Scientific Journal of Riga Technical University
Computer Science. Technologies of Computer Control

2010
Volume 42

The Search Algorithm of Optimal Time Series Model for Forecasting Latvian GDP

Aleksandrs Bezrucko, *Riga Technical University*

***Abstract*: The search algorithm of optimal time series model for forecasting Latvian GDP. In this scientific paper there is developed algorithm for a finding optimal time series model for GDP forecasting. Latvian GDP statistical data with quarterly observation frequency is taken as a time series. ARMA Analysis of Latvian GDP Time series is performed and described. The set of model has been constructed. For check of quality of models Residual tests are performed and models are compared between themselves. Using econometric software EViews 6.0 forecasts for best models are made and results are compared with real data of last three squares of year 2009.**

***Keywords*: time series, GDP (Gross Domestic Product), ARMA (Autoregressive Moving Average) Analysis, Residual tests, Serial Correlation, Heteroskedasticity.**

I. INTRODUCTION

The analysis and forecasting of Gross Domestic Product was an actual object of research for any time and any modern country. These researches are consisting of many objective and subjective factors. For forecasting econometrists can not use only statistical methods but need also take in consideration a lot of economical and political events.

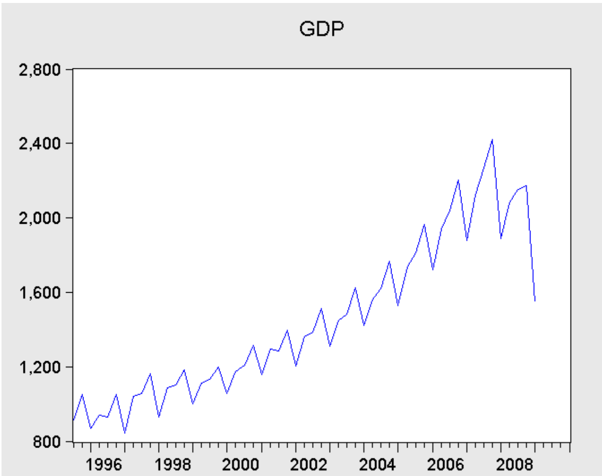

Fig. 1. Latvian GDP (Lats) 1995Q1-2009Q1

Working with this scientific paper different methods of econometrical modeling are analyzed. For example analysis methods for German GDP forecast that are described by Lutkepohl in "Applied Time Series Analysis" [1]. Lutkepohl is describing different ways of ARMA and Residual analysis of time series. In this scientific paper author use known methods of statistical analysis of time series for forecasting Latvian GDP. Using a computer software author perform search for the best models for a concrete time series. Based on analysis of these models a search algorithm of optimal model is build.

For a finding optimal model of forecasting Latvian Gross Domestic Product two different cases of Latvian GDP series with quarterly observation frequency is taken. The first case is quarterly Latvian GDP series in levels (Latvian lats) and second case is same data in percentage growth. The GDP series are given in Figure 1. The time series length is T = 57. The time series is taken from first quarter of year 1996 till year first quarter of year 2009. All searches and forecasts are made using econometrical software EViews 6.0.

II. EVIEWS SOFTWARE

EViews offers academic researchers, corporations, government agencies, and students access to powerful statistical, forecasting, and modelling tools through an innovative, easy-to-use object-oriented interface.

A combination of power and ease-of-use make EViews software the ideal package for anyone working with time series, cross-section, or longitudinal data. With EViews, is possible quickly and efficiently manage data, perform econometric and statistical analysis, generate forecasts or model simulations, and produce high quality graphs and tables for publication or inclusion in other applications.

EViews supports a wide range of basic statistical analyses, encompassing everything from simple descriptive statistics to parametric and nonparametric hypothesis tests.

Basic descriptive statistics are quickly and easily computed over an entire sample, by a categorization based on one or more variables, or by cross-section or period in panel or pooled data. Hypothesis tests on mean, median and variance may be carried out, including testing against specific values, testing for equality between series, or testing for equality within a single series when classified by other variables (allowing you to perform one-way ANOVA). Tools for covariance and factor analysis allow you to examine the relationships between variables.

EViews easy visualize the distribution of data using histograms, theoretical distribution, kernel density, or cumulative distribution, survivor, and quantile plots. QQ-plots (quantile-quantile plots) may be used to compare the distribution of a pair of series, or the distribution of a single series against a variety of theoretical distributions.

EViews also produces scatter plots with curve fitting using ordinary, transformation, kernel, and nearest neighbor regression. [4]

III. Working Plan

1. Using EViews 6.0 software performs ARMA analysis of Time series. Find models with best statistical criteria's.
2. Perform an analysis of criteria. Find the best models.
3. Perform residual tests. Three criteria are analyzed: Akaike info, Schwarz and Hannan-Quinn criteria's.
4. Make Forecast for time series for each model. Compare it with real data.
5. Conclusions.

IV. Analysis Description

A. The analysis of criteria

At the first stage of a choice of the best model, 3 criteria are analyzed: Akaike info, Schwarz and Hannan-Quinn. The best model has the minimal values. R-squared statistic is also present. In this stage models with best criteria are taken. ARMA Analysis is realized on EViews program language and statistical criteria's are a result of the program (Fig.2).

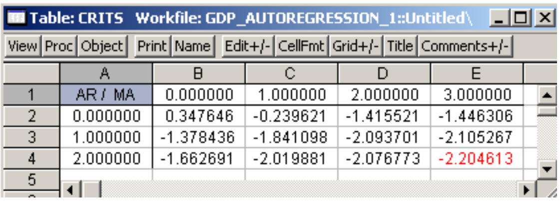
Table: CRITS Workfile: GDP_AUTOREGRESSION_1::Untitled\
View | Proc | Object | Print | Name | Edit+/- | CellFmt | Grid+/- | Title | Comments+/-

	A	B	C	D	E
1	AR / MA	0.000000	1.000000	2.000000	3.000000
2	0.000000	0.347646	-0.239621	-1.415521	-1.446306
3	1.000000	-1.378436	-1.841098	-2.093701	-2.105267
4	2.000000	-1.662691	-2.019881	-2.076773	-2.204613
5					

Fig. 2. ARMA Analysis in EViews 6.0

The R-squared (R^2) statistic measures the success of the regression in predicting the values of the dependent variable within the sample. In standard settings, R^2 may be interpreted as the fraction of the variance of the dependent variable explained by the independent variables. The statistic will equal one if the regression fits perfectly, and zero if it fits no better than the simple mean of the dependent variable. It can be negative for a number of reasons. For example, if the regression does not have an intercept or constant, if the regression contains coefficient restrictions, or if the estimation method is two-stage least squares or ARCH.

The Akaike Information Criterion (AIC) is computed as: $AIC = -2l/T + 2k/T$ where l is the log likelihood. The AIC is often used in model selection for non-nested alternatives-smaller values of the AIC are preferred. For example, you can choose the length of a lag distribution by choosing the specification with the lowest value of the AIC.

The Schwarz Criterion (SC) is an alternative to the AIC that imposes a larger penalty for additional coefficients: $SC = -2l/T + (k\log T)/T$

B. Residual tests

The second stage is Residual tests: Serial Correlation LM test, Histogram – Normality test, Hetereskedasticity ARCH test and Correlogram Square Residual test. Models evaluated as passed the test if P-value is higher then 0,1.

Serial Correlation LM test is an alternative to the *Q*-statistics for testing serial correlation. The test belongs to the class of asymptotic (large sample) tests known as Lagrange multiplier (LM) tests Serial Correlation LM test has the higher importance because on this step we are concerning with the possibility that our errors exhibit autocorrelation. LM test check for higher order ARMA errors and is applicable whether or not there are lagged dependent variables.

The null hypothesis of the LM test is that there is no serial correlation up to lag order p, where p is a pre-specified integer. The local alternative is ARMA(r, q) errors, where the number of lag terms p =max(r, q). Note that this alternative includes both AR(p) and MA(p) error processes, so that the test may have power against a variety of alternative autocorrelation structures.

The test tatistic is computed by an auxiliary regression as follows. First, suppose you have estimated the regression;

$$y_t = X_t\beta + \epsilon_t \qquad (1)$$

where b are the estimated coefficients and ϵ are the errors. The test statistic for lag order p is based on the auxiliary regression for the residuals $e = y - X\beta$:

$$e_t = X_t\gamma + \left(\sum_{s=1}^{p} \alpha_s e_{t-s}\right) + v_t \qquad (2)$$

Histogram and normality tests are displays a histogram and descriptive statistics of the residuals, including the Jarque-Bera statistic for testing normality. If the residuals are normally distributed, the histogram should be bell-shaped and the Jarque-Bera statistic should not be significant. The Jarque-Bera statistic has a χ^2 distribution with two degrees of freedom under the null hypothesis of normally distributed errors. [2]

The ARCH test is a Lagrange multiplier (LM) test for autoregressive conditional heteroskedasticity (ARCH) in the residuals. This particular heteroskedasticity specification was motivated by the observation that in many financial time series, the magnitude of residuals appeared to be related to the magnitude of recent residuals. ARCH in itself does not invalidate standard LS inference. However, ignoring ARCH effects may result in loss of efficiency.

The ARCH LM test statistic is computed from an auxiliary test regression. To test the null hypothesis that there is no ARCH up to order q in the residuals, we run the regression:

$$e_t^2 = \beta_0 + \left(\sum_{s=1}^{q} \beta_s e_{t-s}^2\right) + v_t, \qquad (3)$$

where e is the residual. This is a regression of the squared residuals on a constant and lagged squared residuals up to order q. The *F*-statistic is an omitted variable test for the joint significance of all lagged squared residuals. The Obs*R-squared statistic is Engle's LM test statistic, computed as the number of observations times the R^2 from the test regression. The exact finite sample distribution of the *F*-statistic under

H_0 is not known, but the LM test statistic is asymptotically distributed as a $\chi^2(q)$ under quite general conditions.

Correlogram of squared residuals test displays the autocorrelations and partial autocorrelations of the squared residuals up to any specified number of lags and computes the Ljung-Box Q-statistics for the corresponding lags. The correlograms of the squared residuals can be used to check autoregressive conditional heteroskedasticity (ARCH) in the residuals.

If there is no ARCH in the residuals, the autocorrelations and partial autocorrelations should be zero at all lags and the Q-statistics should not be significant inclusion of ARMA terms. [3]

C. Out-Of-Samle Forecasting

The final evaluating test is "Out-Of-Sample Forecasting". In this step forecasts is comparing with real date. Real data that we have is for the last 3 quarters of 2009.

V. Latvian GDP in Levels

The series to be analyzed consist of seasonally not adjusted, quarterly Latvian GDP in levels for the period 1995Q1 – 2009Q1. It is plotted in Figure 1. Constructing a model for the logs is likely to be advantageous because the changes in the log series display a more stable variance than the changes in the original series. Time series in logs is plotted in Figure 3

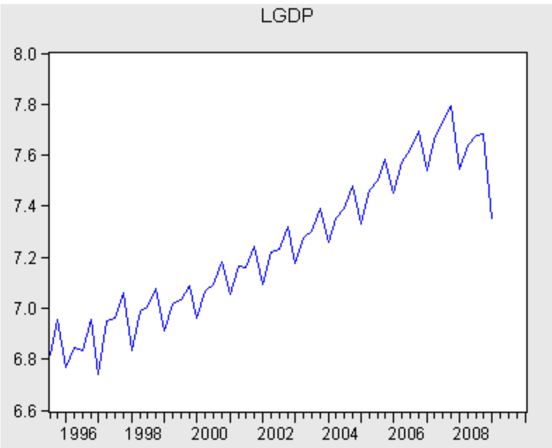

Fig. 3. Latvian GDP in logs

Table 1

The analysis of criteria (Levels)

Nr.	LGDP	Akaike	Schwarz	Han-Quinn
1	Ar(1)	-1.326483	-1.290316	-1.312461
2	Ar(2)	-1.487976	-1.451479	-1.473862
3	Ma(1) C	-0.182363	-0.110677	-0.154503
4	Ma(2) C	-0.143123	-0.071437	-0.115263
5	AR(1) MA(1)	-1.878947	-1.806613	-1.850903
6	AR(1) AR(2) MA(1) MA(2)	-2.087592	-1.941604	-2.031137
7	AR(1) AR(2) SAR(4) MA(1)	-3.746621	-3.595105	-3.688722
8	AR(1) AR(2) SAR(4) MA(4)	-4.18207	-4.030555	-4.124172
9	Trend C	-1.914989	-1.843303	-1.887129
10	AR(1) Trend	-1.294331	-1.221997	-1.266287
11	AR(2) SAR(4)	-3.027406	-2.951648	-2.998456
12	AR(1) SAR(4) MA(4)	-3.646746	-3.534175	-3.603589
13	AR(1) AR(2) SAR(4) MA(4) SEAS(1)	-4.224953	-4.035559	-4.15258
14	AR(1) SAR(4) MA(4) D1997Q2	-4.047836	-3.89774	-3.990293

Best models: Nr. 5,7,8,12,13,14. Other models quit from further evaluation process.

The residual tests results are given in Table 2. Models Nr.13 and Nr.14 are complete all the tests. Residuals graph of Model Nr. 13 is given in Figure 4.

Table 2

Residual Test (Levels)

Nr.	Serial Correlation	Histogram	Heteroskedasticity
5	0.69290	0.00000	0.61360
7	0.69350	0.00000	0.79830
8	0.08070	0.40000	0.53920
12	0.61610	0.00003	0.84840
13	0.09250	0.83015	0.87060
14	0.10670	0.40223	0.85250

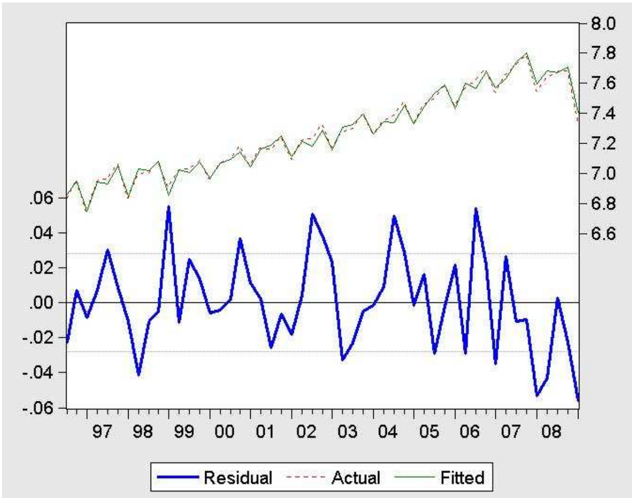

Fig. 4. Residuals graph for Model Nr.13

Worst model residuals are given in Figure 5 for comparison.

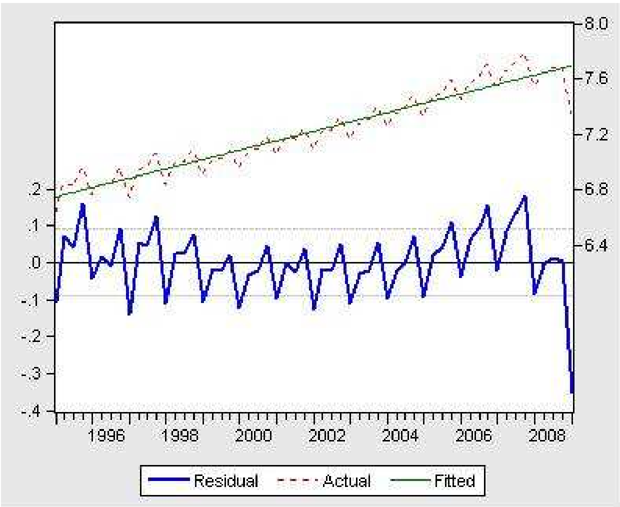

Fig. 5. Residuals graph for Model Nr.9

"Out-Of-Sample Forecasting" test (Table 3) show that forecasts more closes to reality we get from models Nr.12: AR(1) SAR(4) MA(4). Absolute difference (0.037) is minimal in this case. The best model has passed all Residual tests except Normality. The second and third result has models Nr. 7 and Nr.13., also passed almost all residual tests except Nr.7. which also did not pass the Normality test.

TABLE 3

OUT-OF-SAMPLE FORECASTING (LEVELS)

	Forecast - 1 Step - 09q2			Forecast - 2 Step -09q3			Forecast - 3 Step -09q4			
Nr.	LGDP real	LGDPf	Diff	LGDP real	LGDPf	Diff	LGDP real	LGDPf	Diff	| Sum % |
5	0.088	0.435	**-0.347**	0.031	0.022	**0.008**	0.034	0.018	**0.016**	0.371
7	0.088	0.082	**0.007**	0.031	0.017	**0.014**	0.034	-0.012	**0.045**	0.066
8	0.088	0.044	**0.044**	0.031	0.023	**0.007**	0.034	-0.020	**0.054**	0.106
12	0.088	0.103	**-0.014**	0.031	0.042	**-0.012**	0.034	0.023	**0.011**	0.037
13	0.088	0.057	**0.031**	0.031	0.017	**0.014**	0.034	-0.021	**0.055**	0.099
14	0.088	0.050	**0.039**	0.031	0.013	**0.018**	0.034	-0.019	**0.053**	0.109

VI. LATVIAN GDP IN PERCENTAGE GROWTH

All evaluations are made with Latvian GDP. The differences and the log difference of time series are plotted in Figure 6. The log difference display a more stable variance than the changes in the original series that why it is taken.

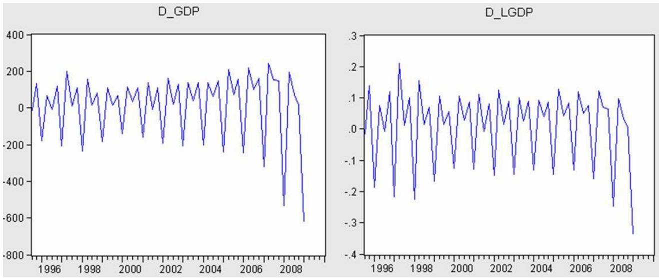

Fig. 6. Time series in differences and in log differences

TABLE 4

THE ANALYSIS OF CRITERIA (DIFFERENCE)

Nr.	d (LGDP)	Akaike	Schwarz	Han-Quinn
1	Ar(1)	-1.659793	-1.623296	-1.645679
2	Ar(2)	-1.379102	-1.342269	-1.364897
3	Ma(1) C	-1.876758	-1.804424	-1.848714
4	Ma(2) C	-1.309161	-1.236827	-1.281117
5	AR(1) MA(1)	-1.712348	-1.639354	-1.684121
6	AR(1) AR(2) MA(1) MA(2)	-2.033619	-1.886286	-1.976798
7	AR(1) AR(2) SAR(4) MA(1)	-3.728444	-3.575482	-3.670195
8	AR(1) AR(2) SAR(4) MA(4)	-4.197451	-4.044489	-4.139202
9	Trend C	-1.306934	-1.2346	-1.278891
10	AR(1) Trend	-1.636801	-1.563807	-1.608574
11	AR(2) SAR(4)	-3.803588	-3.727107	-3.774463
12	AR(1) SAR(4) MA(4)	-4.191773	-4.078136	-4.148349
13	AR(1) AR(2) SAR(4) MA(4) SEAS(1)	-4.307962	-4.116759	-4.235151
14	AR(1) SAR(4) MA(4) D1997Q2	-4.273152	-4.121636	-4.215253

Models with best criteria: Nr. 7, 8, 11, 12, 13, 14. Other models quit from further evaluation process.

TABLE 5

RESIDUAL TEST (DIFFERENCE)

Nr.	Serial Correlation	Histogram	Heteroskedasticity: ARCH
7	0.6938	0	0.825
8	0.6895	0.393373	0.3453
11	0.8852	0	0.017538
12	0.0503	0.421976	0.6305
13	0.64	0.560093	0.2823
14	0.5554	0.997972	0.7499

Models Nr. 8, 13, 14 are complete all the tests. Model Nr. 12 also have good statistic.

"Out-Of-Sample" forecasting test (Table 6) show that forecasts more closes to reality we get from models Nr.11: AR(2) SAR(4). Absolute difference (0.037) is minimal in this case. This model did not complete the histogram test, but pass all other residual tests. Second result has model Nr.7, which has same problem with Normality test. Third results has model Nr.14 – this model pass all residual tests.

TABLE 6

OUT-OF-SAMPLE FORECASTING (DIFFERENCE)

	Forecast - 1 Step - 09q2			Forecast - 2 Step -09q3			Forecast - 3 Step -09q4			
Nr.	dLGDP real	dLGDPf	Diff	dLGDP real	dLGDPf	Diff	LGDP real	LGDPf	Diff	| Sum % |
7	0.088	0.084	**0.004**	0.031	0.022	**0.009**	0.034	0.004	**0.030**	0.043
8	0.088	0.034	**0.054**	0.031	0.019	**0.012**	0.034	-0.020	**0.054**	0.120
11	0.088	0.091	**-0.003**	0.031	0.024	**0.007**	0.034	0.007	**0.027**	0.037
12	0.088	0.045	**0.043**	0.031	0.024	**0.007**	0.034	-0.020	**0.054**	0.104
13	0.088	0.022	**0.067**	0.031	0.009	**0.022**	0.034	-0.021	**0.054**	0.143
14	0.088	0.070	**0.018**	0.031	0.011	**0.020**	0.034	-0.020	**0.054**	0.091

VII. THE SEARCH ALGORITHM

The search algorithm was shown at Figure 7. Step by step it can be described as:

Input data: GDP Time series

1. Constructing of ARMA models in levels and in differences separately. Since this point the model is divided on two branches and the subsequent steps are carried out in parallel for levels and for differences.

2. ARMA Analysis

3. Performing Residual tests

4. If during residual tests probability value is less then 10% model quit from further evaluation. We can not trust to this model.

5. If P Value is higher then 10% go throw and perform forecasting for specified periods of time.

6. Comparing forecasting data with real data. Make evaluation.

7. The analysis of the results. Two branches of models come back to one point.

Output data: Best model for the GDP Forecasts.

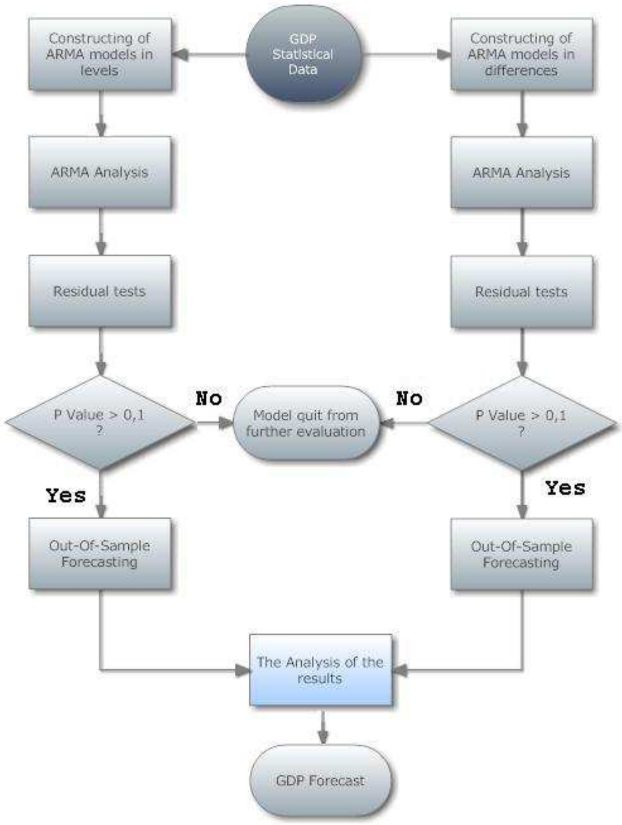

Figure 7 The search algorithm

VIII. CONCLUSION

In the scientific paper author describe search algorithm of optimal time series. Using statistical modeling, the econometric analysis of Latvian Gross Domestic Product is made. Different cases of constructing model are made in Latvian lats (in levels) and in percentage growth (in difference).

Comparison of 1 and 2 cases shows that case in levels and in differences gave approximately same result 3.7% deviation of real data in absolute value for forecasts for 3 steps in future. It is seems that it is does not matter which way to use. But of course it is not the right design. It is very important to look all the result together and understand how they are calculated and evaluated. If a model give a best forecast for one, two or three steps separately it is does not mean that this model will be best in other cases. Figure 7 shows the algorithm of searching optimal model for forecasting.

REFERENCES

[1] Lutkepohl H., Kratzig M. Applied Time Series Econometrics. – Cambridge: Cambridge University press, 2004, 323 pp.
[2] Nosko V., Econometrics: introduction to regression analysis of time series. – Moscow: Moscow Institute of Physics and Technology , 2002, 273 pp.
[3] Molcanov I., Arzenovski S. Statistical methods of forecasting. – Rostov-na-Donu: Rostov State University, 2001, 74 pp.
[4] EViews software documentation: Quantitative Micro Software– http://www.eviews.com – Resource is described 2010, 30 of august.
[5] Latvian GDP Statistical Data: The Bank of Latvia – http://www.bank.lv – Resource is described 2010, 30 of august.

Aleksandrs Bezrucko was born in Riga in 1982. He was educated at the Riga Technical University in Latvia where he earned Bachelor of Computer Science, Engineer in Information Technology and Master of Business Management degrees. Aleksandrs Bezrucko is currently completing his research for PhD at the Riga Technical University in the field of Mathematical Statistics and its Applications.
In 2006-2009 was working in Join Stock Company “SeverStaLat” as IT engineer. Since 2009 is working in e-commerce company “AutoBodyShop” as director.
Email: aleksandr@livas.lv

Aleksandrs Bezručko. Prognozēšanas Latvijas IKP laika rindu optimālā modeļa algoritma atrašana
Zinātniskajā darbā tiek izstrādāts laika rindu optimāla modeļa atrašanas algoritms Latvijas IKP prognozēšanai. Darbā izmantoti statistikas dati, kas iegūti Latvijas Republikas centrālā statistikas pārvaldē un Latvijas Bankā. Latvijas IKP statistikas dati ar gada ceturkšņa izpētes biežumu tiek ņemti laika rindu veidā. Izstrādāta un aprakstīta ARMA laika rindas analīze. Izveidots nepieciešamais modeļu daudzums. Pēc dažādiem statistikas kritērijiem izpētīts modeļu vērtējums, tai skaitā izmantoti Akaike un Švarca kritēriji, kuri ļauj izvēlēties vislabāko modeli no daudzām dažādām specifikācijām. Modeļu precizitātes pārbaudei tika veikti atlikumu autokorelācijas testi, heteroskedastitāte un pārbaudīts atlikumu normalitātes sadalījums. Izmantojot EViews 6.0. programmatūru uz labāko modeļu bāzes izstrādātas prognozes turpmākajiem trīs ceturkšņiem. Veikts prognožu salīdzinājums ar reāli faktiskajiem datiem par pēdējo 2009.gada 3 ceturkšņiem. Izstrādātais algoritms ļauj iegūt kopēju visu izstrādāto modeļu un to rezultātu kopskatu un tāpat iegūtos modeļus salīdzināt pēc līmeņiem un starpības. Darba rezultātā iegūts detalizēts algoritma apraksts un tā grafiskais īstenojums. Iegūtais algoritms var tikt izmantots kā analītisks instruments IKP prognozēšanai dažādās finansu un statistikas iestādes.

Александр Безручко. Алгоритм нахождения оптимальной модели временных рядов для прогнозирования ВВП Латвии
В научной статье разработан алгоритм нахождения оптимальной модели временных рядов для прогнозирования ВВП Латвии. Использованные в статье статистические данные получены в Центральном Статистическом Управлении Латвийской Республики и в Банке Латвии. Статистические данные ВВП Латвии с ежеквартальной частотой наблюдения взяты в качестве временного ряда (в уровнях и в разностях). Выполнен и описан ARMA-анализ временного ряда. Построено определённое множество моделей. Произведена оценка значимости моделей по различным статистическим критериям, в том числе использованы критерии Акайке и Шварца, позволяющие выбирать наилучшую модель из множества различных спецификаций. Для проверки точности моделей были проведены тесты на автокорелляцию остатков, гетероскедастичность и проверена нормальность распределения остатков. Проведено сравнение моделей между собой. Используя программное обеспечение EViews 6.0, на базе лучших моделей, построены прогнозы на три квартала вперёд. Проведено сравнение прогнозов с реальными данными последних трёх кварталов 2009 года. Разработанный алгоритм позволяет получить общую картину всех построенных моделей и их результатов, а также сравнить построение моделей в уровнях и в разностях. В результате работы приведено пошаговое описание алгоритма, а также его графическое изображение. Полученный алгоритм может быть применён как аналитический инструмент прогнозирования ВВП в различных финансовых и статистических учреждениях.