Price Discovery in High Resolution

Joel Hasbrouck

Department of Finance
Stern School of Business
New York University

March 1, 2017

This draft: November 20, 2018

44 West 4th St., New York, NY 10012. Email: jhasbrou@stern.nyu.edu

I am indebted to participants at the 2017 Finance Down Under Conference (University of Melbourne), discussant Jonathan Brogaard, the 2017 Erasmus Liquidity Conference, the 2018 SoFiE Conference (Lugano). I am especially indebted to the reviewers: Carole Comerton-Forde, Fulvio Corsi, Eric Ghysels, and Frank de Jong. All errors remain my own responsibility.

This paper was prepared for the Journal of Financial Econometrics and delivered as the 2018 Halbert White Jr. Memorial Lecture.

A companion computational appendix, programs, and related materials are available at http://people.stern.nyu.edu/jhasbrou/Research/HVAR/HVARindex.html.

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.

Price Discovery in High Resolution

Abstract

US equity market data are currently timestamped to nanosecond precision. This permits models of price dynamics at resolutions sufficient to capture the reactions of the fastest agents. Direct estimation of multivariate time series models at sub-millisecond frequencies nevertheless poses substantial challenges. To facilitate such analyses, this paper applies long distributed lag models, computations that take advantage of the inherent sparsity of price transitions, and bridged modeling. At resolutions ranging from one second down to ten microseconds, I estimate representative models for two stocks (IBM and NVDA) bearing on three topics of current interest. The first analysis examines the extent to which the conventional source of market data (the consolidated tape) accurately reflects the prices observed by agents who subscribe (at additional cost) to direct exchange feeds. At a one-second resolution, the information share of the direct feeds is indistinguishable from that of the consolidated tape. At resolutions of 100 and 10 microseconds, however, the direct feeds are totally dominant, and the consolidated share approaches zero. The second analysis examines the quotes from the primary listing exchange vs. the non-listing exchanges. Here, too, information shares that are essentially indeterminate at one-second resolution become much more distinct at higher resolutions. Although listing exchanges execute about one fifth of the trading volume, their information shares are slightly above one-half. The third analysis examines quotes, lit trades, and dark trades. At a one-second resolution, dark trades appear to have a small, but discernible, information contribution. This vanishes at higher resolutions. Quotes and lit trades essentially account for all price discovery, with information shares of roughly 65% and 35%, respectively.

KEY WORDS: High-resolution, high frequency trading, vector autoregression (VAR), vector error correction models (VECM), polynomial distributed lags, sparsity.

I. Introduction

Many important questions in empirical market microstructure turn on the joint dynamics of bids, asks, and last sale prices. These dynamics are central to identifying innovations in market data, and distinguishing permanent informational effects from transient behaviors. They bear directly on market structures and practices that facilitate (or impede) the incorporation of information into security prices, that is, the process of price discovery. This paper discusses and compares various approaches to modeling these dynamics at natural timescales that range from those of human decision processes down to the reaction times of algorithms. Modern markets convene agents active across the full range of this spectrum.

Although the present paper only considers resolutions down to ten microseconds, the NYSE's TAQ data are currently timestamped to nanosecond precision. Thus, while microstructure data are sometimes described as high frequency, the more striking feature is their high resolution.[1] The enhanced resolution is important because it allows us to identify information sets and feasible strategies that are blurred at coarser timestamps. The identification may derive from physical limits on transmission speeds or deliberate delays ("speed bumps") introduced as a feature of market design.

The paper's approach draws on several econometric themes. The specifications are standard vector autoregression and error correction models (VARs and VECMs), with the usual transformations to obtain random-walk variances and information shares. In microstructure applications, these are usually specified in event time or relatively long (one second or more) intervals of natural time.

To achieve high resolution in a natural time model, the paper follows the heterogeneous autoregressive (HAR) approach used for realized volatility forecasting by Corsi (2009). Essentially, within each timescale, model coefficients are constrained to a small number of values. Corsi's specifications involved daily, weekly and monthly terms, implying a factor of twenty or so between

[1] The NYSE's TAQ Consolidated Quote file for October 1, 1996 contained about 684 thousand records; the file for October 3, 2016, about 670 million records, that is, a factor increase of about one thousand. The timestamp precision over the same period, however, went from seconds to nanoseconds, a factor of one billion.

the shortest and longest components. The components in the present models range between ten seconds and ten microseconds (a factor of one million). Computation relies heavily on sparsity: the number of price updates is typically many orders of magnitude smaller than the number of intervals in the sample. The analysis also investigates computational efficiency achieved in bridging, whereby forecasts from high-resolution models are aggregated and used as inputs for longer-term forecasting based on lower-resolution models (discussed in Bańbura, Giannone, Modugno and Reichlin (2013)).

The full range of specifications are only estimated for one trading day (October 3, 2016) for one NYSE-listed stock (IBM) and one NASDAQ-listed stock (NVIDIA, ticker symbol NVDA). To examine the role of resolution, each analysis is conducted at a range of observation interval widths: from a low-resolution analysis conducted with one-second intervals down to a high-resolution analysis (at ten-microsecond intervals). I also augment the natural-time specifications with corresponding event-time models. The ten-microsecond and event-time specifications are also estimated over a longer 30-day sample, to construct means and standard errors.

The first analysis examines national best bids and offers (NBBOs) constructed from the consolidated tape (historically, the definitive record of market events), versus the NBBOs known to the exchanges' direct subscribers. At a one-second resolution, the alternatives appear to be informationally equivalent, but at the higher resolutions, the dominance of the direct subscribers' information sets becomes total. The second analysis examines the informational contributions of the bids and offers from the listing exchange versus those of all other exchanges. At low resolutions, the two sets of prices are indistinguishable. At the higher resolutions, the listing exchange contributes slightly more information than all other exchanges, combined. This information dominance is perhaps surprising given that the listing exchanges currently account for around one-fifth of executed trading volume. The third analysis examines quotes, lit trades, and dark trades. At a one-second resolution, dark trades appear to have a small, but discernible, information contribution. This vanishes at higher resolutions. Quotes and lit trades account for virtually all price discovery, with information shares of roughly 65% and 35%, respectively. The event-time and high-resolution natural-time results generally agree, but the upper and lower bounds of the event-time information shares are markedly higher.

The paper is organized as follows. Section II presents the model, which is set in natural time. Section III contrasts natural- and event-time approaches to microstructure modeling. Section IV describes the data. The paper then turns to the applications: the analysis of reporting delays (Section V); cross-exchange contributions to price discovery (Section VI); and the relative contributions of quotes and lit and dark trades (Section VII). Section VIII discusses computations based on bridged models. Section IX concludes the paper and indicates further directions.

II. Methodology

A. Representation

Hasbrouck (1995) specializes to microstructure settings the cointegration model of Engle and Granger (1987). The object of interest is a vector time series of prices $p_t = [p_{1t} \quad \cdots \quad p_{nt}]'$ that is assumed to have covariance stationary first differences. The prices are bids, offers, last-sale prices, and so forth, possibly from different markets, but all pertaining to the same security. For this reason, they are cointegrated, possessing one common trend. There are $n-1$ cointegrating vectors, which may be specified without loss of generality as $p_{1t} - p_{2t}, p_{1t} - p_{3t}, \ldots, p_{1t} - p_{nt}$. The cointegrating vectors are zero-mean. The dynamics are represented by vector error correction model (VECM):

$$\Delta p_t = \gamma B p_{t-1} + \phi_1 \Delta p_{t-1} + \phi_2 \Delta p_{t-2} + \cdots + \phi_K \Delta p_{t-K} + \epsilon_t$$

$$\text{where } B = \begin{bmatrix} 1 & \\ \vdots & -I_{n-1} \\ 1 & \end{bmatrix}, E\epsilon_t = 0 \; and \; E\epsilon_t \epsilon_s' = \Omega \; for \; t = s \; and \; 0 \; otherwise \tag{1}$$

The first term on the right-hand side is the error correction term. B defines the (prespecified) cointegrating vectors and γ is the $n \times (n-1)$ matrix of adjustment coefficients. The second term is the usual autoregressive portion of the model. For brevity of notation, the autoregressive part may be written as a polynomial: $\phi(L)p_t = \left[[\phi_1 L + \phi_2 L^2] + \cdots + \phi_K L^K\right]p_t$, where L is lag operator, $L^i p_t = p_{t-i}$. The VECM is assumed to be invertible, possessing a vector moving average (VMA) representation

$$\Delta p_t = \theta(L)\epsilon_t, \tag{2}$$

The VMA representation gives the impulse response function (IRF) subsequent to an arbitrary initial shock. In this application, the IRFs are used to depict the response of all prices to a shock in one of them. The VMA can be computed iteratively from the VECM. Under the assumption of covariance stationarity, the existence of VMA and VECM representations follow from the Wold theorem.[2]

Although the framework is conventional, the present applications exhibit some unusual features. The model is specified in natural time, and the intervals indexed by t are brief, down to 1×10^{-5} seconds (ten microseconds). The series of first differences, Δp_t, is very sparse. Bids, offers and last-sale prices persist in time, but at microsecond timescales they change relatively infrequently. Essentially, at these timescales all price transitions are jumps. Bids and offers are confined to a grid determined by the market's tick size, but the grid of transaction prices is much finer. The order of the VECMs is large, up to $K = 1 \times 10^6$. Importantly, though, none of these features is inherently incompatible with covariance stationarity. Furthermore, the assumed covariance stationarity is unconditional, and so does not rule out the possibilities of conditional refinements associated with time-variation in volatilities, arrival intensities, and so forth.

Equation (1) is a forecasting device, and is not intended to represent the data generating process. Although the Wold Theorem ensures that the ϵ_t are uncorrelated, it is all but impossible for them to be independent. The discreteness of the price realizations can generally only be obtained by allowing for extensive serial dependence in the higher-order moments of ϵ_t.

In addition to the VECM and VMA, the system also possesses a random-walk representation

$$p_t = m_t \iota + s_t \tag{3}$$

where m_t is a scalar random-walk process and s_t is a zero-mean covariance stationary process. It is natural in microstructure applications to identify m_t as the efficient price and s_t as the pricing error. The random-walk representation is not fully identified: m_t and s_t cannot be recovered

[2] Microstructure VECMs are sometimes used in situations where the cointegration relationships arise from no-arbitrage conditions. In such cases, the coefficients in the cointegration vectors will generally differ from ± 1. Specifications also frequently incorporate endogenous variables that are not cointegrated, such as signed orders or prices of other securities. This paper discusses high-resolution modeling for a particular VECM, but the approaches generalize to these other common cases.

without additional assumptions (Beveridge and Nelson (1981); Stock and Watson (1988)). The variance of the random walk increments is identified, however. Let $\theta(1)$ denote the value of $\theta(L)$ evaluated at $L = 1$, that is, the convergent sum, $\theta(1) = I + \theta_1 + \theta_2 + \cdots$. The quantity $\theta(1)\epsilon_t$ is the cumulative long-run predicted price changes implied by an initial shock ϵ_t. Due to the cointegration, the rows of $\theta(1)$ are identical. Intuitively, since all prices in the system refer to the same security, they are all predicted to move, in the long run, by the same amount. Let $\theta(1)_*$ denote any row of $\theta(1)$, and let $w_t = m_t - m_{t-1}$ denote the random-walk increment. Then

$$Var(w_t) = \sigma_w^2 = \theta(1)_* \, \Omega \, \theta(1)'_*. \tag{4}$$

The variance or standard deviation of the random-walk component is an important attribute of the model. If the ϵ_{it} are mutually uncorrelated, then Ω is diagonal. In this case, r.h.s. is a sum of *n* well-defined terms, with the *ith* term driven only by $Var(\epsilon_{it})$. The *i*th information share is $IS_i = Var(\epsilon_{it})/\sigma_w^2$, the proportion of the random-walk variance that is attributed to the innovations in the *i*th price.

If the ϵ_{it} are correlated, then the information shares are not uniquely defined. In such cases, they can be characterized by upper and lower bounds defined by alternative Cholesky factorizations of Ω. Many analyses simply report the midpoints of the range. Contemporaneous correlation in the innovations often arises from time aggregation, however, essentially the blurring of orderings within the interval. Higher resolution can directly alleviate these effects, thereby establishing bounds that are much tighter.

The information share is a popular price discovery measure, but others have been advocated. The alternatives include: the component share (CS, see Harris, McInish, Shoesmith and Wood (1995)); joint use of the IS and the CS, Yan and Zivot (2010); the unobserved components approach of De Jong and Schotman (2010); the tail dependence measure, Grammig and Peter (2013); the information leadership share, Putniņš (2013); and the information percolation share, Hagströmer and Menkveld (2017).[3] The present analysis uses information shares to illustrate the essential features of the high-resolution analyses, but it bears mention that most of the alternative

[3]The information and component shares are discussed at length in the Journal of Financial Market's special issue on price discovery measurement (Baillie, Booth, Tse and Zabotina (2002); de Jong (2002); Harris, McInish and Wood (2002, (2002); Hasbrouck (2002); Lehmann (2002)).

measures are also based on linear multivariate models. To the extent that contemporaneous correlations generally muddy causal and informational attributions, it is likely that these measures might also benefit from higher resolution.

The error correction terms, $\gamma B p_{t-1}$ in (1), are constructed in the usual fashion, as deviations lagged one period. These quantities are sometimes viewed as arising from arbitrageurs who learn about cross-market price discrepancies with delay (see Kumar and Seppi (1994), for example). At the shorter timescales considered here, however, it is likely that the delays are longer than one lag, and that for any given agent the delays differ across markets. If an arbitrageur learns about p_{1t} and p_{2t} with delays of δ_1 and δ_2 periods, her arbitrage flows would be driven by $p_{1,t+\delta_1} - p_{2,t+\delta_2}$. This suggests that variation in information sets might be explored by examining alternative timing in the error correction terms. Cointegration, though, is a long-term property of the system. Engle and Granger (1987) note, "In [the VECM] representation, only the disequilibrium in the previous period is an explanatory variable. However, by rearranging terms, any set of lags of the z [errors] can be written in this form, therefore it permits any kind of gradual adjustment toward the new equilibrium," (p. 255). In short, the convention of defining the disequilibrium as of $t-1$ does not impose any restrictions on the dynamics of the system.

B. Parameterization

With, say, $n = 4$ prices and $K = 1 \times 10^6$ lags, the VECM in (1) possesses over sixteen million coefficients. One approach to reducing the size of the parameter space is to constrain these coefficients to be constant over predefined lag ranges. In the context of realized volatility forecasting, Corsi (2009) suggests a univariate heterogeneous autoregressive (HAR) model. Corsi's specification is heterogeneous in timescales, and forecasts daily realized volatility as a linear function of daily, weekly and monthly components. His equation (8) is:

$$RV_{t+1d}^{(d)} = c + \beta^{(d)} RV_t^{(d)} + \beta^{(w)} RV_t^{(w)} + \beta^{(m)} RV_t^{(m)} + \omega_{t+1d}, \qquad (5)$$

where the *RV*s on the r.h.s. are estimated over daily, weekly, and monthly lags, and the βs are scalar coefficients. It is worth emphasizing that this is a daily forecasting model, and that all components are observed at the highest (daily) frequency. It differs in this respect from typical mixed data

sampling (MIDAS) situations wherein the forecasts may be updated frequently, but some of the predictors are only observed at lower frequencies.

Since realized volatilities are additive, specification (5) can also be expressed solely in terms of the daily RVs:

$$RV_{t+1d}^{(d)} = c + \phi_0 RV_t^{(d)} + \phi_1 RV_{t-1d}^{(d)} + \cdots + \phi_{22} RV_{t-22d}^{(d)} + \omega_{t+1d} \quad (6)$$

The lag over 22 trading days captures the monthly horizon. This is equivalent to (5) if the coefficients are constant within each of the sets $\{\phi_0\}, \{\phi_1, \dots, \phi_4\}$, and $\{\phi_5, \dots, \phi_{22}\}$. That is, the coefficients are constrained to lie on a step function. Despite the equivalence of the two specifications, they might give the appearance of different modeling strategies, with (5) emphasizing smoothing (or pre-averaging) of the data, and (6) suggesting a smoothing of the model parameters.

Bollerslev, Patton and Quaedvlieg (2016) summarize the current state of the HAR model. They note that, "[It] has arguably emerged as the most widely used realized volatility-based forecasting model." It has furthermore served as a useful starting point for extension. From a statistical perspective, the present paper adapts the modeling logic of the HAR model to VAR/VECM settings. The HAR coefficient scheme preserves high resolution at short lags, and tolerates lower resolution at longer lags. The same principle is applied here.

The paper explores sequences of models estimated at progressively higher resolutions. The data sample is defined by a finite interval $(0, T]$ over which each price has a finite number of jumps. This sample is partitioned into subintervals of width d, $d \in \{1\ s, 100\ ms, 10\ ms, 1\ ms, 100\mu s, 10\mu s\}$. Thus, d denotes the resolution of an analysis. In any given interval, the modeled datum is the price established as of the end of the interval. For simplicity, all analyses have the same maximum lag in natural time (10 seconds). In this framework, a representative equation for Δp_{it} in the VECM will have the form

$$\Delta p_{i,td} = \gamma_i B p_{td-1} + \sum_{j=1}^{n} \sum_{k=1}^{M} \phi_k^{ij} \Delta p_{j,td-k} + \epsilon_{it} \quad (7)$$

where $M = 10/d$, and $t = 1, \dots, T/d$ (ignoring the handling of initial values).

For each i, j pair there are M autoregressive coefficients. Suppressing the i, j indices for expositional clarity, they are constrained as follows. For $d = 1s$, ϕ_1 is unrestricted, and $\phi_2 = \cdots = \phi_{10}$. For $d = 0.1s$, ϕ_1 is unrestricted, $\phi_2 = \cdots = \phi_{10}$, and $\phi_{11} = \cdots = \phi_{100}$; for $d = 0.01s$, ϕ_1 is unrestricted, $\phi_2 = \cdots = \phi_{10}$, $\phi_{11} = \cdots = \phi_{100}$, and $\phi_{101} = \cdots = \phi_{1,000}$. Essentially, with each ten-fold increase in resolution, the nearest lag term is replaced by ten higher-resolution terms. Within this latter set, the first coefficient is unrestricted, and the last nine are set to the same value. Table 1 describes the lag intervals covered at each level of resolution. Note that the maximum lag is constant in natural time, ten seconds across all resolutions. Thus, the progression to higher resolution at short lags is not achieved at the expense of progressively lower resolution at long lags.

The motivation for this coefficient arrangement is primarily reduction in the size of the parameter space to achieve computational tractability. Equivalently, though, the restrictions can be viewed as constructing, on the right-hand side of (7), time-aggregations of the high-frequency data. In addition to the HAR model in Corsi (2009), similar terms arise in the HAR-RV-J model (Andersen, Bollerslev and Diebold (2007), and the step-function MIDAS specification (Forsberg and Ghysels (2007),Ghysels, Sinko and Valkanov (2007). Ghysels, Sinko and Valkanov also discuss generalizations based on polynomial distributed lags (PDLs) and beta polynomials.

Although PDLs have been used in microstructure VARs, the lengths involved are typically small. The segments used in Hasbrouck (1995, 2003), for example, extend to 300 lags (five minutes of one-second data). The present applications use lags up to one million. These lengths require special computational approaches. There are two costly computations: assembly of the cross-product matrix and inversion of the VECM to obtain the VMA. Relying on sparsity and the general form of the constraints, the first of these calculations, normally involving nested summations over all lags of all variables, may be reworked to run over the non-zero data values only. The second calculation may not be needed for some purposes: the random-walk volatility and the information share bounds can be determined from the VECM estimates using the representation theorem of Engle and Granger (1987).[4] Only if we wish to construct the impulse response functions is the inversion of the full VECM necessary. This is facilitated computationally by treating long polynomial

[4] This approach is discussed by de Jong (2002) and De Jong and Schotman (2010). I'm grateful to Fulvio Corsi and Frank de Jong on this point.

sums of lagged disturbances as state variables that are updated rather than fully recomputed at each forward iteration of time. Construction of the cross product matrix and the inversion of the VECM are discussed more thoroughly in the computational appendix to this paper (available at http://stern.nyu.edu/~jhasbrou).

The coefficient constraints render estimation of a large multiscale multivariate time series model computationally feasible. This is an important point in its favor. It is also likely, however, to cause some degree of misspecification. There are two immediate concerns. Firstly, the Engle-Granger result regarding the irrelevance of the $t-1$ convention in defining the cointegration errors (noted above) holds only for unrestricted VECMs. Secondly, the errors in the constrained specification may not be uncorrelated, and this will complicate determination of standard errors. These limitations pertain specifically to the model proposed here. The usual general concern with economic inference also applies. That is, the economic implications regarding attribution of information may be incorrect if the expectations formed by market participants differ from those implied by the model.

Statistical models of security prices that span many timescales are generally economically motivated by the observation of clienteles active at different horizons, ranging from high-frequency traders to long-term investors. These interactions implicitly map to latent statistical components that are also differentiated by timescale. Corsi invokes this logic, and refers to Müller, Dacorogna, Davé, Pictet, Olsen and Ward (1993). Crouzet, Dew-Becker and Nathanson (2016) suggest a frequency-domain equilibrium model.

C. Statistical properties of the estimates

The equations that comprise the VAR/VECM are estimated by least squares. It should be noted at the outset that by traditional measures (such as R^2), fit quality is likely to be poor. The situation is analogous to that encountered in ARCH/GARCH models when they are viewed as forecast models for squared or absolute returns. It is more appropriate to view them as forecasts of unobservable conditional variances (Andersen and Bollerslev (1998)). In the present case, the VAR/VECMs are in a sense implicitly forecasting conditional event arrival intensities, where the probability of an event in any one observational interval is extremely small.

Hamilton (1994) and Lütkepohl (2005) discuss the asymptotic distribution of VECM parameter estimates and impulse response functions. (The asymptotic distribution of information shares can be obtained by the delta method.) These results assume, however, that the model is correctly specified. While this might be the case in present applications, it would certainly be prudent to consider robust alternatives. Corsi (2009) reports Newey-West standard errors on his HAR specifications. In principle, Newey-West could be applied to high-resolution models, but the calculations would require calculation of all residuals, a computation that would not benefit from the sparsity of the original data. The same consideration applies to bootstrapping.

Microstructure data samples, however, are quite large. This suggests, following Bartlett (1950), constructing a series of point estimates from non-overlapping subsamples, and basing inference on the distribution of the independent estimates. The present paper adopts this strategy. Typically, though, microstructure applications involve panels of firms tracked over time. In such studies the standard errors of, say, individual firm-day estimates, may be of lesser importance. Tests of broader hypotheses are accomplished by examining coefficients in panel regressions that can easily incorporate firm and time fixed and random effects.

III. Natural and event time

Microstructure data are often modeled in event time. That is, the time index is treated as a counter over trades, quote updates, and so forth. The computations involve a much smaller number of observations and so pose no special computational difficulties. The connection with natural time is severed, but this may be acceptable in many situations, as the economic models underlying the statistical constructs are often set in notional time. Security payoffs are ultimately linked to natural-time physical production and consumption processes, but informational and strategic dynamics are likely to be more flexible.

At daily timescales, price volatility is strongly connected to the pace of trading. Consequently, event time, "trading time" and "business time" have often been chosen as the implicit clock of the trading process (see the discussion in Shephard (2005)).[5] Event-time clocks are widely

[5] If variation in some latent information intensity process simply sped up or slowed down market events, though, we'd expect that the arrival rates of different events (quotes, trades, market orders, limit orders and so on) would rise and fall, proportionately and in unison. An analysis of the NYSE's

used in price discovery analysis. Brogaard, Hendershott and Riordan (2015), for example, estimate a rich event-time model of orders originating from high-frequency traders (HFTs) and non-HFTs. Because lags that are long and variable in clock time may be short and regular in event time, event-time models are generally much more computationally tractable than natural-time models. In view of these considerations, I also estimate here event-time counterparts to the natural-time models.

Event-time specifications may be difficult to interpret, though, if natural time plays a role in the event definition. An event might stem from an action that in the very short run would have been available only to HFTs, but at longer intervals might have been used by anyone. A larger consideration is that an event clock is inherently tied to the event space. Adding events (such as trades or quotes for another security) or refining an existing event (distinguishing trades occurring on different exchanges, for example) changes the clock, possibly changing the event-time separation between unrelated events. Furthermore, it is difficult to map the forecasts from event-time models into natural time.

Event-time modeling suppresses the informational and strategic content of interarrival durations, implicitly suggesting that in the absence of an event nothing of economic consequence is occurring. Many economic models, however, suggest otherwise. In Easley and O'Hara (1992), non-occurrence of trade causes dealers to revise (downward) their conditional beliefs about the probability of an information event.

Midway along the continuum between event- and natural-time approaches are models that treat the location of events in natural time as exogenous, and then, conditional on these arrival times, allow the event-time VAR to incorporate some dependence on interarrival times. Dufour and Engle (2000) suggest an ACD model for trade occurrences. Then, letting T_i denote the interarrival duration for the *i*th trade event, the event-time VAR includes adjustments for $\ln(T_i)$ that effectively weight trades in the distant past less heavily than recent trades. This provides a connection to natural time, but only in one direction. The posting of an aggressively-priced bid, for example, does not influence the arrival rate of orders. This model can model a trade-price dependence that weakens over time. Forecasting, however, requires simulation to generate the event times.

TORQ events (a 1990 sample) suggested that although there was some commonality, time variation in the diverse arrival rates was far from homogenous (Hasbrouck (1999)).

Additionally, if there are different types of events, the approach requires multiple arrival models, or some sort of discrete choice mechanism to generate the event type at a given arrival time.

Timescale properties of economic time series are sometimes characterized using Fourier or wavelet representations.[6] These transformations do not usually highlight the innovations in the series, however, which makes them less attractive for informational attributions. There are also computational complications because the transforms do not preserve the sparsity of the initial data. (The Fourier transform of a series of length *N* will generally have *N* nonzero values even if the original series has only one nonzero value.)

IV. Data

The study examines two actively-traded stocks, IBM and NVIDIA. IBM's primary listing exchange is the NYSE; NVDA's is NASDAQ. IBM is a Dow stock; NVIDIA is not. Some of the results are based on a detailed analysis of October 3, 2016 (the first trading day in the month). The sample was subsequently extended through November 11, 2016 (for a total of thirty trading days). Key statistics are averaged over the thirty daily estimates and reported along with their standard errors. The data are taken from the WRDS SAS copy of the NYSE's Daily TAQ database, (NYSE (2017)). Each trade or quote record contains two timestamps, a participant time and a SIP (securities information processor) time. These are discussed in the next section. Both NASDAQ and the NYSE open trading around 9:30 with a single-price double-sided auction. As the auction mechanism differs substantially from regular continuous trading procedures, the present estimations are confined to the 9:45-16:00 interval. Table 2 reports summary statistics. Quote updates are clearly the most voluminous series, but often only a fraction of these will correspond to substantive changes. For example, of the 893,413 quote updates for NVDA, only 28,128 of these reflected a change to the NBB or NBO. (The other updates were caused by bid or offer changes away from the NBBO or changes to posted sizes.)

[6] Gençay, Selçuk and Whitcher (2002) discuss wavelet analysis of market data. Recent applications include Chinco and Ye (2016) and Hasbrouck (2018).

V. Timestamps and reporting delays

Bartlett and McCrary (2016) discuss in depth the timing conventions in US equity markets. The following material summarizes features essential for present purposes. In a computerized market, the matching engine is the innermost system that implements the essential features of the trading protocol.[7] The matching engine's timestamp is the best available indication of the occurrence time of an event, such as a quote update or execution. It is identified in the TAQ data as the participant timestamp.

The event is transmitted to the participant market's direct subscribers (including members with trading privileges). It is also transmitted (simultaneously, by law) to a securities information processor (SIP), in the present case, the Consolidated Tape Association (CTA), which broadcasts over multiple connections (technically, "multicasts") the event more widely. Traditionally, the CTA record (the consolidated tape) has been viewed as the authoritative source of market data. Prior to June 28, 2013, the CTA timestamp was the multicast dissemination time. Subsequently (and currently) it simply indicates the time when "processing the message was completed [by CTA]". In the TAQ documentation it is simply denoted "Time", but for the sake of clarity it will be referred to herein as the SIP time. Essentially the participant time marks the earliest time that the fastest traders could have learned about the event, while the SIP time marks the event's appearance on the consolidated feed.

Ding, Hanna and Hendershott (2014) examine the lag in CTA reporting. Their data consist of messages received by a vendor located at the BATS data center over a brief period in 2012. The vendor received quote updates via direct and consolidated feeds, allowing DHH to assess latency and to construct two alternative NBBOs. In one sense, the lag in the consolidated feed is small, averaging only of 1.5 milliseconds. DHH also find, however, numerous brief discrepancies (dislocations) in the best bid and offer, depending on which source is used. They conclude that the dislocations are costly for frequent traders.

Bartlett and McCrary (2016) examine the differential lags for the Dow-Jones stocks in a sample (August 2015 through June 2016) based on the participant and SIP times on the TAQ

[7] Authorization, checking for potential trade-through violations, routing, advanced order handling (such as pegged or discretionary orders), drop copy, and similar functions are often handled by auxiliary systems that are distinct from the matching engine.

database. They find that the consolidated quote updates lag the direct feeds by an average 1.1 ms, but that the lag in trade reports is much larger (21.6 ms). The differential lag between trades and quotes is consistent with participants' incentives. A bid that has been raised or an offer that has been lowered must be disseminated to be protected under the Reg NMS order protection rule. A bid that has been lowered or an offer that has been raised signals withdrawal of a previously available price, and if not indicated promptly could give the impression that the market doesn't honor its quotes.[8]

The present analyses are dependent on the precision and synchronization of the timestamps. In the current daily TAQ specification, timestamp fields allow for nanosecond precision (NYSE (2017)). The timestamps for October 3, 2016, however, are generally only populated to microsecond precision. The CTA requires exchanges to synchronize their clocks to UTC to within one hundred microseconds, but this tolerance is generally bettered. Exchange participants indicated to the SEC that, "… absolute clock offset on exchanges averages 36 microseconds," U.S. Securities and Exchange Commission (2016, p. 249). Bartlett and McCrary provide additional discussion.

For each trade and quote record I compute a measure of reporting delays as the difference, denoted δ, between the SIP and participant timestamps, measured in milliseconds. Table 3 summarizes the sample distribution, by stock and by participant. Excepting the FINRA ADF, all of the participants are exchanges. The FINRA ADF is the alternative display facility, a reporting system widely used for dark trades. There is no corresponding system for quotes. In principle, δ shouldn't be negative. A few negative values are present in the sample, however, reflecting clocks that are not precisely synchronized. At most exchanges, the median values are well under one millisecond. With the exception of the FINRA ADF subsample, the distributions are tight.

The present analysis constructs NBBO's based on participant and SIP timestamps. Instead of using these as benchmarks for executions, I study the joint dynamics using the high-resolution VECMs described earlier. The estimates provide a good starting point for other analyses because the effects of the participant/SIP time differentials are easy to understand: the NBB and NBO

[8] The quote/trade timing differential existed well before Reg NMS and the modern era of electronic markets (see, for example, Lee and Ready (1991)).

constructed from SIP times are noisy and delayed signals of the NBB and NBO computed using participant timestamps.

The VECM system comprises four variables. $NBBpart$ and $NBOpart$ are constructed from the participant timestamps. I order all quotes by participant times and build a running record of the bid and offer posted by each exchange. $NBBpart$ and $NBOpart$ are the max and min of the bid and offer over all exchanges' current quotes. Following the usual CTA practice, zero bids or offers indicate that an exchange has withdrawn its quote. When this happens, the exchange is dropped from the max or min calculation (until it posts a valid bid or offer). $NBBsip$ and $NBOsip$ are similarly constructed but using the SIP timestamps. If the reporting delay were a constant δ_0 at all exchanges, we would have $NBBsip_t = NBBpart_{t+\delta_0}$ and $NBOsip_t = NBOpart_{t+\delta_0}$, and the VECM system would be singular. The randomness in the delays removes this determinacy.

I estimate the system in natural time at resolutions of 1.0, 0.1, 0.01, 0.001, 0.0001, and 0.00001 seconds. The parameter estimates are not reported, for the sake of brevity, but they may be characterized as follows. At the lowest (one-second) resolution, the coefficient estimates are noisy, with a few t-statistics around two, but with most insignificant. At higher resolutions, the estimates are generally very significant, with t-statistics in the hundreds, particularly for the shorter lags and the error correction coefficients. Table 5 reports the random-walk volatilities and information shares.

For ease of interpretation, the random-walk volatilities (σ_ws) are scaled to units of dollars per share, per year.[9] An approximate annual return volatility is obtained by dividing the reported σ_w by price per share. Using the average share prices from Table 2, these volatilities are approximately 12% for IBM and 18% for NVDA. The random-walk volatility, σ_w, is in principle a property of the long-run behavior of the system. It should not depend on the modeling of the short-term high-resolution components. This is in fact the case: the estimates vary minimally across resolutions. This is a general property of all the natural time analyses in this paper.

---

[9] The details are as follows. For a given resolution, let d denote the width of an interval in seconds. The units of the Δp_t in the VECM are dollars per share per d. The original units of σ_w^2 are therefore $[(\$/shr)/d]^2$. This is scaled to an annual variance by multiplying by a factor:

$$\frac{250\ trading\ days}{year} \times \frac{6.5\ hours}{trading\ day} \times \frac{3{,}600\ seconds}{hour} \times \frac{1}{d}.$$

The random-walk volatility reported in the table is the square-root of the annual variance.

To highlight the relative contributions of participant-based and SIP-based quotes, the table reports grouped information shares, computed over variable sets $\{NBBpart, NBOpart\}$ and $\{NBBsip, NBOsip\}$. As usual, these shares can only be bounded. At the lowest time resolution, the bounds are uninformative, identical for the participant and SIP groups, and spanning the unit interval. At the highest resolutions, however, the bounds are quite narrow, attributing essentially all the price discovery to the participant data. The tightness of the bounds is closely linked to the innovation correlations in the off-diagonal block $(NBBpart, NBOpart) \times (NBBsip, NBOsip)$. The largest value is near unity in the one-second analysis and close to zero in the ten-microsecond estimates. The overall pattern is clear and sensible: variation that appears contemporaneous over long intervals is easily picked apart at higher resolutions.

As a point of comparison, I estimate event-time specifications. For each price I take the last value in a given microsecond (the precision of the timestamps), then merge and sort on time. After that step, the timestamps are dropped and the data are treated as sequenced events. The event-time VECM includes ten lags of all prices and the coefficients are unrestricted. Table 5 reports the estimated min and max information shares. (Estimated random-walk volatilities are not reported: it is unclear how one might assign an economic interpretation to an event-time volatility.)

The estimated event-time information shares are quite close to their high-resolution natural-time counterparts, and they assign virtually all the leadership to the prices with participant timestamps. This is not surprising because the delay in realizing the SIP prices is random but largely mechanical. It is captured quite adequately in an event-time specification.

Panel B of Table 5 reports summary statistics for daily estimates of the 10-microsecond natural-time and event-time specifications (over the thirty-day sample). These results are consistent with the October 3 results: the information shares attribute virtually all price discovery to the NBBO constructed from participant timestamps. For NVDA the volatility is higher in the longer sample: on the last day (November 11) a positive quarterly report caused a 30% gain.

The random-walk volatility and the information shares characterize the persistent effects of the innovations. The transient dynamics are best illustrated by the impulse response functions. For brevity, a representative example is reported is lieu of the full set of IRFs. Figure 1 depicts, at each of the six resolutions, the cumulative response in $NBBsip$ to a one-dollar shock in $NBBpart$ at

time zero. To clarify the short-run behavior at higher resolutions, the time axis is logarithmic. On a logarithmic scale "time zero" is not displayed. The first update in the one-second analysis occurs at one-second, the first-update in the 0.1 second analysis, at 0.1 seconds, and so forth. Panel A contains the IRFs for IBM; panel B, for NVDA.

The IRFs exhibit some distinctive visual features. They are kinked, a direct consequence of the step functions on which the VECM coefficients are constrained to lie. Many of the IRFs also exhibit a "scalloped" appearance. This is a visual artifact created by the logarithmic timescale. Between step transitions, the IRF is roughly linear. The concavity of the log function induces (on the horizontal axis) a distortion that produces the scalloped paths.

For both stocks the IRFs estimated at one-second resolution differ visibly from the higher-resolution estimates, with strong differences in short- and long-run behavior. This is a reflection of the apparent noisiness in the one-second estimates noted earlier. As we move to higher resolution analyses, however, the IRFs become more consistent.

Adjustment paths exhibit some overshooting and reversion. For the paths associated with the 100ms, 10ms and 1ms resolutions, most of the initial adjustment appears to take place at the first step ahead; the reversion is subsequent. The paths implied by the 100μs and 10 μs resolution analyses, on the other hand, show a gradual initial adjustment. Interestingly, the 10 μs IRF displays no adjustment in the first ten steps (through 100 μs). This is consistent with the finding that the SIP/participant lags found in the earlier studies are around at least 100 μs. Although not unexpected, this result affirms that nothing in the empirical approach inherently attributes adjustment where none is present.

Several other features of the graphs merit comment. It was earlier noted that the long-term behavior, summarized in the random-walk volatility, is unaffected by resolution. The IRFs at different resolutions, however, converge to visibly different initial values, despite having been constructed from the same initial shock. The source of this apparent inconsistency lies in the contemporaneous correlation of the innovations. At lower resolutions, these correlations are substantial. The initial disturbance driving the IRF, in putting a one-dollar shock on $NBBpart$ and setting the other disturbances to zero, ignores the contemporaneous effects that are captured more completely at higher resolutions.

VI. Price discovery across exchanges

Economists and regulators have long debated the merits of fragmented and consolidated financial markets.[10] Numerous studies, therefore, examine relative informational contributions of exchanges and trading mechanisms. Two of the early papers that suggested price discovery measures (Harris, McInish, Shoesmith and Wood (1995) and Hasbrouck (1995)) focused on US equity markets. Recent representative studies cover: Swiss equities, Grünbichler, Kohler and von Wyss (2017); bid and ask quotes for NYSE and Spanish stocks, Pascual and Pascual-Fuster (2014); recent US equity markets, Ozturk, van der Wel and van Dijk (2017). The last study examines cross-sectional and intraday time variation in its price discovery measures.

In the 1990's NYSE samples studied by Harris et al and Hasbrouck, the NYSE's informational contribution is overwhelming. (Harris et al find the NYSE's adjustment to other exchanges' prices to be small; Hasbrouck finds an average NYSE information share of 91.3%.) In that era, however, the NYSE dominated trading in its listed securities. In the Hasbrouck 1993 sample, the average NYSE market share is 84.3% by share volume. It is somewhat lower by number of trades (54.6%), but these were viewed as originating primarily from uninformed retail traders. Thus, to a good approximation, the NYSE information share mirrored its share of trading volume.

In more recent years, and particularly post Reg NMS, activity is more dispersed, and the market shares of listing exchanges are lower. In their 2013 sample, Ozturk et al find the NYSE's share of NYSE-listed securities to be 30.7% by trades and 27.5% by volume. Their corresponding figures for NASDAQ (in NASDAQ-listed stocks) are 40.9% and 42.7%. It would be reasonable to conjecture that the listing exchanges' information production is commensurately low. Ozturk et al report average NYSE information shares of 49.7% (allowing for intraday variation) and 61.9% (with no intraday variation). The corresponding figures for NASDAQ are 39.7% and 35.3%. Thus, the NYSE's information share is about twice as large as its trading market share, and NASDAQ's information share is comparable to its trading share.

Table 4 Panel A summarizes the trading activity by market center for IBM (NYSE-listed) and NVDA (NASDAQ-listed). NYSE activity in IBM comprises 18.0% (by trade count) and 20.8% (by

[10] Recent works include Hatheway, Kwan and Zheng (2017); Kwan, Masulis and McInish (2015); O'Hara and Ye (2011).

trade value). NASDAQ's corresponding figures for NVDA are 27.4% and 22.0%. These market shares are well below the Ozturk et al sample averages noted above. For both stocks, the largest reporting "market" is the FINRA ADF, which will be discussed in the next section.

For both stocks, using participant timestamps, I extract the bid and offer from the listing exchange, denoted $BidLex$ and $AskLex$. I also synthesize a best bid and offer from the remaining exchanges, denoted $BidOther$ and $AskOther$. I then estimate a four-variable VECM consisting of $\{BidLex, AskLex, BidOther, AskOther\}$ and compute information share bounds, at all resolution levels. For clarity in presentation, the information shares are grouped as listing exchange and other.

Table 6 summarizes the analyses. Across resolutions (Panel A), the random-walk volatilities are fairly constant. They are also very close to the corresponding estimates in Table 5. This is not surprising: all variables are prices; all systems are estimated over the same sample period. The modeled prices differ in the particulars of their construction, but they are all cointegrated, and therefore exhibit similar long-run behavior.

As in the previous analysis, the information share bounds are very wide at a one-second resolution, and tighten considerably in the high-resolution analyses. The striking finding here is that even though the primary listing exchanges' market shares are around 20%, the information shares are still over 50%. This may be related to the presence at the primary exchanges of designated market makers (DMMs). Ye, Clark-Joseph and Zi (2017) find that liquidity deteriorates significantly when trading is interrupted on primary listing exchanges, and attribute this to the absence of DMMs.

The IS bounds are noticeably wider, though, in the event time analysis. The precise reasons for this are unclear, but the result underscores the fundamental differences between natural- and event-time frameworks. The event-time framework generally compresses the distance (on the time scale) between events, and discards natural time intervals over which no price change occurs. Events widely separated in natural time may be contiguous in event time. In natural time, a zero price change will generally have a non-zero residual associated with it, which will make its due contribution to the innovation variances. In short, the estimated innovations are very different in the two frameworks. These differences may extend to the off-diagonal elements of the innovation

covariance matrix (which determine the upper and lower bounds). Panel B reports means and standard errors over the full 30-day sample, and these are in line with the one-day results.

VII. Quotes and trades (lit and dark)

Determining the relative information contributions of trades and quotes is a third question of ongoing importance. The classic asymmetric information models take the view that all informational advantage is held by some liquidity demanders, and that the information is partially revealed by their trades (Glosten and Milgrom (1985); Kyle (1985)). This perspective suggests a clean dichotomy: public information is reflected in the quotes, and private information, in the trades. This was always regarded as an oversimplification. Experimental evidence suggests that informed traders will post limit orders (Bloomfield, O'Hara and Saar (2005)). Furthermore, the traditional liquidity suppliers (dealers and specialists and so forth) have been partially displaced by high-frequency traders. The latter are commonly viewed as possessing better information on short-term information asymmetries. If they are acting as liquidity suppliers, they can avoid being picked off, and can update their bids and offers more promptly. This should enhance liquidity and the information content of their quotes. They can quickly become liquidity demanders, however, picking off limit orders posted by others. In this capacity they are essentially active informed agents and their information enters the market through their trades. Brogaard, Hendershott and Riordan (2015) find that information shares of quotes have generally increased, and that the bids and offers of high frequency traders (HFTs) are more informative than those of non-HFTs.

Related questions concern dark trades, that is, trades occurring at prices where the executing market has not posted a visible bid or offer. Hendershott and Jones (2005) examine a natural experiment in which Island, normally a “lit” market, was prevented from disseminating its quotes. Although its market share of trading activity fell substantially, the information share of its trades remained high. In other situations, dark markets are often thought to favor uninformed order flow (such as retail traders or passive institutions). This leaves the “lit” exchanges more exposed to informed traders, weakening their incentives to post visible bids and offers. Using a VAR specification estimated at one-second intervals, Comerton-Forde and Putniņš (2015) find evidence consistent with this effect.

Most (but not all) dark trades are reported to FINRA's ADF (the alternate display facility). The breakdown of trading activity in Table 4, Panel A shows that the ADF is largest reporting channel. To investigate the information content of quotes, lit, and dark trades. I estimate four-price VECMs that include the NBB, NBO, last sale price on a lit trade ($tradeLit$), and the last sale price on dark (ADF) trade ($tradeDark$). Use of last sale prices removes the need to "sign" the trade as buyer- or seller- initiated.

Table 7 reports random-walk volatilities and information share bounds. As noted in the previous analyses, the random-walk volatilities remain stable across resolutions and across analyses. The information share bounds are constructed with the NBB and NBO combined as one group. As in the other analyses, the bounds corresponding to the one-second resolution are wide. Between trades and quotes at one second, we can't tell which is informationally larger. The bounds also admit an informational contribution from dark trades. The findings sharpen considerably at higher resolutions of natural time. Quotes clearly dominate trades, and the contributions from dark trades essentially vanish. These findings hold for both high-resolution natural-time and event-time analyses.

Other results, however, differ between natural and event time. In the event-time analyses, the information shares for quotes are somewhat lower and the min/max bounds are wider. The means and standard errors in the thirty-day sample (Panel B) suggests that the natural/event time differences would be statistically significant in most cases.

Note that even at ten microsecond resolution, the information shares of quotes and trades are not completely resolved. In the timestamp and exchange analyses, the upper and lower bounds collapsed as the resolution increased, suggesting definitive attributions. In the present analysis, though, the gap between the bounds persists (around 6% for IBM, 4% for NVDA). This can be partially attributed to a mechanical effect. The execution of a buy order that takes the entire quantity available at the market's offer price, for example, obviously causes a contemporaneous withdrawal of that offer and, if additional quantities are posted deeper in the book, a revised offer.

Appealing to this mechanism, early microstructure analyses often assigned contemporaneous precedence to the trade. In this view, following the logic of the sequential trade models, bids and offers were determined in an equilibrium involving dealers or limit order traders.

Quotes could be updated following new public information, but if a marketable order arrived, it was assumed to be the proximate cause of whatever quote revision immediately followed. This convention could be imposed econometrically by including contemporaneous trades as explanatory variables in quote-revision equations, or (in price discovery analyses) restricting Cholesky factorizations to those orderings that assigned precedence to trades. Current trading practices and order types, however, don't offer such clear guidance. Both hidden and discretionary orders, for example, can easily lead to interactions in which a limit order submitted with passive intent turns out to be marketable (executable) on arrival. Causal attributions in these cases might potentially be resolved given the original orders, but not from the bid, ask and trade outcomes.[11]

VIII. Bridging approximations

A. Principles

The high-resolution VECM models implemented to this point achieve parsimony through coefficient step functions at a range of timescales. All computation of estimates, impulse response functions, and variance decompositions nevertheless occurs at the highest resolution. The computational effort here is substantial: at $10\mu s$ resolution, 500-second forecasts involves fifty million forward iterations of multivariate models with long coefficient lags. As the forecasts of levels (prices) involve accumulations of small differences, numerical accuracy may also be a concern. Bridging approximations combine analyses at different timescales, using high resolution models to characterize short timescales, and lower resolution models (with fewer iterations) for long term forecasts.

Bridging is commonly used to forecast macroeconomic variables in mixed frequency situations. For example, when one series is observed monthly and another, quarterly, monthly

[11] In the case of hidden orders, suppose that the (visible) bid and offer are $10.00 and $10.10, and there's a hidden sell order for 100 shares at 10.01. A trader intending to simply improve the bid submits an order to buy 300 shares limit 10.02. There is an execution of 100 shares at 10.01 (against the hidden order), and the remaining 200 shares become the new bid at 10.02. That is, the incoming, presumptively passive, limit order both causes the execution and sets an improved bid. A discretionary order is a passive limit order that is automatically canceled and replaced with a marketable order when the opposing quote moves within the discretionary range. (The BATS webpage contains links to some animated examples.) A limit buy order intended to improve the bid, might be executed on arrival if the limit price lies within the range of a discretionary sell order.

predictions of the first series are aggregated to form quarterly forecasts, which are then used as inputs to the forecasting of the quarterly series (Bańbura, Giannone, Modugno and Reichlin (2013)). Because the high- and low-frequency forecast models are distinct (as opposed to being different representations of a single unified model), Bańbura et al describe the bridging approach as "partial". The present applications would not generally be considered mixed-frequency, since all variables are observed at the highest frequency. The motivation here is computational efficiency.

As an overview, the bridging scheme starts with a low-order high-resolution VECM. In response to a high-frequency shock, the forecasts from this VECM are then time-aggregated and used as starting values for forecasting a coarser VECM. These forecasts are also time-aggregated, passed on as starting values to the next-coarser VECM, and so forth, stopping with long-run forecasts at the lowest resolution. These forecasts are approximations to those based on a correctly-specified high-resolution model. Because the shocks ultimately originate at the highest frequency, however, the high-frequency resolution is preserved. The algorithm suggests substantial computational gains in forecasting. Moreover, since the component VECM models might have relatively short lag structures, there may be computational gains in estimation as well.

More specifically, consider construction of a long-horizon high-resolution VMA from a sequence of VECM specifications of the form (1) where the lag length is $K = 10$ at all resolutions. Starting at the highest resolution, say $d = 10\mu s$, we estimate the first 10-lag VECM. Given an initial shock e_0, the cumulative forecast ten periods ahead is $\psi_{10\mu s,10}e_0$. We then estimate a 10-lag VECM at the next coarser timescale, $d = 100\mu s$. The initial conditions for forecasting at this coarser scale are given by two points: the value at time zero and the forecast from the finer scale at $100\mu s$, that is, e_0 and $\psi_{10\mu s,10}e_0$. Denote the coarser ten-period price forecast by $\psi_{100\mu s,10}e_0$. Then e_0 and $\psi_{100\mu s,10}e_0$ establish the initial conditions for forecasting at resolution $d = 1ms$, and so on. By construction, the long-term bridged forecasts are linear in the initial shock. This supports variance decompositions of long-term behavior based on high-resolution disturbance covariances.

If forecasting is the sole objective, the situation involves the familiar trade-off between computational effort and forecast error. If the aim is characterization of price discovery, though, derived measures (like information shares) are imbued with economic content. In this case, the nature and consequences of misspecification warrant deeper examination.

By way of illustration, I consider a special case of model (1) with $n = 2$ prices and $K = 100$ lags. For expositional convenience, the time units are nominally "seconds". The autoregressive coefficient matrices are diagonal, and they are scaled as suggested in the discussion of equation (7): $\phi_1 = 0.1I$; $\phi_k = 0.2I$ for $k = 2, \ldots, 10$; $\phi_k = 0.005I$ for $k = 11, \ldots, 100$. The ranges correspond to two timescales: 10 seconds ("short") and 100 seconds ("long"). The AR coefficients sum to 0.73 (0.28 from the first ten lags; 0.45 from the last ninety lags). The disturbances are i.i.d. $N(0,1)$. (This analysis does not feature sparse jumps; the focus is on bridging.) The adjustment coefficients are $\gamma = [-0.01 \quad 0.02]$. The half-life of the motion in p_1 toward p_2 is about 69 seconds, and that of p_2 toward $p_1$is about 34 seconds, so the adjustments occur over both long and short timescales. Although the autoregressive coefficient structures are identical for both prices, the difference in the adjustment coefficients leads to different information shares. Since the adjustment of p_2 toward p_1 is stronger than the reverse, p_1 will be informationally dominant.

I simulate a series of one million observations, and estimate three models:

- The correctly-specified model, which corresponds exactly to the data generating process, with one-second observations and 100 lags ("short and long")
- A truncated one-second model, with only 10 lags ("short")
- A ten-second time-aggregated model with 10 lags: 10-second price changes are computed (sampling at times $t = 10,20, \ldots.$) and the VECM has ten lags ("long").

The one-second/ten-lag model is clearly misspecified. The 10-second/10-lag model does not correspond to the DGP, but it does correspond to a time-aggregated skip-sampled VECM, and so, in a sense, could still considered correct. The estimates from the long and short models are used to construct a bridged model.

The differences in the models are most clearly illustrated by their impulse response functions. Figure 4 depicts for each model the IRF in p_1 following a one-unit shock to p_1. In the IRF corresponding to the correctly-specified model ("short and long"), the positive autoregressive coefficients impart an upward momentum to the price. The kinks at 10 and 100 seconds arise from the coefficient changes at those lags. As usual in these graphs, the scalloped appearance is an artifact of the logarithmic timescale.

The IRF corresponding to the short model is initially elevated (relative to the correct model) because the estimated short coefficients are picking up variation due to the omitted long terms. The omission of the long terms also leads to reversion in the IRF. The IRF in the long model more closely resembles the correct model, but since the short terms are omitted, the long-horizon price forecast lies below that of the correct model. The bridged IRF is essentially a splicing of the short and long IRF's. Comparing the short, long, and bridged IRFs, the bridged IRF forecast most closely resembles that from the correct model.

Table 8 summarizes price discovery analyses based on the four IRFs. The estimates corresponding to the correct model imply a random-walk volatility of $\sigma_w = 2.745$ ("dollars per second"). The information share bounds are tight. To three decimal places, the min and max of the p_1 information share are identically 0.798; those of the p_2 information share are 0.202. The short model substantially underestimates the random-walk volatility (1.116), which is consistent with behavior of its IRF in Figure 4. The information share bounds are not as tight as the correct model, and they are biased in favor of p_1. The long model estimate of the random-walk variance is close to the correct model. (This consistency is not surprising: random-walk variance is a long-run property of the price series.) The information share bounds, however, are much wider (at 0.121): the time-aggregation in the ten-second innovations induces correlation. The bridged model achieves the best overall approximation to the correct model, with a modest upward bias on the random-walk variance, and tight bounds on the information shares.

Bridging specifications in macro mixed-frequency applications are mostly driven by the observation frequencies of the series being bridged. In the present applications, though, where all data are available at the highest resolution, and bridging is motivated by computational expedience, the bridging scheme can be much more flexible. The short model in the example was used to forecast one step ahead (at the coarser timescale), but forecasts two or more steps ahead might be better starting points for the next timescale.[12]

[12] Alternatively, forecasts and impulse response functions might be estimated by direct estimation of multistep forecasts. In the present analyses, for an IRF k steps ahead this would entail the projection of p_{t+k} on p_{t-1}, p_{t-2}, …, and the error correction term Bp_{t-1}. This approach is developed for VARs by Jordà (2005), and extended to VECMs by Chong, Jordà and Taylor (2012). Each forecast horizon, though, requires reestimation of the full system. Additionally Marcellino, Stock and Watson

B. Applications

This section applies bridging techniques to the three analyses described in Sections V, VI, and VII. The key questions are whether the bridged estimates are close to the regular (fully iterated) estimates presented earlier, and whether they achieve significant computational savings.

The details of the bridging procedure are as follows. At each resolution, specification (7) is estimated with ϕ_1 unrestricted, $\phi_2 = \cdots = \phi_{10}$; $\phi_{11} = \cdots = \phi_{100}$; $\phi_{101} = \cdots = \phi_{1000}$. Then, using the highest-resolution VECM ($10\mu s$), and given an initial disturbance e_0, I forecast $p_1, \ldots, p_{100}$. The skip-sample from this set, $p_{10}, p_{20}, \ldots, p_{100}$ provides ten starting values for the forecasts based on the $100\mu s$ VEC. I forecast 100 steps ahead (at the $100\mu s$ resolution), and skip-sample to obtain starting values for the $1ms$-resolution VECM, and so forth. I stop with the 0.1 second VECM and forecast out 5,000 periods (500 seconds). This is a somewhat richer specification than that used for the simulated model in the preceding subsection. The logic of the procedure suggests that high-resolution specifications and forecasts are likely to perform better than the low-resolution equivalents. The present specification uses lags that are longer than strictly necessary, and the forecasts are constructed one hundred steps ahead (rather than ten).

Figure 5 depicts the components of a bridged impulse response function. The situation is that considered in Section V, a four-price model consisting of bids and offers formed from participant (exchange) and SIP (reporting) timestamps. The IRF depicts the bridged response in the SIP best bid, subsequent to a one-unit shock in the participant best bid, for IBM. In principle, the IRF corresponds to the $10\mu s$ line in Figure 1, Panel A. In Figure 5, the components of the bridged IRF are displaced slightly in the vertical dimension to clarify the intervals of overlap. The general features of the bridged and regular IRFs are very similar. At longer horizons, however, the bridged IRF is slightly lower than detailed. (At 500 seconds, the bridged IRF is 0.641 and the detailed IRF is 0.647.)

Table 9 presents estimates of bridged random-walk volatilities and information shares, for both stocks and each of the three analyses considered in Sections V, VI, and VII. Table 9

(2006) note that in macroeconomic applications, long-term forecasts formed by iterating forecasts based on short-term models generally outperform direct long-term forecasts.

contains only the highest resolution estimates, corresponding to the $10\mu s$ estimates in Tables 6, 7, and 8. The two sets of estimates are not identical, but are very similar.

I now consider computational aspects. As a rough indication, both detail and bridged analyses were implemented in Matlab using identical levels of parallel processing. (The construction of cross-product matrices and IRFs were spread over four processors.) The results for the IBM participant/SIP timestamps are indicative. The detail analysis at $10\mu s$ resolution required approximately 40 minutes to estimate and 133 minutes to construct the 500-second IRFs. The bridged analysis required approximated 20 minutes to estimate the (shorter) VECM specifications at resolutions of $100ms$, $10ms$, $1ms$, $100\mu s$, and $10\mu s$, and under 10 seconds to build the bridged IRFs. In all, the computational time for the bridged analysis is around 12% of the time required for the detail analysis. It is likely that gains would be larger at higher resolutions.

In summary, bridging appears to yield close approximations to forecasts constructed directly, at substantial computational savings.

IX. Conclusion

Although modern market data are commonly described as high-frequency, they are for many purposes better characterized as high-resolution. That is, the precision and accuracy of their time stamps allow them to be ordered at microsecond and nanosecond timescales. Determining the joint dynamics over extremely short intervals is key to resolving the strategies of agents, such as high frequency traders, who can react at these horizons. Another aspect of these strategies, though, is the provision of liquidity to agents operating at timescales considerably longer. A human day-trader might take a few seconds to react; an institutional trader might implement a purchase or sale over a day or longer. Estimating a model in natural time that captures short- and long-term components of these agents' actions, however, poses formidable challenges.

The analysis is based on standard linear VAR/VECM specifications. These are viewed as forecasting models. Due discreteness and other features, they are unlikely to be very representative of the data generating process. The analysis relies on two sources of simplification and tractability. The first involves coefficient constraints that are tied to timescale, following Corsi (2009). The specification allows high resolution at short lags, and lower resolution at longer lags. The second

simplification follows from the sparsity of the data, which facilitates the accumulation of least-squares cross-product matrices using a straightforward approach that involves only the non-zero price-change observations. Supplemental results suggest that bridging approaches will yield further computational efficiencies.

For two stocks (IBM and NVDA), the paper estimates three models of multiple cointegrated prices directed at representative problems in the microstructure literature. Each model is estimated over progressively finer (higher) levels of resolution, ranging from one second down to ten microseconds. Across resolutions, the model estimates behave sensibly. The random-walk volatilities, presumably long-term properties of the securities, are essentially unchanged in the passage to higher resolutions. Short-term effects, though, are much more clearly distinguished. This is particularly apparent in the analysis of information shares. Estimates of these parameters are often determined only within lower and upper bounds, and in practice the bounds are frequently wide. This is certainly the case for the systems estimated here. At one-second resolution, the bounds lie at the extremes of possibility (zero and one-hundred percent). As resolution increases, however, the bounds converge dramatically, in most cases. The systems for which the bounds don't collapse are those involving trades and quote revisions that appear to be contemporaneous even at resolutions of ten microseconds. Identification of causality in these situations is likely to be more reliant on economic analysis than on time stamps.

The main alternative to high-resolution natural time specification is event-time modeling, wherein the time index is a sequential counter of events. Event-time models in the present situations offer mixed results. The broad conclusions regarding information shares are unchanged: quotes using participant timestamps are much more informative that those based on SIP timestamps; the information shares of listing exchanges are larger than their volume shares; and, quotes dominate trades. In view of these similarities and their relative computational simplicity, event-time models should probably be preferred for exploratory analysis. They are not, however, equivalent to the high-resolution natural-time models. Estimates of information shares differ to an extent that may be important in many analyses. Additionally, the min/max bounds of information shares are generally wider in the event-time specifications. The correspondence between natural- and event-time specifications is worthy of further investigation.

The time stamps on modern market data can potentially identify strategies operating below the threshold of cross-market reaction times. For example, competing exchanges, market makers and algorithmic traders have long used autoquote algorithms or pegged orders to ensure that their bids and offers maintain some desired offset relative to other bids and offers. A buyer on some other market might peg her limit price to the bid on the primary listing exchange minus, say, $0.02. Near-simultaneous changes in quotes on different venues are often attributed to these strategies. They are, however, contingent on one player's observation of another's move. If the reaction occurs within an interval shorter than the physical limits of transmission, these strategies can be ruled out.

A multi-market strategy pursued by a single agent does not need to be reactive. The simplest examples involve selective delays in order origination. If Exchange A is 100 microseconds distant from Exchange B, a trader on Exchange B who wants to trade "simultaneously" at both exchanges can send an order to A, wait 100 microseconds, and then submit to B. This technique, applicable to marketable and nonmarketable orders, was implemented in IEX's THOR system, described in the SEC's order approving IEX's exchange registration (U.S. Securities and Exchange Commission (2016)) and *Flash Boys* (Lewis (2014)). The models proposed here can potentially distinguish single-agent delay-based behaviors from cross-market reactive strategies.

References

Andersen, Torben G, and Tim Bollerslev, 1998, Answering the skeptics: Yes, standard volatility models do provide accurate forecasts, *International economic review* 885-905.

Andersen, Torben G., Tim Bollerslev, and Francis X. Diebold, 2007, Roughing It Up: Including Jump Components in the Measurement, Modeling, and Forecasting of Return Volatility, *The Review of Economics and Statistics* 89, 701-720.

Baillie, Richard T., G. Geoffrey Booth, Yiuman Tse, and Tatyana Zabotina, 2002, Price discovery and common factor models, *Journal of Financial Markets* 5, 309-322.

Bańbura, Marta, Domenico Giannone, Michele Modugno, and Lucrezia Reichlin, 2013, Chapter 4 - Now-Casting and the Real-Time Data Flow, in Elliott Graham, and Timmermann Allan, eds.: *Handbook of Economic Forecasting* (Elsevier).

Bartlett, Maurice S., 1950, Periodogram analysis and continuous spectra, *Biometrika* 37, 1-16.

Bartlett, Robert P., III, and Justin McCrary, 2016, How rigged are stock markets: evidence from microsecond timestamps (University of California, Berkeley).

Beveridge, Stephen, and Charles R. Nelson, 1981, A new approach to decomposition of economic time series into permanent and transitory components with particular attention to measurement of the 'business cycle', *Journal of Monetary Economics* 7, 151-174.

Bloomfield, Robert, Maureen O'Hara, and Gideon Saar, 2005, The "make or take" decision in an electronic market: evidence on the evolution of liquidity, *Journal of Financial Economics* 75, 165-199.

Bollerslev, Tim, Andrew J. Patton, and Roger Quaedvlieg, 2016, Modeling and forecasting (un)reliable realized covariances for more reliable financial decisions.

Brogaard, Jonathan, Terrence Hendershott, and Ryan Riordan, 2015, Price discovery without trading: evidence from limit orders (SSRN).

Chinco, Alex, and Mao Ye, 2016, Investment-horizon spillovers: evidence from decomposing trading volume variance https://ssrn.com/abstract=2544738.

Chong, Yanping, Òscar Jordà, and Alan M Taylor, 2012, The Harrod–Balassa–Samuelson Hypothesis: Real Exchange Rates And Their Long-Run Equilibrium, *International Economic Review* 53, 609-634.

Comerton-Forde, Carole, and Tālis J. Putniņš, 2015, Dark trading and price discovery, *Journal of Financial Economics* 118, 70-92.

Corsi, Fulvio, 2009, A Simple Approximate Long-Memory Model of Realized Volatility, *Journal of Financial Econometrics* 7, 174-196.

Crouzet, Nicolas, Ian Dew-Becker, and Charles Nathanson, 2016, A Model of Multi-Frequency Trade (Northwestern University Working Paper).

de Jong, Frank, 2002, Measures of contributions to price discovery: a comparison, *Journal of Financial Markets* 5, 323-327.

De Jong, Frank, and Peter C. Schotman, 2010, Price Discovery in Fragmented Markets, *Journal of Financial Econometrics* 8, 1-28.

Ding, Shengwei, John Hanna, and Terrence Hendershott, 2014, How slow is the NBBO? A comparison with direct data feeds, *The Financial Review* 49, 313-332.

Dufour, Alfonso, and Robert F. Engle, 2000, Time and the Price Impact of a Trade, *The Journal of Finance* 55, 2467-2498.

Easley, David, and Maureen O'Hara, 1992, Time and the process of security price adjustment, *Journal of Finance* 47, 576-605.

Engle, Robert F., and C. W. J. Granger, 1987, Co-Integration and Error Correction: Representation, Estimation, and Testing, *Econometrica* 55, 251-276.

Forsberg, Lars, and Eric Ghysels, 2007, Why do absolute returns predict volatility so well?, *Journal of Financial Econometrics* 5, 31-67.

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).

Ghysels, Eric, Arthur Sinko, and Rossen Valkanov, 2007, MIDAS Regressions: Further Results and New Directions, *Econometric Reviews* 26, 53-90.

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

Grammig, J., and F. J. Peter, 2013, Telltale Tails: A New Approach to Estimating Unique Market Information Shares, *Journal of Financial and Quantitative Analysis* 48, 459-488.

Grünbichler, Andreas, Alexander Kohler, and Rico von Wyss, 2017, Equity Market Fragmentation in the Swiss Market, in Reto Francioni, and Robert A. Schwartz, eds.: *Equity Markets in Transition: The Value Chain, Price Discovery, Regulation, and Beyond* (Springer International Publishing, Cham).

Hagströmer, Björn, and Albert J. Menkveld, 2017, A network map of information percolation (SSRN).

Hamilton, James D., 1994. *Time Series Analysis* (Princeton University Press, Princeton).

Harris, Frederick H. deB, Thomas H. McInish, Gary L. Shoesmith, and Robert A. Wood, 1995, Cointegration, error correction, and price discovery on informationally linked security markets, *Journal of Financial and Quantitative Analysis* 30, 563-579.

Harris, Frederick H. deB, Thomas H. McInish, and Robert A. Wood, 2002, Common factor components vs. information shares: A reply, *Journal of Financial Markets* 5, 341-348.

Harris, Frederick H. deB, Thomas H. McInish, and Robert A. Wood, 2002, Security price adjustment across exchanges: an investigation of common factor components for Dow stocks, *Journal of Financial Markets* 5, 277-308.

Hasbrouck, Joel, 1995, One security, many markets: Determining the contributions to price discovery, *Journal of Finance* 50, 1175-99.

Hasbrouck, Joel, 1999, Trading fast and slow: security market events in real time (NYU Working Paper), https://ssrn.com/abstract=1296401.

Hasbrouck, Joel, 2002, Stalking the "efficient price" in market microstructure specifications: an overview, *Journal of Financial Markets* 5, 329-339.

Hasbrouck, Joel, 2003, Intraday price formation in US equity index markets, *Journal of Finance* 58, 2375-2399.

Hasbrouck, Joel, 2018, High frequency quoting: short-term volatility in bids and offers, *Journal of Financial and Quantitative Analysis (forthcoming)*.

Hatheway, Frank, Amy Kwan, and Hui Zheng, 2017, An Empirical Analysis of Market Segmentation on US Equity Markets, *Journal of Financial and Quantitative Analysis* 52, 2399-2427.

Hendershott, Terrence, and Charles M. Jones, 2005, Island goes dark: transparency, fragmentation and regulation, *Review of Financial Studies* 18, 743-793.

Jordà, Òscar, 2005, Estimation and Inference of Impulse Responses by Local Projections, *American Economic Review* 95, 161-182.

Kumar, Praveen, and Duane J. Seppi, 1994, Information and Index Arbitrage, *The Journal of Business* 67, 481-509.

Kwan, Amy, Ronald Masulis, and Thomas H. McInish, 2015, Trading rules, competition for order flow and market fragmentation, *Journal of Financial Economics* 115, 330-348.

Kyle, Albert S., 1985, Continuous auctions and insider trading, *Econometrica* 53, 1315-1336.

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

Lehmann, Bruce N., 2002, Some desiderata for the measurement of price discovery across markets, *Journal of Financial Markets* 5, 259-276.

Lewis, Michael, 2014. *Flash Boys* (W. W. Norton & Company).

Lütkepohl, Helmut, 2005. *New introduction to multiple time series analysis* (Springer Science & Business Media).

Marcellino, Massimiliano, James H Stock, and Mark W Watson, 2006, A comparison of direct and iterated multistep AR methods for forecasting macroeconomic time series, *Journal of econometrics* 135, 499-526.

Müller, Ulrich A, Michel M Dacorogna, Rakhal D Davé, Olivier V Pictet, Richard B Olsen, and J Robert Ward, 1993, Fractals and intrinsic time: A challenge to econometricians, *Unpublished manuscript, Olsen & Associates, Zürich*.

NYSE, 2017, Daily TAQ Client Specification v. 3.0a https://www.nyse.com/publicdocs/nyse/data/Daily_TAQ_Client_Spec_v3.0a.pdf.

O'Hara, Maureen, and Mao Ye, 2011, Is market fragmentation harming market quality?, *Journal of Financial Economics* 100, 459-474.

Ozturk, Sait R., Michel van der Wel, and Dick van Dijk, 2017, Intraday price discovery in fragmented markets, *Journal of Financial Markets* 32, 28-48.

Pascual, Roberto, and Bartolomé Pascual-Fuster, 2014, The relative contribution of ask and bid quotes to price discovery, *Journal of Financial Markets* 20, 129-150.

Putniņš, Tālis J, 2013, What do price discovery metrics really measure?, *Journal of Empirical Finance* 23, 68-83.

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

Stock, James H., and Mark W. Watson, 1988, Variable trends in economic time series, *Journal of Economic Perspectives* 2, 147-174.

U.S. Securities and Exchange Commission, 2016, In the Matter of the Application of Investors' Exchange, LLC for Registration as a National Securities Exchange.

U.S. Securities and Exchange Commission, 2016, Joint Industry Plan; Notice of Filing of the National Market System Plan Governing the Consolidated Audit Trail https://www.sec.gov/rules/sro/nms/2016/34-77724.pdf.

Yan, Bingcheng, and Eric Zivot, 2010, A structural analysis of price discovery measures, *Journal of Financial Markets* 13, 1-19.

Ye, Mao, Adam Clark-Joseph, and Chao Zi, 2017, Designated market makers still matter: evidence from two natural experiments, *Journal of Financial Economics* forthcoming.

Table 1. Resolution and autoregressive coefficient structure

The autoregressive coefficients follow a step function that is constant within the given range. In the one-second analysis, the lagged intervals have endpoints located at 1, 2, …, 10 seconds. The coefficient at lag 1 varies without restriction; the coefficients at lags 2-10 have the same value. In the 0.1 second (100 ms) analysis, the lagged intervals have endpoints located at 0.1, 0.2, …, 10.0 seconds. The coefficient at lag 1 is unrestricted; coefficients at lags 2-10 have the same value; coefficients at lags 11-100 have the same value.

Resolution (seconds)	Coefficient ranges						
1.0						[1.0]	(1.0,10]
0.1					[0.1]	(0.1,1.0]	(1.0,10]
0.01				[0.01]	(0.01,0.1]	(0.1,1.0]	(1.0,10]
0.001			[0.001]	(0.001,0.01]	(0.01,0.1]	(0.1,1.0]	(1.0,10]
0.0001		[0.0001]	(0.0001,0.001]	(0.001,0.01]	(0.01,0.1]	(0.1,1.0]	(1.0,10]
0.00001	[0.00001]	(0.00001,0.0001]	(0.0001,0.001]	(0.001,0.01]	(0.01,0.1]	(0.1,1.0]	(1.0,10]

Table 2. Summary statistics for trades and quotes

The sample is all trade and quote records for IBM and NVIDIA on the daily TAQ file for October 3, 2016, between 9:45 and 16:00.

	Trades				Quotes
	N	Avg. price	Avg. size (shares)	Avg. size (value)	N
IBM	22,282	$157.60	82.0	$12,930.10	314,324
NVDA	41,724	$68.69	117.1	$8,045.87	893,413

Table 3. Reporting delays

The daily TAQ trade and quote records have a participant timestamp that is inserted by the participant's matching engine, and a SIP timestamp that is inserted when the event has been processed by the securities information processor, prior to its multicast. The difference between these timestamps is $\delta = SIP\ time - participant\ time$, in milliseconds. Panel A summarizes the distribution of δ by participant for trades; panel B, for quotes.

Panel A. Distribution of δ for transactions

				Quantiles					
Symbol	Exchange	N	Min	1%	10%	50%	90%	99%	Max
IBM	BX (NASDAQ)	2,250	0.82	0.84	0.88	0.96	1.88	4.01	7.14
	Bats BYX	1,041	0.45	0.48	0.51	0.56	0.72	2.42	3.95
	Bats BZX	2,062	0.45	0.48	0.52	0.57	0.71	1.92	4.19
	EDGA	412	0.48	0.49	0.53	0.58	0.74	2.17	2.52
	EDGX	1,848	0.46	0.49	0.52	0.57	0.74	3.56	6.47
	FINRA ADF	4,338	-31.95	1.81	2.51	6.28	22.34	946.18	5,625.47
	IEX	428	0.53	0.55	0.58	0.62	1.81	4.32	9.11
	NASDAQ	4,028	0.80	0.83	0.86	0.95	2.24	4.92	30.84
	NASDAQ PSX	82	0.85	0.85	0.87	0.95	2.20	3.74	3.74
	NYSE	4,004	0.22	0.24	0.27	0.31	0.48	1.73	4.80
	NYSE ARCA	1,789	0.16	0.18	0.20	0.24	0.35	1.35	4.04
	All	22,282	-31.95	0.20	0.28	0.85	6.18	127.02	5,625.47
NVDA	BX (NASDAQ)	1,801	0.19	0.23	0.27	0.35	1.37	3.84	7.89
	Bats BYX	2,362	0.37	0.42	0.46	0.54	0.85	2.68	181.63
	Bats BZX	4,972	0.36	0.41	0.46	0.53	0.84	2.86	6.13
	EDGA	1,703	0.40	0.44	0.48	0.56	1.09	2.83	6.38
	EDGX	4,731	0.40	0.43	0.47	0.55	0.84	3.37	12.00
	FINRA ADF	8,684	-16.63	-1.76	1.88	5.63	122.04	1,166.36	11,034.81
	IEX	413	0.50	0.50	0.54	0.61	2.05	3.92	4.19
	NASDAQ	11,415	0.16	0.21	0.25	0.33	1.05	4.20	8.28
	NASDAQ PSX	516	0.21	0.23	0.27	0.34	0.85	3.25	5.07
	NYSE ARCA	5,056	0.77	0.82	0.86	0.93	1.04	1.50	5.25
	NYSE MKT	71	0.89	0.89	0.93	1.01	1.20	3.60	3.60
	All	41,724	-16.63	0.22	0.29	0.58	5.80	216.68	11,034.81

Table 3. Reporting delays (continued)

Panel B. Distribution of δ for quote updates.

				Quantiles					
	Exchange	N	Min	1%	10%	50%	90%	99%	Max
IBM	BX (NASDAQ)	35,116	-0.75	0.80	0.83	0.88	2.46	9.27	45.29
	Bats BYX	37,692	-1.25	0.39	0.41	0.45	1.02	4.83	18.32
	Bats BZX	38,674	-1.18	0.39	0.41	0.44	0.55	6.44	67.84
	EDGA	13,284	-1.17	0.40	0.42	0.45	0.62	4.48	19.14
	EDGX	11,283	-1.56	0.40	0.42	0.45	0.78	6.82	34.46
	IEX	32,125	-1.66	0.51	0.54	0.60	2.34	6.68	42.43
	NASDAQ	28,383	-0.74	0.80	0.83	0.89	4.10	16.76	75.16
	NASDAQ PSX	13,523	-0.76	0.80	0.83	0.87	1.02	6.18	37.48
	NYSE	83,556	-2.10	0.21	0.23	0.29	2.25	8.14	36.33
	NYSE ARCA	20,688	-1.69	0.20	0.23	0.27	0.43	3.78	39.51
	All	314,324	-2.10	0.21	0.25	0.47	1.77	8.24	75.16
NVDA	BX (NASDAQ)	75,268	0.04	0.22	0.26	0.34	0.73	12.05	73.62
	Bats BYX	114,100	0.21	0.39	0.43	0.51	0.82	8.88	53.47
	Bats BZX	139,994	0.22	0.39	0.42	0.50	0.89	11.87	77.34
	EDGA	53,516	0.26	0.40	0.43	0.50	0.82	8.01	48.25
	EDGX	75,539	0.22	0.40	0.43	0.51	0.87	9.44	52.22
	IEX	6,696	0.46	0.48	0.52	0.61	1.55	14.09	54.92
	NASDAQ	190,813	0.07	0.25	0.29	0.39	1.03	7.92	37.43
	NASDAQ PSX	42,246	0.06	0.25	0.28	0.35	0.59	6.12	39.30
	NYSE ARCA	166,343	0.67	0.83	0.87	0.94	1.31	11.32	86.17
	NYSE MKT	28,898	0.82	0.91	0.95	1.00	1.10	2.84	33.35
	All	893,413	0.04	0.25	0.32	0.51	1.06	9.50	86.17

Table 4. Trades and quotes by market center

The sample is all trade and quote records for IBM and NVIDIA for October 3, 2016, between 9:45 and 16:00. Panel A reports summary statistics for trades, including the distribution across exchanges of trade counts and total value. "FINRA ADF" refers to FINRA's Alternative Display Facility, a reporting channel for trades that do not take place on exchanges.

Panel A. Transactions

	Exchange	N	Percent	Total Value	Percent
IBM	BX (NASDAQ)	2,250	10.1	16,237,124	5.6
	Bats BYX	1,041	4.7	11,195,595	3.9
	Bats BZX	2,062	9.3	18,703,805	6.5
	EDGA	412	1.8	4,956,413	1.7
	EDGX	1,848	8.3	24,483,852	8.5
	FINRA ADF	4,338	19.5	77,266,592	26.8
	IEX	428	1.9	7,177,992	2.5
	NASDAQ	4,028	18.1	45,466,059	15.8
	NASDAQ PSX	82	0.4	853,433	0.3
	NYSE	4,004	18.0	59,796,122	20.8
	NYSE ARCA	1,789	8.0	21,971,478	7.6
	All	22,282	100.0	288,108,465	100.0
NVDA	BX (NASDAQ)	1,801	4.3	11,411,852	3.4
	Bats BYX	2,362	5.7	13,691,110	4.1
	Bats BZX	4,972	11.9	29,725,917	8.9
	EDGA	1,703	4.1	11,224,610	3.3
	EDGX	4,731	11.3	33,763,188	10.1
	FINRA ADF	8,684	20.8	120,015,775	35.8
	IEX	413	1.0	2,958,065	0.9
	NASDAQ	11,415	27.4	73,846,144	22.0
	NASDAQ PSX	516	1.2	3,523,688	1.0
	NYSE ARCA	5,056	12.1	34,963,303	10.4
	NYSE MKT	71	0.2	582,066	0.2
	All	41,724	100.0	335,705,717	100.0

Panel B. Quote updates

	Exchange	N	Percent
IBM	BX (NASDAQ)	35,116	11.2
	Bats BYX	37,692	12.0
	Bats BZX	38,674	12.3
	EDGA	13,284	4.2
	EDGX	11,283	3.6
	IEX	32,125	10.2
	NASDAQ	28,383	9.0
	NASDAQ PSX	13,523	4.3
	NYSE	83,556	26.6
	NYSE ARCA	20,688	6.6
	All	314,324	100.0
NVDA	BX (NASDAQ)	75,268	8.4
	Bats BYX	114,100	12.8
	Bats BZX	139,994	15.7
	EDGA	53,516	6.0
	EDGX	75,539	8.5
	IEX	6,696	0.7
	NASDAQ	190,813	21.4
	NASDAQ PSX	42,246	4.7
	NYSE ARCA	166,343	18.6
	NYSE MKT	28,898	3.2
	All	893,413	100.0

Table 5. Information shares using participant and SIP timestamps.

Panel A reports point estimates based on quotes for IBM and NVIDIA, October 3, 2016, between 9:45 and 16:00. The specification is a four-variable VECM estimated at the indicated resolution. The variables are National Best Bids (NBB) and National Best Offers (NBO) constructed using participant exchanges and the securities information processors (SIPs). The resolution refers to the interval width. Prices are taken as of the end of the interval. Bid and offer prices are dollars per share. The random-walk volatility is scaled as dollars per year. Panel B contains summary statistics (means and standard errors) for daily estimates over a thirty-trading-day sample.

Panel A. October 3, 2016

			Information share			
	Resolution	Random-walk	$\{NBBpart, NBOpart\}$		$\{NBBsip, NBOsip\}$	
	(seconds)	volatility (σ_w)	Min	Max	Min	Max
IBM	1.0	19.065	0.002	1.000	0.000	0.998
	0.1	19.154	0.017	0.999	0.001	0.983
	0.01	19.138	0.142	0.998	0.002	0.858
	0.001	19.121	0.827	0.998	0.002	0.173
	0.0001	19.121	0.999	0.999	0.001	0.001
	0.00001	19.124	0.999	0.999	0.001	0.001
	Event time		0.987	1.000	0.000	0.013
NVDA	1.0	12.246	0.000	1.000	0.000	1.000
	0.1	12.135	0.009	1.000	0.000	0.991
	0.01	12.068	0.082	1.000	0.000	0.918
	0.001	12.057	0.685	1.000	0.000	0.315
	0.0001	12.049	0.997	0.999	0.001	0.003
	0.00001	12.046	0.999	0.999	0.001	0.001
	Event time		0.982	0.998	0.002	0.018

Panel B. Means and standard errors of daily estimates (October 3, 2016 to November 11, 2016)

			Information share			
	Resolution	Random-walk	$\{NBBpart, NBOpart\}$		$\{NBBsip, NBOsip\}$	
	(seconds)	volatility (σ_w)	Min	Max	Min	Max
IBM	0.00001	19.824 (1.031)	0.997 (0.001)	0.997 (0.001)	0.003 (0.001)	0.003 (0.001)
	Event time		0.988 (0.001)	1.000 (<0.001)	0.000 (<0.001)	0.012 (0.001)
NVDA	0.00001	15.461 (1.518)	0.997 (0.001)	0.997 (0.001)	0.003 (0.001)	0.003 (0.001)
	Event time		0.979 (0.001)	0.999 (<0.001)	0.001 (<0.001)	0.021 (0.001)

Table 6. Information shares: primary listing exchange vs. all others

Panel A reports point estimates based on all quotes for IBM and NVIDIA, October 3, 2016, between 9:45 and 16:00. The specification is a four-variable VECM estimated at the indicated resolution. The variables are the bid and offer on the primary listing exchange ($BidLEX$ and $AskLEX$) and the best bid and offer on the non-listing exchanges ($BidOther$ and $AskOther$), using participant timestamps. The resolution refers to the interval width. Prices are taken as of the end of the interval. Bid and offer prices are dollars per share. The random-walk volatility is scaled as dollars per year. Panel B contains summary statistics (means and standard errors) for daily estimates over a thirty-trading-day sample.

Panel A. October 3, 2016

			Information share			
	Resolution	Random-walk	$\{BidLEX, AskLEX\}$		$\{BidOther, AskOther\}$	
	(seconds)	volatility (σ_w)	Min	Max	Min	Max
IBM	1.0	19.187	0.136	0.932	0.068	0.864
	0.1	19.269	0.215	0.850	0.150	0.785
	0.01	19.243	0.295	0.792	0.208	0.705
	0.001	19.218	0.435	0.668	0.332	0.565
	0.0001	19.127	0.519	0.535	0.465	0.481
	0.00001	19.135	0.525	0.526	0.474	0.475
	Event time		0.458	0.555	0.445	0.542
NVDA	1.0	12.299	0.067	0.914	0.086	0.933
	0.1	12.147	0.131	0.820	0.180	0.869
	0.01	12.070	0.173	0.795	0.205	0.827
	0.001	12.061	0.281	0.717	0.283	0.719
	0.0001	11.998	0.493	0.560	0.440	0.507
	0.00001	11.939	0.533	0.539	0.461	0.467
	Event time		0.472	0.643	0.357	0.528

Panel B. Means and standard errors of daily estimates (October 3, 2016 to November 11, 2016)

			Information share			
	Resolution	Random-walk	$\{BidLEX, AskLEX\}$		$\{BidOther, AskOther\}$	
	(seconds)	volatility (σ_w)	Min	Max	Min	Max
IBM	0.00001	19.859	0.401	0.403	0.597	0.599
		(1.035)	(0.012)	(0.012)	(0.012)	(0.012)
	Event time		0.344	0.434	0.566	0.656
			(0.011)	(0.013)	(0.013)	(0.011)
NVDA	0.00001	15.301	0.535	0.542	0.458	0.465
		(1.509)	(0.010)	(0.010)	(0.010)	(0.010)
	Event time		0.495	0.628	0.372	0.505
			(0.012)	(0.015)	(0.015)	(0.012)

Table 7. Information shares: quotes, lit trades, and dark trades

Panel A reports point estimates based on all trades and quotes for IBM and NVIDIA, October 3, 2016, between 9:45 and 16:00. The specification is a four-variable VECM estimated at the indicated resolution. The variables are the national best bid and offer (NBB and NBO), the last recorded price of an execution reported by lit exchange ($Trades\ (lit)$), and the last recorded price of an execution reported on FINRA's ADF ($Trades\ (dark)$), constructed using participant timestamps. The resolution refers to the interval width. Prices are taken as of the end of the interval. Bid and offer prices are dollars per share. The random-walk volatility is scaled as dollars per year. Panel B contains summary statistics (means and standard errors) for daily estimates over a thirty-trading-day sample.

Panel A. October 3, 2016

			Information share					
	Resolution	Random-walk	$\{NBB, NBO\}$		$Trades\ (lit)$		$Trades\ (dark)$	
	(seconds)	volatility (σ_w)	Min	Max	Min	Max	Min	Max
IBM	1.0	18.645	0.390	0.979	0.021	0.603	0.000	0.028
	0.1	18.707	0.479	0.960	0.038	0.515	0.002	0.012
	0.01	18.706	0.525	0.916	0.081	0.471	0.002	0.006
	0.001	18.653	0.570	0.794	0.204	0.428	0.002	0.002
	0.0001	18.551	0.580	0.661	0.337	0.418	0.002	0.002
	0.00001	18.554	0.581	0.643	0.355	0.416	0.002	0.002
	Event time		0.533	0.606	0.393	0.465	0.001	0.002
NVDA	1.0	12.444	0.450	0.982	0.016	0.541	0.001	0.048
	0.1	12.339	0.580	0.962	0.036	0.416	0.001	0.012
	0.01	12.296	0.635	0.941	0.058	0.363	0.001	0.004
	0.001	12.278	0.690	0.881	0.117	0.308	0.001	0.002
	0.0001	12.214	0.707	0.763	0.235	0.291	0.002	0.002
	0.00001	12.199	0.708	0.744	0.254	0.289	0.002	0.002
	Event time		0.674	0.782	0.218	0.326	0.000	0.000

Panel B. Means and standard errors of daily estimates (October 3, 2016 to November 11, 2016)

			Information share					
	Resolution	Random-walk	$\{NBB, NBO\}$		$Trades\ (lit)$		$Trades\ (dark)$	
	(seconds)	volatility (σ_w)	Min	Max	Min	Max	Min	Max
IBM	0.00001	19.650	0.630	0.683	0.313	0.366	0.004	0.004
		(1.013)	(0.012)	(0.011)	(0.011)	(0.012)	(0.001)	(0.001)
	Event time		0.510	0.596	0.401	0.486	0.003	0.004
			(0.015)	(0.016)	(0.015)	(0.015)	(0.001)	(0.001)
NVDA	0.00001	15.307	0.691	0.726	0.271	0.305	0.003	0.003
		(1.504)	(0.009)	(0.008)	(0.008)	(0.009)	(0.001)	(0.001)
	Event time		0.602	0.716	0.279	0.395	0.003	0.005
			(0.009)	(0.009)	(0.009)	(0.009)	(0.001)	(0.001)

Table 8. Bridged information share estimates in a simulated model

The simulated model is:

$$\Delta p_t = \gamma B p_{t-1} + \phi_1 \Delta p_{t-1} + \phi_2 \Delta p_{t-2} + \cdots + \phi_K \Delta p_{t-K} + \epsilon_t,$$

where $p_t = [p_{1t} \quad p_{2t}]', B = [1 \quad -1], \gamma = [-0.01 \quad 0.02]$, and $\epsilon_t \sim N(0, I_2)$. The $K = 100$ autoregressive coefficient matrices are: $\phi_1 = 0.1I$; $\phi_k = 0.2I$ for $k = 2, \ldots, 10$; $\phi_k = 0.005I$ for $k = 11, \ldots, 100$. Based on one million simulated observations, three models are estimated: the short model is truncated at $K = 10$; the long model also has $K = 10$ lags, but it is applied to prices sampled every ten periods; the short and long model corresponds to the correct specification. For each of the three models, the vector moving average representation (VMA) is computed through 1,000 periods forward, and I compute the random-walk variance and bounds on the information shares. In the bridged specification, the VMA is constructed by taking the ten-step ahead forecasts from the short model as the starting values for forecasting the long model.

		Information shares			
		p_1		p_2	
Specification	Random-walk volatility (σ_w)	Min	Max	Min	Max
Short and long	2.745	0.798	0.798	0.202	0.202
Short	1.116	0.835	0.839	0.161	0.165
Long	2.713	0.719	0.840	0.160	0.281
Bridged	2.827	0.806	0.810	0.190	0.194

Table 9. Bridged estimates of random-walk volatilities and information shares

In each analysis, short-lag VECMs are estimated at resolutions $d \in \{10\mu s, 100\mu s, 1ms, 10ms, 100ms\}$. VMA representations are constructed from bridged impulse response functions (short-term high-resolution forecasts are propagated as starting values for longer-term, lower resolution forecasts). The final VMAs have a forecast horizon of 500 seconds and an implied resolution of $10\mu s$. In Panel A, $NBBpart$ and $NBOpart$ are the (national best) bid and offer formed using participant (exchange) time stamps; $NBBsip$ and $NBOsip$ use timestamps from the Securities Information Processor. In Panel B, $BidLEX$ and $AskLEX$ are the bid and ask from the primary listing exchange; $BidOther$ and $AskOther$ are the best bid and ask constructed from all other exchanges. In Panel C, NBB and NBO are the National Best Bid and Offer (constructed across all exchanges), $Trades\ (lit)$ is the last-sale price on any lit exchange, $Trades\ (dark)$ is the last sale price reported under exchange code D. The random-walk volatility is scaled as dollars per year. Bounds on information shares are computed over all ordering permutations of the innovations.

Panel A. Participant and SIP timestamps

		Information shares			
	Random-walk	$\{NBBpart, NBOpart\}$		$\{NBBsip, NBOsip\}$	
	volatility (σ_w)	Min	Max	Min	Max
IBM	19.68	0.997	0.998	0.002	0.003
NVDA	11.45	0.971	0.972	0.028	0.029

Panel B. Primary listing exchange vs. all others

		Information shares			
	Random-walk	$\{BidLEX, AskLEX\}$		$\{BidOther, AskOther\}$	
	volatility (σ_w)	Min	Max	Min	Max
IBM	19.12	0.521	0.523	0.477	0.478
NVDA	12.02	0.521	0.526	0.474	0.479

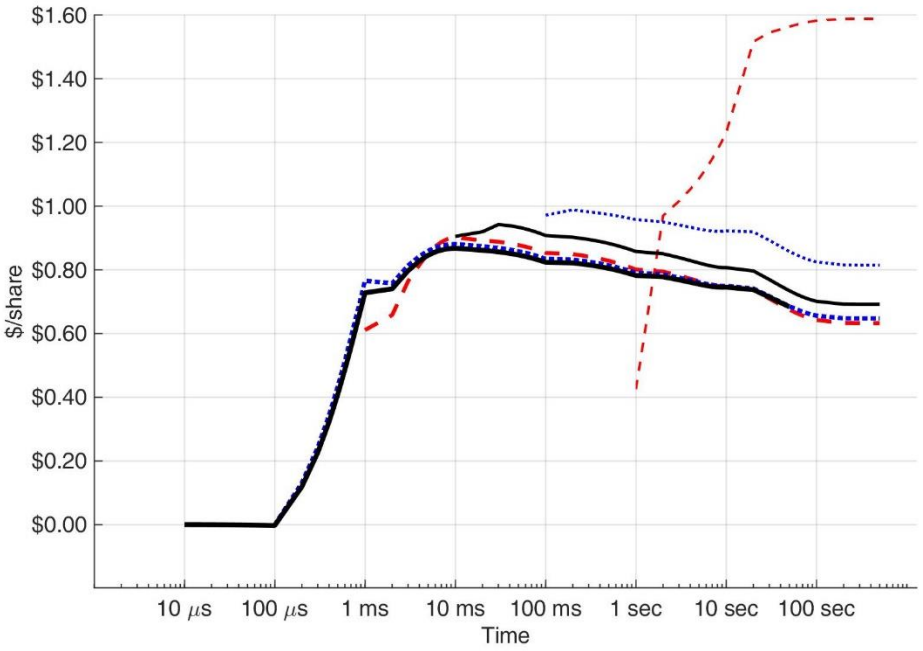
Panel C. Quotes, lit trades and dark trades

		Information shares					
	Random-walk	$\{NBB, NBO\}$		$Trades\ (lit)$		$Trades\ (dark)$	
	volatility (σ_w)	Min	Max	Min	Max	Min	Max
IBM	18.50	0.587	0.648	0.350	0.411	0.002	0.002
NVDA	12.27	0.726	0.760	0.234	0.272	0.002	0.002

Figure 1. Impulse response functions, participant and SIP timestamps

The VECM system comprises national best bids (NBBs) and offers (NBOs) constructed from participant and SIP timestamps. The system is estimated at six resolutions, ranging from one second down to ten microseconds. IRFs depict the cumulative response in the SIP-based NBB following a time-zero one dollar shock to the participant-based NBB. In the figure, the IRF estimated at a one-second resolution begins at one second, and similarly for the finer resolutions.

Panel A. IBM



Panel B. NVDA

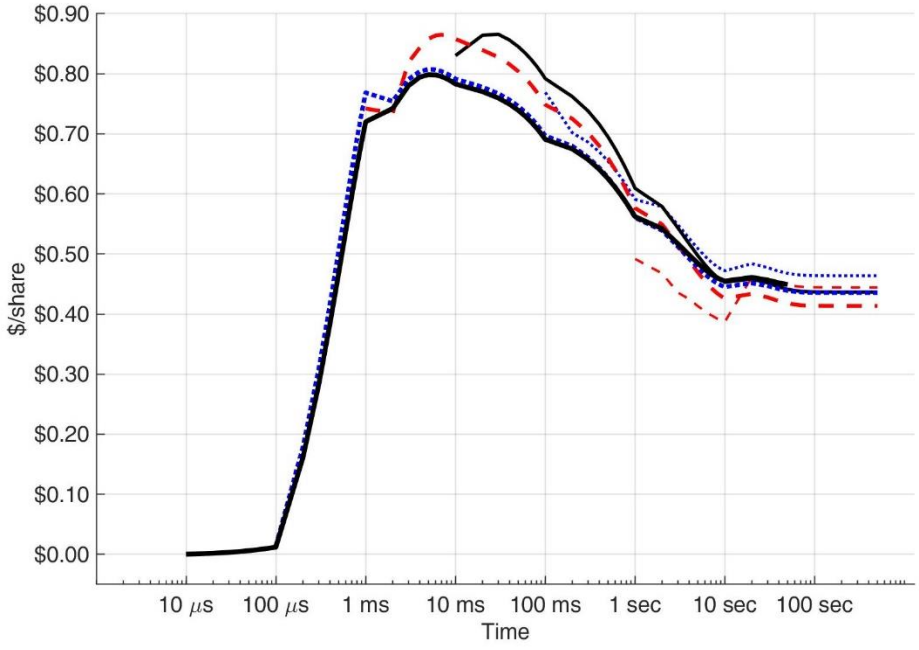
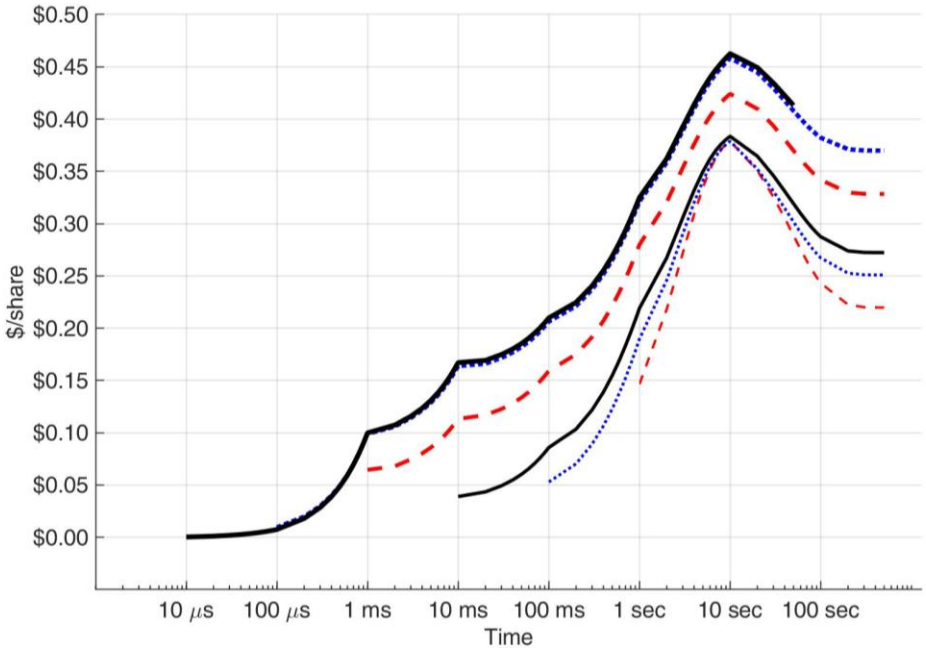


Figure 2. Impulse response functions, primary listing exchanges vs. others

The VECM system comprises four price variables: the best bid and offer disseminated by the primary listing exchange (NYSE for IBM, NASDAQ for NVDA), and the best bid and offer constructed over all other exchanges (that is, ex the primary listing exchange). The system is estimated at six resolutions, ranging from one second down to ten microseconds. IRFs depict the cumulative response in the best bid ex primary following a time-zero one dollar shock bid on the primary listing exchange. In the figure, the IRF estimated at a one-second resolution begins at one second, and similarly for the finer resolutions.

Panel A. IBM



Panel B. NVDA

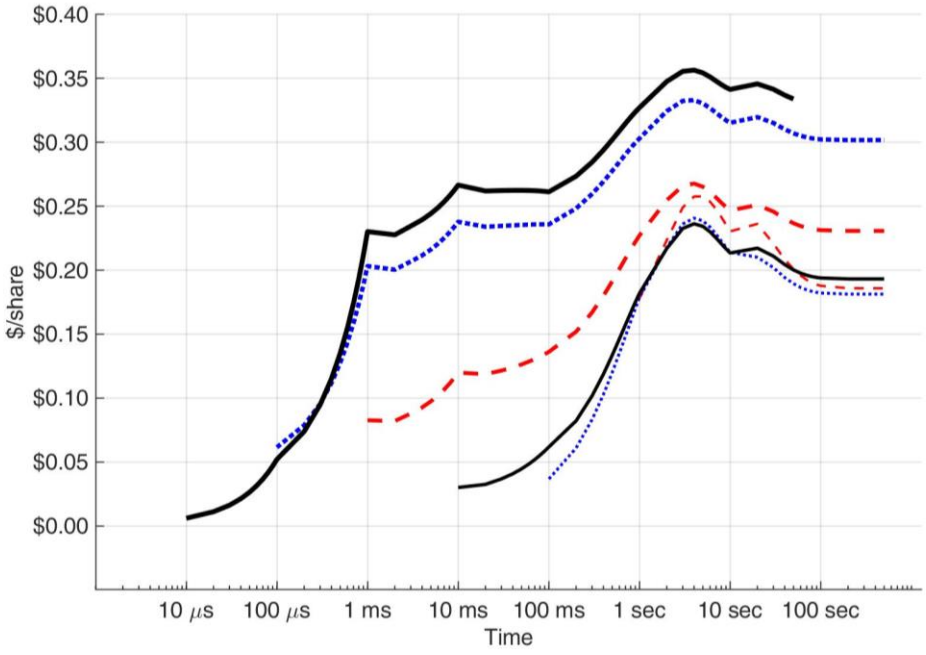
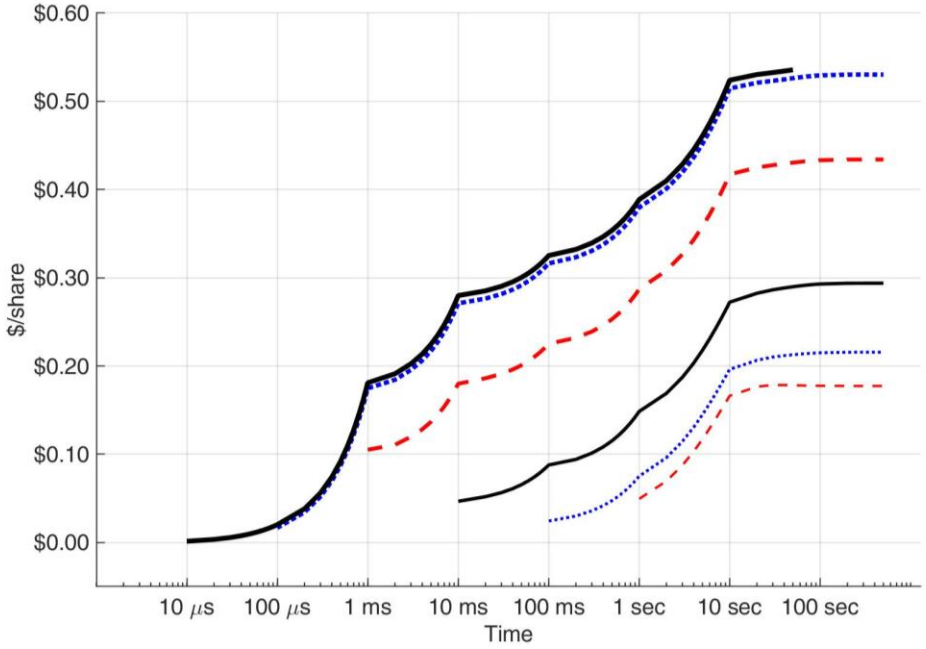


Figure 3. Impulse response functions, last sale prices and quotes.

The VECM system comprises four price variables: the national best bid and offer (NBB and NBO), the last sale price for an execution on a lit exchange, and the last sale price for a dark execution (an execution reported on FINRA's ADF). The system is estimated at six resolutions, ranging from one second down to ten microseconds. IRFs depict the cumulative response in the NBB following a time-zero one-dollar shock to the last sale price on a lit exchange. In the figure, the IRF estimated at a one-second resolution begins at one second, and similarly for the finer resolutions.

Panel A. IBM



Panel B. NVDA

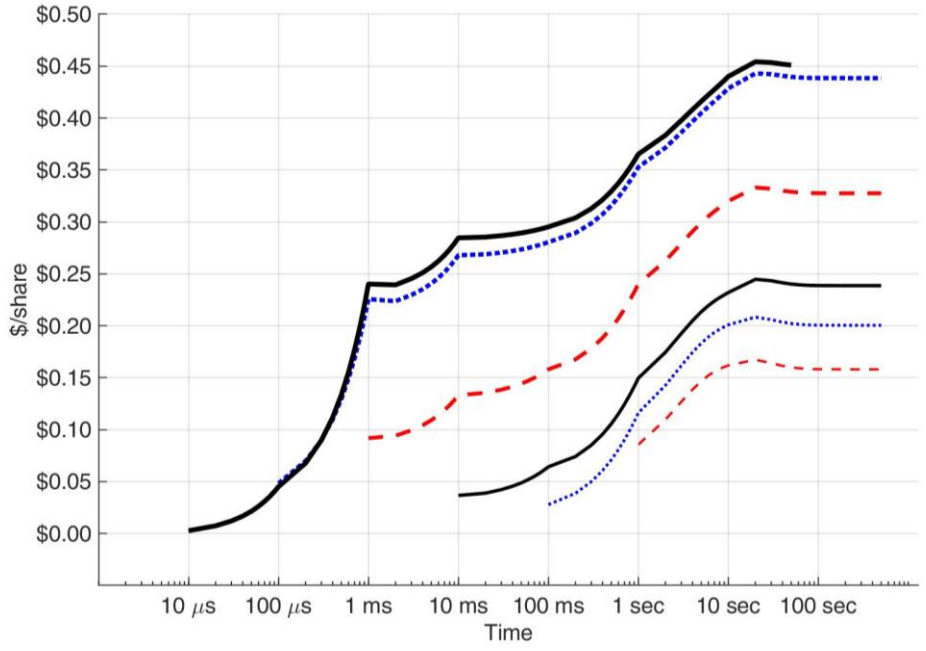


Figure 4. Regular and bridged impulse response functions for a simulated model.

The simulated model is

$$\Delta p_t = \gamma B p_{t-1} + \phi_1 \Delta p_{t-1} + \phi_2 \Delta p_{t-2} + \cdots + \phi_K \Delta p_{t-K} + \epsilon_t,$$

where $p_t = [p_{1t} \quad p_{2t}]', B = [1 \quad -1], \gamma = [-0.01 \quad 0.02]$, and $\epsilon_t \sim N(0, I_2)$. The $K = 100$ autoregressive coefficient matrices are: $\phi_1 = 0.1I$; $\phi_k = 0.2I$ for $k = 2, \ldots, 10$; $\phi_k = 0.005I$ for $k = 11, \ldots, 100$. Based on one million simulated observations, three models are estimated: the short model is truncated at $K = 10$; the long model also has $K = 10$ lags, but it is applied to prices sampled every ten periods; the short and long model corresponds to the correct specification. The figure depicts the dynamics in p_1 subsequent to a one-unit shock to p_1, for each of the three models. The bridged response is computed by taking the ten-step ahead forecast from the short model as the starting value for forecasting the long model.

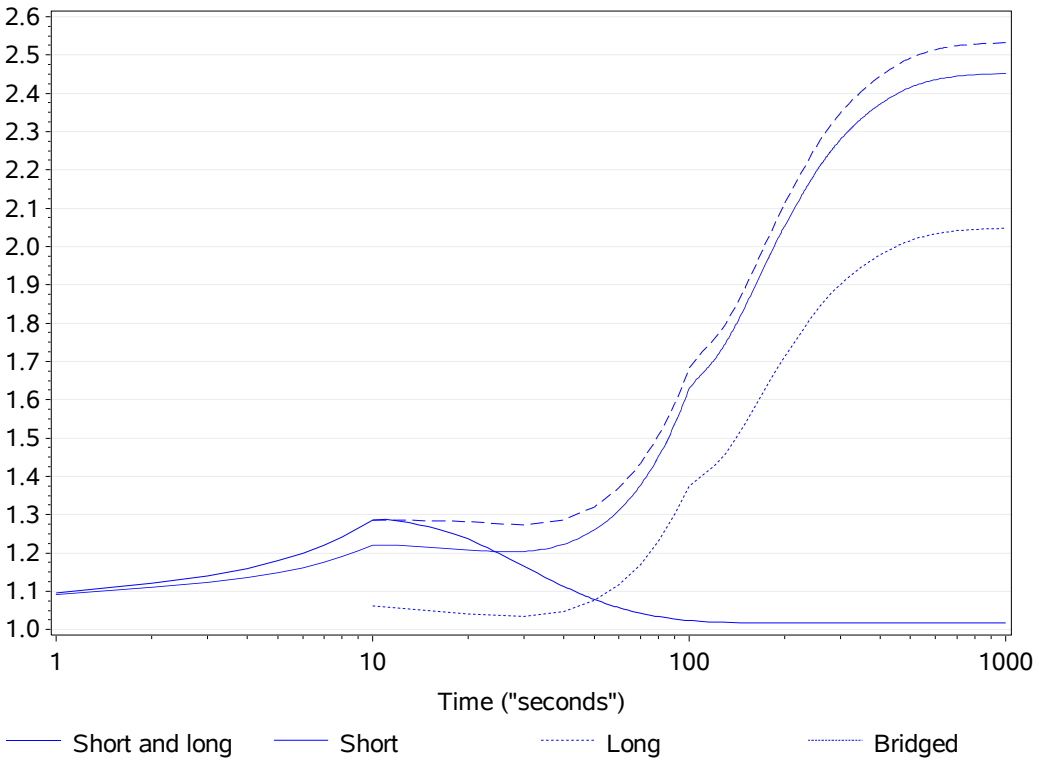


Figure 5. The bridged impulse response function for IBM quotes

The data consist of (national best) bids and offers based on participant and SIP timestamps: $\{NBBpart, NBOpart, NBBsip, NBOsip\}$. The figure depicts the response in $NBBsip$ subsequent to a one-unit shock in $NBBpart$, constructed by bridging forecasts across different timescales. At each resolution $d \in \{10\mu s, 100\mu s, 1ms, 10ms, 100ms\}$ I estimate a VECM. Beginning at $d = 10\mu s$, I forecast one hundred periods ahead, and use these as starting values for forecasts based on the $d = 100\mu s$ model, and so forth. The IRF estimated at a ten microsecond resolution begins at one microsecond, and similarly for the coarser resolutions. Adjoining forecasts are slightly shifted to clarify the periods of overlap.