High Frequency Variability and Microstructure Bias

Adam Sykulski *
Dept. of Mathematics,
Imperial College London
SW7 2AZ, London
ams03@ic.ac.uk

Sofia Olhede †
Dept. of Statistical Science
University College London
WC1 E6BT, London
s.olhede@ucl.ac.uk

Grigorios Pavliotis ‡
Dept. of Mathematics,
Imperial College London
SW7 2AZ, London
g.pavliotis@ic.ac.uk

Abstract

This paper treats the multiscale estimation of the integrated volatility of an Itô process immersed in high-frequency correlated noise. The multiscale structure of the problem is modelled explicitly, and the multiscale ratio is used to quantify energy contributions from the noise, estimated using the Whittle likelihood. This problem becomes more complex as we allow the noise structure greater flexibility, and multiscale properties of the estimation are discussed via a simulation study.

1 Introduction

The estimation of properties of continuous time stochastic processes, whose observation is immersed in high frequency nuisance structure is required in many different fields of application, for example molecular biology and finance. Various methods have been proposed to alleviate bias introduced into the estimation from high frequency nuisance structure, see for example [1–4]. Commonly the model of the observed process is as the process of interest X_{t_i} superimposed with noise ϵ_{t_i}, or

$$Y_{t_i} = X_{t_i} + \epsilon_{t_i}, \tag{1}$$

where Y_{t_i} is the observed process, X_{t_i} the unobserved component of interests, and ϵ_{t_i} is the microstructure noise effect. We model X_t, the process of interest with a suitable stochastic differential equation. For example, the Heston model is specified [5] by

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

where $\nu_t = \sigma_t^2$, and B_t and W_t are correlated 1-D Brownian motions.

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 $\{Y_{t_i}\}$. Different methods have been proposed for determining the properties of X_{t_i}. An outstanding problem is proposing more robust inference methods. [3] has relaxed the assumptions of [1], to include inference of processes with jumps. Another possible direction of development is to include more complicated noise scenarios, namely allowing for correlation between observations. The main issue with such relaxations, is that as the permitted structure of X_t and ϵ_t become less stylized, it naturally becomes harder to separate energy due to the high frequency nuisance component from the process of interest.

Sykulski *et al.* [4] have proposed inference for multiscale processes based on using the discrete Fourier transform. Fourier domain estimators have also been used for estimating noisy Itô processes, see [6], but the main innovation of Sykulski *et al.* was to present a theoretical framework for Harmonizable processes [7, 8] of interest, and an automatic procedure for estimating the nuisance structure was proposed. The Whittle likelihood was used to estimate the energy level of the process of interest, as well as the noise contamination. The method was shown to perform well under various signal to noise scenarios, as well as path lengths, see [4].

*www.ecs.soton.ac.uk/people/as07r
†www.homepages.ucl.ac.uk/~ucaksc0/
‡www.ma.ic.ac.uk/~pavl/

The results of [4] or [1] are only appropriate when the noise is white. We shall in contrast in this paper discuss possible extensions of the multiscale estimators to the case of more complicated market microstructure, and illustrate the performance of the estimator in various noise scenarios.

2 Multiscale Estimation

In the absence of noise a suitable estimator of the integrated volatility, $\langle X, X\rangle_T = \int_0^T \sigma_t^2\, dt$, can be specified from the quadratic variation of the process $\{Y_t\}$. In the presence of market microstructure noise this is no longer true and it is necessary to employ a different estimation procedure. For ease of exposition we denote the difference process $Z_{t_i} - Z_{t_{i-1}}$ by $U_{t_i}^{(Z)}$ where $Z = X, Y$ or ϵ. The Loève spectrum [7,8] of $U_{t_i}^{(Z)}$ will be denoted $S^{(Z)}(f_k, f_k)$, and we note that the observed quadratic variation can be rewritten as:

$$\widehat{\langle X, X\rangle}_T^{(b)} = \sum_{i=0}^{N-1} \left(U_{t_i}^{(Y)}\right)^2 = \sum_{k=-N/2}^{N/2-1} \left|J^{(Y)}(f_k)\right|^2 \tag{3a}$$

$$\widehat{S}^{(Y)}(f_k, f_k) = \left|J^{(Y)}(f_k)\right|^2, \quad J^{(Y)}(f_k) = \frac{1}{\sqrt{N}} \sum_{j=1}^{N-1} U_{t_j}^{(Y)} e^{-2\pi i t_j f_k}. \tag{3b}$$

with $f_k = \frac{k}{T}$. $\widehat{S}^{(Y)}(f_k, f_k)$ is the periodogram estimator, see [9], and normally has a single argument because the covariance of the Fourier Transform at two fixed frequencies is asymptotically equivalent to zero for a stationary process. We note directly from [4] that the bias of the estimator $\widehat{\langle X, X\rangle}_T^{(b)}$ is conveniently expressed in the Fourier domain by the observation that

$$\mathrm{E}\left\{\widehat{\langle X, X\rangle}_T^{(b)}\right\} = \sum_{k=-N/2}^{N/2-1} S^{(X)}(f_k, f_k) + \sigma_\epsilon^2 \sum_{k=-N/2}^{N/2-1} \left|2\sin(\pi f_k \Delta t)\right|^2 + O(N^\alpha) + O\left(N^{1-\alpha}\right). \tag{4}$$

The error terms follow from assumptions regarding the spectral properties of the process X_t, and are detailed in [4]. These assumptions determine the value of α. It is clear from eqn (4) that the influence of the noise increases for larger frequencies, and that the relative magnitude of $S^{(X)}(f_k, f_k)$ to $\sigma_\epsilon^2 \left|2\sin(\pi f_k \Delta t)\right|^2$ at frequency f_k will determine the need for bias correction at f_k.

Sykulski *et al* proposed to measure the average energy of $U_t^{(X)}$ across frequencies, and determine the energy of $U_t^{(\epsilon)}$, using the form of the white noise spectrum. Despite $U_t^{(X)}$ assumed harmonizable and not necessarily stationary, with appropriate assumptions regarding the spectral correlation of the process, it is appropriate to use the Whittle likelihood, see [10], to determine the relative energy of the two processes across scales. Instead of using eqn (3b) to estimate the spectral contributions of the process of interest, a shrinkage estimator of $S^{(X)}(f_k, f_k)$ was therefore proposed in [4]:

$$\widehat{S}^{(X)}(f_k, f_k; L_k) = L_k \widehat{S}^{(Y)}(f_k, f_k). \tag{5}$$

$0 \le L_k \le 1$ is referred to as the 'multiscale ratio' and its optimal form for perfect bias correction when ϵ_{t_i} is white noise is given by:

$$L_k = \frac{S^{(X)}(f_k, f_k)}{S^{(X)}(f_k, f_k) + \sigma_\epsilon^2 \left|2\sin(\pi f_k \Delta t)\right|^2}. \tag{6}$$

Of course this assumes perfect knowledge of $S^{(X)}(f_k, f_k)$ and is not a realizable estimator. Instead typical contributions of $S^{(X)}(f_k, f_k)$ across frequencies were considered, and the multiscale ratio replaced by a sort of average ratio corresponding to

$$\overline{L}_k = \frac{\sigma_X^2}{\sigma_X^2 + \sigma_\epsilon^2 \left|2\sin(\pi f_k \Delta t)\right|^2}. \tag{7}$$

The justification for this choice is discussed in Sykulski *et al*. We estimate the parameters of $\overline{L}_k$ by maximising a pseudo-likelihood namely the multiscale Whittle likelihood defined in parameter

$\boldsymbol{\sigma} = \begin{pmatrix} \sigma_\epsilon^2 & \sigma_X^2 \end{pmatrix}$

$$\ell(\boldsymbol{\sigma}) = -\sum_{k=1}^{N/2-1} \log\left(\sigma_X^2 + \sigma_\epsilon^2 \left|2\sin(\pi f_k \Delta t)\right|^2\right) - \sum_{k=1}^{N/2-1} \frac{\widehat{S}^{(Y)}(f_k, f_k)}{\sigma_X^2 + \sigma_\epsilon^2 \left|2\sin(\pi f_k \Delta t)\right|^2}.$$

If $\left\{U_t^{(X)}\right\}$ is a stationary process, then the full Whittle likelihood (with σ_X^2 replaced by $S^{(X)}(f_k, f_k)$) will approximate the time-domain likelihood of the sample, under suitable regularity conditions, see [11].

The bias corrected estimator of the integrated volatility for an estimated $\hat{L}_K$ sequence becomes

$$\widehat{\langle X, X\rangle}_T^{(m_1)} = \sum_{k=-N/2}^{N/2-1} \widehat{S}^{(X)}(f_k, f_k; \widehat{L}_k). \tag{8}$$

In Sykulski *et al.* it was shown that the estimates of σ_X^2 and σ_ϵ^2 produced suitable $\widehat{L}_k$ such that bias corrected estimators of $S^{(X)}(f_k, f_k)$ with suitable properties were defined. Unfortunately it is not always reasonable to model the high frequency structure as white, and so more subtle modelling needs to be used when the noise is more complicated.

3 Correlated Noise

A key issue is treating correlation in the error terms. A reasonable relaxation of modelling ϵ_{t_i} as white would correspond to ϵ_{t_i} stationary. Stationary processes can be conveniently represented in terms of aggregations of uncorrelated white processes, using the Wold decomposition theorem [12][p. 187]. We may therefore write the zero-mean observation ϵ_{t_i} as

$$\epsilon_{t_i} = \sum_{j=0}^{\infty} \theta_{t_j} \eta_{t_i - t_j}, \tag{9}$$

where $\theta_{t_0} \equiv 1$, $\sum_j \theta_{t_j}^2 < \infty$, and $\{\eta_{t_n}\}$ satisfies $\mathrm{E}\left[\eta_{t_n}\right] = 0$ and $\mathrm{E}\left[\eta_{t_n}\eta_{t_m}\right] = \sigma_\eta^2 \delta_{n,m}$. Common practise would involve approximating the distribution by a finite number of elements in the sum, and thus truncate eqn (9) to $q \in \mathbb{Z}$. We therefore model the noise as a Moving Average (MA) process specified by

$$\epsilon_{t_i} = \eta_{t_i} + \sum_{k=1}^{q} \theta_{t_k} \eta_{t_{i-k}}, \tag{10}$$

and the spectral density function [9] of ϵ_{t_i} takes the form:

$$S^*(f; \boldsymbol{\theta}, \sigma_\eta^2) = \sigma_\eta^2 \left| 1 + \sum_{k=1}^{q} \theta_k e^{2i\pi f k} \right|^2. \tag{11}$$

In this case our spectral model for ϵ_{t_i} changes to a Loève spectrum of

$$S^{(\epsilon)}(f, f) = S^*(f; \boldsymbol{\theta}, \sigma_\eta^2) |2\sin\left(\pi f \Delta t\right)|^2. \tag{12}$$

Two possible methods now exist for treating the nuisance function of $S^*(f; \boldsymbol{\theta}, \sigma_\eta^2)$: we can use the method of Sykulski *et al.* directly without adjustment, assuming the variability of $S^*(f; \boldsymbol{\theta}, \sigma_\eta^2)$ to be moderate or we could adjust the methodology to encompass a parametric model for the noise, replacing $\sigma_\epsilon^2 |2\sin\left(\pi f \Delta t\right)|^2$ by $S^*(f; \boldsymbol{\theta}, \sigma_\eta^2)|2\sin\left(\pi f \Delta t\right)|^2$ when treating the frequency structure of the micro structure noise.

For a fixed and specified value of q, we may therefore estimate the parameters of the MA, using the Whittle likelihood, but where now $\sigma_\epsilon^2 |2\sin\left(\pi f \Delta t\right)|^2$ is replaced by $S^*(f)|2\sin\left(\pi f \Delta t\right)|^2$. We thus get a multiscale likelihood [1] given with $\boldsymbol{\sigma} = \begin{pmatrix}\sigma_\eta^2 & \sigma_X^2\end{pmatrix}$ by

$$\ell(\boldsymbol{\sigma},\boldsymbol{\theta}) = -\sum_{k=1}^{N/2-1} \log\left(\sigma_X^2 + S^*(f_k;\boldsymbol{\theta},\sigma_\eta^2)\left|2\sin(\pi f_k \Delta t)\right|^2\right) - \sum_{k=1}^{N/2-1} \frac{\widehat{S}^{(Y)}(f_k,f_k)}{\sigma_X^2 + S^*(f_k;\boldsymbol{\theta},\sigma_\eta^2)\left|2\sin(\pi f_k \Delta t)\right|^2}. \tag{13}$$

and the augmented multiscale ratio is defined by

$$\overline{L}_k^{(a)} = \frac{\sigma_X^2}{\sigma_X^2 + S^*(f_k;\boldsymbol{\theta},\sigma_\eta^2)\left|2\sin(\pi f_k \Delta t)\right|^2}. \tag{14}$$

If q is not assumed known, then model choice methods can also be applied to determine the value of q, such as applying the modified Akaike AIC [12][p. 287], and adding $2n(q+2)/[n-q-3]$ to minus two times the log multiscale likelihood, and minimizing this objective function. Some care must be applied as the Akaike AIC is known to overestimate the number of parameters, and BIC or some other model choice method may be applied. For a chosen value of q once we augment the estimation of σ_X^2 and σ_ϵ^2 with that of $\{\theta_{t_k}\}$, then we can estimate the noise spectrum and hence the multiscale ratio. This will yield an augmented estimator of the integrated volatility, replacing the parameters by their estimators in eqn (14), that we denote by $\widehat{L}_k^{(a)}$. Our new estimator then takes the form

$$\widehat{\langle X, X\rangle}_T^{(a)} = \sum_{k=-N/2}^{N/2-1} \widehat{S}^{(X)}(f_k,f_k;\widehat{L}_k^{(a)}). \tag{15}$$

This form both takes the high frequency structure into account, and permits the high frequency structure to be more dynamic than is the case of simple white noise nuisance structure.

4 Examples

We investigate the simple case of

$$\epsilon_{t_i} = \eta_{t_i} + \theta_1 \eta_{t_{i-1}}, \tag{16}$$

where then $S^*(f) = 1 + \theta_1^2 + 2\theta_1\cos(2\pi f)$. Clearly it is of interest to investigate the effect of the variability of $S^*(f)$ on the multiscale estimation procedure. We note that setting $\theta_1 = 0$ recovers the white noise structure investigated in [4] and [1]. It is therefore of interest to compare our estimators over a range of values for θ_1 to study the effect of additional variability in the spectrum of the nuisance structure in the estimation of the integrated volatility. This is not a full study of the complete effects of complicated high-frequency structure superimposed on the process of interest: this study is intended to demonstrate the adverse effects of a more dynamic nuisance structure, and the potential of correcting for such effects using the multiscale structure of the process of interest.

We demonstrate the performance of our multiscale estimators of integrated volatility using the Heston model defined in eqn (2), with the same parameter values as used in [4] and [1], except this time we generate the microstructure noise process by eqn (16). Our new estimator $\widehat{\langle X, X\rangle}_T^{(a)}$, requires estimation of the parameters $(\sigma_X^2, \sigma_\epsilon^2, \theta_1)$ and this is done separately for each path using the MATLAB function fmincon on eqn (13). Figures 1(a) and 1(b) show the approximated σ_X^2 and $S^*(f_k;\theta_1,\sigma_\eta^2)\left|2\sin(\pi f_k\Delta t)\right|^2$ (in white) plotted over the periodograms $\widehat{S}^{(X)}(f_k,f_k)$ and $\widehat{S}^{(\epsilon)}(f_k,f_k)$ for one simulated path, where $\theta_1 = 0.5$. The parameters $(\sigma_X^2, \sigma_\epsilon^2, \theta_1)$ seem to have been approximated well, as the approximated spectral densities follow the shape of their respective periodograms. Figure 1(c) shows the corresponding multiscale ratio $\widehat{L}_k^{(a)}$ (in white) plotted over an unrealizable estimate of L_k:

$$\widetilde{L}_k = \frac{\widehat{S}^{(X)}(f_k,f_k)}{\widehat{S}^{(X)}(f_k,f_k) + \widehat{S}^{(\epsilon)}(f_k,f_k)}. \tag{17}$$

[1] Note that $\ell(\boldsymbol{\sigma},\boldsymbol{\theta})$ is not strictly speaking a likelihood, see the full discussion in Sykulski et al. [4], but can for all intents and purposes be treated as such in this context.

Table 1: Root Mean Square Error (RMSE) for the different estimators of the integrated volatility, over different values of θ_1. The RMSEs are averaged over 7,500 paths.

RMSE$\{\cdot\}$	$\widehat{\langle X,X\rangle}_T^{(b)}$	$\widehat{\langle X,X\rangle}_T^{(s_1)}$	$\widehat{\langle X,X\rangle}_T^{(m_1)}$	$\widehat{\langle X,X\rangle}_T^{(a)}$	$\widehat{\langle X,X\rangle}_T^{(u)}$
$\theta_1 = -1$	3.51×10^{-2}	4.82×10^{-4}	7.35×10^{-5}	1.52×10^{-5}	1.46×10^{-5}
$\theta_1 = -0.75$	2.71×10^{-2}	3.62×10^{-4}	7.14×10^{-5}	1.56×10^{-5}	1.44×10^{-5}
$\theta_1 = -0.5$	2.05×10^{-2}	2.42×10^{-4}	6.40×10^{-5}	1.57×10^{-5}	1.44×10^{-5}
$\theta_1 = -0.25$	1.54×10^{-2}	1.21×10^{-4}	4.58×10^{-5}	1.60×10^{-5}	1.44×10^{-5}
$\theta_1 = 0$	1.17×10^{-2}	1.67×10^{-5}	1.61×10^{-5}	1.62×10^{-5}	1.43×10^{-5}
$\theta_1 = 0.25$	9.51×10^{-3}	1.22×10^{-4}	1.18×10^{-4}	1.67×10^{-5}	1.45×10^{-5}
$\theta_1 = 0.5$	8.78×10^{-3}	2.41×10^{-4}	4.67×10^{-4}	1.70×10^{-5}	1.43×10^{-5}
$\theta_1 = 0.75$	9.51×10^{-3}	3.62×10^{-4}	2.13×10^{-3}	1.74×10^{-5}	1.44×10^{-5}
$\theta_1 = 1$	1.17×10^{-2}	4.82×10^{-4}	9.82×10^{-3}	1.67×10^{-5}	1.45×10^{-5}

Our multiscale ratio provides a good estimate to L_k and will remove the noise microstructure from the correct frequencies by shrinkage. Figure 1(d) shows $\widehat{L}_k^{(a)}\widehat{S}^{(Y)}(f_k, f_k)$; the energy has been shrunk at frequencies affected by the microstructure noise and the spectrum is a good approximation to $\widehat{S}^{(X)}(f_k, f_k)$, which in turn should lead to a good approximation of the integrated volatility, compare with Figure 1(a). Figures 2(a) and 2(b) show two more estimated multiscale ratios $\widehat{L}_k^{(a)}$ (in white), but this time with $\theta_1 = -0.5$ and $\theta_1 = 1$ respectively. The multiscale estimator appears to correctly detect the correlation of noise in the process, as well as the magnitude of the signal to noise ratio. Note that for $\theta_1 = -0.5$ we shrink the estimated Loève spectrum at an increasing rate for high frequencies, whilst for $\theta_1 = 1$ we shrink in a highly non-monotone fashion across frequencies.

We investigate the performance of our new estimator against the estimators developed in [4] and [1] using Monte Carlo simulations. A range of values for θ_1 are used to investigate the effect of correlated noise. For each value of θ_1 we generated 7,500 simulated paths. Table I displays the results of our simulation, where the errors are calculated using a Riemann sum approximation on the X_t process (see [4] for details). Along with the performance of our new estimator $\widehat{\langle X,X\rangle}_T^{(a)}$ (eqn (15)), we include the performance of the estimator from [4], $\widehat{\langle X,X\rangle}_T^{(m_1)}$ (eqn (8)) and the best un-biased estimator developed in [1], $\widehat{\langle X,X\rangle}_T^{(s_1)}$. Naturally we do not aim to compare our estimator for correlated noise structure with that of [1,4], as these were not developed for correlated noise, but more include these to show the necessity of treating correlation in the microstructure. Furthermore, had our Whittle estimators been sufficiently poor, then the variability of the estimated multiscale ratio would have made our proposed procedure unsuitable. We also include for reference, the biased estimator in eqn (3a), $\widehat{\langle X,X\rangle}_T^{(b)}$ (the quadratic variation on Y_t) and the unobservable unbiased estimator

$$\widehat{\langle X,X\rangle}_T^{(u)} = \sum_{i=0}^{N-1} \left(U_{t_i}^{(X)}\right)^2 \tag{18}$$

the quadratic variation on X_t, which in some sense is the best estimator that can be achieved.

The table shows that the new estimator performs remarkably well under different values of θ_1. In fact the RMSE of the estimator is very close to that of the unobservable quadratic variation, the best measure in the absence of market microstructure. The loss of efficiency by the more flexible model when $\theta_1 = 0$ is marginal whilst when $\theta_1 = 1$ the RMSE has decreased by a factor of a 500 compared to [4], and by a factor of 30 compared to [1], whilst if $\theta_1 = -1$ the RMSE has decreased by a factor of a 5 compared to [4], and by a factor of 30 compared to [1]. The small and consistent RMSE is due to the successful bias removal of the augmented multiscale estimator, where the low mean square error of the estimators of $(\sigma_X^2, \sigma_\epsilon^2, \theta_1)$, ensures that the bias in the estimated Loève spectrum of the process of interest is removed efficiently. Figure 3 shows the distribution of the estimates of these parameters over the 7,500 simulated paths for $\theta_1 = 0.5$; the estimation procedure is unbiased and has reasonably low variance.

The estimators $\widehat{\langle X, X \rangle}_T^{(s_1)}$ and $\widehat{\langle X, X \rangle}_T^{(m_1)}$ are inconsistent when additional structure is permitted in the noise. We stress that these are estimators based on assumptions of white noise, and their strong performance in this instance ($\theta_1 = 0$) is apparent. As we move away from white noise, $\widehat{\langle X, X \rangle}_T^{(s_1)}$ and $\widehat{\langle X, X \rangle}_T^{(m_1)}$ overcompensate for the noise when θ_1 in near minus one and undercompensate when θ_1 is near one. This happens because as the value of θ changes, taking values between minus one and one the spectral properties of the noise process change quite markedly with the appropriate shrinkage factor changing form in a corresponding fashion. For negative values of θ_1 the multiscale ratio, and smaller positive values of θ_1 the augmented multiscale ratio is decreasing at higher frequencies, whilst when θ_1 approaches one the multiscale ratio is not monotone (see Figures 1(c), 2(a) and 2(b)). $\widehat{\langle X, X \rangle}_T^{(m_1)}$ seems to still perform well for negative θ_1 values (note how the spectral form of the noise process is still largely the same shape) but performs disastrously for positive θ_1 values, due to the larger energy at lower frequencies that the estimator fails to remove. $\widehat{\langle X, X \rangle}_T^{(s_1)}$ suffers equivalent loss of performance as θ_1 moves away from zero in each direction; for such a time-domain estimator to perform better in these instances, the optimal subsampling rate of the estimator would have to be re-calibrated to incorporate the correlated noise. Nevertheless, all the estimators perform better than the noise polluted and biased estimator of the quadratic variation on Y_t, $\widehat{\langle X, X \rangle}_T^{(b)}$.

5 Conclusions

This paper has proposed extending the multiscale estimation methods of Sykulski *et al* for integrated volatility to include the case of stationary high frequency nuisance structure. It was found that naively applying estimators designed for the case of uncorrelated noise did not perform well. By modelling the nuisance structure as a Moving Average process, better bias correction could be applied at each frequency, and this substantially improved our estimator of the integrated volatility. Despite greater flexibility, the performance of the estimator did not deteriorate in terms of mean square error, which could have been a possible outcome. Note that the multiscale methods did not include parametric modelling of $\{X_t\}$ only approximating its multiscale nature. Future avenues of investigation includes rigorous model choice procedures, and the application of Bayesian estimation methods to naturally incorporate the multiscale ratio by Hierarchical modelling. Multiscale modelling shows great promise for designing inference methods for continuous time processes, by the increase in precision and power from investigating properties directly scale-by-scale.

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] G. A. Pavliotis and A. M. Stuart, “Parameter estimation for multiscale diffusions”, *J. Stat. Phys.*, vol. 127, pp. 741–781, 2007.

[3] J. Fan and Y. Wang, “Multi-scale jump and volatility analysis for high-frequency financial data”, *J. of the American Statistical Association*, vol. 102, pp. 1349–1362, 2007.

[4] A. Sykulski, S. C. Olhede, and G. Pavliotis, “Multiscale inference for high-frequency data”, Tech. Rep. 290, Department of Statistical Science, University College London, arxiv.org/abs/0803.0392, 2008.

[5] 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.

[6] M. E. Mancino and S. Sanfelici, “Robustness of fourier estimators of integrated volatility in the presence of microstructure noise”, Tech. Rep., University of Firenze, 2006.

[7] M. Loève, *Probability theory. I*, Springer-Verlag, New York, fourth edition, 1977, Graduate Texts in Mathematics, Vol. 45.

[8] M. Loève, *Probability theory. II*, Springer-Verlag, New York, fourth edition, 1978, Graduate Texts in Mathematics, Vol. 46.

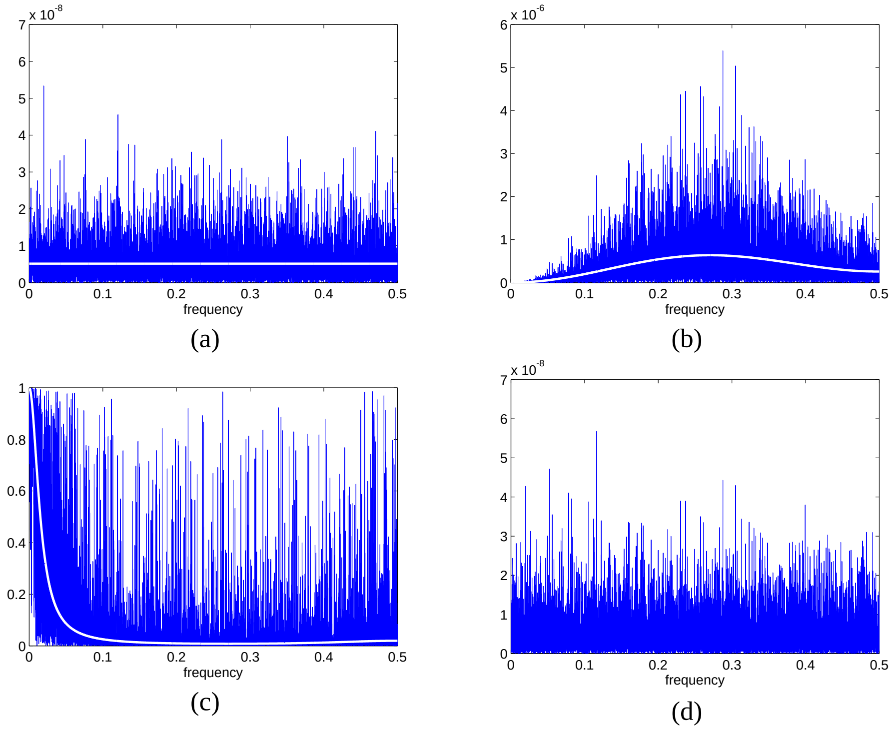


Figure 1: (a) The periodogram of a realisation of $U_t^{(X)}$ (solid line), (b) of a realisation of $U_t^{(\epsilon)}$ (solid line) with the Whittle estimates superimposed (white solid line), (c) the estimate of L_k from the raw periodograms of the unobserved processes (solid line) with the Whittle estimate $\widehat{L_k}$ superimposed (white solid line) and (d) the bias corrected estimator of the periodogram of $U_t^{(X)}$, using $\widehat{L}_k$. $\theta_1 = 0.5$ in this example. Notice the different scales in the four figures. Estimated spectra are here plotted on a linear scale for ease of comparison to the effect of applying L_k.

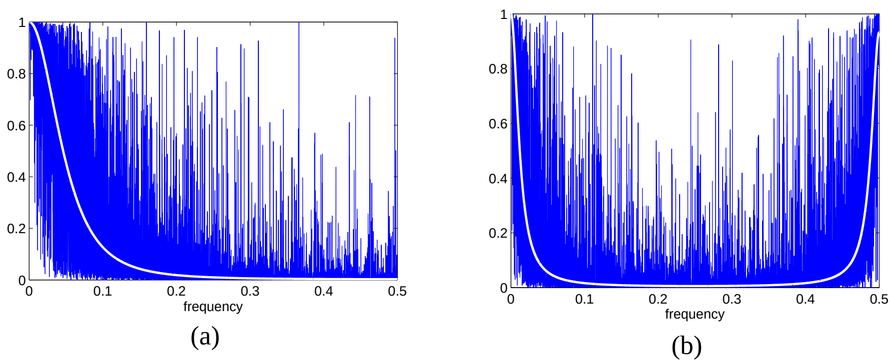
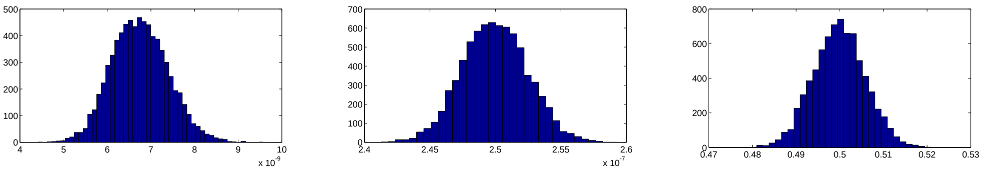


Figure 2: The estimate of L_k from the raw estimated spectra of the unobserved processes (solid line) with the Whittle estimate $\widehat{L_k}$ (white solid line) superimposed for (a) $\theta_1 = -0.5$ and (b) $\theta_1 = 1$. Notice the non-monotone structure of the multiscale ratio in the second case.



(a) (b) (c)

Figure 3: Histograms showing the distribution of the estimators of the parameters in $\widehat{L}_k$, for (a) σ_X^2, (b) σ_ϵ^2 and (c) θ_1 (where the true value of θ_1 is 0.5) over 7,500 simulated paths.

[9] D. Brillinger, *Time series: data analysis and theory*, Society for Industrial and Applied Mathematics, Philadelphia, USA, 2001.

[10] J. Beran, *Statistics for long-memory Processes*, Chapman and Hall, London, 1994.

[11] K. O Dzhamparidze and A. M. Yaglom, "Spectrum parameter estimation in time series analysis", in *Developments in Statistics*, PR Krishnaiah, Ed., vol. 4, pp. 1–181. New York: Academic Press., 1983.

[12] Peter J. Brockwell and Richard A. Davis, *Time Series: Theory and Methods*, Springer, Berlin, Germany, 1991.