

A frequency domain analysis of the error distribution from noisy high-frequency data

By JINYUAN CHANG

*School of Statistics, Southwestern University of Finance and Economics, Chengdu,
Sichuan 611130, China*
changjinyuan@swufe.edu.cn

AUORE DELAIGLE

*School of Mathematics and Statistics, University of Melbourne, Parkville, Victoria 3010,
Australia*
aurored@unimelb.edu.au

PETER HALL

*School of Mathematics and Statistics, University of Melbourne, Parkville, Victoria 3010,
Australia*

AND CHENG YONG TANG

*Department of Statistical Science, Temple University, 1810 North 13th Street, Philadelphia,
Pennsylvania 19122, U.S.A.*
yongtang@temple.edu

SUMMARY

Data observed at a high sampling frequency are typically assumed to be an additive composite of a relatively slow-varying continuous-time component, a latent stochastic process or smooth random function, and measurement error. Supposing that the latent component is an Itô diffusion process, we propose to estimate the measurement error density function by applying a deconvolution technique with appropriate localization. Our estimator, which does not require equally-spaced observed times, is consistent and minimax rate-optimal. We also investigate estimators of the moments of the error distribution and their properties, propose a frequency domain estimator for the integrated volatility of the underlying stochastic process, and show that it achieves the optimal convergence rate. Simulations and an application to real data validate our analysis.

Some key words: Deconvolution; Fourier transform; Functional data; High-frequency data; Measurement error; Smoothing.

1. INTRODUCTION

High-frequency data, observed sequentially at small time intervals, arise in various settings and applications. For example, in social and behavioural investigations, such data are often collected in so-called intensive longitudinal studies; see [Bolger & Laurenceau \(2013\)](#). In functional data analysis, the observations are often considered to be of one of two types: the dense setting,

which corresponds to high-frequency data, and the sparse setting, where the data are of low frequency; see, among others, Müller et al. (2011) and Wang et al. (2016). In finance, analysing high-frequency intraday transaction data has received increasing attention; see Hautsch (2012), Aït-Sahalia & Jacod (2014) and the references therein.

High-frequency observations are often contaminated by measurement errors. For example, in the dense functional data setting, it is common to assume that the observed discrete data are a noisy version of an underlying unknown smooth curve. In finance too, high-frequency data are often regarded as a noisy version of a latent continuous-time stochastic process, observed at consecutive discrete time-points. The latent process is considered to have a continuous path, and the measurement error represents the market microstructure noise; see Aït-Sahalia & Yu (2009). Since increasing the sampling frequency implies smaller sampling errors caused by the discretization of the underlying continuous-time process, we might expect that high-frequency data would enable more accurate inference. However, as the sampling frequency increases, the difference between nearby observations is dominated by random noise, which makes standard methods of inference inconsistent; see, for example, Zhang et al. (2005). Therefore, a main concern in high-frequency financial data analysis has been to find ways of recovering the signal of quantities of interest from noisy high-frequency observations; see Aït-Sahalia & Jacod (2014).

Revealing the distributional properties of the measurement errors is crucial for recovering the signal of the continuous-time process from noisy high-frequency data. First, many estimation procedures require distributional assumptions on the measurement errors. Second, statistical inference, including hypothesis testing and confidence set estimation, inevitably involves unknown nuisance parameters determined by the distribution of the measurement errors; see, for example, Zhang (2006), Aït-Sahalia et al. (2010), Xiu (2010) and Liu & Tang (2014).

The measurement errors themselves sometimes contain useful information, both theoretical and practical, so that successfully recovering the measurement error distribution can improve our understanding of data structures. For example, Dufrenot et al. (2014) argued that the microstructure noise can help us understand financial crises, and Aït-Sahalia & Yu (2009) made a connection between the microstructure noise and the liquidity of stocks. In longitudinal studies, the measurement errors can help to reveal interesting characteristics of a population, such as the way in which individuals misreport their diet. Jacod et al. (2017) recently highlighted the importance of statistical properties of the measurement errors and studied the estimation of moments. Despite these important developments, to our knowledge, as yet no method has been proposed for estimating the entire distribution of the errors.

In this paper we consider frequency domain analysis of high-frequency data, focusing on the measurement errors contaminating continuous-time processes. In high-frequency observations, the relative magnitude of the changes in values of the underlying continuous-time process is small. Compared with the measurement errors, it becomes negligible, locally in a small neighbourhood of a given time-point. As a result, estimating the error distribution shares some common features with nonparametric measurement error problems with repeated measurements studied in Delaigle et al. (2008), and with nonparametric estimation from aggregated data studied by Meister (2007) and Delaigle & Zhou (2015), where the estimation techniques require working in the Fourier domain.

Motivated by this, we propose to estimate the characteristic function of the measurement errors by locally averaging the empirical characteristic functions of the changes in values of the high-frequency data. We obtain a nonparametric estimator of the probability density function of the measurement errors and show that it is consistent and minimax rate-optimal. We propose a simple method for consistently estimating the moments of the measurement errors. Using our estimator of the characteristic function of the errors, we develop a new rate-optimal multi-scale

frequency domain estimator of the integrated volatility of the stochastic process, a key quantity of interest in high-frequency financial data analysis.

In what follows, for two sequences of real numbers $\{a_n\}$ and $\{b_n\}$, we write $a_n \asymp b_n$ if there exist positive constants c_1 and c_2 such that $c_1 \leq a_n/b_n \leq c_2$ for all $n \geq 1$. We denote by $f * g$ the convolution of two functions f and g , defined by $f * g(s) = \int f(s - \tau)g(\tau) d\tau$.

2. METHODOLOGY

2.1. Model and data

We are interested in a continuous-time process $(X_t)_{t \in [0, T]}$ observed at high frequency with $T > 0$. We assume that X_t follows a diffusion process

$$dX_t = \mu_t dt + \sigma_t dB_t, \tag{1}$$

where the drift μ_t is a locally bounded and progressively measurable process, σ_t is a positive and locally bounded Itô semimartingale, and B_t is a standard Brownian motion. The process σ_t^2 represents the volatility of the process X_t at time t , and is often investigated in its integrated form $\int_0^T \sigma_t^2 dt$, called the integrated volatility.

Remark 1. The model (1) is often used when $X_t = \log S_t$, where S_t denotes the price process of an equity; see, for example, [Zhang et al. \(2005\)](#). It is also used to model applications in biology, physics and many other fields; see [Olhede et al. \(2009\)](#). All theoretical properties of our estimators are derived under this model, but the methods developed in § 2.2 and 2.4 could be applied to other types of processes X_t , such as the smooth ones typically encountered in the functional data literature. The key property that makes our methods consistent is the continuity of the underlying process X_t , but the convergence rates of our estimators depend on more specific assumptions, such as those implied by the model (1).

Our data are observed on a generic discrete grid $\mathcal{G} = \{t_0, \dots, t_n\}$ of time-points where, without loss of generality, we let $t_0 = 0$ and $t_n = T$. The observed data are contaminated by additive measurement errors, so that what we observe is a sample $\{Y_{t_j}\}_{j=0}^n$ where

$$Y_{t_j} = X_{t_j} + U_{t_j}. \tag{2}$$

A conventional assumption when analysing noisy high-frequency data is that the random measurement error U_t is independent of X_t ; see [Aït-Sahalia et al. \(2010\)](#), [Xiu \(2010\)](#) and [Liu & Tang \(2014\)](#). This is also a standard assumption in the measurement error and the functional data literature; see [Carroll & Hall \(1988\)](#), [Stefanski & Carroll \(1990\)](#) and [Wang et al. \(2016\)](#). Likewise, we make the following assumption.

Assumption 1. The errors $\{U_{t_j}\}_{j=0}^n$ are independently and identically distributed with unknown density f_U , and are independent of the process $(X_t)_{t \in [0, T]}$.

We are interested in deriving statistical properties of the noise term U_{t_j} when T is fixed and the frequency of observations increases, that is, when $\max_{1 \leq j \leq n} \Delta t_j \rightarrow 0$ as $n \rightarrow \infty$, where $\Delta t_j = t_j - t_{j-1}$ for $j = 1, \dots, n$. Here, the time-points t_j do not need to be equispaced. Formally, we make the following assumption.

Assumption 2. As $n \rightarrow \infty$, $\min_{1 \leq j \leq n} \Delta t_j / \max_{1 \leq j \leq n} \Delta t_j$ is uniformly bounded away from zero and infinity.

Throughout, we use f^{Ft} to denote the Fourier transform of a function f and make the following assumption on the characteristic function of the errors, which is almost always assumed in the related nonparametric measurement error literature:

Assumption 3. f_U^{Ft} is real-valued and does not vanish at any point on the real line.

2.2. Estimating the error density f_U

Motivated by our discussion in § 1, we wish to estimate the error density f_U . At a given t_j , if we had access to repeated noisy measurements of X_{t_j} , say $Y_{t_j,k} = X_{t_j} + U_{t_j,k}$ for $k = 1, \dots, r$, where the $U_{t_j,k}$ s are independent and each $U_{t_j,k} \sim f_U$, then for $\ell \neq k$ we would have $(Y_{t_j,\ell} - Y_{t_j,k}) \sim f_U * f_U$. Under Assumption 3, using the approach in Delaigle et al. (2008) we could estimate f_U^{Ft} by the square root of the empirical characteristic function of the $(Y_{t_j,\ell} - Y_{t_j,k})$ s; then, by Fourier inversion, we could deduce an estimator of f_U .

However, for high-frequency data, at each given t_j we have access to only one contaminated measurement Y_{t_j} . Therefore, the above technique cannot be applied. But since $(X_t)_{t \in [0, T]}$ is a continuous-time and continuous-path stochastic process, $|X_{t+h} - X_t| \rightarrow 0$ almost surely as $h \rightarrow 0$. Thus, the collection of observations $\{Y_{t_\ell}\}$, where t_ℓ lies in a small neighbourhood \mathcal{N} of t_j , can be approximately viewed as repeated measurements of X_{t_j} contaminated by independently and identically distributed errors $\{U_{t_\ell}\}$. As the sampling frequency increases, we have multiple observations in smaller and smaller neighbourhoods \mathcal{N} , which suggests that the density of $Y_{t_\ell} - Y_{t_j}$, for $t_j \neq t_\ell \in \mathcal{N}$, gets closer and closer to $f_U * f_U$. Therefore, we can expect that as the sample frequency increases, the approach suggested by Delaigle et al. (2008), applied to the $(Y_{t_\ell} - Y_{t_j})$ s for t_ℓ and t_j sufficiently close, can provide an increasingly accurate estimator of f_U .

We shall prove in § 2.3 that this heuristic is correct as long as the t_ℓ s and t_j s are carefully chosen, which we characterize through a distance ξ . For \mathcal{G} defined in § 2.1 and $\xi > 0$, we define

$$S_j = \{t_\ell \in \mathcal{G} : |t_\ell - t_j| \leq \xi \text{ and } \ell \neq j\} \quad (j = 0, \dots, n) \tag{3}$$

and denote by N_j the number of points in S_j . For a fixed T , Assumption 2 implies that $\min_{1 \leq j \leq n} \Delta t_j \asymp \max_{1 \leq j \leq n} \Delta t_j \asymp n^{-1}$, so that $\max_{1 \leq j \leq n} N_j \asymp \min_{1 \leq j \leq n} N_j \asymp n\xi$.

Following the discussion above and recalling Assumption 3, for a given ξ we define our estimator of $f_U^{\text{Ft}}(s)$ by the square root of the real part of the empirical characteristic function of the difference between nearby Y_{t_j} s:

$$\hat{f}_{U,1}^{\text{Ft}}(s; \xi) = \left| \frac{1}{N(\xi)} \sum_{j=0}^n \sum_{t_\ell \in S_j} \cos\{s(Y_{t_\ell} - Y_{t_j})\} \right|^{1/2}, \tag{4}$$

where $N(\xi) = \sum_{j=0}^n N_j$. Here ξ can be viewed as a parameter controlling the trade-off between the bias and the variance of $\hat{f}_{U,1}^{\text{Ft}}$: a smaller ξ gives a smaller bias but also results in a smaller $N(\xi)$ so that the variance is higher. On the other hand, a larger ξ induces a lower variance but comes at the price of a larger bias due to the contribution from the dynamics in X_t . The choice of ξ in practice will be discussed in § 3.1.

It follows from the Fourier inversion theorem that $f_U(x) = (2\pi)^{-1} \int \exp(-isx) f_U^{\text{Ft}}(s) ds$, where $i^2 = -1$. We can obtain an estimator of f_U by replacing $f_U^{\text{Ft}}(s)$ with $\hat{f}_{U,1}^{\text{Ft}}(s, \xi)$ in this

integral. However, since $\hat{f}_{U,1}^{\text{Ft}}(s, \xi)$ is an empirical characteristic function, it is unreliable when $|s|$ is large. For the integral to exist, $\hat{f}_{U,1}^{\text{Ft}}(s, \xi)$ needs to be multiplied by a regularizing factor that puts less weight on large $|s|$. As the sample size increases, $\hat{f}_{U,1}^{\text{Ft}}(s, \xi)$ becomes more reliable, and this should be accounted for by letting the weight depend on the sample size. Using standard kernel smoothing techniques, this can be implemented by taking

$$\hat{f}_{U,2}^{\text{Ft}}(s; \xi) = \hat{f}_{U,1}^{\text{Ft}}(s; \xi) \mathcal{K}^{\text{Ft}}(sh),$$

where \mathcal{K}^{Ft} is the Fourier transform of a kernel function \mathcal{K} and $h > 0$ is a bandwidth parameter that satisfies $h \rightarrow 0$ as $n \rightarrow \infty$. Then, we define our estimator of $f_U(x)$ by

$$\hat{f}_U(x; \xi) = \frac{1}{2\pi} \int \exp(-isx) \hat{f}_{U,1}^{\text{Ft}}(s; \xi) \mathcal{K}^{\text{Ft}}(sh) ds = \frac{1}{2\pi} \int \exp(-isx) \hat{f}_{U,2}^{\text{Ft}}(s; \xi) ds. \quad (5)$$

For consistency of a kernel density estimator, the kernel function \mathcal{K} needs to be symmetric and integrate to unity. In practice, the estimator is often not very sensitive to the choice of kernel compared with the choice of bandwidth. Popular choices are the Gaussian kernel, i.e., the standard normal density, and the sinc kernel, whose Fourier transform is $\mathcal{K}^{\text{Ft}}(s) = I(|s| \leq 1)$ where $I(\cdot)$ denotes the indicator function. In practice, an advantage of the Gaussian kernel is that it can produce visually attractive, smoother estimators than the sinc kernel. The sinc kernel is more advantageous analytically; see § 2.3.

2.3. Properties of the density estimator

Assume that the continuous-time process $(X_t)_{t \in [0, T]}$ in (1) belongs to the class $\mathcal{X}(C_1)$ for some $C_1 > 0$, where

$$\mathcal{X}(C_1) = \left\{ (X_t)_{t \in [0, T]} : X_t \text{ satisfies (1) with } \sup_{0 \leq t \leq T} E(\mu_t^4) \leq C_1 \text{ and } \sup_{0 \leq t \leq T} E(\sigma_t^4) \leq C_1 \right\},$$

and that f_U belongs to the class

$$\mathcal{F}_1(\alpha, C_2) = \{ f : |f^{\text{Ft}}(s)| \leq C_2(1 + |s|)^{-\alpha} \text{ for all real } s \}$$

for some constants $\alpha > 0$ and $C_2 > 1$. This class is rich; for example, it contains the functions that have at least $\alpha - 1$ square-integrable derivatives. Characterizing error distributions through their Fourier transforms is standard in nonparametric measurement error problems because it is key to deriving precise asymptotic properties of the estimators.

We consider the sinc kernel \mathcal{K} introduced in the paragraph following (5). Using this kernel simplifies our presentation of the theoretical derivations in two respects: its Fourier transform simplifies calculations, and it is a so-called infinite-order kernel, which implies that the bias of the resulting nonparametric curve estimators depends only on the smoothness of the target curve. Thus the sinc kernel has adaptive properties and automatically ensures optimal convergence rates. In contrast, the bias of estimators based on finite-order kernels, such as the Gaussian kernel, depends on both the order of the kernel and the smoothness of the target curve, which means that various smoothness subcases need to be considered when deriving properties.

For any two square-integrable functions f and g , let $\|f - g\|_2 = (\int |f - g|^2)^{1/2}$. Proposition 1 gives the convergence rate of \hat{f}_U , defined in (5), to the true density function f_U .

PROPOSITION 1. Let $(X_t)_{t \in [0, T]} \in \mathcal{X}(C_1)$ and assume that the errors $\{U_{t_j}\}_{j=0}^n$ satisfy Assumptions 1 and 3 and that $f_U \in \mathcal{F}_1(\alpha, C_2)$. Let $\mathcal{P}_1(\alpha, C_1, C_2)$ denote the collection of models for $(Y_t)_{t \in [0, T]}$ such that $Y_t = X_t + U_t$. Under Assumption 2 and with the sinc kernel \mathcal{K} , if $\alpha > 3/2$, then for some uniform constant $C > 0$ depending only on α, C_1 and C_2 ,

$$\sup_{\mathcal{P}_1(\alpha, C_1, C_2)} E(\|\hat{f}_U - f_U\|_2^2) \leq C(n^{-1}\xi^{-1/2}h^{-1} + n^{-1/2} + \xi + h^{2\alpha-1}).$$

Proposition 1 shows that the L_2 convergence rate of \hat{f}_U to f_U is affected by ξ , the length of each block S_j , and the bandwidth h . The next theorem shows that for appropriate choices of ξ and h , the convergence rate attains $n^{-1/2}$.

THEOREM 1. If the conditions of Proposition 1 hold and we take $\xi \asymp n^{-\delta_1}$ and $h \asymp n^{-\delta_2}$, where $\delta_1 > 0$ and $\delta_2 > 0$ are such that $\delta_1 + 2\delta_2 \leq 1$, $\delta_1 \geq 1/2$ and $\delta_2 \geq (4\alpha - 2)^{-1}$, then for some uniform constant $C > 0$ depending only on α, C_1 and C_2 ,

$$\sup_{\mathcal{P}_1(\alpha, C_1, C_2)} E(\|\hat{f}_U - f_U\|_2^2) \leq Cn^{-1/2}.$$

From Theorem 1 we learn that as long as $\alpha > 3/2$, the convergence rate of \hat{f}_U does not depend on α . This is strikingly different from standard nonparametric density estimation problems, where convergence rates typically depend on the smoothness of the target density: the smoother the density, the faster the rate. For example, if we had access to the $(U_{t_\ell} - U_{t_j})$ s directly instead of just $Y_{t_\ell} - Y_{t_j} = X_{t_\ell} - X_{t_j} + U_{t_\ell} - U_{t_j}$, then we could apply the technique suggested by Meister (2007) and the convergence rate would increase with α . However, in our case, the nuisance contribution due to the $(X_{t_\ell} - X_{t_j})$ s makes it impossible to reach rates faster than $n^{-1/2}$, even if α is very large. This is demonstrated in the next theorem, which shows that the $n^{-1/2}$ rate derived in Theorem 1 is minimax optimal.

THEOREM 2. Denote by $\check{\mathcal{F}}$ the class of all measurable functionals of the data. Under the conditions in Proposition 1, for some uniform constant $C > 0$ depending only on α, C_1 and C_2 ,

$$\inf_{\hat{f} \in \check{\mathcal{F}}} \sup_{\mathcal{P}_1(\alpha, C_1, C_2)} E(\|\hat{f} - f_U\|_2^2) \geq Cn^{-1/2}.$$

2.4. Estimating the moments of the microstructure noise

We can deduce estimators of the moments of the microstructure noise U_{t_j} from the density estimator derived in § 2.2, but proceeding in that way is unnecessarily complex. Recall from § 2.2 that when $t_\ell \neq t_j$ are close, $Y_{t_\ell} - Y_{t_j}$ behaves approximately like $U_{t_\ell} - U_{t_j} \sim f_{\tilde{U}} = f_U * f_U$, where U and \tilde{U} denote generic random variables with the same distributions as, respectively, U_{t_j} and $U_{t_\ell} - U_{t_j}$. This suggests that we could estimate the moments of \tilde{U} by the empirical moments of $Y_{t_\ell} - Y_{t_j}$ and, from these, deduce estimators of the moments of U .

For each integer $k \geq 1$, let $M_{U,k}$ and $M_{\tilde{U},k}$ denote the k th moments of U and \tilde{U} , respectively. Since f_U is symmetric, $M_{U,2k-1}$ and $M_{\tilde{U},2k-1}$ are equal to zero for all $k \geq 1$, and we only need to estimate even-order moments. For each $k \geq 1$, we start by estimating $M_{\tilde{U},2k}$ by

$$\hat{M}_{\tilde{U},2k}(\xi) = \frac{1}{N(\xi)} \sum_{j=0}^n \sum_{t_\ell \in S_j} (Y_{t_\ell} - Y_{t_j})^{2k}.$$

This is directly connected to our frequency domain analysis: it is easily proved that $\hat{M}_{\tilde{U},2k}(\xi) = (-i)^{2k} \{\hat{f}_{\tilde{U}}^{\text{Ft}}(0; \xi)\}^{(2k)}$, where $\hat{f}_{\tilde{U}}^{\text{Ft}} = (\hat{f}_{U,1}^{\text{Ft}})^2$ is an estimator of $f_{\tilde{U}}^{\text{Ft}}$, with $\hat{f}_{U,1}^{\text{Ft}}$ as in (4).

Exploiting the fact that $U_{t_\ell} - U_{t_j} \sim f_{\tilde{U}}$, we can write $M_{\tilde{U},2k} = \sum_{j=0}^k C_{2k}^{2j} M_{U,2j} M_{U,2k-2j}$ where $C_{2k}^{2j} = (2k)! / \{(2j)!(2k - 2j)!\}$, from which it can be deduced that

$$M_{U,2k} = \frac{1}{2} \left(M_{\tilde{U},2k} - \sum_{j=1}^{k-1} C_{2k}^{2j} M_{U,2j} M_{U,2k-2j} \right).$$

Therefore, we can use an iterative procedure to estimate the $M_{U,2k}$ s. First, for $k = 1$, we take $\hat{M}_{U,2}(\xi) = \hat{M}_{\tilde{U},2}(\xi)/2$. Then, for $k > 1$, given $\hat{M}_{U,2}(\xi), \dots, \hat{M}_{U,2(k-1)}(\xi)$ we take

$$\hat{M}_{U,2k}(\xi) = \frac{1}{2} \left\{ \hat{M}_{\tilde{U},2k}(\xi) - \sum_{j=1}^{k-1} C_{2k}^{2j} \hat{M}_{U,2j}(\xi) \hat{M}_{U,2k-2j}(\xi) \right\}. \tag{6}$$

Remark 2. When $k = 1$, $M_{U,2} = M_{\tilde{U},2}/2$ is equal to the variance of U_t , and our estimator is very similar to the so-called difference-based variance estimator often employed in related nonparametric regression problems; see, for example, Buckley et al. (1988) and Hall et al. (1990).

The next theorem establishes the convergence rate of $\hat{M}_{U,2k}(\xi)$. Its proof follows from the convergence rates of the $\hat{M}_{\tilde{U},2l}(\xi)$ s.

THEOREM 3. *Under Assumptions 1–3, for any integer $k \geq 1$, if $E(\int_0^T \mu_s^{2k} ds) < \infty$ and $E(\int_0^T \sigma_s^{2k} ds) < \infty$ hold and if there exists $p \in (2, 3]$ such that $M_{U,2kp} < \infty$, then $\hat{M}_{U,2k}(\xi) = M_{U,2k} + O_p(n^{-1/2})$ provided that $\xi = o[n^{-p/(2(p-1))}]$.*

Next, we derive the asymptotic joint distribution of the proposed moment estimators. Let $W = (W_1, \dots, W_k)^T$ be a random vector with a $N(0, \Sigma_k)$ distribution, where the (l_1, l_2) th element $(l_1, l_2 = 1, \dots, k)$ of Σ_k is equal to

$$e_{l_1 l_2} = \lim_{n \rightarrow \infty} E \left(\left[\sum_{j=0}^n \sum_{t_\ell \in S_j} \frac{(U_{t_\ell} - U_{t_j})^{2l_1} - M_{\tilde{U},2l_1}}{\{N(\xi) V_{2l_1}(\xi)\}^{1/2}} \right] \left[\sum_{j=0}^n \sum_{t_\ell \in S_j} \frac{(U_{t_\ell} - U_{t_j})^{2l_2} - M_{\tilde{U},2l_2}}{\{N(\xi) V_{2l_2}(\xi)\}^{1/2}} \right] \right),$$

with $V_{2l}(\xi) = \text{var}[\{N(\xi)\}^{-1/2} \sum_{j=0}^n \sum_{t_\ell \in S_j} \{(U_{t_\ell} - U_{t_j})^{2l} - M_{\tilde{U},2l}\}]$. Recalling that $N(\xi) \asymp n^2 \xi$ and noting that $V_{2l}(\xi) \asymp n \xi$, let

$$a_l = \lim_{n \rightarrow \infty} \{n V_{2l}(\xi) / N(\xi)\}^{1/2} / 2 \quad (l = 1, \dots, k).$$

The next theorem establishes the asymptotic joint distribution of our moment estimators. It can be used to derive confidence regions for $(M_{U,2}, \dots, M_{U,2k})^T$.

THEOREM 4. *Under Assumptions 1–3, for any integer $k \geq 1$, if $E(\int_0^T \mu_s^{2k} ds) < \infty$ and $E(\int_0^T \sigma_s^{2k} ds) < \infty$ and if there exists $p \in (2, 3]$ such that $M_{U,2kp} < \infty$, then*

$$n^{1/2} \{\hat{M}_{U,2}(\xi) - M_{U,2}, \dots, \hat{M}_{U,2k}(\xi) - M_{U,2k}\}^T \rightarrow Q = (Q_1, \dots, Q_k)^T$$

in distribution as $n \rightarrow \infty$, provided that $\xi = o(n^{-p/\{2(p-1)\}})$, where

$$Q_1 = a_1 W_1, \quad Q_l = a_l W_l - \sum_{j=1}^{l-1} C_{2l}^{2j} M_{U,2l-2j} Q_j \quad (2 \leq l \leq k).$$

2.5. *Efficient integrated volatility estimation*

We have demonstrated that our frequency domain analysis can be used to estimate the error density, which is difficult to estimate in the time domain. Since frequency domain approaches are unusual in high-frequency financial data analysis, a natural question is whether they can lead to an efficient estimator of the integrated volatility $\int_0^T \sigma_t^2 dt$, with σ_t as in (1). The integrated volatility is a key quantity of interest in high-frequency financial data analysis; it represents the variability of a process over time. It is well known that in cases like ours where the data are observed with microstructure noise, the integrated volatility cannot be estimated using standard procedures, which are dominated by contributions from the noise. In such situations, one way of removing the bias caused by the noise is through multi-scale techniques; see [Olhede et al. \(2009\)](#) for a very nice description. [Zhang \(2006\)](#) and [Tao et al. \(2013\)](#) have successfully applied those methods to correct the bias of estimators in the time domain, and [Olhede et al. \(2009\)](#) have proposed a consistent discrete frequency domain estimator in the case where the data are observed at equispaced times. Below we show that these techniques can be applied in our continuous frequency domain context too, even if the observation times are not restricted to be equispaced.

The real part of the empirical characteristic function $n^{-1} \sum_{j=1}^n \exp\{is(Y_{t_j} - Y_{t_{j-1}})\}$ is such that

$$\sum_{j=1}^n \cos\{s(Y_{t_j} - Y_{t_{j-1}})\} = \sum_{j=1}^n \cos\{s(U_{t_j} - U_{t_{j-1}})\} - \frac{s^2}{2} f_U^{Ft}(s) \int_0^T \sigma_t^2 dt + O_p(n^{-1/2}). \quad (7)$$

The second term on the right-hand side of (7) contains the integrated volatility, but the first term dominates because its mean is $nf_U^{Ft}(s)$. This suggests that the integrated volatility could be estimated from $\sum_{j=1}^n \exp\{is(Y_{t_j} - Y_{t_{j-1}})\}$ if we could eliminate that first term. This can be done by applying, to the frequency domain, the multi-scale technique used by [Zhang \(2006\)](#) and [Tao et al. \(2013\)](#). We define a function $G(s)$ which combines the empirical characteristic functions calculated at different sampling frequencies in such a way as to eliminate the first term on the right-hand side of (7) while keeping the second. For $N = \lfloor (n + 1)^{1/2} \rfloor$, we define

$$G(s) = \sum_{m=1}^N a_m \hat{\phi}^{K_m}(s) + \zeta \{ \hat{\phi}^{K_1}(s) - \hat{\phi}^{K_2}(s) \}$$

where, as in [Zhang \(2006\)](#), $K_m = m$, $a_m = 12K_m(m - N/2 - 1/2) / \{N(N^2 - 1)\}$, $\zeta = K_1 K_2 / \{(n + 1)(K_2 - K_1)\}$, and $\hat{\phi}^{K_m}(s) = K_m^{-1} \sum_{\ell=K_m}^n \exp\{is(Y_{t_\ell} - Y_{t_{\ell-K_m}})\}$. We could also select K_m , a_m and ζ as in [Tao et al. \(2013\)](#).

The following proposition shows that the real part of $G(s)$, $\text{Re}\{G(s)\}$, can be used to approximate the second term of (7).

PROPOSITION 2. Under Assumptions 1–3, if $E(U_t^2) < \infty$, then there exist second-order differentiable functions $\tau_1(s)$ and $\tau_2(s)$ such that for any $s \in \mathbb{R}$,

$$\left| \operatorname{Re}\{G(s)\} + \frac{s^2}{2} f_{\hat{U}}^{\text{Ft}}(s) \int_0^T \sigma_t^2 dt \right| = \tau_1(s) O_p(n^{-1/4}) + \tau_2(s) O_p(n^{-1/2}),$$

where the terms $O_p(n^{-1/4})$ and $O_p(n^{-1/2})$ are independent of s , and $\lim_{s \rightarrow 0} |\tau_1''(s)| \leq C$ and $\lim_{s \rightarrow 0} |\tau_2''(s)| \leq C$ for some positive constant C .

Since the function $G(s)$ depends only on the data, it is completely known. Moreover, we have seen in § 2.4 that we could estimate $f_{\hat{U}}^{\text{Ft}}(s)$ by $\hat{f}_{\hat{U}}^{\text{Ft}}(s; \xi) = \{\hat{f}_{U,1}^{\text{Ft}}(s; \xi)\}^2$. Finally, although the proposition holds for all $s \in \mathbb{R}$, the remainders are smaller when $f_{\hat{U}}^{\text{Ft}}(s)$ is close to 1, especially since $\hat{f}_{\hat{U}}^{\text{Ft}}(s; \xi)$ is more reliable in that case too. Therefore, for $\hat{f}_{\hat{U}}^{\text{Ft}}(s; \xi)$ close to 1, Proposition 2 can be used to compute an estimator of the integrated volatility, $\int_0^T \sigma_t^2 dt$. We propose a regression-type approach as follows. For some s_1, \dots, s_m such that $\hat{f}_{\hat{U}}^{\text{Ft}}(s; \xi)$ is close to 1, we consider the fixed design regression problem

$$\operatorname{Re}\{G(s_j)\} = -\frac{s_j^2}{2} \hat{f}_{\hat{U}}^{\text{Ft}}(s_j; \xi) \cdot \beta + \epsilon_j \quad (j = 1, \dots, m),$$

where ϵ_j represents the regression error and $\beta = \int_0^T \sigma_t^2 dt$. Applying a linear regression of $\operatorname{Re}\{G(s_j)\}$ on $-s_j^2 \hat{f}_{\hat{U}}^{\text{Ft}}(s_j; \xi)/2$, we estimate $\int_0^T \sigma_t^2 dt$ by $\hat{\beta}$, the least squares estimator of β .

For any fixed $s \in \mathbb{R}$, it can be shown that $|\hat{f}_{\hat{U}}^{\text{Ft}}(s; \xi) - f_{\hat{U}}^{\text{Ft}}(s)| = O_p(n^{-1/2} + \xi)$. If we select $\xi = O(n^{-1/4})$ in (3), then Proposition 2 still holds if we replace $f_{\hat{U}}^{\text{Ft}}(s)$ by $\hat{f}_{\hat{U}}^{\text{Ft}}(s; \xi)$. The next result establishes the convergence rate of $\hat{\beta}$.

THEOREM 5. Under Assumptions 1–3, if $E(U_t^2) < \infty$ and we select $\xi = O(n^{-1/4})$, then

$$\hat{\beta} - \int_0^T \sigma_t^2 dt = O_p(n^{-1/4}).$$

The convergence rate stated in Theorem 5 is optimal in the sense of Gloter & Jacod (2001) and is the same as for the time domain estimator of Tao et al. (2013). Hence, our frequency domain method is rate-efficient for estimating the integrated volatility.

3. NUMERICAL STUDY

3.1. Practical implementation of the density estimator

To compute the density estimator \hat{f}_U in (5), we need to choose the bandwidth h and the parameter ξ . In our problem, doing this is much more complex than for standard nonparametric density estimators, since, unlike in that case, we do not have direct access to data from our target f_U ; therefore we cannot use existing smoothing parameter selection methods, which all require noise-free data. Moreover, unlike in standard nonparametric problems, we do not have access to a formula measuring the distance between f_U and its estimator.

Similar difficulties arise in the classical errors-in-variables problem, where one is interested in a density f_V but observes only data on $W = V + \varepsilon$, where $V \sim f_V$ is independent of $\varepsilon \sim f_\varepsilon$ with f_ε known. [Delaigle & Hall \(2008\)](#) proposed choosing the bandwidth h using a method called simulation approximation. Instead of computing h for f_V , they approximate h by extrapolating the bandwidths for estimating two other densities, f_1 and f_2 , that are related to f_V . The rationale of the extrapolation scheme is to exploit the analogous relationships between f_1 , f_2 and f_V .

Our problem is different because our estimator of f_U is completely different from that in [Delaigle & Hall \(2008\)](#), and we do not know f_X . Therefore we cannot apply their method directly; nevertheless, here we propose a method in the same spirit. In particular, we consider two density functions $f_1(\cdot) = 2f_U(\sqrt{2}\cdot) * f_U(\sqrt{2}\cdot)$ and $f_2 = 2f_1(\sqrt{2}\cdot) * f_1(\sqrt{2}\cdot)$, and observe that the way in which f_2 and f_1 are connected mimics the way in which f_1 and f_U are connected. As shown in the Appendix, data from both f_1 and f_2 can be made accessible, so that one may perform bandwidth selection for estimating them. We then choose the bandwidth for estimating f_U by an extrapolation with a ratio adjustment from the bandwidths for estimating f_1 and f_2 .

The algorithms for the bandwidth selection are given in the Appendix, and Algorithm A1 summarizes the main steps. Specifically, for $k = 1, 2$, our procedure requires the construction of variables $\Delta Y_{j,\ell}^k$ and time-points t_j^k , which are defined in, respectively, Algorithms A2 and A3. Step (c) of Algorithm A1 requires choosing values of (h, ξ) , say (h_k, ξ_k) for $k = 1, 2$, for estimating f_k by \hat{f}_k , where \hat{f}_k denotes our density estimator in (5) applied to the $\Delta Y_{j,\ell}^k$ s.

The idea is that if we knew f_1 , we would choose (h_1, ξ_1) by minimizing the integrated squared error of \hat{f}_1 , i.e., $(h_1, \xi_1) = \arg \min_{(h,\xi)} \int \{\hat{f}_1(x; \xi) - f_1(x)\}^2 dx$. In practice we do not know f_1 , but we can construct a relatively accurate estimator of it, namely the standard kernel density estimator \tilde{f}_1 of f_1 applied to the data $\Delta_{Y,j} = (Y_{t_{j+1}} - Y_{t_j})/\sqrt{2} \approx (U_{t_{j+1}} - U_{t_j})/\sqrt{2}$ ($j = 0, \dots, n-1$). This is computed by using the Gaussian kernel with bandwidth selected by the method of [Sheather & Jones \(1991\)](#). Using arguments similar to those in [Delaigle \(2008\)](#), under mild conditions, $\tilde{f}_1(x) = f_1(x) + O_p(n^{-2/5})$, whereas with the best possible choice of (h, ξ) , $\hat{f}_1(x) = f_1(x) + O_p(n^{-1/4})$, where the rate $n^{-1/4}$ cannot be improved. Thus, \tilde{f}_1 converges to f_1 faster than \hat{f}_1 does. This motivates us to approximate (h_1, ξ_1) defined above by

$$(h_1, \xi_1) = \arg \min_{(h,\xi)} \int \{\hat{f}_1(x; \xi) - \tilde{f}_1(x)\}^2 dx. \tag{8}$$

Paralleling the arguments in [Delaigle & Hall \(2008\)](#), it is more important to extrapolate the bandwidth h than ξ . Motivated by their results, we take $\xi_2 = \xi_1$. To choose h_2 , let \tilde{f}_2 be the standard kernel density estimator with the Gaussian kernel and bandwidth selected by the method of [Sheather & Jones \(1991\)](#), applied to the data $\Delta_{Y,j,2} = \{\Delta_{Y,j} - \Delta_{Y,k(j)}\}/\sqrt{2}$, where $k(j)$ is chosen at random from $0, \dots, n-1$. We choose $\Delta_{Y,j,2}$ in this way rather than $\Delta_{Y,j,2} = (\Delta_{Y,j} - \Delta_{Y,j+2})/\sqrt{2}$ to prevent accumulated residual X_t effects. Then we take

$$h_2 = \arg \min_h \int \{\hat{f}_2(x; \xi_1) - \tilde{f}_2(x)\}^2 dx. \tag{9}$$

Since \tilde{f}_1 and \tilde{f}_2 converge faster than \hat{f}_1 and \hat{f}_2 , they can be computed using less data than the latter. Therefore, when the time-points are widely unequally spaced, to compute the \tilde{f}_k s we suggest using only a fraction, say one-quarter, of the $\Delta_{Y,j}$ which correspond to the smallest $t_j - t_{j+1}$ values; that is, we use less but more accurate data for computing the \tilde{f}_k s.

Finally, as described in step (d) of Algorithm A1, we obtain our bandwidth for estimating f_U by an extrapolation with a ratio adjustment, $\hat{h} = h_1^2/h_2$, and we take $\hat{\xi} = \xi_1$. This method is not guaranteed to give the best possible bandwidth for estimating f_U , but it provides a sensible approximation for a problem that seems otherwise very hard, if not impossible, to solve. Theorem 1 implies that we have a lot of flexibility in choosing h , but it is impossible to know whether our bandwidth lies in the optimal range without knowing the exact orders of h_1 and h_2 . However, we cannot determine these orders without deriving complex second-order asymptotic results.

3.2. Simulations

We applied our method to data simulated from stochastic volatility models. We generated the data Y_{t_0}, \dots, Y_{t_n} as in (2). Following the convention that a financial year has 252 active days, we took $t \in [0, T]$ with $T = 1/252$, which represents one day of financial activity. We took time-points every Δs seconds, where $\Delta s = 30, 5$ or 1. Using the convention of 6.5 business hours in a trading day, this means that we took the t_j s to be equally spaced by $\Delta s/(252 \times 60 \times 60 \times 6.5)$ and that n was equal to $60 \times 60 \times 6.5/\Delta s$. We generated the microstructure noise U_t according to a normal or scaled t distribution, and for the X_t we used the Heston model

$$dX_t = \sigma_t dB_t, \quad d\sigma_t^2 = \kappa(\tau - \sigma_t^2) dt + \gamma\sigma_t dW_t,$$

where $E(dB_t dW_t) = \rho dt$ and κ, τ, γ and ρ are parameters. As in Ait-Sahalia & Yu (2009), we set the drift part of X_t to zero. The effect of the drift function is asymptotically negligible; see, for example, Xiu (2010). We considered two models, with values similar to those used by Ait-Sahalia & Yu (2009), which reflect practical scenarios in finance (see also Xiu, 2010; Liu & Tang, 2014): (i) $(\kappa, \tau, \gamma, \rho) = (6, 0.16, 0.5, -0.6)$; (ii) $(\kappa, \tau, \gamma, \rho) = (4, 0.09, 0.3, -0.75)$. In each case we took $X_0 = \log(100)$ and considered $U_t \sim N(0, \sigma_U^2)$ and $U_t \sim \sigma_U t(8)$, where $\sigma_U = 0.001$ and $\sigma_U = 0.005$. Typical X_t and Y_t paths for each model, plotted in the Supplementary Material, show that the X_t paths have smaller variation in model (ii) than in model (i). The X_t paths with smaller variation have less nuisance impact on estimators of U_t -related quantities. Thus, estimating the moments and density of U_t should be easier in model (ii).

In each setting, we generated 1000 samples of the form Y_{t_0}, \dots, Y_{t_n} and applied our estimator of the density f_U to each sample, thus obtaining 1000 density estimators \hat{f}_U computed as in (5). We chose the smoothing parameters as in § 3.1, and used the sinc kernel defined below (5). However, while this kernel guarantees optimal theoretical properties, in practice it produces negative wiggles in the tails, which we truncate to zero since f_U is a density. In the Supplementary Material, we show the results obtained when using the Gaussian kernel, which suggest that overall the sinc kernel works better, but the Gaussian kernel produces more attractive estimators in the tails. In cases where the sample has ties, for example when the sample size is large and the data are observed with only a few significant digits, as in our real-data example, the wiggles of the sinc kernel cause it to perform poorly, and significantly better results can be obtained by using the Gaussian kernel; see § 3.3.

For each estimator we computed the integrated squared error $\int (\hat{f}_U - f_U)^2$, and their median and first and third quartiles are reported in Table 1. In Fig. 1 we plot, for selected settings with normal error, the estimated curves \hat{f}_U computed from the samples corresponding to those three quartiles. Our results indicate that our density estimator works well. For a given setting, error densities with larger variances are easier to estimate. Figure 1 shows that our estimator improves as the sample size increases, that is, as Δs decreases. The estimated densities are better in model (ii) than in

Table 1. Median integrated squared error [first quartile, third quartile] of \hat{f}_U in (5), calculated from 1000 simulated samples from models (i) and (ii)

Δs	σ_U	Normal errors		Student errors	
		Model (i)	Model (ii)	Model (i)	Model (ii)
30	0.005	0.29 [0.13, 0.59]	0.26 [0.10, 0.58]	0.37 [0.19, 0.79]	0.36 [0.16, 0.76]
	0.001	9.98 [5.39, 16.6]	5.63 [2.90, 10.2]	9.82 [5.8, 15.59]	6.14 [3.28, 10.4]
5	0.005	0.10 [0.04, 0.27]	0.10 [0.04, 0.29]	0.12 [0.06, 0.27]	0.14 [0.07, 0.33]
	0.001	1.46 [0.64, 3.21]	1.02 [0.41, 2.18]	1.59 [0.78, 2.97]	1.16 [0.56, 2.19]
1	0.005	0.04 [0.01, 0.14]	0.04 [0.01, 0.15]	0.05 [0.02, 0.14]	0.05 [0.02, 0.13]
	0.001	0.32 [0.12, 1.01]	0.25 [0.10, 0.82]	0.42 [0.20, 0.95]	0.36 [0.15, 0.86]

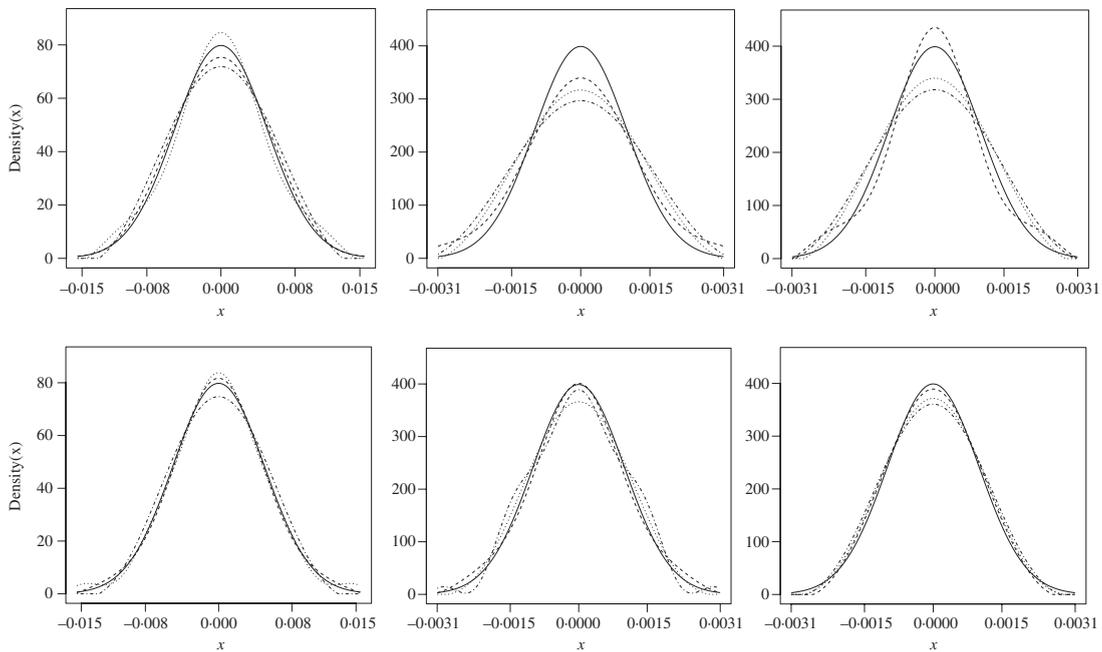


Fig. 1. The estimator $\hat{f}_U(x)$ in (5) in the case of normal errors, for three samples corresponding to the first (---), second (···) and third (-·-) quartiles of the integrated squared errors of estimators computed from data under model (i) with $\sigma_U = 0.005$ (left panels) and with $\sigma_U = 0.001$ (middle panels), and under model (ii) with $\sigma_U = 0.001$ (right panels), when $\Delta s = 30$ (upper panels) and when $\Delta s = 5$ (lower panels). In each panel the solid curve depicts the true $f_U(x)$.

model (i), as expected. While it is difficult to compare estimators of different target densities, the figures suggest that the difficulty in estimating the error densities depends more on the smoothness of X_t than on the error type. This reflects the fact that the smoothness of the error density has no first-order impact on the quality of estimators, as indicated by Theorems 1 and 2.

We also applied the moment estimators from § 2.4 to data simulated from a rescaled version of our two models; the rescaling was applied to avoid working with numbers all numerically rounded to zero. Specifically, we replaced $(\tau, \gamma, \sigma_U^2, X_0)$ in our models by $(c^2\tau, c\gamma, c^2\sigma_U^2, cX_0)$, where $c = 100$. We present the results in Table 2. These results indicate that, as expected, performance improves as the sample size increases. Moreover, the performance is best for lower-order moments, for higher noise levels, and for model (ii). For a given error variance, the moments are easier to recover when the errors have a rescaled Student distribution than when they have a normal distribution.

Table 2. Bias ($\times 10^2$), with standard deviation ($\times 10^2$) in parentheses, of $(\hat{M}_{U,2k} - M_{U,2k})/M_{U,2k}$ of our estimator $\hat{M}_{U,2k}$ in (6), calculated from 1000 simulated samples from models (i) and (ii) and computed using $\xi = t_2 - t_1$, for $k = 1$ and 2

Errors	Model	σ_U	$\Delta s = 30$		$\Delta s = 5$		$\Delta s = 1$	
			$k = 1$	$k = 2$	$k = 1$	$k = 2$	$k = 1$	$k = 2$
Normal	(i)	0.005	1.7 (6.4)	3 (18)	0.27 (2.5)	0.53 (7)	0.08 (1.1)	0.15 (3.3)
		0.001	41 (15)	99 (51)	6.8 (3.6)	14 (9.6)	1.4 (1.2)	2.8 (3.5)
	(ii)	0.005	1 (6.3)	1.6 (18)	0.15 (2.5)	0.29 (7)	0.06 (1.1)	0.11 (3.3)
		0.001	23 (10)	51 (31)	3.8 (2.9)	7.8 (8.1)	0.79 (1.2)	1.6 (3.4)
Student	(i)	0.005	1.4 (7.7)	0.5 (37)	0.28 (3)	0.13 (17)	0.03 (1.4)	-0.2 (9.9)
		0.001	30 (14)	45 (48)	5.2 (3.6)	6.9 (18)	1 (1.4)	1.1 (9.9)
	(ii)	0.005	0.9 (7.7)	-0.2 (37)	0.2 (3)	0.01 (17)	0.01 (1.4)	-0.23 (9.9)
		0.001	17 (10)	23 (43)	3 (3.2)	3.8 (17)	0.57 (1.4)	0.49 (9.9)

Finally, we applied our volatility estimator in each of our four settings and compared it with the estimator of Zhang (2006). For our method we took $\xi = t_2 - t_1$, and to choose s_1, \dots, s_m we took $m = 50$ equispaced points located between 0 and S , where S is the largest number for which $\hat{f}_U^{Ft}(S; \xi) \geq 0.99$. The results are presented in Table 3. In the Supplementary Material, we also show the first three quartiles of 100 times the relative absolute deviation, $|\hat{\beta} - \beta|/\beta$, of both estimators. Together, the results indicate that the two estimators give similar results overall. In the case where the error variance was large, our estimator tended to work a little better, and in the case where the error variance was small, the estimator of Zhang (2006) worked a little better. Specifically, when the error variance was small, theirs was a little less biased, but ours was a little less variable. As expected, the performance of both estimators improved as the sample size increased. The estimators worked a little better when the error variance was small: they were a little more biased but significantly less variable. The integrated volatility is an X_t -related quantity and is therefore easier to estimate when there are less errors contaminating the X_t . Depending on the situation, either the Student errors or the normal errors yielded better results.

3.3. Real-data analysis

We applied our procedure to analysing the high-frequency price data of Microsoft Corporation on the ten trading days from 19 March to 2 April 2013, available from the Trade and Quote database. We took the Y_t s to equal the log prices. Following Barndorff-Nielsen et al. (2011), we pre-processed the data by deleting entries that have zero or negative prices, deleting entries with negative correlation indicator, deleting entries with a letter code in COND, except for E or F, deleting entries outside the period 9:30 a.m. to 4:00 p.m., and using the median price if there were multiple entries at the same time. In this example the sample size is very large, which entails a large number of ties among the Y_t s. As a result, the sinc kernel produced overly wiggly estimates, regarding the data as coming from a multimodal density with modes located at the ties. The oscillation problems of this kernel are well known, but here they are exacerbated by the ties, so we used the standard Gaussian kernel, which is less affected by such issues. The only small adjustment we had to make was to break the ties by adding a small perturbation $\epsilon_t \sim N(0, a_j^2)$ to the $\Delta_{Y,j}$ s when computing the bandwidth of the standard kernel estimator $\tilde{f}_1(x)$ used in our bandwidth selection procedure, where $2a_j$ was equal to the maximum of the distance between each $\Delta_{Y,j}$ and its first smaller and larger nearest neighbours.

Table 3. Bias ($\times 10^2$), with standard deviation ($\times 10^2$) in parentheses, of $(\hat{\beta} - \beta)/\beta$ using our estimator of $\int_0^T \sigma_t^2 dt$ from §2.5, denoted by Ours, and the estimator of Zhang (2006), denoted by Zhang, calculated from 1000 simulated samples from models (i) and (ii)

σ_U	Δs	$\hat{\beta}$	Normal errors		Student errors	
			Model (i)	Model (ii)	Model (i)	Model (ii)
0.005	30	Ours	-1.12 (25.9)	0.42 (29.6)	-1.50 (28.0)	0.71 (33.6)
		Zhang	-2.87 (26.0)	-2.91 (29.7)	-3.89 (28.2)	-3.74 (33.7)
	5	Ours	-1.13 (17.5)	-0.65 (20.1)	-0.78 (18.3)	-0.53 (21.8)
		Zhang	-1.38 (17.5)	-1.18 (20.1)	-1.15 (18.3)	-1.25 (21.8)
	1	Ours	-0.61 (11.2)	-0.78 (12.7)	-0.59 (12.1)	-0.59 (14.4)
		Zhang	-0.64 (11.2)	-0.87 (12.7)	-0.65 (12.1)	-0.73 (14.5)
0.001	30	Ours	-5.29 (21.4)	-4.38 (21.8)	-6.25 (21.6)	-5.40 (22.1)
		Zhang	-2.80 (22.3)	-2.82 (22.4)	-4.25 (22.5)	-4.21 (22.7)
	5	Ours	-3.17 (14.4)	-2.54 (14.7)	-2.30 (14.5)	-1.80 (14.7)
		Zhang	-1.73 (14.8)	-1.71 (14.9)	-1.17 (14.8)	-1.16 (14.9)
	1	Ours	-1.02 (9.62)	-0.72 (9.71)	-1.15 (9.67)	-0.91 (9.76)
		Zhang	-0.31 (9.75)	-0.32 (9.79)	-0.63 (9.76)	-0.62 (9.80)

Figure 2 shows the error densities estimated by our method for three trading days in 2013: 20 March, 28 March and 2 April. In this example, the magnitude of the errors is about 10^{-3} smaller than that of the log prices themselves, but their aggregated impact on quantities such as integrated volatility is substantial. For example, for those three trading days the realized volatility was respectively 3.5, 4.5 and 3.5 times 10^{-4} , which is dominated by contributions from the errors. Indeed, for the same days, our error-corrected estimator was respectively 0.5, 0.5 and 0.3 times 10^{-4} , and the estimator of Zhang (2006) was respectively 0.5, 0.6 and 0.4 times 10^{-4} .

Interestingly, even for this short period of ten trading days, the distributions of the errors were quite different, especially in their tails. Since heavier tails can be linked with higher levels of variation and may affect the properties of the moments, it would be interesting to further investigate the tails of the error distributions of high-frequency financial data. Different tail behaviour may also be associated with different trading or market conditions on different days. For example, the behaviour of the error distributions may differ on days when the whole market or certain industrial segments such as IT are roaring. Hence, further investigations into empirical features of this kind connecting the error distributions with practical market conditions can be helpful for gaining better understanding of the microstructure noise in high-frequency financial data.

4. DISCUSSION

As a first attempt at a deconvolution approach with a contaminated slow-varying continuous process, our error density estimator method and the technical analysis in this paper assume that the distribution of errors in the high-frequency data is symmetric. Our method and analysis have shown potential for revealing various features of high-frequency financial data concerning the error distributions. Relaxing the symmetry assumption is possible and of interest, but challenging. The corresponding frequency domain method is under investigation. Our analysis focuses on the univariate setting for an individual stock or trading instrument. An interesting open question would be to extend the frequency domain analysis to a multivariate context, which would help

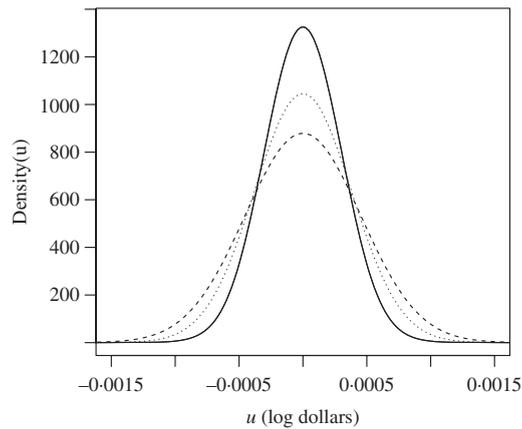


Fig. 2. Estimated densities of the errors contaminating the log prices of the Microsoft Corporation data for 20 March (solid), 28 March (dotted) and 2 April (dashed) 2013.

reveal the dependence structure between errors contaminating different stock prices. However, the problem is significantly more difficult because of issues such as high data dimensionality and asynchronous trading records in multivariate high-frequency data. We hope to address this problem in the future.

ACKNOWLEDGEMENT

We are grateful to the editor, an associate editor and two referees for their helpful suggestions. Chang was supported in part by the Fundamental Research Funds for the Central Universities of China, the National Natural Science Foundation of China, and the Center of Statistical Research and the Joint Lab of Data Science and Business Intelligence at Southwestern University of Finance and Economics. Delaigle was supported by a Future Fellowship and a Discovery Project from the Australian Research Council. Hall was supported by a Laureate Fellowship and a Discovery Project from the Australian Research Council. Tang acknowledges support from the U.S. National Science Foundation.

The main part of this work was completed while Jinyuan Chang was a postdoc under Peter Hall's supervision. Peter was a generous mentor and friend with a warm heart. He will be greatly missed.

SUPPLEMENTARY MATERIAL

[Supplementary material](#) available at *Biometrika* online includes a description of the simulation extrapolation bandwidth of [Delaigle & Hall \(2008\)](#), additional simulation results, a construction of confidence regions for the moment estimators of § 2.4, and all the proofs.

APPENDIX

Algorithm A1. Overview of our bandwidth selection procedure.

- (a) Find f_1 and f_2 so that the relationship between f_1 and f_2 mimics that between f_U and f_1 and so that f_1 and f_2 can be accurately estimated using our data.

- (b) For our estimator in (5), f_U is estimated from $\Delta Y_{j,\ell} = Y_{t_{j+\ell}} - Y_{t_j} \approx U_{t_{j+\ell}} - U_{t_j}$, with $|t_{j+\ell} - t_j|$ being small, $U_{t_{j+\ell}} \sim f_U$ and $U_{t_j} \sim f_U$ being independent. To imitate this, for $k = 1, 2$ we construct versions of $\Delta Y_{j,\ell}$, say $\Delta Y_{j,\ell}^k$, such that if $|t_{j+\ell}^k - t_j^k|$ is small, $\Delta Y_{j,\ell}^k \approx U_{t_{j+\ell}^k}^k - U_{t_j^k}^k$, where $U_{t_{j+\ell}^k}^k \sim f_k$ and $U_{t_j^k}^k \sim f_k$ are independent.
- (c) For $k = 1, 2$, we can compute a bandwidth h_k and a ξ_k for estimating f_k using our procedure in (5) applied to the data $\Delta Y_{j,\ell}^k$; see (8) and (9).
- (d) Step (a) suggests that $h/h_1 \approx h_1/h_2$, so that we can take $\hat{h} = h_1^2/h_2$.

Algorithm A2. Constructing $\Delta Y_{j,\ell}^1$ and t_j^1 .

- (a) For $j = 0, \dots, n-1$, let $t_j^1 = (t_j + t_{j+1})/2$ and $Y_{1,t_j^1}^1 = (Y_{t_j} + Y_{t_{j+1}})/\sqrt{2}$. For $\ell > 1$ and $j = 0, \dots, n-\ell-1$, take $\Delta Y_{j,\ell}^1 = Y_{1,t_{j+\ell}^1}^1 - Y_{1,t_j^1}^1 \approx U_{1,t_{j+\ell}^1}^1 - U_{1,t_j^1}^1$, where $U_{1,t_j^1}^1 = (U_{t_j} + U_{t_{j+1}})/\sqrt{2} \sim f_1$. This does not work for $\ell = 1$ because $Y_{1,t_{j+1}^1}^1 - Y_{1,t_j^1}^1 \approx (U_{t_{j+2}} - U_{t_j})/\sqrt{2} \not\sim f_1 * f_1$; we suggest taking $\Delta Y_{j,1}^1$ as in (b).
- (b) For $j = 0, \dots, n-2$, let $t_j^1 = (t_j + t_{j+2})/2$ and $Y_{2,t_j^1}^1 = (Y_{t_j} + Y_{t_{j+2}})/\sqrt{2}$. Take $\Delta Y_{j,1}^1 = Y_{2,t_{j+1}^1}^1 - Y_{2,t_j^1}^1 \approx U_{2,t_{j+1}^1}^1 - U_{2,t_j^1}^1$ with $U_{2,t_j^1}^1 = (U_{t_j} + U_{t_{j+2}})/\sqrt{2} \sim f_1$.

Algorithm A3. Constructing $\Delta Y_{j,\ell}^2$ and t_j^2 .

- Step 1.* For $j = 0, \dots, n-3$, let $Y_{4,t_j^2}^2 = \sum_{k=0}^3 Y_{t_{j+k}}/2$ and $t_j^2 = \sum_{k=0}^3 t_{j+k}/4$. For $\ell \geq 4$ and $j = 0, \dots, n-\ell-3$, take $\Delta Y_{j,\ell}^2 = Y_{4,t_{j+\ell}^2}^2 - Y_{4,t_j^2}^2 \approx U_{4,t_{j+\ell}^2}^2 - U_{4,t_j^2}^2$, where $U_{4,t_j^2}^2 = \sum_{k=0}^3 U_{t_{j+k}}/2 \sim f_2$. This does not work for $\ell = 1, 2, 3$ for similar reasons as in Algorithm A2; for $\ell = 1, 2, 3$, we suggest taking $\Delta Y_{j,\ell}^2$ as in Steps 2 and 3.
- Step 2.* For $j = 0, \dots, n-6$, let $Y_{3,t_j^2}^2 = \sum_{k=0,1,2,6} Y_{t_{j+k}}/2$ and $t_j^2 = \sum_{k=0,1,2,6} t_{j+k}/4$. For $j = 0, \dots, n-5$, let $Y_{2,t_j^2}^2 = \sum_{k=0,1,4,5} Y_{t_{j+k}}/2$ and $t_j^2 = \sum_{k=0,1,4,5} t_{j+k}/4$. For $j = 0, \dots, n-6$, let $Y_{1,t_j^2}^2 = \sum_{k=0}^3 Y_{t_{j+2k}}/2$ and $t_j^2 = \sum_{k=0}^3 t_{j+2k}/4$.
- Step 3.* For $j = 0, \dots, n-9$, take $\Delta Y_{j,3}^2 = Y_{3,t_{j+3}^2}^2 - Y_{3,t_j^2}^2 \approx U_{3,t_{j+3}^2}^2 - U_{3,t_j^2}^2$, where $U_{3,t_j^2}^2 = \sum_{k=0,1,2,6} U_{t_{j+k}}/2 \sim f_2$. For $j = 0, \dots, n-7$ and $\ell = 1, 2$, take $\Delta Y_{j,\ell}^2 = Y_{\ell,t_{j+\ell}^2}^2 - Y_{\ell,t_j^2}^2 \approx U_{\ell,t_{j+\ell}^2}^2 - U_{\ell,t_j^2}^2$, where $U_{1,t_j^2}^2 = \sum_{k=0}^3 U_{t_{j+2k}}/2 \sim f_2$ and $U_{2,t_j^2}^2 = \sum_{k=0,1,4,5} U_{t_{j+k}}/2 \sim f_2$.

REFERENCES

- AIT-SAHALIA, Y., FAN, J. & XIU, D. (2010). High-frequency covariance estimates with noisy and asynchronous financial data. *J. Am. Statist. Assoc.* **105**, 1504–17.
- AIT-SAHALIA, Y. & JACOD, J. (2014). *High-Frequency Financial Econometrics*. Princeton: Princeton University Press.
- AIT-SAHALIA, Y. & YU, J. (2009). High frequency market microstructure noise estimates and liquidity measures. *Ann. Appl. Statist.* **3**, 422–57.
- BARNDORFF-NIELSEN, O. E., HANSEN, P. R., LUNDE, A. & SHEPHARD, N. (2011). Multivariate realized kernels: Consistent positive semi-definite estimators of the covariation of equity prices with noise and non-synchronous trading. *J. Economet.* **162**, 149–69.
- BOLGER, N. & LAURENCEAU, J.-P. (2013). *Intensive Longitudinal Methods: An Introduction to Diary and Experience Sampling Research*. New York: Guilford Press.

- BUCKLEY, M. J., EAGLESON, G. K. & SILVERMAN, B. W. (1988). The estimation of residual variance in nonparametric regression. *Biometrika* **75**, 189–99.
- CARROLL, R. J. & HALL, P. (1988). Optimal rates of convergence for deconvolving a density. *J. Am. Statist. Assoc.* **83**, 1184–6.
- DELAIGLE, A. (2008). An alternative view of the deconvolution problem. *Statist. Sinica* **18**, 1025–45.
- DELAIGLE, A. & HALL, P. (2008). Using SIMEX for smoothing-parameter choice in errors-in-variables problems. *J. Am. Statist. Assoc.* **103**, 280–7.
- DELAIGLE, A., HALL, P. & MEISTER, A. (2008). On deconvolution with repeated measurements. *Ann. Statist.* **36**, 665–85.
- DELAIGLE, A. & ZHOU, W.-X. (2015). Nonparametric and parametric estimators of prevalence from group testing data with aggregated covariates. *J. Am. Statist. Assoc.* **110**, 1785–96.
- DUFRENOT, G., JAWADI, F. & LOUHICHI, W. (2014). *Market Microstructure and Nonlinear Dynamics: Keeping Financial Crisis in Context*. New York: Springer.
- GLOTER, A. & JACOD, J. (2001). Diffusions with measurement errors: I. Local asymptotic normality. *Eur. Ser. Appl. Indust. Math.* **5**, 225–42.
- HALL, P., KAY, J. W. & TITTERINGTON, D. M. (1990). Asymptotically optimal difference-based estimation of variance in nonparametric regression. *Biometrika* **77**, 521–8.
- HAUTSCH, N. (2012). *Econometrics of Financial High-Frequency Data*. New York: Springer.
- JACOD, J., LI, Y. & ZHENG, X. (2017). Statistical properties of microstructure noise. *Econometrica* **85**, 1133–74.
- LIU, C. & TANG, C. Y. (2014). A quasi-maximum likelihood approach for integrated covariance matrix estimation with high-frequency data. *J. Economet.* **180**, 217–32.
- MEISTER, A. (2007). Optimal convergence rates of density estimation from grouped data. *Statist. Prob. Lett.* **77**, 1091–7.
- MÜLLER, H.-G., SEN, R. & STADTMÜLLER, U. (2011). Functional data analysis for volatility. *J. Economet.* **165**, 233–45.
- OLHEDE, S. C., SYKULSKI, A. M. & PAVLIOTIS, G. A. (2009). Frequency domain estimation of integrated volatility for Itô processes in the presence of market-microstructure noise. *Multiscale Mod. Simul.* **8**, 393–427.
- SHEATHER, S. J. & JONES, M. C. (1991). A reliable data-based bandwidth selection method for kernel density estimation. *J. R. Statist. Soc. B* **53**, 683–90.
- STEFANSKI, L. A. & CARROLL, R. J. (1990). Deconvoluting kernel density estimators. *Statistics* **21**, 169–84.
- TAO, M., WANG, Y. & ZHOU, H. H. (2013). Optimal sparse volatility matrix estimation for high dimensional Itô processes with measurement errors. *Ann. Statist.* **41**, 1816–64.
- WANG, J. -L., CHIOU, J.-M. & MÜLLER, H. G. (2016). Review of functional data analysis. *Ann. Rev. Statist. Appl.* **3**, 257–95.
- XIU, D. (2010). Quasi-maximum likelihood estimation of volatility with high-frequency data. *J. Economet.* **159**, 235–50.
- ZHANG, L. (2006). Efficient estimation of stochastic volatility using noisy observations: A multi-scale approach. *Bernoulli* **12**, 1019–43.
- ZHANG, L., MYKLAND, P. A. & AÏT-SAHALIA, Y. (2005). A tale of two time scales: Determining integrated volatility with noisy high-frequency data. *J. Am. Statist. Assoc.* **100**, 1394–411.

[Received on 10 April 2017. Editorial decision on 10 December 2017]