



Density prediction of stock index returns using GARCH models: Frequentist or Bayesian estimation?

Lennart F. Hoogerheide^a, David Ardia^{b,*}, Nienke Corré^c

^a Department of Econometrics, Vrije Universiteit Amsterdam, The Netherlands

^b aeris CAPITAL AG, Switzerland

^c Bain & Company, Netherlands, LLC, The Netherlands

ARTICLE INFO

Article history:

Received 31 January 2011

Received in revised form

28 February 2012

Accepted 27 March 2012

Available online 3 April 2012

JEL classification:

C2
C11
C13

Keywords:

GARCH
Bayesian
KLIC
Censored likelihood

ABSTRACT

Using GARCH models for density prediction of stock index returns, a comparison is provided between frequentist and Bayesian estimation. No significant difference is found between qualities of whole density forecasts, whereas the Bayesian approach exhibits significantly better left-tail forecast accuracy.

© 2012 Elsevier B.V. All rights reserved.

1. Introduction

The modeling of volatility has been at the heart of financial econometrics for decades. One of its corner-stones is the GARCH model of Bollerslev (1986). Popular extensions include the Exponential GARCH (EGARCH) model by Nelson (1991) and the GJR model by Glosten et al. (1993).

The common approach for estimation in GARCH-type models is Maximum Likelihood (ML). The use of ML is appealing because of the intuitive principle underlying the approach and its ease of implementation. Nevertheless, ML estimation entails some difficulties in practice due to the numerical optimization which involves inequality constraints. The Bayesian approach offers an attractive alternative which enables small sample results, robust estimation, model discrimination and probabilistic statements on (possibly nonlinear functions of) model parameters (see Ardia, 2008).

* Correspondence to: Churerstrasse 70, CH 8808 Pfäffikon SZ, Switzerland. Tel.: +41 555 111 222.

E-mail addresses: l.f.hoogerheide@vu.nl (L.F. Hoogerheide), david.ardia@tolomeo-capital.com, david.ardia@unifr.ch (D. Ardia), nienke.corre@bain.com (N. Corré).

The comparison of frequentist and Bayesian GARCH has received little attention in research. In a study by Nakatsuma and Tsurumi (1996) some attention is paid to the comparison. Based on a small sample Monte Carlo experiment, they conclude that the Bayesian approach performs better than the frequentist approach in terms of smaller mean square errors of the posterior mean versus the MLE of the parameters in an ARMA–GARCH model. Their comparison is however not extended to an empirical application or to the evaluation of density prediction of returns.

This note aims at comparing the performance of well-known GARCH models via frequentist and Bayesian approaches. We evaluate the forecasting performance of the GARCH, GJR and EGARCH model using index data. The performance of the models is evaluated based on the relative predictive accuracy; both the entire density and the left tail, which is more relevant for risk management purposes, are considered. Our results indicate that the Bayesian approach outperforms its frequentist counterpart when forecasting the left-tail of the return's distribution.

2. Frequentist and Bayesian estimation of GARCH models

We consider the GARCH(1,1), GJR(1,1) and EGARCH(1,1) models, with Student-*t* innovations to account for conditional excess

kurtosis (see Geweke and Amisano, 2010). The mean equation is defined as

$$y_t = \mu + \varepsilon_t \quad t = 1, 2, \dots, T$$

$$\varepsilon_t | I_{t-1} \sim \mathcal{S} \left(0, h_t \frac{\nu - 2}{\nu}, \nu \right), \quad (1)$$

with y_t the log-return at time t and where \mathcal{S} is used for the Student- t distribution. The sequence of innovations, given past information I_{t-1} , follows a Student- t distribution with mean zero, variance h_t and ν degrees of freedom. The restriction $\nu > 2$ guarantees existence of the conditional variance.

In the GARCH(1,1) model the conditional variance h_t is a linear function of the past squared innovation and conditional variance:

$$h_t = \omega + \alpha \varepsilon_{t-1}^2 + \beta h_{t-1}. \quad (2)$$

The GJR(1,1) model is similar, except for the indicator variable added to allow for asymmetric effects depending on the sign of the innovation (i.e., typically to allow for a larger effect of a negative shock than a positive shock of the same magnitude):

$$h_t = \omega + (\alpha_1 \mathbb{I}_{\{\varepsilon_{t-1} \geq 0\}} + \alpha_2 \mathbb{I}_{\{\varepsilon_{t-1} < 0\}}) \varepsilon_{t-1}^2 + \beta h_{t-1}. \quad (3)$$

The indicator variable $\mathbb{I}_{\{\bullet\}}$ takes on a value of one if the condition holds and zero otherwise. Finally, the EGARCH(1,1) model is given by:

$$\log(h_t) = \omega + \alpha \frac{|\varepsilon_{t-1}|}{\sqrt{h_{t-1}}} + \gamma \frac{\varepsilon_{t-1}}{\sqrt{h_{t-1}}} + \beta \log(h_{t-1}), \quad (4)$$

where the term $\varepsilon_{t-1}/\sqrt{h_{t-1}}$ is known as the standardized residual; the occurrence of this term without $|\cdot|$ allows for the asymmetric behavior of the conditional variance.

For the GARCH and GJR models the positivity constraints imposed on the model parameters in the variance equation are $\omega > 0, \alpha, \beta \geq 0$.

In what follows, we regroup the model parameters into θ and the observed log-returns into $y = (y_1, \dots, y_T)'$. The likelihood function is given by $p(y|\theta)$ which is the product of the conditional density values at y_t .

In the frequentist approach θ is assumed to be a fixed unknown constant, which is typically estimated by ML. The Bayesian approach incorporates the likelihood in a broader framework. Parameters in θ are assumed to be stochastic variables described by a posterior density $p(\theta|y)$ given by Bayes' theorem as

$$p(\theta|y) = \frac{p(y|\theta)p(\theta)}{\int p(y|\theta)p(\theta)d\theta}, \quad (5)$$

where $p(\theta)$ is the prior density for θ , reflecting the prior beliefs before having observed the data. For the parameters μ, ω, α and β we use non-informative flat (uniform) prior on the parameter domain. For ν , we use a non-informative proper prior (as an improper prior would yield an improper posterior, see Bauwens and Lubrano (1998)): a translated exponential density as in Geweke (1993); $\nu \sim \lambda \exp(-\lambda(\nu - 2))$ with $\lambda = 0.05$. We adopt an independence chain Metropolis–Hastings algorithm (with Student- t candidate distribution around the posterior mode, with scale matrix equal to minus the inverse Hessian of the log-posterior at the mode, and four degrees of freedom for fat tails) to draw random samples from the posterior density: 25,000 draws after a burn-in of 5000 draws. Acceptance rates were always around 50%, indicating the good performance of the posterior simulator. Prediction results under a more flexible candidate distribution, the mixture of Student- t distributions of Hoogerheide et al. (2007), were qualitatively the same. Further, results stay almost the same for somewhat different values of λ .

3. Tests for difference in forecasting performance

This research focuses on one-day-ahead density forecasts for daily returns rather than point forecasts. These density forecasts provide an estimate of the probability distribution of future returns. Our objective is to find the density forecast that comes closest to the true but unobserved density for daily returns by comparing the relative performance of two competing density forecasts denoted by $\hat{f}_t(y_{t+1})$ and $\hat{g}_t(y_{t+1})$. Given (1) the forecast densities can be expressed as

$$y_{t+1} | \theta, \mathcal{M}, I_t \sim \mathcal{S}(\mu, h_{t+1}, \nu), \quad (6)$$

where the forecast density is conditional on θ and model \mathcal{M} . In the frequentist approach we simply use the ML estimator $\hat{\theta}_{MLE}$, implicitly ignoring parameter uncertainty. On the other hand, the Bayesian approach takes into account parameter uncertainty, which is integrated out: the Bayesian density forecast is obtained as the mean of the density forecasts corresponding to the posterior draws.

To compare the performance of the different forecast densities we follow Mitchel and Hall (2005) and Diks et al. (2008) and use the well-known Kullback–Leibler information criterion (KLIC), see Amisano and Giacomini (2007). Using the KLIC a loss differential series is constructed with $d_{t+1} = \ln \hat{g}_t(y_{t+1}) - \ln \hat{f}_t(y_{t+1})$, which can be interpreted as the difference in the score of both models. Obviously, a high score is preferred as this implies that the model assigns a high 'probability' to the observed value. Given a sequence of density forecasts and corresponding realizations for the observed return, densities are compared based on the average score \bar{d} . Under the null hypothesis of equal forecast accuracy the loss differential should be zero on average. This can be tested by a Diebold and Mariano (1995) type test statistic with robust Newey–West standard error.

We also consider the evaluation of forecast performance in the left-tail of the forecast density, which is of special interest for risk management applications. For testing, we rely on the censored likelihood (CSL) scoring rule by Diks et al. (2008). For a particular model the CSL equals the total non-tail probability $(1 - \hat{F}_t(r))$ if y_{t+1} is not a tail event and the forecasted density value $\hat{f}_t(y_{t+1})$ if y_{t+1} is a tail event. Hence the CSL score function is obtained as

$$CSL(\hat{f}_t | y_{t+1}) = \mathbb{I}_{\{y_{t+1} \leq r\}} \ln \hat{f}_t(y_{t+1}) + \mathbb{I}_{\{y_{t+1} > r\}} \ln(1 - \hat{F}_t(r)), \quad (7)$$

where $\hat{F}_t(r)$ is the Student- t cumulative density at time t evaluated at the threshold r . This term implies that the shape of the forecasted density is ignored for observations outside the region of interest.

To evaluate the CSL of competing models a procedure similar to the evaluation procedure based on KLIC scores is employed. This left-tail test is a natural extension of the well-known KLIC-based test for the whole density forecasts (for $r \rightarrow \infty$ it obviously reduces to the same test). The CDF term $\mathbb{I}_{\{y_{t+1} > r\}} \ln(1 - \hat{F}_t(r))$ is required, since its exclusion would yield a criterion that would always favor density forecasts with a fatter left-tail, even if the true density were thin-tailed like the normal distribution.

Note that a proper threshold value r needs to be specified, determining which observations pertain to the left tail. For a fair evaluation, the number of pdf-terms and CDF-terms included in the CSL should be the same across all competing models such that a model-independent threshold is required. Though the threshold is allowed to be time-varying, the threshold is fixed at -2.5% in this study, which is the 7.7% to 8.8% quantile of the unconditional return distribution during the out-of-sample period for the S&P 500 and Nikkei 225.

Table 1

Diebold and Mariano (1995) *t*-statistic (using Newey–West standard error) corresponding to average loss differential for KLIC (whole density) or CSL (left tail of density), based on 600 density forecasts of the GARCH(1,1), GJR(1,1) and EGARCH(1,1) models for daily log-returns of the S&P 500 and Nikkei 225. Positive (negative) values signify better forecast performance of the model specified in the corresponding row (column).

		Frequentist (MLE)			Bayesian (MH)		
		GARCH	GJR	EGARCH	GARCH	GJR	EGARCH
KLIC							
S&P 500							
Frequentist (MLE)	GARCH		−2.632*	−1.602	1.007		
	GJR	2.632*		0.409		1.685	
	EGARCH	1.602	−0.409				−0.881
Bayesian (MH)	GARCH	−1.007				−2.376*	−4.053**
	GJR		−1.685		2.376*		−2.732**
	EGARCH			0.881	4.053**	2.732**	
Nikkei 225							
Frequentist (MLE)	GARCH		−2.622**	−0.477	−0.041		
	GJR	2.622**		0.220		0.373	
	EGARCH	0.477	−0.220				−0.494
Bayesian (MH)	GARCH	0.041				−2.033*	−2.011*
	GJR		−0.373		2.033*		−0.408
	EGARCH			0.494	2.011*	0.408	
CSL							
S&P 500							
Frequentist (MLE)	GARCH		−2.143*	−0.373	−5.274**		
	GJR	2.143*		1.454		−5.480**	
	EGARCH	0.373	−1.454				−4.227**
Bayesian (MH)	GARCH	5.274**				−1.686	−0.798
	GJR		5.480**		1.686		0.246
	EGARCH			4.227**	0.798	−0.246	
Nikkei 225							
Frequentist (MLE)	GARCH		−1.506	−0.007	−5.277**		
	GJR	1.506		0.243		−5.054**	
	EGARCH	0.007	−0.243				−5.422**
Bayesian (MH)	GARCH	5.277**				−1.743	−0.875
	GJR		5.054**		1.743		0.916
	EGARCH			5.422**	0.875	−0.916	

* Two-sided significance at 5%.

** Two-sided significance at 1%.

4. Results and conclusions

Log-returns of the S&P 500 and Nikkei 225 stock indices are examined. Models are estimated using a ten-year rolling window. The out-of-sample period consists of 600 daily forecasts, ranging from January 1, 2008, to April 19, 2010.

The top panel of Table 1 reports the results of the pairwise forecast accuracy comparison of competing density forecasts as measured by the KLIC. Superior predictive accuracy of the model/method in the row (as compared with the model/method in the column) is indicated by a positive value.

Our primary interest lies in the comparison of the frequentist and Bayesian approach. Across all models and data sets the test statistic is not significant, implying that the null hypothesis of equal predictive accuracy of the frequentist and Bayesian approach cannot be rejected.

The testing approach also offers a straightforward framework to compare the predictive accuracy of the different GARCH-type models. Note that the Bayesian approach is more able to signal significant differences across different model specifications. Only for the Bayesian density forecasts, the EGARCH model significantly outperforms the standard GARCH model (for both data sets). For the S&P 500 data, the Bayesian density predictions even yield a significant difference between the somewhat similar GJR and EGARCH models.

The bottom panel of Table 1 reports results for the left-tail forecast accuracy quantified by the CSL. The null hypothesis of equal predictive accuracy of the frequentist and Bayesian approach in the left tail is convincingly rejected for all models and data sets in favor of the Bayesian approach. The difference is mainly caused by the first term on the right-hand side of (7). Our finding makes

intuitively sense: Bayesian inference has been often found to be more appropriate in small samples, where it is more important that the present parameter uncertainty is taken into account. Although we have more than 2500 observations, we still have few observations from the (extreme) left tail. Especially the degrees of freedom parameter ν may be strongly influenced by a few extreme observations in the estimation window. In addition, note that the differences in CSL between the models are less pronounced (if any difference exists) than the differences between frequentist and Bayesian estimation, which further stresses the clearness of the superiority of Bayesian inference in this application.

We have also computed the 95% Value at Risk (VaR) forecasts from the different methods. There is no significant difference between the VaR coverage percentages, as considered by Christoffersen (1998), from the frequentist and Bayesian approach. This is arguably caused by the huge reduction of the test's power when using only a 0/1 variable indicating whether the return exceeds the predicted percentile instead of a (censored) density.

Our results suggest that in risk management applications the Bayesian approach should be preferred, because of its superior predictive accuracy in the left tail. This result is remarkable, since we use large data sets (over 2500 observations) and non-informative priors. For smaller data sets or informative priors the difference between Bayesian and frequentist inference may be even larger. In future research we will focus on different data sets and priors, alternative model specifications (e.g., different error distributions such as the skewed *t* distribution), model combination (see Ardia and Hoogerheide, 2010), and multi-step-ahead forecasting. In the latter case, the importance sampling method of Hoogerheide and Van Dijk (2010) may provide a substantial reduction of the computing time.

Acknowledgments

The authors acknowledge Guido Bolliger, Michel Dubois, Dusan Isakov, Michael Rockinger, Dick van Dijk and an anonymous referee for useful comments. The views expressed in this paper are the sole responsibility of the authors and do not necessarily reflect those of Bain & Company, aeris CAPITAL AG, or any of their affiliates. Any remaining errors or shortcomings are the authors' responsibility.

References

- Amisano, G., Giacomini, R., 2007. Comparing density forecasts via weighted likelihood ratio tests. *Journal of Business and Economic Statistics* 25, 177–190.
- Ardia, D., 2008. Financial Risk Management with Bayesian Estimation of GARCH Models: Theory and Applications. In: *Lecture Notes in Economics and Mathematical Systems*, vol. 612. Springer-Verlag, Berlin, Germany.
- Ardia, D., Hoogerheide, L.F., 2010. Efficient Bayesian estimation and combination of GARCH-type models. In: Böcker, K. (Ed.), *Rethinking Risk Measurement and Reporting – Volume II: Examples and Applications from Finance*. Risk Books, London (Chapter 1).
- Bauwens, L., Lubrano, M., 1998. Bayesian inference on GARCH models using the Gibbs sampler. *Econometrics Journal* 1, C23–C46.
- Bollerslev, T., 1986. Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* 31 (3), 307–327.
- Christoffersen, P.F., 1998. Evaluating interval forecasts. *International Economic Review* 39 (4), 841–862.
- Diebold, F.X., Mariano, R.S., 1995. Comparing predictive accuracy. *Journal of Business & Economic Statistics* 13 (3), 253–263.
- Diks, C., Panchenko, V., Van Dijk, D., 2008. Partial likelihood-based scoring rules for evaluating density forecasts in tails. Tinbergen Institute Discussion Paper TI 2008-050/4.
- Geweke, J.F., 1993. Bayesian treatment of the independent Student-t linear model. *Journal of Applied Econometrics* 8 (S1), S19–S40.
- Geweke, J., Amisano, G., 2010. Comparing and evaluating Bayesian predictive distributions of asset returns. *International Journal of Forecasting* 26, 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.
- Hoogerheide, L.F., Kaashoek, J.F., Van Dijk, H.K., 2007. On the shape of posterior densities and credible sets in instrumental variable regression models with reduced rank: an application of flexible sampling methods using neural networks. *Journal of Econometrics* 139 (1), 154–180.
- Hoogerheide, L.F., Van Dijk, H.K., 2010. Bayesian forecasting of Value at Risk and Expected Shortfall using adaptive importance sampling. *International Journal of Forecasting* 26, 231–247.
- Mitchel, J., Hall, S.G., 2005. Evaluating, comparing and combining density forecasts using the KLIC with an application to the Bank of England and NIESR 'fan' charts of inflation. *Oxford Bulletin of Economics and Statistics* 67, 995–1033.
- Nakatsuma, T., Tsurumi, H., 1996. ARMA-GARCH models: Bayes Estimation Versus MLE, and Bayes Non-stationarity Test. Working Paper No. 9619. Department of Economics, Rutgers University, New Jersey.
- Nelson, D.B., 1991. Conditional heteroskedasticity in asset returns: a new approach. *Econometrica* 59 (2), 347–370.