A Strategy to Improve the Survey of Professional Forecasters (SPF) Predictions Using Bias-Corrected-Accelerated (BCA) Bootstrap Forecast Intervals

Mihaela Bratu (Simionescu)
Academy of Economic Studies, Faculty of Cybernetics, Statistics and Economic Informatics, Bucharest, Romania,
mihaela_mb1@yahoo.com

Abstract

Purpose – The purpose of the research is to find a suitable strategy to improve the SPF predictions for inflation

Design/methodology/approach – some alternative forecasts for the annual rate of change for the HICP for EU were developed, their accuracy was evaluated, using proper accuracy measures, and it was compared with the accuracy of Survey of Professional Forecasters (SPF) predictions.

Findings – a synergy of strategies (combined forecasts, BCA bootstrap technique and the method of regression models) proved to be a good strategy to improve the forecasts accuracy on the horizon January 2010-May 2012

Practical implications – the proposed strategy to get more accurate predictions than SPF or naïve ones will help the government in improving the decisions at macroeconomic level and the economic agents in having a better planning activity

Research limitations – the proposed strategy of improving the forecasts accuracy is dependent by the particularities of the data series and it can not be recommended for all types of predictions

Originality/ value – the proposed strategy is an original contribution of the author in literature, the BCA bootstrap intervals never being used in literature to get better predictions. The historical accuracy method used to construct new forecasts is also an original proposal of the author in literature that was also applied in this research.

Keywords: forecasts accuracy; combined forecasts; naïve forecasts; SPF; synergy, BCA bootstrap intervals, historical accuracy (errors) method

JEL Classification: C54, E37

1. Introduction

In addition to economic analysis, the elaboration of forecasts is an essential aspect that conducts the way of developing the activity al macroeconomic level. But any forecast must be accompanied by macroeconomic explanations of its performance. The purpose of this evaluation is related to different aspects: the improvement of the model on which the forecast was based, adjustment of gouverment policies, the planning of results. Basically, performance evaluation in this context refers directly to the degree of trust confered to the prediction. Although the literature on forecasting methods and techniques used in describing the evolution of an economic phenomenon is particularly rich, surprisingly, few researchers have dealt with the methods used to improve the measurement of forecast uncertainty. The aspect is important, because the macroeconomic predictions must not be easily accepted, taking into account the negative consequences of macroeconomic forecasts failures, consequences that affect the state policies. The decisions of economic policy are based on these forecasts. Hence, there is an evident interest of improving their performance.

In literature there are 3 directions in evaluating the performance of macroeconomic forecasts: accuracy, bias and efficiency.A large number of articles have considered the problem of comparing the accuracy measures, contributions in the field are related of names like: Leith and Tanner-1990, Makridakis-1993, Yokum and Armstrong-1995, Tashman-2000, Makridakis and Hibon-2000, Koehler, Martin and Witt -2002, Hyndman -2006 and Witt -2002, Hyndman-2006.

Meese and Rogoff's paper, „Empirical exchange rate models of the seventies", remains the starting point for many researches on the comparing of accuracy and bias. Recently, Genrea, Kenny, Meylera and Timmermann (2013) made forecasts combinations starting from SPF predictions for ECB and using performance-based weighting, trimmed averages, principal components analysis, Bayesian shrinkage, least squares estimates of optimal weights. Only for the inflation rate there was a strong evidence of improving the forecasts accuracy with respect to the equally weighted average prediction.

A detailed presentation of the literature in the domain will be made in the section reserved for this.

A country is directly interested of improving its macroeconomic forecasts in order to get more accurate results and to get better monetary or governmental policy decisions. The companies will be interested in knowing the future evolution of the indicators to improve its planning activity.

The strategies to improve the forecasts accuracy are not clearly defined in literature, the researchers being interested in choosing some subjective practices resulted from experience. Therefore, we will propose some quantitative strategies that will prove their efficiency in accordance to the specificity of the data series.

2. Forecasts accuracy in literature

Forecast accuracy is a large chapter in the literature related to the evaluation of forecasts uncertainty. There are two methods used in comparing the prediction quality: vertical methods (eg, mean squared error) and horizontal methods (such as distance in time). An exhaustive presentation of the problem taking into account all the achievements in literature is impossible, but will outline some important conclusions.

In literature, there are several traditional ways of measurement, which can be ranked according to the dependence or independence of measurement scale. A complete classification is made by RJ Hyndman and AB Koehler (2005) in their reference study in the field, "Another Look at Measures of Forecast Accuracy ".

Hyndman and Koehler introduce in this class of errors "Mean Absolute Scaled Error" (MASE) in order to compare the accuracy of forecasts of more time series.

Other authors, like Fildes R. and Steckler H. (2000) use another criterion to classify accuracy measures. If we consider, $\hat{X}_t(k)$ the predicted value after k periods from the origin time t, then the error at future time (t+k) is: $e_t(t+k)$. Indicators used to evaluate the forecast accuracy can be classified according to their usage. Thus, the forecast accuracy measurement can be done independently or by comparison with another forecast.

A. Independent measures of accuracy

In this case, it is usually used a loss function, but we can also choose the distance criterion proposed by Granger and Jeon for evaluating forecasts based on economic models. The most used indicators are:

a) Mean Square Error (MSE)
b) Root Mean Squared Error (RMSE)
c) Generalized Forecast Error Second Moment (GFESM)
d) Mean Absolute Percentage Error (MAPE)
e) Symmetric Median Absolute Percent Error (SMAPE)
f) Mean error (ME)
g) Mean absolute error (MAE).

In practice, the most used measures of forecast error are:

– Root Mean Squared Error (RMSE) $RMSE = \sqrt{\frac{1}{n}\sum_{j=1}^{n} e_t^2(T_0 + j, k)}$

– Mean error (ME) $ME = \frac{1}{n}\sum_{j=1}^{n} e_t(T_0 + j, k)$

The sign of indicator value provides important information: if it has a positive value, then the current value of the variable was underestimated, which means expected average values too small. A negative value of the indicator shows expected values too high on average.

– Mean absolute error (MAE) $\mathrm{MAE} = \frac{1}{n}\sum_{j=1}^{n} \left| e_t(T_0 + j, k) \right|$

These measures of accuracy have some disadvantages. For example, RMSE is affected by outliers. Armstrong and Collopy stresses that these measures are not independent of the unit of measurement, unless if they are expressed as percentage. Fair, Jenkins, Diebold and Baillie show that these measures include average errors with different degrees of variability. The purpose of using these indicators is related to the characterization of distribution errors. Clements and Hendry have proposed a generalized version of the RMSE based on errors intercorrelation, when at least two series of macroeconomic data are used. If we have two forecasts with the same mean absolute error, RMSE penalizes the one with the biggest errors.

B. Measures for the evaluation of the relative accuracy of forecasts

Relative accuracy measures are related to the comparison of the forecast with a forecast of reference, found in the literature as the 'benchmark forecast' or 'naive forecast. However, it remains a subjective step to choose the forecast used for comparison Problems may occur in this case are related to these aspects: the existence of outliers or inappropriate choice of models used for predictions and the emergence of shocks. A first measure of relative accuracy is Theil's U statistic, which uses as reference forecast the last observed value recorded in the data series. Collopy and Armstrong have proposed instead of U a new similar indicator (RAE). Thompson improved MSE indicator, suggesting a statistically determined MSE- log mean squared error ratio.

A common practice is to compare the forecast errors with those based on a random-walk. "Naïve model" method assumes that the variable value in the next period is equal to the one recorded at actual moment. U-Theil proposed the calculation of U, that takes into account both changes in the negative and the positive sense of an indicator:

$$U = \sqrt{\frac{\sum (X_{t+k} - \hat{X}_t(k))^2}{\sum X_{t+k}^2}}.$$

U Theil's statistic is calculated in two variants by the Australian Tresorery in order to evaluate the forecasts accuracy.

The following notations are used:
a – the registered results
p – the predicted results
t – reference time
e – the error (e=a-p)
n – number of time periods

$$U_1 = \frac{\sqrt{\sum_{t=1}^{n} (a_t - p_t)^2}}{\sqrt{\sum_{t=1}^{n} a_t^2} + \sqrt{\sum_{t=1}^{n} p_t^2}}$$

The more closer of zero is, the forecasts accuracy is higher.

$$U_2 = \sqrt{\frac{\sum_{t=1}^{n-1} (\frac{p_{t+1} - a_{t+1}}{a_t})^2}{\sum_{t=1}^{n-1} (\frac{a_{t+1} - a_t}{a_t})^2}}$$

If U2=1=> there are not differences in terms of accuracy between the two forecasts to compare

If U2<1=> the forecast to compare has a higher degree of accuracy than the naive one

If U2>1=> the forecast to compare has a lower degree of accuracy than the naive one.

Hyndman and Koehler proposed scale errors based on the mean absolute error of a naive forecasting method. MAE serves therefore, as denominator. Using this method, it is generated the one-step-ahead forecast. Scale error is defined as:

$es_t = \dfrac{e_t}{\frac{1}{n-1}\sum_{i=2}^{n}|X_i - X_{i-1}|}$ and mean absolute scale error as: MASE= medie | es_t |.

Naive forecast values are considered to be the current ones recorded during the previous period. MASE is used both to compare forecast methods applied to a given set of data and also to compare the accuracy of several series. If the scale error is less than 1, the compared forecast is better than the reference one (naïve forecast).

One of the business objectives in forecasting was empirical validation. Research groups around the world made comparisons between different methods of forecasting. In literature the results are known as "M-competition". Ex-ante forecast errors for 21 methods were compared with predictions based on 1001 economic series. Accuracy criteria used in the M competition were: central tendency error (APE median), MSE, which gives more weight to larger error, MAPE, which is the basic measure. This is the measure recommended in reference books in forecast accuracy domain, written by Hanke and Reitsch or Bowerman, O'Connell and Koehler.

Recent studies target accuracy analysis using as comparison criterion different models used in making predictions or the analysis of forecasted values for the same macroeconomic indicators registered in several countries.

Ericsson NR (1992) shows that the parameters stability and mean square error of prediction are two key measures in evaluation of forecast accuracy, but they are not sufficient and it is necessary the introduction of a new statistical test.

Considering the AR (1) process, which is represented as y t = βy t-1 + u t, Hoque A., Magnus JR and Pesaran B. (1988) show that for small values of β the prediction mean square error is a decreasing function in comparison with the number of forecast periods.

CWJ Granger and Y. Jeon (2003) consider four models for U.S. inflation: a univariate model, a model based on an indicator used to measure inflation, a univariate model based on the two previous models and a bivariate model. Applying the mean square error criterion, the best prediction made is the one based on an autoregressive model of order 1 (AR (1)). Applying distance-time method, the best model is the one based on an indicator used to measure the inflation.

Ledolter J. (2006) compares the mean square error of ex-post and ex ante forecasts of regression models with transfer function with the mean square error of univariate models that ignore the covariance and show superiority of predictions based on transfer functions.

T. Teräsvirta, van Dijk D., Medeiros MC (2005) examine the accuracy of forecasts based on linear autoregressive models, autoregressive with smooth transition (STAR) and neural networks (neural network-NN) time series for 47 months of the macroeconomic variables of G7 economies. For each model is used a dynamic specification and it is showed that STAR models generate better forecasts than linear autoregressive ones.

Neural networks over long horizon forecast generated better predictions than the models using an approach from specific to general.

U. Heilemann and Stekler H. (2007) explain why macroeconomic forecast accuracy in the last 50 years in G7 has not improved. The first explanation refers to the critic brought to macroeconometrics models and to forecasting models, and the second one is related to the unrealistic expectations of forecast accuracy. Problemes related to the forecasts bias, data quality, the forecast process, predicted indicators, the relationship between forecast accuracy and forecast horizon are analyzed.

Ruth K. (2008), using the empirical studies, obtained forecasts with a higher degree of accuracy for European macroeconomic variables by combining specific sub-groups predictions in comparison with forecasts based on a single model for the whole Union.

Gorr WL (2009) showed that the univariate method of prediction is suitable for normal conditions of forecasting while using conventional measures for accuracy, but multivariate models are recommended for predicting exceptional conditions when ROC curve is used to measure accuracy.

Dovern J. and J. Weisser (2011) used a broad set of individual forecasts to analyze four macroeconomic variables in G7 countries. Analyzing accuracy, bias and forecasts efficiency, resulted large discrepancies between countries and also in the the same country for different variables. In general, the forecasts are biased and only a fraction of GDP forecasts are closer to the results registered in reality.

Reeve and Vigfusson (2011) compared the performance of forecasts based on futures, choosing as a reference model a random walk and a random walk with drift.

In Netherlands, experts make predictions starting from the macroeconomic model used by the Netherlands Bureau for Economic Policy Analysis (CPB). For the period 1997-2008 was reconstructed the model of the experts macroeconomic variables evolution and it was compared with the base model. The conclusions of Franses PH, Kranendonk HC şi Lanser D. (2011) were that the CPB model forecasts are in general biased and with a higher degree of accuracy.

Franses, McAleer and Legerstee (2012) evaluated two forecasts based on three different methods: the two forecasts are based on different econometric models, one of the prediction is based on an econometric model and the other uses a model and also the intuition, both forecasts are the result of econometric models and intuition.

Deschamps and Bianchi (2012) concluded that there are large differences between macroeconomic forecasts for China regarding the accuracy measures for consumption and investment, GDP and inflation. The slow adjustment to structural shocks generated biased predictions, the information being utilized relatively inefficient.

Allan (2012) obtained a good accuracy for the OECD forecasts combined with outturn values of GDP growth for G7 countries between 1984 and 2010. The same author mentioned two groups of accuracy techniques used in assessing the predictions: quantitative forecasts accuracy statistics and qualitative accuracy methods. In our study we are interested by the first category of techniques that is used to evaluate the accuracy of an institution or to compare the accuracy of different predictions.

Clements and Galvao (2012) proved using empirical data that a mixed data-frequency sampling (MIDAS) approach can improve the accuracy of inflation and GDP growth predictions at short horizons (less than one year).

Clark and Ravazzolo (2012) compared, in terms of accuracy, the forecasts based on Bayesian autoregressive model and Bayesian vector autoregressive one with volatility that is variable in time. The most important macroeconomic variables were chosen for USA and England, the results showing a better accuracy of predictions based on AR and VAR with stochastic variance.

Many studies in literature refer to the combining of two methods based on the same model (such as eg bayesian mediation model), but French and Insura point out that a combination between model predictions and expert assessments has not been proposed yet.

3. The assessment the forecasts accuracy

The Survey of Professional Forecasters provided quarterly survey of macroeconomic predictions since 1968, being conducted since 1990 by the Federal Reserve Bank of Philadelphia.

The monthly data for the annual HICP (harmonized index of prices) inflation rate are published by Eurostat and the predictions are made by SPF (Survey of Professional Forecasters) for January 2010–May 2012.

The monthly data for the annual rate of change for the HICP is not stationary, being necessary to differentiate the data. The stationarized data series for January 1997–December 2009 follows a random walk process: $\Delta ir_t = 0{,}339 \cdot \Delta ir_{t-1} + \varepsilon_t$. Starting from this econometric model, the predictions for January 2010-May 2012 are made.

The strategies proposed to improve the SPF predictions are: the combined forecasts, the historical accuracy method

We refer to the most used combination approaches used in order to improve the forecasts accuracy:

- optimal combination (OPT), with weak results according to Timmermann (2006);
- equal-weights-scheme (EW);
- inverse MSE weighting scheme (INV).

Bates and Granger (1969) considered two predictions p_{1t} and p_{2t} for the same variable Xt, derived h periods ago. We admit that the two predictions are made for the same macroeconomic variable. If the forecasts are unbiased, the error is calculated as:

$e_{i,t} = X_{i,t} - p_{i,t}$. The errors follow a normal distribution of parameters 0 and σ_i^2.

σ_i^2 – the errors variance

Xt – the realization of the variable at time t

pt – the prediction at time t

If ρ is the correlation between the errors, then their covariance is $\sigma_{12} = \rho \cdot \sigma_1 \cdot \sigma_2$. The linear combination of the two predictions is a weighted average:

$c_t = m \cdot p_{1t} + (1 - m) \cdot p_{2t}$.

c_t – the combined prediction

m – the coefficient for the combined forecast

The error of the combined forecast is: $e_{c,t} = m \cdot e_{1t} + (1 - m) \cdot e_{2t.}$ The mean of the combined forecast is zero and the variance is:

$\sigma_c^2 = m^2 \cdot \sigma_1^2 + (1 - m)^2 \cdot \sigma_{2t}^2 + 2 \cdot m \cdot (1- m) \cdot \sigma_{12.}$ By minimizing the error variance, the optimal value for m is determined (m_{opt}):

$$m_{opt} = \frac{\sigma_2^2 - \sigma_{12}}{\sigma_1^2 + \sigma_2^2 - 2 \cdot \sigma_{12}}.$$

The individual forecasts are inversely weighted to their relative mean squared forecast error (MSE) resulting INV. In this case, the inverse weight (m_{inv}) is:

$$m_{inv} = \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2}.$$

Equally weighted combined forecasts (EW) are gotten when the same weights are given to all models.

We can build new forecasts starting from a regression model that explains the registered values of the rate of change using the SPF values. This is second strategy proposed by author. The regression uses time series from 1997–2010 to make predictions for 2010–May 2012. Two valid regression models were selected: M1 and M2.

EFFECTIV = 2.127022766 - 0.05534008024*f_SPF
EFFECTIV = 1.689861546 + 0.6027484692*(1/f_SPF)

where EFFECTIVE – the registered (effective) values of annual inflation change in HICP

f_SPF – the forecasts made by SPF

The new forecasts are got starting from these regression models and knowing the SPF values.

Table 1: Indicators of forecasts accuracy (January 2010–May 2012)

Accuracy indicators	Predictions based on random walk	SPF predictions	Combined foreca sts (OPT scheme)	Combined forecasts (INV scheme)	Combined forecasts (EW scheme)	Mean of the forecasts	Median of the forecasts	Forecasts based on M1	Forecasts based on M2
RMSE	0,634	0,204	0,231	0,271	0,221	0,281	0,231	0,833	0,422
ME	-0,521	-0,018	-0,113	-0,171	-0,094	-0,183	-0,113	-0,617	0,332
MAE	0,534	0,157	0,172	0,204	0,165	0,214	0,172	0,702	0,362
MAPE	0,223	0,065	0,070	0,082	0,067	0,087	0,070	0,247	0,154
U1	0,129	0,038	0,044	0,052	0,042	0,054	0,044	0,178	0,074
U2	3,195	1,068	1,226	1,426	1,177	1,477	1,226	2,948	1,946

Source: processing of data provided by Eurostat and SPF

The SPF forecasts are the best ones, because of the low values for all accuracy indicators. All the predictions are overestimated, the ME values being negative. In average the SPF errors differ with 6,5% from the registered values. All the mentioned predictions are not better than the naïve ones, because of the values greater than 1 for U2 statistics. The median of forecasts is equal to the optimal combined prediction on the entire forecasting horizon.

Another strategy of accuracy improvement is based on the bias-corrected-accelerated (BCA) bootstrap technique which supposes the use of BCA method from Resampling Stats. Another modality to generate confidence intervals (CI) is, according to Efron and Tibshirani (1993), to build bootstrap percentile intervals. Actually, the resample mean values are repeated and scored in this case, and 100 replications are done. The steps of BCA bootstrap method are described by Lunneborg (2000), who calculated the accelaration estimate starting from jacknifed estimates. Then, a bootstrap sampling was generated starting from the initial sample and the bias was estimated. Finally, the z scores from the normal repartition are included to build the BCA confidence interval.

The bias-corrected-accelerated interval (BCA) is a complex bootstrap technique used to construct confidence intervals.

In this case, Davison and Hinkley (1997) showed that estimates for bias and acceleration are provided using the initial sample and the bootstrap samples. The acceleration is the way in which the variance modifies when the parameter of the population increases. The bias is computed as difference between the real value of the parameter and the median of parameter sampling repartition.

We computed the BCA bootstrap intervals for all the proposed forecasts: the three types of combined forecasts, their mean and media and the forecasts based on M1 and M2 models. The inferior limit and the superior one of these intervals are chosen as point predictions and the accuracy was evaluated.

The values of U1 and U2 statistics are presented in the following table, the lower limits of BCA intervals providing an improvement of forecasts accuracy. These point forecasts are even better than the naïve ones. Actually, a synergy of methods is used to grow the degree of accuracy. The previous two strategies (the combined forecasts, their mean and median and the predictions based on the two regression models) are associated with the bootstrap technique.

Table 2: U1 and U2 indicators of forecasts accuracy (January 2010–May 2012)

Forecasts based on:	U1	U2
Lower limit of the BCA bootstrap intervals	0.001983	0.003308
Upper limit of the BCA bootstrap intervals	0.71386	0.85648182

The evolution of the forecasts based on the BCA bootstrap method and the registered inflation rate are presented in the following figure. In the first period of the forecasting horizon the predictions are closer of the lower limits, while in the second one they are closer of the superior limits.

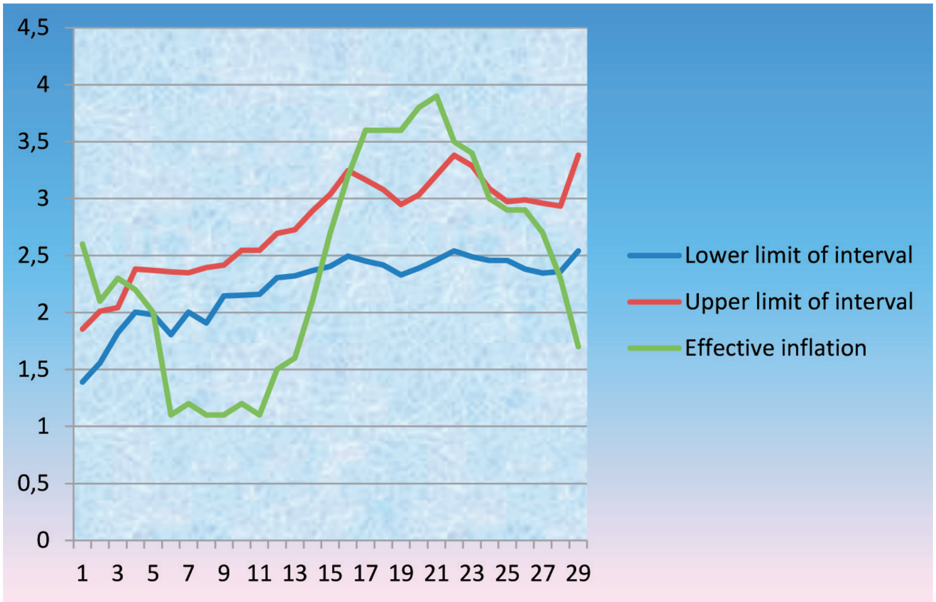

Figure 1: The evolution of point forecasts based on BCA bootstrap intervals and the effective inflation

Another interesting strategy is, according to Bratu (2012) to build new predictions considering that these have as MPE, the mean percentage error, or other accuracy indicator registered for 1997-2009. We used the MPE of SPF predictions or of forecasts based on the AR(1) model. We can replace MPE with the other indicators (ME, MAE, RMSE).

$$MPE = \frac{X_{t+1} - X_t}{X_t} \Rightarrow \frac{X_{t+1}}{X_t} - 1 = MPE \Rightarrow X_{t+1} = (MPE + 1) \cdot X_t$$

$$MPE = \frac{X_{t+1} - X_t}{X_t} \Rightarrow \frac{X_{t+1}}{X_t} - 1 = MPE \Rightarrow X_{t+1} = (MPE + 1) \cdot X_t$$

$$MAE1 = X_{t+1} - X_t \Rightarrow X_{t+1} = MAE1 + X_t$$

$$MAE2 = -X_{t+1} + X_t \Rightarrow X_{t+1} = -MAE2 + X_t$$

$$RMSE^2 = X_{t+1} - X_t \Rightarrow X_{t+1} = RMSE^2 + X_t$$

Table 3: Accuracy indicators for predictions of annual change of HICP (1997–2009)

	ME	MAE	RMSE	MPE
SPF forecasts	-0,021	0,403	0,518	-0,023

To build the predictions for 2010–May 2012 we take into account the accuracy indicator for 1997-2009 and the previous SPF forecasted value, but all the predictions have a lower degree of accuracy than SPF forecasts and the random walk. All the new predictions are overestimated with a rather high degree of accuracy, because of the negative values of ME. In this case, we have an improvement of SPF forecasts according to all accuracy indicators for the predictions based on ME and the previous registered value for the annual change of price index. However, these predictions are not better than the naïve forecasts.

Table 4: Accuracy indicators for forecasts based on a historical accuracy indicator (January 2010–May 2011)

Accuracy indicators	Forecasts based on ME and SPF previous prediction	Forecasts based on MAE1 and SPF previous predictions	Forecasts based on MAE2 and SPF previous predictions	Forecasts based on RMSE and SPF previous prediction	Forecasts based on MPE and SPF previous predictions	SPF prediction
RMSE	0,787	0,508	1,119	0,577	0,799	0,204
ME	-0,638	-0,215	-1,020	-0,348	-0,664	-0,018
MAE	0,649	0,424	1,020	0,471	0,668	0,157
MAPE	0,275	0,178	0,424	0,199	0,282	0,065
U1	0,0163	0,097	0,250	0,113	0,167	0,038
U2	4,067	2,701	5,548	3,078	4,089	1,068

A rather low degree of accuracy was registered for predictions based on MPE and the previous predicted value of SPF. All the new forecasts, excepting those based on MAE1, are overestimated.

The method based on BCA bootstrap forecasts intervals gave the best results, being a good strategy of improving the accuracy of the SPF predictions.

4. Conclusions

Many researchers are interested in making predictions for macroeconomic variables, but few of them studied the accuracy of their forecasts. The problem is essential, especially in crisis periods, because from many forecasts made for the same indicator only one or few are the most accurate.

The accuracy indicators of ex-post forecasts gives us a hint about the way we will chose to build better forecasts, according to the indicator we want to have the lowest value. In this study, the accuracy of SPF forecasts for monthly annual rate of change for HICP was evaluated and some strategies to improve the accuracy were proposed. It seems that the classical approaches from literature didn't improve the accuracy, but the empirical strategy proposed by Bratu (2012) for USA gave good results for EU. So, the author proposed an original strategy based on BCA bootstrap technique to improve the SPF forecasts. We also had an improvement of SPF forecasts according to U1 indicator for the predictions based on ME and the previous registered value for the annual change of HICP.

The lower and upper limits of BCA bootstrap intervals are treated as point forecasts. The intervals were constructed for the combined predictions, their median and mode and those based on the proposed econometric models. Actually, the predictions built starting from this technique of synergy are the most accurate, outperforming even the naïve predictions. The BCA bootstrap method gave the best forecasts, this application in this domain being an original contribution of the author. A considerably improvement of the accuracy was also got for predictions based on mean error of SPF expectations for 1997-2009 and the previous registered value, according to U1 Theil's statistic indicator.

In conclusion, macroeconomic forecasts evaluation is necessary to inform the public about the way in which SPF or other institution predicted the economic phenomenon. Further, the people interested in the prediction of a variable will chose a certain strategy to improve the SPF predictions, according to historical approaches.

5. References

Allan G. (2012), Evaluating the usefulness of forecasts of relative growth, Strathclyde, Discussion Papers in Economics, No. 12–14.

Armstrong, J. S. & Collopy, F. (2000). Another Error Measure for Selection of the Best Forecasting Method: The Unbiased Absolute Percentage Error. *International Journal of Forecasting,* Vol. 8, no. 2, pp. 69–80.

Armstrong, J. S. & Fildes, R. (1995). On the selection of Error Measures for Comparisons Among Forecasting Methods. *Journal of Forecasting*, Vol. 14, no. 3, pp. 67–71.

Ashley, R. (2003), Statistically significant forecasting improvements: how much out-of-sample data is likely necessary ?. *International Journal of Forecasting*, Vol. 19, no. 2, pp. 229–239.

Bates, J., & C. W. J. Granger (1969). The Combination of Forecasts, *Operations Research Quarterly*, Vol. 20, no. 4, pp. 451–468.

Bratu, M. (2012). *Strategies to Improve the Accuracy of Macroeconomic Forecasts in USA*, Germany: LAP LAMBERT Academic Publishing, ISBN-10: 3848403196, ISBN-13**:** 978–3848403196.

Clark, T. E. & McCraken, M. W. (2006). Forecast with Small Macroeconomic VARS in the Presence of Instabilities, *The Federal Reserve Bank of Kansas City*, Economic Research Department.

Clark T.E., Ravazzolo F. (2012), The Macroeconomic Forecasting Performance of Autoregressive Models with Alternative Specifications of Time-Varying Volatility, *FRB of Cleveland,* Working Paper No. 12–18.

Clements M.P., Galvao A.B. (2012), Macroeconomic Forecasting with Mixed Frequency Data: Forecasting US output growth and inflation, *The Warwick Economics Research Paper Series (TWERPS) from University of Warwick, Department of Economics*, no. 773

Clements, M. P. & Hendry, D. F. (1995). Forecasting in cointegrated systems. *Journal of Applied Econometrics*, Vol. 10,no. 2, pp. 127–146.

Diebold, F.X. & Mariano, R. (1995). Comparing Predictive Accuracy. *Journal of Business and Economic Statistics*, Vol. 13, no. 1, pp. 253–265.

Davison A.C., Hinkley D.V. (1997), *Bootstrap methods and their applications*, Cambridge University Press.

Deschamps, B. and Bianchi, P. (2012), An evaluation of Chinese macroeconomic forecasts**,** Journal of Chinese Economics and Business Studies, Volume 10, Number 3, 1 August 2012, pp. 229–246(18).

Diebold, F.X. (1998). The Past, Present and Future of Macroeconomic Forecasting, *Journal of Economic Perspectives*, Vol. 12, no. 2, pp. 175–192.

Dobrescu E. (2006). Macromodels of the Romanian Transition Economy. Bucureşti: *Expert Publishing House.*

Dobrescu E. (2003). Macroeconomic estimators for the Romanian "Pre-accesion Economic Program" (the 2003 version). *Romanian Journal of Economic Forecasting,* 2003, volume 4, issue 4, supliment.

Dovern, J. & Weisser J. (2011). Accuracy, unbiasedness and efficiency of professional macroeconomic forecasts: An empirical comparison for the G7. *International Journal of Forecasting*, Vol. 27, no. 2, pp. 452–465.

Efron B., Tibshirani R. (1993), *An introduction to the bootstrap*, CRC Press

Fildes, R. & Steckler, H. 2000. The State of Macroeconomic Forecasting. Lancaster University EC3/99, George Washington University, *Center for Economic Research*, Discussion Paper No. 99-04, pp. 33–45.

Franses, P.H., McAleer, M. and Legerstee, R. (2012) *Evaluating Macroeconomic Forecasts: A Concise Review of Some Recent Developments.*

Genrea V., Kenny G., Meylera A., Timmermann A. (2013), Combining expert forecasts: Can anything beat the simple average?, I*nternational Journal of Forecasting*, Volume 29, Issue 1, January–March 2013, p. 108–121.

Gorr, W. L. 2009. Forecast accuracy measures for exception reporting using receiver operating characteristic curves. *International Journal of Forecasting*, Vol. 25, no. 1, pp. 48–61.

Granger, C. W. J. & Jeon, Y. (2003). Comparing forecasts of inflation using time distance. *International Journal of Forecasting*, Vol. 19, no. 3, pp. 339–349.

Heilemann, U. & Stekler, H. (2007). Introduction to "The future of macroeconomic forecasting". *International Journal of Forecasting*, 23(2), p. 159–165.

Hyndman, R. J. & Koehler, A.B. 2006. Another Look at Measures of Forecast Accuracy. *International Journal of Forecasting*, 22 (4), p. 679–688.

Lunneborg C. (2000), *Data analysis of resampling*, Duxbury Press, p. 164

Makridakis,, S. (1984). *Forecasting: Methods and Applications*, New York: Wiley & Sons, pp. 122.

Makridakis, S., Wheelwright, S.C. & Hyndman R.J. (1998). *Forecasting: Methods and Applications*. third edition. John Wiley & Sons, New York, pp. 642.

Reeve, T.A. and Vigfusson, R.J. (2011), *Evaluating the forecasting performance of commodity future prices*, Board of Governors System, International Finance Discussion Papers, No. 1025.

Ruth, K. 2008. Macroeconomic forecasting in the EMU: Does disaggregate modeling improve forecast accuracy ?. *Journal of Policy Modeling*, Vol. 30, no. 3, pp. 417–429.

Teräsvirta, T., van Dijk, D. & Medeiros, M.C. (2005). Linear models, smooth transition autoregressions, and neural networks for forecasting, macroeconomic time series: A re-examination. *International Journal of Forecasting*, Vol. 21, no. 4, pp. 755–774.

Timmermann, A. (2006). Forecast Combinations, chap. 4, pp. 135-196,

Handbook of Economic Forecasting. G. Elliott, C. Granger, and A. Timmermann, Elsevier.

EUROSTAT WEBSITE http://sdw.ecb.europa.eu/browseExplanation.do?sk=122.ICP.M.U2.N.000000.4.ANR&node=SEARCHRESULTS&trans=N, date: June 2012

APPENDIX 1

Combined forecasts based on random walk process and SPF predictions on the forecasting horizon 2010–May 2012

Month	Combined forecasts (%) (OPT scheme)	Combined forecasts (%) (INV scheme)	Combined forecasts (%) (EW scheme)	Mean of the forecasts (%)	Median of the forecasts (%)	Forecasts based on M1	Forecasts based on M2
ian.10	1,450	1,314	1,492	1,285	1,450	2,099	1,991
feb.10	1,605	1,504	1,636	1,483	1,605	2,077	2,232
mar.10	1,332	1,271	1,351	1,257	1,332	2,077	2,232
apr.10	1,912	1,754	1,962	1,719	1,912	2,083	2,172
mai.10	2,034	1,972	2,053	1,958	2,034	2,039	2,654
iun.10	2,006	1,948	2,024	1,936	2,006	2,039	2,654
iul.10	1,807	1,792	1,812	1,789	1,807	2,033	2,714
aug.10	2,042	1,965	2,066	1,948	2,042	2,044	2,594
sep.10	1,916	1,885	1,926	1,879	1,916	2,033	2,714
oct.10	2,251	2,158	2,280	2,138	2,251	2,039	2,654
nov.10	2,225	2,179	2,239	2,169	2,225	2,022	2,835
dec.10	2,225	2,179	2,239	2,169	2,225	2,022	2,835
ian.11	2,660	2,552	2,694	2,528	2,660	2,022	2,835
feb.11	2,606	2,548	2,624	2,536	2,606	2,005	3,016
mar.11	2,842	2,765	2,866	2,748	2,842	2,000	3,076
apr.11	3,023	2,935	3,051	2,915	3,023	1,994	3,136
mai.11	3,242	3,165	3,266	3,148	3,242	1,978	3,317
iun.11	3,097	3,055	3,111	3,046	3,097	1,972	3,377
iul.11	2,997	2,955	3,011	2,946	2,997	1,978	3,317
aug.11	2,807	2,792	2,812	2,789	2,807	1,978	3,317
sep.11	2,952	2,902	2,968	2,891	2,952	1,983	3,257
oct.11	3,232	3,128	3,265	3,105	3,232	1,989	3,197
nov.11	3,352	3,302	3,368	3,291	3,352	1,961	3,498
dec.11	3,216	3,185	3,226	3,179	3,216	1,961	3,498
ian.12	2,917	2,929	2,914	2,932	2,917	1,961	3,498
feb.12	2,835	2,816	2,841	2,812	2,835	1,978	3,317
mar.12	2,862	2,839	2,870	2,834	2,862	1,978	3,317
apr.12	2,844	2,809	2,854	2,801	2,844	1,983	3,257
mai.12	2,645	2,653	2,642	2,654	2,645	1,978	3,317

Source: own calculations using Excel

APPENDIX 2

BCA bootstrap intervals for the chosen forecasts on the forecasting horizon 2010–May 2012

Month	Lower limit of interval	Upper limit of interval
ian.10	1.3894285	1.85514285
feb.10	1.5587142	2.01077827
mar.10	1.8217142	2.04493584
apr.10	2.002	2.381357048
mai.10	1.9811428	2.37145911
iun.10	1.8048812	2.35571428
iul.10	2.0039740	2.34985714
aug.10	1.91	2.39617296
sep.10	2.1477142	2.41539632
oct.10	2.152861	2.5465714
nov.10	2.1587272	2.54515782
dec.10	2.3069086	2.69471428
ian.11	2.3227632	2.72828571
feb.11	2.3684096	2.89125171
mar.11	2.4048802	3.03857142
apr.11	2.4957564	3.24495055
mai.11	2.4522991	3.165
iun.11	2.4167427	3.07842857
iul.11	2.3311815	2.94961091
aug.11	2.3903985	3.03149661
sep.11	2.4630711	3.20857142
oct.11	2.5412857	3.38073111
nov.11	2.4898770	3.28828571
dec.11	2.4562064	3.08685714
ian.12	2.4573309	2.97171428
feb.12	2.3810979	2.98812602
mar.12	2.3471747	2.95842857
apr.12	2.3604285	2.9341428
mai.12	2.5412857	3.38073111

Source: own calculations using Excel