

Praktični aspekti modeliranja
stacionarnih vremenskih serija
(Boks-Dženkinsova strategija modeliranja)

Zorica Mladenović

1

Autori metodologije

- Box and Jenkins, *Time Series Analysis: Forecasting and Control*, 1976, II izdanje
- Britanski statističari:
 - G.E.P. Box (1919-2013), započeo studije hemije
 - G.M. Jenkins (1933-1982), dipl. matematičar
- Autori knjige V izdanja iz 2015:
Box, Jenkins, Reinsel and Ljung

2

Osnove

- **Cilj:** izbor ARMA modela koji na zadovoljavajući način opisuje kretanje konkretne vremenske serije.

- **Polazna osnova:** ARMA(p,q) model

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \dots + \phi_p X_{t-p} + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \dots - \theta_q e_{t-q}$$

- **Uobičajeni naziv:** izgradnja ARMA modela (ARMA modeliranje)
- *Boks: Svi modeli su pogrešni, samo su neki korisni.*

3

Osnove II

- Pristup se sastoji od tri faze:
 1. identifikacija modela
 2. ocena parametara modela i
 3. provera adekvatnosti modela.
- Iterativna procedura

5

I Faza identifikacije modela

- Potrebno je izabrati užu klasu ARMA modela za koju pretpostavljamo da predstavlja potencijalni generator skupa podataka.
- Ključno pitanje:
 Koliki je red autoregresione i komponente pokretnih proseka?
- Ključni princip:
 Jednostavnost (ekonomičnost)
- Ključni metodološki okvir:
 Analiza obične i parcijalne autokorelacione funkcije

6

Model	Obična autokorelaciona funkcija	Parcijalna autokorelaciona funkcija
AR(p)	Vrednosti opadaju po eksponencijalnoj, ekspon. oscilatornoj ili sinusoidnoj putanji.	$\phi_{11} \neq 0, \phi_{22} \neq 0, \dots, \phi_{pp} = \phi_p$ $\phi_{kk} = 0$ za $k > p$.
MA(q)	$\rho_1 \neq 0, \rho_2 \neq 0, \dots, \rho_q \neq 0,$ $\rho_k = 0$ za $k > q$.	Vrednosti opadaju po eksponencijalnoj, eksponen. oscilatornoj ili sinusoidnoj putanji.
ARMA(p,q)	Vrednosti opadaju. Prvih q koeficijenata zavisi od parametara AR i MA dela, a za docnje veće od q koeficijenti ponašaju kao kod AR modela.	Vrednosti opadaju. Prvih p koeficijenata zavisi od AR i MA komponente. Za docnje veće od p koeficijenti slede putanju sličnu kao kod MA modela.

7

II Faza ocene parametara modela

- Metod običnih najmanjih kvadrata se može koristiti u oceni parametara AR modela.
- Za ocenu parametara MA i ARMA modela koristi se metod nelinearnih najmanjih kvadrata koji se zasniva na upotrebi metoda numeričke optimizacije.

8

III Faza provere adekvatnosti modela

1. Da li je model saglasan sa podacima?
Da li su reziduali normalno raspodeljeni i neautokorelisani?
2. Da li je izbor AR i MA komponente optimalan?
Da li je model istovremeno ekonomičan i dovoljno precizan?

9

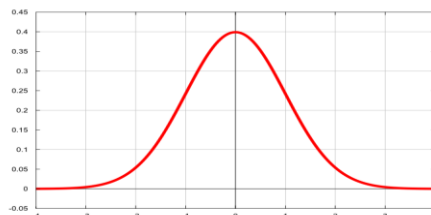
III-1. Analiza reziduala

- Normalnost
- Autokorelacija

10

Testiranje normalnosti reziduala u ocenjenom modelu

- Uobičajena pretpostavka: slučajna greška ima normalnu raspodelu, $e: N(0, \sigma^2)$
- Zbirno dejstvo velikog broja sporadičnih i nesistematičnih faktora opravdano je modelirati normalnom raspodelom. Otuda i pretpostavka.



11

Šta ako slučajna greška nema normalnu raspodelu?

- Ukoliko je samo ova pretpostavka narušena, tada se dobijaju pouzdane ocene.
- Međutim,
 - Testiranje hipoteza je nepouzđano.
 - Vrednosti t-odnosa i F-odnosa su netačne.
 - Verovatno postoji greška u specifikaciji modela.
- Zaključak:
 - Postupak statističkog zaključivanja je pogrešan.
 - Pretpostavka je vitalni deo specifikacije modela.

12

Provera validnosti pretpostavke da slučajna greška ima normalnu raspodelu

- *Neformalni* (grafički) pristup
- *Formalni* pristup

13

Neformalni (grafički) pristup

- ***Histogram***
- Grafički prikaz učestalosti pojavljivanja podataka (reziduala) u pojedinim grupnim intervalima
 - x-osa: celokupni raspon vrednosti date promenljive deli se na određeni broj podintervala jednake širine
 - y-osa: broj pojavljivanja podataka u svakom od podintervala

14

Formalni pristup

- Primena različitih test-statistika
- Najpopularniji test normalnosti je **Žark-Bera** (engl. Jarque-Bera) test
 - Oznaka: JB
 - Zasniva se na ocenama koeficijenata kojima se opisuju svojstva raspodele

15

Koeficijenti za deskripciju raspodele

- Empirijska raspodela često se opisuje na osnovu dva koeficijenta: *asimetrije* i *sploštenosti*.
- Koeficijent asimetrije meri stepen u kojem raspodela nije simetrična u odnosu na srednju vrednost.
- Raspodela može biti
 - Simetrična
 - Asimetrična
 - Ulevo (negativna)
 - Udesno (pozitivna)

16

Koeficijent asimetrije (engl. *skewness*)

Oznaka koeficijenta: α_3

Vrednost koeficijenta α_3	Tip raspodele
Nula	Simetrična
Veća od nule	Asimetrična udesno
Manja od nule	Asimetrična ulevo

18

Koeficijenti za deskripciju raspodele II

- Koeficijent spljoštenosti meri težinu (debljinu) repova raspodele
 - Svojstva repova raspodele iskazuju se u odnosu na normalnu raspodelu
- Repovi raspodele mogu biti
 - *Iste težine* kao kod normalne raspodele
 - *Teži* od repova normalne raspodele
 - *Lakši* od repova normalne raspodele

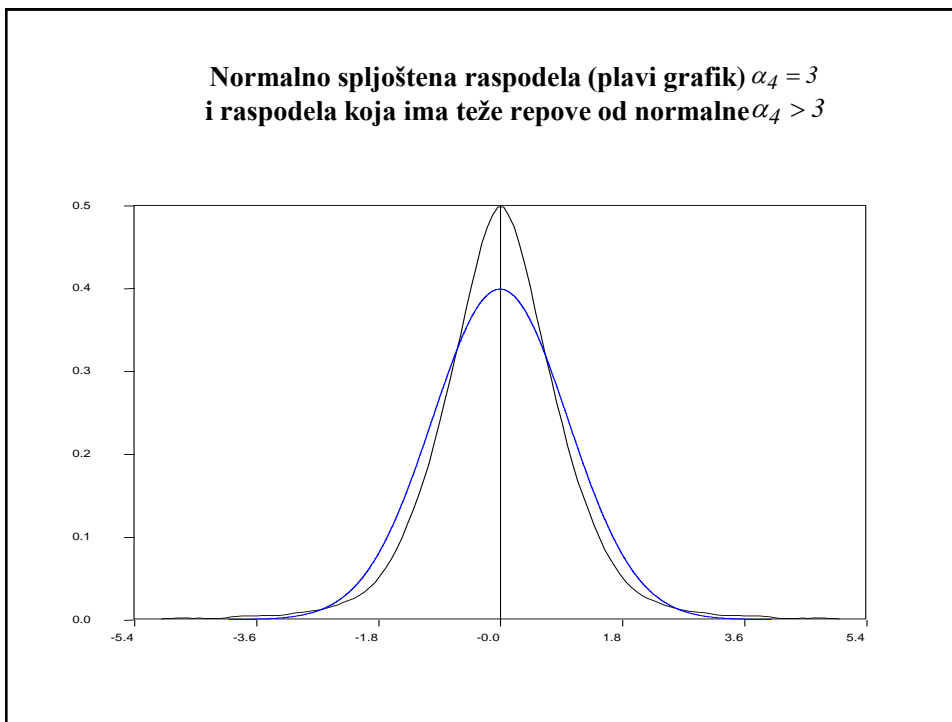
19

Koeficijent spljoštenosti (engl. *kurtosis*)

Oznaka koeficijenta: α_4

Vrednost koeficijenta α_4	Težina repova
Tri	Odgovara normalnoj raspodeli
Veća od tri	Veći deo jedinične verovatnoće je pod repovima nego kod repova N raspodele (prisustvo ekstremnih opservacija)
Manja od tri	Manji deo jedinične verovatnoće je pod repovima nego kod repova N raspodele

20



21

Koeficijent asimetrije	Koeficijent spljoštenosti
$\alpha_3 = 0$ za N raspodelu	$\alpha_4 = 3$ za N raspodelu
$\hat{\alpha}_3 : N\left(0, \frac{6}{T}\right)$	$\hat{\alpha}_4 : N\left(3, \frac{24}{T}\right)$
$\sqrt{\frac{T}{6}}\hat{\alpha}_3 : N(0, 1)$	$\sqrt{\frac{T}{24}}(\hat{\alpha}_4 - 3) : N(0, 1)$
<p><u>Nulta hipoteza</u>: data raspodela je normalna, odnosno vrednosti koeficijenata asimetrije i spljoštenosti su redom 0 i 3.</p> <p><u>Alternativna hipoteza</u>: data raspodela nije normalna.</p> <p>Statistika:</p> $JB = \frac{T}{6} \left[\hat{\alpha}_3^2 + \frac{(\hat{\alpha}_4 - 3)^2}{4} \right] : \chi_2^2$ <p>Odgovarajuća kritična vrednost na nivou značajnosti 5% je 5.99.</p>	

22

Šta raditi u slučaju da raspodela odstupa od normalne?

- Ne postoji jedinstveno rešenje zato što je odstupanje od normalnosti posledica pogrešne specifikacije modela.
- Često se modifikuje polazna specifikacija uključivanjem promenljivih kojima se eksplicitno modeliraju ekstremni događaji.
 - Takve promenljive se nazivaju veštačke promenljive.

23

Testiranje autokorelacije reziduala u ocenjenom modelu

- Da li postoji autokorelacija na određenoj, k-toj, docnji?
($H_0: \rho_k=0$)
- Da li postoji autokorelacija na svim docnjama do **K**-te?
($H_0: \rho_1 = \rho_2 = \dots = \rho_K = 0$).

24

Da li postoji autokorelacija na određenoj, k-toj, doznji? ($H_0: \rho_k=0$)

Validnost nulte hipoteze $H_0: \rho_k=0$ protiv alternativne $H_1: \rho_k \neq 0$ se testira tako što se proverava da li ocena autokorelacionog koeficijenta na doznji k serije reziduala, pripada intervalu $(-1.96/\sqrt{T}, 1.96/\sqrt{T})$.

25

Da li postoji autokorelacija na svim doznjama do **K**-te?

($H_0: \rho_1 = \rho_2 = \dots = \rho_K = 0$).

- Validnost nulte hipoteze $H_0: \rho_1 = \rho_2 = \dots = \rho_K = 0$ se testira protiv alternativne da je bar jedan od prvih **K** autokorelacionih koeficijenata serije reziduala različit od nule.
- Boks-Pirsova i Boks-Ljungova statistika.
- Boks-Ljungova statistika:

$$Q(\mathbf{K}) = BLj(\mathbf{K}) = T(T + 2) \sum_{i=1}^K \frac{\hat{\rho}_i^2}{T - i} : \chi_{\mathbf{K}-p-q}^2$$

26

Šta raditi u slučaju da autokorelacija postoji u rezidualima?

- To je znak da ARMA model nije dobro postavljen, odnosno da odabir objašnjavajućih promenljivih nije adekvatan.
- Neophodno je redefinisati izbor AR i MA komponenti u skladu sa evidentiranim tipom (redom) autokorelacije i potom oceniti novi model.

27

III-2. Optimalan izbor parametara modela (Informacioni kriterijum)

Informacioni kriterijum predstavlja zbir dva elementa

1. Element koji je funkcija neobjašnjenog varijabiliteta modela
2. Element kojim se sankcioniše gubitak u broju stepeni slobode zbog povećanja broja parametara za ocenjivanje

$$IC(p, q) = \ln s^2 + g \frac{p + q}{T}$$

g je nenegativna kaznena funkcija
 s^2 je ocena varijanse slučajne greške modela

28

Informacioni kriterijum

- Sabirci u informacionoj funkciji različito reaguju na povećanje p i q :
 - Ocena varijanse sl. greške modela se smanjuje
 - Kaznena komponenta se povećava
- **Cilj: izbor kombinacije p i q koja minimizira vrednost informacionog kriterijuma**

$$IC(p, q) = \ln s^2 + g \frac{p + q}{T}$$

29

Informacioni kriterijum II

Funkcija g	Kaznena komponenta	Naziv	Oznaka
2	$2(p+q)/T$	Akaikeov	AIC
$\ln T$	$(\ln T)(p+q)/T$	Švarcov	SC / SIC
$2\ln(\ln T)$	$2(\ln(\ln T))^* (p+q)/T$	Hana-Kvinov	HQC / HQIC

30

Dodatni kriterijumi u fazi
provere adekvatnosti modela

1. Namerno proširenje ARMA modela
2. Analiza preciznosti u prognoziranju
3. ...
4. ...

31

Primer: Modeliranje godišnje
stope rasta BDP u Srbiji

Podaci: vežba2.wf1

Kvartalni nominalni BDP za: 2005q1 – 2019q4 (www.nbs.rs)

Godišnja stopa rasta BDP se razmatra za: 2006q1 – 2019q4
($T=56$).

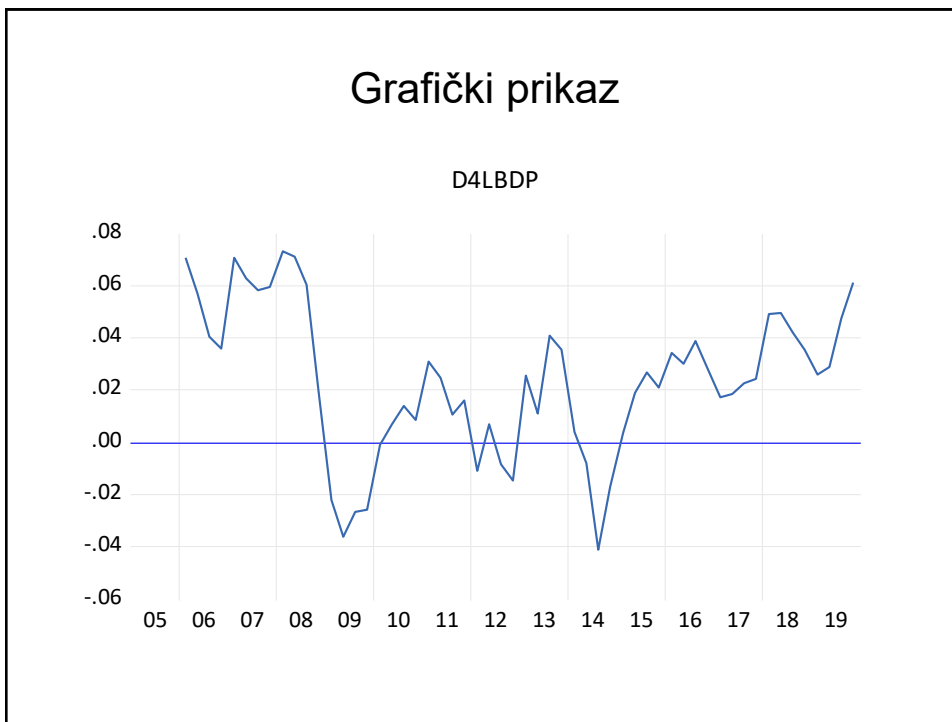
$d4\ln bdp = \ln bdp - \ln bdp(-4)$, $\ln bdp = \log(bdp)$.

Analiziramo ocene obične i parcijalne autokorelacione funkcije
(uzoračke korelograme).

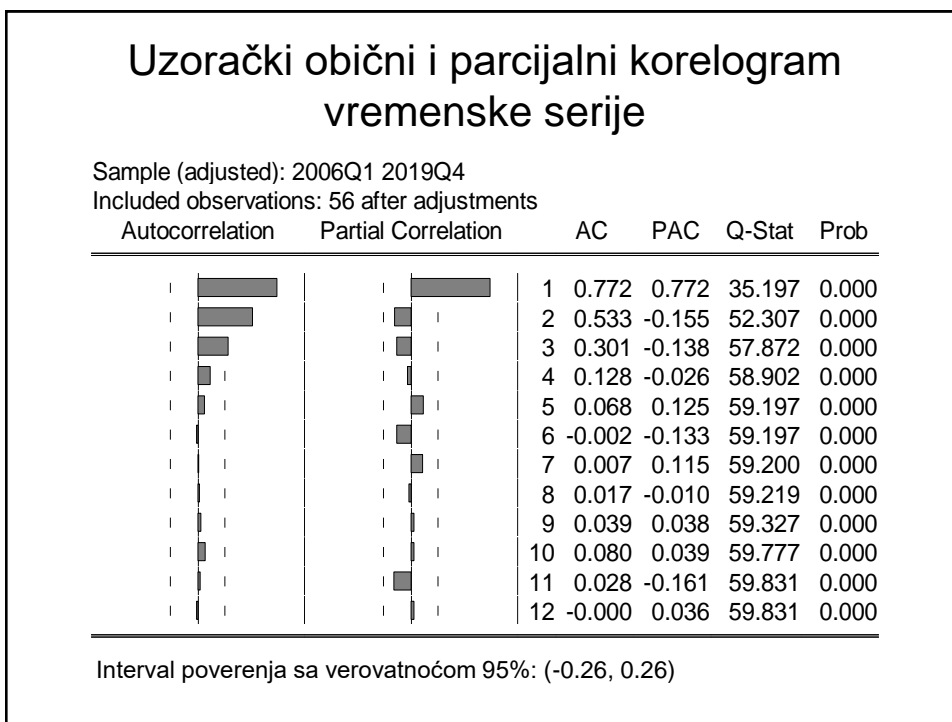
Identifikujemo dva modela na početku: ARMA(1,0) i ARMA(0,3)

Dodatno modeliranje zasniva se na uključivanju veštačke
promenljive ($V_{2006Q1}=1$ za 2006Q1-2008Q3 i 0 ostalo).

32



33



34

Model 1: ARMA(1,0)

Dependent Variable: D4LGDP
 Method: ARMA Conditional Least Squares (Gauss-Newton / Marquardt steps)

Sample (adjusted): 2006Q2 2019Q4
 Included observations: 55 after adjustments
 Convergence achieved after 2 iterations
 Coefficient covariance computed using outer product of gradients

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.022186	0.011166	1.987023	0.0521
AR(1)	0.796926	0.080841	9.857953	0.0000

R-squared	0.647088	Mean dependent var	0.022852
Adjusted R-squared	0.640430	S.D. dependent var	0.028031
S.E. of regression	0.016808	Akaike info criterion	-5.298179
Sum squared resid	0.014974	Schwarz criterion	-5.225185
Log likelihood	147.6999	Hannan-Quinn criter.	-5.269951
F-statistic	97.17924	Durbin-Watson stat	1.635409
Prob(F-statistic)	0.000000		

Inverted AR Roots	.80
-------------------	-----

35

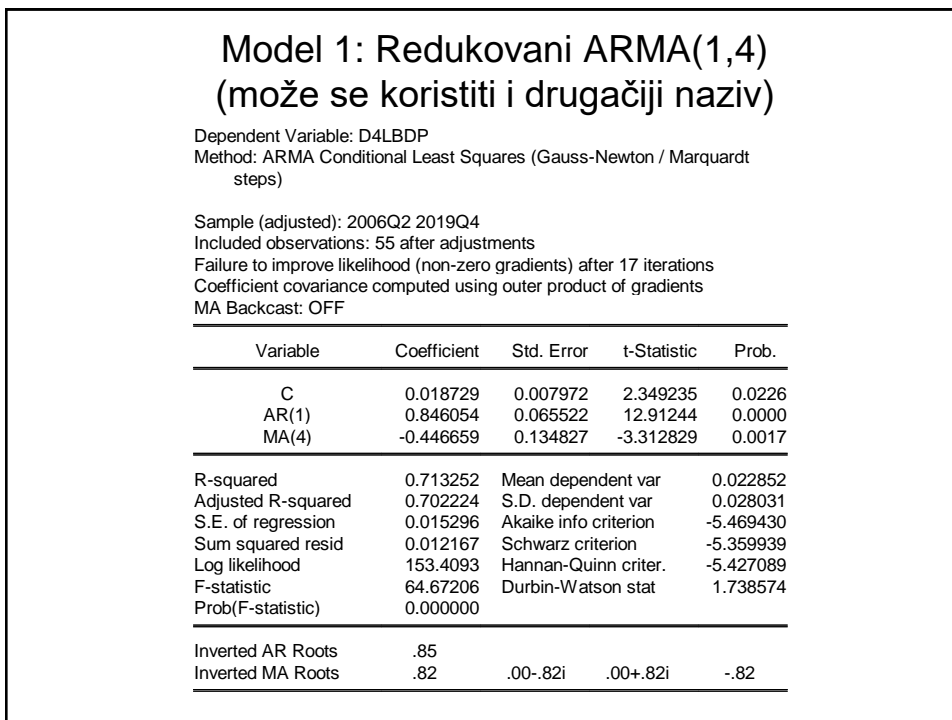
Model 1: ARMA(1,0)

Uzorački obični i parcijalni korelogram **reziduala:** Postoji autokorelacija na docnji 4 i zbirna!

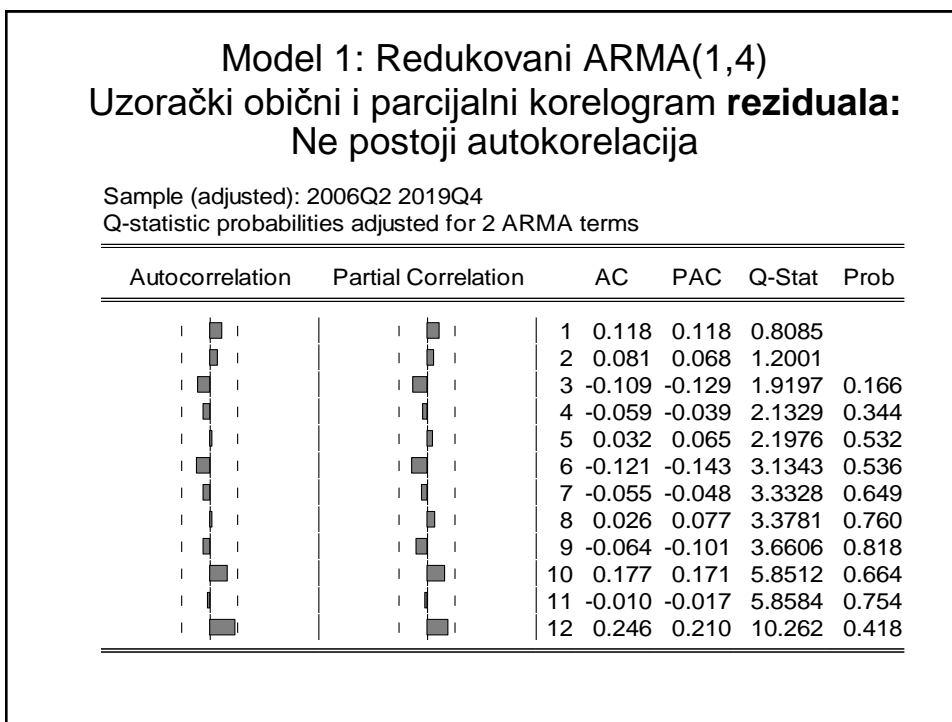
Sample (adjusted): 2006Q2 2019Q4
 Q-statistic probabilities adjusted for 1 ARMA term

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob	
		1	0.170	0.170	1.6770	
		2	0.119	0.093	2.5112	0.113
		3	-0.093	-0.132	3.0347	0.219
		4	-0.385	-0.381	12.124	0.007
		5	0.020	0.188	12.149	0.016
		6	-0.206	-0.177	14.868	0.011
		7	-0.018	-0.066	14.890	0.021
		8	-0.046	-0.148	15.034	0.036
		9	-0.062	0.042	15.294	0.054
		10	0.237	0.149	19.214	0.023
		11	0.046	-0.024	19.364	0.036
		12	0.245	0.131	23.754	0.014

36



37



38

Model 2: ARMA(0,3)

Dependent Variable: D4LBDP
 Method: ARMA Conditional Least Squares (Gauss-Newton / Marquardt steps)

Sample (adjusted): 2006Q1 2019Q4
 Included observations: 56 after adjustments
 Failure to improve likelihood (non-zero gradients) after 11 iterations
 Coefficient covariance computed using outer product of gradients
 MA Backcast: OFF

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.032578	0.006962	4.679582	0.0000
MA(1)	0.945240	0.125437	7.535579	0.0000
MA(2)	0.782287	0.147865	5.290552	0.0000
MA(3)	0.452571	0.126311	3.583001	0.0007
R-squared	0.652501	Mean dependent var	0.023704	
Adjusted R-squared	0.632453	S.D. dependent var	0.028498	
S.E. of regression	0.017277	Akaike info criterion	-5.210118	
Sum squared resid	0.015522	Schwarz criterion	-5.065450	
Log likelihood	149.8833	Hannan-Quinn criter.	-5.154030	
F-statistic	32.54695	Durbin-Watson stat	1.811053	
Prob(F-statistic)	0.000000			
Inverted MA Roots	-.11-.78i	-.11+.78i	-.73	

39

Model 3: Redukovani ARMA(1,4) sa veštačkom promenljivom

Dependent Variable: D4LBDP
 Method: ARMA Conditional Least Squares (Gauss-Newton / Marquardt steps)

Sample (adjusted): 2006Q2 2019Q4
 Included observations: 55 after adjustments
 Failure to improve likelihood (non-zero gradients) after 19 iterations
 Coefficient covariance computed using outer product of gradients
 MA Backcast: OFF

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.016565	0.007396	2.239697	0.0295
D2006Q1	0.034611	0.013594	2.546102	0.0140
AR(1)	0.845194	0.088202	9.582471	0.0000
MA(4)	-0.461975	0.138514	-3.335230	0.0016
R-squared	0.746135	Mean dependent var	0.022852	
Adjusted R-squared	0.731202	S.D. dependent var	0.028031	
S.E. of regression	0.014533	Akaike info criterion	-5.554867	
Sum squared resid	0.010771	Schwarz criterion	-5.408879	
Log likelihood	156.7588	Hannan-Quinn criter.	-5.498412	
F-statistic	49.96479	Durbin-Watson stat	1.837305	
Prob(F-statistic)	0.000000			
Inverted AR Roots	.85			
Inverted MA Roots	.82	.00-.82i	.00+.82i	-.82

40

Sumarni prikaz ocenjenih modela			
Model	1. Redukovani ARMA(1,4)	2. ARMA(0,3)	3. Redukovani ARMA(1,4) + veš. prom.
SC	-5.3599	-5.0655	-5.4089
Stand. greška modela	0.015296	0.017277	0.01453
Q(4)	2.13(0.34)	0.95(0.33)	0.72(0.70)
Q(12)	10.26(0.42)	6.29(0.71)	11.58(0.31)
JB	0.95(0.62)	4.64(0.10)	1.38(0.50)

41