THE ASSESSMENT AND IMPROVEMENT OF THE ACCURACY FOR THE FORECAST INTERVALS

Mihaela Bratu (Simionescu)[*]

Abstract

The objective of this research is to present some accuracy measures associated to forecast intervals, taken into account the fact that in literature some specific accuracy indicators for this type of prediction have not been proposed yet. For the quarterly inflation rate provided by the National Bank of Romania, forecast intervals were built on the horizon 2010-2012. According to the number of intervals that include the real value and to an econometric procedure based on DUMMY variables, the intervals based on historical errors (RMSE- root mean squared errors) are better than those based on BCA bootstrap procedure. However, the new indicator proposed in this paper as a measure of global accuracy, M indicator, the forecast intervals based on BCA bootstraping are more accurate than the intervals based on historical RMSE. Bayesian intervals were constructed for quarterly USA inflation in 2012 using aprioristic information, but the smaller intervals did not imply an increase in the degree of accuracy.

Keyword: forecast intervals, accuracy, uncertainty, BCA bootstrap intervals, indicator M
JEL Classification: C10, C14, L6

I. Introduction

The use of point forecasts can not cover the uncertainty analysis that affects any process. Therefore, the decisions making and the forecasting process can not be correctly established taken into account the point predictions. The degree of uncertainty is diminished by considering the forecast intervals.

[*]*Academy of Economic Studies, Faculty of Cybernetics, Statistics and Economic Informatics.*
E-mail: mihaela_mb1@yahoo.com

A forecast interval includes the inferior limit and the superior one. Between these bounds the registered value may lie with a certain probability. Actually, the prediction interval is an estimate of a future value for a random variable, this value being unknown.
In order to construct a confindence interval there is not a general method, excepting the predictions based on a suitable probability model, according to Chatfield (2001) that showed that in this case the predictionsÕerrors could be assessed. The entire probability distribution associated to the future value of an indicator is called in literature as Ńprobability forecastŇ or Ńdensity forecastŇ.
In this paper we are interested in proposing some measures of accuracy in order to compare prediction intervals based on different methods. A global measure of accuracy is proposed in order to conciliate the existence of two sources of uncertainty in forecasting: the length of the forecast intervals and the position of the registered value compared to the limits of the intervals or the centers of the forecast intervals.

II. Literature

Chatfield (1993) pointed out that the forecasts should be expressed as prediction intervals, which is a coomon way to illustrate the uncertainty. For each future value some probabilities can be associated. Fair (2000) claimed that the possibility of an economic crisis must be included in the forecast interval.
The construction of prediction intervals and the repartitionsÕdetermination developed quite late in literature, some articles in this field being provided by Giordani and Villiani, Cogley, Adolfson, Clark and Jore. The model should include variances deviation in time in order to build a forecast interval with a certain probability.
Kjellberg and Villani (2010) specified the advantages and disadvantages of forecasts based on models and of those proposed by experts. Predictions that use models present the complex relationships between variables, the possible mistakes in forecasting being easily identified. However, it is quite difficult to adapt the model to recent economic changes. The forecast intervals are in most cases too narrow, because it is not taken into consideration the uncertainty regarding the model specification, a disadvantage met also in the expertsÕintervals. But the expert evaluations change immediately to any change of information related to the forecasted variable. A low degree of transparency characterizes the professional forecasters appreciations, being quite difficult to use many explanatory variables outside an explicit model.
Chatfield (1993) made a retrospective presentation of the methods used to build a forecasts interval. Cogley, Morozov and Sargent (2005) used the Bayesian forecast intervals for variables in the monetary system.
Christoffersen (2004) describe the way to evaluate the forecast intervals, the methods for assessing the forecasts density being introduced by Diebold, Gunther and Tat (1998) who made later the extinction for bivariate data. Wallis (2003) proposed tests chi-squared and independent tests for assessing forecasts intervals. Unlike other methods of building prediction intervals that are specified in literature, the Bayesian ones also analyze the impact of estimator error on interval. Even if Stock and Watson specified the conditional distribution function for k-steps-ahead forecasts, their approach was developed by Hansen (2005), who built asymptotic forecasts intervals to include the uncertainty determined by the parameter estimator.

Borbely and Meier (2003) proposed a procedure for assessing three types of predictable uncertainty, computing point forecasts and density forecasts. Demetrescu (2007) built optimal prediction intervals using an asymmetric loss function.
Wang and Wu (2012) built forecast intervals for the exchange rate, getting a smaller lenght of the intervals when fundamental models were used compared to the random walk. The authors made a bootstrap inference to construct the prediction intervals.
Fischer, Garcia-Barzana, Tillmann and Winker (2012) assessed the predictions accuracy for intervals, using the classical measures of accuracy using the midpoints of the intervals.
Tulip and Wallance (2012) used the historical errors to construct confidence intervals for GDP and inflation forecasts in Australia, observing a high degree of uncertainty. Anderson, Meerschaert and Zhang (2012) made gaussian forecast intervals using periodical autoregressive moving average models.

III. The accuracy assessment of the forecast intervals

Forecast intervals based on historical errors method supposes the assumption that the forecast error series is normally distributed of nullmean and a standard deviation corresponding to the root mean squared error (RMSE) of the historical forecast errors. For a probability of (1-┌), prediction interval is computed as:

$$(X_t(k) - z_{\alpha/2} \cdot RMSE(k),\ X_t(k) + z_{\alpha/2} \cdot RMSE(k)), k = 1,...,K \quad (1).$$

$X_t(k)$ - point forecast for variable X at time t for the future moment (t+k)
The forecast error is computed as the difference between the registered value (actual value) and the predicted one. In this study the quarterly inflation rate provided by the National Bank of Romania in 2000-2012 are used. The monthly registered inflation rates during 2000-2011 are published by the National Institute of Statistics. Starting from this data set, the quarterly values are determined, being expressed in prices of December 1999, seasonally adjusted (using moving average method) and differentiated in order to get stationary data series for an autoregressive-model.
The Augmented Dickey-Fuller test was applied to the adjusted and differentiated data series, the property of stationary being put in evidence for a critical thresold of 5%. The one-quarter-ahead variant of forecasts was chosen, being necessary the modelsÔupdate.

Tabel 1

Econometric model (auto-regressive models of order 1) for quarterly inflation rate in Romania

Data series horizon	AR(1) model
2000-2011	$\Delta ri_t = -0.729 + 0.672 \cdot \Delta ri_{t-1} + \varepsilon_t$
2000-2012 T1	$\Delta ri_t = -0.768 + 0.673 \cdot \Delta ri_{t-1} + \varepsilon_t$
2000-2012 T2	$\Delta ri_t = -0.722 + 0.671 \cdot \Delta ri_{t-1} + \varepsilon_t$
2000-2012 T3	$\Delta ri_t = -0.483 + 0.668 \cdot \Delta ri_{t-1} + \varepsilon_t$

Source: own estimations

Another method to construct forecast intervals is bias-corrected-accelerated bootstrap technique. This resampling method supposes the replication of sampleÕs predictions for a large number of tries. A proxy population, which is an artifical one, is built starting from this sample. The bias-corrected-accelerated interval is actually a confidence interval, Davison and Hinkley (1997) showing that the estimators for bias and acceleration are gotten using the initial sample and the bootstrap samples.

In practice, the calibration is used as method for improving the acuracy of forecast intervals. The order 2 corrections of the intervalsÕlimits are suitable second order accuracy corrections, according to Efron and Tibshirani (1993). The interval range could serve as a rudimentary measure of the estimation precision.

The National Bank of Romania periodically revises the quarterly forecasts for the inflation rate. The forecast intervals for the average predictions of a certain quarter were built, taken into account the different variants proposed at different time moments and using BCA bootstrap technique with 1000 replications. The next table presents the prediction intervals for quarterly inflation rate in Romania on the horizon 2010-2012. The BCA bootstrap method was applied using an add-in for Excel, called ***Resampling Stat.***

Table 2

Forecast intervals for quarterly inflation rate in Romania using historical errors method and BCA bootstrap technique (2010-2012)

Quarter	**Forecast intervals based on the historical errors method**		**Forecast intervals based on the BCA bootstrap method**		**Effective value**
	Lower limit	Upper limit	Lower limit	Upper limit	
2010T1	3.5882	6.0117	4	4.5	4.63
2010T2	2.7120	3.8879	4	4.9	4.36
2010T3	5.6395	8.9605	6	7.7667	7.4966
2010T4	4.9452	19.7347	8	8.5	7.8566
2011T1	8.1958	9.8041	6.7	7	7.5325
2011T2	2.8952	4.1048	8	8.7	8.2264
2011T3	3.0255	4.6744	4.8	5	4.1817
2011T4	3.8137	4.1862	3.3	4	3.5999
2012T1	0.1299	3.8100	1.97	2.3	2.4
2012T2	0.6282	4.1717	2.4	3	2
2012T3	-0.2573	7.2573	3.5	5	5.3
2012T4	0.0199	10.1800	5.1	6.5	3.33

Source: own computations

The assessment of forecast intervals accuracy is rather difficult, because of two problems: some intervlas are small and these not include the effective value, while other intervals have a large range and the probability to include the registered value is higher.

The ideal case is a small interval that includes the real value. In literature a specific accuracy indicator for forecast intervals has not been proposed yet. A simple procedure that brings essential information is represented by the number of intrevals that contains the registered value of the predicted indicator. One could use the percentage expression of the number of Ňcorrectŉ intervals out of the total number.

The intervals for inflation rate predicted by the National Bank of Romania (NBR) based on the historical errors method are better than those based on bootstraping procedure, 8 out of 12 intervals including the effective value on the horizon 2010-2012, compared to 4 successes when BCA bootstrap procedure is applied. Actually, more than half of the intervals based on RMSE (66,67%) contain the real values. The disadvantage is done by the large lenght of the intervals compared to those based on bootstraping.

It is also important to make the comparison of two intervals based on two different methods. Three different situations could be identified: none of the intervals includes the real value of the predicted variable (we can colsider that it is better the interval with the lowest distance between the real value and the intervalŝ center/ a limit of the interval), only one of the intervals contains the real value, both intervals includes the registered value (we consider the best that interval with the smallest lenght.

It is interesting to compute a measure of intervals accuracy for those intervals that exclude the real value on the forecasting horizon 2010-2012. we can made the following classification: correct intervals (that include the real value) and incorrect intervals (that exclude the real value). Furtherly, an econometric demarche, as a novelty, is proposed to compare the two methods from the point of view of the values that are not in the intervals.

Two DUMMY variables are considered regarding the unsuitable cases. Thus, we use the variables D1 and D2 in the models. D1 takes the value 1, if the real value is lower than the inferior limit of the forecast interval and it takes the value zero for the rest of the cases. D2 is 1 if the real value is greater than the superior limit and it is zero for all the other cases. The following models were estimated by the coefficientsŌ bootstraping (1000 replications). For these models we are interested only in the value of the standard error of the regression:

$$ri_t = 4.4507 + 3.834 \cdot D1(li_t - ri_t) + 0.901 \cdot D2(ri_t - ls_t)$$

(the standard error of the regression for the intervals based on RMSE is 2.014456)

$$ri_t = 5.1279 - 0.1497 \cdot D1(li_t - ri_t) + 0.981 \cdot D2(ri_t - ls_t)$$

(the standard error of the regression for the intervals based on BCA bootstrap procedure is 2.412078)

ri_t- rate of effective/real inflation for quarter t

li_t- inferior limit of the forecast interval for quarter t

ls_t – superior limit of the forecast interval for quarter t

The coefficients of error are computed using the standard error of the regression that is divided to the average of the quarter inflation rates on the forecasting horizon 2010-2012. A coefficient of error of 3.968% was gotten for the intervals based on RMSE and a coefficient of 4.751% for the intervals based on bootstraping procedure, fact that implies the superiority of the prediction intervals based on the first method.

However, it is important to compute a measure of global uncertainty or an indicator of global accuracy for all the determined intervals, even if tese contain or not the real value.

Therefore, we proposed a new indicator, called M indicator, that is computed as a sum of errors for two cases: when the real value is not the interval and when the real value is in the forecast interval. For the first case, it is calculated the root mean squared of the deviations between the effective value and the inferior limit (if the real value is lower than the inferior limit) and the difference between the real value and the superior limit (if the real value is greater than the upper limit). This root mean squared of the deviations can be considered a modified RMSE, because the reporting isnot done according to a certain limit of the intervals (inferior or superior limit), but in a variable way so as to have a minimum distance betwwen the real value and a limit. This indicator was denoted by RMSE*.
In order to get an indicator as coefficient of variation, this REMSE* is divided by the deviationsÔaverage (errors average). For the second case, when the effective value is placed in the interval, the root squared average of the deviations using the minimum of the distance between the inferior limit and the real value adn respectively, the difference between the supeior limit and the real value. This squared mean deviation is denoted by RMSE** and it is divided by the average of minimum distances. Using the previous explications, the following formula can be utilized in order to compute the indicator M, as a measure of global accuracy of the forecast intervals:

$$M = \frac{RMSE*}{media\ abaterilor\ minime1} + \frac{RMSE**}{media\ abaterilor\ minime2} \qquad (2)$$

If the indicator M for intervals on a horizon is less that the value for intervals based on another method, the first procedure gives better results. The M measure is approximately 2.146 for intervals based on RMSE and it is 1.738 for intervals based on BCA bootstrap method. Therefore, the forecast intervals based on the second method are better than those based on historical errors. The problem that appears here is related to the explications for the diffrences in conclusions when more ways of comparison are applied. An initial evaluation shows that more values are placed in the intervals based on RMSE. On the other hand, even if less values are placed in the intervals based on the second method, these intervals are smaller and the real deviations are in average smaller than in the case of the first type of forecast intervals. The limit of the M indicator proposed by me is that not all the values in the interval are taken into account in order to compute it, but only some specific values (the limits or the center of the interval). Actually, an infinite values are in an interval. If we work with the assumption of a normal distribution, we can take into account the intervalsÔcenters, but our intervals are not symmetrical and the hypothesis of a normal repartion is not checked.
So, the intervals lenght can be considered an important source of uncertainty in forecasting. Higher the intervalĜ lenght is, more the uncertainty increases.
In conclusion, there are two approaches in determining the best method for constructing forecast intervals: a simple approach, that neglects the lenght of the intervals of source of uncertainty and that considers correct in the same degree two intervals that includes the real value but that have different lenghts (the accuracy indicators in this case are the number of correct intervals, their weights in the total and the coefficient of error associated to regression models based on DUMMY variables) and a global approach that includes also the uncertainty associated to the lenghts of forecast intervals.

We proposed other measures for the precision of the prediction intervals. We can determine the differences between the registered value of inflation rate in each quarter and the lower limit/ uper limit/ center of each prediction interval, the differences being denoted as d1, d2 and d3. we can compute also the averages of these differences or the averages of the absolute differences on the forecasting horizon.

Table 3

The computation of average indicators of differences for forecast intervals of the quarterly inflation rate (horizon 2010-2012)

The average differences of the milits/centers deviations compared to the real value	**Intervals based on historical RMSE**	**Intervals based on BCA bootstrapping**
$\overline{d1}$	2.13149	**0.26197**
$\overline{d2}$	-2.1558	**-0.5210**
$\overline{d3}$	**-0.0121**	-0.1296
$\overline{\lvert d1 \rvert}$	2.2776	**0.7505**
$\overline{\lvert d2 \rvert}$	2.9214	**0.6981**
$\overline{\lvert d3 \rvert}$	1.4363	**0.6280**

Source: own computations

The average differences of the deviations for the inflation rate compared to inferior limits, superior limits and centers of the intervals have lower values for the intervals based on BCA bootstrap technique. Only the average difference based on the centers of the intervals are lower for the intervals based on RMSE.

The forecast intervals can also be computed using the bayesian theory that admits different degrees of uncertainty generated by the quantity of information that is known. We started from the data based on surveys that are published by SPF (Survey of Professional Forecasters) for quarterly inflation rate in USA. For each quarter the respondents propose a certain predicted value for the inflation rate, the number of values for each quarter being different. For the first quarter of 2012, 45 responses were registered and for the next quarters 39, 48 and respectively 39 responses. The parameter for which the forecast intervals are computed is the average of the quarterlu inflation rate. The averages based on the samples from each quarter, the selection dispersions and the intervals computed for the average and guaranteed with 95% are daplied in Table.

In bayesian theory terms, we can assess in what degree the aprioristic information increases the accuracy of the estimation in order to get a smaller interval. We considered first of all the first the case of total uncertainty (stage 1). We selected then 5 values in the sample of responses, for which the average, the dispersion and the forecast interval are computed. This is the first measurement of the uncertainty (stage 2). 20 values are selected from the samples and the forecast intervals are computed, the lenght being smaller (stage 3).

The function of revised density function is computed using the two types of information related to the average. A weighted average is computed, the weights being the inverses of the dispersions. The information with a higher degree of accuracy, that has a lower dispersion, receives a higher weight. The last interval that combines the information of the last two intervals corresponds to the stage 4 of the proposed bayesian approach.

Table 4

Bayesian forecast intervals for quarterly inflation rate (%) in USA in 2012 using the SPF appreciations

Time period	Selection average	Selection dispersion	Bayesian forecast intervals	Weights (g1 and g2)
STAGE 1				
2012: Q1	0.8711	0.00083	0.8144- 0.9277	
2012: Q2	2.5386	0.00098	2.4772- 2.6001	
2012: Q3	0.7846	0.0017	0.7032- 0.8661	
2012: Q4	2.2615	0.000386	2.2230- 2.3001	
STAGE 2				
2012: Q1	0.64	0.26	(-2.136)- (-1.496)	3.846154
2012: Q2	2.737635437	0.237635437	*2.0779- 3.3973	4.208127
2012: Q3	0.472912175	0.327087825	(-0.4351)- (+1.3809)	3.057283
2012: Q4	2.24	0.06	*2.0734- 2.4065	16.66667
STAGE 3				
2012: Q1	0.835	0.065	0.6989- 0.9710*	15.38462
2012: Q2	2.575408859	0.060707129	2.4483- 2.7024	16.47253
2012: Q3	0.763228044	0.101016122	0.5518- 0.9746*	9.89941
2012: Q4	2.225	0.036903145	2.1477- 2.3022	27.09796
STAGE 4				
2012: Q1	0.7960	0.052	0.6941- 0.8979	
2012: Q2	2.6084	0.048354	2.6136- 2.7031	
2012: Q3	0.6947	0.07718	0.5434- 0.8459	
2012: Q4	2.2307	0.02285	2.1859- 2.2754	

Source: own sources

The necessary weights to construct the intervals in the last stage (g1 and g2) are computed as the inverses of dispersions in stage 2 and stage 3. The average is computed as arithmetic mean of the averages in stage 2 and 3, weighted with g1 and g2 and the dispersions are the inverse of the sum of weights (1/(g1+g2)).

One could observe that the intervals in the last stage are smaller than those based on aprioristic information or those based on SPF expertsÕ appreciations. The real registered values of the inflation rates in 2012 are: 2.5%, 0.7%, 1.6%, respectively 1.2%, fact tat shows that none of the rates are located in the specified intervals. We used an asterisk to show the intervals for which the effective values are closer of the inferior limit (the intervals for quarters 2 and 4 in the second stage) and, respectively, of the superior limit (intervals for quarters 1 and 3 computed in the third stage).

IV. Conclusions

In this article some improvements were brought by making comparisons regarding the accuracy between intervals based on different methods. We introduced the indicator M that solves two problems related to some sources of uncertainty: the lenght of the intervals and the location of the real value compared to the limits. The values of indicator M recommend the intervals based on BCA bootstrap procedure as better than those based on historical error method for the quarterly inflation rate provided by national Bank of Romania on the horizon 2010-2012. If we take into account only the criterion of appurtenance of the real value to the forecast interval essential information is omitted and the intervals based on RMSE are incorrectly chosen as the best. The value of M indicator recommends the intervals based on BCA bootstrap technique as superior to those based on the historical RMSE measures for the quarterly NBR inflation rate on 2010-2012.

Forecast intervals based on bayesian approach were built for the quarterly inflation in USA in order to reduce the lenght of the intervals. The comparisons put in evidence an increase of uncertainty degree because the real values were not placed in these narrow intervals.

References

Anderson, P.L. Meerschaert, M.M. and Zhang, K., 2012. Forecasting with prediction intervals for periodic autoregressive moving average models. *Journal of Time Series Analysis,* 34(2), pp.187-193.

BorbŽly, D. Meier, C.P., 2003. Macroeconomic interval forecasting: the case of assessing the risk of deflation in Germany. *Kieler Arbeitspapiere*, No.1153.

Chatfield, C., 1993. Calculating interval forecasts. *Journal of Business and Economic Statistics,* 11(3), pp.121-144.

Chatfield, C., 2001. Prediction intervals for time series in Principles of Forecasting: A Handbook for Practitioners and Researchers, Armstrong, J.S. (ed.), Norwall, MA: Kluwer Academic Publishers, 475-494.

Christoffersen, P., 2004. Evaluating Interval Forecasts. *International Economic Review*, 39(4). Symposium on Forecasting and Empirical Methods in Macroeconomics and Finance.

Cogley, T. Morozov and Sargent, T., 2005. Bayesian Prediction Intervals in Evolving Monetary Systems, *Journal of Economic Dynamics and Control*, 29(1), pp.56-63.

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

Demetrescu, M., 2007. Optimal forecast intervals under asymmetric loss. *Journal of Forecast*ing, 26(3), pp.227Ď238. doi: 10.1002/for.1019.

Diebold, F.X. Gunther, T.A. and Tay, A.S., 1998. Evaluating Density Forecasts with Applications to Financial Risk Management. International Economic Review, volume 39, No. 4, Symposium on Forecasting and Empirical Methods in Macroeconomics and Finance (Nov., 1998), 863-883.

Efron, B. and Tibshirani, R., 1993. An introduction to the bootstrap, CRC Press, 51-60.

Hansen, B., 2005. Interval Forecasts and Parameter Uncertainty. *Journal of Econometrics,* 135, pp.377-398.

Kjellberg, D. and Villani, M., 2010. The RiksbankŐs communication of macroeconomic uncertainty. *Economic Review,* 1, pp.2-49.

Knźppe, I.M. and Schultefrankenfeld, G., 2008. How informative are macroeconomic risk forecasts? An examination of the Bank of EnglandŐs inflation forecasts, *Discussion Paper, Series 1: Economic Studies,* 1(14), pp.3-44.

Fair, R. 2000. Fed Policy and the Effects of a Stock Market Crash on the Economy, *Business Economics*, April 2000, 7-14.

Fischer, H. Garc'a-Bçrzana, M. Tillmann, P. and Winker, P., 2012. Evaluating FOMC forecast ranges: an interval data approach. *MAGKS Papers on Economics* 201213, Philipps-UniversitŠt Marburg, Faculty of Business Administration and Economics, Department of Economics (Volkswirtschaftliche Abteilung).

Tulip, P. Wallance, S., 2012. Estimates of Uncertainty around the RBA's Forecasts. Reserve Bank of Australia, *Research Discussion Paper*, 2012-07.

Wallis, K.F., 2003. Chi-squared tests of interval and density forecasts and the Bank of England's fan charts. *International Journal of Forecasting,* 19, pp.165Ď175.

Wang, J. and Wu, J., 2012. The Taylor Rule and Forecast Intervals for Exchange Rates. *Journal of Money, Credit and Banking*, 44, pp.103Ď144.