AUTOREGRESSZIÓ ÉS IDŐSOROELEMZÉS

Dr. Sipos Tibor
Dr. Török Ádám
Szabó Zsombor

MŰEGYETEM 1782

KUKG

BUDAPESTI MŰSZAKI ÉS GAZDASÁGTUDOMÁNYI EGYETEM
Közlekedésmérnöki és Járműmérnöki Kar

Tartalom

Autoregresszió

Idősorelemzés

Tartalom

Autoregresszió

Idősorelemzés

Autokorreláció

- A Gauss-Markov feltételek egyike, hogy u_i, u_j függetlenek $\forall i \neq j$
- Amennyiben ez nem teljesül autokorrelációról beszélhetünk
- Tfh u_i és u_{i-k} $(k < i, k \in \mathbb{N})$ nem függetlenek, ekkor k-ad rendű autokorrelációról beszélhetünk
- Ezt ρ_k-val jelöljük
- Könnyen belátható, hogy n megfigyelés esetén $n - 1$ autokorrelációt definiálhatunk
- Gyakorlatban igyekszünk minél kevesebb bevezetésével leírni a teljes sort

Durbin–Watson-próba

- A Durbin–Watson-próba (DW-próba) az autokorreláció tesztelésére szolgál
- Három alkalmazhatósági feltétel:
  - Csak elsőrendű autokorrelációra
  - A próba bizonyos próbastatisztika értékek esetén nem vezet döntésre
  - Nem alkalmazható, ha a modellben van késleltetett magyarázóváltozó
- Próbastatisztika

$$d = \frac{\sum_{t=2}^{n} (u_t - u_{t-1})^2}{\sum_{t=1}^{n} u_t^2} = \frac{\sum u_t^2 + \sum u_{t-1}^2 - 2 \sum u_t u_{t-1}}{\sum u_t^2}$$

Autokorreláció

- Nagyminta ($n \to \infty$) esetén
  - $\sum u_t^2 \cong \sum u_{t-1}^2$
  - $\bar{u}_t \cong \bar{u}_{t-1} = \bar{u}$
- Gauss–Markov-tétel miatt
  - $u \sim \mathcal{N}(0, \sigma)$
  - $\bar{u} = 0$

$$\rho = r_{u_t, u_{t-1}} = \frac{\sum\left((u_t - \bar{u}_t)(u_{t-1} - \bar{u}_{t-1})\right)}{\sqrt{\sum(u_t - \bar{u}_t)^2 \sum(u_{t-1} - \bar{u}_{t-1})^2}}$$

$$\rho = \frac{\sum u_t u_{t-1}}{\sum u_t^2}$$

DW-próba

- Nagyminta esetén ($\sum u_t^2 \cong \sum u_{t-1}^2$)

$$d = \frac{\sum u_t^2 + \sum u_{t-1}^2 - 2\sum u_t u_{t-1}}{\sum u_t^2}$$

$$d = 2\left(1 - \frac{\sum u_t u_{t-1}}{\sum u_t^2}\right)$$

$$\rho = \frac{\sum u_t u_{t-1}}{\sum u_t^2}$$

$$d \cong 2(1-\rho)$$

- Tehát, ha
  - $\rho = 1 \Rightarrow d = 0$
  - $\rho = -1 \Rightarrow d = 4$
  - $\rho = 0 \Rightarrow d = 2$

DW-próba tesztelése

- $H_0: \rho = 0$
- $H_1: \rho > 0$, tehát pozitív autokorrelációra tesztelünk
- n és a magyarázóváltozók számának függvényében definiáltak d_L és d_U értékeket
  - $d < d_L$: H_0 elutasításra kerül
  - $d > d_U$: H_0 nem kerül elutasításra
  - $d_L < d < d_U$: a próba alapján nem tudunk dönteni
- Negatív autokorreláció tesztelése esetén $4 - d$ alkalmazandó

Autókorreláció kezelése

- Amennyiben az autokorrelálatlanságot el kell utasítani, abban az esetben két lehetőség van
  - Első differenciák alapján történő becslés
  - Autoregressziós modell
- Ökölszabály: ha $d < R^2$, akkor az első differenciáltak alapján érdemes becsülni az egyenletet

Első differenciáltak alapján történő becslés

- Ebben az esetben az $(y_t - y_{t-1})$ változót becsüljük az $(x_t - x_{t-1})$
- Vegyük a következő regressziós modellt

$$y_t = \alpha + \beta x_t + u_t$$
$$y_{t-1} = \alpha + \beta x_{t-1} + u_{t-1}$$

  - Ekkor kivonás után a következő egyenlet adódik

$$(y_t - y_{t-1}) = \beta(x_t - x_{t-1}) + (u_t - u_{t-1})$$

  - Ezt az egyenletet lehet OLS-el becsülni (feltételezve, hogy a $(u_t - u_{t-1})$ hibatagok függetlenek
  - Figyelni kell rá, hogy ezt a modellt konstans tag nélkül kell becsülni

Első differenciáltak alapján történő becslés

- Fontos kérdés, az R^2 értékek kérdésköre
- A kiinduló modell, valamint az első differenciált modell esetében a függő változó nem ugyanaz, tehát az R^2 értékek direktben nem összehasonlíthatók
- ! R_D^2 egy mutató, ami összehasonlítható az eredeti R^2 értékkel

$$\frac{1-R_D^2}{1-R_1^2} = d\,\frac{RSS_0}{RSS_1}\frac{n-k-1}{n-k}$$

  - RSS_0: reziduális négyzetösszeg az eredeti egyenletre
  - RSS_1: reziduális négyzetösszeg az első differenciáltra
  - R_1^2: az első differenciált egyenlet R^2 értéke

Elsőrendű autokorreláció

- Ebben az esetben a regressziós függvényt $(y_t - \rho y_{t-1})$ és $(x_t - \rho x_{t-1})$ között írjuk fel
- Ekkor a hibatagok elsőrendűen autoregresszívek

$$u_t = \rho u_{t-1} + e_t$$

- Ezen típusú modelleket AR(1) modelleknek is nevezik, ahol az AR az autoregresszivitásra utal, míg az 1-es az elsőrendűségre

Paraméterbecslés AR(1) esetén

- Vegyük a következő AR(1) modellt

$$y_t = \alpha + \beta x_t + u_t \ \forall t = 1..T$$
$$u_t = \rho u_{t-1} + e_t$$

- Az első egyenlet $t-1$-re felírva és ρ-val megszorozva

$$\rho y_{t-1} = \rho\alpha + \rho\beta x_{t-1} + \rho u_{t-1}$$
$$y_t - \rho y_{t-1} = \alpha(1-\rho) + \beta(x_t - \rho x_{t-1}) + e_t$$

- Mivel e_t nem autokorrelált, állandó varianciájú, ezért az OLS alkalmazható
- Ekkor $\forall t = 2..T$
  - $y_t^* = y_t - \rho y_{t-1}$
  - $x_t^* = x_t - \rho x_{t-1}$
- $t = 1$ esetén
  - $y_1^* = \sqrt{1-\rho}\, y_1$
  - $x_1^* = \sqrt{1-\rho}\, x_1$
- A gyakorlatban azonban ρ értéke ismeretlen, becsülni kell
  - Iteratív eljárások
  - Rácspontos keresési eljárások

Iteratív eljárások

- Cochrane-Orcutt eljárás
  - OLS eljárással megbecsüljük az u_t értékeket
  - Ez alapján becsülhető ρ

$$\rho = \frac{\sum u_t u_{t-1}}{\sum u_t^2}$$

- Durbin módszer
  - Az előző dián levezetett egyenlet felírható a következő módon

$$y_t = \alpha(1-\rho) + \rho y_{t-1} + \beta x_t - \beta\rho x_{t-1} + e_t$$

Rácspontos keresési eljárások

- Hildreth és Lu eljárása
  - Meghatározzuk y_t^* és x_t^* értékeket különböző ρ értékekre
  - Ahol a modell RSS értéke a legkisebb az a ρ optimális értéke
- ML eljárás
  - Felírjuk a koncentrált likelihood függvényt

$$\ln L = const - \frac{T}{2}\ln\sigma_e^2 + \frac{1}{2}\ln(1-\rho^2) - \frac{Q}{2\sigma_e^2}$$

$$Q = \sum[y_t - \rho y_{t-1} - \alpha(1-\rho) - \beta x_t + \beta\rho x_{t-1}]^2$$

$$\sigma_e^2 = var(e_t)$$

von Neumann-hányados

- A DW-próba alkalmazásának számos megkötése van
- Emiatt több más próbát is alkalmaznak az elsőrendű autokorreláció vizsgálatára
- Neumann-hányados

$$\frac{\delta^2}{s^2} = \frac{\sum_{t=2}^{n} \frac{(e_t - e_{t-1})^2}{n-1}}{\sum_{t=1}^{n} \frac{(e_t - \bar{e})^2}{n}}$$

- Fontos megjegyezni, hogy e_t-nek meg kell felelnie a Gauss–Markov-tétel követelményeinek, így ehhez klasszikus OLS nem alkalmazható
- Szignifikancia-teszt u-próbával

$$\frac{\delta^2}{s^2} \sim \mathcal{N}\left(\frac{2n}{n-1}, \frac{4n^2(n-2)}{(n+1)(n-1)^3}\right)$$

Tartalom

Autoregresszió

Idősorelemzés

Idősorelemzés – Bevezetés

- Az autokorreláció és autoregresszió leggyakrabban az idősorok elemzésében kap szerepet
- Az idősorok elemzésénél alapelv, hogy az idősorok négy fő komponensből állnak
  - Trend: hosszabb időszakon keresztül tartósan érvényesülő tendencia
  - Ciklus: trend feletti vagy alatti tartósabb nem szabályos mozgás, mely megfigyelésére csak hosszabb idősorok alkalmasak
  - Szezonalitás: rövid távú ingadozás, amely szabályos időközönként azonos irányba befolyásolja az idősorértékeket
  - Véletlen ingadozás
- Az idősor adatok esetén nagyon fontos, hogy mindig a korábbi értékből következik a későbbi, ami nagy mértékben megkönnyíti az elemzéseket

Indexszámok

- Idősoradatok esetén indexszámok értelmezhetők
  - Bázisviszonyszám (b_i): egy kitüntetett időpont adatához (x_0) viszonyítás

$$b_i = \frac{x_i}{x_0} \; \forall i$$

  - Láncviszonyszám (l_i): az előző időszak adatához viszonyítás

$$l_i = \frac{x_i}{x_{i-1}} \; \forall i$$

- A kettő átszámítható egymásba

$$l_i = \frac{b_i}{b_{i-1}}$$

$$b_i = \prod_{j=1}^{i} l_i$$

Valószínűségszámítási alapok

- ! X és Y diszkrét valószínűségi változó
- Várható érték: $\mathbf{E}(X) = \sum_{i=1}^{\infty} x_i p_i$, ahol $p_i = \mathbf{P}(X = x_i)$
- Variancia (szórásnégyzet)

$$Var(X) = E(X^2) - \big(E(X)\big)^2$$

- Szórás: $D(X) = \sqrt{Var(X)}$
- Kovariancia: két egymástól különböző val. vált. együttmozgása

$$Cov(X,Y) = E\Big(\big(X - E(X)\big)\big(Y - E(Y)\big)\Big)$$

$$Cov(X,X) = Var(X)$$

- Korreláció

$$R(X,Y) = \frac{Cov(X,Y)}{D(X)D(Y)}$$

Stacionaritás

- ! X_t egy t diszkrét változótól függő idősor
- Definiáljuk a következő értékeket:
  - $E(X_t) = \mu(t)$
  - $Var(X_t) = \sigma^2(t)$
  - $\gamma(t_1, t_2) = Cov(X_{t1}, X_{t2})$ autokovariancia
- Szigorú stacionaritás: bármely $X(t_1), X(t_2), \ldots$ $X(t_n)$ n elemű megfigyeléssorozat együttes eloszlása megegyezik $X(t_{1+k}), X(t_{2+k}), \ldots$ $X(t_{n+k})$ együttes eloszlásával

Szigorú stacionaritás

- Tfh. $n = 2$
- $X(t_1), X(t_2)$ és $X(t_{1+k}), X(t_{2+k})$ együttes eloszlása megegyezik
- A szigorú stacionaritás megköveteli, hogy $\forall k$ érték esetén teljesüljön, tehát $!\, k = t_2 - t_1$
- Ekkor igaz lesz, hogy $X(t_1), X(t_{1+k})$ és $X(t_2), X(t_{2+k})$ együttes eloszlása is megegyezik, tehát az eloszlás csak k-tól függ
- $!\, \gamma(k) = Cov(X_t, X_{t+k})$ autokovariancia függvény
- $\rho(k) = \frac{\gamma(k)}{\gamma(0)}$ autokorrelációs függvény

Stacionaritás

- A szigorú stacionaritás a legritkább esetben teljesül, ezért definiáljuk a gyenge stacionaritást
- Ezt szokták stacionaritásnak nevezni
- Gyenge stacionaritás
  - Várható érték állandó: $E(X_t) = \mu$
  - Az autokovariancia függvény csak a késleltetéstől függ: $Cov(X_t, X_{t+k}) = \gamma(k)\ \forall t, k$

Nevezetes idősorok

- Tisztán véletlen folyamatok (fehér zaj)
  - Független, azonos eloszlású diszkrét időközönként megfigyelt idősor
  - $\gamma(k) = Cov(X_t, X_{t+k}) = 0 \ \forall k \neq 0$
  - $\rho(k) = \begin{cases} 1, & k = 0 \\ 0, & k \neq 0 \end{cases}$
- Véletlen bolyongás
  - ! ε_t tisztán véletlen folyamat μ várható értékkel és σ^2 varianciával
  - Az $\{X_t\}$ folyamat véletlen bolyongás, ha

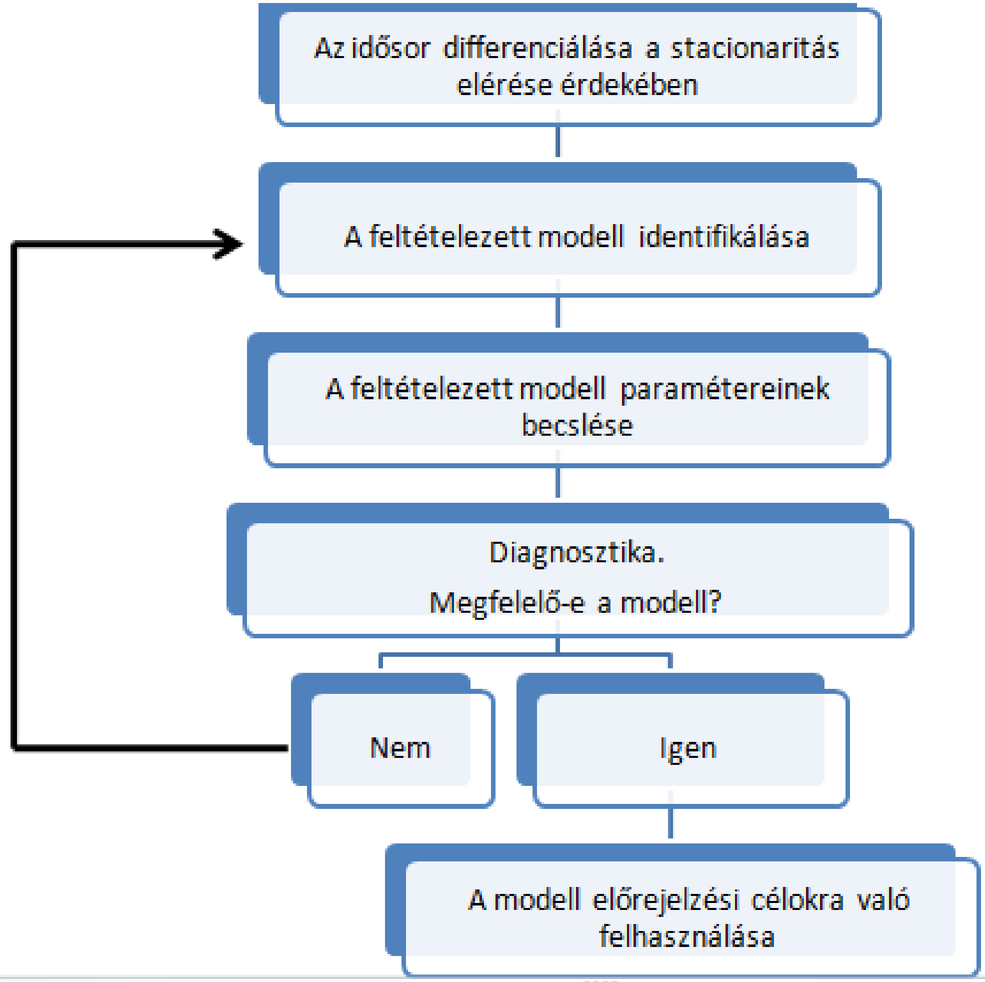
$$X_t = X_{t-1} + \varepsilon_t \ \forall t$$

Nevezetes idősorok

- Mozgóátlag folyamok
  - ! ε_t tisztán véletlen folyamat μ várható értékkel és σ^2 varianciával
  - $X_t = \beta_0\varepsilon_t + \beta_1\varepsilon_{t-1} + \beta_2\varepsilon_{t-2} + \cdots + \beta_m\varepsilon_{t-m}$ idősort m-ed rendű mozgóátlag-folyamatnak (MA(m)) nevezzük
- Autoregresszív folyamok
  - $X_t = \alpha_1 X_{t-1} + \alpha_2 X_{t-2} + \cdots + \alpha_r X_{t-r} + \varepsilon_t$ idősort r-ed rendű autoregresszív folyamatnak (AR(r)) nevezzük
- Autoregresszív mozgóátlag folyamok
  - $X_t = \alpha_1 X_{t-1} + \alpha_2 X_{t-2} + \cdots + \alpha_r X_{t-r} + \varepsilon_t + \beta_1\varepsilon_{t-1} + \beta_2\varepsilon_{t-2} + \cdots + \beta_m\varepsilon_{t-m}$
  - ARMA(r, m)

Box–Jenkins-féle modellezés

1. Az idősor differenciálása a stacionaritás eléréséig
   - MA modellekkel
   - Differenciálással (ARIMA)
2. Felírjuk a megfelelő ARMA(m, r) modellt (m és r meghatározása)
3. ARMA modell paraméterbecslése
   - ML becsléssel
   - Rácspontos kereséssel
4. Diagnosztika
5. Előrejelzés



Ellenőrző kérdések

- Minek a tesztelésére szolgál a Durbin-Watson próba, és milyen korlátai vannak?
- Milyen egyirányú összefüggés fedezhető fel az idősoros adatokban és mi ennek a következménye?
- Mi a stacionáris idősorok legfontosabb tulajdonsága?

KÖSZÖNÖM A MEGTISZTELŐ FIGYELMÜKET!

Dr. Sipos Tibor
Dr. Török Ádám
Szabó Zsombor

MŰEGYETEM 1782

KUKG