Model Uncertainty and Forecasting, a Practitioner Point of View *

B. Bellone[†], E. Michaux[‡§]

February 2006
Revised November 2006

Abstract

Should we run one regression forecast? We confront the Bayesian Model Averaging (BMA) with two major automatic forecasting procedures: the dynamic factor model (DFA) and the "general-to-specific" (GETS) algorithm, to gauge whether introducing Bayesian model uncertainty significantly improves forecasting. We consider those methods, including the "Median BMA" strategy to forecast quarterly US GDP growth on an original monthly basis. All models consistently select the same coincident and leading series, giving a robustness check of the best US GDP predictors. We also assess relative model performance in and out-of-sample using forecast accuracy tests and directional change metrics. Results suggest that the simple BMA approach, the GETS strategy or the DFA model tend to be equivalent.

Key Word: Automatic Model Selection, Bayesian Model Averaging, GETS, Factor models.

JEL Classification: C11, C51, C52, C53

*We are grateful to Eric Dubois, for his comments, his support and permanent improvement of the GROCER toolbox. We benefited from useful comments on a preliminary version of this paper thanks to Annabelle Mourougane and Florian Pelgrin. We would also like to thank Nathalie Cremades-Triaire, Adrien Binet and Mana Tchitsaz, for their help in the earlier beginning of this research, during a six-month ENSAE working group, dedicated to study BMA models. Programs used in this paper are available on request. The views expressed in this working paper are those of the authors and do not necessarily reflect those of the OECD or Halbis Capital Management.

[†]OECD, Economics Department

[‡]Halbis Capital Management

[§]Correspondence to : Emmanuel Michaux, Halbis Capital Management, France, E-mail : emmanuel.michaux@halbis.com

1 Introduction

How should forecasting models be selected? Do methods introducing Bayesian model uncertainty significantly improve forecasting? A fast growing literature has recently emerged on Bayesian Model Averaging (BMA). This method, initiated by Raftery et al. (1997), offers new solutions to select linear forecasting models as the authors first confronted this new approach with the *Stepwise* Method. Fernandez et al. (2001b) studying *growth regressions* conclude that the BMA is the best way of selecting meaningful regressors. Jacobson and Karlsson (2004), Wright (2003) assess BMA forecasting performances and conclude that BMA does surpass other competing methods. However, some major issues remain. First, the models used as benchmarks (autororegressive or random walk models) are based on a minimal set of information and modelling. As the random walk model is not likely to be always the best forecasting tool, those approachs might not fit the forecasters, who expect to measure whether a new method challenges the predictive performance of their own reference model. Second, those papers do not produce statistical tests of model dominance such as Diebold and Mariano (1995), to gauge relative predictive performances.

In this paper, we first confront two BMA procedures[1] (Raftery et al. (1997), Barbieri and Berger (2004)) with two major methods based on "one regression model beats all" principle. On one hand, we examine whether large information data set treatments proposed by Stock and Watson (1998) through dynamic factor analysis (DFA) outclass the BMA. On the other hand, we challenge the BMA results with the automatic model selection procedure called *general-to-specific* (GETS), as introduced by Hoover et Perez (1999) and developped by Krolzig and Hendry (2001). Whereas this latter methodology was not designed to cope with large data sets of exogenous variables, Hendry and Krolzig (2005) showed that this algorithm could be extended, providing that multicolinearity might be sufficiently controlled. We provide, as an empirical application, a forecast of the American quarterly GDP, with a specific treatment on monthly hard data and survey indicators.

The paper is organized as follows. The second section introduces the three different approachs (BMA, DFA "à la Stock and Watson" and GETS). We discuss their theoretical background and potential caveats. The third section presents the data with the related original monthly approach. It also provides the empirical results on model selection. The fourth section studies relative forecasting performances through accuracy tests. The last section concludes.

2 Forecasting with many predictors: a brief review

2.1 Model uncertainty and the Bayesian Model Averaging

Is there a best model to describe economic interrelations? Extracting a single model as a unique selection from a subest of regressors is the typical approach. Yet, as underlined by Brock et al. (2004), this ignores a major source of uncertainty: uncertainty about the model itself. These authors consider that, *"especially in macroeconomics, model uncertainty is sufficiently severe that disparate models should be regarded as potential candidates for the*

[1] All estimates are replicable with the `bma`, `automatic` and `dfasw` functions, developed under Scilab, by Michaux and Dubois. All programs but `dfasw`, are included in GROCER, version 1.2, the Eric Dubois'econometric toolbox, see Dubois (2004), which is donwloadable on http://dubois.ensae.net/grocer.html. Scilab is a free program, available at www.scilab.org.

true or best model". This argument may be related to theoretical competing representations of a same model (backward or forward looking expectations for instance, or because different way of modelling dynamics in a given linear framework). Model averaging represents a natural response to these issues. As recalled in Jacobson and Karlsson (2004), the Bayesian Model Averaging allows a parallel treatment of model and parameter uncertainty. It also supports the idea that forecast performance can be improved by combining forecasts from different models with a rigorous statistical foundation.

2.1.1 The BMA approach, a reminder

Be $y = (y_1, ..., y_T)'$ the quantity of interest and $\{X_1, ..., X_n\}$ a set of n potential regressors, where $X_i = (x_{i,1}, ..., x_{i,T})'$. We look for a representation in a linear framework, where $X^* = \{X_1^*, ..., X_q^*\}$ is a $(q, 1)$ subset of $\{X_1, ..., X_n\}$ [2] such as:[3]

$$y_{t+h} = X_t^* \beta^* + \varepsilon_{t+h}$$

Let us define a given model M_i and $\mathcal{M}$ the finite set of models M_γ. In theory, it includes 2^n different elements. Here, $\gamma \in \{0,1\}^n$ is a summarized representation of the regressors entering the model. For all $j = 1, ..., n$, $\gamma_j = 1$ (resp. $\gamma_j = 0$) means that X_j does belong (resp. not) to the model. Said differently, M_γ is a short representation of the following model:

$$y|\beta_\gamma, \sigma^2 \sim \mathcal{N}(X_\gamma \beta_\gamma, \sigma^2 I) \tag{1}$$

with $\theta = (\beta_\gamma, \sigma^2)$ the unknown vector of parameters.

Given the set $\mathcal{M} = \{M_1, ..., M_{2^n}\}$ of possible models, we may associate $p(M_\gamma)$ and $p(\theta_\gamma|M_\gamma)$ the prior probability for a given model and distribution of the related parameters. Be $L(y|\theta_\gamma, M_\gamma)$ the likelihood related to the model (1) then, the posterior probability of the model is given by:

$$\begin{aligned} p(M_{\gamma 0}|y) &= \frac{m(y|M_{\gamma 0})p(M_{\gamma 0})}{\sum_\gamma m(y|M_\gamma)p(M_\gamma)} \qquad (2) \\ &= \left[\sum_\gamma \frac{P(M_\gamma)}{P(M_{\gamma 0})}\frac{m(y|M_\gamma)}{m(y|M_{\gamma 0})}\right]^{-1} \\ &= \left[\sum_\gamma \frac{P(M_\gamma)}{P(M_{\gamma 0})} B_{\gamma 0,\gamma}\right]^{-1} \end{aligned}$$

where $m(y|M_\gamma)$ is the marginal likelihood, conditional to the model M_γ,

$$m(y|M_\gamma) = \int_\Theta L(y|\theta_\gamma, M_\gamma)p(\theta_\gamma|M_\gamma)d\theta_\gamma \tag{3}$$

and $B_{\gamma 0,\gamma}$ is the Bayes factor.[4]

[2] All BMA model specifications include a constant term. It should be thus excluded from the permutation process. Note that the subset dimension q is random and may vary between 1 and n.

[3] h is limited to $\{0, 1\}$ to avoid serial correlation issues

[4] The Bayes factor may be reconsidered such as $B_{\gamma 0,\gamma} = \frac{m(y|M_\gamma)}{m(y|M_{\gamma 0})} = \frac{P(M_{\gamma 0}|y)}{P(M_\gamma|y)} / \frac{P(M_{\gamma 0})}{P(M_\gamma)}$.

The posterior probability that the variable i belongs to the "true" model:

$$p(x_i|y) = \sum_{\gamma} \mathbf{1}(x_i \in M_\gamma) p(M_\gamma|y) \tag{4}$$

where $\mathbf{1}(x_i \in M_\gamma)$ is one if x_i is included in model γ or zero otherwise. If we come back to our initial objective, the mininum mean squared error forecast of our quantity of interest (here y_{t+h}) taking into account bayesian uncertainty is given by the following average:[5]

$$\mathrm{E}(y_{t+h}|y) = \sum_{\gamma} \mathrm{E}(y_{t+h}|y, M_\gamma) p(M_\gamma|y) \tag{5}$$

Under the Bayesian approach, there is no specific rule to qualify series according to their posterior distribution and to reduce models. Barbieri and Berger (2004) show that for selection among normal linear models, the optimal predictive model is often the *median probability model*. This model includes variables with posterior probability greater than or equal to 1/2 of being in a model. We propose this approach as a parcimonious benchmark for the BMA in our empirical estimates. From a practitioner point of view, this approach offers a remarkable advantage: it delivers a unique model with results clearly interpretable and easy to communicate.

2.1.2 Prior, bayesian estimates and limits

In this paper, we stick to the the Fernandez et al. (2001a) approach using g-priors and considering only models with a constant. We use non-informative priors for the ordinary regression parameters through g-prior:

$$\beta_\gamma|\sigma^2, M_\gamma \sim \mathcal{N}\left(0, \sigma^2 \left(g_\gamma X'_\gamma X_\gamma\right)^{-1}\right) \tag{6}$$

with g_γ accounting for the share of information,[6] which is *a priori* available in the studied sample.[7] As Raftery et al. (1997), Fernandez et al. (2001a) consider the same prior for the constant vector of the matrix β_γ. The mean is defined as follows: $\mu_\gamma = (m_1, 0, ..., 0)$, with a non informative law $p(\mu) \propto 1$. We follow the common g-prior specification: $g_\gamma = g = q^{-2}$ as $n \leq q^2$.

Contrary to Raftery et al. (1997), where prior laws are conjugate and hyperparameters have a considerable influence on the posterior distribution and the Bayes factor, *prior distribution* of σ^2 is non informative: $p(\sigma) \propto \sigma^{-2}$. We may derive, in this framework, the posterior distribution:

$$p(Y|\mu_\gamma, g_\gamma, X_\gamma, \gamma) \propto \left(\frac{g_\gamma}{g_\gamma+1}\right)^{q_\gamma/2} \times \left[\frac{1}{g_\gamma+1} y' M_{X_\gamma} y + \frac{g_\gamma}{g_\gamma+1}(y-\overline{y}\mathbf{1})'(y-\overline{y}\mathbf{1})\right]^{\frac{n-1}{2}}$$

[5] The coefficient $\hat{\beta}^j_{\mathrm{BMA}}$ related to the regressor j may be extracted in the same way:

$$\hat{\beta}^j_{\mathrm{BMA}} = \mathrm{E}\left(\beta^j_\gamma|y\right) = \sum_\gamma \hat{\beta}^j_\gamma p(M_\gamma|y)$$

[6] In this formulation of the prior X_γ is supposed to be strictly exogenous and residuals are assumed to be *iid.* g-prior does not allow for lagged endogenous variable, and is not compatible with GLS like variance-covariance matrix. This can be seen as a straightforward limit of the method when used for forcasting time series.

[7] For instance, if $g = 0.1$, 10% of information of the sample is taken into account.

where $M_X = I - X(X'X)^{-1}X'$. As Raftery et al. (1997), we use a MCMC algorithm called "MC3" to approximate the distribution of M_γ and to produce BMA estimates.

Bayesian estimation strategies may suffer from serious caveats. First, calculations may be extremly costly and may require strong assumptions, especially when dealing with Bayesian priors. Second, endogenous lagged series that could correct any kind of serial correlation or a high level of persistence in the series, can not be included in the set of regressors because strict exogeneity is assumed.[8] This may seriously reduce the scope of the information set. Third, exploring a too large number of potential models to minimise the RMSE is also an evident limit of the BMA approach. All these arguments call for a confrontation with more tractable methods.

2.2 Approximate dynamic factor models

Following Stock and Watson (2002), diffusion indexes and the dynamic factor models offer an alternative method to incorporate all relevant information. Dynamic factor models collect an information from a large set of indicators. They rival models with one or multiple indicators. In this case, it is assumed that the large panel of time series at date t follows a factor structure, namely that $X_t = \Lambda F_t + \epsilon_t$, where F_t, of dimension $(T \times k)$ with k smaller than the number of variables n, are (a relatively small number of) unobserved factors that summarize the systematic information in the data set (see Stock and Watson, (1998) for details). In a second step F_t, is directly introduced as a common unobserved factor in the following equation:

$$y_{t+h} = \beta' F_t + \varepsilon_{t+h}$$

We extract factors from contemporaneous and lagged variables as in Stock and Watson (1998).

From a practioner point of view, the factor model raises the issue of interpretability. To a certain extent, it is a specific constrained OLS. Considering that the k first factors are linear combinations of the regressors, the preceding model can indeed be written accordingly:

$$y_{t+h} = \beta' F_t + c + \varepsilon_{t+h} = \beta' \left((x_t, ...x_{t-k})' W \right) + c + \varepsilon_{t+h}, \tag{7}$$

with W is the $(n \times k)$ derived matrix of communalities from the principal component analysis, on the stacked regressors. Why should the "best model" be initially constrained? The *general-to-specific* approach with its multi-path reduction process challenges this view.

2.3 Model selection and the *general-to-specific* algorithm

Hoover and Perez (1999) advocate a totally different approach through the LSE methodology, the so called *general-to-specific* model selection. Starting from a general dynamic statistical model, automatic statistical tests and procedures are used to reduce its complexity by eliminating statistically-insignificant variables, provided that the model passes chosen specification tests. This algorithm should lead to a few models which can be selected by encompassing tests. An information criterion, such as Akaike (AIC), may be used to discriminate final models.

[8] Unless we strongly constraint the prior distribution and the structure of the variance, we cannot derive the closed form for the posterior of such a model. Relaxing these assumption can be made at the cost of an increase in computational time.

Econometricians may test a limited number of models, whereas there are potentially 2^n challengers. Hendry and Krolzig (2001), show through Monte-Carlo experiments that, starting from an information set including the "true model", the GETS algorithm leads to the right specification, providing that diagnostic tests and significance levels are correctly selected. Krolzig and Hendry (2005) give hints to deal with intractable problems such as regressions with more regressors than observations in regression analysis or with perfectly colinear regressors. Readers may refer to Krolzig and Hendry (2001) for a full presentation and to Dubois (2004) who provides a simplified introduction to the algorithm. Note that, in this exercise, lagged dependent variables are not included, because the BMA specification cannot support endogenous lagged series as regressors. This partially constraints the potential set of information.[9]

GETS estimations are made under the *liberal strategy* and *conservative strategy* The main difference lie in the p-value rejection of tests: the first approach (resp. the second) is based on a 1% (resp. 5%) threshold. See Krozlig and Hendry (2005) for further explanations. Specification tests are Doornick and Hansen normality test, Chow predictive failure test for a break at 50% and 90% of the sample, LM test of 1 to 4 order residual autocorrelation and quadratic in regressors hederoscedasticity test.

3 Empirical results

3.1 Forecasting US GDP on a monthly basis

As described in Appendix, we start from a set of 45 regressors based upon 19 US variables.[10] Time series were selected from prior empirical estimates and a large review of the literature. A first preselection from a larger database was needed, because of calculation capacity constraints related to the Bayesian or GETS algorithms. Colinearity issues and efficiency were also at stake. For instance, employment series were dropped out because of more robuts results using the unemployment series and labor market surveys.

With 45 regressors, the number of potential models is comprehensive: 2^{45} combinations might be investigated.[11] That is more than 35 trillion specifications to explore. Most related series belong to the Conference Board's set of coincident and leading indicators. We added diffusion indices series of the ISM, the Help Wanted Advertising Index, consumer goods and non defense goods manufacturing orders series.[12] See Appendix, page 13 for a detailed presentation of the transformations, description and mnemonics related to the data set. The dependent variable is the NIPA US real quarterly growth rate of the gross domestic product

[9]In our exercise serial correlation is controlled by adding exogenous lagged series.

[10]Appendix, page 13 provides a brief description of the 19 series, label series, mnemonics, date of publication in relation to the reference date, transformations for stationarity. The series were extracted from the Thomson Datastream Economics database (september 2005 vintage). We limit our estimates to 2003Q4 to limit risks related to revision of NIPA preliminary estimates.

[11]The size of our dataset may look small compared with huge samples, common in the litterature. As shown by Boivin and Ng (2006), a large database may not always be desirable, as the presence of variables sharing the same redundant information is likely. This raises risk of rising serial correlation due to idiosyncratic components, which is detrimental to the variability of the principal component. In our experience, we may avoid this particular caveat, by first differencing the selected series, which dramatically reduces multicolinearity.

[12]The Help wanted Index might have been a major predictor over the past 40 years. Because of a recent structural break, may be due to a technological switch on the internet of job ads since 2003, there has been rising uncertainty about its future predictive power. This does not challenge either our methodological approach, either the present results.

(GDP, July 2005 vintage), chained link at 1996 prices, available on the BEA web site. All series are prior transformed with the log(.) operator, except ISM diffusion indices and interest rates. They are first-differenced to limit multi-colinearity risks. There is one exception : following the litterature, the term spread between the 10 year treasury bond and the 3 month Fed fund interest rate has been smoothed on a quarterly basis (quarterly average).

We provide a current (resp. next quarter) US GDP growth forecast using monthly information available the second month of the current quarter. There are two justifications for such a choice. First, because of a carry-over effect, more than 80 percent of the information to deduce the current quarter growth rate is available on the second month.[13] Second, Dubois and Michaux (2006) showed, although in a different context, that starting simulations from the second month rather than from the third month does not deteriorate future forecasts. Monthly regressors are thus described in the form of $x_{m,q}$ where $m = 1, 2, 3$ is the first, (resp. second and third) month of the current quarter q (resp. $q - k$ lagged by k quarters). Note that variables are integrated in accordance with their publication delay. All forecasting comparisons are performed using out-of-sample simulations.[14] First, in sample estimates are conducted between 1970-Q1 and 1990-Q4, then recursive out-of-sample estimates are conducted for 1990-Q1 to 2003-Q4 to produce forecasts for the current and the next quarter as well.[15]

3.2 Estimates and model selection

First, we examine the DFA results. We run dynamic factor analysis (DFA) on the panel of regressors (45 contemporaneous and lagged series) to extract three factors (see Table 4 in Appendix page 14). A rapid analysis leads to the following conclusion: the first factor is highly correlated with two major leading indicators (the term spread lagged by two quarters and the first month housing starts lagged by one quarter) and with major coincident indicators (the industrial production index, manufacturing new orders, manufacturing sales trade, and series related to employment such as the unemployment rate, the jobs hard to get, the help wanted advertising index series).[16] So, it provides a description of coincident activity focusing mainly on output and employment.[17]

Table 6 reports estimations from GETS starting from the same 45 series data set.[18] We note that manufacturing sales trade and industrial production series seem the most informative series. Interestingly, the third month of the previous quarter (resp. the first month of the current quarter) outcomes play a major role in the forecast. This may be due to "carry over effects" at the beginning of each quarter. Note that financial variables play a minor role in the coincident model. With no surprise, the next quarter models include long leading

[13]Taking into account the weights associated to the available past monthly growth rates the second month of a quarter, this is exactly equal to 88 per cent, that is $1 - \frac{1/3}{(2/3+4/3+1)}$.

[14]All models are estimated on data samples, which are prior to each forecast period.

[15]We could not conduct real time estimates, as a complete data base was not available. To our knowledge, a real time approach could only be restricted to the data base of Croushore and Starck (1999), reducing the size of the information set.

[16]Series available the third month of the past quarter and to a lesser extent the first month of the current quarter seem to be the more correlated to the first factor.

[17]The second and third factors are less interpretable. The second axis seems to reflect a dynamical dimension opposing current series to lagged ones. The third axis could be interpreted as a specific "consumption driven factor", with the lagged manufacturing sales trade and consumer expectations the stronger communalities.

[18]Dealing with futur quarter forecasts, selected final models residuals are not normal according to Doornick-Hansen test.

series (the term spread and money supply), short leading indicators (S&P 500, housing starts, initial claims, ISM supplier deliveries) but also more coincident indicators (the help wanted advertising index, industrial production and sales).

BMA estimates are produced in table 5.[19] They are consistent with the first two methods. However, results tends to be more discriminant. For the current quarter model, only a few series show posterior probabilities of inclusion higher than 15%: industrial production and sales trade (resp. Help wanted, ISM, M2, spread, for the future quarter). Following Barbieri and Berger (2004) criterion, selecting posterior probabilities of inclusion superior to 50% would lead to stick only to the following series: the industrial production for the current quarter (resp. the help wanted series and the term spread for the next quarter).

3.3 Model accuracy forecasting performance

How could performances of automatic selection procedures be compared? We focus on out-of-sample prediction errors. To derive out-of-sample forecasts, each models (BMA, DFA, GETS) is estimated in sample on an expanding window, then the four next quarters are forecasted, and so on... In DFA models, the number of factors is selected according to Bai and Ng (2002) IC_{p2} information criteria. We then consider a forecast accuracy test and a directional accuracy test, with the BMA results as the benchmark in all samples.

As the MSE comparisons do not provide a formal statistical indication of whether one model is significantly better than another, we investigate relative forecasting performance with a modified version of the Diebold and Mariano (1995) test proposed by Harvey et al. (1997). Table 1 reports MSE and modified DM statistic estimates.

	BMA	GETS *lib.*[1]	GETS *cons.*	DFA	median
Current quarter forecasts					
MSE	0.21	0.25	0.20	0.21	0.20
DM Stat.		1.05[2]	-0.27	0.07	-0.31
Future quarter forecasts					
MSE	0.31	0.37	0.30	0.24	0.41
DM Stat.		0.68	-0.32	-1.51**	2.83**

(1) *lib.* and *cons.* stand respectively for *liberal* and *conservative* GETS estimation strategy

(2) BMA is considered as the competing model

Statisticaly signifcant at (*) 5% (**) 10% level

Table 1: Diebold-Mariano forecast accuracy test

We complement this "quantitative" accuracy forecast tests by a more "qualitative one". We follow the Pesaran and Timmermann (1992) test. The authors propose a non-parametric statistic to evaluate forecast directional accuracy over any given sample period. Said it differently, they determine which model is able to signal accurately growth acceleration or

[19]The markov chain was run 3,000,000 steps with 1,000,000 burnin, visiting around 27000 models. To ensure the principle of parcimony and limit risks of overfitting, we limit the maximum number of explanatory variables allowed in each regression to 10.

deceleration without wrong or distorted signals.[20] Directional accuracy tests are provided in Table 2.

	Current quarter		**Futur quarter**	
	S_n	p-value	S_n	p-value
BMA	1.80	0.04	-1.30	0.90
GETS lib.[1]	0.73	0.23	-1.30	0.90
GETS cons.	1.81	0.03	-0.92	0.90
DFA	-1.04	0.85	-1.35	0.81
median	1.83	0.03	-0.43	0.67

(1) *lib.* and *cons.* stand respectively for *liberal* and *conservative* GETS estimation strategy

Table 2: Pesaran-Timmermann test of directional change

Figures 1 and 2 show the out-of-sample outcomes of each model, for the current and next quarters. Whereas current quarter forecasts seem to convincingly measure GDP growth, with no surprise, future quarter forecasts are more bumpy.

Two main conclusions arise. Looking first at current forecast estimates, no model can be discriminated. On future quarter forecasts, no model significantly outperforms the BMA except the dynamic factor model, at a 10% level. The "Optimal median" of Barbieri and Berger (2004) is the only approach to be outclassed by the standard BMA, still at a 10% level. Second, considering a 5% probability of rejection under the null, BMA, the *conservative* GETS and the "median model" tend significantly to indicate the right directions on current and future growth. On the whole, the BMA and *conservative* GETS strategies fare quite well, when compared with other automatic selection models. Yet, they are not significantly outclassed by the dynamic factor model.

[20]This test measures whether there is a significant difference between the observed probability of a correctly signed forecast and the estimate of what the probability would be under the null that the forecast and the outcome are independent. On this assumption, independence implies that the forecasting series has no power in predicting the direction of the dependent variable. The corresponding statistic is:

$$S_n = \frac{(\hat{p} - \hat{p_*})^2}{\mathrm{V}(\hat{p}) - \mathrm{V}(\hat{p}_*)} \sim \mathcal{N}(0,1)$$

where $\hat{p}$ is the sample estimate of the probability of a correctly signed forecast and $\hat{p}_*$ is the estimate of its expectation obtained under the null.

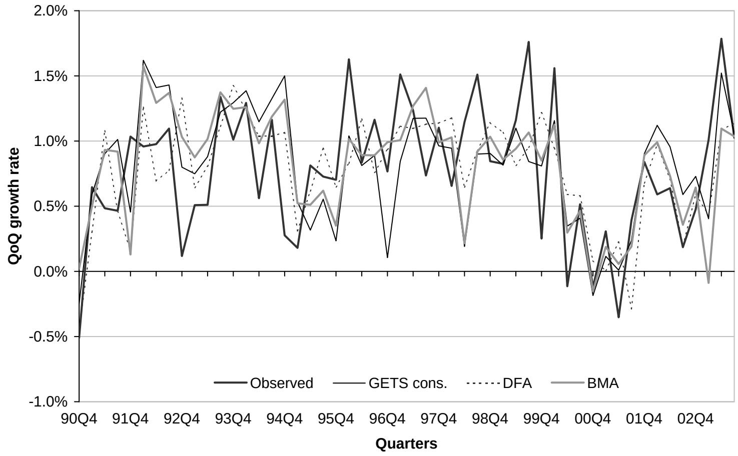

Figure 1: Current out-of-sample forecast of US quarterly GDP - Second month of the current quarter

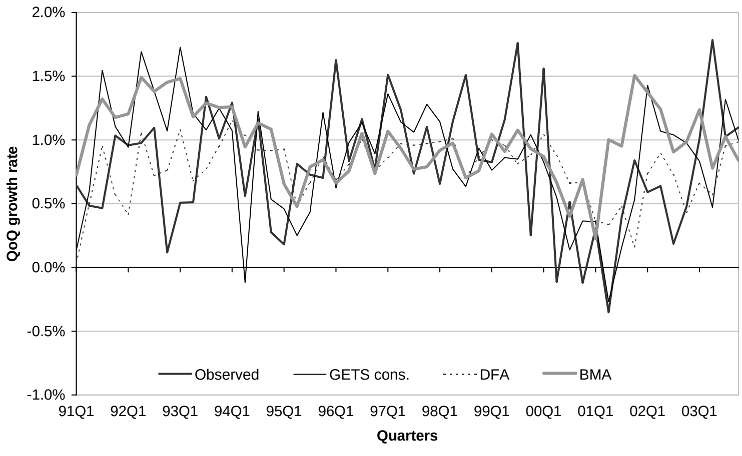

Figure 2: Out-of-sample forecast of future US quarterly GDP - Second month of the current quarter

4 Conclusion

In this paper, we have evaluated forecasting performances of three automatic selection procedures of growing interest. Contrary to standard procedures comparing models to naive benchmarks, we challenge the Bayesian Model Averaging (BMA) against three leading edge approachs: the *general-to-specific* automatic, dynamic factor models, and the "median BMA". From an operational point of view, these methods are more likely to be used by forecasters than the random walk model. Interestingly, empirical results on US GDP growth forecasts cannot significantly discriminate any approach. All models consistently select the same series, such as the industrial production index, the sales in manufacturing and trade, the help wanted advertising index, and the term spread. This gives a serious indication about their robustness as good predictors of US GDP. Second, all methods but the DFA, significantly signal growth accelerations and decelerations at all horizons. Third, all methods are statistically equivalent for short term forecasts. The dynamic factor model seems slightly to dominate on a longer forecasting horizon, but only at a 10% level. In short, the standard BMA based on model uncertainty, and the *general-to-specific* strategy should gain more attention from forecasters. However, a big sample size of potential predictors may limit the use of such models. With information sets including more than 40 series, the factor model (DFA), though less interpretable, is competitive.

References

Bai, J. & Ng, S. (2002). Determining the number of factors in approximate factor model. *Econometrica*, 70, 191-221.

Barbieri, M. & Berger, J.O. (2004). Optimal predictive model selection. *The Annals of Statistics*, 32, 870-897.

Boivin, J. & Ng, S. (2006). Are more data always better for factor analysis? *Journal of Econometrics*, 132, 169-194.

Brock W. A., S. N. Durlauf & West, K. D. (2004). Model uncertainty and policy evaluation: Some theory and empirics, *NBER Working Paper*, 10916.

Dubois, E.(2004). GROCER 1.0: An econometric toolbox for Scilab, mimeo and library available at http://dubois.ensae.net.

Dubois, E. & Michaux, E. (2006). Étalonnages à l'aide d'enquêtes de conjoncture: de nouveaux résultats. *Économie et Prévision*, 172, 11-28.

Diebold, F.X. & Mariano, R.S. (1995). Comparing predictive accuracy. *Journal of Business & Economics Statistics*, 13, 253-263.

Fernández, C., E. Ley & Steel, F. J. (2001a). Benchmark priors for bayesian model averaging. *Journal of Econometrics*, 100, 381-427.

Fernández, C., E. Ley & Steel, F. J. (2001b). Model uncertainty in cross-country growth regressions. *Journal of Applied Econometrics*, 16, 563-56.

Harvey, D. I., S. J. Leybourne & Newbold, P. (1997). Testing the equality of prediction mean squared errors. *International Journal of Forecasting*, 13, 281-291.

Hendry, D. F. & Krolzig, H-M. (2005). The properties of automatic GETS modelling. *The Economic Journal*, 115, 32-61.

Hoover, K. D. & Perez, S. J. (1999). Data mining reconsidered: Encompassing and general-to-specific approach to specification search. *Econometrics Journal*, 2, 167-191.

Jacobson, T. & Karlsson, S. (2004). Finding good predictors for inflation: A bayesian model averaging approach. *Journal of Forecasting*, 23, 479-496.

Krolzig, H-M. & Hendry, D. F. (2001). Computer automation of general-to-specific model selection procedures. *Journal of Economic Dynamics and Control*, 25, 831-866.

Pesaran, M. H. & Timmermann, A. (1992). A simple nonparametric test of predictive performance. *Journal of Business & Economic Statistics*, 10, 461-465.

Raftery, A. E. , D. Madigan & Hoeting, J. A. (1997). Bayesian model averaging for linear regression models. *Journal of the American Statistical Association*, 92, 179-191.

Stock, J. H. & Watson, M. (1998). Diffusion indexes. *NBER Working Paper*, 6702.

Stock, J. H. & Watson, M. (2002). Macroeconomic forecasting using diffusion indexes. *Journal of Business & Economic Statistics*, 20, 147-162.

Wright, J. H. (2003). Bayesian model averaging and exchange rates forecasts. Board of Governors of the Federal System, *International Finance Discussion Papers*, 779.

Appendix A: the data set

names	mnemonics	Publishing lag	Transformation
Average weekly initial claims unemployement insurance (SA*)	incla	one month	$\Delta \log(.)$
Average weekly hours manufacturing (SA)	avh	one month	$\Delta \log(.)$
Help wanted advertising index (SA)	helpw	one month	$\Delta \log(.)$
Index of consumer expectations (SA)	consexp	one month	$\Delta \log(.)$
Industrial production index (SA)	ipi	one month	$\Delta \log(.)$
ISM purchasing managers index (SA)	ism	end of the month	Δ
ISM supplier deliveries (SA)	ismsup	end of the month	Δ
ISM new orders minus inventories (SA)	ismordms	end of the month	Δ
Jobs hard to get minus plentifull (SA)	jhgmp	end oh the month	Δ
M2 money supply (SA)	m2	end oh the month	Δ
Manufacturing new orders, consumer goods & mtrl (volume, SA)	ordt	one month	$\Delta \log(.)$
Manufacturing new orders, nondefense capital goods (volume, SA)	ord	one month	$\Delta \log(.)$
New private housing units started (AR, SA)	houstart	one month	$\Delta \log(.)$
Personal income less transfer payments (SA)	pinc	one month	$\Delta \log(.)$
Sales, manufacturing & trades (SA)	sal	one month	$\Delta \log(.)$
S&P500 index (monthly average of daily price)	sp	end of the month	$\Delta \log(.)$
Treasury 10 year yield adjusted to constant maturity minus 3 month Fed funds	spread	end of the month	No
Unemployment rate (SA)	u	one month	$\Delta \log(.)$

(*) SA: seasonaly adjusted, NSA: not seasonaly adjusted

Table 3: Data description

Appendix B : estimation results

Variable	F_1 (15%*)	F_2 (8%)	F_3 (5%)
spread_{q-2}	0.60	-0.01	0.05
$\text{dlhoustart}_{1,q}$	0.14	0.36	0.13
$\text{dlhoustart}_{3,q-1}$	0.09	0.56	-0.12
$\text{dlhoustart}_{2,q-1}$	0.17	0.27	-0.10
$\text{dlhoustart}_{1,q-1}$	0.37	-0.04	0.21
$\text{dlsp}_{2,q}$	-0.11	0.04	0.09
$\text{dlsp}_{1,q}$	0.00	0.23	0.20
$\text{dlsp}_{3,q-1}$	0.35	0.45	0.01
$\text{dlsp}_{2,q-1}$	0.09	0.38	0.30
$\text{dlsp}_{1,q-1}$	0.22	0.17	0.26
$\text{dlm2}_{1,q}$	0.22	0.34	0.02
$\text{dlm2}_{3,q-1}$	0.18	0.38	0.16
$\text{dlm2}_{2,q-1}$	0.43	0.22	0.29
$\text{dlm2}_{1,q-1}$	0.39	0.03	0.13
$\text{dlordt}_{1,q}$	0.32	0.52	0.07
$\text{dlordt}_{3,q-1}$	0.60	0.13	-0.19
$\text{dlordt}_{2,q-1}$	0.43	-0.14	-0.38
$\text{dlordt}_{1,q-1}$	0.34	-0.31	0.54
$\text{dism}_{2,q}$	0.09	0.69	0.09
$\text{dism}_{1,q}$	0.29	0.56	0.05
$\text{dism}_{3,q-1}$	0.44	0.38	-0.29
$\text{dism}_{2,q-1}$	0.41	-0.11	0.47
$\text{dism}_{1,q-1}$	0.42	-0.28	0.19
$\text{dlavh}_{1,q}$	0.21	0.06	0.11
$\text{dlavh}_{3,q-1}$	0.35	0.17	0.08
$\text{dlavh}_{2,q-1}$	0.19	-0.17	-0.41
$\text{dlavh}_{1,q-1}$	0.13	-0.05	0.52
$\text{du}_{1,q}$	-0.64	-0.24	0.17
$\text{du}_{3,q-1}$	-0.56	0.27	0.05
$\text{du}_{2,q-1}$	-0.40	0.36	0.02
$\text{du}_{1,q-1}$	-0.45	0.44	-0.07
$\text{djhgmp}_{2,q}$	-0.61	-0.18	0.20
$\text{djhgmp}_{1,q}$	-0.45	-0.28	-0.12
$\text{djhgmp}_{3,q-1}$	-0.57	0.33	0.04
$\text{djhgmp}_{2,q-1}$	-0.59	0.15	0.02
$\text{djhgmp}_{1,q-1}$	-0.45	0.27	0.01
$\text{dlincla}_{2,q}$	-0.13	-0.37	-0.08
$\text{dlincla}_{1,q}$	-0.36	-0.45	-0.14
$\text{dlincla}_{3,q-1}$	-0.53	-0.29	0.24
$\text{dlincla}_{2,q-1}$	-0.47	0.03	0.10
$\text{dlincla}_{1,q-1}$	-0.37	0.34	-0.36

Variable	F_1 (15%*)	F_2 (8%)	F_3 (5%)
$\text{dlhelpw}_{1,q}$	0.56	0.21	-0.18
$\text{dlhelpw}_{3,q-1}$	0.48	0.09	0.06
$\text{dlhelpw}_{2,q-1}$	0.55	-0.11	0.00
$\text{dlhelpw}_{1,q-1}$	0.48	-0.50	-0.17
$\text{dlipi}_{1,q}$	0.70	0.26	-0.04
$\text{dlipi}_{3,q-1}$	0.73	-0.16	-0.16
$\text{dlipi}_{2,q-1}$	0.60	-0.46	-0.33
$\text{dlipi}_{1,q-1}$	0.46	-0.52	0.42
$\text{dlpinc}_{1,q}$	0.22	0.03	-0.11
$\text{dlpinc}_{3,q-1}$	0.41	-0.02	0.01
$\text{dlpinc}_{2,q-1}$	0.62	-0.22	-0.16
$\text{dlpinc}_{1,q-1}$	0.15	-0.25	0.20
$\text{dlord}_{1,q}$	-0.02	0.02	0.14
$\text{dlord}_{3,q-1}$	0.27	-0.01	-0.04
$\text{dlord}_{2,q-1}$	0.22	-0.02	-0.21
$\text{dlord}_{1,q-1}$	0.00	-0.15	0.14
$\text{dlsal}_{1,q}$	0.26	0.40	0.13
$\text{dlsal}_{3,q-1}$	0.69	-0.01	-0.09
$\text{dlsal}_{2,q-1}$	0.36	-0.05	-0.52
$\text{dlsal}_{1,q-1}$	0.24	-0.31	0.68
$\text{dlconsexp}_{1,q}$	0.12	0.13	0.14
$\text{dlconsexp}_{3,q-1}$	0.11	0.55	0.10
$\text{dlconsexp}_{2,q-1}$	0.26	0.12	0.25
$\text{dlconsexp}_{1,q-1}$	0.23	-0.03	-0.08
$\text{dismordms}_{2,q}$	-0.21	0.27	0.09
$\text{dismordms}_{1,q}$	-0.12	0.48	0.15
$\text{dismordms}_{3,q-1}$	0.10	0.28	-0.44
$\text{dismordms}_{2,q-1}$	0.10	0.26	0.33
$\text{dismordms}_{1,q-1}$	0.12	-0.10	0.14
$\text{dismsup}_{2,q}$	0.18	0.30	-0.01
$\text{dismsup}_{1,q}$	0.43	0.29	0.08
$\text{dismsup}_{3,q-1}$	0.40	0.09	0.16
$\text{dismsup}_{2,q-1}$	0.30	-0.13	0.01
$\text{dismsup}_{1,q-1}$	0.48	-0.22	0.09

(*) Percentage of total variance explained by each factor

Table 4: Correlation with factors

Variable	Current quarter		Future quarter	
	Coeff.*	Post. prob.	Coeff.	Post. prob.
spread$_{q-2}$	0.01	11.34	0.11	67.49
dlhoustart$_{1,q}$	0.08	4.47	0.10	4.81
dlhoustart$_{3,q-1}$	0.01	0.93	0.29	10.58
dlhoustart$_{2,q-1}$	0.04	2.41	0.03	1.21
dlhoustart$_{1,q-1}$	0.00	0.53	0.00	0.31
dlsp$_{2,q}$	0.01	0.51	0.01	0.84
dlsp$_{1,q}$	0.31	8.78	0.40	8.08
dlsp$_{3,q-1}$	0.02	1.18	0.41	8.88
dlsp$_{2,q-1}$	0.02	1.18	0.01	0.51
dlsp$_{1,q-1}$	0.07	2.60	0.01	0.52
dlm2$_{1,q}$	0.33	1.63	0.61	1.79
dlm2$_{3,q-1}$	1.59	6.14	10.99	24.29
dlm2$_{2,q-1}$	0.58	2.55	26.72	47.73
dlm2$_{1,q-1}$	0.55	2.70	0.05	0.59
dlordt$_{1,q}$	0.02	0.77	0.00	0.26
dlordt$_{3,q-1}$	0.46	6.02	0.00	0.42
dlordt$_{2,q-1}$	-0.05	1.38	0.04	0.78
dlordt$_{1,q-1}$	0.02	0.97	-0.01	0.34
dism$_{2,q}$	0.01	17.83	0.00	3.03
dism$_{1,q}$	0.00	3.17	0.00	0.68
dism$_{3,q-1}$	0.00	0.59	0.04	47.29
dism$_{2,q-1}$	0.00	0.56	0.00	0.31
dism$_{1,q-1}$	0.00	0.72	0.00	0.43
dlavh$_{1,q}$	0.00	0.44	-0.04	0.44
dlavh$_{3,q-1}$	0.21	1.38	0.02	0.39
dlavh$_{2,q-1}$	0.59	4.89	-0.03	0.37
dlavh$_{1,q-1}$	-0.15	1.59	0.01	0.31
du$_{1,q}$1	0.00	0.68	0.00	0.40
du$_{3,q-1}$	0.00	0.59	-0.01	1.16
du$_{2,q-1}$	-0.01	2.11	0.00	0.75
du$_{1,q-1}$	0.00	0.64	0.00	0.36
djhgmp$_{2,q}$	0.00	9.58	0.00	1.98
djhgmp$_{1,q}$	0.00	1.70	0.00	8.60
djhgmp$_{3,q-1}$	0.00	0.72	0.00	0.39
djhgmp$_{2,q-1}$	0.00	2.10	0.00	0.78
djhgmp$_{1,q-1}$	0.00	4.26	0.00	0.49
dlincla$_{2,q}$	-0.07	3.91	-1.30	35.25
dlincla$_{1,q}$	-0.06	3.54	0.00	0.28
dlincla$_{3,q-1}$	0.00	0.67	-0.01	0.53
dlincla$_{2,q-1}$	-0.02	1.55	0.03	1.58
dlincla$_{1,q-1}$	0.00	0.69	-0.01	0.60

(*) Coefficients are multiplied by 100

Variable	Current quarter		Future quarter	
	Coeff.*	Post. prob.	Coeff.	Post. prob.
dlhelpw$_{1,q}$	0.02	1.22	0.05	1.44
dlhelpw$_{3,q-1}$	0.04	1.77	0.10	2.35
dlhelpw$_{2,q-1}$	0.00	0.51	4.36	67.81
dlhelpw$_{1,q-1}$	0.02	1.22	0.00	0.39
dlipi$_{1,q}$	60.24	99.99	-0.26	1.66
dlipi$_{3,q-1}$	30.58	84.28	0.00	0.32
dlipi$_{2,q-1}$	2.16	11.17	-0.50	2.90
dlipi$_{1,q-1}$	-0.01	0.60	0.00	0.34
dlpinc$_{1,q}$	0.10	1.29	0.01	0.31
dlpinc$_{3,q-1}$	0.00	0.56	0.01	0.21
dlpinc$_{2,q-1}$	0.43	1.88	0.14	0.68
dlpinc$_{1,q-1}$	0.00	0.55	0.06	0.99
dlord$_{1,q}$	0.00	0.65	0.00	0.20
dlord$_{3,q-1}$	0.00	0.48	0.00	0.10
dlord$_{2,q-1}$	0.19	10.46	0.00	0.30
dlord$_{1,q-1}$	0.00	0.56	0.00	0.28
dlsal$_{1,q}$	1.38	9.24	0.00	0.09
dlsal$_{3,q-1}$	4.55	23.70	0.01	0.12
dlsal$_{2,q-1}$	0.05	0.82	0.00	0.09
dlsal$_{1,q-1}$	-0.02	0.62	-0.01	0.16
dlconsexp$_{1,q}$	0.01	0.90	0.00	0.10
dlconsexp$_{3,q-1}$	0.00	0.48	0.00	0.18
dlconsexp$_{2,q-1}$	0.00	0.45	0.00	0.03
dlconsexp$_{1,q-1}$	0.00	0.42	0.00	0.16
dismordms$_{2,q}$	0.00	1.94	0.00	0.00
dismordms$_{1,q}$	0.00	0.59	0.00	0.01
dismordms$_{3,q-1}$	0.00	0.61	0.00	0.01
dismordms$_{2,q-1}$	0.00	0.38	0.00	0.00
dismordms$_{1,q-1}$	0.00	0.64	0.00	0.00
dismsup$_{2,q}$	0.00	0.66	0.00	0.00
dismsup$_{1,q}$	0.00	1.06	0.00	0.05
dismsup$_{3,q-1}$	0.00	0.63	0.00	0.00
dismsup$_{2,q-1}$	0.00	0.32	0.00	0.00
dismsup$_{1,q-1}$	0.00	0.43	0.00	0.00

Table 5: Estimation results of BMA

Liberal strategy					
Current quarter			Future quarter		
Variable	Coeff.	t-stat.	Variable	Coeff.	t-stat.
spread_{q-2}	0.00	2.35	spread_{q-1}	0.00	2.88
$\text{dlsp}_{1,q}$	0.03	2.45	$\text{dlhoustart}_{1,q}$	0.01	1.74
$\text{dlsp}_{1,q-1}$	0.03	2.12	$\text{dlsp}_{1,q}$	0.05	2.65
$\text{dljhgmp}_{3,q-1}$	0.00	2.29	$\text{dlm2}_{1,q}$	0.27	2.03
$\text{dlincla}_{1,q}$	-0.02	-2.12	$\text{dism}_{3,q-1}$	0.00	3.35
$\text{dlipi}_{1,q}$	0.51	6.69	$\text{djhgmp}_{1,q}$	0.00	-1.77
$\text{dlipi}_{3,q-1}$	0.29	3.25	$\text{dlincla}_{2,q}$	-0.02	-1.91
$\text{dlipi}_{2,q-1}$	0.14	2.01	$\text{dlhelpw}_{3,q-1}$	0.06	2.84
$\text{dlsal}_{3,q-1}$	0.11	2.01	$\text{dlhelpw}_{2,q-1}$	0.08	4.72
constant	0.00	8.91	$\text{dlipi}_{1,q}$	-0.30	-3.31
			$\text{dismsup}_{1,q}$	0.00	1.27
			$\text{dismsup}_{1,q-1}$	0.00	-2.66
			constant	0.01	9.65
Conservative strategy					
spread_{q-2}	0.00	2.17	spread_{q-1}	0.00	2.86
$\text{dlsp}_{1,q}$	0.03	2.07	$\text{dlhoustart}_{1,q}$	0.02	2.62
$\text{dlsp}_{1,q-1}$	0.03	2.25	$\text{dlsp}_{1,q}$	0.06	3.07
$\text{dlipi}_{1,q}$	0.40	4.53	$\text{dism}_{3,q-1}$	0.00	3.98
$\text{dlipi}_{3,q-1}$	0.24	2.71	$\text{dlincla}_{2,q}$	-0.03	-2.33
$\text{dlipi}_{2,q-1}$	0.11	1.57	$\text{dlhelpw}_{3,q-1}$	0.06	2.99
$\text{dlsal}_{1,q}$	0.14	2.55	$\text{dlhelpw}_{2,q-1}$	0.09	4.81
$\text{dlsal}_{3,q-1}$	0.15	2.55	$\text{dlipi}_{1,q}$	-0.31	-3.35
constant	0.00	9.26	constant	0.01	10.82

Table 6: Estimation results of GETS