

SARIMA-KALMAN MODEL ZA PREDVIĐANJE OBIMA POŠTANSKIH USLUGA

Nikola Knežević¹, Nebojša Bojović¹, Miloš Milenković¹, Nataša Glišović²

¹Univerzitet u Beogradu – Saobraćajni fakultet

²Državni Univerzitet u Novom Pazaru, Departman za matematičke nauke

Sadržaj: U radu je predstavljen pristup prognoziranju obima poštanskih usluga primenom modela sezonskih autoregresivnih integrisanih pokretnih sredina u stanje-prostor formi. Utvrđeno je da vremenska serija realizacija koje se odnose na ukupan mesečni obim pismonosnih pošiljaka u JP PTT Saobraćaja “Srbija” može adekvatno biti opisana SARIMA (0,1,1)(0,1,1)₁₂ modelom. Identifikovani model je preveden u stanje-prostor okvir, a zatim je Kalmanov filter primenjen za predviđanje ukupnog broja pismonosnih pošiljaka u narednih 6 meseci. Kalmanov postupak je predstavljen kao podesan pristup za predikciju SARIMA procesa u stanje-prostor formi.

Ključne reči: prognoziranje, vremenske serije, SARIMA, Kalman, pismonosne pošiljke

1. Uvod

Proces prognoziranja predstavlja predviđanje budućeg izlaza različitih poslovnih fenomena. Predviđanje obuhvata budućnost poslovanja u celini, budućnost postojeće ili predložene linije proizvoda/usluga i budućnost privredne grane u kojoj se nalazi određena kompanija. Kompanija mora da razume i upotrebi prognoziranje kako bi dobila odgovore na brojna pitanja koja se odnose na ključne determinante performansi kao što su profit, tražnja, trošak proizvodnje i ostale. Ovo pomaže kompaniji da se adekvatno pripremi za budućnost. Na ovaj način organizacija može da načini valjane planove koji će omogućiti ostvarenje vizije i misije kompanije. Ako je prognoziranje cilj analize vremenskih serija onda se polazi od raspoloživih podataka iz prošlosti na osnovu kojih se formuliše i ocenjuje model vremenske serije koji se potom koristi za predviđanje budućih vrednosti serije. Pri tome se koristi niz statističkih testova i kriterijuma kojima se verifikuje valjanost ocenjenog modela.

U ovom radu za analizu date vremenske serije primenjena je klasa autoregresivnih modela pokretnih sredina ARMA(p,q). Kod ove klase modela pretpostavka je da tekuća vrednost (član) serije zavisi od vrednosti prethodnih članova serije, tekuće vrednosti slučajnog procesa i prethodnih vrednosti slučajnog procesa. Ova klasa modela predstavlja kombinaciju autoregresivnog modela i modela pokretnih sredina. Kod vremenskih serija kod kojih se uočava uticaj trenda, ciklične ili sezone

komponente, primena ovih modela podrazumeva prethodno odstranjivanje njihovog uticaja. Za otklanjanje uticaja sistematskih komponenti iz vremenske serije koristi se operator diferenciranja. Ovim postupkom diferenciranja, dobija se klasa ARIMA(p,d,q) modela, kod kojih se originalne vrednosti serije zamenjuju određenim diferencijama. U novije vreme, stanje-prostor interpretacija i Kalmanov filter imaju značajno učešće u analizama vremenskih serija. Kalmanov filter predstavlja unificirani pristup predikciji i estimaciji svih procesa koji se mogu predstaviti u stanje-prostor formi i zasniva se na skupu matematičkih izraza koji obezbeđuje efikasan rekurzivni postupak za estimaciju stanja procesa tako da se minimizira srednja kvadratna greška.

U ovom radu analiziran je obim poštanskih aktivnosti u JP PTT Saobraćaja "Srbija". Preciznije, na osnovu raspoloživih istorijskih podataka (Januar 2004 – Jun 2013) predložen je proces sezonskih autoregresivnih integrisanih pokretnih sredina za modelovanje ukupnog broja pismonosnih pošiljaka. U cilju prognoziranja broja pismonosnih pošiljaka u periodu Jul 2013.-Decembar 2013. godine, estimovani SARIMA model je predstavljen u stanje-prostor okviru, a zatim su Kalmanove rekurzije primenjene.

Rad je organizovan na sledeći način. Osnovne karakteristike ARIMA i SARIMA modela date su u drugom poglavlju. Stanje-prostor interpretacija SARIMA modela je predstavljena u trećem poglavlju. U četvrtom poglavlju, opisan je metod Kalmanovog filtera za predikciju i estimaciju procesa datih u stanje-prostor interpretaciji. Efikasnost predloženog metoda je demonstrirana modelovanjem ukupnog broja pismonosnih pošiljaka u petom poglavlju. Poslednje, šesto poglavlje sadrži zaključna razmatranja i pravce budućih istraživanja.

Rad je rezultat istraživanja na projektu MNTR036022 "Upravljanje kritičnom infrastrukturom za održivi razvoj u poštanskom, komunikacionom i železničkom sektoru Republike Srbije" koji finansira Ministarstvo za nauku Republike Srbije.

2. ARIMA i SARIMA modeli

Autoregresivni integrisani model pokretnih sredina (ARIMA) je metod prvi put primenjen od strane Box-a i Jenkins-a (1976) i predstavlja najpopularniji metod za modelovanje univarijantnih vremenskih serija. Ovaj model je nastao iz autoregresivnog modela (AR), modela pokretnih sredina (MA) i kombinacije AR i MA modela, odnosno ARMA modela (Suhartono, 2011). Sezonske autoregresivne integrisane pokretne sredine se koriste za modelovanje vremenskih serija sa trendom, sezonskim obrascem i kratkoročnim korelacijama. Uopšteni oblik $SARIMA(p, d, q) \times (P, D, Q)_s$ modela za niz Y_t se može formulisati kao (Box et al., 2008; Cryer and Chan, 2008):

$$\phi_p(B)\Phi_P(B^s)(1-B)^d(1-B^s)^D Y_t = \theta_q(B)\Theta_Q(B^s)\varepsilon_t \quad (1)$$

gde s predstavlja periodičnost (sezonalnost) i ε_t je beli šum.

$$\phi_p(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p \quad (2)$$

$$\Phi_P(B^s) = 1 - \Phi_1 B^s - \Phi_2 B^{2s} - \dots - \Phi_P B^{Ps} \quad (3)$$

su nesezonski i sezonski autoregresivni (AR) polinomijalni član reda p and P .

$$\theta_q(B) = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q \quad (4)$$

$$\Theta_Q(B^s) = 1 - \Theta_1 B^s - \Theta_2 B^{2s} - \dots - \Theta_Q B^{Qs} \quad (5)$$

su nesezonske i sezonske pokretne sredine (MA) reda q i Q . $(1-B)^d$ je operator nesezonskog diferenciranja reda d koji se koristi za eliminisanje polinomijalnih trendova dok se operator sezonskog diferenciranja $(1-B^s)^D$ reda D koristi za eliminisanje sezonskih varijacija. B je operator zaostajanja (realizacije, pomeranja unazad) čiji se uticaj na vremensku seriju Y_t može predstaviti kao $B^d Y_t = Y_{t-d}$.

Faza identifikacije modela ima subjektivni karakter s obzirom na to da se zasniva isključivo na grafičkim interpretacijama ACF/PACF procena. Ovaj pristup identifikaciji modela zahteva značajne analitičke veštine i detaljno poznavanje vremenske serije i predstavlja poteškoću u slučajevima složenih ARMA procesa (Prista i ostali, 2011). Kako bi se izbegla subjektivnost identifikacije modela Box-Jenkins metodom i unapredio process određivanja reda ARMA procesa, razvijeni su brojni kriterijumi za selekciju modela (De Gooijer et al., 1985). Najčešće se koriste informacioni kriterijumi kao što su AIC (Akaike informacioni kriterijum), BIC (Bajesov informacioni kriterijum) i normalizovana verzija Bajesovog informacionog kriterijuma. Ovi informacioni kriterijumi se primenjuju za procenu fita nelinearnih modela i obuhvatanje većeg broja parametara u modelu. Sastoje se od prirodnog logaritma srednje kvadratne greške (MSE – Mean Squared Error) i kazne za broj parametara koji su procenjeni:

$$AIC = T \ln(MSE) + 2k \quad (6)$$

$$BIC = T \ln(MSE) + k \ln(T) \quad (7)$$

$$Normalized\ BIC = \ln(MSE) + k \frac{\ln(T)}{T} \quad (8)$$

Gde je T broj realizacija, k je broj parametara u modelu $k = p + q + P + Q + 1$.

3. Stanje-prostor forma SARIMA modela

Stanje-prostor forma pruža unificirano predstavljanje velikog broja linearnih Gausovih modela vremenskih serija uključujući ARMA i UC modele (Hidayanto et al., 2010). Gausova stanje-prostor forma se sastoji od jednačine tranzicije za $m \times 1$ vektor stanja α_t i jednačina merenja za $N \times 1$ vektor realizacija y_t za $t = 1, \dots, n$. Model se može formulisati kao (Durbin and Koopman, 2001):

$$\alpha_{t+1} = T_t \alpha_t + R_t \eta_t, \quad \eta_t \sim N(0, Q_t), \quad t = 1, \dots, n \quad (9)$$

$$y_t = Z_t \alpha_t + \varepsilon_t, \quad \varepsilon_t \sim N(0, H_t), \quad \alpha_1 \sim N(a_1, P_1) \quad (10)$$

gde η_t i ε_t predstavljaju članove greške koji su pretpostavljeni kao međusobno nezavisni i nezavisni od vremenske serije u svim vremenskim periodima. Matrice T_t, R_t, Z_t, Q_t i H_t predstavljaju matrice sistema stanje-prostor. Vektor početnog stanja je α_1 sa vektorom očekivanja a_1 i matricom varijansi P_1 . Model (9)-(10) je linearni sa Gausovim poremećajima. Prema tome, stanje-prostor model se može posmatrati klasičnim metodama zasnovanim na Kalmanovom filteru (Anderson and Moore, 1979; Durbin and Koopman, 2001). Matrica varijansi P vektora početnog stanja α_1 može sadržati difuzne elemente u slučaju postojanja nestacionarnih komponenata u vektoru α_t . U ovom slučaju, kao dodatak Kalmanovom filteru koriste se metode difuzne inicijalizacije za procenu egzaktno ili difuzne funkcije (de Jong, 1991; Koopman, 1997).

Stanje-prostor postavka ARMA/ARIMA modela je data u prethodnoj dekadi (Brockwell and Davis, 2002; Box, Jenkins and Reinsel, 2008). SARIMA modeli se mogu dobiti konstruisanjem ARMA modela za stacionarne diferencirane vremenske serije $y_t^* = (1-B)^d (1-B^s)^D y_t$ i definisanjem nestacionarnih promenljivih kao y_{t-i} i $(1-B)^d y_{t-i}$ u vektoru stanja. y_t^* je sezonski $ARMA(p^*, q^*)$ proces sa $p^* = p + SP$ i $q^* = q + SQ$. Podesno konstruisan vektor stanja ovog ARMA procesa za $d = D = 0$ se može definisati kao α_t^* , tako da je $y_t^* = y_t$. U ovom slučaju postoji

$$\alpha_t^* = (y_t^*, \phi_2 y_{t-1}^* + \dots + \phi_p \Phi_p y_{t-p^*+1} + \theta_1 \varepsilon_t + \dots + \theta_q \Theta_q \varepsilon_{t-q^*+1}, \phi_3 y_{t-1}^* + \dots + \phi_p \Phi_p y_{t-p^*+2} + \theta_2 \varepsilon_t + \dots + \theta_q \Theta_q \varepsilon_{t-q^*+2}, \dots, \phi_p \Phi_p y_{t-1} + \theta_q \Theta_q \varepsilon_t) \quad (11)$$

Pri čemu je dimenzija vektora α_t^* jednaka $m = \max(p^*, q^* + 1)$. Potpun vektor stanja α_t ima dimenziju $SD + d + m$, i za slučaj $d = 1$ i $D = 1$ se može formulisati kao:

$$\alpha_t = (y_{t-1}, (1-B)y_{t-1}, \dots, (1-B)y_{t-S}, \alpha_t^*)^T \quad (12)$$

Pri čemu se y_t^* u vektoru stanja α_t^* menja u skladu sa redom prvog i sezonskog diferenciranja d i D , dok struktura α_t^* ostaje ista (Hindrayanto, 2010). MA parametri su obuhvaćeni vektorom poremećaja,

$$H_t \varepsilon_t = (0_{1 \times (SD+d)}, \varepsilon_{t+1}, \theta_1 \varepsilon_{t+1}, \dots, \theta_{m-1} \varepsilon_{t+1})^T \quad (13)$$

Matrica tranzicije T_t ima dimenzije $(SD + d + m) \times (SD + d + m)$, Z_t je vektor vrste dimenzija je $1 \times (SD + d + m)$. Za sezonske modele sa $d = D = 1$, matrice T_t i Z_t se mogu definisati na sledeći način:

$$T_t = \begin{pmatrix} 1 & 0_{1 \times (S-1)} & 1 & 1 & 0 & 0 & \dots & 0 \\ 0 & 0_{1 \times (S-1)} & 1 & 1 & 0 & 0 & \dots & 0 \\ 0 & I_{S-1} & 0 & 0 & 0 & 0 & \dots & 0 \\ 0 & 0_{1 \times (S-1)} & 0 & \phi_1 & 1 & 0 & \dots & 0 \\ 0 & 0_{1 \times (S-1)} & 0 & \phi_2 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0_{1 \times (S-1)} & 0 & \phi_{m-1} & 0 & 0 & \dots & 1 \\ 0 & 0_{1 \times (S-1)} & 0 & \phi_m & 0 & 0 & \dots & 0 \end{pmatrix} \quad (14)$$

$$Z_t = (1, 0_{1 \times (S-1)}, 1, 1, 0, 0, \dots, 0) \quad (15)$$

Matrica identiteta $r \times r$ se označava sa I_r dok matrica nula ima dimenzije $r \times c$ i označena je kao $0_{r \times c}$.

4. Primena Kalmanovih rekurzija za proces prognoziranja

Kalmanov filter omogućava unificirani pristup za predikciju i estimaciju svih procesa koji se mogu predstaviti u stanje-prostor formi. Klasične Kalmanove rekurzije su uvedene od strane Rudolfa E. Kalmana 1960. godine (Kalman, 1960). Osnovni cilj je da se odredi uslovna raspodela parametra α_{t+1} na osnovu opservacija $Y_t = \{y_1, y_2, \dots, y_t\}$. Kako su sve raspodele normalne, uslovne raspodele određenih podskupova promenljivih za date ostale podskupove promenljivih su takođe normalne. Potrebna raspodela je prema

tome određena poznavanjem $a_{t+1} = E(\alpha_{t+1} | Y_t)$ i $P_{t+1} = Var(\alpha_{t+1} | Y_t)$. Pretpostavljeno je da α_t za dato Y_{t-1} ima normalnu raspodelu $N(a_t, P_t)$. Rekurzivni postupak za određivanje a_{t+1} i P_{t+1} na osnovu a_t i P_t je sledeći:

Pošto je $\alpha_{t+1} = T_t \alpha_t + R_t \eta_t$, postoji

$$a_{t+1} = E(T_t \alpha_t + R_t \eta_t | Y_t) = T_t E(\alpha_t | Y_t) \quad (16)$$

$$P_{t+1} = Var(T_t \alpha_t + R_t \eta_t | Y_t) = T_t Var(\alpha_t | Y_t) T_t' + R_t Q_t R_t' \quad (17)$$

za $t = 1, \dots, n$. Neka je

$$v_t = y_t - E(y_t | Y_{t-1}) = y_t - E(Z_t \alpha_t + \varepsilon_t | Y_{t-1}) = y_t - Z_t a_t \quad (18)$$

Pri čemu v_t predstavlja grešku prognoze jedan korak unapred za y_t uz dato Y_{t-1} .

$$\begin{aligned} E(\alpha_t | Y_t) &= E(\alpha_t | Y_{t-1}, v_t) \\ &= E(\alpha_t | Y_{t-1}) + Cov(\alpha_t, v_t) [Var(v_t)]^{-1} v_t \\ &= a_t + M_t F_t^{-1} v_t \end{aligned} \quad (19)$$

Pri čemu je $M_t = Cov(\alpha_t, v_t) = P_t Z_t'$, $F_t = Var(v_t) = Z_t P_t Z_t' + \sigma^2$ i $E(\alpha_t | Y_{t-1}) = a_t$.

Zamenom izraza (16) u (19):

$$a_{t+1} = T_t a_t + T_t M_t F_t^{-1} v_t = T_t a_t + K_t v_t \quad (20)$$

Gde je

$$K_t = T_t M_t F_t^{-1} = T_t P_t Z_t' F_t^{-1} \quad (21)$$

a_{t+1} se određuje kao linearna funkcija prethodne vrednosti a_t i v_t , greške prognoze y_t za dato Y_{t-1} .

Pošto je

$$Var(\alpha_t | Y_t) = P_t - P_t Z_t' F_t^{-1} Z_t P_t \quad (22)$$

Zamena izraza (22) u (17) daje

$$P_{t+1} = T_t P_t L_t' + R_t Q_t R_t' \quad (23)$$

pri čemu je $L_t = T_t - K_t Z_t'$

Rekurzije (20)-(23) obrazuju Kalmanov filter za model (9)-(10). Ove rekurzije omogućavaju ažuriranje poznavanja sistema sa svakom novom opservacijom (Durbin i Koopman, 2001).

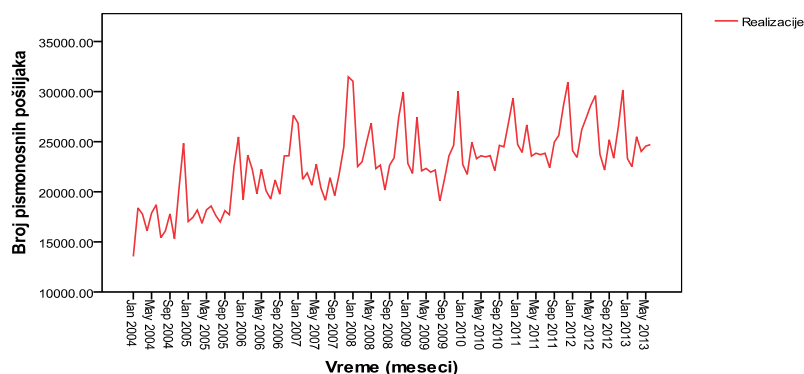
Neophodno je načiniti određena proširenja klasičnih Kalmanovih rekurzija usled delimično difuznog vektora početnog stanja u razmotrenom SARIMA modelu (Durbin and Koopman, 2001; Peng and Aston, 2006). U opštem slučaju, prognoze budućih vrednosti stanja za h perioda unapred se rekurzivno dobijaju kao $a_{t+h} = T_{t+h-1} a_{t+h-1}$ uz matricu kovarijansi $P_{t+1} = T_{t+h-1} P_{t+h-1} T_{t+h-1}' + R_{t+h-1} Q_{t+h-1} R_{t+h-1}'$.

5. Primena Kalmanovih rekurzija za predviđanje obima poštanskih aktivnosti u JP PTT Saobraćaja "Srbija"

5.1. SARIMA proces modelovanja

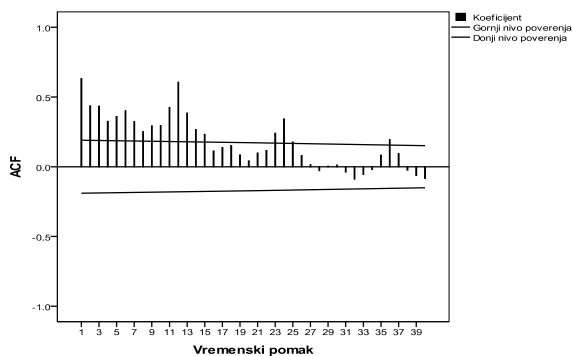
U ovom odeljku predstavljen je pristup Kalmanovog filterisanja za modelovanje obima poštanskih aktivnosti za slučaj JP PTT Saobraćaja "Srbija". Za testiranje različitih SARIMA modela za prognoziranje ukupnog broja pismonosnih pošiljaka za slučaj JP

PTT Saobraćaja “Srbija”, korišćena je vremenska serija ukupnog mesečnog obima pismonosnih pošiljaka – unutrašnje otpremljene i promet sa inostranstvom (RZS, 2013). Skup podataka obuhvata period od januara 2004. godine do juna 2012. godine (114 mesečnih realizacija). Vremenske serije su predstavljene na Slici 1. Prvih 108 realizacija se koristi za estimaciju SARIMA modela, dok su realizacije za poslednjih 6 meseci upotrebljene za procenu performansi izabranog SARIMA modela.



Slika 1. Vremenska serija ukupnog broja pismonosnih pošiljaka za period Januar 2004 – Jun 2013

Velike vrednosti funkcije autokorelacije zabeležene su na vremenskim pomacima 1-15, 23-25 i 36 sa vrednostima 0.632, 0.437, 0.434, 0.326, 0.360, 0.402, 0.325, 0.252, 0.293, 0.295, 0.426, 0.607, 0.385, 0.267, 0.232, 0.240, 0.343, 0.177 i 0.194 (Slika 2.). Postepeno opadanje vrednosti autokorelacije na prva tri vremenska pomaka ukazuje na postojanje dugoročnog trenda. Samim tim, postoji potreba da se strukturom SARIMA modela obuhvati i član prvog diferenciranja ($d = 1$). Takođe su registrovane velike vrednosti autokorelacije na godišnjim pomacima koje ukazuju na potrebu da se sezonsko diferenciranje obuhvati modelom ($S = 12, D = 1$).



Slika 2. Funkcija autokorelacije vremenske serije

Prema tome, $SARIMA(p,1,q) \times (P,1,Q)_{12}$ je izabran kao osnovna struktura SARIMA alternativnog skupa modela. Od svih modela ovog skupa,

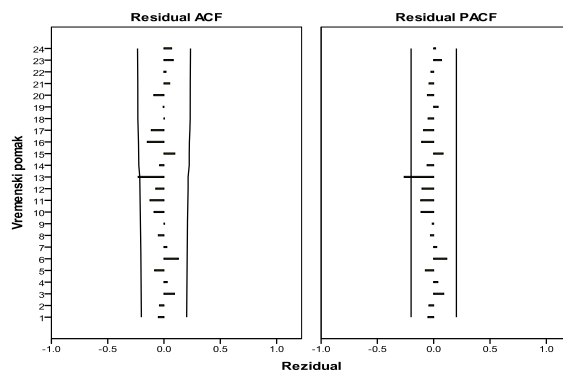
SARIMA(0,1,1)×(0,1,1)₁₂, poznat i kao “*airline*” model predstavlja najbolji model, sa najnižom vrednošću normalizovanog BIC-a od 15.119 i prosečnom apsolutnom procentualnom greškom 5.412 (Tabela 1.). Model obuhvata 77.4% ukupnih varijacija u vremenskoj seriji (R-squared). Parametri modela su značajni o čemu govori i P-vrednost < 0.05. Model ima sledeću jednačinu:

$$(1 - B)(1 - B^{12})Y_t = (1 - 0.756B)(1 - 0.677B^{12})Z_t \quad (24)$$

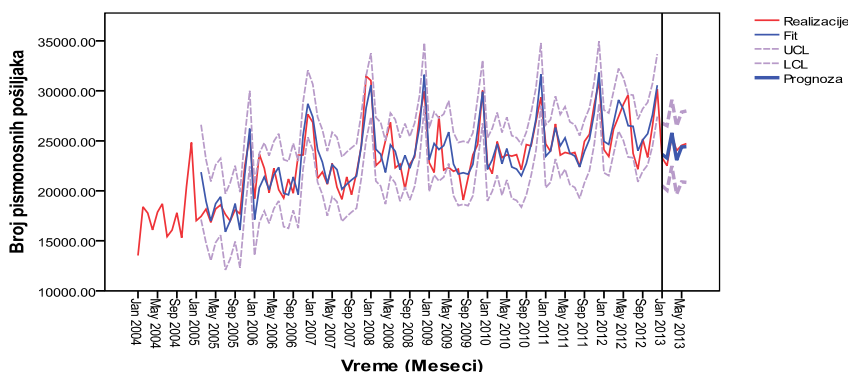
Tabela 1. Normalizovani BIC, srednja apsolutna procentualna greška (MAPE) i stacionarna R-squared vrednost alternativnih SARIMA modela

Models	MAPE	Normalizovana BIC	Stacionarna R-squared
SARIMA(0,1,1)×(0,1,1) ₁₂	5.412	15.119	0.655
SARIMA(0,1,1)×(0,1,0) ₁₂	7.272	15.544	0.253
SARIMA(0,1,0)×(0,1,1) ₁₂	7.684	15.667	0.155
SARIMA(1,1,1)×(0,1,1) ₁₂	6.549	15.412	0.415
SARIMA(1,1,1)×(1,1,1) ₁₂	6.502	15.457	0.421

Dijagnostičke provere ukazuju na zaključak da je izabrani SARIMA model stacionaran, invertibilan i nema redundantne parametre. Reziduali su strogo slučajni (Ljung-Box Q = 18.428, P-vrednost = 0.299 > 0.05), tako da ne postoji značajna autokorelacija između reziduala na različitim vremenskim pomacima (Slika 3.), osim na jednom vremenskom pomaku. Svi testovi su sprovedeni za nivo značajnosti $\alpha = 0.05$. Slika 4. sadrži grafički prikaz realizovanih i fitovanih vrednosti tokom perioda estimacije i validacije modela.



Slika 3. ACF i PACF reziduala fitovanog SARIMA modela



Slika 4. Grafički prikaz izlaza SARIMA modela

5.2. Kalmanove rekurzije

Identifikovani SARIMA model je preveden u stanje-prostor okvir u cilju prognoziranja obima pismonosnih pošiljaka u periodu Jul 2013. – Decembar 2013. SARIMA model je predstavljen preko jednačina stanja i merenja za Kalmanov model. Eksplicitna jednačina odabranog SARIMA(0,1,1) × (0,1,1)₁₂ modela se dobija na sledeći način.

Za $s = 1, \dots, 12$, vektor stanja α_t SARIMA(0,1,1)(0,1,1)₁₂ modela je definisan kao

$$\alpha_t = (y_{t-1}, (1-B)y_{t-1}, \dots, y_{t-12}, (1-B)(1-B)^{12}y_t, 0.756\varepsilon_t + 0.677\varepsilon_{t-11} + 0.512\varepsilon_{t-12}, 0.677\varepsilon_{t-10} + 0.512\varepsilon_{t-11}, 0.512\varepsilon_{t-10}, \dots, 0.512\varepsilon_{t-1}, 0.512\varepsilon_t) \quad (25)$$

Odgovarajući vektor poremećaja je

$$H_t \varepsilon_t = (O_{1 \times 13}, \varepsilon_{t+1}, 0.756\varepsilon_{t+1}, O_{1 \times 11}, 0.512\varepsilon_{t+1}) \quad (26)$$

Matrica tranzicije T_t je dimenzija 27×27 , dok su T_t i Z_t date kao:

$$T_t = \begin{bmatrix} T_a & T_b \\ T_c & T_d \\ T_e & T_f \end{bmatrix}, T_a = \begin{bmatrix} 1 & O_{1 \times 11} & 1 & 1 \\ 0 & O_{1 \times 11} & 1 & 1 \\ O_{11 \times 1} & I_{11} & 0 & 0 \end{bmatrix}, T_b = [O_{13 \times 13}], T_c = [O_{13 \times 14}] \quad (27)$$

$$T_d = [I_{13}], T_e = [O_{1 \times 14}], T_f = [O_{1 \times 13}], Z_t = [1 \ O_{1 \times 11} \ 1 \ 1 \ O_{1 \times 13}]$$

Matrica varijansi-kovarijansi poremećaja stanja je data sa:

$$H_t H_t^i = \begin{bmatrix} O_{13 \times 13} & O_{13 \times 14} \\ O_{14 \times 13} & H_t^* H_t^{*i} \end{bmatrix} \quad (28)$$

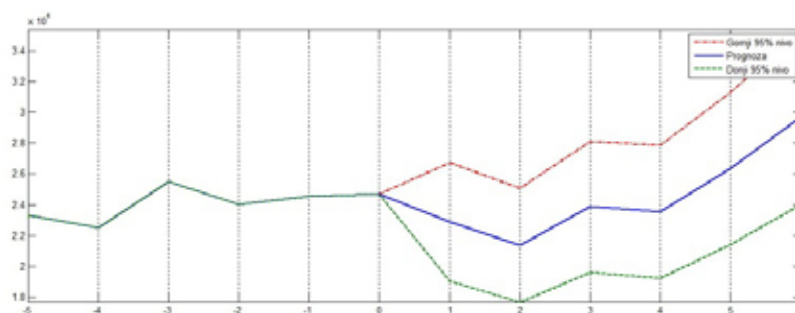
Pri čemu je stacionarni deo ove matrice varijansi, dat matricom $H_t^* H_t^{*i}$ dimenzija 14×14 ,

$$H_t^* H_t^{*i} = \begin{bmatrix} 1 & 0.756 & O_{1 \times 11} & 0.677 \\ 0.756 & 0.572 & O_{1 \times 11} & 0.512 \\ 0 & 0 & O_{11 \times 11} & 0 \\ 0.677 & 0.512 & O_{1 \times 11} & 0.458 \end{bmatrix} \quad (29)$$

Očekivana vrednost početnog vektora stanja je data sa $a = E(\alpha_1) = O_{27 \times 1}$ i odgovarajuća matrica varijansi je

$$P = \begin{bmatrix} kI_{13} & O_{13 \times 14} \\ O_{14 \times 13} & P_{14 \times 14}^* \end{bmatrix} \text{ with } k \rightarrow \infty \quad (30)$$

Matrica $P_{14 \times 14}^*$ je nezavisna matrica varijansi za stacionarni deo vektora stanja. Proces prognoziranja je sproveden u Matlab-u korišćenjem SSM Toolbox-a (Gomez, 2012) u kome su primenjeni odgovarajući metod difuzne inicijalizacije za matricu kovarijansi stanja i Kalmanov filter. Slika 5. prikazuje prognozirane vrednosti ukupnog broja pismonosnih pošiljaka u JP PTT Saobraćaja "Srbija" za period Jul 2013.-Decembar 2013., dobijene primenom SARIMA-Kalman modela.



Slika 5. Prognoza ukupnog broja pismonosnih pošiljaka za period Jul 2013.-Decembar 2013.

6. Zaključak

U ovom radu predložen je sofisticirani matematički metod za analizu obima poštanskih aktivnosti. SARIMA(0,1,1)(0,1,1)₁₂ model se pokazao kao najpodesniji (među izabranim alternativnim SARIMA modelima) za modelovanje ukupnog broja pismonosnih pošiljaka u JP PTT Saobraćaja "Srbija". Ovaj model je preveden u stanje-prostor formu u cilju primene Kalmanovih rekurzija za prognoziranje ukupnog broja pismonosnih pošiljaka u budućem periodu. Međutim, iskustveno je dokazano da prognoze zasnovane na kombinaciji demografskih studija i tržišnih istraživanja daju bolje rezultate. Samim tim, u cilju poboljšanja predloženog modela biće razmotreni efikasniji strukturni dinamički modeli kojima se eksplicitno tretira trend, sezonske i ciklične varijacije u stanje-prostor formi koji će takođe sadržati i dodatne informacije iz tržišnih istraživanja i demografskih studija.

Literatura

- [1] G. Box, and M. Jenkins, Time series analysis: forecasting and control. San Francisko: Holden-Day, 1976.
- [2] Suhartono, "Time Series Forecasting by using Seasonal Autoregressive Integrated Moving Average: Subset, Multiplicative or Additive Model," Journal of Mathematics and Statistics, Vol. 7, No. 1, pp. 20-27, 2011.
- [3] G. E. Box, G. M. Jenkins, and G. C. Reinsel, Time Series Analysis, Forecasting and Control. New Jersey: John Wiley and Sons, 2008.
- [4] J. D. Cryer, and K. S. Chan, Time Series Analysis: with Application in R. New York: Springer, 2008

- [5] N. Prista, N. Diawara, M. Jose Costa, and C. Jones, "Use of SARIMA models to assess data poor fisheries: a case study with a sciaenid fishery off Portugal," *Fishery Bulletin*, Vol. 109, No. 2, pp. 170-185, 2011.
- [6] J.G. De Gooijer, B. Abraham, A. Gould, and L. Robinson, "Methods for determining the order of an autoregressive-moving average process: a survey," *International Statistical Review*, Vol. 53, pp. 301-329, 1985.
- [7] I. Hindrayanto, S. J. Koopman, and M. Ooms, "Exact maximum likelihood estimation for non-stationary periodic time series models," *Computational Statistics & Data Analysis*, Vol. 54, No. 11, pp. 2641-2654, 2010.
- [8] B. D. Anderson, and J. B. Moore, *Optimal Filtering*. Englewood Cliffs: Prentice-Hall, 1979.
- [9] J. Durbin, and S. Koopman, *Time series analysis by state space methods*. Oxford University Press, 2001.
- [10] P. De Jong, "The diffuse Kalman filter," *The Annals of Statistics*, Vol. 19, pp. 1073-1083, 1991.
- [11] S. Koopman, "Exact Initial Kalman Filtering and Smoothing for Nonstationary Time Series Models," *Journal of the American Statistical Association*, Vol. 92, No. 440, pp. 1630-1638, 1997.
- [12] P. Brockwell, and R. Davis, *Introduction to time series and forecasting*. Springer Verlag, 2002.
- [13] R. Kalman, "A new approach to linear filtering and prediction problems," *Journal of basic Engineering*, Vol. 82, pp. 35-45, 1960.
- [14] J. Peng, and J. Aston, "The State Space Models Toolbox for MATLAB," *Journal of Statistical Software*, Vol. 41, No. 6, pp. 1-25, 2006.
- [15] RZS, "Saobraćaj, skladištenje i veze - godišnji statistički izveštaj," *Zavod za statistiku, Republika Srbija*, 2013.
- [16] V. Gomez, "SSMMATLAB, a Set of MATLAB Programs for the Statistical Analysis of State-Space Models," Online at <http://www.sepg.pap.minhap.gob.es/sitios/sgpg/en-GB/Presupuestos/Documentacion/Paginas/SSMMATLAB.aspx>, 2012.

Abstract: *This paper presents a forecasting approach of postal services volume by use of the Seasonal Autoregressive Integrated Moving Average (SARIMA) models in state-space form. It was found that the recurring pattern of the total monthly volume of letter mailing services is well described by SARIMA (0,1,1)(0,1,1)₁₂ model. The identified model was incorporated in the state space framework where classical Kalman recursion is applied for calculation of forecasting values. The Kalman procedure is presented as an elegant approach for prediction of SARIMA processes in state space form.*

Keywords: *forecasting, time series, SARIMA, Kalman filtering, postal services*

SARIMA-KALMAN MODEL FOR POSTAL SERVICES VOLUME FORECASTING

Nikola Knežević, Nebojša Bojović, Miloš Milenković, Nataša Glišović