Stochastic volatility: Modeling the latent process empirically

Erik Bølviken

Department of Mathematics, University of Oslo
& The Norwegian Computing Center
PO Box 1053, N-0316 Oslo, Norway
Email: erikb@math.uio.no

September 1, 2001

1 Introduction

Log-returns of share prices and other financial data are autocorrelated through their volatility. Numerous mathematical models have been invented to deal with this; many belonging to the very popular ARCH/GARHCH type originating with Engel (1982). An alternative approach is to link volatility to an underlying, unobserved state of the market, say to its degree of agitation, nervousness and optimism. As these attributes fluctuate, then so does the volatility. The precise formulation mathematically is in terms of a stochastic process; see e.g. Breidt, Crato and de Lima (1998), Pitt and Shepard (1999) or Aguillar and West (2000) for some fairly recent contributions. One of the challenges is that this state process is only *latent*. How it develops is only observed indirectly through its impact on the prices quoted.

There is an enormous scientific literature in engineering and statistics on the theory and applications of models with latent processes. A relevant reference is West and Harrison (1997). Attempts of using such models in finance have been through fully specified parametric models; see e.g., the references above. One model is used for the latent process and another one for its relationship to the data. The purpose of this article is to point out that it should be possible to deduce some of the features of the latent process from the sole assumption that it is *stationary*. It is then assumed that the process does not systematically alter its behavior with time, which seems a reasonable view to take on the stock market in a long-run perspective.

The statistical and numerical features of such an approach will be outlined in Section 2 and estimates of autocorrelation functions of financial index series given in Section 3. Note the similarity with the elementary statistical methods for directly observed time series in textbooks like Shumway (1988). It is then routinely recommended that autocorrelation functions should be identified prior to the fitting of parametric models. What is done here is to suggest a similar procedure for time series that are only observed indirectly. We shall only consider univariate series, but the same technique could be used for studying relationships between different series based on different latent processes that were strongly interrelated. Another extension is to introduce a feedback between the latent process and the prices quoted. This is briefly discussed in Section 4.

2 Statistical methods

2.1 The model

Let y_k be the log-return at time k, i.e. $y_k = \log(x_k/x_{k-1})$ where x_k is the actual price of the equity. The standard stochastic volatility model by means of a latent process $\{s_k\}$ is to specify

$$y_k = \sigma_k \varepsilon_k, \qquad \sigma_k = \sigma_0 \exp(\alpha s_k). \tag{1}$$

Here σ_k is the volatility at time k, which fluctuates around the fixed σ_0 according to the realizations of $\{s_k\}$, the parameter α defining the size of the oscillations. The trading at time k is represented by the random term ε_k, which has unit variance and also zero mean if we ignore any (small) drift in the long run. The sequence $\varepsilon_1, \varepsilon_2, \ldots$ should be regarded as an an independent one; the whole point behind the introduction of the latent process is that the memory in the log-returns is captured by it.

It will be assumed that $\{s_k\}$ is Gaussian. This condition lacks conviction, but it is at least a reasonable baseline case under which to carry out the present analyses; see also Section 4. Usually a parametric model is imposed on $\{s_k\}$. That is where we differ in that only *stationarity* is assumed. The autocorrelation function

$$\rho_l = \mathrm{cor}(s_k, s_{k+l}), \qquad l = 1, 2 \ldots \tag{2}$$

is our target. Note that it may without loss of generality be assumed that $\{s_k\}$ has zero mean and unit variance; if not, these parameters are absorbed into σ_0 and α.

A Gaussian model will be introduced for ε_k as well, but that is much less restrictive than might appear at first glance, since there is an important indeterminateness in the set-up. This is dealt with next.

2.2 An ambiguity

Any Gaussian stationary process with zero mean and unit variance may be written

$$s_k = b s'_k + (1 - b^2)^{1/2} \omega_k, \tag{3}$$

where $\{s'_k\}$ and $\{\omega_k\}$ are mutually independent processes, the latter consisting of independent terms ω_k. Both these processes are to have zero mean/unit variance, and b is a parameter. The expression (3) may be inserted into (1). This yields

$$y_k = \sigma'_0 \exp(\alpha' s'_k) \varepsilon'_k \tag{4}$$

where

$$\alpha' = \alpha b \tag{5}$$

$$\sigma'_0 = \sigma_0 \exp\{2\alpha^2 (1 - b^2)\} \tag{6}$$

$$\varepsilon'_k = \varepsilon_k \exp\{\alpha (1 - b^2)^{1/2} \omega_k - 2\alpha^2 (1 - b^2)\}, \tag{7}$$

From its appearance (4) is a model of the same type as (1), and the way σ_0 and ε'_k are defined ensures that the two models are actually identical, except for the distribution of the heavy-tailed distribution ε'_k being non-Gaussian. To see this, first note that $\{\varepsilon'_k\}$ is an independent process and is independent from $\{s'_k\}$ as well. Moreover, since $E(\theta\omega_k) = \exp(2\theta^2)$, it follows from (7) that $\mathrm{var}(\varepsilon'_k) = 1$ and clearly $E(\varepsilon'_k) = 0$. The kurtosis of ε'_k turns out to be $3\{\exp(4\alpha^2(1-b^2))-1\}$, which grows from zero for $b=1$ to some maximum for $b=0$.

The argument shows that by changing the latent process according to (3) we change the distribution of the process $\{\varepsilon'_k\}$. In all other ways the two models (1) and (4) are equal. We shall in Section 3 see that this opens for two equally valid interpretations of the results obtained for the financial series considered there.

2.3 Pseudo-likelihood

The estimation of the autocorrelation function (2) from an observed set of log-returns $y_1, \ldots, y_n$ can not be obtained from their full likelihood, since a full model for the latent process is not available, but a so-called pseudo-likelihood technique works. Let $f_0(y_k)$ be the probability density function of each observation y_k and $f_l(y_k, y_{k+l})$ the joint density of pairs at intervals l. Note that the stationarity assumption means that neither of these functions depend on k. Consider, in particular,

$$\lambda_0 = \sum_{k=1}^{n} \log\{f_0(y_k)\}, \tag{8}$$

which is the ordinary likelihood for the log-returns if the autocorrelation in the stochastic volatility is ignored. It is possible to obtain from λ_0 consistent estimates of the parameters σ_0 and α defining the marginal distribution f_0 as long as the latent process does not carry memory of infinite length.

Information on the autocorrelation ρ_l at lag l rests with the pairs (y_k, y_{k+l}) for $k = 1, 2, \ldots,$ and the bivariate likelihood

$$\lambda_l = \sum_{k=1}^{n-l} \log\{f_l(y_k, y_{k+l})\}, \tag{9}$$

which can be maximized with respect to ρ_l for an estimate. The technicalities are outlined in the appendix. When considering the feedback effect in Section 4 below each λ_l depends on more than one of the parameters, and it is necessary to consider

$$\Lambda_L = \sum_{l=1}^{L} \lambda_l \tag{10}$$

for some maximal lag L chosen. This criterion then has to be maximized jointly in all the autocorrelations, using modern numerical software; see e.g, for example Press, Teukolsky, Vetterling and Flannery (1992).

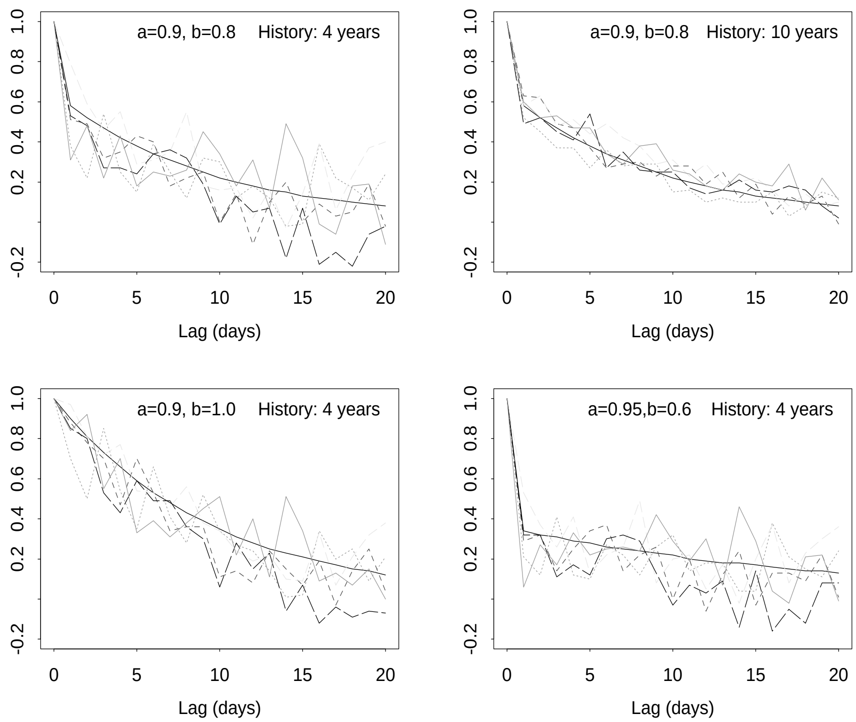

Figure 1: Estimates of the autocorrelations based on simulated data.

2.4 Verification

The purpose of this section is to examine through simulated data how long the time series must be to recover the essential structure of the autocorrelation function and indicate the random error in such estimates. Monte Carlo series of the form (1) where generated. The latent process $\{s_k\}$ was defined by (3) where

$$s'_k = as'_{k-1} + (1-a^2)^{1/2}\eta_k \tag{11}$$

is an autoregressive process of order one. The three series $\{\varepsilon_k\}$ in (1), $\{\omega_k\}$ in (3) and $\{\eta_k\}$ in (11) were all Gaussian with zero mean/unit variance and mutually independent. The true auto-correlation function of $\{s_k\}$ is then

$$\rho_l = b^2 a^l, \qquad l = 1, 2, \ldots \tag{12}$$

An alternative interpretation of the model for $\{s_k\}$ defined by (3) and (11) is as a first order ARMA in both the AR and MA part; see Schumway (1988). Further clarification will be given in Section 3.

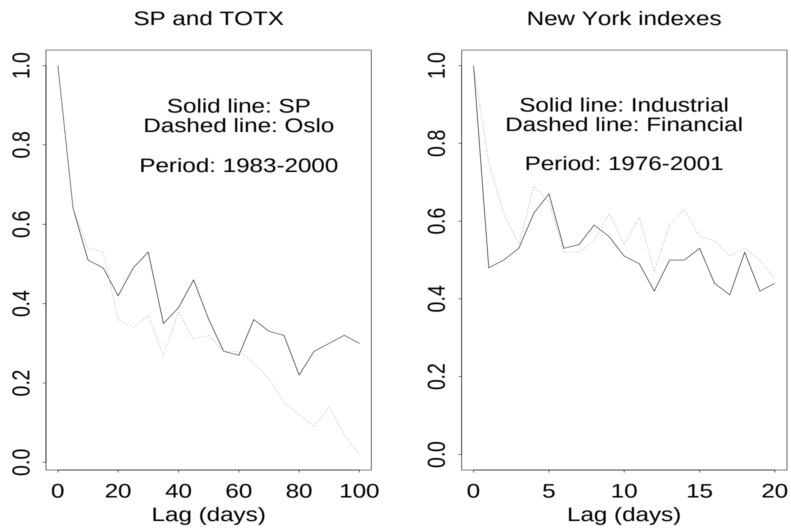

Figure 2: Estimates of the autocorrelations based on real index data.

The parameters chosen for the experiment are characteristic for daily index data. We took $\sigma_0 = 0.01$ and $\alpha = 0.5$ and varied a and b as shown in the panel in Figure 1. The true autocorrelation function is the smooth, solid curve in the middle of the plots. Estimates based on the five replicated Monte Carlo data oscillates around it. Note that random error is quite small in the upper right corner of the panel based on 5000 daily observations, corresponding to a data history of ten years. But even four years in the other examples seem sufficient to gauge the main structure of the underlying phenomenon.

3 Results

The technique has been applied to the industrial and financial index of the stock exchange of New York for the period from January 1th 1976 to July 1th 2001 and to the Standard & Poor index (denoted SP) and the total index (TOTX) of the Oslo (Norway) stock exchange. The latter two were from January 1th 1983 to April 1th 2000. All indexes were daily, and they were detrended by subtracting prior to the analysis the overall mean of their log-returns during the period.

Estimates of the autocorrelations based on all the data are shown in Figure 2. For the SP and TOTX financial indexes only each fifth day have been plotted. The long memory in the latent process is evident everywhere, but it is shorter in the much smaller financial community in Norway where it approaches zero after $80 - 100$ days. The shape of all the autocorrelation functions is consistent with the model in Section 2.4; see (12). The decay factor a is rather close to one, say at least 0.98 the American indexes, but smaller for the

		Period				
		83-87	87-90	91-97	97-00	83-00
SP	σ_0	.007	.008	.005	.010	.008
	α	.37	.54	.40	.41	.50
TOTX	σ_0	.008	.010	.009	.009	.009
	α	.45	.55	.42	.55	.50

Table 1: Estimates of the parameters σ_0 and α for the SP and TOTX indexes.

Norwegian one. The parameter b could be around 0.6.

There are two alternative positions to be taken on the interpretation of these results. They were derived from an assumption that the daily variation in log-returns, i.e. the process $\{\varepsilon_k\}$ in (1), are Gaussian. The latent process then follows the model in Section 2.4 with a value of b considerably smaller than one. This is *not* a Markov model, but it contains a core $\{s_k'\}$ which *is* of the Markov type. According to this view the log-returns are influenced by a latent process which hides daily fluctuations that are not absorbed into the basic attitude of the market as it is perceived over time. Alternatively, these daily oscillations in investor frame of mind is indistinguishable from other types of daily randomness and may be absorbed into them. This yields the model (4) with a non-Gaussian, heavy-tailed distribution for $\{\varepsilon_k\}$ and a Markov model for the latent process. Now the attitude of today completely determines how the degree of agitation, anxiety and so on is likely to develop tomorrow

We have investigated the stability of the estimates by dividing the data into four equal parts and estimating each quarter of the data separately. The series were split within the middle of the year, and no thought was given to the plausibility of the partitions from an economic point of view. The estimates of the parameters σ_0 and α defining the marginal distribution of log-returns are shown in Tables 1 and 2. Their variation between periods follows each other closely. The autocorrelation function estimated in each quarter are shown for the SP and TOTX indexes in Figure 3. The discrepancy between periods and the random variation is now considerable, but not necessarily inconsistent with the random variation obtained for the simulated data; see Figure 1 lower right, in particular, which is the one where the underlying model is closest to the real data. More work is needed to judge random error in Figure 3 properly and evaluate the stability of estimates between periods.

4 Extensions

The method presented can be extended in several directions. It appears interesting to examine cross-correlations between latent processes underlying *different* series. The aim would then be to identify simplifying structures, and it might confirm studies in Ball and Thorus (2000) and Longin and Solnik (2001) which argues that correlations between log-returns

		Period				
		76-82	83-89	90-96	96-01	76-01
Industrial	σ_0	.008	.008	.006	.008	.007
	α	.29	.53	.46	.44	.46
Financial	σ_0	.008	.008	.006	.011	.008
	α	.37	.56	.45	.41	.51

Table 2: Estimates of the parameters σ_0 and α for the New York industrial and financial indexes.

vary with time. Such effects could also be the consequence of using skew distributions for some of the variables defining the model.

Another possibility is the introduction of feedback effects. Most applications of latent process modeling assume the two processes $\{s_k\}$ and $\{\varepsilon_k\}$ in (1) to be stochastically independent, but surely this is not so obvious in the present situation. If s_k is to represent the attitude of the operators in the market, conceivably they might revise their views by the results of the current trading. The simplest way to represent this idea within the non-parametric set-up used here is to introduce the additional parameter

$$\tau = \text{cor}(s_k, \varepsilon_{k-1} | s_{k-1}), \tag{13}$$

which expresses that the *change* in the market view on risk from one day to another is influenced by the trading that has taken place. This is the only change in the relationship between the two processes $\{\varepsilon_k\}$ and $\{s_k\}$; i.e. each ε_{k-1} is still independent of the history $s_{k-1}, s_{k-2}, \ldots$ and *conditionally* independent of the future $s_{k+1}, s_{k+2}, \ldots$ given s_k.

A preliminary study on the feasibility of this will now be presented. The detailed methodological development is around the lines outlined, albeit more complicated, and are skipped, but the results of a small simulation study seems worthwhile to give. The model in Section 2.4 was used exactly as there, except for the error term η_k on the right in (11) now having to be made *dependent* of ε_{k-1} in (1). It can be proved that their correlation must be

$$\text{cor}(\eta_k, \varepsilon_{k-1}) = \tau b^{-1} \{(1 - a^2 b^4)/(1 - a^2)\}^{1/2} \tag{14}$$

to ensure that (13) is satisfied. The experiments were replicated 100 times with daily series of 1000 observations (i.e. 4 years) with parameter values $\sigma_0 = 0.01$, $\alpha = 0.5$, $a = 0.9$ and $b = 0.8$ and using $\tau = 0$ and $\tau = 0.3$. The estimates of τ turned out to be unbiased to two decimals in both cases, with a standard deviation slightly below 0.05. Thus it seems possible to detect feedback effects from the machinery presented. A preliminary try on the ST and TOTX financial indexes gave estimates of τ, roughly to the order of 0.1. Whether this is large enough to be of any importance has not been investigated.

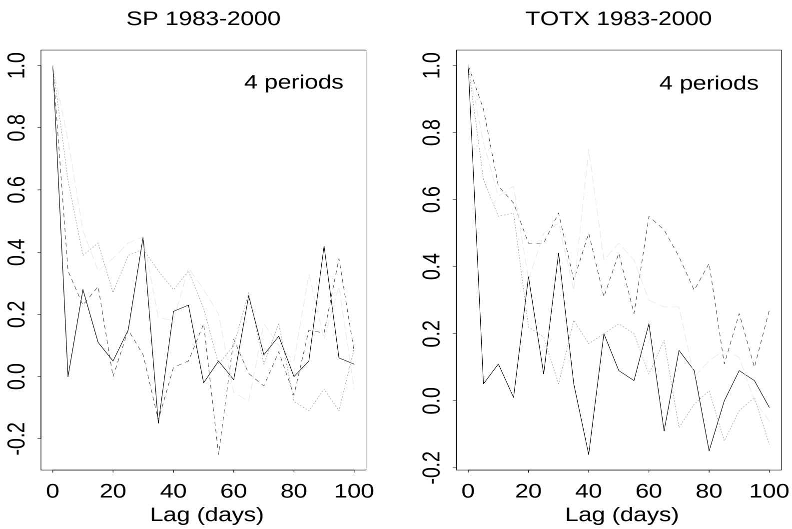

Figure 3: Estimates of the autocorrelations in 4 different periods.

5 Conclusions

It has been suggested that empirical modeling of latent processes in models with dynamic volatility and correlations can be carried out by means of a simple pseudo-likelihood technique. The autocorrelation functions of four financial index series were examined. It was shown that a first order autoregressive process, i.e. a Markov model, provides a good description if the day-to-day variation in log-returns is heavy-tailed. An alternative interpretation based on Gaussian daily variation was also given. Now the latent process followed an ARMA model which meant that the attitude of the market contained some random component, settled on a daily basis and not transferable to the next day and beyond. It was impossible to choose between these two explanations on empirical grounds.

6 References

Aiguillar, O. and West, M. (2000). Bayesian dynamic factor analysis and portfolio allocation. J Business & Economic Statistics, 18, 338-357.

Ball, C.A. and Torus, W.N. (2000). Stochastic correlation across international stock markets. J Empirical Finance, 7, 373-388.

Breidt, F.J., Crato, N. and de Lima, P. (1998). The detection and estimation of long memory in stochastic volatility. J Econometrics, 83, 325-348.

Longin, F. and Solnik, B. (2001). Extreme correlation of international equity markets. J Finance (forthcoming).

Pitt, M.K. and Shepard, N. (1999). Time-varying covariances: a factor analytic stochastic volatility approach. In Bayesian Statistics 6, 547-570. Oxford: Oxford University Press.

Press W.H., Teukolsky, S.A., Vetterling, W.T. and Flannery, P. (1992). Numerical recipes in C. Cambridge: Cambridge University Press.

Shumway, R.H. (1988). Applied statistical time series analysis. Englewood Cliffs: Prentice Hall.

West, M. and Harrison J. (1997). Bayesian forecasting and dynamic models. New York: Springer.

A Details to section 2.3

The marginal and pairwise probability density functions f_0 and f_l in (8) and (9) are under the assumptions in Section 2.1 computed as follows. Let $f_0(y|s)$ be the conditional density of $y = y_k$ given $s = s_k$. Since $\sigma(s) = \sigma_0 \exp(\alpha s)$ is the conditional volatility under the model (1), it follows that $f_0(y|s) = \phi_{\sigma(s)}(y)$, writing $\phi_\sigma(y) = (2\pi\sigma)^{-1/2} \exp(-y^2/2\sigma^2)$ to denote the centered Gaussian density. But then

$$f_0(y) = \int_{-\infty}^{\infty} \phi_{\sigma(s)}(y)\phi_1(s)ds,$$

which is easily computed by numerical integration.

For the pairwise density function $f_l(y, y_l)$ suppose that $l > 0$ and note that y and y_l are conditionally independent given s and s_l. Hence

$$f_l(y, y_l|s, s_l) = f_0(y_l|s_l)f_0(y|s),$$

and

$$f_l(y, y_l) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \phi_{\sigma(s_l)}(y_l)\phi_{\sigma(s)}(y)\phi(s, s_l; \rho_l)dsds_l$$

where $\phi(s, s_l; \rho_l)$ is the bivariate Gaussian density with zero means and unit variances and correlation ρ_l. Again numerical integration is required.