

Základy ekonometrie

XII. Modely panelových dat

1 Souhrnný model

2 Modely individuálních vlivů

- Model fixních vlivů
- Model náhodných vlivů
- Rozšíření modelů individuálních vlivů

Úvod a značení

- Panelová data = časová i průřezová dimenze (státy, firmy, domácnosti → jednotlivci resp. jednotky – individuals).
- Y_{it} = pozorování závisle proměnné pro i -tého jednotlivce v čase t .
- X_{it} = pozorování vysvětlující proměnné pro i -tého jednotlivce v čase t .
- $i = 1, \dots, N$
- $t = 1, \dots, T$

Obsah tématu

1 Souhrnný model

2 Modely individuálních vlivů

- Model fixních vlivů
- Model náhodných vlivů
- Rozšíření modelů individuálních vlivů

Úvod

- *Pooled model*.
- Všechna pozorování jako jeden model:

$$Y_{it} = \alpha + \beta X_{it} + \epsilon_{it}$$

- Při splnění klasických předpokladů OLS je BLUE.
- V případě heteroskedasticity metody *GLS*.
- Nic nového × souhrnný model nevhodný pro většinu aplikací.

Obsah tématu

1 Souhrnný model

2 Modely individuálních vlivů

- Model fixních vlivů
- Model náhodných vlivů
- Rozšíření modelů individuálních vlivů

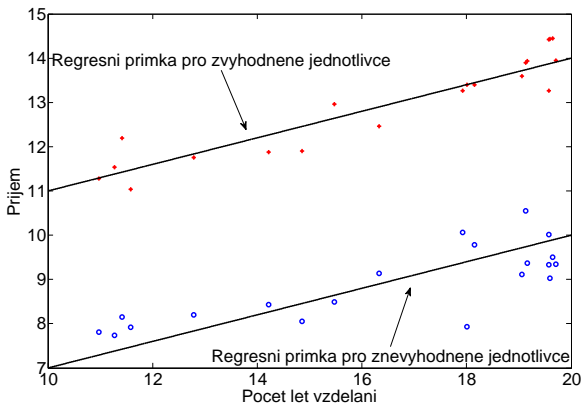
Úvod

- *Individual effects models: fixed effects model a random effects model.*
- Příklad z ekonomie práce (pro motivaci): Y = příjem; X = roky vzdělání.
- Data z průzkumu v průběhu několika let.
- Model:

$$Y_{it} = \alpha + \beta X_{it} + \epsilon_{it}.$$

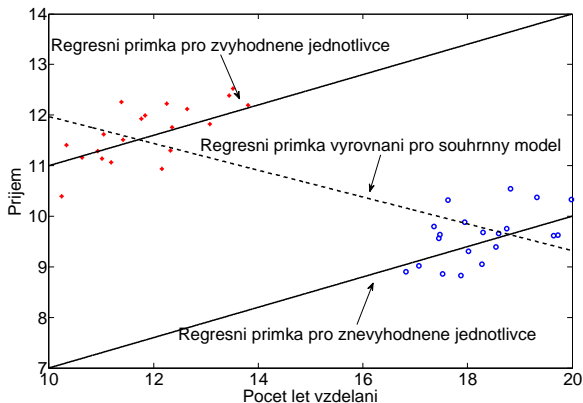
- Stejný vztah pro všechny jednotlivce (α = predikovaný příjem nevzdělaného člověka).

Ilustrace problému souhrnného modelu I



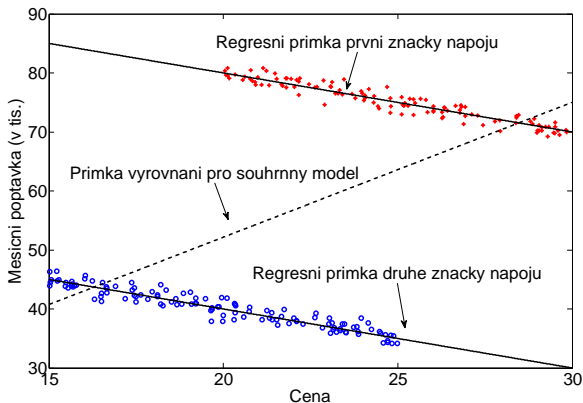
Obrázek: Vztah mezi příjmem a vzděláním pro dva jednotlivce.

Ilustrace problému souhrnného modelu II



Obrázek: Vztah mezi příjmem a vzděláním pro dva jednotlivce.

Další ilustrace problému souhrnného modelu



Obrázek: Vztah mezi cenou a poptávkou po dvou druzích nápojů.

Model individuálních vlivů

- Jediná vysvětlující proměnná:

$$Y_{it} = \alpha_i + \beta X_{it} + \epsilon_{it}.$$

- α_i = individuální vliv (individual effect).
- Panelová data = bohatší datová báze a možnost práce s heterogenitou analyzovaných subjektů.

Obsah tématu

1 Souhrnný model

2 Modely individuálních vlivů

- Model fixních vlivů
- Model náhodných vlivů
- Rozšíření modelů individuálních vlivů

Úvod

- Využití umělých proměnných.
- Konstrukce N umělých proměnných, $D^{(j)}$ pro $j = 1, \dots, N$.
- $D_{it}^{(j)} = 1$ pro j -tého jednotlivce a 0 jinak (TN rozměrný vektor).
- Příklad: $N = 4$ a $T = 2$

$$D^{(3)} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 1 \\ 0 \\ 0 \end{bmatrix} .$$

Odhad

- Regrese:

$$Y_{it} = \alpha_1 D_{it}^{(1)} + \alpha_2 D_{it}^{(2)} + \dots + \alpha_N D_{Nt}^{(N)} + \beta X_{it} + \epsilon_{it}.$$

- Lineární regresní model (hodně vysvětlujících proměnných) – zobecnění i pro „vícenásobnou“ regresi.
- Standardní analýza a odhad.

Interpretace

- Standardní.
- Regresní přímka pro i -tého jednotlivce:

$$Y_{it} = \alpha_i + \beta X_{it} + \epsilon_{it}$$

- Specifikace individuálního vlivu.

Problém

- Regrese s obrovským množstvím vysvětlujících proměnných.
- Např. data z průzkumu $N = 5000$ osob v průběhu $T = 5$ let.
- Vícenásobná regrese: odhad $N + k$ koeficientů.
- V regresním modelu při dané informaci v datech: růst počtu parametrů \rightarrow snížení přesnosti odhadu.
- Přesný odhad pro $N = 100$ a $k = 2 \times N = 10$ a $k = 2$; nemožné pro $k \geq N$.
- V modelu fixních vlivů: pokud T malé \rightarrow stejný problém s přesností odhadu.

Možné řešení

- Práce s diferencemi: $\Delta Y_{it} = Y_{it} - Y_{i,t-1}$.

$$\Delta Y_{it} = \beta \Delta X_{it} + \Delta \epsilon_{it}.$$

- Jen jeden resp. k koeficientů k odhadu.
- Pokud $\Delta \epsilon_{it}$ splňují klasické předpoklady \rightarrow OLS je BLUE.
- Pokud ϵ_{it} splňuje klasické předpoklady $\rightarrow \Delta \epsilon_{it}$ nesplňuje $\rightarrow \text{cov}(\Delta \epsilon_{it}, \Delta \epsilon_{i,t-1}) \neq 0$.
- Autokorelaci konzistentní estimátor: Neweyho-Westův estimátor (Whitney K. Newey a Kenneth D. West).
- Newey, West (1987): A Simple, Positive Semi-definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix, *Econometrica*, vol. 55(3), 703–708.

Další problémy s diferencováním

- Počáteční podmínky.
- Hlavně však koeficienty proměnných neměnicích se v čase (příklad výnosů ze vzdělání: počet let studia rodičů, umělé proměnné vyjadřující rasu či pohlaví apod.).
- Pokud Z_i v čase neměnná vysvětlující proměnná:

$$Y_{it} = \alpha_i + \beta_1 Z_i + \beta_2 X_{it} + \epsilon_{it}$$

- Diferencování:

$$\Delta Y_{it} = \Delta \beta_2 X_{it} + \Delta \epsilon_{it}.$$

- Neexistuje způsob odhadu β_1 .
- V programech obvyklý estimátor pracující s centrovanými hodnotami pro každou jednotku (stejný problém s odhadem vlivů v čase neměnných proměnných).

Testování hypotéz

- Regresní model \rightarrow jakýkoliv test hypotéz.
- Pokud klasické předpoklady: t -statistiky, F -statistiky pro sdružené hypotézy.

$$H_0 : \alpha_1 = \dots = \alpha_N$$

- $N - 1$ restrikcí!

$$F = \frac{(R_{UR}^2 - R_R^2)/(N - 1)}{(1 - R_{UR}^2)/(TN - N - k)}.$$

Obsah tématu

1 Souhrnný model

2 Modely individuálních vlivů

- Model fixních vlivů
- **Model náhodných vlivů**
- Rozšíření modelů individuálních vlivů

Úvod

- Nepoužívá umělé proměnné → předpoklad, že individuální vliv je náhodná veličina (random effects model).

$$Y_{it} = \alpha_i + \beta X_{it} + u_{it}$$

- $\alpha_i = \alpha + v_i$ (náhodná veličina).
- Alternativně:

$$Y_{it} = \alpha + \beta X_{it} + \epsilon_{it}$$

- $\epsilon_{it} = v_i + u_{it}$.
- Netřeba N umělých proměnných!

Předpoklady

- Obvykle: u_{it} splňuje klasické předpoklady $\rightarrow N(0, \sigma_u^2)$ (idiosyncratický rozptyl).
- Obvyklý předpoklad: $v_{it} \sim N(0, \sigma_v^2)$ (průřezový rozptyl).
- Vlastnosti ϵ_{it} :

$$E(\epsilon_{it}) = 0,$$

$$\text{var}(\epsilon_{it}) = \sigma_u^2 + \sigma_v^2,$$

$$\text{cov}(\epsilon_{it}, \epsilon_{jt}) = 0 \quad i \neq j,$$

$$\text{cov}(\epsilon_{it}, \epsilon_{js}) = 0 \quad i \neq j, s \neq t,$$

$$\text{cov}(\epsilon_{it}, \epsilon_{is}) = \sigma_v^2 \quad s \neq t,$$

Analýza

- Pro téhož jednotlivce jsou náhodné složky korelovány v čase.
- OLS není BLUE \rightarrow GLS.
- Při OLS chybné použití $var(\hat{\beta}) \rightarrow$ nekorektní použití t -testů apod.
- Dobré ekonometrické programy spočítají korektní směrodatnou odchylku (různé GLS estimátory).
- Možnost OLS odhadu s *panel-corrected standard errors*.
- Při nesplnění klasických předpokladů pro u_{it} (heteroskedasticita a autokorelace) \rightarrow použití GLS estimátorů kombinujících myšlenky z částí věnovaných daným problémům.

Testování hypotéz

- Standardní přístup (při korektních odhadech rozptylu GLS estimátoru).
- Test použití souhrnného modelu nebo modelu náhodných vlivů.
- Test $H_0 : \sigma_u^2 = 0$ vhodnosti souhrnného modelu (Breuschův-Paganův LM test).
- Test věrohodnostního poměru; obvykleji test Lagrangeových multiplikátorů (nutný jen odhad souhrnného modelu).
- Pokud ekonometrický software tyto testy neumí, možno OLS odhad souhrnného modelu a statistika:

$$LM = \frac{TN}{2(T-1)} \left[\frac{\sum_{i=1}^N [\sum_{t=1}^T \hat{\epsilon}_{it}]^2}{\sum_{i=1}^N \sum_{t=1}^T \hat{\epsilon}_{it}^2} - 1 \right]^2$$

- $\hat{\epsilon}_{it}$ pro $i = 1, \dots, N$ a $t = 1, \dots, T$; OLS rezidua ze souhrnného modelu.
- χ_1^2 rozdělení.

Další problémy

- Model náhodných vlivů: možnost korelace ϵ_{it} a vysvětlujících proměnných \rightarrow potřeba metody instrumentálních proměnných.
- Např.: výnosy ze vzdělání a nepozorovaný talent (korelován s příjmem i roky vzdělání).
- V modelu panelových dat: nepozorovaný talent skrze α_i (výhoda panelových dat – analýza heterogenity).
- Model náhodných vlivů: individuální vliv skrze chybu regrese! (mnohdy automatická korelace s vysvětlující proměnnou)
- Neplatí pro model fixních vlivů.

Hausmanův test

- RE model pokud individuální vliv nekorelován s náhodnou složkou, jinak FE model nebo IV estimátor pro model náhodných vlivů.
- Hausmanův test vhodnosti RE modelu.
- H_0 : individuální vliv nekorelován s žádnou z vysvětlujících proměnných.
- Myšlenka: pokud H_0 pravdivá, jsou RE i FE estimátory konzistentní.
- Pokud H_0 neplatí: RE nekonzistentní \rightarrow odlišný od FE.
- Přesný vztah komplikovaný \times přibližná podoba pracuje s $(\hat{\beta}_{RE} - \hat{\beta}_{FE})^2$; χ^2_{k-1} rozdělení (k vysvětlujících proměnných nezahrnujících úroveňovou konstantu ani umělé proměnné v rámci FE estimátoru).

IV odhad v modelu náhodných vlivů

- Komplikované.
- Populární Hausmanův-Taylorův estimátor a Arellanův-Bondův estimátor (estimátor metody zobecněných momentů, aplikovaný v dynamických panelech).
- Možnost jejich použití v praxi (ekonometrické programy a snadnost použití).
- Intuice: viz např. Koop (2008).
- Rozdělení proměnných: v čase proměnné a nekorelované $X_{it}^{(1)}$; v čase proměnné a korelované $X_{it}^{(2)}$; v čase neměnné a nekorelované $Z_i^{(1)}$; v čase neměnné a korelované $Z_i^{(2)}$
- Počet těchto proměnných: K_1, K_2, L_1, L_2 ; potřeba $K_2 + L_2$ instrumentálních proměnných.

Obsah tématu

1 Souhrnný model

2 Modely individuálních vlivů

- Model fixních vlivů
- Model náhodných vlivů
- Rozšíření modelů individuálních vlivů

Časový vliv

- Zachycení odlišností v každém čase (např. hospodářský cyklus):

$$Y_{it} = \alpha_i + \gamma_t + \beta X_{it} + u_{it}$$

- Model fixních vlivů: zavedení umělých proměnných pro každý časový okamžik.
- Model náhodných vlivů: $\gamma_t = \gamma + \omega_t$

$$Y_{it} = (\alpha + \gamma) + \beta X_{it} + \epsilon_{it}$$

kde $\epsilon_{it} = v_i + \omega_t + u_{it}$.

- Analogické problémy → specifická podoba GLS estimátoru pro korelované náhodné složky.

Metody časových řad pro panelová data

- Metody testů panelových jednotkových kořenů a panelových kointegračních testů.
- Např.

$$Y_{it} = \alpha + \rho Y_{i,t-1} + \beta_1 X_{it}^{(1)} + \beta_2 X_{it}^{(2)} + \beta_3 Z_i^{(1)} + \beta_4 Z_i^{(2)} + \epsilon_{it}$$

- $Y_{i,t-1}$ velmi pravděpodobně korelováno s náhodnou složkou (obsahuje individuální vliv $v_i \rightarrow$ obsažen v ϵ_{it} i $\epsilon_{i,t-1}$ – obsaženo ve výrazu pro $Y_{i,t-1}$).
- Potřeba IV estimátoru (zmíněné dříve).

Model náhodných koeficientů

- Příklad pro jedinou vysvětlující proměnnou:

$$Y_{it} = \alpha_i + \beta_i X_{it} + \epsilon_{it}$$

- Možnost různých mezních vlivů → další rozměr heterogenity.
- Např. v marketingu: spotřebitelé reagují různě na reklamní akce (někteří loajální své značce ⇒ nulová reakce na speciální nabídku pro nový produkt; někteří ochotni více experimentovat s novými produkty ⇒ promo akce na nový produkt vede z jejich strany ke koupi tohoto produktu).
- Různá β_s v aplikacích z marketingu.
- Příklad výnosů ze vzdělání: různý stupeň výhod z dodatečného školního vzdělávání.

Model náhodných koeficientů – pokračování

- Různé koeficienty pro různé jednotlivce.
- Proč ne samostatné regrese?
- Pokud není velké $T \rightarrow N$ regresí nemusí dát přesné odhady.
- Vícenásobná regrese: odhad $N(k + 1)$ koeficientů (k je počet vysvětlujících proměnných).
- Model náhodných koeficientů: různé koeficienty \times dodává určitou společnou strukturu pro všechny jednotlivce, kdy β_s je ze společného rozdělení.
- Řada možných rozdělení (např. normální rozdělení); příklad z marketingu pro dva různé typy spotřebitelů \rightarrow rozdělení umožňující dvě možné hodnoty koeficientu sklonu.
- Odhad v rámci některých programů; v současnosti velký zájem o tyto modely a jejich technickou analýzu (populární bayesovské metody a tzv. hierarchické modely).