Přednáška III
AKM I

Lukáš Frýd

EVROPSKÁ UNIE
Evropské strukturální a investiční fondy
Operační program Výzkum, vývoj a vzdělávání

MINISTERSTVO ŠKOLSTVÍ, MLÁDEŽE A TĚLOVÝCHOVY

Přepoklady KLM a Gauss – Markov teorém

KLM
Klasický lineární model

Blue odhad - GM

Nezkreslený odhad

1) Lineární v parametrech

2) NÁHODNÝM výběrem

3) $E(\epsilon|X) = E(\epsilon) = 0$

4) Není dokonalá multikolinearita

5) $Var(\epsilon|X) = E(\epsilon\epsilon') = \sigma^2 I$

6) $\epsilon|X \sim N(0, \sigma^2 I)$

Potom platí, že:

$$b \sim N(\beta, \sigma^2 (X'X)^{-1})$$

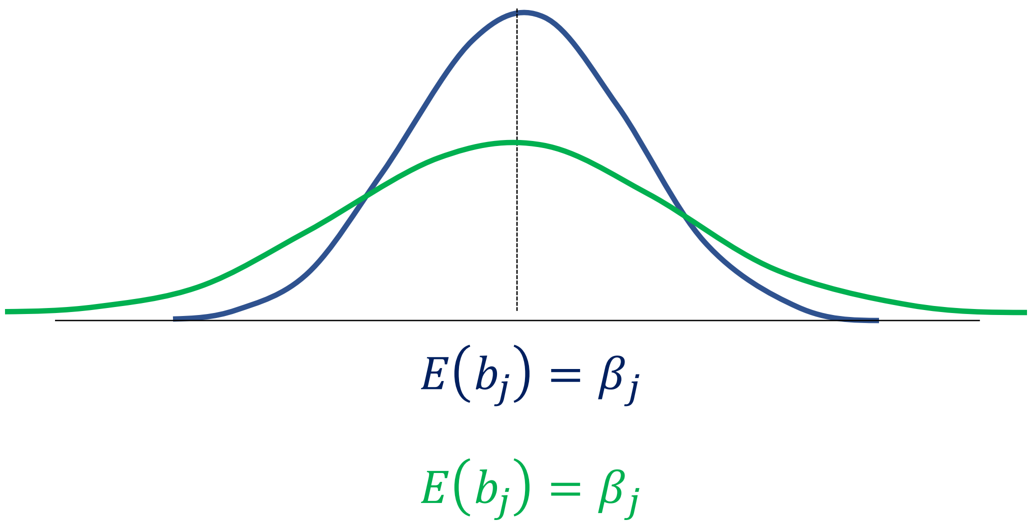
LRM 5. Homoskedasticita

Nás samozřejmě zajímá jaký rozptyl má výběrové rozdělení pro b
A tedy i rozptyl. Odhad rozptylu je dán:

$$Var(b_j) = \frac{\sigma^2}{\sum(x_j - \overline{x_j})^2 (1 - R_j^2)}$$

σ^2 –rozptyl náhodné složky, neznám, musím také odhadnout

$$\widehat{\sigma^2} = \frac{1}{n-k-1} \sum e^2$$



Požaduji, aby platilo: $E(\widehat{\sigma^2}) = \sigma^2$

$$1)\ Var(b_j) = \frac{\widehat{\sigma^2}}{\sum\left(x_j - \overline{x_j}\right)^2 (1 - R_j^2)}$$

Odhad rozptylu náhodné složky a $Var(b)$ je nezkreslený a konzistentní pokud platí $Var(\epsilon|X) = \sigma^2 I$
Tedy požaduji homoskedasticitu

V případě heteroskedasticity bude odhad $Var(b)$ vychýlený a nekonzistentní.

Rozptyl $Var(b_j)$ se vyskytuje jak u t-testu, tak u testů pro hodnocení sdružené významnosti (F, Wald)
Tedy pokud není splněn předpoklad homoskedasticity, tak nelze brát výsledky těchto testů jako relevantní!!!

Homoskedasticita

Podmíněný rozptyl je konstatní

$$Var(\varepsilon|x) = \sigma^2 I$$

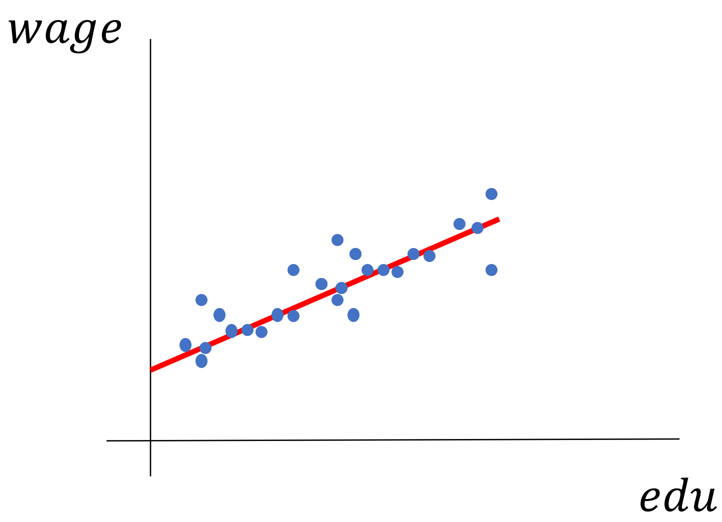
Rozptyl ε, který je dán x, se nemění v závislosti se změnami x

Rozptyl y, který je dán x, se nemění v závislosti se změnami x

Platí

$$Var(\varepsilon|x) = Var(y|x) = \sigma^2 I$$

$$Var(\varepsilon_i|edu_i) = \sigma^2$$



Heteroskedasticita

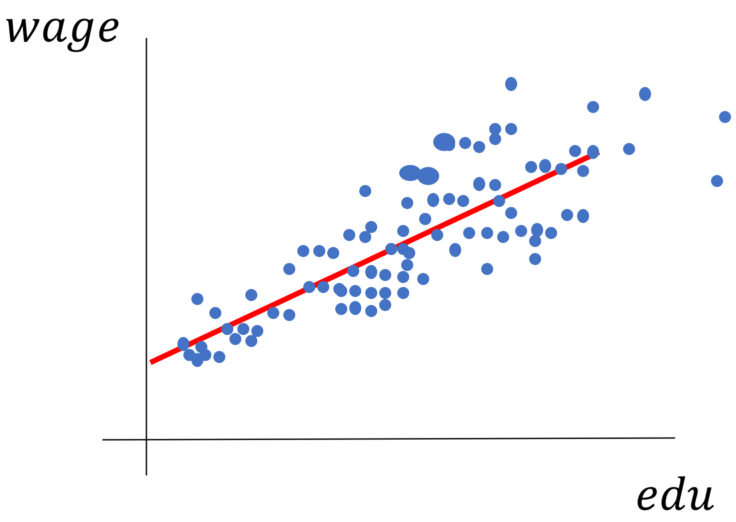
Podmíněný rozptyl NENÍ konstatní

Rozptyl ε, který je dán x, se mění v závislosti se změnami x

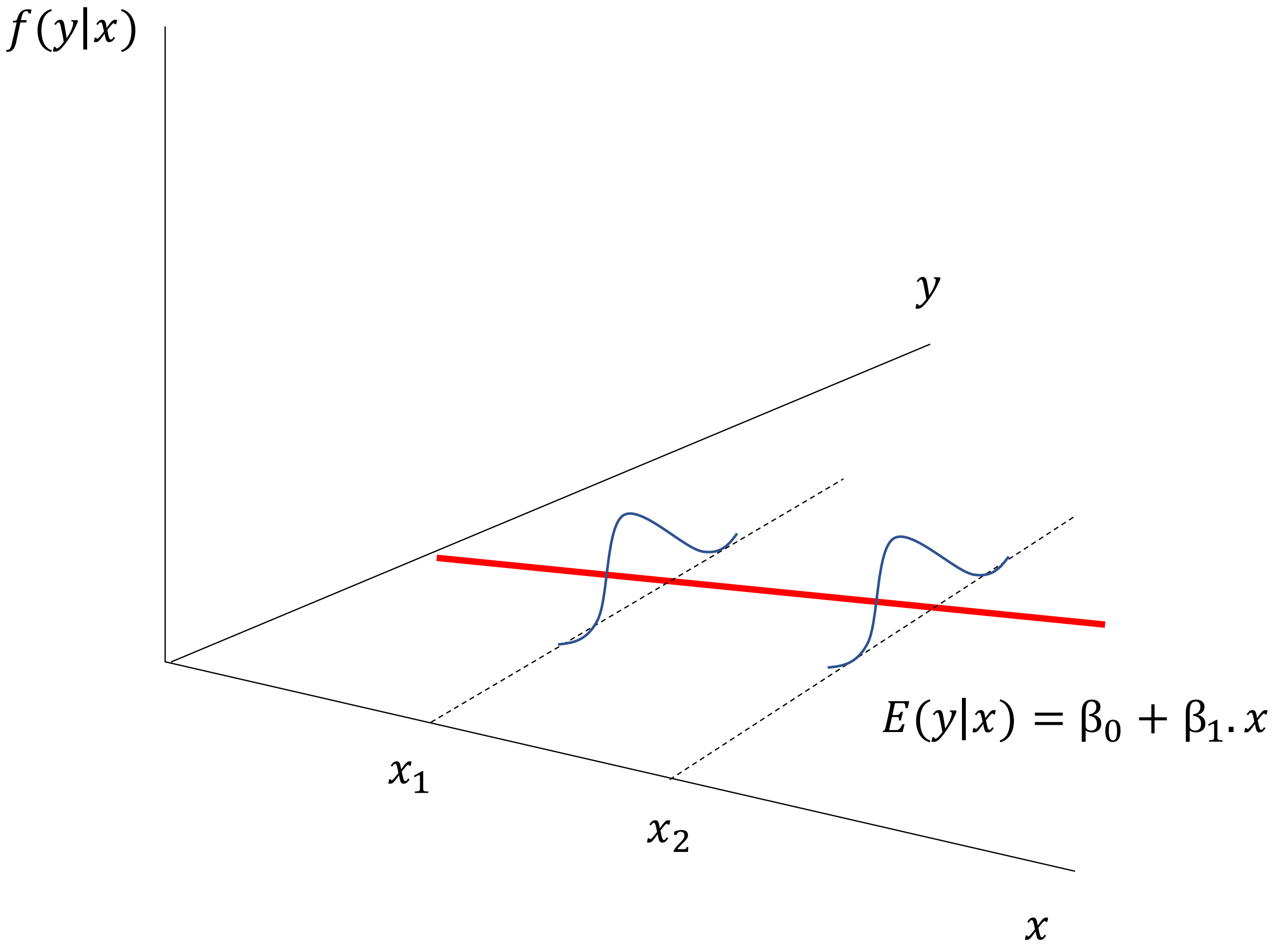
Rozptyl y, který je dán x, se mění v závislosti se změnami x

$$Var(\varepsilon_i|x_i) = \sigma_i^2$$

Podmíněný rozptyl náhodné složky se bude měnit v závislosti na hodnotě nezávislé proměnné (x)



Jak se mění rozptyl y, když se mění (roste) x?

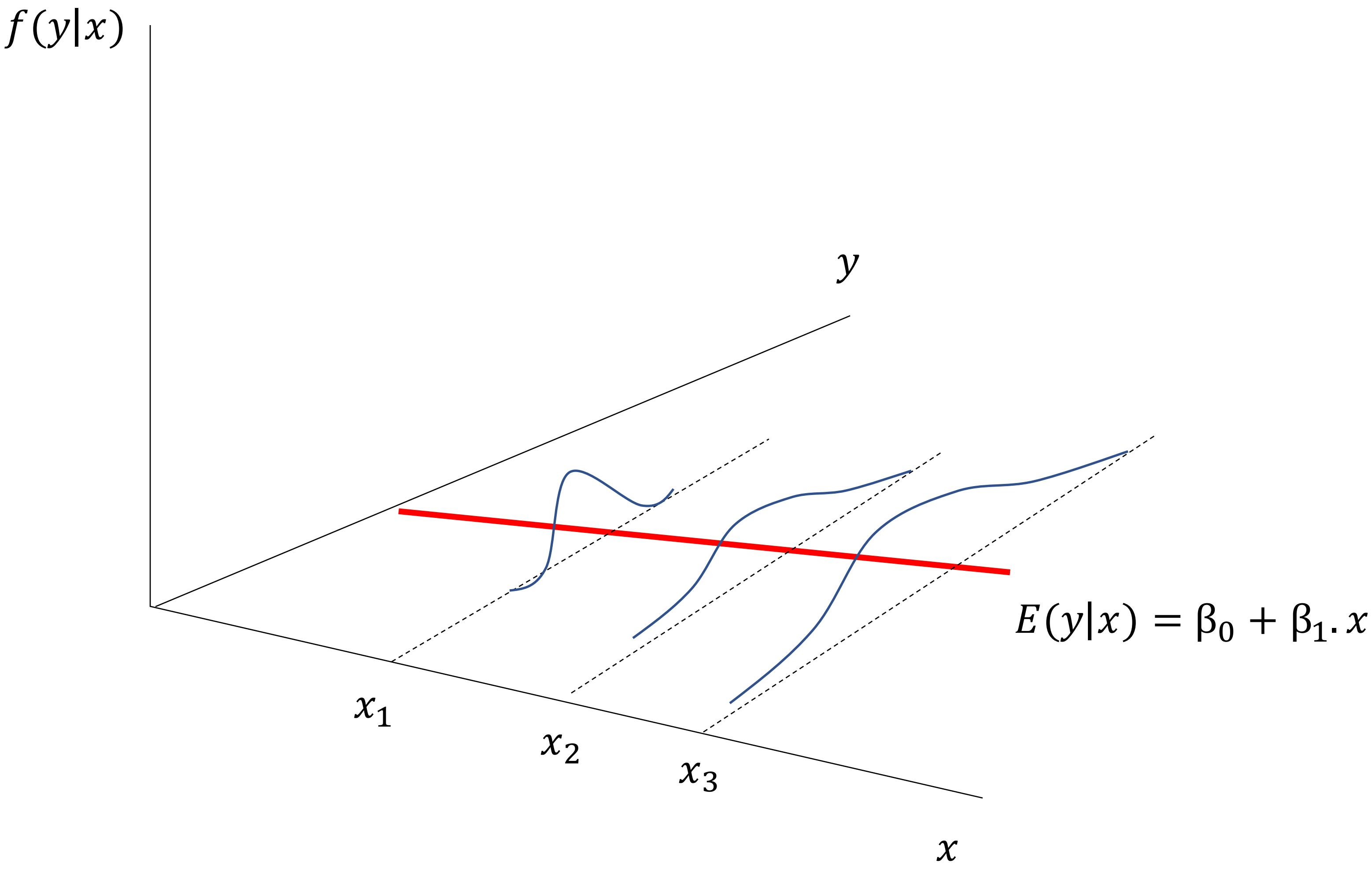


Platí

$Var(\varepsilon|x) = Var(y|x) = \sigma^2 I$

Jak se mění rozptyl y, když se mění (roste) x?

$$Var(\varepsilon_i|x_i) = \sigma_i^2$$



Rozptyl náhodné složky – kovarianční matice

Pro jednoduchost budeme mít pouze 3 pozorování.

$$Var(\varepsilon|X) = E(\varepsilon\varepsilon'|X) = E\left[\left.\begin{pmatrix}\epsilon_1\\ \epsilon_2\\ \epsilon_3\end{pmatrix}.\begin{pmatrix}\epsilon_1 & \epsilon_2 & \epsilon_3\end{pmatrix}\right|X\right] = E\left[\begin{pmatrix}\epsilon_1\epsilon_1 & \epsilon_1\epsilon_2 & \epsilon_1\epsilon_3\\ \epsilon_2\epsilon_1 & \epsilon_2\epsilon_2 & \epsilon_2\epsilon_3\\ \epsilon_3\epsilon_1 & \epsilon_3\epsilon_2 & \epsilon_3\epsilon_3\end{pmatrix}|X\right] =$$

$$= \begin{bmatrix}E(\epsilon_1\epsilon_1|X) & E(\epsilon_1\epsilon_2|X) & E(\epsilon_1\epsilon_3|X)\\ E(\epsilon_2\epsilon_1|X) & E(\epsilon_2\epsilon_2|X) & E(\epsilon_2\epsilon_3|X)\\ E(\epsilon_3\epsilon_1|X) & E(\epsilon_3\epsilon_2|X) & E(\epsilon_3\epsilon_3|X)\end{bmatrix} =$$

$$seriová\ závislost/nezávislost$$

$$= \begin{bmatrix}Var(\epsilon_1) & Cov(\epsilon_1,\epsilon_2) & Cov(\epsilon_1,\epsilon_3)\\ Cov(\epsilon_2,\epsilon_1) & Var(\epsilon_2) & Cov(\epsilon_2,\epsilon_3)\\ Cov(\epsilon_3,\epsilon_1) & Cov(\epsilon_3,\epsilon_2) & Var(\epsilon_3)\end{bmatrix}$$

$$\begin{bmatrix}\sigma^2 & 0 & 0\\ 0 & \sigma^2 & 0\\ 0 & 0 & \sigma^2\end{bmatrix} = \boldsymbol{\sigma^2 I}$$

$$= \begin{bmatrix}\sigma^2 & Cov(\epsilon_1,\epsilon_2) & Cov(\epsilon_1,\epsilon_3)\\ Cov(\epsilon_2,\epsilon_1) & \sigma^2 & Cov(\epsilon_2,\epsilon_3)\\ Cov(\epsilon_3,\epsilon_1) & Cov(\epsilon_3,\epsilon_2) & \sigma^2\end{bmatrix}$$

Co jsou prvky mimo diagonálu?

$$Cov(x,y) = \big(x - E(x)\big).(y - E(y))$$

$$Cov(\epsilon_2,\epsilon_1) = (\epsilon_2 - E(\epsilon_2)).(\epsilon_1 - E(\epsilon_1))$$

$$Cov(\epsilon_2,\epsilon_1) = \epsilon_2.\epsilon_1 = 0$$

$$Var(b|X) = Var\left[\boldsymbol{\beta} + (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\epsilon \,\middle|\, X\right]$$

$$Var(b|X) = Var\left[(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\epsilon \,\middle|\, X\right]$$

$$Var(b|X) = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'Var[\epsilon\,|X]\left((\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\right)'$$

$$= (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'Var[\epsilon|X]\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}$$

$$= (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\sigma}^2\boldsymbol{I}\,\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}$$

$$(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\,\boldsymbol{X} = \boldsymbol{I}$$

$$\boldsymbol{\sigma^2 - konstanta, lze\ vytknout\ p\check{r}ed}$$

$$= \boldsymbol{\sigma}^2(\boldsymbol{X}'\boldsymbol{X})^{-1}$$

$$Var[\epsilon|X] \neq \boldsymbol{\sigma}^2\boldsymbol{I}$$

$$Var(b|X) \neq \boldsymbol{\sigma}^2(\boldsymbol{X}'\boldsymbol{X})^{-1}$$

Při heteroskedasticitě je rozptyl OLS estimatoru $\boldsymbol{b}$:

$$Var(b|X) = (X'X)^{-1}X'\sigma^2\Omega\, X(X'X)^{-1}$$

Sandwich covariance matrix

Tedy v případě heteroskedasticity již nelze získat nevychýlený a konzistentní odhad $Var(b|X)$
Pomocí $\widehat{\boldsymbol{\sigma}^2}(\boldsymbol{X}'\boldsymbol{X})^{-1}$

Příčiny existence heteroskedasticity

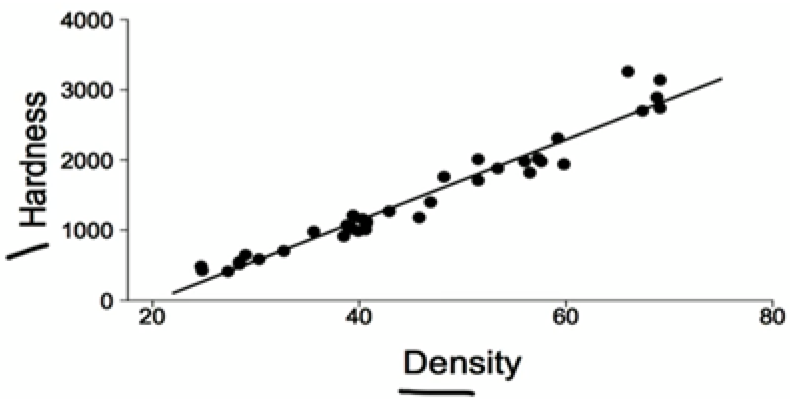
Průřezová data
Značně odlišné hodnoty v jednom modelu
Variabilita závislé proměnné (y) závisí na některé nezávislé proměnné (x)

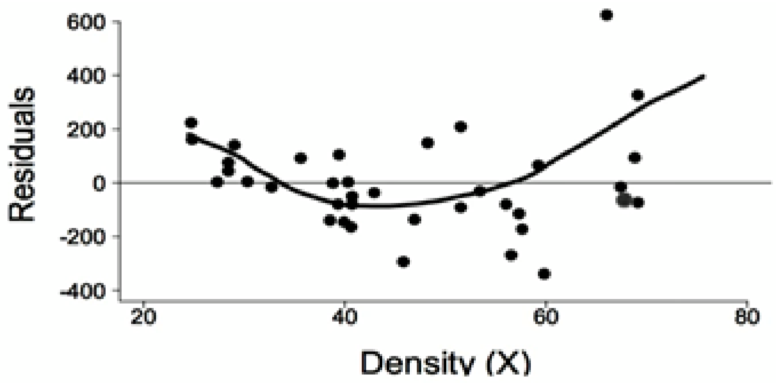
Chybná specifikace modelu
Vynechání „důležité" nezávislé proměnné
Nevhodná funkční forma modelu
OBJEVENÍ HETEROSKEDASTICITY NÁS MŮŽE NAVÉST NA NOVÝ MODEL!

$$\log(wage) = \beta_0 + \beta_1 educ + \epsilon$$

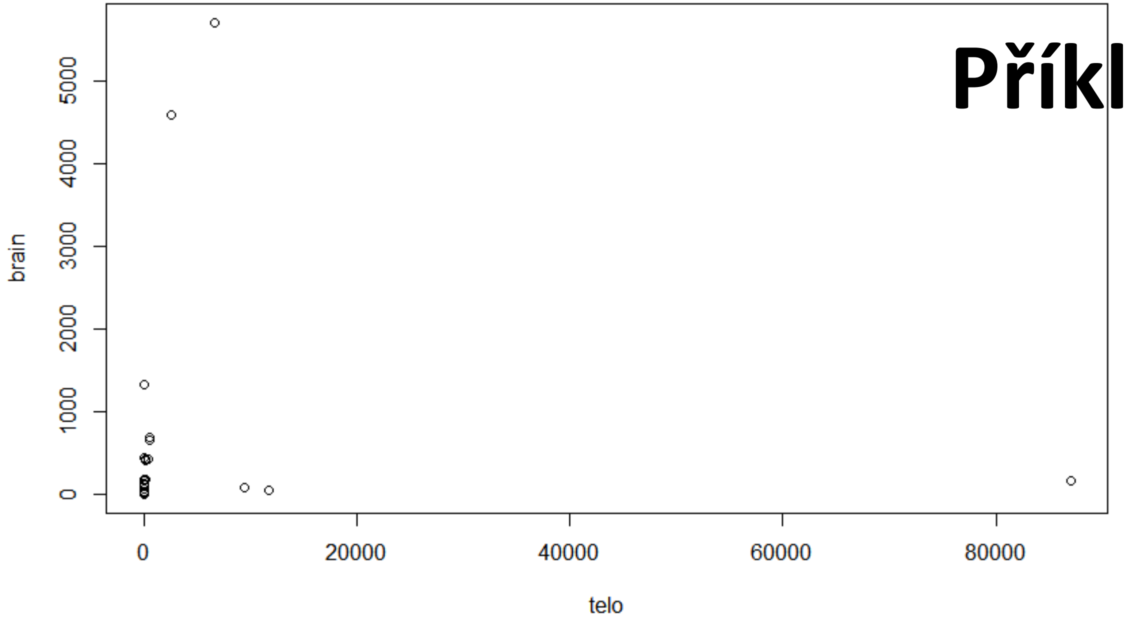
$$wage = \beta_0 + \beta_1 educ + \epsilon$$

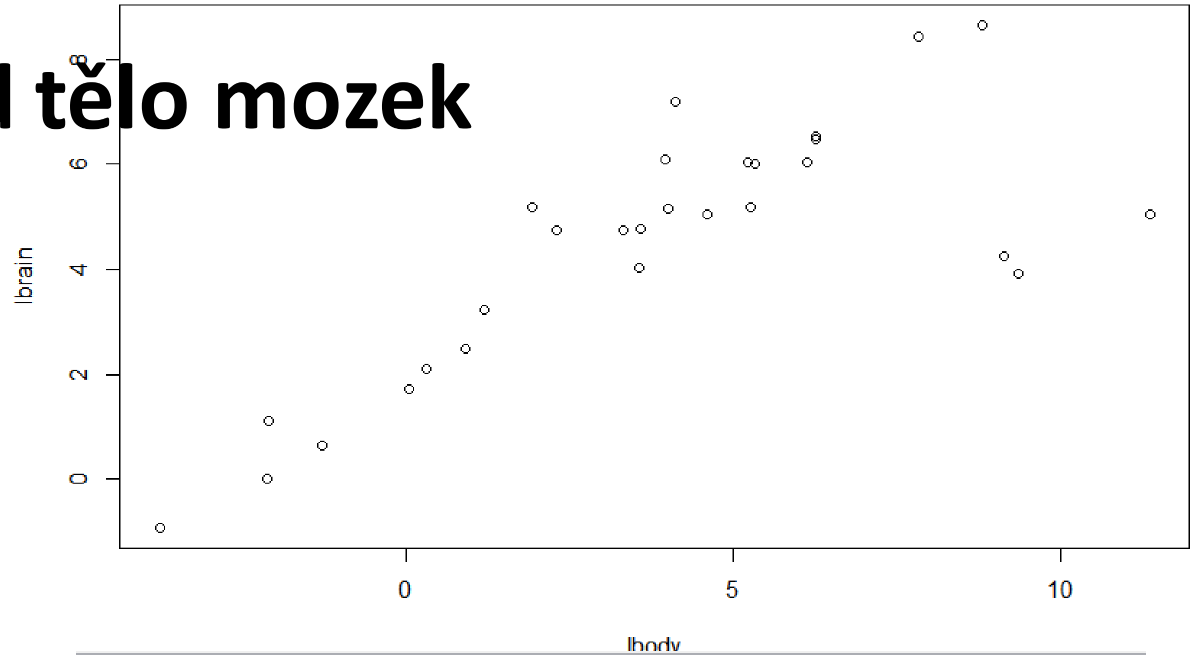
Chyby v měření
S rostoucí hodnotou endogenní proměnné dochází ke kumulaci chyb měření
to zvyšuje rozptyl endogenní proměnné a tedy i rozptyl reziduí





Příklad tělo mozek





Soubor Upravit Testy Uložit Grafy Analýza LaTeX

Model 1: OLS, za použití pozorování 1-28
Závisle proměnná: Brain_g

	koeficient	směr. chyba	t-podíl	p-hodnota	
const	576,372	265,912	2,168	0,0395	**
Body_kg	-0,000432637	0,0158853	-0,02724	0,9785	

Střední hodnota závisle proměnné	574,5214
Sm. odchylka závisle proměnné	1334,929
Součet čtverců reziduí	48113598
Sm. chyba regrese	1360,339
Koeficient determinace	0,000029
Adjustovaný koeficient determinace	-0,038432
F(1, 26)	0,000742
P-hodnota(F)	0,978480
Logaritmus věrohodnosti	-240,7265
Akaikovo kritérium	485,4529
Schwarzovo kritérium	488,1174
Hannan-Quinnovo kritétium	486,2675

zde je poznámka o zkratkách statistik modelu

Whiteův test heteroskedasticity -

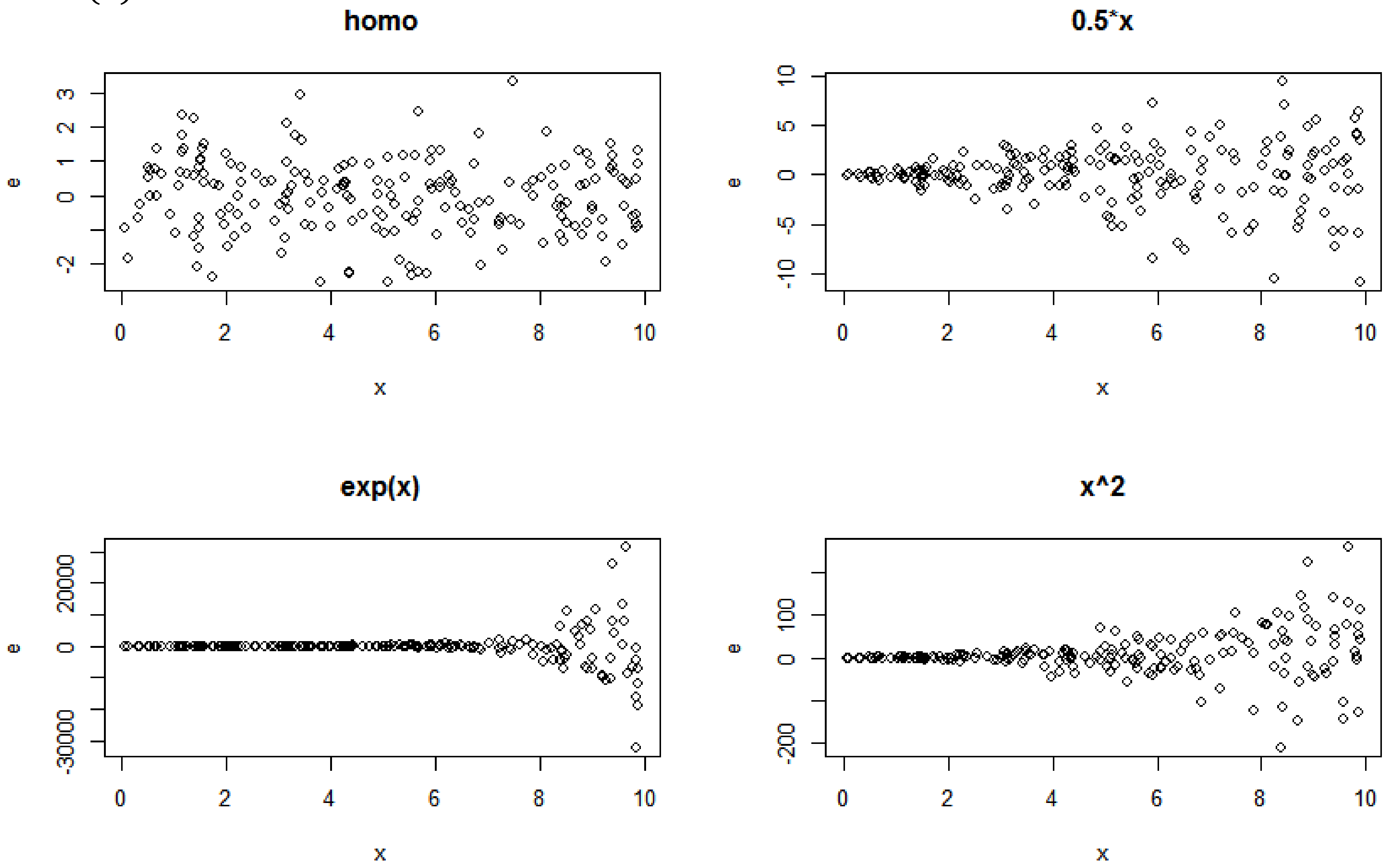
Model 2: OLS, za použití pozorování 1-28
Závisle proměnná: lbrain

	koeficient	směr. chyba	t-podíl	p-hodnota	
const	2,55490	0,413137	6,184	1,53e-06	***
lbody	0,495995	0,0781694	6,345	1,02e-06	***

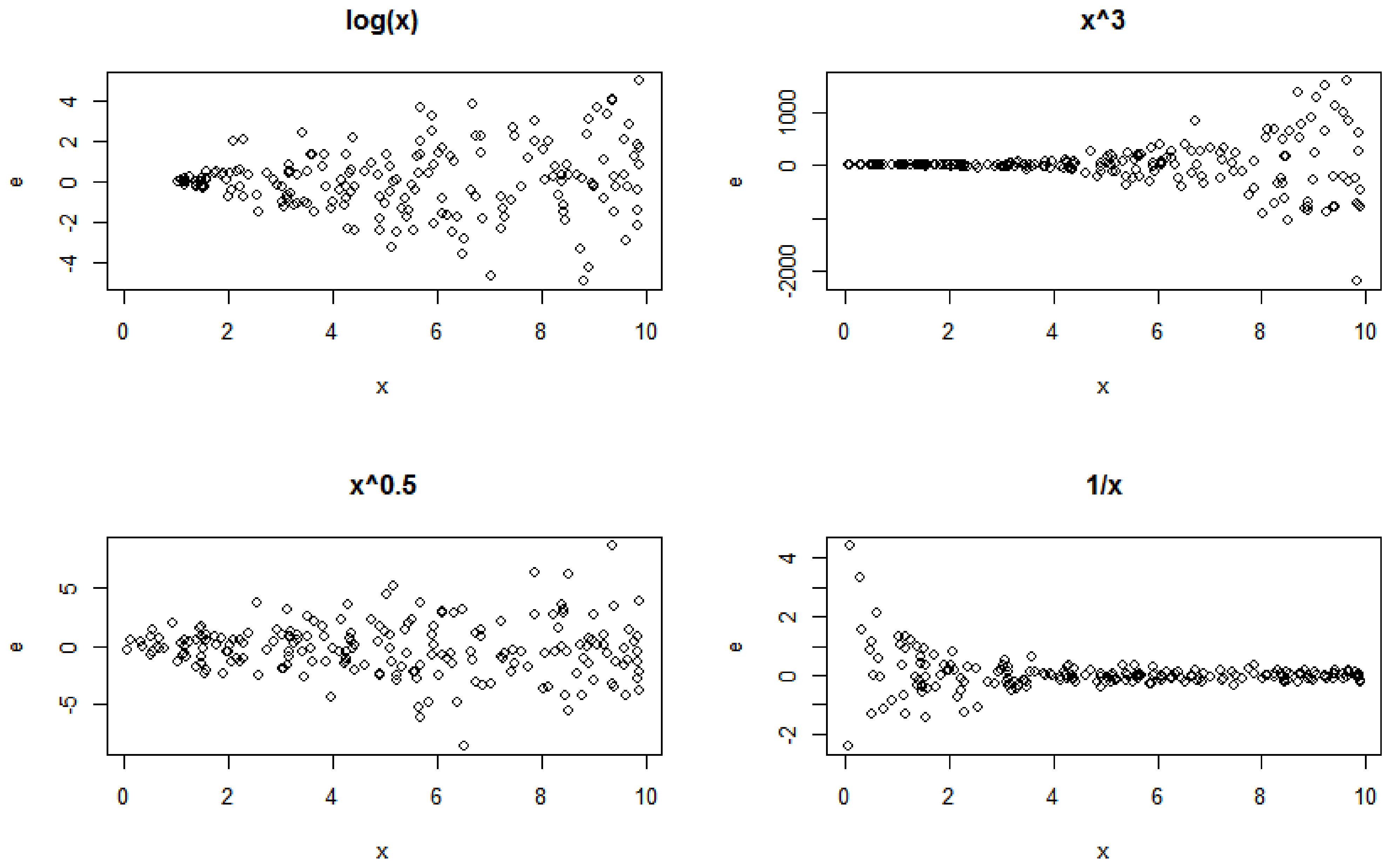
Střední hodnota závisle proměnné	4,425446
Sm. odchylka závisle proměnné	2,399282
Součet čtverců reziduí	60,98799
Sm. chyba regrese	1,531565
Koeficient determinace	0,607610
Adjustovaný koeficient determinace	0,592518
F(1, 26)	40,26063
P-hodnota(F)	1,02e-06
Logaritmus věrohodnosti	-50,62889
Akaikovo kritérium	105,2578
Schwarzovo kritérium	107,9222
Hannan-Quinnovo kritétium	106,0723

zde je poznámka o zkratkách statistik modelu

$$Var(\epsilon|x) = \sigma^2 . h(x)$$



$$Var(\epsilon|x) = \sigma^2 . h(x)$$



Důsledky heteroskedasticity

$Var[\epsilon|X] \neq \boldsymbol{\sigma^2 I}$ $\quad E(b) = \beta \quad t = \frac{b_j}{sd(b_j)}$

Nelze použít t-test, F-test, LM-test
Konfidenční intervaly

$Var(b) \neq \boldsymbol{\sigma^2 (X'X)^{-1}}$

NEMÁ DOPAD NA ODHAD PARAMETRŮ β – NEOVLIVNÍ ZKRESLENOST/NEZKRESLENOST pro konečný počet pozorování
Pro velký počet pozorování „large sample“ – nemá vliv na konzistenci odhadu

1) OLS odhad již není BLUE – má vliv na vydatnost odhadu
Tím, že pozoruje určitý „vzorec“ jak residua (náhodná složka) rostou/klesají
tak jsme nepostihli všechny informace v modelu a proto nemůže být BLUE (best)
Bude jiný lepší (BLUE) estimator, který postihne „tuto informaci“

2) **Odhad $\boldsymbol{Var(b)}$ bude nekonzistentní.**
My vlastně používáme špatný estimátor.
Dopad na t,F,LM test, intervaly spolehlivosti. Jejich výsledky budou zavádějící. Viz předpoklady daných testů

Heteroskedasticita v průřezových datech

$$Var(\epsilon|x) = Var(y|x)$$

Homoskedasticita

$$sav = \beta_0 + \beta_1 inc + \epsilon \qquad Var(\epsilon|inc) = \sigma^2$$

$$Var(\varepsilon_i|x_i) = \sigma_i^2$$

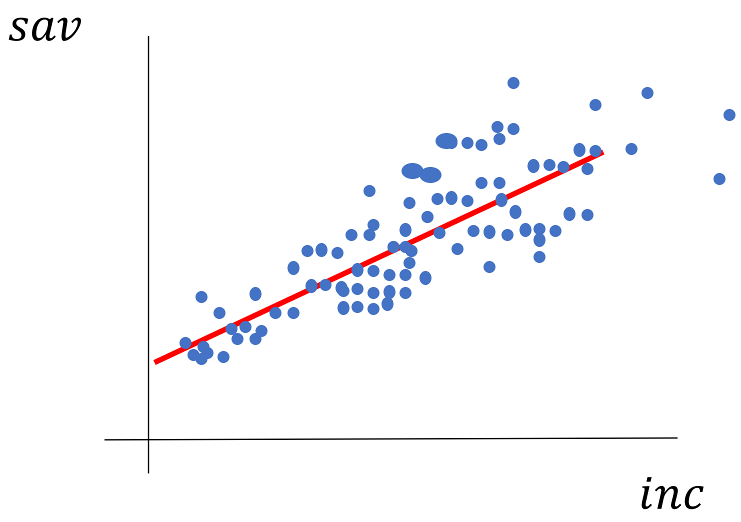
$$Var(\varepsilon_i|x_i) = h(x_i)$$

Heteroskedasticita

$$Var(\epsilon|inc) = \sigma^2 . h(inc)$$

$$Var(\epsilon|X) = h(X)$$

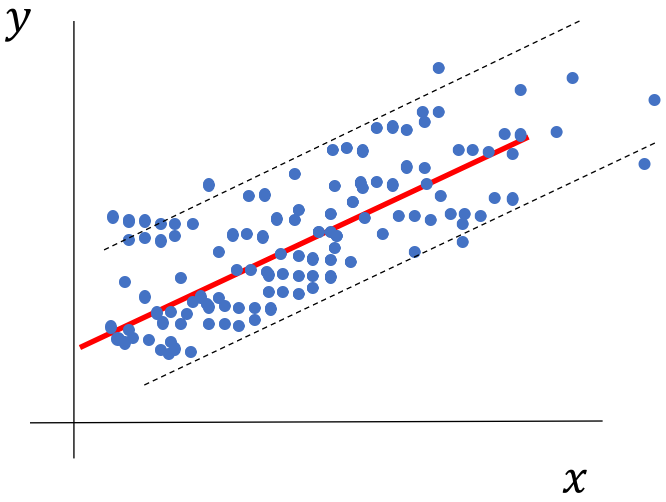
$$Var(sav|inc) = \sigma^2 . h(inc)$$

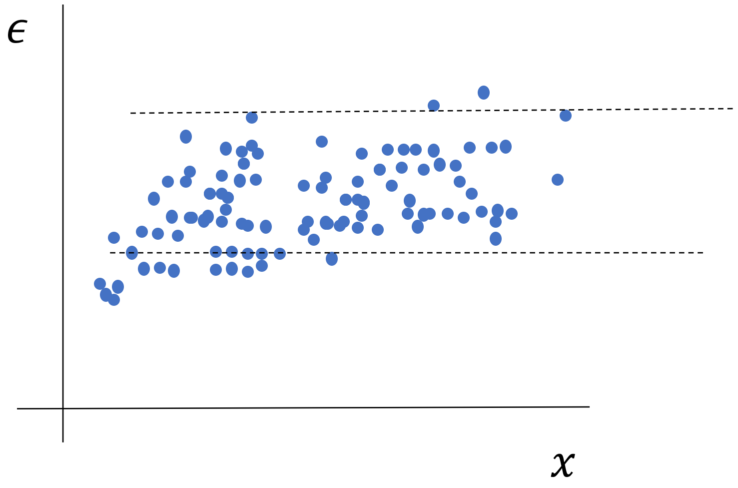


S rostoucím příjmem roste variabilita úspor –např. více možností
S rostoucím vzděláním roste možnost uplatnění a tedy i mezd

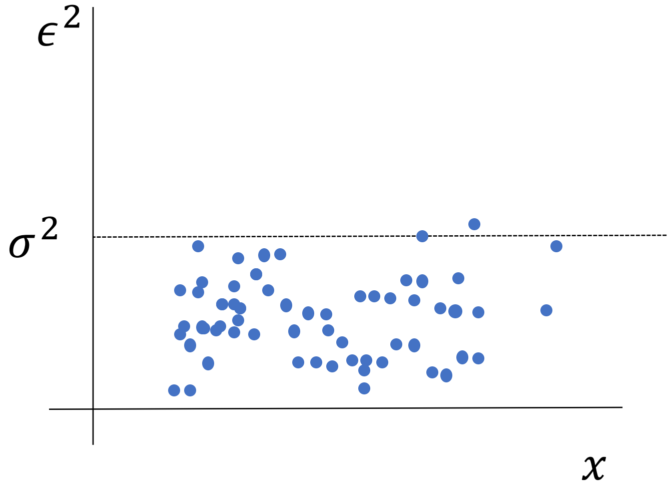
$$Var(\epsilon|x) = Var(y|x)$$

$$Var(\epsilon|x) = \sigma^2 \qquad Var(sav|inc) = \sigma^2 \qquad Var(\epsilon|inc) = \sigma^2$$





Homoskedasticita

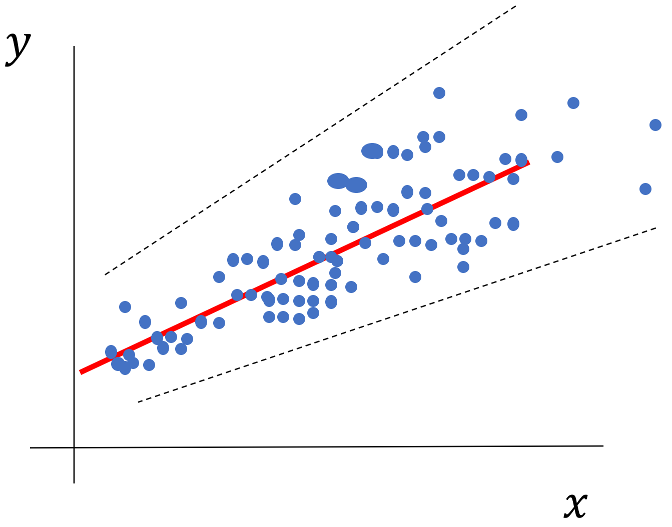


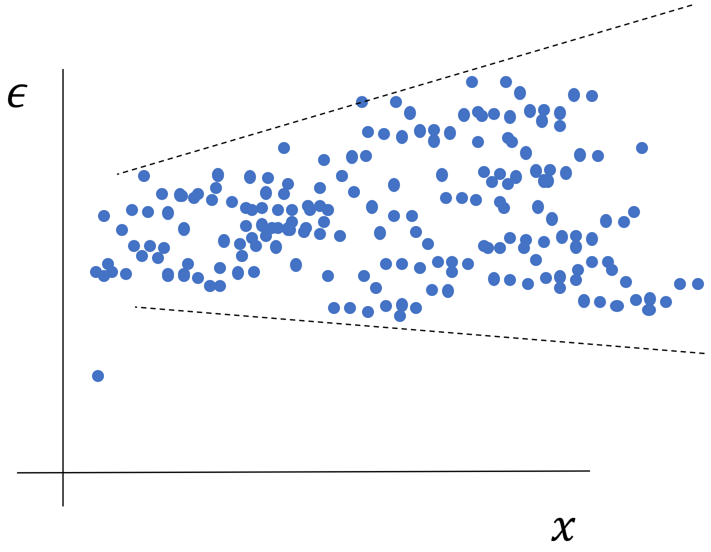
$Var(\epsilon|x) = Var(y|x)$

$Var(\epsilon|x) = \sigma^2.h(x)$

$Var(sav|inc) = \sigma^2.h(inc)$

$Var(\epsilon|inc) = \sigma^2.h(inc)$





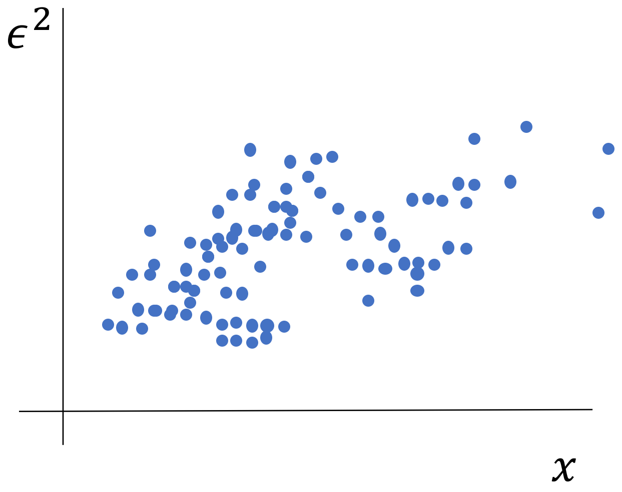
Heteroskedasticita

Nejprve předpokládejme, že známe $h(x)$

$h(x) = x \qquad Var(\epsilon|x) = \sigma^2.x$

$h(x) = x^2 \qquad Var(\epsilon|x) = \sigma^2.x^2$

$h(x) = e^x \qquad Var(\epsilon|x) = \sigma^2.e^x$



Testování heteroskedasticity

Základní myšlenka – ovlivňují nezávislé (X) proměnné rozptyl náhodné složky?

$$y = \beta_0 + \beta_1 . x_1 + \beta_2 . x_2 + \cdots + \beta_k . x_k + \varepsilon$$

$$H0:\ Var(\epsilon | x_1, x_2, \dots, x_k) = \sigma^2 I$$

$$H0:\ Var(\epsilon_i | x_i) = \sigma^2$$

$$Var(\epsilon | X) = E(\epsilon^2 | X) = E(\epsilon^2) = \sigma^2$$

Reg.fce

$$\epsilon^2 = \gamma_0 + \gamma_1 . x_1 + \gamma_2 . x_2 + \cdots + \gamma_k . x_k + v$$

Breusch–Pagan Test

$$H0:\ \gamma_1 = \gamma_2 = \cdots = \gamma_k = 0$$

F-test/LM test

$$e^2 = \gamma_0 + \gamma_1 . x_1 + \gamma_2 . x_2 + \cdots + \gamma_k . x_k + v$$

$$H1: non\ H1$$

H1: Parametry jsou sdruženě signifikantní, tedy alespoň jedna z proměnných ovlivňuje rozptyl residuí – heteroskedasticita

White test

Pro BP test jsme si řekli, že nepostihne nelineární vztah mezi rozptylem náhodné složky a X
Jedno z možných řešení White test

$$y = \beta_0 + \beta_1 . x_1 + \beta_2 . x_2 + \beta_3 . x_3 + \varepsilon$$

$$x_i x_j \; kdy \; i \neq j$$
$$cross - products$$

Pomocná regrese

$$e^2 = \alpha_0 + \alpha_1 x_1 + \alpha_2 x_2 + \alpha_3 x_3 + \alpha_4 x_1^2 + \alpha_5 x_2^2 + \alpha_6 x_3^2 + \alpha_7 x_1 x_2 + \alpha_8 x_1 x_3 + \alpha_9 x_2 x_3 + v$$

Testujeme koeficient determinace R2
Využití F testu, nebo LM testu

$$n. R^2 \sim \chi^2 (k-1)$$

$H0: homoskedasticita$
$H1: heteroskedasticita$

V případě „více“ proměnných snížený počet stupňů volnosti -10 x a 30 pozorování

$$e^2 = \alpha_0 + \alpha_1 \hat{y} + \alpha_2\, \hat{y}^2 + v$$

Postup:

1) Odhadnu rovnici a získám residua (e) a nafitované hodnoty $\hat{y}$
2) Pomocná regrese z kvadrátů residuí na nafitované hodnoty + jejich kvadráty
3) Vyhodnotím pomocí LM/F-testu

Co dělat v případě přítomnosti heteroskedasticity?

Vy jste se učili, že je možné použít metodu zobecněných nejmenších čtverců (GLS)
ALE, GLS má silný a v praxi téměř nereálný předpoklad, že známe tvar heteroskedasticity - $Var(\epsilon|X) = \sigma^2 h(X)$

1) **Zkuste se zamyslet**

Matematická logika, ekonomická teorie. Specifikoval jsem správně tvar modelu? Viz příčiny hetero.
Mohu zkusit použít logaritmus, pokud to dává smysl.

2) **Pokud úprava modelu nevyjde**

Jedinou možností jak získat konzistentní odhady $Var(b)$ je použití tzv. robustních odhadů chyb.

Jelikož případná heteroskedasticita může mít velmi citelný vliv na $Var(b)$, je možné místo OLS použít tzv. FGLS (Feasible general least squares)

Metoda FGLS vychází z GLS, kdy však neznáme $h(X)$

Robustní odhad chyb

Robustní myslíme, že je rezistentní vůči „něčemu“
Zde vůči heteroskedasticitě

$$Var(\epsilon|X) = \boldsymbol{\sigma^2\Omega}$$

$$Var(b) = (\boldsymbol{X'X})^{-1}\boldsymbol{X'\sigma^2\Omega\, X}(\boldsymbol{X'X})^{-1}$$

Heteroskedasticity-consistent, heteroscedasticity robust..

Sandwich estimator

$$b \sim N(\beta, \Sigma_b) \qquad \Sigma_b = A^{-1}BA^{-1}$$

Získat konzistentní odhad pro A není těžké $\hat{A} = \frac{X^T X}{n}$
Estimátor poskytující konzistentní odhad matice B odvodil White

$$\hat{B} = \frac{1}{n}\sum_{i=1}^{n} e_i^2 x_i^T x_i$$

Dnes existuje více konzistentních estimátorů matice B

- HC0 - White
- HC1
- HC2
- HC3

$$Avar(b) = \frac{\widehat{\Sigma_b}}{n} = (\boldsymbol{X'X})^{-1}\left(\sum_{i=1}^{n} e_i^2 x_i^T x_i\right)(\boldsymbol{X'X})^{-1}$$

$$Avar(b) = (\boldsymbol{X'X})^{-1}\,\widehat{\boldsymbol{X'\Omega X}}(\boldsymbol{X'X})^{-1}$$

Asymptoticky jsou odhady ekvivalentní, liší se však pro „malé“ vzorky

Zobecněná metoda nejmenších čtverců (GLS)

Pokud pomocí testů zjistíme přítomnost heteroskedasticity – můžeme se pokusit o její nápravu

Pro odstranění heteroskedasticity však musíme znát její formu - určíme funkční vztah závislosti

Potom zde představená metoda (WLS) povede k vydatnějšímu odhadu než OLS

Získáme tak „nový“ t-test a F-test, který bude mít právě t/F rozdělení
(pokud je přítomna heteroskedasticita, OLS nám neposkytne t/F test, který má t/F rozdělení)

Vážená metoda nejmenších čtverců (WLS)
je jedním ze speciálních případů
zobecněné metody nejmenších čtverců (GLS)

Vy jste se učili GLS i pro případ přítomnosti autokorelace.

Vážená metoda nejmenších čtverců (GLS)

„Víme“

$$sav = \beta_0 + \beta_1 inc + \epsilon \qquad Var(\varepsilon|inc) = \sigma^2 h(x) > 0 \qquad Var(\varepsilon_i|inc_i) = \sigma^2 inc_i^2$$

S rostoucím příjmem roste variabilita úspor–např. více možností

$$Var(\varepsilon|X) = E\left[\left.\begin{matrix} \varepsilon_1^2 & \varepsilon_1\varepsilon_2 & \varepsilon_1\varepsilon_3 \\ \varepsilon_2\varepsilon_1 & \varepsilon_3^2 & \varepsilon_2\varepsilon_3 \\ \varepsilon_3\varepsilon_1 & \varepsilon_3\varepsilon_2 & \varepsilon_3^2 \end{matrix}\right| inc\right] = \begin{bmatrix} E[\varepsilon_1^2|inc] & 0 & 0 \\ 0 & E[\varepsilon_2^2|inc] & 0 \\ 0 & 0 & E[\varepsilon_3^2|inc] \end{bmatrix} =$$

$$= \sigma^2 \begin{bmatrix} inc_1^2 & 0 & 0 \\ 0 & inc_2^2 & 0 \\ 0 & 0 & inc_3^2 \end{bmatrix} = \sigma^2\,\Omega \qquad \Omega = \begin{bmatrix} inc_1^2 & 0 & 0 \\ 0 & inc_2^2 & 0 \\ 0 & 0 & inc_3^2 \end{bmatrix}$$

$$Var(\epsilon|X) = \boldsymbol{\sigma^2\Omega}$$

$$sav = \beta_0 + \beta_1 inc + \epsilon \qquad Var(\epsilon|inc) = \sigma^2 inc^2$$

$$\frac{sav}{inc} = \frac{\beta_0}{inc} + \frac{\beta_1 inc}{inc} + \frac{\epsilon}{inc}$$

$$Var\left(\frac{\epsilon}{inc}\middle| inc\right) = \frac{1}{inc^2} Var(\epsilon|inc) = \frac{1}{inc^2} \sigma^2 inc^2 = \sigma^2$$

Takto upravený rozptyl již nevykazuje heteroskedasticitu

$y = \beta_0 + \beta_1 x + \epsilon$ $\quad Var(\varepsilon|inc) = \sigma^2 h(x)$ $\quad předpoklad, že\ znám\ tvar\ h(x)$

$$\frac{y}{\sqrt{h(x)}} = \frac{\beta_0}{\sqrt{h(x)}} + \frac{\beta_1 x}{\sqrt{h(x)}} + \frac{\epsilon}{\sqrt{h(x)}}$$

$$Var\left(\frac{\epsilon}{\sqrt{h(x)}}\middle|x\right) = \frac{1}{h(x)} Var(\epsilon|x) = \frac{1}{h(x)} \sigma^2 h(x) = \sigma^2$$

Postup maticově

1) Model odhadneme pomocí MNČ
2) Vyhodnotíme, zda-li se v modelu vyskytuje heteroskedasticita/autokorelace
3) Zjistit zda-li se nejedná o vynechanou proměnnou, nelineární vztah atd.
   (případy existence heteroskedasticity/autokorelace)
4) Nalezneme vhodnou transformační matici T
5) Upravíme model
6) Upravený model odhadneme pomocí MNČ

$$y = X\beta + \varepsilon \qquad Var(\varepsilon|x) = \sigma^2\Omega$$

$$Var(\varepsilon|X) = \begin{bmatrix} E[\varepsilon_1^2|X] & 0 & 0 \\ 0 & E[\varepsilon_2^2|X] & 0 \\ 0 & 0 & E[\varepsilon_3^2|X] \end{bmatrix} = \sigma^2.\Omega$$

Mějme matici ***T*** pro kterou platí: $T'T = \Omega^{-1}$

$$\Omega = \begin{bmatrix} h(X) & 0 & 0 \\ 0 & h(X) & 0 \\ 0 & 0 & h(X) \end{bmatrix}$$

$$T.y = TX\beta + T.\varepsilon$$

$$y^* = X^*\beta + \epsilon^*$$

$$Var(A.x) = A.Var(x).A'$$

$$Var(\epsilon^*|X) = Var(T.\varepsilon|x) = T.\sigma^2\Omega.T' = \sigma^2 T.\Omega.T' = \sigma^2 T(T'T)^{-1}T' = \sigma^2.I$$

$$I = T.\Omega.T'$$

$$T^{\prime}T = \Omega^{-1}$$

$$T = \begin{bmatrix} \frac{1}{\sqrt{h(X)}} & 0 & 0 \\ 0 & \cdots & 0 \\ 0 & 0 & \frac{1}{\sqrt{h(X)}} \end{bmatrix}$$

$$\Omega = \begin{bmatrix} h(X) & 0 & 0 \\ 0 & h(X) & 0 \\ 0 & 0 & h(X) \end{bmatrix}$$

transformační matice T

$$y = \beta_0 + \beta_1 x + \epsilon \qquad Var(\varepsilon|inc) = \sigma^2 h(x)$$

předpoklad, že znám tvar h(x)

$$T.y = TX\beta + T.\varepsilon$$

$$\frac{y}{\sqrt{h(x)}} = \frac{\beta_0}{\sqrt{h(x)}} + \frac{\beta_1 x}{\sqrt{h(x)}} + \frac{\epsilon}{\sqrt{h(x)}}$$

$$Var(T.\varepsilon|x) = T.\sigma^2\Omega.T' = \sigma^2 T.\Omega.T' = \sigma^2.I$$

$$Var\left(\frac{\epsilon}{\sqrt{h(x)}}\middle|x\right) = \frac{1}{h(x)}Var(\epsilon|x) = \frac{1}{h(x)}\sigma^2 h(x) = \sigma^2$$

$$T.y = TX\beta + T.\varepsilon \qquad y^* = X^*\beta + \epsilon^* \qquad b_{OLS} = (X^{*\prime}X^*)^{-1}.X^{\prime *}y^*$$

$$b_{GLS} = (X'T'TX)^{-1}.X'T'Ty = (X'\Omega^{-1}X)^{-1}.X'\Omega^{-1}y$$

$$b_{GLS} = (X'\Omega^{-1}X)^{-1}.X'\Omega^{-1}y$$

Aitken estimator
GLS estimator
ZMNČ

$$Var(b_{GLS}|X) = Var[(X'\Omega^{-1}X)^{-1}.X'\Omega^{-1}\epsilon]$$
$$= (X'\Omega^{-1}X)^{-1}.X'\Omega^{-1}Var(\epsilon|X)\Omega^{-1}X.(X'\Omega^{-1}X)^{-1} = \sigma^2(X'\Omega^{-1}X)^{-1}$$

$$Var(b_{GLS}|X) = \sigma^2(X'\Omega^{-1}X)^{-1}$$

$$Var(b_{GM-OLS}|X) = \sigma^2(X'X)^{-1}$$

Rozlišovat rozptyly výběrového rozdělení estimátoru OLS

$$Var(b_{NonGM-OLS}|X) = \sigma^2(X'X)^{-1}.X'.\Omega.X.(X'X)^{-1}$$

$$Q_{OLS} = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \dots min$$

$$Q_{GLS} = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} \frac{(y_i - \hat{y}_i)^2}{h(x)_i} \dots min$$

Je to to samé, jako když rovnici roznásobím $\frac{1}{\sqrt{h(x)}}$ a následně provedu odhad pomocí OLS

$$b_{GLS} = (X'\Omega^{-1}X)^{-1}.X'\Omega^{-1}y$$

My transformujeme regresní rovnici
Následně odhadneme metodou nejmenších čtverců
Toto je metoda ZMNČ, GLS
Tento odhad je již zase BLUE!
Zpětně substituujeme pro interpretaci parametrů

$$\frac{1}{\sqrt{h(x)}}$$

$$\epsilon^* \sim N(0, \sigma^{2*})$$

$$T.y = TX\beta + T.\varepsilon$$

$$y^* = X^*\beta + \epsilon^*$$

$$y^* = X^*b_{GLS} + e^*$$

$$\widehat{\sigma^{2*}} = \frac{1}{n-k-1}\Sigma e^{*2}$$

Již opět platí $E(\widehat{\sigma^{2*}}) = \sigma^{2*}$

Opět můžeme použít t-test, F-test

Feasible general least square (FGLS)

Je přítomná heteroskedasticita, co tedy dělat?

Použít robustní odhad chyb.
Co GLS?
Nelze, neznáme tvar heteroskedasticity
Ale můžeme použít FGLS
FGLS poskytuje asymptoticky vydatnější odhady, než OLS

Cílem je odhadnout tvar $h(X)$ z $Var(\epsilon|X) = \sigma^2 h(X)$

FGLS podle White
1. Odhadneme regresní model a uložíme residua (e)
2. Upravíme residua do tvaru $log(e^2)$ a provedeme pomocnou regresi
3. Z bodu 2 získáme nafitované hodnoty $(\hat{g})$ a upravíme tak, že $\hat{h} = \exp(\hat{g})$
4. Transformujeme rovnici 1.

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + \epsilon$$

$$Var(\epsilon|X) = \sigma^2 \exp(\gamma_0 + \gamma_1 x_1 + \gamma_2 x_2 + .. + \gamma_k x_k)$$

$$h(X) = \exp(\gamma_0 + \gamma_1 x_1 + \gamma_2 x_2 + .. + \gamma_k x_k)$$

$$\epsilon^2 = \sigma^2 \exp(\gamma_0 + \gamma_1 x_1 + \gamma_2 x_2 + .. + \gamma_k x_k)\, u \qquad u \sim IID(0, \Sigma_u)$$

$$log(e^2) = \underbrace{\log \sigma^2 + \gamma_0}_{\alpha_0} + \gamma_1 x_1 + \gamma_2 x_2 + .. + \gamma_k x_k + v$$

Odhad vah je pak

$$\hat{h} = exp\left(\widehat{log(e^2)}\right)$$

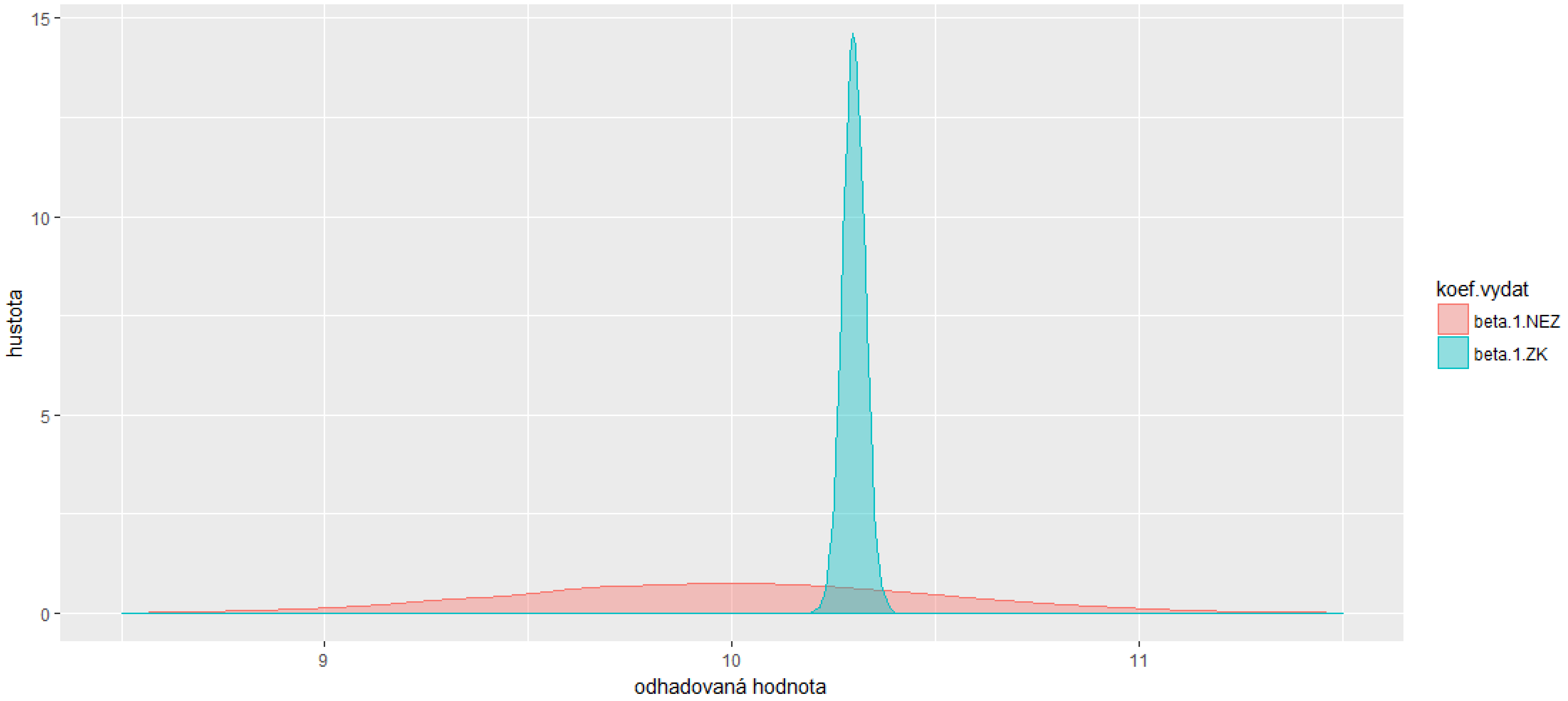
$$\frac{y}{\sqrt{\hat{h}}} = \frac{\beta_0}{\sqrt{\hat{h}}} + \beta_1 \frac{x_1}{\sqrt{\hat{h}}} + \beta_2 \frac{x_2}{\sqrt{\hat{h}}} + \cdots + \beta_k \frac{x_k}{\sqrt{\hat{h}}} + \frac{\epsilon}{\sqrt{\hat{h}}}$$

Nezbavíme se heteroskedasticity!!!
Ale je pravděpodobné, že snížíme její dopad

Odhady pomocí FGLS jsou konzistentní
Ne však nevychýlené.

Tade off vychýlenost vs. vydatnost

Tade off vychýlenost vs. vydatnost



Heteroskedasticita praktické otázky

1) Když máme robustní odhad chyb, proč testujeme model na přítomnost heteroskedasticity?
Robustní odhady jsou pouze asymtoticky validní. To má dopad i na t,F, LM testy.
Ty mají potom pouze asymptoticky t,F,χ^2 rozdělení – tedy v případě málo pozorování, můžeme být „hodně mimo“ **UKAŽ R!**

2) **Máme používat FGLS v případě přítomnosti heteroskedasticity?**
Nelze jednoznačně říci, zdali skutečně FGLS poskytuje vydatnější odhady, nežli OLS. Asymptoticky však ano.
Záleží jakou funkci h(x) použijeme.
Ukážeme si na cviku případ silné heteroskedasticity

Vždy zkuste více testů heteroskedasticity. Pokud vyjde alespoň u jednoho z nich, že můžeme zamítnout hypotézu o homoskedasticite, tak automaticky berte, že je přítomna heteroskedasticita, přestože to jiný test nepotvrzuje.

Pokud vyjde p-value u testů „blízko“ 10%, opět raději použíte robustní odhad chyb.
Například () pomocí simulací zjistil, že je vždy lepší použít robustní odhady chyb.
Jiní např. s tímto závěrem zase nesouhlasí.

Zkuste například porovnat výsledky t-testu pro „normální“ a pro robustní odhad.

Pozor na předpoklady dalších testů – bude často potřeba použít jejich robustní verzi

Heteroskedasticity robust F,LM-test

EVROPSKÁ UNIE
Evropské strukturální a investiční fondy
Operační program Výzkum, vývoj a vzdělávání

MINISTERSTVO ŠKOLSTVÍ, MLÁDEŽE A TĚLOVÝCHOVY