CSCanada
Advances in Natural Science
Vol. 4, No. 2, 2011, pp. 22-25
DOI:10.3968/j.ans.1715787020110402.5z723

ISSN 1715-7862 [PRINT]
ISSN 1715-7870 [ONLINE]
www.cscanada.net
www.cscanada.org

The Application of Bayesian Model Averaging in Macroeconomy

HUA Peng[1,2,*]; DU Zifang[2]

[1]College of Information Science and Engineering, Shandong University of Science and Technology, 266510, China
[2]Center for Applied Statistics, Renmin University of China, Beijing, 100872, China
[*]Corresponding author.
Email: hp0532@163.com

Supported by "SUST Spring Bud", NO.2009AZZ161. Humanities and Social Sciences Project of Ministry of Education, NO. 2009JJD910001.

Received 17 September 2011; accepted 19 November 2011

Abstract
Bayesian Model Averaging is a weighted averaging method based on posterior distribution. It considers comprehensively the prior and sample information of model and parameter, reduces the model uncertainty. Bayesian Model Averaging improves statistical inference accuracy and provides improved out-of-sample predictive performance. In this paper, we outline the details of the Bayesian model averaging principle, introduce the application of Bayesian Model Averaging in macroeconomy and give an example about the application of Bayesian Model Averaging in GDP research.

Key words: Bayesian Model Averaging; Model uncertainty; Macroeconomy

HUA Peng, DU Zifang (2011). The Application of Bayesian Model Averaging in Macroeconomy. *Advances in Natural Science, 4*(2), 22-25. Available from: URL: http://www.cscanada.net/index.php/ans/article/view/j.ans.1715787020110402.5z723
DOI: http://dx.doi.org/10.3968/j.ans.1715787020110402.5z723.

INTRODUCTION
GDP is one of the most important indicator reflect the situation of a nation's economy. It provides the basis to make economic policy, its importance is self-evident. Domestic and foreign scholars have done a lots of research theory about GDP. But the difficulty is still how to build the best model to predict accurately. The usual practice is to establish the regression or time series model, select the significant variables, the criteria to determine the optimal model is Akaike Information Criterion, Bayesian Information Criterion. But the debate focus is which kind of standards is best. Recent years, a new model and variable selection method named Bayesian Model Averaging(BMA)[1] has become the research focus. It has become latest branch of Bayesian statistical theory. BMA is a weighted averaging method based on posterior distribution, it combines all possible considered models and solve the uncertainty of single model. Therefore, in the macroeconomy research, it's better to use BMA than the regression and time series methods.

1. THE APPLICATION OF BMA IN MACROECONOMY RESEARCH

1.1 Principles of BMA
Bayesian model averaging is essentially a model selection method. Madigan and Rafery gave the following forms of BMA in 199[2]: Suppose Δ is the quantity of interest, such as an effect size, a future observable, or the utility of a course of action, then its posterior distribution given data D is

$$P(\Delta|D)=\sum_{k=1}^{K}P(\Delta|M_k,D)P(M_k|D) \tag{1}$$

where $P(\Delta|M_k, D)$ is the conditional distribution of Δ given data D and M_k, $k=1,\cdots,K$ are the model considered, $P(M_k|D)$ is posterior probability of model given data D

$$P(M_k|D)=\frac{P(D|M_k)P(M_k)}{\sum_{j=1}^{K}P(D|M_j)P(M_j)} \tag{2}$$

where $P(M_k)$ is the prior probability of model M_k, $P(D|M_k)$ is the marginal likelihood of model M_k

$$P(D|M_k) = \int\cdots\int P(D|\theta, M_k)P(\theta|M_k)d\theta \tag{3}$$

$\theta=(\theta_1,\cdots, \theta_n)$ is the vector of parameters of model M_k, $P(\theta|M_k)$ is the prior density of θ under model M_k, i.e. the prior information of parameters, $P(D|\theta, M_k)$ is the likelihood of data D given model M_k and vector of parameters θ. The posterior mean and variance of Δ are as follows:[9][10]

$$E[\Delta|D] = \sum_{k=0}^{K} E[\Delta|D, M_k]P(M_k|D) \tag{4}$$

$$Var[\Delta|D] = \sum_{k=0}^{K} (Var[\Delta|D, M_k] + E[\Delta|D, M_k]^2)P(M_k|D) - E[\Delta|D]^2 \tag{5}$$

We can drawn a conclusion from equation (1) that the posterior distribution of Δ is actually an average of the posterior distributions under each of the models considered, weighted by their posterior model probability $P(M_k|D)$.

1.2 The Application of BMA in Macroeconomic Panel Data

In the few macroeconomic study literature using BMA, Gary Koop & Simon Potter studied the GDP and inflation of United States using BMA. They analyzed 162 time series, and their analysis indicates that BMA do outperform autoregressive models in forecasting both GDP and inflation, but only narrowly and at short horizons. They attribute these findings to the presence of structural instability and the fact that lags of the dependent variable seem contain most of the information relevant for forecasting. The basic model used in their paper is:

$$y_{t+1} = \alpha(L)y_t + \gamma(L)w_t + \varepsilon_{t+1} \tag{6}$$

for t =1 ,.., T where y_t is a scalar dependent variable, w_t is a k_w vector of explanatory variables and $\alpha(L)$ and $\gamma(L)$ are polynomials in the lag operator of dimension p_1 and p_2. In many macroeconomic applications, standard methods for statistical inference in (6) are inappropriate since the number of explanatory variables is so high. In (6) we have $p_1+p_2\times k_w$ explanatory variables. In the application in Stock and Watson[6], k_w=215. In this kind of situation, we estimate directly by (6) will be very imprecise, because a lot of explanatory variables is not significant. According to statistical methods, we examine a series of variables and get a final model. But from the perspective of Bayesian statistics, this strategy ignores the model uncertainty and posterior information. BMA method avoids these shortcomings. On the weighted average of posterior distribution of models, BMA avoids the uncertainty caused by a single model. Rewritten (6) as:

$$y=X\beta+\varepsilon \tag{7}$$

where $y=(y_2\cdots y_{T+1})$, $\varepsilon=(\varepsilon_1\cdots\varepsilon_t)$, X to be the T × K matrix containing all potential lagged explanatory variables, the t[th] row of y contains data available at time t +1 while the t[th] row of X contains data available at time t. Considered three kinds of priors:

$$p(M_r) = \prod_{j=1}^{K} \theta_j^{\delta_j}(1-\theta_j)^{\delta_j} \tag{8}$$

for j=1,$\cdots$, K, θ_j is the prior of explanatory variables in models. is indicator variable, it equals 1 if variable j included in model, otherwise it equals zero. The noninformative prior is $\theta_j=\frac{1}{2}$, this means $p(M_r)=\frac{1}{R}$, for r=1, $\cdots$, R. one might expect factors corresponding to higher eigenvalues of $X'X$ to be more relevant than factors with low eigenvalues. We can incorporate this by allowing θ_j to v_j depend on ,the j[th] largest eigenvalue of $X'X$, and take $\theta_j=\frac{v_j}{v_1}$

The third type prior is called 99.9% prior which sets $\theta_j=\frac{1}{2}$ for j=1,$\cdots$, $K_{99.9}$, where 99.9% of the variation in X is contained in the first $K_{99.9}$ factors. The 99.9% prior has good performance in reduction in the root mean squared forecast error(RMSE). It can reduce RMSEs for GDP by roughly 5% while reduce for CPI by roughly 10%. BMA do outperform autoregressive models in forecasting both GDP and inflation, but only narrowly and at short horizons. This attributes to the presence of structural instability and the fact that lags of the dependent variable seem contain most of the information relevant for forecasting.

1.3 The Application of BMA in Economic Growth

Using Bayesian Model Averaging Method and provincial data from 1990 to 2007, Wang Liang and Liu jinquan[7] discern and analyze effectiveness and robustness of some factors which affect China's economic growth in the long run. The result shows six explanatory variables including high education, industrialization, openness to outside, eastern regional advantage, consumption and openness to inside are long- term determinate factors of China's economic growth having long- run, continuous and stable impact on it. Three other explanatory factors including scale of city, middle regional advantage, and initial economic conditions also have certain explanatory abilities. In addition, from perspective of explanatory variables' marginal impact on economic growth, industrialization has the strongest marginal impact which followed by consumption and openness to outside.

2. EXAMPLE

Wu jinglin and Ding yuechao[8] have studied the per capita GDP using primary industry, secondary industry, tertiary industry, household consumption expenditure, per capita annual income of urban household, per capita annual income of rural household, foreign exchange reserves, total value of imports and exports, taxes, final consumption expenditure, total investment in fixed assets. They got equation by stepwise regression:

per capita GDP of next year =1539.445-3.208664 per capita annual income of rural household +0.9568986 per capita annual income of urban household -0.01710913 total value of imports and exports -0.1345639 taxes -0.08447628 primary industry +0.2637986 secondary industry

And the value of regression equation test is 6681.772, F critical value at significance level of 0.05 is 3.094613, F critical value at significance level of 0.01 is 5.069211. Therefore, they concluded the regression equation was significant.

This conclusion only shows the equation is significant. The variables in equations are not given a significant test results, and did not give P value. Even variables in equations are significant in statistical sense, there are still obvious problems in economic sense. For example, the regressive coefficient of per capita income of rural household and the primary industry is negative, the economic sense is unreasonable. This is the instability caused by a single equation.

In this paper, we use R 2.13.0 and call R package (bicreg)[9] for Bayesian model averaging to estimate the impact of 11 variables on per capita GDP estimate. 31 models are selected. Table 1 outlines the best 5 models.

Values in table 1 present the estimate value of variable coefficient. It can be seen from the results that it's better than the stepwise regression. secondary industry, household consumption expenditure, per capita annual income of urban household, per capita annual income of rural household, final consumption expenditure, total investment in fixed assets are selected in 5 models respectively.

The variable selection in all 31 models as shown in figure 1

Table 1
The 5 Largest Model Posterior Probability by BMA

Model	Constant	Primary industry	Secondary industry	Tertiary industry	Household consumption expenditure	Per capita annual income of rural household	Per capita annual income of urban household
Model 1	364.38		0.124				
Model 2	310.4		0.19				
Model 3	219.29		0.125		0.54		
Model 4	125.96		0.128			0.835	
Model 5	86.13		0.099				0.51

(continued)

Model	Foreign exchange reserves	Total value of imports and exports	Taxes	Final consumption expenditure	Total investment in fixed assets	Posterior model probability
Model 1				0.03		0.078
Model 2					-0.039	0.078
Model 3						0.078
Model 4						0.078
Model 5						0.078

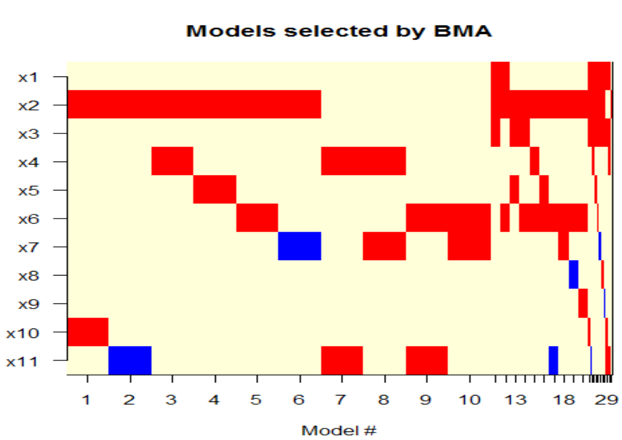

Figure 1
31 Models Selected by BMA

BMA considers the model uncertainty, improves statistical inference accuracy. We believe that with further research, BMA will have more and more widely application and expected it will play an increasingly important role in macroeconomic research.

REFERENCES

[1] Hoeting, J. A., Raftery, M. D. & Volinsky, A. E. (1999). Bayesian Model Averaging: A Tutorial. *Statistical Science, 14*(4), 382-401.

[2] Madigan, D. & Raftery, A. E. (1994). Model Selection and Accounting for Model Uncertainty in Graphical Models Using Occam's Window. *J. Amer. Statist. Assoc*, (89), 1535-1546.

[3] Raftery, A. E. (1993). Bayesian Model Selection in Structural

Equation Models. In K. Bollen & J. Long (Eds.), *Testing Structural Equation Models* (pp. 163-180), London: Sage.
[4] Draper, D. (1995). Assessment and Propagation of Model Uncertainty. *Journal of the Royal Statistical Society Series B., 57*(1), 45-97.
[5] Gary, Koo.p & Simon, Potter (2003). *Forecasting in Large Macroeconomic Panels Using Bayesian Model Averaging*. Federal Reserve Bank of New York Staff Reports no. 163.
[6] Stock, J. & Watson, M. (2002). Macroeconomic Forecasting Using Diffusion Indexes. *Journal of Business and Economic Statistics,* (20), 147-162.
[7] WANG Liang, LIU Jinquan (2010). Determinants of Economic Growth in China: Empirical Studies Based on Bayesian Averaging Model. *Statistics & Information Forum, 25*(9), 3-7.
[8] WU jinglin & DING yuechao (2009). Several Characteristics of China's Economic Indicators and Per Capita GDP Forecast. *Modern Economy,* (3), 5-8.
[9] Raftery A.E., Painter I.S., & Volinsky, C.T. (November, 2005). BMA: An R Package for Bayesian Model Averaging. *R News, 5*(2), 2-8.

