Measurement of Volatility of Diffusion Processes with Noisy High Frequency Data

[1]**K. Oya**

[1]Graduate School of Economics, Osaka University,
560-0043, Machikaneyama Toyonaka, Osaka, Japan. E-Mail: kosuke@econ.osaka-u.ac.jp

Keywords: *High frequency data; volatility; market microstructure noise; Fourier analysis*

EXTENDED ABSTRACT

A measurement volatility of return process should be the primary object of traders and practitioners in financial market for management of their portfolios and making trading decisions. The realized volatility is the representative estimator of (integrated) volatility and is computed from historical data of the return. The sampling interval of the return plays a key role in computing the realized volatility. It is commonly believed in empirical finance literature that the return process should not be sampled too often in a fixed period. The realized volatility results become biased if the sampling interval is chosen to be too small although the realized volatility computed from the high frequency data should be the reliable estimate from a statistical point of view.

One source of the bias is the unequally spaced time series from the non-synchronous trading. The method based on Fourier analysis proposed in Malliavin and Mancino (2002) is very effective approach since it does not require the even spaced time series. However, recent studies find that the market microstructure closely relates with the bias of volatility estimator. The microstructure issues become more pronounced when the data are sampled at finer intervals. The empirical studies suggest that the contamination due to market microstructure makes the bias in the volatility estimator. The property about the estimator based on Fourier analysis has not been examined so far when the observed sample contains market microstructure noise.

This research shows that the estimator based on Fourier analysis is also biased when the observed sample contains microstructure noise and proposes the bias corrected estimation method. We compared the properties of these estimators using Monte Carlo experiments. The results summarized in Table1. In Table 1, $[Y,Y]_T^{(all)}$, $\widehat{\mathcal{IV}}_T^{(RV)}$ and $\widehat{\mathcal{IV}}_T^{(F)}$ are the relative errors of the realized volatility, the bias corrected realized volatility proposed in Zhang *et al* (2004) and the bias corrected estimate proposed in this research, respectively. Figure 1 shows that the empirical distribution of the relative errors of estimates. We find that the realized volatility commonly used fails to estimate the true integrated volatility when the returns are sampled at high frequency with microstructure noise and the proposed estimator in this research has the smallest mean squared error among others. This characteristic is also found in the unevenly sampled case.

Table 1. Evenly sampled case

	$[Y,Y]_T^{(all)}$	$\widehat{\mathcal{IV}}_T^{(RV)}$	$\widehat{\mathcal{IV}}_T^{(F)}$
Mean	1.1×10^4	4.540	1.964
SD	7.9×10^3	13.541	3.445
MSE	1.8×10^8	203.968	15.727

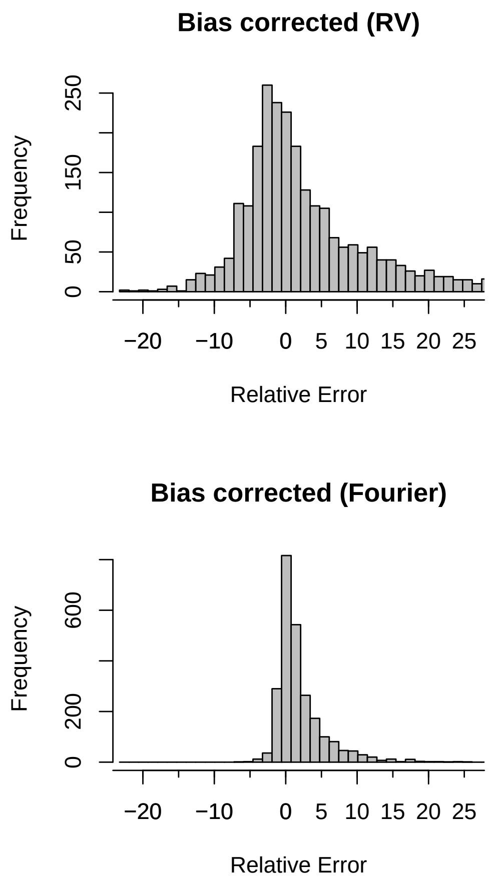

Figure 1. Evenly sampled with Microstructure Noise

1. INTRODUCTION

A measurement volatility of return process should be the primary object of traders and practitioners in financial market for management of their portfolios and making trading decisions. The realized volatility is the representative estimator of (integrated) volatility and is computed from historical data of the return. The sampling interval of the return plays a key role in computing the realized volatility. It is commonly believed in empirical finance literature that the return process should not be sampled too often in a fixed period. The realized volatility results become biased if the sampling interval is chosen to be too small although the realized volatility computed from the high frequency data should be the reliable estimate from a statistical point of view.

One source of the bias is the unequally spaced time series from the non-synchronous trading. The method based on Fourier analysis proposed in Malliavin and Mancino (2002) is very effective approach since it does not require the even spaced time series. However, recent studies find that the market microstructure closely relates with the bias of volatility estimator. The microstructure issues become more pronounced when the data are sampled at finer intervals. The empirical studies suggest that the contamination due to market microstructure makes the bias in the volatility estimator. The property about the estimator based on Fourier analysis has not been examined so far when the observed sample contains market microstructure noise. This paper shows that the estimator based on Fourier analysis is also biased when the observed sample contains microstructure noise, proposes the bias corrected estimation method, and compares the properties of these estimators using Monte Carlo experiments.

The outline of this paper is as follows. Section 2 details the return process and market microstructure noise. In section 3, we summarize the estimation methods. In section 4, some Monte Carlo experiments are conducted to compare the estimators. Section 5 contains concluding remarks.

2. MODEL

2.1 Diffusion Process and Volatility

Let S_t be the price process of some financial asset and the return process $X_t = \ln S_t$ follows a continuous univariate diffusion process.

$$dX_t = \mu_t dt + \sigma_t dW_t, \quad t \in [0,T] \tag{1}$$

where W_t is a standard Brownian motion. The instantaneous variance σ_t^2 of the process X_t and the drift coefficient μ_t also follow stochastic processes. Our main objective is to estimate the integrated volatility $\int_0^T \sigma_t^2 dt$ using $X_t \in [0,T]$ nonparametricly. It has been reported empirically that the nonparametric estimate of $\int_0^T \sigma_t^2 dt$ with high frequency data is not necessarily robust. We introduce the market microstructure noise in the next subsection to treat this phenomenon.

2.2 Market Microstructure Noise

Zhang *et al* (2004) gives the theoretical grounds why and where the standard volatility estimator fails when the data are sampled at the highest frequencies and proposes the estimation method to correct the bias of estimator. Their contention is that the contamination due to market microstructure noise is the same as what statisticians usually call "observation error". Let t_i be the i-th transaction in time period $[0,T]$ and ε_{t_i} be the microstructure noise. We assume that this microstructure noise ε_{t_i} is independent with the process X_{t_i}. Then the observation Y_{t_i} consists of two random variables as follow

$$Y_{t_i} = X_{t_i} + \varepsilon_{t_i}, \quad t_i \in [0,T]. \tag{2}$$

It is noted that the intervals of the transactions $t_i \in [0,T]$, $i = 1, \cdots, N$ are not necessarily equally and regularly spaced.

3. ESTIMATION METHOD

3.1 Using Sum of Squared Return

Empirical researchers and practitioners commonly use the sum of squared return $\sum_{t_i}(Y_{t_{i+1}} - Y_{t_i})^2$ as an estimator of the integrated volatility $\int_0^T \sigma_t^2 dt$ and called "realized volatility". We define the realized volatility as

$$[Y,\ Y]_T^{(all)} = \sum_{t_i}(Y_{t_{i+1}} - Y_{t_i})^2. \tag{3}$$

The realized volatility $[Y,Y]_T^{(all)}$ is a consistent estimator if we observe the return process without the microstructure noise.

$$\text{plim}[Y,\ Y]_T^{(all)} = \int_0^T \sigma_t^2 dt. \tag{4}$$

This convergence is attained as n goes to infinity where n is the sample size over sample period $[0,T]$. In this case

$$\lim_{n\to\infty} \sup_{i<n}(t_i - t_{i-1}) = 0.$$

However the realized volatility $[Y,Y]_T^{(all)}$ is a biased estimator of $\int_0^T \sigma_t^2 dt$ when the microstructure noise exists. It is obvious that $[Y,Y]_T^{(all)}$ converges to the true integrated volatility plus the variance of microstructure noises. The magnitude of the variance of microstructure noise dominates the integrated volatility. Zhang *et al* (2004) shows that the effect of ignoring microstructure noise on $[Y,Y]_T^{(all)}$ is

$$[Y,\ Y]_T^{(all)} = 2n\,\mathrm{E}[\varepsilon^2]\ +\ O_p(n^{1/2}). \tag{5}$$

This fact tells us that the realized volatility which is commonly used in practice does not estimate the integrated volatility $\int_0^T \sigma_t^2 dt$ when the returns are sampled at very high frequency, that is n is very large.

Zhang *et al* (2004) also proposed several bias corrected estimators. The first best approach consists of the realized volatility $[Y,Y]_T^{(all)}$ and the second best estimator which is described below. The second best estimator is constructed by averaging the estimators $[Y,Y]_T^{(k)}$ across K grids of average size $\bar{n}$. The full grid $\mathcal{G}$, $\mathcal{G} = \{t_0,\cdots,t_n\}$ is partitioned into K non-overlapping subgrids $\mathcal{G}^{(k)} = \{t_{k-1},t_{k-1+K},t_{k-1+2K},\cdots,t_{k-1+n_kK}\}$ for $k = 1,\cdots,K$ and n_k is the integer making t_{k-1+n_kK} the last element in $\mathcal{G}^{(k)}$. Let $[Y,Y]_T^{(k)}$ be a realized volatility using only subsampled observations $Y_t \in \mathcal{G}^{(k)}$. The second best estimator that proposed by Zhang *et al* (2004) is

$$[Y,\ Y]_T^{(avg)} = \frac{1}{K}\sum_{k=1}^{K}[Y,\ Y]_T^{(k)} \tag{6}$$

Let $\mathcal{IV}_T$ and $\widehat{\mathcal{IV}}_T^{(RV)}$ be $\int_0^T \sigma_t^2 dt$ and its estimator using realized volatility with full and subgrid sample.

Then the bias corrected estimator of $\mathcal{IV}_T$ is

$$\widehat{\mathcal{IV}}_T^{(RV)} = [Y,Y]_T^{(avg)}\ -\ \frac{\bar{n}}{n}[Y,Y]_T^{(all)} \tag{7}$$

where $\bar{n} = (n-K+1)/K$, $K = cn^{2/3}$ and c is some constant. Zhang *et al* (2004) derived the optimal choice of c^* which includes unknown parameters that have to be estimated. See Zhang *et al* (2004) for details.

3.2 Method based on Fourier Analysis

Malliavin and Mancino (2002) proposed the other type of estimation method that based on the Fourier series analysis. Barucci and Renò (2002) examined that the properties of the estimator and confirmed that the estimator is almost unbiased and its variance is smaller than that of the sum of squared returns when the returns are unevenly sampled at high frequency and there is no microstructure noise.

The return process is

$$dX_t = \mu_t dt\ +\ \sigma_t dW_t,\quad t\in[0,T].$$

We normalize the time window $[0,T]$ to $[0,2\pi]$. The Fourier coefficients of dX_t are

$$a_0(dX) = \frac{1}{2\pi}\int_0^{2\pi} dX_t \tag{8}$$

$$a_k(dX) = \frac{1}{\pi}\int_0^{2\pi}\cos(kt)dX_t \tag{9}$$

$$b_k(dX) = \frac{1}{\pi}\int_0^{2\pi}\sin(kt)dX_t, \tag{10}$$

where $k \geq 1$. Malliavin and Mancino (2002) proposes an estimator of $\mathcal{IV}_T$ using the Fourier coefficients of dX_t

$$\mathcal{F}^{(all)}(X) = \frac{\pi^2}{n+1-n_0}\sum_{k=n_0}^{n}(a_k^2(dX)+b_k^2(dX)) \tag{11}$$

and shows its consistency

$$\mathcal{IV}_T = 2\pi\ a_0(\sigma^2) = \lim_{n\to\infty}\mathcal{F}^{(all)}(X). \tag{12}$$

Using the integration by parts, we have

$$\begin{aligned} a_k(dX) &= \frac{1}{\pi}\int_0^{2\pi}\cos(kt)dX_t \\ &= \frac{X_{2\pi}-X_0}{\pi}\ +\ \frac{k}{\pi}\int_0^{2\pi}\sin(kt)X_t dt \end{aligned}$$

In fact, we only have the discrete sample X_{t_i},$i = 1,\cdots,N$. The standard interpolation methods to represent X_t, $t\in[t_i,t_{i+1}]$ are linear interpolation and previous-tick interpolation. We apply the previous-tick interpolation, that is $X_t = X_{t_i}$ for $t\in[t_i,t_{i+1}]$, for the integration in the above equation. $b_k(dX)$ is also obtained in the same manner.

However the proposed method in Malliavin and Mancino (2002) does not take into account the effect of microstructure noise. As described above, the observed process is

$$Y_{t_i} = X_{t_i}\ +\ \varepsilon_{t_i}.$$

It is easy to see that

$$\lim_{n\to\infty}\mathcal{F}^{(all)}(Y) = \mathcal{IV}_T\ +\ (\text{bias term}) \tag{13}$$

since the assumption about independency between the process X_{t_i} and the noise ε_{t_i}.

3.3 Bias Correction

We propose the bias correction method in the same way of Zhang *et al* (2004) since the amount of the bias in the estimation by Fourier method is the same as in the estimation by realized volatility.

First we use all observations to obtain the estimates using Fourier series analysis described in the previous subsection. Second, the average estimator is constructed by averaging the subgrid estimators $\mathcal{F}^{(k)}(Y)$ across K grids. Then the bias corrected estimator can be defined as

$$\widehat{\mathcal{IV}}_T^{(F)} = \mathcal{F}^{(avg)}(Y) \; - \; \frac{\bar{n}}{n}\mathcal{F}^{(all)}(Y). \qquad (14)$$

To see the properties of estimators described in this section, we conduct the series of Monte Carlo in the next section.

4. MONTE CARLO EXPERIMENT

4.1 Design

We conduct the series of Monte Carlo experiment to compare the properties of the estimates described in the previous section. The four cases are considered here. There is no market microstructure noise in case 1 (evenly sampled) and case 2 (unevenly sampled). We consider the microstructure noise effect in case 3 (evenly sampled) and case 4 (unevenly sampled).

The stochastic volatility model of Heston (1993) is used as data generating process which is the same model as in Zhang *et al* (2004).

$$dX_t \;=\; (\mu - \sigma_t^2/2)dt \;+\; \sigma_t dW_t^{(1)} \qquad (15)$$
$$d\sigma_t^2 \;=\; \kappa(\alpha - \sigma_t^2)dt \;+\; \gamma\sigma_t dW_t^{(2)} \qquad (16)$$

The parameters $(\mu, \kappa, \alpha, \gamma)$ and the correlation coefficient ρ between $W_t^{(1)}$ and $W_t^{(2)}$ are assumed to be constant. We set the same values for the parameters as in Zhang *et al* (2004), $\mu = 0.05$, $\kappa = 5$, $\alpha = 0.04$, $\gamma = 0.5$ and $\rho = -0.5$. Further we assume that the microstructure noise ε_t is Gaussian and small and we set $(\mathrm{E}[\varepsilon_t^2])^{1/2} = 0.0005$. We estimate the volatility over T = 1 day using the sample path. Actually T is equal to 1/252 since parameter values are annualized. One day consists of 6.5 hours of open trading. There is 23400 sec(=6.5 h$\times$ 60 min $\times$ 60 sec) in one day.

We generate two types of high frequency sample paths. One is evenly sampled and the other is unevenly sampled. The sampling interval is one second in the evenly sampled case and the sampling interval is treated as random in the unevenly sampled case. The random sampling interval is drawn from an exponential distribution with mean 10 sec. We set $c = 1$ for determining the number of subgrids K.

We use the relative error

$$\frac{estimate \; - \; \int_0^T \sigma_t^2 dt}{\int_0^T \sigma_t^2 dt} \qquad (17)$$

to evaluate the estimate.

We summarize the sample properties of the relative error such as mean, standard deviation and mean square error in each Table.

The estimates to be compared are the realized volatility with full sample $[Y,Y]_T^{(all)}$, the bias corrected realized volatility $\widehat{\mathcal{IV}}_T^{(RV)}$, the estimator based on Fourier analysis $\mathcal{F}^{(all)}$ and its bias correction $\widehat{\mathcal{IV}}_T^{(F)}$.

4.2 No Microstructure Noise

Table 1 shows that the returns are sampled evenly spaced at high frequency without microstructure noise. The realized volatility $[Y,Y]_T^{(all)}$ is almost same as the true integrated volatility since we can directly observe the process X_{t_i}. The bias corrected estimator $\widehat{\mathcal{IV}}_T^{(RV)}$ introduces some fluctuation and the estimator based on Fourier analysis is well-behaved. Figure 1 and Figure 2 show the distributions of the relative errors of the bias corrected realized volatility and the bias corrected estimate based on Fourier analysis proposed in this paper.

Table 1. Evenly sampled case

	$[Y,Y]_T^{(all)}$	$\widehat{\mathcal{IV}}_T^{(RV)}$	$\mathcal{F}^{(all)}$	$\widehat{\mathcal{IV}}_T^{(F)}$
Mean	0.00	10.38	2.23	2.23
SD	0.00	14.25	3.37	3.36
MSE	0.00	310.75	16.31	16.27

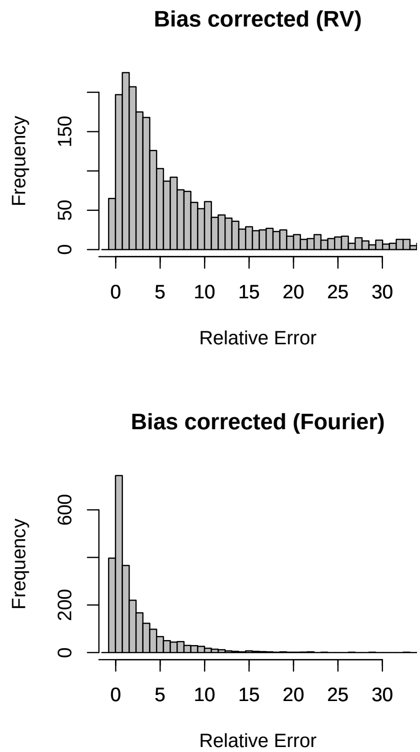

Figure 1. Evenly sampled with No Microstructure Noise

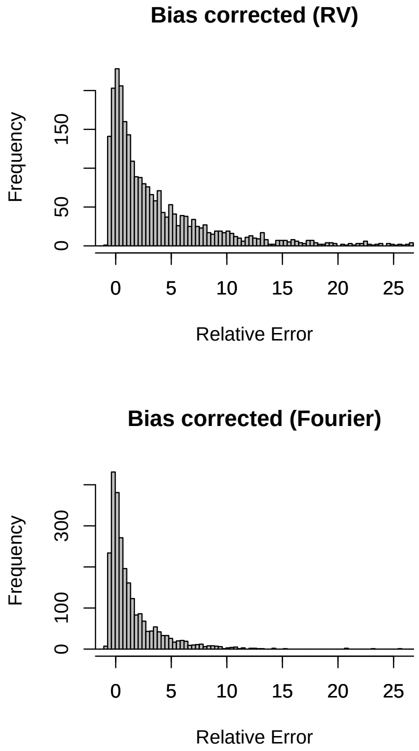

Figure 2. Unevenly sampled with No Microstructure Noise

In the unevenly sampled case, we can see a similar characteristic as in that of the evenly sampled case in Table 2. This fact is also found in Barucci and Renò (2002). The results in the case of no microstructure noise show that the bias corrected estimate proposed in this paper works well even when there is no microstructure noise.

Table 2. Unevenly sampled case

	$[Y,Y]_T^{(all)}$	$\widehat{\mathcal{IV}}_T^{(RV)}$	$\mathcal{F}^{(all)}$	$\widehat{\mathcal{IV}}_T^{(F)}$
Mean	0.000	4.001	1.415	1.397
SD	0.050	5.877	2.398	2.383
MSE	0.002	50.544	7.755	7.632

4.3 Noisy High Frequency Data

The properties of estimates when the observed return process is contaminated with the microstructure noise are summarized in Table 3 and Table 4. It is obvious that the realized volatility commonly used fails to estimate the true integrated volatility when the returns are sampled at high frequency with microstructure noise. Figure 3 shows the distributions of the relative errors of the bias corrected realized volatility and the bias corrected estimate based on Fourier analysis proposed in this paper. The distributions of the relative errors of the estimates without bias correction are not reported here. Their properties can be seen from Table 3 and Table 4. The proposed estimation method in this paper has the smallest mean squared error among other estimators. This characteristic is also found in the unevenly sampled case. The sample properties are summarized in Table 4. The number of sample is smaller than that of evenly sampled case since the sampling interval is randomly generated by the exponential distribution with mean 10 sec. Figure 4 shows that the estimate by Fourier method with bias correction is more reliable than the method proposed

Table 3. Evenly sampled case

	$[Y,Y]_T^{(all)}$	$\widehat{\mathcal{IV}}_T^{(RV)}$	$\mathcal{F}^{(all)}$	$\widehat{\mathcal{IV}}_T^{(F)}$
Mean	1.1×10^4	4.540	15.042	1.964
SD	7.9×10^3	13.541	10.691	3.445
MSE	1.8×10^8	203.968	340.541	15.727

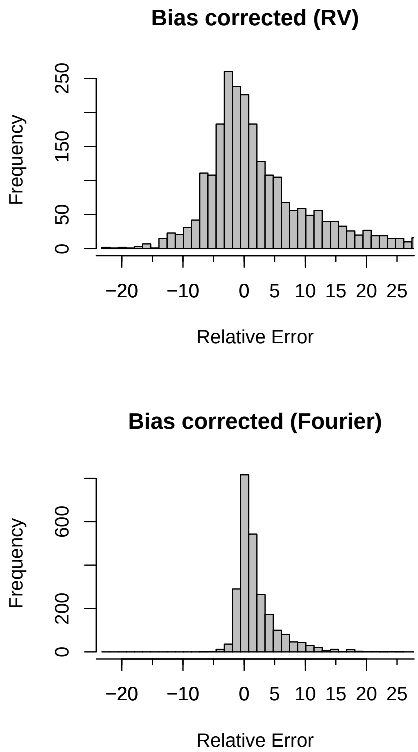

Figure 3. Evenly sampled with Microstructure Noise

Table 4. Unevenly Sampled case

	$[Y,Y]_T^{(all)}$	$\widehat{\mathcal{IV}}_T^{(RV)}$	$\mathcal{F}^{(all)}$	$\widehat{\mathcal{IV}}_T^{(F)}$
Mean	1.0×10^3	1.609	7.407	1.638
SD	7.3×10^2	5.737	5.487	2.535
MSE	1.7×10^6	35.507	84.961	9.109

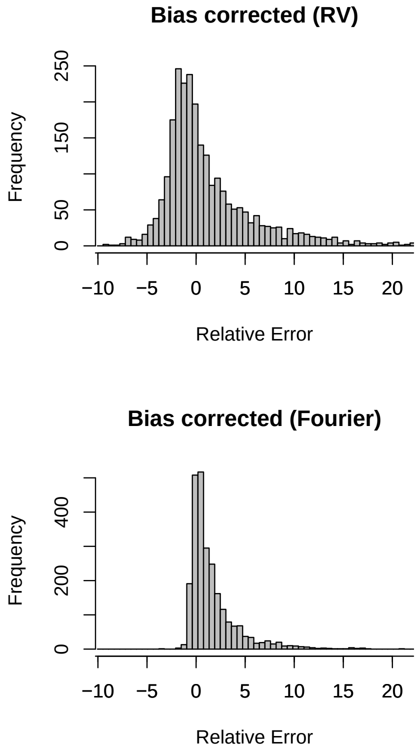

Figure 4. Unevenly sampled with Microstructure Noise

in Zhang *et al* (2004).

5. CONCLUSIONS

This paper has examined the properties of estimators of the integrated volatility when the market microstructure noise exists. We find that the realized volatility commonly used fails to estimate the true integrated volatility and the estimator based on Fourier analysis is also biased when the returns are sampled at high frequency with microstructure noise. We propose the bias corrected estimator that based on Fourier analysis. The proposed bias corrected estimator has the smallest mean squared error among other estimators and works well even when the market microstructure noise does not exist.

6. ACKNOWLEDGEMENT

This research was partially supported by the Ministry of Education, Culture, Sports, Science and Technology: Grants-in-Aid for Scientific Reserach (A)(1),15200022, 2005.

7. REFERENCES

Barucci, E. and R. Renò (2002), On measuring volatility of diffusion processes with high frequency data, *Economics Letters*, **74**, 371-378.

Heston, S.(1993), A Closed-Form Solution for Options with Stochastic Volatility with Applications to Bounds and Currency Options, *Review of Financial Studies*,**6**, 327-343.

Malliavin, P. and M.E. Mancino (2002), Fourier series method for measurement of multivariate volatilities, *Financial Stochastics*, **6**, 49-61.

Zhang, L., Mykland, P. and Y. Aït-Sahalia (2004), A Tale of Two Time Scales: Determining Integrated Volatility with Noisy High-Frequency Data, forthcoming in *Journal of the American Statistical Association*