

# Základy ekonometrie

## VII. Metoda instrumentálních proměnných

# Obsah tématu

- 1 Nezávislá náhodná vysvětlující proměnná
- 2 Korelovaná náhodná vysvětlující proměnná
  - Metoda instrumentálních proměnných
  - Hausmanův test
  - Sarganův test
  - Test slabých instrumentů
- 3 Příčiny korelace
  - Chyby měření
  - Model simultánních rovnic
- 4 Příklad instrumentů

- Experimentální vědy  $\times$  ekonomie  $\Rightarrow$  nenáhodnost vysvětlujících proměnných?
- Náhodnost vysvětlující proměnné: nekorelovanost  $\times$  korelovanost s náhodnou složkou.
- Různý vliv na kvalitu OLS estimátoru  $\rightarrow$  metoda instrumentálních proměnných.
- Asymptotické výsledky!
- Koncept **konzistentního estimátoru**.

- Jednoduchý regresní model:

$$Y_i = \beta X_i + \epsilon_i.$$

- Modifikované klasické předpoklady:

- 1  $E(\epsilon_i) = 0$ .
- 2  $\text{var}(\epsilon_i) = \sigma^2$ .
- 3  $\text{cov}(\epsilon_i, \epsilon_j) = 0$  pro  $i \neq j$ .
- 4  $\epsilon_i$  má normální rozdělení.
- 5  $X_i$  je náhodná proměnná.

- Klíčové vlastnosti  $X_i$ .

# Obsah tématu

- 1 Nezávislá náhodná vysvětlující proměnná
- 2 Korelovaná náhodná vysvětlující proměnná
  - Metoda instrumentálních proměnných
  - Hausmanův test
  - Sarganův test
  - Test slabých instrumentů
- 3 Příčiny korelace
  - Chyby měření
  - Model simultánních rovnic
- 4 Příklad instrumentů

# Značení

- $X_i$  jsou  $i = 1, \dots, N$  je i.i.d. náhodné veličiny:

$$E(X_i) = \mu_X,$$
$$\text{var}(X_i) = \sigma_X^2.$$

- Vysvětlující proměnná a regresní chyba jsou vzájemně nezávislé  $\Rightarrow \epsilon_i$  nekorelováno s jakoukoli funkcí  $X_j$ .

# Východiska

- Při splnění všech klasických předpokladů:

$$\hat{\beta} \sim N\left(\beta, \frac{\sigma^2}{\sum X_i^2}\right).$$

- S využitím asymptotické teorie: asymptotická (aproximativní platnost).
- OLS estimátor:

$$\hat{\beta} = \frac{\sum_{i=1}^N X_i Y_i}{\sum_{i=1}^N X_i^2},$$

alternativně

$$\hat{\beta} = \beta + \frac{\sum X_i \epsilon_i}{\sum X_i^2}.$$

# Nestrannost

- Důkaz:

$$\begin{aligned} E(\hat{\beta}) &= \beta + E\left(\frac{\sum X_i \epsilon_i}{\sum X_i^2}\right) \\ &= \beta + E\left(\sum \left[\frac{X_i}{\sum X_i^2}\right] \epsilon_i\right) \\ &= \beta + \sum E\left(\frac{X_i}{\sum X_i^2}\right) E(\epsilon_i) \\ &= \beta. \end{aligned}$$



## Další aspekty

- Asymptotické rozdělení OLS estimátoru:

$$N\left(\beta, \frac{\sigma^2}{\sum X_i^2}\right).$$

- Aproximativně stejné výsledky jako dříve.

# Obsah tématu

- 1 Nezávislá náhodná vysvětlující proměnná
- 2 Korelovaná náhodná vysvětlující proměnná
  - Metoda instrumentálních proměnných
  - Hausmanův test
  - Sarganův test
  - Test slabých instrumentů
- 3 Příčiny korelace
  - Chyby měření
  - Model simultánních rovnic
- 4 Příklad instrumentů

# Úvod

- Vysvětlující proměnná a chyba regrese jsou vzájemně korelovány:

$$\text{cov}(X_i, \epsilon_i) \neq 0.$$

- OLS estimátor je vychýlený a potřebujeme jiný estimátor: estimátor metody instrumentálních proměnných (IV estimátor).
- Asymptotická vlastnost: OLS je nekonzistentní.
- Vychýlenost:

$$E(\hat{\beta}) = \beta + E\left(\frac{\sum X_i \epsilon_i}{\sum X_i^2}\right).$$

- Intuice:  $E(\sum X_i \epsilon_i) = \sum E(X_i \epsilon_i) = \sum \text{cov}(X_i, \epsilon_i) \neq 0.$

## Důkaz nekonzistence OLS estimátoru

$$\begin{aligned}
 \text{plim}(\hat{\beta}) &= \text{plim}\left(\beta + \frac{\sum X_i \epsilon_i}{\sum X_i^2}\right) \\
 &= \beta + \text{plim}\left(\frac{\sum X_i \epsilon_i}{\sum X_i^2}\right) \text{ dle Slutského teorému} \\
 &= \beta + \text{plim}\left(\frac{\frac{1}{N} \sum X_i \epsilon_i}{\frac{1}{N} \sum X_i^2}\right) \\
 &= \beta + \frac{\text{plim}\left(\frac{1}{N} \sum X_i \epsilon_i\right)}{\text{plim}\left(\frac{1}{N} \sum X_i^2\right)} \text{ dle Slutského teorému.}
 \end{aligned}$$

- Zákon velkých čísel:  $\text{plim}\left(\frac{1}{N} \sum X_i^2\right) = E(X_i^2) = \sigma_X^2 + \mu_X^2$ .

## Důkaz nekonzistence OLS estimátoru (dokončení)

$$\text{plim}(\hat{\beta}) = \beta + \frac{\text{plim}\left(\frac{1}{N} \sum X_i \epsilon_i\right)}{\sigma_X^2 + \mu_X^2}$$

- Zákon velkých čísel pro  $\text{plim}\left(\frac{1}{N} \sum X_i \epsilon_i\right)$ :

$$\text{plim}\left(\frac{1}{N} \sum X_i \epsilon_i\right) = E(X_i \epsilon_i) = \text{cov}(X_i, \epsilon_i) \neq 0,$$

tedy

$$\text{plim}(\hat{\beta}) = \beta + \frac{\text{cov}(X_i, \epsilon_i)}{\sigma_X^2 + \mu_X^2} \neq \beta.$$

# Obsah tématu

- 1 Nezávislá náhodná vysvětlující proměnná
- 2 **Korelovaná náhodná vysvětlující proměnná**
  - **Metoda instrumentálních proměnných**
  - Hausmanův test
  - Sarganův test
  - Test slabých instrumentů
- 3 Příčiny korelace
  - Chyby měření
  - Model simultánních rovnic
- 4 Příklad instrumentů

# Instrumentální proměnná

- *Instrumentální proměnná (instrument)*,  $Z$ .
- Náhodná veličina, která je nekorelovaná s chybou regrese, ale která je korelována s vysvětlující proměnnou.
- Značení:

$$E(Z_i) = \mu_Z,$$
$$\text{var}(Z_i) = \sigma_Z^2.$$

- Nekorelovanost s chybou regrese:  $\text{cov}(Z_i, \epsilon_i) = 0$ .
- Korelovanost s vysvětlující proměnnou:  
 $\text{cov}(X_i, Z_i) = E(X_i Z_i) - \mu_Z \cdot \mu_X = \sigma_{XZ} \neq 0$ .

# IV estimátor

$$\hat{\beta}_{IV} = \frac{\sum_{i=1}^N Z_i Y_i}{\sum X_i Z_i}.$$

- Konzistentní estimátor, výsledky odhadu využitelné při konstrukci intervalů spolehlivosti a testování hypotéz.
- Asymptotické vlastnosti:

$$\hat{\beta}_{IV} \sim N \left( \beta, \frac{(\sigma_Z^2 + \mu_Z^2)\sigma^2}{N(\sigma_{XZ} + \mu_X\mu_Z)^2} \right).$$



# Důkaz konzistence IV estimátoru

$$\text{plim}(\hat{\beta}_{IV}) = \text{plim}\left(\beta + \frac{\sum Z_i \epsilon_i}{\sum X_i Z_i}\right) = \beta + \text{plim}\left(\frac{\sum Z_i \epsilon_i}{\sum X_i Z_i}\right)$$

dle Slutského teoremu

$$= \beta + \text{plim}\left(\frac{\frac{1}{N} \sum Z_i \epsilon_i}{\frac{1}{N} \sum X_i Z_i}\right) = \beta + \frac{\text{plim}\left(\frac{1}{N} \sum Z_i \epsilon_i\right)}{\text{plim}\left(\frac{1}{N} \sum X_i Z_i\right)}$$

dle Slutského teoremu.

- Zákon velkých čísel:  $\text{plim}\left(\frac{1}{N} \sum Z_i \epsilon_i\right) = E(Z_i \epsilon_i) = 0$ .
- Zákon velkých čísel:  $\text{plim}\left(\frac{1}{N} \sum X_i Z_i\right) = E(X_i Z_i) = \sigma_{XZ} + \mu_X \mu_Z$ .

$$\text{plim}(\hat{\beta}_{IV}) = \beta + \frac{0}{\sigma_{XZ} + \mu_X \mu_Z} = \beta,$$

## Důkaz asymptotické normality IV estimátoru

$$\sqrt{N}(\hat{\beta}_{IV} - \beta) = \sqrt{N} \frac{\sum Z_i \epsilon_i}{\sum X_i Z_i} = \sqrt{N} \frac{\frac{1}{N} \sum Z_i \epsilon_i}{\frac{1}{N} \sum X_i Z_i}.$$

- Centrální limitní věta: pro  $N \rightarrow \infty$

$$\sqrt{N} \frac{1}{N} \sum Z_i \epsilon_i \sim N(0, \text{var}(Z_i \epsilon_i)).$$

- Z nekorelovanosti:  $\text{var}(Z_i \epsilon_i) = (\sigma_Z^2 + \mu_Z^2) \sigma^2$ .
- Platí:  $\text{plim} \left( \frac{1}{N} \sum X_i Z_i \right) = \sigma_{XZ} + \mu_X \mu_Z$ .
- S využitím Cramerova teorému:

$$\sqrt{N}(\hat{\beta}_{IV} - \beta) \xrightarrow{N} N \left( 0, \frac{(\sigma_Z^2 + \mu_Z^2) \sigma^2}{(\sigma_{XZ} + \mu_X \mu_Z)^2} \right).$$

# Základní otázky

- „Co když máme model vícenásobné regrese zahrnující více než jednu vysvětlující proměnnou?“
  - Jeden instrument pro každou korelovanou vysvětlující proměnnou (pro nekorelované vysvětlující proměnné jsou instrumenty tyto samotné proměnné).
  - Vztah pro IV estimátor (bez maticové algebry komplikovaný):

$$(Z'X)^{-1}Z'y.$$

- „Co když máme více instrumentálních proměnných, než je potřeba?“
  - Použít jeden z nich.
  - Použít všechny  $\rightarrow$  GIVE (generalized instrumental variables estimator) resp. TSLS (two-stages least squares).

## GIVE estimátor

- Příklad: jednoduchá regrese,  $X$  korelováno s náhodnou složkou, dva instrumenty ( $Z_1$  a  $Z_2$ ).
- Výchozí regrese vysvětlující proměnné na všechny instrumenty:

$$X_i = \gamma_0 + \gamma_1 Z_{1i} + \gamma_2 Z_{2i} + u_i,$$

- Vyrovnané hodnoty:

$$\hat{X}_i = \hat{\gamma}_0 + \hat{\gamma}_1 Z_{1i} + \hat{\gamma}_2 Z_{2i}.$$

- Lze ukázat, že  $\hat{X}$  je nekorelována s náhodnými složkami původní regrese  $\Rightarrow$  vhodný instrument.
- GIVE (konzistentní estimátor):

$$\hat{\beta}_{GIVE} = \frac{\sum_{i=1}^N \hat{X}_i Y_i}{\sum_{i=1}^N X_i \hat{X}_i}.$$

# Obsah tématu

- 1 Nezávislá náhodná vysvětlující proměnná
- 2 **Korelovaná náhodná vysvětlující proměnná**
  - Metoda instrumentálních proměnných
  - **Hausmanův test**
  - Sarganův test
  - Test slabých instrumentů
- 3 Příčiny korelace
  - Chyby měření
  - Model simultánních rovnic
- 4 Příklad instrumentů

# Motivace

- Test korelace vysvětlující proměnné s náhodnou složkou.
- V případě nekorelovanosti: OLS je BLUE.
- Jen intuitivní vysvětlení.

# Motivace

- $H_0$ : vysvětlující proměnné v modelu vícenásobné regrese jsou nekorelovány s náhodnou složkou.
- Myšlenka: pokud  $H_0$  platí, jsou OLS a IV estimátory konzistentní (podobné výsledky)  $\times$  pokud  $H_0$  neplatí  $\rightarrow$  odlišné odhady.
- Testová statistika měřící rozdíly mezi  $\hat{\beta}$  a  $\hat{\beta}_{IV}$ .
- Provedení i s využitím OLS metod: příklad jednoduché regrese s jediným instrumentem.
- Model:  $Y_i = \alpha + \beta X_i + \epsilon_i$ .
- Hausmanův test: regrese  $Y_i = \alpha + \beta X_i + \gamma Z_i + \epsilon_i$ .
- Ekvivalentní standardnímu  $t$ -testu hypotézy  $H_0 : \gamma = 0$  (pokud  $Z$  statisticky významné  $\rightarrow$  zamítnutí  $H_0$ ).

## Příklad vícenásobné regrese.

- GIVE, tři vysvětlující proměnné ( $X_1$ ,  $X_2$  a  $X_3$ ), dvě korelovány ( $X_2$  a  $X_3$ ), dva instrumenty pro  $X_2$  ( $Z_1$  a  $Z_2$ ) a tři pro  $X_3$  ( $Z_3$ ,  $Z_4$  a  $Z_5$ ).
  - 1 OLS regrese  $X_2$  na úrovnovou konstantu,  $Z_1$  a  $Z_2 \rightarrow \hat{X}_2$ .
  - 2 OLS regrese  $X_3$  na úrovnovou konstantu,  $Z_3$ ,  $Z_4$  a  $Z_5 \rightarrow \hat{X}_3$ .
  - 3 OLS regrese  $Y$  na úrovnovou konstantu,  $X_1$ ,  $X_2$ ,  $X_3$ ,  $\hat{X}_2$  a  $\hat{X}_3$ .
  - 4  $F$ -test hypotézy, že koeficienty u  $\hat{X}_2$  a  $\hat{X}_3$  současně nulové.
  - 5 Pokud  $F$ -test zamítá nulovou hypotézu  $\rightarrow$  GIVE  $\times$  jinak OLS.



# Obsah tématu

- 1 Nezávislá náhodná vysvětlující proměnná
- 2 **Korelovaná náhodná vysvětlující proměnná**
  - Metoda instrumentálních proměnných
  - Hausmanův test
  - **Sarganův test**
  - Test slabých instrumentů
- 3 Příčiny korelace
  - Chyby měření
  - Model simultánních rovnic
- 4 Příklad instrumentů

# Motivace

- Jak poznat, že  $Z$  je platný instrument?
- Komplikované: korelace s  $X$  snadná  $\times$  nekorelovanost s náhodnou složkou.
- Pokud jen jeden potenciální instrument ke každé proměnné  $\rightarrow$  neproveditelný test validity.

# Intuice

$$Y_i = \beta X_i + \epsilon_i$$

- Jediný potenciální instrument,  $Z \rightarrow \text{test } \text{cov}(Z, \epsilon_i) = 0$ .
- Jaká rezidua použít?
- $\hat{\epsilon}_i^{IV} = Y_i - \hat{\beta}_{IV} X_i$  nevhodná (možná nekonzistence, protože  $Z$  nemusí být platný instrument  $\times$  to chceme testovat).
- OLS rezidua nevhodná (možná nekonzistence + Hausmanův test nepoužitelný, protože  $Z$  nemusí být platný instrument).
- Problém testování validity instrumentů.
- Mnohdy instrumenty nabízí ekonomická teorie.

# Sarganův test

- Pokud více instrumentů než vysvětlujících proměnných korelovaných s náhodnou složkou.
- Jeden z testů → odvození nad rámec kurzu Základy ekonometrie.
- Postup ( $k$  vysvětlujících potenciálně korelovaných proměnných a  $r$  instrumentů, kdy  $r > k$ ):
  - 1 Regrese  $Y$  na úroňovou konstantu,  $X_1, \dots, X_k$  využitím zobecněného IV estimátoru a získání  $\hat{\epsilon}_i^{IV}$ .
  - 2 OLS regrese IV reziduí,  $\hat{\epsilon}_i^{IV}$ , na úroňovou konstantu a všechny instrumenty,  $Z_1, \dots, Z_r$  a získání koeficientu determinace,  $R^2$ .
  - 3 Sarganova testová statistika  $NR^2$  + kritické hodnoty z  $\chi^2(r - k)$  rozdělení.
- Test Lagrangeových multiplikátorů (asymptotické rozdělení) + nulová hypotéza o nekorelovanosti.

# Obsah tématu

- 1 Nezávislá náhodná vysvětlující proměnná
- 2 **Korelovaná náhodná vysvětlující proměnná**
  - Metoda instrumentálních proměnných
  - Hausmanův test
  - Sarganův test
  - **Test slabých instrumentů**
- 3 Příčiny korelace
  - Chyby měření
  - Model simultánních rovnic
- 4 Příklad instrumentů

# Test slabých instrumentů – jeden instrument

- Slabé instrumenty  $\rightarrow$  IV estimátor zkreslený a vysoké směrodatné odchylky.

$$y = \beta_1 + \beta_2 x_2 + \dots + \beta_G x_G + \beta_{G+1} x_{G+1} + \epsilon$$

- $x_2, \dots, x_G$  exogenní a nekorelované s  $\epsilon$ ,  $x_{G+1}$  endogenní.
- Jediný instrument  $z_1$ .
- Otázka síly vztahu mezi  $x_{G+1}$  a  $z_1$  (po zahrnutí vlivu ostatních exogenních proměnných).

$$x_{G+1} = \gamma_1 + \gamma_2 x_2 + \dots + \gamma_G x_G + \theta_1 z_1 + u.$$

- $H_0 : \theta_1 = 0$ , oproti  $H_1 : \theta_1 \neq 0$ .
- Zamítnutí na hladině významnosti  $\alpha = 0.05$  nedostatečné pro ověření síly instrumentu  $\rightarrow$  „výrazné“ zamítnutí  $\rightarrow$  hrubé pravidlo  $F$ -statistika menší než 10 nebo  $t$ -statistika menší než 3.3 vede k závěru, že instrument je slabý.

# Test slabých instrumentů – více instrumentů

- Pro počet instrumentů  $L > 1$ :

$$x_{G+1} = \gamma_1 + \gamma_2 x_2 + \dots + \gamma_G x_G + \theta_1 z_1 + \dots + \theta_L z_L + u.$$

- Test  $H_0 : \theta_1 = 0, \theta_2 = 0, \dots, \theta_L = 0$  oproti alternativě, že alespoň jeden je nenulový.
- $F$ -test s podobným hrubým pravidlem (menší než 10 vede ke slabým instrumentům).

## Test slabých instrumentů – více proměnných

- Pro model s více než jednou endogenní proměnnou musí být počet instrumentů větší nebo roven jejich počtu.
- $F$ -test pro každou z regresních rovnic.
- Individuální  $F$ -statistiky neodpovídají všeobecné síle instrumentů  $\longleftrightarrow$  pokud každé  $F > 10$  neznamena to, že nemáme problém slabých proměnných.
- Craggova-Donaldova statistika.
- Alternativní diagnostika: porovnat standardní chyby IV odhadu a OLS odhadu (slabé instrumenty vedou k vyšším standardním odchylkám IV estimátoru).
- Co dělat při slabých instrumentech? Obtížná odpověď, v současnosti předmětem výzkumu (možnost vyhodit z množiny instrumentů ty nejslabší, pokud je dostatek instrumentů).



# Craggova-Donaldova statistika

- Cragg, Donald (1993).
- $G$  exogenních proměnných,  $B$  endogenních proměnných,  $L \geq B$  instrumentů.

$$\text{Cragg-Donald } F = [(N - G - B)/L]/[r_B^2/(1 - r_B^2)],$$

- $r_B$  nejmenší kanonická korelace.
- Nulová hypotéza: instrumenty jsou slabé (oproti alternativě, že nejsou).
- Kritické hodnoty tabelovány: Stock a Yogo (2005).
- Kritéria dle relativního zkreslení (*relative bias*) IV estimátoru vzhledem k OLS estimátoru nebo dle míry zamítnutí (*rejection rate*) resp. významnosti testu (*test size*).

# Obsah tématu

- 1 Nezávislá náhodná vysvětlující proměnná
- 2 Korelovaná náhodná vysvětlující proměnná
  - Metoda instrumentálních proměnných
  - Hausmanův test
  - Sarganův test
  - Test slabých instrumentů
- 3 Příčiny korelace
  - Chyby měření
  - Model simultánních rovnic
- 4 Příklad instrumentů

# Úvod

- Proč existuje korelace vysvětlujících proměnných s náhodnou složkou?
- Jak vybírat instrumenty?
- Nelehký úkol: v časových řadách jsou instrumenty obvykle zpožděné proměnné.

# Obsah tématu

- 1 Nezávislá náhodná vysvětlující proměnná
- 2 Korelovaná náhodná vysvětlující proměnná
  - Metoda instrumentálních proměnných
  - Hausmanův test
  - Sarganův test
  - Test slabých instrumentů
- 3 Příčiny korelace
  - Chyby měření
  - Model simultánních rovnic
- 4 Příklad instrumentů

# Značení

$$Y_i = \beta X_i + \epsilon_i$$

- Pozorujeme:  $X_i^* = X_i + \nu_i$ , kde  $\nu_i$  je i.i.d. s nulovou střední hodnotou, rozptylem  $\sigma_\nu^2$  a je nezávislá na  $\epsilon_i$ .
- Není problém s chybou měření ve vysvětlované proměnné (proč?).
- Problém u vysvětlujících proměnných.

# Zdůvodnění korelace

- Model:

$$\begin{aligned} Y_i &= \beta(X_i^* - \nu_i) + \epsilon_i \\ &= \beta X_i^* + \epsilon_i^*, \end{aligned}$$

- $\epsilon_i^* = \epsilon_i - \beta \nu_i$ .

$$\begin{aligned} \text{cov}(X_i^*, \epsilon_i^*) &= E(X_i^* \epsilon_i^*) \\ &= E[(X_i + \nu_i)(\epsilon_i - \beta \nu_i)] \\ &= -\beta \sigma_\nu^2. \end{aligned}$$

# Obsah tématu

- 1 Nezávislá náhodná vysvětlující proměnná
- 2 Korelovaná náhodná vysvětlující proměnná
  - Metoda instrumentálních proměnných
  - Hausmanův test
  - Sarganův test
  - Test slabých instrumentů
- 3 Příčiny korelace
  - Chyby měření
  - **Model simultánních rovnic**
- 4 Příklad instrumentů

# Úvod

- *Endogenní* proměnná: determinována uvnitř modelu, který nás zajímá.
- *Exogenní* proměnná: determinována mimo model.
- Doposud vysvětlující proměnné exogenní.
- Pokud vysvětlující proměnné endogenní → možná korelace s náhodnou složkou.



# Model nabídky a poptávky

- Model simultánních rovnic (všechno endogenní proměnné).
- Poptávková křivka:

$$Q_D = \beta_D P + \epsilon_D.$$

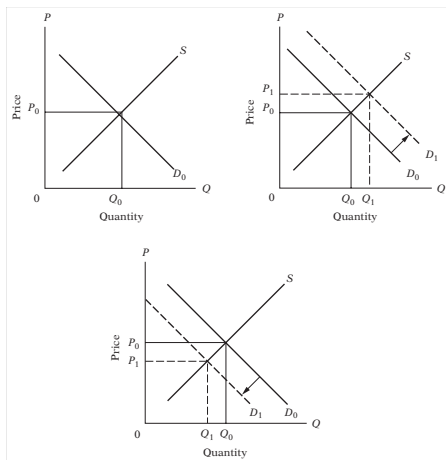
- Nabídková křivka:

$$Q_S = \beta_S P + \epsilon_S.$$

- Rovnováha (rovnice identity):

$$Q_D = Q_S.$$

# Endogenita ceny



Zdroj: Gujarati, Porter (2009) – Basic econometrics.

# Ekonometrický problém

- Údaje o množství a cenách + regrese.
- OLS odhady parametrů sklonu  $\times$  odhady  $\beta_D$  nebo  $\beta_S$ ?
- Nejsme schopni odhadnout.
- Předpoklad závislosti poptávky na důchodu,  $I$  (exogenní):

$$Q_D = \beta_D P + \gamma I + \epsilon_D.$$

- Klasické předpoklady pro náhodné složky.

# Strukturální a redukováaná forma modelu

- *Strukturální forma modelu*: vycházejí z ekonomické teorie.
- *Redukovaná forma modelu*: endogenní proměnné jen na levých stranách.
- Z identity  $Q = Q_D = Q_S$ :

$$\begin{aligned}P &= \frac{-\gamma}{\beta_D - \beta_S} I + \frac{\epsilon_S - \epsilon_D}{\beta_D - \beta_S} \\ &= \pi_1 I + \epsilon_1 \\ Q &= \beta_S(\pi_1 I + \epsilon_1) + \epsilon_S \\ &= \beta_S \pi_1 I + \beta_S \epsilon_1 + \epsilon_S \\ &= \pi_2 I + \epsilon_2.\end{aligned}$$

## Vychýlenost a nekonzistence OLS odhadu

- OLS regrese  $Q$  na  $P$  a  $I$  (poptávková křivka).
- Vysvětlující proměnná  $P$  a chyby regrese,  $\epsilon_D$ , korelovány!

$$\begin{aligned} \text{cov}(P, \epsilon_D) &= E(P\epsilon_D) \\ &= E[(\pi_1 I + \epsilon_1)\epsilon_D] \\ &= E\left[\left(\frac{\epsilon_S - \epsilon_D}{\beta_D - \beta_S}\right)\epsilon_D\right] \\ &= \frac{-\sigma_D^2}{\beta_D - \beta_S} \\ &\neq 0, \end{aligned}$$

- Aplikace přímého OLS odhadu na poptávkovou (i nabídkovou křivku)  
→ vychýlené a nekonzistentní odhady.

# Odhad parametrů redukované formy

- Chyby regrese redukované formy: klasické předpoklady.
- Jediná vysvětlující proměnná: exogenní proměnná  $\Rightarrow$  OLS je BLUE.
- $\hat{\pi}_1$  a  $\hat{\pi}_2 \times$  jaké jsou odhady parametrů strukturální formy modelu?

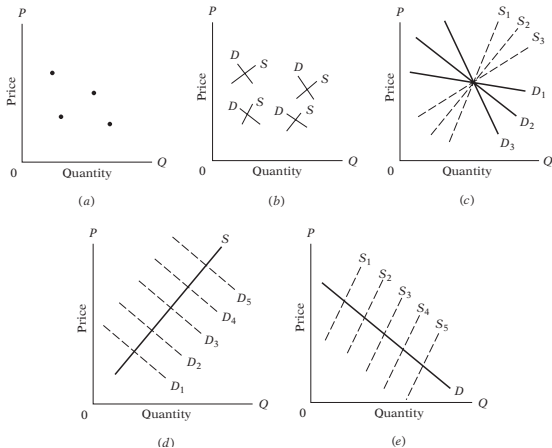
# Vztah strukturálních a redukovaných parametrů

- Někdy možné koeficienty redukované formy použít k odhadu koeficientů strukturální formy.
- V našem příkladu:  $\pi_2 = \beta_S \pi_1 \rightarrow$

$$\hat{\beta}_S = \frac{\hat{\pi}_2}{\hat{\pi}_1}.$$

- Postup = tzv. *nepřímé nejmenší čtverce (indirect least squares)*.
- Pokud rovnice *přesně identifikovatelná*, další možnosti *přeidentifikovatelnost a neidentifikovatelnost*.

# Ilustrace identifikovatelnosti



Zdroj: Gujarati, Porter (2009) – Basic econometrics.



## Vztah k IV estimátoru

- Metoda nepřímých nejmenších čtverců v našem případě = IV odhad nabídkové křivky s  $I$  jako instrumentem (snadno odvoditelné).
- Do skupiny IV estimátorů: *dvoustupňová metoda nejmenších čtverců* (*two-stage least squares – TSLS*), pro přeidentifikovatelné rovnice v rámci modelu simultánních rovnic.
- Nexistuje způsob, jak s využitím estimátoru nepřímých nejmenších čtverců získat odhad parametru  $\beta_D$ !
- Proč? Důchod nevystupuje v křivce nabídky → pokud máme exogenní proměnnou, která není obsažena v některé z rovnicí, lze ji použít jako instrumentální proměnnou pro tuto rovnici.

# Obsah tématu

- 1 Nezávislá náhodná vysvětlující proměnná
- 2 Korelovaná náhodná vysvětlující proměnná
  - Metoda instrumentálních proměnných
  - Hausmanův test
  - Sarganův test
  - Test slabých instrumentů
- 3 Příčiny korelace
  - Chyby měření
  - Model simultánních rovnic
- 4 Příklad instrumentů

# Model výnosů ze vzdělání

- Returns to schooling.
- Závisle proměnná  $Y$  = důchod, a vysvětlující proměnné  $X$  = počet let vzdělávání ve škole a další vysvětlující proměnné (např. pracovní zkušenosti, věk, typ zaměstnání apod.).
- Pravděpodobná korelace  $X$  s náhodnou složkou – osobní kvality jednotlivce (talent, štěstí apod.) v náhodné složce ovlivňují délku vzdělávání.
- OLS vychýlený!

# Přístupy k řešení

- První přístup: nalezení proxy proměnných (např. testy inteligence).
- Druhý přístup: metoda instrumentálních proměnných.
- Instrumenty: charakteristiky rodičů nebo sourozenců (ovlivňuje rozhodování o počtu let vzdělání  $\times$  není přímý vliv na příjem), geografické místo, kde daná osoba žije (komunita, kde je vysoká škola).