Non-parametric estimation of historical volatility

Article *in* Quantitative Finance · August 2004

DOI: 10.1080/14697680400008692 · Source: RePEc

John Randal
Victoria University of Wellington

Martin Lally
Victoria University of Wellington

Non–parametric volatility estimation

Martin Lally
Victoria University
PO Box 600
Wellington, New Zealand
Martin.Lally@vuw.ac.nz

John Randal
Victoria University
PO Box 600
Wellington, New Zealand
John.Randal@vuw.ac.nz

Peter Thomson
Statistics Research Associates Ltd
PO Box 12 649
Wellington, New Zealand
peter@statsresearch.co.nz

1. Introduction

Evolving volatility is a dominant feature observed in most financial time series and a key parameter used in option pricing and many other financial risk analyses. Although there is now an extensive literature on the estimation of parametric volatility models (see Engle (1982), Taylor (1986), Bollerslev, Chou and Kroner (1992), Harvey, Ruiz and Shephard (1994), Shephard (1996) and Barndorff–Nielsen and Shephard (2001) for example) less attention has been paid to simpler non–parametric alternatives. Exceptions include Aït–Sahalia (1996), Anderson and Grier (1992), and Andersen, Bollerslev, Diebold and Labys (2001) for example. More closely related to this paper is the work of Turner and Weigel (1992) who analyse the volatility of the daily returns of the S&P 500 and Dow Jones indices using the interquartile range as well as other measures of volatility. However such estimates needs to be rescaled to estimate the variance of the underlying data since this is the predominant measure of volatility in financial applications due to its use in standard option pricing and portfolio optimisation methodologies.

In this paper we present preliminary findings on the construction and properties of non–parametric estimators of time–varying volatility where volatility is assumed to be a measure of variance. Our focus is on financial data and, in particular, the daily returns of market prices such as equities, market indices, exchange rates etc, where daily returns are the first differences of the logarithms of the prices. Thus each daily return measures the continuously compounded rate of return on the asset, index etc over the day concerned. It is noted that non–parametric volatility estimators are particularly appropriate for extracting and understanding historical volatility prior to fitting any more sophisticated parametric model. They also provide robust benchmarks for testing the forecasting and in–sample performance of competing parametric procedures.

Our objective is to construct non–parametric volatility estimators that have simple structure, are cheap to compute, and are tailored to the typically heavy–tailed distributions met in practice. In particular we seek procedures that are robust to distributional assumptions, resistant to outliers, and have a sound statistical basis with reasonable precision properties. The estimation procedures that we consider construct local robust scale estimates (not necessarily estimates of standard deviation) based on finite moving–averages of the squared deviations of the time series from its local level. The moving–average weights are selected with reference to a target family of heavy–tailed distributions exemplified by the t–distribution, and the span of the moving–average is chosen so that the volatility is approximately constant within the local time window concerned. Finally, a global correction factor is applied to the local scale estimates to provide estimates of time–varying volatility.

Underpinning this approach are two key assumptions. The first assumes that volatility generally changes smoothly over time and, in particular, is locally constant over the local time windows within which estimation takes place. This seems a reasonable assumption for the most part; without it reliable volatility estimation would be difficult to achieve. It also accords with the long–memory dependence of squared daily returns and absolute returns observed in practice (see Ding, Granger and Engle (1993), Granger and Ding (1995), Rydén, Teräsvirta and Åsbrink (1998) for example). However this framework does not properly account for discontinuous breaks in volatility which may occur in practice (see Lamoureux and Lastrapes (1990), Hamilton and Susmel (1994), McConnell and Perez–Quiros (2000) for example) although our methodology could no doubt be adapted to better identify such changes.

The second key assumption is that the daily returns of financial time series have heavy–tailed distributions that are better approximated by a t–distribution with low degrees of freedom than a Gaussian distribution. There are many studies, Fama (1965) being the first, that support the general heavy–tailed hypothesis which would appear to be a ubiquitous feature of financial data. A number of candidate distributions have been proposed of which the t–distribution is a common choice. See, for example, Blattberg and Gonedes (1974), Harvey, Ruiz and Shephard (1994), Hurst and Platen (1997), Liesenfeld and Jung (2000) and Barndorff–Nielsen and Shephard (2001) among many others. Typically the degrees of freedom ν of the t–distribution found in such studies range between 3 and 9. Note, however, that the t_ν distribution has infinite moments of order k when $k \geq \nu$. Thus $\nu \geq 3$ ensures finite variance and $\nu \geq 5$ ensures finite kurtosis.

Subject to these assumptions, the global correction factor is computed from the sample variance of the original data standardised by the local scale estimates. Although the heavy tailed distributions of the returns make the moving sample variance an unreliable local volatility estimate, particularly if the moving time window is small, we shall assume that the sample variance is a reliable estimator of the variance of the underlying heavy–tailed distribution in large samples of the order of the length of the data. In essence we assume that the extreme values of the heavy–tailed distribution will appear representatively in large samples where they will not distort the sample variance. However, in small samples of the order of the moving time window, such extreme observations will be over–represented when they occur and can severely distort the sample variance.

With these points in mind we now choose to model a time series of (daily) returns R_t as

$$R_t = \sigma_t \epsilon_t \tag{1}$$

where $R_t = \ln S_t - \ln S_{t-1}$ and S_t is the underlying time series of prices concerned. The ϵ_t are assumed to be independently and identically distributed with mean zero and unit variance. The condition $E(\epsilon_t^2) = 1$ serves to identify the volatility σ_t^2 which is assumed to be a strictly positive, smoothly–varying function of time so that R_t has mean zero and variance σ_t^2. In fact R_t will almost always have a non–zero mean level that is typically very small in relation to $\sigma_t \epsilon_t$ since it represents the mean return over just one day. However we shall assume that R_t has been corrected for such an evolving mean level if appropriate. As mentioned before, the R_t will typically be heavy–tailed so that the common distribution of the ϵ_t will be heavy–tailed also.

In Section 2 we review methods for robust estimation of scale in the case where σ_t is a constant and in Section 3 we use this information to construct robust and resistant estimates of evolving volatility. Section 4 presents a preliminary assessment of these estimators using simulated data and also applies the estimators to stock price data.

2. Robust scale estimation

An excellent review of robust and resistant scale estimation methods is given in Iglewicz (1983) who builds on the Monte Carlo studies of Gross (1976) and Lax (1975). In a subsequent paper, Lax (1985) reported on a more extensive Monte Carlo study on the finite–sample performance of 17 robust scale estimators for heavy–tailed symmetric distributions. The 17 scale estimators considered had been identified as promising or commonly used from earlier studies. The general conclusion of the Lax (1985) study was that appropriately tuned A–estimators of scale were more robust for heavy–tailed symmetric distributions than the other estimators considered.

Given n independent and identically distributed observations $X_1, \ldots, X_n$, the A–estimator of scale with biweight weight function is given by

$$s_A^2 = \frac{k_A^2}{n-1}\sum_{t=1}^{n}(1-u_t^2)^4 e_t^2 \qquad (2)$$

where

$$k_A^{-1} = \frac{1}{n}\sum_{t=1}^{n}(1-u_t^2)(1-5u_t^2), \qquad u_t = \frac{e_t}{cs_0}, \qquad e_t = \begin{cases} X_t - M & (|X_t - M| \le cs_0) \\ 0 & (|X_t - M| > cs_0) \end{cases}$$

and M is an M–estimator of location. Here c is a positive tuning constant that depends on the choice of auxiliary scale estimate $s_0 > 0$, and s_0 is generally taken to be the median absolute deviation of the X_t from the sample median

$$MAD = \text{median}\{|X_t - \text{median}X_t|\}.$$

An A–estimator of scale is based on the sample analogue of the asymptotic variance of an M–estimator of location (see Iglewicz (1983) or Lax (1985) for further details). The estimator (2) with $c = 9$ and $s_0 = MAD$ performed best in the Lax (1985) Monte Carlo study. Note that for Gaussian data $E(MAD)/0.6745$ is an unbiased estimator of the standard deviation so that, in this case, (2) with $s_0 = MAD$ and $c = 9$ will ignore observations that lie more than 6 standard deviations from the location estimate M. Typically M is the sample median although other robust estimates of location can be used. In the financial time series context considered in Section 3 the level of the time series will typically be determined by a robust smoothing procedure such as *loess* (see Cleveland, Grosse and Shyu (1992)), but other smoothers based on M–estimates of location are also possible.

To better identify appropriate values for c in (2) we re–worked the Lax (1985) study for a selection of scale estimators of interest. Given the significant improvements in speed and accuracy of modern computing that have taken place since the mid 1980s, a larger scale, more accurate study was able to be undertaken. The results of this particular study will be reported elsewhere. However, as before, the A–estimators performed best, but with $c = 10$ and $c = 11$ emerging as better choices for c in (2) than the value $c = 9$ found by Lax (1985). Optimal estimators were taken as those whose logarithms had the largest triefficiency (see Beaton and Tukey (1974)). As well as being a symmetrising transformation, the use of logarithms allows direct comparison between the reliability of scale estimators whose expected values differ. Here triefficiency is measured as the minimum of the efficiencies obtained in the three cases where the data follows the Gaussian, one–wild and slash distributions. These three "corner" distributions have varying degrees of heavy tail behaviour and are meant to encompass most situations met in practice. In particular, the slash distribution has infinite variance and heavier tails than the target family of t_ν distributions ($\nu \ge 3$) that we have in mind here.

We now develop an alternative estimator to (2) which is more closely tailored to the t–distribution. Let $X_1, \ldots, X_n$ denote the n independent and identically distributed random variables given by

$$X_t = \frac{\sigma Z_t}{\sqrt{S_t}} \qquad (t = 1, \ldots, n)$$

where the Z_t are independent $N(0,1)$ random variables that are independent of the S_t, the S_t are independent random variables on $[0, \infty)$ each with distribution function $F_S(s)$, and $E(X_t^2) = \sigma^2$. We have in mind the case where S_t is proportional to a χ^2_ν random variable so that the X_t are scaled t_ν random variables with $\nu \geq 3$ to ensure finite variance. However, for the moment, we pursue the more general case of arbitrary distribution function $F_S(s)$ and consider estimating σ^2 using maximum likelihood. Note that, in almost all cases of practical interest, the maximum likelihood estimator of σ^2 will have to be obtained using iterative optimisation procedures and it is the moving–average equivalent of such procedures that we seek.

Bearing these points in mind we now use the EM algorithm (Dempster, Laird and Rubin (1977)) to construct an iterative formula for estimating σ^2. Ignoring additive terms that are not functions of the parameter σ^2, the log-likelihood of the complete data $\mathbf{X} = (X_1, \ldots, X_n)$, $\mathbf{S} = (S_1, \ldots, S_n)$ is proportional to

$$\log \tilde{L} = -\log \sigma^2 - \frac{1}{\sigma^2} \frac{1}{n} \sum_{t=1}^{n} S_t X_t^2.$$

Maximising $E_0(\log \tilde{L} | \mathbf{X})$ with respect to σ^2 yields

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{t=1}^{n} E_0(S_t | \mathbf{X}) X_t^2 \tag{3}$$

where the expectation is with respect to the conditional distribution evaluated at σ_0^2, a previous estimate of σ^2. It can be shown that

$$E_0(S_t | \mathbf{X}) = Q\left(\tfrac{1}{2}\left(\frac{X_t}{\sigma_0}\right)^2\right) \tag{4}$$

where $Q(t) = -d \log M(t)/dt$ and $M(t)$ is the Laplace transform

$$M(t) = \int_0^\infty e^{-ts} \sqrt{s} dF(s) \qquad (t \geq 0).$$

Note that (3) has the same structure as (2) since both involve averaging squared observations weighted using weights that are a function of an auxiliary or preliminary estimate of σ^2. In this sense both estimators can be regarded as M–estimators of scale.

Now consider the case where the S_t are $\chi^2_\nu/(\nu - 2)$ random variables so that the X_t are scaled t_ν random variables with $E(X_t^2) = \sigma^2$. Then (3) becomes

$$\hat{\sigma}^2_\nu = \frac{\nu + 1}{\nu - 2} \frac{1}{n} \sum_{t=1}^{n} \left(1 + \frac{X_t^2}{(\nu - 2)\sigma_0^2}\right)^{-1} X_t^2. \tag{5}$$

For given ν this estimator will need to be iterated to obtain the maximum likelihood estimator of σ^2. It is the moving-average equivalent of this estimator that we choose to base our volatility estimation procedure on.

Other possible choices could be considered for the distribution of S_t that involve censoring to minimize the impact of outliers, mixed distributions with point mass at zero to account for sticky prices, and mixture distributions among other candidate distributions chosen to exemplify the target family of distributions under study. In some circumstances it may also be possible to infer the underlying distribution of S_t from procedures where $Q(t)$ has been specified. These remain topics for further research.

3. Local volatility estimation

We now consider the time series of daily returns R_t $(t = 1, \ldots, T)$ defined by (1) with heavy–tailed ϵ_t and where the R_t have been corrected for evolving level if appropriate. Natural time series estimators of the evolving volatility σ_t^2 based on (2), (3) and (5) are the finite moving averages of span $n = 2r + 1$ given by

$$\hat{\sigma}_t^2 = \sum_{j=-r}^{r} w_j q_{t+j} R_{t+j}^2 \tag{6}$$

where the smoothness weights w_j satisfy $\sum_{j=-r}^{r} w_j = 1$ and the robustness weights

$$q_t = Q\left(\tfrac{1}{2}\left(\frac{R_t}{s_t}\right)^2\right)$$

depend on a prior estimate s_t^2 of σ_t^2. Here the function $Q(t)$ is as defined for (2), (3) or (5). In practice the estimates will involve iteration so that initial estimates of σ_t^2 are successively refined. If (6) is based on (2) then s_t will be the moving MAD and only two iterations through the data are required, one to determine s_t and the other to determine the final estimate of σ_t^2. In the case of (5) one could iterate until approximate convergence using the moving sample variance as the initial estimate of σ_t^2 (a limited simulation study shows that a total of 4 iterations is usually sufficient) or use one iteration with s_t^2 estimated by (2) yielding a total of 3 iterations.

However these estimators, together with those based on the moving standard deviation and moving MAD, will typically estimate σ_t^2/τ where τ is a positive constant and $\tau \neq 1$. To correct for this bias we multiply through (6) by $\hat{\tau}$ where

$$\hat{\tau} = \frac{1}{T}\sum_{t=1}^{T}\left(\frac{R_t}{\hat{\sigma}_t}\right)^2. \tag{7}$$

This estimator is just the sample variance of the scale adjusted returns and, although not a robust or reliable estimator in small samples, it can be expected to provide a reliable estimate of the variance for very large samples. As noted in Section 1, the extreme values of the heavy–tailed distribution are assumed to appear representatively in large samples of the order of the series length T, but will be over–represented when they occur in small samples of the order of the span n of the moving estimation window.

4. Simulations and data analysis

Using simulated and real data, we now consider the relative performances of the local volatility estimators based on the standard deviation, MAD, biweight A–estimator, and t–estimator respectively. We selected $\nu = 5$ for the t–estimator since the t_5 distribution has heaviest tails among the t–distributions with finite variance and kurtosis. In all cases the iterated t–estimator was initialised by the local sample variance and iterated a further 3 times to produce a final estimate.

For the simulation study we simulated 270 returns from the scaled t_ν distribution with unit variance and with $\nu = 3, 5, 9$ to represent varying degrees of heavy tailed behaviour. The series length was chosen to roughly represent a calendar year of trading days allowing for end effects. Our estimators were based on moving windows of span $n = 21$ with uniform weights $w_j = 1/21$. The latter were selected since our concern at this stage was with the precision of the estimators rather than their smoothness. The impact of the smoothness weights w_j on the properties of the estimators, among other issues, remains to be investigated as part of a more extensive simulation study. The 270 t_ν returns were then multiplied by the smooth volatility function σ_t where

$$\sigma_t^2 = 9e^{\sin(\pi t/125)}.$$

The four estimates of σ_t^2 were calculated and scaled so that the standardised returns had unit sample variance. They were assessed by their mean absolute error (MAE $= \frac{1}{N}\sum|\sigma_t^2 - \hat{\sigma}_t^2|$) and mean absolute proportionate error (MAPE $= \frac{1}{N}\sum|\sigma_t^2 - \hat{\sigma}_t^2|/\sigma_t^2$) where the latter were computed over the $N = 250$ volatility estimates available. These statistics were then averaged over 200 independent realisations of the time series, for each of the three distributions, and the four estimates to yield the results in Table 1.

The results are self evident. The iterated t–estimator performed best, even in the cases where the underlying distribution was not the t_5 distribution. As expected, the volatility estimator based on the moving standard deviation performed reasonably well for $\nu = 9$, but its performance deteriorated as ν decreased. The biweight A–estimator performed reasonably well in all cases. It might be expected that the A–estimator would have a comparable or better performance than the iterated t–estimator for heavy tail data not well–approximated by a t–distribution. However this has not yet been verified.

Finally we turn to examine the performance of these estimators on actual stock data. Coca-Cola Amatil Ltd (CCL) is among the largest and most actively traded companies listed

Table 1. The mean absolute error (MAE) and mean absolute proportionate error (MAPE) of four local volatility estimators estimating a smoothly varying volatility function over moving windows of span 21. The simulated returns have scaled t_ν–distributions with $\nu = 3, 5, 9$.

	t_3		t_5		t_9	
Estimator	MAE	MAPE	MAE	MAPE	MAE	MAPE
Standard deviation	6.01	0.801	4.45	0.441	3.28	0.338
Median absolute deviation	6.06	0.638	6.05	0.520	5.47	0.479
Biweight A-estimator with $c = 10$	4.86	0.568	4.04	0.402	3.38	0.344
Iterated t-estimator with $\nu = 5$	4.02	0.539	3.30	0.347	2.83	0.292

on the Australian Stock Exchange. Daily closing price data for this stock, for the 500 trading days preceeding 11 September 2000, was analysed. A time series plot of the daily return data features periods of low volatility and periods of high volatility, and a small number of extreme returns.

Calculation of evolving volatility for CCL using the moving standard deviation produced volatility estimates which were badly affected not only by the large returns, but also by the many small returns. The resulting fluctuations in the volatility estimates gave a distribution of standardised returns that was not well approximated by the Gaussian distribution since it had a sharp peak and values outside four standard deviations from the mean. Estimating scale using the A-estimator also resulted in standardised returns that were not well approximated by the normal distribution. However the distribution of these standardised returns had a smoother peak (the generally lower volatility estimates had not brought so many returns close to zero), and a distribution that was reasonably well–approximated by a t_ν distribution with $\nu = 5.55$. On this basis, parametric volatility estimation based on the t_5 distribution should improve the quality of the volatility estimates.

A plot against time of the three estimates of evolving volatility based on the standard deviation, A–estimator and iterated t–estimator showed that the t–estimator was the most stable. It typically adopted a compromise position between the standard deviation and the A–estimator, but closer to the A–estimator. The impact of extreme returns was clearly evident on the standard deviation and the A–estimator often appeared to discount such returns too heavily. The superior performance of the iterated t–estimator was to be expected given the results of the simulation study reported in Table 1 and the fact that the distribution of standardised returns was well–approximated by a t–distribution.

5. Conclusions

This paper considers non–parametric estimation of evolving volatility in the context of heavy–tailed distributions of returns. A new robust time series estimation procedure based on finite moving averages and the t–distribution has also been introduced. Our preliminary findings indicate that local volatility estimation based on the biweight A–estimator with tuning parameter $c = 10$ is a reliable estimator in all situations. However local volatility estimation based on the iterated t–estimator performed best in the many cases where the distribution of returns is well–approximated by a t–distribution.

REFERENCES

Aït–Sahalia, Y. (1996). Nonparametric pricing of interest rate derivative securities. *Econometrica*, **64**, 527–560.

Andersen, T. G., Bollerslev, T., Diebold, F.X. and Labys, P. (2001). The distribution of realized exchange rate volatility. *J. Am. Stat. Assoc.*, **96**, 42–55.

Anderson, M. and Grier, D.A. (1992). Robust, non-parametric measures of exchange rate variability. *Applied Economics*, **24**, 951–958.

Barndorff–Nielsen, O.E. and Shephard, N. (2001). Non-Gaussian- Ornstein–Uhlenbeck–based models and some of their uses in financial economics. *J. R. Statist. Soc. B* **63**, 167–241.

Beaton and Tukey (1974). The fitting of power series, meaning polynomials, illustrated on band–spectroscopic data. *Technometrics* **16**, 147–185.

Blattberg, R. and Gonedes, N. (1974) A comparison of the stable and Student distributions as statistical models for stock prices. *J. Bus.*, **47**, 244–280.

Bollerslev, T., Chou, R.Y. and Kroner, K.F. (1992). ARCH modeling in finance: A review of the theory and empirical evidence. *J. Econometrics*, **52**, 5–59.

Cleveland, W.S., Grosse, E. and Shyu, W.M. (1992). Local regression models. In *Statistical Models in S* (eds Chambers, J.M. and Hastie, T.J.), Wadsworth & Brooks/Cole, California.

Dempster, A.P., Laird, N.M., and Rubin, D.B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *J. R. Statist. Soc. B*, **39**, 1–38.

Ding, Z., Granger, C.W.J. and Engle, R.F. (1993). A long memory property of stock market returns and a new model. *J. Empirical Finance*, **1**, 83–106.

Engle, R. (1982). Autoregressive conditional heteroskedasticity with estimates of the variance of the United Kingdom inflation. *Econometrica*, **50**, 987–1007.

Fama, E. (1965). The behaviour of stock market prices. *J. Bus.*, **38**, 34–105.

Granger, C.W.J. and Ding, Z. (1995). Some properties of absolute returns, an alternative measure of risk. *Ann. Econ. Statist.*, **40**, 67–91.

Gross, A.M. (1976). Confidence interval robustness with long-tailed symmetric distributions. *J. Am. Stat. Assoc.*, **71**, 409-416.

Hamilton and Susmel (1994). Autoregressive conditional heteroskedasticity and changes in regime. *J. Econometrics*, **64**, 307–333.

Harvey, A., Ruiz, E. and Shephard, N. (1994). Multivariate stochastic variance models. *Rev. Econ. Stud.*, **61**, 247–264.

Hurst, S.R. and Platen, E. (1997). The marginal distribution of returns and volatility. In L_1*–Statistical Procedures and Related Topics* (ed. Dodge, Y.), IMS Lecture Notes–Monograph Series **31**, California.

Iglewicz, B. (1983). Robust scale estimates. In *Understanding Robust and Exploratory Data Analysis*, John Wiley, New York.

Lamoureux, C.G. and Lastrapes, W.D. (1990). Persistance in variance, structural change, and the GARCH model. *J. Bus. Econ. Statist.*, **8**, 225–234.

Lax, D.A. (1975). An interim report of a Monte Carlo study of robust estimators of width. Technical Report **93**, Series 2, Department of Statistics, Princeton University.

Lax, D.A. (1985). Robust estimators of scale: finite-sample performance in long-tailed symmetric distributions. *J. Am. Stat. Assoc.*, **80**, 736–741.

Liesenfeld, R. and Jung, R.C. (2000). Stochastic volatility models: conditional normality versus heavy-tailed distributions. *J. of Appl. Econometrics*, **15**, 137–160.

McConnell, M.M and Perez–Quiros, G. (2000). Output fluctuations in the United States: What has changed since the early 1980's? *Am. Econ. Rev.*, **90**, 1464–1476.

Rydén, T., Teräsvirta, T. and Åsbrink, S. (1998). Stylized facts of daily return series and the hidden Markov model. *J. of Appl. Econometrics*, **13**, 217–244.

Shephard (1996). Statistical aspects of ARCH and stochastic volatility. In *Time Series Models in Econometrics, Finance and Other Fields* (eds Cox, D.R., Hinkley, D.V. and Barndorff–Nielsen, O.E.), Chapman and Hall, London.

Taylor, S. (1986). *Modelling Financial Time Series.* John Wiley, New York.

Turner, A.L. and Weigel, E.J. (1992). Daily stock market vvolatility: 1928-1989. *Man. Sci.*, **38**, 1586-1609.

RESUME

Dr Martin Lally is Senior Lecturer in Finance at Victoria University of Wellington, New Zealand. His research interests are in valuation and cost of capital.

John Randal is a Teaching Fellow in the School of Economics and Finance at Victoria University of Wellington, New Zealand. His current research interests include non-parametric volatility estimation, robust scale estimation, and parametric stock price modelling.

Dr Peter Thomson is a director of Statistics Research Associates Ltd, Wellington, New Zealand. His research interests are in the general area of time series analysis including forecasting, financial time series, seasonal adjustment and signal processing. This research typically has an applied focus with application to economic, financial and geophysical time series.