Working Paper 2006 | 7

A Bayesian method of forecast averaging for models known only by their historic outputs: an application to the BCRA´s REM

Pedro Elosegui / Francisco Lepone
George McCandless
BCRA

August, 2006

ie | BCRA

Investigaciones Económicas
Banco Central
de la República Argentina

Banco Central de la República Argentina
ie | Investigaciones Económicas

August, 2006
ISSN 1850-3977
Electronic Edition

Reconquista 266, C1003ABF
C.A. de Buenos Aires, Argentina
Tel: (5411) 4348-3719/21
Fax: (5411) 4000-1257
Email: investig@bcra.gov.ar
Pag.Web: www.bcra.gov.ar

A Bayesian method of forecast averaging for models known only by their historic outputs: An application to the BCRA's REM

Dr. Pedro Elosegui Dr. Francisco Lepone
Dr. George McCandless
Subgerencia General de Investigaciones Economicas
Banco Central de la República Argentina

August 31, 2006

Abstract

Similar to other Central Banks, the BCRA publishes monthly a REM that summaries the forecasts and projections of a group of economic analysts and consultants who volunteer to participate in the program. The BCRA publishes only the median, and the standard deviation of the sample received. The logic for using these statistics is that all participants are to be treated equally.

Under the assumption that some forecasters have better underlying models than others, one might be able to improve the accuracy of the aggregate forecast by giving greater priority to those who have historically predicted better. The BCRA does not have access to the models used to make the predictions, only the forecasts are provided. An averaging method that puts higher weights on the predictions of those forecasters who have done best in the past should be able to produce a better aggregate forecast. The problem is how to determine these weights. In this paper, we develop a Bayesian averaging method that can do that well.

The aggregate forecasts that come from our Bayesian averaging provides statistically better forecasts than the mean, median, best model, five best models and other methods traditionally used. In particular, the method developed in this paper is much better at detecting changes in the trends of the variables.

The aggregate predictions published from the REM provide information that is useful, not only for monetary and economic policy decisions, but also for the consumption and business decisions of private economic agents. Improving these forecasts are of benefit to all members of the economy.

1 Introduction

Similar to other Central Banks, the BCRA periodically publishes a Relevamiento de Expectativas de Mercado (REM) which summarizes short and medium term macroeconomic forecasts and projections of the group of economic analysts and consultants who volunteer to participate in the program. In part to protect the confidential nature of the forecasts that the analysts provide to the central bank, only a few principal statistics of the forecast sample are published. These statistics can provide the public and the central bank authorities with relevant information on the professional consensus of the process of important macroeconomic variables. This information can be useful for making decisions on monetary and economic policy as well as for private individuals making their own business and consumption decisions.

The short and medium term variables that are surveyed by the REM fall into five categories: price indices, financial and monetary variables, indicators of economic activity, international trade and exchange rates, and the central government's budget. The short term forecasts are taken monthly and involve one and two month ahead projections. The medium term variables are quarterly or yearly, again with forecasts for two periods. Nominal GDP and the CPI are also published as December on December forecasts. The internet page of the BCRA publishes the principal statistics of the sample, including the means medians, and standard deviations for each variable. In addition, each month the BCRA publishes the names for the firms that produced the three top forecasts in each category.

The objective of the present paper is to develop a methodological tool to calculate a Bayesian average of the forecasts of the REM. Such a forecast would complement the information already provided. In particular, the calculation of a Bayesian average would permit weighting the various forecasts based on the history of the underlying models and their relative forecasting success. The object is to have a weighted forecast that should be able to predict better than the median, which is currently used.

Crucial to the potential success of a Bayesian averaging methodology is the assumption that some forecasting firms have better underlying models than do others. If this assumption holds, an averaging method that puts higher weights on the predictions of those forecasters who have done best in the past will be able to produce a better aggregate forecast. The problem is how to determine these weights. We do not have access to the models, the participants in the REM only provide the BCRA with their predictions. The maximum amount of observations is relatively small, less than 30 data points and the sample is not balanced (although balanced subsets exist and can be extracted from the data). Bayesian techniques provide a method for using the data to calculate weights for the aggregate forecast and, potentially, for using additional (prior) information for finding those weights. We assume that the forecast errors have a likelihood with a normal distribution of zero mean. The variance of this normal distribution is important in determining how much each firm contributes to the aggregate forecast. Too small a variance and only one firm is chosen, too large

a variance and the aggregate is a simple average. We choose a value for the variance of this likelihood function so as to minimize the in-sample aggregate forecast error. We also discuss other choices for the variance and potential priors.

In sections 2, 3, 4, and 5, we present aspects of the statistical theory that permits calculating the averages of the forecasts based on Bayesian weights and simulations to demonstrate some properties of the method. In section 6, we apply these techniques to the data that comes from the REM and we compare the results of our weighting system with the one currently in use.

2 Basic averaging

Each month a set A of forecasting firms submit to the central bank one month ahead forecasts of a particular variable. Let $y_{j,t+1}$ be the one step ahead forecast made at date t by firm $j \in A$ for that variable. We assume that each firm j has a model which we call M_j about which we know nothing except a history of past predictions, $Y_{j,t} = \{y_{j,1}, y_{j,2}, ...y_{j,t}\}$. We interested in getting a best average estimate of the one step ahead forecast. Our data is the history of realizations of the variable $y^t = \{y_1, y_2, ...y_t\}$, where y_t is the realization of the variable at date t. Since we have the history of forecasts of each firm and the realizations of the variable y^t, we can construct the history of the actual error of each forecasting firm: $\varepsilon_{j,t} = y_t - y_{j,t}$ is the error in the forecast for the value at date t and $\Gamma_{j,t} = \{\varepsilon_{j,1}, \varepsilon_{j,2}, ...\varepsilon_{j,t}\}$ is the full history of errors available at date t.

We want to calculate an aggregate forecast of the form,

$$E_t\left(y_{t+1}\right) = \sum_{j \in A} E_t\left(y_{j,t+1} | M_j, y^t\right) g\left(M_j | y^t\right),$$

where the expected value of the one step ahead forecast of each firm j at time t (which we assume to be the value that they announce to the Central Bank) is given by $E_t\left(y_{j,t+1} | M_j, y^t\right)$. The value, $g\left(M_j | y^t\right)$, is the posterior weight applied to firm j's forecast. The posterior weight used for the forecast of firm j's model is equal to

$$g\left(M_j | y^t\right) = \frac{f\left(y^t | M_j\right) g\left(M_j\right)}{f\left(y^t\right)}.$$

In this equation, $g\left(M_j\right)$ is the prior probability that we assume for the model j, $f\left(y^t | M_j\right)$ is the likelihood of model j given the data, $f\left(y^t\right)$ is the probability of the data, and $g\left(M_j | y^t\right)$ is the posterior probability for model j given the data and the prior. The likelihood is also called the conditional probability of the data given the model. The posterior is also called the conditional probability of the model given the data.[1] With a finite set A of models, the probability of

[1] A well know condition of probability is that a joint distribution $f(x, y)$ is equal to a conditional probability, $f(x|y)$ or $f(y|x)$, times the respective marginal probability, $g(y)$ or $g(x)$. This gives ,$f(x, y) = f(x|y)g(y) = f(y|x)g(x)$ which can be rewritten as Bayes rule, $f(y|x) = \frac{f(x|y)g(y)}{g(x)}$.

the data is simply

$$f\left(y^t\right) = \sum_{j \in A} f\left(y^t | M_j\right) g\left(M_j\right).$$

2.1 A flat prior

We begin with an uninformative (flat) prior over the models so if there are n firms in A, then $g\left(M_j\right) = 1/n$.

Our problem is what distribution to assume for the likelihood, $f\left(y^t | M_j\right)$. Since we have chosen to think of a model as the history of the predictions made by that model, we choose to assume that the likelihood function is over the errors of the predictions of that model with respect to the realizations (the data). We assume that this likelihood function is normal, with zero mean and, initially, with a variance, σ^2. We test a range variances and these put different weights on the different models. Below we give an example where we choose the variance of the likelihood function to minimize the quadratic aggregate forecast error.

We assume that the likelihood function is normal, so if we consider σ^2 as known and we have T observations on the prediction errors of firms j, $f\left(y^t | M_j\right)$ is represented by

$$f\left(y^T | M_j\right) = \left[2\pi\sigma\right]^{-.5T} \exp\left[-.5 \frac{\sum_{t=1}^{T} \left(\varepsilon_{j,t} - 0\right)^2}{\sigma^2}\right].$$

The weight applied to the forecast of firm j is then equal to

$$\frac{f\left(y^T | M_j\right)}{\sum_{k \in A} f\left(y^T | M_k\right)}.$$

It should be obvious from this definition that the weights sum to one.

2.2 An alternative prior

The above development of an improved weighting for averaging the reported forecasts from the REM were done using a flat prior. If the precision of forecasts for firms are correlated across variables, improved averaging weights for variables that are of primary interest to the Central Bank could be found using the posteriors of other variables as priors.

The basic assumption is that some firms have forecasting models that are, in general, better than others. This implies that there should be a correlation across prediction errors: firms that have lower prediction errors in one variable would tend to have lower prediction errors in other variables. This characteristic of the firms can be exploited in determining priors.

Suppose that the Central Bank is very interested in the forecasts of a limited set of variables and is willing to accept somewhat less precision is the forecasts of some others. For a typical central bank, forecasts of inflation and interest

rates are normally of primary concern. This is particularly true today when many central banks use inflation targeting as their policy objective. While still important, precision in forecasts or real variables, such as output, are less so.

If precision in output forecasts, for example, is considered somewhat less important, one can find posteriors for the weights across firms in finding the "best" average forecast using the kind of flat prior described in the first section. Even with flat priors, these posteriors might be very good if the data assigns the weighting averages well. Unless the posteriors grant equal weight to every firm, the posteriors weights found using flat priors should be superior to those currently being used.

For inflation, one can use the posterior weights from the output averaging as priors to the calculation of the posterior weights for the inflation or interest rate averaging. This will improve the weights if the output weights contain information about how well the firms predict inflation or interest rates.

Since forecasts of nominal interest rates usually involve (formally or informally) forecasts of inflation rates, it would not be appropriate to use the posteriors of one of these two variables as a prior for the other. This is because one would be using, in some sense, some of the same data for finding the prior as for finding the posterior. For a similar reason, one would not want to use an iterative process where one finds the posterior for output, say, using a flat prior, use this as a prior for inflation, and then begin a cycle using the posterior for inflation as a prior for output, and the resulting posterior for output as a prior for inflation until the two converge. With this kind of iteration process, what should be data for determining the inflation weights is turning up in the prior for these inflation weights. The output posteriors (the priors for inflation) were found using a previous round of posteriors for inflation, so information from the inflation data is in the output posteriors that are used as the inflation priors.

To find a better posterior weighting scheme, we use an assumption that there is correlation in the *precision* of the forecasts of the different variables by a firm. This is different from the correlation among the forecasts that a firm makes of the *variables* (such as that which might appear between the forecast of inflation and that of nominal interest rates) or the correlation among the variables themselves.

More specifically, let $y^{t,i} = \left\{y_1^i, y_2^i, ...y_t^i\right\}$ be the realizations of variable i and let $Y_{j,t}^i = \left\{y_{j,1}^i, y_{j,2}^i, ...y_{j,t}^i\right\}$ be the one set ahead forecasts of variable i made by firm $j \in A$. We choose one variable, i^*, and calculate the posterior weight for each firm j using

$$g\left(M_j|y^{t,i^*}\right) = \frac{f\left(y^{t,i^*}|M_j\right) g\left(M_j\right)}{f\left(y^{t,i^*}\right)},$$

where

$$f\left(y^{t,i^*}\right) = \sum_{j \in A} f\left(y^{t,i^*}|M_j\right) g\left(M_j\right).$$

Now, using the $g\left(M_j|y^{t,i^*}\right)'$s as priors, for some other variable $i \neq i^*$, we cal-

culate the "improved" posterior weights as

$$g\left(M_j|y^{t,i}\right) = \frac{f\left(y^{t,i}|M_j\right) g\left(M_j|y^{t,i^*}\right)}{f\left(y^{t,i}\right)},$$

where

$$f\left(y^{t,i}\right) = \sum_{j \in A} f\left(y^{t,i}|M_j\right) g\left(M_j|y^{t,i^*}\right).$$

This new prior is informative, where the information comes from the firms ability to predict some other variable.

3 Choosing the variance of the likelihood function

We need to make a number of assumption about the mega-properties of the model. Among these properties is the form of the common likelihood function, $f\left(y^t|M_j\right)$, that we use in the equation

$$g\left(M_j|y^t\right) = \frac{f\left(y^t|M_j\right) g\left(M_j\right)}{f\left(y^t\right)}$$

that gives us the weights that we apply to each model[2]. We assume that the errors are normally distributed, that it has a mean of zero when measured in terms of the errors between the predictions of the models and the observed data, and that its variance is the same across models. Given these assumptions, the likelihood function should really be written

$$f\left(y^t|M_j, \sigma^2, \mu = 0\right).$$

The posterior distributions and the resulting weights depend on the choice of σ^2. For that reason, it would be best to express the weights conditional on the mega-parameters as

$$g\left(M_j|y^t, \sigma^2, \mu = 0\right)$$

For the artificial data set used in this paper, Figure 1 shows the calculated weights for a range of .01 to 3 for the variance of the likelihood function. Each line in the figure shows the weights applied to a given model as the value of σ^2 changes. When σ^2 is very small, only the single best model is given any weight. As σ^2 grows, more models are added, until all models are given the same weight as σ^2 approaches ∞.

The problem of which value of σ^2 to use in an exercise does not have an obvious answer.

The method that we recommend in this paper solves a clearly defined optimization problem. One uses the precious predictions and the data and finds the

[2] The only thing we know about each model is the set of forecasts that it has made. The forecasts are the model.

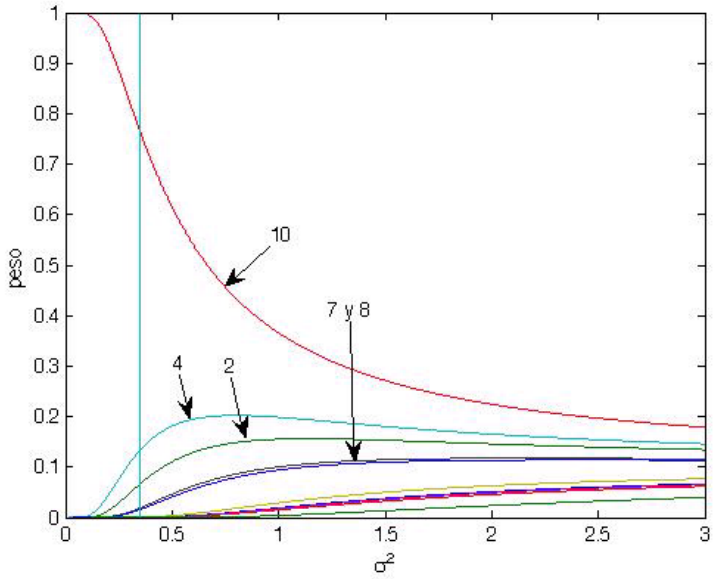

Figure 1: Weights for Bayesian averaging as function of σ^2

$\widehat{\sigma}^2$ whose resulting weights minimize the squared errors of the aggregate predictions made using those weights and the separate models' predictions. Since finding optimal prediction weights is, after all, the real objective of the exercise, this methodology is difficult to fault.

For the set of observations on realizations of the data up to time t, y^t, the past forecasts of this variable by the Q forecasting firms, $\{y_{j,s-1}\}_{s=0}^{t-1}$, and a value for σ^2 in the permitted domain, we get a set of weights $w_{j,t} = g\left(M_j|y^t, \sigma^2, \mu = 0\right)$, for $j = 1, ..., Q$. For this value of σ^2, one can find the in sample weighted forecasts from

$$E_t\left(y_{s+1}\right) = \sum_{j=1}^{Q} w_{j,t} y_{j,s},$$

for $s = 0, ..., t-1$, where $E_t\left(y_{s+1}\right)$ is used to indicate the weighted in sample predictions using historic firm predictions but averaged using weights that were found using all the data up to period t. We search for the value of σ^2 that minimizes the squared in sample forecast errors,

$$\sum_{s=1}^{T} \left[y_{s+1} - \sum_{j=1}^{Q} w_{j,T} y_{j,s} \right]^2.$$

For the simulated data, the values of this function for a range of σ^2 from .04

	M_1	M_2	M_3	M_4	M_5
$\boldsymbol{\sigma}_i$	0.284	0.197	0.292	0.184	0.287
	M_6	M_7	M_8	M_9	M_{10}
$\boldsymbol{\sigma}_i$	0.269	0.218	0.221	0.334	0.147

Table 1: Standard deviation of models.

to.81 (for σ from .2 to .9) are shown in Figure 3. The minimum of the squared errors occurs at $\widehat{\sigma}^2 = .352$ (or, at $\widehat{\sigma} = .5933$).

The vertical line in Figure 1 is at the $\widehat{\sigma}^2$ that was found in the optimization exercise. The numbers for the lines correspond to the different models. One can see that the weights indicated by where the vertical line crosses the weight lines are exactly those given in in Table 2.

4 Testing the method

In order to illustrate the benefits of the Bayesian averaging method that we are recommending, we generate a data set of simple artificial predictions, and use this to compare the Bayesian method with five other methods usually employed to combine forecasts. The other methods used are simple average of all models, simple average of the top five models, the direct choice of the best model, the median of the forecasts distribution and a "method of pooling" average to be described below. The artificial data set is comprised of ten "firms" that are assumed to have a model that is predicting a variable with a constant value of one. The predictions for firm i have a distribution $N(1, \sigma_i^2)$, where the σ_i^2 are chosen from a uniform distribution over $[0, 0.25]$. The models are very simple, each firm's predictions are generated from normal distribution but the variances are different from firm to firm. For each simulation, we have 101 observations for each of the 10 firms and use the first 100 observations to calculate the weights (in the Bayesian pooling methods) and find averages for the six methods for the 101^{th} observation. For the example discussed here, the standard errors, σ_is, are given in Table 1.

Figure 2 shows the forecasts generated by sampling the normal distributions (the model) of each firm. Since the data being predicted is a constant at 1, models M_2, M_4 and M_{10} are better than the rest because of the lower σ_is associated with their forecasts.

For the Bayesian averaging, we choose a non-informative prior over the different models: one that gives the same probability $1/k$ (where $k = 10$ is the number of firms) to each firm's forecast. A non-informative prior assumes that the forecasts of each firm have the same quality. Later we will make some considerations concerning alternative priors.

As it was mentioned in previous sections for the Bayesian posterior weights, we need to decide on a value to be assigned to σ_0, i.e., the value for the standard deviation in the likelihood functions. It turns out that σ_0 is a crucial parameter in the determination of the weights (and, consequently, in the efficiency of the

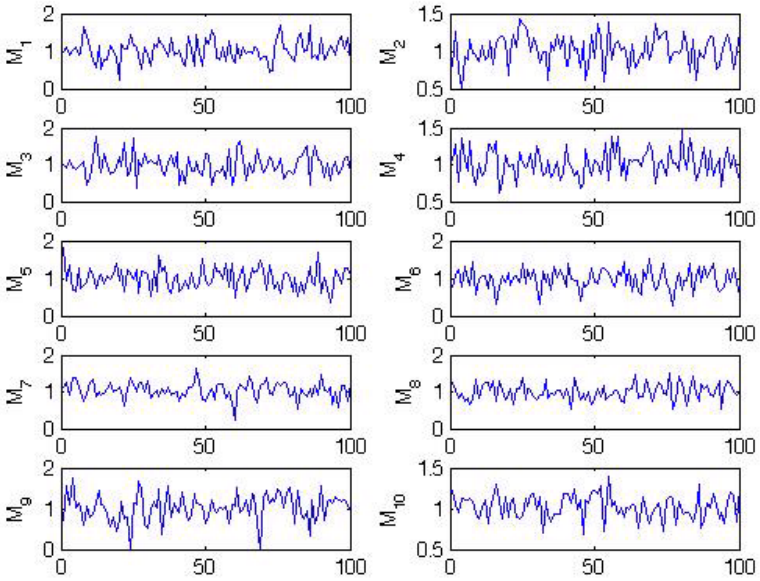

Figure 2: Forecasts corresponding to each model.

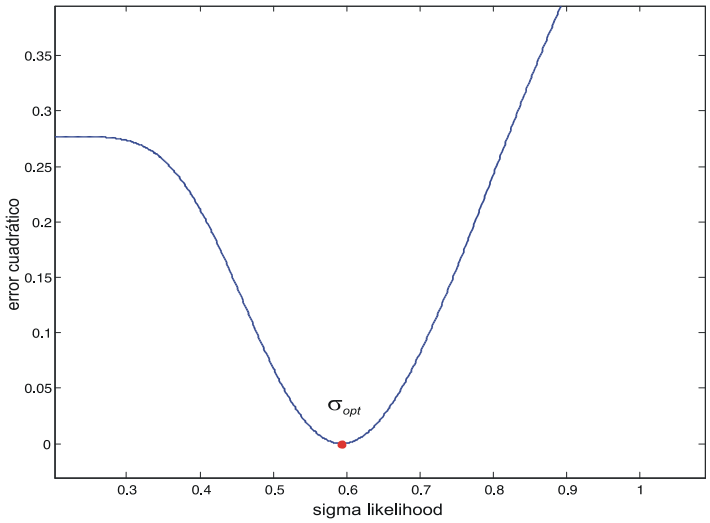

Figure 3: Cumulative quadratic errors for Bayesian forecasts as a function of σ_0 parameter.

wb_1	wb_2	wb_3	wb_4	wb_5
2.0×10^{-4}	0.0685	1.0×10^{-4}	0.1373	1.61×10^{-4}
wb_6	wb_7	wb_8	wb_9	wb_{10}
5.91×10^{-4}	0.016	0.0174	2.40×10^{-6}	0.76

Table 2: Bayesian weights for the models given an optimal sigma

Bayesian forecast). The reason for this is that σ_0 controls both the height and width of the normal distribution. A value for σ_0 that is too big will make all the forecast errors have similar probabilities and the resulting likelihood functions for different models will be almost indistinguishable. A value of σ_0 that is too small will be very restrictive and will select only that model with the smallest errors, ignoring all the other models. Taken this into account, our goal is to employ a value of σ_{opt} for which the historic (the 100 first points) quadratic errors of Bayesian forecasts are minimized. Figure 3 shows the quadratic error of Bayesian forecasts as a function of σ_0 generated from our simulated sample.

The minimization of this function gives an optimum value of $\sigma_{opt} = 0.5933$ for the present example. With this value of σ_{opt} the Bayesian weights for the models are computed. These are shown in Table 2.

Once the weights are obtained, it is straightforward to calculate the Bayesian average forecast ($\overline{y}_b$). As stated above, we also include five other methodologies: 1) a simple arithmetic average of all firms, 2) the method of pooling proposed by Bates and Granger [2][3] for the combination of forecasts, 3) the forecast given by

[3]Synthetically, the method chooses weights for the average to minimize the variance of the

	Forecast
Bayesian average ($\overline{y}_{bay}$)	1.00
Arithmetic average ($\overline{y}_{arith}$)	0.892
Min-Variance average ($\overline{y}_{\min-var}$)	0.912
Best model	1.053
Top 5	0.883
Median	0.924

Table 3: Forecast using different averaging methods.

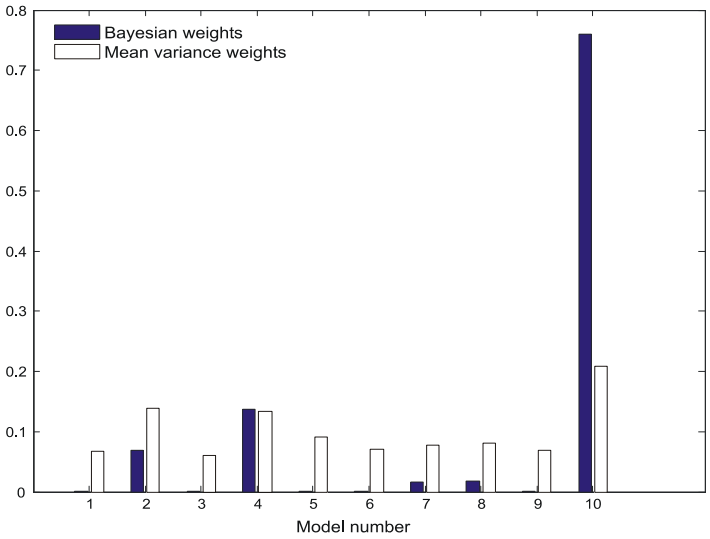

Figure 4: Bayesian and min-variance weights for the example considered here.

the best model[4], 4) an arithmetic average of the top 5 firms, and 5) the median of the distribution of all forecasts. Table 3 shows the results of each averaging method.

From Table 3 it is clear that while simple arithmetic and min-variance forecasts from the example have errors of $\sim 8.8\%$ and $\sim 10.8\%$ respectively, the Bayesian forecast is exact. Figure 4, shows the Bayesian and min-variance weights for comparison purposes.

In order to compare the relative efficiency of the six methods, we performed a test using ten thousand independent forecasting exercises. The characteristics on each run was identical to the described example (with the variance for the ten firms coming from new draws of an uniform $[0, 0.25]$ distribution). The results showed that in the 71.1% of the cases the Bayesian averaging method

errors of the resulting averaged past forecasts.

[4]For best model we understand the one with minimal variance.

Method	% forecast error	Std.	Index.	Ranking
Bayesian	0.72%	0.0176	1.4275	1
Simple Arithmetic	6.91%	0.0874	4.8196	6
Min. Variance	2.35%	0.0364	3.0548	2
Best model	3.30%	0.0557	3.2546	3
Top 5	5.06%	0.0675	4.2985	5
Median	4.73%	0.0670	4.1450	4

Table 4: Average absolute porcentage error and standard deviation of errors.

Method	Index. ($n_t = 15$)	Index. ($n_t = 20$)	Index. ($n_t = 30$)
Bayesian	1.85	1.78	1.68
Simple Arithmetic	4.54	4.66	4.77
Min. Variance	3.63	3.23	3.10
Best model	2.87	3.05	3.10
Top 5	4.24	4.30	4.3
Median	3.82	4.10	4.03

Table 5: Efficiency of the methods as a function of the number of training points.

gave the best forecast, the best model method ranked first 11.6% of the times, the min-variance averaging was first 5.4% of the times, the top 5 arithmetic averaging 4.9% of the times, the median method 4.0% of the times, and the simple arithmetic averaging the remaining 3.0%.

There are different ways of assessing the methods overall performance, we show three 1) the average absolute percentage error of forecasts, 2) the standard deviation and 3) a performance index[5] ranging from 1 (best) to 6 (worst). Table 4 shows the results together with the overall ranking.

From Table 4 it is clear that Bayesian averaging is by far the best method, followed by the min-variance and the best model method.

One interesting question is how sensitive is the performance of the methods to the length of the data set? We tested relative performances as a function of the number of past forecast ($n_t = 15, 20, 30$). The results shows that, although the efficiency is an increasing function of the data number, the Bayesian method is always considerable more efficient than its alternatives (Table 5).

The percentage of times that the Bayesian method gave the best forecast was 52% ($n_t = 15$), 56% ($n_t = 20$) and 62% ($n_t = 30$). On the opposite side, i.e. when n_t is big, the index score of the method tends to a lower bound of approximately 1.25.

[5] The index is computed in the following way. The relative ranking position in each forecast exercise is punctuated from 1 to 6. Then the ranking position of model M_i is calculated as:

$$\sum_k p_{ik} \times k$$

where p_{ik} is the probability of model i having a ranking k. Obviously, the lower the index value, the better.

n	Informative Prior	Flat Prior
3	85.7%	14.3%
5	80.9%	19.1%
10	76.5%	23.5%
20	72.6%	27.4%
30	64.8%	35.2%
40	63.4%	36.6%
50	60.6%	39.4%
60	61.7%	38.3%
100	53.1%	46.9%
200	49.85%	50.15%
500	48.0%	52.0%

Table 6: Percentage of time that each method gave the best forecast.

Another interesting question to be considered is when Bayesian weights become stable? In other words how many training data are needed to produce approximately constant weights? Attempting to answer this question we devised the following experiment: starting from a common random seed we generated sets of forecast of different length and computed their respective Bayesian weights. The results are shown in Figure 5. It can be seen that if we have few training points, they are not informative enough and the weights need to accommodate when new information come. When the amount of training data is big the weights stabilize to their optimal level. From figure 5, we can conclude that between 20 and 30 data points are enough for the weights to reach levels similar to their asymptotic values. Figure 6 shows the Bayesian forecasts as a function of the number of training data.

In order to evaluate the importance of whether using or not informative priors, we devised the following exercise: 1) from an initial sample of 40 data points and using flat priors, compute posteriors to be used as priors in the following steps, 2) for different values of n, generate one thousand independent forecast sets, 3) for each set generate posterior using the priors calculated in 1, 4) compute Bayesian weights using the two sets of posteriors (obtained in 2 and 4) and quantify the relative performance in forecasting the data. Table 6 shows the results. As may be expected, having a reasonable prior helps for reduced data sets. As an example, for a sample of 20 data the efficiency over a flat prior increases by a factor of 2.65. As the number of training data increases the prior becomes less and less important. It is interesting to note that the use of an not very good prior can be even harmful, this case happen when de data set is large. The rational is the following, the prior is calculated from a set of 40 data point, so wen we have 100 or more training data we have enough information by itself and the use of a poor prior may be disturbing.

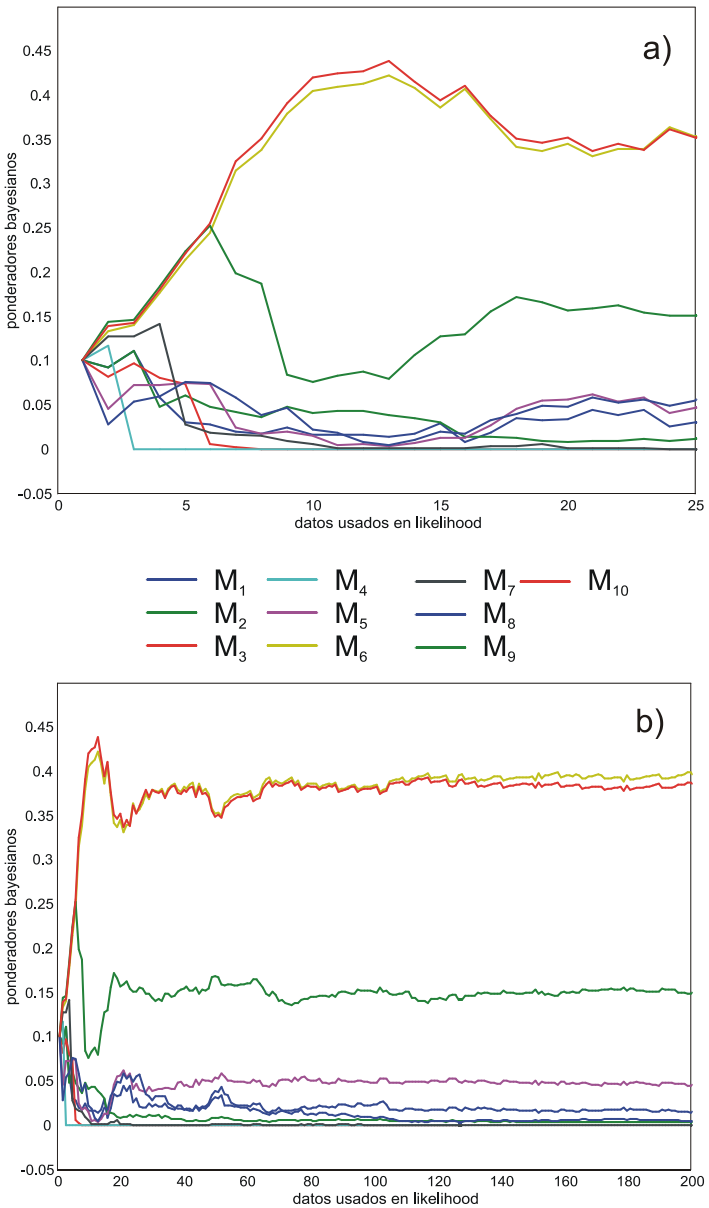

Figure 5: Stability of Bayesian weights as a function of the training data.

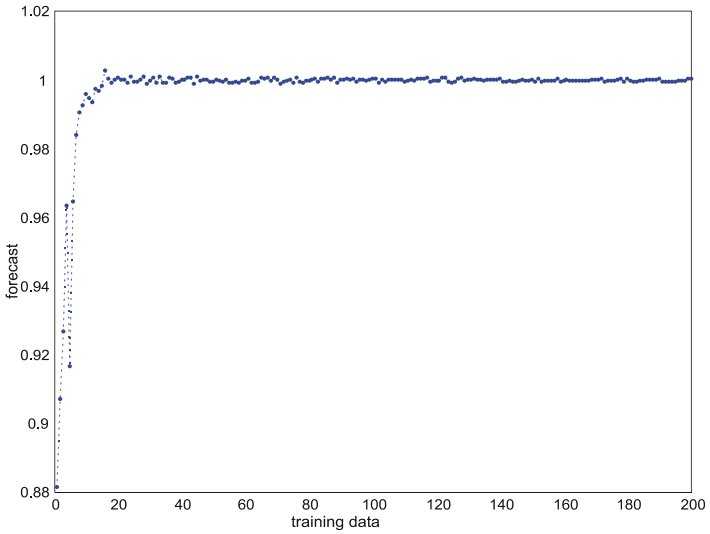

Figure 6: Bayesian weight forecast as a function of the training data.

No. data	19	20	21	22	23	24	25	26
No. Models	29	27	25	24	23	17	15	12

Table 7: Number of participants as a function of the number of complete answers for 2-month-ahead mensual inflation.

5 An application to REM

In this section we show the implementation of the method to real data sets taken from REM. As was mentioned in the introduction, we consider that each of the participants have their own independent model from which they extract expected values for the forecasted variables. We take the forecast series of one and two month ahead monthly inflation from February 2004 to March 2006.

There are a few issues we need to consider before starting the analysis. First, since the REM started at the beginning of 2004, the history of forecasts is relatively short, between 26 and 27 data points. Second, the total number of participants is formally 65, but only a fraction of them has consistently participated in all the periods (21 for one-month-ahead monthly inflation and 12 for two-month-ahead monthly inflation, see Table 7). Third, the method, as was described up to here, assumes a complete set of data. For incomplete samples some technicalities arises that makes the computation of Bayesian weights more cumbersome. Although some efforts have been made in that direction, they are not within the scope of the present paper.

Therefore in what follows, we only show results corresponding to complete

π_{2mth}	real-value	Bayesian avg	Median	Arithmetic avg
Oct-2005	0.8%	0.75%	0.7%	0.69%
Nov-2005	1.2%	1.0%	0.5%	0.583%
Dec-2005	1.1%	0.95%	0.95%	0.95%
Jan-2006	1.3%	1.285%	1.35%	1.275%
Feb-2006	0.4%	0.8%	1.0%	1.03%
March-2006	1.2%	1.208%	1.25%	1.208%

Table 8: Last six 2-month-ahead monthly inflation forecasts.

π_{1mth}	real-value	Bayesian avg	Median	Arithmetic avg
Oct-2005	0.8%	0.78%	0.8%	0.7762%
Nov-2005	1.2%	0.8%	0.7%	0.715%
Dec-2005	1.1%	1.1%	1.2%	1.15%
Jan-2006	1.3%	1.3%	1.3%	1.37%
Feb-2006	0.4%	0.8%	1.0%	0.97%
March-2006	1.2%	1.2%	1.2%	1.19%

Table 9: Efficiency of the methods as a function of the number of training points.

samples, i.e. we reduce the sets of participant[6]. Table 8 shows the results corresponding to the last six two-month-ahead monthly inflation forecasts when the set of 12 participants is employed. In all cases but one, Bayesian averaging gives the best answer to the inflation realized two months later. But what is more remarkable, Bayesian averaging proves to be very good at identifying a change in trend. November-2005 and February-2006 illustrate this ability. In the first case, general expectations were well bellow ($\sim 0.5\%$) the realized value (1.2%), but the Bayesian forecast gave a value of 1.0%. In the second case, the opposite happened, expectations were above the realized value and the Bayesian methodology partially corrected the misperception.

Table 9 shows the results of the last six one-month-ahead monthly inflation forecasts when the set of 21 participants is employed. Again, in all the cases but one, Bayesian averaging gives the best forecast for the inflation realized one month later. The early identification of a change in trend is also present here, although general expectations already started to adjust.

6 Conclusion

Bayesian model averaging can be used to improve the aggregate forecasts generated from a collection of individual model forecasts, even if the models that were used to generate these forecasts are unknown. The part of this process that is least clear is the choice of the likelihood function to be used for generating

[6] In Appendix A we show results for two month ahead mensual inflation for extended sets of participants with some data created artificially. The results show no significant differences.

the Bayesian weights. We work with a normal distribution for the likelihood function and, representing our objective, set the mean of the errors to zero. The remaining crucial variable is the variance of the normal likelihood. We choose this as the value which minimizes the sum of squared errors of the aggregate forecast with respect to the realized values. This method produces a forecast that is statistically superior to five other commonly used methods of producing aggregate forecasts. We explore the characteristics of the Bayesian averaging method for relative efficiency, stability, and the importance of informed priors.

We apply the method to the BCRA's REM data set and find that, even with the short sample that we have, the Bayesian aggregate forecasts dominate the median and arithmetic average. Since the method is able to pick out those REM forecasters who seem to have the best underlying models, it is clearly superior at capturing turning points.

References

[1] BCRA, (2006) *Metodología del Relevamiento de Expectativas de Mercado* (REM), February.

[2] Bates, J.M. and Granger, C.W.J. (1969): "The combination of forecasts". *Operational Research Quarterly* 20, p. 451-468.

[3] Canova, Fabio (2006) *Methods for Applied Macroeconomic Research*, Princeton University Press, Princeton, forthcoming.

[4] Jacobson, Tor, and Sune Karlsson, (2004) "Finding good predictors for inflation: a Bayesian model averaging approach," *Journal of Forecasting*, John Wiley & Sons, Ltd., 23(7), p. 479-496.

[5] MacKay, David J. C. (2003) *Information Theory, Inference, and Learning Algorithms*, Cambridge University Press, Cambridge.

A Completing samples artificially

In general, when there are missing data different statistical factor model techniques are available for solving the problem. However these methods are not reliable when the complete sub-sample size is small compared to the missing data. In the present case, we opted by completing the data generating random values taken from a normal distribution centered in the realized quantity and with a variance equal to the calculated variance of the corresponding participant. We show below two examples considering sets of 19 and 23 artificially completed models As can be see, the results do not differ substantially from the obtained with the reduced 12 set of participants.

π_{2mth} (23 Mod)	real-value	Bayesian avg	Median	Arithmetic avg
Oct-2005	0.8%	0.764%	0.7%	0.68%
Nov-2005	1.2%	1.022%	0.7%	0.69%
Dec-2005	1.1%	0.977%	1.0%	0.78%
Jan-2006	1.3%	1.279%	1.3%	1.25%
Feb-2006	0.4%	0.8%	1.0%	1.01%
March-2006	1.2%	1.2%	1.2%	1.17%

Table 10: Efficiency of the methods as a function of the number of training points.

π_{2mth} (29 Mod)	real-value	Bayesian avg	Median	Arithmetic avg
Oct-2005	0.8%	0.73%	0.7%	0.72%
Nov-2005	1.2%	1.0%	0.7%	0.74%
Dec-2005	1.1%	1.0%	1.0%	1.0%
Jan-2006	1.3%	1.23%	1.2%	1.23%
Feb-2006	0.4%	0.79%	1.0%	1.03%
March-2006	1.2%	1.194%	1.1%	1.19%

Table 11: Efficiency of the methods as a function of the number of training points.