An application of Bayesian Model Averaging to forecast inflation in Colombia

Eliana González

Banco de la República
Colombia

May, 2010

Contents

Contents

Contents

Contents

Contents

Contents

Contents

Motivation

- For Central Banks it is important to count with accurate predictions of future inflation in order to determine the monetary policy stance and make appropriate decisions.
  Alternatives:
  1. Select the 'best' model in terms of fit or forecasting? structural or econometrics models? or select a suite of forecasting models?.
  2. Chose among a large set of potential predictors the ones that better help to explain the dynamics of inflation and predict it in the future?
- Need to summarize the available information into a single output that capture all the relevant information from each individual forecast: Forecast combination
- Combined forecast produces smaller forecast error than any individual forecast, (Bates and Granger [1969], Newbold and Granger [1974],)
- The easy way: equal weighted average, but might be biased and affected by extreme values.
- others alternatives of forecast combination

Motivation

- For Central Banks it is important to count with accurate predictions of future inflation in order to determine the monetary policy stance and make appropriate decisions.
  Alternatives:
  1. Select the 'best' model in terms of fit or forecasting? structural or econometrics models? or select a suite of forecasting models?.
  2. Chose among a large set of potential predictors the ones that better help to explain the dynamics of inflation and predict it in the future?
- Need to summarize the available information into a single output that capture all the relevant information from each individual forecast: Forecast combination
- Combined forecast produces smaller forecast error than any individual forecast, (Bates and Granger [1969], Newbold and Granger [1974],)
- The easy way: equal weighted average, but might be biased and affected by extreme values.
- others alternatives of forecast combination

Motivation

- For Central Banks it is important to count with accurate predictions of future inflation in order to determine the monetary policy stance and make appropriate decisions.
  Alternatives:
  1. Select the 'best' model in terms of fit or forecasting? structural or econometrics models? or select a suite of forecasting models?.
  2. Chose among a large set of potential predictors the ones that better help to explain the dynamics of inflation and predict it in the future?
- Need to summarize the available information into a single output that capture all the relevant information from each individual forecast: Forecast combination
- Combined forecast produces smaller forecast error than any individual forecast, (Bates and Granger [1969], Newbold and Granger [1974],)
- The easy way: equal weighted average, but might be biased and affected by extreme values.
- others alternatives of forecast combination

Motivation

- For Central Banks it is important to count with accurate predictions of future inflation in order to determine the monetary policy stance and make appropriate decisions.
  Alternatives:
  1. Select the 'best' model in terms of fit or forecasting? structural or econometrics models? or select a suite of forecasting models?.
  2. Chose among a large set of potential predictors the ones that better help to explain the dynamics of inflation and predict it in the future?
- Need to summarize the available information into a single output that capture all the relevant information from each individual forecast: Forecast combination
- Combined forecast produces smaller forecast error than any individual forecast, (Bates and Granger [1969], Newbold and Granger [1974],)
- The easy way: equal weighted average, but might be biased and affected by extreme values.
- others alternatives of forecast combination

Motivation

- For Central Banks it is important to count with accurate predictions of future inflation in order to determine the monetary policy stance and make appropriate decisions.
  Alternatives:
  1. Select the 'best' model in terms of fit or forecasting? structural or econometrics models? or select a suite of forecasting models?.
  2. Chose among a large set of potential predictors the ones that better help to explain the dynamics of inflation and predict it in the future?
- Need to summarize the available information into a single output that capture all the relevant information from each individual forecast: Forecast combination
- Combined forecast produces smaller forecast error than any individual forecast, (Bates and Granger [1969], Newbold and Granger [1974],)
- The easy way: equal weighted average, but might be biased and affected by extreme values.
- others alternatives of forecast combination

Motivation

- For Central Banks it is important to count with accurate predictions of future inflation in order to determine the monetary policy stance and make appropriate decisions.
  Alternatives:
  1. Select the 'best' model in terms of fit or forecasting? structural or econometrics models? or select a suite of forecasting models?.
  2. Chose among a large set of potential predictors the ones that better help to explain the dynamics of inflation and predict it in the future?
- Need to summarize the available information into a single output that capture all the relevant information from each individual forecast: Forecast combination
- Combined forecast produces smaller forecast error than any individual forecast, (Bates and Granger [1969], Newbold and Granger [1974],)
- The easy way: equal weighted average, but might be biased and affected by extreme values.
- others alternatives of forecast combination

Motivation

- For Central Banks it is important to count with accurate predictions of future inflation in order to determine the monetary policy stance and make appropriate decisions.
  Alternatives:
  1. Select the 'best' model in terms of fit or forecasting? structural or econometrics models? or select a suite of forecasting models?.
  2. Chose among a large set of potential predictors the ones that better help to explain the dynamics of inflation and predict it in the future?
- Need to summarize the available information into a single output that capture all the relevant information from each individual forecast: Forecast combination
- Combined forecast produces smaller forecast error than any individual forecast, (Bates and Granger [1969], Newbold and Granger [1974],)
- The easy way: equal weighted average, but might be biased and affected by extreme values.
- others alternatives of forecast combination

Contents

Contents

Description of the methodology

BMA is

- A procedure of variable and model selection
- Based on uncertainty about the true data generating process
- Based on the Bayes theorem

Description of the methodology

BMA is

- A procedure of variable and model selection
- Based on uncertainty about the true data generating process
- Based on the Bayes theorem

Description of the methodology

BMA is

- A procedure of variable and model selection
- Based on uncertainty about the true data generating process
- Based on the Bayes theorem

Description of the methodology. Continue...

Procedure

- From a set of M models, $\mathbf{M}_1, \ldots, \mathbf{M}_M$.
- with prior belief about the probability of each model being the true one, $P(\mathbf{M}_i)$ for $i = 1, \ldots, M$,
- Given the observed data $\mathbf{Y}$
- Using Bayes theorem, the posterior probability of each model being the true one is given by:

$$P(\mathbf{M}_i/\mathbf{Y}) = \frac{m(\mathbf{Y}/\mathbf{M}_i)P(\mathbf{M}_i)}{\sum_{i=1}^{M} m(\mathbf{Y}/\mathbf{M}_i)P(\mathbf{M}_i)} \tag{1}$$

where $m(\mathbf{Y}/\mathbf{M}_i)$ is the marginal likelihood of model i defined as

$$m(\mathbf{Y}/\mathbf{M}_i) = \int L(\mathbf{Y}/\Theta_i, \mathbf{M}_i)P(\Theta_i/\mathbf{M}_i)d\Theta_i \tag{2}$$

where L is the likelihood and $P(\Theta_i/\mathbf{M}_i)$ is the posterior density of the parameter vector of model i.

Description of the methodology. Continue...

Procedure

- From a set of M models, $\mathbf{M}_1, \ldots, \mathbf{M}_M$.
- with prior belief about the probability of each model being the true one, $P(\mathbf{M}_i)$ for $i = 1, \ldots, M$,
- Given the observed data $\mathbf{Y}$
- Using Bayes theorem, the posterior probability of each model being the true one is given by:

$$P(\mathbf{M}_i/\mathbf{Y}) = \frac{m(\mathbf{Y}/\mathbf{M}_i)P(\mathbf{M}_i)}{\sum_{i=1}^{M} m(\mathbf{Y}/\mathbf{M}_i)P(\mathbf{M}_i)} \quad (1)$$

where $m(\mathbf{Y}/\mathbf{M}_i)$ is the marginal likelihood of model i defined as

$$m(\mathbf{Y}/\mathbf{M}_i) = \int L(\mathbf{Y}/\Theta_i, \mathbf{M}_i)P(\Theta_i/\mathbf{M}_i)d\Theta_i \quad (2)$$

where L is the likelihood and $P(\Theta_i/\mathbf{M}_i)$ is the posterior density of the parameter vector of model i.

Description of the methodology. Continue...

Procedure

- From a set of M models, $\mathbf{M}_1, \ldots, \mathbf{M}_M$.
- with prior belief about the probability of each model being the true one, $P(\mathbf{M}_i)$ for $i = 1, \ldots, M$,
- Given the observed data $\mathbf{Y}$
- Using Bayes theorem, the posterior probability of each model being the true one is given by:

$$P(\mathbf{M}_i/\mathbf{Y}) = \frac{m(\mathbf{Y}/\mathbf{M}_i)P(\mathbf{M}_i)}{\sum_{i=1}^{M} m(\mathbf{Y}/\mathbf{M}_i)P(\mathbf{M}_i)} \tag{1}$$

where $m(\mathbf{Y}/\mathbf{M}_i)$ is the marginal likelihood of model i defined as

$$m(\mathbf{Y}/\mathbf{M}_i) = \int L(\mathbf{Y}/\Theta_i, \mathbf{M}_i)P(\Theta_i/\mathbf{M}_i)d\Theta_i \tag{2}$$

where L is the likelihood and $P(\Theta_i/\mathbf{M}_i)$ is the posterior density of the parameter vector of model i.

Description of the methodology. Continue...

Procedure

- From a set of M models, $\mathbf{M}_1, \ldots, \mathbf{M}_M$.
- with prior belief about the probability of each model being the true one, $P(\mathbf{M}_i)$ for $i = 1, \ldots, M$,
- Given the observed data $\mathbf{Y}$
- Using Bayes theorem, the posterior probability of each model being the true one is given by:

$$P(\mathbf{M}_i/\mathbf{Y}) = \frac{m(\mathbf{Y}/\mathbf{M}_i)P(\mathbf{M}_i)}{\sum_{i=1}^{M} m(\mathbf{Y}/\mathbf{M}_i)P(\mathbf{M}_i)} \tag{1}$$

where $m(\mathbf{Y}/\mathbf{M}_i)$ is the marginal likelihood of model i defined as

$$m(\mathbf{Y}/\mathbf{M}_i) = \int L(\mathbf{Y}/\Theta_i, \mathbf{M}_i)P(\Theta_i/\mathbf{M}_i)d\Theta_i \tag{2}$$

where L is the likelihood and $P(\Theta_i/\mathbf{M}_i)$ is the posterior density of the parameter vector of model i.

Description of the methodology. Continue...

Procedure

- From a set of M models, $\mathbf{M}_1, \ldots, \mathbf{M}_M$.
- with prior belief about the probability of each model being the true one, $P(\mathbf{M}_i)$ for $i = 1, \ldots, M$,
- Given the observed data $\mathbf{Y}$
- Using Bayes theorem, the posterior probability of each model being the true one is given by:

$$P(\mathbf{M}_i/\mathbf{Y}) = \frac{m(\mathbf{Y}/\mathbf{M}_i)P(\mathbf{M}_i)}{\sum_{i=1}^{M} m(\mathbf{Y}/\mathbf{M}_i)P(\mathbf{M}_i)} \quad (1)$$

where $m(\mathbf{Y}/\mathbf{M}_i)$ is the marginal likelihood of model i defined as

$$m(\mathbf{Y}/\mathbf{M}_i) = \int L(\mathbf{Y}/\Theta_i, \mathbf{M}_i)P(\Theta_i/\mathbf{M}_i)d\Theta_i \quad (2)$$

where L is the likelihood and $P(\Theta_i/\mathbf{M}_i)$ is the posterior density of the parameter vector of model i.

Description of the methodology. Continue...

- For a quantity of interest Δ its posterior distribution is the weighted average of the posterior distributions under each of the available models

$$P(\Delta/\mathbf{Y}) = \sum_{i=1}^{M} P(\Delta/\mathbf{Y}, \mathbf{M}_i) P(\mathbf{M}_i/\mathbf{Y}) \tag{3}$$

- For a function $g(\Delta)$ its posterior distribution is give by

$$E(g(\Delta)/\mathbf{Y}) = \sum_{i=1}^{M} E(g(\Delta)/\mathbf{Y}, \mathbf{M}_i) P(\mathbf{M}_i/\mathbf{Y}) \tag{4}$$

- For the forecast $\tilde{Y}_{t+h} = E(Y_{t+h}/Y_t)$, the optimal forecast combination is the weighted average of the forecasts generated by each model.

$$E(Y_{t+h}/Y_t) = \sum_{i=1}^{M} E(Y_{t+h}/Y_t, \mathbf{M}_i) P(\mathbf{M}_i/\mathbf{Y}) \tag{5}$$

- For a quantity of interest Δ its posterior distribution is the weighted average of the posterior distributions under each of the available models

$$P(\Delta/\mathbf{Y}) = \sum_{i=1}^{M} P(\Delta/\mathbf{Y}, \mathbf{M}_i) P(\mathbf{M}_i/\mathbf{Y}) \tag{3}$$

- For a function $g(\Delta)$ its posterior distribution is give by

$$E(g(\Delta)/\mathbf{Y}) = \sum_{i=1}^{M} E(g(\Delta)/\mathbf{Y}, \mathbf{M}_i) P(\mathbf{M}_i/\mathbf{Y}) \tag{4}$$

- For the forecast $\tilde{Y}_{t+h} = E(Y_{t+h}/Y_t)$, the optimal forecast combination is the weighted average of the forecasts generated by each model.

$$E(Y_{t+h}/Y_t) = \sum_{i=1}^{M} E(Y_{t+h}/Y_t, \mathbf{M}_i) P(\mathbf{M}_i/\mathbf{Y}) \tag{5}$$

Description of the methodology. Continue...

- For a quantity of interest Δ its posterior distribution is the weighted average of the posterior distributions under each of the available models

$$P(\Delta/\mathbf{Y}) = \sum_{i=1}^{M} P(\Delta/\mathbf{Y},\mathbf{M}_i)P(\mathbf{M}_i/\mathbf{Y}) \tag{3}$$

- For a function $g(\Delta)$ its posterior distribution is give by

$$E(g(\Delta)/\mathbf{Y}) = \sum_{i=1}^{M} E(g(\Delta)/\mathbf{Y},\mathbf{M}_i)P(\mathbf{M}_i/\mathbf{Y}) \tag{4}$$

- For the forecast $\tilde{Y}_{t+h} = E(Y_{t+h}/Y_t)$, the optimal forecast combination is the weighted average of the forecasts generated by each model.

$$E(Y_{t+h}/Y_t) = \sum_{i=1}^{M} E(Y_{t+h}/Y_t,\mathbf{M}_i)P(\mathbf{M}_i/\mathbf{Y}) \tag{5}$$

Description of the methodology. Continue...

- When considering the case of variable selection, the posterior probability that variable j is included in the true model is given by

$$p(X_j/\mathbf{Y}) = \sum_{i=1}^{M} I(X_j \in \mathbf{M}_i)P(\mathbf{M}_i/\mathbf{Y}) \tag{6}$$

where $I(X_j \in \mathbf{M}_i)$ is an indicator variable, taking value of one when variable X_j is in model $\mathbf{M}_i$ and zero otherwise.

Contents

Linear Regression models $$Y = Z\gamma + \epsilon$$
where $Y = \{y_1, \cdots, y_T\}$, $\gamma = (\alpha, \beta')'$, $Z = (1, X)$ contains explanatory variables and $\epsilon \sim N(0, \sigma_\epsilon^2 I)$

- priors for the parameters
  - for the variance the Jeffrey's non-informative prior
$$p(\sigma_\epsilon^2) \propto \frac{1}{\sigma_\epsilon^2} \tag{7}$$
  - The prior distribution for the vector parameter γ/σ_ϵ^2 is the g-prior distribution, (Zellner,1986)
$$p(\gamma/\sigma_\epsilon^2, \mathbf{M}) \sim N_{k+1}(0, c\sigma_\epsilon^2(Z'Z)^{-1}) \tag{8}$$
with
$$c = \begin{cases} K^2 & \text{if } T \leq K^2 \\ T & \text{if } T > K^2 \end{cases} \tag{9}$$
as suggested by Fernandez et al, 2001.

Linear Regression models $$Y = Z\gamma + \epsilon$$
where $Y = \{y_1, \cdots, y_T\}$, $\gamma = (\alpha, \beta')'$, $Z = (1, X)$ contains explanatory variables and $\epsilon \sim N(0, \sigma_\epsilon^2 I)$

- priors for the parameters
  - for the variance the Jeffrey's non-informative prior

$$p(\sigma_\epsilon^2) \propto \frac{1}{\sigma_\epsilon^2} \tag{7}$$

  - The prior distribution for the vector parameter γ/σ_ϵ^2 is the g-prior distribution, (Zellner,1986)

$$p(\gamma/\sigma_\epsilon^2, \mathbf{M}) \sim N_{k+1}(0, c\sigma_\epsilon^2(Z'Z)^{-1}) \tag{8}$$

with

$$c = \begin{cases} K^2 & \text{if } T \leq K^2 \\ T & \text{if } T > K^2 \end{cases} \tag{9}$$

as suggested by Fernandez et al, 2001.

Linear Regression models $$Y = Z\gamma + \epsilon$$
where $Y = \{y_1, \cdots, y_T\}$, $\gamma = (\alpha, \beta')'$, $Z = (1, X)$ contains explanatory variables and $\epsilon \sim N(0, \sigma_\epsilon^2 I)$

- priors for the parameters
  - for the variance the Jeffrey's non-informative prior
$$p(\sigma_\epsilon^2) \propto \frac{1}{\sigma_\epsilon^2} \tag{7}$$
  - The prior distribution for the vector parameter γ/σ_ϵ^2 is the g-prior distribution, (Zellner,1986)
$$p(\gamma/\sigma_\epsilon^2, \mathbf{M}) \sim N_{k+1}(0, c\sigma_\epsilon^2 (Z'Z)^{-1}) \tag{8}$$
with
$$c = \begin{cases} K^2 & \text{if } T \leq K^2 \\ T & \text{if } T > K^2 \end{cases} \tag{9}$$
as suggested by Fernandez et al, 2001.

Priors

- priors for the models

$$P(\mathbf{M}_i) \propto \delta^{k_i}(1-\delta)^{K-k_i} \tag{10}$$

  - K : maximum number of variables allowed in a model
  - k_i : number of variables included in model i
  - δ : is set such that the expected model size is equal to some prior. In particular, when $\delta = 0{,}5$ the prior model probability is the same for each model.

Priors

- priors for the models

$$P(\mathbf{M}_i) \propto \delta^{k_i}(1-\delta)^{K-k_i} \qquad (10)$$

  - K : maximum number of variables allowed in a model
  - k_i : number of variables included in model i
  - δ : is set such that the expected model size is equal to some prior. In particular, when $\delta = 0{,}5$ the prior model probability is the same for each model.

Priors

- priors for the models

$$P(\mathbf{M}_i) \propto \delta^{k_i}(1-\delta)^{K-k_i} \tag{10}$$

  - K : maximum number of variables allowed in a model
  - k_i : number of variables included in model i
  - δ : is set such that the expected model size is equal to some prior. In particular, when $\delta = 0{,}5$ the prior model probability is the same for each model.

Priors

- priors for the models

$$P(\mathbf{M}_i) \propto \delta^{k_i}(1-\delta)^{K-k_i} \tag{10}$$

  - K : maximum number of variables allowed in a model
  - k_i : number of variables included in model i
  - δ : is set such that the expected model size is equal to some prior. In particular, when $\delta = 0{,}5$ the prior model probability is the same for each model.

Contents

Marginal likelihood

These set of priors lead to the posterior on the parameters

$$p(\gamma/\mathbf{Y}) \sim t_{k_i+1}(\gamma_1, S_1, M_1, \upsilon_1) \tag{11}$$

where $\gamma_1 = \frac{c}{c+1}\hat{\gamma}$ and $\hat{\gamma}$ is the OLS estimate, $\upsilon_1 = T - 1$

$$S_1 = \frac{c}{c+1}(Y - Z\hat{\gamma})'(Y - Z\hat{\gamma}) + \frac{1}{c+1}Y'Y \tag{12}$$

$$M_1 = \frac{c+1}{c}Z'Z \tag{13}$$

This leads to the marginal likelihood, which is also a multivariate t-distribution

$$m(\mathbf{Y}/\mathbf{M}) \propto (c+1)^{-(k+1)/2} S_1^{(-T/2)} \tag{14}$$

Contents

Predictive likelihood

- Avoid in$-$sample overfitting problem
- Based on the predictive ability of the model

The full sample $(\mathbf{y}_1, \cdots, \mathbf{y}_T)$ is split into two parts:

- Y^* with T_1 observations: used to obtain the posterior distribution on the parameters
- $\tilde{Y}$ with T_2 observations: used to evaluate the model performance.

The posterior predictive likelihood $P(\tilde{Y}/Y^*, \mathbf{M}_i)$ is given by

$$P(\tilde{Y}/Y^*, \mathbf{M}_i) = \int L(\tilde{Y}/\Theta_i, Y^*, \mathbf{M}_i) P(\Theta_i/Y^*, \mathbf{M}_i) d\Theta_i \tag{15}$$

Predictive likelihood

- Avoid in−sample overfitting problem
- Based on the predictive ability of the model

The full sample $(\mathbf{y}_1, \cdots, \mathbf{y}_T)$ is split into two parts:

- Y^* with T_1 observations: used to obtain the posterior distribution on the parameters
- $\tilde{Y}$ with T_2 observations: used to evaluate the model performance.

The posterior predictive likelihood $P(\tilde{Y}/Y^*, \mathbf{M}_i)$ is given by

$$P(\tilde{Y}/Y^*, \mathbf{M}_i) = \int L(\tilde{Y}/\Theta_i, Y^*, \mathbf{M}_i)P(\Theta_i/Y^*, \mathbf{M}_i)d\Theta_i \qquad (15)$$

Predictive likelihood

- Avoid in−sample overfitting problem
- Based on the predictive ability of the model

The full sample $(\mathbf{y}_1, \cdots, \mathbf{y}_T)$ is split into two parts:

- Y^* with T_1 observations: used to obtain the posterior distribution on the parameters
- $\tilde{Y}$ with T_2 observations: used to evaluate the model performance.

The posterior predictive likelihood $P(\tilde{Y}/Y^*, \mathbf{M}_i)$ is given by

$$P(\tilde{Y}/Y^*, \mathbf{M}_i) = \int L(\tilde{Y}/\Theta_i, Y^*, \mathbf{M}_i) P(\Theta_i/Y^*, \mathbf{M}_i) d\Theta_i \qquad (15)$$

Predictive likelihood

- Avoid in−sample overfitting problem
- Based on the predictive ability of the model

The full sample $(\mathbf{y}_1, \cdots, \mathbf{y}_T)$ is split into two parts:

- Y^* with T_1 observations: used to obtain the posterior distribution on the parameters
- $\tilde{Y}$ with T_2 observations: used to evaluate the model performance.

The posterior predictive likelihood $P(\tilde{Y}/Y^*, \mathbf{M}_i)$ is given by

$$P(\tilde{Y}/Y^*, \mathbf{M}_i) = \int L(\tilde{Y}/\Theta_i, Y^*, \mathbf{M}_i) P(\Theta_i/Y^*, \mathbf{M}_i) d\Theta_i \tag{15}$$

Predictive likelihood ...

Under the priors above, the predictive density of $\tilde{Y} = (\mathbf{y}_{T_1+1}, \cdots, \mathbf{y}_T)$ is

$$P(\tilde{Y}/\tilde{Z}, Z^*, Y^*, \gamma^*, \sigma_\epsilon^2) \sim N_{T_2}(\tilde{Z}\gamma^*, \sigma_\epsilon^2 I_{T_2}) \tag{16}$$

- $\tilde{Z}$: out-sample matrix of explanatory variables
- γ^* : parameter vector estimated with the training sample Y^*.

The predictive posterior density of $\tilde{Y}$ is a multivariate t-student distribution.

$$\tilde{Y}/\tilde{Z}, Z^*, Y^* \sim \mathbf{t}_{T_2}(\tilde{Z}\gamma_1, S^*, (I_{T_2} + \tilde{Z}(M^*)^{-1}\tilde{Z}')^{-1}, T_1) \tag{17}$$

with density function

$$P(\tilde{Y}/\tilde{Z}, Z^*, Y^*) \propto \frac{{S^*}^{T_1/2} \left|M^*\right|^{1/2}}{\left|M^* + \tilde{Z}'\tilde{Z}\right|^{1/2}} \times [S^* + (\tilde{Y} - \tilde{Z}\gamma_1)'(I_{T_2} + \tilde{Z}(M^*)^{-1}\tilde{Z})^{-1}(\tilde{Y} - \tilde{Z}\gamma_1)]^{-T/2} \tag{18}$$

where S^*, γ_1 and M^* are defined as in the case of the marginal likelihood but calculated over the sample Y^*.

Predictive likelihood ...

Under the priors above, the predictive density of $\tilde{Y} = (\mathbf{y}_{T_1+1}, \cdots, \mathbf{y}_T)$ is

$$P(\tilde{Y}/\tilde{Z}, Z^*, Y^*, \gamma^*, \sigma_\epsilon^2) \sim N_{T_2}(\tilde{Z}\gamma^*, \sigma_\epsilon^2 I_{T_2}) \tag{16}$$

- $\tilde{Z}$: out-sample matrix of explanatory variables
- γ^* : parameter vector estimated with the training sample Y^*.

The predictive posterior density of $\tilde{Y}$ is a multivariate t-student distribution.

$$\tilde{Y}/\tilde{Z}, Z^*, Y^* \sim \mathbf{t}_{T_2}(\tilde{Z}\gamma_1, S^*, (I_{T_2} + \tilde{Z}(M^*)^{-1}\tilde{Z}')^{-1}, T_1) \tag{17}$$

with density function

$$P(\tilde{Y}/\tilde{Z}, Z^*, Y^*) \propto \frac{{S^*}^{T_1/2} \left|M^*\right|^{1/2}}{\left|M^* + \tilde{Z}'\tilde{Z}\right|^{1/2}} \times [S^* + (\tilde{Y} - \tilde{Z}\gamma_1)'(I_{T_2} + \tilde{Z}(M^*)^{-1}\tilde{Z})^{-1}(\tilde{Y} - \tilde{Z}\gamma_1)]^{-T/2} \tag{18}$$

where S^*, γ_1 and M^* are defined as in the case of the marginal likelihood but calculated over the sample Y^*.

Posterior probabilities...

For each model

- Using marginal likelihood

$$P(\mathbf{M}_i/\mathbf{Y}) = \frac{m(\mathbf{Y}/\mathbf{M}_i)P(\mathbf{M}_i)}{\sum_{m=1}^{M} m(\mathbf{Y}/\mathbf{M}_i)P(\mathbf{M}_i)} \tag{19}$$

- Using predictive likelihood

$$P(\mathbf{M}_i/\tilde{Y}, Y^*) = \frac{P(\tilde{Y}/Y^*, \mathbf{M}_i)P(\mathbf{M}_i)}{\sum_{m=1}^{M} P(\tilde{Y}/Y^*, \mathbf{M}_i)P(\mathbf{M}_i)} \tag{20}$$

Posterior probabilities...

For each model

- Using marginal likelihood

$$P(\mathbf{M}_i/\mathbf{Y}) = \frac{m(\mathbf{Y}/\mathbf{M}_i)P(\mathbf{M}_i)}{\sum_{m=1}^{M} m(\mathbf{Y}/\mathbf{M}_i)P(\mathbf{M}_i)} \tag{19}$$

- Using predictive likelihood

$$P(\mathbf{M}_i/\tilde{Y}, Y^*) = \frac{P(\tilde{Y}/Y^*, \mathbf{M}_i)P(\mathbf{M}_i)}{\sum_{m=1}^{M} P(\tilde{Y}/Y^*, \mathbf{M}_i)P(\mathbf{M}_i)} \tag{20}$$

Contents

Model space

- The idea is to find a set of M "good" models from a large set of K potential predictors. The model space contains 2^K possible models.
- Restrict the model space considering models containing up to k covariates. The model space contains $(\sum_{j=0}^{k} \binom{K}{j})$ possible models. But still too many models to evaluate.
- Jacobson and Karlsson, (2002), suggested to use MCMC algorithms to visit models with non-negligible posterior probabilities.
- Reversible jump MCMC algoritm, (Green,1995).

Model space

- The idea is to find a set of M "good" models from a large set of K potential predictors. The model space contains 2^K possible models.
- Restrict the model space considering models containing up to k covariates. The model space contains $\left(\sum_{j=0}^{k} \binom{K}{j}\right)$ possible models. But still too many models to evaluate.
- Jacobson and Karlsson, (2002), suggested to use MCMC algorithms to visit models with non-negligible posterior probabilities.
- Reversible jump MCMC algoritm, (Green,1995).

Model space

- The idea is to find a set of M "good" models from a large set of K potential predictors. The model space contains 2^K possible models.
- Restrict the model space considering models containing up to k covariates. The model space contains $(\sum_{j=0}^{k} \binom{K}{j})$ possible models. But still too many models to evaluate.
- Jacobson and Karlsson, (2002), suggested to use MCMC algorithms to visit models with non-negligible posterior probabilities.
- Reversible jump MCMC algoritm, (Green,1995).

Model space

- The idea is to find a set of M "good" models from a large set of K potential predictors. The model space contains 2^K possible models.
- Restrict the model space considering models containing up to k covariates. The model space contains $(\sum_{j=0}^{k} \binom{K}{j})$ possible models. But still too many models to evaluate.
- Jacobson and Karlsson, (2002), suggested to use MCMC algorithms to visit models with non-negligible posterior probabilities.
- Reversible jump MCMC algoritm, (Green,1995).

RJMCMC algorithm

Procedure

From an initial state of the chain $(\theta_{\mathbf{M}}, \mathbf{M})$

- Propose a jump from model $\mathbf{M}$ to model $\mathbf{M}^*$ with probability $j(\mathbf{M}^*/\mathbf{M})$
- Generate a vector u from a proposal density $\mathbf{q}(u/\mathbf{M}, \mathbf{M}^*)$
- Set $(\theta_{\mathbf{M}^*}, u^*) = g_{\mathbf{M},\mathbf{M}^*}(\theta_{\mathbf{M}}, u)$, where g is a specified invertible function and u, u^* satisfy $dim(u) + dim(\theta_{\mathbf{M}}) = dim(u^*) + dim(\theta_{\mathbf{M}^*})$
- Accept the move with probability

$$\alpha = \text{mín}\left\{1, \frac{L(\mathbf{Y}/\Theta_{\mathbf{M}^*}, \mathbf{M}^*)P(\Theta_{\mathbf{M}^*}/\mathbf{M}^*)P(\mathbf{M}^*)j(\mathbf{M}/\mathbf{M}^*)q(u^*/\Theta_{\mathbf{M}^*}, \mathbf{M}^*, \mathbf{M})}{L(\mathbf{Y}/\Theta_{\mathbf{M}}, \mathbf{M})P(\Theta_{\mathbf{M}}/\mathbf{M})P(\mathbf{M})j(\mathbf{M}^*/\mathbf{M})q(u/\Theta_{\mathbf{M}}, \mathbf{M}, \mathbf{M}^*)} \times \left|\frac{\partial g_{\mathbf{M},\mathbf{M}^*}(\theta_{\mathbf{M}}, u)}{\partial(\theta_{\mathbf{M}}, u)}\right|\right\} \tag{21}$$

and set $\mathbf{M} = \mathbf{M}^*$ if the move is accepted.

RJMCMC algorithm

Procedure

From an initial state of the chain $(\theta_{\mathbf{M}}, \mathbf{M})$

- Propose a jump from model $\mathbf{M}$ to model $\mathbf{M}^*$ with probability $j(\mathbf{M}^*/\mathbf{M})$
- Generate a vector u from a proposal density $\mathbf{q}(u/\mathbf{M}, \mathbf{M}^*)$
- Set $(\theta_{\mathbf{M}^*}, u^*) = g_{\mathbf{M},\mathbf{M}^*}(\theta_{\mathbf{M}}, u)$, where g is a specified invertible function and u, u^* satisfy $dim(u) + dim(\theta_{\mathbf{M}}) = dim(u^*) + dim(\theta_{\mathbf{M}^*})$
- Accept the move with probability

$$\alpha = \text{mín}\left\{1, \frac{L(\mathbf{Y}/\Theta_{\mathbf{M}^*}, \mathbf{M}^*)P(\Theta_{\mathbf{M}^*}/\mathbf{M}^*)P(\mathbf{M}^*)j(\mathbf{M}/\mathbf{M}^*)q(u^*/\Theta_{\mathbf{M}^*}, \mathbf{M}^*, \mathbf{M})}{L(\mathbf{Y}/\Theta_{\mathbf{M}}, \mathbf{M})P(\Theta_{\mathbf{M}}/\mathbf{M})P(\mathbf{M})j(\mathbf{M}^*/\mathbf{M})q(u/\Theta_{\mathbf{M}}, \mathbf{M}, \mathbf{M}^*)} \times \left|\frac{\partial g_{\mathbf{M},\mathbf{M}^*}(\theta_{\mathbf{M}}, u)}{\partial(\theta_{\mathbf{M}}, u)}\right|\right\} \quad (21)$$

  and set $\mathbf{M} = \mathbf{M}^*$ if the move is accepted.

RJMCMC algorithm

Procedure

From an initial state of the chain $(\theta_{\mathbf{M}}, \mathbf{M})$

- Propose a jump from model $\mathbf{M}$ to model $\mathbf{M}^*$ with probability $j(\mathbf{M}^*/\mathbf{M})$
- Generate a vector u from a proposal density $\mathbf{q}(u/\mathbf{M}, \mathbf{M}^*)$
- Set $(\theta_{\mathbf{M}^*}, u^*) = g_{\mathbf{M},\mathbf{M}^*}(\theta_{\mathbf{M}}, u)$, where g is a specified invertible function and u, u^* satisfy $dim(u) + dim(\theta_{\mathbf{M}}) = dim(u^*) + dim(\theta_{\mathbf{M}^*})$
- Accept the move with probability

$$\alpha = \text{mín}\left\{1, \frac{L(\mathbf{Y}/\Theta_{\mathbf{M}^*}, \mathbf{M}^*)P(\Theta_{\mathbf{M}^*}/\mathbf{M}^*)P(\mathbf{M}^*)j(\mathbf{M}/\mathbf{M}^*)q(u^*/\Theta_{\mathbf{M}^*}, \mathbf{M}^*, \mathbf{M})}{L(\mathbf{Y}/\Theta_{\mathbf{M}}, \mathbf{M})P(\Theta_{\mathbf{M}}/\mathbf{M})P(\mathbf{M})j(\mathbf{M}^*/\mathbf{M})q(u/\Theta_{\mathbf{M}}, \mathbf{M}, \mathbf{M}^*)} \times \left|\frac{\partial g_{\mathbf{M},\mathbf{M}^*}(\theta_{\mathbf{M}}, u)}{\partial(\theta_{\mathbf{M}}, u)}\right|\right\} \quad (21)$$

and set $\mathbf{M} = \mathbf{M}^*$ if the move is accepted.

RJMCMC algorithm

Procedure

From an initial state of the chain $(\theta_{\mathbf{M}}, \mathbf{M})$

- Propose a jump from model $\mathbf{M}$ to model $\mathbf{M}^*$ with probability $j(\mathbf{M}^*/\mathbf{M})$
- Generate a vector u from a proposal density $\mathbf{q}(u/\mathbf{M}, \mathbf{M}^*)$
- Set $(\theta_{\mathbf{M}^*}, u^*) = g_{\mathbf{M},\mathbf{M}^*}(\theta_{\mathbf{M}}, u)$, where g is a specified invertible function and u, u^* satisfy $dim(u) + dim(\theta_{\mathbf{M}}) = dim(u^*) + dim(\theta_{\mathbf{M}^*})$
- Accept the move with probability

$$\alpha = \text{mín}\left\{1, \frac{L(\mathbf{Y}/\Theta_{\mathbf{M}^*}, \mathbf{M}^*)P(\Theta_{\mathbf{M}^*}/\mathbf{M}^*)P(\mathbf{M}^*)j(\mathbf{M}/\mathbf{M}^*)q(u^*/\Theta_{\mathbf{M}^*}, \mathbf{M}^*, \mathbf{M})}{L(\mathbf{Y}/\Theta_{\mathbf{M}}, \mathbf{M})P(\Theta_{\mathbf{M}}/\mathbf{M})P(\mathbf{M})j(\mathbf{M}^*/\mathbf{M})q(u/\Theta_{\mathbf{M}}, \mathbf{M}, \mathbf{M}^*)} \times \left|\frac{\partial g_{\mathbf{M},\mathbf{M}^*}(\theta_{\mathbf{M}}, u)}{\partial(\theta_{\mathbf{M}}, u)}\right|\right\} \qquad (21)$$

  and set $\mathbf{M} = \mathbf{M}^*$ if the move is accepted.

RJMCMC algorithm

Procedure

From an initial state of the chain $(\theta_{\mathbf{M}}, \mathbf{M})$

- Propose a jump from model $\mathbf{M}$ to model $\mathbf{M}^*$ with probability $j(\mathbf{M}^*/\mathbf{M})$
- Generate a vector u from a proposal density $\mathbf{q}(u/\mathbf{M}, \mathbf{M}^*)$
- Set $(\theta_{\mathbf{M}^*}, u^*) = g_{\mathbf{M},\mathbf{M}^*}(\theta_{\mathbf{M}}, u)$, where g is a specified invertible function and u, u^* satisfy $dim(u) + dim(\theta_{\mathbf{M}}) = dim(u^*) + dim(\theta_{\mathbf{M}^*})$
- Accept the move with probability

$$\alpha = \text{mín}\left\{1, \frac{L(\mathbf{Y}/\Theta_{\mathbf{M}^*}, \mathbf{M}^*)P(\Theta_{\mathbf{M}^*}/\mathbf{M}^*)P(\mathbf{M}^*)j(\mathbf{M}/\mathbf{M}^*)q(u^*/\Theta_{\mathbf{M}^*}, \mathbf{M}^*, \mathbf{M})}{L(\mathbf{Y}/\Theta_{\mathbf{M}}, \mathbf{M})P(\Theta_{\mathbf{M}}/\mathbf{M})P(\mathbf{M})j(\mathbf{M}^*/\mathbf{M})q(u/\Theta_{\mathbf{M}}, \mathbf{M}, \mathbf{M}^*)} \times \left| \frac{\partial g_{\mathbf{M},\mathbf{M}^*}(\theta_{\mathbf{M}}, u)}{\partial(\theta_{\mathbf{M}}, u)} \right| \right\} \quad (21)$$

  and set $\mathbf{M} = \mathbf{M}^*$ if the move is accepted.

RJMCMC algorithm ...

Procedure

Following the works of Jacobson and Karlsson,(2002) and Eklund and Karlsson, (2005), the algorithm simplifies

- condidering local moves only:
  1. add or drop jump. $j(M/M^*) = j(M^*/M) = \frac{1}{K}$, with K the number of available variables.
  2. swap jump. $j(M/M^*) = j(M^*/M) = \frac{1}{k(K-k)}$, with k the number of variables in the model.
- If all parameters of the proposed model are generated from a proposal distribution, then
  - $(\theta_{\mathbf{M}^*}, u^*) = (u, \theta_{\mathbf{M}})$ with $dim(\theta_{\mathbf{M}}) = dim(u^*)$ and $dim(\theta_{\mathbf{M}^*}) = dim(u)$
  - the Jacobian

$$\frac{\partial g_{\mathbf{M},\mathbf{M}^*}(\theta_{\mathbf{M}}, u)}{\partial(\theta_{\mathbf{M}}, u)} = 1 \qquad (22)$$

RJMCMC algorithm ...

Procedure

Following the works of Jacobson and Karlsson,(2002) and Eklund and Karlsson, (2005), the algorithm simplifies

- condidering local moves only:
  1. add or drop jump. $j(M/M^*) = j(M^*/M) = \frac{1}{K}$, with K the number of available variables.
  2. swap jump. $j(M/M^*) = j(M^*/M) = \frac{1}{k(K-k)}$, with k the number of variables in the model.
- If all parameters of the proposed model are generated from a proposal distribution, then
  - $(\theta_{M^*}, u^*) = (u, \theta_M)$ with $dim(\theta_M) = dim(u^*)$ and $dim(\theta_{M^*}) = dim(u)$
  - the Jacobian

$$\frac{\partial g_{M,M^*}(\theta_M, u)}{\partial(\theta_M, u)} = 1 \qquad (22)$$

RJMCMC algorithm ...

Procedure

Following the works of Jacobson and Karlsson,(2002) and Eklund and Karlsson, (2005), the algorithm simplifies

- condidering local moves only:
  1. add or drop jump. $j(M/M^*) = j(M^*/M) = \frac{1}{K}$, with K the number of available variables.
  2. swap jump. $j(M/M^*) = j(M^*/M) = \frac{1}{k(K-k)}$, with k the number of variables in the model.
- If all parameters of the proposed model are generated from a proposal distribution, then
  - $(\theta_{\mathbf{M}^*}, u^*) = (u, \theta_{\mathbf{M}})$ with $dim(\theta_{\mathbf{M}}) = dim(u^*)$ and $dim(\theta_{\mathbf{M}^*}) = dim(u)$
  - the Jacobian

$$\frac{\partial g_{\mathbf{M},\mathbf{M}^*}(\theta_{\mathbf{M}}, u)}{\partial(\theta_{\mathbf{M}}, u)} = 1 \quad (22)$$

RJMCMC algorithm ...

Procedure

Following the works of Jacobson and Karlsson,(2002) and Eklund and Karlsson, (2005), the algorithm simplifies

- condidering local moves only:
  1. add or drop jump. $j(M/M^*) = j(M^*/M) = \frac{1}{K}$, with K the number of available variables.
  2. swap jump. $j(M/M^*) = j(M^*/M) = \frac{1}{k(K-k)}$, with k the number of variables in the model.
- If all parameters of the proposed model are generated from a proposal distribution, then
  - $(\theta_{M^*}, u^*) = (u, \theta_M)$ with $dim(\theta_M) = dim(u^*)$ and $dim(\theta_{M^*}) = dim(u)$
  - the Jacobian

$$\frac{\partial g_{M,M^*}(\theta_M, u)}{\partial(\theta_M, u)} = 1 \qquad (22)$$

RJMCMC algorithm ...

Procedure

Following the works of Jacobson and Karlsson,(2002) and Eklund and Karlsson, (2005), the algorithm simplifies

- condidering local moves only:
  1. add or drop jump. $j(M/M^*) = j(M^*/M) = \frac{1}{K}$, with K the number of available variables.
  2. swap jump. $j(M/M^*) = j(M^*/M) = \frac{1}{k(K-k)}$, with k the number of variables in the model.
- If all parameters of the proposed model are generated from a proposal distribution, then
  - $(\theta_{\mathbf{M}^*}, u^*) = (u, \theta_{\mathbf{M}})$ with $dim(\theta_{\mathbf{M}}) = dim(u^*)$ and $dim(\theta_{\mathbf{M}^*}) = dim(u)$
  - the Jacobian

$$\frac{\partial g_{\mathbf{M},\mathbf{M}^*}(\theta_{\mathbf{M}}, u)}{\partial(\theta_{\mathbf{M}}, u)} = 1 \qquad (22)$$

RJMCMC algorithm ...

Procedure

Following the works of Jacobson and Karlsson,(2002) and Eklund and Karlsson, (2005), the algorithm simplifies

- condidering local moves only:
  1. add or drop jump. $j(M/M^*) = j(M^*/M) = \frac{1}{K}$, with K the number of available variables.
  2. swap jump. $j(M/M^*) = j(M^*/M) = \frac{1}{k(K-k)}$, with k the number of variables in the model.
- If all parameters of the proposed model are generated from a proposal distribution, then
  - $(\theta_{\mathbf{M}^*}, u^*) = (u, \theta_{\mathbf{M}})$ with $dim(\theta_{\mathbf{M}}) = dim(u^*)$ and $dim(\theta_{\mathbf{M}^*}) = dim(u)$
  - the Jacobian

$$\frac{\partial g_{\mathbf{M},\mathbf{M}^*}(\theta_{\mathbf{M}}, u)}{\partial(\theta_{\mathbf{M}}, u)} = 1 \qquad (22)$$

RJMCMC algorithm ...

Procedure

Following the works of Jacobson and Karlsson,(2002) and Eklund and Karlsson, (2005), the algorithm simplifies

- condidering local moves only:
  1. add or drop jump. $j(M/M^*) = j(M^*/M) = \frac{1}{K}$, with K the number of available variables.
  2. swap jump. $j(M/M^*) = j(M^*/M) = \frac{1}{k(K-k)}$, with k the number of variables in the model.
- If all parameters of the proposed model are generated from a proposal distribution, then
  - $(\theta_{\mathbf{M}^*}, u^*) = (u, \theta_{\mathbf{M}})$ with $dim(\theta_{\mathbf{M}}) = dim(u^*)$ and $dim(\theta_{\mathbf{M}^*}) = dim(u)$
  - the Jacobian

$$\frac{\partial g_{\mathbf{M},\mathbf{M}^*}(\theta_{\mathbf{M}}, u)}{\partial(\theta_{\mathbf{M}}, u)} = 1 \tag{22}$$

RJMCMC algorithm ...

Procedure ...

- if considering the posterior distribution of θ_M, $P(\theta_M/\mathbf{Y}, M)$ as the proposal distribution for the parameters space, then the acceptance probability of the move from M to M^* simplifies further to

$$\alpha = \text{mín}\left\{1, \frac{L(\mathbf{Y}/\Theta_i, \mathbf{M}^*)P(\Theta_i/\mathbf{M}^*)P(\mathbf{M}^*)}{L(\mathbf{Y}/\Theta_i, \mathbf{M})P(\Theta_i/\mathbf{M})P(\mathbf{M})}\right\} \quad (23)$$

or

$$\alpha = \text{mín}\left\{1, \frac{m(\mathbf{Y}/\mathbf{M}^*)P(\mathbf{M}^*)}{m(\mathbf{Y}/\mathbf{M})P(\mathbf{M})}\right\} \quad (24)$$

RJMCMC algorithm ...

Procedure ...

- if considering the posterior distribution of θ_M, $P(\theta_M/\mathbf{Y}, M)$ as the proposal distribution for the parameters space, then the acceptance probability of the move from M to M^* simplifies further to

$$\alpha = \text{mín}\left\{1, \frac{L(\mathbf{Y}/\Theta_i, \mathbf{M}^*)P(\Theta_i/\mathbf{M}^*)P(\mathbf{M}^*)}{L(\mathbf{Y}/\Theta_i, \mathbf{M})P(\Theta_i/\mathbf{M})P(\mathbf{M})}\right\} \tag{23}$$

or

$$\alpha = \text{mín}\left\{1, \frac{m(\mathbf{Y}/\mathbf{M}^*)P(\mathbf{M}^*)}{m(\mathbf{Y}/\mathbf{M})P(\mathbf{M})}\right\} \tag{24}$$

Contents

Empirical Application

Data

- The variable of interest is inflation measured as the twelve-month growth rate of total CPI.
- Dataset consists on 73 macroeconomic variables seasonally adjusted and transformed as annual growths or twelve-month differences.
  1. Economic activity variables: employment, wages, imports, exports, production, expectations about production ... (26 series)
  2. Prices: total CPI and PPI and components, expectations about inflation (23 series)
  3. Monetary, credit and exchange rate variables: monetary aggregates, interest rates, nominal and real exchange rates, terms of trade ...(24 series)
- Full Sample: from Nov−1999 to Dec−2009.
- For marginal likelihood: $Y = \left\{ \mathbf{y}_{Nov/1999}, \cdots, \mathbf{y}_{Dec/2007} \right\}$
- For predictive likelihood:
  small training sample $Y^* = \left\{ \mathbf{y}_{Nov/1999}, \cdots, \mathbf{y}_{Dec/2004} \right\}$,
  hold-out sample $\tilde{Y} = \left\{ \mathbf{y}_{Jan/2005}, \cdots, \mathbf{y}_{Dec/2007} \right\}$.
- Initial model space: 2^{73} possible models

- The variable of interest is inflation measured as the twelve-month growth rate of total CPI.
- Dataset consists on 73 macroeconomic variables seasonally adjusted and transformed as annual growths or twelve-month differences.
  1. Economic activity variables: employment, wages, imports, exports, production, expectations about production ... (26 series)
  2. Prices: total CPI and PPI and components, expectations about inflation (23 series)
  3. Monetary, credit and exchange rate variables: monetary aggregates, interest rates, nominal and real exchange rates, terms of trade ...(24 series)
- Full Sample: from Nov−1999 to Dec−2009.
- For marginal likelihood: $Y = \left\{ \mathbf{y}_{Nov/1999}, \cdots, \mathbf{y}_{Dec/2007} \right\}$
- For predictive likelihood:
  small training sample $Y^* = \left\{ \mathbf{y}_{Nov/1999}, \cdots, \mathbf{y}_{Dec/2004} \right\}$,
  hold-out sample $\tilde{Y} = \left\{ \mathbf{y}_{Jan/2005}, \cdots, \mathbf{y}_{Dec/2007} \right\}$.
- Initial model space: 2^{73} possible models

Empirical Application

Data

- The variable of interest is inflation measured as the twelve-month growth rate of total CPI.
- Dataset consists on 73 macroeconomic variables seasonally adjusted and transformed as annual growths or twelve-month differences.
    1. Economic activity variables: employment, wages, imports, exports, production, expectations about production ... (26 series)
    2. Prices: total CPI and PPI and components, expectations about inflation (23 series)
    3. Monetary, credit and exchange rate variables: monetary aggregates, interest rates, nominal and real exchange rates, terms of trade ...(24 series)
- Full Sample: from Nov−1999 to Dec−2009.
- For marginal likelihood: $Y = \left\{ \mathbf{y}_{Nov/1999}, \cdots, \mathbf{y}_{Dec/2007} \right\}$
- For predictive likelihood:
  small training sample $Y^* = \left\{ \mathbf{y}_{Nov/1999}, \cdots, \mathbf{y}_{Dec/2004} \right\}$,
  hold-out sample $\tilde{Y} = \left\{ \mathbf{y}_{Jan/2005}, \cdots, \mathbf{y}_{Dec/2007} \right\}$.
- Initial model space: 2^{73} possible models

- The variable of interest is inflation measured as the twelve-month growth rate of total CPI.
- Dataset consists on 73 macroeconomic variables seasonally adjusted and transformed as annual growths or twelve-month differences.
  1. Economic activity variables: employment, wages, imports, exports, production, expectations about production ... (26 series)
  2. Prices: total CPI and PPI and components, expectations about inflation (23 series)
  3. Monetary, credit and exchange rate variables: monetary aggregates, interest rates, nominal and real exchange rates, terms of trade ...(24 series)
- Full Sample: from Nov−1999 to Dec−2009.
- For marginal likelihood: $Y = \left\{ \mathbf{y}_{Nov/1999}, \cdots, \mathbf{y}_{Dec/2007} \right\}$
- For predictive likelihood:
  small training sample $Y^* = \left\{ \mathbf{y}_{Nov/1999}, \cdots, \mathbf{y}_{Dec/2004} \right\}$,
  hold-out sample $\tilde{Y} = \left\{ \mathbf{y}_{Jan/2005}, \cdots, \mathbf{y}_{Dec/2007} \right\}$.
- Initial model space: 2^{73} possible models

- The variable of interest is inflation measured as the twelve-month growth rate of total CPI.
- Dataset consists on 73 macroeconomic variables seasonally adjusted and transformed as annual growths or twelve-month differences.
  1. Economic activity variables: employment, wages, imports, exports, production, expectations about production ... (26 series)
  2. Prices: total CPI and PPI and components, expectations about inflation (23 series)
  3. Monetary, credit and exchange rate variables: monetary aggregates, interest rates, nominal and real exchange rates, terms of trade ...(24 series)
- Full Sample: from Nov−1999 to Dec−2009.
- For marginal likelihood: $Y = \left\{ \mathbf{y}_{Nov/1999}, \cdots, \mathbf{y}_{Dec/2007} \right\}$
- For predictive likelihood:
  small training sample $Y^* = \left\{ \mathbf{y}_{Nov/1999}, \cdots, \mathbf{y}_{Dec/2004} \right\}$,
  hold-out sample $\tilde{Y} = \left\{ \mathbf{y}_{Jan/2005}, \cdots, \mathbf{y}_{Dec/2007} \right\}$.
- Initial model space: 2^{73} possible models

Empirical Application

Data

- The variable of interest is inflation measured as the twelve-month growth rate of total CPI.
- Dataset consists on 73 macroeconomic variables seasonally adjusted and transformed as annual growths or twelve-month differences.
  1. Economic activity variables: employment, wages, imports, exports, production, expectations about production ... (26 series)
  2. Prices: total CPI and PPI and components, expectations about inflation (23 series)
  3. Monetary, credit and exchange rate variables: monetary aggregates, interest rates, nominal and real exchange rates, terms of trade ...(24 series)
- Full Sample: from Nov−1999 to Dec−2009.
- For marginal likelihood: $Y = \left\{\mathbf{y}_{Nov/1999}, \cdots, \mathbf{y}_{Dec/2007}\right\}$
- For predictive likelihood:
  small training sample $Y^* = \left\{\mathbf{y}_{Nov/1999}, \cdots, \mathbf{y}_{Dec/2004}\right\}$,
  hold-out sample $\tilde{Y} = \left\{\mathbf{y}_{Jan/2005}, \cdots, \mathbf{y}_{Dec/2007}\right\}$.
- Initial model space: 2^{73} possible models

Empirical Application
Data

- The variable of interest is inflation measured as the twelve-month growth rate of total CPI.
- Dataset consists on 73 macroeconomic variables seasonally adjusted and transformed as annual growths or twelve-month differences.
  1. Economic activity variables: employment, wages, imports, exports, production, expectations about production ... (26 series)
  2. Prices: total CPI and PPI and components, expectations about inflation (23 series)
  3. Monetary, credit and exchange rate variables: monetary aggregates, interest rates, nominal and real exchange rates, terms of trade ...(24 series)
- Full Sample: from Nov−1999 to Dec−2009.
- For marginal likelihood: $Y = \left\{ \mathbf{y}_{Nov/1999}, \cdots, \mathbf{y}_{Dec/2007} \right\}$
- For predictive likelihood:
  small training sample $Y^* = \left\{ \mathbf{y}_{Nov/1999}, \cdots, \mathbf{y}_{Dec/2004} \right\}$,
  hold-out sample $\tilde{Y} = \left\{ \mathbf{y}_{Jan/2005}, \cdots, \mathbf{y}_{Dec/2007} \right\}$.
- Initial model space: 2^{73} possible models

Empirical Application
Data

- The variable of interest is inflation measured as the twelve-month growth rate of total CPI.
- Dataset consists on 73 macroeconomic variables seasonally adjusted and transformed as annual growths or twelve-month differences.
  1. Economic activity variables: employment, wages, imports, exports, production, expectations about production ... (26 series)
  2. Prices: total CPI and PPI and components, expectations about inflation (23 series)
  3. Monetary, credit and exchange rate variables: monetary aggregates, interest rates, nominal and real exchange rates, terms of trade ...(24 series)
- Full Sample: from Nov$-$1999 to Dec$-$2009.
- For marginal likelihood: $Y = \left\{ \mathbf{y}_{Nov/1999}, \cdots, \mathbf{y}_{Dec/2007} \right\}$
- For predictive likelihood:
  small training sample $Y^* = \left\{ \mathbf{y}_{Nov/1999}, \cdots, \mathbf{y}_{Dec/2004} \right\}$,
  hold-out sample $\tilde{Y} = \left\{ \mathbf{y}_{Jan/2005}, \cdots, \mathbf{y}_{Dec/2007} \right\}$.
- Initial model space: 2^{73} possible models

- The variable of interest is inflation measured as the twelve-month growth rate of total CPI.
- Dataset consists on 73 macroeconomic variables seasonally adjusted and transformed as annual growths or twelve-month differences.
  1. Economic activity variables: employment, wages, imports, exports, production, expectations about production ... (26 series)
  2. Prices: total CPI and PPI and components, expectations about inflation (23 series)
  3. Monetary, credit and exchange rate variables: monetary aggregates, interest rates, nominal and real exchange rates, terms of trade ...(24 series)
- Full Sample: from Nov$-$1999 to Dec$-$2009.
- For marginal likelihood: $Y = \left\{ \mathbf{y}_{Nov/1999}, \cdots, \mathbf{y}_{Dec/2007} \right\}$
- For predictive likelihood:
  small training sample $Y^* = \left\{ \mathbf{y}_{Nov/1999}, \cdots, \mathbf{y}_{Dec/2004} \right\}$,
  hold-out sample $\tilde{Y} = \left\{ \mathbf{y}_{Jan/2005}, \cdots, \mathbf{y}_{Dec/2007} \right\}$.
- Initial model space: 2^{73} possible models

- First stage: Variable selection

1. Models of size up to 5 variables are considered. Model space: $\sum_{j=0}^{5}\binom{73}{j} \approx$ 16 million models.
2. $\delta = 0{,}065$

$$Y_{t+h} = \alpha + \sum_{j=0}^{5} Z_{j,t}\gamma_j + \epsilon_{t+h} \tag{25}$$

3. initial state of the chain randomly chosen (model size and explanatory variables)
4. 7 million draws (the first 2 million left out)
5. consider the 20 variables with higher posterior inclusion probabilities

- First stage: Variable selection

1. Models of size up to 5 variables are considered. Model space: $\sum_{j=0}^{5}\binom{73}{j} \approx$ 16 million models.
2. $\delta = 0{,}065$

$$Y_{t+h} = \alpha + \sum_{j=0}^{5} Z_{j,t}\gamma_j + \epsilon_{t+h} \tag{25}$$

3. initial state of the chain randomly chosen (model size and explanatory variables)
4. 7 million draws (the first 2 million left out)
5. consider the 20 variables with higher posterior inclusion probabilities

- First stage: Variable selection
  1. Models of size up to 5 variables are considered. Model space: $\sum_{j=0}^{5} \binom{73}{j} \approx$ 16 million models.
  2. $\delta = 0{,}065$

$$Y_{t+h} = \alpha + \sum_{j=0}^{5} Z_{j,t}\gamma_j + \epsilon_{t+h} \tag{25}$$

  3. initial state of the chain randomly chosen (model size and explanatory variables)
  4. 7 million draws (the first 2 million left out)
  5. consider the 20 variables with higher posterior inclusion probabilities

- First stage: Variable selection

1. Models of size up to 5 variables are considered. Model space: $\sum_{j=0}^{5}\binom{73}{j} \approx$ 16 million models.
2. $\delta = 0{,}065$

$$Y_{t+h} = \alpha + \sum_{j=0}^{5} Z_{j,t}\gamma_j + \epsilon_{t+h} \tag{25}$$

3. initial state of the chain randomly chosen (model size and explanatory variables)
4. 7 million draws (the first 2 million left out)
5. consider the 20 variables with higher posterior inclusion probabilities

- First stage: Variable selection

 1. Models of size up to 5 variables are considered. Model space: $\sum_{j=0}^{5} \binom{73}{j} \approx$ 16 million models.
 2. $\delta = 0{,}065$

$$Y_{t+h} = \alpha + \sum_{j=0}^{5} Z_{j,t}\gamma_j + \epsilon_{t+h} \tag{25}$$

 3. initial state of the chain randomly chosen (model size and explanatory variables)
 4. 7 million draws (the first 2 million left out)
 5. consider the 20 variables with higher posterior inclusion probabilities

- First stage: Variable selection
  1. Models of size up to 5 variables are considered. Model space: $\sum_{j=0}^{5} \binom{73}{j} \approx$ 16 million models.
  2. $\delta = 0{,}065$

$$Y_{t+h} = \alpha + \sum_{j=0}^{5} Z_{j,t}\gamma_j + \epsilon_{t+h} \tag{25}$$

  3. initial state of the chain randomly chosen (model size and explanatory variables)
  4. 7 million draws (the first 2 million left out)
  5. consider the 20 variables with higher posterior inclusion probabilities

- Second stage: Model selection
  1. Include 2 lags of each pre-selected variable to the dataset (60 potential predictors)
  2. Models of size up to 8 variables are considered. Model space: $\sum_{j=0}^{8} \binom{60}{j} \approx$ 3000 million models.
  3. $\delta = 0{,}13$

$$Y_{t+h} = \gamma_0 + \sum_{j=0}^{8} Z_{j,t-i}\gamma_j + \upsilon_{t+h}, i = 0, 1, 2 \qquad (26)$$

  4. initial state of the chain randomly chosen (model size and explanatory variables)
  5. 11 million draws (the first 1 million left out)
  6. consider the 20 models with higher posterior probabilities

Empirical Application ...

Implementation of BMA

- Second stage: Model selection
  1. Include 2 lags of each pre-selected variable to the dataset (60 potential predictors)
  2. Models of size up to 8 variables are considered. Model space: $\sum_{j=0}^{8} \binom{60}{j} \approx$ 3000 million models.
  3. $\delta = 0{,}13$

$$Y_{t+h} = \gamma_0 + \sum_{j=0}^{8} Z_{j,t-i}\gamma_j + \upsilon_{t+h}, i = 0, 1, 2 \tag{26}$$

  4. initial state of the chain randomly chosen (model size and explanatory variables)
  5. 11 million draws (the first 1 million left out)
  6. consider the 20 models with higher posterior probabilities

- Second stage: Model selection
  1. Include 2 lags of each pre-selected variable to the dataset (60 potential predictors)
  2. Models of size up to 8 variables are considered. Model space: $\sum_{j=0}^{8} \binom{60}{j} \approx$ 3000 million models.
  3. $\delta = 0{,}13$

$$Y_{t+h} = \gamma_0 + \sum_{j=0}^{8} Z_{j,t-i}\gamma_j + \upsilon_{t+h}, i = 0, 1, 2 \qquad (26)$$

  4. initial state of the chain randomly chosen (model size and explanatory variables)
  5. 11 million draws (the first 1 million left out)
  6. consider the 20 models with higher posterior probabilities

- Second stage: Model selection
  1. Include 2 lags of each pre-selected variable to the dataset (60 potential predictors)
  2. Models of size up to 8 variables are considered. Model space: $\sum_{j=0}^{8} \binom{60}{j} \approx$ 3000 million models.
  3. $\delta = 0{,}13$

$$Y_{t+h} = \gamma_0 + \sum_{j=0}^{8} Z_{j,t-i}\gamma_j + \upsilon_{t+h}, i = 0, 1, 2 \tag{26}$$

  4. initial state of the chain randomly chosen (model size and explanatory variables)
  5. 11 million draws (the first 1 million left out)
  6. consider the 20 models with higher posterior probabilities

- Second stage: Model selection
  1. Include 2 lags of each pre-selected variable to the dataset (60 potential predictors)
  2. Models of size up to 8 variables are considered. Model space: $\sum_{j=0}^{8} \binom{60}{j} \approx$ 3000 million models.
  3. $\delta = 0{,}13$

$$Y_{t+h} = \gamma_0 + \sum_{j=0}^{8} Z_{j,t-i}\gamma_j + \upsilon_{t+h}, i = 0, 1, 2 \tag{26}$$

  4. initial state of the chain randomly chosen (model size and explanatory variables)
  5. 11 million draws (the first 1 million left out)
  6. consider the 20 models with higher posterior probabilities

- Second stage: Model selection
  1. Include 2 lags of each pre-selected variable to the dataset (60 potential predictors)
  2. Models of size up to 8 variables are considered. Model space: $\sum_{j=0}^{8}\binom{60}{j} \approx$ 3000 million models.
  3. $\delta = 0{,}13$

$$Y_{t+h} = \gamma_0 + \sum_{j=0}^{8} Z_{j,t-i}\gamma_j + \upsilon_{t+h}, i = 0, 1, 2 \tag{26}$$

  4. initial state of the chain randomly chosen (model size and explanatory variables)
  5. 11 million draws (the first 1 million left out)
  6. consider the 20 models with higher posterior probabilities

- Second stage: Model selection
  1. Include 2 lags of each pre-selected variable to the dataset (60 potential predictors)
  2. Models of size up to 8 variables are considered. Model space: $\sum_{j=0}^{8}\binom{60}{j} \approx$ 3000 million models.
  3. $\delta = 0{,}13$

$$Y_{t+h} = \gamma_0 + \sum_{j=0}^{8} Z_{j,t-i}\gamma_j + \upsilon_{t+h}, i = 0, 1, 2 \tag{26}$$

  4. initial state of the chain randomly chosen (model size and explanatory variables)
  5. 11 million draws (the first 1 million left out)
  6. consider the 20 models with higher posterior probabilities

- Recursive forecasts for the period Jan−2008 to Dec−2009.
- Forecast horizon : from one to twelve months ahead
- weights of individual models change for each forecasting period
- Evaluation criterion : RMSE
- Evaluation and comparison of individual forecasts, BMA−pl, BMA−ml, ITMA, simple average and dynamic factor model.
- The benchmark forecast for comparison is the random walk forecast.

Empirical Application

Forecast evaluation

- Recursive forecasts for the period Jan−2008 to Dec−2009.
- Forecast horizon : from one to twelve months ahead
- weights of individual models change for each forecasting period
- Evaluation criterion : RMSE
- Evaluation and comparison of individual forecasts, BMA−pl, BMA−ml, ITMA, simple average and dynamic factor model.
- The benchmark forecast for comparison is the random walk forecast.

Empirical Application

Forecast evaluation

- Recursive forecasts for the period Jan−2008 to Dec−2009.
- Forecast horizon : from one to twelve months ahead
- weights of individual models change for each forecasting period
- Evaluation criterion : RMSE
- Evaluation and comparison of individual forecasts, BMA−pl, BMA−ml, ITMA, simple average and dynamic factor model.
- The benchmark forecast for comparison is the random walk forecast.

- Recursive forecasts for the period Jan−2008 to Dec−2009.
- Forecast horizon : from one to twelve months ahead
- weights of individual models change for each forecasting period
- Evaluation criterion : RMSE
- Evaluation and comparison of individual forecasts, BMA−pl, BMA−ml, ITMA, simple average and dynamic factor model.
- The benchmark forecast for comparison is the random walk forecast.

Empirical Application

Forecast evaluation

- Recursive forecasts for the period Jan−2008 to Dec−2009.
- Forecast horizon : from one to twelve months ahead
- weights of individual models change for each forecasting period
- Evaluation criterion : RMSE
- Evaluation and comparison of individual forecasts, BMA−pl, BMA−ml, ITMA, simple average and dynamic factor model.
- The benchmark forecast for comparison is the random walk forecast.

- Recursive forecasts for the period Jan−2008 to Dec−2009.
- Forecast horizon : from one to twelve months ahead
- weights of individual models change for each forecasting period
- Evaluation criterion : RMSE
- Evaluation and comparison of individual forecasts, BMA−pl, BMA−ml, ITMA, simple average and dynamic factor model.
- The benchmark forecast for comparison is the random walk forecast.

- For ITMA methodology, the models used in the combination where chosen as:
  1. the same selected models for BMA−pl
  2. the 20 best models according to the out-sample AIC criteria.

- For Dynamic factors model the number of factors and the lags are chosen by BIC criterion.
- A maximum of 6 factors is considered.

Empirical Application

Forecast evaluation

- For ITMA methodology, the models used in the combination where chosen as:
  1. the same selected models for BMA−pl
  2. the 20 best models according to the out-sample AIC criteria.

- For Dynamic factors model the number of factors and the lags are chosen by BIC criterion.
- A maximum of 6 factors is considered.

- For ITMA methodology, the models used in the combination where chosen as:
  1. the same selected models for BMA$-$pl
  2. the 20 best models according to the out-sample AIC criteria.

- For Dynamic factors model the number of factors and the lags are chosen by BIC criterion.
- A maximum of 6 factors is considered.

- For ITMA methodology, the models used in the combination where chosen as:
  1. the same selected models for BMA−pl
  2. the 20 best models according to the out-sample AIC criteria.

- For Dynamic factors model the number of factors and the lags are chosen by BIC criterion.
- A maximum of 6 factors is considered.

- For ITMA methodology, the models used in the combination where chosen as:
    1. the same selected models for BMA−pl
    2. the 20 best models according to the out-sample AIC criteria.

- For Dynamic factors model the number of factors and the lags are chosen by BIC criterion.
- A maximum of 6 factors is considered.

- For most horizons the forecasting errors are autocorrelated,
- An AR model was estimated for the forecasting errors series.

$$\mathbf{e}_{t+h/t} = \varphi_0 + \varphi_1 \mathbf{e}_{t+h-1/t} + \vartheta_{t+h} \tag{27}$$

- Boostrapping samples were drawn over the residuals of that model ϑ_{t+h}.
- Using the residual sample and the parameter estimates, the forecasting error sample series was constructed, $\mathbf{e}^i_{t+h/t} = \hat{\varphi}_0 + \hat{\varphi}_1 \mathbf{e}_{t+h-1/t} + \vartheta^i_{t+h}, i = 1, \cdots, 500.$
- The sample size equals the number of out-sample forecast available for each horizon.

Criterion:

- For each sample, calculate the RMSE relative to the RMSE of the random walk forecast.
- Calculate proportion of samples for which a reduction of at least 5 % in the relative RMSE is observed.

- For most horizons the forecasting errors are autocorrelated,
- An AR model was estimated for the forecasting errors series.

$$\mathbf{e}_{t+h/t} = \varphi_0 + \varphi_1 \mathbf{e}_{t+h-1/t} + \vartheta_{t+h} \tag{27}$$

- Boostrapping samples were drawn over the residuals of that model ϑ_{t+h}.
- Using the residual sample and the parameter estimates, the forecasting error sample series was constructed, $\mathbf{e}^i_{t+h/t} = \hat{\varphi}_0 + \hat{\varphi}_1 \mathbf{e}_{t+h-1/t} + \vartheta^i_{t+h}, i = 1, \cdots, 500$.
- The sample size equals the number of out-sample forecast available for each horizon.

Criterion:

- For each sample, calculate the RMSE relative to the RMSE of the random walk forecast.
- Calculate proportion of samples for which a reduction of at least 5 % in the relative RMSE is observed.

- For most horizons the forecasting errors are autocorrelated,
- An AR model was estimated for the forecasting errors series.

$$\mathbf{e}_{t+h/t} = \varphi_0 + \varphi_1 \mathbf{e}_{t+h-1/t} + \vartheta_{t+h} \tag{27}$$

- Boostrapping samples were drawn over the residuals of that model ϑ_{t+h}.
- Using the residual sample and the parameter estimates, the forecasting error sample series was constructed, $\mathbf{e}^i_{t+h/t} = \hat{\varphi}_0 + \hat{\varphi}_1 \mathbf{e}_{t+h-1/t} + \vartheta^i_{t+h}, i = 1, \cdots, 500$.
- The sample size equals the number of out-sample forecast available for each horizon.

Criterion:

- For each sample, calculate the RMSE relative to the RMSE of the random walk forecast.
- Calculate proportion of samples for which a reduction of at least 5 % in the relative RMSE is observed.

- For most horizons the forecasting errors are autocorrelated,
- An AR model was estimated for the forecasting errors series.

$$\mathbf{e}_{t+h/t} = \varphi_0 + \varphi_1 \mathbf{e}_{t+h-1/t} + \vartheta_{t+h} \tag{27}$$

- Boostrapping samples were drawn over the residuals of that model ϑ_{t+h}.
- Using the residual sample and the parameter estimates, the forecasting error sample series was constructed, $\mathbf{e}^i_{t+h/t} = \hat{\varphi}_0 + \hat{\varphi}_1 \mathbf{e}_{t+h-1/t} + \vartheta^i_{t+h}, i = 1, \cdots, 500$.
- The sample size equals the number of out-sample forecast available for each horizon.

Criterion:

- For each sample, calculate the RMSE relative to the RMSE of the random walk forecast.
- Calculate proportion of samples for which a reduction of at least 5 % in the relative RMSE is observed.

- For most horizons the forecasting errors are autocorrelated,
- An AR model was estimated for the forecasting errors series.

$$\mathbf{e}_{t+h/t} = \varphi_0 + \varphi_1 \mathbf{e}_{t+h-1/t} + \vartheta_{t+h} \tag{27}$$

- Boostrapping samples were drawn over the residuals of that model ϑ_{t+h}.
- Using the residual sample and the parameter estimates, the forecasting error sample series was constructed, $\mathbf{e}^i_{t+h/t} = \hat{\varphi}_0 + \hat{\varphi}_1 \mathbf{e}_{t+h-1/t} + \vartheta^i_{t+h}, i = 1, \cdots, 500$.
- The sample size equals the number of out-sample forecast available for each horizon.

Criterion:

- For each sample, calculate the RMSE relative to the RMSE of the random walk forecast.
- Calculate proportion of samples for which a reduction of at least 5 % in the relative RMSE is observed.

- For most horizons the forecasting errors are autocorrelated,
- An AR model was estimated for the forecasting errors series.

$$\mathbf{e}_{t+h/t} = \varphi_0 + \varphi_1 \mathbf{e}_{t+h-1/t} + \vartheta_{t+h} \tag{27}$$

- Boostrapping samples were drawn over the residuals of that model ϑ_{t+h}.
- Using the residual sample and the parameter estimates, the forecasting error sample series was constructed, $\mathbf{e}^i_{t+h/t} = \hat{\varphi}_0 + \hat{\varphi}_1 \mathbf{e}_{t+h-1/t} + \vartheta^i_{t+h}, i = 1, \cdots, 500$.
- The sample size equals the number of out-sample forecast available for each horizon.

Criterion:

- For each sample, calculate the RMSE relative to the RMSE of the random walk forecast.
- Calculate proportion of samples for which a reduction of at least 5 % in the relative RMSE is observed.

- For most horizons the forecasting errors are autocorrelated,
- An AR model was estimated for the forecasting errors series.

$$\mathbf{e}_{t+h/t} = \varphi_0 + \varphi_1 \mathbf{e}_{t+h-1/t} + \vartheta_{t+h} \tag{27}$$

- Boostrapping samples were drawn over the residuals of that model ϑ_{t+h}.
- Using the residual sample and the parameter estimates, the forecasting error sample series was constructed, $\mathbf{e}^i_{t+h/t} = \hat{\varphi}_0 + \hat{\varphi}_1 \mathbf{e}_{t+h-1/t} + \vartheta^i_{t+h}, i = 1, \cdots, 500$.
- The sample size equals the number of out-sample forecast available for each horizon.

Criterion:

- For each sample, calculate the RMSE relative to the RMSE of the random walk forecast.
- Calculate proportion of samples for which a reduction of at least 5 % in the relative RMSE is observed.

- For most horizons the forecasting errors are autocorrelated,
- An AR model was estimated for the forecasting errors series.

$$\mathbf{e}_{t+h/t} = \varphi_0 + \varphi_1 \mathbf{e}_{t+h-1/t} + \vartheta_{t+h} \tag{27}$$

- Boostrapping samples were drawn over the residuals of that model ϑ_{t+h}.
- Using the residual sample and the parameter estimates, the forecasting error sample series was constructed, $\mathbf{e}^i_{t+h/t} = \hat{\varphi}_0 + \hat{\varphi}_1 \mathbf{e}_{t+h-1/t} + \vartheta^i_{t+h}, i = 1, \cdots, 500$.
- The sample size equals the number of out-sample forecast available for each horizon.

Criterion:

- For each sample, calculate the RMSE relative to the RMSE of the random walk forecast.
- Calculate proportion of samples for which a reduction of at least 5 % in the relative RMSE is observed.

Contents

Contents

Information Theoretical Model Averaging

- Based on the AIC information criterion
  $AIC = 2\frac{k}{T} - 2\frac{l}{T} \propto 2\frac{k}{T} + \log\left(\frac{\sigma_\epsilon^2}{T}\right)$
  where l is the log likelihood, k is the number of parameters in the model and $\sigma_\epsilon^2 = \frac{1}{T}\sum_{t=1}^{T}\epsilon_t^2$ is the estimate of the variance of the residuals.
- The difference between the AIC criteria of a pair of models is an unbiased estimator of the difference of the KL distance for the two models. The KL distance, (Kullback and Leibler,1951), is defined as

$$\mathbf{I(f,g)} = \int f(x)\log\left(\frac{f(x)}{g(x/\hat{\theta})}\right)dx \tag{28}$$

  where $f(x)$ is the unknown data generating process, $g(x/\hat{\theta})$ is the given model and $\hat{\theta}$ is the estimate of the parameters.

Information Theoretical Model Averaging

- Based on the AIC information criterion
  $AIC = 2\frac{k}{T} - 2\frac{l}{T} \propto 2\frac{k}{T} + \log\left(\frac{\sigma_\epsilon^2}{T}\right)$
  where l is the log likelihood, k is the number of parameters in the model and $\sigma_\epsilon^2 = \frac{1}{T}\sum_{t=1}^{T}\epsilon_t^2$ is the estimate of the variance of the residuals.
- The difference between the AIC criteria of a pair of models is an unbiased estimator of the difference of the KL distance for the two models. The KL distance, (Kullback and Leibler,1951), is defined as

$$\mathbf{I(f,g)} = \int f(x)\log\left(\frac{f(x)}{g(x/\hat{\theta})}\right)dx \tag{28}$$

  where $f(x)$ is the unknown data generating process, $g(x/\hat{\theta})$ is the given model and $\hat{\theta}$ is the estimate of the parameters.

Information Theoretical Averaging. Continue...

- There are M forecasting models, ranking according to the KL distance.
- $\exp(-\frac{1}{2}\psi_i)$ with $\psi_i = AIC_i - \text{mín}\, AIC_j$, can be interpreted as the weight of model i to be the KL best model given that there is a certain model in set of forecasting models (M) which is the KL best model for the available data.
- The final optimal weights for model i in the set (M) is given by

$$\omega_i = \frac{\exp(-\frac{1}{2}\psi_i)}{\sum_{j=1}^{M} \exp(-\frac{1}{2}\psi_j)} \tag{29}$$

such that $\sum_{i=1}^{M} \omega_i = 1$

- There are M forecasting models, ranking according to the KL distance.
- $\exp(-\frac{1}{2}\psi_i)$ with $\psi_i = AIC_i - \min AIC_j$, can be interpreted as the weight of model i to be the KL best model given that there is a certain model in set of forecasting models (M) which is the KL best model for the available data.
- The final optimal weights for model i in the set (M) is given by

$$\omega_i = \frac{\exp(-\frac{1}{2}\psi_i)}{\sum_{j=1}^{M} \exp(-\frac{1}{2}\psi_j)} \tag{29}$$

such that $\sum_{i=1}^{M} \omega_i = 1$

Information Theoretical Averaging. Continue...

- There are M forecasting models, ranking according to the KL distance.
- $\exp(-\frac{1}{2}\psi_i)$ with $\psi_i = AIC_i - \text{mín}\, AIC_j$, can be interpreted as the weight of model i to be the KL best model given that there is a certain model in set of forecasting models (M) which is the KL best model for the available data.
- The final optimal weights for model i in the set (M) is given by

$$\omega_i = \frac{\exp(-\frac{1}{2}\psi_i)}{\sum_{j=1}^{M} \exp(-\frac{1}{2}\psi_j)} \tag{29}$$

such that $\sum_{i=1}^{M} \omega_i = 1$

- Kapetanios, et al, (2007) suggested an extension to this approach, by using the sum of square out-sample forecasting errors instead of the in-sample sum of square residuals, to construct the AIC criterion.
- The idea is to replace $\hat{\sigma}_\epsilon^2 = \frac{1}{T_1} \sum_{t=1}^{T_1} \epsilon_t^2$ by

  $\tilde{\sigma}_\epsilon^2 = \frac{1}{T_2} \sum_{t=T_1+1}^{T} (\mathbf{Y}_t - \hat{\mathbf{Y}}_{t/t-h}^i)^2,$

  where $T = T_1 + T_2$ is the full sample size, T_1 is the in-sample size, T_2 is the out-sample size and $\hat{\mathbf{Y}}_{t/t-h}^i$ is the i forecast for period t with information up to $(t-h)$, $i = 1, \cdots, M$.
- the weights ω's change for each forecast horizon, h.
- the forecast combination is given by

$$\hat{\mathbf{Y}}_{T+h} = \hat{\omega}_{1,h} \hat{\mathbf{Y}}_{T+h/T}^1 + \cdots + \hat{\omega}_{K,h} \hat{\mathbf{Y}}_{T+h/T}^M \tag{30}$$

- Kapetanios, et al, (2007) suggested an extension to this approach, by using the sum of square out-sample forecasting errors instead of the in-sample sum of square residuals, to construct the AIC criterion.
- The idea is to replace $\hat{\sigma}_\epsilon^2 = \frac{1}{T_1}\sum_{t=1}^{T_1} \epsilon_t^2$ by

  $\tilde{\sigma}_\epsilon^2 = \frac{1}{T_2}\sum_{t=T_1+1}^{T} (\mathbf{Y}_t - \hat{\mathbf{Y}}^i_{t/t-h})^2$,
  where $T = T_1 + T_2$ is the full sample size, T_1 is the in-sample size, T_2 is the out-sample size and $\hat{\mathbf{Y}}^i_{t/t-h}$ is the i forecast for period t with information up to $(t-h)$, $i = 1, \cdots, M$.
- the weights ω's change for each forecast horizon, h.
- the forecast combination is given by

$$\hat{\mathbf{Y}}_{T+h} = \hat{\omega}_{1,h}\hat{\mathbf{Y}}^1_{T+h/T} + \cdots + \hat{\omega}_{K,h}\hat{\mathbf{Y}}^M_{T+h/T} \tag{30}$$

Information Theoretical Averaging. Continue...

- Kapetanios, et al, (2007) suggested an extension to this approach, by using the sum of square out-sample forecasting errors instead of the in-sample sum of square residuals, to construct the AIC criterion.
- The idea is to replace $\hat{\sigma}_\epsilon^2 = \frac{1}{T_1}\sum_{t=1}^{T_1} \epsilon_t^2$ by

  $\tilde{\sigma}_\epsilon^2 = \frac{1}{T_2}\sum_{t=T_1+1}^{T} (\mathbf{Y}_t - \hat{\mathbf{Y}}^i_{t/t-h})^2,$
  where $T = T_1 + T_2$ is the full sample size, T_1 is the in-sample size, T_2 is the out-sample size and $\hat{\mathbf{Y}}^i_{t/t-h}$ is the i forecast for period t with information up to $(t-h)$, $i = 1, \cdots, M$.
- the weights ω's change for each forecast horizon, h.
- the forecast combination is given by

$$\hat{\mathbf{Y}}_{T+h} = \hat{\omega}_{1,h}\hat{\mathbf{Y}}^1_{T+h/T} + \cdots + \hat{\omega}_{K,h}\hat{\mathbf{Y}}^M_{T+h/T} \tag{30}$$

Information Theoretical Averaging. Continue...

- Kapetanios, et al, (2007) suggested an extension to this approach, by using the sum of square out-sample forecasting errors instead of the in-sample sum of square residuals, to construct the AIC criterion.
- The idea is to replace $\hat{\sigma}^2_\epsilon = \frac{1}{T_1} \sum_{t=1}^{T_1} \epsilon_t^2$ by

  $\tilde{\sigma}^2_\epsilon = \frac{1}{T_2} \sum_{t=T_1+1}^{T} (\mathbf{Y}_t - \hat{\mathbf{Y}}^i_{t/t-h})^2$,
  where $T = T_1 + T_2$ is the full sample size, T_1 is the in-sample size, T_2 is the out-sample size and $\hat{\mathbf{Y}}^i_{t/t-h}$ is the i forecast for period t with information up to $(t-h)$, $i = 1, \cdots, M$.
- the weights ω's change for each forecast horizon, h.
- the forecast combination is given by

$$\hat{\mathbf{Y}}_{T+h} = \hat{\omega}_{1,h}\hat{\mathbf{Y}}^1_{T+h/T} + \cdots + \hat{\omega}_{K,h}\hat{\mathbf{Y}}^M_{T+h/T} \tag{30}$$

Contents

Simple Averaging

Equal weights for each individual forecast

$$\hat{\mathbf{Y}}_{T+h} = \frac{1}{M} \sum_{i=1}^{M} \hat{\mathbf{Y}}^{i}_{T+h/T} \tag{31}$$

Contents

Dynamic Factors model

- methodology of forecasting with many predictors
- From a set of N variables, extract K common factors using principal components (Stock and Watson, 2002)
- The factors are linear combinations of the variables (eigenvectors associated to the largest eingenvalues of the VAR-COV matrix)
- The forecasting model is of the form:

$$\mathbf{Y}_{t+h} = \gamma_0 + \sum_{k=1}^{K} \mathbf{f}^k_{t-j/T_1} \gamma_k + \eta_{t+h} \tag{32}$$

  The factors may enter in the model with time t and/or with some lag(s) j and are estimated with information up to t.
- A forecasting model is estimated for each horizon h.
- The forecast is given by

$$E(\mathbf{Y}_{T+h}/T) = \hat{\gamma}_0 + \sum_{k=1}^{K} \mathbf{f}^k_{T-j/T} \hat{\gamma}_k \tag{33}$$

Dynamic Factors model

- methodology of forecasting with many predictors
- From a set of N variables, extract K common factors using principal components (Stock and Watson, 2002)
- The factors are linear combinations of the variables (eigenvectors associated to the largest eingenvalues of the VAR-COV matrix)
- The forecasting model is of the form:

$$\mathbf{Y}_{t+h} = \gamma_0 + \sum_{k=1}^{K} \mathbf{f}^k_{t-j/T_1} \gamma_k + \eta_{t+h} \tag{32}$$

  The factors may enter in the model with time t and/or with some lag(s) j and are estimated with information up to t.
- A forecasting model is estimated for each horizon h.
- The forecast is given by

$$E(\mathbf{Y}_{T+h}/T) = \hat{\gamma}_0 + \sum_{k=1}^{K} \mathbf{f}^k_{T-j/T} \hat{\gamma}_k \tag{33}$$

Dynamic Factors model

- methodology of forecasting with many predictors
- From a set of N variables, extract K common factors using principal components (Stock and Watson, 2002)
- The factors are linear combinations of the variables (eigenvectors associated to the largest eingenvalues of the VAR-COV matrix)
- The forecasting model is of the form:

$$\mathbf{Y}_{t+h} = \gamma_0 + \sum_{k=1}^{K} \mathbf{f}^k_{t-j/T_1} \gamma_k + \eta_{t+h} \tag{32}$$

  The factors may enter in the model with time t and/or with some lag(s) j and are estimated with information up to t.
- A forecasting model is estimated for each horizon h.
- The forecast is given by

$$E(\mathbf{Y}_{T+h}/T) = \hat{\gamma}_0 + \sum_{k=1}^{K} \mathbf{f}^k_{T-j/T} \hat{\gamma}_k \tag{33}$$

Dynamic Factors model

- methodology of forecasting with many predictors
- From a set of N variables, extract K common factors using principal components (Stock and Watson, 2002)
- The factors are linear combinations of the variables (eigenvectors associated to the largest eingenvalues of the VAR-COV matrix)
- The forecasting model is of the form:

$$\mathbf{Y}_{t+h} = \gamma_0 + \sum_{k=1}^{K} \mathbf{f}^{k}_{t-j/T_1} \gamma_k + \eta_{t+h} \tag{32}$$

  The factors may enter in the model with time t and/or with some lag(s) j and are estimated with information up to t.
- A forecasting model is estimated for each horizon h.
- The forecast is given by

$$E(\mathbf{Y}_{T+h}/T) = \hat{\gamma}_0 + \sum_{k=1}^{K} \mathbf{f}^{k}_{T-j/T} \hat{\gamma}_k \tag{33}$$

Dynamic Factors model

- methodology of forecasting with many predictors
- From a set of N variables, extract K common factors using principal components (Stock and Watson, 2002)
- The factors are linear combinations of the variables (eigenvectors associated to the largest eingenvalues of the VAR-COV matrix)
- The forecasting model is of the form:

$$\mathbf{Y}_{t+h} = \gamma_0 + \sum_{k=1}^{K} \mathbf{f}^k_{t-j/T_1} \gamma_k + \eta_{t+h} \tag{32}$$

  The factors may enter in the model with time t and/or with some lag(s) j and are estimated with information up to t.
- A forecasting model is estimated for each horizon h.
- The forecast is given by

$$E(\mathbf{Y}_{T+h}/T) = \hat{\gamma}_0 + \sum_{k=1}^{K} \mathbf{f}^k_{T-j/T} \hat{\gamma}_k \tag{33}$$

Dynamic Factors model

- methodology of forecasting with many predictors
- From a set of N variables, extract K common factors using principal components (Stock and Watson, 2002)
- The factors are linear combinations of the variables (eigenvectors associated to the largest eingenvalues of the VAR-COV matrix)
- The forecasting model is of the form:

$$\mathbf{Y}_{t+h} = \gamma_0 + \sum_{k=1}^{K} \mathbf{f}^{k}_{t-j/T_1} \gamma_k + \eta_{t+h} \tag{32}$$

  The factors may enter in the model with time t and/or with some lag(s) j and are estimated with information up to t.
- A forecasting model is estimated for each horizon h.
- The forecast is given by

$$E(\mathbf{Y}_{T+h}/T) = \hat{\gamma}_0 + \sum_{k=1}^{K} \mathbf{f}^{k}_{T-j/T} \hat{\gamma}_k \tag{33}$$

Contents


Total CPI Inflation in Colombia
11
10
9
8
7
6
5
4
3
2
1
Nov-99
May-00
Nov-00
May-01
Nov-01
May-02
Nov-02
May-03
Nov-03
May-04
Nov-04
May-05
Nov-05
May-06
Nov-06
May-07
Nov-07
May-08
Nov-08
May-09
Nov-09

Variables with higher posterior probabilities - Predictive Likelihood

h=1		h=3		h=6		h=9		h=12	
CPI	0.999	GASAL	0.960	GASAL	0.972	TCNMPROM	0.463	GAOTGA	0.998
RESNETAS	0.334	GAOTGA	0.546	IPI	0.694	GASAL	0.302	AEA	0.992
TERMINTE	0.149	ISNIMAOB	0.441	GAVES	0.316	CDT90DBA	0.247	NCREGUL	0.738
MBISI	0.146	IPI	0.438	GAVIV	0.258	MBISI	0.197	TERMINTE	0.305
NCNOTRAN	0.127	NCTRAN	0.290	TIBPROME	0.225	GAEDU	0.156	GATRAN	0.304
IPI	0.118	RESNETAS	0.237	TCNMPROM	0.225	CHBRUTA	0.153	UECFINAL	0.250
DTFNO90D	0.091	EFECTIV	0.149	MBKMATCO	0.161	DEPCTAHO	0.117	PCNOVIS	0.080
MBCDUR	0.088	GATRAN	0.098	ITCRIPPN	0.143	DEPCTCOR	0.116	CREDBR	0.061
ISNIMAEM	0.084	CRDOBPRI	0.093	MBCNODU	0.111	AEMIN	0.114	EFECTIV	0.054
ISNCOMIN	0.084	MBISI	0.074	GAEDU	0.110	UEFORK	0.112	EXPSITEC	0.049
DEPCTCOR	0.070	UECFINAL	0.072	ITCRIPCT	0.102	PCVIS	0.103	GAVES	0.044
MBKSA	0.067	AEA	0.069	TASACTIV	0.084	GAVIV	0.100	CDT90DBA	0.044
ITCRIPPN	0.063	MBKEQTRA	0.060	CDT90DBA	0.080	TASACTIV	0.098	EXISTEN	0.041
TIBPROME	0.062	VOLACTPE	0.051	UEFORK	0.072	ISNIMAEM	0.096	M3	0.035
MBKSI	0.061	GAVIV	0.045	CAPINDE	0.066	MBICOMLU	0.094	CRBTES	0.034
ITCRIPPT	0.058	CAPINDE	0.044	ITCRIPCN	0.061	UEMATCO	0.093	BASEMON	0.031
MBKEQTRA	0.055	ACTPROD	0.042	NCREGUL	0.054	EXPAUMPR	0.088	MBICOMLU	0.027
CRBBAN	0.054	NCREGUL	0.042	AEIMAN	0.045	M1	0.084	GASAL	0.026
EXPAUMPR	0.053	M1	0.038	EXPPRO	0.042	NCTRAN	0.082	DEPCTAHO	0.025
CREDBR	0.052	PPI	0.038	GACUL	0.042	CREDBR	0.081	GAEDU	0.020

Numbers correspond to the posterior inclusion probabilities

Variables with higher posterior probability - Marginal likelihood

h=1		h=3		h=6		h=9		h=12	
ISNCOMIN	0.984	GAOTGA	0.692	EXPSITEC	0.817	ISNIMAOB	0.906	CDT90DBA	1.000
CPI	0.942	GAVIV	0.570	TASACTIV	0.739	EFECTIV	0.835	TASACTIV	1.000
TERMINTE	0.543	CRBTES	0.505	ITCRIPCN	0.539	PPI	0.772	GAOTGA	1.000
VOLACTPE	0.506	VOLACTPE	0.491	ITCRIPCT	0.539	NCTRAN	0.728	UECINTER	1.000
NCTRAN	0.496	MBISI	0.419	ITCRIPPN	0.497	ITCRIPCN	0.508	ISNIMAOB	0.998
GASAL	0.407	TASACTIV	0.408	ITCRIPPT	0.460	ITCRIPPT	0.313	GAVES	0.002
ITCRIPPT	0.161	CDT90DBA	0.407	EXPAUMPR	0.337	GAOTGA	0.098	TOTALDEP	0.000
UECFINAL	0.142	DTFNO90D	0.270	SECONOM	0.121	DEPCTAHO	0.097	MBKSI	0.000
PPI	0.133	AEMIN	0.205	GAVES	0.068	DTFNO90D	0.096	PPI	0.000
DTFNO90D	0.107	M1	0.197	CDT90DBA	0.067	PBPRODCO	0.091	CRDOBPRI	0.000
NCNOTRAN	0.085	ITCRIPPT	0.178	PBM	0.063	ITCRIPPN	0.072	GAVIV	0.000
ITCRIPCT	0.054	PBPRODCO	0.178	UEFORK	0.059	AEIMAN	0.070	NCTRAN	0.000
AEA	0.042	UEMATCO	0.082	TCNMPROM	0.038	ISRSINT	0.062	GATRAN	0.000
M1	0.030	TERMINTE	0.081	CRDOBPRI	0.029	RESNETAS	0.040	MBKMATCO	0.000
M3	0.028	UECINTER	0.066	EFECTIV	0.028	UECFINAL	0.035	CRBCORP	0.000
AEMIN	0.027	NCTRAN	0.065	GATRAN	0.028	PCNOVIS	0.025	TCNMPROM	0.000
CRDOBPRI	0.027	AEIMAN	0.034	UEMATCO	0.025	NCREGUL	0.022	DTFNO90D	0.000
ITCRIPCN	0.026	GAEDU	0.032	EXPPRO	0.025	EXISTEN	0.019	M1	0.000
PBM	0.018	DEPCTCOR	0.013	AEA	0.017	SECONOM	0.016	M2	0.000
GAOTGA	0.014	GACUL	0.013	TIBPROME	0.014	EXPPRO	0.016	UEFORK	0.000

Numbers correspond to the posterior inclusion probabilities

Models with higher posterior probability - Predictive likelihood - h=12

Variable	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	M11	M12	M13	M14	M15	M16	M17	M18	M19	M20
GAOTGA	0,1,2	0	1	0,1	0	0	0,1	0	0,1	1	0,1	0	0	0,2	1	0,1	0	0	0	1
AEA	0	0,2	0,2	0,2	0,2	0	0	0	0	0,2	0	0,1	0	0,2	2	0	0,2	0,1	0	0
NCREGUL	1	0	0	1	0	1		1	1	0	1	0		0	0	0	0	0		1
UECFINAL		2	0,2	0	2	0	1	0		2		1	0	0	2		2	1		
TERMINTE	0	1				0					0	0	0			0		0		
BASEMON		1			1	1	1						1			1	1			
DEPCTAHO				1			2		0	1									0	0
EXPSITEC				1				1	2			2					1			2
GATRAN		2					0						1						1	2
EFECTIV			2		2	2		2							2					
GAVES							0						0		1				1	
M3											1			2		2				2
CDT90DBA								0			1						0			
EXISTEN										2				2					0,2	
PCNOVIS			0												1					
MBICOMLU						1														
CREDBR															2					
GASAL	0,1																			
Posterior Prob.	0.120	0.070	0.059	0.058	0.055	0.049	0.049	0.047	0.046	0.044	0.042	0.042	0.042	0.041	0.040	0.040	0.040	0.040	0.039	0.038

* the numbers in cells are the lags of each variable in the model

Forecast Evaluation - RSME

	h=1	h=2	h=3	h=4	h=5	h=6	h=7	h=8	h=9	h=10	h=11	h=12
BMA-pl	***0.338***	0.807	1.777	***1.083***	1.114	***0.874***	***1.000***	1.560	1.804	1.022	0.962	1.021
BMA-ml	0.457	1.066	1.169	1.601	1.557	1.667	1.880	2.201	1.153	1.518	1.167	1.284
ITMA1	***0.358***	1.007	1.819	1.636	1.762	1.355	***1.527***	2.017	2.110	1.402	1.050	1.088
ITMA2	***0.360***	***0.641***	0.950	***0.916***	0.856	***0.746***	***0.789***	0.932	0.913	0.724	0.671	0.739
Simple average	***0.358***	1.007	1.819	1.635	1.762	1.355	***1.527***	2.017	2.110	1.402	1.050	1.088
Dynamic factors	***0.307***	***0.728***	***1.071***	1.400	1.732	1.974	2.003	2.097	2.203	2.213	2.189	2.326
Random Walk	0.473	0.851	1.175	1.457	1.742	2.024	2.278	2.473	2.608	2.712	2.789	2.801

Numbers in bold and italic correspond to the cases where MDM test for equal forecast ability compared to the random walk forecast is rejected

BMA-pl refers to BMA combination using predictive likelihood

BMA-ml refers to BMA combination using marginal likelihood

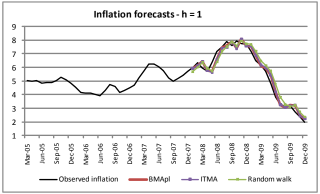
ITMA1 refers to the information theoretic model averaging combination of models selected by BMA

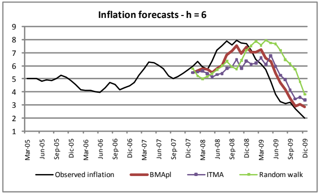
ITMA2 refers to the information theoretic model averaging combination of models selected by ITMA

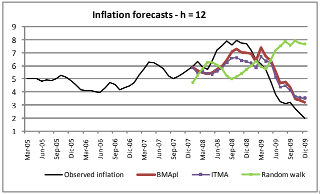
Bootstrapping forecasting errors

Forecast Combination	h=1	h=2	h=3	h=4	h=5	h=6	h=7	h=8	h=9	h=10	h=11	h=12
BMA_PL	***1.000***	0.528	0.154	0.806	***0.924***	***0.994***	***0.990***	0.898	0.734	***1.000***	***0.998***	***0.998***
BMA_ml	0.458	0.304	0.360	0.374	0.578	0.692	0.578	0.456	0.290	0.542	***0.970***	***0.988***
ITMA1	***0.986***	0.352	0.140	0.284	0.450	0.770	0.756	0.520	0.542	0.488	***0.950***	***0.946***
ITMA2	***0.996***	0.884	0.710	0.768	***0.982***	***1.000***	***1.000***	***0.998***	***1.000***	***1.000***	***1.000***	***0.998***
simple average	***0.990***	0.348	0.150	0.304	0.396	0.798	0.760	0.542	0.574	0.480	***0.944***	***0.940***
Dinamic Factors	***1.000***	***0.984***	***0.960***	0.550	0.810	0.726	0.602	0.390	0.320	0.280	0.356	0.356

% of samples with reduction in RMSE of at least 5% relative to the random walk forecast.


Inflation forecasts - h = 1
9
8
7
6
5
4
3
2
1
Mar-05 Jun-05 Sep-05 Dec-05 Mar-06 Jun-06 Sep-06 Dec-06 Mar-07 Jun-07 Sep-07 Dec-07 Mar-08 Jun-08 Sep-08 Dec-08 Mar-09 Jun-09 Sep-09 Dec-09
Observed inflation
BMApl
ITMA
Random walk


Inflation forecasts - h = 6
9
8
7
6
5
4
3
2
1
Mar-05 Jun-05 Sep-05 Dic-05 Mar-06 Jun-06 Sep-06 Dic-06 Mar-07 Jun-07 Sep-07 Dic-07 Mar-08 Jun-08 Sep-08 Dic-08 Mar-09 Jun-09 Sep-09 Dic-09
Observed inflation
BMApl
ITMA
Random walk


Inflation forecasts - h = 12
9
8
7
6
5
4
3
2
1
Mar-05 Jun-05 Sep-05 Dic-05 Mar-06 Jun-06 Sep-06 Dic-06 Mar-07 Jun-07 Sep-07 Dic-07 Mar-08 Jun-08 Sep-08 Dic-08 Mar-09 Jun-09 Sep-09 Dic-09
Observed inflation
BMApl
ITMA
Random walk

Contents

Concluding Remarks

- An alternative approach of forecasting based on a large dataset of potential predictors is implemented for the Colombian inflation.
- BMA is a useful and consistent way to select variables and models with high predictive power.
- Variables and models chosen as good predictors differ whether they are selected using marginal or predictive likelihood.
- BMA outperforms the random walk forecast and simple average combination and is a good competitor of the frequentist forecast combination, ITMA.

Concluding Remarks

- An alternative approach of forecasting based on a large dataset of potential predictors is implemented for the Colombian inflation.
- BMA is a useful and consistent way to select variables and models with high predictive power.
- Variables and models chosen as good predictors differ whether they are selected using marginal or predictive likelihood.
- BMA outperforms the random walk forecast and simple average combination and is a good competitor of the frequentist forecast combination, ITMA.

Concluding Remarks

- An alternative approach of forecasting based on a large dataset of potential predictors is implemented for the Colombian inflation.
- BMA is a useful and consistent way to select variables and models with high predictive power.
- Variables and models chosen as good predictors differ whether they are selected using marginal or predictive likelihood.
- BMA outperforms the random walk forecast and simple average combination and is a good competitor of the frequentist forecast combination, ITMA.

Concluding Remarks

- An alternative approach of forecasting based on a large dataset of potential predictors is implemented for the Colombian inflation.
- BMA is a useful and consistent way to select variables and models with high predictive power.
- Variables and models chosen as good predictors differ whether they are selected using marginal or predictive likelihood.
- BMA outperforms the random walk forecast and simple average combination and is a good competitor of the frequentist forecast combination, ITMA.

Concluding Remarks ...

- The gain of using BMA in reducing the forecasting error is observed as the horizon increases, what is very helpful for our purpose of forecasting inflation in the medium term.
- Forecast combinations whose weigths are based on the predictive ability of the models reduces the forecasting error relative to combinations whose weights are based on the fit of the model.
- BMA based on predictive likelihood is for some horizons better than ITMA when both combinations are made over the models selected by BMA
- Selecting models by ITMA criteria, the combined forecast obtained by the BMA weights performs better for most horizons.
- The model selection by ITMA criteria, implicitly is using Bayesian techniques when visiting the models for which the out-sample AIC criterion is evaluated. This makes the comparison unfair.

Concluding Remarks ...

- The gain of using BMA in reducing the forecasting error is observed as the horizon increases, what is very helpful for our purpose of forecasting inflation in the medium term.
- Forecast combinations whose weigths are based on the predictive ability of the models reduces the forecasting error relative to combinations whose weights are based on the fit of the model.
- BMA based on predictive likelihood is for some horizons better than ITMA when both combinations are made over the models selected by BMA
- Selecting models by ITMA criteria, the combined forecast obtained by the BMA weights performs better for most horizons.
- The model selection by ITMA criteria, implicitly is using Bayesian techniques when visiting the models for which the out-sample AIC criterion is evaluated. This makes the comparison unfair.

Concluding Remarks ...

- The gain of using BMA in reducing the forecasting error is observed as the horizon increases, what is very helpful for our purpose of forecasting inflation in the medium term.
- Forecast combinations whose weigths are based on the predictive ability of the models reduces the forecasting error relative to combinations whose weights are based on the fit of the model.
- BMA based on predictive likelihood is for some horizons better than ITMA when both combinations are made over the models selected by BMA
- Selecting models by ITMA criteria, the combined forecast obtained by the BMA weights performs better for most horizons.
- The model selection by ITMA criteria, implicitly is using Bayesian techniques when visiting the models for which the out-sample AIC criterion is evaluated. This makes the comparison unfair.

Concluding Remarks ...

- The gain of using BMA in reducing the forecasting error is observed as the horizon increases, what is very helpful for our purpose of forecasting inflation in the medium term.
- Forecast combinations whose weigths are based on the predictive ability of the models reduces the forecasting error relative to combinations whose weights are based on the fit of the model.
- BMA based on predictive likelihood is for some horizons better than ITMA when both combinations are made over the models selected by BMA
- Selecting models by ITMA criteria, the combined forecast obtained by the BMA weights performs better for most horizons.
- The model selection by ITMA criteria, implicitly is using Bayesian techniques when visiting the models for which the out-sample AIC criterion is evaluated. This makes the comparison unfair.

Concluding Remarks ...

- The gain of using BMA in reducing the forecasting error is observed as the horizon increases, what is very helpful for our purpose of forecasting inflation in the medium term.
- Forecast combinations whose weigths are based on the predictive ability of the models reduces the forecasting error relative to combinations whose weights are based on the fit of the model.
- BMA based on predictive likelihood is for some horizons better than ITMA when both combinations are made over the models selected by BMA
- Selecting models by ITMA criteria, the combined forecast obtained by the BMA weights performs better for most horizons.
- The model selection by ITMA criteria, implicitly is using Bayesian techniques when visiting the models for which the out-sample AIC criterion is evaluated. This makes the comparison unfair.

Concluding Remarks ...

- The variables chosen as best predictors for inflation have change to some extent over the last two years, especially for the predictive likelihood (40 % of the variables are selected with the full sample). For marginal likelihood, it seems that the forces driving inflation have not changed over time (70 % of the variables are selected with the full sample).

Concluding Remarks ...

for future research,

- How often the selection of variables and models should be done in order to continue applying this methodology on a regular basis, given that the results are influenced by the sample, specially when using predictive likelihood.
- Which priors and which algorithm should be used to select the variables and models, having into account the findings of Ohara and Sillampaa,(2009), "the performance of the method depends on the priors and how it is implemented"
- How many models to combine?
- Which transformation should be used for the response variable and the predictors?.

Concluding Remarks ...

for future research,

- How often the selection of variables and models should be done in order to continue applying this methodology on a regular basis, given that the results are influenced by the sample, specially when using predictive likelihood.
- Which priors and which algorithm should be used to select the variables and models, having into account the findings of Ohara and Sillampaa,(2009), "the performance of the method depends on the priors and how it is implemented"
- How many models to combine?
- Which transformation should be used for the response variable and the predictors?.

Concluding Remarks ...

for future research,

- How often the selection of variables and models should be done in order to continue applying this methodology on a regular basis, given that the results are influenced by the sample, specially when using predictive likelihood.
- Which priors and which algorithm should be used to select the variables and models, having into account the findings of Ohara and Sillampaa,(2009), "the performance of the method depends on the priors and how it is implemented"
- How many models to combine?
- Which transformation should be used for the response variable and the predictors?.

Concluding Remarks ...

for future research,

- How often the selection of variables and models should be done in order to continue applying this methodology on a regular basis, given that the results are influenced by the sample, specially when using predictive likelihood.
- Which priors and which algorithm should be used to select the variables and models, having into account the findings of Ohara and Sillampaa,(2009), "the performance of the method depends on the priors and how it is implemented"
- How many models to combine?
- Which transformation should be used for the response variable and the predictors?.

END.

Thanks!