the lowest average RMSE (open box) on the first two rows. It is statistically indistinguishable from High–Low relative to percent effective spread and significantly better than all other proxies relative to percent quoted spread. Again, Closing Percent Quoted Spread, FHT, and High–Low are the top three percent-cost proxies on both rows. As in panels A, B, and C, Closing Percent Quoted Spread provides enormous performance gains over Zeros, LOT Mixed, etc.

High–Low has the lowest average RMSE (open box) on the last two rows. It is significantly better than all other proxies relative to percent realized spread and percent price impact. Overall, Closing Percent Quoted Spread is closest to the level of percent effective spread and percent quoted spread, whereas High–Low is closest to the level of percent realized spread and percent price impact.

Figure 4 graphs the global average level of the top three percent-cost proxies (Closing Percent Quoted Spread, FHT, and High–Low) and four percent-cost benchmarks from 1996 to 2014. Closing Percent Quoted Spread is very close in both level and pattern to the Percent Quoted Spread Benchmark throughout the sample period. And both of them follow a relatively similar pattern to the Percent Effective Spread Benchmark, except that the level of the latter is approximately 0.5% lower. FHT follows the pattern of Percent Effective Spread well, except that the level is sometimes lower. The Percent Realized Spread Benchmark and the Percent Price Impact Benchmark, which by definition sum up to the Percent Effective Spread Benchmark, are typically nearly equal. Thus, their level is approximately half of the level of the Percent Effective Spread Benchmark. High–Low is typically much closer to the level of the Percent Realized Spread Benchmark and the Percent Price Impact Benchmark than to the level of Percent Effective Spread Benchmark.

To summarize Table VI Panels A–D, Closing Percent Quoted Spread strongly dominates all other monthly percent-cost proxies and provides enormous performance gains over Zeros, LOT Mixed, etc. It is highly correlated with all four percent-cost benchmarks—in the cross-section, portfolio time-series, and individual stock time-series. It does the best job of capturing the level of percent effective spread and percent quoted spread, whereas High–

22 We test whether RMSEs are statistically significant using the *U*-statistic developed by Theil (1966). Here, if $U^2 = 1$, then the proxy has zero ability to predict the benchmark (like a $R^2 = 0$). If $U^2 = 0$, then the proxy perfectly predicts the benchmark (like a $R^2 = 1$). We test if U^2 is significantly less than 1 based on an *F* distribution where the number of degrees of freedom for both the numerator and the denominator is the sample size.

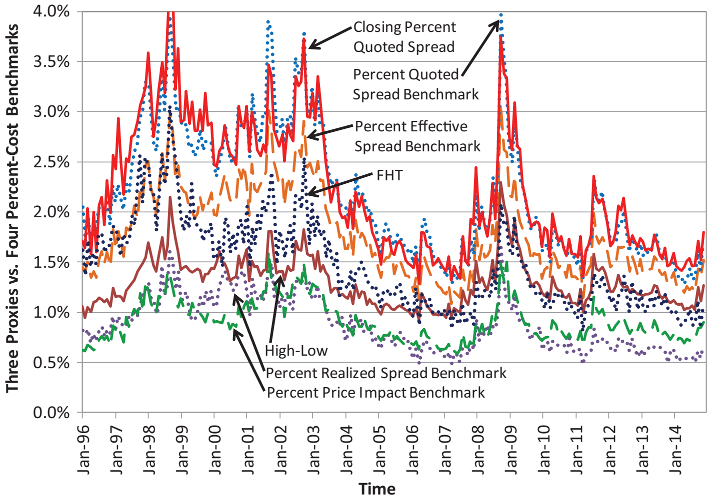

Figure 4. Global average of three percent-cost proxies and four percent-cost benchmarks over time.

Low does the best job of capturing the level of percent realized spread and percent price impact.

6. Monthly Cost-per-Dollar-Volume Results

The global overview continues with Panels E–H, which report the global performance of thirteen monthly cost-per-dollar-volume proxies compared to the single monthly cost-per-dollar-volume benchmark (lambda). Panel E reports the average cross-sectional correlation for each monthly cost-per-dollar-volume proxy compared to monthly lambda. Closing Percent Quoted Spread Impact has the highest correlation (0.565) and that is statistically higher than the correlation of any other proxy. In terms of economic magnitude, five proxies have correlations of 0.5 or higher: Closing Percent Quoted Spread Impact, FHT Impact, High–Low Impact, LOT Mixed Impact, and Amihud.

Figure 5 plots the global average of the cross-sectional correlations of these five cost-per-dollar-volume proxies with lambda from 1996 to 2014. The global average of the cross-sectional correlations of all five proxies (Closing Percent Quoted Spread Impact, FHT Impact, High–Low Impact, LOT Mixed Impact, and Amihud) are nearly identical over the entire sample. The correlations are typically in the 0.45–0.70 range over the entire sample period. In other words, the economic performance of these five proxies is nearly the same throughout the sample period.

Table VI Panel F reports the time-series correlation between each portfolio of detrended (i.e., first differences of) cost-per-dollar-volume proxy and the portfolio detrended lambda. Closing Percent Quoted Spread has the highest correlation (0.427) and the five proxies mentioned above have similar economic magnitudes with correlations of 0.4 or higher.

Panel G reports the individual stock time-series correlation between each detrended cost-per-dollar-volume proxy and detrended lambda. Closing Percent Quoted Spread Impact has the highest correlation (0.315) and the five proxies mentioned above have similar economic magnitudes with correlations of 0.2 or above.

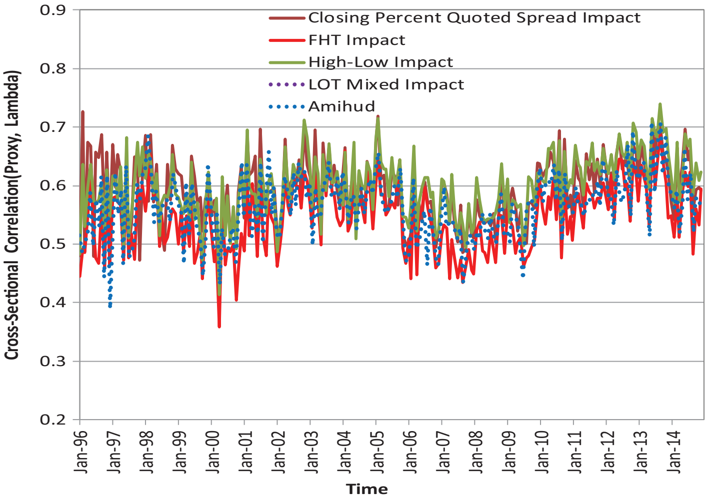

Figure 5. Global average of cross-sectional correlations (proxy, lambda) over time figure.

Panel H reports the ratio of the average RMSE between each cost-per-dollar-volume proxy and lambda divided by the median of lambda. The lowest ratio is Pastor and Stambaugh at 14.6. The rest of the cost-per-dollar-volume proxies have a ratio of 323 or greater. In other words, the average error is an order of magnitude larger than the mean of lambda itself. Thus, we conclude that none of the cost-per-dollar-volume proxies capture the level of lambda.

Figure 6 graphs the global average level of five cost-per-dollar-volume proxies and lambda from 1996-2014. The *y*-axis is on a log scale because the values of the cost-per-dollar-volume proxies and lambda differ by many orders of magnitude. It is visually clear that all five proxies are correlated with lambda. However, considering the log scale of the *y*-axis, it is immediately clear that none of the proxies is on the same order of magnitude as lambda. In other words, there is more than a 10× difference in level between the proxies and lambda nearly all the time throughout the sample period.

To summarize Table VI Panels E–H we find that five monthly proxies (Closing Percent Quoted Spread Impact, FHT Impact, High–Low Impact, LOT Mixed Impact, and Amihud) do nearly as well economically in all three Panels E–G.[23] All five are highly correlated with monthly lambda, but none captures its level.

7. Monthly Proxy Robustness Checks

7.1 By Time Period

Next we examine the robustness of our results by time period. Table VII reports the global performance of liquidity proxies compared to liquidity benchmarks for three time periods: 1996–2007 (primary sample), 2008–09 (financial crisis period), and 2010–14 (post-financial crisis).

23 In an unreported test using an additional criteria of across sample, across treatment, and across stock filters robustness we find Closing Percent Quoted Spread Impact, Amihud, and High–Low Impact perform better among the top five monthly lambda proxies.

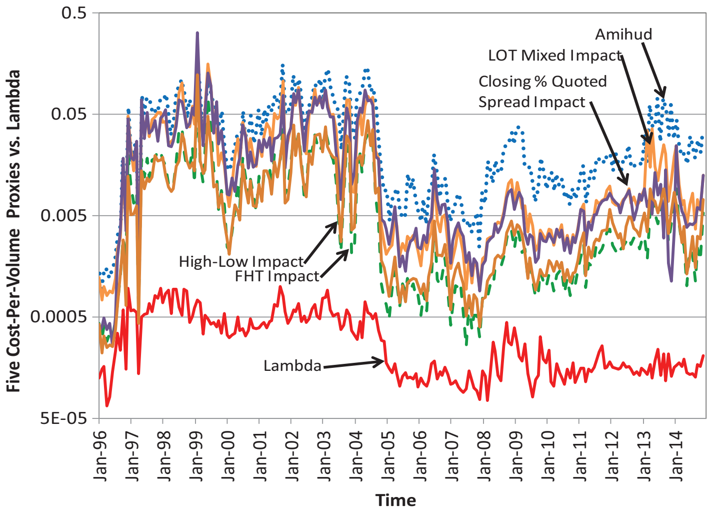

Figure 6. Global average of five cost-per-dollar-volume proxies and lambda over time.

Panels A–D report the performance of monthly percent-cost proxies compared to the percent-cost benchmark percent effective spread. On all four dimensions of performance (average cross-sectional correlations, portfolio time-series correlations, individual stock time-series correlation, and average RMSE), Closing Percent Quoted Spread has the highest correlation (open box) for all three time periods or the lowest average RMSE (open box) for all three time periods. In ten of the twelve rows in Panels A–D, the Closing Percent Quoted Spread correlation (average RMSE) is statistically higher (lower) than the correlation (average RMSE) of any other proxy. In all twelve rows of Panels A–D, Closing Percent Quoted Spread provides large performance gains over any other proxy.

Panels E–H report the performance of monthly cost-per-dollar-volume proxies compared to the cost-per-dollar-volume benchmark lambda. Five cost-per-dollar-volume proxies (Closing Percent Quoted Spread Impact, FHT Impact, High–Low Impact, LOT Mixed Impact, and Amihud) have average cross-sectional correlations of 0.48 or higher in all three time periods, detrended portfolio time-series correlations of 0.36 or higher in all three time periods, and detrended individual stock time-series correlations of 0.15 or higher in all three time periods. All of the cost-per-dollar-volume proxies have a ratio of average RMSE over the mean of lambda of 7 or greater. Thus, all five cost-per-dollar-volume proxies are highly correlated with monthly lambda in all three periods, but none captures its level.

In summary, our monthly proxy results are robust by time period.

7.2 Developed Versus Emerging Countries

Next we examine the robustness of our results in developed countries versus emerging countries. We designate the developing countries as Australia, Austria, Belgium, Canada, Denmark, France, Finland, Germany, Ireland, Italy, Japan, the Netherlands, New Zealand, Norway, Portugal, Spain, Sweden, Switzerland, the UK, and the USA. All other countries in our sample are designated as emerging. Table VIII Panels A–D report monthly percent-cost proxies compared to monthly percent effective spread and Panels E–G report monthly cost-per-dollar-volume proxies compared to monthly lambda.

Table VII. Global performance of liquidity proxies compared to liquidity benchmarks by time period

The percent-cost benchmark (percent effective spread) and a cost-per-dollar-volume benchmark (slope of the price function lambda) are calculated from every trade and corresponding BBO quote in the SIRCA TRTH database for a sample stock-month. All percent-cost proxies and cost-per-dollar-volume proxies are calculated from daily stock price data for a sample stock-month. The primary sample spans forty-two exchanges around the world from 1996 to 2007. It consists of all stock-months with at least five positive-volume days and eleven nonzero return days. The secondary sample spans forty-two exchanges (thirty stocks per exchange) with the same activity filters from 2008 to 2014. An open box means the highest correlation or the lowest average RMSE in the row. Shaded boxes mean correlations that are statistically indistinguishable from the highest correlation or average RMSEs that are statistically indistinguishable from the lowest average RMSE in the row at the 5% level. Bold-faced numbers are statistically different from zero or proxies have predictive power that is significant at the 5% level.

	Roll	Ext Roll	Eff Tick	LOT Mix	LOT Y	FHT	Zeros	Zeros2	High–Low	Closing Quo Sprd (%)
Panel A: Average cross-sectional correlation of monthly percent-cost proxies compared to percent effective spread										
1996–2007	**0.224**	**0.362**	**0.349**	**0.534**	**0.526**	**0.587**	**0.406**	**0.210**	**0.564**	**0.799**
2008–2009	**0.214**	**0.383**	**0.380**	**0.594**	**0.596**	**0.619**	**0.521**	**0.304**	**0.565**	**0.819**
2010–2014	**0.258**	**0.420**	**0.432**	**0.610**	**0.611**	**0.628**	**0.505**	**0.360**	**0.624**	**0.850**
Panel B: Portfolio time-series correlation of monthly detrended percent-cost proxies compared to detrended percent effective spread										
1996–2007	0.059	**0.454**	**0.404**	**0.426**	**0.451**	**0.531**	0.066	−0.081	**0.616**	**0.764**
2008–2009	−0.021	**0.388**	0.268	**0.487**	**0.525**	**0.542**	0.088	−0.078	**0.609**	**0.718**
2010–2014	0.086	**0.350**	**0.256**	**0.403**	**0.444**	**0.466**	0.148	0.012	**0.578**	**0.727**
Panel C: Individual stock time-series correlation of monthly detrended percent-cost proxies compared to detrended percent effective spread										
1996–2007	0.040	0.114	0.080	**0.144**	0.134	**0.157**	0.031	−0.028	**0.223**	**0.429**
2008–2009	−0.002	0.101	0.114	0.132	0.119	0.125	0.020	−0.017	0.266	**0.432**
2010–2014	0.034	0.111	0.085	0.111	0.097	0.105	0.006	−0.039	**0.246**	**0.417**
Panel D: Average RMSE of monthly percent-cost proxies compared to percent effective spread										
1996–2007	0.0259	0.0228	0.0280	0.0402	0.0266	**0.0172**	0.1850	0.1461	**0.0164**	**0.0152**
2008–2009	0.0270	0.0227	0.0281	0.0339	0.0220	**0.0178**	0.1542	0.1169	**0.0160**	**0.0148**
2010–2014	0.0210	0.0174	0.0219	0.0275	0.0164	**0.0145**	0.1551	0.1275	**0.0121**	**0.0109**

(continued)

Table VII. Continued

	Roll Imp	Ext Roll Imp	Eff Tick Imp	LOT Mix Imp	LOT Y Imp	FHT Imp	Zeros Imp	Zeros2 Imp	High–Low Imp	Clos Quo Sprd Imp (%)	Amihud	Pastor and Stambaugh	Amivest
Panel E: Average cross-sectional correlation of monthly cost-per-dollar-volume proxies compared to lambda													
1996–2007	0.249	0.450	0.416	0.539	0.494	0.524	0.494	0.455	0.554	0.565	0.515	0.041	−0.208
2008–2009	0.183	0.439	0.422	0.512	0.477	0.486	0.464	0.413	0.542	0.532	0.509	0.058	−0.237
2010–2014	0.254	0.532	0.510	0.612	0.568	0.570	0.538	0.511	0.638	0.624	0.592	0.034	−0.275
Panel F: Portfolio time-series correlation of monthly detrended cost-per-dollar-volume proxies compared to detrended lambda													
1996–2007	0.132	0.387	0.347	0.426	0.398	0.416	0.296	0.275	0.410	0.427	0.418	0.082	−0.210
2008–2009	0.023	0.302	0.305	0.367	0.347	0.381	0.267	0.167	0.393	0.424	0.425	0.033	−0.201
2010–2014	0.136	0.354	0.294	0.372	0.359	0.389	0.315	0.241	0.454	0.416	0.400	0.072	−0.113
Panel G: Individual stock time-series correlation of monthly detrended cost-per-dollar-volume proxies compared to detrended lambda													
1996–2007	0.049	0.159	0.165	0.227	0.185	0.208	0.156	0.126	0.295	0.315	0.234	0.013	−0.121
2008–2009	−0.009	0.150	0.186	0.193	0.148	0.158	0.115	0.087	0.311	0.304	0.289	0.017	−0.154
2010–2014	0.041	0.182	0.180	0.200	0.151	0.170	0.112	0.103	0.310	0.308	0.256	0.015	−0.157
Panel H: Average RMSE of cost-per-dollar-volume proxies compared to lambda/median of lambda													
1996–2007	381.6	474.7	326.0	882.8	547.6	347.6	1419.3	564.4	323.7	696.8	1642.2	14.6	4.E+06
2008–2009	108.3	179.9	134.0	350.0	201.5	163.4	1087.2	403.5	137.5	288.6	1384.2	16.3	2.E+07
2010–2014	83.0	154.0	236.5	494.7	261.4	212.7	1420.7	629.0	185.7	299.4	1292.8	7.3	2.E+07

Table VIII. Liquidity proxy performance in developed countries versus emerging countries (1996–2007)

A percent-cost benchmark (percent effective spread) and a cost-per-dollar-volume benchmark (slope of the price function lambda) are calculated from every trade and corresponding BBO quote in the SIRCA TRTH database for a sample stock-month. All percent-cost proxies and cost-per-dollar-volume proxies are calculated from daily stock price data for a sample stock-month. The primary sample spans forty-two exchanges around the world from 1996 to 2007. It consists of all stock-months with at least five positive-volume days and eleven nonzero return days. An open box means the highest correlation or the lowest average RMSE in the row. Shaded boxes mean correlations that are statistically indistinguishable from the highest correlation or average RMSEs that are statistically indistinguishable from the lowest average RMSE in the row at the 5% level. Bold-faced numbers are statistically different from zero or proxies have predictive power that is significant at the 5% level.

	Roll	Ext Roll	Eff Tick	LOT Mix	LOT Y	FHT	Zeros	Zeros2	High–Low	Closing Quo Sprd (%)
Panel A: Average cross-sectional correlation of monthly percent-cost proxies when compared to percent effective spread										
Developed	**0.259**	**0.399**	**0.414**	**0.573**	**0.561**	**0.634**	**0.437**	**0.269**	**0.599**	$\boxed{\textbf{0.820}}$
Emerging	**0.185**	**0.321**	**0.278**	**0.490**	**0.488**	**0.535**	**0.373**	**0.146**	**0.526**	$\boxed{\textbf{0.777}}$
Panel B: Portfolio time-series correlation of monthly detrended percent-cost proxies when compared to detrended percent effective spread										
Developed	0.078	**0.499**	**0.429**	**0.397**	**0.446**	**0.524**	0.013	−0.117	**0.630**	$\boxed{\textbf{0.744}}$
Emerging	0.038	**0.404**	**0.376**	**0.457**	**0.456**	**0.537**	**0.124**	−0.042	**0.601**	$\boxed{\textbf{0.785}}$
Panel C: Individual stock time-series correlation of monthly detrended percent-cost proxies when compared to detrended percent effective spread										
Developed	0.038	0.114	0.086	0.134	0.112	0.135	−0.007	−0.042	**0.215**	$\boxed{\textbf{0.382}}$
Emerging	0.042	0.113	0.074	**0.155**	**0.158**	**0.181**	0.072	−0.012	**0.230**	$\boxed{\textbf{0.480}}$
Panel D: Average RMSE of monthly percent-cost proxies when compared to percent effective spread										
Developed	0.0227	0.0203	0.0244	0.0368	0.0244	**0.0149**	0.1794	0.1447	**0.0140**	$\boxed{\textbf{0.0129}}$
Emerging	0.0295	0.0255	0.0319	0.0439	0.0291	**0.0198**	0.1912	0.1476	**0.0191**	$\boxed{\textbf{0.0177}}$

(continued)

Table VIII. Continued

	Roll Imp	Ext Roll Imp	Eff Tick Imp	LOT Mix Imp	LOT Y Imp	FHT Imp	Zeros Imp	Zeros2 Imp	High–Low Imp	Clos Quo Sprd Imp (%)	Amihud	Pastor and Stambaugh	Amivest
Panel E: Average cross-sectional correlation of monthly cost-per-dollar-volume proxies when compared to lambda													
Developed	**0.287**	**0.456**	**0.430**	**0.552**	**0.508**	**0.533**	**0.491**	**0.467**	**0.546**	**0.567**	**0.528**	**0.050**	**−0.191**
Emerging	**0.207**	**0.443**	**0.401**	**0.526**	**0.479**	**0.514**	**0.496**	**0.441**	**0.562**	**0.564**	**0.501**	**0.032**	**−0.227**
Panel F: Portfolio time-series correlation of monthly detrended cost-per-dollar-volume proxies compared to detrended lambda													
Developed	**0.146**	**0.364**	**0.339**	**0.398**	**0.370**	**0.392**	**0.267**	**0.268**	**0.376**	**0.409**	**0.380**	**0.104**	**−0.170**
Emerging	**0.117**	**0.412**	**0.355**	**0.456**	**0.429**	**0.442**	**0.328**	**0.282**	**0.447**	**0.448**	**0.460**	**0.058**	**−0.253**
Panel G: Individual stock time-series correlation of monthly detrended cost-per-dollar-volume proxies compared to detrended lambda													
Developed	0.047	0.128	**0.147**	**0.191**	**0.148**	**0.165**	0.105	0.086	**0.252**	**0.269**	**0.204**	0.012	−0.109
Emerging	0.052	**0.193**	**0.184**	**0.267**	**0.225**	**0.255**	**0.212**	**0.170**	**0.341**	**0.367**	**0.268**	0.014	−0.135
Panel H: Average RMSE of cost-per-dollar-volume proxies compared to lambda/median of lambda													
Developed	86.5	105.5	115.5	240.6	181.1	96.1	602.0	304.3	79.4	261.3	724.7	8.6	6.4E+06
Emerging	74.7	93.6	59.3	168.3	100.7	66.2	246.0	90.3	62.7	124.3	279.5	2.2	3.8E+04

Panels A–C report that Closing Percent Quoted Spread has the highest correlation in all six rows and is significantly higher than the correlation of any other proxy in one developed country row and in all three emerging country rows. Panel D reports that Closing Percent Quoted Spread has the lowest average RMSE, is significantly lower than any other proxy in developed countries except for FHT and High–Low, and is significantly lower than any other proxy in emerging countries.

Panels E–H report the performance of monthly cost-per-dollar-volume proxies compared to the cost-per-dollar-volume benchmark lambda. Five cost-per-dollar-volume proxies (Closing Percent Quoted Spread Impact, FHT Impact, High–Low Impact, LOT Mixed Impact, and Amihud) have average cross-sectional correlations of 0.5 or higher in both developed and emerging countries, detrended portfolio time-series correlations of 0.37 or higher in both developed and emerging countries, and detrended individual stock time-series correlations of 0.16 or higher in both developed and emerging countries. All of the cost-per-dollar-volume proxies have a ratio of average RMSE over the median of lambda of 2 or greater. Thus, all five cost-per-dollar-volume proxies are highly correlated with monthly lambda in both types of countries, but none captures its level.

In summary, our monthly proxy results are robust in both developed and emerging countries.

8. Daily Percent-Cost Results

Table IX provides an overview of daily liquidity proxies. The only proxies that change daily are High–low and Closing Quoted spread. Panels A–C compares the two daily percent-cost proxies with daily percent-cost benchmarks. Panel A compares both daily proxies to four percent-cost benchmarks on a global basis, Panel B compares both daily proxies with daily percent effective spread in developed and emerging countries, and Panel C compares both daily proxies by time period. We find essentially the same pattern as the monthly results. Daily Closing Percent Quoted Spread strongly dominates daily High–Low. Its correlations with all four daily percent-cost benchmarks are surprisingly high (i.e., they are only modestly diminished compared to the analogous monthly proxy correlations). From the average RMSE columns, we see that it does the best job of capturing the level of daily percent effective spread and daily percent quoted spread, whereas daily High–Low does the best job of capturing the level of daily percent realized spread and daily percent price impact.

In summary, our daily percent-cost results follow the same pattern as the monthly results. That is, daily Closing Percent Quoted Spread is strongly dominant as the best daily proxy. Indeed, its correlations that only modestly diminished compared to the analogous monthly proxy correlations.

9. Daily Cost-per-Dollar-Volume Results

Table IX Panels D–F analyze four daily cost-per-dollar-volume proxies relative to daily lambda. Daily Amihud wins the majority of contests. The average cross-sectional correlations remain strong with daily Amihud turning in a correlation of 0.460. However, the detrended portfolio time-series correlations and detrended individual time-series correlations drop to a poor level of only 0.038 and 0.142, respectively. In summary, daily Amihud is strongly correlated with daily lambda in the cross-section, but not in the time-series and does not capture its level.

Table IX. The performance of daily liquidity proxies compared to daily liquidity benchmarks on a global, developed, emerging, and time period basis

The percent-cost benchmarks (percent effective spread, percent quoted spread, percent realized spread, and percent price impact) and a cost-per-dollar-volume benchmark (slope of the price function lambda) are calculated from every trade and corresponding BBO quote in the SIRCA TRTH database for a sample stock-day. All percent-cost proxies and cost-per-dollar-volume proxies are calculated from daily stock data for a sample stock-day. The primary sample spans forty-two exchanges around the world from 1996 to 2007. The secondary sample spans forty-two exchanges (thirty stocks per exchange) from 2008 to 2014. An open box means the highest correlation or the lowest average RMSE among the compared proxies. Shaded boxes mean correlations that are statistically indistinguishable from the highest correlation or average RMSEs that are statistically indistinguishable from the lowest average RMSE among the compared proxies at the 1% level. Bold-faced numbers are statistically different from zero or have predictive power that is significant at the 1% level.

	Average cross-sectional corr.		Detrended portfolio time-series corr.		Detr. individual stock time-series corr.		Average RMSE	
	High–Low	Closing Quo Spr (%)	High–Low	Closing Quo Spr (%)	High–Low	Closing Quo Spr (%)	High–Low	Closing Quo Spr (%)
Panel A: Daily percent-cost proxies compared to four daily percent-cost benchmarks on a global basis								
Effective Spread (%)	**0.305**	**0.686**	−0.013	**0.379**	−0.006	**0.222**	0.0181	**0.0160**
Quoted Spread (%)	**0.335**	**0.728**	0.089	**0.445**	0.024	**0.233**	0.0262	**0.0181**
Realized Spread (%)	**0.234**	**0.495**	0.017	**0.218**	0.023	0.108	0.0179	0.0211
Price Impact (%)	**0.130**	**0.343**	−0.055	**0.229**	−0.035	0.094	0.0164	0.0237
Panel B: Daily percent-cost proxies compared to daily percent effective spread for developed versus emerging countries								
Developed	**0.334**	**0.710**	**0.037**	**0.318**	−0.010	**0.203**	0.0149	**0.0131**
Emerging	**0.273**	**0.659**	−0.069	**0.446**	−0.003	**0.242**	0.0216	**0.0192**
Panel C: Daily percent-cost proxies compared to daily percent effective spread by time period								
2008–2009	**0.274**	**0.697**	**0.471**	**0.700**	−0.007	0.198	0.0218	**0.0178**
2010–2014	**0.347**	**0.730**	**0.387**	**0.629**	−0.003	0.203	0.0163	**0.0134**

(continued)

Table IX. Continued

	Average cross-sectional correlation				Detrended portfolio time-series corr.				Detr. individual stock time-series corr.				Average RMSE/median lambda			
	High–Low Imp	Clos Quo Spr Imp (%)	Amihud	Amivest	High–Low Imp	Clos Quo Spr Imp (%)	Amihud	Amivest	High–Low Imp	Clos Quo Spr Imp (%)	Amihud	Amivest	High–Low Imp	Clos Quo Spr Imp (%)	Amihud	Amivest
Panel D: Daily cost-per-dollar-volume proxies compared to daily lambda on a global basis																
Lambda	**0.372**	**0.446**	**0.460**	**−0.275**	−0.051	0.004	0.038	0.003	0.078	0.133	0.142	-0.054	5.45	5.43	32.28	5.6E+15
Panel E: Daily cost-per-dollar-volume proxies compared to daily lambda for developed versus emerging countries																
Developed	**0.379**	**0.456**	**0.478**	**−0.283**	−0.067	−0.026	0.032	−0.003	0.060	0.108	0.133	−0.055	1.37	1.37	6.15	3.1E+13
Emerging	**0.364**	**0.439**	**0.437**	**−0.231**	−0.034	0.036	0.045	0.010	0.098	0.161	0.152	−0.054	9.93	9.91	61.01	1.2E+16
Panel F: Daily cost-per-dollar-volume proxies compared to daily lambda by time period																
2008–2009	**0.371**	**0.428**	**0.422**	**−0.320**	−0.017	−0.005	0.015	−0.014	0.088	0.126	0.104	−0.062	6.77	6.75	62.4	2.4E+16
2010–2014	**0.461**	**0.523**	**0.521**	**−0.386**	−0.039	−0.023	−0.019	−0.005	0.079	0.119	0.115	−0.062	5.21	5.21	27.3	2.1E+16

10. Conclusion

We run horseraces of monthly and daily liquidity proxies constructed from low-frequency stock data against liquidity benchmarks computed from high-frequency data on forty-two exchanges over 19 years. We find that the best liquidity proxies for global research are:

1. Closing Percent Quoted Spread is the best monthly percent-cost proxy. If Closing Percent Quoted Spread is not sufficiently available for a given research purpose, we find that the High–Low and FHT proxies are the next best monthly percent-cost proxies.
2. Amihud, Closing Percent Quoted Spread Impact, LOT Mixed Impact, High–Low Impact, and FHT Impact are tied as the best monthly cost-per-dollar-volume proxy.
3. The daily version of Closing Percent Quoted Spread is the best daily percent-cost proxy.
4. The daily version of Amihud is the best daily cost-per-dollar-volume proxy.

Supplementary Material

Supplementary data are available at *Review of Finance* online.

Appendix: Existing Low-Frequency Proxies

$$\text{Roll} = \begin{cases} 2\sqrt{-\text{Cov}(\Delta P_t, \Delta P_{t-1})}/\bar{P} & \text{when Cov}(\Delta P_t, \Delta P_{t-1}) < 0 \\ 0 & \text{when Cov}(\Delta P_t, \Delta P_{t-1}) \geq 0, \end{cases}$$

where $\bar{P}$ is the average price in a given stock-month.

$$\text{Extended Roll} = \begin{cases} 2\sqrt{-\text{Cov}(\Delta P_t^*, \Delta P_{t+1}^*)}/\bar{P} & \text{when Cov}(\Delta P_t^*, \Delta P_{t+1}^*) < 0 \\ 0 & \text{when Cov}(\Delta P_t^*, \Delta P_{t+1}^*) > 0, \end{cases}$$

where $\Delta P_t^* = z_t \cdot P_{t-1}$and z_t is the residual from $ar_t - r_f = \alpha + \beta(r_{mt} - r_f) + z_t$.

Effective Tick $= \frac{\sum_{j=1}^{J} \hat{\gamma}_j s_j}{\bar{P}_i}$ on a \$1/8th price grid is:

$$F_j = \frac{N_j}{\sum_{j=1}^{J} N_j} \text{ for } j = 1, 2, \ldots, J;\ U_j = \begin{cases} 2F_j & j = 1 \\ 2F_j - F_{j-1} & j = 2, 3, \ldots, J-1; \\ F_j - F_{j-1} & j = J \end{cases}$$

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

where F_j is the probability of trades on prices corresponding to the j-thspread, U_j be the unconstrained probability of the j-th spread, $\hat{\gamma}_j$ be the constrained probability of the j-th spread, and s_j is the j-th spread. The decimal price grid formula is in Appendix A of Holden (2009). Detailed examples are at www.kelley.iu.edu/cholden/examples.pdf.

$LOT\ Mixed = \alpha_2 - \alpha_1$, where $\alpha_2(\alpha_1)$ is the trans cost to buy (sell) and is estimated using :

$$\underset{\alpha_1,\alpha_2,\beta,\sigma}{\text{Max}} \begin{cases} \prod_1 \frac{1}{\sigma} n\left[\frac{R_t + \alpha_1 - \beta R_{mt}}{\sigma}\right] \\ \times \prod_0 \left[N\left(\frac{\alpha_2 - \beta R_{mt}}{\sigma}\right) - N\left(\frac{\alpha_1 - \beta R_{mt}}{\sigma}\right)\right] \\ \times \prod_2 \frac{1}{\sigma} n\left[\frac{R_t + \alpha_2 - \beta R_{mt}}{\sigma}\right] \end{cases}$$

where $R_t(R_{mt})$ is the own return (market return), σ is the return volatility, and β is the stock's market sensitivity, $S.T.\ \ \alpha_1 \leq 0, \alpha_2 \geq 0, \beta \geq 0$, and $\sigma \geq 0$. *LOT Mixed* is capped at a max value of 1.5.

Region 0 is $R_{jt} = 0$, region 1 is $R_{jt} \neq 0$ and $R_{mt} > 0$, and region 2 is $R_{jt} \neq 0$ and $R_{mt} < 0$.

$LOT\ Y - split = \alpha_2 - \alpha_1$ where everything is the same as LOT Mixed, except that region 0 is $R_{jt} = 0$, region 1 is $R_{jt} > 0$, and region 2 is $R_{jt} < 0$ and no upper bound cap is imposed.

Zeros$=\frac{\text{ZRD}}{\text{TD+NTD}}$, where ZRD = the number of zero returns days, TD = number of trading days, and NTD = number of no-trade days in a given stock-month.

High − Low=Average$\left(\frac{2(e^{\alpha_t}-1)}{1+e^{\alpha_t}}\right)$; where $\alpha_t = \frac{\sqrt{2\beta_t}-\sqrt{\beta_t}}{3-2\sqrt{2}} - \sqrt{\frac{\gamma_t}{3-2\sqrt{2}}}$, β_t is the sum over 2 days of the squared daily log(High/Low), and γ_t is the squared log(High/Low) where the High (Low) value is over 2 days.

$$\text{Closing Percent Quoted Spread}=\text{Average}\left(\frac{\text{Closing Ask}_t - \text{Closing Bid}_t}{(\text{Closing Ask}_t + \text{Closing Bid}_t)/2}\right)$$

Amihud=Average$\left(\frac{|r_t|}{\text{Dollar Volume}_t}\right)$, where the average is computed over positive volume days only and where r_t is the stock return on day t and Dollar Volume$_t$ is the US dollar value of volume on day t.

The ten cost-per-dollar-volume measures below for month i (or in some cases day i) are based on the "Extended Amihud" class of proxies as defined in Goyenko, Holden, and Trzcinka (2009), section 5.2:

Roll Impact$_i$ = Roll$_i$/(Average Daily US Dollar Value of Local Volume)$_i$.
Extended Roll Impact$_i$ = Extended Roll$_i$/(Average Daily US Dollar Value of Local Volume)$_i$.
Effective Tick Impact$_i$ = Effective Tick$_i$/(Average Daily US Dollar Value of Local Volume)$_i$.
LOT Mixed Impact$_i$ = LOT Mixed$_i$/(Average Daily US Dollar Value of Local Volume)$_i$.
LOT Y-Split Impact$_i$ = LOT Y-Split$_i$/(Average Daily US Dollar Value of Local Volume)$_i$.
FHT Impact$_i$ = FHT$_i$/(Average Daily US Dollar Value of Local Volume)$_i$.
Zeros Impact$_i$ = Zeros$_i$/(Average Daily US Dollar Value of Local Volume)$_i$.
Zeros2 Impact$_i$ = Zero2$_i$/(Average Daily US Dollar Value of Local Volume)$_i$.
High–Low Impact$_i$ = High–Low$_i$/(Average Daily US Dollar Value of Local Volume)$_i$.
Closing Percent Quoted Spread Impact$_i$ = Closing Percent Quoted Spread$_i$
/(Average Daily US Dollar Value of Local Volume)$_i$.

Pastor and Stambaugh=Γ, from the regression: $r^e_{t+1} = \theta + \varphi r_t + \Gamma \text{sign}(r^e_t)(\text{Volume}_t) + \varepsilon_t$, where r^e_t is the stock's excess return above the market portfolio on day t, θ is the intercept,

φ and Γ are regression coefficients, and ε_t is the error term. For LOT Mixed, LOT Y-Split, and Pastor and Stambaugh that require a market return, we use the local country value-weighted market portfolio.

$$\text{Amivest}=\text{Average}\left(\frac{\text{Volume}_t}{|r_t|}\right).$$

References

Amihud, Y. (2002) Illiquidity and stock returns: cross section and time-series effects, *Journal of Financial Markets* **5**, 31–56.

Asness, C., Moskowitz, T., and Pedersen, L. (2013) Value and momentum everywhere, *Journal of Finance* **68**, 929–985.

Attig, N., Gadhoum, Y., and Lang, L. (2003) Bid-ask spread, asymmetric information and ultimate ownership. Unpublished working paper, Laval University, University of Quebec, Chinese University of Hong Kong.

Bailey, W., Karolyi, A., and Salva, C. (2006) The economic consequences of increased disclosure: evidence from international cross-listings, *Journal of Financial Economics* **81**, 175–213.

Beber, A. and Pagano, M. (2013) Short-selling bans around the world: evidence from the 2007-09 crisis, *Journal of Finance* **68**, 343–381.

Bekaert, G., Harvey, C., and Lundblad, C. (2007) Liquidity and expected returns: lessons from emerging markets, *Review of Financial Studies* **20**, 1783–1831.

Bekaert, G., Harvey, C., Lundblad, C., and Siegel, S. (2014) Political risk spreads. NBER working paper, Columbia University, Duke University, University of North Carolina at Chapel Hill.

Bhattacharya, U., Daouk, H., Jorgenson, B., and Kehr, C. (2000) When an event is not an event: the curious case of an emerging market, *Journal of Financial Economics* **55**, 69–101.

Bundgaard, K. and Ahm, J. (2012) Secondary stock market liquidity and the cost of issuing seasoned equity – European evidence. Unpublished working paper, Copenhagen Business School.

Chan, J., Jain, R., and Xia, Y. (2008) Market segmentation, liquidity spillover, and closed-end country fund discounts, *Journal of Financial Markets* **11**, 377–399.

Chung, K. and Zhang, H. (2014) A simple approximation of intraday spreads using daily data, *Journal of Financial Markets* **17**, 94–120.

Clark, A. (2011) Revamping liquidity measures: improving investability in emerging and frontier market indices and their related ETFs. Unpublished working paper, Lipper, a Thompson Reuters company.

Corwin, S. and Schultz, P. (2012) A Simple way to estimate bid-ask spreads from daily high and low prices, *Journal of Finance* **67**, 719–759.

DeNicolo, G. and Ivaschenko, I. (2009) Global liquidity, risk premiums, and growth opportunities. Unpublished working paper, International Monetary Fund.

Edmans, A., Fang, V., and Zur, E. (2013) The effect of liquidity on governance, *Review of Financial Studies* **26**, 1443–1482.

Erten, I. and Okay, N. (2012) Deciphering liquidity risk on the Istanbul Stock Exchange. unpublished working paper, Bogazici University.

Fama, E. and MacBeth, J. (1973) Risk, return, and equilibrium: empirical tests, *Journal of Political Economy* **81**, 607–636.

Gersl, A. and Komarkov, Z. (2009) Liquidity risk and bank's bidding behavior: evidence from the global financial crisis. Unpublished working paper, Charles University, Czech National Bank.

Gomez-Puig, M. (2006) Size matters for liquidity: evidence from EMU sovereign yields spreads, *Economic Letters* **90**, 156–162.

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

Griffin, J., Hirschey, N., and Kelly, P. (2011) How important is the financial media in global markets?, *Review of Financial Studies* **24**, 3941–3992.

Griffin, J., Kelly, P., and Nardari, F. (2010) Do market efficiency measures yield correct inferences? A comparison of developed and emerging markets, *Review of Financial Studies* **23**, 3225–3277.

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

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

Hearn, B. (2014) The political institutional and firm governance determinants of liquidity: evidence from North Africa and the Arab Spring, *Journal of International Financial Markets, Institutions & Money* **31**, 127–158.

Hearn, B., Piesse, J., and Strange, R. (2010) Market liquidity and stock size premia in emerging financial markets: the implications for foreign investment, *International Business Review* **19**, 489–501.

Henkel, S. (2008) Is global illiquidity contagious? Contagion and cross-market commonality in liquidity. Unpublished working paper, Indiana University.

Henkel, S., Jain, P., and Lundblad, C. (2008) Liquidity dynamics and stock market automation. Unpublished working paper, Indiana University, University of Memphis, University of North Carolina.

Hennessy, J. and Patterson, D. (2012) *Computer Architecture: A Quantitative Approach*, Elsevier Morgan Kaufmann, Waltham, Massachusetts.

Holden, C. (2009) New low-frequency liquidity measures, *Journal of Financial Markets* **12**, 778–813.

Huang, R. and Stoll, H. (1996) Dealer versus auction markets: a paired comparison of execution costs on NASDAQ and the NYSE, *Journal of Financial Economics* **41**, 313–357.

Ince, O. and Porter, R. (2006) Individual equity return data From Thomson Datastream: handle with care!, *Journal of Financial Research* **29**, 463–479.

Jain, P. (2005) Financial market design and the equity premium: electronic versus floor trading, *Journal of Finance* **60**, 2955–2985.

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

Karnaukh, N., Renaldo, A., and Soderlind, P. (2015) Understanding FX liquidity, *Review of Financial Studies* **28**, 3073–3108.

Kyle, A. (1985) Continuous auctions and insider trading, *Econometrica* **53**, 1315–1335.

LaFond, R., Lang, M., and Skaife, H. (2007) Earnings smoothing, governance and liquidity: international evidence. Unpublished working paper, Massachusetts Institute of Technology, University of North Carolina, University of Wisconsin.

Lang, M., Lins, K., and Maffett, M. (2012) Transparency, liquidity, and valuation: international evidence on when transparency matters most, *Journal of Accounting Research* **50**, 729–774.

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

Lee, H., Tseng, Y., and Yang, C. (2014) Commonality in liquidity, liquidity distribution, and financial crisis: evidence from country ETFs, *Pacific-Basin Finance Journal* **29**, 35–58.

Lee, K. (2011) The world price of liquidity risk, *Journal of Financial Economics* **99**, 136–191.

Lesmond, D. (2005) Liquidity of emerging markets, *Journal of Financial Economics* **77**, 411–452.

Lesmond, D., Ogden, J., and Trzcinka, C. (1999) A new estimate of transaction costs, *Review of Financial Studies* **12**, 1113–1141.

Levine, R. and Schmukler, S. (2006) Internationalization and stock market liquidity, *Review of Finance* **10**, 153–187.

Liang, S. and Wei, K. C. (2006) Liquidity risk and expected returns around the world, *Journal of Banking and Finance* **36**, 3274–3288.

Marshall, B., Nguyen, N., and Visaltanachoti, N. (2012) Commodity liquidity measurement and transaction cost, *Review of Financial Studies* **25**, 599–638.

Marshall, B., Nguyen, N., and Visaltanachoti, N. (2013) Liquidity measurement in frontier markets, *Journal of International Financial Markets, Institutions and Money* **27**, 1–12.

Pastor, L. and Stambaugh, R. (2003) Liquidity risk and expected stock returns, *Journal of Political Economy* **111**, 642–685.

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

Schestag, R., Schuster, P., and Uhrig-Homburg, M. (2016) Measuring liquidity in bond markets, *Review of Financial Studies* **29**, 1170–1219.

Stahel, C. (2005) Is there a global liquidity factor? Unpublished working paper, George Mason University.

Swinscow, T. (1997) *Statistics at Square One*, 9th edn, BMJ Publishing Group, London.

Theil, H. (1966) *Applied Economic Forecasts*, North-Holland, Amsterdam.