School of Economics Working Paper
2019-02

Price discovery in a continuous-time setting

by   Gustavo Fruet Dias*
      Marcelo Fernandes**
      Cristina Mabel Scherrer***

*School of Economics, University of East Anglia and
CREATES
**Sao Paulo School of Economics, FGV
***Norwich Business School, University of East Anglia and
CREATES

Price discovery in a continuous-time setting

Gustavo Fruet Dias*

School of Economics, University of East Anglia and CREATES

Marcelo Fernandes

Sao Paulo School of Economics, FGV

Cristina Mabel Scherrer

Norwich Business School, University of East Anglia and CREATES

* Corresponding author: 2.27, Registry, University of East Anglia, Norwich Research Park, Norwich, Norfolk, NR4 7TJ, UK. Tel. : +44 (0) 1603 592941. E-mail address: g.fruet-dias@uea.ac.uk.

Abstract: We formulate a continuous-time price discovery model and investigate how the standard price discovery measures vary with respect to the sampling interval. We find that the component share measure is invariant to the sampling interval, and hence, discrete-sampled prices suffice to identify the continuous-time component share. In contrast, information share estimates are not comparable across different sampling intervals because the contemporaneous correlation between markets increases in magnitude as the sampling interval grows. We show how to back out the continuous-time information share from discrete-sampled prices under certain assumptions on the contemporaneous correlation. We assess our continuous-time model by comparing the estimates of the (continuous-time) component and information shares at different sampling intervals for 30 stocks in the US. We find that both price discovery measures are typically stable across the different sampling intervals, suggesting that our continuous-time price discovery model fits the data very well.

1   Introduction

Equity markets in the US and Europe have experienced an ongoing process of market fragmentation following regulatory policy changes that aimed to increase competition (see Menkveld, 2014; O'Hara, 2015; Menkveld, 2016; among others).[1] The immediate consequence of this process was that the market share of the listing exchanges decreased dramatically, while the new entrants have captured significant order flow. High-frequency trading has contributed to scatter quotes across the different exchanges and has also made markets much quicker, with time scales of microseconds or even nanoseconds (O'Hara, 2015; Hasbrouck, 2018). In turn, quotes, trades, and information are now dispersed across a variety of exchanges and markets that are populated by players possessing different strategic behaviours. It is thus important to investigate how exchanges and markets impound information on the efficient price of securities in an environment of highly competitive fragmented markets that operate in extremely fast time frames.

This paper addresses this issue by formulating a price discovery model in continuous time. In particular, we first investigate how the standard price discovery measures in discrete time vary with the sampling interval and then show how to recover the continuous-time price discovery measures from discrete-sampled prices. There are essentially two standard price discovery measures: information share (IS) and component share (CS). The former is ascribed to Hasbrouck (1995), gauging the contribution of each market/venue to the total variation in the efficient price innovation (see, for instance, Grammig, Melvin and Schlag, 2005; Lien and Shrestha, 2009; de Jong and Schotman, 2010; Fernandes and Scherrer, 2018). The latter results from the application of the permanent-transitory decomposition of Gonzalo and Granger (1995) and Gonzalo and Ng (2001) to price discovery analysis (see, among others Booth, So and Tseh, 1999; Chu, Hsieh and Tse, 1999; Figuerola-Ferretti and Gonzalo, 2010).

Both measures rely on the estimation of a vector error-correction model (VECM) for price changes in discrete time, with the speed-of-adjustment parameter playing a major role in the price discovery analysis. Baillie, Booth, Tse and Zabotina (2002), de Jong (2002) and Yan and Zivot (2010) provide a formal comparison between CS and IS measures in a discrete-time setting. They show they render similar results (up to a different normalization) if market innovations are contemporaneously uncorrelated and market-specific variances are similar in value. However,

these conditions are highly unlikely to hold in practice because time aggregation usually leads to stronger contemporaneous correlation across markets. In addition, market-specific variances often differ markedly for sufficiently small sampling intervals.

To examine both component and information shares in a continuous-time setting, we assume that the prices of a homogeneous asset traded at k different venues follow a reduced-rank multivariate Ornstein-Uhlenbeck process. We investigate how the component and information shares vary with respect to the sampling interval under exact discretization. We find that the component share is invariant, whereas the information share converges to the value of $1/k$ as the sampling interval increases. This finding essentially means that price discovery analyses at different sampling intervals are not directly comparable if based on discrete-time information shares.

To circumvent this issue, we show how to back out a continuous-time IS measure from discrete-sampled prices that is more appropriate for markets that operate with extremely fast time frames. The latter setting is perhaps the main motivation for Hasbrouck (2018) to examine price discovery in higher resolutions, given that the contemporaneous correlation between markets is still nonzero in the one-second resolution. Next, we extend the early comparison between the CS and IS measures in discrete time (Baillie et al., 2002; de Jong, 2002; and Yan and Zivot, 2010) to the continuous-time setting. We show that the conditions under which they yield similar results in discrete time become less likely to hold as the sampling interval increases.

We empirically assess the implications of our continuous-time price discovery model by comparing the continuous-time CS and IS estimates across different sampling intervals using data from 30 stocks in the US. We find that, although their standard errors increase, the CS estimates remain stable for both thickly and thinly traded stocks as we increase the sampling interval from one to up to five minutes. More formally, our bootstrap-based tests do not reject the null hypothesis of CS invariance to the sampling interval, indicating that our continuous-time price discovery model fits the data very well. We unveil similar evidence for the continuous-time IS measures of heavily traded stocks. Although market leadership inference does not change much, the pointwise estimates of the continuous-time IS measures vary significantly across sampling intervals because illiquidity compromises the estimation of the contemporaneous correlation between markets in lower resolutions.

The remainder of this paper proceeds as follows. Section 2 describes the continuous-time setting for price discovery, whereas Section 3 discusses how the sampling interval affects the CS and IS measures. Section 4 tests our continuous-time price discovery model by checking whether the (continuous-time) CS and IS measures change as we increase the sampling interval from one to five minutes. In addition, we investigate whether the relative informativeness of NYSE- and Nasdaq-listed stock prices change with data frequency. Section 5 offers some concluding remarks. Finally, the appendix not only outlines the residual-based bootstrap procedure we employ to carry out the formal tests in Section 4 but also reports some simulation results.

2 A continuous-time setting for price discovery

In this section, we propose a continuous-time model for price discovery. Let prices for a given asset that trades on multiple venues follow the process

$$\mathrm{d}P_t = \Pi P_t \, \mathrm{d}t + C \, \mathrm{d}W_t, \qquad \text{with} \quad P_0 = p_0, \tag{1}$$

where $P_t = (p_{1,t}, \ldots, p_{k,t})'$ is a $k \times 1$ vector of log prices with k denoting the number of trading venues, $\Pi = \alpha\beta'$ is a $k \times k$ reduced-rank matrix with rank equal to $r = k - 1$, α and β are $k \times r$ full-rank matrices, W is a $k \times 1$ vector of Brownian motion, and C is a $k \times k$ matrix such that the covariance matrix $\Sigma = CC'$ is positive definite. Prices in the different markets should not drift apart much, oscillating around the (latent) efficient price, as they refer to the same asset. Accordingly, there are $k - 1$ cointegrating relationships ($r = k - 1$), with log prices sharing the asset's efficient price as the single common stochastic trend. We assume without loss of generality that β is known and takes the form of $\beta = (I_r, -\iota_r)$, where ι_r denotes a $r \times 1$ unit vector. In turn, α determines how quickly each market reacts to deviations from the long-run equilibria $\beta'P_t$.

The solution to (1) is a homogenous Gaussian Markov process given by

$$P_{t+\delta} = \exp(\delta\Pi) \left[P_t + \int_0^\delta \exp(-u\Pi)C \, \mathrm{d}W_u \right], \tag{2}$$

giving way to a homoskedastic Gaussian VAR(1) process in discrete time. Due to the reduced rank, it will also admit a homoskedastic Gaussian VECM(0) representation as in Hasbrouck (1995). Despite the restricted lag structure in the discrete-time VECM, the reduced-rank Ornstein-Uhlenbeck (OU) process in (1) provides a useful framework to study the dynamics

5

of a single asset traded at multiple venues, given that it ensures that the stochastic trend is a martingale and that returns follow an infinitive VMA process. Nevertheless, it is possible to contemplate reduced-rank continuous-time processes that yield more general lag structures to their discrete-time counterparts by applying the Laplace transform function to the lag operator as in Nguenang (2016).[2] See also Cochrane (2012) for more details on ARMA processes in continuous time.

We assume that prices are observed regularly and equidistantly over the unit interval $[0, 1]$ that characterizes, say, one trading day (calendar-time sampling, as discussed in Hansen and Lunde, 2006). Denote each interval in $[0, 1]$ as $[t_{i-1}, t_i]$, where $i = 1, 2, \ldots, n$ and n is the total number of intervals such that $0 = t_0 < t_1 < \ldots < t_n = 1$. The length of each interval is $\delta = t_i - t_{i-1} = 1/n$ in $[0, 1]$. For instance, the usual trading day in the US market lasts for 6.5 hours (23,400 seconds), and thus, sampling one observation per minute yields $n = 390$ intraday observations, with $\delta = 1/390$. Denoting by $\exp(A)$ the matrix exponential of a $k \times k$ matrix A such that $\exp(A) = \sum_{\ell=0}^{\infty} \frac{1}{\ell!} A^\ell$, the exact discretization of (1) at interval length δ reads

$$\Delta P_{t_i} = \Pi_\delta P_{t_{i-1}} + \varepsilon_{t_i}, \tag{3}$$

where $\Pi_\delta = \alpha_\delta \beta'$ and $\alpha_\delta = \alpha(\beta'\alpha)^{-1} \left[\exp(\delta\beta'\alpha) - I_r \right]$, with I_r denoting a r-dimensional identity matrix, and P_{t_i} is a $k \times 1$ vector of log prices observed at discrete time. The innovation ε_{t_i} is iid Gaussian with zero mean and covariance matrix given by $\Sigma_\delta = \int_0^\delta \exp(u\Pi)\Sigma \exp(u\Pi') \, du$.

Kessler and Rahbek (2004) provide the conditions under which the mapping given $\theta = (\Pi, \Sigma) \overset{\psi}{\mapsto} \psi(\theta) = (\Pi_\delta, \Sigma_\delta)$ is unique, θ is identifiable, and the space spanned by the columns of α is equal to the one spanned by the columns of α_δ. Specifically, if all eigenvalues of Π are real and no elementary divisor of Π occurs more than once, Proposition 1 in Kessler and Rahbek (2004) shows that the mapping ψ is injective and θ is identifiable. It is important to note that temporal aggregation preserves the cointegration rank, i.e., $\operatorname{rank}\Pi_\delta = \operatorname{rank}\Pi$, and that the definition of (co)integration for OU processes in continuous time is consistent with the definition in discrete time (Kessler and Rahbek, 2004). Therefore, one may conduct inferences about rank and cointegrating space using discrete-time procedures and then interpret the results in a continuous-time setting.

To compute price discovery measures, one must decompose the price vector into a permanent

6

$I(1)$ component and a transitory covariance-stationary $I(0)$ component. There are essentially two alternative decompositions in the price discovery literature: Gonzalo and Granger's (1995) permanent-transitory decomposition (PTD) and the Granger representation theorem (GRT). Although they have different implications for the permanent component in a general VECM(p) setting, the resulting price discovery measures coincide up to a scale factor (de Jong, 2002), due to their common reliance on the orthogonal complements of β and α (or α_δ if in discrete time).

Let A be a $k \times r$ matrix with full column rank $r \leq k$. We define the orthogonal complement of A as any matrix $A_\perp$ with dimensions $k \times (k - r)$ and rank $k - r$ such that $A'A_\perp = 0$.[3] In particular, let the $k \times 1$ vectors $\alpha_\perp$ and $\alpha_{\delta,\perp}$ respectively denote the orthogonal complements of the vector of speed-of-adjustment parameters in continuous and discrete time, whereas we fix the orthogonal complement of $\beta = (I_r, -\iota_r)$ to $\beta_\perp = \iota_r$ (any multiple would do). These orthogonal complements relate to the nonstationary directions of the processes and hence to the permanent component that reflects the efficient price.

The PTD posits that $P_{t_i} = a_f f_{t_i} + a_z z_{t_i}$, where the common factor f_{t_i} is a linear combination of the elements of P_{t_i}. Identification requires the absence of long-run Granger causality from the transitory component z_{t_i} to f_{t_i}, implying $a_f = \beta_\perp$ and $f_{t_i} = \left(\alpha'_{\delta,\perp}\beta_\perp\right)^{-1}\alpha'_{\delta,\perp}P_{t_i}$. The GRT extends the Beveridge-Nelson decomposition to a multivariate setting (Stock and Watson, 1988; Johansen, 1991; Hansen, 2005), resulting in a permanent component that follows a random walk process with serially uncorrelated increments (and, hence, a martingale). In turn, the transitory term admits a covariance-stationary VMA(∞) representation, thus inheriting all the remaining serial correlation.

In general, the permanent components of the PTD and GRT differ, given that there is no guarantee that the former is a martingale. However, in the particular case of (3), there is no autocorrelation in $\Delta f_{t_i} = \left(\alpha'_{\delta,\perp}\beta_\perp\right)^{-1}\alpha'_{\delta,\perp}\Delta P_{t_i}$ because $\alpha'_{\delta,\perp}\Delta P_{t_i} = \alpha'_{\delta,\perp}\varepsilon_{t_i}$ given that $\alpha'_{\delta,\perp}\alpha_\delta = 0$. Accordingly, both decompositions yield the same martingale process for the efficient price underlying (3). This means there is no instantaneous risk premium, which is, in practice, a very reasonable assumption for intraday returns (see discussion in Hansen and Lunde, 2006).

To formally establish a bridge between the continuous- and discrete-time price discovery measures, we first establish that the GRT holds in both settings. Kessler and Rahbek (2001) show that the GRT follows in continuous time directly by assuming that α and β have full

column ranks r, $\beta'\alpha$ has full rank r, and all eigenvalues of $\beta'\alpha$ have negative real parts. Under these assumptions,

$$I_k = \beta_\perp(\alpha'_\perp\beta_\perp)^{-1}\alpha'_\perp + \alpha(\beta'\alpha)^{-1}\beta', \qquad (4)$$

and hence a straightforward application of (4) to the solution of (1) yields

$$P_t = \Xi(CW_t + P_0) + \eta_t, \qquad (5)$$

where $\Xi = \beta_\perp\left(\alpha'_\perp\beta_\perp\right)^{-1}\alpha'_\perp$, P_0 contains initial values, and $\eta_t = \alpha\left(\beta'\alpha\right)^{-1}Z_t$, with $Z_t = \beta'P_t$ denoting a stationary OU process given by $\mathrm{d}Z_t = \beta'\alpha Z_t\,\mathrm{d}t + \beta'C\,\mathrm{d}W_t$.

The exact discretization in (3) is such that

(i) prices are not explosive in that the roots of the characteristic polynomial $|I_k - (\Pi_\delta + I_k)z| = 0$ are either outside the unit circle or equal to one;

(ii) $\Pi_\delta = \alpha_\delta\beta'$ has reduced rank $r < k$, with both α_δ and β having full column rank r; and

(iii) the number of unit roots equals $k - r$.

It then follows from Johansen's (1995) Theorem 4.2 that the GRT holds in discrete time irrespective of the distribution of the innovations in (3). Moreover, conditions (i) to (iii) ensure that the projection identity in (4) holds in discrete time, namely,

$$I_k = \beta_\perp(\alpha'_{\delta,\perp}\beta_\perp)^{-1}\alpha'_{\delta,\perp} + \alpha_\delta(\beta'\alpha_\delta)^{-1}\beta'. \qquad (6)$$

Pre-multiplying (3) by β' yields the r-dimensional stationary process $\beta'P_{t_i} = \sum_{h=0}^\infty(I_r + \beta'\alpha_\delta)^h\beta'\varepsilon_{t_{i-h}}$, whereas pre-multiplying by $\alpha'_{\delta,\perp}$ entails the nonstationary component $\alpha'_{\delta,\perp}P_{t_i} = \alpha'_{\delta,\perp}P_{t_0} + \sum_{i=1}^n\alpha'_{\delta,\perp}\varepsilon_{t_i}$.

Using the projection identity in (6), the GRT in discrete time reads

$$P_{t_i} = \Xi_\delta\sum_{h=1}^i\varepsilon_{t_h} + \sum_{h=0}^\infty\Upsilon_{\delta,h}\varepsilon_{t_{i-h}} + \Xi_\delta P_{t_0}, \qquad (7)$$

or, alternatively,

$$\Delta P_{t_i} = \Xi_\delta\varepsilon_{t_i} + \left[\Upsilon_{\delta,0}\varepsilon_{t_i} + \sum_{h=1}^\infty(\Upsilon_{\delta,h} - \Upsilon_{\delta,h-1})\,\varepsilon_{t_{i-h}}\right], \qquad (8)$$

where $\Xi_\delta = \beta_\perp(\alpha'_{\delta,\perp}\beta_\perp)^{-1}\alpha'_{\delta,\perp}$, $\Upsilon_{\delta,h} = (I_k - \Xi_\delta)(I_k + \alpha_\delta\beta')^h$ such that $\sum_{h=0}^\infty\Upsilon_{\delta,h}\varepsilon_{t_{i-h}}$ is a stationary process, and P_{t_0} is a vector of initial values.[4] The stochastic common trend given by

the first term on the right-hand side of (7) reflects the efficient price of the asset. As $\beta_\perp = \iota_r$, not only does Ξ_δ have common rows, but the efficient price also relates to a weighted-average of the prices at the different trading venues. Finally, as mentioned above, the stochastic trend given by $\left(\alpha'_{\delta,\perp}\beta_\perp\right)^{-1}\alpha'_{\delta,\perp}\sum_{h=1}^{i}\varepsilon_{t_h}$ forms a martingale sequence of efficient prices in that its difference $\left(\alpha'_{\delta,\perp}\beta_\perp\right)^{-1}\alpha'_{\delta,\perp}\varepsilon_{t_i}$ is serially uncorrelated.

It is apparent from the above discussion that α_δ and $\alpha_{\delta,\perp}$ are key parameters in any price discovery analysis. For instance, α_δ reflects the adjustment that each market implements such that their prices do not deviate from the latent efficient price. The closer $\alpha_{\delta,m}$ is to zero, the less the market m adjusts to the efficient price. In particular, $\alpha_{\delta,m} = 0$ means that the price at market m coincides with the efficient price, therefore leading the price discovery.

3 The effect of the sampling interval

A natural step forward is to investigate how the sampling interval affects the component and information shares. In this section, we first establish how both measures behave under temporal aggregation starting from the continuous-time setting. We then show how to identify them in continuous time from prices at any sampling interval. Next, we compare the information content of the CS and IS measures in continuous time and wrap up the discussion with a simple example. For simplicity, we consider a single asset traded on two trading venues (i.e., $k = 2$ and $r = 1$).

3.1 Component share

The component share relies on the orthogonal complement of α_δ, namely, $\alpha_{\delta,\perp}$ such that $\alpha'_{\delta,\perp}\alpha_\delta = 0$ (see, among others, Booth et al., 1999; Chu et al., 1999; Harris, McInish and Wood, 2002; Hansen and Lunde, 2006). Because $\alpha_{\delta,\perp}$ is not unique, one typically imposes $\alpha_{\delta,\perp,1} + \alpha_{\delta,\perp,2} = 1$. Whereas α_δ corresponds to the stationary direction of the process in (3), $\alpha_{\delta,\perp}$ relates to the nonstationary direction. Thus, $\alpha_{\delta,\perp}$ is a natural quantity to assess how the efficient price relates to each market innovation. The market with the highest $\alpha_{\delta,\perp}$ has the lowest need to adjust towards the latent efficient price, and hence, it is the one that leads the price discovery process.

Under the normalization $\alpha_{\delta,\perp,1} + \alpha_{\delta,\perp,2} = 1$, it follows from the exact discretization of the reduced-rank OU process in (3) that

$$\alpha_{\delta,\perp} = \left(\frac{\alpha_{\delta,2}}{\alpha_{\delta,2} - \alpha_{\delta,1}}, -\frac{\alpha_{\delta,1}}{\alpha_{\delta,2} - \alpha_{\delta,1}}\right)' = \left(\frac{\alpha_2}{\alpha_2 - \alpha_1}, -\frac{\alpha_1}{\alpha_2 - \alpha_1}\right)', \tag{9}$$

given that $(\beta'\alpha)^{-1}[\exp(\delta\beta'\alpha) - I_r]$ cancels out for appearing in both numerators and denominators. It is now clear that $\alpha_{\delta,\perp}$ is invariant to the sampling interval in that $\alpha_{\delta,\perp} = \alpha_\perp$ for any $0 < \delta < 1$.[5]

Allowing for a stochastic covariance matrix Σ_t does not affect the CS measure. In particular, exact discretization would still yield Markovian price processes, with the same autoregressive matrix as in (2). See, for instance, Proposition 1 in Nguenang (2016). As a result, CS remains so that $\alpha_\perp = \alpha_{\delta,\perp}$ for any $0 < \delta < 1$. This means that identification and inference of the continuous-time price discovery measure arise directly from estimating $\alpha_{\delta,\perp}$ at any sampling interval. From an empirical perspective, (9) allows us to learn about the continuous-time price discovery mechanism even if using data at a higher sampling interval.

3.2 Information share

Hasbrouck's (1995) IS measure gives the share of each market contribution to the total variance of the efficient price (see, among others, Baillie et al., 2002; de Jong, 2002; Grammig, Melvin and Schlag, 2005; Yan and Zivot, 2010). Using the exact discretization of (1), the IS measure of a given market $m \in \{1,2\}$ for $0 < \delta < 1$ is

$$IS_{\delta,m} = \frac{[\xi_\delta C_\delta]_m^2}{\xi_\delta \Sigma_\delta \xi_\delta'}, \tag{10}$$

where $\Sigma_\delta = C_\delta C_\delta' = \int_0^\delta \exp(u\Pi)\Sigma \exp(u\Pi') \, du$, ξ_δ is the common row of Ξ_δ in (7) that follows from $\beta_\perp = (1,1)'$, and $[\,\cdot\,]_m$ denotes the mth element of a vector.

We now investigate the effect of δ on the IS measure. Considering the two-market case, we can re-write (10) as a function of the market-specific variances and correlations. Specifically, let

$$\Sigma_\delta = \begin{pmatrix} \sigma_{\delta,1}^2 & \sigma_{\delta,1}\sigma_{\delta,2}\rho_\delta \\ \sigma_{\delta,1}\sigma_{\delta,2}\rho_\delta & \sigma_{\delta,2}^2 \end{pmatrix} \quad \underline{C}_\delta = \begin{pmatrix} \sigma_{\delta,1} & 0 \\ \sigma_{\delta,2}\rho_\delta & \sigma_{\delta,2}\sqrt{1-\rho_\delta^2} \end{pmatrix} \quad \bar{C}_\delta = \begin{pmatrix} \sigma_{\delta,1}\sqrt{1-\rho_\delta^2} & \sigma_{\delta,1}\rho_\delta \\ 0 & \sigma_{\delta,2} \end{pmatrix},$$

where $\underline{C}_\delta$ and $\bar{C}_\delta$ are the Cholesky decompositions of Σ_δ resulting from the different orderings of the variables such that $\Sigma_\delta = \underline{C}_\delta \underline{C}_\delta' = \bar{C}_\delta \bar{C}_\delta'$. We denote the variance in market $m \in \{1,2\}$ by $\sigma_{\delta,m}^2$ and the contemporaneous correlation between the two markets by ρ_δ for any δ.

As the Cholesky decomposition depends on the ordering of the variables, most studies average the maximum and minimum IS measures across all possible orderings. As $\alpha_{\delta,\perp,m} = \alpha_{\perp,m}$ for

any $0 < \delta < 1$, the average IS measure in a given market $m \in \{1, 2\}$ for $0 < \delta < 1$ then reads

$$\overline{IS}_{\delta,m} = \frac{1}{2}\left(\frac{[\xi_\delta C_\delta]^2_m}{\xi_\delta \Sigma_\delta \xi'_\delta} + \frac{[\xi_\delta \bar{C}_\delta]^2_m}{\xi_\delta \Sigma_\delta \xi'_\delta}\right) = \begin{cases} \frac{(\alpha_{\perp,1}\sigma_{\delta,1}+\alpha_{\perp,2}\sigma_{\delta,2}\rho_\delta)^2+\alpha^2_{\perp,1}\sigma^2_{\delta,1}(1-\rho^2_\delta)}{2(\alpha^2_{\perp,1}\sigma^2_{\delta,1}+\alpha^2_{\perp,2}\sigma^2_{\delta,2}+2\alpha_{\perp,1}\alpha_{\perp,2}\sigma_{\delta,1}\sigma_{\delta,2}\rho_\delta)}, & \text{if } m = 1, \\[2em] \frac{(\alpha_{\perp,2}\sigma_{\delta,2}+\alpha_{\perp,1}\sigma_{\delta,1}\rho_\delta)^2+\alpha^2_{\perp,2}\sigma^2_{\delta,2}(1-\rho^2_\delta)}{2(\alpha^2_{\perp,1}\sigma^2_{\delta,1}+\alpha^2_{\perp,2}\sigma^2_{\delta,2}+2\alpha_{\perp,1}\alpha_{\perp,2}\sigma_{\delta,1}\sigma_{\delta,2}\rho_\delta)}, & \text{if } m = 2. \end{cases} \tag{11}$$

It is apparent from (11) that, unlike the component share, $\overline{IS}_{\delta,m}$ is not invariant to the sampling interval because the market-specific variances and correlation across markets depend on δ. In particular, the contemporaneous correlation absorbs most of the lead-lag patterns as δ increases because both markets have now sufficient time to impound the news. In fact, exact discretization yields $|\rho_\delta| \to 1$ as $\delta \to 1$, and thus, $\lim_{\delta \to 1} \overline{IS}_{\delta,1} = \lim_{\delta \to 1} \overline{IS}_{\delta,2} = 1/2$. Accordingly, a fair comparison of IS measures should account for the sampling interval.

A possible solution of how to avoid the IS measures in discrete time to converge to $1/2$ consists of using data sampled in higher resolutions (e.g., in milliseconds). Hasbrouck (2018) finds that the bounds of the discrete-time IS measures are very wide at the one-second sampling interval. Although they become narrower as resolution increases, they become sufficiently tight only at the sampling frequency of one observation per ten microseconds. This happens because markets are currently not only very fast but also highly interconnected due to the rise of high-frequency trading and statistical cross-market arbitrage (see, among others, Menkveld, 2014, 2016; O'Hara, 2015). As such, the contemporaneous correlation between markets is nonzero even at the very high frequency, such that the IS measures converge to $1/2$ at faster rates.

A natural step forward is to define an IS measure in continuous time. Using the GRT in (5), the continuous-time IS measure of a given market $m \in (1, 2)$ reads

$$IS_m = \frac{[\xi C]^2_m}{\xi \Sigma \xi'}, \tag{12}$$

where $\Sigma = CC'$, ξ is the common row of Ξ in (5), and $[.]_m$ denotes the mth element of a vector. As in the discrete-time setting, the decomposition $\Sigma = CC'$ is not unique, so we average the IS measures over all possible orderings. The average continuous-time IS measure in a given market $m \in (1, 2)$ then reads

$$\overline{IS}_m = \frac{1}{2}\left(\frac{[\xi C]^2_m}{\xi \Sigma \xi'} + \frac{[\xi \bar{C}]^2_m}{\xi \Sigma \xi'}\right), \tag{13}$$

where $\underset{\sim}{C}$ and $\bar{C}$ are the Cholesky decompositions of Σ resulting from the different orderings of the variables such that $\Sigma = \underset{\sim}{C}\underset{\sim}{C}' = \bar{C}\bar{C}'$.

The continuous-time IS measure is immune to the sampling interval and, most importantly, fits well the current trading environment (i.e., fast and furious markets). Finally, it is possible to make inferences about the continuous-time IS measure in (13) using discrete-sampled prices. Indeed, $\hat{\xi}$ readily follows from the VECM parameter estimates and $\widehat{\Sigma}$ from realized measures of covariation (Barndorff-Nielsen and Shephard, 2004; Barndorff-Nielsen, Hansen, Lunde and Shephard, 2011).

Assuming a time-varying covariance matrix now matters because the exact discretization of IS depends heavily on how the stochastic process that governs the dynamics of the instantaneous covariance matrix aggregates over time. As the standard practice in the literature is to estimate daily IS measures, an alternative is to compute the continuous-time IS measures based on daily integrated covariance matrices: $\Sigma_d = \int_d^{d+1} \exp(u\Pi)\Sigma_u \exp(u\Pi')\,\mathrm{d}u$, where $d = 1,\ldots,D$ with D denoting the number of days in the sample. Simulation results in Appendix B indicate that the median estimates are very close to the IS measures that the exact discretization of (1) would imply, so that they still converge to 1/2 as the sampling interval increases due to the increase in the contemporaneous correlation.

3.3   A continuous-time comparison of CS and IS

Next, we investigate the relationship between the CS and IS measures in continuous time. First, rewrite the continuous-time IS measure in a given market $m \in (1,2)$ as

$$
IS_m = \begin{cases}
\dfrac{(\alpha_{\perp,1}\sigma_1 + \alpha_{\perp,2}\sigma_2\rho)^2 + \alpha_{\perp,1}^2\sigma_1^2(1-\rho^2)}{2\big(\alpha_{\perp,1}^2\sigma_1^2 + \alpha_{\perp,2}^2\sigma_2^2 + 2\alpha_{\perp,1}\alpha_{\perp,2}\sigma_1\sigma_2\rho\big)}, & \text{if } m = 1, \\[2ex]
\dfrac{(\alpha_{\perp,2}\sigma_2 + \alpha_{\perp,1}\sigma_1\rho)^2 + \alpha_{\perp,2}^2\sigma_2^2(1-\rho^2)}{2\big(\alpha_{\perp,1}^2\sigma_1^2 + \alpha_{\perp,2}^2\sigma_2^2 + 2\alpha_{\perp,1}\alpha_{\perp,2}\sigma_1\sigma_2\rho\big)}, & \text{if } m = 2.
\end{cases}
\tag{14}
$$

where σ_1^2 and σ_2^2 are the market-specific variances and ρ denotes the correlation between the two markets in continuous time. There is a direct link between CS and IS only if $\rho = 0$ and $\sigma_1 = \sigma_2$. In this case, the information share in continuous time reads

$$
IS_m = \frac{\alpha_{\perp,m}^2\sigma_m^2}{\alpha_{\perp,1}^2\sigma_1^2 + \alpha_{\perp,2}^2\sigma_2^2}, \qquad m = 1, 2.
\tag{15}
$$

This essentially reproduces in continuous time the results that Baillie et al. (2002) and de Jong (2002) obtain in discrete time by imposing $\rho_\delta = 0$ and $\sigma_{\delta,1} = \sigma_{\delta,2}$. It is worth stressing, however, that these conditions are highly unlikely to hold in either continuous or discrete time. Given that $\lim_{\delta \to 1} |\rho_\delta| = 1$, it becomes increasingly more unrealistic to assume that $\rho_\delta = 0$ as

the sampling interval increases. On the other hand, the market-specific variances should diverge as the sampling interval declines due to the different market microstructure features and trading clientele (Dias, Scherrer and Papailias, 2016). We thus conclude that the continuous-time CS and IS measures will most often produce different empirical results.

3.4 A simple illustration

In this section, we illustrate our theoretical findings using a continuous-time version of the Hasbrouck's (1995) example. Suppose that a homogeneous asset trades on two markets. Market 1 is the leading trading venue with prices fully reflecting the efficient price, whereas the price on market 2 reacts to deviations with respect to the (efficient) price on market 1. The absence of adjustment in market 1 ($\alpha_1 = 0$) implies that $p_{1,t}$ coincides with the stochastic trend, and hence, price discovery measures should identify market 1 as the sole contributor to the price discovery process, regardless of the sampling interval.

In this setting, prices cointegrate with $\alpha = (0, \alpha_2)'$ and $\beta' = (1, -1)$:

$$\mathrm{d}\begin{pmatrix} p_{1,t} \\ p_{2,t} \end{pmatrix} = \begin{pmatrix} 0 \\ \alpha_2 \end{pmatrix} \begin{pmatrix} 1 & -1 \end{pmatrix} \begin{pmatrix} p_{1,t} \\ p_{2,t} \end{pmatrix} \mathrm{d}t + C \, \mathrm{d}W_t, \tag{16}$$

where C and W are defined as in (1). In particular, we entertain the exact discretization of (16) for $\sigma_1^2 = \sigma_2^2$ and $\rho \in \{0, 0.3, 0.5, 0.7, 0.9\}$, with δ ranging from $1/23, 400$ to $1/13$ (implying frequencies of one observation per second to one observation per 30 minutes in a trading day of 6.5 hours).

The left panel of Figure 1 displays the CS and IS measures for market 1, whereas the right panel plots the contemporaneous correlation at each sampling interval. As expected, discretization affects the price discovery measures in different manners. While CS is completely immune to the sampling interval, IS converges to $1/2$ as δ increases because $|\rho_\delta|$ converges to one. If the sampling interval is large enough, information hits both markets at the same time, preventing the IS measure from identifying market 1 in (16) as the unique contributor to the price discovery process. As expected, given that $\sigma_1^2 = \sigma_2^2$, CS and IS yield the same result in continuous time only in the absence of contemporaneous correlation between markets.

4 Empirical application: testing the continuous-time price discovery model implications

Our continuous-time setting has sharp testable implications for price discovery measures in discrete time. First, it implies that the CS measure is invariant to the sampling interval, even in the presence of a stochastic instantaneous covariance matrix. This means that we should find no significant differences in CS estimates across sampling intervals. Second, under correct specification of the continuous-time model, the same should apply to the continuous-time IS measures we compute from different sampling intervals. In the following subsection, we first describe the data and estimation details and then discuss our empirical findings.

4.1 Data description

Our high-frequency dataset refers to 30 stocks that differ markedly in terms of industry, listing venue, and trading activity. We group them into three subsamples: 10 Nasdaq-listed stocks, 10 thickly traded NYSE-listed stocks, and 10 thinly traded NYSE-listed stocks. The first group consists of a random sample from the Nasdaq-100 stock market index constituents that have been trading since January 2013: Adobe Systems (ADBE), Align Technology (ALGN), Amazon.com (AMZN), CA Technologies (CA), Expedia (EXPE), Alphabet (GOOG), Micron Technology (MU), Starbucks Corporation (SBUX), Vodafone Group (VOD), and Wendy's (WEN).

As for the actively traded NYSE-listed stocks, we select at random from the S&P 500 index constituents: Bank of America (BAC), General Electric (GE), Hewlett-Packard (HPQ), International Business Machines (IBM), J.C. Penney Company (JCP), JP Morgan Chase (JPM), Coca-Cola Company (KO), Altria Group (MO), Verizon Communications (VZ), and ExxonMobil (XOM). Finally, we randomly select 10 less-liquid NYSE-listed stocks from the Russell 1000 index constituents: Canon (CAJ), Cooper Companies (COO), Dolby Laboratories (DLB), Dun & Bradstreet Corp (DNB), Corporate Office Properties Trust (OFC), Regal Beloit Corp (RBC), Everest Re Group (RE), RPC Inc (RES), Rollins (ROL), and Thor Industries (THO). These stocks exhibit, on average, 70% less trading intensity, as measured by the number of trades, than the actively traded stocks in the previous group.

We extract quote data from TAQ for the year of 2013 because our simulation study in Appendix B shows that reliable inference requires a large sample period if markets operate in

extremely fast time frames. We implement the same cleaning filters as in Barndorff-Nielsen, Hansen, Lunde and Shephard (2009), excluding any observation with a zero quote, negative bid-ask spread, or outside the main trading hours (9:30 to 16:00). We also discard any data point, either with a bid-ask spread higher than 50 times the median spread on that day or with a midquote deviating by more than 10 mean absolute deviations from the median value of the last 50 observations. We then consider the median bid and ask quotes in the presence of multiple ticks for a given second and synchronize NYSE/Arca and Nasdaq midquotes by sampling at regularly spaced intervals of 1 to 5 minutes. Table 1 provides the final number of time-series observations for each stock before and after data handling.

4.2 Estimation of the continuous-time price discovery measures

Because we observe prices intradaily over different days, denote by $T = nD$ the total number of observations, where n is the number of intraday observations and D is the number of trading days. It then follows that $\Delta P_\tau = \alpha_\delta \beta' P_{\tau-1} + \varepsilon_\tau$, with $\tau = 1, \ldots, T$. Consistent estimation of the CS measure requires only a consistent estimator of the speed-of-adjustment parameters for any sampling interval given that $\beta = (1, -1)'$ is known. The continuous-time IS measure also calls for a consistent estimator of the covariance matrix.

As standard in the literature, we estimate α_δ by least squares. In view that $\beta = (1, -1)'$, we impose that $\widehat{\alpha}_{\delta,1} \leq 0$ and $\widehat{\alpha}_{\delta,2} \geq 0$, so that the roots of $|I_2 - (\widehat{\alpha}_\delta \beta' + I_2) z| = 0$ are either outside the unit circle or equal to one for every $\delta \in (1/390, 1/195, 1/130, 1/97, 1/78)$. The asymptotic normality of $\widehat{\alpha}_\delta$ follows from standard regularity conditions (see, for instance, Proposition 7.1 in Lütkepohl, 2007), and so we back out the limiting distribution of the continuous-time CS measures by the delta method. In particular, it follows from (9) that the Jacobian matrix of partial derivatives reads

$$\frac{\partial \alpha_\perp}{\partial \alpha_\delta'} = \begin{pmatrix} \frac{\alpha_{\delta,2}}{\left(-\alpha_{\delta,1}+\alpha_{\delta,2}\right)^2} & -\frac{\alpha_{\delta,1}}{\left(-\alpha_{\delta,1}+\alpha_{\delta,2}\right)^2} \\ -\frac{\alpha_{\delta,2}}{\left(-\alpha_{\delta,1}+\alpha_{\delta,2}\right)^2} & \frac{\alpha_{\delta,1}}{\left(-\alpha_{\delta,1}+\alpha_{\delta,2}\right)^2} \end{pmatrix}, \quad \text{for} \quad \delta = 1/390, 1/195, 1/130, 1/97, 1/78. \quad (17)$$

By construction, the asymptotic standard errors of $\widehat{\alpha}_{\delta,\perp,1}$ and $\widehat{\alpha}_{\delta,\perp,2}$ are identical.

To obtain daily IS measures in continuous time, we estimate the integrated covariance matrix over a day employing Barndorff-Nielsen and Shephard's (2004) realized covariance estimator $\widehat{\Sigma}_{\delta,d}$

using prices at the sampling interval δ.[6] The resulting daily continuous-time IS estimate is

$$\overline{IS}(\widehat{\alpha}_{\delta,\perp}, \widehat{\Sigma}_{\delta,d}) = \frac{1}{2} \left(\frac{\left[\widehat{\alpha}'_{\delta,\perp}\widehat{C}_{\delta,d}\right]^2_m}{\widehat{\alpha}'_{\delta,\perp}\widehat{\Sigma}_{\delta,d}\widehat{\alpha}_{\delta,\perp}} + \frac{\left[\widehat{\alpha}'_{\delta,\perp}\widehat{\overline{C}}_{\delta,d}\right]^2_m}{\widehat{\alpha}'_{\delta,\perp}\widehat{\Sigma}_{\delta,d}\widehat{\alpha}_{\delta,\perp}} \right) \tag{18}$$

where $\widehat{C}_{\delta,d}$ and $\widehat{\overline{C}}_{\delta,d}$ are the Cholesky decompositions of $\widehat{\Sigma}_{\delta,d}$ resulting from the different orderings of the variables such that $\widehat{\Sigma}_{\delta,d} = \widehat{C}_{\delta,d}\widehat{C}'_{\delta,d} = \widehat{\overline{C}}_{\delta,d}\widehat{\overline{C}}'_{\delta,d}$.

Before discussing the IS and CS estimates, it is important to provide some details about model selection and specification. As expected, we find only one cointegrating vector for every pair of stock prices using Johansen's maximum eigenvalue and trace tests at the 1% significance level. In addition, the Ljung-Box test with 30 lags cannot reject the absence of residual autocorrelation at the 1% significance level for every stock.[7] More importantly, we find only weak evidence against the martingale property of the efficient price implied by the GRT in (7). In particular, we test for the presence of serial correlation in $(\alpha'_{\delta,\perp}\beta_\perp)^{-1}\alpha'_{\delta,\perp}\varepsilon_\tau$ using the heteroskedastic-robust Ljung-Box tests with 1, 10, and 30 lags. Across all sampling intervals, we cannot reject the null hypothesis of no serial correlation at the 1% significance level for approximately 80% of the stocks, regardless of how many lags we contemplate. This proportion increases to over 85% if restricting attention to thickly traded stocks, irrespective of the sampling interval.[8]

4.3   Are the estimates indeed stable across sampling intervals?

In this section, we test whether the continuous-time price discovery measures computed from prices sampled at alternative sampling intervals are indeed stable using the residual bootstrap resampling method we describe in Appendix A. Specifically, we first construct bootstrap-based confidence intervals for the continuous-time price discovery measures using bootstrapped data at 2-, 3-, 4-, and 5-minute sampling intervals, based on the continuous-time estimates using prices at the 1-minute sampling interval. We then reject the null hypothesis that $\alpha_{1/390,\perp} = \alpha_{\delta,\perp}$ for $\delta \in \{1/195, 1/130, 1/97, 1/78\}$ if $\widehat{\alpha}_{\delta,\perp}$ lies outside the bootstrap-based confidence interval for a given percentile $p \in \{1\%, 5\%, 10\%\}$.

Table 2 reveals the CS estimates and their asymptotic-based standard errors across sampling intervals, as well as whether the bootstrap-based test rejects the null hypothesis that $\alpha_{1/390,\perp} = \alpha_{\delta,\perp}$ for $\delta \in \{1/195, 1/130, 1/97, 1/78\}$. In particular, we denote by *, **, and *** rejections

at the 10%, 5%, and 1% significance levels, respectively. The results provide strong evidence in favor of our continuous-time price discovery model irrespective of industry, listing venue, and trading activity. Indeed, we cannot reject the null hypothesis that the CS measures are the same across the different sampling intervals for almost all stocks. The continuous-time CS estimates are extremely similar, especially for the Nasdaq-listed stocks and for the less-liquid stocks.

Market leadership remains the same across the alternative sampling interval for virtually every stock in our sample. The listing exchange is the main driver for every stock, apart from JPM, VOD and WEN, regardless of the sampling interval. This effect is stronger for less-liquid stocks, whose average NYSE contribution to price discovery is approximately 88% across sampling intervals (in contrast to 71% and 74% for heavily traded Nasdaq- and NYSE-listed stocks, respectively). To conclude, we find strong statistical evidence supporting the continuous-time price discovery model put forward in the paper in that the CS measures are invariant to the sampling interval and these continuous-time price discovery measures can be inferred from discrete sampled prices.

In line with the simulations in Appendix B, we find that the standard errors generally increase with the sampling interval, reflecting that the noise-to-signal ratio increases with the sampling interval. In addition, we also find a substantial difference in the magnitude of the standard errors across heavily and thinly traded stocks. This finding suggests that the faster the adjustment, the higher the noise-to-signal ratio at larger sampling intervals. In particular, identification of the continuous-time CS measure is remarkably good for the thinly traded stocks up to the frequency of 1 observation per 5 minutes.

Table 3 reports the different estimates of the continuous-time IS measures. We document the median and standard errors (in brackets) of the daily continuous-time IS measures over the entire sample period. Test results now refer to the null hypothesis that $\overline{IS}(\widehat{\alpha}_{1/390,\perp}, \widehat{\Sigma}_{1/390,d}) = \overline{IS}(\widehat{\alpha}_{\delta,\perp}, \widehat{\Sigma}_{\delta,d})$ for $\delta \in \{1/195, 1/130, 1/97, 1/78\}$.

For virtually every heavily traded stock, we do not reject the null hypotheses that the IS measures are the same across the different sampling intervals. We find that their continuous-time IS measures are mostly equal to $1/2$, regardless of the listing exchange. This happens essentially because the estimates of the continuous-time correlation between markets are, on average, 0.96, 0.98, 0.99, 0.99, and 0.99 for the 1-, 2-, 3-, 4-, and 5-minute sampling intervals,

respectively.[9] This finding is consistent with our simulation study, whose results indicate the upward bias in the estimation of the continuous-time correlation increases with the sampling interval. The continuous-time IS estimates at the one-minute sampling interval seem reasonable, however. They indicate the same leading markets as the CS estimates, with the listing exchange again playing a major role. In addition, these estimates are also in line with Hasbrouck's (2018) estimates in higher resolutions for IBM and NVIDIA, despite the differences in sample period and sampling interval. For instance, Hasbrouck's (2018)'s average estimate of the NYSE information share for IBM at the 10-microsecond interval is 0.52, exactly the same as ours.

A different picture arises for thinly traded stocks. We reject the null hypothesis that the continuous-time IS measures using data at different sampling intervals are the same for all stocks except for CAJ and ROL. These stocks appear biased towards $1/2$ for sampling frequencies lower than one observation per minute. We believe that this result occurs mainly because of the bias in the estimation of the instantaneous correlation. The latter estimates monotonically increase from 0.61 to 0.86 as the sampling interval grows from 1 to 5 minutes. This finding is clearly in line with our simulation results in Appendix B, which document a very similar impact of the bias in the correlation estimation on the continuous-time IS measures. Nevertheless, we find that market leadership inference based on the IS estimates agrees with the indications given by the CS estimates, regardless of the sampling interval.

5 Conclusion

This paper entertains price discovery in a continuous-time setting. We first show that the component share measure of price discovery is invariant to the length of the discretization interval, allowing us to make inference on the continuous-time price discovery mechanism from discrete sampled prices. This finding is in contrast with Hasbrouck's (1995) information share, which depends on the contemporaneous correlation across markets, which naturally increases in magnitude as the sampling interval increases. As a result, the discrete-time IS measure converges to $1/k$ as the sampling interval increases at a rate that increases with the speed-of-adjustment parameter α. We nonetheless show how to recover the continuous-time counterpart of the IS measure, which is naturally immune to temporal aggregation.

A testable implication of our continuous-time price discovery model is that the continuous-

time CS and IS estimates should not differ across sampling intervals. We empirically test this implication for a panel of 30 stocks listed on NYSE and Nasdaq. Overall, we find that the continuous-time CS estimates are very stable across sampling intervals, whereas the continuous-time IS estimates seem invariant to the sampling interval only for the thickly traded stocks. Both price discovery measures nonetheless agree that the listing exchange typically leads the price discovery process.

Acknowledgments

References

Baillie, R. T., Booth, G. G., Tse, Y., Zabotina, T., 2002, Price discovery and common factor models, Journal of Financial Markets 5, 309–321.

Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A., Shephard, N., 2008, Designing realized kernels to measure the ex post variation of equity prices in the presence of noise, Econometrica 76, 1481–1536.

Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A., Shephard, N., 2009, Realized kernels in practice: Trades and quotes, Econometrics Journal 12, C1–C32.

Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A., Shephard, N., 2011, Multivariate realised kernels: Consistent positive semi-definite estimators of the covariation of equity prices with noise and non-synchronous trading, Journal of Econometrics 162, 149–169.

Barndorff-Nielsen, O. E., Shephard, N., 2004, Econometric analysis of realized covariation: High frequency based covariance, regression and correlation in financial economics, Econometrica 72, 885–925.

Bibinger, M., Hautsch, N., Malec, P., Reiss, M., 2019, Estimating the spot covariation of asset prices: Statistical theory and empirical evidence, Journal of Business and Economic Statistics 37, 419–435.

Booth, G. G., So, R. W., Tseh, Y., 1999, Price discovery in the German equity index derivatives markets, Journal of Futures Markets 19, 619–643.

Chu, Q. C., Hsieh, W. G., Tse, Y., 1999, Price discovery on the S&P 500 index markets: An analysis of spot index, index futures and SPDRs, International Review of Financial Analysis 8, 21–34.

Cochrane, J. H., 2012, Continuous-time linear models, Foundations and Trends in Finance 6, 165–219.

de Jong, F., 2002, Measures of contributions to price discovery: A comparison, Journal of Financial Markets 5, 323–327.

de Jong, F., Schotman, P. C., 2010, Price discovery in fragmented markets, Journal of Financial Econometrics 8, 1–28.

Dias, G. F., Scherrer, C., Papailias, F., 2016, Volatility discovery, Technical Report 07, CREATES Research Papers.

Fernandes, M., Scherrer, C. M., 2018, Price discovery in dual-class shares across multiple markets, Journal of Futures Markets 38, 129–155.

Figuerola-Ferretti, I., Gonzalo, J., 2010, Modelling and measuring price discovery in commodity markets, Journal of Econometrics 158, 95–107.

Gonçalves, S., Meddahi, N., 2009, Bootstrapping realized volatility, Econometrica 77, 283–306.

Gonzalo, J., Granger, C., 1995, Estimation of common long-memory components in cointegrated systems, Journal of Business and Economic Statistics 13, 27–35.

Gonzalo, J., Ng, S., 2001, A systematic framework for analyzing the dynamic effects of permanent and transitory shocks, Journal of Economics Dynamics and Control 25, 1527–1546.

Grammig, J., Melvin, M., Schlag, C., 2005, Internationally cross-listed stock prices during overlapping trading hours: Price discovery and exchange rate effects, Journal of Empirical Finance 12, 139–164.

Hansen, P. R., 2005, Granger's representation theorem: A closed-form expression for I(1) processes, Econometrics Journal 8, 23–38.

Hansen, P. R., Lunde, A., 2006, Realized variance and market microstructure noise, Journal of Business and Economic Statistics 24, 127–161.

Harris, F., McInish, T. H., Wood, R. A., 2002, Security price adjustment across exchanges: An investigation of common factor components for Dow stocks, Journal of Financial Markets 5, 277–308.

Hasbrouck, J., 1995, One security, many markets: Determining the contributions to price discovery, Journal of Finance 50, 1175–1198.

Hasbrouck, J., 2018, Price discovery in high resolution, Halbert White Jr. Memorial JFEC invited lecture, forthcoming at Journal of Financial Econometrics.

Huang, X., Tauchen, G., 2005, The relative contribution of jumps to total price variance, Journal of Financial Econometrics 3, 456–499.

Jacquier, E., Polson, N. G., Rossi, P. E., 1994, Bayesian analysis of stochastic volatility models, Journal of Business and Economic Statistics 12, 371–389.

Johansen, S., 1991, Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models, Econometrica 59, 1551–1580.

Johansen, S., 1995, Likelihood-Based Inference in Cointegrated Vector Autoregressive Models, Oxford University Press.

Kessler, M., Rahbek, A., 2001, Asymptotic likelihood based inference for co-integrated homogenous Gaussian diffusions, Scandinavian Journal of Statistics 28, 455–470.

Kessler, M., Rahbek, A., 2004, Identification and inference for multivariate cointegrated and ergodic Gaussian diffusions, Statistical Inference for Stochastic Processes 7, 137–151.

Lien, D., Shrestha, K., 2009, A new information share measure, Journal of Futures Markets 4, 377–395.

Lütkepohl, H., 2007, New Introduction to Multiple Time Series Analysis, Springer- Verlag.

Menkveld, A. J., 2014, High-frequency traders and market structure, Financial Review 49, 333–344.

Menkveld, A. J., 2016, The economics of high-frequency trading: Taking stock, Annual Review of Financial Economics 8, 1–24.

Nguenang, C., 2016, Evaluating market's contribution to price discovery in high-frequency for co-listed assets, working paper, Toulouse School of Economics.

O'Hara, M., 2015, High frequency market microstructure, Journal of Financial Economics 116, 257–270.

Pesaran, H., Shin, Y., 1998, Generalized impulse response analysis in linear multivariate models, Economics Letters 58, 17–29.

Stock, J. H., Watson, M. W., 1988, Testing for common trends, Journal of the American Statistical Association 83, 1097–1107.

Yan, B., Zivot, E., 2010, A structural analysis of price discovery measures, Journal of Financial Markets 13, 1–19.

Zu, Y., Boswijk, H. P., 2014, Estimating spot volatility with high-frequency financial data, Journal of Econometrics 181, 117 – 135.

Footnotes

[1] Specifically, Regulation ATS (alternative trading systems; RegATS) in 2000, the Regulation National Market System (Reg NMS) in 2007 in the US, and the Markets in Financial Instruments Directive (MiFiDin) in 2007 in Europe set the foundation for the existence of multiple trading venues linked together and competing for liquidity and trades.

[2] Nguenang (2016) assumes a baseline process similar to (1), but with a stochastic covariance matrix, to introduce a version of the information share measure based on Pesaran and Shin's (1998) generalized impulse response function in both continuous and discrete times. However, he does not examine how it changes once moving from continuous to discrete time.

[3] Alternatively, denoting by $sp(A)$ the subspace in $\mathbb{R}^k$ spanned by the columns of A, $A_\perp$ is any matrix with dimensions $k \times (k - r)$ such that $sp(A_\perp) = sp(A)_\perp$.

[4] Apart from closed-form expressions for the MA component, Hansen (2005) also provides an alternative proof for the GRT (see Theorem 1 and Corollary 2 herein).

[5] It is straightforward to consider k markets. Consider $\alpha_{\delta,\perp} = \alpha_\perp - \alpha (\alpha'_\delta \alpha)^{-1} \alpha'_\delta \alpha_\perp$, so that $\alpha'_\delta \alpha_{\delta,\perp} = 0$. It follows from $\alpha_\delta = \alpha \kappa_\delta$, with $\kappa_\delta = (\beta'\alpha)^{-1} [\exp(\delta\beta'\alpha) - I_r]$, that $\alpha_{\delta,\perp} = \alpha_\perp - \alpha (\alpha'\alpha)^{-1} \alpha' \alpha_\perp = \alpha_\perp$ for any $0 < \delta < 1$, given that $\alpha'\alpha_\perp = 0$.

[6] Alternatively, we could estimate the instantaneous covariance matrix using realized measures of spot variance and covariation (Bibinger, Hautsch, Malec and Reiss, 2019; Zu and Boswijk, 2014) to obtain intraday IS estimates. Given that the standard practice in the price discovery literature is to estimate daily VECMs (under the implicit assumption that the covariance matrix is constant within each day), we consider the realized covariance estimator of the integrated covariance matrix as a compromise.

[7] The only stock for which we find evidence of residual autocorrelation at the 5% significance level is ALGN, with a p-value of 0.03.

[8] The results of the cointegration analyses and specification diagnostics are available upon request.

[9] These very high values reflect that information becomes available very quickly to market participants in each market because markets are currently very fast and interconnected (O'Hara, 2015).

Table 1: Data description

We report summary statistics for raw and cleaned data for Nasdaq and NYSE/Arca. For the Nasdaq-listed stocks in the top panel, we report the number of Nasdaq and Arca observations, whereas we consider Nasdaq and NYSE for the heavily- and thinly-traded stocks listed at NYSE. In particular, we display the number of quotes (in millions) for each stock on the two trading venues before any cleaning filter (raw data) as well as after the implementation of the cleaning procedure (clean data). We also report the daily average number of quotes (in thousands) for both trading venues. The sample period spans from January 2013 to December 2013 and totals 251 trading days.

	raw ('000,000)		clean ('000,000)		obs per day ('000)	
	Nasdaq	NYSE/Arca	Nasdaq	NYSE/Arca	Nasdaq	NYSE/Arca
Nasdaq-listed stocks						
ADBE	28.66	9.08	2.65	2.01	10.54	7.99
ALGN	6.28	2.64	1.28	0.98	5.09	3.91
AMZN	12.29	12.75	2.31	2.15	9.20	8.58
CA	22.74	6.84	2.50	1.69	9.94	6.72
EXPE	14.78	5.93	1.99	1.65	7.93	6.58
GOOG	9.81	11.00	1.98	1.82	7.90	7.23
MU	59.96	15.42	3.21	2.70	12.78	10.75
SBUX	27.72	10.14	2.94	2.23	11.72	8.87
VOD	43.01	15.69	2.59	2.12	10.30	8.43
WEN	9.56	3.62	1.59	1.26	6.33	5.03
heavily-traded NYSE-listed stocks						
BAC	62.92	91.36	3.88	4.53	15.47	18.04
GE	47.01	72.42	3.37	4.22	13.41	16.83
HPQ	56.47	50.00	3.09	3.89	12.31	15.48
IBM	10.52	20.35	2.28	3.38	9.07	13.48
JCP	33.96	30.46	2.50	3.04	9.97	12.13
JPM	107.50	119.33	4.40	4.98	17.54	19.84
KO	46.92	42.14	3.18	3.81	12.68	15.17
MO	29.51	34.27	2.79	3.52	11.12	14.01
VZ	50.37	46.18	3.37	4.02	13.41	16.02
XOM	64.11	80.53	4.17	4.89	16.61	19.50
thinly-traded NYSE-listed stocks						
CAJ	3.21	2.88	0.87	0.89	3.47	3.56
COO	2.19	4.95	0.82	1.24	3.25	4.94
DLB	1.78	4.04	0.58	1.14	2.32	4.54
DNB	2.48	4.67	0.84	1.20	3.34	4.80
OFC	3.37	5.05	0.94	1.28	3.75	5.12
RBC	1.64	3.59	0.64	1.01	2.56	4.02
RE	1.73	4.65	0.71	1.26	2.81	5.04
RES	2.46	4.18	0.69	1.17	2.75	4.67
ROL	1.16	6.29	0.48	0.95	1.90	3.78
THO	3.32	5.27	0.97	1.22	3.85	4.86

Figure 1: Information share and contemporaneous correlation as sampling interval increases

The first plot displays the information and component shares of market 1 in continuous time and for δ ranging from $1/23,400$ (frequency of one observation per second) to $1/13$ (frequency of one observation per 30 minutes). The second plot depicts the exact correlation across markets at each sampling interval. We consider a range of values for the correlation across markets in continuous time, and then compute their discrete-time counterparts using the exact discretization of the reduced-rank OU process as in (3), with $\alpha = (0, 522)'$ so that $\alpha_{1/390} = (0, 0.74)'$ at the 1-minute sampling interval.
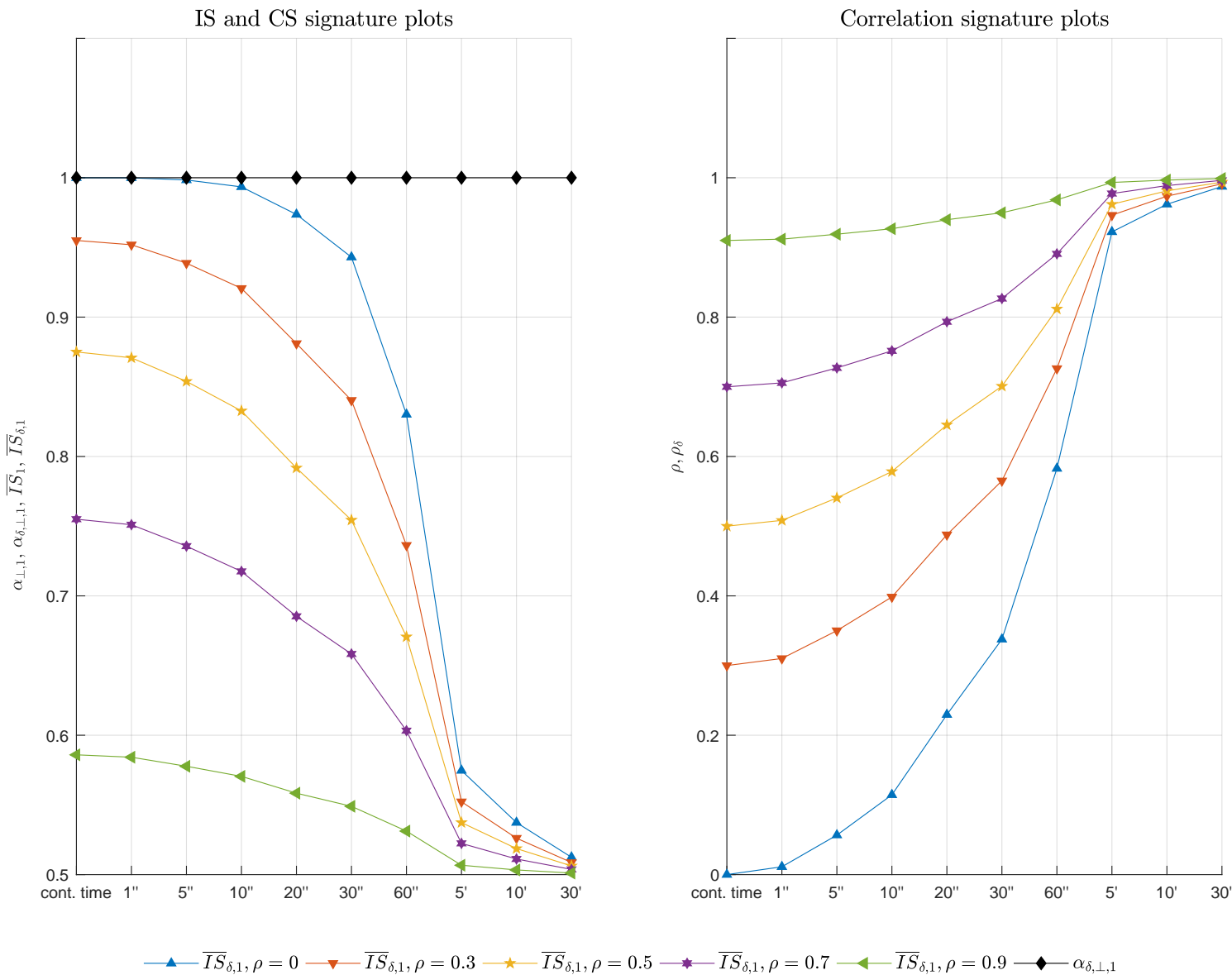
Table 2: Estimates of the continuous-time CS measures at alternative sampling intervals

We report the LS estimates and asymptotic standard errors (in brackets) of the continuous-time CS measure of Nasdaq computed from prices at alternative sampling intervals. Specifically, we denote $\widehat{\alpha}_{1/390,1,\perp}$, $\widehat{\alpha}_{1/195,1,\perp}$, $\widehat{\alpha}_{1/130,1,\perp}$, $\widehat{\alpha}_{1/97,1,\perp}$, $\widehat{\alpha}_{1/78,1,\perp}$ the estimates of the continuous-time CS measure of Nasdaq computed from prices sampled at one-, two-, three-, four-, and five-minutes intervals, respectively. The continuous-time CS measures are normalized such that the element wise estimates sum up to one, i.e., $\widehat{\alpha}_{\delta,\perp,1} + \widehat{\alpha}_{\delta,\perp,2} = 1$ with subscripts 1 and 2 denoting Nasdaq and NYSE/Arca, respectively. Finally, *, **, and *** denote rejections at the 10%, 5%, and 1% significance levels, respectively, of the bootstrap-based invariance test with a null hypothesis of $\alpha_{1/390,\perp,1} = \alpha_{\delta,\perp,1}$ for $\delta \in \{1/195, 1/130, 1/97, 1/78\}$.

	$\widehat{\alpha}_{1/390,1,\perp}$		$\widehat{\alpha}_{1/195,1,\perp}$		$\widehat{\alpha}_{1/130,1,\perp}$		$\widehat{\alpha}_{1/97,1,\perp}$		$\widehat{\alpha}_{1/78,1,\perp}$	
ADBE	0.67	(0.04)	0.73	(0.08)	0.73	(0.12)	0.78	(0.11)	0.83	(0.15)
ALGN	0.81	(0.08)	0.84	(0.11)	0.82	(0.06)	0.88	(0.09)	0.85	(0.10)
AMZN	0.70	(0.03)	0.70	(0.04)	0.71	(0.06)	0.66	(0.07)	0.67	(0.09)
CA	0.74	(0.06)	0.68	(0.11)	0.86	(0.15)	0.77	(0.16)	1.00**	(0.30)
EXPE	1.00	(0.14)	1.00	(0.21)	1.00	(0.29)	1.00	(0.42)	1.00	(0.39)
GOOG	0.62	(0.04)	0.62	(0.04)	0.65	(0.06)	0.61	(0.07)	0.70	(0.08)
MU	0.61	(0.11)	0.67	(0.12)	0.70	(0.14)	0.77	(0.20)	0.80	(0.17)
SBUX	0.77	(0.05)	0.73	(0.07)	0.75	(0.08)	0.60*	(0.10)	0.78	(0.10)
VOD	0.31	(0.19)	0.66**	(0.12)	0.41	(0.18)	0.38	(0.24)	0.75*	(0.19)
WEN	0.37	(0.14)	0.42	(0.15)	0.37	(0.17)	0.33	(0.15)	0.46	(0.15)
BAC	0.20	(0.09)	0.15	(0.14)	0.00	(0.21)	0.00	(0.27)	0.16	(0.29)
GE	0.18	(0.09)	0.13	(0.14)	0.18	(0.18)	0.25	(0.21)	0.32	(0.26)
HPQ	0.29	(0.09)	0.18	(0.15)	0.00*	(0.18)	0.30	(0.21)	0.00***	(0.24)
IBM	0.32	(0.13)	0.30	(0.05)	0.46*	(0.21)	0.22	(0.08)	0.24	(0.08)
JCP	0.43	(0.09)	0.48	(0.14)	0.31	(0.17)	0.41	(0.25)	0.31	(0.23)
JPM	0.55	(0.10)	0.62	(0.16)	0.60	(0.22)	0.33	(0.28)	0.40	(0.34)
KO	0.44	(0.13)	0.08	(0.12)	0.24	(0.28)	0.00**	(0.14)	0.03	(0.17)
MO	0.45	(0.09)	0.40	(0.12)	0.39	(0.16)	0.28	(0.20)	0.29	(0.24)
VZ	0.23	(0.10)	0.06	(0.10)	0.38	(0.14)	0.23	(0.12)	0.16	(0.12)
XOM	0.33	(0.07)	0.15*	(0.14)	0.07**	(0.18)	0.26	(0.22)	0.09	(0.26)
CAJ	0.29	(0.02)	0.32	(0.04)	0.33	(0.04)	0.29	(0.05)	0.32	(0.05)
COO	0.07	(0.01)	0.04**	(0.01)	0.06	(0.02)	0.05	(0.02)	0.02***	(0.02)
DLB	0.10	(0.01)	0.12	(0.02)	0.11	(0.02)	0.10	(0.03)	0.16**	(0.03)
DNB	0.16	(0.02)	0.17	(0.03)	0.18	(0.03)	0.14	(0.04)	0.20	(0.04)
OFC	0.09	(0.01)	0.10	(0.02)	0.14*	(0.03)	0.07	(0.03)	0.09	(0.04)
RBC	0.04	(0.01)	0.04	(0.02)	0.05	(0.01)	0.03	(0.02)	0.06	(0.02)
RE	0.05	(0.01)	0.05	(0.01)	0.06	(0.02)	0.04	(0.02)	0.04	(0.03)
RES	0.19	(0.02)	0.18	(0.03)	0.18	(0.04)	0.22	(0.03)	0.21	(0.05)
ROL	0.07	(0.01)	0.06	(0.01)	0.06	(0.02)	0.05	(0.02)	0.06	(0.02)
THO	0.13	(0.01)	0.13	(0.02)	0.12	(0.02)	0.12	(0.03)	0.12	(0.02)

Table 3: Estimates of the continuous-time IS measures at alternative sampling intervals

We report the median of the daily estimates and their standard errors (in brackets) of the continuous-time IS measure of Nasdaq computed from prices at alternative sampling intervals. Specifically, we denote $\overline{IS}(\widehat{\alpha}_{1/390,\perp}, \widehat{\Sigma}_{1/390,d})_1$, $\overline{IS}(\widehat{\alpha}_{1/195,\perp}, \widehat{\Sigma}_{1/195,d})_1$, $\overline{IS}(\widehat{\alpha}_{1/130,\perp}, \widehat{\Sigma}_{1/130,d})_1$, $\overline{IS}(\widehat{\alpha}_{1/97,\perp}, \widehat{\Sigma}_{1/97,d})_1$, $\overline{IS}(\widehat{\alpha}_{1/78,\perp}, \widehat{\Sigma}_{1/78,d})_1$ the estimates of the continuous-time IS measure of Nasdaq computed from prices sampled at one-, two-, three-, four-, and five-minutes intervals, respectively. The continuous-time IS measures sum up to one, i.e., $\overline{IS}(\widehat{\alpha}_{\delta,\perp}, \widehat{\Sigma}_{\delta,d})_1 + \widehat{\alpha}_{\delta,\perp}, \widehat{\Sigma}_{\delta,d})_2 = 1$ with subscripts 1 and 2 denoting Nasdaq and NYSE/Arca, respectively. Finally, *, **, and *** denote rejections at the 10%, 5%, and 1% significance levels, respectively, of the bootstrap-based invariance test with a null hypothesis of $\overline{IS}(\widehat{\alpha}_{1/390,\perp}, \widehat{\Sigma}_{1/390,d})_1 = \overline{IS}(\widehat{\alpha}_{\delta,\perp}, \widehat{\Sigma}_{\delta,d})_1$ for $\delta \in \{1/195, 1/130, 1/97, 1/78\}$.

	$\overline{IS}(\widehat{\alpha}_{1/390,\perp}, \widehat{\Sigma}_{1/390,d})_1$		$\overline{IS}(\widehat{\alpha}_{1/195,\perp}, \widehat{\Sigma}_{1/195,d})_1$		$\overline{IS}(\widehat{\alpha}_{1/130,\perp}, \widehat{\Sigma}_{1/130,d})_1$		$\overline{IS}(\widehat{\alpha}_{1/97,\perp}, \widehat{\Sigma}_{1/97,d})_1$		$\overline{IS}(\widehat{\alpha}_{1/78,\perp}, \widehat{\Sigma}_{1/78,d})_1$	
ADBE	0.51	(0.01)	0.51	(0.01)	0.50	(0.01)	0.50	(0.01)	0.50	(0.01)
ALGN	0.63	(0.06)	0.59	(0.06)	0.56**	(0.05)	0.55***	(0.06)	0.54**	(0.06)
AMZN	0.53	(0.02)	0.51	(0.02)	0.51	(0.01)	0.51	(0.02)	0.51	(0.01)
CA	0.51	(0.02)	0.51	(0.01)	0.51	(0.02)	0.50	(0.01)	0.51	(0.02)
EXPE	0.56	(0.04)	0.53	(0.03)	0.52	(0.03)	0.52	(0.02)	0.51	(0.03)
GOOG	0.52	(0.01)	0.51	(0.01)	0.51	(0.01)	0.51	(0.01)	0.51	(0.01)
MU	0.50	(0.01)	0.50	(0.01)	0.50	(0.01)	0.50	(0.02)	0.50	(0.02)
SBUX	0.52	(0.02)	0.51	(0.01)	0.51	(0.01)	0.50*	(0.00)	0.50	(0.01)
VOD	0.49	(0.02)	0.50**	(0.01)	0.50	(0.00)	0.50	(0.00)	0.50	(0.01)
WEN	0.49	(0.02)	0.50	(0.02)	0.50	(0.02)	0.50	(0.03)	0.50	(0.03)
BAC	0.49	(0.01)	0.50	(0.01)	0.50*	(0.01)	0.50*	(0.01)	0.50	(0.01)
GE	0.49	(0.01)	0.50	(0.01)	0.50	(0.01)	0.50	(0.01)	0.50	(0.00)
HPQ	0.50	(0.00)	0.50	(0.01)	0.50	(0.01)	0.50	(0.00)	0.50	(0.01)
IBM	0.48	(0.01)	0.49	(0.01)	0.50**	(0.01)	0.49	(0.01)	0.49	(0.01)
JCP	0.50	(0.00)	0.50	(0.00)	0.50	(0.00)	0.50	(0.00)	0.50	(0.00)
JPM	0.50	(0.00)	0.50	(0.00)	0.50	(0.00)	0.50	(0.00)	0.50	(0.00)
KO	0.50	(0.00)	0.50	(0.01)	0.50	(0.01)	0.50	(0.02)	0.50	(0.02)
MO	0.50	(0.00)	0.50	(0.00)	0.50	(0.00)	0.50	(0.00)	0.50	(0.00)
VZ	0.49	(0.01)	0.50	(0.03)	0.50	(0.00)	0.50	(0.01)	0.50	(0.02)
XOM	0.50	(0.00)	0.50	(0.01)	0.50	(0.01)	0.50	(0.00)	0.50	(0.01)
CAJ	0.38	(0.05)	0.42	(0.05)	0.45	(0.04)	0.44	(0.05)	0.46	(0.04)
COO	0.18	(0.10)	0.23	(0.12)	0.31***	(0.12)	0.32***	(0.13)	0.37***	(0.13)
DLB	0.27	(0.07)	0.33***	(0.07)	0.36***	(0.08)	0.39***	(0.07)	0.42***	(0.06)
DNB	0.32	(0.06)	0.38***	(0.05)	0.42***	(0.05)	0.43***	(0.06)	0.44***	(0.04)
OFC	0.29	(0.08)	0.37***	(0.08)	0.42***	(0.06)	0.42***	(0.08)	0.44***	(0.08)
RBC	0.20	(0.10)	0.27	(0.12)	0.34***	(0.12)	0.36***	(0.13)	0.40***	(0.12)
RE	0.22	(0.09)	0.30**	(0.10)	0.36***	(0.11)	0.37***	(0.11)	0.39***	(0.11)
RES	0.37	(0.06)	0.41*	(0.06)	0.43***	(0.04)	0.45***	(0.04)	0.46***	(0.03)
ROL	0.12	(0.07)	0.15	(0.09)	0.19	(0.10)	0.20	(0.10)	0.24***	(0.10)
THO	0.27	(0.07)	0.32*	(0.07)	0.37***	(0.08)	0.39***	(0.08)	0.41***	(0.07)

Appendices

A Bootstrap algorithm

This section details how we construct bootstrap-based confidence intervals to test for the equality of the continuous-time price discovery measures across sampling intervals. See, for instance, Appendix D in Lütkepohl (2007) for more details on resampling methods.

The steps of the bootstrap algorithm are as follows.

(i) Estimate the daily integrated covariance matrix $\Sigma_d = \int_d^{d+1} \exp(u\Pi)\Sigma_u \exp(u\Pi') \, du$ using the realized covariance estimator based on intraday prices at the 1-minute sampling interval. Denote these estimates as $\widehat{\Sigma}_{1/390,d}$, with $d = 1, \ldots, D$.

(ii) After estimating $\alpha_{1/390}$ by least squares, compute the residuals $\widehat{\varepsilon}_\tau$ for $\tau = 1, \ldots, T$ and their sample covariance matrix $\widehat{\Sigma}_{1/390}$.

(iii) Inverting the matrix exponential operator in the exact discretization of (1) evaluated at $\widehat{\alpha}_{1/390}$ yields $\widehat{\Pi} = \widehat{\alpha}\beta' = 390 \log\left(\widehat{\alpha}_{1/390}\, \beta' + I_k\right)$, where $Z = \log(A)$ if Z is such that $\exp(Z) = A$ for any square matrix A. The first column of $\widehat{\Pi}$ then yields $\widehat{\alpha}$.

(iv) Using the exact discretization of (1) and estimates obtained from prices sampled at the 1-minute sampling interval, compute $\alpha_\delta(\widehat{\alpha}) = \widehat{\alpha}(\beta'\widehat{\alpha})^{-1} \left[\exp(\delta\beta'\widehat{\alpha}) - I_r\right]$, and $\widehat{\Sigma}_{\delta,d}(\widehat{\alpha}, \widehat{\Sigma}_{1/390,d}) = \int_0^\delta \exp(u\widehat{\Pi})\widehat{\Sigma}_{1/390,d} \exp(u\widehat{\Pi})' \, du$ for $\delta \in \{1/195, 1/130, 1/97, 1/78\}$ and $d = 1, \ldots, D$.

(v) Demean the residuals to obtain $\widehat{\varepsilon}_\tau - \bar{\varepsilon}_T$, where $\bar{\varepsilon}_T$ denotes the sample average of $\widehat{\varepsilon}_\tau$ over time.

(vii) Standardize the demeaned residuals to obtain $\widetilde{\varepsilon}_\tau = \widehat{\Sigma}_{1/390,d}^{-1/2}(\widehat{\alpha}, \widehat{\Sigma}_{1/390,d})\, (\widehat{\varepsilon}_\tau - \bar{\varepsilon}_T)$ for each $d = 1, \ldots, D$.

(vi) Draw randomly with replacement from the standardized centered residuals $\{\widetilde{\varepsilon}_\tau;\, \tau = 1, \ldots, T\}$ to form B bootstrap samples $\left\{\varepsilon_\tau^{(b)};\, \tau = 1, \ldots, T\right\}$, with $b = 1, \ldots, B$.

(viii) Compute recursively bootstrap samples of the k-dimensional price vector as

$$P_{\delta,\tau}^{(b)} = \alpha_\delta(\widehat{\alpha})\beta' P_{\delta,\tau-1}^{(b)} + \widehat{\Sigma}_{\delta,d}^{1/2}(\widehat{\alpha}, \widehat{\Sigma}_{1/390,d})\varepsilon_\tau^{(b)}, \qquad \text{with} \quad P_{\delta,0} = P_0,$$

for $\delta \in \{1/390, 1/195, 1/130, 1/97, 1/78\}$, $\tau = 1, \ldots, T$, $d = 1, \ldots, D$, and $b = 1, \ldots, B$.

(ix) Compute the continuous-time CS and IS measures for each bootstrap sample at the different sampling intervals. For the daily IS measures, compute their median value over the sample period.

(x) Compute the bootstrap-based confidence intervals using the $p/2$- and $(1-p)/2$-quantiles of the B continuous-time CS and IS estimates, with $p \in \{1\%, 5\%, 10\%\}$.

B Simulation study

Let prices follow

$$\mathrm{d}P_t = \Pi P_t \, \mathrm{d}t + C_t \, \mathrm{d}W_t, \qquad \text{with} \quad P_0 = p_0, \tag{B.1}$$

where C_t is a cadlag stochastic volatility process that evolves intradaily such that the covariance matrix $\Sigma_t = C_t C_t'$ is positive definite. Although the exact discretization in Section 2 no longer holds, the price process remains Markovian, with a discrete-time solution given by

$$P_{t+\delta} = \exp(\delta\Pi)\left[P_t + \int_0^{\delta} \exp(-u\Pi)C_u \, \mathrm{d}W_u\right]. \tag{B.2}$$

Given that allowing for stochastic covariation does not affect the autoregressive matrix $\exp(\delta\Pi)$, CS continues to remain invariant to the sampling interval such that $\alpha_\perp = \alpha_{\delta,\perp}$ for any $0 < \delta < 1$. In contrast, the exact discretization of the IS measure depends on the specification of the stochastic process that governs the dynamics of the instantaneous covariance matrix.

We next illustrate how the price discovery measures change with the sampling interval by simulating the price process in (B.1), with $k = 2$ trading venues, $\alpha = (0, 522)'$ and $\beta' = (1, -1)$ as in (16). As for C_t, we assume that the instantaneous correlation is constant and takes values $\rho \in \{0, 0.3, 0.5, 0.7, 0.9\}$, whereas the market-specific variances follow a single-factor stochastic volatility (SV1F) process as in, e.g., Huang and Tauchen (2005), Gonçalves and Meddahi (2009) and Barndorff-Nielsen, Hansen, Lunde and Shephard (2008). We let $\sigma_{m,t}^2 = \exp(\varsigma_0 + \varsigma_1 V_{m,t})$, with $\mathrm{d}V_{m,t} = \gamma_V V_{m,t} \, \mathrm{d}t + \mathrm{d}B_{m,t}$ for $m = 1, 2$. In addition, we consider that $\mathrm{corr}(\mathrm{d}W_t, \mathrm{d}B_{m,t}) = \nu_W$ for $m = 1, 2$ and that $\mathrm{corr}(\mathrm{d}B_{1,t}, \mathrm{d}B_{2,t}) = \nu_B$, where ν_W and ν_B are (possibly nonzero) constants. Note that $(\varsigma_0, \varsigma_1, \gamma_V, \nu_W)$ are the same for both markets, so that their variance processes exhibit the same time series properties and unconditional distributions.

We set $\varsigma_1 = 0.125$, $\gamma_V = -0.025$, and $\nu_W = -0.30$ as in Barndorff-Nielsen et al. (2008), whereas we impose $\varsigma_0 = \varsigma_1^2/(2\gamma_V)$, so that the instantaneous variances integrate to one in both

markets. In addition, we fix ν_B to 0.95. Overall, this specification is convenient because it makes the results directly comparable to the ones in Figure 1, given that the IS is essentially driven by the amount of correlation if the market-specific variances do not differ considerably. We run simulations using an iterative method based on an Euler scheme that accounts for the exact discretization of the stochastic volatility process to obtain prices at the 1-second sampling interval. We set $n = 23,400$ to match the number of seconds on a trading day of 6.5 hours.

For each replication, we estimate the speed-of-adjustment parameters by least squares and then compute the corresponding CS measures at the 1-, 5-, 10-, 30-, 60-, 300-, 600-, and 1,800-second intervals, corresponding to samples of 23,400, 4,680, 2,340, 1,170, 780, 390, 78, 39, and 13 observations, respectively. We compute the continuous-time IS measure using realized covariances and $\alpha_\perp$ estimates at the highest frequency (i.e., one observation per second). Finally, for the sake of comparison, we also estimate the traditional discrete-time IS measures based on the sample covariance matrix as if it did not change over time.

Figure B.1 displays the box plots of the CS and IS estimates as well as those of the contemporaneous correlation between markets at the different sampling intervals across 50,000 replications. The patterns are very similar to those in Figure 1, confirming that the CS measures are invariant to the sampling interval, with their median values coinciding with the continuous-time CS, regardless of the contemporaneous correlation between markets. The median IS estimates are very close to the IS measures that the exact discretization of (16) would imply in that they still converge to $1/2$ as the sampling interval increases due to the increase in the discrete-time contemporaneous correlation.

Next, we assess the finite sample performance of the estimates of the continuous-time CS and IS measures at increasing sampling intervals for different timespans. As before, we simulate prices on $k = 2$ trading venues according to the same specification of (B.1) at the 1-second sampling interval for $D \in \{1, 22, 252\}$ trading days and then sample them at fixed intervals of 1 to 5 minutes.[10] To fully reflect the empirical environment of markets operating at extremely fast time frames, we fix the continuous-time speed-of-adjustment parameters as the median estimate for the heavily traded NYSE-listed stocks in Section 4 adjusted so that the first market impounds all the information to the efficient price: $\alpha = (0, 96)'$ or, equivalently, $\alpha_{1/390} = (0, 0.22)'$ at the 1-minute sampling interval.

To accommodate multiple trading days, we allow the daily integrated variances to follow

$$\ln \sigma^2_{m,d} = \phi_0 + \phi_1 \ln \sigma^2_{m,d-1} + \varsigma \upsilon_{m,d}, \qquad \text{for} \quad m = 1, 2 \quad \text{and} \quad d = 1, \dots, D \tag{B.3}$$

with $\sigma^2_{m,d}$ denoting the market-specific daily integrated variances. The volatility innovations $\upsilon_{m,d}$ are Gaussian white noises with a constant correlation of 0.95 and unit variances. As in Jacquier, Polson and Rossi (1994), we fix the autoregressive parameter to 0.98 and calibrate ϕ_0 and ς in (B.3) such that the expected annual volatility is 20% and the coefficient of variation given by $\mathbb{V}(\sigma^2_{m,d})/\mathbb{E}(\sigma^2_{m,d}) = \exp\left(\varsigma/(1 - \phi_1^2)\right) - 1$ is equal to 1/2. We then allow the instantaneous market-specific variances integrate to $\sigma^2_{m,d}$ with $m = 1, 2$.

We estimate α_δ for $\delta \in \{1/390, 1/195, 1/130, 1/97, 1/78\}$ by LS using every day in the sample and then compute the (continuous-time) CS estimates. By construction, as we normalize the elements of $\widehat{\alpha}_{\delta,\perp}$ to sum up to one, their bias and standard errors have exactly the same magnitude. As for the IS measures, we compute daily estimates of continuous-time IS measures using $\widehat{\alpha}_{\delta,\perp}$ and $\widehat{\Sigma}_{\delta,d}$ for $\delta \in \{1/390, 1/195, 1/130, 1/97, 1/78\}$.

Tables B.1 and B.2 document the bias and root mean squared error (RMSE) of the CS and IS measures over 1,000 replications. The sample biases of the CS estimates are very close to zero in magnitude, regardless of the sampling interval. As expected, the magnitude of the bias decreases uniformly as the number of days D increases, whereas it seems to increase with the amount of correlation between markets and with the sampling interval. This finding suggests that it becomes more difficult to estimate the speed-of-adjustment parameters from prices sampled at lower frequencies, especially if correlation is high, due to a higher noise-to-signal ratio.

Table B.1 also reveals that there is a sharp decline in the RMSE as D increases. For $D = 1$, the RMSE is uniformly large over the correlation value and alternative sampling intervals because estimating daily VECM models becomes more difficult not only at lower frequencies (due to the smaller sample sizes) but also for highly correlated markets (due to the lesser amount of information). The improvement in the RMSE as the timespan increases to $D = 22$ or $D = 252$ indicates that the reliable estimation of very high speed-of-adjustment parameters, which characterize low noise-to-signal ratios, call for a substantially large sample size. Furthermore, we find that the RMSE increases with the sampling interval because the identification of the speed-of-adjustment parameters weakens as the sampling interval increases and, hence, the LS

estimates become noisier. Overall, our simulation results indicate that data for one trading year suffice to ensure reliable CS estimates.

Table B.2 displays the results for the continuous-time IS estimates based on prices sampled at increasing sampling intervals. Unlike the CS measures, the IS measures are severely biased for prices sampled at frequencies lower than one observation per minute. As expected, the realized variances $\widehat{\sigma}_{m,d}^2$ and $\widehat{\alpha}_{\delta,\perp}$ are unbiased, implying that the upward bias in the estimation of the constant correlation between markets biases the continuous-time IS estimates towards $1/2$. Moreover, the bias increases with the sampling interval, meaning that the drift contaminates the identification of the instantaneous correlation coefficient as the sampling interval increases. Finally, the RMSE figures are driven mainly by the magnitude of the bias, given the relatively small standard errors.

Figure B.1: Information share and contemporaneous correlation as sampling interval increases with stochastic covariance matrix

The first plot in the upper panel displays the theoretical information share based on the exact discretization of (1) and the box plots of the estimates of the IS measures at the different sample intervals based on the simulations of (B.1). The second plot on the upper panel portrays the exact correlation across markets in continuous time and the box plots of the corresponding estimates at the different sampling intervals. Finally, the bottom panel plots depict the theoretical continuous-time CS measures (solid markers in black) and the box plot of the estimates obtained at the different sampling intervals from the simulations. As for the continuous-time parameters, we set $\alpha = (0, 522)'$ so that $\alpha_{1/390} = (0, 0.74)'$ at the 1-minute sampling interval. The edges of boxes in the box plots refer to the 25% and 75% percentiles, whereas the black dot inside a white circle represents the median.
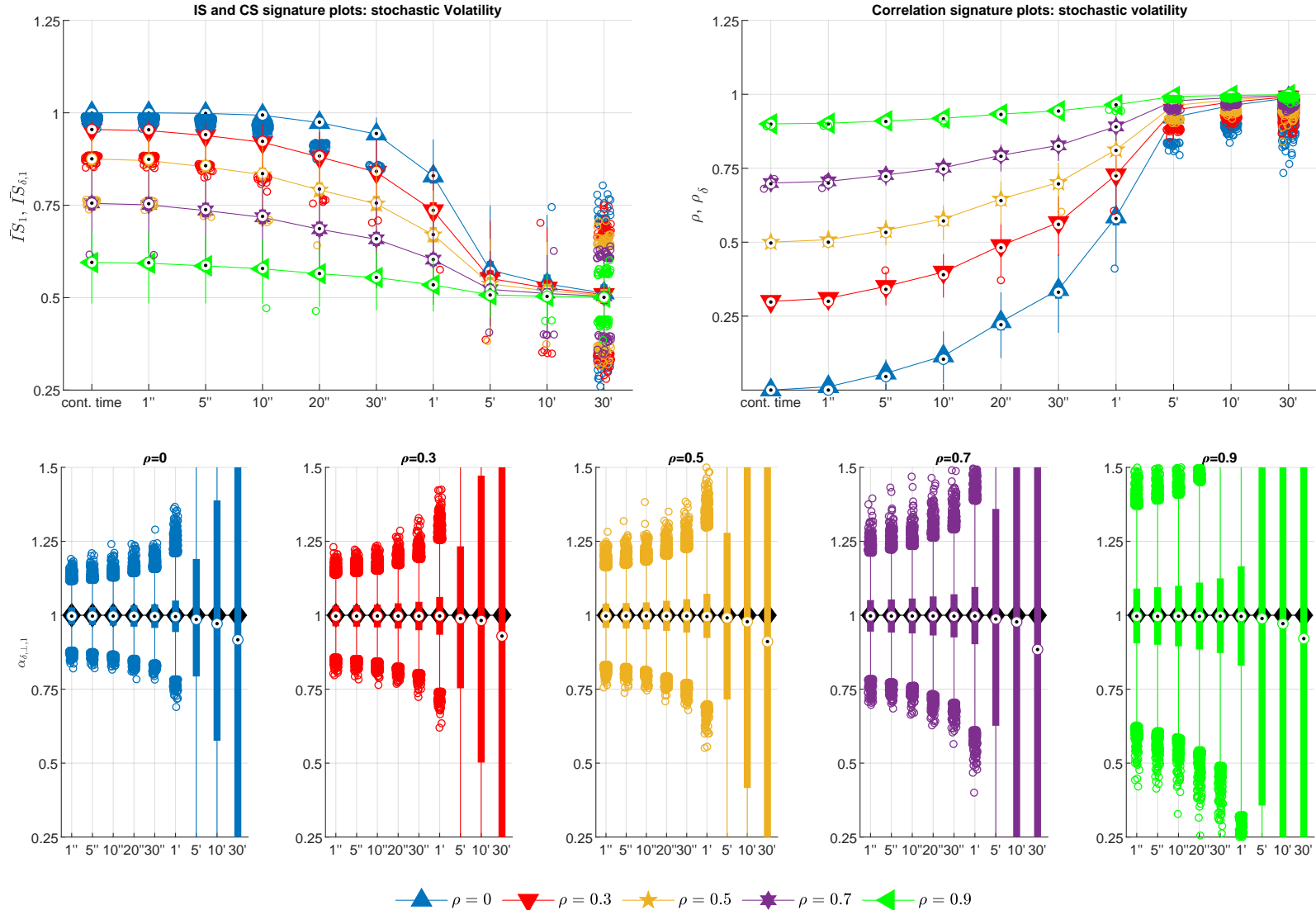
Table B.1: Estimation performance of component share measure

We report the bias and root mean squared error (RMSE) for the continuous-time information and component share measures for sample sizes of $D = 1$, $D = 22$, and $D = 252$ days over 1,000 replications. Estimates of the continuous-time information share measure for market one, $\widehat{IS}_1$, are constructed by combining the estimates of the integrated variance and $\alpha_\perp$ obtained with the realized variance (RV) and LS estimators, respectively, from prices sampled at one observation per minute. Estimates of the component share measure of market one, $\alpha_{\delta,\perp,1}$, are computed with the LS estimator from prices at sampling frequencies ranging from one observation per 1 minute to one observation per 5 minutes: $\delta \in \{1/390, 1/195, 1/130, 1/97, 1/78\}$. The instantaneous correlation between markets ρ ranges from 0 to 0.90 and the diagonal elements of the continuous-time covariance matrix follow a single-factor stochastic volatility (SV1F).

	ρ	$100 \times$ bias$(\widehat{\alpha}_{\perp,1})$					RMSE$(\widehat{\alpha}_{\perp,1})$				
		$\delta = 1/390$	$\delta = 1/195$	$\delta = 1/130$	$\delta = 1/97$	$\delta = 1/78$	$\delta = 1/390$	$\delta = 1/195$	$\delta = 1/130$	$\delta = 1/97$	$\delta = 1/78$
$D = 1$	0.00	1.28	1.50	1.56	0.67	1.23	0.11	0.13	0.14	0.16	0.18
	0.30	1.36	1.24	1.67	1.91	2.03	0.14	0.16	0.18	0.20	0.22
	0.50	1.44	1.75	1.76	2.35	2.01	0.16	0.18	0.20	0.23	0.25
	0.70	0.53	0.09	0.34	-0.52	0.70	0.21	0.24	0.26	0.30	0.35
	0.90	-0.17	-0.66	0.66	2.02	3.52	0.38	0.41	0.47	0.54	0.58
$D = 22$	0.00	0.00	0.01	-0.01	0.02	-0.03	0.02	0.03	0.03	0.03	0.04
	0.30	0.06	0.00	0.05	-0.02	-0.02	0.03	0.03	0.04	0.04	0.05
	0.50	-0.03	0.01	-0.05	0.04	0.09	0.03	0.04	0.04	0.05	0.06
	0.70	-0.06	-0.05	-0.11	0.04	-0.26	0.05	0.05	0.06	0.06	0.07
	0.90	0.12	0.10	-0.42	-0.05	-0.18	0.08	0.08	0.10	0.11	0.12
$D = 252$	0.00	-0.01	-0.02	-0.02	0.01	-0.04	0.01	0.01	0.01	0.01	0.01
	0.30	0.02	0.01	0.00	0.03	0.01	0.01	0.01	0.01	0.01	0.01
	0.50	0.02	0.01	-0.01	0.01	-0.01	0.01	0.01	0.01	0.01	0.02
	0.70	0.00	-0.01	0.00	0.04	-0.08	0.01	0.02	0.02	0.02	0.02
	0.90	0.03	0.05	0.00	0.11	0.06	0.02	0.03	0.03	0.03	0.04

Table B.2: Estimation performance of information measure

We report the bias and root mean squared error (RMSE) for the continuous-time information and component share measures for sample sizes of $D = 1$, $D = 22$, and $D = 252$ days over 1,000 replications. Estimates of the continuous-time information share measure for market one, $\widehat{IS}_1$, are constructed by combining the estimates of the integrated variance and $\alpha_\perp$ obtained with the realized variance (RV) and LS estimators, respectively, from prices sampled at one observation per minute. Estimates of the component share measure of market one, $\alpha_{\delta,\perp,1}$, are computed with the LS estimator from prices at sampling frequencies ranging from one observation per 1 minute to one observation per 5 minutes: $\delta \in \{1/390, 1/195, 1/130, 1/97, 1/78\}$. The instantaneous correlation between markets ρ ranges from 0 to 0.90 and the diagonal elements of the continuous-time covariance matrix follow a single-factor stochastic volatility (SV1F).

	ρ	$100 \times \text{bias}(\widehat{IS}_1)$					$\text{RMSE}(\widehat{IS}_1)$				
		$\delta = 1/390$	$\delta = 1/195$	$\delta = 1/130$	$\delta = 1/97$	$\delta = 1/78$	$\delta = 1/390$	$\delta = 1/195$	$\delta = 1/130$	$\delta = 1/97$	$\delta = 1/78$
$D = 1$	0.00	1.91	3.70	5.85	8.49	10.97	0.03	0.05	0.07	0.10	0.12
	0.30	4.35	7.39	10.45	12.99	15.35	0.07	0.09	0.12	0.15	0.17
	0.50	4.39	7.23	9.97	12.21	14.36	0.08	0.10	0.12	0.14	0.16
	0.70	3.83	6.00	7.78	9.79	11.07	0.08	0.09	0.11	0.12	0.13
	0.90	1.50	2.37	3.15	3.91	4.53	0.06	0.06	0.07	0.07	0.07
$D = 22$	0.00	0.80	2.46	4.63	7.02	9.43	0.01	0.03	0.05	0.08	0.10
	0.30	2.79	5.63	8.41	11.06	13.51	0.03	0.06	0.09	0.12	0.14
	0.50	2.92	5.76	8.38	10.74	12.89	0.04	0.07	0.09	0.12	0.14
	0.70	2.36	4.54	6.51	8.14	9.65	0.03	0.05	0.07	0.09	0.10
	0.90	1.05	1.90	2.68	3.28	3.87	0.02	0.02	0.03	0.04	0.04
$D = 252$	0.00	0.75	2.41	4.56	6.96	9.37	0.01	0.03	0.05	0.08	0.10
	0.30	2.72	5.57	8.35	10.98	13.45	0.03	0.06	0.09	0.12	0.14
	0.50	2.97	5.81	8.41	10.75	12.86	0.04	0.06	0.09	0.11	0.14
	0.70	2.37	4.54	6.46	8.14	9.63	0.03	0.05	0.07	0.09	0.10
	0.90	0.99	1.88	2.63	3.28	3.87	0.01	0.02	0.03	0.04	0.04