

# Az exogenitás és sérülése

Ferenci Tamás  
tamas.ferenci@medstat.hu

Utoljára frissítve: 2023. május 9.

# Tartalom

## 1 Erős exogenitás

# Tartalom

## 1 Erős exogenitás

# Emlékeztetőül

- Az erős exogenitási feltevés sérülésének három gyakorlatban tipikus esete van: kihagyott változó okozta torzítás (confounding!), mérési hiba, szimultaneitás
- Az utóbbi kettő meghaladja a mostani kereteket, így a továbbiakban az elsővel fogunk foglalkozni

# Emlékeztetőül

- Az erős exogenitási feltevés sérülésének három gyakorlatban tipikus esete van: kihagyott változó okozta torzítás (confounding!), mérési hiba, szimultaneitás
- Az utóbbi kettő meghaladja a mostani kereteket, így a továbbiakban az elsővel fogunk foglalkozni

# Változó bevonásának hatása a modellre

Vessük össze ezt a két (demonstráció kedvéért igen kicsi) modellt az esettanulmány feladatára:

$$\widehat{\text{KiadEFt}} = 339,746 + 0,637354 \text{ JovEFt}$$

(13,783)      (0,0064924)

$$T = 8314 \quad \bar{R}^2 = 0,5369 \quad F(1, 8312) = 9637,2 \quad \hat{\sigma} = 662,02$$

(standard errors in parentheses)

$$\widehat{\text{KiadEFt}} = 283,172 + 0,616911 \text{ JovEFt} + 34,1727 \text{ TLetszam}$$

(16,988)      (0,0074136)      (6,0199)

$$T = 8314 \quad \bar{R}^2 = 0,5386 \quad F(2, 8311) = 4852,8 \quad \hat{\sigma} = 660,78$$

(standard errors in parentheses)

Miért változott meg a jövedelem becsült koefficiense?

# Változó bevonásának hatása a modellre

- Mondjuk, hogy a bővebb modell írja le a valóságos helyzetet (a gyakorlatban ezt persze soha nem tudhatjuk, filozófiai kérdés)
- Azaz a valós helyzet a második regresszió
- Az érdekes, hogy ez alapján *előre* meg tudjuk mondani, hogy az első regresszióban mi lesz a jövedelem együttthatója! (... és ebből persze a változás okát is rögtön le tudjuk olvasni)
- A jövedelem ugyanis nem csak a kiadásra hat sztochasztikusan, hanem összefügg a taglétszámmal is:

$$\widehat{\text{TLetszam}} = 1,65553 + 0,000598206 \text{ JovEFt}$$

(0,025067)      (1,1807e-005)

$$T = 8314 \quad \bar{R}^2 = 0,2359 \quad F(1, 8312) = 2566,9 \quad \hat{\sigma} = 1,2040$$

(standard errors in parentheses)

# Változó bevonásának hatása a modellre

- Mondjuk, hogy a bővebb modell írja le a valóságos helyzetet (a gyakorlatban ezt persze soha nem tudhatjuk, filozófiai kérdés)
- Azaz a valós helyzet a második regresszió
- Az érdekes, hogy ez alapján *előre* meg tudjuk mondani, hogy az első regresszióban mi lesz a jövedelem együtthatója! (... és ebből persze a változás okát is rögtön le tudjuk olvasni)
- A jövedelem ugyanis nem csak a kiadásra hat sztochasztikusan, hanem összefügg a taglétszámmal is:

$$\widehat{\text{TLetszam}} = 1,65553 + 0,000598206 \text{ JovEFt}$$

(0,025067)      (1,1807e-005)

$$T = 8314 \quad \bar{R}^2 = 0,2359 \quad F(1, 8312) = 2566,9 \quad \hat{\sigma} = 1,2040$$

(standard errors in parentheses)



# Változó bevonásának hatása a modellre

- Mondjuk, hogy a bővebb modell írja le a valóságos helyzetet (a gyakorlatban ezt persze soha nem tudhatjuk, filozófiai kérdés)
- Azaz a valós helyzet a második regresszió
- Az érdekes, hogy ez alapján *előre* meg tudjuk mondani, hogy az első regresszióban mi lesz a jövedelem együttthatója! (... és ebből persze a változás okát is rögtön le tudjuk olvasni)
- A jövedelem ugyanis nem csak a kiadásra hat sztochasztikusan, hanem összefügg a taglétszámmal is:

$$\widehat{\text{TLetszam}} = 1,65553 + 0,000598206 \text{ JovEFt}$$

(0,025067)      (1,1807e-005)

$$T = 8314 \quad \bar{R}^2 = 0,2359 \quad F(1, 8312) = 2566,9 \quad \hat{\sigma} = 1,2040$$

(standard errors in parentheses)

# Változó bevonásának hatása a modellre

- Mondjuk, hogy a bővebb modell írja le a valóságos helyzetet (a gyakorlatban ezt persze soha nem tudhatjuk, filozófiai kérdés)
- Azaz a valós helyzet a második regresszió
- Az érdekes, hogy ez alapján *előre* meg tudjuk mondani, hogy az első regresszióban mi lesz a jövedelem együttthatója! (... és ebből persze a változás okát is rögtön le tudjuk olvasni)
- A jövedelem ugyanis nem csak a kiadásra hat sztochasztikusan, hanem összefügg a taglétszámmal is:

$$\widehat{\text{TLetszam}} = 1,65553 + 0,000598206 \text{ JovEFt}$$

(0,025067)      (1,1807e-005)

$$T = 8314 \quad \bar{R}^2 = 0,2359 \quad F(1, 8312) = 2566,9 \quad \hat{\sigma} = 1,2040$$

(standard errors in parentheses)

# Változó bevonásának hatása a modellre

- Ebből összerakhatjuk a szűkebb regresszióban a jövedelem együtthatóját:

$$0,637 = 0,617 + 0,000598 \cdot 34,17$$

- A bővebb modellben az együttható 0,617: ennyi a jövedelem direkt hatása (ha egy egységgel nő stb.), és itt véget is ér a sztori, mert a bővebb modellben a taglétszámot állandó értéken tartjuk (v.ö.: c.p.) ezért nincs jelentősége a taglétszám és a jövedelem közti sztochasztikus kapcsolatnak
- A szűkebb modellben viszont a jövedelem egységnyi növekedése a taglétszámot is növeli tendenciájában, a növekvő taglétszám viszont (*önmagában* is!) növeli a kiadást, ez lesz az indirekt hatás
- Teljes hatás = direkt hatás + indirekt hatás(ok)

# Változó bevonásának hatása a modellre

- Ebből összerakhatjuk a szűkebb regresszióban a jövedelem együtthatóját:

$$0,637 = 0,617 + 0,000598 \cdot 34,17$$

- A bővebb modellben az együttható 0,617: ennyi a jövedelem direkt hatása (ha egy egységgel nő stb.), és itt véget is ér a sztori, mert a bővebb modellben a taglétszámot állandó értéken tartjuk (v.ö.: c.p.) ezért nincs jelentősége a taglétszám és a jövedelem közti sztochasztikus kapcsolatnak
- A szűkebb modellben viszont a jövedelem egységnyi növekedése a taglétszámot is növeli tendenciájában, a növekvő taglétszám viszont (*önmagában* is!) növeli a kiadást, ez lesz az indirekt hatás
- Teljes hatás = direkt hatás + indirekt hatás(ok)

# Változó bevonásának hatása a modellre

- Ebből összerakhatjuk a szűkebb regresszióban a jövedelem együtthatóját:

$$0,637 = 0,617 + 0,000598 \cdot 34,17$$

- A bővebb modellben az együttható 0,617: ennyi a jövedelem direkt hatása (ha egy egységgel nő stb.), és itt véget is ér a sztori, mert a bővebb modellben a taglétszámot állandó értéken tartjuk (v.ö.: c.p.) ezért nincs jelentősége a taglétszám és a jövedelem közti sztochasztikus kapcsolatnak
- A szűkebb modellben viszont a jövedelem egységnyi növekedése a taglétszámot is növeli tendenciájában, a növekvő taglétszám viszont (*önmagában* is!) növeli a kiadást, ez lesz az indirekt hatás
- Teljes hatás = direkt hatás + indirekt hatás(ok)

# Változó bevonásának hatása a modellre

- Ebből összerakhatjuk a szűkebb regresszióban a jövedelem együtthatóját:

$$0,637 = 0,617 + 0,000598 \cdot 34,17$$

- A bővebb modellben az együttható 0,617: ennyi a jövedelem direkt hatása (ha egy egységgel nő stb.), és itt véget is ér a sztori, mert a bővebb modellben a taglétszámot állandó értéken tartjuk (v.ö.: c.p.) ezért nincs jelentősége a taglétszám és a jövedelem közti sztochasztikus kapcsolatnak
- A szűkebb modellben viszont a jövedelem egységnyi növekedése a taglétszámot is növeli tendenciájában, a növekvő taglétszám viszont (*önmagában* is!) növeli a kiadást, ez lesz az indirekt hatás
- Teljes hatás = direkt hatás + indirekt hatás(ok)

# Változó bevonásának hatása a modellre

- A szűkebb regresszióban nem tudjuk *izolálni* a taglétszám hatását: ha a jövedelem nő, az a bővebb modellben nem társul a taglétszám növekedésével (v.ö. a paraméter c.p. értelmezésével), a szűkebb modellben viszont igen (hiszen ott nem endogén változó a taglétszám) → a szűkebb modellben a kihagyott változón keresztül terjedő hatások is *beépülnek* az együtthatóba
- Azaz: a bővebb regresszióval, az új változó bevonásával védekeztünk a confounding ellen (kiszűrtük a hatását: kontrolláltunk az újonnan bevont változóra)
- A gyakorlatban persze nem tudhatjuk, hogy mi a „kihagyott változó”

# Változó bevonásának hatása a modellre

- A szűkebb regresszióban nem tudjuk *izolálni* a taglétszám hatását: ha a jövedelem nő, az a bővebb modellben nem társul a taglétszám növekedésével (v.ö. a paraméter c.p. értelmezésével), a szűkebb modellben viszont igen (hiszen ott nem endogén változó a taglétszám) → a szűkebb modellben a kihagyott változón keresztül terjedő hatások is *beépülnek* az együtthatóba
- Azaz: a bővebb regresszióval, az új változó bevonásával védekeztünk a confounding ellen (kiszűrtük a hatását: kontrolláltunk az újonnan bevont változóra)
- A gyakorlatban persze nem tudhatjuk, hogy mi a „kihagyott változó”



# Változó bevonásának hatása a modellre

- A szűkebb regresszióban nem tudjuk *izolálni* a taglétszám hatását: ha a jövedelem nő, az a bővebb modellben nem társul a taglétszám növekedésével (v.ö. a paraméter c.p. értelmezésével), a szűkebb modellben viszont igen (hiszen ott nem endogén változó a taglétszám) → a szűkebb modellben a kihagyott változón keresztül terjedő hatások is *beépülnek* az együtthatóba
- Azaz: a bővebb regresszióval, az új változó bevonásával védekeztünk a confounding ellen (kiszűrtük a hatását: kontrolláltunk az újonnan bevont változóra)
- A gyakorlatban persze nem tudhatjuk, hogy mi a „kihagyott változó”

# A specifikációs torzítás és iránya

- Akkor van tehát kihagyott változó okozta torzítás, ha *egyszerre* fennáll két feltétel: a kihagyott változónak van – önmagában – hatása az eredményváltozóra (tehát a  $\beta$ -ja nem nulla), és korrelált a bennmaradt magyarázó változóval
- Ebből adódóan a torzítás iránya negatív és pozitív is lehet attól függően, hogy ez a két tényező milyen előjelű
- Belátható, hogy többváltozós esetben, ha csak egy változónak is van endogenitási baja, akkor is torzított lesz az összes változó becsült koefficiense

# A specifikációs torzítás és iránya

- Akkor van tehát kihagyott változó okozta torzítás, ha *egyszerre* fennáll két feltétel: a kihagyott változónak van – önmagában – hatása az eredményváltozóra (tehát a  $\beta$ -ja nem nulla), és korrelált a bennmaradt magyarázó változóval
- Ebből adódóan a torzítás iránya negatív és pozitív is lehet attól függően, hogy ez a két tényező milyen előjelű
- Belátható, hogy többváltozós esetben, ha csak egy változónak is van endogenitási baja, akkor is torzított lesz az összes változó becsült koefficiense

# A specifikációs torzítás és iránya

- Akkor van tehát kihagyott változó okozta torzítás, ha *egyszerre* fennáll két feltétel: a kihagyott változónak van – önmagában – hatása az eredményváltozóra (tehát a  $\beta$ -ja nem nulla), és korrelált a bennmaradt magyarázó változóval
- Ebből adódóan a torzítás iránya negatív és pozitív is lehet attól függően, hogy ez a két tényező milyen előjelű
- Belátható, hogy többváltozós esetben, ha csak egy változónak is van endogenitási baja, akkor is torzított lesz az összes változó becsült koefficiense