Relative Forecasting Performance of Volatility Models: Monte Carlo Evidence

Thomas Lux* and Leonardo Morales-Arias†

December 2009

Abstract

A Monte Carlo (MC) experiment is conducted to study the forecasting performance of a variety of volatility models under alternative data generating processes (DGPs). The models included in the MC study are the (Fractionally Integrated) Generalized Autoregressive Conditional Heteroskedasticity models ((FI)GARCH), the Stochastic Volatility model (SV) and the Markov-switching Multifractal model (MSM). The MC study enables to compare the relative forecasting performance of models, which account for different characterizations of the latent volatility process: specifications which incorporate short/long memory, autoregressive components, stochastic shocks, Markov-switching and multifractality. Forecasts are evaluated by means of Mean Squared Errors (MSE), Mean Absolute Errors (MAE) and Value-at-Risk (VaR) diagnostics. Furthermore, complementarities between models are explored via forecast combinations. The results show that (i) the MSM model best forecasts volatility under any other alternative characterization of the latent volatility process and (ii) forecast combinations provide a systematic improvement upon forecasts of single models.

JEL Classification: C22, C53

Keywords: Monte Carlo simulations, volatility forecasting, long memory, multifractality, stochastic volatility, forecast combinations, Value-at-Risk.

*University of Kiel (Chair of Monetary Economics and International Finance) and Kiel Institute for the World Economy, Germany. Email: lux@economics.uni-kiel.de.

†Corresponding author. University of Kiel (Chair of Monetary Economics and International Finance) and Kiel Institute for the World Economy. Correspondence: Department of Economics, University of Kiel, Olshausenstrasse 40, 24118 Kiel, Germany. Email: moralesarias@economics.uni-kiel.de. Tel: +49(0)431-880-3171. Fax: +49(0)431-880-4383. Financial support from the Deutsche Forschungsgemeinschaft is gratefully acknowledged. We would also like to thank L. Calvet and A. Fisher and other participants of the 3rd International Conference on Computational and Financial Econometrics, Limassol, Cyprus, October 2009 for helpful comments and suggestions. Excellent comments and suggestions from H. Herwartz are particularly acknowledged. The usual disclaimer applies.

1 Introduction

Volatility forecasting has become one of the major cornerstones of the financial econometrics literature. A very large body of studies has investigated the performance of various volatility models for forecasting and their applications to e.g. risk management, portfolio selection, option pricing etc. (cf. Andersen et al. (2005a) for a recent review of volatility modeling and Poon and Granger (2003) for a review on volatility forecasting). In general, it has been difficult to outperform the standard Generalized Autoregressive Conditionally Heteroskedastic (GARCH) model at short horizons (Hansen and Lunde, 2005). However, models that account for important salient characteristics of asset returns such as long memory, regime-switching, stochastic shocks and multifractality should be able to generate better out-of-sample forecasts. For instance, the literature on volatility forecasting has shown that accounting for long memory in the GARCH framework by means of the (Fractionally Integrated) GARCH model or other long memory models may improve volatility forecasting at long horizons (Vilasuso, 2002; Zumbach, 2004; Lux and Kaizoji, 2007).

Moreover, while the literature on Stochastic Volatility (SV) modeling has concentrated mostly on searching for various efficient methods for the estimation of the parameters (cf. Andersen et al. (1999) for performance comparisons), studies that compare variants of SV models to (FI)GARCH models have also found promising evidence in favor of improved forecasting accuracy. For instance, SV variants such as ARFIMA models adapted to realized volatility have been found to outperform (FI)GARCH models in terms of forecasting volatility (Andersen et al., 2003, 2005b; Lux and Kaizoji, 2007). There is also evidence that SV models with regime-switching features may provide improved volatility forecasts (Lu and Perron, 2008).

Recently the Markov-Switching Multifractal (MSM) models have been proposed as another alternative mechanism to model and forecast volatility. The MSM models are adapted versions of the Multifractal Model of Asset Returns (MMAR) due originally to Calvet et al. (1997) and inspired by the work of Mandelbrot (1974, 1999). The notion of *multifractality* refers to the variations in the *scaling behavior* of various moments or to different degrees of long-term dependence of various moments. The MSM models account for multifractal volatility via their built-in hierarchical, multiplicative structure with heterogeneous components (Calvet and

Fisher, 2004). Pertinent empirical findings on multifractality have been reported in several studies by economists and physicists so that this feature now counts as a well established stylized fact of financial markets (Ding et al., 1993; Lux, 1996; Mills, 1997; Lobato and Savin, 1998; Schmitt et al., 1999; Vassilicos et al., 2004). There is also direct evidence in favor of the hierarchical structure of multifractal cascade models in phenomenological analyses of volatility dynamics at different levels of time aggregation (Muller, 1997). There is already a handful of studies that compare the forecasting performance of the MSM models against models from the GARCH and SV families in empirical applications (Calvet and Fisher, 2004; Lux and Kaizoji, 2007; Lux, 2008; Lux and Morales-Arias, 2010). Across a variety of financial markets, these studies show very promising performance of multifractal volatility models, often leading to forecast gains against their more time-honored counterparts (e.g. GARCH, FIGARCH, SV).

So far, comparative studies of (FI)GARCH, SV and MSM have been conducted almost exclusively via the assessment of their forecasting performance for various financial data. As it seems, however, there is in general a lack of systematic Monte Carlo evidence on the relative performance of volatility models for in-sample fitting and out-of-sample forecasting under various alternative data generating mechanisms. In contrast, a certain literature exists on the substitutability of time series models such as ARMA and ARFIMA. For instance, some studies show that suitably adapted ARMA models can yield a forecasting performance comparable to that of a 'true' underlying ARFIMA model (Basak et al., 2001; Man, 2003). Other contributions demonstrate that a 'true' ARFIMA model can be relatively well approximated by an ARMA model if the degree of long-term dependence (i.e. the value of the fractional differentiation parameter) is low (Crato and Ray, 1996; Brodsky and Hurvich, 1999). In any case, these results show that at least in certain cases, using a misspecified model might not be detrimental for forecasting purposes. In how far one could obtain reasonable volatility forecasts with the 'wrong' model is not known. While we might expect a certain replication of the ARFIMA versus ARMA findings for the FIGARCH versus GARCH case, the other processes considered (MSM and SV) are too different in their structural properties to come up with any educated guess on the relative in-sample and out-of-sample outcomes.

In this study we shed light on the behavior of misspecified volatility models by conducting a comprehensive computational analysis with synthetic data. Our experiment consists in simulating data from the (FI)GARCH, SV and MSM models and then forecasting volatility under the alternative DGPs. The MC study enables us to compare the relative forecasting performance of models, which account for quite different characterizations of the latent volatility process: specifications which incorporate short/long memory, autoregressive components, stochastic shocks, Markov-switching and multifractality. Since we have a large number of parameters to choose from in our computational experiment, we have restricted our attention to realistic settings. To this end, we took as our benchmark parameter sets, the mean group estimates from a large sample of stock indices at the country level (a total of 48 countries). Note that by calibrating models with mean estimates from international stock markets, we evaluate, as a by-product, the 'theoretical' capabilities of the models considered in forecasting international asset volatility.

In addition, it seems of interest in our set up to uncover the forecasting complementarities of volatility models that exploit different facets of the unobserved volatility process. Indeed, a 'hybrid' model that accounts for many of the important features of various models might be a promising avenue for improvement of volatility forecasting and thus of risk management strategies. Alternatively, a robust model such as the MSM which parsimoniously reproduces many of the stylized facts of asset returns (including some of those in the (FI)GARCH and SV models) might have enough flexibility 'by itself'. In addition, forecast combination strategies of the (many) existing models could be a simple and elegant way to achieve better forecasts from 'hybrid' specifications and to 'hedge' against forecasting with the 'wrong' model (Patton and Sheppard, 2009). In this paper we contribute to the latter issues by constructing forecast combinations from the various models considered to shed light on their complementarities.

To preview some of our main results, we find that: (i) the MSM model seems to forecast volatility better than any other model (save for the true DGP) under any other alternative characterization of the unobserved volatility process and (ii) forecast combinations constructed from forecasts generated by the various models considered provide a clear improvement upon forecasts of one single misspecified model.

The paper is organized as follows. The next section introduces the volatility models considered for the comparative forecasting analysis. Section 3 addresses the Monte Carlo design and the results of our study. The last section concludes with some final remarks.

2 The models

In this section we briefly discuss the volatility models of interest for our study. In general, models of volatility formalize the following specification of financial returns:

$$\Delta p_t = \upsilon_t + \sigma_t u_t, \tag{1}$$

where $\Delta p_t = \ln P_t - \ln P_{t-1}$, $\ln P_t$ is the log asset price, $\upsilon_t = E_{t-1}\Delta p_t$ is the conditional mean of the return series, σ_t is the volatility process and $u_t \sim N(0,1)$. The υ_t component of asset returns can be specified depending on the data generating process assumed for the asset pricing model. Defining $x_t = \Delta p_t - \upsilon_t$, the 'centered' returns are modelled as

$$x_t = \sigma_t u_t. \tag{2}$$

From the above general framework of volatility different parametric and non-parametric representations can be assumed for the volatility process σ_t, e.g. specifications that model different stylized facts of asset return data (i.e. autoregressive volatility, regime-switching, multifractality, long memory and stochastic shocks). In this study we consider four distinct models describing σ_t: the more time-honored GARCH, the FIGARCH model, the SV model and the MSM model. For simplicity, in this study we assume $\upsilon_t = 0$. In what follows we shortly discuss the volatility models considered for our analyses.

2.1 Generalized Autoregressive Conditional Heteroskedasticity models

In a seminal study of conditional heteroskedasticity in economic variables, Engle (1982) introduced the Autoregressive Conditional Heteroskedasticity (ARCH) model. The (Generalized) ARCH model was subsequently proposed by Bollerslev (1986) as a generalization of the ARCH model. The former model generalizes the latter by accounting for autoregressive features of volatility. The GARCH(1,1) model assumes that the volatility dynamics are governed by

$$\sigma_t^2 = \omega + \alpha x_{t-1}^2 + \beta \sigma_{t-1}^2, \tag{3}$$

where the restrictions on the parameters are $\omega > 0, \alpha, \beta \geq 0$ and $\alpha + \beta < 1$. The s-step ahead forecast representation of the GARCH(1,1) is given by

$$\hat{\sigma}_{t+s}^2 = \hat{\sigma}_g^2 + (\hat{\alpha} + \hat{\beta})^{s-1} \left[\hat{\sigma}_{t+1}^2 - \hat{\sigma}_g^2 \right], \tag{4}$$

where $\hat{\sigma}_g^2 = \hat{\omega}(1 - \hat{\alpha} - \hat{\beta})^{-1}$. Various extensions to the GARCH model have been proposed in the financial econometrics literature. One of the major additions to the GARCH-type models are the models that allow for long-memory in the specification of volatility dynamics. The FIGARCH model introduced by Baillie et al. (1996) accounts for 'genuine' long memory in the GARCH specification by means of fractional differences. As in the case of (3) we restrict our attention to one lag in both the autoregressive term and in the moving average term. The FIGARCH(1,d,1) is given by

$$\sigma_t^2 = \omega + \left[1 - \beta L - (1 - \delta L)(1 - L)^d \right] x_t^2 + \beta \sigma_{t-1}^2, \tag{5}$$

where L is the lag operator, d is the parameter of fractional differentiation and the restrictions on the parameters are $\beta - d \leq \delta \leq (2 - d)/3$ and $d(\delta - 2^{-1}(1 - d)) \leq \beta(d - \beta + \delta)$ (Baillie et al., 1996). The main difference of the FIGARCH model to the GARCH model is that the Binomial expansion of the fractional difference operator introduces an infinite number of lags with hyperbolically decaying coefficients for $0 < d < 1$. In the case of $d = 0$, the FIGARCH model reduces to the standard GARCH(1,1) model. Note that in practice, the infinite number of lags with hyperbolically decaying coefficients introduced by the Binomial expansion of the fractional difference operator must be truncated. We employ a lag truncation at 1000 steps as in Lux and Kaizoji (2007). The s-period ahead forecasts of the FIGARCH(1,d,1) model can be obtained most easily by recursive substitution, i.e.

$$\hat{\sigma}_{t+s}^2 = \hat{\omega}(1 - \hat{\beta})^{-1} + \eta(L) \hat{\sigma}_{t+s-1}^2, \tag{6}$$

where $\eta(L) = 1-(1-\hat{\beta}L)^{-1}(1-\hat{\delta}L)(1-L)^{\hat{d}}$ can be calculated from the recursions $\eta_1 = \hat{\delta}-\hat{\beta}+\hat{d}$, $\eta_j = \hat{\beta}\eta_{j-1} + [(j-1-\hat{d})j^{-1} - \hat{\delta}]\pi_{j-1}$, where $\pi_j \equiv \pi_{j-1}(j-1-\hat{d})j^{-1}$ are the coefficients in the MacLaurin series expansion of the fractional differencing operator $(1-L)^d$. Both GARCH and FIGARCH are *unifractal* models. While GARCH exhibits only short-term dependence (i.e. exponential decay of autocorrelations of moments) FIGARCH has a homogeneous hyperbolic decay of the autocorrelation of its moments characterized by the parameter of fractional differentiation d. GARCH and FIGARCH models are typically estimated via Quasi Maximum Likelihood (QML).

2.2 Stochastic Volatility model

In contrast to the deterministic volatility dynamics of the (FI)GARCH family, the SV model first introduced by Taylor (1999a,b), explicitly allows for an unobservable stochastic shock in the conditional variance. However, this feature also makes the SV model difficult to implement since its latent volatility process is stochastic and enters the model for volatility nonlinearly. The SV model can account for autoregressive volatility as in the (FI)GARCH models and unobservable stochastic shocks to volatility which may induce 'apparent' regime-switching (Lu and Perron, 2008).

In this study we consider the simple stationary SV model which is given by

$$\sigma_t^2 = \exp[h_t], \tag{7}$$

where

$$h_t = \kappa + \varphi h_{t-1} + \tau\varepsilon_t,\ \varepsilon_t \sim N(0,1). \tag{8}$$

Moreover, ε_t is assumed to be generated independently of u_t, $|\varphi| < 1$ and $\tau > 0$ is a scaling factor. Several estimation approaches have been proposed for the simple SV model (and its variants) including Generalized Method of Moments (GMM), Efficient Method of Moments (EMM), QML, or Markov Chain Monte Carlo (MCMC) methods (Melino and Turnbull, 1990; Harvey et al., 1994; Kim et al., 1997; Gallant et al., 1997; Liesenfeld and Richard, 2003).

We consider the relatively simple and robust QML approach proposed by Ruiz (1994) which outperforms the GMM approach and which saves on computational time in comparison to MCMC or other Bayesian methods. This requires transforming x_t in (2) by taking logarithms of the squares to obtain the linear model:

$$\ln x_t^2 = \mathrm{E}[\ln u_t^2] + h_t + \xi_t, \tag{9}$$

where $\xi_t = \ln u_t^2 - \mathrm{E}[\ln u_t^2]$ is a non-Gaussian, zero mean, white noise disturbance term and its statistical properties depend on the distribution of u_t^2. Model (9) coupled with (8) form a linear state space model. If u_t is Normally distributed with mean zero and unit variance then the mean and variance of $\ln u_t^2$ are $\psi(0.5) - \ln(0.5) \approx -1.27$ and $\pi^2/2$, respectively where $\psi(\bullet)$ is the Digamma function. Model (9) can be estimated by QML together with the Kalman filter by treating ξ_t as though it were $N(0, \pi^2/2)$. Estimates $\hat{h}_t$ can be obtained via Kalman smoothing. The s-step ahead forecast representation of the SV model is given by

$$\hat{h}_{t+s} = \hat{h} + \varphi^{s-1}(\hat{h}_{t+1} - \hat{h}), \tag{10}$$

where $\hat{h} = \hat{\kappa}(1 - \hat{\varphi})^{-1}$ and thus we may obtain $\hat{\sigma}_{t+s}^2 = \exp(\hat{h}_{t+s} + 0.5\hat{\sigma}_h^2)$ via (7).

2.3 Markov-switching Multifractal model

The MSM model is a causal analog of the earlier combinatorial Multifractal Model of Asset Returns (MMAR) due originally to Calvet et al. (1997). In contrast to (FI)GARCH and SV models, the MSM model can accommodate, by its very construction, the feature of *multifractality* via its hierarchical, multiplicative structure with heterogeneous components. In addition to *multifractality*, MSM models are able to reproduce characteristics of asset return data such as 'apparent' long memory. That is, depending on the number of volatility components, a pre-asymptotic hyperbolic decay of the autocorrelation in the MSM model might be so pronounced as to be practically indistinguishable from 'true' long memory (Liu et al., 2007).

Instantaneous volatility in the MSM model is determined by the product of k independent volatility components or multipliers $M_t^{(1)}, M_t^{(2)}, \ldots, M_t^{(k)}$ and a scale factor σ:

$$\sigma_t^2 = \sigma^2 \prod_{i=1}^{k} M_t^{(i)}. \tag{11}$$

Following the basic hierarchical principle of the multifractal approach, each volatility component will be renewed at time t with a probability γ_i depending on its rank within the hierarchy of multipliers and remains unchanged with probability $1 - \gamma_i$. Calvet and Fisher (2001) demonstrate convergence of the discrete MSM model to a continuous-time Poissonian limit under the following specification of transition probabilities:

$$\gamma_i = 1 - (1 - \gamma_k)^{(b^{i-k})}, \tag{12}$$

with γ_k and b parameters to be estimated. However, previous applications have often used pre-specified parameters $\gamma_k = 0.5$ and $b = 2$ in equation (12) which reduces the number of estimated parameters to two (Lux, 2008). The MSM model is fully specified once we have determined the number k of volatility components and their distribution. In the small body of available literature, the multipliers $M_t^{(i)}$ have been assumed to follow either a Binomial or a Lognormal distribution. Since one would normalize the distribution so that $E[M_t^{(i)}] = 1$, only one parameter has to be estimated for the distribution of volatility components. Taking into account the scale parameter, σ, we end up with a very parsimonious family of stochastic processes that is parameterized by only two parameters although the number of states k could be arbitrary large . An in-depth analysis of the MSM model can be found in Calvet and Fisher (2004).

We focus on the Binomial MSM which can be estimated via ML or GMM (Calvet and Fisher, 2004; Lux, 2008). In the ML procedure we can obtain optimal forecasts via Bayesian updating of the conditional probabilities $\Omega_t = \mathcal{P}(M_t = m^i | x_1, ..., x_t)$ for the unobserved volatility states $m^i, i = 1, ..., 2^k$. However, because of the computational demands of repeated updating of state probabilities via a $2^k \times 2^k$ matrix, ML estimation for MSM models is very time consuming for values of k beyond 8. As a fast and reasonably accurate alternative, a GMM approach has been proposed by Lux (2008). This alternative estimation procedure uses various powers of log differences of the original data in order to avoid biases due to the close-to-long-memory nature of the process. Out-of-sample forecasting of the MSM model estimated via GMM is performed via best linear forecasts (Brockwell and Davis, 1991, c.5) together with the generalized Levinson-Durbin algorithm developed by Brockwell and Dahlhaus (2004). We first have to consider the zero-mean time series:

$$X_t = x_t^2 - E[x_t^2] = x_t^2 - \widehat{\sigma}^2, \tag{13}$$

where $\widehat{\sigma}$ is the estimate of the scale factor σ. Assuming that the data of interest follow a stationary process $\{X_t\}$ with mean zero, the best linear s-step forecasts are obtained as

$$\widehat{X}_{n+s} = \sum_{i=1}^{n} \phi_{ni}^{(s)} X_{n+1-i} = \boldsymbol{\phi}_n^{(s)} \mathbf{X}_n, \tag{14}$$

where the vectors of weights $\boldsymbol{\phi}_n^{(s)} = (\phi_{n1}^{(s)}, \phi_{n2}^{(s)}, ..., \phi_{nn}^{(s)})'$ can be obtained from the analytical autocovariances of X_t at lags s and beyond. More precisely, $\boldsymbol{\phi}_n^{(s)}$ are any solution of $\boldsymbol{\Psi}_n \boldsymbol{\phi}_n^{(s)} = \boldsymbol{\kappa}_n^{(s)}$ where $\boldsymbol{\kappa}_n^{(s)} = (\kappa_{n1}^{(s)}, \kappa_{n2}^{(s)}, ..., \kappa_{nn}^{(s)})'$ denote the autocovariance of X_t and $\boldsymbol{\Psi}_n = [\kappa(i-j)]_{i,j=1,...,n}$ is the variance-covariance matrix. In what follows we describe in detail the Monte Carlo set up for the analysis of the relative forecasting performance of the GARCH, FIGARCH, SV and MSM volatility models.

insert Table 1 around here

3 Monte Carlo analysis

In the next subsection we describe the MC set up designed for our analyses. In the subsequent two subsections, the in-sample and out-of-sample results of the MC experiments are discussed, respectively.

3.1 The Monte Carlo design

The Monte Carlo experiment performed in this paper basically poses the questions 'what would be the in-sample estimates obtained for (say) GARCH given that the true model describing the data generating process is MSM?' and analogously, 'how much better or worse would one forecast volatility if one forecasts with (say) GARCH given that the true data generating process is MSM?'. More precisely, the exercise consists in simulating synthetic data $\tilde{x}_t$ with the volatility specifications $\sigma_t = \{\text{GARCH}, \text{FIGARCH}, \text{SV}, \text{MSM}\}$ in (3), (5), (7) and (11), respectively, and performing in-sample and out-of-sample analyses by using the estimation and forecasting algorithm of a particular model for σ_t with the alternative DGPs. Thus, the exercise allows us to contrast the relative in-sample and out-of-sample performance of volatility models vis-à-vis each other.

In order to have a realistic calibration, the set of parameters corresponding to the (FI)GARCH, SV and MSM models introduced in the previous section are taken from the Mean Group (MG) values for $N = 48$ all-share stock market indices at the country level (Table 1). MG estimates are simply obtained by averaging single market estimates. The data has been standardized prior to estimation and the MG estimates have been rounded off for simplicity. The sample runs from 06/01/1998 to 12/31/2007 at the daily frequency which leads to 2,500 (return) observations. The data is obtained from Datastream and the countries were chosen upon data availability for the sample period covered.

When calibrating the models for the subsequent MC simulations, we also make sure to set the scaling parameters such that the unconditional variance of the simulated model is equal to one. More precisely, the parameter space for the GARCH model is set to $\theta = \left\{\omega, \hat{\beta}, \hat{\alpha}\right\} = \{1 - 0.86 - 0.11, 0.86, 0.11\}$. In the case of the FIGARCH model the unconditional variance is not defined. However, in practice we have to approximate the fractional difference operator by a truncation lag which we have set to 1000. This, in fact, guarantees existence of a limit conditional variance (see Chang (2002) for this feature of the FIGARCH model). Thus, the parameter space of the FIGARCH model is set to $\theta = \left\{\omega, \hat{\beta}, \hat{\delta}, \hat{d}\right\} = \left\{1 - \lambda_{\hat{d}}(1), 0.45, 0.13, 0.42\right\}$ where $\lambda_d(L) = (1 - L)^d \approx \sum_{p=0}^{P} \lambda_{d,p} \cdot L^p$ so that the summation to the cut-off $P = 1000$ instead of the infinite theoretical sum leads to a non-vanishing $\lambda_d(1) \neq 0$.

In the case of the SV model the unconditional variance is given by $\exp(h + 0.5\sigma_h^2)$ where $h = \kappa(1 - \varphi)^{-1}$ and $\sigma_h^2 = \tau^2(1 - \varphi^2)^{-1}$ so we set $\theta = \{\kappa, \hat{\varphi}, \hat{\tau}\} = \left\{-0.5 \times 0.46^2 \times (1 - 0.89^2)^{-1} \times (1 - 0.89), 0.89, 0.46\right\}$. For the MSM model we set $\theta =$

$\{\sigma, \hat{m}_0, k\} = \{1, 1.46, 8\}$. Note that for the MSM model we account for only a very moderate number of multipliers $k = 8$. We do so because ML estimation for MSM models becomes cumbersome for $k > 8$. Note, however, that forecasting performance might nevertheless still improve for $k > 8$ and proximity to temporal scaling of empirical data might be closer (Liu et al., 2007; Lux, 2008). Our choice of the specification $k = 8$ is, therefore, a relatively conservative one.

We use time lengths of $T = 5,000$ and $T = 10,000$ to simulate the series from which half is employed for in-sample estimation and the other half for out-of-sample evaluation. We consider forecast horizons $s = 1$, 20, 50 and 100 periods ahead and a total of 400 MC repetitions are performed for in-sample estimation and forecasting. To evaluate forecasts we employ mean squared forecast errors (MSE) and mean absolute forecast errors (MAE). MSE and MAE measures for each of the models are given in percentage of the MSE and MAE measures of a naive forecast from the in-sample variance of the simulated series which has been set to one by construction. Thus, MSE and MAE values below one indicate a superior performance of a particular model against 'historical' volatility.

To define absolute and relative measures formally, let '0' and '•' indicate a benchmark (historical volatility) and a particular competing volatility model ((FI)GARCH, SV, MSM), respectively. Forecast errors for $\tau = T/2 + 1, ..., T - s$ are given by:

$$\hat{e}_\tau(0) = \sigma^2_\tau - \hat{\sigma}^2, \;\; \hat{e}_\tau(\bullet) = \sigma^2_\tau - \hat{\sigma}^2_\tau, \tag{15}$$

where σ^2_τ are the (simulated) squared returns ($\tilde{x}^2_t$), $\hat{\sigma}^2$ is the historical volatility estimate (in-sample variance of squared returns which is set to one) and $\hat{\sigma}^2_\tau$ is the volatility forecast of the competing model ((FI)GARCH, SV, MSM). The MSE and MAE of the benchmark (historical volatility) specification are:

$$\bar{d}(0) = \mathcal{T}^{-1} \sum_\tau d_\tau(0), \;\; d_\tau(0) = \hat{e}_\tau(0)^2 \text{ or } d_\tau(0) = |\hat{e}_\tau(0)|, \tag{16}$$

with $\mathcal{T}$ the number of out-of-sample observations. The average performance of a competing model specification is given in relation to $\bar{d}(0)$, obtaining relative MSEs or MAEs as:

$$\overline{dr}(\bullet) = \frac{\bar{d}(\bullet)}{\bar{d}(0)},\ \bar{d}(\bullet) = \mathcal{T}^{-1}\sum_{\tau} d_\tau(\bullet). \tag{17}$$

In addition to MSE and MAE measures, we also report quantile Value-at-Risk (VaR) hits as a non-parametric and 'scale free' measure of the reliability of extreme forecast realizations. From an economic perspective this allows us to assess the appropriateness of the various models at hand for risk management purposes. In what follows, the value-at-risk of $\{\tilde{x}_\tau\}_{\tau=1}^{\mathcal{T}}$ conditional on information set $\Omega_{\tau-1}$ with coverage α is denoted $\text{VaR}_\tau(\alpha)$. Formally $\text{VaR}_\tau(\alpha)$ is the quantile such that,

$$\text{Prob}[\tilde{x}_\tau < -\text{VaR}_\tau(\alpha)|\Omega_{\tau-1}] = \alpha. \tag{18}$$

Starting from the above definition for $\text{VaR}_\tau(\alpha)$, we may define the so-called *hit* process,

$$\text{Hit}_\tau(\alpha) \equiv \mathcal{I}(\tilde{x}_\tau < -\text{VaR}_\tau(\alpha)|\Omega_{\tau-1}), \tag{19}$$

where $\mathcal{I}(.)$ is an indicator function. Diagnosing $\text{VaR}_\tau(\alpha)$ estimates consists in testing the unconditional coverage hypothesis. Along the latter lines, an appropriate model for $\text{VaR}_\tau(\alpha)$ evaluation should have an unconditional coverage of α. In what follows, we discuss the results of the Monte Carlo experiments.

insert Table 2 around here

3.2 In-sample results

In-sample results of the MC experiment are displayed in Table 2. For the MSM models we present ML (MSM1) and GMM (MSM2) estimates which allows us to compare the relative performance of both estimators when the true DGP is not MSM. It should also hint at whether the moment conditions based on autocorrelations of absolute log returns used in the MSM models estimated via GMM would be useful to estimate other volatility models that share similar characteristics (e.g. FIGARCH and SV). A more detailed comparison between ML and

GMM estimation for the MSM models can be found in Lux (2008). As it turns out, in-sample results for $T = 2,500$ and $T = 5,000$ do not differ substantially. Thus, to save on space, the following discussion is confined to the case $T = 2,500$ in order to be consistent with the sample size available for the empirical estimation.

For the Binomial MSM model the value of the crucial 'fractality' parameter m_0 is 1.46, which indicates a relatively high degree of heterogeneity of volatility components (Lux, 2008). As expected, the MSM fractality parameter m_0 is highest under the MSM DGP but lower under the alternative DGPs of (FI)GARCH and SV. In fact, no other volatility specification can mimic the degree of heterogeneity of the MSM. Interestingly, GMM estimation in the MSM framework results in estimates of the multifractal parameter m_0 which are noticeably different from ML estimation when the true DGP is not MSM. Overall, the model which yields the highest ML/GMM estimates of m_0 is SV, suggesting that the SV DGP resembles regime-switching processes more than the (FI)GARCH DGP (Lu and Perron, 2008).

GARCH parameter estimates are practically the same under the GARCH and FIGARCH DGPs. The MSM DGP yields a relatively high persistence of past volatility according to the GARCH estimate $\bar{\beta} = 0.76$ suggesting that the MSM model generates a high degree of volatility clustering in asset return volatility. Interestingly, the GARCH model obtains a lower autoregressive parameter under the SV model ($\bar{\beta} = 0.69$) than under the MSM model ($\bar{\beta} = 0.76$).

Considering the GARCH model as the true DPG, the FIGARCH model yields a degree of fractional integration $\bar{d} = 0.48$ which is not too far apart from the 'true' FIGARCH value $\hat{d} = 0.42$. The latter outcome is similar to previous findings that the GARCH DGP can induce some degree of 'apparent' long memory (Baillie et al., 1996). As expected, the MSM mimics a high degree of 'true' long memory as given by the estimated parameter of fractional differentiation of the FIGARCH model, $\bar{d} = 0.69$. Indeed, the latter estimate is high in comparison to previous empirical applications where the latter parameter has usually been less than 0.5 (Baillie et al., 1996; Lux and Kaizoji, 2007). This outcome confirms previous findings that 'apparent' long memory processes induced from multifractal and/or regime-switching DGPs could be easily confused with 'genuine' ones (Granger and Terasvirta, 1999; Liu et al., 2007). As it seems, the SV DGP generates the lowest degree of 'apparent' long memory as the FIGARCH model yields the lowest value for the parameter of differentiation under the former model ($\bar{d} = 0.35$).

Lastly, we find that the SV model estimates a high level of autoregressive volatility (about $\bar{\varphi} = 0.96$) when the DGP is GARCH and this estimate is almost identical when the DPG is FIGARCH ($\bar{\varphi} = 0.97$) and not too far apart from the one obtained under the MSM DGP ($\bar{\varphi} = 0.93$). Interestingly, the estimated standard deviation of the volatility shock estimated by the SV model under the MSM DGP ($\bar{\tau} = 0.46$) is identical to the 'true' value ($\hat{\tau} = 0.46$) and to the estimated value when SV is the true DGP ($\hat{\tau} = 0.47$). The latter result hints at a certain proximity of stochastic renewals of volatility and regime-switching which is explicitly modelled in the MSM framework.

Overall we find that the MSM model can mimic many of the features of other models such as autoregressive volatility, genuine long memory and stochastic shocks while other models cannot mimic the degree of heterogeneity of the MSM. In the following sections we analyze what the latter outcome implies for out-of-sample forecasting.

insert Tables 3, 4 and 5 around here

3.3 Out-of-sample results

3.3.1 Single models

We now turn to the discussion of our out-of-sample analyses which are presented in Tables 3, 4 and 5. Table 3 displays forecasting results with the true parameter values (see Table 1) used for calibration while Tables 4 and 5 display results with estimated parameters (see Table 2). As expected, using estimated parameters as opposed to true ones usually deteriorates forecasts in terms of MSE, MAE or VaR quantile hits. However, most models produce qualitatively similar forecasts with true parameters or estimated ones. Nevertheless, Table 3 provides a good benchmark for the 'theoretical' forecasting capabilities of the volatility models considered.

Considering the MSM model as true DGP, the GARCH model produces the least accurate forecasts in terms of MSEs and MAEs when compared against the FIGARCH and SV models. The latter result is intuitive as the MSM model accounts for characteristics such as (apparent) long memory and stochastic volatility (via the stochastic renewal of multipliers $M_t^{(i)}$) which are also accounted for by the FIGARCH (i.e. long memory) and SV (i.e. stochastic shocks) models but ignored by the GARCH model. Comparing forecasts of FIGARCH and SV models, the SV model can forecast multifractal volatility better.

Taking the GARCH as the true DGP, the MSM1 model seems to offer better forecast accuracy in terms of MSE and MAE when compared against the FIGARCH and SV models. When the true DGP is FIGARCH and thus 'genuine' long memory is taken into consideration, the SV (MSM1) model produces the least (most) accurate forecasts in terms of MSEs and MAEs out of the other three non-FIGARCH models. Interestingly, the MSM model which accounts for 'apparent' long memory works relatively well when the true DGP entails genuine long memory via the FIGARCH specification, while the FIGARCH forecasting performance is dismal when the true DGP is MSM.

Lastly, out of the three non-SV models considered, we obtain that the FIGARCH (MSM1) performs worse (best) in terms of MSE and MAE when the true DGP is SV. Thus, it seems that the MSM model can best forecast stochastic volatility when compared against the other non-SV models considered. The latter result is intuitive as the MSM accounts for stochastic shocks to the latent volatility process whereas the (FI)GARCH models do not. Note, however, that forecastability of SV realisations is very limited beyond short horizons (the same applies to GARCH time series).

insert Tables 6 and 7 around here

In general, when comparing the models' forecasts for our synthetic time series, the results indicate that the MSM model seems to best forecast future volatility independent of the DGP of the model in terms of MSEs and MAEs. Our results, are also robust with respect to the unconditional VaR coverage of the forecasts. Tables 6 and 7 show the results on 99% and 95% quantile VaR hits with estimated parameters. When evaluating quantile VaR hits, we find qualitatively similar results as with MSEs and MAEs. That is, independent of the true DGP, the MSM model seems to forecasts VaR quantiles relatively well, while the (FI)GARCH models do less so when the true DGP is MSM or SV. The GARCH model yields good VaR forecasts at lower horizons independent of the true DGP but its forecasting capabilities deteriorate substantially at longer horizons. The latter finding is of high importance for practitioners as it indicates that the MSM model can produce accurate forecasts even when the analyst does not have a clear idea of which model to employ when adopting (say) Value-at-Risk measures for risk management strategies.

insert Figure 1 and 2 around here

The results discussed so far can be easily appreciated in the boxplots displayed in Figures 1 and 2 corresponding to the MSEs and 95% VaR hits over the 400 simulations. Results are qualitatively the same for MAEs and 99% VaR hits and can be provided upon request. The boxplots basically show that, in terms of MSEs, the MSM can perform significantly better than other models when the true DGP is MSM and not significantly worse than other models when the true DGP is not MSM. The same results also apply for the 95% VaR hits at higher horizons.

Our results point out that the GARCH model falls behind all other models under the other alternative volatility specifications, particularly at longer horizons. The same applies to the FIGARCH specification when the true DGP is other than FIGARCH. However, non-GARCH models such as MSM and SV perform reasonably well when the GARCH or FIGARCH models are the true DGP. Overall, we find that the models considered exhibit different degrees of flexibility under alternative DGPs that exploit salient features of financial data such as long-term dependence, stochastic shocks and regime-switching. Thus, it seems of interest to analyze whether complementarities exist between the various models at hand and whether forecasting performance can be improved via combinations of single models, an issue explored below.

3.3.2 Combined forecasts

A particular insight from the methodological literature on forecasting is that it is often preferable to combine alternative forecasts in a linear fashion and thereby obtain a new predictor (Granger, 1989; Aiolfi and Timmermann, 2006; Patton and Sheppard, 2009; Costantini and Pappalardo, 2009). We analyze forecast complementarities of (FI)GARCH, MSM and SV models by addressing the performance of combined forecasts. The forecast combinations are computed by assigning each single forecast a weight equal to a model's empirical frequency of minimizing the absolute or squared forecast error over *realized past* forecasts. We update the weighting scheme over the 20 most recent forecast errors so that despite linear combinations of forecasts, the influence of various components is allowed to change over time via flexible weights.

We consider two sets of forecast combinations CO1 and CO2. The former forecast combination contains all four models (GARCH, FIGARCH, SV, MSM) while the latter excludes the true model. Thus, CO1 allows us to uncover whether we can improve upon forecasts of single models by combining forecasts of all models. Similarly, CO2 allows us to analyze how well we would be able to improve upon forecasts from single models when we combine forecasts of models that do not include the true DGP.

insert Figures 3 and 4 around here

In terms of MSE and MAEs we find that CO1 improves forecasts of single models that are not the true DGP (Tables 4 and 5). Forecasts of the true DGP are even improved at longer horizons for the FIGARCH specification. Combined forecasts CO2 show, as expected, a deterioration in comparison to CO1. However, the results remain qualitatively similar to those of CO1 as we find that forecasts of single models are generally improved. Considering VaR diagnostics by means of quantile hits, we find again qualitatively similar results as for MSE and MAE evaluations. That is, combining forecasts seems to provide an improvement in forecasting VaR in relation to several of the single models considered (Tables 6 and 7). Interestingly, the results on forecast combinations also seem to suggest different behavior for MSM and other models: For MSM, combined forecasts are somewhat worse than those based on the true DGP, whereas for other models, combined forecasts are qualitatively similar to those of the true DGP.

Overall, forecast combinations significantly improve upon single models' forecasts when forecasting with the 'wrong' model (see Figures 3 and 4). Thus the MC results on the forecast combinations confirm the findings of the previous section and of recent empirical studies (Patton and Sheppard (2009); Lux and Morales-Arias (2010)). That is, 'hybrid' specifications between various volatility models could potentially improve out-of-sample forecasting. However, the MSM model seems to be quite robust 'by itself' independent of the true DGP or forecast combination strategies.

4 Conclusion

This paper has analyzed the relative performance of various volatility models for in-sample fitting and out-of-sample forecasting via Monte Carlo simulations with synthetic data taken from the volatility specifications $\sigma_t = \{\text{GARCH}, \text{FIGARCH}, \text{SV}, \text{MSM}\}$. The GARCH, FIGARCH and SV models are some of the most popular models of the volatility literature and each of them accounts for different facets of asset return data such as short/long memory, autoregressive components and stochastic shocks. The MSM models are a new addition to the set of volatility models available in the literature, whose DGP can reproduce various stylized facts of asset return data that traditional models ignore such as multifractality. They can also account for 'apparent' long memory, regime-switching and volatility shocks. Thus comparing all the models for σ_t with synthetic data, can give some new insights as to the most appropriate model(s) to forecast volatility even when the true model is different.

The relative in-sample estimation shows that the level of heterogeneity generated by non-MSM models is low. However, the MSM model is able to mimic a high degree of 'genuine' long memory as given by the estimated parameter of fractional differentiation in the FIGARCH DGP. The MSM DGP also produces a relatively high short-run persistence of volatility as estimated by the GARCH and SV models. The latter in-sample results for the MSM model are in line with the out-of-sample results where we obtain, in general, that the MSM model seems to best forecast volatility under any other alternative specification of the unobserved volatility process (save for the true process). In terms of relative MSE and MAE measures, the MSM model typically yields forecasts that are not too far apart from the true model's forecasts. The opposite is true for the GARCH and FIGARCH model which produce relative MSE and MAE that are significantly higher than those of the original models with stochastic shocks (MSM, SV). Results are qualitatively similar when considering 95% quantile VaR hits.

Our MC results point out that 'hybrid' specifications between (FI)GARCH, SV and MSM models could potentially improve out-of-sample forecasting. A closer look at the latter possibility by means of forecast combinations leads to sizeable improvements in terms of MSEs, MAEs and quantile VaR hits when forecasting with the 'wrong' model. Thus, alternative forecasting algorithms considering forecast combinations between the models analyzed here (and possibly others) seem a promising avenue to improve out-of-sample forecasting power of volatility models. We leave these issues for future exploration.

References

Aiolfi, M. and Timmermann, A. (2006). Persistence in forecasting performance and conditional combination strategies. *Journal of Econometrics*, 135:31–53.

Andersen, T., Bollerslev, T., Christoffersen, P., and Diebold, F. (2005a). Volatility forecasting. *PIER Working Paper No. 05-011.*

Andersen, T., Bollerslev, T., Diebold, F., and Labys, P. (2003). Modeling and forecasting realized volatility. *Econometrica*, 71:579–625.

Andersen, T., Bollerslev, T., and Meddahi, N. (2005b). Correcting the errors: volatility forecast evaluation using high frequency data and realized volatilities. *Econometrica*, 73:279–296.

Andersen, T., Chung, H., and Sorensen, B. (1999). Efficient method of moments estimation of a stochastic volatility model: a Monte Carlo study. *Journal of Econometrics*, 91:61–87.

Baillie, R., Bollerslev, T., and Mikkelsen, H. (1996). Fractionally Itegrated Generalized Autoregressive Conditional Heteroskedasticty. *Journal of Econometrics*, 74:3–30.

Basak, G., Chan, N., and Palma, W. (2001). The approximation of long-memory processes by an ARMA model. *Journal of Forecasting*, 20:367–389.

Bollerslev, T. (1986). Generalized Autoregressive Conditional Heteroskedasticity. *Journal of Econometrics*, 31:307–1986.

Brockwell, P. and Dahlhaus, R. (2004). Generalized Levinson−Durbin and Burg algorithms. *Journal of Econometrics*, 118:129–144.

Brockwell, P. and Davis, R. (1991). *Time Series: Theory and Methods*. Berlin: Springer.

Brodsky, J. and Hurvich, C. (1999). Multi-step forecasting for long-memory processes. *Journal of Forecasting*, 18:59–75.

Calvet, L. and Fisher, A. (2001). Forecasting multifractal volatility. *Journal of Econometrics*, 105:27–58.

Calvet, L. and Fisher, A. (2004). Regime-switching and the estimation of multifractal processes. *Journal of Financial Econometrics*, 2:44–83.

Calvet, L., Fisher, A., and Mandelbrot, B. (1997). A multifractal model of asset returns. *Cowles Foundation Discussion Papers: Yale University.*

Chang, C. (2002). Estimating the fractionally integrated GARCH model. *Manuscript: Academia Sinica.*

Costantini, M. and Pappalardo, C. (2009). A hierarchical procedure for the combination of forecasts. *International Journal of Forecasting*, in press.

Crato, N. and Ray, B. (1996). Model selection and forecasting for long-range dependent processes. *Journal of Forecasting*, 15:107–125.

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

Engle, R. (1982). Autoregressive Conditional Heteroskedasticity with estimates of the variance of the United Kingdom. *Econometrica*, 50:987–1006.

Gallant, A., Hsieh, D., and Tauchen, G. (1997). Estimation of stochastic volatility models with diagnostics. *Journal of Econometrics*, 81:159–192.

Granger, C. (1989). Combining forecasts–twenty years later. *Journal of Forecasting*, 8:167–173.

Granger, C. and Terasvirta, T. (1999). A simple nonlinear time series model with misleading linear properties. *Economic Letters*, 62:161–165.

Hansen, P. and Lunde, A. (2005). A forecast comparison of volatility models: does anything beat a GARCH(1,1)? *Journal of Applied Econometrics*, 20:873–889.

Harvey, A., Ruiz, E., and Shephard, N. (1994). Multivariate stochastic variance models. *Review of Economic Studies*, 61:289–306.

Kim, S., Sheppard, N., and Chibb, S. (1997). Stochastic volatility: likelihood inference and comparison with ARCH models. *Review of Economic Studies*, 81:159–192.

Liesenfeld, R. and Richard, J. (2003). Univariate and multivariate stochastic volatility models: estimation and diagnostics. *Journal of Empirical Finance*, pages 505–531.

Liu, R., di Matteo, T., and Lux, T. (2007). True and apparent scaling: The proximities of the Markov-switching multifractal model to long-range dependence. *Physica A*, 383:35–42.

Lobato, I. and Savin, N. (1998). Real and spurious long-memory properties of stock market data. *Journal of Business and Economics Statistics*, 16:261–283.

Lu, Y. and Perron, P. (2008). Modeling and forecasting stock return volatility using a random level shift model. *Working paper: Boston University*, pages 1–36.

Lux, T. (1996). Long-term stochastic dependence in financial prices: Evidence from the German stock market. *Applied Economics Letters*, 3:701–706.

Lux, T. (2008). The Markov-switching multifractal model of asset returns: GMM estimation and linear forecasting of volatility. *Journal of Business and Economic Statistics*, 26:194–210.

Lux, T. and Kaizoji, T. (2007). Forecasting volatility and volume in the Tokyo stock market: Long memory, fractality and regime switching. *Journal of Economic Dynamics and Control*, 31:1808–1843.

Lux, T. and Morales-Arias, L. (2010). Forecasting volatility under fractality, regime-switching, long memory and student-t innovations. *Computational Statistics and Data Analysis*, forthcoming.

Man, K. (2003). Long memory time series and short term forecasts. *International Journal of Forecasting*, 19:477–491.

Mandelbrot, B. (1974). Intermittent turbulence in self-similar cascades: Divergence of high moments and dimension of the carrier. *Journal of Fluid Dynamics*, 62:331–358.

Mandelbrot, B. (1999). A multifractal walk down Wall Street. *Scientific American*, 1:50–53.

Melino, A. and Turnbull, S. (1990). Pricing foreign currency options with stochastic volatility. *Journal of Econometrics*, 45:239–265.

Mills, T. (1997). Stylized facts of the temporal and distributional properties of daily FTSE returns. *Applied Financial Economics*, 7:599–604.

Muller, U. (1997). Volatilities of different time resolutions: Analyzing the dynamics of market components. *Journal of Empirical Finance*, 4:213–239.

Patton, A. and Sheppard, K. (2009). Optimal combinations of realised volatility estimators. *International Journal of Forecasting*, 25:218–238.

Poon, S. and Granger, C. (2003). Forecasting volatility in financial markets: A review. *Journal of Economic Literature*, 26:478–539.

Ruiz, E. (1994). Quasi-maximum likelihood estimation of stochastic volatility models. *Journal of Econometrics*, 63:289–306.

Schmitt, F., Schertzer, D., and Lovejoy, S. (1999). Multifractal analysis of foreign exchange data. *Applied Stochastic Models and Data Analysis*, 15:29–53.

Taylor, S. (1999a). *Financial returns modelled by the product of two stochastic processes-a study of daily sugar prices.* in: Time Series Analysis: Theory and Practice 1, ed. Anderson, O.D.. Amsterdam: North Holland.

Taylor, S. (1999b). *Modeling Financial Time Series.* Chichester: Wiley.

Vassilicos, J., Demos, A., and Tata, F. (2004). *No evidence of chaos but some evidence of multifractals in the foreign exchange and the stock market.* in: Applications of Fractals and Chaos, eds. A. Crilly, and R. Earnshaw, and H. Jones. Berlin: Springer.

Vilasuso, J. (2002). Forecasting exchange rate volatility. *Economics Letters*, 76:59–64.

Zumbach, G. (2004). Volatility processes and volatility forecast with long memory. *Quantitative Finance*, 4:70–86.

	MSM		GARCH			FIGARCH				SV		
Param.	$\hat{m}_0$	$\hat{\sigma}$	$\hat{\omega}$	$\hat{\beta}$	$\hat{\alpha}$	$\hat{\omega}$	$\hat{\beta}$	$\hat{\delta}$	$\hat{d}$	$\hat{\kappa}$	$\hat{\varphi}$	$\hat{\tau}$
MG	1.460	1.000	0.024	0.860	0.110	0.044	0.450	0.130	0.420	-0.080	0.890	0.460
SE	(0.018)	(0.021)	(0.003)	(0.013)	(0.010)	(0.005)	(0.045)	(0.045)	(0.020)	(0.016)	(0.023)	(0.056)
Max	1.782	1.993	0.113	0.967	0.376	0.154	0.901	0.954	0.754	-0.004	0.990	1.851
Min	1.000	0.923	0.000	0.452	0.032	0.000	-0.475	-0.710	0.039	-0.470	0.325	0.106

Table 1: In-sample empirical parameter estimates. Mean Group (MG) estimates of parameters for $N = 48$ all-share stock market indices at the country level with standard error (SE) in parentheses. Min (Max) is the minimum (maximum) value of the estimated parameter in the cross-section. MSM estimation is done via GMM to reduce on computational costs. Countries: Argentina, Australia, Austria, Belgium, Brazil, Canada, China, Chile, Colombia, Cyprus, Denmark, Finland, France, Germany, Greece, Hong Kong, Hungary, India, Indonesia, Ireland, Israel, Italy, Japan, Korea, Luxembourg, Netherlands, New Zealand, Norway, Pakistan, Peru, Philippine, Poland, Portugal, Romania, Russia, Singapore, South Africa, Spain, Sri Lanka, Sweden, Switzerland, Taiwan, Thailand, Turkey, United Kingdom, United States, Venezuela.

Estim.	Sim.	MSM				GARCH				FIGARCH				SV			
MSM1	$\bar{\theta}$	$\bar{m}_0$	$\bar{\sigma}$			$\bar{m}_0$	$\bar{\sigma}$			$\bar{m}_0$	$\bar{\sigma}$			$\bar{m}_0$	$\bar{\sigma}$		
	$T = 2,500$	1.459	1.016	—	—	1.218	1.018	—	—	1.221	1.049	—	—	1.385	1.065	—	—
		(0.018)	(0.131)			(0.026)	(0.101)			(0.032)	(0.242)			(0.020)	(0.071)		
	$T = 5,000$	1.459	1.005	—	—	1.220	1.031	—	—	1.228	1.049	—	—	1.387	1.062	—	—
		(0.013)	(0.096)			(0.028)	(0.087)			(0.023)	(0.191)			(0.014)	(0.047)		
MSM2	$\bar{\theta}$	$\bar{m}_0$	$\bar{\sigma}$			$\bar{m}_0$	$\bar{\sigma}$			$\bar{m}_0$	$\bar{\sigma}$			$\bar{m}_0$	$\bar{\sigma}$		
	$T = 2,500$	1.429	1.000	—	—	1.027	0.994	—	—	1.035	1.045	—	—	1.323	0.998	—	—
		(0.118)	(0.140)			(0.072)	(0.082)			(0.081)	(0.250)			(0.159)	(0.057)		
	$T = 5,000$	1.448	0.998	—	—	1.026	0.994	—	—	1.032	1.043	—	—	1.363	0.996	—	—
		(0.078)	(0.102)			(0.066)	(0.055)			(0.073)	(0.199)			(0.113)	(0.039)		
GARCH	$\bar{\theta}$	$\bar{\omega}$	$\bar{\beta}$	$\bar{\alpha}$		$\bar{\omega}$	$\bar{\beta}$	$\bar{\alpha}$		$\bar{\omega}$	$\bar{\beta}$	$\bar{\alpha}$		$\bar{\omega}$	$\bar{\beta}$	$\bar{\alpha}$	
	$T = 2,500$	0.032	0.761	0.234	—	0.032	0.858	0.110	—	0.030	0.863	0.109	—	0.089	0.691	0.240	—
		(0.017)	(0.053)	(0.051)		(0.007)	(0.018)	(0.015)		(0.014)	(0.026)	(0.018)		(0.024)	(0.044)	(0.039)	
	$T = 5,000$	0.027	0.768	0.230	—	0.031	0.858	0.111	—	0.024	0.868	0.109	—	0.084	0.701	0.235	—
		(0.010)	(0.036)	(0.036)		(0.005)	(0.013)	(0.011)		(0.009)	(0.019)	(0.012)		(0.016)	(0.030)	(0.025)	
FIGARCH	$\bar{\theta}$	$\bar{\omega}$	$\bar{\beta}$	$\bar{\delta}$	$\bar{d}$	$\bar{\omega}$	$\bar{\beta}$	$\bar{\delta}$	$\bar{d}$	$\bar{\omega}$	$\bar{\beta}$	$\bar{\delta}$	$\bar{d}$	$\bar{\omega}$	$\bar{\beta}$	$\bar{\delta}$	$\bar{d}$
	$T = 2,500$	0.036	0.593	0.185	0.689	0.032	0.649	0.288	0.477	0.044	0.432	0.111	0.421	0.084	0.608	0.514	0.348
		(0.024)	(0.226)	(0.305)	(0.231)	(0.042)	(0.154)	(0.301)	(0.252)	(0.021)	(0.136)	(0.098)	(0.073)	(0.030)	(0.185)	(0.472)	(0.384)
	$T = 5,000$	0.034	0.588	0.169	0.694	0.028	0.672	0.310	0.482	0.039	0.443	0.122	0.421	0.081	0.616	0.508	0.357
		(0.018)	(0.177)	(0.234)	(0.195)	(0.007)	(0.123)	(0.303)	(0.250)	(0.011)	(0.079)	(0.058)	(0.050)	(0.020)	(0.150)	(0.465)	(0.388)
SV	$\bar{\theta}$	$\bar{\kappa}$	$\bar{\varphi}$	$\bar{\tau}$		$\bar{\kappa}$	$\bar{\varphi}$	$\bar{\tau}$		$\bar{\kappa}$	$\bar{\varphi}$	$\bar{\tau}$		$\bar{\kappa}$	$\bar{\varphi}$	$\bar{\tau}$	
	$T = 2,500$	-0.062	0.933	0.460	—	-0.007	0.963	0.174	—	-0.006	0.970	0.154	—	-0.061	0.882	0.474	—
		(0.031)	(0.023)	(0.079)		(0.006)	(0.015)	(0.039)		(0.014)	(0.020)	(0.049)		(0.022)	(0.031)	(0.081)	
	$T = 5,000$	-0.058	0.939	0.447	—	-0.006	0.966	0.172	—	-0.004	0.977	0.144	—	-0.059	0.885	0.468	—
		(0.019)	(0.016)	(0.057)		(0.004)	(0.009)	(0.025)		(0.009)	(0.013)	(0.032)		(0.013)	(0.021)	(0.054)	

Table 2: In-sample Monte Carlo parameter estimates. Note: Estim. (Sim.) in rows (columns) denote the estimated (simulated) model, $\bar{\theta}$ is the mean value of the parameters in 400 Monte Carlo runs with finite sample standard errors (FSSE) in parentheses. A total of $T = 5,000$ and $T = 10,000$ is used to simulate the series and half is employed for in-sample estimation and the other half for out-of-sample evaluation (hence $T = 2,500$ and $T = 5,000$).

Forc.	Sim.	MSM1				MSM2				GARCH				FIGARCH				SV			
	h	1	20	50	100	1	20	50	100	1	20	50	100	1	20	50	100	1	20	50	100
MSE	$T = 2,500$	0.884	0.966	0.980	0.987	0.894	0.970	0.982	0.987	0.860	0.960	0.989	0.992	0.805	0.888	0.911	0.924	0.945	0.988	0.998	0.999
		(0.044)	(0.037)	(0.037)	(0.036)	(0.046)	(0.027)	(0.021)	(0.015)	(0.058)	(0.042)	(0.031)	(0.031)	(0.123)	(0.133)	(0.136)	(0.140)	(0.023)	(0.005)	(0.004)	(0.004)
	$T = 5,000$	0.893	0.973	0.986	0.993	0.904	0.977	0.988	0.993	0.852	0.955	0.990	0.995	0.801	0.887	0.912	0.927	0.945	0.988	0.998	0.999
		(0.026)	(0.015)	(0.013)	(0.012)	(0.031)	(0.014)	(0.009)	(0.006)	(0.050)	(0.035)	(0.023)	(0.022)	(0.107)	(0.115)	(0.117)	(0.117)	(0.016)	(0.002)	(0.002)	(0.002)
MAE	$T = 2,500$	0.862	0.955	0.978	0.990	0.880	0.964	0.983	0.993	0.925	0.978	0.997	1.000	0.881	0.929	0.947	0.960	0.955	0.998	0.999	0.999
		(0.143)	(0.143)	(0.139)	(0.136)	(0.116)	(0.083)	(0.061)	(0.044)	(0.072)	(0.070)	(0.068)	(0.068)	(0.175)	(0.180)	(0.181)	(0.182)	(0.052)	(0.047)	(0.047)	(0.047)
	$T = 5,000$	0.866	0.960	0.982	0.994	0.883	0.967	0.986	0.995	0.924	0.976	0.995	0.998	0.886	0.935	0.952	0.965	0.956	0.998	0.999	0.999
		(0.103)	(0.102)	(0.099)	(0.098)	(0.084)	(0.059)	(0.042)	(0.028)	(0.054)	(0.053)	(0.052)	(0.052)	(0.157)	(0.162)	(0.161)	(0.161)	(0.038)	(0.034)	(0.034)	(0.034)
99%	$T = 2,500$	0.017	0.019	0.019	0.019	0.015	0.018	0.019	0.019	0.010	0.013	0.014	0.014	0.010	0.013	0.014	0.014	0.013	0.018	0.018	0.018
		(0.003)	(0.004)	(0.005)	(0.006)	(0.004)	(0.006)	(0.007)	(0.008)	(0.002)	(0.004)	(0.005)	(0.005)	(0.002)	(0.003)	(0.004)	(0.006)	(0.003)	(0.004)	(0.004)	(0.004)
	$T = 5,000$	0.017	0.019	0.019	0.019	0.015	0.018	0.018	0.019	0.010	0.013	0.014	0.014	0.010	0.013	0.013	0.014	0.012	0.017	0.017	0.017
		(0.002)	(0.003)	(0.003)	(0.004)	(0.003)	(0.004)	(0.005)	(0.006)	(0.001)	(0.002)	(0.003)	(0.004)	(0.001)	(0.002)	(0.003)	(0.004)	(0.002)	(0.003)	(0.003)	(0.003)
95%	$T = 2,500$	0.045	0.042	0.042	0.041	0.039	0.041	0.041	0.041	0.051	0.048	0.047	0.047	0.050	0.048	0.047	0.047	0.038	0.045	0.045	0.045
		(0.004)	(0.006)	(0.008)	(0.010)	(0.007)	(0.010)	(0.012)	(0.014)	(0.004)	(0.007)	(0.009)	(0.010)	(0.004)	(0.006)	(0.008)	(0.011)	(0.004)	(0.006)	(0.006)	(0.006)
	$T = 5,000$	0.045	0.042	0.041	0.041	0.039	0.041	0.041	0.041	0.051	0.048	0.047	0.047	0.051	0.048	0.047	0.047	0.037	0.045	0.045	0.045
		(0.003)	(0.004)	(0.005)	(0.006)	(0.005)	(0.007)	(0.009)	(0.010)	(0.003)	(0.004)	(0.006)	(0.007)	(0.003)	(0.004)	(0.006)	(0.007)	(0.003)	(0.004)	(0.004)	(0.004)

Table 3: Out-of-sample Monte Carlo forecast evaluation with true parameters. Note: Forc. (Sim.) in rows (columns) denote the forecasting (simulated) model. The table displays average mean squared forecast errors (MSE) and mean absolute forecast errors (MAE) using the true parameters of a particular model in percentage of MSE from a naive forecast using the in-sample variance of the simulated series over 400 Monte Carlo runs for horizon h. The table also displays average 99% and 95% quantile hits with the true parameter values of a particular model over 400 Monte Carlo runs for horizon h. Finite sample standard errors (FSSE) in parentheses. A total of $T = 5,000$ and $T = 10,000$ is used to simulate the series and half is employed for in-sample estimation and the other half for out-of-sample evaluation (hence $T = 2,500$ and $T = 5,000$).

Forc.	Sim.	MSM				GARCH				FIGARCH				SV			
	h	1	20	50	100	1	20	50	100	1	20	50	100	1	20	50	100
MSM1	$T = 2,500$	0.885	0.968	0.984	0.992	0.888	0.971	0.996	1.002	0.839	0.914	0.940	0.961	0.924	1.016	1.011	1.007
		(0.043)	(0.033)	(0.030)	(0.026)	(0.037)	(0.026)	(0.021)	(0.022)	(0.118)	(0.111)	(0.095)	(0.074)	(0.023)	(0.013)	(0.011)	(0.011)
	$T = 5,000$	0.893	0.974	0.988	0.995	0.887	0.974	0.999	1.006	0.843	0.922	0.949	0.969	0.924	1.015	1.011	1.006
		(0.026)	(0.014)	(0.010)	(0.008)	(0.023)	(0.016)	(0.020)	(0.027)	(0.085)	(0.080)	(0.070)	(0.060)	(0.018)	(0.009)	(0.007)	(0.005)
MSM2	$T = 2,500$	0.901	0.974	0.987	0.993	0.984	0.995	0.999	1.000	0.970	0.983	0.988	0.992	0.941	1.006	1.003	1.001
		(0.046)	(0.027)	(0.021)	(0.015)	(0.045)	(0.020)	(0.013)	(0.010)	(0.078)	(0.053)	(0.041)	(0.031)	(0.035)	(0.009)	(0.006)	(0.004)
	$T = 5,000$	0.907	0.979	0.991	0.996	0.983	0.995	0.999	1.000	0.971	0.984	0.990	0.993	0.935	1.008	1.005	1.001
		(0.031)	(0.014)	(0.009)	(0.006)	(0.042)	(0.014)	(0.006)	(0.004)	(0.071)	(0.046)	(0.034)	(0.025)	(0.026)	(0.006)	(0.004)	(0.002)
GARCH	$T = 2,500$	0.947	1.160	1.375	1.886	0.862	0.963	0.996	1.001	0.813	0.915	0.958	0.991	0.956	1.041	1.055	1.077
		(0.052)	(0.103)	(0.354)	(0.818)	(0.057)	(0.038)	(0.021)	(0.016)	(0.120)	(0.109)	(0.082)	(0.063)	(0.043)	(0.061)	(0.122)	(0.226)
	$T = 5,000$	0.951	1.167	1.346	1.794	0.854	0.958	0.994	0.999	0.807	0.908	0.948	0.983	0.954	1.027	1.026	1.032
		(0.037)	(0.064)	(0.201)	(0.583)	(0.049)	(0.034)	(0.021)	(0.020)	(0.106)	(0.105)	(0.085)	(0.059)	(0.030)	(0.031)	(0.046)	(0.076)
FIGARCH	$T = 2,500$	0.947	1.135	1.273	1.512	0.869	0.991	1.062	1.126	0.809	0.892	0.917	0.935	0.971	1.156	1.405	1.743
		(0.059)	(0.373)	(0.548)	(0.780)	(0.069)	(0.065)	(0.104)	(0.245)	(0.123)	(0.129)	(0.128)	(0.125)	(0.048)	(0.213)	(0.685)	(1.111)
	$T = 5,000$	0.947	1.121	1.222	1.369	0.858	0.983	1.056	1.113	0.803	0.889	0.915	0.932	0.970	1.144	1.363	1.735
		(0.040)	(0.350)	(0.509)	(0.615)	(0.050)	(0.046)	(0.081)	(0.192)	(0.107)	(0.114)	(0.114)	(0.113)	(0.037)	(0.176)	(0.579)	(1.072)
SV	$T = 2,500$	0.976	0.985	0.997	1.000	0.918	0.983	1.002	1.005	0.880	0.944	0.972	0.993	0.951	0.998	0.999	0.999
		(0.102)	(0.018)	(0.014)	(0.014)	(0.051)	(0.028)	(0.018)	(0.018)	(0.107)	(0.092)	(0.073)	(0.062)	(0.027)	(0.004)	(0.004)	(0.004)
	$T = 5,000$	0.968	0.985	0.998	1.000	0.911	0.980	1.001	1.004	0.873	0.936	0.962	0.983	0.948	0.998	0.999	0.999
		(0.061)	(0.009)	(0.006)	(0.005)	(0.041)	(0.021)	(0.012)	(0.011)	(0.095)	(0.087)	(0.075)	(0.060)	(0.018)	(0.002)	(0.001)	(0.001)
CO1	$T = 2,500$	0.904	0.978	0.974	0.969	0.880	0.951	0.978	0.985	0.834	0.896	0.917	0.931	0.932	1.006	0.994	0.994
		(0.040)	(0.041)	(0.038)	(0.023)	(0.047)	(0.037)	(0.027)	(0.021)	(0.108)	(0.110)	(0.108)	(0.104)	(0.033)	(0.017)	(0.011)	(0.014)
	$T = 5,000$	0.907	0.983	0.975	0.968	0.876	0.950	0.979	0.986	0.833	0.895	0.916	0.929	0.929	1.005	0.991	0.991
		(0.031)	(0.040)	(0.022)	(0.016)	(0.039)	(0.032)	(0.025)	(0.023)	(0.096)	(0.100)	(0.101)	(0.100)	(0.023)	(0.012)	(0.007)	(0.010)
CO2	$T = 2,500$	0.922	0.987	0.981	0.975	0.882	0.954	0.981	0.987	0.834	0.909	0.947	0.972	0.936	1.010	0.997	0.996
		(0.044)	(0.043)	(0.039)	(0.027)	(0.047)	(0.035)	(0.023)	(0.018)	(0.117)	(0.106)	(0.085)	(0.064)	(0.034)	(0.018)	(0.009)	(0.013)
	$T = 5,000$	0.924	0.988	0.978	0.971	0.878	0.952	0.981	0.987	0.830	0.903	0.942	0.969	0.937	1.011	0.996	0.994
		(0.034)	(0.044)	(0.022)	(0.017)	(0.039)	(0.032)	(0.025)	(0.023)	(0.103)	(0.101)	(0.087)	(0.064)	(0.025)	(0.014)	(0.006)	(0.009)

Table 4: Out-of-sample Monte Carlo mean squared errors with estimated parameters. Note: Forc. (Sim.) in rows (columns) denote the forecasting (simulated) model. The table displays average mean squared forecast errors (MSE) in percentage of MSE from a naive forecast using the in-sample variance of the simulated series over 400 Monte Carlo runs for horizon h. Finite sample standard errors (FSSE) in parentheses. A total of $T = 5,000$ and $T = 10,000$ is used to simulate the series and half is employed for in-sample estimation and the other half for out-of-sample evaluation (hence $T = 2,500$ and $T = 5,000$).

Forc.	Sim.	MSM				GARCH				FIGARCH				SV			
	h	1	20	50	100	1	20	50	100	1	20	50	100	1	20	50	100
MSM1	$T = 2,500$	0.860	0.956	0.982	0.998	0.930	0.980	0.999	1.009	0.885	0.931	0.953	0.971	0.946	1.052	1.060	1.064
		(0.131)	(0.110)	(0.094)	(0.081)	(0.056)	(0.044)	(0.041)	(0.043)	(0.140)	(0.114)	(0.092)	(0.073)	(0.047)	(0.040)	(0.035)	(0.035)
	$T = 5,000$	0.864	0.959	0.982	0.995	0.934	0.987	1.009	1.019	0.883	0.930	0.953	0.971	0.949	1.054	1.062	1.063
		(0.093)	(0.075)	(0.061)	(0.052)	(0.042)	(0.040)	(0.045)	(0.054)	(0.117)	(0.092)	(0.073)	(0.057)	(0.031)	(0.026)	(0.023)	(0.022)
MSM2	$T = 2,500$	0.886	0.964	0.982	0.991	0.990	0.996	0.998	0.999	0.980	0.988	0.991	0.994	0.950	1.005	1.003	1.001
		(0.116)	(0.083)	(0.061)	(0.044)	(0.037)	(0.026)	(0.022)	(0.019)	(0.068)	(0.049)	(0.038)	(0.028)	(0.048)	(0.035)	(0.026)	(0.019)
	$T = 5,000$	0.885	0.966	0.984	0.993	0.990	0.996	0.998	0.998	0.980	0.987	0.991	0.993	0.946	1.007	1.004	1.001
		(0.084)	(0.059)	(0.042)	(0.028)	(0.024)	(0.015)	(0.013)	(0.012)	(0.059)	(0.042)	(0.031)	(0.022)	(0.033)	(0.024)	(0.018)	(0.013)
GARCH	$T = 2,500$	0.922	1.281	1.723	2.402	0.925	0.978	1.000	1.006	0.883	0.942	0.977	1.006	0.960	1.145	1.208	1.239
		(0.142)	(0.173)	(0.379)	(0.717)	(0.063)	(0.046)	(0.032)	(0.028)	(0.154)	(0.116)	(0.080)	(0.059)	(0.044)	(0.100)	(0.205)	(0.282)
	$T = 5,000$	0.933	1.289	1.731	2.459	0.924	0.976	0.997	1.001	0.887	0.941	0.972	1.003	0.962	1.130	1.171	1.185
		(0.105)	(0.129)	(0.264)	(0.555)	(0.049)	(0.038)	(0.031)	(0.030)	(0.142)	(0.116)	(0.083)	(0.054)	(0.033)	(0.058)	(0.102)	(0.138)
FIGARCH	$T = 2,500$	0.914	1.224	1.524	1.902	0.941	1.033	1.106	1.182	0.883	0.933	0.954	0.973	0.978	1.307	1.629	1.955
		(0.150)	(0.336)	(0.559)	(0.755)	(0.075)	(0.077)	(0.112)	(0.208)	(0.170)	(0.165)	(0.158)	(0.151)	(0.054)	(0.262)	(0.694)	(1.081)
	$T = 5,000$	0.920	1.204	1.472	1.791	0.939	1.031	1.105	1.180	0.887	0.935	0.955	0.970	0.982	1.307	1.618	1.977
		(0.105)	(0.250)	(0.497)	(0.609)	(0.051)	(0.054)	(0.092)	(0.180)	(0.154)	(0.152)	(0.147)	(0.141)	(0.043)	(0.245)	(0.644)	(1.072)
SV	$T = 2,500$	1.072	0.975	0.980	0.984	1.016	1.027	1.031	1.033	0.970	0.988	1.001	1.015	0.970	0.999	1.000	1.000
		(0.128)	(0.056)	(0.046)	(0.045)	(0.045)	(0.032)	(0.029)	(0.030)	(0.126)	(0.097)	(0.073)	(0.060)	(0.054)	(0.032)	(0.032)	(0.032)
	$T = 5,000$	1.091	0.972	0.972	0.975	1.019	1.029	1.032	1.033	0.975	0.987	0.995	1.007	0.971	0.999	1.000	1.000
		(0.096)	(0.040)	(0.032)	(0.031)	(0.030)	(0.022)	(0.020)	(0.020)	(0.123)	(0.099)	(0.076)	(0.057)	(0.036)	(0.021)	(0.021)	(0.021)
CO1	$T = 2,500$	0.906	0.986	1.013	1.045	0.925	0.968	0.995	1.011	0.885	0.916	0.932	0.947	0.960	1.041	1.053	1.070
		(0.123)	(0.091)	(0.068)	(0.051)	(0.054)	(0.044)	(0.034)	(0.029)	(0.133)	(0.122)	(0.113)	(0.103)	(0.046)	(0.048)	(0.049)	(0.057)
	$T = 5,000$	0.916	0.994	1.014	1.046	0.928	0.970	0.998	1.013	0.887	0.914	0.927	0.939	0.960	1.040	1.053	1.074
		(0.093)	(0.066)	(0.048)	(0.040)	(0.042)	(0.034)	(0.029)	(0.029)	(0.125)	(0.119)	(0.113)	(0.108)	(0.032)	(0.039)	(0.043)	(0.053)
CO2	$T = 2,500$	0.934	1.015	1.041	1.067	0.930	0.975	1.001	1.015	0.884	0.929	0.961	0.983	0.943	1.049	1.056	1.068
		(0.128)	(0.091)	(0.070)	(0.058)	(0.054)	(0.042)	(0.030)	(0.027)	(0.133)	(0.104)	(0.078)	(0.056)	(0.043)	(0.045)	(0.045)	(0.054)
	$T = 5,000$	0.944	1.018	1.034	1.059	0.933	0.978	1.004	1.017	0.887	0.927	0.957	0.981	0.948	1.052	1.058	1.069
		(0.098)	(0.067)	(0.049)	(0.044)	(0.042)	(0.034)	(0.028)	(0.028)	(0.121)	(0.101)	(0.077)	(0.052)	(0.034)	(0.042)	(0.044)	(0.055)

Table 5: Out-of-sample Monte Carlo mean absolute errors with estimated parameters. Note: Forc. (Sim.) in rows (columns) denote the forecasting (simulated) model. The table displays average mean squared forecast errors (MAE) in percentage of MAE from a naive forecast using the in-sample variance of the simulated series over 400 Monte Carlo runs for horizon h. Finite sample standard errors (FSSE) in parentheses. A total of $T = 5,000$ and $T = 10,000$ is used to simulate the series and half is employed for in-sample estimation and the other half for out-of-sample evaluation (hence $T = 2,500$ and $T = 5,000$).

Forc.	Sim.	MSM				GARCH				FIGARCH				SV			
	h	1	20	50	100	1	20	50	100	1	20	50	100	1	20	50	100
MSM1	$T = 2,500$	0.017 (0.003)	0.019 (0.005)	0.019 (0.006)	0.019 (0.008)	0.009 (0.003)	0.014 (0.005)	0.015 (0.006)	0.015 (0.007)	0.011 (0.008)	0.016 (0.014)	0.018 (0.016)	0.019 (0.019)	0.016 (0.003)	0.018 (0.004)	0.016 (0.004)	0.015 (0.004)
	$T = 5,000$	0.017 (0.002)	0.019 (0.004)	0.019 (0.004)	0.019 (0.005)	0.009 (0.002)	0.013 (0.004)	0.014 (0.004)	0.014 (0.005)	0.010 (0.005)	0.014 (0.009)	0.016 (0.011)	0.017 (0.013)	0.016 (0.002)	0.018 (0.003)	0.016 (0.003)	0.015 (0.003)
MSM2	$T = 2,500$	0.015 (0.004)	0.018 (0.006)	0.019 (0.007)	0.020 (0.008)	0.014 (0.007)	0.015 (0.007)	0.015 (0.007)	0.015 (0.007)	0.019 (0.020)	0.020 (0.020)	0.020 (0.020)	0.020 (0.021)	0.016 (0.003)	0.019 (0.004)	0.018 (0.004)	0.018 (0.004)
	$T = 5,000$	0.015 (0.003)	0.018 (0.004)	0.019 (0.005)	0.019 (0.006)	0.014 (0.005)	0.014 (0.005)	0.014 (0.005)	0.014 (0.005)	0.017 (0.014)	0.018 (0.014)	0.018 (0.015)	0.018 (0.015)	0.016 (0.002)	0.019 (0.003)	0.018 (0.003)	0.018 (0.003)
GARCH	$T = 2,500$	0.018 (0.004)	0.013 (0.006)	0.008 (0.005)	0.005 (0.005)	0.010 (0.002)	0.013 (0.004)	0.014 (0.006)	0.014 (0.007)	0.010 (0.004)	0.015 (0.010)	0.017 (0.015)	0.018 (0.019)	0.016 (0.003)	0.013 (0.004)	0.011 (0.005)	0.011 (0.005)
	$T = 5,000$	0.018 (0.003)	0.013 (0.004)	0.008 (0.003)	0.004 (0.003)	0.010 (0.002)	0.013 (0.003)	0.014 (0.004)	0.014 (0.004)	0.010 (0.003)	0.014 (0.007)	0.016 (0.011)	0.017 (0.014)	0.016 (0.002)	0.013 (0.003)	0.011 (0.003)	0.011 (0.004)
FIGARCH	$T = 2,500$	0.018 (0.003)	0.014 (0.005)	0.011 (0.006)	0.008 (0.006)	0.010 (0.002)	0.012 (0.004)	0.012 (0.005)	0.010 (0.006)	0.010 (0.002)	0.013 (0.004)	0.013 (0.006)	0.014 (0.007)	0.016 (0.003)	0.011 (0.005)	0.008 (0.006)	0.007 (0.006)
	$T = 5,000$	0.018 (0.003)	0.015 (0.004)	0.011 (0.004)	0.008 (0.004)	0.009 (0.002)	0.011 (0.003)	0.011 (0.004)	0.010 (0.005)	0.010 (0.002)	0.013 (0.003)	0.013 (0.004)	0.014 (0.005)	0.016 (0.002)	0.011 (0.004)	0.008 (0.005)	0.007 (0.006)
SV	$T = 2,500$	0.012 (0.004)	0.020 (0.009)	0.022 (0.011)	0.022 (0.012)	0.007 (0.003)	0.011 (0.005)	0.013 (0.006)	0.013 (0.007)	0.008 (0.007)	0.013 (0.012)	0.016 (0.016)	0.017 (0.018)	0.013 (0.003)	0.018 (0.005)	0.018 (0.005)	0.018 (0.005)
	$T = 5,000$	0.011 (0.003)	0.019 (0.006)	0.021 (0.007)	0.021 (0.007)	0.006 (0.002)	0.010 (0.003)	0.012 (0.004)	0.012 (0.004)	0.007 (0.004)	0.011 (0.007)	0.013 (0.010)	0.014 (0.012)	0.013 (0.002)	0.018 (0.003)	0.018 (0.003)	0.018 (0.003)
CO1	$T = 2,500$	0.015 (0.003)	0.017 (0.006)	0.016 (0.006)	0.014 (0.006)	0.009 (0.003)	0.012 (0.005)	0.012 (0.005)	0.012 (0.006)	0.011 (0.006)	0.014 (0.009)	0.015 (0.010)	0.015 (0.012)	0.014 (0.003)	0.016 (0.004)	0.014 (0.004)	0.013 (0.005)
	$T = 5,000$	0.015 (0.002)	0.017 (0.004)	0.015 (0.004)	0.013 (0.004)	0.009 (0.002)	0.012 (0.003)	0.012 (0.004)	0.012 (0.004)	0.010 (0.005)	0.013 (0.006)	0.014 (0.007)	0.015 (0.008)	0.014 (0.002)	0.016 (0.003)	0.014 (0.003)	0.013 (0.004)
CO2	$T = 2,500$	0.015 (0.003)	0.017 (0.006)	0.015 (0.007)	0.013 (0.006)	0.009 (0.003)	0.012 (0.005)	0.012 (0.005)	0.012 (0.006)	0.011 (0.008)	0.015 (0.012)	0.017 (0.016)	0.018 (0.019)	0.016 (0.003)	0.016 (0.004)	0.014 (0.004)	0.013 (0.004)
	$T = 5,000$	0.015 (0.002)	0.017 (0.004)	0.015 (0.004)	0.013 (0.004)	0.009 (0.002)	0.012 (0.003)	0.012 (0.004)	0.012 (0.004)	0.011 (0.006)	0.014 (0.009)	0.016 (0.012)	0.017 (0.014)	0.016 (0.002)	0.016 (0.003)	0.014 (0.004)	0.013 (0.004)

Table 6: Monte Carlo quantile Value-at-Risk hits with estimated parameters. Note: Forc. (Sim.) in rows (columns) denote the forecasting (simulated) model. The table displays average 99% quantile hits over 400 Monte Carlo runs for horizon h. Finite sample standard errors (FSSE) in parentheses. A total of $T = 5,000$ and $T = 10,000$ is used to simulate the series and half is employed for in-sample estimation and the other half for out-of-sample evaluation (hence $T = 2,500$ and $T = 5,000$).

Forc.	Sim.	MSM				GARCH				FIGARCH				SV			
	h	1	20	50	100	1	20	50	100	1	20	50	100	1	20	50	100
MSM1	$T = 2,500$	0.045 (0.005)	0.042 (0.009)	0.042 (0.011)	0.042 (0.013)	0.047 (0.007)	0.049 (0.010)	0.049 (0.012)	0.048 (0.013)	0.049 (0.017)	0.052 (0.024)	0.052 (0.029)	0.053 (0.032)	0.046 (0.004)	0.045 (0.006)	0.042 (0.007)	0.040 (0.007)
	$T = 5,000$	0.045 (0.003)	0.042 (0.006)	0.042 (0.008)	0.041 (0.009)	0.046 (0.005)	0.047 (0.008)	0.047 (0.009)	0.046 (0.010)	0.048 (0.012)	0.049 (0.018)	0.050 (0.022)	0.050 (0.025)	0.046 (0.003)	0.045 (0.004)	0.042 (0.004)	0.040 (0.005)
MSM2	$T = 2,500$	0.039 (0.007)	0.041 (0.010)	0.042 (0.012)	0.042 (0.014)	0.048 (0.014)	0.049 (0.014)	0.049 (0.014)	0.049 (0.014)	0.054 (0.034)	0.055 (0.034)	0.055 (0.035)	0.055 (0.036)	0.045 (0.005)	0.047 (0.007)	0.046 (0.007)	0.046 (0.008)
	$T = 5,000$	0.039 (0.005)	0.041 (0.007)	0.041 (0.009)	0.042 (0.010)	0.047 (0.010)	0.048 (0.010)	0.048 (0.010)	0.048 (0.010)	0.051 (0.026)	0.051 (0.027)	0.051 (0.027)	0.051 (0.028)	0.045 (0.004)	0.048 (0.004)	0.046 (0.004)	0.046 (0.005)
GARCH	$T = 2,500$	0.046 (0.007)	0.031 (0.010)	0.020 (0.010)	0.014 (0.010)	0.051 (0.005)	0.049 (0.009)	0.048 (0.012)	0.048 (0.014)	0.050 (0.011)	0.050 (0.021)	0.050 (0.029)	0.051 (0.034)	0.046 (0.005)	0.035 (0.008)	0.032 (0.010)	0.031 (0.011)
	$T = 5,000$	0.046 (0.005)	0.031 (0.007)	0.020 (0.007)	0.012 (0.006)	0.051 (0.004)	0.048 (0.006)	0.047 (0.008)	0.047 (0.010)	0.051 (0.008)	0.050 (0.015)	0.049 (0.022)	0.049 (0.027)	0.047 (0.003)	0.036 (0.006)	0.032 (0.007)	0.032 (0.007)
FIGARCH	$T = 2,500$	0.046 (0.006)	0.033 (0.010)	0.025 (0.011)	0.019 (0.011)	0.049 (0.005)	0.044 (0.008)	0.041 (0.011)	0.037 (0.014)	0.050 (0.006)	0.048 (0.009)	0.047 (0.012)	0.046 (0.015)	0.045 (0.005)	0.031 (0.010)	0.024 (0.013)	0.021 (0.015)
	$T = 5,000$	0.047 (0.004)	0.034 (0.007)	0.026 (0.008)	0.020 (0.009)	0.049 (0.004)	0.044 (0.006)	0.040 (0.009)	0.036 (0.012)	0.051 (0.004)	0.048 (0.006)	0.047 (0.008)	0.046 (0.011)	0.046 (0.003)	0.030 (0.009)	0.024 (0.012)	0.021 (0.014)
SV	$T = 2,500$	0.031 (0.008)	0.043 (0.016)	0.045 (0.018)	0.045 (0.019)	0.037 (0.009)	0.042 (0.011)	0.043 (0.013)	0.043 (0.014)	0.038 (0.016)	0.045 (0.024)	0.047 (0.029)	0.049 (0.033)	0.038 (0.006)	0.045 (0.009)	0.045 (0.009)	0.045 (0.009)
	$T = 5,000$	0.030 (0.005)	0.042 (0.010)	0.044 (0.011)	0.045 (0.012)	0.036 (0.006)	0.041 (0.007)	0.042 (0.009)	0.042 (0.009)	0.035 (0.011)	0.041 (0.016)	0.043 (0.021)	0.044 (0.024)	0.038 (0.004)	0.045 (0.006)	0.045 (0.006)	0.045 (0.006)
CO1	$T = 2,500$	0.041 (0.006)	0.040 (0.010)	0.036 (0.011)	0.033 (0.011)	0.047 (0.007)	0.046 (0.010)	0.045 (0.012)	0.043 (0.013)	0.049 (0.015)	0.050 (0.019)	0.050 (0.021)	0.050 (0.024)	0.043 (0.005)	0.041 (0.007)	0.038 (0.008)	0.037 (0.009)
	$T = 5,000$	0.041 (0.004)	0.039 (0.007)	0.036 (0.008)	0.032 (0.007)	0.047 (0.005)	0.047 (0.007)	0.045 (0.008)	0.043 (0.009)	0.049 (0.011)	0.049 (0.014)	0.049 (0.016)	0.049 (0.017)	0.043 (0.003)	0.042 (0.005)	0.038 (0.006)	0.036 (0.007)
CO2	$T = 2,500$	0.041 (0.006)	0.038 (0.011)	0.035 (0.012)	0.032 (0.012)	0.046 (0.007)	0.045 (0.010)	0.044 (0.012)	0.043 (0.013)	0.049 (0.018)	0.051 (0.025)	0.051 (0.030)	0.052 (0.034)	0.045 (0.004)	0.041 (0.007)	0.038 (0.008)	0.037 (0.008)
	$T = 5,000$	0.041 (0.004)	0.038 (0.007)	0.035 (0.008)	0.032 (0.008)	0.046 (0.005)	0.045 (0.007)	0.044 (0.008)	0.043 (0.009)	0.049 (0.014)	0.050 (0.019)	0.050 (0.024)	0.051 (0.027)	0.046 (0.003)	0.042 (0.005)	0.039 (0.006)	0.037 (0.007)

Table 7: Monte Carlo quantile Value-at-Risk hits with estimated parameters. Note: Forc. (Sim.) in rows (columns) denote the forecasting (simulated) model. The table displays average 95% quantile hits over 400 Monte Carlo runs for horizon h. Finite sample standard errors (FSSE) in parentheses. A total of $T = 5,000$ and $T = 10,000$ is used to simulate the series and half is employed for in-sample estimation and the other half for out-of-sample evaluation (hence $T = 2,500$ and $T = 5,000$).

Figure 1: Boxplots of MSE for the MSM (with ML), GARCH, FIGARCH and SV forecasting models (rows) under the alternative DGPs (columns). The box has lines at the lower quartile, median, and upper quartile values. The whiskers show the extent of the rest of the data. Outliers are data with values beyond the ends of the whiskers. Boxplots are in the same scale for easy comparability.

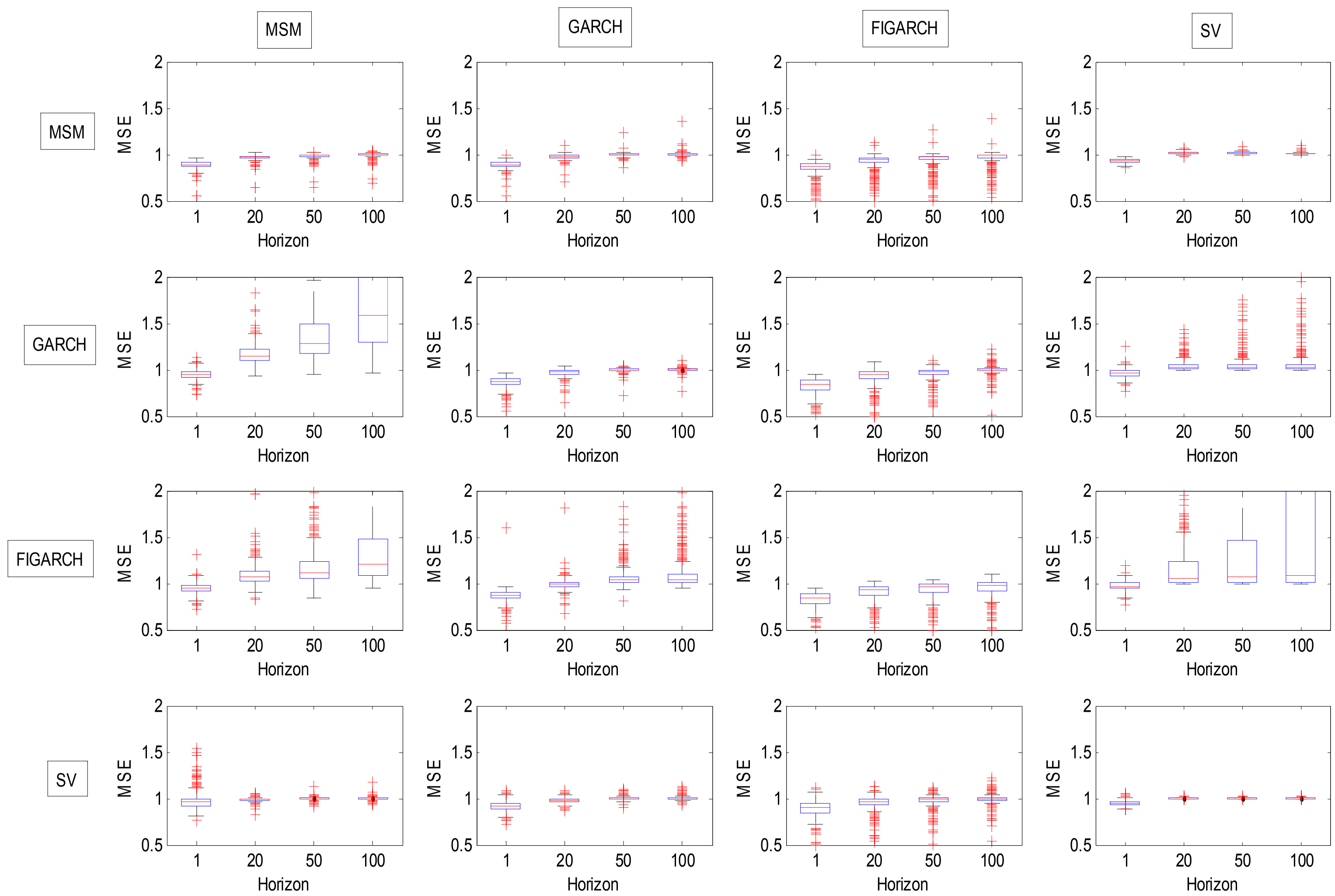

Figure 2: Boxplots of 95% quantile VaR hits for the MSM, GARCH, FIGARCH and SV forecasting models (rows) under the alternative GDPs (columns). Boxplots are in the same scale for easy comparability.

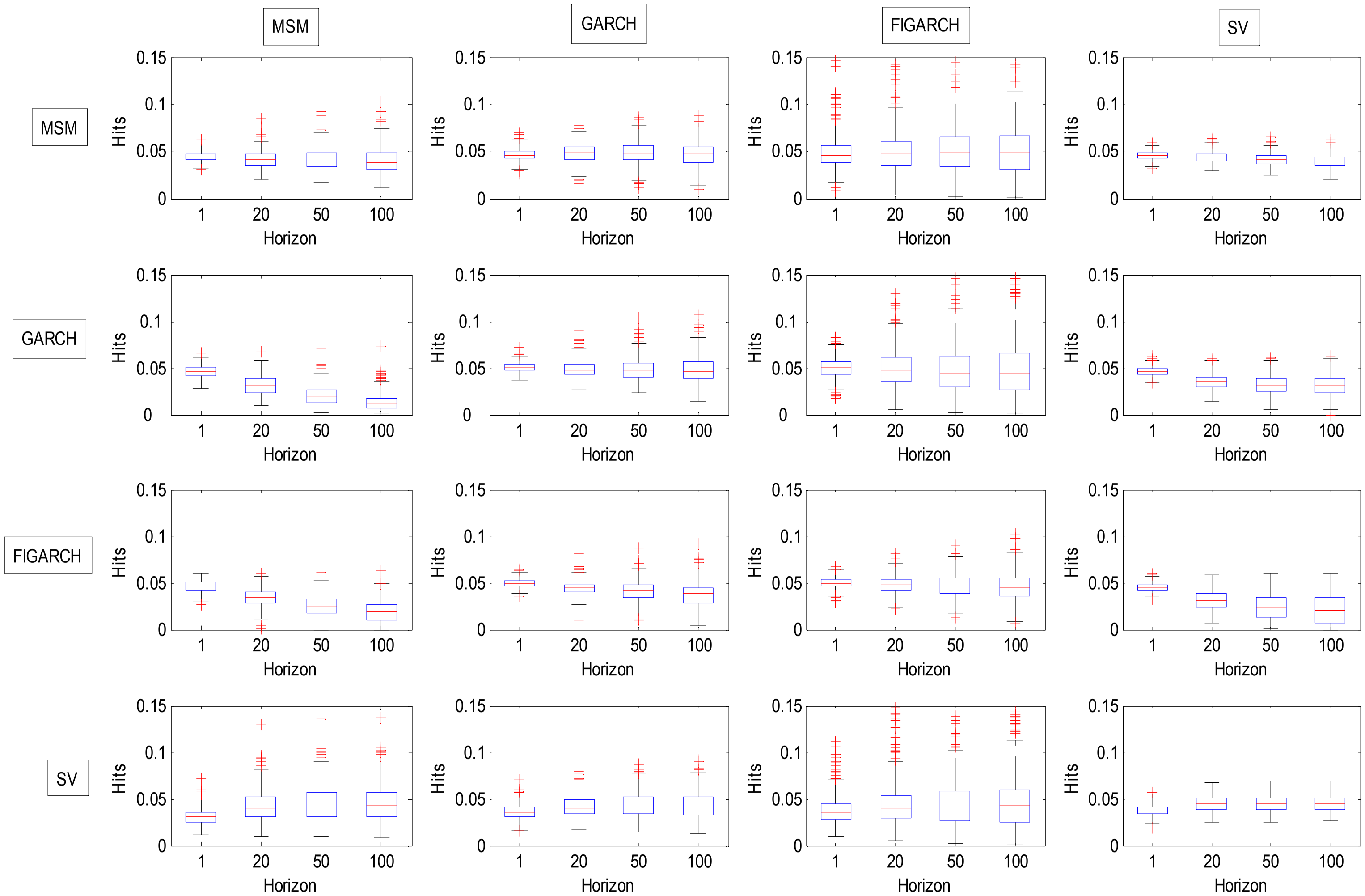

Figure 3: Boxplots of MSE for the MSM, GARCH and CO2 models (rows) under alternative GDPs (columns). Boxplots are in the same scale for easy comparability.

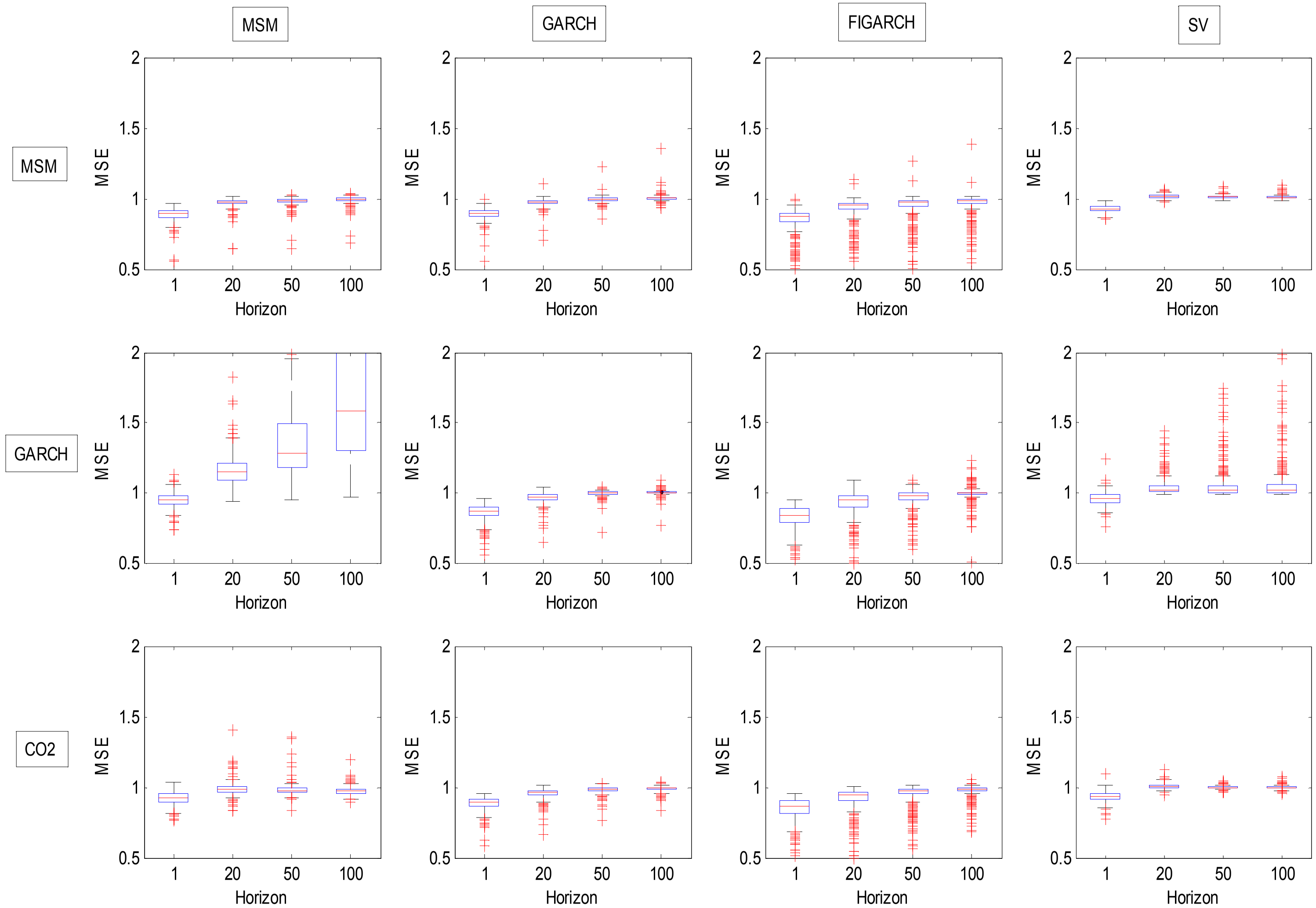

Figure 4: Boxplots of 95% quantile VaR hits for the MSM, GARCH and CO2 models (rows) under alternative GDPs (columns). Boxplots are in the same scale for easy comparability.