

LATVIAN GDP: TIME SERIES FORECASTING USING VECTOR AUTO REGRESSION

BEZRUCKO Aleksandrs, (LV)

Abstract: The target goal of this work is to develop a methodology of forecasting Latvian GDP using ARMA (AutoRegressive-Moving-Average) and VAR methods. The paper follows up with the papers published in the proceedings of the APLIMAT journal in 2011 – see [1]. The algorithm is developed for finding optimal time series model for GDP forecasting. Latvian GDP data with quarterly observation frequency is taken as time series. ARMA and VAR Analysis of Latvian GDP, M2X and inflation indicators time series is performed. The set of models has been constructed. In order to check the accuracy of models, different residual tests are performed: autocorrelation, Portmaneteau, heteroscedasticity and normality of residual distribution. Models are compared in their forecast quality.

Keywords and phrases: time series, Gross Domestic Product, Inflation, VAR (Vector Auto Regression), Residual tests, Serial Correlation

1 Introduction

The urgency of use of VAR models for an estimation of macroeconomic indicators doesn't cause today any doubts. It has been proved by And Sims (one of the main promoters of the use of vector autoregression in empirical macroeconomics, and contributed to the development of Bayes estimators for vector autoregression) in the works on use of VAR-models, which have received the The Nobel Prize in Economic Sciences in 2011 “*for their empirical research on cause and effect in the macroeconomy*”.

The analysis and forecast of GDP for any time and any country is important task for economists, policy makers and entrepreneurs. These researches are consisting of many objective and subjective factors.

The base of this work is taken from previous paper [1] and a research has continued taken into account Vector Auto Regression. Different methods of econometrical modelling have been analyzed. For example, analysis methods for German GDP forecast that are described by Lutkepohl in “Applied Time Series Analysis” [2]. Lutkepohl described different ways of ARMA and Residual analysis of time series. In this paper author uses familiar methods of statistical analysis of time series for forecasting Latvian GDP. Computer software enabled the author to perform the search for

the best models for certain time series. Based on the analysis of these models, a search algorithm of optimal model is created. A lot of attention is given to use VAR-models.

In order to find an optimal model of forecasting Latvian Gross Domestic Product, three different cases of Latvian GDP series with quarterly observation frequency are taken. The first case is quarterly Latvian GDP series in levels (Latvian lats), the second case is same data in percentage growth and in the third case different VAR models are build using Latvian GDP, Inflation and M2X cash mass data. The GDP series are given in Figure 1, Comparison with M2X and inflation data are given in Figure 2. The time series length is $T = 57$. The time series is taken from the first quarter of year 1995 till the first quarter of year 2009. All searches and forecasts are made using econometrical software EViews 6.0.

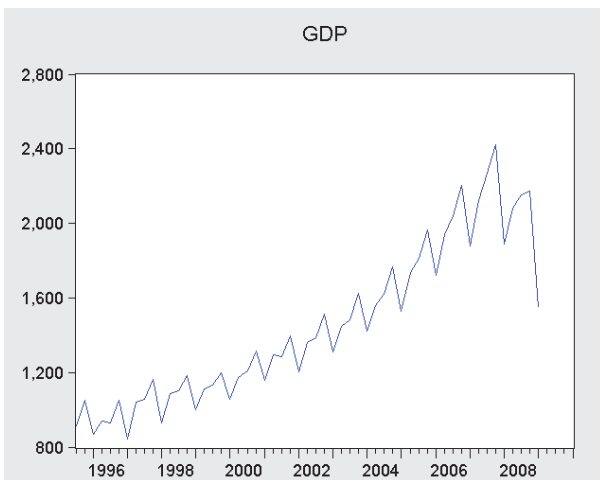


Figure 1 Latvian GDP (Lats) 1995Q1-2009Q1

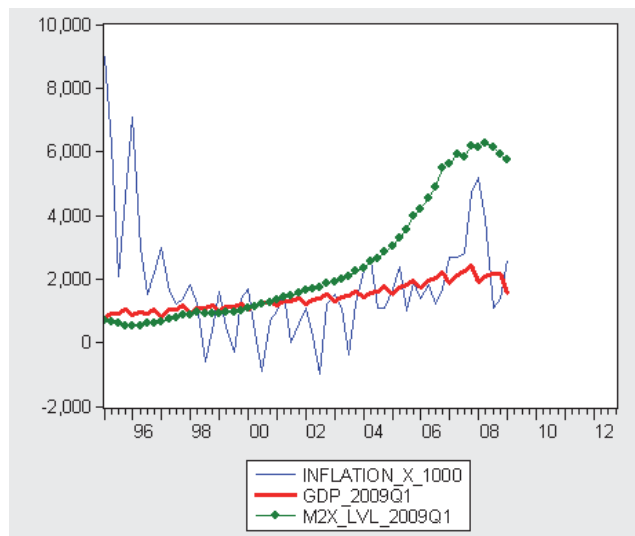


Figure 2 Inflation (%) / GDP (Lats) / M2X (Lats) 1995Q1-2009Q1

2 Analysis description

2.1. The analysis of criteria

ARMA case: At the first stage of choosing the best model, 3 criteria are analyzed: Akaike info, Schwarz and Hannan-Quinn.

VAR case: 2 criteria are analyzed: Akaike info and Schwarz

Constructing of ARMA and VAR models occurs in parallel, but separately from each other. The best model has minimal values. At this stage models with best criteria are taken. ARMA and VAR Analysis are performed in EViews program language and statistical criteria represent the result of the program (Fig.3).

Table: CRITS Workfile: GDP_AUTOREGRESSION_1::Untitled\						
View	Proc	Object	Print	Name	Edit+/-	CellFmt
		A	B	C	D	E
1		AR / MA	0.000000	1.000000	2.000000	3.000000
2			0.347646	-0.239621	-1.415521	-1.446306
3			-1.378436	-1.841098	-2.093701	-2.105267
4			-1.662691	-2.019881	-2.076773	-2.204613
5						

Figure 3 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: $SC = -2l/T + (k \log T)/T$

2.2. Residual tests

The second stage is represented by Residual tests:

ARMA case: Serial Correlation LM test, Histogram – Normality test, Heteroskedasticity ARCH test and Correlogram Square Residual test. Models have passed the test if P Value is higher than 0,1.

VAR case: VAR Residual Serial correlation LM test (Fig. 4.) and VAR Residual Portmanteau Tests for Autocorrelation.

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 \mathbf{p} , where \mathbf{p} is a pre-specified integer. The local alternative is ARMA(r,q) errors, where the number of lag terms $\mathbf{p} = \max(r,q)$. 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 statistic is computed by an auxiliary regression as follows. First, suppose you have estimated the regression:

$$y_i \beta + e_i$$

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

$$e_i = X_i \gamma + \left(\sum_{s=1}^{\mathbf{p}} \alpha_s e_{i-s} \right) + v_i$$

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:

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

where e_t 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. [2]

VAR Residual Serial Correlation LM T...
 Null Hypothesis: no serial correlation ...
 Date: 12/14/11 Time: 23:24
 Sample: 1995Q1 2012Q4
 Included observations: 55

Lags	LM-Stat	Prob
1	48.38674	0.0000
2	21.90037	0.0092
3	8.646018	0.4706
4	22.88185	0.0065
5	13.97551	0.1232
6	10.34989	0.3229
7	7.698525	0.5648
8	12.41758	0.1908
9	6.160805	0.7237
10	6.560951	0.6827
11	6.727117	0.6655
12	9.181946	0.4207

Probs from chi-square with 9 df.

Figure 4 VAR Residual Serial Correlation in EViews 6.0

Autocorrelation LM Test (Fig.4.)

Reports the multivariate LM test statistics for residual serial correlation up to the specified order. The test statistic for lag order h is computed by running an auxiliary regression of the residuals u_t on the original right-hand regressors and the lagged residual u_{t-h} , where the missing first h values of u_{t-h} are filled with zeros. See Johansen (1995, p. 22) for the formula of the LM statistic. Under the null hypothesis of no serial correlation of order h , the LM statistic is asymptotically distributed χ^2 with k^2 degrees of freedom.

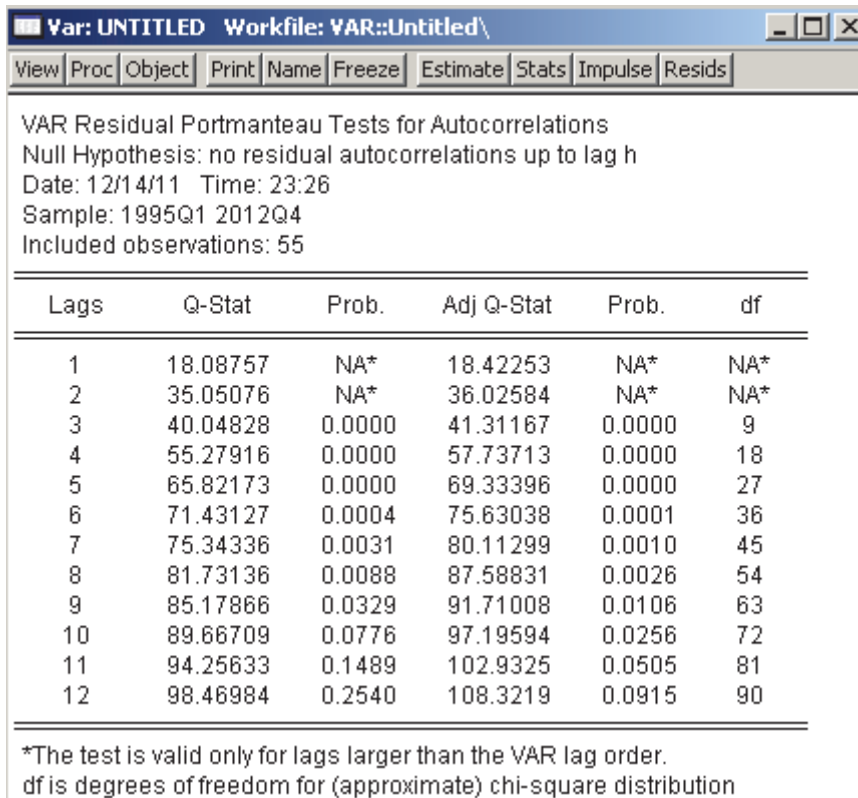


Figure 5 Portmanteau Tests in EViews 6.0

Portmanteau Autocorrelation Test (Fig.5.)

Computes the multivariate Box-Pierce/Ljung-Box Q -statistics for residual serial correlation up to the specified order (see Lütkepohl, 1991, 4.4.21 & 4.4.23 for details). We report both the Q -statistics and the adjusted Q -statistics (with a small sample correction). Under the null hypothesis of no serial correlation up to lag h , both statistics are *approximately* distributed χ^2 with degrees of freedom $k^2(h - p)$ where p is the VAR lag order. The asymptotic distribution is approximate in the sense that it requires the MA coefficients to be zero for lags $i > h - p$. Therefore, this approximation will be poor if the roots of the AR polynomial are close to one and h is small. In fact, the degrees of freedom becomes negative for $h < p$.

2.3. Out-Of-Sample Forecasting

The final evaluation test is “Out-Of-Sample Forecasting”. At this stage forecasts are compared to real data that we have for the period of the last 3 quarters of 2009.

3 The search algorithm

The search algorithm is shown in Figure 7. Step-by-step description looks as follows:
Input data: GDP Time series

1. Construction of ARMA models in levels and in differences and VAR models separately. Starting from this point the model is divided in two branches and the subsequent steps are carried out in parallel for levels and for differences.
2. Model ARMA Analysis.
3. Performing Residual tests.
4. If during residual tests probability value is less than 10%, the model is excluded from further evaluation. This is not a reliable model.
5. If P Value is higher than 10%, the forecast for specified periods of time is performed.
6. Out of Sample forecasting test. Comparing forecast data with real data.
7. The analysis of the results. Output data: Best model for the GDP Forecasts.

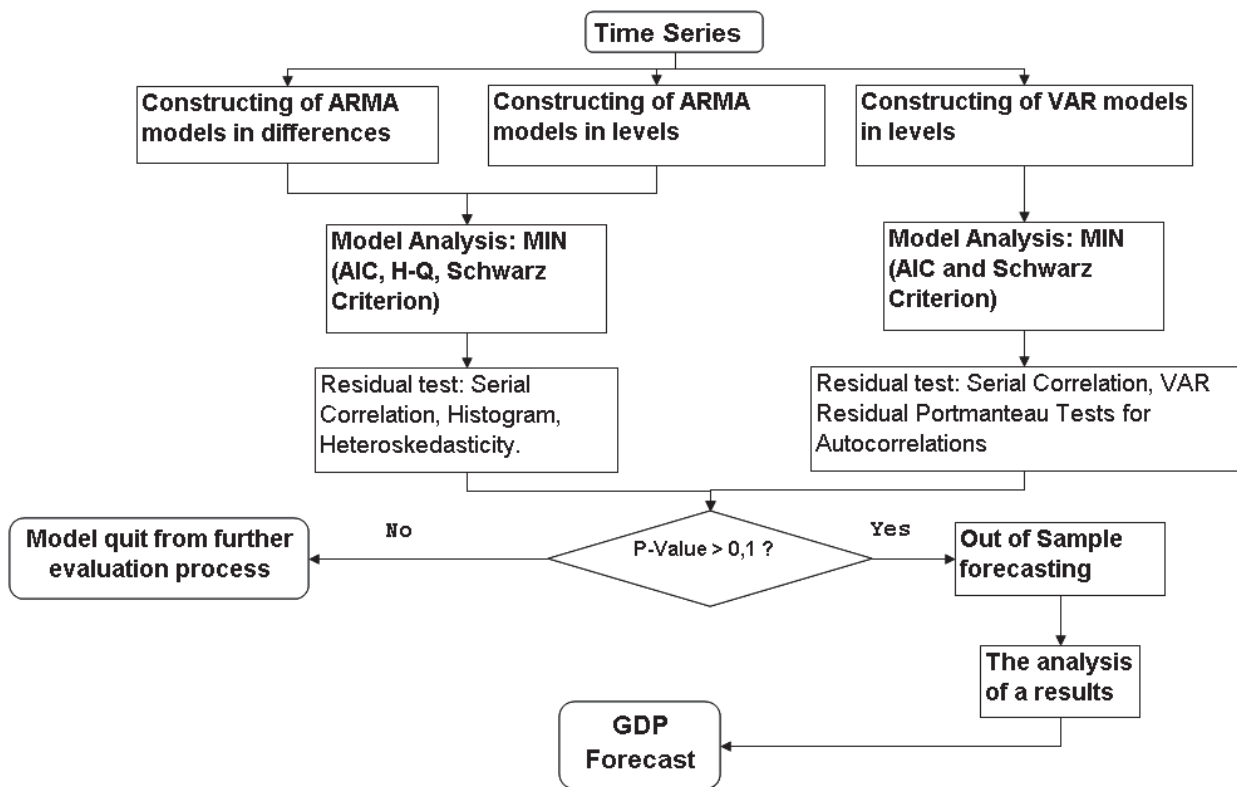


Figure 6 The search algorithm

4 **Latvian GDP in levels**

The analyzed series consist of seasonally unadjusted Latvian quarterly GDP in levels for the period of 1995Q1 – 2009Q1. It is represented in Figure 1. Constructing a model for the logs is more advantageous because the changes in the log series display a more stable variance than the changes in the original series. Time series in logs are shown in Figure 7.

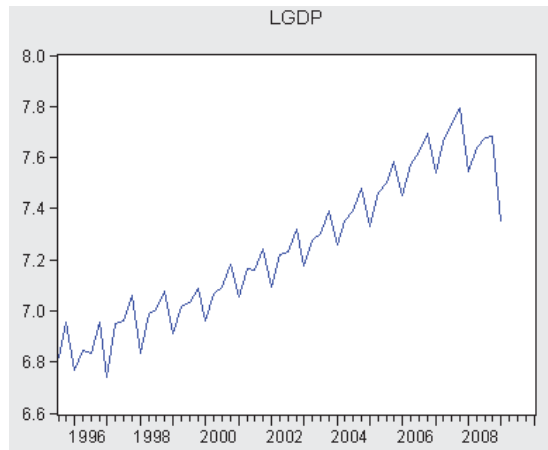


Figure 7 Latvian GDP in logs

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 are excluded from further evaluation process. The residual test results are given in Table 2. Models Nr.13 and Nr.14 have undergone all 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

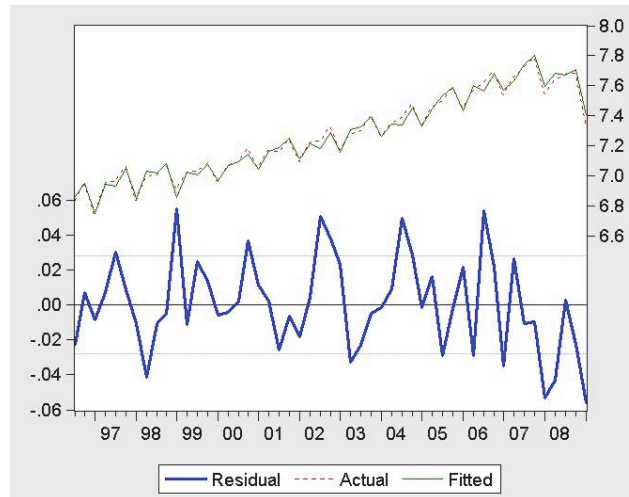


Figure 8 Residuals graph for Model Nr.13

The worst model residuals are given in Figure 5 for comparison.

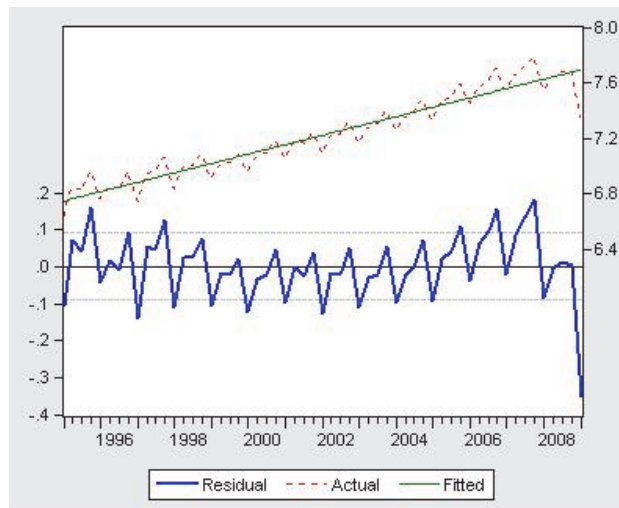


Figure 9 Residuals graph for Model Nr.9

“Out-Of-Sample Forecasting” test (Table 3) shows that the most real forecasts are gained 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 the third results are shown by models Nr. 7 and Nr.13., which have passed almost all residual tests except Nr.7., which also did not pass the Normality test.

Table 3

Out-Of-Sample Forecasting (Levels)

Nr.	Forecast - 1 Step - 09q2			Forecast - 2 Step - 09q3			Forecast - 3 Step - 09q4			Sum %
	LGDP real	LGDPf	Diff	LGDP real	LGDPf	Diff	LGDP real	LGDPf	Diff	
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

5 Latvian GDP in percentage growth

All evaluations are made with Latvian GDP. The differences and the log difference of time series are shown in Figure 6. The log difference displays a more stable variance than the changes in the original series.

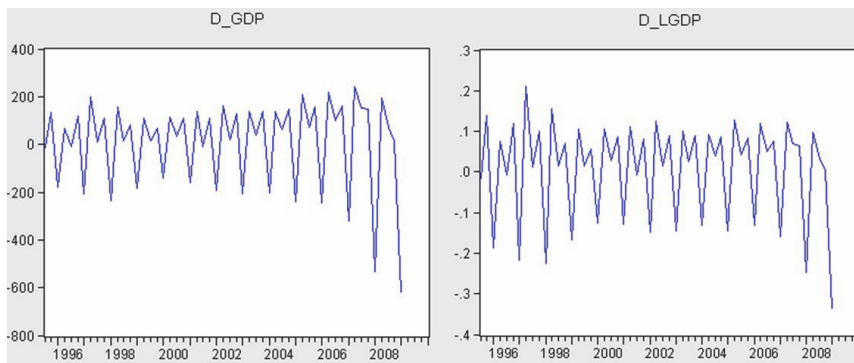


Figure 1 Residuals of time series in differences and in log differences

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 the best criteria: Nr. 7, 8, 11, 12, 13, 14. Other models are excluded 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 have completed all tests. Model Nr. 12 also has good statistic. “Out-Of-Sample” forecasting test (Table 6) shows that the most real forecasts are gained 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 passed all other residual tests. The second result belongs to model Nr.7, which has the same problem with Normality test. Third result is shown by model Nr.14 – this model passed all residual tests.

Table 6

Out-Of-Sample Forecasting (Difference)

Nr.	Forecast - 1 Step -09q2			Forecast - 2 Step -09q3			Forecast - 3 Step -09q4			Sum %
	dLGDP real	dLGDPf	Diff	dLGDP real	dLGDPf	Diff	LGDP real	LGDPf	Diff	
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

6 Vector auto regression

The analyzed series consist of seasonally unadjusted Latvian GDP, Inflation and M2X cash mass data indicators in for the period of 1995Q1 – 2009Q1 with quarterly observation. It is represented in Figure 2. M2X and GDP data are constructed in logs. Constructing a model for the logs is more advantageous because the changes in the log series display a more stable variance than the changes in the original series. Inflation time series is taken as is, because in some points it has negative values. Time series are shown in Figure 2.

Table 7

The analysis of criteria and residual tests for VAR models

Nr.	GDP	Akaike	Schwarz	Serial Correlation	VAR Residual Portmanteau Tests for Autocorrelations
1	C	-3.19764	-2.94216	0.42070	0.09150
2	C Trend	-2.18555	-1.89358	0.29130	0.07320
3	C SEAS(1)	-3.48239	-3.19041	0.99600	0.09250
4	C SEAS(1) SEAS(2) SEAS(3)	-3.67567	-3.3107	0.94820	0.14090
5	C Trend SEAS(1) SEAS(2) SEAS(3)	-3.6687	-3.26723	0.83680	0.09700
6	C Trend SEAS(1) SEAS(2)	-3.54695	-3.18198	0.92760	0.07230
7	C SEAS(1) SEAS(2)	-3.4629	-3.13442	0.9883	0.1457
8	C SEAS(2) SEAS(4)	-3.26908	-2.94061	0.893	0.1733
9	Trend	-1.97041	-1.71493	0.00640	0.00000
10	Trend SEAS(1)	-3.56915	-3.27717	0.60660	0.00790
11	Trend SEAS(1)	-3.56854	-3.24007	0.9648	0.0524

Model analysis and the residual test results are given in Table 7. After carrying out of the analysis of model it is appreciable that best models are Nr. 4,5,10. In the residual tests case best model are Nr. 4,7,8. In Figure 7 we can see residual analysis for model Nr.1 that are performed in EViews software.

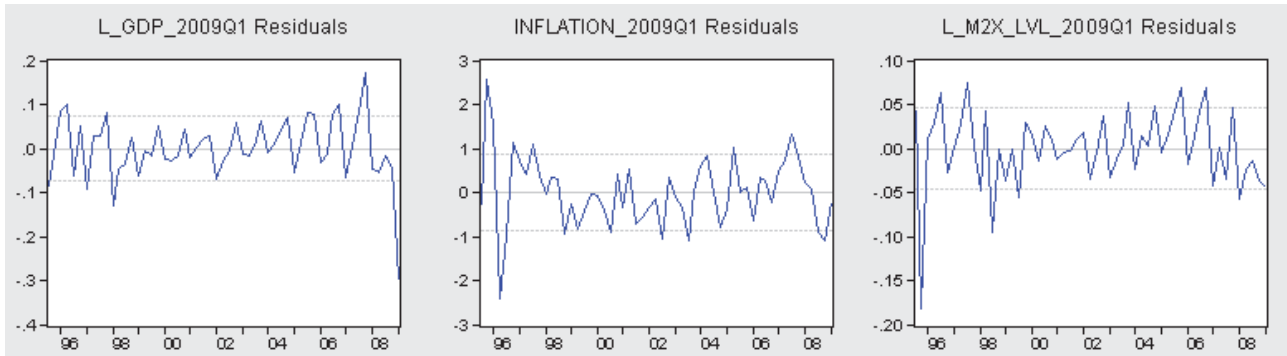


Figure 7 Residuals of time series in VAR model (Nr.1.)

Table 8

Out-Of-Sample Forecasting (VAR)

Nr.	Forecast - 1 Step - 2009q2			Forecast - 2 Step - 2009q3			Forecast - 3 Step - 2009q4			Sum %
	LGDP real	LGDPf	Diff	dLGDP re	dLGDPf	Diff	LGDP rea	LGDPf	Diff	
1	0.08837	0.13140	-0.04303	0.03073	0.07745	-0.04672	0.03364	0.04118	-0.00754	0.09729
2	0.08837	0.36690	-0.27853	0.03073	-0.00229	0.03302	0.03364	0.00154	0.03211	0.34366
3	0.08837	0.14402	-0.05566	0.03073	-0.00366	0.03439	0.03364	0.04070	-0.00706	0.09710
4	0.08837	0.18926	-0.10089	0.03073	0.00038	0.03035	0.03364	0.07325	-0.03961	0.17084
5	0.08837	0.16749	-0.07912	0.03073	-0.01937	0.05010	0.03364	0.07205	-0.03840	0.16763
6	0.08837	0.13501	-0.04664	0.03073	-0.06039	0.09111	0.03364	0.03107	0.00257	0.14032
7	0.08837	0.17461	-0.08624	0.03073	-0.02411	0.05484	0.03364	0.06105	-0.02741	0.16849
8	0.08837	0.29866	-0.21030	0.03073	0.03009	0.00064	0.03364	0.01722	0.01642	0.22735
9	0.08837	0.18843	-0.10006	0.03073	-0.13286	0.16358	0.03364	0.10985	-0.07620	0.33985
10	0.08837	0.14931	-0.06094	0.03073	0.00365	0.02707	0.03364	0.04039	-0.00675	0.09476
11	0.08837	0.10749	-0.01912	0.03073	-0.04182	0.07255	0.03364	0.01343	0.02022	0.11188

“Out-Of-Sample Forecasting” test (Table 8) shows that the most real forecasts are gained from models Nr. 10. The absolute difference is equal to 9,476 percents. The second and the third results are shown by models Nr. 3 and Nr.1. The model Nr.1 has high results in the analysis of models (Akaike and Schwarz Criterion) and in Serial Correlation but did not pass the Portmanteau test.

Conclusion

In the given paper the search algorithm of optimal time series is described. With a help of statistical modelling the econometric analysis of Latvian GDP is done. Different cases of constructing model are made in levels, in percentage growth and constructing a Vector Auto Regression model.

Comparison of 1 and 2 cases shows that case in levels and in differences gave approximately the same result – 3.7% deviation from real data in absolute value for forecasts for 3 steps in future.

Comparison of ARMA and VAR models shows that ARMA models win the competition. If we analyse tables of results of models in all three cases we will see that VAR models have the best dispersion of results.

The conclusion is that very important to analyze all the results and understand how they are calculated and evaluated. Data received as a result of this work will be used for the further researches in a scope of Vector Auto Regress for forecasting of macroeconomic variables.

References

- [1] BEZRUCKO A. Latvian GDP: The Optimal Time Series Forecasting Algorithm. – (APLIMAT Proceeding and Journal (2011), pp 477-486)
- [2] LUTKEPOHL H., KRATZIG M. Applied Time Series Econometrics. – Cambridge: Cambridge University press, 2004, 323 pp.
- [3] EViews software documentation: Quantitative Micro Software– <http://www.eviews.com> – Resource is described 2010, 30 of august.
- [4] NOSKO V., Econometrics: introduction to regression analysis of time series. – Moscow: Moscow Institute of Physics and Technology , 2002, 273 pp.
- [5] MOLCANOV I., ARZENOVSKI S. Statistical methods of forecasting. – Rostov-na-Donu: Rostov State University, 2001, 74 pp.

Current address

Bezrucko Aleksandrs, Mgr.

Chair of the theory of probability and the mathematical statistics
Faculty of Computer Science and Information Technology
Riga Technical University, 1/4 Meza Street, LV-1048 Riga, Latvia
e-mail: bezrucko@gmail.com