UNIVERZITET U NOVOM SADU
PRIRODNO-MATEMATIČKI FAKULTET
DEPARTMAN ZA
MATEMATIKU I INFORMATIKU

Milana Tedić

Autokorelacija i autlajeri u predviđanju vrednosti kursa

- master rad -

Novi Sad, 2011.

SADRŽAJ

PREDGOVOR

Tema ovog master rada je predviđanje vrednosti deviznog kursa. Predviđanje neke vrednosti na osnovu istorijskih podataka je od izuzetne važnosti u raznim oblastima. U tu svrhu se koristi proces autoregresivnih integrisanih proseka (ARIMA), koji su u velikoj meri zastupljeni u ovom radu.

Podaci deviznih kurseva su podaci vremenskih serija. Vremenska serija je skup podataka o vrednostima neke promenljive za niz uzastopnih perioda. Proučavani su devizni kursevi dolar-evro i dolar- jen i za svaki od ovih kurseva izvršeno je ispitivanje autokorelacije u rezidualima, na osnovu čega su određeni parametri, a zatim ocenjeni koeficijenti ARIMA modela, pa je zatim izabranim ARIMA modelom izvršeno predviđanje deviznih kurseva pomoću programskog paketa Statgraphics.

Vremenske serije su često kontaminirane povremenim autlajerima. Autlajer je opažanje koje je udaljeno od ostalih opažanja. Ovo opažanje je prouzrokovano nekim neobičnim faktorima. U radu je prikazan algoritam za detekciju autlajera u deviznim kursevima u programskom paketu MATLAB. U odbacivanju autlajera treba biti pažljiv jer, u nekim slučajevima oni su posledice nelinearnosti, ili loše specificiranog modela. Na kraju su samo empirijski testirani i nelinearni modeli.

Najsrdačnije se zahvaljujem svom mentoru Dr Zorani Lužanin na usmeravanju i stručnoj pomoći tokom izrade ovog master rada.

Novi Sad, 23. 10.2011. Milana Tedić

1. Regresijski model

Najjednostavniji oblik stohastičkog odnosa između dve promenljive *X* i *Y* zove se prost linearni regresijski model. Taj se model formalno izražava u obliku

$$Y_i = \alpha + \beta X_i + \varepsilon_i \quad i = 1, \dots, n \tag{1.1}$$

u kojem se *Y* zove zavisna promenljiva, *X* nezavisna promenljiva i ε slučajno odstupanje u kojem su $\alpha\ i\ \beta$ nepoznati parametri regresije. Vrednosti promenljivih *X* i *Y* su opazive a vrednosti ε nisu.

Ukoliko postoji niz numeričkih podataka u kojima je svaki od njih povezan sa posebnim trenutkom u vremenu tada govorimo o vremenskoj seriji. To mogu biti podaci kao: mesečna nezaposlenost, nedeljne mere ponude novca itd, vrednost kursa, temperatura vazduha. Analiza jednodimenzionalnog niza podataka je nazvana *analizom univarijantnih vremenskih serija*. Analiza skupova podataka za isti niz vremenskih perioda je nazvana *analizom multivarijantnih vremenskih serija* (analiza na osnovu mesečnih podataka, veze između nezaposlenosti, nivoa cena, ponude novca itd). Svrha analize vremenskih serija je da prouči dinamiku ili vremensku strukturu podataka.

Stohastička priroda regresijskog modela podrazumeva da za svaku vrednost promenljive *X* postoji cela raspodela verovatnoća za vrednosti promenljive *Y*. To znači da se vrednost promenljive *Y* nikada ne može tačno predvideti. Neizvesnost se pojavljuje zbog prisutnosti slučajnog odstupanja ε.

Potpuna specifikacija regresijskog modela pored oblika regresijske jednačine date u (1.1) uključuje i specifikaciju raspodele verovatnoće odstupanja i određivanje vrednosti nezavisne promenljive. Tu informaciju daju osnovne pretpostavke:

Normalnost $\varepsilon_i, i = 1, \dots, n$ ima normalnu raspodelu. (1.2)

Sredina je jednaka nuli $E(\varepsilon_i) = 0, i = 1, \dots, n.$ (1.3)

Homoskedatičnost $Var(\varepsilon_i) = \sigma^2, i = 1, \dots, n.$ (1.4)

$$(1.2) - (1.4) \Leftrightarrow \varepsilon_i : N(0, \sigma^2)$$

Odsutnost autokorelacije $Cov(\varepsilon_i, \varepsilon_j) = 0 \ \ (i \neq j), i, j = 1, \dots, n.$ (1.5)

Nestohastičnost promenljive X X je nestohastička promenljiva sa (1.6)

fiksnim vrednostima u ponovljenim uzorcima i takva da je, za bilo koji uzorak veličine *n*,

$$\frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^2$$

različita od 0 i da je njena granična vrednost konačni broj kada $n \to \infty$.

To je takozvani klasični normalni linearni regresijski model koji određuje polazište za pretežni deo rada u ekonometrijskoj teoriji.

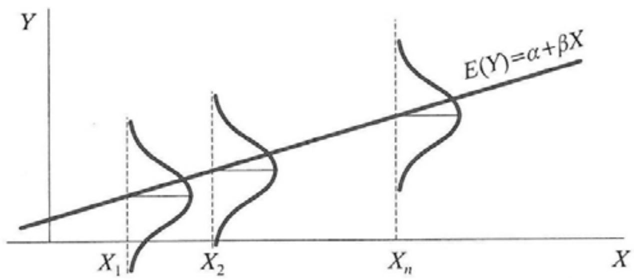
Pretpostavke na kojima se temelji klasični normalni linearni regresijski model koriste se pri izvođenju ocenjivača parametara regresije. Pošto se pretpostavlja da je odstupanje normalno raspodeljeno i da ima sredinu jednaku nuli, varijansa σ^2 je jedini nepoznat parametar te raspodele. Sledi da model opisan izrazima (1.1)-(1.6) ima tri nepoznata parametra α, β i σ^2.

Sada se može odrediti raspodela verovatnoće zavisne promenljive Y

$$E(Y_i) = E(\alpha + \beta X_i + \varepsilon_i) = \alpha + \beta X_i \quad (1.7)$$

$$Var(Y_i) = E[Y_i - E(Y_i)]^2 = E[(\alpha + \beta X_i + \varepsilon_i) - (\alpha + \beta X_i)]^2 = E(\varepsilon_i^2) = \sigma^2 \quad (1.8)$$

Iz jednačine (1.1) sledi da je Y_i slučajna promenljiva jer je ε_i slučajna promenljiva, a pošto je linearna transformacija normalne raspodele normalna raspodela sledi $Y_i \sim N(\alpha + \beta X_i, \sigma^2)$.



Grafik 1

Dalje, budući da je za $i \neq j$,

$$Cov(Y_i, Y_j) = E[Y_i - E(Y_i)][Y_j - E(Y_j)] = E(\varepsilon_i \varepsilon_j) = 0 \quad prema\ (1.3)\ i\ (1.5)$$

$Y_1, Y_2, \cdots, Y_n$ možemo posmatrati kao skup *n* normalno i nezavisno raspodeljenih promenljivih. Međutim te promenljive nisu identično raspodeljene jer imaju različite sredine.

Jednačina (1.7), koja daje očekivanu vrednost promenljive Y za svaku vrednost promenljive X je *regresijska linija populacije*. Odsečak te linije, α, meri srednju vrednost promenljive Y koja odgovara vrednosti nula promenljive X. Nagib linije, β, meri promenu srednje vrednosti promenljive Y koja odgovara jedinici promene vrednosti promenljive X. Pošto su vrednosti tih parametara nepoznate, nepoznata je regresijska linija populacije. Kada se ocene vrednosti

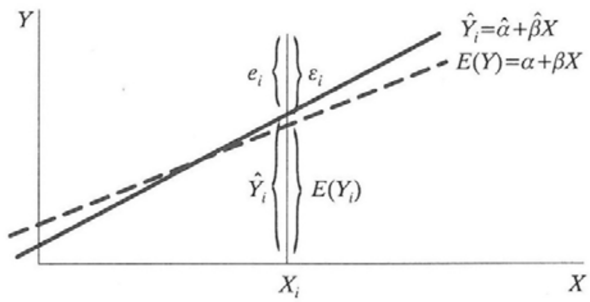
α i β, dobijamo *regresijsku liniju uzorka* koja služi kao ocena regresijske linije populacije. Ako se α i β ocene sa $\hat{\alpha}$ i $\hat{\beta}$, tada je regresijska linija uzorka

$$\hat{Y}_i = \hat{\alpha} + \hat{\beta}X_{i,} \tag{1.9}$$

gde je $\hat{Y}_i$ prilagođena vrednost promenljive Y_i. Većina, ako ne sve, opaženih vrednosti promenljive Y neće tačno ležati na regresijskoj liniji populacije ili uzorka, pa će se vrednosti Y_i i $\hat{Y}_i$ razlikovati. Ta se razlika zove ostatak (*rezidual*) i označava se sa e_i. Stoga moramo razlikovati sledeće:

$$Y_i = \alpha + \beta X_i + \varepsilon_i \quad \text{(populacija)}$$

$$Y_i = \hat{\alpha} + \hat{\beta}X_i + e_i \quad \text{(uzorak).}$$



Grafik 2

1.1. Višestruka regresija

$$Y_i = \beta_1 + \beta_2 X_{i2} + \beta_3 X_{i3} + \cdots + \beta_K X_{iK} + \varepsilon_i \tag{1.10}$$

gde Y označava zavisnu promenljivu, $X_{ik}, k = 2, \dots, K$označavaju nezavisne promenljive, a ε je slučajno odstupanje.

Osnovne pretpostavke

$\varepsilon_{i,} i = 1, \dots, n$ ima normalnu raspodelu (1.11)

$E(\varepsilon_i) = 0, i = 1, \dots, n$ (1.12)

$Var(\varepsilon_i) = \sigma^2, i = 1, \dots, n$ (1.13)

$Cov(\varepsilon_{i,}\varepsilon_j) = 0 \ \ (i \neq j), i, j = 1, \dots, n$ (1.14)

Sve nezavisne promenljive su nestohastičke. Svaka promenljiva je sa fiksnim (1.15) vrednostima u ponovljenim uzorcima i takva da je za bilo koji uzorak veličine n, $\sum_{i=1}^{n}(X_{ik} - \bar{X}_k)^2/n$ različita od nule. Granična vrednost tog izraza, kada $n \to \infty$ je konačni broj za svako $k = 2,3, \dots, K$.

Broj opažanja veći je od broja koeficijenata koje treba oceniti. (1.16)

Ne postoji egzaktna linearna veza između nezavisnih promenljivih (1.17)

Pretpostavke (1.11)-(1.17) čine klasični normalni linearni regresijski model za višestruku regresiju.

Sredina promenljive Y_i je

$$E(Y_i) = E(\beta_1 + \beta_2 X_{i2} + \beta_3 X_{i3} + \cdots + \beta_k X_{iK} + \varepsilon_i) = \beta_1 + \beta_2 X_{i2} + \beta_3 X_{i3} + \cdots + \beta_k X_{iK} \tag{1.18}$$

Varijansa promenljive Y_i je

$$Var(Y_i) = E[Y_i - E(Y_i)]^2 = E(\varepsilon_i^2) = \sigma^2 \tag{1.19}$$

Klasični normalni linearni regresijski model obično se predstavlja matričnim zapisom. Jednačina (1.10) se može napisati kao

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon} \tag{1.20}$$

$$\boldsymbol{y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}, \quad \boldsymbol{X} = \begin{bmatrix} X_{11} & X_{12} \cdots X_{1K} \\ X_{21} & X_{22} \cdots X_{2K} \\ & \vdots \\ X_{n1} & X_{n2} \cdots X_{nK} \end{bmatrix}, \quad \widehat{\boldsymbol{\beta}} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_K \end{bmatrix}, \quad \boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

1.2 . Ocenjivanje parametara regresije

Ocenjivanje se vrši pomoću metoda najmanjih kvadrata, metodom momenata, metodom maksimalne verodostojnosti i metodom najboljih linearnih nepristrasnih ocenjivača (BLUE). Ovde ćemo opisati metod najmanjih kvadrata.

1.2.1. Ocenjivanje metodom najmanjih kvadrata

Princip ocenjivanja metodom najmanjih kvadrata ima za posledicu minimalnu sumu kvadriranih odstupanja opaženih vrednosti od njihove sredine. Moramo minimizirati sumu S datu sa

$$S = \sum_{i=1}^{n} (Y_i - \beta_1 - \beta_2 X_{i2} - \beta_3 X_{i3} - \cdots - \beta_K X_{iK})^2.$$

Diferencirajući S s obzirom na $\beta_1, \beta_2, \ldots, \beta_K$ dobijamo:

$$\frac{\partial S}{\partial \beta_1} = -2 \sum_{i=1}^{n} (Y_i - \beta_1 - \beta_2 X_{i2} - \beta_3 X_{i3} - \cdots - \beta_K X_{iK}),$$

$$\frac{\partial S}{\partial \beta_2} = -2 \sum_{i=1}^{n} X_{i2}\,(Y_i - \beta_1 - \beta_2 X_{i2} - \beta_3 X_{i3} - \cdots - \beta_K X_{iK}),$$

$$\vdots$$

$$\frac{\partial S}{\partial \beta_K} = -2\sum_{i=1}^{n} X_{iK}\,(Y_i - \beta_1 - \beta_2 X_{i2} - \beta_3 X_{i3} - \cdots - \beta_K X_{iK}).$$

Izjednačavanjem sa nulom i sređivanjem dobijamo normalne jednačine:

$$\sum_{i=1}^{n} Y_i = \hat{\beta}_1 n + \hat{\beta}_2 \sum_{i=1}^{n} X_{i2} + \hat{\beta}_3 \sum_{i=1}^{n} X_{i3} + \cdots + \hat{\beta}_K \sum_{i=1}^{n} X_{iK},$$

$$\sum_{i=1}^{n} X_{i2}\, Y_i = \hat{\beta}_1 \sum_{i=1}^{n} X_{i2} + \hat{\beta}_2 \sum_{i=1}^{n} X_{i2}^2 + \hat{\beta}_3 \sum_{i=1}^{n} X_{i2}\, X_{i3} + \cdots + \hat{\beta}_K \sum_{i=1}^{n} X_{i2}\, X_{iK},$$

$$\vdots$$

$$\sum_{i=1}^{n} X_{iK}\, Y_i = \hat{\beta}_1 \sum_{i=1}^{n} X_{iK} + \hat{\beta}_2 \sum_{i=1}^{n} X_{i2} X_{iK} + \hat{\beta}_3 \sum_{i=1}^{n} X_{i3} X_{iK} + \cdots + \hat{\beta}_K \sum_{i=1}^{n} X_{iK}^2.$$

Prve dve jednačine se mogu napisati kao:

$$\hat{\beta}_1 = \bar{Y} - \hat{\beta}_2 \bar{X}_2 - \hat{\beta}_3 \bar{X}_3 - \cdots - \hat{\beta}_K \bar{X}_K,$$

gde je $\bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i$ i $\bar{X}_k = \frac{1}{n}\sum_{i=1}^{n} X_{ik} \quad (k = 2,3,\dots,K)$

Uvrštavanjem ovih izraza u preostale normalne jednačine daje

$$m_{Y2} = m_{22}\hat{\beta}_2 + m_{23}\hat{\beta}_3 + \cdots + m_{2K}\hat{\beta}_K,$$

$$m_{Y3} = m_{23}\hat{\beta}_2 + m_{33}\hat{\beta}_3 + \cdots + m_{3K}\hat{\beta}_K,$$

$$\vdots$$

$$m_{YK} = m_{2K}\hat{\beta}_2 + m_{3K}\hat{\beta}_3 + \cdots + m_{KK}\hat{\beta}_K,$$

gde je

$$m_{Yk} = \textstyle\sum_{i=1}^{n}(Y_i - \bar{Y})\,(X_{ik} - \bar{X}_k)$$

i

$$m_{jk} = \sum_{i=1}^{n}\left(X_{ij} - \bar{X}_j\right)(X_{ik} - \bar{X}_k) \qquad (j,k = 2,3,\dots,K).$$

Ove se jednačine mogu rešiti po $\hat{\beta}_2, \hat{\beta}_3,\dots,\hat{\beta}_K$. Za slučaj gde je $K = 3$, rešenje je sledeće:

$$\hat{\beta}_2 = \frac{\begin{vmatrix} m_{Y2} & m_{23} \\ m_{Y3} & m_{33} \end{vmatrix}}{\begin{vmatrix} m_{22} & m_{23} \\ m_{23} & m_{33} \end{vmatrix}} = \frac{m_{Y2}m_{33} - m_{Y3}m_{23}}{m_{22}m_{33} - m_{23}^2}$$

$$\hat{\beta}_3 = \frac{\begin{vmatrix} m_{22} & m_{Y2} \\ m_{23} & m_{Y3} \end{vmatrix}}{\begin{vmatrix} m_{22} & m_{23} \\ m_{23} & m_{33} \end{vmatrix}} = \frac{m_{Y3}m_{22} - m_{Y2}m_{23}}{m_{22}m_{33} - m_{23}^2}$$

Normalne jednačine metode najmanjih kvadrata mogu se izraziti u matričnom zapisu

$$(\boldsymbol{X}'\boldsymbol{y}) = (\boldsymbol{X}'\boldsymbol{X})\widehat{\boldsymbol{\beta}},$$

gde je

$$(\boldsymbol{X}'\boldsymbol{y}) = \begin{bmatrix} \sum_i Y_i \\ \sum_i X_{i2}Y_i \\ \vdots \\ \sum_i X_{iK}Y_i \end{bmatrix}, \ (\boldsymbol{X}'\boldsymbol{X}) = \begin{bmatrix} n & \sum_i X_{i2} \cdots \sum_i X_{iK} \\ \sum_i X_{i2} & \sum_i X_{i2}^2 \cdots \sum_i X_{i2}X_{iK} \\ & \vdots \\ \sum_i X_{iK} & \sum_i X_{i2}\, X_{iK} \cdots \sum_i X_{iK}^2 \end{bmatrix}, \ \widehat{\boldsymbol{\beta}} = \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \vdots \\ \hat{\beta}_K \end{bmatrix}$$

Rešenje za $\widehat{\boldsymbol{\beta}}$ tada postaje

$$\widehat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}(\boldsymbol{X}'\boldsymbol{y}) \tag{1.21}$$

Uvrštavanjem **y** iz izraza (1.20) u formulu za $\widehat{\boldsymbol{\beta}}$ u (1.21) dobija se

$$\widehat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\,(\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}) = \widehat{\boldsymbol{\beta}} + (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\varepsilon} \tag{1.22}$$

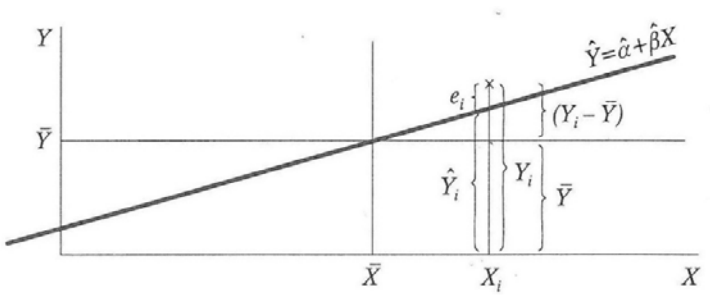
Tada je matrica varijansi i kovarijansi koeficijenata regresije dobijenih metodom najmanjih kvadrata sledećeg oblika

$$\begin{aligned} E\big(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}\big)\big(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}\big)' &= E(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1} \\ &= (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}')\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1} \\ &= \sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1} \qquad (1.23) \end{aligned}$$

Iz (1.22) i (1.23) sledi

$$\widehat{\boldsymbol{\beta}} \sim N[\boldsymbol{\beta}, \sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1}]$$

Razlaganje varijanse promenljive Y iz uzorka



Grafik 3

$$\sum_i (Y_i - \bar{Y})^2 = \sum_i \left(\hat{Y}_i - \bar{Y}\right)^2 + \sum_i \left(Y_i - \hat{Y}_i\right)^2$$

↑	↑	↑
SST	“ ovoliko “ je objasnio naš model	ostalo je neobjašnjeno, pripisujemo greški
↑	SSE	SSR
total	↑	↑
	regression explained	*error residual*

$SST = SSE + SSR$

$$1 = \frac{SSE}{SST} + \frac{SSR}{SST}$$

$$R^2 = \frac{SSE}{SST}$$

R^2 se naziva *koeficijent determinacije* i predstavlja meru uspešnosti prilagođavanja. To je udeo varijacija promenljive Y koje se mogu pripisati varijacijama promenljive X. Kao mera se obično upotrebaljava za opis koliko se regresiona linija uzorka dobro prilagođava registrovanim podacima.

$$0 \leq R^2 \leq 1$$

Za $R^2 = 0$ prilagođavanje je najslabije, a za $R^2 = 1$ prilagođavanje je najbolje.

2. Autokorelisana odstupanja

Na početku rada smo pomenuli sve standardne pretpostavke klasičnog normalnog linearnog regresijskog modela. Standardne pretpostavke za mnoge statističke postupke nisu ispunjene u realnoj situaciji. Regresijske jednačine ocenjene iz podataka vremenskih serija često karakterišu autokorelisana odstupanja, posebno kada je razdoblje opažanja kratko i u ovom delu rada ćemo pažnju posvetiti tom problemu.

Prema pretpostavci klasičnog normalnog linearnog modela imamo

$$Cov(\varepsilon_i, \varepsilon_j) = E[\varepsilon_i - E(\varepsilon_i)][\varepsilon_j - E(\varepsilon_j)] = 0 \quad (\forall\, i \neq j).$$

Budući da se pretpostavlja da su sredine od ε_i i ε_j jednake nuli, to znači da je

$$E(\varepsilon_i \varepsilon_j) = 0.$$

Kada se uzme u obzir pretpostavka normalnosti, pretpostavka da je kovarijansa između ε_i i ε_j jednaka nuli takođe znači da su ε_i i ε_j nezavisni. Ta je osobina regresijskih odstupanja poznata kao neautokorelisanost. To znači da odstupanje koje se javlja u jednoj tački opažanja nije korelisano ni sa jednim drugim odstupanjem. To takođe znači da se, kod opažanja tokom vremena, učinak odstupanja koje se javlja u jednom razdoblju ne prenosi u drugo razdoblje.

U slučaju autkorelisanih odstupanja, uz promenljive se stavlja indeks *t* (za „vreme“) umesto oznake *i* koja se upotrebljava u opštem slučaju.

Kada su odstupanja autokorelisana, imamo

$$E(\varepsilon_t \varepsilon_{t-s}) \neq 0 \quad (t > s).$$

Taj izraz znači da je odstupanje koje se dogodilo u vremenu *t* povezano sa odstupanjem u vremenu *(t - s)*. Posledice autokorelacije za ocenjivanje mogu se najbolje utvrditi ako preciznije specificiramo prirodu autokorelacije. Većina je radova u tom kontekstu izvedena na pretpostavci da regresijsko odstupanje sledi autoregresivni model prvog reda u oznaci AR(1) i koji ćemo u nastavku detaljnije predstaviti.

2.1. Autoregresivna odstupanja prvog reda

Kod autoregresivnih odstupanja prvog reda, odstupanja se generišu prema sledećoj šemi:

$$\varepsilon_t = \rho\varepsilon_{t-1} + u_t \quad (\forall t), \qquad (2.1)$$

gde je ρ parametar čija je apsolutna vrednost manja od jedan i u_t normalna nezavisno raspodeljena promenljiva sa sredinom nula i varijansom σ_u^2. Uz to, promenljiva u_t ne zavisi od promenljive ε_{t-1}.

Drugim rečima

$$u_t \sim N(0, \sigma_u^2) \quad (\forall t),$$

$$E(u_t, u_s) = 0 \quad (\forall t \neq s),$$

$$E(u_t \varepsilon_{t-1}) = 0 \quad (\forall t).$$

U analizi vremenskih serija, u_t je poznat kao *savršeno čisti šum.*

Iz jednačine (2.1) postupnim uvrštavanjem za $\varepsilon_{t-1}, \varepsilon_{t-2}, \dots, \varepsilon_1$ dobijamo

$$\begin{aligned}
\varepsilon_t &= \rho\varepsilon_{t-1} + u_t \\
&= \rho(\rho\varepsilon_{t-2} + u_{t-1}) + u_t \\
&= \rho^2\varepsilon_{t-2} + \rho u_{t-1} + u_t \\
&= \rho^2(\rho\varepsilon_{t-3} + u_{t-2}) + \rho u_{t-1} + u_t \\
&= \rho^3\varepsilon_{t-3} + \rho^2 u_{t-2} + \rho u_{t-1} + u_t \\
&\vdots \\
&= \rho^s\varepsilon_{t-s} + \rho^{s-1}u_{t-s+1} + \rho^{s-2}u_{t-s+2} + \cdots + \rho u_{t-1} + u_t.
\end{aligned}$$

Budući da $\rho^s \to 0$ kad $s \to \infty$, možemo pisati

$$\varepsilon_t = \sum_{s=0}^{\infty} \rho^s\, u_{t-s.} \qquad (2.2)$$

Budući da je $|\rho| < 1$, ρ^2 će imati manju apsolutnu vrednost od ρ, ρ^3 će biti manje od ρ^2 i tako dalje. To znači da se učinak lagovane promenljive *u* smanjuje što idemo više unazad i da će konačno potpuno iščeznuti.

$$\begin{aligned}
Var(\varepsilon_t) &= Var(u_t) + \rho^2 Var(u_{t-1}) + \rho^4 Var(u_{t-2}) + \cdots \qquad (2.3) \\
&= \sigma_u^2(1 + \rho^2 + \rho^4 + \cdots),
\end{aligned}$$

ili

$$\sigma^2 = \frac{\sigma_u^2}{1-\rho^2} \qquad (2.3a)$$

gde je $\sigma^2 = Var(\varepsilon_t)$. Nadalje,

$$\begin{aligned} Cov(\varepsilon_t, \varepsilon_{t-1}) &= E(u_t + \rho u_{t-1} + \rho^2 u_{t-2} + \cdots) \\ &\quad \times (u_{t-1} + \rho u_{t-2} + \rho^2 u_{t-3} + \cdots) \\ &= \rho\sigma_u^2 + \rho^3\sigma_u^2 + \rho^5\sigma_u^2 + \cdots \\ &= \frac{\rho\sigma_u^2}{1-\rho^2} \\ &= \rho\sigma^2 \qquad \text{prema (2.3a).} \end{aligned} \tag{2.4}$$

Slično,

$$Cov(\varepsilon_t, \varepsilon_{t-2}) = \rho^2\sigma^2,$$

$$Cov(\varepsilon_t, \varepsilon_{t-3}) = \rho^3\sigma^2,$$

i uopšteno

$$Cov(\varepsilon_t, \varepsilon_{t-s}) = \rho^s\sigma^2. \tag{2.5}$$

Iz (2.4) sledi

$$\rho = \frac{Cov(\varepsilon_t, \varepsilon_{t-1})}{\sigma^2}$$

pa pošto je $\sigma^2 = Var(\varepsilon_t) = Var(\varepsilon_{t-1})$ sledi

$$\rho = \frac{Cov(\varepsilon_t, \varepsilon_{t-1})}{\sqrt{Var(\varepsilon_t)}\sqrt{Var(\varepsilon_{t-1})}}$$

Izraz u kojem se kovarijansa dve promenljive deli proizvodom standardnih devijacija tih promenljivih poznat je kao *koeficijent korelacije* između dve promenljive.

Kada je ρ jednak nuli imamo

$$\varepsilon_t = u_t,$$

$$Var(\varepsilon_t) = \sigma_u^2.$$

Budući da je u_t normalno i nezavisno raspodeljena promenljiva sa sredinom nula i konstantnom varijansom, ispunjene su sve osnovne pretpostavke o promenljivoj ε.

Ako koeficijente korelacije između ε_t i ε_{t-1}, ε_t i ε_{t-2} i tako dalje posmatramo kao funkciju koja sadrži vremenski lag, ta funkcija se zove *autokorelaciona funkcija*. Grafik te funkcije se zove *korelogram*. U slučaju autoregresivnog procesa prvog reda, autokorelaciona funkcija je opadajuća ako je ρ pozitivan i ima prigušene oscilacije kada je ρ negativan. Procesi kao taj u kojima ni varijansa od ε_t ni autokorelacija između ε_t i ε_{t-s} ne zavise o t zovu se *stacionarni* procesi.

Sada uvodimo pojednostavljenje. Prikaz promenljive ε_t u (2.1) uključuje pretpostavku da je proces generisanja promenljive ε_t započeo u $t = -\infty$. Pošto uzorak započinje sa prvim opažanjem u $t = 1$, prikladan način da se uzme u obzir učinak odstupanja pre prvog opažanja je da se prvo odstupanje u uzorku specificira kao

$$\varepsilon_1 = \frac{u_1}{\sqrt{1-\rho^2}}, \qquad (2.6)$$

i ostavi da se preostala odstupanja uzorka generišu u skladu sa (2.1). Važi da je

$$E(\varepsilon_1) = 0$$

$$Var(\varepsilon_1) = \frac{\sigma_u^2}{1-\rho^2} = \sigma^2$$

$$Cov(\varepsilon_1, \varepsilon_2) = E[\varepsilon_1(\rho\varepsilon_1 + u_2)] = \rho\sigma^2$$

$$Cov(\varepsilon_1, \varepsilon_3) = E[\varepsilon_1(\rho^2\varepsilon_1 + \rho u_2 + u_1)] = \rho^2\sigma^2$$

i tako dalje. Stoga su sve prethodno specificirane karakteristike promenljivih $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n$ sačuvane.

2.2. Posledice autokorelacije

Sada ispitujemo svojstva ocenjivača od α i β dobijenih metodom najmanjih kvadrata u modelu jednostavne regresije $Y_t = \alpha + \beta X_t + \varepsilon_t$ kada je odstupanje ε_t autoregresivno.

Teorema 1:

1. Ocenjivači metode najmanjih kvadrata (OLS) još uvek su linearni i nepristrasni.
2. Oni nisu efikasni u poređenju sa procedurama koje uzimaju u obzir autokorelaciju. Kraće, ocenjivači metode najmanjih kvadrata nisu najbolji linearni nepristrasni ocenjivači (BLUE)
3. Ocenjene varijanse OLS ocenjivača su pristrasne. Ponekad uobičajene formule za izračunavanje varijansi i standardnih grešaka OLS ocenjivača daju niže vrednosti od tačnih varijansi i standardnih grešaka.

4. Uobičajeni *t* i *F* testovi nisu uopšteno pouzdani.

5. Uobičajena formula za izračunavanje ocene varijanse, naime $\hat{\sigma}^2 = SSR/d.f.$ (rezidualna suma kvadrata/ stepeni slobode), je pristrasan ocenjivač parametra σ^2.

6. Kao posledica, konvencionalno izračunat R^2 može biti nepouzdana mera pravog R^2

7. Konvencionalno izračunate varijanse i standardne greške predviđanja mogu takođe biti neefikasne.

<u>*Dokaz:*</u>

Nepristrasnost ocenjivača metode najmanjih kvadrata

Ocenjivač od β dobijen metodom najmanjih kvadrata je

$$\hat{\beta} = \beta + \frac{\sum (X_t - \bar{X})\varepsilon_t}{\sum (X_t - \bar{X})^2}$$
$$E(\hat{\beta}) = \beta + \frac{\sum (X_t - \bar{X})E(\varepsilon_t)}{\sum (X_t - \bar{X})^2} = \beta$$

Ocenjivač od α dobijen metodom najmanjih kvadrata je

$$\hat{\alpha} = \bar{Y} - \hat{\beta}\bar{X} = (\alpha + \beta\bar{X} + \bar{\varepsilon}) - \hat{\beta}\bar{X}$$
$$E(\hat{\alpha}) = \alpha + \beta\bar{X} + E(\bar{\varepsilon}) - E(\hat{\beta})\bar{X} = \alpha$$

Ocenjivači metode najmanjih kvadrata nisu BLUE

Ovu osobinu ćemo dokazati izvodeći formule BLUE za autoregresivni slučaj i upoređujući ih s formulama metode najmanjih kvadrata. Najjednostavniji način izvođenja BLUE regresijskih koeficijenata je transformisanje početne regresijske jednačine sa autoregresivnim odstupanjima u ekvivalentnu jednačinu u kojoj su odstupanja nezavisna. Početna jednačina $Y_t = \alpha + \beta X_t + \varepsilon_t$ se pomnoži sa ρ i laguje jedan period da bi se dobilo

$$\rho Y_{t-1} = \alpha\rho + \beta\rho X_{t-1} + \rho\varepsilon_{t-1}.$$

Oduzimanjem druge jednačine od prve dobijamo

$$Y_t - \rho Y_{t-1} = \alpha(1-\rho) + \beta(X_t - \rho X_{t-1}) + \varepsilon_t - \rho\varepsilon_{t-1}.$$

Pošto važi da je $\varepsilon_t = \rho\varepsilon_{t-1} + u_t$ možemo pisati

$$Y_t - \rho Y_{t-1} = \alpha(1-\rho) + \beta(X_t - \rho X_{t-1}) + u_t \qquad (t = 2,3,\dots,n) \qquad (2.7)$$

Transformacija kojom autoregresivno odstupanje zamenjujemo klasičnim odstupanjem u_t se zove *Cochrane- Orcuttova* transformacija. Jedini njen nedostatak je da se u procesu transformacije gubi jedno opažanje koje je povezano sa u_1. Da bi se u_1 uključilo, početnu jednačinu $Y_1 = \alpha + \beta X_1 + \varepsilon_1$ u skladu sa (2.6), posmatramo kao

$$Y_1 = \alpha + \beta X_1 + \frac{u_1}{\sqrt{1-\rho^2}} \,/\cdot \sqrt{1-\rho^2}$$

Sledi da je

$$Y_1\sqrt{1-\rho^2} = \alpha\sqrt{1-\rho^2} + \beta X_1\sqrt{1-\rho^2} + u_{1.}$$

Početna se jednačina tada može izraziti kao

$$Y_t^* = \alpha W^* + \beta X_t^* + u_t \qquad (t = 1,2,\dots,n) \qquad (2.8)$$

gde je za $t = 1$,

$$Y_t^* = Y_t\sqrt{1-\rho^2}, W^* = \sqrt{1-\rho^2}, X_t^* = X_t\sqrt{1-\rho^2}$$

i za $t = 2,3,\dots,n$

$$Y_t^* = Y_t - \rho Y_{t-1}, W^* = 1-\rho, X_t^* = X_t - \rho X_{t-1}.$$

Transformacija (2.8) poznata je kao *Prais- Winstenova* transformacija.

Pošto su objašnjavajuće promenljive u (2.8) nestohastičke i u_t zadovoljava sve osnovne pretpostavke, zadovoljeni su svi preduslovi za jednakost ocenjivača dobijenih metodom najmanjih kvadrata (OLS) i najboljih linearnih nepristrasnih ocenjivača (BLUE) od α i β. Primenjujući metodu najmanjih kvadrata na (2.8) dobijamo sledeće normalne jednačine metode najmanjih kvadrata:

$$\sum W^* Y_t^* = \tilde{\alpha}\sum W^{*2} + \tilde{\beta}\sum W^* X_t^*,$$

$$\sum X_t^* Y_t^* = \tilde{\alpha}\sum W^* X_t^* + \tilde{\beta}\sum X_t^{*2} \qquad (2.9)$$

gde je

$$\sum_{t=1}^{n} W^* Y_t^* = (1-\rho^2)Y_1 + (1-\rho)\sum_{t=2}^{n}(Y_t - \rho Y_{t-1})$$

$$\sum W^{*2} = (1-\rho^2) + (n-1)(1-\rho^2)$$

$$\sum_{t=1}^{n} W^* X_t^* = (1-\rho^2)X_1 + (1-\rho)\sum_{t=2}^{n}(X_t - \rho X_{t-1})$$

$$\sum_{t=1}^{n} X_t^* Y_t^* = (1-\rho^2)X_1 Y_1 + \sum_{t=2}^{n}(X_t - \rho X_{t-1})(Y_t - \rho Y_{t-1})$$

$$\sum_{t=1}^{n} X_t^{*2} = (1-\rho^2){X_1}^2 + \sum_{t=2}^{n}(X_t - \rho X_{t-1})^2.$$

Formule za $\tilde{\alpha}$ i $\tilde{\beta}$ će sadržati parametar ρ . Pošto ocenjivači dobijeni metodom najmanjih kvadrata ne sadrže ρ, oni nisu BLUE kad je odstupanje autoregresivno. Iz ovoga sledi da ocenjivači dobijeni metodom najmanjih kvadrata nemaju najmanju varijansu od svih nepristrasnih ocenjivača i da u skladu sa tim nisu efikasni.

Pristrasnost ocenjenih varijansi ocenjvača metode najmanjih kvadrata

Formula metode najmanjih kvadrata za ocenjivanje varijanse od $\hat{\beta}$ je

$$s_{\hat{\beta}}^2 = \frac{\hat{\sigma}^2}{\sum(X_t - \bar{X})^2}$$

gde je $\hat{\sigma}^2$ ocenjivač od σ^2 definisan kao suma kvadrata reziduala metode najmanjih kvadrata podeljena sa $(n-2)$. Pošto je $\sum(X_t - \bar{X})^2$ nestohastička, ostaje samo $\hat{\sigma}^2$. Stoga imamo

$$\hat{\sigma}^2 = \frac{1}{n-2}\sum_t \left((Y_t - \bar{Y}) - \hat{\beta}(X_t - \bar{X})\right)^2$$

$$= \frac{1}{n-2}\sum_t \left(\beta(X_t - \bar{X}) + (\varepsilon_t - \bar{\varepsilon}) - \hat{\beta}(X_t - \bar{X})\right)^2$$

$$= \frac{1}{n-2}\sum_t \left[-(\hat{\beta} - \beta)(X_t - \bar{X}) + (\varepsilon_t - \bar{\varepsilon})\right]^2$$

$$= \frac{1}{n-2}\left[(\hat{\beta} - \beta)^2 \sum_t (X_t - \bar{X})^2 + \sum_t (\varepsilon_t - \bar{\varepsilon})^2 - 2(\hat{\beta} - \beta)\sum_t (X_t - \bar{X})(\varepsilon_t - \bar{\varepsilon})\right]$$

$$= \frac{1}{n-2}\left[\sum_t (\varepsilon_t - \bar{\varepsilon})^2 - \left(\hat{\beta} - \beta\right)^2 \sum_t (X_t - \bar{X})^2\right]$$

i

$$E(\hat{\sigma}^2) = \frac{1}{n-2}\left[E\left(\sum_t (\varepsilon_t - \bar{\varepsilon})^2\right) - \left(\sum_t (X_t - \bar{X})^2\right) Var\left(\hat{\beta}\right)\right]$$

Dalje je

$$E\left(\sum_t (\varepsilon_t - \bar{\varepsilon})^2\right) = E\left(\left(\sum_t \varepsilon_t^2\right) - n\bar{\varepsilon}^2\right)$$

$$= n\sigma^2 - nE(\bar{\varepsilon}^2)$$

Asimptotski je $nE(\bar{\varepsilon}^2) = \sigma_u^2/(1-\rho)^2 = \sigma^2(1+\rho)/(1-\rho)$. Stoga važi da kada je n veliko

$$E(\hat{\sigma}^2) \cong \frac{1}{n}\left[n\sigma^2 - \frac{\sigma^2(1+\rho)}{(1-\rho)} - \left(\sum_t (X_t - \bar{X})^2\right) Var\left(\hat{\beta}\right)\right]$$

$$\cong \sigma^2 - \frac{\left(\sum_t (X_t - \bar{X})^2\right) Var\left(\hat{\beta}\right)}{n} - \frac{\sigma^2(1+\rho)}{n(1-\rho)}$$

$$\cong \sigma^2 - \left(\frac{1}{n}\sum_t (X_t - \bar{X})^2\right) Var\left(\hat{\beta}\right)$$

Iz ovoga sledi da je

$$E\left(s_{\hat{\beta}}^2\right) = E\left(\frac{\hat{\sigma}^2}{\sum (X_t - \bar{X})^2}\right)$$

$$\cong \frac{\sigma^2}{\sum (X_t - \bar{X})^2}$$

a varijansa od $\hat{\beta}$ je

$$Var\left(\hat{\beta}\right) = E\left[\frac{\sum (X_t - \bar{X})\varepsilon_t}{\sum (X_t - \bar{X})^2}\right]^2$$

$$= \frac{1}{\left(\sum (X_t - \bar{X})^2\right)^2} E\left[\sum_t (X_t - \bar{X})^2 \varepsilon_t^2 + 2\sum_{s<t} (X_t - \bar{X})\varepsilon_t (X_{t-s} - \bar{X})\varepsilon_{t-s}\right]$$

$$= \frac{\sigma^2}{(\sum(X_t - \bar{X})^2)^2} \left[\sum_t (X_t - \bar{X})^2 + 2 \sum_{s<t} (X_t - \bar{X})(X_{t-s} - \bar{X}) \rho^s \right]$$

$$= \frac{\sigma^2}{\sum(X_t - \bar{X})^2} + \frac{2\sigma^2}{(\sum(X_t - \bar{X})^2)^2} \left[\rho \sum_{t=2}^{n} (X_t - \bar{X})(X_{t-1} - \bar{X}) \right.$$
$$\left. + \rho^2 \sum_{t=3}^{n} (X_t - \bar{X})(X_{t-2} - \bar{X}) + \cdots \right]$$

Pošto se izrazi za $E\left(s_{\hat{\beta}}^2\right)$ i $Var(\hat{\beta})$ razlikuju konvencionalno izračunati ocenjivač varijanse od $\hat{\beta}$ je pristrasan kada su odstupanja autoregresivna.

2.3 . Testovi za odsutnost autokorelacije

Ako ne znamo, ili ne želimo da pretpostavimo da li je regresijsko odstupanje autokorelisano, moramo se orijentisati na informaciju iz uzorka. U okviru autoregresije prvoga reda možemo testirati hipotezu o neautoregresiji

$$H_0: \rho = 0,$$

nasuprot jednostranoj ili dvostranoj alternativi. Uobičajena alternativna hipoteza u ekonomskim vezama je ona da postoji pozitivna autoregresija, tj.

$$H_1: \rho > 0.$$

Neki od uobičajenih testova za autokorelaciju su:

1) **Durbin-Watson *d* Test**

Durbin-Watson statistika je test statistika koja se koristi za otkrivanje prisustva autokorelacije u rezidualima iz regresione analize. Durbin i Watson (1950, 1951) su primenili ovu statistiku na reziduale najmanjih kvadrata i razvili kritične vrednosti testova za nultu hipotezu da greške nisu autokorelisane protiv alternative da one slede autoregresivni proces prvog reda.

Da bismo primenili test izračunavamo vrednost test veličine d određene sa

$$d = \frac{\sum_{t=2}^{n} (e_t - e_{t-1})^2}{\sum_{t=1}^{n} {e_t}^2}$$

gde su e reziduali obične metode najmanjih kvadrata. Ako je alternativna hipoteza da postoji pozitivna autoregresija, važe pravila

1. Odbaciti H_0 ako je $d < d_L$.
2. Ne odbaciti H_0 ako je $d > d_U$.
3. Test je neodređen ako je $d_L \leq d \leq d_U$.

Vrednosti d_L (za „ donju granicu “) i d_U (za „ gornju granicu “) navedene su u tablici koju su konstruisali Durbin i Watson. Ako je alternativna hipoteza dvostrana, tj.

$(H_1: \rho \neq 0)$ pravila su sledeća

1. Odbaciti H_0 ako je $d < d_L$ ili ako je $d > 4 - d_L$.
2. Ne odbaciti H_0 ako je $d_U < d < 4 - d_U$.
3. Test je neodređen ako je $d_L \leq d \leq d_U$ ili ako je $4 - d_U \leq d \leq 4 - d_L$.

Budući da je

$$d = \frac{\sum_{t=2}^{n} e_t^2}{\sum_{t=1}^{n} e_t^2} + \frac{\sum_{t=2}^{n} e_{t-1}^2}{\sum_{t=1}^{n} e_t^2} - 2\frac{\sum_{t=2}^{n} e_t\, e_{t-1}}{\sum_{t=1}^{n} e_t^2}$$

Sledi

$$d \approx 2(1 - \hat{\rho})$$

odnosno

$$\hat{\rho} = \frac{\sum_{t=2}^{n} e_t e_{t-1}}{\sum_{t=1}^{n} {e_t}^2}$$

Vrednosti ρ	Vrednosti d
$\rho = -1$ (perfektna negativna korelacija)	$d = 4$
$\rho = 0$ (nema autokorelacije)	$d = 2$
$\rho = 1$ (perfektna pozitivna korelacija)	$d = 0$

Tabela 1

$$0 \leq d \leq 4$$

$$\hat{\rho} \approx 1 - \frac{d}{2}$$

Durbin-Watsonov test nije primenjiv na regresijske jednačine u kojima se umesto nezavisne promenljive pojavljuje lagovana vrednost zavisne promenljive ili u kojoj nema konstantnog člana.

Ukoliko se lagovane zavisne promenljive nalaze među nezavisnim promenljivama i greške su autokorelisane tada nezavisne promenljive više nisu nekorelisane sa greškama.

Razmatramo model

$$Y_t = \alpha Y_{t-1} + \beta X_t + \varepsilon_t \qquad (2.10)$$

$$\varepsilon_t = \rho\varepsilon_{t-1} + u_t \qquad (2.11)$$

u_t su nezavisne sa sredinom 0 i varijansom σ^2 i $|\alpha| < 1, |\rho| < 1$. Zato što ε_t zavisi od ε_{t-1} i Y_{t-1} zavisi od ε_{t-1}, dve promenljive Y_{t-1} i ε_t biće korelisane. Ocenjivač najmanjih kvadrata $\hat{\alpha}$ biće nekonzistentan. Važi da je

$$plim\ \hat{\alpha} = \alpha + A$$

i

$plim\ \hat{\rho} = \rho - A,$

gde $plim$ predstavlja konvergenciju u verovatnoći.

Definicija: Niz $\{Z_n\}$ slučajnih promenljivih konvergira u verovatnoći ka Z ako za svako $\varepsilon > 0,\ lim_{n\to\infty} P\{Z_n - Z \geq \varepsilon\} = 0$, što se označava sa $p\ lim_{n\to\infty} Z_n = Z$.

gde je

$$A = \frac{\rho\sigma_x^2\ \sigma_e^2}{(1-\alpha\rho)D}$$

$$D = Var(Y_{-1})Var(X) - Cov^2(Y_{-1}, X) > 0$$

$$\sigma_x^2 = Var(X)\ \ \sigma_\varepsilon^2\ = Var(\varepsilon)$$

Stoga ako je ρ pozitivno, ocena parametra α je pristrasna nagore i ocena od ρ je pristrasna nadole. Stoga Durbin-Watson statistika, koja je $\sim 2(1-\hat{\rho})$ je pristrasna ka 2 i neće biti otkrivena svaka značajna autokorelacija čak i ako su greške autokorelisane.

2) Testiranje odsutnosti autoregresije u modelu sa lagovanom zavisnom promenljivom

h test

Ovaj test može biti opisan u sklopu sledećeg regresionog modela:

$$Y_t = \alpha + \beta X_t + \gamma Y_{t-1} + \varepsilon_t \quad (t = 2,3,\cdots,n)$$

$$\varepsilon_t = \rho\varepsilon_{t-1} + u_{t.}$$

Testira se hipoteza da je $\rho = 0$ nasuprot jednostranoj ili dvostranoj alternativi. Test statistika na kojoj se temelji h test je

$$h = \left(1 - \frac{d}{2}\right)\sqrt{\frac{n}{1 - ns_{\hat{\gamma}}^2}}$$

gde je d uobičajena Durbin-Watsonova test statistika i $s_{\hat{\gamma}}^2$ metodom kvadrata ocenjena varijansa od γ. Prema nultoj hipotezi h ima $N(0,1)$ raspodelu. Ovaj se test ne može upotrebiti kada je $ns_{\hat{\gamma}}^2 > 1$ jer se dobije negativna vrednost pod korenom.

m test

$m\ test$ se sastoji od izračunavanja reziduala metode najmanjih kvadrata i primene metode najmanjih kvadrata na

$$e_t = \beta_1 + \beta_2 X_t + \beta_3 Y_{t-1} + \beta_4 e_{t-1} + greška \tag{2.12}$$

i testiranje signifikantnosti ocenjenog koeficijenta uz e_{t-1} pomoću standardnog *t*- testa

$$\frac{\hat{\beta}_k - \beta_k}{s_{\hat{\beta}_k}} : t_{n-K} \qquad (k = 1,2,\dots,K)$$

gde je $s_{\hat{\beta}_k}$ ocenjena standardna greška od $\hat{\beta}_k$. Za dvostrani test sa nivoom značajnosti λ i sa $(n-k)$ stepeni slobode oblast prihvatanja se definiše kao

$$-t_{n-K,\lambda/2} \leq \frac{\hat{\beta}_k - \beta_k}{s_{\hat{\beta}_k}} \leq t_{n-K,\lambda/2}.$$

Ovi testovi će biti detaljnije objašnjeni na primeru.

Primer 1: Detektovanje autokorelacije

Posmatrali smo mesečne podatke za devizne kurseve dolar- evro i dolar- jen u periodu od Januara 2000 do Decembra 2010.

Slučaj sa jednom lagovanom promenljivom

Kurs dolar- evro

$$\hat{Y}_t = 0{,}00828907 + 0{,}988311 * Y_{t-1} + e_t$$

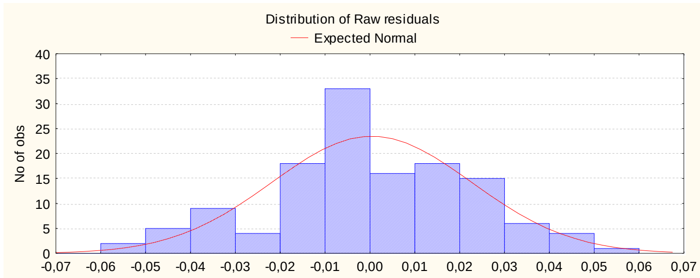
(0,011210) (0,012838)

Durbin-Watson *d* Test

$$d = \frac{\sum_{t=2}^{n}(e_t - e_{t-1})^2}{\sum_{t=1}^{n} {e_t}^2}$$

U našem slučaju $d = \frac{0{,}087930}{0{,}066857} = 1{,}315199$

$$\hat{\rho} = 0{,}342401$$



Slika 1: Histogram reziduala za kurs dolar-evro

$$h = \left(1 - \frac{1{,}315199}{2}\right)\sqrt{\frac{131}{1 - 131 * \ 0{,}012838^2}}$$

$h = 7{,}60916$, a pošto je kritična vrednost standardne normalne raspodele na 5 % 1,645, nulta hipoteza se odbacuje na osnovu ovog testa.

m test je sledeći e_t = 0,007456 - 0,008904 Y_{t-1} +0,342499 $e_{t-1} + greška$

(0,010680) (0,012246) (0,084112)

Pošto je $0{,}342499/0{,}084112 = 4{,}07194$ i t_{128} za 5 % je 1,645, nulta hipoteza o neautoregresiji se odbacuje.

Kurs dolar-jen

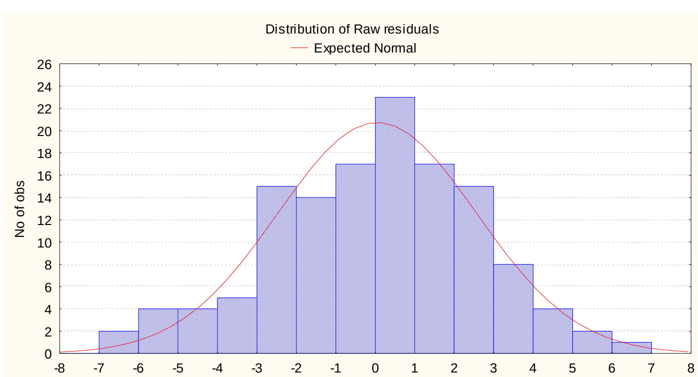
$$\hat{Y}_t = 0{,}397417 + 0{,}99486 * Y_{t-1} + e_t$$

(2,195125) (0,019853)

Durbin- Watson *d* test

U našem slučaju $d = \frac{1439{,}058}{861{,}4031} = 1{,}670598$

$$\hat{\rho} = 0{,}164701$$



Slika 2: Histogram reziduala kursa dolar-jen

$$h = \left(1 - \frac{1{,}670598}{2}\right)\sqrt{\frac{131}{1 - 131 * \ 0{,}019853^2}}$$

h = 1,935724, a pošto je kritična vrednost standardne normalne raspodele na 5 % 1,645, nulta hipoteza se odbacuje na temelju ovog testa.

m test je:

e_t = 0,762680 – 0,007220 Y_{t-1} + 0,161433 e_{t-1} + $greška$

(2,223038) (0,020103) (0,089123)

Pošto je $0{,}161433/0{,}089123 = 1{,}811351$ i kritična vrednost t_{131} za 5 % je 1,645, nulta hipoteza se odbacuje i prema ovom testu.

Za slučaj sa jednom lagovanom promenljivom zaključujemo da postoji autokorelacija prvog reda u deviznim kursevima dolar- evro i dolar- jen.

Slučaj sa dve lagovane promenljive

Kurs dolar- evro

$$\hat{Y}_t = 0{,}0129063 + 1{,}32191 * Y_{t-1} - 0{,}338495 * Y_{t-2} + e_t$$

(0,010597) (0,083137) (0,083129)

$$d = \frac{0{,}109389}{0{,}058204} = 1{,}879402$$

$$\hat{\rho} = 0{,}060299$$

$h = 2{,}156448$, a pošto je kritična vrednost standardne normalne raspodele za pozitivnu alternativu na nivou od 5 % 1,645 nulta hipoteza se odbacuje.

Kurs dolar- jen

$$\hat{Y}_t = 1{,}09594 + 1{,}14907 * Y_{t-1} - 0{,}160603 * Y_{t-2} + e_t$$

(2,215161) (0,086716) (0,088665)

m test :

$$e_t = 0{,}033893 + 0{,}579770 * Y_{t-1} - 0{,}578785 * Y_{t-2} - 0{,}578390 * e_{t-1}$$

(2,214181) (0,552517) (0,551033) (0,559453)

Pošto je $0{,}578390/0{,}559453 = 1{,}033849$, a kritična vrednost za t_{130} je 1,645 pa nema razloga da odbacimo hipotezu H_0.

Slučaj sa tri lagovane promenljive

Kurs dolar- evro

$$\hat{e}_t = 0{,}002394 - 0{,}127343 * Y_{t-1} + 0{,}155203 * Y_{t-2} - 0{,}031121 * Y_{t-3}$$

(0,012286) (0,482991) (0,639564) (0,1796889)

$$+0{,}125294 * e_{t-1} + greška$$

(0,490585)

Pošto je $0{,}125294/0{,}490585 = 0{,}255397$, a kritična vrednost za t_{129} je 1,645 pa nema razloga da odbacimo hipotezu H_0.

Kurs dolar- jen

$$\hat{Y}_t = 1{,}763716 + 1{,}150452 * Y_{t-1} - 0{,}074775 * Y_{t-2} - 0{,}092892 * Y_{t-3} + e_t$$

(2,273175) (0,088538) (0,133537) (0,089580)

m test

$$\hat{e}_t = 0{,}969492 - 0{,}740264 * Y_{t-1} + 0{,}867686 * Y_{t-2} - 0{,}137118 * Y_{t-3} + 0{,}742512 * e_{t-1}$$

(2,586846) (0,972869) (1,136733) (0,194885) (0,977077)

Pošto je $0{,}742512/0{,}977077 = 0{,}759932$, a kritična vrednost za t_{129} je 1,645 pa nema razloga da odbacimo hipotezu H_0.

Za slučaj sa tri lagovane promenljive nema dokaza da postoji autokorelacija za kurs dolar- evro odnosno dolar- jen.

2.4 . Odstupanja pokretnih proseka prvog reda

U nekim je okolnostima regresijsko odstupanje prikladnije modelirati procesom pokretnih proseka nego autoregresivnim procesom. Najjednostavniji oblik procesa pokretnih proseka je MA (1). On je predstavljen kao

$$\varepsilon_t = u_t + \theta u_{t-1} \qquad \text{(za svako t)},$$

gde u_t zadovoljava sve klasične pretpostavke. Važi da je:

$$E(\varepsilon_t) = 0$$

$$Var(\varepsilon_t) = \frac{\sigma_u^2}{1 + \theta^2}$$

$$Cov(\varepsilon_t, \varepsilon_{t-s}) = \theta \quad (za\ s = 1)$$

$$Cov(\varepsilon_t, \varepsilon_{t-s}) = 0 \quad (za\ ostale\ vrednosti)$$

Pošto je

$$u_t = \varepsilon_t - \theta u_{t-1,}$$

uzastopnom zamenom za $u_{t-1,}u_{t-2}, \cdots$, dobijamo

$$u_t = \sum_{s=0}^{\infty} (-\theta)^s\, \varepsilon_{t-s}^s \tag{2.19}$$

Pošto je koeficijent korelacije između ε_t i ε_{t-1}

$$\rho = \frac{\theta}{1+\theta^2}$$

parametar prosečnog pokreta θ može se izraziti kao

$$\theta = \frac{1 \pm \sqrt{1-4\rho^2}}{2\rho}$$

Stoga se zahteva da je $(1-4\rho^2) \geq 0$, ili, ekvivalentno, da je $|\rho| \leq \frac{1}{2}$.

Ocenjivanje regresijskih koeficijenata modela sa odstupanjima MA (1) zasniva se na izrazu (2.19) izraženom u obliku

$$u_t = \varepsilon_t - \theta\varepsilon_{t-1} + \cdots + (-\theta)^{t-1}\varepsilon_1 + (-\theta)^t u_0 \qquad (\ 2.19\ a)$$

Taj se rezultat upotrebljava za transformisanje regresijske jednačine tako da se pokretno prosečno odstupanje ε_t zameni klasičnim odstupanjem u_t. Pristup poznat kao *približno maksimalno verodostojno ocenjivanje* uključuje pretpostavku da je u_0 jednako nuli. To dovodi do sledeće transformisane jednačine:

$$Y_t^* = \alpha W_t^* + \beta X_t^* + u_t, \qquad (t = 1,2,\cdots,n)$$

gde su

$$Y_t^* = Y_t - \theta Y_{t-1} + \cdots + (-\theta)^{t-1} Y_1,$$

$$W_t^* = 1 - \theta + \cdots + (-\theta)^{t-1},$$

$$X_t^* = X_t - \theta X_{t-1} + \cdots + (-\theta)^{t-1} X_1,$$

ili, ekvivalentno,

$$Y_t^* = Y_t - \theta Y_{t-1}^* \qquad (Y_0^* = 0),$$

$$W_t^* = W_t - \theta W_{t-1}^* \quad (W_0^* = 0),$$

$$X_t^* = X_t - \theta X_{t-1}^* \qquad (X_0^* = 0).$$

Maksimiziranje funkcije verodostojnosti uz pretpostavku da je $u_0 = 0$ ekvivalentno je minimiziranju

$$S = \sum (Y_t^* - \alpha W_t^* - \beta X_t^*)^2$$

s obzirom na α, β i ρ.

2.5. Mešoviti proces autoregresivnih odstupanja i pokretnih proseka

Formulacija procesa generisanja odstupanja tokom vremena koja omogućava i autoregresivnu i šemu sa pokretnim prosecima poznata je *mešoviti proces autoregresivnih odstupanja i pokretnih proseka*. Kad je autoregresija reda p i pokretni prosek reda q, proces se kraće označava sa ARMA (p, q). On je definisan kao

$$\varepsilon_t = \rho_1 \varepsilon_{t-1} + \rho_2 \varepsilon_{t-2} + \cdots + \rho_p \varepsilon_{t-p} + u_t + \theta_1 u_{t-1} + \theta_2 u_{t-2} + \cdots + \theta_q u_{t-q.}$$

Tipično je da su u praksi p i q nespecificirani i da se određuju, zajedno sa ρ i θ, primenom *Box- Jenkinsove metode* na reziduale metode najmanjih kvadrata.

Ove modele ćemo detaljnije izučavati u sledećem delu rada.

3. Analiza vremenskih serija

Ranije je uvedena definicija vremenske serije, a sada ćemo se baviti njenim osobinama.

Sa teoretske tačke gledišta vremenska serija je familija slučajnih promenljivih $\{Y_t, t \in T\}$, koja se naziva stohastički proces.

Stroga stacionarnost vremenskih serija

1. Očekivanje $\mu(t) = E(Y_t)$
2. Varijansa $\sigma^2(t) = Var(Y_t)$
3. Autokovarijanse $\gamma(t_1, t_2) = Cov(Y_{t1}, Y_{t2})$.

Vremenska serija je *strogo stacionarna* (stacionarna u užem smislu) ako je zajednička raspodela svakog skupa od *n* opažanja $Y(t_1), Y(t_2), \cdots Y(t_n)$ ista kao zajednička raspodela od $Y(t_1 + k), Y(t_2 + k), \cdots Y(t_n + k)$ za svako *n* i *k*.

Prethodna definicija važi za sve vrednosti *n*. Zamenjujući $n = 1$, dobijamo $\mu(t) = \mu$ i $\sigma^2(t) = \sigma^2$ za svako *t*. Ako zamenimo $n = 2$, dobijamo da je zajednička raspodela od $Y(t_1)$ i $Y(t_2)$ ista kao zajednička raspodela od $Y(t_1 + k)$ i $Y(t_2 + k)$. Ako napišemo $t_1 + k = t_2$, vidimo da je ovo isto kao raspodela od $Y(t_2)$ i $Y(t_2 + k)$. Stoga, ovo zavisi samo od razlike $(t_2 - t_1)$ koja je nazvana *lag*. Iz ovoga sledi da autokovarijansnu funkciju $\gamma(t_1, t_2)$ možemo napisati kao $\gamma(k)$, gde je $k = t_2 - t_1$ lag. Stoga $\gamma(k) = Cov[Y(t), Y(t + k)]$ je autokovarijansni koeficijent na lagu *k*. $\gamma(k)$ se zove *autokovarijansna funkcija*.

Pošto je $Var\, Y(t) = Var\, Y(t + k) = \sigma^2 = \gamma(0)$, definisaćemo autokorelacioni koeficijent $\rho(k)$ na lagu *k* kao

$$\rho(k) = \frac{\gamma(k)}{\gamma(0)}$$

$\rho(k)$ se zove autokorelaciona funkcija.

Vremenska serija je *slabo stacionarna* (stacionarna u širem smislu) ako je njeno očekivanje konstantno i autokorelaciona funkcija zavisi samo od laga, tj.

$$E[Y(t)] = \mu \ \text{ i } Cov[Y(t), Y(t + k)] = \gamma(k)$$

Osobine autokorelacione funkcije

1. Autokorelaciona funkcija je parna funkcija na lagu *k*, npr. $\rho(k) = \rho(-k)$. Ova osobina sledi iz rezultata
   $\gamma(k) = Cov(Y_t, Y_{t+k}) = Cov(Y_{t-k}, Y_t)$ koji je zbog stacionarnosti jednak sa $\gamma(-k)$

2. Za datu autokorelacionu funkciju postojaće samo jedan normalni proces.

Modeli koji se koriste u modelovanju vremenskih serija mogu biti različitog oblika i predstavljaju različite stohastičke procese. Ti modeli su: (1) beli šum, (2) slučajni hod, (3) proces pokretnih proseka (moving average (MA) process) , (4) autoregresivni (AR) proces, (5) mešoviti proces autoregresivnih odstupanja i pokretnih proseka (an autoregressive moving average (ARMA) process), (6) proces autoregresivnih integrisanih pokretnih proseka (an autoregressive integrated moving average (ARIMA) process) itd.

3.1. Beli šum

Ovo je diskretni proces $\{Y_t\}$ koji se sastoji od nezavisnih identično raspodeljenih slučajnih promenljivih.

$$Y_t = \varepsilon_t$$

Ima konstantno očekivanje i varijansu, a autokovarijansna funkcija je

$$\gamma(k) = Cov(Y_t, Y_{t+k}) = 0 \; za \; k \neq 0$$

Autokorelaciona funkcija je data sa

$$\rho(k) = \begin{cases} 1 & za \; k = 0 \\ 0 & za \; k \neq 0 \end{cases}$$

Beli šum se uopšteno koristi za generisanje odgovora na impulse. On se još koristi kao osnova nekih generatora slučajnih brojeva.

3.2. Slučajni hod

Pretpostavimo da je $\{\varepsilon_t\}$ beli šum sa sredinom μ i varijansom σ^2. Tada će proces $\{Y_t\}$ biti sličajni hod ako važi

$$Y_t = Y_{t-1} + \varepsilon_t$$

Pretpostavimo sada da je Y_0 jednako sa nula. Tada je proces:

$$Y_1 = \varepsilon_1$$

$$Y_2 = Y_1 + \varepsilon_2 = \varepsilon_1 + \varepsilon_2 \; itd.$$

Važi da je

$$Y_t = \sum_{i=1}^{t} \varepsilon_i$$

Stoga $E(Y_t) = t\mu$ i $var(Y_t) = t\sigma^2$. Pošto se sredina i varijansa menjaju sa *t*, proces je nestacionaran.

3.3. Proces pokretnih proseka

U toku ovog rada već smo započeli proučavanje odstupanja pokretnih proseka pa ćemo sada detaljnije proučavati ovaj proces.Pretpostavimo da je $\{\varepsilon_t\}$ beli šum sa očekivanjem μ i varijansom σ^2. Tada je proces $\{Y_t\}$ definisan sa

$$Y_t = \theta_0\varepsilon_t + \theta_1\varepsilon_{t-1} + \cdots + \theta_m\varepsilon_{t-m}$$

nazvan proces pokretnih preseka reda *m* i označava se sa MA(*m*). Pošto su ε neopažene promenljive, skaliramo ih tako da je $\theta_0 = 1$. Pošto je $E(\varepsilon_t) = 0$ za sve *t*, imamo da je $E(Y_t) = 0$. Takođe, $Var(Y_t) = (\sum_{i=0}^{m} \theta_i^2)\sigma^2$ pošto su ε_i nezavisne sa zajedničkom varijansom σ^2.

$$\gamma(k) = Cov(Y_t, Y_{t-k})$$

$$= \begin{cases} \sigma^2 \sum_{i=0}^{m-k} \theta_i\theta_{i+k} & za\ k = 0,1,2,\dots,m \\ 0 & za\ k > m \end{cases}$$

Razmatranjem $Cov(Y_t, Y_{t+k})$ dobija se isti izraz pa stoga važi $\gamma(-k) = \gamma(k)$.

Autokorelaciona funkcija može biti dobijena deljenjem $\gamma(k)$ sa $Var(Y_t)$. Za MA proces, $\rho(k) = 0$ za $k > m$, tj. ona je nula za lagove veće od reda procesa. Pošto je γ_k nezavisno od *t*, MA(*m*) proces je slabo stacionaran.

Da bi se olakšala notacija koristimo *operator laga L*. On je definisan sa $L^j Y_t = Y_{t-j}$ za sve *j*. Stoga $LY_t = Y_{t-1}, L^2 Y_t = Y_{t-2,} L^{-1} Y_t = Y_{t+1}$, itd.

Sa ovom notacijom MA(*m*) proces može biti napisan kao (pošto je $\theta_0 = 1$)

$$Y_t = (1 + \theta_1 L + \theta_2 L^2 + \cdots + \theta_m L^m)\varepsilon_t = \theta(L)\varepsilon_t$$

Polinom u L ima *m* korena i možemo pisati

$$Y_t = (1 - \pi_1 L)(1 - \pi_2 L) \cdots (1 - \pi_m L)\varepsilon_t$$

gde su $\pi_1, \pi_2, \dots, \pi_m$ koreni jednačine

$$Z^m + \theta_1 Z^{m-1} + \cdots + \theta_m = 0$$

Posle ocenjivanja modela možemo računati reziduale iz $\varepsilon_t = [\theta(L)]^{-1} y_t$ tako da $[\theta(L)]^{-1}$ konvergira. Ovaj uslov je nazvan *uslov invertibilnosti*. Uslov za invertibilnost je da $|\pi_i| < 1$ za sve *i*. Ovo implicira da MA (*m*) proces može biti napisan kao AR(∞) proces.

3.4 . Autoregresivni proces

Na početku rada samo proučavali autoregresivna odstupanja pa ćemo sada definisati autoregresivni proces. Pretpostavimo da je $\{\varepsilon_t\}$ beli šum sa očekivanjem μ i varijansom σ^2. Tada je proces $\{Y_t\}$ dat sa

$$Y_t = \alpha_1 Y_{t-1} + \alpha_2 Y_{t-2} + \cdots + \alpha_r Y_{t-r} + \varepsilon_t \tag{3.1}$$

nazvan autoregresivni proces reda *r* i označava se sa AR(*r*).

U terminima operatora laga L, AR proces (3.1) može biti napisan kao

$$Y_t = (\alpha_1 L + \alpha_2 L^2 + \cdots + \alpha_r L^r) Y_t + \varepsilon_t$$

ili

$$(1 - \alpha_1 L - \alpha_2 L^2 - \cdots - \alpha_r L^r) Y_t = \varepsilon_t \tag{3.2}$$

ili

$$Y_t = \frac{1}{(1 - \alpha_1 L - \alpha_2 L^2 - \cdots - \alpha_r L^r)} \varepsilon_t$$

$$= \frac{1}{(1 - \pi_1 L)(1 - \pi_2 L) \cdots (1 - \pi_r L)} \varepsilon_t$$

gde su $\pi_1, \pi_2, \dots, \pi_r$ koreni jednačine

$$Z^r - \alpha_1 Z^{r-1} - \alpha_2 Z^{r-2} - \cdots - \alpha_r = 0$$

Uslov da bi razvoj jednačine (3.2) bio validan i varijansa od Y_t konačna je da $|\pi_i| < 1$ za sve *i*.

Da bi se dobila autokovarijansna funkcija, pretpostavimo da je proces stacionaran i posmatramo $\rho(k)$. Da bismo ovo dobili množimo jednačinu (3.1) sa Y_{t-k}, uzimamo očekivanja na svim uslovima i delimo sa $Var(Y_t)$ koja je po pretpostavci konačna. Ovo nam daje

$$\rho(k) = \alpha_1 \rho(k-1) + \cdots + \alpha_r \rho(k-r)$$

Zamenjujući $k = 1,2, \dots, r$ i označavajući $\rho(k) = \rho(-k)$ dobijamo jednačine za određivanje *r* parametara $\alpha_1, \alpha_2, \dots, \alpha_r$. Ove jednačine su poznate kao *Yule- Walker jednačine*.

Za AR (2) proces, procedura je sledeća

$$Y_t = \alpha_1 Y_{t-1} + \alpha_2 Y_{t-2} + \varepsilon_t$$

π_1 i π_2 su koreni jednačine

$$Z^2 - \alpha_1 Z - \alpha_2 = 0$$

Stoga $|\pi_i| < 1$ implicira da

$$\left|\frac{\alpha_1 \pm \sqrt{\alpha_1^2 + 4\alpha_2}}{2}\right| < 1$$

Ovo daje

$$\alpha_1 + \alpha_2 < 1$$

$$\alpha_1 - \alpha_2 > 1$$

$$|\alpha_2| < 1$$

U slučaju AR (2) procesa možemo dobiti $\rho(k)$ rekurzivno koristeći Yule- Walker jednačine.

Važi da je

$$\rho(0) = 1 \text{ i } \rho(1) = \alpha_1\rho(0) + \alpha_2\rho(-1)$$

$$= \alpha_1\rho(0) + \alpha_2\rho(1)$$

ili

$$\rho(1) = \frac{\alpha_1}{1 - \alpha_2}$$

Stoga

$$\rho(2) = \alpha_1\rho(1) + \alpha_2\rho(0)$$

$$= \frac{\alpha_1^2}{1 - \alpha_2} + \alpha_2$$

$$\rho(3) = \alpha_1\rho(2) + \alpha_2\rho(1)$$

$$= \frac{\alpha_1(\alpha_1^2 + \alpha_2)}{1 - \alpha_2} + \alpha_1\alpha_2$$

itd.

3.5 . Mešoviti proces autoregresivnih odstupanja i pokretnih proseka

Ovi modeli su kombinacija AR i MA modela i nazivaju se mešoviti modeli autoregresivnih odstupanja i pokretnih proseka (ARMA modeli). ARMA (p, q) model je definisan sa

$$Y_t = \alpha_1 Y_{t-1} + \cdots + \alpha_p Y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q}$$

gde je $\{\varepsilon_t\}$ beli šum sa sredinom nula i varijansom σ^2.

Korišćenjem operatora laga L, možemo ovo napisati kao

$$\Phi(L)Y_t = \varphi(L)\varepsilon_t$$

gde su $\Phi(L)$ i $\varphi(L)$ polinomi reda p i q, respektivno, definisani kao

$$\Phi(L) = 1 - \alpha_1 L - \alpha_2 L^2 - \cdots - \alpha_p L^p$$

$$\varphi(L) = 1 + \theta_1 L + \theta_2 L^2 + \cdots + \theta_q L^q$$

Za stacionarnost zahtevamo da koreni od $\Phi(L) = 0$ leže van jediničnog kruga. Za invertibilnost MA komponente, zahtevamo da koreni od $\varphi(L)$ leže van jediničnog kruga.

Autokorelaciona i autokovarijansna funkcija za ARMA model su mnogo komplikovanije nego za AR i MA model.

Izvođenje autokorelacione funkcije za ARMA (1, 1) proces:

$$Y_t = \alpha_1 Y_{t-1} + \varepsilon_t + \theta_1 \varepsilon_{t-1}$$

U terminima operatora laga L ovo može biti napisano kao

$$Y_t - \alpha_1 Y_{t-1} = \varepsilon_t + \theta_1 \varepsilon_{t-1}$$

ili

$$(1 - \alpha_1 L)Y_t = (1 + \theta_1 L)\varepsilon_t$$

ili

$$Y_t = \frac{1 + \theta L}{1 - \alpha L}\varepsilon_t$$

$$= (1 + \theta L)(1 + \alpha L + \alpha^2 L^2 + \cdots)\varepsilon_t$$

$$= [1 + (\alpha + \theta)L + \alpha(\alpha + \theta)L^2 + \alpha^2(\alpha + \theta)L^3 + \cdots]\varepsilon_t$$

Pošto je ε_t beli šum sa varijansom σ^2 dobijamo

$$Var(Y_t) = [1 + (\alpha + \theta)^2 + \alpha^2(\alpha + \theta)^2 + \cdots]\sigma^2$$

$$= \left(1 + \frac{(\alpha + \theta)^2}{1 - \alpha^2}\right)\sigma^2 = \frac{1 + \theta^2 + 2\alpha\theta}{1 - \alpha^2}\sigma^2$$

$$Cov(Y_t, Y_{t-1}) = [(\alpha + \theta) + \alpha(\alpha + \theta)^2 + \alpha^3(\alpha + \theta)^2 + \cdots \sigma^2]$$

$$= \left(\alpha + \theta + \frac{(\alpha + \theta)^2 \alpha}{1 - \alpha^2}\right)\sigma^2$$

$$= \frac{(\alpha + \theta)(1 + \alpha\theta)}{1 - \alpha^2}\sigma^2$$

Stoga

$$\rho(1) = \frac{Cov(Y_t, Y_{t-1})}{Var(Y_t)} = \frac{(\alpha + \theta)(1 + \alpha\theta)}{1 + \theta^2 + 2\alpha\theta}$$

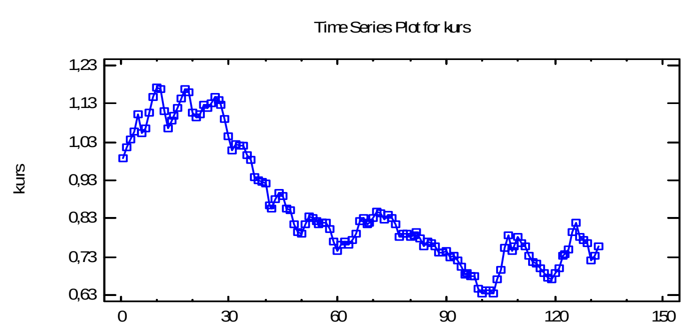
Sukcesivne vrednosti od $\rho(k)$ mogu biti dobijene iz rekurentne relacije $\rho(k) = \alpha\rho(k-1)$ za $k \geq 2$. Za AR (1) proces sa $\rho(1) = \alpha$, $\rho(1)$ za ARMA (1, 1) proces je ili $> \alpha$ ili $< \alpha$ što zavisi od toga da li je $\theta > 0$ ili $\theta < 0$, respektivno.

3.6. Proces autoregresivnih integrisanih pokretnih proseka

U praksi, većina vremenskih serija su nestacionarne. Procedura koja je često korišćena za konvertovanje nestacionarne serije u stacionarnu seriju je sukcesivno diferenciranje. Definišemo operator $\Delta = 1 - L$, tako da $\Delta Y_t = Y_t - Y_{t-1}$, $\Delta^2 Y_t = (Y_t - Y_{t-1}) - (Y_{t-1} - Y_{t-2})$, itd. Pretpostavimo da je $\Delta^d Y_t$ stacionarna serija koja može biti reprezentovana sa ARMA (*p*,*q*) modelom. Tada se može reći da Y_t može biti reprezentovano modelom autoregresivnih integrisanih pokretnih preseka ARIMA (*p*, *d*, *q*).

Primer 2: ARIMA modeli u Statgraphics-u

Kurs dolar- evro



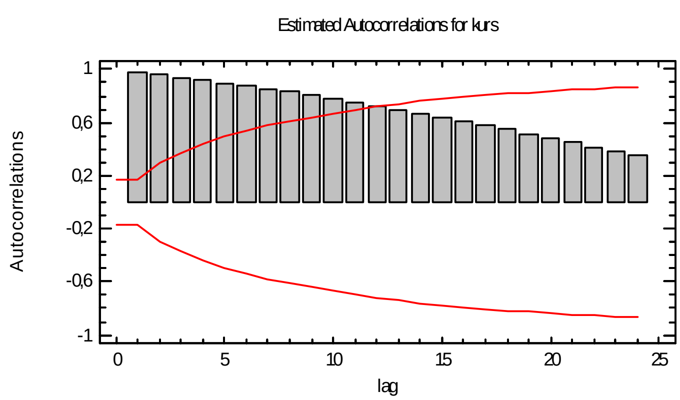
Slika 3: Promene kursa u periodu od Januara 2000 do Decembra 2010

Estimated Autocorrelations for kurs

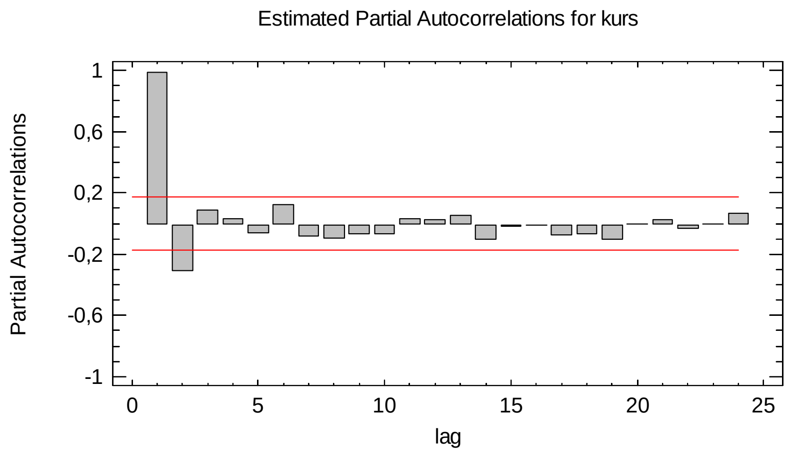
			Lower 95,0%	*Upper 95,0%*
Lag	*Autocorrelation*	*Stnd. Error*	*Prob. Limit*	*Prob. Limit*
1	0,985024	0,0870388	-0,170593	0,170593
2	0,960989	0,149254	-0,292534	0,292534
3	0,936972	0,190445	-0,373266	0,373266
4	0,914952	0,222646	-0,436378	0,436378
5	0,892421	0,249509	-0,48903	0,48903
6	0,872719	0,27262	-0,534327	0,534327
7	0,853145	0,293022	-0,574313	0,574313
8	0,830481	0,311271	-0,610082	0,610082
9	0,803779	0,327628	-0,64214	0,64214
10	0,774358	0,342241	-0,670781	0,670781
11	0,745078	0,355266	-0,69631	0,69631
12	0,717681	0,366913	-0,719138	0,719138
13	0,693497	0,377398	-0,739688	0,739688
14	0,668573	0,386932	-0,758374	0,758374
15	0,642296	0,395587	-0,775337	0,775337
16	0,614846	0,40341	-0,79067	0,79067
17	0,585336	0,410448	-0,804464	0,804464
18	0,552434	0,416723	-0,816765	0,816765
19	0,516571	0,422235	-0,827567	0,827567
20	0,481005	0,426996	-0,836898	0,836898
21	0,447626	0,431081	-0,844905	0,844905
22	0,415186	0,434588	-0,851779	0,851779
23	0,38282	0,437583	-0,857648	0,857648
24	0,352627	0,440113	-0,862607	0,862607

Tabela 2: Ocenjene autokorelacije za reziduale kursa dolar- evro

U ovom slučaju, 11 od 24 koeficijenata autokorelacije su statistički značajni na nivou poverenja od 95 %, sa najvećom vrednošću na prvom lagu što implicira da serija možda nije u potpunosti slučajna. Beli šum za ARIMA modele mora biti stacionaran, što znači da su očekivana vrednost serije i njena autokovarijansna funkcija nezavisni od vremena. Standardan način da se proveri nestacionarnost je da se nacrta serija i njena autokorelaciona i parcijalna autokorelaciona funkcija.



Slika 4: Autokorelaciona funkcija za kurs dolar-evro



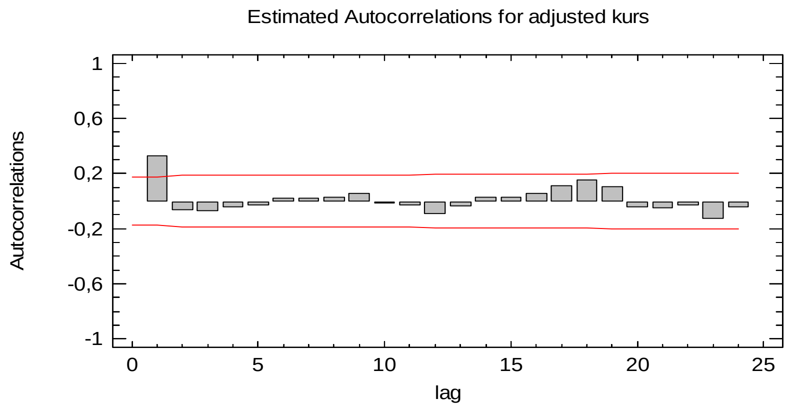
Slika 5: Parcijalna autokorelaciona funkcija za kurs dolar-evro

Ako je serija nestacionarna, njena autokorelaciona funkcija će polako opadati i serija se mora tranformisati u stacionarnu pre nego što se primene ARIMA modeli. To radimo tako što diferenciramo seriju. Pošto je to ovde slučaj prvo diferenciramo seriju.

Estimated Autocorrelations for adjusted kurs

			Lower 95,0%	*Upper 95,0%*
Lag	*Autocorrelation*	*Stnd. Error*	*Prob. Limit*	*Prob. Limit*
1	0,32733	0,0873704	-0,171243	0,171243
2	-0,0685129	0,0962777	-0,188701	0,188701
3	-0,076991	0,0966491	-0,189429	0,189429
4	-0,052172	0,0971162	-0,190345	0,190345
5	-0,0358649	0,0973299	-0,190763	0,190763
6	0,0202508	0,0974307	-0,190961	0,190961
7	0,0219271	0,0974629	-0,191024	0,191024
8	0,0259532	0,0975005	-0,191098	0,191098
9	0,0547886	0,0975532	-0,191201	0,191201
10	-0,0198418	0,0977878	-0,191661	0,191661
11	-0,0375339	0,0978186	-0,191721	0,191721
12	-0,0944237	0,0979284	-0,191937	0,191937
13	-0,0390674	0,098621	-0,193294	0,193294
14	0,0287118	0,0987391	-0,193525	0,193525
15	0,0268003	0,0988028	-0,19365	0,19365
16	0,0576774	0,0988582	-0,193759	0,193759
17	0,108346	0,0991148	-0,194262	0,194262
18	0,151423	0,100015	-0,196026	0,196026
19	0,101841	0,10175	-0,199426	0,199426
20	-0,0497999	0,102525	-0,200946	0,200946
21	-0,05394	0,102709	-0,201307	0,201307
22	-0,0343687	0,102925	-0,201731	0,201731
23	-0,135607	0,103013	-0,201902	0,201902
24	-0,0507415	0,104367	-0,204556	0,204556

Tabela 3: Ocenjene autokorelacije reziduala za deferenciranu vremensku seriju

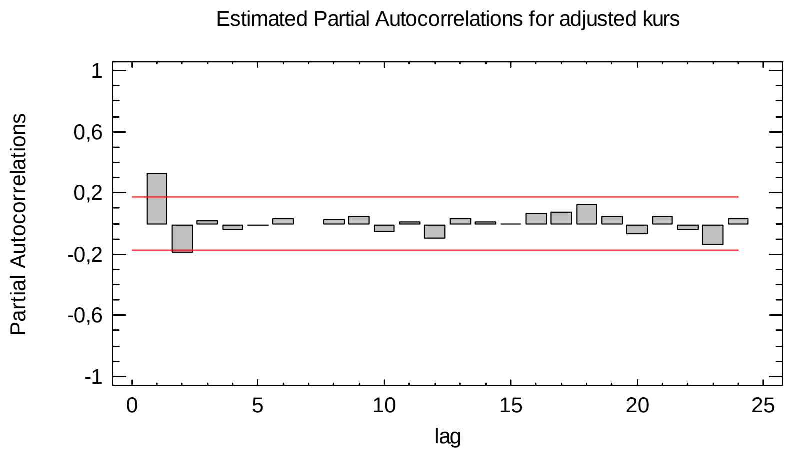


Slika 6: Ocenjene autokorelacije za diferenciranu vremensku seriju

Estimated Partial Autocorrelations for adjusted kurs

	Partial		*Lower 95,0%*	*Upper 95,0%*
Lag	*Autocorrelation*	*Stnd. Error*	*Prob. Limit*	*Prob. Limit*
1	0,32733	0,0873704	-0,171243	0,171243
2	-0,196737	0,0873704	-0,171243	0,171243
3	0,0165975	0,0873704	-0,171243	0,171243
4	-0,0479202	0,0873704	-0,171243	0,171243
5	-0,0157619	0,0873704	-0,171243	0,171243
6	0,0328512	0,0873704	-0,171243	0,171243
7	-0,00868268	0,0873704	-0,171243	0,171243
8	0,0278815	0,0873704	-0,171243	0,171243
9	0,0450147	0,0873704	-0,171243	0,171243
10	-0,0583764	0,0873704	-0,171243	0,171243
11	0,00822622	0,0873704	-0,171243	0,171243
12	-0,101906	0,0873704	-0,171243	0,171243
13	0,0318802	0,0873704	-0,171243	0,171243
14	0,0116895	0,0873704	-0,171243	0,171243
15	-0,00670605	0,0873704	-0,171243	0,171243
16	0,0637332	0,0873704	-0,171243	0,171243
17	0,0746744	0,0873704	-0,171243	0,171243
18	0,121578	0,0873704	-0,171243	0,171243
19	0,0490467	0,0873704	-0,171243	0,171243
20	-0,0767915	0,0873704	-0,171243	0,171243
21	0,0436841	0,0873704	-0,171243	0,171243
22	-0,0485391	0,0873704	-0,171243	0,171243
23	-0,145881	0,0873704	-0,171243	0,171243
24	0,0347758	0,0873704	-0,171243	0,171243

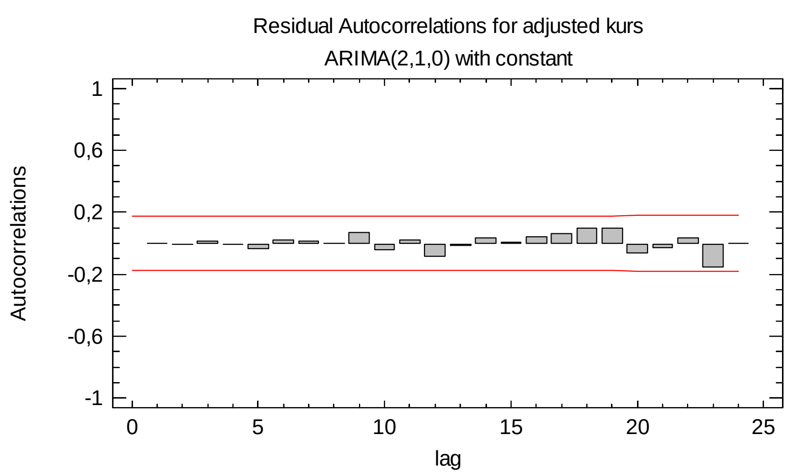
Tabela 4: Ocenjene parcijalne autokorelacije reziduala za deferenciranu vremensku seriju



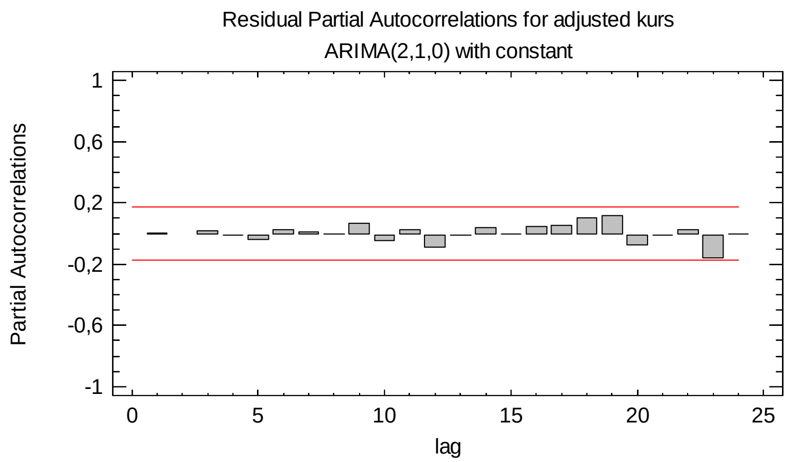
Slika 7: Ocenjene parcijalne autokorelacije za diferenciranu vremensku seriju

Na osnovu oblika autokorelacione i parcijalne autokorelacione funkcije sledi da bi odgovarajući model bio ARIMA(2,1,0) model.

Ako izabrani ARIMA model dobro fituje podatke, reziduali bi trebalo da budu slučajni i svi „ pravougaonici“ bi trebalo da ostanu unutar granica. Ako neki „ pravougaonik“ prevazilazi granicu to je znak statistički značajne autokorelacije između reziduala.



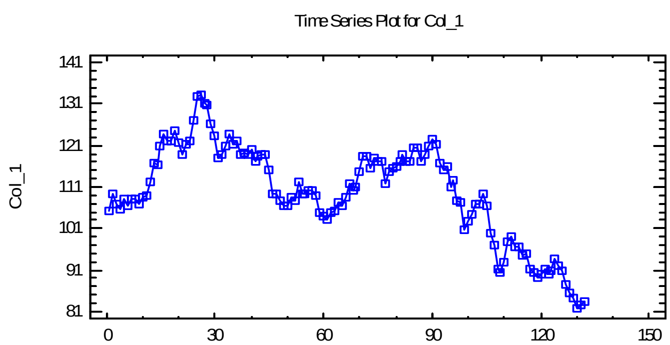
Slika 8: Rezidualne autokorelacije za ARIMA (2,1,0) model



Slika 9: Rezidualne parcijalne autokorelacije za ARIMA (2,1,0) model

Pošto su ovde svi „ pravougaonici “ unutar granica još jednom potvrđujemo da je model ARIMA (2,1,0) odgovarajući.

Kurs dolar-jen



Slika10: promene kursa u periodu od Januara 2000 do Decembra 2010

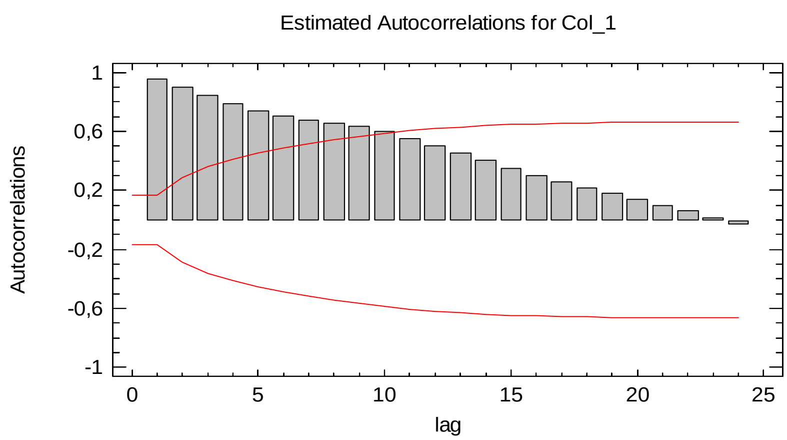
Estimated Autocorrelations for Col_1

			Lower 95,0%	*Upper 95,0%*
Lag	*Autocorrelation*	*Stnd. Error*	*Prob. Limit*	*Prob. Limit*
1	0,954995	0,0870388	-0,170593	0,170593
2	0,902385	0,146267	-0,286679	0,286679
3	0,842926	0,183663	-0,359973	0,359973
4	0,789334	0,210944	-0,413444	0,413444
5	0,740803	0,232245	-0,455192	0,455192
6	0,704909	0,249505	-0,489021	0,489021
7	0,67927	0,264162	-0,517748	0,517748
8	0,658199	0,277078	-0,543065	0,543065

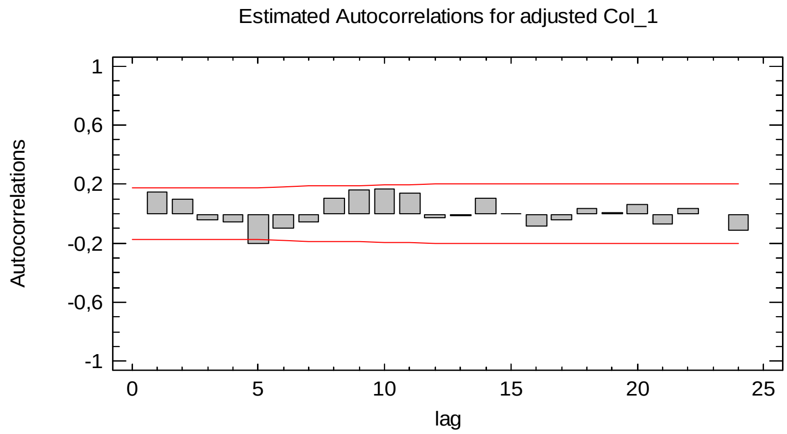
9	0,633895	0,28868	-0,565804	0,565804
10	0,597656	0,29904	-0,586108	0,586108
11	0,552234	0,307956	-0,603583	0,603583
12	0,502033	0,315369	-0,618112	0,618112
13	0,452623	0,321366	-0,629867	0,629867
14	0,40255	0,32616	-0,639262	0,639262
15	0,350154	0,329902	-0,646597	0,646597
16	0,29964	0,332706	-0,652092	0,652092
17	0,257969	0,334744	-0,656087	0,656087
18	0,217641	0,336247	-0,659032	0,659032
19	0,179262	0,337312	-0,661121	0,661121
20	0,139473	0,338033	-0,662534	0,662534
21	0,0991268	0,338469	-0,663388	0,663388
22	0,0613402	0,338689	-0,663819	0,663819
23	0,014313	0,338773	-0,663984	0,663984
24	-0,0355328	0,338777	-0,663992	0,663992

Tabela 5: Ocenjene autokorelacije za reziduale kursa dolar- jen

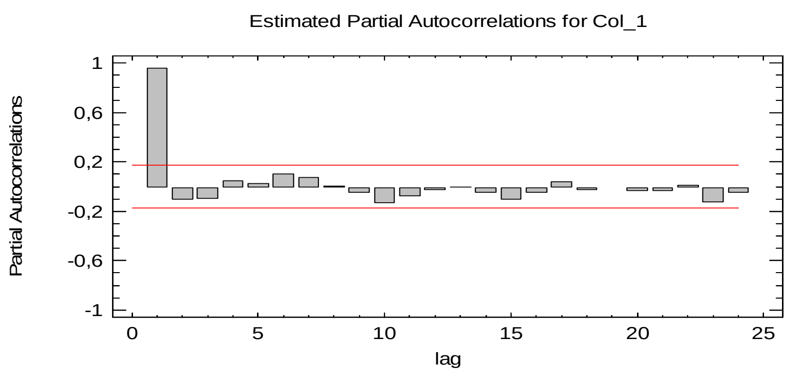
U ovom slučaju, 10 od 24 koeficijenata autokorelacije su statistički značajni na nivou poverenja od 95 %, sa najvećom vrednošću na prvom lagu što implicira da serija možda nije u potpunosti slučajna pa je moramo diferencirati.



Slika 11: Autokorelaciona funkcija za kurs dolar-jen



Slika 12: Ocenjene autokorelacije za diferenciranu seriju kursa dolar- jen



Slika 13: Ocenjene parcijalne autokorelacije za diferenciranu seriju kursa dolar- jen.

Iz oblika auotkorelacione i parcijalne autokorelacione funkcije se dobija da je model slučajnog hoda najadekvatniji za podatke kursa dolar-jen.

4. Ocenjivanje modela

4.1 . Ocenjivanje AR modela

Ocenjivanje AR modela je direktno. Ocenjujemo ih metodom najmanjih kvadrata minimiziranjem $\sum \varepsilon_t^2$.

Pre ocenjivanja modela treba odrediti red autoregresije, *p*. Jedna od najčešćih metoda za izbor reda *p* je Akaike FPE kriterijum.

Biramo red p minimiziranjem

$$FPE = \hat{\sigma}_p^2 \left(1 + \frac{1+p}{n}\right)$$

gde je $\hat{\sigma}_p^2$ ocena od $\sigma^2 = var(\varepsilon_t)$ kada je red autoregresionog procesa *p*, tj.,

$\hat{\sigma}_p^2 = SSR/(n-p-1)$.

4.2. Ocenjivanje MA modela

Razmatramo MA(2) model

$$Y_t = \mu + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2}$$

U slučaju MA modela, ne možemo odrediti $\sum \varepsilon_t^2$ kao jednostavnu funkciju opaženih *y* i parametara u AR modelima. Može se primeniti *grid- search procedura* predložena od Box i Jenkins. U ovoj proceduri računamo ε_t pomoću sukcesivnih zamena za svaku vrednost od (θ_1, θ_2) dajući neke početne vrednosti, $\mu = \bar{y}$ i $\varepsilon_0 = \varepsilon_{-1} = 0$. Tada za MA(2) model važi,

$$\varepsilon_1 = y_1 - \mu$$

$$\varepsilon_2 = y_2 - \mu - \theta_1 \varepsilon_1$$

$$\varepsilon_t = y_t - \mu - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} \;\; za\; t \geq 3$$

Ove sukcesivne vrednosti od ε_t mogu biti generisane i $\sum \varepsilon_t^2$ može biti izračunato za svaki skup vrednosti (θ_1, θ_2).

Ova procedura nije primenljiva ako imamo puno parametara u MA procesu.

4.3. Ocenjivanje ARMA modela

Ponovo se primenjuje grid- search procedura za MA komponentu. Razmatramo ARMA(2, 2) model

$$Y_t = \alpha_1 Y_{t-1} + \alpha_2 Y_{t-2} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2}$$

Ovo može biti napisano kao

$$(1-\alpha_1 L-\alpha_2 L^2)Y_t=\varepsilon_t+\theta_1\varepsilon_{t-1}+\theta_2\varepsilon_{t-2}$$

ili

$$Y_t=\frac{1}{1-\alpha_1 L-\alpha_2 L^2}(\varepsilon_t+\theta_1\varepsilon_{t-1}+\theta_2\varepsilon_{t-2}) \qquad (4.1)$$

Neka je

$$Z_t=\frac{1}{1-\alpha_1 L-\alpha_2 L^2}\varepsilon_t$$

Dalje važi da je

$$Z_t-\alpha_1 Z_{t-1}-\alpha_2 Z_{t-2}=\varepsilon_t \qquad (4.2)$$

Iz jednačine (4.1) imamo

$$Y_t=Z_t+\theta_1 Z_{t-1}+\theta_2 Z_{t-2}$$

ili

$$Z_t=Y_t-\theta_1 Z_{t-1}-\theta_2 Z_{t-2}$$

Grid- search procedura je sledeća: Počinjemo sa $Z_0=Z_{-1}=0$, generišemo sukcesivne vrednosti od Z_t za različite skupove vrednosti za (θ_1,θ_2) u regionu datim sa

$$\theta_1+\theta_2>-1$$

$$\theta_2-\theta_1>-1$$

$$|\theta_2|<1$$

kao sledeće:

$$Z_1=Y_1$$

$$Z_2=Y_2-\theta_1 Z_1$$

$$Z_t=Y_t-\theta_1 Z_{t-1}-\theta_2 Z_{t-2} \quad za\ t\geq 3$$

Mi tada koristimo generisane Z_t za ocenjivanje parametara (α_1,α_2) u jednačini (4.2) pomoću metode najmanjih kvadrata. Mi biramo one vrednosti (θ_1,θ_2) koje minimiziraju $\sum\hat{\varepsilon}_t^2$. Odgovarajuće vrednosti $\hat{\alpha}_1$ i $\hat{\alpha}_2$ daju ocene od α_1 i α_2.

Za ARIMA modele procedura opisana iznad se koristi posle sukcesivnog diferenciranja serije sve dok ne postane stacionarna.

4.3.1. Reziduali iz ARMA modela

Posle dobijanja ocena parametara $(\alpha_1, \alpha_2, \theta_1, \theta_2)$, dobijamo predviđene reziduale iz jednačine (4.2). Imamo

$$\hat{\varepsilon}_t = \hat{Z}_t - \alpha_1 \hat{Z}_{t-1} - \alpha_2 \hat{Z}_{t-2}$$

$\hat{Z}_t$ su dobijene iz Y i konačnih ocena $\hat{\theta}_1$ i $\hat{\theta}_2$.

Alternativni način dobijanja reziduala je rešavanje ARMA (2, 2) modela proširenjem izraza u terminima operatora laga L. Primetimo da model $(1 - \alpha_1 L - \alpha_2 L^2) y_t = (1 + \theta_1 L + \theta_2 L^2) \varepsilon_t$ daje

$$\varepsilon_t = (1 + \theta_1 L + \theta_2 L^2)^{-1} (1 - \alpha_1 L - \alpha_2 L^2) y_t$$

Pošto je ovo stepena serija u L, moramo je napisati kao

$$(1 + \gamma_1 L + \gamma_2 L^2 + \gamma_3 L^3 + \gamma_4 L^4 + \cdots) y_t$$

Moramo pronaći γ_i. Ovo može biti učinjeno pomoću $(1 + \theta_1 L + \theta_2 L^2)(1 + \gamma_1 L + \gamma_2 L^2 + \gamma_3 L^3 + \gamma_4 L^4 + \cdots) = (1 - \alpha_1 L - \alpha_2 L^2)$ i izjednačavanjem koeficijenata od stepena L. Ovo daje

koeficijent od $L = \theta_1 + \gamma_1 = -\alpha_1$ ili $\gamma_1 = -(\alpha_1 + \theta_1)$

koeficijent od $L^2 = \theta_2 + \theta_1 \gamma_1 + \gamma_2 = -\alpha_2$ ili $\gamma_2 = -(\alpha_2 + \theta_2) + \theta_1 (\alpha_1 + \theta_1)$

koeficijent od $L^3 = \theta_1 \gamma_2 + \theta_2 \gamma_1 + \gamma_3 = 0$ ili $\gamma_3 = -(\theta_1 \gamma_2 + \theta_2 \gamma_1)$

Ostatak od γ može biti dobijen rekurzivno iz relacije

koeficijent od $L^{j+1} = \gamma_{j+1} = -\left(\theta_1 \gamma_j + \theta_2 \gamma_{j-1}\right)$

Kada dobijemo γ_j, možemo napisati $\varepsilon_t = y_t + \gamma_1 y_{t-1} + \gamma_2 y_{t-2} + \cdots$ i dobiti ocenjene reziduale $\hat{\varepsilon}_t$ zamenjujući $\hat{\gamma}$ umesto γ.

<u>Ocenjivanje kursa dolar- evro</u>

Ocenjujemo ARIMA (2, 1, 0) model

ARIMA Model Summary

Parameter	*Estimate*	*Stnd. Error*	*t*	*P-value*
AR(1)	0,396141	0,0865175	4,57874	0,0000
AR(2)	-0,19837	0,0865972	-2,29072	0,0236
Mean	-0,00173535	0,00230225	-0,753764	0,4524
Constant	-0,00139215			

Backforecasting: yes
Estimated white noise variance = 0,000449829 with 128 degrees of freedom
Estimated white noise standard deviation = 0,0212092

Tabela 6: Ocenjeni koeficijenti ARIMA (2,1,0) modela

4.4 . Procenjivanje adekvatnosti modela

Za procenjivanje adekvatnosti primenjuju se dva kriterujuma: Akaike informacioni kriterijum (AIC) i Bajesovski informacioni kriterijum (BIC).

Ako je p ukupan broj ocenjenih parametara, važi

$$\mathrm{AIC}(p) = n\,\log\hat{\sigma}_p^2 + 2p$$

$$\mathrm{BIC}(p) = n\,\log\hat{\sigma}_p^2 + p\log n$$

Ovde je n veličina uzorka. Ako je SSR rezidualna suma kvadrata, $\sum \hat{\varepsilon}_t^2$, tada je $\hat{\sigma}_p^2 = SSR/(n-p)$. Ako razmatramo nekoliko ARMA modela, biramo onaj sa najmanjim AIC ili BIC. (Dva kriterijuma mogu voditi ka različitim zaključcima). Zatim treba proveriti obrazac autokorelacije reziduala da bi bili sigurni da ne postoji autokorelacija. Može se posmatrati autokorelacija prvog reda u rezidualima. Kod autoregresivnih modela koristimo Durbin h- test ili m test.

<u>Kurs dolar- evro</u>

AIC kriterijum za ARIMA (2, 1, 0) model je $-7{,}68107$

BIC kriterijum za ARIMA (2, 1, 0) model je $-7{,}63739$

5. Predviđanje iz ARIMA modela

Predviđanje iz ARMA (2, 2) modela

Pretpostavimo da imamo ocenjeni model sa *n* opažanja i želimo da predvidimo y_{n+k}. Ovo je tzv. prognoza unapred za *k* perioda. Ona je označena sa $\hat{y}_{n,k}$, gde *n* daje vremenski period kada je prognoza načinjena, a k označava vremenski period unapred za koji je prognoza načinjena. Počinjemo sa $k = 1$ tako da nam treba prognoza od y_{n+1} u periodu n. Imamo

$$y_{n+1} = \alpha_1 y_n + \alpha_2 y_{n-1} + \varepsilon_{n+1} + \theta_1 \varepsilon_n + \theta_2 \varepsilon_{n-1}$$

Opažamo y_n i y_{n-1}, a ε_n i ε_{n-1} možemo zameniti predviđenim rezidualima. Jedino što je nepoznato je ε_{n+1} i njega zamenjujemo njegovom očekivanom vrednošću nula. Stoga

$$\hat{y}_{n,1} = \hat{\alpha}_1 y_n + \hat{\alpha}_2 y_{n-1} + \hat{\theta}_1 \hat{\varepsilon}_n + \hat{\theta}_2 \hat{\varepsilon}_{n-1}.$$

Neka je sada $k = 2$. Važi da je

$$y_{n+2} = \alpha_1 y_{n+1} + \alpha_2 y_n + \varepsilon_{n+2} + \theta_1 \varepsilon_{n+1} + \theta_2 \varepsilon_n$$

Zamenjujemo ε_{n+2} i ε_{n+1} njihovim očekivanim vrednostima nula. y_{n+1} nije poznato, ali imamo prognozu $\hat{y}_{n,1}$. Stoga dobijamo

$$\hat{y}_{n,2} = \hat{\alpha}_1 y_{n,1} + \hat{\alpha}_2 y_n + \hat{\theta}_2 \hat{\varepsilon}_n.$$

Procedura je sledeća:

1. Napišemo izraz za y_{n+k}
2. Zamenimo sve buduće vrednosti y_{n+j} $(j > 0, j < k)$ njihovim prognozama
3. Zamenimo sve ε_{n+j} $(j > 0)$ vrednošću nula.
4. Zamenimo sve ε_{n-j} $(j \leq 0)$ predviđenim rezidualima.

Predviđanje kursa dolar- evro

Forecasting - kurs
Data variable: kurs

Number of observations = 132
Start index = 1,0
Sampling interval = 1,0

Forecast Summary
Nonseasonal differencing of order: 1
Forecast model selected: ARIMA(2,1,0) with constant
Number of forecasts generated: 36
Number of periods withheld for validation: 12

	Estimation	*Validation*
Statistic	*Period*	*Period*
RMSE	0,0207075	0,0254593
MAE	0,0160596	0,0213281
MAPE	1,84534	2,82334
ME	-0,000223887	0,00630568
MPE	-0,0499672	0,820999

Tabela 7: Test statistike za period ocenjivanja i potvrđivanja

ARIMA Model Summary

Parameter	*Estimate*	*Stnd. Error*	*t*	*P-value*
AR(1)	0,401588	0,0907854	4,42348	0,000021
AR(2)	-0,20031	0,0906903	-2,20872	0,028974
Mean	-0,00222813	0,00236291	-0,942958	0,347479
Constant	-0,00177965			

Backforecasting: yes
Estimated white noise variance = 0,000429493 with 128 degrees of freedom
Estimated white noise standard deviation = 0,0207242
Number of iterations: 1

Tabela 8: Ocenjeni koeficijenti za ARIMA(2,1,0) model

Svaka od test statistika je zasnovana na greškama prognoze jedan korak unapred, koje su razlike između vrednosti podataka u vremenu t i prognoze te vrednosti načinjene u vremenu t-1. Prve tri test statistike mere opseg grešaka, bolji model daće manju vrednost. Poslednje dve test statistike mere pristrasnost. Bolji model daće vrednosti bliske nuli. Tabela pokazuje test statistike i za period ocenjivanja i za period potvrđivanja. Ako su rezultati znatno lošiji u periodu potvrđivanja, to znači da model najverovatnije neće tako biti dobar u predviđanju budućih vrednosti kao drugi očekivani modeli.

$F_t(k)$ = prognoza za vreme $t+k$ načinjena u vremenu t

$\mathrm{e_t} =$ greška prognoziranja jedan period unapred koja je izračunata iz

$$\mathrm{e_t = Y_t - F_{t-1}(1)}$$

Pošto su m opažanja na kraju vremenske serije zadržana u svrhe potvrđivanja, test statistike su sledeće:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{m} e_{n+i}^2}{m}} \quad - \textit{the root mean squared error}$$

$$MAPE = 100\frac{\sum_{i=1}^{m} |e_{n+i}/Y_{t+i}|}{m}\% \quad - \textit{the mean absolute percentage error}$$

$$MAE = \frac{\sum_{i=1}^{m} |e_{n+i}|}{m} \quad - \textit{the mean absolute error}$$

$$ME = \frac{\sum_{i=1}^{m} e_{n+i}}{m} \quad - \textit{the mean error}$$

$$MPE = 100\frac{\sum_{i=1}^{m}\frac{e_{n+i}}{Y_{n+i}}}{m} - the\ mean\ percentage\ error$$

Forecast Table for kurs
Model: ARIMA(2,1,0) with constant
V = withheld for validation

Period	*Data*	*Forecast*	*Residual*	
1,0	0,987			
2,0	1,016	0,988622	0,0273776	
3,0	1,036	1,02469	0,0113075	
4,0	1,058	1,03644	0,0215569	
5,0	1,102	1,06105	0,0409509	
6,0	1,053	1,11348	-0,0604834	
7,0	1,065	1,02273	0,0422711	
8,0	1,106	1,07785	0,0281454	
9,0	1,148	1,11828	0,0297183	
10,0	1,172	1,15487	0,0171257	
11,0	1,168	1,17145	-0,00344545	
12,0	1,111	1,15981	-0,0488066	
13,0	1,065	1,08713	-0,0221311	
14,0	1,085	1,05616	0,0288351	
15,0	1,101	1,10047	0,000533661	
16,0	1,119	1,10164	0,0173604	
17,0	1,144	1,12124	0,022756	
18,0	1,171	1,14865	0,0223455	
19,0	1,161	1,17506	-0,0140555	
20,0	1,109	1,1498	-0,0407961	
21,0	1,096	1,08834	0,00765912	
22,0	1,104	1,09942	0,0045842	
23,0	1,126	1,10804	0,0179629	
24,0	1,12	1,13145	-0,0114528	
25,0	1,133	1,1114	0,021596	
26,0	1,149	1,13764	0,0113572	
27,0	1,141	1,15104	-0,0100417	
28,0	1,128	1,1328	-0,00480269	
29,0	1,089	1,1226	-0,0336022	
30,0	1,045	1,07416	-0,0291624	
31,0	1,008	1,03336	-0,0253626	
32,0	1,023	1,00018	0,0228248	
33,0	1,019	1,03466	-0,0156556	
34,0	1,019	1,01261	0,00639065	
35,0	0,998	1,01802	-0,0200216	
36,0	0,982	0,987787	-0,005787	
37,0	0,94	0,978001	-0,0380014	
38,0	0,928	0,924559	0,00344139	
39,0	0,927	0,929814	-0,00281429	
40,0	0,92	0,927222	-0,00722247	
41,0	0,864	0,91561	-0,0516095	
42,0	0,857	0,841134	0,0158664	
43,0	0,88	0,863627	0,0163734	
44,0	0,897	0,888859	0,00814097	
45,0	0,889	0,89744	-0,00844022	
46,0	0,855	0,880602	-0,0256024	
47,0	0,853	0,841169	0,0118312	
48,0	0,813	0,857228	-0,0442277	
49,0	0,793	0,795557	-0,00255745	
50,0	0,791	0,791201	-0,00020097	
51,0	0,816	0,792423	0,0235766	
52,0	0,834	0,824661	0,00933934	
53,0	0,832	0,834441	-0,00244119	
54,0	0,823	0,825812	-0,0028116	

55,0	0,815	0,818007	-0,00300667	
56,0	0,82	0,81181	0,00818957	
57,0	0,818	0,821831	-0,00383076	
58,0	0,801	0,814416	-0,0134156	
59,0	0,769	0,792794	-0,023794	
60,0	0,745	0,757775	-0,0127748	
61,0	0,763	0,739992	0,0230079	
62,0	0,768	0,773256	-0,00525635	
63,0	0,759	0,764623	-0,00562271	
64,0	0,773	0,752605	0,0203955	
65,0	0,789	0,778645	0,0103546	
66,0	0,823	0,790841	0,0321586	
67,0	0,83	0,831669	-0,00166938	
68,0	0,813	0,824221	-0,0112209	
69,0	0,817	0,802991	0,0140088	
70,0	0,831	0,820232	0,010768	
71,0	0,848	0,834041	0,0139587	
72,0	0,843	0,850243	-0,00724301	
73,0	0,825	0,835807	-0,0108071	
74,0	0,838	0,816993	0,0210067	
75,0	0,831	0,845047	-0,0140466	
76,0	0,814	0,823805	-0,00980521	
77,0	0,783	0,806796	-0,0237955	
78,0	0,79	0,772176	0,0178236	
79,0	0,788	0,797241	-0,00924106	
80,0	0,781	0,784015	-0,003015	
81,0	0,785	0,77681	0,00819015	
82,0	0,792	0,786229	0,00577114	
83,0	0,776	0,79223	-0,0162302	
84,0	0,758	0,766393	-0,00839277	
85,0	0,769	0,752197	0,0168033	
86,0	0,764	0,775243	-0,0112434	
87,0	0,755	0,758009	-0,003009	
88,0	0,74	0,750608	-0,0106076	
89,0	0,74	0,733999	0,00600069	
90,0	0,745	0,741225	0,00377501	
91,0	0,729	0,745228	-0,0162283	
92,0	0,734	0,719793	0,0142066	
93,0	0,719	0,737433	-0,0184332	
94,0	0,703	0,710195	-0,00719498	
95,0	0,681	0,6978	-0,0167996	
96,0	0,687	0,67359	0,0134096	
97,0	0,68	0,692037	-0,0120367	
98,0	0,679	0,674207	0,00479263	
99,0	0,644	0,678221	-0,0342209	
100,0	0,635	0,628365	0,00663492	
101,0	0,643	0,636617	0,00638311	
102,0	0,643	0,646236	-0,00323583	
103,0	0,635	0,639618	-0,00461787	
104,0	0,669	0,630008	0,0389924	
105,0	0,697	0,682477	0,0145232	
106,0	0,753	0,699654	0,0533457	
107,0	0,786	0,768101	0,0178994	
108,0	0,745	0,786255	-0,0412554	
109,0	0,758	0,720145	0,037855	
110,0	0,781	0,769654	0,0113463	
111,0	0,767	0,785853	-0,0188528	
112,0	0,757	0,754991	0,002009	
113,0	0,732	0,754009	-0,0220088	
114,0	0,714	0,722184	-0,00818375	
115,0	0,71	0,71	4,96999E-7	
116,0	0,701	0,71022	-0,00921957	
117,0	0,687	0,696407	-0,00940729	

118,0	0,675	0,681401	-0,0064009	
119,0	0,67	0,671206	-0,00120563	
120,0	0,686	0,668616	0,0173839	
121,0	0,7	0,691647	0,0083527	V
122,0	0,731	0,700638	0,0303624	V
123,0	0,737	0,738865	-0,00186523	V
124,0	0,747	0,73142	0,0155797	V
125,0	0,795	0,748034	0,0469656	V
126,0	0,819	0,810493	0,00850653	V
127,0	0,782	0,817244	-0,0352436	V
128,0	0,775	0,760554	0,0144458	V
129,0	0,764	0,777821	-0,0138207	V
130,0	0,72	0,759205	-0,039205	V
131,0	0,734	0,702754	0,0312461	V
132,0	0,757	0,746656	0,0103438	V

		Lower 95,0%	*Upper 95,0%*
Period	*Forecast*	*Limit*	*Limit*
133,0	0,761653	0,720646	0,802659
134,0	0,757134	0,686531	0,827737
135,0	0,752608	0,662571	0,842645
136,0	0,749916	0,645976	0,853856
137,0	0,747962	0,632329	0,863595
138,0	0,745936	0,619575	0,872298
139,0	0,743735	0,60735	0,88012
140,0	0,741477	0,595725	0,887229
141,0	0,739231	0,584691	0,893772
142,0	0,737002	0,574157	0,899847
143,0	0,734777	0,564033	0,905521
144,0	0,732551	0,554256	0,910846
145,0	0,730322	0,544783	0,915862
146,0	0,728094	0,535582	0,920606
147,0	0,725866	0,526626	0,925106
148,0	0,723638	0,517889	0,929386
149,0	0,72141	0,509352	0,933467
150,0	0,719181	0,500998	0,937365
151,0	0,716953	0,492811	0,941096
152,0	0,714725	0,484778	0,944672
153,0	0,712497	0,476888	0,948106
154,0	0,710269	0,469131	0,951406
155,0	0,708041	0,461498	0,954583
156,0	0,705813	0,453982	0,957644
157,0	0,703584	0,446573	0,960596
158,0	0,701356	0,439268	0,963445
159,0	0,699128	0,432058	0,966198
160,0	0,6969	0,42494	0,96886
161,0	0,694672	0,417909	0,971435
162,0	0,692444	0,410959	0,973929
163,0	0,690216	0,404087	0,976344
164,0	0,687988	0,39729	0,978686
165,0	0,685759	0,390563	0,980956
166,0	0,683531	0,383903	0,983159
167,0	0,681303	0,377309	0,985298
168,0	0,679075	0,370776	0,987375

Tabela 9: Prognoziranje vrednosti kursa dolar-evro

Model Comparison
Data variable: kurs
Number of observations = 132
Start index = 1,0
Sampling interval = 1,0
Number of periods withheld for validation: 12

Models
(A) ARIMA(2,1,0) with constant
(B) Linear trend = 1,11906 + -0,00413479 t
(C) Simple moving average of 3 terms
(D) Simple exponential smoothing with alpha = 0,9999
(E) Brown's linear exp. smoothing with alpha = 0,7795

Estimation Period

Model	*RMSE*	*MAE*	*MAPE*	*ME*	*MPE*
(A)	0,0207075	0,0160596	1,84534	-0,000223887	-0,0499672
(B)	0,0676194	0,055175	6,38599	4,16334E-17	-0,500228
(C)	0,0325482	0,0249687	2,85184	-0,00585185	-0,781579
(D)	0,022327	0,0175089	2,006	-0,00250862	-0,334645
(E)	0,0247651	0,0191031	2,1853	0,000447682	0,0684301

Tabela 10: Test statistike izračunate za period ocenjivanja

Model	*RMSE*	*RUNS*	*RUNM*	*AUTO*	*MEAN*	*VAR*
(A)	0,0207075	OK	OK	OK	OK	*
(B)	0,0676194	***	***	***	OK	**
(C)	0,0325482	***	***	***	OK	**
(D)	0,022327	**	OK	OK	OK	**
(E)	0,0247651	*	OK	OK	OK	**

Tabela 11: Testovi za ispitivanje modela

Validation Period

Model	*RMSE*	*MAE*	*MAPE*	*ME*	*MPE*
(A)	0,0254593	0,0213281	2,82334	0,00630568	0,820999
(B)	0,163516	0,15908	20,9067	0,15908	20,9067
(C)	0,0363344	0,0322778	4,26237	0,0107222	1,36231
(D)	0,0264156	0,0224177	2,95788	0,0059172	0,760216
(E)	0,0294494	0,0222705	2,94714	0,000900868	0,13445

Tabela 12: Test statistike izračunate za period potvrđivanja

RUNM = računa broj puta koliko serija se pomera iznad ili ispod svoje medijane i taj je broj upoređen sa očekivanom vrednošću za vremensku seriju koja je slučajna. Mala *p*- vrednost pokazuje da reziduali nisu u potpunosti slučajni.

AUTO = Box – Pierce test: konstruiše se test statistika zasnovana na prvih *k* rezidualnih autokorelacija izračunavajući

$$Q = n\sum_{i=1}^{k} r_i^2$$

Test statistika je upoređena sa hi- kvadrat raspodelom sa *k* stepeni slobode. Mala *p*- vrednost pokazuje da reziduali nisu u potpunosti slučajni.

MEAN = test za razliku u očekivanju između prve i druge polovine

VAR = test za razliku u varijansi između prve i druge polovine

OK = nije značajno ($p >= 0{,}05$)

$*$=marginalno značajno ($0{,}01 < p <= 0{,}05$)

$**$= značajno ($0{,}001 < p <= 0{,}01$)

$***$= visoko značajno ($p <= 0{,}001$)

ARIMA model i za period ocenjivanja i za period potvrđivanja ima najmanju RMSE , MAE i MAPE. Izabrani ARIMA model prolazi 4 testa (4 znaka OK) od ukupno 5 pa sledi da je model najverovatnije odgovarajući za predviđanje kursa. Model nije prošao jedino VAR test.

6. Detekcija autlajera

Otkrivanje autlajera je primarni korak u mnogim data- mining aplikacijama. Detekcija autlajera za data- mining je često zasnovana na merama udaljenosti, klaster i prostornim metodama.

U mnogim zadacima analize podataka velik broj promenljivih je bilo evidentirano ili uzorkovano. Jedan od prvih koraka ka dobijanju koherentne analize je otkrivanje udaljenih opažanja. Iako su autlajeri često smatrani greškom ili šumom, oni mogu nositi važnu informaciju. Otkriveni autlajeri mogu voditi ka pogrešnoj specifikaciji modela, pristrasnoj oceni parametara i netačnim rezultatima. Stoga je važno identifikovati autlajere pre modelovanja i analize.

Tačna definicija autlajera često zavisi od sakrivenih pretpostavki u pogledu na strukturu podataka i metod detekcije. Hawkins (Hawkins, 1980) definiše autlajer kao opažanje koje mnogo odstupa od drugih opažanja kao sumnja koja je bila generisana različitim mehanizmom. Barnet i Lewis (Barnett i Lewis, 1994) ukazuju da udaljeno opažanje, ili autlajer, je opažanje koje deluje da odstupa upadljivo od drugih članova uzorka u kom se događa. Johnson (Johnson, 1992) definiše autlajer kao opažanje u skupu podataka koje deluje da je nekonzistentno sa ostatkom skupa podataka.

Metode otkrivanja autlajera su predlagane za brojne aplikacije, kao što su otkrivanje prevara sa kreditnim karticama, kliničke studije, analize neregularnosti glasanja, čišćenje podataka, upad na mreže, predviđanje vremena, geografski informacioni sistemi, i drugi data- mining zadaci.

Metode otkrivanja autlajera mogu biti podeljene na univarijantne i multivarijantne metode. Pored ovog metode otkrivanja autlajera se mogu podeliti na parametarske (statističke) metode i neparametarske metode. Statističke parametarske metode ili pretpostavljaju poznatu osnovnu raspodelu opažanja ili su zasnovane na statističkim ocenama nepoznatih parametara raspodele. Ove metode označavaju kao autlajere ona opažanja koja odstupaju od pretpostavki modela. One su često neodgovarajuće za višedimenzionalne skupove podataka i za proizvoljne skupove podataka bez prethodnog znanja o fundamentalnoj raspodeli podataka.

Unutar klase neparametarskih metoda detekcije autlajera može se postaviti odvojen skup data-mining metoda takođe nazvanih metode zasnovane na razdaljini. Ove metode su zasnovane na lokalnim merama udaljenosti i sposobne su da rukuju sa velikim bazama podataka. Druga klasa metoda detekcije autlajera je osnovana na klaster tehnikama, gde klaster male veličine može biti smatran grupisanim autlajerima.

Prethodno deli podatke u dva disjunktna skupa: autlajere i one koji nisu autlajeri. Poslednje nudi rangiranje pomoću dodeljivanja svakog podatka autlajerskom klasifikacionom faktoru koji reflektuje njegov stepen udaljenosti. Druga povezana klasa metoda se sastoji od tehnika detekcije prostornih autlajera. Ove metode traže ekstremna opažanja ili lokalne nestabilnosti sa respektom na susedne vrednosti, iako ova opažanja možda neće biti značajno različita od cele populacije.

6.1. Univarijantne statističke metode

Većina od najranijih univarijantnih metoda za detekciju autlajera se oslanja na pretpostavku o poznatoj osnovnoj raspodeli podataka za koju se pretpostavlja da će biti identički i nezavisno raspodeljena. Mnogi testovi neslaganja (disonance) za otkrivanje univarijantnih autlajera dalje pretpostavljaju da su raspodela parametara i tip očekivanih autlajera poznati, mada su ove pretpostavke često narušene u realnosti za data- mining aplikacije.

Centralna pretpostavka u statistički zasnovanim metodama za otkrivanje autlajera je generisani model koji dozvoljava malom broju opažanja da budu slučajno uzorkovani iz raspodela $G_1, \dots, G_k$, koje se razlikuju od ciljne raspodele F, za koju se često uzima da je normalna raspodela $N(\mu, \sigma^2)$. Problem identifikacije autlajera je tada prenet na problem identifikovanja onih opažanja koja leže u tzv. autlajerskom regionu.

Za svaki koeficijent poverenja α , $0 < \alpha < 1$, α- autlajer region $N(\mu, \sigma^2)$ raspodele je definisan sa

$$out(\alpha, \mu, \sigma^2) = \{x : |x - \mu| > z_{1-\alpha/2}\sigma\} \qquad (6.1)$$

gde je z_q kvantil $N(0,1)$ raspodele. Broj x je α- autlajer sa respektom na F ako

$x \in out(\alpha, \mu, \sigma^2)$. Iako je tradicionalno normalna raspodela korišćena kao ciljna raspodela, ova definicija može biti proširena na svaku unimodalnu simetričnu raspodelu sa pozitivnom funkcijom gustine, uključujući i multivarijantan slučaj.

6.1.1. Jednokoračne nasuprot sekvencijalnih procedura

Davies i Gather (Davies i Gather, 1993) prave važnu razliku između jednokoračnih i sekvencijalnih procedura za detekciju autlajera. Jednokoračne procedure identifikuju sve autlajere kao suprotnost sukcesivnoj eliminaciji ili dodavanju podataka. U sekvencijalnoj proceduri, na svakom koraku, jedno opažanje je testirano da li je autlajer.

Sa respektom na jednačinu (6.1), zajedničko pravilo za nalaženje autlajerskog regiona u jednokoračnom identifikatoru je dato sa

$$out(\alpha_n, \hat{\mu}_n, \hat{\sigma}_n^2) = \{x : |x - \hat{\mu}_n| > g(n, \alpha_n)\hat{\sigma}_n\} \qquad (6.2)$$

gde je n veličina uzorka, $\hat{\mu}_n$, $\hat{\sigma}_n$ su ocenjena sredina i standardna devijacija ciljne raspodele zasnovane na uzorku, α_n označava koeficijent poverenja koji prati korekciju za višestruke testove upoređenja i $g(n, \alpha_n)$ definiše granicu (kritičan broj standardne devijacije) autlajerskog regiona.

Tradicionalno, $\hat{\mu}_n$, $\hat{\sigma}_n$ su ocenjene sa uzoračkom sredinom, $\bar{x}_n$ i uzoračkom standardnom devijacijom, S_n. Pošto su ove ocene jako pogođene prisustvom autlajera, mnoge procedure često ih zamenjuju sa drugim, robustnijim ocenama. Korekcija višestrukog upoređivanja je korišćena kada je nekoliko statističkih testova izvedeno istovremeno. Dok data α- vrednost može biti odgovarajuća za odlučivanje da li jednostruko opažanje leži u autlajerskom regionu (npr., jednostruko poređenje), ovo nije slučaj za skup nekoliko poređenja. Da bi se izbegle podmetnute pozitivne α- vrednosti mora biti smanjeno računanje za brojna izvedena poređenja. Jedna od popularnih i jednostavnih korekcija koristi $\alpha_n = 1 - (1 - \alpha)^{1/n}$. Kritična vrednost $g(n, \alpha_n)$ je često određena numeričkim procedurama, kao što su Monte Carlo simulacije za različite veličine uzorka.

6.1.2. Unutrašnje i spoljašnje procedure

Sekvencijalni identifikatori mogu biti dalje klasifikovani na unutrašnje ili spoljašnje procedure. U unutrašnjem testiranju, u svakom koraku procedure „ najekstremnije opažanje “, npr., jedno opažanje sa najvećom merom udaljenosti je testirano da li je autlajer. Ako je opažanje deklarisano kao autlajer, ono se briše iz skupa podataka i procedura se ponavlja. Ako je ono proglašeno za neudaljeno opažanje, procedura se okončava.

U spoljašnjim procedurama testiranja, uzorak opažanja je prvo smanjen na manji uzorak, a ukljonjena opažanja se čuvaju u rezervoaru. Statistike su izračunate na osnovu smanjenog uzorka i tada se uklonjena opažanja u rezervoaru testiraju obrnutim redom da bi se pokazalo da li su oni autlajeri. Ako je opažanje deklarisano kao autlajer, ono se briše iz rezervoara. Ako je opažanje deklarisano kao neudaljeno opažanje, ono se briše iz rezervoara, dodaje se smanjenom uzorku, statistike se ponovo izračunavaju i procedura se ponavlja sa novim opažanjem. Spoljašnja procedura testiranja je okončana kada nema više preostalih opažanja u rezervoaru.

6.1.3. Univarijantne robustne mere

Tadicionalno, uzoračka sredina i uzoračka varijansa daju dobre ocene za lokaciju podataka i oblik podataka ako oni nisu kontaminirani autlajerima. Kada je baza podataka kontaminirana, ovi parametri mogu odstupati i značajno delovati na performansu detekcije autlajera.

Hampel (Hampel, 1971 ; Hampel, 1974) je uveo koncept tačke sloma, kao mere robustnosti ocenjivača protiv autlajera. Tačka sloma je definisana kao najmanji procenat autlajera koji može prouzrokovati da ocenjivač primi proizvoljno velike vrednosti. Stoga, što ocenjivač ima veću tačku sloma robustniji je. Npr., uzoračka sredina ima tačku sloma od $1/n$ pošto jedno veće opažanje može učiniti da uzoračka sredina i varijansa pređu svaku granicu. Prema tome, Hampel je predlagao medijanu i medijansku apsolutnu devijaciju (MAD) kao robustne ocene lokacije. Hampel identifikator je često praktično veoma efikasan. Drugi rad koji naglašava problem robustnih ocenjivača je predložen od Tukey (Tukey, 1977). Tukey je uveo Boxplot kao grafički prikaz na kom autlajeri mogu biti naznačeni. Boxplot je zasnovan na kvadrantima raspodele. Prvi i treći kavadrant , Q_1 i Q_3, su korišćeni za dobijanje robustnih mera za sredinu, $\hat{\mu}_n = (Q_1 + Q_3)/2$ i standardnu devijaciju, $\hat{\sigma}_n = Q_3 - Q_1$. Drugo popularno rešenje za dobijanje robustnih mera je da se zameni sredina sa medijanom i izračuna standardna devijacija zasnovana na $(1 - \alpha)$ procentu podataka, gde je uobičajeno α; 5%.

6.1.4. Statistički kontrolni procesi (SPC)

Polje statističkih kontrolnih procesa (SPC) je blisko povezano sa univarijantnim metodama otkrivanja autlajera.

Ben- Gal et. al (Gal *et al*., 2003) kategoriše SPC metode prema dva kriterijuma:

1) metode za nezavisne podatke nasuprot metoda za zavisne podatke

2) metode koje su model- specifične nasuprot metoda koje su model- uopštene

Metode specifičnih modela zahtevaju a priori pretpostavke o karakteristikama procesa, uobičajeno definisanih osnovnom analitičkom raspodelom ili izrazom bliskog oblika. Metode uopštenih modela pokušavaju da ocene osnovni model sa minimumom a priori pretpostavki.

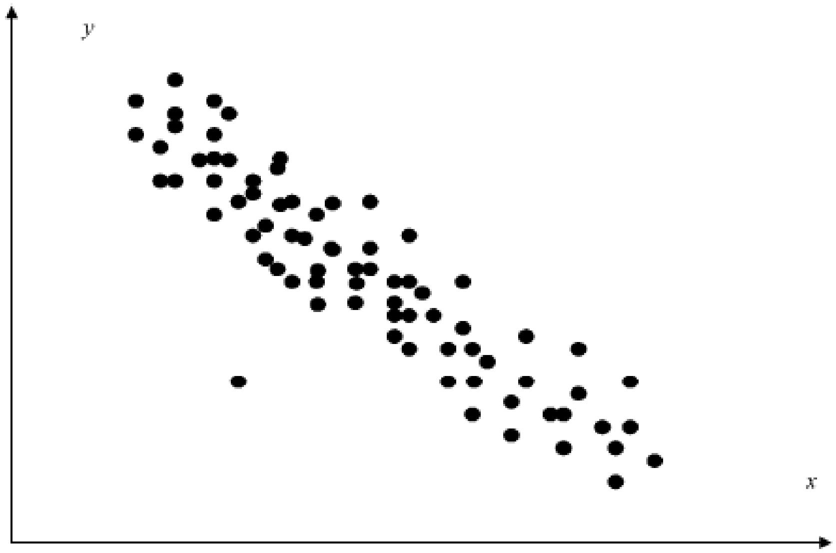
Većina model- specifičnih metoda za zavisne podatke su zasnovane na vremenskim serijama. Često, fundamentalni princip ovih metoda je sledeći: naći model vremenske serije koji može najbolje uhvatiti autokorelacioni proces, upotrebiti ovaj model za filtrisanje podataka i tada primeniti tradicionalne SPC šeme za reziduale.

Specijalno, ARIMA familija modela koju smo već proučavali u ovom radu je implementirana za ocenjivanje i filtrisanje autokorelacionog procesa. Pod sigurnim pretpostavkama, reziduali ARIMA modela su nezavisni i aproksimativno normalno raspodeljeni, na kojim tradicionalan SPC može biti primenjen. Uopšteno važi da ARIMA modeli, najčešće oni jednostavni kao AR model mogu efikasno opisati široku raznovrsnost industrijskih procesa.

Model- specifične metode za zavisne podatke mogu biti dalje podeljene na parametarski-zavisne metode koji zahtevaju eksplicitno ocenjivanje parametara modela i parametarski-slobodne metode u kojima su parametri modela samo implicitno izvedeni, ako su izvedeni uopšte.

6.2. Multivarijantna detekcija autlajera

U mnogim slučajevima multivarijabilna opažanja ne mogu biti otkrivena kao autlajeri kada je svaka promenljiva razmatrana nezavisno. Otkrivanje autlajera je moguće samo kada je multivarijantna analiza izvedena i interakcije između različitih promenljivih su poređene unutar klase podataka. Jednostavan primer je prikazan na Grafiku 3 koji predstavlja podatke koji imaju dve mere na dvodimenzionalnom prostoru. Donje levo opažanje je multivarijantan autlajer, ali nije univarijantan. Kada razmatramo svaku meru posebno sa respektom na širenje vrednosti duž x i y ose, možemo videti da one padaju blizu centra univarijantne raspodele. Test za autlajere mora uzeti u obzir veze između dve promenljive koje u ovom slučaju deluju abnormalne.



Grafik 4: Dvo- dimenzionalni prostor sa jednim udaljenim opažanjem (donji levi ugao)

Skupovi podataka sa višestrukim autlajerima ili klasterima autlajera su predmet za maskirane (MASKING) i preplavljene (SWAMPING) efekte.

Efekat maskiranja: Jedan autlajer maskira drugi autlajer, ako drugi autlajer može biti smatran autlajerom sam po sebi, ali ne u prisustvu prvog autlajera. Stoga, posle brisanja prvog autlajera drugi primer se pojavio kao autlajer. Maskiranje se dešava kada klaster udaljenih opažanja iskošava ocene sredine i kovarijanse ka sebi i rezultujuća distanca udaljenih opažanja od sredine je mala.

Efekat preplavljivanja: Jedan autlajer preplavljuje drugo opažanje, ako poslednji može biti smatran autlajerom samo u prisustvu prvog. Drugim rečima, posle brisanja prvog autlajera drugo opažanje postaje neudaljeno opažanje. Preplavljivanje se dešava kada grupa udaljenih opažanja iskošava ocene sredine i kovarijanse ka sebi i daleko od drugih ne udaljenih opažanja i rezultujuća distanca od ovih opažanja do sredine je velika, što čini da oni izgledaju kao autlajeri.

6.2.1. Statističke metode za multivarijantnu detekciju autlajera

Procedure multivarijantne detekcije autlajera mogu biti podeljene na statističke metode koje su zasnovane na ocenjenim parametrima raspodele i data- mining povezanim metodama koje su bez parametara.

Statističke metode za multivarijantnu detekciju autlajera često ukazuju na ona opažanja koja su locirana relativno daleko od centra raspodele podataka. Nekoliko mera udaljenosti mogu biti implementirane za takav zadatak. Mahalanobis distanca je dobro poznat kriterijum koji zavisi od ocenjenih parametara multivarijantne raspodele.

Datih *n* opažanja iz p- dimenzionalnog skupa podataka označavaju vektor uzoračke sredine sa $\bar{\boldsymbol{X}}_n$ i matricu uzoračke kovarijanse sa $\boldsymbol{V}_n$, gde je

$$\boldsymbol{V}_n = \frac{1}{n-1}\sum_{i=1}^{n}(\boldsymbol{X}_i - \overline{\boldsymbol{X}}_n)\,(\boldsymbol{X}_i - \overline{\boldsymbol{X}}_n)^T \qquad (6.3)$$

Mahalanobis distanca za svaki multivarijantan podatak $i,\ i = 1, \dots, n,$ je označena sa M_i i data sa

$$M_i = \left(\sum_{i=1}^{n}(\boldsymbol{X}_i - \overline{\boldsymbol{X}}_n)^T\,\boldsymbol{V}_n^{-1}(\boldsymbol{X}_i - \overline{\boldsymbol{X}}_n)\right)^{1/2} \qquad (6.4)$$

Prema tome, ona opažanja sa velikom Mahalanobis distancom su indikovana kao autlajeri. Efekti maskiranja i preplavljivanja imaju važnu ulogu u adekvatnosti Mahalanobis distance kao kriterijuma za otkrivanje autlajera. Naime, efekti maskiranja mogu smanjiti Mahalanobis distancu autlajera. Ovo se može dešavati, npr., kada mali klaster autlajera privlači $\overline{\boldsymbol{X}}_n$ i podiže $\boldsymbol{V}_n$ u svom smeru. Sa druge strane efekti preplavljivanja mogu povećati Mahalanobis distancu ne udaljenih opažanja. Npr., kada mali klaster autlajera privlači $\overline{\boldsymbol{X}}_n$ i podiže $\boldsymbol{V}_n$ daleko od obrasca većine opažanja.

6.2.2. Multivarijantne robustne mere

Kao i u jedno- dimenzionalnim procedurama, sredina raspodele (koja meri lokaciju) i varijansa- kovarijansa (koje mere oblik) su dve najviše korišćene statistike za analizu podataka u prisustvu autlajera. Upotreba robustnih ocena parametara multidimenzionalne raspodele često može poboljšati performansu procedura detekcije u prisustvu autlajera. Hadi (Hadi, 1992) adresira ovaj problem i predlaže da se zameni vektor sredine sa vektorom medijane promenljive i da se izračuna kovarijansna matrica za podskup onih opažanja koja imaju najmanju Mahalanobis distancu. Caussinus i Roiz (Caussinus i Roiz, 1990) predlažu robustnu ocenu za kovarijansnu matricu, koja je zasnovana na težinskim opažanjima prema njihovoj udaljenosti od centra.

7. Algoritam pokretnih prozora

U ovom radu ćemo dalje proučavati kombinaciju klaster tehnika sa algoritmom pokretnih prozora da bi se filtrirali autlajerski podaci finansijskih tržišta. Ovaj algoritam zavisi od 3 parametra koji mogu zavisiti od tipa tržišta ili finasijskog instrumenta.

Finansijska tržišta su važan proces u mehanizmu prenošenja monetarne politike. Vremenske serije finansijskih tržišta su često pogođene sa izostavljenim ili pogrešnim opažanjima ili nepoznatim spoljašnjim događajima koji mogu imati veliki uticaj na individualna opažanja. Ova sumnjiva opažanja ili autlajeri se ne otkrivaju lako. U ovom radu prikazujemo i testiramo metod za detekciju autlajera koji se primenjuje na dnevnim podacima. Ovaj metod omogućava da se posle analiziranja tipa autlajera dobiju čisti podaci finasijskog tržišta.

Dacorogna, Gencay, M$\ddot{u}$ller, Olsen i Pictet (2001) i Brownless i Gallo (2005) skiciraju dva slična algoritma za otkrivanje autlajera. Dva algoritma se sastoje od susednih opažanja, nazvanih filtrišući prozor, koji je neophodan za procenjivanje pouzdanosti jednostrukog opažanja. Takav prozor može se povećavati ili smanjivati u zavisnosti od kvaliteta podataka i volatilnosti serije. Ideja algoritma je da oceni validnost novog opažanja na osnovu njegove relativne udaljenosti od susednih najbližih validnih prošlih opažanja.

U oba slučaja numerički metodi sa problemom konvergencije su izbegnuti, pa algoritam proizvodi dobro- definisane rezultate u svim situacijama. U oba slučaja algoritam može biti opisan kao sledeći:

Neka je T vremenski raspon serije x koja treba da bude filtrirana

for $i = 1{:}k$

$f_i = x_i$

end

until $i \leq T$

compute $z_i = \left(f_i - \bar{f}_{[i-1:i-k]}\right) / \sqrt{s^2_{[i-1:i-k]} + \lambda}$

if $|z_i| < threshold\ then\ f_i = x_i$ *and increment i*

else $f_i = x_{i-1}$ *and increment i*

end

end

Za svako opažanje u uzorku izračunat je *z*- rezultat i to se koristi za odlučivanje da li da prihvatimo opažanje ili ne. U *z*- rezultat formuli $\bar{f}_{[i-1:i-k]}$ i $s^2_{[i-1:i-k]}$ reprezentuju pokretni prosek i pokretnu varijansu prethodnih k vrednosti, respektivno. *Z*- rezultat je dobijen oduzimanjem prosečne vrednosti poslednjih k validnih opažanja od i- tog vrednosnog nivoa i deljenjem sa faktorom koji zavisi od uzoračke varijanse i poslednjih k validnih opažanja i vrednosti λ koja izbegava slučajeve u kojim je imenilac 0 ili blizak ovoj vrednosti.

Izabrani algoritam je sekvencionalan i iterativan. On koristi postojeću bazu filtrisanih informacija za novo opažanje i minimalan iznos ažuriranja je neophodan. Rezultat algoritma je nova filtrisana serija f_i. Činjenica da algoritam zamenjuje moguće autlajere sa prethodnim opažanjem ne znači da će sumnjiva vrednost biti zamenjena punom vrednošću, ali da algoritam za otktrivanje autlajera umesto eliminisanja *i*- tog opažanja zamenjuje njega sa prethodnim opažanjem kao deo računice pokretnog prozora. Kada je algoritam završen, i- to opažanje može opet uzeti svoju početnu vrednost i biti označeno kao autlajer. Opažanje koje prevazilazi prag može alternativno biti eliminisano umesto da bude zamenjeno prethodnom vrednošću.

Algoritam je testiran i programiran koriščenjem programskog paketa MATLAB.

Za svaku seriju treba pronaći kombinaciju algoritamskih parametara koji minimiziraju odnos detekcije autlajera. Da bi se pronašla ova kombinacija izračunava se odnos detekcije i broj autlajera koji su bili otkriveni. Najbolja kombinacija ne treba da bude jedinstvena za sve serije, ali treba da bude robustna u smislu da ona treba da otkrije autlajer nezavisno od njegovog mesta ili veličine.

Da bi smo bili sposobni da procenimo stvarnu efikasnost filtrišućeg algoritma definišemo metod zasnovan na statistici kažnjavanja.

Neka je f_i filtrirana serija koja može još uvek sadržati dodatne autlajere i x_i prvobitna serija

If $f_i > x_i + stdv(x_i)$ *then* $d_i = f_i - \left(x_i + stdv(x_i)\right)$

else if $f_i < x_i - stdv(x_i)$ $d_i = \left(x_i - stdv(x_i)\right) - f_i$

else $d_i = 0$

end if

Svaka od vrednosti d_i predstavlja udaljenost od jednog intervala standardne devijacije do filtrisane tačke, ali samo u slučaju gde filtrisana tačka prevazilazi interval definisan sa jednom standardnom devijacijom iz prvobitnih podataka.

Mi na kraju definišemo

$$D_i = \frac{1}{T}\sum_{j=1}^{T}\left(d_j^i\right)^2$$

gde je T ukupan broj opažanja serije x_i . Prilagođavanje algoritamskih parametara je tada zasnovano na statistici kažnjavanja D_i.

Za svaku od serija pokušaćemo da pronađemo algoritamske parametre λ, k i prag koji minimiziraju D_i.

Primer 3: Detekcija autlajera

Posmatrali smo dnevne podatke za kurseve dolar- evro i dolar- jen u periodu od 01.06.2011. do 08.09.2011. godine.

```
function [fi zi mov_mean mov_std Dstat outlier outlier_num] = mwfa()

x=reshape(xx,6,[]);

%  Definisanje default vrednosti

if nargin <2,

    k = 5;

    eps = 3;

    lambda = 0.02;

elseif nargin == 3,

    k = 5;

    eps = 3;

    lambda = 0.02;

elseif nargin == 4,

    eps = 3;

    lambda = 0.02;

elseif nargin == 5

    lambda = 0.02;

end


%  Proveravanje validnosti ulaznih vrednosti

if ~isnumeric(x)

    error('Input x must be a numeric array');

end

[n, c] = size(x);
```

```
fi = x;
% Podaci u 'x' su organizovani tako da su kolone vremenske serije, a redovi vremenski
% intervali.
zi = nan(n,c);
  mov_mean = nan(n,c);
  mov_std = nan(n,c);
  outlier_mat = zeros(n,c);
  d = zeros(n,c);

  for i = k+1: n
    mov_mean(i,:) = mean(fi(i-k:i-1,:));
    zi(i,:) = fi(i,:) - mov_mean(i,:);
    mov_std(i,:) = std(fi(i-k:i-1,:));
    temp=abs(zi(i,:)) > (eps .* mov_std(i,:) + lambda);
    fi(i,temp) = x(i-1,temp);
    outlier_mat(i,:) = temp*1;
    temp1 = fi(i,:) >= (x(i,:) + mov_std(i,:));
    temp2 = fi(i,:) < (x(i,:) - mov_std(i,:));
    d(i,temp1)= fi(i,temp1)-(x(i,temp1)+mov_std(i,temp1));
    d(i,temp2)= (x(i,temp2)-mov_std(i,temp2))-fi(i,temp2);
  end
  Dstat = mean(d.^2);

  [i1,j1] = find(outlier_mat);
  if ~isempty(i1),
    outlier = [cellstr(strcat('Series', num2str(j1)))];
    outlier_num = [i1 j1];
```

```
% outlier - ćelija koja specificira broj serije (broj kolone u 'x' u kojoj su
% potencijalni autlajeri smešteni)
% outlier_num - matrica koja obezbeđuje broj vrste i kolone od vrednosti u 'x' koja je
% smatrana potencijalnim autlajerom.
 else
  outlier = ('No outliers have been identified!');
  outlier_num = ('No outliers have been identified!');
 end
```

Posle izvršenja programa u programskom paketu MATLAB zaključujemo da u deviznom kursu dolar- evro ne postoje autlajeri, dok u kursu dolar- jen postoji jedan potencijalni autlajer.

8. Nelinearni modeli

8.1. Polinomni model

Za kurs dolar- evro sa mesečnim podacima fitujemo polinomni model drugog reda. Jednačina fitovanog modela je

$$Y_t = 0{,}0807941 + 0{,}823961 * Y_{t-1} + 0{,}0901379 * Y_{t-1}^2$$

		Standard	*T*	
Parameter	*Estimate*	*Error*	*Statistic*	*P-Value*
CONSTANT	0,0807941	0,0815931	0,990208	0,3239
Yminus jedan	0,823961	0,183644	4,48673	0,0000
Yminus jedan^2	0,0901379	0,100473	0,897137	0,3713

R-squared = 97,8831 percent
R-squared (adjusted for d.f.) = 97,85 percent
Standard Error of Est. = 0,0227827
Mean absolute error = 0,0179504
Durbin-Watson statistic = 1,32263 (P=0,0000)
Lag 1 residual autocorrelation = 0,32598

Tabela 13: Ocene parametara polinomnog modela za kurs dolar- evro

U određivanju da li je red polinoma odgovarajući, posmatramo da li je *p*- vrednost najvećeg reda polinoma veća ili jednaka sa 0,05. Pošto je u ovom slučaju $p = 0{,}3713$, ovaj uslov nije statistički značajan na nivou poverenja od 95 % , pa treba razmotriti smanjivanje reda polinoma za 1.

Polinomni model reda 1

$$Y_t = 0{,}00828907 + 0{,}988311 * Y_{t-1}$$

		Standard	*T*	
Parameter	*Estimate*	*Error*	*Statistic*	*P-Value*
CONSTANT	0,00828907	0,0112096	0,739464	0,4610
Yminus jedan	0,988311	0,0128376	76,9858	0,0000

Tabela 14: Ocene parametara polinomnog modela reda 1 za kurs dolar-evro

Pošto je sada *p*- vrednost najvećeg reda polinoma manja od 0,05, red polinoma je odgovarajući.

Za kurs dolar- jen sa mesečnim podacima fitujemo polinomni model drugog reda, takođe.

$$Y_t = -18{,}5228 + 1{,}35241 * Y_{t-1} - 0{,}00166894 * Y_{t-1}^2$$

		Standard	*T*	
Parameter	*Estimate*	*Error*	*Statistic*	*P-Value*
CONSTANT	-18,5228	16,6807	-1,11044	0,2689
dolar minus jedan	1,35241	0,313118	4,31916	0,0000
dolar minus jedan^2	-0,00166894	0,00145863	-1,14419	0,2547

R-squared = 95,1634 percent
R-squared (adjusted for d.f.) = 95,0878 percent
Standard Error of Est. = 2,581
Mean absolute error = 2,04454
Durbin-Watson statistic = 1,66467 (P=0,0273)
Lag 1 residual autocorrelation = 0,156875

Tabela 15: Ocene parametara polinomnog modela za kurs dolar- jen

Pošto je i u ovom slučaju *p*- vrednost najvećeg reda polinoma veća ili jednaka sa 0,05, treba razmotriti smanjivanje reda polinoma za jedan.

Polinomni model reda 1

$$Y_t = 0{,}397417 + 0{,}99486 * Y_{t-1}$$

Parameter	*Estimate*	*Standard Error*	*T Statistic*	*P-Value*
CONSTANT	0,397417	2,19512	0,181045	0,8566
dolar minus jedan	0,99486	0,0198529	50,1116	0,0000

Tabela 16: Ocene parametara polinomnog modela reda 1 za kurs dolar-jen

Pošto je sada *p*- vrednost najvećeg reda polinoma manja od 0,05, red polinoma je odgovarajući.

8.2. Eksponencijalni model

Kurs dolar- evro

$$e^{Y_t} = 0{,}023614 + 0{,}988364 * e^{Y_{t-1}}$$

Summary Statistics; DV: e**V1 (Kurs sa mesecnim podacima)

Multiple R	0,989
Multiple R2	0,978
Adjusted R2	0,978
F(1,129)	5701,245
p	0,000
Std.Err. of Estimate	0,058

Tabela 17: Karakteristike ocenjenog modela

Durbin-Watson d (Kurs sa mesecnim podacima) and serial correlation of residuals

	Durbin- Watson d	**Serial corr.**
Estimate	1,284780	0,348229

Tabela 18: Durbin-Watson statistika za eksponencijalni model kursa dolar-evro

8.3. Logaritamski model

Kurs dolar- jen

$$\ln Y_t = -0,002999 + 1,000258 * \ln Y_{t-1}$$

Summary Statistics; DV: LN-V1 (Kurs sa mesecnim podacima 2)	
	Value
Multiple R	0,977
Multiple R2	0,955
Adjusted R2	0,954
F(1,129)	2721,448
p	0,000
Std.Err. of Estimate	0,024

Tabela 19: Karakteristike ocenjenog modela

Durbin-Watson d (Kurs sa mesecnim podacima 2) and serial correlation of residuals		
	Durbin-Watson d	**Serial corr.**
Estimate	1,661723	0,157488

Tabela 20: Durbin-Watson statistika za logaritamski model kursa dolar-jen

8.4. Recipročni model

Kurs dolar- evro

$$\frac{1}{Y_t} = 0,020163 + 0,985145 * \frac{1}{Y_{t-1}}$$

Summary Statistics; DV: 1/V1 (Kurs sa mesecnim podacima)	
	Value
Multiple R	0,987
Multiple R2	0,974
Adjusted R2	0,974
F(1,129)	4900,695
p	0,000
Std.Err. of Estimate	0,032

Tabela 21: Karakteristike ocenjenog modela

Durbin-Watson d (Kurs sa mesecnim podacima) and serial correlation of residuals		
	Durbin-Watson d	**Serial corr.**
Estimate	1,341601	0,322553

Tabela 22: Durbin-Watson statistika za recipročni model kursa dolar-evro

Kurs dolar- jen

$$\frac{1}{Y_t} = -0{,}000032 + 1{,}005552 * \frac{1}{Y_{t-1}}$$

Summary Statistics; DV: 1/V1 (Kurs sa mesecnim podacima 2)	
	Value
Multiple R	0,979
Multiple R2	0,957
Adjusted R2	0,957
F(1,129)	2904,281
p	0,000
Std.Err. of Estimate	0,000

Tabela 23: Karakteristike ocenjenog modela

Durbin-Watson d (Kurs sa mesecnim podacima 2) and serial correlation of residuals		
	Durbin- Watson d	**Serial corr.**
Estimate	1,641567	0,166797

Tabela 24: Durbin-Watson statistika za recipročni model kursa dolar-jen

ZAKLJUČAK

U ovom radu su posmatrani devizni kursevi dolar- evro i dolar- jen. Kao što je već spomenuto predviđanje deviznih kurseva se vrši pomoću familije ARIMA modela.

ARIMA modeli su veoma značajni u statistici i ekonometriji, a posebno u analizi vremenskih serija. Ovi modeli su primenjivani na podatke vremenskih serija da bi se bolje razumeli podaci ili da bi se predvidele buduće vrednosti. Posebne klase ARIMA modela su detaljnije izučavane u radu.

Analiza vremenskih serija obuhvata metode za analiziranje podataka vremenskih serija da bi se dobile značajne statistike i druge karakteristike podataka. Predviđanje vremenskih serija je upotreba modela za predviđanje budućih događaja na osnovu prošlih događaja, da bi se predvideli podaci pre nego što su oni opaženi. Metode za analizu vremenskih serija mogu biti podeljene u dve klase: metode zasnovane na učestalosti i metode zasnovane na vremenu. Metode zasnovane na vremenu uključuju analizu autokorelacija i unakrsnih korelacija.

Posle ispitivanja ARIMA modela za devizne kurseve, ispitivali smo da li su vremenske serije deviznih kurseva dolar- evro i dolar- jen kontaminirane autlajerima i došli do zaključka da u deviznom kursu dolar- evro ne postoje autlajeri, dok u deviznom kursu dolar- jen postoji jedan potencijalni autlajer.

DODATAK

Podaci za devizne kurseve dolar- evro i dolar- jen koji su korišćeni tokom izrade primera.

Dolar- evro

Januar 2000	0,987
Februar 2000	1,016
Mart 2000	1,036
April 2000	1,058
Maj 2000	1,102
Jun 2000	1,053
Jul 2000	1,065
Avgust 2000	1,106
Septembar 2000	1,148
Oktobar 2000	1,172
Novembar 2000	1,168
Decembar 2000	1,111
Januar 2001	1,065
Februar 2001	1,085
Mart 2001	1,101
April 2001	1,119
Maj 2001	1,144
Jun 2001	1,171
Jul 2001	1,161
Avgust 2001	1,109
Septembar 2001	1,096
Oktobar 2001	1,104
Novembar 2001	1,126
Decembar 2001	1,12
Januar 2002	1,133
Februar 2002	1,149
Mart 2002	1,141
Novembar 2001	1,126
Decembar 2001	1,12
Januar 2002	1,133
Februar 2002	1,149
Mart 2002	1,141
April 2002	1,128
Maj 2002	1,089
Jun 2002	1,045
Jul 2002	1,008
Avgust 2002	1,023
Septembar 2002	1,019
Oktobar 2002	1,019
Novembar 2002	0,998
Decembar 2002	0,982
Januar 2003	0,94

Februar 2003	0,928
Mart 2003	0,927
April 2003	0,92
Maj 2003	0,864
Jun 2003	0,857
Jul 2003	0,88
Avgust 2003	0,897
Septembar 2003	0,889
Oktobar 2003	0,855
Novembar 2003	0,853
Decembar 2003	0,813
Januar 2004	0,793
Februar 2004	0,791
Mart 2004	0,816
April 2004	0,834
Maj 2004	0,832
Jun 2004	0,823
Jul 2004	0,815
Avgust 2004	0,82
Septembar 2004	0,818
Oktobar 2004	0,801
Novembar 2004	0,769
Decembar 2004	0,745
Januar 2005	0,763
Februar 2005	0,768
Mart 2005	0,759
April 2005	0,773
Maj 2005	0,789
Jun 2005	0,823
Jul 2005	0,83
Avgust 2005	0,813
Septembar 2005	0,817
Oktobar 2005	0,831
Novembar 2005	0,848
Decembar 2005	0,843
Januar 2006	0,825
Februar 2006	0,838
Mart 2006	0,831
April 2006	0,814
Maj 2006	0,783
Jun 2006	0,79
Jul 2006	0,788
Avgust 2006	0,781
Septembar 2006	0,785
Oktobar 2006	0,792
Novembar 2006	0,776
Decembar 2006	0,758
Januar 2007	0,769

Februar 2007	0,764
Mart 2007	0,755
April 2007	0,74
Maj 2007	0,74
Jun 2007	0,745
Jul 2007	0,729
Avgust 2007	0,734
Septembar 2007	0,719
Oktobar 2007	0,703
Novembar 2007	0,681
Decembar 2007	0,687
Januar 2008	0,68
Februar 2008	0,679
Mart 2008	0,644
April 2008	0,635
Maj 2008	0,643
Jun 2008	0,643
Jul 2008	0,635
Avgust 2008	0,669
Septembar 2008	0,697
Oktobar 2008	0,753
Novembar 2008	0,786
Decembar 2008	0,745
Januar 2009	0,758
Februar 2009	0,781
Mart 2009	0,767
April 2009	0,757
Maj 2009	0,732
Jun 2009	0,714
Jul 2009	0,71
Avgust 2009	0,701
Septembar 2009	0,687
Oktobar 2009	0,675
Novembar 2009	0,67
Decembar 2009	0,686
Januar 2010	0,7
Februar 2010	0,731
Mart 2010	0,737
April 2010	0,747
Maj 2010	0,795
Jun 2010	0,819
Jul 2010	0,782
Avgust 2010	0,775
Septembar 2010	0,764
Oktobar 2010	0,72
Novembar 2010	0,734
Decembar 2010	0,757

Dolar- jen

Januar 2000	105,269
Februar 2000	109,36
Mart 2000	106,756
April 2000	105,583
Maj 2000	108,185
Jun 2000	106,313
Jul 2000	107,976
Avgust 2000	108,244
Septembar 2000	106,86
Oktobar 2000	108,411
Novembar 2000	108,874
Decembar 2000	112,189
Januar 2001	116,837
Februar 2001	116,166
Mart 2001	121
April 2001	123,78
Maj 2001	121,944
Jun 2001	122,205
Jul 2001	124,669
Avgust 2001	121,624
Septembar 2001	118,874
Oktobar 2001	121,363
Novembar 2001	122,312
Decembar 2001	127,033
Januar 2002	132,8
Februar 2002	133,455
Mart 2002	131,24
April 2002	130,773
Maj 2002	126,344
Jun 2002	123,255
Jul 2002	118,015
Avgust 2002	118,982
Septembar 2002	120,833
Oktobar 2002	123,909
Novembar 2002	121,448
Decembar 2002	122,006
Januar 2003	118,805
Februar 2003	119,425
Mart 2003	118,652
April 2003	119,89
Maj 2003	117,377
Jun 2003	118,339
Jul 2003	118,626
Avgust 2003	118,712
Septembar 2003	114,95
Oktobar 2003	109,5
Novembar 2003	109,165

Decembar 2003	107,795
Januar 2004	106,297
Februar 2004	106,649
Mart 2004	108,513
April 2004	107,622
Maj 2004	112,092
Jun 2004	109,468
Jul 2004	109,462
Avgust 2004	110,255
Septembar 2004	110,068
Oktobar 2004	108,787
Novembar 2004	104,705
Decembar 2004	103,846
Januar 2005	103,362
Februar 2005	104,99
Mart 2005	105,277
April 2005	107,255
Maj 2005	106,627
Jun 2005	108,707
Jul 2005	111,956
Avgust 2005	110,211
Septembar 2005	111,105
Oktobar 2005	114,847
Novembar 2005	118,456
Decembar 2005	118,374
Januar 2006	115,478
Februar 2006	117,895
Mart 2006	117,288
April 2006	117,073
Maj 2006	111,757
Jun 2006	114,668
Jul 2006	115,662
Avgust 2006	115,913
Septembar 2006	117,186
Oktobar 2006	118,678
Novembar 2006	117,285
Decembar 2006	117,371
Januar 2007	120,374
Februar 2007	120,465
Mart 2007	117,304
April 2007	118,904
Maj 2007	120,813
Jun 2007	122,677
Jul 2007	121,479
Avgust 2007	116,698
Septembar 2007	115,052
Oktobar 2007	115,897
Novembar 2007	110,921

Decembar 2007	112,52
Januar 2008	107,886
Februar 2008	107,114
Mart 2008	100,707
April 2008	102,608
Maj 2008	104,347
Jun 2008	106,934
Jul 2008	106,844
Avgust 2008	109,331
Septembar 2008	106,645
Oktobar 2008	100,072
Novembar 2008	96,861
Decembar 2008	91,31
Januar 2009	90,27
Februar 2009	92,825
Mart 2009	97,805
April 2009	98,901
Maj 2009	96,531
Jun 2009	96,627
Jul 2009	94,314
Avgust 2009	94,886
Septembar 2009	91,34
Oktobar 2009	90,301
Novembar 2009	89,144
Decembar 2009	89,917
Januar 2010	91,124
Februar 2010	90,112
Mart 2010	90,706
April 2010	93,515
Maj 2010	91,97
Jun 2010	90,83
Jul 2010	87,496
Avgust 2010	85,38
Septembar 2010	84,372
Oktobar 2010	81,803
Novembar 2010	82,552
Decembar 2010	83,27

LITERATURA

[1] Jan Kmenta, *Počela ekonometrije- Drugo izdanje*, MATE doo., Zagreb, 1997.

[2] G.S. Maddala, *Introduction to Econometrics- Third Edition*, JOHN WILEY& SONS, LTD, 2001.

[3] Ben – Gal I., *Outlier Detection,* In: Maimon O. And Rockah L. (Eds.) Data Mining and Knowledge Discovery Handbook: A Complete Guide for Practiotioners and Researches, Kluwer Academic Publishers, 2005.

[4] Josep Maria Puigvert Gutiėrrez, Josep Fortiana Gregori, *Clustering techniques applied to outlier detection of finacial market series using a moving window filtering algorithm,* Working Paper Series, NO 948, October 2008.

[5] www.mathworks.com

[6] Statgraphics help

Biografija

Milana Tedić rođena je 24. oktobra 1986. godine u Novom Sadu. Završila je Osnovnu školu „ Petefi Šandor " u Novom Sadu i prirodno-matematički smer Gimnazije „Isidora Sekulić " u Novom Sadu. Po završetku gimnazije, 2005. godine upisala je smer matematika finansija na Prirodno-matematičkom fakultetu u Novom Sadu i diplomirala 16.10.2009. godine. Iste godine upisuje master studije na smeru Primenjena matematika- matematika finansija. Položila je sve ispite predviđene nastavnim planom i programom.

UNIVERZITET U NOVOM SADU

PRIRODNO - MATEMATIČKI FAKULTET

KLJUČNA DOKUMENTACIJSKA INFORMACIJA

Redni broj: RBR

Identifikacioni broj:

IBR

Tip dokumentacije: Monografska dokumentacija

TD

Tip zapisa: Tekstualni Stampani materijal

TZ

Vrsta rada: Master teza

VR

Autor: Milana Tedić

AU

Mentor: Dr Zorana Lužanin

MN

Naslov rada: Autokorelacija i autlajeri u predviđanju vrednosti kursa

MR

Jezik publikacije: *Srpski (latinica)*

JP

Jezik izvoda: *s / e*

JI

Master rad : Autokorelacija i autlajeri u predviđanju vrednosti kursa

Zemlja publikovanja: Republika Srbija

ZP

Uže geografsko područje: Vojvodina

UGP

Godina: 2011

GO

Izdavač: Autorski reprint

IZ

Mesto i adresa: Trg Dositeja Obradovića 4, 21000 Novi Sad

MA

Fizički opis rada: *8 poglavlja/83 strane/0 lit. citata/ 24 tabele/12 slika/ 4 grafika*

1 prilog

FO

Naučna oblast: Matematika

NO

Naučna disciplina: Primenjena matematika

ND

Ključne reči: vremenske serije, autokorelacija, ARIMA(p,d,q) modeli, autlajeri

PO

UDK:

Master rad : Autokorelacija i autlajeri u predviđanju vrednosti kursa

Čuva se: u biblioteci Departmana za matematiku i informatiku Prirodno- matematičkog fakulteta u Novom Sadu

ČU

Važna napomena:

VN

Izvod: Tema ovog rada je predviđanje vrednosti deviznih kurseva. Proučavani su devizni kursevi dolar- evro i dolar- jen i za svaki od ovih kurseva ispitivana je autokorelacija u rezidualima i ocenjeni su koeficijenti ARIMA modela, pa je zatim izabranim ARIMA modelom vršeno predviđanje kurseva. U radu je prikazan algoritam za detekciju autlajera u deviznim kursevima u programskom paketu MATLAB.

IZ

Datum prihvatanja teme od strane NN veća: 26. 08. 2010.

DP

Datum odbrane:

DO

Članovi komisije:

KO

Predsednik: Dr Nataša Krejić, redovni profesor

Član: Dr Zorana Lužanin, redovni profesor

Član: Dr Dora Seleši, docent

UNIVERSITY OF NOVI SAD

FACULTY OF SCIENCE KEY

WORDS DOCUMENTATION

Accession number:

ANO

Identification number:

INO

Document type: Monograph type

DT

Type of record: Printed text

TR

Contents Code:

CC

Author: Milana Tedić

AU

Mentor: Dr Zorana Lužanin

MN

Title: Autocorrelation and outliers in predicting exchange rates

XI

Language of text: Serbian(Latin)

LT

Language of abstract: s/e

LA

Country of publication: Republic of Serbia

CP

Locality of publication: Vojvodina

LP

Publication year: 2011

PY

Publisher: Author's reprint

PU

Publ. place: Trg Dositeja Obradovića 4, 21000 Novi Sad

PP

Physical description: (*8 chapters, 83 pages, 0 references, 24 tables, 12 pictures, 4 graphs 1 contribution)*

PD

Scientific field: Mathematics

SF

Scientific discipline: Econometrics

Key words: time series, autocorrelation, ARIMA(p,d,q) models, outliers

UC:

Holding data: Department of Mathematics and Informatics' Library, Faculty of Science, Novi Sad

HD Note:

Abstract: The topic of this paper is predicting future values of exchange rates. We studied exchange rates dollar-euro and dollar-yen and examined autocorrelation of residuals and ARIMA models for this exchange rates. We investigate outliers in this exchange rates with specific algorithm in MATLAB.

AB

Accepted by the Scientific Board on: 26.08.2010.

Defended:

Thesis defend board: Dr Nataša Krejić, full professor

Member: Dr Zorana Lužanin, full professor

Member: Dr Dora Seleši, docent