Value-at-Risk Performance of Stochastic and ARCH Type Volatility Models: New Evidence

Binh Do *

March 20, 2007

Abstract

This paper evaluates the effectiveness of selected volatility models in forecasting Value-at-Risk (VaR) for 1-day and 10-day horizons. The latter is the actual reporting horizon required by the Basel Committee on Banking Supervision, but not considered in existing studies. The autoregressive stochastic volatility (Taylor, 1982) is found to be less effective than simpler ARCH type models such as the RiskMetrics and GARCH models. 10-day VaR forecasting is shown to be a difficult task by construction. Various schemes to construct this forecast are proposed and evaluated. In particular, a scheme that uses return time series of 10-day intervals transforms the 10-day VaR forecasting into 1-step-ahead forecasting and shows considerable improvement in accuracy although the result might be due to inclusion of the 1987 crash in training the models.

*Department of Accounting and Finance, Monash University, Australia

1 Introduction

Value-at-Risk is widely regarded as the standard measure of market risk, and recognised by the Basel Committee on Banking Supervision in its guideline document on capital measurement and capital standards (BIS, 2004). In this guideline, institutions have the choice to adopt their own internal models to measure and report their daily VaR statistics that reflect their downside risk position. This development has met with a growing literature that backtests VaR forecasting methods, with the majority focusing on testing alternative models of volatility. These include Giot and Laurent (2003, 2004), Eberlein, Kallsen and Kristen (2003), Huang and Lin (2004), Brooks and Persand (2004, and Raggi and Bordinon (2006). As typical in evaluative studies, mixed results are reported with little consensus achieved with respect to the "most adequate" model for VaR forecasts. The fact that these studies evaluate different pools of models makes it difficult to draw meaningful conclusions in this area. For example, Huang and Lin (2004) and Raggi and Bordignon (2006) find evidence supportive of the asymmetric APARCH model (Ding, Granger and Engle, 1993). Brooks and Persand (2004) conclude a univariate GARCH(1,1) is preferred over multivariate GARCH and EGARCH formulations. Berkowitz and O'Brien (2004) also find GARCH to be superior to internal models used by leading US banks. Giot and Laurent (2004) show that an adequate ARCH model is as good as a model based on realized volatility. As a rare consensus, most studies conclude fat tail distributions such as the t-distribution are the choice for modeling innovations in the return process.

This paper seeks to contribute to this VaR backtesting literature by extending in two directions. The principal extension is to evaluate the effectiveness of a select of volatility models in forecasting 10-day VaR, in order to coincide with the reporting requirement actually stipulated in the Basel framework. Almost all existing studies investigate 1-

day VaR, with the exception of Brooks and Persand (2003), which do not consider SV models. Clearly, based on daily time series, the 1-day VaR computation is based on the one-step-ahead volatility forecast, which is by ARCH construction, deterministic, hence its celebrated success. In contrast, forecasting VaRs for 10-day horizons is much more challenging and likely to encounter two problems: reduced accuracy of VaR estimates in terms of the number of violations, and the serial dependency in violations. These problems are known as unconditional coverage and independence, respectively, in the backtesting literature (c.f eg Christoffersen, 1998). Indeed, by construction, tomorrow's 10-day forecast cannot improve upon today's violation which is not observed until 10 days later. The consequence is that violations are particularly concentrated during periods of prolonged market turbulence, resulting in serial violation as a by-product. When there is such a "mismatch" between the frequency of information update and the forecast horizon, solutions tend to be suboptimal and adhoc. This study empirically tests various schemes to compute the 10-day VaR and suggests one that helps alleviate some of these problems.

The second theme of this paper is to revisit the debate between GARCH models and stochastic volatility models. Although explaining similar stylized facts in the financial market, these two classes of models have different and non-equivalent formulations, making them an attractive subject for empirical studies. Whilst GARCH models (introduced by Engle, 1982 and generalised by Bollerslev, 1986) formulate volatility as a deterministic function of past information, hence observable, stochastic volatility (SV) models, on the other hand, treat volatility as random through time and unobservable. One particular SV model that has been attracting considerable interest is the discrete time, autoregressive model (Taylor, 1982). Although this model is found to better describe historical data than many GARCH models (see Jacquier, Polson and Rossi, 1994 and Shephard, 1996), its econometrics is much more complex than GARCH models. As will be reviewed in the body of this paper, estimators of this model are necessarily approximate and simulation based. Furthermore, the model is less suitable for pricing purposes mainly because of its discrete time setup. A natural question is therefore, whether this complicated model is worth considering in the context of VaR forecasting. Perhaps due to the model's economet-ric complexity, its empirical studies are few, amongst which Raggi and Bordignon (2006) find it to be effective whereas Eberlein et al (2003) find evidence against this model. This paper revisits the comparison between this stochastic volatility model and a select of rep-resentative GARCH models in both 1-day VaR and 10-day VaR forecasting. Furthermore, by including a simple yet commercially popular GARCH model, RiskMetrics (RiskMet-rics Group, 2001) in the analysis, this paper seeks to show whether complicated modeling yields decisive advantage over simpler alternatives in this risk management area.

The remainder of this paper is organised as follows. Section 2 describes four volatility models that are used to compute VaR. Section 3 explains implementation issues in con-structing VaR forecasts. Section 4 discuss estimation techniques, with a special focus on the SV model. Section 5 presents the empirical analysis. Section 6 concludes.

2 Alternative Volatility Models

It has been long documented that daily returns in financial markets exhibit three stylized facts (Taylor, 2005, Chapter 4). First, there is no correlation between returns for different days (*unpredictability*). Second, the return distribution is not normal, with the presence of more extreme occurences than suggested by the normal distribution (*fat tails*). Third, the correlations between the magnitudes of returns on nearby days are positive and significant, a phenomenon known as *volatility clustering*.

These stylized facts can be generally described by the following specification:

$$y_t^* = \mu_t + \sigma_t z_t \tag{1}$$

where y_t^* is day t's return, μ_t is its conditional mean, σ_t is the conditional standard deviation and z_t is an i.i.d noise of mean zero and variance of 1.[1] Since μ_t is approximately zero for daily returns, regardless of conditional specification, one often lets it equal the sample mean, as assumed throughout this paper, and work instead with the excess return process y_t[2] :

$$y_t = \sigma_t z_t \tag{2}$$

Volatility models pertain to parametric formulations of σ_t as a function of information up to time t. By expressing volatility as a function of its past, coupled with the iid noise term, one obtains a data generating process that is serially uncorrelated, but is correlated at its squared level. In addition, extreme values are possible when the volatility is high. One popular formulation is GARCH(1,1) (Bollerslev, 1986):

$$\sigma_t^2 = a_0 + a_1 y_{t-1}^2 + b_1 \sigma_{t-1}^2 \tag{3}$$

This model has a very simple intuition: the variance of today's return is the weighted average of three components, a long run permanent variance, yesterday's forecast of variance and new information that was not incorporated in yesterday's forecast. Information older than yesterday's can be included by adding more lag terms to (2), which becomes GARCH(p,q). Empirical research shows that GARCH models with more lag terms than GARCH(1,1) do not gain significant incremental benefit, which is intuitive given information is quickly factored in securities prices, in line with efficiency market (Taylor, 2005).

[1] Equation (1) can be expressed in continuous time as $dy_t^* = \mu_t dt + \sigma_t dB_t$ with B denoting the standard Brownian motion

[2] More elaborate specifications of μ_t can be found in Taylor (1994) and the references therein.

ARCH models are those where $b_i = 0$ for all i. It is straightforward to verify that when σ_t is GARCH(1,1), the stochastic process $\{y_t\}$ is uncorrelated, $\{y_t^2\}$ is correlated, the excess kurtosis of y_t is positive, hence satisfying the three stylized facts mentioned above.

RiskMetrics model (RiskMetrics Group, 2001), on the other hand, lets the conditional variance forecast be the exponentially weighted moving average of past squared returns, so as to attach increasing importance to more recent information:

$$\sigma_t^2 = \frac{\sum_{\tau=0}^{\infty} \lambda^\tau y_{t-\tau}^2}{\sum_{\tau=0}^{\infty} \lambda^\tau} = (1-\lambda) y_{t-1}^2 + \lambda \sigma_{t-1}^2 \tag{4}$$

RiskMetrics is a version of Integrated GARCH(1,1), or IGARCH(1,1) investigated in Engle and Bollerslev (1986) with the constant term $a_0 = 0$. With only 1 parameter (instead of 3 for GARCH(1,1)), whose value is often preset within the range of 0.94 to 0.96 obtained from empirical backtesting, RiskMetrics commands certain popularity in commercial applications.

Extensions to simple GARCH models seek to capture other stylized facts such as asymmetry or leverage, a phenomenon that past negative shocks tend to have deeper impact on current conditional volatility than past positive shocks (Black, 1976, French, Schwert and Stambaugh, 1987). One such extension that encompasses many GARCH models is the Asymmetric power ARCH, or APARCH, introduced in Ding, Granger and Engle (1993):

$$\sigma_t^\delta = a_0 + a_1(|y_{t-1}| - \gamma y_{t-1})^\delta + b_1 \sigma_{t-1}^\delta \tag{5}$$

It is clear that with a positive γ, a negative excess return in the previous day will increase today's conditional volatility, the extent of which depends on the power factor δ to be determined from the data. This model is widely investigated in VaR related studies (Giot and Laurent, 2002, 2003, Raggi and Bordignon, 2006) where it is found to be effective (in one-step-ahead forecasting).

In all these GARCH models, the conditional volatility is a deterministic function of past information. SV models are a completely different setup, where the volatility is a function of its past plus a separate random noise. A well studied SV model that first appears in Taylor (1982) and Tauchen and Pitts (1983) proposes the logarithm of volatility follow an AR(1) process:

$$x_t = \alpha + \phi x_{t-1} + \sigma_v \eta_t \tag{6}$$

where $x_t = log\sigma_t^2$ and $\eta_t \sim \mathcal{N}(0,1)$ and the two noise terms η and z in (1) are independent. Thinking of (6) in terms of a discretised Vasicek model, it can be seen that the process mean reverts to $\frac{\alpha}{1-\phi}$ at a speed of $-log\phi$, and that volatility follows a lognormal distribution. It can also be verified that the resulting process $\{y_t\}$ in Equation (2.2) meets the three stylized facts discussed above. Whilst (6) is similar to (3), it is not equivalent. The (log transform of) volatility forecast in (6) not only depends on a constant long term level and its immediate past level, but also a random element representing a new shock to volatility over and above past and current information, such as news flows or trading volume. Model (6) is well studied in the econometric literature (cf. Melino and Turnbull (1990), Jacquier, Polson and Rossi (1994), Harvey et al (1994), Andersen and Sorensen (1996), Kim, Shephard and Chib (1998), Meyer and Yu (2000), Chib, Nardari and Shephard (2002), Jacquier et al (2004), Yu (2004) and Omori, Chib and Shephard (2006)). However, empirical financial applications have also emerged, such as Eberlein et al (2003) and Raggi and Bordignon (2006).

This study evaluates the out-of-sample effectiveness of four models (3), (4), (5), and (6) in estimating regulatory VaR. These models are representative of existing volatility literature: RiskMetrics is the most commercially applied model, GARCH(1,1) is the "workhorse" of volatility literature, APARCH(1,1) seems to be the most adequate VaR model, and the standard SV model is the most studied SV specification. As VaR applications focus on tails behaviour, this study considers both normal and Student-t distributions for the noise z_t in (1), hence in effect, 8 models in total are examined. [3]

3 VaR Implementation

VaR is a single number that summarises a portfolio's potential loss over a given period. Mathematically, a portfolio's h-day VaR at time t, with $1-\alpha$) confidence level, is defined to be the α quantile of the portfolio's profit and loss (P&L) distribution over a h-day horizon:

$$VaR_t(\alpha, h) = -F_h^{-1}(\alpha|I_t) \tag{7}$$

where $F_h^{-1}(\alpha|I_t)$ represents the quantile function (or inverse of the P&L function $F_h(.|I_t)$ which varies over time as the information set I_t, changes. The negative sign ensures the resulting VaR is a positive number, as per market convention. A 99% 1 day VaR of $30,000 for a $1,000,000 portfolio means that one can expect to experience a 1 day loss exceeding $30,000 once out of every 100 days. A 95% 10 day VaR of $50,000 for a $1,000,000 portfolio means that one can expect a 10 day cumulative loss exceeding $50,000 five times out of every 100 intervals of 10 days, or five times every four years.

When the daily P&L on the 1 dollar portfolio is assumed to be generated from (1) and any of the GARCH type models, its conditional distribution $F_h(.|I_t)$ has the same form as that of the error z, except that the mean is μ and variance σ^2_{t+1}. As such, the

[3] This study does not consider models that include the leverage effect as the focus is on VaR forecasts for long positions only. Similarly, models that account for jumps in volatility are excluded which should not endanger the completeness of the study since Raggi and Bordignon (2006) find jump models are inferior to simpler GARCH/SV models in a VaR backtesting context.

1-day VaR is deterministically computed as follows:

$$VaR_t(\alpha, 1) = \mu + \sigma_{t+1} Z^{-1}(\alpha) \tag{8}$$

where Z^{-1} is quantile function of z which, in the case of normality, evaluates to -1.64 and -2.33 for $\alpha = 0.05$ and 0.01 respectively. However, when VaRs for an horizon h greater than 1 are desired, the total P&L is now $Y^*_{t+h} = \sum_{i=1}^{h} y^*_{t+1}$, and its distribution $F_h(.|I_t)$ is generally not available, as such the VaRs must be approximated in some way.[4] The existing literature is silent on how to compute multi-period VaRs.

The standard solution to this issue which is suggested in the Basel document, is to use a 1-day VaR scaled up by a factor of $\sqrt{h}$ on the assumption of constant volatility. A major issue with this approach lies in the implicit assumption that the 10-day period return has the same (conditional) distribution form as the the 1-day period return, the only difference being that the former's variance (and mean) are now scaled up. However, it can be easily shown that this is not the case. This point is referred to in Engle (2003) as the asymmetry in variance for multi-period returns: although each period has a symmetric distribution, the multi-period return distribution will be asymmetric. Whilst this approximation is crude, it has been used widely in the market.

Another way to approximate the multi-period variance is to assume the conditional distribution has the same form as z, with its mean and variance matching those of the true distribution of Y^*_{t+1}, ie ignoring higher moments. In other words,

$$VaR_t(\alpha, h) = h\mu + A_{t+h} Z^{-1}(\alpha) \tag{9}$$

[4]To see how complex this distribution can be, let us consider the case of GARCH(1,1) and $h = 2$. In this case $Y^*_{t+2} = 2\mu + \sqrt{a_0 + a_1(z^2_{t+1} + b_1)\sigma^2_{t+1}} z_{t+2}$, with z^2_{t+1} and z_{t+2} being independent standard normal variables. Whilst Y^*_{t+2} has known moments, its distribution function is difficult to get, let alone its quantile function.

where $A_{t+h} = \sqrt{\sum_{i=1}^{h} E[\sigma_{t+i}^2 | I_t]}$, or the square root of the sum of one period conditional variances. Since the terms $E[\sigma_{t+i}^2 | I_t]$ is the optimal forecast of σ_{t+i}^2, this method will be referred to as the *optimal forecast* method.

Alternatively, a natural method is to use simulation to approximate the distribution of Y_{t+h}^* so that VaR is taken from the α quantile of that empirical distribution. This involves simulating a large number of iid sequences and taking sums of the resulting returns. As the number of sequences increases, the simulated total returns approximates well, the true distribution. This is referred to as the *Monte Carlo* method.

When SV models such as (6) are used, even 1-day VaR is not analytically available because $F_1(.|I_t)$ is not known. The pair (2) and (6) constitutes a nonlinear state space model, where the hidden state is x and observation is y. The distribution $F_1(.|I_t)$ corresponds to the density $p(y_{t+1}|Y_t = \{y_1, t_2, ..., y_t\})$ which only admits a tractable form in special cases. One such case is when the state space is both Gaussian and linear, where the density is normal, with mean and variance computable by Kalman filter (as discussed in Chapter 2). Needless to say, h-day VaR is even more difficult because it involves a multi-day return distribution. At this juncture, it should be clear that the SV model, by construction, puts itself in disadvantage compared to GARCH models, in performing 1-day VaR estimation. For the former, the task requires approximation or simulation, for the latter, it is a deterministic calculation. For h-day VaR estimation, the disadvantage is amplified. Both require simulation (unless other approximations as suggested above are adopted for GARCH models). However, SV models require 2 sets of simulated sequences, one for $\{z_{t+1}, z_{t+2}, ..., z_{t+h}\}$ and one for $\{x_{t+1}, x_{t+2}, ..., x_{t+h}\}$, whereas GARCH models require only 1 set, namely for $\{z_{t+1}, z_{t+2}, ..., z_{t+h}\}$. Furthermore, whilst simulating the iid z is simple, simulating paths of x involves simulating the stochastic process in (6), conditioned on the information set I_t. This means the simulation must initialise from

$\hat{x}_t \approx E[x_t|Y_t]$ which implies the need for a filtering step. Of course, the optimal Kalman filter is not applicable for this model, for the reason outlined above. Other sub-optimal filters, for example, particle filtering need to be employed instead.

4 Econometric Methodologies

The previous section identifies three main econometric tasks: estimating GARCH models, estimating the SV model, and filtering the SV model. Following is a detailed discussion of methods to deal with those tasks.

4.1 MLE Estimation of GARCH models

For GARCH models, the log-likelihood $logL(y|\Psi)$ (Ψ being the parameter vector as usual) can be written as:

$$logL(y|\Psi) = \sum_{i=1}^{N} log[p(y_i|I_{i-1})] \tag{10}$$

where N being the length of the time series. For a normal z,

$$log[p(y_t|I_{t-1}, \Psi)] = -\frac{1}{2}\left[log(2\pi) + log\sigma_t^2 + \frac{y_t^2}{\sigma_t^2}\right] \tag{11}$$

For a t-distributed z with v degrees of freedom, following Bollerslev (1987),

$$log[p(y_t|I_{t-1}, \Psi)] = log\left[\Gamma\left(\frac{v+1}{2}\right)\Gamma\left(\frac{v}{2}\right)^{-1}\left((v-2)\sigma_t^2\right)^{-1/2}\left(1 + \frac{y_t^2}{\sigma_t^2(v-2)}\right)^{-\frac{v+1}{2}}\right] \tag{12}$$

with $\Gamma(.)$ denoting the Gamma function. In both (15) and (16), σ_t is a deterministic function of Ψ as defined in (3), (4) and (5) for GARCH(1,1), RiskMetrics, and APARCH(1,1), respectively. Therefore, MLE estimates of these models can be obtained by maximizing

(15) or (16) using appropriate numerical procedures. Many statistical packages exist that perform GARCH estimation. This thesis uses the GARCH package developed by Laurent and Peters (2002), which is written in Ox, a matrix programming language that has the speed of C++.

4.2 Estimating the SV model

Let us now focus on the estimation of the state space model (2) and (6) where the error noise z is first assumed to be standard normal. The case of the t-distribution is discussed at the end of this section. In the normal case, (2) and (6) constitute a state space model that is conditionally Gaussian in both equations but nonlinear in the observation equation. As the likelihood $L(y|\Psi) = \int p(y|x,\Psi)p(x|\Psi)dx$ is not tractable, MLE is not possible. Many methods have been proposed to estimate this model, see Broto and Ruiz (2004) for a comprehensive survey. Below is a brief review of three relatively established methods: Quasi MLE (QML), Generalised Method of Moments (GMM) and Markov Chain Monte Carlo (MCMC).

QML estimation of the model, introduced in Harvey et al (1994) works on the linear formulation of the observation equation:

$$u_t = logy_t^2 = x_t + logz_t^2 \tag{13}$$

where $logz_t^2$ follows a log gamma distribution with $a = 1/2$ and $b = 2$ (Johnson, Kotz and Balakrishnan, 1997), or equivalently a log chi squared distribution with 1 degree of freedom. Hereafter, we refer to this distribution as $log\chi_1^2$. It has a mean of -1.27 and variance of $\pi^2/2$ (Ambramovitz and Stegun, 1970). The QML approach then approximates $logz_t^2$ by a normal random variable that matches the former's mean and variance. MLE is then applied to the resulting Gaussian and linear model, yielding QML estimates.

Monte Carlo simulation in Ruiz (1994) shows a considerable bias in the QML estimates both in small samples in the order of 500 observations and for small values of σ (0.3). Ruiz (1994) suggests the bias might be due to the fact that the parameter is close to the boundary of its permissible space. Nonetheless, this bias is particularly concerning because the daily subperiod sample size in this study can be that small, and the typical σ for daily time series is found to be around 0.15 to 0.25 (Jacquier et al, 1994, Broto and Ruiz, 2004). In addition, the effect of the biased estimates can be further amplified when applied to obtain the filtered estimate of x_t, which is necessary for volatility forecast.

GMM essentially matches the population moments with sample moments, using more moments than the number of parameters. The method is particularly suitable for the SV model considered here because analytical expressions are available for a large number of moments (Appendix A, Jacquier et al, 1994). Apart from the inevitable loss of efficiency by using a finite number of moments to match a distribution, one practical problem with GMM in general is how many and which moments to use. Andersen and Sorensen (1996) suggest fourteen moments are appropriate for this model, although they encounter some problem when $\beta = 0.98$, which is rather typical for daily data. Shephard (1996) lists several criticisms of GMM in the SV application.

This study employs MCMC, rated as one of the best estimation tools for the SV model. See Andersen, Chung and Sorensen (1999) for a comparison of various methods in a Monte Carlo setting. MCMC, as reviewed in Chapter 3, seeks to construct exactly the conditional density, or in Bayesian language, posterior density, by repeatedly sampling from a Markov chain whose invariant distribution is the target density of interest. Two primary sampling concepts are Metropolis-Hastings and Gibbs sampler. Direct application of MCMC to estimate parameters of the SV model is not possible because $p(\Psi|y) = \frac{p(y|\Psi)p(\Psi)}{p(y)} = \frac{p(y|\Psi)p(\Psi)}{\int p(y|\Psi)p(\Psi)d\Psi}$ and the likelihood $p(y|\Psi)$ is not tractable. The solution is to focus instead on the joint posterior $p(x, \Psi|y)$ where, using the Gibbs sampler, draws from the posterior can be obtained by alternating between sampling from the full conditional $p(\Psi|x, h)$ and from $p(x|\Psi, h)$. Samples from this chain will converge to the true posterior density so that point estimates, for example the mean, can be computed based on these samples. The most difficult part of an MCMC algorithm is to derive the posterior expressions and develop an efficient sampling scheme for each of them. There can be many sampling schemes and efficient ones are those that exploit the special structure of the model to speed up convergence. Whilst sampling from the parameter posterior can be relatively straightforward by relying on what is known as standard linear theory $p(\alpha|x, y, \phi, \sigma_v)$, sampling from $p(x|y, \Psi)$ is much harder because x is a high dimension vector, and the joint density is not known.

Sampling from $p(\Psi|x, y)$ is done by, once again, alternating between $p(\alpha|x, y, \phi, \sigma_v)$, $p(\phi|x, y, \alpha, \sigma_v)$ and $p(\sigma_v|x, y, \phi, \alpha)$. Jacquier et al (1994) specify multivariate normal priors for α and ϕ and inverse gamma for σ_v. The standard linear theory shows that the associated full posterior for the parameters are also multivariate normal and inverse gamma, respecitively, hence the name conjugate priors. Kim et al (1998) sample ϕ indirectly from $\phi^* = \frac{1}{2}\phi$ which is in turn assigned a Beta distribution, implying ϕ ranges from -1 to 1. With this Beta prior for ϕ^*, Kim et al show that ϕ can be drawn using rejection sampling.

A more challenging task is to sample from $p(x|y, \Psi)$ where $x = \{x_1, x_2, ..., x_N\}$. The econometric literature to date offers two main sampling schemes, one proposed in Jacquier et al (1994) and the other in Kim et al (1998). The former repeatedly samples from $p(x_i|x_1, ..., x_{i-1}, x_{i+1}, y, \Psi)$ which is the same as $p(x_i|x_{i-1}, x_{i+1}, y, \Psi)$ due to the Markovian structure. The latter works on the linear formulation (17) and approximates the $log z_t^2$ by a mixture of seven Gaussian distributions to match its moments. Then, by conditioning on the latent mixture component indicator s_t, $t = 1, 2, ..., N$ which is now an additional state variable, the resulting state space model is Gaussian and linear, such that drawing x can be done in one single move by appealing to a smoother version of Kalman filter, for example the Raunch-Tung-Streusel algorithm. Kim et al suggest their method converges faster than the multi move one in Jacquier et al.

Either method is rather complicated to implement and is highly model dependent. For example, a change in prior specifications requires re-coding and debugging. This operational issue is the main drawback of MCMC which has been highlighted in Chapter 3. Fortunately, there is now a general Bayesian inference package, called BUGS (Bayesian inference Using Gibbs Sampling). The package, developed since 1989 at Cambridge University, is freely available at $http : //www.mrc - bsu.cam.ac.uk/bugs$, and fully documented in (Spiegelhalter, Thomas, Best and Gilks, 1996). It is an all-purpose piece of Bayesian software that takes away the need to compute conditional posteriors. All the user has to do is to specify the prior for each of the random variables, and the likelihood of the observations conditional on these random effects. BUGS will then select, by using a small expert system, a suitable sampling scheme ranging from conjugacy to adaptive rejection sampling (Gilks and Wild, 1992) to Metropolis-Hastings. In addition, the software is supported by a separate convergence diagnostic package called CODA, written by Best, Cowles and Vines (1995) in R language, available at one of downloadable packages at $http : //cran.r - project.org/$.

4.3 Filtering the SV model

As pointed out previously, the problem of computing daily VaR is recursive: each day, a new VaR is computed, based on the latest return. In the signal processing literature, this exercise is called on-line estimation. Bayesian MCMC is an off-line technique: it is inference based on a fixed set of observations such that updating new information prompts re-estimation. Although the MCMC smoothed estimate for the last state x_N is also the same as the filtered estimate, using MCMC for recursive exercises is not efficient. What is needed instead is a filtering algorithm that can recursively estimate the current estimate based on yesterday's estimate, without the need to go back to the whole history of observations. One such algorithm is the Kalman filter applied on model (2) and (17) where $logz_t^2$ is approximated by a Gaussian noise as in Harvey et al (1994)'s QML procedure. A potentially more attractive solution is particle filtering, in particular, Auxiliary Particle Filter (APF) (Pitt and Shephard, 1999), reviewed in Chapter 2. Particle filtering, as comprehensively discussed in Doucet et al (2001) is a general filtering algorithm that does not make assumptions on linearity nor Gaussianity.

Briefly, particle filtering employs three numerical techniques to recursively estimate the state variable given current and past observations, and model parameters. These techniques are importance sampling, sequential importance sampling and resampling. Essentially at each time step, a filtered estimate is approximated by a large number of samples, or particles, drawn from a carefully selected distribution and these particles are weighted according to their likelihood values. The key to a particle filter therefore is to identify a good density from which particles are sampled. Pitt and Shephard (1999) discuss "adaptive" densities, or those that take into account latest information so that the issue of degeneracy is minimized. One such density is $p(x_{t+1}, k|Y_{t+1}) \propto p(y_{t+1}|\mu_{t+1}^k)p(x_{t+1}|x_t^k)$, where k is an index, and μ_{t+1}^k represents the mean, or mode, of the distribution $p(x_{t+1}|x_t^k)$. This gives rise to the basic Auxiliary Particle Filter (APF) by Pitt and Shephard (1999). Intuitively, this method seeks to simulate particles from those "parents" that are likely to produce "children" that fit the new observation. APF has been applied with great success to filter SV models (Pitt and Shephard, 1999) and to estimate the likelihood function for

Bayes factor computation (Kim et al 1998, Omori et al, 2006, and Raggi and Bordignon, 2006). The filtering task in the following empirical analysis adopts the APF.

5 Empirical Analysis

5.1 Data and stylized facts

This empirical analysis employs daily time series from two equity indices, the U.S S&P 500 (8 May 1985-22 December 2006) and All Ordinaries (21 June 1985-22 December/2006). Both are obtained from Datastream. The comparison amongst the four volatility models is performed on the sub-sample 4 March 2003-22 December 2006. During this period, model re-estimation is done every 50 days, with the first estimation based on the first two years of daily data.[5] This construction implies 9 re-estimations of the models, and 448 and 459 observations for out-of-sample evaluation for the S&P 500 and All Ordinaries index, respectively. The full dataset is used for rolling estimations of 10 daily return models (i.e observations are time series of 10 trading day returns) for 10 day VaR computation. This implies 448 and 460 re-estimations for S&P 500 and All Ordinaries respectively, each for 1 daily VaR estimate. Therefore, there are also 448 and 460 observations, for S&P 500 and All Ordinaries respectively, for out-of-sample evaluation of the 10 day data models.

[5]This temporal lag of 50 days is suggested in Giot and Laurent (2003) and Raggi and Bordignon (2006). Shorter lags have been experimented and show very little differences.

Table 3: Descriptive Statistics of Data Sample

	Full Sample 1985-2006		Sub-sample 2003-2006	
	S&P 500	All Ord.	S&P 500	All Ord.
Mean	0.04	0.03	0.05	0.07
Std Deviation	1.06	0.93	0.75	0.61
Skewness	-0.37	-0.55	0.07	-0.23
Kurtosis	8.13	9.93.	4.51	5.03
Jarque-Bera statistic	6,110**	11,146**	92.57**	174**

(*) and (**) denote rejection of normality at 5% and 1% significance, respectively

Table 3 reports descriptive statistics for this dataset. The ex-crash full sample still exhibits fat-tailedness, and the feature is more pronounced for the All Ordinaries time series. This may suggest the crash had more lingering effect on the Australian market than the US market, reflected in the longer period of high volatility post October. This "volatility premium" also prevails in the sub-sample data which encompasses a local correction in October 2005 and the recent May-July 2006 turbulence induced by global inflation fears. Both time series feature rather symmetrical distributions, although the Australian time series is more skewed. Figure 3 plots return and two volatility estimates, GARCH(1,1) and SV, for the sub-sample period where VaR backtesting is applied. Time variation and clustering of volatility are in evidence. Also, the wider range of return distribution suggests heavy tail distributions may be more appropriate for the All Ordinary Index than for the S&P 500 for this dataset. Furthermore, note that the two volatility estimates are almost indistinguishable. This implies the difference between GARCH and SV models in historical fitting, if any, is very fine.

Table 4 presents the estimation result for the eight models being considered, for the S&P 500 and the All Ordinaries Index. For each model, the mean and standard deviation of parameter estimates are computed across nine overlapping estimation periods. In addition, the mean and standard deviation of the log-likelihoods are also presented. The estimates are within the expected range except for the degree of freedom estimates for the APARCH-

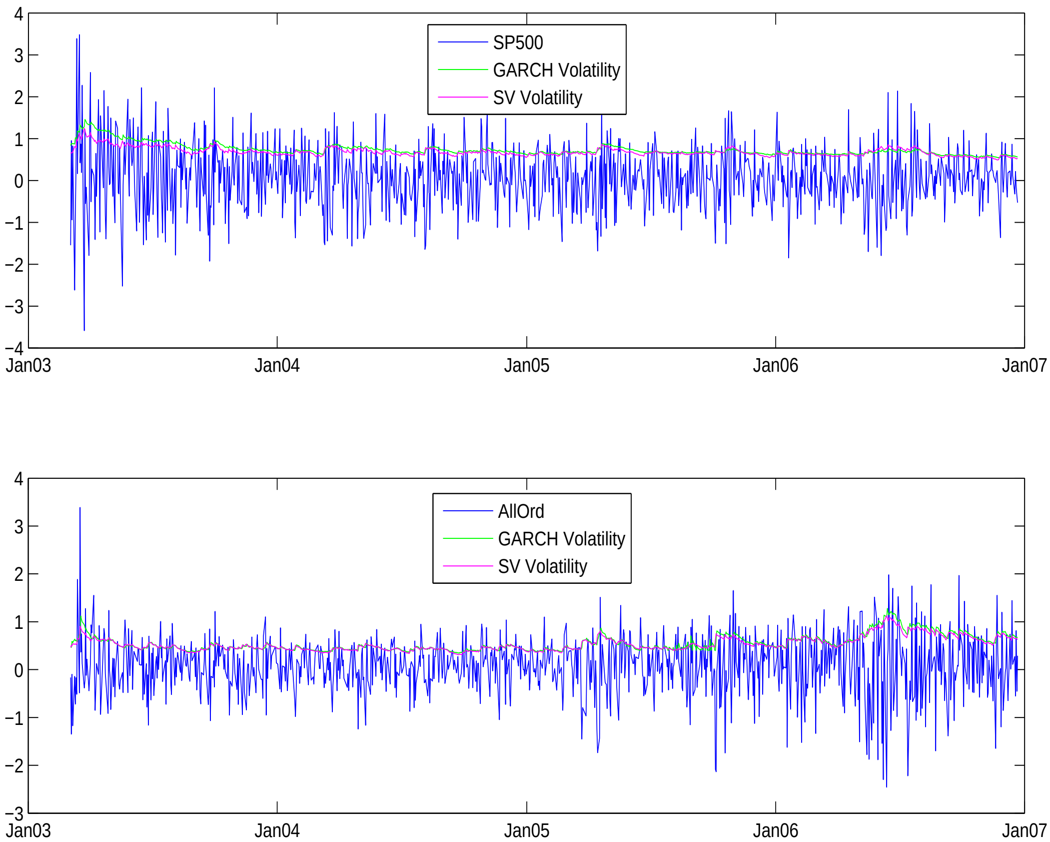

Figure 1: Return Data and Estimated Volatility

t model which are abnormally high for both time series. These results raise concern on reliability of the MLE estimates in this model where the optimisation is over as many as 7 parameters (including the constant term in the mean equation). This is despite the result that the APARCH models are amongst the best fitting models in terms of the log-likelihood statistics. In this aspect, the SV model (with a normal distribution) best describes the return time series, a result that is consistent with other studies such as Hsieh, 1991, Jacquier, Polson and Rossi, 1994, Danielson, 1994, and Shephard 1996.

Table 4: Estimation Results on Eight Volatility Models

	GARCH	GARCH-t	RiskMetrics	RiskMetrics-t	APARCH	APARCH-t	SV	SV-t
Panel A: U.S S&P 500 (March 2003 - December 2006) (standard deviation in parenthesis)								
a_0	0.035	0.035			0.033	0.073		
	(0.016)	(0.016)			(0.018)	(0.098)		
a_1	0.035	0.034	0.060	0.060	0.027	0.014		
	(0.009)	(0.008)	NA	NA	(0.014)	(0.046)		
b_1	0.888	0.889	0.094	0.094	0.896	0.846		
	(0.037)	(0.037)	NA	NA	(0.049)	(0.099)		
γ					0.973	0.706		
					(0.042)	0.608		
δ					1.885	1.920		
					(0.695)	(0.883)		
α							-0.044	-0.110
							(0.006)	(0.029)
ϕ							0.945	0.877
							(0.002)	(0.033)
σ_v							0.129	0.104
							(0.007)	(0.006)
v		495.1		534.2		$6.5\text{x}10^{11}$		19.5
		(457.9)		(566.8)		($3.2\text{x}10^{11}$)		(0.6)
log-llh	-528.4	-528.4	-533.9	-533.6	-521.5	-520.6	-511.5	-530.6
	(33.6)	(33.6)	(32.4)	(32.6)	(32.6)	(32.9)		
Panel B: Australian All Ordinaries Index (March 2003 - December 2006) (standard deviation in parenthesis)								
a_0	0.029	0.043			0.046	0.038		
	(0.047)	(0.063)			(0.029)	(0.026)		
a_1	0.093	0.100	0.060	0.060	0.069	0.068		
	(0.046)	(0.056)	NA	NA	(0.027)	(0.027)		
b_1	0.776	0.704	0.940	0.940	0.818	0.832		
	(0.278)	(0.362)	NA	NA	(0.058)	(0.055)		
γ					0.925	0.886		
					(0.116)	(0.138)		
δ					1.688	1.743		
					(1.215)	(1.022)		
α							-0.059	-0.060
							(0.028)	(0.030)
ϕ							0.961	0.963
							(0.015)	(0.015)
σ_v							0.140	0.127
							(0.006)	(0.004)
v		13.4		10.7		$1.7\text{x}10^{11}$		18.1
		(6.2)		(1.3)		($3.5\text{x}10^{11}$		(1.0)
log-llh	-373.9	-370.3	-380.1	-374.9	-367.3	-364.3	-342.5	-355.9
	(60.8)	(60.7)	(59.9)	(60.1)	(62.6)	(61.6)		

5.2 Hypothesis based Backtests

Throughout this empirical analysis, the hypothesis testing procedures by Kupiec (1995) and Christoffersen (1998) are adopted to backtest alternative volatility models. These procedures test two aspects of VaR forecast performance: unconditional coverage and independence. The first aspect concerns with the extent to which the actual violation rate is sufficiently close to the predicted one. The second aspect deals with the degree of correlation across violations, with the principle that with a good model, there should not be a systematic pattern in the series of violations. The unconditional coverage test draws on the insight that if a model accurately predicts VaR, the number of violations should follow an independent binomial distribution with parameter $p = \alpha$, α being the predicted violation rate. The independence test specifically detects a first order Markov chain behaviour in adjacent day violations. As the tests are formulated with the null hypotheses being that the rate of violations is as predicted and that they are independent, a p-value that is close to zero suggests rejection.

5.3 1-day Value-at-Risk

Table 5 reports test statistics for 1-day VaR backtesting performance of the 8 models under investigation. Associated p-values are in square brackets.An immediate observation that emerges from the table is that for the S&P 500, all models are adequate in VaR forecasting, at both 99% and 95% coverage, for both unconditional coverage (i.e accuracy) and independence. As the time series is rather well behaving as noted above, this result suggests any reasonable model of volatility can well capture the downside risk in a normal market. For all models except APARCH, the formulations with a t-distribution generate more accurate forecasts. However, the improvement is marginal, suggesting fat-tailedness is not particularly dominant in for this sample. Amongst the four models, the SV (with a t-distribution) reports the most accurate violation rate, at 5.25% and 1.31% for 95% and 99% coverage, respectively. Of course these statistics do not imply statistical superiority in hypothesis testing. The RiskMetrics-t model is also accurate, reporting violation rates at 3.94% and 1.31% respectively.

On the other hand, the All Ordinary Index is more difficult for VaR forecasts, with all models but RiskMetrics-t and GARCH-t being rejected at either of, or both, unconditional coverage and independence. Interestingly, both formulations of the SV model are rejected at both 95% and 99% coverage. This finding is against those in Raggi and Bordignon (2006) where the SV model is found to be adequate especially at 99% coverage. On the other hand, it corroborates results from Eberlein et al (2003) that conclude the SV model is rejected at 99% level. [6]It has been noted previously that by construction, SV models are at a disadvantage when it comes to one-day-ahead forecasting because they require a simulation step for predicting the one-step-ahead distribution.

In contrast, the GARCH-t model and RiskMetrics-t model are not rejected for the conditional coverage test, at both levels of significance. The result on the GARCH model is not surprising because not only the model is known to nest important aspects of many other ARCH specifications, but also the finding corroborates those in existing studies including Berkowitz and O'Brien (2003)and Brooks and Persand (2004). [7] The result on the RiskMetrics model with a t-distribution is much more interesting because the model is very simple and requires minimal econometric effort (as only the degree of

[6]Note however that the models in the studies being compared are not identical. The SV model in Raggi and Bordignon (2006) includes the leverage effect whereas that in Eberlein et al (2003) adopts a hyperbolic distribution for fat-tailedness instead of a t-distribution.

[7]These two studies do not consider the same pool of models as the one in this paper. Nevertheless, the general conclusion arising from them is supportive of the GARCH model.

freedom for the t-distribution is to be estimated). Note that previous studies do not consider RiskMetrics-t. Huang and Lin (2004) explicitly compare RiskMetrics with normality assumption against APARCH (t and normal) and find RiskMetrics underestimates the risk, a result that is corroborated by Table 4. One possible explanation for the RiskMetrics success in this analysis might be due to its parsinomious nature leading to a simpler optimisation task, hence lower estimation risk. This advantage is particularly attractive for time series that exhibit considerable outliers such as the All Ordinary Index because MLE solutions for such processes may be difficult to obtain, or not global. This operational issue may also explain the poor performance of APARCH which requires estimation of up to 7 parameters. For example, the estimated parameter of fat-tailedness for APARCH-t is unusually large for some estimation periods, in the order of millions (implying normality), suggesting the numerical optimiser might not be reliable for this data. The results on APARCH stand against those in Huang and Lin (2004) and Raggi and Bordignon (2006) which support the normal version of the model.

Table 5: Backtests of 1-day VaR

		S&P 500					All Ordinaries			
	Confidence level	Violation rate	Uncond. cover	Independ.	Cond. cover	Violation rate	Uncond. cover	Independ.	Cond. cover
RiskMetrics	95%	5.25%	0.06	0.41	0.47	7.63%	5.78*	0.68	6.43*
			[0.807]	[0.520]	[0.789]		[0.016]	[0.408]	[0.040]
RiskMetrics-t	95%	3.94%	1.17	1.76	2.93	5.66%	0.41	3.48	3.89
			[0.280]	[0.184]	[0.231]		[0.522]	[0.062]	[0.143]
GARCH	95%	5.03%	0.00	2.35	2.35	8.28%	8.75**	0.53	9.00*
			[0.974]	[0.125]	[0.309]		[0.003]	[0.615]	[0.011]
GARCH-t	95%	5.03%	0.00	0.57	0.57	5.88%	0.71	3.04	3.76
			[0.974]	[0.452]	[0.753]		[0.398]	[0.081]	[0.153]
APARCH	95%	4.81%	0.03	2.74	2.77	9.1%	13.51**	0.37	13.90**
			[0.854]	[0.098]	[0.250]		[0.000]	[0.534]	[0.000]
APARCH-t	95%	5.69%	0.44	1.39	1.83	7.63%	5.78*	1.96	7.74*
			[0.508]	[0.238]	[0.400]		[0.016]	[0.161]	[0.021]
SV	95%	5.47%	0.21	1.68	1.88	7.84%	6.71*	0.52	7.22*
			[0.649]	[0.195]	[0.390]		[0.010]	[0.473]	[0.027]
SV-t	95%	5.25%	0.06	2.00	2.06	8.06%	7.70**	0.37	8.07*
			[0.807]	[0.158]	[0.358]		[0.006]	[0.542]	[0.018]
RiskMetrics	99%	1.75%	2.12	0.29	2.41	3.05%	12.60**	3.41	16.01**
			[0.145]	[0.593]	[0.300]		[0.000]	[0.065]	[0.000]
RiskMetrics-t	99%	1.31%	0.41	0.16	0.57	1.31%	0.40	3.55	3.95
			[0.521]	[0.689]	[0.752]		[0.528]	[0.060]	[0.139]
GARCH	99%	1.31%	0.41	0.16	0.57	4.14%	25.62**	1.47	27.09**
			[0.521]	[0.689]	[0.752]		[0.000]	[0.225]	[0.000]
GARCH-t	99%	1.31%	0.41	0.16	0.57	1.53%	1.10	2.93	4.03
			[0.521]	[0.689]	[0.752]		[0.294]	[0.087]	[0.133]
APARCH	99%	1.75%	2.12	0.29	2.41	4.14%	25.62**	1.47	27.09**
			[0.145]	[0.593]	[0.300]		[0.000]	[0.225]	[0.000]
APARCH-t	99%	1.75%	2.12	0.29	2.41	2.83%	10.40**	3.95*	14.35**
			[0.145]	[0.593]	[0.300]		[0.001]	[0.047]	[0.000]
SV	99%	1.31%	0.41	0.16	0.57	3.70%	20.04**	2.12	22.16**
			[0.521]	[0.689]	[0.752]		[0.000]	[0.146]	[0.000]
SV-t	99%	1.31%	0.41	0.16	0.57	3.05%	12.60**	3.41	16.01**
			[0.521]	[0.689]	[0.752]		[0.000]	[0.065]	[0.000]

(*) and (**) indicate rejection of the null at 5% and 1% significance levels, respectively

5.4 10-day Value-at-Risk

The main focus of this paper is on forecast of 10-day VaR. As noted previously, this is an extremely difficult problem by construction because multi-day disturbances will likely result in serial violations, jeopardizing both accuracy and independence. The popular technique in practice is to scale the 1-day VaR by a factor of square root of 10. Results for this implementation are reported in Table 6. Another possible approach is to use optimal forecast of the variance, defined in section 4:

$$E[\Sigma_{t+h}^2|I_t] = \sum_{i=1}^{h} E[\sigma_{t+i}^2|I_t] \tag{14}$$

This approach is available only for GARCH type models. For the GARCH model,

$$E[\sigma_{t+i}^2|I_t] = a_0 \sum_{j=0}^{j-1}(a_1+b_1)^j + (a_1+b_1)^{i-1}(a_1 y_t^2 + b_1\sigma_t^2) \tag{15}$$

For APARCH models, there exist a recursive formula for $E[\sigma_{t+i}^\delta|I_t]$ (see Peters and Laurent, 2001):

$$E[\sigma_{t+i}^\delta|I_t] = a_0 + (a_1\kappa + b_1)E[\sigma_{t+i-1}^\delta|I_t] \tag{16}$$

where

$$E[\sigma_{t+1}^\delta|I_t] = a_0 + a_1(|y_t| - \gamma y_t)^\delta + b_1\sigma_t^\delta \tag{17}$$

For normal error:

$$\kappa = \frac{1}{\sqrt{2\pi}}[(1+\gamma)^\delta + (1-\gamma)^\delta]2^{\frac{\delta-1}{2}}\Gamma(\frac{\delta+1}{2}) \tag{18}$$

and for t-distributed error:

$$\kappa = [(1+\gamma)^\delta + (1-\gamma)^\delta]\Gamma(\frac{\delta+1}{2})\Gamma(\frac{v}{2})(v-2)^{\frac{\delta+1}{2}}/[2\sqrt{(v-2)\pi}\Gamma(\frac{v}{2}] \tag{19}$$

Optimal forecasts of variance for RiskMetrics is the same as the variance estimate based on time scaling rule. Table 7 summarises backtest results for 10-day VaR forecast based on optimal forecast.

Another alternative is to simulate the variance and return as discussed in section 4. Results are summarized in Table 8. A final implementation considered here is to employ time series of 10-daily returns such that daily 10-day VaR estimates are simply based on one-step-ahead volatility forecasts. Estimation is rolling, based on 500 observations for each estimation, with the first estimation period ending on the same day as that of the daily data estimation. This implementation therefore necessitates a very long sample spanning the 1987 crash for both markets. Results for this implementation are summarized in Table 9. This implementation is not performed for SV models as computational cost would be prohibitive.

From Table 6-9, it can be seen that no model passes the independence test, regardless of the choice of variance forecast. That is, serial violation is inevitable in turbulent conditions that last several days. Very likely better modeling will not be the solution. Instead, the modeler might want to focus on unconditional coverage and hope to pass the test by trying to restrict violations to periods of prolonged turbulence. Note that the actual backtest formula mandated in the Basel framework does not directly penalize correlated violations. Alternatively, the fact that the problem is insurmountable by construction suggests the 10-day rule is not practical and meaningless. However, critiquing regulators policy is beyond the scope of this empirical study.

Returning to the empirical results, let us now focus instead on the test of unconditional coverage. Regardless of 10-day implementation methods, all models pass this test for the "well-behaving" S&P time series, but fail for the All Ordinaries index, where violations are concentrated around the two turbulent periods October 2005 and May-July 2006, possibly due to the perculiarity of the sample. Implementations based on time scaling and optimal forecasts are ineffective, with rejection reported for all GARCH and SV models. When conditional variance is obtained by simulation, RiskMetrics is the better model, able to accurately (in statistical sense) compute VaR at 95% coverage (RiskMetrics-normal) and 99% coverage (RiskMetrics-t). Predictably, the SV models are ineffective for 10-day forecasts due to the need to simulate two random paths.

When the forecast is based on 10-daily returns as reported in Table 9, all models report less violations compared to the other implementations. In particular, GARCH-normal, APARCH-normal and APARCH-t all pass the unconditional coverage test, with RiskMetrics-t and GARCH-t passing the test for the 99% coverage. The main reason for this success is because of the 1987 crash inclusion which ensures heavy tails are adequately accounted for, hence higher VaR forecast than otherwise. In fact, when the implementation is repeated on the same sample that excludes the crash effect [8], more violations are reported and none of the models pass the test. The fact that the APARCH model "suddenly" becomes adequate in this implementation setting is interesting. A closer look at the parameter estimates show that MLE results are more reasonable. For example, the estimated parameter of the t-distribution for APARCH-t is very stable and reasonable for heavy tail distributions. It has a mean of 7.2 and standard deviation of 0.6. Overall, there is evidence of improvement in forecasts of multi-period variance by using data that includes extreme outcomes. All models however do not pass independence test as violations, albeit reduced in number, remain concentrated around the two turbulent periods.

[8] This is done by replacing the month surrounding the crash by a return series generated randomly from a normal distribution using the sample mean and variance

Table 6: Backtests of 10-day VaR based on time scaling

	S&P 500					All Ordinaries			
	Confidence level	Violation rate	Uncond. cover	Independ.	Cond. cover	Violation rate	Uncond. cover	Independ.	Cond. cover
RiskMetrics	95%	5.13%	0.02	72.19**	72.20**	9.11%	13.02**	144.79**	157.81**
			[0.897]	[0.000]	[0.000]		[0.000]	[0.000]	[0.000]
RiskMetrics-t	95%	3.35%	2.90	68.51**	71.40**	8.44%	9.40**	153.16**	162.56**
			[0.089]	[0.000]	[0.000]		[0.002]	[0.000]	[0.000]
GARCH	95%	4.46%	0.28	40.79**	41.07**	10.00%	18.59**	123.33**	141.92**
			[0.597]	[0.000]	[0.000]		[0.000]	[0.000]	[0.000]
GARCH-t	95%	4.46%	0.28	40.79**	41.07**	8.67%	10.55**	136.42**	146.97**
			[0.597]	[0.000]	[0.000]		[0.000]	[0.000]	[0.000]
APARCH	95%	4.46%	0.28	48.84**	49.12**	11.11%	26.66**	133.12**	159.78**
			[0.597]	[0.000]	[0.000]		[0.000]	[0.000]	[0.000]
APARCH-t	95%	4.24%	0.57	43.58**	44.15**	9.78%	17.12**	119.55**	136.67**
			[0.450]	[0.000]	[0.000]		[0.000]	[0.000]	[0.000]
SV	95%	4.91%	0.01	58.44**	58.45**	9.78%	17.12**	128.29**	145.41**
			[0.931]	[0.000]	[0.000]		[0.000]	[0.000]	[0.000]
SV-t	95%	5.80%	0.58	77.80**	78.38**	9.33%	14.33**	129.52**	143.86**
			[0.446]	[0.000]	[0.000]		[0.000]	[0.000]	[0.000]
RiskMetrics	99%	2.01%	3.56	30.22**	33.78**	4.89%	35.52**	86.73**	122.25**
			[0.059]	[0.000]	[0.000]		[0.000]	[0.000]	[0.000]
RiskMetrics-t	99%	1.34%	0.47	30.43**	30.91**	2.22%	5.04*	27.31**	32.35**
			[0.493]	[0.000]	[0.000]		[0.025]	[0.000]	[0.000]
GARCH	99%	0.89%	0.05	27.01**	27.06**	5.78%	49.26**	107.54**	156.80**
			[0.816]	[0.000]	[0.000]		[0.000]	[0.000]	[0.000]
GARCH-t	99%	0.89%	0.05	27.01**	27.06**	2.22%	5.04*	27.31**	32.35**
			[0.816]	[0.000]	[0.000]		[0.025]	[0.000]	[0.000]
APARCH	99%	1.34%	0.47	30.44**	30.91**	6.22%	56.64**	96.85**	153.48**
			[0.493]	[0.000]	[0.000]		[0.000]	[0.000]	[0.000]
APARCH-t	99%	1.12%	0.06	22.56**	22.62**	4.67%	32.32**	71.42**	103.74**
			[0.809]	[0.000]	[0.000]		[0.000]	[0.000]	[0.000]
SV	99%	1.34%	0.47	30.44**	30.91**	3.78%	20.54**	59.35**	79.89**
			[0.493]	[0.000]	[0.000]		[0.000]	[0.000]	[0.000]
SV-t	99%	0.89%	0.57	43.58**	44.15**	4.00%	23.32**	55.44**	78.76**
			[0.450]	[0.000]	[0.000]		[0.000]	[0.000]	[0.000]

(*) and (**) indicate rejection of the null at 5% and 1% significance levels, respectively

Table 7: Backtests of 10-day VaR based on optimal variance forecast

	S&P 500					All Ordinaries			
	Confidence level	Violation rate	Uncond. cover	Independ.	Cond. cover	Violation rate	Uncond. cover	Independ.	Cond. cover
RiskMetrics	95%	5.13%	0.02	72.19**	72.20**	9.11%	13.02**	144.79**	157.81**
			[0.900]	[0.000]	[0.000]		[0.000]	[0.000]	[0.000]
RiskMetrics-t	95%	3.35%	2.90	68.51**	71.41**	8.44%	9.40**	153.16**	162.56**
			[0.089]	[0.000]	[0.000]		[0.002]	[0.000]	[0.000]
GARCH	95%	4.46%	0.28	40.79**	41.07**	10.89%	24.95**	129.53**	154.49**
			[0.600]	[0.000]	[0.000]		[0.000]	[0.000]	[0.000]
GARCH-t	95%	4.46%	0.28	40.79**	41.07**	8.67%	10.55**	136.42**	146.97**
			[0.089]	[0.000]	[0.000]		[0.001]	[0.000]	[0.000]
APARCH	95%	4.24%	0.57	43.58**	44.15**	11.56%	30.21**	131.75**	161.96**
			[0.450]	[0.000]	[0.000]		[0.000]	[0.000]	[0.000]
APARCH-t	95%	NA	NA	NA	NA	7.78%	6.30*	129.20**	135.50**
			NA	NA	NA		[0.012]	[0.000]	[0.000]
RiskMetrics	99%	2.01%	3.56	30.22**	33.78**	4.89%	35.52**	86.73**	122.25**
			[0.059]	[0.000]	[0.000]		[0.000]	[0.000]	[0.000]
RiskMetrics-t	99%	1.34%	0.47	30.44**	30.91**	2.22%	5.04*	27.31**	32.35**
			[0.493]	[0.000]	[0.000]		[0.025]	[0.000]	[0.000]
GARCH	99%	0.89%	0.05	27.01**	27.06**	5.78%	49.26**	107.54**	156.80**
			[0.816]	[0.000]	[0.000]		[0.000]	[0.000]	[0.000]
GARCH-t	99%	0.89%	0.05	27.01**	27.06**	2.67%	8.67**	39.66**	48.33**
			[0.816]	[0.000]	[0.000]		[0.003]	[0.000]	[0.000]
APARCH	99%	1.34%	0.471	30.44**	30.91**	7.11%	72.28**	125.98**	198.26**
			[0.493]	[0.000]	[0.000]		[0.000]	[0.000]	[0.000]
APARCH-t	99%	NA	NA	NA	NA	2.67%	8.67**	49.82**	58.49**
			NA	NA	NA		[0.003]	[0.000]	[0.000]

(*) and (**) indicate rejection of the null at 5% and 1% significance levels, respectively

Table 8: Backtests of 10-day VaR based on simulation

		S&P 500				All Ordinaries			
	Confidence level	Violation rate	Uncond. cover	Independ.	Cond. cover	Violation rate	Uncond. cover	Independ.	Cond. cover
RiskMetrics	95%	5.36%	0.12	77.18**	77.29**	6.89%	3.04	92.13**	95.17**
			[0.732]	[0.000]	[0.000]		[0.081]	[0.000]	[0.000]
RiskMetrics-t	95%	3.13%	3.80	74.17**	77.97**	7.33%	4.54*	141.73**	146.27**
			[0.051]	[0.000]	[0.000]		[0.032]	[0.000]	[0.000]
GARCH	95%	4.46%	0.28	40.79**	41.07**	10.67%	23.29**	134.51**	157.80**
			[0.597]	[0.000]	[0.000]		[0.000]	[0.000]	[0.000]
GARCH-t	95%	4.46%	0.28	40.79**	41.07**	8.67%	10.55**	136.42**	146.97**
			[0.597]	[0.000]	[0.000]		[0.001]	[0.000]	[0.000]
APARCH	95%	4.24%	0.57	43.58**	44.15**	11.56%	30.21**	140.20**	170.40**
			[0.450]	[0.000]	[0.000]		[0.000]	[0.000]	[0.000]
APARCH-t	95%	4.24%	0.57	43.58**	44.15**	10.44%	21.68**	130.82**	152.49**
			[0.450]	[0.000]	[0.000]		[0.000]	[0.000]	[0.000]
SV	95%	4.91%	0.01	58.44**	58.45**	10.22%	20.11**	127.09**	147.20**
			[0.931]	[0.000]	[0.000]		[0.000]	[0.000]	[0.000]
SV-t	95%	6.03%	0.94	73.79**	74.73**	10.44%	21.68**	130.82**	152.49**
			[0.333]	[0.000]	[0.000]		[0.000]	[0.000]	[0.000]
RiskMetrics	99%	1.56%	1.22	38.14**	39.36**	2.44%	6.76**	67.35**	74.11**
			[0.269]	[0.000]	[0.000]		[0.009]	[0.000]	[0.000]
RiskMetrics-t	99%	1.34%	0.47	30.43**	30.90**	1.56%	1.20	9.01**	10.21**
			[0.493]	[0.000]	[0.000]		[0.273]	[0.003]	[0.006]
GARCH	99%	0.89%	0.05	27.01**	27.06**	5.56%	45.70**	92.00**	137.70**
			[0.816]	[0.000]	[0.000]		[0.000]	[0.000]	[0.000]
GARCH-t	99%	0.89%	0.05	27.01**	27.06**	2.18%	10.75**	36.42**	47.17**
			[0.816]	[0.000]	[0.000]		[0.001]	[0.000]	[0.000]
APARCH	99%	1.12%	0.06	35.69**	35.75**	6.44%	60.44**	101.61**	162.05**
			[0.809]	[0.000]	[0.000]		[0.000]	[0.000]	[0.000]
APARCH-t	99%	0.89%	0.05	27.01**	27.06**	5.11%	38.82**	72.32**	111.15**
			[0.816]	[0.000]	[0.000]		[0.000]	[0.000]	[0.000]
SV	99%	0.89%	0.05	27.01**	27.06**	4.44%	29.21**	66.15**	95.36**
			[0.816]	[0.000]	[0.000]		[0.000]	[0.000]	[0.000]
SV-t	99%	1.12%	0.06	22.56**	22.62**	4.22%	26.21**	60.82**	87.03**
			[0.809]	[0.000]	[0.000]		[0.000]	[0.000]	[0.000]

(*) and (**) indicate rejection of the null at 5% and 1% significance levels, respectively

Table 9: Backtests of 10-day VaR based on 10-daily time series									
	S&P 500					All Ordinaries			
	Confidence level	Violation rate	Uncond. cover	Independ.	Cond. cover	Violation rate	Uncond. cover	Independ.	Cond. cover
RiskMetrics	95%	4.91%	0.01	50.34**	50.35**	8.00%	7.27**	104.94**	112.21**
			[0.931]	[0.000]	[0.000]		[0.007]	[0.000]	[0.000]
RiskMetrics-t	95%	3.57%	2.13	44.48**	46.61**	8.22%	8.31**	127.89**	136.20**
			[0.145]	[0.000]	[0.000]		[0.004]	[0.000]	[0.000]
GARCH	95%	4.24%	0.57	51.80**	52.37**	5.56%	0.28	73.17**	73.45**
			[0.450]	[0.000]	[0.000]		[0.595]	[0.000]	[0.000]
GARCH-t	95%	2.01%	10.81**	40.78**	51.59**	2.89%	4.95*	36.42**	41.37**
			[0.001]	[0.000]	[0.000]		[0.026]	[0.000]	[0.000]
APARCH	95%	2.46%	7.46**	43.32**	50.77**	5.56%	0.28	64.63**	64.92**
			[0.006]	[0.000]	[0.000]		[0.595]	[0.000]	[0.000]
APARCH-t	95%	1.79%	12.81**	33.75**	46.55**	4.00%	1.01	38.66**	39.68**
			[0.000]	[0.000]	[0.000]		[0.314]	[0.000]	[0.000]
RiskMetrics	99%	1.56%	1.22	26.61**	27.84**	2.89%	10.75**	36.42**	47.17**
			[0.269]	[0.000]	[0.000]		[0.001]	[0.000]	[0.000]
RiskMetrics-t	99%	0.67%	0.56	18.00**	18.56**	2.00%	3.52	30.26**	33.78**
			[0.455]	[0.000]	[0.000]		[0.061]	[0.000]	[0.000]
GARCH	99%	1.34%	0.47	30.44**	30.91**	2.00%	3.52	30.26**	33.78**
			[0.493]	[0.000]	[0.000]		[0.061]	[0.000]	[0.000]
GARCH-t	99%	NA	NA	NA	NA	0.67%	0.57	6.59*	7.16*
			NA	NA	NA		[0.449]	[0.010]	[0.028]
APARCH	99%	0.89%	0.05	27.01**	27.06**	1.56%	1.20	26.65**	27.85**
			[0.816]	[0.000]	[0.000]		[0.273]	[0.000]	[0.000]
APARCH-t	99%	NA	NA	NA	NA	0.67%	0.57	18.01**	18.59**
			NA	NA	NA		[0.449]	[0.000]	[0.000]

(*) and (**) indicate rejection of the null at 5% and 1% significance levels, respectively

6 Conclusion

This paper compares the effectiveness in forecasting Value-at-Risk of the autoregressive stochastic volatility model and three representatives of the ARCH typed family: GARCH

(1,1), RiskMetrics and APARCH(1,1). The analysis is conducted on two datasets based on the U.S S&P 500 and the Australian All Ordinary Index respectively, with the former characteristically representative of normal market conditions and the latter more volatile conditions. The findings suggest that for 1-day-ahead VaR forecasting, any reasonable volatility model can perform well in normal market conditions that can be described as having few periods of prolonged turbulence such as the studied sample of the S&P index. However, when the VaR forecast is performed for a market that exhibits multiple periods of persistent volatility, such as the selected sample of the Australian index, differences in the models' performance emerge, with the stochastic volatility model outperformed by the simpler RiskMetrics and GARCH models that incorporate heavy tails in their return processes. This inferiority also holds true when comparing 10-day VaR forecasts. Given the demonstrable complexity in implementing this stochastic model, and its unsuitability for pricing purposes, one can start questioning the model's usefulness at all in finance, as well as the still growing econometric literature that studies this model.

In contrast, the much simpler, estimation-free, RiskMetrics model, has shown to be the most effective on the whole, for this particular dataset. Moreover, RiskMetrics that incorporate a heavy tail distribution, appears to perform better than one with a normal distribution, once again corroborating previous studies that support the use of heavy tail distributions in modeling asset returns. Regarding RiskMetrics's relative success, one possible reason might be that the model is not exposed to the operational issues faced by the other ARCH models that need to be estimated. Indeed, unstable optimisation that is particularly severe for outliers rich samples, may jeopadise the model's performance.

In 10-day VaR forecasting, which is actually required by the Basel Committee, although neither implementation procedures considered are effective in addressing serial violation, there is encouraging evidence that the accuracy of the forecasting can be improved by either simulating the volatility and return processes, or using 10-daily return time series. The latter approach is found to generate more conservative forecasts, possibly because the lengthened sample encloses the 1987 crash. This enclosure is entirely consistent with the general sensible practice to be conservative in measuring risk. Naturally, too much conservatism, a practice that is suggested of the US institutions in (Berkowitz and O'Brien, 2002), however, might penalise the firm in the form of overprovisioning for risk. This issue in the context of 10-day VaR can be a topic of future research.

References

[1] Abramowitz, M. and Stegun, I. (1972) *Handbook of Mathematical Functions*, U.S Government Printing Office, Washington D.C.

[2] Andersen, T. and Sorensen, B. (1996) "GMM Estimation of a Stochastic Volatility Model: a Monte Carlo Study", *Journal of Business and Economic Statistics*, Vol. 14, pp. 328-352.

[3] Berkowitz, J. and O'Brien, J. (2003) "How Accurate are Value-at-Risk Models at Commercial Banks", *Journal of Finance*, Vol. 57, pp. 1093-1111.

[4] Bollerslev, T. (1986) "Generalized Autoregressive Conditional Heteroscedasticity", *Journal of Econometrics*, Vol. 31, pp. 307-327.

[5] Brooks, C. and Persand, G. (2003) "Volatility Forecasting for Risk Management", *Journal of Forecasting*, Vol. 22, pp. 1-22.

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

[7] Christoffersen, P. (1998) "Evaluating Interval Forecasts", *International Economic Review*, Vol. 39, pp. 841-862.

[8] Christoffersen, P and Pelletier, D. (2004) "Backtesting Value at Risk: A Duration-Based Approach", *Journal of Financial Econometrics*, Vol. 2, pp. 84-108.

[9] Eberlein, E., Kallsen, J and Kristen, J. (2003) "Risk Management Based on Stochastic Volatility", *Journal of Risk*, Vol. 5, pp. 19-44.

[10] Engle, R.F. (1982) "Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation", *Econometrica*, Vol. (50), pp. 987-1007.

[11] Ding, Z., Granger, C. and Engle, R. (1993) "A Long Memory Property of Stock Market Returns and a New Model", *Journal of Empirical Finance*, Vol. 1, pp. 83-106.

[12] Doucet, A., de Freitas, N., and Gordon (2001) *Sequential Monte Carlo Methods in Practice*, Springer, New York.

[13] Giot, P. and Laurent, S. (2003) "Value-at-Risk for Long and Short Trading Positions", *Journal of Applied Econometrics*, Vol. 18, pp. 641-664.

[14] Giot, P. and Laurent, S. (2004) "Modeling Daily Value-at-Risk Using Realized Volatility and ARCH Type Models", *Journal of Empirical Finance*, Vol. 11, pp. 379-398.

[15] Harvey, A., Ruiz, E. and Shephard, N. (1994) "Multivariate Stochastic Variance Models", *Review of Economic Studies*, Vol. 61, pp 247-264.

[16] Huang, Y. and Lin, B-J. (2004) "Value-at-Risk Analysis for Taiwan Stock Index Futures: Fat Tail and Conditional Asymmetries in Return Innovations", *Review of Quantitative Finance and Accounting*, Vol. 22, pp. 79-95.

[17] Jacquier, E., Polson, N. and Rossi, P. (1994) "Bayesian Analysis of Stochastic Volatility Models", *Journal of Business and Economic Statistics*, Vol. 12, pp. 371-417.

[18] Jacquier, E., Polson, N. and Rossi, P. (2004) "Bayesian Analysis of Stochastic Volatility Models with Fat Tails and Correlated Errors", *Journal of Econometrics*, Vol. 122, pp. 185-212.

[19] Johnson, N., Kotz, S. and Balakrishnan, N. (1995) *Continuous Univariate Distributions*, Wiley, New York.

[20] Jorion, P. (2007) *Value at Risk: The New Benchmark for Managing Financial Risk*, McGraw-Hill, New York.

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

[22] Kupiec, P. (1995) "Techniques for Verifying the Accuracy of Risk Management Methods", *Journal of Derivatives*, Vol. 2, pp. 73-84.

[23] Melino, A. and Turnbull, S. (1990) "Pricing Foreign Currency Options with Stochastic Volatility", *Journal of Econometrics*, Vol. 45, ppl. 239-265.

[24] Meyer, R. and Yu, J. (2000) "BUGS for a Bayesian Analysis of Stochastic Volatility Models", *Econometrics Journal*, Vol. 3, pp. 198-215.

[25] Omori, Y., Chib, S. and Shephard, N. (2006) "Stochastic Volatility with Leverage: Fast and Efficient Likelihood Inference", *Journal of Econometrics*, in press.

[26] Raggi, D. and Bordignon, S. (2006) "Comparing Stochastic Volatility Models Through Monte Carlo Simulations", *Computational Statistics and Data Analysis*, Vol. 50, pp. 1678-1699.

[27] RiskMetrics Group (2001) *Return to RiskMetrics: The Evolution of a Standard*, www.riskmetrics.com

[28] Taylor, S. (1982) "Financial Returns Modelled by the Product of Two Stochastic Processes, a Study of Daily Sugar Prices 1961-1979". In *Time Series Analysis: Theory and Practice 1* (ed. O.D. Anderson), pp. 203-226.

[29] Taylor, S. (2005) *Asset Price Dynamics, Volatility, and Prediction*, Princeton University Press, Princeton.

[30] Yu, J. (2005) "On Leverage in a Stochastic Volatility Model", *Journal of Econometrics*, Vol. 127, pp. 165-178.