PREDICTION OF LITHUANIAN GDP: ARE REGRESSION MODELS OR TIME SERIES MODELS BETTER?

Alina Stundziene

Kaunas University of Technology, Lithuania

crossref http://dx.doi.org/10.5755/j01.em.18.4.5041

Abstract

The problem of Lithuanian GDP prediction is relevant. There are several institutions, such as Statistics Lithuania, state's central bank, other banks that constantly announce their predictions of GDP. Frequently the forecasts of different institutions vary because they use different methods. The main purpose of the paper is to investigate whether regression models made of the monthly published economic indicators or time series models are better for Lithuanian GDP prediction.

The changes of Lithuanian GDP as well as many other economic indicators that have impact on GDP are published quarterly. Prediction of quarterly economic indicators as it was done by the most researchers can be related to greater errors comparing with the prediction models that are made according to the monthly data. Monthly data can ensure that the newest information is used for prediction of GDP and show how the state's economy is changing in the current quarter, that's why it can reduce the error of prediction.

The research is based on the economic data that is measured and published monthly by Statistics Lithuania (154 ratios at all). Various linear and non-linear regression models are made in order to find the best model for Lithuanian GDP prediction. The results of regression models are also compared with the results got by ARIMA (time series) models. The analysis showed that the regression models made of monthly published economic indicators may be better than time series models for prediction or Lithuanian GDP.

Paper type: *Research paper.*

Keywords: *Lithuanian GDP, forecasting, regression models, time series models.*

JEL Classification: *C51, E27, O47.*

1. Introduction

Prediction of the state's economic situation is a very important but not an easy task. Many things such as competitiveness of the state, production, sales, profit of the firm and human welfare, their living quality depend on the economics. It soaks to all stratum of the state. Companies plan their production, number of workers, payments, investments according to the expectations of the state's economy. The government plans the collection of means to the state's budget, the changes in tax system, expenditures, investment and so on. The situation of economics affects almost everybody. Depending on their expectations each person can change its habits of consumption, choose between savings and investment.

It is obvious that macroeconomic indicators are bound up with others, so they cannot be analysed separately. If one indicator changes the other indicators react as well. So regression models can be one of the possibilities to analyse these relations. Also we must remember that the trend of the past data does not guarantee the same trend in the future that's why advanced time series models can be useful to predict the tendency of one or another indicator.

There are several institutions, such as department of statistics, state's central bank other banks that constantly announce their predictions of macroeconomic indicators. Frequently the forecasts of different institutions vary usually because of different methods of forecasting they employ.

The object of this article is to create appropriate model for prediction of Lithuanian GDP. There are several Lithuanian researchers who applied different methods for the prediction of Lithuanian GDP but no one can be called as the best. Commonly two things must be coordinated: complexity and precision of the model. Usually the more complicated the model, the better prediction can be obtained, but because of its complexity not everybody may be able to use it, and on the contrary, if there is a simple model the precision of it is not high. Thus, the goal of this article is to find a model that is not complicated but quite precise for the prediction of the Lithuanian GDP.

Everybody knows that quarterly data releases of GDP are available only with delay so the prediction (or correction of prediction taking account the newest published data) of GDP for the next quarter can also be made with delay. Most institutions are interested in information about the real-time situation so frequently published information is the most useful. That's why the monthly published data were chosen for this research and that is the difference from the other research that makes forecasts of GDP according to the quarterly data.

The research methods applied in this article are logical and comparative analysis of scientific literature, simple and multiple linear regression models, and time series models. The results of the research were obtained by one of the advanced statistical software packages EViews 8.

2. Methods employed in economic modelling

Definition of the economic model

The basis for economic analyses at the decision-making level is focused on mathematical modelling of the real economic phenomena (Chvatalova, Simberova, 2011). An economic model is a set of equations which describes how the economy or some part of it functions (Almon, 2008). Macroeconomic modelling means looking at the time series of economic data such as the national income and product accounts, employment, interest rates, money supply, exchange rates, prices and others. A model should incorporate and test the understanding of how the economy works. Its equations should make sense. It should be possible to test how adequate the model is by running it over the past and seeing how well it can reproduce the history. It should be possible to analyse the effects of policies by changing some of the model's assumptions and rerunning history with the changed assumptions. Finally, it should be useful not only for the policy analysis but also for the prediction. The builder of the model may also improve his or her understanding of the economy by studying the errors of the forecast (Clopper, 2012).

Model building is a serious business. Many models that are used by forecasters look quite good. But expectations are seldom rational, and when it becomes clear that the model is not the panacea that can cure all the ills of a business, some disillusion set in.

Description of the methods applied for the economic analysis and prediction

Two kinds of prediction methodologies can be identified in the literature: the methods based on parametric modelling, and the methods based on non-parametric techniques. The former group of methods includes the linear autoregressive models, the non-linear, SETAR-STAR, Markov switching and other models. The latter one includes Kernels method, nearest neighbour method, neural network and wavelet methods (Guegan, Rakotomarolahy, 2010).

Growing use of various linear autoregressive models has arisen the discussions about the prediction of GDP. Times series models, usually ARIMA models, are applied for economic modelling in many countries, for example Taht (2008) has applied it for Estonian GDP prediction, Turturian (2007) used it for Romanian GDP prediction , Andrei and Bugudui E. (2011) applied it for US economy, Klučik and Juriová (2010) forecasted Slovakian GDP and so on. These models are also used by European Central Bank (Barhoumi and others, 2008).

Structural vector autoregressive (SVAR) models are also still popular among economists. SVAR is a multivariate, linear representation of a vector of observations on its own lags. SVAR models are used by economists to recover the economic shocks from observables by imposing a minimum of assumptions compatible with a large class of models.

The bridge equations method is another approach that combines linearity and aggregation (Diron, 2008). In that case, the object is to diminish the number of economic indicators. The bridge models provide short-run projections of quarterly time series by using the information available on indicators at monthly frequency (the model thus creates a "bridge"). In general, these models are used for one or, at most, two-quarter forecasts ahead. The approach relies on selected quantitative and qualitative monthly indicators and on the specification of a statistical equation that links these indicators with the quarterly series that has to be projected.

Frequently information on the indicator is not available for all months of the quarter (for all forecasting horizon). In this case a preliminary projection of the monthly indicator is conducted. This is done by the means of other indicators or through the autoregressive models (ARIMA) projecting the monthly indicator on the basis of its past dynamics (Baffigi and others, 2004). Exceptions include interest rates, calculated as implicit rates drawn from the yield curve and the exchange rates, whose expectations are obtained from the uncovered interest rate parity condition applied on US and Euro Area interest rates.

Bridge equations are used by many institutions and have been studied in various papers (Baffigi, Golinelli & Parigi, 2004; Runstler & Sedillot, 2003; Barhoumi and others, 2008). The results of Diron's (2008) research are widely used in Central Banks. Her method is based on a limited number of economic indicators which are plugged in eight linear equations from which an estimate of GDP is obtained. Her method associates the bridge equations and forecasts combinations incorporating a large number of economic activities including different single forecasts based on production sectors, survey data, financial variables and leading index constructed from large number of economic indicators. Her method is competitive comparing with the methods that include a huge number of indicators.

Traditional bridge equations can handle only few variables. Recently, Giannone, Reichlin & Sala (2004) and Giannone, Reichlin & Small (2005) have proposed to use the factors extracted from the large monthly datasets to perform bridging which exploit a large number of indicators within the same model (bridging with factors). They propose to use the Kalman filter to estimate the factors and handle missing data.

Factor models are also applied for GDP prediction, for example Schumacher (2005) used it for prediction of German GDP, Cheung and Demers (2007) applied it for prediction of Canadian GDP, Camacho and Martinez-Martin (2012) used it for prediction of US economics. Methods that have been used in the Eurosystem include the principal component estimator of the factors (Stock & Watson, 2002) and the frequency domain-based two-step estimator of Forni et al. (2005).

Runstler and others (2008) have performed an evaluation of models used in central banks for computing early estimates of current quarter GDP and short-term forecasts of next-quarter GDP. The main finding obtained for the euro area countries was that the bridge models, which timely exploit monthly releases, fare considerably better than the quarterly models and dynamic factor models, which exploit a large number of releases, do generally better than the traditional bridge equations.

Recently the forecasts of GDP based on the microeconomic foundation appear with the so called dynamic stochastic general equilibrium models (Smets & Wouters, 2004). Today they are mostly employed for the evaluation of economic policy. The analysis of international trade, taxes, the policy of state's investments, the qualification of labour force, employment problems, money policy, regional development, integration in economic unions, application of the new technologies, control of protection of nature and others are the spheres of application of general equilibrium models (Tamosiunas, 1999). Nevertheless the practical experience of using modern DSGE models for the scenario analysis in the policymaking institutions is relatively limited and diverse across the individual institutions. Still, the adoption of the DSGE approach to the scenario analysis is, in many respects, an ongoing process and there is much to be learned about the design, implementation and communication of DSGE model-based scenario analysis.

As the research of scientists shows that modelling by DSGE models is not yet satisfactory, a lot of national central banks continue to use the traditional macroeconomic models along with the DSGE models, thus recognizing advantages and limitations of both modelling approaches. The linear ARIMA or VAR models remain the benchmarks in the literature.

It is impossible to find the perfect model in practice because of the uncertainty. Uncertainty plays an important role in many areas of economic behaviour that's why uncertainty is also inherent to forecasting (Boero, Smith & Wallis, 2008). Several sources of uncertainty can be identified. First, uncertainty comes from the forecasting model itself: misspecification, imprecision of the estimated parameters, data errors and revisions. Second, "real-time" forecasting uncertainty depends on the economic conditions and their impact on the size of possible forecasting errors at a given point in time (Laurent, Kozluk, 2012).

Bratu (2012) states some important strategies that can be used in practice in order to improve the accuracy of forecasts. One of these strategies is building combined forecasts in different variants. As there is no one the most suitable method, many countries build a set of different models for prediction of quarterly GDP. For example ARIMA models with the seasonal components and indicator models, similar to the bridge models are used in Albania (Celiku, Kristo, Boka, 2009).

The experience of the prediction of Lithuanian GDP

Macro-econometric modelling in the post-Soviet Lithuania became a topical area. There were a lot of discussions of macro-econometric modelling alternatives, specificity of modelling the Lithuanian economy. Moreover, very important questions were raised: will the created empirically adequate macro-econometric model will be useful in further developments of macro-econometric modelling, e.g., the foreign modelling experience overview and the analysis of Lithuanian specifics were used in the mathematical model of Lithuanian economy. Kropas (1998) and Vilpišauskas (2001) were one of the first analysts in economic modelling.

Lithuania has only few developed macro-econometric models, which are able to make forecasts. One is used by government and others are used by several banks (Stankevičienė, Gruodis, Lokutijevskij, Urbaitė, 2012). Institute of Economics with the help of other institutions has built a medium-sized macro-econometric sectoral model of the Lithuanian economy called LITMOD. A central element in the model is a 12-sector input-output table of the Lithuanian economy facilitating the analyses of structural changes (Celov and others, 2005).

Various individual research in prediction of Lithuanian GDP can also be found. For example Rukšėnaitė (2010) made the analysis of 41 indicators in order to say which are the most important for GDP. She pointed the turnover of retail trade, except of motor vehicles and motorcycles, car fuel trade, export and turnover of retail trade in food and beverage service as the most correlated indicators with GDP. Virbukaitė (2011) has employed VAR method for the prediction of Lithuanian GDP. Although her model was quite accurate, she chose several endogenous variables herself without analyzing their significance to Lithuanian GDP.

Lithuanian scientists started to use DSGE model in 1996, in co-operation with the Academies of Science of Estonia and Latvia as well as the Erasmus University in the Netherlands. Karpavicius and Vilkas (1997) were the first that applied DSGE model for Lithuanian economics. Karpavicius (2008) has properly calibrated the DSGE model for recent Lithuanian data. Together with Vetlov (2008) he had analyzed the economics' impact of the 2006-2008 personal income tax (PIT) reform in Lithuania. Karpavicius (2009) has also examined the effects of the fiscal instruments, namely labour tax, capital tax, consumption tax, transfers to households, and government spending, on Lithuanian economy and welfare assuming balanced government budget using DSGE model.

DSGE models are generally devised for mature economies that are in the vicinity of the steady state of their economic development. In this case, the analysis of impulse responses and simulations is reasonable and policy-relevant. In contrast, many economies, Lithuanian as well, might be decades away from its steady state which could create some doubts regarding the reliability of results.

Mentioned models are restricted and make predictions only for the short or medium runs. Of course maintaining a macro economic model requires a proper management of huge amounts of information, trained and highly skilled personnel able to deal with complex computations and persisting problems, and good knowledge of the countries' macro economy. This is one of the main reasons why there are only few models that are used in Lithuania nowadays and one of the reasons to develop a new one, which will provide a sophisticated approach to the future scenarios of the economy, which can be analyzed afterwards.

3. Model creation for the prediction of Lithuanian GDP

Regression model creation for the prediction of Lithuanian GDP

In order to exploit the opportunity to employ the newest information, economic data that are measured and published monthly by Statistics Lithuania were chosen for creation a new regression model. 154 ratios were collected as independent variables. All they can be grouped as macroeconomics data (for example price indices), population and social statistics (population, migration, labour market data), business statistics (for example production by various economic activities), trade and transport statistics. The analysis was made for the years of 2004-2012. Since dependent variable – GDP – is calculated and published for the quarters, all monthly data of 154 ratios were also recalculated in order to get quarterly data.

In order to get a good regression model the data of the variables must satisfy several requirements: they must be normally distributed and can't have the outliers. The preliminary analysis of the indicators (dependent and independent variables) shows that the data of 38 independent variables are not normally distributed. Various non-linear transformations were used in order to get the normal distribution of these data. Seven variables became normally distributed after $x_i' = x_i^2$ transformation, other seven variables were transformed by $x_i' = \ln(x_i)$ and 24 indicators were rejected from the further analysis because there was impossible to find the transformation that makes them normally distributed. The test of outliers showed that some variables have outliers, but these data can't be rejected from the further analysis because of they possibility of recurrence.

The correlation analysis showed that several independent variables are strongly correlated with GDP. They are:

- Turnover (VAT excluded) of retail trade, except of motor vehicles and motorcycles, at current prices, LTL thousand (X_{127});
- Turnover (VAT excluded) of food and beverage service activities, at current prices, LTL thousand (X_{129});
- Passenger arrivals and departures at airports, in thousands (X_{153});
- Industrial production (VAT and excises excluded) of manufacturing at current prices, LTL thousand (X_{118});
- Imports, LTL thousand (X_{141});
- Industrial production (VAT and excises excluded) at current prices, LTL thousand (X_{112});
- Imports, seasonally adjusted, LTL thousand (X_{142});
- Industrial production (VAT and excises excluded) of manufacturing at constant 2005 prices, LTL thousand (X_{117}).

The correlation coefficient between GDP and any of these indices exceed 0.85. The correlation between GDP and turnover (VAT excluded) of retail trade, except of motor vehicles and motorcycles is the strongest and equals to 0.97. Scatter plot (see Figure 1) also shows the strong linear relation between these indicators.

The simple linear regression model between these indicators can be written as:

$$Y = 1388.58 + 0.00373 \cdot X_{127}. \quad (1)$$

Here Y is GDP, X_{127} is turnover (VAT excluded) of retail trade, except of motor vehicles and motorcycles. The coefficient of determination of such model is equal to 0.94. So it is quite good result if we want 94% of the accuracy of the model.

Greater accuracy can be obtained by making multiple regression model. Stepwise-forwards regression method was chosen for the model formation. EViews lets to perform the automatic variable selection using stepwise regression. It allows some or all of the variables in a standard linear multivariate regression to be chosen automatically, using various statistical criteria, from a set of variables. This time the probability of not putting the variable into the regression model (p-value) was chosen for the stopping criteria.

Stepwise-forwards regression formation method begins with no additional regressors in the regression and then adds the variable with the lowest p-value. The variable with the next lowest p-value given that the first variable has already been chosen, is then added. Next both of the added variables are checked against the backwards p-value criterion. Any variable whose p-value is higher than the criterion is removed.

Once the removal step has been performed, the next variable is added. At this, and each successive addition to the model, all the previously added variables are checked against the backwards criterion and possibly removed. The stepwise-forwards routine ends when the lowest p-value of the variables not yet included is greater than the specified forwards stopping criteria (Startz, 2013).

P-values for the forward and backward procedures were set to 0.05 in this research. All independent variables (primary and transformed if primary data were not normally distributed) were specified as potential variables that can be put into the regression model, except 24 indicators that were rejected during the preliminary analysis because transformation that makes these indicators normally distributed could not be found. Following the stepwise selection process, EViews reports the results of the final regression, i.e. the regression of the selected variables on the dependent variable. The results show that GDP can be forecasted by such the model:

$$Y = 0.003074*X_{127} + 165.55*X_{32} + 5.591532*X_{153} + 0.200558*X_{56}^2 - 1.848777*X_{109} + 140.0275*X_2 - 0.000137*X_{120}. \quad (2)$$

Here Y is GDP, X_{127} is the index of turnover (VAT excluded) of retail trade, except of motor vehicles and motorcycles, X_{32} is the index of unemployment, X_{153} is the index of passenger arrivals and departures at airports, X_{56}^2 is the transformed (squared) index of price expectations during the next twelve months (business tendency survey results), X_{109} is the index of price expectations of services during the next two-three months (business tendency survey results), X_2 is the index of the change in consumer prices comparing with the previous quarter, X_{120} – is the index of industrial production of electricity, gas, steam and air conditioning supply at constant 2005 prices.

The accuracy of such model is quite high – adjusted coefficient of determination is equal to 0.99, but the probabilities of Student statistics of the last three independent variables are higher than significant level (0.05). It means that these variables are not significant in this regression model and must be thrown away. At first the most insignificant ratio X_{109} is thrown from the list of potential independent variables and the stepwise-forward procedure is repeated. Simpler regression model is got for prediction of GDP at this time:

$$Y = 0.003075*X_{127} + 153.1835*X_{32} + 5.78856*X_{153} + 0.198661*X_{56}^2. \quad (3)$$

All these independent variables are significant (probabilities of Student criteria are lower than 0.05) and the precision of the model is very high – 99%.

Now the assumption of linearity must be checked. The scatter plots of each independent variable in the pair of dependent variable show that X_{32} and X_{56}^2 are not linear with Y, so these indexes are also not suitable for multiple regression model (see Figure 1). Correlation coefficients between them also show the existence of non-linearity.

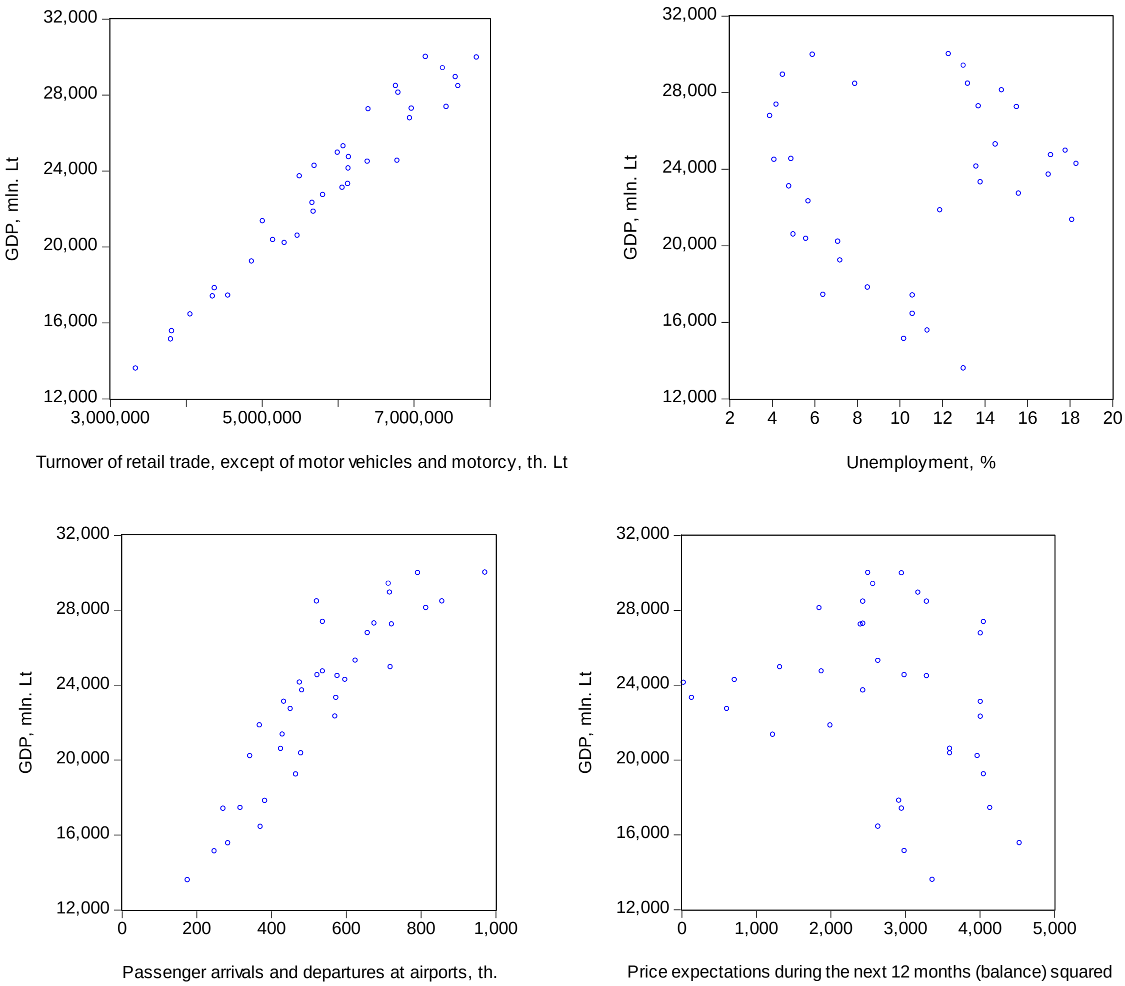

Figure 1. Scatter plots between Y and X_{127}, X_{32}, X_{153}, X_{56}^{2}

Source: created by the author using EViews

So variables X_{32} and X_{56}^{2} are thrown from the list of potential independent variables and the stepwise-forward procedure is repeated. New regression model is got

$$Y = 0.002578*X_{127} + 0.275726*X_{147} + 5.980537*X_{153} + 0.053899*X_{124} \quad (4)$$

and only two ratios X_{127} and X_{153} are the same as in the previous model. Two other variables X_{147} (goods loading) and X_{124} (industrial production (VAT and excises excluded) of water supply, sewerage, waste management and remediation activities, at current prices) are added into the model. The model evaluation screen shows that all these independent variables are significant and the model is accurate in 98,7%. The linearity analysis also shows that all these four independent variables are linear dependent with GDP. The correlation coefficients between them are showed in Table 1.

Table 1. Correlation matrix

	Y	**X_{124}**	**X_{127}**	**X_{147}**	**X_{153}**
Y	1	0.821	0.971	0.773	0.898
X_{124}	0.821	1	0.731	0.891	0.752
X_{127}	0.971	0.731	1	0.666	0.813
X_{147}	0.773	0.891	0.666	1	0.749
X_{153}	0.898	0.752	0.813	0.749	1

Source: created by the author

The correlation matrix shows that there is strong linear relation between dependent and independent variables. But it also shows that there is strong linear relationship between independent variables that can point the problem of multicollinearity. Multicollinearity occurs when there is a linear relationship among one or more of the independent variables. When multicollinearity exists the estimation of the model can be problematic. The inclusion of both independent variables that are collinear adds no more information to the model than the inclusion of just one of them.

The existence of multicollinearity can be checked by calculating Variance Inflation Factor (VIF). If VIF for one of the variables is around or greater than 5, there is collinearity associated with that variable and one of these variables must be removed from the regression model. At first all possible regression models between X_j ($j = 124, 127, 147, 153$) and other X_j, ($i = 124, 127, 147, 153$) $i \neq j$ must be created, then the significant model must be found and adjusted coefficient of determination (R^2) of such model must be estimated. Then VIF is calculated by the formula:

$$VIF_j = \frac{1}{1 - R_j^2} \tag{5}$$

Results are shown in Table 2.

Table 2. Evaluation of multicollinearity

Dependent variable	Independent variables	Significant model	R^2	VIF
X_{124}	X_{127}, X_{147}, X_{153}	X_{124}=f(X_{127}, X_{147})	0.828	5.81
X_{127}	X_{124}, X_{147}, X_{153}	X_{127}=f(X_{153})	0.661	2.95
X_{147}	X_{124}, X_{127}, X_{153}	X_{147}=f(X_{124})	0.794	4.85
X_{153}	X_{124}, X_{127}, X_{147}	X_{153}=f(X_{127}, X_{147})	0.739	3.83

Source: created by the author

VIF shows that independent variables X_{127}, X_{147} and X_{153} are not collinear, but multicollinearity exist among X_{124} and X_{127}, X_{147} as VIF is higher than 5. That means one of these variables must be thrown from the regression model. Collinearity between X_{147} and X_{124} can also be a problem as VIF is close to 5. The variable that is less significant for the regression must be eliminated from the model. Variable selection summary in the model evaluation screen can help to answer to this question. The selection summary shows that X_{127} was added the first into the model, so it is the most significant for the regression and X_{124} was the last that was put into the model, so this variable is significant at least for this regression model. That's why X_{124} is removed from the regression model and the stepwise-forward procedure is repeated. The results are shown in Figure 2.

Dependent Variable: Y
Method: Stepwise Regression
Included observations: 36
No always included regressors
Number of search regressors: 126
Selection method: Stepwise forwards
Stopping criterion: p-value forwards/backwards = 0.05/0.05

	Coefficient	Std. Error	t-Statistic	Prob.
X_{127}	0.002826	0.000118	23.99243	0.0000
X_{147}	0.473039	0.070265	6.732246	0.0000
X_{153}	4.573546	0.887709	5.152076	0.0000

R-squared	0.984463	Mean dependent var	23241.40
Adjusted R-squared	0.983521	S.D. dependent var	4551.830
S.E. of regression	584.3223	Akaike info criterion	15.65844
Sum squared resid	11267276	Schwarz criterion	15.79040
Log likelihood	-278.8519	Hannan-Quinn criter.	15.70450
Durbin-Watson stat	1.724717		

Selection Summary

Added X_{127}
Added X_{147}
Added X_{153}

Figure 2. Multiple regression model evaluation by stepwise-forward method after rejection of not significant, non-linear X_{32}, X_{56}^{2} and multicollinear X_{124} variables

Source: created by the author using EViews

As all the assumptions mentioned above are satisfied, the regression model creation for prediction of GDP has been finished and it is:

$$Y = 0.002826 * X_{127} + 0.473039 * X_{147} + 4.573546 * X_{153}. \quad (6)$$

The real tendency (“Actual” line) and predicted values of GDP (“Fitted” line) using (6) model are shown in Figure 3. The residuals are also plotted in this graph.

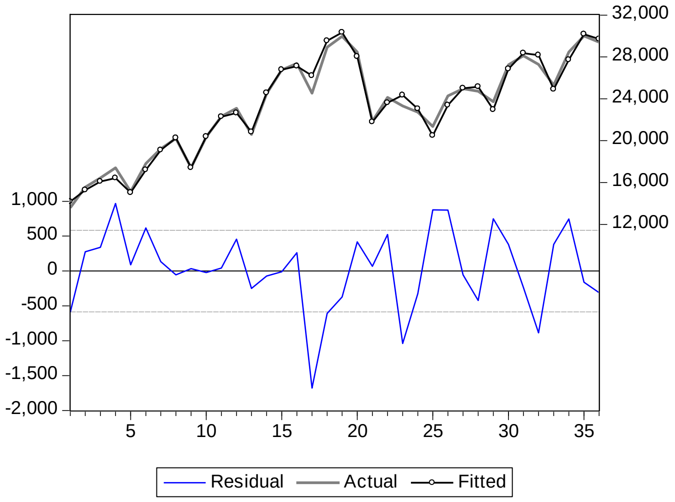

Figure 3. The real tendency of GDP, predicted values of GDP and residuals

Source: created by the author using EViews

In order to say that this model is suitable for prediction with other values of independent variables (that were not used for model creation), several assumptions for the residuals must be checked:

- the mean of the residuals must be equal to zero;
- residuals must be normally distributed;
- they must be uncorrelated;
- they must be homoscedastic.

The residual graph (Figure 3) shows that residuals are distributed in the interval across the zero and the average value of the residuals is close to 0. Jargue-Bera criterion was chosen for the normality test. In this case it is equal to 4.15 and the probability of this test is 0.13. As it is higher than significant level (0.05) it means that the residuals of this regression model are normally distributed.

Breusch-Godfrey test can be used for evaluation of correlation between residuals. $n \cdot R^2$ statistics (n – number of observations, R^2 – coefficient of determination) and $\chi^2_{0.05}$ criterion are used for this test. The comparison of the values of calculated and critical $\chi^2_{0.05}$ shows that the residuals are not correlated. The existence of homoscedasticity is checked by Breusch-Pagan-Godfrey test. It also uses $n \cdot R^2$ and $\chi^2_{0.05}$ statistics. The comparison of these values shows that the residuals are homoscedastic.

So the analysis of residuals confirmed that the (6) regression model is suitable for prediction of GDP.

Time series model creation for the prediction of Lithuanian GDP

Time series models are usually used when the values of the variable are changing accidentally over the time and there is hard to find the influential factors in order to make the regression model or it is not significant. The most popular method in this case is integrated autoregressive moving average (ARIMA) method.

The first step in time series model formation is stationarity test for the process of the tendency of GDP. It can be done by augmented Dickey-Fuller unit root test. It lets to analyze three types of processes:

- the process without trend and intercept,
- the process with intercept but no trend,
- the process with trend and intercept.

Augmented Dickey-Fuller test statistic for the first process (no trend and no intercept) is equal to 1.01 while test's critical value at the significant level of 0.05 is smaller (-1.95). It means that the primary process of the tendency of GDP is not stationary. Similar results are got when analyzing another two processes. It means that the series of GDP must be differentiated. The ΔY_t values are calculated for this purpose:

$$\Delta Y_t = Y_t - Y_{t-1} \tag{7}$$

This change is first order differential. After the first differentiation augmented Dickey-Fuller test statistic for the process that hasn't trend and intercept becomes equal to -2.24 while test's critical value is -1.95. So the process after the first differentiation becomes stationary. The results shows that the model

$$\Delta^2 Y_t = -0.649 \cdot \Delta Y_{t-1} + 0.064 \cdot \Delta^2 Y_{t-1} - 0.084 \cdot \Delta^2 Y_{t-2} - 0.198 \cdot \Delta^2 Y_{t-2} + 0.56 \cdot \Delta^2 Y_{t-4} \tag{8}$$

can be used for prediction of GDP. Here $\Delta^2 Y_t$ is the second order differential that is calculated as

$$\Delta^2 Y_t = Y_t - 2 \cdot Y_{t-1} + Y_{t-2} \, . \tag{9}$$

The accuracy of such model is 90%. If more precise model is needed, ARIMA(p,d,q) model can be used. As the process of the tendency of GDP becomes stationary after the first order differentiation, ARIMA(p,1,q) model can be used. The comparison of various ARIMA models by changing the values of p and q is shown in Table 3.

Table 3. The comparison of ARIMA(p,d,q) models

Model	The probability of Fisher criterion	$\overline{R}^2$	AIC criterion	SIC criterion
ARIMA(1,1,0)	0.310	0.032	18.42	18.51
ARIMA(0,1,1)	0.094	0.083	18.35	18.43
ARIMA(1,1,1)	0.000	0.505	17.81	17.94
.....	.....	.....	.....	......
ARIMA(2,1,2)	0.000	0.682	17.40	17.53
ARIMA(3,1,1)	0.000	0.632	17.70	17.93
ARIMA(4,1,3)	0.000	0.854	16.97	17.34
ARIMA(5,1,4)	0.000	0.851	16.82	17.05
.....	.....	.....	.....	......
ARIMA(1,2,0)	0.007	0.210	19.11	19.20
ARIMA(0,2,1)	0.000	0.566	18.48	18.57
ARIMA(3,2,3)	0.000	0.915	17.25	17.57
ARIMA(4,2,4)	0.000	0.891	17.29	17.47
.....	.....	.....	.....	......
ARIMA(1,3,1)	0.000	0.712	19.26	19.40
ARIMA(1,3,2)	0.000	0.821	18.78	18.92
ARIMA(3,3,3)	0.000	0.952	17.48	17.77
ARIMA(3,3,4)	0.000	0.971	17.12	17.35
.....	.....	.....	.....	......
ARIMA(3,4,1)	0.000	0.980	17.79	18.02

Source: created by the author

Most ARIMA models for the first order integrated (differenciated) process are not significant (Fisher criterion is higher than 0.05). Only models with high values of p and q are significant, but the accuracy of such models is not high. That is why higher order differentiation is needed. If the accuracy of 98% must be sought as it was got in multiple regression model, the fourth order differentiation is needed. In such case ARIMA(3,4,1) is the simplest significant model. Evaluation results of this model are shown in Figure 4.

Dependent Variable: D(Y,4)
Method: Least Squares
Sample (adjusted): 2005Q4 2012Q4
Included observations: 29 after adjustments
Convergence achieved after 40 iterations
MA Backcast: 2005Q3

	Coefficient	Std. Error	t-Statistic	Prob.
C	-2.956977	9.022231	-0.327743	0.7459
AR(1)	-1.013411	0.064734	-15.65504	0.0000
AR(2)	-1.035097	0.080343	-12.88356	0.0000
AR(3)	-0.968307	0.064099	-15.10652	0.0000
MA(1)	-0.997492	2.22E-05	-44833.74	0.0000

R-squared	0.982887	Mean dependent var	197.2552
Adjusted R-squared	0.980034	S.D. dependent var	11539.06
S.E. of regression	1630.463	Akaike info criterion	17.78670
Sum squared resid	63801804	Schwarz criterion	18.02244
Log likelihood	-252.9072	Hannan-Quinn criter.	17.86053
F-statistic	344.6045	Durbin-Watson stat	2.294147
Prob(F-statistic)	0.000000		

Inverted AR Roots	-.02+1.00i	-.02-1.00i	-.97
Inverted MA Roots	1.00		

Figure 4. Evaluation results of the model ARIMA(3,4,1)
Source: created by the author using EViews

So the time series model for prediction of GDP can be

$$\Delta^4 Y_t = -2.957 - 1.0134 \cdot \Delta Y_{t-1} - 1.035 \cdot \Delta Y_{t-2} - 0.968 \cdot \Delta Y_{t-3} - \varepsilon_t - 0.997 \varepsilon_{t-1}. \quad (10)$$

As there is fourth order differentiation ($\Delta^4 Y_t$) the equation of Y_t calculation is complicated. So the prediction is possible only with statistical software.

The last step is to check the residuals. They must satisfy four assumptions that were mentioned while analyzing the multiple regression model. The analysis of residuals shows that the mean is close to 0, they are normally distributed, homoscedastic and uncorrelated so ARIMA(3,4,1) model is suitable for the prediction.

Forecast of Lithuanian GDP

Multiple regression model (6) and ARIMA(3,4,1) model (10) were used to forecast Lithuanian GDP for the first quarter of 2013 year. The results are that it will be 26,47 bln. LTL by multiple regression model and 27.79 bln. LTL by ARIMA(3,4,1) model while the real GDP was 26.32 bln. LTL. The comparison of forecasts and the real value of GDP shows that multiple regression model is more precise for prediction of GDP. All these graphs are showed in Figure 5.

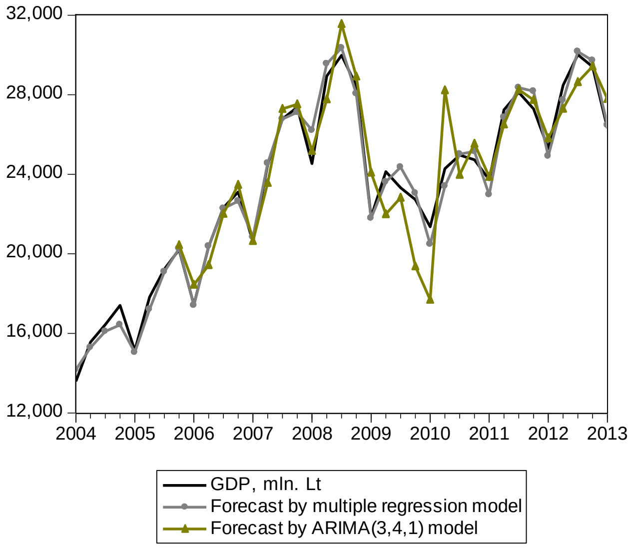

Figure 5. The comparison of forecasts by multiple regression model (5) and ARIMA(3,4,1) and the real GDP

Source: created by the author using EViews

4. Discussion

Official estimates of GDP are released with a considerable delay. The preliminary evaluation of the previous quarter is published after a month, and the exact evaluation is notified only after four months. Meanwhile, the economic analysis must rely on the monthly indicators, which arrive within the quarter such as, e.g. industrial production, retail sales and trade, surveys, monetary and financial data. That’s why Lithuanian Statistics data that are published every month were chosen for the prediction of the Lithuanian GDP model creation.

The analysis of suitability of the multiple regression models and times series models (ARIMA models) for the prediction of Lithuanian GDP shows that the multiple regression model is more appropriate for this purpose. Lithuanian GDP can be forecasted by three indicators: turnover (VAT excluded) of retail trade, except of motor vehicles and motorcycles, at current prices (in LTL thousands), passenger arrivals and departures at airports (in thousands) and goods loading (in thousand tones). In the process of the prediction of GDP these indicators must be forecasted for the quarter before GDP could be forecasted for the same quarter. Thus, the precision of the forecast by the multiple regression model depends on the accuracy of prediction of these three independent indicators. The good thing is that these indicators are calculated and published every month, so the prediction during the quarter can be corrected after the data of intermediate months are published. The construction of an appropriate model for the prediction of three independent indicators mentioned above will be the matter of the next research.

References

Almon, C. (2008). The Craft of Economic Modeling. Part 1. Fifth edition. *Interindustry Economic Research Fund, Inc.*

Andrei, E. A., Bugudui, E. (2011). Econometric Modeling of GDP Time Series. *Theoretical and Applied Economics,* vol. XVIII, no. 10(563), p. 91-98.

Angelini, E., Camba-Méndez, G., Giannone, D., Rünstler, G., Reichlin, L. (2008). Short-term Forecasts of Euro Area GDP Growth. *European Central Bank*. Working Paper Series, no. 949 / October 2008.

Baffigi, A., Golinelli, R., Parigi, G. (2004). Bridge models to forecast the euro area GDP. *International Journal of Forecasting.* vol. 20 (3), p. 447-460. http://dx.doi.org/10.1016/S0169-2070(03)00067-0

Barhoumi, K., Benk, S., Cristadoro, R., Reijer, A.D., Jakaitiene, A., Jelonek, P., Rua, A., Rünstler, G., Ruth, K., Nieuwenhuyze, C.V. (2008). Short-term forecasting of GDP using large monthly datasets: a pseudo real-time forecast evaluation exercise. *European Central Bank*. Occasional Paper Series, no 84 / April 2008.

Barhoumi, K., Brunhes-Lesage, V., Darné, O., Ferrara, L., Pluyaud, B., Rouvreau, B. (2008). Monthly Forecasting of French GDP: a Revised Version of the Optim Model. *Banque De France.*

Batchelor, R. A. (2007). Bias in macroeconomic forecasts. *International Journal of Forecasting*, 23, p. 189-203. http://dx.doi.org/10.1016/j.ijforecast.2007.01.004

Boero, G., Smith, J., Wallis K. F. (2008). Uncertainty and Disagreement in Economic Prediction: the Bank of England Survey of External Forecasters. *Economic Journal*, July. http://dx.doi.org/10.1111/j.1468-0297.2008.02162.x

Bratu, M. (2012). Improving the accuracy of consensus forecasts for the EURO area. *Review of Applied Socio- Economic Research.* ISSN: 2247-6172; ISSN-L: 2247-6172, vol. 4, issue 2/2012, p. 15.

Camacho, M., Martinez-Martin, J. (2012). Real-time forecasting US GDP from small-scale factor models. *BBVA Research.* 12/10 Working Papers, no 12/10. June 2012.

Celiku, E., Kristo, E., Boka, M. (2009). Modelling the quarterly GDP - role of economic and surveys indicators. *Bank of Albania.* ISBN: 978-99956-42-18-1

Celov, D., Vilkas, E., Grinderslev, D., Andersen, F. M. (2005). Macro-econometric model for Lithuania. *Economic Modelling,* 22(4), July, p. 707–719. http://dx.doi.org/10.1016/j.econmod.2005.05.001

Cheung, C., Demers, F. (2007). Evaluating Forecasts from Factor Models for Canadian GDP Growth and Core Inflation. *Bank of Canada.* Working Paper, 2007-8. ISSN 1701-9397

Chvatalova, Z., Simberova, I. (2011). Economic Phenomena via Mathematical Modelling in Maple System**.** *Business, Management and Education*. ISSN 2029-7491 print / ISSN 2029-6169 online, no. 9(2), p. 260–276. http://dx.doi.org/10.3846/bme.2011.18

Clopper, A. (2012). The Craft of Economic Modelling. Fifth edition. *Interindustry Economic Research Fund, Inc.* p. 149.

Diron, M. (2008). Short-term forecasts of Euro area real GDP growth: an assessment of real-time performance based on vintage data. *Journal of Forecasting,* no. 27, p. 371-390. http://dx.doi.org/10.1002/for.1067

Forni, M., Hallin, M., Lippi, M., Reichlin, L. (2005). The generalized dynamic factor model: one-sided estimation and forecasting. *Journal of the American Statistical Association*, no. 100, p. 830-840. http://dx.doi.org/10.1198/016214504000002050

Giannone, D., Reichlin, L., Sala, L. (2004). Monetary Policy in Real Time. *NBER Macroeconomics Annual 2004. MIT Press,* p. 161-200.

Giannone, D., Reichlin, L., Small, D. (2005). Nowcasting GDP and inflation: the real-time informational content of macroeconomic data releases. *Federal Reserve Bank Finance and Economics Discussion Series 2005-42*.

Guegan, D., Rakotomarolahy, P. (2010). Alternative methods for forecasting GDP. *CES Working papers. Documents de Travail du Centre d'Economie de la Sorbonne, 2010.65.* ISSN : 1955-611X. halshs-00505165, version 1.

Karpavicius, S. (2008). Kalibruotas Lietuvos ekonomikos dinaminis stochastinis bendrosios pusiausvyros modelis. *Pinigų studijos,* Nr. 2, p. 22-46.

Karpavicius, S. (2009). The Effects of Fiscal Instruments on the Economy of Lithuania. *Lietuvos Bankas.* Working Paper Series No 4.

Karpavicius, S., Vetlov I. (2008). Personal Income Tax Reform: Macroeconomic and Welfare Implications. *Lietuvos Bankas.* Working Paper Series No 2.

Klučik, M., Juriová, J. (2010). Slowdown or Recession? Forecasts Based on Composite Leading Indicator. *Central European Journal of Economic Modelling and Econometrics*, vol. 2. p. 17-36 .

Laurent, T., Kozluk, T. (2012), Measuring GDP Forecast Uncertainty Using Quantile Regressions. *OECD Economics Department Working Papers*, No. 978. http://dx.doi.org/10.1787/5k95xd76jvvg-en.

Rukšėnaitė, J. (2010). Sudėtiniai rodikliai – papildomas matas Lietuvos ekonomikai vertinti. *Lietuvos statistikos darbai,* 2010 (49).

Runstler, G., Barhoumi, K., Benk, S., Cristadoro, R., Den Reijer, A., Jakaitiene, A., Jelonek, P., Rua, A., Ruth, K., Van Nieuwenhuyze, C. (2008). Short-Term Forecasting of GDP Using Large Monthly Datasets: a Pseudo Real-Time Forecast Evaluation Exercise. *Lietuvos Bankas.* Working Paper Series No 1.

Runstler, G., & Sedillot, F. (2003). Short-term estimates of euro area real GDP by means of monthly data. *ECB Working Paper,* no 276.

Schumacher, C. (2005). Forecasting German GDP using alternative factor models based on large datasets. *Deutsche Bundesbank.* Discussion Paper. Series 1: Economic Studies. No 24/2005.

Smets, F., Wouters R. (2004). Forecasting with a Bayesian DSGE model: An application to the Euro area. *Journal of Common Market Studies,* no. 42, p. 841–867. http://dx.doi.org/10.1111/j.0021-9886.2004.00532.x

Stankevičienė, J., Gruodis, A., Lokutijevskij, A., Urbaitė, Ū. (2012). Building Lithuanian Macro-Econometric Model: Forecast of Average Wages and Unemployment Rate. *Intellectual Economics.* ISSN 1822-8011, vol. 6, no. 1(13), p. 754–775.

Startz, R. (2013). EViews Illustrated for Version 8. *IHS Global Inc.* p. 434.

Stock, J. H., Watson, M. (2002). Forecasting using principal components from a large number of predictors. *Journal of the American Statistical Association,* no. 97, p. 1167-1179. http://dx.doi.org/10.1198/016214502388618960

Tamosiunas, R. (1999). Bendrosios pusiausvyros modelių taikymo galimybės. *Pinigų studijos,* Nr. 1, p. 30-45.

Taht, M. (2008). Forecasting of Economic Growth in Statistics Estonia: Practical Aspects. 5th EUROSTAT colloquium on modern tools for business cycle analysis. Jointly organized by EUROSTAT and European University Institute (Florence, Italy). Luxembourg, 29th September – 1st October 2008.

Turturean, C. I. (2007). The Econometric Modeling of the Dynamics of the Romanian Economic Growth During the Period 1990-2004. *Analele Stiintifice ale Universitatii "Alexandru Ioan Cuza" din Iasi - Stiinte Economice*, vol. 54, p. 160-166. http://ssrn.com/abstract=1048141 or http://dx.doi.org/10.2139/ssrn.1048141.

Vilkas, E. (1997). Ekonominės pusiausvyros skaičiavimai: Lietuvos energetika. *Pinigų studijos,* Nr. 1, p. 51-60.

Virbukaitė, S. (2011). Vektorinės autoregresijos (VAR) modelių taikymas ekonomikos procesams analizuoti. *14-osios Lietuvos jaunųjų mokslininkų konferencijos „Mokslas – Lietuvos ateitis" 2011 metų teminės konferencijos straipsnių rinkinys*. ISSN 2029-7149 online. ISBN 978-9955-28-836-7.