Univerza v Ljubljani

EKONOMSKA FAKULTETA

2. Model multiple regresije

doc. dr. Miroslav Verbič

miroslav.verbic@ef.uni-lj.si

www.miroslav-verbic.si

EFMD EQUIS ACCREDITED

AACSB ACCREDITED

Ljubljana, februar 2014

Univerza v Ljubljani

EKONOMSKA FAKULTETA

2.1 Populacijski regresijski model in regresijski model vzorčnih podatkov

EFMD EQUIS ACCREDITED

AACSB ACCREDITED

Modeli in spremenljivke

MODEL MULTIPLE REGRESIJE
MULTIVARIATNI REGRESIJSKI MODEL

Modeli ene enačbe

odvisna spremenljivka	pojasnjevalne spremenljivke
pojasnjena spremenljivka	nepojasnjene spremenljivke
regresand	regresorji
prediktand	prediktorji
odzivna spremenljivka	kontrolne spremenljivke

Modeli več enačb

endogena(e) spremenljivka(e)	eksogene spremenljivke

Označevanje spremenljivk

$y_i, (y_t)$ – *vrednost odvisne spremenljivke pri i-ti (t-ti) opazovani enoti*

$x_{ji}, (x_{jt})$ – *vrednost j-te pojasnjevalne spremenljivke pri i-ti (t-ti) opazovani enoti*

$$i = 1, 2, \ldots, n \quad (t = 1, 2, \ldots, T)$$
$$j = 1, 2, \ldots, k$$

n (T) – število opazovanih enot

k – število parametrov

Populacijski regresijski model

POPULACIJSKI REGRESIJSKI MODEL
PROCES GENERIRANJA PODATKOV (DGP)

Primer: $PROIZVOD = f\,(DELO,\ KAPITAL)$

$$E(Q \mid L, K) = f\,(L, K)$$

Pogojna pričakovana vrednost odvisne spremenljivke je funkcija pojasnjevalnih spremenljivk

Pričakovana vrednost = matematično upanje = povprečna vrednost slučajne spremenljivke pri neskončno velikem številu merjenj ali realizacij

Populacijski regresijski model

Linearni populacijski regresijski model

(regresijski model = regresijska funkcija = regresijska enačba = regresija)

$$E(Q_i | L_i, K_i) = \beta_1 + \beta_2 L_i + \beta_3 K_i$$

$$Q_i = E(Q_i | L_i, K_i) + u_i$$

oziroma

$$u_i = Q_i - E(Q_i | L_i, K_i)$$

$$Q_i = \beta_1 + \beta_2 L_i + \beta_3 K_i + u_i$$

u_i = slučajna spremenljivka (odkloni)

Populacijski regresijski model

Razlogi za vključevanje u_i:

- Najprej je to nadomestilo za vse tiste spremenljivke, ki vplivajo na odvisno spremenljivko, pa *niso vključene med pojasnjevalne spremenljivke* (navedimo jih nekaj za naš primer). Pogosto je temu vzrok tudi pomanjkljiva ali nedorečena ekonomska teorija.
- Tudi če so nekatere spremenljivke sprejete in spoznane kot pomembne pojasnjevalne spremenljivke, jih pri specifikaciji modela ne moremo upoštevati, ker jih je *težko številčno izraziti ali pa zanje dobiti podatke* (npr. okus ali navade; v našem primeru denimo “idiosinkratično znanje").
- Če obstaja velika verjetnost, da je skupni učinek večine zanemarjenih pojasnjevalnih spremenljivk majhen in nepomemben in predvsem *nesistematičen*, potem lahko te vplive obravnavamo kot naključne in tedaj moramo v model vključiti spremenljivko *u*.

Populacijski regresijski model

- Tudi če bi uspeli pri specifikaciji modela upoštevati vse relevantne pojasnjevalne spremenljivke, bi še vedno ostali določeni naključni elementi (lahko rečemo *pravi slučajni vplivi*) pri vrednostih odvisne spremenljivke.
- Čeprav klasični regresijski model predpostavlja, da so vrednosti spremenljivk *izmerjene oziroma ugotovljene brez napak*, v praksi pogosto to ne velja.
- Pri specifikaciji regresijskega modela naj bi se držali znanega pravila, ki pravi, da naj ima preprostejši model vedno prednost pred zapletenejšim vse dotlej, dokler se ne dokaže, da je zaradi tega neustrezen (t.i. načelo Occamove britve).

Populacijski regresijski model

SPLOŠNI POPULACIJSKI REGRESIJSKI MODEL (PRM):

$$E(y|x_{1i},\ldots,x_{ki}) = f(x_{1i},\ldots,x_{ki})$$

oziroma

$$y_i = E(y|x_{1i},\ldots,x_{ki}) + u_i$$

Populacijski regresijski model

Linearni populacijski regresijski model:

$$E(y|x_{1i},\dots,x_{ki}) = \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_k x_{ki}$$

$$E(y|x_{2i},\dots,x_{ki}) = \beta_1 + \beta_2 x_{2i} \dots + \beta_k x_{ki}$$

oziroma

$$y_i = \beta_1 x_{1i} + \beta_2 x_{21} + \dots + \beta_k x_{ki} + u_i$$

$$y_i = \beta_1 + \beta_2 x_{2i} + \dots + \beta_k x_{ki} + u_i$$

β_j – *parcialni regresijski koeficient* j-**te pojasnjevalne premenljivke**

u_i – *slučajna spremenljivka (odkloni) pri i*-**ti opazovani enoti**

Vzorčni regresijski model

LINEARNI VZORČNI REGRESIJSKI MODEL (VRM):

$$y_i = b_1 + b_2 x_{2i} + \ldots + b_k x_{ki} + e_i$$
$$\hat{y}_i = b_1 + b_2 x_{2i} + \ldots + b_k x_{ki}$$
$$y_i = \hat{y}_i + e_i$$

$\hat{y}_i$ – cenilka pogojne pričakovane vrednosti $E(y|x_{1i},\ldots,x_{ki})$

$b_1,\ldots,b_k$ – cenilke regresijskih koeficientov $\beta_1,\ldots,\beta_k$

e_i – ostanki (reziduali) vzorčnega regresijskega modela

Izraz cenilka (estimator) se v statistični literaturi uporablja za vzorčne statistike. To so obrazci ali postopki, ki povejo, kako oceniti parametre za populacijo na podlagi vzorčnih podatkov (podatkov slučajnega vzorca). Določeno, specifično številčno vrednost parametra, ki smo jo izračunali s pomočjo cenilke, pa bomo imenovali ocena (estimate) parametra.

Vzorčni regresijski model

Linearni regresijski model vzorčnih podatkov:

$$\hat{y}_i = b_1 x_{1i} + b_2 x_{2i} + \ldots + b_k x_{ki}$$

$$\hat{y}_i = b_1 + b_2 x_{2i} + \ldots + b_k x_{ki}$$

oziroma

$$y_i = b_1 x_{1i} + b_2 x_{2i} + \ldots + b_k x_{ki} + e_i$$

$$y_i = b_1 + b_2 x_{2i} + \ldots + b_k x_{ki} + e_i$$

Univerza v Ljubljani

EKONOMSKA FAKULTETA

2.2 Ocenjevanje populacijskega regresijskega modela in metoda najmanjših kvadratov

EFMD EQUIS ACCREDITED

AACSB ACCREDITED

Motivacija

"Edini način do situacije, v kateri bo naša znanost lahko nudila uporabne nasvete v širšem obsegu politikom in podjetnikom, vodi skozi kvantitativno delo. Dokler nismo sposobni pretvoriti naših argumentov v številke, glas naše znanosti, čeprav lahko občasno pomaga preprečiti velike napake, ne bo nikoli slišan s strani praktičnih ljudi. Le-ti so, instiktivno, ekonometriki, vsakdo od njih, v njihovem nezaupanju do vsega, kar ni podvrženo eksaktnemu preverjanju."

Joseph A. Schumpeter: "The Common Sense of Econometrics", *Econometrica*, 1, 1933, str. 12.

Ocenjevanje PRM

Kako naj cenilka minimizira ostanke VRM:

- **vsota ostankov naj bo najmanjša možna?**
- **vsota absolutnih vrednosti ostankov naj bo najmanjša možna?**
- **vsota kvadratov vrednosti ostankov naj bo najmanjša možna?**

Metoda najmanjših kvadratov – MNKVD (OLS, LS)

"Odkril" jo je nemški matematik C. F. Gauss (pri 16. letih). Objavil 1809. leta in ji dal dokončno obliko 1823. leta

Ocenjevanje PRM

Carl Friederich Gauss (1777 – 1855)

Ocenjevanje PRM

$$y_i = b_1 + b_2 x_{2i} + \ldots + b_k x_{ki} + e_i = \hat{y}_i + e_i$$

Ostanki (reziduali) vzorčnega regresijskega modela so enaki:

$$e_i = y_i - \hat{y}_i =$$

$$= y_i - b_1 - b_2 x_{2i} \ldots - b_k x_{ki}$$

$$\sum \mathrm{e}_\mathrm{i}^2 = \sum (y_i - b_1 - b_2 x_{2i} \ldots - b_k x_{ki})^2$$

Očitno je, da je ta vsota funkcija (S) cenilk regresijskih koeficientov, torej:

$$\sum \mathrm{e}_\mathrm{i}^2 = S(b_1, \ldots, b_k)$$

Metoda najmanjših kvadratov

Linearni vzorčni regresijski model (VRM):

$$y_i = b_1 + b_2 x_{2i} + \ldots + b_k x_{ki} + e_i$$

$$\begin{aligned} y_1 &= b_1 + b_2 x_{21} + \ldots + b_k x_{k1} + e_1 \\ y_2 &= b_1 + b_2 x_{22} + \ldots + b_k x_{k2} + e_2 \\ &\vdots \qquad\qquad \vdots \\ y_n &= b_1 + b_2 x_{2n} + \ldots + b_k x_{kn} + e_n \end{aligned}$$

Metoda najmanjših kvadratov

Matrični zapis VRM in izpeljava cenilke MNKVD:

$$\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{e}$$

$$\mathbf{e} = \mathbf{y} - \mathbf{X}\mathbf{b}$$

$$\mathbf{e}^T\mathbf{e} = (\mathbf{y} - \mathbf{X}\mathbf{b})^T(\mathbf{y} - \mathbf{X}\mathbf{b})$$

Metoda najmanjših kvadratov

$$\frac{\partial \mathbf{e}^{\mathrm{T}}\mathbf{e}}{\partial \mathbf{b}} = -2\mathbf{X}^{\mathrm{T}}\mathbf{y} + 2\mathbf{X}^{\mathrm{T}}\mathbf{X}\mathbf{b} = \mathbf{0}$$

$$\mathbf{X}^{\mathrm{T}}\mathbf{y} = \mathbf{X}^{\mathrm{T}}\mathbf{X}\mathbf{b}$$

Cenilka regresijskih koeficientov po metodi najmanjših kvadratov:

$$\boxed{\mathbf{b} = (\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{y}}$$

Značilnosti metode najmanjših kvadratov

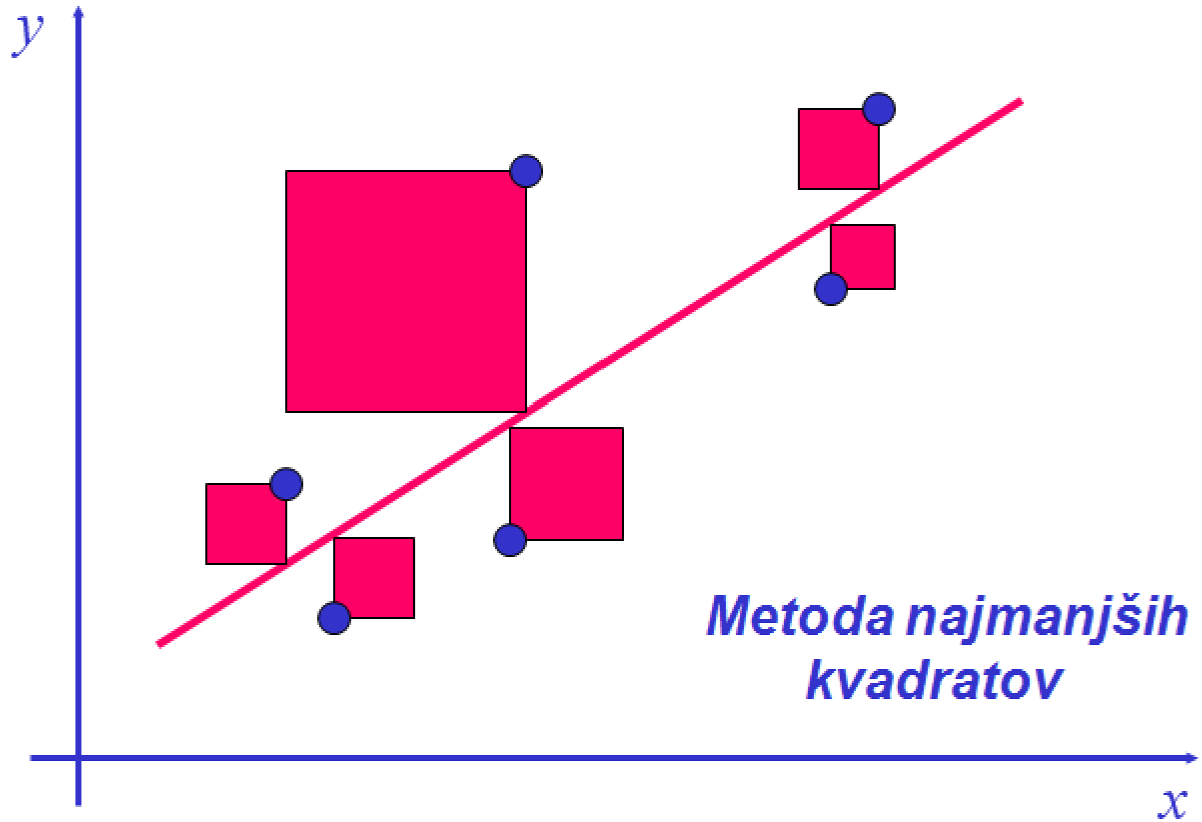
1. $\mathbf{X}^{\mathrm{T}}\mathbf{e} = \mathbf{0}$

2. $\hat{\mathbf{y}}^{\mathrm{T}}\mathbf{e} = \mathbf{0}$

3. $\overline{\hat{y}} = \overline{y}$

4. $\sum_i e_i = 0$

Grafična ponazoritev MNKVD



Univerza v Ljubljani

EKONOMSKA FAKULTETA

2.3 Predpostavke metode najmanjših kvadratov

EFMD EQUIS ACCREDITED

AACSB ACCREDITED

1. predpostavka PRM

Linearnost regresijskega modela:

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

2. predpostavka PRM

Fiksne (nestohastične) vrednosti pojasnjevalnih spremenljivk pri ponovitvah vzorcev:

POGOJNA REGRESIJSKA ANALIZA

3. predpostavka PRM

Ničelna povprečna vrednost u_i:

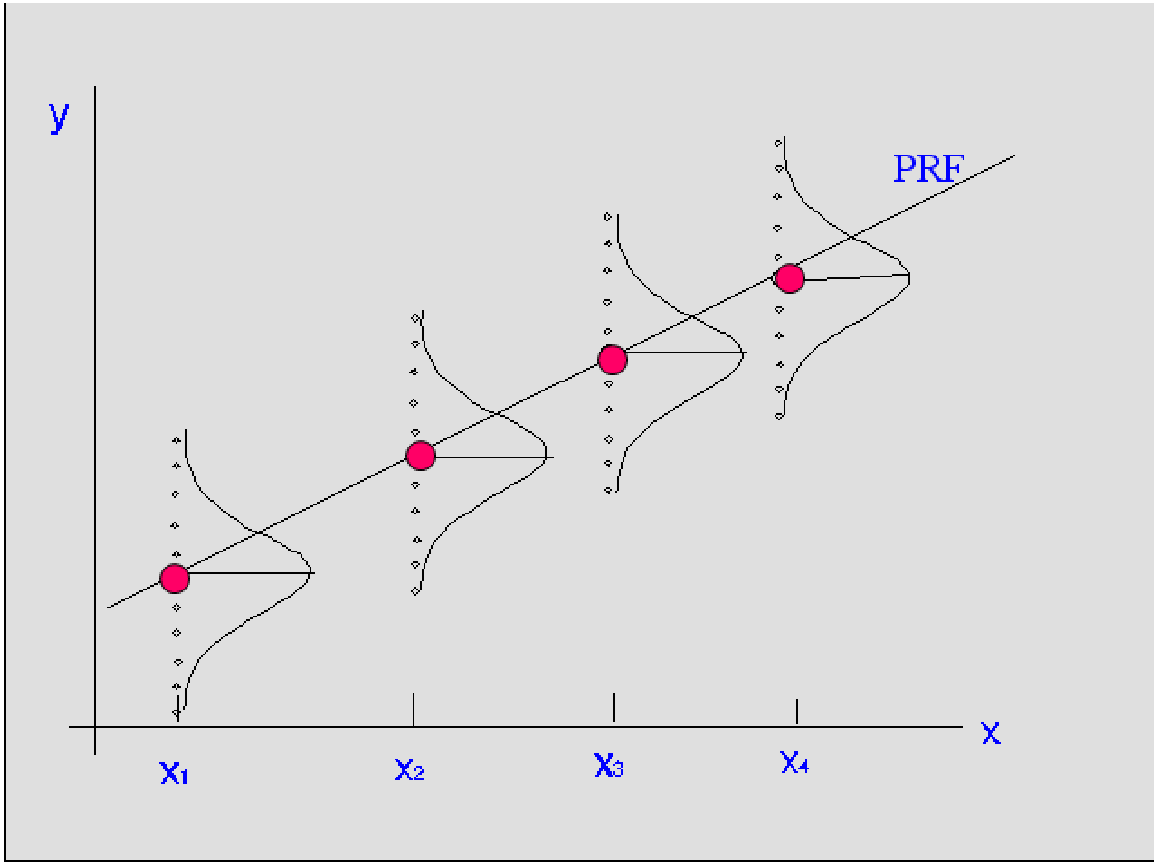
$$E(u_i \mid x_{2i},\ldots,x_{ki}) = 0 \quad \text{ali na kratko } E(u_i) = 0;$$

v tem primeru velja:

$$E(y_i \mid x_{2i},\ldots,x_{ki}) = \beta_1 + \beta_2 x_{2i} + \ldots + \beta_k x_{ki}.$$

3. predpostavka PRM

Porazdelitev slučajnih spremenljivk y in u:



4. predpostavka PRM

Homoskedastičnost:

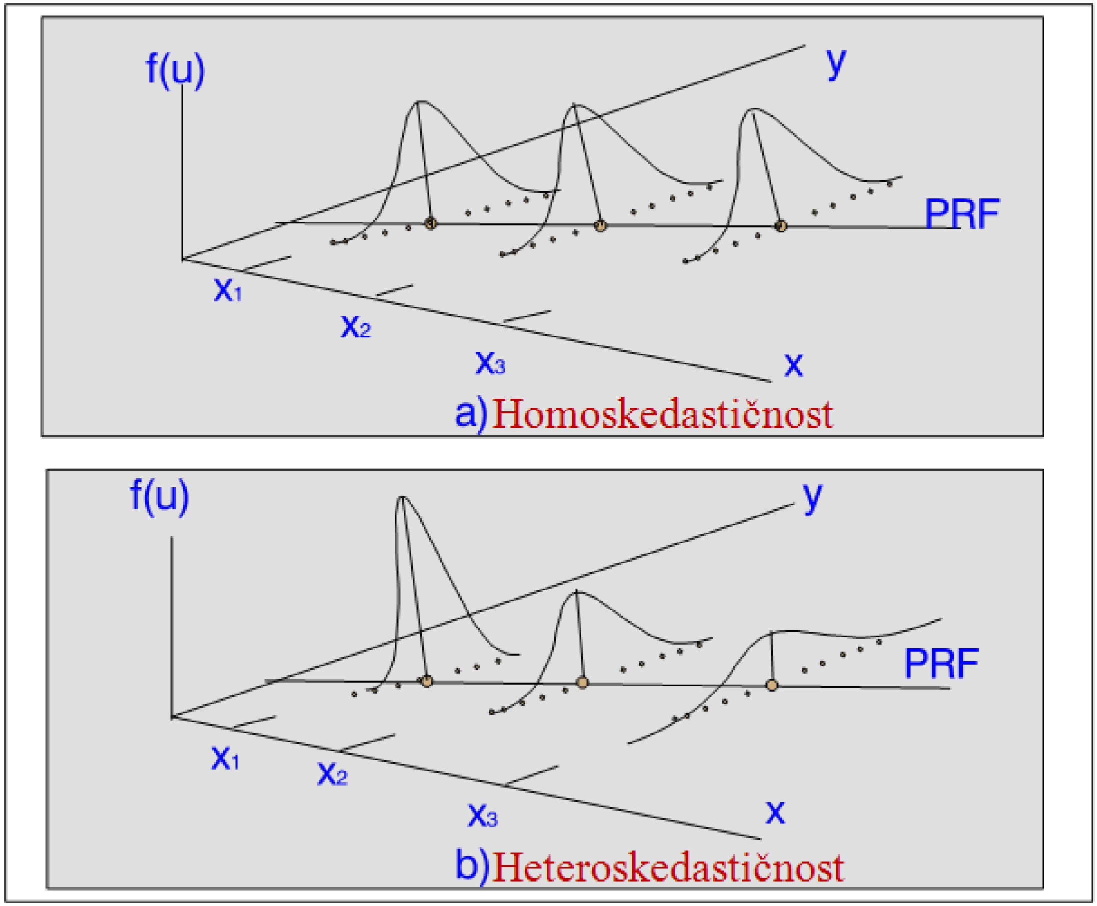
$$Var(u_i|x_i) = E\left[\left(u_i - E(u_i|x_i)\right)^2 \middle| x_i\right] = E\left[u_i^2 \middle| x_i\right] = E(u_i^2) = \sigma^2$$

oziroma:

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

4. predpostavka PRM

Variabilnost slučajne spremenljivke *u*:



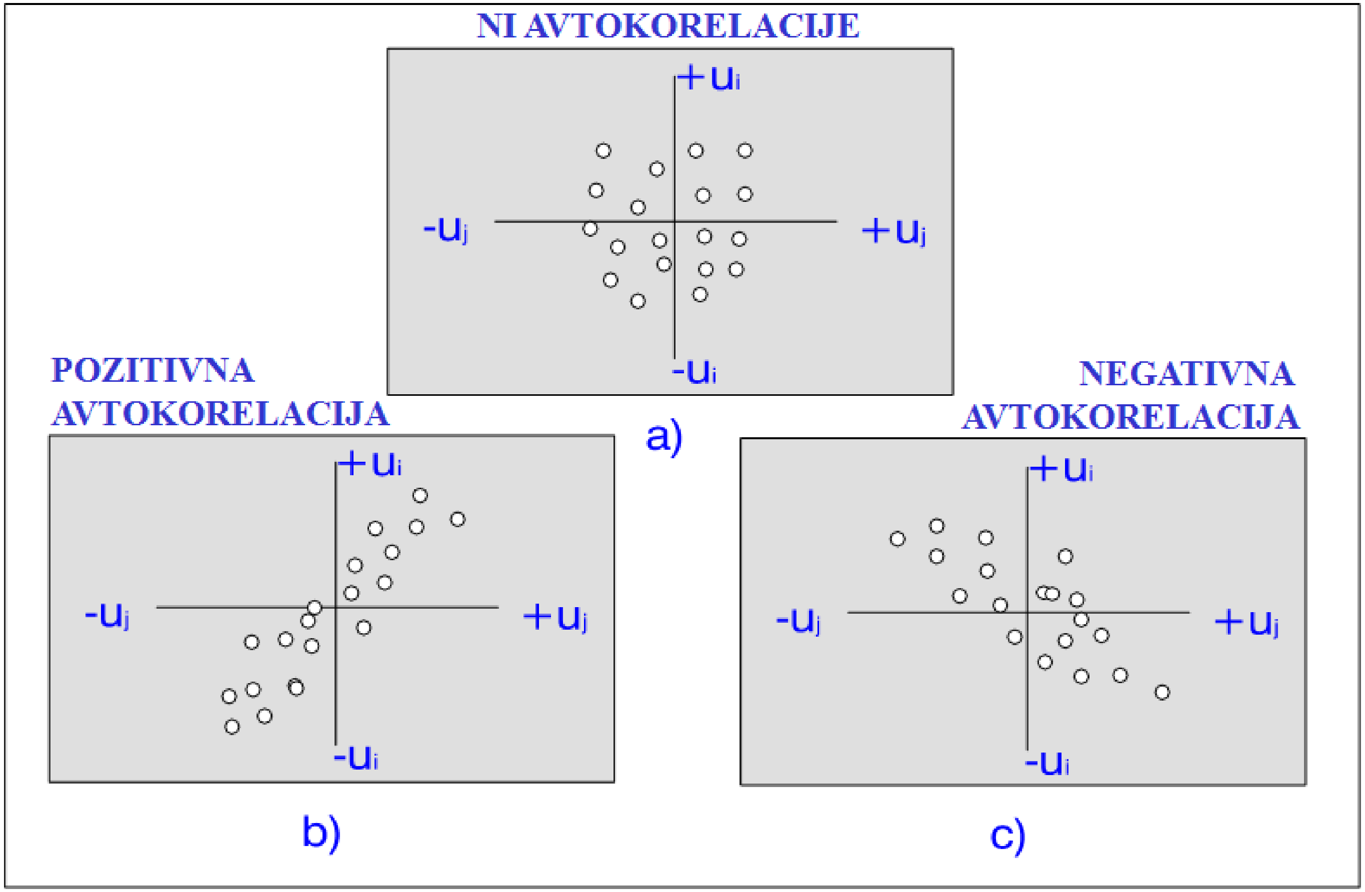
5. predpostavka PRM

Odsotnost avtokorelacije:

$$Cov(u_i, u_j \mid x_i, x_j) = 0; \quad i \neq j$$

5. predpostavka PRM

Razsevni diagrami za vrednosti slučajne spremenljivke u:



6. predpostavka PRM

Nekoreliranost med pojasnjevalnimi spremenljivkami in slučajno spremenljivko *u*:

$$Cov(x_2, u) = Cov(x_3, u) = \ldots = Cov(x_k, u) = 0$$

7. predpostavka PRM

Število opazovanj mora presegati število ocenjenih parametrov oziroma pojasnjevalnih spremenljivk:

$$n > k$$

8. predpostavka PRM

Variabilnost vrednosti pojasnjevalnih spremenljivk:

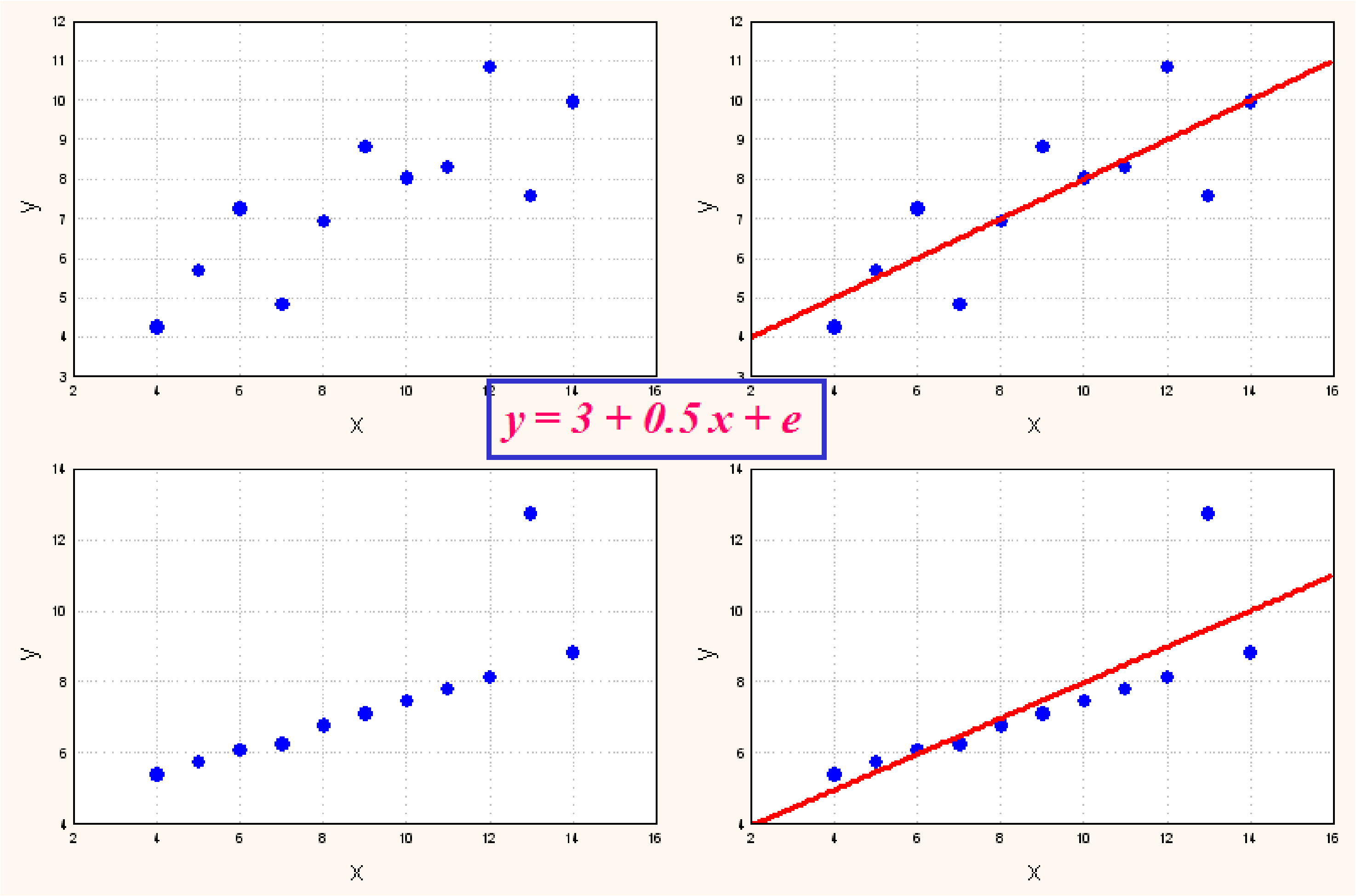
var(*X*) je končno pozitivno število

9. predpostavka PRM

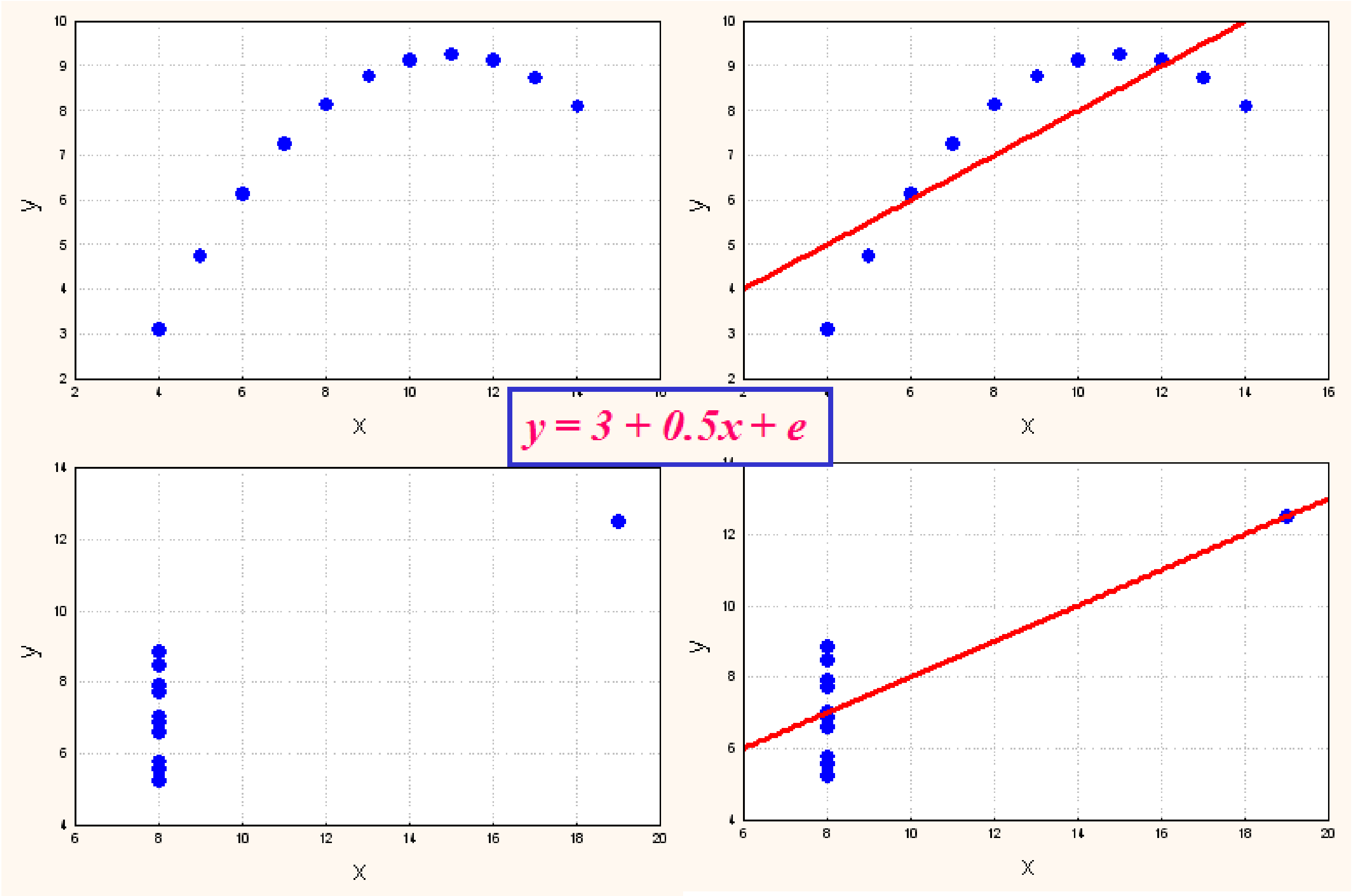
Regresijski model je pravilno specificiran:

- **vključene vse relevantne pojasnjevalne spremenljivke**
- **izbrana ustrezna funkcijska oblika modela**

9. predpostavka PRM



9. predpostavka PRM



$y = 3 + 0.5x + e$

10. predpostavka PRM

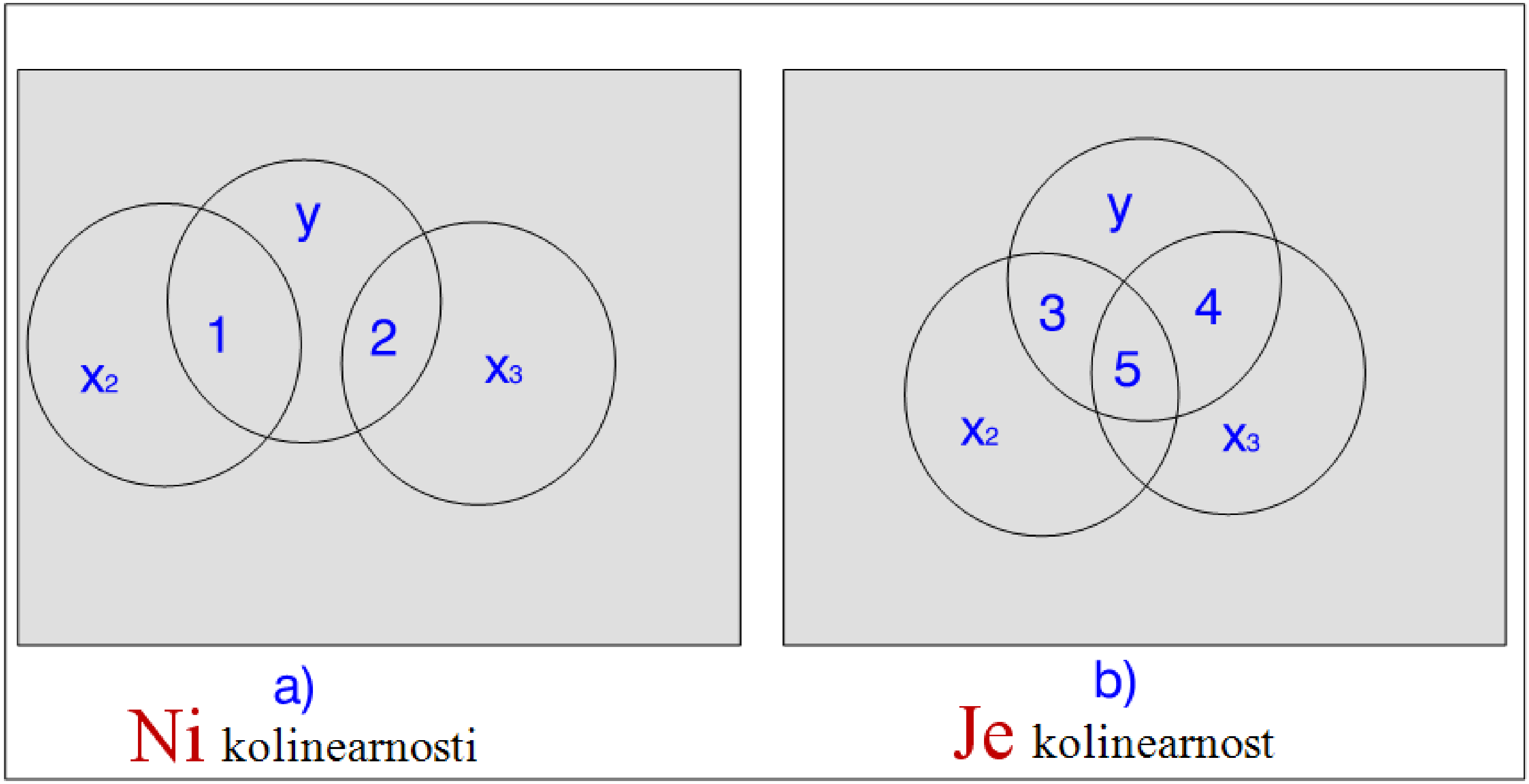
Odsotnost popolne multikolinearnosti:

med pojasnjevalnimi spremenljivkami ne obstaja popolna *linearna* odvisnost oblike:

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

10. predpostavka PRM

Analogija z Vennovimi diagrami:



Operacionalizacija PRM

Slučajna spremenljivka *u* je normalno porazdeljena:

$$u_i \sim N(0, \sigma_u^2)$$

Posledično je odvisna spremenljivka *y* normalno porazdeljena slučajna spremenljivka:

$$y_i \sim N(\beta_1 x_{1i} + \ldots + \beta_k x_{ki}, \sigma_u^2)$$

Univerza v Ljubljani

EKONOMSKA FAKULTETA

2.4 Vzorčne značilnosti in lastnosti metode najmanjših kvadratov

EFMD EQUIS ACCREDITED

AACSB ACCREDITED

Motivacija

- Z metodo najmanjših kvadratov dobljene ocene regresijskih koeficientov so *slučajne spremenljivke*. Kako ugotoviti njihovo povprečno vrednost, varianco, kovariance in verjetnostno porazdelitev?

- Metoda najmanjših kvadratov je le ena od možnih cenilk regresijskih koeficientov. Kako *dobra je izbrana cenilka* v primerjavi z ostalimi možnimi metodami?

- Kako je z *zanesljivostjo ocen* regresijskih koeficientov? Kako dobre so ocene, ki smo jih dobili na podlagi le enega vzorca?

Vzorčne značilnosti MNKVD

ZNAČILNOST A: LINEARNOST

$$\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$$

Vrednosti ocen regresijskih koeficientov na podlagi vzorčnih podatkov so linearna kombinacija vzorčnih vrednosti odvisne spremenljivke *y*.

ZNAČILNOST B: KONSISTENTNOST

$$\mathbf{b} \xrightarrow{n \to \infty} \beta$$

Vrednosti ocen regresijskih koeficientov težijo k pravim vrednostim parametrov, ko se povečuje vzorec.

Vzorčne značilnosti MNKVD

ZNAČILNOST C: NEPRISTRANSKOST

$$E(\mathbf{b}) = \boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'E(\mathbf{u})$$

$$= \boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{0} = \boldsymbol{\beta}$$

Vrednosti ocen regresijskih koeficientov so nepristranske ocene njihovih populacijskih vrednosti.

Vzorčne značilnosti MNKVD

ZNAČILNOST D: UČINKOVITOST

$$Var-Cov(\mathbf{b}) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}$$

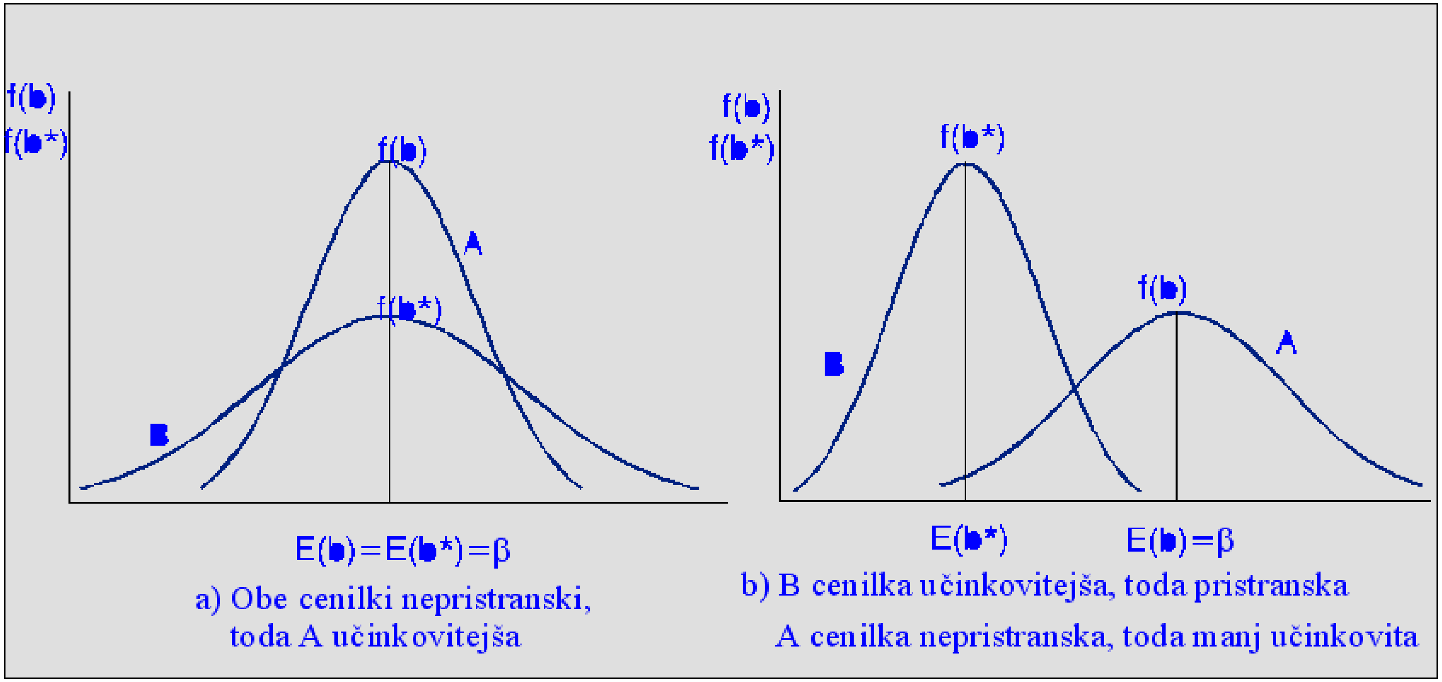
Vrednosti ocen regresijskih koeficientov so učinkovite.

$$Var-Cov(\mathbf{b}) = \begin{bmatrix} Var(b_1) & Cov(b_1,b_2) & \cdots & Cov(b_1,b_k) \\ Cov(b_1,b_2) & Var(b_2) & \cdots & Cov(b_2,b_k) \\ \vdots & \vdots & \ddots & \vdots \\ Cov(b_k,b_1) & Cov(b_k,b_2) & \cdots & Var(b_k) \end{bmatrix}$$

Vzorčne značilnosti MNKVD

Porazdelitev ocen regresijskih koeficientov na podlagi velikega števila ponovljenih vzorcev je normalna:

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



Vzorčne značilnosti MNKVD

Varianca cenilke regresijskega koeficienta:

$$Var(b_2) = \frac{\sigma^2}{nVar(x_2)} \cdot \frac{1}{(1 - r^2_{x_2 x_3})}$$

Varianca je tem manjša:

1. čim večji je vzorec (n), torej čim več opazovanih enot vključuje izračun ocene regresijskih koeficientov;
2. čim večja je variabilnost (varianca) pojasnjevalne spremenljivke x;
3. čim manjša je variabilnost (varianca) slučajne spremenljivke u;
4. čim manjša je linearna odvisnost med pojasnjevalnimi spremenljivkami.

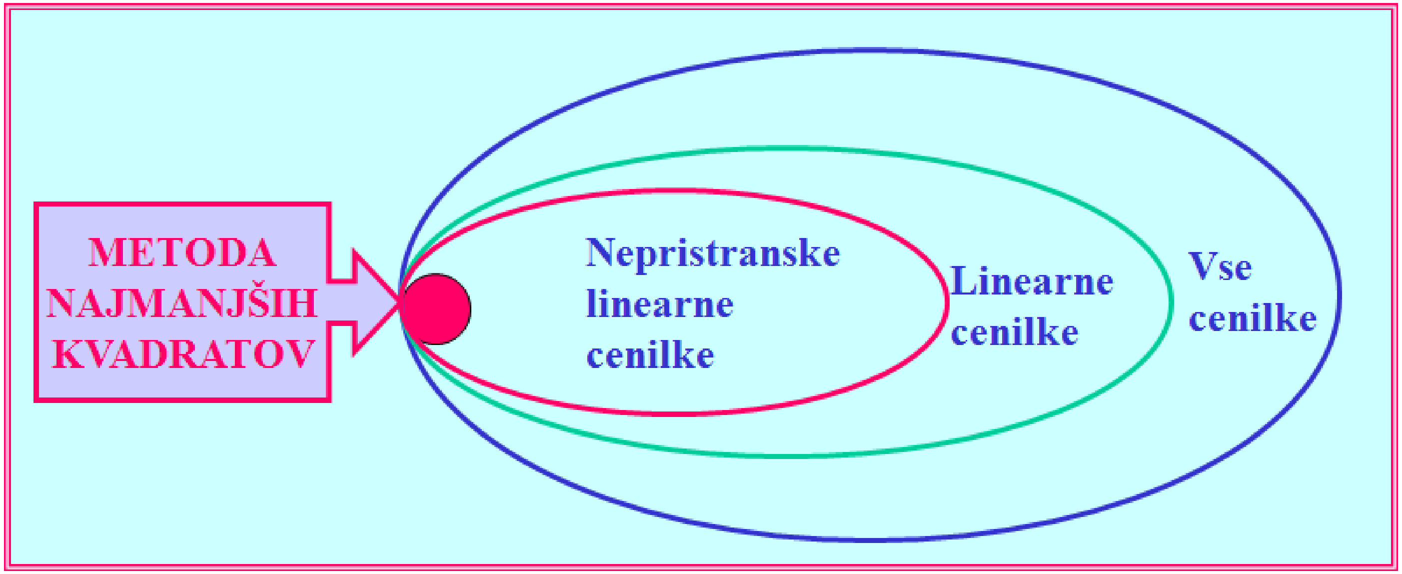
Gauss–Markov teorem

Metoda najmanjših kvadratov je

NEpristranska **NA**jboljša **LI**nearna **CE**nilka

NENALICE

(**B**est **L**inear **U**nbiased **E**stimator - **BLUE**)



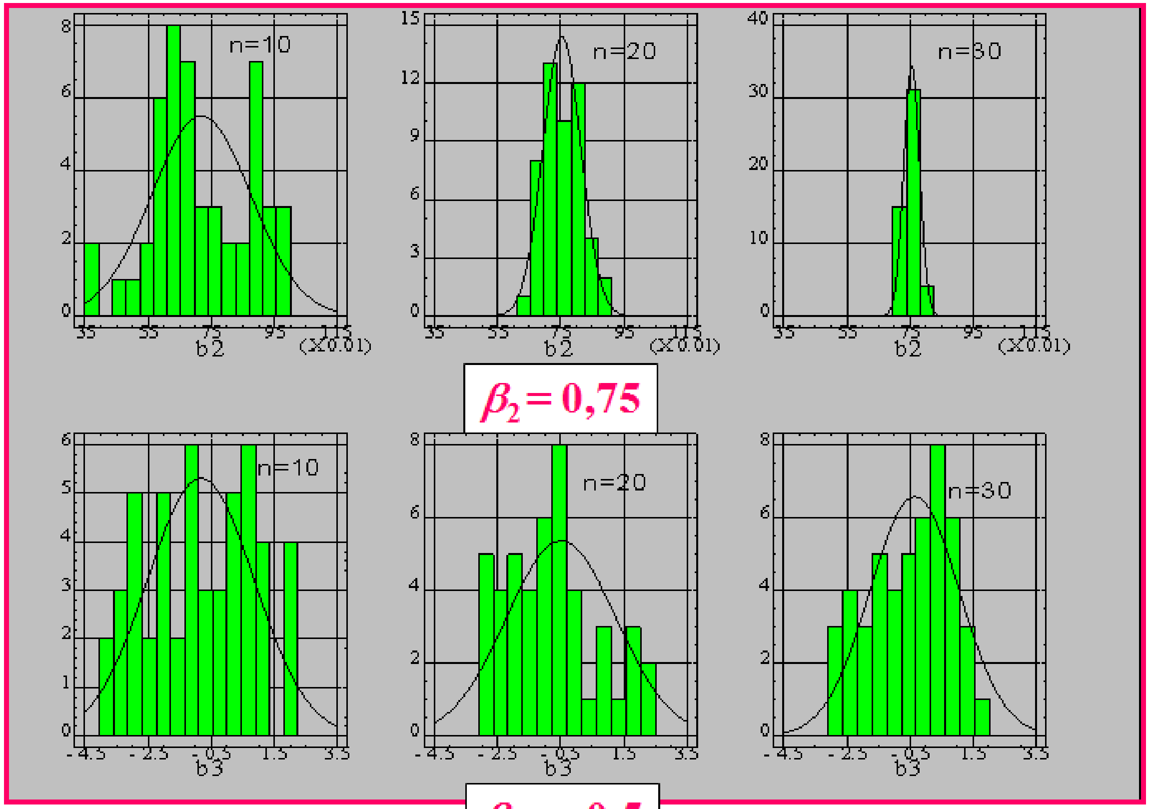
Monte Carlo eksperimenti

Monte Carlo eksperimenti
(Los Alamos – Manhattan)

- **Določimo konkretno obliko PRM;**
- **Določimo vrednosti regresijskih koeficientov, vrednosti pojasnjevalnih spremenljivk in izračunamo vrednosti odvisne spremenljivke;**
- **Pripravimo "popravljene" vrednosti odvisne spremenljivke za večje število vzorcev tako, da jim pri vsakem vzorcu prištejemo naključne vrednosti slučajne spremenljivke *u*;**
- **Za vsak od pripravljenih vzorcev izračunamo ocene regresijskih koeficientov z MNKVD;**
- **Pripravimo porazdelitve ocen regresijskih koeficientov in jih analiziramo.**

Monte Carlo eksperimenti

Porazdelitev ocen regresijskih koeficientov
(50 vzorčnih modelov; $n = 10$, $n = 20$, $n = 30$)



$\beta_2 = 0{,}75$

$\beta_3 = -0{,}5$

Cenilka variance slučajne spremenljivke *u*

$$\mathbf{e} = \mathbf{y} - \mathbf{X}\mathbf{b}$$

$$E(\mathbf{e}^{\mathrm{T}}\mathbf{e}) = \sigma^2 (n-k)$$

$$E\left(\frac{\mathbf{e}^{\mathrm{T}}\mathbf{e}}{n-k}\right) = \sigma^2$$

$$\frac{\mathbf{e}^{\mathrm{T}}\mathbf{e}}{n-k} = s_e^2$$

s_e^2 nepristranska cenilka σ^2

Cenilka var-cov matrike **b**

$$\mathrm{var-cov}(\mathbf{b}) = s_e^2 (\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}$$

pri čemer je

$$s_e^2 = \frac{\mathbf{e}^{\mathrm{T}}\mathbf{e}}{n-k} = \frac{(\mathbf{y}-\mathbf{Xb})^{\mathrm{T}}(\mathbf{y}-\mathbf{Xb})}{n-k}$$

$$\boxed{s_e^2 = \frac{\mathbf{y}^{\mathrm{T}}\mathbf{y} - \mathbf{y}^{\mathrm{T}}\mathbf{Xb}}{n-k}}$$

Porazdelitev ocen regr. koeficientov

$$\mathbf{b} \sim N\left[\boldsymbol{\beta}, \sigma^2 (\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\right]$$

$$b_j \sim N(\beta_j, Var(b_j))$$

$$z = \frac{b_j - \beta_j}{\sqrt{Var(b_j)}} \sim N(0,1)$$

Porazdelitev ocen regr. koeficientov

Izračun *t*–statistike na podlagi vzorčnih podatkov

$$t_j = \frac{b_j - \beta_j}{\sqrt{\text{var}(b_j)}} = \frac{b_j - \beta_j}{se(b_j)}$$

Porazdelitev ocen regr. koeficientov

William Sealy Gosset (1876 – 1937)

Porazdelitev ocen regr. koeficientov

Studentova *t*–porazdelitev:

$$se(b_j) = \sqrt{\operatorname{var}(b_j)} = \sqrt{s_e^2 \left[(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1} \right]_{jj}}$$

$$\boxed{t_j = \frac{b_j - \beta_j}{se(b_j)} \sim t_{(n-k)}}$$

Univerza v Ljubljani

EKONOMSKA FAKULTETA

2.5 Mere primernosti oziroma zanesljivosti regresijskega modela

EFMD EQUIS ACCREDITED

AACSB ACCREDITED

Standardna napaka ocene regresije

Standardna napaka ocene regresijskega modela
Standardna napaka ocene regresije

$$s_e = \sqrt{s_e^2} = \sqrt{\frac{\mathbf{e}^{\mathrm{T}}\mathbf{e}}{n-k}}$$

Koeficient variacije

$$KV = \frac{s_e}{\bar{y}} \quad \text{ali} \quad KV\% = \frac{s_e}{\bar{y}}100$$

Analiza variance (ANOVA)

Dekompozicija vsote kvadratov (*VK*)
[angl. *sum of squares* – *SS*]:

$$SVK = PVK + NVK$$

$$SVK = \sum_{i=1}^{n} (y_i - \bar{y})^2 = \sum_{i=1}^{n} y_i^2 - n\bar{y}^2 = \mathbf{y}'\mathbf{y} - n\bar{y}^2$$

$$PVK = \sum_{i=1}^{n} \left(\hat{y}_i - \bar{y}\right)^2 = \sum_{i=1}^{n} \hat{y}_i^2 - n\bar{y}^2 = \hat{\mathbf{y}}'\hat{\mathbf{y}} - n\bar{y}^2$$

$$PVK = \mathbf{b}'\mathbf{X}'\mathbf{y} - n\bar{y}^2 = \mathbf{y}'\mathbf{X}\mathbf{b} - n\bar{y}^2$$

$$NVK = \sum_{i=1}^{n} \left(y_i - \hat{y}_i\right)^2 = \sum_{i=1}^{n} e_i^2 = \mathbf{e}'\mathbf{e} = \mathbf{y}'\mathbf{y} - \hat{\mathbf{y}}'\hat{\mathbf{y}}$$

Determinacijski koeficient

Determinacijski koeficient multiple regresije
Multipli determinacijski koeficient

$$R^2 = \frac{\text{PVK}}{\text{SVK}} = 1 - \frac{\text{NVK}}{\text{SVK}}$$

$$R^2 = \frac{\hat{\mathbf{y}}'\hat{\mathbf{y}} - n\bar{y}^2}{\mathbf{y}'\mathbf{y} - n\bar{y}^2} = \frac{\mathbf{b}'\mathbf{X}'\mathbf{y} - n\bar{y}^2}{\mathbf{y}'\mathbf{y} - n\bar{y}^2}$$

$$\boxed{R_*^2 = \frac{\hat{\mathbf{y}}'\hat{\mathbf{y}}}{\mathbf{y}'\mathbf{y}} = \frac{\mathbf{b}'\mathbf{X}'\mathbf{y}}{\mathbf{y}'\mathbf{y}}}$$

$$R^2 = (r_{y\hat{y}})^2 = r_{y\hat{y}}^2$$

Determinacijski koeficient

Popravljeni determinacijski koeficient (H. Theil, 1971)

SVK = PVK + NVK

SVK = (SVK - NVK) + NVK

Stopinje prostosti:

$$(n - 1) = ((n - 1) - (n - k)) + (n - k)$$

$$(n - 1) = (k - 1) + (n - k)$$

$$\bar{R}^2 = 1 - \frac{\dfrac{\text{NVK}}{(n-k)}}{\dfrac{\text{SVK}}{(n-1)}} = 1 - \frac{\text{NVK}}{\text{SVK}} \frac{(n-1)}{(n-k)}$$

$$\boxed{\bar{R}^2 = 1 - (1 - R^2)\frac{(n-1)}{(n-k)}}$$

Determinacijski koeficient

Slabosti determinacijskega in popravljenega determinacijskega koeficienta:

- visoka vrednost determinacijskega koeficienta še ne pomeni, da smo v model vključili prave pojasnjevalne spremenljivke;
- vrednosti determinacijskih koeficientov niso primerljive med modeli z različno definirano odvisno spremenljivko;
- vrednosti determinacijskih koeficientov so v splošnem večje pri modelih časovnih vrst kot pri modelih presečnih podatkov;
- nizka vrednost determinacijskega koeficienta ne pomeni, da model ne vključuje pravih, pomembnih pojasnjevalnih spremenljivk.

Testiranje modela kot celote

Testiranje statistične značilnosti regresijskega modela kot celote:

R. A. Fisher je 1922. leta ugotovil, da se razmerje med pojasnjeno in nepojasnjeno varianco, ob upoštevanju stopinj prostosti, porazdeljuje v posebni porazdelitvi, po njem imenovani ***F*–porazdelitvi**.

Testiranje modela kot celote

Ronald Aylmer Fisher (1890 – 1962)

Testiranje modela kot celote

$$F = \frac{\dfrac{\text{PVK}}{k-1}}{\dfrac{\text{NVK}}{n-k}} \sim F_{(k-1,n-k)}$$

$$F = \frac{\dfrac{\text{PVK}}{SVK(k-1)}}{\dfrac{\text{NVK}}{SVK(n-k)}} = \frac{R^2/(k-1)}{(1-R^2)/(n-k)} \sim F_{(k-1,n-k)}$$

Testiranje modela kot celote

$$F = \frac{R^2/(k-1)}{(1-R^2)/(n-k)}$$

$$F > F_k$$

$$R^2 > R^2{}_k$$

$$R_k^2 = \frac{F_k(k-1)}{F_k(k-1)+(n-k)}$$

Funkcija verjetja in informacijski kriteriji

Vrednost logaritma funkcije verjetja

$$\ln L = -\frac{n}{2}\left[\ln(2\pi) + \ln(\frac{\text{NVK}}{n}) + 1\right]$$

Akaikejev informacijski kriterij (AIC)

$$AIC = -2 \ln L + 2k$$

Schwarzov kriterij (SC ali BIC)

$$SC = -2 \ln L + k \ln(n)$$

Univerza v Ljubljani

EKONOMSKA FAKULTETA

2.6 Razlaga regresijskih koeficientov

EFMD EQUIS ACCREDITED

AACSB ACCREDITED

Predstavitev rezultatov ocenjevanja

Zapis oziroma predstavitev rezultatov ocenjevanja regresijskega modela

$$y = \beta_1 + \beta_2 x_2 + \cdots + \beta_k x_k + u$$

$$\begin{array}{ccccccc}
\hat{y} & = & b_1 & + & b_2 x_2 & + \quad \cdots & + b_k x_k \\
 & & (se(b_1)) & & (se(b_2)) & & (se(b_k)) \\
 & & (t_1) & & (t_2) & \cdots & (t_k) \\
 & & (p_1) & & (p_2) & \cdots & (p_k)
\end{array}$$

$$n = \ldots. \qquad R^2 = \ldots \qquad \bar{R}^2 = \ldots\ldots$$

$$s_e = \ldots.. \qquad (F; DW; h)$$

Izpis rezultatov v programskem paketu

Izpis rezultatov ocenjevanja regresijskega modela linearne produkcijske funkcije

```
. regress q l k

      Source |       SS       df       MS              Number of obs =      81
-------------+------------------------------           F(  2,    78) =   52.90
       Model |  6.9350e+12     2  3.4675e+12           Prob > F      =  0.0000
    Residual |  5.1130e+12    78  6.5551e+10           R-squared     =  0.5756
-------------+------------------------------           Adj R-squared =  0.5647
       Total |  1.2048e+13    80  1.5060e+11           Root MSE      = 2.6e+05

------------------------------------------------------------------------------
           q |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           l |   9687.383   3640.852     2.66   0.009     2439.003    16935.76
           k |    2.27941   .7553228     3.02   0.003      .775678    3.783142
       _cons |  -11875.29   34865.13    -0.34   0.734    -81286.43    57535.85
------------------------------------------------------------------------------
```

Izpis rezultatov v programskem paketu

Izpis rezultatov ocenjevanja regresijskega modela linearizirane Cobb-Douglasove produkcijske funkcije

```
. regress lq ll lk

      Source |       SS       df       MS              Number of obs =      81
-------------+------------------------------           F(  2,    78) =  190.75
       Model |  178.261263     2  89.1306313           Prob > F      =  0.0000
    Residual |    36.44752    78  .467275898           R-squared     =  0.8302
-------------+------------------------------           Adj R-squared =  0.8259
       Total |  214.708783    80  2.68385978           Root MSE      =  .68358

------------------------------------------------------------------------------
          lq |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          ll |   .9645479   .1199229     8.04   0.000     .7257997    1.203296
          lk |   .1885438   .0673358     2.80   0.006     .0544886     .322599
       _cons |   7.546026   .4617465    16.34   0.000      6.62676    8.465293
------------------------------------------------------------------------------
```

Definicija regresijskega koeficienta

$$E(y|x_2,\dots,x_k) = \beta_1 + \beta_2 x_2 \dots + \beta_k x_k$$

$$\beta_j = \frac{\partial E(y|x_2,\dots,x_k)}{\partial x_j} \quad ; \quad j = 2,3,\dots,k$$

Multipli regresijski koeficienti
Koeficienti multiple regresije
Parcialni regresijski koeficienti ali parcialni smerni koeficienti

$$\beta_1 = E(y|x_2 = 0,\dots,x_k = 0)$$

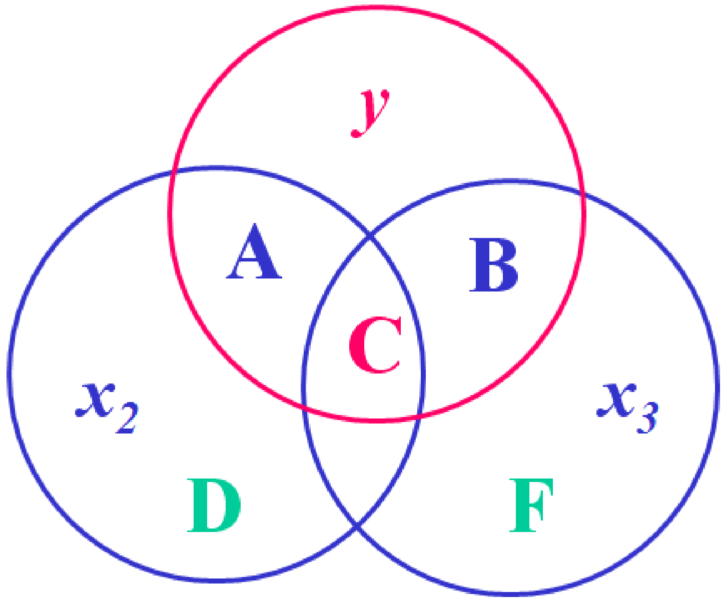
Definicija regresijskega koeficienta

Regresijski koeficienti predstavljajo
čiste, neposredne ali neto učinke

$$\hat{y}_i = b_1 + b_2 x_{2i} + b_3 x_{3i}$$

$$\hat{x}_{2i} = c_1 + c_2 x_{3i}$$

$$e_{2i} = x_{2i} - \hat{x}_{2i}$$



$$\hat{x}_{3i} = f_1 + f_2 x_{2i}$$

$$e_{3i} = x_{3i} - \hat{x}_{3i}$$

$$\hat{y}_i = d_1 + d_2 e_{2i} + d_3 x_{3i}$$

$$\hat{y}_i = g_1 + g_2 x_{2i} + g_3 e_{3i}$$

Dekompozicija skupnih vplivov

Vplivi pojasnjevalne spremenljivke x_2

Neposredni (čisti, direktni) vpliv : $b_2 = d_2$

Posredni (indirektni) vpliv
(vpliv x_2 na x_3 in preko nje na y) : $f_2 \cdot b_3$

Skupni (celotni) vpliv : $g_2 = d_2 + f_2 \cdot b_3$

Vplivi pojasnjevalne spremenljivke x_3

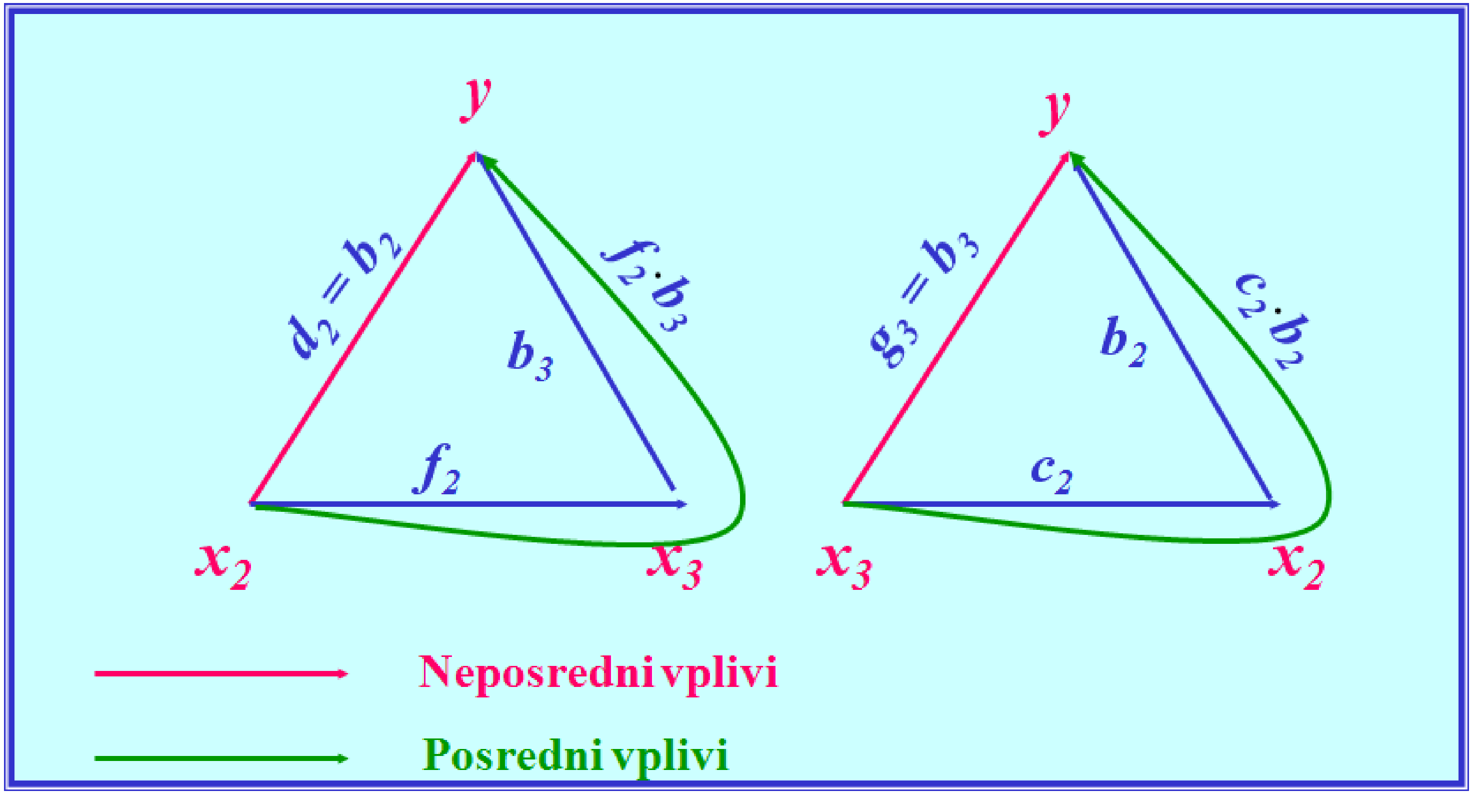
Neposredni (čisti, direktni) vpliv : $b_3 = g_3$

Posredni (indirektni) vpliv
(vpliv x_3 na x_2 in preko nje na y) : $c_2 \cdot b_2$

Skupni (celotni) vpliv : $d_3 = g_3 + c_2 \cdot b_2$

Dekompozicija skupnih vplivov

Grafična ponazoritev neposrednih in posrednih vplivov pojasnjevalnih spremenljivk na odvisno spremenljivko



Frisch – Waugh (– Lovell) teorem

Frisch – Waugh teorem
(Econometrica, 1933)

$$y_i = b_1 + b_2 x_{2i} + b_3 x_{3i} + e_i$$

$$y_i = c_1 + c_2 x_{2i} + e_{y_i} \quad \Rightarrow \quad \tilde{y}_i = e_{y_i} = y_i - \hat{y}_i$$

$$x_{3i} = d_1 + d_2 x_{2i} + e_{x_{3i}} \quad \Rightarrow \quad \tilde{x}_{3i} = e_{x_{3i}} = x_{3i} - \hat{x}_{3i}$$

$$\tilde{y}_i = g_1 + b_3 \tilde{x}_{3i} + e_i$$

Frisch – Waugh (– Lovell) teorem

PRIMER UPORABE:

Metoda individualnega trenda

Prvi korak

Iz proučevanih časovnih vrst izločimo "motečo" sestavino

Drugi korak

Ocenimo regresijski model, v katerem nastopajo "očiščene" časovne vrste

Frisch – Waugh (– Lovell) teorem

$$y_t = \beta_1 + \beta_2 x_{2t} + \beta_3 x_{3t} + \ldots + \beta_k x_{kt} + u_t$$

Prvi korak

$$y_t = c_1 + c_2 t + c_3 t^2 + \ldots + e_{y_t} \quad ; \quad e_{y_t} = y_t - \hat{y}_t$$

$$x_{2t} = c_{21} + c_{22} t + c_{23} t^2 + \ldots + e_{x_{2t}} \quad ; \quad e_{x_{2t}} = x_{2t} - \hat{x}_{2t}$$

$$\ldots$$

$$x_{kt} = c_{k1} + c_{k2} t + c_{k3} t^2 + \ldots + e_{x_{kt}} \quad ; \quad e_{x_{kt}} = x_{kt} - \hat{x}_{kt}$$

Drugi korak

$$e_{y_t} = b_1 + b_2 e_{x_{2t}} + b_3 e_{x_{3t}} + \ldots + b_k e_{x_{kt}} + e_t$$

Univerza v Ljubljani

EKONOMSKA FAKULTETA

2. Model multiple regresije

doc. dr. Miroslav Verbič

miroslav.verbic@ef.uni-lj.si

www.miroslav-verbic.si

EFMD EQUIS ACCREDITED

AACSB ACCREDITED

Ljubljana, februar 2014