The Term Structure of Liquidity Provision[*]

Jennifer Conrad
Kenan Flagler Business School
University of North Carolina at Chapel Hill
j_conrad@kenan-flagler.unc.edu

Sunil Wahal
WP Carey School of Business
Arizona State University
Tempe, AZ 85287
Sunil.Wahal@asu.edu

September 2016

The Term Structure of Liquidity Provision

Abstract

We examine the term structure of liquidity provision in all stocks from 100 milliseconds to 600 seconds after each trade for 2000-2015. At a one second horizon, the aggregate price of liquidity provision, net of losses to information, fell from 17 basis points of total dollar volume in 2000 to 1.5 basis points in 2015. Regulatory changes, technological shocks, and changes in the industrial organization of markets are associated with declines in the price of liquidity provision, and with changes in the term structure. Over the 16-year period, the slope of the term structure is increasingly steep, consistent with intense non-price (i.e. speed) competition. The term structure and profitability of market-making is closely tied to market, rather than idiosyncratic, risk. This is consistent with electronic market makers managing liquidity provision in large well-diversified portfolios of securities.

1. Introduction

Canonical models of the price formation process incorporate inventory effects, adverse selection, order-processing costs, competition between liquidity providers, and strategic behavior (see, e.g., Stoll (1978), Ho and Stoll (1981, 1983), Glosten and Milgrom (1985) and Kyle (1989)). A central tenet of these models is that price changes associated with trading contain a permanent component attributable to information, and a temporary component associated with liquidity provision. The temporary component is the net revenue to liquidity providers, and generates reversals in transaction prices.[1] The decomposition of price changes into permanent and temporary components has been characterized in various forms by Glosten (1987), Glosten and Harris (1988), Stoll (1989), Huang and Stoll (1996), Kraus and Stoll (1972), and many others.[2]

The standard microstructure approach to separating short horizon price changes into permanent and temporary components is to compute an effective spread as the difference between a transaction price and its true value at time t, and then decompose the effective spread into the price impact of the trade due to information, and the realized spread. The latter represents the net profit to liquidity provision. This decomposition relies crucially on an estimate of value at some point in the future ($t+\tau$): if one uses the quote midpoint as a proxy for value, then price impact is estimated as the movement in quote midpoint from the trade at time t to $t+\tau$, while the realized spread is calculated as the movement from the transaction price at time t to the quote midpoint at $t+\tau$.

In this paper, we examine realized spreads over different values of τ, which reflect the increasing speed of quotations and trading. There are a number of reasons why understanding the term structure of profits to liquidity provision is important. First, although many models focus on speed differentials among traders and their effect on profits (for example, Biais, Foucault and Moinas (2015), Foucault, Hombert and Roşu (2016), and Roşu (2015) and others), theory offers no guidance with regard to the appropriate horizon to measure profits. As Huang

---

[1] Net revenues remove permanent price impact. Assuming fixed cost structures, we follow the existing literature and refer to net revenues as profits. We do not mean to imply that these are excess profits. Rather, in a Grossman and Stiglitz (1980) sense, they can be thought of as the equilibrium profits necessary to entice liquidity providers to provide market making services.
[2] For a recent discussion of price formation in a broader context, see Duffie (2010).

1

and Stoll (1996) point out, if τ is too short, future prices may not reflect the reversal from transaction price to fundamental value because prices will continue to be affected by the price impact of (subsequent) trades that are part of the same *ex ante* latent demand. When orders are algorithmically broken up into correlated trades this is a significant concern.[3] On the other hand, if τ is too large, considerable noise enters into the estimation. Second, market-making horizons depend on the nature of competition between liquidity providers. A long literature in industrial organization moves beyond Bertrand competition and considers the role of non-price attributes (for a summary, see Tirole (1988)). The underlying "product" in liquidity provision is homogenous suggesting that absent product differentiation, speed may be important.[4] Market makers can compete by posting quotes more quickly in order to capture uninformed order flow, by removing quotes more rapidly to avoid adverse selection, or by being quicker to remove liquidity when unwinding inventory. Latency is thus a competitive disadvantage because of inferior queue positioning in electronic limit order books. However, if the minimum price variation (i.e. tick size) is binding so that market makers cannot compete by shrinking quoted spreads, the nature of the competition for speed may change. The implication of non-price competition is that to measure profits and better understand competition between liquidity suppliers, it is important to understand the term structure of realized spreads. Third, the slope of the term structure can provide insight into the incentives for liquidity demanders to trade rapidly. Since realized spreads are directly related to price impact, the slope of the term structure of realized spreads provides information about the speed with which trading information is incorporated into prices. If post-trade quote stabilization is rapid, the information content of trades is also reflected rapidly in market prices. This influences the incentives to trade quickly by liquidity demanders who possess information, such as the fast (HFT) traders in Biais, Foucault and Moinas (2015), Foucault, Hombert and Roşu (2016), and Roşu (2015). Speed

---

[3] For example, Implementation Shortfall (IS) algorithms break up orders to minimize price impact. Other algorithms (VWAP, TWAP, etc.) similarly break up orders to maintain a consistent percent of volume (POV) profile over the life of the parent order.

[4] Other forms of non-price competition, such as anonymity (offered by dark pools) or quantity (reflected in increasing quote sizes) could also be important. Our focus is on price and time, but we do not mean to imply that other dimensions are irrelevant. One might also consider spatial models of competition such as those of Hotelling (1929) and Salop (1979) to be relevant because of the role of co-located servers in modern exchanges. However, co-location is simply a means to an end: speed is the primitive competitive tool and co-location is merely a way to enhance speed.

competition among both liquidity demanders and suppliers can also generate speed competition among trading venues, endogenously resulting in increased fragmentation (see, e.g., Pagnotta and Philippon (2015)).

We study the term structure of realized spreads for all trades in US stocks between 2000 and July 2015. The full cross-section allows us to draw inferences about aggregate liquidity provision and examine its cross-sectional variation. A long time series helps us understand the evolution of liquidity provision through a variety of regulatory changes and market structures, and most importantly, the transition from manual to electronic markets with open limit order books. For the 2000-2009 period, we measure realized spreads starting at one second after each trade and ending at 600 seconds after the trade. For the 2010-2015 period, we exploit the availability of millisecond timestamps and also measure realized spreads at 100 millisecond intervals within the first second.

We document substantial changes in liquidity provision through time. Aggregate realized spreads at any given horizon decline sharply over the sample period. The aggregate dollar value of realized spreads one second after trading is over $47 billion in 2000. Scaling by total dollar volume, this implies that the aggregate profit to liquidity provision at this horizon is about 17 basis points of dollar volume. The subsequent decline is not monotonic. Aggregate realized spreads at the same horizon decline every year thereafter, reaching a nadir of about 0.16 basis points of dollar volume in 2012. After 2012, aggregate realized spreads rise steadily to about 1.5 basis points of dollar volume in 2015. We consider two plausible and non-mutually exclusive reasons for this pattern. First, it is possible that intense competition between liquidity providers drives down aggregate market-making profits to below marginal cost levels in 2012. Because aggregate market making profits cannot be negative in the long run, equilibrium requires market consolidation and/or repricing of market-making services. The fact that aggregate spreads are negative at horizons beyond one second in 2012, and that we observe a rise in aggregate realized spreads after 2012, is consistent with this explanation. Second, it is possible that market makers in the latter part of the sample period largely operate at horizons that are shorter than one second. The data are consistent with this explanation as well. In 2012, aggregate realized spreads at a

horizon of 100 milliseconds are about 2 basis points of volume before declining to 0.16 basis points at the one-second horizon mentioned above.

We examine the time-series of average realized spreads scaled by their volatility (which we call Sharpe ratios) for the 2010-2015 period. Aside from the obvious advantage of looking at risk-adjusted profits, this analysis is informative since the increase in quote volatility associated with high frequency trading (Conrad et al. (2015), Hasbrouck (2015)) may increase the risk of liquidity provision. Similar to aggregate profits, Sharpe ratios decrease rapidly as horizons increase, and in the time series, also hit a low in 2012. Both aggregate realized spreads and Sharpe ratios suggest that investment in speed is necessary to profitably supply liquidity in the face of rapidly changing market prices. For instance, the average annualized Sharpe ratio in 2010 at the one-second horizon is approximately 3.0. To achieve the same Sharpe ratio in 2011, average market-making speed must double. Of course, it may be that the required investment in speed not only reflects competitive forces but also a wasteful arms race (see, for example, Stiglitz (2014) and Budish, Crampton and Shim (2015)). Regardless of the welfare issues associated with higher frequency liquidity provision, our results suggest that competition can provide individual liquidity providers with an economic rationale for investment in speed.

To understand the cross-sectional and time series nature of the term structure, we estimate regressions in which the intercepts can be interpreted as an average effective spread (or an instantaneous realized spread), and where the slope represents the decay in net profits over the trading horizon. The ratio of the intercept to slope can be thought of as an average zero-profit horizon for liquidity providers, or equivalently, the horizon at which information associated with trading is fully incorporated into prices. Quotes reflect information much faster than the five minutes that was the de facto standard for measuring realized spreads in the literature for many years.[5] In 2000, the zero-profit horizon is 96 and 128 seconds for large and small stocks, respectively. By 2015, this horizon is reduced to 3 and 18 seconds, respectively. These estimates are most likely upper bounds of the horizons that market participants consider when

---

[5] This is consistent with the observations in Wahal (2012), and O'Hara, Saar and Zhong (2015), who note that "In the era of high-frequency trading, the appropriate interval in our view [to measure the components of the spread] should be much shorter, perhaps on the order of 5 seconds." It is also worth noting that under Rule 11Ac1-5 (which requires that exchanges provide information about the quality of order execution), the SEC mandates that realized spreads be calculated using a five-minute horizon.

entering orders. For liquidity providers to actually turn over inventory at these horizons, they have to jockey for queue position in the limit order book, which means orders must be generated and transmitted much more quickly. Similarly, for liquidity extractors to impose adverse selection on others, trades must extract liquidity before quotes fully reflect information, again implying higher trading speeds.

We also examine the determinants of time series variation in the term structure of liquidity provision. We expect the nature of competition between liquidity suppliers to evolve due to changes in market structure, technology and regulation. These changes consist of episodic shocks to the trading environment, as well as a secular trend. We classify shocks into three categories: (a) market events such as the Flash Crash that are largely unanticipated, (b) consolidation events such as conversions of ECNs to exchanges and mergers between trading venues, and (c) regulatory changes such as decimalization or the ban in short-sales of financial stocks. Many (but not all) such shocks are associated with significant changes in intercepts and slopes. Over the entire time series, we observe a general steepening of the term structure, and the decay in market making profits at any particular horizon is especially rapid in the last few years.

Our results shed light on another shift in the nature of liquidity provision during this period. In early models of liquidity provision, idiosyncratic risk plays a key role as a measure of the risk of inventory, and as a measure of private information. We find very little variation in intercepts and slopes estimated from an entire trading day across idiosyncratic risk quintiles. In the last half hour of trading, however, intercepts increase monotonically and slopes become systematically steeper across idiosyncratic risk quintiles. This suggests that the risk of carrying inventory overnight causes an increase in the compensation required to provide liquidity, as well as an increase in the importance of speed near the end of the day. This is in stark contrast to standard inventory models (e.g. Ho and Stoll (1981)) in which dealer risk *decreases* towards the end of the day: the intuition in those models is that market makers can liquidate portfolios at relatively certain prices at the end of the trading horizon. In contrast, our results indicate that market maker risk aversion is particularly high towards the end of the day, consistent with the behavior of the inventory-averse fast trader in Roşu (2016).

One reason that idiosyncratic or total risk plays a key role in early models of liquidity provision is that market makers in these models, such as NYSE specialists and NASDAQ dealers, specialize in small subsets of securities. Changes in trading technology such as the introduction of Direct Market Access (DMA), and the prevalence of open electronic limit order books, means that any market participant can now provide liquidity. Such participants can take formal market making roles (e.g., designated market makers (DMMs) or secondary liquidity providers (SLPs) on the NYSE), or not, as is the case with high frequency trading firms that endogenously choose to post resting quotes. Regardless, modern trading architecture and infrastructure allows such participants to algorithmically provide liquidity in thousands, instead of dozens, of stocks. The implication is that the role of idiosyncratic risk should decline since it is largely diversifiable.[6] Using double sorts on stock level idiosyncratic volatility and VIX, we continue to find evidence that idiosyncratic risk matters in the last half hour of trading. However, market risk appears to have a more pronounced effect on the term structure and profitability of market making. In small stocks over the entire trading day, the term structure steepens during high volatility periods. For both large and small stocks, holding average idiosyncratic volatility constant, zero profit holding periods in the last half hour of trading decline by as much as 50 percent between low and high VIX days. In high VIX periods, the compensation required by liquidity providers increases, and the speed with which information is reflected in prices also increases – realized spreads to market-making deteriorate very rapidly in high volatility environments. This is consistent with the results in Nagel (2012) who shows that VIX is a reliable predictor of the returns to liquidity provision.

Finally, we investigate large orders and correlated trading. In non-electronic markets, large institutional orders were "worked" by dealers and broken into smaller pieces. In electronic markets, trading algorithms break up large orders into numerous (correlated) submissions and trades. Without observing parent orders that represent latent demand, it is difficult to ascertain the total price impact of an order – or its mirror image, the realized spread for an entire order.

---

[6] The fact that diversification is important is liquidity provision is also evident from the fact that many modern market making firms diversify across asset classes and borders. Large market making firms provide liquidity services in stocks, derivatives and other instruments because of economies of scale in the basic technology, cross-market arbitrage opportunities, and diversification benefits. They are also global in scope, providing the same services in many countries.

We infer the realized spreads of large orders by estimating term structure regressions for stocks in order imbalance groups over non-overlapping 10 minute windows. Average intercepts in the large imbalance groups are systematically higher for both buys and sells. Slopes are also systematically steeper for high imbalance groups, often doubling in magnitude. Thus, when order flow is correlated, the compensation required for liquidity provision rises and the rate at which information is incorporated into prices also rises. Consistent with the time series evidence, the price of providing liquidity for these orders declines from 2010 to 2013, even for the largest imbalance groups. There is a modest increase in the cost of large orders in 2014-2015, but the costs, and speed of reversals, are still smaller at the end of our sample period than in 2010. Once again, both the intercepts and slopes are related to market risk, so that instantaneous realized spreads increase and zero-profit horizons drop on high VIX days.

The remainder of the paper is organized as follows. In Section 2, we describe our sample and basic measurement approach. We discuss aggregate results in Section 3, and present cross-sectional and time series tests in Section 4. Section 5 concludes.

2. Sample Construction and Measurement

2.1 Data and Sample

We use two sources of TAQ data. For the period from 2000-2009, we use the standard monthly TAQ data in which quotes and trades are time-stamped to the second. For 2010-2015, we use the daily TAQ data in which quote and trades contain millisecond timestamps.[7] There are obvious advantages of working with data that have millisecond resolution. We avoid conflation in signing trades, a process necessary for computing effective spreads and realized spreads. More importantly, if the term structure of post-trade returns is sharply downward sloping at less than a one second horizon, millisecond granularity is a necessity.

In processing monthly TAQ data, we remove quotes with mode equal to 4, 7, 9, 11, 13, 14, 15, 19, 20, 27, 28 and trades with correction indicators not equal to 0, 1 or 2. We also remove sale condition codes that are O, Z, B, T, L G, W, J and K, quotes or trades before or after

---

[7] We end the sample on July 31, 2015 because after that the data are reported with micro-second time stamps, increasing computational complexity.

trading hours, and locked or crossed quotes. In the daily millisecond TAQ data, we also employ BBO qualifying conditions and symbol suffixes to filter the data.

We use an algorithm provided by WRDS (TAQ-CRSP Link Table, Wharton Research Data Service, 2010) that generates a linking table between CRSP Permno and TAQ Tickers. Using this table, we create a universe of firms with TAQ data, keeping only firms with CRSP share codes 10 or 11 and exchange codes 1, 2, and 3. To ensure that small infrequently traded firms do not unduly influence our results, we remove firms with a market value of equity less than $100 million or a share price less than $1 at the beginning of the month.

We display most of our results separately for small and large capitalization stocks because size is so strongly correlated with liquidity variables. We employ the prior month's NYSE size breakpoints from Ken French's website for size assignments. On average, we sample more than 3,000 stocks, which represent over 95 percent of aggregate U.S. market capitalization.

2.2 Measuring Realized Spreads

We measure effective (percentage) half spreads as follows:

$$es_{jt} = q_{jt}(p_{jt} - m_{jt})/m_{jt} \tag{1}$$

where q_{jt} is equal to +1 for buyer-initiated trades and -1 for seller-initiated trades, p_{jt} is the transaction price and m_{jt} is the prevailing quote midpoint. Towards the latter part of the 2000-2009 period, signing of trades is noisy because quotes and trades move quicker than second time stamps. We use the Holden and Jacobsen (2014) approach of interpolating time-stamps to minimize errors.

We calculate realized spreads as follows:

$$rs_{jt\tau} = q_{jt}(p_{jt} - m_{j,t+\tau})/m_{jt} \tag{2}$$

where $m_{j,t+\tau}$ is the quote midpoint τ periods after the trade. For the 2000-2009 period, we use $\tau = 1, 2, 3, 4, 5, 10, 15, 20, 30, 60, 120, 240, 360, 480, 600$ seconds. We start measurement at one second after each trade because that is the shortest measurement interval available that allows comparisons over the entire time series. For the 2010-2015 period we exploit the increased time resolution of the data: in addition to the above horizons, we also calculate realized spreads at 100

millisecond intervals till the first second. This allows us to measure extremely rapid changes in realized spreads.[8]

3. Aggregate Results

3.1 Aggregate Effective Spreads

Although our primary interest is in the term structure of realized spreads, we start by characterizing effective spreads. We do so because they serve as a useful benchmark to understand the costs to liquidity extractors, and because realized spreads are computed from effective spreads.

Figure 1 shows daily dollar volume weighted average effective spreads for all trades in all stocks in our sample from 2000-2015. To offer some perspective on the evolution of equity markets over this 16-year period, we annotate various points on the graph with events that are important to market structure. These events are labeled E1 through E11, and roughly coincide with changes to the regulatory environment and market structure. The labels do not represent precise dates since many changes are phased in over several months. A full legend of events appears below the figure.

The decline in effective spreads from the start of the time series is quite apparent. Interestingly, the decline starts prior to the full implementation (i.e., last phase) of decimalization, labelled E1 in the figure and occurring in April 2001. The figure suggests a secular downward trend in effective spreads coincident with replacement of manual trading processes with electronic capability, such as the full introduction of Autoquote on the NYSE (E3). The advent and consolidation of ECNs and crossing networks (such as Instinet's purchase of the Island ECN (E2), Nasdaq's purchase of Instinet (E4), and NYSE's purchase of Archipelago (E5)) also takes place over this period, culminating with full implementation of Regulation NMS (E6) in July 2007. These sets of changes constitute the first phase in the movement towards electronic trading and consolidation in equity markets.

Ignoring the financial crisis period, the post-Reg NMS decline in effective spreads from 2010 to 2015 is much more gradual, roughly coinciding with a second set of consolidation in

---

[8] We do not compute realized spreads at equally-spaced time increments because of computational considerations.

trading venues and drops in latency. For instance, during this period, BATS and Direct Edge converted from ECNs to exchanges (E7, E9), while ICE purchased NYSE Euronext (E10) and BATS merged with Direct Edge (E11). By the end of the sample period, weighted average effective spreads are approximately 2 basis points, almost 12 times lower than in 2000. In aggregate terms, this represents an enormous reduction in trading costs for liquidity seekers.[9]

In addition to these long-term changes in markets and effective spreads, the figure also shows volatility in effective spreads that correspond to economically significant events in markets. For instance, E8 shows a spike in effective spreads the day after the Flash Crash. Other (unlabeled) events typically correspond to days with large positive or negative market returns.

3.2 Aggregate Realized Spreads

Liquidity providers do not earn effective spreads because of losses to information. Netting out these losses, we compute dollar realized spreads (the dollar equivalent of equation 2) at various horizons for all trades in a day. At each horizon, we then calculate the volume-weighted sum of the dollar realized spread for each day. Summing over all trading days in a calendar year provides a dollar value for realized spreads for all trades in a year. To account for time series variation in trading volume, we normalize each year's summed realized spread by total dollar trading volume.

The results of this exercise appear in Figure 2. The graph illustrates several important features of data. First, the term structure of realized spreads is reliably downward sloping. This is true in the early part of the sample period as well as in more recent periods. The slope of the term structure varies over time, an issue we investigate in depth later in the paper. Second, the level of aggregate realized spreads at any given horizon declines sharply from 2000 through 2012. At the one-second horizon, aggregate realized spreads as a proportion of dollar volume are over 17 basis points in 2000. This declines to about 0.16 basis points in 2012, indicating an

---

[9] This reduction might, but does not necessarily, imply that transaction costs for large institutional trades that demand liquidity have also gone down. This is because when large orders are broken up into small trades, the cumulative price impact of constituent trades could have risen to offset the decline in effective spreads. We examine this issue by looking at the relative spreads associated with large order imbalances later in the paper.

enormous reduction in the aggregate price of liquidity provision. After 2012, we observe a slight but steady increase in aggregate realized spreads. By 2015, the aggregate price of liquidity provision measured at the one-second horizon reaches 1.46 basis points, similar to that in 2010. This broad trend suggests intense competition in liquidity provision. It is possible that this intense competition temporarily reduces aggregate market-making profits to below marginal cost levels, perhaps generating industry consolidation and re-pricing of market-making services.[10] Without detailed cost data, it is impossible to verify this hypothesis but the fact that aggregate realized spreads rise after 2012 is consistent with the conjecture.

Equally plausible is the possibility that competition between market makers is also manifested through speed so that industry profits are severely mis-measured at a one-second horizon, particularly in the latter part of the sample period. Indeed, in 2012, at horizons beyond one second, aggregate realized spreads are negative. To investigate the role of sub-second speed, we calculate and report aggregate realized spreads at 100 millisecond intervals in the first second after trading for 2010-2015 in Figure 3. This figure effectively "zooms in" on the last six years of the sample while exploiting millisecond timestamps. Figure 3 confirms that in all six years, the term structure of realized spreads measured continues to be downward sloping. In 2010, the aggregate price of liquidity provision 100 milliseconds after trades is about two basis points of aggregate volume. This declines to a nadir of 0.4 basis points in 2012, and then increases each year thereafter to 1.3 basis points in 2015. These results suggest that sub-second speed plays a vital role in the profits of liquidity provision, especially after 2010.

3.3 Aggregate Sharpe Ratios

It is particularly important to account for the risk of market-making because high frequency trading is associated with considerable quote volatility, which influences execution risk (Conrad et al. (2015) and Hasbrouck (2015)). We standardize realized spreads by their volatility and call this measure the Sharpe ratio of liquidity provision. To calculate this ratio, we first compute the dollar-volume weighted average realized spread for all trades in all stocks in a

---

[10] Several of the consolidation events in Figure 1, such as the purchase of NYSE Euronext by ICE, and the merger between BATS and Direct Edge, are representative of industry consolidation.

one-minute interval. The Sharpe ratio is then the average realized spread, scaled by the standard deviation of the realized spread for the trades in that minute (for various values of τ). We compute the dollar-volume weighted average Sharpe ratio across the 390 trading minutes in a day to generate a daily time series average. This procedure effectively takes the perspective of the aggregate market maker who facilitates all trades, rebuilding portfolios every minute.

Annualized Sharpe ratios for each horizon are reported in Table 1. In the 2010-2015 period, the annualized Sharpe ratio at 100 milliseconds is very large at 6.92. By way of comparison, Nagel (2012) reports an annualized Sharpe ratio of 8.44 based on a rolling five-day transaction price-based reversal strategy for 1998-2010. At the one-second mark, the average Sharpe ratio declines to 1.09, and becomes zero or negative at the 15 to 20 second horizon. This suggests that liquidity providers must operate over extremely short horizons to remain profitable on a risk-adjusted basis, and that they face a market where trading information is assimilated into prices over equally short horizons.

The table also shows sub-second Sharpe ratios across the individual years 2010-2015. As with the realized spreads in Figure 3, Sharpe ratios decline steadily from 2010 to 2012. At the one-second horizon, for example, Sharpe ratios decline by over 90 percent, from approximately 2.87 in 2010 to 0.15 in 2012. In 2013 and 2014, they rise to 0.66 and 0.67, respectively, before again falling to 0.15 in 2015. For the most part, these calendar year patterns mirror those observed in aggregate realized spreads. Another way to interpret the results is to compare the horizons over which liquidity providers must operate to generate approximately the same Sharpe ratio from year to year. For example, at the one-second horizon, to achieve the same Sharpe ratio in 2011 as in 2010 (of approximately 3), the horizon of the liquidity provider must halve, from 1 second in 2010 to 500 milliseconds in 2011. From the liquidity providers' point of view, the market speed required to generate the same risk-adjusted return doubled over this one year interval. That, of course, requires investment in data feeds and technology. Such investment has been characterized as a wasteful arms race by some (Stiglitz (2014), Budish, Crampton, and Shim (2015)). While we do not construct such a welfare analysis, these results provide an economic rationale for such investments on the part of individual liquidity providers.

4. Time Series and Cross-Sectional Variation in Term Structure

Our interest in cross-sectional variation arises from the fact that, in addition to large differences in liquidity across large and small market capitalization securities, many microstructure models posit a specialist who makes markets in a single, or at most, a handful of securities. In such models, idiosyncratic risk plays a key role, whether due to inventory or information. There are two reasons for our interest in time series variation. First, we expect the nature of competition between liquidity suppliers to change over time, from the monopolist specialist first modeled by Garman (1976), to competitive market makers considered by Glosten and Milgrom (1985) and Easley and O'Hara (1987), and finally to endogenous liquidity suppliers in a public limit order book (Glosten (1994)). Second, the use of technology in market making means that market makers can provide liquidity in hundreds or perhaps thousands of securities. As a consequence, the role of idiosyncratic risk in realized spreads and their term structure may diminish over time. In contrast, we expect the role of non-diversifiable market risk to become more important.

4.1 Baseline Methods

We take a simple reduced form approach to characterize the term structure that accommodates differing timestamps in the data. Our basic model regresses realized spreads on the horizon τ, while letting the relation between the realized spread and the horizon change over various horizons as well as at different times of the day. When using millisecond data for the 2010-2015 sample period, we estimate the following regression for each stock-day (suppressing subscripts for stock i and day t):

$$rs_{it} = \alpha + \beta\tau + \gamma_1 I^{1<\tau<30} + \gamma_2 \tau * I^{1<\tau<30} + \gamma_3 I^{\tau>30} + \gamma_4 \tau * I^{\tau>30} + \gamma_5 I^{t>3:30} + \gamma_6 \tau * I^{t>3:30} + \varepsilon \qquad (3)$$

where I denotes indicator variables for horizons or time periods indicated by the superscript and $\tau*I$ denotes interaction terms. This specification allows for the slope of the term structure to vary between horizons less than one second (in the millisecond data), horizons between one and 30

seconds, and horizons longer than 30 seconds. We also allow the relationship to change in the last half hour of the trading day.

The advantage of this regression approach is that it is parsimonious and captures the term structure with two terms, the intercept (α) and the slope (β).[11] We can think of the intercept as the effective spread (since $\tau=0$), or equivalently, the instantaneous realized spread, expressed in basis points. The slope represents the decay in the net realized spread, or profits to the liquidity provider, over time. Since the term structure is downward sloping, if we divide the intercept by the absolute value of the slope, we can consider this to be the zero-profit horizon – the interval in which liquidity providers must turn around trades to earn non-negative profits.

When using trades from the last half hour of trading, we estimate a modified version of equation (3) in which we drop γ_5 and γ_6. For the 2000-2009 sample period, we drop the indicator variables and interaction terms corresponding to γ_1 and γ_2 since we do not have millisecond data for that period. We caution, however, that the lack of millisecond timestamps may cause an upward bias in β (i.e., a shallower slope to the term structure) since portions of the price response may well occur within a second. For instance, suppose the term structure of realized spreads in 2009 is similar that in 2010. Estimates of slopes (β's) from second timestamps are inevitably flatter than those using millisecond timestamps because quote changes within the first second have already occurred. As a consequence, the slope estimates prior to 2010 should be interpreted with some caution. We provide more detail on this below.

4.2 Time Series Variation

Table 2 shows average intercepts and slopes. We compute daily cross-sectional averages for large and small capitalization stocks, and then report time series averages and standard errors for each year. Unsurprisingly, intercepts are substantially larger for small capitalization stocks, indicating that the price of providing liquidity is higher for these securities. For both large and small stocks, intercepts decline noticeably and fairly steadily from 2000 to about 2007,

---

[11] To test for the influence of non-linearity, we also estimate these regressions with squared terms on the right hand side for a subsample. The squared term has little influence on the intercept and slope, so we elect to estimate simpler specifications. One could also imagine more complicated specifications (including hazard functions), or some way to find the optimal (profit-maximizing) horizon. However, given the size of the dataset, such alternatives involve enormous computational complexity and time.

consistent with the data presented in Figure 2. We also observe a spike in the cost of liquidity provision during the financial crisis.

Dividing intercepts by the slope estimates indicates that in 2000, the zero-profit horizon for large and small stocks is about 96 and 128 seconds, respectively. Even pre-decimalization, these data suggest that calculating realized spreads using a 300 second horizon may have been unnecessarily long. Zero-profit horizons decline steadily from 2000 to 2006. The advent of high frequency trading from about 2007 onwards means that the slopes estimated using data with one second timestamps appear flatter simply because they miss sub-second intervals when quotes are moving. This is most obvious in 2009, where the slope for large stocks is only -0.02. In contrast, in the very next year (2010), the slope estimated with millisecond data is -0.43. Indeed, in the last six years of the time series, average slopes are quite large and negative. Through the period in which we have millisecond data, the implied zero profit horizon is approximately 5 to 7 seconds for large stocks and varies from 7 to 20 seconds for small stocks.

The secular changes in Table 2 highlight the long-term decline in the price of liquidity, as seen in the intercepts, and the increasing steepness of the term structure, as observed in the slopes. We also investigate the role of regulatory changes, market events, and consolidations in the industry. To do so, we compile a list of major changes in each of these categories. We do not presume that these events are exhaustive, but we believe that they capture major events in each category and represent shocks to the trading environment that may influence the term structure. Table 3 shows the list of important regulatory changes (Panel A) and market consolidation events (Panel B). For each change, we define a pre-event period from days -20 to -1 of the first instance of the change, and a post-event period as days +1 to +20 after the full implementation of the change. For instance, Autoquote, studied by Hendershott, Jones and Menkveld (2011), was implemented between January 29 and May 27, 2003. The pre-event period for this event is the 20-day period prior to January 29, 2003 and the post-event period is the 20-day period after May 27, 2003. We compute event-time average intercepts and slopes for each stock before and after each event. The table reports cross-sectional averages of the intercept and slope, and after-minus-before differences in the intercept and slope, with associated *t*-statistics.

The intercepts in Panel A show that decimalization and the introduction of Autoquote are associated with large declines in the instantaneous price of liquidity provision. Both events show some flattening of the slope, indicating that subsequent to these regulatory changes, profits of liquidity providers are slower to deteriorate. The fact that realized spreads do not deteriorate as quickly after the introduction of Autoquote may be consistent with the evidence in Hendershott, Jones and Menkveld (2011) that the realized spreads of liquidity providers increase, at least temporarily, after Autoquote. In the intervals around the implementation of Reg. NMS, we observe a modest increase in the intercept, or the instantaneous price of liquidity provision, and a very small increase in the steepness of the term structure. The short-sale ban and the introduction of limit-up limit-down rules (LULD), on the other hand, are relatively rapid regulatory changes. The increase in the intercept around the short-sale ban is large and positive, at 4.89 basis points. The slope changes from -0.08 before the ban to -0.14 after the ban, implying a faster deterioration in market-making profits and, therefore, an increasing focus on speed.[12] In contrast, LULD appears to have no influence on either intercepts or slopes.

Panel B shows similar statistics for a variety of market consolidation events. Nasdaq's purchase of the Instinet ECN, NYSE's purchase of the Archipelago ECN (forming NYSE-Arca), the formation of the Direct Exchange, NYSE's merger with Euronext and the merger of BATS and DirectEdge all appear to be related to declines in intercepts – a decline in the instantaneous price of liquidity provision. Slope changes, however, are much smaller. Indeed only Instinet's purchase of the Island ECN (in the early days of market consolidation), and the merger between BATS and DirectEdge, show a non-trivial steepening of the slope of the term structure.

We also examine four key market events: the Flash Crash of May 6, 2010, the Quant Crash of August 7, 2007, the Lehman Bankruptcy of September 15, 2008, and the US debt downgrade of August 5, 2011. Since market events are unanticipated, we can focus on tighter windows around these events. Panels A and B of Figure 4 show average intercepts and slopes for each of these events from day -5 to day 0. With the exception of the Quant Crash, average intercepts show large increases on the day of the event. For instance, on the day of the debt

---

[12] In computing event time averages for the short-sale ban, we use only the financial stocks affected by the ban. For all other events, we use the full cross-section of stocks.

downgrade, the average intercept increases from about 6 basis points on day -1 to 8.28 basis points on day 0. Similarly, with the exception of the Quant Crash, average slopes also become steeper. Again illustrating with the debt downgrade, the average slope on day -1 is -0.87 and drops to -1.18 on day 0. Overall, during the periods of market stress represented by these events, liquidity providers increase spreads, and market speed increases. In a subsequent section of the paper, we analyze the relation between market volatility and liquidity provision over the entire sample period.

4.3 Idiosyncratic Risk

In standard microstructure models, idiosyncratic or total risk plays an important role for risk-averse liquidity providers. We wish to examine how variation in idiosyncratic risk affects the price of liquidity provision through time. Since term structure slopes are estimated more precisely with millisecond timestamps, we confine our attention here to the 2010-2015 period. We measure idiosyncratic risk as the standard deviation of 5-minute transaction price returns for each stock-day. We place stocks in standard deviation quintiles based on contemporaneous breakpoints, separately for large and small capitalization stocks. Analyzing large and small market capitalization securities separately prevents high idiosyncratic risk quintiles from being dominated by small stocks. We compute cross-sectional averages of intercepts and slopes from the term structure regressions for all stocks in a size and standard deviation quintile. Time series averages of intercepts and slopes are presented in Table 4 with standard errors in parentheses.

In the 2010-2015 period, there is very little variation in intercepts and slopes between low and high standard deviation stocks. In large stocks, across the entire six-year period, the intercept declines by 0.3 basis points and the slope is virtually unchanged as we move from low to high standard deviation groups. Scaling the average intercept by the slope suggests that the zero-profit horizon is a little more than 5 seconds after a trade for the low standard deviation quintile, and about 4.6 seconds for the high standard deviation quintile. In small stocks, we see the same result: there is virtually no difference between the intercepts and slopes in low and high standard deviation quintiles. There is some time series variation in these results for both large and small securities, but it does not appear to be large or systematic.

We also examine this relation in the last 30 minutes of trading. From a theoretical perspective, models such as Ho and Stoll (1981) suggest that towards the end of the dealer's horizon, spreads decline since the risk of the securities over the remaining horizon decreases. Assuming that the dealer's horizon in a high-frequency trading environment is the end of the day, this implies, counter-factually, that spreads are lower towards the end of the day. From an empirical perspective, it is well known that both volume and volatility spike in the last half hour of the trading day, implying shifts in both demand and supply curves. As a consequence, closing prices are far from certain, implying that a dealer wishing to "go home flat" is likely to be more averse to holding inventory towards the end of the day (see, for example, Cushing and Madhavan (2000)).

Table 5 shows intercepts and slopes from term structure regressions estimated only from trades in the last half hour of the trading day (3:30 – 4:00 PM EST). Unlike Table 4, there is considerable variation in slopes and intercepts between low and high standard deviation quintiles. The average intercept for the low standard deviation quintile for large stocks is 0.85 but more than doubles to 1.83 for the high standard deviation quintile. Similarly, the slope increases from -0.11 to -0.23. Note that these changes combine to leave the zero-profit horizon essentially unchanged at the end of the day. In small stocks, the average intercept rises from 4.55 basis points to 15.33 basis points, and the slope increases from -0.24 to -0.92; we observe a slight reduction in the zero-profit horizon of small stocks at the end of the day. Overall, these results suggest that there are two aspects to liquidity provision and management of risk in the last half hour of trading. First, the steepening of the term structure suggests that prices move more quickly, and as a consequence, the speed of liquidity provision is more important at the end of the day, consistent with the inventory-averse fast traders modelled in Roşu (2016). Second, and related, the difference in intercepts implies that market makers require higher *ex ante* compensation for idiosyncratic risk as the market close approaches.

4.4 Market Risk

Specialists and single-stock market makers in non-electronic markets form the centerpiece of early models of the price formation process. Market participants in electronic

limit order books, however, can engage in algorithmic liquidity provision over a large number of securities, effectively diversifying away much of the idiosyncratic risk. Such liquidity providers are still subject to market risk. Models such as Adrian and Shin (2010) and Brunnermeier and Pedersen (2009), and the evidence in Nagel (2012) suggest that the risk-bearing capacity of liquidity providers is sensitive to measures of aggregate risk. We use the VIX to measure market risk and classify each day into low, medium and high VIX days based on the 33rd and 67th percentile of the distribution of closing VIX values constructed over the entire sample period. We then examine the variation in average intercepts and slopes for these groups.

As with the analysis of idiosyncratic risk, we examine the effect of market risk in the 2010-2015 period, but also extend the analysis to the 2000-2009 interval. As discussed earlier, the latter data do not have millisecond timestamps, which results in smaller (absolute value) estimates of slope. Notwithstanding this complication, examining this earlier period may allow us to distinguish the effects of market risk more precisely, since this subperiod includes particularly large changes in market volatility. Another advantage of including the earlier years is that they encompass periods with significant amounts of non-electronic trading, providing an interesting contrast.

Table 6 shows average intercepts and slopes from both the 2010-2015 (Panel A) and 2000-2009 (Panel B) sample periods. Panel A shows that when slopes and intercepts are estimated using trades over the entire day, we observe very little variation in average intercepts or slopes between low and high VIX periods for large stocks. In small stocks, intercepts decline modestly from 10.71 basis points in low VIX days to 9.34 basis points in high VIX days. However, the slopes for small stocks also become steeper in high volatility environments, declining from -0.52 to -1.24. Combining the two, the zero profit turnaround time for a trade in small stocks drops sharply from 20.59 seconds to 7.53 seconds between low and high VIX days.

In the last half hour of the trading day, average intercepts are consistently larger on high VIX days and slopes are substantially steeper for both large and small market capitalization securities. In large stocks, the average intercept rises from 0.89 basis points to 1.49 basis points while the slope increases from -0.08 to -0.30. The implied zero profit time falls from 11.12 seconds to 4.96 seconds. In small stocks, the equivalent drop is from 27.55 seconds to 10.52

seconds. The magnitudes of these effects are similar to the effects of differences in idiosyncratic risk at the end of the day, and the interpretation is similar. That is, these results suggest that on high VIX days, at least in the last half hour of the trading day, market makers require higher compensation and realized profits are much more sensitive to speed.

Panel B reports equivalent statistics for the 2000-2009 period. Average intercepts are systematically higher in 2000-2009 (Panel B) than in 2010-2015. This is expected, given the secular decline in realized spreads throughout our sample period. More importantly, the variation in intercepts between low and high VIX days is quite large and economically significant. For example, in large stocks, the average intercept is 1.91 basis points in low VIX days and 6.76 basis points in high VIX days. To put this in economic perspective, the differential in the price of liquidity provision between high and low volatility trading days is similar in magnitude to the decline in the intercept over a 10-year period, between 2002 and 2012. The difference for small stocks is even bigger. On low VIX days, the intercept is 7.44 basis points, while on high VIX days it is 26.54 basis points. The differences in slopes are also sizeable. In large stocks, the average slope triples from -0.03 to -0.09, and in small stocks, it doubles from -0.12 to -0.25. In the last half hour of the trading day, the intercepts are again higher across the board and the differences in the slopes are substantially larger. In general, we observe results that are qualitatively similar in 2000-2009 to those in 2000-2015. Riskier environments, measured in terms of VIX or in terms of the proximity of the market close, are associated with large differences in estimates of (instantaneous) realized spreads and more rapid deterioration in profits over short horizons. In these high-risk environments, the price of liquidity provision increases, speed advantages are larger, and trading information is assimilated into market prices more rapidly.

4.5 The Interaction Between Idiosyncratic and Market Risk

In the previous section, we presented evidence that idiosyncratic risk affects liquidity provision but only at the end of the day, and that periods of high market volatility also change the price of liquidity and the effects of speed. In this section, we investigate the marginal impact of idiosyncratic versus market risk using two approaches. The first is a simple (independent)

double sort on idiosyncratic risk and the VIX. Table 7 shows average intercepts and slopes for standard deviation quintiles and VIX terciles. Panels A and B show estimates for large and small stocks respectively, based on all trades in a day. Holding VIX levels approximately constant, we do not observe any clear pattern in average intercepts or slopes for large stocks across idiosyncratic risk quintiles in Panel A. Similarly, in idiosyncratic risk quintiles, we do not observe any monotonic pattern as the volatility of the market environment changes. Combined, in large stocks and across the entire trading day, we see no systematic evidence that either idiosyncratic or market risk affect the price and speed of liquidity provision.

In small stocks, across the full day of trading, both intercepts and slopes are largely the same across all standard deviation quintiles. However, when we hold idiosyncratic risk approximately constant and examine variation in average intercepts and slopes on low, medium, and high VIX days, we see larger effects. Although there is no systematic variation in intercepts, slopes are significantly more negative on high VIX days. For instance, in the low (high) standard deviation quintile, the average slope is -0.51 (-0.50) on low VIX days and rises to -1.21 (-1.28) on high VIX days.

Panels C and D show average intercepts and slopes based on trades in the last half hour of trading for large and small stocks respectively. As with single sorts on idiosyncratic risk and market risk, inferences about the term structure are much sharper towards the end of the trading day. Holding market risk approximately constant, intercepts and slopes increase monotonically across standard deviation quintiles for both large and small stocks. In large stocks, on low (high) VIX days, the average intercept in the low standard deviation quintile is 0.62 (1.23) basis points, rising to 1.54 (2.31) for the high standard deviation quintile. Similarly on low (high) VIX days, the average slope in the low standard deviation quintile is -0.05 (-0.26), but declines to -0.14 (-0.42) for the high standard deviation quintile. However, note that these changes in intercepts and slopes across idiosyncratic risk quintiles leave the zero-profit horizon relatively unchanged. Specifically, on low (high) VIX days, the zero profit horizon changes only slightly, from 12.40 seconds (4.65 seconds) in the low standard deviation quintile to 11.85 seconds (5.21 seconds) in the high standard deviation quintile. Thus, the implied changes in zero profit holding periods from changes in idiosyncratic risk appear to be small.

In contrast, holding idiosyncratic risk approximately constant, the changes in intercepts, slopes and implied holding periods are much larger when market risk changes. Focusing again on large stocks, in the low standard deviation quintile, the intercept rises from 0.62 on low VIX days to 1.23 on high VIX days, and for the same groups, slopes change from -0.05 to -0.26. In the high standard deviation group, intercepts rise from 1.54 to 2.31, and the corresponding slopes change from -0.14 to -0.42. The implied zero-profit holding period in the low standard deviation group on low VIX days is 12.40 seconds and drops to 4.65 seconds on high VIX days. For the high standard deviation group, the equivalent drop is from 11.85 seconds to 5.21 seconds. In contrast to the effect of idiosyncratic risk, variation in the market risk environment appear to have more substantial effects on liquidity provision. Higher market risk is associated with higher instantaneous realized spreads, more rapid reversals and faster assimilation of information into prices. There appears to be more variation in the term structure related to market, rather than idiosyncratic, risk.

Variation in intercepts and slopes could come from other stock characteristics that are correlated with the term structure. Separate tests for large and small firms can control for some of these differences, at least to the extent that unobserved characteristics covary with firm size. However, we can also reduce the noise in the idiosyncratic versus market risk comparisons by using the time series of each stock's idiosyncratic risk. To do so, we place each stock-day in a standard deviation quintile based on the time series distribution of its standard deviation. In other words, the standard deviation quintile breakpoints are different for each stock and based on its own (time series) distribution of volatility, rather than breakpoints across stocks. Since stocks can migrate from small to large capitalization groups and vice versa, we only compute average intercepts and slopes for stocks that remain in a large or small stock group.

Table 8 reports the results of this sorting process. Panels A and B show average intercepts and slopes for large and small stocks respectively, based on all trades in a day. Panels C and D show equivalent statistics for the last half hour of trading. The results are qualitatively similar to those in Table 6. Market risk appears to have a larger effect on liquidity provision, and this effect is particularly important in the last half-hour of the trading day.

4.6 The Role of Tick Size

The minimum price variation (or tick size) can play an important role in the term structure of realized spreads. If the minimum tick size is not binding, then electronic market makers can compete by narrowing the quoted spread. If the minimum tick size is binding then liquidity providers can compete by offering greater size at the quoted prices, a perspective studied by Harris (2013) and O'Hara, Saar and Zhong (2015). Our interest, however, is in whether market makers compete on the dimension of latency. To do so, we estimate the following variation of equation (3) for each stock-day in the 2010-2015 period:

$$rs_{it} = \alpha + \beta\tau + \gamma_1 I^{1<\tau<30} + \gamma_2\tau * I^{1<\tau<30} + \gamma_3 I^{\tau>30} + \gamma_4\tau * I^{\tau>30} + \gamma_5 IP + \gamma_6\tau * IP + \varepsilon \qquad (4)$$

where we add an indicator variable (*IP*) if the quoted spread is equal to a penny, and an interaction term with IP and the horizon τ. We estimate the regression using all trades for each stock-day, but include only those stocks which have at least 500 trades in a day, and for which at least 20 percent of trades occur when the quoted spread is a penny. These conditions reduce the noise in estimates of γ_5 and γ_6. Because the tick size is more likely to be binding for low price stocks, we split the sample of large and small stocks into low, medium and high price categories in which the daily closing price is below $10, between $10 and $35, and greater than $35 respectively.

Table 9 shows average intercepts, slopes and coefficients on the penny indicator and interaction variables (γ_5 and γ_6). For medium and high priced stocks, the coefficient on the penny indicator is negative: when the quoted spread is at the minimum level of one penny for these securities, the effective spread (or equivalently, the instantaneous realized spread) is lower. This likely reflects the intraday variation in spreads for stocks in which the tick size is not frequently binding. The more interesting case is for those liquid, large capitalization stocks which have low prices, and for which the penny tick may be more likely to bind. With the exception of 2010, average intercepts for low-priced, large stocks are generally lower than those for medium or high price stocks. The coefficient on the penny indicator, however, is large and positive for this same group; the relative price of liquidity provision for these stocks, when the spread is a penny,

increases substantially. The implication is that the penny tick is binding for these stocks and that the price of liquidity may be artificially high.

More interestingly, while the effects on the slope are generally negative, they are relatively small. As a result, when the penny tick is binding, the zero-profit horizon for the liquidity provider increases for large-capitalization, low-priced stocks for every year in our sample. For example, in 2011, when the penny tick is not binding the implied zero profit horizon is 1.47 seconds in low price large stocks. For the same group, the zero profit horizon rises to 11.80 seconds when the penny tick binds. In fact, when the penny tick binds, the zero-profit horizon also increases for medium-priced large stocks in the last three years in our sample. This suggests that markets for mid-priced securities may have become sufficiently liquid in the latter part of our sample that the penny tick is more likely to bind and, as a consequence, liquidity providers are competing on latency for these stocks, as well.

For small market capitalization securities, which are generally less liquid, average intercepts are higher, and average slopes are more negative than large cap securities. We observe an increase in the zero-profit horizon in only two cases: low-priced securities in 2011 and 2013. Except for those two cases, the coefficient on the penny indicator is negative and the coefficient on the interaction term tends to be positive and quite small. Thus, in small market capitalization securities, there is only weak evidence that liquidity providers compete on latency where the quoted spread is a penny.

Overall, we find some evidence in large capitalization, low-priced securities that the lower bound of one penny in quoted spreads is binding; in those cases, the price of liquidity is significantly higher with little evidence that profits deteriorate more rapidly. In these cases, the zero-profit horizon becomes significantly longer – so that liquidity demanders as well as liquidity suppliers have the luxury of time.

4.7 Order Imbalance

Our analysis to this point treats trades as independent. While this assumption may be true for small trades, we expect that throughout our sample period, large trades could cause dependence in orders. For example, in non-electronic (manual) trading, large institutional orders

were "worked", generating a sequence of correlated trades. In electronic markets, trading algorithms break up large orders into a number of smaller submissions and trades. Correlated trades influence price impact, and by construction, realized spreads. Without observing *ex ante* expressions of demand (e.g. the parent order in an algorithm), however, it is impossible to precisely "roll-up" a sequence of trades to estimate total price impact, or equivalently, estimate the term structure of realized spreads for large orders. The solution we employ in order to estimate the realized spreads of large orders in our sample is approximate but simple. For each stock, we compute order imbalances in (non-overlapping) 10-minute windows as the dollar value of buyer-initiated trades minus the dollar value of seller-initiated trades, scaled by total dollar volume in the interval. In each 10-minute window, stocks with buy or sell imbalances are placed in separate terciles (labeled low, medium and high). These terciles are reformed every 10 minutes, and we estimate term structure regressions using all trades in an order imbalance group in a day. A stock can therefore appear in different order imbalance groups in a day but its trades correspond to a given imbalance group within a 10 minute period.

Table 10 shows intercepts and slopes for order imbalance groups for the 2010-2015 sample period, as well as for low, medium and high VIX days. Intercepts rise monotonically from low to high imbalance groups. For example, the intercept in the low sell-imbalance group is 1.35 basis points and rises to 3.74 basis points in the high sell-imbalance group. Average intercepts for buy imbalance groups are quite similar. Slopes also become steeper in high imbalance groups, increasing from -0.29 in the low sell-imbalance group to -0.44 in the high sell imbalance group. Compared to the overall sample results in Table 2, however, the incremental price of liquidity provision in periods with high order imbalances seems surprisingly modest. The average intercept (across buys and sells) for the highest imbalance groups across all stocks and all years is on the order of 1-3 basis points higher than the intercept for trades in large market capitalization securities in Table 2. Moreover, the differential between the intercepts appears to be declining through time (although not monotonically), with the largest differential in 2010 and the smallest in 2014. Overall, these results suggest that the time-series patterns in the price and term structure of liquidity that we observe in the full sample of trades is robust to periods with high order imbalances.

Consistent with earlier results, average intercepts and slopes estimated over the entire five-year period across order imbalance groups are systematically different in low, medium and high VIX periods. In low, medium and high imbalance groups, intercepts rise systematically and slopes become steeper with VIX levels. For example, in the high sell-imbalance group, the average intercept rises from 3.31 basis points in the low VIX period to 4.81 basis points in the high VIX period. For the same groups, slopes change from -0.31 to -0.65. For these groups, the implied zero profit horizon drops from 10.67 seconds on low VIX days to 7.40 seconds on high VIX days. Benchmarking to the results for large stocks in Table 5, liquidity provision for large orders in Table 10 appears to be more sensitive to variations in volatility. Using the terminology in Korajczyk and Murphy (2015), providing liquidity to stressful orders in a stressed market (or faster-moving market) commands a higher price.

It is also interesting to note the patterns in the estimates of intercepts and slopes across each year of this sub-sample. Both the intercepts and the (absolute value of the) slopes decline monotonically from 2010 through 2013, then rebound slightly in 2014-2015. These changes through 2013 appear to indicate that the market for liquidity provision, even for relatively large orders, is becoming more competitive. Recall that Figure 2 indicates that aggregate realized spreads in 2012 and 2013 are negative for horizons longer than 10 to 15 seconds. The rebound in the magnitude of realized spreads in 2014-2015 is also observed in Figure 2, similar to the regression estimates for 2014-2015 in Table 10.

5. Conclusion

We investigate the short horizon term structure of liquidity provision for all stocks between 2000 and 2015. The aggregate price of liquidity provision measured at one second after each trade falls from about 17 basis points of total dollar volume in 2000 to 1.5 basis points in 2015. We also report increasing steepness in the slope of the term structure over time, implying quicker post-trade quote stabilization. This suggests a substantial increase in the need for speed to make liquidity provision economically viable. In turn, this implies an increase in the speed with which trading information is reflected into prices. In contrast to standard models of market making, idiosyncratic volatility matters relatively little to the price of liquidity provision,

particularly in the later part of the sample period. Instead, non-diversifiable market risk is related to the *ex ante* compensation required to provide liquidity, as well as the slope of the term structure, most likely because electronic market makers are well diversified across securities. Overall, the data suggest a massive change in the industry of liquidity provision driven by technology, competition between liquidity providers in electronic limit order books, and appetite for risk.

References

Adrian, T., and H.S. Shin, 2010, Liquidity and leverage, *Journal of Financial Intermediation* 19, 418-437.

Biais, Bruno, Thierry Foucault and Sophie Moinas, 2015, Equilibrium fast trading, *Journal of Financial Economics* 116, 292-313.

Brunnermeier, M.K., and L.H. Pedersen, 2009, Market liquidity and funding liquidity, *Review of Financial Studies* 22, 2201-2238.

Budish, Eric, Peter Cramton and John Shim, 2015, The high-frequency trading arms race: Frequent batch auctions as a market design response, *Quarterly Journal of Economics* 130, 1547-1621.

Conrad, Jennifer, Sunil Wahal and Jin Xiang, 2015, High frequency quoting, trading and the efficiency of prices, *Journal of Financial Economics* 116, 271-291.

Cushing, David and Ananth Madhavan, 2000, Stock returns and trading at the close, *Journal of Financial Markets* 3, 45-67.

Duffie, Darrell, 2010, Presidential address: Asset price dynamics with slow-moving capital, *Journal of Finance* 65, 1237-1269

Easley, David and Maureen O'Hara, 1987, Price, trade size and information in securities markets, *Journal of Financial Economics* 19, 69-90.

Foucault, Thierry, Johan Hombert and Ioanid Roşu, 2016, News trading and speed, *Journal of Finance* 71, 335-381.

Garman, Mark, 1976, Market microstructure, *Journal of Financial Economics* 3, 257-275.

Glosten, Lawrence R., and Paul R. Milgrom, 1985, Bid, ask and transaction prices in a specialist market with heterogeneously informed traders, *Journal of Financial Economics* 14, 71-100.

Glosten, Lawrence, 1987, Components of the bid-ask spread and the statistical properties of transaction prices, *Journal of Finance* 42, 1293-1307.

Glosten, Lawrence and Lawrence Harris, 1988, Estimating the components of the bid/ask spread, *Journal of Financial Economics* 21, 123-142.

Glosten, Lawrence, 1994, Is the open limit order book inevitable? *Journal of Finance* 49, 1127-1161.

Grossman, Sanford and Joseph Stiglitz, 1980, On the impossibility of informally efficient markets, *American Economic Review* 70, 393-408.

Hasbrouck, Joel 2015, High frequency quoting: short-term volatility in bids and offers, forthcoming *Journal of Financial and Quantitative Analysis*.

Hendershott, Terrance, Charles Jones and Albert Menkveld, 2011, Does algorithmic trading improve liquidity, *Journal of Finance* 66, 1-33.

Ho, Thomas and Hans Stoll, 1981, Optimal dealer pricing under transactions and return uncertainty, *Journal of Financial Economics* 9, 319-381.

Hotelling, Harold, 1929, Stability in competition, *Economic Journal* 39, 41-57.

Huang, Roger and Hans Stoll, 1996, Dealer versus auction markets: A paired comparison of execution costs on Nasdaq and the NYSE, *Journal of Financial Economics* 41, 313-357.

Korajczyk, R. and D. Murphy, 2015, High-frequency market making to large institutional trades, Working Paper, Northwestern University.

Kraus, A., and H. Stoll, 1972, Price impacts of block trading on the New York Stock Exchange, *Journal of Finance* 27, 569-588.

Kyle, A.S., 1989, Informed Speculation with Imperfect Competition, *Review of Economic Studies* 56, 317-355.

Nagel, Stefan, 2012, Evaporating liquidity, *Review of Financial Studies* 25, 2005-2039.

O'Hara, M., G. Saar and Z. Zhong, Relative Tick Size and the Trading Environment, 2015, Working Paper, Cornell University.

Pagnotta, Emiliano S., and Thomas Philippon, 2015, Competing on speed, Working Paper, New York University.

Roşu, Ioanid, 2015, Fast and slow informed trading, working paper, HEC Paris.

Salop, Steven C, 1979, Monopolistic competition with outside goods, *Bell Journal of Economics* 10, 141-156.

Stoll, Hans R., 1989, Inferring the components of the bid-ask spread: Theory and empirical tests, *Journal of Finance* 44, 115-134.

Stiglitz, Joseph, 2014, Tapping the brakes: are less active markets safer and better for the economy? Unpublished working paper. Columbia University, New York, NY.

Tirole, Jean, 1988, The theory of industrial organization, MIT Press.

Wahal, Sunil, 2012, Electronic markets and trading algorithms, *Journal of Trading*, 26-36.

Table 1

Average annualized Sharpe ratios

We compute the dollar-volume weighted average realized spread for all trades across all stocks in a one minute interval. We then calculate the Sharpe Ratio as the ratio of the average realized spread to the standard deviation of the realized spread for every horizon τ. This gives the Sharpe Ratio for all trades across all stocks in that minute. We take the dollar volume weighted average across the day to generate a daily time series of Sharpe Ratios. The table reports average annualized Sharpe Ratios computed from the daily time series. Horizons are reported in seconds.

Horizon (τ)	2010-2015	2010	2011	2012	2013	2014	2015
0.1	6.92	18.18	9.24	2.56	3.85	3.87	1.54
0.2	4.03	10.26	5.49	1.21	2.47	2.49	0.91
0.3	2.93	7.29	4.11	0.79	1.86	1.89	0.66
0.4	2.35	5.77	3.40	0.62	1.50	1.53	0.51
0.5	1.99	4.87	2.92	0.49	1.26	1.29	0.40
0.6	1.70	4.24	2.53	0.37	1.07	1.09	0.30
0.7	1.49	3.78	2.24	0.29	0.93	0.95	0.24
0.8	1.34	3.42	2.01	0.24	0.83	0.84	0.19
0.9	1.21	3.13	1.82	0.19	0.74	0.75	0.15
1.0	1.09	2.87	1.65	0.15	0.66	0.67	0.11
2.0	0.53	1.63	0.86	-0.05	0.28	0.28	-0.05
3.0	0.33	1.13	0.57	-0.10	0.14	0.15	-0.10
4.0	0.22	0.86	0.41	-0.12	0.08	0.08	-0.12
5.0	0.15	0.68	0.31	-0.13	0.03	0.04	-0.13
10.0	0.03	0.31	0.12	-0.13	-0.04	-0.04	-0.13
15.0	0.00	0.19	0.07	-0.12	-0.06	-0.05	-0.11
20.0	-0.02	0.13	0.04	-0.10	-0.06	-0.05	-0.10
30.0	-0.02	0.08	0.02	-0.08	-0.05	-0.04	-0.08
60.0	-0.02	0.04	0.01	-0.05	-0.04	-0.03	-0.05
120.0	-0.01	0.02	0.01	-0.03	-0.02	-0.01	-0.03
240.0	0.00	0.01	0.01	-0.01	-0.01	-0.01	-0.01
360.0	0.00	0.01	0.00	-0.01	-0.01	0.00	-0.01
480.0	0.00	0.01	0.00	-0.01	0.00	0.00	-0.01
600.0	0.00	0.01	0.00	0.00	0.00	0.00	-0.01

Table 2

Average annual intercepts and slopes from term structure regressions

We calculate realized spreads for all trades as $rs_{jt} = q_{jt}(p_{jt} - m_{j,t+\tau})/m_{jt}$ where p_{jt} is the transaction price, m_{jt} is the prevailing midpoint, q_{jt} is an indicator equal to +1 (-1) for buyer-initiated (seller-initiated) trades and τ takes on values from 100 milliseconds to 600 seconds after the trade. In the 2000-2009 period where data are time-stamped to the second, the values of τ used to calculate realized spreads start at 1 second. For the 2010-2015 period, τ takes on values from 100 milliseconds to 600 seconds after the trade. For the 2000-2009 period, we estimate the following term structure regression using all trades in a day:

$$rs_{it} = \alpha + \beta\tau + \gamma_1 I^{\tau>30} + \gamma_2\tau*I^{\tau>30} + \gamma_3 I^{t>3:30} + \gamma_4\tau*I^{t>3:30} + \varepsilon$$

For the 2010-2015 period, we estimate the following term structure regressions using all trades in a day:

$$rs_{it} = \alpha + \beta\tau + \gamma_1 I^{1<\tau<30} + \gamma_2 I^{\tau>30} + \gamma_3\tau*I^{1<\tau<30} + \gamma_4\tau*I^{\tau>30} + \gamma_5 I^{t>3:30} + \gamma_6\tau*I^{t>3:30} + \varepsilon$$

The table reports time series average intercepts (α) and slopes (β) as well as standard errors.

	Large Stocks		Small Stocks	
	Intercept	Slope	Intercept	Slope
2000	19.34	-0.20	51.28	-0.40
	(0.20)	(0.02)	(2.38)	(0.03)
2001	9.25	-0.13	24.34	-0.19
	(0.18)	(0.01)	(2.32)	(0.04)
2002	7.07	-0.13	35.67	-0.35
	(1.11)	(0.03)	(5.98)	(0.04)
2003	3.61	-0.06	14.18	-0.10
	(0.09)	(0.01)	(1.33)	(0.04)
2004	2.61	-0.04	11.42	-0.10
	(0.10)	(0.01)	(0.74)	(0.04)
2005	2.31	-0.04	7.94	-0.14
	(0.17)	(0.01)	(0.65)	(0.01)
2006	1.83	-0.04	6.70	-0.11
	(0.11)	(0.01)	(0.37)	(0.01)
2007	1.63	-0.02	6.32	-0.10
	(0.22)	(0.01)	(0.94)	(0.01)
2008	1.77	-0.03	10.53	-0.14
	(0.05)	(0.01)	(0.99)	(0.06)
2009	2.55	-0.02	9.04	-0.12
	(0.95)	(0.01)	(0.68)	(0.05)
2010	2.27	-0.43	9.05	-1.31
	(0.08)	(0.02)	(0.38)	(0.05)
2011	2.03	-0.32	7.85	-1.02
	(0.19)	(0.01)	(0.08)	(0.01)
2012	2.11	-0.36	8.32	-0.70
	(0.30)	(0.02)	(0.08)	(0.01)
2013	1.72	-0.25	9.63	-0.47
	(0.11)	(0.01)	(0.14)	(0.01)
2014	3.09	-0.53	11.95	-0.57
	(0.47)	(0.03)	(0.09)	(0.02)
2015	1.92	-0.65	10.63	-0.57
	(0.40)	(0.14)	(0.18)	(0.02)

Table 3

Average intercepts and slopes before and after regulatory and market consolidation events

For each major regulatory change or market consolidation event, we define the pre-event period as days -20 to -1 relative to the first implementation of the change. For phased-in regulatory changes this is sometimes the period before the first phase. The post-event period is days +1 to +20 after the last dates of the implementation (e.g. the last phase of a regulatory change). We compute time series average intercepts and slopes in the pre- and post-event period for each stock, and report the cross-sectional average in the columns below. We use the full sample of stocks for all events with the exception of the short-sale ban. For the short-sale ban, we use only the financial stocks affected by the ban.

	Intercept				Slope			
	Pre-Event	Post-Event	Change	T-statistic	Pre-Event	Post-Event	Change	T-statistic
Panel A: Major regulatory events								
Decimalization	29.33	19.33	-10.01	-7.22	-0.26	-0.22	0.04	3.49
Autoquote	15.07	10.09	-4.97	-8.03	-0.21	-0.12	0.08	6.73
Reg. NMS	5.45	6.74	1.28	12.82	-0.05	-0.06	-0.01	-10.21
Short-sale ban	6.41	11.3	4.89	33.45	-0.08	-0.14	-0.07	-40.00
LULD	6.53	6.54	0.01	0.17	-0.29	-0.31	-0.01	-2.77
Panel B: Major market consolidation events								
Instinet + Island	19.26	20.71	1.44	1.71	-0.22	-0.29	-0.06	-5.86
Nasdaq + Instinet	7.61	7.10	-0.50	-2.19	-0.10	-0.10	0.01	1.24
NYSE + Arca	7.20	6.29	-0.91	-4.96	-0.10	-0.09	0.01	3.34
BATS Exchange	10.88	11.05	0.17	1.59	-0.14	-0.15	-0.01	-4.15
Direct Exchange	6.29	6.19	-0.10	-2.68	-0.96	-0.98	-0.02	-2.36
NYSE + Euronext	6.21	5.51	-0.70	-6.85	-0.06	-0.06	0.01	4.81
ICE + NYSE	7.58	7.58	-0.01	-0.24	-0.32	-0.3	0.01	2.79
BATS + DirectEdge	8.92	8.71	-0.20	-2.62	-0.34	-0.38	-0.04	-5.21

Table 4

Average intercepts and slopes from term structure regressions for idiosyncratic risk quintiles, 2010-2015

We calculate realized spreads for all trades as $rs_{jt} = q_{jt}(p_{jt} - m_{j,t+\tau})/m_{jt}$ where p_{jt} is the transaction price, m_{jt} is the prevailing midpoint, q_{jt} is an indicator equal to +1 (-1) for buyer-initiated (seller-initiated) trades and τ takes on values from 100 milliseconds to 600 seconds after the trade. For each stock day, we estimate the following term structure regressions using all trades:

$$rs_{it} = \alpha + \beta\tau + \gamma_1 I^{1<\tau<30} + \gamma_2 I^{\tau>30} + \gamma_3 \tau * I^{1<\tau<30} + \gamma_4 \tau * I^{\tau>30} + \gamma_5 I^{t>3:30} + \gamma_6 \tau * I^{t>3:30} + \varepsilon$$

On each day, stocks are placed in large and small cap groups based on NYSE breakpoints from the prior month. We calculate the standard deviation of 5 minute returns for each stock-day and place stocks in standard deviation quintiles based on contemporaneous breakpoints separately for large and small stocks. The table reports time series average intercepts (α) and slopes (β) as well as standard errors.

	Intercept					Slope				
	Low Std. Deviation	2	3	4	High Std. Deviation	Low Std. Deviation	2	3	4	High Std. Deviation
	Panel A: Large Stocks									
2010	1.82	1.57	2.66	2.74	2.55	-0.44	-0.40	-0.43	-0.45	-0.59
	(0.12)	(0.13)	(0.25)	(0.05)	(0.15)	(0.05)	(0.02)	(0.02)	(0.02)	(0.03)
2011	1.45	2.40	1.83	2.86	1.63	-0.31	-0.36	-0.30	-0.33	-0.33
	(0.40)	(0.52)	(0.45)	(0.52)	(0.08)	(0.02)	(0.02)	(0.02)	(0.02)	(0.04)
2012	1.70	3.65	1.61	1.81	1.80	-0.41	-0.36	-0.34	-0.39	-0.32
	(0.04)	(0.05)	(1.49)	(0.22)	(0.14)	(0.03)	(0.02)	(0.03)	(0.03)	(0.08)
2013	1.80	1.65	1.65	1.99	1.52	-0.20	-0.28	-0.22	-0.26	-0.28
	(0.03)	(0.50)	(0.11)	(0.08)	(0.23)	(0.02)	(0.02)	(0.02)	(0.02)	(0.04)
2014	4.67	2.55	2.76	2.98	2.46	-0.42	-0.54	-0.55	-0.52	-0.58
	(0.06)	(2.25)	(0.16)	(0.45)	(0.21)	(0.02)	(0.07)	(0.05)	(0.03)	(0.15)
2015	2.36	1.64	1.82	2.07	1.73	-0.73	-0.53	-0.58	-1.08	-0.34
	(0.13)	(0.08)	(0.11)	(0.09)	(0.06)	(0.24)	(0.20)	(0.20)	(0.61)	(0.03)
All	2.29	2.28	2.07	2.43	1.96	-0.40	-0.40	-0.40	-0.46	-0.42
Years	(0.39)	(0.29)	(0.13)	(0.18)	(0.08)	(0.03)	(0.03)	(0.03)	(0.07)	(0.03)
	Panel B: Small Stocks									
2010	9.08	8.70	8.80	9.64	9.00	-1.31	-1.28	-1.35	-1.34	-1.33
	(0.46)	(0.41)	(0.38)	(0.48)	(0.42)	(0.05)	(0.05)	(0.06)	(0.07)	(0.07)
2011	7.53	7.74	8.04	8.20	7.91	-1.00	-1.03	-1.00	-1.03	-1.04
	(0.08)	(0.13)	(0.08)	(0.09)	(0.15)	(0.02)	(0.04)	(0.02)	(0.03)	(0.02)
2012	7.77	8.15	8.48	8.48	7.99	-0.64	-0.69	-0.72	-0.70	-0.67
	(0.11)	(0.10)	(0.08)	(0.16)	(0.14)	(0.03)	(0.02)	(0.04)	(0.03)	(0.02)
2013	9.75	9.68	9.75	9.65	9.29	-0.50	-0.47	-0.46	-0.47	-0.44
	(0.59)	(0.18)	(0.14)	(0.25)	(0.18)	(0.01)	(0.02)	(0.02)	(0.02)	(0.01)
2014	12.27	12.10	12.40	13.05	11.95	-0.55	-0.55	-0.55	-0.64	-0.62
	(0.14)	(0.14)	(0.17)	(0.17)	(0.13)	(0.02)	(0.02)	(0.02)	(0.05)	(0.04)
2015	10.3	10.01	10.50	11.35	11.15	-0.54	-0.61	-0.50	-0.64	-0.54
	(0.11)	(0.11)	(0.12)	(0.14)	(0.33)	(0.03)	(0.03)	(0.02)	(0.07)	(0.03)
All	9.36	9.35	9.60	9.96	9.43	-0.71	-0.77	-0.78	-0.81	-0.79
Years	(0.10)	(0.10)	(0.14)	(0.12)	(0.10)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)

Table 5

Average intercepts and slopes from term structure regressions for idiosyncratic risk quintiles for 3:30 – 4:00 PM EST, 2010-2015

We calculate realized spreads for all trades in the last half hour of the trading day as $rs_{jt} = q_{jt}(p_{jt} - m_{j,t+\tau})/m_{jt}$

where p_{jt} is the transaction price, m_{jt} is the prevailing midpoint, q_{jt} is an indicator equal to +1 (-1) for buyer-initiated (seller-initiated) trades and τ takes on values from 100 milliseconds to 600 seconds after the trade. For each stock day, we estimate the following term structure regressions using all trades:

$$rs_{it} = \alpha + \beta\tau + \gamma_1 I^{1<\tau<30} + \gamma_2 I^{\tau>30} + \gamma_3\tau*I^{1<\tau<30} + \gamma_4\tau*I^{\tau>30} + \gamma_5 I^{t>3:30} + \gamma_6\tau*I^{t>3:30} + \varepsilon$$

On each day, stocks are placed in large and small cap groups based on NYSE breakpoints from the prior month. We calculate the standard deviation of 5 minute returns for each stock-day and place stocks in standard deviation quintiles based on contemporaneous breakpoints separately for large and small stocks. The table reports time series average intercepts (α) and slopes (β) as well as standard errors.

	Intercept					Slope				
	Low Std. Deviation	2	3	4	High Std. Deviation	Low Std. Deviation	2	3	4	High Std. Deviation
Panel A: Large Stocks										
2010	1.27	1.23	1.31	1.52	2.50	-0.23	-0.24	-0.26	-0.31	-0.43
	(0.09)	(0.04)	(0.03)	(0.07)	(0.09)	(0.03)	(0.01)	(0.01)	(0.02)	(0.02)
2011	0.94	0.96	1.03	1.19	1.86	-0.13	-0.16	-0.17	-0.19	-0.26
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
2012	0.88	0.86	0.93	1.08	1.71	-0.08	-0.09	-0.10	-0.12	-0.20
	(0.06)	(0.01)	(0.01)	(0.01)	(0.01)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
2013	0.69	0.72	0.79	0.92	1.57	-0.07	-0.08	-0.09	-0.10	-0.15
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
2014	0.60	0.62	0.70	0.85	1.18	-0.05	-0.06	-0.07	-0.08	-0.13
	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
2015	0.57	0.60	0.67	0.82	1.59	-0.05	-0.06	-0.07	-0.09	-0.14
	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
All	0.85	0.85	0.92	1.08	1.83	-0.11	-0.12	-0.13	-0.16	-0.23
Years	(0.02)	(0.01)	(0.01)	(0.02)	(0.02)	(0.01)	(0.00)	(0.00)	(0.00)	(0.01)
Panel B: Small Stocks										
2010	5.51	5.57	6.42	8.39	14.91	-0.46	-0.54	-0.67	-0.85	-1.58
	(0.42)	(0.33)	(0.39)	(0.44)	(0.54)	(0.07)	(0.04)	(0.07)	(0.05)	(0.07)
2011	4.56	4.56	5.34	7.10	13.91	-0.32	-0.39	-0.47	-0.60	-1.24
	(0.12)	(0.12)	(0.07)	(0.10)	(0.26)	(0.01)	(0.01)	(0.10)	(0.01)	(0.02)
2012	4.21	4.18	5.08	6.84	13.84	-0.19	-0.24	-0.31	-0.43	-0.87
	(0.04)	(0.03)	(0.04)	(0.06)	(0.13)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
2013	4.14	4.12	5.39	7.56	14.60	-0.17	-0.20	-0.26	-0.34	-0.63
	(0.05)	(0.08)	(0.13)	(0.15)	(0.16)	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)
2014	4.40	4.22	5.56	8.33	18.00	-0.13	-0.16	-0.21	-0.28	-0.49
	(0.06)	(0.03)	(0.04)	(0.05)	(0.12)	(0.00)	(0.00)	(0.00)	(0.01)	(0.01)
2015	4.46	4.01	5.19	7.53	17.73	-0.14	-0.16	-0.22	-0.27	-0.47
	(0.12)	(0.04)	(0.04)	(0.08)	(0.12)	(0.00)	(0.00)	(0.00)	(0.01)	(0.01)
All	4.55	4.48	5.52	7.63	15.33	-0.24	-0.29	-0.37	-0.47	-0.92
Years	(0.09)	(0.07)	(0.08)	(0.10)	(0.14)	(0.01)	(0.01)	(0.02)	(0.01)	(0.02)

Table 6

Average intercepts and slopes from term structure regressions for VIX terciles

We calculate realized spreads for all trades in the last half hour of the trading day as $rs_{jt} = q_{jt}(p_{jt} - m_{j,t+\tau})/m_{jt}$ where p_{jt} is the transaction price, m_{jt} is the prevailing midpoint, q_{jt} is an indicator equal to +1 (-1) for buyer-initiated (seller-initiated) trades and τ takes on values from 100 milliseconds to 600 seconds after the trade. In the 2000-2009 period where data are time-stamped to the second, the values of τ used to calculate realized spreads start at 1 second. For the 2010-2014 period, τ takes on values from 100 milliseconds to 600 seconds after the trade. For the 2000-2009 period, we estimate the following term structure regression using all trades in a day:

$$rs_{it} = \alpha + \beta\tau + \gamma_1 I^{\tau>30} + \gamma_2\tau*I^{\tau>30} + \gamma_3 I^{t>3:30} + \gamma_4\tau*I^{t>3:30} + \varepsilon$$

For the same period, when estimating regressions using trades from the last half hour, γ_3 and γ_4 are omitted. For the 2010-2014 period, we estimate the following term structure regressions using all trades in a day:

$$rs_{it} = \alpha + \beta\tau + \gamma_1 I^{1<\tau<30} + \gamma_2 I^{\tau>30} + \gamma_3\tau*I^{1<\tau<30} + \gamma_4\tau*I^{\tau>30} + \gamma_5 I^{t>3:30} + \gamma_6\tau*I^{t>3:30} + \varepsilon$$

For the same period, when estimating regressions using trades from the last half hour, γ_5 and γ_6 are omitted. The table reports time series average intercepts (α) and slopes (β) as well as standard errors. Each day is classified into low, medium and high VIX categories based on the 33rd and 67th percentile.

	Large Stocks		Small Stocks	
	Intercept	Slope	Intercept	Slope
Panel A: 2010-2015				
Slopes and intercepts from entire trading day				
Low VIX	2.17	-0.47	10.71	-0.52
	(0.08)	(0.05)	(0.11)	(0.01)
Medium VIX	2.21	-0.37	9.00	-0.82
	(0.19)	(0.01)	(0.07)	(0.01)
High VIX	2.31	-0.45	9.34	-1.24
	(0.23)	(0.03)	(0.08)	(0.04)
Slopes and estimates from 3:30 – 4:00 PM				
Low VIX	0.89	-0.08	7.44	-0.27
	(0.01)	(0.00)	(0.05)	(0.01)
Medium VIX	1.13	-0.15	7.17	-0.48
	(0.01)	(0.00)	(0.05)	(0.01)
High VIX	1.49	-0.30	8.84	-0.84
	(0.08)	(0.01)	(0.53)	(0.09)
Panel B: 2000-2009				
Slopes and intercepts from entire trading day				
Low VIX	1.91	-0.03	7.44	-0.12
	(0.13)	(0.00)	(0.39)	(0.01)
Medium VIX	5.66	-0.07	16.44	-0.15
	(0.27)	(0.00)	(0.68)	(0.01)
High VIX	6.76	-0.09	26.54	-0.25
	(0.40)	(0.01)	(2.12)	(0.03)
Slopes and estimates from 3:30 – 4:00 PM				
Low VIX	2.19	-0.02	10.78	-0.08
	(0.04)	(0.00)	(0.20)	(0.01)
Medium VIX	6.13	-0.04	21.84	-0.02
	(0.19)	(0.00)	(2.51)	(0.10)
High VIX	7.43	-0.07	30.92	-0.23
	(0.39)	(0.02)	(1.72)	(0.08)

Table 7

Average intercepts and slopes from term structure regressions for double sorts on idiosyncratic risk and VIX, 2010-2015

We calculate realized spreads for all trades as $rs_{jt} = q_{jt}(p_{jt} - m_{j,t+\tau})/m_{jt}$ where p_{jt} is the transaction price, m_{jt} is the prevailing midpoint, q_{jt} is an indicator equal to +1 (-1) for buyer-initiated (seller-initiated) trades and τ takes on values from 100 milliseconds to 600 seconds after the trade. For each stock day, we estimate the following term structure regressions using all trades:

$$rs_{it} = \alpha + \beta\tau + \gamma_1 I^{1<\tau<30} + \gamma_2 I^{\tau>30} + \gamma_3\tau * I^{1<\tau<30} + \gamma_4\tau * I^{\tau>30} + \gamma_5 I^{t>3:30} + \gamma_6\tau * I^{t>3:30} + \varepsilon$$

When estimating regressions using trades from the last half hour, γ_5 and γ_6 are omitted. On each day, stocks are placed in large and small cap groups based on NYSE breakpoints from the prior month. We calculate the standard deviation of 5 minute returns for each stock-day and place stocks in standard deviation quintiles based on contemporaneous breakpoints separately for large and small stocks. Each day is classified into low, medium and high VIX categories based on the 33rd and 67th percentile. The table reports time series average intercepts (α) and slopes (β) as well as standard errors.

	Intercept					Slope				
	Low Std. Deviation	2	3	4	High Std. Deviation	Low Std. Deviation	2	3	4	High Std. Deviation
Panel A: Large Stocks, Full Day Estimates										
Low VIX	2.19	2.02	2.27	2.34	2.01	-0.32	-0.39	-0.52	-0.66	-0.46
	(0.04)	(0.40)	(0.13)	(0.11)	(0.21)	(0.01)	(0.05)	(0.04)	(0.02)	(0.12)
Med VIX	2.47	2.41	2.05	2.24	1.86	-0.43	-0.38	-0.32	-0.36	-0.36
	(0.05)	(1.17)	(0.77)	(0.27)	(0.09)	(0.01)	(0.01)	(0.01)	(0.01)	(0.04)
High VIX	1.74	2.32	1.77	3.45	2.24	-0.40	-0.49	-0.40	-0.44	-0.52
	(0.23)	(0.18)	(0.27)	(0.28)	(0.11)	(0.03)	(0.02)	(0.02)	(0.02)	(0.02)
Panel B: Small Stocks, Full Day Estimates										
Low VIX	10.50	10.51	10.35	11.28	10.42	-0.51	-0.50	-0.50	-0.50	-0.50
	(0.46)	(0.15)	(0.15)	(0.21)	(0.16)	(0.01)	(0.01)	(0.02)	(0.04)	(0.01)
Med VIX	8.87	8.81	9.04	9.37	8.95	-0.80	-0.84	-0.81	-0.83	-0.82
	(0.11)	(0.12)	(0.11)	(0.13)	(0.11)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
High VIX	9.02	9.22	9.34	9.74	9.40	-1.21	-1.19	-1.26	-1.25	-1.28
	(0.27)	(0.23)	(0.23)	(0.28)	(0.25)	(0.03)	(0.04)	(0.04)	(0.05)	(0.05)
Panel C: Large Stocks, Estimates from 3:30 – 4:00 PM										
Low VIX	0.62	0.66	0.74	0.88	1.54	-0.05	-0.06	-0.07	-0.08	-0.14
	(0.01)	(0.00)	(0.01)	(0.01)	(0.01)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Med VIX	0.87	0.87	0.95	1.09	1.85	-0.10	-0.12	-0.14	-0.16	-0.23
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
High VIX	1.23	1.17	1.23	1.47	2.31	-0.26	-0.25	-0.27	-0.32	-0.42
	(0.06)	(0.02)	(0.02)	(0.04)	(0.06)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)
Panel D: Small Stocks, Estimates from 3:30 – 4:00 PM										
Low VIX	4.16	4.08	5.32	7.78	16.11	-0.14	-0.17	-0.22	-0.30	-0.53
	(0.04)	(0.06)	(0.09)	(0.11)	(0.14)	(0.00)	(0.00)	(0.00)	(0.01)	(0.01)
Med VIX	4.36	4.34	5.32	7.30	14.65	-0.24	-0.29	-0.37	-0.50	-0.96
	(0.07)	(0.06)	(0.05)	(0.04)	(0.15)	(0.00)	(0.00)	(0.01)	(0.01)	(0.02)
High VIX	6.21	5.88	6.79	8.73	16.61	-0.53	-0.57	-0.71	-0.84	-1.55
	(0.24)	(0.19)	(0.19)	(0.26)	(0.32)	(0.04)	(0.02)	(0.04)	(0.03)	(0.04)

Table 8

Average intercepts and slopes from term structure regressions for double sorts on own-stock standard deviation and VIX

We calculate realized spreads for all trades as $rs_{jt} = q_{jt}(p_{jt} - m_{j,t+\tau})/m_{jt}$ where p_{jt} is the transaction price, m_{jt} is the prevailing midpoint, q_{jt} is an indicator equal to +1 (-1) for buyer-initiated (seller-initiated) trades and τ takes on values from 100 milliseconds to 600 seconds after the trade. For each stock day, we estimate the following term structure regressions using all trades:

$$rs_{it} = \alpha + \beta\tau + \gamma_1 I^{1<\tau<30} + \gamma_2 I^{\tau>30} + \gamma_3\tau * I^{1<\tau<30} + \gamma_4\tau * I^{\tau>30} + \gamma_5 I^{t>3:30} + \gamma_6\tau * I^{t>3:30} + \varepsilon$$

When estimating regressions using trades from the last half hour, γ_5 and γ_6 are omitted. On each day, stocks are placed in large and small cap groups based on NYSE breakpoints from the prior month. We calculate the standard deviation of 5 minute returns for each stock-day. Each stock-day is placed in a standard deviation quintile based on the time series of its own standard deviation estimates. Each day is also classified into low, medium and high VIX categories based on the 33rd and 67th percentile. Stocks that migrate from small to large capitalization groups (or vice versa) are omitted. The table reports time series average intercepts (α) and slopes (β) as well as standard errors.

	Intercept					Slope				
	Low Std. Deviation	2	3	4	High Std. Deviation	Low Std. Deviation	2	3	4	High Std. Deviation
Panel A: Large Stocks, Full Day Estimates										
Low VIX	1.57	1.76	1.63	1.59	1.92	-0.30	-0.43	-0.37	-0.44	-0.45
	(0.15)	(0.26)	(0.30)	(0.29)	(0.28)	(0.06)	(0.06)	(0.08)	(0.17)	(0.06)
Med VIX	1.67	1.46	1.72	1.47	1.52	-0.39	-0.32	-0.36	-0.29	-0.29
	(0.61)	(0.83)	(0.10)	(0.18)	(0.29)	(0.06)	(0.03)	(0.04)	(0.03)	(0.03)
High VIX	2.09	1.72	1.67	2.36	1.71	-0.42	-0.40	-0.38	-0.75	-0.39
	(0.14)	(0.20)	(0.10)	(0.63)	(0.17)	(0.03)	(0.09)	(0.04)	(0.16)	(0.01)
Panel B: Small Stocks, Full Day Estimates										
Low VIX	16.66	16.64	16.58	15.69	16.13	-0.62	-0.77	-0.74	-0.65	-0.66
	(0.63)	(1.22)	(0.75)	(0.73)	(0.89)	(0.05)	(0.03)	(0.04)	(0.10)	(0.04)
Med VIX	16.07	16.49	16.56	16.34	17.41	-1.14	-1.25	-1.32	-1.21	-1.11
	(0.40)	(1.79)	(0.57)	(0.45)	(0.87)	(0.04)	(0.18)	(0.05)	(0.03)	(0.06)
High VIX	15.02	15.75	15.99	17.79	15.25	-1.53	-1.90	-1.63	-3.25	-1.55
	(0.40)	(0.64)	(0.77)	(1.07)	(0.49)	(0.06)	(0.08)	(0.05)	(0.86)	(0.11)
Panel C: Large Stocks, Estimates from 3:30 – 4:00 PM										
Low VIX	0.70	0.73	0.76	0.80	0.94	-0.05	-0.07	-0.08	-0.11	-0.19
	(0.03)	(0.03)	(0.03)	(0.03)	(0.05)	(0.00)	(0.00)	(0.00)	(0.00)	(0.02)
Med VIX	0.93	0.95	0.99	1.04	1.18	-0.10	-0.13	-0.15	-0.18	-0.24
	(0.03)	(0.03)	(0.04)	(0.05)	(0.07)	(0.00)	(0.00)	(0.00)	(0.00)	(0.01)
High VIX	1.18	1.17	1.20	1.71	1.50	-0.24	-0.23	-0.25	-0.27	-0.35
	(0.05)	(0.05)	(0.06)	(0.10)	(0.09)	(0.04)	(0.01)	(0.01)	(0.02)	(0.04)
Panel D: Small Stocks, Estimates from 3:30 – 4:00 PM										
Low VIX	10.11	10.48	11.12	13.67	16.96	-0.19	-0.29	-0.36	-0.41	-0.60
	(0.78)	(0.62)	(0.59)	(0.83)	(0.50)	(0.02)	(0.01)	(0.02)	(0.04)	(0.03)
Med VIX	9.98	10.19	11.58	13.03	19.23	-0.32	-0.50	-0.62	-0.84	-1.41
	(0.36)	(0.22)	(0.26)	(0.31)	(0.74)	(0.02)	(0.02)	(0.03)	(0.04)	(0.05)
High VIX	10.96	10.85	11.73	13.42	17.98	-0.53	-0.70	-0.88	-1.07	-1.56
	(0.30)	(0.33)	(0.34)	(0.39)	(0.70)	(0.03)	(0.06)	(0.04)	(0.06)	(0.16)

Table 9

Intercepts and slopes from term structure regressions for stocks sorted on size and price level

We calculate realized spreads for all trades as $rs_{jt} = q_{jt}(p_{jt} - m_{j,t+\tau})/m_{jt}$ where p_{jt} is the transaction price, m_{jt} is the prevailing midpoint, q_{jt} is an indicator equal to +1 (-1) for buyer-initiated (seller-initiated) trades and τ takes on values from 100 milliseconds to 600 seconds after the trade. We estimate the following term structure regressions using all trades in a day for stocks that have at least 500 trades in the day and where at least 20 percent of trades occur when the quoted spread is a penny:

$$rs_{it} = \alpha + \beta\tau + \gamma_1 I^{1<\tau<30} + \gamma_2 I^{\tau>30} + \gamma_3\tau*I^{1<\tau<30} + \gamma_4\tau*I^{\tau>30} + \gamma_5 IP + \gamma_6\tau*IP + \varepsilon$$

The table average intercepts (α), slopes (β), the penny indicator (γ_5) and the penny interaction term (γ_6). Low price level stocks are those with closing prices below \$10, medium price level stocks have prices between \$10 and \$35, and high price level stocks are those with prices greater than \$35.

Year	Price Level	Large Stocks				Small Stocks			
		Intercept	Slope	Penny Indicator	Penny Interaction	Intercept	Slope	Penny Indicator	Penny Interaction
2010	Low	1.85	-0.87	4.35	-0.0054	8.96	-1.19	-0.03	-0.0037
	Medium	1.39	-0.36	-0.12	-0.0022	3.06	-0.56	-1.79	0.0004
	High	1.41	-0.33	-0.71	0.0016	4.32	-0.52	-2.95	0.0052
2011	Low	0.66	-0.45	4.71	-0.0051	7.12	-0.91	1.01	-0.0041
	Medium	1.36	-0.21	-0.32	0.0001	3.11	-0.38	-1.74	0.0009
	High	1.12	-0.23	-0.52	0.0003	3.34	-0.31	-2.22	0.0032
2012	Low	0.50	-0.29	3.41	-0.0031	7.77	-0.61	-1.11	-0.0004
	Medium	0.99	-0.19	-0.03	-0.0001	3.69	-0.24	-2.42	0.0033
	High	1.41	-0.27	-0.63	0.0007	2.99	-0.43	-1.83	0.0018
2013	Low	0.22	-0.26	3.49	-0.0031	5.93	-0.39	1.37	-0.0012
	Medium	0.57	-0.12	0.41	-0.0007	3.23	-0.17	-1.68	0.0035
	High	0.83	-0.16	-0.25	0.0002	2.57	-0.19	-1.09	0.0014
2014	Low	0.07	-0.24	3.05	-0.0022	7.56	-0.42	-1.03	0.0009
	Medium	0.61	-0.14	0.37	-0.0002	2.62	-0.21	-1.28	0.0014
	High	1.01	-0.29	-0.06	-0.0001	2.57	-0.19	-1.09	0.0014
2015	Low	-0.09	-0.25	3.31	-0.0031	7.38	-0.39	-0.88	0.0021
	Medium	0.43	-0.11	0.41	-0.0001	2.22	-0.19	-1.09	0.0015
	High	0.91	-0.13	-0.35	0.0006	1.69	-0.23	-1.04	0.0013

Table 10

Intercepts and slopes from term structure regressions for order imbalance groups

Each trades is signed as a buyer- or seller-initiated trade. For each stock, we compute order imbalance in non-overlapping 10 minute windows as the dollar value of buyer-initiated trades minus the dollar value of sell-initiated trades, scaled by the total dollar volume. In each 10 minute window, stocks with sell imbalance are placed in three groups labeled low, medium and high. Stocks with buy imbalance are placed in equivalent groups. Order imbalance breakpoints and groups are reformed every 10 minutes. We estimate term structure regressions using all trades in an order imbalance group in a day (a stock can appear in different order imbalance groups in a day but its trades correspond to a given imbalance group within a 10 minute period). The sample period is 2010-2015.

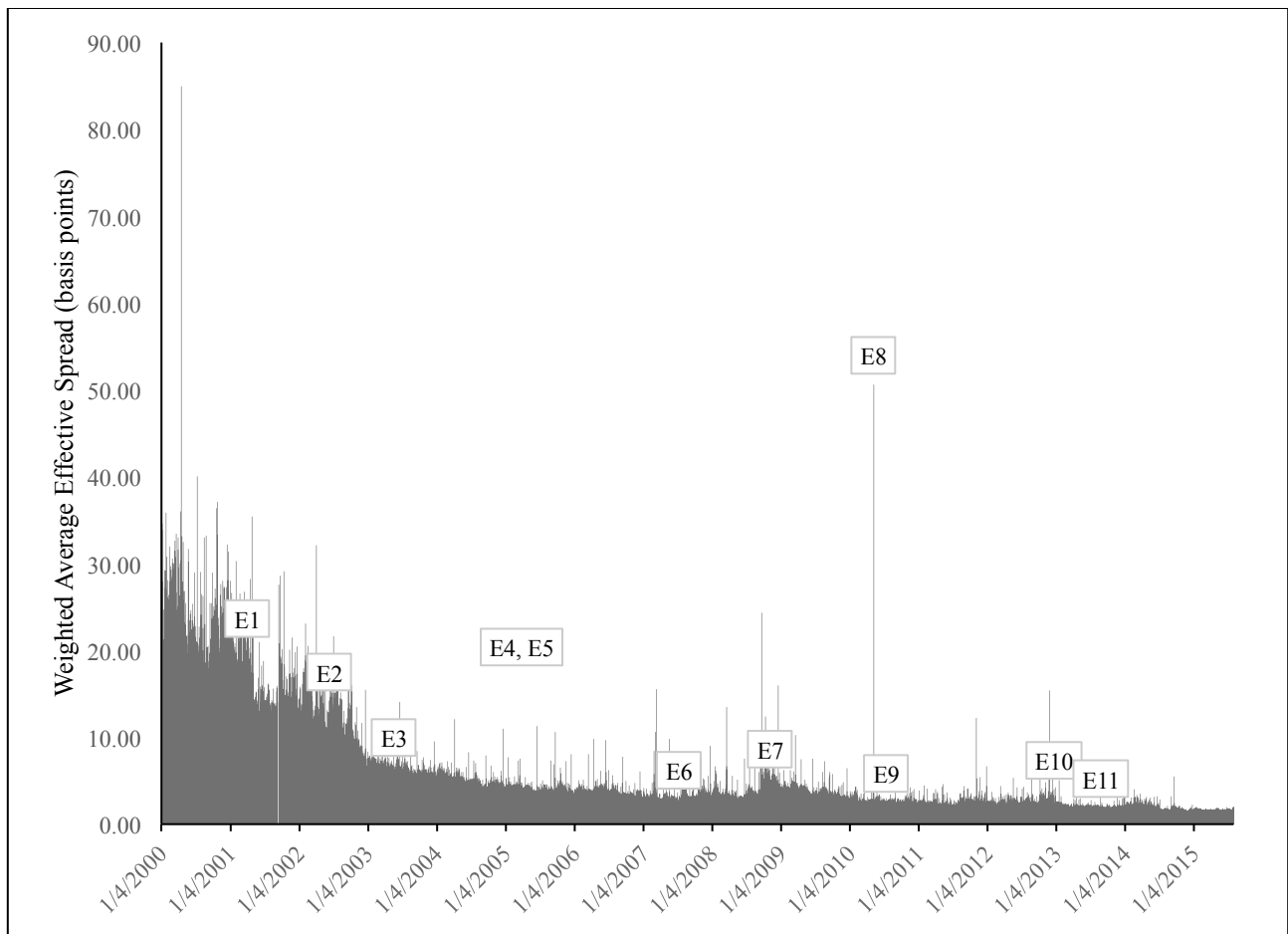| | Intercept | | | | | | Slope | | | | | |
	Sell Imbalance			Buy Imbalance			Sell Imbalance			Buy Imbalance		
	High	Med.	Low	Low	Med.	High	High	Med.	Low	Low	Med.	High
2010	4.98	2.23	1.64	1.50	2.06	5.12	-0.73	-0.51	-0.44	-0.43	-0.52	-0.77
	(0.29)	(0.09)	(0.05)	(0.13)	(0.23)	(0.44)	(0.01)	(0.02)	(0.02)	(0.02)	(0.03)	(0.03)
2011	3.96	1.76	1.31	1.32	1.75	3.83	-0.51	-0.31	-0.28	-0.26	-0.31	-0.51
	(0.03)	(0.02)	(0.01)	(0.01)	(0.02)	(0.04)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
2012	3.12	1.62	1.32	1.31	1.62	3.06	-0.36	-0.27	-0.27	-0.24	-0.27	-0.37
	(0.04)	(0.02)	(0.02)	(0.01)	(0.02)	(0.03)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
2013	3.05	1.47	1.14	1.18	1.50	3.01	-0.25	-0.19	-0.17	-0.17	-0.20	-0.27
	(0.03)	(0.02)	(0.01)	(0.02)	(0.02)	(0.03)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
2014	3.62	1.72	1.34	1.33	1.71	3.47	-0.37	-0.31	-0.28	-0.27	-0.31	-0.35
	(0.04)	(0.03)	(0.02)	(0.02)	(0.02)	(0.05)	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.01)
2015	3.63	1.49	1.52	1.98	1.47	4.00	-0.32	-0.22	-0.19	-0.20	-0.30	-0.35
	(0.03)	(0.05)	(0.47)	(0.63)	(0.03)	(0.15)	(0.01)	(0.02)	(0.01)	(0.01)	(0.02)	(0.01)
2010-2015	3.74	1.76	1.35	1.36	1.13	3.71	-0.44	-0.32	-0.29	-0.27	-0.32	-0.46
	(0.06)	(0.02)	(0.01)	(0.03)	(0.05)	(0.09)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Low VIX	3.31	1.60	1.26	1.37	1.61	3.31	-0.31	-0.25	-0.24	-0.23	-0.26	-0.32
	(0.03)	(0.02)	(0.02)	(0.02)	(0.02)	(0.04)	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.01)
Med. VIX	3.68	1.74	1.33	1.32	1.68	3.67	-0.45	-0.32	-0.29	-0.27	-0.32	-0.46
	(0.03)	(0.02)	(0.01)	(0.04)	(0.07)	(0.08)	(0.01)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)
High VIX	4.81	2.07	1.46	1.49	2.07	4.71	-0.65	-0.44	-0.38	-0.37	-0.45	-0.71
	(0.40)	(0.14)	(0.07)	(0.08)	(0.19)	(0.51)	(0.02)	(0.03)	(0.02)	(0.03)	(0.04)	(0.05)

Figure 1: The figure shows the dollar volume weighted average effective spread for all stocks in a day. The sample period is 2000-2015. The figure shows 10 important market structure events, labeled E1 through E11. These events are listed below.

E1: Full implementation of decimalization.
E2: Instinet purchase of Island ECN
E3: Full introduction of Autoquote on NYSE
E4: Nasdaq purchase of Instinet
E5: NYSE purchase of Archipelago
E6: Full implementation of Regulation NMS
E7: BATS converted from ECN to Exchange
E8: One day after Flash Crash
E9: Direct Edge converted to Exchange
E10: ICE purchase of NYSE Euronext
E11: BATS to merge with Direct Edge
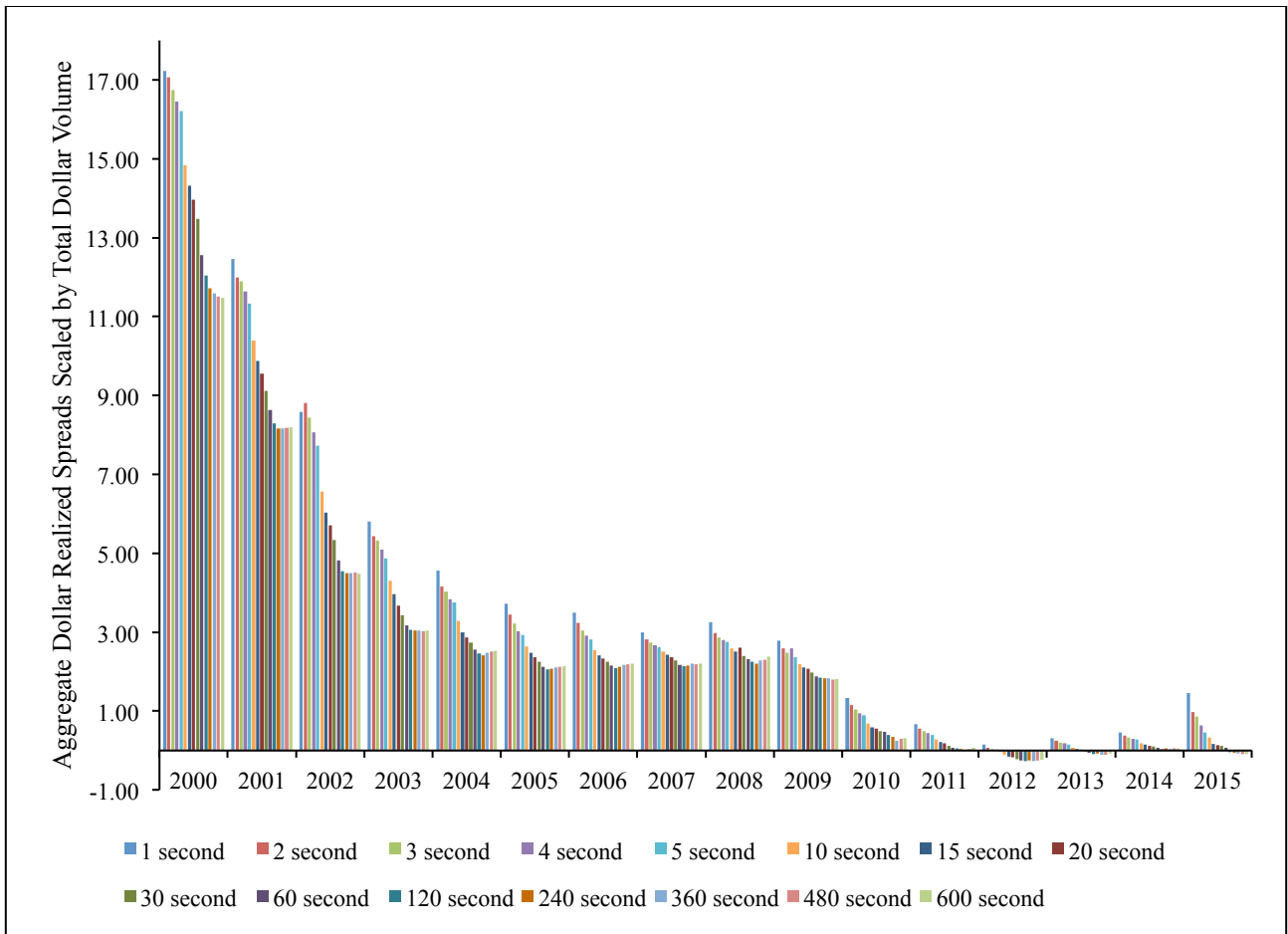
Figure 2: Dollar realized spreads for each trade are calculated at the horizons shown in the figure. Aggregate dollar realized spreads are scaled by total dollar volume and shown in the figure.
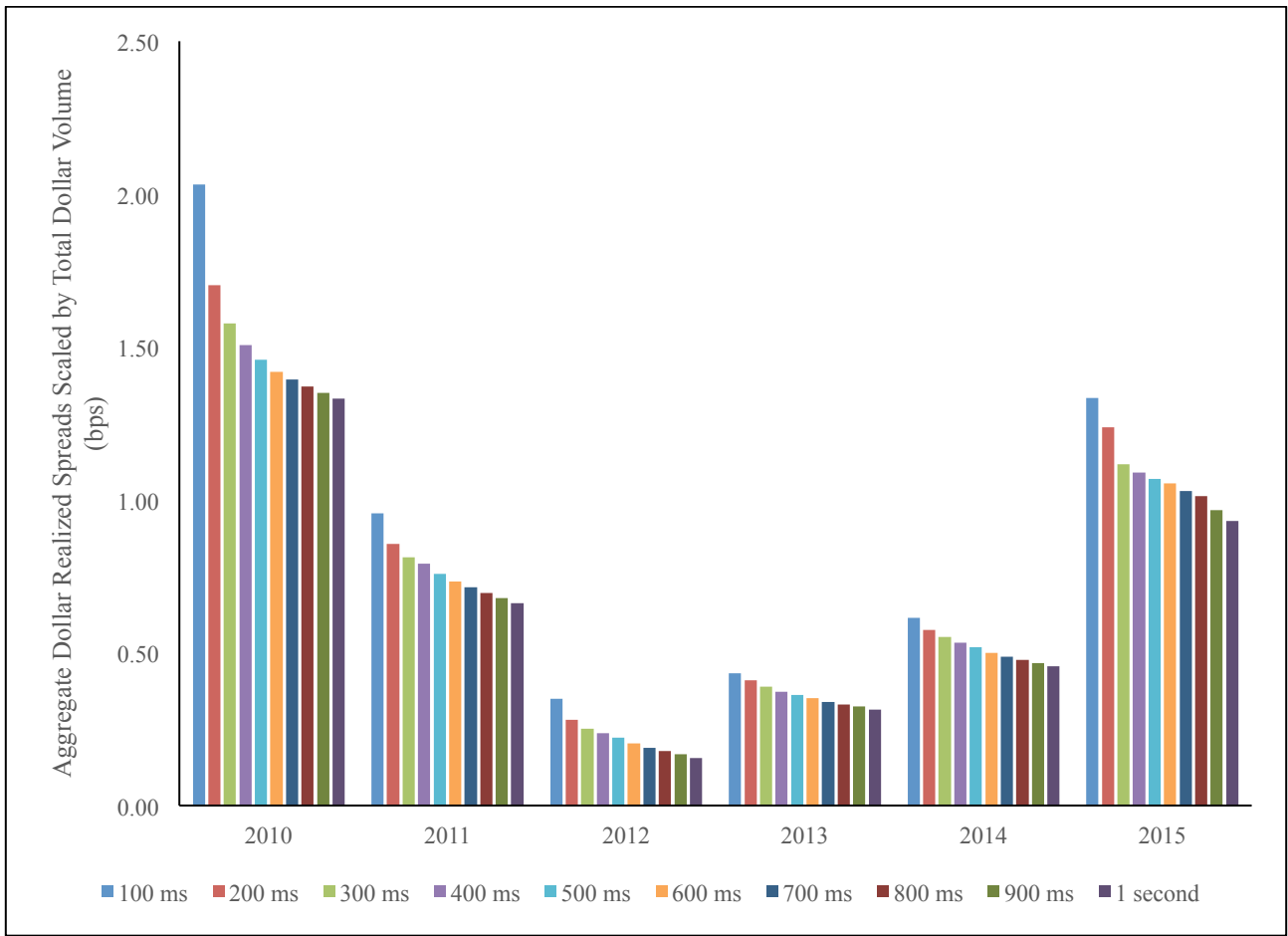
Figure 3: Dollar realized spreads for each trade are calculated at the horizons shown in the figure. Aggregate dollar realized spreads are scaled by total dollar volume and shown in the figure.
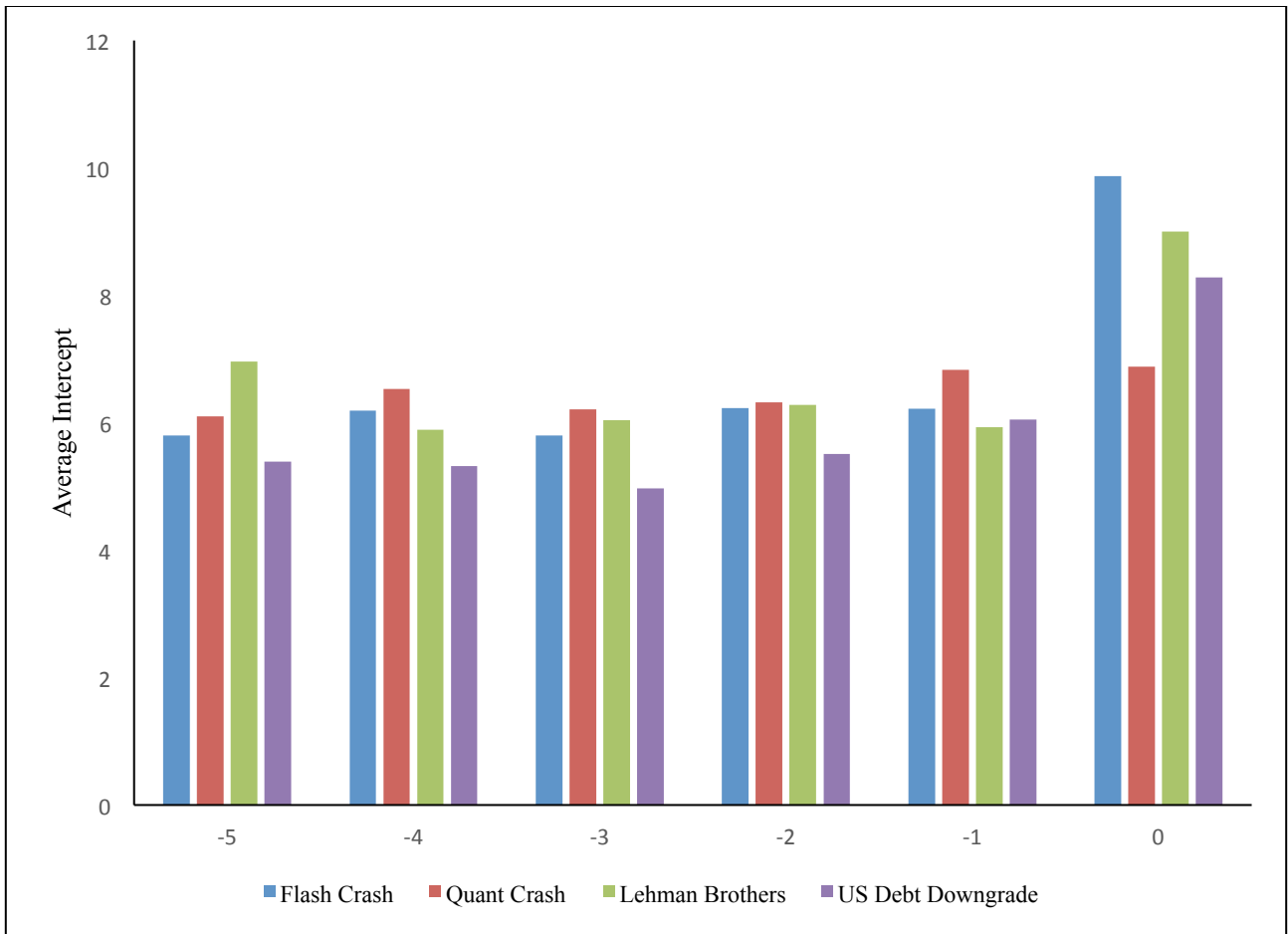
Figure 4, Panel A: For the Flash Crash (May 6, 2010), the Quant Crash (August 7, 2007), the Lehman Bankruptcy (September 5, 2008) and the US Debt Downgrade (August 5, 2011), the figure shows average intercepts across all stocks from day -5 to day 0.
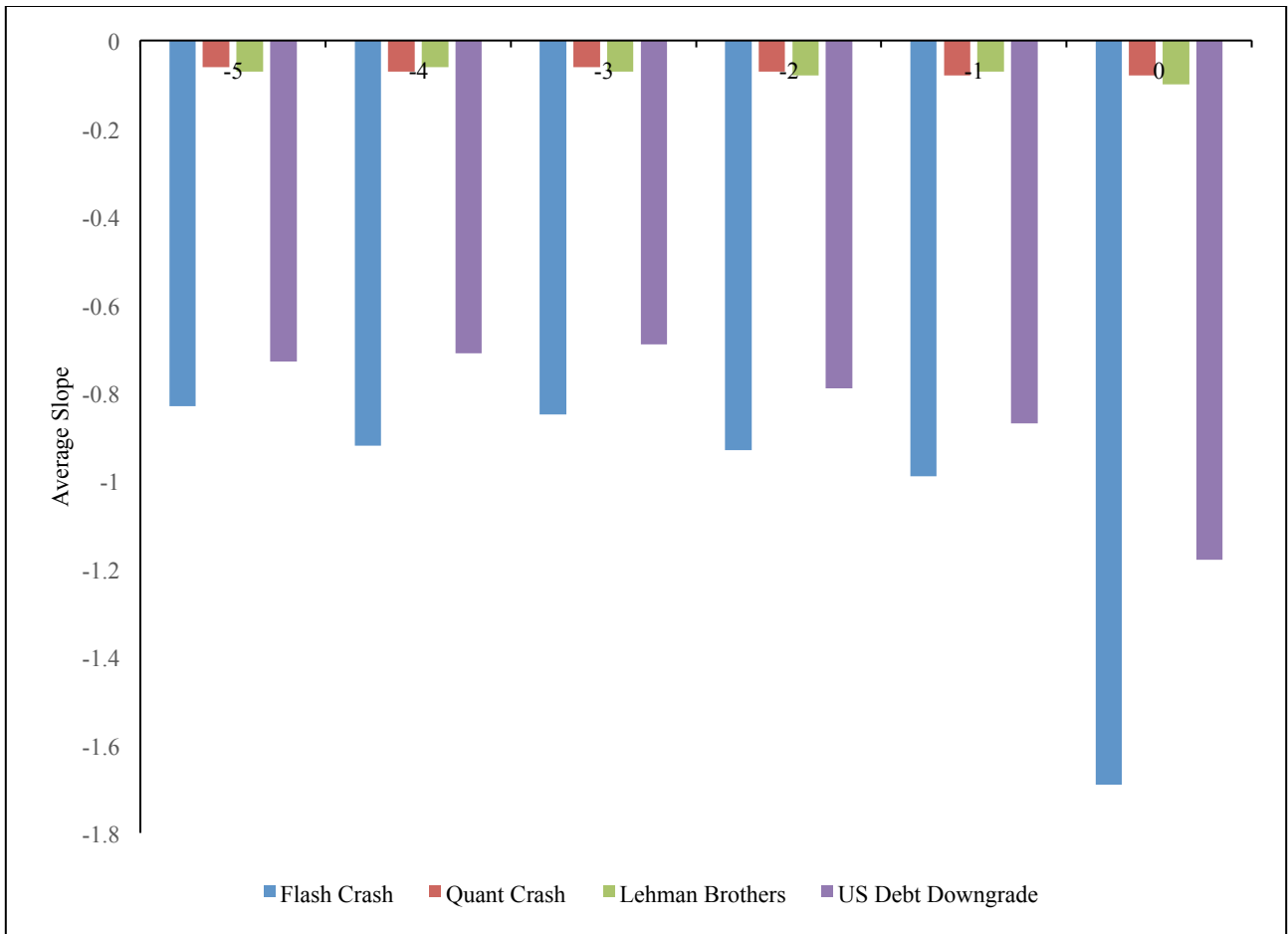
Figure 4, Panel B: For the Flash Crash (May 6, 2010), the Quant Crash (August 7, 2007), the Lehman Bankruptcy (September 5, 2008) and the US Debt Downgrade (August 5, 2011), the figure shows average slopes across all stocks from day -5 to day 0.