Censoring in Metropolis - Hastings sampling algorithm. Left-tail prediction with GARCH models.

Lukasz T. Gatarek*, Koen Hooning, Lennart F. Hoogerheide and Herman K. Van Dijk

October 16, 2011

Abstract

Abstract: The estension of the Metropolis-Hastings algorithm is proposed by replacing the standard likelihood with the censored likelihood for the acceptance-rejection weights evaluation. Further the Bayesian Model Averaging is extended by replacing the predictive likelihood by the censored predictive likelihood. The use of censored Metropolis-Hastings and censored predictive likelihood offers a focus on the left tail of the distirbution. We perform an extensive simulation study to investigate the ability of these methods to outperform the standard sampling algorithm and traditional Bayesian Model Averaging techniques with application to the GARCH models and Value-at-Risk methods (coverage probability).

Keywords:

*Econometric Institute and Tinbergen Institute, Erasmus University Rotterdam, P.O. Box 1738, NL-3000 DR Rotterdam, The Netherlands. E-mail: gatarek@tlen.pl

1 Introduction

The Bayesian approach assumes that only probabilistic expression for the parameters in a model can be given. The inference is based on the information presently available which consists of the probability function of the dataset according to the statistical model and possibly some prior knowledge. By combining this information the posterior density functions of model parameters can be defined. The posterior densities describe the present states of mind concerning the parameters. The posterior densities can be used for model combination. This approach assume that for statistical inference based on individual model is abandoned in favour the average of distinct model specifications weighted by their posterior model probability. To account for the potential problems related to impact of prior diffuseness for model selection the standard likelihood in evaluation of posterior model probabilities is often replaced by the predictive likelihood. We extend this approach in that the predictive likelihood is replaced by the censored predictve lieklihood. It can be useful in situations where the focus concerns the left tail of the distribution. A potential field of implementation of such algorithms is Value-at-Risk analysis where the emphasis is to estimate the probability of extreme outcomes located in the left tail of the distribution.

We propose a new approach to the Metropolis-Hastings sampling algorithm for posterior distribution. We censor the likelihood for evaluation of acceptance-rejection weights in this algorithm. As a consequence the Bayesian model averaging (BMA) is based upon Bayes factors evalauted with censored likelihood. We consider both standard and predictive Bayes factors. The approach is tested within GARCH models that play important role in Value-at-Risk analysis. Apart from simple simualtion experiment we consider deep empirical analysis with abundant model specifications and complete set of Bayesian model averages. Due to specifics of Value-at-Risk analysis the methodology is evaluated via measures of accuracy of predictive distribution.

2 Metropolis-Hastings algorithm

The Metropolis-Hastings (M-H) algorithm developed in [Hastings, 1970] and [Metropolis et al., 1973] can be used to sample from posterior distributions if the closed form expressions for posterior density are not available. The algorithm is based on the theory of Markov chains and is usually constructed as follows:

Step 1. Calculate the maximum likelihood estimator $\hat{\theta}_{ML}$ of parameter θ in a statistical model.

Step 2. Choose a certain candidate density $q(\theta)$ centred around $\hat{\theta}_{ML}$.

Step 3. Define the weight function

$$w(\theta) = \frac{k(\theta)}{q(\theta)}, \tag{1.1}$$

where $k(\theta)$ is the posterior density kernel given by the product of the likelihood and the prior density kernel, i.e. $k(\theta) = p(y|\theta)p(\theta)$ and where $q(\theta)$ is the candidate density.

Step 4. Construct a Markov chain by generating draws θ^i from the candidate density $q(\theta)$ and accepting or rejecting them based on a transition probability. If the candidate draw is accepted the chain moves to the new value, else the chain stays in the current state. The transition probability of the generated candidate draws θ^i is given by

$$\alpha(\theta^{i-1}, \theta^i) = \min\left\{\frac{w(\theta^i)}{w(\theta^{i-1})}, 1\right\}. \tag{1.2}$$

Draws from the Markov chain generated according to this algorithm are considered as draws from the posterior distribution of the model parameters. To remove the influence of the starting point of the Markov chain a *burn-in* period can be used.

For financial models the candidate density should be selected among the fat-tailed distributions so that it covers the entire posterior density space. For GARCH models that we consider it is natural to select a Student-t distribution with low degrees of freedom.

3 Censored M-H

In some applied statistical problems the attention is mainly paid to the left tail of the likelihood. We construct the Metropolis-Hastings algorithm that explores this part of the distribution. To that end we apply the logarithmic censored likelihood scoring rule. [Diks et al, 2008] describe a logarithmic censored likelihood function (CSL) to be used in order to define a scoring rule with a focus on a particular part of the density forecast. We apply this function to censor the likelihood of the data. It is given by

$$S^{csl}(\hat{f}_t; y_t) = \mathrm{I}(y_t \leq r) \log \hat{f}_t(y_t) + \mathrm{I}(y_t > r) \log \left(1 - \hat{F}_t(r)\right), \tag{1.3}$$

where $\hat{f}_t(y_t)$ is the density function of $y_t,\ t \in 1, \dots, T$ and r indicates the region of interest, in this case specified for values less than or equal to r. Using this function we can extend the standard M-H algorithm to censored M-H algorithm where the censored likelihood function is given by

$$\mathcal{L}^{csl}(\theta | y_1, ..., y_T) = p^{csl}(y|\theta) = \prod_{t=1}^{T} p^{csl}(y_t | y_1, ... y_{t-1}, \theta), \tag{1.4}$$

where the standard likelihood $p(y_t|y_1, ...y_{t-1}, \theta)$ is replaced by censored likelihood $p^{csl}(y_t|y_1, ...y_{t-1}, \theta)$ specified as the exponential of S^{csl} in (1.3) evaluated for y_t and a certain value of r. For a fair comparison of the different models, r should be equal for all models at each point in time. In the simulation and empirical studies that are presented in this paper parameter r is equal to 20% and 30% Value at Risk level based on the corresponding quantile of the empirical CDF Φ of the in sample returns. These levels are further referred to as the cut-off points.

To construct the censored M-H algorithm in ***Step 3.*** of the standard algorithm M-H the likelihood $p(y|\theta)$ is replaced by censored likelihood $p^{csl}(y|\theta)$. Then we simulate from a CSL-based posterior (censored posterior). The candidate density $q(\theta)$ remains unchanged compared to the standard algorithm.

4 Bayesian Model Averaging (BMA)

With Bayesian approach to model estimation it is possible to construct efficient forecast combinations. These methods are advocated when the selection of single model specification is embarked with high risk of misspecification or the analyst wishes to efficiently combine predictive power of many distinct specifications rather than to adhere to a single model indicated as an optimal choice according to specified evalaution criterion.

In Bayesian econometrics we can calculate the posterior odds ratio to discriminate between two models. If the two models M_1 and M_2 are considered, given data y, their model probabilities are respectively given by

$$\Pr(M_1|y) = \frac{p(M_1, y)}{p(y)} = \frac{p(y|M_1)\Pr(M_1)}{p(y|M_1)\Pr(M_1) + p(y|M_2)\Pr(M_2)}, \tag{1.5}$$

and

$$\Pr(M_2|y) = \frac{p(M_2, y)}{p(y)} = \frac{p(y|M_2)\Pr(M_2)}{p(y|M_1)\Pr(M_1) + p(y|M_2)\Pr(M_2)}. \tag{1.6}$$

Combining these two equations we can define the posterior odds ratio as

$$\frac{\Pr(M_1|y)}{\Pr(M_2|y)} = \frac{\Pr(M_1)}{\Pr(M_2)}\frac{p(y|M_1)}{p(y|M_2)}, \tag{1.7}$$

where $\Pr(M_1)/\Pr(M_2)$ is the reffered to as a prior odds ratio and $p(y|M_1)/p(y|M_2)$ is the Bayes factor. If there is no *a priori* preference for one of the two models the prior odds ratio is set to 1. Then the posterior odds ratio simplifies to the Bayes factor which is given by the ratio of the marginal likelihoods of the data under the two different models. If we extend this principle to M models we find the standard approach to forecast combination using BMA in which the posterior probabilities of the models, given the data y, are given by

$$p(M_i|y) = \frac{p(y|M_i)\Pr(M_i)}{\sum_{j=1}^{M} p(y|M_j)\Pr(M_j)}. \tag{1.8}$$

Using these posterior model probabilities one can construct the 'minimum mean squared error one-period-ahead forecast' $\mathrm{E}[y_{T+1}|y]$ by

$$\mathrm{E}[y_{T+1}|y] = \sum_{i=1}^{M} \mathrm{E}[y_{T+1}|y, M_i]p(M_i|y), \tag{1.9}$$

which is simply an average of the 'minimum mean squared error one-day-ahead forecasts' under each of the models, weighted by the posterior model probabilities.

From the above we learn that the marginal likelihood $p(y|M_i)$ plays an important role in defining the posterior model probabilities in the BMA. We showed that $p(y|M_i)$ is given by $p(y|M_i) = \int p(y|\theta_i, M_i)p(\theta_i|M_i)d\theta_i$. Thus we find that the priors of the parameters $p(\theta_i|M_i)$ in the different models influence the posterior model probabilities. If one possesses highly informative, tight priors $p(\theta_i|M_i)$, the BMA will result in usable posterior model probabilities. If not, the BMA may result in unreliable posterior model probabilities. The reason of this phenomenon is known in the statistical literature as the Bartlett's paradox, seeh10.

The Bartlett's paradox may be interpreted as the fact that if we spread too much prior probability mass in the prior of say model 1 $p(\theta_1|M_1)$ over silly values, i.e. we make the prior wide enough as compared to the priors of the other models $p(\theta_i|M_i)$, we can make the posterior probability of model 1

$$p(M_1|y) = \frac{p(y|M_1)\Pr(M_1)}{\sum_{j=1}^{M} p(y|M_j)\Pr(M_j)} \tag{1.10}$$

as small as we want, independent of the information in the data y, see also [Hoogerheide, 2010]. Therefore, another method is needed to compute posterior model probabilities if no highly informative priors for all the models are available. This method makes use of the predictive likelihood.

4.1 BMA using predictive likelihood (predictive BMA)

According to the Barlett's paradox one should not use the marginal likelihood in BMA if no highly informative priors for all models can be defined. Instead one should use the predictive likelihood given by

$$p(\tilde{y}|y^*, M_i) = \int p(\tilde{y}|\theta_i, y^*, M_i)p(\theta_i|y^*, M_i)d\theta_i, \tag{1.11}$$

where $y^* = (y_1, ..., y_m)$ and $\tilde{y} = (y_{m+1}, ..., y_T)$. We see that the predictive likelihood is similar to the marginal likelihood but instead of a non-informative prior a prior based on the first m values of the dataset is used. This prior is given by the posterior density $p(\theta_i|y^*, M_i)$ based on the observations $y^* = (y_1, ..., y_m)$. According to Bayes' rule this posterior density is given by

$$p(\theta_i|y^*, M_i) = \frac{p(y^*|\theta_i, M_i)p(\theta_i|M_i)}{\int p(y^*|\theta_i, M_i)p(\theta_i|M_i)d\theta_i}. \tag{1.12}$$

Substituting this into (1.11) yields

$$p(\tilde{y}|y^*, M_i) = \frac{\int p(\tilde{y}|\theta_i, y^*, M_i)p(y^*|\theta_i, M_i)p(\theta_i|M_i)d\theta_i}{\int p(y^*|\theta_i, M_i)p(\theta_i|M_i)d\theta_i}, \tag{1.13}$$

which is equal to

$$p(\tilde{y}|y^*, M_i) = \frac{\int p(y|\theta_i, M_i)p(\theta_i|M_i)d\theta_i}{\int p(y^*|\theta_i, M_i)p(\theta_i|M_i)d\theta_i}. \tag{1.14}$$

Now it becomes clear that the predictive likelihood is simply given by the ratio of the marginal likelihood of all observations over the marginal likelihood for the first m observations. This way the first m observations are used to delete the completely silly values from the original non-informative prior $p(\theta_i|M_i)$. Using the predictive likelihood the posterior model probabilities in the BMA can now be defined as

$$p(M_i|y) = \frac{p(\tilde{y}|y^*, M_i)\Pr(M_i)}{\sum_{j=1}^{M} p(\tilde{y}|y^*, M_j)\Pr(M_j)}. \tag{1.15}$$

However it is often not possible to evaluate the integrals for calculation of marginal (and thus predictive) likelihoods analytically. Luckily, the simulation method of Metropolis and Hastings comes with a rather easy solution. The marginal likelihood

$$p(y|M_i) = \int p(y|\theta_i, M_i)p(\theta_i|M_i)d\theta_i, \tag{1.16}$$

gives

$$p(y|M_i) = \int \frac{p(y|\theta_i, M_i)p(\theta_i|M_i)}{q(\theta_i)}q(\theta_i)d\theta_i, \tag{1.17}$$

which is equal to

$$p(y|M_i) = \mathrm{E}\left[\frac{p(y|\tilde{\theta}_i, M_i)p(\tilde{\theta}_i|M_i)}{q(\tilde{\theta}_i)}\right], \tag{1.18}$$

where $q(\tilde{\theta}_i)$ is the candidate density defined in Step 2 of the M-H algorithm and $\tilde{\theta}_i$ has this candidate density. The numerical estimation for the expectation is obviously given by

$$\hat{p}(y|M_i) = \frac{1}{M}\sum_{j=1}^{M}\frac{p(y|\tilde{\theta}_i^j, M_i)p(\tilde{\theta}_i^j|M_i)}{q(\tilde{\theta}_i^j)} = \frac{1}{M}\sum_{j=1}^{M} w(\tilde{\theta}_i^j), \tag{1.19}$$

where $\tilde{\theta}_i^j$ are draws from the candidate density $q(\tilde{\theta}_i)$ and $w(\tilde{\theta}_i)$ is the weight function as defined in Step 3 of the MH algorithm. Note that (1.19) actually is the Importance sampling estimator of the marginal likelihood.

One question that remains when using the predictive likelihood is how to divide the data in a training sample y^* and a hold-out sample $\tilde{y}$. [Gelfand and Dey, 1994] give an overview of different options. However following [Eklund and Karlsseon, 2007] we simply use the equal sample split; the first part of the data, the training sample y^*, is used to obtain posterior distributions that are used to update the priors to construct more reliable prior distribution. The updated prior distributions are then used for assessing the fit of the model to the data $\tilde{y}$ and for model selection procedures via Bayes factors. By setting m equal to half the number of observations in y we assure that both the training and the hold-out sample contain enough data to obtain reliable posterior distributions.

4.2 BMA using censored predictive likelihood

In BMA using predictive likelihood the entire posterior density $p(y|\theta_i, M_i)$ is considered as equally important. However, in empirical applications the focus of the analyst is to correctly describe a particular part of the density. For example, in risk management only the downside risk described by the left tail of the density is of crucial importance in order to obtain correct Value-at-Risk forecasts. As a consequence it seems natural to focus on the correct description of the left tail of the density.

Using function 1.3 to construct the censored M-H algorithm, we can extend both the BMA using standard and predictive likelihood into BMA using censored likelihood resulting in a new Bayesian Model Averaging methods that weights the models according to their tail likelihood. This is achieved by replacing the predictive likelihood in (1.11) by the censored predictive likelihood defined as

$$p^{csl}(\tilde{y}|y^*, M_i) = \int p^{csl}(\tilde{y}|\theta_i, y^*, M_i) p(\theta_i|y^*, M_i) d\theta_i. \tag{1.20}$$

The distribution $p(\theta_i|y^*, M_i)$ remains unaltered and is given by (1.12). Substituting this into (1.20) results in

$$p^{csl}(\tilde{y}|y^*, M_i) = \frac{\int p^{csl}(\tilde{y}|\theta_i, y^*, M_i) p(y^*|\theta_i, M_i) p(\theta_i|M_i) d\theta_i}{\int p(y^*|\theta_i, M_i) p(\theta_i|M_i) d\theta_i}, \tag{1.21}$$

which in this case is clearly not equal to

$$p^{csl}(\tilde{y}|y^*, M_i) = \frac{\int p^{csl}(y|\theta_i, M_i) p(\theta_i|M_i) d\theta_i}{\int p(y^*|\theta_i, M_i) p(\theta_i|M_i) d\theta_i}, \tag{1.22}$$

because

$$p^{csl}(\tilde{y}|\theta_i, y^*, M_i) p(y^*|\theta_i, M_i) \neq p^{csl}(y|\theta_i, M_i). \tag{1.23}$$

As a consequence, $p^{csl}(\tilde{y}|y^*, M_i)$ is no longer expressed by the ratio of the marginals which can be estimated numerically as in (1.19). Instead, the numerator in (1.21) can be numerically estimated by

$$\frac{1}{M} \sum_{j=1}^{M} \frac{p^{csl}(\tilde{y}|\tilde{\theta}_i^j, y^*, M_i) p(y^*|\tilde{\theta}_i^j, M_i) p(\tilde{\theta}_i^j|M_i)}{q(\tilde{\theta}_i^j)}, \tag{1.24}$$

where $q(\tilde{\theta}_i)$ is the candidate density for the whole dataset y defined in Step 2 of the M-H algorithm and $\tilde{\theta}_i$ has this candidate density. The denominator still corresponds to the marginal likelihood of y^* and can thus be estimated as

$$\hat{p}(y^*|M_i) = \frac{1}{M}\sum_{j=1}^{M} \frac{p(y^*|\tilde{\theta}_i^{\,j}, M_i)p(\tilde{\theta}_i^{\,j}|M_i)}{q(\tilde{\theta}_i^{\,j})} = \frac{1}{M}\sum_{j=1}^{M} w(\tilde{\theta}_i^{\,j}), \tag{1.25}$$

where $q(\tilde{\theta}_i)$ is the candidate density for the first m observations of the dataset, i.e. y^*, as defined in Step 2 of the M-H algorithm and $\tilde{\theta}_i$ has this candidate density.

Having defined the censored predictive likelihood the model weights according to the BMA using censored predictive likelihood can be defined as

$$\hat{p}^{csl}(M_i|y) = \frac{\hat{p}^{csl}(\tilde{y}|y^*, M_i)\Pr(M_i)}{\sum_{j=1}^{M} \hat{p}^{csl}(\tilde{y}|y^*, M_j)\Pr(M_j)}. \tag{1.26}$$

5 Simulation study

VaR modelling techniques require accuracy in forecasting the tails of a returns density rather than in the main body of the the return distribution. Therefore we focus on the ability of a forecast to predict returns in the lower tail, i.e. the extreme losses in a portfolio. The prediction accuracy is evaluated via the coverage probability. The coverage probability is defined as the probability associated with a given interval of a distribution. For example the coverage probability of a 5% tail of a certain distribution is 5%. The coverage probabilites are designed to evaluate the accuracy of a forecast for a specific interval of the distribution.

In this simulation study we work with the Student-t GARCH(p, q) as defined in [Bollerslev, 1986]. Our focus conserns the left tail of the distribution. The DGP is based on Student-t distributed innovations. However to emphasize the role of the left tail of the distribution we assume the following distribution of innovations

$$\varepsilon \sim I\{Poisson(\lambda) = 0\} \times t(df) \; + \; Poisson(\lambda) \times \; \Phi^{-1}_{t(df)}. \tag{1.27}$$

If conditon $\{Poisson(\lambda) = 0\}$ is met the innovation is drawn from Student-t with df degrees of freedom. In the contrary, if $Poisson(\lambda) \neq 0$ the innovation is sampled from the left tail of the $t(df)$ distribution determined by cumulated probability $Uniform(0, 1) \times \alpha$, which can maximally be drawn as α. $\Phi_{t(df)}$ denotes the c.d.f. of the $t(df)$ distribution. The parameter α controls for the area of the Student-t distribution that is of particular interest in our simulation. By increasing λ appropriately we can control for the number of extreme outcomes for the DGP. The parametric part of the model is specified as GARCH$(2, 2)$ model with constant. All 5 parameters of the models are set to 0.07. The other parameters in the simulation are: $alpha = 0.1$, $df = 8$ $\lambda = 0.55$.

To focus on the left tail we estimate the first 10 quantiles of the predictive distribution. We consider 3 options of the M-H sampling algorithm

1. use 20% pdf's and 80% (1-CDF)'s information, thus with cut-off at 20% of the entire dataset;

2. use 30% pdf's and 70% (1-CDF)'s information, thus with cut-off at 20% of the entire dataset;

3. use 100% pdf's information (no cut-off) - standard, uncensroed M-H algorithm.

Simulation setup

1. For given $p = 2$ and $q = 2$ simulate 1000 observations from GARCH(2,2) model for daily returns.

2. Set $i = 0$ to initialize the rolling window prediction experiment.

3. Fix the estimation sample $1 + i, \ldots, T/2 + i$ and estimate all achievable GARCH (p,q) models for $p = 1, 2$ and $q = 1, 2$ with accurate M-H sampling procedure.

4. For each model and for Bayesian model avergae over all those models construct the predictive distribution for period $T/2 + i + 1$. (Every draw from the sampler results in distinct predicted value for return in $T/2 + 1$. The distribution of these forecasts defines the predictive distribution.)

5. Estimate the quantiles $1, \ldots, 10$ of the predictive distribution for $T/2 + i + 1$.

6. Compare the actual (realized) 1-period ahead returns with the quantiles of predictive distirbution (VaR estimates) and record the exceedances.

7. Return to step 3 and repeat with the estimation sample rolled one observation forward ($i := i + 1$).

8. Repeat steps $3 - 6$ until the entire training sample $T/2, , T$ is exhausted.

9. For each quantile considered sum up the exceedances recorded over $T/2$ experiments defined by sequence of Steps $3 - 6$. In the ideal case the number of exceedances divided by sample size $T/2$ (known alternatively as the Value-at-Risk 'violation rate') shall be equal to the nominal probability defined by the quantile.

10. Return to step 1 and repeat the entire experiment 100 times. The sequence of 100 exceedance levels recorded for each Value-at-Risk level provides for the distribution of the 'violation rate'. Compute the average 'violation rate'.

11. Compute the squared errors between the average 'violation rate' and the nominal 'violation rate' for each quantile considered. The average of these squared errors results in MSE that is considered a measure to evaluate the models and model averages.

This exerise is performed for censored M-H with cut-offs of 20 and 30% as well as for uncensored M-H algorithm. We consider both standard and predictive approach to Bayes factors for Bayesian model averaging.

In Figures 1, 3, 2 and 4 the MSE defined in the setup of the simulation experimetn is presented for each considered model specification. As expected, for the true model, GARCH$(2,2)$ the censored M-H is outperformed by the standard M-H algorithm, see Figure 4, becuase for the true model it is optimal to use all observations to learn about the model parameters. For GARCH$(1,1)$, GARCH$(1,2)$ and GARCH$(2,1)$ we observed higher accuracy of predictive distribution based on censored M-H sampler.

In Figure 6 we present the result of applying central idea of this paper, the Censored Predictive Bayesian Model Averaging. It is contrasted with standard Predictive Bayesian Model Averaging method. Moreover as the benchmark we also apply the standard Bayesian Model Averaging, based on simple posterior probability. Alike the predictive BMA, the standard BMA is also applied in two variants: censored and uncensored. The outcome of applying the standard methods is depicted in Figure 5. We observe that for both the standard BMA and predictive BMA methods, censoring the distribution impacts the accuracy of the predictive distribution.

In Table 1 we present the average squared deviation between the observed 'violation

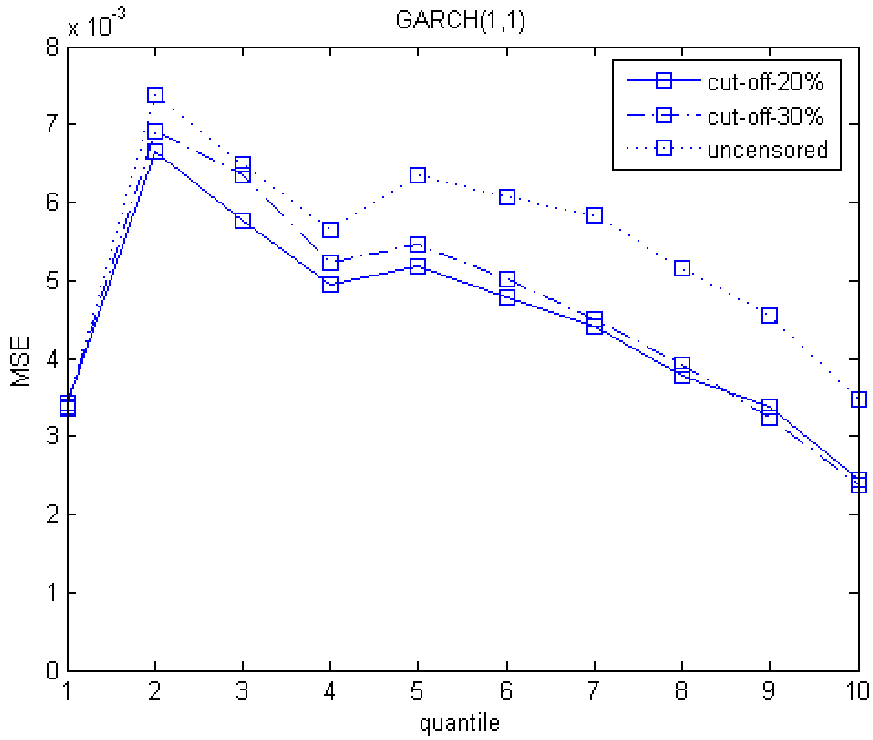

Figure 1: MSE between the observed 'violation percentage' and the nominal 'violation percentage' under GARCH$(1, 1)$.

percentage' and the 'perfect' nominal 'violation percentage' for all model specifications to enable compariosn across the models. We observe that for censored M-H the predictive BMA method based is an unquestioned winner among all misspecified models. It has substantial advantage over standard BMA and all individual models. The true model GARCH$(2, 2)$ is outperforming predictive BMA approach but the advantage is not substantial. In case of uncensored M-H the true model is leading followed by individual models. The BMA methods are able to outperform only the simplest model. This exercise indicate that for estimation of the left tail of the predictive distribution the censoring is particularly advocated jointly with the predictive BMA technique.

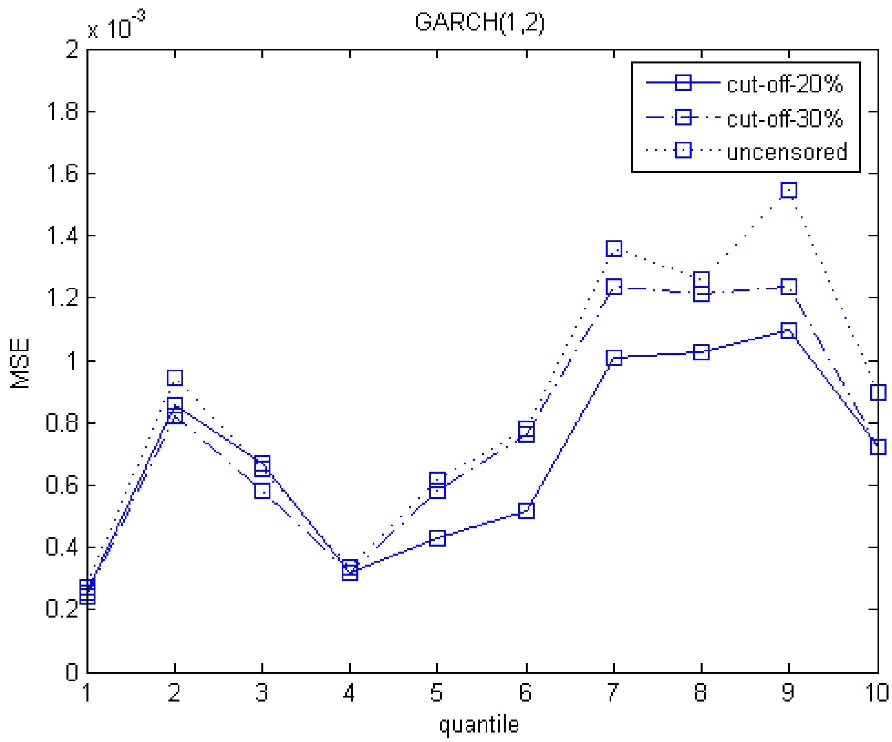

Figure 2: MSE between the observed 'violation percentage' and the nominal 'violation percentage' under GARCH$(1, 2)$.

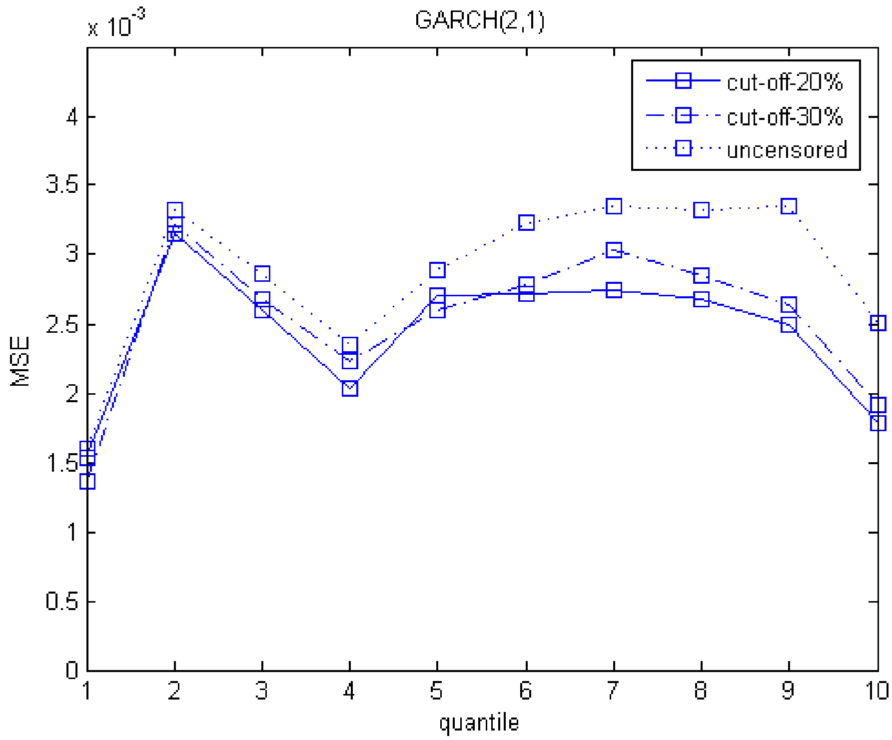

Figure 3: MSE between the observed 'violation percentage' and the nominal 'violation percentage' under GARCH(2, 1).

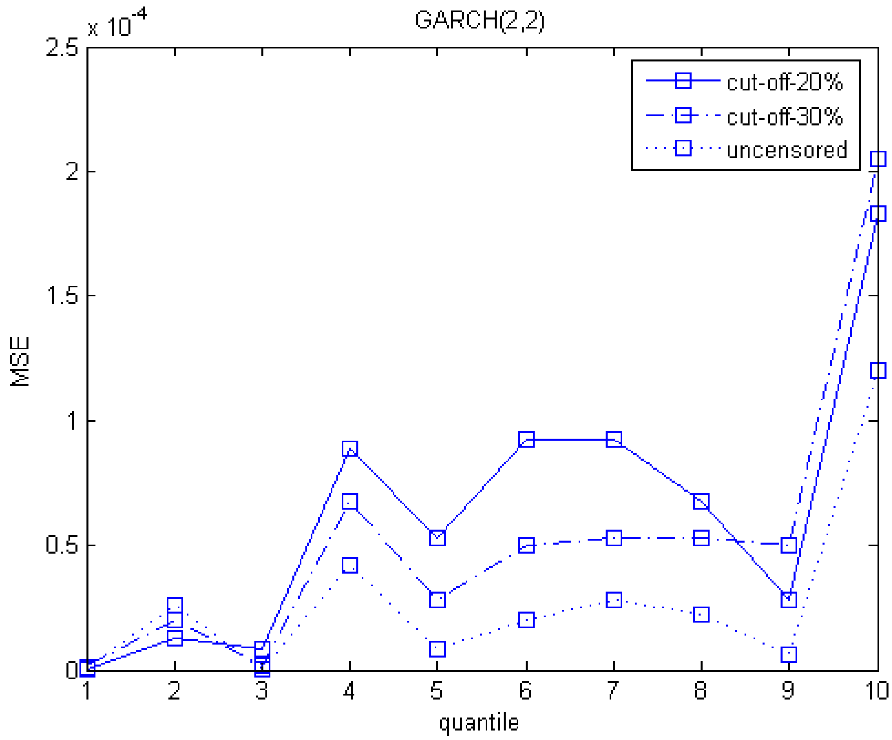

Figure 4: MSE between the observed 'violation percentage' and the nominal 'violation percentage' under GARCH$(2, 2)$.

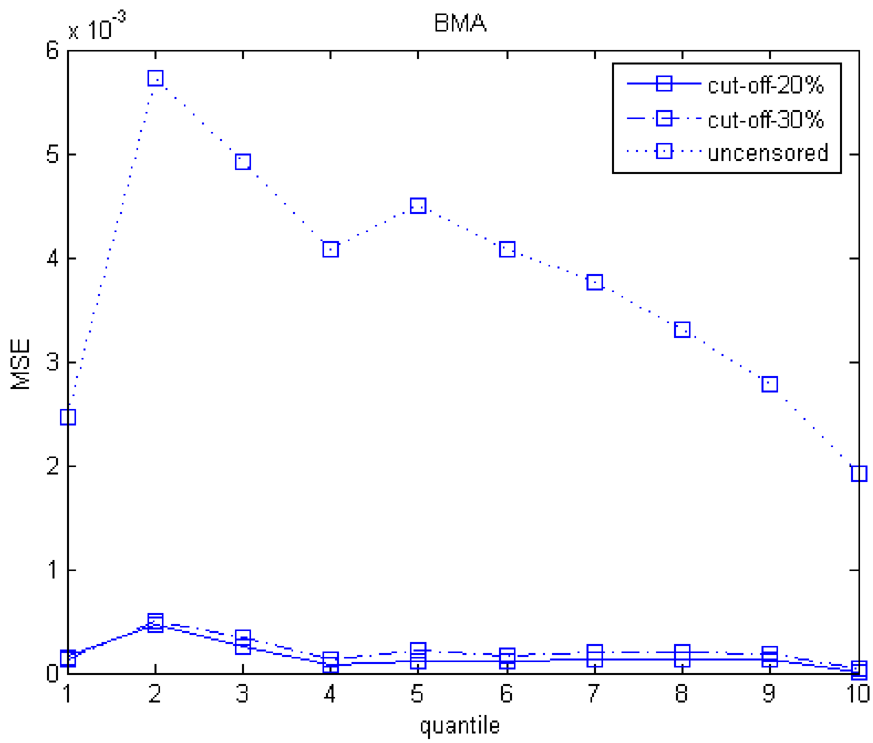

Figure 5: MSE between the observed 'violation percentage' and the nominal 'violation percentage' under censored BMA.

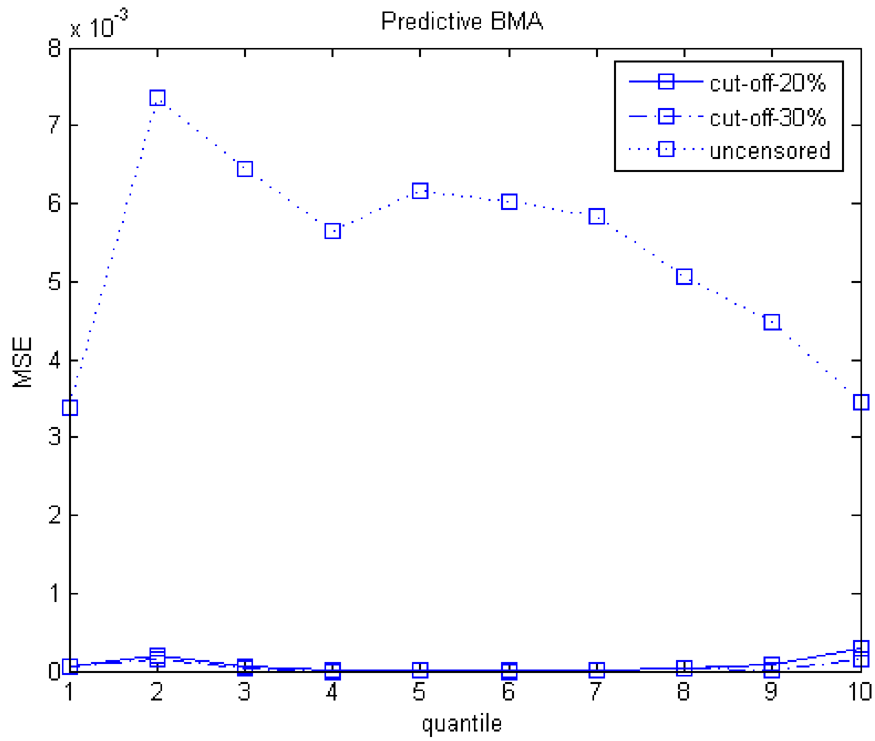

Figure 6: MSE between the observed 'violation percentage' and the nominal 'violation percentage' under predictive censored BMA.

$MSE_{90\%,\ldots,99\%}$			
	20%	30%	uncensored
GARCH(1,1)	0.04475	0.04640	0.05431
GARCH(1,2)	0.00690	0.00771	0.00866
GARCH(2,1)	0.02443	0.02529	0.02876
GARCH(2,2)	0.00062	0.00052	0.00027
BMA	0.00160	0.00210	0.03760
Predictive BMA	0.00080	0.00040	0.05390

Table 1: MSE between the 'violation percentage' and the nominal 'violation percentage' in simulation experiment. Both standard and censored M-H with cut-off of 20% and 30% are considered for all model specifications.

6 Empirical analysis

In the empirical analysis we consider univariate timeseries of daily returns corresponding to S&P500 stock market index recorded over a period of 3 years: 1st July 2008-30th June 2011. We consider a rolling sample of daily returns $1+i,\ldots,500+i$ with $i \in 1,\ldots,500$ to estimate the model and predict the observation $500+i$. The setup of the experiment is similar to the setup of simulation exercise, however a significantly broader spectrum of models is considered. Apart from simple GARCH models we also consider TGARCH(p,q,o) specifications with $p=1,\ldots,4$, $q=1,\ldots,4$ and $o=1,\ldots,4$. By definition the TGARCH specification encompasses the standard GARCH models what is equivalent with $o=4$. Altogether we consider $4*4*5=80$ distinct model specifications. The performance of individual models is compared with a set of Bayesian model averages.

Setup of empirical exrercise

1. Set $T = 500$.
2. Set $i = 0$.
3. Fix the estimation sample $1 + i, \dots, T/2 + i$ and estimate all the GARCH(p,q,o) models with $p = 1, \dots, 4$, $q = 1, \dots, 4$ and $o = 1, \dots, 4$, with the appropriate M-H sampling algorithm.
4. Order the models according to their Bayes factors. Consider 79 Bayesian model averages with $2, \dots, 80$ components constructed sequentially according to the ranking of models.
5. For every model specification and for every model average (in total $80 + 79$ specifications) construct the predictive distribution for one-day-ahead, i.e. $T/2 + i + 1$.
6. Estimate quantiles $1, \dots, 10$ of the predictive distribution.
7. Compare the actual (realized) return realized in $T/2 + i + 1$ with the quantiles of the predictive distribution and record the exceedances.
8. Return to step 3 and repeat steps $3 - 7$ with the estimation sample rolled one observation forward ($i := i + 1$). Repeat this procedure until the entire sample of available data is exhausted.
9. For each quantile considered count the exceedances over entire training sample $T/2+1, \dots, T$.
10. Compute the squared errors between the observed and nominal 'violation rates'. The average of these squared errors results in MSE that evaluates botj the individual

models and model avergaes.

These exerise is performed with censored M-H algorithm with cut-offs of 20 and 30% as well as with uncensored M-H algorithm. For each apporach we consider both standard and predictive approach to Bayes factors for Bayesian model averaging. It should be emphasized that in each period i the components of respective model averages can differ due to changing ranking of models with respect to Bayes factors. Apart form Bayesian model averages and individual model specifications in each period i we also consider the model with highest posterior probability (specification is varying from period to period according to the ranking of Bayes factors). The exceedances are recorded and the 'violation rate' is also computed for this specification.

In Table **??** the evaluation of individaul models with MSE criterion is presented. The individaul models are numbered $1-80$. We observe that the censored M-H sampling algorithm is able to outperfrom the uncensored M-H procedure for most individual specifications considered. However our primary interest concerns performance of model averages and comparison of predictive likelihood and standard likelihood approach in evalaution of Bayes factors for model selection. In Figure 7 we present the Bayes factors evaluated for all 80 models. For every model 500 evalautions of Bayes factors are plotted each for another interval i in line with setup of rolling window prediction experiment. The left panels of Figure 7 present the standard Bayes factors computed under diffuse flat prior. Irrespective of the M-H algorithm (censored or uncensored specification) the Bayes factors are clearly directly proportional to the number of components. It suggests that the model selection procedure might not be reliable based on this approach to evaluate the Bayes factors. It indicates the necessity of prior regularization. To that end in the right panel we present the Bayes factors under predictive likelihood. We observe that the number of variables in the model is no longer a key factor in evaluation of Bayes factor. Based on both standard and predictive Bayes factors in Tables 3 and 4, respectively, we present the evaluation of the model averaging specifications together with the highest posterior probability specifications. Due to abundancy of distinct model averaging specifications, ranging from 2 up to 80 components, information available in Table 3 and 4 is presented in Figure 8. The MSE criterion is presented as a function of the number of components in the Bayesian model averaging. The general picture suggests that the number of components in the Bayesian model average is an important factor heavily impacting the accuracy of predictive distribution. It is highly beneficial for the accuracy in left tail of predictive distribution. The MSE is negatively proportional to the number of components in the model average irrespectively of sampler and approach to compute the Bayes factors. We observe stable outperformance of uncensored M-H sampling algorithm by its censored counterparts irrespective of the cut-off tresholds considered. The results indicate that censoring is able to highly improve the accuracy of prediction in the left tail. It is particularly advantegous jointly with predictive likelihood approach to Bayes factors. The comparison of standard and predictive likelihood depicted in Figure 9 indicates the advantages of evaluation of Bayes factors with predictive likelihood. This method is able to outperform the standard likelihood approach irrespectively of the sampling algorithm applied. Figure 9 suggests that the advantage of predictive over standard likelihood is particularly pronounced in case of censored algorithm (top and middle panels of Figure 9). Generally it can be concluded that the censored approach is particularly advocated in combination with predictive likelihood in evalaution of model probabilites for Bayesian model averaging.

References

Anderson, T. W., (1984), *An Introduction to Multivariate Statistical Analysis*, New York: Wiley.

Bollerslev T., (1986) ,Generalized Autoregressive Conditional Heteroskedasticity, *Journal of Econometrics*, 31(3), 307-327.

Diks, C., Panchenko, V., Van Dijk, D., (2008), Partial Likelihood-Based Scoring Rules for Evaluating Density Forecasts in Tails,*Tinbergen Institute Discussion Series*, TI 2008-050/4.

Eklund, J., Karlsson, S., (2007), Forecast Combination and Model Averaging Using Predictive Measures, *Econometric Reviews*, 26 (2-4), 329-363.

Gelfand, A.E., Dey, D.K., (1994), Bayesian model choice: asymptotics and exact calculations, *J. Royal Statist. Soc. B*, 56(3), 501-514.

Hastings, W.K., (1970), Monte Carlo Sampling Methods Using Markov Chains and Their Applications, *Biometrika*, 57(1), 97109.

Hoogerheide L.F., (2010), Why should one use the predictive likelihood for computing model probabilities?, unpublished manuscript.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., Teller, E., (1973), Equation of state calculation by fast computing machines, *Journal of Chemical Physics*, 21(6), 10871092.

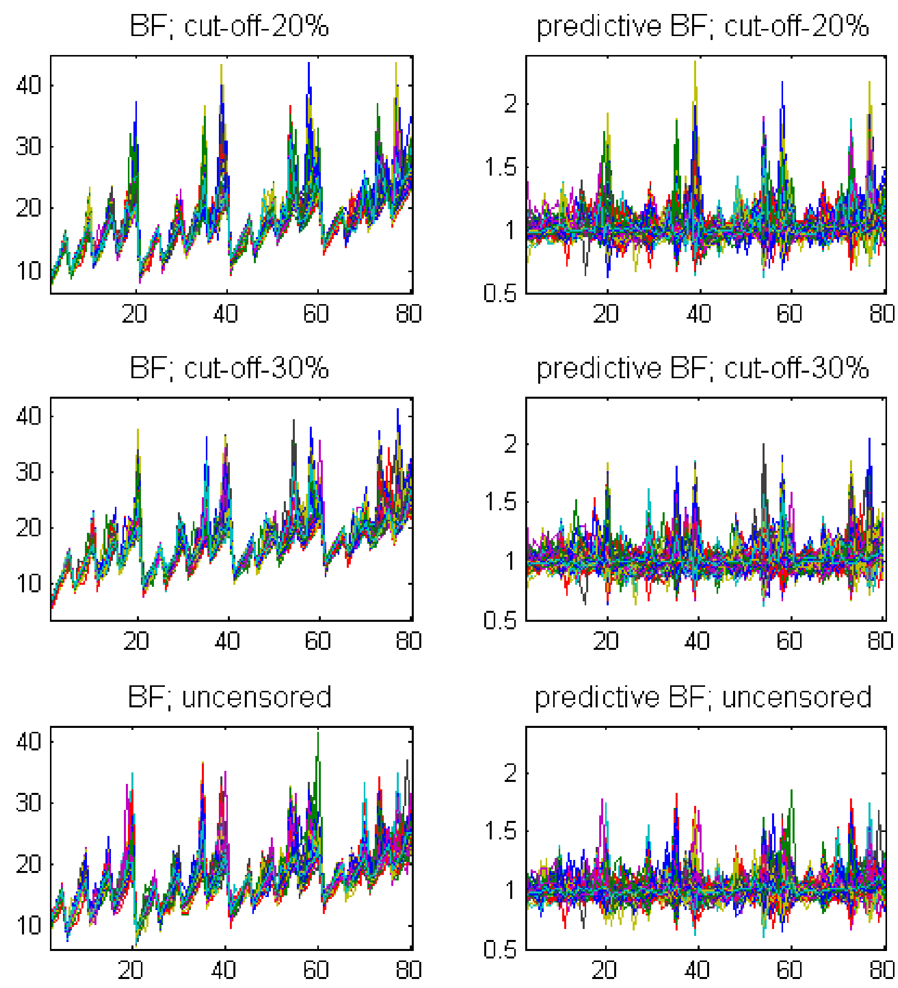

Figure 7: Bayes factors based on standard and predictive likelihood approach.

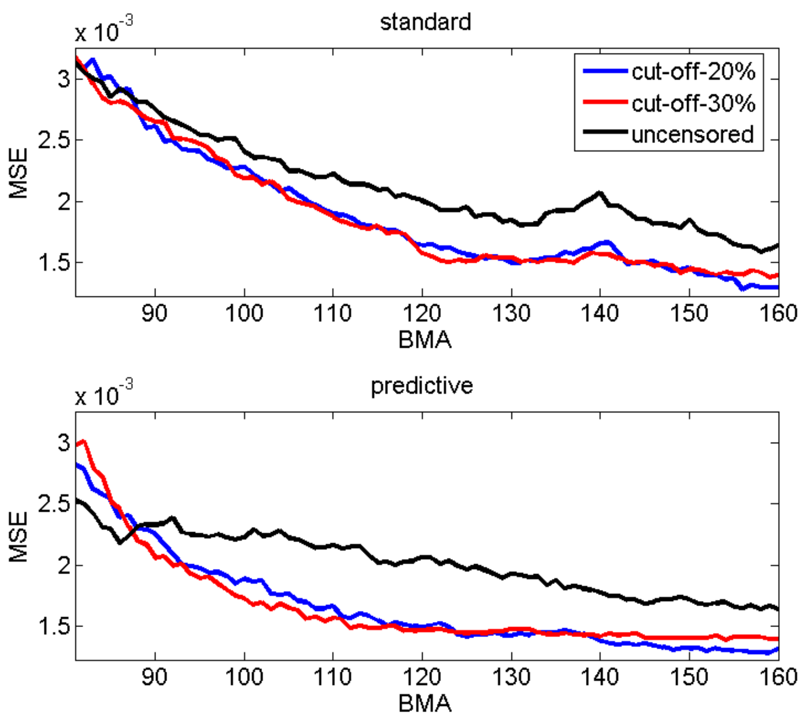

Figure 8: MSE between the observed 'violation percentage' and the nominal 'violation percentage' for BMA (top panel) and predictive BMA (bottom panel) specifications.

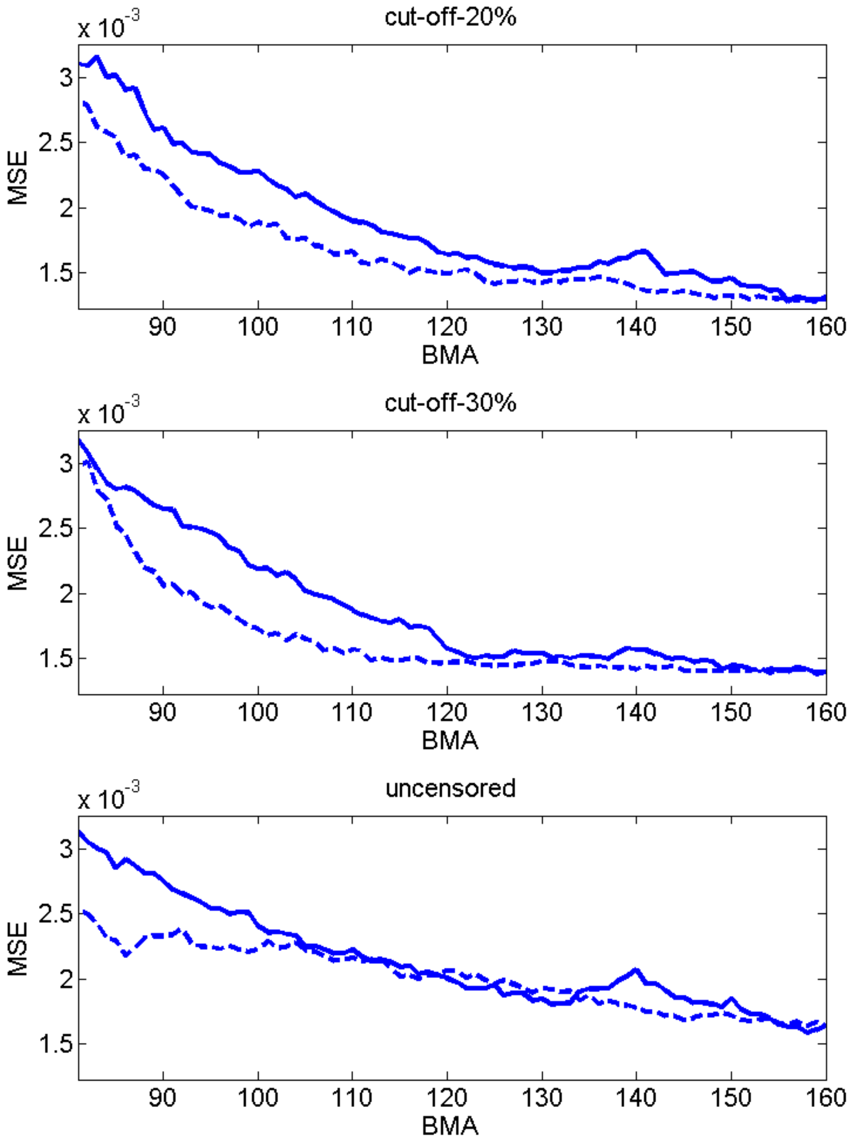

Figure 9: Comparison of standard (solid line) and predictive (dashed line) BMA in terms of MSE.

no.	model	20%	30%	uncens.	no.	model	20%	30%	uncens.
1	TGARCH(1,1,0)	0.0021	0.00535	0.0001	41	TGARCH(3,1,0)	0.0165	0.02143	0.01009
2	TGARCH(1,1,1)	5e-005	0.0001	4e-005	42	TGARCH(3,1,1)	0.00426	0.0041	0.0029
3	TGARCH(1,1,2)	0.00298	0.00414	0.00596	43	TGARCH(3,1,2)	0.00195	0.0013	0.00058
4	TGARCH(1,1,3)	0.00013	0.00023	0.00073	44	TGARCH(3,1,3)	0.00431	0.00681	0.00223
5	TGARCH(1,1,4)	0.0003	0.00025	5e-005	45	TGARCH(3,1,4)	0.00269	0.00454	0.00122
6	TGARCH(1,2,0)	0.00017	0.00022	0.00161	46	TGARCH(3,2,0)	0.00042	0.00059	1e-005
7	TGARCH(1,2,1)	0.00028	0.00027	0.00069	47	TGARCH(3,2,1)	0.00135	0.00146	0.002
8	TGARCH(1,2,2)	4e-005	4e-005	4e-005	48	TGARCH(3,2,2)	0.00026	0.00021	0.00032
9	TGARCH(1,2,3)	0.00211	0.0021	0.00202	49	TGARCH(3,2,3)	0.00034	0.00027	0.00052
10	TGARCH(1,2,4)	0.00021	0.00049	0.00097	50	TGARCH(3,2,4)	0.00047	0.00053	0.00115
11	TGARCH(1,3,0)	0.00122	0.00095	0.00153	51	TGARCH(3,3,0)	0.00316	0.00285	0.00324
12	TGARCH(1,3,1)	0.00105	0.00126	0.00222	52	TGARCH(3,3,1)	0.00127	0.00118	0.00112
13	TGARCH(1,3,2)	0.00262	0.00316	0.00332	53	TGARCH(3,3,2)	0.00133	0.00151	0.00146
14	TGARCH(1,3,3)	0.00133	0.0013	0.00169	54	TGARCH(3,3,3)	0.00232	0.00256	0.00216
15	TGARCH(1,3,4)	0.00178	0.00145	0.00225	55	TGARCH(3,3,4)	0.00152	0.00159	0.00182
16	TGARCH(1,4,0)	0.00178	0.0017	0.00222	56	TGARCH(3,4,0)	0.00264	0.0023	0.00234
17	TGARCH(1,4,1)	0.00345	0.00341	0.00349	57	TGARCH(3,4,1)	0.00282	0.00283	0.00265
18	TGARCH(1,4,2)	0.00249	0.0025	0.00287	58	TGARCH(3,4,2)	0.00275	0.00263	0.00278
19	TGARCH(1,4,3)	0.00263	0.00251	0.0031	59	TGARCH(3,4,3)	0.00183	0.0019	0.0019
20	TGARCH(1,4,4)	0.00273	0.00259	0.00317	60	TGARCH(3,4,4)	0.00313	0.00317	0.00305
21	TGARCH(2,1,0)	0.00693	0.01455	0.00048	61	TGARCH(4,1,0)	0.00896	0.01133	0.00127
22	TGARCH(2,1,1)	0.01163	0.01608	0.00565	62	TGARCH(4,1,1)	0.00194	0.0027	0.00014
23	TGARCH(2,1,2)	0.0079	0.00922	0.00082	63	TGARCH(4,1,2)	0.00695	0.00586	0.00174
24	TGARCH(2,1,3)	0.00182	0.0019	0.0001	64	TGARCH(4,1,3)	0.00329	0.00516	0.00108
25	TGARCH(2,1,4)	0.00588	0.00587	0.00239	65	TGARCH(4,1,4)	0.00295	0.00541	0.00187
26	TGARCH(2,2,0)	0.00019	0.00087	0.0003	66	TGARCH(4,2,0)	0.00113	0.0007	0.00175
27	TGARCH(2,2,1)	8e-005	0.00044	2e-005	67	TGARCH(4,2,1)	0.00044	0.00021	0.00076
28	TGARCH(2,2,2)	0.00086	0.00122	0.00189	68	TGARCH(4,2,2)	0.00012	0.00019	0.00061
29	TGARCH(2,2,3)	0.00036	0.0002	0.00062	69	TGARCH(4,2,3)	0.00053	0.00033	0.0012
30	TGARCH(2,2,4)	0.00013	9e-005	0.00061	70	TGARCH(4,2,4)	0.00031	0.00017	0.00069
31	TGARCH(2,3,0)	0.0015	0.0014	0.00171	71	TGARCH(4,3,0)	0.00117	0.00103	0.00136
32	TGARCH(2,3,1)	0.00318	0.00298	0.00354	72	TGARCH(4,3,1)	0.00153	0.00122	0.00169
33	TGARCH(2,3,2)	0.00106	0.00104	0.00124	73	TGARCH(4,3,2)	0.00209	0.0022	0.00251
34	TGARCH(2,3,3)	0.00134	0.0014	0.0018	74	TGARCH(4,3,3)	0.00133	0.00123	0.00149
35	TGARCH(2,3,4)	0.00234	0.0026	0.00253	75	TGARCH(4,3,4)	0.0024	0.00242	0.00281
36	TGARCH(2,4,0)	0.00355	0.00351	0.00337	76	TGARCH(4,4,0)	0.0029	0.00276	0.00305
37	TGARCH(2,4,1)	0.00262	0.00238	0.0025	77	TGARCH(4,4,1)	0.00317	0.00278	0.00286
38	TGARCH(2,4,2)	0.00245	0.00254	0.0029	78	TGARCH(4,4,2)	0.00195	0.00206	0.00222
39	TGARCH(2,4,3)	0.00279	0.00266	0.00295	79	TGARCH(4,4,3)	0.00291	0.00319	0.00306
40	TGARCH(2,4,4)	0.00206	0.0022	0.00247	80	TGARCH(4,4,4)	0.00303	0.00287	0.00315

Table 2: MSE between the observed and the nominal 'violation percentage' for 80 different TGARCH specifications for S& P500 returns.

no.	model	20%	30%	uncens.	no.	model	20%	30%	uncens.
81	h.p.p.	0.00311	0.00318	0.00313	121	BMA(41)	0.00165	0.00155	0.00197
82	BMA(2)	0.00309	0.00309	0.00305	122	BMA(42)	0.00161	0.00151	0.00193
83	BMA(3)	0.00317	0.00296	0.003	123	BMA(43)	0.00162	0.0015	0.00193
84	BMA(4)	0.00301	0.00285	0.00298	124	BMA(44)	0.00158	0.00152	0.00193
85	BMA(5)	0.00302	0.0028	0.00285	125	BMA(45)	0.00156	0.00151	0.00195
86	BMA(6)	0.0029	0.00282	0.00292	126	BMA(46)	0.00155	0.00151	0.00187
87	BMA(7)	0.00292	0.00279	0.00287	127	BMA(47)	0.00153	0.00156	0.00189
88	BMA(8)	0.00273	0.00273	0.00282	128	BMA(48)	0.00154	0.00154	0.00189
89	BMA(9)	0.00259	0.00268	0.00282	129	BMA(49)	0.00153	0.00154	0.00183
90	BMA(10)	0.00261	0.00265	0.00275	130	BMA(50)	0.0015	0.00154	0.00185
91	BMA(11)	0.00249	0.00265	0.00269	131	BMA(51)	0.0015	0.0015	0.0018
92	BMA(12)	0.0025	0.00252	0.00265	132	BMA(52)	0.00151	0.00151	0.00181
93	BMA(13)	0.00243	0.00251	0.00262	133	BMA(53)	0.00151	0.0015	0.00181
94	BMA(14)	0.00242	0.0025	0.00259	134	BMA(54)	0.00154	0.00151	0.0019
95	BMA(15)	0.00241	0.00247	0.00254	135	BMA(55)	0.00154	0.00151	0.00192
96	BMA(16)	0.00234	0.00243	0.00254	136	BMA(56)	0.00158	0.00151	0.00192
97	BMA(17)	0.00232	0.00235	0.0025	137	BMA(57)	0.00156	0.00149	0.00192
98	BMA(18)	0.00227	0.00232	0.00252	138	BMA(58)	0.0016	0.00154	0.00198
99	BMA(19)	0.00227	0.00221	0.00251	139	BMA(59)	0.00161	0.00158	0.00202
100	BMA(20)	0.00228	0.00219	0.0024	140	BMA(60)	0.00166	0.00156	0.00207
101	BMA(21)	0.00222	0.00219	0.00236	141	BMA(61)	0.00166	0.00156	0.00196
102	BMA(22)	0.00217	0.00214	0.00236	142	BMA(62)	0.00158	0.00153	0.00196
103	BMA(23)	0.00214	0.00216	0.00235	143	BMA(63)	0.00149	0.00151	0.00191
104	BMA(24)	0.00208	0.00211	0.00234	144	BMA(64)	0.00149	0.00149	0.00186
105	BMA(25)	0.0021	0.00202	0.00225	145	BMA(65)	0.0015	0.00149	0.00186
106	BMA(26)	0.00205	0.00199	0.00225	146	BMA(66)	0.00151	0.00147	0.00181
107	BMA(27)	0.00201	0.00197	0.00222	147	BMA(67)	0.00146	0.00148	0.00181
108	BMA(28)	0.00197	0.00195	0.00219	148	BMA(68)	0.00143	0.00147	0.0018
109	BMA(29)	0.00192	0.00191	0.00219	149	BMA(69)	0.00143	0.00141	0.00178
110	BMA(30)	0.00189	0.00187	0.00223	150	BMA(70)	0.00146	0.00144	0.00185
111	BMA(31)	0.00188	0.00183	0.00217	151	BMA(71)	0.0014	0.00144	0.00177
112	BMA(32)	0.00186	0.00181	0.00214	152	BMA(72)	0.00139	0.00141	0.00173
113	BMA(33)	0.00181	0.00178	0.00214	153	BMA(73)	0.00139	0.0014	0.00173
114	BMA(34)	0.0018	0.00177	0.00213	154	BMA(74)	0.00136	0.00141	0.0017
115	BMA(35)	0.00178	0.00179	0.00209	155	BMA(75)	0.00136	0.0014	0.00165
116	BMA(36)	0.00176	0.00174	0.0021	156	BMA(76)	0.00128	0.0014	0.00163
117	BMA(37)	0.00176	0.00175	0.00203	157	BMA(77)	0.00131	0.00143	0.00163
118	BMA(38)	0.00172	0.00173	0.00206	158	BMA(78)	0.00129	0.00142	0.00158
119	BMA(39)	0.00166	0.00165	0.00203	159	BMA(79)	0.00129	0.00138	0.0016
120	BMA(40)	0.00164	0.00157	0.002	160	BMA(80)	0.00129	0.00139	0.00164

Table 3: Mean of mean squared error statistics between the observed 'violation percentage' and the nominal 'violation percentage' for DAX index: the highest posterior probability specification (h.p.p.) and 79 BMAs. BMA(k) refers to averaging over k models.

no.	model	20%	30%	uncens.	no.	model	20%	30%	uncens.
81	h.p.p.	0.00282	0.00298	0.00254	121	BMA(41)	0.00151	0.00147	0.00206
82	BMA(2)	0.00279	0.00302	0.0025	122	BMA(42)	0.00152	0.00147	0.00201
83	BMA(3)	0.00262	0.00279	0.00243	123	BMA(43)	0.00149	0.00146	0.00204
84	BMA(4)	0.00258	0.00272	0.00231	124	BMA(44)	0.00143	0.00145	0.00199
85	BMA(5)	0.00254	0.00251	0.0023	125	BMA(45)	0.00141	0.00144	0.00196
86	BMA(6)	0.00239	0.00246	0.00217	126	BMA(46)	0.00143	0.00145	0.00199
87	BMA(7)	0.00241	0.00231	0.00224	127	BMA(47)	0.00143	0.00144	0.00196
88	BMA(8)	0.0023	0.0022	0.00231	128	BMA(48)	0.00144	0.00145	0.00193
89	BMA(9)	0.00229	0.00217	0.00233	129	BMA(49)	0.00142	0.00145	0.00189
90	BMA(10)	0.00225	0.00206	0.00233	130	BMA(50)	0.00142	0.00147	0.00193
91	BMA(11)	0.00217	0.00207	0.00234	131	BMA(51)	0.00144	0.00147	0.00192
92	BMA(12)	0.00209	0.002	0.00239	132	BMA(52)	0.00142	0.00147	0.0019
93	BMA(13)	0.002	0.00201	0.00226	133	BMA(53)	0.00145	0.00145	0.0019
94	BMA(14)	0.00199	0.00193	0.00225	134	BMA(54)	0.00145	0.00143	0.00183
95	BMA(15)	0.00198	0.00189	0.00225	135	BMA(55)	0.00145	0.00143	0.00188
96	BMA(16)	0.00193	0.00191	0.00223	136	BMA(56)	0.00146	0.00144	0.0018
97	BMA(17)	0.00194	0.00184	0.00225	137	BMA(57)	0.00145	0.00142	0.00183
98	BMA(18)	0.00191	0.0018	0.00222	138	BMA(58)	0.00143	0.00142	0.0018
99	BMA(19)	0.00185	0.00175	0.0022	139	BMA(59)	0.00141	0.00143	0.00179
100	BMA(20)	0.00189	0.00172	0.00222	140	BMA(60)	0.00137	0.00141	0.00177
101	BMA(21)	0.00186	0.00167	0.00229	141	BMA(61)	0.00136	0.00144	0.00175
102	BMA(22)	0.00187	0.00169	0.00224	142	BMA(62)	0.00135	0.00143	0.00175
103	BMA(23)	0.00176	0.00163	0.00224	143	BMA(63)	0.00136	0.00143	0.00171
104	BMA(24)	0.00175	0.00168	0.00228	144	BMA(64)	0.00134	0.00144	0.00171
105	BMA(25)	0.00176	0.00165	0.00222	145	BMA(65)	0.00136	0.0014	0.00168
106	BMA(26)	0.0017	0.00163	0.00221	146	BMA(66)	0.00134	0.0014	0.0017
107	BMA(27)	0.0017	0.00156	0.00217	147	BMA(67)	0.00132	0.0014	0.00172
108	BMA(28)	0.00164	0.00158	0.00214	148	BMA(68)	0.00131	0.0014	0.00172
109	BMA(29)	0.00164	0.00154	0.00214	149	BMA(69)	0.00132	0.0014	0.00174
110	BMA(30)	0.00166	0.00157	0.00216	150	BMA(70)	0.00132	0.0014	0.00172
111	BMA(31)	0.00157	0.00155	0.00214	151	BMA(71)	0.00132	0.0014	0.00168
112	BMA(32)	0.00157	0.00148	0.00215	152	BMA(72)	0.00129	0.0014	0.00167
113	BMA(33)	0.0016	0.0015	0.00215	153	BMA(73)	0.00131	0.00142	0.00169
114	BMA(34)	0.00158	0.00148	0.00208	154	BMA(74)	0.0013	0.00138	0.0017
115	BMA(35)	0.00154	0.00148	0.00202	155	BMA(75)	0.0013	0.00141	0.00166
116	BMA(36)	0.0015	0.0015	0.00203	156	BMA(76)	0.0013	0.00141	0.00165
117	BMA(37)	0.00153	0.0015	0.002	157	BMA(77)	0.00128	0.00141	0.00167
118	BMA(38)	0.00151	0.00146	0.00203	158	BMA(78)	0.00128	0.00141	0.00164
119	BMA(39)	0.00149	0.00146	0.00203	159	BMA(79)	0.00128	0.00139	0.00167
120	BMA(40)	0.00149	0.00146	0.00206	160	BMA(80)	0.00131	0.00139	0.00164

Table 4: Mean of mean squared error statistics between the observed 'violation percentage' and the nominal 'violation percentage' for DAX index: the highest posterior probability specification (h.p.p.) and 79 predictive BMAs. Predictive BMA(k) refers to averaging over k models.