

Idősorok szűrése

Ferenci Tamás

tamas.ferenci@medstat.hu

Utoljára frissítve: 2023. május 12.

Tartalom

Tartalomjegyzék

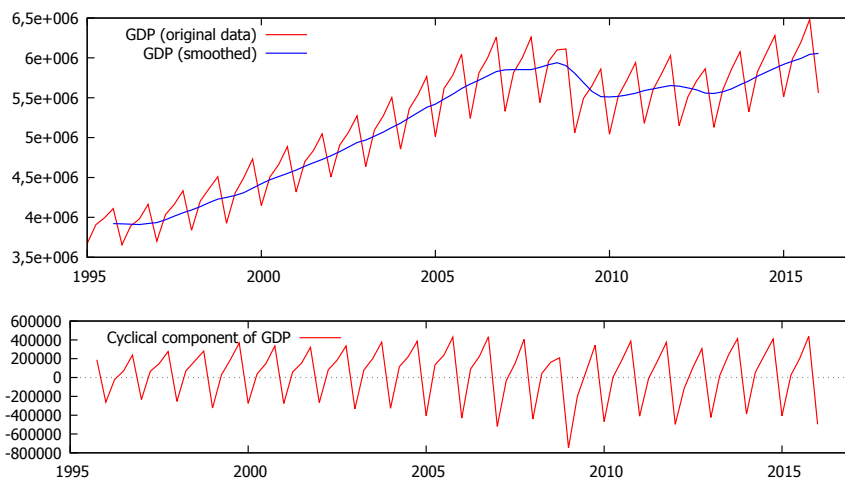
1	Idősorok szűrése	1
---	------------------	---

1. Idősorok szűrése

Célunk

- Szeretnénk elkülöníteni a trendet és a ciklikus komponenseket (szezón + ciklus); ez számos közgazdasági kérdésnél fontos feladat
- Egy lehetséges megoldás: „paraméteres szűrés”, azaz előírjuk a függvényformát és regresszióval megbecsüljük
- Igazából ezt megtettük az előbb is: megadtuk a trendet (lineáris vagy kvadratikus), megadtuk a szezonálisitást (ezt nem-paraméteresen), és a látott reziduum a kiszűrt ciklus (+zaj) volt, ha van ilyen
- De ez függ a függvényforma-választás helyességén; nem lehetne ilyen feltevések nélkül is megtenni?
- Hogyne, például átlagoljunk ki évenként!
- Nem a legjobb, abrupt ugrások év végén, átlagoljunk inkább folyamatosan, mintegy csúszóablakot végigtolva (így is mindig négy különböző szezon lesz benne!)

Motiváló példa



Mozgóátlagolás

- Ez volt az (egyszerű) **mozgóátlag**:

$$y'_t = \frac{y_t + y_{t-1} + y_{t-2} + \dots + y_{t-p}}{p + 1}$$

- (Néha nem visszafelé átlagolnak, hanem az eredmény az átlagolt ablak közepén van („centrált” mozgóátlag), a dolognak nincs nagy jelentősége: ez talán kicsit jobban néz ki, viszont jövőbeli megfigyeléseket is igényel)
- Ez azt jelenti, hogy minden megfigyelés ugyanolyan súlyú; kézenfekvő gondolat, hogy a régebbiek kevésbé számítsanak
- Például:

$$y'_t = \frac{py_t + (p - 1)y_{t-1} + (p - 2)y_{t-2} + \dots + y_{t-(p-1)}}{p + (p - 1) + (p - 2) + \dots + 1}$$

- Ez az ún. súlyozott mozgóátlag

Exponenciális mozgóátlag

- Ökonometriában gyakoribb az exponenciális súlyozás:

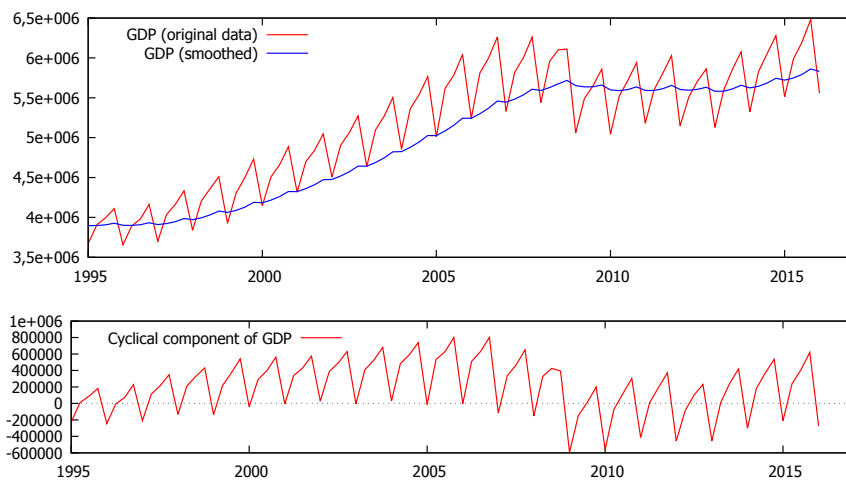
$$y'_t = \alpha y_t + (1 - \alpha) y'_{t-1},$$

hiszen ez nyilván annak felel meg, hogy

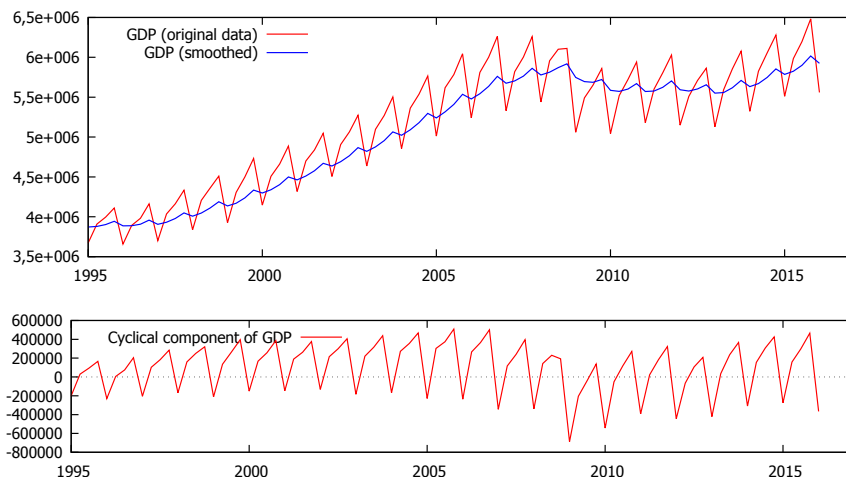
$$y'_t = \alpha y_t + (1 - \alpha) \alpha y_{t-1} + (1 - \alpha)^2 \alpha y_{t-2} + \dots + (1 - \alpha)^{t-2} \alpha y_2 + (1 - \alpha)^{t-1} \alpha y_1 + (1 - \alpha)^t \alpha y_0$$

- (Az y_0 kezdőértéket nekünk kell megadni, a tipikus választások: y_1 , az első néhány megfigyelés átlaga, az egész idősor átlaga; a következő példákban az első 4 megfigyelés átlaga került alkalmazásra)
- Tehát az ablak egyre nyúlik (végig az egész tartomány), és a súlyok exponenciálisan csengenek le

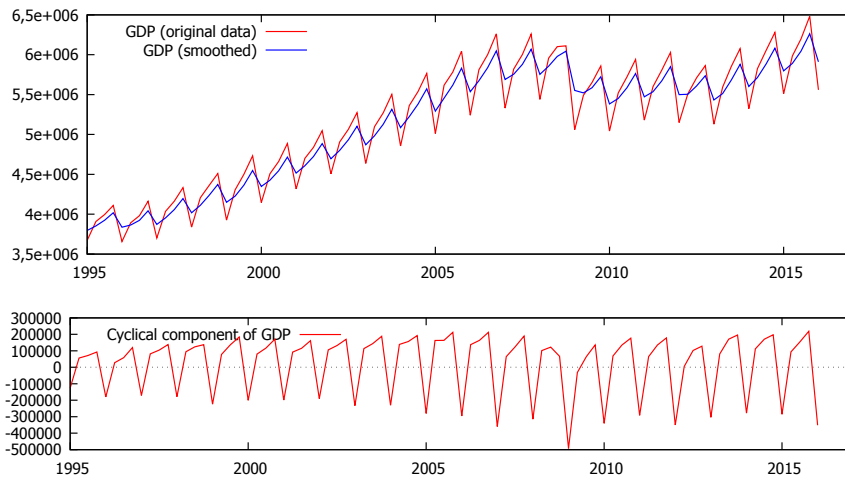
Negyedéves GDP exponenciális mozgóátlagolása, $\alpha = 0,1$



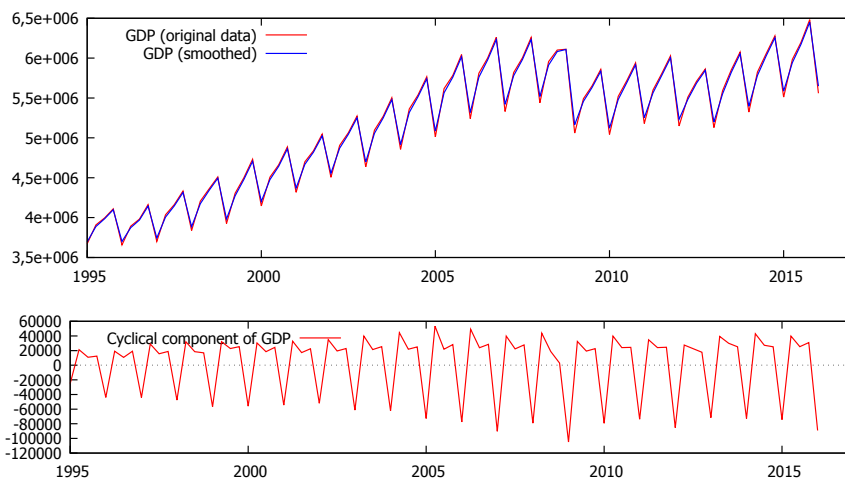
Negyedéves GDP exponenciális mozgóátlagolása, $\alpha = 0,2$



Negyedéves GDP exponenciális mozgóátlagolása, $\alpha = 0,5$



Negyedéves GDP exponenciális mozgóátlolása, $\alpha = 0,9$



Lineáris szűrő

- Az egyszerű és a súlyozott mozgóátlag speciális esete annak, hogy

$$y'_t = a_0 y_t + a_1 y_{t-1} + \dots + a_p y_{t-p} = \sum_{i=0}^p a_i y_{t-i}$$

- Például egyszerű mozgóátlagra $a_i = 1/(p+1)$
- Lényegében diszkrét konvolúció
- Ezt hívjuk lineáris szűrőnek, a tulajdonságait értelemszerűen teljes mértékben meghatározzák az a_0, a_1, \dots, a_p szűrőegyütthatók

- Roppant érdekes kérdés, hogy a szűrt idősor hogyan néz ki az eredetihez képest a szűrőegyütthatók függvényében, lehet otthon kísérletezgetni...

Hodrick–Prescott-szűrő

- Különösen makroökonómiában népszerű
- Alapgondolat: a trend (g_t) követi az idősort (azaz az $y_t - g_t$ kicsi), de nem nagyon ugrándoza (azaz g_t sima)
- A megoldandó feladat:

$$\min_{\{g_t\}_{t=1}^T} \sum_{t=1}^T (y_t - g_t)^2 + \lambda \sum_{t=2}^{T-1} [(g_{t+1} - g_t) - (g_t - g_{t-1})]^2$$

- Első tag: mennyire követi jól az idősort a trend, második tag: mennyire „rángatózik” a trend
- A λ együttható határozza meg a két szempont egymáshoz viszonyított fontosságát ($\lambda = 0$: a trend nem kell, hogy sima legyen \rightarrow pontosan az idősor lesz; $\lambda \rightarrow \infty$: trend legyen tökéletesen sima, nem érdekes, hogy mennyire követi az idősort \rightarrow egyenes lesz)
- Negyedéves adatokra a tipikus választás $\lambda = 1600$

A második tagra azt írtuk, hogy „mennyire rángatózik a trend”; később majd pontosabban is fogjuk látni, hogy ez micsoda, mert van egy nagyon konkrét tartalma.

A HP-szűrő matematikája I.

- Az érdekes az, hogy a fenti minimalizációs feladatnak van zárt alakú megoldása!
- Legyen \mathbf{y} az y_t -k, \mathbf{g} a g_t -k vektora és legyen

$$\mathbf{Q}_{(T-2) \times T} = \begin{pmatrix} 1 & -2 & 1 & \dots & 0 & 0 & 0 \\ 0 & 1 & -2 & \dots & 0 & 0 & 0 \\ 0 & 0 & 1 & \dots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & -2 & 1 & 0 \\ 0 & 0 & 0 & \dots & 1 & -2 & 1 \end{pmatrix},$$

akkor a megoldandó feladat

$$\min_{\mathbf{g}} (\mathbf{y} - \mathbf{g})^T (\mathbf{y} - \mathbf{g}) + \lambda (\mathbf{Q}\mathbf{g})^T (\mathbf{Q}\mathbf{g})$$

A HP-szűrő matematikája II.

- Deriválva \mathbf{g} szerint:

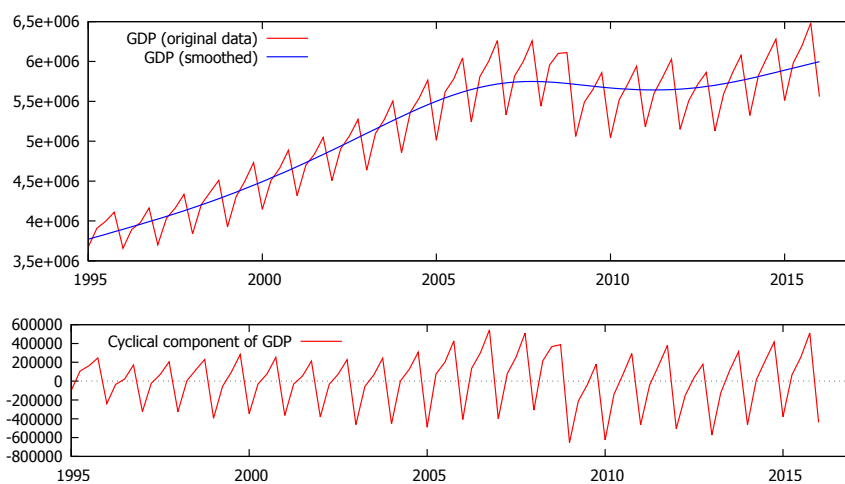
$$-2\mathbf{y} + 2\mathbf{g} + 2\lambda\mathbf{Q}^T\mathbf{Q}\mathbf{g}$$

- Egyenlővé téve nullával és megoldva (a másodrendű feltételek teljesülnek ahhoz, hogy ez tényleg minimum legyen):

$$-2\mathbf{y} + 2\mathbf{g} + 2\lambda\mathbf{Q}^T\mathbf{Q}\mathbf{g} = 0 \Rightarrow \hat{\mathbf{g}} = (\mathbf{I} + \lambda\mathbf{Q}^T\mathbf{Q})^{-1}\mathbf{y}$$

- (A megoldás mindegyik megfigyeléstől függ)

Negyedéves GDP HP-szűrése, $\lambda = 1600$



A HP-szűrő kritikája

Érdekes olvasmány: Hamilton, James D. „Why You Should Never Use the Hodrick-Prescott Filter.” (2016). <http://econweb.ucsd.edu/~jhamilto/hp.pdf>.

Egyéb szűrők

- Több szűrő viselkedése alapvetően frekvencia-tartományon érthető meg
- Trend/ciklus szétválasztás: aluláteresztő szűrés (mert a trend az, ami lassan változik, a ciklus az, ami gyorsabban, persze kérdés, hogy hol a határ)
- Igazából már az egyszerű mozgóátlag is egy aluláteresztő szűrő volt!
- Vagy: LOESS-használata (STL-dekompozíció)