Bayes factors for Assessing Dynamic Quantile Forecasts

Richard Gerlach[1]*, and Cathy W.S. Chen[2]

[1] The University of Sydney Business School, Australia.

[2] Department of Statistics, Feng Chia University, Taiwan.

Abstract

This paper proposes Bayesian evaluation and Bayes factor methods for assessing dynamic forecasts of quantile levels. Bayes factor analogues of several popular frequentist tests for independence and correct coverage of quantile forecasts are developed. Multivariate quadrature methods and the standard asymmetric Laplace quantile likelihood function are employed, when analytic formulas are not available, to obtain relevant marginal likelihoods. The proposed methods are extensively assessed via simulations and compared to the relevant frequentist testing procedures. For the empirical study, the favoured methods from the simulation study are applied to test the adequacy of a range of forecasting methods, including from nonlinear and threshold GARCH models, for Value at Risk (VaR) in several financial market data series.

Key words:Bayesian Hypothesis testing; asymmetric-Laplace distribution; Value-at-Risk; quantile regression; Bayes factor.

1 Introduction

Following Basel II, Value-at-Risk (VaR) is popular and widely used in practice for risk management and capital allocation. Risk managers should regularly back-test the risk

*Corresponding author: Richard Gerlach. Building H04, Discipline of Business Analytics, The University of Sydney, NSW, Australia, 2006. Fax: Email: richard.gerlach@sydney.edu.au

models being used. Four well-known formal back-testing methods are the unconditional coverage (UC) test of Kupiec (1995), the conditional coverage (CC) test of Christoffersen (1998), the dynamic quantile (DQ) test of Engle and Manganelli (2004), and the quantile regression test method of Gaglianone et al. (2011).

Bayesian methods are widely employed for forecasting Value at Risk (VaR); see e.g. Chen et al. (2011), Gerlach and Chen (2008) and Hoogerheide and van Dijk (2010), etc. However, when back-testing, these papers revert to classical or frequentist methods and/or tests, or use informal criteria, to compare VaR models. The major goal of this paper is to propose back-testing methods that are within a Bayesian framework.

2 Bayes Factors and Hypothesis Testing

Hypothesis testing and model comparison problems in a Bayesian framework are tackled via posterior credible intervals or via Bayes factors (BFs). BFs are estimated via marginal likelihoods: $p(y|M_k) = \int p(y|\theta, M_k)p(\theta|M_k)d\theta$ where we choose model M_k over M_j if $\text{BF} = \frac{p(y|M_k)}{p(y|M_j)} > 1$. BFs can also be employed in hypothesis testing of $\theta = \theta_0$, where the hypothesis is rejected if $\frac{p(y|\theta_0,M)}{p(y|M)} < 1$.

2.1 UC and CC Bayesian testing

The null hypothesis in the UC test is $H_0 : \alpha = \alpha^*$, where α^* is the nominal quantile level. Under a binomial $\text{Bin}(m, \alpha)$ distribution, and employing a conjugate $\text{Beta}(a, b)$ prior, a $\text{Beta}(m_1 + a, m - m_1 + b)$ posterior distribution results for $p(\alpha|I)$. We choose a flat Beta(1,1) prior and simply form the 95% posterior credible interval from a $\text{Beta}(m_1 + 1, m - m_1 + 1)$. We reject $\alpha = \alpha^*$ when α^* is outside the credible interval and label this method "Bp11".

A BF test, analogous to the frequentist UC test is also proposed, where:

$$\text{BFUC} \quad = \quad \frac{\alpha^{m_1}(1-\alpha)^{m-m_1}}{\int \pi^{m_1}(1-\pi)^{m-m_1}p(\pi)d\pi}$$

and $H_0 : \alpha = \alpha^*$ is rejected whenever: $\text{BFUC} = \frac{\alpha^{m_1}(1-\alpha)^{m-m_1}}{B(m_1+1,m-m_1+1)} < 1$, where $B(m_1 + 1, m - m_1 + 1)$ is the standard incomplete Beta integral.

The independence and CC tests of Christofferson (1998) can also have BF analogues. First, the null model is $M_0 : I_t \sim$ i.i.d. Binomial(m, α) vs $M_1 : I_t|I_{t-1} \sim$ Binomial$(m, \pi_{i,j})$, $i, j = 0, 1$, where the alternative model is a two state Markov chain and $\pi_{i,j} = Pr(I_t = j, I_{t-1} = i)$.

$$\text{BFind} = \frac{B(m_1 + 1, m - m_1 + 1)}{B(m_{01} + 1, m_{00} + 1)B(m_{11} + 1, m_{10} + 1)}$$

where we reject the null M_0 if BFind < 1. For the CC test we have:

$$\text{BFCC} = \frac{\alpha^{m_1}(1 - \alpha)^{m - m_1}}{B(m_{01} + 1, m_{00} + 1)B(m_{11} + 1, m_{10} + 1)}$$

where m_{ij} is the number of instances where $I_t = j, I_{t-1} = i$ for $i, j = 0, 1$ and $t = 2, \ldots, m$.

2.2 Bayesian DQ testing

A Bayes factor requires an assumed model and data distribution to produce a likelihood. The DQ test employs the series of "hits" $H_t = I_t - \alpha$, $t = 1, \ldots, n$ and fits a regression:

$$H_t = \beta_0 + \sum_{i=1}^{(q-1)} \beta_i W_{i,t} + \epsilon_t$$

To get a likelihood, a distribution needs to assumed for ϵ_t. The simplest, but non-intuitive, choice is that $\epsilon_t \sim N(0, \sigma^2)$. This leads to $p(H|\beta, \sigma^2) = (2\pi)^{-0.5(m-q-1)}\sigma^{-\frac{m-q-1}{2}} \exp\left(-\frac{1}{2\sigma^2}\sum_{t=q+1}^{m} \epsilon_t^2\right)$. σ^2 is a nuisance parameter here, but under a Jeffrey's prior $p(\sigma^2) \propto \sigma^{-2}$ it can be analytically integrated out. Doing so gives:

$$\text{BFDQ1} = \frac{\left[0.5\sum_{t=q+1}^{m} H_t^2\right]^{-m/2}}{\int \ldots \int \left[0.5\sum_{t=q+1}^{m} \epsilon_t^2\right]^{-m/2} p(\beta) d\beta}.$$

Under a proper Gaussian prior on β, e.g. $\beta|\sigma^2 \sim N(0, C\sigma^2)$ (where C is diagonal with large elements), the denominator can be integrated analytically and BFDQ1 calculated. Under the null all $\beta = 0$, which is rejected whenever BFDQ ¡ 1. We employ the same regressors as in the DQ statistics, giving BFDQ1 and BFDQ4 procedures.

2.3 Bayesian VQR testing

Koenker and Machado (1999) first noted that quantile regression estimation, usually performed by minimising the quantile distance function:

$$\min_{\boldsymbol{\beta}} \sum_t u_t \left[\alpha - I(u_t < 0)\right]$$

here in the context of the regression function $y_t = \beta_0 + \beta_1 \text{VaR}_t + u_t$, is equivalent to a maximum likelihood procedure when assuming $u \sim SL(0, \sigma, \alpha)$, so that:

$$p_\alpha(u) = \frac{\alpha(1-\alpha)}{\sigma} \exp\left[-\left(\frac{u}{\sigma}\right)\right]$$

Again assuming a Jeffrey's prior on σ and integrating gives:

$$p(\boldsymbol{u}|\boldsymbol{\beta}) = \alpha^m (1-\alpha)^m \Gamma(n) \left[\sum_{t=2}^{m} u_t(\alpha - I(u_t < 0))\right]^{-m}$$

i.e. the *integrated* likelihood function. This allows a BFVQ test procedure, as:

$$\text{BFVQ} = \frac{p(\mathbf{u}|\beta_0 = 0, \beta_1 = 1)}{\int\int p(\mathbf{u}|\beta)p(\beta)d\beta_0 d\beta_1}$$

where we reject the null of $\beta_0 = 0, \beta_1 = 1$ whenever BFVQ¡1. The denominator above is a double integral over the real line. We employ a diffuse, proper Gaussian prior on β, then transform to the region $(-1, 1)^2$ and use adaptive quadrature methods to numerically estimate this integral. This takes only a second on a standard laptop using Matlab.

3 Simulations

We take the simulation setting as in Galgionone et al (2011). The true model is a GARCH(1,1), specified as:

$$\sigma_t^2 = 0.1 + 0.1 y_{t-1}^2 + 0.85\sigma_{t-1}^2 \; ; \; y_t = \sigma_t \epsilon_t \, ; \, \epsilon_t \sim N(0,1)$$

where we have $\text{VaR}_{t,\alpha} = \sigma_t \Phi^{-1}(\alpha)$. For power we use an incorrect, but common, historical simulation VaR estimator:

$$\text{HS250}_{t,\alpha} = \hat{Q}_\alpha(y_{t-250}, \ldots, y_{t-1})$$

Table 1: Size and Size-adjusted power for 5% tests of 1-step-ahead quantile forecasts $\alpha = 0.05$.

	Size				Size-adjusted power			
Method	n=250	n=500	n=1000	n=2500	n=250	n=500	n=1000	n=2500
UC	0.059	0.049	0.052	0.054	0.155	0.069	0.027	0.027
BFUC	0.012	0.006	0.003	0.003	0.132	0.063	0.027	0.027
Bp11	0.041	0.045	0.053	0.048	0.146	0.063	0.037	0.039
IND	0.014	0.034	0.084	0.053	0.148	0.174	0.211	0.500
BFIND	0.033	0.025	0.020	0.013	0.161	0.242	0.340	0.559
CC	0.043	0.037	0.060	0.057	0.173	0.167	0.183	0.394
BFCC	0.003	0.003	0.001	0.0008	0.194	0.200	0.245	0.457
IND4	0.072	0.082	0.072	0.069	0.205	0.336	0.583	0.910
BFIND4	0.024	0.028	0.020	0.014	0.257	0.391	0.615	0.914
CC4	0.059	0.070	0.068	0.062	0.253	0.300	0.514	0.872
BFCC4	0.006	0.004	0.002	0.0005	0.289	0.356	0.537	0.880
DQ1	0.050	0.049	0.049	0.048	0.355	0.448	0.621	0.949
BFDQ1	0.008	0.003	0.001	0.0002	0.476	0.550	0.676	0.958
DQ4	0.065	0.056	0.054	0.050	0.341	0.479	0.691	0.971
BFDQ4	0.005	0.001	0.0001	0.0001	0.448	0.540	0.737	0.976
VQR	0.141	0.103	0.078	0.068	0.095	0.202	0.400	0.852
BFVQ	0.634	0.527	0.405	0.223	0.276	0.309	0.411	0.866

using the sample percentile of the last 250 observations as a 1-step-ahead VaR forecast. 5000 replications of data, using sample sizes $m = 250, 500, 1000$ and 2500, are simulated in each case.

Table shows the results for size and size-adjusted power (SAP) for VaR forecasts at $\alpha = 0.05$. SAP is calculated using the empirical 95% quantile, over the 5000 datasets, for the test statistics UC, Ind, CC, DQ1, DQ4 and VQR, and also the same for each BF test, when testing the $\text{HS250}_{t,\alpha}$ forecasts in each dataset. Naturally, there is no reason why 1 should be the 95th percentage point for the sampling distribution of a BF, so size for the BF methods is not relevant, just reported for comparison.

Clearly the DQ1 test is most favoured regarding the nominal size of 5%, closely followed by the UC and Bp11 methods. The VQR method is clearly over-sized, even at $n = 2500$. Regarding power (SAP) the BFDQ methods dominate the other tests at each sample size, with clearly superior performance at $n = 250, 500$ and marginally better power at $n = 1000, 2500$ where the DQ tests become highly competitive. Results are similar for $\alpha = 0.01$.

At both $\alpha = 0.01, 0.05$ the corresponding Bayesian or BF method in most cases has higher size-adjusted power than its competing frequentist analogue.

4 Conclusion

Bayesian methods for assessing and testing forecast accuracy for dynamic quantile forecasts are developed. In the simulation study, at both $\alpha = 0.01, 0.05$ quantile levels, the corresponding Bayesian or BF method in most cases had higher size-adjusted power than its competing frequentist analogue. This suggests that Bayesian and BF methods have much to offer for quantile forecast assessment and testing.

References

Chen, C.W.S., Gerlach, R., Lin, E.M.H., and Lee, W.C.W. (2011) Bayesian forecasting for financial risk management, pre and post the global financial crisis, *Journal of Forecasting* **31**, 661-687, DOI: 10.1002/for.1237.

Christoffersen, P.F. (1998) Evaluating interval forecasts. *International Economic Review* **39**, 841-862.

Engle RF, Manganelli S. (2004) CAViaR: Conditional autoregressive value at risk by regression quantiles. *Journal of Business and Economic Statistic* **22**, 367- 381.

Gaglianone, W.P., Lima, L. R., Linton, O., and Smith, D. R. (2011) Evaluating Value-at-Risk models via quantile regression, *Journal of Business & Economic Statistics* **29**, 150-160.

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.

Kupiec, P. (1995) Techniques for verifying the accuracy of risk measurement models, *Journal of Derivatives* **2**, 173-84.