EKONOMETRIA ECONOMETRICS 32 • 2011

ISSN 1507-3866

Maria M. Kaźmierska-Zatoń
State Higher Vocational School in Skierniewice

Wojciech Zatoń
University of Lodz

MEASURING THE QUALITY OF MULTIPERIOD ROLLING COMPETING FORECASTS

Summary: For multiperiod rolling forecasts for the same phenomena developed by many forecasters, a complex problem arises of the fair evaluation of the quality of these competing forecasts. The score for a single period should then be only a component of an overall assessment of the sequence of forecasts for multiple periods. Assuming the regularity in the preparation of forecasts and their regular revisions for a given period, we get several forecasts made with various advances. In the final evaluation of the whole sequence of forecasts one should take this issue into account, introducing weights for the forecasts with different horizons. The article discusses some aspects of measuring the quality of competing multiperiod rolling forecasts noting the above-mentioned problems and proposing some evaluation systems.

Key words: rolling forecasts, evaluation systems, competing forecasts.

1. Introduction

Preparing forecasts for different phenomena, not only those concerning economic issues, is not an easy task although a wide range of methods are available. This even may create some problems since various methods can generate different forecasts. Therefore, choosing the best one becomes quite a challenge. Moreover, in many cases the forecasts are prepared by authors representing different analytical groups. In our opinion among numerous types of forecasts, multiperiod rolling forecasts occupy a special position. They can be applied in many different areas, on different levels of activity, particularly in financial planning (state budget [Rup 2009], budget of a company [Hauzer 2008]), in making decisions on a macroeconomic scale or investing in financial markets. The analysis of multiperiod rolling forecasts enables to evaluate the reliability of forecasting centres better than any other type of forecast.

In the case of developing the above mentioned forecasts for the same phenomena (categories) prepared by many forecasters, the problem of the fair multi-aspect evaluation of the quality of the entire sequence of forecasts for multiple periods arises. The evaluation of a single period forecast should be then only an element of the overall assessment of the sequence of forecasts for many periods. Assuming the regularity in the preparation of forecasts and their regular updating, we obtain for consecutive periods a number of forecasts for a certain period prepared with various advancement. Thus, in the final evaluation of the entire sequence of forecasts, this ought to be taken into consideration by building a certain scheme of weights for forecasts with different horizons. The article discusses some aspects of measuring the quality of competing multiperiod rolling forecasts. We focus our attention on the above mentioned issues, and we analyze the evaluation systems corresponding with them. We put forward a thesis that in the case of evaluating multiperiod rolling forecasts we should consider the following elements:

- applied measures,
- evaluation criteria,
- varying set of information available in the course of forecast preparation.

Including these elements offers the possibility of multi-aspect creation of evaluation systems which cannot be generalized because of the variety and peculiarity of the forecasted phenomena. The article presents and describes the scheme of developing multiperiod rolling forecasts. Then a number of multiperiod rolling forecasts evaluation systems and the results of their application are presented.

2. The problem

The subject of the study presented in the article is the evaluation of the quality of multiperiod rolling forecasts with different horizons developed by many centres for the same periods at the same time. The quality of these forecasts will be evaluated by their accuracy since this element of the quality can be easily measured and it influences the general utility of the forecasts.

Taking into consideration the number of periods which the forecasts are being prepared for, the permanence of their horizons and how systematically they are developed, four types of forecasts (for a given category in a particular centre) can be distinguished along with the methods of their evaluation:

1. Single one-period forecast, prepared occasionally for a given period.

2. Single multiperiod forecast, prepared occasionally for several periods (a forecast with a horizon longer than one period).

3. One-period rolling forecast, prepared systematically with a moving horizon of a one-period length (e.g. [Borowski 2009] and prior articles – www.borowski.pl/publikacje/], forecasts published by experts of „Miesięcznik Kapitałowy" and „Nowe Życie Gospodarcze" – www.pte.pl).

4. Multiperiod rolling forecasts, prepared systematically for several periods with a moving horizon ([Wilkowicz 2010; Wyżnikiewicz et al. 2010]) or a fixed horizon ("to the wall", e.g. [Player 2009]).

Rolling forecasting consists in a permanent shifting of the forecasting horizon. To make the process of multiperiod rolling forecasts more comprehensible, we present Table 1, where **P** denotes a single forecast prepared in period *i* for period *j*, *k* – a number of periods for which forecasts are made in period *i*.

Table 1. Overall scheme of multiperiod rolling forecasts

Period when forecast is prepared (*i*)	Period a forecast is prepared for (*j*)										
	T+1	T+2	…	T+k	T+k+1	T+k+2	…	T+2k	T+2k+1	T+2k+2	…
T	**P**	**P**	**…**	**P**							
T+1		**P**	**…**	**P**	**P**						
T+2			**…**	**P**	**P**	**P**					
…				**P**	**P**	**P**	**…**				
T+k					**P**	**P**	**…**	**P**			
T+k+1						**P**	**…**	**P**	**P**		
T+k+2							**…**	**P**	**P**	**p**	
…								**P**	**P**	**P**	…

Source: own study.

If we are forecasting in a given period for, e.g. a year, preparing forecasts for every month, then k equals 12. After the first period passed, the series of monthly forecasts is being updated in order to have every month a forecast with an annual horizon. This method of preparing forecasts enables to obtain a series of forecasts for a given period *j* prepared in various periods *i*. This makes rolling forecasts different from single ones.

The ex post assessment of a single one-period forecast's accuracy for a given category is very simple – it is sufficient to calculate the forecast error as the difference between forecast and actual value.

To evaluate the accuracy of a single multiperiod forecast, as well as one-period rolling forecast, a wide range of ex post forecast error measures (e.g. MAE, RMSE, MAPE, MRAE [Hyndman 2006], [Cieślak 2000]) may be applied. An alternative criterion for quality evaluation of one-period rolling forecasts .may be scores arbitrarily granted to an individual centre, depending on the scale of the forecast error in a particular period, or the position in ranking of all centres in this period.

Evaluating becomes more complex when we attempt to compare the accuracy of multiperiod rolling forecasts provided by several centres. Table 2 shows a sample diagram of the preparation of quarterly rolling forecasts of the annual horizon (i.e. for the next four quarters) in the period 2008-2010. Such a scheme is used in the example presented in the third part of the article.

Table 2 refers to the forecasts prepared by one centre for an individual phenomena (category). A sequence of forecasts is understood as a series of forecasts prepared by

Table 2. Diagram of quarterly rolling forecasts with an annual horizon

		Period a forecast is prepared for											
		2008				2009				2010			
Period when forecast is prepared		I q	II q	III q	IV q	I q	II q	III q	IV q	I q	II q	III q	IV q
2008	I q		░	▒	▓	█							
	II q			░	▒	▓	█						
	III q				░	▒	▓	█					
	IV q					░	▒	▓	█				
2009	I q						░	▒	▓	█			
	II q							░	▒	▓	█		
	III q								░	▒	▓	█	
	IV q									░	▒	▓	█
2010	I q										░	▒	▓
	II q											░	▒
	III q												░

░ – forecasts with a horizon length of 1 quarter
▒ – forecasts with a horizon length of 2 quarters
▓ – forecasts with a horizon length of 3 quarters
█ – forecasts with a horizon length of 4 quarters

Source: own study.

a given centre in four successive periods for the same quarter of the year (forecasts in the columns of Table 2). We assume that for the final evaluation of the quality of forecasts developed by all the centers, the forecasts for four quarters of the year are covered (forecasting cycle). Thus, a cycle of forecasts names a set of four series of forecasts prepared for all the quarters of a year. As Table 2 shows, the first evaluation of the entire cycle may refer to the year 2009 (for every quarter four forecasts were prepared).

The quality of the forecasts submitted by a given centre is evaluated on three levels: a single forecast (one-period forecast), a series of forecasts and a cycle of forecasts. We assume that the final evaluation of a cycle of forecasts presented by a given centre has to be put in one number.

To repeat the thesis advanced in the introduction to the article, we think that assessing the results of the process of multiperiod rolling forecasts should meet the following elements:

– applied measures,
– evaluation criteria,
– varying set of information available in the course of forecast preparation.

As to the applied measurement, it has to be decided whether in the whole process of evaluation we use directly absolute or relative deviations of forecasts from actual values or we define ranges of such deviations on the basis of which an arbitrary score scale for evaluation is created.

The main criterion of an individual forecast's accuracy of a given centre is its deviation from the actual value. In the case of the evaluation of competing forecasts, it seems to be justified to place this relation in the background of other centres' forecasts, which will be represented by the so called *consensus forecast*. The measurement of the latter can be the mean or the median of the competing forecasts. The quality of a particular forecast could be then defined not only on the basis of the absolute deviation from the actual value, but also on examining the deviation of the forecast from the consensus forecast.

In the process of multiperiod rolling forecasting, forecasts of different length of horizon are developed (Table 2). Obviously, the accuracy of the forecast with a long horizon (prepared four quarters in advance) ought to be evaluated higher than the one of the forecasts with a short horizon (e.g. one quarter). This is the outcome of the supply and reliability of information at the time the forecast was being developed. This can be easily allowed for by introducing weights attributed to forecast errors within the series of forecasts.

The above consideration proves that evaluation systems of multiperiod rolling forecasts may be of a multi-aspect type, particularly when, apart from the above mentioned factors a special character of the forecasted phenomena is allowed for, e.g. the scale of their changeability. The complexity of the assessment of this kind of forecasts increases when we assume that a given forecasting centre prepares forecasts for many categories. Then, the evaluation of multiperiod rolling forecasts quality of the centre is developed on four levels.

In part 3 of the article we propose several evaluation systems for the multiperiod rolling forecasts for one category developed simultaneously by several centres, and we evaluate their functioning on an example.

3. The example

The forecasted category was GDP growth (percentage change from the same quarter of prior year) in Poland in the period 2008-2010. The evaluation of the cycle of forecasts referred to the year 2009 which, because of the economic crisis, was exceptional and difficult for forecasting centres. We assumed that six centres prepared quarterly rolling forecasts of GDP growth with an annual horizon. The first (o1) and the second (o2) centres developed forecasts using trend extrapolation methods described by models whose parameters were estimated on the basis of empirical data with, respectively, cumulative number of observations (o1), and constant, moving number of observations (o2). The third centre (o3) applied the naive method with seasonality and trend adjustment. It was assumed that the forecasts of the fourth (o4) and fifth (o5) centres result from the opinions of experts. The forecasters from the fourth centre were optimistic and unwillingly lowered the value of the expected GDP

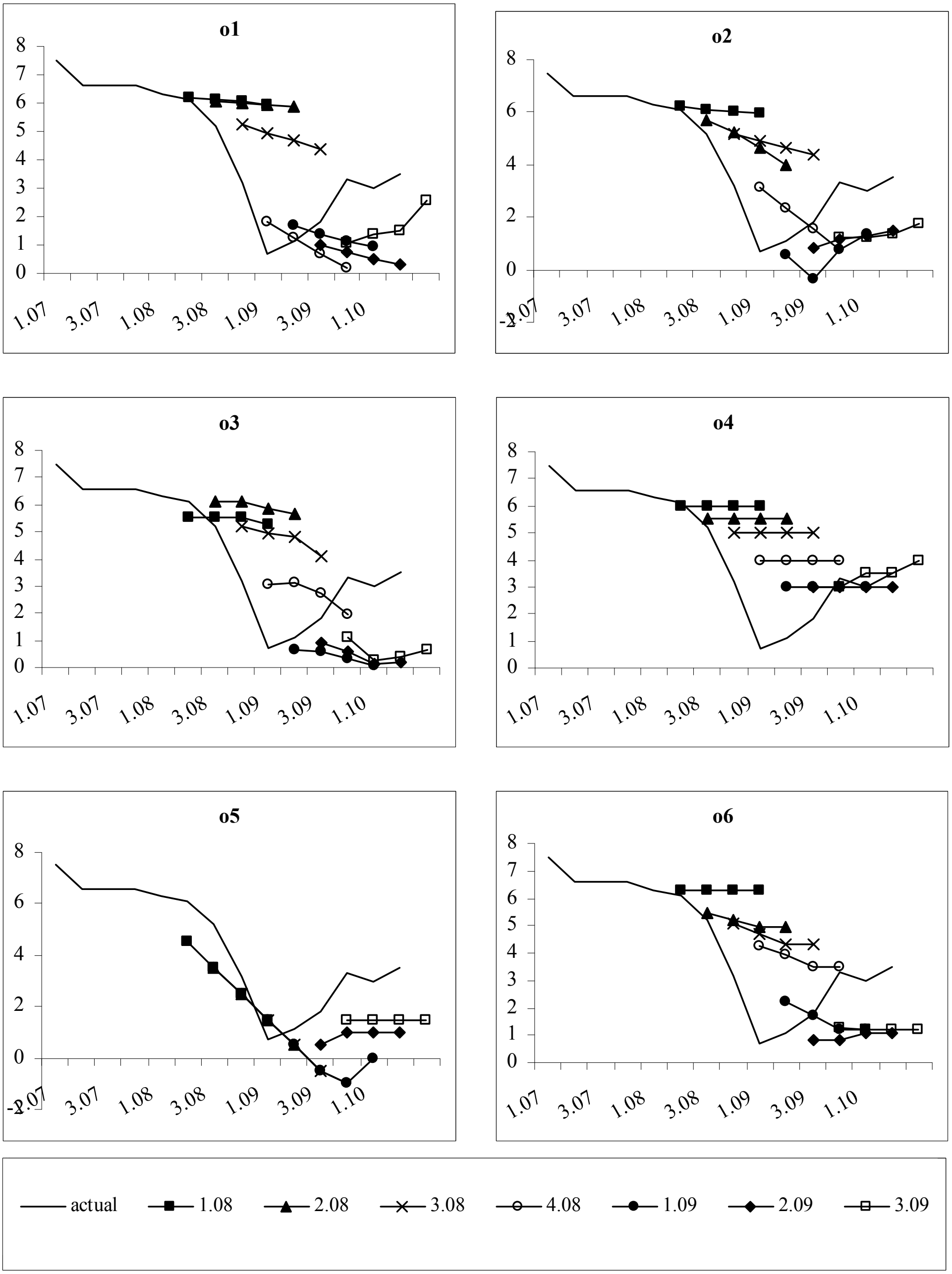

Figure 1. Rolling forecasts of GDP growth (%) by centres, prepared in periods indicated in figure key (e.g. 1.08 denotes first quarter of 2008)

Source: own study.

growth for 2009 and 2010 (the minimum value of forecast in this period was +3%, the minimum actual value was +0,5% in the first quarter of 2009). On the contrary, the forecasters from the fifth centre were pessimistic and, facing the economic crisis, were quickly lowering the forecasts of GDP growth (down to –1% in the fourth quarter of 2010). The analysts from the sixth centre (o6) decided to make their task simpler and applied the method of following the forecasters from other centres. Their forecasts for the successive quarter were set as the average from the forecasts of the remaining centres for the latest quarter.

The selection of methods presented in the example is subjective. Actually, forecasts are often some kind of combination (particularly macroeconomic forecasts which are being prepared by specialized forecasting centres, e.g. [Clemen 1989]). At the same time, we think, and the surveys prove the same, (e.g. [Garczarczyk, Mocek 2010]) that quantitative methods proposed by us, as well as the opinions of experts, are very often used in forecasting on the corporate level.

In detail, GDP growth expected by the centres are presented in Figure 1. The significant variety of forecasts is the result of the way they were prepared. The fact that the values of the forecasts considerably differ from the actual rate may be understood as an illustration of the problem mentioned in the previous part of the article − the influence of the varying set of information available in the course of forecast preparation on forecasts' quality (Figure 2 and 3). The forecasts of the majority of centres, prepared with the longest horizons (i.e. four quarters) are definitely further from actual values than those prepared one quarter in advance.

It is also worth noticing that in the analyzed period a reversion of direction of the actual changes in forecasted category followed, which usually has a negative influence on the accuracy of multiperiod forecasts.

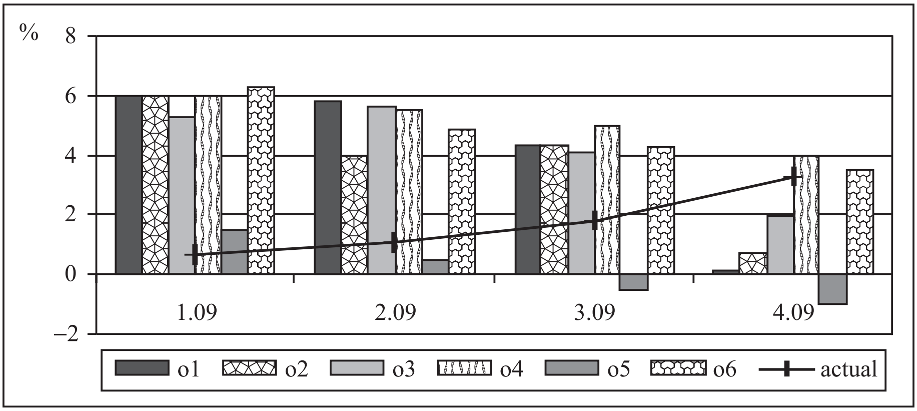

Figure 2. Forecasts of GDP growth (%) by centres, prepared for subsequent quarters of 2009 with maximum horizon length of four quarters.

Source: own study.

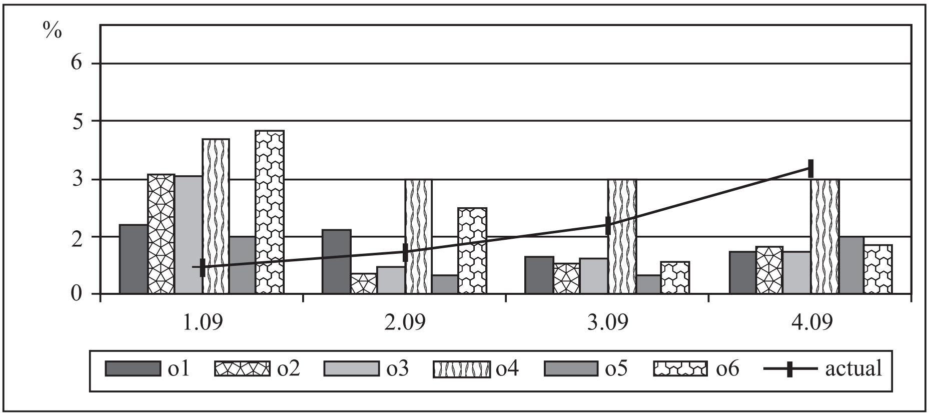

Figure 3. Forecasts of GDP growth (%) by centres, prepared for subsequent quarters of 2009 with minimum horizon length of one quarter

Source: own study.

Table 3. Characteristics of forecasts' evaluation systems.

Variant of evaluation system	Evaluation measure of a single forecast (ESF)	Evaluation measure of a series of forecasts (ESSF)	Evaluation measure of a cycle of forecasts (ECF)
1	percentage error ep_{ij}^{k} (eq. 1)	weighted average of absolute ESF$_i$ MEP_j^k (eq. 2)	arithmetic mean of ESSF$_j$ $MMEP^k$ (eq. 3)
2	as above.	scores according to the scale determined by MEP_j^k (eq. 4)	sum of ESSF$_j$
3	relative percentage error rep_{ij}^{k} (eq. 5)	weighted average of absolute ESF$_i$ $MREP_j^k$ (eq. 6)	arithmetic mean of ESSF$_j$ $MMREP^k$ (eq. 7)
4	scores according to the scale (eq. 8)	weighted average of ESF$_i$	sum of ESSF$_j$
5	scaled absolute error se_{ij}^{k} (eq. 9)	weighted average of absolute ESF$_i$ MSE_j^k (eq. 10)	arithmetic mean of ESSF$_j$ $MMSE^k$ (eq. 11)
6	quasi-standardized error qse_{ij}^{k} (eq. 12)	weighted average of absolute ESF$_i$ $MQSE_j^k$ (eq. 13)	arithmetic mean of ESSF$_j$ $MMQSE^k$ (eq. 14)

Source: own study.

Even a careful analysis of charts showing the forecasts does not allow for a clear identification of the best forecasting centre (i.e. the one which the predictions are closest to actual values). Below, the quality of the forecasts of all the centres has been evaluated on a sample of data of four quarters of 2009 (one cycle of forecasts) using six variants of evaluation systems described in Table 3. In each variant, the final assessment of the quality of a cycle of forecasts of a given forecasting centre is represented by one number and is formed on three levels: a single forecast, a series of forecasts and a cycle of forecasts. In four proposed variants selected ex post, forecast errors measures were used directly in the evaluation process while in two remaining variants arbitrary score scales were applied in addition.

$$ep_{ij}^{k} = \frac{p_{ij}^{k} - y_j}{y_j} \times 100 \qquad (1)$$

where: p_{ij}^{k} is a forecast of centre *k* (*k* = 1, 2, 3, 4, 5, 6) prepared in period *i* (*i* = 1.08, 2.08, 3.08, 4.08) for period *j* (*j* = 1.09, 2.09, 3.09, 4.09).

y_j – actual value in period *j*.

$$MEP_{j}^{k} = \sum_{i} w_i * \left| ep_{ij}^{k} \right| \qquad (2)$$

where: w_i – weights determined arbitrarily. In every variants to evaluate a series of forecasts the following weights were adopted: 0.4 for forecasts with the horizon length of 4 quarters, 0.3 for forecasts with the horizon length of 3 quarters, 0.2 for forecasts with the horizon length of 2 quarters and 0.1 for forecasts with the horizon length of 1 quarter.

$$MMEP^{k} = \frac{1}{4} \sum_{j} MEP_{j}^{k} \qquad (3)$$

$0 \le MEP_{j}^{k} \le 50$ – 4 scores, 50 < $MEP_{j}^{k} \le 100$ – 3 scores, 100 < – 2 scores, $150 < MEP_{j}^{k} \le 200$ – 1 score, $MEP_{j}^{k} > 200$ – 0 scores. (4)

$$rep_{ij}^{k} = \frac{p_{ij}^{k} - y_j}{s_{ij} - y_j} \times 100 \qquad (5)$$

where: s_{ij} is a consensus forecast calculated as an arithmetic mean of all centres' forecasts p_{ij}^{k}. Alternatively the median could be applied instead of the mean.

$$MREP_{j}^{k} = \sum_{i} w_i * \left| rep_{ij}^{k} \right| \qquad (6)$$

The measure defined by the equation above is a kind of modification of known relative error measures such as *MdRAE* (e.g. [Armstrong, Collopy 1992)]).

$$MMREP^{k} = \frac{1}{4} \sum_{j} MREP_{j}^{k} \qquad (7)$$

scores according to the following scale: (8)

- 5 – if $\left| p_{ij}^{k} - y_j \right| \le 20\% SP_{ij}$
- 3 – if $\left| p_{ij}^{k} - y_j \right| > 20\% SP_{ij}$ and $\left| p_{ij}^{k} - y_j \right| < \left| m_{ij} - y_j \right|$

- 1 – if $\left|p_{ij}^{k}-y_{j}\right|>20\% SP_{ij}$ and $\left|p_{ij}^{k}-y_{j}\right|=\left|m_{ij}-y_{j}\right|$
- 0 - if $\left|p_{ij}^{k}-y_{j}\right|>20\% SP_{ij}$ and $\left|p_{ij}^{k}-y_{j}\right|>\left|m_{ij}-y_{j}\right|$

where: SP_{ij} − a standard deviation of forecasts prepared by all centres in period *i* for period *j*,

m_{ij} is a consensus forecast calculated as the median of all centres' forecasts p_{ij}^{k}.

The above rules of score scale building and evaluation system based on this scale (variant 4 in Table 3) are actually applied in the competition for the best macroeconomic analyst run by NBP (The National Bank of Poland) and two newspapers: the Parkiet and the Rzeczpospolita [http://www.nbportal.pl/r/res/edukacja/kryteria.pdf].

$$se_{ij}^{k}=\frac{p_{ij}^{k}-y_{j}}{\frac{1}{n-1}\sum_{t=2}^{n}\left|y_{t}-y_{t-1}\right|} \tag{9}$$

where: *n*-number of forecast in a series of forecasts

$$MSE_{j}^{k}=\sum_{i}w_{i}\times\left|se_{ij}^{k}\right| \tag{10}$$

The measure defined by equation (10) is an analogue to the MASE measure proposed by Hyndman [Hyndman, Koehler 2006]).

$$MMSE^{k}=\frac{1}{4}\sum_{j}MSE_{j}^{k} \tag{11}$$

$$qse_{ij}^{k}=\frac{p_{ij}^{k}-y_{j}}{SP_{ij}} \tag{12}$$

$$MQSE_{j}^{k}=\sum_{i}w_{i}*\left|qse_{ij}^{k}\right| \tag{13}$$

$$MMQSE^{k}=\frac{1}{4}\sum_{j}MQSE_{j}^{k} \tag{14}$$

Accuracy evaluations of the cycle of forecasts for 2009 achieved by various centers are shown in Table 4. The values contained herein may not be directly comparable, however, for two reasons. Firstly, in particular variants of evaluation systems different measures were used. Secondly, a type of measure diversely determines the choice of the most accurate forecasts. In variants 1, 3, 5 and 6 the best centres are those with forecasts that were given the lowest score, while in variants 2 and 4 − the highest. However, the results shown illustrate well the variation in the quality of a cycle of forecasts among centres, which must be regarded as very moderate. Small variations in evaluations for the entire cycle of forecasts compared to a quite large discrepancy of single forecasts arise from a kind of smoothing process based on averaging at two levels − the evaluation of a series of forecasts and a cycle of forecasts. The final judgment may be better formulated after examining Figure 4 and the ranking shown in Table 5. While the position of the best forecasting centre, according to most variants of evaluation system, is granted to centre o5 (pessimistic experts) then the centres' classification behind the first place varies in different variants of the evaluation system. It should be noted, however, that on two occasions, first place in the ranking was occupied by centers other than o5. It is especially interesting to see the success of forecasters from the centre o6 simply mimicking forecasts from other centres (preparing a forecast for the successive quarter as the average from the forecasts of the remaining centres for the latest quarter).

Table 4. Accuracy estimates of a cycle of forecasts of all forecasting centres in different variants of evaluation system in 2009

	Variant of evaluation system					
Centre	*1*	*2*	*3*	*4*	*5*	*6*
o1	276	6	155	4	3	8
o2	254	6	**126**	6	3	8
o3	274	6	149	4	3	8
o4	289	6	202	4	3	7
o5	**100**	**10**	185	7	**2**	**5**
o6	272	7	169	**8**	3	8

Source: own study.

Table 5. Ranking of the forecasting centres

	Variant of evaluation system					
Rank	*1*	*2*	*3*	*4*	*5*	*6*
1	o5	o5	o2	o6	o5	o5
2	o2	o6	o3	o5	o6	o4
3	o6	o2	o1	o2	o2	o6
4	o3	o3	o6	o4	o4	o2
5	o1	o1	o5	o1	o3	o1
6	o4	o4	o4	o3	o1	o3

Source: own study.

The least accurate forecasts were prepared by centres o1 (four times at the penultimate position in the ranking, once at the last one) and o4 (three times at the last place in the rankings, twice at the fourth place). It is worth recalling that these centres were handling the forecasting methods "of different philosophy" – the centre o1 extrapolated trends of cumulative number of observations, while the experts from centre o4 were preparing quite steady optimistic forecasts. Under current economic conditions of 2008-2009 both methods yielded similarly evaluated forecasts.

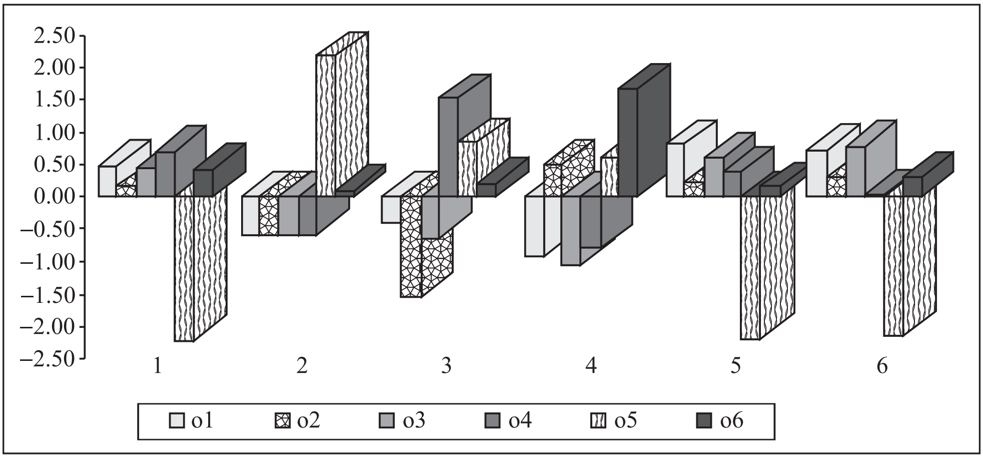

Figure 4. Standardized accuracy estimates of a cycle of forecasts for all forecasting centres by different variants of evaluation system

Source: own study.

Some kind of multi-criteria evaluation was formulated on the basis of standardized accuracy estimates of a cycle of forecasts for all centres. This is simply the sum of standardized accuracy estimates obtained in each variants of an evaluation system (to keep the same direction for the selection of the best centre in all variants, score ratings for variants 2 and 4 were multiplied by –1). The best overall evaluation was obtained by centre o5, and a further order in the overall ranking was as follows: o4, o1, o3, o2, o6.

4. Conclusion

The article shows that the measurement process of the quality of competing multiperiod rolling forecasts is quite complex. It should include the following elements: applied measures, evaluation criteria and the volatility of the stock of existing information at the time when the forecast is formed. On the basis of these factors various evaluations systems can be constructed. As demonstrated in the example, the results obtained in various evaluation systems may differ, i.e. the ranking of the same forecasts may be different upon different evaluation systems. This hampers a clear evaluation of the quality of multiperiod rolling forecasts from competing centres and indication which centre deserves the highest praise. A possible solution to arrive at a more clear conclusion is to use the sum of the standardized accuracy estimates from different evaluation systems. The process of averaging a series of forecasts and a cycle of forecasts at different levels of evaluation, , leads to a smoothed final evaluation even in the presence of the high dispersion of single forecasts.

Literature

Armstrong J.S., Collopy F., *Error measures for generalizing about forecasting methods: Empirical comparisons*, "International Journal of Forecasting" 1992, Vol. 8, no. 1, p. 69-80.

Borowski M., *Odwróć tabele, resort finansów na czele*, "Gazeta Wyborcza" 1czerwca 2009.

Clemen R.T., *Combining forecasts. A review and annotated bibliography*, "International Journal of Forecasting" 1989, no 5, p. 539-383.

Garczarczyk J., Mocek M., *Prognozowanie w firmach w świetle wyników badań ankietowych,* [in:] *Prognozowanie w zarządzaniu firmą*, red. P. Dittmann, E. Szabela-Pasierbińska, Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu nr 103, Wrocław 2010.

Hauzer M., *Planowanie – od adhokracji do powtarzalnego procesu*, "Business Intelligence Magazine" 01/2008.

Hyndman R.J., Koehler A.B., *.Another look at measures of forecast accuracy*, "International Journal of Forecasting" 2006, Vol. 22(4), p. 679-688.

Player S., *Managing through change: The power of rolling forecasts,* "Innovation in Action Series," czerwiec 2009, IBM Cognos Innovation Center.

Prognozowanie gospodarcze. Metody i zastosowania, red. M. Cieślak, PWN, Warszawa 2000.

Rup W., *Przejrzysty i efektywny budżet od 2010 r.*, "Rzeczpospolita" 16 września 2009.

Wilkowicz Ł, *Najlepsi w prognozowaniu*, "Parkiet" 28 maja 2010.

Wyżnikiewicz B., Fundowicz J., Lada K., Łapiński K., Peterlik M., *Stan i prognoza koniunktury gospodarczej,* "Kwartalne Prognozy Makroekonomiczne" kwiecień 2010, nr 66.

www.borowski.pl/publikacje/.

www.IBNGR.pl.

www.nbportal.pl/r/res/edukacja/kryteria.pdf).

www.pte.pl.

POMIAR JAKOŚCI KROCZĄCYCH KONKURENCYJNYCH PROGNOZ WIELOOKRESOWYCH

Streszczenie: W kontekście wielookresowych prognoz kroczących, dla tych samych zjawisk (kategorii) opracowanych przez wielu prognostów, powstaje problem obiektywnej wielopłaszczyznowej oceny jakości ciągu konkurencyjnych prognoz. Ocena dla pojedynczego okresu powinna być wtedy tylko elementem składowym ogólnej oceny za ciąg prognoz dla wielu okresów. Zakładając systematyczność w sporządzaniu prognoz, ich regularne uaktualnianie, otrzymujemy na dany okres np. kilka prognoz sporządzonych z różnym wyprzedzeniem. W ostatecznej ocenie całego ciągu prognoz należy wziąć tę kwestię pod uwagę, budując np. określony schemat wag dla prognoz o różnych horyzontach. W artykule omówiono wybrane aspekty pomiaru jakości konkurencyjnych kroczących prognoz wielookresowych, zwracając uwagę na wyżej przedstawione problemy i proponując systemy ocen, które je uwzględniają.