

The Unemployment Rate Forecasts Evaluation Using New Aggregated Accuracy Indicators

Mihaela Simionescu¹

Abstract: In this study, the unemployment rate forecasts for Romania were assessed using the predictions provided on the horizon 2006-2013 by three experts in forecasting or forecasters (F1, F2 and F3). The absolute and relative accuracy indicators, excepting mean relative absolute error (MRAE) indicated that F3 forecasts are the most accurate on the mentioned horizon. The high value of this indicator brought differences in accuracy hierarchy. New aggregated accuracy indicators were proposed (modified sum of summary statistics- S1, sum of relative accuracy measures- S2 and sum of percentage for directional and sign accuracy- S3). The contradictory results of S1 and S2 were solved by the method of relative distance with respect to the best forecaster that indicated F2 forecasts for unemployment rate forecasts in Romania as the best. It is clearly that F3 outperformed the other experts as directional and sign accuracy. The Diebold-Mariano test identified F1 predictions as the less accurate, but significant accuracy differences were not found between F3 and F2 predictions.

Keywords: forecasts accuracy; forecast error; unemployment rate; Diebold-Mariano test; directional accuracy

JEL Classification: E37; E66

1 Introduction

In this study, the forecasts accuracy was assessed for unemployment rate predictions in Romania provided by three anonymous forecasters (F1, F2 and F3). The novelty of the research compared to previous studies is that new aggregated indicators (S1, S2 and S3) were proposed in order to solve the problem of contradictory results provided by different accuracy measures. However, for this particular case of unemployment rate predictions in Romania different results were obtained, but a multi-criteria ranking method was applied for S1 and S2 measures to select the best forecaster.

The paper is structured as it follows. After a brief literature review, the third section describes the methodological framework, while the forecasts accuracy assessment

¹ Senior Researcher, Institute for Economic Forecasting of the Romanian Academy, Bucharest, Romania, tel. 004021.318.81.48, Corresponding author: mihaela_mb1@yahoo.com.

for unemployment rate in Romania is presented in the fourth section. The last section gives a brief conclusion.

2. Literature Review

There are many international organizations that provide their economic predictions for various countries. The comparisons between forecasts consider these institutions anticipations (OECD, IMF, World Bank, European Commission, SPF etc.) and those of other international organizations, the accuracy assessment being made. The forecast errors for these institutions are in general large and non-systematic. Three international institutions (European Commission- EC, IMF and OECD) made predictions using macroeconomic models, but these forecasts failed to anticipate the downturn from 2007. Other providers of forecasts are statistical institutes, ministries of finance, and private companies like banks or insurance companies.

Literature usually makes comparisons between OECD and IMF forecasts and Consensus Economics ones or private predictions. The accuracy is evaluated according to different criteria: forecasts errors and associated accuracy measures, comparisons with naïve predictions that is based on random walk, directional accuracy evaluation.

For 25 transition countries the EBRD predictions during 1994-2004 improve in accuracy with the progress in transition. These predictions accuracy for late GDP is better than of other institutions with around 0.4 percentage points. The Russian crisis seems to be the only structural break (Krkoska & Teksoz, 2007).

The European Commission's forecasts analyzed on the horizon from 1998 to 2005 are comparable in terms of accuracy with those of Consensus, IMF and OECD for variables like inflation rate, unemployment rate, GDP, total investment, general government balance and current account balance (Melander, Sismanidis & Grenouilleau, 2007) stated.

The forecasts accuracy of the predictions provided by European Commission before and during the recent economic crisis was assessed (González Cabanillas & Terzi, 2012). They compared these forecasts with those provided by Consensus Economics, IMF and OECD. The Commission's forecasts errors have increased because of the low accuracy from 2009 for variables as GDP, inflation rate, government budget balance, and investment.

The strategic behavior of the private forecasters that placed their expectations away from OECD's and IMF's ones, was assessed by experts, this duration of this event being 3 months (Frenkel, Rülke & Zimmermann, 2013).

Greenbook inflation forecasts are more accurate than those of the private forecasts, making comparisons between the predictions provided by Survey of Professional Forecasters, Greenbook and other private forecasters (Liu & Smith, 2014).

The common approach to evaluate the predictions' usefulness consists in the measurement of the error's magnitude, using accuracy measures like mean square error (MSE) (Diebold and Mariano, 2002), or log of the mean squared error ratio (log MSER). However, these measures do not have an economic interpretation and they neglect the presence of outliers. The directional forecasts technique was used for assessing the macroeconomic forecasts by many other authors ((Pesaran & Timmermann, 1994), Artis, 1996), (Öller & Barot, 2000), (Pons, 2001) and (Ashiya, 2006).

3. Methodological Framework

There are different methods used in literature to assess the forecasts accuracy. In practice, there are many cases when some indicators suggest the superiority of certain forecasts while other ones indicate that other predictions are more accurate. Therefore, it is proposed a new methodology to solve this contradiction given by the results of accuracy assessment. The method is based on different types of accuracy measures: statistics based on size errors, coefficients for comparisons and directional accuracy measures. These types of indicators were also used in literature without any aggregation (Melander, Sismanidis & Grenouilleau, 2007).

The prediction error at time t is the simplest indicator based on the comparison of the registered value with the forecasted one and it is denoted by e_t . There are two ways of computing the forecast error if \hat{y}_t is the prediction at time t : $e_t = y_t - \hat{y}_t$ or $e_t = \hat{y}_t - y_t$. Seven out of eleven members from International Institute of Forecasters recommended in a survey the use of the first variant ($e_t = y_t - \hat{y}_t$). This is the most utilized version in literature and it will also be used in this study (Green & Tashman, 2008).

The following summary statistics have been used: root mean squared error, mean squared error, mean error, mean absolute error, mean absolute percentage error. If the horizon length is h and the length of actual data series is n , the indicators are computed as in the following table:

Table 1. Summary statistics for forecasts accuracy

Indicator	Formula
Mean error- ME	$ME = \frac{1}{h} \sum_{t=n+1}^{n+h} (y_t - \hat{y}_t)$
Mean absolute error- MAE	$ME = \frac{1}{h} \sum_{t=n+1}^{n+h} y_t - \hat{y}_t $
Root mean squared error- RMSE	$RMSE = \sqrt{\frac{1}{h} \sum_{t=n+1}^{n+h} (y_t - \hat{y}_t)^2}$
Mean squared error- MSE	$MSE = \frac{1}{h} \sum_{t=n+1}^{n+h} (y_t - \hat{y}_t)^2$
Mean absolute percentage error- MAPE	$MAPE = 100 \cdot \frac{1}{h} \sum_{t=n+1}^{n+h} \left \frac{y_t - \hat{y}_t}{y_t} \right $

The aggregate statistic for comparisons is based on UITheil's statistic, mean relative absolute error, relative RMSE and mean absolute scaled error. $RMSE_b$ is the RMSE for the benchmark. e_t^* is the benchmark error. In our case the benchmark is represented by the naïve projection.

Table 2. Statistics for comparing the forecasts accuracy

Indicator	Formula
UI Theil's statistic	$U_1 = \frac{\sqrt{\sum_{t=n+1}^{n+h} (y_t - \hat{y}_t)^2}}{\sqrt{y_t^2} + \sqrt{\hat{y}_t^2}}$
Mean relative absolute error- MRAE	$MRAE = average\left(\frac{ e_t }{ e_t^* }\right)$
Relative Root mean squared error- RRMSE	$RRMSE = \frac{RMSE}{RMSE_b}$
Mean absolute scaled error-MASE	$MASE = average\left(\frac{e_t}{\frac{1}{n-1} \sum_{t=n+1}^{n+h} y_t - y_{t-1} }\right)$

If ME takes a positive value on the mentioned horizon with the proposed definition of the forecast error, the predictions are underestimated. For negative value of ME

the forecasts are overestimated. For optimal predictions ME is zero, but this value is also met when the errors offset each other perfectly.

MSE penalizes the predictions with high errors. It considers that the high errors are more harmful than the small errors. The positive and the negative errors cannot compensate each other like in the case of ME, which is an advantage for MSE. There is not a superior limit for MSE and it has a different unit of measurement compared to actual data. The null value is the lowest value of the indicator and it is achieved for perfect precision of the forecasts. RMSE is equal or larger than MAE. A higher difference between these two indicators implies a higher errors variance. The errors have the same magnitude if RMSE equals MAE. The minimum value of those measures is 0, but there is not a superior limit for them. A null value for the MAPE expressed as percentage shows a perfect forecast. If MAPE is smaller than 100% the prediction is better than the naïve one. MAPE has no superior limit.

The percentage of sign correct forecasts (PSC) shows how many percent of time is sign of prediction forecasted correctly. Percentage of directional accuracy correct forecasts (PDA) shows if the expert correctly anticipates the increase or decrease of the variable. It measures the ability to correctly predict the turning points. PDA and PSC are located between 0% and 100%. According to Melander et al. (2007) the success rate of the indicators should be greater than 50%.

Table 3. Measures for directional and sign accuracy

Indicator	Formula	Conditions
Percentage of sign correct forecasts- PSC	$PSC = \frac{100}{h} \sum_{t=n+1}^{n+h} z_t$	$z_t = 1, y_t \cdot \hat{y}_t > 0$ $z_t = 0, otherwise$
Percentage of directional accuracy correct forecasts- PDA	$PDA = \frac{100}{h} \sum_{t=n+1}^{n+h} z_t$	$z_t = 1, (y_t - y_{t-1})(\hat{y}_t - y_{t-1}) > 0$ $z_t = 0, otherwise$

The proposed methodology consists in the following steps:

- The computation of sums of summary statistics after the division to each standard deviation (S1);
- The computation of sum of relative accuracy measures (S2);
- The computation of sum of percentage for directional and sign accuracy (S3).

For the first indicator S1, the MSE has been excluded, because it has the same significance as RMSE. S1 and S2 should be as lower as possible, while S3 should be as high as possible. After these measures assessment, the best forecaster is chosen.

$$S_1 = \frac{|ME_t|}{SD_t^{ME_t}} + \frac{MAE_t}{SD_t^{MAE_t}} + \frac{RMSE_t}{SD_t^{RMSE_t}} + \frac{MAPE_t}{SD_t^{MAPE_t}} \quad (1)$$

$$S_2 = U_1 + MRAE + RRMSE + MASE \quad (2)$$

$$S_3 = PSC_t + PDA_t \quad (3)$$

Let us consider the actual values of a variable $\{y_t\}, t = 1, 2, \dots, T$ and two predictions for it $\{\hat{y}_{t1}\}, t = 1, 2, \dots, T$ and $\{\hat{y}_{t2}\}, t = 1, 2, \dots, T$. The prediction errors are computed as: $e_{it} = \hat{y}_{it} - y_t, i=1,2$. The loss function in this case is calculated as:

$$g(y_t, \hat{y}_{it}) = g(\hat{y}_{it} - y_t) = g(e_{it}) \quad (4)$$

In most cases this function is a square-error loss or an absolute error loss function.

Two predictions being given, the loss differential is:

$$d_t = g(e_{1t}) - g(e_{2t}) \quad (5)$$

The two predictions have the same degree of accuracy if the expected value of loss differential is 0.

For Diebold-Mariano (2002) test, the null assumption of equal accuracy checks if the expected value of differential loss is zero: $E(d_t) = 0$. The covariance stationary been given, the distribution of differential average follows a normal distribution. The DM statistic, according to Diebold and Mariano (2012), under null hypothesis is:

$$S_1 = \frac{\bar{d}}{\sqrt{\hat{V}(\bar{d})}} \rightarrow N(0,1)$$

$$\bar{d} = \frac{\sum_{t=1}^n d_t}{n} \quad (6)$$

$$\hat{V}(\bar{d}) = \frac{\hat{y}_0 + 2 \sum_{k=1}^{n-1} \hat{y}_k}{n}$$

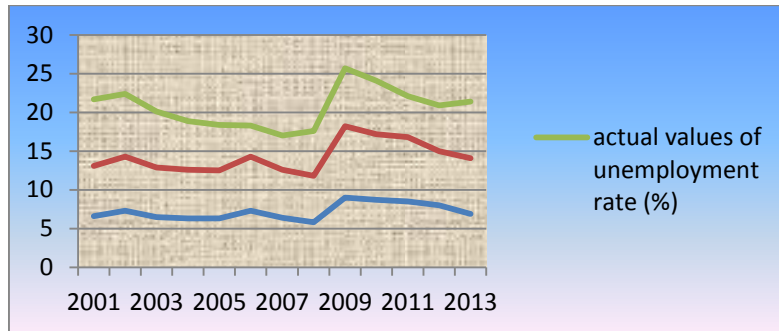
$$\hat{y}_k = \frac{\sum_{t=k+1}^n (d_t - \bar{d})(d_{t-k} - \bar{d})}{n}$$

Instead of estimating the variance we can study the prediction error auto-covariances. This test does not suppose restrictions like forecast errors with normal distribution, independent and contemporaneously uncorrelated predictions errors.

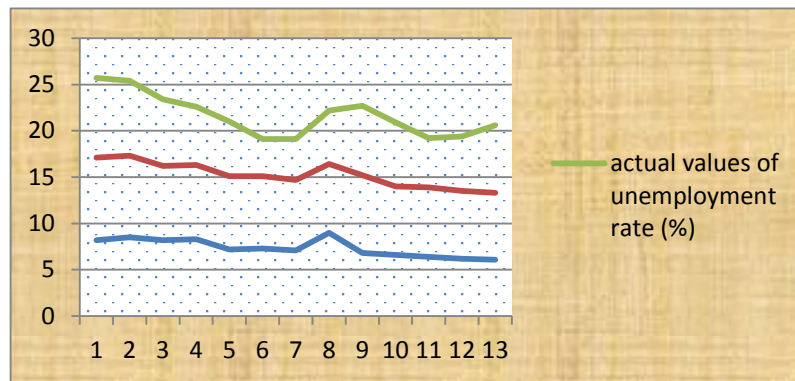
4. The Assessment of Unemployment Rate Forecasts

For the unemployment rate during the economic crisis 2009-2013, we used the predictions provided by the following forecasters: F1, F2 and F3. One-step-ahead forecasts were provided, these predictions being made at the same time. With red and blue line are drawn the predictions at time h and respectively h+1.

F1



F2



F3

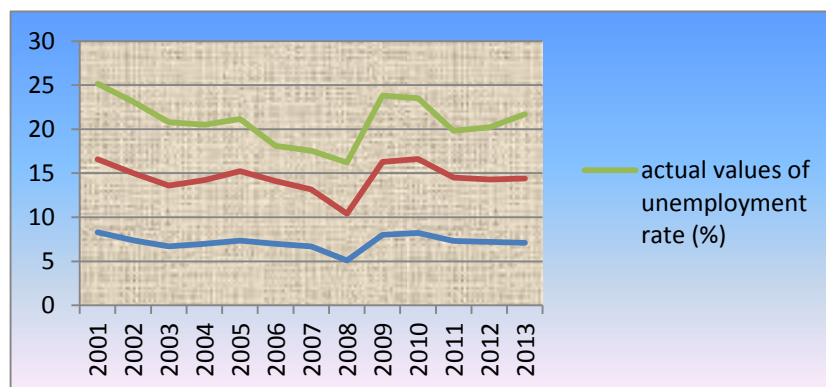


Figure 1. Scenarios for unemployment rate forecasts in Romania

For all the forecasters the spring versions provided higher forecasts errors than the autumn/winter scenarios. This is well explained by the fact that the horizon is smaller

in the second scenario compared to the spring version. The spring versions of the current year made by the F1 and F2 were used for the next year forecasts.

Table 4. The evaluation of accuracy measures for unemployment rate forecasts (2006-2013)

Indicator	F1	F2	F3	
Mean error- ME		-1,4813	0,1563	-0,8313
Mean absolute error- MAE		1,5563	1,3188	1,2438
Root mean squared error- RMSE		1,6986	1,5084	1,3921
Mean squared error- MSE		2,8853	2,2753	1,9378
Mean absolute percentage error- MAPE		14,6959%	11,0105%	11,8670%
U1 Theil's statistic		0,1232	0,1237	0,1058
Mean relative absolute error- MRAE		2,2142	3,2134	7,1259
Relative Root mean squared error- RRMSE		1,0708	0,9509	0,8775
Mean absolute scaled error- MASE		1,1940	1,0290	0,8503
Percentage of sign correct forecasts- PSC		100%	100%	100%
Percentage of directional accuracy correct forecasts- PDA		62,5%	62,5%	75%

According to U1 Theil's statistic, F3 provided the most accurate forecasts. MASE value confirms the superiority of these forecasts that outperformed the naïve predictions. The lowest values for ME, MAE, RMSE and MSE are also registered by these appreciations of unemployment rate evolution. The value for MRAE is very large compared to the other forecasts.

Table 5. The values of S1, S2 and S3 indicators for assessing the accuracy of unemployment rate forecasts (2006-2013)

Indicator	F1	F2	F3	
S1		29,93157	23,72887	23,78
S2		4,6022	5,3170	8,9595
S3		162,5%	162,5%	175%

The lowest value of S1 was registered by F2, while F1 had the smallest value for S2. F3 provided the best forecasts of unemployment rate in terms of directional and sign accuracy. As we can observe each aggregated indicator shows a different expert as the best forecasts provider. Therefore, the multi-criteria ranking is applied to determine the most accurate forecasts. Actually, the MRAE value is the indicator that defaced the good accuracy of F3 predictions.

The method of relative distance with respect to the maximal performance is employed in this study. It is calculated the distance between each prediction and the one with the highest degree of accuracy. The closer the prediction is to the best one, the higher the accuracy is. The method is applied for S1 and S2 for which the performance is judged according to the minimum value. A distance of each forecaster with respect to the one with the best performance is computed for each accuracy indicator. The distance is calculated as a relative indicator of coordination:

$$d_{i\ ind_j} = \frac{ind_i^j}{\{\min abs(ind_i^j)\}_i}, i=1,2,3 \text{ and } j=1, 2. \tag{7}$$

The relative distance computed for each forecaster is presented as a ratio, where the best value for the accuracy indicator for all experts is the denominator.

A geometric mean for the distances of each institution is calculated, its significance being an average relative distance for institution *i*.

$$\bar{d}_i = \sqrt{\prod_{j=1}^2 d_{i\ ind_j}}, i=1,2,3 \tag{8}$$

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

$$loc_i^{\%} = \frac{\bar{d}_i}{\min(d_i)_{i=1,3}} \cdot 100 \tag{9}$$

Table 6. Ranks of Institutions According to the values of S1 and S2 (Method of Relative Distance with Respect to the Best Forecaster)

ACCURACY MEASURE	F1	F2	F3
S1	1,2614	1,0000	1,0022
S2	1,0000	1,1553	1,9468
Average relative distance	1,1231	1,0749	1,3968
Ranks	2	1	3
Location (%)	104.4902	100	129,9499

The results of multi-criteria ranking application show that F2 provided the most accurate forecasts and F3 the less accurate. However, according to S3, F3 is the best forecaster in terms of directional and sign accuracy. The Diebold-Mariano test was employed to check the differences in accuracy between the unemployment rate forecasts of the three experts. The maximum lag is 6 chosen by Schwartz criterion and the Kernel is uniform.

Table 7. The forecasts accuracy comparisons based on Diebold-Mariano test

Comparison	DM statistic value	MSE	Expert with the more accurate forecasts
F1-F2	S(1) = 5.571 p-value = 0.0000	F1 2.885 F2 2.275	F2
F1-F3	S(1) = 12.56 p-value = 0.0000	F1 2.885 F3 1.938	F3
F2-F3	S(1) = .348 p-value = 0.7279	F1 2.275 F3 1.938	No differences between F2 and F3 forecasts

According to Diebold-Mariano test F2 and F3 forecasts are more accurate than F1 predictions, but there are not significant differences in terms of accuracy between F2 and F3 predictions. These results are also presented in Appendix 1. The actual economic crisis explains the decrease in accuracy of the F3 predictions. The econometric models did not take into account all the shocks in the labour market.

5. Conclusions

It is clearly that F3 provided the best forecasts in terms of directional and signed accuracy, but the errors' magnitude is higher than that of the other experts. Our methodology based on aggregated indicators S1 and S2 that were ranked using the method of relative distance with respect to the best expert indicated that F2 forecasts for unemployment rate forecasts in Romania on 2006-2013 were the most accurate. The Diebold-Mariano test identified F1 predictions as the less accurate, but significant accuracy differences were not found between F3 and F2 predictions. A further research may consider another aggregated indicator based on the sum of S1 and S2, taking into account that a lower value will show a better accuracy.

6. Acknowledgement

This article is a result of the project POSDRU/159/1.5/S/137926, Routes of academic excellence in doctoral and post-doctoral research, being co-funded by the European Social Fund through The Sectorial Operational Programme for Human Resources Development 2007-2013, coordinated by The Romanian Academy.

7. References

- Artis, M. J. (1996). How Accurate Are the IMF'S Short-Term Forecasts? Another Examination of Economic Outlook. *Staff studies of the world economic outlook*, Vol. 96, No. 89, pp. 1-94.
- Ashiya, M. (2003). The Directional Accuracy of 15-Months-Ahead Forecasts Made by the IMF. *Applied Economics Letters*, Vol. 10, No. 6, pp. 331-333.
- Diebold, F.X. & Mariano R. (2002). Comparing Predictive Accuracy. *Journal of Business and Economic Statistics*, Vol. 20, No. 1, pp. 134-144.

Frenkel, M., Rülke, J.C. & Zimmermann, L. (2013). Do private sector forecasters chase after IMF or OECD forecasts? *Journal of Macroeconomics*, Vol. 37, No. 1, pp. 217-229.

González Cabanillas, L. & Terzi, A. (2012). The accuracy of the European Commission's forecasts re-examined. *Economic Papers*, Vol. 476, No. 1, pp. 1-53.

Green, K. & Tashman, L. (2008). Should We Define Forecast Error as $e = F - A$ or $e = A - F$? *Foresight: The International Journal of Applied Forecasting*, Vol. 10, No. 1, pp. 38-40.

Liu, D. & Smith, J. K. (2014). Inflation forecasts and core inflation measures: Where is the information on future inflation? *The Quarterly Review of Economics and Finance*, Vol. 54, No. 1, pp. 133-137.

Melander, A., Sismanidis, G. & Grenouilleau, D. (2007). The track record of the Commission's forecasts-an update. *Directorate General Economic and Monetary Affairs (DG ECFIN) Working Paper*, No. 291, pp. 1-110.

Öller L-E & Barot B. (2000). The Accuracy of European Growth and Inflation Forecasts. *International Journal of Forecasting*, Vol. 16, No. 3, pp. 293-315.

Pesaran, M. H. & Timmermann A.G. (1994). A Generalization of the Edited the Non-Parametric Henriksson-Merton Test of Market Timing. *Economics Letters*, Vol. 44, No. 1, pp. 1-7.

Pons, J. (2001). The Rationality of Price Forecasts: A Directional Analysis. *Applied Financial Economics*, Vol. 11, No. 3, pp. 287-290.

APPENDIX 1

Diebold-Mariano test results

Series	MSE
F1	2.883
F2	2.275
Difference	0.61
S1	5.571 (p-value=0.000)

Series	MSE
F1	2.885
F3	1.938
Difference	0.9475
S1	12.56 (p-value=0.000)

Series	MSE
F2	2.275
F3	1.838
Difference	0.61
S1	0.348 (p-value=0.7279)