Univerza v Ljubljani

EKONOMSKA FAKULTETA

6. Preverjanje predpostavk klasičnega regresijskega modela

doc. dr. Miroslav Verbič

miroslav.verbic@ef.uni-lj.si

www.miroslav-verbic.si

EFMD EQUIS ACCREDITED

AACSB ACCREDITED

Ljubljana, februar 2014

Motivacija

Tri zlata pravila ekonometrije so:
testiraj, testiraj in še enkrat testiraj!

D. F. Hendry: "*Econometrics - Alchemy or Science*"
Economica, 1980

Greh je ne vedeti kdaj grešiš.
Nesmiselnemu grehu se je treba izogniti.

E. E. Leamer: "*Let's Take the Con Out of Econometrics*"
AER, 1983

Preverjanje predpostavke v splošnem

PREVERJANJE PREDPOSTAVK KLASIČNEGA LINEARNEGA REGRESIJSKEGA MODELA

A Kaj predpostavka pomeni in kaj so temeljne posledice njenega neizpolnjevanja?

B Kako ugotoviti oziroma odkriti veljavnost oziroma neveljavnost predpostavke?

C Kakšne so možne rešitve v primeru neizpolnjevanja predpostavke?

Univerza v Ljubljani

EKONOMSKA FAKULTETA

6.1 Normalna porazdelitev slučajne spremenljivke

EFMD EQUIS ACCREDITED

AACSB ACCREDITED

Pomen predpostavke

NORMALNA PORAZDELITEV SLUČAJNE SPREMENLJIVKE *u*

A **Kaj predpostavka pomeni in kaj so temeljne posledice njenega neizpolnjevanja?**

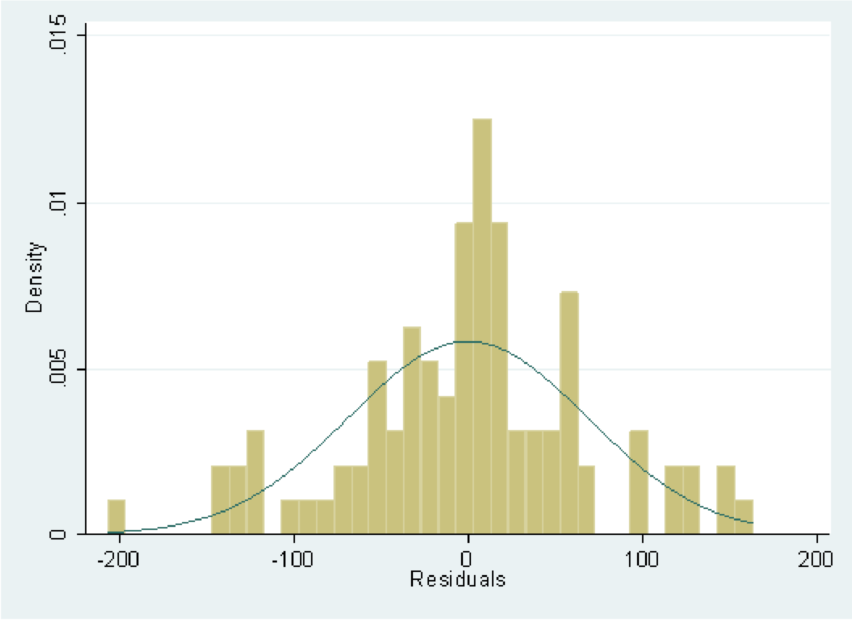
Predpostavka: $u \sim N(0, \sigma_u^2)$ in $y \sim N(\mathbf{X}\beta, \sigma_u^2)$

? **Predpostavka je bistvena za preverjanje domnev (od nje odvisni testi, ki temeljijo na t, F in χ^2 porazdelitvah)!**

Ugotavljanje veljavnosti predpostavke

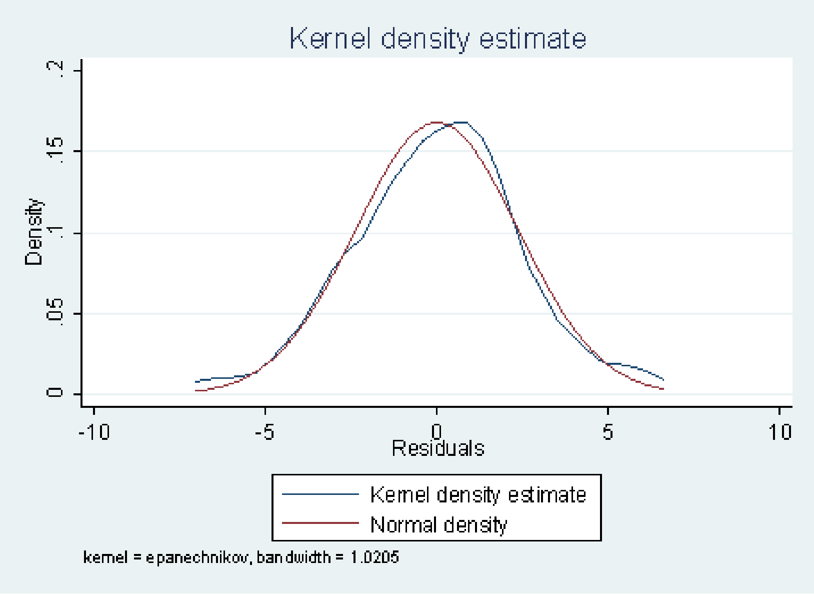
B **Kako ugotoviti oziroma odkriti veljavnost oziroma neveljavnost predpostavke?**

1 **Prikaz porazdelitve ostankov vzorčnega regresijskega modela - histogram ali poligon (standardiziranih) vrednosti**



Ugotavljanje veljavnosti predpostavke

2 **Prikaz porazdelitve ostankov vzorčnega regresijskega modela – diagram jedrne gostote (kernel density)**



3 **Shapiro – Wilkov test. Izračun je bil včasih računsko zahteven, saj smo potrebovali tabelirane vrednosti za preverjanje domnev (Biometrika, 1965).**

Ugotavljanje veljavnosti predpostavke

4 **Testi, ki temeljijo na preveri le asimetrije ali le sploščenosti porazdelitve.**

5 **Jarque – Bera test (1987)**

$$JB = n\left[\frac{S^2}{6} + \frac{(K-3)^2}{24}\right] \sim \chi^2_{(2)}$$

$$S^2 = \frac{\left(\frac{1}{n}\sum e^3\right)^2}{\left(\frac{1}{n}\sum e^2\right)^3} \qquad K = \frac{\frac{1}{n}\sum e^4}{\left(\frac{1}{n}\sum e^2\right)^2}$$

H_0: spremenljivka u je normalno porazdeljena

H_1: spremenljivka u ni normalno porazdeljena

Rešitve neizpolnjevanja predpostavke

C — Možne rešitve v primeru neizpolnjevanja predpostavke?

1 **Uporaba drugih cenilk regresijskih koeficientov - robustnih cenilk (robusticators ali robust regression). Zelo popularna je cenilka:**

$$\min \sum \left| y_i - \hat{y}_i \right|$$

Za cenilko srečamo tudi oznake MSAE, LAR, LAE in LAD.

2 **Transformacije spremenljivk: potenciranje ali logaritmiranje odvisne spremnljivke *y*.**

Univerza v Ljubljani

EKONOMSKA FAKULTETA

6.2 Multikolinearnost

EFMD EQUIS ACCREDITED

AACSB ACCREDITED

Temeljni pojmi

Besedo "multikolinearnost" je v ekonometrijo vpeljal Ragnar Frisch leta 1934, v smislu *popolne, perfektne linearne odvisnosti med pojasnjevalnimi spremenljivkami regresijskega modela.*

☹ **V splošnem pri preučevaju ekonomskih zakonitosti nimamo opraviti z opazovanji, ki bi bili rezultat nadzorovanih eksperimentov. Nimamo možnosti, da bi proces generiranja teh podatkov nadzorovali in tako izmerili vplive posameznih pojavov izolirano od drugih na določeno zakonitost.**

☹ **Ekonomski pojavi so zato v določeni medsebojni odvisnosti. Pričakovati moremo, da so do določene mere ekonomske spremenljivke medsebojno povezane – *kolinearne*, torej se bolj ali manj spreminjajo skupno, na medsebojno povezan način.**

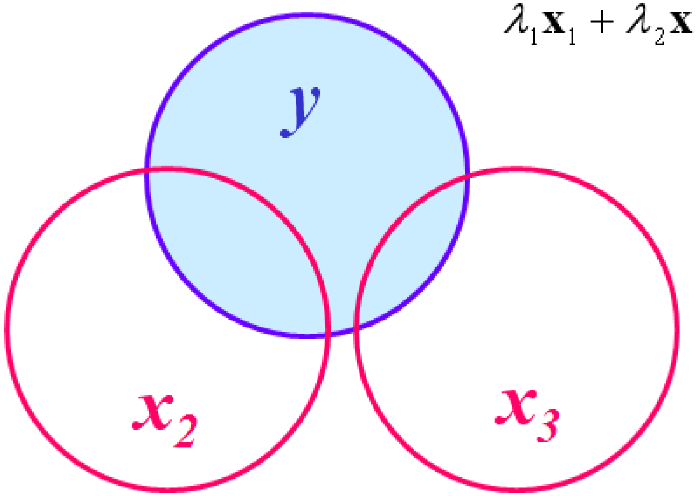
Pomen predpostavke

A **Kaj predpostavka pomeni in kaj so temeljne posledice njenega neizpolnjevanja?**

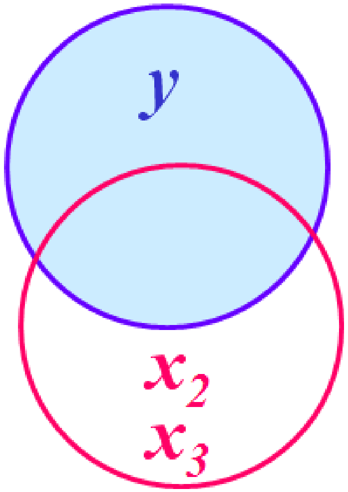
Predpostavka:

Med pojasnjevalnimi spremenljivkami regresijskega modela <u>ne obstaja</u> popolna linearna odvisnost v obliki:

$$\lambda_1 \mathbf{x}_1 + \lambda_2 \mathbf{x}_2 + \ldots + \lambda_k \mathbf{x}_k = 0$$



NI (multi)kolinearnosti



Popolna (multi)kolinearnost

Pomen predpostavke

I. Popolna (perfektna, natančna, eksaktna) multikolinearnost

$\mathbf{x}_2 - 3\mathbf{x}_3 = 0$ ⟹ popolna "kolinearnost": $\mathbf{x}_2 = 3\mathbf{x}_3$

$\mathbf{x}_2 + \mathbf{x}_3 + 2\mathbf{x}_4 = 0$ ⟹ popolna "multikolinearnost": $\mathbf{x}_2 = -\mathbf{x}_3 - 2\mathbf{x}_4$

1. Matrika $(X^T X)$ je singularna, ni možno izračunati njenega inverza.
2. $b = (X^T X)^{-1} X^T y$ ni definiran.
3. Multikolinearnost vključuje le linearno odvisnost med pojasnjevalnimi spremenljivkami, ne pa nelinearne.

Pomen predpostavke

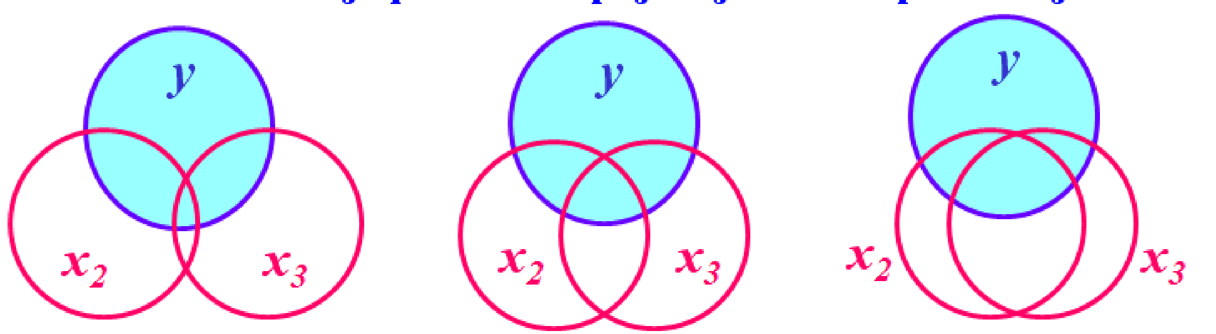
II. Nepopolna multikolinearnost

$$\lambda_1 \mathbf{x}_1 + \lambda_2 \mathbf{x}_2 + \ldots + \lambda_k \mathbf{x}_k + \mathbf{v} = \mathbf{0}$$

Vsako od pojasnjevalnih spremenljivk lahko predstavimo s preostalimi spremenljivkami, npr. za zadnjo velja:

$$\mathbf{x}_k = -\frac{\lambda_1}{\lambda_k}\mathbf{x}_1 - \frac{\lambda_2}{\lambda_k}\mathbf{x}_2 - \ldots - \frac{\lambda_{k-1}}{\lambda_k}\mathbf{x}_{k-1} - \frac{1}{\lambda_k}\mathbf{v}$$

Posamezne pojasnjevalne spremenljivke niso natančne linearne kombinacije preostalih pojasnjevalnih spremenljivk.



Šibka (multi)kolinearnost | Srednja (multi)kolinearnost | Močna (multi)kolinearnost

Pomen predpostavke

- Metoda najmanjših kvadratov, kljub multikolinearnosti, ostaja *NENALICE* regresijskih koeficientov.
- Varianca ocen regresijskih koeficientov vse bolj narašča z naraščanjem multikolinearnosti.
- Varianca ocen regresijskih koeficientov se neposredno odraža pri vrednosti t–statistike. Vse težje zavračamo H_0.
- Ocene regresijskih koeficientov ter njihovih standardnih napak postanejo močno občutljive na spreminjanje specifikacije modela.

Pomen predpostavke

Zaradi multikolinearnosti niso bistveno prizadete vrednosti R^2. Pogosto imamo visoko vrednost R^2, toda ***ocenjeni regresijski koeficinti so pri večini pojasnjevalnih spremenljivk statistično neznačilni.***

Resnost težav multikolinernosti je premo sorazmerna stopnji multikolinearnosti med pojasnjevalnimi spremenljivkami.

Ugotavljanje veljavnosti predpostavke

B **Kako ugotoviti oziroma odkriti veljavnost oziroma neveljavnost predpostavke?**

1. **Predznak enega ali več regresijskih koeficientov je v nasprotju s pričakovanji ekonomske teorije.**

2. **Visoka vrednost R^2 in večina ocenjenih regresijskih koeficientov je statistično neznačilnih.**

3. **Visoke vrednosti korelacijskih koeficientov med spremenljivkami.**

$$r_{x_j x_l}^2 \geq R_{y\hat{y}}^2 \qquad t = \frac{r_{x_j x_l}\sqrt{n-2}}{\sqrt{1-r_{x_j x_l}^2}} \sim t_{n-2}$$

Ugotavljanje veljavnosti predpostavke

4 **Uporaba *F* - statistike izračunane na podlagi "pomožnih regresij".**

$$x_{ki} = \beta_1 + \beta_2 x_{2i} + \ldots + \beta_{k-1} x_{k-1} + v_i$$

$$F_j = \frac{R^2_{x_j | \{x_1,\ldots,x_k\} \setminus x_j} / (k-2)}{\left(1 - R^2_{x_j | \{x_1,\ldots,x_k\} \setminus x_j}\right) / (n-k+1)}$$

$$F_j > F_{(k-2,\, n-k+1,\, \alpha)}$$

5 **Izračun variančnega inflacijskega faktorja – *VIF* (Variance Inflation Factor):**

$$VIF_j = \frac{1}{\left(1 - R^2_{x_j | \{x_1,\ldots,x_k\} \setminus x_j}\right)}$$

Ugotavljanje veljavnosti predpostavke

6

Izračun *števila pogojenosti* (condition number) in *indeksa pogojenosti* (condition index) matrike X^TX pojasnjevalnih spremenljivk:

$$K = \sqrt{\frac{\lambda_{\max}}{\lambda_{\min}}}$$

$$CI_j = \sqrt{\frac{\lambda_{\max}}{\lambda_j}}$$

Rešitve neizpolnjevanja predpostavke

C **Kakšne so možne rešitve v primeru neizpolnjevanja predpostavke?**

1 **Ali sploh kaj storiti?**
Odločitev "*ne storiti ničesar*" je lahko mnogokrat pravilna.

2 **V specifikacijo modela vključimo "a priori" poznavanja o vrednostih regresijskih koeficientov pri eni ali več pojasnjevalnih spremenljivkah.**

3 **Pri ocenjevanju modela "kombiniramo" podatke v časovnih vrstah s presečnimi podatki, danimi za enote v danem trenutku ali razmiku (time series and cross–sectional data).**

4 **Izpustitev ene ali več pojasnjevalnih spremenljivk, in sicer tistih, ki izražajo največjo multikolinearnost z ostalimi.**

Rešitve neizpolnjevanja predpostavke

5 Uporabimo transformacijo spremenljivk, ki naj bi zmanjšala njihovo medsebojno odvisnost (prve diference oziroma logaritmiranje).

6 Kadar je le možno, povečamo velikost vzorca.

7 Uporaba faktorske analize in glavnih komponent.

8 V oporo je lahko tudi postopek posameznih regresijskih modelov. Primerjava regresijskih koeficientov in t–statistik med delnimi in celotnim modelom more dati določene napotke o problemu multikolinearnosti.

V razmislek…

Multikolinearnost ne krši nobene od regresijskih predpostavk. Ocene ostajajo nepristranske in konsistentne, njihove standardne napake pa so pravilno ocenjene. Edini učinek multikolinearnosti je težavnost dobiti ocene regresijskih koeficientov z nizkimi standardnimi napakami.

Vendar pa ima majhno število opazovanj za rezultat podoben učinek, kakor ga ima tudi prisotnost neodvisnih spremenljivk z majhnimi variancami. Vprašanje "Kaj naj naredim glede multikolinearnosti?" je torej podobno vprašanju: "Kaj naj naredim, če nimam veliko opazovanj?"

Statističen odgovor ne obstaja.

C. H. Achen: *Interpreting and Using Regression*, 1982.

Univerza v Ljubljani

EKONOMSKA FAKULTETA

6.3 Heteroskedastičnost

EFMD EQUIS ACCREDITED

AACSB ACCREDITED

Motivacija

Heteroskedastičnost ni nikoli razlog za zavrnitev drugače dobrega modela.

Vendar pa je prav tako ne smemo ignorirati!

N. G. Mankiw & D. N. Gujarati

Pomen predpostavke

HOMOSKEDASTIČNOST

Izvor besede

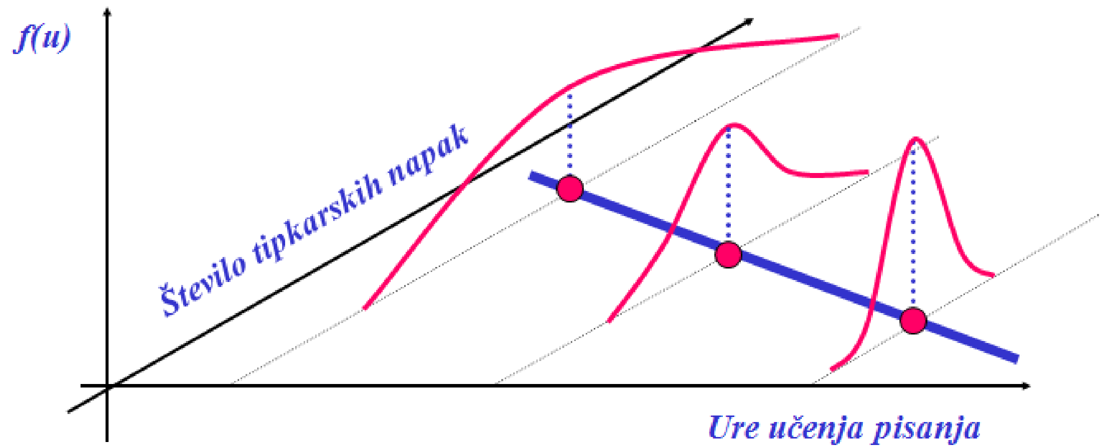
Homo : Skedasticos	Hetero : Skedasticos
Enak : Razpršenost	Neenak (različen) : Razpršenost
Homoscedasticity	Heteroscedasticity
Homoskedastičnost	Heteroskedastičnost

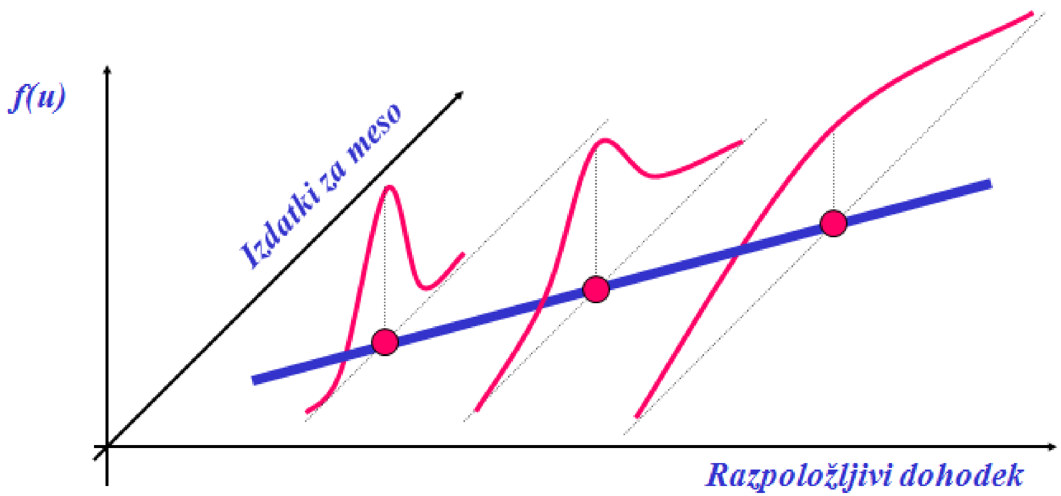
A Kaj predpostavka pomeni in kaj so temeljne posledice njenega neizpolnjevanja?

Predpostavka:

$$Var(u_i | x_i) = E\left[\left(u_i - E(u_i | x_i) \right)^2 \middle| x_i \right] = E\left[u_i^2 \middle| x_i \right] = E(u_i^2) = \sigma^2$$

Pomen predpostavke





Pomen predpostavke

Razlogi za pojav heteroskedastičnosti
(S. Valvanis: Econometrics, McGraw Hill, NY, 1959, str. 48)

1. Odločevalci se sčasoma učijo na napakah, torej naj bi varianca s časom upadala.
2. Mnogo pojavov v času tudi realno narašča; npr. bolj sproščena uporaba dohodkov (dobičkov podjetij) kot posledica gospodarske rasti povzroča naraščanje variabilnosti oziroma variance naključnih vplivov.
3. Variabilnost je mnogokrat posledica slabe organizacije oziroma načina zbiranja podatkov. Pričakovati je, da se s časom kvaliteta podatkov izboljšuje in s tem zmanjšuje varianca slučajnih vplivov.
4. Pri analizah, ki temeljijo na presečnih podatkih, so razlogi za heteroskedastičnost kaj različni – ugotavljamo jih na podlagi vsebine proučevanega pojava. Heteroskedastičnost je pogost pojav predvsem takrat, kadar so razlike v vrednosti posamezne spremenljivke med opazovanimi enotami zelo velike.
5. Pogosto je heteroskedastičnost tudi posledica napačne specifikacije regresijskega modela – "neprava heteroskedastičnost".

Pomen predpostavke

Posledice heteroskedastičnosti

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$$

$$E(\mathbf{u}) = \mathbf{0} \quad \text{toda} \quad E\left(\mathbf{u}\mathbf{u}^{\mathrm{T}}\right) = Var - \operatorname{cov}(\mathbf{u}) = \mathbf{W}$$

1 $$E(\mathbf{b}) = E\left[\left(\mathbf{X}^{\mathrm{T}}\mathbf{X}\right)^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{y}\right] = E\left[\left(\mathbf{X}^{\mathrm{T}}\mathbf{X}\right)^{-1}\mathbf{X}^{\mathrm{T}}(\mathbf{X}\boldsymbol{\beta} + \mathbf{u})\right] =$$

$$= E\left(\boldsymbol{\beta} + \left(\mathbf{X}^{\mathrm{T}}\mathbf{X}\right)^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{u}\right) = \boldsymbol{\beta} + \left(\mathbf{X}^{\mathrm{T}}\mathbf{X}\right)^{-1}\mathbf{X}^{\mathrm{T}}E(\mathbf{u}) = \boldsymbol{\beta}$$

Cenilka regresijskih koeficientov ostaja nepristranska!

Pomen predpostavke

2

$$Var-\operatorname{cov}(\mathbf{b})=E\left[(\mathbf{b}-\boldsymbol{\beta})(\mathbf{b}-\boldsymbol{\beta})^{\mathrm{T}}\right]=$$

$$=E\left[\left(\mathbf{X}^{\mathrm{T}}\mathbf{X}\right)^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{u}\mathbf{u}^{\mathrm{T}}\mathbf{X}\left(\mathbf{X}^{\mathrm{T}}\mathbf{X}\right)^{-1}\right]=$$

$$=\left(\mathbf{X}^{\mathrm{T}}\mathbf{X}\right)^{-1}\mathbf{X}^{\mathrm{T}}E(\mathbf{u}\mathbf{u}^{\mathrm{T}})\mathbf{X}\left(\mathbf{X}^{\mathrm{T}}\mathbf{X}\right)^{-1}=$$

$$=\left(\mathbf{X}^{\mathrm{T}}\mathbf{X}\right)^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{W}\mathbf{X}\left(\mathbf{X}^{\mathrm{T}}\mathbf{X}\right)^{-1}$$

Homoskedastičnost $\Longrightarrow$ $Var-\operatorname{cov}(\mathbf{b})=\sigma^2\left(\mathbf{X}^{\mathrm{T}}\mathbf{X}\right)^{-1}$

Cenilka regresijskih koeficientov ni več najbolj učinkovita!

MNKVD ni več NENALICE, je le NELICE!

Pomen predpostavke

3

Cenilka variance slučajne spremenljivke u je pristranska.

Cenilke varianc in kovarianc cenilk regresijskih koeficientov postanejo pristranske.

Testne statistike za regresijske koeficiente niso več zanesljive!

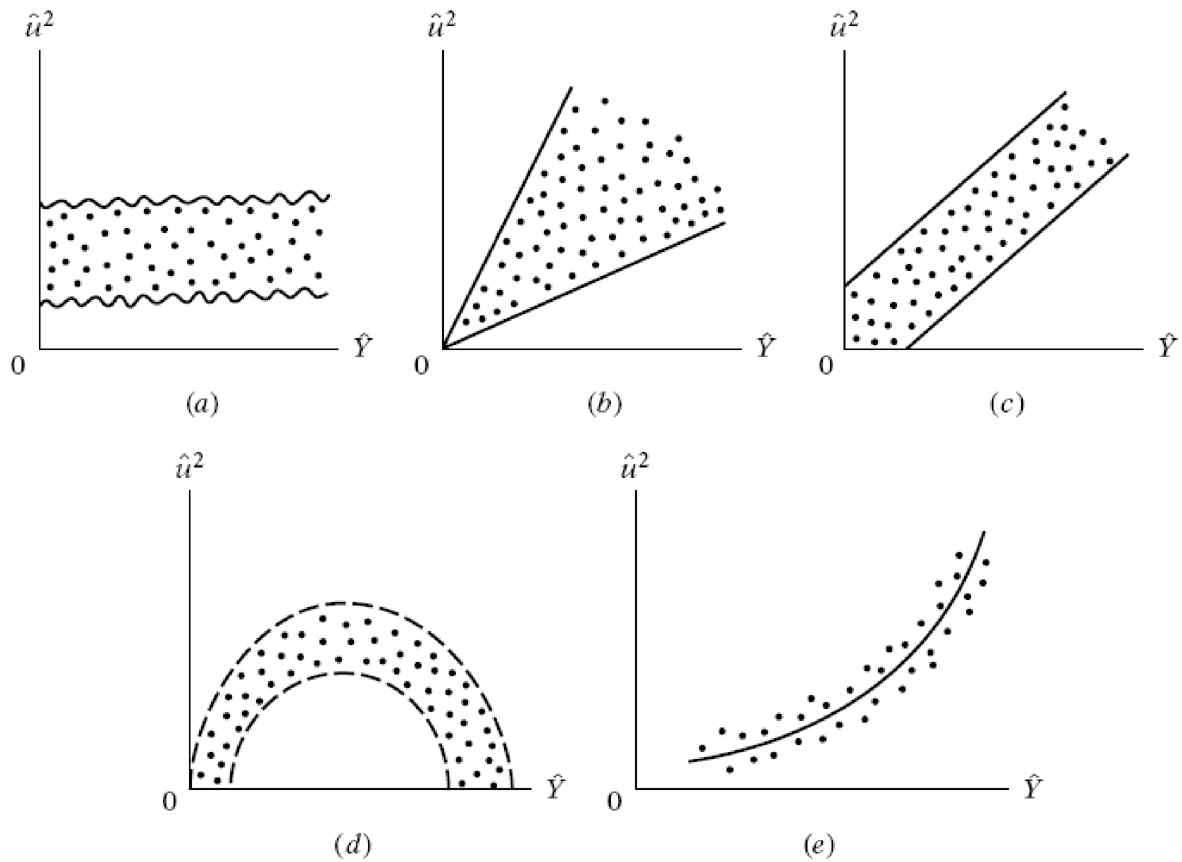
(Ne)veljavnost predpostavke

B&C

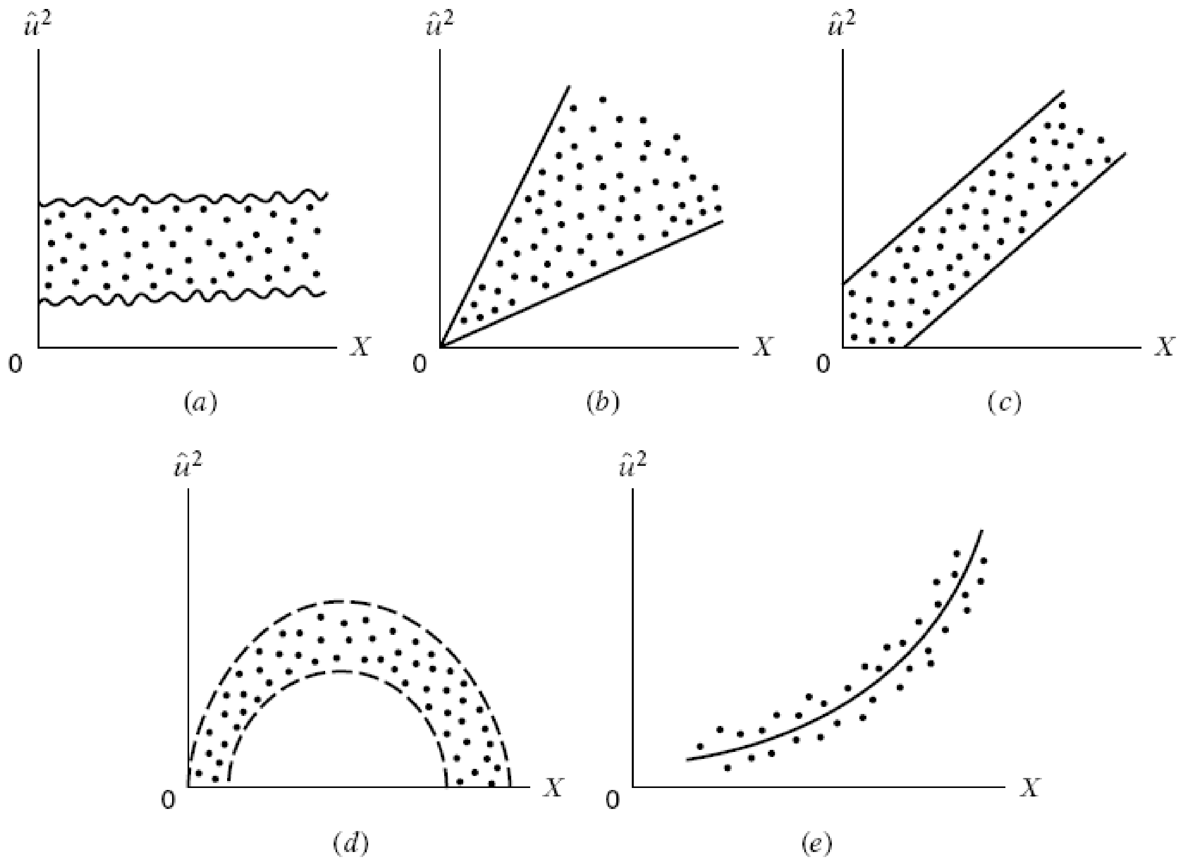
Kako ugotoviti oziroma odkriti veljavnost oziroma neveljavnost predpostavke?
Kakšne so možne rešitve v primeru neizpolnjevanja predpostavke?

(Ne)veljavnost predpostavke

B–1 Grafična metoda odkrivanja heteroskedastičnosti



(Ne)veljavnost predpostavke



(Ne)veljavnost predpostavke

C–1 Odpravljanje posledic heteroskedastičnosti, če velja:

$$Var(u_i) = E(u_i^2) = \sigma_i^2 = \sigma^2 x_i$$

$$u_i \sim (0, \sigma^2 x_i)$$

Model delimo s transformacijskim deliteljem: ⇒ $\sqrt{x_i}$

$$y_i = \beta_1 + \beta_2 x_i + u_i \quad \Rightarrow \quad \frac{y_i}{\sqrt{x_i}} = \beta_1 \frac{1}{\sqrt{x_i}} + \beta_2 \frac{x_i}{\sqrt{x_i}} + \frac{u_i}{\sqrt{x_i}}$$

$$\boxed{Var(u_i^*) = Var\left(\frac{u_i}{\sqrt{x_i}}\right) = E\left[\left(\frac{u_i}{\sqrt{x_i}}\right)^2\right] = \frac{1}{x_i} E(u_i^2) = \frac{1}{x_i}\sigma^2 x_i = \sigma^2}$$

(Ne)veljavnost predpostavke

Nekaj opozoril pri uporabi transformacij osnovnega modela

! **Izračunati moramo transformirane spremenljivke (odvisno in pojasnjevalne).**

! **Preveriti moramo, kako pravilno uporabiti MNKVD (v našem primeru model nima več konstantnega člena).**

! **Transformacija ohranja vsebino in s tem tudi razlago regresijskih koeficientov osnovnega modela. Je le sredstvo za prevedbo heteroskedestičnosti v homoskedestičnost.**

(Ne)veljavnost predpostavke

Generalizirani (Posplošeni) Najmanjši Kvadrati
GNK; PNK
(Generalized Least Squares – GLS)

Generalizirani najmanjši kvadrati so navadni najmanjši kvadrati, uporabljeni na transformiranem modelu, ki zagotavlja izpolnjevanje temeljnih predpostavk MNKVD.

Tehtani Najmanjši Kvadrati – TNK
(Weighted Least Squares – WLS)

(Ne)veljavnost predpostavke

B–2 Parkov test (1966)

a) $$E(u_i^2) = \sigma_i^2 = \sigma^2 x_i^\gamma$$

b) $$H_0 : \gamma = 0 \quad \Rightarrow \quad H_1 : \gamma \neq 0$$

c) $$\ln \sigma_i^2 = \ln \sigma^2 + \gamma \ln x_i + v_i$$

$$\widehat{\ln e_i^2} = g_1 + g_2 \ln x_i$$

d) $$|t(g_2)| \rangle t_{k(n-2)} \quad \Rightarrow$$ **H_0 zavrnemo**

? **Katero spremenljivko vključiti v test?**

(Ne)veljavnost predpostavke

C–2 Odpravljanje vpliva ugotovljene heteroskedastičnosti

$$y_i = \beta_1 + \beta_2 x_i + u_i$$

1 Določimo transformacijski delitelj: $\sqrt{x_i^{\gamma}} = x_i^{\frac{\gamma}{2}}$

$$\frac{y_i}{x_i^{\gamma/2}} = \beta_1 \frac{1}{x_i^{\gamma/2}} + \beta_2 \frac{x_i}{x_i^{\gamma/2}} + \left(\frac{u_i}{x_i^{\gamma/2}}\right)$$

2 $$Var\left(\frac{u_i}{x_i^{\gamma/2}}\right) = E\left[\left(\frac{u_i}{x_i^{\gamma/2}}\right)^2\right] = \frac{1}{x_i^{\gamma}} E\left(u_i^2\right) = \frac{1}{x_i^{\gamma}} \sigma^2 x_i^{\gamma} = \sigma^2$$

3 $$\widehat{y_i^*} = b_1 x_{1i}^* + b_2 x_{2i}^*$$

Regresijski model je brez konstantnega člena!

(Ne)veljavnost predpostavke

B–3 Glejserjev test (1969)

a)

1. $|u_i| = \beta_1 + \beta_2 x_i + v_i$; $E(u_i^2) = \sigma_i^2 = \sigma^2 x_i^2$
2. $|u_i| = \beta_1 + \beta_2 \sqrt{x_i} + v_i$; $E(u_i^2) = \sigma_i^2 = \sigma^2 x_i$
3. $|u_i| = \beta_1 + \beta_2 \frac{1}{x_i} + v_i$; $E(u_i^2) = \sigma_i^2 = \sigma^2 \frac{1}{x_i^2}$
4. $|u_i| = \beta_1 + \beta_2 \frac{1}{\sqrt{x_i}} + v_i$; $E(u_i^2) = \sigma_i^2 = \sigma^2 \frac{1}{x_i}$

b) $H_0 : \beta_2 = 0$; $H_1 : \beta_2 \neq 0$

c) **Izračunamo ostanke proučevanega regresijskega modela in jih uporabimo pri ocenjevanju enega izmed navedenih modelov.**

Npr. za prvi model: $\widehat{|e_i|} = b_1 + b_2 x_i$

d) $|t(b_2)| \rangle t_{k(n-2)}$ $\Rightarrow$ ***H_0* zavrnemo**

(Ne)veljavnost predpostavke

C–3 Odpravljanje vpliva ugotovljene heteroskedastičnosti

$$y_i = \beta_1 + \beta_2 x_i + u_i$$

1 **Določimo transformacijski delitelj**

Za model 1: x_i

$$\frac{y_i}{x_i} = \beta_1 \frac{1}{x_i} + \beta_2 + \left(\frac{u_i}{x_i}\right)$$

Za model 2: $\sqrt{x_i}$

$$\frac{y_i}{\sqrt{x_i}} = \beta_1 \frac{1}{\sqrt{x_i}} + \beta_2 \sqrt{x_i} + \left(\frac{u_i}{\sqrt{x_i}}\right)$$

(Ne)veljavnost predpostavke

2

Model 1: $$E\left[\left(\frac{u_i}{x_i}\right)^2\right]=\frac{1}{x_i^2}E\left(u_i^2\right)=\frac{1}{x_i^2}\sigma^2x_i^2=\sigma^2$$

Model 2: $$E\left[\left(\frac{u_i}{\sqrt{x_i}}\right)^2\right]=\frac{1}{x_i}E\left(u_i^2\right)=\frac{1}{x_i}\sigma^2x_i=\sigma^2$$

3

Model 1: $$\widehat{y_i^*}=b_1x_{1i}^*+b_2$$

OLS cenilka regresijske konstante v transformiranem modelu predstavlja GLS cenilko regresijskega koeficienta osnovnega modela!

Model 2: $$\widehat{y_i^*}=b_1x_{1i}^*+b_2x_{2i}^*$$

Regresijski model je brez konstantnega člena!

(Ne)veljavnost predpostavke

B–4 Goldfeld – Quandtov test (1965)

a) $$E(u_i^2) = \sigma_i^2 = \sigma^2 x_i^2$$

b) Opazovanja razvrstimo glede na vrednosti (od najmanjše do največje) pojasnjevalne spremenljivke, za katero sumimo, da povzroča heteroskedastičnost.

c) V sredini urejenih opazovanj izpustimo določeno število opazovanj (cca 1/3), preostala opazovanja razdelimo v dve skupini in sicer:

$n_1 = n_2 = \dfrac{n-c}{2}$; s c smo označili število izločenih opazovanj.

d) Ocenimo regresijska modela:

$n_1 \longrightarrow NVK_1$ $\quad$ $n_2 \longrightarrow NVK_2$

$$H_0 : \sigma_1^2 = \sigma_2^2 \quad ; \; H_1 : \sigma_1^2 \neq \sigma_2^2$$

e) $$\lambda = \frac{NVK_2 / (n_2 - k)}{NVK_1 / (n_1 - k)} \sim F_{(n_2-k, n_1-k, \alpha)}$$

$\lambda > F_{k(n_2-k, n_1-k, \alpha)} \longrightarrow$ H_0 zavrnemo

(Ne)veljavnost predpostavke

C–4 Odpravljanje vpliva ugotovljene heteroskedastičnosti

$$y_i = \beta_1 + \beta_2 x_i + u_i$$

1 Določimo transformacijski delitelj: x_i

$$\frac{y_i}{x_i} = \beta_1 \frac{1}{x_i} + \beta_2 + \left(\frac{u_i}{x_i}\right)$$

2

$$Var\left(\frac{u_i}{x_i}\right) = E\left[\left(\frac{u_i}{x_i}\right)^2\right] = \frac{1}{x_i^2} E\left(u_i^2\right) = \frac{1}{x_i^2} \sigma^2 x_i^2 = \sigma^2$$

3

$$\widehat{y_i^*} = b_1 x_i^* + b_2$$

OLS cenilka regresijske konstante v transformiranem modelu predstavlja GLS cenilko regresijskega koeficienta osnovnega modela!

(Ne)veljavnost predpostavke

B–5 Breusch – Paganov test (1979)

a) $$\sigma_i^2 = \alpha_1 + \alpha_2 x_{2i} + \ldots + \alpha_m x_{mi}$$

b) Na podlagi ostankov proučevanega regresijskega modela izračunamo njihovo varianco: $$\hat{\sigma}^2 = \frac{\sum e_i^2}{n}$$

c) Ocenimo pomožno regresijo in izračunamo njeno PVK:

$$\widehat{e_i^2} = a_1 + a_2 x_{2i} + \ldots + a_m x_{mi}; \quad PVK = \frac{NVK}{1 - R^2} - NVK$$

d) $H_0: \alpha_2 = \alpha_3 = \ldots = \alpha_m = 0$ $\quad$ H_1: Vsaj ena α je različna od 0

e) $$\theta = \frac{PVK}{2(\hat{\sigma}^2)^2} \sim \chi^2_{(m-1)}$$ $\theta \rangle \chi_k^2 \longrightarrow$ H_0 zavrnemo

(Ne)veljavnost predpostavke

C–5 Odpravljanje vpliva ugotovljene heteroskedastičnosti

1

$$\widehat{e_i^2} = a_1 + a_2 x_{2i} + \ldots + a_m x_{mi}$$

Določimo transformacijski delitelj: $s_i = \sqrt{\widehat{e_i^2}}$

2

$$Var\left(\frac{u_i}{s_i}\right) = E\left[\left(\frac{u_i}{s_i}\right)^2\right] = \frac{1}{s_i^2} E\left(u_i^2\right) = \frac{1}{s_i^2} s_i^2 = 1$$

3

$$\frac{y_i}{s_i} = \beta_1 \frac{1}{s_i} + \beta_2 \frac{x_{2i}}{s_i} + \beta_3 \frac{x_{3i}}{s_i} + \frac{u_i}{s_i}$$

Regresijski model je brez konstante!
Možne so negativne vrednosti s_i^2!

(Ne)veljavnost predpostavke

Kadar $\sqrt{\widehat{e_i^2}}$ ni definiran:

1 Ocenimo pomožno regresijo: $\widehat{\ln e_i^2} = a_1 + a_2 x_{2i} + \ldots + a_m x_{mi}$

Določimo transformacijski delitelj: $s_i = \sqrt{e^{\widehat{\ln e_i^2}}}$

2 3 Nadaljujemo prejšnji postopek.

(Ne)veljavnost predpostavke

B–6 Whiteov test (1980)

$$y_i = \beta_1 + \beta_2 x_{2i} + \beta_3 x_{3i} + u_i$$

a) $$\sigma_i^2 = \alpha_1 + \alpha_2 x_{2i} + \alpha_3 x_{3i} + \alpha_4 x_{2i}^2 + \alpha_5 x_{3i}^2 + \alpha_6 x_{2i} x_{3i}$$

b) Ocenimo regresijski model, ugotovimo ostanke e_i ter izračunamo

$$e_i^2 = a_1 + a_2 x_{2i} + a_3 x_{3i} + a_4 x_{2i}^2 + a_5 x_{3i}^2 + a_6 x_{2i} x_{3i} + v_i$$

c) $$H_0 : \alpha_2 = \alpha_3 = \ldots = \alpha_m = 0$$

H_1: Vsaj ena α je različna od 0

d) $$\theta(W) = nR^2 \sim \chi^2_{(m-1)}$$

$\theta \rangle \chi^2_k$ → H_0 zavrnemo

(Ne)veljavnost predpostavke

C–6 Odpravljanje vpliva ugotovljene heteroskedastičnosti

1 Določimo transformacijski delitelj: $s_i = \sqrt{\widehat{e_i^2}}$

2 $$Var\left(\frac{u_i}{s_i}\right) = E\left[\left(\frac{u_i}{s_i}\right)^2\right] = \frac{1}{s_i^2} E\left(u_i^2\right) = \frac{1}{s_i^2} s_i^2 = 1$$

3 $$\frac{y_i}{s_i} = \beta_1 \frac{1}{s_i} + \beta_2 \frac{x_{2i}}{s_i} + \beta_3 \frac{x_{3i}}{s_i} + \frac{u_i}{s_i}$$

Regresijski model je brez konstante!
Možne so negativne vrednosti s_i^2!
Pozorni moramo biti na število stopinj prostosti!

(Ne)veljavnost predpostavke

C–7 Logaritemska transformacija

Logaritemska transformacija zmanjša variabilnost vseh premenljivk v modelu, tudi odvisne in s tem tudi vrednosti slučajne spremenljivke u_i.

Na transformiranem modelu nato uporabimo ustrezni test heteroskedastičnosti in po potrebi ukrepamo.

Na kaj moramo biti pozorni…

PRI UPORABI PRIKAZANIH POSTOPKOV VELJA UPOŠTEVATI:

- Pri multipli regresiji se moramo odločiti, katera oz. katere spremenljivke povzročajo heteroskedastičnost (analiziramo razsevne diagrame e_i^2 z vsemi spremenljivkami).
- Logaritemska transformacija ni vedno uporabna (upoštevanje ekonomske teorije in $(y_i, x_i) > 0$).
- Problem nesmiselne povezave. Pogosto se zgodi, da izhodiščni model ne odraža odvisnosti med odvisno in pojasnjevalnimi spremenljivkami, medtem ko transformirani model to odvisnost potrjuje.
- Kadar nimamo oz. ne poznamo vrednosti σ_u^2, veljajo omenjene procedure testiranja prisotnosti heteroskedastičnosti strogo gledano le za velike vzorce.

Uporaba robustnih standardnih napak

Izpis rezultatov ocenjevanja regresijskega modela povpraševanja po denarju

```
. regress hm1 ppr rvp rvv czp

      Source |       SS       df       MS              Number of obs =      96
-------------+------------------------------           F(  4,    91) =  527.72
       Model |  11431132.5     4  2857783.12           Prob > F      =  0.0000
    Residual |  492791.936    91    5415.296           R-squared     =  0.9587
-------------+------------------------------           Adj R-squared =  0.9569
       Total |  11923924.4    95  125514.994           Root MSE      =  73.589

------------------------------------------------------------------------------
         hm1 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         ppr |   1.697766    .513892     3.30   0.001     .6769831     2.71855
         rvp |  -311.6847   45.25178    -6.89   0.000    -401.5718   -221.7976
         rvv |  -11.57513    5.33166    -2.17   0.033    -22.16582    -.98444
         czp |   11.50168   1.472604     7.81   0.000     8.576535    14.42683
       _cons |  -229.2038   125.2134    -1.83   0.070    -477.9248    19.51725
------------------------------------------------------------------------------
```

Uporaba robustnih standardnih napak

```
. whitetst

White's general test statistic :  53.83009  Chi-sq(14)  P-value =  1.4e-06

. regress hm1 ppr rvp rvv czp, robust

Linear regression                                      Number of obs =      96
                                                       F(  4,    91) = 1000.25
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.9587
                                                       Root MSE      =  73.589

------------------------------------------------------------------------------
             |               Robust
         hm1 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         ppr |   1.697766   .5633882     3.01   0.003     .5786649    2.816868
         rvp |  -311.6847   44.33028    -7.03   0.000    -399.7413   -223.6281
         rvv |  -11.57513   3.532513    -3.28   0.001    -18.59203   -4.558225
         czp |   11.50168    1.32376     8.69   0.000     8.872196    14.13117
       _cons |  -229.2038   58.25138    -3.93   0.000     -344.913   -113.4945
------------------------------------------------------------------------------
```

Univerza v Ljubljani

EKONOMSKA FAKULTETA

6.4 Avtokorelacija

EFMD EQUIS ACCREDITED

AACSB ACCREDITED

Temeljni pojmi

Pomen besede

Avtokorelacija je določena kot korelacija med členi serije (vrste) opazovanj urejenih po času (v časovnih vrstah) ali prostoru (pri presečnih podatkih).

Serijska korelacija, ima isti pomen kot avtokorelacija, toda velja za vzorčne podatke (A Dictionary of Statistical Terms, str. 10 in 176).

Nekateri avtorji ju ne razlikujejo v omenjenem smislu temveč:

Serijsko korelacijo poimenujejo korelacijo med dvema različnima časovnima vrstama, pri čemer je ena odložena za določeno število časovnih enot:

$u_1, u_2, \ldots, u_{n-1}$, ter $v_2, v_3, \ldots, v_n$

Pomen predpostavke

A **Kaj predpostavka pomeni in kaj so temeljne posledice njenega neizpolnjevanja?**

Predpostavka

$$Cov\left(u_i,u_j \middle| x_i,x_j\right)=0 \quad \Longleftrightarrow \quad Cov\left(u_i,u_j \middle| x_i,x_j\right)\neq 0$$

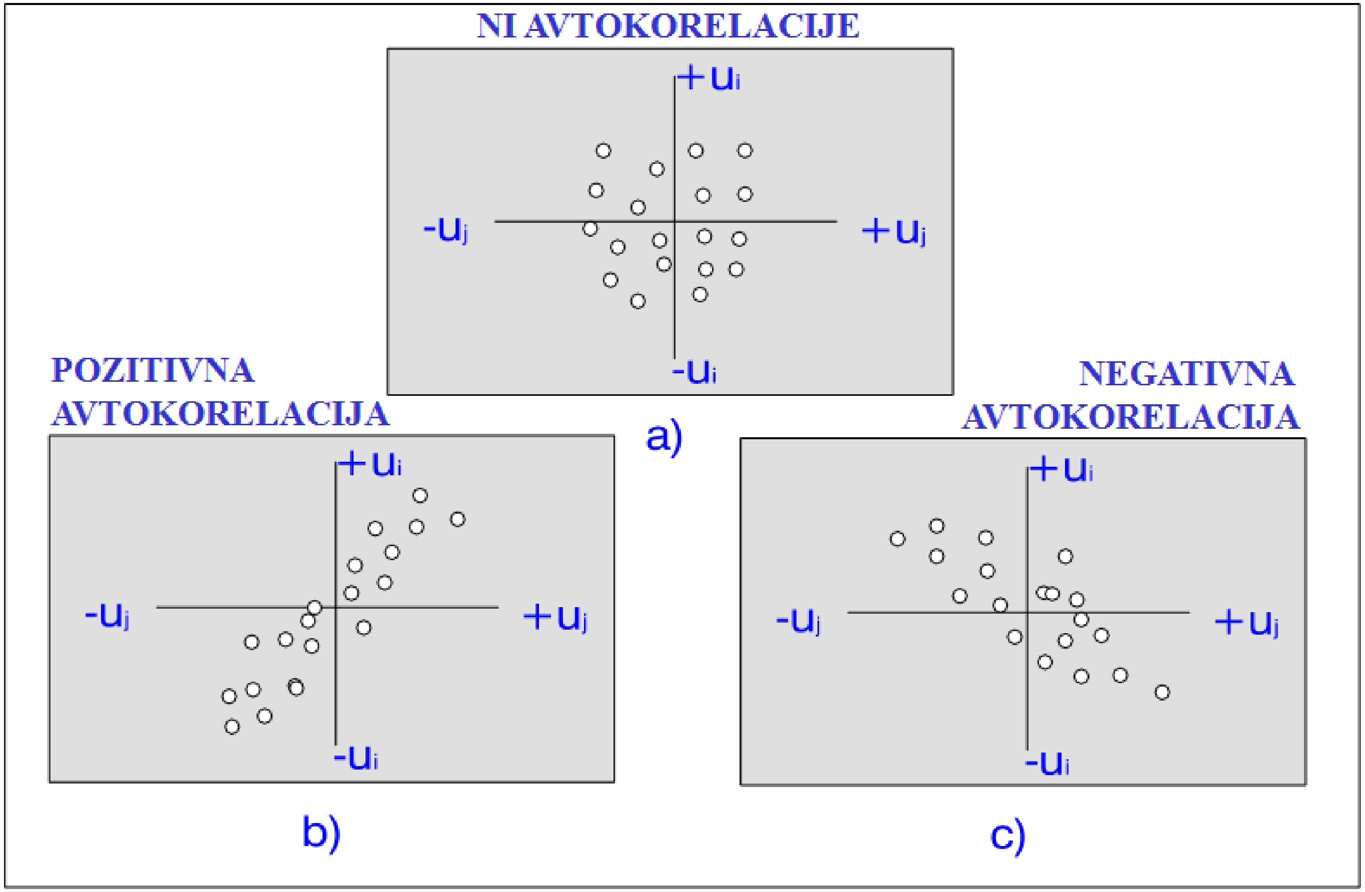
$$i \neq j \qquad\qquad i \neq j$$

NI avtokorelacije | **JE avtokorelacija**

Razlikujemo med *pravo* (pristno, čisto) *avtokorelacijo* in *nepravo* (nepristno) *avtokorelacijo*. O slednji govorimo, kadar je le-ta posledica napačne specifikacije regresijskega modela.

Pomen predpostavke

Razsevni diagrami za vrednosti slučajne spremenljivke u:



Pomen predpostavke

Razlogi za pojav avtokorelacije:

1. **Pri večini časovnih vrst v ekonomiji velja *prikrita inercija*, to je razvoj v času v odvisnosti od pojava v predhodni ali predhodnih časovnih enotah. To je tudi razlog njihovemu cikličnemu razvoju.**

2. ***Napake pri specifikaciji modela*. Pogosto v empiričnih raziskavah izhajamo iz najbolj splošnega (teoretično) opredeljenega modela. Pri tem lahko "izpade" pomembna pojasnjevalna spremenljivka. Temu pravimo "izločitvena" specifikacijska napaka (excluded-variable specification bias). To kaj rado povzroči problem avtokorelacije. Npr.:**

$$y_t = \beta_1 + \beta_2 x_{2t} + \beta_3 x_{3t} + \beta_4 x_{4t} + u_t$$

Količina (y_t), **Cena** (x_{2t}), **Dohodek** (x_{3t}), **Cena substituta ali kompl. blaga** (x_{4t})

$$y_t = \beta_1 + \beta_2 x_{2t} + \beta_3 x_{3t} + v_t \Rightarrow v_t = \beta_4 x_{4t} + u_t$$

Pomen predpostavke

3 **Napaka pri specifikaciji modela zaradi (napačne) *neprave funkcijske oblike* regresijskega modela.**

$$y_i = \beta_1 + \beta_2 x_{2i} + \beta_3 x_{2i}^2 + u_i$$

Mejni stroški **Proizvodnja** **Proizvodnja**

$$y_i = \beta_1 + \beta_2 x_{2i} + v_i \Rightarrow v_i = \beta_3 x_{2i}^2 + u_i$$

4 **Neupoštevanje *odloženih spremenljivk* oziroma odlogov pri spremenljivkah.**

Pomen predpostavke

5 *Uporaba transformacij* **pri preoblikovanju časovnih vrst (uporaba povprečij, vsot, trendov in drugih interpolacij).**

6 *Nestacionarnost časovne serije*, **kar pomeni, da so prvi in drugi momenti časovne serije spremenljivi v času.**

Če sta odvisna in pojasnjevalna spremenljivka nestacionarni, je zelo verjetno, da je tudi slučajna spremenljivka nestacionarna, v modelu pa imamo avtokorelacijo.

Pomen predpostavke

Predpostavke o vrstah avtokorelacije

1. **Avtokorelacija prvega reda (avtoregresijska shema prvega reda)**

$$u_t = \rho_1 u_{t-1} + \varepsilon_t \Rightarrow \text{AR(1)}$$

$-1 < \rho_1 < 1$ $\rho = \rho_1$ – koeficient avtokorelacije prvega reda

2. **Avtokorelacija drugega reda**

$$u_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + \varepsilon_t \Rightarrow \text{AR(2)}$$

3. **Avtokorelacija p-tega reda**

$$u_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + \ldots + \rho_p u_{t-p} + \varepsilon_t \Rightarrow \text{AR(p)}$$

Pomen predpostavke

Posledice avtokorelacije

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$$

$$E(\mathbf{u}) = \mathbf{0} \quad \text{toda} \quad E\left(\mathbf{u}\mathbf{u}^{\mathrm{T}}\right) = Var - \operatorname{cov}(\mathbf{u}) = \mathbf{W}$$

1 $$E(\mathbf{b}) = E\left[\left(\mathbf{X}^{\mathrm{T}}\mathbf{X}\right)^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{y}\right] = E\left[\left(\mathbf{X}^{\mathrm{T}}\mathbf{X}\right)^{-1}\mathbf{X}^{\mathrm{T}}(\mathbf{X}\boldsymbol{\beta} + \mathbf{u})\right] =$$

$$= E\left(\boldsymbol{\beta} + \left(\mathbf{X}^{\mathrm{T}}\mathbf{X}\right)^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{u}\right) = \boldsymbol{\beta} + \left(\mathbf{X}^{\mathrm{T}}\mathbf{X}\right)^{-1}\mathbf{X}^{\mathrm{T}}E(\mathbf{u}) = \boldsymbol{\beta}$$

Cenilka regresijskih koeficientov ostaja nepristranska!

Pomen predpostavke

2

$$Var-\mathrm{cov}\left(\mathbf{b}\right)=E\left[\left(\mathbf{b}-\boldsymbol{\beta}\right)\left(\mathbf{b}-\boldsymbol{\beta}\right)^{\mathrm{T}}\right]=$$

$$=E\left[\left(\mathbf{X}^{\mathrm{T}}\mathbf{X}\right)^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{u}\mathbf{u}^{\mathrm{T}}\mathbf{X}\left(\mathbf{X}^{\mathrm{T}}\mathbf{X}\right)^{-1}\right]=$$

$$=\left(\mathbf{X}^{\mathrm{T}}\mathbf{X}\right)^{-1}\mathbf{X}^{\mathrm{T}}E(\mathbf{u}\mathbf{u}^{\mathrm{T}})\mathbf{X}\left(\mathbf{X}^{\mathrm{T}}\mathbf{X}\right)^{-1}=$$

$$=\left(\mathbf{X}^{\mathrm{T}}\mathbf{X}\right)^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{W}\mathbf{X}\left(\mathbf{X}^{\mathrm{T}}\mathbf{X}\right)^{-1}$$

Ni avtokorelacije $\Rightarrow$ $Var-\mathrm{cov}\left(\mathbf{b}\right)=\sigma^2\left(\mathbf{X}^{\mathrm{T}}\mathbf{X}\right)^{-1}$

Cenilka regresijskih koeficientov ni več najbolj učinkovita!

MNKVD ni več NENALICE, je le NELICE!

Pomen predpostavke

3

Cenilka variance slučajne spremenljivke *u* je pristranska.

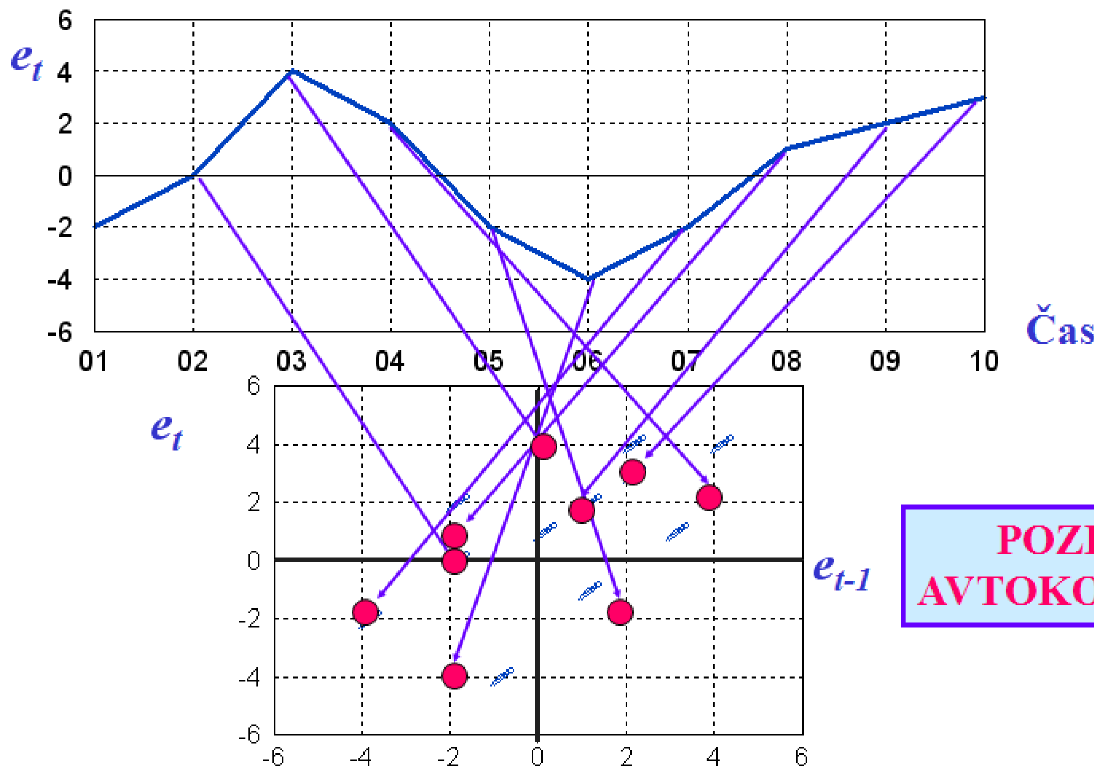
Cenilke varianc in kovarianc cenilk regresijskih koeficientov postanejo pristranske.

Testne statistike za regresijske koeficiente niso več zanesljive!

Ugotavljanje veljavnosti predpostavke

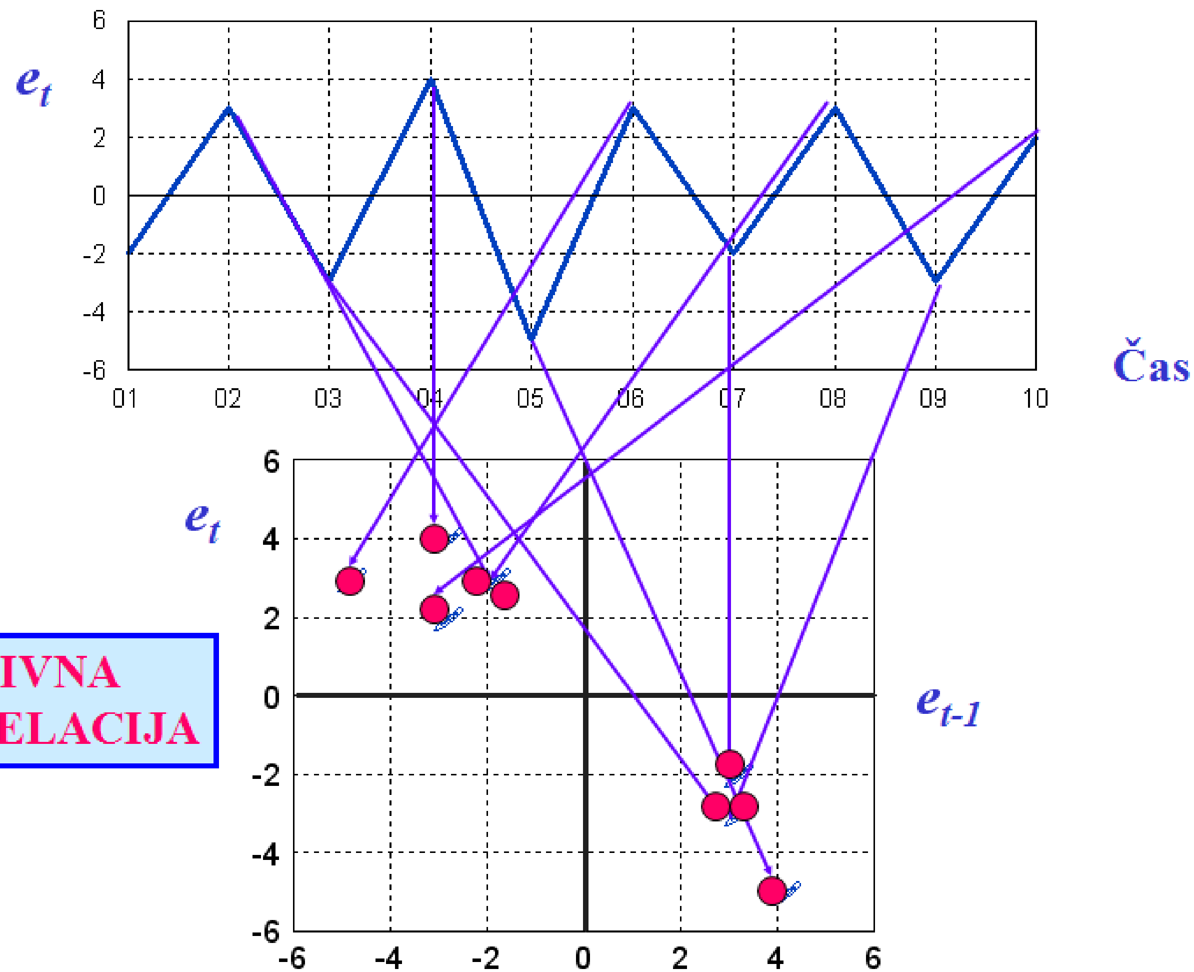
B **Kako ugotoviti oziroma odkriti veljavnost oziroma neveljavnost predpostavke?**

1. Grafična metoda



POZITIVNA AVTOKORELACIJA

Ugotavljanje veljavnosti predpostavke



NEGATIVNA AVTOKORELACIJA

Ugotavljanje veljavnosti predpostavke

2. Test sekvenc (The runs test)
Gearyjev test (1970)

1 **Na podlagi ostankov VRM izračunamo**

$$E(r) = \frac{2N_1N_2}{N_1+N_2} + 1 \qquad \sigma_r^2 = \frac{2N_1N_2(2N_1N_2 - N_1 - N_2)}{(N_1+N_2)^2(N_1+N_2-1)}$$

Sekvenca: neprekinjen niz enako predznačenih ostankov

r – število sprememb predznaka ostankov (e_t)

N_1 – število pozitivnih vrednosti ostankov

N_2 – število negativnih vrednosti ostankov

2 **Izračunamo interval zaupanja**

$$\Pr(E(r) - 1{,}96\sigma_r \leq r \leq E(r) + 1{,}96\sigma_r) = 0{,}95$$

Ugotavljanje veljavnosti predpostavke

3. Test asociacije rezidualov modela χ^2 – test

1

Na podlagi rezidualov modela pripravimo tabelo

Reziduali	Pozitivne vrednosti v t	Negativne vrednosti v t	
Pozitivne vrednosti v $t-1$	$f_{11}\ (+\,+)$	$f_{12}\ (+\,-)$	$f_{1.}$
Negativne vrednosti v $t-1$	$f_{21}\ (-\,+)$	$f_{22}\ (-\,-)$	$f_{2.}$
	$f_{.1}$	$f_{.2}$	$n-1$

2

$$f'_{ij} = \frac{f_{i.}f_{.j}}{n-1} \Rightarrow \chi^2 = \sum \frac{(f_{ij} - f'_{ij})^2}{f'_{ij}}$$

$$\chi^2 \sim \chi^2_{(m=1,\,\alpha)}$$

Ugotavljanje veljavnosti predpostavke

4. Durbin – Watsonov *d* test (1951)

1 **Definicija *d (DW)* statistike**

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

2 **Predpostavke**

a) Pojasnjevalne spremenljivke so nestohastične.
b) Preverjamo avtokorelacijo prvega reda.
c) Regresijski model ne vključuje odložene(nih) vrednosti odvisne spremenljivke.
d) Podatki se nanašajo na zaporedne člene (ni manjkajočih podatkov).
e) Regresijski model vsebuje regresijsko konstanto.
f) Velja predpostavka normalne porazdelitve *u*.

Ugotavljanje veljavnosti predpostavke

3 **Lastnosti *d (DW)* statistike**

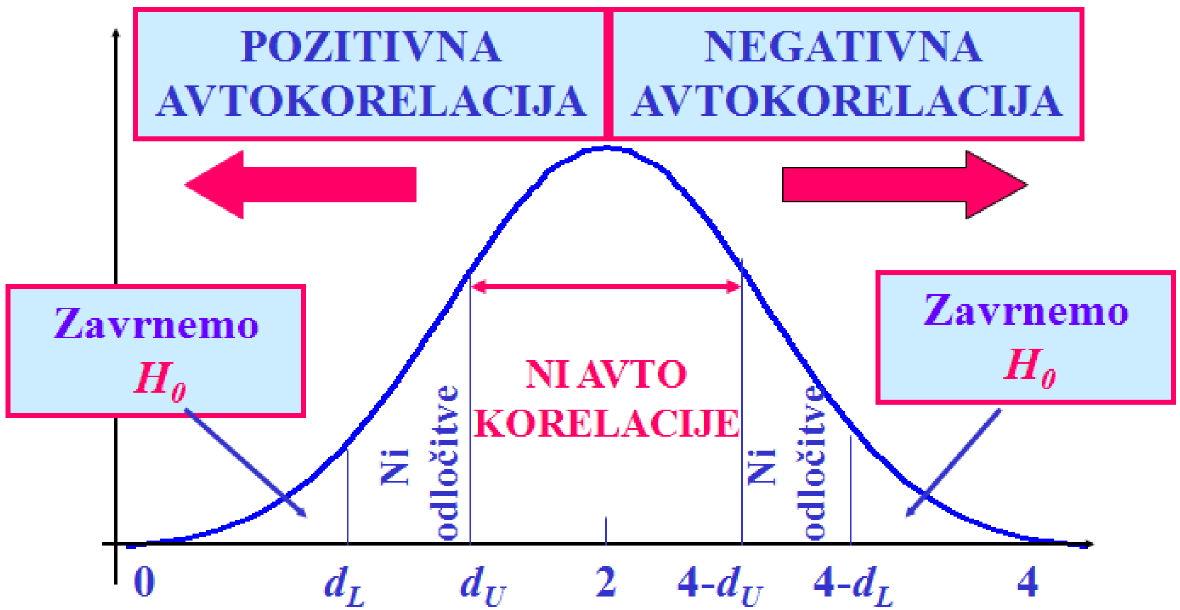
$$d = \frac{\sum e_t^2 + \sum e_{t-1}^2 - 2\sum e_t e_{t-1}}{\sum e_t^2}$$

$$d \cong \frac{2\sum e_t^2 - 2\sum e_t e_{t-1}}{\sum e_t^2} \cong 2\left(1 - \frac{\sum e_t e_{t-1}}{\sum e_t^2}\right)$$

$$\frac{\sum e_t e_{t-1}}{\sum e_t^2} \cong \hat{\rho}_1$$

$$d \cong 2(1 - \hat{\rho}_1) \Rightarrow 0 \le d \le 4$$

Ugotavljanje veljavnosti predpostavke



$H_0: \rho_1 \leq 0$	$H_0: \rho_1 = 0$	$H_0: \rho_1 \geq 0$
$H_1: \rho_1 > 0$	$H_1: \rho_1 \neq 0$	$H_1: \rho_1 < 0$

Ničelna domneva	Odločitev	Če velja
Ni poz. avtokorelacije	Zavrnemo	$0 < d < d_L$
	Ni odločitve	$d_L \leq d \leq d_U$
Ni neg. avtokorelacije	Zavrnemo	$4 - d_L < d < 4$
	Ni odločitve	$4 - d_U \leq d \leq 4 - d_L$
Ni avtokorelacije	Ne zavrnemo domneve	$d_U < d < 4 - d_U$

Ugotavljanje veljavnosti predpostavke

5. Wallisov test (1972)

1 **Avtokorelacija 4. reda**

$$u_t = \rho_4 u_{t-4} + \varepsilon_t \Rightarrow \text{AR (4)}$$

2 **Testna statistika**

$$d_4 = W_4 = DW_4 = \frac{\sum_{t=5}^{n}(e_t - e_{t-4})^2}{\sum_{t=1}^{n} e_t^2}$$

a) **Specifikacija modela ne vsebuje nepravih spremenljivk**

b) **Specifikacija modela vsebuje neprave spremenljivke**

Ugotavljanje veljavnosti predpostavke

6. Durbinova *h* statistika (1970)

1 **Regresijski model vključuje med pojasnjevalnimi spremenljivkami tudi odloženo odvisno spremenljivko**

$$y_t = \beta_1 + \beta_2 x_{2t} + \cdots + \beta_k x_{kt} + \beta_{k+1} y_{t-1} + u_t$$

2 **Testna statistika**

$$h = \hat{\rho}_1 \sqrt{\frac{n'}{1 - n'(\mathrm{var}(b_{k+1}))}} \quad \textit{oziroma} \quad h \cong (1 - 0,5d)\sqrt{\frac{n'}{1 - n'(\mathrm{var}(b_{k+1}))}}$$

$$h \sim AN(0,1)$$

$h > 1,96$; v modelu je prisotna pozitivna avtokorelacija (AR(1))

$h < -1,96$; v modelu je prisotna negativna avtokorelacija (AR(1))

$-1,96 < h < 1,96$; v modelu ni prisotna avtokorelacija (AR(1))

Ugotavljanje veljavnosti predpostavke

7. Breusch – Godfreyjev test (1978) (test Lagrangeovega multiplikatorja)

1 **Prepostavlja se avtokorelacija poljubnega reda**

$$u_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + \cdots + \rho_p u_{t-p} + \varepsilon_t$$

$H_0: \rho_1 = \rho_2 = \cdots = \rho_p = 0$ → H_1: **Vsaj en ρ ni enak 0**

2 **Testna statistika**

Ocenimo pomožno regresijo

$$\hat{e}_t = b_1 + b_2 x_{2t} + \cdots + b_k x_{kt} + \hat{\rho}_1 e_{t-1} + \cdots + \hat{\rho}_p e_{t-p}$$

$$LM = (n-p)R^2 \sim \chi^2_{(p)}$$

- Regresijski model lahko vključuje tudi odložene vrednosti odvisne spremenljivke
- Dolžina odloga ni poznana vnaprej
- Težave s stopinjami prostosti

V razmislek…

Zakaj imamo toliko testov za avtokorelacijo?

Odgovor je: "Zato, ker doslej še noben posamezen test ni bil enotno proglašen za najboljšega (statistično najmočnejšega) in je zato analitik še vedno v nezavidljivem položaju, ko mora proučiti množico testnih procedur za odkrivanje prisotnosti oziroma strukture avtokorelacije."

Podoben argument seveda velja za različne teste heteroskedastičnosti, ki smo jih že proučevali.

D. N. Gujarati & D. C. Porter: *Basic Econometrics, 5th Edition*, 2009.

Rešitve neizpolnjevanja predpostavke

C Možne rešitve v primeru neizpolnjevanja predpostavke?

I. Znana je oblika avtokorelacije – znana je vrednost koeficienta(ov) avtokorelacije

Primer:

Avtokorelacija prvega reda in regresijski model z eno pojasnjevalno spremenljivko

$$y_t = \beta_1 + \beta_2 x_{2t} + u_t$$

$$AR(1): \quad u_t = \rho_1 u_{t-1} + \varepsilon_t$$

Rešitve neizpolnjevanja predpostavke

$$u_t - \rho_1 u_{t-1} = \varepsilon_t$$

$$y_t - \beta_1 - \beta_2 x_{2t} - \rho_1 y_{t-1} + \rho_1 \beta_1 + \rho_1 \beta_2 x_{2t-1} = \varepsilon_t$$

$$y_t - \rho_1 y_{t-1} = \beta_1 (1 - \rho_1) + \beta_2 (x_{2t} - \rho_1 x_{2t-1}) + \varepsilon_t$$

$$y_t^* = \beta_1^* + \beta_2 x_{2t}^* + \varepsilon_t$$

Generalizirani (posplošeni) najmanjši kvadrati
PNK (GLS)

Generalizirana (posplošena ali kvazi) diferenčna enačba
PDE (GDE)

Prais – Winstenova transformacija $\Rightarrow$ $y_1\sqrt{1-\rho_1^2}$; $x_{j1}\sqrt{1-\rho_1^2}$

Rešitve neizpolnjevanja predpostavke

II. **Vrednost koeficient(ov) avtokorelacije ni znana**

Koeficient(e) avtokorelacije prvega (in višjih redov) reda ocenimo!

1. Metoda prvih diferenc, AR(1)

Pristop pride v poštev, kadar je:
– vrednost koeficienta avtokorelacije zelo visoka oziroma
– vrednost Durbin-Watsonove d statistike zelo nizka.

Maddala (1992): $d < R^2$

Rešitve neizpolnjevanja predpostavke

$$u_t - \rho_1 u_{t-1} = \varepsilon_t$$

$$y_t - \beta_1 - \beta_2 x_{2t} - \rho_1 y_{t-1} + \rho_1 \beta_1 + \rho_1 \beta_2 x_{2t-1} = \varepsilon_t$$

$$y_t - \rho_1 y_{t-1} = \beta_1 (1 - \rho_1) + \beta_2 (x_{2t} - \rho x_{2t-1}) + \varepsilon_t; \quad \rho_1 \approx 1$$

$$y_t - 1 y_{t-1} = \beta_1 (1 - 1) + \beta_2 (x_{2t} - 1 x_{2t-1}) + \varepsilon_t$$

$$\Delta y_t = \beta_2 \Delta x_{2t} + \varepsilon_t$$

Transformirani regresijski model je brez konstantnega člena!

Rešitve neizpolnjevanja predpostavke

2. Ocena ρ na podlagi vrednosti Durbin-Watsonove d statistike, AR(1)

$$d \cong 2(1-\hat{\rho}_1) \Rightarrow \hat{\rho}_1 \cong 1-\frac{d}{2} \quad \Rightarrow \quad \text{PDE}$$

3. Ocena ρ na podlagi rezidualov, AR(1) in AR(p)

$$e_t = \hat{\rho}_1 e_{t-1} + v_t; \quad \hat{\rho}_1 \quad \Rightarrow \quad \text{PDE}$$

$$e_t = \hat{\rho}_p e_{t-p} + v_t; \quad \hat{\rho}_p \quad \Rightarrow \quad \text{PDE}$$

Rešitve neizpolnjevanja predpostavke

4. Iterativne metode ocenjevanja ρ, AR(1)
Cochrane – Orcuttova iterativna procedura (1949)

a) Ocenimo specificirani regresijski model (predpostavimo, da je bivariatni in da je avtokorelacija prvega reda)

b) Na podlagi ostankov modela ocenimo vrednost koeficienta avtokorelacije z regresijskim modelom

$$e_t = \hat{\rho}_1 e_{t-1} + v_t$$

- število opazovanih enot je $(n - 1)$
- regresijski model nima konstantnega člena

c) Ocenjeno vrednost koeficienta avtokorelacije uporabimo za zapis transformiranega izhodiščnega modela

$$y_t - \hat{\rho}_1 y_{t-1} = \beta_1(1-\hat{\rho}_1) + \beta_2(x_{2t} - \hat{\rho}_1 x_{2t-1}) + v_t$$

$$y_t^* = \beta_1^* + \beta_2 x_{2t}^* + v_t$$

Rešitve neizpolnjevanja predpostavke

d) Na podlagi ostankov transformiranega modela pod c) ocenimo

$$e'_t = \hat{\rho}'_1 e'_{t-1} + v'_t$$

e) Novo ocenjeno vrednost koeficienta avtokorelacije uporabimo za zapis transformiranega izhodiščnega modela

$$y_t - \hat{\rho}'_1 y_{t-1} = \beta_1(1-\hat{\rho}'_1) + \beta_2(x_{2t} - \hat{\rho}'_1 x_{2t-1}) + v'_t$$

$$y'^*_t = \beta'^*_1 + \beta_2 x'^*_{2t} + v'_t$$

f) Se ocenjeni vrednosti koeficientov avtokorelacije razlikujeta za več kot 0,01 ali 0,005?

NE — Končamo iterativni postopek

DA — Vrnemo se na korak d)

Rešitve neizpolnjevanja predpostavke

5. Box-Jenkinsovo ARIMA modeliranje, AR(p)

Uporabimo ARMAX(p, q, k) model, pri čemer moramo opredeliti:

– odvisno spremenljivko y;

– pojasnjevalne spremenljivke x_j, $j = 1, \ldots, k$;

– avtoregresijske člene AR(p) za tiste odloge, pri katerih ne moremo zavrniti ničelne predpostavke o odsotnosti avtokorelacije;

– ne pa (nujno) tudi členov drsečih povprečij MA(q).

Uporaba HAC standardnih napak

Izpis rezultatov ocenjevanja regresijskega modela povpraševanja po denarju

```
. regress hm1 ppr rvp rvv czp

      Source |       SS       df       MS              Number of obs =      96
-------------+------------------------------           F(  4,    91) =  527.72
       Model |  11431132.5     4  2857783.12           Prob > F      =  0.0000
    Residual |  492791.936    91    5415.296           R-squared     =  0.9587
-------------+------------------------------           Adj R-squared =  0.9569
       Total |  11923924.4    95  125514.994           Root MSE      =  73.589

------------------------------------------------------------------------------
         hm1 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         ppr |   1.697766    .513892     3.30   0.001     .6769831     2.71855
         rvp |  -311.6847   45.25178    -6.89   0.000    -401.5718   -221.7976
         rvv |  -11.57513    5.33166    -2.17   0.033    -22.16582    -.98444
         czp |   11.50168   1.472604     7.81   0.000     8.576535    14.42683
       _cons |  -229.2038   125.2134    -1.83   0.070    -477.9248    19.51725
------------------------------------------------------------------------------
```

Uporaba HAC standardnih napak

```
. estat bgodfrey, lags(1 2 3 4 5 6 7 8 9 10 11 12)

Breusch-Godfrey LM test for autocorrelation
---------------------------------------------------------------------------
    lags(p)  |          chi2               df                 Prob > chi2
-------------+-------------------------------------------------------------
       1     |         70.771               1                   0.0000
       2     |         70.773               2                   0.0000
       3     |         71.397               3                   0.0000
       4     |         71.398               4                   0.0000
       5     |         71.639               5                   0.0000
       6     |         71.641               6                   0.0000
       7     |         73.403               7                   0.0000
       8     |         73.536               8                   0.0000
       9     |         74.066               9                   0.0000
      10     |         74.112              10                   0.0000
      11     |         74.164              11                   0.0000
      12     |         76.742              12                   0.0000
---------------------------------------------------------------------------
                        H0: no serial correlation
```

Uporaba HAC standardnih napak

```
. newey hm1 ppr rvp rvv czp, lag(12)

Regression with Newey-West standard errors          Number of obs  =         96
maximum lag: 12                                     F(  4,     91) =     228.95
                                                    Prob > F       =     0.0000

------------------------------------------------------------------------------
             |             Newey-West
         hm1 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         ppr |   1.697766   .3914995     4.34   0.000     .9201006    2.475432
         rvp |  -311.6847   97.22559    -3.21   0.002    -504.8114    -118.558
         rvv |  -11.57513   7.305648    -1.58   0.117    -26.0869     2.936644
         czp |   11.50168   1.012866    11.36   0.000     9.489749    13.51362
       _cons |  -229.2038   118.4651    -1.93   0.056      -464.52    6.112544
------------------------------------------------------------------------------
```

Univerza *v Ljubljani*

EKONOMSKA FAKULTETA

6.5 Napotki ocenjevalcem in uporabnikom regresijskih modelov

EFMD EQUIS ACCREDITED

AACSB ACCREDITED

Motivacija

**Kdor ne dela nobenih napak,
ponavadi ne počne ničesar.**

W. C. Magee

Napotki ocenjevalcem in uporabnikom

Kaj gre lahko narobe ?	Kaj so posledice tega?	Kako to odkrijemo (ugotovimo)?	Kako to popravimo (odpravimo)?
Izpuščen konstantni člen regresijskega modela	➢ Pristranske ocene b_j. ➢ Nezanesljiv t–test.	➢ Prevera izpeljave in zapisa regresijskega modela.	➢ Vključitev konstantnega člena kot nadomestilo za vse ostale vplive.
Izpuščena relevantna spremenljivka	➢ Pristranske in nekonsistentne ocene b_j.	➢ Na podlogi teorije nepričakovani predznaki b_j. ➢ Nizka vrednost R^2.	➢ Vključitev relevantne spr. ali primerne nadomestne, če zanjo ni podatkov.
Vključena irelevantna spremenljivka	➢ Zmanjšana vrednost R^2_{adj}. ➢ Večje standardne napake in nizke vrednosti testnih statistik.	➢ Teorija. ➢ Testne statistike. ➢ Njena izključitev lahko vpliva na reg. koeficiente ostalih spremenljivk.	➢ Prevera ekonomske teorije. ➢ Izključitev spr., če ni eksplicitno zahtevana s strani ekonomske teorije.
Funkcijska oblika modela	➢ Zmanjšana zanesljivost modela. ➢ Pristranske in nekonsistentne ocene b_j.	➢ RESET test. ➢ Box-Cox test. ➢ Analiza razsevnega diagrama.	➢ Transformacija ene ali več spremenljivk ➢ Prevera skladnosti z ekonomsko teorijo.

Napotki ocenjevalcem in uporabnikom

Kaj gre lahko narobe ?	Kaj so posledice tega?	Kako to odkrijemo (ugotovimo)?	Kako to popravimo (odpravimo)?
Multi-kolinearnost	➢ Regresijski koeficienti niso pristranski. ➢ Visoke njihove standardne napake in nizke vrednosti *t*–statistike.	Testi: ➢ Kleinovo pravilo, ➢ *F*–test, ➢ *VIF* in *Tol*, ➢ število in indeksi pogojenosti.	➢ Prevera in previdna izločitev "prave" spremenljivke. ➢ Oblikovanje agregatne spremenljivke ➢ "Ne storiti ničesar" lahko še najboljše.
Hetero-skedastičnost	➢ Regresijski koeficienti niso pristranski. ➢ Standardne napake so pristranske in niso najnižje možne. ➢ Testne statistike so nezanesljive.	Testi: ➢ grafična metoda, ➢ Park, ➢ Glejser, ➢ Goldfeld-Quandt, ➢ Breusch-Pagan, ➢ White.	➢ Izboljšati specifikacijo. ➢ Transformacija spremenljivk. ➢ Uporaba PNK–TNK. ➢ Uporaba robustne cenilke variance.
Avtokorelacija	➢ Regresijski koeficienti niso pristranski. ➢ Standardne napake so pristranske in niso najnižje možne. ➢ Testne statistike so nezanesljive. ➢ Previsoke vrednosti R^2.	Testi: ➢ grafična metoda, ➢ test sekvenc, ➢ test asociacije, ➢ Durbin–Watson *d*, ➢ Wallis, ➢ Durbin *h* statistika, ➢ Breusch–Godfrey.	➢ Izboljšati specifikacijo. ➢ Uporaba PNK–PDE. ➢ Uporaba HAC cenilke variance. ➢ Uporaba Box-Jenkinsove ARIMA metodologije.

Heteroskedastičnost in avtokorelacija

Predpostavka	Homoskedastičnost	Odsotnost avtokorelacije
Oblika predpostavke	$Var(u_i) = E(u_i^2) = \sigma_u^2$	$Cov(u_i, u_j \mid x_i, x_j) = 0; \quad i \neq j$
Posledice predpostavke	$E(\mathbf{b}) = \boldsymbol{\beta}$	
	$\mathrm{Var}-\mathrm{cov}(\mathbf{u}) = \begin{bmatrix} \sigma_u^2 & 0 & \cdots & 0 \\ 0 & \sigma_u^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_u^2 \end{bmatrix} = \sigma_u^2 \cdot \mathbf{I}$	
	$\mathrm{Var}-\mathrm{cov}(\mathbf{b}) = \sigma_u^2 \cdot (\mathbf{X}'\mathbf{X})^{-1}$	

Heteroskedastičnost in avtokorelacija

Problem	Heteroskedastičnost	Avtokorelacija
Oblika problema	$Var(u_i) = E(u_i^2) = \sigma_{u_i}^2 \neq \sigma_u^2$	$Cov(u_i, u_j \mid x_i, x_j) \neq 0; \quad i \neq j$
Posledice problema	$E(\mathbf{b}) = \boldsymbol{\beta}$	
	$\mathrm{Var}-\mathrm{cov}(\mathbf{u}) = \mathbf{W} \neq \sigma_u^2 \cdot \mathbf{I}$	
	$\mathbf{W} = \begin{bmatrix} \sigma_{u_1}^2 & 0 & \cdots & 0 \\ 0 & \sigma_{u_2}^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_{u_n}^2 \end{bmatrix}$	$\mathbf{W} = \frac{\sigma_\varepsilon^2}{1-\rho_1^2} \begin{bmatrix} 1 & \rho_1 & \cdots & \rho_1^{T-1} \\ \rho_1 & 1 & \cdots & \rho_1^{T-2} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_1^{T-1} & \rho_1^{T-2} & \cdots & 1 \end{bmatrix}$
	$\mathrm{Var}-\mathrm{cov}(\mathbf{b}) = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \neq \sigma_u^2 \cdot (\mathbf{X}'\mathbf{X})^{-1}$	
Pomen za rezultate	Cenilka regresijskih koeficientov ostaja nepristanska. Cenilka regresijskih koeficientov ni več najbolj učinkovita. Cenilka variance slučajne spremenljivke postane pristranska. Cenilke varianc in kovarianc ocen regresijskih koeficientov so pristranske. $\Downarrow$ Nezanesljivo statistično sklepanje.	

Univerza v Ljubljani
EKONOMSKA FAKULTETA

Heteroskedastičnost in avtokorelacija

Problem	Heteroskedastičnost	Avtokorelacija
Pojavnost problema	Modeli presečnih podatkov; tudi modeli časovnih serij	Modeli časovnih serij; redko modeli presečnih podatkov
Odkrivanje problema	➢ grafična metoda ➢ Parkov test ➢ Glejserjev test ➢ Goldfeld–Quandtov test ➢ Breusch–Paganov test ➢ Harvey–Godfreyev test ➢ Whiteov test	➢ grafična metoda ➢ test sekvenc ➢ test asociacije ostankov ➢ Durbin–Watsonov test ➢ Wallisov test ➢ Durbinov *h*–test ➢ Ljung–Boxov test ➢ Breusch–Godfreyev test
Najcelovitejši pristop	Whiteov test	Breusch–Godfreyev test

EFMD EQUIS ACCREDITED

AACSB ACCREDITED

Heteroskedastičnost in avtokorelacija

Problem	Heteroskedastičnost	Avtokorelacija
Odpravljanje (ugotovljenega) problema	Izboljšanje slabe specifikacije modela (odpravlja nepravo heteroskedastičnost in nepravo avtokorelacijo)	
	Uporaba cenilk posplošenih najmanjših kvadratov (PNK):	
	cenilka tehtanih najmanjših kvadratov (TNK)	cenilka posplošene diferenčne enačbe (PDE): ➢ dvostopenjski postopek ➢ iterativna procedura (CORC)
	V kolikor ugotovimo točno obliko problema, odpravimo vse njegove zgoraj navedene neugodne posledice.	

Heteroskedastičnost in avtokorelacija

Problem	Heteroskedastičnost	Avtokorelacija
Odpravljanje (ugotovljenega) problema	Uporaba robustnih cenilk variance (Huber/White cenilka variance) $u \sim IID$ Cenilka sprosti predpostavko o identični porazdeljenosti.	Uporaba HAC cenilk variance (Newey–West robustna cenilka variance) $u \sim IID$ Cenilka sprosti obe predpostavki (o neodvisnosti in identični porazdeljenosti).
	Postopek ne vpliva na vrednosti ocen regresijskih koeficientov. Standardne napake spet postanejo nepristranske. Cenilka regresijskih koeficientov ne postane nujno spet najbolj učinkovita (standardne napake niso spet nujno najnižje možne).	
	Transformacija spremenljivk	AR(I)MAX metodologija

Univerza v Ljubljani

EKONOMSKA FAKULTETA

6. Preverjanje predpostavk klasičnega regresijskega modela

doc. dr. Miroslav Verbič

miroslav.verbic@ef.uni-lj.si

www.miroslav-verbic.si

EFMD EQUIS ACCREDITED

AACSB ACCREDITED

Ljubljana, februar 2014