Macrothink Institute™

Business and Economic Research
ISSN 2162-4860
2013, Vol. 3, No. 1

New Methods of Evaluating the Forecasts Accuracy: A Case Study for USA Inflation

Mihaela Bratu (Simionescu) (Corresponding author)

Dept. of Statistics and Econometrics, Faculty of Cybernetics, Statistics and Economic Informatics, Academy of Economic Studies, Bucharest, Romania

E-mail: mihaela_mb1@yahoo.com

Received: December 21, 2012 Accepted: January 4, 2013

doi:10.5296/ber.v3i1.3103 URL: http://dx.doi.org/10.5296/ber.v3i1.3103

Abstract

Using the two-year inflation forecasts provided by CBO, Blue Chips and Administration for USA on the forecasting horizon 1982-2011, the accuracy of forecasts was assessed. According to U1 Theil's statistic, the CBO projections are the best, followed by Administration and Blue Chips predictions. A new accuracy measure is proposed to be introduce in literature (ratio of radicals of sum of squared errors) in order to compare the forecasts with the naïve ones. The same hierarchy of institutions, according to accuracy criterion is obtained if all the computed accuracy indicators are taking into account using ranks method. According to the relative distance method with respect to the better institution and other methods (binary logistic regression and some non-parametric tests (Wilcoxon and Kruskall-Wallis tests)) the following rank is gotten: Administration, CBO and Blue Chips. All these methods (multi-criteria ranking, logistic regression and non-parametric tests) were not mentioned before in literature as possible ways of comparing the forecasts accuracy, but most of them gave better results than the classical U1, because these methods take into consideration more aspects regarding the accuracy measurement. Some empirical strategies of improving the forecasts accuracy were applied (combined forecasts, smoothed predicted values based on Holt-Winters technique, Hodrick-Prescott, Baxter King and Christiano-Fitzegerald filters), getting more accurate predictions only for the combined forecasts of Blue Chips and Administration using inverse MSE scheme (the highest improvement) and equally weighted scheme, but also for forecasts based on CBO and Blue Chips using optimal scheme.

Keywords: forecasts, Accuracy, Logistic regression, Multi-criteria ranking, Non-parametric tests

1. Introduction

The assessing of macroeconomic forecasts accuracy became a more important problem, especially in the context of alternative predictions offered by more institution and in the context of the economic crisis. If the government has to take decisions and establish policies when there are many economic problems, it has to reduce as much as possible the uncertainty of forecasts used in the decisional process by choosing the most accurate ones on the historical forecasting horizon. The problem is statistically solved by making comparisons between predictions. There are classical measures of accuracy for comparisons, like famous U Theil's statistic, but some other methods are proposed in this article. As alternative to U Theilț's statistic we proposed a new indicator that was denoted by " ratio of radicals of sum of squared errors". These are common statistical means that were not used before in literature in the context of assessing the predictions accuracy: multi-criteria ranking, logistic regression and non-parametric tests

2. New Methods Used in Comparing the Forecasts Accuracy: Multi-Criteria Ranking, Nonparametric Tests and Logistic Regression

Most international institutions provide their own macroeconomic forecasts. It is interesting that many researchers compare the predictions of those institutions (Melander for European Commission, Vogel for OECD, Timmermann for IMF) with registered values and those of other international organizations, but it is omitted the comparison with official predictions of government.

Abreu (2011) evaluated the performance of macroeconomic forecasts made by IMF, European Commission and OECD and two private institutions (Consensus Economics and The Economist). The author analized the directional accuracy and the ability of predicting an eventual economic crisis.

Bratu (2012 a) evaluated the accuracy of some macroeocnomic predictions for Romania made by the Institute of Economic Forecasting and the National Commission of Prognosis, the second institution getting more accurate forecasts for: inflation, unemployment, GDP deflator, export rate and exchange rate on the horizon 2004-2011.

Novotny F. & Rakova M.D. (2012) assessed the accuracy of macroeconomic forecasts made by Consensus for the Czech Republic, observing an improvement in accuracy from a year to another on the horizon 1994-2009. The authors also proposed a regression for comparing the predictions.

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

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

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

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

To make comparisons between forecasts we propose a new methodology that was not mentioned before in literature: multi-criteria ranking, logistic regression and non-parametric tests.

Two methods of multi-criteria ranking (ranks method and the method of relative distance with respect to the maximal performance) are utilized in order to select the institution that provided the best forecasts on the horizon 1982-2011 taking into account at the same time all computed measures of accuracy.

If we consider $\hat{X}_{t(k)}$ the predicted value after k periods from the origin time t, then the error at future time (t+k) is: $e_t(t+k)$. This is the difference between the registered value and the predicted one.

The indicators for evaluating the forecasts accuracy that will be taken into consideration when the multi-criteria ranking is used are:

- Root Mean Squared Error (RMSE)

$$RMSE = \sqrt{\frac{1}{n}\sum_{j=1}^{n} e_X^2(T_0 + j, k)}$$

- Mean error (ME)

$$ME = \frac{1}{n}\sum_{j=1}^{n} e_X(T_0 + j, k)$$

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

- Mean absolute error (MAE)

$$MAE = \frac{1}{n}\sum_{j=1}^{n} \left| e_X(T_0 + j, k) \right|$$

These measures of accuracy have some disadvantages. For example, RMSE is affected by outliers. Armstrong and Collopy stresses that these measures are not independent of the unit of measurement, unless if they are expressed as percentage. If we have two forecasts with the same mean absolute error, RMSE penalizes the one with the biggest errors.

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

U Theil's statistic can be computed in two variants, specified also by the Australian Tresorery.

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

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

A value close to zero for U_1 implies a higher accuracy.

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

If U_2 =1=> there are not differences in terms of accuracy between the two forecasts to compare

If U_2 <1=> the forecast to compare has a higher degree of accuracy than the naive one

If U_2 >1=> the forecast to compare has a lower degree of accuracy than the naive one

For comparisons with the naive forecasts a new indicator is proposed and computed: ratio of radicals of sum of squared errors (RRSSE)= $\frac{\sqrt[n]{\sum_{t=1}^{n} e_t^2}}{\sqrt[n-1]{\sum_{t=2}^{n}(x_t - x_{t-1})^2}}$.

Ranks 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 four institutions that made forecasts. The rank for each institution is denoted by: $(r_{i_{ind_j}})$, i=1,2,3,4 and ind_j –accuracy indicator j. We chose 5 indicators: mean error, mean absolute error, root mean squared error, U1 and U2.

2. If the ranks assigned to each institution are sum up, the score to each of them is computed.

$$S_{i=\sum_{j=1}^{5}(r_{i_{ind_j}})}, \text{ i=1,2,3,4}$$

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

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 (institution) with respect to the one with the best performance is computed. The distance is calculated as a relative indicator of coordination:

$$d_{i_{ind_j}}=\frac{ind_i^j}{\{\min abs(ind_i^j)_{i=1,..,4}}, \text{ i=1,2,3,4 and j=1,2,..,5}$$

The relative distance computed for each institution is a ratio, where the denominator is the best value for the accuracy indicator for all institutions.

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

$$\overline{d_i}=\sqrt[5]{\prod_{j=1}^{5} d_{i_{ind_j}}}, \quad \text{i=1,2,3,4}$$

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

$$loc_i^{\%}=\frac{\overline{d_i}}{\min(d_i)_{i=\overline{1,4}}}\cdot 100$$

Wilcoxon Signed Ranks test or other nonparametric tests could be used when the series distribution is not normal or when the type of repartition is not known. These parametric procedures give good results regarding differences between populations when the assumptions regarding the population are not fulfilled.

For our sample we do not know the shape of distribution, the appliance of nonparametric tests being the best choice.

Wilcoxon Signed Ranks test is used to compare the means of two populations or to make inferences about the population mean. It is based on a random selection and a symmetric distribution. The variable is measured on a scale that is at least an ordinal one. The hypothesis set supposes that the null assumption (H0) implies the equality between the medians of the two samples ($\theta = \theta_0$). If H0 is accepted, then the samples come from the same population. The alternative hypothesis supposes that there are differences between samples medians ($\theta \neq \theta_0, \theta > \theta_0$ or $\theta < \theta_0$).

Using the sample with "n" observations (x1,x2,…,xn), the difference scores are computed: $D_i = x_i - \theta_0$, $i = 1,2,\dots,n$.

The difference scores data series is ascending ordered and ranks are assigned to $|D_i|$, $i = 1,2,\dots,n$. Null scores are removed from analysis.

We denoted by W the sum of the positive ranks. The test statistic is:

$$Z = \frac{W - \frac{n(n+1)}{4}}{\sqrt{\frac{n(n+1)(2n+1)}{24}}}$$

If the p-value corresponding to this statistic is less than 0.05, then the H0 is rejected. So, the samples come from different populations or there are statistically significant differences between the two samples (or the two variables in the analysis).

In our research all the conclusions drawn using the data from our survey are guaranteed with a probability of 95% (a significance level of 0.05).

Kruskall-Wallis test is used to compare two or more populations. The null hypothesis tests that the medians of the populations are equal. If at least two medians are different the alternative hypothesis is accepted. Actually, in this case, we do not have reasons to accept the null hypothesis. The interest variable is not necessarily numerical one and the factors that may affect it are independent. The assumptions of normality and homoscedasticity are not checked when the test is applied. If the statistic test is lower than 0.05, the null hypothesis is rejected and the alternative one is accepted. So, there are differences between the populations or the interest variable does not depend on the specified factors.

For samples of low volume you may wish to refer to tables of the statistic but the chi-square approximation is better in most cases than Kruskal-Wallis test, according to Conover (1999).

Another method of comparing forecasts accuracy is based on logistic regression.

Logistic regression measured the impact of more independent (exogenous) characteristics that appear simultaneously in order to predict membership of one or other category of the two

dependent characteristics. The dependent variable is categorical and the exogenous ones are categorical or a mix of continuous and categorical.

There some important advantages of the logistic or multinomial regression: some assumptions are not taken into consideration (the errors non-correlation, the normality or homoscedasticity of the independent variables). However, the maximum likelihood estimation method is used instead of ordinary least squares.

The dependent variable takes values between 0 and 1 and the independent one (X) takes real values. With p is denoted the probability that a case is in a certain category. The odds ratio of an event (likelihood ratio) are p/(1-p). The log of the odds ratio is $ln\frac{p}{1-p} = b_0 + b_1X + \varepsilon$. The parameters that should be estimated are b_0 *and* b_1. The error is denoted by ε.

The p is determined as: $p = \frac{e^{b_0+b_1X+\varepsilon}}{1+e^{b_0+b_1X+\varepsilon}}$. The estimated of the parameters are: $\widehat{b_0}$ *and* $\widehat{b_1}$.

An odds ratio (OR) equaled to one shows that if X increase by an unit, the odds remain the same. In other words, X does not influence the dependent variable Y.

An OR higher than 1 implies the following: an increase by one unit in the exogenous variable determines a growth by $e^{\widehat{b_1}}$ in the level of the dependent variable. If OR is lower than 1, we will have a decrease by $e^{\widehat{b_1}}$ in the dependent variable.

We calculate the absolute errors for each year in the forecasting horizon. We test is each error differs significantly from a fixed value. We choose a threshold of 0.5%. We test each predicted value to establish if it differs significantly from the threshold. A binary variable is created:

error_significance= $\begin{cases} 1, & significant\ error \\ 0, & non-significant\ error \end{cases}$

A logistic regression is built starting from this binary variable and the predictions made by an institution. For each forecasted value of the inflation rate we associated the significance of the error.

3. Comparisons between Inflation Forecasts Made By CBO, Blue Chips and Administration

In this study we used the forecasted values of two-year average inflation in the Consumer Price Index made for USA by CBO, Blue Chips and Administration. The forecasting horizon is 1976-2011.

Armstrong and Fildes (1995) showed that it is not sufficient to use a single measure of accuracy. Therefore, more accuracy indicators were computed for the three types of forecasts on the specified horizon.

Table 1. The accuracy of forecasts made by CBO, Blue Chips and Administration for two-year average inflation in the Consumer Price Index (1982-2011)

Accuracy indicator	CBO	Blue Chips	Administration
ME	-0.19667	-0.28	-0.05
MAE	0.63	0.713333	0.636667
RMSE	0.828855	0.876356	0.818332
U1	0.125171	0.131009	0.126933
RRSSE	0.8475	1.02026	0.8583
U2	0.951844	1.043638	0.98383

Author's computations using Excel

The best predictions are provided by CBO, according to U1 value, but for some accuracy indicators Administration managed to provide smaller errors (ME and RMSE). The CBO's and Administration's predictions are better than the naïve ones, according to U1 and to the new indicator proposed by us (RRSSE).

The two methods of multi-criteria ranking are applied in order to take into account all the values of the computed accuracy indicators.

Table 2. The ranks of institutions according to the accuracy measures (ranks method)

Criteria	CBO	Blue Chips	Administration
ME	2	3	1
MAE	1	3	2
RMSE	2	3	1
U1	1	3	2
U2	1	3	2
Sum of ranks	7	15	8
Final ranks	1	3	2

Author's computations using Excel

According to ranks method, the best predictions belong to CBO and the less accurate to Blue Chips. This method provided the same hierarchy as U1 values.

Table 3. The ranks of institutions according to the accuracy measures (method of relative distance with respect to the best institution)

Criteria	CBO	Blue Chips	Administration
ME	3.9334	5.6	1
MAE	1	1.132275	1.010583
RMSE	1.012859	1.070905	1
U1	1	1.032111	1.014077
U2	1	1.096438	1.033604
Average relative distance	1.318449	1.503558	1.011578
Ranks	2	3	1
Location (%)	130.34%	148.63%	100.00%

Author's computations using Excel

This method is better than the ranks one and it shows that Administration provided the most accurate predictions. It is certain that Blue Chips forecasts are less accurate. In literature the method of relative distance is considered better than the ranks one.

First of all we can check the dependencies between the effective values and the predicted ones using non-parametric tests.

The assumptions of the tests are:

H0 (null hypothesis): there are not significant differences between the predicted values and the registered ones

H1 (alternative hypothesis): there are significant differences between the predicted values and the registered ones

A value greater than 0.05 for the significance value or p-value (when the probability of guarantee the results is 95%) implies the acceptance of null hypothesis. The higher the value is greater than 0.05 the more insignificant are the differences between predictions and real values on inflation and consequently the forecasts accuracy is higher.

The non-parametric tests show the following with a probability of 95%:

- There are not significant differences between CBO forecasts and the registered values;
- There are not significant differences between Blue Chips forecasts and the registered values, but the Significance indicator is smaller; this implies that CBO predictions are better than the Blue Chips ones;
- There are not significant differences between Administration forecasts and the registered values, but the Significance indicator value is higher than the Significance values of the other two predictions; this implies that Administration expectations are better than CBO ones and even better than Blue Chips results. The results of non-parametric tests applied in SAS are presented in **Appendix 1**.

So, the hierarchy given by the application of non-parametric tests is: Administration, CBO and Blue Chips.

The regression equation is given by the table from SPSS outputs called: "Variables in the Equation". The tables from SPSS are presented in **Appendix 2**. The odds of having a non-significant error for CBO forecasts increase with 63.6%, while the chances for Blue Chips and Administration grow with 37.8%, respectively, 142.6%. So, the best forecasts are provided by Administration. Blue Chips predictions are less accurate than all the other estimations. The results of binary logistic regression are the same as those provided by relative distance method.

4. Strategies to Improve the Forecasts Accuracy

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

The combined forecasts are another possible strategy of getting more accurate predictions. 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 f1;t and f2;t, for the same variable Xt, 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: $c_t = m \cdot f_{1t} + (1-m) \cdot f_{2t}$. The error of the combined forecast is: $e_{c,t} = m \cdot e_{1t} + (1-m) \cdot e_{2t}$. The mean of the combined forecast is zero and the variance is:

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

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

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

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

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

The U Theil's statistics were computed for the combined forecasts based on the three schemes, the results being shown in the following table (Table):

Table 4. The accuracy of combined forecasts for two-year average inflation in the Consumer Price Index (1982-2011)

Accuracy indicator	CBO+ Blue Chips	CBO+ Administration	Blue Chips + Administration
U1 (optimal scheme)	0.123573	0.138802	*0.129965*
U2 (optimal scheme)	1.083943	1.122845	1.073063
U1 (inverse MSE scheme)	*0.126368*	*0.126938*	0.117619
U2 (inverse MSE scheme)	1.048741	1.141105	1.307234
U1 (equally weighted scheme)	0.125982	*0.128894*	0.119605
U2 (equally weighted scheme)	0.985934	1.047126	1.177012

Author's computations using Excel

Only the combined forecasts based on CBO expectations and Blue Chips ones using EW scheme are better than the naïve predictions. Improvements in the accuracy were brought by the combined forecasts based on the following schemes:

- ✓ Optimal scheme: CBO+ Blue Chips predictions;
- ✓ Inverse MSE scheme: Blue Chips+ Administration predictions;
- ✓ Equally weighted scheme: Blue Chips+ Administration predictions.

Some combined forecasts (Blue Chips+ Administration predictions using OPT scheme, CBO+ Blue Chips predictions, CBO+ Administration forecasts using INV scheme, CBO+ Administration forecasts using EW scheme) are better than the predictions made by Blue Chips.

Another technique used by Bratu (Simionescu) (2013) to improve the forecasts accuracy is the application of filters to the predicted data. The author recommends also the use of exponential smoothing methods like Holts Winters.

The Hodrick–Prescott (HP) filter is very used in macroeconomics to extract the trend of the data series and separate the cyclical component of the time series. The smoothed data gotten are more sensitive to long term changes.

The initial data series is composed of trend and cyclical component:

$$inf_t = tr_t + c_t.$$

Hodrick & Prescott (1997) suggest the solve of the minimization problem:

$$\min_{\{tr_t\}t=1,T} \sum_{t=1}^{T}(inf_t - tr_t)^2 + \gamma \sum_{t=2}^{T-1}(\nabla^2 tr_{t+1})^2$$

γ- penalty parameter

The solution to the above equation can be written, according to Hyeongwoo as:

$$inf_t = (\gamma F + I_T) \cdot tr_T$$

inf_t- vector of the initial data series of the inflation rate

$$F=\begin{bmatrix} 1 & -2 & 1 & 0 & \dots & & & & & & \\ -2 & 5 & -4 & 1 & 0 & \dots & \cdots & & & & 0 \\ 1 & -4 & 6 & -4 & 1 & 0 & & & & & \\ & & \vdots & & & & \ddots & & \vdots & & \\ & & & & & & \cdots & & & & \dots \\ & & & & & & & 0 & 1 & -4 & 6 & -4 & 1 & 0 \\ & & & & & & & 0 & 0 & 1 & -4 & 6 & -4 & 1 \\ 0 & & 0 & & & & \dots & & 0 & 0 & 1 & -4 & 5 & -2 \end{bmatrix}$$

The trend is calculated as: $tr_T = [(\gamma \cdot F + I_T)]^{-1} \cdot inf_T$.

Razzak (1997) proved that the Hodrick-Prescott filter acts as true `filter' at the end of the

sample and as a `smoother' over the sample. The output gap from the true filter generates better out-of-sample predictions of inflation.

Christiano &Fitzgerald (2003) explained that Band-Pass filter is used to determine that component of the chronological series that is situated within a specific band of frequencies. (Baxter & King, 1995) built a bandpass filter of order K, where K-finite. If the analyzed time series is a random walk, its spectrum of a Band- Pass filter is:

$$f(\omega) = \left|\alpha^{I(1)}(\omega)\right|^2 \cdot \frac{\sigma^2}{2\pi}, where\ \omega\varepsilon[-\pi;\pi]$$

$$\left|\alpha^{I(1)}(\omega)\right|^2 = \left(2 \cdot \sin(\frac{\omega}{2}) \cdot \sum_{j=1}^{K} a_j \cdot \sum_{h=-(j-1)}^{j-1} (j-|h|) \cdot \cos(\omega h)\right)^2$$

The peak that shows a spurious cycle is smaller in case of a Band Pass filter in comparison with the Hodrick-Prescott one.

$\left|\alpha^{I(1)}(\omega)\right|^2$ is the ptf of the filter.

Christiano-Fitzegerald filter (CF filter) is an asymmetric one and it converges on long run to an optimal filter. It has a steep frequency response function at the limits of the band. The CF filter is computed, according to Christiano &Fitzgerald (2003), as:

$$c_t = B_0 \cdot inf_t + B_1 \cdot inf_{t+1} + \cdots + B_{T-1-t} \cdot inf_{T-1} + B_1 \cdot inf_{t-1} + \cdots + B_{t-2} \cdot inf_2 + B_{t-1} \cdot inf_1$$

$B_0 = \frac{b-a}{\pi}$, a=$\frac{2\pi}{p_u}$, b=$\frac{2\pi}{p_l}$

p_u, p_l- parameters that are cut-off cycle length in month

c- cycle term

$B_j = \sin(jb) - \sin(ja), j \geq 1$

$$B_k = -\frac{1}{2} \cdot B_0 - \sum_{j=1}^{k-1} B_j$$

Holt-Winters Simple exponential smoothing method is recommended for data series with linear trend and without seasonal variations, the forecast being determined as:

$$\inf_{n+k} = a + b \times k\ .$$

$$a_n = \alpha \times \inf_n + (1-\alpha) \times (a_{n-1} + b_{n-1})$$

$$b_n = \beta \cdot (a_n - a_{n-1}) + (1-\beta) \cdot b_{n-1}$$

Finally, the prediction value on horizon k is:

$$\hat{\inf}_{n+k} = \hat{a}_n + \hat{b}_n \times k$$

Table 5. The accuracy of smoothed forecasts for two-year average inflation in the Consumer Price Index (1982-2011)

Accuracy indicator	CBO	Blue Chips	Administration
U1 (Holt-Winters technique)	*0.126021*	0.131898	*0.128072*
U2 ((Holt-Winters technique)	1.209814	1.243206	1.249674
U1 (Hodrick-Prescott filter)	*0.125967*	0.131755	*0.127681*
U2 (Hodrick-Prescott filter)	1.088852	1.118875	1.002278
U1 (Baxter King filter)	0.238421	0.252416	0.236235
U2 (Baxter King filter)	4.293729	4.292725	4.312521
U1 (Christiano-Fitzegerald filter)	0.200792	0.210874	0.199279
U2 (Christiano-Fitzegerald filter)	3.023635	2.986817	3.104637

Author's computations using Excel

The smoothed values are not more accurate than the corresponding predictions. The smoothed CBO and Administration forecasts are still better than Blue Chips expectations when techniques are applied (Holt-Winters method and HP filter).

5. Conclusion

This research brings into attention the problem of improving the assessment of forecasts accuracy, giving also some possible strategies of getting more accurate predictions.

According to U1 statistic and ranks method, the hierarchy of institutions that forecasted between the two-year inflation in 1982-2011 is: CBO, Administration and Blue Chips. The relative distance method with respect to the better institution, the logistic regression, the non-parametric tests provided the following ranking: Administration, CBO and Blue Chips. The highest improvement in accuracy was brought by the combined forecasts of Blue Chips and Administration using inverse MSE scheme. The smoothed predicted values based on Holt-Winters technique, Hodrick-Prescott, Baxter King and Christiano-Fitzegerald filters din not improve the forecasts accuracy.

The novelty of this research is given by the application of some statistic methods to compare the predictions accuracy, these methods never being mentioned in literature in this context. The results of the new approach are better than those provided by the U Theil's statistic, because more aspects of accuracy problem are taken into account.

References

Abreu I. (2011). *International organizations' vs. private analysts' forecasts: an Evaluation.* http://www.bportugal.pt/en-US/BdP%20Publications%20Research/wp201120.pdf(October 27, 2012)

Bates, J., & Granger, C. W. J. (1969). The Combination of Forecasts. *Operations Research Quarterly*, *20*, 451-468. http://dx.doi.org/10.1057/jors.1969.103

Bratu, M. (2012a). Macroeconomic Forecasts Accuracy in Romania. *Review of Economic Studies and Research Virgil Madgearu*, *2*, 99-120.

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

Bratu (Simionescu) Mihaela (2013). Filters or Holt Winters technique to improve the forecasts for USA inflation rate *?*. *Acta Universitatis Danubius. Œconomica*, 9, 1/2013 (forthcoming)

Christiano, L. J. & Fitzgerald, T. J. (2003). The Band Pass Filter, *International Economic Review, 44*, 435-465. http://dx.doi.org/10.1111/1468-2354.t01-1-00076

Clark T.E. & Ravazzolo F. (2012). The Macroeconomic Forecasting Performance of Autoregressive Models with Alternative Specifications of Time-Varying Volatility. *FRB of Cleveland, 3*, 12-18

Clements M. P.& Galvao A. B. (2012). Macroeconomic Forecasting with Mixed Frequency Data: Forecasting US output growth and inflation. *The Warwick Economics Research Paper Series (TWERPS) from University of Warwick, Department of Economics*, *773*, 10-29.

Conover W. J. (1999). *Practical nonparametric statistics*, Wiley series in probability and statistics, New York: Wiley.

Deschamps, B. and Bianchi, P. (2012), An evaluation of Chinese macroeconomic forecasts, *Journal of Chinese Economics and Business Studies*, *10*, 229-246. http://dx.doi.org/10.1080/14765284.2012.699704

Genrea V., Kenny G., Meylera A. & Timmermann A. (2013). Combining expert forecasts: Can anything beat the simple average?. *International Journal of Forecasting*, *29*, 108–121. http://dx.doi.org/10.1016/j.ijforecast.2012.06.004

Hodrick R. & Prescott E. C. (1997). Postwar U.S. Business Cycles: An empirical investigation. *Journal of Money, Credit and Banking*, *16*, 84-90.

Novotny F. & Rakova M. D. (2012). Assessment of Consensus forecasts accuracy: The Czech National Bank perspective. *Economic Research Bulletin*, *10*, 10-13.

Razzak W. (1997). The Hodrick-Prescott technique: A smoother versus a filter: An application to New Zealand GDP. *Economics Letters*, 57, 163–168. http://dx.doi.org/10.1016/S0165-1765(97)00178-X

Appendix

Appendix 1. The results of non-parametric tests in SAS

The NPAR1WAY Procedure

Wilcoxon Scores (Rank Sums) for Variable cbo

Classified by Variable real

real	N	Sum of Scores	Expected Under H0	Std Dev Under H0	Mean Score
4.6	1	30.00	15.50	8.638094	30.000000
3.8	2	48.50	31.00	12.003639	24.250000
3.9	1	28.50	15.50	8.638094	28.500000
2.7	2	37.50	31.00	12.003639	18.750000
2.8	3	46.00	46.50	14.436484	15.333333
4.4	1	26.50	15.50	8.638094	26.500000
5.1	1	28.50	15.50	8.638094	28.500000
4.8	1	23.50	15.50	8.638094	23.500000
3.6	1	25.00	15.50	8.638094	25.000000
3	3	32.00	46.50	14.436484	10.666667
2.9	2	37.00	31.00	12.003639	18.500000
2.6	1	16.50	15.50	8.638094	16.500000
1.9	3	26.50	46.50	14.436484	8.833333
3.1	1	8.50	15.50	8.638094	8.500000
2.2	1	14.00	15.50	8.638094	14.000000
2.5	1	5.50	15.50	8.638094	5.500000
3.3	2	6.00	31.00	12.003639	3.000000
1.7	1	5.50	15.50	8.638094	5.500000
2.1	1	8.50	15.50	8.638094	8.500000
2.3	1	11.00	15.50	8.638094	11.000000

Average scores were used for ties.

Kruskal-Wallis Test

Chi-Square		24.0090
DF		19
Asymptotic Pr >	Chi-Square	0.1958
Exact	Pr >= Chi-Square	0.203

The NPAR1WAY Procedure

Wilcoxon Scores (Rank Sums) for Variable blue

Classified by Variable real

real	N	Sum of Scores	Expected Under H0	Std Dev Under H0	Mean Score
4.6	1	30.00	15.50	8.647736	30.000000
3.8	2	49.00	31.00	12.017038	24.500000
3.9	1	29.00	15.50	8.647736	29.000000
2.7	2	40.50	31.00	12.017038	20.250000

2.8	3	44.00	46.50	14.452598	14.666667
4.4	1	24.50	15.50	8.647736	24.500000
5.1	1	27.00	15.50	8.647736	27.000000
4.8	1	23.00	15.50	8.647736	23.000000
3.6	1	26.00	15.50	8.647736	26.000000
3	3	31.50	46.50	14.452598	10.500000
2.9	2	36.50	31.00	12.017038	18.250000
2.6	1	12.50	15.50	8.647736	12.500000
1.9	3	23.00	46.50	14.452598	7.666667
3.1	1	8.50	15.50	8.647736	8.500000
2.2	1	8.50	15.50	8.647736	8.500000
2.5	1	4.50	15.50	8.647736	4.500000
3.3	2	9.50	31.00	12.017038	4.750000
1.7	1	10.50	15.50	8.647736	10.500000
2.1	1	12.50	15.50	8.647736	12.500000
2.3	1	14.50	15.50	8.647736	14.500000

Average scores were used for ties.

Kruskal-Wallis Test

Chi-Square		22.6182
DF		19
Asymptotic Pr >	Chi-Square	0.2546
Exact	Pr >= Chi-Square	0.354

The SAS System

The NPAR1WAY Procedure

Wilcoxon Scores (Rank Sums) for Variable administration

Classified by Variable real

		Sum of	Expected	Std Dev	Mean
real	N	Scores	Under H0	Under H0	Score
ffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffff					
4.6	1	30.00	15.50	8.647736	30.000000
3.8	2	50.00	31.00	12.017038	25.000000
3.9	1	27.00	15.50	8.647736	27.000000
2.7	2	43.00	31.00	12.017038	21.500000
2.8	3	44.50	46.50	14.452598	14.833333
4.4	1	25.50	15.50	8.647736	25.500000
5.1	1	22.00	15.50	8.647736	22.000000
4.8	1	24.00	15.50	8.647736	24.000000
3.6	1	28.00	15.50	8.647736	28.000000
3	3	34.00	46.50	14.452598	11.333333
2.9	2	37.00	31.00	12.017038	18.500000
2.6	1	16.00	15.50	8.647736	16.000000
1.9	3	20.50	46.50	14.452598	6.833333
3.1	1	11.00	15.50	8.647736	11.000000
2.2	1	12.00	15.50	8.647736	12.000000
2.5	1	4.50	15.50	8.647736	4.500000

3.3	2	11.00	31.00	12.017038	5.500000
1.7	1	6.50	15.50	8.647736	6.500000
2.1	1	8.50	15.50	8.647736	8.500000
2.3	1	10.00	15.50	8.647736	10.000000

Average scores were used for ties.

Kruskal-Wallis Test

Chi-Square	22.9634
DF	19
Asymptotic Pr > Chi-Square	0.2390
Exact Pr >= Chi-Square	0.243

Appendix 2. The results of logistic regression in SPSS

Variables in the Equation

		B	S.E.	Wald	Df	Sig.	Exp(B)
Step 1[a]	Cbo	.492	.353	1.948	1	.0163	1.636
	Constant	-1.868	1.201	2.419	1	.120	.154

a. Variable(s) entered on step 1: cbo.

Variables in the Equation

		B	S.E.	Wald	Df	Sig.	Exp(B)
Step 1[a]	Blue	.320	.347	.851	1	.0356	1.378
	Constant	-.923	1.191	.600	1	.438	.397

a. Variable(s) entered on step 1: blue.

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1[a]	administration	.754	.437	2.973	1	.0085	2.426
	Constant	-2.300	1.362	2.853	1	.091	.100

a. Variable(s) entered on step 1: administration.