School of Economics Working Paper
2024-04

University of East Anglia

SCHOOL OF ECONOMICS

Integrated Variance Estimation for Assets Traded in Multiple Venues

Gustavo Fruet Dias *
Karsten Schweiker **

* University of East Anglia
** University of Hohenheim

School of Economics
University of East Anglia
Norwich Research Park
Norwich NR4 7TJ
United Kingdom
www.uea.ac.uk/economics

Integrated Variance Estimation for Assets Traded in Multiple Venues

Gustavo Fruet Dias Karsten Schweikert*

University of East Anglia & CREATES University of Hohenheim

[Latest update: April 1, 2024]

Abstract

In this paper, we identify a novel form of multiplicative market microstructure noise, referred to as *fragmentation noise*, which arises when the same asset is traded across multiple venues. We demonstrate that conventional estimators, such as realized variance and other well-established noise-robust methods, yield inconsistent estimates in the presence of fragmentation noise. To address this estimation issue, we propose a two-step estimator. In the first step, we model prices in different trading venues using a vector error correction model, leveraging its common trend representation to estimate the efficient price of the asset. In the second step, we compute the realized variance estimator using the estimates of the efficient price. We derive the asymptotic distribution of our proposed two-step estimator and conduct comprehensive simulation experiments. An application to the constituents of the DJIA reveals that our two-step estimator outperforms or performs on par with the univariate estimators under consideration.

Keywords: High-frequency data, Ornstein-Uhlenbeck process, Cointegration, Realized variance, Realized kernel estimators, Market microstructure, Price discovery
JEL Classification: C12; C15; G14

*Address: University of Hohenheim, Core Facility Hohenheim & Institute of Economics, Schloss Hohenheim 1 C, 70593 Stuttgart, Germany, e-mail: *karsten.schweikert@uni-hohenheim.de*

1 Introduction

The availability of high-frequency financial data has led to a remarkable increase in research activity in the field of return volatility estimation (see, among others, Andersen et al., 2001, 2003; Barndorff-Nielsen et al., 2008). The current state-of-the-art approaches for estimating the integrated variance rely on univariate estimators that account for (dependent) market microstructure noise, jumps, and other features of high-frequency financial data (see, e.g. the existing surveys by McAleer and Medeiros, 2008; Aït-Sahalia and Jacod, 2014). However, a feature that has received little attention in this literature is the increasing fragmentation of markets. For example, equities in the U.S. are traded not only on their listing exchanges but also on a number of alternative exchanges and trading platforms. Despite the law-of-one-price principle stating that prices for the same asset should be tied to the asset's latent efficient price, deviations between prices on different trading venues are prevalent at high frequencies. This is mainly due to the differences in market microstructure such as trading venues' levels of transparency, trading clientele, pricing models, liquidity and market design (see, e.g., O'Hara, 2015; Menkveld, 2016). These factors ultimately reflect on the speed at which markets impound information to the latent efficient price and the properties of the market microstructure noise in high-frequency prices (see, e.g., Hasbrouck, 1995, 2003). In this research, we investigate the effect of market fragmentation on the estimation of the integrated variance.

It is well-known that if the efficient asset price is perfectly observed and follows a continuous semimartingale process, then its integrated variance can be straightforwardly estimated using the sum of squared intra-day returns: the realized variance (RV) estimator (Andersen et al., 2001). In the absence of fragmentation and market microstructure noise, the RV estimator is consistent and asymptotically distributed as a mixed normal (Barndorff-Nielsen and Shepard, 2002). However, in the presence of market microstructure noise, the RV estimator is inconsistent, leading to the development of alternative estimators that are robust to different forms of market microstructure noise (see, for example, Christensen and Podolskij, 2007; Xiu, 2010; Barndorff-Nielsen et al., 2011b, for more details). The statistical properties of the market microstructure noise are chief to determine the choice of estimator. For example, Hansen and Lunde (2006) find that the noise is often time-dependent and correlated with the efficient price, while Bandi and Russell (2006) further discuss the importance of the market structure for characterizing the noise component. On addressing the market fragmentation issue, the conventional practice in the literature is to solely sample data from a single trading venue, typically the listing exchange, when computing estimates of the integrated variance. However, in a fragmented market setting, we show that the price-adjustment behaviour to correct deviations from the long-run equilibrium (latent efficient price) induces a specific form of drift-dependent multiplicative market microstructure noise in addition to bid-ask bounce, informational content and discreteness of price changes, strategic component of the order flow, and other factors that are commonly discussed in the literature (see, for example, Zhou, 1996; Aït-Sahalia and Jacod, 2014; Aït-Sahalia and Xiu, 2019). We refer to this new type of noise as *fragmentation noise*. While univariate noise-robust estimators such as Barndorff-Nielsen et al. (2011a)'s realized kernel can address the (endogenous) additive noise associated with typical market microstructure effects, we demonstrate that this class of estimators falls short in addressing the multiplicative noise originating from assets traded across multiple trading venues.

The fragmentation noise we examine relates to a more contemporary strand of microstructure models that highlight the impact of learning on shaping the market microstructure noise. For instance, Andersen et al. (2022) introduce a single-venue microstructure model with a lagged price adjustment mechanism to account for persistent return autocorrelation and endogeneity. They argue that risk-averse agents trade with incomplete information leading to temporary mispricing which in turn leads to speculative trading that pulls the price back towards efficiency. The concept of learning is central to many microstructure models in the literature (see, e.g., Diebold and Strasser, 2013).[1] These models often use an idealized market with a single market maker that learns about the efficient price through trades with informed and uninformed traders. We propose a new explanation for the fragmentation noise in line with these more recent models and the well established class of partial adjustment models (see, e.g., Amihud and Mendelson, 1987; Scherrer, 2021). This explanation is based on the concept of heterogeneous beliefs about the future value of assets among various groups of agents in different market venues. These differing beliefs drive the markets' contribution to price discovery, as observed prices in the different trading venues are expressed as a function of the fundamental value of the asset, their lagged values, and transitory terms.

We specifically address the estimation issue posed by fragmentation noise by proposing a two-step estimator of the integrated variance that leverages data from multiple

[1]The sequential trade models in Glosten and Milgrom (1985) and Easley and O'Hara (1987), as well as the continuous auction models in Kyle (1985) are theoretical foundations for these microstructure models.

exchanges and trading platforms. This strategy allows us to exploit information from the drift of the multivariate diffusion process driving the price process to obtain consistent estimates of the integrated variance. We start by recognizing that prices traded at multiple trading venues cointegrate. This feature of the price process is a cornerstone in the price discovery literature (see, for instance, Hasbrouck, 1995, 2002; De Jong, 2002; Dias et al., 2021) and enables the identification of the common stochastic trend of market prices: the asset's latent efficient price. For this purpose, we apply a vector error correction model (VECM) to prices from different trading venues and estimate the orthogonal complement of the drift. Next, we implement the well-known Stock-Watson decomposition and estimate the efficient price which satisfies the martingale property. For the second step, the returns of this efficient price can form the basis for a conventional RV estimation that is then robust to the fragmentation noise. Hansen and Lunde (2006) employ a VECM to extract the efficient price from transaction prices and bid-ask quotes to investigate the properties of the market microstructure noise and how a change in the efficient price dynamically affects bid, ask, and transaction prices. Differently, our approach uses prices from multiple exchanges to derive a two-step multivariate estimator of the integrated variance in the context of fragmented markets. Additionally, we also establish the estimator's asymptotic properties. More specifically, given that the initial step of our two-step estimator entails estimating the drift of a diffusion process, our asymptotic theory relies on both long span and in-fill asymptotics (see, for instance, Bandi and Phillips, 2003). We also conduct extensive simulation experiments to show that our multivariate estimator outperforms commonly used univariate estimators.

We make several contributions to the literature. Firstly, we identify a novel form of noise referred to as fragmentation noise. We show that this type of noise originates from assets traded across multiple markets and exhibits a multiplicative form. Consequently, both noise-robust univariate and standard multivariate estimators yield inconsistent estimates of the integrated variance. Secondly, we propose a theoretical framework for the estimation of the integrated variance in situations where an asset is actively traded in multiple venues. The estimator can be implemented straightforwardly and it is especially relevant for market practitioners and applied researchers who can now utilize information from all trading venues to estimate the integrated variance, instead of relying on an ad-hoc choice of a single source of information. Thirdly, we derive the asymptotic distribution of our proposed two-step estimator and conduct simulation experiments to analyse its small sample properties. Fourth, we bridge the gap between the realized measures and price discovery literature. By examining the contributions of each market to the price discovery process, we can anticipate when the multivariate estimator is more likely to be more effective than the univariate estimators that are prevalent in the literature.[2] Finally, we apply our two-step estimator to a dataset comprising the 30 constituents of the DJIA, traded across five different trading platforms in the U.S. Our results suggest that the two-step estimator consistently provides reliable estimates in the presence of fragmentation noise, with its estimates never being statistically inferior to those of its competitors. Hence, our approach offers a robust means of estimating the integrated variance and can be extended to other financial markets and assets.[3]

The remainder of this paper is organized as follows. In Section 2, we outline a continuous-time framework for prices in fragmented markets and identify the effect of the fragmentation noise on the well-established noise-robust estimators. Next, we introduce our two-step multivariate estimator and establish its asymptotic theory. In Section 3, we conduct an extensive Monte Carlo simulation study using a data-generating process that includes both constant and stochastic volatility processes. In Section 4, we present an empirical application that shows how incorporating price information from fragmented markets can improve integrated variance estimation. Section 5 summarizes our findings and outlines objectives for future research.

2 Methodology

2.1 Model and basic properties

If an asset is traded on different exchanges simultaneously, each market impounds new information into the efficient price which is believed to exist as a latent common factor driving the observed prices. We model these multiple log-prices in a continuous time framework as a cointegrated multivariate Ornstein-Uhlenbeck (OU) type process (Phillips, 1991; Corradi, 1997; Kessler and Rahbek, 2001, 2004; Dias et al., 2021, 2022)

[2]While the literature is not conclusive on the choice of markets for univariate RV estimation, typically either the listing exchange or the most liquid market is preferred (Martens and van Dijk, 2007; Corsi et al., 2008).

[3]We repeat the same analysis using several spot-futures pairs, including major stock indices, Bitcoin and gold. To conserve space, we relegate the results to the Supplementary Material. In line with our primary results, we find that using multiple markets for the estimation of the integrated variance is more effective than relying on single-market based estimators.

satisfying the following stochastic differential equation (SDE):

$$\begin{aligned} dP(t) &= \Pi P(t)dt + \Sigma(t)dW(t), \qquad t \geq 0, \\ P(0) &= P_0, \end{aligned} \tag{1}$$

where $P(t) = (p_1(t), \ldots, p_N(t))'$ is a $N \times 1$ vector of log prices, $P(0) = P_0$ are initial values, Π is a $N \times N$ reduced-rank parameter matrix with rank equal to $r = N - 1$, W is a $N \times 1$ vector of standard Brownian motion and $\Sigma(t)$ is a $N \times N$ matrix such that the covariance matrix $\Omega(t) = \Sigma(t)\Sigma(t)'$ is positive definite for all $t \geq 0$. Note the important difference to the aforementioned studies that we do not assume the covariance matrix to be constant, instead we allow for a more realistic stochastic covolatility process (see Section 6.4 in Chambers, 2011, for a similar extension).[4] More specifically, we work under the following class of covolatility processes as defined in Assumption 1.

Assumption 1. *(a) $\Sigma(t)$ has elements that are all càdlàg.*
(b) For all $t < \infty$, it holds that $\int_0^t \Omega^{kk}(u)du < \infty$, where $\Omega^{kl}(t)$ denotes the (k,l)-th element of the $\Omega(t)$ process.
(c) $\Sigma(t) \perp W(t)$ for all t.

The first two conditions are standard regularity conditions and are chosen analogously to Barndorff-Nielsen and Shephard (2004). Assumption 1 (c) allows us to condition on the observed volatility path, which is referred in the finance literature as the "no-leverage" case (Aït-Sahalia and Jacod, 2014).

The dynamics of the process are governed by the reduced-rank matrix $\Pi = \alpha\beta'$, where α is an $N \times r$ matrix that contains the adjustment coefficients, and β is a $N \times r$ matrix whose space spanned by its r columns defines the cointegrating vectors. Setting the cointegration rank to $N - 1$ requires the price system to share a single common stochastic trend which we define as the efficient price for the underlying asset. Furthermore, we specify $\beta' = (I_r, -\iota_r)$, where ι_r denotes a $r \times 1$ vector of ones to model pairwise law-of-one-price relationships among the variables (see, among others, Hasbrouck, 1995; De Jong, 2002). We summarize the standard regularity conditions for the cointegrated OU type process (Kessler and Rahbek, 2001; Dias et al., 2022) as follows:

[4] The solution to the SDE in (1) is the homogeneous Gaussian Markov process $P(t) = \exp(t\Pi)\left[P_0 + \int_0^t \exp(-u\Pi)\Sigma(u)dW(u)\right]$, where $\exp(M) = \sum_{l=0}^{\infty} \frac{M^l}{l!}$ is the matrix exponential.

Assumption 2. *(a) All eigenvalues of* Π *are real and no elementary divisor of* Π *occurs more than once.*
(b) α *and* β *have full column ranks* r *and* $\beta' = (I_r, -\iota_r)$. $\beta'\alpha$ *has full rank* r *and all of its eigenvalues have negative real parts, i.e.* $Re(\beta'\alpha) < 0$.

The cointegrated system in (1) can equivalently be written in a common trend representation (Kessler and Rahbek, 2001),

$$P(t) \;=\; \beta_\perp P^*(t) + \Psi(1)P_0 + \zeta(t), \tag{2}$$

where $P^*(t) = (\alpha_\perp{}'\beta_\perp)^{-1}\alpha_\perp{}'\Sigma(t)W(t)$, $\Psi(1) = \beta_\perp(\alpha_\perp{}'\beta_\perp)^{-1}\alpha_\perp$, $\beta_\perp$ and $\alpha_\perp$ denote the orthogonal complements of β and α, respectively, such that $\beta_\perp^{'}\beta = 0$ and $\alpha_\perp^{'}\alpha = 0$, $0 < \delta < 1$, $\zeta(t) = \alpha(\beta'\alpha)^{-1}Z(t)$ and Z is the OU type diffusion process that admits a stationary solution, such that $dZ(t) = \beta'\alpha Z(t)dt + \beta'\Sigma(t)dW(t)$. As is the case of α and β, their orthogonal complements are also not unique. Without loss of generality, we address this point by imposing the following normalizations: $\beta_\perp = \iota_N$ and $\alpha_\perp^{'}\iota_N = 1$. These normalizations imply $\alpha_\perp'\beta_\perp = 1$ and $P^*(t) = \alpha_\perp{}'\Sigma(t)W(t)$, highlighting the importance of $\alpha_\perp$ in our analysis. Specifically, $\alpha_\perp$ shows how the efficient price relates to the market innovations. In the price discovery literature, $\alpha_\perp$ is known as the component share measure, indicating the relative informational importance of the markets. Specifically, the market with the largest element of $\alpha_\perp$ is the most important on impounding information to the efficient price, thereby leading the price discovery process (De Jong, 2002). The Granger decomposition in (2) is key to identify the latent efficient price and ultimately its quadratic variation. In view of Assumption 2, $\beta_\perp = \iota_N$ implies that P^* is common to all markets. Moreover, P^* exhibits the martingale property, fulfilling the required non-arbitrage conditions. Based on these properties, P^* emerges as the natural definition of the latent efficient price. The stationary transitory component $\zeta(t)$ accounts for transitory innovations that impact the market-specific prices but have no permanent effect on the efficient price.

Next, we assume that market prices are observed on a grid of discrete time points. We consider intra-day data sampled equidistantly in calendar time over T days. More specifically, we adopt Phillips and Yu (2009)'s flexible notation that accommodates both long span and in-fill asymptotic analysis. For any day t, the observation times are given by $t_i = ((t-1)m + i)\delta$, $i = 1, \ldots, m$ and $t = 1, \ldots, T$, where $\delta = (1/m)$ denotes the sampling interval, m is the number of intraday observations, and T is the time span of the data. Note that this construction allows for the definition of T non-overlapping sub-samples, where each sub-sample consists of m observations over the interval $[t_1, t_m]$, $t = 1, \dots, T$. For example, 5-min data sampled over the usual 6.5 hours trading day for 252 trading days yields $m = 78$ intra-day observations, an interval length of $\delta = 1/78$, and $Tm = 19656$ observations in total.

It then follows that the exact discretization of (1) results in a linear VECM process as

$$\Delta P_{t_i} = \Pi_\delta P_{t_{i-1}} + \epsilon_{t_i}, \qquad i = 1, \dots, m, \quad t = 1, \dots, T, \tag{3}$$

where P_{t_i} is a $N \times 1$ vector of log-prices, $\Pi_\delta = \alpha_\delta \beta'$, $\alpha_\delta = \alpha(\beta'\alpha)^{-1}[\exp(\delta\beta'\alpha) - I_r]$ and I_r is the $r \times r$ identity matrix.[5] The innovation ϵ_{t_i} is mean-zero Gaussian with covariance matrix $\Omega_{\delta,t_i} = \int_0^\delta \exp(u\Pi)\Omega(t_i - u)\exp(u\Pi')du$. We use the subscript δ to indicate that the values of α_δ and Ω_{δ,t_i} are functions of the sampling interval. In contrast, the value of β remains constant across different choices for the length of the sampling intervals. Kessler and Rahbek (2001, 2004) further show that discrete sampling of the continuous time process preserves the cointegration properties. More specifically, the cointegration rank and their informational ordering remains unchanged.

Finally, we obtain the discrete-time counterpart of the common trend representation of the VECM in Equation (3) as follows

$$P_{t_i} \;\; = \;\; \beta_\perp P^*_{t_i} + \zeta_{t_i} + P_0, \tag{4}$$

where $P^*_{t_i} = \alpha'_{\delta,\perp}\left(\sum_{j=1}^{t-1}\sum_{h=1}^{m}\epsilon_{j_h} + \sum_{h=1}^{i}\epsilon_{t_h}\right)$, $\zeta_{t_i} = \sum_{h=0}^{i-1}\widetilde{C}_h\epsilon_{t_{i-h}} + \sum_{j=1}^{t-1}\sum_{h=1}^{m}\widetilde{C}_{(j-1)m+(i-1)+h}\epsilon_{(t-j)_{(m-h+1)}}$ with $\widetilde{C}_\ell = \alpha_\delta(\beta'\alpha_\delta)^{-1}(I_r + \beta'\alpha_\delta)^\ell\beta'$, $\ell = 0, 1, \dots (t-1)m + (i-1)$, and P_0 is a vector of initial values. Consistent with the continuous-time Granger representation presented in (2), $P^*_{t_i}$ is shared among all price variables and maintains the martingale property, i.e. its increments $\alpha'_{\delta,\perp}\epsilon_{t_i}$ are serially uncorrelated by construction. Therefore, $P^*_{t_i}$ serves as the natural definition of the efficient price in discrete time.[6] Furthermore, the representation in (4) also satisfies the class of partial adjustment models originally introduced by Amihud and Mendelson

[5] While the cointegrated OU type of process yields a discrete time model without short-run dynamics, it still provides a useful framework for financial data. More general lag structures of the discrete time counterparts can be created by related continuous time models, for example by applying the Laplace transform function to the lag operator as in Nguenang (2016).

[6] Note that the stochastic trend in (4) takes a more simplified form than the usual Granger representation because we normalize $\beta_\perp$ and $\alpha_\perp$ as $\beta_\perp = \iota_N$ and $\iota'_N\alpha_\perp = 1$, respectively. This normalization condition ensures that $(\alpha_{\delta,\perp}{}'\beta_\perp)^{-1} = 1$, further simplifying the form of the stochastic trend.

(1987) (see also Scherrer, 2021, for a detailed discussion on the identification of permanent and transitory shocks). In what follows, we will use the common stochastic trend implied by the Granger representation in (4) as the building block for formulating our multivariate estimator for the integrated variance of the efficient price.

2.2 The fragmentation noise

Prices of assets traded in multiple venues exhibit a more intricate dependence structure compared to those traded in a single venue. We can infer from the VMA representation of (3) that the high-frequency returns are serially correlated. Under the assumption that only one common stochastic trend drives the system, at least $N-1$ components of ΔP_{t_i} are serially correlated.[7] In the single market setting, the serial correlation in the intra-day returns are by far most commonly modelled as originated from an additive market microstructure noise, i.e., the observed prices are equal to the efficient price plus the market microstructure noise (Hansen and Lunde, 2006). Differently, the serial correlation and the identification of the latent efficient price in our setting emerges from the drift and its orthogonal complements, respectively. To appreciate the latter, multiply both sides of (4) by $\alpha'_{\delta,\perp}$ and rearrange terms to write the efficient price as

$$P^*_{t_i} = \alpha'_{\delta,\perp} \left(P_{t_i} - P_0\right). \tag{5}$$

The expression in (5) indicates a different source of deterministic market microstructure noise that takes a multiplicative form and depends on the relative informational importance of each market, $\alpha_{\delta,\perp}$. We term this dependence structure fragmentation noise, and we now explore it in the context of integrated variance estimation.

Our goal is to estimate the quadratic variation of the latent efficient price. To this extent, we start by defining the quadratic variation of P in (1) as

$$\langle P, P' \rangle_{t-1:t} = \int_{t-1}^{t} \Sigma(u)\Sigma(u)' du = \int_{t-1}^{t} \Omega(u) du, \tag{6}$$

where $\langle P, P' \rangle_{t-1:t} = \text{plim}_{\max\{t_{i+1}-t_i\}\to 0} \sum_{i=1}^{m} \left(P_{t_{i+1}} - P_{t_i}\right)\left(P_{t_{i+1}} - P_{t_i}\right)'$ and the diagonal elements of $\langle P, P' \rangle_{t-1:t}$ are the market-specific quadratic variations. Because the drift is slower moving that the diffusion component, it has zero quadratic variation. This

[7] In the Supplementary Material, we study the autocovariance structure under the assumption of a constant contemporaneous covariance matrix of the innovation terms.

means that the long-run equilibrium relationship between market prices and the speed of adjustment towards the common efficient price cannot be identified by the quadratic variation process, and has no effect on $\langle P, P' \rangle_{t-1:t}$. We formalize this feature from the realized covariance class of estimators in Assumption 3 below.

Assumption 3. *As $m \to \infty$, there exists an estimator denoted by $\widehat{\langle P, P' \rangle}_{t-1:t}$ such that $\widehat{\langle P, P' \rangle}_{t-1:t} - \langle P, P' \rangle_{t-1:t} = O_p(m^{-\varphi})$, with $\varphi \leq 1/2$.*

Assumption 3 simply states that $\widehat{\langle P, P' \rangle}_{t-1:t}$ is a consistent estimator of the quadratic variation of P with a convergence rate denoted by φ. This class of consistent estimators includes, among others, the well-known realized covariance estimator with $\varphi = 1/2$ (Barndorff-Nielsen and Shephard, 2004), the additive-noise robust pre-averaging estimator that achieves the optimal rate $\varphi = 1/4$ (Christensen et al., 2010), and the dependent noise-robust realized kernel estimators with $\varphi = 1/5$ (Barndorff-Nielsen et al., 2011a).

Similarly, from the efficient price process defined in (2), it follows that the quadratic variation of the efficient price P^* reads

$$\langle P^*, P^{*\prime} \rangle_{t-1:t} = \alpha'_\perp \langle P, P' \rangle_{t-1:t} \alpha_\perp = \int_{t-1}^{t} \alpha'_\perp \Omega(u) \alpha_\perp du, \tag{7}$$

where $\langle P^*, P^{*\prime} \rangle_{t-1:t} = \operatorname{plim}_{\max\{t_{i+1}-t_i\} \to 0} \sum_{i=1}^{m} \left(P^*_{t_{i+1}} - P^*_{t_i} \right)^2$ and $\iota'_N \alpha_\perp = 1$. Hence, the quadratic variation of the efficient price, henceforth the integrated variance, is a function of the market-specific quadratic variations and their covariations, weighted by the informational ordering in $\alpha'_\perp$.

It is worth reiterating that the price process in (1) is a semimartingale. This means that when using the estimators that satisfy Assumption 3, consistent estimates of $\langle P, P' \rangle_{t-1:t}$ are obtained instead of $\langle P^*, P^{*\prime} \rangle_{t-1:t}$. Similarly, employing a noise-robust univariate estimator on a single market among the N trading venues will only estimate the market-specific quadratic variations, that is the diagonal element of $\langle P, P' \rangle_{t-1:t}$ associated with that particular market. These findings suggest that using estimators that control for additive market microstructure noise alone is not sufficient to obtain consistent estimates of $\langle P^*, P^{*\prime} \rangle_{t-1:t}$.

The expression for $\langle P^*, P^{*\prime} \rangle_{t-1:t}$ and the representation in (5) hints that $\alpha_\perp$ represents a different type of multiplicative noise that summarizes the degree of informational market fragmentation. To illustrate the significance of $\alpha_{\delta,\perp}$ as a measure of informational market fragmentation in discrete time, consider the efficient price

$P^*_{t_i} = \alpha'_{\delta,\perp}\left(\sum_{j=1}^{t-1}\sum_{h=1}^{m}\epsilon_{j_h} + \sum_{h=1}^{i}\epsilon_{t_h}\right)$. If $\alpha_{\delta,\perp} = k(n)$, where $n \in [1,\ldots,N]$ and $k(n)$ is an $N \times 1$ vector with its nth element equal to one and all other entries equal to zero, then only market n contributes to the efficient price, indicating that markets are not informationally fragmented. In this case, applying any estimator to prices sampled from market n will produce consistent estimates of the integrated variance. However, in a more realistic scenario where more than one market contributes to the efficient price, it folows that $\alpha_{\delta,\perp} \neq k(n)$, thereby indicating that markets are informationally fragmented. In this case, $\alpha_{\delta,\perp}$ acts as a multiplicative fragmentation noise that is a function of the drift and therefore cannot be directly identified from the quadratic variation of the market prices P. Finally, it is important to note that even when a sparse sampling frequency is chosen for the analysis, where other microstructure effects become less significant (such as 1-minute or 5-minute sampling intervals), the fragmentation noise remains present. This is because $\alpha_{\delta,\perp}$ is independent of the sampling frequency, $\alpha_{\delta,\perp} = \alpha_\perp$ for $0 < \delta < 1$ (Dias et al., 2021). Theorem 1 formalizes the effect of fragmentation noise on the asymptotic bias of estimates of $\langle P, P'\rangle_{t-1:t}$.

Theorem 1. *Suppose Assumptions 1, 2 and 3 hold. If $m \to \infty$, then*

$$\widehat{\langle P, P'\rangle}^{n,n}_{t-1:t} - \langle P^*, P^{*\prime}\rangle_{t-1:t} = \left(1-\alpha^2_{\perp,n}\right)\langle P, P'\rangle^{n,n}_{t-1:t} - \sum_{i=1, i\neq n}^{N}\alpha^2_{\perp,i}\langle P, P'\rangle^{i,i}_{t-1:t} - 2\sum_{i=1}^{N-1}\sum_{j=i+1}^{N}\alpha_{\perp,i}\alpha_{\perp,j}\langle P, P'\rangle^{i,j}_{t-1:t} + O_p(m^{-\varphi}),$$

where $n = 1,\ldots,N$ and $\langle P, P'\rangle^{i,j}_{t-1:t}$ denotes the (i,j)-th element of $\langle P, P'\rangle_{t-1:t}$.

Theorem 1 presents three important results. First, if a single market (market n) is exclusively responsible for impounding information into the efficient price, i.e., $\alpha_\perp = k(n)$, then $\widehat{\langle P, P'\rangle}^{n,n}_{t-1:t} - \langle P^*, P^{*\prime}\rangle_{t-1:t} = O_p(m^{-\varphi})$, that is consistent estimation of $\langle P, P'\rangle^{n,n}_{t-1:t}$ implies consistent estimation of the integrated variance.

Second, if more than one market contributes to the efficient price such that two or more elements of $\alpha_\perp$ are non-zero, $\Omega(t)$ is restricted such that $\langle P, P'\rangle^{n,n}_{t-1:t} = \langle P, P'\rangle^{\bar{n},\bar{n}}_{t-1:t}$ for all $\bar{n} \neq n$ and $n, \bar{n} = 1,\ldots N$ and the correlation between markets remain the same across all market combinations, then the following limit holds:

$$\lim_{\rho\to 1}\widehat{\langle P, P'\rangle}^{n,n}_{t-1:t} - \langle P^*, P^{*\prime}\rangle_{t-1:t} = O_p(m^{-\varphi}), \tag{8}$$

where ρ denotes the correlation between any market combination. Furthermore, if

$\rho = 0$ and $\langle P, P' \rangle^{n,n}_{t-1:t} = \langle P, P' \rangle^{\bar{n},\bar{n}}_{t-1:t}$ for all $\bar{n} \neq n$ and $n, \bar{n} = 1, \ldots N$, then $\widehat{\langle P, P' \rangle}^{\bar{n},\bar{n}}_{t-1:t} - \langle P^*, P^{*\prime} \rangle_{t-1:t} = O_p(m^{-\varphi})$ also holds. It is important to highlight that the first two sets of results hold under rather strong additional assumptions, which are unlikely to hold in practice. In this context, we analyze the asymptotic bias under more general and realistic conditions, leading us to the next result.

Third, if more than one market contributes to the efficient price and no additional restrictions are imposed on $\Omega(t)$, then $\widehat{\langle P, P' \rangle}^{n,n}_{t-1:t} - \langle P^*, P^{*\prime} \rangle_{t-1:t} = O_p(1)$. This implies that the consistent estimation of $\langle P, P' \rangle^{n,n}_{t-1:t}$ does not necessarily result in the consistent estimation of $\langle P^*, P^{*\prime} \rangle_{t-1:t}$. The severity of the asymptotic bias depends on the parametrization of α, particularly its implied informational ordering, and the stochastic covolatility process $\Omega(t)$. In summary, Theorem 1 provides theoretical evidence that fragmentation noise generally leads to inconsistent estimation of the integrated variance of the efficient price.

To address this issue, we propose a two-step estimator, which consistently estimates $\langle P^*, P^{*\prime} \rangle_{t-1:t}$. The key insight is to recognize that $\alpha_\perp$ cannot be identified from the quadratic variation of P, and hence it is estimated parametrically in the first step. In the second step, the integrated variance is estimated based on the estimates of P^*. The benefits of the multivariate estimation framework we adopt is directly related to the present market microstructure. To identify favorable settings, we direct market practitioners to the price discovery literature. This literature offers empirical measures like component shares and information shares, which help determine the informational ordering in the price discovery process and assess the presence of fragmentation noise (Hasbrouck, 1995; Harris et al., 2002). More specifically, the component share measure coincides with the $\alpha_\perp$ and, in contrast to information shares, they are invariant to the sampling frequency (Dias et al., 2021).

2.3 Estimator and asymptotic theory

To obtain a consistent estimator of the integrated variance under fragmentation noise, we propose to incorporate available price information from all venues and parametrically estimate the drift in (1). Our estimator consists of two main steps. First, we use the full sample to estimate the discrete-time VECM coefficients in (3), i.e. α_δ. Notably, if the diffusion matrix in (1) is constant, parameters in the drift can be identified from discretely observed data (Kessler and Rahbek, 2004). More generally, under Assumptions 1 and 2, α remains identifiable when we condition on the observed volatility path and β is known. Next, because the conditional heteroskedasticity in our model is of unknown form, the log-likelihood function for the joint estimation of the drift coefficient Π_δ and the complete covolatility process Ω_{δ,t_i}, for $t = 1, \ldots, T$, $i = 1, \ldots, m$ is generally intractable. Alternatively, because we are only interested in the parametric estimation of the drift coefficient (the quadratic variation is estimated nonparametrically in the second step), α_δ can be estimated by least squares. Specifically, identification of the drift also requires the time span to diverge, and the natural least-squares criterion minimizes the function:

$$Q(\Pi_\delta) = \sum_{t=1}^{T} \sum_{i=1}^{m} \left(\Delta P_{t_i} - \tilde{\alpha}_\delta \beta' P_{t_{i-1}} \right)^2, \tag{9}$$

with respect to $\tilde{\alpha}_\delta$, thereby yielding $\hat{\alpha}_\delta$. The least squares-based estimator of $\alpha_\perp$ can be computed as in Hansen and Lunde (2006) such that the normalization $\iota_N' \hat{\alpha}_\perp = 1$ holds. Next, $\widehat{\alpha}_\delta$ is used to compute the residuals $\widehat{\epsilon}_{t_i}$, and ultimately the efficient price returns $\Delta \widehat{P}^*_{t_i} = \widehat{\alpha}'_{\delta,\perp} \widehat{\epsilon}_{t_i}$.

Second, we rely on the RV estimator to take advantage of the available intra-day price information to consistently estimate the integrated variance (see, among others, Andersen et al., 2001; Barndorff-Nielsen and Shepard, 2002). The reason for using this method is that the estimated efficient price conforms to a martingale process. Consequently, we can leverage the theoretical outcomes for the RV estimator presented in Barndorff-Nielsen and Shepard (2002) to establish the complete asymptotic theory for our two-step estimator. Our two-step estimator then reads:

$$RV^{(m)}_{GRT,t} = \sum_{i=1}^{m} (\hat{\alpha}'_{\delta,\perp} \widehat{\epsilon}_{t_i})^2, \quad t = 1, \ldots, T. \tag{10}$$

Alternatively, we could replace the estimator in (10) by the asymptotically equivalent estimator $\sum_{i=1}^{m} (\hat{\alpha}'_{\delta,\perp} \Delta P_{t_i})^2$, as $\sum_{i=1}^{m} (\Delta P_{t_i} \Delta P'_{t_i}) \xrightarrow{p} \langle P, P' \rangle^{n,n}_{t-1:t} = \int_{t-1}^{t} \Omega_u du$ for $m \to \infty$ (Kessler and Rahbek, 2001).

We now focus on establishing the limiting distribution of $RV^{(m)}_{GRT,t}$. In the literature on integrated variance estimation, in-fill asymptotics are commonly employed to derive the asymptotic distribution of RV and other alternative noise-robust estimators (see, among others, Andersen et al., 2001; Hansen and Lunde, 2006; Barndorff-Nielsen et al., 2008). The typical asymptotic framework involves limiting the data to a finite time span, such as a single trading day, and then increasing the sampling frequency, resulting in $m \to \infty$. In contrast, our case involves first estimating the drift coefficient of a diffusion process. It is well-known that the least squares estimator in such settings is only consistent under the conditions $T \to \infty$ and $T\delta \to 0$ (Dorogovtsev, 1978). Furthermore, Prakasa Rao (1983) demonstrates that the least squares estimator is asymptotically normal and efficient if the conditions $\delta \to 0$, $T \to \infty$, and $T\delta^2 \to 0$ hold.[8] Generating data from the cointegrated OU type process means that the adjustment coefficients of the exactly discretized VECM take frequency-specific values and increasing the sampling frequency does not provide enough observations to estimate α_δ consistently. However, assuming in-fill and long span asymptotics, i.e., letting $T \to \infty$, $\delta \to 0$ and maintaining the condition $T\delta^2 \to 0$, ensures that $\hat{\alpha}_\delta$ is consistent and efficient. This follows because with an increasing number of days T, we now have sufficient observations to estimate α_δ for each choice of α, despite α_δ changing as $\delta \to 0$. Theorem 2 formalizes the asymptotic distribution of our two-step estimator.[9]

Theorem 2. *Under Assumption 1 and Assumption 2, it holds that*

$$\frac{\left(\sum_{i=1}^{m}(\hat{\alpha}'_{\delta,\perp}\Delta P_{t_i})^2 - \int_{t-1}^{t}\alpha'_\perp\Omega(u)\alpha_\perp du\right)}{\sqrt{\frac{2}{3\delta}\sum_{i=1}^{m}(\hat{\alpha}'_{\delta,\perp}\Delta P_{t_i})^4}} \xrightarrow{d} \mathcal{N}(0,1),$$

for $T \to \infty$, $\delta \to 0$, and $T\delta^2 \to 0$.

Remark 1. We use the realized quarticity in the denominator to derive a statistically feasible result (except for $\int_{t-1}^{t}\alpha'_\perp\Omega(u)\alpha_\perp du$ being random and unknown) which allows us to construct confidence bands for $\int_{t-1}^{t}\alpha'_\perp\Omega(u)\alpha_\perp du$. Finally, it is worth noting that we have presented Theorem 2 in terms of $\hat{\alpha}'_{\delta,\perp}\Delta P_{t_i}$ instead of the efficient price returns $\hat{\alpha}'_{\delta,\perp}\widehat{\epsilon}_{t_i}$. This choice is justified by their asymptotic equivalence and it aligns with the broader literature.

To investigate the quality of the asymptotic approximation in finite samples, we generate data according to the cointegrated OU type process in Equation (1) with $\alpha = (-5, 5)$ and $\Omega(t)$ specified as in the data-generating process defined in our simulation exercise in Subsection 3.3. The results are depicted in Figure B1 and Figure B2 in

[8] Prakasa Rao's 1983 so-called experimental design condition is further relaxed in Yoshida (1992) and Kessler (1997).

[9] The corresponding result under the assumption of a constant contemporaneous covariance matrix is given in the Supplementary Material.

the Appendix. Both figures show that the estimator slightly underestimates the true integrated variance at lower frequencies which makes its studentized distribution left-skewed. The bias vanishes with increasing sampling frequency. While we find that the quality of the approximation mostly hinges on the sampling frequency, the results for different combinations of T and δ, reported in Table B1 in the Appendix, clearly show that the rate $T\delta^2$ needs to tend to zero for consistency.

3 Simulation study

3.1 Quadratic estimators

In this section, we assess the relative performance of out two-step estimator under different data-generating process. Before we turn to our simulation experiments, we need to define the set of alternative univariate estimators that can be compared with our proposed multivariate approach (see, for instance, Liu et al., 2015). Since we limit our analysis to continuous processes, we do not include estimators explicitly addressing jump activity (see, for example, Andersen et al., 2007, for a discussion of jumps in financial time series). Apart from the RV estimator, we predominantly focus on different realized kernel estimators for the integrated variance that are robust to different types of market microstructure noise contaminations.[10] We use the notation $r_{n,t_i} = P_{n,t_i} - P_{n,t_{i-1}}$ for intra-day returns of market n, $n = 1, \ldots, N$, sampled equidistantly in calendar time at a fixed interval δ.

The first class of estimators we consider is the RV estimator (Andersen et al., 2001):

$$RV_{n,t} = \sum_{i=1}^{m} r_{n,t_i}^2, \quad n = 1, \ldots, N. \tag{11}$$

This estimator simply adds up the squared intra-day returns at a given sampling interval δ to measure the integrated variance of any market n for day t.[11] The RV estimator consistently estimates market n's quadratic variation, i.e. $RV_{n,t} \xrightarrow{p} \langle P, P' \rangle_{t-1:t}^{n,n}$ as

[10] Liu et al. (2015) discuss additional estimators. For example, the pre-averaging estimator by Jacod et al. (2009) is closely related to the class of realized kernel estimators. Other viable alternatives like the class of range-based estimators (Martens and van Dijk, 2007; Christensen and Podolskij, 2007) could, in principle, be added but are excluded because they follow a different principle to estimate the integrated variance.

[11] Bandi and Russell (2008) discuss the choice of optimal sampling frequency for the simple RV estimator as a bias-variance trade-off.

$m \to \infty$ (Barndorff-Nielsen and Shepard, 2002). Alternatively, the realized covariance (RCov) estimator, denoted as $RCov_t = \sum_{i=1}^{m} \Delta P_{t_i} \Delta P'_{t_i}$, consistently estimates $\langle P, P' \rangle_{t-1:t}$ (Barndorff-Nielsen and Shephard, 2004). Both the univariate and multivariate realized variance estimators attain the optimal convergence rate $m^{1/2}$, but they are biased and inconsistent estimators for the integrated variance in the presence of market microstructure noise. Additionally, in view of Theorem 1, if $\alpha_\perp \neq k(n)$ and $\rho \neq 0$, both the RV and RCov estimators are inconsistent estimators of the integrated variance.

In our second class of estimators, we examine two kernel-based univariate methods proposed by Hansen and Lunde (2006). These estimators capture the effects of serial correlation in high frequency returns induced by additive market microstructure noise. More specifically, they read

$$RV_{n,t}^{AC} = \sum_{i=1}^{m} r_{n,t_i}^2 + 2 \sum_{h=1}^{H} \frac{m}{(m-h)} \sum_{i=1}^{m-h} r_{n,t_i} r_{n,t_{i+h}}, \tag{12}$$

and

$$RV_{n,t}^{NW} = \sum_{i=1}^{m} r_{n,t_i}^2 + 2 \sum_{h=1}^{H} \left(1 - \frac{h}{(H+1)}\right) \sum_{i=1}^{m-h} r_{n,t_i} r_{n,t_{i+h}}, \tag{13}$$

where H is a frequency-dependent truncation parameter for the covariance terms. The first term on the right side of (12) and (13) is the classical RV estimator and the second term is a bias correction. The first estimator scales the h-th autocorrelation by $m/(m-h)$ to compensate for the 'missing' autocovariance terms at the end of each day. The upward scaling has the drawback that it increases the variance of the estimator. The optimal bandwidth for the estimators to balance bias and variance is given by $H = \lceil (4m/100)^{2/9} \rceil$. Both estimators are robust to the presence of endogenous noise in the sense that they are unbiased for this general type of noise.

Finally, we extend our analysis to the generalized realized kernel estimator of Barndorff-Nielsen et al. (2008, 2011a). The realized kernel estimator for trading day t is given by:

$$RK_{n,t} = \sum_{i=1}^{m} r_{n,t_i}^2 + \sum_{h=1}^{H} k\left(\frac{h-1}{H}\right)(\eta_h + \eta_{-h}), \quad \eta_h = \sum_{i=1}^{m-h} r_{n,t_i} r_{n,t_{i+h}}, \tag{14}$$

where $k(\cdot)$ is either the modified Tukey-Hanning kernel or Parzen kernel function. For our simulations and the empirical application, we denote the realized kernel estimator with modified Tukey-Hanning kernel by RK^{MTH} and with Parzen kernel by RK^{P}.

The optimal choice of the bandwidth H is discussed in Barndorff-Nielsen et al. (2009). Barndorff-Nielsen et al. (2008) show that realized kernel estimators are unbiased and consistent. Estimators based on 'flat-top' kernel functions like the Bartlett or modified Tukey-Hanning converge at a faster rate than those based on the 'non-flat-top' Parzen kernel, but are not guaranteed to be non-negative. However, both kernel choices are robust to time-dependent and endogenous noise. An analysis of the finite sample performance of realized kernels is provided by Bandi and Russell (2011).[12]

3.2 Constant volatility: reduced-form models

In the first part of our simulation study, we use reduced-form models, similar to the two-market example in De Jong (2002), to show how the properties of our estimators change depending on the parametrization of the discrete time VECM process. More precisely, we investigate in which situations the multivariate estimator performs better than the univariate estimators and assess the improvement depending on the given market structure. Using a parametric reduced-form model, we can easily compute the true integrated variance and evaluate the estimators' performance in those situations. For this matter, we generate 5-min data ($m = 78$, NYSE trading hours from 9:30 to 16:00) over $T = 1826$ trading days (five years) according to the bivariate VECM,

$$\begin{bmatrix} \Delta P_{1,t_i} \\ \Delta P_{2,t_i} \end{bmatrix} = \begin{bmatrix} \alpha_{\delta,1} \\ \alpha_{\delta,2} \end{bmatrix} (P_{1,t_{i-1}} - P_{2,t_{i-1}}) + \begin{bmatrix} \epsilon_{1,t_i} \\ \epsilon_{2,t_i} \end{bmatrix}, \qquad i = 1, \ldots, m, \quad t = 1, \ldots, T, \quad (15)$$

where the error term $\epsilon_{t_i} = (\epsilon_{1,t_i}, \epsilon_{2,t_i})'$ is mean-zero i.i.d. Gaussian with constant covariance matrix

$$\Omega_\delta = \begin{pmatrix} \tau \cdot 6.5 \cdot 10^{-7} & 5.0 \cdot 10^{-8} \\ 5.0 \cdot 10^{-8} & 6.5 \cdot 10^{-7} \end{pmatrix}, \quad (16)$$

and τ is a scaling factor for the first markets' innovation variance. For each configuration of the DGP that we study, we have the knowledge of the true efficient price and market-specific contributions to the integrated variance. This information enables us to compute the bias associated with the different estimators. The covariance matrix Ω_δ and the values of α_δ are chosen in accordance with our empirical application.

The first configuration (DGP I) is given by $\alpha = (-0.5, 0.5)'$, i.e., both markets contribute equally to maintain the cointegration relationship. This yields the corresponding

[12] See also Liu et al. (2015) for a discussion about the performance of alternative realized measures estimators.

adjustment coefficients $\alpha_\delta \approx (-0.0063, 0.0063)'$ for 5-min data and $\alpha_{\delta,\perp} = (0.5, 0.5)'$. Furthermore, we choose $\tau = 1.25$ to have slightly unequal variance levels of the idiosyncratic innovation processes in the two markets. With this specification, the first market generates additional relevant information for the efficient prices of the asset. In light of the presence of the fragmentation noise and the result in Theorem 1, it is noteworthy that the estimators discussed in Section 3.1 satisfy a univariate version of Assumption 3 and hence do not consistently estimate the integrated variance in this DGP. More specifically, the integrated variance and the quadratic variation of P in DGP I are given by

$$\langle P^*, P^{*\prime}\rangle_{t-1:t} = 3.086 \cdot 10^{-5} \quad \text{and} \quad \langle P, P'\rangle_{t-1:t} = \begin{pmatrix} 6.436 \cdot 10^{-5} & 3.780 \cdot 10^{-6} \\ 3.780 \cdot 10^{-6} & 5.152 \cdot 10^{-5} \end{pmatrix}. \tag{17}$$

The results for the first configuration with a 5-minute sampling interval are reported on the left panel of Table 1. Specifically, for each estimator, we present the average bias across all replications along with the corresponding RMSE (in brackets). Our simple two-step RV estimator performs best for estimating the integrated variance in terms of bias and RMSE. This outcome is expected, as we consistently estimate $\alpha_\perp$ in the first step and apply the RV estimator to the consistent estimates of the efficient price, which, in turn, satisfies the martingale property. The simulation results confirm our theoretical findings in Theorem 2, demonstrating that the R_{GRT} estimator consistently estimates the integrated variance. The second and third set of results on the left panel in Table 1 display the results associated with alternative univariate estimators computed using prices from markets one and two, respectively. Overall, we find that the RV estimator is a consistent estimator of the respective diagonal element of the quadratic variation of P and, therefore, fails to consistently estimate the integrated variance. The kernel-based estimators unsuccessfully attempt to control for some of the fragmentation noise, and their estimates remain largely biased irrespective of the lag truncation H we adopt. Overall, the first set of results confirms our theoretical findings, namely that estimators that satisfy Assumption 3 are not robust to fragmentation noise and hence deliver inconsistent estimates of $\langle P^*, P^{*\prime}\rangle_{t-1:t}$ (Theorem 1), whereas our two-step estimator successfully estimates the integrated variance (Theorem 2).

The second parametrization (DGP II) sets $\alpha = (-0.9, 0.1)'$ with the corresponding 5-minute adjustment coefficients and orthogonal complement given by $\alpha_\delta = (-0.0113, 0.0013)'$ and $\alpha_{\delta,\perp} = (0.1, 0.9)'$, respectively. The latter implies that the second market is the most relevant in impounding information to the efficient price and thus leads the price discovery process. We set $\tau = 1$ to obtain the same variance level of the idiosyncratic innovation process for both markets. It holds that the integrated variance of the efficient price and the quadratic variation of P then read

$$\langle P^*, P^{*\prime} \rangle_{t-1:t} = 4.282 \cdot 10^{-5} \quad \text{and} \quad \langle P, P' \rangle_{t-1:t} = \begin{pmatrix} 5.184 \cdot 10^{-5} & 3.896 \cdot 10^{-6} \\ 3.896 \cdot 10^{-6} & 5.136 \cdot 10^{-5} \end{pmatrix}. \tag{18}$$

Similarly to the first parametrization, we see clear performance gains from our multivariate estimator in terms of bias and efficiency. The bias associated with the RV_{GRT} estimator is close to zero, which reinforces the theoretical properties established in Theorem 2. On contrary, the alternative univariate estimators display substantial biases across the board. More specifically, while the realized kernel estimators seem to account for some of the fragmentation noise by increasing the lag truncation parameter, the results for both market prices are still severely biased even at $H = 20$. Overall, we find that univariate estimators are not able to separate the fragmentation noise from the efficient price process if both markets contribute to the price discovery process. As highlighted in Theorem 1, if one price is exogenous, it coincides with the efficient price in this two-market model and we see no gains from using the multivariate approach.[13]

3.3 Stochastic volatility process

We now turn to the second part of our simulation study, where we adopt more realistic assumptions about the price process. Specifically, we relax the constant volatility assumption used for the first part of our simulation study and parameterize the market-specific spot variances as following a single-factor stochastic volatility process with serially correlated daily integrated variances.

Following Dias et al. (2021), we assume that two markets trade the same asset, and have the same parametrization of the stochastic volatility processes. This structure is superimposed by the informational ordering of both markets via the α coefficients of the cointegrated OU type process, which means that different signals generated in each market must be incorporated in the efficient price of the asset. More specifically, the prices follow the model in (1) and the contemporaneous correlation between both

[13]Results for this case are not reported but can be obtained from the authors upon request.

innovation processes is constant and takes values $\rho \in \{0.70, 0.90, 0.95\}$.[14] The market-specific spot variances follow a single-factor stochastic volatility process as in Huang and Tauchen (2005) (see also Barndorff-Nielsen et al., 2008; Goncalves and Meddahi, 2009, for similar data-generating processes). The spot variances evolve according to the process

$$\sigma_j^2(t) = \exp(\varsigma_0 + \varsigma_1 V_j(t)), \tag{19}$$

with

$$dV_j(t) = \gamma_V V_j(t)dt + dB_j(t), \;\; j = 1, 2. \tag{20}$$

We follow Barndorff-Nielsen et al. (2008) and choose $\varsigma_0 = 0$, $\varsigma_1 = 0.125$, $\gamma_V = -0.025$, $corr(dW(t), dB_j(t)) = -0.30$, and $corr(dB_1(t), dB_2(t)) = 0.95$. The daily integrated market-specific variances follow the AR(1) process outlined in Jacquier et al. (1994):

$$\ln \sigma_{j,t}^2 = \varrho_0 + \varrho_1 \ln \sigma_{j,t}^2 + \varsigma \upsilon_{j,t}, \quad j = 1, 2, \quad t = 1, \ldots, T, \tag{21}$$

where the volatility innovations $\upsilon_t = (\upsilon_{1,t}, \upsilon_{2,t})'$ are generated by a two-dimensional Gaussian white noise process with constant correlation of 0.95 and unit variances. The autoregressive parameter is set to 0.98, while the coefficients ϱ_0 and ς are calibrated such that the expected variance is $5.5 \cdot 10^{-5}$, i.e., corresponds approximately to the values observed in our empirical application.

Table 2 displays the results. In a nutshell, our RV_{GRT} estimator always performs better than any of the univariate estimators we consider both in terms of bias and RMSE. This holds irrespective of the $\alpha_\perp$ and correlation between markets. In line with the result in Theorem 1, we find results even more in favor of our multivariate estimator when we choose a lower value of the contemporaneous correlation (e.g. choosing $\rho = 0.7$). Consistent with our previous simulation results, we find that estimators that satisfy Assumption 3 are unable to consistently estimate the integrated variance in the presence of fragmentation noise.

In summary, our simulation study demonstrates that the RV_{GRT} consistently estimates the integrated variance and significantly outperforms all competitors. Additionally, in line with the results in Theorem 1, all the univariate estimators we consider yield inconsistent estimates of the integrated variance. Moreover, introducing a more

[14] We report our main results for $\rho = 0.9$ corresponding to a medium value observed in our empirical application. The results for alternative specifications of the contemporaneous correlation parameter are reported in the Supplementary Material.

realistic stochastic covolatility process does not change the properties of our estimator, which are mainly driven by informational ordering determined by the orthogonal complement of the speed-of-adjustment coefficients of the cointegrated OU type process and the contemporaneous correlation between the innovations.

4 Empirical application

4.1 Data

Our primary analysis focuses on stock prices for the 30 constituents of the DJIA. The dataset spans 359 days, from January 4, 2021, to June 06, 2022, encompassing the five main exchanges in the U.S.: NYSE (N), Nasdaq (T), Arca (P), Bats EDGX (K), and Bats BZX (Z). We retrieve quote data from TAQ and apply the same cleaning filters as in Barndorff-Nielsen et al. (2009). This cleaning procedure involves discarding observations with a zero quote, negative bid-ask spread, or outside the main trading hours (9:30 to 16:00). Additionally, we exclude data points with a bid-ask spread exceeding 50 times the median spread on that day or with a midquote deviating by more than 10 mean absolute deviations from a rolling centered median of 50 observations. In cases of multiple ticks within the same second, we use the median bid and ask quotes. We synchronize midquotes from both trading venues using the refresh-time approach and aggregate them into regularly spaced intervals of 1 second and 5 minutes.

4.2 Testing framework

We evaluate the accuracy of estimators of the integrated variance using the testing framework developed by Patton (2011). This approach is designed to account for the latent nature of the integrated variance and enables a comparative evaluation of estimators against a benchmark: the two-step estimator RV_{GRT}. To conduct this assessment, we employ a data-driven ranking of RV estimators, allowing us to gauge the performance of the multivariate approach against the same set of univariate estimators introduced in our simulation study.

Denote $\boldsymbol{X}_t$ a vector collecting the benchmark estimator $X_{0,t}$ along our 5 different univariate estimators for daily integrated variance obtained for each market, such that $\boldsymbol{X}_t = (X_{0,t}, X_{1,t}, \ldots, X_{5N,t})'$. Our target variable (the latent integrated variance) is denoted as θ_t. For a suitably chosen loss function L, we evaluate the null hypotheses

$$H_0^s : E[L(\theta_t, X_{0,t})] = E[L(\theta_t, X_{s,t})], \quad s = 1, 2, \ldots, 5N, \tag{22}$$

against the alternatives

$$H_1^s : E[L(\theta_t, X_{0,t})] > E[L(\theta_t, X_{s,t})], \tag{23}$$

or

$$H_2^s : E[L(\theta_t, X_{0,t})] < E[L(\theta_t, X_{s,t})]. \tag{24}$$

Consequently, our null hypotheses are equal accuracy of the respective estimator compared with the benchmark estimator. In our empirical application in Subsection 4.3, we designate RV_{GRT} as our benchmark estimator, such that $X_{0,t} = RV_{GRT}$ in the context of the null and alternative hyptheses above. Following Patton (2011) and Liu et al. (2015), we adopt the QLIKE loss function,

$$L(x, y) = \frac{x}{y} - \log\left(\frac{x}{y}\right) - 1, \tag{25}$$

to evaluate the equal accuracy of a set of estimators. Specifically, they find that tests based on the QLIKE loss function have higher power than those based on the MSE.[15] Because θ_t represents the latent integrated variance, we must identify an unbiased estimator for it that serves as a proxy for the target variable. Moreover, it is crucial that the estimation errors of the proxy and the estimators $\boldsymbol{X}_t$ are uncorrelated. As demonstrated by Patton (2011), the leads of commonly used variance estimators, such as daily squared returns or 5-minute RV, can serve as appropriate proxies. However, as outlined in Section 2, the univariate estimator can be inconsistent depending on the configuration of the fragmentation noise. Therefore, we utilize leads of the daily squared return of the estimated efficient price as the proxy, though results for alternative proxies yield similar outcomes. To test the null hypotheses in (22), we refer to the stepwise multiple testing method proposed by Romano and Wolf (2005) which controls the family-wise error rate. We generate 1000 bootstrap replications from the stationary bootstrap of Politis and Romano (1994) with an average block size of 10 days to evaluate the test. The choice of the average block size has to be made on the basis of the time series' persistence and our choice is quite conservative in this regard.

[15]Results for the alternative MSE loss function can be obtained from the authors upon request.

4.3 DJIA's constituents in multiple trading venues

We begin by examining the data aggregated at a 5-minute interval. The initial step in the RV_{GRT} estimator involves estimating $\alpha_\perp$. For this purpose, we employ a VEC(0) process, as outlined in (3), and estimate α by least squares using the 359 days in our sample, meeting the long-span dimension requirement outlined in Theorem 2. Subsequently, following the procedure by Hansen and Lunde (2006), we calculate $\widehat{\alpha}_{\delta,\perp}$ to satisfy the normalization $\iota_N'\widehat{\alpha}_{\delta,\perp} = 1$.

The results are presented in Table 3. In general, a single market does not emerge as the sole driver of the price discovery process, indicating the presence of fragmentation noise across most of the stocks in our sample. The listing exchanges (N and T) are generally identified as the primary contributors to the latent efficient price, whereas the Z and K trading venues usually appear as the least important to the price discovery process. As previously discussed in Subsection 2.2, the price discovery analysis has the potential to hint whether our multivariate approach is beneficial for estimating the integrated variance. In view of the result in Theorem 1, this appears to be the case in this setting, as $\widehat{\alpha}_\perp \neq k(n)$ with $n = 1, \ldots, N$ indicates that the competitor estimators are likely to be inconsistent.

Next, we proceed to the second step of our RV_{GRT} estimator, applying the RV estimator to the estimates of the efficient price computed using $\widehat{\alpha}_{\delta,\perp}$. We then conduct the testing procedure outlined in Subsection 4.2 at the 5% significance level with $k = 1$ (number of false rejections). The QLIKE distances from the RV_{GRT} estimator are reported in Table 4, where $\triangle$ ($\blacktriangle$) denotes a (significant) positive distance, and $\triangledown$ ($\blacktriangledown$) denotes a (significant) negative distance.

The main takeaway from Table 4 is that the RV_{GRT} estimator is never outperformed by any of our competitors, i.e. we cannot reject the null hypothesis at the 5% significance level in favour of H_1^s, which implies a negative distance from the RV_{GRT} estimator. This strongly suggests that considering multiple trading venues to compute the integrated variance is a reliable approach with substantial potential for favorable outcomes.

In terms of the performance of individual estimators, the RV_{GRT} estimator consistently outperforms the kernel-based estimators, particularly in those trading venues that contribute less to the price discovery process. This outcome is somewhat expected, given that these estimators are designed to work at much higher frequencies (1-minute, 1-second, and tick-by-tick sampling intervals), and these trading venues are less informative to the price discovery process. Meanwhile, the standard univariate RV estimators applied on the 5-minute interval are difficult to beat across all exchanges. This is in line with the recent literature emphasizing the robust finite sample performance of the RV estimator (Liu et al., 2015). This observation holds regardless of the importance that trading venues have in the price discovery process. We identify two potential factors contributing to this result in the 5-minute dataset. First, despite being consistent, the least squares estimator of $\alpha_{\delta,\perp}$ tends not to perform well at low frequencies such as the 5-minute interval (Dias et al., 2021), potentially deteriorating the finite sample performance of the RV_{GRT} estimator. This is because the signal at these lower frequencies is already too weak, given that markets in these very liquid stocks update information at seconds or even microseconds intervals, leading to potentially very imprecise estimates (high standard errors). Second, the correlation between markets in discrete time tends to be very high at this low frequency, negatively affecting the performance of the multivariate RV estimator. In other words, in a setting of extremely fast adjusting prices, the multivariate realized variance estimator delivers upward-biased estimates of the continuous-time correlation between markets. This may artificially bring the RV_{GRT} estimates closer to their univariate competitors (see auxiliary result in Theorem 1), contributing to make it more difficult to reject the H_0^s. The solution to this issue would be to increase the sampling frequency for all estimators, thereby obtaining a more accurate comparison between the alternative estimators. However, the RV estimator is not robust to the additive market microstructure noise that is usually present in these higher frequencies, preventing us from adopting such a strategy. In this context, we explicitly address the estimation issue of $\alpha_\perp$ in the RV_{GRT} estimator and increase the sampling frequency for all the kernel-based estimators.

We begin by recognizing that $\alpha_{\delta,\perp} = \alpha_\perp$ for all $0 < \delta < 1$ (Dias et al., 2021). This implies that we can sample prices at shorter intervals (e.g., 1 second) and leverage the stronger signal at this sampling frequency to estimate α_δ by least squares. Subsequently, we utilize these estimates to calculate $\widehat{\alpha}_{\delta,\perp}$, which remains a consistent estimate of $\alpha_\perp$ irrespective of the sampling interval as long as the number of trading days is large enough. Finally, we use the higher-frequency based $\widehat{\alpha}_{\delta,\perp}$ to compute the estimate of the efficient price at the 5-minute sampling frequency, aiming to mitigate the impact of additive market microstructure noise. The results for the estimates of $\alpha_{\delta,\perp}$ for the five different trading venues, computed from prices sampled at 1-second intervals, are displayed in the right-hand panel of Table 3. Across all trading venues, the estimates of $\alpha_{\delta,\perp}$ consistently display much smaller standard errors than those obtained from prices sampled at a 5-minute interval. This finding is in line with Dias et al. (2021) and confirms that the signal used for estimating the speed-of-adjustment parameter α_δ weakens as the length of the sampling interval increases. Similar to the results obtained from prices sampled at a 5-minute interval, we find that the listing venues (N and T) are usually more important on impounding information to the efficient price; and virtually all trading venues actively participate in the price discovery process. These observations strongly suggest the presence of fragmentation noise, implying that the univariate estimators are likely to be inconsistent estimators of the integrated variance.

Next, Table 5 presents the results for Patton's 2011 accuracy test, with RV_{GRT} as the benchmark estimator. Similar to Table 4, we report the QLIKE distances from the RV_{GRT} estimator. Again, our main conclusion remains unchanged, indicating that the QLIKE distances from the RV_{GRT} estimator are never statistically significantly negative. This result validates our previous finding that employing the RV_{GRT} estimator is essentially a safe choice amid the presence of fragmentation noise with the potential to yield more accurate results in finite samples than those based on univariate estimators. More specifically, we find that employing estimates of $\alpha_{\delta,\perp}$ computed from prices sampled at 1-second interval markedly improves the performance of the RV_{GRT} estimator compared to the usual RV estimator. Specifically, the RV_{GRT} estimator improves its performance with respect to the RV estimators, particularly in those trading venues with the smallest contribution to the price discovery process (K, P, and Z). Additionally, the kernel-based estimators substantially improve their performance by adopting prices sampled at the 1-second interval. It is noteworthy that despite using a much richer dataset (prices sampled at a 1-second interval) these estimators are unable to statistically outperform the RV_{GRT} estimator.

To conclude, we find evidence of fragmentation noise across all constituents of the DJIA index. This type of noise, in addition to the usual market microstructure noise observed in high-frequency data, may result in inconsistent estimates of the integrated variance of the efficient price. This scenario justifies the superior performance of the RV_{GRT} estimator compared to both RV and high-frequency kernel-based estimators. Specifically, we recommend computing the RV_{GRT} estimator using estimates of $\alpha_{\delta,\perp}$ derived from prices sampled at a 1-second interval, as they have shown to yield more accurate results compared to those relying solely on 5-minute interval data.

5 Conclusion

This paper identifies a novel source of market microstructure noise: the fragmentation of markets. This type of noise arises in assets that trade simultaneously across multiple markets or derivative markets. The importance of fragmentation noise becomes evident with the proliferation of trading venues in recent years, resulting in quotes being scattered across exchanges, alternative trading systems, and electronic communication networks (see, among others, O'Hara, 2015; Menkveld, 2016). More specifically, our theoretical results demonstrate that fragmentation noise deviates from the conventional class of additive market microstructure noise. Consequently, in the presence of fragmentation noise, both the RV and noise-robust estimators are no longer consistent estimators of the integrated variance. We then put forward a new two-step estimator that consistently estimates the integrated variance in the presence of fragmentation noise. We establish the limiting distribution of the RV_{GRT} estimator under mixed asymptotics.

The results from our simulations and empirical application demonstrate that the proposed multivariate estimation approach performs superior to or on par with a wide range of univariate estimators for assets actively traded in more than one venue. Unlike the commonly applied univariate RV and noise-robust estimators, which typically rely on data from major listing exchanges due to data availability constraints, our findings highlight the importance of explicitly addressing the fragmentation noise and using information from alternative trading platforms.

Further research on the subject could be directed at generalizing the approach presented here. Since we limit our analysis to continuous processes in this paper, one could focus on multivariate estimation that takes jumps into account. Particularly, it could be interesting to investigate how (co-)jumps affect the parameter estimates for the VECM and the overall robustness of the multivariate RV estimation. Moreover, it might be necessary in some empirical applications to relax the assumption of constant adjustment coefficients over the sampling period. A practical solution for this usually consists of estimating the VECM anew for each trading day. However, this approach violates the necessary assumptions needed for consistency. More elegant approaches for VECMs with time-varying coefficients that are able to incorporate intra-day data from multiple trading days are proposed by Krolzig (1997) (Markov-switching VECM) or Bierens and Martins (2010) (Time-varying VECM), among others.

Acknowledgements

The authors thank Thomas Dimpfl, Konstantin Kuck, Monica Gehde-Trapp, Robert Jung, Joachim Grammig, Roxana Halbleib, Cristina Scherrer as well as the participants of the 8th Finance Workshop in Ulm, the University of Freiburg's Doctoral Seminar, the SoFiE 2022 Annual Conference, the Econometric Society 2022 Australasia Meeting, the AMES 2022 Tokyo, and the Joint Econometrics Seminar in Hohenheim for comments and suggestions. Further, they thank Domenic Franjic for excellent research assistance. Access to Refinitiv Tick History, provided by the Hohenheim Datalab (DALAHO), is gratefully acknowledged.

References

Aït-Sahalia, Y., Jacod, J., 2014. High-Frequency Financial Econometrics. Princeton University Press.

Aït-Sahalia, Y., Xiu, D., 2019. A Hausman Test for the Presence of Market Microstructure Noise in High Frequency Data. Journal of Econometrics 211, 176–205. doi:10.10.1016/j.jeconom.2018.12.013.

Amihud, Y., Mendelson, H., 1987. Trading mechanisms and stock returns: an empirical investigation. Journal of Finance 42, 533–553. doi:10.2307/2328369.

Andersen, T.G., Archakov, I., Cebiroglu, G., Hautsch, N., 2022. Local mispricing and microstructural noise: A parametric perspective. Journal of Econometrics 230, 510–534. doi:https://doi.org/10.1016/j.jeconom.2021.06.006.

Andersen, T.G., Bollerslev, T., Diebold, F.X., 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. doi:10.1162/rest.89.4.701.

Andersen, T.G., Bollerslev, T., Diebold, F.X., Labys, P., 2001. The distribution of realized exchange rate volatility. Journal of the American Statistical Association 96, 42–55. doi:10.1198/016214501750332965.

Andersen, T.G., Bollerslev, T., Diebold, F.X., Labys, P., 2003. Modeling and forecasting realized volatility. Econometrica 71, 579–625. doi:10.1111/1468-0262.00418.

Bandi, F.M., Phillips, P.C.B., 2003. Fully nonparametric estimation of scalar diffusion models. Econometrica 71, 241–283. doi:10.1111/1468-0262.00395.

Bandi, F.M., Russell, J.R., 2006. [Realized Variance and Market Microstructure Noise]: Comment. Journal of Business & Economic Statistics 24, 167–173. doi:10.1198/07350010600000.

Bandi, F.M., Russell, J.R., 2008. Microstructure noise, realized variance, and optimal sampling. Review of Economic Studies 75, 339–369. doi:10.1111/j.1467-937X.2008.00474.x.

Bandi, F.M., Russell, J.R., 2011. Market microstructure noise, integrated variance estimators, and the accuracy of asymptotic approximations. Journal of Econometrics 160, 145–159. doi:10.1016/j.jeconom.2010.03.027.

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. doi:10.3982/ecta6495.

Barndorff-Nielsen, O.E., Hansen, P.R., Lunde, A., Shephard, N., 2009. Realized kernels in practice: Trades and quotes. Econometrics Journal 12, 1–32. doi:10.1111/j.1368-423X.2008.00275.x.

Barndorff-Nielsen, O.E., Hansen, P.R., Lunde, A., Shephard, N., 2011a. 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. doi:10.1016/j.jeconom.2010.07.009.

Barndorff-Nielsen, O.E., Hansen, P.R., Lunde, A., Shephard, N., 2011b. Subsampling realised kernels. Journal of Econometrics 160, 204–219. doi:10.1016/j.jeconom.2010.03.031.

Barndorff-Nielsen, O.E., Shepard, N., 2002. Econometric analysis of realised volatility and its use in estimating stochastic volatility models. Journal of the Royal Statistical Society B 64, 253–280. doi:10.1111/1467-9868.00336.

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. doi:10.1111/j.1468-0262.2004.00515.x.

Bierens, H.J., Martins, L.F., 2010. Time-Varying Cointegration. Econometric Theory 26, 1453–1490. doi:10.1017/S0266466609990648.

Chambers, M.J., 2011. Cointegration and sampling frequency. Econometrics Journal 14, 156–185. doi:10.1111/j.1368-423X.2010.00329.x.

Christensen, K., Kinnebrock, S., Podolskij, M., 2010. Pre-averaging estimators of the ex-post covariance matrix in noisy diffusion models with non-synchronous data. Journal of Econometrics 159, 116–133. doi:https://doi.org/10.1016/j.jeconom.2010.05.001.

Christensen, K., Podolskij, M., 2007. Realized range-based estimation of integrated variance. Journal of Econometrics 141, 323–349. doi:10.1016/j.jeconom.2006.06.012.

Corradi, V., 1997. Comovements Between Diffusion Processes. Econometric Theory 13, 646–666. doi:10.1017/s0266466600006113.

Corsi, F., Mittnik, S., Pigorsch, C., Pigorsch, U., 2008. The volatility of realized volatility. Econometric Reviews 27, 46–78. doi:10.1080/07474930701853616.

De Jong, F., 2002. Measures of contributions to price discovery: A comparison. Journal of Financial Markets 5, 323–327. doi:10.1016/S1386-4181(02)00028-9.

Dias, G.F., Fernandes, M., Scherrer, C.M., 2021. Price Discovery in a Continuous-Time Setting. Journal of Financial Econometrics 19, 985–1008. doi:10.1093/jjfinec/nbz030.

Dias, G.F., Fernandes, M., Scherrer, C.M., 2022. Price discovery with a richer market microstructure noise. SSRN Working Paper , 1–72doi:10.2139/ssrn.3864966.

Diebold, F.X., Strasser, G., 2013. On the Correlation Structure of Microstructure Noise: A Financial Economic Approach. Review of Economic Studies 80, 1304–1337. doi:10.1093/restud/rdt008.

Dorogovtsev, A.Y., 1978. The consistency of an estimate of a parameter of a stochastic differential equation. Theory of Probability and Mathematical Statistics 10, 73–82.

Easley, D., O'Hara, M., 1987. Price, trade size, and information in securities markets. Journal of Financial Economics 19, 69–90. doi:10.1016/0304-405X(87)90029-8.

Glosten, L.R., Milgrom, P.R., 1985. Bid, ask and transaction prices in a specialist market with heterogeneously informed traders. Journal of Financial Economics 14, 71–100. doi:10.1016/0304-405X(85)90044-3.

Goncalves, S., Meddahi, N., 2009. Bootstrapping Realized Volatility. Econometrica 77, 283–306. doi:10.3982/ecta5971.

Hafner, C.M., Herwartz, H., 2009. Testing for linear vector autoregressive dynamics under multivariate generalized autoregressive heteroskedasticity. Statistica Neerlandica 63, 294–323. doi:10.1111/j.1467-9574.2009.00424.x.

Hansen, P.R., Lunde, A., 2006. Realized variance and market microstructure noise. Journal of Business and Economic Statistics 24, 127–161. doi:10.1198/073500106000000071.

Harris, F.H.d., 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. doi:10.1016/S1386-4181(01)00017-9.

Hasbrouck, J., 1995. One security, many markets: Determining the contributions to price discovery. Journal of Finance 50, 1175–1199. doi:10.2307/2329348.

Hasbrouck, J., 2002. Stalking the efficient price in market microstructure specifications: An overview. Journal of Financial Markets 5, 329–339. doi:10.1016/S1386-4181(02)00029-0.

Hasbrouck, J., 2003. Intraday price formation in u.s. equity index markets. The Journal of Finance 58, 2375–2400. doi:10.1046/j.1540-6261.2003.00609.x.

Huang, X., Tauchen, G., 2005. The relative contribution of jumps to total price variance. Journal of Financial Econometrics 3, 456–499. doi:10.1093/jjfinec/nbi025.

Jacod, J., Li, Y., Mykland, P.A., Podolskij, M., Vetter, M., 2009. Microstructure noise in the continuous case: The pre-averaging approach. Stochastic Processes and their Applications 119, 2249–2276. doi:10.1016/j.spa.2008.11.004.

Jacquier, E., Polson, N.G., Rossi, P.E., 1994. Bayesian analysis of stochastic volatility models. Journal of Business & Economic Statistics 20, 69–87. doi:10.1198/073500102753410408.

Kessler, M., 1997. Estimation of an ergodic diffusion from discrete observations. Scandinavian Journal of Statistics 24, 211–229. doi:10.1111/1467-9469.00059.

Kessler, M., Rahbek, A., 2001. Asymptotic Likelihood Based Inference for Co-integrated Homogenous Gaussian Diffusions. Scandinavian Journal of Statistics 28, 455–470. doi:10.1111/1467-9469.00248.

Kessler, M., Rahbek, A., 2004. Identification and Inference for Multivariate Cointegrated and Ergodic Gaussian Diffusions. Statistical Inference for Stochastic Processes 7, 137–151. doi:10.1023/b:sisp.0000026044.28647.56.

Krolzig, H.M., 1997. Markov-Switching Vector Autoregressions: Modelling, Statistical Inference, and Application to Business Cycle Analysis. Springer, New York.

Kyle, A.S., 1985. Continuous Auctions and Insider Trading. Econometrica 53, 1315–1335. doi:10.2307/1913210.

Liu, L.Y., Patton, A.J., Sheppard, K., 2015. Does anything beat 5-minute RV? A comparison of realized measures across multiple asset classes. Journal of Econometrics 187, 293–311. doi:10.1016/j.jeconom.2015.02.008.

Martens, M., van Dijk, D., 2007. Measuring volatility with the realized range. Journal of Econometrics 138, 181–207. doi:10.1016/j.jeconom.2006.05.019.

McAleer, M., Medeiros, M.C., 2008. Realized Volatility: A Review. Econometric Reviews 27, 10–45. doi:10.1080/07474930701853509.

Menkveld, A.J., 2016. The Economics of High-Frequency Trading: Taking Stock. Annual Review of Financial Economics 8, 1–24. doi:10.1146/annurev-financial-121415-033010.

Nguenang, C., 2016. Evaluating Market's Contribution to Price Discovery in High-Frequency for Co-Listed Assets. SSRN Working Paper doi:10.2139/ssrn.2788879.

O'Hara, M., 2015. High frequency market microstructure. Journal of Financial Economics 116, 257–270. doi:10.1016/j.jfineco.2015.01.003.

Patton, A.J., 2011. Data-based ranking of realised volatility estimators. Journal of Econometrics 161, 284–303. doi:10.1016/j.jeconom.2010.12.010.

Phillips, P.C.B., 1991. Error Correction and Long-Run Equilibrium in Continuous Time. Econometrica 59, 967–980. doi:10.2307/2938169.

Phillips, P.C.B., Yu, J., 2009. A two-stage realized volatility approach to estimation of diffusion processes with discrete data. Journal of Econometrics 150, 139–150. doi:10.1016/j.jeconom.2008.12.006.

Politis, D.N., Romano, J.P., 1994. The Stationary Bootstrap. Journal of the American Statistical Association 89, 1303–1313. doi:10.2307/2290993.

Prakasa Rao, B.L.S., 1983. Asymptotic theory for non-linear least squares estimator for diffusion processes. Statistics: A Journal of Theoretical and Applied Statistics 14, 195–209. doi:10.1080/02331888308801695.

Romano, J.P., Wolf, M., 2005. Stepwise multiple testing as formalized data snooping. Econometrica 73, 1237–1282. doi:10.1111/j.1468-0262.2005.00615.x.

Scherrer, C.M., 2021. Information processing on equity prices and exchange rate for cross-listed stocks. Journal of FInancial Markets 54. doi:10.1016/j.finmar.2021.100634.

White, H., 1980. A Heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. Econometrica 48, 817–838. doi:10.2307/1912934.

Xiu, D., 2010. Quasi-maximum likelihood estimation of volatility with high frequency data. Journal of Econometrics 159, 235–250. doi:https://doi.org/10.1016/j.jeconom.2010.07.002.

Yoshida, N., 1992. Estimation for diffusion processes from discrete observation. Journal of Multivariate Analysis 41, 220–242. doi:10.1016/0047-259X(92)90068-Q.

Zhou, B., 1996. High-frequency data and volatility in foreign-exchange rates. Journal of Business and Economic Statistics 14, 45–52. doi:10.1080/07350015.1996.10524628.

Table 1: Simulation Results: Bias and relative efficiency of RV_{GRT} estimator for the cointegrated OU process

	DGP I			DGP II		
		$\hat{P}^*$			$\hat{P}^*$	
RV_{GRT}	**-0.046 (0.473)**	–	–	**-0.108 (0.677)**	–	–
		P_1			P_1	
RV	3.297 (3.445)	–	–	0.964 (1.266)	–	–
	$H=3$	$H=8$	$H=20$	$H=3$	$H=8$	$H=20$
RV_{AC}	3.029 (3.398)	2.687 (3.332)	2.104 (3.496)	0.865 (1.565)	0.754 (1.938)	0.555 (2.846)
RV_{NW}	3.163 (3.372)	2.994 (3.293)	2.685 (3.121)	0.915 (1.338)	0.862 (1.475)	0.756 (1.710)
RK_{MTH}	3.271 (3.760)	3.021 (3.803)	2.519 (3.824)	1.065 (1.881)	0.979 (2.209)	0.816 (2.779)
RK_P	3.235 (3.766)	2.950 (3.835)	2.366 (3.866)	1.050 (1.925)	0.958 (2.325)	0.764 (2.978)
		P_2			P_2	
RV	2.038 (2.196)	–	–	0.777 (1.121)	–	–
	$H=3$	$H=8$	$H=20$	$H=3$	$H=8$	$H=20$
RV_{AC}	1.885 (2.263)	1.693 (2.377)	1.320 (2.692)	0.712 (1.458)	0.640 (1.897)	0.464 (2.746)
RV_{NW}	1.969 (2.188)	1.859 (2.189)	1.673 (2.140)	0.751 (1.217)	0.703 (1.357)	0.622 (1.615)
RK_{MTH}	2.121 (2.604)	1.951 (2.706)	1.638 (2.895)	0.944 (1.789)	0.861 (2.108)	0.725 (2.696)
RK_P	2.098 (2.622)	1.902 (2.750)	1.546 (2.986)	0.932 (1.840)	0.839 (2.215)	0.685 (2.902)

Note: The table reports the estimated average bias ($\times 10^5$) over the full sampling period with the RMSE ($\times 10^5$) given in parentheses. Data are drawn from DGP I with $\alpha = (-0.5, 0.5)'$ and $\tau = 1.25$ and from DGP II with $\alpha = (-0.9, 0.1)'$ and $\tau = 1$. H denotes the lag truncation parameter used for realized kernel estimators. The estimators with the lowest bias and RMSE are marked in boldface. Finally, the true values, scaled up by 10^5, read $\langle P^*, P^{*\prime} \rangle_{t-1:t} = 3.086$, $\langle P, P' \rangle^{1,1}_{t-1:t} = 6.436$ and $\langle P, P' \rangle^{2,2}_{t-1:t} = 5.152$ for DGP I and $\langle P^*, P^{*\prime} \rangle_{t-1:t} = 4.282$, $\langle P, P' \rangle^{1,1}_{t-1:t} = 5.184$ and $\langle P, P' \rangle^{2,2}_{t-1:t} = 5.136$ for DGP II.

Table 2: Simulation Results: Bias and relative efficiency of RV_{GRT} estimator for the cointegrated OU type of process with stochastic volatility

$\alpha_\perp$	RV_{GRT}	$RV^{(1)}$	$RV^{(1)}_{AC}$	$RV^{(1)}_{NW}$	$RK^{(1)}_{MTH}$	$RK^{(1)}_{P}$	$RV^{(2)}$	$RV^{(2)}_{AC}$	$RV^{(2)}_{NW}$	$RK^{(2)}_{MTH}$	$RK^{(2)}_{P}$
					Panel A: Bias						
$(0.1, 0.9)'$	**-0.039**	0.100	0.090	0.103	0.324	0.324	0.090	0.073	0.087	0.304	0.302
$(0.2, 0.8)'$	**-0.031**	0.172	0.157	0.172	0.391	0.390	0.167	0.146	0.162	0.377	0.373
$(0.3, 0.7)'$	**-0.024**	0.222	0.204	0.221	0.438	0.437	0.223	0.198	0.216	0.429	0.426
$(0.4, 0.6)'$	**-0.017**	0.252	0.232	0.249	0.465	0.465	0.258	0.231	0.250	0.462	0.458
$(0.5, 0.5)'$	**-0.011**	0.260	0.240	0.258	0.473	0.473	0.272	0.244	0.264	0.476	0.472
					Panel B: RMSE						
$(0.1, 0.9)'$	**0.268**	0.274	0.445	0.330	0.524	0.547	0.276	0.445	0.335	0.535	0.556
$(0.2, 0.8)'$	**0.265**	0.276	0.446	0.333	0.528	0.550	0.279	0.445	0.338	0.539	0.559
$(0.3, 0.7)'$	**0.261**	0.279	0.446	0.335	0.531	0.552	0.283	0.446	0.341	0.542	0.562
$(0.4, 0.6)'$	**0.259**	0.281	0.447	0.336	0.532	0.554	0.286	0.447	0.342	0.544	0.565
$(0.5, 0.5)'$	**0.259**	0.281	0.447	0.336	0.533	0.554	0.288	0.448	0.344	0.545	0.566

This table reports the average bias and RMSE ($\times 10^5$) for several RV estimators over the full sampling period. The results are based on 1000 draws from the stochastic volatility two-market model with $\rho = 0.9$. The superscript (1) and (2) denote the estimators for the price series $P_{1,t}$ and $P_{2,t}$, respectively. The lag truncation parameter for realized kernel estimators is set optimally according to the rule-of-thumb suggested in Hansen and Lunde (2006) and Barndorff-Nielsen et al. (2008). The estimators with the lowest bias and RMSE are marked in boldface.

Table 3: Component shares and residual correlation for the DJIA stocks

	5-min data						1-sec data					
	N	T	K	P	Z	$\bar{\rho}$	N	T	K	P	Z	$\bar{\rho}$
AAPL	0.195	−0.155	−0.203	0.726	0.437	0.995	0.022	0.303	0.232	0.226	0.217	0.876
	(0.189)	(0.163)	(0.178)	(0.164)	(0.097)		(0.001)	(0.001)	(0.001)	(0.001)	(0.003)	
AMGN	0.111	0.503	0.071	0.077	0.239	0.874	0.052	0.633	0.071	0.115	0.129	0.384
	(0.009)	(0.008)	(0.008)	(0.007)	(0.015)		(0.002)	(0.001)	(0.002)	(0.001)	(0.002)	
AXP	0.471	0.450	0.051	0.057	−0.029	0.951	0.365	0.327	0.062	0.114	0.133	0.553
	(0.014)	(0.014)	(0.012)	(0.014)	(0.024)		(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	
BA	0.813	0.332	−0.005	−0.041	−0.098	0.989	0.574	0.130	0.170	0.059	0.067	0.620
	(0.044)	(0.050)	(0.048)	(0.039)	(0.039)		(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	
CAT	0.950	0.052	−0.027	0.011	0.015	0.945	0.564	0.182	0.073	0.023	0.157	0.472
	(0.038)	(0.035)	(0.025)	(0.016)	(0.028)		(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	
CRM	0.869	0.260	−0.001	−0.167	0.040	0.977	0.424	0.260	0.066	0.172	0.079	0.565
	(0.044)	(0.052)	(0.034)	(0.035)	(0.032)		(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	
CSCO	0.168	−0.319	0.409	0.433	0.309	0.988	0.121	0.234	0.186	0.224	0.235	0.877
	(0.039)	(0.038)	(0.043)	(0.043)	(0.043)		(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	
CVX	0.922	−0.098	0.087	−0.006	0.095	0.991	0.423	0.237	0.134	0.080	0.126	0.862
	(0.054)	(0.073)	(0.062)	(0.056)	(0.051)		(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	
DIS	0.914	0.027	0.139	−0.014	−0.066	0.987	0.341	0.280	0.059	0.077	0.242	0.773
	(0.04)	(0.055)	(0.044)	(0.039)	(0.039)		(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	
DOW	0.768	0.099	−0.024	0.025	0.132	0.972	0.538	0.294	0.054	0.043	0.072	0.640
	(0.039)	(0.033)	(0.029)	(0.023)	(0.033)		(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	
GS	0.912	0.023	0.063	0.005	−0.003	0.933	0.550	0.237	0.065	0.028	0.120	0.390
	(0.034)	(0.03)	(0.017)	(0.018)	(0.022)		(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	
HD	0.669	0.115	0.021	0.077	0.118	0.952	0.424	0.220	0.114	0.071	0.171	0.420
	(0.027)	(0.02)	(0.018)	(0.015)	(0.028)		(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	
HON	0.132	0.538	0.142	0.097	0.091	0.898	0.080	0.537	0.074	0.035	0.275	0.378
	(0.012)	(0.008)	(0.007)	(0.007)	(0.022)		(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	
IBM	0.870	0.060	−0.030	0.012	0.089	0.970	0.592	0.185	0.041	0.069	0.112	0.618
	(0.045)	(0.038)	(0.027)	(0.03)	(0.028)		(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	
INTC	0.088	0.619	0.105	−0.026	0.214	0.995	0.140	0.207	0.281	0.233	0.139	0.916
	(0.064)	(0.046)	(0.063)	(0.044)	(0.055)		(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	
JNJ	0.931	−0.059	0.101	0.024	0.003	0.964	0.553	0.216	0.044	0.070	0.117	0.626
	(0.038)	(0.034)	(0.024)	(0.023)	(0.028)		(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	
JPM	0.926	0.097	0.190	−0.278	0.065	0.990	0.348	0.382	0.113	0.059	0.098	0.849
	(0.07)	(0.08)	(0.061)	(0.054)	(0.044)		(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	
KO	0.980	−0.168	0.246	−0.018	−0.040	0.985	0.323	0.168	0.173	0.196	0.140	0.867
	(0.053)	(0.056)	(0.045)	(0.05)	(0.033)		(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	
MCD	0.935	0.001	0.075	−0.014	0.004	0.922	0.680	0.114	0.052	0.082	0.073	0.434
	(0.028)	(0.018)	(0.017)	(0.020)	(0.020)		(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	
MMM	0.892	0.028	0.032	0.039	0.009	0.905	0.671	0.152	0.056	0.025	0.096	0.410
	(0.025)	(0.018)	(0.015)	(0.014)	(0.019)		(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	
MRK	0.983	−0.230	0.008	0.078	0.162	0.987	0.400	0.275	0.088	0.061	0.176	0.846
	(0.064)	(0.071)	(0.05)	(0.056)	(0.044)		(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	
MSFT	0.099	0.547	0.270	0.115	−0.031	0.991	0.018	0.634	0.072	0.132	0.144	0.866
	(0.042)	(0.037)	(0.044)	(0.025)	(0.058)		(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	
NKE	0.908	−0.131	0.201	−0.054	0.075	0.980	0.636	0.134	0.049	0.066	0.116	0.705
	(0.055)	(0.051)	(0.036)	(0.036)	(0.039)		(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	
PG	0.912	−0.015	0.024	0.017	0.061	0.963	0.549	0.225	0.045	0.073	0.108	0.625
	(0.034)	(0.037)	(0.024)	(0.022)	(0.032)		(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	
TRV	0.677	0.166	0.030	0.022	0.105	0.801	0.628	0.197	0.030	0.030	0.115	0.364
	(0.014)	(0.012)	(0.008)	(0.008)	(0.015)		(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	
UNH	0.909	−0.047	0.030	0.046	0.063	0.891	0.452	0.233	0.086	0.105	0.124	0.375
	(0.024)	(0.017)	(0.015)	(0.016)	(0.017)		(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	
V	0.977	−0.108	−0.031	0.036	0.127	0.974	0.722	0.126	0.040	0.077	0.035	0.581
	(0.060)	(0.053)	(0.034)	(0.040)	(0.030)		(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	
VZ	0.779	0.141	0.187	−0.101	−0.006	0.986	0.286	0.121	0.258	0.181	0.154	0.900
	(0.043)	(0.044)	(0.045)	(0.044)	(0.031)		(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	
WBA	−0.100	0.586	0.149	0.092	0.273	0.956	0.029	0.654	0.099	0.110	0.108	0.612
	(0.019)	(0.016)	(0.016)	(0.009)	(0.024)		(0.002)	(0.001)	(0.001)	(0.001)	(0.002)	
WMT	0.685	0.159	0.076	−0.054	0.134	0.977	0.524	0.274	0.066	0.066	0.070	0.654
	(0.038)	(0.037)	(0.030)	(0.024)	(0.030)		(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	

This tables reports the component share of the cross-listed DJIA stocks at the 5-min and 1-sec sampling frequencies. The standard errors of the component shares are given in parentheses (computed with the delta-method). The average residual correlation across all markets is denoted with $\bar{\rho}$.

Table 4: Tests of equal accuracy for DJIA stocks: 5-minute sampling interval

DJIA Stocks

	RV_N	RV_N^{AC}	RV_N^{NW}	RK_N^{MTH}	RK_N^P	RV_T	RV_T^{AC}	RV_T^{NW}	RK_T^{MTH}	RK_T^P	RV_K	RV_K^{AC}
AAPL	▽	▲	△	▲	▲	▽	▲	▲	▲	▲	▽	▲
AMGN	△	△	△	△	△	△	△	△	△	△	△	▽
AXP	△	△	△	▲	▲	▽	△	▽	▲	▲	△	△
BA	△	▲	▲	▲	▲	▽	▲	▲	▲	▲	△	▲
CAT	▽	△	△	△	△	▽	△	▽	△	△	▽	△
CRM	▽	△	△	△	△	△	▲	△	△	△	▽	△
CSCO	△	▲	△	▲	▲	△	▲	△	▲	▲	△	▲
CVX	▽	▲	△	▲	▲	▽	▲	△	▲	▲	▽	▲
DIS	△	▲	▲	▲	▲	▽	▲	▲	▲	▲	▽	▲
DOW	△	△	△	△	△	▽	△	△	△	△	▽	△
GS	▽	△	△	▲	▲	▽	△	△	▲	▲	△	△
HD	△	▲	△	▲	▲	▽	▲	△	▲	▲	▽	△
HON	△	△	△	△	△	△	△	△	△	△	▽	△
IBM	▽	△	△	▲	▲	△	△	△	▲	▲	▽	△
INTC	▽	△	△	△	△	△	△	△	△	△	△	△
JNJ	▽	△	△	▲	▲	△	△	△	▲	▲	△	△
JPM	▽	▲	△	▲	▲	▽	△	△	▲	▲	▽	▲
KO	▽	△	△	△	△	▽	△	△	△	△	▽	△
MCD	▽	△	△	△	△	△	△	△	△	△	▽	△
MMM	▽	▲	△	▲	▲	△	▲	△	▲	▲	△	▲
MRK	△	△	△	△	△	△	△	△	△	△	△	△
MSFT	▽	△	△	△	△	△	△	△	△	△	▽	△
NKE	▽	△	△	▲	▲	▽	△	△	▲	▲	▽	▲
PG	▽	▲	▲	▲	▲	△	▲	▲	▲	▲	▽	▲
TRV	▽	△	△	▲	▲	△	△	△	▲	▲	▲	▲
UNH	▽	△	△	▲	▲	△	△	△	▲	▲	△	△
V	△	△	△	▲	▲	▽	△	△	▲	▲	▽	△
VZ	▽	△	△	▲	▲	△	△	△	▲	▲	△	△
WBA	△	△	△	▲	▲	▽	▽	▽	△	△	△	△
WMT	△	▲	△	▲	▲	△	▲	△	▲	▲	▽	▲

	RV_K^{NW}	RK_K^{MTH}	RK_K^P	RV_P	RV_P^{AC}	RV_P^{NW}	RK_P^{MTH}	RK_P^P	RV_Z	RV_Z^{AC}	RV_Z^{NW}	RK_Z^{MTH}	RK_Z^P
AAPL	▲	▲	▲	▽	▲	▲	▲	▲	△	▲	▲	▲	▲
AMGN	▽	△	△	△	△	△	△	△	△	▽	▽	▽	▽
AXP	△	▲	▲	▽	△	▽	△	△	▽	△	△	▲	▲
BA	▲	▲	▲	▽	▲	▲	▲	▲	▽	▲	▲	▲	▲
CAT	▽	△	△	▽	▽	▽	△	△	▽	△	△	△	△
CRM	△	△	△	▽	△	△	△	△	▽	△	△	△	△
CSCO	▲	▲	▲	▽	△	△	▲	▲	△	▲	△	▲	▲
CVX	▲	▲	▲	▽	▲	△	▲	▲	▽	▲	▲	▲	▲
DIS	▲	▲	▲	△	▲	▲	▲	▲	△	▲	▲	▲	▲
DOW	△	△	△	▽	△	▽	△	△	▽	△	▽	△	△
GS	△	▲	▲	▽	▽	▽	▲	▲	▽	△	▽	▲	▲
HD	△	▲	▲	▽	△	▽	▲	▲	▽	▲	△	▲	▲
HON	▽	△	△	▲	△	△	▲	▲	▽	△	△	△	△
IBM	△	▲	▲	△	△	△	▲	▲	△	△	△	▲	▲
INTC	△	△	△	▽	△	△	△	△	▽	△	△	△	△
JNJ	△	△	△	△	△	△	▲	▲	▽	△	△	△	△
JPM	△	▲	▲	▽	△	△	▲	▲	▽	▲	△	▲	▲
KO	△	△	△	▽	△	△	△	△	▽	△	▽	△	△
MCD	△	△	△	△	△	△	△	△	△	△	△	△	△
MMM	△	▲	▲	▲	▲	▲	▲	▲	△	▲	△	▲	▲
MRK	△	△	△	▽	△	△	△	△	▽	△	△	△	△
MSFT	△	△	△	△	△	△	△	△	▽	△	△	△	△
NKE	△	▲	▲	▽	△	△	▲	▲	▽	△	△	▲	▲
PG	△	▲	▲	△	▲	▲	▲	▲	△	▲	▲	▲	▲
TRV	△	▲	▲	▲	△	▲	▲	▲	△	△	△	▲	▲
UNH	△	▲	▲	▽	▽	▽	▲	▲	△	△	△	▲	▲
V	△	▲	▲	△	△	△	▲	▲	▽	△	▽	△	△
VZ	△	▲	▲	△	△	△	▲	▲	△	▲	△	▲	▲
WBA	▽	△	△	△	▽	▽	△	△	△	△	▽	△	△
WMT	△	▲	▲	△	▲	▲	▲	▲	▽	△	△	▲	▲

This table reports the QLIKE distances from the benchmark estimator (RV_{GRT}). △ (▲) denotes a (significant) positive distance and ▽ (▼) denotes a (significant) negative distance, respectively. For example, ▲ means that the RV_{GRT} estimator significantly outperforms a given competitor. The squared daily return of the estimated efficient price serves as the proxy for the daily integrated variance. The test is conducted at the 5% significance level and $k = 1$ (number of false rejections). We employ a stationary bootstrap (1000 draws) for the evaluation of our test statistics with an average block length of ten days. The RV_{GRT}, RV and the kernel-based estimators are all computed using data sampled at 5-minute intervals.

Table 5: Tests of equal accuracy for DJIA stocks: mixed sampling intervals

DJIA Stocks																									
	RV_N	RV_N^{AC}	RV_N^{NW}	RK_N^{MTH}	RK_N^P	RV_T	RV_T^{AC}	RV_T^{NW}	RK_T^{MTH}	RK_T^P	RV_K	RV_K^{AC}	RV_K^{NW}	RK_K^{MTH}	RK_K^P	RV_P	RV_P^{AC}	RV_P^{NW}	RK_P^{MTH}	RK_P^P	RV_Z	RV_Z^{AC}	RV_Z^{NW}	RK_Z^{MTH}	RK_Z^P
AAPL	▲	▲	▲	△	△	▽	▽	▽	△	△	▽	▽	▽	△	△	▽	▽	▽	△	△	▽	▽	▽	△	△
AMGN	▲	▲	▲	△	△	△	▽	▽	▽	▽	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
AXP	△	▽	▽	▽	▽	△	▽	▽	▽	▽	▲	△	△	△	△	▲	△	△	▽	▽	△	△	△	▽	▽
BA	▽	▽	▽	▽	▽	△	▽	▽	▽	▽	▽	▽	▽	▽	▽	▲	△	△	▽	▽	▲	△	△	▽	▽
CAT	▽	▽	▽	▽	▽	▲	▽	△	▽	▽	▲	▲	▲	△	△	▲	▲	▲	▲	▲	△	▽	▽	▽	▽
CRM	△	▽	▽	▽	▽	△	▽	▽	▽	▽	▲	△	▲	△	△	▲	△	△	▽	▽	▲	△	▲	▽	▽
CSCO	△	▽	▽	▽	▽	▽	▽	▽	▽	▽	▽	▽	▽	△	△	▽	▽	▽	▽	▽	▽	▽	▽	▽	▽
CVX	▽	▽	▽	△	△	▽	▽	▽	▽	▽	▽	▽	▽	▽	▽	△	▽	△	▽	▽	△	▽	▽	▽	▽
DIS	▽	▽	▽	▽	▽	▽	▽	▽	▽	▽	△	▽	▽	▽	▽	△	▽	△	▽	▽	▽	▽	▽	▽	▽
DOW	△	▽	▽	△	△	△	△	△	△	△	▲	▲	▲	△	△	▲	▲	▲	▲	▲	▲	▲	▲	△	△
GS	△	▽	▽	▽	▽	▲	△	▲	▽	▽	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	△	△
HD	△	▽	△	△	△	▲	△	△	△	△	▲	△	▲	△	△	▲	▲	▲	△	△	▲	△	△	▽	▽
HON	▲	▲	▲	▲	▲	▲	▽	▽	▽	▽	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	△	▲	▽	▽
IBM	▽	▽	▽	▽	▽	△	▽	△	▽	▽	▲	▲	▲	△	△	▲	△	▲	△	△	▲	△	△	▽	▽
INTC	△	▽	▽	▽	▽	△	▽	▽	▽	▽	▽	▽	▽	▽	▽	△	▽	▽	▽	▽	△	▽	▽	△	△
JNJ	△	▽	▽	▽	▽	▲	△	△	▽	▽	▲	▲	▲	△	△	▲	▲	▲	△	△	▲	△	△	△	△
JPM	▽	▽	▽	▽	▽	▽	▽	▽	▽	▽	▽	▽	▽	▽	▽	▽	▽	▽	▽	▽	▽	▽	▽	▽	▽
KO	△	▽	▽	▽	▽	▽	▽	▽	▽	▽	△	△	△	△	△	△	▽	▽	▽	▽	△	▽	△	△	△
MCD	△	▽	▽	▽	▽	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
MMM	△	▽	▽	▽	▽	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
MRK	▽	▽	▽	▽	▽	▽	▽	▽	▽	▽	△	△	△	▽	▽	▲	△	△	△	△	▽	▽	▽	▽	▽
MSFT	▲	△	△	▽	▽	▽	▽	▽	△	△	▽	▽	▽	▽	▽	▽	▽	▽	▽	▽	▽	▽	▽	▽	▽
NKE	▽	▽	▽	▽	▽	▽	▽	▽	▽	▽	△	▽	▽	▽	▽	△	▽	▽	▽	▽	▽	▽	▽	▽	▽
PG	▽	▽	▽	▽	▽	△	△	△	▽	▽	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	△	▲	△	△
TRV	△	▽	▽	▽	▽	▲	▲	▲	△	△	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	△	△
UNH	▲	△	△	▽	▽	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
V	▽	▽	▽	▽	▽	△	▽	▽	▽	▽	▲	△	△	▽	▽	△	▽	△	▽	▽	▲	▲	▲	△	△
VZ	△	▽	▽	△	△	△	▽	▽	△	△	△	▽	△	△	△	△	▽	▽	△	△	△	△	△	△	△
WBA	▲	▲	▲	▲	▲	▽	▽	▽	▽	▽	▲	▲	▲	△	△	▲	▲	▲	△	△	▲	▲	▲	▽	▽
WMT	▽	▽	▽	▽	▽	△	▽	△	▽	▽	▲	▲	▲	△	△	▲	▲	▲	▲	▲	▲	▲	▲	△	△

This table reports the QLIKE distances from the benchmark estimator (RV_{GRT}). △ (▲) denotes a (significant) positive distance and ▽ (▼) denotes a (significant) negative distance, respectively. For example, ▲ means that the RV_{GRT} estimator significantly outperforms a given competitor. The squared daily return of the estimated efficient price serves as the proxy for the daily integrated variance. The test is conducted at the 5% significance level and $k = 1$ (number of false rejections). We employ a stationary bootstrap (1000 draws) for the evaluation of our test statistics with an average block length of ten days. The RV_{GRT} estimator is computed by estimating $\alpha_\perp$ using data sampled at 1-second intervals for the first step, while the second step utilizes estimates of the efficient price returns sampled at 5-minute intervals. The RV and the kernel-based estimators are computed using data sampled at 5-minute and 1-second intervals, respectively.

A Proofs

Lemma 1. Under Assumption 1 and Assumption 2, $T \to \infty$, $\delta \to 0$, and $T\delta^2 \to 0$, it holds that

$$\text{(i)} \qquad \sqrt{Tm} V_{Tm} W_{Tm}^{-1/2} \operatorname{vec}(\hat{\alpha}_\delta - \alpha_\delta) \xrightarrow{d} \mathcal{N}(0, I_{(N-1)^2 N^2}),$$

and

$$\text{(ii)} \qquad \hat{\alpha}_{\delta,\perp} \xrightarrow{p} \alpha_\perp.$$

Proof of Lemma 1. If β is known, we can rewrite the discretely sampled VECM as follows

$$\begin{aligned} \Delta P_{t_i} &= \Pi_\delta P_{t_{i-1}} + \epsilon_{t_i}, \qquad i = 1, \ldots, m, \quad t = 1, \ldots, T \\ &= \alpha_\delta \beta' P_{t_{i-1}} + \epsilon_{t_i} \\ &= \alpha_\delta z_{t_{i-1}} + \epsilon_{t_i}, \end{aligned} \tag{26}$$

where ϵ_{t_i} is mean-zero Gaussian with covariance matrix $\Omega_{\delta,t_i} = \int_0^\delta \exp(u\Pi)\Omega(t_i - u)\exp(u\Pi')du$. Similar to Dorogovtsev (1978) and Prakasa Rao (1983), it can be shown that the least squares estimator α_δ is consistent under these conditions. Now, we derive the asymptotic distribution of $\hat{\alpha}_\delta$ under heteroskedasticity of unknown form.

According to Hafner and Herwartz (2009) it holds that

$$\lim_{T,m\to\infty} \frac{1}{Tm} \sum_{t=1}^{T} \sum_{i=1}^{m} E\left[(z_{t_{i-1}} z'_{t_{i-1}}) \otimes (\epsilon_{t_i} \epsilon'_{t_i})\right] = W, \tag{27}$$

with some finite, positive definite matrix W. Under our assumptions for the cointegrated OU type process, it further holds that

$$\lim_{T,m\to\infty} \frac{1}{Tm} \sum_{t=1}^{T} \sum_{i=1}^{m} E\left[(z_{t_{i-1}} z'_{t_{i-1}})\right] = V, \tag{28}$$

for a non-singular V.

Then, using the consistent estimators

$$W_{Tm} = \frac{1}{Tm} \sum_{t=1}^{T} \sum_{i=1}^{m} (z_{t_{i-1}} z'_{t_{i-1}}) \otimes (\epsilon_{t_i} \epsilon'_{t_i}), \tag{29}$$

and

$$V_{Tm} = \frac{1}{Tm} \sum_{t=1}^{T} \sum_{i=1}^{m} (z_{t_{i-1}} z'_{t_{i-1}}) \otimes I_N, \tag{30}$$

we follow White (1980) and Hafner and Herwartz (2009), and obtain the following convergence result

$$\sqrt{Tm} V_{Tm} W_{Tm}^{-1/2} \ \mathrm{vec}(\hat{\alpha}_\delta - \alpha_\delta) \xrightarrow{d} \mathcal{N}(0, I_{(N-1)^2 N^2}), \tag{31}$$

or equivalently

$$\sqrt{Tm} \ \mathrm{vec}(\hat{\alpha}_\delta - \alpha_\delta) \xrightarrow{d} \mathcal{N}(0, V_{Tm}^{-1} W_{Tm} V_{Tm}^{-1}). \tag{32}$$

This concludes the proof for part (i).

Without loss of generality, we choose $\hat{\alpha}_{\delta,\perp} = \alpha_{\delta,\perp} - \alpha_\delta (\hat{\alpha}'_\delta \alpha_\delta)^{-1} \hat{\alpha}'_\delta \alpha_{\delta,\perp}$, then $\hat{\alpha}'_\delta \hat{\alpha}_{\delta,\perp} = 0$ and

$$\hat{\alpha}_{\delta,\perp} - \alpha_{\delta,\perp} = -\hat{\bar{\alpha}}_\delta (\hat{\alpha}_\delta - \alpha_\delta)' \alpha_{\delta,\perp}, \tag{33}$$

where $\hat{\bar{\alpha}}_\delta = \alpha_\delta (\hat{\alpha}'_\delta \alpha_\delta)^{-1}$. Using the property $\hat{\alpha}_\delta \xrightarrow{p} \alpha_\delta$, it follows from the Continuous Mapping Theorem that $\hat{\bar{\alpha}}_\delta \xrightarrow{p} \bar{\alpha}_\delta = \alpha_\delta (\alpha'_\delta \alpha_\delta)^{-1}$ and finally $\hat{\alpha}_\perp = \hat{\alpha}_{\delta,\perp} \xrightarrow{p} \alpha_{\delta,\perp} = \alpha_\perp$.

□

Proof of Theorem 1. Let $\mathbb{K} = (\mathrm{vech}\,(\mathbb{K}_1)\,, \ldots, \mathrm{vech}\,(\mathbb{K}_N))'$ be a $N \times \frac{1}{2}N(N+1)$ matrix with $\mathbb{K}_N$ denoting a $N \times N$ matrix that has the (n,n)-th $n = 1, \ldots, N$ entry equal to one and all the remaining entries equal to zero, and $\mathbb{L}$ be a $\frac{1}{2}N(N+1) \times N^2$ elimination matrix such that, for any symmetric matrix A, $\mathrm{vech}\,(A) = \mathbb{L}\,\mathrm{vec}\,(A)$. First, write the diagonal elements of $\widehat{\langle P, P' \rangle}_{t-1:t}$ as $\mathbb{K}\mathbb{L}\,\mathrm{vec}\left(\widehat{\langle P, P' \rangle}_{t-1:t}\right)$. Next, write the difference between the market-specific estimates of the integrated variances and the integrated variance of the efficient price as $\mathbb{K}\mathbb{L}\,\mathrm{vec}\left(\widehat{\langle P, P' \rangle}_{t-1:t} - \beta_\perp \langle P^*, P^{*\prime} \rangle_{t-1:t} \beta'_\perp\right)$.

It then follows that

$$
\begin{aligned}
\widehat{\langle P, P' \rangle}^{n,n}_{t-1:t} - \langle P^*, P^{*\prime} \rangle_{t-1:t} &= k'(n)\mathbb{K}\mathbb{L}\,\mathrm{vec}\left(\widehat{\langle P, P' \rangle}_{t-1:t} - \beta_\perp \langle P^*, P^{*\prime} \rangle_{t-1:t} \beta'_\perp\right) \\
&= k'(n)\mathbb{K}\mathbb{L}\,\mathrm{vec}\left(\widehat{\langle P, P' \rangle}_{t-1:t} - \beta_\perp \alpha'_\perp \langle P, P' \rangle_{t-1:t} \alpha_\perp \beta'_\perp\right) \\
&= k'(n)\mathbb{K}\mathbb{L}\left[\mathrm{vec}\left(\widehat{\langle P, P' \rangle}_{t-1:t}\right)\right. \\
&\qquad \left. - (\beta_\perp \alpha'_\perp) \otimes (\beta_\perp \alpha'_\perp)\,\mathrm{vec}\left(\langle P, P' \rangle_{t-1:t}\right)\right], \\
&= k'(n)\mathbb{K}\mathbb{L}\left[I_{N^2} - (\beta_\perp \alpha'_\perp) \otimes (\beta_\perp \alpha'_\perp)\right] \\
&\qquad \times \mathrm{vec}\left(\langle P, P' \rangle_{t-1:t}\right) + O_p(m^{-\varphi}), \\
&= k'(n) \begin{pmatrix}
(1 - \alpha^2_{\perp,1}) \langle P, P' \rangle^{1,1}_{t-1:t} - \sum\limits_{\substack{i=1,\\ i \neq 1}}^{N} \alpha^2_{\perp,i} \langle P, P' \rangle^{i,i}_{t-1:t} \\
(1 - \alpha^2_{\perp,2}) \langle P, P' \rangle^{2,2}_{t-1:t} - \sum\limits_{\substack{i=1,\\ i \neq 2}}^{N} \alpha^2_{\perp,i} \langle P, P' \rangle^{i,i}_{t-1:t} \\
\vdots \\
(1 - \alpha^2_{\perp,N}) \langle P, P' \rangle^{N,N}_{t-1:t} - \sum\limits_{\substack{i=1,\\ i \neq N}}^{N} \alpha^2_{\perp,i} \langle P, P' \rangle^{i,i}_{t-1:t}
\end{pmatrix} \\
&\qquad - k'(n) \begin{pmatrix}
2 \sum\limits_{i=1}^{N-1} \sum\limits_{j=i+1}^{N} \alpha_{\perp,i}\alpha_{\perp,j} \langle P, P' \rangle^{i,j}_{t-1:t} \\
2 \sum\limits_{i=1}^{N-1} \sum\limits_{j=i+1}^{N} \alpha_{\perp,i}\alpha_{\perp,j} \langle P, P' \rangle^{i,j}_{t-1:t} \\
\vdots \\
2 \sum\limits_{i=1}^{N-1} \sum\limits_{j=i+1}^{N} \alpha_{\perp,i}\alpha_{\perp,j} \langle P, P' \rangle^{i,j}_{t-1:t}
\end{pmatrix} \\
&\qquad + O_p(m^{-\varphi}). \qquad (34)
\end{aligned}
$$

It then follows that (34) reduces to

$$
\begin{aligned}
\widehat{\langle P, P' \rangle}^{n,n}_{t-1:t} - \langle P^*, P^{*\prime} \rangle_{t-1:t} =& (1 - \alpha^2_{\perp,n}) \langle P, P' \rangle^{n,n}_{t-1:t} - \sum_{i=1, i \neq n}^{N} \alpha^2_{\perp,i} \langle P, P' \rangle^{i,i}_{t-1:t} + \\
& - 2 \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \alpha_{\perp,i}\alpha_{\perp,j} \langle P, P' \rangle^{i,j}_{t-1:t} + O_p(m^{-\varphi}), \qquad (35)
\end{aligned}
$$

$n = 1, \ldots, N$, which proves the theorem. Furthermore, if $\langle P, P' \rangle^{n,n}_{t-1:t} = \langle P, P' \rangle^{\bar{n},\bar{n}}_{t-1:t}$ for all $\bar{n} \neq n$ and $\bar{n} = 1, \ldots N$ and the correlation between markets remain the same across all market combinations, then Equation 35 simplifies to

$$\widehat{\langle P, P' \rangle}_{t-1:t}^{n,n} - \langle P^*, P^{*\prime} \rangle_{t-1:t} = \widehat{\langle P, P' \rangle}_{t-1:t}^{n,n} - \rho \left(\alpha_{\perp,1} + \ldots + \alpha_{\perp,N} \right)^2 \widehat{\langle P, P' \rangle}_{t-1:t}^{n,n} + O_p(m^{-\varphi}). \tag{36}$$

It then follows that the

$$\lim_{\rho \to 1} \widehat{\langle P, P' \rangle}_{t-1:t}^{n,n} - \langle P^*, P^{*\prime} \rangle_{t-1:t} = O_p(m^{-\varphi}).$$

□

Proof of Theorem 2. Assumption 2 ensures that only one common stochastic trend (the efficient price) exists, i.e., we restrict our analysis to the case that α, β are $N \times (N-1)$ matrices. This implies that $\alpha_{\delta,\perp}$, $\beta_\perp$ are $N \times 1$ vectors and $RV_{GRT,t}^{(m)}$ are scalar random variables defined by

$$\begin{aligned} RV_{GRT,t}^{(m)} &= \sum_{i=(t-1)/\delta}^{t/\delta} (\hat{\alpha}_{\delta,\perp}' \hat{\epsilon}_i)^2, \qquad t = 1, \ldots, T \\ &= \sum_{i=(t-1)/\delta}^{t/\delta} (\hat{\alpha}_{\delta,\perp}' \Delta P_i)^2 \\ &= \sum_{i=(t-1)/\delta}^{t/\delta} (\alpha_\perp' \Delta P_i)^2 + O_p((Tm)^{-1/2}), \end{aligned} \tag{37}$$

where $\hat{\epsilon}_{t_i} = \Delta P_{t_i} - \hat{\Pi}_\delta P_{t_{i-1}}$ and $\hat{\alpha}_{\delta,\perp}' \iota_N = 1$. The last equality follows from Lemma 1 (ii), stating that the discretely sampled least squares estimator of $\alpha_\perp$ is consistent, and Slutsky's Theorem. We further note that

$$\lim_{m \to \infty} \sum_{i=1}^{m} E(\epsilon_{t_i} \epsilon_{t_i}') = \int_{t-1}^{t} \Omega(u) du = \lim_{m \to \infty} \sum_{i=1}^{m} E(\Delta P_{t_i} \Delta P_{t_i}'), \tag{38}$$

defines the daily quadratic variation of the process.

Since the random variables $\hat{\alpha}_{\delta,\perp}' \Delta P_{t_i}$ are scalar under our assumptions and uncorrelated by construction, we can use the univariate results in Barndorff-Nielsen and Shepard (2002) to derive the estimators asymptotic distribution. Analogously to the proof of Theorem 1 in Barndorff-Nielsen and Shepard (2002), it then follows that

$$m^{1/2}\frac{\left(\sum\limits_{i=(t-1)/\delta}^{t/\delta}(\hat{\alpha}'_{\delta,\perp}\Delta P_{t_i})^2-\int\limits_{t-1}^{t}\alpha'_{\perp}\Omega(u)\alpha_{\perp}du\right)}{\sqrt{2\int\limits_{t-1}^{t}\alpha'_{\perp}\Omega(u)\alpha_{\perp}\alpha'_{\perp}\Omega(u)\alpha_{\perp}du}}\xrightarrow{d}\mathcal{N}\left(0,1\right), \tag{39}$$

for $m\to\infty$. Since the estimator is scaled by another unknown quantity, this result is not statistically feasible. Alternatively, one can use the convergence result for the realized quarticity (Barndorff-Nielsen and Shepard, 2002)

$$\frac{1}{3\delta}\sum_{i=(t-1)/\delta}^{t/\delta}(\hat{\alpha}'_{\delta,\perp}\Delta P_{t_i})^4\xrightarrow{p}\int_{t-1}^{t}\alpha'_{\perp}\Omega(u)\alpha_{\perp}\alpha'_{\perp}\Omega(u)\alpha_{\perp}du, \tag{40}$$

and write

$$\frac{\left(\sum\limits_{i=(t-1)/\delta}^{t/\delta}(\hat{\alpha}'_{\delta,\perp}\Delta P_{t_i})^2-\int\limits_{t-1}^{t}\alpha'_{\perp}\Omega(u)\alpha_{\perp}du\right)}{\sqrt{\frac{2}{3\delta}\sum\limits_{i=(t-1)/\delta}^{t/\delta}(\hat{\alpha}'_{\delta,\perp}\Delta P_{t_i})^4}}\xrightarrow{d}\mathcal{N}\left(0,1\right). \tag{41}$$

□

B Appendix

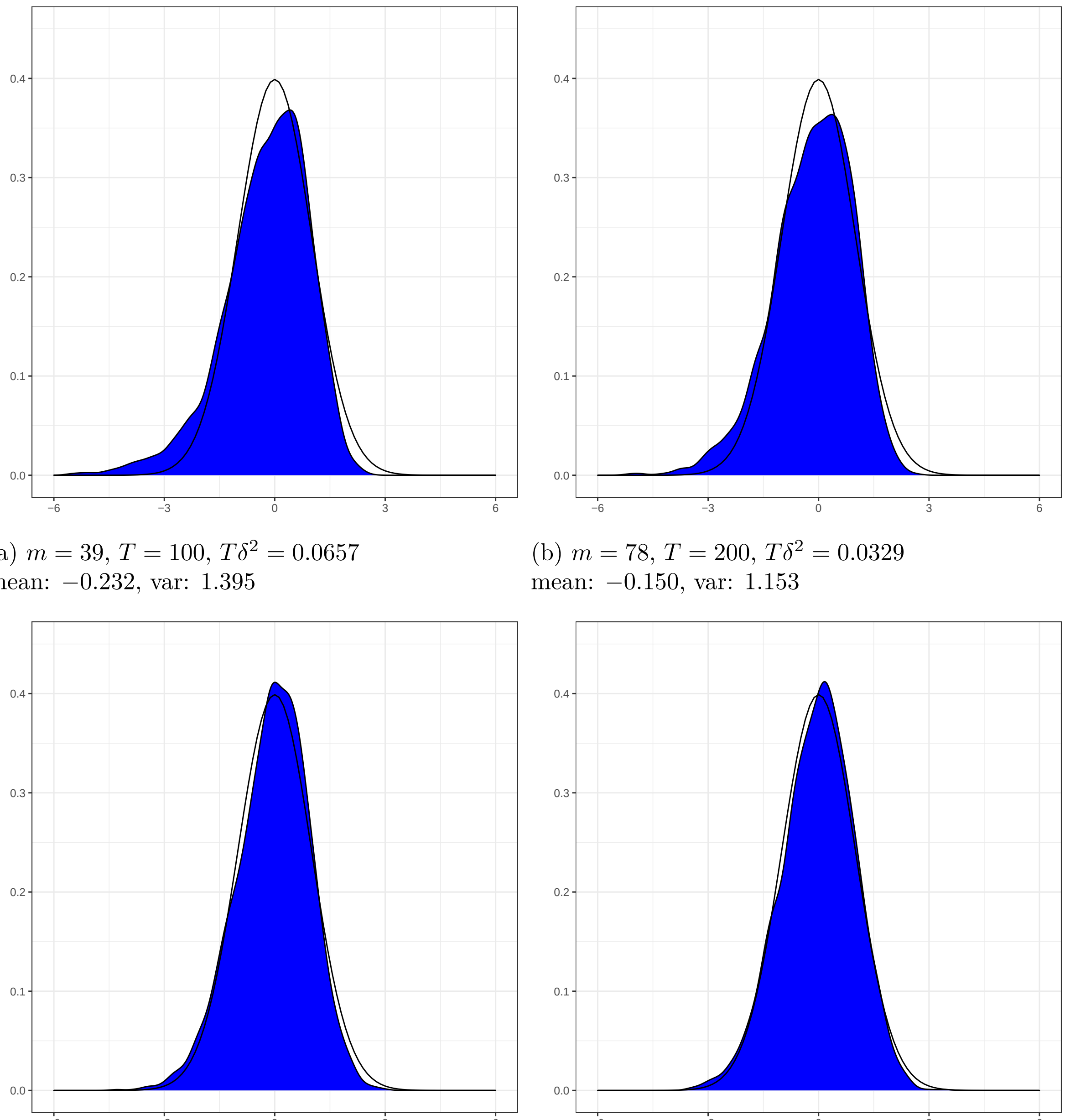
Table B1: Estimated mean and variance of the standardized $RV^{(m)}_{GRT,t}$

Panel A: mean					
δ \ T	50	100	200	400	800
1/39	-0.237	-0.232	-0.235	-0.236	-0.234
1/78	-0.128	-0.135	-0.150	-0.147	-0.151
1/390	-0.049	-0.046	-0.050	-0.053	-0.047
1/780	0.098	0.125	-0.017	0.006	-0.017
1/2340	1.062	0.027	0.149	0.116	0.014

Panel B: variance					
δ \ T	50	100	200	400	800
1/39	1.403	1.395	1.400	1.402	1.398
1/78	1.137	1.142	1.153	1.150	1.154
1/390	0.962	0.961	0.964	0.964	0.964
1/780	0.959	0.956	0.972	0.970	0.972
1/2340	0.893	0.970	0.961	0.963	0.971

Note: This table reports result based on 2,500 draws from the stochastic volatility two-market model with $\rho = 0$.

Figure B1: Asymptotic distribution of $RV_{GRT,t}^{(m)}$, $R = 2,500$, $\alpha = (-5, 5)'$



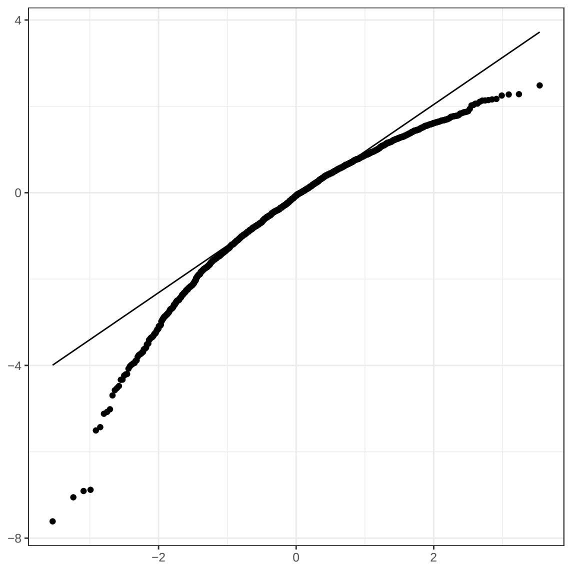
(a) $m = 39$, $T = 100$, $T\delta^2 = 0.0657$
mean: -0.232, var: 1.395

(b) $m = 78$, $T = 200$, $T\delta^2 = 0.0329$
mean: -0.150, var: 1.153

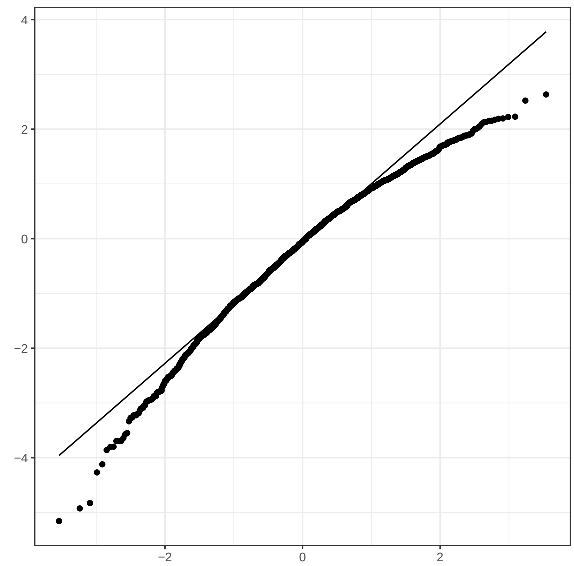
(c) $m = 390$, $T = 400$, $T\delta^2 = 0.0026$
mean: -0.053, var: 0.964

(d) $m = 780$, $T = 800$, $T\delta^2 = 0.0013$
mean: -0.017, var: 0.972

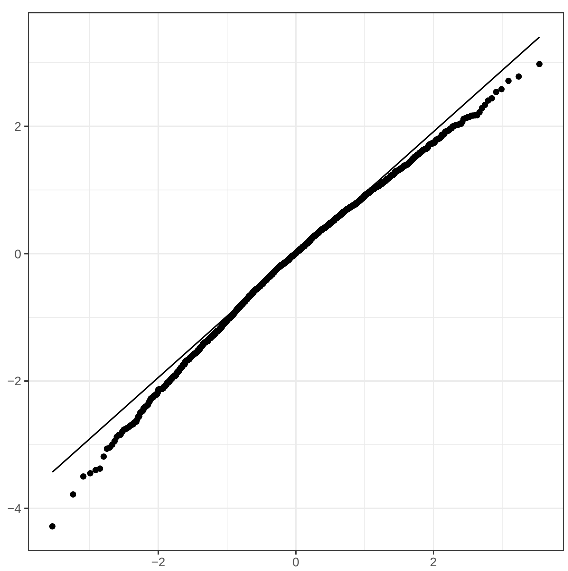
Figure B2: QQ-plots for $RV^{(m)}_{GRT,t}$, $R = 2,500$, $\alpha = (-5, 5)'$



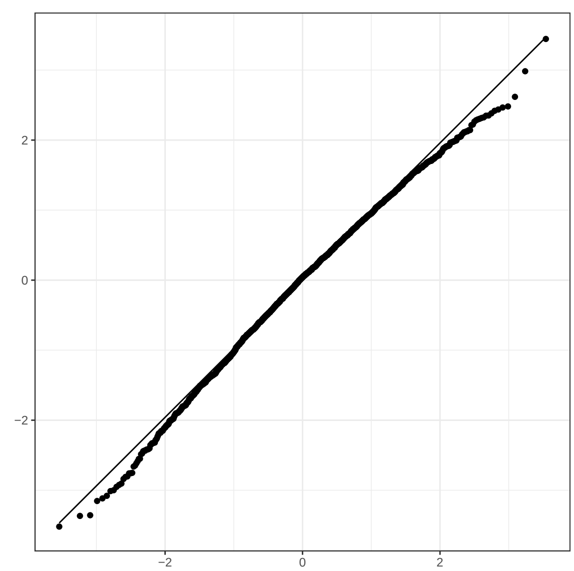
(a) $m = 39$, $T = 100$, $T\delta^2 = 0.0657$
mean: -0.232, var: 1.395



(b) $m = 78$, $T = 200$, $T\delta^2 = 0.0329$
mean: -0.150, var: 1.153



(c) $m = 390$, $T = 400$, $T\delta^2 = 0.0026$
mean: -0.053, var: 0.964



(d) $m = 780$, $T = 800$, $T\delta^2 = 0.0013$
mean: -0.017, var: 0.972