Uriel Moreira Silva

COMPARING CONDITIONAL AND STOCHASTIC VOLATILITY MODELS: GOODNESS OF FIT, FORECASTING AND VALUE-AT-RISK

Belo Horizonte - MG
19/02/2016

Uriel Moreira Silva

COMPARING CONDITIONAL AND STOCHASTIC VOLATILITY MODELS: GOODNESS OF FIT, FORECASTING AND VALUE-AT-RISK

Dissertação apresentada ao
Departamento de Estatística da UFMG
como requisito parcial para a obtenção
do título de Mestre em Estatística.

Advisor: Thiago Rezende dos Santos
Co-advisor: Frank Magalhães de Pinho

Belo Horizonte - MG
19/02/2016

ao meu irmão, Glauco

AGRADECIMENTOS

Primeiramente agradeço ao meu orientador Thiago e ao meu co-orientador Frank, cujo apoio e paciência foram cruciais nessa jornada e sem os quais esse trabalho não seria possível. Agradeço também ao Ivair e à Glaura pela leitura cuidadosa do manuscrito e pelos comentários e recomendações feitas para a melhora do trabalho.

Agradeço profundamente à toda minha família, em especial à minha mãe Maria Bernadete, ao meu pai Evandro e à minha noiva Alice; o amor e suporte de vocês é o fundamento da minha vida. Serei sempre grato ao meu irmão Glauco, grande companheiro e enorme influência em mim tanto como pessoa quanto como profissional.

No departamento de Estatística da UFMG, onde esse trabalho de mestrado foi realizado, agradeço profundamente à todos os meus professores e colegas pelo acolhimento, companheirismo e pelas sempre agradáveis discussões e conversas. Agradeço também aos meus professores e colegas de graduação da Economia no Ibmec MG, com os quais aprendi tanto e fui inspirado à fazer parte do meio acadêmico.

Finalmente, agradeço aos meus outros amigos e a todas às pessoas que de alguma forma contribuíram para a produção desse trabalho e da minha formação. Não citarei nomes por receio de omitir alguém.

A produção desse trabalho e meu mestrado foram totalmente financiados pela CAPES, à qual dedico um agradecimento especial. O desenvolvimento e divulgação de minha pesquisa foram financiados parcialmente pela FAPEMIG e pelo instituto CNPq, à quem também sou grato.

RESUMO

Nesse trabalho uma comparação das três famílias de volatilidade *Autoregressive Conditional Heteroskedasticity* (ARCH), *Stochastic Volatility* (SV) e *Non-Gaussian State Space Models* (NGSSM) é feita de acordo com três diferentes métricas: ajuste, previsão e *Value-at-Risk* (VaR). Procedimentos de inferência sobre a distribuição *Skew Generalized Error* são detalhados. Os respectivos critérios de avaliação usados para cada métrica são o Critério de Informação de Akaike, Erro Quadrático Médio das previsões um passo à frente e Cobertura Incondicional do VaR um passo à frente. A amostra utilizada é composta por séries de retornos diários (Ibovespa, Hang Seng Index, Merval Index e S&PTSX Index) de Janeiro de 2000 até Janeiro de 2016 ou 4000 observações, das quais 3000 são utilizadas para estimação e 1000 são reservadas para previsão e avaliação do VaR. As estimativas obtidas servem de base para a condução de um experimento de simulação envolvendo 1000 replicações de séries com o mesmo número de observações para estimação e previsão dos dados de retorno.

Resultados das simulações indicam que o modelo SV apresenta consistentemente o melhor desempenho quanto ao ajuste e previsão, ficando atrás apenas do APARCH na avaliação do VaR um passo à frente. Conclusões para o EGARCH e o NGSSM são mistas: quanto ao ajuste, o APARCH fica em segundo, o NGSSM em terceiro e o EGARCH em último; quanto à previsão, o EGARCH fica em segundo, o APARCH em terceiro e o NGSSM em último; quanto ao VaR, o APARCH fica em primeiro, o EGARCH em terceiro e o NGSSM em último. O tempo de CPU gasto na estimação de cada modelo também é reportado e comparado: tomando o NGSSM como base, a estimação do modelo SV demora 82 vezes mais, enquanto a estimação do APARCH demora 4 vezes mais e o EGARCH 2 vezes mais.

Palavras-chave: Heterocedasticidade Condicional, Modelos de Espaços de Estados Não-Gaussianos, Distribuição *Skew Generalized Error*, Distribuição *Asymmetric Exponential Power*

ABSTRACT

In this work a comparison of three families of volatility models, namely the Autoregressive Conditional Heteroskedasticity (ARCH), Stochastic Volatility (SV) and Non-Gaussian State Space Models (NGSSM) is made according to three different metrics: goodness of fit, forecasting and assessing Value-at-Risk (VaR). Inference procedures under the flexible Skew Generalized Error family of distributions is detailed. Respective evaluation criteria used for these metrics are the Akaike Information Criterion, Mean Squared Error of one-step-ahead forecasts and Unconditional Coverage of one-step-ahead VaR. The data used are daily asset return series (Ibovespa, Hang Seng Index, Merval Index and S&PTSX Index) from Jan-2000 to Jan-2016, or roughly 4000 observations, from which 3000 are used for estimation and 1000 are reserved for forecasting and VaR evaluation. Parameter estimates serve as basis to conduct a simulation experiment which consists of 1000 replications of series with the same number of observations for estimation and forecasting as the return data.

Simulation results indicate that the Stochastic Volatility model consistently outperforms competing specifications in goodness of fit and forecasting, and ranks second (right after the APARCH) in assessing the out-of-sample VaR. Conclusions for the EGARCH and NGSSM are mixed: in goodness of fit performance, the APARCH ranks second, the NGSSM ranks third and the EGARCH ranks last; in forecasting performance, the EGARCH is second, the APARCH third and the NGSSM last; in VaR assessment, the APARCH ranks first, the EGARCH third and the NGSSM last. CPU time spent on the estimation of each model is also reported and compared: taking the NGSSM as the benchmark, estimation of the SV model takes about 82 times as long, while APARCH estimation takes about 4 times and EGARCH estimation about 2 times.

Keywords: Conditional Heteroskedasticity, Non-Gaussian State Space Models, Skew Generalized Error Distribution, Asymmetric Exponential Power Distribution

INDEX

1. INTRODUCTION

Volatility plays a key role in Finance, representing the risk of an asset. It is therefore the basis for risk management, portfolio optimization, assessing the Value-at-Risk, and the pricing of options, futures and derivatives. Volatility is also important in Economics, where risk-averse agents will require premia for more volatile operations, and where it can also be used as a measure of how agents' decisions and preferences change over time, or even to assess the non-constant variability of economic variables such as output, money supply and inflation.

Performing statistical inference for the volatility is an issue that has received special attention. Since volatility is non-observable, estimating it requires a specific set of techniques. Engle (1982) proposed a conditioning argument that solves the problem for deterministic volatility dynamics. Specifically, in his Autoregressive Conditional Heteroskedasticity (ARCH) model, volatility is a function of past squared values of the original time series. Therefore, by conditioning on the immediate past information set, volatility is effectively an observable quantity. Later, Bollerslev (1986) proposed the Generalized ARCH (GARCH) model, in which the volatility is also allowed to depend on its own past, analogously to an Autoregressive Moving Average (ARMA) model. However, the GARCH specification is limited; according to Carnero et al. (2004), GARCH models require additional distribution assumptions and parameter values that make the model close to nonstationary in order to reproduce the behavior in daily series of financial returns.

An alternative to ARCH-type models is the Stochastic Volatility (SV) family of models, of which the first was proposed by Taylor (1982) as a first-order stochastic autoregressive process for the volatility. SV models offer a natural economic interpretation of volatility and are easier to connect with continuous-time diffusion models, which are often used in financial theory to represent the behaviour of financial returns. They are also found to be more flexible than ARCH-type models; Carnero et al. (2004) show that the basic SV model is more appropriate than the GARCH model in reproducing main empirical properties of daily returns. Another advantage is that statistical properties of SV models are simpler to derive when compared to models from the ARCH family, using elementary properties of stochastic processes.

However, the Stochastic Volatility family has a serious drawback: estimation techniques for its models are much more complicated than for ARCH models, even for the canonical autoregressive model with Gaussian innovations of Taylor (1982). There is a whole body of literature dedicated to the estimation of SV models, which is reviewed extensively in the paper by Broto and Ruiz (2004) and in the book by Bauwens et al. (2012).

An interesting alternative to both SV and ARCH-type models is the Non-Gaussian State Space Models (NGSSM) family proposed by Gamerman et al. (2013). The NGSSM in the volatility context is essentially a local scale model (a multiplicative local level model) where the dynamic level has a Beta evolution. This evolution may seem restrictive at first, but it allows for exact likelihood inference, filtering and smoothing; furthermore, observations are allowed to follow a whole plethora of distributions as long as they can be written in a specific form. Examples of distributions nested within the NGSSM family include the Normal, Laplace, Rayleigh, Poisson, Weibull and Generalized Gamma, as well as the heavy-tailed distributions included in the extension by Pinho et al. (2016), such as the Frechet, Levy, Log-gamma, Log-normal and Skew-Generalized Error Distribution (Skew-GED).

Since it is a state space model, the NGSSM allows the volatility to have its own stochastic process, as in the SV family. Furthermore, likelihood inference for these models is straightforward,

since the marginal likelihood of hyperparameters is available in closed form. Therefore, the NGSSM family seems to capture advantages of both SV and ARCH models: it allows for flexible specifications, but it is also computationally simple. Furthermore, when compared to the traditional lognormal SV model by Taylor (1982) and the GARCH model by Bollerslev (1986), both of which have 3 parameters, the corresponding Gaussian NGSSM only has 1 parameter to be estimated.

However, due to the fact that the NGSSM family is relatively recent, there have been few comparison studies between these three families. The works of Pinho and Santos (2013) and Pinho et al. (2016) suggest that the NGSSM family performs better than the ARCH and even than the SV model for the series taken into account. In detail, Pinho and Santos (2013) compare the fit of NGSSM (assuming various distributions) and the Asymmetric Power ARCH (APARCH) model of Ding et al. (1993) for series of daily returns of financial indexes. Their conclusion is that NGSSM outperforms APARCH when goodness-of-fit is evaluated using both Akaike and Bayesian Information Criteria (AIC and BIC, respectively) and loglikelihood values.

The work of Pinho et al. (2016) proceeds further comparing fit and forecasting performance of NGSSM, GARCH, Exponential GARCH (EGARCH) [Nelson (1991)] and log-t Stochastic Volatility also for series of daily returns of financial indexes. They conclude that the NGSSM outperforms the GARCH, EGARCH and log-t SV in fit by means of the AIC, BIC and loglikelihood and also in forecasting by means of the Square Root of Mean Squared Error (SQRMSE), calculated for 5 pseudo-out-of-sample one-step ahead forecasts.

Most of the literature on volatility models consists of new proposals along with a limited comparison between the new model and a few already established others. Examples include Chan and Gray (2006), which introduce the AR-GARCH-EVT model and compares it with parametric and nonparametric VaR approaches in a forecasting and conditional/unconditional coverage context using electricity return data; Omori et al. (2007), which extend the MCMC-based estimation approach proposed by Kim et al. (1998) to allow for a leverage effect and compares it to competing SV specifications on the basis of the marginal likelihood using japanese stock return data and Deschamps (2011), which introduces a new version of the local scale model of Shephard (1994a) and compares it with t-GARCH and lognormal SV models by means of Bayes factors using exchange rate and stock return data.

However, some papers that exclusively concern themselves with comparing models from different families can also be found. In a forecasting context, examples include Hansen and Lunde (2005), which compare a variety of ARCH-type models by means of tests for Superior Predictive Ability and Reality Check for data snooping using daily exchange rate and intraday stock return data and Iltuzer and Tas (2013), which compare naive volatility estimation approaches with ARCH-type and SV models on the basis of Superior Predictive Ability, Reality Check and Model Confidence Set in different forecast horizons using stock return data.

In a Value-at-Risk context, examples include So and Yu (2006), which compare ARCH-type models by assessing VaR estimation accuracy at various confidence levels on long and short positions using stock return and exchange rate data and Angelidis et al. (2004), which also compare ARCH-type models for stock return data, but assessing one-step-ahead VaR forecasts.

Finally, in the goodness of fit context, examples include Nakajima (2012), which compares SV and ARCH models by means of the marginal likelihood using daily individual securities, stock return and exchange rate data and Silva et al. (2015), which compares ARCH models on the basis of the Akaike Information Criterion using daily stock return data.

A thorough review of the current state of the literature on volatility model comparison does

not suggest a clear prevalence of any family of models, even if the interest is limited to one specific criterion. This work is expected to fill in this suggested gap in the literature, providing relevant empirical results and an extensive simulation experiment specifically designed for the comparison of the most utilized volatility models in practice so far. Another important point of interest to the applied use of these models is a comparison of computational time, especially in financial markets, where the number of assets in a portfolio tend to be very large.

The objectives of this work are summarized below, in no particular order.

1. Provide an accessible reference for the properties and inference techniques for the families of models presented here, specifically when a skewed and leptokurtic distribution (such as the Skew-GED) is assumed for the error terms.

2. Determine a family of models as being the most adequate for each metric: goodness of fit, forecasting and Value-at-Risk.

3. Draw conclusions about which features/stylized facts influence model performance the most, overall and for each criteria.

4. Establish a trade-off between accuracy and computational efficiency between models.

The next section details the relevant volatility families and their respective inference procedures, as well as important stylized facts of financial data and respective model evaluation criteria.

2. VOLATILITY MODELS

Consider a stochastic process $\{X_t\}_{t=0}^{\infty}$ with conditional mean $\mathbb{E}(X_t|\Psi_{t-1})$, in which $\mathbb{E}$ denotes the expectation operator and $\Psi_t = (X_0, x_1, \ldots, x_t)'$ is the information set of X_t, with X_0 denoting previously available information about the process and x_t denoting a realization of X_t. Volatility models are commonly written as a product of two independent stochastic processes, such as

$$\begin{aligned} &x_t - \mathbb{E}(X_t|\Psi_{t-1}) = y_t = \sigma_t \epsilon_t, && \epsilon_t \sim (0,1), \\ &\sigma_t = \sigma(\mathcal{F}_{t-1}) \end{aligned} \tag{1}$$

for $t = 1, \ldots, n$, where y_t is a realization of $\{Y_t\}_{t=0}^{\infty}$ and ϵ_t is a white noise. The volatility $\sigma_t > 0$ can be represented by any measurable positive function of the sigma-algebra generated by Ψ_{t-1}, denoted by $\mathcal{F}_{t-1}$. This set includes not only past values of X_t but also those of Y_t and past volatility values.

The volatility σ_t rescales the conditional distribution of y_t for each time t while allowing for an underlying constant scale $\mathbb{E}[\sigma_t] = \sigma_*$, which is assumed finite and constant over time. That is, the law of y_t obeys $\mathbb{V}[y_t|\psi_t] = \sigma_t^2$ and $\mathbb{V}[y_t] = \sigma_*^2$, where $\mathbb{V}$ denotes the variance operator, $\psi_t = (Y_0, y_1, \ldots, y_t)'$ denotes the information set of Y_t and Y_0 denotes previously available information. Under these assumptions, a non-constant conditional variance is consistent with first and second-order stationarity of Y_t as is a non-constant conditional mean. Note that although $y_t|\psi_{t-1}$ is serially uncorrelated, it is not serially independent since its variance is a function of the past; this is an important point and plays a crucial role in the identification of volatility models.

If a time series exhibits non-constant conditional variance, it is said to be conditionally heteroskedastic. This type of heteroskedasticity induces unconditional but not necessarily conditional leptokurtosis. For example, if we assume ϵ_t has a mesokurtic distribution (such as the Gaussian),

the conditional kurtosis equals 3, but the unconditional kurtosis is $\mathbb{K}[y_t] = \frac{3\,\mathbb{E}[\sigma_t^4]}{(\mathbb{E}[\sigma_t^2])^2}$ which is greater than or equal to 3 by Jensen's inequality. In essence, this means that if volatility is time-varying, the unconditional distribution of y_t will have higher probability for outliers, even if the distribution of ϵ_t does not.

When the error term is also leptokurtic, the unconditional tails of y_t will be even thicker. This behavior is relevant in practice, since volatility models are usually employed in data exhibiting not only leptorkutic but also skewed behavior [see. Lambert and Laurent (2002)]. In this case, the assumption of normality might be too restrictive; it can be relaxed by instead assuming that the error term follows a Skew Generalized Error Distribution (or Asymmetric Power Exponential Distribution). Reparameterized to have zero mean and unit variance, the density function of a Skew-GED variate is

$$f(x) = \frac{\nu}{\tau\Gamma(1/\nu)}\frac{\kappa}{1+\kappa^2}\exp\left\{-\left[\left(\frac{\kappa(x-\pi)^+}{\tau}\right)^\nu + \left(\frac{(x-\pi)^-}{\tau\kappa}\right)^\nu\right]\right\},\ x\in\mathbb{R},$$

where $\nu > 0$ is a shape/tail thickness parameter, $\kappa > 0$ is an asymmetry parameter, $\tau = \left[\frac{\Gamma(3/\nu)}{\Gamma(1/\nu)}\frac{1+\kappa^6}{\kappa^2(1+\kappa^2)} - \frac{\Gamma^2(2/\nu)}{\Gamma^2(1/\nu)}\frac{(1-\kappa^2)^2}{\kappa^2}\right]^{-1/2} > 0$ is a scale parameter, $\pi = -\tau\left(\frac{1}{\kappa}-\kappa\right) \in \mathbb{R}$ is a location parameter, $x^+ = x\mathbb{I}\{x \geq 0\}$ and $x^- = -x\mathbb{I}\{x \leq 0\}$, with $\mathbb{I}\{x \in A\}$ denoting the indicator function of x in set A.

We denote a normalized Skew-GED random variable by $X \sim \text{SGED}(\kappa, \nu)$. This distribution includes several others such as the Gaussian and Laplace - as well as their skewed versions - as special cases; see Table 1 for details. The parameterization of the Skew-GED used here is due Ayebo and Kozubowski (2003).

If $\epsilon_t \sim \text{SGED}(\kappa, \nu)$, applying the Jacobian transformation to $\epsilon_t = y_t/\sigma_t$ gives the density of y_t,

$$\frac{1}{\sigma_t}f\left(\frac{y_t}{\sigma_t}\right) = \frac{1}{\sigma_t}\frac{\nu}{\tau\Gamma(1/\nu)}\frac{\kappa}{1+\kappa^2}\exp\left\{-\frac{1}{\sigma_t^\nu}\left[\left(\frac{\kappa(y_t-\pi)^+}{\tau}\right)^\nu + \left(\frac{(y_t-\pi)^-}{\tau\kappa}\right)^\nu\right]\right\},\ y_t\in\mathbb{R}. \quad (2)$$

Now, in order to simplify notation, define Y_t and S_t to be the row vectors of realizations and volatilities up to time t, i.e. $Y_t = (Y_0, y_1, \ldots, y_t)'$ and $S_t = (\sigma_1, \ldots, \sigma_t)'$. The joint loglikelihood of (Y_n, S_n) of a sample of size n is

$$\log\mathbb{L}(\varphi|Y_n, S_n) = \sum_{t=1}^{n}\left\{-\log\sigma_t + \log\frac{\nu}{\tau\Gamma(1/\nu)}\frac{\kappa}{1+\kappa^2} - \frac{1}{\sigma_t^\nu}\left[\left(\frac{\kappa(y_t-\pi)^+}{\tau}\right)^\nu + \left(\frac{(y_t-\pi)^-}{\tau\kappa}\right)^\nu\right]\right\}, \quad (3)$$

where φ is a $p \times 1$ vector of parameters which includes κ, ν and hyperparameters of the volatility.

Since performing an inverse probability transformation on a Skew-GED variate is not possible, generating draws from this distribution might seem difficult. However, Ayebo and Kozubowski (2003) exploit the relationship between the Skew-GED and the Gamma distribution to derive a simple pseudorandom number generator algorithm, stated in Algorithm 1 for convenience.

Although the probability density function of a Skew-GED variate is available in analytical form, its distribution and quantile functions are not; they depend respectively on the incomplete Gamma function and its inverse, both of which must be evaluated numerically. A simple solution to this problem is to use Algorithm 1 to draw samples from the Skew-GED distribution and then calculate its empirical distribution function and quantiles. A simulation size of 10000 draws seems to be enough for most practical applications.

Algorithm 1 Skew-GED random number generator.

draw $G \sim \text{Gamma}(1/\nu, 1)$
draw $U \sim \text{Uniform}(0, 1)$
if $U < \frac{\kappa^2}{1+\kappa^2}$ **then**
    set $I = \frac{1}{\kappa}$
else
    set $I = -\kappa$
end if
return $X = \pi + \tau I G^{1/\nu} \sim \text{SGED}(\kappa, \nu)$

Table 1. Special cases of the skew-GED distribution.

Distribution	ν	κ	Kurtosis	Skewness
skew-GED	any	any	any	skewed
GED	any	$= 1$	any	symmetric
skew-Normal	$= 2$	any	mesokurtic	skewed
Normal	$= 2$	$= 1$	mesokurtic	symmetric
skew-Laplace	$= 1$	any	leptokurtic	skewed
Laplace	$= 1$	$= 1$	leptokurtic	symmetric
Uniform	$\to \infty$	any	platykurtic	symmetric

A vast body of research on volatility concerns financial time series, specially daily asset return data. It might be therefore relevant to study their behavior beforehand when considering models for the volatility. Stylized facts found in financial data usually serve as a starting point for proposing model extensions. The pioneer work of Mandelbrot (1963) is perhaps the most important example; he noted that large (small) returns are followed by large (small) returns, giving rise to temporal clusters in their variability. He named this behavior "volatility clustering", and it essentially refers to the fact that there is an autoregressive dependence in the volatility, which is a fundamental property of volatility models. Another observation made by the same author is that the distribution of returns are usually leptokurtic, being more propense to outliers than a Gaussian distribution.

Other important stylized facts about financial returns are the leverage effect and long memory in the volatility. The leverage effect was first discovered by Black (1976), which found that volatility responds asymmetrically to negative and positive returns of the same magnitude. This finding is in accordance with the financial theory that a decrease in the price of an asset leads to an increase in its debt/equity ratio (financial leverage) and therefore in its volatility (financial risk), in addition to the increase in the risk that occurs due to the (absolute) variation of returns. The presence of long memory dependence in the volatility was first noted by Ding et al. (1993), which found that the daily absolute returns (a common proxy for the volatility) of the S&P500 presented positive autocorrelations of lag up to and above the order of 2500 and proceeded to propose a model in order to capture this and a myriad other stylized facts.

2.1. Model Evaluation Criteria

The most straightforward way to compare the goodness of fit between volatility models is to use the joint loglikelihood function (3). However, this is not adequate since model complexity varies between specifications; rather, an information criterion, which is a loss function of both the

likelihood and the number of estimated parameters should be used. The one adopted in this work is the Akaike information criterion (AIC), due Akaike (1974).

The AIC expresses the information lost - measured by the Kullback-Leibler divergence - when approximating the true data generating process by an estimated model. Therefore, minimizing AIC is equivalent to obtaining the best fit. Scaled for sample size, its expression is

$$\mathrm{AIC}(p) = -\frac{2}{n}\bigg(\log \mathbb{L}(\varphi|Y_n, S_n) - p\bigg),$$

where n is the sample size, $\mathbb{L}(\varphi|Y_n, S_n)$ is the joint likelihood in (3) and p is the number of model parameters.

Evaluating the goodness of fit of a model is referred to as in-sample evaluation, reflecting the fact that only available information in the moment of estimation is used. However, when there is an interest in predicting future values, out-of-sample evaluation techniques should be used; they are also named "pseudo-out-of-sample" in order to reflect the fact that some observations are treated as unknown for model estimation but are subsequently used for evaluation.

The out-of-sample comparisons in this work are done in two separate contexts: forecasting volatility and Value-at-Risk. Forecasts in general are obtained as a minimization of a loss function, which penalizes deviations from the true value. A common loss function is the quadratic, expressed as $L[\sigma_t(h)] = \mathbb{E}\{[\sigma_t(h) - \sigma_{t+h}]^2|\psi_t\}$ where $\sigma_t(h)$ is the h-step ahead forecast. It is a well-known result [Hamilton (1994)] that $L[\sigma_t(h)]$ is minimal at $\tilde{\sigma}_t(h) = \mathbb{E}(\sigma_{t+h}|\psi_t)$, the conditional expectation of σ_{t+h} over the information set ψ_t and henceforth denoted by $\sigma_{t+h|t}$.

Forecasts of an estimated model are obtained by straightforward substitution of the populational parameter vector φ by its estimate $\hat{\varphi}$ in the forecast equation $\sigma_{t+h|t}$; the estimated forecast is denoted by $\hat{\sigma}_{t+h|t}$. The quadratic loss function above can then be used to compare the overall forecasting performance across models; its sample counterpart, known as mean squared error (MSE), is expressed by

$$\mathrm{MSE}(k) = \frac{1}{k}\sum_{t=n}^{n+k}(\hat{\sigma}_{t+h|t} - \check{\sigma}_t)^2$$

where k is the number of performed forecasts, n is the sample size used for estimation and $\check{\sigma}_t$ is the true volatility. Usually in practice the true volatility is not available, and must be replaced by a proxy; the absolute demeaned returns is fairly adequate for this purpose, as it has the same unconditional expectation as the volatility.

The Value-at-Risk (VaR) is a very useful tool in risk management. It is defined as the loss corresponding to the α%th percentile of the distribution of returns over the next N days. In other words, it measures the loss over the next N days that is exceeded only α% of the time. The N-day h-step-ahead out-of-sample VaR is expressed as

$$\mathrm{VaR}_{t+h|t}(N, \alpha) = -\sqrt{N}\hat{\sigma}_{t+h|t}\mathrm{SGED}_\alpha(\kappa, \nu),$$

where $\mathrm{SGED}_\alpha(\kappa, \nu)$ denotes the α% quantile of the Skew-GED distribution. As stated before, it is a good idea to calculate this quantile by taking the empirical quantile of a pseudorandom sample drawn using Algorithm 1. Mathematically, the predicted VaR is essentially the α% quantile of the distribution of returns, scaled by the forecasted volatility and the square root of the number of days. Since the VaR expresses a positive loss, the negative of this quantile is used.

A simple method to evaluate out-of-sample Value-at-Risk is the unconditional VaR coverage,

which is based on Christoffersen (1998). First consider the following indicator function, also known as hit function,

$$\mathbb{H}_{t+h|t}(N,\alpha) = \begin{cases} 1, & y_{t+h} \leq -\text{VaR}_{t+h|t}(N,\alpha) \\ 0, & y_{t+h} > -\text{VaR}_{t+h|t}(N,\alpha) \end{cases}$$

defined for $t = n, \ldots, n+k$. If the predicted VaR is correct, it is expected that $\mathbb{P}[\mathbb{H}_{t+h|t}(N,\alpha) = 1] = \alpha$. Therefore, the quantity

$$\text{UC}(k) = |\alpha - \hat{\alpha}|,$$

where $\hat{\alpha} = \frac{1}{k}\sum_{t=n}^{n+k} \mathbb{H}_{t+h|t}(N,\alpha)$ is the estimated probability that the loss exceeds the predicted VaR, can be used to assess peformance between competing models. Although the sign of the difference $(\alpha - \hat{\alpha})$ is usually ignored, it has a financial interpretation: if the unconditional coverage is positive (negative), the VaR is said to be conservative (risky), since the loss is being underestimated (overestimated).

In the remainder of this section, we present the three volatility families used in this work.

2.2. Autoregressive Conditional Heteroskedasticity

Introduced in Engle (1982), the Autoregressive Conditional Heteroskedasticity (ARCH) is a model for the square of the volatility, which is a function of past squared returns. The main assumption made in the ARCH family is that by conditioning on the information set ψ_{t-1}, volatility at time t is an observable volatility. This essentially means that, once the past of a time series is known, its next-period volatility is deterministic. ARCH-type models are also commonly referred to as conditional volatility models.

The canonical model in the ARCH family is the Generalized ARCH (GARCH) proposed in Bollerslev (1986), which is an extension of the original ARCH model to allow for the squared volatility to also depend on its past values. While the original ARCH model allows for an autoregressive representation in the squared returns, the GARCH allows for an autoregressive moving-average representation.

Two specifications for the volatility in the ARCH family are considered here: the Asymmetric Power ARCH (APARCH) of Ding et al. (1993) and Exponential Generalized ARCH (EGARCH) of Nelson (1991). The APARCH nests at least 9 other popular ARCH models (see Table 2) and the EGARCH is closely related to the Stochastic Volatility model of the next subsection. Since most of the literature on ARCH models consider only one-period (Markovian) dependence on the volatility, that is the case which is presented here.

The APARCH is defined as

$$\begin{aligned} y_t &= \sigma_t \epsilon_t = (\sigma_t^\delta)^{1/\delta}\epsilon_t, & \epsilon_t &\sim \text{SGED}(\kappa,\nu) \\ \sigma_t^\delta &= \omega + \alpha(|y_{t-1}| - \gamma y_{t-1})^\delta + \beta\sigma_{t-1}^\delta, \end{aligned} \tag{4}$$

for $t = 1, \ldots, n$, where $\omega > 0$ is a constant, $\alpha \geq 0$ is an autoregressive parameter, $-1 < \gamma < 1$ is a leverage effect parameter, $\beta \geq 0$ is a moving average parameter and $\delta \geq 0$ is a power transformation parameter. These parameter constraints are necessary only to ensure that σ_t^δ is positive.

In regard to stylized facts, the APARCH is capable of not only reproducing volatility clustering and leptokurticity, but also the leverage effect and even a limited form of long memory. To see

how the parameter γ captures leverage, let $\beta = 0$ and $\delta = 2$. Then,

$$\sigma_t^\delta = \begin{cases} \omega + \alpha(1-\gamma)^2 y_{t-1}^2, & y_{t-1} \geq 0, \\ \omega + \alpha(1+\gamma)^2 y_{t-1}^2, & y_{t-1} \leq 0. \end{cases}$$

Therefore, if $\gamma > 0$, negative returns increase the volatility more than positive returns of the same magnitude, which is precisely what the definition of leverage requires.

A last stylized fact captured by the APARCH is long memory in the volatility. Although the precise definition of long memory within volatility models is given in Baillie et al. (1996) as an analogue of Autoregressive Fractionally Integrated Moving Average (ARFIMA) models, Ding et al. (1993) adopt the long memory definition of a slower/hyperbolic decay of autocorrelations, and show that the APARCH is capable of reproducing that behavior for certain values of δ. This parameter allows for greater flexibility within the volatility specification compared to competing ARCH models, since it relaxes the usual assumption that volatility must be expressed either as a conditional standard deviation ($\delta = 1$) or variance ($\delta = 2$).

One-step-ahead predictions under the APARCH model are given by

$$\sigma_{t+1|t}^\delta = \omega + \alpha(|y_t| - \gamma y_t)^\delta + \beta \sigma_t^\delta. \tag{5}$$

The EGARCH is defined as

$$\begin{aligned} y_t &= \sigma_t \epsilon_t = \exp(0.50 \log \sigma_t^2)\epsilon_t, & \epsilon_t &\sim \text{SGED}(\kappa, \nu) \\ \log \sigma_t^2 &= \omega + \theta \epsilon_{t-1} + \gamma(|\epsilon_{t-1}| - \mathbb{E}|\epsilon_{t-1}|) + \beta \log \sigma_{t-1}^2, \end{aligned} \tag{6}$$

for $t = 1, \ldots, n$, where ω is a constant, θ is a leverage effect parameter, γ is a magnitude change parameter, β is a moving average parameter and $\mathbb{E}|\epsilon_t| = \frac{1}{\Gamma(1/\nu)} \frac{\kappa}{1+\kappa^2} \left[\frac{\tau \Gamma(2/\alpha)(1+\kappa^4)}{\kappa^2} + \frac{\pi \Gamma(1/\alpha)(1-\kappa^2)}{\kappa} \right]$ is the expectation of the absolute value of ϵ_t under the Skew-GED distribution.

The logarithmic transformation in the EGARCH ensures that the volatility σ_t is always positive, and therefore there are no positivity constraints on the parameters. However, in order for the model to properly reproduce the volatility clustering property, it is required that $-\gamma < \theta < \gamma$ and $\beta \geq 0$. Under these conditions, large (small) innovations will increase (decrease) volatility, as the definition of clustering requires.

Regarding other stylized facts, the EGARCH is capable of capturing the leverage effect through the parameter θ. If $\beta = 0$ and $\gamma = 0$, $\log \sigma_t^2 = \omega + \theta \epsilon_{t-1}$. That is, for $\theta < 0$, $\log \sigma_t^2$ is larger (smaller) than its mean if ϵ_{t-1} (and thefore y_{t-1}) is negative (positive). The behavior captured by the parameter γ is also noteworthy: provided that $\gamma > 0$, innovations larger (smaller) than their expectation increase (decrease) volatility.

One-step-ahead forecasts under the EGARCH model are given by

$$\log \sigma_{t+1|t}^2 = \omega + \theta \epsilon_t + \gamma(|\epsilon_t| - \mathbb{E}|\epsilon_t|) + \beta \log \sigma_t^2. \tag{7}$$

Estimation of ARCH-type models by maximum likelihood is rather straightforward, and proceeds as follows: given a sample $Y_n = (Y_0, y_1, \ldots, y_n)'$, set quantities at $t = 0$ at their unconditional expectations and write the volatility recursively for $t = 1, \ldots, n$ using (4) for the APARCH and (6) for the EGARCH. An explicit expression for the joint loglikelihood in (3) as a function of the parameter vector φ is then obtained, and it can be maximized using a numerical optimization

Table 2. Special cases of the APARCH model.

Model	δ	$\gamma(1)$	$\beta(L)$	Author
APARCH	any	any	any	Ding et al. (1993)
NARCH	any	$= 0$	$= 0$	Higgins and Bera (1992)
AGARCH	$= 2$	any	any	Meitz and Saikkonen (2011)
GJR-GARCH	$= 2$	any	any	Glosten et al. (1993)
GARCH	$= 2$	$= 0$	any	Bollerslev (1986)
ARCH	$= 2$	$= 0$	$= 0$	Engle (1982)
TGARCH	$= 1$	any	any	Zakoian (1994)
TARCH	$= 1$	any	$= 0$	Rabemananjara and Zakoian (1993)
Taylor/Schwert	$= 1$	$= 0$	any	Taylor (1986), Schwert (1990)
log-ARCH	$\to 0$	$= 0$	$= 0$	Geweke (1986), Pantula (1986)

algorithm such as the one proposed independently by Broyden (1970), Goldfarb (1970), Fletcher (1970) and Shanno (1970), henceforth denoted as BFGS. For the APARCH $\varphi = (\omega, \alpha, \beta, \gamma, \delta, \kappa, \nu)'$ and for the EGARCH $\varphi = (\omega, \theta, \gamma, \beta, \kappa, \nu)'$.

Due to the practice of setting unobservable quantities at $t = 0$ at their unconditional expectations, some authors refer to the likelihood maximization procedure described above as conditional (or approximate) maximum likelihood estimation. Having such a simple inference procedure is a major comparative advantage of ARCH models, and what makes them so relevant in practice.

2.3. Stochastic Volatility

In the Stochastic Volatility (SV) family, volatility is driven by its own stochastic process. Models in this family have often been used in mathematical Finance to reproduce the behavior of prices in the stock market. Although the model exact origins are somewhat uncertain, the first discrete-time version of the SV was proposed by Taylor (1982). In this canonical version, the log-squared volatility follows a first-order autoregressive process with Gaussian innovations; therefore, it is usually referred to as the lognormal SV model. A slight generalization of this model allowing for returns to be Skew-GED distributed can be defined as

$$\begin{aligned} y_t &= \sigma_t \epsilon_t = \exp(h_t/2)\epsilon_t, & \epsilon_t &\sim \text{SGED}(\kappa, \nu) \\ h_{t+1} &= \mu + \phi h_t + \sigma_\eta \eta_t, & \eta_t &\sim \text{Normal}(0, 1), \end{aligned} \tag{8}$$

for $t = 1, \ldots, n$, where $h_t = \log(\sigma_t^2)$ is the log-squared volatility, ϵ_t and η_t are serially and mutually independent, μ is a constant, ϕ is an autoregressive parameter and $\sigma_\eta \geq 0$ is a scale parameter. The model is initialized with $h_0 = 0$, i.e. $h_1 \sim \text{Normal}(\mu, \sigma_\eta^2)$.

As in the EGARCH, the logarithmic transformation ensures that the volatility is always positive for any parameter values. Furthermore, when $\sigma_\eta = 0$, the SV model is equivalent to the EGARCH with $\phi = \beta$ and $\theta = \gamma = 0$. Since it is essentially an AR(1) model for the log-squared volatility, properties of the SV are straightforward to derive.

The basic SV model presented in (8) is able to reproduce volatility clustering and leptokurticity, but not long memory or the leverage effect. The leverage effect extension of the lognormal SV was first proposed by Harvey and Shephard (1996), and consists of allowing the previous observation disturbance ϵ_{t-1} and the current state disturbance η_t to be correlated, by assuming that their joint distribution is multivariate normal with non-diagonal covariance matrix. The long memory

version of the lognormal SV was proposed independently by Harvey (1998) and Breidt et al. (1998), which instead of an AR(1) considered an ARFIMA(1, d, 0) process for the volatility. However, both generalizations require that observations be normally distributed; extending these models for Skew-GED distributed returns is out of the scope of this work and therefore they will not be considered here.

Statistical inference for Stochastic Volatility models is considerably more complex than it is for ARCH models. There have been a myriad techniques proposed to estimate the canon lognormal SV; see the excellent reviews on the subject by Broto and Ruiz (2004) and more recently by Bauwens et al. (2012). The main problem in estimating SV models is that volatility is a function not only of its past but also of the stochastic process η_t. Therefore, even after conditioning on past information, volatility is still an unobservable quantity. In order to illustrate that point further, consider the marginal likelihood of Y_n,

$$\mathbb{L}(\varphi|Y_n) = f(Y_n|\varphi) = \int_S f(Y_n, H_n|\varphi)dH_n = \int_S f(Y_n|H_n, \varphi)f(H_n|\varphi)dH_n, \tag{9}$$

where $H_n = (h_1, \ldots, h_n)'$ is the joint vector of log-squared volatilities and $S = (0, \infty)^n$ is its corresponding support.

While the expression for $f(Y_n|H_n, \varphi)$ is straightforward, the marginal distribution $f(H_n|\varphi)$ is not available in analytical form. Approximating this distribution - and therefore the joint likelihood - is the main estimation problem in the Stochastic Volatility family. An interesting solution is the importance sampling technique for non-Gaussian and nonlinear state space models proposed independently by Durbin and Koopman (1997) and Shephard and Pitt (1997) and which has been considerably improved upon in the textbook treatment given by Durbin and Koopman (2012). It is sometimes referred to as Monte Carlo Maximum Likelihood Estimation.

Although the ideas of importance sampling are relatively simple, the estimation process is considerably more complex. First, consider the following linear Gaussian model,

$$\begin{aligned} x_t &= h_t + \epsilon_t, & \epsilon_t &\sim \text{Normal}(0, A_t), \\ h_{t+1} &= \mu + \phi h_t + \sigma_\eta \eta_t, & \eta_t &\sim \text{Normal}(0, 1), \end{aligned} \tag{10}$$

for $t = 1, \ldots, n$. The model is initialized with $h_0 = 0$, i.e. $h_1 \sim \text{Normal}(\mu, \sigma_\eta^2)$. Denote by g the densities associated with this linear state space model, and notice that the state equation in (10) is the same as that of the Stochastic Volatility model (8), implying that $f(H_n) = g(H_n)$.

Now, using the fact that $\frac{f(H_n,Y_n)}{g(H_n,Y_n)} = \frac{f(Y_n|H_n)f(H_n)}{g(Y_n|H_n)g(H_n)} = \frac{f(Y_n|H_n)}{g(Y_n|H_n)}$, rewrite the joint likelihood in (9) as

$$\mathbb{L}(\varphi|Y_n) = \int_S f(H_n, Y_n)dH_n = g(Y_n)\int_S \frac{f(H_n, Y_n)}{g(H_n, Y_n)}g(H_n|Y_n)dH_n = \mathbb{L}_g(\varphi|Y_n)\mathbb{E}_g\left[\frac{f(Y_n|H_n)}{g(Y_n|H_n)}\right], \tag{11}$$

where $g(H_n|Y_n)$ is the smoothed Gaussian density, $\mathbb{L}_g(\varphi|Y_n) = g(Y_n)$ is the marginal Gaussian likelihood and $\mathbb{E}_g$ denotes expectation with respect to $g(H_n|Y_n)$. The dependence of f and g on φ was supressed to simplify notation.

Expression (11) was first proposed by Durbin and Koopman (1997) and essentially defines the joint non-Gaussian likelihood (9) as an adjustment to a simple Gaussian density. This adjusment term is readily estimable by importance sampling by taking $g(H_n|Y_n)$ as the importance density. As will be made clear in what follows, drawing from this distribution is relatively straightforward,

and therefore expression (11) is easily manageable.

In order to completely determine $g(H_n|Y_n)$ one must first appropriately specify x_t and A_t for $t = 1, \ldots, n$. Examination of the non-Gaussian likelihood (11) suggests that $g(Y_n|H_n)$ should be as close as possible to $f(Y_n|H_n)$ in a neighborhood of their mode. Durbin and Koopman (1997) suggest using a Laplace approximation of the log-ratio of these densities to determine x_t, A_t and the mode vector of log-squared volatilites H_n, denoted by $\hat{H}_n$. Conditional independence of (H_n, Y_n) under both Gaussian and non-Gaussian densities implies that $\frac{f(Y_n|H_n)}{g(Y_n|H_n)} = \prod_{t=1}^{n} \frac{f(y_t|h_t)}{g(y_t|h_t)}$; therefore, the log-ratio between these two densities is given by

$$l(H_n) = \sum_{t=1}^{n} \left[\log f(y_t|h_t) - \log g(y_t|h_t)\right],$$

where the dependence of l on Y_n was dropped since it is assumed to be fixed when determining the mode.

It follows that x_t and A_t are obtained as the solutions of $\partial l(h_t)/\partial h_t = 0$ and $\partial^2 l(h_t)/\partial h_t^2 = 0$ at $h_t = \hat{h}_t$, for $t = 1, \ldots, n$. That is,

$$A_t = \frac{4}{\nu^2} \exp\left(\frac{\nu \hat{h}_t}{2}\right) \left[\left(\frac{\kappa(y_t - \pi)^+}{\tau}\right)^{\nu} + \left(\frac{(y_t - \pi)^-}{\kappa\tau}\right)^{\nu}\right]^{-1} \quad \text{and} \quad x_t = \hat{h}_t - \frac{1}{2}A_t + \frac{2}{\nu}. \tag{12}$$

However, the expressions obtained for x_t and A_t are still functions of the unknown mode $\hat{h}_t$. Therefore, a Newton-Raphson procedure must be employed to iteratively solve for this mode. Exploiting the linear and Gaussian structure of $g(H_n|Y_n)$ allows the use of the Kalman filter and smoother to calculate the next guess of the iterative procedure; this is computationally more efficient than using the Newton-Raphson algorithm directly. Note that this is only possible due to equality between mean and mode under the Normal distribution.

Denote by $a_{t|t-1} = \mathbb{E}[h_t|Y_{t-1}]$ and $P_{t|t-1} = \mathbb{V}[h_t|Y_{t-1}]$ the filtered estimate of h_t and its respective variance, by $a_{t|t} = \mathbb{E}[h_t|Y_t]$ and $P_{t|t} = \mathbb{V}[h_t|Y_t]$ the updated estimate of h_t and its respective variance and by $a_{t|n} = \mathbb{E}[h_t|Y_n]$ and $P_{t|n} = \mathbb{V}[h_t|Y_n]$ the smoothed estimate of h_t and its respective variance. The Kalman filter and smoother for the linear state space model (10) is given in Algorithm 2.

Algorithm 2 Kalman filter and smoother for the linear state space model (10).

1: initalize $a_{0|0} = \mu$ and $P_{0|0} = 10^7$
2: **for** $t = 1, \ldots, n$ **do**
3: calculate $a_{t|t-1} = \phi a_{t-1|t-1} + \mu$ and $P_{t|t-1} = \phi^2 P_{t-1|t-1} + \sigma_\eta^2$
4: calculate $F_t = P_{t|t-1} + A_t$, $a_{t|t} = a_{t|t-1} + P_{t|t-1}F_t^{-1}(x_t - a_{t|t-1})$ and $P_{t|t} = P_{t|t-1} - P_{t|t-1}F_t^{-1}P_{t|t-1}$
5: **end for**
6: **for** $t = n-1, \ldots, 1$ **do**
7: calculate $P_t^* = \phi P_{t|t} P_{t+1|t}$, $a_{t|n} = a_{t|t} + P_t^*(a_{t+1|n} - \phi a_{t|t})$ and $P_{t|n} = P_{t|t} + P_{t|t}^*(P_{t+1|n} - P_{t+1|t})P_{t|t}^*$.
8: **end for**
9: **return** $a_{t|t}$, $P_{t|t}$, $a_{t|n}$, $P_{t|n}$, $t = 1, \ldots, n$

To complete the iterative process to obtain the mode, the initialization conditions must be specified. Note that the observation equation in the Stochastic Volatility model (8) implies that $\mathbb{E}[y_t^2] = \mathbb{E}[h_t]$, since $\mathbb{E}[\epsilon_t^2] = 1$. Therefore, taking $h_t \approx \log y_t^2$ and substituting it for $\hat{h}_t$ in (12)

yields the initial values for x_t and A_t. The entire process of obtaining the mode is summarized in Algorithm 3.

Algorithm 3 Iterative process to approximate the mode of $f(H_n|Y_n)$.

1: initialize $A_t = \frac{4}{\nu^2}\exp\left(\frac{\nu \log y_t^2}{2}\right)\left[\left(\frac{\kappa(y_t-\pi)^+}{\tau}\right)^\nu + \left(\frac{(y_t-\pi)^-}{\kappa\tau}\right)^\nu\right]^{-1}$ and $x_t = \log y_t^2 - \frac{1}{2}A_t + \frac{2}{\nu}$
2: compute a first guess $\tilde{h}_t = a_{t|n}$, $t = 1, \ldots, n$ using Algorithm 2
3: **repeat**
4: compute the next guess $\tilde{h}_t^+ = a_{t|n}$, $t = 1, \ldots, n$ using Algorithm 2
5: calculate x_t and A_t by taking $\tilde{h}_t^+ = \hat{h}_t$ in (12)
6: **until** convergence
7: **return** $\hat{H}_n = (\hat{h}_1, \ldots, \hat{h}_n)'$

Convergence to the mode is usually attained with 10 iterations or less. After obtaining the mode, the importance density $g(H_n|Y_n)$ is completely determined and it is possible to draw from it using the simulation procedure of Shephard (1994b), summarized in Algorithm 4. It should be clear that the mode and random draws obtained from the importance density are conditional on $\varphi = (\mu, \phi, \sigma_\eta, \kappa, \nu)'$ known; the situation where φ is estimated is considered below.

Algorithm 4 Simulation procedure for drawing from $g(H_n|Y_n)$.

1: compute $a_{t|t}, a_{t|t-1}, P_t t|t-1$ and $P_{t|t}$, $t = 1, \ldots, n$ using steps 1-5 of Algorithm 2.
2: draw $h_n|Y_n \sim \text{Normal}(a_{n|n}, P_{n|n})$.
3: **for** $t = n-1, \ldots, 1$ **do**
4: draw $h_t|h_{t+1} \sim \text{Normal}\left\{a_{t|t} + \frac{\phi P_{t|t}(h_{t+1}-a_{t+1|t})}{P_{t+1|t}}, P_{t|t} - \frac{\phi^2 P_{t|t}^2}{P_{t+1|t}}\right\}$
5: **end for**
6: **return** $H_n = (h_1, \ldots, h_n)'$

When simulating from $g(H_n|Y_n)$, computational efficiency can be increased by employing antithetic variables. As Durbin and Koopman (2012) defines, an antithetic variable is a function of a random drawn of H_n which is equiprobable with H_n and which increases the efficiency of the estimation when included in the drawing process. The first antithetic variable used here is $\check{H}_n = 2\hat{H}_n - H_n$; since H_n is Gaussian with mean $\hat{H}_n$, it is straightforward to verify that $\check{H}_n$ has the same distribution as H_n. Whenever this antithetic variable is used, the simulation sample is said to be balanced for location.

The second antithetic variable used here was developed by Durbin and Koopman (1997). Let U_n be the vector of the n Normal$(0, 1)$ variables used in the simulation procedure to generate H_n and let $c = U_n'U_n$; then $c \sim \chi_n^2$, where χ_n^2 denotes a chi-squared distribution with n degrees of freedom. For a given value of c let $q = \mathbb{P}(\chi_n^2 < c) = F(c)$ be the distribution function of c and $\acute{c} = F^{-1}(1-q)$ be the quantile function of c; moreover, note that c and $\acute{c}$ have the same distribution. Noting that c and $(H_n - \hat{H}_n)/\sqrt{c}$ are independently distributed, two additional antithetic variables can be constructed: $\acute{H}_n = \hat{H}_n + \sqrt{\frac{\acute{c}}{c}}(H_n - \hat{H}_n)$ and $\grave{H}_n = \hat{H}_n + \sqrt{\frac{\acute{c}}{c}}(\check{H}_n - \hat{H}_n)$. When these two antithetics are used, the simulation sample is said to be balanced for scale. By using both types of antithetic variables a set of four equiprobable values of H_n are obtained for each run of the simulation procedure, yielding a simulation sample which is balanced for both location and scale.

The primary objective of importance sampling here is to estimate the unknown parameter vector φ. In order to accomplish that, it is first necessary to estimate the joint likelihood (11) by simulation and maximize it numerically with respect to φ. The estimate of the logarithm of (11)

is

$$\log \hat{\mathbb{L}}(\varphi|Y_n) = \log \mathbb{L}_g(\varphi|Y_n) + \log \bar{w} + \frac{S_w^2}{2N\bar{w}^2}, \tag{13}$$

where N is the number of simulations and $\bar{w} = (1/N)\sum_{i=1}^{N} w_i$, with $w_i = w(H_n^{(i)}, Y_n) = f(H_n^{(i)}, Y_n)/g(H_n^{(i)}, Y_n)$ denoting the importance weights and $H_n^{(i)} = (h_1^{(i)}, \ldots, h_n^{(i)})'$ denoting the vector of log-squared volatilities generated for $i = 1, \ldots, N$. The last term is a necessary bias correction when taking the logarithm of the likelihood, since $\mathbb{E}[\log \bar{w}] \neq \log \mathbb{E}_g[w_i]$ and its numerator is given by $S_w^2 = \frac{1}{N-1}\sum_{i=1}^{N}(w_i - \bar{w})^2$; see Durbin and Koopman (1997) for a detailed proof.

The Gaussian loglikelihood $\log \mathbb{L}_g(\varphi|Y_n)$ can be evaluated by the Kalman filter using the predicitve error decomposition. That is, after obtaining $a_{t|t}$ and F_t from steps 1-5 of Algorithm 2, compute

$$\log \mathbb{L}_g(\varphi|Y_n) = -\frac{1}{2}\sum_{t=1}^{n}\left[\log 2\pi + \log F_t + \frac{(x_t - a_t)^2}{F_t}\right]. \tag{14}$$

The only quantity remaining to be determined for the estimation process are the importance weights w_i. Since $f(y_t|h_t^{(i)})$ is given in (2) by taking $\sigma_t = \exp(h_t^{(i)}/2)$ and $g(y_t|h_t^{(i)}) = \text{Normal}(h_t^{(i)}, A_t)$,

$$w_i = \exp\left\{ -\frac{1}{2}\sum_{t=1}^{n}\left[h_t^{(i)} + \frac{2}{\exp(\nu h_t^{(i)}/2)}\left[\left(\frac{\kappa(y_t - \pi)^+}{\tau}\right)^{\nu} + \left(\frac{(y_t - \pi)^-}{\kappa\tau}\right)^{\nu}\right]\right.\right.$$
$$\left.\left. - 2\log\left[\frac{\nu}{\tau\Gamma(1/\nu)}\frac{\kappa}{1+\kappa^2}\right] - \log 2\pi - \log A_t - \left(\frac{x_t - h_t^{(i)}}{\sqrt{A_t}}\right)^2\right]\right\}. \tag{15}$$

After drawing N samples of H_n and calculating the loglikelihood estimate (13) it is possible to maximize it with respect to φ using an iterative maximization process such as the BFGS. Each iteration of the procedure is started with an initial guess of φ given by first maximizing the approximate joint loglikelihood

$$\log \hat{\mathbb{L}}(\varphi|Y_n) \approx \log \mathbb{L}_g(\varphi|Y_n) + \log w(\hat{H}_n), \tag{16}$$

where $w(\hat{H}_n) = f(\hat{H}_n, Y_n)/g(\hat{H}_n, Y_n)$.

The entire estimation procedure of the Stochastic Volatility model (8) using importance sampling can be summarized in Algorithm 5

Algorithm 5 Importance sampling estimation of φ.

1: initialize with a guess $\tilde{\varphi}$
2: **repeat**
3: compute $\hat{H}_n = (\hat{h}_1, \ldots, \hat{h}_n)'$ using Algorithm 3 conditional on $\tilde{\varphi}$.
4: compute an intermediate guess $\check{\varphi}$ by maximizing the approximate loglikelihood (16) using the BFGS numerical procedure conditional on $\tilde{\varphi}$
5: sample $H_n^{(i)}$, $i = 1, \ldots, N$ using Algorithm 4 and calculate the antithetic variables $\check{H}_n$, $\acute{H}_n$ and $\grave{H}_n$ conditional on $\check{\varphi}$
6: evaluate the Gaussian loglikelihood in (13) using steps 1-5 of Algorithm 2 conditional on $\check{\varphi}$
7: compute importance weights w_i, $i = 1, \ldots, N$ using (15) conditional on $\check{\varphi}$
8: estimate the joint loglikelihood (13) and compute a new guess $\tilde{\varphi}$ using the BFGS algorithm conditional on $\check{\varphi}$
9: **until** convergence to $\hat{\varphi} = \arg\max_{\varphi \in \Phi} \mathbb{L}(\varphi|Y_n)$
10: **return** $\hat{\varphi}$

For numerical stability of the optimization process, Durbin and Koopman (2012) suggest using the same random numbers to generate the states from the importance density in step 5 of Algorithm 5 for each value of φ. This will ensure that the loglikelihood is a smooth function of the parameters.

Two other functions of interest to estimate using importance sampling are the smoothed states $a_{t|n} = \mathbb{E}(h_t|Y_n)$ and their respective variances $P_{t|n} = \mathbb{V}(h_t|Y_n)$, for $t = 1, \ldots, n$. A simple manipulation of conditional densities similar to that done in (11) yields their respective expressions,

$$a_{t|n} = \frac{\mathbb{E}_g[h_t w(h_t, Y_n)]}{\mathbb{E}_g[w(h_t, Y_n)]} \qquad \text{and} \qquad P_{t|n} = \frac{\mathbb{E}_g[h_t^2 w(h_t, Y_n)]}{\mathbb{E}_g[w(h_t, Y_n)]} - a_{t|n}^2. \tag{17}$$

Conditional on $\hat{\varphi}$, their respective estimates are given by

$$\hat{a}_{t|n} = \sum_{i=1}^{N} \tilde{w}_t^{(i)} h_t^{(i)} \qquad \text{and} \qquad \hat{P}_{t|n} = \sum_{i=1}^{N} \tilde{w}_t^{(i)} h_t^{(i)2} - \hat{a}_{t|n}^2, \tag{18}$$

for $t = 1, \ldots, n$, where $\tilde{w}_t^{(i)} = w(h_t^{(i)}, Y_n) / \sum_{i=1}^{N} w(H_n^{(i)}, Y_n)$ are the normalized importance weights.

One-step-ahead forecasts under the SV model are given by

$$h_{t+1|t} = \mu + \phi h_t. \tag{19}$$

2.4. Non-Gaussian State Space Models

The last and most recent family of models considered here is the Non-Gaussian State Space Models (NGSSM) family proposed in Gamerman et al. (2013) and extended in Pinho et al. (2016) to include heavy-tailed distributions. In the volatility context, the NGSSM is essentially a dynamic scale model with Beta innovations, allowing for a variety of distributions for the observations, including most members of the exponential family. It is defined as

$$\begin{aligned} y_t &= \sigma_t \epsilon_t = \lambda_t^{-1/\nu} \epsilon_t, & \epsilon_t &\sim \text{SGED}(\kappa, \nu) \\ \lambda_{t+1} &= w_{t+1}^{-1} \lambda_t \varsigma_{t+1}, & \varsigma_{t+1}|Y_t &\sim \text{Beta}\left(w a_t, (1 - w_{t+1}) a_t\right), \end{aligned} \tag{20}$$

for $t = 1, \ldots, n$, where $\lambda_t = \sigma_t^{-\nu} > 0$ is the dynamic level, $0 < w \leq 1$ is an autoregressive parameter, a_t is defined in (22) and $w_t = \exp\{\Psi(w a_{t-1}) - \Psi(a_{t-1})\}$ with $\Psi(\cdot)$ denoting the digamma function. The model is initialized with $\lambda_0|Y_0 \sim \text{Gamma}(a_0, b_0)$ where a_0 and b_0 are positive arbitrary constants.

Since the Non-Gaussian State Space Models are relatively recent in the literature, the only known extension proposed so far in this family is the heavy-tailed distribution and scale modelling generalization proposed by Pinho et al. (2016). Therefore, although it has a very flexible form, in terms of stylized facts the basic NGSSM model (20) can only reproduce volatility clustering and leptokurticity.

The role of the parameter w is similar to that of a discount factor: only $100w_t\%$ of the information (in terms of precision) is retained from one period to another. To see this, note that $\mathbb{V}(\lambda_t|Y_{t-1}) = w_t^{-1}\mathbb{V}(\lambda_{t-1}|Y_{t-1})$. Note also that $\mathbb{E}(\lambda_t|Y_{t-1}) = \mathbb{E}(\lambda_{t-1}|Y_{t-1})$; that is, the conditional mean of the dynamic level does not change over time.

Another interesting fact about the NGSSM is that the multiplicative evolution equation can be rewritten as a logarithmic random walk with Beta-scaled innovations. That is, after applying

the logarithm on both sides of the evolution equation in (20), it becomes

$$\log(\lambda_t) = \log(\lambda_{t-1}) + \varsigma_t^*,$$

where $\varsigma_t^* = \log(\varsigma_t/w_t) \in \mathbb{R}$. This equation is very similar to that of a local level model, therefore reinforcing the idea that the NGSSM has by definition a nonstationary variance (though it has a stationary mean).

Analogous to the SV model, in the NGSSM volatility is also driven by its own stochastic process. However, in terms of inference the latter has several advantages over the former. Part 3 of Theorem 1 in Gamerman et al. (2013) states that under the NGSSM it is possible to analytically integrate out the volatilities in (9), thus allowing for exact likelihood inference. This procedure yields a marginal loglikelihood which depends only on φ and Y_n, given by

$$\begin{aligned}\log \mathbb{L}(\varphi|Y_n) = \sum_{t=1}^{n} \Bigg\{ & \log\Gamma(1/\nu + a_{t|t-1}) + \log\left[\frac{\nu}{\tau\Gamma(1/\nu)}\frac{\kappa}{1+\kappa^2}\right] + a_{t|t-1}\log b_{t|t-1} - \log\Gamma(a_{t|t-1}) \\ & - (1/\nu + a_{t|t-1})\log\left[\left(\frac{\kappa(y_t-\pi)^+}{\tau}\right)^{\nu} + \left(\frac{(y_t-\pi)^-}{\kappa\tau}\right)\nu + b_{t|t-1}\right]\Bigg\}, \end{aligned} \tag{21}$$

where

$$\begin{aligned} a_{t|t-1} &= w a_{t-1}, & b_{t|t-1} &= w_t b_{t-1}, \\ a_t &= a_{t|t-1} + 1/\nu, & b_t &= b_{t|t-1} + \left[\left(\frac{\kappa(y_t-\pi)^+}{\tau}\right)^{\nu} + \left(\frac{(y_t-\pi)^-}{\kappa\tau}\right)^{\nu}\right]. \end{aligned} \tag{22}$$

Estimation of $\varphi = (w, \kappa, \nu)'$ proceeds by straightforward maximization of the loglikelihood in (21) using the BFGS algorithm. After estimating φ, it is possible to use the results from Theorem 2 and Parts 1 and 2 of Theorem 1 in Gamerman et al. (2013) to obtain exact filtered and smoothed estimates of the joint vector $L_n = (\lambda_1, \ldots, \lambda_n)'$. The filtering procedure is given in Algorithm 6 and the smoothing procedure is given in Algorithm 7.

Algorithm 6 Non-Gaussian State Space Model filter.

1: initalize $a_0 = 100$ and $b_0 = 100$
2: **for** $t = 1, \ldots, n$ **do**
3: calculate $a_{t|t-1} = w a_{t-1}$ and $b_{t|t-1} = w_t b_{t-1}$
4: calculate $a_t = a_{t|t-1} + 1/\nu$ and $b_t = b_{t|t-1} + \left[\left(\frac{\kappa(y_t-\pi)^+}{\tau}\right)^{\nu} + \left(\frac{(y_t-\pi)^-}{\kappa\tau}\right)^{\nu}\right]$
5: draw $\lambda_t|Y_t, \varphi \sim \text{Gamma}(a_t, b_t)$
6: **end for**
7: **return** $L_n = (\lambda_1, \ldots, \lambda_n)'$

Algorithm 7 Non-Gaussian State Space Model smoother.

1: compute λ_t, a_t and b_t, $t = 1, \ldots, n$ using Algoritm
2: draw $\lambda_n|Y_n, \varphi \sim \text{Gamma}(a_n, b_n)$
3: **for** $t = n-1, \ldots, 1$ **do**
4: draw $\lambda_{t|t+1} = \lambda_t - w_t\lambda_{t+1}|\lambda_{t+1}, Y_t, \varphi \sim \text{Gamma}((1-w_t)a_t, b_t)$.
5: calculate $\lambda_t = \lambda_{t|t+1} + w_t\lambda_{t+1}$
6: **end for**
7: **return** $L_n = (\lambda_1, \ldots, \lambda_n)'$

The one-step-ahead predictive distribution of the dynamic level is given by

$$\lambda_{t+1}|Y_t, \varphi \sim \text{Gamma}(a_{t+1|t}, b_{t+1|t}), \tag{23}$$

where $a_{t+1|t} = w_t a_t$ and $b_{t+1|t} = w_t b_t$. Forecasts are calculated by taking a summary measure of this distribution; its expectation is given by

$$\hat{\lambda}_{t+1|t} = \frac{a_{t+1|t}}{b_{t+1|t}}. \tag{24}$$

3. APPLICATION

In this section an application concerning a sample of 4 daily asset return series, from Jan-2000 to Jan-2016 is made to illustrate the volatility models and evaluation criteria presented so far. The assets are:

1. Ibovespa (BOVESPA): an index composed by a theoretical portfolio with the stocks that accounted for 80% of the volume traded in the last 12 months and that were traded at least on 80% of the trading days in the BM&F Bovespa Stock Exchange. In average the components of Ibovespa represent 70% of all the stock value traded.

2. Hang Seng Index (HANGSENG): a freefloat-adjusted market capitalization-weighted stock market index in Hong Kong. It is used to record and monitor daily changes of the largest companies on the Hong Kong stock market and is the main indicator of the overall market performance in Hong Kong. Its 50 constituent companies represent about 58% of the capitalisation of the Hong Kong Stock Exchange.

3. Merval Index (MERVAL): a price-weighted index, calculated as the market value of a portfolio of stocks selected using the 80% volume and 80% trading days criteria in the last semester in the Buenos Aires Stock Exchange.

4. S&P/TSX Composite Index (SPTSX): an index of the stock (equity) prices of the largest companies on the Toronto Stock Exchange as measured by market capitalization. The listed companies in this index account for about 70% of market capitalization for all Canadian-based companies listed.

Taking p_t to represent the price of the asset (index value) at time t, daily return series are calculated as $y_t = 100 \times [\log(p_t) - \log(p_{t-1})]$ minus its mean. Of the 4000 collected observations, $k = 1000$ - or about 4 years, from 2012 to 2016 - are reserved for forecasting evaluation, while the remaining $n = 3000$ - or about the remaining 12 years, from 2000 to 2012 - are used for model estimation and loglikelihood evaluation.

Table 3 presents summary statistics and autocorrelation tests for the return data. Some observations are in order:

· The "Portmanteau" statistic reported on the table refers to the nonparametric test proposed in Francq and Zakoian (2000) to test for joint absence of autocorrelation. The test is robust to higher moment dependency in the data, such as conditional heteroskedasticity. The null hypothesis is that the autocorrelations from orders 1 through m are jointly equal zero. Here $m \approx \log(n) = 8$, as recommended in Tsay (2010) to maximize the power of the test.

Table 3. Summary statistics for the logreturns series.

Series	BOVESPA	HANGSENG	MERVAL	SPTSX
Mean	0.0216	-0.0002	-0.0242	0.0026
Variance	3.7584	2.7477	4.6981	1.5055
Skewness	-0.1086	-0.0602	-0.1295	-0.6643
Kurtosis	6.6278	10.3158	7.9756	11.9047
Maximum	13.6509	13.4028	16.0407	9.3632
Minimum	-12.1217	-13.5860	-13.0274	-9.7949
Jarque-Bera	1650.9800	6691.9100	3102.8900	10132.2000
Portmanteau	8.4943	2.9338	10.4455	11.6315
Box-Pierce	1318.8600	1570.4400	1247.1500	2089.6200
Chi-squared(2) 5% quantile:	5.9915			
Chi-squared(8) 5% quantile:	15.5073			

· The "Box-Pierce" statistic refers to the conventional Box-Pierce test, but applied to the square of the series. As proved in Francq and Zakoian (2010), it is equivalent to the LM statistic to test for conditional heteroskedasticity. The null hypothesis is that the autocorrelations of the squares from orders 1 through m are jointly equal zero. As in the previous test, $m \approx \log(n) = 8$.

· Results from the first portmanteau test indicate that at the 5% level there is no information in the conditional mean, whereas the second portmanteau test indicate that there is in fact temporal information in the conditional variance, for all series. This is ideal for an application of volatility models, since it is only necessary to model the conditional variance.

· The Jarque-Bera statistic follows a Chi-Squared(2) distribution, of which the 5% quantile is also present in the table. The null hypothesis for this is test is that data comes from a Normal distribution, and it essentially compares deviations of skewness and kurtosis from the Gaussian (which are respectively 0 and 3) with that of the data. The hypothesis of gaussianity is rejected for all 4 series.

· In addition to non-gaussianity, the presence of significant negative skewness and excess kurtosis exhibited in these data are also stylized facts of financial series, as discussed in section 2 of this work.

In accordance to these points, Figure 1 illustrates the time series, standardized density, autocorrelation function[1] (ACF) and ACF of squares for all 4 asset returns. The ACF plots include a nonparametric significance band, used in the portmanteau test and the ACF squared plots include the standard Bartlett significance bands, used in the Box-Pierce test.

Parameter estimates and respective confidence intervals for all 4 volatility models presented are contained in Table 4, and their respective CPU time spent[2] are presented in Table 5. Some interesting points to note are:

· In the APARCH, EGARCH and NGSSM models, the asymmetry parameter κ is always greater than 1 and the tail thickness parameter ν is always between 1 and 2. The corresponding distribution in this case is skewed to the right and is between a skew-Laplace and a skew-Normal in terms of tail thickness.

[1]The maximum lag chosen to display the ACF is $m \approx \min(10 \times \log 10(n), n-1) = 35$.
[2]Computations were performed in a Core i7 Series 3 CPU.

Table 4. Estimated parameters and confidence intervals (in brackets) for the logreturn data.

Series	BOVESPA	HANGSENG	MERVAL	SPTSX
APARCH				
ω	0.058622	0.018694	0.13529	0.01639
	[0.0342;0.1006]	[0.012;0.0291]	[0.0705;0.2595]	[0.0113;0.0237]
α	0.062633	0.065541	0.088835	0.051107
	[0.0432;0.0907]	[0.0511;0.0841]	[0.0619;0.1275]	[0.0318;0.0822]
β	0.91762	0.93508	0.86968	0.93338
	[0.893;0.9429]	[0.9203;0.9501]	[0.8335;0.9075]	[0.9152;0.9519]
γ	0.58883	0.50395	0.19022	0.8101
	[0.2832;0.7859]	[0.2984;0.6648]	[0.0798;0.2961]	[-0.1653;0.9844]
δ	1.3852	1.17	2.2051	1.3754
	[0.9518;2.0161]	[0.8287;1.6518]	[1.5833;3.0711]	[1.0295;1.8377]
κ	1.0559	1.0465	1.0556	1.1154
	[1.0165;1.0968]	[1.0093;1.0851]	[1.0143;1.0986]	[1.0756;1.1567]
ν	1.7018	1.4687	1.2887	1.5771
	[1.581;1.8319]	[1.359;1.5873]	[1.2047;1.3786]	[1.4663;1.6962]
EGARCH				
ω	0.091424	0.055073	0.079945	0.047208
	[0.0635;0.1193]	[0.0389;0.0713]	[0.052;0.1079]	[0.032;0.0624]
β	0.97593	0.98731	0.9719	0.98538
	[0.9612;0.9851]	[0.9803;0.9918]	[0.9549;0.9826]	[0.9785;0.9901]
θ	-0.077507	-0.06125	-0.051785	-0.080142
	[-0.1007;-0.0543]	[-0.0813;-0.0412]	[-0.0747;-0.0288]	[-0.1016;-0.0587]
γ	0.12986	0.12707	0.19286	0.11266
	[0.0947;0.1651]	[0.0985;0.1556]	[0.1479;0.2379]	[0.0822;0.1431]
κ	1.0556	1.0467	1.0535	1.1128
	[1.0162;1.0965]	[1.01;1.0848]	[1.0184;1.0898]	[1.073;1.1541]
ν	1.6912	1.4615	1.2631	1.5637
	[1.5716;1.82]	[1.3527;1.5789]	[1.1823;1.3495]	[1.4536;1.6821]
NGSSM				
w	0.94114	0.95087	0.92875	0.94284
	[0.928;0.952]	[0.9394;0.9603]	[0.9138;0.9413]	[0.9294;0.9539]
κ	1.0316	1.0404	1.0429	1.115
	[0.9927;1.072]	[1.0055;1.0766]	[1.0026;1.0848]	[1.0767;1.1546]
ν	1.8205	1.4842	1.4745	1.7125
	[1.6724;1.9816]	[1.3702;1.6077]	[1.3596;1.5992]	[1.5747;1.8625]
SV				
μ	0.019837	0.0040052	0.038969	-0.0024313
	[0.0078;0.0318]	[-0.0009;0.0089]	[0.0203;0.0576]	[-0.0075;0.0026]
ϕ	0.98102	0.99159	0.96472	0.98868
	[0.9705;0.9916]	[0.9861;0.9971]	[0.9492;0.9803]	[0.9818;0.9956]
σ_η	0.1263	0.11061	0.23097	0.13582
	[0.096;0.1566]	[0.0874;0.1338]	[0.1847;0.2773]	[0.1061;0.1655]

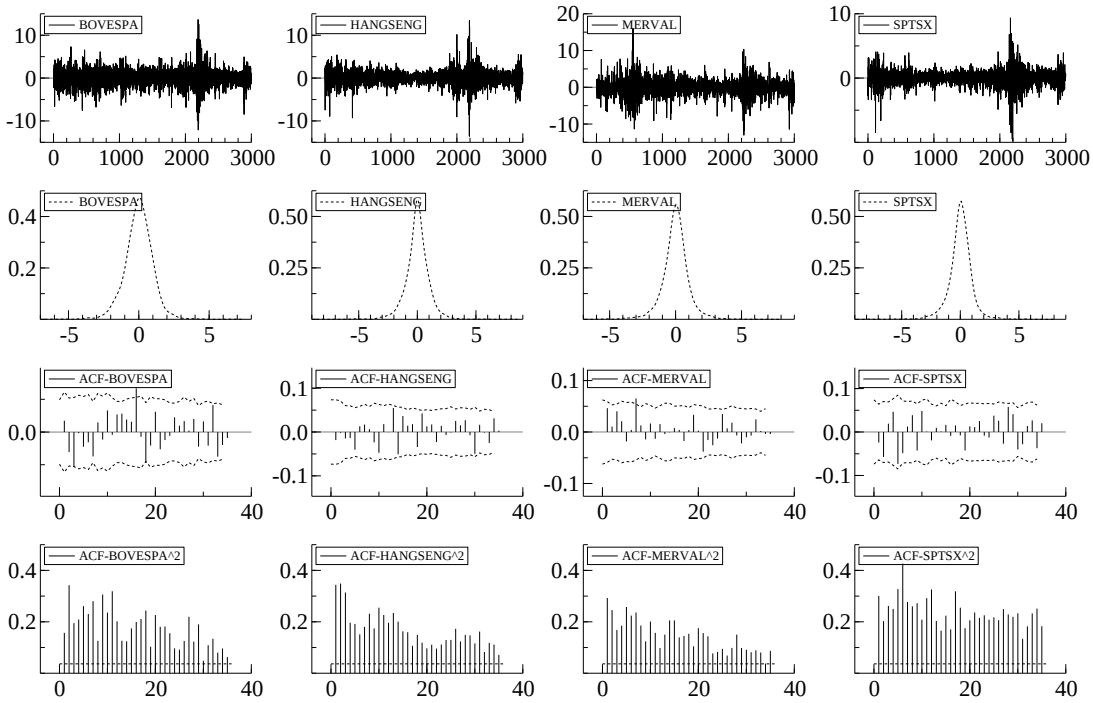

Figure 1. Time series, standardized densities, ACF and ACF-squared of logreturn data.

Table 5. CPU time spent (in seconds) in the estimation of each model.

Series	BOVESPA	HANGSENG	MERVAL	SPTSX
APARCH	3.094	2.797	3.235	3.453
EGARCH	1.829	1.703	1.687	1.438
NGSSM	0.859	0.828	0.86	0.953
SV	71.082	73.129	66.05	67.206

· In the SV model the hypotheses $\kappa = 1$ and $\nu = 2$ could not be rejected at the 5% level, for all 4 series. Although the lack of skewness in the unconditional distribution of returns is somewhat surprising, unconditional mesokurtosis is actually expected in a SV model. As discussed earlier, the work of Carnero et al. (2004) argues that in comparison to ARCH-type models the lognormal SV can reproduce a much wider range of behavior (especially excess unconditional leptokurtosis) without the inclusion of additional parameters. Since when $\kappa = 1$ and $\nu = 2$ the skew-GED distribution reduces to the Gaussian, the re-estimated SV model under these assumptions actually corresponds to the lognormal SV. Computational efficiency gains to relaxing the skew-GED assumption are substantial, reducing CPU time by a factor greater than 5.

· There is a vast difference between computational time amongst models: estimating the NGSSM takes the least amount of time, and it is followed in this sense by the EGARCH, the APARCH and the SV. Although the absolute difference in seconds might not appear large, the relative difference is striking. For example, for the BOVESPA series, the SV model takes roughly 83 times as long as the NGSSM model, the APARCH takes about 4 times and the EGARCH about 2 times. Relative computational efficiency is especially important when the number of series considered increases exponentially, as is the case in a large portfolio management.

· All models display a strong persistence in the volatility; this is measured in the APARCH by $\alpha + \beta$, in the EGARCH by β, in the NGSSM by w and in the SV by ϕ. This is also a stylized fact of financial series and it is to be expected. However, it is worth noticing that the hypothesis of stationarity (tested by veryfing that the value 1 is contained in the confidence intervals for these parameters) can not be rejected at 5% for any of these models - except the NGSSM, which is nonstationary by construction - for any of the series.

· The leverage effect is significant in the EGARCH for all 4 series, and in APARCH for all but the SPTSX. The estimated coefficients also have the expected sign: positive γ in the APARCH and negative θ in the EGARCH.

After estimating model parameters, smoothed volatility estimates can also be obtained for each model. They are contained in Figure 2 for each model and series, along with the corresponding proxy for the true volatility (absolute value of logreturns). Another quantity of interest is the one-step-ahead forecast of the volatility, which is calculated recursively for the $k = 1000$ reserved observations for all models and series, and shown in Figure 3 along with the absolute returns for these observations. When calculating the forecasts, it is assumed that the sample size $n = 3000$ used for estimation is enough to ensure that estimated parameters remain relatively constant over time, so that the models do not need to be reestimated after each new observation is included in the sample. The smoothed and one-step-ahead volatility forecasts seem to closely reproduce the observed behavior patterns in the data.

Although the above information is useful to illustrate and understand the behavior of volatility and financial series in general, it contributes little to the issue of model comparison. The adequate statistical tools for that end are the evaluation criteria introduced in section 2.1. Table 6 contains the computed criteria for all estimated models so far.

The logic of comparison here is to first rank the models according to how they perform according to each individual criterion, and then if possible stablish an overall rank. The criteria are:

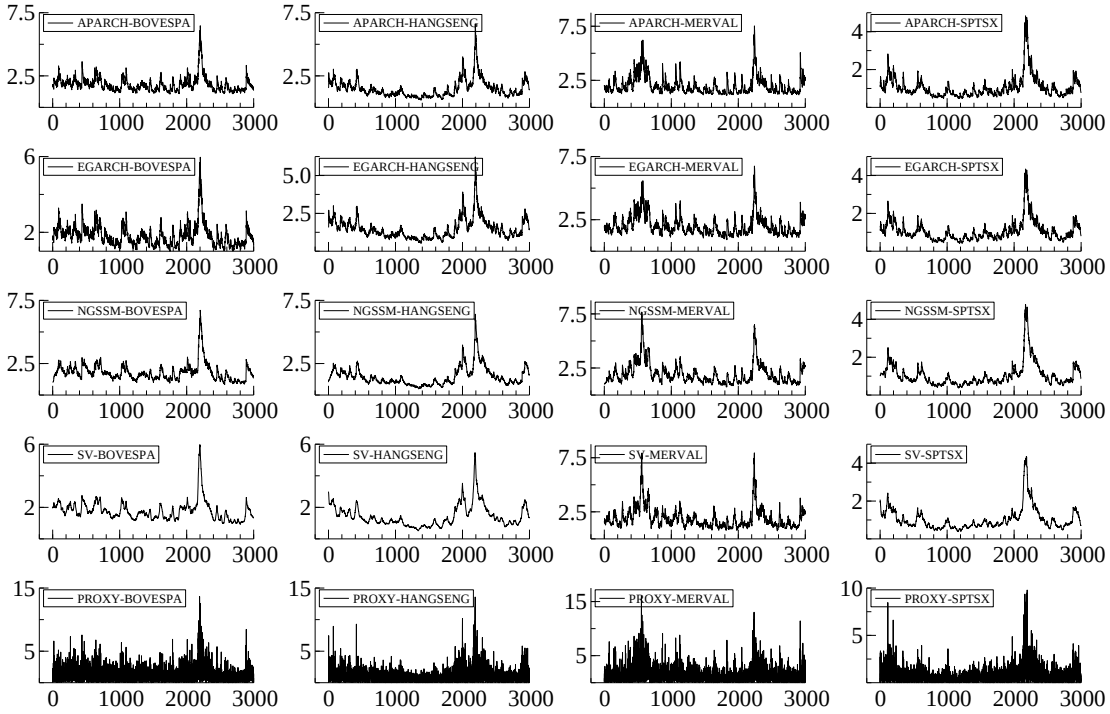

Figure 2. Smoothed volatility estimates and absolute returns.

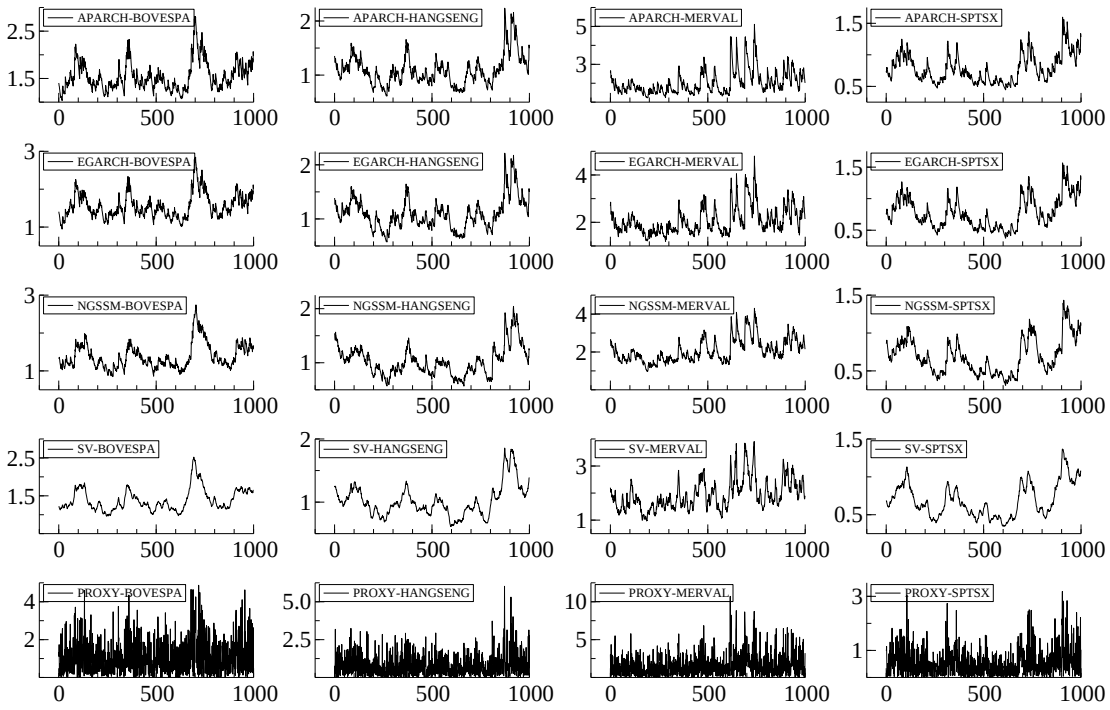

Figure 3. One-step ahead volatility forecasts and absolute returns.

Table 6. Evaluation criteria (to be minimized) for the estimated models.

Series	BOVESPA	HANGSENG	MERVAL	SPTSX
AIC(p)				
APARCH	3.9511	3.4220	4.0433	2.7342
EGARCH	4.0107	3.4509	4.0609	2.7694
NGSSM	4.0139	3.4327	4.0222	2.7644
SV	3.8578	3.3439	3.8855	2.6494
MSE(k)				
APARCH	0.9717	0.6059	2.2828	0.2736
EGARCH	0.9887	0.6115	2.3003	0.2731
NGSSM	0.8854	0.5933	2.3722	0.2520
SV	0.7985	0.5256	1.8387	0.2199
UC(k)				
APARCH	0.0170	0.0030	0.0040	0.0060
EGARCH	0.0190	0.0040	0.0020	0.0120
NGSSM	0.0080	0.0120	0.0000	0.0170
SV	0.0080	0.0000	0.0010	0.0070

1. Akaike Information Criteria - measures the overall goodness-of-fit of a model as a measure of the expected information lost when approximating the true DGP with an estimated model, and its calculation involves the maximized loglikelihood and a penalization for model complexity (represented by the number of estimated parameters). It indicates that the SV outperforms all other models for all series, that the APARCH outperforms the NGSSM in 3 out of 4 cases (with MERVAL being the exception), and that the EGARCH is always outperformed by competing specifications (although there is a draw between EGARCH and NGSSM).

2. Minimum Squared Error - measures forecasting accuracy for one-step-ahead forecasts. The proxy used as the true volatility to calculate the MSE was the absolute returns from the reserved observation set, which is of size $k = 1000$. Also, it is assumed that the sample size used of 3000 is enough to ensure that the estimated parameters remain relatively constant over time, so that the models do not need to be reestimated after each new observation is included in the sample. The comparison of MSE values suggests that the SV model also outperforms all other competing specifications in forecasting accuracy, being followed by the NGSSM which outperforms the APARCH 3 out of 4 times (again with MERVAL being the exception), by the APARCH which outperforms the EGARCH 3 out of 4 times (with SPTSX being the exception, although these rankings are remarkably close). The NGSSM also outperforms the EGARCH 3 out of 4 times, with MERVAL being again the exception.

3. Unconditional Value-at-Risk Coverage - an important and widely used measure of financial risk of an asset. The VaR is computed at the level of 5% and involves one-day-ahead operations. A number of $k = 1000$ observations are used for the estimation and evaluation of pseudo-out-of-sample VaR, and the reported values represent absolute deviations from the $\alpha = 5\%$ quantile. Evidence for this criterion is mixed: the SV outperforms all other models at best 2 out of 4 times (3 out of 4 compared to the NGSSM and the EGARCH but 4 out of 4 compared to the APARCH), as well as the EGARCH and the NGSSM, with the APARCH

coming last. If each model is compared individually, the SV model comes out as the favorite, followed by a tie between the EGARCH and the NGSSM, and finally by the APARCH.

Since evidence from the criteria values is somewhat mixed, specially in the out-of-sample VaR assessment, a more robust procedure to compare estimated models is desirable. With that in mind, the next section contains a simulation experiment drawing from the estimates in this section; the logic is that the increased number of replications will minimize the error attributed to randomness and produce a more consistent ranking amongst model performances.

4. SIMULATION EXPERIMENT

The simulations are performed by taking previous estimated parameter sets as the true parameter sets. Since there are 4 different sets and 4 volatility specifications - APARCH, EGARCH, NGSSM and SV - there are thus a total 16 Data Generating Processes (DGPs) to take into account. Each DGP is replicated $M = 1000$ times, and each series has a size of $n = 3000$ observations, plus $k = 1000$ reserved for forecasting. In all simulations a burn-in of 7000 observations is used in the drawing process to alleviate dependence from initial values. The specific parameter values used for simulation are restated for convenience in Table 7.

Table 7. Simulation experiment parameter sets.

Set	A	B	C	D
APARCH				
ω	0.0586	0.0187	0.1353	0.0164
α	0.0626	0.0655	0.0888	0.0511
β	0.9176	0.9351	0.8697	0.9334
γ	0.5888	0.5040	0.1902	0.8101
δ	1.3852	1.1700	2.2051	1.3754
κ	1.0559	1.0465	1.0556	1.1154
ν	1.7018	1.4687	1.2887	1.5771
EGARCH				
ω	0.0914	0.0551	0.0799	0.0472
β	0.9759	0.9873	0.9719	0.9854
θ	-0.0775	-0.0613	-0.0518	-0.0801
γ	0.1299	0.1271	0.1929	0.1127
κ	1.0556	1.0467	1.0535	1.1128
ν	1.6912	1.4615	1.2631	1.5637
NGSSM				
w	0.9411	0.9509	0.9288	0.9428
κ	1.0316	1.0404	1.0429	1.1150
ν	1.8205	1.4842	1.4745	1.7125
SV				
μ	0.0198	0.0040	0.0390	-0.0024
ϕ	0.9810	0.9916	0.9647	0.9887
σ_η	0.1263	0.1106	0.2310	0.1358

In order to ensure that experiment results do not arise from bias in the estimation, it is necessary to assess the accuracy of the maximum likelihood estimators used. To that end, an interesting

quantity is the percentage bias, defined as

$$\text{Bias}(\%) = 100 \times \frac{\hat{\varphi} - \varphi}{|\varphi|},$$

where $\hat{\varphi}$ is the estimated parameter vector and φ is the true parameter vector. The advantage of using the percentage bias is that it by definition describes a density, allowing any descriptive measure such as the mean or the median to be calculated from it.

Table 8. Percentage bias under parameter set A: mean, median and SD.

	Mean	Median	SD
APARCH			
ω	-21.7940	-27.7360	9.9769
α	-10.4080	-22.7710	8.3136
β	-4.0045	-0.6268	3.1595
γ	-1.1203	-0.2002	2.4897
δ	2.9287	0.9933	5.2934
κ	12.3450	10.9170	11.1300
ν	25.2890	23.4410	14.1440
EGARCH			
ω	-12.7130	-6.5440	9.7709
β	-7.8141	-0.4723	9.5222
θ	-2.5709	0.4339	4.1677
γ	1.3404	4.0327	5.4771
κ	5.9813	29.1850	8.5568
ν	12.5610	31.0220	11.4630
NGSSM			
w	-6.5933	-4.6899	2.1003
κ	-1.1592	-0.2937	0.7310
ν	10.6340	6.4233	5.3036
SV			
μ	-66.8240	-0.3598	91.1320
ϕ	1.0016	0.3747	5.5772
σ_η	45.0100	153.4000	59.7100

Since conclusions about the estimation bias are the same for all parameter sets, only the results for the first set is presented here. It should be clear that it only makes sense to access accuracy under the same DGP with which it was simulated from; therefore, when assessing bias, only the respective DGP of a model is taken into account for each estimated model.

Figure 4 contains the percentage bias ordinate corresponding to each estimated parameter in each model for all 1000 replications; their respective mean, median and standard deviation are contained in Table 8. According to maximum likelihood estimation assumptions, the percentage bias should have an asymptotical Normal distribution with zero mean and unit variance. Although most estimated parameters do seem to satisfy that criteria, the ones directly unrelated to the autoregressive dependence of the volatility present statistically significant means and fatter distribution tails; they are the constants ω in the APARCH and EGARCH and μ in the SV and the tail thickness parameter ν on the APARCH, EGARCH and the NGSSM. Despite this fact, the

overall accuracy of the maximum likelihood estimators employed here seem satisfying.

Since one of the objectives of this work is to asses the relative computational time benchmarks across different volatility families, Table 9 present the CPU time spent on the estimation of each model under each parameter set and DGP. Note that the median relative CPU time spent is roughly the same as that reported in Table 5 in the application; the SV takes close to 82 times as much CPU time as the NGSSM, while the APARCH takes about 4 times and the EGARCH about 2 times.

The order in which the simulation experiment results are presented is similar to the previous section: first an overall comparison of each evaluation criteria (AIC, MSE and Unconditional VaR Coverage) is present, followed by a respective detailed comparison.

Table 9. CPU time (in seconds) spent on the estimation of each set of models.

Set	A	B	C	D
APARCH DGP				
APARCH	2906.84	2751.09	2970.50	3627.05
EGARCH	1522.04	1374.91	1379.86	1270.06
NGSSM	706.71	693.95	679.24	710.06
SV	86450.80	85942.00	82806.00	88259.10
EGARCH DGP				
APARCH	2727.31	2720.58	2574.38	3219.69
EGARCH	1558.80	1409.14	1411.75	1290.26
NGSSM	690.47	701.00	667.94	691.28
SV	97481.80	88060.00	86418.70	82344.50
NGSSM DGP				
APARCH	2652.29	2648.63	2627.06	2565.79
EGARCH	1612.64	1456.83	1408.21	1296.18
NGSSM	697.87	703.28	670.93	711.07
SV	85190.10	88053.10	82887.20	87330.60
SV DGP				
APARCH	2640.45	2640.40	2585.16	2553.98
EGARCH	1618.54	1445.77	1409.31	1292.39
NGSSM	699.28	699.25	666.11	690.40
SV	100416.00	85679.00	83310.30	81029.50

Starting with goodness of fit, Table 10 contains the percentage of times a single model outperforms all others; that is, the percentage of times an estimated model's AIC is smaller or equal than its competing specifications'. In this overall comparison, the SV consistently outperforms the APARCH, EGARCH and NGSSM - its AIC was smaller in 100% samples of all 16 DGPs. Although this result might seem somewhat surprising, since it is expected that a model will outperform all others under the simulation of its own DGP, it illustrates the range of behavior that can be reproduced by the Stochastic Volatility model albeit it having a small number of parameters.

The overall comparison, although illustrative, lacks in depth and contains no information about relative rankings between the other families of models. Therefore, Table 13 provides a separate and detailed comparison of the goodness of fit of each model: each cell represents the percentage of times the AIC of a model was smaller or equal than its direct competing model. For example, the first row and second column contains the percentage of times that the APARCH model's AIC

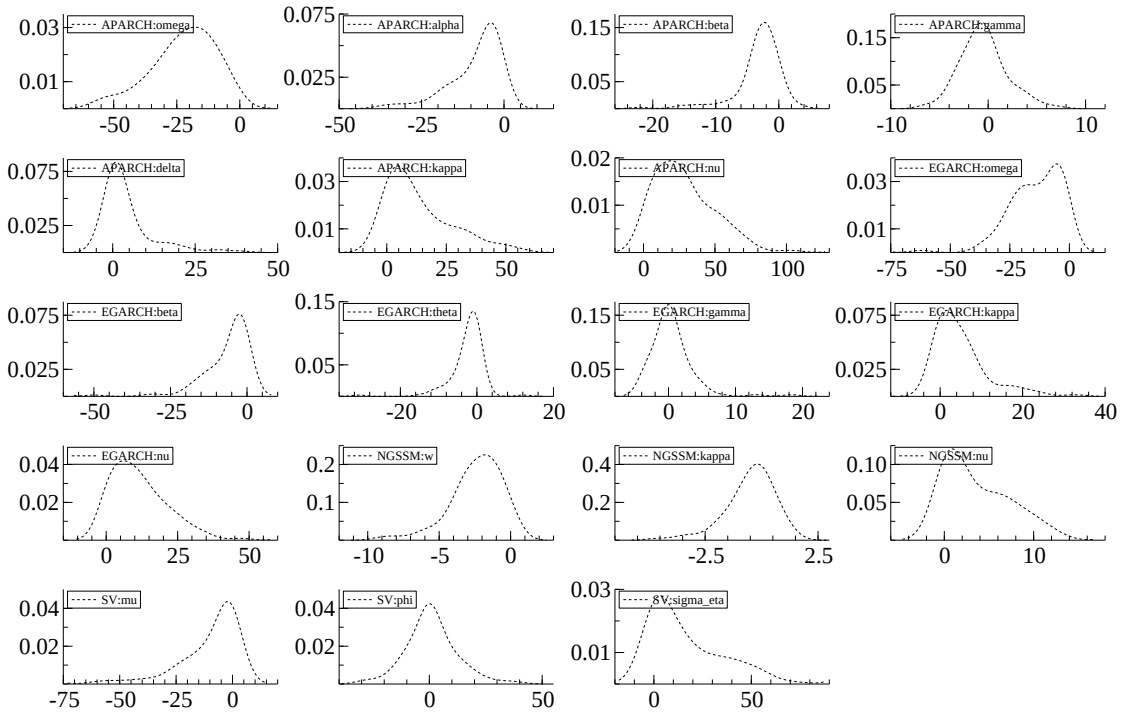

Figure 4. Standardized estimated parameter densities under parameter set A.

was smaller than or equal the EGARCH's under the APARCH DGP and the first parameter set. Since this measure of performance also includes equality between criteria values, diagonals of each block-matrix of comparisons are always equal to 100%.

In the detailed comparison of AIC values the SV outperforms all other models 100% of times, under each DGP and parameter set. The APARCH consistenly outperforms the NGSSM (although not as often as the SV) and the EGARCH, and the NGSSM often outperforms the EGARCH. It can therefore be concluded that an appropriate ranking reflecting goodness of fit between models is, in descending order: SV, APARCH, NGSSM and EGARCH.

Next are the forecast comparisons, based on the Mean Squared Error. Table 11 is the analogue of Table 10 and Table 14 is the analogue of Table 13; respectively, they contain the overall and detailed comparison of MSE values calculated for each DGP and parameter set of the simulation experiment. As in the application made in the previous section, the overall comparison points to the SV as the best forecasting model, outperforming the others 100% of times under all DGPs and parameter sets.

The detailed comparison also points out the SV as the best forecastig model 100% of times, with the EGARCH coming second by outperforming both the APARCH and the NGSSM. The difference in relative performance between the APARCH and the NGSSM is narrow: in average, the APARCH outperforms the NGSSM only 52.87% of times. It is interesting to note, however, that the performance gap between the EGARCH and the NGSSM is not so large as the gap between the EGARCH and the APARCH. As expected, the relative performance of a model is significantly better under its respective DGP. The ranking suggested by the forecasting criterion is, in descending order: SV, EGARCH, APARCH and NGSSM.

The final comparison contained here concerns the out-of-sample Value-at-Risk, of which the

Table 10. Overall number of times (%) of AIC minimization.

Set	A	B	C	D
APARCH DGP				
APARCH	0.00	0.00	0.00	0.00
EGARCH	0.00	0.00	0.00	0.00
NGSSM	0.00	0.00	0.00	0.00
SV	100.00	100.00	100.00	100.00
EGARCH DGP				
APARCH	0.00	0.00	0.00	0.00
EGARCH	0.00	0.00	0.00	0.00
NGSSM	0.00	0.00	0.00	0.00
SV	100.00	100.00	100.00	100.00
NGSSM DGP				
APARCH	0.00	0.00	0.00	0.00
EGARCH	0.00	0.00	0.00	0.00
NGSSM	0.00	0.00	0.00	0.00
SV	100.00	100.00	100.00	100.00
SV DGP				
APARCH	0.00	0.00	0.00	0.00
EGARCH	0.00	0.00	0.00	0.00
NGSSM	0.00	0.00	0.00	0.00
SV	100.00	100.00	100.00	100.00

unconditional coverage of 1-day one-step-ahead VaR forecast is taken as evaluation criteria. As before, an overall comparison is presented in Table 12 and a detailed comparison is presented in Table 15. Unlike the results based on AIC and MSE, however, the overall comparison is much more heterogeneous; the average percentage of times a model outperforms all others in all DGPs and parameter sets is 54.31% for the APARCH, 39.69% for the SV, 5.38% for the EGARCH and 0.63% for the NGSSM.

In line with these results, the detailed comparison in Table 15 is also fairly heterogeneous: there is a stark difference between performances under the various parameter sets. Although models still perform significantly better under their own DGP, the APARCH consistently outperforms the EGARCH and the NGSSM, but not the SV; the latter is outperformed by the former only 58.50% of times. The Stochastic Volatility model also consistently outperforms the EGARCH and the NGSSM, and the NGSSM is often outperformed by all others. The suggested ranking is, in descending order: APARCH, SV, EGARCH and NGSSM.

The next section contains a brief discussion of results from both the present and previous section and concludes this work.

5. CONCLUSION

In this work a comparison of three families of volatility models, namely the Autoregressive Conditional Heteroskedasticity (ARCH), Stochastic Volatility (SV) and Non-Gaussian State Space Models (NGSSM) was made according to three different metrics: goodness of fit, forecasting and assessing Value-at-Risk (VaR). The models' inference procedures under the flexible Skew Gener-

Table 11. Overall MSE comparison: number of times (in %) a single model outperformed all others in a DGP.

Set	A	B	C	D
APARCH DGP				
APARCH	0.00	0.00	0.00	0.00
EGARCH	0.00	0.00	0.00	0.00
NGSSM	0.00	0.00	0.00	0.00
SV	100.00	100.00	100.00	100.00
EGARCH DGP				
APARCH	0.00	0.00	0.00	0.00
EGARCH	0.00	0.00	0.00	0.00
NGSSM	0.00	0.00	0.00	0.00
SV	100.00	100.00	100.00	100.00
NGSSM DGP				
APARCH	0.00	0.00	0.00	0.00
EGARCH	0.00	0.00	0.00	0.00
NGSSM	0.00	0.00	0.00	0.00
SV	100.00	100.00	100.00	100.00
SV DGP				
APARCH	0.00	0.00	0.00	0.00
EGARCH	0.00	0.00	0.00	0.00
NGSSM	0.00	0.00	0.00	0.00
SV	100.00	100.00	100.00	100.00

alized Error family of distributions were detailed. Respective evaluation criteria used for these metrics were the Akaike Information Criterion, Mean Squared Error of one-step-ahead forecasts and Unconditional Coverage of 1-day one-step-ahead forecast VaR. The data used were daily asset return series (Ibovespa, Hang Seng Index, Merval Index and S&PTSX Index) from Jan-2000 to Jan-2016, or roughly 4000 observations, from which 3000 were used for estimation and 1000 were reserved for forecasting and VaR evaluation. Parameter estimates served as basis to conduct a simulation experiment which consisted of 1000 replications of series with size 3000 plus 1000 observations reserved for forecasting.

Although some interesting conclusions can be drawn from the application, some of its evidence regarding model performance were somewhat mixed, creating the demand for the simulation experiment. Important results from the simulation experiment are:

- Estimates of parameters which are not directly related to the autoregressive dependence of the volatility, specifically model constants and the tail thickness parameter of the Skew-GED are substantially more biased than the rest. This is probably a reflection of the distant relationship these parameters have to the data, which makes their identification difficult.

- Relative CPU time spent on model estimation is fairly stable between the application the simulation experiment. Taking the NGSSM as benchmark, SV, APARCH and EGARCH estimation are slower by a respective factor of about 82, 4 and 2.

- The suggested rankings according to goodness of fit, forecasting and VaR assesment performance are, in descending order:

Table 12. Overall Unconditional VaR Coverage comparison: number of times (in %) a single model outperformed all others in a DGP.

Set	A	B	C	D
APARCH DGP				
APARCH	34.00	52.00	71.00	47.00
EGARCH	4.00	7.00	17.00	10.00
NGSSM	0.00	0.00	0.00	1.00
SV	62.00	41.00	12.00	42.00
EGARCH DGP				
APARCH	34.00	49.00	74.00	50.00
EGARCH	6.00	2.00	4.00	9.00
NGSSM	0.00	0.00	1.00	2.00
SV	60.00	49.00	21.00	39.00
NGSSM DGP				
APARCH	53.00	49.00	72.00	55.00
EGARCH	2.00	4.00	2.00	11.00
NGSSM	1.00	0.00	0.00	1.00
SV	44.00	47.00	26.00	33.00
SV DGP				
APARCH	45.00	49.00	73.00	62.00
EGARCH	2.00	2.00	3.00	1.00
NGSSM	3.00	1.00	0.00	0.00
SV	50.00	48.00	24.00	37.00

- AIC: SV, APARCH, NGSSM, EGARCH.
- MSE: SV, EGARCH, APARCH, NGSSM.
- UC: APARCH, SV, EGARCH, NGSSM.

It is convenient to restate here the objectives of this work, in order to ascertain whether or not they were successfully attained. They are:

1. Provide an accessible reference for the properties and inference techniques for the families of models presented here, specifically when a skewed and leptokurtic distribution (such as the Skew-GED) is assumed for the error terms.
2. Determine a family of models as being the most adequate for each metric: goodness of fit, forecasting and Value-at-Risk.
3. Draw conclusions about which features/stylized facts influence model performance the most, overall and for each criteria.
4. Establish a trade-off between accuracy and computational efficiency between models.

The second section of this work is dedicated entirely to objective 1, while sections 3 and 4 are dedicated to objective 2, although they also seem to fulfill the requirements for objectives 3 and 4; the empirical results and simulation experiment not only ascertain a ranking according to the evaluation criteria used but in doing so illustrate the relative importance of captured stylized facts and computational efficiency.

For example, the relative rankings indicate that for obtaining a good fit and accurate forecasts the extra flexibility introduced by an additional source of error in the SV model might be more important than the functional form flexibility of the APARCH or even the leverage effect captured by both EGARCH and APARCH models. The relative rankings also indicate that the most CPU time-consuming model (Stochastic Volatility) has an overall best performance in goodness of fit and forecasting and is ranked second in VaR assesment; however, there is no consistent difference between the fastest models APARCH, EGARCH and NGSSM. Since relative CPU time consumed between the latter group is not nearly as striking between the SV and alternatives, this serves as illustration that although a large increase in model complexity (as measured by CPU time spent) leads to an improvement in accuracy, a small increase does not seem to have any effect.

This conclusion would not be complete if the limitations of the results obtained here were not acknowledged: a comparison based only on criteria values does not allow one to perform statistical inference in the sense of precising whether the difference between two criteria values is significant or not. There should be no confusion between these two goals; the methodology used here provides the answer to how one should rank models across different criteria, and not to how significant is the difference between these criteria.

REFERENCES

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6):716–723.

Angelidis, T., Benos, A., and Degiannakis, S. (2004). The use of garch models in var estimation. *Statistical Methodology*, 1(1–2):105 – 128.

Ayebo, A. and Kozubowski, T. J. (2003). An asymmetric generalization of gaussian and laplace laws. *Journal of Probability and Statistical Science*, 1.

Baillie, R. T., Bollerslev, T., and Mikkelsen, H. O. (1996). Fractionally integrated generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 74(1):3–30.

Bauwens, L., Hafner, C., and Laurent, S. (2012). *Volatility Models and Their Applications*. Wiley.

Black, F. (1976). Studies of stock price volatility changes. *Proceedings from the American Statistical Association, Business and Economic Statistics Section 177– 181.*, 6(2).

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31(3):307–327.

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

Broto, C. and Ruiz, E. (2004). Estimation methods for stochastic volatility models: A survey. *Journal of Economic Surveys*, 18(5).

Broyden, C. G. (1970). The convergence of a class of double-rank minimization algorithms 1. general considerations. *IMA Journal of Applied Mathematics*, 6(1):76–90.

Carnero, M. A., Pena, D., and Ruiz, E. (2004). Persistence and kurtosis in garch and stochastic volatility models. *Journal of Financial Econometrics*, 2(2):319–342.

Chan, K. F. and Gray, P. (2006). Using extreme value theory to measure value-at-risk for daily electricity spot prices. *International Journal of Forecasting*, 22.

Christoffersen, P. (1998). Evaluating interval forecasts. *International Economic Review*, 39.

Deschamps, P. J. (2011). Bayesian estimation of an extended local scale stochastic volatility model. *Journal of Econometrics*, 162.

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

Durbin, J. and Koopman, S. (2012). *Time Series Analysis by State Space Methods: Second Edition*. Oxford Statistical Science Series. OUP Oxford.

Durbin, J. and Koopman, S. J. (1997). Monte carlo maximum likelihood estimation for non-gaussian state space models. *Biometrika*, 84(3):669–684.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica*, 50(4):987–1007.

Fletcher, R. (1970). A new approach to variable metric algorithms. *The Computer Journal*, 13(3):317–322.

Francq, C. and Zakoian, J.-M. (2000). Estimating Stochastic Volatility Models : A New Approach Based on ARMA Representations. Working Papers 2000-47, Centre de Recherche en Economie et Statistique.

Francq, C. and Zakoian, J.-M. (2010). *GARCH Models: Structure, Statistical Inference and Financial Applications*. Wiley.

Gamerman, D., Santos, T. R., and Franco, G. C. (2013). A non-gaussian family of state-space models with exact marginal likelihood. *Journal of Time Series Analysis*, 34.

Geweke, J. (1986). Modelling the persistence of conditional variance: A comment. *Econometric Reviews*, 5.

Glosten, L. R., Jagannathan, R., and Runkle, D. E. (1993). On the relation between the expected value and the volatility of the nominal excess return on stocks. *Journal of Finance*, 48(5):1779–1801.

Goldfarb, D. (1970). A family of variable-metric methods derived by variational means. *Mathematics of Computation*, 24:23–26.

Hamilton, J. D. (1994). *Time series analysis*. Princeton.

Hansen, P. R. and Lunde, A. (2005). A forecast comparison of volatility models: Does anything beat a garch(1,1)? *Journal of Applied Econometrics*, 20.

Harvey, A. C. (1998). Long memory in stochastic volatility. *In J. Knight and S. Satchell (eds.), Forecasting Volatility in Financial Markets*.

Harvey, A. C. and Shephard, N. G. (1996). Estimation of an asymmetric stochastic volatility model for asset returns. *Journal of Business and Economic Statistics*, 14.

Higgins, M. L. and Bera, A. K. (1992). A class of nonlinear arch models. *International Economic Review*, 33.

Iltuzer, Z. and Tas, O. (2013). The forecasting performances of volatility models in emerging stock markets: Is a generalization really possible? *Journal of Applied Finance & Banking*, 3(2).

Kim, S., Shephard, N., and Chib, S. (1998). Stochastic volatility: Likelihood inference and comparison with arch models. *The Review of Economic Studies*, 65(3).

Lambert, P. and Laurent, S. (2002). Modelling skewness dynamics in series of financial data using skewed location-scale distributions. *Working paper*.

Mandelbrot, B. (1963). The variation of certain speculative prices. *The Journal of Business*, 36:394.

Meitz, M. and Saikkonen, P. (2011). Parameter estimation in nonlinear ar-garch models. *Econometric Theory*, 27.

Nakajima, J. (2012). Bayesian analysis of generalized autoregressive conditional heteroskedasticity and stochastic volatility: Modelling leverage, jumps and heavy-tails for financial time series. *The Japanese Economic Review*, 63(1).

Nelson, D. B. (1991). Conditional heteroskedasticity in asset returns: A new approach. *Econometrica*, 59(2):347–70.

Omori, Y., Chib, S., Shephard, N., and Nakajima, J. (2007). Stochastic volatility with leverage: Fast and efficient likelihood inference. *Journal of Econometrics*, 140.

Pantula, S. (1986). Modelling the persistence of conditional variance: A comment. *Econometric Reviews*, 5.

Pinho, F. M., Franco, G. C., and Silva, R. S. (2016). Modeling volatility using state space models with heavy tailed distributions. *Mathematics and Computers in Simulation*, 119:108 – 127.

Pinho, F. M. and Santos, T. R. (2013). Volatility of the european stock market indices during the global financial crisis - a new proposal of stochastic volatility. *Journal of Statistical and Econometric Methods*, 2(2):107–126.

Rabemananjara, R. and Zakoian, J. M. (1993). Threshold arch models and asymmetries in volatility. *Journal of Applied Econometrics*, 8.

Schwert, W. (1990). Stock volatility and the crash of '87. *Review of Financial Studies*, 3.

Shanno, D. F. (1970). Conditioning of quasi-newton methods for function minimization. *Mathematics of Computation*, 24:647–656.

Shephard, N. (1994a). Local scale models: State space alternative to integrated garch processes. *Journal of Econometrics*, 60.

Shephard, N. (1994b). Partial non-gaussian state space. *Biometrika*, 81(1):115–131.

Shephard, N. and Pitt, M. K. (1997). Likelihood analysis of non-gaussian measurement time series. *Biometrika*, 84(3):653–667.

Silva, U. M., Pinho, F. M., and Salvato, M. A. (2015). Um estudo comparativo de modelos de heterocedasticidade condicional. Working paper, Universidade Federal de Minas Gerais - UFMG.

So, M. K. P. and Yu, P. L. H. (2006). Empirical analysis of garch models in value at risk estimation. *Journal of International Financial Markets, Institutions & Money*, 16.

Taylor, S. J. (1982). Financial returns modelled by the product of two stochastic processes, a study of daily sugar prices 1961-79. *In O. D. Anderson (Ed.), Time Series Analysis: Theory and Practice*, 1.

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

Tsay, R. S. (2010). *Analysis of Financial Time Series*. Wiley.

Zakoian, J.-M. (1994). Threshold heteroskedastic models. *Journal of Economic Dynamics and Control*, 18(5):931–955.

Table 13. Detailed AIC comparison: number of times (in %) the model in a row outperformed the corresponding model in a column.

Set	A				B			
APARCH DGP								
	APARCH	EGARCH	NGSSM	SV	APARCH	EGARCH	NGSSM	SV
APARCH	100.00	100.00	100.00	0.00	100.00	100.00	66.00	0.00
EGARCH	0.00	100.00	16.00	0.00	0.00	100.00	0.00	0.00
NGSSM	0.00	84.00	100.00	0.00	34.00	100.00	100.00	0.00
SV	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
EGARCH DGP								
	APARCH	EGARCH	NGSSM	SV	APARCH	EGARCH	NGSSM	SV
APARCH	100.00	100.00	100.00	0.00	100.00	100.00	63.00	0.00
EGARCH	0.00	100.00	4.00	0.00	0.00	100.00	0.00	0.00
NGSSM	0.00	96.00	100.00	0.00	37.00	100.00	100.00	0.00
SV	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
NGSSM DGP								
	APARCH	EGARCH	NGSSM	SV	APARCH	EGARCH	NGSSM	SV
APARCH	100.00	100.00	100.00	0.00	100.00	100.00	100.00	0.00
EGARCH	0.00	100.00	1.00	0.00	0.00	100.00	1.00	0.00
NGSSM	0.00	99.00	100.00	0.00	0.00	99.00	100.00	0.00
SV	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
SV DGP								
	APARCH	EGARCH	NGSSM	SV	APARCH	EGARCH	NGSSM	SV
APARCH	100.00	100.00	100.00	0.00	100.00	100.00	100.00	0.00
EGARCH	0.00	100.00	1.00	0.00	0.00	100.00	0.00	0.00
NGSSM	0.00	99.00	100.00	0.00	0.00	100.00	100.00	0.00
SV	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
Set	C				D			
APARCH DGP								
	APARCH	EGARCH	NGSSM	SV	APARCH	EGARCH	NGSSM	SV
APARCH	100.00	100.00	8.00	0.00	100.00	100.00	95.00	0.00
EGARCH	0.00	100.00	1.00	0.00	0.00	100.00	0.00	0.00
NGSSM	92.00	99.00	100.00	0.00	5.00	100.00	100.00	0.00
SV	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
EGARCH DGP								
	APARCH	EGARCH	NGSSM	SV	APARCH	EGARCH	NGSSM	SV
APARCH	100.00	100.00	1.00	0.00	100.00	100.00	90.00	0.00
EGARCH	0.00	100.00	0.00	0.00	0.00	100.00	2.00	0.00
NGSSM	99.00	100.00	100.00	0.00	10.00	98.00	100.00	0.00
SV	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
NGSSM DGP								
	APARCH	EGARCH	NGSSM	SV	APARCH	EGARCH	NGSSM	SV
APARCH	100.00	100.00	92.00	0.00	100.00	100.00	99.00	0.00
EGARCH	0.00	100.00	4.00	0.00	0.00	100.00	4.00	0.00
NGSSM	8.00	96.00	100.00	0.00	1.00	96.00	100.00	0.00
SV	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
SV DGP								
	APARCH	EGARCH	NGSSM	SV	APARCH	EGARCH	NGSSM	SV
APARCH	100.00	100.00	79.00	0.00	100.00	100.00	99.00	0.00
EGARCH	0.00	100.00	2.00	0.00	0.00	100.00	0.00	0.00
NGSSM	21.00	98.00	100.00	0.00	1.00	100.00	100.00	0.00
SV	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00

Table 14. Detailed MSE comparison: number of times (in %) the model in a row outperformed the corresponding model in a column.

Set	A				B			
APARCH DGP								
	APARCH	EGARCH	NGSSM	SV	APARCH	EGARCH	NGSSM	SV
APARCH	100.00	43.00	79.00	0.00	100.00	45.00	69.00	0.00
EGARCH	57.00	100.00	77.00	0.00	55.00	100.00	66.00	0.00
NGSSM	21.00	23.00	100.00	0.00	31.00	34.00	100.00	0.00
SV	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
EGARCH DGP								
	APARCH	EGARCH	NGSSM	SV	APARCH	EGARCH	NGSSM	SV
APARCH	100.00	27.00	70.00	0.00	100.00	36.00	68.00	0.00
EGARCH	73.00	100.00	72.00	0.00	64.00	100.00	68.00	0.00
NGSSM	30.00	28.00	100.00	0.00	32.00	32.00	100.00	0.00
SV	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
NGSSM DGP								
	APARCH	EGARCH	NGSSM	SV	APARCH	EGARCH	NGSSM	SV
APARCH	100.00	35.00	44.00	0.00	100.00	24.00	37.00	0.00
EGARCH	65.00	100.00	49.00	0.00	76.00	100.00	57.00	0.00
NGSSM	56.00	51.00	100.00	0.00	63.00	43.00	100.00	0.00
SV	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
SV DGP								
	APARCH	EGARCH	NGSSM	SV	APARCH	EGARCH	NGSSM	SV
APARCH	100.00	38.00	48.00	0.00	100.00	32.00	30.00	0.00
EGARCH	62.00	100.00	56.00	0.00	68.00	100.00	40.00	0.00
NGSSM	52.00	44.00	100.00	0.00	70.00	60.00	100.00	0.00
SV	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
Set	C				D			
APARCH DGP								
	APARCH	EGARCH	NGSSM	SV	APARCH	EGARCH	NGSSM	SV
APARCH	100.00	45.00	48.00	0.00	100.00	44.00	72.00	0.00
EGARCH	55.00	100.00	54.00	0.00	56.00	100.00	69.00	0.00
NGSSM	52.00	46.00	100.00	0.00	28.00	31.00	100.00	0.00
SV	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
EGARCH DGP								
	APARCH	EGARCH	NGSSM	SV	APARCH	EGARCH	NGSSM	SV
APARCH	100.00	26.00	64.00	0.00	100.00	23.00	70.00	0.00
EGARCH	74.00	100.00	66.00	0.00	77.00	100.00	73.00	0.00
NGSSM	36.00	34.00	100.00	0.00	30.00	27.00	100.00	0.00
SV	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
NGSSM DGP								
	APARCH	EGARCH	NGSSM	SV	APARCH	EGARCH	NGSSM	SV
APARCH	100.00	13.00	47.00	0.00	100.00	27.00	33.00	0.00
EGARCH	87.00	100.00	65.00	0.00	73.00	100.00	56.00	0.00
NGSSM	53.00	35.00	100.00	0.00	67.00	44.00	100.00	0.00
SV	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
SV DGP								
	APARCH	EGARCH	NGSSM	SV	APARCH	EGARCH	NGSSM	SV
APARCH	100.00	17.00	41.00	0.00	100.00	30.00	26.00	0.00
EGARCH	83.00	100.00	54.00	0.00	70.00	100.00	39.00	0.00
NGSSM	59.00	46.00	100.00	0.00	74.00	61.00	100.00	0.00
SV	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00

Table 15. Detailed Unconditional VaR Coverage comparison: number of times (in %) the model in a row outperformed the corresponding model in a column.

Set	A				B			
APARCH DGP								
	APARCH	EGARCH	NGSSM	SV	APARCH	EGARCH	NGSSM	SV
APARCH	100.00	87.00	97.00	37.00	100.00	91.00	98.00	58.00
EGARCH	21.00	100.00	90.00	18.00	21.00	100.00	83.00	39.00
NGSSM	5.00	16.00	100.00	0.00	3.00	18.00	100.00	1.00
SV	71.00	83.00	100.00	100.00	49.00	67.00	99.00	100.00
EGARCH DGP								
	APARCH	EGARCH	NGSSM	SV	APARCH	EGARCH	NGSSM	SV
APARCH	100.00	87.00	98.00	36.00	100.00	95.00	96.00	51.00
EGARCH	18.00	100.00	94.00	22.00	9.00	100.00	84.00	28.00
NGSSM	2.00	8.00	100.00	1.00	5.00	19.00	100.00	6.00
SV	71.00	84.00	100.00	100.00	56.00	78.00	95.00	100.00
NGSSM DGP								
	APARCH	EGARCH	NGSSM	SV	APARCH	EGARCH	NGSSM	SV
APARCH	100.00	92.00	97.00	54.00	100.00	91.00	98.00	51.00
EGARCH	11.00	100.00	84.00	29.00	21.00	100.00	89.00	29.00
NGSSM	4.00	18.00	100.00	3.00	4.00	14.00	100.00	3.00
SV	52.00	76.00	97.00	100.00	57.00	80.00	99.00	100.00
SV DGP								
	APARCH	EGARCH	NGSSM	SV	APARCH	EGARCH	NGSSM	SV
APARCH	100.00	98.00	95.00	47.00	100.00	98.00	99.00	51.00
EGARCH	9.00	100.00	85.00	14.00	9.00	100.00	87.00	26.00
NGSSM	6.00	18.00	100.00	6.00	1.00	18.00	100.00	1.00
SV	58.00	89.00	96.00	100.00	55.00	79.00	100.00	100.00
Set	C				D			
APARCH DGP								
	APARCH	EGARCH	NGSSM	SV	APARCH	EGARCH	NGSSM	SV
APARCH	100.00	80.00	99.00	85.00	100.00	83.00	95.00	53.00
EGARCH	30.00	100.00	96.00	71.00	28.00	100.00	86.00	39.00
NGSSM	1.00	6.00	100.00	9.00	10.00	17.00	100.00	7.00
SV	18.00	32.00	96.00	100.00	54.00	64.00	94.00	100.00
EGARCH DGP								
	APARCH	EGARCH	NGSSM	SV	APARCH	EGARCH	NGSSM	SV
APARCH	100.00	93.00	97.00	78.00	100.00	85.00	94.00	59.00
EGARCH	18.00	100.00	94.00	64.00	29.00	100.00	91.00	44.00
NGSSM	3.00	8.00	100.00	10.00	10.00	11.00	100.00	2.00
SV	24.00	39.00	92.00	100.00	51.00	66.00	98.00	100.00
NGSSM DGP								
	APARCH	EGARCH	NGSSM	SV	APARCH	EGARCH	NGSSM	SV
APARCH	100.00	97.00	100.00	74.00	100.00	84.00	98.00	63.00
EGARCH	13.00	100.00	98.00	51.00	24.00	100.00	91.00	52.00
NGSSM	0.00	2.00	100.00	0.00	3.00	10.00	100.00	8.00
SV	28.00	55.00	100.00	100.00	45.00	54.00	96.00	100.00
SV DGP								
	APARCH	EGARCH	NGSSM	SV	APARCH	EGARCH	NGSSM	SV
APARCH	100.00	95.00	100.00	76.00	100.00	96.00	99.00	63.00
EGARCH	7.00	100.00	96.00	48.00	8.00	100.00	96.00	27.00
NGSSM	0.00	6.00	100.00	2.00	1.00	12.00	100.00	1.00
SV	30.00	55.00	98.00	100.00	44.00	79.00	99.00	100.00