

Regressziós modellek alapjai	Olvasási idő: 20 perc 	Készítette: Dr. habil. Kiss Gábor Dávid 
---	---	---

Bevezetés

Cikk vagy szakdolgozat írása során az elméleti, irodalom feldolgozó fejezet során az a célunk, hogy bevezessük azokat a változókat és modelleket, amelyekkel majd mi is dolgozni fogunk. Ez alapján be tudjuk mutatni, hogy a szakirodalomban melyek a már ismert tények és a még nem vizsgált területek. Éppen ezért az elméleti fejezet végére egyfajta konklúzióként be kell illeszteni egy „elméleti modell” névre hallgató alfejezetet, ahol bemutatásra kerül az általunk vizsgált modell (esetleg annak különböző alváltozatai), a változók és az egyes változóknál várt koefficiensek előjele és annak értelmezése. Ezzel nagyban megkönnyítjük a bíráló munkáját és a kapott eredményeinket is könnyebben tudjuk bemutatni²³.

Általános bemeneti és kimeneti követelmények

- Stacioner bemenet: ADF-teszt $p < 0.05$
 - Differenciálást követően: $I(1)$
- Hasonló skála
 - Logaritmizálás (ha $P > 0$)
 - leskálázás 10^{-n} -al való szorzással
 - standardizálás: Z-score
- Korlátozás: a változók száma, az idősor hossza és a visszatekintés: $t * v - v^2 * l > 0$
 - t: idősor hossza (évek, negyedévek száma)
 - l: késleltetés (lag) száma
 - v: változók száma
 - általános szabály: kb 30 időegység hosszúságú idősort érdemes használni normális eloszlású hibatagokhoz, ugyanakkor érdemes 100 alatt maradni, hogy szignifikáns eredményeket kaphassunk.
- Kimenet (hibatagok – residuals - diagnosztikája):
 - Autokorreláció hiánya:
 - Durbin-Watson $\sim 2 \rightarrow 1.85 < D-W < 2.15$
 - Ljung-Box $p > 0.05$
 - Model-specifikus teszt (pl. panel regressziók)
 - Mit tegyünk autokorreláció esetén?
 - Késleltetés-szám (lag) emelése

²³ Az ökonometriához kapcsolódó blogok gyűjteménye: <http://www.ncer.edu.au/resources/econometrics-blogs.php>

- Először: regresszor (X) változónál
 - Utána: a vizsgált (y) változót is késleltetjük
- Ha még így se, akkor keressük a felelős változót és azt kivesszük; szezonálisan kiigazított változót használunk (vagy H-P filterezzük)
- Lag-szám meghatározás: AIC, BIC értékek minimumának keresésével (pl. VAR esetében használják) – de ettől még nem biztos, hogy jó (autokorrelálatlan hibatagú és szignifikáns változókkal teli) lesz!
- Normális eloszlás:
 - Jarque-Bera teszt $p > 0.05$
 - Mit tegyünk a normális eloszlás hiánya esetén?
 - dummy változók: outlierok, exogén sokkok (recesszió, válság), rezsim változások (árfolyam-rendszer).
 - késleltetés növelése
 - skálázás, logaritmizálás átgondolása
 - panel-regressziók esetén: országspecifikus sokkok reprezentálása dummyval
- Hibatagok homoszkedaszticitása
 - ARCH-LM $p > 0.05$
- Főszabályként mind az autokorreláció hiányának, mind a normális eloszlásnak teljesülnie kell, azonban historikus idősorok tesztelése esetén az autokorrelálatlanságra, előrejelzés esetén a normális eloszlás fenntartására kell nagyobb hangsúlyt fektetnünk!
- Szignifikancia: társadalomtudományokban $p < 0.1$ *; $p < 0.05$ **; $p < 0.01$ ***
 - dummy változókkal érdemes reprezentálni az outlierokat
- Determináltság (R^2): makro-modelleknél túl sok változó hat a változókra, így ez szükségszerűen alacsony nálunk, emiatt nem is szokták publikálni komoly (pl. US FED) helyeken sem.
- Dummy változó: rezsimváltás, recesszió, sokk, exogén hatások
 - Akkor is be kell őket vonni a regresszióba, ha egyébként nem szignifikánsak!

Két lépéses regresszió (TSLS), endogenitás

- Endogenitás problémája: $y = \text{konst.} + a x_1 + b x_2 + \text{hiba}$
 - Endogenitás lép fel, ha x_1 és y egyaránt okozhatja a másikat (nem állapítható meg egyértelmű oksági hierarchia a változók között); hasonlóan problémás, ha x_1 és x_2 között lépne fel. $\rightarrow x_1$ és a *hibatag* korrelált lesz
 - Javítási mechanizmus: két lépéses regressziót alkalmazunk (TSLS, two-step), instrumentális változó (IV): $y = \text{konst.} + a x_1 + b x_2 + c Z + \text{hiba2}$
 - IV korrelált a x_1 -el és korrelálatlan a *hibataggal*
- Képlet:

$$\begin{aligned} \circ dY_1 &= \text{const.} + \beta_1 dY_2 + \beta_{2:N} dX_{2:N} + \beta_{N+1} dI_1 + \varepsilon_1 \\ \circ dY_2 &= \text{const.} + \beta_1 dY_1 + \beta_{2:N} dX_{2:N} + \beta_{N+1} dI_2 + u_1 \end{aligned}$$

Matlab-kód: TSLS

(JPL toolbox: <https://www.spatial-econometrics.com/>):

```
clear
data=xlsread('szakdog_a_tsls.xlsx')
cd 'C:\Users\kiss.gabor\Documents\MATLAB\MATLAB\JPL_toolbox'
nobs=length(data);
```

```
y=data(1:end-1,5); %Y1
y1=data(1:end-1,10); %Y2
x1=[data(2:end,5) data(1:end-1,11:13)]; %all X
i1=data(1:end-1,3); %I1
i2=data(1:end-1,4); %I2
iota=ones(nobs-1,1); %constant
xexog=[iota x1];
xall=[iota x1 i1 i2]; %const. X I1 I2
results=tsls(y,y1,xexog,xall)
prt(results)
```

Források

- Gábor, Tamás ; Kiss, Gábor Dávid ; Kovács, Péter (2012): A monetáris sterilizáció hatékonysága és költségei Kínában. *KÖZGAZDASÁGI SZEMLE* 59 : 2 pp. 164-188. , 25 p.
- LaSage J. P. (1999): *Applied Econometrics using MATLAB*. University of Toledo <https://www.spatial-econometrics.com/html/mbook.pdf> pp. 290.

Önellenőrző kérdések

1. Miért fontos, hogy stacioner legyen a regresszió bemenete?
2. Milyen elvárásaink vannak egy regresszió hibatagjának autokorrelációját és eloszlását illetően?
3. Mi a teendő, ha a regressziós modellünk hibatagja autokorrelált?
4. Mi az endogenitás és miért jelenthet problémát?
5. Miért lehet szükséges dummy-változókkal reprezentálni a válságos időszakokat?

