УЧЕНЫЕ ЗАПИСКИ КАЗАНСКОГО УНИВЕРСИТЕТА.
СЕРИЯ ФИЗИКО-МАТЕМАТИЧЕСКИЕ НАУКИ

2018, Т. 160, кн. 2
С. 350–356

ISSN 2541-7746 (Print)
ISSN 2500-2198 (Online)

UDK 519.248

IMPROVING THE ACCURACY OF MACROECONOMIC TIME SERIES FORECAST BY INCORPORATING FUNCTIONAL DEPENDENCIES BETWEEN THEM

N.A. Moiseev

Plekhanov Russian University of Economics, Moscow, 117997 Russia

Abstract

A parametric approach to forecasting vectors of macroeconomic indicators, which incorporates functional dependencies between them, has been considered in this paper. As it is possible to functionally bind together most indicators, we believe that this information can help to substantially decrease their forecast error. In this paper, we have proposed to readjust the traditionally obtained forecasts given the known analytical form of the relationship between the considered indicators by the maximum likelihood method. We have also derived a standard form of the readjusted probability density function for each analyzed indicator by normalizing its marginal distribution. In order to prove the efficiency of the proposed method, an empirical out-of-sample investigation has been carried out regarding a simple example for such macroeconomic indicators as gross domestic product (GDP), GDP deflator, and GDP in constant prices.

Keywords: regression analysis, GDP, inflation, monetary base, maximum likelihood method, probability density function, functional dependencies of macroeconomic indicators

Introduction

Accurate forecasting of macroeconomic indicators is a crucial task for both government or corporations and investors, because it helps to elaborate strategies for effective development and risk management. To obtain point and interval forecasts of some macroeconomic time series, most econometricians use regression analysis. In particular, they apply linear regression models with the optimization method of ordinary least squares (OLS). Nowadays, time series of macroeconomic processes are forecasted in the vast majority of cases by modeling separately each indicator under consideration even through using such advanced techniques as partial least squares, nonlinear regression, model averaging, etc.; see [1]. However, it is possible to increase the accuracy of such forecasts by incorporating functional dependencies between modeled macroeconomic indicators. Ideas that are in some way similar to the one discussed in this paper have already been published in some statistical journals and are based on regression models with multiple responses. Such models consist of several equations with the assumption that there is a certain degree of correlation between the modeled target variables. Several papers have covered the biresponse (involving only two dependent variables) nonparametric regression, including [2–5] with smoothing the spline approach, as well as [6] with the local polynomial approach. Basically, the purpose of multi-response modeling is to obtain a better model than the one from single response modeling, given that such a model does not only consider the predictors' influence on a response, but also the relationship between responses. Representation of the relationship between responses is usually expressed in the form of a variance-covariance matrix, which is used as a weighting in the model parameter estimation. The maximum effect from this modeling method is achieved in case of a quite strong correlation between the considered responses, which is explicitly shown in [7, 8], and [9]. This approach is usually applied to panel or cross-sectional data in the field of medicine (see, e.g., [2–4, 10, 11]), sociology (see [4]), and also without any special application as in paper [5]. However, in the field of economics approaches, sharing the same idea as multi-response regression models imply can be successfully used as well. The reasoning under this statement lies in the fact that many macroeconomic indicators are bound by some functional dependency, the understanding of which makes it possible to use this information to reduce the forecast error. In this paper, we consider a very simple case of this relationship between macroeconomic processes, which involves inflation, gross domestic product (GDP), and real gross domestic product (RGDP). However, even constructing this simple model may help to significantly improve the accuracy of forecasts, which clearly offers a huge potential of the proposed ideas for predicting time series of economic processes.

1. Readjustment of traditional forecasts

Let $\{y_t, \mathbf{X}_t : t = 1, 2, \dots, n\}$ be a considered real-valued sample, where y_t is a target variable and $\mathbf{X}_t = \{1, x_{1t}, x_{2t}, \dots, x_{mt}\}$ is a countable dataset of possible explanatory variables. Suppose we can compute a linear regression model as follows:

$$y_t = \mathbf{X}_t\mathbf{B} + e_t \quad \text{or} \quad \hat{y}_t = \mathbf{X}_t\mathbf{B},$$

where e_t is an observed residual of the model at time t and $\mathbf{B}$ is a column vector of parameters, which is computed straightforward according to the OLS method as below:

$$\mathbf{B} = (\mathbf{X}^\mathbf{T}\mathbf{X})^{-1}\mathbf{X}^\mathbf{T}\mathbf{Y}, \tag{1}$$

where

$$\mathbf{X} = \begin{pmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \\ \vdots \\ \mathbf{X}_n \end{pmatrix} \quad \text{and} \quad \mathbf{Y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}.$$

The model has an unobserved error term ϵ_t which is subject to the following assumptions:

Proposition 1. *Strict exogeneity, i.e.,* $E(\epsilon_t|\mathbf{X}) = 0$

Proposition 2. *Homoskedasticity, i.e.,* $E(\epsilon_t^2|\mathbf{X}) = \sigma^2$

Proposition 3. *Normality, i.e.,* $\epsilon_t \sim N(0, \sigma^2)$

Proposition 4. *No perfect multicollinearity, i.e.,* $\mathbf{X}^\mathbf{T}\mathbf{X}$ *is a positive-definite matrix*

Proposition 5. *No autocorrelation, i.e.,* $cov(\epsilon_i; \epsilon_i) = 0 \;\; \forall\, i \neq j$

Then, the probability density function (pdf) for one-step-ahead value of the target variable y_{n+1} is subject to Student's location-scale distribution with $\nu = n - m - 1$ degrees of freedom.

$$\Psi(y_{n+1}) = \frac{\Gamma\left(\dfrac{\nu+1}{2}\right)}{s\sqrt{\nu\pi}\,\Gamma\left(\dfrac{\nu}{2}\right)} \left[\frac{\nu + \left(\dfrac{y_{n+1} - \hat{y}_{n+1}}{s}\right)^2}{\nu}\right]^{-(\nu+1)/2}, \tag{2}$$

where

$$\hat{y}_{n+1} = \mathbf{X}_{n+1}\mathbf{B} \tag{3}$$

is location parameter,

$$s = \sqrt{\frac{\mathbf{e}^{\mathbf{T}}\mathbf{e}}{n-m-1}\Big(1 + \mathbf{X}_{n+1}(\mathbf{X}^{\mathbf{T}}\mathbf{X})\mathbf{X}^{\mathbf{T}}_{n+1}\Big)} \tag{4}$$

is scale parameter and $\mathbf{e} = (e_1, e_2, \ldots, e_n)^T$ is a column-vector of residuals observed for the model. Now suppose we have a set of target variables $y_t^{(1)}, y_t^{(2)}, \ldots, y_t^{(K)}$, each modeled by a data set $\mathbf{X}^{(1)}, \mathbf{X}^{(2)}, \ldots, \mathbf{X}^{(K)}$, respectively. Besides, all the considered target variables are bound with functional dependency of some form

$$y_t^{(i)} = f_i\big(y_t^{(1)}, \ldots, y_t^{(i-1)}, y_t^{(i+1)}, \ldots, y_t^{(K)}\big), \tag{5}$$

where subscript i denotes the function that expresses $y_t^{(i)}$ in terms of other target variables.

Then, it is possible to readjust each forecasted value of target variables taking into account all obtained pdfs and known functional dependency between target variables. It is proposed that this procedure is performed using the maximum likelihood method, where we maximize the product of the considered pdfs. Below we present a joint likelihood computed using f_1.

$$\mathrm{LH} = \Psi_1(f_1[y_{n+1}^{(2)}, y_{n+1}^{(3)}, \ldots, y_{n+1}^{(K)}]) \cdot \Psi_2(y_{n+1}^{(2)}) \cdots \Psi_K(y_{n+1}^{(K)})$$

To simplify the computational process, we shall use log-likelihood as displayed below:

$$\text{log-LH} = \ln\{\Psi_1(f_1[y_{n+1}^{(2)}, y_{n+1}^{(3)}, \ldots, y_{n+1}^{(K)}])\} + \ln\{\Psi_2(y_{n+1}^{(2)})\} + \cdots + \ln\{\Psi_K(y_{n+1}^{(K)})\} \tag{6}$$

Thus, the values of the considered target variables that maximize expression (6) appear to be more precise than the initial ones. It is also possible to obtain readjusted pdf for all the considered target variables by computing the normalized marginal probability distribution as shown below:

$$\mathrm{pdf}\,\big(y_{n+1}^{(i)}\big) = \frac{\Omega\big(y_{n+1}^{(i)}\big)}{\int\limits_{-\infty}^{\infty} \Omega\big(y_{n+1}^{(i)}\big)\,\mathrm{d}y_{n+1}^{(i)}}, \tag{7}$$

where

$$\Omega\big(y_{n+1}^{(i)}\big) = \int \cdots \int\limits_{-\infty}^{\infty} \Psi_1(y_{n+1}^{(1)}) \cdots \Psi_{i-1}(f_{i-1}[y_{n+1}^{(1)}, y_{n+1}^{(2)}, \ldots, y_{n+1}^{(i-2)}, y_{n+1}^{(i)}, \ldots, y_{n+1}^{(K)}]) \times$$
$$\times\, \Psi_i(y_{n+1}^{(i)}) \cdots \Psi_K(y_{n+1}^{(K)})\,\mathrm{d}y_{n+1}^{(1)} \cdots \mathrm{d}y_{n+1}^{(i-1)}\mathrm{d}y_{n+1}^{(i+1)} \cdots \mathrm{d}y_{n+1}^{(K)}. \tag{8}$$

Hence, we conclude that knowing the functional dependency between the considered macroeconomic indicators can be a helpful source of information and facilitates improvement of the classical regression models.

Table 1. Mean squared realized forecast error for GDP deflator, RGDP, and GDP with and without readjustment

Number of observations	Without readjustment Price index	Real GDP index	GDP index	With readjustment Price index	Real GDP index	GDP index
20	0.9705	0.1103	0.9757	**0.9123**	**0.1064**	**0.9702**
30	0.8295	0.0857	0.8598	**0.7985**	**0.0837**	**0.8362**
40	0.7695	0.0775	0.8373	**0.7444**	**0.0769**	**0.7971**
50	0.6878	0.0779	0.7502	**0.6592**	**0.0779**	**0.7181**
60	0.6662	0.0792	0.7364	**0.6341**	**0.0791**	**0.7009**
70	0.6654	0.0837	0.7683	**0.6443**	**0.0828**	**0.7217**
80	0.6562	0.0887	0.7718	**0.6481**	**0.0887**	**0.7199**
90	0.6516	**0.0866**	0.7941	**0.6461**	0.0867	**0.7388**
100	0.5814	**0.0776**	0.6822	**0.5694**	0.0792	**0.6537**
110	0.5294	**0.0577**	0.6628	**0.5247**	0.0581	**0.6242**
120	0.5324	**0.0511**	0.6626	**0.5208**	0.0521	**0.6256**

2. Empirical testing

In order to prove the efficiency of the proposed method, we carried out an empirical experiment with one-step-ahead out-of-sample forecasting of three macroeconomic indicators: gross domestic product (GDP), GDP deflator, and real GDP or GDP in constant prices for the USA. Our data set comprises historical data on these indicators from Q1.1947 to Q3.2016. From basic principles of macroeconomics we know that

$$I_{pq} = \frac{\sum p_1 q_1}{\sum p_0 q_0}, \quad I_p = \frac{\sum p_1 q_1}{\sum p_0 q_1}, \quad I_q = \frac{\sum p_0 q_1}{\sum p_0 q_0} \quad => \quad I_{pq} = I_p \times I_q,$$

where I_{pq} is the GDP index, I_p is the GDP deflator computed by the Paasche index, and I_q is the RGDP Laspeyres index.

Then, our three target variables will be $y^{(1)} = (I_p - 1) \cdot 100\%$, $y^{(2)} = (I_q - 1) \cdot 100\%$, and $y^{(3)} = (I_{pq} - 1) \cdot 100\% = [(1 + y^{(1)} \div 100) \cdot (1 + y^{(2)} \div 100) - 1] \cdot 100\%$. Each of these target variables are forecasted by linear fourth order autoregression as below:

$$y_t^{(i)} = b_0^{(i)} + b_1^{(i)} y_{t-1}^{(i)} + b_2^{(i)} y_{t-2}^{(i)} + b_3^{(3)} y_{t-3}^{(i)} + b_4^{(i)} y_{t-4}^{(i)} + e_t^{(i)} \tag{9}$$

In our empirical experiment, we compared the mean squared realized error for the predicted values with and without readjustment procedure for different number of observations in a data frame. The results are presented in Table 1. As one can see from this table, nearly all readjusted forecasts have a smaller mean squared realized error (highlighted in bold) than those without readjustment. In the cases when the original forecast demonstrated better performance, its mean squared realized forecast error was insignificantly smaller than that for the proposed method (maximal difference was around 2%). The fact that the proposed readjustment did not work perfectly for GDP in constant prices is explained by the significantly smaller forecast error compared to the deflator and GDP. Similarly, the obtained distributions for more volatile variables are not able to considerably correct the less volatile variable, because they contain insufficient data for it. In this respect, the proposed method would surely work well for all variables if they have more or less similar predictability. Nevertheless, the mean squared realized forecast error for GDP deflator was reduced by 3% on average, as well as by 1% for RGDP and by 5% for GDP, respectively.

In order to demonstrate how the readjusted probability density differs from the original one, we also provide a graphical example from our empirical investigation, see Fig. 1.

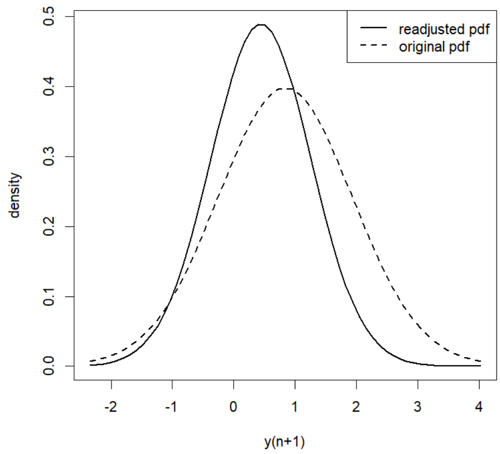

Fig. 1. Original and readjusted probability density function (pdf)

From Fig. 1 one can clearly observe that the newly computed pdf not only shifts an expected value for y_{n+1} after readjustment, but also shrinks its original pdf, which ensures better prediction. Thus, we managed to obtain considerable corrections of the initial forecasted values even with such a simple example and, due to that, to substantially increase the forecast accuracy of the considered macroeconomic indicators. Notably, it is possible to bind together many more indicators, such as ultimate consumption, balance of trade, unemployment, money velocity, monetary base, industrial production, etc. By doing that, one can create a complex macroeconomic forecasting model that would predict incorporated indicators being in accordance with each other and having greater efficiency than when forecasting each indicator separately.

Conclusions

In this paper, we discuss the idea that knowing the functional dependency between forecasted macroeconomic data can help to improve the accuracy of obtained predictions and to avoid any kind of disaccordance in forecasted values. The point is that if each indicator under consideration is modeled separately by its own regression equation, the obtained forecasts very often contradict each other, thereby making impossible the simultaneous occurrence of all predicted values in the modeled vector of indicators. Therefore, we propose to take into account additional information from the functional relationship between the data to bring initial forecasts into accordance. This procedure can be implemented by the maximum likelihood method. Furthermore, we derive a standard form for the readjusted pdf by using the normalized marginal distribution of the variable with regard to the preliminary obtained distributions of all analyzed target variables. This readjusted pdf can be used to compute the interval forecast with a given significance level. Our out-of-sample empirical experiment testifies in favor of the proposed method, even in case when the simplest macroeconomical relationship between just three indicators (GDP deflator, GDP, and GDP in constant prices) is taken into account. Therefore, we conclude that the proposed method can be considered as promising and useful for forecasting a set of balance macroeconomic indicators, because it helps to substantially reduce the forecast error.

Acknowledgements. The study was supported by the Russian Foundation for Basic Research (project no. 18-310-20008).

References

1. Moiseev N.A. Linear model averaging by minimizing mean-squared forecast error unbiased estimator. *Model Assisted Stat. Appl.*, 2016, vol. 11, no. 4, pp. 325–338. doi: 10.3233/MAS-160376.
2. Wang Y., Guo W., Brown M.B. Spline smoothing for bivariate data with application to association between hormones. *Stat. Sin.*, 2000, vol. 10, no. 2, pp. 377–397.
3. Chen H., Wang Y. A penalized spline approach to functional mixed effects model analysis. *Biometrics*, 2010, vol. 67, no. 3, pp. 861–870. doi: 10.1111/j.1541-0420.2010.01524.x.
4. Welsh A.H., Yee T.W. Local regression for vector responses. *J. Stat. Plann. Inference*, 2006, vol. 136, no. 9, pp. 3007–3031. doi: 10.1016/j.jspi.2004.01.024.
5. Lestari B., Budiantara I.N., Sunaryo S., Mashuri M., Spline estimator in multi-response nonparametric regression model with unequal correlation of errors. *J. Math. Stat.*, 2010, vol. 6, no. 3, pp. 327–332.
6. Chamidah N., Budiantara I.N., Sunaryo S., Zain I. Designing of child growth chart based on multi-response local polynomial modeling. *J. Math. Stat.*, 2012, vol. 8, no. 3, pp. 342–247.
7. Ruchstuhl A., Welsh A.H., Carroll R.J. Nonparametric function estimation of the relationship between two repeatedly measured variables. *Stat. Sin.*, 2000, vol. 10, pp. 51–71.
8. Welsh A.H., Lin X., Carroll R.J. Marginal longitudinal nonparametric regression: Locality and efficiency of spline and kernel methods. *J. Am. Stat. Assoc,*, 2002, vol. 97, no. 458, pp. 482–493.
9. Guo W. Functional mixed effects models. *Biometrics*, 2002, vol. 58, no. 1, pp. 121–128. doi: 10.1111/j.0006-341X.2002.00121.x.
10. Antoniadis A., Sapatinas T., Estimation and inference in functional mixed-effects models. *Comput. Stat. Data Anal.*, 2007, vol. 51, no. 10, pp. 4793–4813. doi: 10.1016/j.csda.2006.09.038.
11. Krishtafovich V., Krishtafovich D., Belkin Y., Gubarev R. Histological identification of animal protein ingredients. *Pak. J. Nutr.*, 2016, vol. 15, no. 1, pp. 95–98. doi: 10.3923/pjn.2016.95.98.

Recieved
December 8, 2017

Moiseev Nikita Alexandrovich, Candidate of Economics, Associate Professor of the Department of Mathematical Methods in Economics

Plekhanov Russian University of Economics
per. Stremyannyi 36, Moscow, 117997 Russia
E-mail: *moiseev.na@rea.ru*

УДК 519.248

Повышение точности прогноза временных рядов макроэкономических процессов путем учета функциональных зависимостей между ними

Н.А. Моисеев

Российский экономический университет им. Г.В. Плеханова, г. Москва, 115093, Россия

Аннотация

В статье рассмотрен параметрический подход к прогнозированию векторов макроэкономических показателей, который включает в себя функциональные зависимости между ними. Поскольку большинство таких индикаторов можно связать функционально, предполагается, что использование этой информации может существенно снизить ошибку прогноза. С помощью метода максимального правдоподобия предлагается скорректировать традиционно полученные прогнозы, учитывая известную аналитическую форму взаимосвязи между рассматриваемыми индикаторами. В статье представлен также вывод стандартной формы исправленной функции плотности вероятности для каждого анализируемого индикатора путем нормализации ее маргинального распределения. Эффективности предложенного метода доказана с помощью эмпирического вневыборочного тестирования согласно тривиальному примеру, который включает такие макроэкономические показатели, как валовой внутренний продукт (ВВП), дефлятор ВВП и ВВП, выраженный в постоянных ценах.

Ключевые слова: регрессионный анализ, ВВП, инфляция, денежная база, метод максимального правдоподобия, функция плотности вероятности, функциональные зависимости макроэкономических показателей

Поступила в редакцию
08.12.17

Моисеев Никита Александрович, кандидат экономических наук, доцент кафедры математических методов в экономике

Российский экономический университет им. Г.В. Плеханова
Стремянный пер., д. 36, г. Москва, 115093, Россия
E-mail: *moiseev.na@rea.ru*

***For citation*:** Moiseev N.A. Improving the accuracy of macroeconomic time series forecast by incorporating functional dependencies between them. *Uchenye Zapiski Kazanskogo Universiteta. Seriya Fiziko-Matematicheskie Nauki*, 2018, vol. 160, no. 2, pp. 350–356.

***Для цитирования*:** *Moiseev N.A.* Improving the accuracy of macroeconomic time series forecast by incorporating functional dependencies between them // Учен. зап. Казан. ун-та. Сер. Физ.-матем. науки. – 2018. – Т. 160, кн. 2. – С. 350–356.