Idősorok elemzése 2.

Identifikáció, Paraméterbecslés, Spektrálelemzés.

2018. november 5. és 12.

Ismétlés, összefoglalás

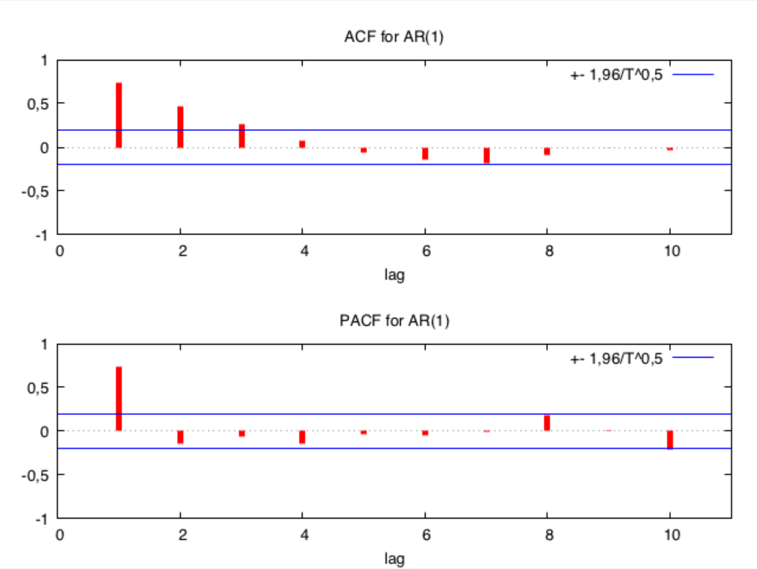
Box-Jenkins modellezés:

- **Stacionarizálás I.: trend szűrése***
- **Stacionarizálás II.: szezonalitás szűrése***
- Az ARMA modell p és q rendjének becslése az ACF és PACF segítségével - IDENTIFIKÁCIÓ
- Paraméterbecslés
- Modelldiagnosztika

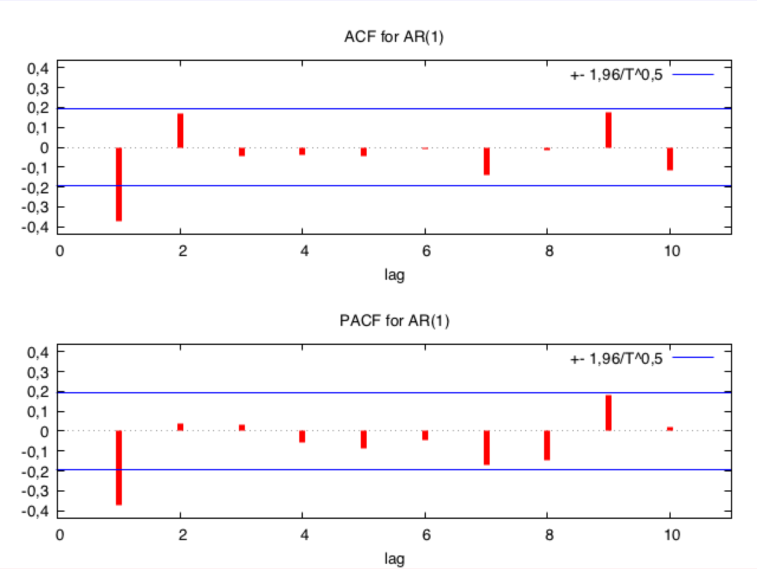
* A rejtett periódusok vizsgálata és a stacionaritást ellenőrző ADF tesztek még hiányoznak, ezeket később tárgyaljuk.

IDENTIFIKÁCIÓ

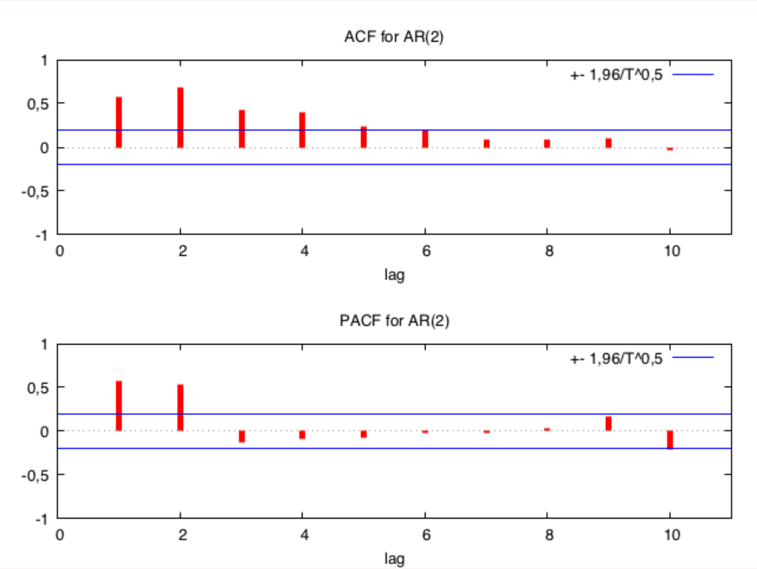
ACF, PACF: AR(1) $\alpha_1 = 0,7$



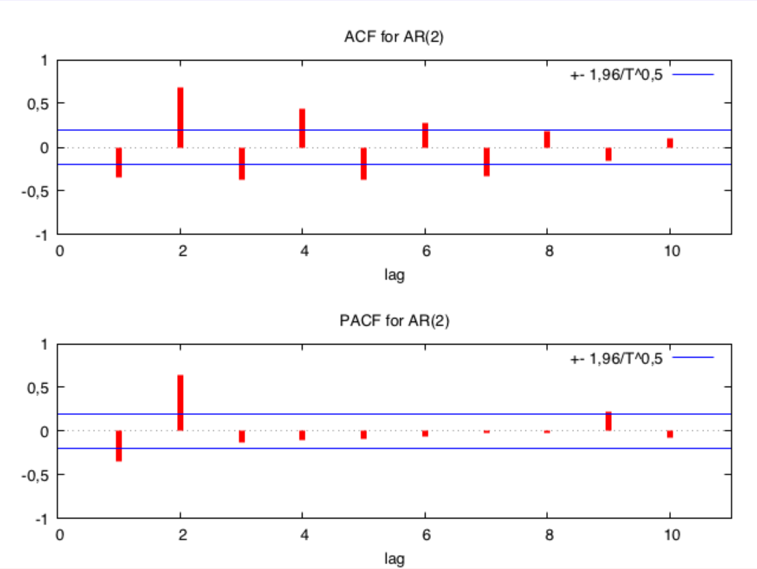
ACF, PACF: AR(1) $\alpha_1 = -0,5$



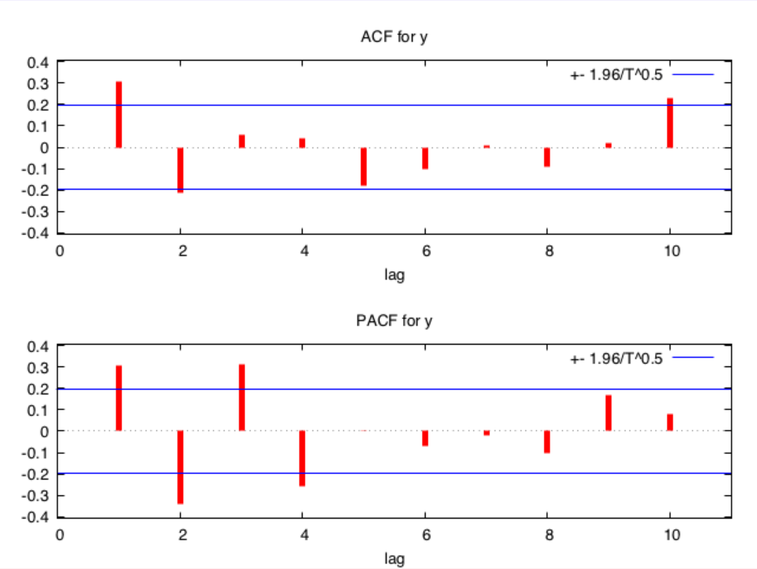
ACF, PACF: AR(2) $\alpha_1 = 0,2,\ \alpha_2 = 0,4$



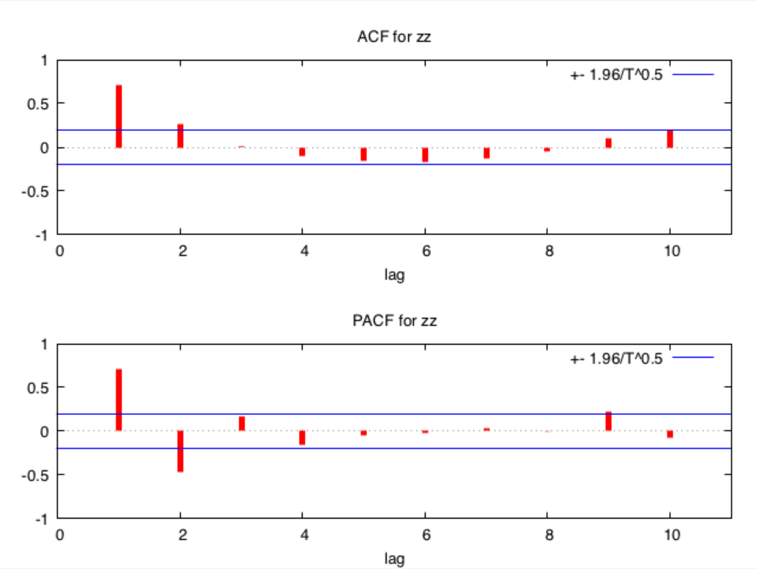
ACF, PACF: AR(2) $\alpha_1 = -0,2$, $\alpha_2 = 0,4$



ACF, PACF: MA(2)


ACF for y
+- 1.96/T^0.5
0.4
0.3
0.2
0.1
0
-0.1
-0.2
-0.3
-0.4
0
2
4
6
8
10
lag
PACF for y
+- 1.96/T^0.5
0.4
0.3
0.2
0.1
0
-0.1
-0.2
-0.3
-0.4
0
2
4
6
8
10
lag

ACF, PACF: ARMA(2, 2)



Az ARMA modell beazonosítása

Tehát az ACF és PACF függvények grafikonjai alapján valószínűsíteni lehet az alkalmas ARMA modell rendjét.

Modell	ACF	PACF
AR(p)	lecseng	eltűnik a p. tag után
MA(q)	eltűnik a q. tag után	lecseng
ARMA(p, q)	eltűnik a q. tag után	eltűnik a p. tag után
fehérzaj, véletlen foly.	nincs szig. érték	nincs szig. érték

- Tehát ha az ACF nulla közelében marad egy adott késleltetés után, akkor MA rendjének ez lesz a megfelelő választás.
- Ha a PACF nulla közelében marad egy adott késleltetés után, akkor AR rendjének ez lesz a megfelelő választás.
- Ha a fenti esetek egyike sem áll fenn, de egy idő után mind az ACF, mind a PACF nullához közeli értékeket vesz fel, akkor ARMA modellt választunk.

PARAMÉTERBECSLÉS

AR folyamatok paraméterbecslése

Legyen

$$y_t = \alpha_1^* y_{t-1} + \ldots + \alpha_p^* y_{t-p} + \varepsilon_t, \quad t = 1, \ldots, T, \tag{1}$$

ahol $\alpha_1^*, \ldots, \alpha_p^* \in \mathbb{R}$ a folyamat igazi paraméterei, (ε_t) pedig fehérzaj folyamat. Jelölje $\theta^* = (\alpha_1^*, \ldots, \alpha_p^*)^T$ az igazi paraméterek vektorát, $x_t = (y_{t-1}, \ldots, y_{t-p})^T$ pedig a megfigyelések vektorát, és legyenek

$$Y = [y_{p+1}, \ldots, y_T]^T, \quad X = [x_{p+1}^T, \ldots, x_T^T]^T, \quad \varepsilon = [\varepsilon_{p+1}, \ldots, \varepsilon_T]^T.$$

Ekkor

$$Y = X \cdot \theta^* + \varepsilon.$$

Jelölje

$$D = \{\theta = (\alpha_1, \ldots, \alpha_p)^T : A(s)\,\text{polinom stabil}\}$$

AR folyamatok paraméterbecslése

1. módszer: LS-becslés. A lineáris modellforma miatt alkalmazható. Ezek alapján adódik, hogy

$$\hat{\theta}_T = (X^T X)^{-1} X^T Y,$$

feltéve, hogy a fenti inverz létezik. A becslés konzisztens lesz.

Rekurzív változat: legyen $U(t) = \frac{1}{t} X(t)^T X(t)$, ahol $X(t) = X$. Ekkor

$$U(t) = U(t-1) + \frac{1}{t}\left(x_t x_t^T - U(t-1)\right),$$

és

$$\hat{\theta}_T = \hat{\theta}_{T-1} + \frac{1}{t} U^{-1}(t) x_t \left(y_t - x_t^T \hat{\theta}_{T-1}\right)$$

Gond: U^{-1} felújítása minden lépésben!

Tétel 1 (Mátrix inverziós lemma).

$$(A + uu^T)^{-1} = A^{-1} - \frac{A^{-1} uu^T A^{-1}}{1 + u^T A^{-1} u}$$

AR folyamatok paraméterbecslése

2. módszer: ML-becslés, ahol most az $(y_1, \ldots, y_T)$ megfigyelések együttes sűrűségfüggvényét kell maximalizálnunk.

Sajnos a módszer a gyakorlatban nem működik, mert nehézkesebb, mint az LS-módszer, ugyanis egy nem-lineáris függvény szélsőértékeit kell megkeresni, ami nem mindig oldható meg könnyen és egyértelműen.

MA folyamatok esetében ugyanez a helyzet, így ezekre a folyamatokra nem használjuk az ML-becslést, helyette minden esetben az LS-becslést alkalmazzuk!

ARMA folyamatok paraméterbecslése

Legyen

$$A(z)^* X_t = B(z)^* \varepsilon_t,$$

ahol

$$A(z)^* = 1 - \alpha_1^* z - \ldots - \alpha_p^* z^p$$
$$B(z)^* = \beta_0^* + \beta_1^* z + \ldots + \beta_q^* z^q,$$

(ε_t) pedig fehérzaj folyamat. Tegyük fel, hogy A^* és B^* relatív prímek. Legyen $\theta^* = (\alpha_1^*, \ldots, \beta_q^*)$, és tegyük fel, hogy

$$D = \{\theta = (\alpha_1, \ldots, \beta_q)^T : A \text{ és } B \text{ stabilak}\}$$

Legyenek adottak az $(X_1, \ldots, X_T)$ megfigyelések, és tegyük fel, hogy $(\varepsilon_t) \sim N(0,1)$. Ekkor $(X_1, \ldots, X_T) \sim \mathcal{N}(\underline{0}, \Sigma(\theta^*))$.

ARMA folyamatok paraméterbecslése

Az LS becslés formálisan alkalmazható lenne, de az MA rész nem fehérzaj volta miatt a becslés aszimptotikusan torzított lesz.

Így ehelyett ML-becslést alkalmazunk:

$$-\log f(X_1, \ldots, X_T; \theta) = \frac{1}{T} \log 2\pi + \frac{1}{2} \log \det \Sigma(\theta) + \frac{1}{2} \underline{X}^T \Sigma^{-1}(\theta) \underline{X}^T$$

formában adott. Ennek becslése viszont reménytelen a fellépő nemlinearitás miatt. Helyette invertáljuk a folyamatot az

$$e_t(\theta) = \frac{A(\theta)}{B(\theta)} X_t$$

egyenlet segítségével. Ekkor megmutatható, hogy az előbb definiált $-\log f(\underline{X}; \theta)$ függvény konstanstól eltekintve megegyezik a

$$V_T(\theta) = \frac{1}{2} \sum_{t=1}^{T} e_t^2(\theta)$$

kifejezéssel. Tehát a feladat ennek minimalizálása a D halmazon.

SPEKTRÁL-ELEMZÉS

Példa - BKV villamosenergia-terhelési görbéje

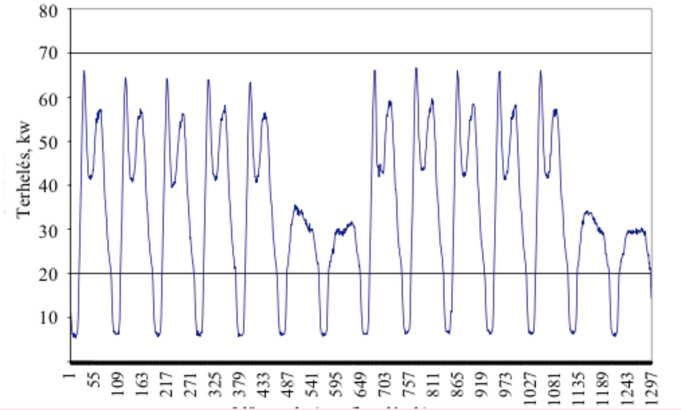


Figure: BKV villamosenergia-terhelési görbéje, negyedóránkénti mérések (2 hét adatai, ezer kw-ban mérve)

Az idősor alakja sugallja a napi periodicitás jelenlétét.

Példa - Fogyasztói árindexek

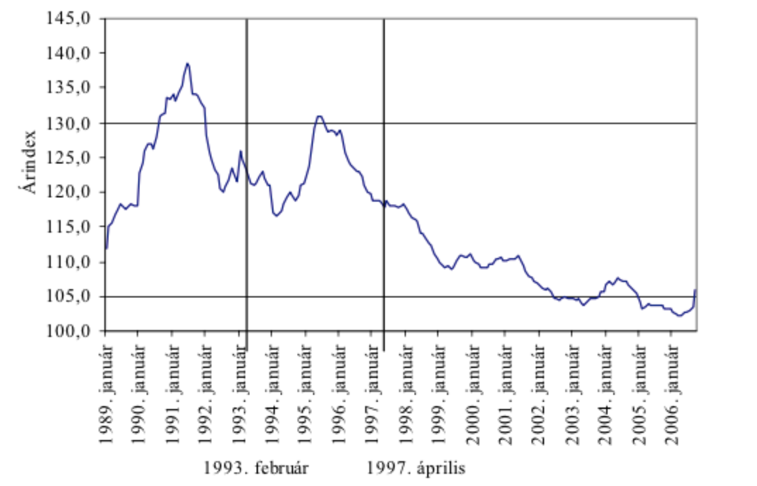


Figure: Fogyasztói árindexek

Spektrál-elemzés

Definíció 1.

Minden $\lambda \in [-\pi, \pi]$ *értékre legyen az* $I_N(\lambda) = (a(\lambda))^2 + (b(\lambda))^2$ *függvény az ún. periodogram függvény, ahol*
$a(\lambda) = \sqrt{\frac{2}{N}} \sum_{t=1}^{N} x_t \cos(\lambda t)$ *és* $b(\lambda) = \sqrt{\frac{2}{N}} \sum_{t=1}^{N} x_t \sin(\lambda t)$. *Azaz*

$$I_N(\lambda) = \frac{2}{N} \left| \sum_{t=1}^{N} x_t e^{-it\lambda} \right|.$$

Tétel 2.

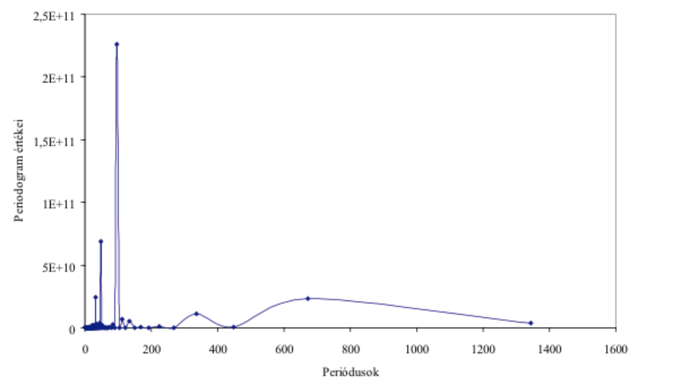
Tegyük fel, hogy a mintabeli megfigyelések száma végtelen tart. Ekkor a periodogram határértékére

$$\lim_{N\to\infty} I_N(\lambda) = \begin{cases} 0, & \lambda \neq \lambda_k \ k = 1, \ldots, m \\ +\infty, & \lambda = \lambda_k \ \textit{valamilyen } k\textit{-ra} \end{cases}$$

adódik. A fenti konvergencia sztochasztikus értelemben értendő.

Numerikus számításoknál a periodogramot a $\lambda = 0, \frac{2\pi}{N}, \frac{4\pi}{N}, \ldots$ ún. Fourier-frekvenciákon értékelik ki. Nagy mintaelemszám esetén a periodogram kiugró értékei nagy valószínűséggel rejtett frekvenciákat jelentenek. Ezek tesztelésére számos próba áll rendelkezésre (pl. Fisher-féle g-próba, Whittle-féle módosított g-próba).

Példa - BKV adatok spektruma



A numerikus adatok alapján megállapítható, hogy a becsült sűrűségfüggvény a 96-os periódusnál (periódus$= 2\pi/\lambda_k$) veszi fel a maximumát, ami a negyedórás megfigyelések miatt egy napos ciklusnak felel meg ($24 \cdot 4 = 96$).

Példa - BKV villamosenergia-terhelési görbéje

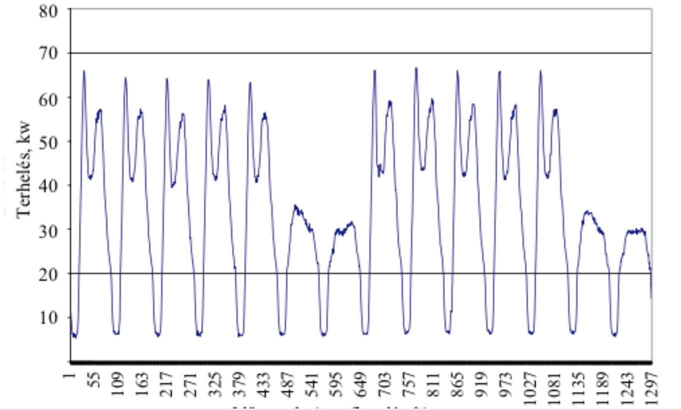
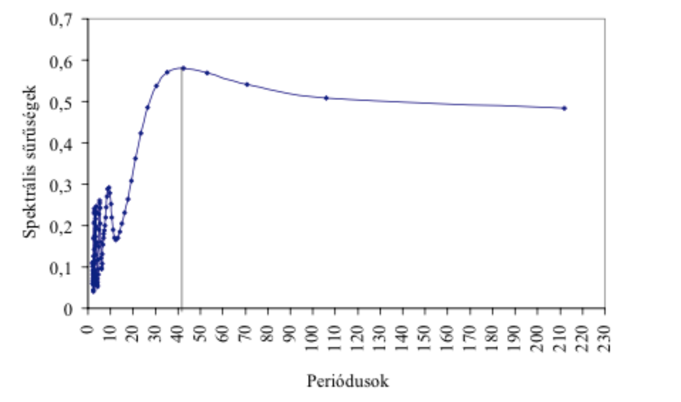


Figure: BKV villamosenergia-terhelési görbéje, negyedóránkénti mérések (2 hét adatai, ezer kw-ban mérve)

Példa - fogyasztói árindexek spektruma



A maximumot a $42,5$ hosszú perióduson veszi fel, ami kb. $3,5$ éves ciklusnak felel meg.

Példa - Fogyasztói árindexek

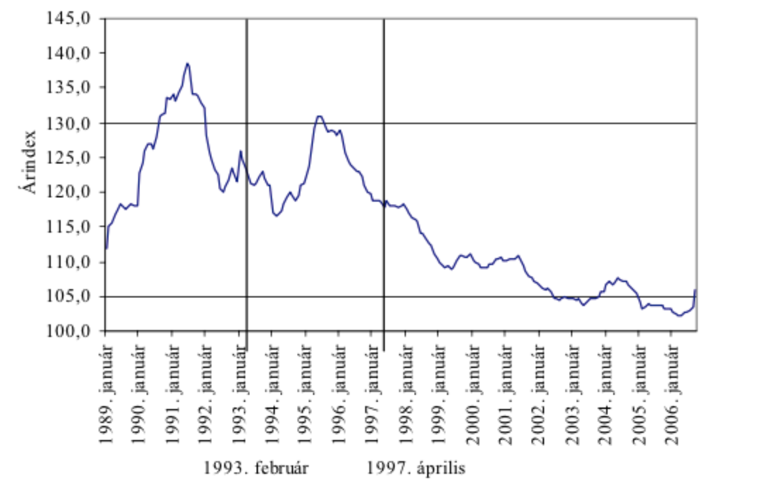


Figure: Fogyasztói árindexek