High Frequency Quoting:
Short-Term Volatility in Bids and Offers

Joel Hasbrouck*

November 13, 2012

This version: August 18, 2014

I have benefited from the comments of Fany DeKlerck, Ramo Gençay, Dale Rosenthal, Gideon Saar, Mao Ye, seminar/conference participants at Aalto University, Baruch College, BI Norwegian Business School, the Conference on Financial Econometrics (Toulouse), the Copenhagen Business School, the Emerging Markets Group (Cass Business School, City University London), the Euronext Paris Conference on High-Frequency Trading (April, 2013), l'Institute Louis Bachelier, Jump Trading, Norges Bank Investment Management, SAC Capital, UC Irvine, Society of Financial Econometrics (Toronto), the University of Illinois at Chicago, the University of Illinois at Champaign, the University of Pennsylvania Econometrics Seminar and Utpal Bhattacharya's doctoral students at the University of Indiana. All errors are my own responsibility.

DISCLAIMER: This research was not specifically supported or funded by any organization. During the period over which this research was developed, I taught (for compensation) in the training program of a firm that engages in high frequency trading, and served as a member (uncompensated) of a CFTC advisory committee on high frequency trading.

I am grateful to Jim Ramsey for originally introducing me to time scale decompositions.

*Department of Finance, Stern School of Business, New York University, 44 West 4th Street, New York, NY 10012 (Tel: 212-998-0310, jhasbrou@stern.nyu.edu).

High-Frequency Quoting:

Short-Term Volatility in Bids and Offers

Abstract

At horizons down to 50 ms, bids and offers in US equity markets exhibit volatility much higher than what is implied by long-term fundamentals. In examining the origins of this volatility, the findings suggest that competitive Edgeworth cycles are more likely than single-agent stuffing, spoofing, and experimentation activity or multiple-agent mixed-strategy behavior. To assess impact, the paper proposes a model wherein traders' random delays (latencies) interact with quote volatility to generate execution price risk and relative latency costs. The estimates imply that traders with latencies longer than 800 ms trade at a 1.8 bp disadvantage relative to faster traders. Finally, over the 2001-2011 period, despite high growth in quote traffic, quote volatility does not display a strong trend.

KEYWORDS: High-frequency trading; high-frequency quoting;

Recent developments in market technology have called attention to the practice of high frequency trading. The term is used broadly in reference to all sorts of fast-paced market activity, not just “trades”, but trades have certainly received the most attention. There are good reasons for this, as trades signify the actual transfers of income streams and risk. Quotes also play a significant role in trading process, however. This paper examines short-term volatility in bids and offers of US equities, a consequence of what might be called high frequency quoting.

By way of illustration, Figure 1 depicts the bid and offer for AEP Industries (a NASDAQ-listed manufacturer of packaging products) on April 29, 2011.[1] In terms of broad price moves, the day is not a particularly volatile one, and the bid and offer quotes are stable for long periods. The placidity is broken, though, by several intervals where the bid undergoes extremely rapid changes. The average price levels, before, during and after these episodes are not dramatically different. Moreover, the volatility is largely one-sided: the bid volatility is associated with an only moderately elevated volatility in the offer quote. Nor is the volatility associated with increased executions. These considerations suggest that the volatility is unrelated to fundamental public or private information. It appears to be an artifact of the trading process.

In the context of the paper’s data sample, the AEPI episode does not represent typical behavior. Nor, however, is it a singular event. It therefore serves to motivate the paper’s key questions. What is the extent of short-term quote volatility? What market practices give rise to it? What does it cost slower traders? Finally, given the current public policy debate surrounding low-latency activity, how has it changed over time?

Quote volatility is often attributed to the supposedly manipulative single-agent practices of quote-stuffing (canceling and submitting orders to produce congestion and/or confusion) and spoofing (briefly exposing quotes that are not intended for execution). Baruch and Glosten (2013) suggest that quote setters may be pursuing mixed (randomized) strategies. Their analysis builds on IO models of price randomization by sellers in product markets. These settings also sometimes exhibit Edgeworth cycles, wherein sellers incrementally undercut each other, reset to a high price, and repeat (Edgeworth (1925); Maskin and Tirole (1988); Noel (2011)). The mixed-strategy and Edgeworth cycle mechanisms offer rational competitive alternative explanations for quote volatility. This paper proposes a partial empirical resolution.

[1] The bid is the National Best Bid (NBB), the maximum bid across all exchanges. The offer is the National Best Offer (NBO), the minimum offer. They are often jointly referred to as the NBBO. Unless otherwise noted, or where clarity requires a distinction, “bid” and “offer” indicate the NBBO.

Quote volatility imposes costs and risks on liquidity demanders. The customary view is that bids and offers represent the terms of immediate trading opportunities, but in today's markets "immediacy" is hypothetical. All agents experience random latency in observing the quotes, formulating their responses, and communicating these decisions to the markets. For marketable orders quote volatility and random latency combine to create execution price uncertainty. This uncertainty extends beyond the marketable orders that are sent directly to the quoting venue: the NBBO in the "lit" US market establishes reference prices for dark trades, a category that includes roughly thirty percent of all U.S. equity trading volume.[2]

If latencies were identically distributed across agents, the execution price uncertainty might well be zero-mean and diversifiable. In practice, however, latency depends on proximity to the market, status (retail vs. institutional, or subscriber/member vs. public customer), and technology. Quote volatility magnifies the consequences of these differences: faster participants can condition on bid and offer states that, from the perspective of slower traders, are subsumed by noise. This paper proposes a model to measure relative latency costs.

Jarrow and Protter (2012), Foucault, Hombert and Rosu (2013), and Biais, Foucault and Moinas (2012) have recently proposed models in which speed confers an advantage. This advantage generally stems from more timely knowledge of fundamental information. The quote volatility considered in this paper comprises both fundamental and transient volatility. The model shows that relative speed can imply transfers from slow to fast traders even when the quote volatility is stationary (as in Figure 1).

This study estimates quote volatility in a broad sample of US equity market data using short-term variances centered about short-term averages of bids and offers. Given that market participants' definitions of "short-term" are likely to diverge, however, the analysis uses the flexible tools of time scale decomposition to estimate bid and offer volatility over horizons ranging from under 50 ms to about 27 minutes. In a 2011 sample, estimates suggest that at low-latency time-scales (roughly one second and lower) quote variances are two or three times the level that can be explained by fundamental movements, implying the presence of substantial stationary components. In a stylized model of trading latencies, the quote volatility estimates suggest that fast market-order

[2] Dark mechanisms do not publish visible bids and offers. They establish buyer-seller matches, either customer-to-customer (as in a crossing network) or dealer-to-customer (as in the case of an internalizing broker-dealer). The matches are priced by reference to the NBBO: generally at the NBBO midpoint in a crossing network, or at the NBB or the NBO in a dealer-to-customer trade.

traders (about one-second and faster) have an expected gain of about $0.003 per share (or about 1.8 basis points) relative to slower traders.

Using these estimates, the analysis reconsiders the mechanisms hypothesized to generate quote volatility. Single-agent and mixed-strategy models generally predict an inverse relation between competition and quote volatility. This paper finds that the empirical association is generally positive, suggesting that these mechanisms are not predominant. Edgeworth cycles are characterized by skewness in price changes (negative for bids, positive for offers). This study finds that quote volatility is associated with more extreme skewness, a result supportive of the Edgeworth mechanism.

To study quote volatility in historical data with truncation or rounding of time stamps to relatively coarse resolution, this paper proposes a novel simulation approach. Application to a sample of 2001-2011 data suggests that quote volatility displays no strong historical upward trend, in contrast to the growth in quote records and similar market data.

Two recent papers also focus on quote volatility. Egginton, Van Ness and Van Ness (2012) investigate quote stuffing, which they define as intense rates of order submission and cancellation, and which they proxy by number of quote updates. In a 2010 sample, they find that standard measures of liquidity are lower during quote-stuffing episodes. In a 2009-2011 sample, however, Conrad, Wahal and Xiang (2014) find that increased quoting activity is associated with improved efficiency and liquidity. The present study uses a different measure of quote volatility, different tools to assess costs and explanations for the source of the volatility, and, in parts of the analysis, a substantially longer sample.

The paper is organized as follows. The next two sections examine the economics of quote volatility, first from the perspective of liquidity suppliers (Section I), and then in terms of the costs borne by liquidity demanders (Section II). Section III connects the framework used to analyze liquidity demanders to the statistical tools used to measure quote volatility. Section IV describes the 2011 millisecond-stamped data used from the primary analysis, and the paper then turns to results: Section V analyzes the variance ratios, and Section VI discusses latency risk and cost estimates. Section VII presents an empirical analysis of the mechanisms generating quote volatility. The historical evidence on quote volatility from 2001 to 2011 is discussed in Section VIII. The connection to recent studies on high frequency trading is explored in Section IX. Section X summarizes the findings and concludes the paper.

I. The economic origins of quote volatility

Bid and offer quotes are set by liquidity suppliers. Why might they pursue volatile strategies? First note that any standard microstructure model can generate quote volatility if we introduce corresponding time variation in the underlying parameters. For example, if the half-spread in the basic Roll (1984) model represents a pass-through of the summary cost of market making, then variation in this cost will induce variation in the bid and ask. Alternatively, in a simple sequential trade model, we might conjecture time variation in the probability of an information event or the proportion of informed traders. This approach, though, has some obvious limitations. For such mechanisms to generate behavior like that depicted in Figure 1, the parameter must be rapidly oscillating, perhaps at sub-second periods. For clearing costs, interest rates, dealer risk aversion, and even informed trading probabilities or fundamental volatility, this sort of variation is implausible.

The field of possible explanations is wider if we consider settings in which the bid and offer are not simply determined in static equilibrium, but reflect instead a multi-move game in which quote updates occur in response to earlier moves by other players or as a device to trigger subsequent moves. Timing is then determined by the reaction speeds of the players (which are very high in the present market environment). The discussion will start with single-agent strategies, and then move on to competitive settings.

When there is one agent bidding or offering, or if the distance to the next best bid or offer is large, the agent has great latitude to maintain or vary the quote. Some practices that might arise in this case, such as quote-stuffing and some forms of spoofing, are often cited as major concerns associated with high frequency trading. Quote-stuffing is the rapid-fire cancellation and repricing of orders to impose higher costs or delays on agents who must process those quotes (Egginton, Van Ness and Van Ness (2012)). Spoofing is defined as entering a bid or offer for which the submitter does not actually desire an execution (U.S. Commodities Futures Trading Commission (2013)). Execution may be discouraged by exposing the quote for such a brief duration that access is essentially impossible. A seller, for example, might submit and quickly cancel an aggressive bid, perhaps to encourage existing bidders to raise their prices or to set a high reference price for a sale in a dark pool.

Single-agent strategies need not, however, have manipulative intent. A single seller, for example, might experiment with different offers in an attempt to discern demand elasticity (Leach and Madhavan (1992); Leach and Madhavan (1993)). In fact, Leach and Madhavan suggest that such experimentation will only occur in a market with a single (monopolistic) dealer. Bids and offers set in the course of experimentation may lose money in expectation, and so multiple dealers give rise to a free rider problem.

In the Leach and Madhavan framework, time is notional. Applying their insights in the present context, one might skeptically question what could be learned from an offer exposed for only, say, fifty milliseconds. The situation can easily be reframed, however. A single seller who experiments by monotonically decreasing the offering price might simply be running an impromptu clock auction.[3] The auction ends when the offer is hit ("mine") or when the seller's reservation price is reached. In the latter event, the seller might consider continuing to post the reservation price, but this would subject her to pick-off risk and the consequent need to monitor her offering price. If the latter costs are high, it may be preferable to withdraw from the market completely, re-entering with another auction at some later time when she may face a new set of buyers.[4]

When there are multiple bidders or offers, alternative explanations for quote volatility are suggested by dynamic models of price setting originally developed for non-financial markets. One line of analysis, following Edgeworth (1925), yields equilibria with price cycles ("Edgeworth cycles", see Maskin and Tirole (1988); Noel (2011)). Another set of analyses investigates mixed strategy equilibria. Varian (1980) considers a product market where at each revision opportunity each seller quotes a price randomly drawn from a stable distribution. Baruch and Glosten (2013) develop a model of a limit order market that exhibits similar equilibria.

In both Edgeworth and mixed strategy models, the driving force is the strategy of undercutting a competitor's price by a small amount. The classic Edgeworth cycle arises in a duopoly with a discrete price grid bounded from above. Starting from this upper bound, the producers alternately undercut each other's prices until the next lower feasible price would lead to a loss. At that point, the best response is to reset the price to the upper bound, and the process starts anew. The incremental undercutting followed by a jump gives a distinctive saw-tooth price path. Edgeworth cycles have been documented in retail gasoline markets. Although generally

---

[3] In a seller's clock auction, the price starts high and then descends at a constant rate until a buyer claims the lot. Also sometimes called a Dutch auction, it has long been used in the wholesale flower market at Aalsmeer (EconPort (2014), for example).

[4] Periodic single-price call auctions have been suggested to alleviate the perceived inequities of high-frequency trading (Budish, Cramton and Shim (2013); Schwartz and Wu (2013)). These proposals envision consolidated auctions coordinated by an exchange. Ad hoc clock auctions, however, can be run unilaterally, with no coordination or consolidation (other than guarantees of access).

studied in the context of competing sellers, Zhang (2005) has noted their occurrence in online advertising auction bidding.

Under slightly different assumptions (most importantly, a continuum of prices), Varian and Baruch and Glosten (BG) examine mixed-strategy equilibria. In the BG model identical sellers make rapid-fire draws of an offer price from a common distribution. The best offer at any given time is the minimum, that is, the first order statistic of the extant sample of draws. Until a trade occurs, the price distribution is time invariant. The sequence of best offers is therefore iid, in contrast to the dependent dynamic of the Edgeworth cycle. The assumption of a continuous price grid plays a major role in establishing the mixed-strategy equilibrium. Any pure strategy (or mixed strategy with mass points) invites a response that undercuts by an infinitesimal amount. At this lower price the responder captures the entire market, rendering the original proposer's strategy suboptimal. BG show that as the number of sellers increases, the offer price schedule converges to that predicted by the competitive (that is, non-strategic) equilibrium. They do not prove that this convergence is monotone, but for most common distributions (including the uniform, exponential and normal), as the sample size increases, the variance of the first order statistic drops. The conjecture that this is a general feature of the mixed strategy models suggests that quote volatility should be inversely related to competition among liquidity suppliers.

Edgeworth cycles may be identified by visual inspection of price plots. More objective identification criteria, though, may be derived from the skewness of the price changes. The sellers' saw-tooth price path involves numerous drops of small magnitude and a smaller number of large price increases, which implies a right-skewed distribution for price changes. Similarly, an Edgeworth cycle on the bid side is marked by numerous small price increases punctuated by large drops, which implies a left-skewed distribution.

These remarks establish the general empirical features of the two mechanisms. The details of their implementation in the empirical analysis are deferred to a later section.

II. The costs of quote volatility for liquidity demanders

A trader who transmits a market order does not achieve an execution until the order actually arrives at the market center. If the transmission delay (latency) is random, quote volatility induces corresponding execution price risk and also places the trader at a disadvantage relative to those with shorter delays.

The situation can be illustrated with a stylized model. Consider an offer price that is evolving over a trading session of eight periods ("seconds"). Market-order buyers are classified according to their time scale, ℓ, which characterizes their latency in the following sense. If a type-ℓ buyer transmits a marketable order at time t, the actual arrival time of the order at the market center is uniformly distributed on the interval $(t, t+\ell)$. The ℓ parameter thus summarizes both the mean and dispersion of the trader's latency. In a sense that will be made more precise below, a type-ℓ trader also possesses information specific to his type. A slow buyer has $\ell = 8$. Given the uniform distribution of his arrival time, he can expect to pay the mean offer price over the session.

The offer path is constructed as a residual, $R(t)$, relative to the session mean, with changes occurring at times $t = 0, 1, \ldots, 7$ (that is, at the beginning of each second). The residual is the sum of short-, medium- and long-term components, corresponding to levels of resolution indexed by $j = 1,2,3$. The component at level j is called the level-j detail and denoted by $D_j(t)$, so $R(t) = \sum_j D_j(t)$. Each $D_j(t)$ in turn is constructed as a random linear combination of Haar transforms. The Haar function is a one-period square wave, defined on the real line as $\psi(x) = +1$ if $0 < x < 1/2$, -1 if $1/2 < x < 1$, and 0 otherwise. In this example, the basis functions at level j are of the $\psi(t,j,k) = 2^{-j/2}\psi\left(2^{-j}x + k + 1\right)$. The full basis set consists of seven functions $\psi(t,j,k)$ for $j = 1, \ldots, 3$ and $k = 1, \ldots, 2^{3-j}$.

Figure 2 depicts this set. The top row contains the four short-run ($j = 1$) basis functions. Each is constant for one second, and then flips sign in the next second. There are four such functions, arranged to cover the eight-period interval without overlap. In the middle row, the medium-term basis functions ($j = 2$) are constant for two periods, flip sign over the next two periods, and also cover the full interval without overlap. In the bottom row, the long-term basis function ($j = 3, k = 1$) is constant for four periods and then flips sign. The duration over which a basis function maintains its positive or negative value is defined as the time-scale of the function, $\tau_j = 2^{j-1}$.

The seven functions constitute a basis for the eight unit segments that define $R(t)$. (Since the offer is being expressed as a deviation from the session mean, $R(t)$ integrates to zero, reducing the degrees of freedom by one.) It is clear by visual inspection that the functions are zero-mean and orthogonal. The amplitudes of the functions reflect a convenient normalization. When the two levels of $\psi(t,j,k)$ are set to $\pm 2^{-j/2}$, $\int \psi(t,j,k)^2 dt = 1$: the basis is orthonormal.

The choice of the Haar basis here is motivated by two considerations. Firstly, it affords a differentiation of the components by time scale that is clearer than would be possible with, say, an innovations representation. Secondly, the procedure used in this section to construct a hypothetical quote time series is essentially reversed in the empirical sections to obtain time-scale decompositions of actual bid and offer series.

The level-*j* detail is constructed as a random linear combination of the level-*j* basis functions: $D_j(t) = \sum_k a_{jk}\, \psi(t, j, k)$. The a_{jk} are random variables independently drawn from a distribution that may depend on *j* (but not *k*). Since all of the level-*j* basis functions have the same time scale, τ_j is also the time scale of the level *j* detail. From the orthogonality of the basis functions the integrated squared residual can be decomposed as the sum of the integrated detail squares: $\int R(t)^2 dt = \sum_j \int D_j(t)^2 dt$. Alternatively, the risk faced by the slow trader decomposes as $Var\big(R(t)\big) = \sum_j v_j^2$, where $v_j^2 \equiv 2^{-j} Var\big(a_{jk}\big)$.

For tractability, I now assume that the a_{jk} are drawn from a two-state distribution. Specifically, for a given value of v_j^2, $a_{jk} \in \{\pm 2^{j/2} v_j\}$ with probability 1/2 for each value. This implies $Var\big(a_{jk}\big) = 2^j v_j^2$, and generates $2^7 = 128$ equiprobable sample paths.

Recall that the slow $\ell = 8$ buyer pays the average price over the interval. Consider a relatively "fast" buyer whose latency is equal to the time scale of the long-term component, $\ell = \tau_3 = 2^2 = 4$ seconds. It is common in models of high-frequency trading to endow faster traders with predictive power. In this model, the informational advantage of an $\ell = \tau_j$ trader consists of observing a_{jk} at the start of the interval defined by the support of the corresponding basis function. The long-term component has one basis function (bottom row of Figure 2), which starts at $t = 0$ and flips sign at $t = 4$. The $\ell = 4$ buyer's optimal order submission time t^* is therefore either 0 or 4. If $a_{j=3,k=1} > 0$, the fast trader knows that the offer is relatively high in the first half of the session and relatively low in the second half, so he submits his buy order at $t^* = 4$. If $a_{j=3,k=1} < 0$, he submits his order at $t^* = 0$. Given the normalization of $\psi(t, j = 3, k = 1)$, his purchase price in either outcome is v_3 below the session average. That is, his incremental gain relative to the slow trader is v_3. He remains subject to short- and medium-term risk, $v_1^2 + v_2^2$.

Now consider a "faster" buyer whose latency is $\ell = \tau_2 = 2^1 = 2$ seconds, and whose optimal order submission time is $t^* \in \{0,2,4,6\}$. Like the $\ell = 4$ buyer, she observes $a_{j=3,k=1}$ and so can narrow her choices down to $t^* \in \{0,2\}$ or $t^* \in \{4,6\}$, implying that her order will arrive in $t \in (0,4)$ or $t \in (4,8)$. She also possesses information about D_2: at $t = 0$ she observes $a_{j=2,k=1}$, and at time $t = 4$, she observes $a_{j=2,k=2}$. If she has chosen to submit her order in $t^* \in \{0,2\}$, the value of $a_{j=2,k=1}$ guides her final choice: $a_{j=2,k=1} < 0 \Rightarrow t^* = 0$ and $a_{j=2,k=1} > 0 \Rightarrow t^* = 2$. If she has narrowed her choice down to $t^* \in \{4,6\}$, then $a_{j=2,k=2} < 0 \Rightarrow t^* = 4$ and $a_{j=2,k=2} > 0 \Rightarrow t^* = 6$.

Her incremental gain, relative to the fast trader is ν_2. She remains exposed to short-term risk.

Continuing, the "fastest" buyer is subject to a latency $\ell = \tau_1 = 2^0 = 1$ second. In addition to the information possessed by all slower buyers, he observes $a_{j=1,k=1}$ at $t = 0$, $a_{j=1,k=2}$ at $t = 2$, and so on. This allows him to time his submission to the optimal second. His incremental gain relative to the faster trader is ν_1.

Jarrow and Protter (2012), Foucault, Hombert and Rosu (2013), and Biais, Foucault and Moinas (2012) suggest that faster traders have advance knowledge of fundamental (that is, value-relevant) information. The volatility mechanisms considered in Section I, however, originate in the liquidity supply process. From a statistical perspective, fundamental information is permanently impounded in the in the random-walk component of security prices, whereas price movements created by the mixed-strategy and Edgeworth mechanisms are stationary and transient. The present specification of the price process is agnostic as to the sources of variation and can accommodate both types. The present model is similar to the aforementioned analyses in that it imbues the faster agents with predictive power. Their forecasting ability, though, is limited to sources of variation at their particular latency/time-scale. It is not necessary that they possess full knowledge of these components. Knowing a_{jk} at the point when the associated basis function first becomes nonzero suffices to determine the location of the minimum offer, at least up to the level-j trader's latency.

The empirical analysis conducted below produces estimates of the detail volatility, ν_j. The models suggests that ν_j can also be interpreted as the incremental expected gain of the level-j trader relative to the level $j+1$ trader, or as the incremental shortfall of the level $j+1$ trader. By implication then, $G_j = \sum_{i=1}^{j} \nu_j$ is the total shortfall of the level $j+1$ trader relative to all faster agents.[5]

[5] G_j possesses an additional interpretation. Most retail orders are executed by market makers who act as counterparty at a price equal to the prevailing bid or ask. If the timing of the prevailing bid or ask can't be verified, the market maker may try to choose the price within an interval that is most favorable to them. Stoll and Schenzler (2006) call this a look-back option. The faster market order buyers in the model are trying to find the minimum average offer. A market maker exploiting a look-back option tries to locate the maximum offer. The max and the min will occur, of course, at different times, but by the symmetry of the process the magnitudes are identical. Thus, G_j can be interpreted as the loss incurred by a market-order buyer trading against a market maker who enjoys a look-back option.

The relative gain associated with a latency differential is hypothetical, an amount that a fast trader might expect to gain versus a slower one, ceteris paribus. In equilibrium, however, these relative advantages might become real trading costs. Suppose that the entire population of market order buyers has latency $\ell = 1$ second, and that the average offer price yields zero-expected profit to sellers. If some traders reduce their latency below one second, that average offer will generate expected losses, as the faster traders are more successful at finding better prices. It might reasonably be conjectured that in equilibrium, sellers will raise their average offer price, thereby effecting a transfer from slow traders to fast traders, in the usual fashion. Broadening the scope of the analysis, there is no necessary implication that the faster traders' advantages are rents, particularly if entry is available to all who purchase the necessary technology.

In summary, timing uncertainty in this model has two effects. It gives rise to a zero-mean risk and to cost differentials, both of which depend on a trader's latency. The estimates presented later in this paper will give insight into the relative importance of these considerations. This evidence will suggest that the zero-mean risk is probably of second-order importance, but that the cost differentials are large enough to warrant attention.

III. Measuring quote volatility

The last section employed time-scale decomposition as a constructive tool, to generate a quote deviation series from primitives related to latency. Time-scale decomposition is more commonly used as a statistical device. Starting with a sample series, details are essentially computed as deviations between local means at different time scales, and their mean-squares are estimates of the detail variances. In principle the computations could be accomplished by any statistical package capable of estimating basic statistics for grouped data.

First proposed in Haar (1910), the Haar representation is now usually treated as a member of a broader class of functions called wavelets. Most statistical results, efficient computational algorithms, and connections to traditional time series analysis are developed in the broader framework of wavelet analysis. Percival and Walden (2000) provide a comprehensive text-book development.[6,7]

Most of the notation and terminology developed thus far conforms to the Percival and Walden usage, but there are two exceptions. The ν_j^2, termed "detail variances" in section II, are more generally called "wavelet variances". Also, $R(t)$, the sum of the detail processes, was described in section II as a "residual". In the wavelet or signal processing setting it is referred to as a "rough". A time scale decomposition generates a family of roughs, computed as sums of the detail processes: $R_1(t) \equiv D_1(t); R_2(t) = D_1(t) + D_2(t); R_3(t) = D_1(t) + D_2(t) + D_3(t)$; and so on.

The statistical framework can be summarized as follows. The series of price levels, p_t, is assumed to possess stationary first differences. (Note that while the stationarity condition is imposed on the first differences, the computations are performed on levels, as in section II.) This condition suffices to define the rough and detail variances $\sigma_j^2 \equiv Var\left(R_j\right)$ and $\nu_j^2 = Var\left(D_j\right)$. In the Section II model, the arrangement of the basis vectors in time corresponds to a discrete wavelet transform (DWT). Although estimates of σ_j^2 and ν_j^2 can be based on sample DWTs, better estimates

---

6 Time scale and multi-resolution decompositions are widely used across many fields. In addition to Percival and Walden, Gençay, Selçuk and Whitcher (2002) discuss economic and financial applications in the broader context of filtering. Nason (2008) discusses time series and other applications of wavelets in statistics. Ramsey (1999); (2002) provides other useful economic and financial perspectives. Walker (2008) is clear and concise, but oriented more toward engineering applications.

7 Studies that apply time scale decompositions in the economic analysis of stock prices loosely fall into two groups. The first set explores time scale aspects of stock comovements. A stock's beta is a summary statistic that reflects short-term linkages (like index membership or trading-clientele effects) and long-term linkages (like earnings or national prosperity). Wavelet analyses can characterize the strength and direction of these horizon-related effects Gençay, Selçuk and Whitcher (2005); In and Kim (2006). Most of these studies use wavelet transforms of stock prices at daily or longer horizons. A second group of studies uses wavelet methods to characterize volatility persistence Dacorogna, Gencay, Muller, Olsen and Pictet (2001); Elder and Jin (2007); Gençay, Selçuk, Gradojevic and Whitcher (2010); Gençay, Selçuk and Whitcher (2002); Høg and Lunde (2003); Teyssière and Abry (2007). These studies generally involve absolute or squared returns at minute or longer horizons. Wavelet methods have also proven useful for jump detection and jump volatility modeling Fan and Wang (2007). Beyond studies where the focus is primarily economic or econometric lie many more analyses where wavelet transforms are employed for ad hoc stock price forecasting Atsalakis and Valavanis (2009); Hsieh, Hsiao and Yeh (2011, for example). An early draft of Hasbrouck and Saar (2013) used wavelet analyses of message count data to locate periods of intense message traffic on NASDAQ's Inet system.

are generally obtained using a variant called the maximal overlap discrete wavelet transform (MODWT). All estimates in this paper employ the MODWT.[8]

The empirical analysis reports estimates of σ_j, ν_j, and the implied gain G_j in cents per share and basis points. For some purposes, though, it is useful to compute statistics that summarize the relative importance of the stationary and random-walk components of the price. This is achieved by variance ratios.

There is a long tradition of variance ratios in empirical market microstructure (Amihud and Mendelson (1987); Barnea (1974); Hasbrouck and Schwartz (1988, among others), among others).[9] Let *K* periods denote some suitably "long" horizon, and *k* periods a shorter horizon. A typical variance ratio compares the variance per period implied at these two time scales:

$$V\Delta_k = \frac{Var(\Delta_k p_t)/k}{Var(\Delta_K p_t)/K} \tag{1}$$

where Δ_k is the *k*-period differencing operator $\Delta_k p_t = p_t - p_{t-k}$. If p_t follows a random-walk at all horizons, $V\Delta_k = 1$ for all *k*. In typical microstructure samples, variance ratios with $k<<K$ are generally elevated due to short-term microstructure effects. The motivation for using a long-term variance in the denominator of $V\Delta_K$ is the desire for an estimate of fundamental volatility, on the assumption that a long-term price change is dominated by informational components.

The end-points of the *K*-period price change remain subject, however, to microstructure effects just as strong (in absolute terms) as those of the short-term price change. That is, both p_t and p_{t-K} are subject to bid-ask bounce, discreteness effects, and so on. A long-term wavelet variance, however, is in principle purged of the short-term variation, and so may serve as a better estimate of fundamental long-term variance. Fan and Gençay (2010) apply this principle to unit root tests based on time scale decompositions. Gençay and Signori (2012) explore the use of variance ratios at different time scales to test for serial correlation. The variance ratios used here are special cases or minor modifications of theirs.

---

[8] The DWT is sensitive to alignment. In the section II example, if we constructed a sample by successively replicating the first eight periods, estimates of σ_j^2 and ν_j^2 would converge to their population values. If the sample were somehow shifted (by dropping the first observation, for example), the alignment of the estimated process would no longer conform to that of the data generating process. Estimates based on the MODWT essentially average over all possible alignments.

[9] Return variance ratios are also used more broadly in economics and finance to characterize deviations from random-walk behavior over longer horizons (Lo and MacKinlay (1988)).

The appendix develops a wavelet variance ratio of the form

$$V_j = \frac{\nu_j^2/\tau_j}{\nu_J^2/\tau_J} = 2^{J-j}\frac{\nu_j^2}{\nu_J^2} \tag{2}$$

where J is the highest level (corresponding to the longest time scale) in the analysis and $j < J$. The corresponding time scales are $\tau_j = 2^{j-1}$ and $\tau_J = 2^{J-1}$, which implies the second equality. Like the price-change variance ratio, the wavelet variance ratios for a random walk are one for all j. Wavelet variances measure variation at a particular time scale. Rough variances measure variation at and below given time scale. A variance ratio can also be based on the rough variance:

$$VR_j = 2^{J-j-1}\frac{\sigma_j^2}{\nu_J^2} \tag{3}$$

Note that while the numerator is a rough variance, the denominator is a wavelet variance. For a random walk, $VR_j = 1$ for all j. A finding of $VR_j > 1$ indicates elevated short-term volatility. (Whereas $V_J = \nu_J^2/\nu_J^2 = 1$ by construction, however, $VR_J = \sigma_J^2/2\nu_J^2$, which need not equal one.) The appendix contains further details.

Security prices at all horizons are a mix of integrated and stationary components. The former are usually identified with persistent fundamental information innovations; the latter, with transient microstructure effects. The former are important to long-term hedging and investment; the latter, to trading and market-making. The dichotomy is sometimes reflected in different statistical tools and models.

In the somewhat distinct literatures that focus on integrated and stationary volatilities, the greatest common concerns arise in the analysis of realized volatility (Andersen, Bollerslev, Diebold and Ebens (2001); Andersen, Bollerslev, Diebold and Labys (2003a); Andersen, Bollerslev, Diebold and Labys (2003b)). Realized volatilities (RVs) are calculated from short-term price changes. They are useful as estimates of fundamental integrated volatility (IV), and typically serve as inputs to longer-term forecasting models. RVs constructed directly from trade, bid and offer prices are typically noisy, however, due to the presence of microstructure components. Local averaging moderates these effects [see Hansen and Lunde (2006) and accompanying comments]. Other approaches are discussed in Aït-Sahalia, Mykland and Zhang (2011); Zhang (2006); Zhang, Mykland and Aït-Sahalia (2005). There is a methodological connection here, in that long-term wavelet variances are computed from short-term averages, much like the pre-averaged inputs to realized volatility.

The present study draws on several themes in the RV literature. The variance ratios employed here serve a purpose similar to the volatility signature plots introduced by Fang (1996), and used by Andersen, Bollerslev, Diebold, and Ebens (2002) and Hansen and Lunde (2006), among others. Hansen and Lunde also articulate the connection between bid-offer comovement and fundamental volatility: since the bid and offer have economic fundamentals in common, divergent movements must be short-term, transient, and unconnected to fundamentals.

One strand in the RV literature emphasizes analysis of multiple time-scales. Zhang, Mykland and Aït-Sahalia (2005) posit a framework consisting of a Brownian motion with time-varying parameters, $dX_t = \mu_t dt + \sigma_t dz$, and a discretely-sampled noisy observation process, $Y_{t_i} = X_{t_i} + \varepsilon_{t_i}$. The Y_{t_i} are viewed as transaction prices, and ε_{t_i} constitute i.i.d. microstructure noise. The objective is estimation of the integrated volatility $\int \sigma_t^2 dt$ over a sample. They propose a two-scale variance estimator in which a long-scale estimate is corrected for bias with an adjustment based on properties of the noise estimated at a short scale. While the present analysis also features multiple time scales, there are major differences in the perspective. In the present situation, execution price risk is caused by volatility in the observed process (the quote, not the underlying latent value); the quote process is right-continuous (and continuously observable); the noise is not necessarily i.i.d. (cf. the AEPI episodes in Figure 1); and, the noise is possibly correlated with the X_t increments.

The paper also departs from the RV literature in other respects. The millisecond time scales employed in this paper are several orders of magnitude shorter than those typically encountered. Most RV studies also focus on relatively liquid assets (index securities, Dow-Jones stocks, etc.). The low-activity securities included in the present paper's samples are important because, due to their larger spreads and fewer participants, they are likely to exhibit relatively strong, persistent and distinctive microstructure-related components.

IV. Sample and data

The analyses are performed for a subsample of US firms using quote data from April, 2011 (the first month of my institution's subscription.) The subsample is constructed from all firms present on the CRSP and TAQ databases from January through April of 2011 with share codes of 10 or 11, with closing prices between two and one thousand dollars, and with a primary listing on the

New York, Amex or NASDAQ exchanges.[10] I compute the daily average dollar volume based on trading in January through March, and randomly select 15 firms from each decile. For brevity, reported results are grouped into quintiles.

The U.S. equity market is highly fragmented, but all exchanges post their quotes to the Consolidated Quote System (CQS).[11] The CQ and NBBO files from the NYSE's daily TAQ dataset used here are definitive transcripts of the consolidated activity, time-stamped to the millisecond.[12] A record in the consolidated quote (CQ) file contains the latest bid and offer originating at a particular exchange. If the bid and offer establish the NBBO this fact is noted on the record. If the CQ record causes the NBBO to change for some other reason, a message is posted to another file (the NBBO file). Thus, the NBBO can be obtained by merging the CQ and NBBO files. It can also be constructed (with a somewhat more involved computation) directly from the CQ file. Spot checks confirm that these two approaches are consistent.

Studies involving TAQ data have traditionally used error filters to suppress quotes that appear spurious. Recent daily TAQ data, though, appear to be much cleaner than older samples. In particular, the NBBO construction provided by the NYSE clearly defines what market participants would have perceived. Some quotes present in the CQ file are not incorporated into the NBBO because they are not firm, indicative or otherwise deemed "not NBBO-eligible". Beyond these exclusions, however, I impose no additional filters for the estimates discussed in this section. Error filters are used, however, in the subsequent historical analysis, and will be discussed in greater detail at that point.

Table I reports summary statistics. Post-Reg NMS US exchanges have become more similar in structures and trading mechanisms. With respect to listing characteristics, though, differences persist. The NYSE "classic" has the largest proportion of high-volume stocks, NYSE Amex has the smallest, and NASDAQ falls in the middle. In some instances, a stock that is present in CRSP and

---

[10] The American Stock Exchange merged with NYSE Euronext in 2008, and was renamed NYSE Amex LLC. In May, 2012, the name was changed to NYSE MKT LLC. For the sake of clarity, it is identified here simply as "Amex".

[11] At the same time that an exchange sends a quote update to the consolidated system, it can also transmit the update on its own subscriber line. For subscribers this can reduce the delay associated with consolidation and retransmission (which is on the order of about five milliseconds). Thus, while the CQS is a widely-used single-source of market data, it is not the fastest. Moreover, bids and offers with sizes under 100 shares are not reported.

[12] The "daily" reference in the Daily TAQ dataset refers to the release frequency. Each morning the NYSE posts files that cover the previous day's trading. The Monthly TAQ dataset, more commonly used by academics is released with a monthly frequency and contains time stamps in seconds.

TAQ's master file is absent from a day's quote file. Stocks missing more than half of the days in the sample are dropped.

Market event counts (trades, quotes, and so forth) display some interesting patterns. There are large numbers of quote records, since one is generated when any market center changes its best bid, best offer, or size at the bid or offer. If the action establishes the bid and offer as the NBBO this fact is noted on the quote record. But if the action causes some other change in the aggregate prices or sizes at the NBBO, an NBBO record is generated. Since many quote records don't induce such a change, there are substantially fewer NBBO records. Finally, many actions might change one of sizes or one side of the quote. Thus, the numbers of NBB and NBO changes are smaller yet.

Volatility and spreads tend to be elevated at the start and end of trading sessions (9:30 to 16:00). To remove the effect of these deterministic effects, I confine the variance estimates to the 9:45 to 15:45 subperiod. Estimates are computed separately for the bid and offer, and then averaged for convenience in presentation. Most samples are winsorized at $\pm$ 5% prior to computation of statistics.

V. Variance ratio estimates

Variance ratios provide a familiar point of departure. Table II summarizes sample averages for the wavelet, rough, and the conventional price-change variance ratios for all time scales. Table III reports results for a subset of time scales (50 ms, 800 ms, and 27.3 minutes), but with standard errors, and for subsamples constructed as quintiles in average dollar trading volume.

The pattern across time scales is similar across all ratios. At shorter time scales, they are substantially greater than one, indicating elevated short-term volatility. At a time scale of 50 ms., all three ratios are in the vicinity of three, suggesting that quotes at this time scale are about three times what one would expect from a random-walk calibrated to 27.3 minute volatility. These estimates attest to the importance of stationary (non-fundamental) volatility as a latency-related concern for traders. As the time scale increases, the means of all three ratios drop towards unity.

Both the wavelet and rough variance ratios are normalized by the same denominator, the wavelet variance over 27 minutes. The numerator of the wavelet variance ratio measures variance only at the indicated time scale. At a time scale of 800 ms ($j = 5$), for example, $V_j = 1.975$ indicates that stochastic components varying at that time scale are 1.975 times the value implied by a random walk. The numerator of the rough variance ratio additionally captures volatility from components at all shorter scales, so at $VR_j = 2.061$, it is slightly higher. At the longest time scale,

$V_{j=16}$ is one by construction, the average $VR_{j=16} = 1.074$, indicating that volatility from all shorter time scales is 7.4% higher than expected (relative to a random walk).

The computation of the price-change variance ratio $V\Delta_j$ differs fundamentally from that of the wavelet and rough variance ratios, and sample values could in principle diverge substantially. In fact, however, the average $V\Delta_j$ is close to the average V_j and VR_j at all time scales (except for $VR_{j=16}$, at the longest time scale). This can be explained by noting that, relative to the wavelet variances in V_j, the price-change variances in $V\Delta_j$ are inflated by shorter-term components. This inflation occurs, though, in both the numerator and denominator of $V\Delta_j$, thus working in offsetting directions.

Across volume quintiles, short-term volatility elevation is much more pronounced in the low-volume firms. At 50 ms., the variance ratios for the lowest volume quintile are between three and four; for the highest volume quintile, about 1.3. This suggests that the spreads for small firms are not only wider on average, but the bid and offer quotes are substantially more volatile.

In an earlier study, Hansen and Lund analyze realized volatility based on bid and offer changes at time scales down to one second Hansen and Lunde (2006). Their Figure 1 depicts realized volatility profiles for tickers AA and MSFT, for 2001 and 2004. For AA in 2004, the RVs are about 2.5 at one second and about 2.2 at 1,200 seconds (20 minutes). This implies a one-second price-change variance ratio of about 1.14. For MSFT in 2004, the one-second and 20-minute RVs are approximately equal, implying a ratio around unity. In the present sample, both AA and MSFT lie in the top volume quintile. The average $V\Delta_j$ at a time scale of 1,800 ms. is 1.146, which is quite similar to the Hansen and Lund values.

In all volume quintiles, as the time scale drops, the variance ratios increase. It might be conjectured that this an inevitable consequence of the way these ratios are defined. To pursue this case, consider a price-change variance ratio computed for a transaction price modeled as a random-walk plus uncorrelated bid-ask bounce. In this case, the price change sampled at interval length h is $\Delta_h p_t = p_t - p_{t-h} = u_t + \epsilon_t - \epsilon_{t-h}$ where u_t is the fundamental (random-walk) price change over the interval and ϵ_t is the bid-ask bounce component at time *t*. If ϵ_t and ϵ_{t-h} are assumed independent, then Zhang, Mykland and Aït-Sahalia (2005) point out that $Var(\Delta_h p_t) = h\sigma_u^2 + 2\sigma_\epsilon^2$, and that in the limit of higher frequency sampling , as $h \to 0$, $Var(\Delta_h p_t) \to 2\sigma_\epsilon^2$. In this limit the short/long variance ratio $Var(\Delta_h p_t)/h\sigma_u^2 \to \infty$.

The driving force behind this divergence, though, is the assumption that the sample is infinitely dense in transactions, i.e., that at the beginning and end of any interval, no matter how short, we can find two trades with independent bid-ask bounce terms. In the case of a frequently-traded stock, this is a tenable assertion with one-minute or even one-second sampling. It is less plausible with 50-ms sampling, or for a low-volume stock.

Bid and asks are step functions, with a finite number of transitions in any given sample. Intuitively, in computing a sample variance of the changes at progressively smaller h, once we move below the resolution at which each change is contained in its own interval, the sum of squares stays fixed. Denote this quantity by $\overline{SS}$. For a fixed clock-time sample of length T, however, the number of intervals in the sample is T/h. The estimated sample variance per interval is $\overline{SS}/(T/h)$, which converges to zero with h. The price-change variance ratio converges to $\overline{SS}/T\sigma_u^2$.

In this situation, the wavelet variance ratio is more precise than the standard price-change ratio. In the process of refining the intervals, we will arrive at a point where the sample wavelet variance is zero (because all changes are captured at longer time scales). Because the wavelet variance ratio is time-scale specific, it will be zero if there is in fact no variation at that time scale.

VI. Volatilities and related estimates

Table IV summarizes the sample means for wavelet volatilities, rough volatilities, and implied cumulative gains in units of cents per share and basis points for all time scales. Table V reports similar statistics for three representative time scales, but with standard errors and also in volume-based subsamples.

The entries in the Table IV row for level $j = 5$, for example, indicate that a market-order trader facing arrival time uncertainty of 800 ms is exposed to wavelet (detail) volatility of ν_5 =\$0.00105 per share, or 0.452 basis points (0.00452%), from variation at an 800 ms time scale. From sources of variation at that time scale and shorter, the trader is exposed to cumulative risk of $\sigma_5 = \$0.00152$ per share, or 0.910 basis points (0.00910%). In moving from shorter to longer time scales, sample means for both σ_j and ν_j increase. A trader whose order arrives at some time within a 27.3-minute window will incur a risk of $\sigma_{j=16} = 28.990\ bp$.

How should the economic significance of these magnitudes be assessed? Many trading fees (such as commissions and clearing fees) are assessed on a per share basis. Access fees, the charges levied by exchanges on taker (aggressor) sides of executions are also assessed per share. US SEC Regulation NMS caps access fees at \$0.003 per share, and in practice most exchanges are close to this level. Exchanges also pay liquidity rebates that transfer a portion of the access fee to the maker (passive) side of the trade, generally about \$0.002 per share. Variations in these fees are regarded as important determinants of order routing decisions. As an alternative benchmark, a recent survey estimates average institutional commissions for US equities at about six basis points (ITG (2014)). The volatilities of latency risk are of comparable magnitude.

If the latency risk is zero-mean, however, the direct economic significance is small. Consider a hypothetical investor with CRRA utility $U(W) = W^{1-\gamma}/(1-\gamma)$ and $\gamma = 2$. The investor holds a risky investment with annual return relative R where $Log(R)$ is normally distributed with mean 0.08 and standard deviation 0.20. The investor buys the investment, holds for one year, and then sells. Latency risk that is normally distributed with mean zero and standard deviation of five basis points (roughly five times larger than the σ_5 estimate) is incurred on both trades. Numerical calculations suggest that relative to the case with no latency risk, the drop in expected utility corresponds to a relative initial wealth penalty of only 3×10^{-4} percent (which is roughly at the limit of computational precision). For a frequent trader, of course, the impact of latency risk might be larger. But even if the investor turns over his portfolio daily, buying and selling in each of a year's 250 trading days, the equivalent wealth penalty is only about one basis point, an amount that would surely be swamped by other trading costs.

Following the model presented in Section II, these estimates can be mapped to costs of latency. It should be borne in mind, however, that while the estimates of σ_j and ν_j are computed in a fairly robust statistical framework, the model mapping these estimates to costs is stylized, and involves a number of additional assumptions about the data generating process, the information possessed by traders, and their strategies. Cost estimates derived from this model should therefore be viewed more tentatively.

The timing model implies that the wavelet volatility ν_j can also be interpreted as the time-scale-specific gain, that is, the relative advantage held by a trader subject to a random latency bounded at $\ell = \tau_j = 50 \times 2^{j-1}$ ms versus a trader who is slower by a factor of two. The cumulative gain G_j measures the relative advantage held by all traders at time scale τ_j and faster. For example, a trader whose latency is $\ell > 800$ ms loses \$0.00303 per share or 1.817 bp to faster traders. The utility loss to slower ($\ell > 800$) traders is small, but non-trivial. If we rework the utility calculations with a twice-a-year cost that is normally distributed with mean of 1.817 and standard deviation 0.910 bp, the implied initial wealth penalty is about 3.7 bp, which is about two-thirds of the average institutional commission. Costs to frequent traders, of course, could be much larger.

Table V reports estimates in trading-volume quintile subsamples. Taking the 800 ms estimates as representative, in moving from the low- to high-volume quintiles, the cumulative gains measured in $0.01 per share tend to increase, while those measure in basis points tend to decrease. Low-volume stocks tend to have high relative volatility and latency costs, with the latter on the order of two or three basis points. Some of the cumulative gain estimates at the longest time scale are quite large: 106.455 bp (1.06 percent) for the lowest volume quintile, or $0.20711 per share for the highest volume quintile. These represent gains, however, relative to traders who are subject to latencies above 27.3 minutes. At such horizons, the assumptions of the timing model, particularly predictability, are more suspect.

Hansen and Lunde (2006) note that to the extent that volatility is fundamental, we would expect bid and offer variation to be perfectly correlated, that is, that a public information revelation would shift both prices by the same amount. On this point, Tables IV and V also report, in the last column, the average wavelet correlation between the bid and offer, i.e., the correlation between bid and offer detail components at a given time scale. In the full sample (Table IV), for example, the correlation between bid and offer details at 800 ms. is 0.471. This is positive (bids and offers do tend to move in the same direction), but it is at best only moderately positive. The average correlation increases with time scale, reaching 0.896 at 27.3 minutes. The corresponding estimates for volume quintile subsamples, however, display great dispersion (Table V). For the lowest quintile, the 50-ms correlation is near zero (at 0.060), 0.166 at 800 ms, and only 0.532 at 27.3 minutes. At sub-second time scales for these stocks, bid and offer movements are essentially uncorrelated. In moving to higher volume subsamples, correlations increase. At the shortest time scale (50 ms), the correlation in the highest volume subsample is 0.553. In general, the weak correlations affirm that quote volatility at short time scales is primarily transient.

VII. The strategies of liquidity suppliers

Section I discussed volatility originating from three actors: single agents pursing stuffing, spoofing, and experimentation/auction strategies; multiple agents following mixed strategies (Baruch and Glosten, BG); and multiple agents generating Edgeworth cycles. This section suggests and implements empirical tests to detect and resolve these mechanisms.

A. Competition

The earlier discussion establishes that the single agent and mixed strategy mechanisms generally suggest a negative relation between volatility and the number of agents competing to set the best bid and offer. “Competing” in this context means that the agent’s quote is at or near the best much of the time, that the agent is closely monitoring market conditions, and that she is ready to quickly revise her quote. Agents whose limit orders are deep in the book or are infrequently monitored are not considered competitive. By definition, the single-agent strategies giving rise to quote volatility only when one agent has the market power to set and change the quote. If there are two or more active quote setters, the Baruch-Glosten model predicts that they will follow mixed strategies. The conjecture that the variance of the minimum or maximum draw from a common distribution generally decreases with the sample size suggests that the variance of the best offer or best bid should also drop as the number of competitors increases.

The TAQ do not identify individuals, and so cannot measure the number of competing quote setters. The data do identify exchanges, however. Exchanges partially aggregate individual activity, and so exchange competition may plausibly proxy for competition among individuals. This study therefore uses Herfindahl-Hirschman Indices (HHIs) computed from reporting exchanges as an inverse proxy for competition among agents. These HHIs are constructed for three variables: time at best; time alone at best; and number of quote improvements.

The time at best HHI variant is defined as follows. For a given firm and time interval, let $m_j^{at\ best}$ denote the number of ms that exchange *j*'s offer is at (that is, either sets or matches) the National best offer.

$$HHI^{at\ best} = \sum_j \left(\frac{m_j^{at\ best}}{\sum_k m_k^{at\ best}} \right)^2 \tag{4}$$

Time alone at best, HHI^{alone}, is defined similarly, but is based on m_j^{alone}, the number of ms that exchange *j*'s offer is alone at (that is, sets) the National best offer. The quote improvement variant, $HHI^{improve}$, is computed using $m_j^{improve}$, the number of quote improvements (reductions in the National best offer) occurring at exchange *j*. Note that for this to occur, the exchange must previously have been alone at the best offer.

The empirical specification is a general linear model of the form

$$HHI_{itd} = X_{itd}\beta + e_{itd} \tag{5}$$

where *HHI* is a placeholder for one of the three HHI measures, $i = 1, \dots, 150$ indexes firms, t indexes ten-minute date/time intervals, and $d \in \{bid, offer\}$ indicates direction. (The bid and offer sides of the market are treated as separate observations). The set of explanatory variables includes control fixed effects (dummies for firm and date/time), and also a measure of high-frequency quoting.

To enhance comparability across firms, the high-frequency quoting measure is constructed as an indicator variable. The HFQ dummy $HFQ_{itd} = 1$ if the rough volatility at level $j = 9$ (for the given firm, date/time and direction) lies at or above the 90th percentile of the distribution (for the firm and direction) over all t. Essentially, HFQ_{itd} is an indicator of high relative extent of quote volatility.

Table VI reports the estimated least squares means of the model. The column labeled "$HFQ = Low$," for example, reports the estimated mean *HHI* for the low HFQ state. The last column reports the implied difference in the *HHI* across the two HFQ states. Across increasing volume subsamples, all HHI's decline: liquidity supply in actively traded stocks is more competitive. Among the three HHI variants, the *alone* and *improve* measures are higher than the *at best* measure. This is not a necessary consequence of the definitions. Over a given interval in a set of *n* exchanges, it is possible that all would match the best quote most of the time, implying $HHI^{at\ best} = \sum(1/n)^2$. If in addition there are many quote improvements, and the first-mover for each improvement is randomly drawn with equal likelihood from the full set, HHI^{alone} and $HHI^{improve}$ would have values similar to $HHI^{at\ best}$.

The difference estimates exhibit striking consistency. For all *HHI* measures and across all samples, the difference is negative and statistically significant. This suggests that increased quote volatility is associated with more competition. The difference is smallest for the high-volume firms, and generally (but not uniformly) higher for the lower-volume quintiles.

B. Edgeworth cycles

The hallmark of an Edgeworth cycle in a typical product market is right-tail skewness in the first difference of the selling price. This occurs because the back-and-forth undercutting generates many small price drops, while the "reset" consists of a single large price increase. Similarly, Edgeworth cycles on the bid side are characterized by negative skewness.

Skewness can be assessed directly via the usual skewness coefficient, defined for the random price changes $skewness = E(\Delta p - E\Delta p)^3/[E(\Delta p - E\Delta p)^2]^{3/2}$. Some studies of Edgeworth cycles also use the normalized mean-less-median, denoted here by $MLM = \left(E\Delta p - \Delta p^{median}\right)/\sigma_{\Delta p}$. For a hypothetical cycle consisting of n one-tick steps down followed by one n-tick step up, $skewness = (n-1)/\sqrt{n}$ and $MLM = 1$ tick. Since *skewness* is increasing in *n*, and the price-change variance is $Var(dp) = n$, it might be conjectured that increasing price volatility necessarily leads to increased *skewness*. This is true for Edgeworth price dynamics, but not necessarily in general: price volatility could be symmetrical.

Because skewness is asymmetric across the bid and offer sides of the market, the high-frequency quoting indicator HFQ and $direction \in \{bid, offer\}$ must be interacted. Table VII reports the results. The results are expressed as implied means (for the dependent variable) under the indicated fixed effects. The direction fixed-effect estimates imply, for example, that unconditionally (that is, disregarding value of HFQ) bid price changes are positively skewed and offer price changes are negatively skewed. The interaction effects imply that the magnitude of the skewness is increased (by about fifty percent) with $HFQ = high$.

C. Discussion

The estimated effects of competition (proxied by the *HHI*s) are not consistent with prevalent single-agent and/or mixed multiple-agent strategies. The skewness estimates are consistent with Edgeworth cycles. Recall that both the mixed-strategy and Edgeworth cycle equilibria involved successive undercutting. There are many differences in the formal assumptions and arguments, but one in particular stands out. The mixed strategy models assume a continuous price space. The possibility of undercutting by an infinitesimal amount prevents the occurrence of mass points in the equilibrium distribution. The Edgeworth cycle equilibria, on the other hand, arise in formal models where the price space is discrete. Thus, the implication that Edgeworth cycles dominate mixed strategies may be a consequence of the relatively large tick size.

As a final note, the analysis here is directed at a broad classification of HFQ activity. It does not rule out, for example, the occurrence of quote-stuffing or spoofing. It merely says that they are not the dominant activity.

VIII. Historical evidence

The recent history of securities trading has been marked by advances in technology and proliferation of regulations governing the use of these advances. It is obvious that technology has enabled strategies that weren't possible in an era of manual markets. One oft-remarked feature of this trend is explosive growth in the number of quote records handled by the consolidated systems. It is logical to extend this observation with a conjecture of a similar increase in quote volatility. This section explores the empirical evidence bearing on this point.

A. Sample and data

The data for this phase of the analysis are drawn from CRSP and *Monthly* TAQ datasets. The sample selection procedure in each year is essentially identical to that described for the 2011 cross-sectional sample. In each year, from all firms present on CRSP and TAQ in April, with share codes in (10 and 11), and with primary listings on the NYSE, Amex and NASDAQ exchanges, I draw fifteen firms from each dollar trading volume decile.[13] Quote data are drawn from TAQ. Table VIII reports summary statistics. The increase in the intensity of trading activity is clearly visible in the trends for median number of trade and quote records. From 2001 to 2011, the average annual compound growth rate is about 23% percent for trades, and about 32% for quotes.

B. Methodology

Most of this paper's analyses rely on the April 2011 sample of daily TAQ data. In the daily TAQ, millisecond time stamps are only available from 2006 onwards. Monthly TAQ data (the standard source used in academic research) is available back to 1993, and the precursor ISSM data go back to the mid-1980s. These data are substantially less expensive than the daily TAQ, and they have a simpler logical structure. The time stamps on the Monthly TAQ and ISSM datasets are reported only to the second, however. This limitation might seem to render these data useless for characterizing sub-second variation, but on closer examination it turns out that the data are actually quite rich and informative.

The usual sampling situation in discrete time series analysis involves either aggregation over periodic intervals (such as quarterly GDP) or point-in-time periodic sampling (such as the end-of-day S&P index). In both cases there is one observation per interval, and in neither case do the data support resolution of components shorter than one interval. In the present situation, however, quote updates occur in continuous time and are disseminated continuously. The one second time-stamps arise as a truncation (or equivalently, a rounding) of the continuous event times. The Monthly TAQ data include all quote records, and it is not uncommon for a second to contain ten or even a hundred quote records.

Assume that all quote updates in a given second arrive as a Poisson process of constant intensity. If the interval $(0,t)$ contains n updates, then the update times have the same distribution as the order statistics in a sample of n independent random variables uniformly distributed on the

[13] As of April, 2001, NASDAQ had not fully implemented decimalization. For this year, I do not sample from stocks that traded in sixteenths.

interval $(0,t)$, Ross (1996, Theorem 2.3.1). Within a one-second interval containing n updates, therefore, we can simulate continuous arrival times by drawing n realizations from the standard uniform distribution, sorting, and assigning them to quotes (in order) as the fractional portions of the arrival times. These simulated time-stamps are essentially random draws from true distribution. This result does not require knowledge of the underlying Poisson arrival intensity.

I make the additional assumption that the quote update times are independent of the updated bid and offer prices. (That is, the "marks" associated with the arrival times are independent of the times.) Then all estimates based on the simulated time stamp series constitute draws from their corresponding posterior distributions. This procedure can be formalized in a Bayesian Markov-Chain Monte Carlo (MCMC) framework. To refine the estimates, we would normally make repeated simulations ("sweeps") over the sample, but due to computational considerations and programming complexity, I make only one draw for each CQ record.

The assumptions underlying this model are unlikely to be completely satisfied in practice. For a time-homogeneous Poisson process, interevent durations are independent. In fact, inter-event times in market data frequently exhibit pronounced serial dependence, and this feature is a staple of the autoregressive conditional duration and stochastic duration literature (Engle and Russell (1998); Hautsch (2004)). In NASDAQ data, Hasbrouck and Saar (2013) show that event times exhibit intra-second deterministic patterns. Subordinated stochastic process models of security prices suggest that transactions (not wall-clock time) are effectively the "clock" of the process Shephard (2005).

The reliability of the randomization approach can be assessed, however, by a simple test. The time-stamps of the data analyzed in the last section are stripped of their millisecond remainders. New millisecond remainders are simulated, the random-time-stamped data are analyzed, and the two sets of estimates (based on true vs. simulated time stamps) are compared. When this procedure was performed, estimates of the ν_j^2 and σ_j^2 parameters were found to be very highly correlated, even at sub-second time scales. (The wavelet bid and ask correlation estimates, however, are more sensitive to alignment, and are therefore not as reliable.)[14]

---

[14] In a sample of n uniform random numbers, the expected values of the n order statistics is $\{\delta, 2\delta, \dots, n\delta\}$ where $\delta = 1/(n+1)$. In working with Monthly TAQ data, Holden and Jacobsen (2013, HJ) suggest assigning sub-second time stamps as $\{\delta/2, 3\delta/2, \dots, 1-\delta/2\}$. HJ show this assignment yields reliable estimates of effective spreads. The two approaches can be shown to follow from different conditioning assumptions. The present result conditions only on the n arrivals

C. Results

In analyzing 2001-2011, the most useful statistics are variance ratios. By construction they are normalized with respect to long-term variance, and over this period there are large swings in market-wide long-term volatility (evident from a cursory examination of the VIX). These would be expected to affect the short term variances as well. Table IX Panel A reports the mean wavelet variance ratios for the shorter time scales. As in the 2011 sample, there is substantial variance inflation relative to the random-walk in all years. Perhaps surprisingly, though, the excess variance is high in all years, including the early years of the decade. The estimates are higher in 2001 than in 2011. The pattern does not suggest an increasing trend.

Given the recent media attention devoted to low-latency activity and the undeniable growth in quote volume, the absence of a strong trend in quote volatility seems surprising. There are several possible explanations. In the first place, "flickering quotes" drew comment well before the start of the sample, in an era when quotes were dominated by human market makers Harris (1999); U.S. Commodities Futures Trading Commission Technology Advisor Committee (2001). Also an artifact of this era is the specialist practice of "gapping" the quotes to indicate larger quantities at worse prices Jennings and Thirumalai (2007). In short, the quotes may have actually been less stable than popular memory holds. The apparent discrepancy between quote volatility and quote volume can be explained by appealing to the increase in market fragmentation and consequent growth in matching quotes.

Exploring this finding further, bid-offer plots for firm-days in each year that correspond to extreme realizations of the variances exhibit an interesting pattern. In later years, these outlier plots tend to resemble the initial AEPI example, with rapid oscillations of relatively low amplitude. In the earlier years, they are more likely to feature small number of prominent spikes associated with a sharply lower bid or elevated offer that persists for a minute or less.

As an example, Figure 3 (Panel A) depicts the NBBO for PRK (Park National Corporation, Amex-listed) on April 6, 2001. At around 10:00 there is a downward spike in the NBB. Shortly after noon there is a sharp drop in the NBB of roughly three dollars and a sharp rise in the NBO of about one dollar. Examination of the CQ record establishes that during this period there are multiple exchanges active in the market, but Amex is the apparent price leader. At 12:02:22, the Amex

---

in one interval, with an unknown arrival intensity. The HJ assignment corresponds to the expected locations under the assumption that the process has a constant arrival intensity.

establishes the NBB at $86.74. At 12:03:11 the Amex drops its bid to $83.63, exposing the NASDAQ bid of $86.68 as the new NBB. At 12:03:16, the NASDAQ bid drops, leaving the Amex's $83.63 as best. Within half a minute, however, the NBB is back at 86.50. The lower bid is not marketed by any special mode flag. It is not a penny ("stub") bid. The size of the bid at two (hundred shares) is typical for the market on that day. A similar sequence of events sends the NBO up a dollar for about one second.

These quotes are not so far off the mark as to be clearly erroneous. We must nevertheless question whether they were "real"? Did they reliably indicate the consensus market values at those instances? Were they accessible for execution? Were they truly the best in the market? There were no trades between 11:38 and 12:13, but if a market order had been entered, would it in fact have been executed at the NBBO?[15] These are meaningful questions because they bear directly on market quality. Ultimately, though, the record is unlikely to provide clear answers. The US equity market in 2001 reflected a blend of human and automated mechanisms, practices and conventions that defies detailed description even at a distance of only twelve years.

Discerning whether or not quote volatility increased over the period, therefore, requires that we sharpen the question. The quote volatility in the initial AEPI example is of high frequency, but low amplitude. This is visually distinct from the spikes of high frequency and high amplitude found in PRK. The latter is sometimes called "pop" noise, in reference to its sound in audio signals Walker (2008). As in the de-noising of audio signals, the goal is to remove the pops from the signals of lower amplitude. The wavelet literature has developed many denoising approaches (see Percival and Walden, Gençay et al, and Walker). When the stochastic properties of the noise and signal processes are known, optimal methods can often be established. In the present case, though, I adopt a simpler method.

Wavelet transforms facilitate the direct computation of smooth and rough components. This process, known as multiresolution analysis, isolates components at different time scales. As an example, Panel B of Figure 3 plots the rough component of the PRK bid at a time scale of 51.2 seconds. It is zero mean by construction, and the spikes are cleanly resolved. On the principle that high frequency quoting (as in the AEPI example) should not be substantially larger than the bid-

[15] The Amex (like the NYSE) had specialists in 2001. Specialists generally had affirmative price continuity obligations that would have discouraged (though not expressly forbidden) trades occurring at prices substantially different from those prevailing immediately before and immediately after. A broker-dealer, however, would not have been subject to this restriction.

offer spread in magnitude, I set acceptance bands at $\pm Min\left(1.5\times\left(average\ spread\right),\$0.25\right)$. The minimum of $0.25 is set to accommodate stocks with very tight spreads. For PRK, the bands are approximately ±$0.33, and they are indicated in the figure by horizontal black lines. Values lying outside of the band are set to the band limits. This clips the high-amplitude peaks, while leaving the low-amplitude components, some of which are highly oscillatory, untouched. The signal (bid or offer) is reconstituted using the clipped rough, and analysis proceeds on this denoised signal. I recompute all estimates for all firms using the denoised bids and offers.

Table IX Panel B reports the wavelet variance ratios for the denoised quotes. The results are striking. In the early years, the variance ratios computed from the denoised quotes are much lower than those computed from the raw data. In later years, however, the reduction associated with the denoising is small. For the 200 ms variance ratio, for example, the 2001 drop is from 2.43 (for the raw quotes) to 1.52 (for the denoised quotes), but the 2011 value only drops from 2.61 to 48.

These results are consistent with the view that the overall level of quote volatility did not change very much over the decade. The nature of the volatility has apparently, however, evolved. In the early years, the volatility was of relatively high amplitude but non-oscillatory. It is removed by the pop-denoising procedure. The procedure does not attenuate the low-amplitude highly oscillatory components, however, which drive quote volatility in the later years. The difference between the raw and denoised ratios generally declines throughout the decade, but the convergence appears to be strongest during the Reg NMS transition period.

The denoising procedure accentuates low-amplitude oscillatory volatility. Since one might expect that this would be tied more closely to low-latency technology, it is sensible to ask whether the denoised volatilities have increased. Table X therefore presents rough volatilities for the denoised quotes in $0.01 per share (Panel A), in basis points (Panel B), and as a variance ratio (Panel C) for a representative subset of time scales. Figure 4 plots these quantities at the 800 ms time scale. The table and figure suggest that neither the $0.01 per share volatility (Panel A) nor the basis point volatility (Panel B) evinces an upward trend. The variance ratio (Panel C) appears to climb from 2001 to 2004, but thereafter drifts distinctly downwards. In summary, the climb that might be expected from cumulative enhancements to trading technology or the growth in quote traffic is conspicuously absent.

IX. High frequency quoting and trading.

Although trading and quoting are different activities, most definitions of algorithmic and high frequency trading encompass many aspects of market behavior (not just executions), and would be presumed to cover quoting as well as trading.[16] In particular, the same technology that makes high frequency executions possible also facilitates the rapid submission, cancellation and repricing of the nonmarketable orders that establish the bid and offer. Quote volatility is not necessarily associated with a high frequency of executions. One can envision regimes where relatively stable quotes are hit intensively when fundamental valuations change, and periods (such as that depicted in Figure 1) where frenetic quoting occurs in the absence of executions. One might nevertheless expect this commonality of technology to link the two activities in practice.

Executions are generally emphasized over quotes when identifying agents as high frequency traders. For example, Kirilenko, Kyle, Samadi and Tuzun (2011) select on high volume and low inventory. The low inventory criterion excludes institutional investors who might use algorithmic techniques to accumulate or liquidate a large position. The NASDAQ HFT dataset uses similar criteria Brogaard (2012); Brogaard, Hendershott and Riordan (2012). Once high frequency traders are identified, their executions and the attributes of these executions lead to direct measures of HF activity in panel samples.

In some situations, however, identifications based on additional, non-trade information are possible. Menkveld (2013) identifies one Chi-X participant on the basis of size and prominence. The Automated Trading Program on the German XETRA system allows and provides incentives for designating an order as algorithmic Hendershott and Riordan (2013). Other studies analyze indirect measures of low-latency activity. Hendershott, Jones, and Menkveld (2011) use NYSE message traffic. Hasbrouck and Saar (2013) suggest strategic runs (order chains) of cancel and replace messages linked at intervals of 100 ms or lower.

Most of these studies find a positive association between low-latency activity and market quality. Low-latency activity, for example, tends to be negatively correlated with as posted and effective spreads, which are inverse measures of market quality. Most also find a zero or negative

---

[16] A CFTC draft definition reads: "High frequency trading is a form of automated trading that employs: (a) algorithms for decision making, order initiation, generation, routing, or execution, for each individual transaction without human direction; (b) low-latency technology that is designed to minimize response times, including proximity and co-location services; (c) high speed connections to markets for order entry; and (d) high message rates (orders, quotes or cancellations)" U.S. Commodities Futures Trading Commission (2011).

association between low-latency activity and volatility, although the constructed volatility measures usually span intervals that are long relative to those of the present paper. With respect to algorithmic or high frequency activity, Hendershott and Riordan (2012) find an insignificantly negative association with the absolute value of the prior 15-minute return; Hasbrouck and Saar (2013) find a negative association with the high-low difference of the quote midpoint over 10-minute intervals.

The time-scaled variance estimates used here clearly aim at a richer characterization of volatility than the high/low or absolute return proxies used in the studies above. The present study does not, on the other hand, attempt to correlate the variance measures with intraday proxies for high frequency trading. One would further suspect, of course, that the ultimate strategic purpose of high frequency quoting is to facilitate a trade or to affect the price of a trade. The mechanics of this are certainly deserving of further research.

The discussion in Section II associates short-term quote volatility with price uncertainty for those who submit marketable orders, use dark mechanisms that price by reference, or face monitoring difficulties. From this perspective, quote volatility is an inverse measure of market quality. Although the present study finds evidence of economically significant and elevated quote volatility, it does not establish a simple connection to technological trends associated with low latency activity.

X. Conclusions

This paper presents a number of new economic and statistical perspectives on volatility in bid and offer quotes. Intuitively, this volatility is measured as the local mean square deviation of the bid or offer about a local mean. The analysis is framed as a time scale decomposition, and draws extensively on the formal structure and results established in that area. In a 2011 sample of millisecond-stamped US equity data, estimates of sub-second high frequency variance for the National Best Bid and Offer (NBBO) are well in excess of what would be expected relative to random-walk volatility estimated over longer intervals. At an 800 ms time scale, for example, the estimated quote volatility is on average close to a basis point, about double what can be explained by fundamental volatility. Furthermore, the correlations between bids and offers at sub-second time scales are positive, but low. That the bid and offer are not moving together also suggests that the volatility is not fundamental.

Under additional assumptions, latency volatility implies cost differentials for agents subject to differing latencies, as faster traders can pick off better prices. As a representative estimate, traders with latencies of 800 ms and faster have an advantage of about 1.8 basis points relative to slower traders. In equilibrium one would expect liquidity providers' losses to fast traders to be passed on to slower traders, but the paper neither formally models this mechanism nor provides evidence directly bearing on it.

Sub-second volatility is comparable in magnitude to access fees and other transaction costs, but since it arises as a zero-mean risk its economic significance for most stocks and most traders is low. The latency differentials, however, imply expected costs for slower traders that are small but meaningful. All in, a hypothetical investor with latency above 800 ms, trading at the beginning and end of the year incurs an expected utility loss equivalent to a wealth penalty of 3.7 basis points.

The public debate on high frequency trading has given prominence to allegedly manipulative single-agent mechanisms (such as spoofing and quote stuffing). The evidence in this study suggests, however, that quote volatility is associated with increased competition. This finding runs counter to the predictions of mixed-strategy models of quote-setting. Quote volatility is also associated with more pronounced skewness in bid and offer changes. This is consistent with Edgeworth cycles commonly found in product markets. In product markets, it should also be noted, these cycles are not generally considered manipulative.

The paper also investigates recent trends and patterns in quote volatility. To facilitate the use of data with time stamps that are truncated or rounded to low resolution, the paper proposes a simple and straightforward simulation strategy. In a 2001-2011 sample of second-stamped TAQ data, the compound annual average growth rate in the number quote (CQ) records is 32%. Quote volatility, however, exhibits no such striking trend. In fact, some of the highest estimates occur in 2001 and 2002, a finding that seems to reflect market-makers "gapping" the quotes. The highest levels of quote volatility occur in 2004-2006.

Many of these findings raise additional questions. The findings on quote volatility presented here are broad characterizations that may on closer examination exhibit diversity as to strategies of liquidity providers and relative costs to liquidity seekers. The volatility in the motivating example, moreover, is concentrated and episodic, which raises questions about conditions that might give rise to these bursts. Finally, the volatility considered here embraces both fundamental value-relevant and transient effects. A resolution (even if partial) would help clarify the role of quotes in price discovery. All of these concerns are worthwhile goals of further research.

References

Aït-Sahalia, Yacine, Per A. Mykland, and Lan Zhang, 2011, Ultra high frequency volatility estimation with dependent microstructure noise, *Journal of Econometrics* 160, 160-175.

Amihud, Yakov, and Haim Mendelson, 1987, Trading mechanisms and stock returns: An empirical investigation, *Journal of Finance* 42, 533-553.

Andersen, Torben G., Tim Bollerslev, Francis X. Diebold, and H. Ebens, 2002, Great realizations, *Risk* 13, 105-108.

Andersen, Torben G., Tim P. Bollerslev, Francis X. Diebold, and H. Ebens, 2001, The distribution of realized stock return volatility, *Journal of Financial Economics* 61, 43-76.

Andersen, Torben G., Tim P. Bollerslev, Francis X. Diebold, and Paul Labys, 2003a, The distribution of realized exchange rate volatility, *Journal of the American Statistical Association* 98, 501-501.

Andersen, Torben G., Tim P. Bollerslev, Francis X. Diebold, and Paul Labys, 2003b, Modeling and forecasting realized volatility, *Econometrica* 71, 579-625.

Atsalakis, George S., and Kimon P. Valavanis, 2009, Surveying stock market forecasting techniques - Part II: Soft computing methods, *Expert Systems with Applications* 36, 5932-5941.

Barnea, Amir, 1974, Performance evaluation of New York Stock Exchange specialists, *Journal of Financial and Quantitative Analysis* 9, 511-535.

Baruch, Shmuel, and Lawrence R. Glosten, 2013, Flickering quotes, (Columbia University).

Biais, Bruno, Thierry Foucault, and Sophie Moinas, 2012, Equilibrium high-frequency trading, (University of Toulouse).

Brogaard, Jonathan, 2012, Essays on High-Frequency Trading, Kellogg School (Northwestern University, Evanston).

Brogaard, Jonathan, Terrence J. Hendershott, and Ryan Riordan, 2012, High frequency trading and price discovery, SSRN eLibrary (University of Washington).

Budish, Eric B., Peter Cramton, and John J. Shim, 2013, The high-frequency trading arms race: frequent batch auctions as a market design response, (SSRN).

Conrad, Jennifer S., Sunil Wahal, and Jin Xiang, 2014, High-frequency quoting, trading, and the efficiency of prices, (Kenan-Flagler School, University of North Carolina).

Dacorogna, Michel M., Ramazan Gencay, Ulrich A. Muller, Richard B. Olsen, and Olivier B. Pictet, 2001. *High-Frequency Finance* (Academic Press, New York).

EconPort, 2014, "Dutch Auction" (entry), http://www.econport.org/econport/request?page=man_auctions_dutchauction, accessed on August 18, 2014.

Edgeworth, Francis Y., 1925, The pure theory of monopoly, in *Papers Relating to Political Economy, Vol. I* (MacMillan, London).

Egginton, Jared, Bonnie F. Van Ness, and Robert A. Van Ness, 2012, Quote stuffing, (University of Mississippi).

Elder, John, and Hyun J. Jin, 2007, Long memory in commodity futures volatility: A wavelet perspective, *Journal of Futures Markets* 27, 411-437.

Engle, Robert F., and Jeffrey R. Russell, 1998, Autoregressive conditional duration: A new model for irregularly spaced transaction data, *Econometrica* 66, 1127-1162.

Fan, Jianqing, and Yazhen Wang, 2007, Multi-scale jump and volatility analysis for high-frequency financial data, *Journal of the American Statistical Association* 102, 1349-1362.

Fan, Yanqin, and Ramazan Gençay, 2010, Unit root tests with wavelets, *Econometric Theory* 26, 1305-1331.

Fang, Y., 1996, Volatility Modeling and Estimation of High-Frequency Data with Gaussian Noise, (MIT Sloan School).

Foucault, Thierry, Johan Hombert, and Ioanid Rosu, 2013, News trading and speed, (HEC Paris).

Gençay, Ramazan, Faruk Selçuk, Nikola Gradojevic, and Brandon Whitcher, 2010, Asymmetry of information flow between volatilities across time scales, *Quantitative Finance* 10, 895-915.

Gençay, Ramazan, Faruk Selçuk, and Brandon Whitcher, 2005, Multiscale systematic risk, *Journal of International Money and Finance* 24, 55-70.

Gençay, Ramazan, Frank Selçuk, and Brandon Whitcher, 2002. *An Introduction to Wavelets and Other Filtering Methods in Finance and Economics* (Academic Press (Elsevier), San Diego).

Gençay, Ramazan, and Danielle Signori, 2012, Multi-scale tests for serial correlation, (Simon Fraser University).

Haar, Alfred, 1910, Zur Theorie der Orthogonalen Funktionensysteme, *Mathematische Annalen* 69, 331-71.

Hansen, Peter R., and Asger Lunde, 2006, Realized variance and market microstructure noise, *Journal of Business & Economic Statistics* 24, 127-161.

Harris, Lawrence E., 1999, Trading in pennies: a survey of the issues, (Marshall School, University of Southern California).

Hasbrouck, Joel, and Gideon Saar, 2013, Low-Latency Trading, *Journal of Financial Markets, forthcoming*.

Hasbrouck, Joel, and Robert A. Schwartz, 1988, Liquidity and execution costs in equity markets, *Journal of Portfolio Management* 14, 10-16.

Hautsch, Nikolaus, 2004. *Modelling Irregularly Spaced Financial Data: Theory and Practice of Dynamic Duration Models* (Springer).

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

Hendershott, Terrence J., and Ryan Riordan, 2012, High Frequency Trading and Price Discovery, *SSRN eLibrary*.

Hendershott, Terrence J., and Ryan Riordan, 2013, Algorithmic trading and the market for liquidity, *Journal of Financial and Quantitative Analysis*.

Høg, Esben, and Asger Lunde, 2003, Wavelet estimation of integrated volatility, (Aarhus School of Business, Department of Information Science).

Holden, Craig W., and Stacey E. Jacobsen, 2013, Liquidity measurement problems in fast, competitive markets: expensive and cheap solutions, *Journal of Finance*.

Hsieh, Tsung-Jung, Hsiao-Fen Hsiao, and Wei-Chang Yeh, 2011, Forecasting stock markets using wavelet transforms and recurrent neural networks: An integrated system based on artificial bee colony algorithm, *Applied Soft Computing* 11, 2510-2525.

In, Francis, and Sangbae Kim, 2006, The hedge ratio and the empirical relationship between the stock and futures markets: a new approach using wavelet analysis, *The Journal of Business* 79, 799-820.

ITG, 2014, Global Cost Review Q1/2014, (ITG Inc.).

Jarrow, Robert A., and Philip Protter, 2012, A dysfunctional role of high-frequency trading in electronic markets, *International Journal of Theoretical and Applied Finance* 15.

Jennings, Robert H., and Ramabhadran S. Thirumalai, 2007, Advertising for liquidity on the New York Stock Exchange, (Kelley School, University of Indiana).

Kirilenko, Andrei A., Albert S. Kyle, Mehrdad Samadi, and Tugkan Tuzun, 2011, The flash crash: the impact of high frequency trading on an electronic market, SSRN eLibrary (CFTC and University of Maryland).

Leach, J. Chris, and Ananth N. Madhavan, 1992, Intertemporal price discovery by market makers: active versus passive learning, *Journal of Financial Intermediation* 2, 207-235.

Leach, J. Chris, and Ananth N. Madhavan, 1993, Price experimentation and security market structure, *Review of Financial Studies* 6, 375-404.

Lo, Andrew W., and A. Craig MacKinlay, 1988, Stock market prices do not follow random walks: Evidence from a simple specification test, *Review of Financial Studies* 1, 41-66.

Maskin, Eric, and Jean Tirole, 1988, A theory of dynamic oligopoly, II: price competition, kinked demand curves, and Edgeworth cycles, *Econometrica* 56, 971-599.

Menkveld, Albert J., 2013, High-frequency trading and the new market-makers, *Journal of Financial Markets, forthcoming*.

Nason, Guy P., 2008. *Wavelet Methods in Statistics with R* (Springer Science+Business Media, LLC, New York).

Noel, Michael D., 2011, Edgeworth price cycles, in Steven N. Durlauf, and Lawrence E. Blume, eds.: *The New Palgrave Dictionary of Economics* (Palgrave Macmillan, Basingstoke).

Percival, Donald B., and Andrew T. Walden, 2000. *Wavelet Methods for Time Series Analysis* (Cambridge University Press, Cambridge).

Ramsey, James B., 1999, The contribution of wavelets to the analysis of economic and financial data, *Philosophical Transactions of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences* 357, 2593-2606.

Ramsey, James B., 2002, Wavelets in economics and finance: past and future, *Studies in Nonlinear Dynamics and Econometrics* 6.

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

Ross, Sheldon M., 1996. *Stochastic Processes* (John Wiley and Sons, Inc., New York).

Schwartz, R. A., and L. R. Wu, 2013, Equity Trading in the Fast Lane: The Staccato Alternative, *Journal of Portfolio Management* 39, 3-6.

Shephard, Neil, 2005, General introduction, in Neil Shephard, ed.: *Stochastic Volatility* (Oxford University Press, Oxford).

Stoll, H. R., and C. Schenzler, 2006, Trades outside the quotes: Reporting delay, trading option, or trade size?, *Journal of Financial Economics* 79, 615-653.

Teyssière, Gilles, and Patrice Abry, 2007, Wavelet analysis of nonlinear long-range dependent processes: applications to financial time series, in Gilles Teyssière, and Alan P. Kirman, eds.: *Long Memory in Economics* (Springer, Berlin).

U.S. Commodities Futures Trading Commission, 2011, Presentation (Working Group 1, Additional Document), Technology Advisor Committee.

U.S. Commodities Futures Trading Commission, 2013, Antidisruptive Practices Authority (Interpretive Guidance and Policy Statement), (Federal Register).

U.S. Commodities Futures Trading Commission Technology Advisor Committee, 2001, Market Access Committee Interim Report.

Varian, Hal, 1980, A model of sales, *American Economic Review* 70, 651-659.

Walker, James S., 2008. *A Primer on Wavelets and Their Scientific Applications* (Chapman and Hall/CRC (Taylor and Francis Group), Boca Raton).

Zhang, Lan, 2006, Efficient Estimation of Stochastic Volatility Using Noisy Observations: A Multi-Scale Approach, *Bernoulli* 12, 1019-1043.

Zhang, Lan, Per A. Mykland, and Yacine Aït-Sahalia, 2005, A Tale of Two Time Scales: Determining Integrated Volatility with Noisy High-Frequency Data, *Journal of the American Statistical Association* 100, 1394-1411.

Zhang, Xiaoquan (Michael), 2005, Finding Edgeworth cycles in online advertising auctions, (Sloan School, MIT, Cambridge).

Appendix: Deviations from averages of random walks

Consider a price series that evolves as $p_t = p_{t-1} + u_t$ where u_t is a white-noise process with unit variance. Without loss of generality, we initialize $p_0 = 0$ and consider the mean-squared deviations over n observations:

$$MSD(n) = \frac{1}{n}\sum_{i=1}^{n} p_i^2 - \left(\frac{1}{n}\sum_{i=1}^{n} p_i\right)^2 = \frac{1}{n}\sum_{i=1}^{n}\left(\sum_{j=1}^{i} u_i\right)^2 - \left(\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{i} u_i\right)^2 \tag{A1}$$

Taking expectations (noting that $Eu_iu_j = 1$ for $i = 1$ and zero otherwise) and simplifying the sums gives

$$E\big(MSD(n)\big) = \frac{n+1}{2} - \frac{(n+1)(2n+1)}{6n} = \frac{n^2-1}{6n} \tag{A2}$$

For the sequence of averaging periods $n_j = n_0 2^j$ for $j = 0, \ldots$, the corresponding sequence of variances is

$$\gamma_j^2 = E\left(MSD(n_j)\right) = \frac{4^j n_0^2 - 1}{3n_0 2^{j+1}} \tag{A3}$$

In moving from $j-1$ to j the incremental change in variance (the wavelet variance) is

$$\nu_j^2 = \gamma_j^2 - \gamma_{j-1}^2 = \frac{4^j n_0^2 + 2}{3n_0 2^{j+1}} \text{ for } j = 1, \ldots \tag{A4}$$

In the special case where $n_0 = 1$, this reduces to Percival and Walden, exercise 8.3, p. 337 (using $\sigma_\epsilon^2 = 1$, and $\tau_j = 2^{j-1}$). The rough variance is usually defined as

$$\sigma_j^2 = \sum\nolimits_{i=1}^{j} \nu_j^2 = \frac{2^{-j-1}\big(2^j - 1\big)\big(2^j n_0^2 - 1\big)}{3n_0} \tag{A5}$$

where the summation starts at $i = 1$. We now reinterpret these results in a slightly expanded framework. Suppose that the original time subscript t indexes periods of Δ time units ("milliseconds") and that the variance per unit time of the u_t process is σ_u^2. Let M denote the averaging period measured in units of time, and correspondingly, $M_j = M_0 2^j$ for $j = 0,1, \ldots$. Then the wavelet and rough variances become

$$\nu_j^2 = \frac{\left(4^j M_0^2 + 2\Delta^2\right)\sigma_u^2}{3M_0 2^{j+2}} \text{ and } \sigma_j^2 = \frac{2^{-j-1}\left(2^j - 1\right)\left(2^j M_0^2 - \Delta^2\right)\sigma_u^2}{3M_0} \tag{A6}$$

In the continuous time limit, as $\Delta \to 0$,

$$\nu_j^2 = \frac{1}{3} 2^{j-2} M_0 \sigma_u^2 \text{ and } \sigma_j^2 = \frac{1}{6}\left(2^j - 1\right) M_0 \sigma_u^2. \tag{A7}$$

The wavelet and rough variance ratios are:

$$V_j = 2^{J-j} \frac{\nu_j^2}{\nu_J^2} = 1 \text{ and } VR_j = \frac{2^{J-1}}{(2^j - 1)} \times \frac{\sigma_j^2}{\nu_j^2} = 1 \tag{A8}$$

Alternatively, the rough variance may be defined to include the variance within the initial averaging interval. In this case, corresponding to equation (A5):

$$\sigma_j^2 = \gamma_0^2 + \sum_{i=1}^{j} \nu_j^2 = \frac{2^{-j-1}\left(4^j n_0^2 - 1\right)}{3n_0}.$$

In the continuous time limit the rough variance ratio becomes

$$VR_j = 2^{J-j-1} \times \frac{\sigma_j^2}{\nu_j^2} = 1. \tag{A9}$$

Table I. Descriptive Statistics

Source: CRSP and Daily TAQ data, April 2011. The sample is 150 firms randomly selected from CRSP with stratification based on average dollar trading volume in the first quarter of 2011, grouped in quintiles by dollar trading volume. NBB is the National Best Bid; NBO the National Best Offer. Except for the counts (first four rows), all table entries are cross-firm medians.

		Dollar trading volume quintile				
	Full sample	1 (low)	2	3	4	5 (high)
No. of firms	149	29	30	30	30	30
NYSE	47	0	5	7	16	19
Amex	6	2	2	0	1	1
NASDAQ	96	27	23	23	13	10
Avg. daily CT records (trades)	1,346	33	431	1,126	3,478	16,987
Avg. daily CQ records (quotes)	24,053	1,067	7,706	24,026	53,080	181,457
Avg. daily NBBO records	7,203	354	3,029	7,543	16,026	46,050
Avg. daily NBB changes	1,265	121	511	1,351	2,415	4,124
Avg. daily NBO changes	1,179	106	460	1,361	2,421	4,214
Avg. price (bid-offer midpoint)	$15.77	$4.76	$5.46	$17.86	$27.76	$51.60
Market capitalization of equity, $Million	$690	$41	$202	$747	$1,502	$8,739

Table II. Quote variance ratios

Estimates for 150 US firms during April, 2011. The wavelet variance ratio is $V_j = 2^{J-j}\nu_j^2/\nu_J^2$ where ν_j^2 is the wavelet variance at level j and $J = 16$ is the highest level in the analysis. The rough variance ratio is $VR_j = 2^{J-j-1}\sigma_j^2/\nu_J^2$ where σ_j^2 is the rough variance at level j. The price-change variance ratio is $V\Delta_j = 2^{J-j}Var\left(\Delta_{\tau_j}p_t\right)/Var\left(\Delta_{\tau_J}p_t\right)$ where Δ_τ is the differencing operator $\Delta_\tau p_t = p_t - p_{t-\tau}$ and $\tau_j = 2^{j-1}$is the level-j time scale. For a random-walk, all variance ratios should be unity for all j. The sample is winsorized at $\pm5\%$. The National Best Bid and Offer are computed from TAQ data; the bid and offer are separately transformed using the Haar basis; the reported variance estimates are averages of the bid and offer variances. The data are time stamped to the millisecond. Prior to transformation, I take the average of the bid or offer over non-overlapping 50 millisecond intervals. Entries for $j = 0$ are variances within the 50 ms intervals.

Level	Time scale	V_j	VR_j	$V\Delta_j$
0	< 50 ms	5.198	5.198	
1	50 ms	2.344	2.465	2.377
2	100 ms	2.257	2.362	2.295
3	200 ms	2.154	2.258	2.192
4	400 ms	2.048	2.162	2.085
5	800 ms	1.975	2.061	2.004
6	1,600 ms	1.888	1.971	1.917
7	3.2 sec	1.774	1.880	1.823
8	6.4 sec	1.657	1.774	1.721
9	12.8 sec	1.544	1.654	1.606
10	25.6 sec	1.449	1.548	1.496
11	51.2 sec	1.353	1.455	1.406
12	102.4 sec	1.269	1.364	1.310
13	3.4 min	1.199	1.289	1.232
14	6.8 min	1.141	1.217	1.168
15	13.7 min	1.078	1.148	1.094
16	27.3 min	1.000	1.074	1.000

Table III. Quote variance ratios in volume subsamples

Estimates for 150 US firms during April, 2011. The wavelet variance ratio is $V_j = 2^{J-j}\nu_j^2/\nu_J^2$ where ν_j^2 is the wavelet variance at level *j* and $J = 16$ is the highest level in the analysis. The rough variance ratio is $VR_j = 2^{J-j-1}\sigma_j^2/\nu_J^2$ where σ_j^2 is the rough variance at level *j*. The price-change variance ratio is $V\Delta_j = 2^{J-j}Var\left(\Delta_{\tau_j}p_t\right)/Var\left(\Delta_{\tau_J}p_t\right)$ where Δ_τ is the differencing operator $\Delta_\tau p_t = p_t - p_{t-\tau}$ and $\tau_j = 2^{j-1}$is the level-*j* time scale. For a random-walk, all variance ratios should be unity for all *j*. All entries are winsorized cross-firm means. The National Best Bid and Offer are computed from TAQ data; the bid and offer are separately transformed using the Haar basis; the reported variance estimates are averages of the bid and offer variances.. Table reports estimates for the full sample and subsamples constructed as quintiles of dollar trading volume.

Sample	*Level, j*	*Time scale*	V_j	VR_j	$V\Delta_j$
Full sample	1	50 ms	2.344	2.465	2.377
			(0.108)	(0.116)	(0.104)
	5	800 ms	1.975	2.061	2.004
			(0.082)	(0.087)	(0.078)
	16	27.3 min	1.000	1.074	1.000
				(0.007)	
1 (low)	1	50 ms	3.648	3.815	3.465
			(0.262)	(0.274)	(0.249)
	5	800 ms	3.096	3.184	2.927
			(0.195)	(0.205)	(0.182)
	16	27.3 min	1.000	1.156	1.000
				(0.017)	
2	1	50 ms	2.831	3.007	2.926
			(0.252)	(0.276)	(0.235)
	5	800 ms	2.300	2.435	2.411
			(0.173)	(0.195)	(0.170)
	16	27.3 min	1.000	1.077	1.000
				(0.018)	
3	1	50 ms	2.268	2.445	2.388
			(0.161)	(0.184)	(0.175)
	5	800 ms	1.930	2.018	1.993
			(0.115)	(0.126)	(0.120)
	16	27.3 min	1.000	1.056	1.000
				(0.010)	
4	1	50 ms	1.645	1.708	1.784
			(0.139)	(0.149)	(0.157)
	5	800 ms	1.385	1.457	1.511
			(0.099)	(0.109)	(0.114)
	16	27.3 min	1.000	1.016	1.000
				(0.009)	
5 (high)	1	50 ms	1.328	1.350	1.322
			(0.049)	(0.050)	(0.038)
	5	800 ms	1.164	1.211	1.179
			(0.042)	(0.043)	(0.035)
	16	27.3 min	1.000	1.065	1.000
				(0.010)	

Table IV. Wavelet volatilities and derived measures

Estimates of time scale variances and related measures for 150 US firms during April, 2011. The wavelet volatilities, ν_j are estimates of the price volatility at the time scale $\tau_j = 50 \times 2^{j-1}$ ms. The rough volatilities, σ_j measure cumulative variation at all time scales $\leq \tau_j$. Both values are scaled to \$0.01/share and (alternatively) basis points. Based on the model of timing advantage developed in section 2, ν_j is also the expected transfer from a trader active at time scale τ_{j+1} to a trader active at time scale τ_j. The cumulative gain $G_j = \sum_{i=1}^{j} \nu_i$ is the expected transfer from a trader active at time scale τ_{j+1} to traders active at all lower time scales. $Corr(D_j^{bid}, D_j^{offer})$ is the correlation between bid and offer detail components at level *j*. All entries are winsorized cross-firm means. The National Best Bid and Offer are computed from TAQ data; the bid and offer are analyzed separately and then averaged. The data are time stamped to the millisecond. Prior to transformation, I take the average of the bid or offer over non-overlapping 50 millisecond intervals. Entries for $j = 0$ are variances within the 50 ms intervals.

		\$0.01 /share			bp, 0.01%			
Level	Time scale	ν_j	σ_j	G_j	ν_j	σ_j	G_j	$Corr(D_j^{bid}, D_j^{offer})$
0	< 50 ms	0.030	0.030	0.000	0.178	0.178	0.000	
1	50 ms	0.028	0.041	0.028	0.171	0.246	0.171	0.312
2	100 ms	0.039	0.057	0.068	0.236	0.342	0.407	0.356
3	200 ms	0.055	0.079	0.122	0.326	0.472	0.732	0.400
4	400 ms	0.076	0.109	0.198	0.452	0.653	1.183	0.437
5	800 ms	0.105	0.152	0.303	0.630	0.910	1.817	0.471
6	1,600 ms	0.144	0.211	0.450	0.870	1.264	2.696	0.509
7	3.2 sec	0.201	0.290	0.654	1.196	1.740	3.894	0.548
8	6.4 sec	0.279	0.402	0.924	1.617	2.376	5.511	0.592
9	12.8 sec	0.384	0.559	1.313	2.206	3.236	7.697	0.640
10	25.6 sec	0.528	0.769	1.850	3.014	4.430	10.688	0.688
11	51.2 sec	0.722	1.059	2.578	4.123	6.066	14.825	0.739
12	102.4 sec	0.992	1.449	3.573	5.655	8.297	20.512	0.787
13	3.4 min	1.374	1.989	4.929	7.751	11.407	28.336	0.828
14	6.8 min	1.906	2.764	6.807	10.627	15.614	39.168	0.859
15	13.7 min	2.657	3.829	9.475	14.548	21.358	53.750	0.882
16	27.3 min	3.626	5.288	13.096	19.633	28.990	73.299	0.896

Table V. Wavelet volatilities and derived measures in volume subsamples

Estimates of time scale variances and related measures for 150 US firms during April, 2011. The wavelet volatilities, ν_j are estimates of the price volatility at the time scale $\tau_j = 50 \times 2^{j-1}$ ms. The rough volatilities, σ_j measure cumulative variation at all time scales $\leq \tau_j$. Based on the model of timing advantage developed in section 2, ν_j is also the expected transfer from a trader active at time scale τ_{j+1} to a trader active at time scale τ_j. The cumulative gain $G_j = \sum_{i=1}^{j} \nu_i$ is the expected transfer from a trader active at time scale τ_{j+1} to traders active at all lower time scales. $Corr(D_j^{bid}, D_j^{offer})$ is the correlation between bid and offer detail components at level *j*. Subsamples are constructed on average daily dollar trading volume. Standard errors are given in parentheses.

			\$0.01 /share			bp, 0.01%			
Sample	Level	Time scale	ν_j	σ_j	G_j	ν_j	σ_j	G_j	$Corr(D_j^{bid}, D_j^{offer})$
Full	1	50 ms	0.028	0.041	0.028	0.171	0.246	0.171	0.312
sample			(0.002)	(0.003)	(0.002)	(0.009)	(0.013)	(0.009)	(0.018)
	5	800 ms	0.105	0.152	0.303	0.630	0.910	1.817	0.471
			(0.007)	(0.009)	(0.019)	(0.032)	(0.047)	(0.093)	(0.019)
	16	27.3 min	3.626	5.288	13.096	19.633	28.990	73.299	0.896
			(0.230)	(0.333)	(0.811)	(0.725)	(1.110)	(2.949)	(0.018)
1 (low)	1	50 ms	0.020	0.029	0.020	0.296	0.426	0.296	0.060
			(0.003)	(0.004)	(0.003)	(0.022)	(0.031)	(0.022)	(0.009)
	5	800 ms	0.074	0.106	0.212	1.101	1.586	3.156	0.166
			(0.011)	(0.016)	(0.032)	(0.079)	(0.116)	(0.231)	(0.015)
	16	27.3 min	1.792	2.736	7.115	26.332	40.416	106.455	0.532
			(0.249)	(0.375)	(0.938)	(1.761)	(2.626)	(6.887)	(0.046)
2	1	50 ms	0.020	0.029	0.020	0.217	0.313	0.217	0.235
			(0.004)	(0.006)	(0.004)	(0.018)	(0.025)	(0.018)	(0.031)
	5	800 ms	0.070	0.103	0.205	0.782	1.135	2.271	0.403
			(0.013)	(0.019)	(0.038)	(0.062)	(0.090)	(0.180)	(0.032)
	16	27.3 min	2.117	3.115	7.873	23.360	34.561	87.989	0.958
			(0.348)	(0.527)	(1.361)	(1.669)	(2.571)	(6.810)	(0.009)
3	1	50 ms	0.028	0.042	0.028	0.142	0.208	0.142	0.308
			(0.004)	(0.005)	(0.004)	(0.012)	(0.017)	(0.012)	(0.026)
	5	800 ms	0.103	0.150	0.300	0.532	0.766	1.528	0.512
			(0.013)	(0.018)	(0.037)	(0.045)	(0.063)	(0.126)	(0.032)
	16	27.3 min	3.409	4.939	12.312	17.461	25.521	64.006	0.993
			(0.417)	(0.599)	(1.481)	(1.254)	(1.939)	(5.102)	(<0.001)
4	1	50 ms	0.036	0.051	0.036	0.113	0.163	0.113	0.406
			(0.004)	(0.006)	(0.004)	(0.006)	(0.009)	(0.006)	(0.030)
	5	800 ms	0.133	0.192	0.383	0.420	0.607	1.214	0.561
			(0.017)	(0.024)	(0.048)	(0.023)	(0.033)	(0.066)	(0.028)
	16	27.3 min	5.003	7.168	17.468	17.396	24.672	59.770	0.998
			(0.566)	(0.816)	(1.983)	(1.142)	(1.602)	(3.786)	(<0.001)
5 (high)	1	50 ms	0.039	0.055	0.039	0.085	0.121	0.085	0.553
			(0.004)	(0.005)	(0.004)	(0.005)	(0.008)	(0.005)	(0.033)
	5	800 ms	0.145	0.209	0.417	0.317	0.458	0.917	0.714
			(0.014)	(0.020)	(0.040)	(0.021)	(0.030)	(0.059)	(0.029)
	16	27.3 min	5.809	8.481	20.711	13.615	19.778	48.278	1.000
			(0.505)	(0.735)	(1.779)	(1.054)	(1.508)	(3.601)	(<0.001)

Table VI.
High-frequency quoting and competition

For three measures of the Herfindahl-Hirschman index (HHI), the table reports mean estimates implied by the linear fixed-effects panel model $(\cdot)_{itd} = X_{itd}\beta + e_{itd}$, where $i = 1, \ldots, 150$ indexes firms, $t = 1, \ldots, 20 \times 36$ indexes 10-minute date/time intervals between 9:45 and 15:45 and trading days in April 2011 (20 days × 36 intervals/day), and direction $d \in \{bid, offer\}$. (The bid and offer sides of the market are treated as separate observations.) All variables are computed at the firm/date/time/direction level, but the *itd* subscripts are suppressed for clarity. $HFQ \in \{low, high\}$ is a high-frequency quoting indicator, set to *high* if the rough variance of the quote (bid or offer, depending on direction) lies at or above 90th percentile of the empirical distribution of rough variances (for the firm *i* and direction *d*). The HHI that measures competition at the best quote is defined as $HHI^{at\ best} = \sum_j \left(m_j^{at\ best} / \sum_k m_k^{at\ best}\right)^2$ where $m_j^{at\ best}$ denotes the number of ms that exchange *j*'s quote sets or matches the National best; HHI^{alone}, is defined similarly, but is based on m_j^{alone}, the number of ms that exchange *j*'s quote is alone at (that is, sets) the National best; $HHI^{improve}$, is computed using $m_j^{improve}$, the number of quote improvements (reductions in the National best offer or increases in the National best bid) occurring at exchange *j*. All specifications include firm and date/time fixed effects. Standard errors are reported in parentheses.

Variable	*Sample*	$HFQ = Low$	$HFQ = High$	$High - Low\ Diff$
$HHI^{at\ best}$	All firms	0.3646	0.3322	-0.0324
		(0.0003)	(0.0010)	(0.0011)
	1 (low)	0.7331	0.6533	-0.0798
		(0.0014)	(0.0032)	(0.0038)
	2	0.3844	0.3370	-0.0474
		(0.0008)	(0.0026)	(0.0027)
	3	0.2954	0.2684	-0.0270
		(0.0005)	(0.0018)	(0.0019)
	4	0.2363	0.2273	-0.0089
		(0.0004)	(0.0012)	(0.0013)
	5 (high)	0.1788	0.1757	-0.0031
		(0.0002)	(0.0007)	(0.0008)

(continued on next page)

Table VI.
High-frequency quoting and competition (Continued)

$Variable$	$Sample$	$HFQ = Low$	$HFQ = High$	$High - Low\ Diff$
HHI^{alone}	All firms	0.7245	0.5988	-0.1256
		(0.0005)	(0.0014)	(0.0015)
	1 (low)	0.9698	0.8582	-0.1117
		(0.0009)	(0.0018)	(0.0021)
	2	0.8302	0.6454	-0.1848
		(0.0012)	(0.0034)	(0.0036)
	3	0.7120	0.5699	-0.1422
		(0.0012)	(0.0036)	(0.0038)
	4	0.5911	0.5101	-0.0810
		(0.0010)	(0.0036)	(0.0038)
	5 (high)	0.5150	0.4400	-0.0751
		(0.0010)	(0.0034)	(0.0037)
$HHI^{improve}$	All firms	0.6152	0.5060	-0.1093
		(0.0007)	(0.0015)	(0.0016)
	1 (low)	0.9301	0.7877	-0.1424
		(0.0034)	(0.0039)	(0.0048)
	2	0.7309	0.5300	-0.2009
		(0.0016)	(0.0040)	(0.0043)
	3	0.5941	0.4732	-0.1208
		(0.0012)	(0.0036)	(0.0039)
	4	0.4522	0.4045	-0.0477
		(0.0009)	(0.0032)	(0.0034)
	5 (high)	0.3835	0.3541	-0.0294
		(0.0007)	(0.0026)	(0.0028)

Table VII
High-frequency quoting and asymmetry in quote changes

For two measures of quote change asymmetry, the table reports mean estimates implied by the linear fixed-effects panel model $(\cdot)_{itd} = X_{itd}\beta + e_{itd}$, where $i = 1, \ldots, 150$ indexes firms, $t = 1, \ldots, 20 \times 36$ indexes 10-minute date/time intervals between 9:45 and 15:45 and trading days in April 2011 (20 days × 36 intervals/day), and direction $d \in \{bid, offer\}$. (The bid and offer sides of the market are treated as separate observations.) *Skewness* is the usual skewness coefficient computed for changes in the NBB or NBO (depending on d). *MLM* is the mean less the median, normalized by (that is, divided by) the standard deviation. $HFQ \in \{low, high\}$ is a high-frequency quoting indicator, set to *high* if the rough variance of the quote (bid or offer, depending on direction) lies at or above 90th percentile of the empirical distribution of rough variances (for the firm *i* and direction *d*). The fixed effects of interest are those associated with direction *d* and the interaction effect $d \times HFQ$. All specifications include fixed effects for firm and date/time. The sample is either all firms, or subsamples constructed as quintiles of average daily dollar volume.
Standard errors are reported in parentheses.

		Direction fixed effect			*Direction × HFQ interaction effect*					
Variable	*Sample*	$d = Bid$	$d = Offer$	*Difference*	$d = Bid$ and $HFQ = low$	$d = Bid$ and $HFQ = high$	*Difference*	$d = Offer$ and $HFQ = low$	$d = Offer$ and $HFQ = high$	*Difference*
Skewness	Full	-0.2134	0.2246	-0.4379	-0.1686	-0.2581	-0.0896	0.1804	0.2687	0.0883
		(0.0029)	(0.0029)	(0.0036)	(0.0024)	(0.0050)	(0.0054)	(0.0024)	(0.0051)	(0.0054)
	1	-0.1716	0.2142	-0.3857	-0.1160	-0.2272	-0.1112	0.1816	0.2467	0.0651
	(low)	(0.0098)	(0.0101)	(0.0111)	(0.0119)	(0.0136)	(0.0164)	(0.0122)	(0.0139)	(0.0166)
	2	-0.2763	0.2693	-0.5456	-0.2146	-0.3380	-0.1234	0.2156	0.3229	0.1072
		(0.0068)	(0.0067)	(0.0093)	(0.0053)	(0.0124)	(0.0134)	(0.0053)	(0.0124)	(0.0134)
	3	-0.3333	0.3396	-0.6729	-0.2736	-0.3929	-0.1193	0.2771	0.4021	0.1249
		(0.0067)	(0.0067)	(0.0091)	(0.0045)	(0.0126)	(0.0134)	(0.0045)	(0.0126)	(0.0134)
	4	-0.1940	0.2054	-0.3994	-0.1523	-0.2358	-0.0835	0.1677	0.2430	0.0753
		(0.0054)	(0.0054)	(0.0071)	(0.0033)	(0.0104)	(0.0111)	(0.0033)	(0.0104)	(0.0110)
	5	-0.0943	0.1131	-0.2073	-0.0864	-0.1021	-0.0158	0.0812	0.1450	0.0638
	(high)	(0.0043)	(0.0043)	(0.0056)	(0.0026)	(0.0084)	(0.0090)	(0.0026)	(0.0084)	(0.0090)

(continued next page)

Table VII
High-frequency quoting and asymmetry in quote changes (continued)

		Direction fixed effect			*Direction × HFQ interaction effect*					
Variable	*Sample*	$d = Bid$	$d = Offer$	*Difference*	$d = Bid$ and $HFQ = low$	$d = Bid$ and $HFQ = high$	*Difference*	$d = Offer$ and $HFQ = low$	$d = Offer$ and $HFQ = high$	*Difference*
MLM	Full	-0.1814	0.1789	-0.3603	-0.1470	-0.2159	-0.0689	0.1509	0.2070	0.0561
		(0.0037)	(0.0038)	(0.0047)	(0.0031)	(0.0066)	(0.0070)	(0.0031)	(0.0066)	(0.0070)
	1	-0.1378	0.1467	-0.2845	-0.1219	-0.1537	-0.0318	0.1434	0.1500	0.0066
	(low)	(0.0090)	(0.0093)	(0.0102)	(0.0110)	(0.0125)	(0.0150)	(0.0112)	(0.0128)	(0.0153)
	2	-0.1546	0.1621	-0.3167	-0.1372	-0.1720	-0.0347	0.1492	0.1749	0.0256
		(0.0067)	(0.0067)	(0.0092)	(0.0052)	(0.0123)	(0.0134)	(0.0053)	(0.0123)	(0.0133)
	3	-0.2404	0.2259	-0.4663	-0.2001	-0.2808	-0.0807	0.1940	0.2578	0.0638
		(0.0065)	(0.0065)	(0.0089)	(0.0044)	(0.0123)	(0.0131)	(0.0044)	(0.0123)	(0.0131)
	4	-0.1988	0.2027	-0.4015	-0.1489	-0.2488	-0.0999	0.1597	0.2457	0.0860
		(0.0077)	(0.0077)	(0.0102)	(0.0047)	(0.0148)	(0.0158)	(0.0047)	(0.0148)	(0.0157)
	5	-0.1597	0.1505	-0.3103	-0.1180	-0.2014	-0.0834	0.1017	0.1994	0.0977
	(high)	(0.0088)	(0.0089)	(0.0114)	(0.0053)	(0.0172)	(0.0184)	(0.0053)	(0.0173)	(0.0185)

Table VIII
Summary statistics for US equities

From the CRSP file, for each year, 2001-2011 and all stocks present in January through April of that year with share codes equal to 10 or 11, I draw 150 firms in a random sample stratified by dollar trading volume in January through March. NBB is the National Best Bid; NBO, the National Best Offer; CT, Consolidated Trade; CQ, Consolidated Quote. Trade and quote counts are from the Monthly TAQ database (one-second time stamps). Except for the number of firms, table entries are cross-firm medians.

	2001	**2002**	**2003**	**2004**	**2005**	**2006**	**2007**	**2008**	**2009**	**2010**	**2011**
No. firms	137	122	141	148	144	150	150	147	145	149	149
NYSE	106	46	51	44	48	44	55	53	56	54	47
Amex	16	4	10	12	8	15	14	6	5	14	6
NASDAQ	15	72	80	92	88	91	81	88	84	81	96
Avg. daily CT records (trades)	167	228	231	399	448	605	970	1,217	1,993	1,141	1,346
Avg. daily CQ records (quotes)	1,525	1,053	1,470	3,917	6,004	7,307	12,521	16,791	41,571	23,530	24,053
Avg. daily NBB changes	128	162	210	514	611	761	772	1,183	1,787	1,468	1,225
Avg. daily NBO changes	127	163	226	545	729	751	789	1,142	1,789	1,461	1,146
Avg. price (bid-offer midpoint)	$20.57	$20.98	$14.41	$16.53	$16.10	$21.14	$15.81	$14.12	$11.25	$16.79	$15.77
Market capitalization of equity, $Million	$976	$410	$205	$352	$348	$411	$480	$411	$382	$490	$690

Table IX
Wavelet variance ratios for US equities, 2001-2011

In each year 2001-2011, 150 US firms are randomly selected from CRSP (stratified by average daily dollar trading volume during the first quarter of the year). Quote records for April are taken from the NYSE Monthly TAQ database. Within each second, quotes are randomly assigned order-preserving millisecond fractional portions. The wavelet variance ratio is $V_j = 2^{J-j}\nu_j^2/\nu_J^2$ where ν_j^2 is the wavelet variance at level j and $J = 16$ is the highest level in the analysis. For a random-walk, the ratio would be unity at all horizons. All entries are cross-firm means. The National Best Bid and Offer are computed from TAQ data; the bid and offer are separately transformed using the Haar basis; variance estimates are formed as the average of the bid and offer variances. Estimates in Panel A are constructed from bids and offers that were filtered for errors, but not otherwise adjusted. Estimates in Panel B are constructed from denoised bids and offers (with short-term peaks clipped). Table reports cross-firm means, winsorized at ±5%.

Panel A. Wavelet variance ratios, V_j, computed from raw bids and offers

Level, j	Time scale	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011
1	50 ms	2.90	2.87	3.59	6.01	5.40	5.89	5.10	4.40	4.15	5.99	3.16
2	100 ms	2.88	2.77	3.55	5.79	5.21	5.68	4.75	3.98	3.81	5.24	2.86
3	200 ms	2.43	2.64	3.50	5.24	4.94	5.33	4.17	3.43	3.42	4.51	2.61
4	400 ms	2.36	2.54	3.44	4.80	4.56	4.78	3.58	2.96	3.02	3.90	2.32
5	800 ms	2.13	2.41	3.27	4.19	4.03	4.30	2.94	2.57	2.53	3.28	2.07
6	1,600 ms	2.02	2.24	3.04	3.47	3.33	3.47	2.61	2.29	2.25	2.77	1.92
7	3.2 sec	1.95	2.12	2.65	2.96	2.86	2.98	2.28	2.12	2.06	2.39	1.79
8	6.4 sec	1.90	1.97	2.43	2.64	2.54	2.61	2.10	1.98	1.88	2.07	1.66

Panel B. Wavelet variance ratios, V_j, computed from denoised bids and offers

Level, j	Time scale	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011
1	50 ms	1.55	2.19	2.88	5.09	4.89	5.61	4.70	4.18	3.53	5.71	2.94
2	100 ms	1.53	2.14	2.85	4.97	4.75	5.41	4.36	3.78	3.22	5.03	2.72
3	200 ms	1.52	2.11	2.80	4.77	4.55	5.07	3.77	3.23	2.92	4.36	2.48
4	400 ms	1.52	2.08	2.73	4.44	4.22	4.52	3.13	2.77	2.65	3.70	2.22
5	800 ms	1.53	2.04	2.64	3.97	3.72	3.81	2.63	2.42	2.35	3.13	2.03
6	1,600 ms	1.56	2.00	2.54	3.28	3.12	3.23	2.31	2.21	2.09	2.62	1.89
7	3.2 sec	1.65	1.96	2.40	2.85	2.69	2.77	2.12	2.06	1.90	2.26	1.77
8	6.4 sec	1.71	1.91	2.24	2.64	2.41	2.48	1.99	1.94	1.75	2.00	1.66

Table X Time scale volatility estimates for US equities, 2001-2011

In each year 2001-2011, 150 US firms are randomly selected from CRSP (stratified by average daily dollar trading volume during the first quarter of the year). Quote records for April are taken from the NYSE Monthly TAQ database. Within each second, quotes are randomly assigned order-preserving millisecond fractional portions. Panel A reports rough volatilities, $\sigma_j = \sqrt{\sum_{i=1}^{j} \nu_j^2}$, in \$0.01 per share; Panel B reports the rough volatility, σ_j, in basis points. Panel C reports rough variance ratios, VR_j. Table entries are cross-firm means and standard errors winsorized at $\pm$5%. All estimates are constructed from denoised bids and offers (with short-term peaks clipped).

Panel A. Rough volatility, σ_j, \$0.01 per share.

Level, j	Time scale	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011
1	50 ms	0.04	0.04	0.03	0.04	0.04	0.04	0.03	0.04	0.04	0.04	0.03
		(<0.01)	(<0.01)	(<0.01)	(<0.01)	(<0.01)	(<0.01)	(<0.01)	(<0.01)	(<0.01)	(<0.01)	(<0.01)
3	200 ms	0.10	0.09	0.07	0.11	0.10	0.11	0.07	0.09	0.11	0.10	0.08
		(<0.01)	(<0.01)	(<0.01)	(<0.01)	(<0.01)	(0.01)	(<0.01)	(<0.01)	(0.01)	(0.01)	(<0.01)
5	800 ms	0.20	0.19	0.15	0.21	0.20	0.22	0.14	0.17	0.22	0.19	0.15
		(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
7	3.2 sec	0.43	0.38	0.29	0.40	0.37	0.40	0.26	0.33	0.40	0.35	0.30
		(0.02)	(0.02)	(0.01)	(0.02)	(0.02)	(0.02)	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)
10	25.6 sec	1.37	1.04	0.79	0.98	0.93	1.01	0.70	0.90	1.07	0.89	0.78
		(0.09)	(0.04)	(0.04)	(0.04)	(0.04)	(0.05)	(0.04)	(0.04)	(0.06)	(0.05)	(0.05)
14	6.8 min	4.70	3.66	2.61	3.39	3.21	3.53	2.53	3.20	3.87	3.17	2.79
		(0.23)	(0.15)	(0.12)	(0.16)	(0.15)	(0.18)	(0.14)	(0.16)	(0.20)	(0.18)	(0.17)
16	27 min	8.78	6.77	4.69	6.10	5.80	6.47	4.74	6.04	7.36	6.03	5.33
		(0.45)	(0.29)	(0.22)	(0.29)	(0.27)	(0.33)	(0.28)	(0.30)	(0.40)	(0.36)	(0.34)

Table 8. Time scale volatility estimates for US equities, 2001-2011 (continued)

Panel B. Rough volatility, σ_j, basis points

Level, j	Time scale	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011
1	50 ms	0.19	0.21	0.21	0.31	0.30	0.22	0.19	0.29	0.38	0.25	0.19
		(0.01)	(0.01)	(0.01)	(0.02)	(0.02)	(0.01)	(0.01)	(0.02)	(0.02)	(0.01)	(0.01)
3	200 ms	0.50	0.55	0.55	0.80	0.78	0.55	0.46	0.71	0.94	0.61	0.47
		(0.02)	(0.03)	(0.02)	(0.06)	(0.06)	(0.03)	(0.03)	(0.04)	(0.05)	(0.03)	(0.02)
5	800 ms	1.06	1.14	1.14	1.59	1.53	1.07	0.89	1.35	1.82	1.15	0.92
		(0.05)	(0.05)	(0.05)	(0.10)	(0.11)	(0.05)	(0.05)	(0.08)	(0.09)	(0.06)	(0.05)
7	3.2 sec	2.22	2.29	2.20	2.82	2.72	1.97	1.61	2.52	3.35	2.08	1.74
		(0.10)	(0.11)	(0.09)	(0.16)	(0.18)	(0.08)	(0.09)	(0.14)	(0.16)	(0.10)	(0.09)
10	25.6 sec	6.59	6.33	5.80	6.93	6.54	4.91	4.12	6.43	8.40	5.04	4.47
		(0.31)	(0.31)	(0.23)	(0.35)	(0.36)	(0.18)	(0.20)	(0.31)	(0.36)	(0.21)	(0.21)
14	6.8 min	23.72	21.84	19.22	21.81	21.05	16.38	14.03	21.84	29.56	17.07	15.76
		(1.04)	(1.05)	(0.70)	(0.91)	(0.95)	(0.51)	(0.57)	(0.83)	(1.18)	(0.63)	(0.67)
16	27.3 min	43.30	40.08	34.55	38.77	37.00	29.40	25.34	39.83	54.23	31.52	29.25
		(1.87)	(1.93)	(1.27)	(1.61)	(1.51)	(0.88)	(0.90)	(1.27)	(1.97)	(1.10)	(1.14)

Panel C. Rough variance ratio, VR_j

Level, j	Time scale	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011
1	50 ms	1.55	2.19	2.88	5.09	4.89	5.61	4.70	4.18	3.53	5.71	2.94
		(0.06)	(0.09)	(0.10)	(0.41)	(0.43)	(0.59)	(0.51)	(0.44)	(0.22)	(0.52)	(0.15)
3	200 ms	1.52	2.13	2.82	4.87	4.67	5.25	4.09	3.55	3.09	4.74	2.61
		(0.06)	(0.08)	(0.10)	(0.38)	(0.41)	(0.55)	(0.42)	(0.35)	(0.18)	(0.40)	(0.13)
5	800 ms	1.53	2.08	2.71	4.29	4.05	4.37	3.09	2.77	2.62	3.63	2.21
		(0.06)	(0.09)	(0.10)	(0.31)	(0.33)	(0.42)	(0.26)	(0.23)	(0.15)	(0.28)	(0.10)
7	3.2 sec	1.61	2.02	2.52	3.33	3.15	3.36	2.42	2.28	2.16	2.70	1.92
		(0.07)	(0.09)	(0.09)	(0.20)	(0.21)	(0.28)	(0.16)	(0.16)	(0.11)	(0.18)	(0.08)
10	25.6 sec	1.74	1.81	2.09	2.52	2.23	2.35	1.88	1.81	1.67	1.86	1.56
		(0.08)	(0.07)	(0.07)	(0.14)	(0.10)	(0.14)	(0.09)	(0.10)	(0.06)	(0.09)	(0.05)
14	6.8 min	1.35	1.35	1.46	1.53	1.46	1.49	1.33	1.29	1.30	1.30	1.22
		(0.03)	(0.03)	(0.03)	(0.04)	(0.03)	(0.04)	(0.03)	(0.03)	(0.02)	(0.03)	(0.02)
16	27.3 min	1.11	1.11	1.15	1.17	1.15	1.16	1.11	1.09	1.10	1.09	1.08
		(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)

Figure 1. The bid and offer for AEPI, April 29, 2011.

National best bid and offer (NBBO) from the NYSE Daily TAQ dataset. The National best bid (bottom line, in blue) is the maximum bid, taken over all market centers reporting to the Consolidated Tape Association; the National best offer (top line, red) is the minimum offer.

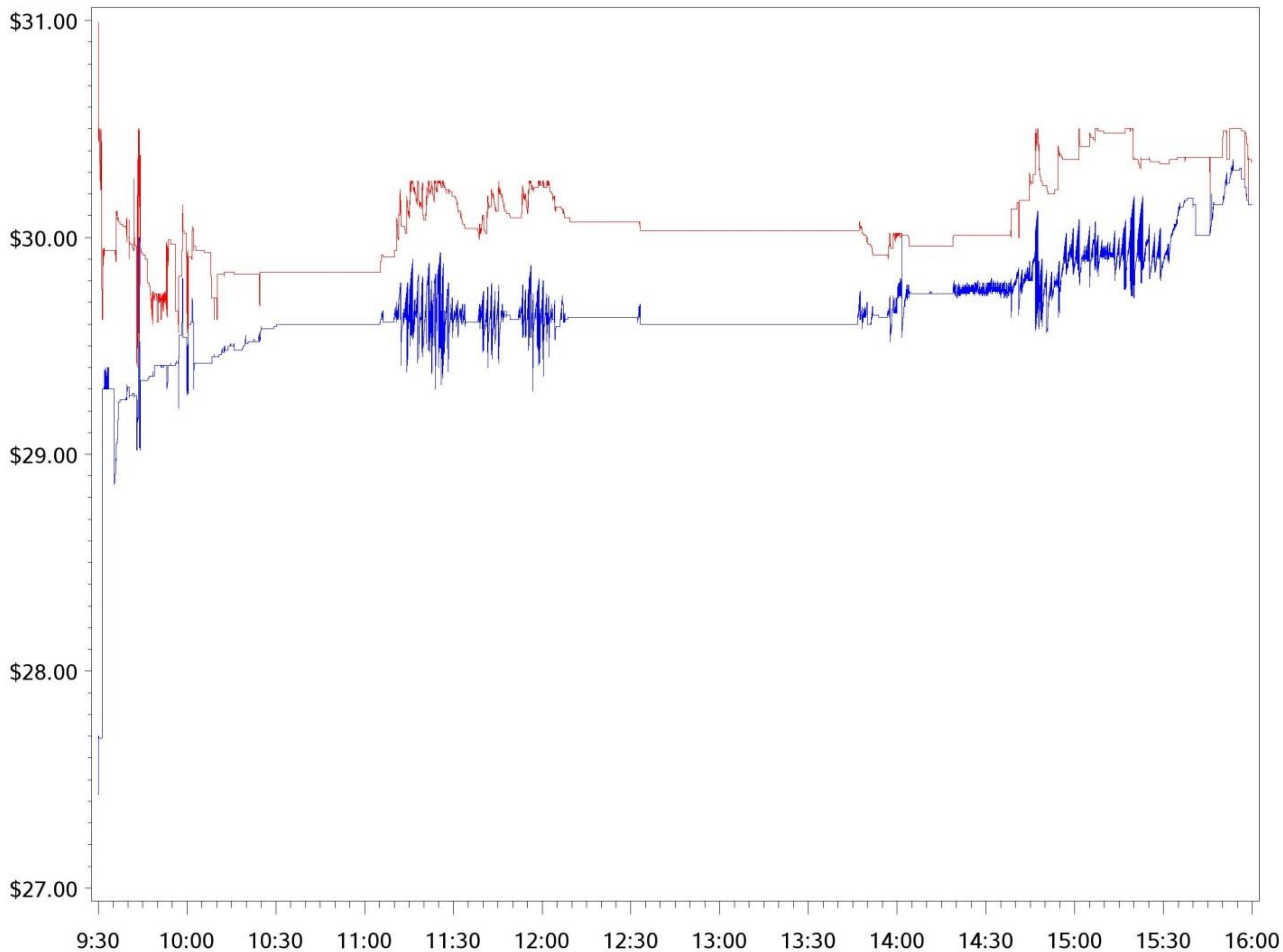


Figure 2. Haar basis for components of offer residual

The Haar function is defined over the real line as $\psi(x) = +1$ if $0 < x < 1/2$, -1 if $1/2 < x < 1$, and 0 otherwise. At level j the basis set contains $k = 1, \ldots, 2^{3-j}$ functions $\psi(t, j, k) = 2^{-j/2}\psi(2^{-j}x + k + 1)$. The top row ($j = 1$) contains the basis functions for the short-term component; the middle row ($j = 2$), the medium term component; and the bottom row ($j = 3$), the long-term component.

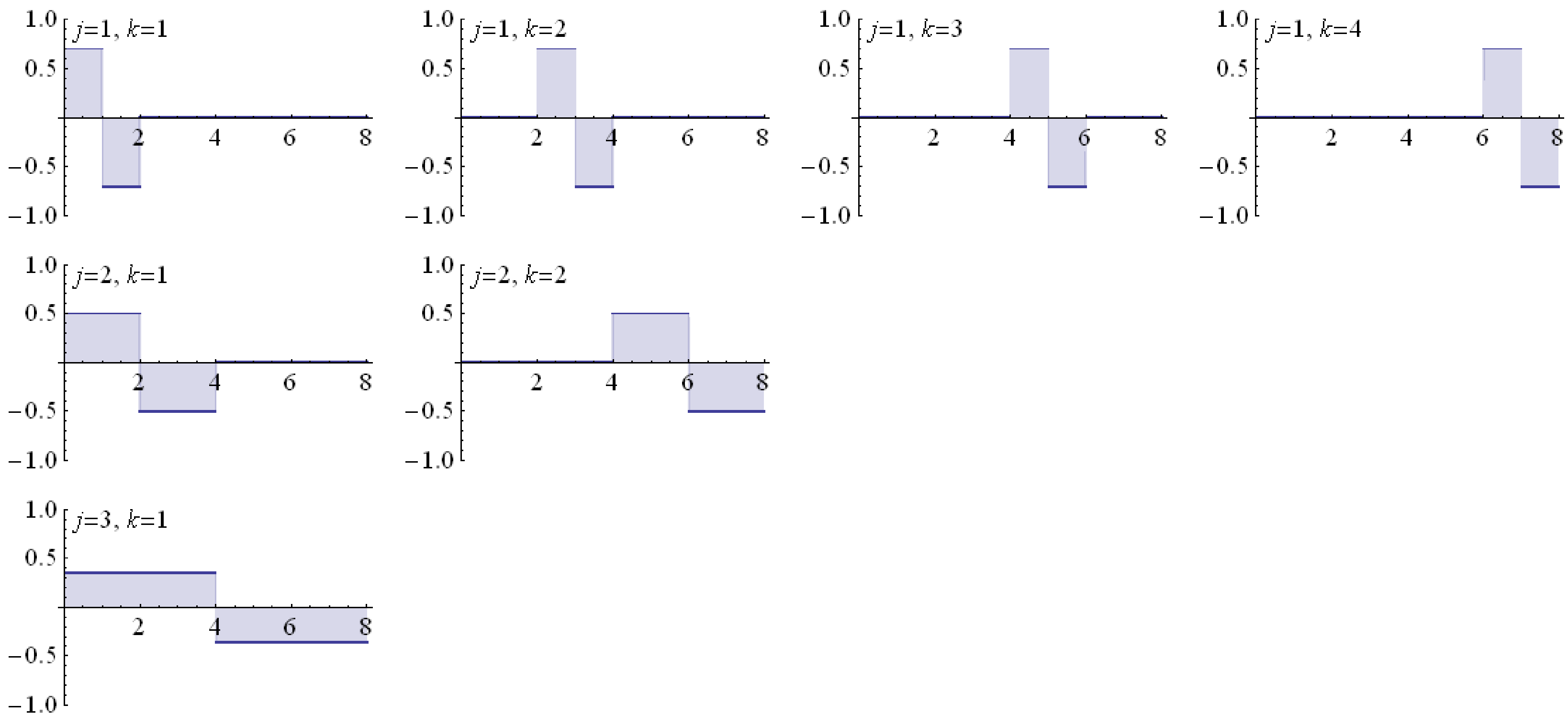
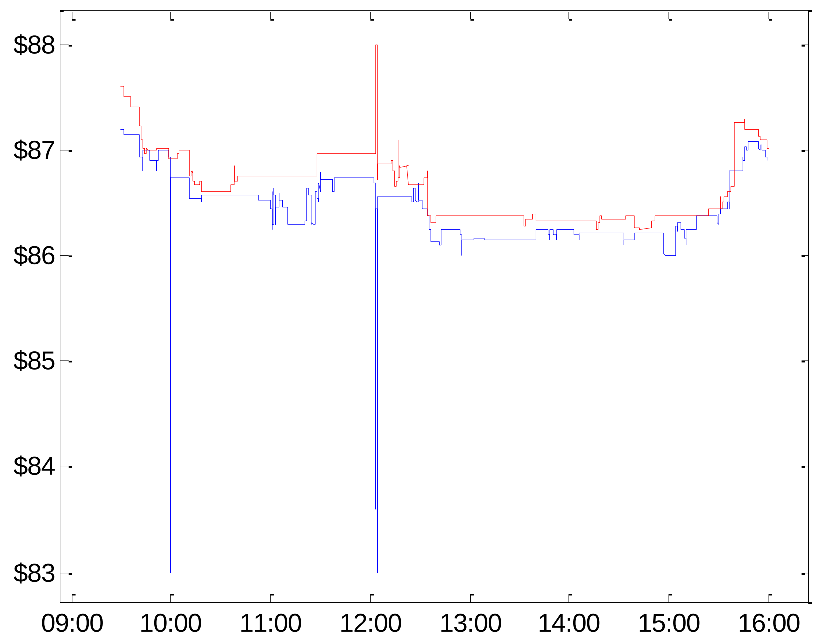


Figure 3 Bid and offer for PRK (Park National Corporation) on April 6, 2001

Panel A. National best bid and offer (NBBO) from the NYSE Daily TAQ dataset. The National best bid (bottom line, in blue) is the maximum bid, taken over all market centers reporting to the Consolidated Tape Association; the National best offer (top line, red) is the minimum offer.



Panel B. Rough component of the National Best bid, constructed from a Haar wavelet transform and comprising components at time scales of 51.2 seconds and lower. The bands demarcate ±$0.33, approximately 150% of the average bid-ask spread for the day.

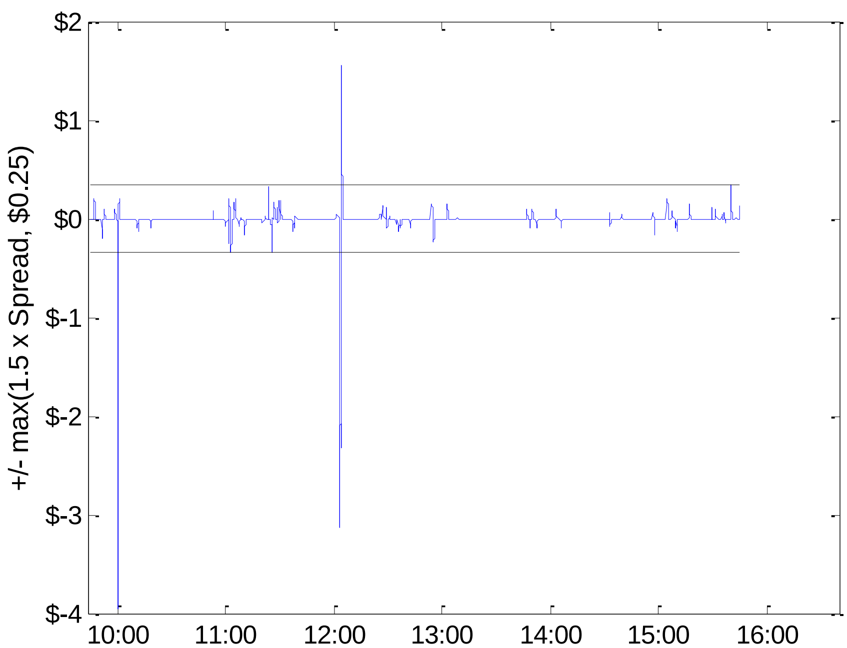
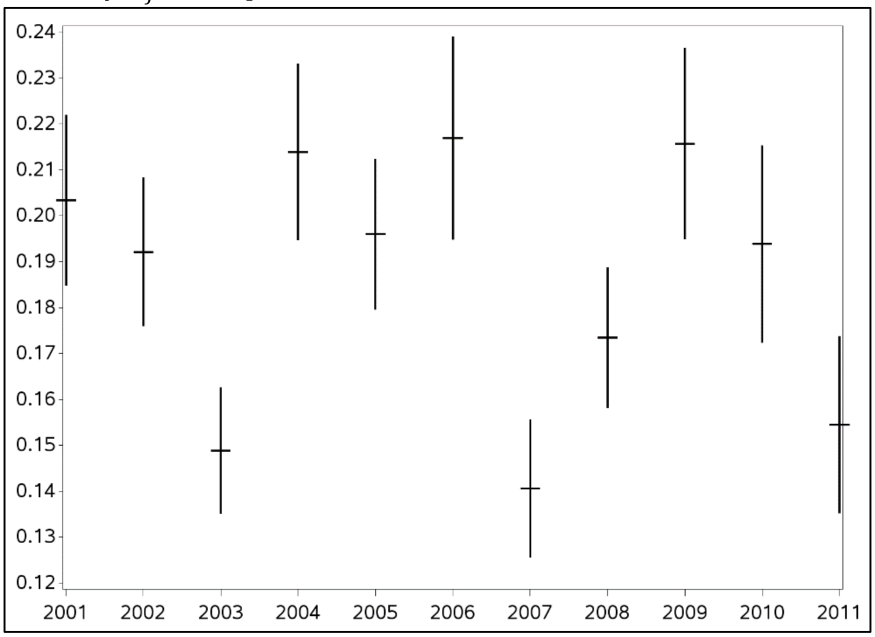


Figure 4.
Time scale volatility estimates for U.S. Equities, 2001-2011

Quote volatility at a time scale of 800 ms., $0.01 per share (Panel A), basis points (Panel B), and as a variance ratio (Panel C). In each year, the horizontal tick marks the mean and the vertical line demarcates the mean ± twice the standard error.

Panel A. Rough volatility, σ_j, $0.01 per share



Panel B Rough volatility, σ_j, basis points

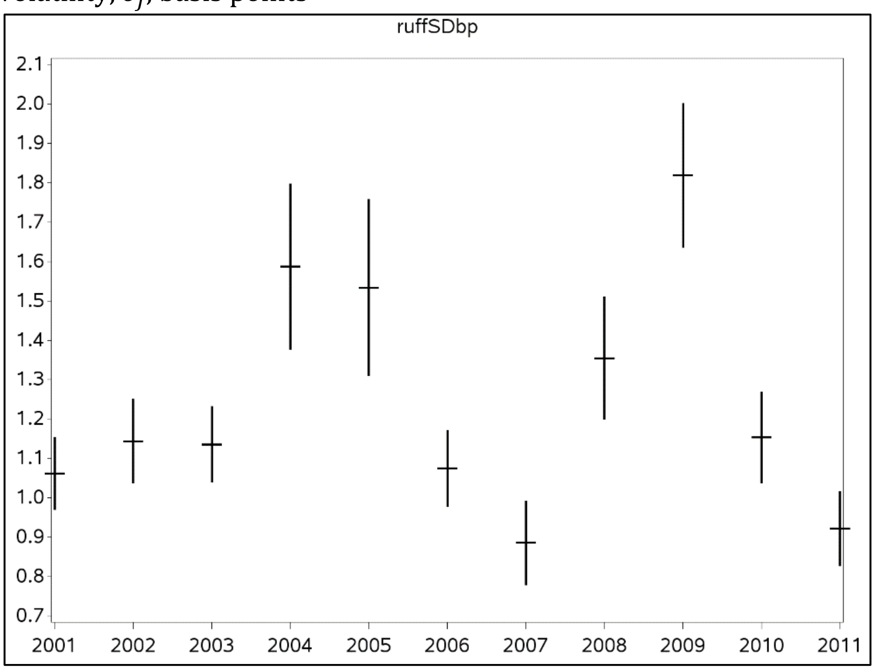


Figure 4. Time scale volatility estimates for U.S. Equities, 2001-2011 (continued)

Panel C. Rough variance ratio, $VR_{j=5}$