Multiscale Inference for High-Frequency Data

Adam Sykulski, Sofia C. Olhede and Grigoris A. Pavliotis

Abstract

This paper proposes a novel multiscale estimator for the integrated volatility of an Itô process with harmonizable increments, in the presence of market microstructure noise. The multiscale structure is modelled frequency-by-frequency and the concept of the multiscale ratio is introduced to quantify the bias in the quadratic variation due to the microstructure noise process. The multiscale ratio is estimated from a single sample path, and a frequency-by-frequency bias correction procedure is proposed. The new method is implemented to estimate the integrated volatility for the Heston model, and the improved performance of our method is illustrated by simulation studies.

KEYWORDS: Bias correction; microstructure; realized volatility; multiscale inference.

I. INTRODUCTION

Many phenomena in finance, engineering and the sciences have multiple characteristic time scales. As examples we mention financial time series, [1], [2], atmosphere/ocean science, [3], [4], and molecular dynamics, [5]–[7]. In such applications one is confronted with data that contain a high frequency component. Hence, the development of appropriate inference methods for high frequency data sets is a very important and challenging problem. High frequency data also often exhibits multiscale characteristics, i.e. disparate structural features associated with different time scales. Such scale disparity has not yet been well investigated, due to sampling limitations.

In financial time series the more recent availability of high frequency observations has uncovered a number of inhomogeneous effects, such as non-synchronous trading, bid-ask spread and other microstructure features, see [2]. It is a standard practice to use an Itô process (i.e. a stochastic differential equation–SDE) to model the financial time series of interest. To account for observed high frequency characteristics, market microstructure noise, an additional noise process is added in the model. Estimation of properties associated with the Itô process from noisy observations has sparked considerable interest, see [1], [2], [8], [9].

An interesting study of the effects of noise or high frequency structure in the observations when estimating integrated volatility can be found in [1]. The initial and quite surprising characteristic of such problems is that simply subsampling the observed process reduces bias in estimation, and it is preferable to subsample rather than to use the full length of the sample. Various strategies have been proposed to improve on the simple subsampled estimator in [1]. The useful properties of subsampling for parameter estimation in multiscale problems was also shown in [10] for data generated by multiscale SDEs.

The inspirational work of Zhang et al. is focused on time-domain understanding of the process, and proposes methods of picking the optimal subsampling rate, as well as aggregating over all subsampled estimators. A final “optimal” estimator is constructed from combining this treatment with a de-biasing step, based on estimating the bias in the estimation; this bias is due to usage of the contaminated set of observations. The elegant construction of Zhang et al. incorporates all structural features of the Itô process and the noise.

Multiscale processes are in many ways more naturally treated in the frequency domain. Processes with many characteristic time scales have different characteristic properties associated with different frequencies. Fourier domain estimators of the integrated volatility have been proposed for observations devoid of microstructure features, see [11]–[13]. Fourier domain estimators have also been used for estimating noisy Itô processes, see [14]. The bias of the quadratic variation can be understood directly in the frequency domain, since the energy associated with each frequency is contaminated by the microstructure noise process. This bias is particularly damaging at high frequencies. In this article we propose a frequency-by-frequency de-biasing procedure to improve the accuracy of the estimation of the integrated volatility.

A. Sykulski and G. A. Pavliotis are with the Department of Mathematics, Imperial College London, South Kensington Campus, SW7 2AZ London, UK (adam.sykulski03@imperial.ac.uk, g.pavliotis@imperial.ac.uk), tel: +44 20 7594 8564, fax: +44 20 7594 8517.

S. C. Olhede is with the Department of Statistical Science, University College London, Gower Street, London WC1 E6BT, UK (s.olhede@ucl.ac.uk), tel: +44 20 7679 8321, fax: +44 20 7383 4703.
[14] argue that due to the multiscale properties of the observed process, a robust estimator of the integrated volatility of \( X_t \) can be obtained by using Fourier domain estimation with truncation at high frequencies and calculating the estimator from less than a full set of Fourier frequencies. Trimming (removal of frequencies) has been used for estimating the memory parameter of long-memory processes, see [15], and this procedure may suffer from a loss of efficiency in estimation depending on the degree and method of data removal, see also [16].

In our work we refrain from assuming that the high frequency structure is only present above a certain cut-off. Instead, we intend to learn from the data to what extent the high frequency structure should be shrunk when estimating the integrated volatility. This will make the method robust to different sampling rates (i.e. not only applicable to high-frequently sampled data) and different noise levels (not only heavily contaminated data). We propose to estimate the spectral contribution of the noise and the Itô processes at each frequency, modelling the multiscale nature of the data explicitly. To implement this, we introduce the concept of the multiscale ratio as an empirical measure of the energy of the Itô process compared to that of the noise at any given frequency. The multiscale ratio is estimated by solving an optimization problem derived from the Whittle likelihood, see [17]. The properties of the estimated multiscale ratio are stated in Theorem 1.

With the estimated multiscale ratio at hand, we can de-bias the empirical estimator of the spectral density and use the full set of frequencies for the estimation of the integrated volatility. We will refer to the resulting estimator of the integrated volatility as the multiscale estimator. The multiscale estimator is unbiased even when the microstructure noise is absent. We determine the properties of the multiscale estimator of the integrated volatility in Theorem 2. We also show that when the incremental process is stationary (and not merely harmonizable), the multiscale estimator has reduced variance.

To illustrate the power of the proposed multiscale estimator we perform simulation studies using the model introduced by [1]. This model consists of the Heston model with additive Gaussian white noise at each point of observation. We investigate the sample properties of the estimator under signal-to-noise scenarios and sample lengths considered by [1]. We also consider less contaminated data as well as smaller data sets. The simulation studies confirm the advantages of using the multiscale estimator. We conclude with discussing extensions and applications of the proposed methodology.

II. ESTIMATION METHODS

A. Description of the Model

In this paper we will study a regularly sampled Itô process with additive white noise superimposed upon it at each observation point \( t_i \). We denote by \( \{X_t\} \) the Itô process and by \( \{Y_{t_i}\} \) the sampled observation process. We assume that the Itô process has mean \( \mu(X) = E(X_t) \), and that the mean corrected process \( X - \mu(X) \) possesses harmonizable increments, see [18]–[20]. The additive noise is assumed to be independent of the noise that drives the Itô process. Our main objective is to estimate the integrated volatility, \( \langle X, X \rangle_T \) of the Itô process \( \{X_t\} \), from the set of observations \( \{Y_{t_i}\} \). The observations and the process are related through

\[
Y_{t_i} = X_{t_i} + \varepsilon_{t_i}, \quad i = 1, 2, \ldots, N, \quad t_i = \frac{i-1}{N} T = (i-1)\Delta t.
\]

\( \{\varepsilon_{t_i}\} \) is the white noise process with variance \( \sigma_t^2 \). In the numerical simulations of section III we shall use the Heston model to generate \( \{X_t\} \). The Heston model is specified by ( [21]):

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

where \( \nu_t = \sigma_t^2 \), and \( B_t \) and \( W_t \) are correlated 1-D Brownian motions. The correlation structure of the two processes will be specified at a later stage. The integrated volatility, which is the quantity that we are interested in estimating from observations \( \{Y_{t_i}\}_{i=1}^N \) is given by

\[
\langle X, X \rangle_T = \int_0^T \sigma_t^2 \, dt.
\]

A process \( X_t \) is harmonizable if it admits a representation in terms of the zero-mean incremental process \( dZ(f) \) of

\[
X_{t_i} - \mu_t^{(X)} = \sqrt{\Delta t} \int_{-\infty}^{\infty} dZ(f) e^{2\pi if_{t_i}},
\]

where the second order structure of \( \{dZ(f)\} \) is given by \( \text{cov}(dZ(f), dZ(f')) = S(f, f') \, df \, df' \). We shall call \( S(f, f') \) the Loève spectrum.
In the absence of market microstructure noise (i.e., when \( Y_{t_i} = X_{t_i} \)) the integrated volatility can be estimated from the quadratic variation of the process \( \{ Y_t \} \). We will see later in this section that in the presence of market microstructure noise this is no longer true and that a different estimation procedure is necessary.

### B. Naive Estimators and the Percival-Rayleigh Theorem

Let \( \{ Y_t \} \) be given by eqn (1), where the noise \( \{ \varepsilon_t \} \) is independent of \( \{ X_{t_i} \} \), is zero-mean and iid. A simple estimator of the integrated volatility of \( \{ X_t \} \) would ignore the multiscale structure of the data and use the realized volatility of the observed process. We define the naive estimator to be

\[
\langle X, X \rangle_T^{(b)} = [Y, Y]_T = \sum_{i=0}^{N-1} (Y_{t_{i+1}} - Y_{t_i})^2.
\]

This estimator is both inconsistent and biased, see [9]. For comparative reasons, define also the realized volatility of the process \( \{ X_t \} \):

\[
\langle X, X \rangle_T^{(u)} = [X, X]_T = \sum_{i=0}^{N-1} (X_{t_{i+1}} - X_{t_i})^2.
\]

This cannot be used in practice as \( \{ X_t \} \) is not directly observed.

We are interested in constructing an estimator which is consistent and unbiased in the limit as the number of observations and the length of the path that we are observing goes to infinity, whereas the distance between subsequent observations goes to zero, i.e. \( T, N \to \infty, \Delta t \to 0 \). We note that this is different from infill asymptotics (\( N \to \infty \) with \( T \) fixed) or \( N \to \infty \) for fixed \( \Delta t \) asymptotics. The study of the asymptotic limit of interest is facilitated by letting \( T = \Delta t N \) with \( \Delta t = O(N^{-\alpha}) \) where \( 0 < \alpha < 1 \). A similar device was used in [22] and in [10].

To be able to derive a Fourier domain estimator, we represent the estimator given by eqn (4) in the frequency domain. Firstly we denote the difference processes \( Z_{t_i} - Z_{t_i-1} \) by \( U_{t_i}^{(Z)} \) where \( Z = X, Y \) or \( \varepsilon \). We define the Discrete Fourier Transforms (DFTs) of \( \{ U_{t_i}^{(Z)} \} \) by:

\[
J^{(Z)}(f_k) = \frac{1}{\sqrt{N}} \sum_{j=1}^{N-1} U_{t_i}^{(Z)} e^{-2\pi i j f_k}, \quad f_k = \frac{k}{T}, \quad Z = X, Y, \varepsilon.
\]

The naive estimator can be rewritten as:

\[
\langle X, X \rangle_T^{(b)} = \sum_{i=0}^{N-1} \left| U_{t_i}^{(Y)} \right|^2 = \sum_{k=-N/2}^{N/2} \left| J^{(Y)}(f_k) \right|^2.
\]

\[
\hat{S}^{(Y)}(f_k) = \left| J^{(Y)}(f_k) \right|^2.
\]

\( \hat{S}^{(Y)}(f_k) \) is the periodogram estimator, see [23], and normally has a single argument because the covariance of two fixed frequencies is asymptotically equivalent to zero for a stationary process. The Percival-Rayleigh relationship in Eqn. (7a) is discussed in a slightly more general setting by [14]. Estimator (7) is inconsistent and biased since it is equivalent to estimator (4). Such a procedure would give an unbiased estimator of the integrated volatility only when \( \sigma^2 = 0 \).

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 the periodogram from \( \{ X_t \} \) can be distinguishable to those emanating from \( \{ \varepsilon_t \} \). We here view the periodogram of \( \{ Y_t \} \) as an estimator of the diagonal of the Loève spectrum of \( \{ X_t \} \). To be able to propose an improved estimator of the the Loève spectrum of \( X_t \) we must establish a modelling framework for \( \{ J^{(Y)}(f_k) \} \).

We shall now develop a frequency domain specification of the bias of the naive estimator. We state the following lemma.

**Lemma 1:** (Frequency Domain Bias of the Naive Estimator) Let \( \{ X_t \} \) be an Itô process with harmonizable increments that have Loève spectrum \( S^{(X)}(f, f') \). We assume that either a) \( S^{(X)}(f, f') = \hat{S}^{(X)}(f) \delta(f - f') \)
where \( \hat{S}^{(X)}(f) \in C^2_b(\mathbb{R}) \), or that b) \( S^{(X)}(f, f) \in C^3_b(\mathbb{R}) \), and \( S^{(X)}(f, f + \nu) \) has two bounded derivatives in \( \nu \) for all \( f \). Additionally assume \( S^{(X)}(f, f') \) decays sufficiently rapidly so that there is a \( r_0 > 0 \) such that \( \int_R^\infty \int_{-\infty}^\infty |S^{(X)}(f, f')|^2 \, df \, df' = \int_{-\infty}^\infty \int_R^\infty |S^{(X)}(f, f')|^2 \, df \, df' = O(1/R^2) \) for all \( r_0 < R < \infty \) (the spectral decay condition). Then the naive estimator given by eqn (7) has an expectation given by:

\[
E\left\{ \langle \hat{X}, \hat{X} \rangle_T^{(b)} \right\} = \sum_{k=-N/2}^{N/2-1} S^{(X)}(f_k, f_k) + \sigma^2_e \sum_{k=-N/2}^{N/2-1} |2\sin(\pi f_k \Delta t)|^2 + O(N^\alpha) + O(N^{1-\alpha}).
\]

**Proof:** See appendix A.

### C. Multiscale Modelling

Obviously to correct the biased estimator we need to correct the usage of \( |J^{(Y)}(f_k)|^2 \) at each frequency. We therefore define a new shrinkage estimator of \( S^{(X)}(f_k, f_k) \) by

\[
\tilde{S}^{(X)}(f_k, f_k; L_k) = L_k \hat{S}^{(Y)}(f_k, f_k).
\]

(8)

0 \leq L_k \leq 1 is referred to as the ‘multiscale ratio’ and its optimal form for perfect bias correction is given by:

\[
L_k = \frac{S^{(X)}(f_k, f_k)}{S^{(X)}(f_k, f_k) + \sigma^2_e |2\sin(\pi f_k \Delta t)|^2}.
\]

(9)

This quantity cannot be calculated without explicit knowledge of \( S^{(X)}(f_k, f_k) \) and \( \sigma^2_e \). These are two objects that we do not know. For the moment ignore this and simply note that, again constraining \( \{U_t^{(X)}\} \) to be harmonizable, see appendix A, then:

\[
E\left\{ \tilde{S}^{(X)}(f_k, f_k; L_k) \right\} = L_k E \left\{ |J^{(Y)}(f_k)|^2 \right\} = S^{(X)}(f_k, f_k) + O(N^{-1+\alpha}) + O(N^{-\alpha}),
\]

and so we can define a new estimator via:

\[
\langle \tilde{X}, \tilde{X} \rangle_T^{(m1)} = \sum_{k=-N/2}^{N/2-1} \tilde{S}^{(X)}(f_k, f_k; L_k)
\]

\[
E\left\{ \langle \tilde{X}, \tilde{X} \rangle_T^{(m1)} \right\} = \langle X, X \rangle_T + O(N^\alpha) + O(N^{1-\alpha}).
\]

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

### D. Estimation of the Multiscale Ratio

For a stationary process, a typical measure of the “energy” of the process \( \{U_t^{(X)}\} \) is its variance given by

\[
\sigma^2_X = \Delta t \int_{-\infty}^\infty S^{(X)}(f, f) \, df.
\]

(10)

\( \sigma^2_X \) is therefore the ‘average frequency contribution’ of the process. If the process is non-stationary then

\[
\frac{1}{N} \sum_{i=0}^{N-1} \text{var} \left\{ U_t^{(X)} \right\} = \Delta t \int_{-\infty}^\infty S^{(X)}(f, f) \, df + O \left( \frac{1}{N} \right),
\]

(11)

and the same interpretation still holds. In Fig 1 we plot the periodograms of the Ilhô and noise processes, \( \hat{S}^{(X)}(f_k, f_k) \), calculated from 100,000 realisations from the Heston model (for a more careful discussion of the simulation study, see section III). The spectrum is almost flat. This is very reasonable, as the integrated nature of the Ilhô process defined from eqn (2) would imply that an equal weighting is given to all frequencies for the differences process.
The noise process will in contrast have a spectrum that is far from flat, and a suitable bias correction would shrink the estimator at higher frequencies. We therefore simplify the multiscale ratio to the following:

$$L_k = \frac{\sigma_X^2}{\sigma_X^2 + \sigma_\varepsilon^2 |2 \sin(\pi f_k \Delta t)|^2}.$$ \hfill (12)

If we adjust the spectrum by $\hat{L}_k$ rather than $L_k$ then from Figure 1 it appears as if this procedure will work as well as using the true $L_k$. We cannot estimate $L_k$ without some simplifying assumptions, as the problem would be overparameterised.

We have now a two-parameter description to how the energy should be adjusted at each frequency. We now only need to determine estimators of $\sigma = (\sigma_X^2, \sigma_\varepsilon^2)$. We propose to implement the estimation using the Whittle likelihood methods (see [17] or [24]). The log-Whittle likelihood can be written as:

$$\ell(\sigma) = \sum_{j=1}^{N/2-1} \log \left[ \frac{1}{S(Y)(f_j, f_j)} e^{-\frac{\hat{S}(Y)(f_j, f_j)}{\sigma_X^2 + \sigma_\varepsilon^2 |2 \sin(\pi f_j \Delta t)|^2}} \right] + O(1)$$

$$= - \sum_{j=1}^{N/2-1} \log \left( S(X)(f_j, f_j) + \sigma_X^2 |2 \sin(\pi f_j \Delta t)|^2 \right) - \sum_{j=1}^{N/2-1} \frac{\hat{S}(Y)(f_j, f_j)}{S(X)(f_j, f_j) + \sigma_X^2 |2 \sin(\pi f_j \Delta t)|^2}.$$ \hfill (13)

If $\{U_t^{(X)}\}$ is a stationary process, then the Whittle likelihood will approximate the time-domain likelihood of the sample, under suitable regularity conditions, see [25]. If $\{U_t^{(X)}\}$ is a harmonizable process but not stationary, then as long as the total contributions of the covariance of the incremental process can be bounded using this likelihood will asymptotically (in $N$) produce suitable estimators, as we shall see from Theorem 2. We now assume that $S^{(X)}(f_j, f_j)$ is contributing with approximately equal magnitude across all frequencies, and use this assumption to determine how much the periodogram should be shrunk.

**Definition 2.1:** (Multiscale Energy Likelihood)

The multiscale energy likelihood is defined as:

$$\ell(\sigma) = - \sum_{j=1}^{N/2-1} \log \left( \sigma_X^2 + \sigma_\varepsilon^2 |2 \sin(\pi f_j \Delta t)|^2 \right) - \sum_{j=1}^{N/2-1} \frac{\hat{S}(Y)(f_j, f_j)}{\sigma_X^2 + \sigma_\varepsilon^2 |2 \sin(\pi f_j \Delta t)|^2}.$$ \hfill (13)

We stress that strictly speaking this is not a (log-)likelihood, but merely a device for determining the multiscale ratio. We then maximise this function in $\sigma$ to obtain a set of estimators $\hat{\sigma}$. We assume that the Whittle likelihood produces suitable estimators of $\sigma$, and will discuss the performance of the estimator again in the examples section.
Theorem 1: (The Estimated Multiscale Ratio)
Assume that \( \{U_i^{(X)}\} \) is a harmonizable stochastic process, and assume that \( S^{(X)}(f_1, f_2) \) satisfies either condition (a) or (b) of Lemma 1, and that there exists \( \Delta f > 0 \) such that
\[
|S^{(X)}(f_1, f_2)| \leq \frac{C}{|f_1 - f_2|^\alpha'}, \quad \alpha' > 1, \quad \forall |f_1 - f_2| > \Delta f
\]  
(14)
(the spectral correlation decay condition). Assume that \( \left\{ \left| U_{i_j}^{(Y)} \right| / \sqrt{\operatorname{var}(U_{i_j}^{(Y)})} \right\} \) is strongly mixing and \( \left\{ \left( U_{i_j}^{(Y)} - \Delta \mu_{i_j}^{(X)} \right)^2 / \operatorname{var}(U_{i_j}^{(Y)}) \right\} \) is uniformly integrable, where
\[
\frac{1}{\sqrt{\operatorname{var}(\sum_{j=1}^N U_{i_j}^{(Y)})}} \to 0, \quad \sup_N \frac{N}{\operatorname{var}(\sum_{j=1}^N U_{i_j}^{(Y)})} < \infty.
\]
Then the estimated multiscale ratio is given by:
\[
\hat{L}_j = \frac{\hat{\sigma}_X^2}{\hat{\sigma}_X^2 + \hat{\sigma}_e^2 |2 \sin(\pi f_j \Delta t)|^2},
\]  
(15)
where \( \hat{\sigma}_X^2 \) and \( \hat{\sigma}_e^2 \) maximise \( \ell(\sigma) \) given in Eqn. (13). \( \hat{L}_j \) satisfies
\[
\hat{L}_j = \frac{\hat{\sigma}_X^2}{\hat{\sigma}_X^2 + \hat{\sigma}_e^2 |2 \sin(\pi f_j \Delta t)|^2} + O \left( \frac{1}{\sqrt{N}} \right) = \mathcal{L}_j + O \left( \frac{1}{\sqrt{N}} \right),
\]  
(16)
where \( \mathcal{L}_j = \frac{\hat{\sigma}_X^2}{\hat{\sigma}_X^2 + \hat{\sigma}_e^2 |2 \sin(\pi f_j \Delta t)|^2} \). For processes \( \{U_i^{(X)}\} \) such that \( S^{(X)}(f, f) \) is not constant \( \sigma = \varsigma \) is the solution of:
\[
0 = \sum_{j=1}^{N/2-1} \frac{S^{(X)}(f_j, f_j) - \sigma_X^2}{\left( \hat{\sigma}_X^2 + \hat{\sigma}_e^2 |2 \sin(\pi f_j \Delta t)|^2 \right)^2}
\]  
(17)
\[
0 = \sum_{j=1}^{N/2-1} \frac{2 \sin(\pi f_j)^2 |S^{(X)}(f_j, f_j) - \sigma_X^2|}{\left( \hat{\sigma}_X^2 + \hat{\sigma}_e^2 |2 \sin(\pi f_j \Delta t)|^2 \right)^2},
\]  
(18)
whilst if \( S^{(X)}(f, f) \) is constant, then \( \varsigma_X^2 = S^{(X)}(f, f) \) and \( \varsigma_e^2 = \sigma_e^2 \).

Proof: See appendix B.

From eqn (17) we may note that \( \varsigma_X^2 \) is a weighted average of the spectral density of \( \{U_i^{(X)}\} \) over the range \([-1/(2\Delta t), 1/(2\Delta t)]\), with a preferential weighting to the low frequencies. Combining eqn (8) with (15) the proposed estimator of the spectral density of \( \{U_i^{(X)}\} \) becomes:
\[
\hat{S}^{(X)}(f_k, f_k; \hat{L}_k) = \hat{L}_k \left| J^{(Y)}(f_k) \right|^2,
\]  
(19)
where \( \hat{L}_k \) is given by eqn (16).

Theorem 2: (The Multiscale Estimator of the Integrated Volatility)
Assume that \( U_i^{(X)} \) is a harmonizable process and that the conditions of Lemma 1 and Theorem 1 are satisfied. The multiscale estimator of the integrated volatility defined by
\[
\langle X, X \rangle_T^{(m)} = \sum_{k=-N/2}^{N/2-1} \hat{S}^{(X)}(f_k, f_k; \hat{L}_k),
\]  
(20)
where \( \hat{S}^{(X)}(f_k, f_k; \hat{L}_k) \) is defined by eqn (19) has a mean and variance given by:

\[
E \{ \langle X, X \rangle_T^{(m)} \} = \sum_{k=-N/2}^{N/2-1} \hat{S}^{(X)}(f_k, f_k) + O \left( N \left[ \hat{\varsigma}^2 - \Delta t \int_{-1/(2\Delta t)}^{1/(2\Delta t)} S^{(X)}(f, f) \, df \right] \right) + O \left( \left( \hat{\varsigma}^2 - \sigma^2 \right) N \right) + O \left( \sqrt{N} \right) + h.o.t.
\]

\[
\text{var} \{ \langle X, X \rangle_T^{(m)} \} = \sum_{k=-N/2}^{N/2-1} \sum_{k_2=-N/2}^{N/2-1} \left[ \mathcal{L}_{k_1} \mathcal{L}_{k_2} + O \left( \frac{1}{\sqrt{N}} \right) \right]
\]

\[
\left( \left| S^{(X)}(f_{k_1}, f_{k_2}) \right|^2 + O(N^{-\alpha}) + O \left( N^{-1+\alpha} \right) \right).
\]

Proof: See appendix C.

This theorem specifies the newly introduced multiscale estimator and its properties.

Note that \( \left[ \hat{\varsigma}^2 - \Delta t \int_{-1/(2\Delta t)}^{1/(2\Delta t)} S^{(X)}(f, f) \, df \right] \) measures the variability of the Itô process compared to the nominal average over the range \([-1/(2\Delta t), 1/(2\Delta t)]\), and if the incremental process \( \{ X_{t_{i+1}} - X_{t_i} \} \) is too variable this will increase the bias in the estimation. If we can additionally assume stationarity of \( \{ U_t^{(X)} \} \) then

\[
\text{var} \{ \langle X, X \rangle_T^{(m)} \} = \sum_{k=-N/2}^{N/2-1} \left[ \mathcal{L}_{k} + O \left( \frac{1}{\sqrt{N}} \right) \right]
\]

\[
\left( \left| S^{(Y)}(f_k, f_k) \right|^2 + O(N^{-1+\alpha}) + O \left( N^{-\alpha} \right) \right)
\]

\[
< \text{var} \{ \langle X, X \rangle_T^{(b)} \} + O \left( \sqrt{N} \right) + O(N^\alpha) + O \left( N^{1-\alpha} \right),
\]

unless \( \zeta_e = 0 \). We thus note in this case that the multiscale estimator has lower variance than the naive method of moments estimator as \( 0 \leq \mathcal{L}_k \leq 1 \).

### III. Monte Carlo Studies – The Heston Model

We shall now demonstrate the performance of the multiscale estimator of the integrated volatility using the Heston model defined in eqn (2). We will use the same parameter values to the ones that were used in [1], namely \( \mu = .05, \kappa = 5, \alpha = .04, \gamma = .5 \) and the correlation coefficient between the two Brownian motions \( B \) and \( W \) is \( \rho = -.5 \). We set \( X_0 = 0 \) and \( \nu_0 = 0.04 \), which is the long time limit of the expectation of the process \( \nu_t \).

To illustrate the multiscale features of the process \( Y_t = X_t + \varepsilon_t \), \( i = 1, 2, \ldots \) defined in eqn (1) we calculate the periodogram of \( U_t^{(X)} \) and \( U_t^{(e)} \) for one simulated path, displayed in Figure 2. Here we have used the same sample length \( T \) and noise intensity \( \sigma^2 \) as in [1]: \( T = 1 \) day and \( \sigma^2 = 0.0005^2 \). The length of the sample path, \( T = 1 \) day or 23,400s with \( \Delta t = 1 \)s, corresponds to one trading day, since we take one trading day to be 6.5h long. Notice the different shape of the two periodograms. The periodogram of \( U_t^{(Y)} \) will not be distinguishable from that of \( U_t^{(e)} \) at higher frequencies, despite the moderate to low intensity of the market microstructure noise. If we observed the two components \( X_t \) and \( \varepsilon_t \) separately, then the multiscale ratio \( L_j \) could be estimated from the periodograms of \( U_t^{(X)} \) and \( U_t^{(e)} \) using the method of moments formula. In this case, we would estimate \( L_j \) by

\[
\bar{L}_j = \frac{\hat{S}^{(X)}(f_j, f_j)}{\hat{S}^{(X)}(f_j, f_j) + \hat{S}^{(e)}(f_j, f_j)}.
\]

The corresponding estimator of the integrated volatility becomes:

\[
\langle X, X \rangle_T^{(m)} = \sum_{k=-N/2}^{N/2-1} \bar{L}_k \hat{S}^{(Y)}(f_k, f_k).
\]

The estimated multiscale ratio \( \bar{L}_j \), for the Heston model with the specified parameters, is plotted in Figure 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 \( \varepsilon_t \) separately. Despite the variability of \( \bar{L}_j \) this multiscale

\[ \lim_{t \to +\infty} \mathbb{E}\nu_t = \kappa \alpha. \]
ratio when multiplied by $\hat{S}^{(Y)}(f_j, f_j)$ will remove microstructure energy from the high frequencies. Consequently, $\hat{L}_j \hat{S}^{(Y)}(f_j, f_j)$ should recover a good approximation to $\hat{S}^{(X)}(f_j, f_j)$ and hence lead to a good estimator of the integrated volatility. Figure 2 displays the estimated multiscale ratio applied to $\hat{S}^{(Y)}(f_j, f_j)$ over one path realisation. This plot suggests that the energy over the high frequencies has been shrunk and that $\tilde{L}_j \hat{S}^{(Y)}(f_j, f_j)$ is a good approximation to $\hat{S}^{(X)}(f_j, f_j)$. It therefore seems not unreasonable to assume that the summation of this function across frequencies should make a good approximation to the integrated volatility.

The parameters ($\hat{\sigma}^2_X$ and $\hat{\sigma}^2_\epsilon$) are estimated separately for each path using the MATLAB function fminsearch on eqn (13). Figure 2 shows the approximated form of $\hat{S}^{(X)}(f_j, f_j)$ and $\hat{S}^{(\epsilon)}(f_j, f_j)$ (in white) plotted over the periodograms $\hat{S}^{(X)}(f_j, f_j)$ and $\hat{S}^{(\epsilon)}(f_j, f_j)$ for one simulated path. The approximated forms of the spectral densities of $U_t^{(X)}$ and $U_t^{(\epsilon)}$ seem to approximate the expectations of their respective periodograms. These approximations should be quite similar to the averaged periodograms of Figure 1; in fact the accuracy of the new estimator depends on how consistently these spectral densities are estimated in the presence of limited information from the sampled process $Y_t$. Figure 3 shows the corresponding estimated multiscale ratio $\hat{L}_j$ (in white) from this simulated path, as defined in eqn (15). The function decays, as expected, so that it will remove the high-frequency noise microstructure in the spectrum of $Y_t$; the ratio is also a good approximation of $L_j$. Figure 2 shows $\hat{L}_j \hat{S}^{(Y)}(f_j, f_j)$, which is again

Fig. 2. The periodogram of a realisation of $U_t^{(X)}$ (top left), a realisation of $U_t^{(\epsilon)}$ (top right) with the Whittle estimates superimposed and of two biased corrected estimators of the periodogram of $U_t^{(X)}$, using $\tilde{L}_j$ (bottom left) and $\hat{L}_j$ (bottom right). Notice the different scales in the four figures.
Fig. 3. The method of moments estimate $\tilde{L}_j$ from a single realisation, with the Whittle estimate (white line) of $L_j$ superimposed.

![Graph showing method of moments estimate](image)

**TABLE I**

|                | Sample bias | Sample variance | Sample RMSE |
|----------------|-------------|-----------------|-------------|
| $\langle X_t, X_t \rangle_T^{(b)}$ | $1.17 \times 10^{-2}$ | $1.79 \times 10^{-8}$ | $1.17 \times 10^{-2}$ |
| $\langle X_t, X_t \rangle_T^{(m_1)}$ | $7.30 \times 10^{-7}$ | $2.78 \times 10^{-10}$ | $1.67 \times 10^{-5}$ |
| $\langle X_t, X_t \rangle_T^{(m_2)}$ | $2.04 \times 10^{-7}$ | $2.46 \times 10^{-10}$ | $1.62 \times 10^{-5}$ |
| $\langle X_t, X_t \rangle_T^{(u)}$ | $2.59 \times 10^{-9}$ | $2.09 \times 10^{-10}$ | $1.45 \times 10^{-5}$ |
| $\langle X_t, X_t \rangle_T^{(m_2)}$ | $3.02 \times 10^{-9}$ | $2.09 \times 10^{-10}$ | $1.44 \times 10^{-5}$ |

**Simulation study comparing the new estimator with the best estimator of [1].**

similar to $\hat{S}(X)(f_j, f_j)$. It would appear that the new estimator has successfully removed the microstructure effect from each frequency.

It is worth noting here that the ratios $L_j$ and $\tilde{L}_j$ quantify the effect of the multiscale structure of the process. If $\sigma^2_\varepsilon$ 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). Therefore the parametric model is only used to estimate $L_j$, and we use the periodogram of $U_t^{(X)}$ to estimate the integrated volatility of $X_t$. So in the case of zero microstructure noise, the estimate would recover the periodogram of the $X_t$ difference process, and so the estimate of the integrated volatility would simply be the realized volatility of the observable process.

### A. Simulation Results

In this section we investigate the performance of the proposed estimator using Monte Carlo simulations. In this study 100,000 simulated paths are generated. For each individual path the periodogram of $U_t^{(X)}$ is estimated by shrinking the periodogram of $U_t^{(Y)}$ (as discussed in the previous subsection) and the periodogram is aggregated over frequencies to provide an estimator of the integrated volatility for each path.

Table I displays the results of our simulation, where biases, variances and errors are calculated using a Riemann sum approximation of the integral

$$\langle X_t, X_t \rangle = \int_0^T \sigma_t^2 \, dt.$$ (24)

The performance of the estimators of the integrated volatility calculated by aggregation are reported in Table I. The two estimators $\langle X, X \rangle_T^{(u)}$ and $\langle X, X \rangle_T^{(m_2)}$ (see equations (5) and (23) 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 [1] (denoted by $\langle X, X \rangle^{(s)}_T$) 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, $\langle X, X \rangle^{(b)}_T$, given in eqn (4) (the fifth-best estimator in [1]).

The table shows that the new estimator, $\langle X, X \rangle^{(m1)}_T$, is competitive with the first-best approach in [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 approach; both estimators involve a bias-correction, [1] perform this globally by weighting different sampling frequencies, whilst we correct locally at each frequency. The realized volatility of $Y_t$ at the highest frequency, produces disastrous results, as expected.

A histogram of the observed bias of the new estimator is plotted in Figure 4 along with a histogram of the observed bias of the first best estimator in [1]. The observed bias of our estimator follows a Gaussian distribution centred at zero, suggesting that this estimator is unbiased. 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 RMSE).
The new estimator requires calculation of the parameters $\tilde{\sigma}_X^2$ and $\tilde{\sigma}_\varepsilon^2$ which will vary over each process due to the limited information given from the $Y_t$ process. The stability of this approximation is of great importance if the estimator is to perform well. Figure 5 shows the distribution of the parameters $\tilde{\sigma}_X^2$ and $\tilde{\sigma}_\varepsilon^2$ over the 100,000 simulated paths. The parameter estimation is quite consistent, with all values estimated within a narrow range. Figure 1 suggests that these estimates are roughly unbiased; as $\sigma_X^2 \approx 6.8 \times 10^{-9}$ and $\sigma_\varepsilon^2 \approx 2.5 \times 10^{-7}$ (as $\sigma_\varepsilon^2 |2\sin(\pi f_j)|^2 \approx 1 \times 10^{-6}$, at $f_j = 0.5$).

### B. Comparing estimators over shorter sample lengths

This section compares our new estimator and the first-best estimator by [1] for a shorter sample length. A shorter sample length will increase the variance of the multiscale ratio ($\tilde{\sigma}_\varepsilon^2$) and the corresponding plot of $\tilde{\sigma}_X^2$. Figure 6 also shows the corresponding estimate of the multiscale ratio $\tilde{L}_j$ (in white) along with the periodograms $\tilde{S}^{(X)}(f_j, f_j)$ and $\tilde{S}^{(\varepsilon)}(f_j, f_j)$ for one simulated path. The estimator still approximates the energy structure of the processes accurately. Figure 6 also shows the corresponding estimate of the multiscale ratio $\tilde{L}_j$ (in white) from this simulated path (together with $\tilde{L}_j$ and the corresponding plot of $\tilde{L}_j S^{(X)}(f_j)$). The new estimator has removed the microstructure noise effect and has formed a good approximation of $\tilde{S}_{U(X)}(f_j)$. The approximation of the spectral densities is still accurate despite the lack of available data.

Table II displays the accuracy of the estimators over the 100,000 simulated paths. The first-best estimator by [1] and the new estimator are once again comparable in performance and both estimates are close to the best attainable RMSE given by the realized volatility on $X_t$.

### C. Comparing estimators with a low-noise process

This section compares the new estimator and the first-best estimator by [1] for smaller levels of microstructure noise. Reducing the microstructure noise will reduce the need to subsample. The first best estimator by [1] will have a higher sampling frequency and the new estimator will reduce its estimate of $\tilde{\sigma}_\varepsilon^2$ accordingly. However, for very small levels of noise, 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 the noise is reduced by a factor of 10, ie. $E(\varepsilon^2) = 5 \times 10^{-5}$. Note that the path length is kept at its original length of $T = 1$ day.

It is interesting once again to see whether the methods developed still model each process accurately. Figure 7 shows the estimates of $\tilde{\sigma}_X^2$ and $\tilde{\sigma}_\varepsilon^2 |\sin(\pi t f_j)|^2$ (in white) along with the periodograms $\tilde{S}^{(X)}(f_j, f_j)$ and $\tilde{S}^{(\varepsilon)}(f_j, f_j)$ for one simulated path along with the corresponding estimate of the multiscale ratio $\tilde{L}_j$ (in white) (plotted over the approximated $\tilde{L}_j$ and the corresponding plot of $\tilde{L}_j S^{(X)}(f_j)$).

The estimation 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 $L_j$ to be more tempered across the high frequencies than it was before, due to the smaller microstructure noise. However, the new estimator has still detected

| $\{X, X\}^{(n)}_T$ | $\{X, X\}^{(n)}_{T1}$ | $\{X, X\}^{(m)}_{T}$ | $\{X, X\}^{(m)}_{T1}$ |
|---------------------|---------------------|---------------------|---------------------|
| $1.17 \times 10^{-3}$ | $6.84 \times 10^{-7}$ | $2.54 \times 10^{-7}$ | $8.03 \times 10^{-9}$ |
| $1.78 \times 10^{-9}$ | $7.34 \times 10^{-12}$ | $6.50 \times 10^{-12}$ | $5.03 \times 10^{-13}$ |
| $1.17 \times 10^{-3}$ | $2.79 \times 10^{-6}$ | $2.56 \times 10^{-6}$ | $7.08 \times 10^{-7}$ |
| $1.17 \times 10^{-3}$ | $2.79 \times 10^{-6}$ | $2.56 \times 10^{-6}$ | $7.08 \times 10^{-7}$ |

**TABLE II**

**SIMULATION STUDY FOR SHORTER SAMPLER LENGTH.**
the small levels of noise in the data. Table III reports on the results of 25,000 simulations performed as before. The first-best estimator of [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 by [1]) (denoted by \( \hat{\langle X, X \rangle} (s^2) \)) was effective; this is simply an estimator that averages estimates calculated from sub-sampled paths at different starting points. The new estimator, was remarkably robust, with RMSE very close to the RMSE of estimators based on the \( X_t \) process. The difference between the estimators using \( Y_t \) and the estimators using \( X_t \) is expected to become smaller with little microstructure noise and this can be seen by the similar order RMSE errors between all estimators; however the new estimator was much closer in performance to the realized 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.

IV. CONCLUSIONS

The problem of estimating the integrated volatility of an Itô process from noisy observations was studied in this paper. It is well known that the presence of market microstructure noise in the model renders the problem of parameter estimation quite subtle, since the standard, naive method of moments estimator is asymptotically biased.
TABLE III
SIMULATION STUDY FOR LOWER MARKET MICROSTRUCTURE NOISE.

|         | Sample bias  | Sample variance | Sample RMSE  |
|---------|--------------|-----------------|--------------|
| $\langle X, X \rangle_T^{(b)}$ | $1.17 \times 10^{-4}$ | $2.11 \times 10^{-10}$ | $1.18 \times 10^{-4}$ |
| $\langle X, X \rangle_T^{(m_1)}$ | $4.21 \times 10^{-6}$ | $2.51 \times 10^{-10}$ | $1.64 \times 10^{-5}$ |
| $\langle X, X \rangle_T^{(m_2)}$ | $2.43 \times 10^{-8}$ | $2.12 \times 10^{-10}$ | $1.46 \times 10^{-5}$ |
| $\langle X, X \rangle_T^{(u)}$ | $1.57 \times 10^{-8}$ | $2.07 \times 10^{-10}$ | $1.44 \times 10^{-5}$ |

Fig. 7. The periodogram of a realisation of $U_t^{(X)}$ (top left), of a realisation of $U_t^{(e)}$ (top right) with the Whittle estimates superimposed, the estimate of $L_j$ (bottom left) with the Whittle estimate of $L_j$ superimposed and the biased corrected estimator of the periodogram of $U_t^{(X)}$ (bottom right), using $\hat{L}_j$. Notice the different scales in the four figures.
Unlike previous works on this problem, see [1], [10], the method for estimating the integrated volatility developed in this paper is based on the frequency domain representation of the Itô process and of 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 model the level of the market microstructure noise parametrically at all frequencies, and 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 multiple scales we used the Whittle likelihood, and quantified the noise contribution by the multiscale ratio. Various properties of the multiscale estimator were determined, see Theorems 1 and 2. As was illustrated by the set of examples, our estimator performs extremely well on data simulated from the Heston model, and is competitive with the methods proposed by [1], 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 quick.

Frequency domain inference is 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. The noise in some applications is correlated and the process of interest might be a Lévy rather than an Itô process, see also [14], [26]. Extensions of frequency domain inference to such scenarios are currently under investigation.

REFERENCES

[1] L. Zhang, P. A. Mykland, and Y. Ait-Sahalia, “A tale of two time scales: Determining integrated volatility with noisy high-frequency data,” J. Am. Stat. Assoc., vol. 100, pp. 1394–1411, 2005.
[2] R. S. Tsay, Analysis of Financial Time Series. New York, USA: John Wiley & Sons, 2005.
[3] A. Majda, I. Timofeyev, and E. V. Eijnden, “A mathematical framework for stochastic climate models,” Comm. Pure Appl. Math., vol. 54, no. 8, pp. 891–974, 2001.
[4] J. M. Lilly and G.-C. Gascard, “Wavelet ridge diagnosis of time-varying elliptical signals with application to an oceanic eddy,” Nonlinear Processes in Geophysics, vol. 13, pp. 467–483, 2006.
[5] E. Meerbach, E. Dittmer, I. Horenko, and C. Schütte, “Multiscale modelling in molecular dynamics: Biomolecular conformational ensembles at metastable states,” in Computer Simulations in Condensed Matter Systems: From Materials to Chemical Biology, M. Ferrario, G. Cicotti, and K. Binder, Eds. Springer, 2006, pp. 475–497.
[6] T. O. F. Conrad, A. Leichtle, A. Hagehn, E. Diederichs, S. Baumann, J. Thiery, and C. Schütte, “Beating the noise: New statistical methods for detecting signals in maldi-tof spectra below noise level,” in Computational Life Sciences II, M. R. Berthold, R. Glen, , and I. Fischer, Eds. Springer, 2006, pp. 119–128.
[7] G. Pavliotis and A. Stuart, Multiscale Methods: Homogenization and Averaging. New York: Springer, 2008.
[8] J.-P. Fouque, G. Papanicolaou, K. R. Sircar, and K. Solna, “Short time-scale in SP-500 volatility,” J. Comp. Finance, vol. 6, pp. 1–23, 2003.
[9] P. R. Hansen and A. Lunde, “Realized variance and market microstructure noise,” J. Business & Economic Statistics, vol. 24, pp. 127–161, 2006.
[10] G. A. Pavliotis and A. M. Stuart, “Parameter estimation for multiscale diffusions,” J. Stat. Phys., vol. 127, pp. 741–781, 2007.
[11] P. R. Hansen and A. Lunde, “A forecast comparison of volatility models: does anything beat a GARCH(1,1)?” Journal of Applied Econometrics, vol. 20, pp. 873–889, 2005.
[12] E. Barucci and R. Reno, “On measuring volatility of diffusion processes with high frequency data,” Economics Letters, vol. 74, pp. 371–378, 2002.
[13] P. Malliavin and M. E. Mancino, “Fourier series for measurements of multivariate volatilities,” Finance and Stochastics, vol. 6, pp. 49–61, 2002.
[14] M. E. Mancino and S. Sanfelici, “Robustness of fourier estimators of integrated volatility in the presence of microstructure noise,” University of Firenze, Tech. Rep., 2006.
[15] P. M. Robinson, “Gaussian semiparametric estimation of long range dependence,” Annals of Statistics, vol. 23, pp. 1630–1661, 1995.
[16] C. M. Hurvich, R. Deo, and J. Brodsky, “The mean square error of Geweke and Porter-Hudak’s estimator of long memory time series,” Journal of Time Series Analysis, vol. 19, pp. 19–46, 1998.
[17] J. Beran, Statistics for long-memory Processes. Chapman and Hall, London, 1994.
[18] M. Loève, Probability theory. I, 4th ed. New York: Springer-Verlag, 1977, graduate Texts in Mathematics, Vol. 45.
[19] ———, Probability theory, II, 4th ed. New York: Springer-Verlag, 1978, graduate Texts in Mathematics, Vol. 46.
[20] T. Kawata, “Fourier analysis of nonstationary stochastic processes,” Transactions of the American Mathematical Society, vol. 118, pp. 276–302, 1965.
[21] Heston, “A closed form solution for options with stochastic volatility with applications to bond and currency options,” Review of Financial Studies, vol. 6, pp. 327–343, 1993.
[22] A. G. Constantine and P. Hall, “Characterizing surface smoothness via estimation of effective fractal dimensions,” J. Royal Statistical Society, vol. 56, pp. 97–113, 1994.
[23] D. Brillinger, Time series: data analysis and theory. Philadelphia, USA: Society for Industrial and Applied Mathematics, 2001.
where \( \{dZ(Z)(f)\} \) is a complex proper process (see [27]) with expectation zero and covariance \( \text{cov} \{dZ(Z)(f),dZ(Z)(f')\} = S(Z)(f,f')dfdf' \). Recall that \( S(Z)(f,f') \) is the Loève spectrum of \( \{U_t(Z)\} \). We know that the noise process admits representation:

\[
U_{t_j}^{(e)} = \sqrt{\Delta t} \int_{-\frac{1}{2}(2\Delta t)}^{\frac{1}{2}(2\Delta t)} dZ^{(e)}(f) e^{2\pi i f t_j} = \sqrt{\Delta t} \int_{-\frac{1}{2}(2\Delta t)}^{\frac{1}{2}(2\Delta t)} dZ^{(e)}(f) \left[1 - e^{-2\pi i f} \right] e^{2\pi i f t_j},
\]

where \( \{dZ^{(e)}(f)\} \) is an orthogonal incremental process with variance \( \sigma^2_e \). Thus \( dZ^{(e)}(f) \) has variance \( \sigma^2_e [2 \sin(\pi f \Delta t)]^2 \), a quantity increasing with \( f \in [-1/(2\Delta t),1/(2\Delta t)] \).

\[
U_{t_j}^{(Y)} = \mu_{t_j}^{(X)} - \mu_{t_{j-1}}^{(X)} + \sqrt{\Delta t} \int_{-\infty}^{\infty} [dZ^{(X)}(f)dZ^{(e)}(f) \mathbf{I}(f \in [-1/(2\Delta t),1/(2\Delta t)])] e^{2\pi i f t_j},
\]

\[
= \mu_{t_j}^{(X)} - \mu_{t_{j-1}}^{(X)} + \sqrt{\Delta t} \int_{-\infty}^{\infty} dZ^{(Y)}(f) e^{2\pi i f t_j},
\]

\[
t_j \in [0,T].
\]

For a harmonizable process \( \{Z_k\} \) with \( f_k = k/T, \ k = -N/2, \ldots, N/2, \ N \) even:

\[
\mathbb{E} \left\{ |r^{(Z)}(f_k)|^2 \right\} = \frac{\Delta t}{N} \mathbb{E} \left\{ \left| \sum_{j=0}^{N-1} U_{t_j}^{(Z)} e^{-2\pi i f_k t_j} \right|^2 \right\}
\]

\[
= \frac{\Delta t}{N} \mathbb{E} \left\{ \left| \sum_{j=0}^{N-1} \int_{-\infty}^{\infty} dZ^{(Z)}(f) e^{2\pi i f t_j} e^{-2\pi i f_k t_j} \right|^2 \right\}
\]

\[
= \frac{\Delta t}{N} \mathbb{E} \left\{ \left| \int_{-\infty}^{\infty} dZ^{(Z)}(f) e^{2\pi i f t_j} \frac{2\sin(\pi (f - f_k)(N-1))}{\sin(\pi f t_j - f_k)} \right|^2 \right\}
\]

\[
= \frac{\Delta t}{N} \mathbb{E} \left\{ \left| \int_{-\infty}^{\infty} dZ^{(Z)}(f) e^{i\pi (f_1 - f_2)(N-1)} \frac{\sin(\pi (f_1 - f_2)(N-1))}{\sin(\pi (f_1 - f_2))} \right|^2 \right\}
\]

\[
\int_{-\infty}^{\infty} dZ^{(Z)}(f_1) e^{i\pi (f_1 - f_2)(N-1)} \Delta t \sin(\pi (f_1 - f_2)(N-1))
\]

\[
= \frac{\Delta t}{N} \int_{-\infty}^{\infty} g^{(Z)}(f_1, f_2) e^{i\pi (f_1 - f_2)(N-1)} \Delta t \sin(\pi (f_1 - f_2)(N-1))
\]

\[
= \frac{\Delta t}{N} \int_{-\infty}^{\infty} g^{(Z)}(f_1, f_2) e^{i\pi (f_1 - f_2)(N-1)} \Delta t \sin(\pi (f_1 - f_2)(N-1))
\]

\[
= \frac{\Delta t}{N} \int_{-\infty}^{\infty} g^{(Z)}(f_1, f_2) e^{i\pi (f_1 - f_2)(N-1)} \Delta t \sin(\pi (f_1 - f_2)(N-1))
\]

\[
= \frac{\Delta t}{N} \int_{-\infty}^{\infty} g^{(Z)}(f_1, f_2) e^{i\pi (f_1 - f_2)(N-1)} \Delta t \sin(\pi (f_1 - f_2)(N-1))
\]

\[
= \frac{\Delta t}{N} \int_{-\infty}^{\infty} g^{(Z)}(f_1, f_2) e^{i\pi (f_1 - f_2)(N-1)} \Delta t \sin(\pi (f_1 - f_2)(N-1))
\]

\[
= \frac{\Delta t}{N} \int_{-\infty}^{\infty} g^{(Z)}(f_1, f_2) e^{i\pi (f_1 - f_2)(N-1)} \Delta t \sin(\pi (f_1 - f_2)(N-1))
\]

\[
= \frac{\Delta t}{N} \int_{-\infty}^{\infty} g^{(Z)}(f_1, f_2) e^{i\pi (f_1 - f_2)(N-1)} \Delta t \sin(\pi (f_1 - f_2)(N-1))
\]
If the process is stationary and \( \hat{S}(Z)(f) \) is differentiable where both derivatives are assumed bounded, and the process satisfies the spectral decay condition:

\[
E \left\{ \left| J^{(Z)}(f_k) \right|^2 \right\} = \frac{\Delta t}{N} \int_{-\infty}^{\infty} S^{(Z)}(f) \sin^2(\pi \Delta t (f - f_k)(N - 1)) \sin(\pi \Delta t (f - f_k)) df_k
\]

\[
= \frac{\Delta t}{N} \int_{-1/2\Delta t}^{1/2\Delta t} S^{(Z)}(f) \frac{\sin^2(\pi \Delta t (f - f_k)(N - 1))}{\sin^2(\pi \Delta t (f - f_k))} + O(N^{-\alpha})
\]

\[
= \Delta t \int_{-(1/2 + f_k \Delta t)(N-1)}^{(1/2 - f_k \Delta t)(N-1)} \frac{\xi}{\Delta t(N-1)} + f_k \frac{\sin^2(\pi \xi)}{\sin^2(\pi \xi/\Delta t(N-1))} \]

\[
\times \frac{d\xi}{(N-1)\Delta t} + O(N^{-\alpha})
\]

\[
= \hat{S}(Z)(f_k) + O(N^{-\alpha}) + O(N^{-1+\alpha}).
\]

If the process \( \{U^{(Z)}_n\} \) is harmonizable we denote the dual frequency spectrum \( S_d^{(Z)}(f, \nu) \) and note that \( S^{(Z)}(f_1, f_2) = S_d^{(Z)}(f_1, f_1 - f_2) \), we can therefore rewrite the integral as:

\[
E \left\{ \left| J^{(Z)}(f_k) \right|^2 \right\} = \frac{\Delta t}{N} \int_{-\infty}^{\infty} S_d^{(Z)}(f_1 + f_k, f_2 + f_k) e^{i\pi (f_1 - f_2)(N-1)\Delta t}
\]

\[
\frac{\sin(\pi \Delta t f_1 N - 1)}{\sin(\pi \Delta t f_1)} \frac{\sin(\pi \Delta t f_2(N - 1))}{\sin(\pi \Delta t f_2)} df_1 df_2
\]

\[
= \Delta t \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} S_d^{(Z)}(f_1 + f_k, f_1 - f_2) e^{i\pi (f_1 - f_2)(N-1)\Delta t}
\]

\[
\frac{\sin(\pi \Delta t f_1 N - 1)}{\sin(\pi \Delta t f_1)} \frac{\sin(\pi \Delta t f_2(N - 1))}{\sin(\pi \Delta t f_2)} df_1 df_2.
\]

We implement a change of variables \( f_1 - f_2 = u_1 \) and then

\[
E \left\{ \left| J^{(Z)}(f_k) \right|^2 \right\} = \Delta t \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} S_d^{(Z)}(f_1 + f_k, u_1) e^{i\pi u_1(N-1)\Delta t}
\]

\[
\frac{\sin(\pi \Delta t f_1 [N - 1])}{\sin(\pi \Delta t f_1)} \frac{\sin(\pi \Delta t (f_1 - u_1)(N - 1))}{\sin(\pi \Delta t (f_1 - u_1))} df_1 du_1
\]

\[
= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} S_d^{(Z)}(f_1 N \Delta t + f_k, u_1) e^{i\pi u_1(N-1)\Delta t}
\]

\[
\frac{\sin(\pi \Delta t \frac{f_1}{N \Delta t} [N - 1])}{\sin(\pi \Delta t \frac{f_1}{N \Delta t})} \frac{\sin(\pi \Delta t \left(\frac{f_1}{N \Delta t} - u_1\right)(N - 1))}{\sin(\pi \Delta t \left(\frac{f_1}{N \Delta t} - u_1\right))} \frac{df_1}{N^2} du_1.
\]

Because of the spectral decay condition we can approximate the above double integral by

\[
E \left\{ \left| J^{(Z)}(f_k) \right|^2 \right\} = \int_{-\infty}^{\infty} S_d^{(Z)}(f_k, u_1) e^{i\pi u_1(N-1)\Delta t} D_{N,\Delta t}(u_1) du_1 + O(N^{-\alpha}) + O(N^{-1+\alpha})
\]

defining

\[
D_{N,\Delta t}(u_1) = \frac{1}{N^2} \int_{-N/\Delta t}^{N/\Delta t} \frac{\sin(\pi f_1)}{\sin(\pi f_1/N)} \frac{\sin(\pi [f_1 - u_1 N \Delta t]/N)}{\sin(\pi [f_1 - u_1 N \Delta t]/N)} df_1
\]

\[
= \begin{cases} 
1 & \text{if } u = 0 \\
O(1/N) & \text{if } u = N^{\alpha''-1}
\end{cases}
\]

for some \( \alpha'' > 0 \). Given the limiting behaviour of \( D_{N,\Delta t}(u_1) \) and the smoothness of \( S^{(Z)}(f, f + \nu) \) in \( \nu \) we deduce that:

\[
E \left\{ \left| J^{(Z)}(f_k) \right|^2 \right\} = S_d^{(Z)}(f_k, 0) + O(N^{-\alpha}) + O(N^{-1+\alpha})
\]

\[
\equiv S^{(Z)}(f_k, f_k) + O(N^{-\alpha}) + O(N^{-1+\alpha}).
\]
An extension of this calculation allows us to note that:

\[
\mathbb{E}\left\{ J(Z)(f_{k_1})J(Z)(f_{k_2}) \right\} = e^{2\pi(f_{k_1} - f_{k_2})(N-1)\Delta t} S(Z)(f_{k_1}, f_{k_2}) + O( N^{-\alpha}) + O( N^{-1+\alpha}),
\]

for \( k_1 \) and \( k_2 \) sufficiently close to each other. We may therefore note that:

\[
\mathbb{E}\left\{ |J(f_k)|^2 \right\} = S(X)(f_k, f_k) + \sigma^2 \left( 2 \sin(\pi f_k \Delta t) \right)^2 + O( N^{-\alpha}) + O( N^{-1+\alpha})
\]

Combining these sets of relationships then yields the desired result.

\[ \]

**B. PROOF OF THEOREM 1**

We differential the multiscale energy likelihood function (13) to obtain

\[
\hat{\ell}_X(\sigma) = \frac{\partial \ell(\sigma)}{\partial \sigma_X} = - \sum_{j=1}^{N/2-1} \frac{1}{\sigma_X^2 + \sigma_e^2} \left( 2 \sin(\pi f_j \Delta t) \right)^2 + \sum_{j=1}^{N/2-1} \frac{\hat{S}(Y)(f_j, f_j)}{\sigma_X^2 + \sigma_e^2} \left( 2 \sin(\pi f_j \Delta t) \right)^2.
\]

Furthermore we note:

\[
\hat{\ell}_e(\sigma) = \frac{\partial \ell(\sigma)}{\partial \sigma_e^2}
\]

\[
= - \sum_{j=1}^{N/2-1} \frac{\left( 2 \sin(\pi f_j \Delta t) \right)^2}{\sigma_X^2 + \sigma_e^2} + \sum_{j=1}^{N/2-1} \frac{\left( 2 \sin(\pi f_j \Delta t) \right)^2 \hat{S}(Y)(f_j, f_j)}{\sigma_X^2 + \sigma_e^2}.
\]

For processes \( \{U_t^{(X)}\} \) such that \( S(X)(f, f) \) is not constant let \( \sigma = \varsigma \) be the solution of:

\[
\sum_{j=1}^{N/2-1} \frac{1}{\sigma_X^2 + \sigma_e^2} = 0 = \sum_{j=1}^{N/2-1} \frac{\left( 2 \sin(\pi f_j \Delta t) \right)^2}{\sigma_X^2 + \sigma_e^2}
\]

\[
\sum_{j=1}^{N/2-1} \frac{\left( 2 \sin(\pi f_j \Delta t) \right)^2}{\sigma_X^2 + \sigma_e^2} = 0 = \sum_{j=1}^{N/2-1} \frac{\left( 2 \sin(\pi f_j \Delta t) \right)^2 \hat{S}(Y)(f_j, f_j)}{\sigma_X^2 + \sigma_e^2}.
\]

If \( S(X)(f, f) \) is constant, then \( \varsigma_X^2 = S(X)(f, f) \) and \( \varsigma_e^2 = \sigma_e^2. \) To show large sample properties of the estimated average energies we Taylor expand the multiscale likelihood:

\[
\hat{\ell}_X(\sigma) = \hat{\ell}_X(\varsigma) + \hat{\ell}_{XX}(\sigma') \left[ \sigma_X^2 - \varsigma_X^2 \right] + \hat{\ell}_{Xe}(\sigma') \left[ \sigma_e^2 - \varsigma_e^2 \right]
\]

\[
\hat{\ell}_e(\sigma) = \hat{\ell}_e(\varsigma) + \hat{\ell}_{eX}(\sigma') \left[ \sigma_X^2 - \varsigma_X^2 \right] + \hat{\ell}_{ee}(\sigma') \left[ \sigma_e^2 - \varsigma_e^2 \right],
\]

for \( \sigma' \) lying in a ball centred at \( \sigma \) at most a distance of \( ||\sigma - \varsigma|| \) away. We therefore note with \( F = \left[ \hat{\ell}_{XX}(\sigma'); \hat{\ell}_{Xe}(\sigma'); \hat{\ell}_{ee}(\sigma') \right] \) that

\[
\left( \sigma_X^2 - \varsigma_X^2 \sigma_e^2 - \varsigma_e^2 \right) = F^{-1} \left( \hat{\ell}_X(\sigma) - \hat{\ell}_X(\varsigma) \right).
\]

\[ \]
Furthermore to leading order:

\[
\text{var} \left( \hat{\ell}_X(\sigma) \right) = \sum_{l=1}^{N/2-1} \sum_{j=1}^{N/2-1} \left( \sigma_X^2 + \sigma_s^2 \right) \left| \sin(\pi f_l \Delta t) \right|^2 \left( \sigma_X^2 + \sigma_s^2 \right) \left| \sin(\pi f_l \Delta t) \right|^2 \quad + O(1)
\]

\[
\text{var} \left( \hat{\ell}_x(\sigma) \right) = \sum_{l=1}^{N/2-1} \sum_{j=1}^{N/2-1} \left( \sigma_X^2 + \sigma_s^2 \right) \left| \sin(\pi f_l \Delta t) \right|^2 \left( \sigma_X^2 + \sigma_s^2 \right) \left| \sin(\pi f_l \Delta t) \right|^2 \quad + O(1)
\]

\[
\text{cov} \left( \hat{\ell}_X(\sigma), \hat{\ell}_x(\sigma) \right) = \sum_{l=1}^{N/2-1} \sum_{j=1}^{N/2-1} \left( \sigma_X^2 + \sigma_s^2 \right) \left| \sin(\pi f_l \Delta t) \right|^2 \left( \sigma_X^2 + \sigma_s^2 \right) \left| \sin(\pi f_l \Delta t) \right|^2 \quad + O(1)
\]

We note that \( J^{(Y)}(f_k) \) is asymptotically Gaussian if we assume \( \{ U_t^{(Y)}(t) - \Delta \mu_{t_j}(X) \} / \sqrt{\text{var}(U_t^{(Y)})} \) is strongly mixing and \( \{ (U_t^{(Y)}(t) - \Delta \mu_{t_j}(X)) / \sqrt{\text{var}(U_t^{(Y)})} \} \) is uniformly integrable, with suitable conditions on the decay of \( \text{var}(U_t^{(Y)}) \), see for example [28]. Whilst this does not produce any results for finite \( N \), we use this to argue that \( J^{(Y)}(f_k) \) will to all intent and purposes be nearly Gaussian for sufficiently large \( N \). We may then utilize Isserlis’ formula (see [29]) to determine the covariance of the periodogram:

\[
\text{cov} \left( \hat{S}(Z)(f_j, f_i), \hat{S}(Z)(f_k, f_k) \right) = \text{cov} \left( \left| J^{(Z)}(f_j) \right|^2, \left| J^{(Z)}(f_k) \right|^2 \right)
\]

\[
= \text{cov} \left( \left| J^{(Z)}(f_j) \right|^2, \left| J^{(Z)}(f_k) \right|^2 \right) - \text{E} \left( \left| J^{(Z)}(f_j) \right|^2 \right) \text{E} \left( \left| J^{(Z)}(f_k) \right|^2 \right)
\]

\[
= \text{E} \left( J^{(Z)}(f_j) J^{(Z)*}(f_k) \right) \text{E} \left( J^{(Z)*}(f_j) J^{(Z)}(f_k) \right)
\]

\[
= \left| S^{(Z)}(f_j, f_k) \right|^2 + O \left( N^{-\alpha} \right) + O \left( N^{-1+\alpha} \right),
\]

using eqn (A-27). This procedure may seem to discount potential error terms in the Gaussian approximation, where we possess no rates of convergence. To avoid this issue we could constrain \( \text{cov} \left( \hat{S}(Z)(f_j, f_i), \hat{S}(Z)(f_k, f_k) \right) \) by modelling the polispectra of \( \{ U_t^{(X)}(t) \} \) directly, see [30], and constraining this decay similarly to the spectral correlation decay condition. For simplicity we have have kept the intuitive reasoning as to the validity of the theorem. Thus it in fact transpires that:

\[
\text{var} \left( \hat{\ell}_X(\sigma) \right) = \sum_{l=1}^{N/2-1} \sum_{j=1}^{N/2-1} \left( \sigma_X^2 + \sigma_s^2 \right) \left| \sin(\pi f_l \Delta t) \right|^2 \left( \sigma_X^2 + \sigma_s^2 \right) \left| \sin(\pi f_l \Delta t) \right|^2 \quad + O(1)
\]

\[
= 2 \sum_{\tau=1}^{N/2-1} \sum_{k=1}^{N/2-\tau} \left( \sigma_X^2 + \sigma_s^2 \right) \left| \sin(\pi f_k \Delta t) \right|^2 \left( \sigma_X^2 + \sigma_s^2 \right) \left| \sin(\pi f_k \Delta t) \right|^2 \quad + O(1)
\]

\[
= O(N) + R_1,
\]

where we may bound \( R_1 \) by:

\[
R_1 = 2 \sum_{\tau=R}^{N/2-\tau} \sum_{k=1}^{N/2-\tau} \left( \sigma_X^2 + \sigma_s^2 \right) \left| \sin(\pi f_k \Delta t) \right|^2 \left( \sigma_X^2 + \sigma_s^2 \right) \left| \sin(\pi f_k \Delta t) \right|^2 \quad + O(1)
\]

\[
\leq 2 \sum_{\tau=R}^{N/2-\tau} \sum_{k=1}^{N/2-\tau} \frac{C^2}{\sigma_X^2 \left| \tau/T \right|^{2\alpha}} + O(1)
\]

\[
= 2 \sum_{\tau=R}^{N/2-1} (N/2 - \tau) \frac{C^2}{\sigma_X^2 \left| \tau/T \right|^{2\alpha}} + O(1)
\]

\[
= 2TC^2T^{1-2\alpha} = O(N(2-2\alpha)(1-\alpha)),
\]
and we therefore need to assume $\alpha > 1$, to make the terms in $R_1$ become negligible. Similarly we can treat \( \text{var} \left( \hat{\ell}_e(\hat{\sigma}) \right) \) and \( \text{cov} \left( \hat{\ell}_X(\hat{\sigma}), \hat{\ell}_e(\hat{\sigma}) \right) \) and deduce that they have variance to leading order $O(N)$. Finally we must consider the observed Fisher information to deduce the distribution of \( \hat{\sigma} \), see for example [31]. Note that:

\[
\hat{\ell}_X(\sigma) = \sum_{j=1}^{N/2-1} \frac{1}{\sigma_X^2 + \sigma^2 [2 \sin(\pi f_j \Delta t)]^2} - 2 \sum_{j=1}^{N/2-1} \frac{\hat{S}(f_j)}{\sigma_X^2 + \sigma^2 [2 \sin(\pi f_j \Delta t)]^2}^3
\]

\[
\hat{\ell}_{ee}(\sigma) = \sum_{j=1}^{N/2-1} \frac{|2 \sin(\pi f_j \Delta t)|^4}{\sigma_X^2 + \sigma^2 [2 \sin(\pi f_j \Delta t)]^2} - 2 \sum_{j=1}^{N/2-1} \frac{|2 \sin(\pi f_j)|^4 \hat{S}(f_j)}{\sigma_X^2 + \sigma^2 [2 \sin(\pi f_j \Delta t)]^2}^3
\]

\[
\hat{\ell}_{Xe}(\sigma) = \sum_{j=1}^{N/2-1} \frac{|2 \sin(\pi f_j \Delta t)|^2}{\sigma_X^2 + \sigma^2 [2 \sin(\pi f_j \Delta t)]^2} - 2 \sum_{j=1}^{N/2-1} \frac{|2 \sin(\pi f_j)|^2 \hat{S}(f_j)}{\sigma_X^2 + \sigma^2 [2 \sin(\pi f_j \Delta t)]^2}^3
\]

Thus

\[
-E \left\{ \hat{\ell}_X(\sigma) \right\} = \sum_{j=1}^{N/2-1} \frac{\sigma^2 [2 \sin(\pi f_j \Delta t)]^2 + 2S(X)(f_j, f_j) - \sigma_X^2}{\left( \sigma_X^2 + \sigma^2 [2 \sin(\pi f_j \Delta t)]^2 \right)^3} = O(N)
\]  \hspace{1cm} (B-36)

\[
-E \left\{ \hat{\ell}_{ee}(\sigma) \right\} = \sum_{j=1}^{N/2-1} \frac{|2 \sin(\pi f_j)|^4 (\sigma_X^2 [2 \sin(\pi f_j \Delta t)]^2 + 2S(X)(f_j, f_j) - \sigma_X^2)}{\left( \sigma_X^2 + \sigma^2 [2 \sin(\pi f_j \Delta t)]^2 \right)^3} = O(N)
\]

\[
-E \left\{ \hat{\ell}_{Xe}(\sigma) \right\} = \sum_{j=1}^{N/2-1} \frac{|2 \sin(\pi f_j)|^2 (\sigma_X^2 [2 \sin(\pi f_j \Delta t)]^2 + 2S(X)(f_j, f_j) - \sigma_X^2)}{\left( \sigma_X^2 + \sigma^2 [2 \sin(\pi f_j \Delta t)]^2 \right)^3} = O(N),
\]

whilst it follows that the variance of the observed Fisher information is also $O(N)$. We can therefore deduce that renormalized versions of the entries of the Fisher information converge in probability to a constant. Thus using Slutsky’s theorem we can deduce that:

\[
\sqrt{N} \left[ \left( \frac{\sigma_X^2}{\sigma^2} \right) - \left( \frac{\sigma_X^2}{\sigma^2} \right) \right] \xrightarrow{L} N(0, V),
\]

where the entries of $V$ can be found from eqns (B-35) and (B-36). There is no point in deriving this form as we cannot estimate it: it will depend on the unknown Loeve spectrum of $\{X_t\}$ that we have chosen not to model explicitly.
We write $\delta^2 = \hat{\sigma}^2 - \sigma^2 = O \left( \frac{1}{\sqrt{N}} \right)$ and $\delta^2 = \hat{\sigma}^2 - \sigma^2 = O \left( \frac{1}{N} \right)$ and so:

$$\langle X, X \rangle_T^{(m)} = \sum_{k=-N/2}^{N/2-1} \hat{S}(X)(f_k, f_k) = \sum_{k=-N/2}^{N/2-1} \hat{L}_k \hat{S}(Y)(f_k, f_k)$$

$$E \left\{ \langle X, X \rangle_T^{(m)} \right\} = \sum_{k=-N/2}^{N/2-1} \left( \frac{\hat{\sigma}^2 + \sigma^2}{2 |2 \sin(\pi f_k \Delta t)|^2} \right) \left[ S(X)(f_k, f_k) + \sigma^2 |2 \sin(\pi f_k \Delta t)|^2 \right]$$

$$+ \sigma^2 |2 \sin(\pi f_k \Delta t)|^2 + O(N^{-\alpha}) + O(N^{-1+\alpha}) + O \left( \frac{1}{\sqrt{N}} \right)$$

$$= \sum_{k=-N/2}^{N/2-1} \left( \frac{\hat{\sigma}^2 + \sigma^2}{2 |2 \sin(\pi f_k \Delta t)|^2} \right) \left[ S(X)(f_k, f_k) + \sigma^2 |2 \sin(\pi f_k \Delta t)|^2 \right]$$

$$+ O \left( \sqrt{N} \right) + O(N^{-\alpha}) + O(N^{-1+\alpha})$$

$$= \sum_{k=-N/2}^{N/2-1} \left( S(X)(f_k, f_k) + \frac{\hat{\sigma}^2 - S(X)(f_k, f_k)}{\hat{\sigma}^2 - S(X)(f_k, f_k)} \right)$$

$$+ O \left( \sqrt{N} \right) + O(N^{-\alpha}) + O(N^{-1+\alpha})$$

$$= \sum_{k=-N/2}^{N/2-1} \left( \frac{\hat{\sigma}^2 - S(X)(f_k, f_k)}{\hat{\sigma}^2 - S(X)(f_k, f_k)} \right)$$

$$+ O \left( \sqrt{N} \right) + O(N^{-\alpha}) + O(N^{-1+\alpha}) + h.o.t. \quad (C-37)$$

The higher order terms in this expression come from the geometric expansion. These terms will be negligible unless the spectrum of $U_t^{(X)}$ is highly variable. Note that the leading order term in eqn (C-37) is $O(N)$, this making terms $O(\sqrt{N})$ of lesser interest. We then find:

$$\langle X, X \rangle_T^{(m)} = \sum_{k=-N/2}^{N/2-1} S(X)(f_k, f_k) + O \left( \sqrt{N} \right) + \left( \frac{1}{N} \sum_{k=-N/2}^{N/2-1} \right) \left[ \hat{\sigma}^2 - S(X)(f_k, f_k) \right]$$

$$+ O \left( \sqrt{N} \right) + O(N^{-\alpha}) + O(N^{-1+\alpha}) + h.o.t. \quad (C-38)$$

Thus the bias of the estimator can be expressed precisely in terms of the difference between the overall energy across all frequencies and $\hat{S}_X$ the solution to eqns (17), and the difference between $\hat{\sigma}$ the solution to eqns (18), and the true noise variance. To get an asymptotically unbiased estimator we need $\hat{\sigma}^2 - \sigma^2 = o(1)$ as well as $\hat{\sigma}^2 - \Delta t \int S(X)(f, f) df = o(1)$. The less deviation we have been these terms the faster may we expect to get perfect debiasing by our proposed frequency correction. We also note that the variance of this estimator is given by:

$$\text{var} \left\{ \langle X, X \rangle_T^{(m)} \right\} = \frac{1}{N-2} \sum_{k_1=-N/2}^{N/2-1} \sum_{k_2=-N/2}^{N/2-1} \left( \hat{\epsilon}_{k_1} \hat{\epsilon}_{k_2} + O \left( \frac{1}{\sqrt{N}} \right) \right) \text{cov} \left\{ [j(1)(f_k)]^2, [j(1)(f_{k_2})]^2 \right\}.$$
Again we argue that \( \{ J^{(Y)}(f_{k_i}) \} \) are nearly Gaussian from the assumptions of theorem 1, and so we use Isserlis’ formula, see eqn ((B-34)), and we substitute this into eqn (C-39). For processes with stationary increments:

\[
\text{var}\left\{ \langle X, X \rangle_{m_1}^{(m_1)} \right\} = \sum_{k_1=-N/2}^{N/2-1} \sum_{k_2=-N/2}^{N/2-1} \mathcal{L}_{k_1} \mathcal{L}_{k_2} + O\left( \frac{1}{\sqrt{N}} \right) \left\{ |S^{(Y)}(f_{k_1}, f_{k_2})|^2 \text{\delta}_{k_1,k_2} + O(N^{-1+\alpha}) + O(N^{-\alpha}) \right\},
\]

and the stated result follows. Again we could here treat the propriety of using Isserlis’ formula a bit more carefully. The term \( \text{cov}\left\{ |J^{(Y)}(f_{k_1})|^2, |J^{(Y)}(f_{k_2})|^2 \right\} \) could be determined directly from the polyspectrum of \( \{ U^{(Y)}_{t_i} \} \), and for processes with stationary increments, the convergence could be made more precise, see for example [23]. To avoid unnecessary technicality, this discussion is not included in the manuscript.