

# Idősorok rendbecslése információelméleti módszerekkel

Diplomamunka

Írta: Darabos Beáta

Alkalmazott matematikus szak

Témavezető: Márkus László, egyetemi docens  
Valószínűségelméleti és Statisztika Tanszék



Eötvös Loránd Tudományegyetem  
Természettudományi Kar

2010

# Tartalomjegyzék

BEVEZETÉS . . . . .	5
<b>1. A modellszelekció elméletei</b>	<b>6</b>
1.1. Modellszelekciós módszerek . . . . .	6
1.1.1. Teszteléses eljárások . . . . .	7
1.1.2. Szelekciós kritériumok . . . . .	7
1.1.3. Egyéb módszerek . . . . .	8
1.2. A szelekciós kritériumok tulajdonságai . . . . .	9
<b>2. Információelméleti kritériumok</b>	<b>11</b>
2.1. A Kullback-Leibler távolság . . . . .	11
2.2. Az Akaike kritérium és kiterjesztései . . . . .	14
2.2.1. Az Akaike információs kritérium . . . . .	15
2.2.2. Az általánosított Akaike kritérium . . . . .	16
2.3. A bayesi kritérium . . . . .	18
2.3.1. A maximum 'a posteriori' szabály . . . . .	18
2.3.2. A bayesi információs kritérium . . . . .	19
2.4. A Hannan-Quinn kritérium . . . . .	21
<b>3. A kritériumok tulajdonságai</b>	<b>24</b>
3.1. Konzisztencia . . . . .	24
3.2. Poszt-modellszelekciós becslések . . . . .	26
3.3. Aszimptotikus hatásosság . . . . .	27
<b>4. Empirikus vizsgálatok</b>	<b>30</b>
4.1. Autoregresszív folyamatok modellezése . . . . .	30
4.1.1. Alkalmazott becslési módszerek . . . . .	32
4.1.2. AR(1) folyamat rendbecslése . . . . .	33
4.1.3. AR(2) folyamatok becslése . . . . .	35

4.2. Rendbecslés ARMA folyamatok esetén . . . . .	37
4.3. A félrespecifikálás esetei . . . . .	40
4.3.1. Autoregresszív folyamat nem Gauss zajjal . . . . .	40
4.3.2. Autoregresszív folyamat ARCH zajjal . . . . .	41
4.3.3. Rezsinváltó modellek . . . . .	42

# Ábrák jegyzéke

4.1.	Helyes rendbecslések aránya $\theta$ függvényében, $n = 500$ . . . . .	34
4.2.	Helyes rendbecslések aránya $\theta$ függvényében, $n = 2000$ . . . . .	34
4.3.	AR(2) folyamat stacionárius pontthalmaza . . . . .	35
4.4.	Helyes rendbecslések aránya a gyökök hosszának függvényében	35
4.5.	Helyes rendbecslések aránya $\theta_1$ függvényében, $\theta_2 = -0.9$ . . .	36
4.6.	Helyes rendbecslések aránya a valós gyök hosszának függvényében . . . . .	37
4.7.	Helyes rendbecslések aránya a $\theta$ paraméter függvényében, $\phi = 0.7$ . . . . .	39
4.8.	Helyes rendbecslések aránya a $\theta$ paraméter függvényében, $\phi = 0.2$ . . . . .	39

# Táblázatok jegyzéke

4.1. Becsült rendek aránya különböző zajok esetén . . . . .	40
4.2. Becsült rendek aránya AR-ARCH folyamat esetén . . . . .	41
4.3. Becsült rendek aránya különböző AR paraméterek mellett . .	42

# BEVEZETÉS

A statisztikai modellalkotás célja a vizsgált folyamatok megismerése és ezáltal azok várható jövőbeli alakulásának becslése. Az utóbbi évtizedekben, elsősorban a számítástechnika fejlődésének köszönhetően, megnőtt az érdeklődés a modellválasztás problémája iránt. A modellszelekció azt jelenti, hogy a megfigyelt adatainkból nem csak egy 'a priori' adott paraméteres modell hiányzó paramétereit becsüljük, hanem magát a modellt is. Ehhez elsőként meg kell határozni az alkalmas modellek halmazát, melyből aztán valamely kritérium szerint a legjobbat választjuk. Hogy mit tekintünk legjobbnak, az a vizsgálat céljától függ.

Idősorok esetében a rendbecslés a modellszelekció egy speciális fajtája. Tekintsük például alkalmas modelleknek az autoregresszív folyamatok családját. Ekkor a modellválasztás problémája leszűkül az autoregressziós egyenletben szereplő legnagyobb késleltetés meghatározására. Másszóval a paraméterek dimenzióját, a folyamat rendjét becsüljük.

A rendbecslési eljárások rendkívül sokfélék, a témával foglalkozó szakirodalom rohamosan bővül. A továbbiakban az információelméleti módszerekre helyezem a hangsúlyt. Egy rövid bevezetés után bemutatom az információs kritériumok családját, köztük az Akaike, bayesi és Hannan-Quinn kritériumokat. Ezt követően szimulációs becslések segítségével vizsgálom ezen szelekciós eljárások tulajdonságait. Céлом egy átfogóbb elemzés nyújtása, mely mind az analitikus, mind az empirikus ismereteket összefoglalva rámutat a kritériumok jó tulajdonságaira, valamint azok korlátaira is.

# 1. fejezet

## A modellszelekció elméletei

### 1.1. Modellszelekciós módszerek

Egy adatsor modellezésekor rendszerint több lehetséges modell közül kell választanunk. Legyenek a megfigyelt adataink  $y = (y_1, \dots, y_n)$  és jelöljük  $\mathcal{M}$ -mel azon modellek halmazát, melyeket illeszthetőnek tartunk az adatsorra. Egy  $\mathcal{M}$ -beli modell tulajdonképpen nem más, mint  $y$ -ra vonatkozó valószínűségi eloszlások gyűjteménye:

$$M = \{P_\theta : \theta \in H\},$$

ahol  $H$  a paraméterter,  $P_\theta$  pedig az  $y$  eloszlása a  $\theta$  paraméter függvényében. Előfordulhat, hogy a fenti halmaz tartalmazza a helyes modellt, azaz valamely  $M \in \mathcal{M}$  modell által leírt  $P_\theta$  eloszlás megfelel az  $y$  adatok valódi eloszlásának.

A modellszelekció egy eljárás, mely a megfigyelt adatokra támaszkodva egy olyan  $\hat{M}$  modell választását eredményezi, mely valamilyen értelemben jól modellezi az adatsort. Amennyiben  $\mathcal{M}$  tartalmazza a helyes modellt, a feladat egyértelmű. A gyakorlatban azonban többnyire nem ismerjük a valódi eloszlást és annak csak egy jó közelítését keressük. Hogy mit tekintünk jó modellnek, leginkább attól függ, mi a végső célunk a folyamat modellezésével. A modellválasztás ugyanis csak az első lépés az elemzési folyamatban, amit a paraméterbecslés, majd az előrejelzés vagy más vizsgálat (pl. kiugró értékek elemzése) követ.

Általában a szelekciós módszer és a paraméterbecslési eljárás szorosan összekapcsolódnak, vagyis az  $M$  modellhez hozzárendelhetjük a  $\hat{\theta}(M)$  függvényt, ami a választott modell függvényében a becsült paramétereket jelöli.

A modellszelekció utáni becslés tehát a  $\hat{\theta}(\hat{M})$  lesz.

Vegyünk példának egy olyan  $y_t$  idősort, ami feltehetően autoregresszív folyamatot követ. Ekkor az illeszthető modellek halmaza az

$$y_t = \theta_1 y_t + \dots + \theta_k y_{t-k} + \epsilon_t$$

alakú egyenletekkel írható le, ahol  $\epsilon_t$  valószínűségi változók független, normális eloszlásúak,  $k \in [0, K]$  pedig a  $\theta$  paramétervektor dimenziója. A modellszelekció problémája ekkor az optimális  $k$  érték, vagyis az idősor rendjének meghatározását jelenti.

### 1.1.1. Teszteléses eljárások

Modellválasztás céljára alkalmazhatunk egyszerű hipotézisvizsgálatot. Tekintsünk elsőként egy két elemű  $\mathcal{M}$  halmazt és tegyük fel, hogy  $M_1$  és  $M_2$  egymásbaágyazott, azaz az  $M_1$  modell  $M_2$  egy speciális eseteként fogható fel. Tegyük fel továbbá, hogy legalább az általánosabb modell helyes, azaz a tényleges eloszlás szerepel a szóbaajövő modellek halmazában. Ekkor a modellszelekció annak a  $H_0$  nullhipotézisnek a tesztje, hogy a valódi eloszlás az  $M_1$  szerinti. Amennyiben  $H_0$ -t elvetjük, az  $M_2$  bővebb modellt érdemes választanunk.

A helyes modell hiányában is alkalmazhatunk teszteket. Ha például a vizsgálatunk célja, hogy minél jobb előrejelzést adjunk, a nullhipotézist módosíthatjuk eszerint:  $H_0$  jelentse, hogy az előrejelzés átlagos négyzetes hibája (mean squared error of prediction) az  $M_1$  modellben kisebb.

Több egymásba ágyazott modell esetén a fenti eljárás kiterjeszhető tesztek láncolatává, az általánostól az egyre specifikusabb modellek felé.

### 1.1.2. Szelekciós kritériumok

A modellszelekciós kritériumok megjelenésével a modellválasztás kérdése hipotézisvizsgálatból becslési problémává alakult át. A kritériumok mindegyike valamely kockázat vagy hiba minimalizálásán alapul, azaz az illeszthető modellek halmazából azt az eloszlást választják, amire nézve az adott hibafüggvény a legkisebb értéket veszi fel.

A legtöbb kritérium a becslült és a valódi folyamat közötti eltérést tekintik kockázatnak. Lineáris modellek esetén <sup>1</sup> a végső előrejelzési hiba (final

---

<sup>1</sup>autoregresszió, lineáris regresszió



prediction error) kritérium az előrejelzés átlagos négyzetes hibáját, míg a Mallows-féle  $C_p$  statisztika az ún. empirikus kockázatot minimalizálja.<sup>2</sup> Szintén elterjedt módszer a Parzen-féle CAT kritérium, ami a spektrálsűrűségfüggvények eltérését, az ún. integrált relatív négyzetes hibát minimalizálja autoregresszív folyamatok esetén.

Az információelméleti módszerek a fentiekkel szemben általánosan kiterjeszthetők bármely modelltípusra. Hibafüggvényként tekintünk a  $P_{\hat{\theta}(M)}$  modelleloszlás és az  $Y$  adatok tényleges eloszlása közötti Kullback-Leibler távolságot. A minimalizálandó függvény az alábbi általános alakba írható:

$$IC(M) = -2 \ln f(y, \hat{\theta}_M) + C(n)k_M, \quad (1.1)$$

ahol  $f$  az adott modellnek megfelelő likelihood-függvény,  $k_M \in \mathbb{Z}$  a modellbeli paraméterek száma (dimenziója),  $\theta_M$  a paramétervektor,  $C(n)$  pedig az ún. büntetőfaktor, ami a megfigyelések számától függ. Az általunk vizsgált információs kritériumok mindössze a  $C(n)$  tényező felírásában különböznek egymástól. A legelterjedtebb módszerek az Akaike információs kritérium (AIC), ahol  $C(n) = 2$ , valamint a Schwarz-féle bayesi információs kritérium (BIC), melyre  $C(n) = \ln n$ .

Szintén információelméleti alapokon nyugszik a különböző struktúrájú modellek összehasonlítására alkalmas minimális leíráshossz (minimum description length), valamint annak bayesi alternatívája a minimális üzenethossz (minimum message length) kritérium. A módszer koncepciója, hogy azt a modellt választja, ami a megfigyelt adatokat a legrövidebb kóddal írja le.

Érdemes megemlíteni még a Spiegelhalter által kidolgozott deviancia információs kritériumot<sup>3</sup>, mely bonyolult hierarchikus modellek esetén az optimális paraméterszám meghatározására szolgál. Ez az eljárás elsősorban akkor hasznos, amikor a paraméterek száma összemérhető a mintaelemszámmal<sup>4</sup>.

### 1.1.3. Egyéb módszerek

A modellszelekciós kritériumok elterjedése előtt a megfelelő modell kiválasztására kvázi-intuitív módszerek álltak rendelkezésre. Autoregresszív folyamatok rendjének becslésére Whittle az alábbi eljárást javasolta<sup>5</sup>: Ábrázoljuk a

<sup>2</sup> $MSE(\hat{Y} - \theta X)$ , ahol  $X$  a lineáris regresszió magyarázó változóit jelöli

<sup>3</sup>lásd Spiegelhalter et al. (2002)

<sup>4</sup>különbben ekvivalens az Akaike kritériummal

<sup>5</sup>lásd Chic (2002)

becsült rend függvényében a becsült reziduális varianciát és válasszuk azt a rendet, melytől kezdve a függvény stagnál. Mindemögött az az egyszerű gondolatmenet húzódik meg, miszerint ha a becsült rend kisebb a valódi rendnél, a reziduális variancia nagyobb lesz, mivel a hiányzó tagok magyarázzák a variancia további részét. Ha a becsült rend eléri vagy meghaladja a rendet, a variancia nem csökken tovább.

Mozgóátlag vagy autoregresszív folyamatok esetén a tapasztalati autokorreláció illetve parciális autokorreláció függvény nyújthat támpontot. Az előbbi esetében ugyanis az autokorreláció a rendnél nagyobb késleltetések esetében 0, míg az utóbbi esetén ugyanez a parciális autokorrelációra teljesül.

Az egyéb módszerek közé sorolhatjuk még a modell-átlagoló eljárásokat, melyek egyetlen modell kiválasztása helyett súlyokat rendelnek a lehetséges modellekhez. További szelekciós módszerekről lásd bővebben Leeb-Pötscher (2006).<sup>6</sup>

## 1.2. A szelekciós kritériumok tulajdonságai

A modellszelekciós kritériumok tulajdonságaival foglalkozó szakirodalom szintén szerteágazó. Az elméletek két nagy témakörbe sorolhatók.

Elsőként tegyük fel, hogy az  $y$  adatvektor valamely ismert eloszlásból való, véges sok paraméterrel. Ekkor feltételezhetjük, hogy az  $\mathcal{M}$  halmaz tartalmazza a valódi eloszlást - ez legyen az  $M$  modell szerinti - és véges. Egy adott szelekciós kritérium által választott  $\hat{M}$ -ot tekinthetjük a valódi  $M$  modell egy becslésének. Egy módszer jóságát tehát mérhetjük annak a valószínűségével, hogy az a helyes modellt választja.

Konzisztensnek nevezzük azon kritériumokat, melyekre a valódi eloszlás detektálásának valószínűsége 1-hez tart, ha a mintaelemszámot növeljük.<sup>7</sup>

$$\lim_{n \rightarrow \infty} P_{n,\theta}(\hat{M} = M) = 1 \quad (1.2)$$

Gyengébb tulajdonság, ha egy kritérium olyan modellt választ, melybe beágyazható a valódi modell, azaz a szűkebb, vagy félrespecifikált modelleket

<sup>6</sup>Léteznek olyan paraméterbecslési módszerek, melyek túlparaméterezés esetén is képesek pontos becslést adni. Lásd pl. Lu-Ju-Chon (2001)

<sup>7</sup>gyenge konzisztencia

kiszűri. Egy módszer konzervatív, ha nem konzisztens, ugyanakkor a félre-specifikálás valószínűsége aszimptotikusan 0.

$$\lim_{n \rightarrow \infty} P_{n,\theta}(\hat{M} \not\supseteq M) = 0, \quad (1.3)$$

ahol  $\supseteq$  az egymásbaágyazottságot jelöli. Egy szelekciós kritérium akkor használható a gyakorlatban, ha legalább a konzervatív tulajdonságot teljesíti.

A másik megközelítés szerint a valódi adatgeneráló folyamat jóval összetettebb az általunk használt modelleknél. Érdemesebb tehát feltenni, hogy  $\mathcal{M}$  nem tartalmazza a helyes modellt. Ekkor a modellbeli eloszlás a valódi eloszlásnak egy közelítéseként interpretálható. Gyakran a paraméterek dimenziója jóval meghaladja a megfigyelések számát, ilyenkor az adatgeneráló folyamatot végtelen dimenziósnek tekinthetjük. Végtelen dimenziós folyamat véges közelítése esetén egy szelekciós kritérium teljesítményét Shibata szerint az előrejelzés átlagos négyzetes hibájával célszerű mérni.

## 2. fejezet

# Információelméleti kritériumok

A most következő fejezetben az információs kritériumok elméleti hátterét mutatjuk be. A kiindulópont mindegyik módszer esetében az információelméletből ismert Kullback-Leibler információ vagy távolság. A modellbeli és a valódi eloszlás különbözőségét a továbbiakban ezzel a távolsággal mérjük. A KL információ eltérő feltételezésekre alapozott becslése különböző kritériumokhoz vezet.

Vezessük be az alábbi jelöléseket:

- $y$  legyen az  $n$  dimenziós (valós) adatvektor, amely a  $Y$  valószínűségi változó egy realizációja,
- $\theta$  legyen a valós paramétervektor,
- $k$  egész legyen a paramétervektor dimenziója, amiről feltételezzük, hogy nem függ  $n$ -től.

A továbbiakban célunk, hogy az ismeretlen  $\theta$  paramétervektor dimenzióját meghatározzuk, azaz a modellszelekció problémáját a  $k$  rend becslésére korlátozzuk.

### 2.1. A Kullback-Leibler távolság

Az információelméletben a Kullback-Leibler divergencia vagy távolság két valószínűségi eloszlás különbözőségét méri. Az egyik tipikusan az elméleti eloszlást, míg a másik ennek egy modelljét reprezentálja. A közöttük lévő távolság felfogható úgy, mint a modellezésből származó információvesztés

vagy hiba. A Kullback-Leibler távolság ugyan nem-negatív, de nem valódi metrika, mivel nem szimmetrikus, azaz megkülönbözteti a modell és modellezett eloszlást.

Tegyük fel, hogy az  $Y$  valószínűségi változó abszolút folytonos,  $f_Y(y)$  jelölje a sűrűségfüggvényt,  $f_Y(y, \theta)$  pedig a paraméterektől függő likelihoodot. A modelltől származó sűrűségfüggvényt jelöljük  $f_M(y)$ -mel. A két függvény Kullback-Leibler távolsága

$$D(f_Y, f_M) = \int f_Y(y) \ln \frac{f_Y(y)}{f_M(y)} dy.$$

Legyen  $E_Y$  az  $f_Y$  szerinti várható érték, így a fenti definíció az alábbi alakra hozható:

$$D(f_Y, f_M) = E_Y(\ln f_Y(y)) - E_Y(\ln f_M(y)). \quad (2.1)$$

A divergencia-függvényre teljesül, hogy

$$D(f_Y, f_M) \geq 0$$

$$D(f_Y, f_M) = 0 \Leftrightarrow f_Y(y) = f_M(y)$$

A fenti definíció azt sugallja, hogy azt a modellt érdemes választanunk, amelyre a modellbeli és a valódi eloszlás Kullback-Leibler távolsága a legkisebb. Az információs kritériumok mindegyike erre a koncepcióra épül. (2.1) első tagja a modellezés szempontjából konstans, a divergencia minimalizálása tehát ekvivalens az alábbi függvény maximalizálásával:

$$I(M) := E_Y(\ln f_M(y))$$

Rendbecslés esetén a lehetséges modellek  $\mathcal{M}$  halmaza az  $M = \{P_{\theta^k} : k \in [0, K]\}$  alakú modellekből áll. Tegyük fel, hogy  $\mathcal{M}$  tartalmazza a valódi eloszlást. Az  $\ln f_M(y)$  tehát az  $\ln f_Y(y, \theta^k)$  likelihoodfüggvény lesz, azaz

$$I(k) := E_Y(\ln f_Y(y, \theta^k)) \quad (2.2)$$

(2.2) kiszámításához a valódi paraméterek nem állnak rendelkezésre, így az  $\ln f_Y(y, \theta^k)$  helyett csak az  $\ln f_Y(y, \hat{\theta}^k)$  függvényt használhatjuk, ahol  $\hat{\theta}^k$  a paraméterek egy becslése adott  $k$  rend mellett. Az információs kritériumok eredetileg maximum-likelihood paraméterbecsléshez kötöttek, így a továbbiakban  $\hat{\theta}^k$  az ML becslést jelöli. <sup>1</sup>

<sup>1</sup>A gyakorlatban sokkal elterjedtebb a könnyben számolható Yule-Walker becslés autoregresszív folyamatok esetén.

Mivel a valódi eloszlás sem ismert,  $E_Y f_Y(y, \hat{\theta}^k)$  helyett annak egy torzítatlan becslését maximalizáljuk:

$$\hat{I}(k) = \ln f_Y(y, \hat{\theta}^k) \quad (2.3)$$

Az a szelekciós módszer, ami (2.3) minimalizálására épül, nem lesz megfelelő. Különösen egymásbaágyazott modellek esetén (a rendbecslés tipikusan ilyen) a  $\hat{I}(k)$  függvény monoton növekedő lesz  $k$ -ban, így mindig a legnagyobb lehetséges rendet,  $K$ -t választanánk. Valójában ez a "kritérium" nem veszi figyelembe a véletlen hatását, és a meglévő adatokra a lehető legprecízebb modellt igyekszik illeszteni, meggátolva ezzel a folyamat valódi tulajdonságainak vizsgálatát (pl. előrejelzésre nem alkalmas).

A fenti eljárás helyett közelítsük az ismeretlen modellbeli likelihoodot annak  $\hat{\theta}^k$  körüli Taylor-sorával:

$$\begin{aligned} \ln f_Y(y, \theta^k) &\approx \ln f_Y(y, \hat{\theta}^k) + (\theta^k - \hat{\theta}^k)^T \frac{\partial \ln f_Y(y, \theta^k)}{\partial \theta^k} \Big|_{\hat{\theta}^k} + \\ &+ \frac{1}{2} (\theta^k - \hat{\theta}^k)^T \frac{\partial^2 \ln f_Y(y, \theta^k)}{(\partial \theta^k) (\partial \theta^k)^T} \Big|_{\hat{\theta}^k} (\theta^k - \hat{\theta}^k) \end{aligned} \quad (2.4)$$

Mivel  $\hat{\theta}$  a maximum likelihood becslés, az első derivált éppen  $\hat{\theta}$ -ban 0.

További közelítésekhez tekintsük a likelihood becslés ismert aszimptotikus tulajdonságait: Tegyük fel, hogy az erős regularitási feltételek teljesülnek, ekkor a paraméterek ML becslése aszimptotikusan normális. A várható érték  $\theta$ -val, a variancia pedig a Cramér-Rao határral lesz egyenlő<sup>2</sup>. A Cramér-Rao határ reciproka a Fisher-féle információs mátrix:

$$J = -E_Y \frac{\partial^2 \ln f_Y(y, \theta^k)}{(\partial \theta^k) (\partial \theta^k)^T}$$

Bizonyos gyenge feltételek mellett elég nagy  $n$ -re teljesül, hogy

$$-\frac{1}{n} \frac{\partial^2 \ln f_Y(y, \theta^k)}{(\partial \theta^k) (\partial \theta^k)^T} \Big|_{\hat{\theta}^k} \simeq \frac{1}{n} J = \mathcal{O}(1) \quad (2.5)$$

Így a likelihood-függvény becslése:

$$\ln f_Y(y, \theta^k) \approx \ln f_Y(y, \hat{\theta}^k) - \frac{1}{2} (\theta^k - \hat{\theta}^k)^T J (\theta^k - \hat{\theta}^k)$$

---

<sup>2</sup>a maximum likelihood becslés konzisztens és aszimptotikusan hatásos

A második tag  $Y$  szerinti várható értékét véve

$$E_Y((\theta^k - \hat{\theta}^k)^T J(\theta^k - \hat{\theta}^k)) = \text{tr}(J E_Y(\theta^k - \hat{\theta}^k)(\theta^k - \hat{\theta}^k)^T) \approx \text{tr}(J J^{-1}) = n$$

Az első tag várható értéke helyett vegyük annak torzítatlan becslését, így a becült Kullback-Leibler távolság

$$\hat{I}(k) = \ln f_Y(y, \hat{\theta}^k) - \frac{k}{2},$$

a belőle származó ún. névtelen kritérium pedig

$$NN(k) = -2 \ln f_Y(y, \hat{\theta}^k) + k \quad (2.6)$$

Ez a függvény már nem lesz monoton növekedő, hiszen a második, ún. büntető tag ellensúlyozza a likelihood növekedését. Ennek ellenére a fenti kritérium a gyakorlatban nem használatos, mivel hajlamos a túlbecslésre. Ezt úgy is interpretálhatjuk, hogy a plusz változó bevezetésével járó büntetés mértéke nem elég nagy.

A továbbiakban bemutatásra kerülő kritériumok alakja hasonló, csak a büntető tagban térnek el.

$$IC(k) = -2 \ln f_Y(y, \hat{\theta}^k) + C(n)k \quad (2.7)$$

A fenti forma elsősorban azért elterjedt, mert normális eloszlású adatok feltételezése mellett az alábbi alakra hozható:

$$IC(k) = n \ln \hat{\sigma}^2 + C(n)k, \quad (2.8)$$

ahol  $\hat{\sigma}^2$  a szórásnégyzet maximum-likelihood becslése.

## 2.2. Az Akaike kritérium és kiterjesztései

A valódi és a becült modell közötti Kullback-Leibler távolság meghatározásához közelítéseket kell alkalmaznunk. Láthattuk, hogy a modell-likelihood Taylor-sorba fejtése nem ad a gyakorlat számára megfelelő közelítést. A fenti gondolatmenet egyik hiányossága, hogy ugyanazt a mintát veszi alapul mind a paraméterbecsléshez, mind a likelihood-függvény kiszámításához. Ezt ki-küszöbölendő, Akaike két egymástól független, azonos eloszlású adatsort feltételez a becsléshez és a kiértékeléshez, ezt nevezzük a kereszt-kiértékelés módszerének.

### 2.2.1. Az Akaike információs kritérium

Legyen  $(x_1, \dots, x_n)$   $Y$ -nak egy  $(y_1, \dots, y_n)$ -től független realizációja. A paraméterek likelihood-becsléséhez ezt a fiktív adatsort használjuk,  $\hat{\theta}_x$  jelölje tehát az  $x$  megfigyelések szerinti ML becslést. Az ismeretlen  $\ln f_Y(y, \theta^k)$  likelihood-függvényt helyettesítsük most az  $E_x(\ln f_Y(y, \hat{\theta}_x^k))$  függvénnyel, ahol  $E_x$  az  $x$  minta szerinti várható értéket jelöli.<sup>3</sup> A minimalizálandó kritérium tehát

$$I(k) := E_y(E_x(\ln f_Y(y, \hat{\theta}_x^k))) \quad (2.9)$$

A fiktív minta szerinti likelihood-becslés nyilván nem áll rendelkezésre, tehát ismét közelítést kell alkalmaznunk. Tekintsük most az  $\ln f(y, \hat{\theta}_x^k)$  Taylor-sorát a  $\hat{\theta}_y^k$  körül:

$$\begin{aligned} \ln f_Y(y, \hat{\theta}_x^k) &\approx \ln f_Y(y, \hat{\theta}_y^k) + (\hat{\theta}_x^k - \hat{\theta}_y^k)^T \frac{\partial \ln f_Y(y, \theta^k)}{\partial \theta^k} \Big|_{\hat{\theta}_y^k} + \\ &+ \frac{1}{2} (\hat{\theta}_x^k - \hat{\theta}_y^k)^T \frac{\partial^2 \ln f_Y(y, \theta^k)}{(\partial \theta^k) (\partial \theta^k)^T} \Big|_{\hat{\theta}_y^k} (\hat{\theta}_y^k - \hat{\theta}_y^k) \end{aligned}$$

Az első derivált természetesen 0, a második deriváltat pedig a korábbiakhoz hasonlóan közelíthetjük a Fisher-információval.

$$\ln f_Y(y, \hat{\theta}_x^k) \approx \ln f_Y(y, \hat{\theta}_y^k) - \frac{1}{2} (\hat{\theta}_x^k - \hat{\theta}_y^k)^T J (\hat{\theta}_x^k - \hat{\theta}_y^k)$$

A második tag várható értékeit véve

$$\begin{aligned} E_y(E_x(\hat{\theta}_x^k - \hat{\theta}_y^k)^T J (\hat{\theta}_x^k - \hat{\theta}_y^k)) &= \\ &= \text{tr}(J E_y E_x((\hat{\theta}_x^k - \theta) - (\hat{\theta}_y^k - \theta))((\hat{\theta}_x^k - \theta) - (\hat{\theta}_y^k - \theta))^T) \approx \\ &\approx \text{tr}(J(J^{-1} + J^{-1})) = 2k \end{aligned}$$

Az első tag  $x$  tekintetében konstans. Az  $y$  szerinti várható érték helyett a szokásos torzítatlan becslést véve kapjuk:

$$\hat{I}(k) = \ln f_Y(y, \hat{\theta}^k) - k,$$

ami a szokásos alakra hozva az Akaike információs kritériumot adja.

$$AIC(k) = -2 \ln f_Y(y, \hat{\theta}^k) + 2k \quad (2.10)$$

<sup>3</sup>Nyilván ez megegyezik a korábbi  $E_Y$ -nal, de itt most szükséges a minták szerinti megkülönböztetés is.



Az AIC kritérium bizonyítottan nem konzisztens, azaz ha feltesszük, hogy a lehetséges modelleink halmaza tartalmazza a valódi eloszlást, akkor a helyes modell detektálásának valószínűsége kisebb 1-nél. Mivel viszont konzervatív, a félrespecifikálás esélye minimális. Rendbecslés esetén ez annyit tesz, hogy a tényleges  $k$  rendet hajlamos felülbecsülni, míg az alulbecslés valószínűsége 0-hoz tart. Formálisan

$$\lim_{n \rightarrow \infty} P(\hat{k}_{AIC} > k) = c > 0$$

$$\lim_{n \rightarrow \infty} P(\hat{k}_{AIC} < k) = 0$$

ahol  $c$  konstans,  $n$  pedig a mintaelemszám.

A fenti tulajdonságok alapján azt mondhatnánk, hogy a büntető tag tehát még mindig nem elég nagy, ennek ellenére az AIC a gyakorlatban mégis hasznos módszernek bizonyul. A valóságban ugyanis ritkán áll fenn az az eset, hogy a lehetséges modellek halmaza tartalmazná a valódi (ismeretlen) eloszlást. Ekkor a célunk egy megfelelő közelítő eloszlást találni az megfigyelt adatainkhoz. Tegyük fel, hogy a folyamat rendje végtelen és annak egy jó véges modelljét keressük. Shibata bebizonyította, hogy az AIC az egy lépéses előrejelzési hiba tekintetben optimális (Shibata-féle aszimptotikus hatásosság), azaz az átlagos négyzetes előrejelzési hiba aszimptotikus értelemben minimális.

Lezárásképpen tekintsük az Akaike kritérium egy módosítását. Az AIC, mint láttuk (2.9) egy aszimptotikusan torzítatlan becsléséből adódik. Lineáris regresszióra létezik nem csak aszimptotikusan torzítatlan becslés, melyből az alábbi kritérium származik:

$$AIC_C(k) = -2 \ln f_Y(y, \hat{\theta}^k) + \frac{2n}{n-k-1} k, \quad (2.11)$$

azaz a büntető tag függ a mintanagyságtól. Az  $AIC_C$  aszimptotikusan meg egyezik az eredeti AIC-vel, azonban véges minta esetén a büntetőtag értéke nagyobb, csökkentve a túlbecslés kockázatát.

### 2.2.2. Az általánosított Akaike kritérium

Az Akaike kritérium általánosításaképp tekintsük az alábbi célfüggvényt:

$$GIC(k) = -2 \ln f_Y(y, \hat{\theta}^k) + \nu k, \quad (2.12)$$

melyben  $\nu$  szintén  $N$ -től független konstans. Empirikus eredmények azt igazolják, hogy a fenti kritérium a teljesítménymérési szemponttól függően  $\nu \in [2, 6]$  esetben bizonyul a legjobbnak. Ha  $\nu = 2$ , az AIC-t kapjuk.

Elméleti megalapozásként a kereszt-kiértékelés módszerét végezzük két olyan független mintával, ahol az  $y$  kiértékelő vektor hossza többszöröse az  $x$  becslő vektorénak.

$$n = \text{hossz}(y) = \rho \text{hossz}(x) \quad (\rho \geq 1)$$

Túlbecslés lényegében akkor áll fenn, amikor a  $\hat{\theta}_x$  likelihood-becslés a mintában lévő zajt is lefedi, azaz az  $f(x, \hat{\theta}_x)$  jóval meghaladja a valódi  $f(x, \theta)$  likelihood értéket. Ilyenkor minél hosszabb a kiértékelő-vektor, az  $f(y, \hat{\theta}_x)$  értéke annál inkább lesz alacsony  $f(y, \theta)$ -hoz képest.

Az előző részbeli levezetés ennek értelmében az alábbiak szerint módosul:

$$\ln f_Y(y, \hat{\theta}_x^k) \approx \ln f_Y(y, \hat{\theta}_y^k) - \frac{1}{2} (\hat{\theta}_x^k - \hat{\theta}_y^k)^T J_y (\hat{\theta}_x^k - \hat{\theta}_y^k)$$

A Fisher információs mátrixokra teljesül, hogy

$$J_y = \rho J_x.$$

A második tag várható értékeit véve

$$\begin{aligned} E_y(E_x(\hat{\theta}_x^k - \hat{\theta}_y^k)^T J_y (\hat{\theta}_x^k - \hat{\theta}_y^k)) &= \\ &= \text{tr}(J_y E_y E_x((\hat{\theta}_x^k - \theta) - (\hat{\theta}_y^k - \theta))((\hat{\theta}_x^k - \theta) - (\hat{\theta}_y^k - \theta))^T) = \\ &= \text{tr}(J_y(\rho J_y^{-1} + J_y^{-1})) = (1 + \rho)k \end{aligned}$$

(2.9) torzítatlan becslése most a két adatvektor hosszának arányában módosul

$$\hat{I} = \ln f_Y(y, \hat{\theta}^k) - \frac{1 + \rho}{2} k$$

Az általánosított Akaike kritérium tehát

$$GIC(k) = -2 \ln f_Y(y, \hat{\theta}^k) + (1 + \rho)k \quad (2.13)$$

Nyilván  $\rho = 1$  esetén (vagyis ha a két minta hossza azonos) az AIC-t kapjuk. Amennyiben  $\rho > 1$ , a büntetőfaktor értéke nagyobb lesz, mint az AIC esetén, így a túlbecslés valószínűsége csökkenthető. Ugyanakkor nincsen egyértelmű szabály, hogy melyik arányt érdemes választani. A kritérium jósága továbbra is függ az adott mintanagyságtól, a mintavételi szabálytól és a teljesítményértékelés módjától.

## 2.3. A bayesi kritérium

A statisztika egyik külön ága a bayesi modellezés. Lényege, hogy úgy tekintünk a modellbeli paraméterekre, mint valószínűségi változókra. A most következő fejezetben ebben a szemléletben közelítjük meg a már bemutatott KL információra épülő elméletet. A paraméterek eloszlását 'a priori' eloszlásnak, míg a mintavétel utáni feltételes eloszlást 'a posteriori' eloszlásnak nevezzük.

Mint látni fogjuk, a kapott kritérium a mintanagyságtól való függést expliciten is tartalmazza, azaz a  $C(n)$  büntetőfaktor nem lesz konstans.

### 2.3.1. A maximum 'a posteriori' szabály

Tekintsük a  $H_k$ ,  $k \in [0, K]$  hipotéziscsaládot, ahol  $H_k$  azt a feltevést jelöli, hogy a modellbeli paraméterek valódi rendje  $k$ .

$$H_k : \theta_k \neq 0, \theta_{k+1} = 0, \dots, \theta_K = 0$$

Nyilvánvaló, hogy  $K$ -tól 0-ig a fenti hipotézisek egymás speciális esetei, tehát egymásbaágyazottak. Emellett feltételezzük, hogy kölcsönösen kizáróak is, azaz közülük egyszerre pontosan egy teljesül.

Tegyük fel most, hogy maga a rend is valószínűségi változó, és jelölje  $f_k(H_k)$  az 'a priori' eloszlást. Továbbá legyen  $f_Y(y|H_k)$  az  $y$  minta sűrűségfüggvénye, feltéve hogy a  $k$ -adik hipotézis teljesül.  $H_k$  'a posteriori' eloszlása a Bayes törvény szerint

$$f_k(H_k|y) = \frac{f_Y(y|H_k) f_k(H_k)}{f_Y(y)}$$

A maximum 'a posteriori' (MAP) szabály, a likelihood elmélethez hasonlóan, azt a hipotézist választja, amelyik az adott minta mellett a legvalószínűbb, azaz a becsült rend az  $\hat{k}$  lesz, amelyik a legnagyobb  $f_k(H_k|y)$  'a posteriori' valószínűséghez tartozik.

Mivel  $f_Y(y)$  konstans  $k$ -ban és általában az 'a priori' eloszlást vehetjük egyenletesnek, azaz  $f_k(H_k) = \frac{1}{K}$ , a MAP szabály az alábbi alakban írható fel:

$$\max_k f_Y(y|H_k) \tag{2.14}$$

A fenti szabály optimális abban az értelemben, hogy maximalizálja a jó választás totális valószínűségét:

$$P(\text{jó választás}) = \sum_{i=1}^K P(H_i \text{ választása} \cap H_i \text{ igaz}).$$

### 2.3.2. A bayesi információs kritérium

Nézzük az előző rész elején bemutatott KL elméletet most bayesi szemléletben. Tegyük fel, hogy  $\theta$  paramétervektor valószínűségi változó  $f_\theta(\theta)$  'a priori' sűrűségfüggvénnyel. Az  $f_\theta(\theta)$ -ról feltesszük, hogy kellően sima a  $\hat{\theta}$  ML becslés egy környezetében és nem függ az  $n$  mintaelemszámtól.  $H_k$  jelentse továbbra is azt a feltevést, hogy a paramétervektor dimenziója  $k$ . Ekkor

$$f_Y(y|H_k) = \int f_Y(y, \theta^k) d\theta^k,$$

ahol  $f_Y(y, \theta^k)$  most az együttes sűrűségfüggvényt jelöli. A korábbi  $f_Y(y, \theta^k)$  függvényt most  $f_Y(y|\theta^k)$  jelöli, mivel a  $\theta$  is valószínűségi változó. Átalakítva kapjuk:

$$f_Y(y|H_k) = \int f_Y(y|\theta^k) f_{\theta,k}(\theta^k) d\theta^k = E_\theta(f_Y(y, \theta)) \quad (2.15)$$

A MAP szabállyal összhangban célunk most is a fenti függvény maximalizálása.

Az ismeretlen  $f_Y(y|\theta)$  sűrűségfüggvény becslésére használjuk fel a (2.4) levezetést.

$$f_Y(y|\theta^k) \approx f_Y(y|\hat{\theta}^k) e^{-\frac{1}{2}(\hat{\theta}^k - \theta^k)^T \hat{J}(\hat{\theta}^k - \theta^k)}$$

ahol  $\hat{J}$  a második deriváltat jelöli:

$$\hat{J} = -\frac{\partial^2 \ln f_Y(y|\theta^k)}{(\partial \theta^k)(\partial \theta^k)^T} \Big|_{\hat{\theta}^k} = \hat{\theta}^k$$

Behelyettesítve a (2.15) középső képletébe kapjuk, hogy

$$f_Y(y|H_k) \approx f_Y(y|\hat{\theta}^k) f_\theta(\hat{\theta}^k) \int e^{-\frac{1}{2}(\hat{\theta}^k - \theta^k)^T \hat{J}(\hat{\theta}^k - \theta^k)} d\theta^k$$

A harmadik tényezőt kiegészítve a normális eloszlás sűrűségfüggvényére

$$\begin{aligned} f_Y(y|H_k) &\approx \frac{(2\pi)^{n/2}}{|\hat{J}|^{1/2}} f_Y(y|\hat{\theta}^k) f_\theta(\hat{\theta}^k) \int \frac{1}{(2\pi)^{n/2} |\hat{J}^{-1}|^{1/2}} e^{-\frac{1}{2}(\hat{\theta}^k - \theta^k)^T \hat{J}(\hat{\theta}^k - \theta^k)} d\theta^k = \\ &= \frac{(2\pi)^{n/2}}{|\hat{J}|^{1/2}} f_Y(y|\hat{\theta}^k) f_\theta(\hat{\theta}^k) \end{aligned}$$

A fenti becslés logaritmusát véve

$$\ln f_Y(y|H_k) \approx \ln f_Y(y|\hat{\theta}^k) + \ln f_\theta(\hat{\theta}^k) + \frac{k}{2} \ln 2\pi - \frac{1}{2} \ln |\hat{J}| \quad (2.16)$$

(2.16) egyelőre függ az 'a priori' eloszlástól, amit eddig tetszőlegesnek feltételeztünk. Viszont ha a mintaelemszámmal tartunk a végtelenbe, a  $|\hat{J}|$  determináns növekszik, míg a két középső tag konstans, tehát gyakorlatilag elhanyagolható. A harmadik tag növekedési ütemét a (2.5) alapján jellemezhetjük. Elég általános feltételek mellett

$$\ln |\hat{J}| = \ln \left| n \frac{1}{n} \hat{J} \right| = k \ln n + \ln \left| \frac{1}{n} \hat{J} \right| = k \ln n + \mathcal{O}(1),$$

mivel  $\hat{J}$   $k \times k$ -s mátrix. A  $\mathcal{O}(1)$  tag, mivel konstans, a maximalizálás szempontjából lényegtelen. Közelítéssel tehát az alábbi becslést kaptuk:

$$\ln f_Y(y|H_k) \approx \ln f(y|\hat{\theta}^k) - \frac{k}{2} \ln n$$

A bayesi információs kritérium azt a rendet választja, mely mellett

$$BIC(k) = -2 \ln f_Y(y|\hat{\theta}^k) + k \ln n \quad (2.17)$$

a legkisebb.

A levezetésből adódik, hogy a bayesi kritérium aszimptotikusan megegyezik a MAP-szabállyal, tehát a BIC elég nagy  $n$  esetén maximalizálja a jó választás totális valószínűségét. Emellett bizonyíthatóan konzisztens, vagyis mind az alulbecslés, mind a túlbecslés valószínűsége 0-hoz tart.

$$\lim_{n \rightarrow \infty} P(\hat{k}_{BIC} = k) = 1$$

## 2.4. A Hannan-Quinn kritérium

A fejezet utolsó részében egy erősen konzisztens rendbecslési kritériumot mutatunk be, mely Hannan és Quinn nevéhez fűződik. Az információs kritériumok általános alakjából kiindulva egy olyan büntető faktort keresünk, amely a mintelemszámban a lehető leghalványabban növekszik úgy, hogy a rendbecslés még konzisztens marad.

Erősen konzisztensnek nevezünk egy rendbecslési módszert, ha

$$P(\lim_{n \rightarrow \infty} \hat{k} = k) = 1 \quad (2.18)$$

Mivel  $k \in \mathcal{Z}$ , a fenti definíció azt jelenti, hogy majdnem minden  $y$  megfigyelés esetén létezik olyan  $N$  mintanagyság, hogy  $n > N$ -re a kritérium a valódi rendet adja becslésként.<sup>4</sup>

Vizsgáljuk speciálisan a  $k$  rendű autoregresszív folyamatokat:

$$y_t = \theta_1 y_{t-1} + \dots + \theta_k y_{t-k} + \epsilon_t, \quad (2.19)$$

ahol  $\epsilon_t$  a független, azonos eloszlású zajkomponenseket jelöli. 0 várható értékű,  $\sigma^2$  varianciájú normális eloszlást feltételezve az információs kritériumok általános alakja

$$IC(k) = \ln \hat{\sigma}^2 + \frac{C(n)}{n} k$$

A Hannan és Quinn által bevezetett kritérium

$$HQIC(k) = \ln \hat{\sigma}^2 + \frac{2c \ln \ln n}{n} k \quad (2.20)$$

ahol  $c > 1$  konstans.

$\hat{\sigma}^2$  kiszámítására a maximum likelihood módszer helyett használjuk most a Yule-Walker egyenleteket.

$$\begin{aligned} 0 &= \hat{\gamma}(i) - \sum_{j=1}^k \hat{\theta}(j) \hat{\gamma}(i-j) \quad i = 1, \dots, k \\ \hat{\sigma}^2 &= \hat{\gamma}(0) - \sum_{j=1}^k \hat{\theta}(j) \hat{\gamma}(j) \end{aligned} \quad (2.21)$$

---

<sup>4</sup>nyilvánvalóan erősebb a (1.2)-ben definiált (gyenge) konzisztenciánál

$\hat{\gamma}$  itt az autokovariancia tapasztalati becslését jelöli.

A Yule-Walker egyenletek szerint becsült paraméterek rekurzívan számolhatók.  $\hat{\theta}^k$  legyen a  $k$  rend esetén becsült paramétervektor,  $\hat{\sigma}_k$  pedig a becsült variancia.  $-\hat{\theta}_k^k$  nem lesz más, mint a  $k$ -adrendű parciális autokorreláció becslése, amelyre Hannan és Quinn az alábbi iterált logaritmus tételt bizonyították.

**2.4.1. Tétel.** *Legyen  $y$  egy (2.19) szerinti autoregresszív folyamat, melyre teljesül, hogy*

$$z^k - \theta_1 z^{k-1} + \dots + \theta_{k-1} z^1 + \theta_k \neq 0, |z| \geq 1$$

valamint  $E(\epsilon_t | \epsilon_{t-1}) = 0$ ,  $E(\epsilon_t^2 | \epsilon_{t-1}) = \sigma^2$  és  $E(\epsilon_t^4) < \infty$ . Ekkor

$$\limsup_{n \rightarrow \infty} \frac{n^{1/2} |\rho_n(m | \hat{m} - 1)|}{2 \ln \ln n^{1/2}} = 1, \forall m > k$$

ahol  $\hat{\rho}_n(m | m - 1)$  a parciális autokorreláció Yule-Walker egyenletek szerinti becslését jelöli  $n$  mintanagyság mellett.<sup>5</sup>

$\hat{\sigma}_k$ -ra az alábbi rekurzió írható fel:

$$\hat{\sigma}_k = (1 - \hat{\theta}_k^2) \hat{\sigma}_{k-1}$$

A Hannan-Quinn kritérium a rekurziót iterálva

$$HQIC(k) = \ln \hat{\sigma}_0^2 + \sum_{j=1}^k \ln(1 - \hat{\rho}^2(j | j - 1)) + \frac{2c \ln \ln n}{n} k$$

alakra hozható.  $\sigma_0^2$  Yule-Walker becslése a reziduálisok négyzetösszegével lesz egyenlő, ami a rend szempontjából konstans. Így a HQIC függvény növekménye adott  $m$  pontban

$$\ln(1 - \hat{\rho}^2(m | m - 1)) + \frac{2c \ln \ln n}{n}$$

Ha  $m$  épp a valódi rend, azaz  $m = k$ , akkor  $\rho(k | k - 1) = \theta_k \neq 0$ , és mivel  $\rho(k | \hat{k} - 1)$  konzisztens becslése  $\rho(k | k - 1)$ -nak, elég nagy  $n$ -re a fenti növekmény negatív. Ebből következik, hogy a függvény aszimptotikus értelemben

<sup>5</sup>bizonyítást lásd Hannan-Quinn (1979)

nem érheti el az abszolút minimumát, ha  $m < k$ .  $m > k$  esetén használjuk az iterált logaritmus tételt. Minden  $\epsilon > 0$ -hoz van olyan 1-valószínűséggel véges  $N$  küszöb, hogy  $\forall n > N$ -re

$$|\rho_n(m|\hat{m} - 1)| < (1 + \epsilon) \frac{2 \ln \ln n^{1/2}}{n^{1/2}}$$

$$\ln(1 - (1 + \epsilon) \frac{2 \ln \ln n}{n}) > \frac{2c \ln \ln n}{n},$$

ha  $\epsilon < c - 1$ . Azaz a Hannan-Quinn kritérium autoregresszív folyamatokra erősen konzisztens rendbecslést ad.

Jól látható, hogy a fejezetben bemutatott kritériumok közül a BIC bünteti legjobban a plusz paraméter bevonását. Az (2.7) szerinti büntető tag  $C(n)_{BIC} = \ln n$ . A  $C(n)_{HQIC} = 2c \ln \ln n$  ennél lassabban növekszik, de 1-nél nagyobb  $c$  szorzó esetén még szigorúan konzisztens. Az AIC és GIC faktora ezzel szemben konstans (nem konzisztensek), míg az AICc csökkenő, és aszimptotikusan az AIC-hez közelít.



## 3. fejezet

# A kritériumok tulajdonságai

A modellszelekción módszerekkel kapcsolatos elméleti kutatások két nagy csoportra oszthatók. Az egyik megközelítésben feltesszük, hogy a lehetséges modellek halmaza tartalmazza a valódi eloszlást. Ekkor egy kritérium jóságát elsősorban az határozza meg, hogy milyen eséllyel becsüli jól a helyes modellt. A másik megfontolás szerint a valódi folyamat jóval összetettebb, a modellszelekció célja pedig egy megfelelően közelítő modell kiválasztása. Ilyenkor egy kritérium jóságát mérhetjük annak előrejelző képességével.

### 3.1. Konzisztencia

Tegyük fel, hogy az illesztendő modelleink halmaza véges és tartalmazza a korrekt eloszlást. Ekkor egy kritérium által választott  $\hat{M}$  modellre tekinthetünk úgy, mint a helyes modell,  $M_0$  egy becslésére. Egy becslőfüggvény jóságát vizsgálhatjuk a statisztikából ismert konzisztencia tulajdonság szempontjából.

Rendbecslés esetén a modellszelekció problémája a helyes paraméterszám,  $k_0$  meghatározására korlátozódik. Ekkor a konzisztencia definíciója:

$$\lim_{n \rightarrow \infty} P(\hat{k} = k_0) = 1 \tag{3.1}$$

Helytelen becslést kétféleképpen kaphatunk, ha felül- illetve alulbecsüljük a paraméterek számát. Amennyiben az alulbecslés valószínűsége 0, a felülbecslésé azonban pozitív, a korábban bevezetett konzervativitás tulaj-

donságról beszélünk.

$$\begin{aligned}\lim_{n \rightarrow \infty} P(\hat{k} < k_0) &= 0 \\ \lim_{n \rightarrow \infty} P(\hat{k} > k_0) &> 0\end{aligned}\tag{3.2}$$

Az (2.7) alakú információs kritériumok esetében az alulbecslés valószínűsége aszimptotikusan 0, amennyiben a büntetőfaktorra teljesül, hogy

$$\frac{C(n)}{n} \rightarrow 0 \quad (n \rightarrow \infty).$$

Ha ezenfelül

$$C(n) \rightarrow \infty \quad (n \rightarrow \infty),$$

a becslés konzisztens lesz. <sup>1</sup> Ebből egyenesen következik, hogy az általunk vizsgált BIC és HQIC kritériumok konzisztensek, míg az AIC és kiterjesztései konzervatívak.

Hannan továbbá azt is bizonyította, hogy ARMA folyamatok esetében mind a bayesi, mind a Hannan-Quinn kritérium erősen konzisztens. <sup>2</sup> Vagyis a  $\hat{p}, \hat{q}$  becslésekre 1 valószínűséggel fennáll, hogy bizonyos  $N$  küszöbtől kezdve  $\hat{p} = p_0, \hat{q} = q_0 \forall n > N$ . A tétel ráadásul az (2.8) formában adott kritériumokra vonatkozik, nem feltételezve a normális eloszlást.

**3.1.1. Tétel.** *Ha  $y(t)$  kauzális és invertálható ARMA( $p, q$ ) folyamat*

$$y_t - \theta_1 y_{t-1} - \dots - \theta_p y_{t-p} = \epsilon_t + \phi_1 \epsilon_{t-1} + \dots + \phi_q \epsilon_{t-q}$$

ahol  $\epsilon_t$ -k független, 0 várható értékű, konstans szórású komponensek, akkor a

$$BIC(p, q) = \ln \hat{\sigma}_{p,q}^2 + \frac{\ln n}{n}(p + q)$$

vagy

$$HQIC(p, q) = \ln \hat{\sigma}_{p,q}^2 + \frac{2c \ln \ln n}{n}(p + q) \quad c > 1$$

minimalizálásával kapott  $\hat{p}, \hat{q}$  becslések erősen konzisztensek.

A gyenge konzisztenciának elégséges feltétele a büntetőtag végtelenbe tartása. Ezzel szemben az Akaike kritérium bizonyítottan felülbecsül.

---

<sup>1</sup>lásd Bierens (2006)

<sup>2</sup>lásd Hannan (1980)

**3.1.2. Tétel.** *Az előző tétel feltételei mellett ha  $P$  és  $Q$  a maximális rendet jelöli,  $\hat{p}, \hat{q}$  pedig*

$$AIC(p, q) = \ln \hat{\sigma}_{p,q}^2 + \frac{2}{n}(p + q)$$

*minimalizálásából adódik,*

$$\lim_{n \rightarrow \infty} P(\hat{p} = p_0, \hat{q} = q) = \pi(q - q_0, Q - q)$$

$$\lim_{n \rightarrow \infty} P(\hat{p} < p_0, \hat{q} = q) = 0,$$

*ha  $P = p_0$ , és*

$$\lim_{n \rightarrow \infty} P(\hat{p} = p, \hat{q} = q_0) = \pi(p - p_0, P - p)$$

$$\lim_{n \rightarrow \infty} P(\hat{p} = p, \hat{q} < q_0) = 0,$$

*ha  $Q = q_0$ .*

$q_0$ -t 0-nak választva a tiszta autoregresszió esetét kapjuk. A tétel azt mondja, hogy ekkor egy tetszőleges  $p$  rend választásának aszimptotikus esélye kizárólag  $p - p_0$  és  $P - p$  függvénye, és ez a függvény 0, ha  $p < p_0$ .

## 3.2. Poszt-modellszelekciós becslések

A modellépítés feladata nem korlátozódik kizárólag a helyes modell típus megtalálására. A modellszelekciót rendszerint paraméterbecslés, majd előrejelzés vagy más statisztikák becslésének kiszámítása követ.

Legyen a vizsgálatunk tárgya a  $\nu$  paraméter. Ekkor a modellszelekciót követő becslésünk  $\nu$ -re vonatkozóan  $\hat{\nu}_{PMS} = \hat{\nu}(\hat{M})$  lesz. Nevezzük ezt poszt-modellszelekciós becslésnek.

$\hat{\nu}(\hat{M})$ -et írhatjuk az alábbi szemléletesebb formába is:

$$\hat{\nu}_{PMS} = \sum_{M \in \mathcal{M}} I(\hat{M} = M) \tag{3.3}$$

$\hat{\nu}_{PMS}$  mint valószínűségi változó nem egyezik meg egyik modellhez tartozó  $\hat{\nu}(M)$  becsléssel sem, hanem ezek egy random konvex kombinációja lesz.

Fontos észrevétel, hogy emiatt a választott modell a hozzátartozó paraméterbecsléssel együtt már nem feltétlenül lesz optimális az eredeti célkitűzéseinkhez mérten. Ha például a rendbecslési eljárást a  $\nu$  statisztika átlagos

négyzetes hibájának minimalizálására alapoztuk, a belőle származó poszt-modellszelekciós becslés nem feltétlenül lesz optimális ebből a szempontból, hiszen  $\hat{\nu}_{PMS}$  kívül esik a  $\{\hat{\nu}(M) : M \in \mathcal{M}\}$  halmazon.

Sőt, szélsőséges esetben létezhet olyan régió a paraméterterben, ahol az  $MSE(\hat{\nu}_{PMS})$  meghaladja bármely  $\hat{\nu}(M)$  hibáját. Konzisztens modellszelekció esetén bizonyított, hogy

$$\limsup_{n \rightarrow \infty} \sup_{\theta} MSE_{n,\theta}(\hat{M}) = \infty. \quad (3.4)$$

azaz ha a legrosszabb eseteket vesszük, a négyzetes hiba a végtelenhez konvergál. Konzervatív esetben ugyanez a határérték véges:

$$\limsup_{n \rightarrow \infty} \sup_{\theta} MSE_{n,\theta}(\hat{M}) < \infty. \quad (3.5)$$

A fenti jelenség oka, hogy a poszt-modellszelekciós becslések véges mintás eloszlása nem egyenletesen konvergál az aszimptotikus eloszláshoz. Erről lásd bővebben Leeb-Pötscher (2000). Konzisztens eljárás esetén tehát előfordulhat, hogy a  $\hat{\nu}_{PMS}$  rosszabb becslést eredményez, mint ha egyszerűen a legbővebb modellt választjuk.

### 3.3. Aszimptotikus hatásosság

Mindeztáig arra az esetre koncentráltunk, amikor létezik helyes modell és az a lehetséges modellek halmazának eleme. A lehetséges modellek halmazát végesnek és a mintaelemszámtól függetlennek tekintettük.

Tegyük fel most, hogy egy olyan folyamatot kell modelleznünk, ahol a paraméterek dimenziója nagyságrendileg meghaladja a megfigyelések számát és a szóba jövő modellek közül egyik sem egyezik meg a valódi eloszlással. Ilyenkor érdemes az adatgeneráló folyamatot végtelen dimenziósnak tekinteni, aminek egy megfelelő véges közelítését keressük.

Az aszimptotikus hatásosság fogalmát Shibata vezette be először. Eszerint végtelen rendű folyamat véges közelítése esetén egy szelekciós kritérium aszimptotikusan hatásos, ha az előrejelzés átlagos négyzetes hibája a lehető legkisebb.

Tekintsünk egy végtelen dimenziós autoregresszív folyamatot

$$Y_t + \alpha_1 Y_{t-1} + \alpha_2 Y_{t-2} + \dots = \epsilon_t, \quad (3.6)$$

ahol  $\epsilon_t$ -k 0 várható értékű,  $\sigma^2$  varianciájú, független azonos eloszlású változók.

Jelölje  $\Gamma$  a végtelen dimenziós autokovariancia-mátrixot. Legyen  $\alpha(k) = (\alpha_1(k), \dots, \alpha_k(k), 0, 0, \dots)$  az eredeti paraméterek vetítése a

$$\|v\|_\Gamma = \sum_{i,j=1}^{\infty} (v_i v_j \gamma_{i,j})^{1/2}$$

norma szerint.  $k$  rendű autoregressziót feltételezve  $y_1, \dots, y_n$  megfigyelések mellett a vetített paraméterek legkisebb négyzetes becslése

$$\hat{\alpha}(k) = (\hat{\alpha}_1(k), \dots, \hat{\alpha}_k(k), 0, 0, \dots).$$

Legyen  $X_t$  az eredeti folyamat egy független realizációja. Az egylépéses előrejelzés átlagos négyzetes hibája

$$Q_n(k) = E((\hat{X}_{t+1} - X_{t+1})^2 | y_1, \dots, y_n) - \sigma^2 \quad (3.7)$$

ahol  $\hat{X}_{t+1} = -\hat{\alpha}_1(k)X_t - \dots - \hat{\alpha}_k(k)X_{t-k}$ .

Tegyük fel, hogy a rendbecslés felső korlátja,  $K_n$  a mintaelemszám növekedésével végtelenbe tart. Jelölje  $L_n(k)$  a (3.7) várható értékét. Legyen  $k_n^*$  az a sorozat, melyre

$$L_n(k_n^*) = \min_{1 \leq k \leq K_n} L_n(k).$$

Ekkor  $\hat{k}$  rendbecslés aszimptotikusan hatásos, ha

$$\frac{Q_n(\hat{k})}{L_n(k_n^*)} \rightarrow 1 \text{ sztochasztikusan, ha } n \rightarrow \infty. \quad (3.8)$$

Shibata bebizonyította, hogy amennyiben  $\epsilon_t$  normális eloszlású, az Akaike kritérium illetve a vele aszimptotikusan ekvivalens módszerek rendelkeznek a fenti tulajdonsággal. Ugyanakkor a BIC és HQIC nem.

Az eredeti definíció szerint a rend- és paraméterbecslés, valamint az előrejelzés két független mintából történik. Ing és Wei megvizsgálták azt az esetet, amikor ugyanazon realizációt használjuk fel mindhárom lépésben.<sup>3</sup> Az AIC ezen feltételek mellett is aszimptotikusan hatásos marad.

0 várható értékű nem Gauss zaj feltételezése esetén Karagrigoriou bizonyítja az AIC-típusú kritériumok aszimptotikus hatásosságát.<sup>4</sup> Ellenben a

<sup>3</sup>lásd Ing-Wei (2005)

<sup>4</sup>lásd Karagrigoriou (1997)

$h$ -lépéses ( $h > 1$ ) előrejelzésen alapuló hasonló definíció értelmében az AIC nem aszimptotikusan hatásos.<sup>5</sup>

Shibata eredménye azon a feltételezésen alapul, hogy a szóbajövő modellek "nem modellezik túl jól" a valódi adatgeneráló folyamatot. Konzisztens modellszelekció esetén a rendbecslés tipikusan nem lesz aszimptotikusan hatásos. Véges dimenziós folyamatok esetében ugyanakkor éppen az ellenkezője igaz. A konzisztens BIC és HQIC pontonként hatásos, míg az AIC típusú kritériumok nem.

Az eddig bemutatott elméletek azt sugallják, hogy ha a valódi eloszlást végtelen dimenziósra feltételezzük, érdemesebb a konzervatív AIC-t, míg véges dimenziós esetben a konzisztens BIC-t vagy HQIC-t alkalmaznunk. A feladat azonban korántsem ilyen egyszerű. A fenti eredmények (beleértve a konzisztenciát is) ugyanis csak pontonként értendők, a konvergencia nem egyenletes a paramétertéren.

Végtelen dimenziós modell véges mintából történő becslése esetén előfordulhat, hogy egy véges modell elég jól közelíti a vizsgált folyamatot, ekkor az aszimptotikus pontonkénti eredmények irrelevánssá válnak. Ellenben véges dimenzió feltételezése mellett, mint látni fogjuk, kis paraméterek esetén a konzervatív becslések hajlamosak alulbecsülni a rendet, ami rossz poszt-modellszelekciós becsléshez vezet.

---

<sup>5</sup>lásd Bhansali (1997)

## 4. fejezet

# Empirikus vizsgálatok

Idősorok rendbecslésével kapcsolatos kutatások egy tekintélyes része szimulációs vizsgálatokra épül. Ezen tanulmányok célja általában több szelekciós módszer összehasonlítása abból a szempontból, hogy különböző modelltípusok mellett mennyiben becsülik jól a valódi rendet.<sup>1</sup>

A szelekciós kritériumokkal kapcsolatos elméleti eredményeket bemutattuk az előző fejezetben. Ezen elméletek a kritériumok aszimptotikus tulajdonságaira koncentrálnak. A jelen fejezet célja egy átfogó képet adni az információs kritériumok viselkedéséről a gyakorlatban (véges minta esetén), kiemelve azok jó tulajdonságait és hiányosságait.

Empirikus elemzések azt mutatják, hogy a valódi rend detektálásának valószínűsége sok tényezőtől függ: milyen hosszú a szimulált idősor, mekkora a valódi rend, milyenek a paraméterek. Minden tényezőt nehéz egyszerre vizsgálni, így mi elsősorban a paraméterekre helyezzük a hangsúlyt.

A továbbiakban a szimulációk futtatásához az R statisztikai programcsomag beépített moduljait használtuk.

### 4.1. Autoregresszív folyamatok modellezése

A modellszelekció témakörében a lineáris folyamatok rendbecslése a legintenzívebben kutatott téma. Ide sorolhatjuk a lineáris regressziót valamint az autoregresszív folyamatokat. Ez utóbbi esetében a rendbecslés célja az autoregresszív egyenletben szereplő legnagyobb késleltetés meghatározása.

---

<sup>1</sup>lásd pl. ARMA modellekre Sen-Shitan (2002), AR modellekre Chic (2002)

Az információs kritériumok AR(p) folyamat esetén az alábbi általános formába írhatók:

$$IC(p) = -2 \ln f_Y(y, \hat{\theta}^k) + C(n) p. \quad (4.1)$$

A rendbecslési eljárás eredetileg kötött a maximum likelihood paraméterbecsléshez. Azaz  $\hat{\theta}^k$  a paraméterek ML becslése. Normális eloszlású zajt feltételezve az általános alak

$$IC(p) = \ln \hat{\sigma}^2 + \frac{C(n)}{n} p, \quad (4.2)$$

ahol  $\hat{\sigma}^2$  a reziduális variancia ML becslése.

A gyakorlatban a maximum likelihood becslés kevésbé kedvelt, mivel elég lassú és gyakran merülnek fel konvergencia-problémák az optimalizálásakor. Autoregressziós folyamat modellezésekor helyette a vele aszimptotikusan ekvivalens Yule-Walker egyenleteket vagy a Burg algoritmust alkalmazhatjuk. Kis mintára a Yule-Walker becslés erősebben torzíthat, viszont túlparameterezett modellek illesztésekor egyértelműen jobb eredményt ad. <sup>2</sup>

Autoregresszió esetén - ellentétben a regressziós folyamatokkal - egy új megfigyelés nem független a korábbiaktól. A reziduálisok kiszámításához szükségünk van legalább annyi elemre, amennyi a becslés rendje. Fontos tehát előre rögzíteni, hogy az  $n$  megfigyelésből mennyit tekintünk relevánsnak, ez legyen  $m = n - \nu$ . A maximális lehetséges rendet jelölje  $K$ .  $\sigma^2$  becslésére több lehetőségünk is van:

$$\hat{\sigma}_k^2 = \frac{1}{\tau} \sum_{\nu+1}^n \hat{e}_t^2(k)$$

ahol  $\tau$  lehet  $n$ ,  $n - k$  vagy  $n - K$  és  $\nu$  lehet  $k$  vagy  $K$ .

A rendbecslés érzékeny  $\sigma^2$  ezen felírásának módjára. Ng és Perron szimulációs vizsgálatokkal igazolják, hogy robusztusabb becslést kapunk, amennyiben a releváns paraméterszám konstans, vagyis  $m = n - K$ .  $\tau$  értékének a mintaelemszámot választhatjuk. <sup>3 4</sup>

<sup>2</sup>lásd Chan et al. (1993)

<sup>3</sup>lásd Ng-Perron (2001)

<sup>4</sup>Egyéb megfontolásokról lásd pl. Basci-Zaman (2002). Ha az  $\hat{y}$  becsléséhez csak az idősor korábbi elemeit vesszük figyelembe, és a reziduálisok szórásbecslését ez alapján végezzük, a kritériumok más tulajdonságot mutatnak.



A paraméterek becsléséhez AR folyamatokra a Yule-Walker egyenleteket használjuk, míg a reziduális szórásnégyzet becslését az R beépített függvénye alapján számoljuk.<sup>5</sup>

#### 4.1.1. Alkalmazott becslési módszerek

Az információs kritériumok AR( $p$ ) folyamat esetén rendre

$$AIC(p) = \ln \hat{\sigma}_p^2 + \frac{2}{n}p \quad (4.3)$$

$$AIC_c(p) = \ln \hat{\sigma}_p^2 + \frac{2}{n-k-1}p \quad (4.4)$$

$$BIC(p) = \ln \hat{\sigma}_p^2 + \frac{\ln n}{n}p \quad (4.5)$$

$$HQIC(p) = \ln \hat{\sigma}_p^2 + \frac{2 \ln \ln n}{n}p \quad (4.6)$$

ahol  $p$  a legnagyobb késleltetést jelöli.

A bemutatott módszereken túl két másik kritériumot is vizsgálunk összehasonlításképp. Egyik legegyszerűbb rendbecslési eljárás a parciális autokorreláció függvény,  $\rho(m)$  vizsgálata. Tudjuk, hogy  $\rho(m) = 0 \forall m > p$ , ebből az alábbi kritérium származtatható:

Becsüljük a parciális autokorrelációt a Yule-Walker egyenletek segítségével. A kiszámításához alkalmazhatjuk pl a Durbin-Levinson algoritmust.<sup>6</sup>  $m > p$  esetén  $\hat{\alpha}_m$  közelítőleg normális eloszlást követ 0 várható értékkel és  $\sqrt{n}$  szórással. Legyen tehát  $\hat{p}$  az a legkisebb egész, amire

$$|\hat{\rho}(m)| < 1.96n^{-1/2}$$

minden  $m > \hat{p}$ -re.

A másik módszer a szintén Akaike nevéhez fűződő végső előrejelzési hiba (továbbiakban FPE) kritérium. Az FPE nem más, mint az egy lépéses előrejelzés átlagos négyzetes hibájának (MSE) becslése egy az eredeti megfigyelésektől független hipotetikus adatsorra (kereszt-kiértékelés). A paramé-

<sup>5</sup>Idősorok szimulálásakor ún. beégetési periódus előzi meg a tényleges idősort. A  $\sigma^2$  becslését általában ennek felhasználásával számítják.

<sup>6</sup>lásd Brockwell-Davis 8. fejezet

terbecslést az eredeti adatokból végezve az előrejelzés MSE-je

$$\begin{aligned} & E(Y_{n+1} - \hat{\theta}_1 Y_n + \dots + \hat{\theta}_p Y_{n-p})^2 = \\ & = E(Y_{n+1} - \theta_1 Y_n + \dots + \theta_p Y_{n-p} - (\hat{\theta}_1 - \theta_1) Y_n + \dots + (\hat{\theta}_p - \theta_p) Y_{n-p})^2 = \\ & = \sigma^2 + E((\hat{\theta} - \theta)^T (Y_{n+1-i} Y_{n+1-i})_{i,j=1}^p (\hat{\theta} - \theta)) \end{aligned}$$

Írjuk át az utolsó tagban szereplő várható értéket a megfigyeléseinkre vonatkozó feltételes várható értékre. Ekkor a két minta függetlensége miatt

$$E(Y_{n+1} - \hat{\theta}_1 Y_n + \dots + \hat{\theta}_p Y_{n-p})^2 = \sigma^2 + E((\hat{\theta} - \theta)^T \Gamma_p (\hat{\theta} - \theta))$$

ahol  $\Gamma_p$  az autokovariancia mátrix.  $n^{1/2}(\hat{\theta} - \theta)$  eloszlása aszimptotikusan normális,  $\sigma^2 \Gamma_p^{-1}$  szórással, amelyből az alábbi közelítés adódik

$$E(Y_{n+1} - \hat{\theta}_1 Y_n + \dots + \hat{\theta}_p Y_{n-p})^2 \approx \sigma^2 + \sigma^2 \frac{p}{n}.$$

$\sigma^2$ -et helyettesítsük a  $\frac{n\hat{\sigma}^2}{n-p}$  becsléssel, ahol  $\hat{\sigma}^2$  a maximum likelihood becslés. Így adódik a kritérium végső alakja:

$$FPE(p) = \hat{\sigma}^2 \frac{n+p}{n-p}. \quad (4.7)$$

#### 4.1.2. AR(1) folyamat rendbecslése

Az autoregressziós egyenletnek

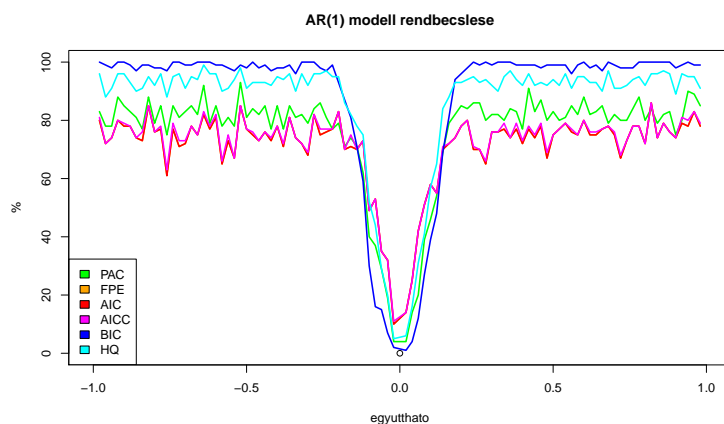
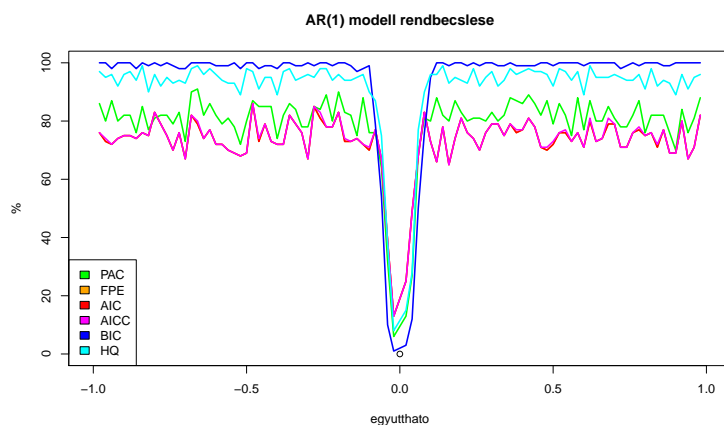
$$Y_t - \theta_1 Y_{t-1} - \dots - \theta_p Y_{t-p} = \epsilon_t \quad (4.8)$$

pontosan akkor létezik egyértelmű (jövőtől független) stacionárius megoldása, ha a karakterisztikus polinom

$$P(x) = x^p - \theta_1 x^{p-1} - \dots - \theta_p \quad (4.9)$$

komplex gyökei az egységkörön belül helyezkednek el.

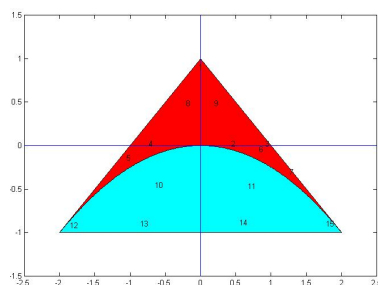
Az AR(1) folyamat esetében ez nyilvánvalóan a  $(-1, 1)$  nyílt intervallum. A paramétereket  $-0.98$ -tól  $0.02$ -es lépésenként változtatva  $0.98$ -ig 100-100 szimulációt futtattunk. A maximális rend 5, a szimulált idősor hossza pedig 500 illetve 2000. Az alábbi két ábra a helyes rendbecslések arányát mutatja százalékban.

4.1. ábra. Helyes rendbecslések aránya  $\theta$  függvényében,  $n = 500$ 4.2. ábra. Helyes rendbecslések aránya  $\theta$  függvényében,  $n = 2000$ 

Az origótól távolabb egyértelműen a BIC teljesít legjobban (közel 100%), utána a HQIC, majd a parciális autokorreláció. Az AIC, AICc és FPE mindössze 70% körül mozog. A 0 körül azonban az ellenkezőjét figyelhetjük meg. Kis paraméterek mellett a BIC nagy eséllyel alulbecsül, így a HQIC és az AIC jobb teljesítményt nyújtanak, habár a 0 környékén egyik sem éri el a 20%-ot. Mivel a BIC konzisztens, a mintaelemszám növekedésével ennek az intervallumnak zsugorodnia kell. Ezt támasztja alá a második ábra.

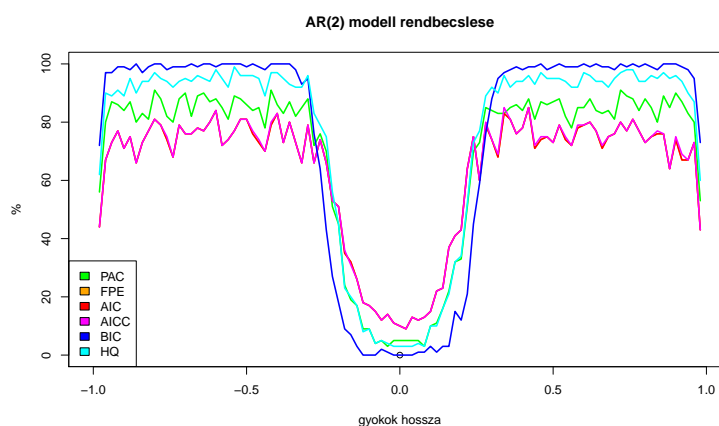
### 4.1.3. AR(2) folyamatok becslése

Az AR(2) egyenletnek egyértelmű stationárius megoldása a két paraméter függvényében az alábbi pontok mellett adódik:



4.3. ábra. AR(2) folyamat stationárius ponthalmaza

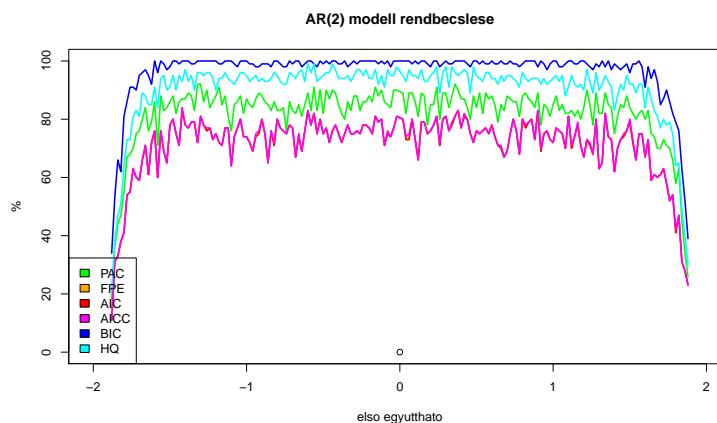
A görbe alatti terület annak az esetnek felel meg, amikor a (4.9) karakterisztikus polinom gyökei komplex számok. A valódi rend detektálásának valószínűségét itt a komplex régióba eső parabolák mentén vizsgáltuk. Vagyis a konjugált gyökök hosszát változtattuk. A futások száma most is 100 volt minden paraméterre, az idősor hossza 2000, a maximális rend 5. A 4.4 ábra azt az esetet ábrázolja, mikor a gyökök szöge  $-30$  és  $30$  fok. A gyökök hosszát most is  $0.02$  léptékkal vizsgáltuk  $-0.98$  és  $0.98$  között.



4.4. ábra. Helyes rendbecslések aránya a gyökök hosszának függvényében

Vizsgáltuk még a  $60$ ,  $90$  és  $0$  fokos esetet is, és azt tapasztaltuk, hogy

a kritériumok a stacionárius régió ábráján 12 illetve 15-tel jelölt területeken gyengébben szerepelnek. Ennek valószínűleg az az oka, hogy közel vagyunk a stacionaritási terület határához. Megnéztük még továbbá, hogyan alakulnak a becslések, ha vízszintes egyenesek mentén vizsgálódunk, vagyis a nagyobbik késleltetés paraméterét rögzítjük. A 4.5. ábra a  $-0.9$ -es paraméter esetét ábrázolja a már ismert felbontással.



4.5. ábra. Helyes rendbecslések aránya  $\theta_1$  függvényében,  $\theta_2 = -0.9$

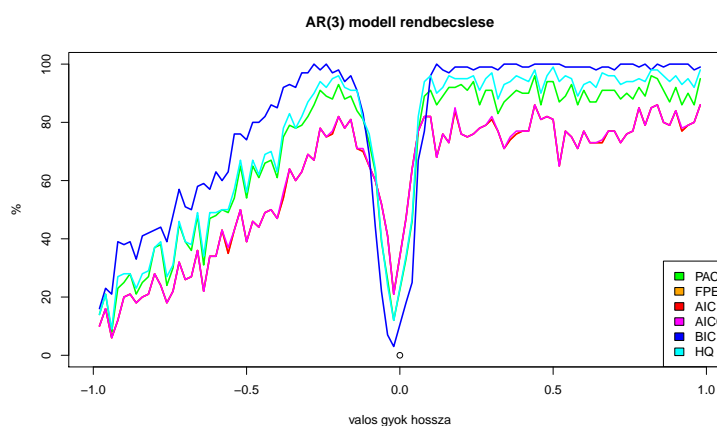
Általánosságban most is a BIC teljesít legjobban, kivéve amikor a komplex gyökök hossza 0-hoz közelít. Ekkor a stacionaritási területen közelítünk az origóhoz, vagyis a paraméterek alacsonyok. A második ábra azt sugallja, hogy a kritériumok nem érzékenyek az első késleltetés paraméterére. Megvizsgáltuk még a  $\theta_2 = 0.1$  és  $\theta_2 = -0.05$  egyeneseket is, ahol szintén ezt tapasztaltuk.  $\theta_2 = 0.1$  mellett a helyes becslések aránya a BIC és a HQIC esetén 90% körüli, míg az AIC 75%. A  $\theta_2 = -0.05$  egyenes mentén viszont az AIC és a HQIC átlagosan 60%, amíg a BIC 40% alatti.

A tapasztalataink alapján azt mondhatjuk, hogy nagy minta esetén a konzisztens rendbecslések jól teljesítenek, kivéve amikor a legnagyobb késleltetés paramétere alacsony. Hogy ezen paraméterre mennyire érzékenyek, az függ a mintelemszámtól. Az Akaike, AICc és FPE kritériumok ezzel ellentétben hajlamosak a túlbecslésre. A helyes becslések aránya sosem érte el a 80%-ot és rendszerint alulmaradnak a parciális autokorreláció vizsgálatára alpozott kvázi-intuitív kritériummal szemben.

Az AIC, AICc és FPE gyakorlatilag ugyanazt az eredményt adták mind-egyik futásnál. Ez korántsem meglepő, hiszen a három kritérium aszimptotikusan ekvivalens, ha a lehetséges modelleink halmaza kevés elemből áll, és a paraméterek száma kicsi a mintához képest.

Chik 2002-es cikkében szintén autoregresszív folyamatokat vizsgál 1-5 valódi rend mellett kis és közepes mintákra. Eredményei a fentiekhez hasonlóak: a BIC és HQIC kritériumok jobban teljesítenek, azonban AR(5) folyamatokra gyakorlatilag egyik becslési módszer sem bizonyul kielégítőnek.

Magasabb rendű folyamatok esetére végezetül álljon itt egy elgondolkodtató ábra. AR(3) modelleket futtattunk a karakterisztikus polinom gyökei szerint paraméterezve. A gyökök szöge 135, -135 és 0 fok. A hosszuk kezdetben 0.98, majd a szokásos léptékkal csökkentjük a valós gyök hosszát -0.98-ig.



4.6. ábra. Helyes rendbecslések aránya a valós gyök hosszának függvényében

## 4.2. Rendbecslés ARMA folyamatok esetén

Autoregresszív mozgóátlag folyamatok rendbecslése hasonló az autoregressziós modellekéhez, itt azonban két paramétervektor dimenzióját kell becsülnünk, tehát a minimalizálást két érték szerint végezzük. Célunk a helyes  $(p,q)$  pár detektálása.

Az információs kritériumok ARMA(p,q) rendbecsléshez

$$AIC(p, q) = \ln \hat{\sigma}_{p,q}^2 + \frac{2}{n}(p + q) \quad (4.10)$$

$$AIC(p, q) = \ln \hat{\sigma}_{p,q}^2 + \frac{2}{n - p - q - 1}(p + q) \quad (4.11)$$

$$BIC(p, q) = \ln \hat{\sigma}_{p,q}^2 + \frac{\ln n}{n}(p + q) \quad (4.12)$$

$$HQIC(p, q) = \ln \hat{\sigma}_{p,q}^2 + \frac{2 \ln \ln n}{n}(p + q) \quad (4.13)$$

ahol  $p$  a legnagyobb késleltetés,  $q$  pedig a mozgóátlagolás rendje.

ARMA(p,q) modellek esetén a paraméterbecslés valamint a reziduálisok kiszámítása összetettebb algoritmusokat igényel. <sup>7</sup> A hosszú futásidő miatt csak ARMA(1,1) modelleket vizsgáltunk különböző AR és MA koefficiensek mellett. A maximális AR és MA rend egyaránt 3.

Kizárólag kauzális és invertálható modelleket vizsgáltunk, hogy a stacionaritási feltétel ne sérüljön. Egy ARMA(p,q) folyamat

$$Y_t + \theta_1 Y_{t-1} + \dots + \theta_p Y_{t-p} = \epsilon_t + \phi_1 \epsilon_{t-1} + \dots + \phi_q \epsilon_{t-q} \quad (4.14)$$

pontosan akkor kauzális és invertálható, ha a

$$P(x) = x^p + \theta_1 x^{p-1} + \dots - \theta_p \quad (4.15)$$

$$Q(x) = x^q + \phi_1 x^{q-1} + \dots + \phi_q \quad (4.16)$$

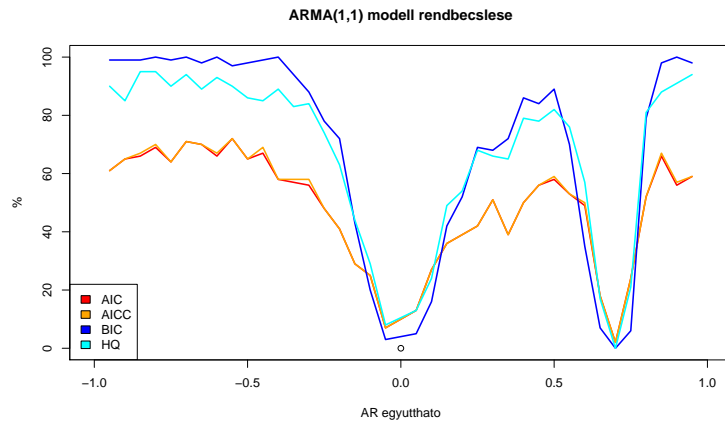
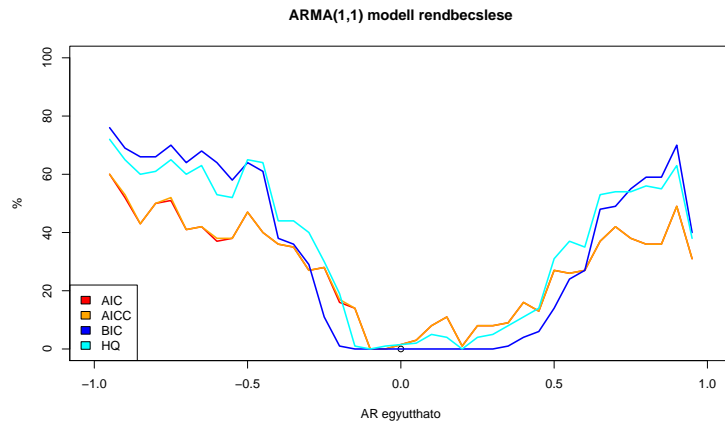
polinomok gyökei mind az egységkörön belül vannak.

A fenti állításnak feltétele, hogy a  $P(x)$  és  $Q(x)$  polinomoknak ne legyen közös gyöke. Amennyiben ez nem áll fenn, és a közös gyökök mind az egységkörön belül helyezkednek el, az ARMA egyenletet gyakorlatilag egyszerűsíthetjük a közös tényezővel (ugyanaz a folyamat lesz a stacionárius megoldás).

Vizsgáljuk most az ARMA(1,1) folyamat rendbecsléseit az AR paraméter függvényében, rögzített MA együttható mellett. A futások száma továbbra is 100, a generált idősorok hossza pedig 1000. Az MA koefficienseket 0.1-esével változtattuk. Az alábbi két ábra a  $\phi = 0.7$  illetve  $\phi = 0.2$  eseteket mutatja.

Mindkét esetben látszik, hogy a helyes becslések aránya kis AR paraméter mellett az AR(1) folyamathoz hasonlóan kicsi. Ugyancsak ezt tapasztaljuk,

<sup>7</sup>pl. innovációs algoritmus a paraméterek előzetes becsléshez, majd ezt követi a maximum likelihood becslés. Lásd bővebben Brockwell-Davis (1986)

4.7. ábra. Helyes rendbecslések aránya a  $\theta$  paraméter függvényében,  $\phi = 0.7$ 4.8. ábra. Helyes rendbecslések aránya a  $\theta$  paraméter függvényében,  $\phi = 0.2$ 

ha a két paraméter megegyezik. Ez ARMA(1,1) modellekre a fent említett közös gyökök esetének felel meg, tehát valójában itt egy ARMA(0,0) folyamatot generáltunk (vagyis egy Gauss zajt), és a kritériumok "alulbecsülik" az eredeti rendeket.

Az AR paraméter mellett az MA is hatással van a becslések jóságára. Hasonlóan  $\theta$ -hoz, ha  $\phi$  értéke alacsony, a helyes becslések aránya csökken, szembetűnőbben a konzisztens kritériumok esetén.

Sen és Shitan az AICc kritérium teljesítményét vizsgálták ARMA folya-



matok esetén. Tiszta autoregresszió vagy mozgóátlag esetében átlagosan 70% körüli volt a helyes becslések aránya, míg valódi ARMA folyamatokra ennél alacsonyabb. Ez egybevág a mi eredményeinkkel is. A konzisztencia tulajdonság ARMA modellek esetében sem egyenletesen értendő a paramétertérren. A konvergencia sebessége tehát függ a paramétereiktől, minél nagyobb mintát választunk, annál kisebb lesz az az intervallum, amelyben a BIC és HQIC kritériumok teljesítménye alacsony.

### 4.3. A félrespecifikálás esetei

Az információs kritériumok levezetéséből adódik, hogy azok aszimptotikusan torzítatlan becslései lesznek a Kullback-Leibler távolságnak - bizonyos feltételek mellett. A (2.8) alak nyilvánvalóan normális eloszlású AR modellekre érvényes, mégis gyakran alkalmazzák olyan esetekben is, ahol a zaj nem feltétlenül normális, avagy heteroszkedasztikus, vagyis ARCH folyamattal van dolgunk.

A dolgozat végéhez érve megmutatjuk milyen hibákhoz vezethet, ha az ismert kritériumokat nem az eredeti feltételek mellett használjuk.

#### 4.3.1. Autoregresszív folyamat nem Gauss zajjal

Tekintsük elsőként azt az esetet, amikor a folyamatot meghajtó zaj eloszlása a normális eloszláshoz képest vastag farkú. Az alábbi táblázatba egy  $\theta = -0.5$  paraméterű AR(1) folyamat szimulációjából kapott eredményeket gyűjtöttük össze. A rendbecslések eloszlását vizsgáltuk Gauss, t(1) illetve Laplace(0,4) eloszlás mellett.

	"N(0,1)"			"t(1)"			"Lap(0,4)"		
	AIC	BIC	HQ	AIC	BIC	HQ	AIC	BIC	HQ
1	75	99	98	90	97	95	74	99	98
2	8	1	2	4	0	1	9	1	1
3	3	0	0	1	1	1	7	0	0
4	5	0	0	0	0	0	2	0	0
5	6	0	0	0	2	3	4	0	0
6	3	0	0	5	2	3	4	0	0

4.1. táblázat. Becsült rendek aránya különböző zajok esetén

Az eredmény meglepő az Akaike kritériumra nézve, ugyanis  $t$  eloszlás mellett a kritérium számottevően jobban teljesít. Ez a jelenség megmarad, ha a paramétert 0.1-re illetve 0.9-re változtatjuk.

A konzisztens kritériumok mindhárom esetben hasonló eredményt adnak, ami nem váratlan. A 3. fejezetben már láttuk, hogy Hannan nem csak normális eloszlású zajoka igazolta a konzisztencia tulajdonságot.

### 4.3.2. Autoregresszív folyamat ARCH zajjal

Autoregresszív feltételes heteroszkedaszticitásról (ARCH) akkor beszélünk, ha az adott folyamatot meghajtó zaj feltételes szórása időben nem állandó.

$$Y_t = f + \epsilon_t$$

$$\epsilon_t = \sigma_t \nu_t$$

$$\sigma^2 = \alpha_0 + \alpha_1 \epsilon_{t-1}^2 + \dots + \alpha_p \epsilon_{t-p}^2 \quad (4.17)$$

A zajkomponens várható értéke és szórása továbbra is konstans marad, viszont a feltételes variancia autoregresszív folyamatot követ. Ez az idősor grájában úgy jelentkezik, hogy a változékonny és a kevésbé változékonny időszakok tömörülnek.

Vizsgáljuk meg, hogyan viselkednek az információs kritériumok, ha egy AR(1) folyamatot ARCH(1) zajjal hajtunk meg. Legyen  $\alpha_0 = 0.0001$  és  $\alpha_2 = 0.8$ . Az AR paraméternek három különböző értéket választottunk.

	"-0.5"			"0.1"			"0.9"		
	AIC	BIC	HQ	AIC	BIC	HQ	AIC	BIC	HQ
0	0	0	0	3	9	2	0	0	0
1	13	57	37	16	42	28	11	51	30
2	16	22	23	17	17	23	12	16	17
3	18	13	16	11	8	14	12	14	17
4	10	5	6	14	11	8	15	6	13
5	14	2	8	8	1	5	6	6	9
6	9	1	6	8	0	4	11	3	6
7	9	0	2	8	1	3	10	0	2
8	2	0	1	4	0	1	5	1	2
9	4	0	1	2	0	0	4	1	2
10	5	0	0	9	1	2	14	2	2

4.2. táblázat. Becsült rendek aránya AR-ARCH folyamat esetén

Látható, hogy mindhárom esetben mindhárom kritérium erősen szór a rendbecslés tekintetében.

Ellenőrzésképpen megnéztük, hogy ha ugyanezen folyamatokat a zaj kevert változatával hajtjuk meg, a rendbecslés javul. Nem meglepő, hiszen ekkor közönséges AR modellel van dolgunk.

### 4.3.3. Rezsimváltó modellek

Végezetül olyan rezsimváltó modelleket vizsgáltunk, melyben az autoregresszív paraméterek két állapotú Markov folyamatot követnek (Markov-Switching modell egy fajtája).

$$y_t = \theta_{S_t}^1 y_{t-1} + \dots + \theta_{S_t}^p y_{t-p}$$

ahol  $S_t$  irreducibilis stacionárius Markov lánc.

A valószínűség-átmenet mátrix:

$$\begin{pmatrix} 0.1 & 0.9 \\ 0.9 & 0.1 \end{pmatrix}$$

A két paramétert pedig változtattuk.

	"0.4,0.6"			"0.3,0.8"			"0.5,-0.5"		
	AIC	BIC	HQ	AIC	BIC	HQ	AIC	BIC	HQ
1	70	97	92	8	60	26	0	0	0
2	11	3	8	40	38	57	64	99	88
3	7	0	0	30	2	16	17	1	10
4	1	0	0	7	0	1	6	0	1
5	5	0	0	8	0	0	4	0	1
6	2	0	0	1	0	0	4	0	0
7	1	0	0	2	0	0	1	0	0
8	3	0	0	4	0	0	3	0	0

4.3. táblázat. Becsült rendek aránya különböző AR paraméterek mellett

Az Akaike kritérium az AR-ARCH modellhez hasonlóan szétesik. Viszont a BIC és HQIC, főleg az utolsó esetben határozottan félrebecsül.

Érdekességként megemlítjük, hogy a fenti rezsimváltó folyamatnak létezik gyenge ARMA reprezentációja, azaz elég általános feltételek mellett az autokovariancia függvény ARMA folyamatot követ. Zhang és Stine bizonyította <sup>8</sup>, hogy az autoregresszió és a mozgóátlagolás rendje egyaránt felülről

<sup>8</sup>lásd Zhang-Stine (1997)

becsülhető a rezsinváltások számával. Ha az ARMA reprezentáció rendje  $p'$  és  $q'$ , valamint a folyamat AR( $p$ ), változó paraméterekkel, a rezsinváltások száma pedig  $p$ , akkor  $p' \leq rp^2$  és  $q' \leq rp^2 - 1$ . Tetszőleges rendbecslési eljárást alkalmazva a tapasztalati autokovariancia-függvényre, becslést adhatunk a rezsinváltások számára (érdemes az alsó korlátot választani).

A tanulság tehát az, hogy óvatosan kell bánnunk a megismert információs kritériumokkal. Egy valós folyamat modellezésekor körültekintően kell eljárni a modell típusánk meghatározásakor. Ebben segítségünkre lehet a folyamat ábrája (ha a változékonyság lassan cseng le, akkor ARCH-csal lehet dolgunk), a tapasztalati autokovariancia és parciális autokovariancia függvények (AR és MA modellek) és még sok más módszer.

Az ARCH és rezsinváltó folyamatok rendbecslése még korántsem annyira kidolgozott, mint az ARMA vagy AR modelleké. Amennyiben jól alátámaszthatóan ARMA vagy AR folyamattal van dolgunk, érdemes több rendbecslési eljárást is alkalmazni, tekintve hogy a konzisztens kritériumok is adhatnak rossz becslést. A Hannan-Quinn kritérium amolyan kompromisszumnak látszik az AIC és a BIC között. Olyan paraméterekre, ahol a BIC jól becsül, a HQIC is közel olyan jó eredményt ad, viszont szélsőséges esetekben kevésbé viselkedik rosszul.

# Irodalomjegyzék

- [1] Akaike, H. (1981) : Likelihood of a Model and Information Criteria, Journal of Econometrics, 16, North-Holland Publishing Company, 3-14. o.
- [2] Basci, S. - Zaman, A. (1998) : Variance Estimates and Model Selection, Working Papers, Bilkent University, Department of Economics
- [3] Banshali, R. J. (1986) : Asymptotically Efficient Selection of the Order by the Criterion Autoregressive Transfer function, Annals of Statistics, 1/14, 315-325. o.
- [4] Banshali, R. J. (1997) : Direct Autoregressive Predictors for Multistep Prediction: Order Selection and Performance Relative to the Plug in Predictors, Statistica Sinica, 7/1997, 425-449 o.
- [5] Bierens, H. J. (2006) : Information Criteria and Model Selection, Pennsylvania State University, jegyzet
- [6] Brockwell, P. J. - Davis, R. A. (1986) : Time-series: Theory and Methods, Springer-Verlag
- [7] Chan, C. - Davis, R. A. - Brockwell, P. J. - Bai, Z. D. (1993) : Order Determination for Autoregressive Processes Using Resampling Methods, Statistica Sinica, 3, 481-500. o.
- [8] Chik, Z. (2002) : Performance of Order Selection Criteria for Short Time Series, Pakistan Journal of Applied Sciences, 2/7, 783-788. o.
- [9] Fitzgibbon, L. J. - Dowe, D. L. - Vahid F. (2004) : Minimum Message Length Autoregressive Model Order Selection, Proceedings of International Conference on Intelligent Sensing and Information Processing

- [10] Hannan, E. J. - B. G. Quinn (1979): The Determination of the Order of an Autoregression, *Journal of the Royal Statistical Society*, 41, 190-195 o.
- [11] Hannan, E. J. (1980): The Estimation of the Order of an ARMA Process, *Annals of Statistics*, 8, 1071-1081 o.
- [12] Ing, C-K. - Wei C-Z. (2005) : Order Selection for Same-realization Predictions in Autoregressive Processes, *The Annals of Statistics*, 5/33, 2423-2474 o.
- [13] Karagrigoriu, A. (1997) : Asymptotic Efficiency of the Order Selection of a Nongaussian AR Process, *Statistica Sinica*, 7/1997, 407-423 o.
- [14] Kearns, M. - Mansour, Y. - Ng, A. Y. - Ron, D. (1995) : An Experimental and Theoretical Comparison of Model Selection Methods, *Proceedings of the Eighth Annual ACM Conference on Computational Learning Theory*
- [15] Leeb, H. - Pötscher, B. M. (2000) : The Finite-Sample Distribution of Post-Model-Selection Estimators, and Uniform Versus Non-Uniform Approximations, *Econometric Theory* 19, 100-142. o.
- [16] Leeb, H. - Pötscher, B. M. (2008) : Model Selection, In *The Handbook Of Financial Time Series*. Springer, New York, 785-821 o.
- [17] Liavas, A.P. - Regalia, P. A. (2001) : On the Behavior of Information Theoretic Criteria for Model Order Selection, *IEEE Transactions on Signal Processing*, 8/49
- [18] Lu, S. - Ju, K. H. - Chon, K. H. (2001) : A New Algorithm for Linear and Nonlinear ARMA Model Parameter Estimation Using Affine Geometry, *IEEE Transactions on Biomedical Engineering*, 10/48
- [19] Mallows, C. L. (2000) : Some Comments on Cp, *Technometrics*, 42/1
- [20] Ng, S. - Perron, P. (2001) : A Note on the Selection of Time Series models, *Boston College Working Papers in Economics*
- [21] Sen, L. K. - Shitan M. (2002) : The Performance of AICc as an Order Selection Criterion in ARMA Time Series Models, *Pertanika Journal of Science and Technology*, 10/1, 25-33 o.

- [22] Spiegelhalter, D. J. - Best, N. G. - Carlin, B. - van der Linde, A. (2002) : Bayesian Measures of Model Complexity and Fit, Journal of the Royal Statistical Society, 4/64, 583-639 o.
- [23] Stoica, P. - Moses, R. L. (2005) : Spectral Analysis of Signals, Prentice Hall, 376-397 o.
- [24] Schwarz, G. (1978) : Estimating the Dimension of a Model, Annals of Statistics, 6, 461-464.
- [25] Tulassay Zs. (2007) : ARCH/GARCH modellek, Budapesti Corvinus Egyetem, Gazdálkodástudományi kar, jegyzet
- [26] Zhang, J. - Stine, R. A. (1997) : Autocovariance Structure of Markov Regime Switching Models and Model Selection, Journal of Time Series Analysis, 2/22, 107-124 o.