4EK211 Základy ekonometrie

Odhad klasického lineárního regresního modelu III

Cvičení 4

Jednoduchý KLRM – funkční formy a interpretace β_1

- Pro jednoduchý LRM v základním tvaru $Y_i = \beta_0 + \beta_1 X_i + u_i$, platí:

Přehled funkčních forem při použití logaritmické transformace dat

Model	Endogenní proměnná	Exogenní proměnná	Interpretace β_1
Lineární (level-level)	Y_i	X_i	$E(\Delta Y_i) = \beta_1 \Delta X_i$
Lineárně logaritmický (level-log)	Y_i	$ln(X_i)$	$E(\Delta Y_i) = (\beta_1/100)\%\Delta X_i$
Logaritmicko lineární (log-level)	$ln(Y_i)$	X_i	$E(\%\Delta Y_i) = (100\beta_1)\Delta X_i$
Dvojlogaritmický (log-log)	$ln(Y_i)$	$ln(X_i)$	$E(\%\Delta Y_i) = \beta_1 \%\Delta X_i$

kde $E(\Delta Y_i)$ je očekávaná změna Y_i a $E(\%\Delta Y_i)$ je očekávaná procentní změna Y_i

Jednoduchý KLRM – funkční formy a interpretace β_1

Příklad 1 (odhadnutý KLRM):

$$\widehat{\ln(wage_i)} = 0{,}584 + 0{,}083\, educ_i$$

kde *wage* je měsíční mzda v USD
educ - vzdělání (ukončené roky standardního studia)

Interpretace:

- Každý ukončený rok studia zvýší očekávané (vyrovnané) příjmy o (100*0,083)%, tj. o 8,3%.
- *POZOR:* Efekt získaného „titulu“ (maturita, tituly Bc., Ing.): uvažujeme-li systematický mzdový bonus za získaný titul, není mzdový přínos stejný u všech roků standardního studia. Obvykle *řešíme zahrnutím umělých proměnných* do modelu: získání/nezískání titulu vyjádříme pomocí nové proměnné typu 1-0 (logická proměnná, kvalitativní exogenní informace). Očekávaný mzdový efekt titulu pak bude vyjádřen příslušným koeficientem β_j u umělé proměnné.

Jednoduchý KLRM – funkční formy a interpretace β_1

Příklad 2 (odhadnutý KLRM):

$$\widehat{\ln(\boldsymbol{CEOwage_i})} = 4{,}822 + 0{,}257 \ln(\boldsymbol{sales_i})$$

kde *CEOwage* je roční mzda ředitele (CEO) sledované firmy, měřeno v tis. USD.

sales - roční obrat (prodeje) sledované firmy, měřeno v milionech USD.

Interpretace:

- Jednoprocentní růst prodejů firmy zvýší očekávané (průměrné, vyrovnané) příjmy ředitele (CEO) o 0,257 %.

Pozn: Interpretace parametrů je smysluplná a platí pro verifikované modely (ekonomicky, statisticky, ekonometricky)

Různé specifikace KLRM – srovnání, testování hypotéz

Kvalita regresního modelu závisí na dvou hlavních faktorech:

- **Shoda modelu s daty** (vysoké R^2, model významný jako celek)
- **Jednoduchost** (Occamova břitva, počet stupňů volnosti, praktická využitelnost).
- **Zde popsaná srovnání dvou (a více) KLRM mají smysl pouze pro danou (jednu) endogenní proměnnou:**
- **Příklad:** „Kvalitu“ LRM ve tvaru:

$$wage_i = \beta_0 + \beta_1\, educ_i + u_i$$

lze přímo srovnat s tvarem:

$$wage_i = \beta_0 + \beta_1\, educ_i + \beta_2\, exper_i + u_i$$

ale nikoliv s LRM:

$$\ln(wage_i) = \beta_0 + \beta_1\, educ_i + u_i$$

Různé specifikace KLRM – srovnání, testování hypotéz

Jednoduchost KLRM je nepřímo úměrná počtu proměnných KLRM: *k*

Ukazatele, které poměřují shodu s daty vůči jednoduchosti modelu:

- **Ke srovnání různých specifikací modelu lze použít:**
  - Korigovaný koeficient vícenásobné determinace $\overline{R}^2$
  - Informační kritéria (např.: *AIC, SIC, HQIC,* atd.)

- **Testování (sdružené) hypotézy o parametrech β_j**
  - *F- test* pro nulová omezení parametrů β_j (speciální případ lineárního omezení)
  - F-test pro lineární omezení parametrů β_j

Různé specifikace KLRM – srovnání, testování hypotéz

- **Korigovaný koeficient vícenásobné determinace**

$$\overline{R}^2 = 1-(1-R^2)\frac{n-1}{n-(k+1)}$$

kde:

$$\boldsymbol{R}^2 = \frac{\boldsymbol{VS\check{C}}}{\boldsymbol{CS\check{C}}} = 1-\frac{\boldsymbol{NS\check{C}}}{\boldsymbol{CS\check{C}}} \quad ; \quad \boldsymbol{R}^2 \in \langle 0,1 \rangle \quad ; \quad \boldsymbol{NS\check{C}} = \sum(\boldsymbol{Y_i}-\hat{\boldsymbol{Y}}_i)^2 \quad ; \quad \boldsymbol{CS\check{C}} = \sum(\boldsymbol{Y_i}-\overline{\boldsymbol{Y}})^2$$

- Zatímco R^2 nikdy neklesne přidáním další vysvětlující proměnné do modelu, $\overline{R}^2$ poklesnout může: v důsledku přidání málo významné exogenní proměnné ($\overline{R}^2$ může nabývat i záporných hodnot).

- Rozhodovací kritérium: maximalizujeme hodnotu $\overline{R}^2$.

Různé specifikace KLRM – srovnání, testování hypotéz

- **Korigovaný koeficient vícenásobné determinace: PŘÍKLAD**

Máme LRM ve tvaru:

$$\boldsymbol{Y_i} = \beta_0 + \beta_1 \, \boldsymbol{X}_{1\boldsymbol{i}} + \beta_2 \, \boldsymbol{X}_{2\boldsymbol{i}} + \beta_3 \, \boldsymbol{X}_{3\boldsymbol{i}} + \boldsymbol{u_i} \quad ; \quad \boldsymbol{n} = 20 \qquad (1)$$

Odhadnutý model má R^2 = 0,8 a je významný z hlediska *t*-testů i *F*-testu.

(2) Ke specifikaci (1) přidám vysvětlující proměnnou X_4 a tím zvýším R^2 na hodnotu 0,82.

(3) Alter.: ke specifikaci (1) přidám exogenní proměnnou Z_1 a tím zvýším R^2 na hodnotu 0,81.

Podle hodnot koeficientu $\overline{R}^2$ vyberte ze tři variant (1), (2) a (3) „nejlepší“ specifikaci LRM pro proměnnou Y_i.

(Řešení: 0,762 vs. 0,772 vs. 0,759)

Různé specifikace KLRM – srovnání, testování hypotéz

Informační kritéria

- Charakterizují závislost shody modelu s daty na počtu vysvětlujících proměnných zahrnutých do modelu (podobně jako $\overline{R}^2$).
- U informačních kritérií (IC) je shoda modelu s daty obvykle dána pomocí NSČ (RSS) a jednoduchost (resp. složitost) modelu se řídí počtem regresních proměnných *k*.
- **POZOR NA DATA MINING**: výběr ekonometrických modelů by měl být založen primárně na ekonomické teorii a teprve v druhé řadě na ukazatelích jako $\overline{R}^2$ nebo IC.
  Cílem ekonometrické analýzy má být získání smysluplných odhadů (pro prognózy, rozhodování, analýzy), které odpovídají ekonomické teorii.

Různé specifikace KLRM – srovnání, testování hypotéz

Akaikeho informační kritérium (*AIC*)

- $AIC = \left(\dfrac{2k}{n}\right) + \ln\left(\dfrac{RSS}{n}\right)$, alternativně: $AIC = e^{\frac{2k}{n}} \cdot \dfrac{RSS}{n}$

kde: *RSS* je součet čtverců reziduí regrese (NSČ),
n - počet pozorování,
k - počet vysvětlujících proměnných modelu.

- **Srovnávat lze jen modely pro danou (jednu) endogenní proměnnou.**
- Vybereme takovou specifikaci (zvolíme sadu exogenních proměnných), pro kterou je spočtená hodnota *AIC* minimální. Pravidlo platí pro logaritmický i exponenciální tvar *AIC*.
- *AIC* je vhodné jen pro modely s úrovňovou konstantou.

Různé specifikace KLRM – srovnání, testování hypotéz

Schwarzovo informační kritérium (*SIC*)

- *SIC* = $\frac{k}{n}\ln(n)+\ln\left(\frac{RSS}{n}\right)$, alternativně: *SIC* = $n^{\frac{k}{n}}.\frac{RSS}{n}$

kde: *RSS* je součet čtverců reziduí regrese (NSČ),
n - počet pozorování,
k - počet vysvětlujících proměnných modelu.

- **Srovnávat lze jen modely pro danou (jednu) endogenní proměnnou.** Kritérium: minimalizujeme *SIC*. Použití *SIC* je vhodné jen pro modely s úrovňovou konstantou.

- *SIC* penalizuje zahrnutí dodatečných vysvětlujících proměnných více než *AIC*: je tedy „přísnější“ (nemusí být „lepší“).

- Nejčastěji používaným IC je *AIC*.

Různé specifikace KLRM – srovnání, testování hypotéz

- *F*-test významnosti LRM jako celku je založen na testovací statistice:

$$F = \frac{R^2}{1-R^2}\frac{n-(k+1)}{k}\bigg|_{H_0} \sim F(k, n-k-1)$$

Pro H_0: R^2 se statisticky významně neliší od nuly,
altern.: LRM jako celek je statisticky nevýznamný,
altern.: $\beta_1 = \ldots \beta_k = 0$; LRM dán pouze úrovňovou konstantou β_0.

H_1: R^2 se statisticky významně liší od nuly,
alt.: LRM jako celek je statisticky významný.

$F > F^*_{(k,n-k-1)}$ … zamítáme H_0 ve prospěch H_1
$F \leq F^*_{(k,n-k-1)}$ … nezamítáme H_0

- Jde o speciální užití obecnějšího testu: ***F*-test na omezení parametrů.**

- <u>*F*-test významnosti LRM jako celku</u> klade ***k*** omezení na parametry: $\beta_1 = \ldots \beta_k = 0$. „Omezen“ zde není pouze parametr β_0.
- „Neomezená rovnice“, tj. testovaný LRM se všemi proměnnými a parametry má (***n*-*k*-1**) stupňů volnosti.
- odtud rozdělení *F*-poměru s (*k*) a (*n*-*k*-1) stupni volnosti.

Různé specifikace KLRM – srovnání, testování hypotéz

- ***F*-test na nulová a lineární omezení parametrů** :

$$F = \frac{(RSS_r - RSS_u)}{q(RSS_u)}(n-k-1)\Bigg|_{H_0} \quad \sim \quad F(k, n-k-1)$$

kde RSS_u je nevysvětlený součet čtverců (NSČ) **neomezené regrese** (unrestricted regression),

RSS_r – NSČ omezené regrese (restricted regression),

q – počet omezení na parametry (v omezené regresi),

(n-k-1) – počet stupňů volnosti **v neomezené regresi**,

F – má rozdělení $F(q, n\text{-}k\text{-}1)$

- Pro vhodně formulované hypotézy H_0, H_1 platí:

  $F > F^*_{(k,n\text{-}k\text{-}1)}$ … zamítáme H_0 ve prospěch H_1

  $F \leq F^*_{(k,n\text{-}k\text{-}1)}$ … nezamítáme H_0

Různé specifikace KLRM – srovnání, testování hypotéz

- ***F*-test na nulová omezení parametrů – PŘÍKLAD:**

1. Máme LRM (výchozí, tzv. **neomezený tvar**):

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 X_{4i} + u_i$$

   Pro n =158 je tento LRM model odhadnut (jako významný) a máme spočteno RSS_u = 1 181.

2. Formulujeme hypotézu H_0: $\beta_2 = \beta_3 = 0$; proti H_1: $\neg H_0$ (máme dvě nulová omezení). **Omezený tvar, který odpovídá H_0:**

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_4 X_{4i} + u_i$$

   Na identické sadě pozorování odhadneme omezený LRM. RSS_r = 3 128.

   Pozn.: $RSS_r \geq RSS_u$; analogicky vždy platí: $R_r^2 \leq R_u^2$
   Pozn.: omezený tvar je v tomto případě tzv. zahnízděný (nested), tvoří speciální případ neomezeného tvaru pro $\beta_2 = 0$ a $\beta_3 = 0$.

Klasický lineární regresní model - verifikace

- ***F*-test na nulová omezení parametrů – PŘÍKLAD (dokončení):**

3. Spočítáme *F-poměr* podle vzorce:

$$F = \frac{(3128 - 1811)}{2(1811)}(158 - 4 - 1) \cong 126$$

Kritická hodnota rozdělení: $F^*(2, 153) = 3{,}05$ … (pro $\alpha = 0{,}05$)

4. Vyhodnocení:

Protože $F > F^*$, zamítám H_0 ve prospěch H_1

5. Interpretace:

Hypotéza H_0 je odmítnuta jako neplatná, protože statisticky významně zhoršuje vlastnosti LRM. Omezený model významně hůře postihuje variabilitu endogenní proměnné ve srovnání s neomezeným LRM.

Pozn.: Pokud by u našeho příkladu H_0 platila („byla správná“): v takovém případě nárůst NSČ u omezeného modelu je nevýznamný a *F* poměr se pohybuje v rozmezí: $\langle 0 ; 3{,}05 \rangle$.

Klasický lineární regresní model - verifikace

- ***F*-test na lineární omezení parametrů** :

Výchozí **neomezený tvar** LRM může být dán například takto:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i$$

Kromě nulových omezení existují další ***základní typy*** lineárních omezení, která lze testovat pomocí F-testu:

- H_0: $\beta_1 = \beta_2$
  H_1: $\neg H_0$ nebo H_1: $\beta_1 < \beta_2$ nebo H_1: $\beta_1 > \beta_2$

- H_0: $\beta_1 = 1 - \beta_2$
  H_1: $\neg H_0$ nebo H_1: $\beta_1 < 1 - \beta_2$ nebo H_1: $\beta_1 > 1 - \beta_2$

- H_0: $\beta_1 = 2{,}7$ (konstanta)
  H_1: $\neg H_0$ nebo H_1: $\beta_1 < 2{,}7$ nebo H_1: $\beta_1 > 2{,}7$

Různé specifikace KLRM – srovnání, testování hypotéz

- ***F*-test na omezení typu H_0: $\beta_1 = \beta_2$:**

Výchozí **neomezený tvar** LRM může být dán například takto:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i$$

Omezený tvar (na základě předpokladu $\beta_1 = \beta_2$) sestavíme:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_1 X_{2i} + \beta_3 X_{3i} + u_i$$

lze psát jako :

$$Y_i = \beta_0 + \beta_1 (X_{1i} + X_{2i}) + \beta_3 X_{3i} + u_i$$

resp. dosazením $Z_i = (X_{1i} + X_{2i})$

získáme výsledný omezený tvar :

$$Y_i = \beta_0 + \beta_1 Z_i + \beta_3 X_{3i} + u_i$$

Odhadneme omezený i neomezený LRM, spočteme jejich NSČ a srovnáme získaný *F*-poměr s kritickou hodnotou $F^*(1, n-k-1)$

Různé specifikace KLRM – srovnání, testování hypotéz

- **Alternativa k *F*-testu na omezení typu H_0: $\beta_1 = \beta_2$:**

  Hypotézu H_0 lze přeformulovat na H_0: $\beta_1 - \beta_2 = 0$ a vyhodnotit pomocí *t*-poměru:

$$t = \frac{b_1 - b_2}{s.e.(b_1 - b_2)}$$

kde :

$$s.e.(b_1 - b_2) = \sqrt{VAR(b_1) + VAR(b_2) - 2COV(b_1, b_2)}$$

Obvyklým způsobem srovnáme *t*-poměr s kritickou hodnotou $t^*(n-k-1)$ pro zvolenou hladinu významnosti (jednostranný či oboustranný test).

Různé specifikace KLRM – srovnání, testování hypotéz

- ***F*-test na omezení typu H_0: $\beta_1 = 1 - \beta_2$:**

Výchozí **neomezený tvar** LRM může být dán například takto:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i$$

Omezený tvar (na základě předpokladu $\beta_1 = 1 - \beta_2$) sestavíme:

$$Y_i = \beta_0 + \beta_1 X_{1i} + (1 - \beta_1) X_{2i} + \beta_3 X_{3i} + u_i$$

lze psát jako:

$$Y_i = \beta_0 + \beta_1 X_{1i} + X_{2i} - \beta_1 X_{2i} + \beta_3 X_{3i} + u_i$$

sdružíme podle β_1:

$$Y_i = \beta_0 + \beta_1 (X_{1i} - X_{2i}) + X_{2i} + \beta_3 X_{3i} + u_i$$

dosadíme $W_i = (X_{1i} - X_{2i})$

$$Y_i = \beta_0 + \beta_1 W_i + X_{2i} + \beta_3 X_{3i} + u_i$$

X_{2i} nemá parametr β a porušuje tvar LRM, proto:

$$(Y_i - X_{2i}) = \quad Y_i^* = \beta_0 + \beta_1 W_i + \beta_3 X_{3i} + u_i$$

Různé specifikace KLRM – srovnání, testování hypotéz

- ***F*-test na omezení typu H_0: $\beta_1 = 1 - \beta_2$ (dokončení):**

Odhadneme výsledný **omezený tvar** LRM:

$$Y_i^* = \beta_0 + \beta_1 W_i + X_{2i} + \beta_3 X_{3i} + u_i$$

Spočítáme NSČ pro omezený a neomezený tvar a srovnáme získaný *F*-poměr s kritickou hodnotou $F^*(1,n-k-1)$.

Aby bylo možné porovnat NSČ neomezeného a omezeného modelu, odhadujeme oba tvary vždy na identické sadě pozorování!

\- Platí i v tomto případě: Y^*_i a W_i jsou jednoznačně dány původními pozorováními neomezeného tvaru LRM.

Různé specifikace KLRM – srovnání, testování hypotéz

- ***F*-test na omezení typu H_0: $\beta_1 = 2{,}7$ (konstanta):**

Výchozí **neomezený tvar** LRM může být dán například takto:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i$$

Omezený tvar (na základě předpokladu $\beta_1 = 2{,}7$) sestavíme:

$$Y_i = \beta_0 + 2{,}7 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i$$

X_{1i} nemá parametr β - nelze zde odhadovat MNČ

Protože X_{1i} porušuje tvar pro LRM, upravíme:

$$(Y_i - 2{,}7 X_{1i}) = \quad Y_i^* = \beta_0 + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i$$

Nakonec opět spočítáme NSČ pro omezený i neomezený tvar a srovnáme získaný *F*-poměr s kritickou hodnotou $F^*(1, n-k-1)$.

Různé specifikace KLRM – srovnání, testování hypotéz

- ***F*-testy na omezení parametrů v GRETLu**:
  Tento typ testů je do GRETLu zabudovaný:
  - Okno s výstupem odhadu: Tests → Linear restrictions

  Testovaná omezení se zadávají formou rovnice:

Obecně:	GRETL
H_0: $\beta_1 = 0$	b[*vn1*] = 0
H_0: $\beta_1 = 2{,}7$ (konstanta)	b[*vn1*] = 2,7
H_0: $\beta_1 = \beta_2$	b[*vn1*] – b[*vn2*] = 0
H_0: $\beta_1 = 1 - \beta_2$	b[*vn1*] + b[*vn2*] = 1
H_0: $\beta_1 = 1 - 2\beta_2$	b[*vn1*] + 2*b[*vn2*] = 1

kde *vn1* a *vn2* zastupují názvy proměnných u parametrů β_1 a β_2