

Idősorok alapvető tulajdonságai	Olvasási idő: 20 perc 	Készítette: Dr. habil. Kiss Gábor Dávid 
--	---	---

Bevezetés

- Az idősorok olyan adat vektorok, amelyek az idő függvényében változnak: $P = (p_1, p_2, \dots, p_t)$
- Az idősor felbontása jellemzően napi, heti, havi, negyed éves és éves. A napi felbontási idősorok mutatják a legtöbb anomáliát az alapmodellhez képest, míg az éves idősorok elég jól közelítik azt.
- Alapmodell: az idősor kövessen normális eloszlást, ne legyen autokorrelált, legyen homoszkedasztikus és gyengén stacioner.
- Ebben a jegyzetben a makrogazdaság és a pénzügyi piacok által generált idősorok elemzésével foglalkozunk.
- A makro- vagy pénzügyi input idősorok (P) regressziós elemzése során elvárjuk, hogy a várható értékük nulla legyen, és hasonló skálán mozogjanak. Ezt az alábbi módokon érhetjük el:
 - differenciált (hozam): $r_t = (P_t - P_{t-1})$ vagy **$r = \text{diff}(P)$** ;
 - A skálázásban segíthet a logaritmizálás a $P > 0$ esetekben, ekkor használhatjuk a logaritmikus differenciáltakat (pl.: tőzsdei árfolyamok, nyersanyag árak, GDP). Jellemzően olyan esetekben, amikor az idősor egyébként is hajlamos exponenciális változásokra: $r_t = (\ln(P_t) - \ln(P_{t-1}))$ vagy **$r = \text{diff}(\log(P))$** ;
 - Amennyiben az idősor felvehet negatív értékeket is és a várható értéke nulla $E(P)=0$, akkor a z-score standardizálás lehet célravezető (pl.: fizetési mérleg statisztikák, kötvényhozamok): $x = \frac{x - \mu}{\sigma}$

Centrális momentumok

- k-adik centrális momentum: $E[(x - \mu)^k]$
- Első momentum a várható érték: átlag, módusz, medián.
 - $\mu = E[(x - \mu)]$
 - Hasonlónak és zérusnak kell lenniük ideális esetben.
 - Ők jelentik a várható értéket.
- Második momentum a szórás (volatilitás).
 - $\sigma = E[(x - \mu)^2]$
 - Az idősor eltérését mutatja a várható értékhez képest.
 - A közgazdaságtanban a kockázatot a számszerűsíthető bizonytalanságok halmazaként definiálják, amelyet jellemzően a szórással lehet társítani.
 - Ehhez hasonlóan, a monetáris politika makrogazdasági hatásainak értékelése során az egyes rezsimek makrogazdasági változók (pl. GDP, kamatláb, infláció) szórását vizsgálják.

- A szórás csak normális eloszlás fennállása mellett írható le véges számként.
- Harmadik momentum: ferdeség vagy aszimmetria (skewness).
 - $\mu_3 = E[(x - \mu)^3]$
 - Ha az idősor hasonló tömegben változik negatív és pozitív irányban, akkor szimmetrikusnak tekinthető.
 - Ha egy deviza jellemzően gyengül, akkor ott nagyobb tömegben jelennek meg csökkenő értékek, így az aszimmetria negatív.
- Negyedik momentum: csúcsosság (kurtosis).
 - $\mu_4 = E[(x - \mu)^4]$
 - Amíg az adatgeneráló folyamat normális eloszlást (véletlent) követ, addig a kurtózisa 3 lesz. Ebben az esetben az eloszlás szélein exponenciális lecsengést tapasztalunk – azaz eleve kis tömegben találunk itt adatokat.
 - Háromnál magasabb kurtózis esetén áll fenn a vastagfarkúság (leptokurtózis, heavy vagy fat tailness) esete, ekkor az extrémnek tekinthető elmozdulások jóval nagyobb tömegben és sokkal komolyabb elmozdulásokat valószínűsíthetően jelennek meg.

Forrás

Greene (2003): pp. 848.

Valószínűségi eloszlások

- Minden mért (tapasztalati) idősor esetében feltételezünk valamilyen adatgeneráló folyamatot a háttérben, amelynek viszont szüksége van valamilyen elméleti eloszlásra¹.

i. Normális eloszlás

Amikor egy kísérlet kimenetelét nagyszámú, egymástól független vagy csak kevésbé függő véletlen tényező határozza meg úgy, hogy az egyes tényezők külön-külön csak kis mértékben járulnak hozzá az összes véletlen hatásából eredő ingadozásokhoz, továbbá az egyes tényezők hatásai egyszerűen összeadódnak, akkor $X \sim N(\mu, \sigma^2)$ **normális (Gauss) valószínűségi eloszlás** lép fel, melynek sűrűségfüggvénye:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}},$$

ahol σ és μ állandóak, végesek és a normális eloszlás paramétereinek nevezzük őket, valamint $\sigma > 0$ (Rényi 1972, Coles 2001, Brockwell 1996).

A sűrűségfüggvény a normál eloszlás esetében gyorsan nullához tart, így az ilyen eloszlásból történő mintavétel során az elemek nagy része a várható értékek közelébe esik.

A normál eloszlás központi (centrális) határeloszlás-tétele kimondja, hogy $x_1, x_2, \dots$ kellően nagyszámú (n) független és azonos eloszlású (*independent and identically distributed - iid*) véletlen változó (melyek közös várható értéke $E(x)$ és közös szórása $D(x)$) véges $x_1 + \dots + x_n$ összeg standardizáltja megközelítőleg normál eloszlást követ.

Az elméleti és gyakorlati pénzügyek eszköztára az elmúlt évtizedekben a hozamok normál eloszlásának feltételezésére épült, ide tartozik többek között a klasszikus portfólió elmélet, a Black-Scholes-Merton-féle opcióárazás, a RiskMetrics variancia-kovariancia eljárása. Ezen elméletek alkalmazását a központi

¹ További érdekességek az eloszlásokról: <https://www.analyticsvidhya.com/blog/2017/09/6-probability-distributions-data-science/>

határeloszlás tétele tette lehetővé, azon feltételezés szerint, hogy a pénzügyi hozamok számos információ és egyéni döntés eredményeként jönnek létre (Dunbar 2002).

Statisztikai szempontból sajátos eredményre vezethet, amennyiben a hatékony piacok elmélete (*efficient market hypothesis*) kapcsán kizárólag Fama 1970-es cikkének fő mondanivalójára hagyatkozunk – ahol a piaci szereplők informáltsága alapján vezette le a gyenge-közepes-erős hatékonyság eseteit. Ez esetben ugyanis a gyenge hatékonyság elvetéséhez is bőven elegendő lenne a hozamok autokorreláltságának kimondása. A cikket elolvasva azonban egyértelmű, hogy Fama az autokorrelálatlanság megkövetelésével csupán kiegészítette a bolyongás-Markov-folyamat-normál-eloszlás gondolati körét, miután a 384. oldalon előbb „fair játék” szükségességét (egyensúlyi várható érték (*equilibrium expected return*) körül ingadozó hozamok bevezetése) mutatja be. Ezt követően a 386. oldalon vázolja a hozamok szubmatringál-jellegét (egy eszköz várható hozama legyen nagyobb vagy egyenlő nullánál – ami nullánál nagyobb esetben a játékos szempontjából „kedvező” játékot takar), illetve bemutatja a bolyongás modelljét – és csak ezt követően fogalmazza meg a hatékonyság különböző formáihoz kötődő piaci feltételeket. A modell tesztelése során többek között például külön kitér a hozamok eloszlásának kérdésére a 399. oldalon. A tőkepiaci hatékonyság kapcsán gyakran hivatkozott **Fama-féle modell** tehát **pusztán kiegészítette** a már meglévő, **piaci hatékonyságot vizsgáló modelleket**. Ezért fordulhat elő, hogy az ökonometriával foglalkozó irodalom a hatékony piacokat automatikusan összekapcsolja a bolyongással (lásd például Alexander (2008) 213. oldal), vagy feltételezi, hogy a vizsgált idősorokat létrehozó sztochasztikus folyamatok mögött kizárólag a véletlen áll (Lütkepohl – Kratzig 2004), esetleg kimondja, hogy a gazdasági folyamatok véletlen folyamatok eredményeként jönnek létre².

ii. Vastagfarkú eloszlások

A tőkepiacokon tapasztalható hozamok normálistól eltérő eloszlása régóta ismert tény szakirodalomban (Borak és mtsai. 2005, Tsay 2005, Dávid 2009). A vastagfarkúságból fakadó problémák súlyosságát leginkább az 1987-es és 1998-as részvénytőkepiaci válságokkal szokás illusztrálni, ahol a hozamok normál eloszlására vonatkozó feltevés okozott több milliárd dolláros veszteségeket – majd a Scholes és Merton nevével fémjelzett Long-Term Capital Management hedge-fund bukását (Dunbar 2002).

A logaritmikus hozamok empirikus eloszlására sokkal inkább illeszthető **hatvány- avagy Pareto eloszlás**, függetlenül a piac típusától, tér- és időbeli karakterisztikáitól Molnár (2006), Gabaix és mtsai. (2003) és Clauset és mtsai. (2009), valamint Jentsch és mtsai. (2006) szerint a hozamok hatványeloszlásának megléte a skálafüggetlen hálózatok jelenlétéről tanúskodik. A hatványeloszlások (*power law distribution*)³ sűrűségfüggvényét Newman (2005) nyomán az alábbi módon fejezhetjük ki: egy folytonos valós változó $p(x)$ dx valószínűséggel veszi fel az x és $x+dx$ intervallum értékeit, ahol $p(x)=Cx^{-\alpha}$ lesz, $\alpha>0$ farokexponens esetén. A x és $x+dx$ intervallum használatából következik, hogy a hatványeloszlások elsődlegesen egy adott eloszlás farkain értelmezhetők – ahogyan ezt a fenti cikk szerzője nyomatékosan ki is emeli.

Borak és mtsai. (2005) szerint emellett még a **stabilis eloszlások** alkalmasak a vastagfarkúság kezelésére. Egy X valószínűségi változót akkor tekintünk α -stabilis eloszlásúnak, amennyiben tetszőleges $X_1, X_2, \dots, X_n$ független, X eloszlású valószínűségi változók esetén létezik c_n és d_n konstans, hogy $X_1+\dots+X_n$ eloszlása c_nX+d_n eloszlásával egyezik meg, és $c_n=n^{1/\alpha}$.

² lásd például: Greene, W. H. 2003: *Econometric Analysis*. Prentice Hall, Upper Saddle River 845. oldalán,

³ A “power-law distribution” fogalmat magyarul egyaránt illetik hatvány-eloszlás illetve hatványtörvény-eloszlás névvel is. Munkám során az előbbit használom.

A szimmetrikus α -stabilis eloszlások túlélési függvényeinek aszimptotikája az alábbi (21):

$$\begin{cases} \lim_{x \rightarrow \infty} x^\alpha P(X > x) = C_\alpha(1 + \beta)\sigma^\alpha \\ \lim_{x \rightarrow -\infty} x^\alpha P(X < -x) = C_\alpha(1 - \beta)\sigma^\alpha \end{cases} \quad (21)$$

ahol $C_\alpha = (2 \int_0^\infty x^{-\alpha} \sin(x) dx)^{-1} = \frac{1}{\pi} \Gamma(\alpha) \sin \frac{\pi\alpha}{2}$, valamint:

- $\alpha \in (0, 2]$ jelöli a farok exponenst (stabilitási index, karakterisztikus exponens). A függvény $\alpha = 2$ esetén normál eloszlást ír le. A p -edik momentuma egy stabil random változónak csak akkor véges, ha $p < \alpha$. $\alpha > 1$ esetén a várható érték véges,
- $\beta \in [-1, 1]$ ferdeségi paraméter nullánál magasabb értéke jelöli az eloszlás jobb oldalán jelentkező hosszabb farkat,
- $\sigma > 0$ skála paraméter az eloszlás szélességét határozza meg.

A stabilis eloszlások részvénytársi, kötvénypiaci, devizapiaci, ingatlanpiaci és nyersanyagpiaci idősorok hozamaira történő illeszthetősége kétségtelenül jobb, mint a normál eloszlásé, azonban Borak és mtsai. (2005) kiemeli, hogy bizonyos esetben a stabilis eloszlás a farok (így a kockázat) túlbecsülésével jár. További hátrányként Tsay (2005) a stabilis eloszlások végtelen varianciáját emeli ki, ami a normál eloszlással szemben határozott hátrány.

A ferde **Student-T eloszlás** esetén mind a kurtózis, mind az aszimmetria paraméterezhető, ami hasznos lehet például folyamatos szimulációja során (például a később bemutatásra kerülő GARCH-modellek esetén) az m és ν paraméterek segítségével:

$$p(\varepsilon | \nu, \mu, \lambda) = \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})} \left(\frac{\lambda}{\pi\nu}\right)^{\frac{1}{2}} \left(1 + \frac{\lambda(\varepsilon - \mu)^2}{\nu}\right)^{-\frac{\nu+1}{2}}$$

Ahol μ jelöli a móduszt, ν a normális eloszláshoz képest mért szabadságfok és a $\lambda = \frac{1}{\sigma^2}$ az inverz skálázási paraméter.

Forrás:

Bonanno, G. – Lillo, F. – Mantegna, R. (2001): Levels of complexity in financial markets. *Physica A*. 299. pp. 16-27.

Clauset, A. – Shalizi, C. R. – Newman, M. E. J. (2009): Power-law distributions in empirical data. *SIAM Review*. 51. 4. pp. 661-703.

Deutsch H-P. (2009): *Derivatives and Internal Models*. Palgrave and Macmillan, London

Gabaix X. – Gopikrishnan P. – Plerou V. – Stanley H. E. (2003): A theory of power-law distributions in financial market fluctuations. *Nature*. 423 pp. 267-270.

Alapstatisztikák

A leíró statisztikák alkalmazásának célja, hogy ellenőrizzük vizsgált idősorok megfelelését az alapmodell elvárásainak – ami a hozamok normál eloszlását, gyenge stacionaritását, autokorrelálatlanságát és homoszkedaszticitását feltételezi.

i. **Normális eloszlás: Jarque-Bera teszt ($p > 0,05$)**

Az alapmodell feltételezi a hozamok várható érték körüli szóródását, és gyors lecsengését. Egy **normális eloszlás** esetében a lecsengés dinamikáját jelző farkak exponenciálisak, középtételei nagyrészt fedik egymást, illetve kis szórással rendelkeznek (Jentsch és mtsai. 2006).

Idősorok elemzése során a **Jarque-Bera teszt**re támaszkodunk, ahol az 5% alatti p érték a normál eloszlás elvetését jelenti⁴. A teszt az r_t^s asztandardizált hozamok harmadik és negyedik momentumainak (csúcosság és aszimmetria) standardizált normál eloszlással vett konzisztenciájának vizsgálatán alapul T elemszám mellett (22):

$$H_0: E(r_t^s)^3 = 0 \text{ és } E(r_t^s)^4 = 3, \text{ míg } H_1: E(r_t^s)^3 \neq 0 \text{ vagy } E(r_t^s)^4 \neq 3.$$

$$JB = \frac{T}{6} [T^{-1} \sum_{t=1}^T (r_t^s)^3]^2 + \frac{T}{24} [T^{-1} \sum_{t=1}^T (r_t^s)^4 - 3]^2 \quad (22)$$

A tesztstatisztika aszimptotikus $\chi^2(2)$ eloszlással rendelkezik, amennyiben a nullhipotézis elfogadható, míg a JB nagy értéke esetén a nullhipotézis elvethető (Lütkepohl 2004).

A Matlab egyik legfőbb előnye, hogy képes egy számolási feladat ismételt elvégzésére – esetünkben a Jarque-Bera teszt elvégzésére az összes oszlop esetében. A teszt kimenetét a könnyebb áttekinthetőség érdekében egy leíró statisztikákat gyűjtő mátrixban helyezem el. A $H=0$ esete ($p > 5\%$) normál eloszlásra, míg a **$H=1$ eset ($p < 5\%$) a normál eloszlás hiányára** utal.

ii. **Stacionaritás: ADF-teszt ($p < 0,05$)**

Lütkepohl és Kratzig (2004) illetve Brockwell és Davis (2002) könyvében definiálja a **stacionaritás erős, kovariancia (gyenge)** és az **aszimptotikus** változatait. Az **erős stacionaritást** egy $X_1, X_2, \dots$ diszkrét idejű véletlen folyamat együttes eloszlásából szokás levezetni, amennyiben bármilyen egész $\{i_1, i_2, \dots, i_k\}$ -ra és bármilyen m egészre igaz, hogy az $(X_{i_1}, X_{i_2}, \dots, X_{i_k})$ és $(X_{i_1+m}, X_{i_2+m}, \dots, X_{i_k+m})$ együttes eloszlása megegyezik. Ezért: $E[X_{i_1}, X_{i_2}, \dots, X_{i_k}] = E[X_{i_1+m}, X_{i_2+m}, \dots, X_{i_k+m}]$. Amennyiben X^k jelöli az $X_t, t \in \mathbb{Z}$ folyamat k -eltolású ($k \in \mathbb{Z}$) folyamatát, az X^k és az X_t a véges dimenziós eloszlás folytán ekvivalens lesz. **Kovariancia (gyenge) stacionaritásról** beszélhetünk abban az esetben, amennyiben az idősor első és a második momentuma explicit módon nem függ az időtől⁵. Tehát $E(x_t) = \mu_t$ minden $t \in T$ esetre, illetve $E[(x_t - \mu_t)(x_{t-m} - \mu_t)] = \gamma_t$ minden $t \in T$ illetve $t - h \in T$ esetre. Az **aszimptotikus stacionaritás** lép fel abban az esetben, ha az idősor egy előre meghatározott időpontban indul és ezt követően némi időre van szüksége a momentumok stabilizálódásához – a folyamat ez esetben a kezdőpont változtatásával stacionáriussá tehető (Lütkepohl és Kratzig 2004, 11-12. oldal, Brockwell és Davis 2002, 15. oldal). Egy idősort i -ed rendű integrálnak tekintünk, és $I(i)$ -gyel jelölünk, ha bár önmaga nem stacioner, ám az i -edik differenciáltja már az.

A bolyongás csupán az integrált folyamatok olyan speciális eseteként fogható fel Alexander (2008) szerint, ahol az első differenciált *iid*. Általánosan megfogalmazva, egy integrált folyamat első differenciáltja (23) rendelkezhet autokorrelált és mozgóátlag-komponensekkel – amennyiben stacioner:

$$R_t \sim I(1) \Leftrightarrow r_t = \alpha + r_{t-1} + \varepsilon_t, \text{ amennyiben } \varepsilon_t \sim I(0), \quad (23)$$

ahol $I(0)$ jelöli az idősor stacionerségét (Alexander 2008, 213. oldal).

⁴ Általánosságban elmondható, hogy a $p > 0,05$ értékek a tesztek nullhipotéziseinek elfogadását jelentik.

⁵ az angol „time-invariant” magyar megfelelőjeként az *idő-invariáns*t használom

A nem stacionárius folyamatokat a trendstacionárius (determinisztikus trend) és az egységgyök (sztochasztikus trend, differencia-stacionárius) folyamatok alkotják Darvas (2005) szerint. A trendstacionárius folyamatok (24) esetében a trend idősből történő kivonása stacioner folyamatot eredményez:

$$r_t = \mu + \delta t + \varepsilon_t, \text{ ahol } \varepsilon_t \sim iid(0, \sigma^2), E(r_t) = \mu + \delta t, V(r_t) = \sigma^2. \quad (24)$$

Az egységgyök folyamat (25) esetében az alábbi esettel van dolgunk:

$$r_t = \delta + \beta r_{t-1} + \varepsilon_t, \text{ ahol } \varepsilon_t \sim iid(0, \sigma^2), E(r_t) = r_0 + \delta t, V(r_t) = t\sigma^2, \quad (25)$$

ahol $r_t = \beta_1 r_{t-1} + \beta_2 r_{t-2} + \dots + \beta_p r_{t-p} + w_t$ esetet feltételezve a paraméterekből képzett $(1 - \beta_1 z + \beta_2 z^2 + \dots + \beta_p z^p) = 0$ polinom gyökei az egységkörön kívül helyezkednek el, de közöttük van 1 abszolút értékű.

Az **ADF(q) teszt** nullhipotézise szerint a vizsgált idősor nem gyengén stacioner, míg az alternatív hipotézis szerint gyenge stacionaritást mutat (26):

$$H_0: R_t \sim I(1) \text{ vs. } H_1: R_t \sim I(0). \quad (26)$$

A teszt során a (25)-ben ábrázolt AR(q) folyamatból indulunk ki, feltételezve, hogy $\rho_1 = 1 + \beta$. Amennyiben ugyanis β értéke nulla, ρ_1 nem fér bele az egységkörbe, így egységgyököt találunk. A q számú késleltetés célja, hogy az ε_t hibatagok autokorreláltságát elkerüljük (27):

$$\Delta r_t = \alpha + \beta r_{t-1} + \varrho_1 \Delta r_{t-1} + \dots + \varrho_q r_{t-q} + \varepsilon_t. \quad (27)$$

Amennyiben az ADF megfigyelt **t értéke** (*value of test statistic*), illetve első, vagy második integrálja **magasabb**, mint az 1%, 5% vagy 10%-os szint esetén megadott **kritikus érték**, akkor az egységgyök hipotézise nem vethető el (**nem stacioner** az idősor). Egy I(1) vagy I(2) eredmény a piacok gyengébb hatékonyságára utal. (Lütkepohl 2004)

Strukturális törések illetve hirtelen ugrások esetén azonban az ADF teszt hajlamos az elsőfajú hibára – akkor is stacionaritásra utal, amikor az idősor látványosan magán viseli a heteroszkedaszticitás jegyeit. További tapasztalat, hogy az alap tőkepiaci idősor adataiból számított logaritmusos hozamok már hajlamosak a stacionaritásra. A Matlab esetében alkalmazott módszer során a H=0 eset a stacionaritás hiányára, míg a H=1 stacionaritásra utal.

iii. Autokorreláció hiánya: Ljung-Box teszt ($p > 0,05$)

Egy hatékony piacon az eszköz hozama nem jósolható meg, és nem lehet autokorrelált – az **autokorreláció** vizsgálat egy eszköz a gyenge hatékonyságának vizsgálatára (Tsay 2005).

Autokorrelációról (szeriális korrelációról) abban az esetben beszélünk, ha az Y_t idősor értékei korrelálnak ugyanezen idősor korábbi értékeivel. Ekkor az idősor elemei között fellépő *sztochasztikus kapcsolatot* autokorrelációnak nevezzük, és a kapcsolat szorosságát autokorrelációs együtthatóval mérhetjük. A k-adrendű autokorrelációs együttható (ρ_k) (28) az egymástól k időegységnyi távolságra álló elempárok korrelációját mutatja:

$$\rho_k = \frac{C(Y_t, Y_{t+k})}{\sigma_{Y_t} \sigma_{Y_{t+k}}}, \quad (28)$$

ahol $c(Y_t, Y_{t+k})$ az Y_t és Y_{t+k} változók kovarianciája ($t=1, 2, \dots, n-k$), a σ_{Y_t} és $\sigma_{Y_{t+k}}$ a megfelelő szórások. Speciálisan $k=1$ esetén az idősor szomszédos elemei közötti korrelációt az elsőrendű autokorrelációs

együttható mutatja. Egy regressziós modell autokorrelálatlan, ha a különböző megfigyelésekhez tartozó reziduális változók korrelálatlanok. (Katona-Lengyel 1999)

Az autokorreláció meglétének teszteléséhez Portmenteau tesztet, vagy annak a módosított változatát, a **Ljung-Box tesztet** lehet alkalmazni. Ennek nullhipotézise (29) szerint adott késleltetés mellett nincs autokorreláció, szemben az alternatív hipotézissel, amely szerint van.

$$H_0: \rho_{r,1} = \dots = \rho_{r,h} = 0,$$

$$H_1: \rho_{r,i} \neq 0 \text{ } i = 1, \dots, h \text{ legalább egy esetben,} \quad (29)$$

ahol $\rho_{r,i} = \text{korrr}(r_t, r_{t-i})$ jelöli az idősor autokorrelációját.

A tesztstatisztika (30) az alábbi formát veszi fel:

$$Q_h = T \sum_{j=1}^h \rho_{r,j}^2,$$

$$\text{ahol } \rho_{r,j} = T^{-1} \sum_{t=j+1}^T r_t^s r_{t-j}^s. \quad (30)$$

A tesztstatisztika megközelítően egy $\chi^2(h)$ eloszlást vesz fel, amennyiben a nullhipotézis igaz. Q_h magas értékei mellett a nullhipotézis elvethető. A χ^2 eloszlás kikötése miatt a bevont elemek száma (h) nem lehet túl alacsony, sem túl magas. Ezt a problémát a Ljung-Box-féle módosítással (31) lehet áthidalni, ahol a χ^2 becslés sokkal alkalmasabb a probléma kezelésére:

$$Q_{(m)}^* = T(T+2) \sum_{j=1}^h \frac{1}{T-j} \rho_{r,j}^2 \approx \chi^2(h). \quad (31)$$

A nullhipotézist ez esetben akkor vetjük el, amennyiben $Q(m) > \chi_{\alpha}^2$, ahol χ_{α}^2 jelöli a $100(1-\alpha)$ dik percentilisét a h szabadságfokú χ^2 eloszlásnak. (Lütkepohl 2004, Chan 2002) A $H=0$ eset ($p>5\%$) az autokorrelálatlanságra, míg a **$H=1$ ($p<5\%$) autokorrelációra utal.**

iv. Homoszkedaszticitás: ARCH-LM teszt ($p>0,05$)

Statisztikai modellek gyakori feltétele a részsokaságok azonos varianciája, a homoszkedaszticitás. Az **ARCH-LM teszt a homoszkedaszticitást** méri egy ARCH(p) modell (32) illesztésével

$$r_t^2 = \beta_0 + \beta_1 r_{t-1}^2 + \dots + \beta_p r_{t-p}^2 + \varepsilon_t, \quad (32)$$

ahol a nullhipotézis (33) ellenőrzése

$$H_0: \beta_1 = \dots = \beta_p = 0 \text{ szemben a}$$

$$H_1: \beta_1 \neq 0 \text{ vagy } \dots \beta_p \neq 0. \quad (33)$$

Az LM teszt értéke a fenti regresszió R^2 koeficienséből vezethető le: $ARCH_{LM}(q) = TR^2$. Mindez $\chi^2(p)$ eloszlást vesz fel megközelítőleg, amennyiben teljesül a feltételes heteroszkedaszticitás nullhipotézise. A tesztstatisztika ennél magasabb értéke a nullhipotézis elvetését jelenti, ARCH-ra utalva a változók értékében. (Lütkepohl 2004) A $H=0$ értéke homoszkedaszticitásra ($p>5\%$), míg **$H=1$ ($p<5\%$) heteroszkedaszticitásra utal.**

Alapstatisztikák Matlabban

Egy szinkronizált, folytonos, adathiánytól mentes mátrixot tartalmazó *input.xls* Excel-fájlnak be lehet olvasni a *munka* névre hallgató munkalapját, majd az MFE-toolboxot megnyitva kiszámítjuk a logaritmikus differenciáltját. Ezt követően meghatározzuk a momentumait, teszteljük a normális

eloszlását, autokorreláltságot, homoszkedaszticitását és gyenge stacionaritását. A kapott eredményeket az *input* mátrix minden oszlopáról az *alap_stat* mátrix megfelelő sorába gyűjtjük ki, automatizálttá téve az alapstatisztikák kiszámítását.

```
clear
data=xlsread('input.xls','munka');
cd 'c:\ documents\matlab\mfetoolbox'
hozam=diff(log(data));
T=size(hozam);
for j=1:T(1,2)
    % 1. momentumok
    alap_stat(j,1)=mean(hozam(:,j));
    alap_stat(j,2)=std(hozam(:,j));
    alap_stat(j,3)=skewness(hozam(:,j));
    alap_stat(j,4)=kurtosis(hozam(:,j));
    % 2. normális eloszlás tesztelése (H=0 normál eloszlás)
    [statjstjc,alap_stat(j,5),H] = jarquebera(hozam(:,j));
    % 3. autokorreláció tesztelése (p<5% autokorreláció)
    [q, pval] = ljungbox(hozam(:,j), 2);
    alap_stat(j,6)=pval(2,1);
    % 4. heteroszkedaszticitás ARCH-LM teszttel (p<5% heteroszked)
    [stat2,pval2]=lmtest1(hozam(:,j)-mean(hozam(:,j)), 2);
    alap_stat(j,7)=pval2(2,1);
    % 5. ADF teszt (p<5%: stacioner)
    [adfstat,pval3,critval,resjd,lags]=augdfautolag(hozam(:,j),0,10);
    alap_stat(j,8)=pval3(1,1);
end
```

Forrás

Lütkepohl, H. and Kratzig, M. (2004) Applied Time Series Econometrics. Cambridge University Press, Cambridge. pp. 11-12.

Brockwell – Davis (2002) pp. 15.

Hodrick–Prescott (H-P) filter

- Hodrick–Prescott dekompozíció: a ciklikus komponensek kiszűrése az idősből a λ szorzó módosításával.
 - éves adat: $\lambda = \frac{1600}{4^4} = 19,75$
 - negyedéves adat: $\lambda = 1600$
 - havi adat: $\lambda = \frac{1600}{3^4} = 6,25$
- Feltételezve, hogy az idősor a g_t növekvő trend és a c_t ciklikus komponens nulla várható értékű összege $p_t = g_t + c_t$, az alábbi minimalizáció segít meghatározni a g_t értékét:
$$\text{Min}_{\{g_t\}_{t=1}^T} \{ \sum_{t=1}^T c_t^2 + \lambda \sum_{t=1}^T [(g_t - g_{t-1}) - (g_{t-1} - g_{t-2})]^2 \},$$
 - ahol a λ simítási paraméter és pozitív szám, amely korlátozza a trend-komponens változékonyságát. Ahogyan λ közelíti a végtelent, a trend-komponens úgy közelíti a lineáris időtrendet: $g_0 + \beta t$, ahol $\beta \approx (g_{t+1} - g_t)$ (Hodrick and Prescott 1997).
- Output-gap kiszámítása során alkalmazhatunk H-P filterezést a potenciális output megállapításához, figyelembe véve az idősor elején és végén fellépő torzításokat.
- Matlab (MFE toolbox):


```
[trend_l,cyclic] = hp_filter(y,1600)
[trend_s,cyclic] = hp_filter(y,100)
opg=trend_l-trend_s
```

Források

- Hodrick, R. J., Prescott, E. C. (1997): Postwar U.S. Business Cycles: An Empirical Investigation. *Journal of Money, Credit and Banking*. 29, 1, pp. 1-16
- Ravn, Morten; Uhlig, Harald (2002): On adjusting the Hodrick–Prescott filter for the frequency of observations. *The Review of Economics and Statistics*. 84, 2, pp. 371.

Önellenőrző kérdések

1. Mekkora az négy momentum elvárt értéke?
2. Mit reprezentál a Jarque-Bera-teszt 0,05 alatti p-értéke?
3. Mire következtethetünk akkor, ha a Ljung-Boksz-teszt p-értéke 0,05 alatti?
4. Miért fontos, hogy az ARCH-LM-teszt p-értéke 0,05 felett legyen?
5. Szükséges-e differenciálnunk az idősort, ha az ADF-teszt p-értéke 0,01?

Jelen tananyag a Szegedi Tudományegyetemen készült az Európai Unió támogatásával. Projekt azonosító: EFOP-3.4.3-16-2016-00014

