

PRIMENA BOX-JENKINS PRISTUPA ZA MODELOVANJE OBIMA POŠTANSKIH USLUGA *

Nikola Knežević, Nebojša Bojović, Miloš Milenković, Dragana Macura
Univerzitet u Beogradu – Saobraćajni fakultet

Sadržaj: *U tržišnim uslovima privređivanja uspešno poslovanje kompanije zavisi od praćenja, analize i predviđanja, kako rezultata, tako i najvažnijih faktora koji na njih utiču. Analiza stanja i predviđanje mogu biti zasnovani na uređenom nizu podataka u jednakim vremenskim intervalima, odnosno na analizi vremenskih serija posmatranih pojava. U ovom radu, analiziran je proces isporuke preporučenih pošiljaka i za određenu vremensku seriju realizacija tokom perioda 2006.-2010. godine načinjen odgovarajući ARIMA model za prognoziranje broja preporučenih pošiljaka za izabranu poštansku upravu JP PTT Saobraćaja "Srbija" tokom 2011. godine.*

Ključne reči: *prognoziranje, vremenske serije, ARIMA, poštanske usluge*

1. Uvod

Prognoziranje je proces na osnovu kog se kompanije pripremaju za budućnost. Proces prognoziranja predstavlja predviđanje budućeg izlaza različitih poslovnih fenomena. Predviđanje obuhvata budućnost poslovanja u celini, budućnost postojeće ili predložene linije proizvod/usluga, i budućnost privredne grane u kojoj se nalazi određena kompanija. Na osnovu procesa prognoziranja moguće je dati odgovor na neka od sledećih pitanja:

- Koliki profit će kompanija ostvariti?
- Kolika tražnja će postojati za određeni proizvod/uslugu?
- Koliko će koštati proizvodnja određenog proizvoda/usluge?
- Koliko će novčanih sredstava kompanija morati da pozajmi?

Kompanija mora da razume i upotrebi prognoziranje kako bi dobila odgovore na ova važna pitanja. Ovo pomaže kompaniji da se pripremi za budućnost. Takođe, pomaže organizaciji da načini planove koji će omogućiti ostvarenje vizije i misije kompanije.

Ako je prognoziranje cilj analize vremenskih serija polazi se 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 u analizi i predviđanju 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 je kombinacija autoregresivnog modela i modela pokretnih sredina. Kod vremenskih serija kod kojih se uočava uticaj trend, ciklične ili sezone

* Ovaj tekst je rezultat rada na projektu 36022: "Upravljanje kritičnom infrastrukturom za održivi razvoj u poštanskom, komunikacionom i železničkom sektoru Republike Srbije", koji se realizuje uz finansijsku podršku Ministarstva za nauku i tehnološki razvoj Republike Srbije.

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 ovom radu, detaljno je izložen ARIMA model, zatim isti razvijen za usluge poštanskih pošiljaka JP PTT Saobraćaja "Srbija" na osnovu podataka o realizovanom broju preporučenih pošiljaka u periodu između 2006. i 2010. godine i konačno, primenjen za prognoziranje obima preporučenih pošiljaka tokom 2011. godine. Proces modelovanja i prognoziranja sproveden je korišćenjem probne verzije softverskog paketa IBM SPSS Statistics.

Rad je organizovan na sledeći način. U drugom poglavlju date su teorijske osnove analize vremenskih serija i predstavljena Box-Jenkins metodologija za analiziranje i modelovanje vremenskih serija. U trećem poglavlju primenjena je predstavljena metodologija za modelovanje obima poštanskih usluga za Javno preduzeće PTT saobraćaja „Srbija”. Četvrto poglavlje sadrži zaključna razmatranja i pravce budućih istraživanja.

Rad je nastao kao 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. Teorijske osnove analize vremenskih serija

Vremenska serija predstavlja skup vrednosti neprekidne promenljive Y (Y_1, Y_2, \dots, Y_n) uređenih u odnosu na diskretnu indeks promenljivu t ($1, 2, \dots, n$). Termin "vremenska serija" potiče iz ekonometrijskih istraživanja gde indeks promenljiva predstavlja vremenske intervale. Međutim, treba napomenuti da direktna povezanost sa vremenom nije neophodna, odnosno, različito značenje može imati indeks promenljiva. U datoj vremenskoj seriji moguće je prepoznati [1]:

- Pravilnu, dugoročnu komponentu varijabilnosti, poznatu kao trend, koja predstavlja celokupan razvoj vremenske serije;
- Pravilnu, kratkoročnu komponentu čiji oblik nastaje periodično u određenim intervalima indeks promenljive, poznatu kao sezonalnost;
- $AR(p)$ autoregresivnu komponentu reda p , koja povezuje svaku vrednost $Z_t = Y_t - (\text{trend i sezonalnost})$ sa p prethodnih Z vrednosti, u skladu sa sledećim linearnim odnosom:

$$Z_t = \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + \dots + \phi_p Z_{t-p} + \varepsilon_t \quad (1)$$

pri čemu $\phi_i (i = 1, \dots, p)$ predstavljaju nepoznate parametre koje treba proceniti, dok ε_t predstavlja grešku.

- $MA(q)$ komponentu pokretne sredine reda q , koja povezuje svaku vrednost Z_t sa q reziduala svih prethodnih q procena Z :

$$Z_t = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q} \quad (2)$$

pri čemu $\theta_i (i = 1, \dots, q)$ su parametri koji trebaju biti procenjeni.

Autoregresivna komponenta i komponenta pokretne sredine se mogu kombinovati u jedan autoregresivni model pokretne sredine ARMA(p,q):

$$Z_t = \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + \dots + \phi_p Z_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q} \quad (3)$$

U preliminarnim analizama vremenske serije korisno je posebno proceniti dugoročne i kratkoročne periodične komponente, koje su važne za definisanje pravilne strukture vremenske serije. Trend komponenta se može proceniti fitovanjem pravilne, polinomijalne ili složenije opšte funkcije. Sezonska komponenta se može proceniti postupkom sezonske dekompozicije, koji proračunava sezonski indeks zasnovan na odnosu realizovanih vrednosti i pokretne sredine. U konačnoj fazi modelovanja serije, i trend i sezonska komponenta se moraju integrisati u ARMA(p,q) proces. Za trend, integracija se dobija korišćenjem diferentnog linearnog operatora (∇) koji je definisan sa $\nabla Y_t = Y_t - Y_{t-1}$. Jednom primenjen ∇ operator koriguje linearni rastući trend vremenske serije, dok njegova ponovna upotreba d puta koriguje trend koji se može fitovati polinomom d-tog reda. Stacionarna serija Z_t se može dobiti kao d-ta razlika (∇^d) od Y_t ,

$$Z_t = \nabla^d Y_t \quad (4)$$

odnosno može se modelovati jednim ARMA(p,q) procesom. Kombinovana upotreba ∇ operatora i ARMA(p,q) procesa rezultira jednim ARIMA(p,d,q) modelom.

Box-Jenkins modeli mogu da obuhvate i sezonske autoregresivne članove i sezonske članove pokretne sredine. U fazi identifikacije modela, cilj je utvrditi sezonalnost, i ako postoji potrebno je identifikovati red za sezonske autoregresivne članove i članove pokretne sredine [2]. Sezonski ARIMA (p,d,q) model je jedan ARIMA (p,d,q) model čiji reziduali ε_t mogu dalje biti modelovani jednom ARIMA (P,D,Q)_s strukturom sa linearnim operatorima (P,D,Q) koji su funkcije B^s operatora. B je operator povratne sprege (na primer, $BY_t = Y_{t-1}$, $B^2 Y_t = Y_{t-2}$, ..., $B^s Y_t = Y_{t-s}$), dok s predstavlja sezonski vremenski pomak. Sezonski ARIMA, odnosno ARIMA(p,d,q)(P,D,Q)_s model je definisan kao [3]:

$$\phi_p(B)\delta_p(B^s)\nabla^d\nabla_s^D Y_t = \chi_Q(B^s)\theta_q(B)\varepsilon_t \quad (5)$$

gde je

$$\phi_p(B) = 1 - \phi_1 B - \dots - \phi_p B^p \quad (6)$$

$$\theta_q(B) = 1 - \theta_1 B - \dots - \theta_q B^q \quad (7)$$

$$\delta_p(B^s) = 1 - \delta_1 B^s - \dots - \delta_p B^{sp} \quad (8)$$

$$\chi_Q(B^s) = 1 - \chi_1 B^s - \dots - \chi_Q B^{sQ} \quad (9)$$

Gde su δ i ϕ sezonski i nesezonski autoregresivni parametri, dok χ i θ predstavljaju sezonske i nesezonske parametre pokretne sredine. Parametri p i q predstavljaju redove nesezonskih autoregresivnih i parametara pokretne sredine, dok P i Q predstavljaju redove sezonskih autoregresivnih i parametara pokretne sredine. Takođe, d i D označavaju nesezonsku i sezonsku razliku.

2.1. Box-Jenkins metodologija za analiziranje i modelovanje vremenskih serija

Postoji tri osnovne faze u analiziranju i modelovanju Box-Jenkins modela vremenskih serija [4]:

- Identifikacija modela
- Estimacija modela
- Validacija modela

Identifikacija modela

Faza identifikacije je najvažnija i najsloženija. U ovoj fazi potrebno je odrediti adekvatan model iz ARIMA familije modela. Opšti Box-Jenkins model obuhvata diferentne operatore, autoregresivne članove, članove pokretne sredine, sezonske diferentne operatore, sezonske autoregresivne članove i sezonske članove pokretne sredine.

Stacionarnost u Box-Jenkins modelima

U okviru faze identifikacije potrebno je proveriti postojanje stacionarnosti i sezonalnosti vremenske serije. Box-Jenkins model pretpostavlja da je vremenska serija stacionarna.

Stacionarna vremenska serija ima konstantnu sredinu, konstantnu varijansu i konstantnu autokorelaciju. Box i Jenkins predlažu diferenciranje nestacionarnih serija jednom ili više puta kako bi se postigla stacionarnost. Diferenciranjem se dobija ARIMA model, sa "I" koje predstavlja integraciju. Ako je prva razlika realizacija vremenske serije $Y_t - Y_{t-1}$ stacionarna, tada je vremenska serija integrisana reda 1. Takođe, ako je u vremenskoj seriji prisutna sezonalnost, neophodno je identifikovati red sezonskih autoregresivnih članova i sezonskih članova pokretne sredine [5].

Nakon što su stacionarnost i sezonalnost utvrđeni, sada je potrebno identifikovati red (p i q) autoregresivnog i člana pokretne sredine. Osnovni alati za ovo su grafik funkcije autokorelacije (ACF) i parcijalne funkcije autokorelacije (PACF) koji se porede sa teorijskim ponašanjem ovih funkcija. Za AR(1) proces, funkcija autokorelacije bi trebala da ima eksponencijalno opadajući oblik. AR procesi višeg reda su obično miks eksponencijalno opadajućih i prigušenih sinusoidnih komponenata. Za autoregresivne procese višeg reda neophodno je funkciju autokorelacije dopuniti funkcijom parcijalne autokorelacije. Parcijalna autokorelacija AR(p) procesa postaje 0 na vremenskom pomaku p+1 i više. Podatak je AR(p) ako: ACF konstantno opada, dok PACF naglo nestaje posle p vremenskih pomaka. Funkcija autokorelacije MA(q) procesa postaje 0 na vremenskom pomaku q+1 i više. Podatak je MA(q) ako: ACF naglo nestaje posle q vremenskih pomaka, i PACF opada konstantno.

Estimacija modela

Za identifikovani model, naredni korak je rekurzivni postupak estimovanja parametara modela, odnosno fitovanje modela. Osnovni pristupi fitovanju Box-Jenkins modela su metod nelinearnih najmanjih kvadrata i metod maksimalne verodostojnosti.

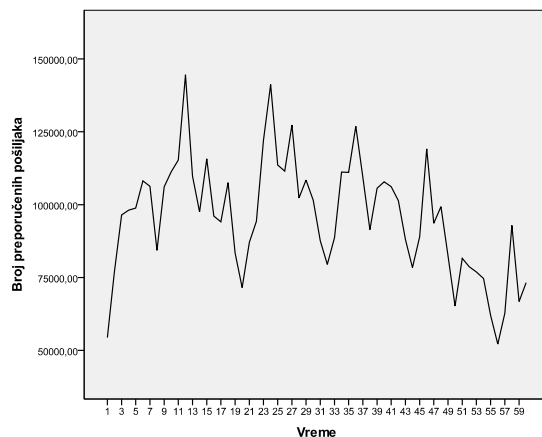
Validacija modela

Nakon što je odgovarajući model usvojen, i njegovi parametri estimovani, Box-Jenkins metodologija zahteva proveru reziduala, odnosno razlike aktuelnih vrednosti i onih estimovanih od strane modela. Ako su reziduali slučajni, model je odgovarajući. Nekoliko testova (na primer, Box-Pierce statistika) su predloženi kao pomoć za određivanje slučajnosti reziduala. U suprotnom, neophodno je fitovati bolji model. Odnosno, potrebno je vratiti se na fazu identifikacije modela i probati sa pronalaženjem boljeg modela.

3. Formiranje ARIMA modela za prognoziranje obima poštanskih usluga JP PTT Saobraćaja "Srbija"

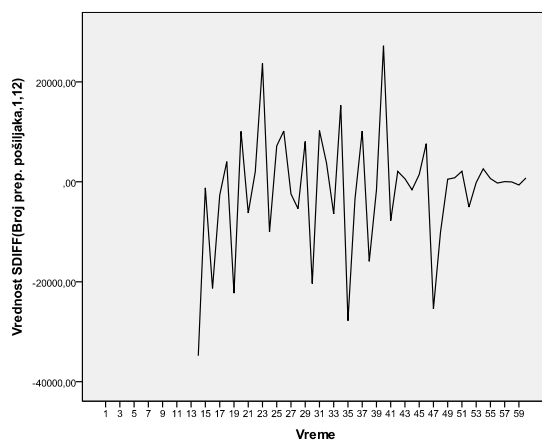
U ovom radu izvršeno je prognoziranje budućeg obima preporučenih pošiljaka za dostavu na teritoriji odabrane jedinice poštanske mreže.

Prema tome, promenljiva koja se koristi u analizi je mesečni obim preporučenih pošiljaka u periodu od 2006. godine do 2010. godine. Grafički prikaz realizovanog obima običnih pošiljaka za analizirani period dat je na slici 1.



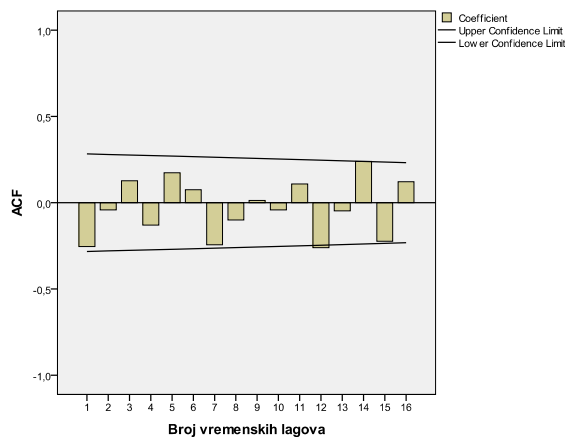
Slika 1. *Obim realizovanih preporučenih pošiljaka u periodu između 2006. i 2010. godine*

Kao što se sa slike može uočiti, vremenska serija je nestacionarna. Stacionarna vremenska serija ima konstantnu sredinu i konstantnu varijansu. Nestacionarnost je korigovana odgovarajućim diferenciranjem podataka. U ovom slučaju razlika 1. reda i sezonska razlika 1. reda su dovoljne za postizanje stacionarnosti (Slika 2.).

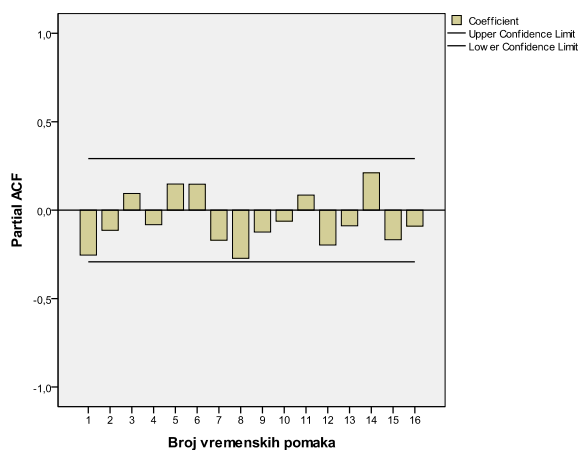


Slika 2. *Stacionarna vremenska serija obima realizovanih preporučenih pošiljaka*

U narednoj fazi Box-Jenkins metodologije potrebno je utvrditi red modela, odnosno identifikovati vrednosti p i q parametara. Za određivanje reda modela, neophodno je odrediti koeficijente autokorelacije i parcijalne autokorelacije promenljive. Na slici 2, prikazana je funkcija autokorelacije, dok je na slici 3. predstavljena funkcija parcijalne autokorelacije.



Slika 3. Funkcija autokorelacije diferenciranih podataka



Slika 4. Funkcija parcijalne autokorelacije diferenciranih podataka

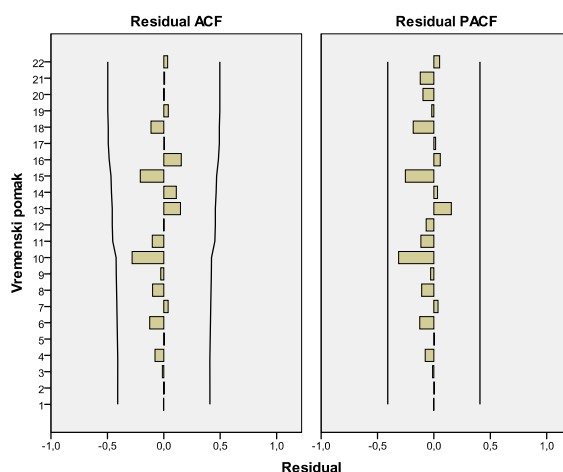
Poređenjem oblika funkcije autokorelacije i funkcije parcijalne autokorelacije u odnosu na usvojene teorijske karakteristike ovih funkcija za različite modele (Tabela 1.) može se zaključiti da je model tipa $ARIMA(0,1,1)(1,1,0)$ ¹², odnosno $p=0$, $d=1$, $q=1$ i $P=1$, $D=1$ i $Q=0$, pri čemu P , D , i Q predstavljaju sezonske komponente ARIMA modela čiji se red utvrđuje na isti način kao u slučaju osnovnih komponenti.

Tabela 1. Teorijske karakteristike ACF i PACF za modele iz ARIMA familije

Vrsta modela	Klasičan oblik ACF	Klasičan oblik PACF
AR(p)	Opada eksponencijalno	Naglo opada posle p vremenskih pomaka
MA(q)	Naglo opada posle q vremenskih pomaka	Opada eksponencijalno
ARMA(p,q)	Opada eksponencijalno	Opada eksponencijalno

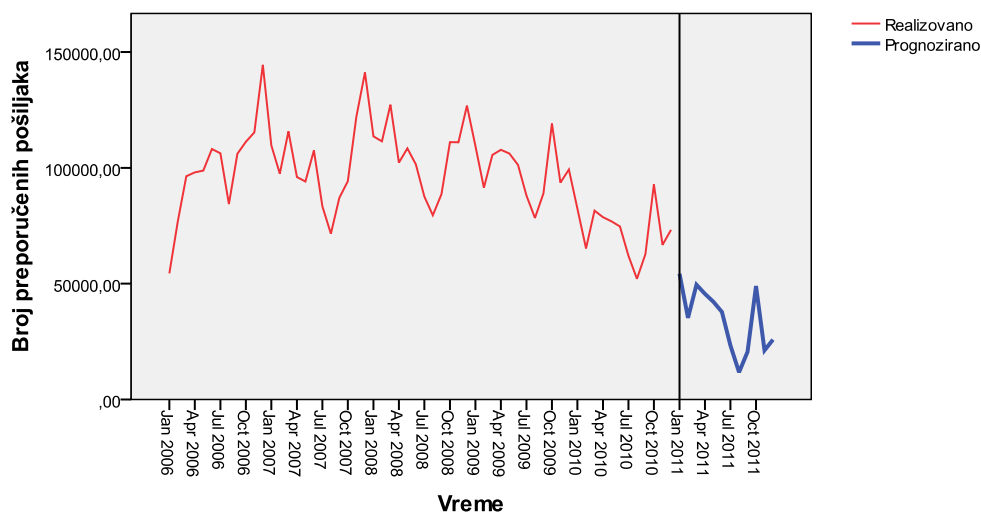
Verifikacija izabranog modela sprovedena je analizom autokorelacija i parcijalnih autokorelacija reziduala različitih redova. Za ovu svrhu, različite korelacije do 14 vremenskih pomaka su određene i testirane preko Box-Ljung testa. Kao što može videti na Slici 5., nijedna od ovih korelacija nije značajno veća od nule. Ovo dokazuje da je izabrani ARIMA model odgovarajući model za fitovanje i prognoziranje broja preporučenih pošiljaka. Prema tome, izabrani model je $ARIMA(0,1,1)(1,1,0)^{12}$, koji se može formulirati kao:

$$(1 + 0.268B^{12})(1 - B)(1 - B^{12})Y_t = -1537.020 + (1 + 0.133B)\varepsilon_t \quad (10)$$



Slika 5. ACF i PACF reziduala fitovanog ARIMA modela

Odabrani model je upotrebljen za prognozu kretanja broja preporučenih pošiljaka za period januar 2011. – decembar 2011. godine. Na slici 6. je prikazana projekcija vremenske serije za izabrani period prognoziranja.



Slika 6. Prognozirane vrednosti broja preporučenih pošiljaka za 2011. godinu

4. Zaključak

U radu je analizirana mogućnost izbora odgovarajućeg ARIMA modela koji poseduje veliku moć predviđanja. Identifikovan je okvir za ARIMA modelovanje koji obuhvata sledeće korake: prikupljanje podataka i ispitivanje, određivanje reda integracije, identifikaciju modela, validaciju i procenu performansi prognoze. Usvojen je Box-Jenkins pristup prognoziranja poznat kao ARIMA modelovanje u kome su vremenske serije izražene preko prošlih vrednosti (autoregresivna komponenta) i tekućih i prošlih vrednosti „belog šuma“ člana greške (komponenta pokretne sredine). U funkcionisanju poštanskog sistema od izuzetne važnosti je, između ostalog, i što preciznije prognoziranje budućeg obima pošiljaka. U postupku prognoziranja potrebno je pored faktora (ekonomskih, poštanskih, društvenih i tehnoloških) uzeti u obzir i kretanja obima poštanskih pošiljaka u prethodnom periodu. Istraživanje je zasnovano na mesečnim podacima o broju preporučenih pošiljaka za izabranu poštansku upravu u okviru JP PTT Saobraćaja „Srbija“ za period 2006-2010. godina. Kvalitetno prognoziranje obima pošiljaka na dostavi u narednom periodu omogućava menadžerima u Pošti da na adekvatan način planiraju buduće potrebe kadrova, radi efikasnog realizovanja poslova dostave. Na osnovu Box-Jenkins metodologije izabran je najbolji model za prognoziranje broja preporučenih pošiljaka tokom 2011. godine.

Literatura

- [1] Box, G. E. P., Jenkins, G. M. and Reinsel, G. C (1994). Time Series Analysis, Forecasting and Control, 3rd ed., Prentice Hall, Englewood Cliffs.
- [2] Salam M.A., Salam S., Feridun M. (2006). Forecasting Inflation in Developing Nations: The Case of Pakistan. International Research Journal of Finance and Economics, 3, 138-159.
- [3] Dobre I, Alexandru A. A. (2008). Modelling unemployment rate using Box-Jenkins procedure. Journal of Applied Quantitative Methods, 3(2), 156-166.
- [4] Montgomery D.C., Jennings C.L., Kulahci M. (2007). Introduction to Time Series Analysis and Forecasting. John Wiley & Sons, Inc., New Jersey.
- [5] Enders W. (2009). Applied Econometric Time Series. John Wiley & Sons, Inc., New Jersey.

Abstract. *In modern market conditions successful business of a company depends on monitoring, analysis, and predicting of business results and the most important factors influencing on these results. State analysis and prediction can be based on an ordered set of data in equally distributed time periods, or in other words, on analysis of time series of data. In this paper, a process of registered letters distribution has been analyzed for a certain time period from 2004. to 2008. and a good fitted ARIMA model for forecasting the volume of registered letters has been determined for a selected postal unit of P.E. of PTT Communications “Srbija”*

Keywords: *forecasting, time series, ARIMA, postal services*

POST SERVICE VOLUME MODELLING USING BOX-JENKINS PROCEDURE

Nikola Knezevic, Nebojsa Bojovic, Milos Milenkovic, Dragana Macura