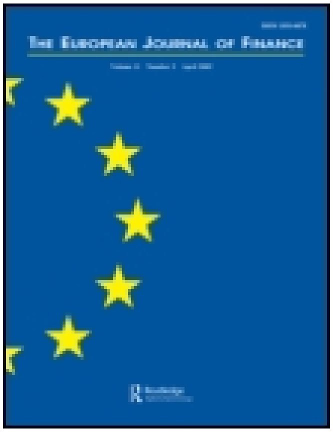

The European Journal of Finance
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/rejf20

# Stochastic Volatility and GARCH: a Comparison Based on UK Stock Data

Chiara Pederzoli [a]

[a] University of Milano-Bicocca , Milano, Italy

Published online: 08 Feb 2011.

*The European Journal of Finance*
*Vol. 12, No. 1, 41–59, January 2006*

# Stochastic Volatility and GARCH: a Comparison Based on UK Stock Data

CHIARA PEDERZOLI
*University of Milano-Bicocca, Milano, Italy*



ABSTRACT *This paper compares two types of volatility models for returns, ARCH-type and stochastic volatility (SV) models, both from a theoretical and an empirical point of view. In particular a GARCH(1,1) model, an EGARCH(1,1) model and a log-normal AR(1) stochastic volatility model are considered. The three 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 to evaluate how well the models fit the data. The models are used to obtain daily volatility forecasts and these volatilities are used to estimate the "VaR" on a simple one-unit position on FTSE100. The VaR accuracy is tested by means of a backtest. While the results do not lead to a straightforward preference between GARCH(1,1) and SV, the EGARCH shows the best performance.*

KEY WORDS: Volatility models, stochastic volatility, GARCH, value at risk

## 1. Introduction

In many financial applications, the specification of a model to represent the behaviour of returns 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 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 than 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 out-of-sample predictive performance, the evidence is not clearly conclusive. For example, Gonzalez-Rivera *et al.* (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

*Correspondence Address:* Chiara Pederzoli, University of Milano-Bicocca, Piazza dell' Ateneo Nuovo, 1, 20126 Milano, Italy. Email: chiara.pederzoli@unimib.it

1351-847X Print/1466-4364 Online/06/010041–19 © 2006 Taylor & Francis
DOI: 10.1080/13518470500039121

GARCH in a VaR framework, while SV is preferred in option pricing. 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, three 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 alternative models are evaluated by means of backtest.

The paper is organized 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 three models, a log-normal first order autoregressive (AR(1)) SV, a GARCH(1,1) and an EGARCH(1,1) are estimated. Some graphical analyses are omitted from the paper and are available on request. The third section presents the implementation of VaR estimation and related backtest using alternatively the volatilities coming from the different models estimated. The final section provides conclusions.



## 2. 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

$$r_t = \sigma_t \varepsilon_t \quad \varepsilon_t \sim IID(0;1) \tag{1}$$

In Equation 1 $\varepsilon_t$ is a zero-mean white noise often assumed to be normal and $\sigma_t$ is the time-varying volatility. This model handles the dynamics in the variance, while ignoring the possible dynamics in the mean, which is a common choice when dealing with daily returns (e.g. the RiskMetrics model by J. P. Morgan, which is a standard for VaR estimation). Assuming that $\varepsilon_t$ is a normal white noise, the returns conditional on $\sigma_t$ are normal. While the normality is often assumed for the conditional distribution, by modelling $\sigma_t$ as being time varying the unconditional distribution is leptokurtic. Different specifications for $\sigma_t$ 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 $\sigma_t$ 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 $\sigma_t^2 = f(r_{t-1}^2, \ldots, r_{t-p}^2, \sigma_{t-1}^2, \ldots, \sigma_{t-q}^2)$. The simple or moving average models are special cases. One of the most appealing features of the observation-driven models is that the one-step-ahead forecast density is defined explicitly; assuming normality:

$$r_t | R_{t-1} \sim N(0; \sigma_t^2) \quad R_{t-1} = (r_1, \ldots, r_{t-1}) \tag{2}$$

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)

$$\begin{aligned} &(3.A) \quad r_t = \sigma_t \varepsilon_t \quad \varepsilon_t \sim NID(0; 1) \\ &(3.B) \quad \sigma_t^2 = \gamma + \alpha r_{t-1}^2 + \beta \sigma_{t-1}^2 \end{aligned} \tag{3}$$

Since $\varepsilon_t$ 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 $\alpha + \beta < 1$ 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 $\gamma/(1-\alpha-\beta)$ and kurtosis greater than 3 (i.e. leptokurtic distribution). Similar features apply to the generic ARCH(p) and GARCH(p,q) models. Among several modifications to the standard GARCH models, Nelson (1991) developed a very successful asymmetric GARCH model, the Exponential GARCH (EGARCH), which accounts for asymmetric impact of returns on conditional variance. For an EGARCH(1,1) the Equation 3b is modified to

$$\ln \sigma_t^2 = \gamma + \beta \ln \sigma_{t-1}^2 + \alpha \left| \frac{r_{t-1}}{\sigma_{t-1}} \right| + \omega \frac{r_{t-1}}{\sigma_{t-1}} \tag{4}$$

The parameter $\omega$ quantifies the asymmetry. The logarithmic formulation ensures a positive conditional variance.

The normality assumption makes estimating an ARCH-type model very easy by maximum likelihood. The estimation becomes more complex if the normality assumption is relaxed: most commonly a Student-t distribution for $\varepsilon_t$ is assumed, which increases the ability of the model to capture the fat tails of the actual distribution of the returns (Baillie and Bollerslev, 1989). 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 characterizing GARCH compared to SV models.

Parameter-driven models, represented by SV models, define the volatility as a stochastic process. Unlike the case of GARCH, a contemporaneous innovation appears in the variance equation, accounting for the random new information that characterizes financial markets. The variance is unobservable but it can be estimated. The conditional normality still holds, but the forecast density $r_t|R_{t-1}$ is not defined explicitly.

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

$$\begin{aligned} &(5.A) \quad r_t = \sigma_t \varepsilon_t \quad \sigma_t = \exp(h_t/2) \qquad \varepsilon_t \sim N(0;1) \\ &(5.B) \quad h_t = \gamma + \phi h_{t-1} + \sigma_\eta \eta_t \quad \eta_t \sim N(0;1) \end{aligned} \tag{5}$$

The logarithmic formulation ensures the positiveness of the variance. In the simplest case $\varepsilon_t$ and $\eta_t$ 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 $\phi$ represents the persistence of the log-variance: if $|\phi| < 1$ the log-variance is stationary. Assuming this stationarity condition, the log-variance follows an autoregressive process AR(1) with unconditional mean $E(h_t) = \gamma/(1-\phi)$ and unconditional variance $Var(h_t) = \sigma_\eta^2/(1-\phi^2)$. As in the case of GARCH the returns are conditionally normal:

$$r_t|h_t \sim N(0; \exp(h_t)) \tag{6}$$

The main difference compared to GARCH is that in the SV model $f(r_t|\sigma_t = \exp(h_t/2)) \neq f(r_t|R_{t-1})$ 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

$$f(r_t|R_{t-1}) = \int f(r_t|h_t)f(h_t|h_{t-1})f(h_{t-1}|R_{t-1})dh_t dh_{t-1} \tag{7}$$

where $f(r_t|h_t)$ is a normal as in (6), and $f(h_t|h_{t-1})$ is a normal with mean $\gamma + \phi h_{t-1}$ and variance $\sigma_\eta^2$. The unconditional distribution of the return is a non-standard one 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 is higher than 3. 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 log-variance. A mixture of normal distributions has fat tails by definition. While the returns are white noise by definition, if the stationarity condition $|\phi| < 1$ is satisfied the squared returns dynamic is driven by an autocorrelation function very close to that of an ARMA(1,1) process, 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:

(1) high kurtosis (fat tails);
(2) low 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 intractability of the predictive density $f(r_t|R_{t-1})$.

### *2.1 SV Estimation and Testing*

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. 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 and Shephard (1998) (henceforth CKS98).

The log-variance can be computed using the full sample $R_T = (r_1, \ldots, r_t, \ldots, r_T)$, in which case the estimate $\exp(h_t/2)|R_T$ is a "smoothed volatility", or it can be based on the observations up to the considered period producing the "filtered volatility"[1] $\exp(h_t/2)|R_t$. MCMC in the context of SV estimation is used within a Bayesian framework and it consists in drawing correlated samples (Markov Chain) from the required distributions. MCMC provides both the estimates of the parameters and of the smoothed volatility from a single algorithm, by simulating the conditional densities $f(h_t|R_T)$ $t = 1, \ldots, T$ and $f(\theta|R_T)$ and taking the averages (posterior mean) as estimates. As pointed out in Jaquier *et al.* (1994), a Markov Chain sampler can be built by splitting the joint posterior density $f(h, \theta|R_T)$ in the two marginal densities $f(\theta|R_T, h)$ and $f(h|R_T, \theta)$ and then alternating them in the simulation.

While in the case of ARCH-type models the volatility 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 volatility filtering procedure assumes known values for the parameters $\theta$ of the model: filtering the volatility $\exp(h_t/2)|R_t$ means finding the filtering density $f(h_t|R_t, \theta)$. 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. 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 $\sigma_t|R_{t-1}$ is different from the filtered one $\sigma_t|R_t$, as the filtered volatility is computed through the filter and the forecast one is then obtained from the transition Equation 5b by simulation.

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

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

$$f(h_{t+1}|R_t) = \int f(h_{t+1}|h_t) f(h_t|R_t) dh_t \tag{8}$$

2. Use Bayes theorem to compute the filtering density:

$$f(h_{t+1}|R_{t+1}) = \frac{f(r_{t+1}|h_{t+1}) \times f(h_{t+1}|R_t)}{f(r_{t+1}|R_t)} \propto f(r_{t+1}|h_{t+1}) \times f(h_{t+1}|R_t) \tag{9}$$

   where

$$f(r_{t+1}|R_t) = \int f(r_{t+1}|h_{t+1}) f(h_{t+1}|R_t) dh_{t+1} \tag{10}$$

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 $h_t|R_t$ in the SV model, by "particles", that is by a finite number $M$ of values $h_t^1, \ldots, h_t^M$, associated with discrete probabilities $\pi_t^1, \ldots, \pi_t^M$. The filtering density is approximated with the discrete set of values, which are considered like samples from that density. This allows computing the prediction density in Equation 8 with the approximation

$$\hat{f}(h_{t+1}|R_t) = \sum_{j=1}^{M} \pi_t^j f(h_{t+1}|h_t^j) \tag{11}$$

Clearly the higher is the number of simulated values $M$ the more accurate is the filtering sampler. The density in Equation 11 is an "empirical prediction density", that can be used to formulate an "empirical filtering density":

$$\hat{f}(h_{t+1}|R_{t+1}) \propto f(r_{t+1}|h_{t+1}) \times \sum_{j=1}^{M} \pi_t^j f(h_{t+1}|h_t^j) \tag{12}$$

In the literature the probabilities $\pi_t^1, \ldots, \pi_t^M$ are usually assumed to be equal, *i.e.* $\pi_t^j = 1/M$ $\forall j = 1, \ldots, M$. 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 $f(h_{t+1}|h_t^j)\, j = 1, \ldots, M$. The output of the filtering algorithm consists of $M$ simulated values for each $h_t|R_t\ t = 1, \ldots, T$. Analogously $M$ values for each forecast log-variance

$h_{t+1}|R_t$ can be computed as draws from $h^j_{t+1}|h^j_t \sim N(\hat{\mu} + \hat{\phi}(h^j_t - \hat{\mu}); \hat{\sigma}^2_\eta)$. The mean of the simulation samples are considered as estimates of the filtered and forecast volatility.

In GARCH models diagnostic tests are generally based on the standardized returns $y_t = r_t/\sigma_t$ which should be *IID* $N(0;1)$. According to the model defined in Equation 5, since the conditional returns $r_t|h_t$ are distributed as normal $N(0; \exp(h_t))$, the standardized returns $y_t = r_t/\exp(h_t/2)$ should be distributed as *IID* $N(0;1)$ as in the case of GARCH. Unlike the case of GARCH, however, in the SV filtering framework the final output for the log-variance is not just a series of number, but it is an $n \times m$ matrix containing $m$ draws from the distribution of $h_t$ for every $t = 1, \ldots, n$. In the literature about SV estimated by MCMC, tests consider the full simulated distribution of the variance. Considering the $M$ draws on $h_{t+1}$ from the predictive density, the probability of $y^2_{t+1}$ being less than its observed value is

$$\Pr(y^2_{t+1} \leq y^{2oss}_{t+1}|R_t, \theta) = \frac{1}{M}\sum_{j=1}^{M} \Pr(y^2_{t+1} \leq y^{2oss}_{t+1}|h^j_{t+1}, \theta) = u^M_{t+1} \qquad (13)$$

The random variable $u^M_{t+1}$ converges to an IID uniform (0,1) random variable for $M \to \infty$ if the model is correctly specified. Diagnostic tests can be performed on the variable

$$v^M_{t+1} = \Phi^{-1}(u^M_{t+1}) \quad t = 1, \ldots, n-1 \qquad (14)$$

The variable $v^M_{t+1}$ should turn out to be N(0;1) if the SV model has been correctly defined and estimated.

## 3. GARCH and SV Estimation on UK Stock Data

### 3.1 The Data

The data consist of FTSE100 daily prices covering the period 01/01/1990 to 31/12/2001.[2] The date 11 September 2001 is included in the observations: the corresponding return is high in absolute value but its magnitude is not very different from other episodes along the series. The volatility models deal with returns: the series of FTSE100 (continuously-compounded) return is presented in Fig. 1.

Figure 1 suggests that the returns are moving around an approximately zero mean with time-varying clustering volatility. Table 1 presents some basic statistics describing the data. The sample mean, very close to zero, supports the assumption made in the return model. The Augmented Dickey–Fuller (ADF) test shows that the series is stationary. The negative skewness and the high kurtosis, and the consequent rejection of the normality hypothesis by the Jarque–Bera (JB) test, confirm the common empirical finding that daily returns are far from being Gaussian.

According to the market efficiency hypothesis, the returns are expected to be serially uncorrelated: the series considered here presents little evidence of autocorrelation.[3] 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 6.05 and 61.87 respectively, the hypothesis of no autocorrelation is strongly rejected at any lag tested and the Ljung–Box (LB) test shows strong autocorrelation in the squared returns (Table 1). The correlogram in Fig. 2 graphically supports this result.

Moreover, an ARCH test shows the presence of “ARCH effects” in the returns (Table 1). The features of the FTSE100 return series, as highlighted above, suggest the use of a model in which

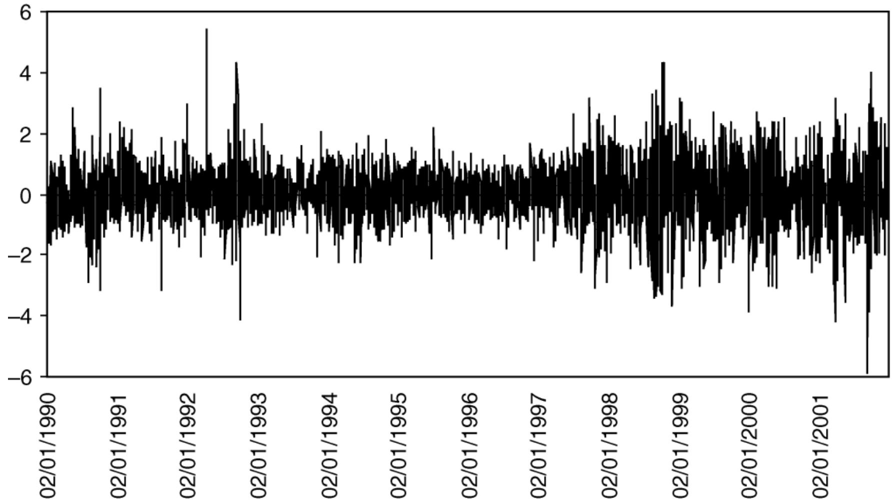


**Figure 1.** FTSE 100 % daily returns series: 01/01/1990 to 31/12/2001

the volatility can change over time. Three simple models, a GARCH(1,1), an EGARCH(1,1) and a log-normal AR(1) SV, will be estimated and compared in the next two sections. GARCH(1,1) is one of the most popular GARCH models used in applications. Hansen and Lunde (2001), analysing DM–$ exchange rate, find that, despite its simplicity, GARCH(1,1) is not outperformed by other more complex ARCH/GARCH-type models. Hence this simple model is considered as a benchmark. The EGARCH is also very popular (see Granger and Poon, 2003 for a review of the empirical literature on volatility forecasting).

The first 11 years have been taken as the estimation sample, while the last year of data has been used as out of sample period for volatility forecasting and VaR estimation. While the comparison of the out-of-sample forecasting ability, presented in a later section, is based on "VaR" estimation

**Table 1.** FTSE100 daily per cent returns basic statistics in full sample

| | |
|---|---|
| Mean | 0.0253 |
| Std. deviation | 0.9849 |
| Excess kurtosis | 2.0626 |
| Skewness | −0.0767 |
| Minimum | −5.8853 |
| Maximum | 5.4396 |
| ADF (5) | −24.36 |
| JB | 541 (0.000) |
| ARCH (5) | 66 (0.000) |
| LB (30) $r^2$ | 1531 (0.000) |

ADF (5) is the Augmented Dickey–Fuller test for unit root with five lags. JB is the Jarque–Bera test for normality. ARCH(5) is the test for ARCH effects with five lags. LB (30) $r^2$ is the Ljung–Box test for serial correlation with 30 lags, applied to the squared returns.

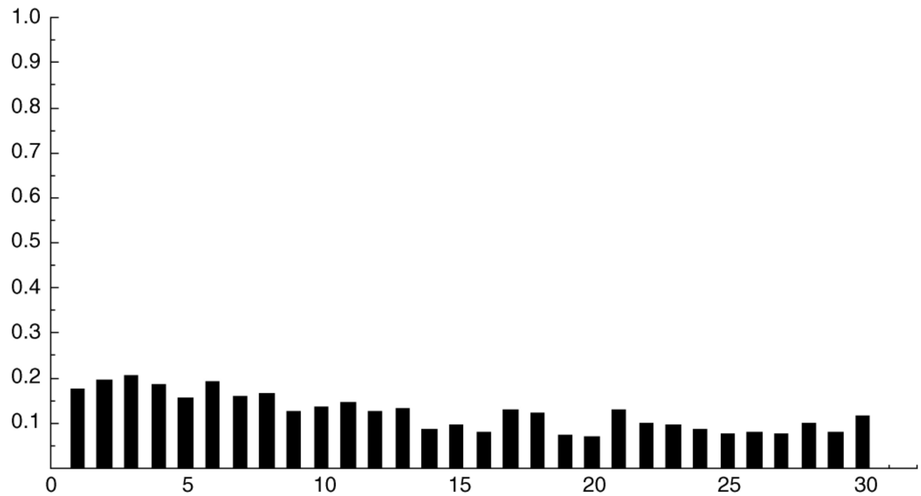


**Figure 2.** FTSE100 squared returns correlogram

and backtesting, in the following, an in-sample comparison of the models is performed by means of some standard diagnostic tests. Since the diagnostic tests are based on the estimation sample, *i.e.* 01/01/1990 to 31/12/2000, Table 2 presents the features of the returns on this period and some relevant statistics.

The ARCH tests in Table 2 reject the null hypothesis highlighting the presence of ARCH effects in the returns. The LB test applied to the squared returns confirms the presence of non-linear dependence in the data. The LB is also reported for returns: the value is much lower as expected since the dependence is mainly non-linear.

According to the model considered in this paper, the standardized returns $y_t = r_t/\hat{\sigma}_t$, defined by the estimated conditional volatility $\hat{\sigma}_t$, should be independent and distributed as standard normal: based on this observation, the standardized returns from the estimated models are tested. Table 6 at the end of this section present the results for all the models. In particular the excess

**Table 2.** FTSE100 daily per cent returns basic statistics in estimation sample

| | |
|---|---|
| Mean | 0.03398 |
| Std. deviation | 0.94161 |
| Excess kurtosis | 1.8128 |
| JB | 381 (0.000) |
| ARCH (1) | 48.2 (0.000) |
| ARCH (5) | 39.8 (0.000) |
| LB (30) $r$ | 70.1 (0.000) |
| LB (30) $r^2$ | 1254.5 (0.000) |

JB is the Jarque–Bera test for normality. ARCH(1) and ARCH(5) are the tests for ARCH effects with one and five lags. LB (30) $r$ is the Ljung–Box test for serial correlation with 30 lags, applied to the returns. LB (30) $r^2$ is the Ljung–Box test for serial correlation with 30 lags, applied to the squared returns.

kurtosis and the JB are computed on the standardized returns to highlight the ability of the models to capture fat tails. The LB tests are applied to the squared standardized returns to check their independence. In line with the literature, the LB is also applied to the standardized returns. The ARCH tests are computed to check for residual ARCH effects in the standardized returns. The models are also compared through the Bayesian Information Criterion (BIC) for model selection.

### *3.2 GARCH Estimation*

The GARCH(1,1) as in Equation 3 is estimated by maximum likelihood[4] and the results are presented in Table 3.

The sum of the two estimated coefficients $\alpha$ and $\beta$ (persistence coefficient) is very close to one, meaning that 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 the RiskMetrics model, which can be interpreted as an Integrated GARCH (IGARCH), *i.e.* the model obtained from Equation 3 by setting $\alpha + \beta = 1$. The conditional variance is estimated and forecast one-step-ahead through Equation 3b. The model succeeds in accounting for second order dependence since the squared standardized returns present no significant autocorrelation, consistently with the LB. Even if the kurtosis has been reduced by standardizing the returns, the normality hypothesis is still rejected (Table 6). This is not surprising, since most of the literature on ARCH-type models (e.g. Hsieh, 1991) conclude that GARCH can only partly account for fat tails in the distribution of the returns. The LB statistic proves the success of the GARCH(1,1) in capturing the non-linear dependence: the squared standardized returns are in fact independent. The ARCH tests show that there are no residual ARCH effects in the standardized returns.

The EGARCH(1,1) as in Equation 4 is estimated again by maximum likelihood and the estimation results are presented in Table 4.

The stationarity condition is satisfied since $\hat{\beta} < 1$. The asymmetry coefficient $\omega$ is significantly different from zero, hence supporting the existence of an asymmetric impact of returns on conditional variance and the consequent value of using an asymmetric model. The correlation in

**Table 3.** GARCH(1,1) parameters estimation

| Parameter | Estimate | Std. Error | $p$-value |
|---|---|---|---|
| $\gamma$ | 0.00838 | 0.00237 | 0.0004 |
| $\alpha$ | 0.05077 | 0.00609 | 0.0000 |
| $\beta$ | 0.94028 | 0.00729 | 0.0000 |
| $\alpha + \beta$ | 0.991 | | |

**Table 4.** EGARCH(1,1) parameters estimation

| Parameter | Estimate | Std. Error | $p$-value |
|---|---|---|---|
| $\gamma$ | −0.06669 | 0.009529 | 0.000 |
| $\beta$ | 0.99041 | 0.002185 | 0.000 |
| $\alpha$ | 0.08406 | 0.011952 | 0.000 |
| $\omega$ | −0.04257 | 0.006529 | 0.000 |

the squared standardized returns is definitely reduced and the LB supports their independence, analogously to the case of GARCH(1,1).

The diagnostic tests in Table 6 suggest that the EGARCH(1,1) performs better than the standard GARCH(1,1) in fitting the data: in particular the excess kurtosis is reduced more significantly (even if the normality hypothesis is still far from being accepted according to the JB) and the goodness of fit is better according to the BIC. It is hence worthwhile to include the EGARCH model in the out-of-sample performance comparison.

### *3.3 SV Estimation*

The SV model is estimated following the formulation in CKS98: the log-variance equation is slightly different from the model presented in Equation 5:

$$\begin{aligned} &(15.A)\ h_t = \mu + \phi(h_{t-1} - \mu) + \sigma_\eta \eta_t \\ &(15.B)\ h_1 \sim N\left(\mu, \frac{\sigma_\eta^2}{1-\phi^2}\right) \end{aligned} \tag{15}$$

Equation 15b shows the initialization of the log-variance. The two error terms $\varepsilon_t$, $\eta_t$ are assumed to be uncorrelated. The estimation is performed by an MCMC procedure as presented in CKS98.[5]

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. A Gibbs sampler as described in CKS98 is used for the estimation of the parameters. 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 results of the estimation are summed up in Table 5.

The parameter $\mu$ is expressed as $\beta = \exp(\mu/2)$ in CKS98 because this has an economic interpretation as the modal instantaneous volatility. Even if the log-variance is stationary, the coefficient of persistence $\phi$ 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 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 the choice of 50,000 iterations and of a simple algorithm is 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.[6] 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 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

**Table 5.** SV parameters estimation

| Parameter | Estimate | MC Std. Error | Inefficiency | Covariance | | |
|---|---|---|---|---|---|---|
| $\phi \mid r$ | 0.98803 | 0.000237 | 181 | 0.000015 | *−0.6258* | *0.0591* |
| $\sigma_\eta \mid r$ | 0.09724 | 0.001116 | 429 | −0.000028 | 0.000138 | *−0.0370* |
| $\beta = \exp(\mu/2) \mid r$ | 0.85580 | 0.001306 | 15 | 0.000017 | −0.000032 | 0.005399 |

independent draws. The inefficiency factor for the constant scaling factor $\beta = \exp(\mu/2)$ is much lower than for the other two parameters, following the results normally found in other empirical work (e.g. CKS98).[7] The last three columns of Table 5 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 3 gives a graphical illustration of the simulation results.

The upper graphs represent the full sample of iterations for each parameter, which give the simulated marginal densities $f(\phi|r)$, $f(\sigma|r)$, $f(\beta|r)$ respectively in (a), (b) and (c). The three graphs in the middle (d), (e), (f) show the histograms of the simulated marginal densities: the histogram corresponding to $\beta$ (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 $\beta$. The degree of serial correlation in the draws is an indicator of how well the simulation algorithm behaves. While for $\beta$ the autocorrelation is almost inexistent, there is significant autocorrelation for the other two parameters: in particular $\sigma$ presents (h) important serial correlation. This result is just a graphical confirmation of the information already contained in the inefficiency factors. An improved algorithm proposed by CKS98, an offset mixture with appropriately adapted Gibbs sampler (henceforth referred to as "SV mixture"), is also implemented for comparison: the resulting parameter estimates are not very different (0.98812, 0.096085, 0.85378), while there is a significant improvement in the inefficiency factors (103, 211, 1.4) and in the standard errors (0.000187, 0.000927, 0.000392). The algorithm is more complex and hence slower.

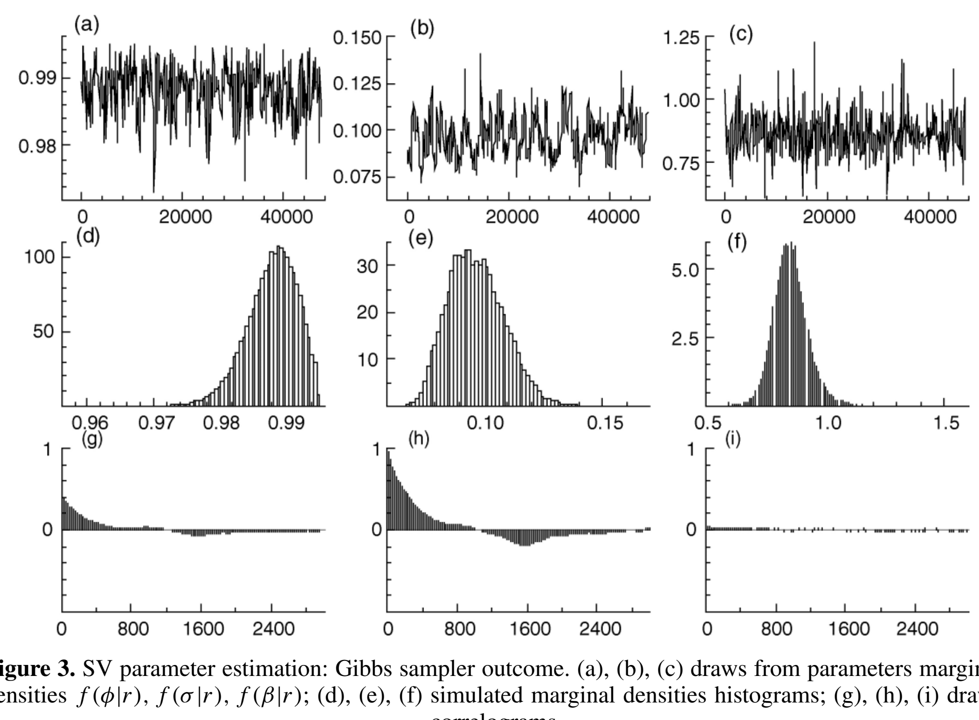


**Figure 3.** SV parameter estimation: Gibbs sampler outcome. (a), (b), (c) draws from parameters marginal densities $f(\phi|r)$, $f(\sigma|r)$, $f(\beta|r)$; (d), (e), (f) simulated marginal densities histograms; (g), (h), (i) draws correlograms

Once the parameters have been estimated, the filtering and forecasting procedure can be applied. The log-variance is initialized according to Equation 15b and a particle filter algorithm as described earlier is applied to sample draws from $h_t|R_t, \theta$ and $h_{t+1}|R_t, \theta$. The number of simulations $M$ is fixed at 2000. The diagnostic tests in Fig. 4 are based on the variable $u^M_{t+1}$defined in Equation 14.

The comparison between graph (a) and (c) in Fig. 4 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. Graph (b) illustrates the series of the transformed data, *i.e.* the series $v^M_{t+1}$ for $t = 1, \ldots, n-1$. 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 CKS98. The normality hypothesis on $v$ is accepted at the 1% level of significance by JB (6.56). The SV estimated with the mixture sample algorithm performs even better, with JB (5.99) accepting normality at the 5%.

In order to compare GARCH and SV performances in sample, the features of the standardized returns $y_t$ are analysed in the SV case as well, by taking the mean of the forecast volatility simulations as an estimate of $\sigma_t$. This is a useful approximation which allows a direct comparison with GARCH and the use of the volatility forecast output for VaR calculation. However this approximation could lead to underestimation of the performances of SV. Analogously to the GARCH case, the diagnostic tests presented in Table 6 show the ability of the model in capturing non-linear dependence structure of the returns. The SV model, estimated with both algorithms, produces standardized returns with excess kurtosis lower than GARCH(1,1) but higher than EGARCH(1,1) and the JB statistics give the same result. Both LB and ARCH tests prove the ability of the model

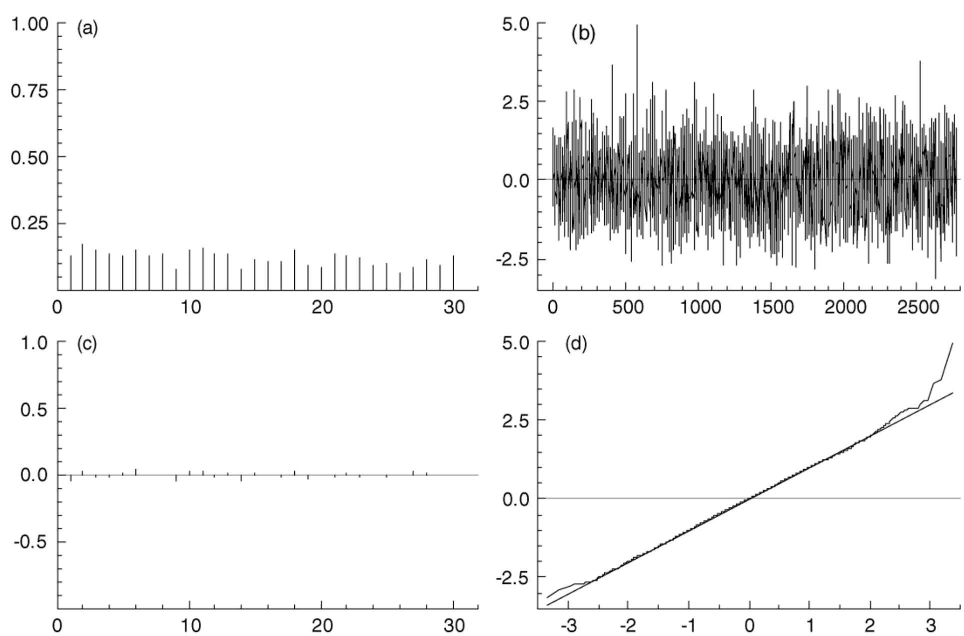


**Figure 4.** SV diagnostic tests. (a), (c) Correlogram of squared returns and of the transformed data $v$; (b) series of transformed data $v$; (c) QQ plot of transformed data $v$

**Table 6.** In-sample comparison based on standardized returns

| | GARCH(1,1) | EGARCH(1,1) | SV | SV mixture |
|---|---|---|---|---|
| Excess kurtosis | 1.096 | 1.033 | 1.071 | 1.081 |
| JB | 141 (0.00) | 125 (0.00) | 134 (0.00) | 137 |
| ARCH (1) | 0.469 (0.49) | 0.211 (0.65) | 0.738 (0.39) | 0.606 (0.44) |
| ARCH (5) | 0.658 (0.65) | 0.827 (0.53) | 0.707 (0.62) | 0.665 (0.65) |
| LB (30) $y$ | 46.23 (0.03) | 46.26 (0.03) | 46.27 (0.03) | 46.81 (0.03) |
| LB (30)$y^2$ | 21.27 (0.88) | 23.90 (0.78) | 22.42 (0.83) | 21.85 (0.86) |
| Log-Likelihood | −3584 | −3567 | −3567 | −3567 |
| BIC | 7192 | 7166 | 7158 | 7158 |

JB is the Jarque–Bera test for normality. ARCH(1) and ARCH(5) are the tests for ARCH effects with one and five lags. LB (30) $y$ is the Ljung–Box test for serial correlation with 30 lags, applied to the standardized returns. LB (30) $y^2$ is the Ljung–Box test for serial correlation with 30 lags, applied to the squared standardised returns. BIC is the Schwartz Bayesian information criterion.

in capturing non-linear dependence: as in the case of GARCH, the squared standardized returns are not autocorrelated and there are no residual ARCH effects. According to the BIC, the SV is preferred to the other models: the Log-Likelihood is the same as in EGARCH(1,1), but the latter requires the estimation of an additional parameter. Hence, from the in-sample comparison it can be concluded that EGARCH(1,1) and SV perform in a similar way and that both perform better than GARCH(1,1). The differences among the models performances are anyway quite slight.

## 4. 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 $\alpha$ 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

$$\Pr(\Delta V_{t+1} \leq -VaR_{t+1} | I_t) = \alpha\% \tag{16}$$

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. 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.

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.

The returns are assumed to be conditionally normal:

$$r_{t+1} | R_t \sim N(0; \sigma^2_{t+1}) \tag{17}$$

By adopting the common approximation $\Delta P_{t+1} \cong P_t r_{t+1}$, VaR can be easily calculated following the standard variance-covariance approach as

$$VaR_{t+1} \cong -\Phi^{-1}(\alpha) \times P_t \times \sigma_{t+1} \tag{18}$$

In this work a confidence level of 95% is used, *i.e.* $\Phi^{-1}(\alpha) = -1.65$, and the one-step-ahead forecast volatility is estimated alternatively through GARCH and SV models. 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

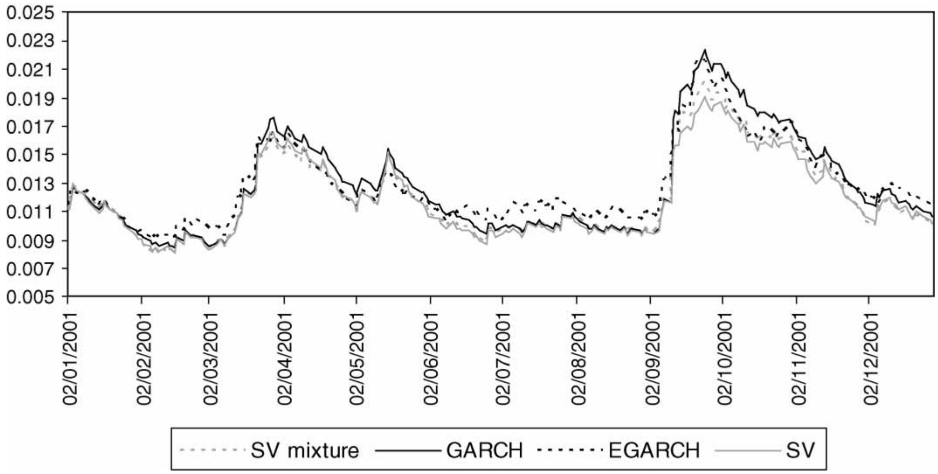


**Figure 5.** Volatility forecast. 'SV mixture' indicates the SV model estimated with the offset mixture algorithm



one-step-ahead. Figure 5 shows the forecast volatility coming from the estimated models: as for SV, both the volatilities coming from the two estimation algorithms are presented. The GARCH volatility is generally higher than the SV volatility, particularly during the period of high volatility following 11 September. An explanation could be related to the persistence coefficient which is higher in GARCH than in SV, that is a typical result as pointed out by Shephard (1996).

VaR is estimated day by day on the last year of data 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 different 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 6 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 11.1%, 10.3%, 13.4% and 13% for GARCH(1,1), EGARCH(1,1), SV and SV mixture respectively. All the models tend to underestimate the extreme price changes, since they all present an actual proportion of misses greater then 10%. The EGARCH however exceeds only slightly the target percentage. It has to be noted that the forecast period includes 11 September 2001: the absolute return is obviously very high on that day producing a miss.[8] 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 models can be modified to account for a greater proportion of leptokurtosis by changing the distribution of $\varepsilon_t$ from a normal to 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 all over because the VaR estimation, according to the

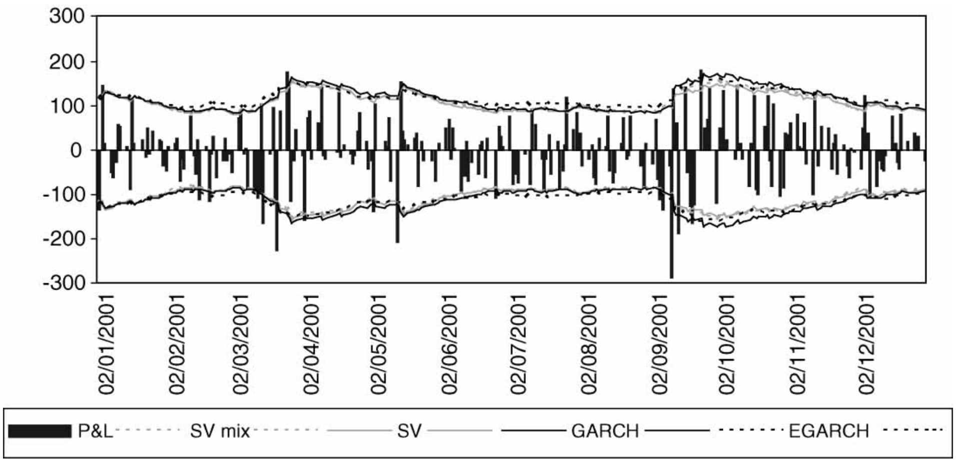


**Figure 6.** VaR and backtest on one-unit FTSE100 portfolio. 'SV mix' indicates the SV model estimated with the offset mixture algorithm. P&L indicates the profits and losses computed for backtest



RiskMetrics methodology, relies on the conditional normality hypothesis, so that changing this hypothesis would change the VaR computation.

Following Christofersen (1998), the goodness of an interval forecast can be formally evaluated on the basis of two main requirements which need to be satisfied:

- Correct coverage: the observed proportion of misses is approximately equal to $\alpha\%$, assuming independence.
- Independence: hits and misses are independently distributed.

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

$$I_t = \begin{cases} 1 & \textit{if } \Delta P_{t+1} \in [-VaR_{t+1}; VaR_{t+1}] \quad (\textit{hit}) \\ 0 & \textit{if } \Delta P_{t+1} \notin [-VaR_{t+1}; VaR_{t+1}] \quad (\textit{miss}) \end{cases} \tag{19}$$

Based on the proportion of misses and hits compared to the predefined probability of 90%, two Likelihood Ratio (LR) tests can be performed, with the null hypothesis of correct coverage (LRcc) and independence (LRind) respectively. Moreover, a joint test (LRjoint) can be built by summing these two statistics: the null is of conditional correct coverage, *i.e.* both correct coverage and independence. The results of the tests are presented in Table 7.

As expected the value of the correct coverage test is much lower in the case of GARCH, and particularly EGARCH(1,1), even if the null of correct coverage is accepted in all cases. Hence GARCH models perform better according to the correct coverage criterion, and a similar conclusion is reached according to the independence criterion: the null of independence is accepted in all cases at the 1% level of significance while it is accepted at the 5% only for the GARCH models. The same holds for the joint test of correct coverage and independence. However EGARCH(1,1) clearly outperforms GARCH(1,1) within the GARCH type models.

From this first comparison the SV comes out defeated. However, there are several points to be made, which lead to a more uncertain conclusion. First, 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

**Table 7.** Likelihood ratio tests on the VaR-defined interval forecast

| | GARCH(1,1) | EGARCH(1,1) | SV | SV mixture |
|---|---|---|---|---|
| LRcc | **0.311** | **0.021** | **3.035** | **2.400** |
| | (0.577) | (0.884) | (0.082) | (0.121) |
| LRind | **2.872** | **2.118** | **4.850** | **5.628** |
| | (0.090) | (0.146) | (0.028) | (0.018) |
| LRjoint | **3.183** | **2.139** | **7.884** | **8.028** |
| | (0.204) | (0.343) | (0.019) | (0.018) |

LRcc is the likelihood ratio test for correct coverage. LRind is the likelihood ratio test for independence. LRjoint is the likelihood ratio test for correct coverage and independence. For each test the number in bold is the $\chi^2$ and the number in parentheses the relative probability.

by applying a filtering algorithm is a vector of $M$ draws $[h^1_{t+1}, h^2_{t+1}, \ldots, h^M_{t+1}]$ 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 $\hat{\sigma}_{t+1}$, but the returns are normally distributed conditionally on the volatility distribution. It could be argued that the returns are normal conditionally on each single realisation from the simulation sample $[h^1_{t+1}, h^2_{t+1}, \ldots, h^M_{t+1}]$:

$$r_{t+1}|h^j_{t+1} \sim N(0; \exp(h^j_{t+1})) \quad \forall j \tag{20}$$

The predictive distribution given the information up to time $t$ could then be approximated through the filtering output as follows:

$$f(r_{t+1}|R_t) = \int f(r_{t+1}|h_{t+1}) f(h_{t+1}|R_t) dh_{t+1} \cong \frac{1}{M} \sum_{j=1}^{M} f_N(r_{t+1}|h^j_{t+1}) \tag{21}$$

Starting again from the definition in Equation 16:

$$\begin{aligned} \Pr(\Delta P_{t+1} < -VaR_{t+1}|R_t) &= \int \Pr(\Delta P_{t+1} < -VaR_{t+1}|h_{t+1}) f(h_{t+1}|R_t) dh_{t+1} \\ &\cong \frac{1}{M} \sum_{j=1}^{M} \Phi\left(\frac{-VaR_{t+1}}{P_t \exp(h^j_{t+1}/2)}\right) \end{aligned} \tag{22}$$

The right-hand side of Equation 22 is a proxy deriving from the application of a particle filter. The conditional distribution $f(h_{t+1}|R_t)$ is in fact approximated by the simulated sample $[h^1_{t+1}, h^2_{t+1}, \ldots, h^M_{t+1}]$ and the probability of each value is $1/M$. By imposing the last term in Equation 22 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 is solved numerically using the Excel solver. Consistently with the *ex ante* expectations, the interval is larger than the one defined by using the average SV volatility. However, the increase in the wideness of the interval is not sufficiently relevant to improve substantially the performances of the SV model in estimating VaR. Second, while the LR test is determined by the number of misses, regardless of the magnitude of the excess loss, another way of comparing the forecasting performances related to VaR is through a loss function. Following Gonzalez-Rivera *et 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:

$$V = \frac{1}{n}\sum_{t=1}^{n} |\Delta P_{t+1} + VaR_{t+1}| \times \left[(1-\alpha)1_{\{\Delta P_{t+1} < -VaR_{t+1}\}} + \alpha 1_{\{\Delta P_{t+1} > -VaR_{t+1}\}}\right] \tag{23}$$

Table 8 shows a partial inversion of the preferences according to this criterion: while the EGARCH(1,1) keeps the best performance, GARCH(1,1) and SV perform very similarly, with a preference for SV if estimated through the more complex offset mixture algorithm (SV mixture).

Third, the results in Table 8 are quite weak since they do not consider the magnitude of the excess loss: even very small excess losses change the results significantly. In particular GARCH(1,1) performance is sensible to the VaR formulation, while the SV mixture results are more robust: the exact[9] VaR has been used for the same purpose and the same tests are shown in Table 9.

The EGARCH(1,1) is still the preferred one, while the comparison between GARCH(1,1) and SV becomes less straightforward. In particular, if GARCH(1,1) and SV estimated by the offset mixture algorithm are compared, the GARCH(1,1) is still preferred according to the correct coverage criterion, while SV mixture performs better in terms of independence. This means that in the case of SV mixture, even if the interval forecast is not wide enough to contain the correct proportion of realizations, the misses cluster less than in the case of GARCH(1,1).

The conclusion that can be drawn from the entire set of tests performed is that there is no relevant difference between GARCH(1,1) and SV performances. Since the SV is computationally more demanding, particularly if estimated through the more successful but more complex offset mixture algorithm, it could be argued that it is not worthwhile to adopt an SV model for VaR purpose. This conclusion is even stronger if the EGARCH(1,1) model is considered, which definitely performs better than the other models.

**Table 8.** VaR-based loss function values

| | GARCH(1,1) | EGARCH(1,1) | SV | SV mixture |
|---|---|---|---|---|
| Loss V | 8.335 | 8.128 | 8.355 | 8.309 |

**Table 9.** LR tests on the exact VaR-defined interval forecast

| | GARCH(1,1) | EGARCH(1,1) | SV | SV mixture |
|---|---|---|---|---|
| LRcc | **0.577** | **0.004** | **4.506** | **2.400** |
| | (0.448) | (0.950) | (0.034) | (0.121) |
| LRind | **6.626** | **2.613** | **7.637** | **5.628** |
| | (0.010) | (0.106) | (0.006) | (0.018) |
| LRjoint | **7.203** | **2.617** | **12.143** | **8.028** |
| | (0.027) | (0.270) | (0.002) | (0.018) |

LRcc is the likelihood ratio test for correct coverage. LRind is the likelihood ratio test for independence. LRjoint is the likelihood ratio test for correct coverage and independence. For each test the number in bold is the $\chi^2$ and the number in parentheses the relative probability.

## 5. Conclusions

Two types of volatility models are discussed and estimated: GARCH and stochastic volatility. Since the SV model presents some theoretical appeal but its estimation is definitely more computationally demanding, this paper compares the GARCH and SV performances within a simple VaR calculation in order to evaluate if the use of SV models in contrast to GARCH is worthwhile. In particular a GARCH(1,1), an EGARCH(1,1) and a log-normal AR(1) SV models are estimated on the FTSE100 index returns.

In the preliminary in-sample comparison, the GARCH(1,1) results outperformed by both EGARCH(1,1) and SV: the main objective of the paper is anyway the out-of-sample comparison based on VaR. The analysis based on VaR does not presents evidence in favour of SV compared to GARCH: on the contrary GARCH models look quite useful in defining an interval forecast, even if all the models tend to underestimate losses, probably due to the common underlying assumption of conditional normality. The EGARCH(1,1) has the best performance in terms of VaR, while a clear preference does not emerge between GARCH(1,1) and SV. It has to be noted that the performance of SV is also related to the algorithm used in the estimation, as it is evident from the Likelihood Ratio tests and the loss function values for SV estimated by a simple Gibbs sampler and an offset mixture algorithm. While the GARCH(1,1) performs in general better than SV estimated by the simpler algorithm, the SV estimated by the offset mixture algorithm performs better in terms of loss function. Moreover, a slight change in the VaR definition makes SV preferable to GARCH also according to one of the two LR tests. This result suggests that the performances of the SV model could be significantly improved by using more sophisticated estimation algorithms. However, this would make the SV estimation even more computationally demanding compared to the simpler GARCH models. The fact that the asymmetric model EGARCH(1,1) has the best performance suggests also a comparison with an asymmetric SV model, which is again more computationally demanding. Some caution should be exercised in reading these conclusions, since they could be related to the particular data set used and since the emerging classification is also related to the techniques adopted to evaluate the performances.

## Acknowledgement

This paper is a revised version of one presented at the XXXII EWGFM in London (April 2003). I wish to thank the conference participants and the two anonymous referees for their comments and suggestions. I am also grateful to Michael Pitt and Ilias Tsiakas for helpful advice. Financial support from MIUR is gratefully acknowledged. The usual caveats apply.

## Notes

[1] The process of computing the log-variance conditional on the observations up to the current time $h_t|R_t$, for $t = 1, \ldots, T$, is called "filtering". In general, given a time series modelled as independent conditionally on an unobserved state, "filtering" means "to learn about the state given contemporaneously available information" (quote from Pitt and Shephard, 1999).

[2] The FTSE100 index prices have been downloaded from Datastream; the bank holidays have been taken out from the series.

[3] Here and in the following most of the correlograms are omitted: these graphs are available on request.

[4] The GARCH models (both GARCH(1,1) and EGARCH(1,1)) are estimated with the software E-Views.

[5] The SV model is estimated through an Ox code related to CKS98. The estimation and filtering algorithms used for CKS98 are available at http://www.nuff.ox.ac.uk/users/shephard/ox/. The appropriate algorithms have been exploited by adapting the code to the context of this paper.

[6] For technical details on the inefficiency factors and the standard errors see CKS98.

[7] In CKS98 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 CKS98. A more complex algorithm is also estimated for comparison.

[8] If the highest absolute returns (e.g. higher than 4%) are considered as outliers and taken out of the series, the effects on the backtest is double: on one hand the misses corresponding to the outliers are clearly avoided. On the other hand, however, the volatility forecast, which is strongly affected by past returns, does not increase as a consequence of the high absolute returns, leading to new misses in the period following outliers, as in the case of 11 September. In the present case, the backtest after eliminating outliers present at least the same number of misses as before.

[9] While the common VaR definition in Equation 18 is based on an approximation, the exact lower and upper limits for the interval forecast are respectively $-[1 - \exp(-1.65\sigma_{t+1})]P_t$ and $[\exp(1.65\sigma_{t+1}) - 1]P_t$.



## References

Baillie, R. T. and Bollerslev, T. (1989) The message in daily exchange rates: a conditional variance tale, *Journal of Business and Economic Statistics*, 7, pp. 297–305.

Bluhm, H. H. W. and 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*, 31, pp. 307–327.

Chib, S., Kim, S. and Shephard, N. (1998) Stochastic volatility: likelihood inference and comparison with ARCH models, *Review of Economic Studies*, 65, pp. 361–393.

Christofersen, P. F. (1998) Evaluating interval forecasts, *International Economic Review*, 39, pp. 842–862.

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

Gonzales-Rivera, G., Lee, T. H. and Mishra, S. (2002) Forecasting volatility: a reality check based on option pricing, utility function, value at risk, and predictive likelihood. University of California, Riverside, Working Paper available at www.gloriamundi.org.

Granger, C. W. J. and Poon, S. H. (2003) Forecasting volatility in financial markets: a review, *Journal of Economic Literature*, 41, pp. 478–539.

Hansen, P. R. and Lunde, A. (2001) A forecast comparison of volatility models: does anything beat a GARCH(1,1)? Brown University.

Hsieh, D. A. (1991) Chaos and non-linear dynamics: application to financial markets, *The Journal of Finance*, 46, pp. 1839–1877.

Jaquier, E., Polson, N. G. and Rossi, P. E. (1994) Bayesian analysis of stochastic volatility models (with discussion), *Journal of Business and Economic Statistics*, 12, pp. 371–417.

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

Nelson, D. B. (1991) Conditional heteroscedasticity in asset returns: a new approach, *Econometrica*, 59, pp. 347–370.

Pitt, M. K. and Shephard, N. (1999) Filtering via simulation: auxiliary particle filter, *Journal of the American Statistical Association*, 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*, pp. 1–67 (London: Chapman & Hall).

Taylor, S. J. (1994) Modelling stochastic volatility, *Mathematical Finance*, 4, pp. 183–204.

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