Stochastic Volatility and GARCH: A comparison based on UK stock data

Chiara Pederzoli[*]

University of Modena and Reggio Emilia

pederzoli.chiara@unimore.it

Abstract

This paper compares two volatility models for returns, i.e. a log-normal AR(1) stochastic volatility (SV) model and a GARCH (1,1) model, both from a theoretical and empirical point of view. The two models are estimated on UK stock data: a series of the British equity index FTSE100 is used to estimate the relevant parameters. Diagnostic tests are implemented in both cases to evaluate how well the models fit the data. Both models are used to get daily volatility forecasts and these volatilities are used to estimate the Value at Risk on a simple one-unit position on FTSE100. The VaR accuracy is tested by means of backtest and appropriate Likelihood Ratio tests: the results do not lead to a straightforward preference for any of the two models.

Keywords: volatility models, stochastic volatility, GARCH, Value at Risk.

[*] This paper is a revised version of my MSc Economics and Finance dissertation at the University of Warwick and was presented at the XXXII Euro Working Group on Financial Modelling conference in London (April 2003). I am grateful to my supervisor Ilias Tsiakas and to Mike Pitt for their helpful advices and to the conference participants for their comments and suggestions. The remaining errors and deficiencies are, needless to say, my own.

Introduction

In many financial applications, the specification of a model to represent returns' behaviour is of crucial importance. According to the traditional market efficiency hypothesis, the returns are defined as zero-mean serially uncorrelated and hence unpredictable random variables, but the empirical evidence suggests that returns, even if linearly independent, show a significant higher order dependency: more precisely, the squared returns are autocorrelated and 'clustering' in returns is very common: this means that volatility changes over time depending on its past values and hence it is predictable.

The issue of modelling returns accounting for time-varying volatility has been widely analysed in financial econometrics literature. Since the introduction by Engle (1982) of the ARCH (Autoregressive Conditional Heteroscedasticity) model, a wide range of extensions and modifications to the original model have been developed. Stochastic volatility models (SV), the most popular of which is due to Taylor (1986), are more sophisticated then ARCH-type models, and from a theoretical point of view they are more appropriate to represent the behaviour of the returns in real financial markets, the main drawback being a more statistically and computationally demanding implementation. As concerns the out-of-sample predictive performances the evidences are not clearly conclusive: for example, Gonzalez-Rivera, Lee, Mishra (2002) find a preference for SV in Value at Risk (VaR) computation. On the other hand Bluhm and Yu (2001) find that SV is not preferred to GARCH in a VaR framework, while SV is preferred in option pricing. Lehar, Scheicher and Schittenkopf (2002) show a preference for GARCH in option pricing, while no notable differences are found in VaR.

The aim of this paper is to contribute to this issue by drawing a VaR-based comparison of SV and GARCH on UK stock data (FTSE100 stock index). To this end, two simple models are estimated and the related volatility forecasts are used to estimate VaR on an artificial portfolio containing a unit position on FTSE100: the performances of the two alternative models are evaluated on the relative backtest.

The paper is organised as follows: the first section presents a brief description of volatility models, focusing on the stochastic volatility models definition and estimation. In the second section the data are presented and two models, a log-normal first order autoregressive (AR(1)) SV and a GARCH(1,1), are estimated. The last section presents the implementation

of VaR estimation and related backtest using alternatively the volatilities coming from the two models estimated. The last section concludes.

1. Volatility Models

The easiest assumption to model daily returns is a zero-mean normal random variable. While the zero mean is a credible assumption that is generally confirmed by financial data, many empirical findings show that the stock returns have a negatively skewed and leptokurtic distribution. The leptokurtosis can be handled by incorporating conditional heteroscedasticity into a Gaussian process. A model commonly adopted for returns is defined by

$$r_t = s_t e_t \quad e_t \sim IID(0;1) \tag{1.1}$$

In (1.1) e_t is a zero-mean white noise often assumed to be normal and s_t is the time-varying volatility.[1]

Assuming that e_t is a normal white noise, the returns conditional on s_t are normal:

$$r_t = s_t e_t \quad e_t \sim NID(0;1) \quad \Rightarrow \quad r_t \mid s_t \sim N\left(0; s_t^2\right) \tag{1.2}$$

While the normality is often assumed for the conditional distribution, by modelling s_t as being time varying the unconditional distribution is leptokurtic. Different specifications for s_t define different volatility models. Following Shephard (1996), two main classes of volatility models can be identified: observation-driven and parameter-driven volatility models.

Observation-driven models define s_t as a deterministic function of past observations of the returns: these are mainly the ARCH-type models, which in the most general formulation define the conditional variance as $s_t^2 = f(r_{t-1}^2,...,r_{t-p}^2, s_{t-1}^2,...,s_{t-q}^2)$. One of the most appealing features of the observation-driven models is that the one-step-ahead forecast density is defined explicitly; assuming normality:

$$r_t \mid R_{t-1} \sim N\left(0; s_t^2\right) \qquad R_{t-1} = (r_1,.....,r_{t-1}) \tag{1.3}$$

[1] This model handles the dynamics in the variance, while ignores the possible dynamics in the mean: even if the autoregressive structure of the return can be significant at least in the first order term, here the mean dynamics is neglected since the purpose of this work is to model the variance.

The ARCH models were introduced by Engle (1982). Many ARCH-type models have been developed, among which one of the most popular is the GARCH(1,1), originally proposed by Bollerslev (1986)

$$\begin{aligned} r_t &= s_t e_t \quad e_t \sim NID(0;1) \\ s_t^2 &= g + a r_{t-1}^2 + b s_{t-1}^2 \end{aligned} \tag{1.4}$$

Since e_t is assumed to be normal, the returns are normally distributed conditionally on the variance which depends on the information up to the previous period. The value of the parameters $a + b < 1$ defines the stationarity condition. While returns are modelled as being serially uncorrelated, the squared returns can be expressed as an autoregressive process (specifically an ARMA(1,1)). The unconditional distribution is symmetric around a zero mean, with variance $g/(1-a-b)$ and kurtosis greater then 3 (i.e.leptokurtic distribution). Similar features apply to the generic ARCH(p) and GARCH(p,q) models.

The normality assumption allows estimating an ARCH-type model very easily by maximum likelihood. The estimation becomes more complex if the normality assumption is relaxed: most commonly a Student-t distribution for e_t is assumed, which increases the ability of the model to capture the fat tails of the actual distribution of the returns (Baillie, Bollerslev (1989)). In this case an additional parameter that represents the degree of freedom of the t distribution has to be estimated.

The forecast of the variance is based on the conditional variance equation: the fact that the one step ahead forecast is fully determined is one of the main simplifications characterising GARCH compared to SV models.

Parameter-driven models, represented by SV models, define the volatility as an unobserved random variable driven by a latent stochastic process, that means a contemporaneous innovation appears in the equation defining volatility. The stochastic component in the log-variance equation accounts for the random new information that characterises financial markets. The statement in (1.2) is still valid, but the forecast density $r_t \mid R_{t-1}$ is not defined explicitly.

The simplest SV model is the log-normal AR(1), due to Taylor (1986), which is formulated as:

$$\begin{aligned} r_t &= s_t e_t \quad s_t = \exp(h_t/2) \quad e_t \sim N(0;1) \\ h_t &= g + f h_{t-1} + s_h h_t \quad h_t \sim N(0;1) \end{aligned} \tag{1.5}$$

The logarithmic formulation ensures the positiveness of the variance. In the simplest case e_t and h_t are assumed to be independent: a correlation between these two terms would heavily complicate the estimation of the model, but would allow the conditional variance to respond asymmetrically to rises and falls in the returns. The parameter f represents the persistence of the log-variance: if $|f|<1$ the log-variance is stationary. Assuming this stationarity condition, the log-variance follows an autoregressive process AR(1) with unconditional mean $E(h_t)=g/1-f$ and unconditional variance $Var(h_t)=s_h^2/1-f^2$.

As in the case of GARCH the returns are conditionally normal:

$$r_t \mid h_t \sim N(0;\exp(h_t)) \tag{1.6}$$

The main difference compared to GARCH is that in the SV model $f(r_t \mid s_t=\exp(h_t/2)) \neq f(r_t \mid R_{t-1})$ which is not defined explicitly and hence makes the likelihood intractable, with the consequence of estimation difficulties. Formally, the distribution of the returns conditional on the series of returns up to the previous period is defined as:

$$f(r_t \mid R_{t-1}) = \int f(r_t \mid h_t) f(h_t \mid h_{t-1}) f(h_{t-1} \mid R_{t-1}) dh_t dh_{t-1} \tag{1.7}$$

where $f(r_t \mid h_t)$ is a normal as in (1.6), and $f(h_t \mid h_{t-1})$ is a normal with mean $g+fh_{t-1}$ and variance s_h^2.

The unconditional distribution of return is a non-standard one[2] as in the case of GARCH. As in GARCH all the odd moments are zero, so that the unconditional distribution is symmetric and centred on zero. Moreover the kurtosis results to be higher then 3. While the returns are a white noise by definition, if the stationarity condition $|f|<1$ is satisfied the squared returns dynamic is driven by an autocorrelation function very close to the one of an ARMA(1,1), as pointed out in Shephard (1996).

Hence both the SV model and the GARCH model are able to explain some common features of the daily returns, which can be summarized as follows:

---

[2] In Taylor (1994) the unconditional distribution of the returns is defined as a lognormal mixture of normal distributions: it can in fact be argued that the returns are normally distributed conditionally on each realization of h_t, and these possible realizations are defined by the stochastic process which drives the logvariance. A mixture of normal distributions has fat tails by definition, so that the unconditional distribution of the return has excess kurtosis whenever the stochastic variance has positive variance and is independent of e_t (see Taylor (1994)). This means that the leptokurtosis of the unconditional distribution, while it derives just from the persistency coefficient $a+b$ in the case of GARCH, it comes both from f and from s_h^2 in the case of SV.

1) High kurtosis (fat tails);

2) Small autocorrelation in level;

3) Positive and statistically significant autocorrelation in the squared returns, slowly decreasing as the time-lag increases.

By contrast to GARCH, SV accounts for contemporaneous shocks affecting volatility, which can be interpreted as the random new information in the stock market.

The main disadvantage of SV compared to GARCH is the difficulty of estimation and of forecasting, due to the untractability of the predictive density $f(r_t \mid R_{t-1})$.

SV Estimation

While the ARCH-type models can be estimated by maximum likelihood due to the explicit definition of the predictive density, the SV model estimation is much more statistically demanding. Many different estimation methods have been developed and analysed: some of them only deliver the estimates of the model's parameters, while other methods also estimate the log-variance h_t. The log-variance can be computed using the full sample $R_T = (r_1,...,r_t,...,r_T)$, in which case the estimate $\exp(h_t/2) \mid R_T$ is called 'smoothed volatility', or it can be based on the observations up to the considered period producing the 'filtered volatility' $\exp(h_t/2) \mid R_t$. Among the most popular estimation methods are the Generalized Method of Moments (GMM), the Quasi Maximum Likelihood Estimation and the Markov Chain Monte Carlo (MCMC) which is implemented in the present work following Chib, Kim, Shephard (1998).

MCMC in the context of SV estimation is used within a Bayesian framework and it consists in drawing correlated samples (Markov Chain) form the required distributions . MCMC provides both the estimates of the parameters and of the smoothed volatility from a unique algorithm, by simulating the conditional densities $f(h_t \mid R_T)$ t=1,...,T and $f(q \mid R_T)$ and taking the averages (posterior mean). As pointed out in Jaquier, Polson, Rossi (1994), a Markov Chain sampler can be built by splitting the joint posterior density $f(h,q \mid R_T)$ in the two marginal densities $f(q \mid R_T, h)$ and $f(h \mid R_T, q)$ and then alternating them in the simulation. The present work implements an MCMC procedure with Gibbs sampler, as presented in Chib, Kim, Shephard (1998).

Volatility Filtering and Forecasting

While in the case of ARCH-type models the forecasting is straightforward since the one-step-ahead variance is fully deterministic, this task is much more statistically demanding in the case of SV models, since the log-variance equation is stochastic.

The process of computing the log-variance conditional on the observations up to the current time $h_t \mid R_t$, for t=1,...,T, is called 'filtering'[3]. Hence the outcome of the filtering is the density function $f(h_t \mid R_t)$. The filtering procedure assumes known values for the parameters q of the model: in the following the parameters will not be mentioned, implicitly assuming that all the densities are meant to be conditional also on the parameters values.

Following Pitt and Shephard (1999), the filtering density $f(h_t \mid R_t)$ is obtained by repeating over time a two-stage procedure:

1. Define the prediction density through the 'transition density' $f(h_{t+1} \mid h_t)$:

$$f(h_{t+1} \mid R_t) = \int f(h_{t+1} \mid h_t) f(h_t \mid R_t) dh_t \tag{1.8}$$

2. Use Bayes theorem to compute the filtering density:

$$f(h_{t+1} \mid R_{t+1}) = \frac{f(r_{t+1} \mid h_{t+1}) \times f(h_{t+1} \mid R_t)}{f(r_{t+1} \mid R_t)} \propto f(r_{t+1} \mid h_{t+1}) \times f(h_{t+1} \mid R_t) \tag{1.9}$$

where

$$f(r_{t+1} \mid R_t) = \int f(r_{t+1} \mid h_{t+1}) f(h_{t+1} \mid R_t) dh_{t+1} \tag{1.10}$$

In the context of SV models, the so called 'particle filters' are commonly used. Particle filters are a particular class of simulation filters which approximate the filtering variable, the log-variance $h_t \mid R_t$ in the SV model, by 'particles', that is by a finite number M of values $h_t^1, ..., h_t^M$, associated with discrete probabilities $p_t^1, ..., p_t^M$. The filtering density is approximated with the discrete set of values $h_t^j \quad j = 1, ..., M$ which are considered like samples from that density. This allows computing the prediction density in (1.8) with the approximation

$$\hat{f}(h_{t+1} \mid R_t) = \sum_{j=1}^{M} p_t^j f(h_{t+1} \mid h_t^j) \tag{1.11}$$

[3] In general, given a time series $y_t \quad t = 1, ..., n$ modelled as independent conditionally on an unobserved state a_t, 'filtering' means "to learn about the state given contemporaneously available information" (quote from Pitt, Shephard 1999).

Clearly the higher is the number of simulated values M the more accurate is the filtering sampler. The density in (1.11) is an 'empirical prediction density', that can be used to formulate an 'empirical filtering density':

$$\hat{f}(h_{t+1} \mid R_{t+1}) \propto f(r_{t+1} \mid h_{t+1}) \times \sum_{j=1}^{M} p_t^{\,j} f(h_{t+1} \mid h_t^j) \tag{1.12}$$

In the literature the probabilities $p_t^{\,1},.....,p_t^{\,M}$ are usually assumed to be equal, i.e. $p_t^{\,j} = 1/M \quad \forall j = 1,...,M$. Usually some accept/reject algorithm is performed to sample h_{t+1}.

Once the structure to filter volatility is available, the one-step-ahead forecast is based on the prediction densities $f(h_{t+1} \mid h_t^j) \quad j = 1,...,M$.

2. GARCH and SV Estimation on UK stock data

The Data

The data consist of FTSE100 daily prices covering the period 01/01/1990 – 31/12/2001.[4] The first eleven years have been considered as estimation sample, while the last year of data has been used as out of sample period for volatility forecasting and VaR estimation. The prices have been downloaded from DataStream.

The volatility models deal with returns, the series of which is presented in Fig. 2.1.[5]

[4] It has to be noted that the 11th September 2001 is included in the observations: the decision of not eliminating it from the sample is due to the consideration that there are other 'outliers' in the sample, similar in absolute magnitude, which have not been eliminated.

[5] The DataStream series excludes the week-ends but contains holidays like Christmas, where the markets are closed and so there are no new prices. On these holidays the prices are set equal to the previous day, so that the corresponding returns turn out to be zero. Following Shephard (1996) the zeros have not been taken out from the series.

Fig. 2.1 FTSE 100 percentage log-returns series: 01/01/1990-31/12/2001

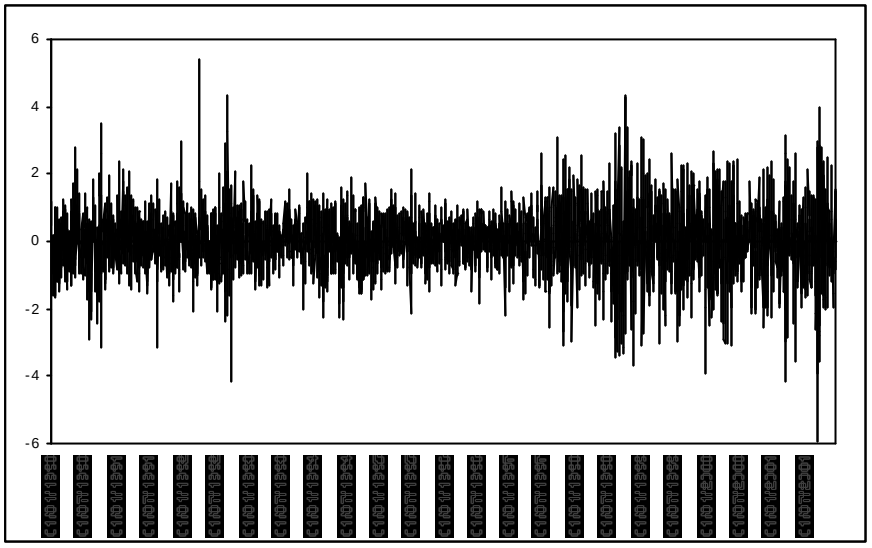


Figure 2.1 suggests that the returns are moving around an approximatively zero mean with time-varying volatility and a clustering phenomenon is quite evident. Table 2.1 presents some basic statistics describing the data. The sample mean, very close to zero, supports the assumption made in the return model. The ADF test gives a strong rejection of the null hypothesis of unit root, meaning that the series is stationary. The negative skewness and the high kurtosis, and the consequent rejection of the normality hypothesis, confirm the common empirical finding that daily returns are very far from being Gaussian.

Table 2.1 Basic Statistics for Percentage Returns

Mean	0.024508
Std. Deviation	0.96967
Skewness	-0.075289
Excess Kurtosis	2.2206
Minimum	-5.8853
Maximum	5.4396
Normality test[6]	Chi^2(2) = 380.08 (0.0000)
Augmented Dickey-Fuller (5 lags)	ADF = -24.55

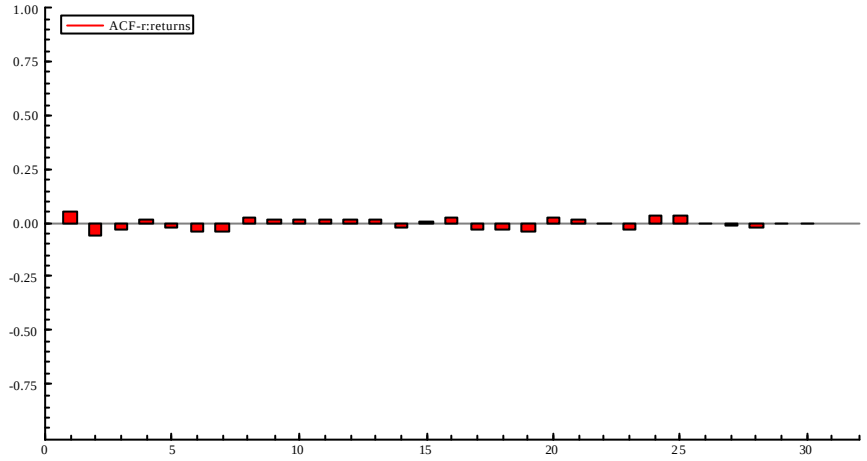
According to the market efficiency hypothesis, the returns are expected to be serially uncorrelated: the series considered here presents significant autocorrelation as table 2.2.A shows, but the values of the F statistic are not very high and the coefficients are quite low as can be seen in figure 2.2.

Table 2.2 Test of serial correlation at different lags

(2.A) H0: no autocorrelation in returns		
Lag	F value	p-value
1	8.5089	0.0036
up to 2	10.209	0.0000
up to 3	7.2972	0.0001
up to 30	2.0592	0.0006

(2.B) H0: no autocorrelation in returns		
Lag	F value	p-value
1	8.5089	0.0036
2	10.791	0.0010
3	2.4511	0.1175
from 3 to 30	1.5637	0.0300

Fig. 2.2 Returns Correlogram



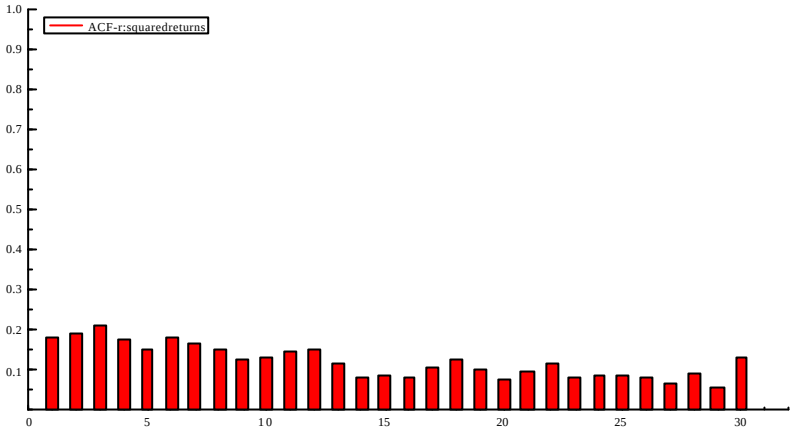
It can be argued that the autocorrelation, even if statistically significant, is not economically significant, at least beyond the first lag: this dynamic is ignored in this work. The autocorrelation is much more important in the squared returns: commonly squared returns have positive skewness, very high kurtosis and a strong positive autocorrelation, and these 'rules' are confirmed on this data set. Skewness and Excess Kurtosis are 6.1151 and 63.288 respectively and the hypothesis of no autocorrelation is strongly rejected at any lag tested.

Table 2.3 Test of autocorrelation

H0: no autocorrelation in squared returns		
Lag	F value	p-value
1	103.68	0.0000
up to 2	95.233	0.0000
up to 3	92.763	0.0000
up to 30	15.315	0.0000

[6] PcGive normality test: modification of Jarque-Bera presented in Doornik and Hansen (1994).

Fig. 2.3 Squared Returns Correlogram



The features of the FTSE 100 returns series as highlighted above suggest to use a model in which the volatility can change over time: this aim can be reached both adopting an ARCH-type model or a stochastic volatility model. Two simple models, a GARCH(1,1) and a log-normal AR(1) SV, will be estimated and discussed in the next two sections.

GARCH Estimation

A simple GARCH(1,1) is implemented, as it is one of the most popular in applications. Hansen and Lunde (2001) find that, despite its simplicity, GARCH(1,1) is not outperformed from other more complex ARCH/GARCH-type models. The first eleven years of the sample are used to estimate the model defined as follows:

$$(2.1.A) \quad r_t = \sigma_t \varepsilon_t \quad \varepsilon_t \sim NID(0;1)$$
$$(2.1.B) \quad \sigma_t^2 = \gamma + \alpha r_{t-1}^2 + \beta \sigma_{t-1}^2 \qquad (2.1).$$

Table 2.4 GARCH(1,1) Estimation Results

parameter	Estimate	Std. Error	p-value
gamma	7.95E-07	3.14E-07	0.117
alpha	0.0491	0.00906	0.002
beta	0.9420	0.01145	0.000
alpha+beta	0.9911		

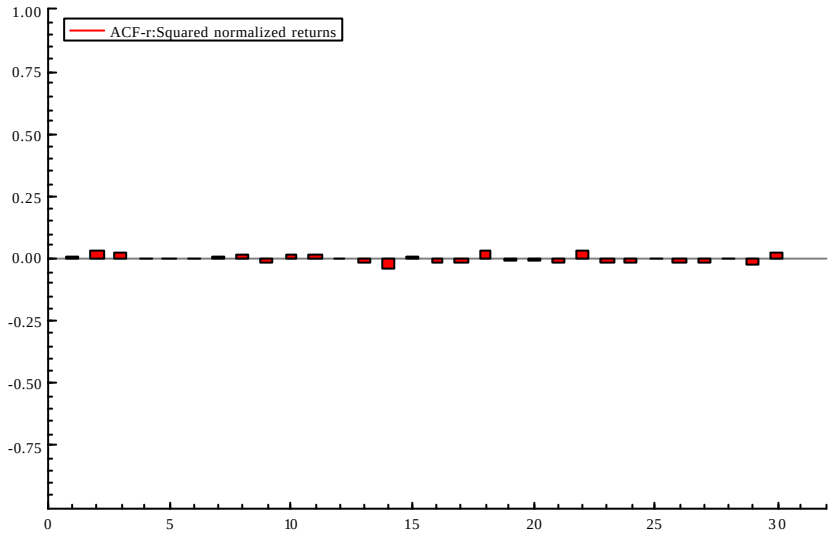
The parameters in table 2.4 are estimated through maximum likelihood[7]. The constant is very small and not significant, while the two coefficients are significant. The sum of the two estimated coefficients α and β (persistency coefficient) is very close to one, meaning that

[7] The estimation has been implemented by PcGive (Ox). The parameters have been restricted to be positive, according to a common practice to avoid negative variances.

even if there is stationarity, the variance is highly persistent. It is interesting to note that the parameter estimates are very close to the fixed parameter values used in RiskMetrics, which can be interpreted as an IGARCH. The conditional variance can be estimated and forecast one-step-ahead through (2.1.B) by adding new information (returns) day by day.

According to the model defined in (2.1), the standardized returns should be distributed as a standard normal: based on this observation, the standardized returns computed with the estimated variance $r_t / \hat{s}_t$ are tested for normality and independence in the squares.

Figure 2.4 Squared Normalized Returns Correlogram



The model succeeds in accounting for second order dependence since the squared standardized returns present no significant autocorrelation, as in figure 2.4 and table 2.5.

Table 2.5 Autocorrelation test on squared standardized returns

H0: no autocorrelation in $r_t^2 / \hat{s}_t^2$		
Lag	F value	p-value
1	0.094221	0.7589
up to 2	1.797	0.1660
up to 3	2.0609	0.1033
up to 30	1.0476	0.3954

The standardized returns present skewness –0.102 and excess kurtosis 1.284. Even if the kurtosis has been reduced by standardizing the returns, the normality hypothesis is still rejected. This is not surprising, since most of the literature on ARCH-type (e.g. Hsieh (1991))

models conclude that a GARCH can only partly account for fat tails in the distribution of the returns.

SV Estimation

The following SV model (from Taylor (1994)) is estimated:

$$(2.2.A)\quad r_t = \exp(h_t/2)e_t \qquad e_t \sim N(0,1);$$
$$(2.2.B)\quad h_t = m + f(h_{t-1} - m) + s_h h_t \qquad h_t \sim N(0,1);\quad e_t, h_t \text{ uncorrelated}; \qquad (2.2)$$
$$(2.2.C)\quad h_1 \sim N\left(m, \frac{s_h^2}{1-f^2}\right)$$

The formulation of equation (2.2.B) follows Chib, Kim, Shephard (1998).[8] The log-variance h_t is modelled as a stationary (assuming $|f|<1$) first order autoregressive process with normal white noise independent from the returns noise. Under the stationarity assumption, the initial value for the log-variance can be drawn from the unconditional distribution defined in (2.2.C).

The first step in estimating the SV model consists in the estimation of the parameters of the log-variance equation (2.2.B): a Gibbs sampler (as described in Chib, Kim, Shephard (1998)) is used in the present work. Secondly the log-variance is filtered and forecast through the application of a particle filter.[9]

Estimation of the parameters

The estimation of the parameters consists in the simulation of a posterior density for each parameter, so that the mean of the simulated distribution can be seen as a posterior mean. Following Chib, Kim, Shephard (1998), the prior distribution for the parameters s_h^2, f and m are respectively an inverse gamma, a beta and a normal.

The main problem using MCMC is that the simulated samples are correlated: many iterations are required to reach the accuracy that could be obtained with independent draws. In this application the first 2,500 iterations (sweeps) have been performed at the beginning without recording the results (burn-in period), in order to ensure that the initial values do not influence the final outcome; then 50,000 sweeps have been performed and recorded. The

[8] The program treats de-meaned returns, but since the mean is quite close to zero the difference can be considered negligible.
[9] The SV model is estimated through an Ox code related to Chib, Kim, Shephard (1998) and available at http://www.nuff.ox.ac.uk/users/shephard/ox/ .

parameters are estimated based on the first eleven years of data and the results are summarised in the table 2.6.

Table 2.6 Estimation Results

parameter	Mean	MC Std Error	Inefficiency	Covariance		
$f \mid r$	0.98708	0.000296	238.24	0.0000174	*-0.669*	*0.0625*
$s_h \mid r$	0.10208	0.00153	601.12	-0.0000379	0.000185	*-0.0648*
$b = \exp(m/2) \mid r$ [10]	0.83814	0.00144	20.56	0.0000181	-0.0000611	0.00482

Even if the log-variance results to be stationary, the coefficient of persistency f is very close to one, indicating that shocks in the log-variance are highly persistent as in the case of GARCH. The numerical standard errors of the sample mean deriving from the Monte Carlo simulation[11] are considered as a measure of the accuracy of the estimates. Clearly the accuracy could be improved by increasing the simulation sample size, i.e. the number of iterations, and by using more complex algorithms: here a choice of 50,000 iterations and of a simple algorithm has been done in order to obtain results in a limited time.

The simulation inefficiency factors measure how well the Markov Chain mixes. The inefficiency factor is defined as the ratio of the numerical variance (i.e. square of the Monte Carlo standard error) and the variance of the sample mean that would derive from drawing independent samples[12]: as independent random draws would be the optimal outcome of the simulation procedure, the most desirable inefficiency factor is the one closest to one. The

[10] The parameter m is expressed $b = \exp(m/2)$ in Chib, Kim, Shephard (1998) because this has an economic interpretation as the modal instantaneous volatility.

[11] To take into account the serial correlatoin between successive samples, the Monte Carlo standard errors are calculated as

$$\hat{R}_{B_M} = 1 + \frac{2B_M}{B_M - 1} \sum_{i=1}^{B_M} K\left(\frac{i}{B_M}\right) \hat{r}(i)$$

where $\hat{r}(i)$ is the estimated autocorrelation of the simulated samples at lag i , B_M is the bandwidth for the correlogram estimation and K is the so called 'Parzen kernel', calculated as

$$K(z) = \begin{cases} 1 - 6z^2 + 6z^3 & for\ z \in [0,0.5] \\ 2(1-z)^3 & for\ z \in [0.5,1] \\ 0 & elsewhere. \end{cases}$$

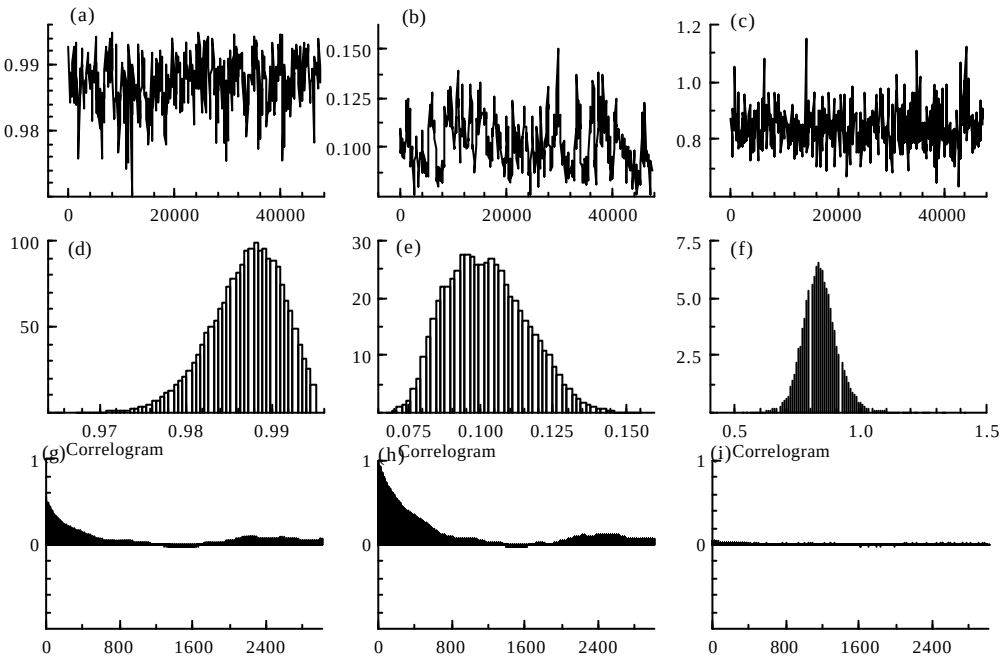
[12] Given a parameter q estimated through the simulation sample average $\hat{q} = E(q)$, the variance of $\hat{q}$ is defined as $\mathrm{var}(\hat{q}) = \frac{\mathrm{var}(q)}{n} \times \sum_{i=-\infty}^{\infty} r_q(i)$, where n is the number of iterations and $r_q(i)$ is the autocorrelation functions of the simulated samples. An independent sampler would give $r_q(i) = 0 \ \forall i \neq 0$, meaning $\mathrm{var}(\hat{q}) = \frac{\mathrm{var}(q)}{n}$, so the inefficiency factor is defined as $\sum_{i=-\infty}^{\infty} r_q(i) = 1 + 2\sum_{i=1}^{\infty} r_q(i)$.

inefficiency factor can be interpreted as the number of times the algorithm needs to be run to produce the same accuracy in the estimate that would derive from independent draws. The inefficiency factors can in general be reduced by increasing the number of iterations[13]. The inefficiency factor for the constant scaling factor $b = \exp(m/2)$ is massively lower then for the other two parameters, following the results normally found in other empirical works (e.g. Kim, Shephard and Chib (1998)[14]).

The last three columns of table 2. contain the parameters' covariance matrix; in the upper triangle the correlations have been reported instead of the covariance, because they give a more explicit information on the relation between the parameters.

Figure 2.5 gives a graphical illustration of the simulation results.

Fig. 2.5 Gibbs sampler outcome



The upper graphs represent the full sample of iterations for each parameter, which give the simulated marginal densities $f(f \mid r)$, $f(s \mid r)$, $f(b \mid r)$ respectively in (a), (b) and (c). The three graphs in the middle (d), (e), (f) show the histograms of the simulated marginal

[13] In this case the magnitudes of these factors are similar to the ones obtained in Kim, Shephard and Chib (1998) with 1,000,000 iterations: this suggests that the number of iterations used should be appropriate.

[14] In Kim, Shephard and Chib (1998) it is shown that more complex estimation algorithms are definitely more efficient (lower inefficiency factors) for all the parameters. In the present work the simplest and quickest algorithm has been chosen, considering that the resulting estimated parameters are anyway very close even using more efficient algorithm in Kim, Shephard and Chib (1998).

densities: the histogram corresponding to b (f) looks perfectly symmetric and very concentrated on the mean, while the other two look slightly skewed. For the simulation sample mean to be a good estimate of the parameter the iterations should be like draws from independent random variables. The correlograms, showed in the bottom graphs, report the best performance for b. The degree of serial correlation in the draws is an indicator of how well the simulation algorithm behaves. While for b the autocorrelation is almost inexistent, there is significant autocorrelation for the other two parameters: in particular s presents (h) very important serial correlation until 800 lags. This result is just a graphical confirmation of the information already contained in the inefficiency factors. An improved algorithm proposed by Chib, Kim, Shephard (1998), an offset mixture with appropriately adapted Gibbs sampler, is also implemented: the resulting parameter estimates are not very different (0.98762, 0.096769, 0.8387), while there is a significant improvement in the inefficiency factors (82, 175, 13) and in the standard errors (0.000174, 0.000848, 0.000351). The algorithm is more complex and hence slower.

Volatility Estimation and Diagnostic Checks

Filtering the volatility $\exp(h_t/2) \mid R_t$ means finding the filtering density $f(h_t \mid R_t, q)$, where the parameters, indicated synthetically as q, are considered constant and known. The information consists of the series of returns up to time t. Once the log-variance has been initialised according to (2.3.C), a particle filter algorithm is applied to sample draws from $h_t \mid R_t, q$ given a sample from $h_{t-1} \mid R_{t-1}, q$: the output consists of M (=2000) simulated values for each $h_t \mid R_t, q$ t=1,...,T. Analogously M values for each forecast log-variance $h_{t+1} \mid R_t, q$ can be computed as draws from $h_{t+1}^j \mid h_t^j \sim N\left(\hat{m} + \hat{f}(h_t^j - \hat{m}); \hat{s}_h^2\right)$. The mean of the simulation samples are considered as estimates of the filtered and forecast volatility[15].
Unlike the case of GARCH where the estimated and the one-step-ahead forecast volatility are computed exactly in the same way since all the components of the variance equation are known in t, in the case of SV the forecast volatility $s_t \mid R_{t-1}$ is different from the filtered one

[15] The expected volatility is calculated as $\hat{s}_t = E\left(\exp(h_t/2) \mid R_t, q\right) = \frac{1}{M}\sum_{j=1}^{M} \exp(h_t^j/2)$ and $\hat{s}_{t+1} = E\left(\exp(h_{t+1}/2) \mid R_t, q\right) = \frac{1}{M}\sum_{j=1}^{M} \exp(h_{t+1}^j/2)$ respectively for t covering the full sample and the number of simulation M equal to 2000.

$s_t \mid R_t$, as the filtered volatility is computed through the particle filter and the forecast one is then obtained from the transition equation (2.3.B) by simulation.

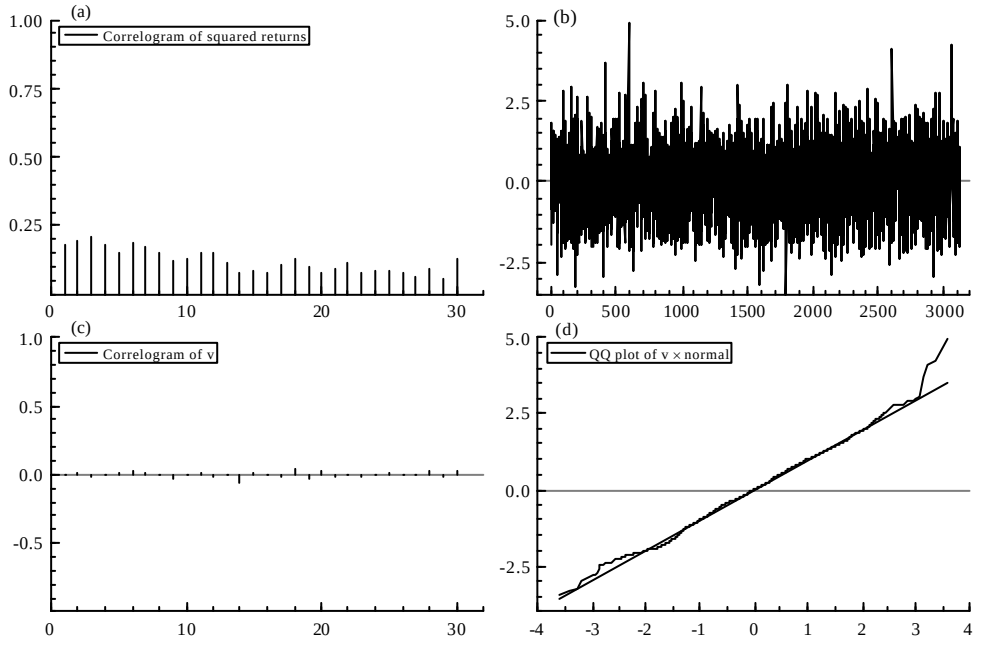
According to the model defined in (2.3), since the conditional returns $r_t \mid h_t$ are distributed as a normal $N(0;\exp(h_t))$, the standardised returns $y_t = r_t / \exp(h_t/2)$ should be distributed as an IID N(0;1). Hence the goodness of the model in representing the data can be evaluated by checking the features of y_t. Unlike the case of GARCH, in the SV filtering framework the final output for the log-variance is not just a series of number, but is an n×m matrix containing m draws from the distribution of h_t for every t=1,...,n.

In the literature about SV tests consider the full simulated distribution of the variance. Considering the M draws on h_{t+1} from the predictive density, the probability of y_{t+1}^2 being less than its observed value is

$$\Pr\left(y_{t+1}^2 \le y_{t+1}^{2oss} \mid R_t, q\right) = \frac{1}{M}\sum_{j=1}^{M} \Pr\left(y_{t+1}^2 \le y_{t+1}^{2oss} \mid h_{t+1}^j, q\right) = u_{t+1}^M \qquad (2.3)$$

The random variable u_{t+1}^M converges to an IID uniform (0,1) random variable for $M \to \infty$ if the model is correctly specified. All the diagnostic tests are based on the variable $v^M{}_{t+1} = \Phi^{-1}(u^M{}_{t+1})$ for t=1,...,n-1 that should turn out to be N(0;1) if the SV model has been correctly defined and estimated.

Fig. 2.7 Diagnostic tests



The comparison between graph (a) and (c) shows that the dependence in the data, which emerges as serial correlation in the squared returns, is captured effectively by the SV model because the correlogram of the transformed variable v is almost flat at zero. Graph (b) illustrates the series of the transformed data, i.e. the series v^{M}_{t+1} for t=1,...,n-1. For the SV model to be correct this series should come out from a standard normal distribution. A QQ plot is presented in graph (d), which compares the actual distribution with a standard normal: the middle values are almost perfectly represented by the normal, but the outcome is worst for extreme values of v: a similar result is found in Chib, Kim, Shephard (1998). The normality hypothesis is accepted on v: the skewness and excess kurtosis are in fact 0.0231 and 0.0456 respectively. These results proved that the SV quite well.

3. Value at Risk

The main issue in VaR estimation is the definition of the one-step-ahead portfolio return distribution: on the future distribution of the profits and losses (portfolio value changes), the quantile corresponding to the confidence level chosen a represents the estimated maximum loss. Formally, the VaR estimated in t for the following period t+1 can be defined to be the quantity such that[16]

$$\Pr\left(\Delta V_{t+1} \le -VaR_{t+1} \mid I_t\right) = a\% \tag{3.1}$$

where I_t is the information available in t, V_t is the portfolio value in t and $\Delta V_{t+1} = V_{t+1} - V_t$ represents the portfolio profits and losses.

In the present work a unit position on FTSE100 has been considered, so that the value of the portfolio is just the index price, and the profits and losses are given by its changes.
Assuming the conditional normality of the returns

$$r_{t+1} \mid R_t \sim N\left(0; s^{2}_{t+1}\right) \tag{3.2}$$

VaR can be easily calculated following the standard variance-covariance approach as[17]

$$VaR_{t+1} \cong -\Phi^{-1}(a) \times P_t \times s_{t+1} \tag{3.3}$$

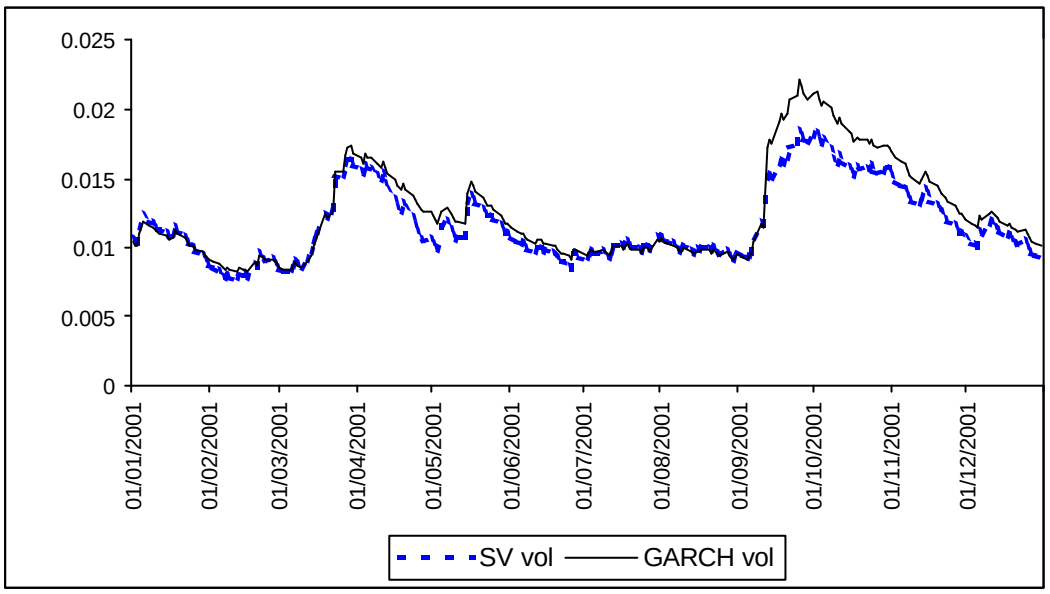
[16] The VaR is taken with the negative sign because it is always calculated as an absolute value, but in this expression it represents a loss.

[17] The common approximation $\Delta P_{t+1} \cong P_t r_{t+1}$ is adopted.

In this work a confidence level of 95% is used, i.e. $\Phi^{-1}(a) = -1.65$, and the one step ahead forecast volatility is estimated alternatively with the GARCH(1,1) and SV as seen above. The conditional volatility in t+1 is estimated on the information up to time t, and the returns are assumed normal conditionally to this estimate.

Hence it is evident the crucial importance of forecasting volatility one-step-ahead. Figure 3.1 shows the forecast volatility coming from the two models: the GARCH volatility is generally higher then the SV volatility, particularly during the period of high volatility following the 11th September. An explanation could be related to the persistency coefficient which is higher in GARCH (alpha+beta=0.991) then in SV (phi=0.987), that is a typical result as pointed out by Shephard (1996).

Fig. 3.1 Volatility Forecast

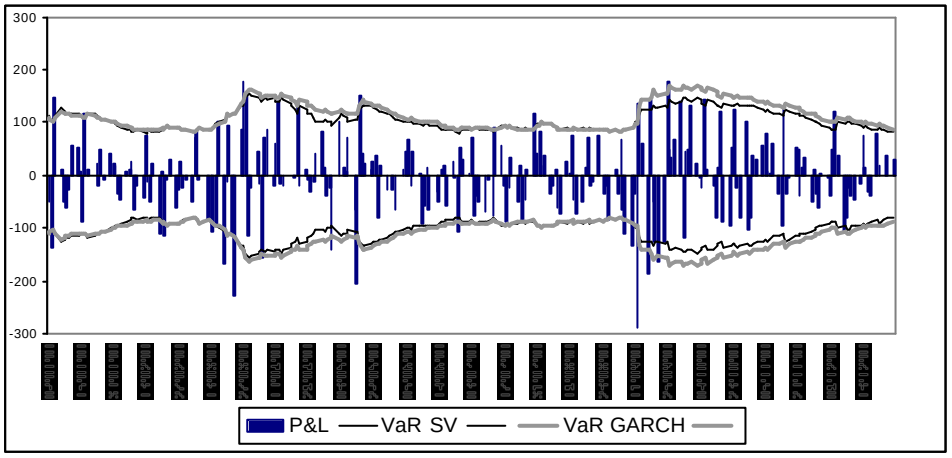


VaR is estimated day by day on the last year of data (01/01/2001-31/12/2001) and the backtest is analysed. Since the return distribution is symmetric around zero, a symmetric 90% confidence interval for each period can be built centred on zero and delimited by $\pm VaR$. Even if the attention commonly focuses on losses, from a statistical point of view profits and losses can be treated in the same way as realizations of the returns, and given the symmetry hypothesis the limit on profits is the same as the limit on losses. The backtest is performed comparing the VaR calculated with both models and the realized price changes: $\pm VaR$ define a symmetric 90% confidence interval, and the attention focuses on the realized proportion of profits and losses which fall outside of this confidence interval (misses). A perfect model

should produce a confidence interval that contains exactly 90% of the realizations, that is 10% of the outcomes should be misses. Figure 3.2 illustrates the outcome of the backtest.

The confidence interval defined by GARCH is in general wider then the one defined by SV. The percentage of misses is in fact about 11.5% for GARCH and 13.8% for SV. Both the models underestimate the extreme price changes, since both present an actual proportion of misses greater then 10%.[18] The basic assumption for both models is the conditional normality of the returns: the unconditional distribution has fat tails, but this is not enough to account for the fat tails of the actual distribution. Both GARCH and SV model can be modified to account for a greater proportion of leptokurtosis by changing the distribution of e_t from a normal in a Student-t: this goes beyond the scope of the present work, because the estimation would become much more complex, especially in the case of SV, and also because the VaR estimation, according to the RiskMetrics methodology, relies on the conditional normality hypothesis, so that changing this hypothesis would change the VaR computation.

Fig. 3.2 VaR and backtest on the last year: 01/01/2001-31/12/2001



A formal comparison of the performances of the two models can be done through appropriate Likelihood Ratio tests. Following Christofersen (1998), the goodness of an interval forecast is evaluated on the basis of two main requirements which need to be satisfied:

- Correct coverage: the observed proportion of misses is approximately equal to a %, assuming independence.

[18] It has to be noted that the forecast period includes the 11th September 2001: the absolute return is obvoiusly very high on that day producing a miss.

- Independence: hits and misses are independently distributed (clustering ruled out).

The implementation of the two mentioned criteria requires the definition of an indicator function:

$$I_t = \begin{cases} 1 & if\ \Delta P_{t+1} \in \left[-VaR_{t+1}; VaR_{t+1}\right] \quad (hit) \\ 0 & if\ \Delta P_{t+1} \notin \left[-VaR_{t+1}; VaR_{t+1}\right] \quad (miss) \end{cases} \tag{3.4}$$

Based on the proportion of misses and hits compared to the predefined probability of 90%, two LR tests can be performed, with the null hypothesis of correct coverage and independence respectively. The results of the two tests, i.e. the tests values and relative p-values, are presented in table 3.1.

Table 3.1 LR tests on the two interval forecast

	GARCH	SV
	Correct Coverage (LR_{cc})	
LR (c_1^2)	**0.6208** (0.431)	**3.7772** (0.052)
	Independence (LR_{ind})	
LR (c_1^2)	**8.8171** (0.003)	**3.8167** (0.051)

As expected the value of the first LR test is much lower in the case of GARCH, even if the null of correct coverage is accepted in both cases. Hence GARCH performs better according to the correct coverage criterion, but the opposite conclusion is reached according to the independence criterion. The null of independence is in fact accepted for SV but rejected for GARCH: this means that in the case of SV, even if the interval forecast is not wide enough to contain the predicted proportion of realizations, the misses are independently distributed over the period considered, that is not true for GARCH. The conclusion is that none of the two models produce an optimal interval forecast: in the case of SV both the hypotheses of correct coverage and independence are accepted but quite weekly, while in the case of GARCH the correct coverage is definitely accepted but the independence is rejected. In conclusion, despite at a first look GARCH appears to perform better, from a statistical point of view the choice between the two models is not straightforward. Moreover, using the results from the improved SV algorithm to estimate variance, the LR tests produce a better output: LRcc=2.3556 (0.118), LRind=3.3437 (0.07). Combining the two tests, a single statistic is obtained which is distributed as a $c^2(2)$. The results are presented in table 3.2. Only the SV model with improved algorithm doesn't reject the joint hypothesis of correct coverage and independence.

Table 3.2 Joint LR test

	LRcc+LRind c_2^2
GARCH	9.4379 (0.009)
SV	7.5939 (0.022)
SV improved	5.7313 (0.057)

It has to be noted that SV VaR has been estimated using the expected value as a volatility forecast: since the outcome of the SV model by applying a filtering algorithm is a vector of M draws $[h_{t+1}^1\ h_{t+1}^2 .....h_{t+1}^M]$ for each period t, a loss of information occurs when taking the average, which could lead to a reduction in the ability of the SV model to capture the fat tails. In fact in the previous section the distribution of the one-step-ahead return is assumed to be normal given the forecast value $\hat{s}_{t+1}$, but the returns are normally distributed conditionally on the volatility distribution. It could be argued that the return are normal conditionally on each single realisation from the simulation sample $[h_{t+1}^1\ h_{t+1}^2 .....h_{t+1}^M]$

$$r_{t+1} \mid h_{t+1}^j \ N(0; \exp(h_{t+1}^{j)})) \quad \forall j \tag{3.5}$$

The predictive distribution given the information up to time t could then be approximated through the filtering output as follows[19]

$$f(r_{t+1} \mid R_t) = \int f(r_{t+1} \mid h_{t+1}) f(h_{t+1} \mid R_t) dh_{t+1} \cong \frac{1}{M} \sum_{j=1}^{M} f_N(r_{t+1} \mid h_{t+1}^j) \tag{3.6}$$

Starting again from the definition in (3.1):

$$\Pr(\Delta P_{t+1} < -VaR_{t+1} \mid R_t) = \int \Pr(\Delta P_{t+1} < -VaR_{t+1} \mid h_{t+1}) f(h_{t+1} \mid R_t) dh_{t+1} \cong \frac{1}{M} \sum_{j=1}^{M} \Phi\left(\frac{-VaR_{t+1}}{P_t \exp(h_{t+1}^j / 2)} \right)$$

(3.7)

The right hand side of (3.7) is an approximation deriving from the application of a particle filter. The conditional distribution $f(h_{t+1} \mid R_t)$ is in fact approximated by the simulated sample $[h_{t+1}^1\ h_{t+1}^2 .....h_{t+1}^M]$ and the probability of each value is 1/M. By imposing the expression in (3.7) to be equal to the predefined probability 0.05 the VaR can be computed for each t in order to build a forecast interval over time. The equation

[19] In Shephard (1996) it is stated that $\Pr(r_{t+1} \le x \mid R_t) = F_{r_{t+1}|R_t}(x) = 1/M \sum_{j=1}^{M} \Pr(r_{t+1} \le x \mid h_{t+1}^j)$

$$\frac{1}{M}\sum_{j=1}^{M}\Phi\left(\frac{-VaR_{t+1}}{P_t \exp(h_{t+1}^{j}/2)}\right)=0.05 \qquad (3.8)$$

is solved numerically using the Excel solver. Coherently with the ex ante expectations, the interval is larger then the one defined before by using the average SV volatility. However the increase in the wideness of the interval is not relevant enough to improve substantially the performances of the SV model in estimating VaR. The percentage of misses only drops to 13.4%, which is still much higher then the percentage of misses in the GARCH case.

Another way of comparing the forecasting performances related to VaR is through a loss function. Following Gonzalez-Rivera at al (2002), a loss function from quantile estimation is adopted: this function focuses only on the loss-side and considers the magnitude of the loss:

$$V=\frac{1}{n}\sum_{t=1}^{n}|\Delta P_{t+1}+VaR_{t+1}|\times\left[(1-a)1_{\{\Delta P_{t+1}<-VaR_{t+1}\}}+a1_{\{\Delta P_{t+1}>-VaR_{t+1}\}}\right] \qquad (3.9)$$

The results are quite similar for the two models with a slight preference for GARCH (GARCH: V=8.32, SV: V=8.33). The improved SV performs better (V=8.26).

Conclusions

Two volatility models are discussed and estimated: GARCH(1,1) and log-normal AR(1) SV. Since the SV estimation is definitely more computationally demanding, this paper analysed their performances within a simple VaR calculation in order to evaluate if the use of SV models in contrast to GARCH is worthwhile. The VaR estimated on FTSE 100 data does not presents clear evidence in favour of SV: on the contrary GARCH, looks quite useful, even if both the models underestimate losses, probably due to the common underlying assumption of conditional normality. The results anyway are not fully conclusive: they could be related to the particular data set used on one hand, and on the estimation procedure on the other hand, as it is evident from the two different algorithms used in SV. Moreover the results are sensible to the particular evaluation techniques adopted to analyse the performances, which can focus on different aspects of the issue.

References

Baillie R. T., Bollerslev T., 1989, "The Message in daily Exchange Rates: a Conditional Variance Tale", *Journal of Business and Economic Statistics*, Vol. 7, pp. 297-305.

Bluhm H. H. W., Yu J., (2001), "Forecasting Volatility: Evidence from the German Stock Market", Working Paper, University of Auckland.

Bollerslev T., 1986, "Generalised Autoregressive Conditional Heteroskedasticity", *Journal of Econometrics*, Vol. 31, pp. 307-327.

Bollerslev T., 1987, "A Conditionally Heteroskedastic Time Series Model for Speculative Prices and Rates of Return", *The Review of Economics and Statistics*, Vol. 69, pp. 542-547.

Bollerslev T., Chou R. Y. and Kroner F. K., 1992, "ARCH Modelling in Finance", *Journal of Econometrics*, Vol. 52, pp. 5-59.

Carter C., Gerlach R., Kohn R., 1999. "Diagnostic for Time Series Analysis", *Journal of Time Series Analysis*, Vol. 20, No. 3, pp. 309-330.

Chib S., Kim S., Shephard N., 1998. "Stochastic Volatility: Likelihood Inference and Comparison with ARCH Models". *Review of Economic Studies*, Vol. 65, pp. 361-393.

Chib S., Nardari F., Shephard N., 2002. "Markov Chain Monte Carlo Methods for Stochastic Volatility Models", *Journal of Econometrics*, Vol. 108, pp. 281-316.

Christofersen P. F.,1998, "Evaluating Interval Forecasts", *International Economic Review*, Vol. 39, pp. 842-862.

Doornik J. A., 1996, "Ox: Object Oriented Matrix Programming", 1.10, London: Chapman & Hall.

Engel J., Gizycki M., 1999, "Value at Risk: on the Stability and Forecasting of the Variance-Covariance Matrix", Research Discussion Paper.

Engle R. F., 1982, "Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of the United Kingdom Inflation", *Econometrica*, Vol. 50, pp. 987-1007.

Gonzales-Rivera G., Lee T. H., Mishra S., (2002), "Forecasting Volatility: A Reality Check Based on Option Pricing, Utility Function, Value at Risk, and Predictive Likelihood", Working Paper available at www.gloriamundi.org.

Granger C. W. J., Poon S. H., (2003) "Forecasting Volatility in Financial Markets: A Review", *Journal of economic Literature*, forthcoming.

Hansen P. R., Lunde A., (2001), "A Forecast Comparison of Volatility Models: Does Anything Beat a GARCH(1,1)?", Brown University.

Hansson B., Hordahl P., (2002), "Forecasting Variance Using Stochastic Volatility and GARCH", *The European Journal of Finance*, forthcoming.

Harvey A. C., 1989, *Forecasting, Structural Time Series Models and Kalman Filter*, Cambridge: Cambridge University Press.

Harvey A. C., 1993, *Time Series Models*, Harvester-Wheatsheaf.

Harvey A. C., Ruiz E., Shephard N., 1994. "Multivariate Stochastic Variance Models", *Review of Economic Studies*, Vol. 61, pp. 247-264.

Hsieh D. A., 1991, "Chaos and Non-linear Dynamics: Application to Financial Markets", *The Journal of Finance*, Vol.46, pp. 1839-1877.

Jaquier E., Polson N. G., Rossi P. E., 1994, "Bayesian Analysis of Stochastic Volatility Models (with discussion) ", *Journal of Business and Economic Statistics*, Vol. 12, pp. 371-417.

Jorion P., 1997, *Value at Risk: the New Benchmark for Controlling Market Risk*, Chicago: Irwin.

J.P.Morgan, 1996, *RiskMetrics Technical Document*, New York.

Lehar A., Scheicher M., Scittenkopf C., (2002), "GARCH vs. Stochastic Volatility: Option Pricing and Risk Management", *Journal of Banking and Finance*, Vol. 26, pp. 323-345.

Pederzoli C., Torricelli C., (1999), "Una Rassegna sui Metodi di Stima del Value at Risk (VaR)", Materiali di Discussione, Università di Modena, N. 263.

Pitt M. K., Shephard N., 1999, "Filtering via Simulation: Auxiliary Particle Filter", *Journal of the American Statistical Association*, Vol. 94, pp. 590-599.

Shephard N., 1996, "Statistical Aspects of ARCH and Stochastic Volatility", in D.R. Cox, O.E. Barndorff-Nielson and D.V. Hinkley (eds.), *Time Series Models in Econometrics, Finance and Other Fields*, London: Chapman & Hall, pp. 1-67.

Shephard N., 1998, "SvPack 2.0: Software for Stochastic Volatility Models", downloadable at http://www.nuff.ox.ac.uk/users/shephard/ox/

Taylor S. J., 1994, "Modelling Stochastic Volatility", *Mathematical Finance*, Vol. 4, pp. 183-204.

Taylor S. J., 1986, *Modelling Financial Time Series*, John Wiley, Chichester.

Verbeek M., 2000, *A Guide to Modern Econometrics*, Wiley.

Yu J., (2002), "Forecasting Volatility in the New Zealand Stock Market", *Applied Financial Economics*, Vol. 12, pp. 193-202.