Do Measures of Liquidity Measure Liquidity?*

Ruslan Y. Goyenko
McGill University

Craig W. Holden
Indiana University

Charles A. Trzcinka**
Indiana University

January 25, 2008

Abstract

Liquidity plays an increasingly important role in empirical asset pricing, market efficiency, and corporate finance. Identifying high quality proxies for liquidity based on daily data only (not intraday data) would permit liquidity to be studied over relatively long timeframes and across many countries. We introduce new liquidity measures. We run horseraces of both monthly and annual liquidity measures. We compare to effective spread, realized spread, and price impact based on both TAQ and Rule 605 data, including the decimals era. We identify the best proxies in each case and find that the new liquidity measures win the majority of horseraces.

JEL classification: C15, G12, G20.

Keywords: Liquidity, transaction costs, effective spread, price impact, asset pricing.

* We thank Utpal Bhattacharya, Andrew Ellul, Joel Hasbrouck, Christian Lundblad, Darius Miller, Xiaoyun Yu, and seminar participants at Indiana University and the Frontiers of Finance Conference in Bonaire, The Netherland Antilles. We also thank Charles Jones for making Dow spreads available. We are solely responsible for any errors. This paper was previously circulated as "Horseraces of Monthly and Annual Liquidity Measures."

** Corresponding author: Charles A. Trzcinka, Kelley School of Business, Indiana University, 1309 E. Tenth St., Bloomington, IN 47405-1701. Tel.: + 1-812-855-9908; fax: + 1-812-855-5875: email address: ctrzcink@ indiana.edu.

1. Introduction

The role of liquidity in empirical finance has grown rapidly over the past five years and has begun to have an effect on conclusions in asset pricing, market efficiency and corporate finance. A number of studies have proposed liquidity measures derived from daily return and volume data as proxies for the liquidity and transactions costs experienced by investors. Researchers rarely test the hypothesis that the liquidity proxies are related to actual transactions costs. The maintained assumption of most studies is that the available liquidity proxies capture the transactions costs of market participants. This assumption has not tested because of the limited availability of actual trading costs. In U.S. markets transactions data are only available since 1983; in many markets transactions data are not available at all. The consequences of not testing liquidity proxies on actual trading data is that there is little consensus on which measures are better and little evidence that any of the proposed measures are related to investor experience. In short, not much is known about whether transactions cost proxies measure what researchers claim they measure.

A handful of studies, Lesmond, Ogden, and Trzcinka (1999), Lesmond (2004), Hasbrouck (2006), test whether a few of the available liquidity proxies as constructed on an annual or quarterly basis from daily return data are related to annual or quarterly liquidity computed from transactions data. Yet the vast majority of the literature that *uses* liquidity proxies is based on *monthly* (or finer) data. This suggests the need to test *monthly* proxies. Furthermore, the literature has proposed different types of liquidity proxies designed to capture different liquidity benchmarks (e.g. effective spread, realized spread, or price impact). Given the limited number of liquidity proxies tested, the limited set of liquidity benchmarks compared to, and the absence of monthly proxies, it is not surprising that there are conflicting views about which measure is better and that there is little assurance that these measures actually capture the transactions costs of market participants.

The purpose of this paper is to address this omission in the literature. We provide a comprehensive study of available measures and introduce three proxies for effective spread and nine price impact measures. We run “horseraces” of liquidity measures testing all widely used proxies and our new ones. We conduct our tests relative to multiple liquidity benchmarks, using multiple high frequency datasets (TAQ and Rule 605 data), using multiple performance metrics, and over a long testing period that includes the decimals regime. We find a close association between many of the measures and actual transactions costs. Some measures are able to precisely estimate the magnitude of effective spread and many are highly correlated with the effective spread and price impact. We can safely assert that the literature has generally not been mistaken in the assumption that liquidity proxies measure liquidity. The new class of measures we introduce in this paper consistently wins a majority of our horseraces. Furthermore, a measure commonly used in the literature, Pastor-Stambaugh Gamma, is clearly dominated by much simpler measures. Further, we find a better alternative to the widely used Amihud measure.

The paper is organized as follows. Section 2 details the empirical design of the paper. Section 3 explains the high-frequency liquidity benchmarks in the horserace. Section 4 explains the low-frequency measures of effective spread in the horserace. Section 5 explains the low-frequency measures of price impact in the horserace. Section 6 describes the datasets and methodology used. Section 7 presents the horserace results. Section 8 concludes the paper.

2. Empirical Design

Our basic hypothesis is that useful monthly and annual liquidity measures going back in time can be constructed from low frequency (daily) stock returns and volume data, where such data are available. For the U.S., daily stock returns and volume data are available from the Center for Research in Security Prices (CRSP) covering NYSE/AMEX firms from 1926 to the present and covering NASDAQ firms from 1983 to the present. For international equity markets, daily stock returns and volume data are available from a wide variety of vendors. For example, Thompson Financial's DataStream provides daily stock returns and volumes covering firms in more than 60 countries from the 1990s to the present for most countries, with coverage for several developed markets extending to the early 1970s.

These tests should be of interest to a broad spectrum of empirical research in financial economics. First, consider applications in the asset pricing literature. For the U.S., Chordia, Roll, and Subrahmanyam (2000) show that various spread measures vary systematically. Goyenko (2006) shows that various effective spread measures are priced. Sadka (2003), Acharya and Pedersen (2004), Pastor and Stambaugh (2003) and Fujimoto and Watanabe (2006) show that various price impact measures are priced. Fujimoto (2003), Hasbrouck (2006), Korajczyk and Sadka (2006), and others test the pricing of *both* effective spread and price impact measures. Bekaert, Harvey and Lundblad (2005) test the pricing of *both* effective spread and price impact measures in emerging markets where liquidity concerns may be more pronounced. All of the above mentioned studies use monthly liquidity estimates. Reliable *monthly* effective spread and price impact measures going back in time and/or across countries are needed in order to determine if these asset pricing relationships hold up across time and space.

Next consider applications in the market efficiency literature. Starting with De Bondt and Thaler (1985), Jegadeesh and Titman (1993, 2001), Chan et al (1996), Rouwenhorst (1998), and many others have found trading strategies that appear to generate significant abnormal returns. Yet, Chordia, Goyal, Sadka, Sadka and Sivakumar (2008) show that the oldest trading strategy in the literature, the post earnings announcement drift, cannot produce returns greater than the Keim and Madhavan (1997) measures. Clearly liquidity measures over time and/or across countries are needed in order to determine if these trading strategies are truly profitable net of a relatively precise measure of cost of trading.

Finally there is a growing need in corporate finance research for useful monthly liquidity measures. Kalev, Pham, Steen (2003), Dennis and Strickland (2003), Cao, Field, and Hanka (2004), Lipson and Mortal (2004a), Schrand and Verrecchia (2004), Lesmond, Lemma, and O'Connor (2005), and many others examine the impact of corporate finance events on stock liquidity. Heflin and Shaw (2000), Lipson and Mortal (2004b), Lerner and Schoar (2004), and many others examine the influence of liquidity on capital structure, security issuance form, and other corporate finance decisions. Liquidity measures over time would expand the potential sample size of this literature. Liquidity measures across countries would greatly extend the potential diversity of international corporate finance environments that this literature could analyze.

In the exploration of the best measures, we compare many liquidity proxies calculated from low-frequency data to sophisticated benchmarks of liquidity calculated from two high-frequency datasets. First, we compare each measure to the effective spread, realized spread, and to two different price impact benchmarks calculated from the NYSE's Trade and Quote (TAQ) dataset from 1993 to 2005. Our monthly benchmarks are computed as monthly averages based on every trade and corresponding BBO[1] quote over the month. Similarly, our annual benchmarks are computed as annual averages based on every trade and corresponding BBO quote over the

[1] BBO means the best bid and offer. It is the highest bid price and lowest ask available for a given stock at a moment in time.

year. Second, we compare each measure to the effective spread for marketable orders[2] and price impact across order sizes[3] calculated from data required to be disclosed under SEC Rule 605 of regulation NMS (formerly regulation 11Ac1-5) from October 2001 to December 2005. Rule 605 requires all exchanges and other market centers to disclose very detailed order-based performance statistics by stock, order type, and order size, providing a cross-check to the TAQ based results.

We run monthly and annual horseraces between twelve effective spread contestants and twelve price impact contestants, gauging their abilities to match the salient features of our high-frequency based benchmarks. While some contestants are well established in the literature, many are being tested for the first time. The newly tested proxies for effective spread (described in detail below) are: the “Effective Tick,” and “Effective Tick2”measures, developed jointly by this paper and Holden (2007); the “Holden” measure from Holden (2007); and “LOT Y-Split” developed by this paper. The other contestants from the previous literature are: “Roll” from Roll (1984); the “Gibbs” measure from Hasbrouck (2004) ; the “LOT Mixed,” “Zeros,” and “Zeros2” measures from Lesmond, Ogden, and Trzcinka (1999); the “Amihud” measure from Amihud (2002); the “Pastor and Stambaugh” measure from Pastor and Stambaugh (2003); and finally the Amivest “Liquidity” ratio.[4]

---

[2] Marketable orders is the combination of market orders and marketable limit orders.

[3] Defined as the difference in the effective spread between large and small orders divided by the difference in the average share size between large and small orders.

[4] The Amihud, Pastor and Stambaugh, and Amivest measures are perhaps more naturally thought of as price impact measures, but the use of these measures in the literature has been more broadly and loosely justified. Therefore, we test these measures relative to *both* effective spread and price impact benchmarks.

Nine of the twelve price impact contestants (also described below) are based on a new class of price impact proxies developed by this paper as extensions of the Amihud measure from Amihud (2002). They are: (1) "Roll Impact," (2) "Effective Tick Impact," (3) "Effective Tick2 Impact," (4) "Holden Impact," (5) "Gibbs Impact," (6) "LOT Mixed Impact," (7) "LOT Y-split Impact," (8) "Zeros Impact," and (9) "Zeros2 Impact." The other three price impact measures we test are: (1) "Amihud" from Amihud (2002), (2) "Pastor and Stambaugh" from Pastor and Stambaugh (2003), and (3) the Amivest "Liquidity" ratio.

Our first performance metric is the average cross-sectional correlation based on individual firms between the low-frequency liquidity proxy and the high-frequency liquidity benchmark (either effective spread or one of the price impact benchmarks). Our second performance metric is the time-series correlation based on an equally-weighted portfolio each month between the liquidity proxy and the liquidity benchmark. Both of these performance metrics are most relevant for asset pricing purposes, where the scale of the low-frequency proxy does *not* matter; rather, all that matters is the magnitude of the correlation. Our third and fourth performance metrics are the prediction error between the liquidity proxy and the liquidity benchmark as measured by mean bias and by the root mean squared error, respectively. They are most relevant for market efficiency and corporate finance tests, where the scale *does* matter, because one wishes to subtract a correctly-scaled proxy for transaction costs.

Hasbrouck (2006) runs annual horseraces between four effective cost measures, comparing each to effective spread and price impact computed from TAQ data from 1993 – 2005. Among the measures he tests, Gibbs dominates as a proxy for annual effective spread and

Illiquidity dominates as a proxy for annual price impact.[5] Lesmond, Ogden, and Trzcinka (1999) run annual horseraces between three liquidity measures, and find that LOT dominates Roll and Zeros. Lesmond (2004) runs quarterly horseraces between five liquidity measures for 23 emerging countries, and finds that LOT dominates Roll, Illiquidity, Liquidity, and Turnover.

Our results can be summarized as follows. In the monthly effective spread horseraces, we find that Holden, Effective Tick, and LOT Y-split are the best overall. In the annual effective spread horseraces, we find that six low-frequency proxies (Roll, Effective Tick, Effective Tick2, Holden, Gibbs, and LOT Y-split) dominate; with Gibbs being the best overall. We also find that in the more recent years, during decimals regime, the performance of all measures deteriorates with the exception of Zeros and Amihud measures.

For the realized spread horseraces, the Amihud measure is the best overall. XXX expand realized spread results a bit XXX

For the 5-minute price impact horseraces, the new class of price impact proxies that we develop dominates the measures in the existing literature; with Holden Impact being the best overall. For the permanent price impact horseraces, there are mixed results.

For the monthly Rule 605 effective spreads horserace, Effective Tick is the best overall. All of the price impact proxies fail to capture the special Rule 605 version of price impact.

3. High Frequency Liquidity Benchmarks

[5] Hasbrouck then extends his basic model to include a latent common liquidity factor for a subsample of stocks. He also estimates his Gibbs measure for all common NYSE/AMEX/NASDAQ stocks from 1926 to present and tests whether liquidity is a priced risk factor. He finds only weak support for the view that effective cost affects expected stock returns, except when interacted with a January season dummy variable.

Our liquidity benchmarks are two measures of effective spread and four measures of price impact (described below). Effective spread is an estimate of the percentage cost of trading for a hypothetical one-shot, round trip transaction of the average trade size used to calculate it.

Our first measure of liquidity is effective spread as calculated from the TAQ database. Specifically, the TAQ effective spread of a particular stock on the k^{th} trade is defined as

$$\text{Effective Spread (TAQ)}_{\text{k}} = 2 \cdot \left| \ln\left(P_k\right) - \ln\left(M_k\right) \right|, \tag{1}$$

where P_k is the price of the k^{th} trade and M_k is the midpoint of the consolidated BBO prevailing at the *time of the* k^{th} *trade*. For a particular stock aggregated over a time interval *i* (either a month or a year), the $\text{Effective Spread (TAQ)}_{\text{i}}$ is the dollar-volume-weighted average of $\text{Effective Spread (TAQ)}_{\text{k}}$ computed over all trades in time interval *i*.

Our second measure of liquidity is effective spread as aggregated from figures in the Rule 605 database. Specifically, the Rule 605 dollar effective spread of a particular stock based on the trade generated by the k^{th} order is defined as

$$\text{\$ Effective Spread (605)}_{\text{k}} = \begin{cases} 2 \cdot \left(P_k - m_k\right) & \text{for marketable buys} \\ 2 \cdot \left(m_k - P_k\right) & \text{for marketable sells} \end{cases}, \tag{2}$$

where m_k is the midpoint of the consolidated BBO prevailing at the *time of receipt* of the k^{th} order at the exchange.[6] For a particular stock aggregated over a time interval *i* (either a month or a year), the $\text{Effective Spread (605)}_{\text{i}}$ is the dollar-volume-weighted average of $\text{\$ Effective Spread (605)}_{\text{k}}$ computed over all trades in time interval *i* and then divided by $\overline{P}_i$, the average price in time interval *i*.

[6] Marketable buys are market buy orders and marketable limit buy orders. Marketable sells are market sell orders and marketable limit sell orders. Effective spreads are not reported for non-marketable limit orders in the 605 data.

In principle, the $\text{Effective Spread (605)}_i$ should be an improvement over the $\text{Effective Spread (TAQ)}_i$. Each market center constructs their Rule 605 figures from *order data*, which is more refined than trade and quote data for several reasons. First, the Rule 605 midpoint is based on the order's *time of receipt*, whereas a TAQ midpoint is based on the trade's time of execution. The order's time of receipt is a closer proxy to the trader's information set at the time of order submission. Second, there is no confusion in the Rule 605 data about buys vs. sells or about marketable orders vs. non-marketable orders. Lee and Radhakrishna (2000) report that the Lee and Ready (1991) method commonly used with TAQ data incorrectly classifies 24% of inside-the-spread trades that have a clear trade initiator. Third, there is no confusion in the Rule 605 data when a marketable buy is crossed with a marketable sell. Lee and Radhakrishna (2000) find that 40% of the trades in their TORQ sample are "non-directional" trades, where a marketable buy and marketable sell are crossed. The Rule 605 data correctly treats this case as two marketable executions (both a marketable buy execution and a marketable sell execution). By contrast, users of TAQ data cannot distinguish nondirectional trades vs. directional trades and usually treat this case as a single marketable execution.[7] Accordingly, the Rule 605 data provide a useful cross-check to the TAQ based results; however, the Rule 605 data are only available from mid-2001, so the comparison is limited to only 51 months in our sample.

The term "price impact" has been used in a variety of ways in the literature and we construct corresponding benchmarks. A static version of price impact is the slope of the price function at a moment in time. Essentially, this is the cost of demanding additional instantaneous

[7] There are downsides to 605 data as well. An order that is re-routed between market centers is double-counted. The 605 data do not include block trades. The SEC is an imperfect monitor of data quality. For more discussion of these issues, see Boehmer, Jennings, and Wei (2003).

liquidity. It can be thought of as the first derivative of the effective spread with respect to the order size. Our third measure of liquidity uses two (aggregated) points on this curve to measure this slope. Specifically, Static Price Impact based on the Rule 605 data of a particular stock over time interval *i* is defined as

$$\text{Static Price Impact (605)}_{\text{i}} = \begin{bmatrix} \left(\text{\$ Effective Spread (605)}_{\text{Big Orders,i}} / \bar{P}_i\right) \\ -\left(\text{\$ Effective Spread (605)}_{\text{Small Orders,i}} / \bar{P}_i\right) \end{bmatrix} / \begin{bmatrix} \left(\text{Ave Trade Size (605)}_{\text{Big Orders,i}}\right) \\ -\left(\text{Ave Trade Size (605)}_{\text{Small Orders,i}}\right) \end{bmatrix}, \tag{3}$$

where Big Orders,i is the set of all orders in the range of 2,000 – 9,999 shares that execute in time interval *i* and Small Orders,i is the set of all orders in the range of 100 – 499 shares that execute in time interval *i*.

Our fourth measure of liquidity introduces a time dimension that is not present in static price impact. Five-minute price impact measures the derivative of the cost of demanding a certain amount of liquidity over five minutes. It may be very different from the analogous curve for demanding the same amount of liquidity immediately. We follow Hasbrouck (2006) by calculating TAQ price impact over time interval *i* as the slope coefficient of the following regression

$$r_n = \left(\text{Five-Minute Price Impact (TAQ)}_{\text{i}}\right) \cdot S_n + u_n, \tag{4}$$

using data from every five-minute period n in time interval *i*.[8] Specifically, r_n is the stock return over the n^{th} five-minute period, S_n is the signed square-root dollar volume over the n^{th} five-

[8] We also used a 15 minute interval with similar results suggesting that our results are independent of the time interval over which we aggregate the data.

minute period $S_n = \sum_k Sign(v_{kn})\sqrt{|v_{kn}|}$, v_{kn} is the signed dollar volume of the k^{th} trade in the n^{th} five-minute period, and u_n is the error term for the n^{th} five-minute period.

Both of price impact measures mentioned so far implicitly include both temporary and permanent price impacts. However, some liquidity proxies (e.g., the "Gamma" price reversal measure of Pastor and Stambaugh 2003) are designed to capture *only* the temporary component. Our fifth measure of liquidity is realized spread from Huang and Stoll (1996), which is the temporary component of the effective spread. Specifically, the TAQ realized spread of a particular stock on the k^{th} trade is defined as

$$\text{Realized Spread}(\text{TAQ}_k) = 2 \cdot \left|\ln(P_k) - \ln(M_{k+5})\right|, \tag{5}$$

where M_{k+5} is the midpoint of the consolidated BBO prevailing *five-minutes after the* k^{th} *trade*. For a particular stock aggregated over a time interval *i* (either a month or a year), the $\text{Realized Spread (TAQ)}_\text{i}$ is the dollar-volume-weighted average of $\text{Realized Spread (TAQ)}_\text{k}$ computed over all trades in time interval *i*.

Finally, another notion of price impact focuses on the change in prices and quotes after a signed trade. A common definition of price impact is the increase (decrease) in the midpoint over a five minute interval beginning at the time of the buyer- (seller-) initiated transaction. This is the permanent price change of a given transaction, or equivalently, the permanent component of the effective spread. Huang and Stoll (1996) show that this permanent price impact is mathematically equal to the effective spread minus the realized spread. Specifically, the TAQ permanent price impact of a particular stock aggregated over a time interval *i* is

$$\begin{aligned}\text{Permanent Price Impact} &= \text{Effective Spread (TAQ)}_\text{i} - \text{Realized Spread (TAQ)}_\text{i} \\ &= 2 \cdot \left|\ln(M_{k+5}) - \ln(M_k)\right|.\end{aligned} \tag{6}$$

4. Low-Frequency Measures of Effective Spread

Eight low-frequency measures of effective spread are explained below. For each measure, we require that the measure always produces a numerical result. In other words, if a given measure can not be computed for a given stock / time period, then a default numerical value must be substituted. We have tested other possibilities and found that our results are not sensitive to the choice of default numerical value.

4.1. Roll

Roll (1984) develops an estimator of the effective spread based on the serial covariance of the change in price as follows. Let V_t be the unobservable fundamental value of the stock on day t. Assume that it evolves as

$$V_t = V_{t-1} + \varepsilon_t , \tag{7}$$

where ε_t is the mean-zero, serially uncorrelated public information shock for day t.

Let P_t be the last observed trade price on day t. Assume it is determined by

$$P_t = V_t + \tfrac{1}{2} S Q_t , \tag{8}$$

where S is the effective spread and Q_t is a buy/sell indicator for the last trade that equals $+1$ for a buy and -1 for a sell. Assume that Q_t is equally likely to be $+1$ or -1, is serially uncorrelated, and is independent of ε_t. Taking the first difference of Eq. (8) and combining it with Eq. (7) yields

$$\Delta P_t = \tfrac{1}{2} S \Delta Q_t + e_t \tag{9}$$

where Δ is the change operator. Given this setup, Roll shows that the serial covariance is

$$Cov\left(\Delta P_t , \Delta P_{t-1}\right) = \tfrac{1}{4} S^2 \tag{10}$$

or equivalently

$$S = 2\sqrt{-Cov\left(\Delta P_t, \Delta P_{t-1}\right)} \,. \tag{11}$$

When the sample serial covariance is positive, the formula above is undefined and so we substitute a default numerical value of zero. Hence, we use a modified version of the Roll estimator

$$Roll = \begin{cases} 2\sqrt{-Cov\left(\Delta P_t, \Delta P_{t-1}\right)} & \text{When } Cov\left(\Delta P_t, \Delta P_{t-1}\right) < 0 \\ 0 & \text{When } Cov\left(\Delta P_t, \Delta P_{t-1}\right) \geq 0 \end{cases}. \tag{12}$$

4.2. Effective Tick

Holden (2007) and this paper jointly develop a proxy of the effective spread based on observable price clustering.[9] Based on the negotiation cost theory of Harris (1991), we assume that trade prices are clustered in order to minimize negotiation costs between potential traders. Let S_t be the realization of the effective spread on the closing trade of day t. Assume that the realization of the spread on the closing trade of day is randomly drawn from a set of possible spreads s_j, $j = 1, 2, \ldots, J$ with corresponding probabilities γ_j, $j = 1, 2, \ldots, J$. By convention, the possible effective spreads $s_1, s_2, \ldots s_J$ are ordered from smallest to largest. For example on a $\$\frac{1}{8}$ price grid, S_t is modeled as having a probability γ_1 of $s_1 = \$\frac{1}{8}$ spread, γ_2 of $s_2 = \$\frac{1}{4}$ spread, γ_3 of $s_3 = \$\frac{1}{2}$ spread, and γ_4 of $s_4 = \$1$ spread.

Following the intuition of Christie and Schultz (1994), we assume that price clustering is completely determined by the spread size. For example, if the spread is $\$\frac{1}{4}$, the model assumes that the bid and ask prices employ only even quarters. The quote could be $\$25\frac{1}{4}$ bid, $\$25\frac{1}{2}$ offered, but never $\$25\frac{3}{8}$ bid, $\$25\frac{5}{8}$ offered. Thus, if odd-eighth transaction prices are observed, one infers that the spread must be $\$\frac{1}{8}$.

[9] Holden (2006) also develops and tests many additional versions of the Effective Tick measure.

This implies that the simple frequency with which closing prices occur in special price clusters can be used to estimate the spread probabilities $\hat{\gamma}_j$, $j = 1, 2, \ldots, J$. For example on a $\$\frac{1}{8}$ fractional price grid, the frequency with which trades occur on odd $\frac{1}{8}$s, odd $\frac{1}{4}$s, odd $\frac{1}{2}$s, and whole dollars can be used to estimate the probability of a $\$\frac{1}{8}$ spread, $\$\frac{1}{4}$ spread, $\$\frac{1}{2}$ spread, and a \$1 spread. Similarly for a decimal price grid, the frequency of off pennies, off nickels, off dimes, off half dollars and whole dollars can be used to estimate the probability of a penny spread, nickel spread, dime spread, quarter spread, and a whole dollar spread. Let N_j be the empirical number of special trade prices corresponding to the j^{th} spread $(j = 1, 2, \ldots, J)$ from positive-trade days in the time interval. In the $\$\frac{1}{8}$ price grid example (where $J = 4$), N_1 through N_4 are the empirical number of odd $\frac{1}{8}$ prices, the number of odd $\frac{1}{4}$ prices, the number of odd $\frac{1}{2}$ prices, and the number of whole dollar prices, respectively.

Let F_j be the empirical probabilities of special trade prices corresponding to the j^{th} spread $(j = 1, 2, \ldots, J)$. These empirical probabilities are computed as

$$F_j = \frac{N_j}{\sum_{j=1}^{J} N_j} \quad \text{for } j = 1, 2, \ldots, J \,. \tag{13}$$

Let U_j be the unconstrained probability of the j^{th} spread $(j = 1, 2, \ldots, J)$. The unconstrained probability of the effective spread is

$$U_j = \begin{cases} 2F_j & j = 1 \\ 2F_j - F_{j-1} & j = 2, 3, \ldots, J-1\,. \\ F_j - F_{j-1} & j = J. \end{cases} \tag{14}$$

The effective tick model directly assumes price clustering (i.e., a higher frequency on rounder increments). However, in small samples it is possible that *reverse* price clustering may be realized (i.e., a lower frequency on rounder increments). Reverse price clustering unintentionally causes the unconstrained probability of one or more effective spread sizes to go above 1 or below 0. Thus, constraints are added to generate proper probabilities. Let $\hat{\gamma}_j$ be the *constrained* probability of the jth spread $\left(j = 1, 2, \ldots, J \right)$. It is computed in order from smallest to largest as follows

$$\hat{\gamma}_j = \begin{cases} Min\left[Max\left\{ U_j, 0 \right\}, 1 \right] & j = 1 \\ Min\left[Max\left\{ U_j, 0 \right\}, 1 - \sum_{k=1}^{j-1} \hat{\gamma}_k \right] & j = 2, 3, \ldots, J. \end{cases} \tag{15}$$

Finally, the effective tick measure is simply a probability-weighted average of each effective spread size divided by $\overline{P}_i$, the average price in time interval *i*.

$$Effective \;\; Tick = \frac{\sum_{j=1}^{J} \hat{\gamma}_j s_j}{\overline{P}_i} \tag{16}$$

A second version, called *Effective Tick2*, is otherwise the same except that it uses the daily prices from *all* days, rather than just positive volume days only. The difference between the two versions depends on the usefulness of the no trade prices.

4.3. Holden

Holden (2007) develops a model that uses both serial correlation (like the Roll measure) and price clustering (like Effective Tick) to estimate the effective spread. Indeed, the Holden model formally nests both the Roll model and the Effective Tick model as special cases. His model is based on modifying the model of Huang and Stoll (1997). Huang and Stoll develop a generalized model of components of the bid-ask spread. A by-product of the Holden model is a two-way decomposition of the bid-ask spread as estimated from low-frequency data.

Holden begins by modifying the Huang and Stoll model to account for changing spreads linked to price clustering. Just like the Effective Tick model above, he specifies a random probability of jumping each period among multiple spreads that are linked price cluster regimes.

Next, he derives a price change process that is a natural extension of Eq. (9) above

$$\Delta P_t = \tfrac{1}{2} S_t Q_t - (1-\lambda) \tfrac{1}{2} S_{t-1} Q_{t-1} + e_t \tag{17}$$

where the effective spread S_t is allowed to change each day and λ is the percentage of the half-spread attributable to the sum of adverse selection and inventory holding costs. Conversely, $1-\lambda$ is the percentage of the half-spread attributable to order processing costs.[10] The public information shock e_t is assumed to be normally distributed with a mean $\bar{e}$ and a standard deviation σ_e.

Let μ be the probability of a trading day and $1-\mu$ be the probability of a non-trading day. Consider a $\$\frac{1}{8}$ price grid where S_t has a probability γ_1 of $s_1 = \$\frac{1}{8}$ spread, γ_2 of $s_2 = \$\frac{1}{4}$ spread, γ_3 of $s_3 = \$\frac{1}{2}$ spread, and γ_4 of $s_4 = \$1$ spread. Of course, the spread probabilities must sum to one: $\sum_{j=1}^{J} \gamma_j = 1$. The Holden spread proxy is just the weighted-average of the possible spreads

$$\text{Holden} \equiv S_H = \sum_{j=1}^{J} \gamma_j s_j \, . \tag{18}$$

Define the variable C_t as the observable price cluster on day t. Specifically, on a zero volume day, let $C_t = 0$. On a positive volume day, let clusters $C_t = 1,\ 2,\ 3,$ and 4 correspond to when the trade price is on odd $\frac{1}{8}$s, odd $\frac{1}{4}$s, odd $\frac{1}{2}$s, and whole dollars, respectively. Define $\hat{Q}_t$

[10] This component also includes any liquidity provider rents due to market power or price discreteness.

as a buy/sell/zero volume indicator on day d that equals +1 for a buy, -1 for a sell, and 0 for a zero volume day. Define the *unobserved* signed half spread on day t as $H_t = \frac{1}{2} S_t \hat{Q}_t$. Considering all spread and indicator combinations, there are nine possible values of the signed half spread H_t : $\$\frac{1}{2}, \$\frac{1}{4}, \$\frac{1}{8}, \$\frac{1}{16}, \$0, -\$\frac{1}{16}, -\$\frac{1}{8}, -\$\frac{1}{4}, -\$\frac{1}{2}$.

On three successive trading days, we observe a price triplet (P_t, P_{t+1}, P_{t+2}), which corresponds to a price cluster triplet (C_t, C_{t+1}, C_{t+2}). Define H as the set of all half spread triplets (H_t, H_{t+1}, H_{t+2}) that are feasible given the observed price cluster triplet.[11] For a given a set of parameter values $(\mu, \gamma_1, \gamma_2, S_H, \overline{e}, \sigma_e, \lambda)$, Holden calculates the likelihood of the price triplet

$$\begin{aligned}&\Pr\left(P_t, P_{t+1}, P_{t+2} \mid \mu, \gamma_1, \gamma_2, S_H, \overline{e}, \sigma_e, \lambda\right)\\&= \sum_{(H_t, H_{t+1}, H_{t+2}) \in H} \left\{ \begin{array}{l} \Pr(C_t) \cdot \Pr(C_{t+1}) \cdot \Pr(C_{t+2}) \cdot \Pr(H_t \mid C_t) \cdot \Pr(H_{t+1} \mid C_{t+1}) \cdot \Pr(H_{t+2} \mid C_{t+2}) \\ \cdot n\left(P_{t+1} - H_{t+1} - \left(P_t - (1-\lambda) H_t\right)\right) \cdot n\left(P_{t+2} - H_{t+2} - \left(P_t - (1-\lambda) H_{t+1}\right)\right) \end{array} \right\}, \end{aligned} \quad (19)$$

where $n(\ )$ is the normal density with a mean of $\overline{e}$ and a standard deviation of σ_e. Using three prices at a time allows the serial correlation of the price changes to be picked up, but avoids the combinatoric explosion of feasible half spread combinations that would result if all observations were used at the same time.

---

[11] For example, suppose that the price $P_t = \$25\,1/8$, which is an odd eighth that corresponds to price cluster $C_t = 1$. For this price cluster there is only feasible spread $S_t = \$\,1/8$. Thus, there are only two feasible value of the signed half spreads $H_t \in \{\$\,1/16, -\$\,1/16\}$. Similarly, P_{t+1} and P_{t+2} imply the feasible values of the signed half spreads H_{t+1} and H_{t+2}. Taking all combinations of the feasible values on each day, yield the set of feasible half spread triplets.

Taking the log of Eq. (19), the likelihood function is the sum of the log likelihoods of all price triplets in the time period of aggregation

$$\sum_{t=1}^{T-2} Ln\left(\Pr\left(P_t, P_{t+1}, P_{t+2} \mid \mu, \gamma_1, \gamma_2, S_H, \overline{e}, \sigma_e, \lambda\right)\right), \tag{20}$$

where T is the number of days in the time period of aggregation. The likelihood function is maximized by choice of the parameters $\mu, \gamma_1, \gamma_2, S_H, \overline{e}, \sigma_e, \lambda$ subject to the constraints that $\gamma_1, \gamma_2, \gamma_3, \gamma_4, \mu, S_H, \sigma_e,$ and λ are greater than or equal to zero and the constraints that $\gamma_1, \gamma_2, \gamma_3, \gamma_4, \mu,$ and λ are less than or equal to one.[12]

4.4. Gibbs

Hasbrouck (2004) develops a method for Bayesian estimation of the Roll model. Specifically, he performs Gibbs sampler estimation of the Roll model using prices from all days. Hasbrouck assumes that the public information shock ε_d in the Roll model is normally distributed with a mean of zero and a variance of σ_ε^2. He denotes the half spread in Roll model as $c \equiv \frac{1}{2} S$.

Hasbrouck uses the Gibbs sampler to numerically estimate the model parameters $\left\{c, \sigma_\varepsilon^2\right\}$, the latent buy/sell/no-trade indicators $Q = \left\{Q_1, Q_2, \ldots, Q_T\right\}$, and the latent "efficient prices" $V = \left\{V_1, V_2, \ldots, V_T\right\}$, where T is the number of days in the time interval. Specifically, the Gibbs

---

[12] The constraints $\gamma_3 \geq 0$ and $\gamma_3 \leq 1$ can be expressed as a function of the parameters to be estimated $\left(\mu, \gamma_1, \gamma_2, S_H, \overline{e}, \sigma_e, \lambda\right)$ as: $2\left[1 - S - \gamma_1\left(\frac{7}{8}\right) - \gamma_2\left(\frac{3}{4}\right)\right] \geq 0$ and $2\left[1 - S - \gamma_1\left(\frac{7}{8}\right) - \gamma_2\left(\frac{3}{4}\right)\right] \leq 1$, respectively. Similarly, the constraints $\gamma_4 \geq 0$ and $\gamma_4 \leq 1$ can be expressed as: $1 - \gamma_1 - \gamma_2 - 2\left[1 - S - \gamma_1\left(\frac{7}{8}\right) - \gamma_2\left(\frac{3}{4}\right)\right] \geq 0$ and $1 - \gamma_1 - \gamma_2 - 2\left[1 - S - \gamma_1\left(\frac{7}{8}\right) - \gamma_2\left(\frac{3}{4}\right)\right] \leq 1$, respectively.

sampler is an iterative process with three steps for each cycle (or "sweep"). First, given the sample of prices P from all days in the time interval, starting values for Q, a prior for c, and a prior for σ_ε^2, estimate c using a Bayesian regression that is restricted to the positive domain. Second, given P, Q, the prior for σ_ε^2, and the updated estimate of c, estimate the residuals and make a new draw of σ_ε^2 from an inverted gamma distribution. Third, given P, the updated estimate of c, and the new draw of σ_ε^2, make new draws of Q and V.

Hasbrouck runs 1,000 sweeps of the sampler. He discards the first 200 as burn-in and takes the mean of the c values in the remaining 800 sweeps as the final estimate of c.

Hasbrouck also extends the Gibbs measure to account for discreteness, clustering, and asymmetric information. The Gibbs measure has also been used by Hasbrouck (1999a, 2004). Hasbrouck generously provides programming code to compute Gibbs on his web site and we directly use his code without modification of the main routines for both monthly and annual computations.

4.5. LOT

Lesmond, Ogden, and Trzcinka (1999) develop an estimator of the effective spread based on the idea of informed trading on non-zero return days and the absence of informed trading on zero return days. A standard "market model" relationship holds on non-zero return days, but a flat horizontal segment applies on zero return days.

The LOT model assumes that the unobserved "true return" R_{jt}^* on a stock *j* on day *t* is given by

$$R_{jt}^* = \beta_j R_{mt} + \varepsilon_{jt} , \tag{21}$$

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 a mean of zero and a variance of σ_j^2. Let $\alpha_{1j} \le 0$ be the percent transaction cost of selling stock *j* and let $\alpha_{2j} \ge 0$ be the percent transaction cost of buying stock *j*. Then, the observed return R_{jt} on a stock *j* is given by

$$\begin{aligned} R_{jt} &= R_{jt}^* - \alpha_{1j} && \text{when } R_{jt}^* < \alpha_{1j} \\ R_{jt} &= R_{jt}^* && \text{when } \alpha_{1j} < R_{jt}^* < \alpha_{2j} \\ R_{jt} &= R_{jt}^* - \alpha_{2j} && \text{when } \alpha_{2j} < R_{jt}^*. \end{aligned} \tag{22}$$

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

$$LOT = \alpha_{j2} - \alpha_{j1}. \tag{23}$$

Lesmond, Ogden, and Trzcinka develop the following maximum likelihood estimator of the model's parameters

$$\begin{aligned} & L\left(\alpha_{1j}, \alpha_{2j}, \beta_j, \sigma_j \mid R_{jt}, R_{mt}\right) \\ & = \prod_1 \frac{1}{\sigma_j} n\left[\frac{R_{jt} + \alpha_{1j} - \beta_j R_{mt}}{\sigma_j}\right] \\ & \times \prod_0 \left[N\left(\frac{\alpha_{2j} - \beta_j R_{mt}}{\sigma_j}\right) - N\left(\frac{\alpha_{1j} - \beta_j R_{mt}}{\sigma_j}\right)\right] \\ & \times \prod_2 \frac{1}{\sigma_j} n\left[\frac{R_{jt} + \alpha_{2j} - \beta_j R_{mt}}{\sigma_j}\right] \\ & S.T. \quad \alpha_{j1} \ge 0, \alpha_{j2} \le 0, \beta_j \ge 0, \sigma_j \ge 0, \end{aligned} \tag{24}$$

where $N(\ )$ is the cumulative normal distribution.

A very important issue of the LOT measure is the definition of the three regions over which the estimation is done. The original LOT (1999) measure, which we call LOT Mixed, broke out the three regions based on both the X-variable and Y-variable. That is, 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 $R_{mt} < 0$. In this paper we develop an alternative measure, which we call LOT Y-split, that breaks out the three regions based on the Y-variable. That is, region 0 is $R_{jt} = 0$, region 1 is $R_{jt} > 0$, and region 2 is $R_{jt} < 0$. Interestingly, LOT Y-split and LOT Mixed sometimes produce very different results, so it is worth tracking both of them.

4.6. Zeros

Lesmond, Ogden, and Trzcinka (1999) develop the proportion of days with zero returns as a proxy for liquidity. There are two key arguments that support this measure. First, stocks with lower liquidity are more likely to have zero volume days and thus more likely to have nothing-going-on, zero return days. Second, stocks with higher transaction costs have less private information acquisition (because it is more difficult to overcome higher transaction costs) and thus, even on positive volume days, they are more likely to have no-information-revelation, zero return days.

Lesmond, Ogden, and Trzcinka define the proportion of days with zero returns as

$$\text{Zeros} = (\text{\# of days with zero returns})/T \qquad (25)$$

An alternative version, called Zeros2, is defined as

$$\text{Zeros2} = (\text{\# of positive volume days with zero return})/T \qquad (26)$$

"T" is the number trading days in a month. For emerging markets, the Zeros measure has been used by Bekaert, Harvey, and Lundblad (2005).

4.7. Other Proxies

Three additional proxies are tested: (1) "Illiquidity" from Amihud (2002), (2) "Gamma" from Pastor and Stambaugh (2002), and (3) the (Amivest) "Liquidity" ratio. These measures are intended to be proxies for price impact, but for completeness we test them for correlation with effective spread and realized spread as well. All three are described below.

5. Low-Frequency Measures of Price Impact

Next, we explain ten low-frequency measures of price impact. As before, we require that each measure always produce a numerical result.

5.1. Amihud

Amihud (2002) develops a price impact measure which represents the "daily price response associated with one dollar of trading volume." Specifically, he uses the following ratio

$$\text{Illiquidity } = Average\left(\frac{|r_t|}{Volume_t}\right), \tag{27}$$

where r_t is the stock return on day *t* and $Volume_t$ is the dollar volume on day *t*. The average is calculated over all positive-volume days, since the ratio is undefined for zero volume days.

5.2. The Extended Amihud Proxies

We develop a new class of price impact proxies by extending the Amihud measure. We start with the Amihud base model. Then we decompose the total return in the base model numerator into a liquidity component and a non-liquidity component. This is done by dividing both sides of Eq. (9) by P_{t-1} to obtain

$$r_t = \frac{\frac{1}{2} S\Delta Q_t}{P_{t-1}} + \frac{e_t}{P_{t-1}} . \tag{28}$$

where the first term on the right is the liquidity component and the second term is the non-liquidity component. Recall that $\frac{1}{2} S\Delta Q_t$ is the change in the signed effective half spread and ε_t is the mean-zero, serially uncorrelated public information shock for day t. Substituting Eq. (28) into Eq. (27), we get

$$Average\left(\frac{\left|\frac{\frac{1}{2}S\Delta Q_t}{P_{t-1}}+\frac{e_t}{P_{t-1}}\right|}{Volume_t}\right). \quad (29)$$

By assumption, the random variable ε_t is independent of the liquidity component. Therefore, we drop the non-liquidity component in order to measure the *liquidity costs associated with one dollar of trading volume*, as

$$Average\left(\frac{\left|\frac{\frac{1}{2}S\Delta Q_t}{P_{t-1}}\right|}{Volume_t}\right). \quad (30)$$

Essentially, this eliminates a noise term that is unrelated to the variable of interest. The average numerator value is very close to the percent effective half spread. Since we don't observe the numerator in low-frequency datasets, we construct an extended Amihud proxy for time interval *i* by using a proxy for percent effective spread over time interval *i* as follows

$$\text{Extended Amihud Proxy}_{\text{i}} = \frac{\text{Proxy for Percent Effective Spread}_{\text{i}}}{\text{Average Volume}_{\text{i}}}, \quad (31)$$

where the whole spread convention is used instead of the half spread convention.[13]

[13] Eq. (31) can be obtained in an alternative manner. First use Amihud and Mendelson (1986) to decompose the total return into two pieces $r_t = r_t^* + \mu S_i$, where r^* is spread-adjusted return (gross return minus spread-adjustment) and μS_i is the expected liquidation costs (or spread adjustment), which is equal to the product of the liquidation probability per unit time, μ, by the percentage spread, S_i for time interval *i*. Assume that the holding period of the marginal investor is equal to the time interval *i*, so that $\mu = 1$. By assumption, spread adjusted return, r^*, does not contain liquidity component. Therefore, to measure the liquidity costs associated with one dollar of trading volume,

This defines a class of price impact proxies depending on what particular proxy for percent effective spread is used. For example, one member of this class is called the Roll Impact measure for time interval *i*, which uses Roll measure for time interval *i* and the average daily dollar volume over time interval *i* as follows

$$\text{Roll Impact}_i = \frac{\text{Roll}_i}{\text{Average Volume}_i}. \quad (32)$$

We test nine versions of this class of price impact measures based on seven proxies for percent effective spread. The seven measures we test are: Roll Impact, Effective Tick Impact, Effective Tick2 Impact, Holden Impact, Gibbs Impact, LOT Mixed Impact, and LOT Y-split Impact, Zeros Impact, Zeros2 Impact.

5.3. Pastor and Stambaugh

Pastor and Stambaugh (2002) develop a measure Gamma of price impact by running the following regression

$$r_{t+1}^e = \theta + \phi r_t + (Gamma)\, sign\left(r_t^e\right)\left(Volume_t\right) + \varepsilon_t \quad (33)$$

where r_t^e is the stock's excess return above the CRSP value-weighted market return on day *t* and $Volume_t$ is the dollar volume on day *t*. Intuitively, Gamma measures the reverse of the previous day's order flow shock. Gamma should have a negative sign. The larger the absolute value of Gamma, then the larger the implied price impact.

5.4. Amivest Liquidity

The Amivest Liquidity ratio is a measure of price impact

we are left with the ratio $Average\left(S_i / Volume_i\right)$. The average numerator value is just the percent effective spread. We approximate this with Eq. (31).

$$\text{Liquidity} = Average\left(\frac{Volume_t}{|r_t|}\right). \tag{34}$$

The average is calculated over all non-zero return days, since the ratio is undefined for zero return days. A larger value of Liquidity implies a lower price impact. This measure has been used by Cooper, Groth, and Avera (1985), Amihud, Mendelson, and Lauterback (1997), and Berkman and Eleswarapu (1998) and others.

6. Data

To compute our effective spread and price impact benchmarks, we use two high-frequency datasets. First, we used the NYSE TAQ data from 1993 to 2005. Because of the enormous size of the TAQ data, we select a random sample. Following the methodology of Hasbrouck (2006), a stock must meet five criteria to be eligible: (1) it has to be a common stock, (2) it has to be present on the first and last TAQ master file for the year, (3) it has to have the NYSE, AMEX or NASDAQ as the primary listing exchange, (4) it does not change primary exchange, ticker symbol or cusip over the year, and (5) has to be listed in CRSP. We randomly select 400 stocks each year from the universe of eligible stocks in 1993. Rolling forward, if any of the 1993 selections is not eligible in 1994, then we randomly draw a replacement for from the universe of eligible stocks in 1994. We continue rolling forward over a 13 year span. Thus, we have 5,200 stock-years. We use the same selected stocks for the monthly measures. We loose a tiny number of observations in extremely illiquid stocks. Specifically, we loose 30 stock-months because there are insufficient trades (2 or less) on positive-volume days to run the Bayesian regression that is part of the Gibbs measure. This results in 62,100 stock months from TAQ.

Second, we use data that are required to be disclosed under Rule 605 of regulation NMS (formerly regulation 11Ac1-5) from October 2001 to December 2005. The data is collected and manually assembled from the Transaction Auditing Group, Inc (www.tagaudit.com) for October 2001 to December 2005. We used the same stocks as above. Data on NYSE / AMEX firms were taken from their respective market center statistics. Data on NASDAQ firms are aggregated by volume-weighting the disclosed statistics from the following market centers: SOES, all ECNs (Archipelago (ARCA), Instinet (INET), Island (ISLD), NexTrade (NTRD), and Redibook (REDI)), and the top ten NASDAQ market makers[14] (Schwab (SCHB), Brutt (BRUT), Goldman Sachs. (GSCO), Knight (NITE and TRIM), GVR (GVRC), B-Trade (BTRD), Lehman Brothers (LEHM), Credit Suisse First Boston (FBCO), Merrill Lynch (MLCO), and J.P. Morgan (JPMS)).

To compute our low-frequency liquidity measures, we used the Daily Stock database from CRSP over the same time periods. We notice that the analytic-formula proxies (Roll, Effective Tick, Effective Tick2, Zeros, Zeros2, Illiquidity, Gamma, and Liquidity) are fast to compute. By contrast, the single measure, numerically-iterated proxies (Gibbs, LOT Mixed, and LOT Y-split) are slower to compute as is the combination measure, Holden, which is the most computationally intensive. In perspective, all low-frequency proxies, with the exception of the Holden measure, are faster to compute than their high-frequency counterparts.

Table 1 provides some descriptive statistics. Panel A describes monthly effective spread benchmarks and proxies calculated from 1993-2005 TAQ data. The high-frequency benchmark, Effective Spread (TAQ), has a mean of 0.029 and a median of 0.016. Looking across the effective spreads proxies, we see that Roll, Effective Tick, Effective Tick 2, Holden, Gibbs, and LOT Y-split are approximately the same in magnitude as the benchmark. LOT Mixed is approximately double the benchmark. The rest of the low-frequency measures are of completely different orders of magnitude. Panel B describes annual effective spread benchmarks and proxies, where the picture is essentially the same. The same measures in panel B and A are

[14] The top ten list is based on NASDAQ composite volume for the month of March 2004 at www.nasdaqtrader.com.

approximately the same magnitude as the benchmark. Panel C describes monthly effective spread benchmarks and proxies calculated from 10/2001 – 12/2005 Rule 605 data. Effective Spread (605), has a mean of 0.015 and a median of 0.006. Again, the low-frequency proxies have essentially the same magnitude relationships as in Panel A.

Panel D describes monthly price impact benchmarks and proxies calculated from 1993-2005 TAQ data. The high-frequency benchmark, Price Impact (TAQ), has a mean of 130.425 and a median of 15.793, after multiplying by 1,000,000. Looking at the means and medians of the price impact proxies, we see that none of the proxies are of the same order of magnitude as the benchmark. The same holds true in Panel E for annual price impact proxies. Panel F describes monthly price impact benchmarks and proxies calculated from 10/2001 – 12/2005 Rule 605 data. Price Impact (605) has a mean of 1.016 and a median of .326, after multiplying by 1,000,000. Again, none of the price impact proxies are of the same order of magnitude.

Panel G breaks down the firms by exchange. Roughly 68% are listed on NASDAQ, 25% on the NYSE, and the rest on AMEX. This break down is nearly the same as the eligible universe of TAQ and Rule 605 stock symbols.

7. Results

7.1. Monthly Effective Spread Results

Table 2 provides monthly effective spread evidence. It compares effective spread measures calculated from daily prices and volumes each month (e.g., using a maximum of 23 daily prices and volumes per month) with monthly effective spread calculated from the TAQ data (e.g. a volume-weighted average of the effective spread of every trade and corresponding BBO quote over the month).

Panel A reports the average cross-sectional correlation of each low-frequency effective spread measure with the effective spread calculated from TAQ. This is computed, in the spirit of Fama-MacBeth, by: (1) calculating for each month separately, the cross-section correlation across all 400 firms and then (2) calculating the average correlation value over all 156 months. We find that six measures, Effective Tick, Effective Tick2, Holden, Gibbs, LOT Mixed, and LOT Y-split, have average cross-sectional correlations greater than 0.600. The Holden measure has highest average cross-sectional correlation at 0.682.

We test whether the average cross-sectional correlations are different from each other in Tables 2 – 8 by computing a t-test of monthly differences in correlations. The standard errors for the differences are Newy West s described below every month (or year) and then calculate the average t-value across months (or years) in the sample.

Table 2, panel A shows that Gibbs, Holden, Lot Mixed and Lot Y-split correlations are insignificantly different from each other. Said differently, considering the measure with the highest correlation, Holden, we show that Gibbs, LOT Mixed, and LOT Y-split are *inside* of its 95% confidence region.

Next, we form equally-weighted portfolios across all 400 stocks in a given month. Specifically, we compute a portfolio liquidity measure in month i by taking the average of that liquidity measure over all 400 stocks in month i. Panel B reports the time-series correlation over 156 months of each low-frequency effective spread portfolio measure with the effective spread portfolio measure calculated from TAQ Asset pricing researchers may be especially interested in the time-series correlations since so much of asset pricing research involves forming portfolios and exploring co-movement over time. Panel B results may differ from Panel A results, not only because they are computed over the time-series vs. across the cross-section, but also because some measurement error that affects individual stocks may be diversified away in portfolios. Consistent with a diversification effect, we find relatively high time-series correlations. Six measures, Roll, Effective Tick, Effective Tick2, Holden, Gibbs, and LOT Y-split, have time-series correlations greater than 0.900. The Holden measure has the highest time-series correlation at 0.951 and no measure is in its 95% confidence interval. All of the time-series correlations are significantly different from zero.

To look at consistency of the measures, we break down the time-series correlations by sub-periods in Panel C. Specifically, we use the same portfolio liquidity measures as above, but compute time-series correlations for three subperiods that closely correspond to minimum tick-size regimes. The subperiods are 1993-1996, 1997-2000, and 2001-2005, which relate to the minimum tick-size regimes of \$1/8th, \$1/16th, and \$.01, respectively. Consistent with Panel B, the same six measures, Roll, Effective Tick, Effective Tick2, Holden, Gibbs, and LOT Y-split, do consistently well in each subperiod. All six measures have time-series correlations greater than 0.900 in 1993-1996, in the interval [0.663, 0.886] in 1997-2000, and greater than 0.863 in 2001-2005. It is not clear why all six measures did worse during the \$1/16th years. Gibbs has the highest correlation in 1993-1996, Effective Tick is the highest in 1997-2000, and Roll is the highest in 2001-2005.

We form decile portfolios stratified by firm size (market capitalization) and by effective spread to check the robustness of the measures. For firm size, we sort the 400 stocks *each month* by market capitalization, assigning the first 40 stocks with the smallest size to portfolio 1, and so on. Each portfolio is equally weighted. Panel D reports the time-series correlation of size decile portfolios. Four measures do quite well across the decile portfolios. Effective Tick, Effective Tick2, Holden, and LOT Y-split have high and statistically significant time-series correlations overall with mildly lower correlations for larger size portfolios. By contrast, Roll and Gibbs do very poorly with the larger firms in Portfolios 7 – 10. Specifically, they obtain time-series correlations of 0.400 or lower, which appears to be a serious robustness problem. They do much better with the small and medium-size firms in Portfolios 1 – 6. All measures do much worse than their own average with the largest firms in Portfolio 10.

Next, we form decile portfolios stratified by effective spread in the same manner as above, assigning the 40 stocks with the lowest effective spread to portfolio 1, and so on. Each portfolio is equally weighted. Panel E reports the time-series correlations of these decile portfolios. Consistent with Panel D, the same four measures, Effective Tick, Effective Tick2, Holden, and LOT Y-split, do quite well. All four have high and statistically significant time-series correlations overall with mildly lower correlations in lower effective spread portfolios. By contrast, Roll and Gibbs do very poorly in Portfolios 1 – 4. Specifically, they obtain time-series correlations lower than 0.330, which is also a serious robustness problem. Undoubtedly, there is a great deal of overlap between these low effective spread portfolios and the large size portfolios. Roll and Gibbs do far better in Portfolios 6 – 10. Nearly all measures do worse than their own average with the lowest effective spread firms in Portfolio 1. So large sizes and small effective spreads are the most challenging firms to proxy with low-frequency effective spread measures.

Finally, we calculate our second performance metric: the prediction error between the low-frequency effective spread measures and effective spread as calculated from TAQ. Panel F reports two measures of prediction error: (1) mean bias (e.g., the difference between the low-frequency mean and the high-frequency mean) and (2) the root mean squared error. The mean bias is for all 62,100 firm months. The root mean squared error is calculated every month and then averaged over 156 months. We exclude the Zeros, Zeros2, Amihud, Pastor and Stambaugh, and Amivest measures from these tests because they are measured in different units than the effective spread. We find that Roll, Effective Tick, Effective Tick2, Holden, Gibbs, and LOT Y-split have relatively small biases, ranging from -0.0018 to -0.0126, that are significantly different from zero based on a T-test. Roll has the smallest bias. Roll, Effective Tick, Effective Tick2, Gibbs and Holden have relatively low root mean squared errors ranging from 0.0286 to .0321. Holden and Gibbs have the lowest root mean squared error, which are not significantly different from each other based on a paired t-test.

Summarizing the monthly effective spread evidence in table 2, we generally conclude that low frequency measures that are designed to estimate effective spread, do, in fact, provide accurate measures of effective spread computed from TAQ data. These measures are highly correlated at the firm level and portfolio level; and provide low bias and small mean squared error. Not surprisingly, we find that measures intended to capture other features of transactions cost, Amihud, Pastor and Stambaugh, and Amivest, do a poor job of estimating effective spread. We find that using zero returns is inferior to all other measures designed to capture effective spread. While zero returns do show significant correlations and root mean squared error predictions, they provide estimates that are significantly lower than other measures. We think of "winning" as providing high and consistent correlations along with low bias and low root mean squared error. Clearly Holden, Effective Tick, and LOT-Y split fit this definition. Roll and Gibbs do well in many cases, but they are not consistent. Roll and Gibbs have periods of much lower correlation (1997-2000) and sub-samples that are much lower (large cap stocks and low effective spread stocks) than the other measures. Holden, which is a non-linear combination of Roll and

Effective Tick, has the highest correlation in both the cross-sectional and time series tests; and its mean squared error is the lowest. This suggests that there are gains from combining measures.[15]

7.2. Annual Effective Spread Results

Table 3 provides annual effective spread evidence. We evaluate the ability of effective spread measures calculated from daily prices and volumes each year (e.g., using a maximum of 254 daily prices and volumes per year) to capture the salient features of annual effective spreads calculated from the TAQ data (e.g. a volume-weighted average of the effective spread of every trade and corresponding BBO quote over the year).

Panel A reports the average cross-sectional correlation of each low-frequency effective spread measure with the effective spread calculated from TAQ. Again, the average cross-sectional correlation is computed in the spirit of Fama-Macbeth. We find that Effective Tick, Holden, and Gibbs have correlations greater than 0.700 and are statistically significant. Gibbs has the highest correlation at 0.779 and is statistically significantly higher than any other measure. Excellent performance by Gibbs in annual effective spread horseraces is consistent with what Hasbrouck (2006) finds.

Next, we form equally-weighted portfolios across all 400 stocks in given year. Each portfolio liquidity measure for a given year is the simple average of that liquidity measure over all 400 stocks in that year. Panel B reports the time-series correlation over 13 years of each low-frequency effective spread portfolio measure with the effective spread portfolio calculated from TAQ. As before, we find that portfolio correlations are much higher than individual stock correlations. Six measures, Roll, Effective Tick, Effective Tick2, Holden, Gibbs, and LOT Y-split, have time-series correlations greater than 0.930. Gibbs has the highest time-series

[15] See Holden (2006) for more development and testing of this point.

correlation at an impressive 0.991. Gibbs is insignificantly different from Roll. Therefore, Gibbs and Roll are the top leadership group on this performance dimension.

Finally, we pool all 5,200 firm-year observations and calculate the prediction error between the low-frequency annual effective spread measures and annual effective spread as calculated from TAQ. Panel C reports these results. Roll, Effective Tick, Effective Tick2, Holden, Gibbs, and LOT Y-split have a small bias ranging from -0.0064 to 0.0076. All are significantly different from zero. Roll, Effective Tick, Effective Tick2, Holden, Gibbs, and LOT Y-split have low root mean squared errors ranging from 0.016 to 0.045 which are statistically different from zero. Gibbs has the lowest RMSE at 0.016 and is statistically significantly lower than any other measure.

Summarizing the annual effective spread evidence in Table 3, we again generally conclude that low frequency measures that are designed to estimate effective spread provide accurate measures of effective spread computed from TAQ data. These measures are highly correlated at the firm level and portfolio level; and provide low bias and small mean squared error. Not surprisingly, we find that measures intended to capture other features of transactions cost, Amihud, Pastor and Stambaugh, and Amivest, do a poor job of estimating effective spread. Six measures dominate, in the sense of having a high and consistent correlation with a low bias and mean squared error, namely Roll, Effective Tick, Effective Tick2, Holden, Gibbs, and LOT Y-split. Overall, Gibbs has the highest correlation in both the cross-sectional and time series tests; and its bias and mean squared error are the lowest.

7.3. Monthly Price Impact Results

Table 4 provides monthly price impact evidence, comparing price impact measures calculated from daily prices and volumes each month with three monthly price impact benchmarks (5-minute price impact, realized spread, and permanent price impact) calculated from TAQ data.

Panel A reports the average cross-sectional correlation of each low-frequency price impact measure with each price impact benchmark. If we look at the measure with the largest correlation and then consider the measures within its confidence interval we get a picture of the superior measures. Amihud has the highest correlation with the 5-minute price impact at 0.317 and is insignificantly different from Roll Impact, Effective Tick Impact, Effective Tick2 Impact, Holden Impact, Gibbs Impact, LOT Mixed Impact, and LOT Y-split Impact. Therefore, all eight measures are in the top leadership group for this horserace. For the realized spread, Amihud has the highest correlation at 0.551 and is statistically significantly higher than any other measure. Therefore, Amihud wins this horserace, since no other measure is in its confidence interval. For permanent price impact, Effective Tick Impact has the highest average cross-sectional correlation at 0.145 and is insignificantly different from Roll Impact, Effective Tick2 Impact, Holden Impact, Gibbs Impact, LOT Mixed Impact, LOT Y-split Impact, and Amihud. Therefore, all eight measures are in the top leadership group for this horserace.

Next, we form equally-weighted portfolios across all 400 stocks in given month. Panel B reports the time-series correlation over 156 months of each low-frequency price impact portfolio measure with each price impact benchmark portfolio calculated from TAQ. As before, most portfolio correlations are higher than the individual stock correlations. Roll Impact has the highest correlation with 5-minute price impact at 0.562 and is insignificantly different from Gibbs Impact, Holden Impact, Lot Mixed Impact and Zeros Impact. Therefore, these four measures are in the top leadership group for this horserace. Roll Impact has the highest correlation with realized spread at 0.686 and is insignificantly different from Holden Impact, Lot

M Impact, Zeros Impact, Zeros 2 Impact and Amihud. These five measures are in the top leadership group for this horserace. Zeros2 Impact has the highest correlation with permanent price impact at 0.623 and is insignificantly different from Holden Impact, LOT Mixed Impact, and Zeros Impact. Therefore these six measures are the top leadership group for this horserace.

Finally, we pool all 62,100 firm-month observations and calculate the prediction error between the low-frequency monthly price impact measures and monthly price impact as calculated from TAQ. Panel C reports that all of the price impact proxies have a huge bias. Excluding the especially extreme Liquidity measure, all of the other price impact proxies have a bias in the range from -130.604 to -118.214 which compares to a mean value of 130.45 for price impact calculated from TAQ. This huge bias is consistent with the observation from Table 1, Panel D that the mean and median values of all of the price impact proxies are too small by an order of magnitude. None of the price impact proxies even comes close to being on the right scale. The mean squared error is not appropriate if the two variables are in different units as these price impact measures. We omit the RMSE calculation.

The overall Table 4 monthly price impact results are summarized together with the overall Table 5 annual price impact results.

7.4. Annual Price Impact Results

Table 5 provides annual price impact evidence. It compares price impact measures calculated from daily prices and volumes each year with three annual price impact benchmarks (5-minute price impact, realized spread, and permanent price impact) calculated from the TAQ data. In general, the annual correlations are much higher than the monthly correlations.

Panel A reports the average cross-sectional correlations of each low-frequency price impact measure with the three price impact benchmarks calculated from TAQ. For the 5 minute price impact, Effective Tick2 has the highest correlation with 5-minute price impact at 0.687 and is insignificantly different than Holden Impact, LOT Mixed Impact and Amihud. Like the monthly result, Amihud has the highest correlation with annual realized spread at 0.571 which is statistically significantly higher than any other measure.

Surprisingly, all of the price impact measures have statistically significant *negative* correlations with annual permanent price impact. By contrast, the monthly price impact measures in Table 4 are *positively* correlated with monthly permanent price impact. A possible explanation for this result is that the permanent component reverses out over a long enough period of time. This is consistent with the Pastor and Stambaugh (2003) argument that all price impact is temporary over some horizon. This might be an avenue for future research.

Next, we form equally-weighted portfolios across all 400 stocks in given year. Panel B of Table 5 reports the time-series correlation over 13 years of each low-frequency price impact portfolio measure with price impact portfolio calculated from TAQ. We find that price impact portfolio correlations are much higher than the price impact correlations for individual stocks. All nine measures from the new class of price impact measures have relatively high correlations in the range 0.875 to 0.963 and are statistically significant. Overall, Roll Impact has the highest correlation and Holden Impact along with Gibbs Impact are in its confidence interval.

Finally, we pool all 5,198 firm-year observations and calculate the mean prediction error between the low-frequency annual price impact measures and annual price impact as calculated from TAQ. Panel C reports that all of the annual price impact proxies, like the monthly measures, are biased to the point of being poor measures of the magnitude of price impact. Excluding the more extreme Liquidity measure, all of the other price impact proxies have a bias in the range from -70.266 to -57.406, which compares to a mean value of 70.285 for annual price impact calculated from TAQ. Like the monthly case, none of the annual price impact proxies even comes close to being on the right scale. We again do not compute the mean squared error statistics because of this reason.

Summarizing the 5-minute price impact horseraces of both tables 4 and 5, the new class of price impact proxies developed in this paper is generally better than the price impact measure widely used in the literature, namely the Amihud measure. Using an effective spread measure divided by dollar volume gives a correlation coefficient with all measures of price impact that is either better than the Amihud, Gamma and Amivest measures or is statistically insignificant from the Amihud measure. Overall, Holden Impact is the best single measure being in the top leadership group of all correlation tests. None of these measures are very good at estimating the magnitude of 5-minute price impact.

Summarizing the realized spread horseraces of both tables 4 and 5, Amihud is the best at the cross-sectional correlation tests, but the new class of price impact proxies is in the leadership group of the time-series correlation tests. Overall, Amihud is the best single measure being in the top leadership group of three of the four correlation tests and standing by itself in two of them. None of these measures is very good at estimating the magnitude of realized spread.

Summarizing the permanent price impact horseraces of both table 4 and 5, the new class of price impact proxies is the leadership group for monthly tests, but it is surprising that the correlations are negative for annual tests. None of these measures is very good at estimating the magnitude of permanent price impact. Dividing a measure of effective spread by dollar volume is clearly better than simply dividing returns by dollar volume.

7.5. Rule 605 Results

As discussed above, the new Rule 605 data allows us to test the robustness of our previous results by using a completely different high-frequency database. Accordingly, Table 6 presented comparable evidence for the Rule 605 data from October 2001 to December 2005. Panels A, B, and C compare effective spread proxies with effective spread calculated from the Rule 605 data. Panels D, E, and F compares price impact proxies with static price impact calculated from Rule 605 data.

The Rule 605 results, presented in Panel A, are relatively similar to the TAQ results. The same six measures have relatively high joint time-series cross-sectional correlations in nearly the same range as before and are statistically significant. Amihud has the highest correlation 0.533 and the Effective Tick, Holden, and LOT Mixed are in its 95% confidence interval.

The time-series correlations presented in Panel B for the Rule 605 data are also generally similar to the TAQ results. The correlations of the portfolios are much higher than the cross sectional correlations. A top measure for the cross-section, Effective Tick, has the highest correlation and Effective Tick2 along with Zeros are in its 95% confidence interval.

Next, Rule 605 results presented in Panel C on the prediction error are again roughly similar to Table 2. Effective Tick2 has the smallest bias and is statistically significantly smaller than any other measure. Gibbs has the smallest RMSE and is insignificantly different from Effective Tick, Effective Tick 2, and Holden. Summarizing Panels A - C, the monthly Rule 605 effective spreads results show that low frequency measures computed from daily returns are able to capture effective spreads reported by the market centers. Overall, Effective Tick is the best single proxy of Rule 605 effective spread.

In panel D, we present evidence on price impact for the Rule 605 data. Recall that TAQ price impact is calculated from a regression, whereas Rule 605 price impact is calculated as the difference between the effective spreads associated with large and small orders, divided by the difference between large and small order shares. Thus, it is not especially surprising to see a very different set of price impact results, presented in Panel D, for the Rule 605 data in comparison to the TAQ data. Essentially, all of the joint time-series cross-sectional correlations between the price impact proxies and price impact calculated from Rule 605 are insignificantly different from zero. All of the proxies fail to pick up Rule 605 price impact on individual stocks. In Panel E, we get similar results that nothing is significant. Finally, Panel F reports the prediction errors of the price impact proxies with respect to price impact calculated from Rule 605. All of the price impact proxies have a large bias. Summarizing Panels D - F, all of the price impact proxies fail to capture the special Rule 605 version of price impact.

As we described above, the Rule 605 price impact measure is the ratio of two differences. The numerator is the difference in percentage spread of "big" orders less the percentage spread of "small" orders. The denominator is the average trade size of big orders less the average trade size of small orders. The definition of "big" is in the range 2,000 to 9,999 shares and "small" is 100 to 500 shares. The SEC data excludes block trades (trades of more than 10,000 shares) and is averaged over a month. There simply may not be any price impact in these data. Very large trades are more likely to move a stock than smaller trades and taking ratios of aggregate data may obscure whatever price impact there actually is in these data.

Table 6 shows that actual effective spread data reported by the market centers can be accurately estimated using measures computed from daily returns. The table also shows that neither the new price impact measures developed in this paper nor the existing measures used in the literature are significantly associated with whatever price impact is in the Rule 605 data.

7.6. Results By Exchange

For robustness, we explore the degree to which our results vary across exchanges. In Table 7, we break out the monthly effective spread and price impact evidence by exchange, sorting firms into two groups based on NYSE/AMEX and NASDAQ. In Panel A, with respect to joint time-series cross-sectional correlations with effective spreads, the every measure, except Gibbs and Roll,[16] shows a lower correlation for NASDAQ stocks than for NYSE stocks. The largest differences are associated with the Effective Tick and Holden measures where the first digit of the correlation coefficient changes. In contrast, the time series averages show that the measures do better for NASDAQ stocks than NYSE. Finally, the price impact measures are mixed across exchange. The clear conclusion from this table is that the exchange does not matter very much and should not be a factor in developing low-frequency measures of transactions cost.

7.7. Results By Year

Our next robustness check is to explore how our results vary across time. Specifically, Table 8 breaks out the monthly effective spread and price impact evidence by year. Panels A and B report the time variation of cross-sectional correlations and RMSE. In each month there are 400 observations for a correlation and RMSE which are averaged over the year. The two panels tell opposite stories. Panel A shows that the cross-sectional correlations decrease over time for seven measures (Roll, Effective Tick, Effective Tick 2, Holden, Gibbs, LOT Mixed, and LOT Y-split). The decline is strongest during the decimal era (2001-2005). By contrast, the Amihud measure does not decline over time and joins the leadership group in the decimal era only. This result contrasts with Table 2, Panel C result that the $1/8 era and decimal era had very high time-series correlations, but the $1/16 era had somewhat lower time-series correlations. In Panel B, all measures improve in their ability to predict the effective spread. LOT Mixed has an RMSE that

[16] Schultz (2000) estimates the Roll measure using intraday TAQ data. He finds that the intraday Roll measure is a very accurate estimate of effective spread, because various biases in the Roll measure tend to offset each other in his NASDAQ sample.

is 81% more accurate in 2005 than in 1993. The RMSE in the last year is basically the same number for all but Zeros, Zeros2 and price impact measures. The mean squared error is the square of the bias plus the variance of the estimator. The fact that the correlation coefficient has fallen but the errors are smaller is the result of a lower bias and smaller variance of the measure.

In Panels C, D and E we present the average correlations between the price impact measure and the three high frequency measures of price impact used in this paper. Generally, the measures are statistically significant in all tables and demonstrate considerable volatility in Panel C (5 min price impact), stability in Panel D (realized spread) and deterioration in Panel E (perm. price impact).

7.8. Dow Jones Data

Our final robustness test is to test the effective spread measures out-of-sample. We examine the stocks in the Dow Jones Industrial Average from 1962 to 2000[17]. The "high frequency" measure of transactions cost is the average quoted bid-ask spread constructed by Jones. For every year we compute each of the low frequency measures for each of the 30 Dow stocks and then equally weight the measures across stocks for the year since the historical spreads for the Dow stocks are available only on an annual basis. Table 9 shows the results. The biggest surprise is the large negative and significantly negative correlation coefficients of the Roll and Gibbs measures. The Roll time-series correlation is -0.642 and the Gibb time-series correlation is -0.395. Of course, the Dow Jones stock are large capitalization stocks with low effective spreads. In that respect, the poor *annual* performance of Roll and Gibbs with the Dow Jones stocks is very consistent with the poor *monthly* performance of Roll and Gibbs with large capitalization deciles and low effective spread deciles in Table 2, panels D and E.

[17] We thank Charles Jones for these data.

As a double-check on this result, we estimated the average autocovariance of daily price changes for each stock. Whenever we have positive autocovariance we change it to a zero value which is exactly the way we constructed the Roll measure. We then correlated the average absolute value of the autocovariance with the spread and found there is a -55% correlation. Thus, in this sample of large, liquid stocks, the lower the spread the higher the absolute value of the autocovariance. This is the opposite relationship supposed by Roll who argued that liquid stocks should have lower autocovariance than illiquid stocks.

For the other measures in Table 9, the correlations between the average measure and the average quoted spread are generally smaller than the time series portfolio correlations of Table 3 panel B, but they are still large and significant. Effective Tick, Effective Tick2, and Holden all have time-series correlations greater than 0.840 and are statistically insignificantly different from each other. Figure 1 shows the time series for the quoted spread of the Dow Jones portfolio and the low frequency measures Holden, LOT Y-split and Effective Tick. These data generated the correlations of table 9. The low frequency measures track the quoted spread very well especially at the end of the sample. The conclusion of table 9 and Figure 1 is that the measures are useful on a different sample of stocks over a different time period.

8. Discussion of Results and Conclusion

The purpose of this paper is to test the hypothesis that low-frequency measures of transactions cost, measured *monthly* and *annually*, can usefully estimate high frequency measures and if so, which measures are the best. We compare all prior proxies and develop three new measures of effective spread and nine new measures of price impact. We use a sample of 400 randomly selected stocks (because of the enormous size of the high-frequency datasets) over the period 1993 to 2005. We compute the effective spread and several measures of price impact from two high-frequency datasets: TAQ and Rule 605 data required to be disclosed by market centers by the SEC. We then compute the low-frequency measures from daily return and volume available on CRSP on a monthly and annual basis.

The evidence is overwhelming that both monthly and annual low-frequency measures usefully capture high-frequency measures of transactions costs. In many applications the correlations are high enough and the mean-squared error low enough, so that the effort of using high frequency measures is simply not worth the cost. The only real question of this paper is which measure should a researcher use? The answer depends on what, exactly, the researcher wants to measure.

For monthly effective spread, we find three measures dominate the remaining nine in correlations and mean squared prediction error. The simplest of the dominant measures is the analytic "Effective Tick." The most computationally intensive is the "Holden" measure. Intermediate in computational requirements is LOT Y-split. The Holden measure does better than the others but all provide statistically significant and useful measures, high correlations and low root mean squared error, regardless of the database we used. For annual effective spread, we find that six low-frequency proxies (Roll, Effective Tick, Effective Tick2, Holden, Gibbs, and LOT Y-split) dominate; with Gibbs being the best overall. For the 5-minute price impact horseraces, the new class of price impact proxies that we develop dominates the measures in the existing literature; with Holden Impact being the best overall. For the realized spread horseraces, the Amihud measure is the best overall. For the permanent price impact horseraces, there are mixed results We find that no low frequency measure is very good at estimated the magnitude of any version of price impact.

For the monthly Rule 605 effective spreads horserace, Effective Tick is the best overall. All of the price impact proxies fail to capture the special Rule 605 version of price impact.

We conduct several robustness checks on these conclusions. First, we examine the pattern of these measures over time. Second, we examine whether a NYSE or NASDAQ listing matters. Finally we, test the ability of these measures to predict the average spread on Dow stocks going from 1962 to 2000. The conclusions are generally true in these tests. The measures vary over time in their ability to capture high-frequency measures but the dominant measures over time are the same group. The exchange listing does not matter and the low frequency measures do well in predicting the quoted spreads on Dow stocks.

As with any empirical paper several caveats should be mentioned. First, using a random sample in this paper means that caution should be used in applying these measures to other samples or other time periods. Second, we do not know whether the measures are effective on international data especially to those stocks with extremely thin trading. Both limitations suggest avenues for future research. With these limitations in mind, we think the results of this paper are strong enough so that using the low-frequency proxies to extend asset pricing, market efficiency, and corporate finance research back in time and around the world is a step that the finance literature needs to take.

References

Acharya, V., Pedersen L., 2004. Asset pricing with liquidity risk. Journal of Financial Economics, Article in press.

Amihud, Y., 2002. Illiquidity and Stock Returns: Cross section and time-series effects. Journal of Financial Markets 5, 31-56.

Amihud, Y., Mendelson H., 1986, Asset Pricing and the Bid-Ask Spread, Journal of Financial Economics 17, 223-249.

Amihud, Y., Mendelson H., Lauterbach B., 1997. Market microstructure and securities values: evidence from the Tel Aviv Exchange. Journal of Financial Economics 45, 365-390.

Blalock, H., 1972. Social Statistics. McGraw-Hill, New York.

Bekaert, G., Harvey C., Lundblad C., 2005. Liquidity and expected returns: lessons from Emerging markets. Unpublished working paper. Indiana University.

Berkman, H., Eleswarapu V., 1998. Short-term traders and liquidity: a test using Bombay Stock Exchange data. Journal of Financial Economics 47, 339-355.

Boehmer, E., Jennings R., Wei L., 2003. Public disclosure and private decisions: the case of equity market execution quality. Unpublished working paper. Indiana University.

Cao, C., Field, L., Hanka, G., 2004. Does insider trading impair market liquidity? Evidence from IPO lockup expirations. Journal of Financial and Quantitative Analysis 39, 25-46.

Chan, L.K.C., Jegadeesh N., Lakonishok J., 1996. Momentum strategies. Journal of Finance 51, 1681-1713.

Chordia, T., Roll R., Subrahmanyam A., 2000. Commonality in liquidity. Journal of Financial Economics, 56, 3-28.

Christie, W., Schultz, P., 1994. Why Do NASDAQ Market Makers Avoid Odd-Eighth

Quotes? Journal of Finance 49, 1813-1840.

Cooper, S., Groth K., Avera W., 1985. Liquidity, exchange listing and common stock Performance. Journal of Economics and Business 37, 19-33.

De Bondt, W. Thaler R., 1985. Does the stock market overreact? Journal of Finance 40, 793–805.

Dennis, P., Strickland, D., 2003. The effect of stock splits on liquidity: evidence from shareholder ownership composition. Journal of Financial Research 26, 355-370.

Fujimoto**,** A., 2003. Macroeconomic sources of systematic liquidity. Unpublished working paper. Yale University.

Fujimoto, A., Watanabe, M., 2006. Time-Varying Liquidity Risk And The Cross Section Of Stock Returns. Unpublished working paper. Rice University.

Geweke, J., 1996. Monte Carlo Simulation and Numerical Integration. In: Amman H. et al. (Eds), Handbook of Computational Economics, Vol. 1. Elsevier, Amsterdam.

Goyenko, R., 2006. Stock And Bond Pricing With Liquidity Risk. Unpublished working paper. Indiana University.

Harris, L., 1991. Stock Price Clustering and Discreteness. Review of Financial Studies 4, 389-415.

Hasbrouck, J., 1999a. The dynamics of discrete bid and ask quotes. Journal of Finance 54, 2109-2142.

Hasbrouck, J., 1999b. Security bid/ask dynamics with discreteness and clustering: simple strategies for modeling and estimation. Journal of Financial Markets 2, 1-28.

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., 2006. Trading costs and returns for US equities: the evidence from daily data. Unpublished working paper. New York University.

Helfin, F., Shaw, K., 2000. Blockholder ownership and market liquidity. Journal of Financial and Quantitative Analysis 35, 621-633.

Henker, T., Wang, J., 2006. On the importance of timing specifications in market microstructure research. Journal of Financial Markets 9, 162-179.

Holden, C., 2007. New low-frequency liquidity measures. Unpublished working paper. Indiana University.

Huang, R., 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.

Huang, R., Stoll, H., 1997. The components of the bid-ask spread: A general approach. Review of Financial Studies 10, 995-1034.

Jegadeesh, N., Titman S., 1993. Returns to buying winners and selling losers: implications for market efficiency. Journal of Finance 48, 65–92.

Jegadeesh, N., Titman, S., 2001. Profitability of momentum strategies: an evaluation of alternative explanations. Journal of Finance 56, 699–720.

Judd, K., 1998. Numerical Methods in Economics. MIT Press, Cambridge, Massachusetts.

Jones, C., 2002. A Century of Stock Market Liquidity and Trading Costs. Unpublished working paper. Columbia University.

Kalev, P., Pham, P., Steen, A., 2003. Underpricing, stock allocation, ownership structure and post-listing liquidity of newly listed firms. Journal of Banking and Finance 27, 919-947.

Korajczyk R., Sadka R., 2006. Pricing The Commonality Across Alternative Measures Of Liquidity. Unpublished working paper. Northwestern University.

Lee C., Radhakrishna, B., 2000. Inferring investor behavior: Evidence from TORQ data. Journal of Financial Markets 3, 83-112.

Lee C., Ready M., 1991. Inferring trade direction from intraday data. Journal of Finance 46, 733-746.

Lerner, J., Schoar A., 2004. The illiquidity puzzle: theory and evidence from private equity. Journal of Financial Economics 72, 3-40.

Lesmond, D., 2004. Liquidity of emerging markets. forthcoming in the Journal of Financial Economics.

Lesmond, D., Lemma S., O'Connor P., 2005. Leverage recapitalizations and liquidity. Unpublished working paper. Tulane University.

Lesmond, D., Ogden J., Trzcinka C., 1999. A new estimate of transaction costs. Review of Financial Studies 12, 1113-1141.

Lipson, M., Mortal S., 2004a. Liquidity and firm characteristics: evidence from mergers and acquisitions. Unpublished working paper. University of Georgia.

Lipson, M., Mortal S., 2004b. Capital structure decision and equity market liquidity. Unpublished working paper. University of Georgia.

Pastor, L., 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.

Rouwenhorst, G., 1998. International momentum strategies. Journal of Finance 53, 267–84.

Sadka, R., 2003. Liquidity risk and asset pricing. Unpublished working paper. Northwestern University.

Schrand, C., Verrecchia, R., 2004. Disclosure choice and cost of capital: evidence from underpricing in initial public offerings. Unpublished working paper. University of Pennsylvania.

Swinscow, T., 1997. Statistics at Square One, 9th ed. BMJ Publishing Group, London.

Theil, H., 1966. Applied economic forecasts. North Holland, Amsterdam

Table 1

Descriptive Statistics

The benchmarks Effective Spread (TAQ), 5-Minute Price Impact (TAQ), Realized Spread (TAQ), and Permanent Price Impact (TAQ) are calculated from every trade and corresponding BBO quote in TAQ for a given firm-month or firm-year. The benchmarks Effective Spread (605) and Static Price Impact (605) are calculated from data required to be disclosed under SEC Rule 605 (formerly 11Ac1-5) for a given firm-month or firm-year. All Effective Spread Proxies and Price Impact Proxies are calculated from daily price and volume data for a given firm-month or firm-year. The Effective Spread Proxies are: Roll from Roll (1984), Effective Tick and Effective Tick 2 from this paper and Holden (2006), Holden from Holden (2006), Gibbs from Hasbrouck (2004), LOT Mixed, Zeros, and Zeros2 from Lesmond, Odgen, and Trzcinka (1999), LOT Y-split from this paper, Amihud from Amihud (2002), Pastor and Stambaugh from Pastor and Stambaugh (2002), and the Amivest Liquidity ratio. The Price Impact Proxies are: Roll Impact, Effective Tick Impact, Effective Tick2 Impact, Holden Impact, Gibbs Impact, LOT Mixed Impact, LOT Y-split Impact, Zeros Impact, and Zeros2 Impact from this paper, Amihud from Amihud (2002), Pastor and Stambaugh from Pastor and Stambaugh (2002), and the Amivest Liquidity ratio.

	Effective Spread Benchmarks		Effective Spread Proxies								
	Effective Spread (TAQ)	Effective Spread (605)	Roll	Effective Tick	Effective Tick2	Holden	Gibbs	LOT Mixed	LOT Y-split	Zeros	Zeros2
Panel A: Monthly, 1993-2005, using a TAQ Benchmark											
Average	0.029	-	0.027	0.017	0.016	0.018	0.018	0.056	0.023	0.143	0.127
Std Dev	0.040	-	0.037	0.032	0.030	0.030	0.021	0.089	0.051	0.147	0.130
Min	0.000	-	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Median	0.016	-	0.016	0.008	0.007	0.009	0.012	0.031	0.009	0.095	0.095
Max	1.605	-	0.906	0.929	0.949	0.917	0.673	1.000	1.000	0.909	0.909
Panel B: Annual, 1993-2005, using a TAQ Benchmark											
Average	0.019	-	0.025	0.013	0.013	0.014	0.014	0.074	0.027	0.145	0.128
Std Dev	0.028	-	0.032	0.019	0.018	0.019	0.018	0.117	0.061	0.126	0.101
Min	0.000	-	0.000	0.000	0.000	0.000	0.001	0.000	0.000	0.000	0.000
Median	0.011	-	0.016	0.007	0.007	0.008	0.007	0.039	0.011	0.115	0.109
Max	0.620	-	0.327	0.289	0.340	0.269	0.190	1.787	1.119	0.917	0.653
Panel C: Monthly, 10/2001-12/2005, using a 605 Benchmark											
Average	0.018	0.015	0.019	0.006	0.005	0.007	0.013	0.025	0.006	0.049	0.046
Std Dev	0.026	0.033	0.028	0.015	0.014	0.014	0.015	0.040	0.018	0.073	0.069
Min	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Median	0.009	0.006	0.012	0.002	0.002	0.003	0.009	0.014	0.000	0.000	0.000
Max	0.503	0.948	0.906	0.425	0.447	0.482	0.393	1.000	0.581	0.667	0.667

	Price Impact Benchmarks				Price Impact Proxies											
	5-Min Price Impact (TAQ)	Realized Spread (TAQ)	Perm Price Impact (TAQ)	Static Price Impact (605)	Roll Impact	Effective Tick Impact	Effective Tick2 Impact	Holden Impact	Gibbs Impact	LOT Mixed Impact	LOT Y-split Impact	Zeros Impact	Zero2 Impact	Amihud	Pastor & Stambaugh	Amivest Liquidity
Panel D: Monthly, 1993-2005, using a TAQ Benchmark																
Average	130.425	0.029	-0.0002	-	3.816	4.587	4.049	4.068	3.626	12.211	9.295	20.917	7.782	6.314	-0.179	639355
Std Dev	2446.202	0.039	0.011	-	57.617	154.809	147.568	93.306	75.851	288.448	284.875	305.990	102.754	91.957	10.129	155561102
Min	-41544.120	0.000	-0.728	-	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	-1508.411	0.000
Median	15.793	0.017	-0.0005	-	0.015	0.020	0.019	0.024	0.029	0.074	0.018	0.202	0.148	0.104	-0.00003	26.622
Max	398507	1.585	0.511	-	6978	32742	32742	16371	11399	42000	42000	38000	21000	14160	798	38762898699
Panel E: Annual, 1993-2005, using a TAQ Benchmark																
Average	70.285	0.030	-0.011	-	2.045	1.569	1.335	1.353	1.486	6.604	4.346	12.879	4.972	6.307	0.018	586002.888
Std Dev	300.430	0.042	0.024	-	17.937	25.274	22.932	13.734	14.257	87.651	70.645	191.552	31.360	46.973	0.292	41202127.133
Min	-10943.480	0.001	-0.772	-	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	-5.598	0.007
Median	15.535	0.018	-0.005	-	0.015	0.015	0.015	0.017	0.014	0.089	0.023	0.237	0.236	0.148	0.0001	36.563
Max	7655.088	1.392	0.133	-	834.616	1644.99	1504.080	581.405	578.151	5381.836	3826.47	11554.83	1424.65	1681.365	8.436	2970331874
Panel F: Monthly, 10/2001-12/2005, using a 605 Benchmark																
Average	-	-	-	1.016	1.600	1.057	0.985	0.875	1.071	2.659	1.213	5.713	2.963	4.046	0.025	2066923
Std Dev	-	-	-	31.278	19.639	28.910	39.373	12.177	15.198	40.269	27.799	125.983	20.595	66.740	3.446	280924448
Min	-	-	-	-1491.101	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	-91.366	0.002
Median	-	-	-	0.326	0.003	0.002	0.002	0.004	0.012	0.013	0.000	0.000	0.000	0.034	-0.00004	94.631
Max	-	-	-	2407.128	1525.001	3590.67	5229.895	699.319	1372.41	3773.920	3255.19	15587.53	894.38	7245.073	408.992	38762898699

* All price impact measures are multiplied by 1,000,000, except for Liquidity which is divided by 1,000,000.

Panel G: Observations Classified By Exchange Listing

Data	Total	NYSE	AMEX	Nasdaq
Monthly TAQ, 1993-2005	62,100	15,536	4,431	42,133
Annual TAQ, 1993-2005	5,198	1,295	370	3,533
Monthly 605, 10/2001-12/2005	19,039	5,167	1,633	12,239

Table 2

Monthly Effective Spread Proxies Compared To A TAQ Benchmark

The benchmark Effective Spread (TAQ) is calculated from every trade and corresponding BBO quote in TAQ for a given firm-month. All Effective Spread Proxies are calculated from daily price and volume data for a given firm-month. The Effective Spread Proxies are: Roll from Roll (1984), Effective Tick and Effective Tick 2 from this paper and Holden (2006), Holden from Holden (2006), Gibbs from Hasbrouck (2004), LOT Mixed, Zeros, and Zeros2 from Lesmond, Odgen, and Trzcinka (1999), LOT Y-split from this paper, Amihud from Amihud (2002), Pastor and Stambaugh from Pastor and Stambaugh (2002), and the Amivest Liquidity ratio. Bold numbers are statistically significant at the 5% level. * means that the statistic is significantly different at the 5% level from all other statistics in the same row.

	Roll	Effective Tick	Effective Tick2	Holden	Gibbs	LOT Mixed	LOT Y-split	Zeros	Zeros2	Amihud	Pastor & Stambaugh	Amivest Liquidity
Panel A: Average Cross-sectional Correlation based on Individual Firms (400 Firms per Cross-sectional Correlation; Average of 156 Months)												
Effective Spread (TAQ)	**0.560**	**0.626**	**0.614**	**0.682**	**0.667**	**0.651**	**0.644**	**0.427**	**0.308**	**0.571**	**-0.118**	**-0.136**
Insignificantly different from	Eff Tick Eff Tick2 Amihud	Roll Eff Tick2 LOT Mix LOT Y Amihud	Roll Eff Tick Gibbs LOT Mix LOT Y Amihud	Gibbs LOT Mix LOT Y	Holden	Eff Tick Eff Tick2 Holden LOT Y	Eff Tick Eff Tick2 Holden LOT Mix	*	*	Roll Eff Tick Eff Tick2	Amivest	Pas/Stam
Panel B: Time-series Correlations based on an Equally-weighted Portfolio in Each Month (156 Months)												
Effective Spread (TAQ)	**0.925**	**0.941**	**0.939**	**0.951**	**0.905**	**0.722**	**0.931**	**0.874**	**0.860**	**0.608**	**-0.366**	**-0.145**
Insignificantly different from	Eff Tick Eff Tick2 LOT Y	Eff Tick2 Roll	Eff Tick Roll	*	*	*	Roll	*	*	*	*	*
Panel C: Pure Time-series Correlations based on Equally-weighted Portfolio by Sub-periods (48, 48, and 60 Months, Respectively)												
1993 - 1996 Effective Spread (TAQ)	**0.901**	**0.918**	**0.930**	**0.932**	**0.936**	**0.407**	**0.909**	**0.769**	**0.706**	**0.476**	**-0.364**	-0.160
1997 - 2000 Effective Spread (TAQ)	**0.703**	**0.886**	**0.885**	**0.882**	**0.663**	**0.306**	**0.812**	**0.304**	0.227	**0.539**	-0.226	-0.045
2001 - 2005 Effective Spread (TAQ)	**0.933**	**0.887**	**0.884**	**0.904**	**0.913**	**0.896**	**0.863**	**0.730**	**0.665**	**0.833**	**-0.215**	-0.183
Panel D: Pure Time-series Correlations based on Equally-weighted Decile Portfolios stratified by Firm Size (156 Months)												
Portfolio 1: Smallest Size (Eff.Spread TAQ)	**0.924**	**0.930**	**0.916**	**0.930**	**0.958**	**0.842**	**0.921**	**0.823**	**0.747**	**0.514**	**-0.292**	**-0.169**
Portfolio 2	**0.920**	**0.890**	**0.887**	**0.909**	**0.944**	**0.752**	**0.891**	**0.774**	**0.703**	**0.822**	**-0.399**	-0.022
Portfolio 3	**0.904**	**0.912**	**0.912**	**0.923**	**0.930**	**0.732**	**0.904**	**0.837**	**0.824**	**0.845**	**-0.273**	-0.116
Portfolio 4	**0.863**	**0.861**	**0.857**	**0.870**	**0.852**	**0.505**	**0.844**	**0.789**	**0.750**	**0.732**	**-0.251**	**-0.782**
Portfolio 5	**0.806**	**0.862**	**0.858**	**0.889**	**0.789**	**0.662**	**0.848**	**0.809**	**0.791**	**0.775**	**-0.210**	0.068
Portfolio 6	**0.767**	**0.903**	**0.899**	**0.928**	**0.686**	**0.590**	**0.895**	**0.832**	**0.819**	**0.823**	**-0.182**	**-0.208**
Portfolio 7	**0.367**	**0.897**	**0.897**	**0.920**	**0.400**	**0.511**	**0.867**	**0.800**	**0.801**	**0.806**	0.130	**-0.148**
Portfolio 8	0.046	**0.705**	**0.706**	**0.712**	**0.204**	**0.435**	**0.683**	**0.628**	**0.625**	**0.639**	**-0.276**	**-0.159**
Portfolio 9	0.126	**0.886**	**0.884**	**0.908**	**0.202**	**0.550**	**0.866**	**0.771**	**0.765**	**0.492**	0.051	**-0.138**
Portfolio 10 Largest Size (Eff. Spread TAQ)	**0.362**	**0.591**	**0.590**	**0.627**	**0.381**	**0.415**	**0.545**	**0.436**	**0.433**	**0.188**	-0.113	-0.130
Average	0.608	0.844	0.841	0.862	0.635	0.599	0.826	0.750	0.726	0.664	-0.182	-0.180

	Roll	Effective Tick	Effective Tick2	Holden	Gibbs	LOT Mixed	LOT Y-split	Zeros	Zeros2	Amihud	Pastor & Stambaugh	Amivest Liquidity
Panel E: Pure Time-series Correlations based on Equally-weighted Decile Portfolios stratified by Effective Spread (156 Months)												
Portfolio 1 Eff.Spread (TAQ) (Low E.S.)	0.097	**0.880**	**0.886**	**0.915**	**0.197**	**0.530**	**0.834**	**0.732**	**0.733**	**0.360**	**0.137**	**-0.134**
Portfolio 2 Eff. Spread (TAQ)	0.018	**0.924**	**0.924**	**0.942**	**0.191**	**0.512**	**0.864**	**0.804**	**0.805**	**0.582**	0.040	**-0.216**
Portfolio 3 Eff. Spread (TAQ)	0.092	**0.944**	**0.943**	**0.957**	**0.150**	**0.630**	**0.919**	**0.864**	**0.861**	**0.691**	0.061	**-0.166**
Portfolio 4 Eff. Spread (TAQ)	**0.299**	**0.938**	**0.935**	**0.949**	**0.322**	**0.491**	**0.909**	**0.888**	**0.883**	**0.639**	**0.158**	**-0.557**
Portfolio 5 Eff. Spread (TAQ)	**0.621**	**0.934**	**0.930**	**0.948**	**0.623**	**0.567**	**0.894**	**0.877**	**0.869**	**0.523**	0.109	0.040
Portfolio 6 Eff. Spread (TAQ)	**0.797**	**0.930**	**0.926**	**0.945**	**0.738**	**0.644**	**0.914**	**0.870**	**0.863**	**0.754**	0.018	**-0.214**
Portfolio 7 Eff. Spread (TAQ)	**0.890**	**0.921**	**0.915**	**0.934**	**0.862**	**0.693**	**0.907**	**0.837**	**0.801**	**0.768**	**-0.144**	-0.119
Portfolio 8 Eff. Spread (TAQ)	**0.917**	**0.906**	**0.908**	**0.930**	**0.918**	**0.623**	**0.906**	**0.821**	**0.784**	**0.783**	0.035	**-0.137**
Portfolio 9 Eff. Spread (TAQ)	**0.911**	**0.897**	**0.903**	**0.914**	**0.942**	**0.719**	**0.906**	**0.821**	**0.763**	**0.776**	**-0.200**	-0.117
Portfolio 10 Eff. Spread (TAQ) (High E.S.)	**0.928**	**0.919**	**0.911**	**0.924**	**0.956**	**0.843**	**0.909**	**0.820**	**0.753**	**0.556**	**-0.366**	**-0.212**
Average	0.557	0.919	0.918	0.936	0.590	0.625	0.896	0.833	0.811	0.643	-0.015	-0.183
Panel F: Prediction Error of Effective Spread (TAQ)												
Mean Bias (based on 62,100 firm/months)	**-0.0018**	**-0.0117**	**-0.0126**	**-0.0107**	**-0.0102**	**0.0275**	**-0.0053**	na	na	na	na	na
Insignificantly different from	*	*	*	*	*	*	*					
Root Mean Squared Error (ave. of 156 mon.)	**0.0321**	**0.0311**	**0.0316**	**0.0286**	**0.0287**	**0.0606**	**0.0342**	na	na	na	na	na
Insignificantly different from	Eff Tick Eff Tick2	Roll Eff Tick2	Roll Eff Tick	Gibbs	Holden	*	*					

Table 3

Annual Effective Spread Proxies Compared To A TAQ Benchmark

The benchmark Effective Spread (TAQ) is calculated from every trade and corresponding BBO quote in TAQ for a given firm-year. All Effective Spread Proxies are calculated from daily price and volume data for a given firm-year. The Effective Spread Proxies are: Roll from Roll (1984), Effective Tick and Effective Tick 2 from this paper and Holden (2006), Holden from Holden (2006), Gibbs from Hasbrouck (2004), LOT Mixed, Zeros, and Zeros2 from Lesmond, Odgen, and Trzcinka (1999), LOT Y-split from this paper, Amihud from Amihud (2002), Pastor and Stambaugh from Pastor and Stambaugh (2002), and the Amivest Liquidity ratio. Bold numbers are statistically significant at the 5% level. * means that the statistic is significantly different at the 5% level from all other statistics in the same row.

	Roll	Effective Tick	Effective Tick2	Holden	Gibbs	LOT Mixed	LOT Y-split	Zeros	Zeros2	Amihud	Pastor & Stambaugh	Amivest Liquidity
Panel A: Average Cross-sectional Correlation based on Individual Firms (400 Firms per Cross-sectional Correlation; Average of 13 Years)												
Effective Spread (TAQ)	**0.623**	**0.707**	**0.650**	**0.745**	**0.779**	**0.545**	**0.535**	**0.590**	**0.481**	**0.594**	**0.152**	**-0.120**
Insignificantly different from	Eff Tick2 LOT Y Zeros Amihud	*	Roll Zeros Amihud	*	*	Zeros Amihud	Roll Zeros Zeros2 Amihud	Roll Eff Tick2 LOT Mix LOT Y Amihud	LOT Y	Roll Eff Tick2 LOT Mix LOT Y Zeros	*	*
Panel B: Pure Time-series Correlations based on an Equally-weighted Portfolio (13 Years)												
Effective Spread (TAQ)	**0.982**	**0.945**	**0.954**	**0.966**	**0.991**	**0.821**	**0.934**	**0.919**	**0.909**	**0.797**	0.193	**-0.490**
Insignificantly different from	Eff Tick2 Holden Gibbs	Eff Tick2 LOT Y	Roll Eff Tick Holden LOT Y	Roll Eff Tick2	Roll	Zeros Zeros2 Amihud	Eff Tick Eff Tick2 Zeros Zeros2	LOT Mix LOT Y	LOT Mix LOT Y	LOT Mix	*	*
Panel C: Prediction Error of Effective Spread (TAQ) based on Individual Firms (5,198 Firm-Years)												
Mean Bias	**0.0055**	**-0.0059**	**-0.0064**	**-0.0054**	**-0.0054**	**0.0544**	**0.0076**	**na**	**na**	**na**	**na**	**na**
Insignificantly different from	*	*	*	Gibbs	Holden	*	*					
Root Mean Square Error	**0.0244**	**0.0184**	**0.0198**	**0.0174**	**0.0160**	**0.0924**	**0.0445**	**na**	**na**	**na**	**na**	**na**
Insignificantly different from	*	*	*	*	*	*	*					

Table 4
Monthly Price Impact Proxies Compared To A TAQ Benchmarks
The benchmarks 5-Minute Price Impact (TAQ), Realized Spread (TAQ), and Permanent Price Impact (TAQ) are calculated from every trade and corresponding BBO quote in TAQ for a given firm-month. All Price Impact Proxies are calculated from daily price and volume data for a given firm-month. The Price Impact Proxies are: Roll Impact, Effective Tick Impact, Effective Tick2 Impact, Holden Impact, Gibbs Impact, LOT Mixed Impact, LOT Y-split Impact, Zeros Impact, and Zeros2 Impact from this paper, Amihud from Amihud (2002), Pastor and Stambaugh from Pastor and Stambaugh (2002), and the Amivest Liquidity ratio. Bold numbers are statistically significant at the 5% level. * means that the statistic is significantly different at the 5% level from all other statistic in the same row.

	Roll Impact	Effective Tick Impact	Effective Tick2 Impact	Holden Impact	Gibbs Impact	LOT Mixed Impact	LOT Y-split Impact	Zeros Impact	Zero2 Impact	Amihud	Pastor & Stambaugh	Amivest Liquidity
Panel A: Average Cross-sectional Correlation based on Individual Firms (400 Firms per Cross-sectional Correlation; Average of 156 Months)												
5-Min Price Impact (TAQ)	**0.288**	**0.296**	**0.305**	**0.305**	**0.296**	**0.309**	**0.294**	**0.278**	**0.259**	**0.317**	-0.064	-0.030
Insignificantly different from	All except Zeros2 Imp Pas/Stam Amivest	All except Zeros2 Imp Pas/Stam Amivest	All except Zeros2 Imp Pas/Stam Amivest	All except Zeros2 Imp Pas/Stam Amivest	All except Zeros Imp Zeros2 Imp Pas/Stam Amivest	All except LOT Y Imp Zeros Imp Zeros2 Imp Pas/Stam Amivest	All except Zeros2 Imp Pas/Stam Amivest	All except Gibbs Imp LOT Mix Imp Pas/Stam Amivest	Zeros Imp	All except Zeros2 Imp Pas/Stam Amivest	Amivest	Pas/Stam
Realized Spread (TAQ)	**0.500**	**0.423**	**0.441**	**0.444**	**0.473**	**0.460**	**0.412**	**0.355**	**0.386**	**0.551**	**-0.114**	**-0.132**
Insignificantly different from	LOT Mix Imp	Eff Tick Imp Holden Imp Gibbs Imp LOT Mix Imp LOT Y Imp Zero2 Imp	Eff Tick Imp Holden Imp Gibbs Imp LOT Mix Imp LOT Y Imp Zero2 Imp	Eff. Tick Imp Eff Tick2 Imp Gibbs Imp LOT Mix Imp LOT Y Imp Zeros 2 Imp	Eff. Tick Imp Eff Tick2 Imp Holden Imp LOT Mix Imp	Roll Imp Eff Tick Imp Eff Tick2 Imp Holden Imp Gibbs Imp	Eff Tick Imp Eff Tick2 Imp Holden Imp Zeroz 2 Imp	Zeros 2 Imp	Eff Tick Imp Eff Tick2 Imp Holden Imp LOT Y Imp Zeros Imp	*	Amivest	Pas/Stam
Permanent Price Impact (TAQ)	**0.132**	**0.145**	**0.144**	**0.142**	**0.126**	**0.141**	**0.132**	**0.131**	**0.136**	**0.135**	-0.031	-0.036
Insignificantly different from	All except Pas/Stam Amivest	All except Pas/Stam Amivest	All except Pas/Stam Amivest	All except Pas/Stam Amivest	All except Pas/Stam Amivest	All except Pas/Stam Amivest	All except Pas/Stam Amivest	All except Pas/Stam Amivest	All except Pas/Stam Amivest	All except Pas/Stam Amivest	Amivest	Pas/Stam
Panel B: Pure Time-series Correlations based on an Equally-weighted Portfolio (156 Months)												
5-Min Price Impact (TAQ)	**0.562**	**0.404**	**0.406**	**0.505**	**0.519**	**0.482**	**0.433**	**0.485**	**0.434**	**0.400**	**-0.192**	-0.062
Insignificantly different from	Holden Imp Gibbs Imp LotM Imp Zeros Imp	Eff Tick2 Imp LOT Mix Imp LOT Y Imp Zero Imp Zero2 Imp	Eff Tick2 Imp LOT Mix Imp LOT Y Imp Zeros Imp Zeros2 Imp	Roll Imp LOT Mix Imp Zeros Imp Zeros2 Imp	Roll Imp Holden Imp LOT Mix Imp Zeros Imp Zeros2 Imp Amihud	Roll Imp Holden Imp Zeros Imp Zeros2 Imp Amihud	Eff Tick Imp Eff Tick2 Imp Zero Imp Zero2 Imp Amihud	All except Pas/Stam Amivest	All except Pas/Stam Amivest	All except Pas/Stam Amivest	*	*
Realized Spread (TAQ)	**0.686**	**0.464**	**0.456**	**0.627**	**0.628**	**0.620**	**0.545**	**0.726**	**0.668**	**0.613**	**-0.362**	**-0.151**
Insignificantly different from	Holden Imp LOT Mix Imp Zeros Imp Zeros2 Imp Amihud	Eff Tick2 Imp	Eff. Tick Imp	Roll Imp Gibbs Imp Zero2 Imp Amihud	Holden Imp Lot M Imp Zero 2 Imp Amihud	Roll Imp Holden Imp Gibbs Imp Zero2 Imp Amihud	Amihud	Roll Imp Zero 2 Imp	Roll Imp Holden Imp Gibbs Imp Lot M Imp Zeros Imp Amihud	Roll Imp Holden Imp Gibbs Imp LOT Mix Imp LOT Y Imp Zeros2 Imp	*	*
Permanent Price Impact (TAQ)	**0.491**	**0.453**	**0.438**	**0.585**	**0.514**	**0.555**	**0.518**	**0.619**	**0.623**	**0.386**	**-0.260**	**-0.066**
Insignificantly different from	All except Zeros Imp Zeros2 Imp Pas/Stam Amivest	Roll Imp Eff Tick2 Imp Gibbs Imp LOT Y Imp Amihud	Roll Imp Eff Tick Imp Gibbs Imp Amihud	Roll Imp Gibbs Imp LOT Mix Imp Zeros Imp Zeros2 Imp	All except Zeros Imp Zeros2 Imp Pas/Stam Amivest	Roll Imp Holden Imp Gibbs Imp Zeros2 Imp	Roll Imp Eff Tick Imp Gibbs Imp Amihud	Holden Imp Zeros2 Imp	Holden Imp LOT Mix Imp Zeros Imp	Roll Imp Eff Tick Imp Eff Tick2 Imp Gibbs Imp LOT Y Imp	Amivest	Pas/Stam

	Roll Impact	Effective Tick Impact	Effective Tick2 Impact	Holden Impact	Gibbs Impact	LOT Mixed Impact	LOT Y-split Impact	Zeros Impact	Zero2 Impact	Amihud	Pastor & Stambaugh	Amivest Liquidity
Panel C: Mean Bias Prediction Error based on Individual Firms (62,100 Firm-Months)*												
5-Min Price Impact (TAQ)	-126.609	-125.838	-126.376	-126.357	-126.799	-118.214	-121.130	-109.508	-122.643	-124.111	-130.604	639,224
Realized Spread (TAQ)	3.788	4.558	4.021	4.039	3.597	12.183	9.266	20.888	7.753	6.286	-0.208	639,355
Permanent Price Impact (TAQ)	3.816	4.587	4.049	4.068	3.626	12.211	9.295	20.917	7.782	6.314	-0.179	639,355

* All price impact measures are multiplied by 1,000,000, except for Liquidity which is divided by 1,000,000.

Table 5

Annual Price Impact Proxies Compared To A TAQ Benchmark

The benchmarks 5-Minute Price Impact (TAQ), Realized Spread (TAQ), and Permanent Price Impact (TAQ) are calculated from every trade and corresponding BBO quote in TAQ for a given firm-month. All Price Impact Proxies are calculated from daily price and volume data for a given firm-month. The Price Impact Proxies are: Roll Impact, Effective Tick Impact, Effective Tick2 Impact, Holden Impact, Gibbs Impact, LOT Mixed Impact, LOT Y-split Impact, Zeros Impact, and Zeros2 Impact from this paper, Amihud from Amihud (2002), Pastor and Stambaugh from Pastor and Stambaugh (2002), and the Amivest Liquidity ratio. Bold numbers are statistically significant at the 5% level. * means that the statistic is significantly different at the 5% level from all other statistic in the same row.

	Roll Impact	Effective Tick Impact	Effective Tick2 Impact	Holden Impact	Gibbs Impact	LOT Mixed Impact	LOT Y-split Impact	Zeros Impact	Zero2 Impact	Amihud	Pastor & Stambaugh	Amivest Liquidity
Panel A: Average Cross-sectional Correlation based on Individual Firms (400 Firms per Cross-sectional Correlation; Average of 13 Years)												
5-Min Price Impact (TAQ)	**0.644**	**0.655**	**0.687**	**0.666**	**0.634**	**0.679**	**0.634**	**0.538**	**0.647**	**0.653**	**0.186**	-0.058
Insignificantly different from	Eff Tick Imp Holden Imp Zeros2 Imp Amihud	Roll Imp Holden Imp Gibbs Imp LOT Mix Imp Zeros2 Imp Amihud	Holden Imp LOT Mix Imp Amihud	Roll Imp Eff Tick Imp Eff Tick2 Imp LOT Mix Imp Zeros2 Amihud	Eff Tick Imp LOT Y Imp Zeros2 Imp Amihud	Eff Tick Imp Eff Tick2 Imp Holden Imp Zeros2 Imp Amihud	Gibbs Imp Zeros2 Imp Amihud	*	Roll Imp Eff Tick Imp Holden Imp Gibbs Imp LOT Mix Imp LOT Y Imp	All except Zeros2 Imp Pas/Stam Amivest	*	*
Realized Spread (TAQ)	**0.501**	**0.434**	**0.459**	**0.468**	**0.479**	**0.490**	**0.430**	**0.357**	**0.445**	**0.571**	**0.179**	**-0.115**
Insignificantly different from	Eff Tick Imp LOT Mix Imp Zeros2 Imp	Gibbs Imp LOT Y Imp Zeros2 Imp	Roll Imp Holden Imp Gibbs Imp LOT Mix Imp LOT Y Imp Zeros2 Imp	Eff Tick2 Imp Gibbs Imp LOT Mix Imp Zeros2 Imp	Eff Tick Imp Eff Tick2 Imp Holden Imp LOT Mix Imp Zeros2 Imp	Roll Imp Eff Tick2 Imp Holden Imp Gibbs Imp Zeros2 Imp	Eff Tick Imp Eff Tick2 Imp Zeros2 Imp	*	Roll Imp Eff Tick Imp Eff Tick2 Imp Holden Imp Gibbs Imp LOT Mix Imp LOT Y Imp	*	*	*
Permanent Price Impact (TAQ)	**-0.270**	**-0.210**	**-0.236**	**-0.227**	**-0.243**	**-0.270**	**-0.222**	**-0.172**	**-0.205**	**-0.332**	**-0.151**	0.066
Insignificantly different from	Eff Tick2 Imp LOT Mix Imp	Holden Imp Gibbs Imp LOT Y Imp Pas/Stam	Roll Imp Holden Imp Gibbs Imp LOT Y Imp Pas/Stam	Eff Tick Imp Eff Tick2 Imp Gibbs Imp LOT Y Imp	Eff Tick Imp Eff Tick2 Imp Holden Imp LOT Y Imp	Roll Imp	Eff Tick Imp Eff Tick2 Imp Holden Imp Gibbs Imp Pas/Stam	Eff Tick Imp Holden Imp LOT Mix Imp Zeros2 Imp Pas/Stam	Eff Tick Imp Eff Tick2 Imp Holden Imp Gibbs Imp LOT Y Imp Zeros Imp Pas/Stam	*	Eff Tick Imp Eff Tick2 Imp LOT Y Im	*

	Roll Impact	Effective Tick Impact	Effective Tick2 Impact	Holden Impact	Gibbs Impact	LOT Mixed Impact	LOT Y-split Impact	Zeros Impact	Zero2 Impact	Amihud	Pastor & Stambaugh	Amivest Liquidity
Panel B: Pure Time-series Correlations based on an Equally-weighted Portfolio (13 Years)												
5-Min Price Impact (TAQ)	**0.963**	**0.876**	**0.888**	**0.928**	**0.949**	**0.875**	**0.891**	**0.881**	**0.933**	**0.914**	0.264	-0.345
Insignificantly different from	Holden Imp Gibbs Imp	Eff Tick2 Imp Holden Imp LOT Mix Imp LOT Y Imp Amihud	Eff Tick Imp Holden Imp LOT Mix Imp LOT Y Imp Amihud	Roll Imp Eff Tick Imp Eff Tick2 Imp Gibbs Imp LOT Mix Imp Amihud	Roll Imp Holden Imp Amihud	Eff Tick Imp Eff Tick2 Imp Holden Imp LOT Y Imp Amihud	Eff Tick Imp Eff Tick2 Imp LOT Mix Imp Amihud	Eff Tick Imp Eff Tick2 Imp LOT Mix Imp LotY Imp Amihud	Eff Tick Imp Eff Tick2 Imp Holden Imp Gibbs Imp LOT Mix Imp Amihud	Eff Tick Imp Eff Tick2 Imp Holden Imp Gibbs Imp LOT Mix Imp LOT Y Imp	Liquidity	Pas/Stam
Realized Spread (TAQ)	**0.904**	**0.864**	**0.863**	**0.892**	**0.894**	**0.855**	**0.868**	**0.879**	**0.921**	**0.829**	0.289	**-0.522**
Insignificantly different from	All except Amihud Pas/Stam Amivest	All except Pas/Stam Amivest	All except Pas/Stam Amivest	All except Pas/Stam Amivest	All except Pas/Stam Amivest	All except Pas/Stam Amivest	All except Pas/Stam Amivest	All except Pas/Stam Amivest	All except LOT Y Imp Pas/Stam Amivest	All except Roll Imp Pas/Stam Amivest	*	*
Permanent Price Impact (TAQ)	-0.427	**-0.513**	**-0.535**	-0.380	-0.369	**-0.533**	-0.440	-0.437	-0.426	**-0.474**	**-0.508**	0.348
Insignificantly different from	Eff Tick2 Imp LOT Mix Imp	Eff Tick2 Imp Holden Imp Gibbs Imp LOT Y Imp Zeros Imp Zeros2 Imp Pas/Stam	Roll Imp Eff Tick Imp Holden Imp Gibbs Imp LOT Mix Imp LOT Y Imp Zeros Imp Pas/Stam	Eff Tick Imp Eff Tick2 Imp Gibbs Imp LOT Y Imp Zero Imp Pas/Stam	Eff Tick Imp Eff Tick2 Imp Holden Imp LOT Y Imp Zeros Imp	Roll Imp Eff Tick2 Imp	Eff Tick Imp Eff Tick2 Imp Holden Imp Gibbs Imp Zeros Imp Pas/Stam	Eff Tick Imp Zeros2 Imp Pas/Stam	All except Roll Imp LOT Mix Imp Amihud Amivest	*	All except Roll Imp Gibb Imp LOT Mix Imp Amihud Amivest	*
Panel C: Mean Bias Prediction Error based on Individual Firms (5,198 Firm-Years)*												
5-Min Price Impact (TAQ)	-68.239	-68.716	-68.950	-68.931	-68.798	-63.681	-65.939	-57.406	-65.312	-63.978	-70.266	585,933
Realized Spread (TAQ)	2.015	1.538	1.305	1.323	1.456	6.573	4.315	12.848	4.942	6.277	-0.012	586,003
Permanent Price Impact (TAQ)	2.057	1.580	1.346	1.364	1.498	6.615	4.357	12.890	4.984	6.318	0.029	586,003

* All price impact measures are multiplied by 1,000,000, except for Liquidity which is divided by 1,000,000.

Table 6

Monthly Effective Spread and Price Impact Proxies Compared To 605 Benchmarks

The benchmarks Effective Spread (605) and Static Price Impact (605) are calculated from data required to be disclosed under SEC Rule 605 (formerly 11Ac1-5) for a given firm-month. All Effective Spread Proxies and Price Impact Proxies are calculated from daily price and volume data for a given firm-month. The Effective Spread Proxies are: Roll from Roll (1984), Effective Tick and Effective Tick 2 from this paper and Holden (2006), Holden from Holden (2006), Gibbs from Hasbrouck (2004), LOT Mixed, Zeros, and Zeros2 from Lesmond, Odgen, and Trzcinka (1999), LOT Y-split from this paper, Amihud from Amihud (2002), Pastor and Stambaugh from Pastor and Stambaugh (2002), and the Amivest Liquidity ratio. The Price Impact Proxies are: Roll Impact, Effective Tick Impact, Effective Tick2 Impact, Holden Impact, Gibbs Impact, LOT Mixed Impact, and LOT Y-split Impact from this paper, Amihud from Amihud (2002), Pastor and Stambaugh from Pastor and Stambaugh (2002), and the Amivest Liquidity ratio.

	Roll	Effective Tick	Effective Tick2	Holden	Gibbs	LOT Mixed	LOT Y-split	Zeros	Zeros2	Amihud	Pastor & Stambaugh	Amivest Liquidity
Panel A: Average Cross-sectional Correlation based on Individual Firms (400 Firms per Cross-sectional Correlation; Average of 51 Months)												
Effective Spread (605)	**0.387**	**0.483**	**0.457**	**0.513**	**0.445**	**0.464**	**0.449**	**0.371**	**0.340**	**0.533**	0.013	-0.075
Insignificantly different from	Eff Tick2 LOT Mix LOT Y Zeros Zeros2	Holden Gibbs LOT Mix LOT Y Amihud	Roll Gibbs LOT Mix LOT Y	Eff Tick Gibbs LOT Mix Amihud	Eff Tick Eff Tick2 Holden LOT Mix LOT Y Zeros	Roll Eff Tick Eff Tick2 Holden Gibbs LOT Y	Roll Eff Tick Eff Tick2 Gibbs LOT Mix	Roll Gibbs Zeros2	Roll Zeros	Eff Tick Holden LOT Mix	Amivest	Pas/Stam
Panel B: Time-series Correlations based on an Equally-weighted Portfolio (51 Months)												
Effective Spread (605)	**0.408**	**0.528**	**0.522**	**0.430**	**0.435**	**0.419**	**0.449**	**0.429**	**0.383**	**0.412**	0.166	-0.136
Insignificantly different from	Holden Gibbs LOT Mix LOT Y Zeros Zeros2 Amihud Pas/Stam	Eff Tick2 Zeros	Eff Tick Zeros	Roll Gibbs LOT Mix LOT Y Zeros Zeros2 Amihud	Roll Holden LOT Mix LOT Y Zeros Zeros2 Amihud	Roll Holden Gibbs LOT Y Zeros Zeros2 Amihud Pas/Stam	Roll Holden Gibbs LOT MIx Zeros Zeros2 Amihud	Roll Eff Tick Eff Tick2 Holden Gibbs LOT Mix LOT Y Amihud Pas/Stam	Roll Holden Gibbs LOT Mix LOT Y Amihud Pas/Stam	Roll Holden Gibbs LOT Mix LOT Y Zeros Zeros2	Roll LOT Mix Zeros Zeros2 Amivest	Pas/Stam
Panel C: Prediction Error of Effective Spread (605) based on Individual Firms (19,039 Firm-Months)												
Mean Bias	**-0.0124**	**0.0217**	**-0.0045**	**-0.0062**	**-0.0123**	**-0.0087**	**0.0074**	**-0.0486**	**-0.0456**	na	na	na
Insignificantly different from	*	*	*	*	*	*	*	*	*			
Root Mean Squared Error	**0.0258**	**0.0243**	**0.0246**	**0.0229**	**0.0222**	**0.0300**	**0.0242**	**0.0721**	**0.0682**	na	na	na
Insignificantly different from	Eff Tick	Roll Gibbs	Gibbs	Gibbs	Eff Tick Eff Tick2 Holden	*	*	*				

	Roll Impact	Effective Tick Impact	Effective Tick2 Impact	Holden Impact	Gibbs Impact	LOT Mixed Impact	LOT Y-split Impact	Zeros Impact	Zeros2 Impact	Amihud	Pastor & Stambaugh	Amivest Liquidity
Panel D: Average Cross-sectional Correlation of Price Impact based on Individual Firms (400 Firms per Cross-sectional Correlation; Average of 51 Months)												
Static Price Impact (605)	0.020	0.024	0.030	0.026	0.031	0.041	0.039	0.069	0.054	0.033	-0.001	-0.022
Insignificantly different from	All except LOT Mix Imp Zeros Imp	All except Zeros2 Imp	All except Zeros2 Imp	All	All except LOT Y Imp Zeros Imp	All except Roll Imp Zeros Imp	All except Gibbs Imp	All except Roll Imp Gibbs Imp LOT Mix Imp Amihud	All except Eff Tick Imp Eff Tick2 Imp	All except Zeros Imp	All	All
Panel E: Pure Time-series Correlations based on an Equally-weighted Portfolio (51 Months)												
Static Price Impact (605)	0.083	0.057	0.057	0.076	0.083	0.087	0.083	0.060	0.033	0.086	0.016	-0.029
Insignificantly different from	All	All	All	All	All	All	All	All	All	All	All	All
Panel F: Mean Bias based on Individual Firms (19,039 Firm-Months)*												
Static Price Impact (605)	0.5840	0.0407	-0.0311	-0.1410	0.0554	1.6425	0.1969	4.6972	1.9465	na	na	na

* All price impact measures are multiplied by 1,000,000, except for Liquidity which is divided by 1,000,000.

Table 7

NYSE/AMEX Vs. NASDAQ Breakdown For Monthly Proxies Compared To TAQ Benchmarks

The benchmarks Effective Spread (TAQ), 5-Minute Price Impact (TAQ), Realized Spread (TAQ), and Permanent Price Impact (TAQ) are calculated from every trade and corresponding BBO quote in TAQ for a given firm-month. All Effective Spread Proxies and Price Impact Proxies are calculated from daily price and volume data for a given firm-month. The Effective Spread Proxies are: Roll from Roll (1984), Effective Tick and Effective Tick 2 from this paper and Holden (2006), and Holden from Holden (2006), Gibbs from Hasbrouck (2004), LOT Mixed, Zeros, and Zeros2 from Lesmond, Odgen, and Trzcinka (1999), LOT Y-split from this paper, Amihud from Amihud (2002), Pastor and Stambaugh from Pastor and Stambaugh (2002), and the Amivest Liquidity ratio. The Price Impact Proxies are: Roll Impact, Effective Tick Impacte, Effective Tick2 Impact, Holden Impact, Gibbs Impact, LOT Mixed Impact, and LOT Y-split Impact from this paper, Amihud from Amihud (2002), Pastor and Stambaughfrom Pastor and Stambaugh (2002), and the Amivest Liquidity ratio.

	Roll	Effective Tick	Effective Tick2	Holden	Gibbs	LOT Mixed	LOT Y-split	Zeros	Zeros2	Amihud	Pastor & Stambaugh	Amivest Liquidity
Panel A: Average Cross-sectional Correlation based on Individual Firms (400 Firms per Cross-sectional Correlation; Average of 156 Months)												
NYSE/AMEX Eff. Spread (TAQ)	**0.503**	**0.745**	**0.752**	**0.782**	**0.638**	**0.705**	**0.690**	**0.507**	**0.472**	**0.633**	-0.006	**-0.176**
Nasdaq Eff. Spread (TAQ)	**0.546**	**0.594**	**0.571**	**0.646**	**0.659**	**0.613**	**0.620**	**0.389**	**0.246**	**0.586**	**-0.124**	**-0.126**
Panel B: Time-series Correlations based on an Equally-weighted Portfolio (156 Months)												
NYSE/AMEX Eff. Spread (TAQ)	**0.770**	**0.617**	**0.616**	**0.654**	**0.810**	**0.428**	**0.541**	**0.144**	0.121	**0.447**	-0.047	-0.106
Nasdaq Eff. Spread (TAQ)	**0.930**	**0.957**	**0.955**	**0.967**	**0.915**	**0.756**	**0.956**	**0.911**	**0.891**	**0.664**	**-0.391**	-0.129

	Roll Impact	Effective Tick Impact	Effective Tick2 Impact	Holden Impact	Gibbs Impact	LOT Mixed Impact	LOT Y-split Impact	Zeros Impact	Zeros2 Impact	Amihud	Pastor & Stambaugh	Amivest Liquidity
Panel C: Ave Cross-sectional Corr based on Individual Firms (400 Firms per Cross-sectional Correlation; Average of 156 Months)												
NYSE/AMEX 5-Min Pri Imp (TAQ)	**0.456**	**0.489**	**0.504**	**0.508**	**0.506**	**0.504**	**0.496**	**0.433**	**0.468**	**0.513**	0.005	-0.065
Nasdaq 5-Min Price Impact (TAQ)	**0.291**	**0.286**	**0.298**	**0.297**	**0.300**	**0.302**	**0.291**	**0.279**	**0.224**	**0.291**	-0.066	-0.028
NYSE/AMEX Rel. Spread (TAQ)	**0.505**	**0.477**	**0.488**	**0.489**	**0.506**	**0.509**	**0.487**	**0.410**	**0.465**	**0.559**	0.006	**-0.167**
Nasdaq Realized Spread (TAQ)	**0.517**	**0.438**	**0.468**	**0.457**	**0.485**	**0.469**	**0.416**	**0.359**	**0.381**	**0.574**	**-0.122**	**-0.118**
NYSE/AMEX Perm Pri Imp (TAQ)	**0.316**	**0.368**	**0.369**	**0.379**	**0.385**	**0.388**	**0.372**	**0.342**	**0.313**	**0.357**	-0.018	**-0.094**
Nasdaq Perm Price Impact (TAQ)	**0.131**	**0.126**	**0.116**	**0.130**	**0.122**	**0.126**	**0.115**	**0.122**	**0.107**	**0.116**	-0.029	-0.045
Panel D: Time-series Correlations based on an Equally-weighted Portfolio (156 Months)												
NYSE/AMEX 5-Min Pri Imp (TAQ)	**0.785**	**0.200**	**0.354**	**0.369**	**0.836**	**0.628**	**0.660**	**0.286**	0.074	**0.190**	**-0.462**	-0.012
Nasdaq 5-Min Price Impact (TAQ)	**0.517**	**0.500**	**0.475**	**0.606**	**0.513**	**0.536**	**0.513**	**0.582**	**0.520**	**0.467**	**-0.169**	-0.069
NYSE/AMEX Rel. Spread (TAQ)	**0.297**	**0.267**	**0.263**	**0.290**	**0.254**	**0.248**	**0.251**	**0.226**	**0.311**	**0.407**	-0.035	-0.085
Nasdaq Realized Spread (TAQ)	**0.743**	**0.473**	**0.466**	**0.649**	**0.669**	**0.642**	**0.578**	**0.759**	**0.647**	**0.671**	**-0.396**	**-0.133**
NYSE/AMEX Perm Pri Imp (TAQ)	0.028	0.097	0.042	**0.208**	0.038	0.122	0.095	0.105	**0.294**	-0.002	-0.023	-0.036
Nasdaq Perm Price Impact (TAQ)	**0.460**	**0.382**	**0.374**	**0.507**	**0.448**	**0.474**	**0.446**	**0.564**	**0.553**	**0.423**	**-0.232**	-0.062

Table 8

Year-By-Year Breakdown For Monthly Proxies Compared To TAQ Benchmarks

The benchmarks Effective Spread (TAQ), 5-Minute Price Impact (TAQ), Realized Spread (TAQ), and Permanent Price Impact Proxies and Price (TAQ) are calculated from every trade and corresponding BBO quote in TAQ for a given firm-month. All Effective Spread and Price Impact Proxies are calculated from daily price and volume data for a given firm-month. The Effective Spread Proxies are: Roll from Roll (1984), Effective Tick, Effective Tick 2 from this paper and Holden (2006), Holden from Holden (2006), Gibbs from Hasbrouck (2004), LOT Mixed, Zeros, and Zeros2 from Lesmond, Odgen, and Trzcinka (1999), LOT Y-split from this paper, Amihud from Amihud (2002), Pastor and Stambaugh from Pastor and Stambaugh (2002), and the Amivest Liquidity ratio. The Price Impact Proxies are: Roll Impact, Effective Tick Impact, Effective Tick2 Impact, Holden Impact, Gibbs Impact, LOT Mixed Impact, and LOT Y-split Impact from this paper, Amihud from Amihud (2002), Pastor and Stambaugh from Pastor and Stambaugh (2002), and the Amivest Liquidity ratio.

	Roll	Effective Tick	Effective Tick2	Holden	Gibbs	LOT Mixed	LOT Y-split	Zeros	Zeros2	Amihud	Pastor & Stambaugh	Amivest Liquidity
Panel A: Average Cross-sectional Correlation based on Individual Firms (400 Firms per Cross-sectional Correlation; 156 Months)												
1993	**0.711**	**0.689**	**0.638**	**0.720**	**0.838**	**0.702**	**0.700**	**0.438**	**0.182**	**0.513**	**-0.306**	**-0.173**
1994	**0.744**	**0.707**	**0.700**	**0.771**	**0.883**	**0.781**	**0.745**	**0.407**	**0.142**	**0.579**	**-0.262**	**-0.193**
1995	**0.714**	**0.647**	**0.637**	**0.741**	**0.845**	**0.718**	**0.718**	**0.401**	**0.202**	**0.614**	**-0.127**	**-0.191**
1996	**0.726**	**0.691**	**0.700**	**0.749**	**0.880**	**0.789**	**0.785**	**0.422**	**0.194**	**0.652**	**-0.256**	**-0.153**
1997	**0.678**	**0.693**	**0.649**	**0.724**	**0.825**	**0.711**	**0.760**	**0.472**	**0.287**	**0.560**	**-0.182**	**-0.162**
1998	**0.554**	**0.696**	**0.693**	**0.757**	**0.675**	**0.691**	**0.708**	**0.464**	**0.314**	**0.650**	-0.003	**-0.148**
1999	**0.550**	**0.663**	**0.667**	**0.724**	**0.681**	**0.660**	**0.657**	**0.452**	**0.374**	**0.531**	**-0.144**	**-0.161**
2000	**0.398**	**0.692**	**0.699**	**0.728**	**0.509**	**0.619**	**0.651**	**0.442**	**0.404**	**0.522**	0.018	**-0.129**
2001	**0.554**	**0.618**	**0.601**	**0.658**	**0.614**	**0.630**	**0.603**	**0.380**	**0.326**	**0.585**	**-0.169**	**-0.106**
2002	**0.528**	**0.575**	**0.553**	**0.650**	**0.670**	**0.658**	**0.604**	**0.429**	**0.387**	**0.565**	-0.070	-0.079
2003	**0.421**	**0.519**	**0.514**	**0.530**	**0.521**	**0.547**	**0.529**	**0.415**	**0.399**	**0.448**	0.000	**-0.106**
2004	**0.358**	**0.506**	**0.495**	**0.566**	**0.400**	**0.479**	**0.458**	**0.399**	**0.391**	**0.606**	0.023	**-0.086**
2005	**0.350**	**0.449**	**0.437**	**0.549**	**0.324**	**0.484**	**0.456**	**0.425**	**0.404**	**0.598**	-0.052	-0.076
Panel B: Root Mean Square Prediction Error of Effective Spread (TAQ) based on Individual Firms (4,777 Firm-Months per year)												
1993	0.045	0.050	**0.053**	0.046	0.046	**0.130**	**0.065**	**0.255**	**0.227**	**226.082**	**23.595**	**1.16E+03**
1994	0.036	0.041	0.041	0.038	0.037	**0.075**	0.045	**0.250**	**0.226**	**141.302**	**12.833**	**9.49E+02**
1995	0.035	0.042	0.042	0.035	0.035	**0.103**	0.048	**0.246**	**0.220**	**42.304**	**12.206**	**1.43E+03**
1996	0.035	0.040	0.039	0.035	0.033	**0.069**	0.040	**0.230**	**0.206**	**41.368**	**6.742**	**2.88E+05**
1997	0.029	0.031	0.033	0.030	0.027	**0.149**	0.034	**0.206**	**0.180**	**39.327**	**12.959**	**2.84E+03**
1998	0.032	0.026	0.025	0.024	0.027	**0.056**	0.028	**0.166**	**0.153**	**28.734**	**2.893**	**4.19E+06**
1999	0.032	0.025	0.024	0.022	0.022	**0.064**	0.031	**0.170**	**0.159**	**17.735**	**1.496**	**4.45E+03**
2000	0.041	0.030	0.030	0.029	0.033	**0.057**	0.034	**0.158**	**0.143**	**14.366**	**0.986**	**8.04E+03**
2001	0.035	0.032	0.033	0.031	0.030	0.048	0.032	**0.118**	**0.102**	**123.135**	**6.549**	**2.46E+04**
2002	0.037	0.036	0.036	0.033	0.030	0.046	0.036	**0.092**	**0.080**	**134.208**	**14.189**	**6.48E+04**
2003	0.029	0.027	0.028	0.027	0.024	0.028	0.027	0.072	0.067	19.217	1.921	3.21E+04
2004	0.020	0.019	0.019	0.018	0.018	0.022	0.019	0.065	0.062	3.315	0.369	1.36E+06
2005	0.020	0.018	0.018	0.016	0.020	0.022	0.017	0.068	0.065	3.415	0.422	5.60E+08

	Roll Impact	Effective Tick Impact	Effective Tick2 Impact	Holden Impact	Gibbs Impact	LOT Mixed Impact	LOT Y-split Impact	Zeros Impact	Zeros2 Impact	Amihud	Pastor & Stambaugh	Amivest Liquidity
Panel C: Ave Cross-sect Corr with **5-Min Price Impact (TAQ)** based on Individual Firms (400 Firms; 156 Months)												
1993	**0.230**	**0.171**	**0.208**	**0.167**	**0.206**	**0.229**	**0.180**	**0.228**	**0.225**	**0.249**	**-0.124**	-0.031
1994	**0.286**	**0.336**	**0.308**	**0.296**	**0.273**	**0.240**	**0.239**	**0.258**	**0.011**	**0.238**	**-0.249**	-0.033
1995	**0.431**	**0.513**	**0.495**	**0.543**	**0.548**	**0.529**	**0.556**	**0.494**	**0.239**	**0.435**	-0.066	-0.032
1996	**0.553**	**0.416**	**0.468**	**0.473**	**0.553**	**0.543**	**0.508**	**0.467**	**0.265**	**0.478**	-0.357	-0.036
1997	**0.158**	**0.225**	**0.203**	**0.225**	**0.175**	**0.139**	**0.135**	**0.185**	**0.131**	**0.271**	**-0.201**	-0.031
1998	**0.450**	**0.477**	**0.510**	**0.487**	**0.429**	**0.443**	**0.410**	**0.460**	**0.423**	**0.405**	**0.082**	-0.033
1999	**0.296**	**0.257**	**0.274**	**0.346**	**0.339**	**0.396**	**0.407**	**0.371**	**0.321**	**0.264**	**0.015**	-0.051
2000	**0.294**	0.297	0.303	0.305	0.283	0.282	0.259	**0.207**	**0.228**	**0.278**	0.030	-0.039
2001	**0.363**	**0.376**	**0.392**	**0.376**	0.409	**0.484**	**0.451**	**0.282**	**0.354**	**0.494**	0.119	-0.030
2002	0.338	0.385	0.383	0.446	0.357	0.363	0.339	0.413	0.432	0.382	-0.090	-0.023
2003	0.149	0.150	0.213	0.140	0.131	0.214	0.211	0.204	0.309	0.183	-0.083	-0.019
2004	**0.093**	**0.074**	**0.045**	**-0.011**	**0.069**	**0.029**	**-0.046**	**-0.087**	**0.212**	**0.272**	**0.080**	**-0.016**
2005	**0.109**	**0.166**	**0.167**	**0.176**	**0.075**	**0.130**	**0.178**	**0.135**	**0.222**	**0.168**	**0.004**	**-0.014**
Panel D: Ave Cross-sect Corr with **Realized Spread (TAQ)** based on Individual Firms (400 Firms; 156 Months)												
1993	**0.520**	**0.428**	**0.408**	**0.458**	**0.499**	**0.477**	**0.455**	**0.424**	**0.391**	**0.494**	**-0.298**	**-0.169**
1994	**0.547**	**0.496**	**0.497**	**0.508**	**0.505**	**0.475**	**0.445**	**0.450**	**0.400**	**0.557**	-0.255	**-0.191**
1995	**0.502**	**0.454**	**0.476**	**0.446**	**0.479**	**0.458**	**0.446**	**0.444**	**0.419**	**0.580**	**-0.124**	**-0.189**
1996	**0.586**	**0.462**	**0.508**	**0.518**	**0.552**	**0.533**	**0.503**	**0.459**	**0.438**	**0.656**	-0.262	**-0.146**
1997	**0.509**	**0.470**	**0.446**	**0.476**	**0.460**	**0.427**	**0.411**	**0.411**	**0.432**	**0.561**	**-0.189**	**-0.156**
1998	**0.574**	**0.492**	**0.494**	**0.501**	**0.543**	**0.513**	**0.465**	**0.440**	**0.421**	**0.642**	0.010	-0.140
1999	**0.524**	**0.471**	**0.496**	**0.504**	**0.498**	**0.500**	**0.484**	**0.371**	**0.453**	**0.535**	-0.146	**-0.150**
2000	**0.416**	**0.352**	**0.377**	**0.358**	**0.364**	**0.381**	**0.326**	**0.192**	**0.340**	**0.498**	0.001	**-0.115**
2001	**0.532**	**0.477**	**0.493**	**0.492**	**0.522**	**0.521**	**0.481**	**0.358**	**0.417**	**0.527**	**-0.143**	-0.107
2002	0.474	0.432	0.455	0.476	0.462	0.453	0.416	0.348	0.416	0.545	-0.044	-0.077
2003	0.371	0.362	0.368	0.374	0.363	0.387	0.336	0.308	0.353	0.424	0.020	-0.108
2004	**0.444**	**0.270**	**0.315**	**0.284**	**0.397**	**0.359**	**0.256**	**0.153**	**0.289**	**0.559**	**0.047**	-0.090
2005	**0.500**	**0.338**	**0.397**	**0.382**	**0.499**	**0.494**	**0.328**	**0.261**	**0.252**	**0.589**	**-0.103**	-0.074
Panel E: Ave Cross-sect Corr with **Perm. Price Impact (TAQ)** based on Individual Firms (400 Firms; 156 Months)												
1993	**0.114**	0.165	0.127	0.152	0.126	**0.159**	0.128	0.136	0.145	**0.190**	-0.106	-0.051
1994	**0.256**	**0.230**	**0.262**	**0.250**	**0.223**	**0.225**	**0.175**	**0.147**	**0.183**	**0.248**	-0.131	-0.054
1995	0.225	**0.193**	**0.175**	**0.222**	**0.193**	**0.218**	**0.210**	**0.194**	**0.162**	0.274	-0.014	-0.056
1996	**0.094**	**0.094**	**0.122**	**0.099**	**0.095**	**0.082**	**0.079**	0.080	**0.101**	0.117	0.035	-0.077
1997	**0.087**	**0.078**	**0.060**	**0.062**	**0.067**	**0.087**	**0.082**	**0.082**	0.072	**0.122**	0.000	-0.059
1998	**0.127**	**0.146**	0.166	0.139	**0.100**	**0.112**	**0.104**	**0.149**	0.198	0.098	-0.056	-0.054
1999	**0.068**	**0.107**	**0.102**	**0.105**	**0.085**	**0.117**	**0.124**	**0.124**	**0.140**	**0.002**	-0.011	-0.036
2000	**0.092**	**0.106**	**0.115**	**0.108**	**0.094**	**0.105**	**0.105**	**0.065**	**0.101**	**0.082**	0.038	-0.041
2001	0.097	0.142	**0.137**	0.146	0.148	0.137	**0.122**	**0.093**	**0.063**	0.143	**-0.073**	-0.002
2002	0.100	0.083	0.082	0.046	0.097	0.096	0.108	0.129	0.057	0.056	-0.072	-0.005
2003	0.165	0.239	0.214	0.217	0.183	0.199	0.190	0.206	0.227	0.176	-0.062	-0.018
2004	0.245	0.224	0.216	0.232	0.240	0.244	0.196	0.198	0.208	0.204	-0.037	-0.001
2005	0.043	0.073	0.099	0.074	-0.015	0.054	0.094	0.100	0.117	0.048	0.082	-0.008

Table 9

Annual Effective Spread Proxies Compared To The Quoted Spread Of The Dow Portfolio 1962 - 2000

The benchmark Quoted Spread (Dow) is the percentage quoted spread of the Dow Jones Industrial Average portfolio for a given year. All Effective Spread Proxies are calculated from daily price and volume data for a given firm-year. The Effective Spread Proxies are: Roll from Roll (1984), Effective Tick and Effective Tick 2 from this paper and Holden (2006), Holden from Holden (2006), Gibbs from Hasbrouck (2004), LOT Mixed, Zeros, and Zeros2 from Lesmond, Odgen, and Trzcinka (1999), LOT Y-split from this paper, Amihud from Amihud (2002), Pastor and Stambaugh from Pastor and Stambaugh (2002), and the Amivest Liquidity ratio. Bold numbers are statistically significant at the 5% level.

	Roll	Effective Tick	Effective Tick2	Holden	Gibbs	LOT Mixed	LOT Y-split	Zeros	Zeros2	Amihud	Pastor & Stambaugh	Amivest Liquidity
Time-series Correlations based on the Dow Jones Industrial Average Portfolio 1962 - 2000 (39 Years)												
Quoted Spread (Dow)	**-0.642**	**0.844**	**0.844**	**0.847**	**-0.395**	**0.427**	**0.778**	**0.741**	**0.741**	**0.596**	0.120	**-0.503**
Insignificantly different from	Amivest	Eff Tick2 Holden	Eff Tick Holden	Eff Tick Eff Tick2	Amivest	Amihud Pas/Stam	Zeros Zeros2	LOTY Zeros2 Amihud	LOTY Zeros Amihud	LOTY Zeros Zeros2	LOT Mix	Roll Gibbs

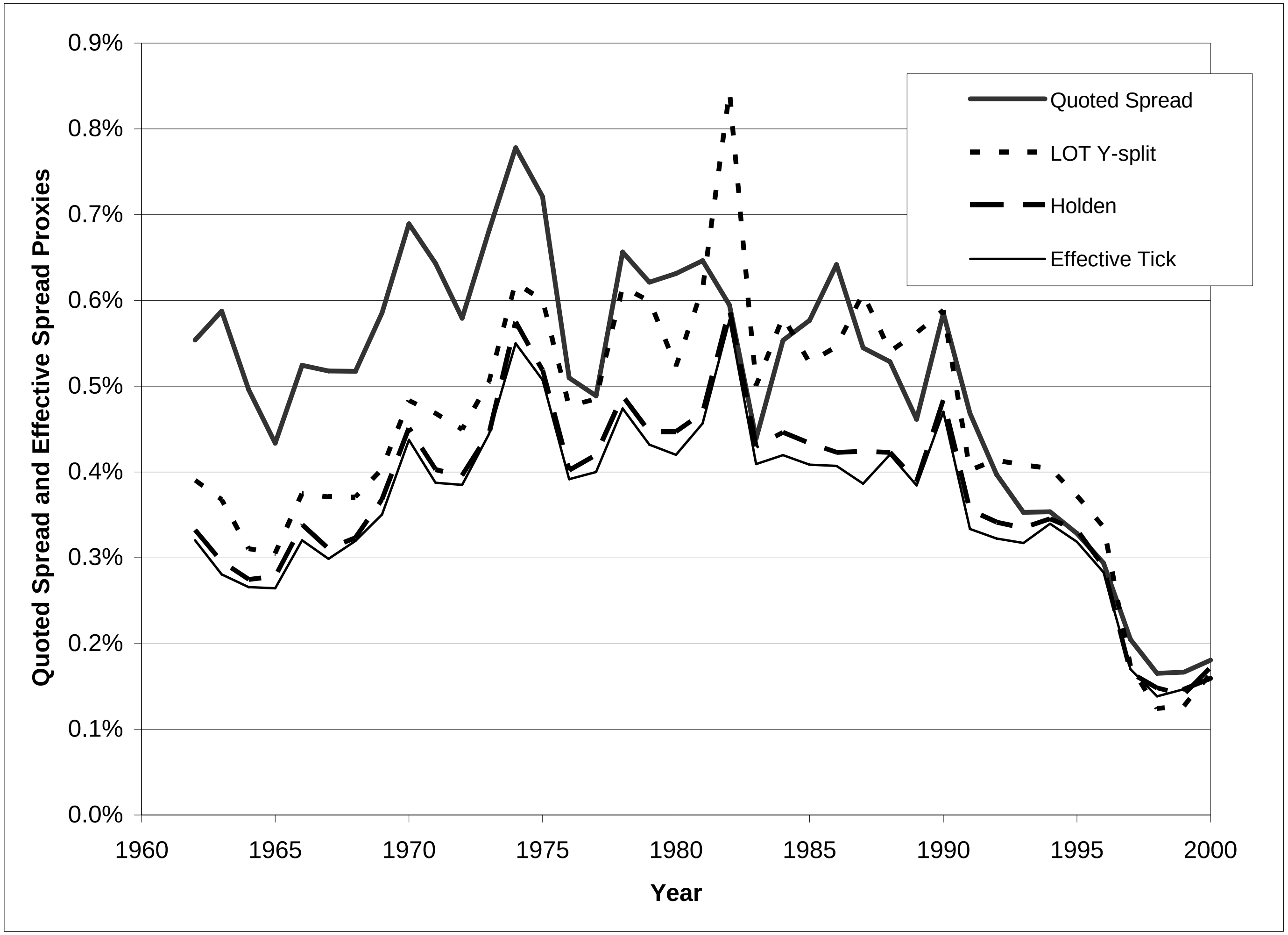

Fig. 1. Dow Portfolio Quoted Spread and Effective Spread Proxies (1962 – 2000).