Article:

Dimitrakopoulos, S orcid.org/0000-0002-0043-180X and Kolossiatis, M (2020) Bayesian analysis of moving average stochastic volatility models: modeling in-mean effects and leverage for financial time series. Econometric Reviews, 39 (4). pp. 319-343. ISSN 0747-4938

https://doi.org/10.1080/07474938.2019.1630075

Bayesian analysis of moving average stochastic volatility models: modelling in-mean effects and leverage for financial time series

Stefanos Dimitrakopoulos[1*] and Michalis Kolossiatis[**]

[*]Economics Division, Leeds University, Leeds, LS6 1AN, UK

[**]Department of Mathematics and Statistics, University of Cyprus, Nicosia, 1678, Cyprus

Abstract

We propose a moving average stochastic volatility in mean model and a moving average stochastic volatility model with leverage. For parameter estimation, we develop efficient Markov chain Monte Carlo algorithms and illustrate our methods, using simulated and real data sets. We compare the proposed specifications against several competing stochastic volatility models, using marginal likelihoods and the observed-data Deviance information criterion. We also perform a forecasting exercise, using predictive likelihoods, the root mean square forecast error and Kullback-Leibler divergence. We find that the moving average stochastic volatility model with leverage better fits the four empirical data sets used.

Keywords: in-mean effects, leverage, Markov chain Monte Carlo, moving average, stochastic volatility

1 Introduction

Stochastic volatility (SV) models (Taylor, 1986) have enjoyed great popularity in modelling financial time series over the last couple of decades. This class of models allows for time-varying variances (heteroscedastic errors), where the log-volatilities follow a first-order stationary autoregressive process.

In financial literature, various extensions of the SV model have been put forward. Among such extensions, the moving average component, the conditional heteroscedasticity in mean and the leverage effect are highlighted as important elements for capturing the behavior of financial data.

[1]Correspondence to: Stefanos Dimitrakopoulos, Economics Division, Leeds University Business School, Leeds University, UK, E-mail: s.dimitrakopoulos@leeds.ac.uk.

The moving average SV (MASV) model was introduced by Chan (2013), who assumed serially dependent errors in the measurement equation. In that paper, the MASV model was applied to inflation data and was found to provide better goodness of fit and out-of-sample forecasts than SV models without the moving average component. Also, a moving average component in SV models is important in modelling crude oil returns (Chan and Grant, 2016a) and S&P500 returns (Chan and Grant, 2016b).

The SV in mean (SVM) model was proposed by Koopman and Hol Uspensky (2002), who incorporated the latent volatility as an additional covariate in the conditional mean of the returns, in order to capture potential volatility feedback effects. Koopman and Hol Uspensky (2002) estimated their model parameters with a simulated maximum likelihood method, while Chan (2017) devised an MCMC algorithm, allowing also for time-varying parameters.

The SV model with leverage (SVL), due to Black (1976), captures the negative correlation between the returns today and the volatility tomorrow; a negative shock to returns at time t will lead to a larger volatility at time $t + 1$. Several MCMC methods have been designed for this model. For example, Omori et al. (2007) proposed an efficient mixture sampler, Omori and Watanabe (2008) developed a block sampler and Chan and Grant (2016b) worked with band and sparse matrix algorithms.

So far, the moving average component, the leverage effect and the conditional heteroscedasticity in mean have been considered separately in the stochastic volatility literature. Our first contribution is that in this paper we merge these strands of literature by setting up two novel moving average SV models for modelling financial return series. The first model specification is the moving average stochastic volatility model with conditional heteroscedasticity in mean, called the MASVM model. The second model specification is the moving average stochastic volatility model with leverage and we name it the MASVL model.

The resulting model specifications are more flexible and less susceptible to estimation bias, compared to their nested versions (MASV, SVL, SVM). For instance, in a simulation study we show that ignoring the moving average component from the MASVM and MASVL models, distorts important parameter estimates, such as the in mean and the leverage effects. Similarly, omitting the leverage or the in mean component from the MASVL and MASVM model, respectively, induces bias in the SV parameter estimates.

In addition, we investigate if the inclusion of both the moving average component and the leverage effect in the context of the SV model provides an improvement in in-sample fitness, as well as a better out-of-sample forecast performance than SV models with only the moving average or only the leverage component. Similarly, we examine if the moving average stochastic volatility in mean model better captures the behavior of return data, in terms of model fit and forecasting, than its nested versions, the MASV and SVM models.

The estimation of the proposed models is nontrivial. For the MASVM model, the volatility term appears both in the conditional mean and variance of the responses, while the errors in the measurement equation are serially dependent. For the MASVL model, the innovation errors are correlated with the moving average errors. As such, the efficient update of volatilities is not an easy task.

In the standard stochastic volatility model, the log-volatilities are latent parameters, leading to an intractable likelihood function. By augmenting the parameter space to include the latent volatilities, the (conditional on the extended parameter vector) likelihood function is then of known form, and standard Bayesian techniques can therefore be applied. Moreover, the Bayesian literature has proposed various efficient MCMC samplers for updating the volatilities, such as the popular algorithms of Shephard and Pitt (1997) and Kim et al. (1998). However, as we discuss in section 3, these samplers can not be directly implemented in the context of our models. To this end, we follow the method of Chan (2017), which is based on the precision sampler of Chan and Jeliazkov (2009).

The proposed models are illustrated with simulated data sets and empirical data sets taken from finance. For each empirical application, we compare the proposed models against several competing SV models that have been used in the financial econometrics literature. We conduct model comparison, using marginal likelihoods and the observed-data deviance information criterion of Chan and Grant (2016b). To evaluate the forecast performance of the proposed models, we compute point and density forecasts, as well as the Kullback-Leibler divergence between the forecast distributions and the kernel density estimate of left-out data.

The rest of the paper is structured as follows. In section 2 we set up the proposed models. In section 3 we outline the MCMC estimation algorithms and the model comparison criteria and the related estimation methods. Section 4 conducts the empirical analysis. Section 5 concludes. An Online Appendix accompanies this paper.

2 Modelling set up

2.1 The moving average stochastic volatility in mean model

Consider the following SV model

$$y_t = \mu + \lambda e^{h_t} + u_t,\ t = 1, ..., T, \tag{1}$$

$$u_t = \epsilon_t + \psi\epsilon_{t-1},\ |\psi| < 1,\ \epsilon_t \sim N(0, e^{h_t}) \tag{2}$$

$$h_{t+1} = \mu_h + \phi(h_t - \mu_h) + \eta_t,\ |\phi| < 1,\ \eta_t \sim N(0, \sigma_\eta^2). \tag{3}$$

In equation (1) y_t is the return, μ is a constant intercept and h_t is the log-volatility at time t. The exponential of log-volatility enters the conditional mean of the observation equation (1) as an additional explanatory variable and the scalar parameter λ captures the magnitude of the volatility feedback effect.

In equation (2) the error term u_t follows a first-order moving average (MA) process. The error terms ϵ_t are independent and identically normally distributed random variable with mean $E(\epsilon_t) = 0$ and variance $Var(\epsilon_t) = e^{h_t}$. This MA(1) process is initialized with $\epsilon_0 = 0$. For stationarity, we require $|\psi| < 1$.

The conditional variance of y_t is $Var(y_t|\mu, \lambda, \psi, \boldsymbol{h}) = e^{h_t} + \psi^2 e^{h_{t-1}}$, where $\boldsymbol{h} = (h_1, ..., h_T)$ is the vector of log-volatilities. Note that the time-varying conditional variance of y_t is attributed to the error variances of the MA process. In addition, even after conditioning on the parameters, the observed responses y_t are still serially correlated.

The volatility process is described by a stationary autoregressive process of order one, given in expression (3). The parameter $|\phi| < 1$ (for stationarity) measures the volatility persistence and the error variance σ_η^2 is the volatility of the log-volatility. The autoregressive process in (3) in initialized with $h_1 \sim N(\mu_h, \sigma_\eta^2/(1-\phi^2))$.

We name the model (1)-(3) the moving average stochastic volatility in mean (MASVM) model. The MASVM model reduces to the moving average stochastic volatility (MASV) model for $\lambda = 0$ and to the stochastic volatility in mean (SVM) model for $\psi = 0$.

2.2 The moving average stochastic volatility model with leverage

The moving average SV model with leverage in given by

$$y_t = \mu + u_t, \ t = 1, ..., T, \tag{4}$$

$$u_t = \epsilon_t + \psi \epsilon_{t-1}, \ |\psi| < 1, \tag{5}$$

$$h_{t+1} = \mu_h + \phi(h_t - \mu_h) + \eta_t, \ |\phi| < 1, \tag{6}$$

where the joint distribution of the errors ϵ_t and η_t is bivariate normal; namely,

$$\begin{pmatrix} \epsilon_t \\ \eta_t \end{pmatrix} \sim N_2\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} e^{h_t} & \rho\sigma_\eta e^{h_t/2} \\ \rho\sigma_\eta e^{h_t/2} & \sigma_\eta^2 \end{pmatrix}\right). \tag{7}$$

The degree of correlation between ϵ_t and η_t is reflected in the parameter ρ, which takes values in the interval $(-1, 1)$. When this correlation parameter is negative ($\rho < 0$), there is the so-called leverage effect; a decrease in the return is followed by an increase in volatility. Such empirical evidence can be found in numerous empirical studies (Harvey and Shephard, 1996; Omori et al. 2007; Nakajima and Omori, 2012). As before, in (4) y_t is the return and μ is a constant intercept. The assumptions for the MA process in (5) and SV process in (6) are the same as in the case of the MASVM model.

The model given by expressions (4)-(7) is referred to as the moving average stochastic volatility model with leverage (MASVL model). We note that the MASVL model includes as special cases the moving average stochastic volatility (MASV) model (for $\rho = 0$) and the stochastic volatility with leverage (SVL) model (for $\psi = 0$).

2.3 Prior specifications

We assume the following priors for the common parameters of the MASVM and MASVL models: $\mu \sim N(\mu_0, V_\mu), \mu_h \sim N(\mu_{h0}, V_{\mu_h}), \phi \sim N(\phi_0, V_\phi)1_{(-1<\phi<1)}, \psi \sim N(\psi_0, V_\psi)1_{(-1<\psi<1)}, \sigma_\eta^2 \sim IG(\nu_h, S_h)$. We assume normal prior distributions for the intercepts μ, μ_h. The prior of ϕ is a normal distribution, truncated to the stationary region of ϕ's parameter space[1]. The same prior is used for the moving average parameter ψ. For σ_η^2 we assume an inverse gamma prior.

For the in-mean parameter λ of the MASVM model we assume a normal prior,

$$\lambda \sim N(\lambda_0, V_\lambda).$$

For the correlation parameter ρ of the MASVL model we assume a normal distribution, truncated in (-1,1):

$$\rho \sim N(\rho_0, V_\rho)1_{(-1<\rho<1)}.$$

As can be seen, the priors for the parameters are assumed to be independent. One, though, could impose joint priors on various parameters. For instance, Jacquier et al. (2004) considered a prior distribution over (ρ, σ_η^2) for the SVL model, while Lopes and Polson (2010) assumed a prior distribution over (ϕ, σ_η^2) for the SV model. See also Leão et al. (2017). In the context of our analysis, one could impose dependence between λ and ϕ. Such a dependence could be justified by the fact that the more persistent volatile market could have higher impact on the effect of volatility on asset returns. This joint prior[2] on (λ, ϕ) is examined in a simulation study; see Online Appendix.

[1] Another choice of prior for ϕ is the (shifted, scaled) beta distribution, as in Nakajima and Omori (2012). Although the beta prior requires no truncation, the normal distribution offers easier interpretability of its hyperparameters. In addition, when using the beta prior, the simulation results remain essentially the same; see Online Appendix.

[2] We thank a referee for pointing out this issue.

3 Posterior analysis

3.1 MCMC algorithm

For such complicated Bayesian models, the posterior distributions of the parameters are not known explicitly, so we resort to the MCMC simulation method. The most challenging part in the design of the MCMC algorithms for our proposed models is the update of the volatility vector. The auxiliary mixture algorithm of Kim et al. (1998) can not be applied to the MASVM model, as the volatility component enters the conditional mean and as such, this model can not be represented by a linear state-space model. On the other hand, the Shephard and Pitt (1997) algorithm can be applied to the MASVM model, but not to the MASVL model, as it assumes that the observation vector and the volatility vector are conditionally independent. Omori and Watanabe (2008) extended the Shephard and Pitt (1997) sampler for SV models with leverage effects.

Here, in order to efficiently sample the volatility vector, we adopt a more direct approach based on the precision sampler of Chan and Jeliazkov (2009), that works only with sparse matrices. As opposed to the methods of Shephard and Pitt (1997) and Omori and Watanabe (2008), this sampler does not utilize Kalman filter techniques. As has been shown (Chan and Jeliazkov, 2009; McCausland et al. 2011), precision-based algorithms are computationally more efficient than standard Kalman-filter-based methods. We examine these claims in a simulation study (see Online Appendix, section 2.3) that compares our algorithms against those based on Shephard and Pitt (1997) for the MASVM model and on Omori and Watanabe (2008) for the MASVL model. Comparison is in terms of both efficiency and computational time.

The parameter vector for the MASVM model is $(\boldsymbol{\theta}, \boldsymbol{h})$, where $\boldsymbol{\theta} = \left(\mu, \mu_h, \phi, \sigma_\eta^2, \psi, \lambda\right)$. All full conditional distributions are of known form, except the ones for ϕ, ψ and $\boldsymbol{h}$. For the first two we use Metropolis-Hastings steps. For the parameter vector $\boldsymbol{h}$, we follow the method of Chan (2017), which is based on the precision sampler of Chan and Jeliazkov (2009). We also note that μ and λ are jointly updated from a bivariate normal distribution.

The parameter vector for the MASVL model is $(\boldsymbol{\theta}, \boldsymbol{h})$, where $\boldsymbol{\theta} = \left(\mu, \mu_h, \phi, \sigma_\eta^2, \psi, \rho\right)$ and $\boldsymbol{h} = (h_1, ..., h_{T+1})$ (notice that here $\boldsymbol{h}$ is of length $T+1$). All the full conditional distributions are of known form, except the ones for ϕ, ψ, ρ and $\boldsymbol{h}$. For the first two we use Metropolis-Hastings steps. For ρ, although its full conditional distribution is not of known form, the fact that this parameter is defined in the region $(-1, 1)$ allows for the use of Griddy-Gibbs sampling (see, for example, Ritter and Tanner (1992)). Finally, we update $\boldsymbol{h}$ similarly to the MASVM model.

The updating steps for the two algorithms, as well as various simulation studies, are presented in the Online Appendix.

3.2 Model comparison

In order to compare the models stated above, we will use two different criteria: the marginal likelihood and the observed-data deviance information criterion (Chan and Grant, 2016b).

3.2.1 Marginal likelihood

This is an in-sample prediction criterion, which measures the model fit to the data in hand (larger values indicate better model fit). For model $\mathcal{M}$ with observed-data likelihood $p(\boldsymbol{y}|\mathcal{M},\boldsymbol{\theta})$, where $\boldsymbol{y}$ is the data vector, and prior $p(\boldsymbol{\theta}|\mathcal{M})$, the marginal likelihood (ML) is defined as

$$p(\boldsymbol{y}|\mathcal{M}) = \int p(\boldsymbol{y}|\mathcal{M},\boldsymbol{\theta})p(\boldsymbol{\theta}|\mathcal{M})d\boldsymbol{\theta}. \tag{8}$$

Unfortunately, for the models we consider, expression (8) does not have closed form. To compute this, we utilize the Importance Sampling (IS) method of Chan and Eisenstat (2015), which itself is based on cross-entropy ideas. An importance sampling estimator of expression (8) is given by

$$\widehat{p(\boldsymbol{y}|\mathcal{M})} = \frac{1}{M_1}\sum_{i=1}^{M_1}\frac{p(\boldsymbol{y}|\mathcal{M},\boldsymbol{\theta}^{(i)})p(\boldsymbol{\theta}^{(i)}|\mathcal{M})}{g_1(\boldsymbol{\theta}^{(i)})}, \tag{9}$$

where $g_1(\cdot)$ is the importance density and $\boldsymbol{\theta}^{(i)}$ is the i^{th} independent draw from $g_1(\cdot)$, for $i = 1, ..., M_1$.

- MASVM model: As mentioned above, $\boldsymbol{\theta} = \left(\mu, \mu_h, \phi, \sigma_\eta^2, \psi, \lambda\right)$ for this model. The function g_1 consists of the product of independent distributions for each parameter, normal for the ones defined in $\mathbb{R}$, truncated normal for those that are defined in $(-1, 1)$ and inverse gamma for the positive ones:

  $$\begin{aligned} g_1(\boldsymbol{\theta}) &= N(\mu;\hat{\mu},S_\mu)\times N(\mu_h;\hat{\mu_h},S_{\mu_h})\times N(\phi;\hat{\phi},S_\phi)1_{(-1<\phi<1)} \\ &\times IG(\sigma_\eta^2;\hat{\sigma_\eta^2},S_{\sigma_\eta^2})\times N(\psi;\hat{\psi},S_\psi)1_{(-1<\psi<1)}\times N(\lambda;\hat{\lambda},S_\lambda), \end{aligned}$$

  where in all the above, $\hat{c}$ and S_c denote the posterior mean and variance for parameter c, respectively, obtained from the MCMC output.

  Having obtained independent draws $\boldsymbol{\theta}^{(1)}$,..., $\boldsymbol{\theta}^{(M_1)}$ from $g_1(\cdot)$ we calculate expression (9). Yet, the observed-data likelihood in (9) is an intractable integral, that is, $p(\boldsymbol{y}|\mathcal{M},\boldsymbol{\theta}) = \int p(\boldsymbol{y}|\mathcal{M},\boldsymbol{\theta},\boldsymbol{h})p(\boldsymbol{h}|\boldsymbol{\theta},\mathcal{M})d\boldsymbol{h}$. To overcome this problem, we use the Importance Sampling method (Chan and Eisenstat, 2015) again. In particular, an Importance Sampling estimator of the observed-data likelihood is

  $$\widehat{p(\boldsymbol{y}|\mathcal{M},\boldsymbol{\theta})} = \frac{1}{M_2}\sum_{i=1}^{M_2}\frac{p(\boldsymbol{y}|\mathcal{M},\boldsymbol{\theta},\boldsymbol{h}^{(i)})p(\boldsymbol{h}^{(i)})|\mathcal{M},\boldsymbol{\theta})}{g_2(\boldsymbol{h}^{(i)})}, \tag{10}$$

  where $\boldsymbol{h}^{(i)}$ are independent samples from another importance density $g_2(\cdot)$, $i = 1, 2, \ldots, M_2$. We want $g_2(\cdot)$ to be close to the conditional distribution of the volatility vector $p(\boldsymbol{h}|\boldsymbol{y},\boldsymbol{\theta})$, which is the theoretical zero-variance importance density. The density $p(\boldsymbol{h}|\boldsymbol{y},\boldsymbol{\theta})$ can be approximated by a Gaussian density (see Online Appendix) and we use this approximate density as the proposal density in the second IS step (in (10)).

- MASVL model: The same principles as above hold for this model, too (where now $\boldsymbol{\theta} = (\mu, \mu_h, \phi, \sigma_\eta^2, \psi, \rho)$). For the distribution of ρ in g_1, we take $N(\rho; \hat{\rho}, S_\rho)1_{(-1<\rho<1)}$.

3.2.2 Observed-data deviance information criterion

An alternative model comparison criterion is the conditional deviance information criterion (DIC) of Spiegelhalter et al. (2002) that accounts for model fit and model complexity. It is defined as

$$DIC = \overline{D(\boldsymbol{\Theta})} + p_D, \tag{11}$$

where $\overline{D(\boldsymbol{\Theta})} = -2\mathbf{E}_{\boldsymbol{\Theta}}[\log f(\boldsymbol{y}|\boldsymbol{\Theta})|\boldsymbol{y}]$ is the posterior mean deviance and $\boldsymbol{\Theta} = (\boldsymbol{\theta}, \boldsymbol{h})$ is the joint vector of parameters and volatilities for the SV models we examine. Model fit is measured by the deviance $D(\boldsymbol{\Theta})$, where $D(\boldsymbol{\Theta}) = -2\log f(\boldsymbol{y}|\boldsymbol{\Theta})$ and $logf(\boldsymbol{y}|\boldsymbol{\Theta})$ is the conditional log-likelihood function. Model complexity is measured by the effective number of model parameters p_D and is defined as

$$p_D = \overline{D(\boldsymbol{\Theta})} - D(\overline{\boldsymbol{\Theta}}), \tag{12}$$

where $D(\overline{\boldsymbol{\Theta}}) = -2\log f(\boldsymbol{y}|\overline{\boldsymbol{\Theta}})$ and $\log f(\boldsymbol{y}|\overline{\boldsymbol{\Theta}})$ is the conditional log-likelihood evaluated at $\overline{\boldsymbol{\Theta}}$, the posterior mean of $\boldsymbol{\Theta}$.

Chan and Grant (2016b) showed that for stochastic volatility models the DIC, which is calculated from the conditional likelihood (where the latent variables are conditioned on), is biased, favouring overfitted models. The DIC which is calculated from the observed-data likelihood does not suffer from this problem. Therefore, we use the observed-data DIC, which is given by

$$DIC_{obs} = -4\mathrm{E}_{\boldsymbol{\theta}}(\log f(\mathbf{y}|\boldsymbol{\theta})) + 2\log f(\mathbf{y}|\hat{\boldsymbol{\theta}}), \tag{13}$$

where $\boldsymbol{\theta}$ is the parameter vector $\hat{\boldsymbol{\theta}}$ is the posterior mode of $\boldsymbol{\theta}$ and $\mathrm{E}_{\boldsymbol{\theta}}(\log f(\mathbf{y}|\boldsymbol{\theta}))$ is the expectation of the logarithm of the observed-data likelihood. The observed-data likelihood is estimated using the importance sampling method described in equation (10). Lower values of DIC_{obs} indicate better model fit.

3.3 Forecast evaluation

We also conduct a recursive out-of-sample forecasting exercise to evaluate the predictive performance of the proposed models. In particular, for each model we compute the one-step ahead predictive likelihood of the observation y_{t_0+1}, conditional on the previous observations $\boldsymbol{y}_{t_0} = (y_1, ..., y_{t_0})$, that is $f(y_{t_0+1}|\boldsymbol{y}_{t_0})$. This predictive likelihood, which is used to evaluate the density forecast for the observation at t_0+1, is given by

$$f(y_{t_0+1}|\boldsymbol{y}_{t_0}) = \int f(y_{t_0+1}|h_{t_0+1}, \boldsymbol{\theta}, h_{t_0}) f(h_{t_0+1}, \boldsymbol{\theta}, h_{t_0}|\boldsymbol{y}_{t_0}) d\boldsymbol{\theta} dh_{t_0} dh_{t_0+1}. \tag{14}$$

This expression can be approximated by Monte Carlo integration, in particular

$$\widehat{f(y_{t_0+1}|\boldsymbol{y}_{t_0})} = \frac{1}{R}\sum_{i=1}^{R} f(y_{t_0+1}|h_{t_0+1}^{(i)}, \boldsymbol{\theta}^{(i)}, h_{t_0}^{(i)}), \tag{15}$$

where $\boldsymbol{\theta}^{(i)}$, $h_{t_0}^{(i)}$ and $h_{t_0+1}^{(i)}$ are draws obtained from the posterior sampler at iteration $i = 1, ..., R$.

For the MASVM model, future log-volatilities h_{t_0+1} are drawn from $N(\mu_h + \phi(h_{t_0} - \mu_h), \sigma_\eta^2)$. For the MASVL model, the h_{t_0+1} value is already available from the MCMC sampler (since for the MCMC algorithm for this model, this value is included in the vector of log-volatilities). In both cases, y_{t_0+1} is a normally distributed random variable given by expression (1) for the MASVM model or expression (4) for the MASVL model (where we use the corresponding h_{t_0+1}).

For the evaluation period $t \in \{t_0+1, ..., T\}$, the sum of the log predictive likelihoods $\sum_{t=t_0}^{T-1} \log f(y_{t+1}|\boldsymbol{y}_t)$ is known as the log predictive score of the model. Higher values entail better (out-of-sample) forecasting ability of the model. We report the log predictive score of the competing models in all our empirical applications.

We also compute the one-step ahead predictive means $E(y_{t+1}|\boldsymbol{y}_t)$, $t \in \{t_0+1, ..., T\}$ as point forecasts. A usual metric to evaluate point forecasts is the root mean squared forecast error (RMSFE) defined as

$$RMSFE = \sqrt{\frac{\sum_{t=t_0}^{T-1}(y_{t+1} - E(y_{t+1}|\boldsymbol{y}_t))^2}{T - t_o}}. \tag{16}$$

Lower values of the RMSFE indicate better point forecasts.

As a forecast density evaluation measure, we calculate the Kullback-Leibler divergence between the forecast density $f(y_{t_0+1}|\boldsymbol{y}_{t_0})$ and the kernel density of the left-out observations $g(y_{t_0+1}, \ldots, y_T)$, as follows:

$$KLD(f,g) = \int f(y_{t_0+1}|\boldsymbol{y}_{t_0}) \log \frac{f(y_{t_0+1}|\boldsymbol{y}_{t_0})}{g(y_{t_0+1}, \ldots, y_T)} dy = E_y[\log f(y_{t_0+1}|\boldsymbol{y}_{t_0})] - E_y[\log g(y_{t_0+1}, \ldots, y_T)] \tag{17}$$

Standard Monte Carlo simulation techniques are used for the approximation of this integral. The smaller the KLD value, the closer f will be to g.

4 Empirical applications

In the following subsections we consider four series of return data and compare the proposed models (MASVM and MASVL) against their nested versions; the MASV, SVL and SVM models. We also consider the stochastic volatility in mean model with leverage (SVML model) that has been considered by Abanto-Valle et al. (2011). This model is described in the Online Appendix. For completeness, we also report the results for the standard stochastic volatility (SV) model.

For each empirical application, we run each model for 100000 iterations, after a burn-in period of 80000 draws. To compute the observed-data likelihood, we sample $M_1 = 1000$ and $M_2 = 50$ draws from the importance densities. Regarding the observed-data DIC, we run 10 parallel chains and then we took the average of these values. To obtain the numerical standard error for the estimated observed-data DIC values, we divided the standard deviation of the sampled DIC estimates by $\sqrt{10}$.

For all models we used the priors of section 2.3. The hyperparameters for the priors are as follows:

$$\mu_0 = 0, V_\mu = 10, \mu_{h0} = 1, V_{\mu_h} = 10, \phi_0 = 0, V_\phi = 1,$$

$$\nu_h = 5, S_h = 0.16, \psi_0 = 0, V_\psi = 100, \lambda_0 = 0, V_\lambda = 10000, \rho_0 = 0, V_\rho = 1.$$

To monitor the performance of our sampling efficiency, we estimated the inefficiency factor (IF) that measures how well the MCMC chain mixes. The IF is defined as $1 + \sum_{s=1}^{\infty} \varrho_s$, where ϱ_s is the sample autocorrelation at lag s; see also Chib (2001). The IF quantifies the relative efficiency loss due to the correlation in the samples obtained. A well designed posterior algorithm will generate low correlation across draws and therefore a low IF.

To check convergence, we also computed the Convergence Diagnostics (CD) statistic of Geweke (1992). This statistic compares draws in the early part of the chain to those in the last part of the chain, so as to detect problems of convergence (after burn-in). Lower absolute values of CD statistic indicate better convergence.

4.1 Application I: Equity Hedge

In our first empirical application we use daily returns on the hedge fund, Equity Hedge, from April 1, 2003 to May 31, 2010. The period of analysis yields T=1870 observations[3]. Figure 1 presents the time series plot of the data.

4.1.1 Estimation results

Figures 2 and 3 display the posterior autocorrelation functions (top panel), the posterior paths (middle panel) and the posterior histograms (bottom panel) for the parameters of the MASVM and MASVL models, respectively. For both these two models the posterior paths are stable and the posterior autocorrelations decay rapidly, suggesting that the proposed MCMC algorithmic schemes are efficient.

In Table 1 we present the results of the posterior means, standard deviations, inefficiency factors (IF) and CD statistics for the seven models in question, using the full sample. From the CD values of

[3]This data set has been used in the textbook of Martin et al. (2012) and can be downloaded from that textbook's website: http://www.cambridge.org/features/econmodelling/exercises.htm.

Table 1, the produced sequences of MCMC draws converge for all parameters of the proposed models (as well as of the rest of the models). Similarly, according to the IF values, the proposed algorithms produce a good mixing of the corresponding MCMC chains.

All the parameters across all models of Table 1 are statistically significant. The posterior estimate (posterior mean) of ψ is similar for the models that incorporate the moving average component (MASVM, MASVL, MASV), with their values being between 0.175 and 0.191. Therefore, there is positive autocorrelation in the observed sequence of Equity Hedge's returns. The 95% credible intervals for the MA parameter were estimated to be (0.1283, 0.22129) and (0.14565, 0.23643) for the MASVM and MASVL models, respectively, and both these two intervals exclude zero. Figures 2 (bottom panel for the MASVM model) and 3 (bottom panel for the MASVL model) tell the same story; the posterior histograms of ψ are concentrated around 0.1. Taken together, these empirical findings suggest the importance of extending the SVM and SVL models to include a moving average process. This conclusion is in agreement with the model comparison results (see next subsection).

For the SVL-type models (MASVL, SVL, SVML), the correlation coefficient ρ was found to be negative, implying the existence of leverage effect in the data. This parameter has the largest absolute value for the MASVL model and the smallest for the SVML model. Hence, the leverage effect is relatively more strong and important in the MASVL model than in the rest SVL-type models. Furthermore, as can be seen from Figure 3 (bottom panel), the posterior histogram of ρ for the MASVL model is situated in the negative range. The 95% credible interval of ρ for this model is (-0.60182 -0.29325) and does not contain zero, an additional indication that the parameter ρ is significant.

The in-mean parameter λ had a negative posterior mean[4] of about -0.4 across all models that contain in-mean effects (MASVM, SVM and SVML). Regarding this coefficient in the MASVM model, its posterior histogram is located in the negative range and its 95% posterior credibility interval was estimated to be (-0.69438, -0.26566). These results highlight that volatility feedback effect should not be ignored when modelling Equity Hedge's returns[5].

According to Koopman and Hol Uspensky (2002), λ measures both the volatility feedback effect and the ex-ante relationship between returns and volatility. The volatility feedback theory is based on two assumptions. The first one is that volatility is persistent and the second one is that there is a positive relation between expected returns and the volatility process. Under these two assumptions, λ is expected to be negative, as is in our case. The intuition is that an unanticipated large shock to the return process (h_t), due to good or bad news, causes investors to expect higher persistent levels

[4]The parameter λ has also been found negative in other studies of stock returns (Koopman and Hol Uspensky, 2002; Abanto-Valle et al. 2011).

[5]In the Online Appendix, we also rerun the MASVM model using the joint prior on (ϕ, λ), without observing any substantial changes in the results.

of volatility in the future (due to the first assumption). As such, due to the second assumption, risk-averse investors require a compensation for this in the form of higher expected future returns (French et al. 1987), which is achieved by a drop in current log returns (y_t).

The volatility feedback effect is one explanation for the asymmetric volatility argument, according to which there is a negative relationship between unexpected returns and innovations to the volatility process. The second explanation is the leverage effect. As noted by French et al. (1987) and Schwert (1989), among others, the leverage alone can not capture the magnitude of the negative relationship. For instance, Campbell and Hentschel (1992) and Bekaert and Wu (2000) found evidence of both volatility feedback and leverage effects. The significance of both λ and ρ parameters in our SVM-type and SVL-type models confirm these empirical findings in the literature.

In Figure 4 we plotted the posterior means of $\exp(h_t)$ for the MASVM model, along with the ones for its nested versions. Under all three models, there is apparent variation in the returns' volatility estimates, suggesting that it was worth allowing for conditional heteroscedasticity. Similar analysis holds for the plotted means of $\exp(h_t)$ for the MASVL model and its nested versions; see Figure 5. Furthermore, from the same figures, we observe that the returns show high volatility towards the end of 2008 and beginning of 2009, the time where the recent Global Crisis began to show down.

4.1.2 Model comparison results

Table 1 reports the logarithm of the estimated marginal likelihoods (log ML) and the estimated observed-data DIC (DIC_{obs}) values, along with their numerical standard errors. Under both model comparison criteria, the MASVL model is the most preferred, as it has the highest log ML and the lowest DIC_{obs}. This indicates that the SV model with both the MA component and leverage increases the in-sample fit more than the SV model with only the MA component or with only the leverage.

Furthermore, based on the reported log ML and DIC_{obs} values, the second best model is the MASVM model that controls for the MA term and in-mean effects. Therefore, the inclusion of both these two factors in the SV model contributes more to model fit than the inclusion of only the MA term or only the in-mean variable.

At this point it is important to note that the significance of the parameters ψ, λ, and ρ in all models supports the ranking of the log marginal likelihood and observed-data DIC, regarding the superiority of the MASVM and MASVL models over their nested versions.

Both criteria also agree that the MASV model is the third best model and outperforms the SVM and SVL models[6], as well as their combination (the SVML model). Therefore, there is strong evidence

[6]This empirical finding is in agreement with the results from the simulated study, according to which the MA part increases the model fit more than the in-mean effect or leverage, when the true models are the two proposed ones.

that the empirical data prefer SV variants with the MA term.

The SVML model, which possesses the fourth best place, is favoured over its nested models (SVL and SVM). The fifth best place is less clear-cut; according to the log ML, the SVL model performs better than the SVM, whereas according to the DIC_{obs}, the SVM is more preferred than the SVL model. The worst model is the SV model.

Lastly, using the observed-data likelihood, we calculated and reported the effective number of parameters $p_{D_{obs}}$ that measures model complexity. When the likelihood dominates prior information, one can show that $p_{D_{obs}}$ is close to the actual number of model parameters, with the difference reflecting the quantity of prior information (Li et al. 2012). The least complex models are the MASVL and the SV models which gave the same $p_{D_{obs}}$ value. The rest of the models yield similar degrees of model complexity, with the highest being for the SVML and SVM models.

4.1.3 Forecasting results

We also compared the forecasting performance of the seven models, using log predictive scores (LPS) and RMSFEs. For this out-of-sample forecasting exercise the evaluation period is from January 11, 2010 to the end of the sample, May 31, 2010 *i.e.* for the last 100 data points. For each such point, we sampled each model's unknown parameters 20000 times, after discarding the first 20000 draws. The results are presented in Table 1.

Regarding density forecasts, the MASVL model is the best model, however its nested version MASV produces the second worst density forecasts, with the worst model being the SV model. The SVL model, the other nested version of the MASVL model, has the second best forecast performance. The SVML is the third best option. So, models that contain the leverage component dominate the ones that do not, in terms of density forecast performance. This result is different from the two model comparison criteria (log ML and DIC_{obs}), which favour MASV-type models (MASVM, MASVL, MASV). Also, the MASVM model does worse, in terms of forecasting performance, than the SVML model but outperforms all the other models.

As far as point forecasts are concerned, the MASVL model delivered the lowest RMSFE, whereas the MASV produced the second best point forecast, followed by the MASVM model that beats the rest of the models. Contrary to the ranking induced by the density forecast results, the RMSFE results seem to be closer to the results obtained from the log ML and DIC_{obs} , in the sense that the latter also support the MASV variants (MASVL, MASVM, MASV) over the rest of the models. As was also the case for the density forecasts, the SV is again the worst point forecast model.

In the same table (last line) we display the results from the Kullback-Leibler divergence. This measure clearly favours the MASVL model, whereas the SV model produces a forecasting density which is the farthest from the kernel density of the left-out observations (as was also the case for the LPS and RMSFE ranking). The second and third best positions, based on the KLD values, are assigned to the MASV and SVL, respectively. The MASVM model is next in ordering and does better than the SVML and SVM models.

In any case, the proposed MASVL model dominates all models, both in terms of goodness of fit and out-of-sample forecasting ability.

4.2 Application II: S&P 500 index

Our second empirical set consists of T=2500 daily returns (values are given in percentage points) on the S&P 500 index over the period of January 2, 1970 to November 21, 1979. Figure 6 plots the returns and Table 2 presents the empirical results.

4.2.1 Estimation results

The intercept μ of the measurement equation was found insignificant for all but the MASVM and SVM models. This is not the case for the MA parameter ψ and the correlation parameter ρ, that were both found to be significant inclusions in the stochastic volatility model. The parameters ruling the stochastic volatility process $(\mu_h, \phi, \sigma_\eta^2)$ are significant across all models. From the reported IF and CD values, as well as the plotted posterior paths and posterior autocorrelations (Figures 7 and 8), there seem to be no mixing or convergence problems with the produced Markov chains.

The posterior mean of ρ is negative, signalling the presence of leverage effects. Furthermore, it is the largest in absolute value for the MASVL model (and the smallest for the SVML model), as in the case of the Equity Hedge returns. The absolute values of leverage effect are larger, and therefore leverage is stronger, in this application, compared to the previous one.

The positive sign of the parameter ψ indicates positive serial error correlation, which is the strongest in the MASVL model. In addition, the estimated value of ψ for the MASV-type models is larger in this empirical application than in the previous one.

In the SVM-type models the estimate of λ is negative[7]. For this parameter in the MASVM model, the zero is barely contained in the 95% credible interval, which is (-0.17724, 0.030911), and its posterior histogram is mainly located in the negative range (Figure 7, bottom panel). Very similar (negative and weak) values of λ have been found in previous studies that have also analysed daily S&P500

[7] In the Online Appendix, we also run the MASVM model using the joint prior on (ϕ, λ), for all the empirical data sets, without observing any substantial changes in the results.

returns (for example, Koopman and Hol Uspensky, 2002). In terms of most parameter estimates, the MASVM and MASV models are very similar.

It is also important to compare the relationship between λ and ϕ between the two empirical studies we considered so far. French et al. (1987), as well as Koopman and Hol Uspensky (2002), pointed out that the volatility feedback effect is larger when the log-volatilities are highly correlated. In the second empirical application ϕ attains larger values than those in the first application and this is in accordance with the larger estimated values of λ in the second data set.

Figures 9 and 10 show the paths of the posterior means of $\exp(h_t)$ for the MASVM and MASVL models and their nested versions, respectively. In either figure, the evolution of $\exp(h_t)$ signals the presence of conditional heteroscedasticity.

4.2.2 Model comparison results

Based on the results of Table 2, both model comparison criteria agree that the model with the best fit to the data is the MASVL model. Also, the MA-type SV models (MASVM, MASVL, MASV) are preferred to the rest of the models. That was also the case in the previous empirical application.

In particular, according to the log marginal likelihood, the second best in-sample fit is achieved by the MASV model and not the MASVM model. On the contrary, the observed-data DIC selects the MASVM model as the second best model. This ambiguity as to which of the MASV and MASVM models is preferred can be attributed to the fact that λ is just barely insignificant. When $\lambda = 0$, the MASVM model is reduced to the MASV model, and the small differences in log marginal likelihood and observed-data DIC values for the two models might be due to statistical error.

For the non-MA variants of the SV model (SVM,SVL, SV and SVML), the SVL model performs better, followed by the SVML model, under both criteria. The worst model is the SVM model according to the log marginal likelihood, or the SV model according to the observed-data DIC.

Based on the $p_{D_{obs}}$ values that measure model complexity, the most complex models are the SVM and SVML, while the least complex models are the MASVL and SV; see Table 2.

4.2.3 Forecasting results

For our forecasting exercise we use as evaluation period the last 100 data points; that is, from July 02, 1979 to the end of the sample. The results for point and density forecasts are also presented in Table 2.

From the reported LPS values, it can be seen that the MASVL is the best density forecast model, followed by the MASV model. The SVM model attains the third best position, with the fourth and fifth position being occupied by the SVML and MASVM models, respectively. The worst two models, in terms of density forecast performance, are the SVL and the SV models.

In terms of point forecasts, the best model is the MASVL. The MASVM is the third best, being outperformed by the SVM model. The MASV and SVML produced the same RMSFEs, occupying fourth place. The worst point forecast model is the SV.

Based on the Kullback-Leibler divergence results, the MASVL model is the winner over the rest of the models, whereas the MASVM model, being in the fourth position, is dominated by the SVL and SVM models. The MASV model produces a forecasting distribution which is further away from the kernel density of the left-out observations than the forecasting distribution of the MASVM model, but does better than the SVML and SV models.

In conclusion, we again see that the proposed MASVL model is the best model, in terms of all the model comparison criteria used.

4.3 Two additional empirical data sets

To further illustrate the proposed Bayesian methodology of this paper we turned our attention to two other types of return data: exchange rate returns and energy returns.

4.3.1 Foreign exchange returns

Figure 11 depicts the daily returns of Philippine peso (PHP) against the US dollar from July 2007 to December 2012. In total, we have $T = 1436$ observations. Table 3 reports the posterior means and standard deviations of model parameters. The IF and CD statistics for the parameters of the proposed models, along with the posterior plots in Figures 12 and 13, show that there are no mixing or convergence problems with the corresponding MCMC algorithms.

Model comparison results (log ML and DIC_{obs}) act in accordance (except for the ranking of the SVL and SV models); see Table 3. Once again, the MASV-type variants outperform the other models; one can find the MASVL, MASV and MASVM models in the first, second and third best position, respectively.

As seen in Table 3, the MASVL model produced the smallest LPS value, whereas the MASVM model was the second worst density forecast model (the worst being the SV model). The SVL gave the second best density forecast, followed by the SVM model, which outperforms the rest of the models.

The best point forecast model is the MASVM model and the second best is the MASVL model. The third position is assigned to the MASV model. The SVM does worse than the MASV model but does better than the SVML, SVL and SV. The latter yielded the largest RMSFE value.

The Kullback-Leibler divergence results favour the MASVL (best) and the MASV (second best) over the rest of the models. Among these models, the SVL model is closer to the kernel density of the left-out observations and the next models in ranking are the MASVM and SVML models. The two models with the furthest distance from that kernel density are the SVM and SV.

Our empirical analysis (Figures 14 and 15) reveals that there were two big spikes in the path of the estimated volatilities for the PHP/USD returns since the advent of the recent financial crisis. The first one took place in 2009 and the second one in mid-2010.

4.3.2 Energy returns

Our last data set consists of weekly petroleum prices of US Gulf Coast Conventional Gasoline Regular, spanning the period from January 3, 1997 to February 6, 2015. These prices were obtained from the US Energy Information Administration and were transformed into petroleum returns, by taking the first difference of the logs of this energy price series and multiplying it by 100. The time series plot of the $T = 936$ energy returns is presented in Figure 16.

The posterior paths and the posterior autocorrelations for the MASVM (Figure 17) and the MASVL (Figure 18) models verify that the corresponding algorithms perform efficiently. The estimation, model comparison and forecasting results are displayed in Table 4.

According to the log ML and DIC_{obs} values, the MASVL model is dominant, in terms of model fit, whereas the MASVM is third, with the second best being the MASV. Hence, the MASV-type models do better than the rest of the models. This behaviour was also observed in the previous empirical applications. The two model comparison criteria diverge only on the ranking of the SVM and SVL models.

In terms of point and density forecasting, the MASVL model dominates again, with the SVL and MASV being the second best models in the LPS and RMSFE ranking, respectively. The results based on the Kullback-Leibler divergence favor the MASVL model, with the MASVM and MASV models possessing the second and third places, respectively.

The estimated volatilities $\exp(h_t)$ for the proposed models and their nested versions are given in Figures 19 and 20. Clearly, the volatility of the weekly returns increases substantially in 2009, as the result of the 2008 global financial crisis.

5 Conclusions

In this paper we proposed two novel Bayesian moving average stochastic volatility models. The first one is a moving average stochastic volatility model with conditional heteroscedasticity in mean and the second is a moving average stochastic volatility model with leverage. Efficient Markov chain Monte Carlo algorithms were designed for inference, model comparison and forecasting for these two models. We demonstrated our methodologies with four empirical applications, involving daily return series. The results revealed that the moving average stochastic volatility model with leverage is preferred in terms of model fit and forecast performance over several competing stochastic volatility models.

Of course, the literature on stochastic volatility models is vast, and so is the number of possible extensions to these models. One extension is to consider the proposed models with Student-t (instead of Gaussian) errors. Another possible direction for future work would be to consider moving average stochastic volatility models, with the addition of jump components.

Another interesting issue with the related empirical applications we used is the evaluation of financial risk.[8] This risk can be quantified, for example, using Value-at-Risk or Expected Shortfall measures. A reliable calculation of financial risk measures (such as Value-at-Risk and Expected Shortfall) in the context of Bayesian SV models could, for example, be achieved by using particle filter methods; see for example the Adaptive Particle MCMC of Yang et al. (2017). In a future paper, it will be interesting, both computationally and empirically, to assess the risk obtained from the proposed models for the above, or similar, data sets.

Finally, an interesting direction is to incorporate random probability measures in stochastic volatility models. This will result in Bayesian semiparametric models, which are useful, as it is known that return data contain asymmetries or exhibit leptokurtic behaviour, both of which can not be captured adequately by parametric stochastic volatility specifications.

References

Abanto-Valle, C. A., Migon, H. S., Lachos, V. H. (2011). Stochastic volatility in mean models with scale mixtures of normal distributions and correlated errors: a Bayesian approach. *Journal of Statistical Planning and Inference* 141: 1875-1887.

Bekaert, G., Wu, G. (2000). Asymmetric volatility and risk in equity markets. *The Review of Financial Studies* 13: 1-42.

Black, F. (1976). Studies of stock market volatility changes. In: *Proceedings of the American Statistical Association, Business and Economic Studies Section*, pp. 177-181.

Campbell, J. Y., Hentschel, L. (1992). No news is good news: an asymmetric model of changing volatility in stock returns. *Journal of Financial Economics* 31: 281-318.

Chan, J. C. C. (2013). Moving average stochastic volatility models with application to inflation

[8]We thank a referee for pointing out this issue.

forecast. *Journal of Econometrics* 176: 162-172.

Chan, J. C. C. (2017). The stochastic volatility in mean model with time-varying parameters: an application to inflation modeling. *Journal of Business & Economic Statistics* 35: 17-28.

Chan, J. C. C., Eisenstat, E. (2015). Marginal likelihood estimation with the cross-entropy method. *Econometric Reviews* 34: 256-285.

Chan, J. C. C., Grant, A. L. (2016a). Modeling energy price dynamics: GARCH versus stochastic volatility. *Energy Economics* 54: 182-189.

Chan, J. C. C., Grant, A. L. (2016b). On the observed-data deviance information criterion for volatility modeling. *Journal of Financial Econometrics* 14: 667-689.

Chan, J. C. C., Jeliazkov, I. (2009). Efficient simulation and integrated likelihood estimation in state space models. *International Journal of Mathematical Modelling and Numerical Optimisation* 1: 101-120.

Chib, S. (2001). Markov chain Monte Carlo methods: computation and inference. In: *Handbook of Econometrics*, Volume 5. Elsevier, pp. 3569-3649.

French, K. R., Schert, W. G., Stambugh, R. F. (1987). Expected stock returns and volatility. *Journal of Financial Economics* 19: 3-29.

Jacquier, E., Rossi, P. E., Polson, N. (2004). Bayesian analysis of stochastic volatility models with fat-tails and correlated errors. *Journal of Econometrics* 122: 185-212.

Geweke, J. (1992). Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. In: *Bayesian Statistics 4*. Oxford: Clarendon Press, pp. 641-649.

Harvey, A. C., Shephard, N. (1996). Estimation of an asymmetric stochastic volatility model for asset returns. *Journal of Business & Economic Statistics* 14: 429-434.

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

Koopman, S. J., Hol Uspensky, E. (2002). The stochastic volatility in mean model: empirical evidence from international stock markets. *Journal of Applied Econometrics* 17: 667-689.

Leão, W. L., Abanto-Valle, C. A., Chen, M. H. (2017). Bayesian analysis of stochastic volatility-in-mean model with leverage and asymmetrically heavy-tailed error using generalized hyperbolic skew Student's t-distribution. *Statistics and its Interface* 10(4): 529-541.

Li, Y., Zeng, T., Yu, J. (2012). Robust deviance information criterion for latent variable models. *SMU Economics and Statistics Working Paper Series.*

Lopes, H. F., Polson, N. G. (2010). Bayesian inference for stochastic volatility modeling. In: *Rethinking Risk Measurement and Reporting: Uncertainty, Bayesian Analysis and Expert Judgement.* Risk Books, pp. 515-551.

Martin, V., Hurn, S., Harris, D. (2012). *Econometric Modelling with Time Series: Specification, Estimation and Testing. Themes in Modern Econometrics.* New York: Cambridge University Press.

McCausland, W. J., Miller, S., Pelletier, D. (2011). Simulation smoothing for state-space models: a computational efficiency analysis. *Computational Statistics and Data Analysis* 55: 199-212.

Nakajima, J., Omori, Y. (2012). Stochastic volatility model with leverage and asymmetrically heavy-tailed error using GH skew Student's t-distribution. *Computational Statistics & Data Analysis* 56: 3690-3704.

Omori, Y., Chib, S., Shephard, N., Nakajima, J. (2007). Stochastic volatility with leverage: fast and efficient likelihood inference. *Journal of Econometrics* 140: 425-449.

Omori, Y., Watanabe, T. (2008). Block sampler and posterior mode estimation for asymmetric stochastic volatility models. *Computational Statistics & Data Analysis* 52: 2892-2910.

Ritter, C., Tanner, M. A. (1992). Facilitating the Gibbs sampler: the Gibbs stopper and the Griddy-Gibbs sampler. *Journal of the American Statistical Association* 87: 861-868.

Schwert, G. W. (1989). Why does stock market volatility change over time?. *The Journal of Finance* 44: 1115-1153.

Shephard, N., Pitt, M. (1997). Likelihood analysis of non-Gaussian measurement time series. *Biometrika* 84: 653-667.

Spiegelhalter, D., Best, N., Carlin, B., Van Der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 64: 583-639.

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

Yang, X., Chatpatanasiri, R., Sattayatham, P. (2017). Value at risk estimation under stochastic volatility models using adaptive PMCMC methods. *Communications in Statistics-Simulation and Computation* 46(9): 7221-7237.

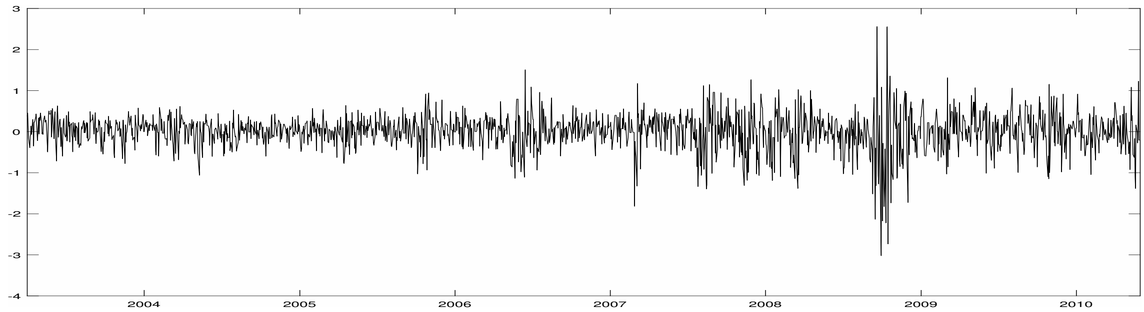

Figure 1: Daily returns on Equity Hedge from April 1, 2003 to May 31, 2010.

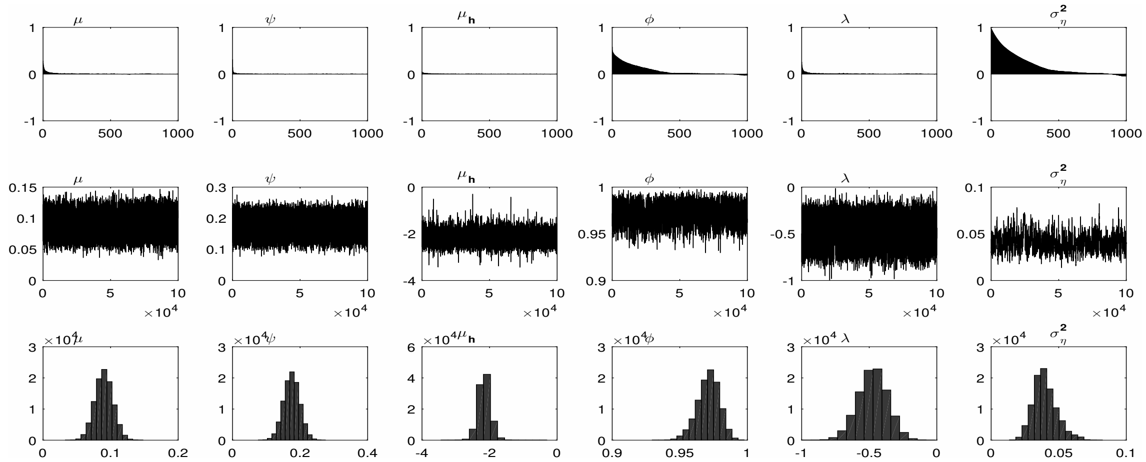

Figure 2: Empirical results (Equity Hedge). Posterior autocorrelations (top), posterior paths (middle) and posterior histograms (bottom) for the parameters of the MASVM model.

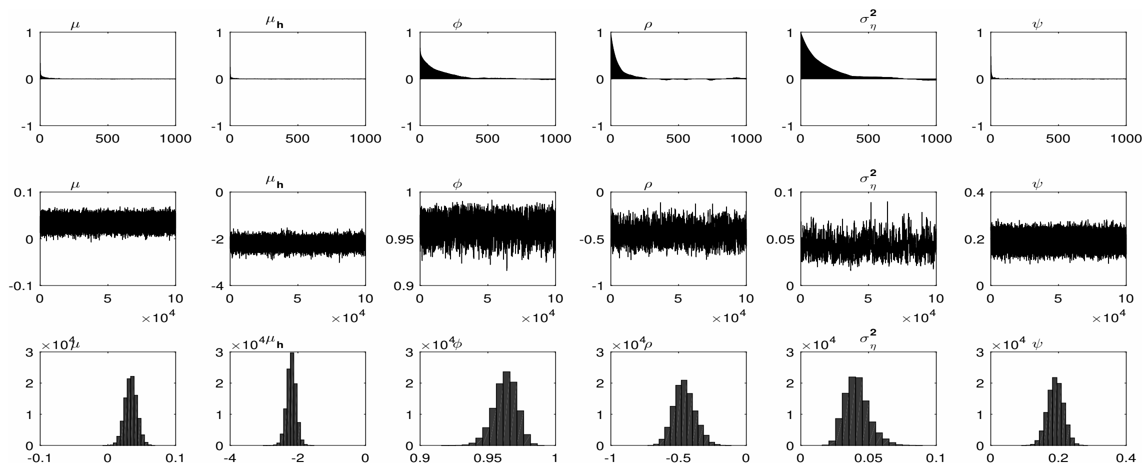

Figure 3: Empirical results (Equity Hedge). Posterior autocorrelations (top), posterior paths (middle) and posterior histograms (bottom) for the parameters of the MASVL model.

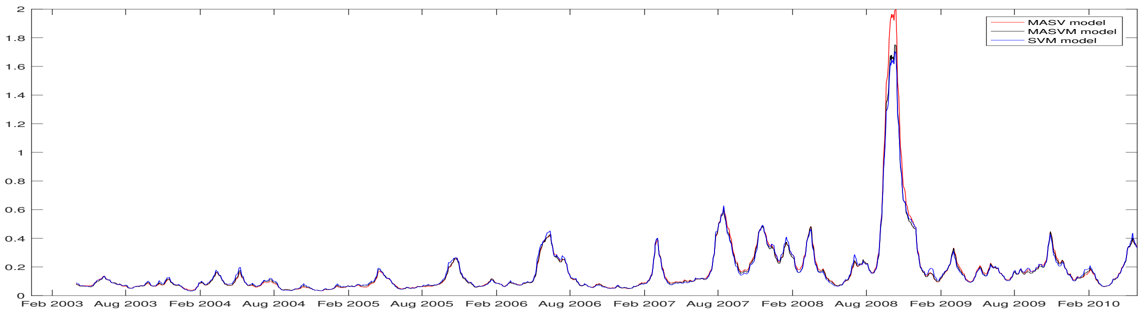

Figure 4: Empirical results (Equity Hedge). Evolution of estimated $\exp(h_t)$ for the MASV (red), MASVM (black) and SVM (blue) models.

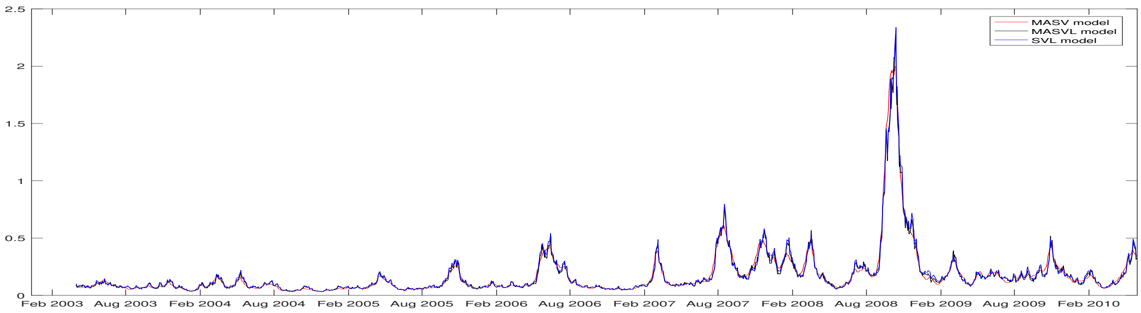

Figure 5: Empirical results (Equity Hedge). Evolution of estimated $\exp(h_t)$ for the MASV (red), MASVL (black) and SVL (blue) models.

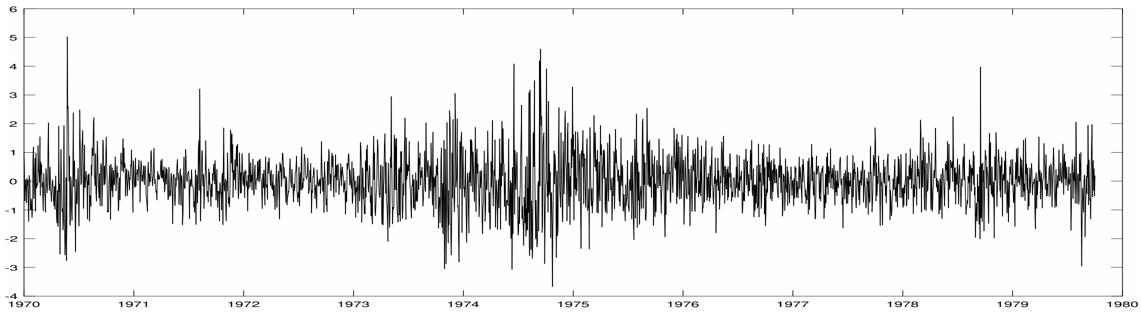

Figure 6: Daily returns on the S&P 500 index from January 2, 1970 to November 21, 1979 (in percentage points).

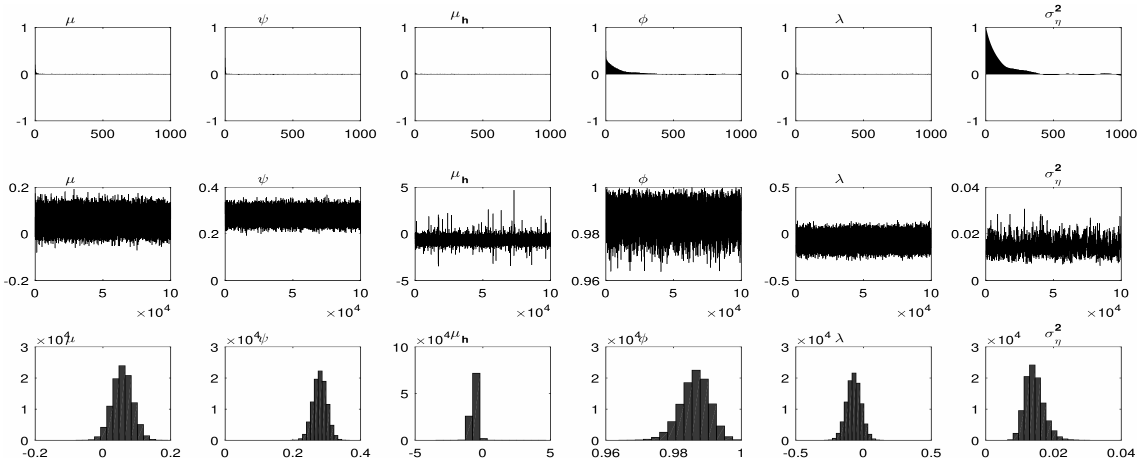

Figure 7: Empirical results (S&P 500 index). Posterior autocorrelations (top), posterior paths (middle) and posterior histograms (bottom) for the parameters of the MASVM model.

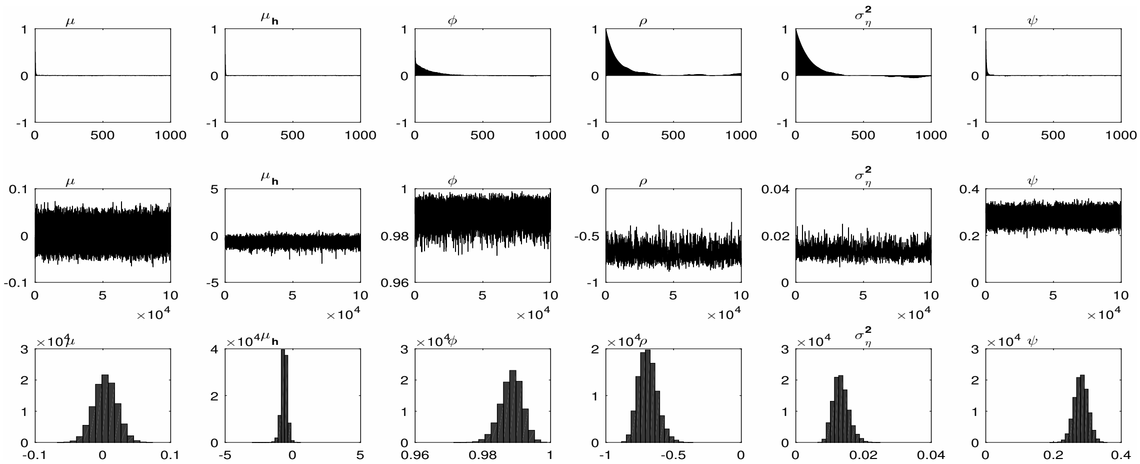

Figure 8: Empirical results (S&P 500 index). Posterior autocorrelations (top), posterior paths (middle) and posterior histograms (bottom) for the parameters of the MASVL model.

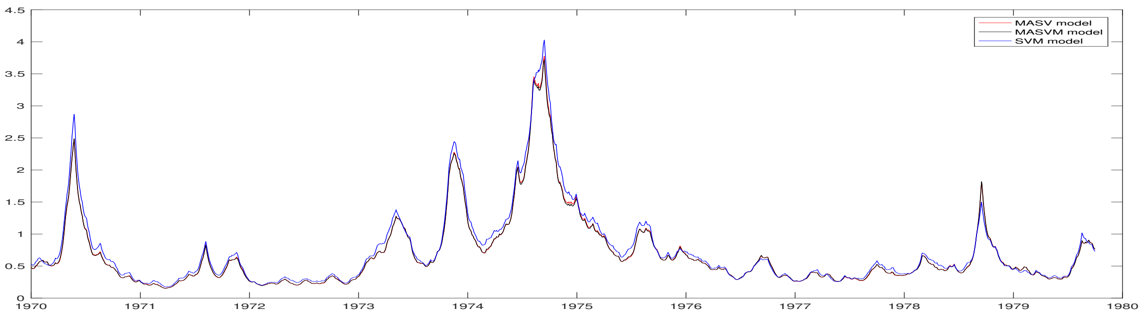

Figure 9: Empirical results (S&P 500 index). Evolution of estimated $\exp(h_t)$ for the MASV (red), MASVM (black) and SVM (blue) models.

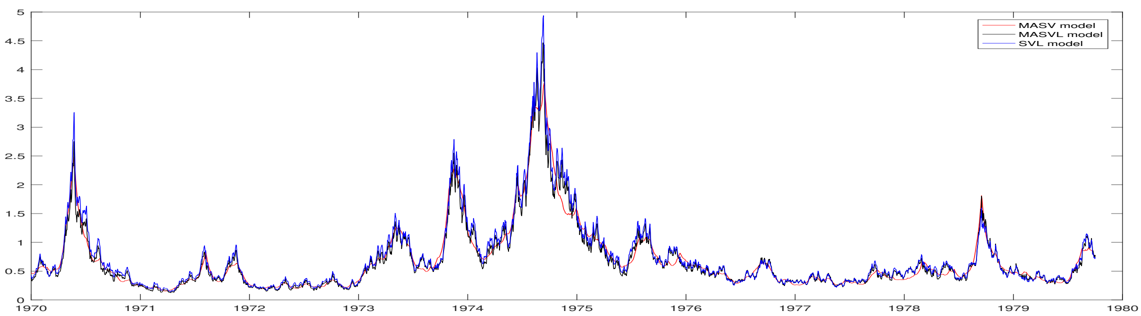

Figure 10: Empirical results (S&P 500 index). Evolution of estimated $\exp(h_t)$ for the MASV (red), MASVL (black) and SVL (blue) models.

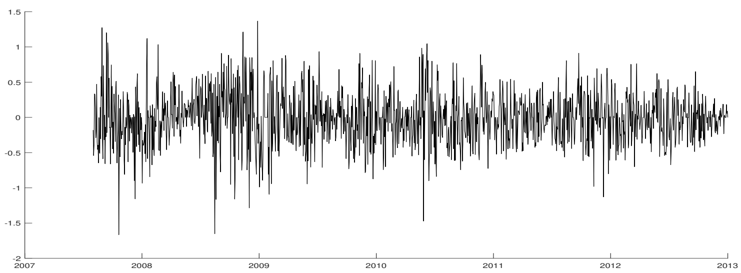

Figure 11: PHP/USD daily returns from July 2007 to December 2012.

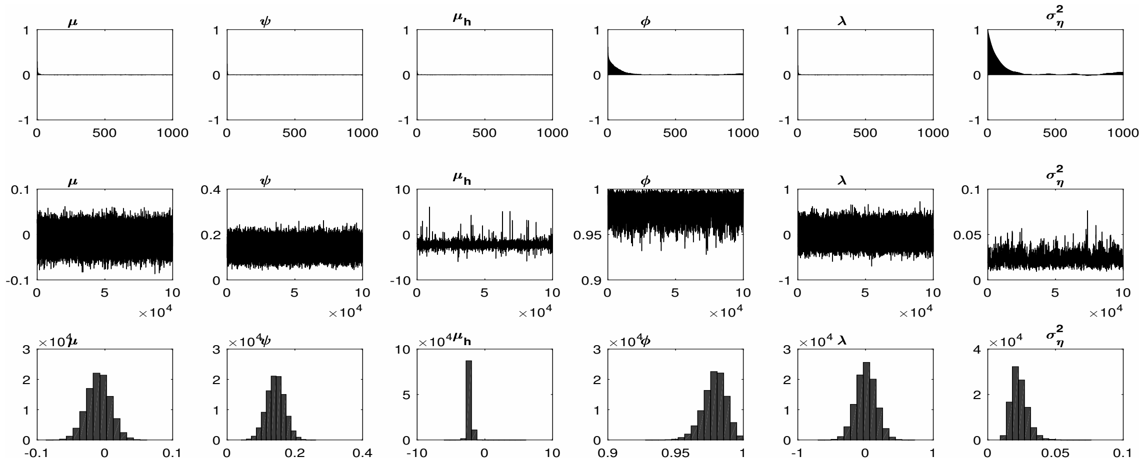

Figure 12: Empirical results (foreign exchange returns). Posterior autocorrelations (top), posterior paths (middle) and posterior histograms (bottom) for the parameters of the MASVM model.

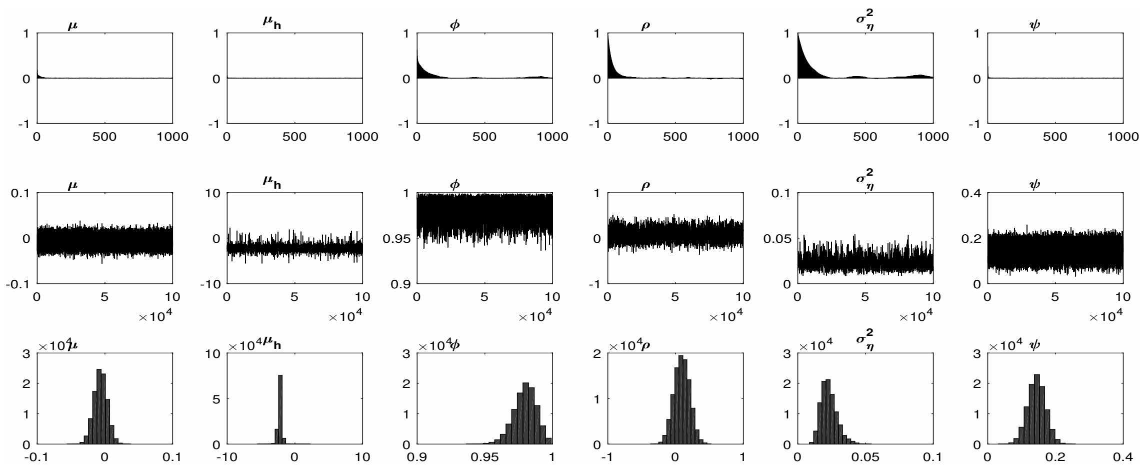

Figure 13: Empirical results (foreign exchange returns). Posterior autocorrelations (top), posterior paths (middle) and posterior histograms (bottom) for the parameters of the MASVL model.

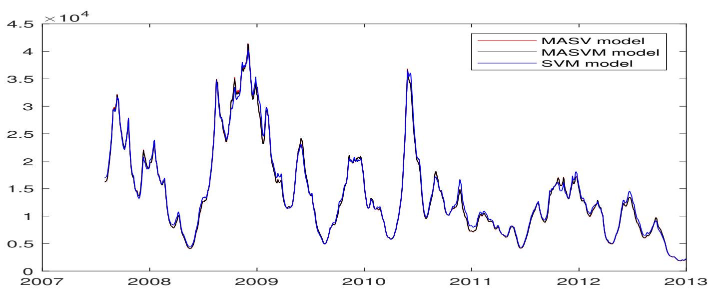

Figure 14: Empirical results (foreign exchange returns). Evolution of estimated $\exp(h_t)$ for the MASV (red), MASVM (black) and SVM (blue) models.

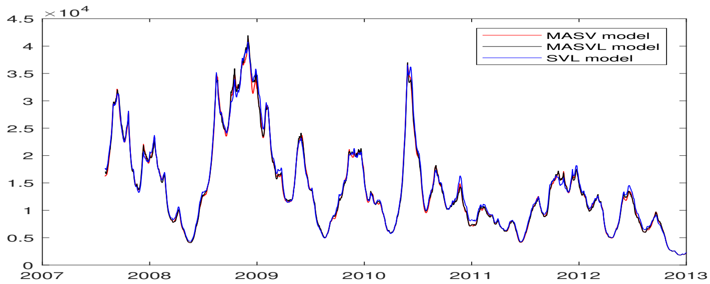

Figure 15: Empirical results (foreign exchange returns). Evolution of estimated $\exp(h_t)$ for the MASV (red), MASVL (black) and SVL (blue) models.

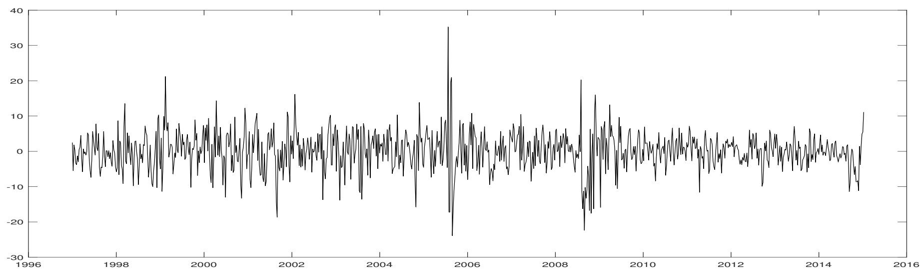

Figure 16: Weekly energy returns from January 3, 1997 to February 6, 2015.

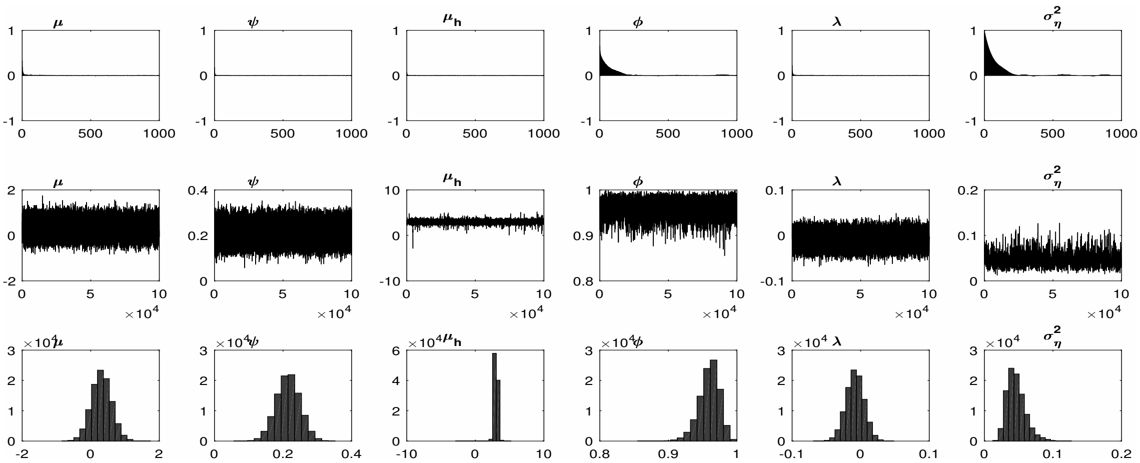

Figure 17: Empirical results (energy returns). Posterior autocorrelations (top), posterior paths (middle) and posterior histograms (bottom) for the parameters of the MASVM model.

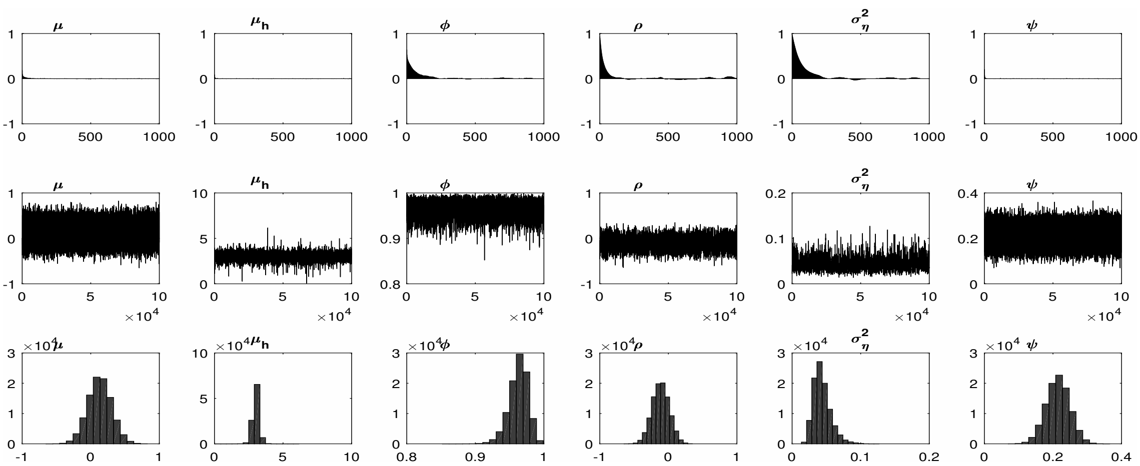

Figure 18: Empirical results (energy returns). Posterior autocorrelations (top), posterior paths (middle) and posterior histograms (bottom) for the parameters of the MASVL model.

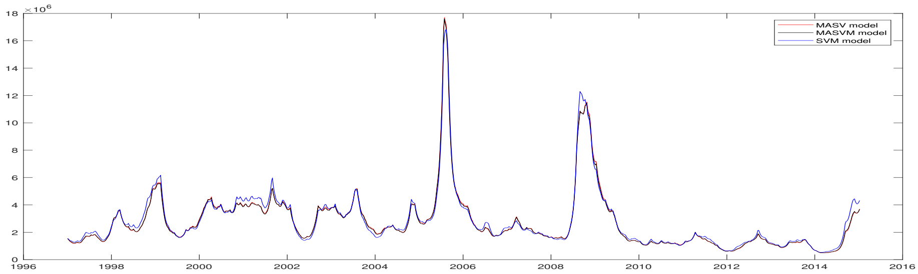

Figure 19: Empirical results (energy returns). Evolution of estimated $\exp(h_t)$ for the MASV (red), MASVM (black) and SVM (blue) models.

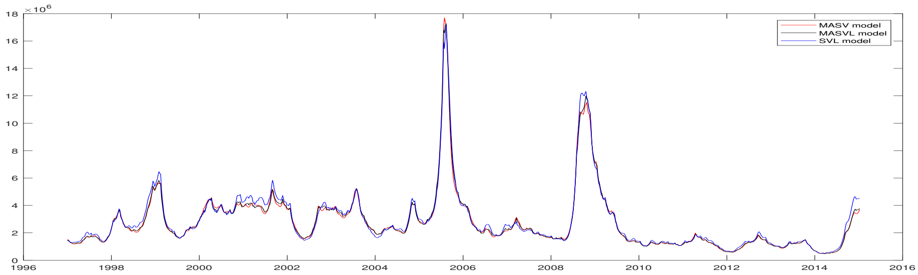

Figure 20: Empirical results (energy returns). Evolution of estimated $\exp(h_t)$ for the MASV (red), MASVL (black) and SVL (blue) models.

Table 1: Empirical results (Equity Hedge). Competing stochastic volatility models

Model	MASVM			MASVL			MASV			SVM			SVL			SV			SVML		
	Mean	IF	CD	Mean	IF	CD	Mean	IF	CD	Mean	IF	CD	Mean	IF	CD	Mean	IF	CD	Mean	IF	CD
μ	0.090*	7.326	0.834	0.033*	4.156	1.722	0.044*	1.626	2.106	0.095*	5.631	1.065	0.038*	3.548	1.502	0.045*	3.0185	3.534	0.080*	14.502	-0.988
	(0.013)			(0.008)			(0.008)			(0.012)			(0.007)			(0.007)			(0.012)		
λ	-0.474*	5.671	0.845							-0.494*	4.585	-0.729							-0.403*	12.031	1.496
	(0.109)									(0.098)									(0.097)		
ψ	0.175*	2.438	-0.060	0.191*	4.493	-1.421	0.185*	2.233	-2.599												
	(0.023)			(0.023)			(0.023)														
μ_h	-2.153*	1.193	-0.046	-2.194*	1.930	-1.268	-2.133*	1.099	2.351	-2.134*	1.089	3.771	-2.172*	1.820	-1.262	-2.099*	1.176	-2.354	-2.126*	1.368	-2.370
	(0.177)			(0.134)			(0.192)			(0.170)			(0.133)			(0.193)			(0.168)		
ϕ	0.971*	146.36	0.569	0.9628*	121.66	-0.637	0.974*	68.482	-0.534	0.969*	82.029	-0.312	0.9626*	135.58	-1.614	0.973*	121.35	-1.772	0.969*	110.96	-1.14
	(0.008)			(0.008)			(0.007)			(0.008)			(0.008)			(0.007)			(0.008)		
ρ				-0.456*	97.21	0.568							-0.399*	83.492	0.496				-0.306*	98.065	1.135
				(0.078)									(0.075)						(0.087)		
σ_η^2	0.039*	367.49	-0.173	0.041*	272.14	0.801	0.036*	168.58	0.406	0.042*	178.32	0.088	0.040*	260.15	1.392	0.037*	352.92	1.734	0.041*	271.13	1.155
	(0.008)			(0.008)			(0.007)			(0.009)			(0.009)			(0.008)			(0.009)		
Log ML	-753.4			-746.1			-756.8			-775.8			-774.5			-783.7			-771.8		
	(0.10)			(0.04)			(0.05)			(0.08)			(0.06)			(0.05)			(0.08)		
DIC_{obs}	1475.6			1468.2			1494.7			1525.6			1531.0			1553.2			1513.2		
	(0.65)			(0.42)			(0.43)			(0.44)			(0.63)			(0.24)			(0.54)		
$p_{D_{obs}}$	13.5			12.8			13.6			13.7			13.2			12.8			13.7		
	(0.55)			(0.42)			(0.26)			(0.43)			(0.63)			(0.23)			(0.54)		
LPS	-58.240			-54.696			-59.909			-59.181			-56.348			-60.346			-56.395		
RMSFE	0.4266			0.4255			0.4256			0.4279			0.4270			0.4287			0.4271		
KLD	0.0065			0.0032			0.0037			0.0074			0.0063			0.0092			0.0066		

*Significant based on the 95% highest posterior density interval. Standard deviation in parentheses (for the estimated parameters). For the Log ML estimates and observed-data DIC (DIC_{obs}) estimates we report their numerical standard errors in parentheses. We also report the estimated effective number of parameters $p_{D_{obs}}$ for each model that was computed, using the observed-data likelihood, along with their numerical standard errors in parentheses. LPS stands for Log Predictive Score. RMSFE stands for root mean squared forecast error. KLD stands for Kullback–Leibler divergence. IF stands for Inefficiency Factor and CD stands for Convergence Diagnostics.

Table 2: Empirical results (S&P500 index). Competing stochastic volatility models

Model	MASVM			MASVL			MASV			SVM			SVL			SV			SVML		
	Mean	IF	CD	Mean	IF	CD	Mean	IF	CD	Mean	IF	CD	Mean	IF	CD	Mean	IF	CD	Mean	IF	CD
μ	0.057*	1.974	-2.229	0.003	3.227	-0.082	0.025	1.207	-0.129	0.061*	2.329	-1.374	0.0130	2.435	0.549	0.025	1.302	0.4251	0.018	11.49	0.544
	(0.029)			(0.016)			(0.017)			(0.025)			(0.013)			(0.014)			(0.024)		
λ	-0.072	1.534	3.066							-0.072	1.785	0.871							-0.011	10.497	-1.138
	(0.052)									(0.041)									(0.041)		
ψ	0.2794*	2.198	-0.190	0.281*	8.330	-2.885	0.2799*	2.208	-0.128												
	(0.020)			(0.019)			(0.020)														
μ_h	-0.642*	1.063	-1.065	-0.6412*	2.85	0.326	-0.6411*	1.085	-0.597	-0.574*	1.051	-0.868	-0.602*	2.169	-0.306	-0.574*	1.071	0.952	-0.587*	1.990	1.382
	(0.205)			(0.211)			(0.207)			(0.208)			(0.197)			(0.211)			(0.223)		
ϕ	0.9862*	44.897	-0.541	0.9884*	50.087	-1.731	0.9863*	44.265	0.295	0.9867*	41.099	0.217	0.9880*	46.089	-1.468	0.986*	29.806	0.495	0.988*	42.569	0.894
	(0.004)			(0.003)			(0.004)			(0.004)			(0.003)			(0.004)			(0.003)		
ρ				-0.689*	162.72	1.766							-0.589*	108.7	2.546				-0.583*	123.53	0.914
				(0.068)									(0.069)						(0.074)		
σ^2_η	0.0143*	172.38	1.032	0.0133*	179.1	1.320	0.0144*	140.62	-0.466	0.0139*	175.03	-0.092	0.012*	166.81	0.978	0.013*	127.3	-0.319	0.012*	195.77	-0.605
	(0.002)			(0.002)			(0.002)			(0.002)			(0.002)			(0.002)			(0.002)		
Log ML	-2838.3			-2808.8			-2831.6			-2921.9			-2895.7			-2915.6			-2903.5		
	(0.05)			(0.04)			(0.08)			(0.07)			(0.05)			(0.03)			(0.08)		
DIC_{obs}	5634.4			5581.5			5636.3			5805.7			5760.5			5807.6			5761.4		
	(0.44)			(0.46)			(0.93)			(0.36)			(0.43)			(0.45)			(1.11)		
$p_{D_{obs}}$	8.8			7.0			9.7			7.8			7.4			7.5			7.4		
	(0.46)			(0.49)			(0.82)			(0.37)			(0.46)			(0.39)			(1.05)		
LPS	-124.6054			-112.6389			-121.5384			-123.1079			-124.7304			-126.5061			-124.5992		
RMSFE	0.7732			0.7730			0.7735			0.7731			0.7738			0.7741			0.7735		
KLD	0.0271			0.0244			0.0275			0.0269			0.0263			0.0283			0.0278		

*Significant based on the 95% highest posterior density interval. Standard deviation in parentheses (for the estimated parameters). For the Log ML estimates and observed-data DIC (DIC_{obs}) estimates we report their numerical standard errors in parentheses. We also report the estimated effective number of parameters $p_{D_{obs}}$ for each model that was computed, using the observed-data likelihood, along with their numerical standard errors in parentheses. LPS stands for Log Predictive Score. RMSFE stands for root mean squared forecast error. KLD stands for Kullback–Leibler divergence. IF stands for Inefficiency Factor and CD stands for Convergence Diagnostics.

Table 3: Empirical results (PHP/USD returns). Competing stochastic volatility models

Model	MASVM			MASVL			MASV			SVM			SVL			SV			SVML		
	Mean	IF	CD	Mean	IF	CD	Mean	IF	CD	Mean	IF	CD	Mean	IF	CD	Mean	IF	CD	Mean	IF	CD
μ	-0.009*	2.427	0.516	-0.006	2.999	0.347	-0.008	1.308	0.242	-0.011	2.870	0.831	-0.008	2.036	-0.286	-0.009	1.431	-1.548	-0.008	12.673	-0.061
	(0.017)			(0.009)			(0.009)			(0.016)			(0.008)			(0.008)			(0.017)		
λ	0.008	1.834	-0.665							0.023	2.196	-0.795							0.003	9.767	-0.236
	(0.147)									(0.132)									(0.137)		
ψ	0.143*	1.721	-0.063	0.144*	1.805	-0.313	0.142*	1.704	-1.411												
	(0.026)			(0.026)			(0.026)														
μ_h	-2.215*	1.283	0.586	-2.208*	1.169	1.057	-2.215*	1.223	0.434	-2.192*	1.188	2.081	-2.190*	1.099	1.30	-2.193*	1.258	-0.789	-2.191*	1.167	-1.260
	(0.264)			(0.264)			(0.276)			(0.260)			(0.24)			(0.262)			(0.253)		
ϕ	0.979*	47.656	1.26	0.979*	50.628	-0.167	0.978*	46.274	-1.376	0.978*	51.617	2.102	0.979*	44.752	-1.015	0.978*	52.485	-0.261	0.978*	44.764	0.278
	(0.008)			(0.008)			(0.008)			(0.008)			(0.008)			(0.008)			(0.008)		
ρ				0.092	61.546	1.449							0.054	40.541	0.711				0.053	43.495	-0.325
				(0.114)									(0.101)						(0.105)		
σ^2_η	0.023*	136.66	-1.173	0.023*	130.69	0.349	0.023*	134.6	1.580	0.022*	144.84	-1.857	0.023*	129.48	0.855	0.023*	150.79	-0.094	0.022*	128.41	-0.291
	(0.006)			(0.005)			(0.006)			(0.005)			(0.005)			(0.006)			(0.005)		
Log ML	-555.4			-545.9			-547.4			-563.7			-558.6			-556.9			-565.3		
	(0.06)			(0.03)			(0.03)			(0.05)			(0.05)			(0.03)			(0.08)		
DIC_{obs}	1075.6			1073.9			1074.5			1102.3			1101.3			1101.4			1102.8		
	(0.49)			(0.30)			(0.60)			(0.30)			(0.12)			(0.30)			(0.23)		
$p_{D_{obs}}$	9.4			8.6			9.3			8.4			7.9			8.6			8.4		
	(0.45)			(0.30)			(0.59)			(0.28)			(0.15)			(0.31)			(0.22)		
LPS	-53.2563			-51.7896			-52.6824			-52.1581			-51.8068			-53.7426			-52.2280		
RMSFE	0.2885			0.2898			0.2901			0.2923			0.2935			0.2936			0.2928		
KLD	0.0650			0.0490			0.0508			0.0747			0.0611			0.0784			0.0695		

*Significant based on the 95% highest posterior density interval. Standard deviation in parentheses (for the estimated parameters). For the Log ML estimates and observed-data DIC (DIC_{obs}) estimates we report their numerical standard errors in parentheses. We also report the estimated effective number of parameters $p_{D_{obs}}$ for each model that was computed, using the observed-data likelihood, along with their numerical standard errors in parentheses. LPS stands for Log Predictive Score. RMSFE stands for root mean squared forecast error. KLD stands for Kullback–Leibler divergence. IF stands for Inefficiency Factor and CD stands for Convergence Diagnostics.

Table 4: Empirical results (energy returns). Competing stochastic volatility models

Model	MASVM			MASVL			MASV			SVM			SVL			SV			SVML		
	Mean	IF	CD	Mean	IF	CD	Mean	IF	CD	Mean	IF	CD	Mean	IF	CD	Mean	IF	CD	Mean	IF	CD
μ	0.313	3.097	-0.125	0.130	2.325	0.790	0.164	1.469	1.188	0.442	4.019	0.940	0.168	2.318	-0.118	0.197	1.720	-0.005	0.373	12.838	2.266
	(0.290)			(0.171)			(0.168)			(0.264)			(0.146)			(0.145)			(0.277)		
λ	-0.008	2.070	-0.792							-0.012	2.904	-1.671							-0.010	10.174	-2.321
	(0.013)									(0.011)									(0.011)		
ψ	0.214*	1.610	-0.058	0.214*	1.653	-0.719	0.215*	1.572	0.137												
	(0.034)			(0.034)			(0.034)														
μ_h	3.011*	1.229	-0.026	3.017*	1.241	-1.744	3.014*	1.204	-0.941	3.051*	1.220	-2.113	3.055*	1.283	0.567	3.053*	1.216	0.100	3.054*	1.250	0.420
	(0.209)			(0.221)			(0.218)			(0.202)			(0.210)			(0.205)			(0.210)		
ϕ	0.959*	65.00	0.327	0.962*	53.928	0.131	0.960*	46.781	-0.691	0.957*	50.275	-0.883	0.960*	57.862	1.504	0.958*	55.817	-1.874	0.959*	57.754	-0.026
	(0.014)			(0.013)			(0.01)			(0.014)			(0.014)			(0.014)			(0.01)		
ρ				-0.111	52.276	0.324							-0.125	39.299	0.338				-0.094	42.683	2.133
				(0.124)									(0.109)						(0.114)		
σ^2_η	0.046*	130.2	-0.39	0.043*	124.21	0.156	0.046*	106.8	0.447	0.048*	121.29	0.627	0.045*	144.85	-1.235	0.048*	149.95	1.696	0.046*	129.29	0.062
	(0.013)			(0.012)			(0.012)			(0.013)			(0.013)			(0.014)			(0.014)		
Log ML	-2818.5			-2804.1			-2806.8			-2831.3			-2823.9			-2822.7			-2832.6		
	(0.05)			(0.04)			(0.03)			(0.08)			(0.04)			(0.04)			(0.06)		
DIC_{obs}	5607.5			5605.2			5605.5			5641.5			5641.6			5640.4			5643.4		
	(0.41)			(0.59)			(0.48)			(0.86)			(0.36)			(0.53)			(0.42)		
$p_{D_{obs}}$	11.6			10.3			10.3			10.0			9.9			8.5			11.5		
	(0.41)			(0.59)			(0.47)			(0.87)			(0.37)			(0.52)			(0.43)		
LPS	-836.9548			-833.5388			-834.4560			-837.1503			-833.9798			-834.7274			-836.5458		
RMSFE	3.6825			3.6796			3.6812			3.6983			3.6821			3.6838			3.6980		
KLD	0.0297			0.0291			0.0309			0.0322			0.0352			0.0317			0.0381		

*Significant based on the 95% highest posterior density interval. Standard deviation in parentheses (for the estimated parameters). For the Log ML estimates and observed-data DIC (DIC_{obs}) estimates we report their numerical standard errors in parentheses. We also report the estimated effective number of parameters $p_{D_{obs}}$ for each model that was computed, using the observed-data likelihood, along with their numerical standard errors in parentheses. LPS stands for Log Predictive Score. RMSFE stands for root mean squared forecast error. KLD stands for Kullback–Leibler divergence. IF stands for Inefficiency Factor and CD stands for Convergence Diagnostics.