Theoretical and Applied Economics
Volume XVIII (2011), No. 10(563), pp. 91-98

Econometric Modeling of GDP Time Series

Elena-Adriana ANDREI
Bucharest Academy of Economic Studies
adrianna_andrei@yahoo.com
Elena BUGUDUI
"Artifex" University of Bucharest
bugudui@yahoo.com

Abstract. *Article aims of time series econometric model of macroeconomic variable GDP in the US economy. Because that is a nonstationary time series, there are used several statistical tests in order to turn into a stationary series. After applying these tests, the time series became stationary and integrated of order I; thus, we use Box-Jenkins procedure for the determination of ARMA. We estimate by OLS the parameters of various models. Performances chosen ARIMA model (1,1,1) are verified on the basis of classical statistical tests and forecasting.*

Keywords: stationary time series; nonstationary time series; statistical tests.

JEL Codes: C22, E01.
REL Codes: 8C, 10G.

In order to set time series, we use macroeconomic variable GDP during 1947 (first quarter) – 2010 (third quarter) in the US economy. Times series values are observed at the same frequency, namely quarterly; each of these values are random variables. Therefore, we can say that GDP is a stochastic process and the actual values observed in period mentioned above are individual achievements of this process.

Descriptive analysis time series gives us informations about the following indicators:

i) time series mean:

$$\bar{y} = \frac{\sum_{t=1}^{T} y_t}{T} = 6{,}372.390$$

ii) time series variance:

$$s^2 = \frac{\sum_{t=1}^{T} \left(y_t - \bar{y} \right)^2}{T-1} = 12{,}735{,}711$$

iii) Jarque-Berra statistic value, which suggest a normal distribution of time series in terms of asymetry and flattening.

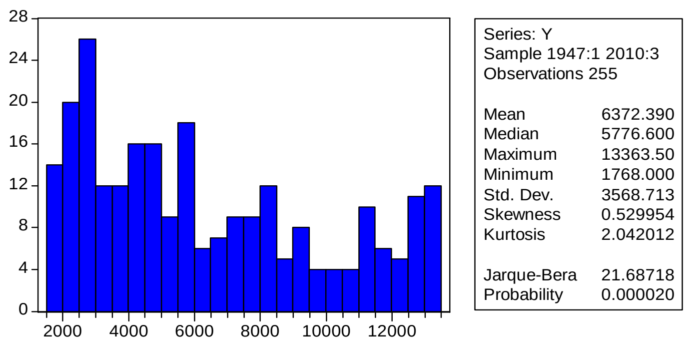


Figure 1. *Histogram and descriptive indicators*

View time series graph shows a linear path with positive slope; therefore it is *nonstationary homogenous type*, characterized by constant changes from one period to another.

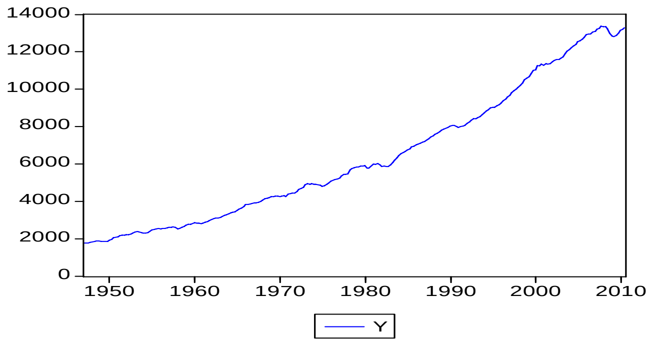


Figure 2. *GDP, US, 1947-I to 2010-III*

In order to check time series stationarity or nonstationarity we performed statistical tests, such as:

- *Unit root Test -Dickey – Fuller Test (DF):*

$\Delta Y_t = \alpha_0 + \gamma \times Y_t + \alpha_1 \times \Delta Y_{t-1} + e_t$

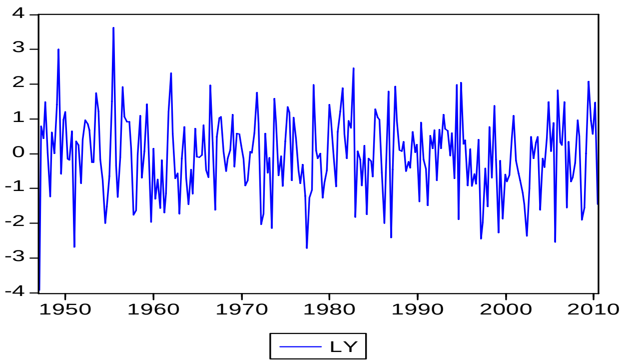


Figure 3. $\ln Y_t$ evolution

We obtain the following results:

$\Delta Y_t = 0.020348 + (-0.01768) Y_{t-1} + 0.353951 \Delta Y_{t-1}$

t = (2.407469) (-1.817024) (6.020099)

R^2 =0.146846 d = 2.062047

Our primary interest is in the t ($=\tau$) value of the Y_{t-1} coefficient. Since $\hat{\tau}$ = -1.82, and critical τ value is 2.87, for the 5 percent confidence threshold (τ >0.05) we can not reject RW null hypothesis.

Therefore, GDP time series is *nonstationary*, i.e. it contain a unit root; this conclusion is supported also by the ρ-value (0.3717) associated with t-statistic, which is higher than 0.005.

- *Autocorrelation function (ACF) and correlogram*

Table 1a

GDP Correlogram, US, 1947-I to 2010-III

Sample: 1947:1 2010:3
Included observations: 255

	AC	PAC	Q-Stat	Prob
1	0.989	0.989	252.25	0.000
2	0.977	-0.016	499.65	0.000
3	0.966	-0.015	742.15	0.000
4	0.954	-0.001	979.82	0.000
5	0.943	0.001	1212.8	0.000
6	0.931	0.000	1441.1	0.000
7	0.920	-0.011	1664.8	0.000
8	0.908	-0.018	1883.7	0.000
9	0.896	-0.028	2097.7	0.000
10	0.884	-0.021	2306.6	0.000
11	0.871	-0.008	2510.5	0.000
12	0.858	-0.021	2709.2	0.000
13	0.846	0.011	2903.0	0.000
14	0.834	0.005	3092.1	0.000
15	0.822	0.010	3276.6	0.000
16	0.810	-0.003	3456.7	0.000
17	0.799	-0.004	3632.3	0.000
18	0.787	-0.004	3803.6	0.000
19	0.776	-0.002	3970.6	0.000
20	0.764	-0.008	4133.4	0.000
21	0.752	-0.008	4291.9	0.000
22	0.741	-0.010	4446.3	0.000
23	0.729	-0.009	4596.4	0.000
24	0.718	0.008	4742.6	0.000
25	0.707	0.000	4884.9	0.000
26	0.695	-0.007	5023.3	0.000
27	0.684	-0.014	5157.8	0.000
28	0.673	-0.019	5288.4	0.000
29	0.661	-0.012	5415.0	0.000
30	0.649	-0.006	5537.7	0.000

The autocorrelation coefficients start at a very high value - $\hat{\rho}_1 = 0.989$, $\hat{\rho}_2 = 0.977$, …, $\hat{\rho}_5 = 0.943$ – and decline very slowly; thus, GDP time series is *nonstationary*.

Since *statistical tests* concluded that time series is nonstationary, the solution is the first differences of GDP time series.

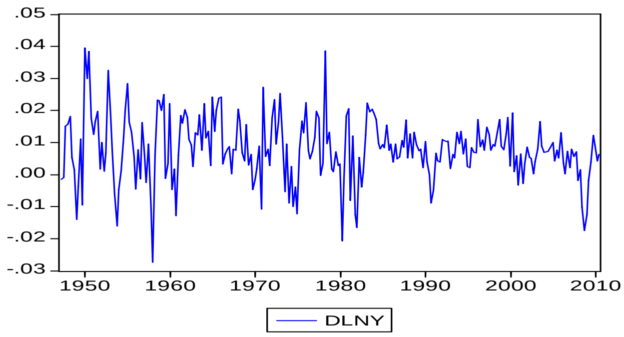


Figure 5. *GDP first differences, US, 1947-I to 2010-III*

$$\Delta Y_t = 0.005049 Y_t - 0.632433 Y_{t-1}$$
$$t = \quad (6.783151) \qquad (-10.79566)$$
$$R^2 = 0.317093 \qquad d = 2.068089$$

The 1 percent critical DF τ value is - 3.456093; since τ < 0.01, we conclude that we can *reject RW null hypothesis* (H_0: $\gamma = 0$). Thus, GDP first differenced is stationary, i.e. time series is I(1).

From first differences correlogram we observe that values of autocorrelation coefficients – $\hat{\rho}_1 = 0.368, \hat{\rho}_2 = 0.211, \quad \hat{\rho}_3 = 0.000, \quad \hat{\rho}_4 = -0.008, \quad \hat{\rho}_5 = -0.142$, decline very quickly, so after the first lag it is likley to have an autoregressive process (AR) component of order p = 1.

Table 1b

First differences time series correlogram

Sample: 1947:1 2010:3
Included observations: 254

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.368	0.368	34.717	0.000
		2	0.211	0.088	46.185	0.000
		3	0.000	-0.120	46.185	0.000
		4	-0.088	-0.083	48.180	0.000
		5	-0.142	-0.075	53.440	0.000
		6	-0.061	0.044	54.431	0.000
		7	-0.049	-0.016	55.051	0.000
		8	-0.022	-0.023	55.184	0.000
		9	0.070	0.087	56.494	0.000
		10	0.071	0.019	57.851	0.000
		11	0.026	-0.041	58.031	0.000
		12	-0.131	-0.175	62.626	0.000
		13	-0.120	-0.018	66.532	0.000
		14	-0.087	0.046	68.571	0.000
		15	-0.085	-0.054	70.515	0.000
		16	0.042	0.079	71.007	0.000
		17	0.053	0.000	71.782	0.000
		18	0.087	0.035	73.879	0.000
		19	0.055	-0.016	74.720	0.000
		20	0.061	0.006	75.738	0.000
		21	-0.085	-0.098	77.744	0.000
		22	-0.063	0.019	78.855	0.000
		23	-0.096	-0.022	81.422	0.000
		24	-0.032	0.014	81.714	0.000
		25	0.032	0.042	81.996	0.000
		26	0.021	-0.047	82.126	0.000

Box-Jenkins procedure takes into account the value of d parameter, equal to the initial time series differences order applied to obtaine a stationary time series.

Since first differences time series in stationary, $\ln Y_t$ time series is $I(1)$, i.e. $d = 1$.

To calculate p and q we estimate model parameters by OLS; estimation results are presented in the table bellow.

Table 2a

US GDP time series estimated models results

	Eq1	Eq2	Eq3	Eq4
c	0.007984	0.008025	0.007999	0.008012
ar(1)	0.367567	0.332860	0.485481	-0.218832
ar(2)		0.088608		0.302587
ma(1)			-0.134935	0.557360
AIC	-6.520910	-6.518205	-6.51786	-6.520263
SCI	-6.492978	-6.476188	-6.47596	-6.464240
DW	2.068089	1.967281	2.020407	1.981934

We choose between competing models Eq3 model, ARIMA (1,1,1), which admits the following representation:

$$Y_t = 0.007999 + 0.485481 Y_{t-1} - 0.134935 e_t$$

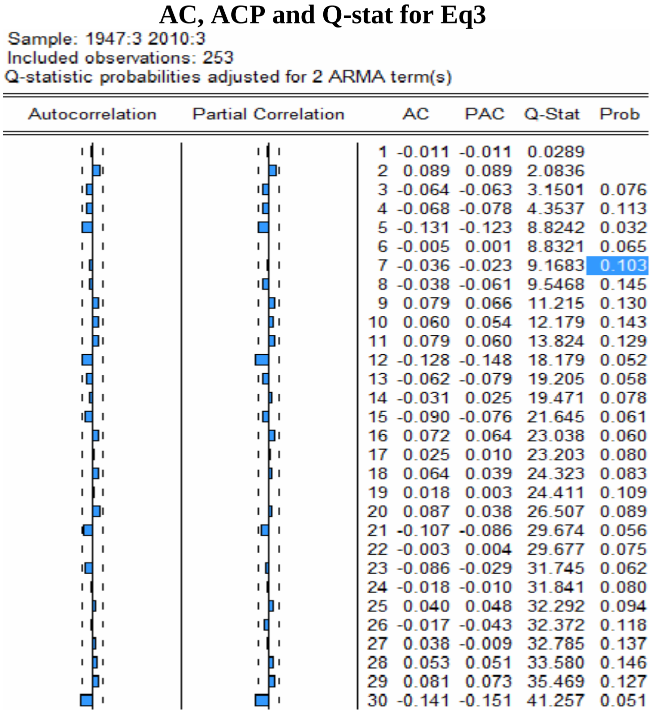
ARIMA model selection was done taking into account: a) the classical criteria (*adjusted R-square, F test*); b) indicators based on information theory (*Akaike and Schwartz statistics*).

For diagnostic checking we analysed:

- *residual correlogram*: autoregressive model defined by Eq3 suggest that residuals series do not present autocorrelation, i.e. is no need to look for another ARIMA model.

Table 2b

AC, ACP and Q-stat for Eq3



Sample: 1947:3 2010:3
Included observations: 253
Q-statistic probabilities adjusted for 2 ARMA term(s)

	AC	PAC	Q-Stat	Prob
1	-0.011	-0.011	0.0289	
2	0.089	0.089	2.0836	
3	-0.064	-0.063	3.1501	0.076
4	-0.068	-0.078	4.3537	0.113
5	-0.131	-0.123	8.8242	0.032
6	-0.005	0.001	8.8321	0.065
7	-0.036	-0.023	9.1683	0.103
8	-0.038	-0.061	9.5468	0.145
9	0.079	0.066	11.215	0.130
10	0.060	0.054	12.179	0.143
11	0.079	0.060	13.824	0.129
12	-0.128	-0.148	18.179	0.052
13	-0.062	-0.079	19.205	0.058
14	-0.031	0.025	19.471	0.078
15	-0.090	-0.076	21.645	0.061
16	0.072	0.064	23.038	0.060
17	0.025	0.010	23.203	0.080
18	0.064	0.039	24.323	0.083
19	0.018	0.003	24.411	0.109
20	0.087	0.038	26.507	0.089
21	-0.107	-0.086	29.674	0.056
22	-0.003	0.004	29.677	0.075
23	-0.086	-0.029	31.745	0.062
24	-0.018	-0.010	31.841	0.080
25	0.040	0.048	32.292	0.094
26	-0.017	-0.043	32.372	0.118
27	0.038	-0.009	32.785	0.137
28	0.053	0.051	33.580	0.146
29	0.081	0.073	35.469	0.127
30	-0.141	-0.151	41.257	0.051

Table 2c

ARIMA model residuals correlogram

Sample: 1947:3 2010:3
Included observations: 253
Q-statistic probabilities adjusted for 2 ARMA term(s)

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.125	0.125	4.0284	
		2	0.170	0.156	11.424	
		3	0.022	-0.017	11.546	0.001
		4	0.154	0.132	17.667	0.000
		5	0.018	-0.014	17.750	0.000
		6	0.039	-0.004	18.149	0.001
		7	-0.007	-0.010	18.162	0.003
		8	0.101	0.084	20.862	0.002
		9	0.087	0.074	22.844	0.002
		10	-0.024	-0.077	22.995	0.003
		11	0.139	0.143	28.177	0.001
		12	0.122	0.091	32.179	0.000
		13	0.046	-0.043	32.741	0.001
		14	0.060	0.051	33.706	0.001
		15	0.123	0.092	37.827	0.000
		16	-0.022	-0.098	37.954	0.001
		17	0.049	0.018	38.621	0.001
		18	-0.029	-0.013	38.856	0.001
		19	0.047	0.007	39.462	0.002
		20	0.066	0.046	40.650	0.002
		21	0.071	0.054	42.033	0.002
		22	-0.035	-0.066	42.381	0.002
		23	0.013	-0.059	42.431	0.004
		24	0.039	0.055	42.867	0.005
		25	0.024	0.008	43.024	0.007
		26	-0.032	-0.093	43.309	0.009
		27	0.035	0.069	43.656	0.012

- *residuals series homoschedasticity*: the phenomenon is not present at residuals series correlogram level;
- *residuals series distribution*: Jarque - Berra statistic value suggests a normal distribution in terms of asymmetry and flattening.

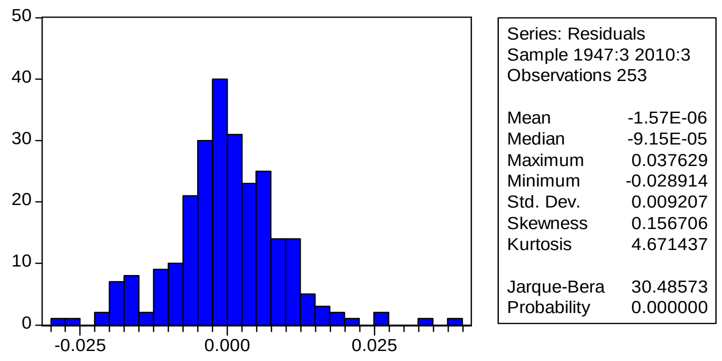


Figure 9. *Jarque-Berra test for Eq3*

Analysed Y_t time series, by initially nonstationary, was transformed into a stationary after applying two transformation: logarithms and first order difference. Using Box-Jenkins methodology we obtained an ARIMA (1,1,1) model. Model validity was justified on the basis of adjusted R-square, F test, Akaike and Schwartz indicators, statistical tests and forecasting.

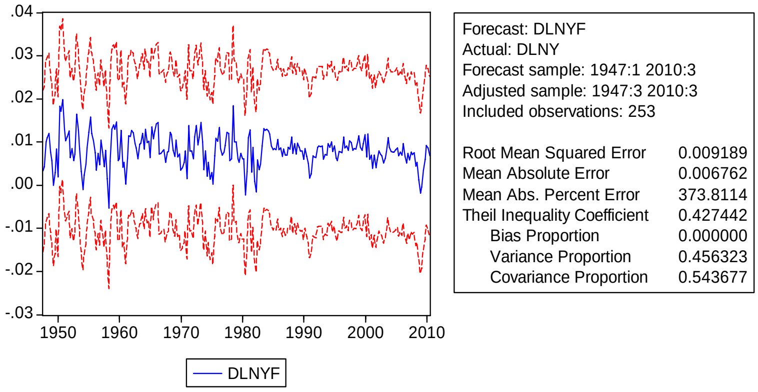


Figure 10. *Forecast Eq3*

Taking into account the loss function value (0.009189), time series prediction is considered accurate for 5 percent significance level and ARIMA (1,1,1) model.

Since R^2 value is very low, 13.9%, we can not successfully use the previous values of the series in order to make predictions with a high degree of accuracy.

References

Andrei, T., Bourbonnais, R. (2008). *Econometrie*, Editura Economică, Bucureşti

Asteriou D., Hall, S. (2011). *Apllied Econometrics*, Palgrave Macmillan

Gujarati (2004). *Basic Econometrics*, 4th Edition, The McGraw-Hill Companies

*** www.bea.gov, US Departement of Commerce, Bureau of Economic Analysis, National Economic Accounts