

*Mihaela Bratu (Simionescu) **

COMBINED FORECASTS USED TO IMPROVE THE PERFORMANCE OF UNEMPLOYMENT RATE FORECASTS IN ROMANIA

The main goals of this research are: the evaluation of unemployment rate forecasts made for Romania by three experts in forecasting for 2001–2011 (E1, E2 and E3) and the proposal of some combined forecasts to improve the predictions performance. The predictions are made at the same time. The forecasts performance implies three directions: accuracy, biasedness and efficiency. In addition to the usual accuracy indicators, multi-criteria ranking methods were used to make a hierarchy of experts according to the forecasts accuracy and the forecasts overall performance, resulting in the following classification: E3, E1 and E2. Besides the combined forecasts based on three classical schemes (optimal scheme, equally-weighted scheme, inverse MSE weighting scheme), the combined predictions proposed by us based on mean error, improved the performance, solving in most of the cases the biasness problem and generating more accurate and efficient forecasts.

Keywords: forecasts, predictions, performance, accuracy, bias, efficiency, multi-criteria ranking methods, combined forecasts

JEL Classification: E21, E27, C51, C53

DOI: 10.15611/aoe.2015.1.04

1. INTRODUCTION

Performance of the forecasts is related to the quality of the provided predictions. Everyone is interested in the best forecasts, but these are the ones that respect the performance criterion. In practice, many people tend to assume that performance means accuracy, but this assumption is not quite correct. Actually, performance assumes three directions: accuracy, biasedness and efficiency. The real motivation for the evaluation and improvement of forecasts performance is related to the fact that nowadays the interest in the better anticipation of future macroeconomic evolution is growing more than ever in the context of the current economic crisis. The decisional process at all levels (from governmental level to microeconomic one) refers mostly to phenomena that will continue in the future. But the

* Romanian Academy, Bucharest, Romania

major problem is that we do not know the future, but we can anticipate it more or less accurately.

The performance of the unemployment rate forecasts should be known to governmental decisional bodies, employment agencies, researchers interested in the labour market and even employees and unemployed people. It is a subject of interest for overall public opinion. Many studies have treated the problem of the performance evaluation of macroeconomic forecasts, but only few of them are related to unemployment predictions. Most of the articles assessed the performance for the GDP, the exchange rate and the inflation rate forecasts.

Many important national and international institutions use their own macroeconomic forecasts. In general, the tendency is to compare the predictions of the institution with those of other famous organizations, but fewer researchers take government expectations as a benchmark.

Allan (2012) improved the accuracy of OECD unemployment forecasts made for G7 countries by applying the combination technique. He used two types of methods to assess accuracy: quantitative techniques and qualitative accuracy methods.

A detailed study regarding unemployment forecasting and predictions performance was made by Barnichon and Nekarda (2012), who proposed a model for the unemployment rate that outperformed in terms of the predictions of the results offered by the classical time series and by the Survey and Professional Forecasters and Federal Reserve Board.

Franses, McAleer and Legerstee (2012) evaluated the performance of unemployment forecasts made by the Federal Reserve Board and Federal Open Market Committee, the Diebold-Mariano test indicating insignificant differences in terms of forecasts accuracy.

Kurita (2010) used an ARFIMA model to make forecasts for Japan's unemployment rate, his predictions outperforming those based on random walk.

Heilemann and Stekler (2007) gave some reasons for the lack of accuracy of G7 predictions in the last 50 years. There is a continuous critique brought to macro-econometrics models and to forecasting techniques, but also the accuracy expectations are not realistic. Other aspects of the forecasts failure are related to: forecasts bias, data quality, the forecasting procedure, type of predicted indicators, the relationship between forecast accuracy and forecast horizon.

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

One of the main goals of our research is to assess the performance of Romania's unemployment rate forecasts using a newly proposed method of evaluating the accuracy, by considering more indicators of accuracy at the same time. Instead of the classical U_1 Theil's coefficient used for making comparisons, we can use multi-criteria methods.

Another important objective of this research is related to the finding of suitable strategies for improving the forecasts performance. The concept of an "empirical strategy for improving the forecasts performance" was introduced in literature by Bratu (2012), who proposed some strategies for the improvement of the USA's macroeconomic variables.

The main goals are achieved by assessing the performance of the unemployment rate forecasts provided for Romania by the experts in forecasting denoted by Expert 1 (E1), Expert 2 (E2) and Expert 3 (E3). The ranking method is applied to make a hierarchy of those experts according to the forecasts performance. The most accurate predictions were provided by E3 and the least accurate by E2. The E1 excels at efficiency, which is a weak point for E3. The biasedness seems to be a problem for all the experts for 2001–2011.

On the other hand, different combined forecasts are built to check if these improved the experts' forecasts. The most used combination approaches were utilized in this study: the optimal combination (OPT), the equal-weights-scheme (EW) and the inverse MSE weighting scheme (INV). We proposed our own combination procedure based on the mean errors on the forecasts horizon, this strategy outperforming the other ones.

2. HIERARCHY OF EXPERTS ACCORDING TO UNEMPLOYMENT FORECASTS PERFORMANCE

The data sets are represented by the predicted values of the annual registered unemployment rate made at the same time for Romania by three experts: E1, E2 and E3. The predictions horizon is 2001–2011. The objective is to assess the accuracy, the biasness and the efficiency of these predictions and determine the best expert with the highest performance.

Armstrong and Fildes (1995) recommended the use of more measures of accuracy. Therefore, more accuracy indicators were calculated for the three categories of predictions for 2001–2011.

Multi-criteria ranking methods (ranking method and the method of relative distance with respect to maximal performance) are used to make the

hierarchy of forecasters according to accuracy criterion. More accuracy measures are taken into account at the same time. Next the same method is applied to make a hierarchy according to forecasts performance by taking into consideration three criteria: accuracy, biasness and efficiency.

The prediction error is computed as the difference between the effective value and the forecasted one of a variable X and is denoted by e_X . For the number of forecasts on the horizon we used the notation “ n ”. The most frequently used statistical measures for assessing the forecasts accuracy, according to Bratu (2012), are:

Root Mean Squared Error (RMSE):

$$RMSE = \sqrt{\frac{1}{n} \sum_{j=1}^n e_X^2},$$

Mean Error (ME):

$$ME = \frac{1}{n} \sum_{j=1}^n e_X,$$

Mean Absolute Error (MAE):

$$MAE = \frac{1}{n} \sum_{j=1}^n |e_X|.$$

RMSE is influenced by outliers. These absolute measures depend on the unit of measurement, this disadvantage being eliminated unless the indicators are expressed as percentage.

U Theil’s statistic, used in making comparisons between predictions, can be used in two variants, as presented also by the Australian Treasury.

The next notations are used:

a – actual/registered value of the variable,

p – value for the predicted variable,

t – time,

e – error (difference between actual value and the forecasted one),

n – number of periods.

U_I takes a value between 0 and 1, a closer value to zero indicating a better accuracy for that prediction. If there are alternative forecasts for the same variable, the one with the lowest value of U_I is the most accurate.

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

If $U_1 = 1 \rightarrow$ there are no differences in terms of accuracy between the two forecasts to compare.

If $U_1 < 1 \rightarrow$ the forecast to compare has a higher degree of accuracy than the naive one.

If $U_1 > 1 \rightarrow$ the forecast to compare has a lower degree of accuracy than the naive one.

Table 1

The accuracy of forecasts made by the three experts for the unemployment rate in Romania (2001–2011)

ACCURACY MEASURE	EXPERT		
	E1	E2	E3
ME	-0.5455	-0.5636	-0.7273
MAE	1.2364	1.6364	1.0909
RMSE	1.4948	1.7633	1.3052
U_1	0.1066	0.1240	0.0920
U_2	1.1575	1.0966	0.9977

Source: own computations

According to all accuracy indicators for forecasts made for 2001–2011, except for the mean error, E3 provided the most accurate predictions for the unemployment rate. Only the forecasts of this expert outperformed the naive predictions based on the random walk. The negative values of the mean error imply too high average predicted values for all the experts. The least accurate forecasts are made by E2.

Ranking method application supposes several steps:

1. Ranks are assigned to each value of an accuracy indicator (the value that indicates the best accuracy receives the rank 1);

The statistical units are the three experts that made forecasts. The rank for each expert is denoted by: $(r_{i_{ind_j}})$, $i=1,2,3$ and ind_j –accuracy indicator j .

We chose 5 indicators: mean error, mean absolute error, root mean squared error, U_1 and U_2 .

2. If the ranks assigned to each expert are added up, the score for each of them is computed.

$$S_i = \sum_{j=1}^5 (r_{i \text{ind}_j}), i=1,2,3.$$

3. The expert with the lowest score has the highest performance and it will get the final rank 1.

Table 2

The ranks of experts according to the accuracy measures (ranking method)

ACCURACY MEASURE	EXPERT		
	E1	E2	E3
ME	1	2	3
MAE	2	3	1
RMSE	2	3	1
U ₁	2	3	1
U ₂	3	2	1
Sum of ranks	10	13	7
Final ranks	2	3	1

Source: own computations

The results of the ranking method are the same as those provided by most accuracy measures, especially U₁ used in making comparisons between forecasts. Actually, if all the calculated accuracy indicators are taken into account at the same time, the following hierarchy was obtained: E3, E1 and E2.

The method of relative distance with respect to the maximal performance is the second way of ranking.

For each accuracy indicator, the distance of each statistical unit (expert) with respect to the one with the best performance is computed. The distance is calculated as a relative indicator of coordination:

$$d_{i \text{ind}_j} = \frac{\text{ind}_i^j}{\{\min \text{abs}(\text{ind}_i^j)\}_{i=1, \dots, 4}}, i=1,2,3 \text{ and } j=1,2, \dots, 5.$$

The relative distance computed for each expert is a ratio where the denominator is the best value for the accuracy indicator for all experts.

The geometric mean for the distances of each expert is calculated, its significance being the average relative distance for expert i .

$$\bar{d}_i = \sqrt[5]{\prod_{j=1}^5 d_{i_{ind_j}}}, \quad i=1,2,3.$$

The final ranks are assigned according to the values of average relative distances. The expert with the lowest average relative distance will take the rank 1. The position (location) of each expert with respect to the one with the best performance is computed as its average relative distance over the lowest average relative distance.

$$loc_i\% = \frac{\bar{d}_i}{\min(d_i)_{i=1,\dots,4}} \cdot 100$$

Table 3

The ranking of experts according to the accuracy measures (method of relative distance with respect to the best expert)

ACCURACY MEASURE	E1	E2	E3
ME	1	1.0332	1.3333
MAE	1.1334	1.5000	1
RMSE	1.1453	1.3510	1
U ₁	1.1587	1.3478	1
U ₂	1.1602	1.0991	1
Average relative distance	1.1178	1.2541	1.0592
Ranks	2	3	1
Location (%)	105.5286	118.3964	100

Source: own computations

The method of relative distance with respect to the best expert gave the same results as the previous methods. The lowest average relative distance was registered by E3 (1.0592).

The Diebold-Mariano test (DM test) is utilized to check if two forecasts have the same accuracy. The following steps are applied:

- The difference between the squared errors of forecasts (e^2) to compare and the squared errors of reference forecasts(e^{*2}) is calculated:

$$d_{t,t} = (e_{t,t}^2) - (e_{t,t}^{*2})$$

- The following model is estimated: $d_{t,t} = a + \varepsilon_t$
- We test if a differs from zero, where the null hypothesis is that $a=0$ (equal forecasts). A p-value less than 0.05 implies the rejection of the null hypothesis for a probability of 95% in guaranteeing the results.

The following variables are computed: d_1 , d_2 and d_3 to make comparisons between E1 and E2 forecasts, E1 and E3 predictions, and respectively E2

and E3 expectations. All the parameters are zero from the statistical point of view, so there are no significant differences between the forecasts provided by the three experts in terms of accuracy. The regression models are estimated in EViews and the results are presented in *Appendix 1*. Thus, the accuracy test showed that there are no significant differences between the forecasts provided by the three experts. If we take into account the results based on accuracy indicators and those of the DM test, we conclude that the best predictions are those of E3, followed by E1 and E2, but the differences between the unemployment rate forecasts are not too big.

By applying qualitative tests for directional accuracy, we check if there is a correct prediction of the change. A test of independence between the effective values and the direction of change can be applied in this situation, the null hypothesis showing the independence. A probability of less than 0.05 implies the rejection of the null hypothesis. All the asymptotic significances are greater than 0.05, according to *Appendix 5*, a fact that makes us conclude that the directional changes in the outcome are independent from the predictions.

Bias in this context implies a zero mean forecast error series. In the literature, rationality tests are used to check if the forecasts are optimal in relation to a certain criterion, if they are biased or ensure a good informational efficiency. The standard test of forecast bias, the Mincer-Zarnowitz test, starts from this model: $A_t = a + b \cdot P_t + e_t$, A_t – current values, P_t – predicted values.

Holden et al. (1990) proposed a modified version of the test, which is based on forecast errors, by testing whether their mean (m) is zero:

$$A_t - P_t = m + e_t.$$

The unbiasedness of the forecasts is tested applying a simple t-test for the following regression:

$$e_{t+1} = a + \varepsilon_{t+1}.$$

We have to test if the parameter a differs or not significantly from zero.

A p-value or probability of less than 0.05 for t test implies the existence of biasedness for those forecasts. The values of probability computed in EViews show that the largest biasedness is provided by the E2 forecasts while the lowest one by E3. The errors for each expert are denoted by e1, e2 and e3 and the tests results are presented in *Appendix 2*.

The efficiency of the forecasts is tested using an F test for the following regression model:

$$y_{t+1} = a + b \cdot y_{t+1,t} + \varepsilon_{t+1}$$

y_{t+1} – the value of the indicator registered for year $(t+1)$

$y_{t+1,t}$ – the forecast of the indicator made at moment t for the period $(t+1)$

E3 provided the best predictions of unemployment rate in terms of efficiency, followed by the E2 and E1.

Fair and Schiller (1989) propose a test in order to compare the efficiency of two forecasts made by two different experts for the same variable. To this end, they consider a simple linear regression model:

$$X_t - X_{t-1} = b_0 + b_1(X_1^p - X_{t-1}) + b_2(X_2^p - X_{t-1})$$

X_t – the value recorded for variable X at time t

X_{t-1} – the value recorded at time $t-1$ for variable X

X_1^p – the predicted value of the first expert

X_2^p – the predicted value of the second expert.

If $b_1 > 0$ and $b_2 = 0$, the second expert provides a relatively inefficient forecast, and the first expert forecast contains, in addition to the information from the first one, an essential piece of information about changes that may occur in the analyzed variable.

If $b_2 > 0$ and $b_1 = 0$, the first expert provides a relatively inefficient forecast.

If both parameters of the regression model are strictly positive, then each expert brings different information through the forecast.

E1 provided relatively efficient forecasts in respect to those of E2, while E2 predictions are “more efficient” than those of E3. On the other hand, E1 and E3 brought different information in their forecasts. In **Appendix 3** the results of the efficiency tests can be viewed.

Another test of weak efficiency regresses the error on a constant term and a lagged forecast error, according to Melander, Sismanidis and Grenoulleau (2007):

$$e_{t+1} = a + b \cdot e_t + \varepsilon_{t+1}$$

The null hypothesis acceptance implies a weak efficiency (bias and/or serial correlation) for a probability higher than 0.05. The results of this test of weak efficiency presented in **Appendix 3** imply the acceptance of a null hypothesis.

The Ljung-Box test is applied to check the persistence of errors. The null hypothesis of the test states the absence of errors autocorrelation. If the probability is less than 0.05, we conclude with a probability of 95% that there is errors autocorrelation.

The errors of IEF are autocorrelated only for a lag equal to one, according to the results presented in *Appendix 4*. There is not any autocorrelation for E2 predictions, while for a lag up to 2, the E1 forecasts are correlated.

The informational efficiency tests applied are stronger than those for testing weak efficiency. Actually, these tests check if the past information was fully utilized. Therefore, we check the dependencies between the forecasts errors and the key variable predictions, or respectively, the past values of a certain variable. The equations models could have the following forms:

$$e_{t+1,t} = a + b \cdot var_{t+1,t} + \varepsilon_{t+1},$$

where var is the forecasted variable and $e_{t+1,t} = a + b \cdot var_{t-1} + \varepsilon_{t+1}$.

The null hypothesis for both tests is: $b=0$.

The parameter differs significantly from zero for E1 and E2 forecasts if the first test is applied. The parameter is zero for E2 forecasts. So, only the E1 and E3 provided efficient predictions from an informational point of view, according to the results presented in *Appendix 4*. The second efficiency test gave other results: E2 and E1 forecasts are efficient, but not the E3 ones.

The three performance criteria are used to determine the best expert using multi-criteria ranking.

Table 4

The ranking of experts according to the performance criteria (ranking method)

Performance criteria	EXPERT		
	E1	E2	E3
Accuracy	2	3	1
Biasness	2	3	1
Informational efficiency	1	3	2
Sum of ranks	5	9	4
Final ranks	2	3	1

Source: own computations using Excel

According to the ranking method, the E3 provided the forecasts with the highest performance and the E2 the worse predictions in terms of performance.

3. COMBINED FORECASTS – A STRATEGY TO IMPROVE THE UNEMPLOYMENT RATE FORECASTS PERFORMANCE

Combined forecasts are a possible strategy of getting more accurate predictions.

Supposing we have a vector of n forecasts, f , that are combined and a linear weighting vector, w , the combined prediction is: $f_c = w' \cdot f$.

De Menezes, Bunn and Taylor (2000) presented seven methods of combining forecasts: simple average (chosen by many researchers), outperformance, optimal, regression, regression with restricted weights, optimal (adaptive) with independence assumption and optimal (adaptive) with restricted weights.

Newbold and Granger (1974) concluded that the assumption of independence between individual forecasts errors performs better than the correlation hypothesis. Schnaars (1986) compared the MSE of seven extrapolation models and three combinations of models for many forecasts and he pointed out that equal weights were the best and combinations generally outperformed the individual models.

De Menezes, Bunn and Taylor (2000) showed that some researchers claimed the superiority of regression methods for combining forecasts (Guerard and Holmen), but others recommended the optimal approach as the best choice (Mills and Stephenson, Clemens, Holden and Peel and Lobo).

Holden et al. (1990) recommended the inclusion of a constant in the regression and restricted the weights on the predictions to add up to one.

In practice two ways are used to generate new forecasts: using different methods or using different data. If there are several sources of data, an average of the extrapolated values can be carried to obtain the combined forecasts. Batchelor and Dua (1995) observed the decrease of the MSE indicator by combining forecasts based on different sources of data and on assumptions, or based on several methods.

Clemens (1989) concluded that equal weighting is the best choice for many types of predictions, especially when experts have to predict changes. When the accuracy is measured by MAE, McNeas (1992) and other

researchers recommend the use of median, while the mean is preferable when RMSE is used as the accuracy indicator.

Klugman (1945) checked that combined judgments are recommended for heterogeneous data more than homogenous ones. Clemens and Winkler (1986) studied the accuracy of GNP forecasts based on four econometric models. The equally weighted combined forecast was better in terms of accuracy than the initial predictions.

Fildes (1991) observed an improvement in accuracy when different types of predictions are equally weighting (experts' forecasts, naïve extrapolation and prediction based on an econometric model).

Armstrong (2001) showed that combined forecasts are useful when there is uncertainty regarding the forecasting context, the selection of the best method or the cost for large errors is high.

The researchers consider that the optimal combination is the one based on the covariance matrix of the prediction errors. Timmermann (2006) took into account only the diagonal elements of this matrix because of the huge number of coefficients to be estimated.

Another problem in combining forecasts is related to the accuracy measure, most of the researchers choosing the mean squared error (MSE). Therefore, the good models that generated predictions with a lower MSE receive a better weight than the ones with less accurate forecasts.

If there are large differences between the prediction error variance of two models, a forecasts combination will not improve too much the degree of accuracy. Stock and Watson (1999) introduced the PLS (predicted least squares) method to build the combined forecasts. This technique consists in giving a weight of one for the best model up to the prediction date and zero for the other ones.

Granger and Jeon (2004) propose the use of models with a close degree of accuracy. Aiolfi and Timmermann (2006) clustered the models into groups according to the MSE indicator and selected the best models for the combined forecast.

When shifts are presented in the economy, it is important to construct an adaptable weights scheme. Granger and Ramanathan (1984) used ARCH effects to predict forecast error variance and then to build the weights. Aiolfi and Timmermann (2006) recommended Markov switching models and time varying parameter models to determine the changing weights.

Clark and McCracken (2009) showed that the combined forecasts of recursive predictions and rolling ones often improved the accuracy by minimizing the mean squared forecast error.

Wallis (2011) analyzed the effects of experts' data on the combinations of point forecasts and the properties of strategies of combining predictions. For density forecasts, the logarithmic method provided better results than the simple linear one.

Genrea, Kenny, Meylera and Timmermann (2013) made forecasts combinations starting with SPF predictions for ECB and using performance-based weighting, trimmed averages, principal components analysis, Bayesian shrinkage, and least squares estimates of optimal weights. Only for the inflation rate was there a strong evidence of improving the forecasts accuracy with respect to the equally weighted average prediction.

Bratu (2012) utilized some strategies to improve the forecasts accuracy (combined predictions, regressions models, historical errors method, application of filters and exponential smoothing techniques).

The most utilized combination approaches are:

- optimal combination (OPT);
- equal-weights-scheme (EW);
- inverse MSE weighting scheme (INV).

Bates and Granger (1969) started from two forecasts $f_{1,t}$ and $f_{2,t}$, for the same variable X_t , derived h periods ago. If the forecasts are unbiased, the error is calculated as: $e_{i,t} = X_{i,t} - f_{i,t}$. The errors follow a normal distribution of parameters 0 and σ_i^2 . 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: $c_t = m \cdot f_{1,t} + (1-m) \cdot f_{2,t}$. The error of the combined forecast is: $e_{c,t} = m \cdot e_{1,t} + (1-m) \cdot e_{2,t}$. The mean of the combined forecast is zero and the variance is: $\sigma_c^2 = m^2 \cdot \sigma_1^2 + (1-m)^2 \cdot \sigma_2^2 + 2 \cdot m \cdot (1-m) \cdot \sigma_{12}$. By minimizing the error variance, the optimal value for m is determined (m_{opt}):

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

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 predictions (EW) are obtained when the same weights are given to all models.

We proposed a new variant of combining forecasts, starting from the empirical results that inform us that E3 forecasts are the best. It is rational to give a higher weight to the best predictions. Therefore an equation is used, the coefficients being determined on empirical bases:

$$\begin{aligned} \text{Combined prediction} = & [1 - |ME(E1_f)|] \cdot E1_f + \dots \\ & + [1 - |ME(E2_f)|] \cdot E2_f + [1 - |ME(E3_f)|] \cdot E3_f \end{aligned}$$

where:

$E1_f$ – $E1$ forecasts,

$E2_f$ – $E2$ forecasts,

$E3_f$ – $E3$ forecasts.

This procedure is based on the mean error of the forecasts. The difference between 1 and the corresponding ME is the weight. U Theil's statistics were computed for the combined forecasts based on the three schemes and on our proposed variant, the results are shown in the following table (Table 5):

Table 5

The accuracy of combined forecasts for unemployment rate (2001–2011)

Accuracy indicator	E1+E2 forecasts	E1+E3 forecasts	E2+E3 forecasts
U ₁ (optimal scheme)	0.0846	0.0666	0.1254
U ₂ (optimal scheme)	0.9867	0.7130	1.1063
U ₁ (inverse MSE scheme)	0.0864	0.0553	0.1105
U ₂ (inverse MSE scheme)	1.0026	0.5888	1.0116
U ₁ (equally weighted scheme)	0.0861	0.0739	0.0888
U ₂ (equally weighted scheme)	0.9207	0.7933	0.9134
U ₁ (proposed method)	0.0835	-	-
U ₂ (equally weighted scheme)	0.08985	-	-
U ₁ (proposed method)	0.0872	0.0946	0.9459
U ₂ (mean of the forecasts)	0.974	0.784	0.8795

Source: author's computations

Our proposed combination method gave the best results, outperforming the classical schemes. Quite good results are provided by the mean of the forecasts offered by the three experts.

The combined forecasts proved to be a good strategy of improving the accuracy when the E1 and E2 forecasts, and respectively, E1 and E3 predictions are combined using OPT and INV schemes. Only if equally weighted scheme is utilized we obtain better forecasts for the combined predictions of E2 and E3. The most accurate forecasts are those resulting from combining the E1 and E3 expectations. All the combined predictions are better than the naïve ones, excepting those of E2 and E3 using the OPT scheme.

We test the biasedness of the combined forecasts. Only the combined forecasts based on the E1 and E3 expectations are biased, all the other predictions being unbiased. So, the combined forecasts are a very good strategy of getting unbiased forecasts.

Each combined forecast based on the INV scheme provided different information if we make comparisons of two forecasts from this group. The combined forecasts of the E1 and E3 and those of the E2 and E3 are relative efficient in respect to the combined predictions of the E1 and E2. These efficient combined forecasts have a better performance than the original ones of the experts concerning efficiency. The predictions based on our combination scheme are informationally efficient.

CONCLUSIONS

In addition to economic analysis, the elaboration of forecasts is an essential aspect that conducts the way of developing the activity at 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, the adjustment of government policies, the planning of results. Basically, performance evaluation in this context refers directly to the degree of trust conferred 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. This aspect is important, because the macroeconomic predictions must not be easily accepted, taking into account the negative consequences of macroeconomic

forecasts failures, and the consequences that affect state policies. The decisions on economic policy are based on these forecasts. Hence, there is evident interest in improving their performance.

In our study we assessed the unemployment forecasts performance for the predictions provided during 2001–2011 by three experts: E1, E2 and E3. The better performance in predicting the unemployment rate of Romania was achieved by E3, followed by E1 and E2. The best accuracy was provided by E3, followed by E1 and E2. This hierarchy resulted from the application of the multi-criteria ranking, but also from the measurement of accuracy indicators, as U_1 , used in making comparisons between forecasts. All the predictions are biased, according to international trends. The lowest bias was registered for E3 expectations, while the largest bias was observed for E2. The E1 provided the forecasts with the highest efficiency, E3 offered the least efficient ones.

Our proposed combination procedure succeeded in providing the best performance improvements. In this case we took into consideration the mean errors of the forecasts, the weights being computed as a complementary measure of ME. The combined forecasts using the three classical schemes are also a good strategy for improving accuracy. These predictions are unbiased, except for the combined forecasts of E1 and E3, which have a higher degree of efficiency.

The forecasts performance should be a priority for the public authorities that use these predictions underlying the decisional process. The combined forecasts are a very good strategy for achieving improvements in performance for predicting the unemployment rate.

REFERENCES

- Aiolfi, M., Timmermann, A., *Structural Breaks and the Performance of Forecast Combinations*. Bocconi University, Mimeo, 2004.
- Allan, G., *Evaluating the Usefulness of Forecasts of Relative Growth*, “Strathclyde, Discussion Papers in Economics”, No. 12-14, 2012.
- Armstrong, J. S., Fildes, R., *On the Selection of Error Measures for Comparisons Among Forecasting Methods*, “Journal of Forecasting”, 14, pp. 67-71, 1995.
- Armstrong, J. S., *Principles of Forecasting: A Handbook for Researchers and Practitioners*. Kluwer Academic Publishers, Norwell, MA, 2001.
- Barnichon, A., Nekarda, C., *The Ins and Outs of Forecasting Unemployment*, “Brookings Papers on Economic Activity”, Fall 2012.
- Batchelor, R., Dua, P., *Forecaster Diversity and the Benefits of Combining Forecasts*, “Management Science”, 41, pp. 68-75, 1995.

- Bates, J., Granger, C. W. J., *The Combination of Forecasts*, "Operations Research Quarterly", 20(4), pp. 451-468, 1969.
- Bratu, M., *Strategies to Improve the Accuracy of Macroeconomic Forecasts in the USA*, LAP LAMBERT Academic Publishing, 2012.
- Clark, T., McCracken, M., *Improving Forecast Accuracy by Combining Recursive and Rolling Forecasts*, "International Economic Review", 2009, Vol. 50, Issue 2, pp. 363-395, 2009.
- Clemens, R. T., *Combining Forecasts: A Review and Annotated Bibliography*, "International Journal of Forecasting", 5, pp. 559-583, 1989.
- Clemens, R. T., Winkler, R. L., *Combining Economic Forecasts*, "Journal of Business and Economic Statistics", 4, pp. 39-46, 1986.
- De Menezes, L. M., Bunn, D. W., Taylor, J. W., *Review of Guidelines for the Use of Combined Forecasts*, "European Journal of Operational Research", 2000, Vol. 120, pp. 190-204, 2000.
- Fildes, R., *Efficient Use of Information in the Formation of Subjective Industry Forecasts*, "Journal of Forecasting", 10, pp. 597-617, 1991.
- Franses, P. H., McAleer, M., Legerstee, R., *Evaluating Macroeconomic Forecasts: A Concise Review of Some Recent Developments*, Working paper, Department of Economics and Finance, University of Canterbury, 2012.
- Genrea, V., Kenny, G., Meylera, A., Timmermann, A., *Combining Expert Forecasts: Can Anything Beat the Simple Average?*, "International Journal of Forecasting", Vol. 29, Issue 1, January–March 2013, pp. 108-121, 2013.
- Granger, C., Jeon, Y., *Thick Modelling*, "Economic Modelling", 21(2), pp. 323-343, 2004.
- Granger, C. W. J., Ramanathan, R., *Improved Methods for Combining Forecasts*, "Journal of Forecasting", 3(2), pp. 197-204, 1984.
- Heilemann, U., Stekler, H., *Introduction to "The Future of Macroeconomic Forecasting"*, "International Journal of Forecasting", 23(2), pp. 159-165, 2007.
- Holden, K., Peel, D. A., Thomson, J. L., *Economic Forecasting: An Introduction*. Cambridge University Press, Cambridge, 1990.
- Klugman, S. F., *Group Judgments for Familiar and Unfamiliar Materials*, "Journal of General Psychology", 32, pp. 103-110, 1945.
- Kurita, T., *A Forecasting Model for Japan's Unemployment Rate*, "Eurasian Journal of Business and Economics", No. 3(5), pp. 127-134, 2010.
- McNees, S. K., *The Uses and Abuses of Consensus' Forecasts*, "Journal of Forecasting", 11, pp. 703-710, 1992.
- Melander, A., Sismanidis, G., Grenoulleau, D., *The Track Record of the Commission's Forecasts—An Update*, "European Economy, Economic Papers", no. 29, 2007.
- Newbold, P., Granger, C.W. J., *Experience with Forecasting Univariate Time Series and the Combination of Forecasts (with Discussion)*, "Journal of the Royal Statistical Society", Series A137, pp. 131-149, 1974.
- Schnaars, S. P., *An Evaluation of Rules for Selecting an Extrapolation Model on Yearly Sales Forecasts*, "Interfaces", 16, pp. 100-107, 1986.

- Stock, J., Watson, M., *A Comparison of Linear and Nonlinear Univariate Models for Forecasting Macroeconomic Time Series* [in:] Engle R., White H., *Cointegration, Causality and Forecasting: A Festschrift in Honour of Clive W. J. Granger*. Chapter 1, pp. 1-44, 1999.
- Timmermann, A., *Forecast Combinations* [in:] Elliott, G., Granger, C., Timmermann, A., *Handbook of Economic Forecasting*. Chapter 4, pp. 135-196, Elsevier, 2006.
- Wallis, K.F., *Combining Forecasts –Forty Years Later*, “Applied Financial Economics”, Vol. 21, Issue 1-2, pp. 33-41, 2011.

Received: November 2012, revised: May 2013

APPENDIX 1

The results of Diebold-Mariano test in EViews

Dependent Variable: D1

Sample: 2001 2011

Included observations: 11

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.874545	1.187738	-0.736312	0.4785
R-squared	0.000000	Mean dependent var		-0.874545
Adjusted R-squared	0.000000	S.D. dependent var		3.939283
S.E. of regression	3.939283	Akaike info criterion		5.666382
Sum squared resid	155.1795	Schwarz criterion		5.702555
Log likelihood	-30.16510	Durbin-Watson stat		1.518619

Dependent Variable: D2

Sample: 2001 2011

Included observations: 11

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.530909	0.624816	0.849704	0.4154
R-squared	0.000000	Mean dependent var		0.530909
Adjusted R-squared	0.000000	S.D. dependent var		2.072281
S.E. of regression	2.072281	Akaike info criterion		4.381685
Sum squared resid	42.94349	Schwarz criterion		4.417857
Log likelihood	-23.09927	Durbin-Watson stat		1.521367

Dependent Variable: D3

Sample: 2001 2011

Included observations: 11

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	1.405455	0.886219	1.585900	0.1438
R-squared	0.000000	Mean dependent var		1.405455
Adjusted R-squared	0.000000	S.D. dependent var		2.939256
S.E. of regression	2.939256	Akaike info criterion		5.080698
Sum squared resid	86.39227	Schwarz criterion		5.116871
Log likelihood	-26.94384	Durbin-Watson stat		1.686150

APPENDIX 2

The results of bias tests

Dependent Variable: e1

Sample: 2001 2011

Included observations: 11

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.545455	0.440116	-1.239341	0.2435
R-squared	0.000000	Mean dependent var		-0.545455
Adjusted R-squared	0.000000	S.D. dependent var		1.459701
S.E. of regression	1.459701	Akaike info criterion		3.680848
Sum squared resid	21.30727	Schwarz criterion		3.717021
Log likelihood	-19.24467	Durbin-Watson stat		0.719003

Dependent Variable: e2

Sample: 2001 2011

Included observations: 11

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.563636	0.528337	-1.066811	0.3111
R-squared	0.000000	Mean dependent var		-0.563636
Adjusted R-squared	0.000000	S.D. dependent var		1.752297
S.E. of regression	1.752297	Akaike info criterion		4.046240
Sum squared resid	30.70545	Schwarz criterion		4.082413
Log likelihood	-21.25432	Durbin-Watson stat		1.144096

Dependent Variable: e3

Sample: 2001 2011

Included observations: 11

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.727273	0.342741	-2.121934	0.0598
R-squared	0.000000	Mean dependent var		-0.727273
Adjusted R-squared	0.000000	S.D. dependent var		1.136742
S.E. of regression	1.136742	Akaike info criterion		3.180717
Sum squared resid	12.92182	Schwarz criterion		3.216889
Log likelihood	-16.49394	Durbin-Watson stat		0.876038

APPENDIX 3

The results of efficient tests

Dependent Variable: UR
 Sample: 2001 2011
 Included observations: 11

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	1.243713	3.121967	0.398375	0.6996
E1	0.749926	0.431692	1.737179	0.1164
R-squared	0.251110	Mean dependent var		6.609091
Adjusted R-squared	0.167900	S.D. dependent var		1.656173
S.E. of regression	1.510753	Akaike info criterion		3.826059
Sum squared resid	20.54136	Schwarz criterion		3.898403
Log likelihood	-19.04332	F-statistic		3.017791
Durbin-Watson stat	0.668732	Prob(F-statistic)		0.116366

Dependent Variable: UR
 Sample: 2001 2011
 Included observations: 11

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	3.306537	1.662447	1.988958	0.0779
E2	0.460432	0.223734	2.057946	0.0697
R-squared	0.319992	Mean dependent var		6.609091
Adjusted R-squared	0.244436	S.D. dependent var		1.656173
S.E. of regression	1.439599	Akaike info criterion		3.729571
Sum squared resid	18.65200	Schwarz criterion		3.801916
Log likelihood	-18.51264	F-statistic		4.235140
Durbin-Watson stat	1.051699	Prob(F-statistic)		0.069712

Dependent Variable: UR
 Sample: 2001 2011
 Included observations: 11

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-3.315701	2.908748	-1.139907	0.2838
E3	1.352822	0.393666	3.436471	0.0074
R-squared	0.567502	Mean dependent var		6.609091
Adjusted R-squared	0.519446	S.D. dependent var		1.656173
S.E. of regression	1.148092	Akaike info criterion		3.277045
Sum squared resid	11.86303	Schwarz criterion		3.349390
Log likelihood	-16.02375	F-statistic		11.80934
Durbin-Watson stat	1.299172	Prob(F-statistic)		0.007432

Dependent Variable: DIFFERENCE
 Sample(adjusted): 2002 2011
 Included observations: 10 after adjusting endpoints

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.522476	0.311075	-1.679582	0.1369
A1	0.494477	0.163304	3.027959	0.0192
A2	0.187598	0.156957	1.195222	0.2709
R-squared	0.687461	Mean dependent var		-0.110000
Adjusted R-squared	0.598165	S.D. dependent var		1.458652
S.E. of regression	0.924647	Akaike info criterion		2.924515
Sum squared resid	5.984803	Schwarz criterion		3.015291
Log likelihood	-11.62258	F-statistic		7.698614
Durbin-Watson stat	1.685137	Prob(F-statistic)		0.017067

Dependent Variable: DIFFERENCE
 Sample(adjusted): 2002 2011
 Included observations: 10 after adjusting endpoints

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.544827	0.337852	-1.612618	0.1509
A1	0.332518	0.396650	0.838315	0.4295
A3	0.285263	0.440949	0.646930	0.5383
R-squared	0.644909	Mean dependent var		-0.110000
Adjusted R-squared	0.543454	S.D. dependent var		1.458652
S.E. of regression	0.985585	Akaike info criterion		3.052161
Sum squared resid	6.799638	Schwarz criterion		3.142937
Log likelihood	-12.26081	F-statistic		6.356627
Durbin-Watson stat	1.438761	Prob(F-statistic)		0.026680

Dependent Variable: DIFFERENCE
 Sample(adjusted): 2002 2011
 Included observations: 10 after adjusting endpoints

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.550912	0.352550	-1.562651	0.1621
A2	0.061708	0.202908	0.304115	0.7699
A3	0.579810	0.234691	2.470519	0.0428
R-squared	0.614355	Mean dependent var		-0.110000
Adjusted R-squared	0.504170	S.D. dependent var		1.458652
S.E. of regression	1.027113	Akaike info criterion		3.134706
Sum squared resid	7.384726	Schwarz criterion		3.225481
Log likelihood	-12.67353	F-statistic		5.575693
Durbin-Watson stat	1.217360	Prob(F-statistic)		0.035617

Weak efficiency tests

Dependent Variable: e1
 Sample(adjusted): 2002 2011
 Included observations: 10 after adjusting endpoints

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.552846	0.355320	-1.555913	0.1583
e1(-1)	0.494307	0.231391	2.136240	0.0652
R-squared	0.363236	Mean dependent var		-0.800000
Adjusted R-squared	0.283640	S.D. dependent var		1.255211
S.E. of regression	1.062386	Akaike info criterion		3.135769
Sum squared resid	9.029317	Schwarz criterion		3.196286
Log likelihood	-13.67884	F-statistic		4.563520
Durbin-Watson stat	2.456784	Prob(F-statistic)		0.065157

Dependent Variable: e2
 Sample(adjusted): 2002 2011
 Included observations: 10 after adjusting endpoints

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.141481	0.657457	-0.215195	0.8350
e2(-1)	0.414903	0.383698	1.081328	0.3111
R-squared	0.127520	Mean dependent var		-0.490000
Adjusted R-squared	0.018461	S.D. dependent var		1.829056
S.E. of regression	1.812094	Akaike info criterion		4.203700
Sum squared resid	26.26949	Schwarz criterion		4.264217
Log likelihood	-19.01850	F-statistic		1.169270
Durbin-Watson stat	1.660361	Prob(F-statistic)		0.311067

Dependent Variable: e3
 Sample(adjusted): 2002 2011
 Included observations: 10 after adjusting endpoints

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.399439	0.412785	-0.967669	0.3615
e3(-1)	0.518748	0.302262	1.716219	0.1245
R-squared	0.269100	Mean dependent var		-0.830000
Adjusted R-squared	0.177737	S.D. dependent var		1.143144
S.E. of regression	1.036588	Akaike info criterion		3.086602
Sum squared resid	8.596116	Schwarz criterion		3.147119
Log likelihood	-13.43301	F-statistic		2.945408
Durbin-Watson stat	1.996888	Prob(F-statistic)		0.124457

Information efficiency tests

Dependent Variable: e1
 Sample: 2001 2011
 Included observations: 11

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	1.243713	3.121967	0.398375	0.6996
E1	-0.250074	0.431692	-0.579289	0.5766
R-squared	0.035946	Mean dependent var		-0.545455
Adjusted R-squared	-0.071171	S.D. dependent var		1.459701
S.E. of regression	1.510753	Akaike info criterion		3.826059
Sum squared resid	20.54136	Schwarz criterion		3.898403
Log likelihood	-19.04332	F-statistic		0.335576
Durbin-Watson stat	0.668732	Prob(F-statistic)		0.576604

Dependent Variable: e2
 Sample: 2001 2011
 Included observations: 11

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	3.306537	1.662447	1.988958	0.0779
E2	-0.539568	0.223734	-2.411650	0.0391
R-squared	0.392551	Mean dependent var		-0.563636
Adjusted R-squared	0.325057	S.D. dependent var		1.752297
S.E. of regression	1.439599	Akaike info criterion		3.729571
Sum squared resid	18.65200	Schwarz criterion		3.801916
Log likelihood	-18.51264	F-statistic		5.816058
Durbin-Watson stat	1.051699	Prob(F-statistic)		0.039143

Dependent Variable: e3
 Sample: 2001 2011
 Included observations: 11

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-3.315701	2.908748	-1.139907	0.2838
E3	0.352822	0.393666	0.896247	0.3935
R-squared	0.081938	Mean dependent var		-0.727273
Adjusted R-squared	-0.020069	S.D. dependent var		1.136742
S.E. of regression	1.148092	Akaike info criterion		3.277045
Sum squared resid	11.86303	Schwarz criterion		3.349390
Log likelihood	-16.02375	F-statistic		0.803258
Durbin-Watson stat	1.299172	Prob(F-statistic)		0.393462

Dependent Variable: e1
 Sample(adjusted): 2002 2011
 Included observations: 10 after adjusting endpoints

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-3.654271	2.766698	-1.320806	0.2231
E1(-1)	0.406591	0.390078	1.042332	0.3277
R-squared	0.119569	Mean dependent var		-0.800000
Adjusted R-squared	0.009515	S.D. dependent var		1.255211
S.E. of regression	1.249225	Akaike info criterion		3.459781
Sum squared resid	12.48451	Schwarz criterion		3.520298
Log likelihood	-15.29891	F-statistic		1.086457
Durbin-Watson stat	0.611476	Prob(F-statistic)		0.327730

Dependent Variable: e2
 Sample(adjusted): 2002 2011
 Included observations: 10 after adjusting endpoints

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.252195	2.394369	0.105328	0.9187
E2(-1)	-0.100842	0.314602	-0.320537	0.7568
R-squared	0.012680	Mean dependent var		-0.490000
Adjusted R-squared	-0.110735	S.D. dependent var		1.829056
S.E. of regression	1.927667	Akaike info criterion		4.327355
Sum squared resid	29.72721	Schwarz criterion		4.387872
Log likelihood	-19.63677	F-statistic		0.102744
Durbin-Watson stat	1.278182	Prob(F-statistic)		0.756774

Dependent Variable: e3
 Sample(adjusted): 2002 2011
 Included observations: 10 after adjusting endpoints

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-7.027908	2.151712	-3.266194	0.0114
E3(-1)	0.843253	0.290478	2.902987	0.0198
R-squared	0.513007	Mean dependent var		-0.830000
Adjusted R-squared	0.452133	S.D. dependent var		1.143144
S.E. of regression	0.846133	Akaike info criterion		2.680576
Sum squared resid	5.727526	Schwarz criterion		2.741093
Log likelihood	-11.40288	F-statistic		8.427336
Durbin-Watson stat	0.766689	Prob(F-statistic)		0.019801

APPENDIX 4

The results of the Ljung-Box test

Sample: 2001 2011

Included observations: 11

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob	
. ****.	. ****.	1	0.484	0.484	3.3444	0.067
. ***.	. * .	2	0.374	0.183	5.5713	0.062
. .	. ** .	3	0.042	-0.263	5.6024	0.133

Sample: 2001 2011

Included observations: 11

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob	
. ** .	. ** .	1	0.295	0.295	1.2424	0.265
. ** .	. *** .	2	-0.299	-0.423	2.6659	0.264
. * .	. ** .	3	-0.071	0.238	2.7572	0.431

Sample: 2001 2011

Included observations: 11

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob	
. ****.	. ****.	1	0.480	0.480	3.2990	0.069
. * .	. * .	2	0.134	-0.125	3.5863	0.166
. ** .	. *** .	3	-0.239	-0.332	4.6052	0.203

APPENDIX 5

The results of the tests for directional accuracy

Test Statistics

	ur	E1
Chi-Square	.818a	1.273b
Df	9	8
Asymp. Sig.	1.000	.996

Test Statistics

	ur	E2
Chi-Square	.818a	.000b
Df	9	10
Asymp. Sig.	1.000	1.000

Test Statistics

	ur	E3
Chi-Square	.818a	1.273b
Df	9	8
Asymp. Sig.	1.000	.996