4EK211 Základy ekonometrie

Heteroskedasticita náhodné složky

Cvičení 6

Gauss-Markovy předpoklady

Náhodná složka: Gaussovy-Markovovy předpoklady

1. $E(\boldsymbol{u}) = 0$

2. $E(\boldsymbol{u}\,\boldsymbol{u}^T) = \boldsymbol{\sigma}^2 \boldsymbol{I}_n$ ← kovarianční matice, tj. rozptyl $\boldsymbol{u}$.
   - konečný a konstantní rozptyl = homoskedasticita
   - prvky na diagonále kovarianční matice konstantní ($\boldsymbol{\sigma}^2$)

     → porušení: heteroskedasticita

   - náhodné složky jsou sériově nezávislé
   - nuly mimo diagonálu kovarianční matice

     → porušení: autokorelace

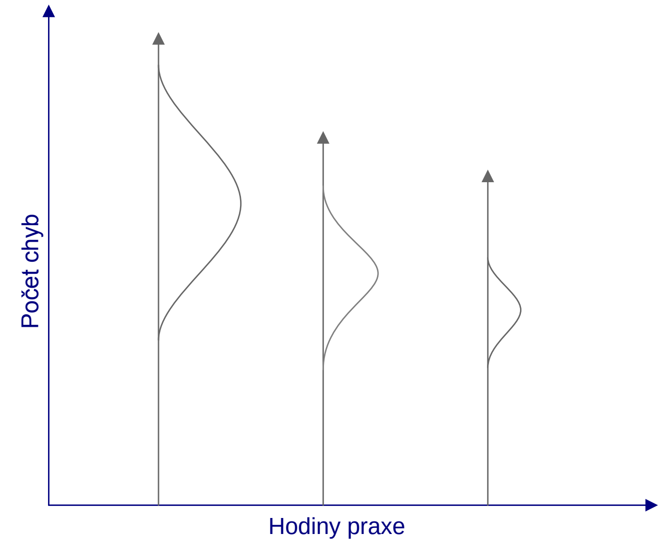
3. $\boldsymbol{X}$ je nestochastická matice – $E(\boldsymbol{X}^T\boldsymbol{u}) = 0$

4. $\boldsymbol{X}$ má plnou hodnost k+1
   - matice $\boldsymbol{X}$ neobsahuje žádné perfektně lineárně závislé sloupce

Heteroskedasticita - obecně

- **Heteroskedasticita**: rozptyl náhodné složky: σ^2 není konečný a konstantní, obvykle σ^2 je funkcí některé exogenní proměnné
- náhodná složka může mít v případě heteroskedasticity odlišný rozptyl pro každé pozorování:

$$E(u_i^2) = \sigma_i^2 \neq \text{konst}$$



Příklad

y = počet chyb při psaní na stroji

x = počet hodin strávených cvičením

y* = *f*(*x*) + *u

- čím více hodin cvičení – tím méně chyb
- rozptyl chybovosti je větší pro skupinu lidí s nižší praxí
  - někdo se učí rychleji a už od počátku dělá méně chyb než ti, kteří se učí pomaleji a na začátku dělají spoustu chyb
  - s rostoucím počtem hodin praxe se schopnosti jednotlivců začínají sbližovat a rozptyl chybovosti se snižuje

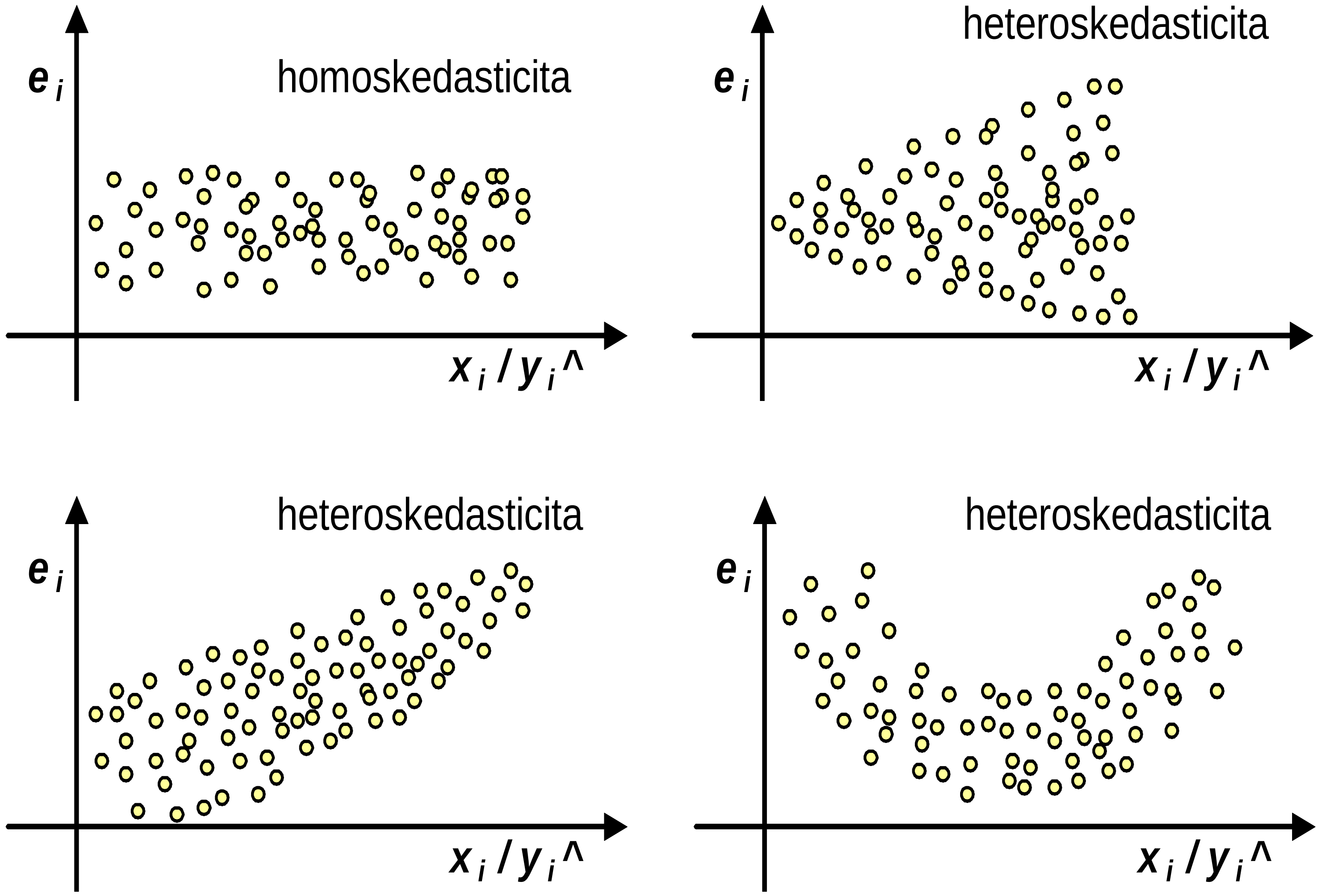
Heteroskedasticita - příčiny

- chybná specifikace modelu
  - vynechání podstatné vysvětlující proměnné
  - nevhodná funkční forma modelu
- odhad z průřezových dat se značnou variabilitou ve výběru
  - variabilita endogenní proměnné (a tedy i reziduí) může být závislá na magnitudě některé exogenní proměnné
- chyby měření
  - s rostoucí hodnotou endogenní proměnné dochází ke kumulaci chyb měření – to zvyšuje rozptyl endogenní proměnné a tedy i rozptyl reziduí
- odhad z upravených dat
  - odhad nikoliv na původních pozorováních, ale např. ze skupinových průměrů získaných z tříděných dat

Heteroskedasticita - důsledky

- bodové odhady parametrů
  - zůstávají **nevychýlené** a **konzistentní**
  - nemají však minimální rozptyl – tj. **nejsou vydatné** a **ani asymptoticky vydatné**
- odhady směrodatných chyb bodových odhadů (s_{bi}) a rozptylu sigma (s^2) jsou **vychýlené**
  - intervalové odhady nejsou spolehlivé
  - statistické testy (*t*-testy, *F*-test) ztrácejí na síle
  - Nelze předem říci, zda je odhadnutá hodnota s^2, resp. s_{bi} podhodnocená nebo nadhodnocená

Heteroskedasticita graficky (eyeballing test)



Heteroskedasticita – neparametrické testy

Spearmanův test korelace pořadí

- zkoumá korelaci pořadí mezi jednou vysvětlující proměnou a rezidui
- test je nutné udělat pro každou vysvětlující proměnnou LRM zvlášť!!!
- vypočítáme koeficient pro konkrétní výběr, pak testujeme statistickou významnost pro „populaci“

Postup

1. Absolutní hodnoty reziduí $|e_i|$ seřadíme vzestupně a očíslujeme
2. Pořadové číslo přiřadíme k původním (tj. nesrovnaným) reziduím
3. Absolutní hodnoty exogenní proměnné $|x_i|$ seřadíme vzestupně a očíslujeme
4. Pořadové číslo přiřadíme k původním (tj. nesrovnaným) hodnotám x_i
5. Spočítáme rozdíly v pořadí reziduí a pozorování:

   d_i = pořadí $|e_i|$ - pořadí $|x_i|$ …. rozdíl pořadí
6. Spočítáme Spearmanův koeficient korelace pořadí:

$$r_{e,x} = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n(n^2-1)}$$

Heteroskedasticita – neparametrické testy

7. Vyhodnocení:

- **$|r_{e,x}|$** $\rightarrow$ **0** (resp. $|r_{e,x}| < 0{,}8 - 0{,}9$)
  … je možné očekávat **homoskedasticitu**
- **$|r_{e,x}|$** $\rightarrow$ **1** (resp. $|r_{e,x}| > 0{,}8 - 0{,}9$)
  … je možné očekávat **heteroskedasticitu**
- test statistické významnosti koeficientu (spočteného pro daný výběr)
- pomocí *t*-statistiky:

$$t = r_{e,x}\sqrt{\frac{n-k-1}{1-r_{e,x}^2}} \approx t_{(n-k-1)}$$

Testovaná hypotéza:

H_0: homoskedasticita

H_1: heteroskedasticita

- vypočtená t hodnota $> t^*_{1-\alpha/2}$ (n-k-1) $\rightarrow$ zamítneme H_0
- vypočtená t hodnota $\leq t^*_{1-\alpha/2}$ (n-k-1) $\rightarrow$ nezamítáme H_0

Heteroskedasticita – neparametrické testy – příklad

Soubor: CV6_PR1.xls

Data: ***y*** = průměrný roční výnos cenného papíru
x = riziko cenného papíru (směrodatná odchylka)

Zadání: Odhadněte závislost průměrného ročního výnosu cenného papíru (y) na riziku (x).
Vyhodnoťte heteroskedasticitu s využitím Spearmanova koeficientu korelace pořadí pro $\alpha = 0{,}05$.

$y_i = \beta_0 + \beta_1 x_i + u_i, \qquad i = 1, 2, \ldots, 10$

Heteroskedasticita – neparametrické testy

Goldfeldův-Quandtův test

Postup:

1. zvolíme statisticky významnou proměnnou X_i a seřadíme datový soubor (všechna pozorování) vzestupně podle této proměnné
2. rozdělíme data na dvě stejné poloviny (přitom kolem „středu“ řady vynecháme $\boldsymbol{q}$ hodnot: $\boldsymbol{q \leq n/4}$): pro „malé“ a „velké“ hodnoty X_i
3. vypočteme stupně volnosti $\boldsymbol{v}$

$$\boldsymbol{v} = \frac{\boldsymbol{n} - \boldsymbol{q}}{2} - \boldsymbol{k} - 1$$

4. vypočteme $\boldsymbol{F(v,v)}$ statistiku (odhad 2 modelů v PcGive)

$$\boldsymbol{F(v,v)} = \frac{\boldsymbol{S}_2}{\boldsymbol{S}_1}, \quad \textbf{kde} \quad \boldsymbol{S}_j = \sum_{j=1,2} \boldsymbol{e}_j^2$$

5. Vyhodnocení - t**estovaná hypotéza:**

H_0: homoskedasticita

H_1: heteroskedasticita

$F(v,v) > F^*(v,v) \rightarrow$ zamítáme H_0 ve prospěch heteroskedasticity

$F(v,v) < F^*(v,v) \rightarrow$ nezamítáme H_0

Heteroskedasticita – neparametrické testy – příklad

Soubor: CV6_PR2.xls

Data: ***y*** = spotřební výdaje (tis. USD/rok)
x = disponibilní příjem (tis. USD/rok)

Zadání: Odhadněte závislost spotřebních výdajů (y) na disponibilním příjmu (x).

Vyhodnoťte heteroskedasticitu

- graficky
- s využitím testu Goldfelda-Quandta pro α = 0,05
- uvažujte logaritmickou transformaci modelu a vyhodnoťte heteroskedasticitu s využitím testu Goldfelda-Quandta pro α = 0,05 (pomůcka – zlogaritmujte proměnnou y a x pomocí funkce **log(VAR)** v PcGivu)

$y_i = \beta_0 + \beta_1 x_i + u_i$, $i = 1, 2,\ldots,30$

Heteroskedasticita – parametrické testy

- testy s pomocnou regresí
- většinou potřebujeme $n \geq 30$

Parkův test

- podle Parka je vztah mezi rozptylem a proměnnou (která způsobuje heteroskedasticitu) následovný (pomocná regrese):

$$\sigma_i^2 = \beta_0 X_i^{\beta_1} e^{v_i}$$

- po zlogaritmování (zajímá nás pouze významnost odhadu β_1):

$$\ln \sigma_i^2 = \beta_0 + \beta_1 \ln X_i + v_i$$

- Rozptyl náhodné složky není pozorovatelný, použijeme pomocnou regresi přes (přímo pozorovatelná) rezidua:

$$\ln e_i^2 = \beta_0 + \beta_1 \ln X_i + v_i$$

- parametry modelu odhadneme pomocí MNČ a t-testem vyhodnotíme významnost β_1 → H_0: homoskedasticita (β_1 není významný)
  H_1: heteroskedasticita (β_1 je významný)

Heteroskedasticita – parametrické testy

Glejserův test

- pomocná regrese: absolutní hodnota reziduí a různé formy závislosti:

$$|e_i| = \beta_0 + \beta_1 X_i + v_i$$

$$|e_i| = \beta_1 X_i + v_i$$

$$|e_i| = \beta_0 + \beta_1 \frac{1}{\sqrt{X_i}} + v_i$$

$$|e_i| = \beta_0 + \beta_1 \sqrt{X_i} + v_i$$

$$|e_i| = \sqrt{\beta_0 + \beta_1 X_i + v_i}$$

$$|e_i| = \sqrt{\beta_0 + \beta_1 X_i^2 + v_i} \qquad i = 1, 2, \ldots, n$$

- parametry modelu odhadneme pomocí MNČ a *t*-testem vyhodnotíme významnost β_1 → H_0: homoskedasticita (β_1 není významný)
  H_1: heteroskedasticita (β_1 je významný)

Heteroskedasticita – parametrické testy

Whiteův test

- testovaný LRM: $Y_i = \beta_0 + \beta_1 X_{1i} + \beta_1 X_{2i} + u_i$

  pomocná regrese:

  $e_t^2 = f(x_1, x_2, x_1^2, x_2^2, x_1 * x_2) + v$

- pomocná regrese obeně:

  $e_t^2 = f$ (exogenní proměnné, jejich čtverce, jejich párové násobky)

- statistika $n^* R^2 \approx \chi^2(k\text{-}1)$
  - n = rozsah souboru
  - k = počet parametrů pomocné regrese

Testovaná hypotéza:

H_0: homoskedasticita H_1: heteroskedasticita

$n^* R^2 > \chi^{*2}_{\alpha}(k\text{-}1)$ ... zamítáme H_0 ve prospěch heteroskedasticity
obvykle Whiteův test vyhodnocujeme pomocí p-value
(na hl. α = 5% zamítáme H_0, je-li p-value menší než 0,05)

Heteroskedasticita – parametrické testy – příklady

Soubor: CV6_PR3.xls

Data: ***vydaje*** = průměrné měsíční výdaje placené kreditní kartou (v USD)
vek = věk (v letech)
prijem **=** příjem (v tis. USD)

Zadání: Odhadněte závislost výdajů na věku a příjmu.
Vyhodnoťte heteroskedasticitu s využitím Whiteova testu pro α = 0,01.

$vydaje_i = \beta_0 + \beta_1 vek_i + \beta_2 prijem_i + u_i, \quad i = 1, 2, \ldots, 72$