Математички факултет
Универзитет у Београду

Мастер рад

Регресиона анализа временских серија са јединичним кореном и коинтеграција

Ментор:
проф др Весна Јевремовић

Студент:
Драгана Богуновић

Београд, 2015.

Садржај

1 Увод

У свакодневном животу се сусрећемо са појавама које имају свој ток у времену и које су предмет изучавања различитих наука. Тако се демографија бави прикупљањем података о годишњим стопама наталитета и морталитета, односно проучавањем природног прираштаја. Економија се бави анализом променљивих као што су цене производа, плате, девизни курс, запосленост, индустријска производња, итд. У метеорологији се сваког сата региструје брзина ветра, прати дневна температура, бележи месечна или годишња количина падавина. Пољопривреда прати годишње кретање приноса појединих пољопривредних култура, као и њихове откупне и продајне цене. Све ове појаве су примери временских серија.

Формална дефиниција временске серије биће дата у следећем поглављу. За сада је довољно рећи да се временска серија може посматрати као уређени низ података, при чему се уређење најчешће врши у односу на време и у једнаким временским интервалима. Избор јединице времена зависи од посматране појаве, па то може бити секунда, минут, сат, дан, месец, година, али и деценија, век или миленијум.

Анализа временских серија представља једну од статистичких дисциплина која бележи најдинамичнији развој последњих деценија. До овога је дошло због развоја саме дисциплине, али и због јаке интеракције са осталим дисциплинама, пре свега са економијом. Развој анализе временских серија је почео тридесетих година двадесетог века, када је почело изучавање временских серија засновано на теорији вероватноће и математичкој статистици. Принципи анализе временских серија одступају од уобичајених претпоставки теорије статистичког закључивања па се временске серије не могу изучавати методама класичне теорије вероватноће и класичне математичке статистике. Наиме, у класичној теорији вероватноће и математичкој статистици је основни појам случајан узорак који представља низ независних случајних величина. Насупрот томе временска серија је низ случајних величина које су најчешће међусобно зависне.

Као што ћемо видети у наредној глави, једна од основних особина коју временска серија може да поседује је стационарност. Слободно речено, серија је стационарна уколико је њено кретање предвидиво током времена, тј. уколико њено кретање испуњава сличан образац понашања током времена. Супротно, ако су параметри кретања временске серије функције временског тренутка, тада је серија нестационарна. Предмет изучавања овог рада је управо анализа нестационарних временских серија.

2 Временске серије

Временска серија се обично дефинише као низ података са одређеним хронолошким или линијским редоследом. Међутим дефиницији се може прићи и са становишта случајних процеса као што ће бити описано у наставку.

2.1 Случајни процеси

До појма случајног процеса долази се проширивањем појма случајне променљиве па уводимо следећих неколико дефиниција.

Дефиниција 2.1.1 *Ω је скуп свих исхода једног експеримента. Елементи овог скупа се означавају са ω_i, $i = 1, 2, 3, ...$ и називају елементарни догађаји.*

Дефиниција 2.1.2 *Подскуп $\mathcal{F}$ партитивног скупа $\mathcal{P}(\Omega)$ је алгебра над Ω ако важе услови:*

- $\Omega \in \mathcal{F}$
- $A \in \mathcal{F} \Rightarrow \overline{A} \in \mathcal{F}$
- $A \in \mathcal{F} \bigwedge B \in \mathcal{F} \Rightarrow A \bigcup B \in \mathcal{F}$.

Дефиниција 2.1.3 *Подскуп $\mathcal{F}$ партитивног скупа $\mathcal{P}(\Omega)$ је σ-алгебра над Ω ако важе услови:*

- $\Omega \in \mathcal{F}$
- $A \in \mathcal{F} \Rightarrow \overline{A} \in \mathcal{F}$
- $\{A_i\}_{i \in \mathbb{N}} \in \mathcal{F} \Rightarrow \bigcup\limits_{i=1}^{\infty} A_i \in \mathcal{F}$

Дефиниција 2.1.4 *Ако је $\mathcal{K}$ нека фамилија подскупова простора исхода Ω($\mathcal{K}$ се још назива и колекција), тада пресек свих σ-алгебри које садрже фамилију $\mathcal{K}$ називамо минимална σ-алгебра генерисана са $\mathcal{K}$ и означавамо је $\sigma[\mathcal{K}]$.*

Аналогно се дефинише минимална алгебра која садржи све скупове дате колекције.

Дефиниција 2.1.5 *Нека је простор исхода једнак скупу реалних бројева $\mathbb{R}$ и нека је колекција $\mathcal{K}$ дата са $\mathcal{K} = \{(a, b] : a, b \in \mathbb{R}, a < b\}$. Минималну алгебру која садржи колекцију $\mathcal{K}$ означавамо са $\mathcal{B}_0$. Минимална σ-алгебра која садржи колекцију $\mathcal{K}$ зове се Борелова σ-алгебра подскупова реалне праве. Означавамо је са $\mathcal{B}$. Њени елементи зову се Борелови скупови на реалној правој.*

Дефиниција 2.1.6 *Нека је Ω простор исхода случајног експеримента и $\mathcal{F}$ σ-алгебра догађаја. Уређени пар $(\Omega, \mathcal{F})$ се зове мерљив простор.*

Дефиниција 2.1.7 *Нека је Ω скуп свих елементарних догађаја и нека је $\mathcal{F}$ σ-алгебра над Ω. Функција $P : \mathcal{F} \to \mathbb{R}$ се зове вероватноћа или вероватносна мера на мерљивом простору $(\Omega, \mathcal{F})$ ако има следећа својства:*

- *својство нормираности - $P(\Omega) = 1$*
- *својство ненегативности - $P(A) \geq 0$, за сваки догађај $A \in \mathcal{F}$*
- *својство σ адитивности вероватноће - ако $\{A_i\}_{i\in\mathbb{N}} \in \mathcal{F}$, $A_i \bigcap A_j = \emptyset$, $i \neq j$, $i = 1, 2, 3, ...$, $j = 1, 2, 3, ...$, онда $P\left(\bigcup\limits_{i=1}^{\infty} A_i\right) = \sum\limits_{i=1}^{\infty} P(A_i)$*

Простор вероватноћа је уређена тројка $(\Omega, \mathcal{F}, P)$.

Дефиниција 2.1.8 *Нека је $(\Omega, \mathcal{F}, P)$ простор вероватноћа. Реално пресликавање $X : \Omega \to \mathbb{R}$ се зове случајна променљива ако $\forall B \in \mathcal{B}$ важи да је $X^{-1}(B) \in \mathcal{F}$ где је $\mathcal{B}$ Борелова σ-алгебра. За X кажемо да је $\mathcal{F}$-мерљиво.*

Дефиниција 2.1.9 *Фамилија случајних променљивих $\{X_t(\omega) : t \in T, \omega \in \Omega\}$ дефинисаних над истим простором вероватноће $(\Omega, \mathcal{F}, P)$ се зове случајни (стохастички) процес са индексним скупом T.*

Приликом записа се често изоставља променљива ω, па се уместо тога случајни процес означава са $\{X_t, t \in T\}$. Процес очигледно зависи од две променљиве t и ω. За фиксирано $t \in T$ случајни процес је једна променљива. За фиксирано $\omega \in \Omega$ функција $X_t(\omega)$, $t \in T$ се зове *трајекторија* случајног процеса. Ако пак фиксирамо обе променљиве X_t је реалан број или једна реализација случајног процеса.

Скуп T из претходне дефиниције се још назива и *параметарски скуп*. Ако је $T = \mathbb{R} = (-\infty, \infty)$, $T = [0, \infty)$ или $T = [a, b] \subseteq \mathbb{R}$, онда се параметар $t \in T$ обично интерпретира као време, а случајна функција $\{X_t, t \in T\}$ зове се *случајан процес са непрекидним временом*. Ако је $T \subseteq \mathbb{Z}$ или $T \subseteq \mathbb{N}$, онда се случајна функција $\{X_t, t \in T\}$ зове *случајан процес са дискретним временом* или *случајни низ*.

Дефиниција 2.1.10 *За сваки природан број n и произвољне елементе $t_1,...,t_n \in T$, $t_1 < t_2 < ... < t_n$, расподела вероватноћа случајног вектора $(X_{t_1}, X_{t_2}, ..., X_{t_n})$ одређена је функцијом расподеле*

$$F_{t_1,...,t_n}(x_1, ..., x_n) = P\{X_{t_1} \leq x_1, ..., X_{t_n} \leq x_n\}.$$

Фамилија свих ових расподела зове се фамилија коначнодимензионих расподела случајног процеса $\{X_t, t \in T\}$.

Фамилија коначнодимензионих расподела задовољава следеће услове:

- **услов симетрије** - ако је $i_1, ..., i_n$ једна пермутација бројева од 1 до n онда важи
$$F_{t_{i_1},...,t_{i_n}}(x_{i_1}, ..., x_{i_n}) = F_{t_1,...,t_n}(x_1, ..., x_n)$$
- **услов сагласности** - ако је $m < n$ за произвољне $t_{m+1}, ..., t_n \in T$ важи
$$F_{t_1,...,t_m,t_{m+1},...,t_n}(x_1, ..., x_m, \infty, ..., \infty) = F_{t_1,...,t_m}(x_1, ..., x_m)$$

Нека је $\{X_t, t \in T\}$ случајан процес. Тада је:

- средња вредност случајног процеса
$$E(X_t) = \mu_t, t \in T$$
- варијанса случајног процеса
$$D(X_t) = \sigma_t^2, t \in T$$
- аутоковаријанса (или краће коваријанса) случајног процеса
$$\gamma(r, s) = cov(X_r, X_s) = E((X_r - E(X_r))(X_s - E(X_s))), r, s \in T$$
- аутокорелација (или краће корелација) случајног процеса
$$\rho(r, s) = \frac{cov(X_r, X_s)}{\sqrt{D(X_r)D(X_s)}}$$

Дефиниција 2.1.11 *Случајан процес је строго стационаран ако су његове коначнодимензионе функције расподеле инваријантне у односу на* t*, односно ако за све* $t_i, t_i + t \in T$*,* $i = 1, 2, ...$ *важи:*
$$F_{t_1,...,t_n}(x_1, ..., x_n) = F_{t_1+t,...,t_n+t}(x_1, ..., x_n).$$

Дефиниција 2.1.12 *Случајан процес* $\{X_t, t \in T\}$ *је слабо стационаран ако су испуњени следећи услови:*

- $E(X_t) = m = const$, $\forall t \in T$
- $E(X_t^2) < \infty$, $\forall t \in T$
- $\gamma(r, s) = \gamma(r + t, s + t)$, $\forall r, s, t \in T$.

Сада можемо увести и формалну дефиницију временских серија.

Дефиниција 2.1.13 *Временска серија је једна реализација случајног процеса.*

У смислу претходне дефиниције однос временске серије и случајног процеса одговара односу узорка и основног скупа у стандардној теорији статистичког закључивања. Као што узорак представља део основног скупа на основу кога се доносе закључци о самом скупу, тако и временска серија представља део случајног процеса који треба да омогући уочавање карактеристика тог процеса.

2.2 Компоненте временске серије

Временске серије се, у зависности од појаве, могу састојати од различитих компоненти. Традиционално се претпоставља да се временска серија може разложити на четири компоненте: тренд $f(t)$, цикличну компоненту $c(t)$, сезонску компоненту $s(t)$ и случајну компоненту $\varepsilon(t)$. У случају када серија садржи све ове компоненте типови модела који се најчешће користе су:

- **адитивни модел** - $X_t = f(t) + s(t) + c(t) + \varepsilon(t)$
- **мултипликативни модел** - $X_t = f(t) \cdot s(t) \cdot c(t) \cdot \varepsilon(t)$
- **мешовити модел** - $X_t = f(t) \cdot s(t) \cdot c(t) + \varepsilon(t)$.

Тренд временске серије описује њено основно понашање у дужем временском периоду, или прецизније у читавом временском периоду за који је временска серија позната. У зависности од тога да ли вредности временске серије током времена систематски расту или опадају, тренд може бити растући или опадајући. Осим тога, тренд може бити детерминистички или стохастички у зависности од тога да ли се промене серије могу предвидети или не.

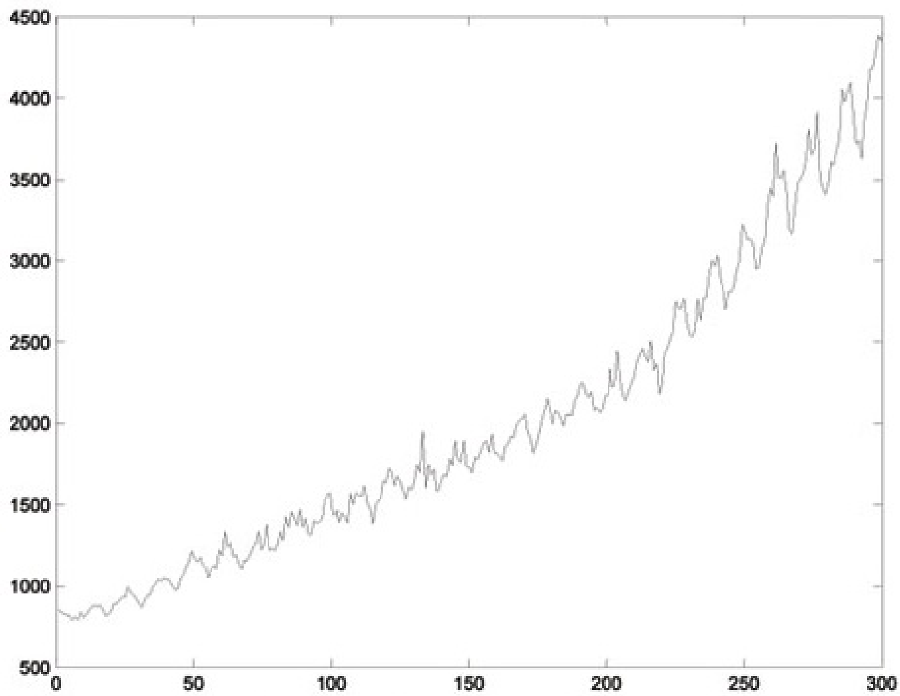

График 2.2.1: **Пример временске серије са трендом:** *Потрошња енергије у Југоисточном Бразилу од јануара 1976. до децембра 2000. (преузето са http : //www.scielo.br)*

У практичним применама већина временских серија представља податке прикупљене у узастопним временским интервалима као што су дани, месеци, квартали, итд. Због тога поједине временске серије испољавају правилности у свом кретању које се јављају у току једне календарске године. За такве серије кажемо да имају сезонску компоненту. Присуство сезонске компоненте утиче на то да постоји већи степен корелације између података добијених у истим месецима различитих година него између оних добијених у суседним месецима исте године.

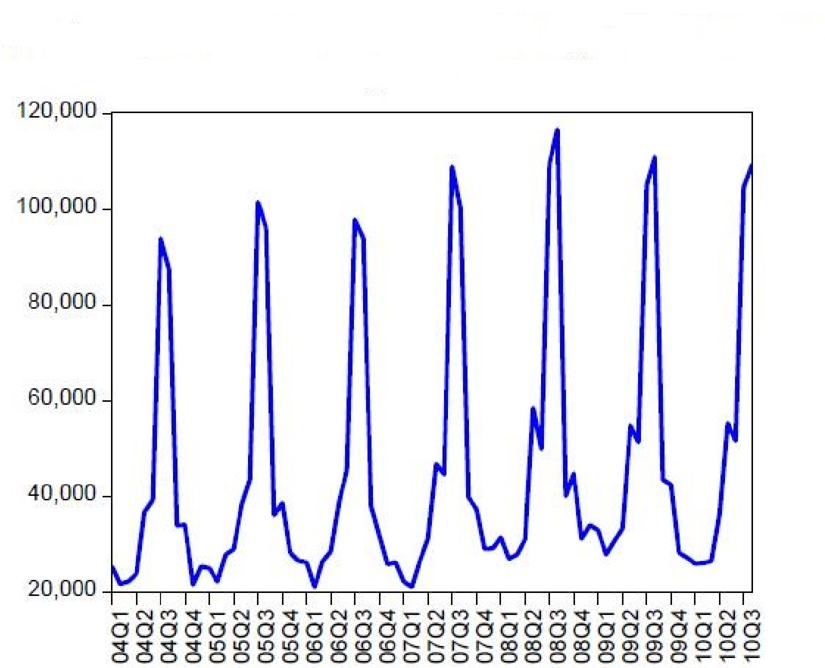

График 2.2.2: **Пример временске серије са сезонском компонентом:** *Број туриста у Републици Македонији од 2004. до 2010. (преузето са http : //mk.wikipedia.org/wiki)*

Циклична компонента представља слично понашање које се јавља у серији током неког периода времена, односно можемо рећи да се у серији запажа периодични карактер промене. Овакве серије имају своју периоду, односно циклус.

Претходне три компоненте се једним именом називају неслучајне компоненте. Ниједна од ових компоненти није обавезна у моделу. Једина компонента коју сваки модел мора да има је случајна компонента. Ова компонента служи за објашњавање непредвидивих флуктуација.

2.3 Класификација временских серија

Можемо извршити класификацију временских серија користећи различите критеријуме. Најједноставнија је подела на непрекидне и дискретне (прекидне) временске серије. Као и код случајних процеса та подела зависи од параметарског скупа. Временска серија је непрекидна ако њене вредности можемо регистровати у сваком временском тренутку. Прекидна временска серија је она чије вредности бележимо у једнаким временским интервалима (дан, сат, месец, година). Прекидну временску серију можемо добити и од непрекидне тако што ћемо бележење резултата посматране појаве вршити само у одређеним временским интервалима.

Кључна подела временских серија јесте подела на стационарне и нестационарне. Слободно речено, серија је стационарна уколико је њено кретање предвидиво током времена, тј. уколико њено кретање испуњава сличан образац понашања током времена. Супротно, ако су параметри кретања временске серије функције временског тренутка, тада је она нестационарна. На следећим сликама се може видети како изгледају стационарне а како нестационарне временске серије.

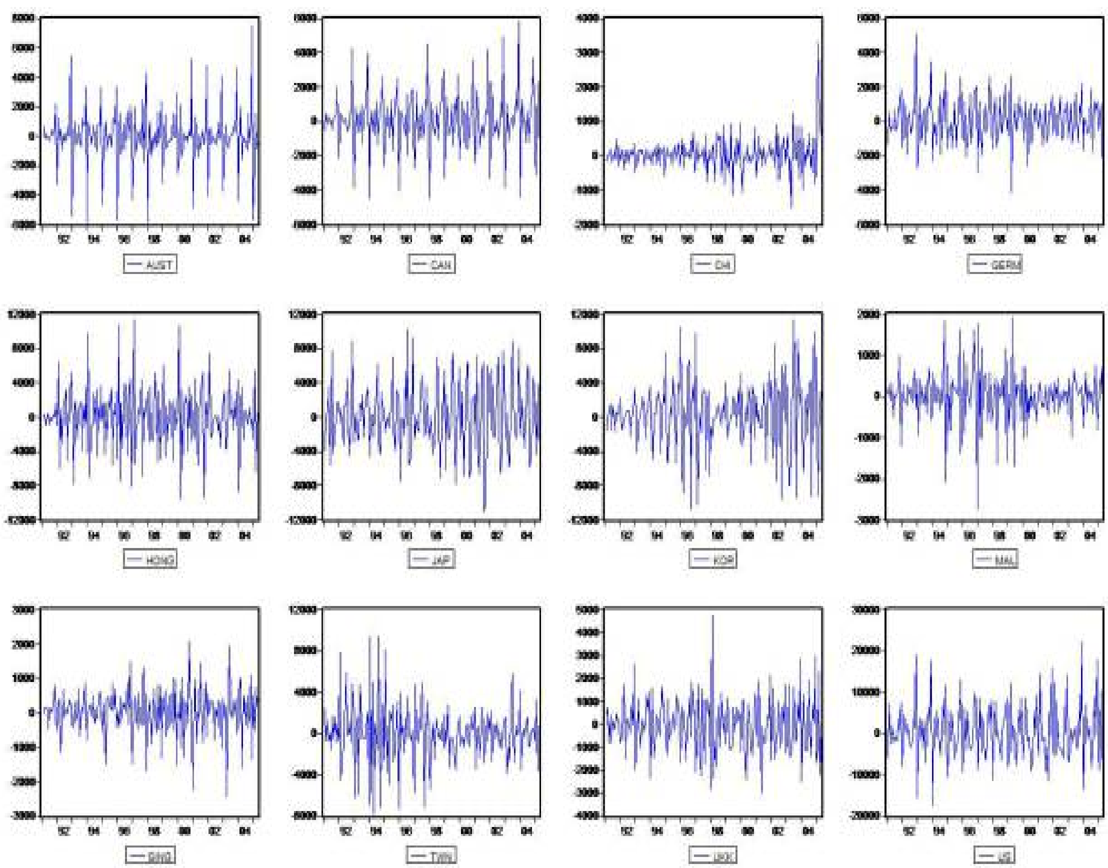

График 2.3.1: **Пример стационарних временских серија**
(преузето са *https://drsifu.wordpress.com/*2012/11/27/*time − series − econometrics*)

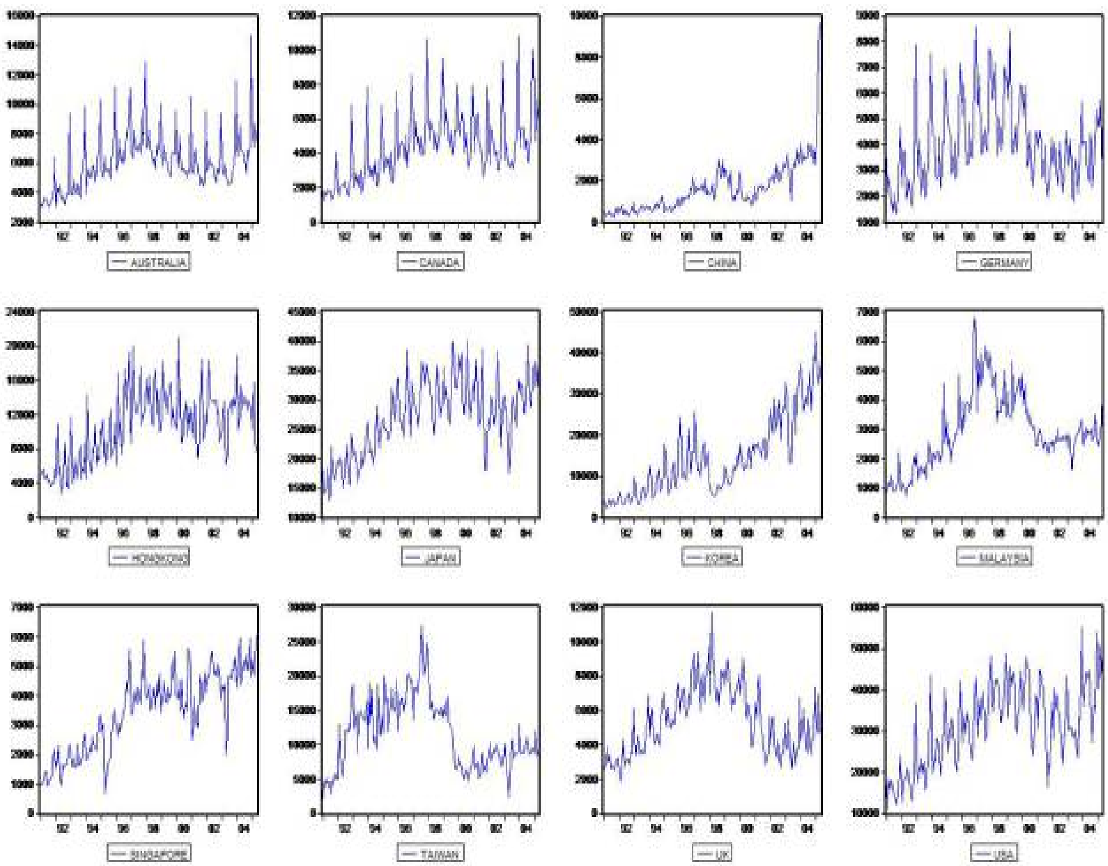

График 2.3.2: **Пример нестационарних временских серија**
(преузето са *https://drsifu.wordpress.com/2012/11/27/time-series-econometrics/*)

Као што је већ поменуто постоје два вида стационарности. У случају дефиниције 2.1.11 серија је стационарна ако се њена својства не мењају транслирањем у времену. То значи да случајне променљиве које припадају тој серији имају исту очекивану вредност, варијансу и моменте вишег реда. Ако пак узмемо дефиницију 2.1.12 у обзир онда за стационарну серију важи да се очекивана вредност и варијанса не мењају током времена док је коваријанса између свака два члана серије само функција временског растојања између њих.

Већина временских серија уопште, а посебно у економији има бар једну особину којом одступа од карактеристичног изгледа стационарне временске серије. Нестационарна временска серија се може трансформисати тако да се добије стационарна временска серија помоћу оператора прве диференце, односно коначних разлика првог реда, за шта се уобичајено, мада неправилно, користи термин диференцирање података.

Претпоставимо да временска серија X_t поседује тренд детерминистичког типа $(a+bt)$ око којег релативно правилно осцилира, што зависи од понашања случајне грешке $e_t : X_t = a+bt+e_t$. Применом оператора прве диференце елиминише се компонента линеарног тренда на следећи начин:

$$\left.\begin{array}{l} X_t = a + bt + e_t \\ X_{t-1} = a + b(t-1) + e_{t-1} \end{array}\right\} \Rightarrow \Delta X_t = X_t - X_{t-1} = b + \Delta e_t.$$

2.4 Неки модели временских серија

Један од начина за изучавање неких појава које су добијене у виду временских серија је изградња модела за такве серије. Основна мотивација за изградњу модела била је жеља да се прогнозирају будуће вредности серије. У наставку ће бити описани неки од основних модела. Сви модели у наставку се односе на временске серије без тренда, сезонске и цикличне компоненте.

2.4.1 Бели шум

Нека је дат низ случајних величина: $X_1, X_2, X_3, ..., X_n$ за које важи:

$$X_t = f(t) + \xi_t, \quad t = \overline{1,n}$$

где је $f(t)$ детерминистичка, неслучајна компонента, а ξ_t, $t = \overline{1,n}$ случајна компонента за коју важи:

$$\begin{aligned} &E\xi_t = 0, \quad t = \overline{1,n} \\ &D\xi_t = \sigma^2, \quad t = \overline{1,n} \\ &Cov(\xi_t, \xi_s) = E\xi_t\xi_s = 0, \quad za \quad sve \quad s \neq t. \end{aligned}$$

Низ ξ_t се назива бели шум. Да је неки низ бели шум означаваћемо са: $\xi_t \perp (0, \sigma^2)$.

2.4.2 Модели покретних просека

Нека је дат низ реалних случајних променљивих:

$$\{\xi_t, t \in D\},\ D = \{0, \pm1, \pm2, ...\}$$

који представља бели шум:

$$\xi_t \perp (0, \sigma^2).$$

Дефинишимо сада низ реалних случајних променљивих $\{X_t, t \in D\}$ на следећи начин:

$$X_t = \sum_{j=0}^{q} a_j \xi_{t-j},\ a_0 \neq 0,\ a_q \neq 0.$$

Ово је серија покретних просека реда q или $MA(q)$ серија (енглески: MA-moving average).

За $MA(q)$ серије важи:

$$\begin{aligned} &EX_t = 0 \\ &DX_t = E\left(\sum_1^q a_j\xi_{t-j}\right)^2 = \sigma^2 \sum_{j=0}^{q} a_j^2 \\ &K_X(t,s) = K(t-s) = K(\tau) = \begin{cases} \sigma^2 \sum\limits_{j=0}^{q-|\tau|} a_j a_{j+|\tau|}, & 0 \leq |\tau| \leq q \\ 0, & |\tau| > q \end{cases} \end{aligned}$$

Одакле закључујемо да је $MA(q)$ стационарна серија.

Најједноставнији пример серије покретних просека је $MA(1)$ серија. Ова серија има следећи облик:

$$X_t = \xi_t + a\xi_{t-1}.$$

2.4.3 Ауторегресиони модели

Нека је дат низ реалних случајних променљивих:

$$\{\xi_t, t \in D\},\ D = \{0, \pm 1, \pm 2, ...\}$$

који представља бели шум и нека је серија $\{X_t\}$ дефинисана на следећи начин:

$$X_t = h_1 X_{t-1} + h_2 X_{t-2} + ... + h_p X_{t-p} + \xi_t,\ t \in D,$$

где су $h_1, h_2, ..., h_p$ реални коефицијенти. Такве серије ћемо записивати и у облику:

$$X_t + b_1 X_{t-1} + ... + b_p X_{t-p} = \xi_t$$

и сматрати да може бити $p \geq 1$. Ово је ауторегресиона серија реда p или $AR(p)$ серија (енглески: AR-autoregression).

За све ауторегресионе серије важи да се могу представити у облику бесконачне серије покретних просека. Тако серију $AR(p)$ можемо представити и на следећи начин:

$$X_t = \sum_{j=0}^{\infty} c_j \xi_{t-j}, \quad t \in D.$$

За ову серију онда важи:

$$\begin{aligned}
&EX_t = \lim_{n\to\infty} EX_{t,n} = \lim_{n\to\infty} \sum_{j=0}^{n} c_j E\xi_{t-j} = 0 \\
&DX_t = \lim_{n\to\infty} EX_{t,n}^2 = \lim_{n\to\infty} \sum_{j=0}^{n} |c_j|^2 E\xi_{t-j}^2 = \sigma^2 \sum_{j=0}^{\infty} cj^2 \\
&K_X(t, t-\tau) = E \sum_{j=0}^{\infty} c_j \xi_{t-j} \cdot \sum_{v=0}^{\infty} c_v \xi_{t-\tau-v} = \\
&= \sum_{j,v=0}^{\infty} c_j c_v E(\xi_{t-j}\xi_{t-\tau-v}) = \sigma^2 \sum_{v=0}^{\infty} c_v c_{\tau+v}, \quad \tau > 0 \\
&K(t, t-\tau) = K(\tau) = K(-\tau).
\end{aligned}$$

Одавде закључујемо да параметри c_j морају бити мањи од 1 по апсолутној вредности како би серија била стационарна, јер је само у том случају сума $\sum_{j=1}^{\infty} c_j^2$ коначна, односно само тад је $DX_t < \infty$.

Најједноставнији пример ауторегресионе серије је $AR(1)$ серија. Ова серија има следећи облик:

$$X_t = \beta X_{t-1} + \xi_t.$$

2.4.4 Ауторегресиони модели покретних просека

Серија $\{X_t, t \in D\}$ реалних случајних величина дефинисана са:

$$\sum_{j=0}^{p} b_j X_{t-j} = \sum_{k=0}^{q} a_k \xi_{t-k},\ t \in D,$$

у којој ћемо сматрати да је $b_0 = 1, b_p \neq 0, a_0 = 1, a_q \neq 0$ и где је:

$$\xi_t \perp (0, \sigma^2),\ t \in D,$$

је ауторегресиона серија покретних просека реда (p, q) или $ARMA(p, q)$ серија.

Најједноставнији облик ових серија је $ARMA(1, 1)$ серија. Ова серија има следећи облик:

$$X_t + b_1 X_{t-1} = \xi_t + a_1 \xi_{t-1}.$$

3 Временске серије са јединичним кореном

Ауторегресионом моделу реда p:

$$X_t = h_1 X_{t-1} + h_2 X_{t-2} + ... + h_p X_{t-p} + \xi_t$$

може се придружити следећа једначина:

$$g^p - h_1 g^{p-1} - h_2 g^{p-2} - ... - h_p = 0.$$

Оваква једначина се обично назива карактеристична једначина $AR(p)$ модела, где су $g_1, g_2, ..., g_p$ корени карактеристичне једначине. Може се показати да важи следеће:

- Уколико су сви корени $g_1, g_2, ..., g_p$ по модулу мањи од 1, онда је серија стационарна
- Уколико је бар један корен $g_i,\ i = 1, 2, ..., p$ по модулу једнак један онда је временска серија нестационарна. Такве временске серије се обично називају временске серије са јединичним кореном.

Посматрајмо следећи модел:

$$X_t = \beta + 1 \cdot X_{t-1} + e_t.$$

Ово је ауторегресиони модел првог реда код кога је ауторегресиони параметар једнак вредности 1. Понашање ове временске серије на дуги рок одређује решење следеће карактеристичне једначине:

$$g - 1 = 0 \Rightarrow g = 1.$$

Дакле корен карактеристичне једначине узима вредност 1, због чега се овакве серије и називају временске серије са јединичним кореном. Број јединичних корена одговара степену интегрисаности временске серије, односно одређује колико је пута потребно диференцирати нестационарну временску серију да би она постала стационарна. Претпоставимо да је број јединичних корена d тада је серија $\Delta^d X_t$ стационарна временска серија.

Можемо дати још један пример временске серије са јединичним кореном. Посматрајмо модел облика:

$$X_t = 2X_{t-1} - X_{t-2} + e_t.$$

Ово је ауторегресиони модел реда 2. Карактеристична једначина овог модела је:

$$g^2 - 2g + 1 = 0 \Leftrightarrow (g-1)^2 = 0 \Rightarrow g_1 = g_2 = 1.$$

Дакле, овај модел има два јединична корена, што значи да је потребно два пута извршити диференцирање да би се добила стационарна временска серија. То се може урадити на следећи начин:

$$\begin{aligned} &X_t = 2X_{t-1} - X_{t-2} + e_t \\ &\underbrace{X_t - X_{t-1}}_{\Delta X_t} = \underbrace{X_{t-1} - X_{t-2}}_{\Delta X_{t-1}} + e_t \end{aligned} \quad \Rightarrow \quad \underbrace{\Delta X_t = \Delta X_{t-1}}_{\Delta X_t - \Delta X_{t-1} = \Delta^2 X_t} + e_t \Rightarrow \quad \Delta^2 X_t = e_t$$

Првим диференцирањем се елиминише први јединични корен али добијена серија $\Delta X_t = \Delta X_{t-1} + e_t$ је и даље нестационарна. Другим диференцирањем се елеминише и други јединични корен и добија се стационарна временска серија $\Delta^2 X_t = e_t$.

3.1 Тестови јединичног корена

У практичном моделирању од великог је значаја утврђивање броја јединичних корена које дата временска серија поседује. У том циљу се користе тестови јединичног корена.

Постоји велики број различитих тестова јединичних корена. Први тест који се појавио дефинисали су Дики и Фулер. Поред овог у раду ће бити описана још два теста јединичног корена.

3.1.1 Дики-Фулеров тест

Овај тест је добио име по својим ауторима. Један је од најчешће коришћених и често се краће назива ДФ тест.

Претпоставимо да је временска серија X_t генерисана $AR(1)$ моделом:

$$X_t = \phi X_{t-1} + e_t$$

и да је њена средња вредност $E(X_t) = 0$. Као што је већ објашњено у претходном поглављу, природа кретања X_t зависи од вредности параметра ϕ. Када је $\phi < 1$ X_t следи стационарну путању, а када је $\phi = 1$ X_t поседује јединични корен.

Према овоме се формирају хипотезе теста које имају следећи облик:

$$H_0 : \phi = 1 \text{ (}X_t \text{ поседује један јединични корен)}$$

$$H_1 : \phi < 1 \text{ (}X_t \text{ је стационарна временска серија).}$$

Нулту хипотезу проверавамо помоћу тест статистике која се рачуна према подацима из узорка обима T: $X_1, X_2, ..., X_T$. Први корак је оцењивање параметра ϕ применом методе најмањих квадрата:

$$\hat{\phi} = \frac{\sum\limits_{t=2}^{T} X_t X_{t-1}}{\sum\limits_{t=2}^{T} X_{t-1}^2}.$$

Следећи корак је израчунавање стандардне грешке оцене $\hat{\phi}$, $s(\hat{\phi})$, на следећи начин:

$$s(\hat{\phi}) = \sqrt{\frac{s^2}{\sum\limits_{t=2}^{T} X_{t-1}^2}},$$

где је:

$$s^2 = \frac{\sum\limits_{t=2}^{T} X_t^2 - \hat{\phi} \sum\limits_{t=2}^{T} X_t X_{t-1}}{(T-1)-1}.$$

Уколико се оцена $\hat{\phi}$ не разликује много од један логично је закључити да је нулта хипотеза тачна. Према овоме можемо образовати тест статистику на следећи начин:

$$\frac{\hat{\phi}-1}{s(\hat{\phi})}.$$

Од тога да ли серија има јединични корен или не зависи и расподела ове тест статистике. Наиме, под претпоставком да је $\hat{\phi}$ оцена параметра стандардног линеарног регресионог модела ова тест статистика би имала t-расподелу. Насупрот томе, ако серија има јединични корен, тада оцена $\hat{\phi}$ нема нормалну расподелу, па ни тест статистика нема t-расподелу.

Да тест статистика не припада класи познатих теоријских расподела, тј. да има нестандардну расподелу, први су показали Дики и Фулер. Према тесту ова расподела је названа расподела ДФ теста јединичног корена. Често се означава и као τ-расподела, па се и сама тест статистика често означава са τ.

Критична вредност τ^k за τ-расподелу се може одредити за сваки обим узорка за ниво значајности 5% помоћу следеће формуле:

$$\tau^k = -1.9393 - \frac{0.398}{\tau}.$$

Према познатим правилима статистичких тестова затим се одређује да ли се нулта хипотеза прихвата или одбија. Наиме, ако је израчуната вредност тест-статистике већа од критичне вредности хипотеза H_0 се одбацује и закључујемо да је серија стационарна. У супротном нулту хипотезу прихватамо и закључујемо да серија има бар један јединични корен.

Уколико се уместо почетног модела користи модел:

$$\Delta X_t = \beta_0 + \beta t + \varphi X_{t-1} + \delta_1 \Delta X_{t-1} + \delta_2 \Delta X_{t-2} + ... + \delta_k \Delta X_{t-k} + e_t$$

израчуната ДФ статистика из овог модела назива се проширена Дики-Фулерова статистика која се обично означава **ADF (Augmented Dickey-Fuller).**Обе статистике, ДФ и АДФ, поседују идентичну асимптотску расподелу и у пракси се користе исте критичне вредности.

3.1.2 КПСС тест

КПСС тест јединичног корена име је добио по почетним словима презимена својих аутора **(Kwiatkowski–Phillips–Schmidt–Shin).** Заснива се на потпуно другачијем приступу од ДФ теста. Посматрајмо модел:

$$X_t = X_0 + \beta t + e_t, \quad t = 1, 2, ..., T \tag{3.1.1}$$

и претпоставимо да e_t поседује један јединични корен:

$$e_t = e_{t-1} + v_t, \quad e_0 = 0 \tag{3.1.2}$$

тако да је сада v_t процес бели шум са средњом вредношћу 0 и варијансом σ_v^2.

Ако сада заменимо 3.1.2 у 3.1.1 добијамо:

$$X_t = X_0 + \beta t + e_t = X_0 + \beta t + e_{t-1} + v_t = X_0 + \beta t + e_{t-2} + v_t + v_{t-1} = ... = X_0 + \beta t + \sum_{j=1}^{t} v_j,$$

где је $t = 1, 2, ...$

Оно од чега ће зависити да ли је серија стационарна или не јесте последњи сабирак, сума случајних грешака $\sum_{j=1}^{t} v_j$. Серија ће бити стационарна једино онда када је варијанса случајне величине v_t једнака нули.

Из ових закључака формирамо хипотезе:

$$H_0 : \sigma_v^2 = 0 \text{ } (X_t \text{ је стационарна временска серија})$$

$$H_1 : \sigma_v^2 > 0 \text{ } (X_t \text{ поседује један јединични корен}).$$

Дакле основни задатак овог теста је заправо оцена варијансе случајне величине v_t.

Када из релације 3.1.1 применом методе најмањих квадрата оценимо параметар β добијамо прво резидуале $\hat{e}_t$, а потом и парцијалне суме резидуала - $S_t = \sum_{i=1}^{t} \hat{e}_i$, за $t = 1, 2, ..., T$. Затим се формира величина $\sum_{t=1}^{T} S_t^2$, која се нормира оценом дугорочне варијансе s_∞^2 случајне компоненте $\sum_{j=1}^{t} v_j$. Одавде изводимо КПСС тест-статистику која је облика:

$$KPSS = \frac{\frac{1}{T^2} \sum_{t=1}^{T} S_t^2}{s_\infty^2}.$$

Величина s_∞^2 представља Њуи-Вестову оцену дугорочне регресије:

$$\sigma_\infty^2 = \lim_{T \to \infty} T^{-1} var(\sum_{t=1}^{T} v_t),$$

која се добија према:

$$\begin{aligned} s_\infty^2 &= \frac{1}{T} \sum_{t=1}^{T} \hat{e}_t^2 + 2 \sum_{j=1}^{L} w_j (\frac{1}{T} \sum_{t=j+1}^{T} \hat{e}_t \hat{e}_{t-j}) \\ &= \frac{\sum_{t=1}^{T} \hat{e}_t^2}{T} \left(1 + 2w_1 \frac{\sum_{t=2}^{T} \hat{e}_t \hat{e}_{t-1}}{\sum_{t=1}^{T} \hat{e}_t^2} + ... + 2w_L \frac{\sum_{t=L+1}^{T} \hat{e}_t \hat{e}_{t-L}}{\sum_{t=1}^{T} \hat{e}_t^2} \right). \end{aligned}$$

Реч је о уобичајеној оцени варијансе случајне грешке модела $s^2 = \sum_{t=1}^{T} \hat{e}_t^2/T$, која се множи датим елементом у загради да би се укључила информација о аутокорелисаности података. Параметар L представља број доцњи и бира се на следећи начин : $L = [4(T/100)^{1/4}]$. Величина w_j представља прозор доцњи и дефинише се као $w_j = 1 - j/(L+1), \quad j = 1, ..., L$ и 0 за остале доцње. На овај начин се претпоставља да аутокорелација линеарно опада током времена и да не постоји за доцње строго веће од L.

3.1.3 ПП тест

ПП **(Phillips–Perron)** тест је дефинисан за исте хипотезе као и код ДФ теста. Тест статистика је постављена тако да представља модификацију ДФ статистике са циљем повећања поузданости тестирања када у регресионом моделу постоји аутокорелација и када је потребно водити рачуна о присуству детерминистичке компоненте. ПП тест статистика се дефинише на следећи начин:

$$Z_t = \frac{s}{s_\infty}\tau_t - 0.5\left(\frac{s_\infty^2 - s^2}{s_\infty^2}\right)\left(\frac{Ts(\hat{\varphi})}{s^2}\right).$$

Са s^2 и s_∞^2 означене су редом обична и дугорочна оцена варијансе случајне грешке модела, док $s(\hat{\varphi})$ представља стандардну грешку оцене $\hat{\varphi}$ у моделу:

$$\Delta X_t = \beta_0 + \beta t + \varphi X_{t-1} + e_t,$$

а τ_T ДФ статистику за исти модел.

У условима важења нулте хипотезе статистика Z_t има расподелу која је асимптотски идентична расподели ДФ теста па се користи исти скуп критичних вредности.

4 Регресиона анализа временских серија са јединичним кореном

4.1 Линеарна регресија

У свакој научној дисциплини основни проблем је утврђивање веза између променљивих величина. Те везе могу бити потпуно одређене. На пример, у физици се може утврдити тачна функционална зависност између удаљености објеката од земље и гравитационе силе.

Међутим, у биолошким и друштвеним наукама суочавамо се са много компликованијим ситуацијама. Овде је много теже одредити тачну повезаност па је потребно користити статистичка изучавања која мере просечне промене једне величине изазване променама друге величине. Регресиона анализа управо има за циљ да утврђује и мери везе таквог типа.

У медицинским истраживањима често се интерес истраживача усмерава ка проблему повезаности међу случајним променљивим. Притом је од посебног интереса могућност прогнозирања или предвиђања вредности једне променљиве на основу других променљивих.

Први тако формулисани проблем потиче од енглеског антрополога Франциса Галтона. Он је заједно са Пирсоном студирао наслеђивање у биологији. Током студија је мерећи висине очева и синова установио неку врсту парадокса. Наиме, приметио је да високи очеви имају високе синове али у просеку не тако високе као што су они сами, и слично, да ниски очеви имају ниске синове али опет у просеку не тако ниске као што су они. Ову тенденцију просека неке карактеристике (у овом случају висине) одабране групе да у следећој генерацији синова тежи ка просеку популације а не просеку њихових очева, Галтон је назвао регресијом, тачније, регресијом према просеку. Да би добио информацију зависности висине синова од висине њихових очева, Пирсон је претпоставио да се та зависност може изразити као функција, $y = f(x)$, при чему је y зависна променљива, односно променљива коју желимо да објаснимо или предвидимо (у Галтоновом примеру висина синова), а x независна променљива, односно променљива коју користимо да објаснимо зависну променљиву (висина очева).

Основна сврха примене регресионе анализе је да се на основу једне, познате променљиве може предвидети вредност друге, непознате променљиве и то из једначине која показује њихову зависност. Зависност може бити линеарна и у том случају се изражава полиномом првог степена или може бити нелинеарна (логаритамска, експоненцијална), па се изражава одговарајућим математичким изразима.

4.1.1 Линеарни регресиони модели

Регресиони модели се деле на просте, у којима постоји само једна зависна и једна независна променљива и вишеструке у којима постоји две или више независних и једна зависна променљива. И једни и други се према природи зависности између променљивих деле на линеарне и нелинеарне.

Прости линеарни регресиони модели се изражавају полиномом првог степена, која за популацију гласи:

$$y = \beta_0 + \beta_1 x + \epsilon.$$

У овој једначини су: y - зависна променљива, x – независна променљива, а ϵ – случајна грешка, која представља разлику између експериментално добијене вредности y и вредности y израчунате из полинома првог степена.

За узорак једначина има следећи облик:

$$y_i = a + bx_i$$

где су a и b непознати параметри које треба оценити. Константа a представља вредност y када је $x = 0$ (вредност у којој права линија сече y осу када је $x = 0$) и назива се одсечак на y оси, а служи за процену популационог параметра β_0. Коефицијент регресије b показује колико се линеарно мења вредност зависне променљиве y ако се независна променљива x промени (повећа или смањи) за јединицу мере. Коефицијент b има позитиван предзнак када се са повећањем x повећава y (зависност између x и y је управо пропорционална), а негативан када се са повећањем x смањује y (обрнуто пропорционална зависност). Ако су обе променљиве изражене у истим димензијама, b представља тангенс угла α који права линија заклапа са x осом.

Метода која се користи да се одреде параметри модела на основу датих тачака назива се метода најмањих квадрата, а базира се на томе да се минимизује сума квадрираних разлика (грешака $= \epsilon$) између добијених и израчунатих вредности y. Суштина је у томе да кроз групу тачака може да се повуче више правих линија које ће на неки начин описивати тај скуп тачака, а "најбоља" је она код које је збир квадрата одступања тачака од одговарајуће праве најмањи могући. Такву праву називамо регресиона права.

Права линија која се према критеријуму најмањих квадрата најбоље уклапа у групу тачака назива се регресиона линија, а једначина која је дефинише назива се регресиона једначина. Регресиона једначина има следеће особине:

- разлика између стварне вредности y и израчунате вредности y је најмања могућа,
- из средње вредности x можемо да израчунамо средњу вредност y,
- када x одступа од средње вредности, можемо да очекујемо и да y одступа од своје средње вредности.

Из регресионе једначине може да се израчуна очекивана вредност y из дате вредности x, односно регресиона једначина може да се користи за "предвиђање" вредности y.

Коефицијенти регресионе једначине израчунавају се према следећим изразима:

$$b = \frac{\sum xy - N\bar{x}\bar{y}}{\sum x^2 - N\bar{x}^2}, \quad a = \bar{y} - b\bar{x}.$$

Величину одступања тачака од праве линије изражава стандардна грешка регресионе праве, која се израчунава према изразу:

$$S_{y,x} = \sqrt{\frac{\sum (y_i - y)^2}{N-2}} = \sqrt{\frac{\sum y^2 - a\sum y - b\sum xy}{N-2}}.$$

Јака линеарна зависност између промењљивих x и y значи да су тачке врло близу праве линије, па је стога сума квадрата одступања добијених вредности y од израчунатих вредности y, мала, односно мала је и стандардна грешка $S_{y,x}$. И обрнуто, када је линеарна зависност између промењљивих x и y слаба, тачке су расуте око праве, сума квадрата одступања тачака од праве је велика и велика је стандардна грешка $S_{y,x}$.

Разлика између добијених вредности y и израчунатих вредности y назива се остатак (резидуал), а пошто је стандардна грешка $S_{y,x}$ мера за величину ових резидуала, њен други назив је резидуална стандардна девијација.

4.2 Линеарна регресија код временских серија са јединичним кореном

Једна од претпоставки класичног линеарног регресионог модела односи се на својство објашњавајућих променљивих да узимају фиксиране вредности из поновљених узорака. Варијанса објашњавајућих променљивих је нула јер оне и нису случајне променљиве. У случају да је поменута претпоставка нетачна не може се доказати да се методом најмањих квадрата добијају конзистентне оцене. Осим тога те оцене немају нормалну расподелу па се не могу користити ни стандардне тест-статистике са Студентовом и χ^2 расподелом.

Ако објашњавајућа променљива има јединични корен, није стационарна, онда њена варијанса расте током времена и таква променљива не узима фиксиране вредности из поновљених узорака. Одавде следи да се класични линеарни регресиони модели не могу користити при изучавању међусобне зависности нестационарних временских серија.

Један од начина оцењивања зависности између нестационарних временских серија је њихова трансформација тако да се добију стационарне временске серије. У ту сврху се најчешће користи већ поменуто диференцирање временских серија. Нека су Y_t и X_t временске серије са једним јединичним кореном онда се класични метод најмањих квадрата може применити на модел облика:

$$\Delta Y_t = \beta_0 + \beta_1 \Delta X_t + \epsilon_t.$$

Оно што представља проблем код оваквог оцењивања је то што резултат не даје везу између Y_t и X_t која у економским истраживањима може да буде од великог значаја. Ова веза се код стационарних временских серија добија једноставно из линеарног регресионог модела, док се у случају серија са јединичним кореном она може добити искључиво на основу концепта коинтеграције који ће бити објашњен у наставку рада.

5 Коинтеграција

Претпоставимо да постоје две временске серије, од којих свака поседује свој стохастички тренд, али се њихово кретање одвија тако да се оне не удаљавају много једна од друге. У извесном смислу је њихово заједничко кретање стационарно. Управо из овог примера потиче појам коинтеграције, који слободно речено подразумева стационарност линеарне комбинације индивидуално нестационарних временских серија.

Упоредимо сада концепт коинтеграције са концептом линеарног регресионог модела. Циљ регресионе анализе је да зависну променљиву представи преко објашњавајућих променљивих тако да необјашњени део буде стационаран, односно процес бели шум. Циљ коинтеграције је да кретање једне временске серије буде објашњено кретањем неких других временских серија тако да необјашњени део буде стационаран. Дакле по својој суштини ова два концепта се не разликују.

5.1 Модел са корекцијом равнотежне грешке

У анализи односа економских временских серија значајно место заузима модел корекције равнотежне грешке. Овај модел представили су Енгле и Грејнџер као начин описивања коинтегрисане временске серије.

Претпоставимо да су дате две временске серије X_t, Y_t које су у релацији $Y_t = \beta_0 + \beta X_t$. Модел са корекцијом равнотежне грешке је облика:

$$\Delta Y_t = \gamma_0(Y_{t-1} - \beta_0 - \beta X_{t-1}) + \gamma_{11}\Delta Y_{t-1} + ... + \gamma_{1k}\Delta Y_{t-k} + \gamma_{21}\Delta X_{t-1} + ... + \gamma_{2k}\Delta X_{t-k} + \delta,$$

где су $\gamma_0, \gamma_{11}, ..., \gamma_{1k}, \gamma_{21}, ..., \gamma_{2k}$ параметри, при чему важи $\gamma_0 < 0$, а δ случајна грешка. Кључни део модела је променљива $(Y_{t-1} - \beta_0 - \beta X_{t-1})$ јер она представља одступање Y_t од равнотежне релације $Y_t = \beta_0 + \beta X_t$ до кога је дошло у периоду $t-1$. Поменута променљива се зато назива равнотежна грешка. У наставку модела се промена Y_t, ΔY_t, током периода t моделира на основу информације колика равнотежна грешка је направљена у периоду $t-1$. Равнотежна грешка има корективну улогу.

Параметар γ_0 се назива параметар прилагођавања јер показује колики део промене Y_t се усклађује у сваком периоду према путањи дугорочне равнотежне везе.

Променљиве $\Delta Y_{t-1}, \Delta Y_{t-2}, ..., \Delta Y_{t-k}, \Delta X_{t-1}, \Delta X_{t-2}, ..., \Delta X_{t-k}$ се користе за описивање такозване краткорочне динамике. Број ових променљивих зависи од структуре корелације почетних временских серија.

Између модела са корекцијом равнотежне грешке и коинтеграције постоји јака веза па важи да ако су временске серије коинтегрисане онда се увек могу представити помоћу овог модела, али и обрнуто, односно ако се једна од временских серија може описати помоћу модела са корекцијом равнотежне грешке онда су оне коинтегрисане [2]. Ова веза представља фундаментални резултат Грејнџер-Јохансенове теореме репрензетације. Према овој теореми модел са корекцијом равнотежне грешке садржи информације о нивоу временских серија, чак и ако су оне нестационарне.

5.2 Тест коинтеграције

Тест коинтеграције су осмислили Енгл и Грејнџер и он је познат под именом двостепена процедура Енгла и Грејнџера. Претпоставка овог теста је да временске серије образују само једну равнотежну релацију. Ова претпоставка је тачна ако су у питању две временске серије, у супротном не мора да важи. Ако се ради о више коинтегрисаних временских серија оне могу образовати и неколико равнотежних релација. У том случају се поступак тестирања разликује.

Претпоставимо да желимо да проверимо да ли су временске серије X_t и Y_t коинтегрисане. Ове временске серије су коинтегрисане ако необјашњени део кретања Y_t представља стационарну компоненту. Та компонента је серија резидуала коју ћемо означавати са r_t. Дакле ако је r_t стационарно временске серије X_t и Y_t су коинтегрисане, док ако је r_t нестационарно ове временске серије нису коинтегрисане. Одавде закључујемо да се тест коинтеграције своди на тест јединичног корена серије резидуала r_t.

Тестирање можемо извршити помоћу ДФ теста јединичног корена где су нам хипотезе дате са:

$$H_0 : r_t \text{ поседује јединични корен}$$

$$H_1 : r_t \text{ је стационарна временска серија.}$$

Оваква верзија ДФ теста се назива ДФ тест серије резидуала и изводи се на исти начин као и ДФ тест, али има једну битну разлику а то је другачија асимптотска расподела. Ова расподела зависи од тога да ли коинтеграциона једначина садржи као детерминистичке компоненте само константу или и константу и тренд, као и од броја објашњавајућих променљивих у полазној једначини.

Ако тест доведе до закључка да су серије коинтегрисане анализа се наставља оцењивањем параметара модела равнотежне грешке. Оцене се могу добити стандардном методом најмањих квадрата и тако добијене оцене су пристрасне и немају нормалну расподелу. Одузимањем стварних вредности и вредности добијених из оцењеног модела добија се серија резидуала.

Коинтегрисане временске серије X_t и Y_t образују дугорочну равнотежну везу облика $Y_t = \beta_0 + \beta X_t + r_t$, где је r_t стационарна серија резидуала.

6 Векторска временска серија

Дефиниција 6.0.1 *Дефинишемо векторску временску серију X_t као:*

$$X_t = \begin{bmatrix} x_{1t} \\ x_{2t} \\ \vdots \\ x_{mt} \end{bmatrix}$$

за $t = 1, 2, ...,$ где су $x_{1t}, x_{2t}, ..., x_{mt}$ временске серије.

Дефиниција 6.0.2 *Векторска временска серија X_t је слабо стационарна ако задовољава следеће услове:*

1. $E(X_t) = \mu = const$, *где је* $\mu = \begin{bmatrix} E(x_{1t}) \\ E(x_{2t}) \\ \vdots \\ E(x_{mt}) \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_m \end{bmatrix}$
2. $cov(x_{it}, x_{js}) = g(t-s),\ i, j = 1, 2, ..., m,\ s = 1, 2, ...$

Дефиниција 6.0.3 *Коваријациона матрица за доцњу k је:*

$$\Gamma_k = E(X_t - \mu)(X_{t-k} - \mu)\prime = \begin{bmatrix} \gamma_{11}(k) & \gamma_{12}(k) & ... & \gamma_{1m}(k) \\ \gamma_{21}(k) & \gamma_{22}(k) & ... & \gamma_{2m}(k) \\ \vdots & & & \\ \gamma_{m1}(k) & \gamma_{m2}(k) & ... & \gamma_{mm}(k) \end{bmatrix}$$

где је:

$$\gamma_{ij}(k) = E(x_{it} - \mu_i)(x_{jt-k} - \mu_j), \quad i, j = 1, 2, ..., m, \quad k = 0, 1, 2, ...$$

Дефиниција 6.0.4 *Означимо са $D = diag[\gamma_{11}(0), \gamma_{22}(0), ..., \gamma_{mm}(0)]$ дијагоналну матрицу чији су елементи варијансе чланова векторске временске серије X_t. Матрица аутокорелационих коефицијената на доцњи k је онда дефиннисана на следећи начин:*

$$\rho_k = D^{-1/2}\Gamma_k D^{-1/2'}.$$

Ако означимо:

$$\rho_k = \begin{bmatrix} \rho_{11}(k) & \rho_{12}(k) & ... & \rho_{1m}(k) \\ \rho_{21}(k) & \rho_{22}(k) & ... & \rho_{2m}(k) \\ \vdots & & & \\ \rho_{m1}(k) & \rho_{m2}(k) & ... & \rho_{mm}(k) \end{bmatrix}$$

онда је:

$$\rho_{ij}(k) = \frac{\gamma_{ij}(k)}{\sqrt{\gamma_{ii}(0)\gamma_{jj}(0)}}.$$

Дефиниција 6.0.5 *Нека је дат узорак обима T векторске временске серије X_t. Оцена $\hat{\Gamma}_k$ коваријационе матрице Γ_k је:*

$$\hat{\Gamma}_k = \frac{1}{T} \sum_{t=k+1}^{T} (X_t - \overline{X})(X_{t-k} - \overline{X}), \quad k = 0, 1, 2, ...$$

где је:

$$\overline{X} = \frac{1}{T} \sum_{t=1}^{T} X_t.$$

Дефиниција 6.0.6 *Оцена $\hat{\rho}_k$ матрице ρ_k је:*

$$\hat{\rho}_k = \hat{D}^{-1/2} \hat{\Gamma}_k \hat{D}^{-1/2'},$$

где је $\hat{D}$ матрица димензије $m \times m$ чији елементи на главној дијагонали представљају оцене варијанси чланова векторске временске серије X_t.

6.1 ВАР модел

У моделовању векторских временских серија најзаступљенији је векторски ауторегресиони модел или скраћено ВАР модел.

Дефиниција 6.1.1 *ВАР модел векторске временске серије X_t, димензије m и реда k има следећи облик:*

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + ... + \phi_k X_{t-k} + \epsilon_t.$$

У претходном моделу параметри $\phi_1, \phi_2, ..., \phi_k$ су матрице димензије $m \times m$. Случајна грешка модела је ϵ_t који је векторски случајни процес бели шум димензије $m \times 1$ и који представља некорелисане временске серије чија је средња вредност 0, а варијанса коначна. Важи следеће:

$$E(\epsilon_t' \epsilon_{t-k}) = \begin{cases} \Sigma, & k = 0 \\ 0, & k \neq 0 \end{cases}$$

где је са Σ означена коваријациона матрица процеса.

Значајност ВАР модела огледа се у чињеници да њега можемо користити када се бавимо анализом коинтегрисаних временских серија са јединичним кореном, а да их претходно не трансформишемо применом оператора диференцирања.

6.2 Коинтеграција у ВАР моделу

Посматрајмо ВАР(2) модел за две временске серије x_t и y_t са јединичним кореном:

$$\begin{bmatrix} x_t \\ y_t \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} x_{t-1} \\ y_{t-1} \end{bmatrix} + \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix} \begin{bmatrix} x_{t-2} \\ y_{t-2} \end{bmatrix} + \begin{bmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{bmatrix}$$

Ако од обе стране једнакости одузмемо $\begin{bmatrix} x_{t-1} \\ y_{t-1} \end{bmatrix}$ и потом користимо релацију $\begin{bmatrix} \Delta x_{t-1} \\ \Delta y_{t-1} \end{bmatrix} = \begin{bmatrix} x_{t-1} \\ y_{t-1} \end{bmatrix} - \begin{bmatrix} x_{t-2} \\ y_{t-2} \end{bmatrix}$ тада добијамо векторску форму модела са корекцијом равнотежне грешке:

$$\begin{bmatrix} \Delta x_t \\ \Delta y_t \end{bmatrix} = \begin{bmatrix} a_{11}+b_{11}-1 & a_{12}+b_{12} \\ a_{21}+b_{21} & a_{22}+b_{22}-1 \end{bmatrix} \begin{bmatrix} x_{t-1} \\ y_{t-1} \end{bmatrix} - \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix} \begin{bmatrix} \Delta x_{t-1} \\ \Delta y_{t-1} \end{bmatrix} + \begin{bmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{bmatrix},$$

односно:

$$\Delta X_t = \Pi X_{t-1} + \Gamma \Delta X_{t-1} + \epsilon_t,$$

где је $\Pi = \Phi_1 + \Phi_2 - I$, а $\Gamma = -\Phi_2$.

Према претпоставци свака од компоненти вектора X_t, x_t и y_t има по један јединични корен. У претходној једначини на левој страни се налази стационарна репрезентација ових временских серија (ΔX_t), док се на десној страни налази њихова нестационарна форма (X_{t-1}), као и стационарна са доцњом првог реда (ΔX_{t-1}). Постоје два решења проблема тако да случајна грешка и даље буде бели шум.

У првом случају ранг матрице Π је нула. Тада се једначина своди на ВАР модел првих диференци.

У другом случају ранг матрице Π једнак је један и матрица се може представити на следећи начин $\Pi = \gamma\beta'$, где су γ и β вектори параметара димензија 2×1, тако да је линеарна комбинација $\beta' X_t = \beta'[x_t y_t]'$ стационарна, односно да су компоненте x_t и y_t коинтегрисане.

Дакле, да би постојала корелација између ΔX_t и X_{t-1} неопходно је неутралисати нестационарност у вектору X_{t-1}. То се може урадити коинтеграцијом компоненти овог вектора.

Посматрајмо ВАР модел реда k и димензије m:

$$X_t = \Phi_1 X_{t-1} + \Phi_2 X_{t-2} + ... + \Phi_k X_{t-k} + \epsilon_t$$

њему одговара већ уведена једначина облика:

$$|I_m - \Phi_1 h - \Phi_2 h^2 - ... - \Phi_k h^k| = 0.$$

Ако су сва решења ове једначине једнака један онда су компоненте временске серије интегрисане реда један.

Векторска форма модела са корекцијом равнотежне грешке дефинише се као:

$$\Delta X_t = \Pi X_{t-1} + \Gamma_1 \Delta X_{t-1} + ... + \Gamma_{k-1} \Delta X_{t-k+1} + \epsilon_t,$$

где је $\Pi = \Phi_1 + \Phi_2 + ... + \Phi_k - I_m$, $\Gamma_j = -\sum\limits_{i=j+1}^{k} \Phi_i$, $j = 1, ..., k-1$.

Како су компоненте у ВАР моделу интегрисане реда један следи да су сва решења претходне једначине једнака један, односно да је вредност детерминанте матрице Π једнака нули. Матрица Π је сингуларна:

$$|I_m - \Phi_1 h - \Phi_2 h^2 - ... - \Phi_k h^k| = 0, \quad h_1 = ... = h_{km} = 1 \Rightarrow$$

$$|I_m - \Phi_1 - ... - \Phi_k| = 0 \Rightarrow |\Pi| = 0.$$

Када је матрица Π сингуларна могућа су два закључка.

1. Матрица Π је нула матрица. У том случају векторски модел са корекцијом равнотежне грешке своди се на ВАР модел првих диференци реда $k-1$ и временске серије у вектору X_t нису коинтегрисане.

2. Ранг матрице Π је мањи од m и већи од нуле. Нека је ранг једнак r, тада је $\Pi = \gamma\beta'$, где су γ и β матрице параметара димензије $m \times r$. У овом случају временске серије у вектору X_t су коинтегрисане.

6.2.1 Тестирање коинтеграције у ВАР моделу и оцене коинтеграционих параметара у ВАР моделу

Најчешће коришћен метод оцењивања коинтеграционих параметара и тестирања коинтеграције у ВАР моделу је такозвана Јохансенова процедура која се заснива на методи максималне веродостојности.

Претпоставимо да су компоненте векторске временске серије димензије m коинтегрисане. Модел са корекцијом равнотежне грешке тада има следећи облик:

$$\Delta X_t = \Pi X_{t-1} + \Gamma_1 \Delta X_{t-1} + ... + \Gamma_{k-1} \Delta X_{t-k+1} + \epsilon_t$$

где је $\Pi = \Phi_1 + \Phi_2 + ... + \Phi_k - I$, $\Gamma_j = -\sum\limits_{i=j+1}^{k} \Phi_i$, $j = 1, ..., k-1$, а ранг коинтеграције је $0 < r < m$, тако да важи $\Pi = \gamma\beta'$.

Како бисмо добили оцене параметара γ и β елиминишемо параметре краткорочне динамике $\Gamma_1, \Gamma_2, ..., \Gamma_{k-1}$. То се остварује на следећи начин:

- оцени се зависност ΔX_t од $\Delta X_{t-1}, \Delta X_{t-2}, ..., \Delta X_{t-k+1}$ применом метода најмањих квадрата и добија се вектор резидуала R_{0t}

- оцени се зависност X_{t-k} од $\Delta X_{t-1}, \Delta X_{t-2}, ..., \Delta X_{t-k+1}$ применом метода најмањих квадрата и добија се вектор резидуала R_{1t}

Сада се полазни модел своди на:

$$R_{0t} = \gamma\beta' R_{1t} + \epsilon_t$$

Оцене параметара модела добијају се на следећи начин:

- параметри матрице γ оцењују се методом најмањих квадрата уз претпоставку да је матрица β позната
- добијена оцена $\hat{\gamma}$ користи се у одговарајућој функцији веродостојности како би се добила оцена за β
- према оцени $\hat{\beta}$ добија се нова и коначна оцена $\hat{\gamma}$.

Важи да је:

$$L_{max}^{-\frac{2}{T}}(\beta) = |\hat{\Sigma}(\beta)| + const$$

као и:

$$\begin{aligned}\hat{\Sigma} &= T^{-1}\sum_{t=1}^{T}(R_{0t} - \gamma\beta' R_{1t})(R_{0t} - \gamma\beta' R_{1t})' \\ &= T^{-1}\left(\sum_{t=1}^{T} R_{0t}R_{0t}' - \sum_{t=1}^{T} R_{0t}R_{1t}'\beta\gamma' - \gamma\beta'\sum_{t=1}^{T} R_{1t}R_{0t}' + \gamma\beta'\sum_{t=1}^{T} R_{1t}R_{1t}'\beta\gamma'\right) \\ &= S_{00} - S_{01}\beta\gamma' - \gamma\beta' S_{10} + \gamma\beta' S_{11}\beta\gamma'\end{aligned}$$

где је:

$$S_{ij} = T^{-1}\sum_{t=1}^{T} R_{it}R_{jt}',\ i,j = 0,1.$$

Дакле као први корак у оцењивању налазимо оцену $\hat{\gamma}$ помоћу методе најмањих квадрата. Њу добијамо као оцену нагиба из регресије у којој је R_{0t} вектор зависне променљиве, док $\beta' R_{1t}$ представља вектор објашњавајућих променљивих. Добија се следећа оцена:

$$\hat{\gamma} = \gamma(\beta) = S_{01}\beta(\beta' S_{11}\beta)^{-1}.$$

Заменом у израз за $\hat{\Sigma}$ добија се:

$$\begin{aligned}\hat{\Sigma}(\beta) = {} & S_{00} - S_{01}\beta(\beta' S_{11}\beta)^{-1}\beta' S_{10} - S_{01}\beta(\beta' S_{11}\beta)^{-1}\beta' S_{10} + \\ & + S_{01}\beta(\beta' S_{11}\beta)^{-1}\beta' S_{11}\beta(\beta' S_{11}\beta)^{-1}\beta' S_{10}\end{aligned}$$

односно:

$$\hat{\Sigma} = S_{00} - S_{01}\beta(\beta' S_{11}\beta)^{-1}\beta' S_{10}.$$

Сада на ред долази други корак у оцењивању у коме се оцењује матрица β применом методе максималне веродостојности.

Знамо да за несингуларне матрице A, B и H важи:

$$\begin{vmatrix} A & B \\ B' & H \end{vmatrix} = |A| \cdot |H - B'A^{-1}B| = |H| \cdot |A - BH^{-1}B'|.$$

Ако сада уведемо смене:

$$
\begin{gathered}
S_{00} = A \\
\beta' S_{11} \beta = H \\
S_{01} \beta = B
\end{gathered}
$$

добијамо:

$$|S_{00}| \cdot |\beta' S_{11} \beta - \beta' S_{10} S_{00}^{-1} S_{01} \beta| = |\beta' S_{11} \beta| \cdot |S_{00} - S_{01} \beta (\beta' S_{11} \beta)^{-1} \beta' S_{10}|.$$

Одавде добијамо следећу оцену:

$$|\hat{\Sigma}(\beta)| = |S_{00}| \frac{|\beta'(S_{11} - S_{10} S_{00}^{-1} S_{01}) \beta|}{|\beta' S_{11} \beta|}.$$

Нека су сада G и F симетричне, квадратне и позитивно дефинитне матрице димензије m. Посматрајмо следећу функцију матрице R:

$$f(R) = \frac{|R' G R|}{|R' F R|}.$$

Максимална вредност функције $f(R)$ добија се одређивањем m карактеристичних вредности и m карактеристичних вектора из:

$$|kF - G| = 0.$$

Ми тражимо вредност за коју:

$$\frac{|\beta'(S_{11} - S_{10} S_{00}^{-1} S_{01}) \beta|}{|\beta' S_{11} \beta|}$$

достиже најмању вредност. Ако уведемо смене:

$$
\begin{gathered}
G = S_{11} - S_{10} S_{00}^{-1} S_{01} \\
F = S_{11} \\
R = \beta
\end{gathered}
$$

у релацију за одређивање максималне вредности функције $f(R)$ добијамо:

$$|k S_{11} - (S_{11} - S_{10} S_{00}^{-1} S_{01})| = 0,$$

односно:

$$|(1-k) S_{11} - S_{10} S_{00}^{-1} S_{01}| = 0.$$

Ако сада уведемо смену $\lambda = 1 - k$ добијамо:

$$|\lambda S_{11} - S_{10} S_{00}^{-1} S_{01}| = 0.$$

Решавањем ове једначине добија се m карактеристичних вредности $\hat{\lambda}_1 > \hat{\lambda}_2 > ... > \hat{\lambda}_m \geq 0$. Одговарајући карактеристични вектори су $v_1, v_2, ..., v_m$.

Тражена оцена матрице β је:

$$\hat{\beta} = (v_1, v_2, ..., v_r)$$

при чему важи:

$$\hat{\beta}' S_{11} \hat{\beta} = I_r.$$

Дакле, сада остаје још да се на основу $\hat{\beta}$ добије коначна оцена за γ. Из:

$$\hat{\gamma} = S_{01}\beta(\beta' S_{11}\beta)^{-1}$$

и оцене $\hat{\beta}$ добија се:

$$\hat{\gamma} = S_{01}\hat{\beta}.$$

Максимална вредност функције веродостојности узорка се изводи према израчунатим карактеристичним вредностима на следећи начин:

$$L_{max}^{-\frac{2}{T}} = |S_{00}| \frac{|\beta'(S_{11} - S_{10}S_{00}^{-1}S_{01})\hat{\beta}|}{|\hat{\beta}' S_{11}\hat{\beta}|} = |S_{00}| \prod_{i=1}^{m} (1 - \hat{\lambda}_i).$$

Карактеристична вредност $\hat{\lambda}_i,\ i = 1, 2, ..., m$ је показатељ корелисаности између линеарне комбинације нивоа временских серија и диференце временских серија. Што је $\hat{\lambda}_i$ ближе један то је линеарна комбинација нестационарних временских серија у већој сагласности са стационарним делом модела. Овако јака корелација је могућа једино када линеарна комбинација нестационарних елемената постаје стационарна. У том случају су посматране временске серије коинтегрисане. У другом случају, ако је $\hat{\lambda}_i = 0$ онда су линеарна комбинација нестационарних елемената и одговарајући стационарни елементи слабо корелисани. У овом случају је линеарна комбинација нестационарних елемената и сама нестационарна, односно коинтегрисаност између посматраних временских серија не постоји.

Сада је потребно утврдити тачан статистички поступак којим ће се тестирати коинтегрисаност временских серија. Нека је r број стационарних линеарних комбинација, а $m - r$ број нестационарних комбинација. Као први корак у тестирању постављају се следеће хипотезе:

$$H_0\colon\ rang(\Pi) = r \text{ (постоји } r \text{ стационарних релација)}$$

$$H_1\colon\ rang(\Pi) = m \text{ (постоји } m \text{ стационарних релација).}$$

Ако претпоставимо да је тачна нулта хипотеза максимална вредност функције веродостојности се добија на основу:

$$L_{max}^{-\frac{2}{T}}[r] = |S_{00}|(1 - \hat{\lambda}_1) \times (1 - \hat{\lambda}_2) \times ... \times (1 - \hat{\lambda}_r).$$

Ако пак претпоставимо да је тачна алтернативна хипотеза максимална вредност функције веродостојности се добија на основу:

$$L_{max}^{-\frac{2}{T}}[m] = |S_{00}|(1 - \hat{\lambda}_1) \times (1 - \hat{\lambda}_2) \times ... \times (1 - \hat{\lambda}_r) \times (1 - \hat{\lambda}_{r+1}) \times ... \times (1 - \hat{\lambda}_m).$$

Као тест статистика користи се количник веродостојности:

$$\begin{aligned} -2\ln Q\left(\frac{H_0}{H_1}\right) &= T\ln\left[\frac{|S_{00}|(1-\hat{\lambda}_1)\times...\times(1-\hat{\lambda}_r)}{|S_{00}|(1-\hat{\lambda}_1)\times...\times(1-\hat{\lambda}_r)\times(1-\hat{\lambda}_{r+1})\times...\times(1-\hat{\lambda}_m)}\right] \\ &= -T\ln\left[(1-\hat{\lambda}_{r+1})\times...\times(1-\hat{\lambda}_m)\right] \\ &= -T\sum_{i=r+1}^{m}\ln(1-\hat{\lambda}_i) \end{aligned}.$$

Овако је дефинисана Јохансенова статистика трага за коју ћемо користити ознаку τ_{m-r}. Дакле:

$$\tau_{m-r} = -T\sum_{i=r+1}^{m}\ln(1-\hat{\lambda}_i).$$

Под условом да је нулта хипотеза тачна ова тест статистика нема χ^2 расподелу већ вишедимензионалну верзију расподеле ДФ теста јединичног корена која зависи од броја стохастичких трендова $m-r$ и детерминистичких компоненти. Асимптотска расподела не зависи од броја доцњи k, па је Јохансенову статистику могуће кориговати тако што се обим узорка T замени са бројем степени слободе сваке једначине ВАР модела, $T-km$. На тај начин добија се следећи облик тест статистике:

$$\tau_{m-r} = -(T-km)\sum_{i=r+1}^{m}\ln(1-\hat{\lambda}_i).$$

Нека су добијене вредности тест статистика: $\tau_m, \tau_{m-1}, ..., \tau_1$, а $K_m, K_{m-1}, ..., K_1$ одговарајуће критичне вредности. Тестирање ћемо почети формулисањем одговарајућих хипотеза:

H_0: $r=0$ (нема коинтеграције)

H_1: $r>0$ (постоји бар једна стационарна релација).

Важи следеће. Ако је $\tau_m < K_m$ онда се нулта хипотеза прихвата и закључујемо да нема коинтеграције. Ако је пак $\tau_m > K_m$ нулта хипотеза се одбацује и закључујемо да постоји најмање један коинтеграциони вектор. Да бисмо утврдили да ли је то тачно један или можда више коинтеграционих вектора потребно је да наставимо тестирање.

Сада дефинишемо следеће хипотезе:

H_0: $r=1$ (постоји једна стационарна комбинација)

H_1: $r>1$ (постоје најмање две стационарне комбинације).

Ако је $\tau_{m-1} < K_{m-1}$ прихватамо нулту хипотезу и завршавамо тестирање. Ако је пак $\tau_{m-1} > K_{m-1}$ тестирање се наставља све до прихватања нулте хипотезе када ће бити одређен тачан број коинтеграционих релација.

6.3 Детерминистичке компоненте у коинтегрисаном ВАР моделу

Векторски модел са корекцијом равнотежне грешке садржи и детерминистичку компоненту. У циљу поједностављивања излагања ова компонента је у претходним поглављима била изостављена. Слично као и у стандардном ВАР моделу и овде се детерминистичка компонента јавља у облику константе или константе и линеарног тренда.

Код коинтеграционих ВАР модела детерминистичке компоненте припадају скупу објашњавајућих променљивих. То важи и за сам модел али и за коинтеграционе релације. Због овога је неопходно објаснити могуће комбинације детерминистичких компоненти у коинтегрисаном ВАР моделу. Постоји 5 могућих комбинација према Јохансеновом приступу. Ове комбинације су дате у следећој табели [2]:

Детерминистичка компонента у:		
полазном вектору X_t	векторској форми модела са корекцијом равнотежне грешке (моделира се ΔX_t)	коинтеграционом вектору
1. Не постоји	Не постоји	Не постоји
2. Константа	Константа (са ограничењима)	Константа
3. Линеарни тренд	Константа (без ограничења)	Не постоји
4. Линеарни тренд	Константа (без ограничења) и линеарни тренд (са ограничењима)	Линеарни тренд
5. Квадратни тренд	Константа и линеарни тренд (без ограничења)	Не постоји

Свака од ових комбинација биће детаљније објашњена.

Комбинација 1:

У моделу са корекцијом равнотежне грешке, као и у коинтеграционим релацијама, не постоје детерминистичке компоненте. Модел има следећи облик:

$$\Delta X_t = \gamma\beta' X_{t-1} + \Gamma_1 \Delta X_{t-1} + ... + \Gamma_{k-1}\Delta X_{t-k+1} + \varepsilon_t.$$

Комбинација 2:

Модел са корекцијом равнотежне грешке садржи константу. У питању је m-димензиони вектор c. У коинтеграционим релацијама $\beta' X_{t-1}$ такође постоји константа.

Оно што је карактеристично је да је константа присутна у моделу са ограничењем припадања коинтеграционој релацији, односно она је у функцији γ па је $c = \gamma c_0$, где c_0 означава вектор константи у коинтеграционим релацијама. Укључивањем константе у коинтеграциону релацију добија се стационарност линеарне комбинације променљивих око нулте средње вредности. На овај начин коинтеграционом комбинацијом неутралишу се не само индивидуални стохастички трендови, већ и појединачни детерминистички трендови. Модел је облика:

$$\Delta X_t = \gamma(\beta' X_{t-1} + c_0) + \Gamma_1 \Delta X_{t-1} + ... + \Gamma_{k-1} \Delta X_{t-k+1} + \varepsilon_t.$$

Комбинација 3:

У моделу са корекцијом равнотежне грешке постоји константа, c. У коинтеграционим релацијама $\beta' X_{t-1}$ нема константе. Константа се у моделу јавља без ограничења. Коинтеграциони вектор показује да је линеарна комбинација стационарна око не-нулте средње вредности. Модел има следећи облик:

$$\Delta X_t = c + \gamma\beta' X_{t-1} + \Gamma_1 \Delta X_{t-1} + ... + \Gamma_{k-1} \Delta X_{t-k+1} + \varepsilon_t.$$

Комбинација 4:

Модел са корекцијом равнотежне грешке садржи константу, c, и линеарни тренд, $c_1 t$. У коинтеграционим релацијама $\beta' X_{t-1}$ постоји само линеарни тренд. У моделу је константа без ограничења припадања коинтеграционој релацији, док је линеарни тренд са ограничењем. Вектор c_1 је у функцији γ, па је $c_1 = \gamma c_{10}$, где је са c_{10} означен вектор параметара линеарног тренда у коинтеграционим релацијама. Присуством линераног тренда у моделу постиже се елиминација и детерминистичког и стохастичког тренда појединачних временских серија. Модел је облика:

$$\Delta X_t = c + \gamma(\beta' X_{t-1} + c_{10} t) + \Gamma_1 \Delta X_{t-1} + ... + \Gamma_{k-1} \Delta X_{t-k+1} + \varepsilon_t.$$

Комбинација 5:

У моделу са корекцијом равнотежне грешке постоје и константа и линеарни тренд. У коинтеграционим релацијама $\beta' X_{t-1}$ детерминистичке компоненте нису присутне. Обе детерминистичке компоненте су у моделу присутне без ограничења. Одсуство линеарног тренда у коинтеграционој релацији показује да временске серије формирају тренд-стационарну комбинацију. Модел има следећи облик:

$$\Delta X_t = c + \gamma\beta' X_{t-1} + \Gamma_1 \Delta X_{t-1} + ... + \Gamma_{k-1} \Delta X_{t-k+1} + \varepsilon_t.$$

Дакле, да резимирамо. У коинтеграционој релацији могу бити присутни константа или линеарни тренд као детерминистичке компоненте. Ако полазне временске серије немају изражен детерминистички раст, онда константу треба уврстити као детерминистичку компоненту у коинтеграциону релацију. Ако пак такав раст постоји у полазним серијама онда је константа свакако присутна у моделу са корекцијом равнотежне грешке као посебна објашњавајућа променљива, док коинтеграциона релација може да нема никакве детерминистичке компоненте или да се у њу укључи линеарни тренд у циљу моделирања нивоа променљивих кроз равнотежну релацију.

Код сваке од могућих наведених комбинација потребан је нови скуп критичних вредности за поступак примене Јохансенове статистике трага. Постоје већ састављене табеле које садрже ове критичне вредности.

6.4 Тестирање линеарних ограничења на коинтеграционе параметре

Понекад се у економским анализама јавља потреба да се провери да ли коинтеграциони параметар узима тачно одређену вредност. У таквим случајевима користи се тест количника веродостојности.

Нулта хипотеза у овом тесту је да сваки од r коинтеграционих вектора β_1, β_2,..., β_r следи s_i ограничења, $i = 1, 2, ..., r$, односно:

$$H_0 : \beta = (\beta_1, \beta_2, ...\beta_r) = (H_1\nu_1, H_2\nu_2, ..., H_r\nu_r).$$

У претходном изразу β је полазна коинтеграциона матрица димензија $m \times r$, $H_1, H_2, ..., H_r$ су матрице познатих парметара димензија $m \times s_i$ којима се дизајнирају ограничења и $\nu_1, \nu_2, ..., \nu_r$ су редом вектори са параметрима који се оцењују димензија $s_i \times 1$.

У наставку ће бити наведени кључни принципи идентификације коинтеграционог система:

1. Сваки коинтеграциони вектор треба да поседује јединствену форму. То значи да коинтеграциони вектор не може да се добије ни линеарном комбинацијом преосталих коинтеграционих вектора у датом простору, ни алгебарском трансформацијом било ког другог вектора.

2. У случају да је број коинтеграционих вектора већи од један неопходно је идентификовати коинтеграциони простор, односно неопходно је издвојити подскупове променљивих које су коинтегрисане. То се постиже према резултатима тестирања ограничења на поједине параметре при чему услов идентификованости мора бити задовољен.

3. Услов идентификованости система еквивалентан је услову идентификованости у систему симултаних једначина.

4. Коинтеграциони скуп је инваријантан на додатна укључивања нових променљивих. Њиховим укључивањем може се евентуално променити број коинтеграционих вектора.

Тестирање линеарних ограничења се може обавити и за параметре матрице прилагођавања γ. Претпоставимо да постоји један коинтеграциони вектор. У овом случају нулта хипотеза ће имати следећи облик:

$$H_0 : \gamma' = (\gamma_1, 0, ..., \gamma_m).$$

Овим се претпоставља да коинтеграциони вектор у моделу прве диференце није значајан за другу по реду од m временских серија. За такву временску серију кажемо да је слабо егзогена у односу на параметре коинтеграционе везе. Она је извор нестационарности у целом систему па је ово тестирање изузетно значајно у економској анализи јер омогућава препознавање оних променљивих чија динамика не зависи од равнотежне путање.

7 Примена коинтеграционе анализе

У овом поглављу биће дат пример коинтегрисаних временских серија. Сви тестови и анализа биће обављени у статистичком програму *R*.

Посматраћемо везу између гдп индекса, потрошње и залиха новца у Сједињеним Америчким Државама у периоду од 1959. до краја 2014. године. Подаци су преузети са сајта банке федералних резерви:

`https://research.stlouisfed.org/fred2/graph/`

Подаци су дати квартално. Гдп индекс, или код нас бдп-бруто домаћи производ, представља укупну производњу роба и услуга оствареној у националној економији, без обзира на власништво. У анализи ћемо користити природни логаритам потрошње.

За почетак проверимо да ли серије задовољавају услове, односно да ли су оне нестационарне док су серије њихових првих диференци стационарне. За ову проверу ћемо користити *PP* и *ADF* тест јединичног корена. Проверићемо прво гдп индекс:

```
> gdp<-read.xlsx("gdp.xlsx",sheetName="Sheet1")
> gdp<-ts(gdp,start=c(1947,1),end=c(2014,10),
+         frequency=3)
> pp.test(gdp)

        Phillips-Perron Unit Root Test

data:  gdp
Dickey-Fuller Z(alpha) = 1.782, Truncation lag parameter = 4,
p-value = 0.99
alternative hypothesis: stationary

Warning message:
In pp.test(gdp) : p-value greater than printed p-value
> adf.test(gdp)

        Augmented Dickey-Fuller Test

data:  gdp
Dickey-Fuller = 3.0273, Lag order = 5, p-value = 0.99
alternative hypothesis: stationary

Warning message:
In adf.test(gdp) : p-value greater than printed p-value
```

```
> pp.test(diff(gdp))

        Phillips-Perron Unit Root Test

data:  diff(gdp)
Dickey-Fuller Z(alpha) = -128.4288, Truncation lag parameter = 4,
p-value = 0.01
alternative hypothesis: stationary

Warning message:
In pp.test(diff(gdp)) : p-value smaller than printed p-value
> adf.test(diff(gdp))

        Augmented Dickey-Fuller Test

data:  diff(gdp)
Dickey-Fuller = -3.8537, Lag order = 5, p-value = 0.01743
alternative hypothesis: stationary
```

Као што видимо у прва два теста добијамо велике p-вредности теста, а како је алтернативна хипотеза у оба случаја стационарност, закључујемо да је серија нестационарна. У друга два теста, у којима тестирамо серију прве диференце, видимо да је p-вредност теста мања од 0.05 па можемо закључити да се нулта хипотеза одбацује, односно серија је стационарна.

Извршићемо исте провере и за потрошњу и залихе новца. Добијамо:

```
> data<-read.xlsx("consumption_q.xlsx",sheetName="Sheet1")
> cons<-ts(data,start=c(1947,1),end=c(2015,1),
+          frequency=3)
> cons<-log(cons)
>
> pp.test(cons)

        Phillips-Perron Unit Root Test

data:  cons
Dickey-Fuller Z(alpha) = -2.5973, Truncation lag parameter = 4,
p-value = 0.9511
alternative hypothesis: stationary
```

```
> adf.test(cons)

        Augmented Dickey-Fuller Test

data:  cons
Dickey-Fuller = -1.9142, Lag order = 5, p-value = 0.612
alternative hypothesis: stationary

>
> pp.test(diff(cons))

        Phillips-Perron Unit Root Test

data:  diff(cons)
Dickey-Fuller Z(alpha) = -238.4561, Truncation lag parameter = 4,
p-value = 0.01
alternative hypothesis: stationary

Warning message:
In pp.test(diff(cons)) : p-value smaller than printed p-value
> adf.test(diff(cons))

        Augmented Dickey-Fuller Test

data:  diff(cons)
Dickey-Fuller = -4.2296, Lag order = 5, p-value = 0.01
alternative hypothesis: stationary

Warning message:
In adf.test(diff(cons)) : p-value smaller than printed p-value


> data<-read.xlsx("money_q.xlsx",sheetName="Sheet1")
> stock<-ts(data,start=c(1959,1),end=c(2014,12),
+            frequency=3)
>
> pp.test(stock)

        Phillips-Perron Unit Root Test

data:  stock
Dickey-Fuller Z(alpha) = -7.8158, Truncation lag parameter = 4,
p-value = 0.6675
alternative hypothesis: stationary
```

```
> adf.test(stock)

        Augmented Dickey-Fuller Test

data:  stock
Dickey-Fuller = -2.4067, Lag order = 5, p-value = 0.4063
alternative hypothesis: stationary

>
> pp.test(diff(stock))

        Phillips-Perron Unit Root Test

data:  diff(stock)
Dickey-Fuller Z(alpha) = -113.1367, Truncation lag parameter = 4,
p-value = 0.01
alternative hypothesis: stationary

Warning message:
In pp.test(diff(stock)) : p-value smaller than printed p-value
> adf.test(diff(stock))

        Augmented Dickey-Fuller Test

data:  diff(stock)
Dickey-Fuller = -4.2642, Lag order = 5, p-value = 0.01
alternative hypothesis: stationary

Warning message:
In adf.test(diff(stock)) : p-value smaller than printed p-value
```

Дакле и ове серије задовољавају потребне услове. Сада можемо тестирати коинтегрисаност ових серија. За то ћемо користити функцију *ca.po* из R-а која извршава Филипс-Оуиларис-ов тест коинтеграције. Филипс и Оулиарис су предложили две тест статистике за тестирање коинтеграције временских серија. Прва се означава са P_u и назива тест статистика количника варијансе, док се друга означава са P_z и назива вишедимензиона траг статистика. Друга тест статистика је значајна и има предност јер резултати теста не зависе од нормализације коинтеграционе једначине. У оба случаја нулта хипотеза теста је да нема коинтеграције, а алтернативна хипотеза је да коинтеграција постоји. Овај тест спада у тестове који се заснивају на резидуалима.

За почетак ћемо спојити ове три серије у један објекат у R-у како би нам било једноставније за рад и направити график тог објекта да бисмо видели како серије заправо изгледају.

```
usacons<-window(cbind(stock,gdp,cons),start=c(1959,1),
end=c(2014,12),frequency=3)
plot(usacons)
```

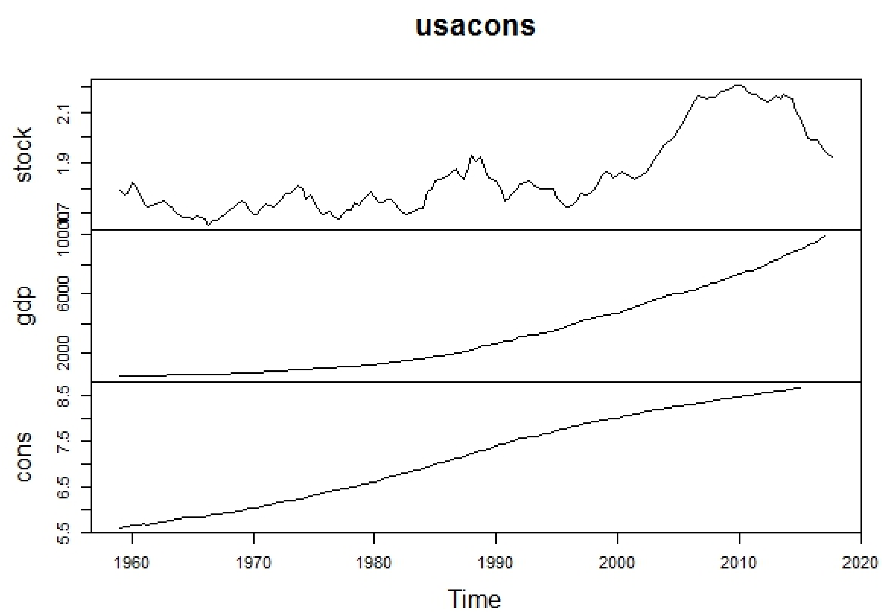

График 7.0.1: **Временске серије залиха новца, гдп индекса и потрошње у САД у периоду 1959-2014.**

Сада можемо извршити и сам тест:

```
> summary(ca.po(usacons,demean='const',type='Pu'))

########################################
# Phillips and Ouliaris Unit Root Test #
########################################

Test of type Pu
detrending of series with constant only


Call:
lm(formula = z[, 1] ~ z[, -1])

Residuals:
      Min        1Q    Median        3Q       Max
-0.147580 -0.045189 -0.009311  0.046518  0.166837
```

```
Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)    2.160e+00  9.796e-02  22.053  < 2e-16 ***
z[, -1]gdp     8.511e-05  6.268e-06  13.579  < 2e-16 ***
z[, -1]cons   -8.083e-02  1.625e-02  -4.974 1.62e-06 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 0.06707 on 166 degrees of freedom
Multiple R-squared:  0.8259,     Adjusted R-squared:  0.8238
F-statistic: 393.6 on 2 and 166 DF,  p-value: < 2.2e-16


Value of test-statistic is: 21.6962

Critical values of Pu are:
                   10pct    5pct    1pct
critical values 33.6955 40.5252 53.8731
```

Као што видимо p-вредност теста је изузетно мала па закључујемо да се нулта хипотеза одбацује, односно ове серију су коинтегрисане.

Можемо поновити тестирање али користећи другу тест статистику P_z, вишедимензиону статистику трага.

```
> summary(ca.po(usacons,demean='const',type='Pz'))

########################################
# Phillips and Ouliaris Unit Root Test #
########################################

Test of type Pz
detrending of series with constant only

Response stock :

Call:
lm(formula = stock ~ zr)
```

```
Residuals:
      Min         1Q     Median         3Q        Max
-0.062426  -0.012136   0.001768   0.013892   0.061918

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 1.775e-02  5.729e-02   0.310    0.757
zrstock     9.697e-01  2.267e-02  42.770   <2e-16 ***
zrgdp       1.178e-07  2.714e-06   0.043    0.965
zrcons      5.550e-03  5.118e-03   1.084    0.280
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 0.01928 on 164 degrees of freedom
Multiple R-squared:  0.9858,     Adjusted R-squared:  0.9855
F-statistic:  3786 on 3 and 164 DF,  p-value: < 2.2e-16


Response gdp :

Call:
lm(formula = gdp ~ zr)

Residuals:
    Min      1Q  Median      3Q     Max
-97.085  -9.587  -1.020   8.590  75.621

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -172.89432   70.06477  -2.468   0.0146 *
zrstock        1.07798   27.72784   0.039   0.9690
zrgdp          1.00177    0.00332 301.752  < 2e-16 ***
zrcons        30.38505    6.25990   4.854  2.8e-06 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 23.58 on 164 degrees of freedom
Multiple R-squared:  0.9999,     Adjusted R-squared:  0.9999
F-statistic: 6.61e+05 on 3 and 164 DF,  p-value: < 2.2e-16
```

```
Response cons :

Call:
lm(formula = cons ~ zr)

Residuals:
       Min         1Q     Median         3Q        Max
-0.0215300 -0.0041505  0.0000593  0.0041458  0.0153121

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -8.685e-02  1.972e-02  -4.405  1.9e-05 ***
zrstock      1.425e-02  7.802e-03   1.827   0.0695 .
zrgdp       -6.400e-06  9.341e-07  -6.852  1.4e-10 ***
zrcons       1.014e+00  1.761e-03 575.560  < 2e-16 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 0.006636 on 164 degrees of freedom
Multiple R-squared:      1,     Adjusted R-squared:      1
F-statistic: 1.232e+06 on 3 and 164 DF,  p-value: < 2.2e-16



Value of test-statistic is: 6.3699

Critical values of Pz are:
                  10pct    5pct     1pct
critical values 80.2034 89.7619 109.4525
```

Видимо да су оцењене све три коинтеграционе везе када се нормализација врши редом у односу на залихе новца, гдп индекс и потрошњу. У сва три случаја p-вредност теста је $< 2.2e - 16$. Како тест статистика упада у критичну област за праг значајности 0.1 и овај тест потврђује да коинтеграција постоји.

8 Статистичке таблице за тестове јединичног корена

Модел коришћен за генерисање података: $y_t = y_{t-1} + \epsilon_t$
1. Оцењени модел: Дики-Фулер $\delta y_t = a_0 + a_1 y_{t-1} + a_2 t + \epsilon_t$
Филипс-Перон $y_t = \tilde{a}_0 + \tilde{a}_1 \tilde{y}_{t-1} + \tilde{a}_2(t - n/2) + \tilde{\epsilon}_t$
Критичне вредности у тестирању хипотезе:

			10%	5%	1%
$a_1 = 0,$	$\tilde{a}_1 = 1$	$t - test$	-3.15	-3.45	-4.04
$a_0 = 0,$	$\tilde{a}_0 = 0$	$t - test$	2.73	3.11	3.78
$a_2 = 0,$	$\tilde{a}_2 = 0$	$t - test$	2.38	2.79	3.53
$a_1 = a_2 = 0,$	$\tilde{a}_1 = 1, \tilde{a}_2 = 0$	$F - test$	5.47	6.49	8.73
$a_0 = a_1 = a_2 = 0,$	$\tilde{a}_0 = \tilde{a}_2 = 0, \tilde{a}_1 = 1$	$F - test$	4.16	4.88	6.50

2. Оцењени модел: Дики-Фулер $\delta y_t = a_0 + a_1 y_{t-1} + \epsilon_t$
Филипс-Перон $y_t = \tilde{a}_0 + \tilde{a}_1 y_{t-1} + \tilde{\epsilon}_t$
Критичне вредности у тестирању хипотезе:

			10%	5%	1%
$a_1 = 0,$	$\tilde{a}_1 = 1$	$t - based$	-2.58	-2.89	-3.51
$a_0 = 0,$	$\tilde{a}_0 = 0$	$t - based$	2.17	2.54	3.22
$a_0 = a_1 = 0,$	$\tilde{a}_0 = 0, \tilde{a}_1 = 1$	$F - based$	3.86	4.71	6.70

3. Оцењени модел: Дики-Фулер $\delta y_t = a_1 y_{t-1} + \epsilon_t$
Филипс-Перон $y_t = \tilde{a}_1 y_{t-1} + \tilde{\epsilon}_t$
Критичне вредности у тестирању хипотезе:

			10%	5%	1%
$a1 = 0,$	$\tilde{a}_1 = 1$	$t - test$	-1.61	-1.95	-2.60

Биографија

Рођена сам 13.12.1990.године у Приштини. Преселила сам се у Ниш 1999. године, а затим у Чачак 2001.године, где сам завршила основну школу и природно-математички смер Гимназије.

Основне студије Математичког факултета у Београду, смер вероватноћа, статистика и актуарска математика, завршила сам 2013. године, са просеком 9.29. Након тога уписала сам мастер студије на истом смеру.

Литература

[1] J. Малишић и В. Јевремовић: *Случајни процеси и временске серије*, Математички факултет, Београд, 2008.

[2] З. Младеновић и А. Нојковић: *Примењена анализа временских серија*, Центар за издавачку делатност Економског факултета у Београду, 2012.

[3] Златко Ковачић: *Анализа временских серија*, Економски факултет, Београд, 1995.

[4] Јован Малишић: *Временске серије*, Математички факултет, Београд, 2002.

[5] Bernhard Pfaff: Analysis of Integrated Series with R and Cointegrated Time, Springer Science+Business Media, 2008.

[6] `http://a-little-book-of-r-for-time-series.readthedocs.org/-en/latest/src/timeseries.html`

[7] `http://cran.r-project.org/`

[8] `https://research.stlouisfed.org/fred2/graph/`