International Journal of Forecasting 34 (2018) 733–747

Contents lists available at ScienceDirect

International Journal of Forecasting

journal homepage: www.elsevier.com/locate/ijforecast

Forecasting risk with Markov-switching GARCH models: A large-scale performance study

David Ardia[a,b,*], Keven Bluteau[a,c], Kris Boudt[c,d], Leopoldo Catania[e]

[a] Institute of Financial Analysis, University of Neuchâtel, Neuchâtel, Switzerland
[b] Department of Finance, Insurance and Real Estate, Laval University, Québec City, Canada
[c] Solvay Business School, Vrije Universiteit Brussel, Belgium
[d] School of Business and Economics, Vrije Universiteit Amsterdam, The Netherlands
[e] Department of Economics and Business Economics, Aarhus University and CREATES, Denmark

ARTICLE INFO

Keywords:
GARCH
MSGARCH
Forecasting performance
Large-scale study
Value-at-risk
Expected shortfall
Risk management

ABSTRACT

We perform a large-scale empirical study in order to compare the forecasting performances of single-regime and Markov-switching GARCH (MSGARCH) models from a risk management perspective. We find that MSGARCH models yield more accurate Value-at-Risk, expected shortfall, and left-tail distribution forecasts than their single-regime counterparts for daily, weekly, and ten-day equity log-returns. Also, our results indicate that accounting for parameter uncertainty improves the left-tail predictions, independently of the inclusion of the Markov-switching mechanism.

1. Introduction

Under the regulation of the Basel Accords, risk managers of financial institutions must make use of state-of-the-art methodologies for monitoring financial risks (Board of Governors of the Federal Reserve Systems, 2012). Clearly, regime-switching time-varying volatility models and Bayesian estimation methods can be considered strong candidates for being classified as state-of-the-art methodologies. However, many academics and practitioners also consider the single-regime volatility model and the use of frequentist estimation via maximum likelihood (ML) as state-of-the-art. Risk managers disagree as to whether the computational complexity of a regime-switching model and the Bayesian estimation method pay off in terms of a higher accuracy of their financial risk monitoring system. We study this question in the context of monitoring the individual risks of a large number of financial assets.

The specification of the conditional volatility process is key among the various building-blocks of any risk management system, especially for short horizons (McNeil, Frey, & Embrechts, 2015). Research on the use of time series models for modeling the volatility has proliferated since the creation of the original ARCH model by Engle (1982) and its generalization by Bollerslev (1986), and multiple extensions of the GARCH scedastic function have been proposed for capturing additional stylized facts that are observed in financial markets, such as nonlinearities, asymmetries, and long-memory properties; see Engle (2004) for a review. These so-called GARCH-type models are essential tools for risk managers today.

An appropriate risk model should be able to accommodate the properties of financial returns. Recent academic studies have shown that many financial assets exhibit structural breaks in their volatility dynamics, and that ignoring this feature can have a big effect on the precision

* Correspondence to: University of Neuchâtel, Rue A.-L. Breguet 2, CH 2000, Neuchâtel, Switzerland.
E-mail addresses: david.ardia@unine.ch (D. Ardia), keven.bluteau@unine.ch (K. Bluteau), kris.boudt@vub.be (K. Boudt), leopoldo.catania@econ.au.dk (L. Catania).

https://doi.org/10.1016/j.ijforecast.2018.05.004
0169-2070/

of volatility forecasts (see e.g. Bauwens, Dufays, & Rombouts, 2014; Lamoureux & Lastrapes, 1990). As was noted by Danielsson (2011), this shortcoming in individual forecasting systems can have systemic consequences. Indeed, he refers to these single-regime volatility models as one of the culprits of the great financial crisis: *"(...) the stochastic process governing market prices is very different during times of stress compared to normal times. We need different models during crisis and non-crisis and need to be careful in drawing conclusions from non-crisis data about what happens in crises and vice versa"*.

One way to address the *switch* in the return process is provided by Markov-switching GARCH models (MSGARCH), whose parameters can change over time according to a discrete latent (i.e., unobservable) variable. These models can adapt quickly to variations in the unconditional volatility level, which improves risk predictions (see e.g. Ardia, 2008; Marcucci, 2005).

The initial studies of Markov-switching autoregressive heteroscedastic models applied to financial time series focused on ARCH specifications, and thus omitted a lagged value of the conditional variance in the variance equation (Cai, 1994; Hamilton & Susmel, 1994). The use of ARCH rather than GARCH dynamics leads to computational tractability in the likelihood calculation. Indeed, Gray (1996) shows that, given a Markov chain with K regimes and T observations, the evaluation of the likelihood of a Markov-switching model with general GARCH dynamics requires integration over all K^T possible paths, rendering the estimation infeasible. While this difficulty is not present in ARCH specifications, the use of lower-order GARCH models tends to offer a more parsimonious representation than higher-order ARCH models.

Dueker (1997), Gray (1996) and Klaassen (2002) tackle the *path-dependence problem* of MSGARCH through approximation, by collapsing the past regime-specific conditional variances based on ad hoc schemes. A further solution is to consider alternatives to traditional maximum likelihood estimation. Bauwens et al. (2014) recommended the use of Bayesian estimation methods that are still feasible through so-called data augmentation and particle MCMC techniques. Augustyniak (2014) relied on a Monte Carlo EM algorithm with importance sampling. In our study, we consider the alternative approach provided by Haas, Mittnik, and Paolella (2004), who let the GARCH process of each state evolves independently of those in the other states. In addition to avoiding the path-dependence problem that arises with traditional maximum likelihood estimation, their model also allows for a clear-cut interpretation of the variance dynamics in each regime.

The first contribution of our paper is to test whether MSGARCH models do indeed provide risk managers with useful tools that can improve their volatility forecasts.[1] We answer this question by performing a large-scale empirical analysis in which we compare the risk forecasting performances of single-regime and Markov-switching GARCH models. We take the perspective of a risk manager who is working for a fund manager and conduct our study on the daily, weekly and ten-day log-returns of a large universe of stocks, equity indices, and foreign exchange rates. Thus, in contrast to Hansen and Lunde (2005), who compare a large number of GARCH-type models on a few series, we focus on a few GARCH and MSGARCH models and a large number of series. For single-regime and Markov-switching specifications, the scedastic specifications that we consider account for different reactions of the conditional volatility to past asset returns. More precisely, we consider both the symmetric GARCH model (Bollerslev, 1986) and the asymmetric GJR model (Glosten, Jagannathan, & Runkle, 1993). These scedastic specifications are integrated into the MSGARCH framework using the approach of Haas et al. (2004). For the (regime-dependent) conditional distributions, we use the symmetric and Fernández and Steel (1998) skewed versions of the normal and Student-t distributions. This leads to a total of sixteen models.

Our second contribution is to test the impact of the estimation method on the performance of the volatility forecasting model. Traditionally, GARCH and MSGARCH models are estimated using a frequentist (typically via ML) approach; see Augustyniak (2014), Haas et al. (2004) and Marcucci (2005). However, several recent studies have argued that a Bayesian approach offers some advantages. For instance, Markov chain Monte Carlo (MCMC) procedures can explore the joint posterior distribution of the model parameters, and parameter uncertainty is integrated into the risk forecasts naturally via the predictive distribution (Ardia, 2008; Ardia, Kolly, & Trottier, 2017; Bauwens, De Backer, & Dufays, 2014; Bauwens, Preminger, & Rombouts, 2010; Geweke & Amisano, 2010).

Combining the sixteen model specifications using the frequentist and Bayesian estimation methods, we obtain 32 possible candidates for a state-of-the-art methodology for monitoring the financial risk. We use an out-of-sample evaluation period of 2,000 days, from (approximately) 2005 to 2016, consisting of daily log-returns. We evaluate the accuracy of the risk prediction models in terms of the Value-at-Risk (VaR), the expected shortfall (ES), and the left-tail (i.e., losses) of the conditional distribution of the assets' returns.

Our empirical results suggest a number of practical insights, which can be summarized as follows. First, we

[1] Our study focuses exclusively on GARCH and MSGARCH models. GARCH is the workhorse model in financial econometrics and has been being investigated for decades. It is used widely by practitioners and academics; see for instance Bams, Blanchard, and Lehnert (2017) and Herwartz (2017). MSGARCH is the most natural and straightforward extension to GARCH. Alternative conditional volatility models include stochastic volatility models (Jacquier, Polson, & Rossi, 1994; Taylor, 1994), realized measure-based conditional volatility models such as HEAVY (Shephard & Sheppard, 2010) or realized GARCH (Hansen, Huang, & Shek, 2011), or even combinations of these (Opschoor, van Dijk, & van der Wel, 2017). Finally, note that our study considers only the (1,1)-lag specification for the GARCH and MSGARCH models. While considering higher orders for (MS)GARCH model specifications has a clear computational cost, the payoff in terms of improvements in forecasting precision may be low. In fact, several studies have shown that increasing the orders does not lead to any substantial improvement in the forecasting performance in the case of predicting the conditional variance of asset returns (see e.g. Hansen & Lunde, 2005). We tested whether this result also holds for our sample and investigated the fits of GARCH(p, q) and GJR(p, 1, q) models over the three universes of stocks, indices and foreign exchange rates, for rolling windows of 1500 points, and selected the best in-sample model via the BIC. We found that the (1,1) specification is selected in the vast majority of cases.

find that MSGARCH models deliver better VaR, ES, and left-tail distribution forecasts than their single-regime counterparts. This is especially true for stock return data. Moreover, the improvements are more pronounced when the Markov-switching mechanism is applied to simple specifications such as the GARCH-Normal model. Second, accounting for parameter uncertainty improves the accuracy of the left-tail predictions, independently of the inclusion of the Markov-switching mechanism. Moreover, larger improvements are observed in the case of single-regime models. Overall, we recommend risk managers to rely on more flexible models and to perform inference accounting for parameter uncertainty.

In addition to showing the good performance of MSGARCH models and Bayesian estimation methods, we refer risk managers to our R package MSGARCH (Ardia, Bluteau, Boudt, Catania, Peterson, & Trottier, 2017; Ardia, Bluteau, Boudt, Catania, & Trottier, forthcoming), which implements MSGARCH models in the R statistical language with efficient C++ code.[2] We hope that this paper and the accompanying package will encourage practitioners and academics in the financial community to use MSGARCH models and Bayesian estimation methods.

The paper proceeds as follows. Model specification, estimation, and forecasting are presented in Section 2. The datasets, testing design, and empirical results are discussed in Section 3. Section 4 concludes.

2. Risk forecasting with Markov-switching GARCH models

A key aspect of quantitative risk management is the modeling of the risk drivers of the securities held by the fund manager. Here, we consider the univariate parametric framework, which computes the desired risk measure in four steps. First, a statistical model which describes the daily log-returns (profit and loss, P&L) dynamics is determined. Second, the model parameters are estimated for a given estimation window. Third, the one-/multi-day-ahead distribution of log-returns is obtained (either analytically or by simulation). Fourth, relevant risk measures such as the Value-at-Risk (VaR) and the expected shortfall (ES) are computed from the distribution. The VaR represents a quantile of the distribution of log-returns at the desired horizon, and the ES is the loss expected when the loss exceeds the VaR level (Jorion, 2006). Risk managers can then allocate their risk capital based on their density or risk measure forecasts, and can also assess the quality of the risk model *ex post* via statistical procedures that are referred to as *backtesting*.

2.1. Model specification

We define $y_t \in \mathbb{R}$ as the log-return of a financial asset at time t. To simplify the exposition, we assume that the log-returns have a zero mean and are not autocorrelated.[3] The general Markov-switching GARCH specification can be expressed as:

$$y_t \mid (s_t = k, \mathcal{I}_{t-1}) \sim \mathcal{D}(0, h_{k,t}, \boldsymbol{\xi}_k)\,, \tag{1}$$

where $\mathcal{D}(0, h_{k,t}, \boldsymbol{\xi}_k)$ is a continuous distribution with a zero mean, time-varying variance $h_{k,t}$, and additional shape parameters (e.g., asymmetry) gathered in the vector $\boldsymbol{\xi}_k$.[4] Furthermore, we assume that the latent variable s_t, defined on the discrete space $\{1, \ldots, K\}$, evolves according to an unobserved first-order ergodic homogeneous Markov chain with transition probability matrix $\mathbf{P} \equiv \{p_{i,j}\}_{i,j=1}^{K}$, where $p_{i,j} \equiv \mathbb{P}[s_t = j \mid s_{t-1} = i]$. We denote the information set up to time $t-1$ by $\mathcal{I}_{t-1}$, that is, $\mathcal{I}_{t-1} \equiv \{y_{t-i}, i > 0\}$. Given the parametrization of $\mathcal{D}(\cdot)$, we have $\mathbb{E}[y_t^2 \mid s_t = k, \mathcal{I}_{t-1}] = h_{k,t}$; that is, $h_{k,t}$ is the variance of y_t conditional on the realization of s_t and the information set $\mathcal{I}_{t-1}$.

As per Haas et al. (2004), the conditional variance of y_t is assumed to follow a GARCH-type model. More precisely, conditional on regime $s_t = k$, $h_{k,t}$ is specified as a function of past returns and the additional regime-dependent vector of parameters $\boldsymbol{\theta}_k$:

$$h_{k,t} \equiv h(y_{t-1}, h_{k,t-1}, \boldsymbol{\theta}_k),$$

where $h(\cdot)$ is a $\mathcal{I}_{t-1}$-measurable function, which defines the filter for the conditional variance and also ensures that it is positive. We further assume that $h_{k,1} \equiv \bar{h}_k$ $(k = 1, \ldots, K)$, where $\bar{h}_k$ is a fixed initial variance level for regime k that we set equal to the unconditional variance in regime k. We obtain different scedastic specifications depending on the form of $h(\cdot)$. For instance, if

$$h_{k,t} \equiv \omega_k + \alpha_k y_{t-1}^2 + \beta_k h_{k,t-1},$$

with $\omega_k > 0$, $\alpha_k > 0$, $\beta_k \geq 0$ and $\alpha_k + \beta_k < 1$ $(k = 1, \ldots, K)$, we obtain the Markov-switching GARCH(1, 1) model presented by Haas et al. (2004).[5] In this case, $\boldsymbol{\theta}_k \equiv (\omega_k, \alpha_k, \beta_k)'$.

Alternative definitions of the function $h(\cdot)$ can be incorporated in the model easily. For instance, we account for the well-known asymmetric reaction of volatility to the sign of past returns (often referred to as the *leverage effect*; see Black 1976) by specifying a Markov-switching GJR(1, 1) model that exploits the volatility specification of Glosten et al. (1993):

$$h_{k,t} \equiv \omega_k + (\alpha_k + \gamma_k \mathbb{I}\{y_{t-1} < 0\})\, y_{t-1}^2 + \beta_k h_{k,t-1},$$

where $\mathbb{I}\{\cdot\}$ is an indicator function that is equal to one if the condition holds, and zero otherwise. In this case, the additional parameter $\gamma_k \geq 0$ controls the asymmetry in the conditional variance process. We have $\boldsymbol{\theta}_k \equiv$

2 Our research project was funded by the 2014 SAS/IIF forecasting research grant, for comparing MSGARCH and GARCH models, and for developing and rendering publicly available the computer code for the estimation of MSGARCH models.

3 In practice, this means that we apply the (MS)GARCH models to de-meaned log-returns, as is explained in Section 3.

4 For $t = 1$, we initialize the regime probabilities and conditional variances at their unconditional levels. For simplicity of exposition, we then use the same notation henceforth for $t = 1$ as for a general t, since no confusion is possible.

5 We require the conditional variance in each regime to be covariance-stationary. This is a stronger condition than that of Haas et al. (2004), but allows us to ensure stationarity for various forms of conditional variance and/or conditional distributions.

$(\omega_k, \alpha_k, \gamma_k, \beta_k)'$. Covariance-stationarity of the variance process conditional on the Markovian state is achieved by imposing $\alpha_k + \beta_k + \kappa_k\gamma_k < 1$, where $\kappa_k \equiv \mathbb{P}[y_t < 0 \mid s_t = k, \mathcal{I}_{t-1}]$. For symmetric distributions, we have $\kappa_k = 1/2$. For skewed distributions, κ_k is obtained following the approach of Trottier and Ardia (2016).

We consider several different choices for $\mathcal{D}(\cdot)$. We start with the standard normal ($\mathcal{N}$) and Student-t ($\mathcal{S}$) distributions, then investigate the benefits of incorporating skewness in our analysis by also considering the standardized skewed versions of $\mathcal{N}$ and $\mathcal{S}$ obtained using the mechanism of Bauwens and Laurent (2005) and Fernández and Steel (1998); see Trottier and Ardia (2016) for more details. We denote the standardized skew-Normal and the skew-Student-t by sk$\mathcal{N}$ and sk$\mathcal{S}$, respectively.

Overall, our model set includes 16 different specifications, recovered as combinations of:

- the number of regimes, $K \in \{1, 2\}$: we label our specification single-regime (SR) when $K = 1$ and Markov-switching (MS) when $K = 2$;
- the conditional variance specification: GARCH(1, 1) and GJR(1, 1);
- the choice of the conditional distribution, $\mathcal{D} \in \{\mathcal{N}, \mathcal{S}, \text{sk}\mathcal{N}, \text{sk}\mathcal{S}\}$.[6]

2.2. Estimation

We estimate the models using either frequentist or Bayesian techniques. Both approaches require the evaluation of the likelihood function.

We write the likelihood function that corresponds to the MSGARCH model specification in Eq. (1) by regrouping the model parameters into $\boldsymbol{\Psi} \equiv (\boldsymbol{\xi}_1, \boldsymbol{\theta}_1, \ldots, \boldsymbol{\xi}_K, \boldsymbol{\theta}_K, \mathbf{P})$. The conditional density of y_t in state $s_t = k$ given $\boldsymbol{\Psi}$ and $\mathcal{I}_{t-1}$ is denoted by $f_{\mathcal{D}}(y_t \mid s_t = k, \boldsymbol{\Psi}, \mathcal{I}_{t-1})$.

By integrating out the state variable s_t, we obtain the density of y_t given $\boldsymbol{\Psi}$ and $\mathcal{I}_{t-1}$ only. The (discrete) integration is obtained as

$$f(y_t \mid \boldsymbol{\Psi}, \mathcal{I}_{t-1}) \equiv \sum_{i=1}^{K}\sum_{j=1}^{K} p_{i,j}\,\eta_{i,t-1} \times f_{\mathcal{D}}(y_t \mid s_t = j, \boldsymbol{\Psi}, \mathcal{I}_{t-1}), \quad (2)$$

where $\eta_{i,t-1} \equiv \mathbb{P}[s_{t-1} = i \mid \boldsymbol{\Psi}, \mathcal{I}_{t-1}]$ is the filtered probability of state i at time $t-1$, and where we recall that $p_{i,j}$ denotes the transition probability of moving from state i to state j. The filtered probabilities $\{\eta_{k,t}; k = 1, \ldots, K; t = 1, \ldots, T\}$ are obtained via the Hamilton filter; see Hamilton (1989) and Hamilton (1994, Ch. 22) for details.

Finally, the likelihood function is obtained from Eq. (2) as follows:

$$\mathcal{L}(\boldsymbol{\Psi} \mid \mathcal{I}_T) \equiv \prod_{t=1}^{T} f(y_t \mid \boldsymbol{\Psi}, \mathcal{I}_{t-1}). \quad (3)$$

The ML estimator $\widehat{\boldsymbol{\Psi}}$ is obtained by maximizing the logarithm of Eq. (3). In the case of the Bayesian estimation, the likelihood function is combined with a prior $f(\boldsymbol{\Psi})$ in order to build the kernel of the posterior distribution $f(\boldsymbol{\Psi} \mid \mathcal{I}_T)$. We build our prior from diffuse independent priors as follows:

$$\begin{aligned}
f(\boldsymbol{\Psi}) &\propto f(\boldsymbol{\theta}_1, \boldsymbol{\xi}_1)\cdots f(\boldsymbol{\theta}_K, \boldsymbol{\xi}_K) f(\mathbf{P})\,\mathbb{I}\{\bar{h}_1 < \cdots < \bar{h}_K\}\\
f(\boldsymbol{\theta}_k, \boldsymbol{\xi}_k) &\propto f(\boldsymbol{\theta}_k) f(\boldsymbol{\xi}_k)\,\mathbb{I}\{(\boldsymbol{\theta}_k, \boldsymbol{\xi}_k) \in \mathcal{CSC}_k\} \quad (k = 1, \ldots, K)\\
f(\boldsymbol{\theta}_k) &\propto f_{\mathcal{N}}(\boldsymbol{\theta}_k; \mathbf{0}, 1000 \times \mathbf{I})\,\mathbb{I}\{\boldsymbol{\theta}_k > \mathbf{0}\} \quad (k = 1, \ldots, K)\\
f(\boldsymbol{\xi}_k) &\propto f_{\mathcal{N}}(\boldsymbol{\xi}_k; \mathbf{0}, 1000 \times \mathbf{I})\,\mathbb{I}\{\xi_{k,1} > 0, \xi_{k,2} > 2\}\\
&\quad (k = 1, \ldots, K)\\
f(\mathbf{P}) &\propto \prod_{i=1}^{K}\left(\prod_{j=1}^{K} p_{i,j}\right)\mathbb{I}\{0 < p_{i,i} < 1\},
\end{aligned}$$

where $\mathbf{0}$ and $\mathbf{I}$ respectively denote a vector of zeros and an identity matrix of appropriate sizes, $f_{\mathcal{N}}(\bullet; \boldsymbol{\mu}, \boldsymbol{\Sigma})$ is the multivariate normal density with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$, $\xi_{k,1}$ is the asymmetry parameter, and $\xi_{k,2}$ is the tail parameter of the skewed Student-t distribution in regime k. The prior density for the transition matrix is obtained by assuming that the K rows are independent and follow a Dirichlet prior with all hyperparameters equal to two. Moreover, $\bar{h}_k \equiv \bar{h}_k(\boldsymbol{\theta}_k, \boldsymbol{\xi}_k)$ is the unconditional variance in regime k and $\mathcal{CSC}_k$ denotes the covariance-stationarity condition in regime k; see Trottier and Ardia (2016). As the posterior is of an unknown form (the normalizing constant is numerically intractable), it must be approximated by simulation techniques. In our case, MCMC draws from the posterior are generated using the adaptive random-walk Metropolis sampler of Vihola (2012). We use 50,000 burn-in draws and build a posterior sample of size 1,000 with the next 50,000 draws, keeping only every 50th draw so as to diminish the autocorrelation in the chain.[7] We ensure positivity and stationarity of the conditional variance in each regime during the estimation for both the frequentist and Bayesian estimation. Moreover, we impose constraints on the parameters so as to ensure that the volatilities under the MSGARCH specification cannot be generated by a single-regime specification. In the case of the frequentist estimation, these constraints are enforced in the likelihood optimization by using mapping functions. For the Bayesian estimation, this is achieved through the prior.

[6] We also tested the asymmetric EGARCH scedastic specification (Nelson, 1991), as well as alternative fat-tailed distributions, such as the Laplace and GED distributions. The performance results were qualitatively similar.

[7] We performed several sensitivity analyses to assess the impact of the estimation setup. First, we changed the hyper-parameter values. Second, we ran longer MCMC chains. Third, we used 10,000 posterior draws instead of 1,000. Finally, we tested an alternative MCMC sampler based on adaptive mixtures of Student-t distributions (Ardia, Hoogerheide, & van Dijk, 2009). The conclusions remained qualitatively similar in all cases. Note that we choose a long burn-in sample size in order to rule out the possibility that our results are affected by non-convergent MCMC chains. For simpler applications where it is easier to check the convergence of the MCMC algorithm, a lower value for the burn-in phase can be chosen to speed up the computation. In the MSGARCH package (Ardia, Bluteau, Boudt, Catania, Peterson et al., forthcoming; Ardia, Bluteau, Boudt, Catania & Trottier, 2017), the default value is set to 5,000.

2.3. Density and downside risk forecasting

The generation of one-step-ahead density and downside risk forecasts (VaR and ES) using MSGARCH models is straightforward. First, note that the one-step-ahead conditional probability density function (PDF) of y_{T+1} is a mixture of K regime-dependent distributions:

$$f(y_{T+1} \mid \mathbf{\Psi}, \mathcal{I}_T) \equiv \sum_{k=1}^{K} \pi_{k,T+1} \times f_{\mathcal{D}}(y_{T+1} \mid s_{T+1} = k, \mathbf{\Psi}, \mathcal{I}_T), \quad (4)$$

with mixing weights $\pi_{k,T+1} \equiv \sum_{i=1}^{K} p_{i,k}\eta_{i,T}$, where $\eta_{i,T} \equiv \mathbb{P}[s_T = i \mid \mathbf{\Psi}, \mathcal{I}_T]$ $(i = 1, \ldots, K)$ are the filtered probabilities at time T. The cumulative density function (CDF) is obtained from Eq. (4) as follows:

$$F(y_{T+1} \mid \mathbf{\Psi}, \mathcal{I}_T) \equiv \int_{-\infty}^{y_{T+1}} f(z \mid \mathbf{\Psi}, \mathcal{I}_T)\mathrm{d}z \,. \quad (5)$$

Within the frequentist framework, the predictive PDF and CDF are simply computed by replacing $\mathbf{\Psi}$ with the ML estimator $\widehat{\mathbf{\Psi}}$ in Eqs. (4) and (5). Within the Bayesian framework, we proceed differently, and integrate out the parameter uncertainty. Given a posterior sample $\{\mathbf{\Psi}^{[m]}, m = 1, \ldots, M\}$, the predictive PDF is obtained as:

$$f(y_{T+1} \mid \mathcal{I}_T) \equiv \int_{\mathbf{\Psi}} f(y_{T+1} \mid \mathbf{\Psi}, \mathcal{I}_T) f(\mathbf{\Psi} \mid \mathcal{I}_T)\mathrm{d}\mathbf{\Psi} \approx \frac{1}{M}\sum_{m=1}^{M} f(y_{T+1} \mid \mathbf{\Psi}^{[m]}, \mathcal{I}_T), \quad (6)$$

while the predictive CDF is given by:

$$F(y_{T+1} \mid \mathcal{I}_T) \equiv \int_{-\infty}^{y_{T+1}} f(z \mid \mathcal{I}_T)\mathrm{d}z \,. \quad (7)$$

For both estimation approaches, the VaR is estimated as a quantile of the predictive density by numerically inverting the predictive CDF. For instance, in the Bayesian framework, the VaR at the α risk level is equal to

$$\mathrm{VaR}^{\alpha}_{T+1} \equiv \inf\{y_{T+1} \in \mathbb{R} \mid F(y_{T+1} \mid \mathcal{I}_T) = \alpha\} \,, \quad (8)$$

while the ES at the α risk level is given by

$$\mathrm{ES}^{\alpha}_{T+1} \equiv \frac{1}{\alpha}\int_{-\infty}^{\mathrm{VaR}^{\alpha}_{T+1}} z f(z|\mathcal{I}_T)\mathrm{d}z \,. \quad (9)$$

In our empirical application, we consider the VaR and the ES at the 1% and 5% risk levels.

When evaluating the risk at a horizon of h periods ahead, we must rely on simulation techniques to obtain the conditional density and downside risk measures, as described by Blasques, Koopman, Łasak, and Lucas (2016), for instance. More specifically, given a model parameter, we generate 25,000 paths of daily log-returns over a horizon of h days.[8] The simulated distribution and the α-quantile thus obtained then serve as estimates of the density and downside risk forecasts of the h-day cumulative log-return.

3. Large-scale empirical study

We use 1,500 log-returns (in percent) for the estimation and run the backtest over 2,000 out-of-sample log-returns for the period from October 10, 2008, to November 17, 2016 (the full dataset starts on December 26, 2002). Each model is estimated on a rolling window basis, and both one-step-ahead and multi-step-ahead cumulative log-returns density forecasts are obtained.[9] We then compute the VaR and the ES from the estimated density at the 1% and 5% risk levels.

3.1. Datasets

We test the performances of the various models on several universes of securities that are typically traded by fund managers:

- A set of 426 stocks, selected by taking the S&P 500 universe index as of November 2016 and omitting all stocks for which either more than 5% of the daily returns are zero or there are fewer than 3,500 daily return observations.
- A set of eleven stock market indices: (1) S&P 500 (US; SPX), (2) FTSE 100 (UK; FTSE), (3) CAC 40 (France; FCHI), (4) DAX 30 (Germany; GDAXI), (5) Nikkei 225 (Japan; N225), (6) Hang Seng (China, HSI), (7) Dow Jones Industrial Average (US; DJI), (8) Euro Stoxx 50 (Europe; STOXX50), (9) KOSPI (South Korea; KS11), (10) S&P/TSX Composite (Canada; GSPTSE), and (11) Swiss Market Index (Switzerland; SSMI).
- A set of eight foreign exchange rates: USD against CAD, DKK, NOK, AUD, CHF, GBP, JPY, and EUR.[10]

The data are retrieved from Datastream. Each price series is expressed in the local currency. We compute the daily percentage log-return series, defined by $x_t \equiv 100 \times \log(P_t/P_{t-1})$, where P_t is the adjusted closing price (value) on day t, then de-mean the returns x_t using an AR(1)-filter, and use those filtered returns, y_t, to estimate and evaluate the precision of the financial risk monitoring systems.

Table 1 reports the summary statistics of the out-of-sample daily, five-day, and ten-day cumulative log-returns for the three asset classes. We report the standard deviation (Std) and the skewness (Skew) and kurtosis (Kurt) coefficients evaluated over the full sample, as well as the

8 With the frequentist estimation, we generate 25,000 paths with a parameter of $\widehat{\mathbf{\Psi}}$, while in the case of the Bayesian estimation, we generate 25 paths for each of the 1,000 values of $\mathbf{\Psi}^{[m]}$ in the posterior sample. We use this number in order to get enough draws from the predictive distribution, as we focus on the left tail. Geweke (1989) shows that the consistent estimation of the predictive distribution does not depend on the number of paths generated from the posterior. Thus, we do indeed converge to the correct predictive distribution with 25 paths. We also verified that increasing the number of simulations has no material impact on the results.

9 The model parameters are updated every ten observations. We selected this frequency in order to speed up the computations. Similar results were obtained for a subset of stocks when the parameters were updated every day. This is also in line with the observation of Ardia and Hoogerheide (2014), who show, in the context of GARCH models, that the performance of VaR forecasts is not affected significantly when moving from a daily updating frequency to a weekly or monthly updating frequency. Note that while the parameters are updated every ten observations, the density and downsides risk measures are computed every day.

10 In the context of foreign exchange rates, left-tail forecasts aim to assess the risk for a foreign investor investing in USD, and thus facing a devaluation of USD.

Table 1
Summary statistics of the return data.

h	Percentile	Std	Skew	Kurt	1% VaR	5% VaR	1% ES	5% ES
Panel A: Stocks (426 series)								
	25th	1.48	−0.39	6.89	−6.55	−3.44	−9.30	−5.53
1	50th	1.89	−0.13	9.24	−5.23	−2.85	−7.31	−4.50
	75th	2.33	0.12	14.10	−4.10	−2.25	−5.68	−3.50
	25th	3.29	−0.42	4.93	−14.60	−7.94	−19.14	−12.11
5	50th	4.21	−0.20	5.87	−11.59	−6.55	−14.84	−9.82
	75th	5.19	0.01	7.53	−9.15	−5.17	−12.00	−7.71
	25th	4.54	−0.49	4.47	−19.99	−10.92	−25.42	−16.54
10	50th	5.76	−0.27	5.30	−15.74	−9.02	−20.28	−13.19
	75th	6.98	−0.05	6.92	−12.43	−7.16	−16.08	−10.46
Panel B: Stock market indices (11 series)								
	25th	1.07	−0.40	6.07	−3.70	−2.37	−4.84	−3.30
1	50th	1.15	−0.23	7.29	−3.39	−1.85	−4.31	−2.78
	75th	1.39	−0.17	10.29	−3.05	−1.77	−4.01	−2.58
	25th	2.42	−0.55	5.04	−8.38	−5.09	−10.65	−7.30
5	50th	2.54	−0.47	6.18	−7.60	−4.22	−9.85	−6.17
	75th	3.09	−0.29	8.22	−6.91	−3.86	−9.22	−5.97
	25th	3.29	−0.79	5.47	−12.32	−7.13	−15.96	−10.22
10	50th	3.43	−0.62	6.31	−10.83	−5.70	−13.92	−8.70
	75th	4.19	−0.55	7.04	−9.99	−5.19	−12.90	−8.22
Panel C: Exchange rates (8 series)								
	25th	0.61	−0.53	4.36	−1.73	−1.07	−2.42	−1.60
1	50th	0.62	−0.08	4.51	−1.62	−1.01	−2.10	−1.42
	75th	0.77	0.05	11.60	−1.56	−0.95	−1.92	−1.34
	25th	1.32	−0.36	3.65	−3.72	−2.39	−5.02	−3.36
5	50th	1.39	−0.05	4.05	−3.48	−2.26	−4.33	−3.03
	75th	1.66	0.08	5.91	−3.07	−2.06	−3.82	−2.77
	25th	1.85	−0.31	3.36	−5.00	−3.43	−6.99	−4.55
10	50th	1.93	−0.10	3.52	−4.78	−3.04	−5.72	−4.06
	75th	2.29	0.13	5.12	−4.64	−2.93	−5.41	−3.94

Notes: This table presents the summary statistics of the (de-meaned) h-day cumulative log-returns for securities in the three asset classes used in our study. We report the standard deviation (Std), skewness (Skew), kurtosis (Kurt), and 1% and 5% historical VaR and ES, on an unconditional basis for the 2,000 out-of-sample observations. For each statistic, we compute the 25th, 50th and 75th percentiles over the whole universe of assets.

historical 1% and 5% VaR and ES levels. We note the higher volatility in all periods for the universe of stocks, followed by indices and exchange rates. All of the securities exhibit negative skewness, with larger values for indices and stocks, while the exchange rates seem to behave more symmetrically. Interestingly, the negative skewness tends to become more pronounced for indices as the horizon grows. Finally, we observe a significant kurtosis for stocks at the daily horizon. Fat tails are also present for indices and exchange rates, but they are less pronounced than for stocks. However, the kurtosis of all asset classes tends to diminish as the horizon grows.

3.2. Forecasting performance tests

We compare the adequacy of each of the 32 models in terms of providing accurate forecasts of the left tail of the conditional distribution and the VaR and ES levels.

3.2.1. Accuracy of VaR predictions

We test the accuracy of the VaR predictions using the so-called *hit* variable, which is a dummy variable that indicates a loss that exceeds the VaR level:

$$I_t^\alpha \equiv \mathbb{I}\{y_t \leq \mathrm{VaR}_t^\alpha\},$$

where VaR_t^α denotes the VaR prediction at a risk level of α for time t, and $\mathbb{I}\{\cdot\}$ is the indicator function, which is equal to one if the condition holds and zero otherwise. If the VaR is specified correctly, then the hit variable has a mean value of α and is distributed independently over time. We test this for the $\alpha = 1\%$ and $\alpha = 5\%$ risk levels using the unconditional coverage (UC) test of Kupiec (1995) and the dynamic quantile (DQ) test of Engle and Manganelli (2004).

The UC test of Kupiec (1995) uses the likelihood ratio statistic to test that the violations have a binomial distribution, with $\mathbb{E}[I_t^\alpha] = \alpha$. Denote the number of exceedances observed on a total of T observations by $x \equiv \sum_{t=1}^T I_t^\alpha$; then, under the null of correct coverage, we have that the test statistic

$$\mathrm{UC}_\alpha \equiv -2\ln\left[(1-\alpha)^{T-x}\alpha^x\right] + 2\ln\left[\left(1-\frac{x}{T}\right)^{T-x}\left(\frac{x}{T}\right)^x\right]$$

is asymptotically chi-square distributed with one degree of freedom.

The DQ test of Engle and Manganelli (2004) is a test of the joint hypothesis that $\mathbb{E}[I_t^\alpha] = \alpha$ and that the hit variables are distributed independently. The implementation of the test involves the de-meaned process $\mathrm{Hit}_t^\alpha \equiv I_t^\alpha - \alpha$. Under a correct model specification, unconditionally and conditionally, Hit_t^α has a zero mean and is serially uncorrelated. The DQ test is then the traditional Wald test of the joint nullity of all coefficients in the following linear regression:

$$\text{Hit}_t^\alpha = \delta_0 + \sum_{l=1}^{L} \delta_l \text{Hit}_{t-l}^\alpha + \delta_{L+1}\text{VaR}_{t-1}^\alpha + \epsilon_t.$$

If we denote the OLS parameter estimates by $\widehat{\delta} \equiv (\widehat{\delta}_0, \ldots, \widehat{\delta}_{L+1})'$ and the corresponding data matrix with the observations for the $L+2$ explanatory variables in the columns by $\mathbf{Z}$, then the DQ test statistic of the null hypothesis of correct unconditional and conditional coverage is

$$\text{DQ}_\alpha \equiv \frac{\widehat{\delta}'\mathbf{Z}'\mathbf{Z}\widehat{\delta}}{\alpha(1-\alpha)}.$$

Like Engle and Manganelli (2004), we choose $L = 4$ lags. Under the null hypothesis of correct unconditional and conditional coverages, we have that DQ_α is asymptotically chi-square distributed with $L+2$ degrees of freedom.[11]

3.2.2. Accuracy of the left-tail distribution

Risk managers care not only about the accuracy of the VaR forecasts, but also about the accuracy of the complete left-tail region of the log-return distribution. This broader view of all losses is central to modern risk management and consistent with the regulatory shift to using the expected shortfall as the risk measure for determining capital requirements starting in 2018 (Basel Committee on Banking Supervision, 2013). We evaluate the effectiveness of MSGARCH models for yielding accurate predictions of the left-tail distribution in three ways.

Our first approach is to compute the weighted average difference of the observed returns with respect to the VaR value, giving higher weights to losses that violate the VaR level. This corresponds to the quantile loss assessment of González-Rivera, Lee, and Mishra (2004) and McAleer and Da Veiga (2008). Formally, given a VaR prediction at risk level α for time t, the associated quantile loss (QL) is defined as

$$\text{QL}_t^\alpha \equiv (\alpha - I_t^\alpha)(y_t - \text{VaR}_t^\alpha).$$

The choice of this loss function for VaR assessment is appropriate because quantiles are elicited by it; that is, when the conditional distribution is static over the sample, the VaR_t^α can be estimated by minimizing the average quantile loss function. Elicitability is useful for model selection, estimation, forecast comparison, and forecast ranking.

Unfortunately, there is no loss function available for which the ES risk measure can be elicited; see for instance Bellini & Bignozzi (2015) and Ziegel (2016). However, Fissler and Ziegel (2016) recently showed that, in the case of a constant conditional distribution, the pair (VaR, ES) can be elicited jointly as the values of v_t and e_t that minimize the sample average of the following loss function:

$$\begin{aligned} &FZ(y_t, v_t, e_t, \alpha, G_1, G_2) \\ &\equiv (I_t^\alpha - \alpha)\left(G_1(v_t) - G_1(y_t) + \frac{1}{\alpha}G_2(e_t)v_t\right) \\ &\quad - G_2(e_t)\left(\frac{1}{\alpha}I_t^\alpha y_t - e_t\right) - \mathcal{G}_2(e_t), \end{aligned}$$

where G_1 is weakly increasing, G_2 is strictly positive and strictly increasing, and $\mathcal{G}_2' = G_2$. In a setup similar to ours, Patton, Ziegel, and Chen (2017) assume the values of VaR and ES to be strictly negative and recommend setting $G_1(x) = 0$ and $G_2(x) = -1/x$. For VaR and ES predictions at a risk level of α for time t, the associated joint loss function (FZL) is then given by:

$$\begin{aligned} \text{FZL}_t^\alpha \equiv\ & \frac{1}{\alpha\, \text{ES}_t^\alpha} I_t^\alpha (y_t - \text{VaR}_t^\alpha) \\ & + \frac{\text{VaR}_t^\alpha}{\text{ES}_t^\alpha} + \log(-\text{ES}_t^\alpha) - 1\,, \end{aligned} \tag{10}$$

for $\text{ES}_t^\alpha \leq \text{VaR}_t^\alpha < 0$. Hence, we use the FZL function as our second evaluation criterion in order to gauge the precision of the VaR and ES downside risk estimates.

The third approach that we consider is to compare the empirical distribution with the predicted conditional distribution through the weighed continuous ranked probability score (wCRPS), introduced by Gneiting and Ranjan (2011) as a generalization of the CRPS scoring rule (Matheson & Winkler, 1976). Following the notation introduced in Section 2, the wCRPS for a forecast at time t is defined as:

$$\text{wCRPS}_t \equiv \int_{\mathbb{R}} \omega(z)(F(z \mid \mathcal{I}_{t-1}) - \mathbb{I}\{y_t \leq z\})^2 \text{d}z,$$

where F is the predictive CDF and $\omega : \mathbb{R} \rightarrow \mathbb{R}^+$ is a continuous weight function, which emphasizes regions of interest in the predictive distribution, such as the tails or the center. Since our focus is on predicting losses, we follow Gneiting and Ranjan (2011) and use the decreasing weight function $\omega(z) \equiv 1 - \Phi(z)$, where Φ is the CDF of a standard Gaussian distribution. This way, discrepancies in the left tail of the return distribution are weighted more heavily than those in the right tail.[12]

For the QL, FZL and wCRPS approaches, we test the statistical significance of the differences in the forecasting performances of two competing models, say models i and

[11] As per Bams et al. (2017), it is possible to add more explanatory variables, such as lagged returns and lagged squared returns, and test the new coefficients jointly. In our case, the results obtained by adding lagged returns or lagged squared returns are qualitatively similar to the simpler specification.

[12] We follow the implementation of Gneiting and Ranjan (2011) and compute wCRPS with the following approximation:

$$\text{wCRPS}_t \approx \frac{z_u - z_l}{M-1} \sum_{m=1}^{M} w(z_m)(F(z_m \mid \mathcal{I}_{t-1}) - \mathbb{I}\{y_t \leq z_m\})^2,$$

where $z_m \equiv z_l + m \times (z_u - z_l)/M$ and z_u and z_l are the upper and lower values, which defines the range of integration. The accuracy of the approximation can be increased to any desired level by M. Setting $z_l = -100$, $z_u = 100$ and $M = 1000$ provides an accurate approximation when working with returns in percentage points. We also tested the triangular integration approach and obtained numerically equivalent results. Alternative specifications of the weights, focusing on the right tail or center of the full distribution, lead to similar conclusions at the one-day forecasting horizon. The results are available from the authors upon request.

j. We do this by first computing the average performance statistics across all securities in the same asset class for each out-of-sample date t. Denote this difference by $\Delta_t^{i-j} \equiv L_t^i - L_t^j$, where L_t^i is the average value of the performance measure (QL, FZL or wCRPS) of all assets within the same asset class. We then test $H_0 : \mathbb{E}[\Delta_t^{i-j}] = 0$ using the standard Diebold and Mariano (1995) test statistic, implemented with the heteroscedasticity and autocorrelation robust (HAC) standard error estimators of Andrews (1991) and Andrews and Monahan (1992). If the null hypothesis is rejected, the sign of the test statistics indicates which model is preferred, on average, for a particular loss measure.

3.3. Results

We now summarize the results regarding our main research question: *Does the additional complexity of Markov-switching and the use of Bayesian estimation methods lead to more accurate out-of-sample downside risk predictions?* We begin by presenting our results regarding the accuracy of the VaR predictions, then use the QL, FZL and wCRPS approaches to evaluate the gains in terms of left-tail predictions.

3.3.1. Effect of model and estimator choice on the accuracy of VaR predictions

We first use the UC test of Kupiec (1995) and the DQ test of Engle and Manganelli (2004) to evaluate the accuracy of each of the 32 methods considered in terms of predicting the VaR at the 5% and 1% levels for the daily returns on the 426 stocks, 11 stock indices and eight exchange rates. For each asset, we obtain the p-values that correspond to the UC and DQ tests computed using 2,000 out-of-sample observations. Table 2 aggregates the results for each asset class by presenting the percentages of assets for which the null hypothesis of correct unconditional and conditional coverage is rejected at the 5% level by the UC and DQ tests, respectively.[13]

Consider the results of the UC test in Panels A and B of Table 2. We find that the validity of the VaR predictions based on the GARCH and GJR skewed Student-t risk models is never rejected at either VaR risk level, regardless of whether SR or MS models and frequentist or Bayesian estimation methods are used. However, this result changes drastically when we consider the more powerful DQ test of correct conditional coverage in Panels C and D. Here, we find clear evidence that the use of MS GJR models leads to lower percentages of rejections of the validity of the VaR prediction for all asset classes. At the 1% risk level, these differences are mostly significant.

Overall, the one-day-ahead backtest results indicate that the MS models outperform the SR models, especially for VaR prediction on equities. Moreover, a GJR specification leads to a substantial reduction in the rejection frequencies. A fat-tailed conditional distribution is of primary importance for both the MS and SR specifications, and delivers excellent results at both risk levels.

Finally, the frequencies of rejections for this analysis are similar between the Bayesian and frequentist estimation methods. More precisely, a t-test for equal average rejections indicates that the differences are insignificant. Thus, we conclude that it is hard to discriminate between the estimation methods based on an analysis of the VaR forecast accuracy.

3.3.2. Effect of model choice on accuracy of left-tail predictions

A further question relates to the way in which model simplification affects the accuracy of the left-tail return prediction. Table 3 reports the standardized differences between the average QL, FZL and wCRPS values of the assets from a given asset class when we switch from a MS specification to a SR specification. The standardization corresponds to the Diebold and Mariano (1995) (DM) test statistic. Negative values indicate out-of-sample evidence of a deterioration in the prediction accuracy when using the SR specification instead of the MS specification. When the standardized value exceeds 2.57 (i.e., the critical value computed using a 1% significance level for a bilateral test based on the asymptotic normal distribution) in absolute value, the statistical significance is highlighted with gray shading.[14] We report only results obtained with the Bayesian framework, as the performances obtained using the Bayesian estimation are either similar to or better than those obtained using the frequentist estimation for both MS and SR models.[15]

The one-step-ahead results for wCRPS favor MS models, with negative values being observed for almost all asset classes and model specifications. The QL, FZL and wCRPS results are consistent with the backtest results, confirming the superior performance of the MS specification for the universe of stocks, while the outperformance is less clear for indices and exchange rates. Indeed, MS is required for indices only when a non-fat-tailed conditional distribution is assumed, while MS is generally not required for exchange rates. Note that the improvements tend to be more pronounced for all assets when the Markov-switching mechanism is applied to simple specifications such as the GARCH-Normal model.

For stocks, the MS specification performs significantly better in terms of the FZL and wCRPS measures at the five-day horizon. The results for the wCRPS measure at the

[13] In the case of stocks, because the universe is large and therefore prone to false positives, the p-values are corrected for Type I error using the false discovery rate (FDR) approach of Benjamini and Hochberg (1995). The FDR correction for a confidence level of q proceeds as follows. For a set of m ordered p-values $p_1 \leq p_2 \leq \cdots \leq p_m$ and corresponding null hypotheses $H_1, H_2, \ldots, H_m$, define υ as the largest value of i for which $p_i \leq \frac{i}{m}q$, then reject all hypotheses H_i for $i = 1, \ldots, \upsilon$.

[14] We take the standard critical value of Diebold and Mariano (1995), as our Markov-switching specifications do not nest the alternative single-regime model due to parameter constraints that require the volatility dynamics to be different numerically in each regime, and each regime to have a non-zero probability. The approach of Clark and McCracken (2001) should be used when comparing nested models.

[15] We find in Section 3.3.3 that the gains from Bayesian estimation relative to frequentist estimation are larger in the case of SR models. Thus, our discussion regarding the gains of MS versus SR models based on the Bayesian estimation results is conservative, in the sense that it gives the SR specifications an advantage.

Table 2
Percentage of assets for which the validity of the VaR predictions is rejected.

	Stocks				Stock market indices				Exchange rates			
	Bayesian		Frequentist		Bayesian		Frequentist		Bayesian		Frequentist	
Model	MS	SR	MS	SR	MS	SR	MS	SR	MS	SR	MS	SR
Panel A: UC 1% VaR												
GARCH $\mathcal{N}$	0.00	26.76	0.23	29.34	72.73	90.91	72.73	90.91	25.00	25.00	25.00	25.00
GARCH sk$\mathcal{N}$	0.00	8.92	0.23	9.62	9.09	63.64	0.00	63.64	0.00	12.50	0.00	12.50
GARCH $\mathcal{S}$	0.00	0.00	0.00	0.00	54.55	45.45	27.27	27.27	25.00	25.00	25.00	12.50
GARCH sk$\mathcal{S}$	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
GJR $\mathcal{N}$	0.00	16.43	0.00	19.48	54.55	90.91	63.64	90.91	25.00	25.00	25.00	37.50
GJR sk$\mathcal{N}$	0.00	3.52	0.00	5.16	0.00	54.55	0.00	45.45	0.00	12.50	0.00	25.00
GJR $\mathcal{S}$	0.00	0.00	0.00	0.00	18.18	36.36	18.18	36.36	12.50	12.50	12.50	12.50
GJR sk$\mathcal{S}$	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Panel B: UC 5% VaR												
GARCH $\mathcal{N}$	0.70	39.20	0.70	38.73	36.36	36.36	27.27	36.36	25.00	50.00	25.00	50.00
GARCH sk$\mathcal{N}$	0.00	41.31	0.00	40.38	0.00	0.00	0.00	0.00	12.50	25.00	0.00	25.00
GARCH $\mathcal{S}$	0.94	1.17	0.70	0.70	54.55	54.55	36.36	54.55	25.00	12.50	25.00	12.50
GARCH sk$\mathcal{S}$	0.23	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
GJR $\mathcal{N}$	0.47	38.73	0.47	36.15	18.18	18.18	36.36	27.27	25.00	37.50	25.00	37.50
GJR sk$\mathcal{N}$	0.00	40.38	0.00	39.91	0.00	0.00	0.00	0.00	12.50	12.50	0.00	12.50
GJR $\mathcal{S}$	1.64	1.64	0.70	0.47	18.18	27.27	18.18	27.27	37.50	37.50	37.50	37.50
GJR sk$\mathcal{S}$	0.00	0.00	0.00	0.00	0.00	18.18	0.00	18.18	0.00	0.00	0.00	0.00
Panel C: DQ 1% VaR												
GARCH $\mathcal{N}$	14.08	53.52	14.32	54.69	63.64	90.91	72.73	90.91	25.00	37.50	12.50	37.50
GARCH sk$\mathcal{N}$	14.08	48.36	15.49	50.00	45.45	63.64	45.45	63.64	12.50	37.50	12.50	37.50
GARCH $\mathcal{S}$	19.95	28.64	16.90	29.34	54.55	63.64	63.64	54.55	25.00	25.00	25.00	25.00
GARCH sk$\mathcal{S}$	18.31	23.94	17.37	24.18	45.45	45.45	36.36	36.36	12.50	25.00	12.50	25.00
GJR $\mathcal{N}$	5.87	32.39	6.10	34.74	18.18	90.91	36.36	90.91	12.50	37.50	12.50	37.50
GJR sk$\mathcal{N}$	5.87	27.00	6.10	28.17	9.09	27.27	9.09	45.45	12.50	25.00	0.00	25.00
GJR $\mathcal{S}$	7.04	10.33	4.46	9.86	18.18	27.27	18.18	18.18	12.50	25.00	12.50	25.00
GJR sk$\mathcal{S}$	5.16	10.33	6.57	11.27	0.00	0.00	0.00	0.00	12.50	12.50	12.50	12.50
Panel D: DQ 5% VaR												
GARCH $\mathcal{N}$	3.52	26.29	3.52	25.82	18.18	9.09	36.36	9.09	0.00	0.00	0.00	0.00
GARCH sk$\mathcal{N}$	3.52	29.81	2.82	30.05	9.09	9.09	9.09	9.09	0.00	0.00	0.00	0.00
GARCH $\mathcal{S}$	1.64	7.75	1.64	8.92	45.45	54.55	36.36	54.55	0.00	0.00	0.00	0.00
GARCH sk$\mathcal{S}$	2.11	6.57	2.82	7.98	9.09	9.09	9.09	9.09	0.00	0.00	0.00	0.00
GJR $\mathcal{N}$	0.00	14.32	0.00	14.55	9.09	9.09	9.09	0.00	0.00	0.00	0.00	0.00
GJR sk$\mathcal{N}$	0.00	15.02	0.00	13.62	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
GJR $\mathcal{S}$	0.00	0.00	0.00	1.17	9.09	0.00	9.09	9.09	12.50	12.50	12.50	12.50
GJR sk$\mathcal{S}$	0.00	0.70	0.00	0.70	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Notes: This table presents the percentages of assets for which the unconditional coverage test (UC, Panels A and B) of Kupiec (1995) and the dynamic quantile test (DQ, Panels C and D) of Engle and Manganelli (2004) reject the null hypothesis of correct unconditional coverage (UC, DQ) and independence of violations (DQ) for the one-step-ahead 1% VaR (Panels A and C) and 5% VaR (Panels B and D) at the 5% significance level. The VaR forecasts are obtained for both Markov-switching (MS) and single-regime (SR) models for the various universes (426 stocks, 11 indices, and 8 exchange rates), and are estimated via Bayesian or frequentist techniques. We highlight in gray the best performing method in all the cases in which, for a given asset class and model specification, the percentages of rejections are significantly different between the MS and SR models at the 5% level. In the case of stocks, rejection frequencies are corrected for Type I error using the FDR approach of Benjamini and Hochberg (1995).

ten-day horizon, and for the QL measure at the five- and ten-day horizons, are mostly insignificant, except for the FZL 5% measure, which favors MS models when a non-fat-tailed conditional distribution is assumed. The MS and SR models perform similarly for the five- and ten-day returns on stock indices. Finally, for exchange rate returns, SR models outperform MS models at the five- and ten-day horizons according to the QL 1% measure, while the differences in QL 5%, FZL, and wCRPS are insignificant.

It is informative to examine whether these gains in forecasting precision are stable across the out-of-sample window. We determine this by displaying in Fig. 1 the cumulative average loss differentials over the whole out-of-sample period for the best performing specification, the GJR skewed Student-t model. Interestingly, we find that MS systematically outperforms SR according to the criteria that are most sensitive to the extreme left tail of the return distribution, namely the FZL (for $\alpha = 1\%$ and $\alpha = 5\%$) and QL (for $\alpha = 1\%$). We also notice that the gains of MS over SR in these cases increase during the last phase of the turbulent period 2008–2012. With regard to wCRPS and QL at $\alpha = 5\%$, we find that MS starts to outperform SR after the end of the turbulent period 2008–2012. We conjecture that this improvement in performance can be explained by the lack of flexibility of the single-regime GARCH specification. As is also evident from the first panel of Fig. 1, the market volatility has changed both its unconditional level and its dependence structure between the two periods 2008–2012 and 2012–2015. Since the estimation window consists of 1,500 observations (approximately seven years), the observations in the period 2008–2012 affect the SR predictions for the entire 2012–2015 forecasting period.

Table 3
Standardized gain in average performance when switching from MS to SR models.

		Stocks					Stock market indices					Exchange rates				
Horizon	Model	QL 1%	QL 5%	FZL 1%	FZL 5%	wCRPS	QL 1%	QL 5%	FZL 1%	FZL 5%	wCRPS	QL 1%	QL 5%	FZL 1%	FZL 5%	wCRPS
$h=1$	GARCH $\mathcal{N}$	−0.60	−4.94	−3.74	−8.54	−9.32	−3.84	0.50	−5.11	−1.59	−4.04	−0.09	0.39	−0.90	0.41	−2.65
	GARCH sk$\mathcal{N}$	−0.25	−4.90	−3.43	−8.39	−9.25	−2.64	0.10	−3.32	−1.18	−3.26	0.95	−0.25	−0.77	0.37	−3.41
	GARCH $\mathcal{S}$	−4.00	−3.55	−3.82	−4.42	−3.41	−1.50	−0.90	−1.70	−1.37	−0.17	1.12	−1.26	1.17	1.08	−2.17
	GARCH sk$\mathcal{S}$	−4.52	−4.20	−3.86	−4.86	−2.79	−2.21	−0.87	−1.69	−0.86	0.22	2.18	−0.56	1.63	1.26	−1.45
	GJR $\mathcal{N}$	−0.63	−6.02	−4.07	−10.74	−9.96	−3.58	0.53	−4.99	−2.3	−4.30	0.64	0.58	−0.98	−0.02	−1.64
	GJR sk$\mathcal{N}$	−0.22	−5.95	−3.76	−10.32	−9.94	−2.04	−0.31	−3.06	−1.40	−3.00	0.79	0.07	−0.87	0.11	−1.88
	GJR $\mathcal{S}$	−3.88	−4.44	−3.23	−4.29	−5.00	−1.80	0.49	−2.26	−1.43	0.11	0.36	−1.03	−0.20	−0.19	−2.35
	GJR sk$\mathcal{S}$	−3.64	−4.17	−3.16	−4.10	−3.44	−0.93	0.61	−1.20	−0.55	0.47	0.92	−0.68	1.32	0.65	−1.66
$h=5$	GARCH $\mathcal{N}$	−0.66	−1.70	−2.16	−7.24	−2.72	−2.11	−1.47	−4.97	−2.48	−1.19	2.69	1.69	−0.52	0.95	0.73
	GARCH sk$\mathcal{N}$	−0.52	−1.67	−1.93	−7.01	−2.65	−2.14	−1.61	−2.69	−1.40	−0.88	2.59	1.17	−0.70	0.43	0.46
	GARCH $\mathcal{S}$	−1.78	−2.27	−2.86	−3.22	−2.68	−0.77	−1.61	−1.52	−2.18	−0.89	2.00	1.53	1.26	1.55	−1.26
	GARCH sk$\mathcal{S}$	−1.70	−2.37	−2.68	−3.23	−2.39	−1.55	−2.72	−0.44	−1.32	−0.32	3.67	0.85	1.02	1.50	−0.93
	GJR $\mathcal{N}$	−0.53	−2.38	−2.24	−8.29	−2.77	0.10	−0.10	−4.91	−3.29	−0.51	3.27	1.15	−1.02	−0.23	0.30
	GJR sk$\mathcal{N}$	−0.30	−2.37	−1.98	−8.37	−2.74	0.62	−1.10	−2.76	−2.44	−0.30	3.70	1.54	−1.09	−0.40	1.15
	GJR $\mathcal{S}$	−1.32	−2.63	−2.46	−2.74	−4.52	−0.21	−1.76	−2.08	−2.85	−0.49	4.08	0.04	0.41	0.75	−2.07
	GJR sk$\mathcal{S}$	−1.14	−1.61	−2.26	−2.75	−3.37	0.26	−0.65	−1.06	−1.40	−0.04	4.60	1.05	1.46	1.11	−1.09
$h=10$	GARCH $\mathcal{N}$	−0.18	−0.82	−1.59	−6.36	−1.93	−1.23	−0.66	−3.96	−2.07	−1.91	1.55	1.18	−0.33	0.99	0.89
	GARCH sk$\mathcal{N}$	−0.14	−0.71	−1.42	−6.32	−1.96	−1.76	−0.80	−2.69	−2.12	−1.35	1.23	1.58	−0.72	0.25	0.86
	GARCH $\mathcal{S}$	−0.83	−1.01	−1.77	−1.90	−1.25	−1.05	−0.79	−1.69	−2.01	−1.95	1.12	1.63	1.24	1.55	−0.59
	GARCH sk$\mathcal{S}$	−0.93	−1.22	−1.53	−2.04	−1.02	−1.19	−0.99	−1.52	−1.60	−1.22	2.78	2.12	1.30	1.54	−0.46
	GJR $\mathcal{N}$	−0.15	−1.24	−1.64	−6.73	−1.91	0.48	−1.08	−3.87	−2.81	−0.92	1.16	1.11	−0.99	−0.41	0.62
	GJR sk$\mathcal{N}$	0.02	−1.22	−1.32	−6.76	−1.71	0.21	−1.44	−1.79	−1.79	−0.87	2.47	1.15	−0.97	−0.11	1.03
	GJR $\mathcal{S}$	−0.53	−1.46	−1.47	−2.10	−4.04	1.16	−1.43	−1.47	−3.32	−1.13	2.92	0.34	0.71	0.89	−1.71
	GJR sk$\mathcal{S}$	−0.48	−1.15	−1.56	−2.44	−2.80	0.72	−2.19	−1.44	−2.17	−1.31	4.55	1.12	1.79	1.17	−1.09

Notes: This table presents the Diebold and Mariano (1995) test statistic for equal average loss between the MS and SR models for forecasting the distribution of h-day cumulative log-returns ($h \in \{1, 5, 10\}$). As loss functions, we consider the QL and FZL measures (at $\alpha = 1\%$ and $\alpha = 5\%$), and the wCRPS measure. Negative values indicate the Markov-switching specification outperforms the single-regime models. We also report statistics which are significantly negative (positive) at the 1% level in light (dark) gray (bilateral test). The multi-step cumulative log-return forecasts are generated using 25,000 simulated paths of daily log-returns. Models are estimated using the Bayesian approach.

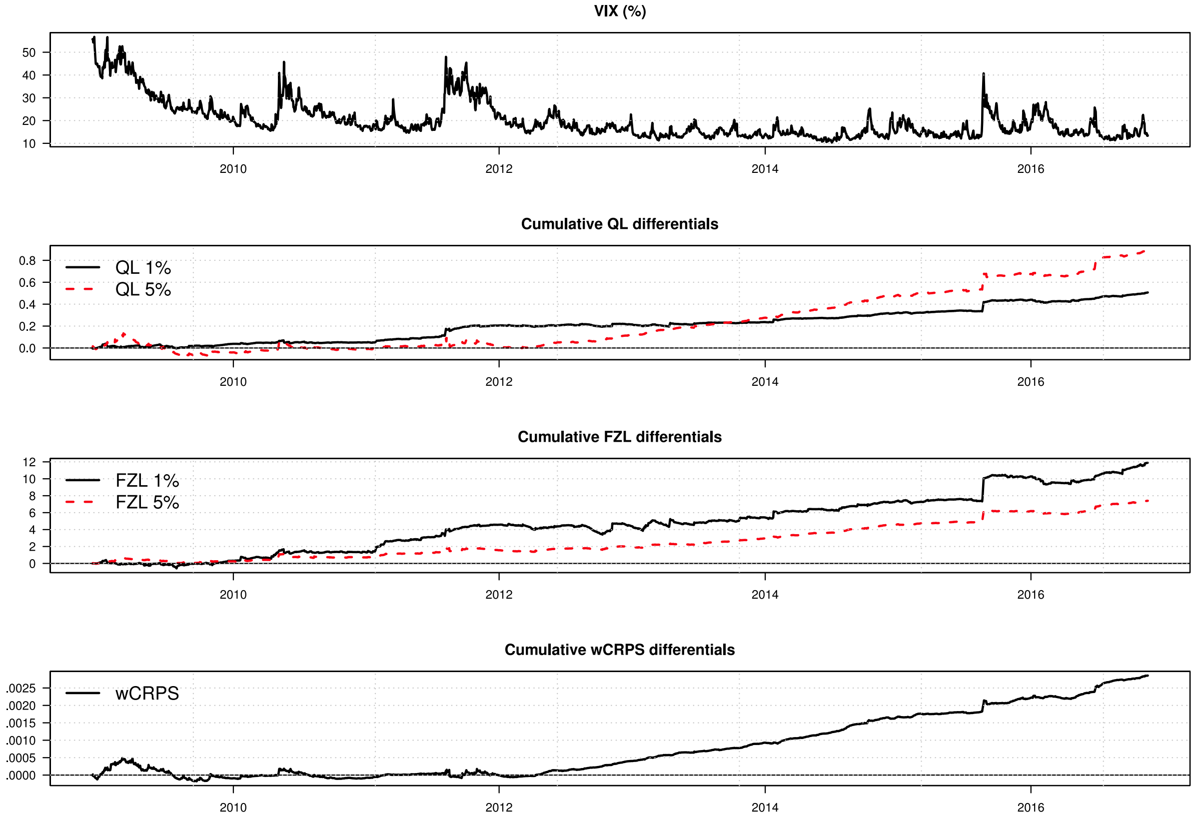

Fig. 1. Cumulative performance. Notes: The top panel presents the evolution of VIX (the Chicago Board of Exchange's volatility index), while the lower panels show the cumulative average loss differentials (QL, FZL and wCRPS) for the 2,000 out-of-sample observations (ranging from December 2008 to November 2016). The comparison is done between the Markov-switching and single-regime GJR skewed Student-t models. A positive value indicates the Markov-switching specification outperforms the alternative. A positive slope indicates outperformance at the corresponding date.

On the other hand, MSGARCH allows the volatility process to adapt more rapidly to changes in regimes, resulting in better risk predictions. This is the case for both the first half of the window, from December 2008 to November 2012, which encompasses the Great Financial Crisis, and the second half of the window, from December 2012 to November 2016, which is a calmer market period.

Table 4 now considers a complete comparison of the wCRPS performances of all MS (in the rows) and SR (in the columns) models. The elements on the diagonal correspond to the wCRPS values reported in Table 3, and are informative about the change in wCRPS when switching from a MS model to a SR model, keeping the same specifications for the conditional variance and the distribution. The analysis of the extra-diagonal elements is informative about the changes in wCRPS when switching from a MS model to a SR model and changing the specification of the volatility model or the density function. In this table, an outperforming MS risk model is a model for which all standardized gains when changing the specification are negative. This is the case for the MS GJR model with skewed Student-t innovations for almost all comparisons. The only exception is for modeling the returns of stock market indices, where it performs similarly to its SR counterpart.

Relative to SR models, MS specifications offer the flexibility of different volatility responses to extreme (positive or negative) and moderately large observations. This feature is desirable in the case where the discretely-observed returns are generated by an underlying continuous-time process with jumps. Such jumps usually correspond to one-off events (as per Boudt, Daníelsson, & Laurent, 2013), and have a less persistent effect on the future volatility (see Andersen, Bollerslev, & Diebold, 2007). As was explained by Laurent, Lecourt, and Palm (2016), this effect can be captured in a SR framework through the use of generalized autoregressive score (GAS) models (also referred to as dynamic conditional score (DCS) models) with fat tails, as was introduced by Creal, Koopman, and Lucas (2013) and Harvey (2013). Clearly, a SR GAS model is simpler computationally than a MS model. Thus, it is relevant to benchmark the MSGARCH model against a GAS model whose specification fits with the assumed density function, which is the Beta-Skew-t-EGARCH(1,1) model introduced by Harvey and Sucarrat (2014) in the case of the skewed Student-t density function. In the case of a fat-tailed conditional distribution, this model yields volatility forecasts that are robust to outliers in the return series (due to one-off events causing price jumps, for instance), and may therefore deliver better downside risk predictions than GARCH; see for instance Ardia, Boudt, and Catania (2018) and Bernardi and Catania (2016).

Table 5 reports the standardized gain in average QL, FZL (at 1% and 5% risk levels) and wCRPS performances when switching from the most flexible MSGARCH model, that is, the MS GJR sk$\mathcal{S}$, to the Beta-Skew-t-EGARCH(1,1) model. For downside risk prediction related to the returns on stocks and stock market indices, the results significantly favor the MS specification when focusing on the 1% largest losses, as can be seen in the QL 1% and FZL 1% columns. The other performance measures are negative, indicating an outperformance of the MS specification, albeit not a statistically significant one. For exchange rates, the results favor the GAS model but are not significant. Thus, we find overall that the MS specification can offer added value in the downside risk prediction of equity investments, relative to the use of GAS models.

3.3.3. *Effect of the estimator choice on the accuracy of left-tail predictions*

Table 6 reports the results for the Bayesian versus frequentist estimation methods in the case of the one-step-ahead QL, FZL and wCRPS measures. Panel A (Panel B) shows the results for MS (SR) models, with a negative (positive) value indicating that the Bayesian estimation outperforms (underperforms) the frequentist estimation. We emphasize cases where the Bayesian estimation outperforms the frequentist approach significantly in light gray. For stocks, the QL 1% and 5% comparisons indicate that Bayesian is preferred over ML, with the difference being significant in the majority of the specifications. The same observation can be made when using the FZL and wCRPS evaluation criteria. For stock indices and exchange rates, the QL, FZL and wCRPS results are in favor of the Bayesian estimation for both the MS and SR models, but the results are less significant than for stocks. Overall, we recommend accounting for parameter uncertainty, especially for stocks data and when the interest focuses on the left tail of the log-returns distribution. The performance gains are especially large for SR models.

3.3.4. *Constrained Markov-switching specifications*

Thus far, our empirical results have highlighted the need for a MS mechanism in GARCH-type models in the case of stocks. We now refine the analysis by examining whether the same gains are achieved when constraining the conditional distribution of the MS specifications to have the same shape parameter across regimes. Hence, we apply the MS mechanism only to the conditional variance. The objective is to determine whether the switches in the variance dynamics are the dominant contributor to the gains in risk forecasting accuracy in the context of MS models.

Table 7 reports the performance measures obtained using the constrained MS models for the various horizons, when the models are estimated via the Bayesian approach.[16] The results are in line with the non-constrained case of Table 3, but less significant. Hence, accounting for structural breaks in the variance dynamics alone improves the risk forecasts at the daily, weekly and ten-day horizons. We improve the performance further if we let the shape parameters depend upon the regime.

[16] The forecasting results obtained via frequentist estimation are qualitatively similar, and are available from the authors upon request.

Table 4
Standardized gain in average performance when switching from MS to SR and changing the specification.

		SR GARCH				SR GJR			
		$\mathcal{N}$	sk$\mathcal{N}$	$\mathcal{S}$	sk$\mathcal{S}$	$\mathcal{N}$	sk$\mathcal{N}$	$\mathcal{S}$	sk$\mathcal{S}$
Panel A: Stocks									
MS GARCH	$\mathcal{N}$	−9.32	−9.56	3.29	3.30	−6.80	−6.85	3.29	3.38
	sk$\mathcal{N}$	−9.00	−9.25	3.60	3.67	−6.60	−6.65	3.42	3.54
	$\mathcal{S}$	−9.01	−9.20	−3.41	−2.99	−7.29	−7.36	−0.14	−0.13
	sk$\mathcal{S}$	−8.86	−9.07	−2.92	−2.79	−7.15	−7.22	0.01	0.04
MS GJR	$\mathcal{N}$	−10.11	−10.26	0.88	0.93	−9.96	−10.25	3.20	3.18
	sk$\mathcal{N}$	−9.88	−10.06	0.88	0.95	−9.64	−9.94	3.33	3.38
	$\mathcal{S}$	−9.73	−9.88	−2.92	−2.76	−9.48	−9.68	−5.00	−4.79
	sk$\mathcal{S}$	−9.57	−9.74	−2.46	−2.34	−9.24	−9.46	−3.19	−3.44
Panel B: Stock market indices									
MS GARCH	$\mathcal{N}$	−4.04	−0.67	3.09	6.00	4.80	7.15	8.15	9.76
	sk$\mathcal{N}$	−5.25	−3.26	−1.04	3.29	3.06	5.46	6.18	8.55
	$\mathcal{S}$	−5.66	−2.90	−0.17	5.09	3.68	6.13	7.17	9.20
	sk$\mathcal{S}$	−6.08	−4.83	−3.52	0.22	2.00	4.39	4.98	7.71
MS GJR	$\mathcal{N}$	−9.65	−7.81	−6.19	−4.26	−4.30	0.33	2.19	4.76
	sk$\mathcal{N}$	−10.39	−9.41	−7.75	−6.35	−5.21	−3.00	−1.80	1.82
	$\mathcal{S}$	−9.79	−8.28	−6.91	−5.11	−4.66	−1.15	0.11	3.92
	sk$\mathcal{S}$	−10.20	−9.53	−8.29	−7.19	−5.34	−3.80	−2.83	0.47
Panel C: Exchange rates									
MS GARCH	$\mathcal{N}$	−2.65	−3.49	5.38	3.95	−2.06	−2.74	3.52	2.81
	sk$\mathcal{N}$	−2.00	−3.41	4.86	5.74	−1.53	−2.45	3.44	3.78
	$\mathcal{S}$	−6.84	−6.53	−2.17	−2.36	−6.09	−6.03	−2.31	−2.45
	sk$\mathcal{S}$	−5.45	−6.29	−0.99	−1.45	−4.81	−5.61	−1.32	−1.73
MS GJR	$\mathcal{N}$	−1.71	−2.33	4.40	3.59	−1.64	−2.35	5.32	3.89
	sk$\mathcal{N}$	−1.13	−1.95	4.26	4.53	−1.02	−1.88	4.53	5.14
	$\mathcal{S}$	−6.02	−6.03	−1.56	−1.68	−6.38	−6.38	−2.35	−2.46
	sk$\mathcal{S}$	−5.05	−5.49	−0.84	−1.21	−5.21	−5.74	−1.35	−1.66

Notes: This table presents the Diebold and Mariano (1995) test statistic of equal average wCRPS between a MS implementation (in rows) and a SR implementation (in columns), for all specifications considered, when forecasting the distribution of one-day-ahead log-returns. We report test statistics computed using robust HAC standard errors. Negative values indicate the Markov-switching specification outperforms the single-regime models. We also report statistics which are significantly negative (positive) at the 1% level in light (dark) gray (bilateral test). Models are estimated with the Bayesian approach.

Table 5
Standardized gain in average performance when switching from the MS GJR sk$\mathcal{S}$ model to the Beta-Skew-t-EGARCH(1,1) model.

	QL 1%	QL 5%	FZL 1%	FZL 5%	wCRPS
Stocks	−3.54	−0.29	−3.72	−1.17	−1.37
Stock market indices	−4.63	−1.52	−1.03	−1.03	−1.25
Exchange rates	−0.03	2.40	−0.04	1.04	0.58

Notes: This table presents the Diebold and Mariano (1995) test statistics of equal average loss between the MS GJR sk$\mathcal{S}$ model and the Beta-Skew-t-EGARCH(1,1) model for forecasting the distribution of one-day-ahead log-returns. As loss functions, we consider the QL and FZL measures (at $\alpha = 1\%$ and $\alpha = 5\%$), and the wCRPS measure. Negative values indicate the Markov-switching specification outperforms the Beta-Skew-t-EGARCH(1,1) model. We report statistics which are significantly negative at the 1% level in light gray (bilateral test). The critical values of the two-sided (one-sided) test are 2.57 (2.33), 1.96 (1.64), and 1.64 (1.28) at the 1%, 5%, and 10% significance levels, respectively. The models are estimated by maximum likelihood.

4. Conclusion

This paper investigates whether MSGARCH models provide risk managers with useful tools for improving the risk forecasts of the sorts of securities that are typically held by fund managers. Moreover, we investigate whether integrating the model's parameter uncertainty within the forecasts, via the Bayesian approach, improves the predictions. Our results and practical advice can be summarized as follows.

First, risk managers should extend their GARCH-type models with Markov-switching specifications in the case

Table 6
Standardized gain in average performance when switching from Bayesian to frequentist estimation.

Model	Stocks QL 1%	Stocks QL 5%	Stocks FZL 1%	Stocks FZL 5%	Stocks wCRPS	Stock market indices QL 1%	Stock market indices QL 5%	Stock market indices FZL 1%	Stock market indices FZL 5%	Stock market indices wCRPS	Exchange rates QL 1%	Exchange rates QL 5%	Exchange rates FZL 1%	Exchange rates FZL 5%	Exchange rates wCRPS
Panel A: Markov-switching GARCH models															
GARCH $\mathcal{N}$	−3.65	−3.26	−2.24	−2.85	−2.24	−0.33	−0.58	0.17	−1.78	−0.25	−1.33	0.99	−1.34	−0.37	−2.02
GARCH sk$\mathcal{N}$	−3.58	−2.93	−2.57	−2.81	−0.60	−1.56	−2.33	−2.05	−2.43	−1.04	−0.82	−1.24	0.97	0.35	−1.04
GARCH $\mathcal{S}$	−2.20	−5.78	−3.12	−4.27	−5.55	0.77	−0.17	−0.89	−0.97	−0.85	−0.78	0.29	−0.69	−1.00	0.35
GARCH sk$\mathcal{S}$	−5.04	−6.88	−6.12	−5.70	−7.04	1.13	−0.52	−0.62	−2.02	−0.58	−1.54	−1.64	−0.35	−1.54	−2.98
GJR $\mathcal{N}$	−1.91	−2.66	−1.59	−2.17	−3.22	−1.21	−2.95	−2.05	−1.81	−2.08	−1.09	−1.38	−0.85	−1.42	−3.61
GJR sk$\mathcal{N}$	−1.83	−3.12	−2.03	−2.44	−2.06	−1.11	−0.84	−2.47	−1.76	−1.40	0.06	−0.32	−0.28	−0.19	−1.17
GJR $\mathcal{S}$	−1.07	−3.11	−2.90	−2.67	−4.48	−1.29	−1.56	−1.66	−2.61	−4.11	−1.75	−2.40	−0.46	−0.51	−4.19
GJR sk$\mathcal{S}$	−3.10	−3.90	−4.67	−2.54	−5.28	−2.95	−2.02	−0.66	−0.46	−3.48	−1.59	−0.38	−0.81	−0.75	−2.50
Panel B: Single-regime GARCH models															
GARCH $\mathcal{N}$	−5.05	−4.23	−6.62	−4.50	−7.84	−2.99	−0.23	−3.40	−3.85	−5.63	−1.59	−0.42	−1.58	0.24	−3.20
GARCH sk$\mathcal{N}$	−4.77	−3.36	−6.59	−4.25	−6.64	−2.55	−1.05	−3.49	−1.98	−4.48	−1.33	−0.86	0.48	−1.00	−4.14
GARCH $\mathcal{S}$	−5.13	−5.08	−5.48	−5.34	−4.93	−1.27	−0.60	−3.01	−1.53	−3.39	−1.41	−1.12	0.15	1.18	−3.76
GARCH sk$\mathcal{S}$	−5.72	−5.40	−5.73	−5.44	−5.18	−2.74	−1.51	0.56	−0.38	−3.87	−2.83	−2.47	−1.74	−1.61	−4.46
GJR $\mathcal{N}$	−5.11	−2.80	−7.64	−3.90	−6.90	−3.65	−2.43	−5.52	−4.87	−5.92	−1.67	−2.70	−1.87	−0.93	−4.14
GJR sk$\mathcal{N}$	−4.55	−2.30	−7.22	−3.02	−5.65	−2.26	−2.03	−2.86	−1.30	−3.94	−0.13	−2.62	−1.53	−0.30	−4.61
GJR $\mathcal{S}$	−3.78	−4.23	−4.90	−4.14	−5.23	−4.13	−2.52	−4.94	−4.00	−4.17	−1.61	−1.96	0.97	−1.29	−4.71
GJR sk$\mathcal{S}$	−3.93	−4.06	−5.32	−4.41	−5.03	−3.82	−1.64	−3.01	−2.01	−3.16	−1.46	−2.24	−0.94	0.74	−4.66

Notes: This table presents the Diebold and Mariano (1995) test statistic of equal average loss between the Bayesian and frequentist estimated models for forecasting the distribution of one-day-ahead log-returns. As loss functions, we consider the QL and FZL measures (at $\alpha = 1\%$ and $\alpha = 5\%$), and the wCRPS measure. Panels A and B report the test statistics when comparing Bayesian and frequentist estimation for the SR and MS specifications, respectively. Negative values indicate the Bayesian estimation method outperforms the frequentist method. We report statistics which are significantly negative (positive) at the 1% level in light (dark) gray (bilateral test).

Table 7
Standardized average gain in performance when switching from constrained MS to SR models.

Horizon	Model	Stocks QL 1%	Stocks QL 5%	Stocks FZL 1%	Stocks FZL 5%	Stocks wCRPS	Stock market indices QL 1%	Stock market indices QL 5%	Stock market indices FZL 1%	Stock market indices FZL 5%	Stock market indices wCRPS	Exchange rates QL 1%	Exchange rates QL 5%	Exchange rates FZL 1%	Exchange rates FZL 5%	Exchange rates wCRPS
$h = 1$	GARCH sk$\mathcal{N}$	−0.44	−5.19	−3.56	−8.80	−9.34	−2.68	0.66	−3.53	−1.00	−3.26	0.27	−0.93	−1.13	−0.31	−3.70
	GARCH $\mathcal{S}$	−2.43	−2.32	−3.49	−3.74	−2.97	−1.53	−1.54	−1.68	−1.51	−1.02	0.73	0.10	−0.08	0.25	−0.53
	GARCH sk$\mathcal{S}$	−2.70	−2.69	−3.83	−4.22	−2.62	−1.70	−0.98	−1.33	−0.65	−1.40	1.62	0.31	−0.01	0.54	−0.23
	GJR sk$\mathcal{N}$	−0.37	−6.39	−3.90	−10.91	−9.92	−1.95	−1.59	−3.15	−2.21	−5.07	0.16	−0.91	−1.23	−0.89	−2.77
	GJR $\mathcal{S}$	−2.64	−2.99	−3.29	−3.80	−4.33	−2.25	−0.48	−2.29	−1.44	−0.64	0.30	−0.60	−0.63	−0.22	−1.04
	GJR sk$\mathcal{S}$	−2.90	−3.20	−3.31	−3.73	−4.10	−1.34	−0.93	−1.38	−0.89	−0.71	0.72	−0.40	0.32	−0.17	−0.97
$h = 5$	GARCH sk$\mathcal{N}$	−0.62	−1.79	−1.98	−7.07	−2.76	−2.05	−0.55	−3.02	−1.67	−0.73	2.75	0.63	−1.08	−0.27	0.18
	GARCH $\mathcal{S}$	−0.48	−0.98	−1.74	−1.93	−1.44	−0.92	−2.23	−1.00	−2.43	−1.72	1.96	0.28	−0.58	0.34	−0.58
	GARCH sk$\mathcal{S}$	−0.81	−1.06	−1.98	−2.07	−1.08	−1.58	−2.34	0.16	−1.11	−1.45	1.25	1.56	−0.54	0.65	0.46
	GJR sk$\mathcal{N}$	−0.44	−2.46	−2.11	−8.61	−2.60	−0.48	−1.49	−3.14	−2.55	−0.49	1.94	0.55	−1.26	−1.00	0.41
	GJR $\mathcal{S}$	−0.41	−1.26	−1.94	−2.29	−3.19	−0.19	−1.03	−2.62	−3.49	−0.87	3.19	1.40	−0.61	0.11	−0.56
	GJR sk$\mathcal{S}$	−0.21	−0.83	−1.71	−1.97	−2.86	0.80	−0.33	−1.59	−1.21	0.02	3.02	1.54	0.63	0.24	−0.92
$h = 10$	GARCH sk$\mathcal{N}$	−0.25	−0.84	−1.48	−6.63	−2.06	−1.92	0.03	−3.43	−2.68	−1.71	1.44	1.55	−0.80	−0.09	0.73
	GARCH $\mathcal{S}$	−0.29	−0.25	−0.86	−1.09	−0.23	−1.74	−1.46	−2.00	−2.85	−2.11	0.17	1.76	−0.48	0.40	0.10
	GARCH sk$\mathcal{S}$	−0.27	0.09	−0.61	−0.93	0.42	−0.40	−1.94	−1.01	−2.13	−2.30	1.20	1.30	1.19	1.70	0.92
	GJR sk$\mathcal{N}$	−0.10	−1.28	−1.43	−6.86	−1.80	−0.74	−1.92	−2.28	−2.49	−1.90	0.66	0.61	−1.06	−0.82	0.58
	GJR $\mathcal{S}$	−0.18	−0.41	−0.97	−1.55	−1.74	−0.53	−1.24	−2.23	−3.21	−1.79	2.36	1.27	−0.04	0.24	−0.25
	GJR sk$\mathcal{S}$	0.01	−0.25	−0.91	−1.29	−1.45	0.12	−1.54	−0.95	−1.60	−1.14	2.86	1.06	−0.36	0.47	−1.21

Notes: This table presents the Diebold and Mariano (1995) test statistic of equal average loss functions between the constrained MS and SR models for forecasting the distribution of h-day cumulative log-returns ($h \in \{1, 5, 10\}$). As loss functions, we consider the QL and FZL measures (at $\alpha = 1\%$ and $\alpha = 5\%$), and the wCRPS measure. We report the test statistics computed with robust HAC standard errors, for the time series in the various universes. Negative values indicate the shape-parameter-constrained Markov-switching specification outperforms its single-regime counterpart. We report statistics which are significantly negative (positive) at the 1% level in light (dark) gray (bilateral test). The multi-step cumulative log-returns forecasts are generated using 25,000 simulated paths of daily log-returns. The models are using the Bayesian approach.

of investments in equities. Indeed, we find that Markov-switching GARCH models deliver better Value-at-Risk, expected shortfall, and left-tail distribution forecasts than their single-regime counterparts, especially for stock return data. Moreover, the improvements are more pronounced when the Markov-switching mechanism is applied to simple specifications such as the GARCH-Normal model.

Second, accounting for parameter uncertainty helps for left-tail predictions independently of the inclusion of the Markov-switching mechanism. Moreover, larger improvements are observed when the parameter uncertainty is included in single-regime models.

Overall, we recommend that risk managers rely on more flexible models and perform inference to account for parameter uncertainty. We have assisted them in implementing these in practice by releasing the open-source R package MSGARCH; see Ardia et al. (2017) and Ardia et al. (forthcoming).

Our research could be extended in several ways. First, our study has considered single-regime versus two-state Markov-switching specifications. Hence, it would be of interest to see whether a third regime leads to superior performances, and whether the optimal number of regimes (according to penalized likelihood information criteria) changes over time and across data sets. Second, additional universes could be considered, such as emerging markets and commodities. Third, one could extend the set of models and compare the performances of MSGARCH and realized volatility models such as the HEAVY model of Shephard and Sheppard (2010). Fourth, as was suggested by a referee, it would be interesting to shed light on the parameter configurations for which the MSGARCH predictions can be expected to yield the greatest improvements in risk forecast precision. An exploratory analysis has shown that a high persistence of at least one state seems to be needed in order to achieve any substantial difference in precision between MSGARCH and single-regime GARCH downside risk forecasts. A definite answer to this question is beyond the scope of this paper. Finally, our analysis only considered financial risk monitoring systems for individual financial assets. The new standard for capital requirements for market risk (Basel Committee on Banking Supervision, 2016) calls for backtesting at both the individual desk level and the aggregate level. Thus, it would be interesting to consider also the impact of choices in modeling dependence. Including these extensions in our current research setup would have increased the (already large) number of models included in the comparison even further, so we have left them as a topic for future work.

Acknowledgments

An earlier version of this paper was circulated under the title "Forecasting performance of Markov-switching GARCH models: A large-scale empirical study". We are grateful to the Editor (Esther Ruiz), the Associate Editor, and two anonymous referees for their useful comments, which improved the paper significantly. We thank Samuel Borms, Peter Carl, Dirk Eddelbuettel, Richard Gerlach, Lennart Hoogerheide, Eliane Maalouf, Brian Peterson, Enrico Schumann, and conference participants at the R/Finance 2017 (Chicago), the 37th International Symposium on Forecasting (Cairns), UseR 2017 (Brussels), Quant Insights 2017 (London), MAFE 2018 (Madrid), and eRum 2018 (Budapest), as well as seminar participants at HEC Liège, Paris–Dauphine, and IAE–AMSE Aix-Marseille. We acknowledge Industrielle-Alliance, the International Institute of Forecasters, Google Summer of Code 2016 and 2017, FQRSC (Grant # 2015-NP-179931) and Fonds de Donations at the University of Neuchâtel for their financial support. We thank Félix-Antoine Fortin and Calcul Québec (clusters Briaree, Colosse, Mammouth and Parallèle II), as well as Laurent Fastnacht and the Institute of Hydrology at the University of Neuchâtel (cluster Galileo), for computational support. All computations have been performed using the R package MSGARCH (Ardia, Bluteau, Boudt, Catania, Peterson et al., forthcoming; Ardia, Bluteau, Boudt, Catania & Trottier, 2017), which is available from the CRAN repository at https://cran.r-project.org/package=MSGARCH.

References

Andersen, T. G., Bollerslev, T., & Diebold, F. X. (2007). Roughing it up: Including jump components in the measurement, modeling, and forecasting of return volatility. *Review of Economics and Statistics*, *89*(4), 701–720.

Andrews, D. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica*, *59*(3), 817–858.

Andrews, D., & Monahan, J. (1992). An improved heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica*, *60*(4), 953–966.

Ardia, D. (2008). *Financial risk management with Bayesian estimation of GARCH models: Theory and applications*. Heidelberg: Springer.

Ardia, D., Bluteau, K., Boudt, K., Catania, L., Peterson, B., & Trottier, D.-A. (2017). MSGARCH: Markov-switching GARCH models in R. URL https://cran.r-project.org/package=MSGARCH.

Ardia, D., Bluteau, K., Boudt, K., Catania, L., & Trottier, D.-A. (2017). Markov-switching GARCH models in R: The MSGARCH package. *Journal of Statistical Software* (forthcoming).

Ardia, D., Boudt, K., & Catania, L. (2018). Downside risk evaluation with the R package GAS. Working paper.

Ardia, D., & Hoogerheide, L. F. (2014). GARCH models for daily stock returns: impact of estimation frequency on Value-at-Risk and expected shortfall forecasts. *Economics Letters*, *123*(2), 187–190.

Ardia, D., Hoogerheide, L. F., & van Dijk, H. K. (2009). Adaptive mixture of Student-*t* distributions as a flexible candidate distribution for efficient simulation: The R package AdMit. *Journal of Statistical Software*, *29*(3), 1–32.

Ardia, D., Kolly, J., & Trottier, D.-A. (2017). The impact of parameter and model uncertainty on market risk predictions from GARCH-type models. *Journal of Forecasting*, *36*(7), 808–823.

Augustyniak, M. (2014). Maximum likelihood estimation of the Markov-switching GARCH model. *Computational Statistics & Data Analysis*, *76*, 61–75.

Bams, D., Blanchard, G., & Lehnert, T. (2017). Volatility measures and Value-at-Risk. *International Journal of Forecasting*, *33*(4), 848–863.

Basel Committee on Banking Supervision (2013). Fundamental review of the trading book: A revised market risk framework. Technical Report 265. Bank of International Settlements.

Basel Committee on Banking Supervision (2016). Minimum capital requirements for market risk. techreport 352. Bank of International Settlements.

Bauwens, L., De Backer, B., & Dufays, A. (2014). A Bayesian method of change-point estimation with recurrent regimes: Application to GARCH models. *Journal of Empirical Finance*, *29*, 207–229.

Bauwens, L., Dufays, A., & Rombouts, J. V. K. (2014). Marginal likelihood for Markov-switching and change-point GARCH models. *Journal of Econometrics*, *178*, 508–522.

Bauwens, L., & Laurent, S. (2005). A new class of multivariate skew densities, with application to generalized autoregressive conditional heteroscedasticity models. *Journal of Business & Economic Statistics*, *23*(3), 346–354.

Bauwens, L., Preminger, A., & Rombouts, J. V. K. (2010). Theory and inference for a Markov switching GARCH model. *The Econometrics Journal*, *13*(2), 218–244.

Bellini, F., & Bignozzi, V. (2015). On elicitable risk measures. *Quantitative Finance*, *15*(5), 725–733.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B.*, *57*(1), 289–300.

Bernardi, M., & Catania, L. (2016). Comparison of value-at-risk models using the MCS approach. *Computational Statistics*, *31*(2), 579–608.

Black, F. (1976). Studies of Stock price volatility changes. In *Proceedings of the 1976 meetings of the american statistical association* (pp. 177–181).

Blasques, F., Koopman, S. J., Łasak, K., & Lucas, A. (2016). In-sample confidence bands and out-of-sample forecast bands for time-varying parameters in observation-driven models. *International Journal of Forecasting*, *32*(3), 875–887.

Board of Governors of the Federal Reserve Systems (2012). 99th Annual Report. Technical Report. Board of Governors of the Federal Reserve Systems.

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

Boudt, K., Daníelsson, J., & Laurent, S. (2013). Robust forecasting of dynamic conditional correlation GARCH models. *International Journal of Forecasting*, *29*(2), 244–257.

Cai, J. (1994). A Markov model of switching-regime ARCH. *Journal of Business & Economic Statistics*, *12*(3), 309–316.

Clark, T. E., & McCracken, M. W. (2001). Tests of equal forecast accuracy and encompassing for nested models. *Journal of Econometrics*, *105*(1), 85–110.

Creal, D., Koopman, S. J., & Lucas, A. (2013). Generalized autoregressive score models with applications. *Journal of Applied Econometrics*, *28*(5), 777–795.

Danielsson, J. (2011). Risk and crises. VoxEU.org URL: http://voxeu.org/article/risk-and-crises-how-models-failed-and-are-failing.

Diebold, F. X., & Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics*, *13*(3), 253–263.
Dueker, M. J. (1997). Markov switching in GARCH processes and mean-reverting stock-market volatility. *Journal of Business & Economic Statistics*, *15*(1), 26–34.
Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica*, *50*(4), 987–1008.
Engle, R. F. (2004). Risk and volatility: Econometric models and financial practice. *American Economic Review*, *94*(3), 405–420.
Engle, R. F., & Manganelli, S. (2004). CAViaR: Conditional autoregressive Value at Risk by regression quantiles. *Journal of Business & Economic Statistics*, *22*(4), 367–381.
Fernández, C., & Steel, M. F. J. (1998). On Bayesian modeling of fat tails and skewness. *Journal of the American Statistical Association*, *93*(441), 359–371.
Fissler, T., & Ziegel, J. F. (2016). Higher order elicitability and Osband's principle. *The Annals of Statistics*, *44*(4), 1680–1707.
Geweke, J. (1989). Exact predictive densities for linear models with ARCH disturbances. *Journal of Econometrics*, *40*(1), 63–86.
Geweke, J., & Amisano, G. (2010). Comparing and evaluating Bayesian predictive distributions of asset returns. *International Journal of Forecasting*, *26*(2), 216–230.
Glosten, L. R., Jagannathan, R., & Runkle, D. E. (1993). On the relation between the expected value and the volatility of the nominal excess return on stocks. *Journal of Finance*, *48*(5), 1779–1801.
Gneiting, T., & Ranjan, R. (2011). Comparing density forecasts using threshold — and quantile — weighted scoring rules. *Journal of Business & Economic Statistics*, *29*(3), 411–422.
González-Rivera, G., Lee, T.-H., & Mishra, S. (2004). Forecasting volatility: A reality check based on option pricing, utility function, Value-At-Risk, and predictive likelihood. *International Journal of Forecasting*, *20*(4), 629–645.
Gray, S. F. (1996). Modeling the conditional distribution of interest rates as a regime-switching process. *Journal of Financial Economics*, *42*(1), 27–62.
Haas, M., Mittnik, S., & Paolella, M. S. (2004). A new approach to Markov-switching GARCH models. *Journal of Financial Econometrics*, *2*(4), 493–530.
Hamilton, J. D. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica*, *57*(2), 357–384.
Hamilton, J. D. (1994). *Time series analysis* (1st ed.). Princeton, USA: Princeton University Press.
Hamilton, J. D., & Susmel, R. (1994). Autoregressive conditional heteroskedasticity and changes in regime. *Journal of Econometrics*, *64*(1–2), 307–333.
Hansen, P. R., Huang, Z., & Shek, H. H. (2011). Realized GARCH: A joint model for returns and realized measures of volatility. *Journal of Applied Econometrics*, *27*(6), 877–906.
Hansen, P. R., & Lunde, A. (2005). A forecast comparison of volatility models: Does anything beat a GARCH(1,1)? *Journal of Applied Econometrics*, *20*(7), 873–889.
Harvey, A. C. (2013). *Dynamic models for volatility and heavy tails: with applications to financial and economic time series, Vol. 52*. Cambridge University Press.
Harvey, A., & Sucarrat, G. (2014). EGARCH models with fat tails, skewness and leverage. *Computational Statistics & Data Analysis*, *76*, 320–338.
Herwartz, H. (2017). Stock return prediction under GARCH – An empirical assessment. *International Journal of Forecasting*, *33*(3), 569–580.
Jacquier, E., Polson, N. G., & Rossi, P. E. (1994). Bayesian analysis of stochastic volatility models. *Journal of Business & Economic Statistics*, *12*(4), 371–389.
Jorion, P. (2006). *Value-at-risk — the new benchmark for managing financial risk* (3rd ed.). McGraw–Hill.
Klaassen, F. (2002). Improving GARCH volatility forecasts with regime-swi;tching GARCH. In *Advances in Markov-switching models* (pp. 223–254). Springer-Verlag.
Kupiec, P. H. (1995). Techniques for verifying the accuracy of risk measurement models. *Journal of Derivatives*, *3*, 73–84.
Lamoureux, C. G., & Lastrapes, W. D. (1990). Persistence in variance, structural change, and the GARCH model. *Journal of Business & Economic Statistics*, *8*(2), 225–234.
Laurent, S., Lecourt, C., & Palm, F. C. (2016). Testing for jumps in conditionally Gaussian ARMA-GARCH models, a robust approach. *Computational Statistics & Data Analysis*, *100*, 383–400.
Marcucci, J. (2005). Forecasting stock market volatility with regime-switching GARCH models. *Studies in Nonlinear Dynamics & Econometrics*, *9*(4), 1558–3708.
Matheson, J. E., & Winkler, R. L. (1976). Scoring rules for continuous probability distributions. *Management Science*, *22*(10), 1087–1096.
McAleer, M., & Da Veiga, B. (2008). Single-index and portfolio models for forecasting Value-at-Risk thresholds. *Journal of Forecasting*, *27*(3), 217–235.
McNeil, A. J., Frey, R., & Embrechts, P. (2015). *Quantitative risk management: concepts, techniques and tools* (2nd ed.). Princeton University Press.
Nelson, D. B. (1991). Conditional heteroskedasticity in asset returns: A new approach. *Econometrica*, *59*(2), 347–370.
Opschoor, A., van Dijk, D., & van der Wel, M. (2017). Combining density forecasts using focused scoring rules. *Journal of Applied Econometrics*, *32*(7), 1298–1313.
Patton, A. J., Ziegel, J. F., & Chen, R. (2017). Dynamic semiparametric models for expected shortfall. Working paper.
Shephard, N., & Sheppard, K. (2010). Realising the future: Forecasting with high-frequency-based volatility (HEAVY) models. *Journal of Applied Econometrics*, *25*(2), 197–231.
Taylor, S. J. (1994). Modeling stochastic volatility: A review and comparative study. *Mathematical Finance*, *4*(2), 183–204.
Trottier, D.-A., & Ardia, D. (2016). Moments of standardized Fernández-Steel skewed distributions: Applications to the estimation of GARCH-type models. *Finance Research Letters*, *18*, 311–316.
Vihola, M. (2012). Robust adaptive Metropolis algorithm with coerced acceptance rate. *Statistics and Computing*, *22*(5), 997–1008.
Ziegel, J. F. (2016). Coherence and elicitability. *Mathematical Finance*, *26*(4), 901–918.