Discussion of the paper
Forecast Accuracy and Economic Gains from Bayesian Model Averaging using Time Varying Weights

L. Hoogerheide, R. Kleijn, F. Ravazzolo, H. K. van Dijk, M. Verbeek

Roberto Casarin
University of Brescia

2nd International Conference in Memory of Carlo Giannini

Rome, 18-19 January

January 19, 2010

Bayesian Model Averaging

The paper deals with Bayesian Model Averaging (BMA) and studies the forecasting performance of different model averaging schemes.

Bayesian Model Averaging

The paper deals with Bayesian Model Averaging (BMA) and studies the forecasting performance of different model averaging schemes.

The posterior probability for model the $\mathcal{M}_k$ (with $k = 1, \ldots, K$) is

$$p(\mathcal{M}_k | y_{1:T}) = \frac{p(y_{1:T}|\mathcal{M}_k)p(\mathcal{M}_k)}{\sum_{r=1}^{K} p(y_{1:T}|\mathcal{M}_r)p(\mathcal{M}_r)}$$

Bayesian Model Averaging

The paper deals with Bayesian Model Averaging (BMA) and studies the forecasting performance of different model averaging schemes.

The posterior probability for model the $\mathcal{M}_k$ (with $k = 1, \dots, K$) is

$$p(\mathcal{M}_k | y_{1:T}) = \frac{p(y_{1:T} | \mathcal{M}_k) p(\mathcal{M}_k)}{\sum_{r=1}^{K} p(y_{1:T} | \mathcal{M}_r) p(\mathcal{M}_r)}$$

The proposed approaches allow for parameter uncertainty, model uncertainty an robust time varying model weights

Model Posterior

- The posterior probability for model $\mathcal{M}_k$ $(k = 1, \ldots, K)$

$$p(\mathcal{M}_k|y_{1:T}) = \frac{p(y_{1:T}|\mathcal{M}_k)p(\mathcal{M}_k)}{\sum_{r=1}^{K} p(y_{1:T}|\mathcal{M}_r)p(\mathcal{M}_r)}$$

Model Posterior

- The posterior probability for model $\mathcal{M}_k$ $(k = 1, \ldots, K)$

$$p(\mathcal{M}_k|y_{1:T}) = \frac{p(y_{1:T}|\mathcal{M}_k)p(\mathcal{M}_k)}{\sum_{r=1}^{K} p(y_{1:T}|\mathcal{M}_r)p(\mathcal{M}_r)}$$

- In terms of Bayes Factors ($K+1$ models, thus $k = 0, \ldots, K$)

$$p(\mathcal{M}_k|y_{1:T}) = \frac{\alpha_k B_{k0}}{\sum_{r=1}^{K} \alpha_r B_{r0}}$$

where $\alpha_k = p(\mathcal{M}_k)/p(\mathcal{M}_0)$ and $B_{0k} = p(y_{1:T}|\mathcal{M}_k)/p(y_{1:T}|\mathcal{M}_0)$

Model Posterior

- The posterior probability for model $\mathcal{M}_k$ $(k = 1, \ldots, K)$

$$p(\mathcal{M}_k|y_{1:T}) = \frac{p(y_{1:T}|\mathcal{M}_k)p(\mathcal{M}_k)}{\sum_{r=1}^{K} p(y_{1:T}|\mathcal{M}_r)p(\mathcal{M}_r)}$$

- In terms of Bayes Factors ($K+1$ models, thus $k = 0, \ldots, K$)

$$p(\mathcal{M}_k|y_{1:T}) = \frac{\alpha_k B_{k0}}{\sum_{r=1}^{K} \alpha_r B_{r0}}$$

where $\alpha_k = p(\mathcal{M}_k)/p(\mathcal{M}_0)$ and $B_{0k} = p(y_{1:T}|\mathcal{M}_k)/p(y_{1:T}|\mathcal{M}_0)$

- In terms of predictive likelihood (the paper is in this framework)

$$p(\mathcal{M}_k|y_{1:T}) = \frac{p(y_T|y_{1:T-1}, \mathcal{M}_k)p(\mathcal{M}_k)}{\sum_{r=1}^{K} p(y_T|y_{1:T-1}, \mathcal{M}_r)p(\mathcal{M}_r)}$$

A Historical Perspective of BMA (see Hoeting,Madigan, Raftery an Volinsky (1999), Stat. Science)

Further references for the literature review in Introduction (pp. 2-3)

A Historical Perspective of BMA (see Hoeting,Madigan, Raftery an Volinsky (1999), Stat. Science)

Further references for the literature review in Introduction (pp. 2-3)

• Barnard, G. A. (1963), New Methods of quality control, JRSS A
First mention of model combination in the statistical literature (airline passenger data)

A Historical Perspective of BMA (see Hoeting,Madigan, Raftery an Volinsky (1999), Stat. Science)

Further references for the literature review in Introduction (pp. 2-3)

• Barnard, G. A. (1963), New Methods of quality control, JRSS A First mention of model combination in the statistical literature (airline passenger data)

• Roberts, H. V. (1965), Probabilistic prediction, JASA Suggests a distribution which combines the opinion of two experts (or models)

A Historical Perspective of BMA (see Hoeting,Madigan, Raftery an Volinsky (1999), Stat. Science)

Further references for the literature review in Introduction (pp. 2-3)

• Barnard, G. A. (1963), New Methods of quality control, JRSS A First mention of model combination in the statistical literature (airline passenger data)

• Roberts, H. V. (1965), Probabilistic prediction, JASA Suggests a distribution which combines the opinion of two experts (or models)

• Bates, J. M. and Granger, C. W. J. (1969), The combination of forecasts, Operational Research Quarterly. Seminal forecasting paper about combining predictions from different models.

A Historical Perspective of BMA (see Hoeting,Madigan, Raftery an Volinsky (1999), Stat. Science)

Further references for the literature review in Introduction (pp. 2-3)

• Barnard, G. A. (1963), New Methods of quality control, JRSS A First mention of model combination in the statistical literature (airline passenger data)

• Roberts, H. V. (1965), Probabilistic prediction, JASA Suggests a distribution which combines the opinion of two experts (or models)

• Bates, J. M. and Granger, C. W. J. (1969), The combination of forecasts, Operational Research Quarterly. Seminal forecasting paper about combining predictions from different models.

• Leamer (1978), Hodges (1987), Draper (1995)...

Alternative Approaches to BMA

Introduction and References (pp. 2-3). Other approaches to BMA: all the stochastic methods that move simultaneously in the model and parameter spaces.

Alternative Approaches to BMA

Introduction and References (pp. 2-3). Other approaches to BMA:

all the stochastic methods that move simultaneously in the model and parameter spaces.

• Markov Chain Monte Carlo Model Comparison (MC^3). See for example Madigan, York (1995) Int. J. Stat. Review, the reversible jump in Green (1995) Bka, the product space search in Carlin and Chib (1995) JRSS B

Alternative Approaches to BMA

Introduction and References (pp. 2-3). Other approaches to BMA: all the stochastic methods that move simultaneously in the model and parameter spaces.

• Markov Chain Monte Carlo Model Comparison (MC^3). See for example Madigan, York (1995) Int. J. Stat. Review, the reversible jump in Green (1995) Bka, the product space search in Carlin and Chib (1995) JRSS B

• Stochastic Search Variable Selection (SSVS) see for example George and McCulloch (1993) JASA and more recently see the model search approach for state space models in Frühwirth-Schnatter and Wagner (2009) JoE.

Forecast Combination Schemes (a $\mathcal{M} - \textit{open}$ approach?)

Section 2, pp. 4-8 of the paper.

Forecast Combination Schemes (a $\mathcal{M} - \mathit{open}$ approach?)

Section 2, pp. 4-8 of the paper.

Bernardo and Smith (1994) suggest that a BMA approach should satisfy at some properties. In particular they propose the following classification

Forecast Combination Schemes (a $\mathcal{M} - \textit{open}$ approach?)

Section 2, pp. 4-8 of the paper.

Bernardo and Smith (1994) suggest that a BMA approach should satisfy at some properties. In particular they propose the following classification

- One know the entire class of models
($\mathcal{M}$-closed perspective)

Forecast Combination Schemes (a $\mathcal{M} - open$ approach?)

Section 2, pp. 4-8 of the paper.

Bernardo and Smith (1994) suggest that a BMA approach should satisfy at some properties. In particular they propose the following classification

- One know the entire class of models
($\mathcal{M}$-closed perspective)
- The model class is not fully known in advance
($\mathcal{M}$-open perspective)

Forecast Combination Schemes (a $\mathcal{M} - \mathit{open}$ approach?)

Section 2, pp. 4-8 of the paper.

Bernardo and Smith (1994) suggest that a BMA approach should satisfy at some properties. In particular they propose the following classification

- One know the entire class of models ($\mathcal{M}$-closed perspective)
- The model class is not fully known in advance ($\mathcal{M}$-open perspective)

and we may expect that a BMA procedure should allow a new model to enter into the pool of models.

Forecast Combination Schemes (a $\mathcal{M}-$ *open* approach?)

Section 2, pp. 4-8 of the paper.

Note that the basic Ocam's Window (Madigan and Raftery (1994) JASA) approach

$$\mathcal{A}'_T = \left\{ \mathcal{M}_k | \frac{\max_r p(\mathcal{M}_r|y_{1:T})}{p(\mathcal{M}_k|y_{1:T})} \leq C \right\}$$

is $\mathcal{M}$-open (at each time iteration a new model can enter and an old model can exit the class of models)

Forecast Combination Schemes (a $\mathcal{M} - open$ approach?)

Section 2, pp. 4-8 of the paper.

The first BMA proposed in the paper is based on the following model

$$y_t = w_0 + \sum_{i=1}^{n} w_i y_{t,i} + u_t$$

with $u_t \sim \mathcal{N}(0, \sigma^2)$ i.i.d.

Forecast Combination Schemes (a $\mathcal{M} - open$ approach?)

Section 2, pp. 4-8 of the paper.

The first BMA proposed in the paper is based on the following model

$$y_t = w_0 + \sum_{i=1}^{n} w_i y_{t,i} + u_t$$

with $u_t \sim \mathcal{N}(0, \sigma^2)$ i.i.d.

It would be interesting to discuss how the proposed BMA approach is related to the Bernardo and Smith (1994) classification. (See next slide!)

Forecast Combination Schemes (a $\mathcal{M} - \textit{open}$ approach?)

p. 7 of the paper

Does the $\mathcal{M}$-open principle brings us to consider the following elements of the BMA procedure?

Forecast Combination Schemes (a $\mathcal{M}-open$ approach?)

p. 7 of the paper

Does the $\mathcal{M}$-open principle brings us to consider the following elements of the BMA procedure?

• The role of the n, that is the number of models in the pool. Could n change over time? (For example consider n_t)

Forecast Combination Schemes (a $\mathcal{M} - open$ approach?)

p. 7 of the paper

Does the $\mathcal{M}$-open principle brings us to consider the following elements of the BMA procedure?

• The role of the n, that is the number of models in the pool. Could n change over time? (For example consider n_t)

• The role of the residual term u_t and of its variance. If the true model does not belong to the pool of models then the residuals should be flexible enough (skewness, kurtosis, autocorrelation, time varying volatility...) to capture the the missing components.

Forecast Combination Schemes (a $\mathcal{M} - open$ approach?)

p. 7 of the paper

Does the $\mathcal{M}$-open principle brings us to consider the following elements of the BMA procedure?

• The role of the n, that is the number of models in the pool. Could n change over time? (For example consider n_t)

• The role of the residual term u_t and of its variance. If the true model does not belong to the pool of models then the residuals should be flexible enough (skewness, kurtosis, autocorrelation, time varying volatility...) to capture the the missing components.

• How to interpret the analysis of the residuals in a BMA context? (may stability tests (e.g. CUSUM test) help?)

Forecast Combination Schemes

p. 7 of the paper. In the time-varying weights scheme

$$y_t = w_{0,t} + \sum_{i=1}^{n} w_{i,t} y_{t,i} + u_t$$

with $u_t \sim \mathcal{N}(0, \sigma^2)$ i.i.d. and $w_t = w_{t-1} + \xi_t$ and $\xi_t \sim \mathcal{N}_{n+1}(0, \Sigma)$
The authors focus on a diagonal structure for Σ and non-diagonal Σ is for future research.

Forecast Combination Schemes

p. 7 of the paper. In the time-varying weights scheme

$$y_t = w_{0,t} + \sum_{i=1}^{n} w_{i,t} y_{t,i} + u_t$$

with $u_t \sim \mathcal{N}(0, \sigma^2)$ i.i.d. and $w_t = w_{t-1} + \xi_t$ and $\xi_t \sim \mathcal{N}_{n+1}(0, \Sigma)$ The authors focus on a diagonal structure for Σ and non-diagonal Σ is for future research.

Could one expect that change in the model weights are related to change in the prediction errors? that is use the following specification

$$\mathbb{E}(u_t \xi_t) = (\lambda_1, \ldots, \lambda_{n+1})'$$

or a more parsimonious model: $\lambda_1 = \ldots = \lambda_{n+1}$.

Forecast Combination Schemes

p. 7 of the paper. In the time-varying weights scheme

Forecast Combination Schemes

p. 7 of the paper. In the time-varying weights scheme

• In Eq. (12) for each Monte Carlo experiment s we will obtain an estimate of $w_{i,t}^{s}$. The error term is $u_t^s \sim \mathcal{N}(0, \sigma^2)$. How the author deal with the fact that σ^2 is constant across the random draws?

Forecast Combination Schemes

p. 7 of the paper. In the time-varying weights scheme

• In Eq. (12) for each Monte Carlo experiment s we will obtain an estimate of $w_{i,t}^s$. The error term is $u_t^s \sim \mathcal{N}(0, \sigma^2)$. How the author deal with the fact that σ^2 is constant across the random draws?

• In the time varying model I would expect (in financial applications for example) that the volatility of the observed values influences the forecast ability of the some models. It could be interesting to have some variables, z_t, in the dynamics of the weights $w_t = w_{t-1} + \beta' z_t + \xi_t$.

Forecast Combination Schemes

p. 8 of the paper. In the robust time-varying weights scheme

Forecast Combination Schemes

p. 8 of the paper. In the robust time-varying weights scheme

- The authors consider robust time-varying weights

$$y_t = w_{0,t} + \sum_{i=1}^{n} w_{i,t} y_{t,i} + u_t$$

with $w_t = w_{t-1} + k_t \odot \xi_t$ and $k_t \in \{0, 1\}$.

Forecast Combination Schemes

p. 8 of the paper. In the robust time-varying weights scheme

- The authors consider robust time-varying weights

$$y_t = w_{0,t} + \sum_{i=1}^{n} w_{i,t} y_{t,i} + u_t$$

with $w_t = w_{t-1} + k_t \odot \xi_t$ and $k_t \in \{0, 1\}$.

- consider a robust scheme instead (or as a further extension) the unobserved $\eta_t \in \{0, 1\}$ influences u_t, e.g.

$$u_t \sim \mathcal{N}(0, \sigma_t^2)$$

with $\sigma_t^2 = \sigma_0^2(1 - \eta_t) + \eta_t \sigma_1^2$

Active Portfolio Performances

p. 11, optimal portfolio. The authors propose to choose the optimal portfolio weights in a utility-based decision problem and use the predictive density (and thus the optimal combination scheme) to approximate the expected utility.. Consider the following points:

Active Portfolio Performances

p. 11, optimal portfolio. The authors propose to choose the optimal portfolio weights in a utility-based decision problem and use the predictive density (and thus the optimal combination scheme) to approximate the expected utility.. Consider the following points:

• The optimal combination scheme could be chosen on the basis of the expected utility function.

Active Portfolio Performances

p. 11, optimal portfolio. The authors propose to choose the optimal portfolio weights in a utility-based decision problem and use the predictive density (and thus the optimal combination scheme) to approximate the expected utility.. Consider the following points:

- The optimal combination scheme could be chosen on the basis of the expected utility function.
- Then the optimal portfolio weights and the optimal combination problems should be solved simultaneously (or iteratively)

Active Portfolio Performances

p. 11, optimal portfolio. The authors propose to choose the optimal portfolio weights in a utility-based decision problem and use the predictive density (and thus the optimal combination scheme) to approximate the expected utility.. Consider the following points:

- The optimal combination scheme could be chosen on the basis of the expected utility function.
- Then the optimal portfolio weights and the optimal combination problems should be solved simultaneously (or iteratively)

In Eq. 24, p. 11. How do the authors choose the number G of independent draws from the predictive density?

Active Portfolio Performances

p. 11, optimal portfolio. The authors propose to choose the optimal portfolio weights in a utility-based decision problem and use the predictive density (and thus the optimal combination scheme) to approximate the expected utility.. Consider the following points:

- The optimal combination scheme could be chosen on the basis of the expected utility function.
- Then the optimal portfolio weights and the optimal combination problems should be solved simultaneously (or iteratively)

In Eq. 24, p. 11. How do the authors choose the number G of independent draws from the predictive density?

In Tab. 1 Panel C, p. 21. Is (should) the comparison between the Sharpe ratio and realized utility be done in statistical terms?