Value at Risk with High Frequency Data

Emilio Barucci[1] and Roberto Renò[2]

[1] Dipartimento di Statistica e Matematica Applicata all'Economia, Università di Pisa

[2] Scuola Normale Superiore, Pisa

Abstract. It has been recently shown in the literature that the use of high frequency data allows very precise estimates of the daily integrated volatility for a continuous time stochastic process. A good forecasting model for daily integrated volatility is crucial for VaR estimates. Traditional models regard volatility as a latent factor; in this paper we model it as an observable quantity through an $AR(n)$ model estimated by ordinary least squares. In spite of its simplicity, this model performs better than traditional models (GARCH(1,1) and Riskmetrics).

1 Introduction

The role played by volatility in most financial applications is crucial, expecially in risk management, where Value at Risk (VaR) estimates are mandatory for regulatory reasons and asset allocation decisions.

In the recent years, literature focused on the role of integrated volatility, see [1,6,7]; the importance of integrated daily volatility in VaR applications relies on two facts. First of all, it has been shown that it is possible to measure it by using intra-day data with very good precision, see [1,8,9], paralleling the use of daily returns in computing monthly volatility, see [21–23]. Second, empirical studies ([4,12]) showed that the distribution of returns divided by the square root of the integrated volatility can be well approximated by a Gaussian distribution with zero mean and variance equal to one. This fact means that VaR estimates are linked to integrated volatility forecasting, since the quantiles of the return distribution can be extracted by a Gaussian distribution with zero mean and variance given by the integrated volatility.

On the other hand, persistence properties displayed by volatility suggest that daily volatility can be forecasted with reliable precision. Typically, volatility models regard it as a latent factor which drives asset prices-returns (ARCH, GARCH models). This is the approach followed also in [10], where high frequency data are used to estimate the dynamical model for latent volatility (FIGARCH model) and to compute VaR.

In this paper, we will model directly the integrated volatility as an observable quantity through a simple $AR(n)$ model. A similar approach has been proposed by [14] and [5] in a multivariate setting. The main difference is provided by the computation method of the integrated volatility. In the above papers, integrated volatility is computed by using an equally spaced high frequency time series (typically five minute returns) as the sum of squared intraday logarithmic returns. Our procedure instead is based on the Fourier analysis methodology proposed in [17].

We will compare the performance of this model to that of the GARCH(1,1) model and to that of the Riskmetrics model proposed in [16], which is very popular among practitioners. We will show that, though the $AR(n)$ is quite simple, it performs better than traditional models in forecasting daily volatility. Our findings suggest that constructing directly a model for volatility based on the measurements of daily integrated volatility, instead of modeling volatility as a latent factor, can be a profitable idea.

The remainder of the paper is organized as follows: Section 2 describes the method to compute daily volatility and the three models we are going to compare; Section 3 presents the results of the comparison and Section 4 concludes.

2 Measuring daily volatility

In what follows, we will deal with 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. Our task is to measure and forecast the so-called integrated volatility, defined as

$$\hat{\sigma}_t^2 = \int_t^{t+1} \sigma^2(s)ds, \tag{2}$$

being one day the time unit. Recently the importance of using intra-day data in order to measure (2) has been repeatedly stressed. In [1] it is shown that using the sum of five-minute squared returns as a volatility estimate, much better results are obtained than using the simple daily squared return. This method has been employed to measure daily volatility of exchange rate time series ([2,4,19]), stock prices ([3]) and index future prices ([18]). On the same topic, see also [11,12].

In this paper, instead of the 5-minutes cumulative squared returns, we will use the integrated volatility estimator proposed in [17] which has been studied in [8,9] where it is shown that it performs better than the estimator proposed in [1]. Here we recall briefly how this volatility estimator is constructed; the interested reader is referred to [17] for details. We normalize the time window [0,T] where the time series is observed to $[0, 2\pi]$. We compute the Fourier coefficients of dp:

$$\begin{array}{l} a_0(dp) = \frac{1}{2\pi}\int_0^{2\pi} dp(t) \\ a_k(dp) = \frac{1}{\pi}\int_0^{2\pi} \cos(kt)dp(t) \\ b_k(dp) = \frac{1}{\pi}\int_0^{2\pi} \sin(kt)dp(t) \quad k \geq 1. \end{array} \tag{3}$$

In [17, 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

$$a_0(\sigma^2) = \lim_{n\to\infty} \frac{\pi}{n+1-n_0} \sum_{s=n_0}^{n} \frac{1}{2}\left(a_s^2(dp) + b_s^2(dp)\right) \tag{4}$$

$$a_k(\sigma^2) = \lim_{n\to\infty} \frac{2\pi}{n+1-n_0} \sum_{s=n_0}^{n} a_s(dp)a_{s+k}(dp) \tag{5}$$

$$b_k(\sigma^2) = \lim_{n\to\infty} \frac{2\pi}{n+1-n_0} \sum_{s=n_0}^{n} a_s(dp)b_{s+k}(dp), \tag{6}$$

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:

$$\sigma^2(t) = \lim_{n\to\infty} \sum_{k=0}^{n} (1 - \frac{k}{n}) \cdot (a_k(\sigma^2)\cos(kt) + b_k(\sigma^2)\sin(kt)). \tag{7}$$

Integrating (7) between 0 and 2π we obtain our integrated volatility estimator:

$$\int_0^{2\pi} \sigma^2(t)dt = 2\pi a_0(\sigma^2). \tag{8}$$

Given a time series of N, not necessarily evenly sampled, observations $(t_i, p(t_i))$, $i = 1, \ldots N$, we will compute the integrals in (3) through *integration by parts*:

$$a_k(dp) = \frac{1}{\pi}\int_0^{2\pi} \cos(kt)dp(t) = \frac{p(2\pi) - p(0)}{\pi} - \frac{k}{\pi}\int_0^{2\pi} \sin(kt)p(t)dt. \tag{9}$$

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 (9) in the interval $[t_i, t_{i+1}]$ becomes

$$\frac{k}{\pi}\int_{t_i}^{t_{i+1}} \sin(kt)p(t)dt = p(t_i)\frac{k}{\pi}\int_{t_i}^{t_{i+1}} \sin(kt)dt = p(t_i)\frac{1}{\pi}\left(cos(kt_i) - cos(kt_{i+1})\right). \tag{10}$$

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). In the computation of (4) we could be tempted to stop the expansion at $N/2$. However, as shown in [8], microstructure effects, mostly due to a negative short-lived autocorrelation in the price returns (see [13,24]) prevents us to include the largest frequencies. The choice of the frequency at which to stop the expansion (4) is largely an empirical matter. The most appealing feature of this method is that it employs all the tick-by-tick observations in the integrated volatility computation. This allows to increase the precision in the estimate of (2) with respect to the sum of squared intraday returns proposed by [1], see [8]. Moreover, we completely avoid measurement biases due to an aggregating procedure of the data, see [9].

The distribution of returns divided by the square root of the daily integrated volatility can be well approximated by a Gaussian with zero mean and variance equal to one, as shown in the empirical studies in [4,12]. Than the problem of extracting the quantiles of the return distribution, i.e. the problem of calculating VaR, amounts to providing a good forecast of the integrated volatility.

Being able to measure daily volatility with good precision, we will try to model integrated volatility as an observed quantity, instead of a latent factor as in GARCH models. We will try with the simplest possible model for the time evolution of the integrated volatility, i.e. an $AR(n)$ model:

$$\hat{\sigma}_t^2 = \sigma_0^2 + \sum_{i=1}^{n} \alpha_i \hat{\sigma}_{t-i}^2 + \varepsilon_t \tag{11}$$

with $\boldsymbol{E}\left[\varepsilon_t\right] = 0,\ \boldsymbol{E}\left[\varepsilon_t^2\right] = \Sigma^2$. The parameters $\sigma_0^2, \Sigma^2, \alpha_1, \ldots, \alpha_n$ can be estimated by ordinary least squares (OLS) and $\hat{\sigma}_t^2$ is measured by the integrated volatility estimator in (8). We will use (11) to forecast future volatility.

We will compare this very simple volatility forecasting model with two models largely used by practitioners, the GARCH(1,1) model, where future volatility is estimated as:

$$\hat{\sigma}_{t+1}^2 = \psi + \alpha \cdot r_t^2 + \beta \cdot \hat{\sigma}_t^2, \tag{12}$$

where $r(t) = p(t) - p(t-1)$ is the daily return, and the model used by RiskMetrics [16], which estimates the future volatility as a sum of past realizations with exponentially declining weights:

$$\hat{\sigma}_{t+1}^2 = \frac{\displaystyle\sum_{i=0}^{M-1} \lambda^i \left(r_{t-i} - \frac{1}{M} \sum_{i=0}^{M-1} r_{t-i} \right)^2}{\displaystyle\sum_{i=0}^{M-1} \lambda^i} \tag{13}$$

where $\lambda = 0.94$.

Other authors try to model the integrated volatility, measuring it via the sum of squared intraday returns. In [14] it is found that a EMA-HAR model for the integrated volatility performs better than Riskmetrics. In [5] a tri-variate vector auto-regression (VAR), which incorporates long memory effects, is fitted on the DM-$ and Y-$ foreign exchange time-series. These authors choose a polynomial of lag 5, and find that this model performs largely better than GARCH(1,1) and Riskmetrics.

3 Results

The data set under study in this paper is the one-year collection of bid-ask quotes of the Deutshche Mark-U.S. Dollar and Japanese Yen-U.S. Dollar exchange rates, as they appeared on the Reuters screen from October, 1^{st} 1992 to September, 30^{th} 1993. Each quote comes with a time stamp rounded to the nearest even second. We define a trading day to start and end 21.00 GMT; the foreign exchange market is active 24 hours per day. We excluded weekends and days with few activity due to main holidays, see [13]. This data set has been recorded and provided by Olsen & Associates, and has been extensively studied, see for example [20,15]. For any quote, we define the price to be the mid-price between the bid and the ask price. In the expansion (4) we use n as in [8], i.e. $n = 500$ for the DM-$ time series and $n = 160$ for the Y-$ time series. We discard days in which the observations are less than 1000 and 320 respectively; we end up with 258 daily volatility measurements for the DM-$ time series and 259 for the Y-$ time series. We divide our samples in 160 days for in-sample model estimate and the remainder for out-of-sample comparison.

Table 1 shows the OLS in-sample estimates of model (11), together with standard errors, the in-sample R^2 and the R^2 adjusted for degrees of freedom. For the model (12) we use the estimates given in [1], i.e. $\psi = 0.022, \alpha = 0.068, \beta = 0.898$ for the DM-$ time series and $\psi = 0.026, \alpha = 0.104, \beta = 0.844$ for the Y-$ time series. For the model (13) we will use $M = 160$, i.e. the largest M at our disposal.

Table 2 compares the forecasting performance. In spite of its simplicity, model (11) performs considerably better than (12) and (13). For the DM-$ time series, it is already true with $n = 1$; however, by increasing the order of the auto-regressive model, we find better results. We interpret this finding as an evidence of long-memory effects in the volatility evolution. For the Y-$ time series it is necessary to employ $n = 2$, while the best result is obtained with $n = 5$. Figure 1 and 2 show the comparison between the integrated out-of-sample volatilities and the forecasts of the three models for the two exchange rate time series. Also visual inspection confirms that the simple $AR(1), AR(2)$ model does a good job in tracking the volatility time series. We interpret the results of our simple exercise as a confirmation that the use of high frequency data in measuring volatility, and directly modeling the integrated volatility dynamics, can substantially improve volatility forecasting, thus Value at Risk estimates.

4 Conclusions and acknowledgements

The importance of volatility measuring in risk management is increasing. Providing a good forecasting model for daily integrated volatility is essential in calculating reliable VaR estimates. Recently, it has been shown that the use

Table 1. OLS in-sample estimates and R^2 of model (11) for the two time series. Between parenthesis, we report the standard error and the R^2 adjusted by the degrees of freedom.

Model	FX	σ_0^2	α_1	α_2	α_3	α_4	α_5	R^2 (R^2_{adj})
AR(1)	DM-$	0.209	0.728					0.245
		(0.048)	(0.051)					(0.245)
AR(2)	DM-$	0.225	0.575	0.122				0.246
		(0.047)	(0.075)	(0.072)				(0.242)
AR(3)	DM-$	0.194	0.541	-0.018	0.209			0.240
		(0.050)	(0.079)	(0.086)	(0.072)			(0.230)
AR(4)	DM-$	0.159	0.487	0.007	0.064	0.215		0.248
		(0.051)	(0.079)	(0.088)	(0.084)	(0.072)		(0.233)
AR(5)	DM-$	0.152	0.461	-0.004	0.068	0.158	0.096	0.250
		(0.053)	(0.082)	(0.088)	(0.088)	(0.084)	(0.074)	(0.231)
AR(1)	Y-$	0.299	0.381					0.150
		(0.043)	(0.074)					(0.150)
AR(2)	Y-$	0.216	0.276	0.267				0.217
		(0.048)	(0.077)	(0.077)				(0.212)
AR(3)	Y-$	0.204	0.272	0.252	0.047			0.219
		(0.051)	(0.081)	(0.080)	(0.080)			(0.209)
AR(4)	Y-$	0.192	0.264	0.250	0.032	0.052		0.221
		(0.054)	(0.081)	(0.084)	(0.083)	(0.080)		(0.206)
AR(5)	Y-$	0.188	0.262	0.248	0.030	0.046	0.023	0.222
		(0.056)	(0.082)	(0.085)	(0.087)	(0.084)	(0.081)	(0.201)

of high frequency data allows to measure daily integrated volatility with high precision. This suggests that, instead of using models in which the volatility is a latent factor, like for example GARCH(1,1), one can try to model directly the dynamics of integrated volatility.

In this paper we tried to build a simple forecasting $AR(n)$ model for integrated volatility, estimated by OLS, and we showed that it performs considerably better than traditional models. We conclude that modeling directly integrated volatility, measured by high frequency data, can be a promising direction for risk management.

The authors wish to acknowledge Olsen & Associates for providing us the data set under study.

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. Andersen, T., Bollerslev, T., Diebold, F. and Labys, P. (1999) The Distribution of Realized Exchange Rate Volatility. *Journal of American Statistical Association* 96, 42-55.

Table 2. RMSE, MAE for the $AR(n)$ model (till the best n plus one), the GARCH(1,1) model and the Riskmetrics model, for the two time series

Model	RMSE	MAE
DM-$ time series		
GARCH(1,1)	0.272	0.199
RiskMetrics	0.267	0.205
AR(1)	0.256	0.179
AR(2)	0.245	0.170
AR(3)	0.241	0.168
AR(4)	0.242	0.167
AR(5)	0.234	0.160
AR(6)	0.229	0.155
AR(7)	0.230	0.156
Y-$ time series		
GARCH(1,1)	0.508	0.346
RiskMetrics	0.486	0.344
AR(1)	0.497	0.344
AR(2)	0.472	0.340
AR(3)	0.467	0.339
AR(4)	0.461	0.335
AR(5)	0.457	0.332
AR(6)	0.458	0.334

3. Andersen, T. Bollerslev, T., Diebold, F. and Ebens, H. (2000) The Distribution of Stock Return Volatility. *Journal of Financial Economics*, forthcoming.
4. Andersen, T. Bollerslev, T., Diebold, F. and Labys, P. (1999), Exchange Rate Returns Standardized by Realized Volatility are (Nearly) Gaussian. *Multinational Finance Journal*, forthcoming.
5. Andersen, T. Bollerslev, T., Diebold, F. and Labys, P. (2001) Modeling and Forecasting Realized Volatility. NBER Working paper.
6. Barndorff-Nielsen, O.E. and Shephard, N. (2000) Non-Gaussian OU based models and some of their uses in financial economics. CAF Working Paper n.37, Aarhus University.
7. Barndorff-Nielsen, O.E. and Shephard, N. (2000) Econometric Analysis of Realised Volatility and it Use in Estimating Levy Based Non-Gaussian OU Type Stochastic Volatility Models. CAF Working Paper n.72, Aarhus University.
8. Barucci, E. and Renò, R. (2000) On measuring volatility and the GARCH forecasting performance. Manuscript, Università di Pisa and Scuola Normale Superiore, Pisa.
9. Barucci, E. and Renò, R. (2000) On Measuring Volatility of diffusion processes with high frequency data. Manuscript, Università di Pisa and Scuola Normale Superiore, Pisa.
10. Beltratti, A. and Morana, C. (1999) Computing Value-at-Risk with High Frequency Data. *Journal of Empirical Finance*, 6, 431-455.
11. Bollen, B. and Inder, B. (1998), A General Volatility Framework and the Generalized Historical Volatility Estimator. Manuscript, Monash University.

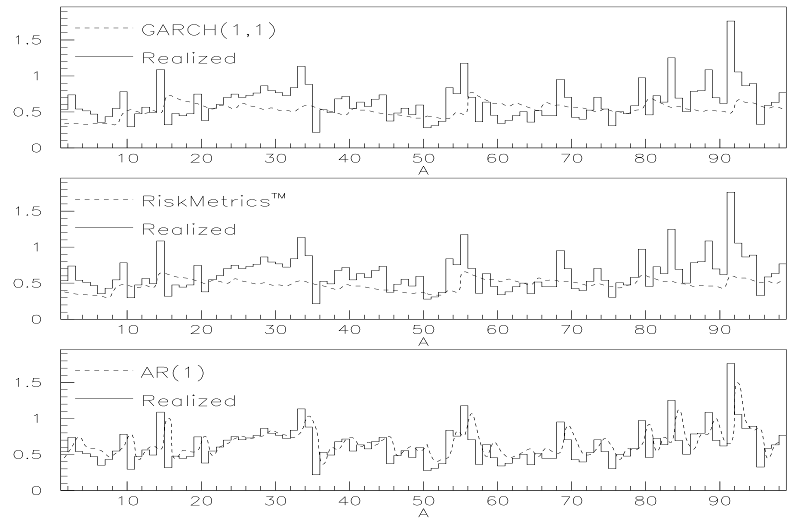

Fig. 1. Out-of-sample measured integrated volatility for the DM-$ time series (solid line) together with the forecast (dashed line) of the GARCH(1,1) model (top), Riskmetrics (center) and $AR(1)$ model (bottom).

12. Bollen, B. and Inder, B. (1999), Ex Post, Unconditional Estimators of Daily Volatility. Manuscript, Monash University.
13. Bollerslev, T., Domowitz, I., Trading patterns and prices in the interbank foreign exchange market. *Journal of Finance*, 48, 1421-1443
14. Corsi, F., Dacorogna, M., Muller, U., Zumbach, G. (2000) High Frequency Data Do Improve Volatility and Risk Estimation. Manuscript, Olsen & Associates.
15. Guillaime, D.M. *et al.* (1997) From the bird's eye to the microscope: A survey of new stylized facts of the intra-daily foreign exchange market. *Finance & Stochastics*, 1, 95-129.
16. J.P.Morgan, Reuters (1996) Riskmetrics - Technical Document, Fourth Edition.
17. P. Malliavin, M. Mancino, (2000) Fourier series method for Measurement of Multivariate Volatilities. Forthcoming *Finance* & *Stochastics*.
18. Martens, M. (2000), Measuring and forecasting stock market volatility using high-frequency data. Manuscript, University of New South Wales
19. Martens, M. (2000), Forecasting daily exchange rate volatility using intraday returns. *Journal of International Money and Finance*, Vol.20, n.1, 1-23.
20. 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.
21. Schwert, G. (1989) Why does stock market volatility change over time? *Journal of Finance*, 44: 1115-1153.

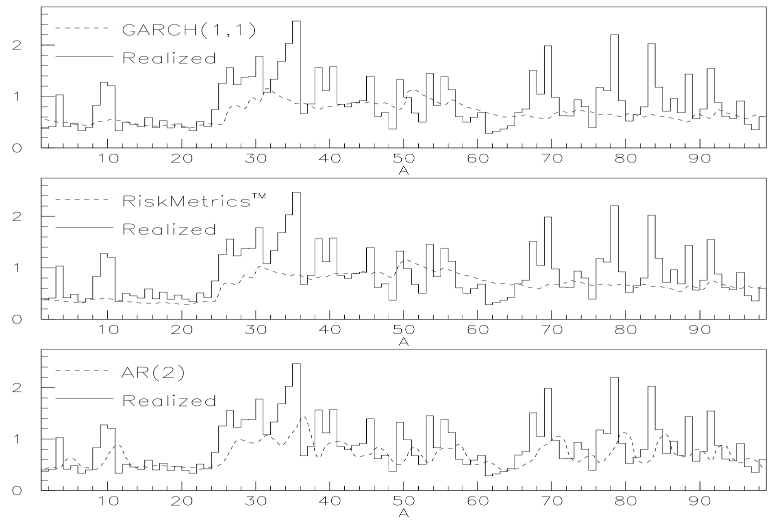

Fig. 2. Out-of-sample measured integrated volatility for the Y-$ time series (solid line) together with the forecast (dashed line) of the GARCH(1,1) model (top), Riskmetrics (center) and $AR(2)$ model (bottom).

22. Schwert, G. (1990) Stock market volatility. *Financial Analyst Journal*, 46: 23-34.
23. Schwert, G. (1998) Stock market volatility: Ten years after the crash. *Brookings-Wharton Papers on Financial Services*, I, 65-114
24. Stoll, H. and Whaley, R. (1990) The Dynamics of Stock Index and Stock Index Futures Returns. *Journal of Financial and Quantitative Analysis*, 25, 441-468