

FREQUENCY DOMAIN ESTIMATION OF INTEGRATED VOLATILITY FOR ITÔ PROCESSES IN THE PRESENCE OF MARKET-MICROSTRUCTURE NOISE*

S. C. OLHEDE[†], A. M. SYKULSKI[‡], AND G. A. PAVLIOTIS[‡]

Abstract. This paper proposes a novel multiscale estimator for the integrated volatility of an Itô process in the presence of market microstructure noise (observation error). The multiscale structure of the observed process is represented frequency by frequency, and the concept of the multiscale ratio is introduced to quantify the bias in the realized integrated volatility due to the observation error. The multiscale ratio is estimated from a single sample path, and a frequency-by-frequency bias correction procedure is proposed, which simultaneously reduces variance. We extend the method to include correlated observation errors and provide the implied time-domain form of the estimation procedure. The new method is implemented to estimate the integrated volatility for the Heston and other models, and the improved performance of our method over existing methods is illustrated by simulation studies.

Key words. bias correction, market microstructure noise, realized volatility, multiscale inference, Whittle likelihood

AMS subject classifications. 62M15, 62M05, 62F10, 60J60

DOI. 10.1137/090756363

1. Introduction. Diffusion (Itô) processes are being used as models in many applications such as physics, biology, finance, and atmosphere/ocean science [4, 15, 18, 24, 31, 33]. Some examples include the stochastic modeling of epidemics, the theory of derivative pricing, and stochastic modeling in oceanography. Most of the stochastic models that are used in applications involve unknown parameters which can be, in principle, estimated from observations of the Itô process, using maximum likelihood or another choice of estimation procedure. The theory of statistical inference for diffusion processes is very well developed [38, 28, 7].

In many cases the observations of a diffusion process are contaminated by high-frequency observation error. Examples include data in econometrics [44] and oceanography [18]. It is important to develop accurate and efficient statistical inference procedures that take into account this contamination of the observed high-frequency data, to ensure well-behaved procedures even in the limit of the sampling period tending to zero. The goal of this paper is to address this issue by developing statistical inference methodologies in the frequency domain.

The problem that we shall study in this paper can be stated as follows: Let X_t be the solution of a scalar stochastic differential equation (SDE)

$$(1.1) \quad dX_t = \mu_t dt + \sigma_t dB_t, \quad X_0 = x_0,$$

where B_t denotes a standard one-dimensional Brownian motion and μ_t, σ_t^2 are stochastic processes, corresponding to the drift and the volatility of the SDE. $\mu_t, \sigma_t,$

*Received by the editors April 16, 2009; accepted for publication (in revised form) September 22, 2009; published electronically December 9, 2009.

<http://www.siam.org/journals/mms/8-2/75636.html>

[†]Department of Statistical Science and Department of Computer Science, University College London, London WC1E6BT, United Kingdom (s.olhede@ucl.ac.uk).

[‡]Department of Mathematics, Imperial College London, South Kensington, London SW72AZ, United Kingdom (adam.sykulski@imperial.ac.uk, g.pavliotis@imperial.ac.uk).

and the Brownian motion B_t are (in general) correlated, and μ_t as well as σ_t may be functions of X_t . It has long been recognized that constant volatility models (i.e., models for which $\sigma_t^2 = \sigma^2$ for all t) cannot explain many of the empirical features of the time series behavior of stock and option prices. Richer classes of models, such as stochastic volatility models, i.e., models where $\{\sigma_t^2\}$ is itself a stochastic process like the one that was introduced by Heston [22], have been used for many years to model time series of stock and option prices [42, 16]. As $\{\sigma_t^2\}$ is a stochastic process, such models are referred to as “stochastic volatility models.”

Let $\{Y_t\}_{j=1}^{N+1}$ be the regularly sampled, noisy observations of the process X_t , or

$$(1.2) \quad Y_{t_j} = X_{t_j} + \varepsilon_{t_j}, \quad j = 1, 2, \dots, N+1, \quad t_j := \frac{j-1}{N}T = (j-1)\Delta t,$$

with ε_{t_j} the observational error. We assume the data is sampled at regularly spaced time points with a sampling period of Δt . The length of the path $T = N\Delta t$ is assumed fixed. The observation noise $\{\varepsilon_{t_j}\}_{j=1}^{N+1}$ is assumed to be a white noise process with variance σ_ε^2 , and is assumed to be independent of the noise that drives the Itô process X_t .

Our main objective is to estimate the *integrated stochastic volatility*

$$(1.3) \quad \langle X, X \rangle_T = \int_0^T \sigma_t^2 dt$$

of the Itô process $\{X_t\}$ from the set of observations $\{Y_{t_j}\}_{j=1}^{N+1}$. The integrated stochastic volatility is a nonparametric measure of volatility for an asset return, which simply summarizes the natural variability of the process over an interval of time. The problem of estimating the integrated stochastic volatility has attracted considerable attention in recent years; see, for example, [2] and the references therein. In this paper we will use the terminology that is common in econometrics and apply our estimator to a stochastic volatility model. We emphasize also that similar problems arise in many other areas where the process of interest is modeled using an SDE, such as, for example, oceanography [18, 1]. In such cases one is interested in estimating the eddy diffusivity from noisy Lagrangian observations. This problem was, for example, studied recently in [11].

We let the *realized integrated volatility* of X_t be written as

$$[X, X]_T = \sum_{j=1}^N (X_{t_{j+1}} - X_{t_j})^2.$$

It is well known [25] that

$$(1.4) \quad \lim_{\Delta t \rightarrow 0} [X, X]_T \equiv \lim_{\Delta t \rightarrow 0} \sum_{j=1}^N (X_{t_{j+1}} - X_{t_j})^2 = \langle X, X \rangle_T,$$

where we keep $N\Delta t = T$ fixed, and where the limit is in probability. Consequently, in the absence of market microstructure noise (i.e., when $Y_{t_j} = X_{t_j}$, $j = 1, \dots, N+1$) the integrated volatility can be estimated from the realized integrated volatility of Y_t [44] of $[Y, Y]_T \equiv \sum_{j=1}^N (Y_{t_{j+1}} - Y_{t_j})^2$. In the presence of market microstructure noise, $\lim_{\Delta t \rightarrow 0} [Y, Y]_T \neq \langle X, X \rangle_T$; see, for example, the discussion in [48]. In fact, the realized integrated volatility is in this case dominated by the energy of the observation

error, and it becomes an estimate of the variance of the noise process σ_ε^2 , rather than an estimate of the integrated volatility. To make this observation from a more intuitive perspective, consider the case where there is no drift in (1.1), i.e., when $\mu = 0$ for all $t \in [0, T]$ (see also the numerical results reported in section 3.2). In this case the increment of the observed process Y_t can directly be written as

$$(1.5) \quad \Delta Y_{t_j} = Y_{t_{j+1}} - Y_{t_j} \approx \sigma_{t_{j+1}} \sqrt{\Delta t} \zeta_j + \sigma_\varepsilon \sqrt{2} \xi_j,$$

where ζ_j and ξ_j are two independent Gaussian white noise processes. From (1.5) one may note that taking first differences of Y_{t_i} at decreasing Δt will not be sufficient to distinguish between the observation noise and the true noise driving the process X_t , and furthermore, as Δt decreases, the dominant contribution to the difference will be $\sigma_\varepsilon \sqrt{2} \xi_j$ rather than $\sigma_{t_{j+1}} \sqrt{\Delta t} \zeta_j$, as the former term does not decrease with decreasing Δt .

A different estimation procedure is therefore necessary when estimating the integrated volatility of X_t from $\{Y_{t_j}\}$, such as increasing Δt to make the first contribution in (1.5) competitive in magnitude (i.e., subsampling the data). This fact was used in [3, 48], where accurate and efficient methods for estimating the integrated volatility for noisy observations were developed. In these papers a method for estimating the integrated volatility which consisted of subsampling at an optimal sampling rate combined with averaging and an appropriate debiasing step was proposed. Various other estimators were suggested in [48, 44, 13, 17, 21] for processes contaminated by high-frequency nuisance structure.

In order to estimate the variances of X_t and ε_t , we need to look at differences over different time scales, since X_t and ε_t have different characteristic time scales. Another natural way to do so at *all* scales is by using the Fourier transform of the discrete increments of the observed process ΔY_{t_j} , as will be further discussed. In addition most of the time-domain methods can be put in a unified framework as linear filtering techniques, i.e., as a convolution with a linear kernel, of some time-domain quadratic function of the data. The understanding of these methods is enhanced by studying them directly in the frequency domain, as convolutions in time are multiplications in frequency. Fourier domain estimators of the integrated volatility have been proposed for observations devoid of microstructure features; see [20, 5, 29]. Fourier domain estimators have also been used for estimating noisy Itô processes (i.e., processes of the form (1.2); see [30, 39, 40]), based on smoothing the time-domain quantities by using only a limited number of frequencies in the reconstruction.

The bias in using the realized integrated volatility of the observed process Y_{t_j} to estimate the integrated volatility of X_t due to the observation noise ε_{t_j} can also be understood directly in the frequency domain, since the energy or variance associated with each frequency is contaminated by the microstructure noise process. This bias is particularly damaging at high frequencies, as we shall demonstrate. In this article we propose a frequency-by-frequency debiasing procedure to improve the accuracy of the estimation of the integrated volatility. The proposed estimation method can also be viewed in the time domain as smoothing the empirical autocovariance of the increments of the process, but where the implied time-domain smoothing kernel is itself estimated from the observed process.

The proposed estimator can be described roughly as follows. Let $\{J_k^{(X)}\}$ denote the discrete Fourier transform of the differenced sampled X_t process (i.e., the discrete derivative of X_t) or of $\Delta X_{t_j} = X_{t_{j+1}} - X_{t_j}$, and similarly for Y_{t_j} and ε_{t_j} . The integrated volatility can be written in terms of the inverse discrete Fourier transform

of the variance of $J_k^{(X)}$. We calculate the bias in the variance of $J_k^{(Y)}$ when using its sample estimator to estimate the variance of $J_k^{(X)}$. The high-frequency coefficients are heavily contaminated by the microstructure noise. With a formula for the bias it is possible to debias the estimated variance of the Fourier transform at every frequency, with the unknown parameters of the bias estimated using the Whittle likelihood [46, 47]. This produces a debiased estimator of the integrated volatility via an aggregation of the estimated variance, and we show also that the variance of the proposed estimator is reduced by the debiasing.

Our estimator shows highly competitive mean square error performance; it also has several advantages over existing estimators. First, it is robust with respect to the signal-to-noise ratio; furthermore, it is easy to formulate and to implement; in addition, it readily generalizes to the case of correlated observation errors (in time). Finally, the properties of our estimator are transparent using frequency domain analysis.

The rest of the paper is organized as follows. In section 2 we introduce our estimator and present some of its properties, stated in Theorems 2.3 and 2.4. We also discuss the time-domain understanding of the proposed method and the extension of the method to the case where the observation noise is correlated. In section 3 we present the results of Monte Carlo simulations for our estimator. Section 4 is reserved for conclusions. Various technical results are included in the appendices.

2. Estimation methods. Let $\{Y_{t_j}\}$ be given by (1.2), where the zero-mean noise $\{\varepsilon_{t_j}\}$ is independent of $\{X_{t_j}\}$ and its variance at any time is equal to σ_ε^2 . The simplest estimator of the integrated volatility of X_t would ignore the high-frequency component of the data and use the realized integrated volatility of the observed process. The realized integrated volatility is growing in magnitude with increasing sample sizes, and we note that this biased and naive estimator has order

$$(2.1) \quad \widehat{\langle X, X \rangle}_T^{(b)} = [Y, Y]_T = \mathcal{O}\left(\frac{1}{\Delta t}\right).$$

This estimator is both inconsistent and biased; see [21]. An estimator is consistent if, as $N \rightarrow \infty$ (or equivalently as $\Delta t = T/N \rightarrow 0$), it converges in probability to the parameter it is used to estimate. It is highly unsatisfactory to obtain an inconsistent estimator, and we shall in this paper seek to remedy the problem of inconsistency when using $\widehat{\langle X, X \rangle}_T^{(b)}$. For comparative purposes, we define also the unbiased estimator of the realized integrated volatility of the sampled process $\{X_{t_j}\}$ as a “gold standard” estimator of integrated volatility:

$$(2.2) \quad \widehat{\langle X, X \rangle}_T^{(u)} = [X, X]_T = \mathcal{O}(N\Delta t) = \mathcal{O}(1).$$

This cannot be used in practice, as X_{t_j} is not directly observed. Both $\widehat{\langle X, X \rangle}_T^{(u)}$ and $\widehat{\langle X, X \rangle}_T^{(b)}$ are estimators of the integrated volatility (quadratic variation) of X .

2.1. Fourier domain properties. We shall start by deriving an alternative representation of (2.1) to motivate further development. First we define the increment process of a sample from a generic time series U_{t_j} , $j = 1, \dots, N + 1$, as $\Delta U_{t_j} = U_{t_{j+1}} - U_{t_j}$, $j = 1, \dots, N$. The discrete Fourier transform of ΔU_{t_j} is denoted by $J_k^{(U)}$ and takes the form [36, p. 206]

$$(2.3) \quad J_k^{(U)} = \sqrt{\frac{1}{N}} \sum_{j=1}^N \Delta U_{t_j} e^{-2\pi i t_j f_k}, \quad f_k = \frac{k}{T}, \quad U = X, Y, \varepsilon.$$

Note the square root normalization; this is not standard in the definition of the discrete Fourier transform, but is a standard normalization in statistics. The reason for this is that unitary transforms have convenient second order properties such as transforming a unit variance vector to another unit variance vector. Our proposed estimator will be based on examining the second order properties of $\{J_k^{(Y)}\}$. $|J_k^{(Y)}|^2$ is the *periodogram* [9] defined for a time series and is an inconsistent estimator of $\text{var}\{J_k^{(Y)}\} = \mathcal{S}_{k,k}^{(X)}$. This variance (when based on an infinite length sample) is generally referred to as the *spectral density* of the process. Inconsistent estimators do not have decreasing variances with increasing sample sizes, and $|J_k^{(Y)}|^2$ will not converge in probability to its expectation, namely, the true spectral density of the first difference of Y_{t_j} . We examine the properties of $\{J_k^{(X)}\}$. We have, with $\bar{\mu}_j = \frac{1}{\Delta t} \int_{(j-1)\Delta t}^{j\Delta t} \mu_s ds$ denoting the local average of μ_t ,

$$\begin{aligned} \Delta X_{t_j} &= \int_{(j-1)\Delta t}^{j\Delta t} [\mu_s ds + \sigma_s dW_s] = \bar{\mu}_j \Delta t + \int_{(j-1)\Delta t}^{j\Delta t} \sigma_s dW_s, \\ J_k^{(X)} &= \sqrt{\frac{1}{N}} \sum_{j=1}^N \Delta X_{t_j} e^{-2i\pi \frac{kj}{N}} = \sqrt{\frac{1}{N}} \sum_{j=1}^N \left[\Delta t \bar{\mu}_j + \int_{(j-1)\Delta t}^{j\Delta t} \sigma_s dW_s \right] e^{-2i\pi \frac{kj}{N}} \\ (2.4) \quad &= \mathcal{O}\left(\Delta t^{1/2} \frac{T}{k}\right) + \sqrt{\frac{1}{N}} \sum_{j=1}^N \int_{(j-1)\Delta t}^{j\Delta t} \sigma_s dW_s e^{-2i\pi \frac{kj}{N}}. \end{aligned}$$

We define

$$\tilde{J}_k^{(X)} = \sqrt{\frac{1}{N}} \sum_{j=1}^N \int_{(j-1)\Delta t}^{j\Delta t} \sigma_s dW_s e^{-2i\pi \frac{kj}{N}},$$

and this to leading order approximates $J_k^{(X)}$ as $\Delta t \rightarrow 0$ for all but a few frequencies, namely, k very small, as indicated by the order terms in (2.4). This follows, as μ_s is an Itô process and has almost surely continuous paths, which implies that

$$(2.5) \quad \Delta t^2 \sum_{k=0}^{N-1} \left| \sqrt{\frac{1}{N}} \sum_{j=1}^N \bar{\mu}_j e^{-2i\pi \frac{kj}{N}} \right|^2 \sim \sum_k \left| \frac{1}{\sqrt{N}} \Delta t N/k \right|^2 = \mathcal{O}(\Delta t),$$

$$(2.6) \quad \Delta t \sum_{k=0}^{N-1} \left| \sqrt{\frac{1}{N}} \sum_{j=1}^N \bar{\mu}_j e^{-2i\pi \frac{kj}{N}} \right|^2 \sim \sum_k \left| \frac{1}{\sqrt{N}} \Delta t N/k \right|^2 = \mathcal{O}(\log(\Delta t) \sqrt{\Delta t}),$$

as $\Delta t \sum_j \bar{\mu}_j e^{-2i\pi \frac{kj}{N}} = \mathcal{O}\left(\frac{T}{k}\right)$. So we only need calculate, to leading order, $\sum_{k=0}^{N-1} |\tilde{J}_k^{(X)}|^2 = \mathcal{O}(1)$ when calculating the properties of $\sum_{k=0}^{N-1} |J_k^{(X)}|^2$ from (2.5) and (2.6). More formally we note that

$$\sum_{k=0}^{N-1} |J_k^{(X)}|^2 = \sum_{k=0}^{N-1} |\tilde{J}_k^{(X)}|^2 + \mathcal{O}(\log(\Delta t) \sqrt{\Delta t}).$$

We need to determine the first and second order structures of $\{\tilde{J}_k^{(X)}\}_k$. In general $\{\tilde{J}_k^{(X)}\}_k$ is a complex-valued random vector, which may not be a sample from a multivariate Gaussian distribution. The covariance matrix of a complex random vector \mathbf{Z}

is given by $\text{cov}\{\mathbf{Z}, \mathbf{Z}\} = \mathbb{E}\{\mathbf{Z}\mathbf{Z}^H\} - \mathbb{E}\{\mathbf{Z}\}\mathbb{E}\{\mathbf{Z}\}^H$ [32, 37]. We have

$$\mathbb{E}\{\tilde{J}_k^{(X)}\} = 0, \quad k = 1, \dots, N - 1.$$

Furthermore, with $\tilde{\mathcal{F}}_{k_1, k_2}^{(X)} = \mathbb{E}\{\tilde{J}_{k_1}^{(X)} \tilde{J}_{k_2}^{(X)*}\}$,

$$\begin{aligned} \tilde{\mathcal{F}}_{k_1, k_2}^{(X)} &= \frac{1}{N} \mathbb{E}\left\{ \sum_{n=1}^N \int_{(n-1)\Delta t}^{n\Delta t} \sum_{l=1}^N \int_{(l-1)\Delta t}^{l\Delta t} \sigma_s dW_s \sigma_t dW_t e^{-2i\pi(\frac{k_1 n}{N} - \frac{k_2 l}{N})} \right\} \\ &= \frac{1}{N} \sum_{n=1}^N \int_{(n-1)\Delta t}^{n\Delta t} \sum_{l=1}^N \int_{(l-1)\Delta t}^{l\Delta t} \mathbb{E}\{\sigma_s dW_s \sigma_t dW_t\} e^{-2i\pi(\frac{k_1 n}{N} - \frac{k_2 l}{N})} \\ &= \frac{1}{N} \sum_{n=1}^N \int_{(n-1)\Delta t}^{n\Delta t} \sum_{l=1}^N \int_{(l-1)\Delta t}^{l\Delta t} \mathbb{E}\{\sigma_s \sigma_t\} \delta_{nl} \delta(t - s) e^{-2i\pi(\frac{k_1 n}{N} - \frac{k_2 l}{N})} ds dt. \end{aligned}$$

In particular we have that

$$\begin{aligned} \tilde{\mathcal{F}}_{k, k}^{(X)} &= \frac{1}{N} \int_0^T \mathbb{E}\{\sigma_s^2\} ds + \mathcal{O}(\Delta t^2) := \bar{\sigma}_X^2 + \mathcal{O}(\Delta t^2) \\ (2.7) \quad &= \frac{\langle X, X \rangle_T}{N} + \mathcal{O}(\Delta t^2), \end{aligned}$$

where the error terms are due to the Riemann approximation to an integral, and thus it follows that

$$(2.8) \quad \sum_{k=0}^{N-1} \tilde{\mathcal{F}}_{k, k}^{(X)} = \int_0^T \mathbb{E}\{\sigma_s^2\} ds + \mathcal{O}(\Delta t).$$

$\bar{\sigma}_X^2$ does not depend on the value of k . Malliavin and Mancino [29], in contrast, under very light assumptions show how the Fourier coefficients of $\{\sigma_t^2\}$ can be calculated from the Fourier coefficients of dX_t , using a Parseval–Rayleigh relationship; see also [40, 30]. We can from (2.7) make a stronger link from the Fourier transform to the integrated volatility than that of the Parseval–Rayleigh relationship, and shall use this “uniformity of energy” to estimate the microstructure bias.

We note that the covariance between different frequencies is given by

$$\begin{aligned} \tilde{\mathcal{F}}_{k_1, k_2}^{(X)} &= \frac{1}{N} \sum_{n=1}^N \int_{(n-1)\Delta t}^{n\Delta t} \mathbb{E}\{\sigma_t^2\} dt e^{-2i\pi n(\frac{k_1}{N} - \frac{k_2}{N})} \\ &= \frac{1}{N} \int_0^T \mathbb{E}\{\sigma_t^2\} dt e^{-2i\pi t(\frac{k_1}{T} - \frac{k_2}{T})} + \mathcal{O}(\Delta t^2). \end{aligned}$$

Let $\Xi(f) := \int_0^T \mathbb{E}\{\sigma_t^2\} e^{-2i\pi ft} dt$. We can bound the size of $\Xi(\frac{k_1}{T} - \frac{k_2}{T})$ as $|k_1 - k_2|$ increases. As $\mathbb{E}\{\sigma_t^2\}$ is smooth in t , the modulus of the covariance can be bounded for increasing $|k_1 - k_2|$, as the Fourier transform $\Xi(\frac{k_1}{T} - \frac{k_2}{T})$ decays proportionally to $|k_1 - k_2|^{-\alpha-1}$, where α is the number of smooth derivatives of $\mathbb{E}\{\sigma_t^2\}$. We can also directly note that the variance of the discrete Fourier transform of the noise is precisely (this is *not* a large sample result)

$$(2.9) \quad \mathcal{I}_{k_1, k_2}^{(\varepsilon)} = \sigma_\varepsilon^2 |2 \sin(\pi f_{k_1} \Delta t)|^2 \delta_{k_1, k_2}$$

by virtue of being the first difference of white noise (see also [8]). The naive estimator can therefore be rewritten, with $\mathcal{S}_{k_1, k_2}^{(Y)} = \text{cov}\{J_{k_1}^{(Y)}, J_{k_2}^{(Y)}\}$, as

$$(2.10a) \quad \widehat{\langle X, X \rangle}_T^{(b)} = \sum_{j=1}^N \Delta Y_{t_j}^2 = \sum_{k=0}^{N-1} \left| J_k^{(Y)} \right|^2,$$

$$(2.10b) \quad \widetilde{\mathcal{S}}_{k,k}^{(Y)} = \left| J_k^{(Y)} \right|^2,$$

$$(2.10c) \quad \mathbb{E} \left\{ \widehat{\langle X, X \rangle}_T^{(u)} \right\} = \sum_{k=0}^{N-1} \widetilde{\mathcal{S}}_{k,k}^{(X)} + \mathcal{O} \left(\log(\Delta t) \sqrt{\Delta t} \right) + \mathcal{O}(\Delta t)$$

$$(2.10d) \quad \equiv \sum_{k=0}^{N-1} \mathcal{S}_{k,k}^{(X)}.$$

Recall that $\widehat{\langle X, X \rangle}_T^{(b)}$ is the biased estimator constructed by treating Y_t as if it were X_t , and $\widehat{\langle X, X \rangle}_T^{(u)}$ is the unbiased estimator constructed from the unobserved X_t ; cf. (2.1) and (2.2). The Parseval–Rayleigh relationship in (2.10a) is discussed in [30] and is used in [29]. We shall now develop a frequency domain specification of the bias of the naive estimator.

LEMMA 2.1 (frequency domain bias of the naive estimator). *Let X_t be an Itô process and assume that the covariance of $J_{k_1}^{(X)}$ and $J_{k_2}^{(X)}$ is $\mathcal{S}_{k_1, k_2}^{(X)}$ with the chosen sampling. Then the naive estimator of the integrated volatility given by (2.10) has an expectation given by*

$$\mathbb{E} \left\{ \widehat{\langle X, X \rangle}_T^{(b)} \right\} = \sum_{k=0}^{N-1} \left(\widetilde{\mathcal{S}}_{k,k}^{(X)} + \sigma_\varepsilon^2 |2 \sin(\pi f_k \Delta t)|^2 \right) + \mathcal{O} \left(\log(\Delta t) \sqrt{\Delta t} \right).$$

Thus

$$(2.11) \quad \begin{aligned} \mathbb{E} \left\{ \widehat{\langle X, X \rangle}_T^{(b)} \right\} &= \mathbb{E} \left\{ \widehat{\langle X, X \rangle}_T^{(u)} \right\} + \sum_{k=0}^{N-1} \sigma_\varepsilon^2 |2 \sin(\pi f_k \Delta t)|^2 + \mathcal{O} \left(\log(\Delta t) \sqrt{\Delta t} \right) \\ &= \mathcal{O}(1) \quad + \mathcal{O}(\Delta t^{-1}) \quad + \mathcal{O} \left(\log(\Delta t) \sqrt{\Delta t} \right). \end{aligned}$$

We note the order of the three terms directly to clarify their behavior when estimating the integrated volatility with decreasing Δt .

Proof. This result follows from the independence of $\{\varepsilon_t\}$ and $\{X_t\}$, combined with (2.8) and (2.9). \square

We notice directly from (2.11) that the relative frequency contribution of ΔX_t and ε_t , i.e., $\mathcal{S}_{k,k}^{(X)}$ compared to the noise contribution $\sigma_\varepsilon^2 |2 \sin(\pi f_k \Delta t)|^2$, determines the inherent bias of $\widehat{\langle X, X \rangle}_T^{(b)}$. Estimator (2.10) is inconsistent and biased since it is equivalent to estimator (2.1), and such a procedure would give an unbiased estimator of the integrated volatility only when $\sigma_\varepsilon^2 = 0$. When the estimator is expressed in the time domain, the microstructure cannot be disentangled from the Itô process. On the other hand, in the frequency domain, from the very nature of a multiscale process, the contributions to $\widetilde{\mathcal{S}}_{k,k}^{(Y)}$ can be disentangled.

2.2. Multiscale modelling. To correct the biased estimator we need to correct the usage of the biased estimator of $\mathcal{S}_{k,k}^{(X)}, \widehat{\mathcal{F}}_{k,k}^{(Y)}$, at each frequency. We therefore define a new *shrinkage*¹ estimator [45, p. 155] of $\mathcal{S}_{k,k}^{(X)}$ for $0 \leq L_k \leq 1$ by

$$(2.12) \quad \widehat{\mathcal{F}}_{k,k}^{(X)}(L_k) = L_k \widehat{\mathcal{F}}_{k,k}^{(Y)}.$$

L_k is referred to as the *multiscale ratio*, and its optimal form for perfect bias correction is for an arbitrary Itô process given by

$$(2.13) \quad L_k = \frac{\mathcal{S}_{k,k}^{(X)}}{\mathcal{S}_{k,k}^{(X)} + \sigma_\varepsilon^2 |2 \sin(\pi f_k \Delta t)|^2}.$$

This quantity *cannot* be calculated without explicit knowledge of $\mathcal{S}_{k,k}^{(X)}$ and σ_ε^2 . The estimator in (2.12) with L_k as in (2.13) therefore yields an *oracle estimator* of $\mathcal{S}_{k,k}^{(X)}$; see also [45, p. 172]. We can use (2.7) to simplify (2.13) to obtain

$$(2.14) \quad L_k = \frac{\bar{\sigma}_X^2}{\bar{\sigma}_X^2 + \sigma_\varepsilon^2 |2 \sin(\pi f_k \Delta t)|^2}.$$

For a fixed $0 \leq L_k \leq 1$

$$\begin{aligned} \mathbb{E} \left\{ \widehat{\mathcal{F}}_{k,k}^{(X)}(L_k) \right\} &= L_k \mathbb{E} \left\{ |J_k^{(Y)}|^2 \right\} \\ &= \bar{\sigma}_X^2 + \mathcal{O} \left(\frac{\Delta t^{3/2}}{k} \right), \end{aligned}$$

where the order terms follow from the continuity of μ_s . We can define a new estimator for the true L_k via

$$(2.15) \quad \widehat{\langle X, X \rangle}_T^{(m_3)} = \sum_{k=0}^{N-1} \widehat{\mathcal{F}}_{k,k}^{(X)}(L_k),$$

where

$$\mathbb{E} \left\{ \widehat{\langle X, X \rangle}_T^{(m_3)} \right\} = \langle X, X \rangle_T + \mathcal{O} \left(\log(\Delta t) \sqrt{\Delta t} \right).$$

Recall that $\langle X, X \rangle_T = O(T) = O(1)$. Consequently, to leading order we can remove the bias from the naive estimator if we know the multiscale ratio, L_k . We shall now develop a multiscale understanding of the process under observation and use this to construct an estimator for the multiscale ratio.

2.3. Estimation of the multiscale ratio. We have a two-parameter description on how the energy should be adjusted at each frequency. We only need to determine estimators of $\sigma = (\bar{\sigma}_X^2, \sigma_\varepsilon^2)$ to make this adjustment. We propose implementing the estimation using the Whittle likelihood methods (see [6] or [12, 46, 47]).

¹A shrinkage estimator reduces the variance of an empirical estimator by shrinking the magnitude of the empirical estimator to achieve a smaller mean square error in estimation. The optimal shrinkage rule, i.e., L_k , is parameter dependent and can only be provided by an oracle. Usually the aim in estimation is to achieve “nearly” the same risk as using the oracle shrinkage rule, without knowledge of the true values of the parameters in the estimation problem.

For a time-domain sample $\Delta \mathbf{Y} = (\Delta Y_{t_1}, \dots, \Delta Y_{t_N})$ that is *stationary*,² if suitable additional conditions are satisfied, then maximizing the Whittle likelihood produces estimators that asymptotically are equivalent to maximum likelihood estimators obtained from maximizing the exact time-domain likelihood of the data; see [19, p. 398] or [12]. For processes that are not stationary, such conditions are in general not met, and so the Whittle likelihood function can only be used as an objective function to construct estimators, but not as a true likelihood.

The Whittle log-likelihood is defined [46, 47] by

$$\begin{aligned}
 (2.16) \quad l(\mathcal{S}) &\equiv \log \left[\prod_{k=1}^{N/2-1} \frac{1}{\mathcal{S}_{k,k}^{(Y)}} e^{-\frac{\widehat{\mathcal{F}}_{k,k}^{(Y)}}{\mathcal{S}_{k,k}^{(Y)}}} \right] \\
 &= - \sum_{k=1}^{N/2-1} \log \left(\mathcal{S}_{k,k}^{(X)} + \sigma_\varepsilon^2 |2 \sin(\pi f_k \Delta t)|^2 \right) - \sum_{k=1}^{N/2-1} \frac{\widehat{\mathcal{F}}_{k,k}^{(Y)}}{\mathcal{S}_{k,k}^{(X)} + \sigma_\varepsilon^2 |2 \sin(\pi f_k \Delta t)|^2}.
 \end{aligned}$$

In our problem, to be able to use the Whittle likelihood, we will need to show for a sample from an SDE of the form (1.1) that the variance of the discrete Fourier transform takes a specified form, and that the covariance at two different frequencies is negligible, which is a combination of (2.7) and Appendix A’s (A.4). Even with such results holding, the Whittle likelihood need not produce asymptotically efficient³ estimators, but can (at worst) be viewed as similar to using least squares estimation even if the data is not Gaussian. We shall show that optimizing (2.16) will asymptotically (in Δt^{-1}) produce unbiased estimators.

DEFINITION 2.2 (multiscale energy likelihood). *The multiscale energy log-likelihood is then defined using (2.7) as*

$$(2.17) \quad l(\boldsymbol{\sigma}) = - \sum_{k=1}^{N/2-1} \log \left(\bar{\sigma}_X^2 + \sigma_\varepsilon^2 |2 \sin(\pi f_k \Delta t)|^2 \right) - \sum_{k=1}^{N/2-1} \frac{\widehat{\mathcal{F}}_{k,k}^{(Y)}}{\bar{\sigma}_X^2 + \sigma_\varepsilon^2 |2 \sin(\pi f_k \Delta t)|^2}.$$

We stress that, strictly speaking, this may not be a (log-)likelihood, but merely a device for determining the multiscale ratio. We maximize this function in $\boldsymbol{\sigma}$ to obtain a set of estimators $\widehat{\boldsymbol{\sigma}}$.

THEOREM 2.3 (estimated multiscale ratio). *The estimated multiscale ratio is given by*

$$(2.18) \quad \widehat{L}_k = \frac{\widehat{\sigma}_X^2}{\widehat{\sigma}_X^2 + \widehat{\sigma}_\varepsilon^2 |2 \sin(\pi f_k \Delta t)|^2},$$

where $\widehat{\sigma}_X^2$ and $\widehat{\sigma}_\varepsilon^2$ maximize $\ell(\boldsymbol{\sigma})$ given in (2.17). \widehat{L}_k satisfies

$$(2.19) \quad \frac{\widehat{L}_k}{L_k} = 1 + \mathcal{O}(\Delta t^{1/4}).$$

²By stationary we here mean second order stationary, i.e., $E\{\Delta Y_{t_k}\}$ is a finite constant and $\text{cov}\{\Delta Y_{t_k}, \Delta Y_{t_{k-l}}\}$ is finite and does not depend on k .

³An asymptotically efficient estimator has variance smaller than or equal to that of all other estimators that are asymptotically unbiased.

Proof. See Appendix A. \square

Combining (2.12) with (2.18) the proposed estimator of the spectral density of $\{\Delta X_t\}$ is

$$(2.20) \quad \widehat{\mathcal{F}}_{k,k}^{(X)}(\widehat{L}_k) = \widehat{L}_k \widehat{\mathcal{F}}_{k,k}^{(Y)},$$

where \widehat{L}_k is given by (2.18).

THEOREM 2.4 (multiscale estimator of the integrated volatility). *Assume that ΔX_{t_j} satisfies the conditions of Lemma 2.1 and Theorem 2.3. The multiscale estimator of the integrated volatility defined by*

$$(2.21) \quad \widehat{\langle X, X \rangle}_T^{(m_1)} = \sum_{k=0}^{N-1} \widehat{\mathcal{F}}_{k,k}^{(X)}(\widehat{L}_k),$$

where $\widehat{\mathcal{F}}_{k,k}^{(X)}(\widehat{L}_k)$ is defined by (2.20), has a mean and variance given by

$$\begin{aligned} \mathbb{E} \left\{ \widehat{\langle X, X \rangle}_T^{(m_1)} \right\} &= \sum_{k=0}^{N-1} \mathcal{F}_{k,k}^{(X)} + \mathcal{O} \left(\log(\Delta t) \sqrt{\Delta t} \right) + \mathcal{O} \left(\Delta t^{1/4} \right) \\ &= \int_0^T \mathbb{E} \{ \sigma_t^2 \} + \mathcal{O} \left(\log(\Delta t) \sqrt{\Delta t} \right) + \mathcal{O} \left(\Delta t^{1/4} \right) \end{aligned}$$

and

$$\text{var} \left\{ \widehat{\langle X, X \rangle}_T^{(m_1)} \right\} = \sum_{k=0}^{N-1} L_k^2 \left| \mathcal{F}_{k,k}^{(Y)} \right|^2 + \mathcal{O}(\Delta t^{1/2}) = \mathcal{O}(\Delta t^{1/2}).$$

Proof. See Appendix B. \square

We also note that

$$(2.22) \quad \begin{aligned} \text{var} \left\{ \widehat{\langle X, X \rangle}_T^{(m_1)} \right\} &= \sum_{k=0}^{N-1} L_k^2 \left| \mathcal{F}_{k,k}^{(Y)} \right|^2 + \mathcal{O}(\Delta t^{1/2}) \\ &< \mathcal{O} \left(\frac{1}{\Delta t} \right) = \text{var} \left\{ \widehat{\langle X, X \rangle}_T^{(b)} \right\}, \end{aligned}$$

unless $\sigma_\varepsilon = 0$. We note that the multiscale estimator has *lower* variance than the naive method of moments estimator $\widehat{\langle X, X \rangle}_T^{(b)}$ due to the fact that $0 \leq L_k \leq 1$. We have thus removed bias *and* simultaneously decreased the variance, the latter effect usually being the main purpose of shrinkage estimators. Note that if we knew the true multiscale ratio L_k and used it rather than our estimate \widehat{L}_k (i.e., used $\widehat{\langle X, X \rangle}_T^{(m_3)}$), then we would expect an estimator from this quantity to recover the same variance as the estimator based on the noise-free observations, with a variance of $\mathcal{O}(\Delta t)$. This loss of efficiency, i.e., between $\widehat{\langle X, X \rangle}_T^{(m_1)}$ and $\widehat{\langle X, X \rangle}_T^{(m_3)}$, is inevitable as we have to estimate L_k . Finally we can also construct a Whittle estimator for the integrated volatility by starting from (2.7) and taking

$$(2.23) \quad \widehat{\langle X, X \rangle}_T^{(w)} = N \widehat{\sigma}_X^2.$$

The sampling properties of $\widehat{\langle X, X \rangle}_T^{(w)}$ are found in Appendix A, and $\widehat{\sigma}_X^2$ is asymptotically unbiased. The results in Appendix A imply that

$$(2.24) \quad \text{var} \left\{ \widehat{\langle X, X \rangle}_T^{(w)} \right\} = T \frac{\sigma_\varepsilon}{\tau_X^{1/2}} 16\tau_X^2 \sqrt{\Delta t}.$$

We see that the variance depends on the length of the time course, the inverse of the signal-to-noise ratio, the square root of the sampling period, and the fourth power of the “average standard deviation” of the X_t process. We may compare the variance of (2.22) with the variance of (2.24) to determine which estimator of $\widehat{\langle X, X \rangle}_T^{(w)}$ and $\widehat{\langle X, X \rangle}_T^{(m_1)}$ is preferable. We shall return to this question of relative performance in the examples section, but intuitively argue that $\widehat{\langle X, X \rangle}_T^{(w)}$ and $\widehat{\langle X, X \rangle}_T^{(m_1)}$ are more or less the same estimator, with the former estimator demanding more intuition to explain.

Of interest is comparing the order of magnitude of the standard deviation of $\widehat{\langle X, X \rangle}_T^{(m_1)}$ and $\widehat{\langle X, X \rangle}_T^{(w)}$ ($\mathcal{O}(\Delta t^{1/4})$) to that of other estimators. For example, the best estimator of Zhang, Mykland, and Ait-Sahalia [48, eq. 9] achieves an $\mathcal{O}(\Delta t^{1/6})$ order of magnitude, which is not asymptotically efficient compared to our $\mathcal{O}(\Delta t^{1/4})$. In our numerical examples the Zhang et al. estimator roughly achieves the same performance as ours, even if it is asymptotically worse performing. If there is no drift, and the discrete Fourier transform has a Gaussian distribution at every frequency, then we may note that the Whittle estimator $\widehat{\langle X, X \rangle}_T^{(w)}$ arises as a maximum likelihood estimator of the integrated volatility, and so, due to general results for maximum likelihood estimators, is asymptotically efficient; see, for example, Severini [41]. For a process with Gaussian Fourier coefficients the effects of the nonzero drift can be bounded, and this will not affect the asymptotic efficiency of the Whittle estimator. The final outstanding question is then for non-Gaussian Fourier coefficients, and we note here that in this case the optimal rate is not (as far as we know) known. However, it seems unlikely that non-Gaussian coefficients would produce a smaller variance than that of Gaussian coefficients, and so we hypothesize that $\mathcal{O}(\Delta t^{1/4})$ is efficient.

2.4. Time-domain representation of the method. We may write the frequency domain estimator of the spectral density of ΔX_t in the time domain to clarify some of its properties. We define

$$\widehat{s}_\tau^{(X)} = \frac{1}{N} \sum_{k=0}^{N-1} \widehat{\mathcal{F}}_{k,k}^{(X)} e^{2i\pi \frac{k\tau}{N}}, \quad \tau \in \mathbb{N},$$

which when ΔX_t is a stationary process corresponds to the estimated autocovariance sequence of ΔX_t using the method of moments estimator [9, Ch. 5]. We then have

$$(2.25) \quad \widehat{\mathcal{F}}_{k,k}^{(X)} = L_k \widehat{\mathcal{F}}_{k,k}^{(Y)}, \quad \widehat{s}_\tau^{(X)} = \sum_u \ell_{\tau-u} \widehat{s}_u^{(Y)},$$

and so the estimated autocovariance of the ΔX_t process, namely, $\widehat{s}_\tau^{(X)}$, is a smoothed version of $\widehat{s}_\tau^{(Y)}$. We can therefore view $\widehat{\mathcal{F}}_{k,k}^{(X)}$ as the Fourier transform of a smoothed version of the autocovariance sequence of ΔY_t . We let

$$(2.26) \quad L(f) = \frac{\overline{\sigma}_X^2}{\overline{\sigma}_X^2 + \sigma_\varepsilon^2 |2 \sin(\pi f \Delta t)|^2},$$

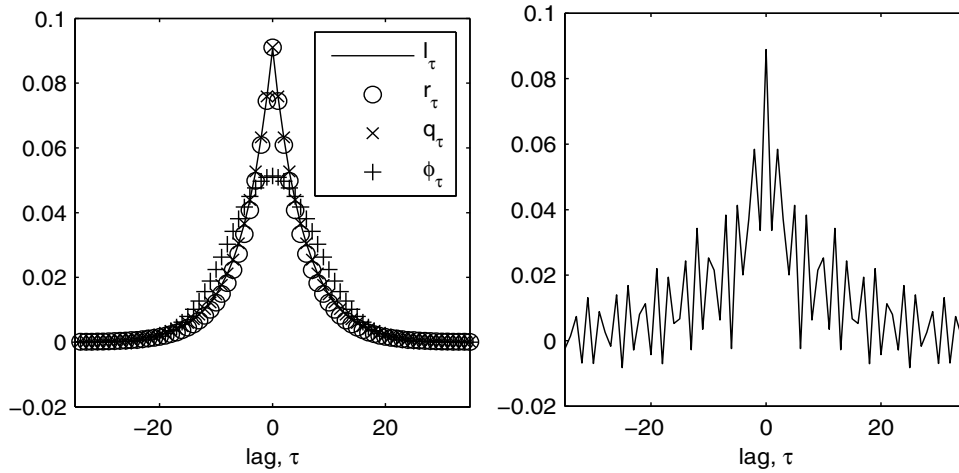


FIG. 2.1. l_τ as well as r_τ and q_τ for a chosen value of the signal-to-noise ratio (left). The approximate weighting functions perfectly mirror the exact calculation. We overlay a Gaussian kernel with the same spread for comparison. l_τ estimated for the MA(6) case (right).

be the continuous analogue of L_k . To find the smoothing kernel we are using, we need to calculate

$$\begin{aligned}
 \ell_\tau &= \frac{1}{N} \sum_{k=0}^{N-1} L_k e^{2i\pi \frac{k\tau}{N}} \\
 (2.27) \quad &= \int_{-\frac{1}{2}}^{\frac{1}{2}} \frac{\bar{\sigma}_X^2}{\bar{\sigma}_X^2 + 4\sigma_\varepsilon^2 \sin^2(\pi f)} e^{2i\pi f\tau} df + \mathcal{O}(\Delta t).
 \end{aligned}$$

Thus utilizing integration in the complex plane (see Appendix C), we obtain that

$$(2.28) \quad \ell_\tau = \begin{cases} \left(\frac{\sigma_\varepsilon^2}{\bar{\sigma}_X^2}\right)^\tau + \mathcal{O}\left(\left(\frac{\sigma_\varepsilon}{\bar{\sigma}_X}\right)^{2\tau+2}\right) & \text{if } \sigma_\varepsilon^2 < \bar{\sigma}_X^2, \\ \frac{\bar{\sigma}_X}{2\sigma_\varepsilon} \left(1 - \frac{\bar{\sigma}_X}{\sigma_\varepsilon}\right)^\tau + \mathcal{O}\left(\frac{\bar{\sigma}_X^2}{2\sigma_\varepsilon^2} \left(1 - \frac{\bar{\sigma}_X}{\sigma_\varepsilon}\right)^\tau\right) & \text{if } \sigma_\varepsilon^2 > \bar{\sigma}_X^2. \end{cases}$$

These are both decreasing sequences in τ . We write $r_\tau = \frac{\bar{\sigma}_X}{2\sigma_\varepsilon} \left(1 - \frac{\bar{\sigma}_X}{\sigma_\varepsilon}\right)^\tau$. If we can additionally assume that $L(f)$ decreases sufficiently rapidly to be near zero by $f = \frac{1}{\pi}$, then we find that

$$\ell_\tau \approx q_\tau = \frac{\bar{\sigma}_X}{2\sigma_\varepsilon} e^{-\frac{\bar{\sigma}_X}{\sigma_\varepsilon} |\tau|}.$$

In the limit of no observation noise ($\frac{\bar{\sigma}_X}{\sigma_\varepsilon} \rightarrow \infty$) this sequence then becomes a delta function centered at $\tau = 0$. Let us plot these functions, i.e., ℓ_τ , r_τ , and q_τ for a chosen case of $\bar{\sigma}_X^2/\sigma_\varepsilon^2 \approx 0.0331$ (the approximate signal-to-noise ratio used in a later example) in Figure 2.1 (left). We see that theory coincides very well with practice, with almost perfect agreement between the three functions. l_τ is, however, a strange choice of kernel, if dictated by the statistical inference problem: it has heavier tails than the common choice of the Gaussian kernel, and is extremely peaked around zero (a Gaussian kernel with the same variance has been overlaid in Figure 2.1). This is

not strange, since we are trying to filter out correlations due to non-Itô behavior, but counter to our intuition about suitable kernel functions, as the differenced Itô process exhibits very little covariance at any lag but zero, the sharp peak at zero is necessary.

We could have replaced the estimator in (2.25) by using another smoothing window estimator with some arbitrary smoothing window $\{h_\tau\}$ with discrete Fourier transform $H(\sigma_f f)$ or by smoothing as follows:

$$(2.29) \quad \overline{s}_\tau^{(X)} = \sum_u h_{t-u} \widehat{s}_u^{(Y)}, \quad \overline{\mathcal{F}}_{k,k}^{(X)} = H(\sigma_f f_k \Delta t) \widehat{\mathcal{F}}_{k,k}^{(Y)}.$$

If we choose to use a Gaussian window in frequency, then

$$H(\sigma_f f_k \Delta t) = e^{-(\pi \sigma_f f_k \Delta t)^2}, \quad f_k = \frac{k}{T},$$

and our estimator becomes

$$\widehat{\langle X, X \rangle}_T^{(m_4)} = \sum_{k=0}^{N-1} \overline{\mathcal{F}}_{k,k}^{(X)}.$$

The expectation of this new estimator is (using a periodic extension of the window)

$$(2.30) \quad \begin{aligned} \mathbb{E} \left\{ \widehat{\langle X, X \rangle}_T^{(m_4)} \right\} &= \sum_{k=-N/2}^{N/2-1} \mathbb{E} \left\{ \overline{\mathcal{F}}_{k,k}^{(X)} \right\} \\ &= \sum_{k=-N/2}^{N/2-1} H(\sigma_f f_k \Delta t) \left(\overline{\sigma}_X^2 + \sigma_\varepsilon^2 |2 \sin(\pi f_k \Delta t)|^2 \right) + R \\ &= \overline{\sigma}_X^2 N \int_{-1/2}^{1/2} H(\sigma_f f) \left\{ 1 + \frac{\sigma_\varepsilon^2}{\overline{\sigma}_X^2} |2 \sin(\pi f)|^2 \right\} df + R. \end{aligned}$$

R denotes remainder terms due to the drift, which can be bounded as before. From this expression we note that it is natural to take $\sigma_f \propto (\sigma_\varepsilon / \overline{\sigma}_X)^\alpha$ for $\alpha > 0$. To obtain an asymptotically unbiased estimator we need to choose

$$(2.31) \quad \int_{-1/2}^{1/2} H(\sigma_f f) \left\{ 1 + \frac{\sigma_\varepsilon^2}{\overline{\sigma}_X^2} |2 \sin(\pi f)|^2 \right\} df = 1.$$

For the smoothing window that exactly matches the form of the contamination, this balance is easily made as the integrand in (2.31) is then unity; if, however, we use a Gaussian window as $H(f)$, then due to the mismatch of the Gaussian and Laplacian functions one needs to decide how to weight the two terms in (2.31). As the noise becomes larger, the second term becomes more important and σ_f^2 will need adjustment to match the noise, becoming more narrow in frequency. It is impossible to exactly match the window to the second term in (2.31) at every frequency. Our Whittle estimator will be required to estimate the smoothing parameter, as this will be a function of σ_ε and $\overline{\sigma}_X$, and so nothing is really gained from using this procedure.

We remark that smoothing the autocovariance sequence $\widehat{s}_\tau^{(Y)}$ is very different from smoothing the estimated variance $(Y_{t_{j+1}} - Y_{t_j})^2$. The former of these two objects is not necessarily a very smooth sequence in time lag τ , while the latter should be a smooth sequence in t_j , the global time index. Conventional smoothing wisdom is therefore not applicable to this problem.

2.5. Correlated errors. In many applications we need to consider correlated observation noise. We assume that despite being dependent the ε_{t_j} is a stationary time series. Stationary processes can be conveniently represented in terms of aggregations of uncorrelated white noise processes, using the Wold decomposition theorem [10, p. 187]. We may therefore write the zero-mean observation ε_{t_j} as

$$(2.32) \quad \varepsilon_{t_j} = \sum_{k=0}^{\infty} \theta_{t_k} \eta_{t_j - t_k},$$

where $\theta_{t_0} \equiv 1$, $\sum_j \theta_{t_j}^2 < \infty$, and $\{\eta_{t_n}\}$ satisfies $E\{\eta_{t_n}\} = 0$ and $E\{\eta_{t_n} \eta_{t_m}\} = \sigma_\eta^2 \delta_{n,m}$, a model also used in [43]. Common practice would involve approximating the variable by a finite number of elements in the sum, and thus we truncate (2.32) to some $q \in \mathbb{Z}$. We therefore model the noise as a moving average (MA) process specified by

$$(2.33) \quad \varepsilon_{t_j} = \eta_{t_j} + \sum_{k=1}^q \theta_{t_k} \eta_{t_j - k},$$

and the covariance of the discrete Fourier transform of the differenced ε_{t_j} process takes the form

$$(2.34) \quad \mathcal{S}_{k,k}^{(\varepsilon)} = \sigma_\eta^2 \left| 1 + \sum_{k=1}^q \theta_k e^{2i\pi f k} \right|^2 |2 \sin(\pi f \Delta t)|^2.$$

This leads to defining a new multiscale ratio replacing $\sigma_\varepsilon^2 |2 \sin(\pi f \Delta t)|^2$ of (2.14) with $\sigma_\eta^2 \left| 1 + \sum_{k=1}^q \theta_k e^{2i\pi f k} \right|^2 |2 \sin(\pi f \Delta t)|^2$. We then obtain a new estimator of $\mathcal{S}_{k,k}^{(X)}$. In general the value of q is not known. To simultaneously implement model selection, we need to penalize the likelihood. We define the corrected Akaike information criterion (AICC) by [10, p. 303] (refer to (2.17) for $l(\boldsymbol{\sigma}, \boldsymbol{\theta})$ with $\sigma_\varepsilon^2 |2 \sin(\pi f \Delta t)|^2$ replaced by $\sigma_\eta^2 \left| 1 + \sum_{k=1}^q \theta_k e^{2i\pi f k} \right|^2 |2 \sin(\pi f \Delta t)|^2$)

$$(2.35) \quad \text{AICC}(\boldsymbol{\theta}) = -2l(\boldsymbol{\sigma}, \boldsymbol{\theta}) + 2 \frac{(p+2)n}{n-p-3}.$$

By minimizing this function, in $\boldsymbol{\sigma}$, $\boldsymbol{\theta}$, and q , we obtain the best fitting model for the noise accounting for overfitting by using the penalty term. It can be shown that $\text{AICC}(\boldsymbol{\theta})$ is (approximately) an unbiased estimator of the expected value of the Kullback–Leibler index. For pure autoregressive models, Hurvich and Tsai [23] have shown that the AICC criterion is asymptotically efficient, and the criterion is better behaved than several other asymptotically efficient methods of model selection. Simulation studies argue that AICC behaves well for mixed autoregressive and MA models also, and Brockwell and Davis [10] advocate its usage in more general settings.

With this method we retrieve a new multiplier that is applied in the Fourier domain, which corresponds to a new smoother in the time domain, where the smoothing window (and its smoothing width) have been automatically chosen by the data. See an example of such a smoothing window, ℓ_τ , in Figure 2.1 (right). Here L_k has been estimated from an Itô process immersed in an MA noise process. The spectrum of the MA has a trough at frequency 0.42. We therefore expect to reinforce oscillations at period $1/0.42 \approx 2.5$, which is evident from the oscillations of the estimated kernel. For more details of this process see section 3.5.

3. Monte Carlo studies. In this section we demonstrate the performance of the multiscale estimator through Monte Carlo simulations. We first describe the debiasing procedure of the estimator for the Heston model using Fourier domain graphs. We then present bias, variance, and mean square error results of various estimators (including the multiscale estimator, the naive estimator, and the first-best estimator developed in [48]) for the Heston model as well as Brownian and Ornstein Uhlenbeck processes. We then consider the case where the sample path in a Heston model is much shorter and another case where the microstructure noise is greatly reduced. Finally, we consider the case of correlated errors and show how a stationary noise process can be captured using model selection methods, and then the integrated volatility can be estimated using the adjusted multiscale estimator.

3.1. The Heston model. The Heston model is specified in [22]:

$$(3.1) \quad dX_t = (\mu - \nu_t/2) dt + \sigma_t dB_t, \quad d\nu_t = \kappa(\alpha - \nu_t) dt + \gamma \nu_t^{1/2} dW_t,$$

where $\nu_t = \sigma_t^2$, and B_t and W_t are correlated one-dimensional Brownian motions. We will use the same parameter values as those used in [48], namely, $\mu = 0.05$, $\kappa = 5$, $\alpha = 0.04$, $\gamma = 0.5$, and the correlation coefficient between the two Brownian motions B and W is $\rho = -0.5$. We set $X_0 = 0$ and $\nu_0 = 0.04$, which is the long time limit of the expectation of the process ν_t .⁴

We calculate $\widehat{\mathcal{F}}_{k,k}^{(X)}$ and $\widehat{\mathcal{F}}_{k,k}^{(\varepsilon)}$ directly from simulated data and average across realizations, producing Figure 3.1, where k is indicated by its frequency $f_k = k/N$, and only plotted for $k = 0, \dots, N/2 - 1$, as the spectrum (or $\mathcal{S}_{k,k}^{(X)}$) is symmetric. We see directly from these plots that (on average, as we show) $\widehat{\mathcal{F}}_{k,k}^{(X)}$ is constant while $\widehat{\mathcal{F}}_{k,k}^{(\varepsilon)}$ is strongly increasing with k , completely dwarfing the other spectrum at large k . Equation (2.8) implies that an equal weighting is given to all frequencies for the differenced Itô process. The noise process will in contrast have a spectrum that

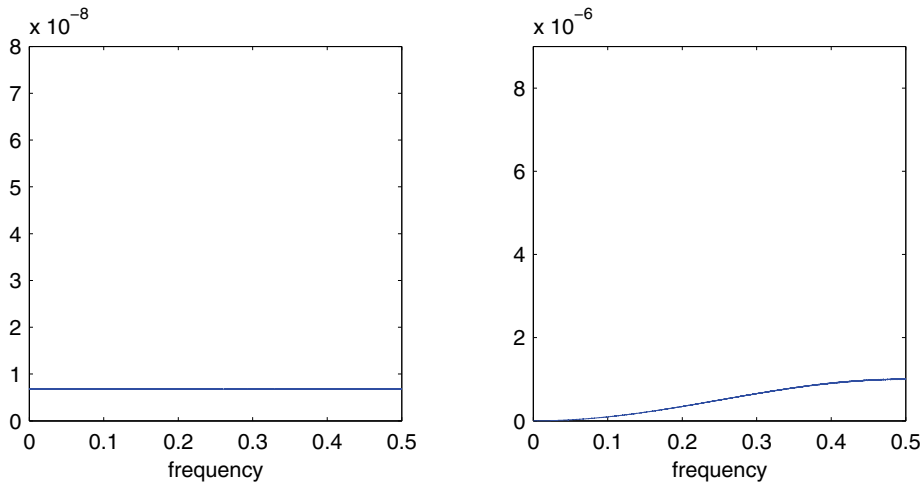


FIG. 3.1. $\widehat{\mathcal{F}}_{k,k}^{(X)}$ (left) and $\widehat{\mathcal{F}}_{k,k}^{(\varepsilon)}$ (right) averaged over 100,000 realizations. Note the different scaling of the y axis in the two figures.

⁴ $\lim_{t \rightarrow +\infty} \mathbb{E}\nu_t = \alpha$.

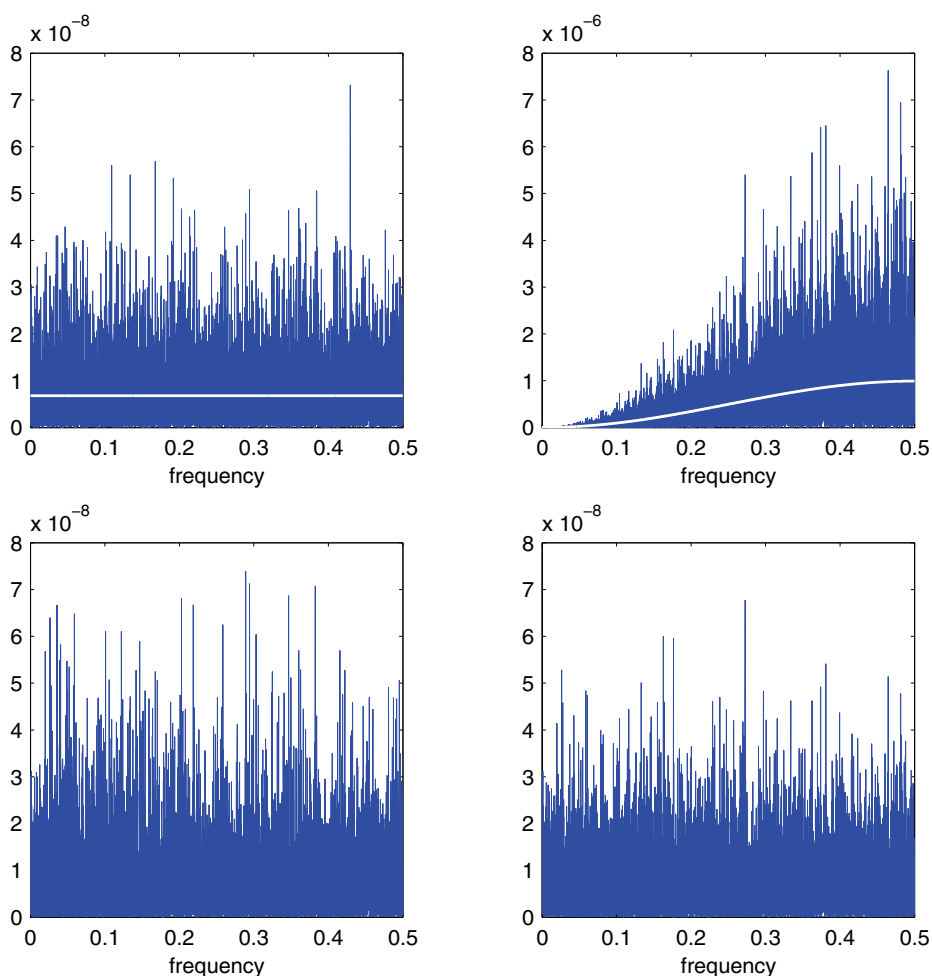


FIG. 3.2. A realization of $\widehat{\mathcal{F}}_{k,k}^{(X)}$ (top left), a realization of $\widehat{\mathcal{F}}_{k,k}^{(\varepsilon)}$ (top right) with the Whittle estimates superimposed, and realizations of two biased corrected estimators of $\mathcal{S}_{k,k}^{(X)}$, using $\widetilde{L}_k \widehat{\mathcal{F}}_{k,k}^{(X)}$ (bottom left) and $\widehat{L}_k \widehat{\mathcal{F}}_{k,k}^{(Y)}$ (bottom right). Notice the different scales in the four figures. Estimated spectra are plotted on a linear scale for ease of comparison to the effect of applying \widehat{L}_k .

is far from flat, and a suitable bias correction would shrink the estimator of $\mathcal{S}_{k,k}^{(X)}$ at higher frequencies.

We also calculate $\widehat{\mathcal{F}}_{k,k}^{(X)}$ and $\widehat{\mathcal{F}}_{k,k}^{(\varepsilon)}$ for one simulated path, displayed in Figure 3.2. Here we have used the same sample length T and noise intensity σ_ε^2 as in [48]: T is one day and $\sigma_\varepsilon^2 = 0.0005^2$. The length of the sample path, one day or 23,400s with $\Delta t = 1$ s, corresponds to one trading day, since we take one trading day to be 6.5 h long. We use, however, the units of a trading year, so $T = 1/252$. Notice the different shapes of the two periodograms. $\widehat{\mathcal{F}}_{k,k}^{(Y)}$ will not be distinguishable from $\widehat{\mathcal{F}}_{k,k}^{(\varepsilon)}$ at higher frequencies, despite the moderate to low intensity of the market microstructure noise.

If we observed the two components X_t and ε_t separately, then the multiscale ratio L_k could be estimated from $\widehat{\mathcal{F}}_{k,k}^{(X)}$ and $\widehat{\mathcal{F}}_{k,k}^{(\varepsilon)}$ using the method of moments formula. In this case, we would estimate L_k by the sample Fourier transform variances

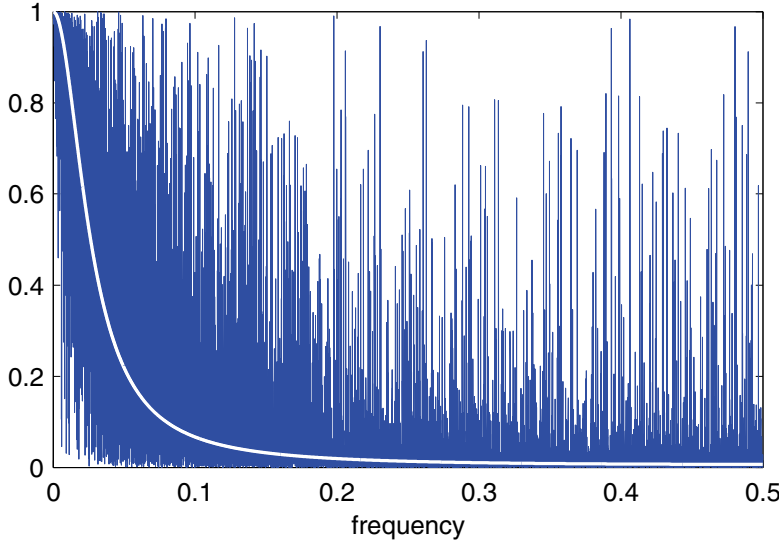


FIG. 3.3. The method of moments estimate \tilde{L}_k from a single realization, with the Whittle estimate (white line) of L_k superimposed.

$$(3.2) \quad \tilde{L}_k = \frac{\widehat{\mathcal{F}}_{k,k}^{(X)}}{\widehat{\mathcal{F}}_{k,k}^{(X)} + \widehat{\mathcal{F}}_{k,k}^{(\varepsilon)}}.$$

The corresponding estimator of the integrated volatility becomes

$$(3.3) \quad \widehat{\langle X, X \rangle}_T^{(m_2)} = \sum_{k=0}^{N-1} \tilde{L}_k \widehat{\mathcal{F}}_{k,k}^{(Y)}.$$

The estimated multiscale ratio \tilde{L}_k , for the Heston model with the specified parameters, is plotted in Figure 3.3.

The multiscale ratio cannot be estimated using the method of moments in realistic scenarios, as we only observe the aggregated process Y_t and not the two processes X_t and ε_t separately. Figure 3.2 displays the estimated multiscale ratio \tilde{L}_k applied to $\widehat{\mathcal{F}}_{k,k}^{(Y)}$ over one path realization. This plot suggests that the energy over the high frequencies has been shrunk and that $\tilde{L}_k \widehat{\mathcal{F}}_{k,k}^{(Y)}$ is a good approximation to $\widehat{\mathcal{F}}_{k,k}^{(X)}$. It therefore seems not unreasonable that the summation of this function across frequencies should make a good approximation to the integrated volatility.

The parameters $(\hat{\sigma}_X^2$ and $\hat{\sigma}_\varepsilon^2)$ are found separately for each path using the MATLAB function `fmincon` on (2.17). Figure 3.2 shows $\hat{\sigma}_X^2$ and $\hat{\sigma}_\varepsilon^2 |2 \sin(\pi f_k \Delta t)|^2$ (in white) plotted over the periodograms $\widehat{\mathcal{F}}_{k,k}^{(X)}$ and $\widehat{\mathcal{F}}_{k,k}^{(\varepsilon)}$ for one simulated path. The approximated values of $\bar{\sigma}_X^2$ and σ_ε^2 are quite similar to the averaged periodograms of Figure 3.1; in fact the accuracy of the new estimator depends on how consistently these parameters are estimated in the presence of limited information from the sampled process Y_t . Figure 3.3 shows the corresponding estimated multiscale ratio \hat{L}_k (in white) from this simulated path, as defined in (2.18). The function decays, as expected, so that it will remove the high-frequency microstructure noise in the spectrum

TABLE 3.1
Simulation study comparing the new estimator with the best estimator of [48].

	Sample bias	Sample variance	Sample RMSE	Defn.
$\widehat{\langle X, X \rangle}_T^{(b)}$	1.17×10^{-2}	1.80×10^{-8}	1.17×10^{-2}	Eqn. (2.1)
$\widehat{\langle X, X \rangle}_T^{(s_1)}$	6.44×10^{-7}	2.76×10^{-10}	1.66×10^{-5}	From [48]
$\widehat{\langle X, X \rangle}_T^{(m_1)}$	2.90×10^{-7}	2.59×10^{-10}	1.61×10^{-5}	Eqn. (2.21)
$\widehat{\langle X, X \rangle}_T^{(w)}$	2.63×10^{-7}	2.59×10^{-10}	1.61×10^{-5}	Eqn. (2.23)
$\widehat{\langle X, X \rangle}_T^{(m_2)}$	1.39×10^{-8}	2.07×10^{-10}	1.44×10^{-5}	Eqn. (3.3)
$\widehat{\langle X, X \rangle}_T^{(u)}$	1.20×10^{-8}	2.06×10^{-10}	1.44×10^{-5}	Eqn. (2.2)

of Y_t ; the ratio is also a good approximation of \tilde{L}_k . Figure 3.2 shows $\widehat{L}_k \widehat{\mathcal{F}}_{k,k}^{(Y)}$, which is again similar to $\widehat{\mathcal{F}}_{k,k}^{(X)}$. It would appear that the new estimator has successfully removed the microstructure effect from each frequency. It is worth noting that the ratios L_k and \tilde{L}_k quantify the effect of the multiscale structure of the process. If σ_ε^2 is zero (i.e., there is no microstructure noise), then no correction will be made to the spectral density function (the ratio will equal 1 at all frequencies). So in the case of zero microstructure noise, the estimate would recover $\widehat{\mathcal{F}}_{k,k}^{(X)}$, and from (2.10) the estimate of the integrated volatility would simply be the realized integrated volatility of the observable process.

We investigate the performance of the multiscale estimator using Monte Carlo simulations. In this study 50,000 simulated paths are generated. Table 3.1 displays the results of our simulation, where biases, variances, and errors are calculated using a Riemann sum approximation of the integral:

$$(3.4) \quad \frac{T}{N} \sum_{i=1}^N \sigma_i^2 = \int_0^T \sigma_t^2 dt.$$

The two estimators $\widehat{\langle X, X \rangle}_T^{(u)}$ and $\widehat{\langle X, X \rangle}_T^{(m_2)}$ (see (2.2) and (3.3), respectively) are both included for comparison, even though these require use of the unobservable X_t process. The performance of the first-best estimator in [48] (denoted by $\widehat{\langle X, X \rangle}_T^{(s_1)}$) is also included as a well-performing and tested estimator using only the Y_t process, as is the naive estimator of the realized volatility on Y_t at the highest frequency, $\widehat{\langle X, X \rangle}_T^{(b)}$, given in (2.1) (the fifth-best estimator in [48]). We also include the performance of $\widehat{\langle X, X \rangle}_T^{(w)}$, defined in (2.23). Note that once the bias is removed the mean square error is now mainly due to the variance of the estimator. Given the magnitude of the bias, achieving this debiasing is vital, as discussed by Zhang, Mykland, and Ait-Sahalia [48, p. 1397].

Table 3.1 shows that the new estimator, $\widehat{\langle X, X \rangle}_T^{(m_1)}$, is competitive with the first-best approach in [48], $\widehat{\langle X, X \rangle}_T^{(s_1)}$, as an estimator of the integrated volatility for the Heston model with the stated parameters. For this simulation the new method performed marginally better. The similar performance of the two estimators is quite remarkable, given their different approaches; both estimators involve a bias correction, though [48] performs this globally by weighting different sampling frequencies, while we correct locally at each frequency. The realized integrated volatility of Y_t at the

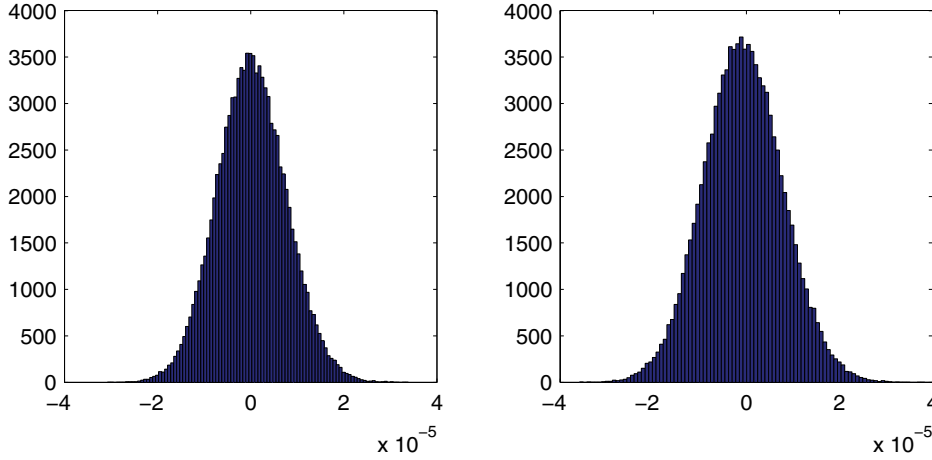


FIG. 3.4. The histograms of the observed bias of the proposed estimator (left), and the first-best estimator (right), over 100,000 sample paths.

highest frequency, $\widehat{\langle X, X \rangle}_T^{(b)}$, produces disastrous results, as expected.

We also note that $\widehat{\langle X, X \rangle}_T^{(m_1)}$ performs more or less identically to $\widehat{\langle X, X \rangle}_T^{(w)}$. These two estimators can almost be used interchangeably due to the invariance property of a maximum likelihood estimator. This observation is born out by our simulation studies, and we henceforth report results only for $\widehat{\langle X, X \rangle}_T^{(m_1)}$. Note that the variance of $\widehat{\langle X, X \rangle}_T^{(w)}$ can be found from (2.24). To compare theory with simulations we note that the average estimated standard deviation is 1.6093×10^{-5} , while the expression for the variance to leading order gives an expression for the standard deviation of $[\text{var}\{\widehat{\langle X, X \rangle}_T^{(w)}\}]^{1/2} = 1.0246 \times 10^{-5}$, using the parameter values of $\bar{\sigma}_X^2 \approx 6.8 \times 10^{-9}$ and $\sigma_\varepsilon^2 \approx 2.5 \times 10^{-7}$. Recall that $T = 1/252$ and $N = 23,400$.

A histogram of the observed bias of the new estimator is plotted in Figure 3.4 along with a histogram of the observed bias of the first-best estimator in [48]. The observed bias of our estimator follows a Gaussian distribution centered at zero, suggesting that this estimator is unbiased, as our results claim. Comparing our estimator to the first-best estimator, it can be seen that the new estimator has similar magnitudes of error also (hence the similar root mean square error (RMSE)).

The new estimator requires calculation of $\hat{\sigma}_X^2$ and $\hat{\sigma}_\varepsilon^2$, which will vary over each process due to the limited information given from the Y_t process. The stability of this estimation is of great importance if the estimator is to perform well. Figure 3.5 shows the distribution of the parameters $\hat{\sigma}_X^2$ and $\hat{\sigma}_\varepsilon^2$ over the simulated paths. The parameter estimation is quite consistent, with all values estimated within a narrow range. Figure 3.1 suggests that these estimates are roughly unbiased, as $\bar{\sigma}_X^2 \approx 6.8 \times 10^{-9}$ and $\sigma_\varepsilon^2 \approx 2.5 \times 10^{-7}$ (as $\sigma_\varepsilon^2 |2 \sin(\pi f_k)|^2 \approx 1 \times 10^{-6}$, at $f_k = 0.5$).

3.2. Brownian process and Ornstein Uhlenbeck process. We repeated our simulations for a Brownian process given by

$$(3.5) \quad dX_t = \sqrt{2\sigma_t^2} dB_t,$$

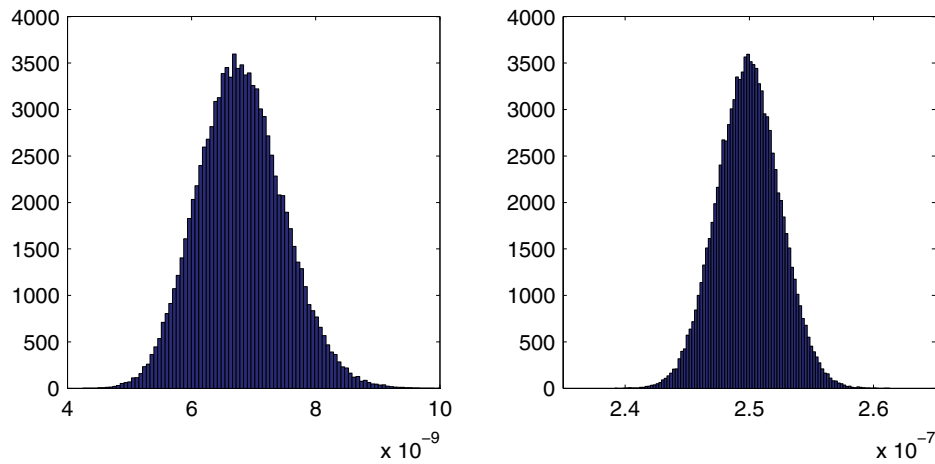


FIG. 3.5. The histograms of the estimated $\bar{\sigma}_X^2$ (left) and σ_ε^2 (right).

TABLE 3.2
Simulation study for the Brownian process.

	Sample bias	Sample variance	Sample RMSE	Defn.
$\widehat{\langle X, X \rangle}_T^{(b)}$	1.17×10^{-2}	1.77×10^{-8}	1.17×10^{-2}	Eqn. (2.1)
$\widehat{\langle X, X \rangle}_T^{(s_1)}$	6.52×10^{-7}	2.68×10^{-11}	5.22×10^{-6}	From [48]
$\widehat{\langle X, X \rangle}_T^{(m_1)}$	3.02×10^{-7}	1.98×10^{-11}	4.46×10^{-6}	Eqn. (2.21)
$\widehat{\langle X, X \rangle}_T^{(m_2)}$	1.96×10^{-9}	6.93×10^{-13}	8.32×10^{-7}	Eqn. (3.3)
$\widehat{\langle X, X \rangle}_T^{(u)}$	3.79×10^{-9}	5.44×10^{-13}	7.38×10^{-7}	Eqn. (2.2)

where $\sigma_t^2 = 0.01$. We otherwise keep the same simulation setup as before with 50,000 simulated paths of length 23,400. The results are displayed in Table 3.2. The new estimator, $\widehat{\langle X, X \rangle}_T^{(m_1)}$, again delivers a marked improvement on the naive estimator, $\widehat{\langle X, X \rangle}_T^{(b)}$, and performs marginally better than the first-best estimator in [48], $\widehat{\langle X, X \rangle}_T^{(s_1)}$.

We also performed a Monte Carlo simulation for the Ornstein Uhlenbeck process given by

$$(3.6) \quad dX_t = -X_t dt + \sqrt{2\sigma_t} dB_t,$$

where also $\sigma_t^2 = 0.01$. Again we retain the same simulation setup and the results are displayed in Table 3.3. The results are almost identical to that of the Brownian process, with the new estimator again outperforming other time-domain estimators.

3.3. Comparing estimators over shorter sample lengths. This section compares estimators for a shorter sample length, which will reduce the benefit of subsampling due to the variance issues of small-length data but will also affect the variance of the multiscale ratio (cf. Theorem 2.3).

TABLE 3.3
Simulation study for the Ornstein Uhlenbeck process.

	Sample bias	Sample variance	Sample RMSE	Defn.
$\widehat{\langle X, X \rangle}_T^{(b)}$	1.17×10^{-2}	1.78×10^{-8}	1.17×10^{-2}	Eqn. (2.1)
$\widehat{\langle X, X \rangle}_T^{(s_1)}$	6.69×10^{-7}	2.66×10^{-11}	5.20×10^{-6}	From [48]
$\widehat{\langle X, X \rangle}_T^{(m_1)}$	2.95×10^{-7}	1.97×10^{-11}	4.44×10^{-6}	Eqn. (2.21)
$\widehat{\langle X, X \rangle}_T^{(m_2)}$	5.09×10^{-9}	6.76×10^{-13}	8.22×10^{-7}	Eqn. (3.3)
$\widehat{\langle X, X \rangle}_T^{(u)}$	6.29×10^{-9}	5.33×10^{-13}	7.30×10^{-7}	Eqn. (2.2)

The simulation setup is exactly the same as before (using the Heston model with the same parameters) except that T , the simulation length, is reduced by a factor of 10 to 0.1 days or 2340s. Before the results of the simulation are reported, it is of interest to see whether the frequency domain methods developed still model each process accurately. Figure 3.6 shows the calculated $\widehat{\sigma}_X^2$ and $\widehat{\sigma}_\varepsilon^2 |\sin(\pi \Delta t f_k)|^2$ (in white) together with the periodograms $\widehat{\mathcal{F}}_{k,k}^{(X)}$ and $\widehat{\mathcal{F}}_{k,k}^{(\varepsilon)}$ for one simulated path. The estimator still approximates the energy structure of the processes accurately. Figure 3.6 also shows the corresponding estimate of the multiscale ratio \widehat{L}_k (in white) from this simulated path (together with \widetilde{L}_k) and the corresponding plot of $\widehat{L}_k \widehat{\mathcal{F}}_{k,k}^{(Y)}$. The new estimator has removed the microstructure noise effect and has formed a good approximation of $\widehat{\mathcal{F}}_{k,k}^{(X)}$. The approximation of the periodograms is still accurate despite the reduction of available data.

Table 3.4 displays the accuracy of the estimators over the 50,000 simulated paths. The first-best estimator in [48], $\widehat{\langle X, X \rangle}_T^{(s_1)}$, and the new estimator, $\widehat{\langle X, X \rangle}_T^{(m_1)}$, are once again comparable in performance, and both estimates are close to the best attainable RMSE, given by $\widehat{\langle X, X \rangle}_T^{(u)}$, the realized integrated volatility on X_t .

3.4. Comparing estimators with a low-noise process. This section compares estimators for smaller levels of microstructure noise. Reducing the microstructure noise will reduce the need to subsample. The first-best estimator in [48], $\widehat{\langle X, X \rangle}_T^{(s_1)}$, will have a higher sampling frequency, and the new estimator will reduce its estimate of $\widehat{\sigma}_\varepsilon^2$ accordingly. For very small levels of noise, however, the first-best estimator will become zero, as the optimal number of samples becomes n (the highest available). This possibility is now examined, using the Heston model as before, with all parameters unchanged, except that the noise is reduced by a factor of 10, i.e., $\sigma_\varepsilon^2 = 0.00005^2$. Note that the path length is kept at its original length of one day.

Figure 3.7 shows the estimates of $\widehat{\sigma}_X^2$ and $\widehat{\sigma}_\varepsilon^2 |2 \sin(\pi \Delta t f_k)|^2$ (in white) along with the periodograms $\widehat{\mathcal{F}}_{k,k}^{(Y)}$ and $\widehat{\mathcal{F}}_{k,k}^{(\varepsilon)}$ for one simulated path along with the corresponding estimate of the multiscale ratio \widehat{L}_k (in white) (plotted over the approximated \widetilde{L}_k) and the corresponding plot of $\widehat{L}_k \widehat{\mathcal{F}}_{k,k}^{(Y)}$. The estimation method works well again; notice how the magnitude of the microstructure noise has been greatly reduced (the scale is now of order 10^{-8} rather than 10^{-6}), causing the multiscale ratio L_k to be more tempered across the high frequencies than it was before, due to the smaller microstructure noise. Nonetheless, the new estimator has still detected the smaller levels of noise in the data.

Table 3.5 reports on the results of 50,000 simulations performed as before. The

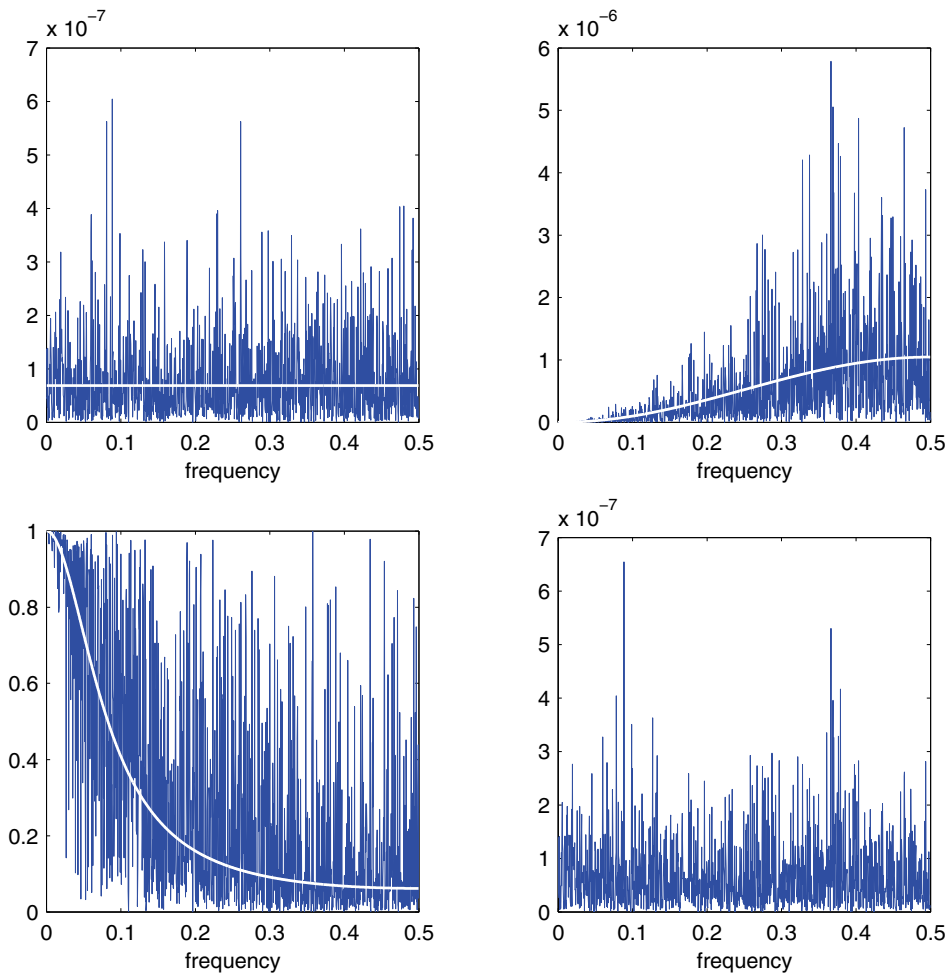


FIG. 3.6. A realization of $\widehat{\mathcal{S}}_{k,k}^{(X)}$ (top left), a realization of $\widehat{\mathcal{S}}_{k,k}^{(\epsilon)}$ (top right) with the Whittle estimates superimposed, the estimate of L_k (bottom left) with the Whittle estimate of L_k superimposed, and the biased corrected estimator of $\mathcal{S}_{k,k}^{(X)}$ using $\widehat{L}_k \widehat{\mathcal{S}}_{k,k}^{(Y)}$ (bottom right). Notice the different scales in the four figures.

TABLE 3.4
Simulation study for shorter sampler length.

	Sample bias	Sample variance	Sample RMSE	Defn.
$\langle \widehat{X}, \widehat{X} \rangle_T^{(b)}$	1.17×10^{-3}	2.29×10^{-9}	1.17×10^{-3}	Eqn. (2.1)
$\langle \widehat{X}, \widehat{X} \rangle_T^{(s_1)}$	1.00×10^{-6}	4.51×10^{-10}	2.13×10^{-5}	From [48]
$\langle \widehat{X}, \widehat{X} \rangle_T^{(m_1)}$	1.84×10^{-7}	4.23×10^{-10}	2.06×10^{-5}	Eqn. (2.21)
$\langle \widehat{X}, \widehat{X} \rangle_T^{(m_2)}$	4.80×10^{-8}	2.42×10^{-10}	1.55×10^{-5}	Eqn. (3.3)
$\langle \widehat{X}, \widehat{X} \rangle_T^{(u)}$	5.27×10^{-8}	2.28×10^{-10}	1.51×10^{-5}	Eqn. (2.2)

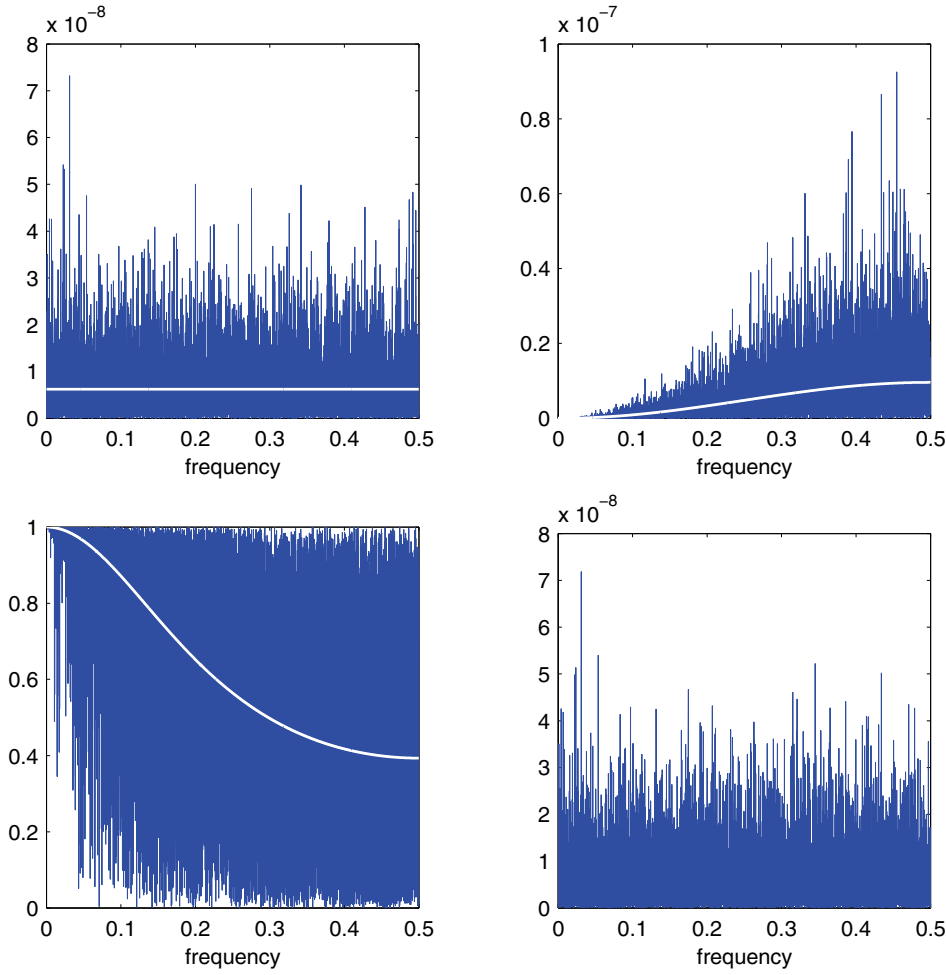


FIG. 3.7. A realization of $\widehat{\mathcal{S}}_{k,k}^{(X)}$ (top left), a realization of $\widehat{\mathcal{S}}_{k,k}^{(\epsilon)}$ (top right) with the Whittle estimates superimposed, the estimate of L_k (bottom left) with the Whittle estimate of L_k superimposed, and the biased corrected estimator of $\mathcal{S}_{k,k}^{(X)}$ using $\widehat{L}_k \widehat{\mathcal{S}}_{k,k}^{(Y)}$ (bottom right). Notice the different scales in the four figures.

TABLE 3.5
Simulation study for lower market microstructure noise.

	Sample bias	Sample variance	Sample RMSE	Defn.
$\widehat{\langle X, X \rangle}_T^{(b)}$	1.17×10^{-4}	2.11×10^{-10}	1.18×10^{-4}	Eqn. (2.1)
$\widehat{\langle X, X \rangle}_T^{(s_2)}$	3.53×10^{-6}	1.00×10^{-9}	3.19×10^{-5}	From [48]
$\widehat{\langle X, X \rangle}_T^{(m_1)}$	7.63×10^{-9}	2.12×10^{-10}	1.46×10^{-5}	Eqn. (2.21)
$\widehat{\langle X, X \rangle}_T^{(m_2)}$	7.91×10^{-9}	2.06×10^{-10}	1.44×10^{-5}	Eqn. (3.3)
$\widehat{\langle X, X \rangle}_T^{(u)}$	9.83×10^{-9}	2.05×10^{-10}	1.43×10^{-5}	Eqn. (2.2)

first-best estimator of [48], $\widehat{\langle X, X \rangle}_T^{(s_1)}$, categorically failed for this model. This is due to the fact that the optimal number of samples was always equal to n , the total number of samples available. Therefore, the first-best estimator was always zero. The second-best estimator in [48], denoted by $\widehat{\langle X, X \rangle}_T^{(s_2)}$, was reasonably effective. This is simply an estimator that averages estimates calculated from subsampled paths at different starting points and is therefore asymptotically biased. The new estimator, $\widehat{\langle X, X \rangle}_T^{(m_1)}$, was remarkably robust, with RMSE very close to the RMSE of estimators based on the X_t process. The difference in performance between estimators using Y_t and estimators using X_t is expected to become smaller with less microstructure noise, and this can be seen by the similar order RMSE errors between all estimators. Nevertheless, the new estimator was much closer in performance to the realized integrated volatility on X_t than it was to any other estimator on Y_t , a result that demonstrates the precision and robustness of this new estimator of integrated volatility.

3.5. Correlated noise. In this section we consider microstructure noise that is correlated. If this process is stationary, the noise process can be modeled as an MA process (as described in section 2.5), and the corresponding parameters can be estimated by maximizing the multiscale Whittle likelihood using (2.14) and (2.34). Figure 3.8 shows the multiscale estimator applied to the Heston model (with the same parameters as before) with a microstructure noise that follows an MA(6) process (parameters given in the caption). The Whittle estimates (in white) form a good approximation of $\widehat{\mathcal{F}}_{k,k}^{(X)}$ and $\widehat{\mathcal{F}}_{k,k}^{(\varepsilon)}$ despite the more complicated nuisance structure. The corresponding estimate of the multiscale ratio \widehat{L}_k (in white) therefore removes energy from the correct frequencies, and the corresponding plot of $\widehat{L}_k \widehat{\mathcal{F}}_{k,k}^{(Y)}$ is a good approximation of $\widehat{\mathcal{F}}_{k,k}^{(X)}$. This is the same noise process and Itô process for which we calculated the optimal smoothing window in section 2.5, and the trough in the noise at about $f = 0.42$ corresponds to the oscillations in the kernel plotted in Figure 2.1.

If the length of the MA(p) process is unknown, then p can be determined using (2.35). In Table 3.6 we show an example with $p = 4$ with parameters $\theta_1 = 0.8$, $\theta_2 = -0.6$, $\theta_3 = 0.1$, $\theta_4 = 0.4$. Clearly $p = 4$ is identified as the best fitting model, yielding nearly perfect estimates of the noise parameters. The estimator is therefore robust at removing the effect of microstructure noise when this process is correlated (and stationary), even if the length of the MA(p) process is not explicitly known.

We also tested our estimator using Monte Carlo simulations in [43] for a variety of MA(1) processes, and the results showed a significant reduction in error compared with not only the naive estimator but also the estimators based on a white noise assumption. Furthermore, the adjusted multiscale estimator performed almost identically to our multiscale estimator when we set $\theta_1 = 0$ and recovered a white noise process, meaning the loss in precision from searching for a parameter unnecessarily was negligible (as to be expected for $q \ll N$). Notice also that in Table 3.6 there appears to be little loss in precision from estimating more parameters in the MA(4) process than is required as θ_p for $p > 4$ is always estimated to be very close to zero. This further demonstrates the robustness and precision of our estimation technique.

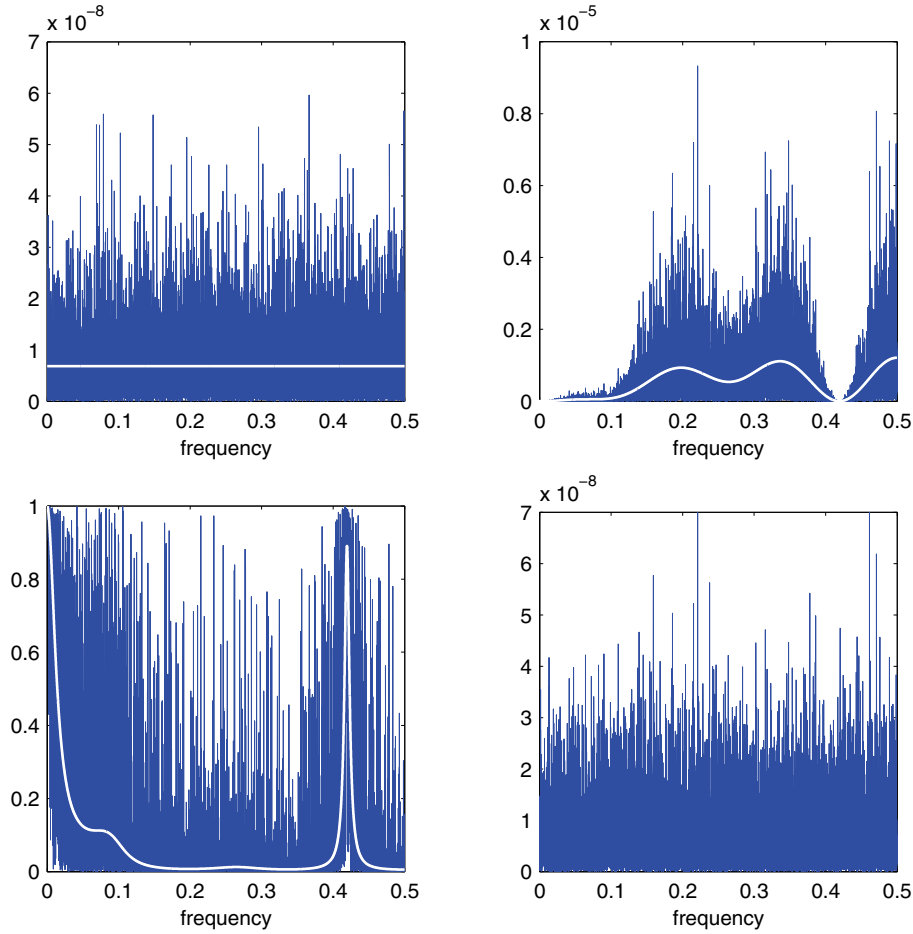


FIG. 3.8. A realization of $\widehat{\mathcal{S}}_{k,k}^{(X)}$ (top left), a realization of $\widehat{\mathcal{S}}_{k,k}^{(\epsilon)}$ (top right) with the Whittle estimates superimposed, the estimate of L_k (bottom left) with the Whittle estimate of L_k superimposed, and the biased corrected estimator of $\mathcal{S}_{k,k}^{(X)}$ using $\widehat{L}_k \widehat{\mathcal{S}}_{k,k}^{(Y)}$ (bottom right). In this example we use an MA(6) with $\theta_1 = 0.5$, $\theta_2 = -0.1$, $\theta_3 = -0.1$, $\theta_4 = 0.2$, $\theta_5 = 0$, and $\theta_6 = 0.4$. Notice the different scales in the four figures.

TABLE 3.6

Values of θ found by modeling the noise process as an MA(p) process for $p = 1, \dots, 8$. Model selection methods (AICC) are used to select which process to model the noise by; in this case the AICC is minimized by selecting an MA(4) with the given parameters. The true noise is indeed an MA(4) process (with parameters $\theta_1 = 0.8$, $\theta_2 = -0.6$, $\theta_3 = 0.1$, $\theta_4 = 0.4$).

MA(p)	θ_1	θ_2	θ_3	θ_4	θ_5	θ_6	θ_7	θ_8	AICC
$p = 1$	0.935								-3.208490×10^5
$p = 2$	0.624	-0.445							-3.239947×10^5
$p = 3$	0.658	-0.459	-0.046						-3.240000×10^5
$p = 4$	0.806	-0.603	-0.101	0.410					-3.262427×10^5
$p = 5$	0.813	-0.606	-0.101	0.411	-0.008				-3.262416×10^5
$p = 6$	0.815	-0.604	-0.097	0.420	-0.003	0.000			-3.262409×10^5
$p = 7$	0.807	-0.613	-0.114	0.413	0.002	-0.002	-0.005		-3.262402×10^5
$p = 8$	0.817	-0.614	-0.128	0.427	0.005	0.011	-0.009	-0.017	-3.262384×10^5

4. Conclusions. The problem of estimating the integrated stochastic volatility of an Itô process from noisy observations was studied in this paper. Unlike most previous works on this problem (see, e.g., [48, 35]), the method for estimating the integrated volatility developed in this paper is based on the frequency domain representation of both the Itô process and the noisy observations. The integrated volatility can be represented as a summation of variation in the process of interest over all frequencies (or scales). In our estimator we adjust the raw sample variance at each frequency. Such an estimator is truly multiscale, as it corrects the estimated energy directly at every scale. In other words, the estimator is debiased *locally at each frequency*, rather than globally.

To estimate the degree of scale separation in the data, we used the Whittle likelihood and quantified the noise contribution by the multiscale ratio. Various properties of the multiscale estimator were determined; see Theorems 2.3 and 2.4. As was illustrated by the set of examples, our estimator performs extremely well on data simulated from the Heston model, and it is competitive with the methods proposed by [48], under varying signal-to-noise and sampling scenarios. The proposed estimator is truly multiscale in nature and adapts automatically to the degree of noise contamination of the data, a clear strength. It is also easily implemented and computationally efficient.

The new estimator for the integrated stochastic volatility can be written as

$$\widehat{\langle X, X \rangle} = \sum_u \ell_{-u} \sum_k (X_{t_{k-u}} - X_{t_{k-u-1}}) (X_{t_k} - X_{t_{k-1}}),$$

where the kernel ℓ_u is given by (2.27). We can compare this estimator with kernel estimators; see [14]. There the estimated increment square ΔX_t^2 is locally smoothed to estimate the diffusion coefficient using a kernel function, $K(\cdot)$. Contrary to this approach we estimate the integrated volatility by smoothing the estimated autocovariance of ΔX_{t_j} . In particular, we use a data-dependent choice of smoothing window. We show that, from a minimum bias perspective, using a Laplace window to smooth is optimal. This data-dependent choice of smoothing window becomes more interesting after relaxing the assumptions on the noise process and treating correlated observation error.

Inference procedures implemented in the frequency domain are still very underdeveloped for problems with a multiscale structure. The modern data deluge has caused an excess of high-frequency observations in a number of application areas, for example, finance and molecular dynamics. More flexible models could also be used for the high-frequency nuisance structure. In this paper we have introduced a new frequency domain-based estimator and applied it to a relatively simple problem, namely, the estimation of the integrated stochastic volatility, for data contaminated by high-frequency noise. There are many extensions and potential applications of the new estimator. Here we list a few which seem interesting to us and which are currently under investigation.

- Parameter estimation for noisily observed SDEs which are driven by more general noise processes, for example, Lévy processes.
- Application of the new estimator to the problem of statistical inference for fast/slow systems of SDEs, of the type studied in [35, 34].
- The combined effects of high-frequency and multiscale structure in the data. A first step in this direction was taken in [11].

Appendix A. Proof of Theorem 2.3. Let the true value of σ be denoted σ^* . We differentiate the multiscale energy likelihood function (2.17) with respect to σ to

obtain

$$\begin{aligned}\dot{\ell}_X(\boldsymbol{\sigma}) &= \frac{\partial \ell(\boldsymbol{\sigma})}{\partial \bar{\sigma}_X^2} = - \sum_{k=1}^{N/2-1} \frac{1}{\bar{\sigma}_X^2 + \sigma_\varepsilon^2 |2 \sin(\pi f_k \Delta t)|^2} + \sum_{k=1}^{N/2-1} \frac{\widehat{\mathcal{F}}_{k,k}^{(Y)}}{\left(\bar{\sigma}_X^2 + \sigma_\varepsilon^2 |2 \sin(\pi f_k \Delta t)|^2\right)^2}, \\ \dot{\ell}_\varepsilon(\boldsymbol{\sigma}) &= \frac{\partial \ell(\boldsymbol{\sigma})}{\partial \sigma_\varepsilon^2} = - \sum_{k=1}^{N/2-1} \frac{|2 \sin(\pi f_k \Delta t)|^2}{\bar{\sigma}_X^2 + \sigma_\varepsilon^2 |2 \sin(\pi f_k \Delta t)|^2} + \sum_{k=1}^{N/2-1} \frac{|2 \sin(\pi f_k \Delta t)|^2 \widehat{\mathcal{F}}_{k,k}^{(Y)}}{\left(\bar{\sigma}_X^2 + \sigma_\varepsilon^2 |2 \sin(\pi f_k \Delta t)|^2\right)^2}.\end{aligned}$$

To remove implicit Δt dependence we let $\tau_X = \bar{\sigma}_X^2 / \Delta t$ and denote derivatives with respect to τ_X by the subscript τ . Then $\dot{\ell}_\tau(\widehat{\boldsymbol{\sigma}}) = \Delta t \dot{\ell}_X(\widehat{\boldsymbol{\sigma}})$, and so on. We calculate the expectation and variance of the score functions evaluated at $\boldsymbol{\sigma}^*$ and find that the bias of $\widehat{\tau}_X$ is order $\mathcal{O}(\Delta t^{1/2} \log(\Delta t))$ and the bias of $\widehat{\sigma}_\varepsilon^2$ is order $\mathcal{O}(\Delta t^2 \log(\Delta t))$. These contributions become negligible, and are of lesser importance compared to the variance.

To show large sample properties, we Taylor expand the multiscale likelihood with $\widehat{\boldsymbol{\sigma}}$ corresponding to the estimated maximum likelihood, and $\boldsymbol{\sigma}'$ is lying between $\widehat{\boldsymbol{\sigma}}$ and $\boldsymbol{\sigma}^*$. Then

$$\begin{aligned}\dot{\ell}_\tau(\widehat{\boldsymbol{\sigma}}) &= \dot{\ell}_\tau(\boldsymbol{\sigma}^*) + \ddot{\ell}_{\tau\tau}(\boldsymbol{\sigma}') [\widehat{\sigma}_X^2 - \sigma_X^{*2}] / \Delta t + \ddot{\ell}_{\tau\varepsilon}(\boldsymbol{\sigma}') [\widehat{\sigma}_\varepsilon^2 - \sigma_\varepsilon^{*2}], \\ \dot{\ell}_\varepsilon(\widehat{\boldsymbol{\sigma}}) &= \dot{\ell}_\varepsilon(\boldsymbol{\sigma}^*) + \ddot{\ell}_{\varepsilon\tau}(\boldsymbol{\sigma}') [\widehat{\sigma}_X^2 - \sigma_X^{*2}] / \Delta t + \ddot{\ell}_{\varepsilon\varepsilon}(\boldsymbol{\sigma}') [\widehat{\sigma}_\varepsilon^2 - \sigma_\varepsilon^{*2}].\end{aligned}$$

We note with the observed Fisher information

$$\mathbf{F} = \begin{bmatrix} \ddot{\ell}_{\tau\tau}(\boldsymbol{\sigma}') & \ddot{\ell}_{\tau\varepsilon}(\boldsymbol{\sigma}') \\ \ddot{\ell}_{\varepsilon\tau}(\boldsymbol{\sigma}') & \ddot{\ell}_{\varepsilon\varepsilon}(\boldsymbol{\sigma}') \end{bmatrix}$$

that

$$(A.1) \quad \begin{pmatrix} (\widehat{\sigma}_X^2 - \sigma_X^{*2}) / \Delta t \\ \widehat{\sigma}_\varepsilon^2 - \sigma_\varepsilon^{*2} \end{pmatrix} = \mathbf{F}^{-1} \begin{pmatrix} \dot{\ell}_\tau(\widehat{\boldsymbol{\sigma}}) - \dot{\ell}_\tau(\boldsymbol{\sigma}^*) \\ \dot{\ell}_\varepsilon(\widehat{\boldsymbol{\sigma}}) - \dot{\ell}_\varepsilon(\boldsymbol{\sigma}^*) \end{pmatrix}.$$

We henceforth ignore the term $J_k^{(\mu)} = J_k^{(X)} - \tilde{J}_k^{(X)}$, as this will not contribute to leading order, and write $J_k^{(X)}$, where formally we would write $\tilde{J}_k^{(X)}$ or $J_k^{(X)}$. We can observe the suitability of this directly from (2.17) and use bounds for $J_k^{(\mu)}$, where we could formally apply these to get bounds on each derivative of $l(\boldsymbol{\sigma})$ (note that we cannot differentiate bounds). To avoid needless technicalities, the details of this approach will not be reported. To leading order

$$\begin{aligned}\text{var} \left(\dot{\ell}_\tau(\boldsymbol{\sigma}) \right) &= \sum_{l=1}^{N/2-1} \sum_{k=1}^{N/2-1} \frac{\Delta t^2 \text{cov} \left(\widehat{\mathcal{F}}_{k,k}^{(Y)}, \widehat{\mathcal{F}}_{l,l}^{(Y)} \right)}{\left(\bar{\sigma}_X^2 + \sigma_\varepsilon^2 |2 \sin(\pi f_k \Delta t)|^2 \right)^2 \left(\bar{\sigma}_X^2 + \sigma_\varepsilon^2 |2 \sin(\pi f_l \Delta t)|^2 \right)^2}, \\ \text{var} \left(\dot{\ell}_\varepsilon(\boldsymbol{\sigma}) \right) &= \sum_{l=1}^{N/2-1} \sum_{k=1}^{N/2-1} \frac{|2 \sin(\pi f_k \Delta t)|^2 |2 \sin(\pi f_l \Delta t)|^2 \text{cov} \left(\widehat{\mathcal{F}}_{k,k}^{(Y)}, \widehat{\mathcal{F}}_{l,l}^{(Y)} \right)}{\left(\bar{\sigma}_X^2 + \sigma_\varepsilon^2 |2 \sin(\pi f_k \Delta t)|^2 \right)^2 \left(\bar{\sigma}_X^2 + \sigma_\varepsilon^2 |2 \sin(\pi f_l \Delta t)|^2 \right)^2}, \\ \text{cov} \left(\dot{\ell}_\tau(\boldsymbol{\sigma}), \dot{\ell}_\varepsilon(\boldsymbol{\sigma}) \right) &= \sum_{l=1}^{N/2-1} \sum_{k=1}^{N/2-1} \frac{\Delta t |2 \sin(\pi f_l \Delta t)|^2 \text{cov} \left(\widehat{\mathcal{F}}_{k,k}^{(Y)}, \widehat{\mathcal{F}}_{l,l}^{(Y)} \right)}{\left(\bar{\sigma}_X^2 + \sigma_\varepsilon^2 |2 \sin(\pi f_k \Delta t)|^2 \right)^2 \left(\bar{\sigma}_X^2 + \sigma_\varepsilon^2 |2 \sin(\pi f_l \Delta t)|^2 \right)^2}.\end{aligned}$$

We now need to calculate $\text{cov}(\widehat{\mathcal{F}}_{k,k}^{(Y)}, \widehat{\mathcal{F}}_{l,l}^{(Y)})$, which is

$$(A.2) \quad \begin{aligned} \text{cov}(\widehat{\mathcal{F}}_{k,k}^{(Y)}, \widehat{\mathcal{F}}_{l,l}^{(Y)}) &= \mathbb{E} \left\{ J_k^{(Y)} [J_k^{(Y)}]^* [J_l^{(Y)}]^* J_l^{(Y)} \right\} - \mathbb{E} \left\{ \widehat{\mathcal{F}}_{k,k}^{(Y)} \right\} \mathbb{E} \left\{ \widehat{\mathcal{F}}_{l,l}^{(Y)} \right\} \\ &= \rho_{kl}^{(Y)} \mathcal{F}_{k,k}^{(Y)} \mathcal{F}_{l,l}^{(Y)}. \end{aligned}$$

Furthermore

$$\begin{aligned} & \mathbb{E} \left\{ J_k^{(Y)} [J_k^{(Y)}]^* [J_l^{(Y)}]^* J_l^{(Y)} \right\} - \mathbb{E} \left\{ J_k^{(Y)} [J_k^{(Y)}]^* \right\} \mathbb{E} \left\{ [J_l^{(Y)}]^* J_l^{(Y)} \right\} \\ &= \mathbb{E} \left\{ (J_k^{(X)} + J_k^{(\varepsilon)}) [(J_k^{(X)} + J_k^{(\varepsilon)})]^* [(J_l^{(X)} + J_l^{(\varepsilon)})]^* (J_l^{(X)} + J_l^{(\varepsilon)}) \right\} \\ & \quad - \mathbb{E} \left\{ J_k^{(Y)} [J_k^{(Y)}]^* \right\} \mathbb{E} \left\{ [J_l^{(Y)}]^* J_l^{(Y)} \right\} \\ &= \text{cov} \left\{ \widehat{\mathcal{F}}_{k,k}^{(X)}, \widehat{\mathcal{F}}_{l,l}^{(X)} \right\} + \text{cov} \left\{ \widehat{\mathcal{F}}_{k,k}^{(\varepsilon)}, \widehat{\mathcal{F}}_{l,l}^{(\varepsilon)} \right\} + \mathcal{F}_{kl}^{(X)} \mathcal{F}_{k,l}^{(\varepsilon)*} + \mathcal{F}_{k,l}^{(X)*} \mathcal{F}_{k,l}^{(\varepsilon)}. \end{aligned}$$

We therefore need to calculate the individual terms of this expression. We note

$$\text{cov} \left\{ \widehat{\mathcal{F}}_{k,k}^{(\varepsilon)}, \widehat{\mathcal{F}}_{l,l}^{(\varepsilon)} \right\} = \delta_{kl} [\mathcal{F}_{k,k}^{(\varepsilon)}]^2, \quad \mathcal{F}_{k,l}^{(X)} \mathcal{F}_{k,l}^{(\varepsilon)*} + \mathcal{F}_{k,l}^{(X)*} \mathcal{F}_{k,l}^{(\varepsilon)} = 2\delta_{kl} \mathcal{F}_{k,k}^{(X)} \mathcal{F}_{k,k}^{(\varepsilon)}.$$

Then it follows that

$$(A.3) \quad \text{cov} \left\{ \widehat{\mathcal{F}}_{k,k}^{(Y)}, \widehat{\mathcal{F}}_{l,l}^{(Y)} \right\} = \text{cov} \left\{ \widehat{\mathcal{F}}_{k,k}^{(X)}, \widehat{\mathcal{F}}_{l,l}^{(X)} \right\} + \delta_{kl} [\mathcal{F}_{k,k}^{(\varepsilon)}]^2 + 2\delta_{kl} \mathcal{F}_{k,k}^{(X)} \mathcal{F}_{k,k}^{(\varepsilon)}.$$

We therefore only need to worry about $\text{cov} \left\{ \widehat{\mathcal{F}}_{k,k}^{(X)}, \widehat{\mathcal{F}}_{l,l}^{(X)} \right\}$. We need

$$\begin{aligned} & \mathbb{E} \left\{ J_k^{(X)} [J_k^{(X)}]^* [J_l^{(X)}]^* J_l^{(X)} \right\} = \frac{1}{N^2} \mathbb{E} \left\{ \sum_{n=1}^N \int_{(n-1)\Delta t}^{n\Delta t} \sigma_s dW_s e^{-2i\pi \frac{kn}{N}} \right. \\ & \times \left. \sum_{p=1}^N \int_{(p-1)\Delta t}^{p\Delta t} \sigma_t dW_t e^{2i\pi \frac{kp}{N}} \sum_{m=1}^N \int_{(m-1)\Delta t}^{m\Delta t} \sigma_u dW_u e^{-2i\pi \frac{lm}{N}} \sum_{w=1}^N \int_{(w-1)\Delta t}^{w\Delta t} \sigma_v dW_v e^{2i\pi \frac{lw}{N}} \right\} \\ & =: \frac{1}{N^2} \sum_{n=1}^N \sum_{p=1}^N \sum_{m=1}^N \sum_{\rho=1}^N \left(e_{kn} e_{kp}^* e_{\ell m}^* e_{\ell \rho} \mathbb{E} \left(M_n M_p M_m M_\rho \right) \right), \end{aligned}$$

where $M_n := \int_{(n-1)\Delta t}^{n\Delta t} \sigma_s dW_s$ and $e_{kn} := e^{-\frac{2i\pi kn}{N}}$. Since Brownian motion has independent increments, we have that $\mathbb{E}(M_n M_p M_m M_\rho) = \mathbb{E}M_n^4$ if $n = p = m = \rho$, $\mathbb{E}(M_n M_k M_m M_\rho) = \mathbb{E}M_n^2 \mathbb{E}M_k^2$ if $n = k, m = \rho$, and $\mathbb{E}(M_n M_p M_m M_\rho) = 0$ otherwise. Consequently,

$$\begin{aligned} \mathbb{E} \left\{ J_k^{(X)} [J_k^{(X)}]^* [J_l^{(X)}]^* J_l^{(X)} \right\} &= \frac{1}{N^2} \sum_{n=1}^N \mathbb{E}M_n^4 + \frac{1}{N^2} \left(\sum_{n=1}^N \mathbb{E}M_n^2 \right)^2 \\ & \quad + \frac{1}{N^2} \sum_{n=1}^N \sum_{p=1}^N e_{kn} e_{\ell n}^* e_{kp}^* e_{\ell p} \mathbb{E}M_n^2 \mathbb{E}M_p^2 \\ & \quad + \frac{1}{N^2} \sum_{n=1}^N \sum_{p=1}^N e_{kn} e_{\ell p}^* e_{kp}^* e_{\ell n} \mathbb{E}M_n^2 \mathbb{E}M_p^2. \end{aligned}$$

We use standard bounds on moments of stochastic integrals [26] to obtain the bound

$$\frac{1}{N^2} \sum_{n=1}^N EM_n^4 \leq \frac{1}{N^2} \sum_{n=1}^N 36\Delta t \int_{(n-1)\Delta t}^{n\Delta t} E\sigma_s^4 ds \leq C(\Delta t)^3,$$

since, by assumption, $E\sigma_s^4 = \mathcal{O}(1)$.⁵ We have

$$\begin{aligned} \rho_{kl}^{(X)} \mathcal{J}_{k,k}^{(X)} \mathcal{J}_{l,l}^{(X)} &= E \left\{ J_k^{(X)} [J_k^{(X)}]^* [J_l^{(X)}]^* J_l^{(X)} \right\} - E \left| J_k^{(X)} \right|^2 E \left| J_l^{(X)} \right|^2 \\ &= \frac{1}{N^2} \int_0^T \int_0^T \left(\cos \left(2\pi(k+l) \left(\frac{s-u}{T} \right) \right) + \cos \left(2\pi(k-l) \left(\frac{s-u}{T} \right) \right) \right) \\ &\quad \times E \left\{ \sigma_s^2 \right\} E \left\{ \sigma_u^2 \right\} dsdu + \mathcal{O}((\Delta t)^3) \\ &= \frac{1}{2N^2} \int_0^T \int_0^T E \left\{ \sigma_s^2 \right\} E \left\{ \sigma_u^2 \right\} \left(e^{2i\pi(k+l)(\frac{s-u}{T})} + e^{-2i\pi(k+l)(\frac{s-u}{T})} \right. \\ &\quad \left. + e^{2i\pi(k-l)(\frac{s-u}{T})} + e^{-2i\pi(k-l)(\frac{s-u}{T})} \right) dsdu + \mathcal{O}((\Delta t)^3) \\ &= \frac{1}{2N^2} \left(\Xi \left(-\frac{k+l}{T} \right) \Xi \left(\frac{k+l}{T} \right) + \Xi \left(\frac{k+l}{T} \right) \Xi \left(-\frac{k+l}{T} \right) \right. \\ &\quad \left. + \Xi \left(-\frac{k-l}{T} \right) \Xi \left(\frac{k-l}{T} \right) + \Xi \left(\frac{k-l}{T} \right) \Xi \left(-\frac{k-l}{T} \right) \right) + \mathcal{O}((\Delta t)^3). \end{aligned}$$

Since $E\sigma_t^2$ is a smooth function of time we can bound the decay of $\Xi(f) \propto \frac{1}{f}$ so that

$$(A.4) \quad \rho_{kl}^{(X)} \mathcal{J}_{k,k}^{(X)} \mathcal{J}_{l,l}^{(X)} = \Delta t^2 \left(\mathcal{O} \left(\frac{1}{(k+l)^2} \right) + \mathcal{O} \left(\frac{1}{(k-l)^2} \right) \right).$$

We combine the foregoing calculations with (A.3):

$$\begin{aligned} \text{var} \left\{ \widehat{\mathcal{F}}_{k,k}^{(Y)} \right\} &= \left(\mathcal{J}_{k,k}^{(X)} + \mathcal{J}_{k,k}^{(\varepsilon)} \right)^2, \\ \text{var} \left(\dot{\ell}_\tau(\widehat{\sigma}) \right) &= \sum_{l=1}^{N/2-1} \sum_{k=1}^{N/2-1} \frac{\Delta t^2 \text{cov} \left(\widehat{\mathcal{F}}_{k,k}^{(Y)}, \widehat{\mathcal{F}}_{l,l}^{(Y)} \right)}{\left(\overline{\sigma}_X^2 + \sigma_\varepsilon^2 |2 \sin(\pi f_k \Delta t)|^2 \right)^2 \left(\overline{\sigma}_X^2 + \sigma_\varepsilon^2 |2 \sin(\pi f_l \Delta t)|^2 \right)^2}. \end{aligned} \tag{A.5}$$

We note that

$$\text{cov} \left(\widehat{\mathcal{F}}_{k,k}^{(Y)}, \widehat{\mathcal{F}}_{l,l}^{(Y)} \right) = \rho_{kl}^{(X)} \mathcal{J}_{k,k}^{(X)} \mathcal{J}_{l,l}^{(X)} + \delta_{kl} \left[\mathcal{J}_{l,l}^{(\varepsilon)} \right]^2 + 2\delta_{kl} \mathcal{J}_{k,k}^{(X)} \mathcal{J}_{k,k}^{(\varepsilon)}.$$

Thus it follows that

$$\begin{aligned} (A.6) \quad \text{var} \left(\dot{\ell}_\tau(\widehat{\sigma}) \right) &= \sum_{l=1}^{N/2-1} \frac{\Delta t^2}{\left(\overline{\sigma}_X^2 + \sigma_\varepsilon^2 |2 \sin(\pi f_l \Delta t)|^2 \right)^2} + C + \mathcal{O}(\log(\Delta t)\Delta t^{-1/4}) \\ &= \mathcal{O}(\Delta t^{-1/2}) + C + \mathcal{O}(\log(\Delta t)\Delta t^{-1/4}). \end{aligned}$$

The extra order terms acknowledge potential effects from the drift. We need to establish the size of C . Using (A.3) we find that

⁵ C in this paper denotes a generic constant, rather than the same constant.

$$\begin{aligned}
 |C| &\leq \sum_{l \neq k}^{N/2-1} \frac{\Delta t^4 C_2 ((k+l)^{-2} + (k-l)^{-2})}{\left(\bar{\sigma}_X^2 + \sigma_\varepsilon^2 |2 \sin(\pi f_k \Delta t)|^2\right)^2 \left(\bar{\sigma}_X^2 + \sigma_\varepsilon^2 |2 \sin(\pi f_l \Delta t)|^2\right)^2} \\
 &\leq 2 \sum_{k=1}^{N/2-1} \sum_{\tau=1}^k \frac{\Delta t^4 C_2 ((2k-\tau)^{-2} + \tau^{-2})}{\left(\bar{\sigma}_X^2 + \sigma_\varepsilon^2 |2 \sin(\pi f_{k-\tau} \Delta t)|^2\right)^2 \left(\bar{\sigma}_X^2 + \sigma_\varepsilon^2 |2 \sin(\pi f_k \Delta t)|^2\right)^2} \\
 &\sim 2 \sum_{k=1}^{N/2-1} \sum_{\tau=1}^k \frac{C_2 ((2k-\tau)^{-2} + \tau^{-2})}{\left(\tau_X^2 + \sigma_\varepsilon^2 |2 \sin(\pi f_{k-\tau} \Delta t)|^2 / \Delta t\right)^2 \left(\tau_X^2 + \sigma_\varepsilon^2 |2 \sin(\pi f_k \Delta t)|^2 / \Delta t\right)^2} \\
 &= \mathcal{O}(\log(\Delta t)).
 \end{aligned}$$

This is negligible in size compared to $\Delta t^{-1/2}$. Similar calculations can bound contributions from the off-diagonals in the other two calculations. Also, as $\bar{\sigma}_X^2 = \tau_X \Delta t$,

(A.7)

$$\begin{aligned}
 -\mathbb{E} \left\{ \ddot{\ell}_{\tau\tau}(\boldsymbol{\sigma}) \right\} &= \sum_{k=1}^{N/2-1} \frac{\Delta t^2}{\left(\bar{\sigma}_X^2 + \sigma_\varepsilon^2 |2 \sin(\pi f_k \Delta t)|^2\right)^2} + \mathcal{O}(\log(\Delta t)) = \mathcal{O}(\Delta t^{-1/2}), \\
 -\mathbb{E} \left\{ \ddot{\ell}_{\varepsilon\varepsilon}(\boldsymbol{\sigma}) \right\} &= \sum_{k=1}^{N/2-1} \frac{|2 \sin(\pi f_k)|^4}{\left(\bar{\sigma}_X^2 + \sigma_\varepsilon^2 |2 \sin(\pi f_k \Delta t)|^2\right)^2} + \mathcal{O}(\log(\Delta t)) = \mathcal{O}(\Delta t^{-1}), \\
 -\mathbb{E} \left\{ \ddot{\ell}_{\tau\varepsilon}(\boldsymbol{\sigma}) \right\} &= \sum_{k=1}^{N/2-1} \frac{\Delta t |2 \sin(\pi f_k)|^2}{\left(\bar{\sigma}_X^2 + \sigma_\varepsilon^2 |2 \sin(\pi f_k \Delta t)|^2\right)^2} + \mathcal{O}(\log(\Delta t)) = \mathcal{O}(\Delta t^{-1/2}).
 \end{aligned}$$

The order terms follow from usual spectral theory on the white noise process, as well as bounds on $J_k^{(\mu)}$. By considering the variance of the observed Fisher information, we can also deduce that renormalized versions of the entries of the observed Fisher information converge in probability to a constant, or

$$\text{diag}(\Delta t^{1/4}, \Delta t^{1/2}) \mathbf{F} \text{diag}(\Delta t^{1/4}, \Delta t^{1/2}) \rightarrow \mathcal{F},$$

and thus using Slutsky's theorem we can deduce that

$$\text{diag}(\Delta t^{-1/4}, \Delta t^{-1/2}) \left[\begin{pmatrix} \hat{\sigma}_X^2 / \Delta t \\ \hat{\sigma}_\varepsilon^2 \end{pmatrix} - \begin{pmatrix} \bar{\sigma}_X^{*2} / \Delta t \\ \sigma_\varepsilon^{*2} \end{pmatrix} \right] \text{diag}(\Delta t^{-1/4}, \Delta t^{-1/2}) \xrightarrow{L} N(\mathbf{0}, \mathcal{F}^{-1}),$$

where the entries of \mathcal{F} can be found from (A.7), (A.5), and (A.6), and

$$\begin{aligned}
 &\text{diag}(\Delta t^{-1/4}, \Delta t^{-1/2}) \text{var} \left\{ \left[\begin{pmatrix} \hat{\sigma}_X^2 / \Delta t \\ \hat{\sigma}_\varepsilon^2 \end{pmatrix} - \begin{pmatrix} \bar{\sigma}_X^{*2} / \Delta t \\ \sigma_\varepsilon^{*2} \end{pmatrix} \right] \right\} \text{diag}(\Delta t^{-1/4}, \Delta t^{-1/2}) \\
 &= \text{diag}(\Delta t^{-1/4}, \Delta t^{-1/2}) \mathbf{F}^{-1} \mathbf{F} \mathbf{F}^{-1} \text{diag}(\Delta t^{-1/4}, \Delta t^{-1/2}) \\
 &= \text{diag}(\Delta t^{-1/4}, \Delta t^{-1/2}) \mathbf{F}^{-1} \text{diag}(\Delta t^{-1/4}, \Delta t^{-1/2}) \rightarrow \mathcal{F}^{-1}.
 \end{aligned}$$

We have

$$(A.8) \quad \mathcal{F} = \begin{pmatrix} \frac{T}{\sigma_\varepsilon 16} & 0 \\ \frac{1}{\tau_X^{3/2}} & 0 \\ 0 & \frac{2T}{\sigma_\varepsilon^2} \end{pmatrix} = \begin{pmatrix} \mathcal{I}_{\tau\tau} & 0 \\ 0 & \mathcal{I}_{\varepsilon\varepsilon} \end{pmatrix}.$$

This expression follows by direct calculation. Asymptotic normality of both $\widehat{\tau}_x$ and $\widehat{\sigma}_\varepsilon^2$ follows by the usual arguments. We can determine the asymptotic variance of $\widehat{\langle X, X \rangle}^{(w)}$ via

$$\begin{aligned} \text{var} \left\{ \widehat{\langle X, X \rangle}^{(w)} \right\} &= T^2 \text{var} \{ \widehat{\tau}_x \} \\ \text{(A.9)} \qquad \qquad \qquad &= T \frac{\sigma_\varepsilon}{\tau_X^{1/2}} 16 \tau_X^2 \sqrt{\Delta t}. \end{aligned}$$

We see that the variance depends on the length of the time course, the inverse of the signal-to-noise ratio, the square root of the sampling period, and the fourth power of the ‘‘average standard deviation’’ of the X_t process.

Appendix B. Proof of Theorem 2.4. We now wish to use these results to deduce properties of $\widehat{\sigma}$, first using the well-known *invariance* of maximum likelihood estimators to transfer the estimators of $\overline{\sigma}_X^2$ and σ_ε^2 to estimators of $\langle X, X \rangle_T$. We therefore take

$$\widehat{\langle X, X \rangle}_T^{(m_1)} = \sum_{k=0}^{N-1} \widehat{\mathcal{F}}_{k,k}^{(X)}(\widehat{L}_k) = \sum_{k=0}^{N-1} \widehat{L}_k \widehat{\mathcal{F}}_{k,k}^{(Y)}.$$

It therefore follows that with $\widehat{\tau}_X = \tau_X + \delta\tau_X$ and $\widehat{\sigma}_\varepsilon^2 = \sigma_\varepsilon^2 + \delta\sigma_\varepsilon^2$

$$\begin{aligned} \text{E} \left\{ \widehat{\langle X, X \rangle}_T^{(m_1)} \right\} &= \sum_{k=0}^{N-1} \text{E} \left\{ \left(\frac{\overline{\sigma}_X^2 + \delta\sigma_X^2}{\overline{\sigma}_X^2 + \delta\sigma_X^2 + (\sigma_\varepsilon^2 + \delta\sigma_\varepsilon^2) |2 \sin(\pi f_k \Delta t)|^2} \right) \widehat{\mathcal{F}}_{k,k}^{(Y)} \right\} \\ &= \sum_{k=0}^{N-1} \text{E} \left\{ \left(\frac{\overline{\sigma}_X^2 + \delta\sigma_X^2}{1 + \frac{[\delta\sigma_X^2 + \delta\sigma_\varepsilon^2] |2 \sin(\pi f_k \Delta t)|^2}{\overline{\sigma}_X^2 + \sigma_\varepsilon^2 |2 \sin(\pi f_k \Delta t)|^2}} \right) \frac{\widehat{\mathcal{F}}_{k,k}^{(Y)}}{\overline{\sigma}_X^2 + \sigma_\varepsilon^2 |2 \sin(\pi f_k \Delta t)|^2} \right\} \\ &= \sum_{k=0}^{N-1} \text{E} \left\{ (\overline{\sigma}_X^2 + \delta\sigma_X^2) \left(1 - \frac{[\delta\sigma_X^2 + \delta\sigma_\varepsilon^2] |2 \sin(\pi f_k \Delta t)|^2}{(\overline{\sigma}_X^2 + \sigma_\varepsilon^2 |2 \sin(\pi f_k \Delta t)|^2)} \right) \frac{\widehat{\mathcal{F}}_{k,k}^{(Y)}}{\overline{\sigma}_X^2 + \sigma_\varepsilon^2 |2 \sin(\pi f_k \Delta t)|^2} \right\} \\ &= \sum_{k=0}^{N-1} \left[\overline{\sigma}_X^2 + \mathcal{O}(\Delta t^{5/4}) \right] + \mathcal{O}(\sqrt{\Delta t} \log(\Delta t)) = \text{E} \{ \langle X, X \rangle_T \} + \mathcal{O}(\sqrt{\Delta t} \log(\Delta t)) \\ &\quad + \mathcal{O}(\sqrt[4]{\Delta t}). \end{aligned}$$

This implies that the estimator is asymptotically unbiased. We can also note that the variance of the new estimator is given by

$$\begin{aligned} \text{var} \left\{ \widehat{\langle X, X \rangle}_T^{(m_1)} \right\} &= \sum_j \sum_k \text{cov} \{ \widehat{L}_j \widehat{\mathcal{F}}_{j,j}^{(Y)}, \widehat{L}_k \widehat{\mathcal{F}}_{k,k}^{(Y)} \} \\ &= \sum_j \sum_k \text{cov} \left\{ \frac{\widehat{L}_j}{L_j} L_j \widehat{\mathcal{F}}_{j,j}^{(Y)}, \frac{\widehat{L}_k}{L_k} L_k \widehat{\mathcal{F}}_{k,k}^{(Y)} \right\} \\ &= \sum_j \sum_k \text{cov} \left\{ \left(1 + \frac{\delta\tau_X}{\tau_X} - \frac{\delta\tau_X \Delta t + \delta\sigma_\varepsilon^2 |2 \sin(\pi f_j \Delta t)|^2}{\tau_X \Delta t + \sigma_\varepsilon^2 |2 \sin(\pi f_j \Delta t)|^2} \right) L_j \widehat{\mathcal{F}}_{j,j}^{(Y)}, \right. \\ &\quad \left. \left(1 + \frac{\delta\tau_X}{\tau_X} - \frac{\delta\tau_X \Delta t + \delta\sigma_\varepsilon^2 |2 \sin(\pi f_k \Delta t)|^2}{\tau_X \Delta t + \sigma_\varepsilon^2 |2 \sin(\pi f_k \Delta t)|^2} \right) L_k \widehat{\mathcal{F}}_{k,k}^{(Y)} \right\}. \end{aligned}$$

Then

$$\begin{aligned}
\text{var} \left\{ \widehat{\langle X, X \rangle}_T^{(m_1)} \right\} &= \sum_j \sum_k \left\{ \text{cov} \{ L_j \widehat{\mathcal{F}}_{j,j}^{(Y)}, L_k \widehat{\mathcal{F}}_{k,k}^{(Y)} \} + \text{cov} \left\{ \frac{\delta \tau_X}{\tau_X} L_j \widehat{\mathcal{F}}_{j,j}^{(Y)}, L_k \widehat{\mathcal{F}}_{k,k}^{(Y)} \right\} \right. \\
&\quad + \text{cov} \left\{ L_j \widehat{\mathcal{F}}_{j,j}^{(Y)}, \frac{\delta \tau_X}{\tau_X} L_k \widehat{\mathcal{F}}_{k,k}^{(Y)} \right\} \\
&\quad - \text{cov} \left\{ \frac{\delta \tau_X \Delta t + \delta \sigma_\varepsilon^2 |2 \sin(\pi f_j \Delta t)|^2}{\tau_X \Delta t + \sigma_\varepsilon^2 |2 \sin(\pi f_j \Delta t)|^2} L_j \widehat{\mathcal{F}}_{j,j}^{(Y)}, L_k \widehat{\mathcal{F}}_{k,k}^{(Y)} \right\} \\
&\quad - \text{cov} \left\{ L_j \widehat{\mathcal{F}}_{j,j}^{(Y)}, \frac{\delta \tau_X \Delta t + \delta \sigma_\varepsilon^2 |2 \sin(\pi f_k \Delta t)|^2}{\tau_X \Delta t + \sigma_\varepsilon^2 |2 \sin(\pi f_k \Delta t)|^2} L_k \widehat{\mathcal{F}}_{k,k}^{(Y)} \right\} \\
&\quad \left. + \text{cov} \left\{ \frac{\delta \tau_X}{\tau_X} L_j \widehat{\mathcal{F}}_{j,j}^{(Y)}, \frac{\delta \tau_X}{\tau_X} L_k \widehat{\mathcal{F}}_{k,k}^{(Y)} \right\} + \dots \right\} \\
&= \sum_j \sum_k \left\{ \delta_{jk} \sigma_X^4 + L_j L_k \text{cov} \left\{ \frac{\delta \tau_X}{\tau_X} \widehat{\mathcal{F}}_{j,j}^{(Y)}, \widehat{\mathcal{F}}_{k,k}^{(Y)} \right\} + \dots \right\}.
\end{aligned}$$

By looking at the individual terms of this expression, and noting that the estimated renormalized variance $\widehat{\tau}_X = \tau_X + \delta \tau_X$ and $\widehat{\sigma}_\varepsilon^2 = \sigma_\varepsilon^2 + \delta \sigma_\varepsilon^2$ are linear combinations of $\widehat{\mathcal{F}}_{k,k}^{(Y)}$, we can deduce the stated order terms by again noting the $\sqrt{\Delta t}$ order of the important terms. However, to leading order, this estimator will perform identically to $\widehat{\langle X, X \rangle}^{(w)}$ in terms of variance.

Appendix C. Proof of time-domain form. The integral can be calculated from first principles using complex variables with $z = e^{2i\pi f}$. Thus $dz/df = 2i\pi z$ or $df = dz/(2i\pi z)$. Equation (2.27) takes the form

$$(C.1) \quad \ell_\tau = \frac{1}{2i\pi} \oint_{|z|=1} \frac{\overline{\sigma}_X^2}{\overline{\sigma}_X^2 z - \sigma_\varepsilon^2 [z-1]^2} z^\tau dz.$$

We need the poles, or

$$\overline{\sigma}_X^2 z - \sigma_\varepsilon^2 [z-1]^2 = 0 \iff z = 1 + \frac{\overline{\sigma}_X^2}{2\sigma_\varepsilon^2} \pm \sqrt{\frac{\overline{\sigma}_X^2}{\sigma_\varepsilon^2} + \frac{\overline{\sigma}_X^4}{4\sigma_\varepsilon^4}} = z^\pm.$$

If $|\frac{\sigma_\varepsilon^2}{\overline{\sigma}_X^2}| < 1$, we have

$$\begin{aligned}
z^- &= 1 + \frac{\overline{\sigma}_X^2}{2\sigma_\varepsilon^2} - \frac{\overline{\sigma}_X^2}{2\sigma_\varepsilon^2} \sqrt{1 + \frac{4\sigma_\varepsilon^2}{\overline{\sigma}_X^2}} \\
&= 1 + \frac{\overline{\sigma}_X^2}{2\sigma_\varepsilon^2} - \frac{\overline{\sigma}_X^2}{2\sigma_\varepsilon^2} \left(1 + \frac{1}{2} \frac{4\sigma_\varepsilon^2}{\overline{\sigma}_X^2} + \frac{1}{4} \frac{(-1)}{2} \left[\frac{4\sigma_\varepsilon^2}{\overline{\sigma}_X^2} \right]^2 + \mathcal{O} \left(\frac{\sigma_\varepsilon^6}{\overline{\sigma}_X^6} \right) \right), \\
&= \frac{\sigma_\varepsilon^2}{\overline{\sigma}_X^2} + \mathcal{O} \left(\frac{\sigma_\varepsilon^4}{\overline{\sigma}_X^4} \right), \\
z^+ &= \frac{\overline{\sigma}_X^2}{\sigma_\varepsilon^2} + \dots
\end{aligned}$$

We then note that

$$\begin{aligned} \ell_\tau &= -\frac{1}{2i\pi} \oint_{|z|=1} \frac{\bar{\sigma}_X^2/\sigma_\varepsilon^2}{-(\bar{\sigma}_X^2/\sigma_\varepsilon^2)z + [z-1]^2} z^\tau dz = -\frac{1}{2i\pi} \oint_{|z|=1} \frac{\bar{\sigma}_X^2/\sigma_\varepsilon^2}{(z-z^-)(z-z^+)} z^\tau dz \\ &= \frac{2i\pi \bar{\sigma}_X^2/\sigma_\varepsilon^2}{2i\pi} \frac{\left(\frac{\sigma_\varepsilon^2}{\sigma_X^2}\right)^\tau}{z^+ - \frac{\sigma_\varepsilon^2}{\sigma_X^2} + \mathcal{O}\left(\frac{\sigma_\varepsilon^4}{\sigma_X^4}\right)} = \left(\frac{\sigma_\varepsilon^2}{\sigma_X^2}\right)^\tau + \mathcal{O}\left(\frac{\sigma_\varepsilon^{2\tau+2}}{\sigma_X^{2\tau+2}}\right). \end{aligned}$$

If on the other hand we consider $\left|\frac{\sigma_\varepsilon^2}{\sigma_X^2}\right| > 1$, which in many scenarios is more realistic, then we find that

$$\begin{aligned} z^- &= 1 + \frac{\bar{\sigma}_X^2}{2\sigma_\varepsilon^2} - \frac{\bar{\sigma}_X}{\sigma_\varepsilon} \sqrt{1 + \frac{\sigma_X^2}{4\Delta t \sigma_\varepsilon^2}} = 1 + \frac{\bar{\sigma}_X^2}{2\sigma_\varepsilon^2} - \frac{\bar{\sigma}_X}{\sigma_\varepsilon} \left(1 + \frac{1}{2} \frac{\sigma_X^2}{4\sigma_\varepsilon^2}\right) \\ &= 1 - \frac{\bar{\sigma}_X}{\sigma_\varepsilon} + \mathcal{O}\left(\frac{\bar{\sigma}_X^2}{\sigma_\varepsilon^2}\right), \\ z^+ &= 1 + \frac{\bar{\sigma}_X}{\sigma_\varepsilon} + \mathcal{O}\left(\frac{\bar{\sigma}_X^2}{\sigma_\varepsilon^2}\right). \end{aligned}$$

In this case we find that

$$\ell_\tau = \bar{\sigma}_X^2/(\sigma_\varepsilon^2) \frac{[1 - \frac{\bar{\sigma}_X}{\sigma_\varepsilon}]^\tau}{2\frac{\bar{\sigma}_X}{\sigma_\varepsilon} + \mathcal{O}\left(\frac{\bar{\sigma}_X^2}{\sigma_\varepsilon^2}\right)} = \frac{\bar{\sigma}_X}{2\sigma_\varepsilon} \left(1 - \frac{\bar{\sigma}_X}{\sigma_\varepsilon}\right)^\tau + \mathcal{O}\left(\frac{\bar{\sigma}_X^2}{2\sigma_\varepsilon^2} \left(1 - \frac{\bar{\sigma}_X}{\sigma_\varepsilon}\right)^\tau\right).$$

In both cases the decay of the filter is geometric. We note that in most practical examples L_k decays very rapidly in k . Therefore, we do not need to integrate between $-1/2$ to $1/2$, and only need to integrate over $-1/\pi$ to $1/\pi$. In this range of f we find that for the smallish remainder term R_3 we have $\sin^2(\pi f) = \pi^2 f^2 + R_3(f\pi)$. Then we note

$$\begin{aligned} \ell_\tau &= \int_{-\frac{1}{\pi}}^{\frac{1}{\pi}} \frac{\bar{\sigma}_X^2}{\bar{\sigma}_X^2 + 4\sigma_\varepsilon^2 \pi^2 f^2 + R_3(f\pi)} e^{2i\pi f\tau} df + C \\ &= \frac{\bar{\sigma}_X}{2\sigma_\varepsilon} \int_{-\infty}^{\infty} \left[2\frac{\bar{\sigma}_X/\sigma_\varepsilon}{\frac{\sigma_X^2}{\sigma_\varepsilon^2} + 4\pi^2 f^2} + R_4(f) \right] e^{2i\pi f\tau} df + C \\ &= \frac{\bar{\sigma}_X}{2\sigma_\varepsilon} e^{-\frac{\bar{\sigma}_X|\tau|}{\sigma_\varepsilon}} + C. \end{aligned}$$

Thus we are smoothing the autocovariance sequence with a smoothing window that becomes a delta function as $\bar{\sigma}_X/\sigma_\varepsilon \rightarrow \infty$. It is reasonable that this nondimensional quantity arises as an important factor.

Acknowledgments. We thank the anonymous referee for his/her extremely careful reading of an earlier version of this paper and for the many useful suggestions. SO would also like to thank Dr. Wenyang Zhang of Bath University for a very enlightening discussion on kernel estimators.

REFERENCES

- [1] R. J. ADLER, P. MÜLLER, AND B. L. ROZOVSKII, *Stochastic Modelling in Physical Oceanography*, Birkhäuser, Berlin, 1996.
- [2] Y. AIT-SAHALIA AND R. KIMMEL, *Maximum likelihood estimation of stochastic volatility models*, *J. Financial Econom.*, 83 (2007), pp. 413–452.
- [3] Y. AIT-SAHALIA, P. A. MYKLAND, AND L. ZHANG, *How often to sample a continuous-time process in the presence of market microstructure noise*, *Rev. Financ. Stud.*, 18 (2005), pp. 351–416.
- [4] E. ALLEN, *Modeling with Itô Stochastic Differential Equations*, Springer-Verlag, Berlin, 2007.
- [5] E. BARUCCI AND R. RENO, *On measuring volatility of diffusion processes with high frequency data*, *Econom. Lett.*, 74 (2002), pp. 371–378.
- [6] J. BERAN, *Statistics for Long-Memory Processes*, Chapman and Hall, London, 1994.
- [7] J. P. N. BISHWAL, *Parameter Estimation in Stochastic Differential Equations*, Springer-Verlag, Berlin, 2008.
- [8] P. BLOOMFIELD, *Fourier Analysis of Time Series*, Wiley-IEEE, New York, 2004.
- [9] D. BRILLINGER, *Time Series: Data Analysis and Theory*, SIAM, Philadelphia, 2001.
- [10] P. J. BROCKWELL AND R. A. DAVIS, *Time Series: Theory and Methods*, Springer-Verlag, New York, 1991.
- [11] C. J. COTTER AND G. A. PAVLIOTIS, *Estimating eddy diffusivities from Lagrangian observations*, *Commun. Math. Sci.*, to appear (2009).
- [12] K. O. DZHAMPARIDZE AND A. M. YAGLOM, *Spectrum parameter estimation in time series analysis*, in *Developments in Statistics*, Vol. 4, P. R. Krishnaiah, ed., Academic Press, New York, 1983, pp. 1–181.
- [13] J. FAN AND Y. WANG, *Multi-scale jump and volatility analysis for high-frequency financial data*, *J. Amer. Statist. Assoc.*, 102 (2007), pp. 1349–1362.
- [14] D. FLORENS-ZMIROU, *On estimating the diffusion coefficient from discrete observations*, *J. Appl. Probab.*, 30 (1993), pp. 790–804.
- [15] A. FRIEDMAN, *Stochastic Differential Equations and Applications*, Dover, London, 2007.
- [16] J.-P. FOUQUE, G. PAPANICOLAOU, AND K. R. SIRCAR, *Derivatives in Financial Markets with Stochastic Volatility*, Cambridge University Press, Cambridge, UK, 2000.
- [17] J.-P. FOUQUE, G. PAPANICOLAOU, K. R. SIRCAR, AND K. SOLNA, *Short time-scale in SP-500 volatility*, *J. Comput. Finance*, 6 (2003), pp. 1–23.
- [18] A. GRIFFA, K. OWENSA, L. PITERBARG, AND B. ROZOVSKII, *Estimates on turbulence parameters from Lagrangian data using a stochastic particle model*, *J. Marine Res.*, 53 (1995), pp. 371–401.
- [19] E. J. HANNAN, *Multiple Time Series*, John Wiley, New York, 1970.
- [20] P. R. HANSEN AND A. LUNDE, *A forecast comparison of volatility models: Does anything beat a GARCH(1, 1)?*, *J. Appl. Econometrics*, 20 (2005), pp. 873–889.
- [21] P. R. HANSEN AND A. LUNDE, *Realized variance and market microstructure noise*, *J. Bus. Econom. Statist.*, 24 (2006), pp. 127–161.
- [22] S. L. HESTON, *A closed form solution for options with stochastic volatility with applications to bond and currency options*, *Rev. Financ. Stud.*, 6 (1993), pp. 327–343.
- [23] C. M. HURVICH AND C.-L. TSAI, *Regression and time series model selection in small samples*, *Biometrika*, 76 (1989), pp. 297–307.
- [24] S. M. IACUS, *Simulation and Inference for Stochastic Differential Equations with R Examples*, Springer-Verlag, New York, 2008.
- [25] J. JACOD AND A. SHIRYAEV, *Limit Theorems for Stochastic Processes*, Springer-Verlag, New York, 2003.
- [26] I. KARATZAS AND S. E. SHREVE, *Brownian Motion and Stochastic Calculus*, Springer-Verlag, New York, 1991.
- [27] Y. A. KUTOYANTS, *Statistical Inference for Ergodic Diffusion Processes*, Springer Series in Statistics, Springer-Verlag, London, 2004.
- [28] Y. A. KUTOYANTS, *Semiparametric estimation for dynamical systems with small noise*, *Math. Methods Statist.*, 7 (1998), pp. 457–465.
- [29] P. MALLIAVIN AND M. E. MANCINO, *Fourier series for measurements of multivariate volatilities*, *Finance Stoch.*, 6 (2002), pp. 49–61.
- [30] M. E. MANCINO AND S. SANFELICI, *Robustness of Fourier Estimators of Integrated Volatility in the Presence of Microstructure Noise*, tech. report, University of Firenze, Firenze, Italy, 2006.
- [31] X. MAO, *Stochastic Differential Equations and Applications*, Horwood, Chichester, UK, 2008.
- [32] F. NEESER AND J. MASSEY, *Proper complex random processes with applications to information theory*, *IEEE Trans. Inform. Theory*, 39 (1993), pp. 1293–1302.

- [33] B. K. OKSENDAL, *Stochastic Differential Equations: An Introduction with Applications*, 6th ed., Springer-Verlag, Berlin, 2003.
- [34] T. PAPAVALIOU, G. A. PAVLIOTIS, AND A. M. STUART, *Maximum likelihood drift estimation for multiscale diffusions*, *Stochastic Process. Appl.*, 119 (2009), pp. 3173–3210.
- [35] G. A. PAVLIOTIS AND A. M. STUART, *Parameter estimation for multiscale diffusions*, *J. Statist. Phys.*, 127 (2007), pp. 741–781.
- [36] D. B. PERCIVAL AND A. T. WALDEN, *Spectral Analysis for Physical Applications*, Cambridge University Press, Cambridge, UK, 1993.
- [37] B. PICINBONO, *Second-order complex random vectors and normal distributions*, *IEEE Trans. Signal Process.*, 44 (1996), pp. 2637–2640.
- [38] B. L. S. PRAKASA RAO, *Statistical Inference for Diffusion Type Processes*, Kendall's Lib. Statist. 8, Edward Arnold, London, 1999.
- [39] R. RENÒ, *Volatility Estimate via Fourier Analysis*, Ph.D. thesis, Scuola Normale Superiore, 2005; available online from <http://www.econ-pol.unisi.it/~reno/tesi.pdf>.
- [40] R. RENÒ, *Nonparametric estimation of the diffusion coefficient of stochastic volatility models*, *Econometric Theory*, 24 (2008), pp. 1174–1206.
- [41] T. SEVERINI, *Likelihood Methods in Statistics*, Oxford University Press, Oxford, 2001.
- [42] N. SHEPHARD, *Stochastic Volatility: Selected Readings*, Oxford University Press, Oxford, 2005.
- [43] A. SYKULSKI, S. OLHEDE, AND G. A. PAVLIOTIS, *High frequency variability and microstructure bias*, in *Inference and Estimation in Probabilistic Time-Series Models*, D. Barber, A. T. Cemgil, and S. Chiappa, eds., Isaac Newton Institute for Mathematical Sciences, Cambridge, UK, 2008; proceedings available from <http://www.newton.ac.uk/programmes/SCH/schw05-papers.pdf>.
- [44] R. S. TSAY, *Analysis of Financial Time Series*, John Wiley, New York, 2005.
- [45] L. WASSERMAN, *All of Nonparametric Statistics*, Springer-Verlag, Berlin, 2007.
- [46] P. WHITTLE, *Estimation and information in stationary time series*, *Ark. Mat.*, 2 (1953), pp. 423–434.
- [47] P. WHITTLE, *Gaussian estimation in stationary time series*, *Bull. Inst. Statist. Inst.*, 39 (1962), pp. 105–129.
- [48] L. ZHANG, P. A. MYKLAND, AND Y. AIT-SAHALIA, *A tale of two time scales: Determining integrated volatility with noisy high-frequency data*, *J. Amer. Statist. Assoc.*, 100 (2005), pp. 1394–1411.