Università degli Studi di Pisa
Dipartimento di Statistica e Matematica
Applicata all'Economia

Report n. 200

On Measuring volatility of diffusion processes with high frequency data

Emilio Barucci e Roberto Renò

Pisa, Dicembre 2000

- Stampato in Proprio -

Via Cosimo Ridolfi, 10 – 56124 PISA – Tel. Segr. Amm. 050 945231 Segr. Stud. 050 945317 Fax 050 945375
Cod. Fisc. 80003670504 - P. IVA 00286820501 - Web http://statmat.ec.unipi.it/ - E-mail: dipstat@ec.unipi.it

On Measuring Volatility of Diffusion Processes with High Frequency Data

Emilio Barucci
Dipartimento di Statistica e Matematica Applicata all'Economia, Università di Pisa

Roberto Renò
Scuola Normale Superiore, Pisa and INFM, Sezione di Palermo, Italy

December 13, 2000

Keywords: Volatility, Forecasting, High frequency data, SR-SARV(1), GARCH models.
JEL Classification (2000): C15, C22, C53.

Abstract

We analyze a recently proposed method to estimate the volatility of a diffusion process with high frequency data. The method is based on Fourier analysis, all observations are included in the computation without any data manipulation. By Monte Carlo experiments, we evaluate its performance in measuring volatility under the assumption that the asset price evolves according to models belonging to the SR-SARV(1) class, which includes GARCH(1,1) as a particular case. We compare the performance of the method to that associated with the cumulative squared intraday returns. The forecasting capability of the models is also evaluated.

1 Introduction

An unbiased estimator of the daily asset price's volatility is provided by the squared daily return, an estimate that has been recognized to be very noisy. To overcome this problem, in [8, 17, 1] it is proposed to use intraday returns in measuring daily volatility, paralleling the use of daily returns in computing monthly volatility pioneered by [16]. Indeed, a much more precise estimate of the daily volatility is obtained by means of the cumulative squared intraday returns. By employing this method to measure ex post volatility, standard GARCH models perform well in forecasting volatility compared to the poor performance obtained with the daily squared return estimate, see [1]. To apply this method, an equally spaced time series is built from unevenly sampled high frequency observations by interpolating the data, thus reducing the number of data and introducing distortions caused by non-synchronous trading, see [9].

In this note we apply the method proposed in [10] to compute the time series volatility. This method is based on Fourier analysis and therefore on the *integration* of the time series rather than on its *differentiation*, as a consequence all the observations are employed without any manipulation of the data. To evaluate the method, we compare its volatility estimate to that obtained through the cumulative squared intraday returns when the asset price dynamics is governed by a model belonging to the $SR - SARV(1)$ class, studied in detail in [12]. The comparison is done through Monte Carlo experiments, and the forecasting performance of the model is addressed. In [2] the forecasting performance of GARCH models is evaluated when this method is employed to measure the ex post volatility of the DM-$ and Y-$ exchange rate.

We show that the cumulative squared intraday returns estimator is biased because of the interpolation procedure used to build an equally spaced time series. The method based on Fourier analysis is almost unbiased and renders a better forecasting performance for the models analyzed.

2 The Method

In what follows we will concentrate on univariate diffusion processes of the kind:

$$dp(t) = \sigma(t)dW(t) + \mu(t)dt, \tag{1}$$

where $W(t)$ is a Brownian motion, $\mu(t), \sigma(t)$ are allowed to be random time dependent functions. The methodology developed in [10] allows us to compute the volatility through Fourier analysis. Here we recall briefly how this volatility estimator is constructed. We normalize the time window [0,T] to $[0, 2\pi]$. We start from the Fourier coefficients of dp:

$$\begin{aligned} a_0(dp) &= \tfrac{1}{2\pi}\int_0^{2\pi} dp(t) \\ a_k(dp) &= \tfrac{1}{\pi}\int_0^{2\pi} \cos(kt)dp(t) \\ b_k(dp) &= \tfrac{1}{\pi}\int_0^{2\pi} \sin(kt)dp(t) \quad k \geq 1. \end{aligned} \tag{2}$$

In [10, Theorem 1.2] it is shown that the Fourier coefficients of $\sigma^2(t)$ can be computed by means of the Fourier coefficients of dp according to

$$a_0(\sigma^2) = \lim_{n\to\infty} \frac{\pi}{n+1-n_0} \sum_{s=n_0}^{n} \frac{1}{2}(a_s^2(dp) + b_s^2(dp)) \tag{3}$$

$$a_k(\sigma^2) = \lim_{n\to\infty} \frac{2\pi}{n+1-n_0} \sum_{s=n_0}^{n} a_s(dp)a_{s+k}(dp) \tag{4}$$

$$b_k(\sigma^2) = \lim_{n\to\infty} \frac{2\pi}{n+1-n_0} \sum_{s=n_0}^{n} b_s(dp)b_{s+k}(dp), \tag{5}$$

then classical results of Fourier theory allows us to reconstruct $\sigma^2(t)$ $\forall t \in [0, 2\pi]$ by the Fourier-Féjer inversion formula:

$$\sigma^2(t) = \lim_{n\to\infty} \sum_{k=0}^{n} (1 - \frac{k}{n}) \cdot (a_k(\sigma^2)\cos(kt) + b_k(\sigma^2)\sin(kt)). \tag{6}$$

In multivariate setting the volatility matrix can be computed in a similar way. Given a time series of N, not necessarily evenly sampled, observations $(t_i, p(t_i))$, $i = 1, \dots N$, we will compute the integrals in (2) through *integration by parts*:

$$a_k(du) = \frac{1}{\pi}\int_0^{2\pi} \cos(kt)dp(t) = \frac{p(2\pi) - p(0)}{\pi} - \frac{k}{\pi}\int_0^{2\pi} \sin(kt)p(t)dt. \tag{7}$$

To implement the method, we need an assumption on how data are connected. Our choice is $p(t)$ equal to $p(t_i)$ in the interval $[t_i, t_{i+1}]$ (piecewise constant). Then, the integral in (7) in the interval $[t_i, t_{i+1}]$ becomes

$$\frac{k}{\pi}\int_{t_i}^{t_{i+1}} \sin(kt)p(t)dt = p(t_i)\frac{k}{\pi}\int_{t_i}^{t_{i+1}} \sin(kt)dt = p(t_i)\frac{1}{\pi}(cos(kt_i) - cos(kt_{i+1})). \tag{8}$$

The smallest wavelength that can be evaluated is twice the smallest distance between two consecutive prices; in the case of equally spaced data, it will correspond to $k = N/2$ (Nyquist frequency): this frequency is the largest that can be evaluated in the expansion (3). In the integration by parts formula (7), the constant term $(p(2\pi) - p(0))/2\pi$ appears in every coefficient; this term can be set to zero by adding in equation (1) the drift term $-\frac{p(2\pi)-p(0)}{2\pi}dt$. This change of variables will not affect the volatility estimate and will remove a possible source of bias.

3 Monte Carlo Simulations

Let $p(t) = \log S(t)$, where $S(t)$ is a generic asset price, and $r_t = p(t) - p(t-1)$. We start with the assumption that the asset price follows the continuous time GARCH model proposed in [14]:

$$\begin{aligned} dp(t) &= \sigma(t)dW_1(t) \\ d\sigma^2(t) &= \theta(\omega - \sigma^2(t))dt + \sqrt{2\lambda\theta}\sigma^2(t)dW_2(t), \end{aligned} \tag{9}$$

where W_1, W_2 are independent Brownian motions. This model is closed under temporal aggregation in a weak sense, see [4], and its discrete time analogous is given by:

$$\begin{aligned} r_t &= \sigma_t\epsilon_t \\ \sigma_{t+1}^2 &= \psi + \alpha \cdot r_t^2 + \beta \cdot \sigma_t^2 \end{aligned} \tag{10}$$

where ϵ_t are i.i.d. Normal distributed random variables. The exact relation between (ψ, α, β) and $(\theta, \omega, \lambda)$ is derived in [4]. We generate high frequency unevenly sampled observations in the following way: we first simulate one day of trading by discretizing (9) with a time step of one second, for a total of 86400 observations per day. Then we extract the observation times in such a way that the time differences are drawn from an exponential distribution with mean equal to $\tau = 45$ seconds. As a result, we will have a data set $(t_k, p(t_k), k = 1, \dots, N)$ with t_k unevenly sampled, and $\sigma(t)$ recorded for every t.

In [1] the integrated volatility $\int_0^1 \sigma^2(t+\tau)d\tau$ is measured it by the sum of squared intra-day returns:

$$\hat{\sigma}^2(m) = \sum_{i=2}^{m}(p(\frac{i}{m}) - p(\frac{i-1}{m}))^2. \tag{11}$$

As described in [13], if p is not observed at time i/m, $p(i/m)$ is given by the linear interpolation of two adjacent observations (one before and one after the time i/m). Theoretically, by increasing the frequency of observations, an arbitrary precision in the estimate of the integrated volatility can be reached. This could not be our case because of the interpolation procedure. In most of the papers estimating volatility with high frequency data, e.g. see [1, 11], (11) is computed with $m = 288$, corresponding to five-minutes returns. In our simulation setting we will also compute it with $m = 144$, corresponding to ten-minutes returns, and $m = 720$ corresponding to two-minutes returns; we don't increase further m since the mean time between transactions is 45 seconds.

The method proposed in [10] provides us with an estimator of the integrated volatility. Integrating equation (6) between 0 and 2π, we have

$$\int_0^{2\pi} \sigma^2(t)dt = 2\pi a_0(\sigma^2), \tag{12}$$

where $a_0(\sigma^2)$ is given by (3). Note that with this method we use all the observations and no data manipulation is needed.

We will evaluate the performance of the estimators (12) and (11) with $m = 144, 288, 720$ by the statistics:

$$\mu = \boldsymbol{E}\, \frac{\int_0^1 \sigma^2(s)ds - \hat{\sigma}^2}{\int_0^1 \sigma^2(s)ds}, \quad std = \left[\boldsymbol{E}\left(\frac{\int_0^1 \sigma^2(s)ds - \hat{\sigma}^2}{\int_0^1 \sigma^2(s)ds}\right)^2\right]^{\frac{1}{2}},$$

where $\hat{\sigma}^2$ is the estimate and $\int_0^1 \sigma^2(s)ds$ is the integrated volatility generated in a simulation, whose value is known in our simulation setting. We will also evaluate the forecasting performance of the model (10), when ex-post volatility is measured by $\hat{\sigma}^2$. This is done by means of the coefficient of multiple determination R^2 of the linear regression

$$\hat{\sigma}^2_{t+1} = a + b \cdot \sigma^2_{t+1} + \varepsilon_t. \tag{13}$$

As a reference time step to build the time series through interpolation we choose five minutes. To evaluate the effect of the interpolation procedure described above, we also consider a ten and a two minutes time step. We recall that without manipulating the data, we should observe smaller μ and std when increasing the frequency. Figure 1 shows the results on simulated time series with $\alpha = 0.7, \beta = 0.25, \psi(1 - \alpha - \beta) = 1$. The ten-minutes estimator provides a downward biased estimate of the integrated volatility, with a standard deviation larger than the bias. The five-minutes is also downward biased, with a standard deviation of the same order of the bias in mean. Increasing further the frequency the estimator is characterized by less variance but a larger bias is observed. This effect can only be due to the data manipulation procedure described above and therefore it can be linked to non-synchronous trading, see [9]. The Fourier estimator is characterized by the smallest bias in mean and by a variance smaller than that of the 5-10 minutes estimate and slightly larger than that of the 2 minutes estimate. To check the robustness of these results, we repeated the Monte Carlo experiments on a grid of values $(\alpha, \beta, \psi = (1-\alpha-\beta)^{-1})$ with 2 and 5 minutes returns. The results, reported in Table 1, can be summarized as follows: the estimator (11) turns out to be downward biased ($\mu > 0$), with a bias increasing with m, while the bias of the Fourier estimator is almost null. If m is chosen in such a way that the bias of (11) is less than its standard deviation, then the Fourier estimate provides a smaller standard deviation. Analyzing the forecasting capability of the discrete time GARCH model (10) we have that the Fourier estimate renders a better performance than the classical estimator

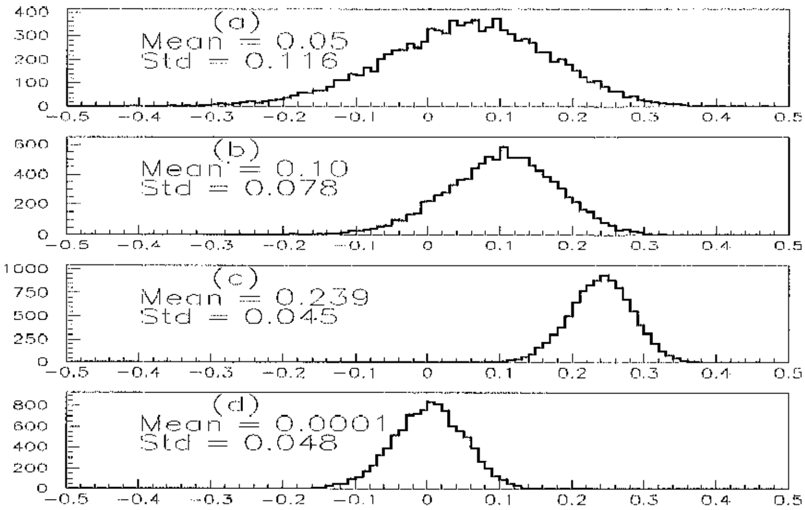

FIGURE 1: Distribution of $\frac{\int_0^1 \sigma^2(t)dt-\hat{\sigma}^2}{\int_0^1 \sigma^2(t)dt}$, where $\hat{\sigma}^2$ are three different estimators of the integrated volatility: (a) estimator (11) with $m = 144$; (b) estimator (11) with $m = 288$; (c) estimator (11) with $m = 720$; (d) Fourier estimator (12). The distribution is computed with 10,000 "daily" replications.

computed with 2 and 5 minutes returns. . For completeness, we checked these results on the following autoregressive diffusion models:

$$\text{NGARCH}(1,1) \text{ model, } [7]: \sigma_{t+1}^2 = \psi + \alpha(r_t - \gamma\sigma_t)^2 + \beta \cdot \sigma_t^2 \tag{14}$$

$$\text{GJR} - \text{GARCH model, } [6]: \sigma_{t+1}^2 = \psi + \alpha \cdot r_t^2 + \beta \cdot \sigma_t^2 + \delta\theta(-r_t)r_t^2 \tag{15}$$

$$\text{EGARCH model, } [15]: \log(\sigma_{t+1}^2) = \psi + \alpha\sigma_t \cdot (|r_t| + \gamma \cdot r_t) + \beta \cdot \log(\sigma_t^2) \tag{16}$$

where the θ-function is given by $\theta(x) = 1$ if $x > 0$ and $\theta(x) = 0$ if $x \leq 0$. All these models fall in the general class of the $SR - SARV(1)$ models, which have the nice property to be closed under temporal aggregation, see [12, 5], so that a continuous diffusion process exists, with the property that the corresponding discrete process is its exact discretization. We simulated these processes with the parameters estimated in [5] and reported in Table 2. Table 2 reports also μ, std, R^2 for each model. The results in Table 2 confirm those obtained with the continuous GARCH model, i.e. the Fourier estimator has a smaller bias and an higher precision than (11); moreover, it provides a better forecasting performance for the model at hand.

4 Conclusions

The importance of measuring volatility with high frequency data has been pointed out repeatedly in the literature. Computing volatility with the cumulative squared intraday returns, an equally spaced time series is built by interpolating the observations, this procedure may

$\beta \downarrow$ $\alpha \rightarrow$		0.05			0.1			0.15			0.2			0.25		
		2′	5′	F	2′	5′	F	2′	5′	F	2′	5′	F	2′	5′	F
0.5	μ	23.9	10.0	0.1	23.9	10.0	0.1	23.9	10.0	0.1	23.9	10.0	0.1	23.9	10.0	0.1
	std	4.4	7.7	4.9	4.5	7.8	5.0	4.5	7.9	5.0	4.6	7.9	5.1	4.6	8.0	5.1
	R^2	4.92	4.79	4.98	9.58	9.44	9.70	14.0	13.9	14.2	18.6	18.5	18.8	23.8	23.8	24.0
0.6	μ	23.9	10.0	0.1	23.9	10.0	0.1	23.9	10.0	0.1	23.9	10.0	0.1	23.9	10.0	0.1
	std	4.4	7.7	4.9	4.5	7.8	4.9	4.5	7.8	5.0	4.5	7.8	5.0	4.5	7.8	5.0
	R^2	6.79	6.62	6.88	13.8	13.7	14.0	21.9	21.8	22.1	32.5	32.5	33.0	44.6	44.5	45.6
0.7	μ	23.9	10.0	0.1	23.9	10.0	0.1	23.9	10.0	0.1	23.9	10.0	0.1	23.9	10.0	0.1
	std	4.4	7.7	4.9	4.4	7.7	4.9	4.4	7.7	4.9	4.5	7.8	4.9	4.5	7.8	4.9
	R^2	9.67	9.43	9.79	21.5	21.3	21.7	38.2	37.9	38.7	51.3	51.1	52.2	48.2	48.1	48.8
0.8	μ	23.9	10.0	0.1	23.9	10.0	0.1	23.9	10.0	0.1						
	std	4.4	7.6	4.9	4.4	7.7	4.9	4.4	7.7	4.9						
	R^2	15.2	14.7	15.3	36.9	36.5	37.2	54.3	53.8	54.5						
0.9	μ	23.9	10.0	0.1												
	std	4.4	7.6	4.8												
	R^2	30.8	29.9	30.9												

TABLE 1: μ, std, R^2 (multiplied by 100) for the three estimators: (11) with $m = 720$ denoted by $2'$, (11) with $m = 188$ denoted by $5'$ and (12) denoted by F, on a grid of values for (α, β) in (10), and $\psi \cdot (1 - \alpha - \beta) = 1$. All the values are computed with 10000 "daily" replications.

	GJR $\psi = 0.0587,\ \alpha = 0.0312$ $\beta = 0.8275,\ \delta = 0.1271$			NGARCH $\psi = 0.0554,\ \alpha = 0.0952$ $\beta = 0.8001,\ \gamma = 0.6048$			EGARCH $\psi = -0.1491,\ \alpha = 0.1786$ $\beta = 0.9512,\ \gamma = -0.4815$		
	2′	5′	F	2′	5′	F	2′	5′	F
μ	23.9	10.0	-0.3	23.9	10.0	-0.5	23.9	10.0	-0.2
std	4.4	7.6	4.9	4.4	7.6	5.0	4.4	7.6	4.9
R^2	37.8	37.0	37.9	47.3	46.4	47.4	46.6	45.5	46.8

TABLE 2: μ, std, R^2 (multiplied by 100) for the three estimators, (11) with $m = 720$ denoted by $2'$, (11) with $m = 288$ denoted by $5'$ and (12) denoted by F, for the diffusion processes (14-15-16) with the reported parameter values. All the values are computed with 10000 "daily" replications.

introduce biases because of non-synchronous trading. In this note we have shown through Monte Carlo simulations that this estimator is biased: increasing the frequency, the variance of the estimator is reduced but the bias in mean increases. The method based on Fourier analysis proposed in this paper does not require any manipulation of the data. Relying upon Monte Carlo simulations of models belonging to the SR-SARV(1) class, which includes the GARCH(1,1), we have shown that this estimator is almost unbiased and its variance is smaller than that of the cumulative squared intraday returns, when the latter is chosen with a reasonable bias in mean. Moreover, when the Fourier method is employed to evaluate the forecasting performance of the models, their performance is better than that obtained by employing the cumulative intraday squared returns.

References

[1] Andersen, T. and Bollerslev, T. (1998) Answering the Skeptics: Yes, Standard Volatility Models do Provide Accurate Forecasts. *International Economic Review*, 39: 885-905.

[2] Barucci, E. and Renò, R. (2000) On measuring volatility and the GARCH forecasting performance. Manuscript, Università di Pisa and Scuola Normale Superiore, Pisa.

[3] Drost, F. and Nijman, T. (1993) Temporal aggregation of GARCH processes. *Econometrica*, 61: 909-927

[4] Drost, F. and Werker, B. (1996) Closing the GARCH Gap: Continuous time GARCH Models. *Journal of Econometrics*, 74: 31-57.

[5] Duan, J.C. (1997) Augmented GARCH(p,q) process and its diffusion limit. *Journal of Econometrics*, 79: 97-127.

[6] Glosten, L., Jagannathan, R., Runkle, D. (1989), On the Relation between the Expected Value of the Volatility of the Nominal Excess Return on Stocks. *Journal of Finance*, 48: 1779-1801.

[7] Engle, R., Ng, V. (1993), Measuring and testing the impact of news on volatility. *Journal of finance*, 48: 1749-1778.

[8] Hsieh, D. (1991) Chaos and nonlinear dynamics: application to financial markets. *Journal of Finance*, 46: 1839-1877.

[9] Lo, A. and MacKinlay, C. (1990) An Econometric Analysis of Nonsynchronous Trading. *Journal of Econometrics*, 45: 181-211.

[10] P. Malliavin, M. Mancino, (2000) Fourier series method for Measurement of Multivariate Volatilities. Forthcoming *Finance & Stochastics.*

[11] Martens, M. (2000), Measuring and forecasting stock market volatility using high-frequency data. Manuscript, University of New South Wales.

[12] Meddahi, N. and Renault, E. (2000), Temporal Aggregation of Volatility Models. Scientific Series 2000s-22, CIRANO, Montreal, Canada.

[13] Muller, P. et al. (1990), Statistical Study of Foreign Exchange Rates, Empirical Evidence of a Price Change Scaling Law and Intraday Analysis. *Journal of Banking and Finance*, 14, 1189-1208.

[14] Nelson, D. (1990) Arch Models as Diffusion Approximations. *Journal of Econometrics*, 45: 7-38.

[15] Nelson, D. (1991) Conditional heteroskedasticity in asset returns: a new approach. *Econometrica*, 59: 347-370.

[16] Schwert, G. (1989) Why does stock market volatility change over time? *Journal of Finance*, 44: 1115-1153.

[17] Schwert, G. (1990) Stock market volatility. *Financial Analyst Journal*, 46: 23-34.