

Determinisztikus időszerelemzés, dekompozíciós időszormodellek: trend, szezonális és ciklus

Ferenci Tamás
tamas.ferenci@medstat.hu

Utoljára frissítve: 2023. május 12.

Tartalom

- 1 Determinisztikus időszorelemzés
 - Alapgondolat
 - Determinisztikus időszormodellezés regresszióval
 - Trend és szezonaritás

A determinisztikus idősorelemzés

- Az idősor alakulása *elvileg* függvényszerűen felírható bizonyos tényezők alapján
- Csak azért nem tudjuk tökéletesen megtenni, mert nem ismerjük e tényezőket, nem tudjuk milyen függvényformával hatnak, nem tudjuk pontosan mérni stb. ezért fogunk hibázni
- De pont: a hibának *csak* ennyi szerepe van...
- ...beállítja az aktuális időszak értékét, és *kész*

Dekompozíciós idősormodellek

- Minderre a legtipikusabb – és egyben legklasszikusabb – példát a **dekompozíciós idősormodellek** jelentik
- A legismertebb additív modell:

$$Y_t = R_t + C_t + S_t + u_t,$$

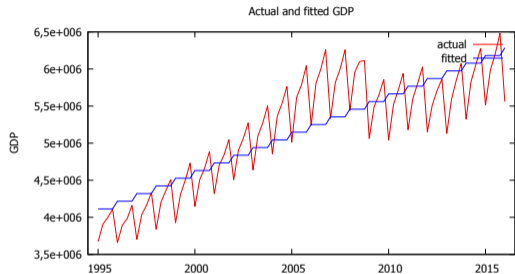
ahol R_t , C_t és S_t a trend, a ciklus és a szezonális t -edik időszakbeli értéke rendre, u_t pedig a már említett eltérésváltozó

- Becslés?

Regresszió alkalmazása

- Az előbbi modell teljesen természetesen becsülhető regresszióval, ha R_t , C_t és S_t helyébe beírjuk a feltételezett – paraméteres – függvényformákat
- (Most tehát mindvégig paraméteres regressziót fogunk használni)
- Legegyszerűbb eset: $R_t = \alpha + \beta t$, $C_t = 0$ és $S_t = 0$ (egyszerű lineáris trend)
- Az így kapott modell OLS-sel becsülhető

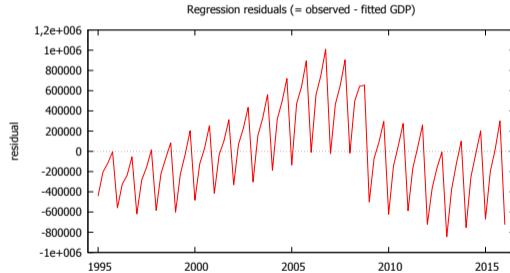
Negyedéves GDP (éves) lineáris trenddel I.



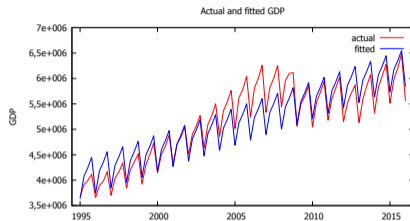
	Coefficient	Std. Error	t-ratio	p-value
const	-2,02165e+008	1,50629e+007	-13,4214	0,0000
EV	103398,	7512,17	13,7641	0,0000
Mean dependent var	5161052	S.D. dependent var	765270,3	
Sum squared resid	1,50e+13	S.E. of regression	424924,3	
R^2	0,695356	Adjusted R^2	0,691686	
$F(1, 83)$	189,4494	P-value(F)	3,94e-23	
Log-likelihood	-1221,169	Akaike criterion	2446,339	
Schwarz criterion	2451,224	Hannan-Quinn	2448,304	
$\hat{\rho}$	0,315141	Durbin-Watson	1,343686	

Negyedéves GDP (éves) lineáris trenddel II.

Mi ezzel a baj? Hibatag jól specifikált? Aligha!



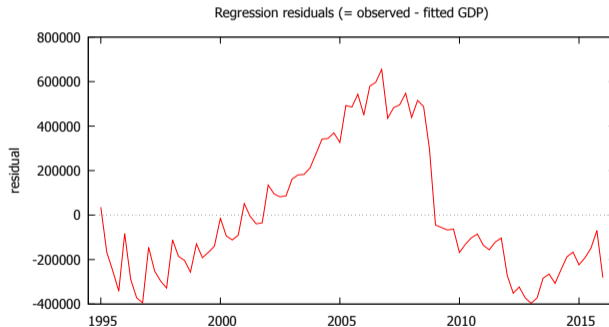
Negyedéves GDP (éves) lineáris trenddel és szezonalitással I.



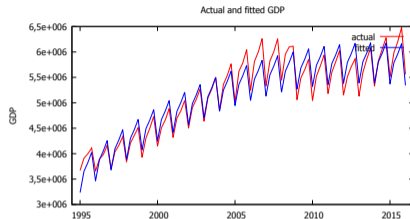
	Coefficient	Std. Error	t-ratio	p-value
const	$-2,04994e+008$	$1,06300e+007$	-19,2845	0,0000
EV	104985,	5301,64	19,8024	0,0000
DNEGYEDEV_1	-815807,	91469,3	-8,9189	0,0000
DNEGYEDEV_2	-375072,	92487,9	-4,0554	0,0001
DNEGYEDEV_3	-203380,	92487,9	-2,1990	0,0308
Mean dependent var	5161052	S.D. dependent var	765270,3	
Sum squared resid	$7,19e+12$	S.E. of regression	299695,1	
R^2	0,853937	Adjusted R^2	0,846634	
$F(4, 80)$	116,9271	P-value(F)	$1,34e-32$	
Log-likelihood	-1189,928	Akaike criterion	2389,855	
Schwarz criterion	2402,068	Hannan-Quinn	2394,768	
$\hat{\rho}$	0,946516	Durbin-Watson	0,116617	

Negyedéves GDP (éves) lineáris trenddel és szezonaritással II.

A szezonaritás jónak tűnik, de az alaptrendet még mindig nem sikerült megragadni:



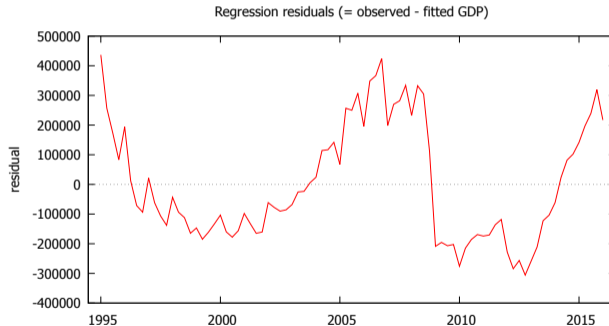
Negyedéves GDP (éves) kvadratikus trenddel és szezonalitással I.



	Coefficient	Std. Error	t-ratio	p-value
const	-2,60273e+010	2,61365e+009	-9,9582	0,0000
EV	2,58613e+007	2,60697e+006	9,9201	0,0000
DNEGYEDEV_1	-792077,	61608,7	-12,8566	0,0000
DNEGYEDEV_2	-375072,	62247,4	-6,0255	0,0000
DNEGYEDEV_3	-203380,	62247,4	-3,2673	0,0016
sq_EV	-6422,59	650,072	-9,8798	0,0000
Mean dependent var	5161052	S.D. dependent var	765270,3	
Sum squared resid	3,21e+12	S.E. of regression	201704,8	
R^2	0,934664	Adjusted R^2	0,930529	
$F(5, 79)$	226,0280	P-value(F)	2,80e-45	
Log-likelihood	-1155,736	Akaike criterion	2323,473	
Schwarz criterion	2338,128	Hannan-Quinn	2329,368	
$\hat{\rho}$	0,889365	Durbin-Watson	0,173391	

Negyedéves GDP (éves) kvadratikus trenddel és szezonalitással II.

Reziduumok kicsit jobbak:



Mindezeken a limitációk

- *Egyrészt* el kell találni a függvényformát
- Persze modelldiagnosztika (az előbb látott grafikus módszerek és tesztek is) ott van
- (Ez igazából már keresztmetszetnél is így volt)
- Pl. a kvadratikus nyilván csak erre az időszakra jó, az általánosítóképessége botrányos lenne
- *Másrészt* a hibetag diagnosztikája bonyolultabbá válik, egy új szempont is megjelenik (autokorreláció) → később még nagyon sokat fogunk róla beszélni

A trend megadása

- **Trend:** „hosszú távú alapirányzat”
- A mostani trend (determinisztikus trend) bármi lehet, amit paraméteres függvényformában megadunk; például:
 - Lineáris trend: $a + bt$
 - Kvadratikus trend: $a + b_1t + b_2t^2$
 - Polinomiális trend: $a + b_1t + b_2t^2 + \dots + b_kt^k$
 - Exponenciális trend: ae^{bt}
 - Aszimptotikus trend: $c + \frac{1}{a+bt}$
 - Logisztikus trend: $\frac{1}{c+e^{a+bt}}$
 - stb. stb.
- (Persze amelyek nem lineáris, ott vagy linearizálni kell vagy – ha ez nem lehetséges – akkor nem OLS-sel becsülni)
- Ezek mind paraméteres trendek voltak, elképzelhető nem-paraméteres trend is, a legismertebb a spline-ok használata (de ne feledjük, annak a becslése kevésbé hatásos, nem kapunk egyetlen vagy néhány számba sűrített – és jó esetben tárgyterületileg értelmezhető – eredményt, valamint az előrejelzés is problémásabb)

Szezonális megadása

- **Szezonális:** „éven belüli mintázat”, exogén módon rögzített hosszúságú, periodikus (vs. **ciklus:** „éven túli”, nem feltétlenül exogén módon adott, ismert hosszúságú)
- A szezonálisnál viszont tipikusabb a nem-paraméteres megadás: minden negyedévnek (hónapnak, félévnek stb.) saját paramétere van
- (Dummy-kkal, ld. később, regressziós keretbe szintén szépen illeszkednek!)
- Persze itt is elképzelhető paraméteres megadás, a legismertebb a trigonometrikus (harmonikus) függvények használata

Dummy-kódolás szezonaritáshoz: referenciakódolás

- Az egyik szezon indikátorát elhagyjuk: **referenciakódolás**

	D_{Q1}	D_{Q2}	D_{Q3}
Q1	1	0	0
Q2	0	1	0
Q3	0	0	1
Q4	0	0	0

- Értelmezés: eltérés a referenciacsoporthoz képest (ami az elhagyott indikátorú csoport)

Dummy-kódolás szezonaritáshoz: kontrasztkódolás I.

- Egy másik népszerű megoldás a **kontrasztkódolás**: viszonyítsunk az *átlaghoz*!
- Ehhez hogyan kell kódolni...?

	C_{Q1}	C_{Q2}	C_{Q3}
Q1	1	0	0
Q2	0	1	0
Q3	0	0	1
Q4	-1	-1	-1

Dummy-kódolás szezonálitáshoz: kontrasztkódolás II.

Mert:

$$\alpha + \beta_{C_{Q1}} + 0 + 0 = \bar{y}_{Q1} \quad (1)$$

$$\alpha + 0 + \beta_{C_{Q2}} + 0 = \bar{y}_{Q2} \quad (2)$$

$$\alpha + 0 + 0 + \beta_{C_{Q3}} = \bar{y}_{Q3} \quad (3)$$

$$\alpha - \beta_{C_{Q1}} - \beta_{C_{Q2}} - \beta_{C_{Q3}} = \bar{y}_{Q4} \quad (4)$$

És így:

- $(1)+(2)+(3)+(4) \Rightarrow 4\alpha = \bar{y}_{Q1} + \bar{y}_{Q2} + \bar{y}_{Q3} + \bar{y}_{Q4} \Rightarrow \alpha$ tényleg a főátlag (mert azonosak voltak a csoportok elemszámai, különben ún. súlyozott kontraszt kellene)
- $(2)+(3)+(4) \Rightarrow 3\alpha - \beta_{C_{Q1}} = \bar{y}_{Q2} + \bar{y}_{Q3} + \bar{y}_{Q4} \Rightarrow \beta_{C_{Q1}} = 3\alpha - (\bar{y}_{Q2} + \bar{y}_{Q3} + \bar{y}_{Q4}) = 3\alpha - (4\alpha - \bar{y}_{Q1}) \Rightarrow \beta_{C_{Q1}} = \bar{y}_{Q1} - \alpha \Rightarrow$ tényleg az átlagtól való eltérés (és hasonlóan a másik kettő)

Dummy-kódolás szezonálitáshoz: egyebek

- Az angol irodalomban az általunk kontrasztkódolásnak nevezett módszert nagyon gyakran „effect coding”-nak nevezik...
- ... a kontraszt pedig az, amikor a csoportok tetszőleges – általunk meghatározott – lineáris kombinációját teszteljük