PREGLEDNI RAD
UDK 338.57:338.45>:519.23

Prof. dr. sc. Zdenka Gogala
Gordan Žanić

EVALUACIJA SKUPINE PROGNOSTIČKIH MODELA PRIMJENOM DVOSTRUKOG EKSPONENCIJALNOG IZGLAĐIVANJA

EVALUATION OF FORECASTING MODELS BY APPLYING DOUBLE EXPONENTIAL SMOOTHING

SAŽETAK: Uspješnost vođenja gospodarske politike jedne zemlje povezana je s uspješnošću prognoza kretanja gospodarskih veličina temeljem relevantnih vremenskih nizova. Predmet ovoga rada je poredbena analiza uspješnosti dvaju pristupa u prognoziranju 11 konkretnih prognostičkih modela. Modeli su primijenjeni u praćenju dinamike serije indeksa cijena industrijskih proizvoda proizvođača u razdoblju od 1997. do 2006. godine u Hrvatskoj te kratkoročnom prognoziranju tih indeksa. Osnovna hipoteza rada je bila da bi Holtov model dvostrukog eksponencijalnog izglađivanja u prisutnosti trenda trebao biti relativno uspješniji od ostalih. No, pokazalo se da je on, unatoč metodološkoj opravdanosti uporabe, manje uspješan od prognostičkog modela temeljenog na linearnom trendu. Unutar skupine Holtovih modela, jedan od njih se pokazao značajno superiornijim prema ostalima.

KLJUČNE RIJEČI: Holtov model eksponencijalnog izglađivanja, ekstrapolacija modelom linearnog trenda, MAPE, MAD.

ABSTRACT: In every economy, it is important for policymakers to create good forecasts that rely on logical methods of manipulating the data that have been generated by historical events. Time-series techniques attempt to account for changes over time by examining patterns, cycles, trends, or using information about previous time periods.

In this paper eleven forecasting models were applied to the time series in producers' price indices 1997-2006, with emphasis on evaluating the efficiency of forecasting models applied.

It was hypothesized that applying Holt's exponential smoothed forecast model to series that exhibit trend should give the best results, but the best results were obtained with the linear trend model. Holt's double exponential smoothed model proved to be less efficient. Surprisingly, the variants of the Holt's model with smoothing constants obtained by trials and errors method outperformed the variant with the "equivalent smoothing constant".

KEY WORDS: Holt's exponentially smoothed forecasting model; linear trend forecasting model; MAPE; MAD.

---

1. UVOD

Donošenje poslovnih odluka je rizično i otežano zbog neizvjesnosti budućih uvjeta poslovanja. Donositelji odluka stoga nastoje prognozirati buduće ekonomske veličine relevantne za njihovo poslovanje. Pritom bi konačni odabir prognostičkog modela u svrhu prognoziranja trebao biti temeljen na poredbenoj analizi statističkih pokazatelja kakvoće prognoza dobivenih pomoću više modela primjerenih prirodi empirijskih podataka.

U ovome se radu izlažu i primjenjuju dvije raznorodne statističke metode prognoziranja budućih vrijednosti jednog gospodarskog vremenskog niza s izraženim trendom. Cilj je bio analizirati, odnosno usporediti uspješnost prikladnog linearnog regresijskog modela trenda te modela dvostrukog eksponencijalnog izglađivanja u prognoziranju serije indeksa cijena industrijskih proizvoda pri proizvođačima u Hrvatskoj temeljem podataka za razdoblje od 1997. do 2006. godine, te zaključiti koji bi od velikog broja modela bio najuspješniji.

U ovome su radu statistički prognostički modeli primijenjeni na skupu sekundarnih podataka preuzetih su iz Statističkog ljetopisa Republike Hrvatske 2007. Podatci se odnose na vremenski niz indeksa cijena industrijskih proizvoda pri proizvođačima u Hrvatskoj, a analize su provedene pomoću softvera MS EXCEL.

Za teorijski dio rada korištena je relevantna svjetska i domaća statistička literatura koja obrađuje metode analize vremenskih nizova, a napose Aczel, Sounderpandian (2006), Gogala (2001), Hanke, Reitsch, (1989), Makridakis, Wheelwright, Hyndman (1998), McClave, Benson, Sincich (2008), Montgomery, Johnson (1976), Šošić (1991a, 1991b, 1991c).

2. PROGNOZIRANJE EKSTRAPOLACIJOM TRENDA

U svakidašnjici se susrećemo s nizom pojava koje nisu konstantne. Točnije, njihove se promjene očituju kroz neki interval. Da se opišu nastale promjene, koriste se različiti modeli koji olakšavaju njihovo objašnjenje. Vremenske serije mogu biti stacionarne, što znači da podatci variraju oko neke zamišljene konstante ili mogu pokazivati izraženije promjene tijekom vremena. Ukoliko te promjene pokazuju neku pravilnost, tj. sustavnu tendenciju razvoja u vremenu koju je moguće predočiti pomoću odgovarajuće matematičke funkcije, tada pojava izražava trend. Trend je funkcija razvoja pojave u vremenu, a najjednostavnija je linearna funkcija čija će se primjena u prognostičke svrhe pokazati u ovome radu.

Vremenska se serija uobičajeno prikazuje kao zbroj nekoliko komponenata:

$$Y = T + C + S + R, \tag{1}$$

pri čemu je:

- T oznaka za dugoročnu tendenciju razvoja pojave u vremenu, tj. *trend* komponentu
- C označuje *cikličku* komponentu kojom su izražena odstupanja od trenda koja se pripisuju općim poslovnim i ekonomskim uvjetima u kojima se pojava odvija,
- S izražava *sezonsku* komponentu kojom se opisuju fluktuacije vremenskog niza koje se ponavljaju u određenim razdobljima (npr. u zimskim su mjesecima aktivnosti u građevinarstvu na najnižoj razini)

- R je *rezidualna* komponenta koja predstavlja neobjašnjeni ostatak pojave, nakon što se glavnina objasni trend komponentom, te cikličkom i sezonskom komponentom. Rezidualna se komponenta može djelomično pripisivati nepredvidivim rijetkim događajima, poput potresa ili terorističkog napada, a djelomično i nepredvidivim akcijama ljudi.

Premda često nije moguće izdvojiti svaku od komponenata (to vrijedi u prvom redu za cikličku komponentu), ovakva formulacija modela, kako je to dano u izrazu (1), pomaže analitičarima da bolje shvate fluktuacije analizirane pojave.

Model (1) naziva se *aditivnim*, za razliku od *multiplikativnog* modela što ga čini umnožak spomenutih komponenata. Multiplikativni model je sljedeći:

$$Y = T \cdot C \cdot S \cdot R \qquad (2)$$

Općenito, trend se izražava kao funkcija vremena jednadžbom (3):

$$Y = f(X) + e \qquad (3)$$

Vremenska serija je dakle u nekom vremenu predstavljena zbrojem funkcijske vrijednosti $f(X)$ i vrijednosti slučajne varijable e.

Simbolom Y označena je zavisna varijabla kojom su predočene vrijednosti vremenske serije, a simbolom X označena je nezavisna varijabla vrijeme, čiji se utjecaj ispituje na razvoj serije Y. Sa e je označena varijabla nepoznatih, neobjašnjenih utjecaja na zavisnu varijablu Y. Model (3) se također naziva *aditivnim,* jer je varijabla e pribrojena matematičkom dijelu modela $f(X)$.

Ako se, primjerice, na temelju grafičkog prikaza pojave Y zaključi da se pojava linearno mijenja s vremenom, tada je prikladan model predočen jednadžbom (4):

$$f(X) = a + bX\,. \qquad (4)$$

Jednadžba linije trenda s procijenjenim parametrima, kao i kod modela linearne regresije, glasi:

$$\hat{Y} = a + bX\,. \qquad (5)$$

Parametri A *i* B ocjenjuju se pomoću procjenitelja a i b koji su dobiveni metodom najmanjih kvadrata. Pretpostavka koju model trenda mora zadovoljiti je svojstvo prosjeka, što znači da zbroj odstupanja opaženih vrijednosti vremenske serije od vrijednosti procijenjenih trend modelom mora biti jednak nuli, a zbroj kvadrata odstupanja mora biti minimalan.

Za svaku jedinicu vremena t jednadžba linearnog trenda s procijenjenim parametrima glasi

$$\hat{y}_t = a + bx_{t,} \qquad (6)$$

pri čemu se procjene parametara računaju pomoću izraza:

$$b = \frac{\sum_{t=11}^{n} x_t y_t - n\overline{x}\overline{y}}{\sum_{t=1}^{n} x_t^2 - n\overline{x}^2} \text{ i } a = \overline{y} - b\overline{x}\,. \qquad (7)$$

Trend vrijednosti su teorijske vrijednosti, tj. očekivane vrijednosti koje bi pojava trebala poprimiti na temelju jednadžbe s procijenjenim parametrima. Za dani niz vrijednosti nezavisne varijable vrijeme *X*, pripadne se trend vrijednosti računaju uvrštavanjem u jednadžbu s procijenjenim parametrima $\hat{Y}$:

$$\begin{aligned} \hat{y}_1 &= a + bx_1 \\ \hat{y}_2 &= a + bx_2 \\ &\;\;\vdots \\ \hat{y}_n &= a + bx_n \end{aligned} \tag{8}$$

Trend vrijednosti služe kao prognostičke vrijednosti pojave u nekom vremenu *t.* Razlike $y_t - \hat{y}_t = e_t$ pokazuju koliko se prognostičke vrijednosti razlikuju od stvarnih, opaženih vrijednosti. Općenito, $y_t - \hat{y}_t = e_t$ predstavlja vrijednosti neobjašnjenih odstupanja od teorijskih vrijednosti koje se nazivaju *rezidualnim*[1] *odstupanjima*. Za njih se pretpostavlja da su normalno distribuirana s očekivanom vrijednosti nula i konstantnom varijancom. Relativna rezidualna odstupanja dobivaju se dijeljenjem rezidualnih odstupanja pripadnom empirijskom vrijednosti zavisne varijable i množenjem kvocijenata sa 100, tj.

$$e_{t,rel} = \frac{e_t}{y_t} \cdot 100 \tag{9}$$

Budući da je pretpostavka modela trenda normalnost distribucije rezidualnih odstupanja, tj. pogrješaka relacije, one se zasebno analiziraju. Grafički prikaz rezidualnih odstupanja, a što je dijagram rasipanja rezidualnih odstupanja s obzirom na varijablu vrijeme, pokazuje je li varijanca pogrješaka konstantna. Temeljem navedenog grafičkog prikaza može se zaključiti o veličini varijance rezidualnih odstupanja. Ukoliko se širina dijagrama rasipanja s porastom vrijednosti varijable *X* ili $\hat{Y}$ povećava ili smanjuje, znači da nije zadovoljena pretpostavka regresijskog modela o konstantnoj varijanci pogrješaka relacije. Taj se problem naziva *heteroskedastičnost.* Ukoliko postoji problem heteroskedastičnosti, u svrhu procjene parametara regresijskog modela, umjesto metode najmanjih kvadrata, primjenjuje se *generalizirana metoda najmanjih kvadrata.*

Normalnost distribucije grješaka relacije također se i grafički provjerava pomoću tzv. *Normal probability plot*-a koji, ukoliko je pretpostavka normalnosti zadovoljena, mora pokazivati raspored točaka približno duž pravca.

U svrhu prognoziranja razine pojave za τ razdoblja, nakon *n*- tog, postupa se po istom obrascu kao i za trend vrijednosti, tj.

$$\hat{y}_{n+\tau} = a + bx_{n+\tau} \tag{10}$$

Pri korištenju trenda u prognostičke svrhe treba procijeniti hoće li se i nakon *n*-tog razdoblja pojava o kojoj se radi, odvijati u istim uvjetima kao i do tada, jer je samo u tom slučaju prognoziranje opravdano. Zbog toga nije preporučljivo prognozirati za dulji vremenski horizont.

[1] Od latinskog: residuum = ostatak.

3. MJERE USPJEŠNOSTI PROGNOZIRANJA

Prognoziranje je to uspješnije, što su razlike opaženih i prognoziranih vrijednosti pojave manje, tj. što su manje razlike prikazane jednadžbom (11)

$$y_t - \hat{y}_t = e_t \tag{11}$$

Da se ustanovi reprezentativnost odabranog trend modela, koriste se pokazatelji koji ukazuju na disperziju oko linije trenda. Kao mjere disperzije, koriste se varijanca trenda, iz nje izvedena standardna devijacija, kao i koeficijent varijacije trenda. Varijanca trenda je aritmetička sredina kvadriranih rezidualnih odstupanja, tj.:

$$\hat{\sigma}^2 = MSD = \frac{\sum_{t=1}^{n}(greške\ prognoze)^2}{n} = \frac{\sum_{t=1}^{n}(rezidua\ln ih\ odstupanja)^2}{n} = \frac{\sum_{t=1}^{n} e_t^2}{n}, \tag{12}$$

pri čemu je MSD (*od engl. Mean Square Error*) tzv. srednja kvadratna pogrješka.

Koeficijent varijacije predstavlja omjer standardne devijacije i aritmetičke sredine zavisne varijable pomnožen sa sto.

Standardna devijacija predstavlja prosječno odstupanje stvarnih vrijednosti zavisne varijable od trend vrijednosti u mjernim jedinicama varijable. Trend model je to reprezentativniji što su oba navedena pokazatelja bliža nuli.

Kao pokazatelj reprezentativnosti modela trenda još se koristi i koeficijent determinacije. On pokazuje proporciju modelom protumačenog dijela ukupnih odstupanja vrijednosti pojave od njene aritmetičke sredine. Model je reprezentativniji što je koeficijent determinacije bliži jedan.

MAD (od engl. *Mean Absolute Deviation)* ili tzv. srednje apsolutno odstupanje je mjera uspješnosti prognoze kojom se računa prosječna veličina prognostičke pogrješke, a dana je izrazom (13):

$$MAD = \frac{\sum_{t=1}^{n}|greške\ prognoze|}{n} = \frac{\sum_{t=1}^{n}|rezidua\ln ih\ odstupanja|}{n} = \frac{\sum_{t=1}^{n}|e_t|}{n}. \tag{13}$$

Pokazatelj MAPE (od engl. *Mean Absolute Percentage Error*), odnosno tzv. srednja apsolutna postotna pogrješka, dan je izrazom (14):

$$MAPE = \frac{\sum_{t=1}^{n}\frac{|st\,\mathrm{var}\,ne\ vrijednosti - prognozirane\ vrijednosti|}{st\,\mathrm{var}\,ne\ vrijednosti}}{n}\cdot 100, \text{ tj.}$$

$$MAPE = \frac{\sum_{t=1}^{n}\frac{|y_t - \hat{y}_t|}{y_t}}{n}\cdot 100 = \frac{\sum_{t=1}^{n}\frac{|e_t|}{y_t}}{n}\cdot 100\ . \tag{14}$$

Pokazatelj *MAPE* pokazuje za koliko posto prognozirane vrijednosti odstupaju od opaženih vrijednosti.

4. METODE EKSPONENCIJALNOG IZGLAĐIVANJA

4.1 Jednostavno eksponencijalno izglađivanje

Jednostavno eksponencijalno izglađivanje je korisna metoda za predviđanje budućih frekvencija vremenske serije kod koje nisu uočeni trend i sezonska komponenta. Kao što i samo ime sugerira, ova je metoda najjednostavnija među metodama eksponencijalnog izglađivanja. Eksponencijalno izglađivanje je proširenje ideje pomičnih prosjeka kojima se izglađuju originalne frekvencije vremenske serije. Izglađeni podatci se koriste za predviđanje budućih vrijednosti vremenske serije. Kod eksponencijalnog izglađivanja, međutim, vrijednosti serije koje su bliže tekućem vremenu, imaju veći utjecaj na prognostičke vrijednosti serije negoli opažanja koja su više udaljena u vremenu.

Eksponencijalno izglađivanje je metoda predviđanja kod koje se predviđanje temelji na vaganom prosjeku sadašnje i prošlih vrijednosti niza. Najveći se ponder pridaje sadašnjem opažanju, manji ponder prethodnom opažanju, još manji ponder opažanju prije prethodnog itd. Ponderi se smanjuju po geometrijskoj progresiji idući od sadašnjeg vremena unatrag. Ponder, označen simbolom α, zapravo je konstanta izglađivanja i ona poprima vrijednosti iz intervala između 0 i 1, tj.

$$0 < \alpha < 1 \ . \qquad (15)$$

Odabirom konstante izglađivanja α, definirana je jednadžba kojom se određuje F_{t+1}, eksponencijalno izglađena vrijednost razdoblja *t*, koja služi kao tražena prognostička vrijednost za razdoblje *t+1* :

$$F_{t+1} = \alpha F_t + \alpha(1-\alpha)F_{t-1} + \alpha(1-\alpha)^2 F_{t-2} + \alpha(1-\alpha)^3 F_{t-3} + ... \qquad (16)$$

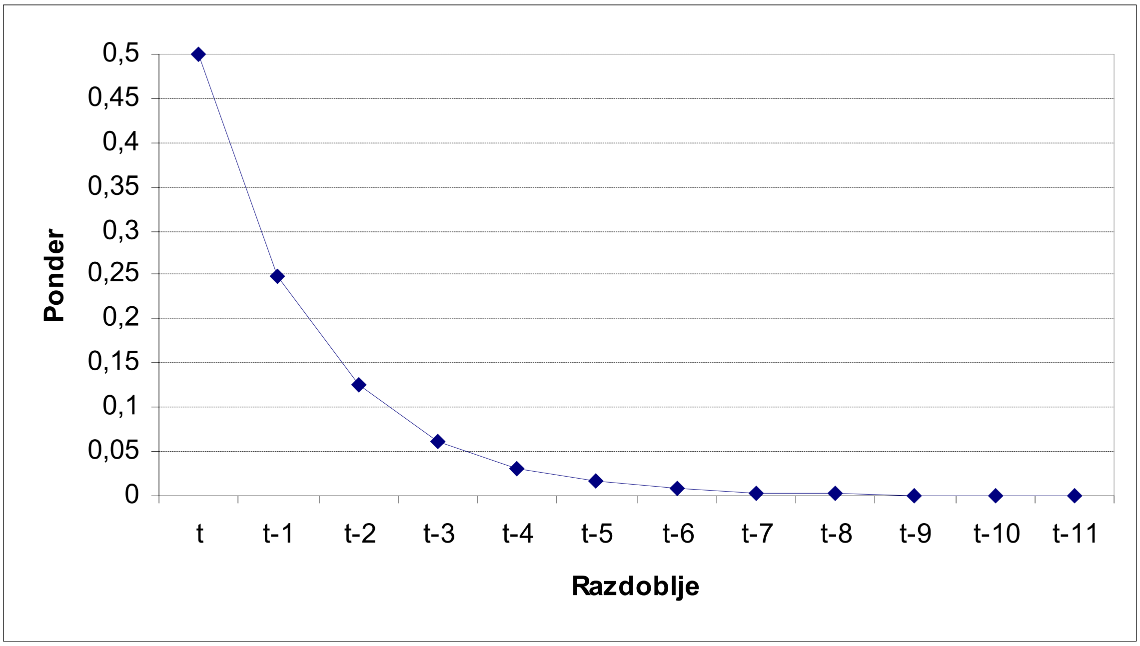
gdje su F_t, F_{t-1}, F_{t-2}, itd. stvarne vrijednosti niza počevši od vremena *t* unatrag.

Niz pondera pomoću kojih se računa prognostička vrijednost F_{t+1} čine:

$$\alpha, \alpha(1-\alpha), \alpha(1-\alpha)^2, \ldots \ . \qquad (17)$$

Ponderi iz izraza (17) eksponencijalno se smanjuju. Primjenom tih pondera, kretanjem unatrag u seriji, svaka njena vrijednost sudjeluje u predviđanju vrijednosti F_{t+1} sa sve manjim ponderom. Uz, primjerice, $\alpha = 0,5$, sljedeći su ponderi $\alpha(1-\alpha) = 0,25$, $\alpha(1-\alpha)^2 = 0,125$, $\alpha(1-\alpha)^3 = 0,0625$, $\alpha(1-\alpha)^4 = 0,03125$, $\alpha(1-\alpha)^5 = 0,015625$, $\alpha(1-\alpha)^6 = 0,007813$ itd. Eksponencijalno opadanje veličine pondera prema nuli je očito, što je prikazano i slikom 1.

Slika 1. Ponderi za konstantu izglađivanja α =0,5



Jednadžba kojom se određuje vrijednost F_{t+1}, može se pisati i u rekurzivnoj formi koju čine prethodna opažanja i prethodne prognostičke vrijednosti. Jednadžba (16) ekvivalentna je sljedećem modelu jednostavnog eksponencijalnog izglađivanja:

$$F_{t+1} = \alpha y_t + (1-\alpha)F_t \tag{17}$$

Prognostička vrijednost za razdoblje $t+1$ može se interpretirati kao vagani prosjek stvarne vrijednosti serije u razdoblju t i prognostičke vrijednosti za razdoblje t (koje je računano u razdoblju $t-1$).

Gornji je model (17) moguće pisati i u sljedećem obliku:

$$F_{t+1} = y_t + (1-\alpha)(F_t - y_t)\ . \tag{18}$$

Rekurzivni oblik jednadžbe pogodan je za računanje niza prognostičkih vrijednosti serije za svako razdoblje t . U jednadžbu se redom uvrštavaju vrijednosti t ($t=1,2,3,,..., n$) i tako dobivaju prognostičke vrijednosti za sljedeća razdoblja $t+1$. Problem jedino predstavlja prognoziranje prve vrijednosti F_1. Uobičajeno vrijedi da je $F_1 = y_1$, tj. prva opažena vrijednost serije služi ujedno kao prognostička vrijednost za prvo razdoblje.

Izbor konstante izglađivanja α ovisi o istraživaču i vrlo je važan. Konstanta izglađivanja α upravlja brojem prošlih realizacija vremenske serije koje utječu na prognostičku vrijednost serije. Što je α veći, to prognostički niz brže reagira na promjene u originalnoj seriji. Male vrijednosti konstante izglađivanja daju veće značenje prijašnjim opažanjima i rezultiraju sporim reagiranjem mehanizma prognoziranja na promjene parametara u modelu vremenske serije. Općenito, što je α manji, to su prognostičke vrijednosti manje osjetljive na promjene originalnih vrijednosti niza. O tome treba voditi računa pri odabiru konstante izglađivanja.

Veće vrijednosti konstante izglađivanja daju ponder samo najsvježijim povijesnim podatcima pa time čine da prognostički mehanizam brže reagira na promjene parametara. Međutim, veća vrijednost konstante izglađivanja može uzrokovati da prognostički mehanizam reagira na slučajne varijacije, a da se zapravo parametri modela vremenske serije nisu promijenili, što nije poželjno. Kao opće pravilo, prema Montgomery, Johnson (1976), vrijednost konstante izglađivanja trebala bi se kretati između 0,1 i 0,3. Uobičajena je praksa niz od nekoliko pokušaja s različitim konstantama izglađivanja, nakon čega se izabire vrijednost konstante α koja optimizira neku od mjera uspješnosti prognoze, kao npr. MAPE (*Mean Absolute Percentage Error*) ili MAD (*Mean Absolute Deviation*). Isti takav pristup može se koristiti i kod kompliciranijih modela.

Ukoliko rezultati niza uzastopnih pokušaja ukazuju na optimalnu vrijednost α veću od 0,3, tada bi trebalo preispitati valjanost modela. Podatci mogu ukazivati na značajnu autokorelaciju, u kom slučaju model eksponencijalnog izglađivanja nije valjan.

4.2. Dvostruko eksponencijalno izglađivanje

Metoda dvostrukog eksponencijalnog izglađivanja se koristi kad podatci ukazuju na postojanje trenda, što se najlakše uočava pomoću grafičkog prikaza serije.

Dvostruko je eksponencijalno izglađivanje slično jednostavnom, osim što se u svakom razdoblju moraju korigirati dvije komponente – razina pojave i trend. Razina pojave je izglađena procjena prosječne razine pojave na kraju svakog razdoblja, dok je trend izglađena procjena prosječnog rasta (pada) na kraju svakog razdoblja. Prema Holtovom modelu dvostruko se eksponencijalno izglađivanje vrši prema obrascu:

$$\begin{aligned} F_{t+1} &= \alpha y_t + (1-\alpha)(F_t + T_t) \qquad 0 \le \alpha \le 1 \\ T_{t+1} &= \gamma(F_{t+1} - F_t) + (1-\gamma)T_t \qquad 0 \le \gamma \le 1 \end{aligned} \tag{19}$$

pri čemu je:

F_{t+1} = eksponencijalno izglađena vrijednost razdoblja t

α = konstanta izglađivanja

y_t = stvarna vrijednost vremenske serije

F_t = prognostička vrijednost za razdoblje t

γ = konstanta izglađivanja za procjenu trenda

T_{t+1} = procjena trenda za razdoblje $t+1$

T_t = procjena trenda za razdoblje t.

Kao što je prethodno spomenuto, metoda dvostrukog eksponencijalnog izglađivanja primjenjuje se kada podatci vremenskog niza slijede trend, a koristi se i za prognoziranje razine pojave nakon posljednjeg promatranog razdoblja. Budući da je:

$$\begin{aligned} F_t &= \alpha y_{t-1} + (1-\alpha)(F_{t-1} + T_{t-1}) \\ T_t &= \gamma(F_t - F_{t-1}) + (1-\gamma)T_{t-1}, \end{aligned}, \tag{20}$$

to je prognoza za razdoblje nakon n-tog, te za τ razdoblja nakon njega kako slijedi:

$$\begin{aligned} F_{t+1} &= F_t + T_t \\ F_{t+\tau} &= F_t + \tau T_t \end{aligned}. \tag{21}$$

Naravno, pritom je potreban oprez, jer se u pravilu radi o prognoziranju na vrlo kratki rok, što znači da prognostički horizont τ ne smije biti velik. Kao i kod svih drugih metoda prognoziranja, uvjet za uspješnost prognoze je odvijanje promatrane pojave u razdoblju nakon *n*-tog u nepromijenjenim uvjetima kao i do tada.

Izbor druge konstante izglađivanja u modelu dvostrukog eksponencijalnog izglađivanja najčešće je rezultat metode pokušaja i pogrješaka. Postoji, međutim, i ideja "ekvivalentne konstante izglađivanja" prema Montgomery i Johnson (1997), koja se za slučaj dvostrukog eksponencijalnog izglađivanja računa prema obrascu:

$$\gamma = 1 - (1-\alpha)^{1/2} \; . \qquad (22)$$

Postoje različite metode odabira početnih vrijednosti za razinu pojave i za trend. Kao F_1 se uobičajeno uzima prva opažena vrijednost serije y_1, dok se za početnu trend komponentu može odabrati 0 (Hanke – Reitsch).U upotrebi su i varijante:

$$\begin{aligned} T_1 &= y_2 - y_1 \\ T_1 &= \left[(y_2 - y_1) + (y_3 - y_2) + (y_4 - y_3)\right]/3 \text{ i} \qquad (23) \\ T_1 &= (y_n - y_1)/n - 1 \; . \end{aligned}$$

5. USPJEŠNOST PROGNOZIRANJA SERIJE INDEKSA CIJENA INDUSTRIJSKIH PROIZVODA

Za primjenu izloženih metoda prognoziranja odabrana je vremenska serija indeksa cijena industrijskih proizvoda pri proizvođačima, prikazana tabelom 1.

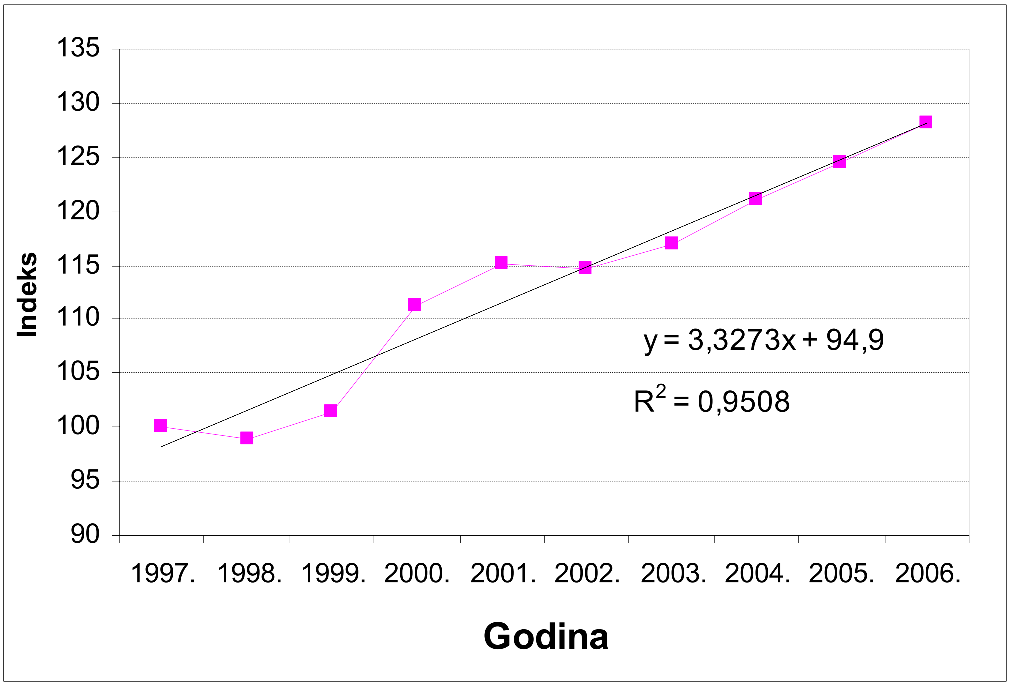
Tabela 1. Indeksi cijena industrijskih proizvoda pri proizvođačima

Godina	Indeksi 1997.=100
1997.	100
1998.	98,8
1999.	101,4
2000.	111,2
2001.	115,2
2002.	114,7
2003.	116,9
2004.	121
2005.	124,6
2006.	128,2

Izvor: Statistički ljetopis RH, str. 186.

Radi se o dinamičkoj vremenskoj seriji što je vidljivo i iz njenog grafičkog prikaza predočenog slikom 2.

Slika 2. Indeksi cijena industrijskih proizvoda, 1997.=100



Podatci o indeksima cijena analizirani su pomoću Microsoft Excela. Provedena regresijska analiza dala je sljedeće rezultate, vidi sliku 3.

Slika 3. Izlazni rezultati regresijske analize

ABSTRACT OUTPUT

Regression Statistics	
Multiple R	0,975099
R Square	0,950818
Adjusted R Square	0,94467
Standard Error	2,430114
Observations	10

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	913,3364	913,3364	154,6598	1,63239E-06
Residual	8	47,24364	5,905455		
Total	9	960,58			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	94,9	1,6601	57,1658	9,73E-12	91,072	98,728
Godina	3,327	0,2675	12,4362	1,63E-06	2,710	3,944

Kao što je vidljivo u grafičkom prikazu serije na slici 2 i na računalnom ispisu na slici 3, radi se o modelu linearnog trenda koji se može zapisati na sljedeći način, kako je to dano izrazom (24):

$$\begin{aligned} &\hat{Y} = 94{,}9 + 3{,}327X \\ &x = 1, 1997. \\ &\text{Jed. za x} = 1\,\text{god.} \end{aligned} \qquad (24)$$

Na temelju jednadžbe (24) vidljivo je da se u razdoblju od 1997. do 2006. godine indeks cijena industrijskih proizvoda pri proizvođačima prosječno godišnje povećavao za 3,327 indeksnih poena. Prema empirijskoj značajnosti reprezentativnost trenda je vrlo velika (p-vrijednost $=1{,}63 \cdot 10^{-6}$, vidi sliku 3).

Također, koeficijent determinacije R^2 =0,975, što znači da je 97,5% svih odstupanja protumačeno jednadžbom trenda.

Kako je postojanje trenda očito, to je za kratkoročno prognoziranje odabran Holtov model dvostrukog eksponencijalnog izglađivanja:

$$\begin{aligned} F_{t+1} &= \alpha y_t + (1-\alpha)F_t + T_t \qquad 0 \le \alpha \le 1 \\ T_{t+1} &= \gamma(F_{t+1} - F_t) + (1-\gamma)T_t \qquad 0 \le \gamma \le 1 \end{aligned} \qquad (25)$$

Izbor konstanti izglađivanja u primijenjenom modelu dvostrukog eksponencijalnog izglađivanja rezultat je metode pokušaja i pogrješaka. U jednom od prezentiranih oblika modela dvostrukog eksponencijalnog izglađivanja (model 2) druga konstanta izglađivanja γ, računana je prema formuli za "ekvivalentnu konstantu izglađivanja", prema Montgomery, Johnson (1997), tj. prema obrascu:

$$\gamma = 1 - (1-\alpha)^{1/2} = 1 - \sqrt{1 - 0{,}10} = 0{,}051 \approx 0{,}05 \qquad (26)$$

Za seriju indeksa cijena industrijskih proizvoda analizirano je nekoliko različitih oblika modela dvostrukog eksponencijalnog izglađivanja dobivenih variranjem konstanti izglađivanja. Uspješnost svake od prognoza ocijenjena je mjerama MAD i MAPE .

Tabela 2. Model 1: $\alpha= 0{,}3$ $\gamma = 0{,}7$

Godina	Indeks	F_t	T_t	Razlike	Apsolutne razlike	Razlika u %
1997.	100	100,00	0,00	0,00	0,00	0,00
1998.	98,8	100,00	0,00	-1,20	1,20	1,21
1999.	101,4	99,64	-0,25	1,76	1,76	1,74
2000.	111,2	99,99	0,17	11,21	11,21	10,08
2001.	115,2	103,47	2,49	11,73	11,73	10,18
2002.	114,7	108,73	4,43	5,97	5,97	5,20
2003.	116,9	113,62	4,75	3,28	3,28	2,80
2004.	121,0	117,93	4,44	3,07	3,07	2,54
2005.	124,6	121,96	4,15	2,64	2,64	2,12
2006.	128,2	125,66	3,84	2,54	2,54	1,98
2007.		**129,11**			**43,38**	**37,85**
MAD =	4,338348					
MAPE=	3,784736					
MAPE[2]=	4,205262					

Izvor: izračun autora.

[2] MAPE je računan 2 puta. U nazivniku MAPE je duljina vremenske serije *n*, tj. 10, dok je za računanje druge varijante, tj. MAPE[1] u nazivniku te mjere *n-1,*, tj. 9. Na taj se način pokušao ukloniti utjecaj prve, proizvoljno odabrane prognostičke vrijednosti na uspješnost prognoze.

Tabela 3. Model 2: α = 0,3 γ = 0,05

Godina	Indeks	F_t	T_t	Razlike	Apsolutne razlike	Razlika u %
1997.	100	100,00	0,00	0,00	0,00	0,00
1998.	98,8	100,00	0,00	-1,20	1,20	1,21
1999.	101,4	99,64	-0,02	1,76	1,76	1,74
2000.	111,2	100,16	0,01	11,04	11,04	9,93
2001.	115,2	103,47	0,17	11,73	11,73	10,18
2002.	114,7	107,11	0,35	7,59	7,59	6,61
2003.	116,9	109,63	0,46	7,27	7,27	6,22
2004.	121,0	112,13	0,56	8,87	8,87	7,33
2005.	124,6	115,18	0,68	9,42	9,42	7,56
2006.	128,2	118,49	0,81	9,71	9,71	7,58
2007.		**121,97**			**68,58**	**58,35**
MAD=	6,857966					
MAPE=	5,835297					
MAPE=	6,483663					

Izvor: izračun autora.

U cilju postizanja boljih prognoza, opisani je postupak još nekoliko puta ponovljen s različitim kombinacijama konstanti izglađivanja. Rezultati svih primijenjenih varijanti prognoziranja dani su u tabeli 6.

Da bi se ocijenila uspješnost primijenjenih prognostičkih modela usporedbe radi, izračunane su i pripadne trend vrijednosti indeksa, te rezidualna odstupanja, rezidualna odstupanja u apsolutnom iznosu, kao i relativna rezidualna odstupanja.

Tabela 4. Model trenda – prognostičke vrijednosti i grješke prognoza

Godina	Indeksi cijena 1997.=100	Trend vrijednosti $\hat{y}_t$	$y_t - \hat{y}_t$	Apsolutne razlike	Razlika u %
1997.	100	98,2	1,8	1,8	1,77
1998.	98,8	101,6	-2,8	2,8	2,79
1999.	101,4	104,9	-3,5	3,5	3,43
2000.	111,2	108,2	3,0	3,0	2,69
2001.	115,2	111,5	3,7	3,7	3,18
2002.	114,7	114,9	-0,2	0,2	0,14
2003.	116,9	118,2	-1,3	1,3	1,10
2004.	121,0	121,5	-0,5	0,5	0,43
2005.	124,6	124,8	-0,2	0,2	0,20
2006.	128,2	128,2	0,0	0,0	0,02
2007.		**131,5**	**0,0**	**16,91**	**15,76**
MAD =	1,69095				
MAPE =	1,575821				
MAPE =	1,750912				

Izvor: izračun autora

Nadalje, zbirni prikaz rezultata analize primjenom različitih prognostičkih modela dan je u tabeli 5. U toj je tabeli s *DEI* označeno dvostruko eksponencijalno izglađivanje.

Tabela 5. Usporedba pokazatelja uspješnosti prognostičkih modela

Metoda	MAD	MAPE	MAPE[1]
DEI, Model 1: $\alpha = 0,3$ $\gamma = 0,7$	4,338348	3,784736	4,205262
DEI, Model 2: $\alpha = 0,3$ $\gamma = 0,05$	6,857966	5,835297	6,483663
DEI, Model 3: $\alpha = 0,5$ $\gamma = 0,5$	3,704388	3,225655	3,584062
DEI, model 4: $\alpha = 0,95$ $\gamma = 0,5$	**3,202882**	**2,7832529**	3,0925032
DEI, model 5: $\alpha = 0,3$ $\gamma = 0,95$	3,807920	3,345071	3,716745
DEI, model 6: $\alpha = 0,5$ $\gamma = 0,95$	3,484564	3,032902	3,369892
DEI, model 7: $\alpha = 0,6$ $\gamma = 0,95$	3,432966	2,977350	3,308166
DEI, model 8: $\alpha = 0,7$ $\gamma = 0,95$	3,400396	2,948242	3,275824
DEI, model 9: $\alpha = 0,8$ $\gamma = 0,95$	3,368077	2,922541	3,247268
DEI, model 10: $\alpha = 0,9$ $\gamma = 0,95$	3,288352	2,856190	3,173544
Linearni trend	1,69095	1,575821	1,750912

Izvor: izračun autora

Usporedbom izračunanih mjera prognoziranja može se zaključiti da linearni trend daje najbolje prognoze, čemu u prilog govore izračunane mjere uspješnosti prognoze MAD i MAPE (1,69, odnosno 1,58). Međutim, graf rezidualnih odstupanja ukazuje na moguću pojavu heteroskedastičnosti, što znači da u ovom slučaju možda nije ispunjen zahtjev konstantne varijance.

Od ovdje analiziranih modela dvostrukog eksponencijalnog izglađivanja, najuspješniji je model 3 s konstantama $\alpha = \gamma = 0,5$. Za taj model MAD iznosi 3,70, a MAPE 3,23 %.

Rezimiraju li se rezultati provedene analize, zaključak je da u slučaju serije indeksa cijena model trenda daje najbolje prognostičke rezultate. To nije iznenađujuće, jer je već iz samog grafičkog prikaza serije indeksa (slika 3) vidljivo da se u ovom slučaju trend poklapa s opaženim vrijednostima, pogotovo s vrijednostima nekoliko posljednjih godina. Eksperimentiranja s različitim kombinacijama konstanti izglađivanja provedena su sa svrhom postizanja boljih prognostičkih rezultata.

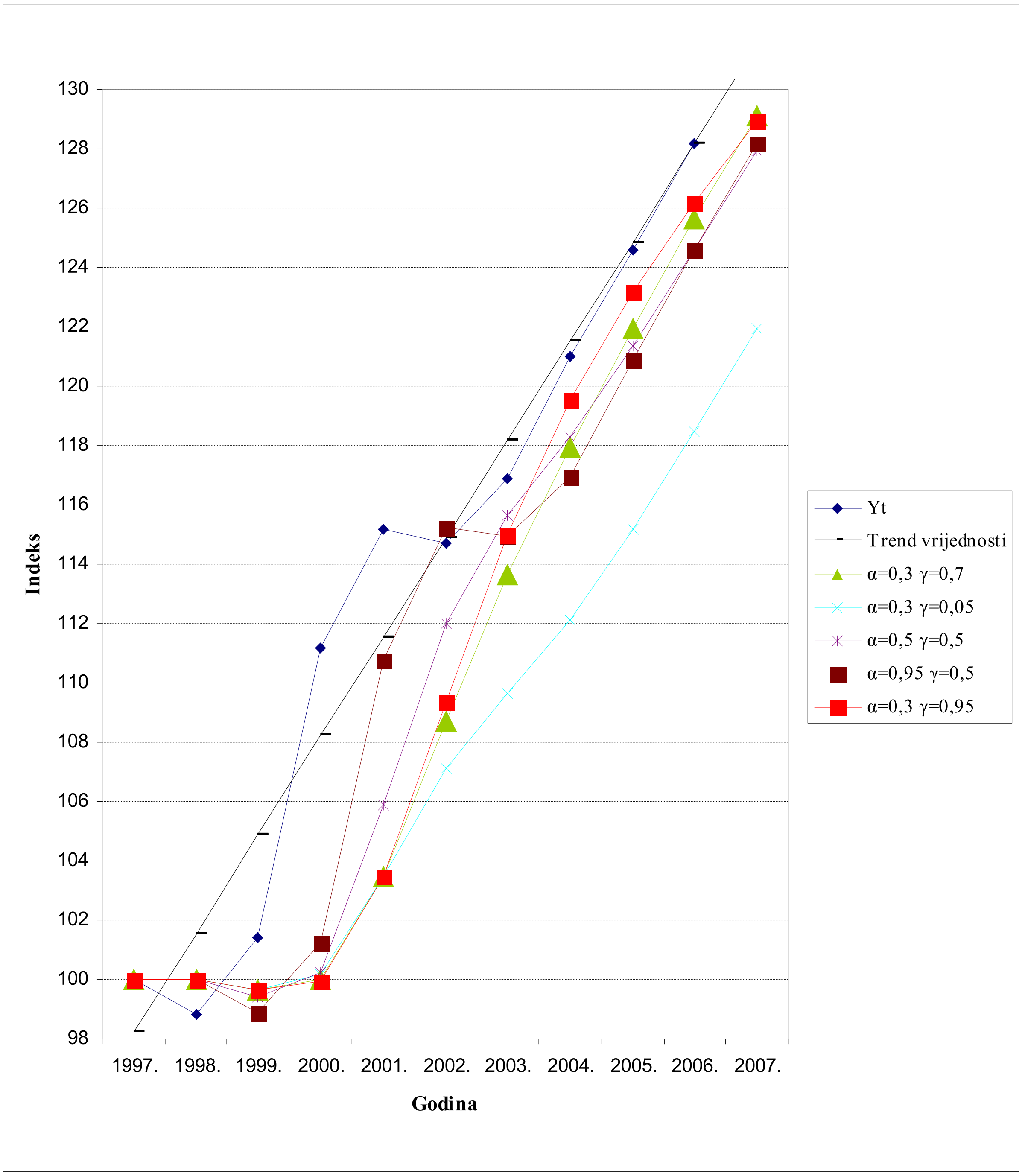
Usporedbom svih rezultata vidi se da je od analiziranih modela dvostrukog eksponencijalnog izglađivanja najbolji model 4, s konstantama izglađivanja $\alpha = 0,95$ $\gamma = 0,5$. Taj model ima MAD =3,202882, te MAPE 2,7832529. Po značajkama, tom je modelu najbliži model 10 s konstantama izglađivanja $\alpha = 0,9$ i $\gamma = 0,95$ uz pokazatelje MAD =3,288352 i MAPE =2,856190. To ipak nije ni približno rezultatima koje daje trend model, čije su prosječne prognostičke pogrješke upola manje.

Rezultati prognoziranja pomoću trend modela i prvih pet verzija modela eksponencijalnog izglađivanja prikazani su i grafički slikom 4.

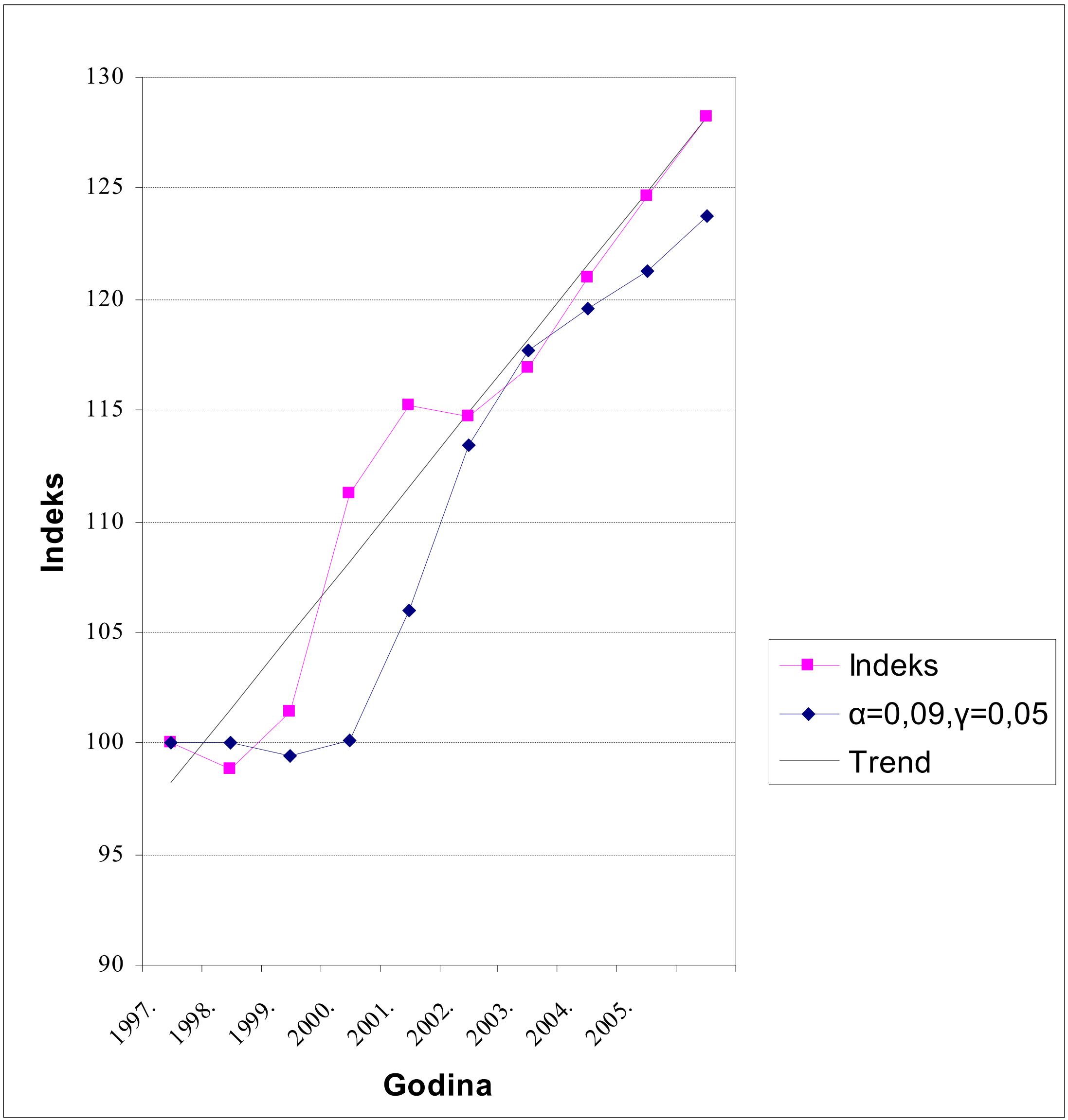
Vidljivo je da se izglađena serija s konstantama izglađivanja $\alpha = 0,95$ i $\gamma = 0,5$ najbolje prilagođuje originalnim vrijednostima indeksa cijena industrijskih proizvoda.

Općenito, verzije modela s relativno velikim konstantama izglađivanja dale su bolje rezultate. Najlošiji rezultati postignuti su modelom 2 uz konstante izglađivanja $\alpha = 0,3$ i $\gamma = 0,05$.

Slika 4. Eksponencijalno izglađeni nizovi za različite vrijednosti konstanti izglađivanja



Slika 5. Eksponencijalno izglađeni niz indeksa proizvođačkih cijena



Na slici 5 vidljivo je da i najbolji prognostički model temeljem eksponencijalnog izglađivanja (onaj sa konstantama α = 0,95 i γ = 0,05) zaostaje za kakvoćom prognoziranja linearnim trendom.

Radi upotpunjenja analize rezultata dobivenih pomoću trend modela, tj. provjere njegove adekvatnosti kao prognostičkog sredstva, proveden je i Durbin-Watsonov test za ispitivanje postojanja autokorelacije grješaka prvog reda, vidi tabelu 6.

Durbin-Watsonov pokazatelj *d* iznosi: $d = \dfrac{\sum_{t=2}^{n}(e_t - e_{t-1})^2}{\sum_{t=1}^{n} e_t^{\,2}} = 1{,}6941669$

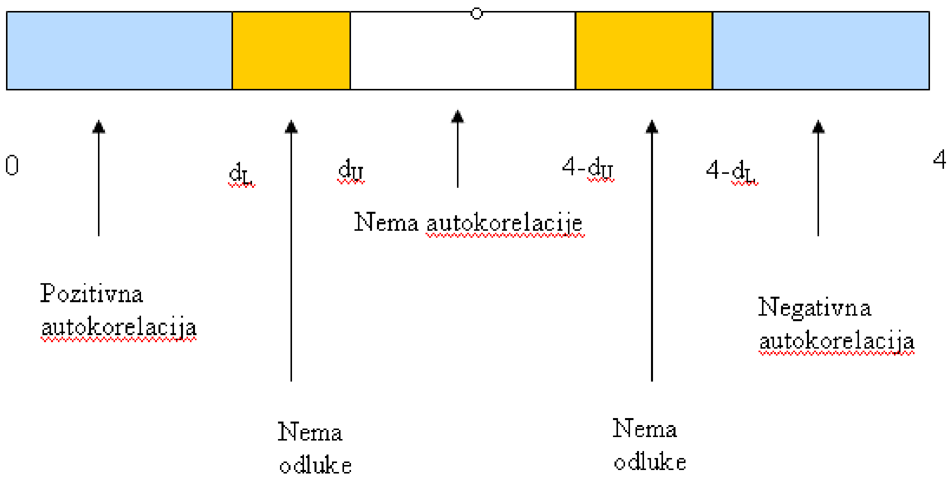
Tabela 6. Elementi za izračun Durbin-Watsonovog pokazatelja

Godina	X_t	Yt	$\hat{y}_t$	e_t	e_t^2	e_t-e_{t-1}	$(e_t$-$e_{t-1})^2$
1997.	1	100	98,2273	1,7727	3,1424653	-	-
1998.	2	98,8	101,5546	-2,7546	7,5878212	-4,5273	20,496445
1999.	3	101,4	104,8819	-3,4819	12,123628	-0,7273	0,5289653
2000.	4	111,2	108,2092	2,9908	8,9448846	6,4727	41,895845
2001.	5	115,2	111,5365	3,6635	13,421232	0,6727	0,4525253
2002.	6	114,7	114,8638	-0,1638	0,0268304	-3,8273	14,648225
2003.	7	116,9	118,1911	-1,2911	1,6669392	-1,1273	1,2708053
2004.	8	121	121,5184	-0,5184	0,2687386	0,7727	0,5970653
2005.	9	124,6	124,8457	-0,2457	0,0603685	0,2727	0,0743653
2006.	10	128,2	128,173	0,027	0,000729	0,2727	0,0743653
	11	1132	131,5003		47,243637		80,038608

Izvor: izračun autora

Kritične vrijednosti Durbin-Watsonovog testa za razinu signifikantnosti α = 0,05 i za *k= 1 (k* = broj nezavisnih varijabli u regresijskom modelu) i za duljinu serije n = 10 iznose: d_L= 0,879 i d_U =1,320. Buduući da (4- d_U) iznosi 2,68, a (4- d_L) iznosi 3,121, na razini značajnosti od 10% ne postoji problem autokorelacije prvog reda, budući da je d_U =1,320 < 1,694 < 4 - d_U = 2,6, vidi primjerice Makridakis, Wheelwright, Hyndman (1998) ili Aczel, Sounderpandian (2006).

Slika 6: Shema za donošenje odluke o postojanju autokorelacije za razinu signifikantnosti 2α[3]



Dobiveni rezultat DW testa ide u prilog korištenju trend modela u prognostičke svrhe, budući da ne postoji problem autokorelacije grješaka relacije.

[3] Vidi Aczel i Sounderpandian [1], str. 565.

7. ZAKLJUČAK

U ovome je radu za vremenski niz podataka o indeksima cijena industrijskih proizvoda pri proizvođačima za Republiku Hrvatsku, 1997.-2006., izvršeno modeliranje temeljem dva pristupa pomoću linearnog regresijskog modela i pomoću Holtovog modela dvostrukog eksponencijalnog izglađivanja. Oba su modela primijenjena u prognostičke svrhe uz vremenski horizont $\tau = 1$.

Holtov model dvostrukog eksponencijalnog izglađivanja primijenjen je s 10 različitih kombinacija konstanti izglađivanja α i γ, a provedena je i ekstrapolacija modelom linearnog trenda. Na kraju je dana poredbena analiza uspješnosti primijenjenih metoda prognoziranja. Prema provedenoj analizi, najbolje je rezultate dao prognostički model ekstrapolacijom linearnog trenda, a nakon njega, po uspješnosti, slijedi Holtov model s proizvoljno odabranim konstantama izglađivanja $\alpha = 0.95$ i $\gamma = 0.05$. Holtov model s ekvivalentnim konstantama izglađivanja dao je najlošije prognostičke rezultate.

Detaljnije, od svih prezentiranih 10 prognostičkih modela eksponencijalnog izglađivanja najlošije rezultate dao je onaj s konstantama izglađivanja $\alpha = 0,3$ $\gamma = 0,05$, a što je suprotno preporuci autora Montgomery i Johnson (1997) koji preferiraju vrijednost konstante do najviše $\alpha = 0,3$. Druga konstanta izglađivanja je izračunana prema obrascu istih autora za tzv. "ekvivalentnu konstantu izglađivanja", tj. primijenjen je obrazac (22): $\gamma = 1-(1-\alpha)^{1/2}$. Najbolji je rezultat postignut metodom pokušaja i pogrješaka te s tako izabranim konstantama izglađivanja $\alpha = 0,95$ i $\gamma = 0,05$, no i taj rezultat zaostaje za rezultatima postignutim pomoću linearnog trenda.

Budući da izbor tehnike prognoziranja ovisi o prognostičaru, to se zaključno može dati naputak da je korisno primijeniti nekoliko modela pri prognoziranju s različitim vrijednostima konstanti izglađivanja. Također, uputno je prognoziranje na kratki rok, jer je posebice u današnje vrijeme brzih promjena u makroekonomskom i mikroekonomskom okruženju rizik na dulji rok veoma velik.

Relativna prognostička inferiornost Holtovog modela pred modelom trenda začuđuje jer se u statističkoj literaturi taj model smatra prikladnim sredstvom za prognoziranje vremenske serije koja sadrži trend. Možda je razlog tome relativna kratkoća ovdje analizirane vremenske serije, kao i male oscilacije među njenim vrijednostima. Autori u tome vide poticaj za daljnja istraživanja u primjeni spomenutih modela na vremenske nizove veće duljine.

LITERATURA

1. Aczel, A. D., Sounderpandian, J.(2006), *Complete Business Statistics,* 6th Edt. New York: McGraw Hill.
2. Dumičić, K., Čeh Časni, A., Gogala, Z. (2008). *Evaluating Holt's Double Exponential Smoothing and Linear Trend Forecasting of Basic Tourism Time Series in Croatia.* 4th International Conference «An Enterprise Odyssey: Tourism – Governance and Entrepreneurship. University of Economics and Business – Zagreb, Cavtat, Croatia.
3. Gogala, Z. (2001). *Osnove statistike*. Zagreb: Sinergija.

4. Hanke, J. E., Reitsch, A. G.,(1989), *Business Forecasting,* 3rd Edt.. Needheim Heights: Allyn and Bacon.
5. Makridakis, S., Wheelwright, S. C., Hyndman, R.J. (1998), *Forecasting, Methods and Applications*, 3rd Edt.. New York: Wiley.
6. McClave, J. T., Benson, P. G., Sincich, T. (2008), *Statistics for Business and Economics*, 10th Edt.. Upper Saddle River, NJ.: Pearson Education, Prentice Hall.
7. Montgomery, D. C, Johnson, L. A. (1976), *Forecasting and Time Series Analysis.* New York: McGraw Hill.
8. *NIST/SEMATECH - e-HandbookofStatisticalMethods*, http://www.itl.nist.gov/div898/handbook/
9. Šošić, I. (1991a), Prognoze pojava s trendom (I), *Ekonomski analitičar,* veljača 1991, str.13-24.
10. Šošić, I. (1991b), Prognoze pojava s trendom (II),*Ekonomski analitičar,* ožujak, 1991, str. 32-39.
11 Šošić, I. (1991c), Prognoze pojava s trendom (III), *Ekonomski analitičar*, travanj, 1991, str 13-20.