

Jövedelmek szerkezeti elemzése az 1968-as magyar generáció adatain

Készítette: Varga Viktor

Témavezető: Berlinger Edina



ELTE-BCE, Biztosítási és Pénzügyi Matematika MSc

Budapest, 2012

1. fejezet

Bevezetés

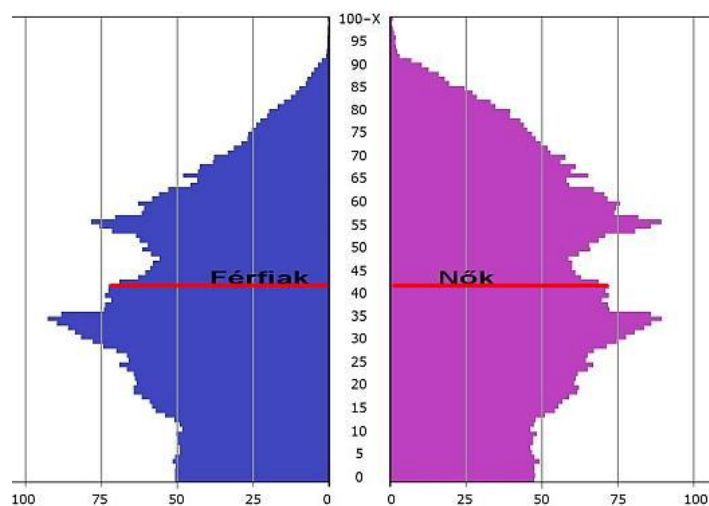
A hosszútávú kockázatkezelés napjaink egyik legfontosabb feladata. A néhány éve kialakult, és sok szempontból még a jelenben is tartó gazdasági válság utáni időszak folyamatainak megértése és előrejelzése mindannyiunk számára nélkülözhetlenné vált. A rövid távú kockázatkezelés módszertana azonban sok szempontból nem nyújt kielégítő választ bizonyos kérdéskörökre. Csak néhány példát említve: a diákhitelezés, a nyugdíjrendszer és az életbiztosítás alapvetően hosszútávú rendszerek, amelyekhez egészen másféle vizsgálat szükséges. Egy azonban közös bennük: mindegyikük a jövedelmekkel függ igen szorosan össze.

A magyar diákhitel törlesztése jövedeleमारányos, az aktív életpálya során fizetendő nyugdíjjárulék úgyszintén, életbiztosítás pedig sok esetben azért köttetik, mert vagy egy esetleges halálesetből adódó jövedelemkiesés negatív következményeit igyekeznek kivédeni, illetve sok esetben egyfajta öngondoskodási motívumként jelentkeznek, amelynek nagysága általában a jövedelmi szinttől függ.

Szakedolgozatom során éppen ezért igyekszem a magyar jövedelmek struktúráját szemügyre venni, és az empirikus elemzés során egyfajta támaszt nyújtani a hosszútávú rendszerekre levonható következtetéseknek. A vizsgálathoz egy olyan korszakot veszünk górcső alá, amely szándékosan nem a

válság éveiben, hanem az azt megelőző „boldog békeidőkben” mutatja egy lehetőleg reprezentatív minta jövedelmeit.

Ezért esett választásom az 1968-ban születettekre, és az 1995 – 2005-ös időszakra (ld: *1.1 ábra* pirossal megjelölt korosztálya). Így a mintában a rendszerváltozás utáni „káosz” és a válságot közvetlenül megelőző időszak nem szerepel. A magyar korfa alapján látható, hogy szándékosan nem a kiugró lélekszámú Ratkó-generációt, illetve azok leszármazottait, hanem egy általános, szélsőségektől mentes, ugyanakkor mégis kellően „népes” generációt elemzünk. A korfa adott ágának időbeli változásától itt most eltekintünk. A vizsgált korosztály jövedelmének alakulását így végigkísérjük 27 és 37 éves koruk között, ezáltal a jövedelemdinamikát az életpálya egyik legfontosabb szakaszában tudjuk megvizsgálni. Továbbá megjegyezzük: az adott generációnak a fent említett 11 év során történő vizsgálata azt eredményezi, hogy a minta jó eséllyel mentes a generációk közötti transzfertől, ezáltal nyújthat bizonyos szempontból hitelesebb képet a teljes magyar lakosság vizsgálatánál.



1.1. ábra. Magyar korfa - 2012 január
Forrás: KSH - Interaktív korfák

Az adatok *APEH*-forrásból származnak, ezáltal elemzésre alkalmasnak és pontosnak tekinthetők, ugyanakkor sok helyen sajnos „hiányosak”, amit majd adott esetben orvosolnunk kell. Főként a számszerű mutatókra, és azok következményeire vagyunk elsősorban kíváncsiak, ezért az elemzéshez szükséges módszertant a gyakorlat számára szükséges fogalmakon keresztül ismertetem. A vizsgálatok elkészítéséhez főként *Excel*-t és *Matlab*-ot használtam.

A motiváció tehát adott: keresztmetszeti és hosszmetzeti elemzéssel megvizsgáljuk, mit lehet mondani a jövedelmek eloszlása és időbeli változása alapján, illetve ezt követően bemutatunk néhány lehetőségét annak, hogy hol és hogyan alkalmazhatjuk eredményeinket.

2. fejezet

A jövedelmek eloszlásvizsgálata

2.1. Bevezetés

A szakirodalomban számos helyen előfordulnak a jövedelmekre és azok eloszlására vonatkozó feltevések. Általában alapvető stilizált ténynek tekinthető a jövedelmek lognormális eloszlása, függetlenül attól, hogy magyar vagy külföldi jövedelmekről beszélünk. Magyar jövedelmeket vizsgálva például többparaméteres lognormális eloszlás viszonylag jól illeszthető volt az 50 évvel ezelőtti jövedelemadatokra (*Éltető-Vita [1982]*). Ez a feltevés azonban csak első közelítésben tekinthető igaznak, és nem pusztán az idő múlása miatt.

A helyzetet talán ahhoz lehetne hasonlítani, ahogyan azt is sokszor feltételezzük, hogy a részvényhozamok alakulása normális eloszlást követ. Ez a feltevés sok esetben kiindulásként helytállónak tekinthető, ráadásul ezáltal a modellek könnyen kezelhetőek, és az így levont következtetések általában nem is állnak messze a valóságtól.

Sajnos azonban pont akkor bizonyul tévesnek ez az alapfeltevés, amikor igazán szükség lenne a pontos eredményekre, például az erősen negatív hozamok valószínűségének megbecslése esetén. Ráadásul hosszabb távon vizsgálódva ez a hiba komoly következményekkel is járhat. Éppen ezért nem támaszkodunk feltétlenül a jövedelmek lognormális eloszlására.

2.2. A lognormalitás elvetése

A jövedelmek tényleges empirikus eloszlását megvizsgálva elsőként a lognormalitás helyességét ellenőrizzük. Ezt a következőképpen végezzük: tekintsük a jövedelmek logaritmusát, és ha valóban teljesülne a jövedelmek lognormális eloszlása, akkor a jövedelmek logaritmusai normális eloszlást követne.

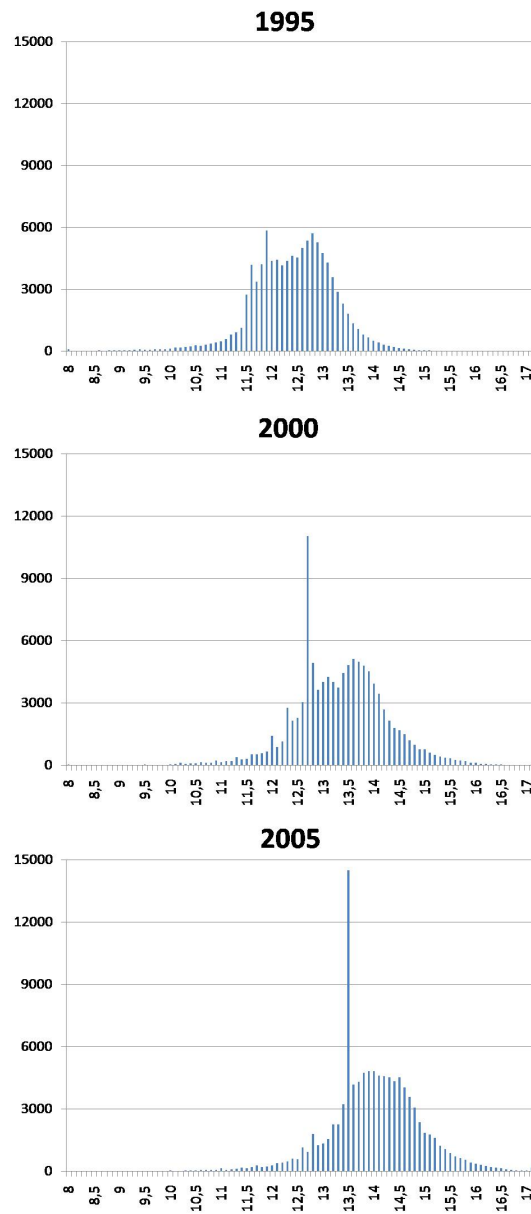
A vizsgált időszak elejét, közepét és a végét tekintve, az 1995-ös, 2000-es és a 2005-ös évek jövedelmei logaritmusának gyakoriságait láthatjuk a *2.1-es ábrán*. Ezekből a grafikonokból is sejthető, hogy a normalitás igencsak kétséges, és nem pusztán a minden évben megjelenő, pontszerűen kiugró gyakoriság miatt. Jól láthatóan az eloszlás igen ferde, ami ugyan az idő előrehaladtával valamelyest mérséklődik, de valójában az eloszlás látványosan eltér a normálistól. Egyedül talán a faroklecsengésről mondhatjuk, hogy valóban hasonlít a normális eloszláséhoz.

Vizuális benyomásainkat számszerűen alá is támasztjuk: a normális eloszlás tesztelésére a *Jarque-Bera tesztet* alkalmazzuk, amely paraméterfüggetlen becslés: a tapasztalati minta ferdeségét és csúcsosságát használja fel, és ezt hasonlítja össze a normális eloszlás viselkedésével. A ferdeséget (S) és a csúcsosságot (K) a következőképpen definiálhatjuk:

$$S = \frac{\sum_{i=1}^n (x_i - \hat{\mu})^3}{n\hat{\sigma}^3}$$

$$K = \frac{\sum_{i=1}^n (x_i - \hat{\mu})^4}{n\hat{\sigma}^4}$$

Itt n jelöli a minta elemszámát, $\hat{\mu}$ a mintából számolt átlagot, $\hat{\sigma}$ pedig a mintából számolt szórást.

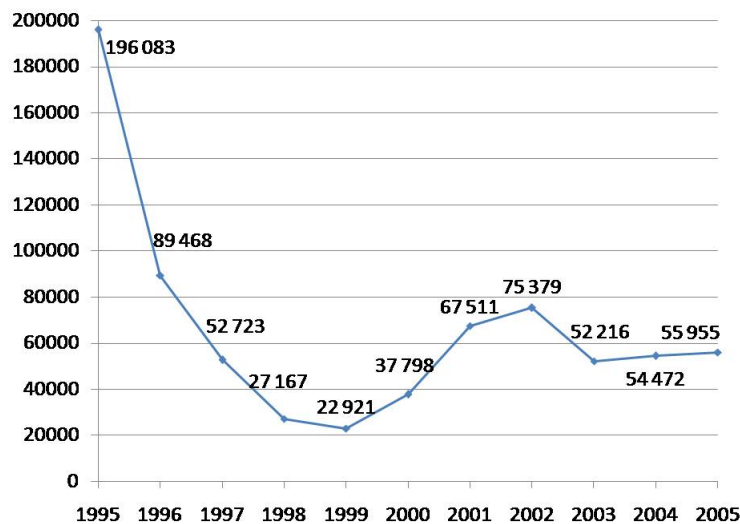


2.1. ábra. A jövedelmek logaritmusainak gyakorisága: 1995, 2000, 2005

A Jarque-Bera tesztstatisztika a következő:

$$JB = \frac{N}{6} \left(S^2 + \frac{(K - 3)^2}{4} \right)$$

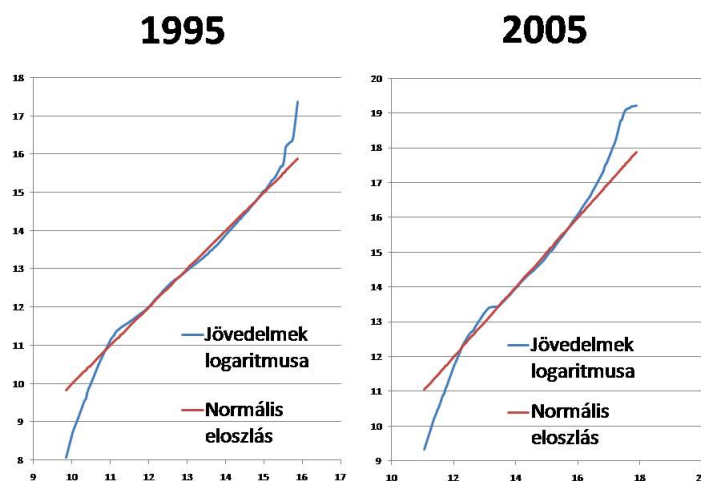
A nullhipotézis, vagyis a normális eloszlás fennállása esetén a fenti statisztika aszimptotikusan χ_2^2 eloszlást követ. Ennek kritikus értéke 95%-os szignifikancia-szinten körülbelül 5,99. Minden évre kiszámolva a 2.2 ábrán láthatjuk a tesztstatisztika értékeit.



2.2. ábra. A Jarque-Bera statisztika értékei: 1995-2005

Látható, hogy a kritikus értékhez képest tízezres nagyságrendű a tesztstatisztika, ezáltal a normalitás, vagyis a jövedelmek eloszlására vonatkozó lognormális eloszlás feltételezése egyértelműen elvethető minden évre. Az igazsághoz hozzátartozik azonban, hogy a Jarque-Bera teszt rendkívül érzékeny az outlierekre, ezért konkrét értékét minden esetben fenntartással kell kezelnünk. Azonban az eltérés olyannyira nagy, hogy feltehetőleg nem ez a fő probléma.

Ezt a megállapítást egy grafikus vizsgálattal, a $Q-Q$ plottal ellenőrizzük: a normális eloszlás percentilisei függvényében ábrázoljuk a tapasztalati minta percentiliseit, illetve a normális eloszlás percentiliseit ugyanazon ábrában. Az utóbbi grafikonja nyilvánvalóan az $f(x) = x$ egyenletű 45 fokos egyenes, tehát lényegében azt vizsgáljuk, szignifikánsan eltér-e ettől az egyenestől a mintánk. Minden évre elvégezve a számolást, hasonló eredményt kapunk a 2.3-as ábrán látható grafikonhoz, amely az 1995-ös, és a 2005-ös évre mutatja a $Q-Q$ plotokat.



2.3. ábra. $Q-Q$ plot: 1995, 2005

Megállapítható, hogy az eloszlás bal széle jelentősen eltér a normális eloszlástól, és a farklecsengés is különbözőséget mutat. Azt gondolhatnánk talán, hogy utóbbi nem annyira számottevő, így ha a jövedelmek teljes eloszlására nem is, de a farklecsengésére már esetleg a lognormális eloszlás lecsengéséhez hasonló viselkedést tételezhetünk fel. Az alábbi okoskodás mutatja azonban, hogy ez sem bizonyul helytállónak.

Az X valószínűségi változó $(\mu; \sigma)$ paraméterű lognormális eloszlású, ha $\ln X$ μ várható értékű, σ szórású normális eloszlású. A lognormális eloszlás sűrűségfüggvénye ebben az esetben az alábbi alakban írható fel:

$$f(x) = \frac{1}{x\sigma\sqrt{2\pi}} \exp\left(-\frac{(\ln x - \mu)^2}{2\sigma^2}\right)$$

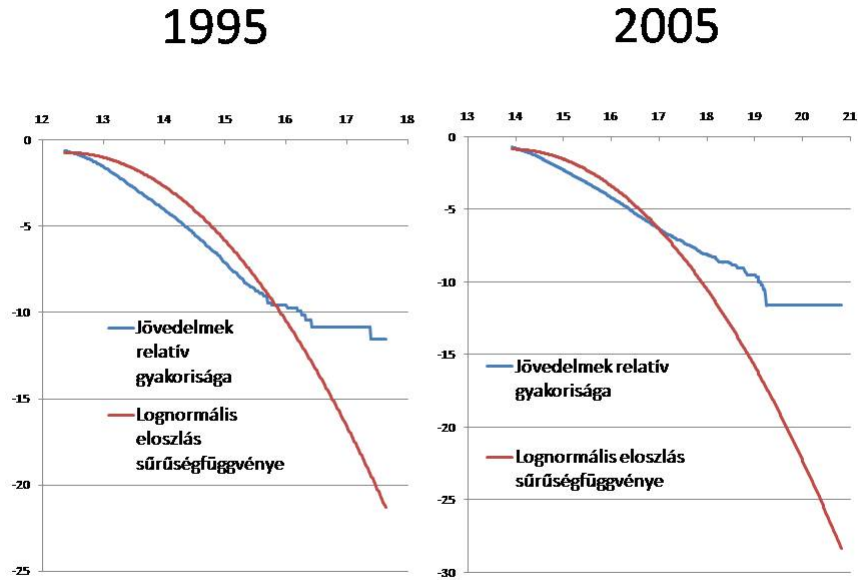
Tehát K -val pozitív konstans jelölve, a fenti kifejezést egyszerűbb alakra hozva:

$$f(x) = K \frac{1}{x} \exp\left(-\frac{\ln^2 x}{2\sigma^2} + \frac{\mu \ln x}{\sigma^2}\right) = K \frac{1}{x} x^{-\frac{\ln x}{2\sigma^2}} x^{\frac{\mu}{\sigma^2}} = K x^{-1 - \frac{\ln x}{2\sigma^2} + \frac{\mu}{\sigma^2}}$$

Így a sűrűségfüggvény logaritmusát felírva, a , b , c -vel alkalmasan megválasztott konstansokat jelölve:

$$\ln f(x) \sim \ln K + \left(-1 - \frac{\ln x}{2\sigma^2} + \frac{\mu}{\sigma^2}\right) \ln x = a \ln^2 x + b \ln x + c$$

A fenti számolás azt jelenti, hogy ha feltesszük a jövedelmek lognormális eloszlását, vagy akár csak azt, hogy egy bizonyos küszöbtől kezdve adataink lognormális eloszlást követnek, akkor a küszöbön túli értékekre a küszöbtúllépések relatív gyakoriságainak logaritmusait a küszöbértékek logaritmusainak függvényében ábrázolva (tehát lényegében az adatokat $\ln - \ln$ skálán tekintve), és felhasználva, hogy $a < 0$, megközelítőleg *parabolikus* esést kellene tapasztalnunk. A 2.4-es ábrán azonban jól látható, hogy az $\ln - \ln$ skálán a lognormális esés parabolikusságához képest lényegében *lineáris* függvény szerinti esést tapasztalunk, így a jövedelmek eloszlása esetén a lognormális faroklecsengés is elvethető.



2.4. ábra. Ln-Ln skálán az 1995 és a 2005-ös jövedelemadatok

A fenti lineáris esés ugyanakkor hatványszerű lecsengést jelent, hiszen:

$$P(X > x) \sim \frac{1}{x^\alpha} \Leftrightarrow f(x) \sim \alpha x^{-\alpha-1}$$

Ezáltal pedig C -vel pozitív konstans jelölve: $\ln f(x) \sim -C \ln x$, azaz $\ln - \ln$ skálán csökkenő lineáris függvényt kapunk, éppen ahogyan adataink mutatják.

Összefoglalva tehát megállapíthatjuk: a jövedelmek eloszlása nem lognormális, sőt a faroklecsengés is lognormális esés helyett hatványszerűnek tekinthető.

A hatványszerű esés kitevőjének megbecsléséhez a *Hill-módszert* alkalmazzuk (ld: *Marossy Zita [2010]* értekezése, illetve *McNeil, Frey, Embrechts [2005]* alapján).

Tegyük fel, hogy az A küszöb felett a jövedelmek eloszlása az alábbi alakú sűrűségfüggvénnyel jellemezhető:

$$f(x) = \frac{\alpha A^\alpha}{x^{1+\alpha}}$$

Legyen N azon értékek száma a mintában, akik meghaladják az A küszöböt. A rendezett mintát jelölje: $\{X_1; X_2; \dots; X_N\}$. A hatványszerű viselkedés kitevőjére maximum-likelihood becslést alkalmazunk. A likelihood-függvény alakja tehát a következő:

$$L(X_1; X_2; \dots; X_N; \alpha; A) = \prod_{i=1}^N f(X_i) = \frac{\alpha^N A^{N\alpha}}{\left(\prod_{i=1}^N X_i\right)^{1+\alpha}}$$

Így a log-likelihood függvény:

$$\ln L = N \ln \alpha + N\alpha \ln A - (1 + \alpha) \sum_{i=1}^N \ln X_i$$

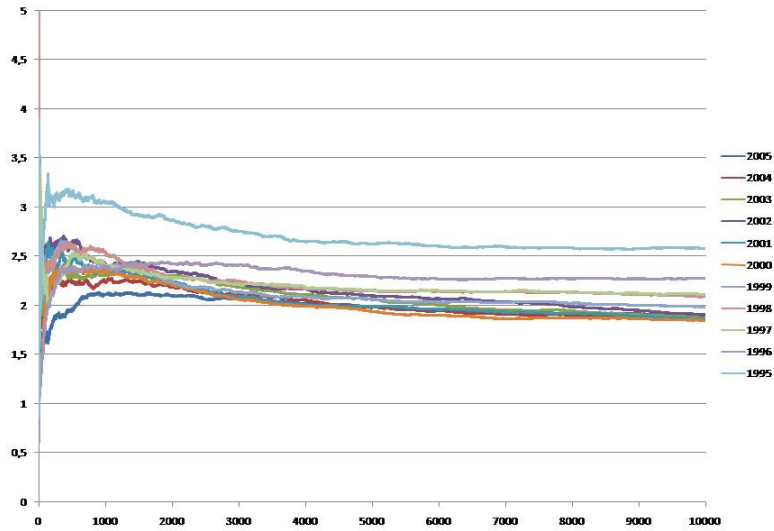
Keressük azt az α -t, amire ez maximális, így α szerint deriválva:

$$\frac{\partial \ln L}{\partial \alpha} = \frac{N}{\alpha} + N \ln A - \sum_{i=1}^N \ln X_i = 0$$

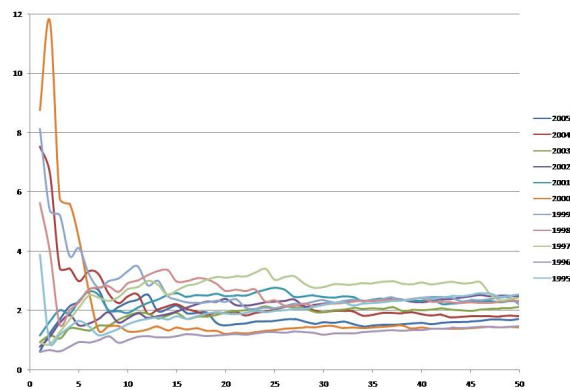
Innen átrendezve megkapjuk az α hatványkitevő becslt értékét:

$$\hat{\alpha} = \frac{N}{\sum_{i=1}^N \frac{\ln X_i}{A}}$$

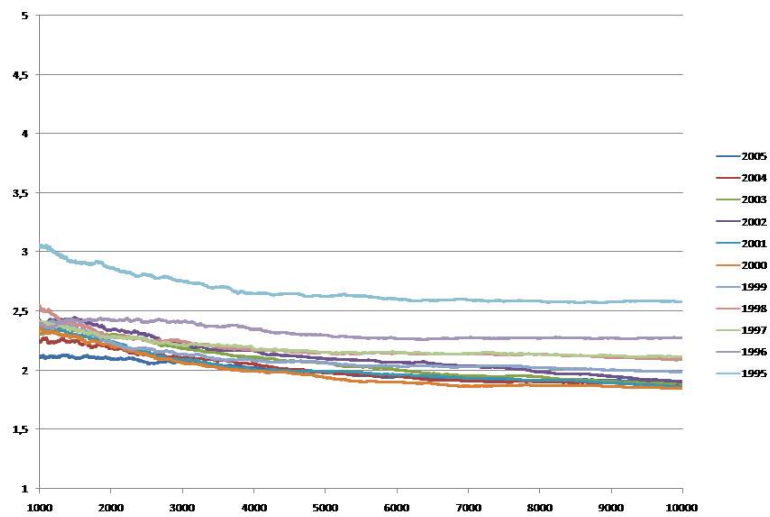
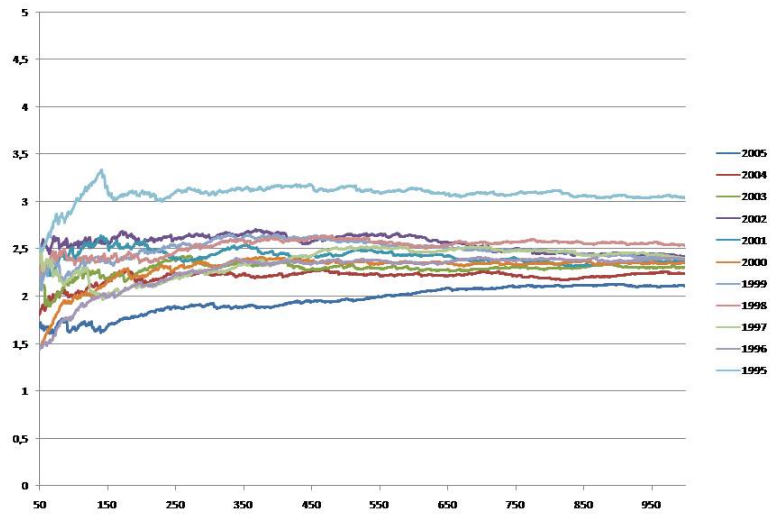
A becslt kitevő függ a küszöb megválasztásától, ezért kiszámoljuk a kitevőt különböző A küszöbértékekre. Az eltérő küszöbértékekhez természetesen eltérő mintaelemszám tartozhat, ezért értelmes, és érdemes vizsgálnunk a becslt kitevőt a mintaelemszám függvényében. Ezt ábrázolva kapjuk a *Hill-plotot*, amelyet egészében (2.5 ábra) és szakaszaiban is megtekinthetünk.



2.5. ábra. A becstelt kitevő a mintaelemszám függvényében (Hill-plot)



2.6. ábra. A Hill-plot első szakasza: az 50 legnagyobb jövedelemre illesztett kitevő



2.7. ábra. A Hill-plot középső és harmadik szakasza

Az első szakaszból, amely a legnagyobb 50 jövedelem (a teljes minta kb. 0,05%-a) logaritmusára illesztett kitevőt mutatja. Jól láthatóan az eloszlás legeslegszielét, vagyis körülbelül a 4 – 6 legnagyobb (extrém) értéket leszámítva a becsült kitevő minden évre rendkívül hamar stabilizálódik (*2.6 ábra*).

Ugyanez a megállapítás érvényes a Hill-plot középső szakaszát tekintve: az 50 – 150 legnagyobb megfigyelést követően a kitevő lényegében konstanssá válik: értéke, az adott évtől függően 2 – 2,5 közötti (*2.7 ábra*).

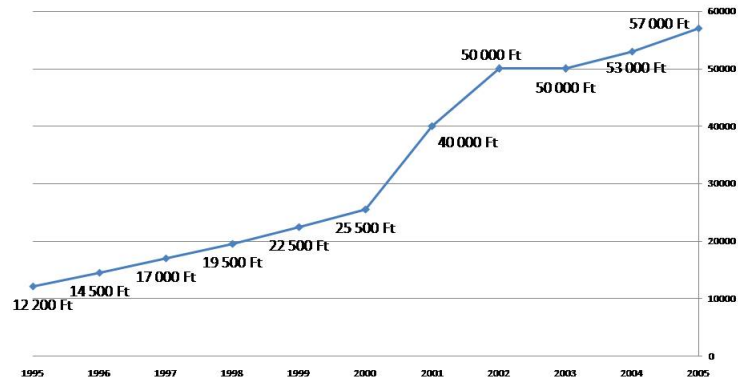
10000 küszöbtúllépésre, azaz körülbelül a teljes minta felső decilisét tekintve a stabilitás megmarad, a kitevő ugyan enyhén csökken, ám ez az érték már további küszöbtúllépésekre, a teljes faroklecsengést tekintve is stabil marad. Érdeemes megfigyelni azt is, hogy az idő előrehaladtával a stabilizálódott kitevő értéke enyhén csökken, tehát az eloszlás idővel egyre vastagabb szélűvé válik (*2.7 ábra*).

2.3. További empirikus tények

A továbbiakban olyan egyszerű megállapításokat teszünk, amelyek számszerű elemzése nehézkes, ezért a megfogalmazásokon túl részletes elemzést nem végzünk, ugyanakkor fontosnak tartjuk, hogy legalább említés szintjén foglalkozzunk velük.

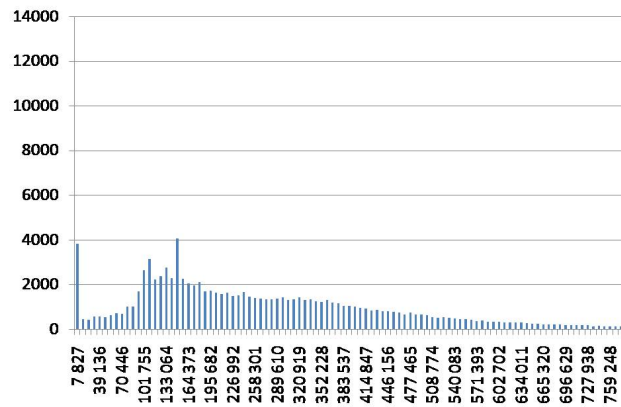
A minimálbér alakulását a teljes időszakra, illetve a jövedelmek gyakoriságát az 1995-ös, a 2000-es és a 2005-ös évekre vonatkozóan láthatjuk az alábbi ábrákon. A jobb megjeleníthetőség végett a jövedelmek adott évi legfelső 5%-át nem ábrázoltuk. Következtetéseinket az alábbiakban foglaljuk össze.

Havi minimálbér

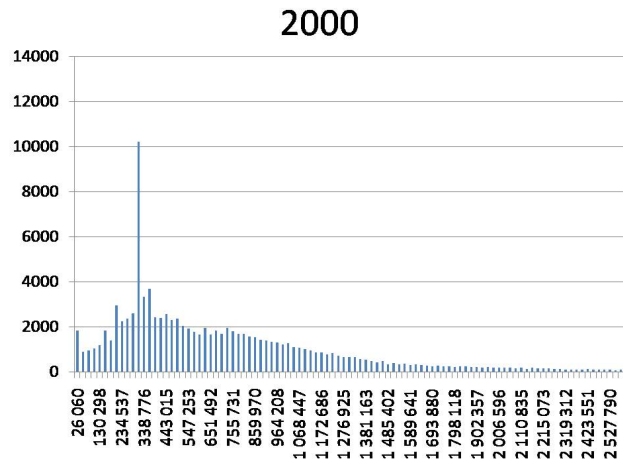


2.8. ábra. Havi minimálbér alakulása 1995-2005
 Forrás: KSH - Minimálbér

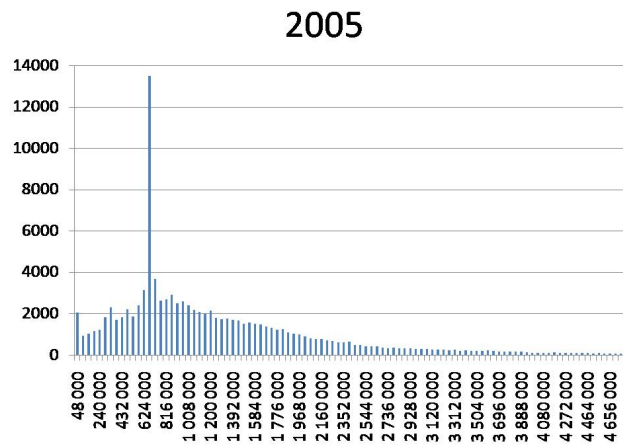
1995



2.9. ábra. Jövedelmek gyakorisága: 1995



2.10. ábra. Jövedelmek gyakorisága: 2000



2.11. ábra. Jövedelmek gyakorisága: 2005

1. *A minimálbér hatása igen jelentős.*

Akármelyik évet is tekintjük, van az eloszlásnak egy olyan pontja, amelyik kiugró értéket mutat. Ez nem más mint a minimálbér megfelelő értéke: összevetve a havi minimálbér időbeli alakulásával (lsd: 2.8 ábra), látható, hogy az éves minimálbér értékére esik igen sok megfigyelésünk, azaz (hivatalosan) igen jelentős a minimálbéren dolgozók száma. Ráadásul arányuk az idő előrehaladtával nő, ezért a jövedelem eloszlását semmiképpen sem tekinthetjük végig folytonosan alakulónak.

Ugyanakkor a minimálbéresek számát, illetve arányát megtippelni és főként előre becsülni igen nehéz lenne. Ami jól látszik, hogy a minimálbér összege lehetne az egyik magyarázó tényező: ha a minimálbér összege az előző évihez képest nagyobb ütemben nőtt, mint a jövedelmek szintjének általános növekedési üteme az adott évre, akkor várhatóan nőni fog a minimálbéresek száma.

2. *A gyakoriságok szakaszonként sem monotonak.*

Ezalatt azt értjük, hogy a minimálbér felénél-kétharmadánál, illetve kétszeresénél az eloszlás a „várható” viselkedésétől eltérően magasabb gyakorisági értékeket mutat fel. Ez azonban másként jelentkezik a két értéknél: a minimálbér felénél-kétharmadánál pontszerű kiugrás figyelhető meg, amit valószínűleg a 4 – 6 órás „félállások”, illetve az alkalmi munkanélküliség magyarázhat, míg a minimálbér nagyjából kétszeresénél ez inkább egy környezetre kihatóan, és kevésbé pontszerűen jelentkezik, ez a diplomás minimálbér hatása miatt lehetséges.

3. *Az extrém kis jövedelműek gyakorisága hirtelen növekszik.*

A gyakorisági értékek jól láthatóan nem folytonosan növekednek a 0-ból indulva, hanem hirtelen „megugranak”: akik pozitív jövedelemmel rendelkeznek, azok esetében (néhány kivételtől eltekintve) a jövedelem-szint rögtön egy bizonyos értéktől indul és onnan növekszik folytonosan, nem pedig a 0-ból indulva.

4. *Az extrém nagy értékek száma rendkívül kicsi.*

A nagy jövedelmi differencia hozománya, hogy a legfelső decilisen belül is találhatunk néhány kiugróan magas értéket. Ezek száma azonban relatíve alacsony, számuk legfeljebb 10 körüli, általában 4 – 6 megfigyelésről van szó, ezeket adott esetben különös figyelemmel kell megvizsgálni.

Azonban körülbelül 100000-es mintanagyságnál a 4 – 6 megfigyelés olyan elenyészőnek tekinthető, hogy emiatt nincs szükség az extrém-érték elméletben használt küszöbtúllépéseket, illetve általánosított Pareto-eloszlást vizsgálni.

A veszteségeloszlások esetében, ehhez kapcsolódóan a kockázatkezelés során sokszor előszeretettel használják a fenti vizsgálatot és eloszlást (lsd. pl: *Embrechts-Klüppelberg-Mikosch [1997]*), azonban számunkra ez nem szükséges, hiszen a fontos alkalmazásoknál nem az extrém nagy jövedelműek adják a fő vizsgálat tárgyát.

2.4. A jövedelmek egyenlőtlensége

Ezután a teljes jövedelem felosztásának egyenlőtlenségét szeretnénk megvizsgálni. Ehhez grafikus és számszerű mutatókat is igénybe veszünk.

1. *Lorenz-görbe*

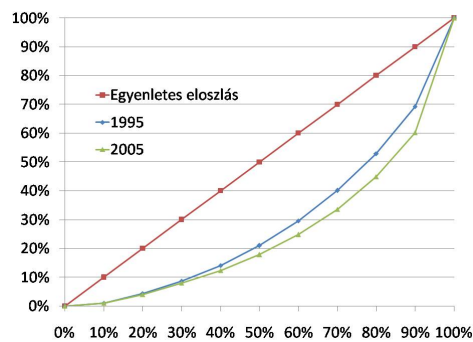
A Lorenz-görbe grafikus formában mutatja a koncentrálttságot, esetünkben a jövedelmek eloszlásának egyenlőtlenségét. Azt ábrázolja, hogy rendezett jövedelem-minta (x_i) esetén a teljes összjövedelemnek mekkora százalékát „birtokolja” a jövedelem szerint rendezett egyének adott alsó hányada (lsd. pl: [20]). A definícióból adódóan ez egy olyan $L : [0; 1] \rightarrow [0, 1]$ függvény görbéje, amely az origóból indul, az $(1; 1)$ pontba érkezik, amely továbbá (jövedelmekről lévén szó) monoton, és konvex:

$$L(x) = L\left(\frac{k}{n}\right) = \frac{\sum_{i=1}^k x_i}{\sum_{i=1}^n x_i}$$

A görbe két szélsőséges helyzetét akkor kapjuk, amikor teljes egyenlőség, illetve teljes egyenlőtlenség valósul meg. Előbbi esetben minden egyén ugyanakkora arányban részesül az összjövedelemből, ezáltal a görbe az $f(x) = x$ egyenletű 45 fokos egyenes lenne. A teljes egyenlőtlenség esetén mindössze 1 egyén birtokolja az összes jövedelmet, így a görbe egy olyan töröttvonal, amely végig konstans 0, kivéve az utolsó ponttól, ahonnan egyenletesen közelíti meg az 1-et. Azaz, n -el jelölve a teljes minta elemszámát, és χ -vel az indikátorfüggvényt:

$$L(x) = (nx + (1 - n)) \chi_{\{x \geq 1 - \frac{1}{n}\}}$$

Az adott évek jövedelmi deciliseire illesztve a Lorenz-görbét, azt tapasztaljuk, hogy egészen minimális az eltérés a görbék között, ezért illusztrálásképpen csak a két szélső évre, tehát 1995-re és 2005-re ábrázoljuk őket (2.12 ábra).



2.12. ábra. Az egyenletes eloszlás, az 1995-ös és a 2005-ös év jövedelmeinek Lorenz-görbéje

Az évek során tehát a görbe keveset változik, ezzel együtt a belőle levonható következtetések is: az összjövedelem felét a felső két decilis, vagyis a felső 20% birtokolja. Bár általában nem feltétlenül teljesül a szimmetria, de esetünkben ez a helyzet: az alsó társadalmi csoportok, azaz a minta „rosszabbik fele” csupán a jövedelem 20%-ának birtokosa. A fentihez rendkívül hasonló görbét láthatunk *Kovács I. [2010]* művében, amely hasonló időszak magyar adataira illeszkedik. Az egyenlőtlenség számszerűsítéséhez a Lorenz-görbe segítségével kiszámítunk egy mutatószámot.

2. Gini-index

Láthattuk, hogy a tapasztalati eloszlás természetesen (mint ahogyan minden más vizsgálódás során) a két szélsőség között mozog, minél egyenlőtlenebb az eloszlás, annál inkább „távolabb” leszünk az $f(x) = x$ egyenestől. Ennek a távolságnak egyfajta mérőeszköze a Gini-index.

Ezt az együtthatót tipikusan jövedelem, illetve vagyon eloszlásának vizsgálatra használják, így nekünk is segítséget nyújthat. Értékét a $[0; 1)$ intervallumból veszi fel: 0 értéket a fent említett egyenletes eloszlás esetén vesz fel, az 1-et pedig akkor venné fel, ha az eloszlás egyetlen pontra koncentrált, és végtelen sok adatunk lenne. Minél nagyobb ez az érték, annál inkább egyenlőtlen a jövedelemszerkezet.

Alapvetően azt méri, hogy a tapasztalati és az egyenletes eloszlás Lorenz-görbéje, vagyis az $\tilde{L}(x) = x$ egyenes mennyire tér el egymástól. Az eltérést a két görbe alatti terület különbsége adja, pontosabban: a Gini-index a két görbe közötti terület és az egyenletes eloszlás Lorenz-görbéje alatti területének a hányadosa. Az utóbbi értéke $\frac{1}{2}$, ezért mindezt formalizálva:

$$GI = 1 - 2 \int_0^1 L(x) dx$$

A fenti képletet átírhatjuk másmilyen alakba is, attól függően, hogy a teljes rendezett minta, a mintában előforduló relatív gyakoriságok, vagy az eloszlásfüggvény alakjából könnyebb az index meghatározása (lsd. pl: [20]). Esetünkben a rendezett mintával történő számolás a legkönnyebb, így tehát x_i -vel jelölve a rendezett minta i . elemét:

$$GI = \frac{1}{n} \left(n + 1 - 2 \frac{\sum_{i=1}^n (n + 1 - i)x_i}{\sum_{i=1}^n x_i} \right) = \frac{2 \sum_{i=1}^n i x_i}{n \sum_{i=1}^n x_i} - \frac{n + 1}{n}$$

Az indexet kétféleképpen is kiszámoljuk, megnézzük decilisenként tekintve és egyéni szinten is. Nyilvánvalóan a decilisek esetén kisebb értéket kapunk, mintha egyénekre lebontva vizsgáljuk, hiszen a fenti integrál kiszámolásához a teljes mintát figyelembe véve n elemű, míg a decilisek esetén 10 elemű közelítőösszeget vettünk, amely nyilvánvalóan durvább felosztást, ezáltal (konvex függvényről lévén szó) kisebb területet eredményez.

Kiszámolva minden évre az indexet, jól látható a 2.13-as ábrán, hogy időben nagyjából konstansnak tekinthető, akármelyik felosztást is tekintjük: az 1995-ös évet leszámítva az értéke nem nagyon változik az idő múlásával. Látható az is továbbá, hogy a decilisre osztással számolt érték már kellően jó közelítésnek tekinthető, így hagyatkozhatunk erre az eredményre.

Kérdéses, hogy hogyan értelmezzük a kapott eredményt. Az index számos előnyös tulajdonsággal bír: úgy mér egyenlőtlenséget, hogy mindez nem függ sem az eloszlástól, sem a mintaelemszámtól, ezért például a jövedelmek lognormalitásának fel nem tévése sem jelent problémát. Minden évre kiszámolva időben tudjuk követni magának az egyenlőtlenségnek az időbeli változását. Ugyanakkor azt nem tudjuk belőle megállapítani, hogy az adott egyenlőtlenség milyen szinten van: a tel-

jes társadalom vagyoni helyzetét, nevezetesen, hogy inkább általános szegénység vagy jólét van-e, azt nem tudjuk meghatározni. Ami az előnye, az a hátránya is egyben: ugyanolyan értékhez többféle eloszlás is tartozhat, tehát magára a jövedelemeloszlásra semmilyen módon nem tudunk következtetni, ugyanakkor megjegyezzük, hogy az index kiszámolásakor nem is ez volt a fő célunk.

3. *Robin Hood-index*

A fent nevezett index szintén az egyenlőtlenség egyik kifejezője, nevéből következően azt mutatja meg, hogy a jövedelem mekkora százalékát kellene átcsoportosítani az egyes kategóriák között úgy, hogy megvalósulhasson a teljes társadalmi egyenlőség, azaz minden kategóriában ugyanakkora legyen az egy főre jutó jövedelem (lsd. pl: [19]).

Jelöljük A_i -vel a rendezett jövedelmek összegét az i . kategóriában, B_i -vel az i . kategóriába eső elemek számát, N pedig jelölje a kategóriák számát. Ekkor az index kiszámítása a következő:

$$RH = \frac{1}{2} \sum_{i=1}^N \left| \frac{A_i}{\sum_{j=1}^N A_j} - \frac{B_i}{\sum_{j=1}^N B_j} \right|$$

Jól láthatóan az index százalékos eredményt ad, hiszen százalékos különbségek abszolút értékeinek összegeként áll elő. a_i -vel az i . kategória relatív jövedelmét, b_i -vel az i . kategória relatív gyakoriságát jelölve (ami decilisekről lévén szó esetünkben mindvégig 10%):

$$RH = \frac{1}{2} \sum_{i=1}^N |a_i - b_i|$$

Megjegyezzük, hogy az $\frac{1}{2}$ -el történő szorzás azért szükséges, mert az átcsoportosítás esetén az átcsoportosítandó jövedelmet kétszer számol-

tuk, egyszer amikor az alacsonyabb kategóriákban lévőknek „odaadtuk”, illetve akkor is, amikor a magasabb jövedelmű egyénektől „elvettük”.

További érdekes következményt kapunk, ha az átcsoportosítandó jövedelmet más alakban írjuk fel. Ha a „minden ember egyenlő” elvet alkalmazzuk, akkor n egyén esetén az egy főre eső jövedelem nagysága $\frac{1}{n}$ lenne. Ehhez képest a szegényebb csoportokban az egyének jövedelme x_i , ahol $x_i < \frac{1}{n}$, így az átcsoportosítandó, „a szegényeknek adandó” jövedelem nagysága (rendezett jövedelem-minta esetén) a $j = \max \{i : x_i < \frac{1}{n}\}$ index-jelöléssel az alábbi (lsd: [18]):

$$\sum_{i=1}^j \left(\frac{1}{n} - x_i \right)$$

Ezt felhasználva az átcsoportosítandó jövedelem arányára $\hat{x} = \frac{j}{n}$ választással a következőt kapjuk:

$$\frac{\sum_{i=1}^j \left(\frac{1}{n} - x_i \right)}{\sum_{i=1}^n x_i} = \frac{\frac{j}{n}}{\sum_{i=1}^n x_i} - \frac{\sum_{i=1}^j x_i}{\sum_{i=1}^n x_i} = \tilde{L}(\hat{x}) - L(\hat{x})$$

$\tilde{L}(x)$ az egyenletes eloszlás Lorenz-görbét jelöli. A definícióból adódóan \hat{x} olyan, amely maximalizálja az $\tilde{L}(x)$ és $L(x)$ távolságát, hiszen $x_i > \frac{1}{n}$ esetén az átcsoportosítandó jövedelemösszegben negatív tagok is megjelennek, ha pedig j nem a maximális olyan index, ami teljesíti a feltételt, akkor további jövedelemegyenlőtlenség figyelhető meg, amely $L(x)$ -nek $\tilde{L}(x)$ -től történő távolodását implikálja.

Összefoglalva tehát: a Robin Hood-index a fenti definíciója mellett az alábbi módon is megadható:

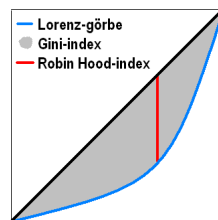
$$RH = \max_x \left(\tilde{L}(x) - L(x) \right)$$

Gini-index											
Év	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005
Decilisek	41,8%	46,7%	45,8%	46,8%	48,4%	49,4%	47,6%	47,3%	48,2%	47,3%	48,6%
Teljes minta	42,8%	48,2%	47,2%	48,2%	49,9%	51,0%	49,1%	48,8%	49,8%	48,9%	50,3%

Robin Hood-index											
Év	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005
Decilisek	30,4%	34,0%	33,8%	34,7%	35,9%	36,8%	35,5%	35,3%	36,1%	35,3%	36,4%

2.13. ábra. A Gini-index értéke decilisek és a teljes minta figyelembe vételével, illetve a Robin-Hood-index decilisek esetén

Kiszámolva a Robin Hood-indexet minden évre, ugyanazt tapasztaljuk, mint a Lorenz-görbe és a Gini-index esetén, nevezetesen: időben nagyjából konstansnak tekinthető az értéke (2.13 ábra). Ez nem meglepő, hiszen a három mutató kapcsolatát a szematikus 2.14-es ábrán tekinthetjük meg.



2.14. ábra. A Lorenz-görbe, a Gini-index és a Robin Hood-index kapcsolata

Természetesen rengeteg egyéb számszerű mutatót definiálhatnánk még, néhányat említve a teljesség igénye nélkül: a legfelső és a legalsó decilis átlagának hányadosa; az átlag feletti és az átlag alatti jövedelmek hányadosa (*Éltető-Frigyes-index*), *Theil-index*, stb. (részletesen lsd. pl: [19]). Azonban nem ez az elsődleges célunk, ráadásul megjegyeznénk, hogy további indexek kiszámolása esetén sem jutnánk alapvetően új következtetésre.

Bármennyire meglepőnek tűnik tehát, az intuíciónkkal ellentétben mégis azt kell, hogy mondjuk: aggregált szinten alapvetően nem növekedtek lényegesen a társadalmi különbségek a fenti magyar korosztály adatait megvizsgálva az adott időszak során.

Érezzük azonban, hogy ezzel a vizsgálat nem érhet véget, főként azért, mert adataink lehetővé teszik, hogy egyéni jövedelempályákat is figyelembe vehessünk, és azt valószínűsíthetjük, hogy a valódi egyenlőtlenséget az eloszlás belső szerkezetében bekövetkezett változás eredményezi. Ezt a vélekedést *Tóth I. GY. [2003]* cikkében is megtaláljuk. Ennek elemzését a következő fejezetben végezzük el.

3. fejezet

A jövedelmek belső struktúrája

3.1. Bevezetés

A jövedelmek eloszlásvizsgálata után most áttérünk az eloszlás belső szerkezetének megvizsgálásához. A panel-szerkezetű adataink lehetővé teszik, hogy nem pusztán az aggregált eloszlást, hanem az egyéni pályákat is figyelembe vegyük. A vizsgálatok módszertanához a Markov-modelleket használjuk fel. Elsőként az egyszerű Markov-lánc modellt illesztjük adatainkra, majd ezután egy általánosabb (és reálisabb) illesztést, a mover-stayer modellt alkalmazzuk. A módszertani felépítést lényegében a Major Klára által szerkesztett Markov-modellek című kötet alapján végeztük (ld: *Major Klára (Szerk.) [2008]*).

A Markov-modellek elméleti megalapozása a múlt század elején kezdődött, gyakorlati alkalmazásuk pedig több mint 50 évre tekint vissza. A nemzetközi szakirodalomban elsősorban területi, regionális vizsgálatokra alkalmazzák, de valójában jól alkalmazható olyan esetekre is, amelyek egy adott minta szerkezeti változását mérik. Különösen népszerűek manapság a jövedelmi dinamikához kapcsolódó alkalmazások. Ennek köszönhető a modell közgazdasági területen történő széles felhasználása. Ez adta számunkra is a fő motivációt az alkalmazáshoz.

3.2. Markov-lánc modell

3.2.1. A modell felépítése

A Markov-lánc lényege, hogy magát a változás mikéntjét, a folyamat dinamikáját ragadjuk meg, tehát számunkra nem az adott állapotok, hanem az *állapot változása* a hangsúlyos. A modell felépítéséhez először tehát meg kell határoznunk, hogy minek a változását vizsgáljuk, és ebből adódóan, hogyan definiáljuk magukat az állapotokat.

A vizsgált objektum a különböző időpontokban megfigyelt sokasági eloszlás, azaz a sokaság megoszlása a vizsgálati jellemző szerint egy adott időpontban. Az állapotok pedig azok a kategóriák, amelyekbe a megfigyelési egységeinket besoroltuk. Formálisan tehát az alábbi definícióinkat fogalmazhatjuk meg.

Sztochasztikus folyamatnak nevezzük a t paramétertől függő valószínűségi változók ξ_t sorozatát. Feltesszük, hogy a ξ_t valószínűségi változó értékeit minden pillanatban az $S = \{x_1; x_2; \dots; x_n\}$ halmazból veszi fel. Ekkor S -et ξ_t állapotterének, x_i -ket pedig a lehetséges állapotoknak nevezzük.

Legyen ξ_t egy diszkrét idejű sztochasztikus folyamat. Ha teljesül, hogy a folyamat jövőbeli állapota csak a jelenen keresztül függ a múlttól, akkor a ξ_t folyamatot *Markov-láncnak* nevezzük. Formálisan, $\forall t$ -re, $x_i \in S$ -re és $i(\tau) \in \{1, 2, \dots, n\}$ esetén:

$$P \left(\xi_{t+1} = x_j^{(t+1)} \left| \bigcap_{\tau=0}^t \left\{ \xi_\tau = x_{i(\tau)}^{(\tau)} \right\} \right. \right) = P \left(\xi_{t+1} = x_j^{(t+1)} \left| \xi_t = x_i^{(t)} \right. \right)$$

A sokasági eloszlást egy olyan vektorral azonosítjuk, ami az adott kategóriához tartozás valószínűségeit tartalmazza: $p = (p_1; p_2; \dots; p_n)$, ahol p_i az i . állapotba tartozás valószínűsége.

Fontos feltevés, hogy véges sok kategória legyen, és a kategóriák uniója, illetve a vizsgálati tartomány között bijekciónak kell lennie, hogy minden

egyén pontosan 1 csoportba tartozzon. Így nyilvánvalóan:

$$\sum_{i=1}^n p_i = 1$$

Ezeket felhasználva definiálhatjuk az átmenet-valószínűség mátrixot.

Sztochasztikus mátrixnak nevezzük az olyan A négyzetes mátrixot, amelynek soraiban szereplő elemei összege 1 és minden eleme nemnegatív, azaz minden i -re és $a_{ij} \geq 0$ -ra:

$$\sum_{j=1}^n a_{ij} = 1$$

Összehasonlítva a definíciót a fentivel, jól látható, hogy a sztochasztikus mátrixok sorai értelmezhetőek eloszlásvektorként. A mátrixszorzás szabályaiból adódóan következik az is, hogy két sztochasztikus mátrix szorzata is sztochasztikus mátrix lesz.

Legyen az A egy olyan $n \times n$ -es sztochasztikus mátrix, amelynek i . sorában és j . oszlopában található p_{ij} eleme azt mutatja meg, hogy a jelenben i . állapotban lévő egyén milyen valószínűséggel lesz a következő időpontban a j . állapotban, azaz: $p_{ij}(t) = P(\xi_{t+1} = j \mid \xi_t = i)$. Ekkor A -t átmenet-valószínűség mátrixnak (*átmenet-mátrixnak*) nevezzük, amelynek elemei tehát feltételes valószínűségek.

A definícióból következően a mátrix sorai az i . állapotból indulás, oszlopai a j . állapotba érkezés valószínűségeit határozzák meg minden lehetséges kimenetelre. Így a sorok például a következő időszakra vonatkozó eloszlást mutatják meg, feltéve, hogy az i . állapotból indultunk.

A sokasági eloszlás p vektorának időtől való függését hangsúlyozva, p_t -vel jelölve a t . időpontra vonatkozó eloszlást, könnyen láthatjuk, hogy a fent definiált A átmenet-mátrix kielégíti a következő rekurziót, ami egyben a Markov-láncok alapegyenlete:

$$p_{t+1} = p_t A$$

A kezdeti eloszlást p_0 -val jelölve, a fenti rekurzió alapján a t . időpontra vonatkozó eloszlás

$$p_t = p_0 A^t$$

alakban írható fel, ahol A definíciójából következően A^t is átmenet-valószínűségi mátrixként értelmezhető, azaz tetszőleges időpontra vonatkozó eloszlás megkapható a kezdeti eloszlás és az átmenet-mátrix hatványának szorzataként, ahol az átmenet-mátrix hatványának a_{ij} eleme azt mutatja meg, hogy t lépés során (tehát t periódus múlva) milyen valószínűséggel leszünk a j . állapotban, feltéve, hogy kezdetben az i . állapotból indultunk.

Érdekes kérdés lehet, hogy általában létezik-e olyan p^* eloszlás, amire teljesül a

$$p^* = p^* A$$

összefüggés, azaz van-e egyensúlyi eloszlás, továbbá ettől függetlenül vizsgálhatjuk azt is, hogy konvergencia mikor teljesül, azaz $k \rightarrow \infty$ esetén léteznek-e a $p_k \rightarrow \hat{p}$ és a $A^k \rightarrow A^*$ határértékek.

Belátható, hogy ha az A mátrix reguláris, akkor a határérték létezik, egyértelmű, és ebben az esetben $p^* = \hat{p}$, továbbá A^* sorai A egyensúlyi eloszlását adják meg, tehát reguláris átmenet-mátrixokra a határeloszlás egyben az egyensúlyi eloszlás lesz (ld: *Stokey – Lucas [1989]*).

3.2.2. Mobilitás

Az átmenet-mátrixok egyik legfőbb jellemzője a belőlük kiolvasható mobilitás. Ennek mérésére sokféle mutatót találhatunk a szakirodalomban. Erről részletes összefoglaló található például *Satya [2009]* írásában. Mi alapvetően két mutatót vizsgálunk meg. Az első a mobilitást a „nem-immobilitás” segítségével adja meg, a második mutató pedig valóban a mobilitást számszerűsíti.

A mátrix főátlójában lévő elemek az adott állapotban maradás valószínűségét adják meg, ezt felhasználva megadhatjuk a folyamatunk mobilitását az adott mátrix segítségével: az A mátrix nyomát $tr(A)$ -val jelölve:

$$\mu(A) = \frac{n - \text{tr}(A)}{n - 1}$$

Jól láthatóan a fenti kifejezés két dologtól függ: a főátló elemeitől, illetve a mátrix méretétől.

$\mu(A) = 0 \Leftrightarrow \text{tr}(A) = n$, ami átmenetmátrixról lévén szó, azt jelenti, hogy $A = I$, azaz A egységmátrix, ami a teljes immobilitást valósítja meg: a folyamat determinisztikusan stagnál.

$\mu(A) = 1 \Leftrightarrow \text{tr}(A) = 1$, ami például akkor valósul meg, amikor a mátrix minden eleme egyenlő, azaz $n \times n$ -es esetben minden eleme $\frac{1}{n}$. Ekkor $\text{tr}(A) = n \cdot \frac{1}{n} = 1$, ami pedig bizonyos értelemben a teljes mobilitást valósítja meg: a jelen ismeretében semmit sem tudunk mondani a jövőről, mert bármely állapotba ugyanolyan eséllyel léphetünk a következő időpontban.

Itt fontos megjegyeznünk, hogy mobilitáson azt értjük, hogy a jelen milyen mértékben determinálja a jövőt. Ezért lehetséges, hogy bár a nyom elméletben lehet 1-nél kisebb is, mi mégsem azt tekintjük abszolút mobil rendszernek, hanem a teljes bizonytalanságot. A gyakorlatban, így a mi vizsgálódásunk esetében is általában $\mu(A) \leq 1$ teljesül.

Minél inkább immobilis a rendszer (azaz az állapotváltás kis valószínűségű), annál nagyobb elemek vannak a főátlóban, így annál nagyobb a mátrix nyoma, ezáltal pedig a fenti tört értéke annál kisebb, ezért a mobilitási mutató annál inkább nagyobb, minél mobilisabb a rendszer.

Ez egy jó jellemzése lehet a folyamatnak és az átmenetmátrix struktúrájának, azonban érdemes megjegyezni, hogy a mérettől függés miatt önmagában ez a szám csak akkor mond valamit, ha a mátrix méretét is figyelembe vesszük a mutatóból levont következtetések során, hiszen ha több kategóriába soroljuk ugyanazt a mintát, akkor ezáltal magasabb mobilitást figyelhetünk meg.

A μ mutató a mobilitást a főátló elemei segítségével ragadta meg, lényegében a főátlóban lévő elemek immobilitást felhasználva valójában azt vizsgálta, mennyire nem immobilis a rendszer. Általában hasznos és jól al-

kalmazható, azonban nem minden esetben pontos, hiszen lényegében teljesen figyelmen kívül hagyja a főátlón kívüli elemeket, ráadásul igen érzékeny a mátrix méretére.

Ugyanolyan mobilitást definiál az alábbi két mátrix, holott szemmel láthatóan semmiképpen sem szeretnénk hasonló mobilitást tulajdonítani nekik:

$$M = \begin{bmatrix} 0,6 & 0,3 & 0,1 \\ 0,3 & 0,5 & 0,2 \\ 0,1 & 0,2 & 0,7 \end{bmatrix} \quad N = \begin{bmatrix} 0,6 & 0 & 0,4 \\ 0,5 & 0,5 & 0 \\ 0 & 0,3 & 0,7 \end{bmatrix}$$

Ezért a mobilitást máshogy is megvizsgáljuk. A gyakorlati elemzések során az egyik leghasznosabb a széles körben alkalmazott Bartholomew-index, amelynek alakja az alábbi:

$$B(M) = \frac{\sum_{i,j}^n |i-j| m_{ij}}{n}$$

Ha $i = j$, akkor az összeg adott tagja 0, ezért valóban csak a főátlón kívüli elemeket vesszük számításba, méghozzá súlyozzuk őket annak megfelelően, milyen távol vannak a főátlótól, azaz szemben a korábbi mutatókkal, a „nem-immobilitás” helyett valóban a mobilitást vizsgáljuk, ráadásul úgy, hogy a mobilitás mértékét is figyelembe vesszük.

Megjegyezzük, hogy $|i-j|$ helyett más súlyozást is használhatnánk, amely a mobilitástól „elvárt” tulajdonságokat kielégíti (ld: *Satya [2009]*). Például az $|i-j|^\alpha$, $\alpha > 1$ is megfelelő választás, amely a változás hatását felerősítve méri.

Mi úgy szeretnénk a mobilitást definiálni, hogy a gyakorlatban tapasztalt jövedelem-dinamikát kifejező mátrixokra értéke 0 és 1 közé essen úgy, hogy 0 értéket a teljes immobilitás, 1 értéket pedig a teljes mobilitás esetén vegyen fel. Teljes mobilitás alatt ugyanúgy a „kiszámíthatatlanságot”, tehát a jövőre vonatkozó teljes bizonytalanságot értjük. Ekkor az átmenet-mátrix ($n \times n$ -es

esetben) a csupa $\frac{1}{n}$ elemű mátrix. Könnyen belátható, hogy ebben az esetben a Bartholomew-index értéke: $\frac{n^2-1}{3n}$.

Jelöljük E -vel a csupa 1 elemű mátrixot, így az új ν mobilitási mutatónkat az $n \times n$ -es A átmenet-mátrixra az alábbi módon adhatjuk meg:

$$\nu(M) = \frac{B(M)}{B(\frac{1}{n}E)} = \frac{B(M)}{\frac{n^2-1}{3n}}$$

Természetesen a fenti mutató értéke 1-nél nagyobb is lehet, azonban a jövedelmek vizsgálatok előforduló átmenet-mátrixok esetében tipikusan 0 és 1 közötti értéket kapunk.

3.2.3. A felosztás szerkezete

A jövedelmek dinamikájának vizsgálatok jól láthatóan a helyzet bizonyos szempontból speciális. Olyan átmeneteket szeretnénk elemezni, amely az eloszlás általános, aggregált tulajdonságainak változásától a lehető legkevésbé függenek. Erre azért van szükségünk, mert az adatok belső szerkezetének elemzésekor a külső és belső tényezők együttes változásának figyelembevétele erősen torzítja az eredményt.

Az eloszlás főbb tulajdonságait és annak változását már külön megvizsgáltuk, most pedig szeretnénk ettől függetlenül megvizsgálni a belső jövedelemszerkezetet. Az eloszlás dinamikájától való függést úgy szüntetjük meg, hogy minden egyes adott évben úgy osztjuk fel jövedelmi kategóriák között az egyéneket, hogy minden csoportba azonos számú megfigyelés essen, a csoportjaink pedig minden évre az aktuális eloszlás ugyanazon részét reprezentálják. Ezzel elérjük azt, hogy egy adott év adott jövedelmi viszonyait megtestesítő csoportját egy tetszőleges másik év ugyanolyan viszonyait kifejező csoportjával hasonlíthassuk össze. Ezt az eloszlásfüggvény inverze szerinti felosztás-struktúra teszi lehetővé. n kategória esetén tehát az j . kategóriát

definiáló intervallum a következő:

$$\left[F_i^{-1} \left(\frac{j-1}{n} \right); F_i^{-1} \left(\frac{j}{n} \right) \right),$$

ahol F_i az i . évi jövedelem (mint valószínűségi változó) eloszlásfüggvénye.

Ejtsünk néhány szót a kategóriák számáról. Olyan sok egyénünk van (minden évre több, mint 100000 adat), hogy nem lehet külön-külön tekinteni minden egyes ember jövedelempályáját, ezért kategóriákat hozunk létre, és az egy kategóriába tartozókat lényegében azonosnak, homogénnek tételezzük fel.

A túl sok kategória nehezen értelmezhetővé teszi a különbségeket: 100 részre osztás esetén pl. a 66. csoportba tartozókat a 67. csoport tagjaitól a valóságban elég nehéz lenne megkülönböztetni. Ráadásul a sok részre osztás azt eredményezi, hogy egy-egy csoportba kevés megfigyelés esik, ezért nehezebb általános következtetéseket levonni.

A kevés felosztás viszont összemoshat olyan különbségeket, amik számunkra szignifikánsak lehetnek. Ezért jó választás lehet a jövedelem-decilisek használata. Statisztikai elemzések során sok egyéb tekintetben is ezek a használatosak. A szakirodalomban jövedelmekkel kapcsolatos vizsgálatokhoz gyakran használt még a kvintilisek (5 egyforma részre osztás) is. Látni fogjuk, hogy számunkra hol az egyik, hol a másik felosztás vizsgálata lesz a célszerűbb.

Tehát pl. 10 csoportba osztás esetén kiszámoljuk minden évre az aktuális jövedelmi deciliseket és ezen decilisek határait. Ezzel azt értük el, hogy minden évre ugyanúgy egy diszkrétizált egyenletes eloszlás jellemzi a mintát, azonban az, hogy az adott évben ki melyik csoportba kerül, az természetesen függ attól, hogyan változnak a jövedelmi határok. Röviden tehát: az eloszlás változása és a csoportok jövedelmi határainak állandósága helyett az eloszlás állandóságát, és a jövedelemhatárok változását tekintjük kiindulópontunknak.

Ez sok szempontból igen hasznosnak bizonyul: ezzel elértük, hogy a változás struktúráját az eloszlás változásától függetlenül meghatározzuk, másrészt egyforma elemszámú csoportok révén a kategóriák azonos súllyal szerepelnek, így a csoportokat reprezentáns elemmel helyettesítve általános összefüggéseket tárhatunk fel a jövedelmi kategóriák között.

3.2.4. A modell illesztése jövedelmekre

Az átmenet-valószínűségi mátrix elemeit könnyedén megbecsülhetjük maximum-likelihood módszerrel. Jelölje n_{ij} azon egyének számát, akik a jelenben az i ., a következő időszakban a j . állapotban vannak. N -el jelölve azon (i, j) párok összességét, amelyekre $p_{ij} > 0$, a likelihood függvény alakja:

$$L = \prod_N p_{ij}^{n_{ij}}$$

Tehát a log-likelihood függvény:

$$\ln L = \sum_N n_{ij} \ln p_{ij}$$

Felhasználva, hogy $\sum_j p_{ij} = 1$, p_{ij} -re a becslésünk:

$$\hat{p}_{ij} = \frac{n_{ij}}{\sum_j n_{ij}}$$

Tehát az átmenet-valószínűségek becslhetőek relatív gyakorisággal: az i . állapotból a j . állapotba történő tényleges átmeneteket felhasználva.

Az átmenet-mátrixokat ki lehet számolni külön-külön minden évre is, azonban hosszútávra vonatkoztatva pontosabb becslést akkor kapunk, ha a teljes adatstruktúrát figyelembe vesszük, azaz minden év adatait felhasználva egyetlen átmenet-mátrixot illesztünk. Ezt elkészítettük 5 és 10 kategóriára bontás (azaz jövedelmi kvintilisek és decilisek) esetén, illetve a mintát férfiakra és nőkre szétbontva külön-külön is.

Az adatbázis hiányossága miatt további kétféle becslést készítettünk. Az egyik esetben a hiányos megfigyeléseket külön kategóriának definiáltuk, ezáltal 6 illetve 11 kategóriánk lett, a másik esetben pedig a becslés során csak a teljesen hiánytalan megfigyelésekkel rendelkező egyéneket vesszük figyelembe. Ezzel ugyan adatokat veszítünk, de még így is több, mint 60000 elemű mintával dolgozunk. Az eredményekből az alábbi következtetéseket vonhatjuk le.

A hiányos esetre becsült 6×6 -os átmenetmátrix bal felső 5×5 -ös blokkját összehasonlítva a hiánytalan adatokból becsült 5×5 -ös mátrixszal, jól láthatóan a mátrix elemei természetesen eltérnek ugyan, de a mátrix szerkezete változatlan: a megfelelő elemek aránya nagyjából hasonlóknak tekinthető (3.1 ábra). Hasonló megállapítást tehetünk a 11×11 -es mátrix felső blokkjának és a 10×10 -es esetnek az összehasonlításából is. Ráadásul a későbbi számolásainkhoz a hiányos állapot definiálása, és az ebből nyert számítások nehezen értelmezhetőek lennének, ezért a hiánytalan adatokból nyert becslést tekintjük mérvadónak, és a továbbiakban ezzel dolgozunk.

	1	2	3	4	5	Üres
1	50,5%	19,1%	8,5%	2,8%	1,6%	17,4%
2	19,6%	53,4%	14,2%	3,7%	2,1%	7,0%
3	7,4%	16,1%	54,8%	14,2%	3,2%	4,3%
4	2,1%	3,1%	15,5%	64,3%	12,4%	2,4%
5	1,7%	2,1%	3,2%	12,7%	78,0%	2,3%
Üres	17,8%	6,9%	4,3%	2,7%	3,0%	65,3%

	1	2	3	4	5
1	56,5%	24,1%	12,6%	4,4%	2,4%
2	17,2%	58,8%	16,9%	4,5%	2,6%
3	5,8%	15,6%	59,2%	15,9%	3,5%
4	1,4%	2,6%	15,1%	68,2%	12,6%
5	1,1%	1,8%	3,0%	13,0%	81,1%

3.1. ábra. A hiányos, illetve a hiánytalan mintára becsült átmenet-mátrixok

A következő megállapításunk ugyanazon minta különböző felosztásainak összevetése, azaz a decilisek és a kvintilisek összehasonlítása (3.2 ábra). Szemügyre véve a két mátrixot észrevehetjük, hogy a decilisekre becsült átmenet-mátrix főátlójában az elemeink jóval kisebbek, ám ez nem meglepő, hiszen 10 kategória esetén nagyobb eséllyel válthatunk jövedelmi csoportot, hiszen a kategóriahatárok közelebb esnek egymáshoz. A mátrix szerkezete azonban itt sem változott: a főátló domináns elemei alatt és felett lévő elemek még viszonylag jelentős mértékű átmenetet prognosztizálnak, azonban a 2, vagy annál nagyobb kategóriaváltás igen kis eséllyel valósulhat meg.

	1	2	3	4	5	6	7	8	9	10
1	47,0%	18,1%	9,6%	10,1%	5,7%	3,5%	2,2%	1,4%	1,1%	1,3%
2	15,7%	34,8%	12,0%	15,2%	9,3%	5,7%	3,2%	1,8%	1,1%	1,2%
3	6,3%	15,9%	34,2%	24,6%	7,9%	4,5%	2,6%	1,5%	1,2%	1,2%
4	3,6%	10,4%	14,0%	44,7%	14,2%	5,6%	3,0%	1,8%	1,2%	1,4%
5	2,1%	6,0%	5,8%	18,1%	39,8%	16,5%	5,8%	2,8%	1,7%	1,5%
6	1,1%	2,7%	2,3%	5,8%	18,4%	43,4%	17,4%	5,0%	2,3%	1,6%
7	0,6%	1,2%	1,0%	2,3%	5,1%	18,5%	47,8%	17,7%	3,8%	1,9%
8	0,4%	0,7%	0,7%	1,3%	2,1%	4,6%	19,1%	51,6%	16,2%	3,2%
9	0,4%	0,6%	0,5%	1,1%	1,3%	2,1%	3,7%	17,7%	60,0%	12,7%
10	0,5%	0,6%	0,6%	1,3%	1,1%	1,4%	1,7%	3,0%	13,4%	76,2%

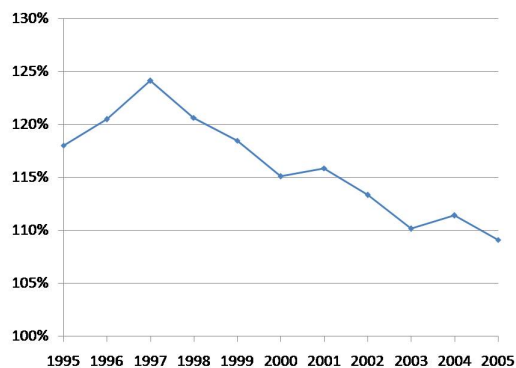
	1	2	3	4	5
1	56,5%	24,1%	12,6%	4,4%	2,4%
2	17,2%	58,8%	16,9%	4,5%	2,6%
3	5,8%	15,6%	59,2%	15,9%	3,5%
4	1,4%	2,6%	15,1%	68,2%	12,6%
5	1,1%	1,8%	3,0%	13,0%	81,1%

3.2. ábra. A decilisek, illetve a kvintilisek esetén becsült átmenet-mátrixok

Érdekességként észrevehetjük, ami talán a decilisekre becsült 10×10 -es mátrixból jobban látszik: a főátlóban lévő elemek, amelyek az immobilitást jelentik, nem egyenlően oszlanak meg a jövedelmi kategóriák között. A szélsőséges, tehát a legjobb és legrosszabb csoportokban magasabb értéket figyelhetünk meg, mint a középső jövedelmi csoportokban, ez a magasabb kategóriákra hatványozottan igaz.

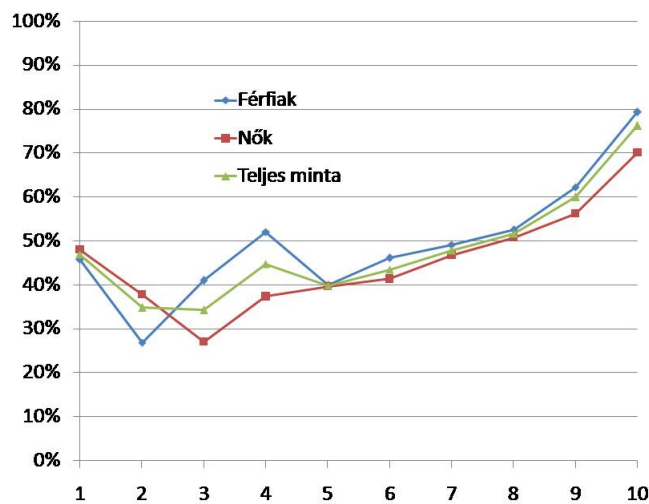
Ez a ferde „mosoly-szerkezet” mutatja, hogy a társadalom legalján és főként legtetején nagyfokú beragadás figyelhető meg: a legszegényebbek és a leggazdagabbak várhatóan megmaradnak saját társadalmi csoportjukban, tehát legalulról felemelkedni, és főként legfelülről visszaesni igen kis eséllyel lehet. Gyaníthatóan ez a jelenség a rendszerváltást követően végig, a vizsgált időszakunk után is fennmarad. Ezáltal „társadalmi-olló” keletkezhet: a szegényebb rétegekből felemelkedni, és a gazdagabbak közül visszaesni jóval nehezebb, miközben a középosztály nagyobb fokú mobilitása azt eredményezheti, hogy egyrészük várhatóan felzárkózik a gazdagabb csoportokhoz, másik részük visszaesik, a középosztály hiánya pedig növelheti a társadalmi különbségeket.

További érdekes dolgot vehetünk észre, ha a mobilitást nemenként is külön-külön vizsgáljuk. Ismert jelenség, hogy a nők jövedelmi viszonyai meghatározó mértékben rosszabbak a férfiakénál: keresetük jelentősen alulmúlja a férfiakét: a vizsgált időszakban a férfiak keresete átlagosan megközelítőleg 30 – 40%-al magasabb volt a nőkéénél. Megjegyezzük azonban, hogy a jövedelmi különbségek a vizsgált időszak során jelentősen csökkentek: a férfiak és a nők átlagos kategóriájának arányából ez kiválóan látszik (3.3 ábra).



3.3. ábra. A férfiak és a nők átlagos kategóriaszámának aránya

A mobilitást tekintve azt vehetjük észre, hogy a felsőbb jövedelmi csoportokban enyhén, a közepesen alacsony jövedelmi csoportokban pedig jelentősen mobilisebbeknek tekintetők a nők (3.4 ábra). Egyedül a két legalacsonyabb kategóriában mondató el, hogy a férfiak nagyobb eséllyel váltanak jövedelmi kategóriát. Jól láthatjuk tehát, hogy az az általános vélekedés, miszerint a nők kevésbé mobilisebbek, nem bizonyul helytállónak. A legalacsonyabb csoportokban sem sokkal immobilisebbek a nők, így a nők alacsonyabb mobilitása is inkább csak a közepesen alacsony kategóriákhoz képest igaz. Ékes példáját kaptuk tehát annak, hogy miért érdemes külön vizsgálni magát az eloszlást és annak belső szerkezetét.



3.4. ábra. Az immobilitás mértéke a teljes mintában, illetve nemenként

Az igazsághoz hozzátartozik azonban, hogy a mobilitás nem feltétlenül jelent jót, sőt: a nők mobilitása az átmenet-mátrixokból is jól láthatóan inkább

enyhén negatív irányban hat: a férfiak immobisebbek ugyan, a nők nagyobb mobilitásában viszont némileg jelentősebb részben a kategóriák közötti lefelé lépés a meghatározó (3.5 ábra).

	1	2	3	4	5	6	7	8	9	10
1	48,1%	20,3%	8,2%	8,8%	5,6%	3,6%	2,3%	1,2%	0,8%	1,0%
2	14,2%	37,9%	10,3%	13,6%	10,1%	6,5%	3,4%	1,7%	1,1%	1,2%
3	6,7%	21,5%	27,0%	22,1%	9,8%	5,9%	3,2%	1,6%	1,2%	1,1%
4	3,6%	14,6%	13,3%	37,4%	16,6%	6,7%	3,5%	2,0%	1,1%	1,2%
5	1,8%	7,3%	6,0%	16,9%	39,7%	16,7%	6,1%	2,8%	1,4%	1,3%
6	0,9%	3,2%	2,2%	6,2%	18,6%	41,4%	18,5%	5,3%	2,1%	1,6%
7	0,5%	1,3%	0,9%	2,1%	6,2%	17,9%	46,7%	18,9%	3,6%	1,9%
8	0,4%	0,8%	0,6%	1,1%	2,5%	5,9%	18,9%	50,7%	15,9%	3,2%
9	0,5%	0,8%	0,6%	1,1%	1,5%	3,0%	5,0%	18,7%	56,2%	12,7%
10	0,7%	1,2%	0,8%	1,6%	1,7%	2,3%	2,7%	4,6%	14,2%	70,1%

	1	2	3	4	5	6	7	8	9	10
1	45,7%	15,6%	11,1%	11,7%	5,7%	3,4%	2,2%	1,6%	1,5%	1,6%
2	19,4%	26,9%	16,4%	19,2%	7,0%	3,8%	2,7%	1,9%	1,3%	1,4%
3	6,0%	10,7%	41,0%	26,9%	6,1%	3,2%	2,1%	1,4%	1,2%	1,3%
4	3,6%	6,3%	14,8%	52,1%	11,7%	4,5%	2,4%	1,7%	1,3%	1,6%
5	2,6%	3,8%	5,5%	20,1%	39,9%	16,1%	5,3%	2,9%	2,0%	1,8%
6	1,3%	1,9%	2,5%	5,3%	18,0%	46,1%	16,1%	4,6%	2,5%	1,6%
7	0,7%	1,0%	1,2%	2,5%	3,9%	19,3%	49,1%	16,3%	4,1%	2,0%
8	0,5%	0,6%	0,8%	1,3%	1,8%	3,5%	19,4%	52,5%	16,4%	3,3%
9	0,4%	0,4%	0,5%	1,1%	1,2%	1,6%	2,8%	17,1%	62,3%	12,7%
10	0,4%	0,4%	0,6%	1,2%	0,8%	1,0%	1,3%	2,2%	13,0%	79,3%

3.5. ábra. A férfiak és a nők decilisekre becsült átmenet-mátrixa

Ha külön-külön tekintjük az átmenet-mátrix felső és alsó háromszögmátrixát, és ennek elemeit a Bartholomew-index során látott súlyozással összeadjuk, akkor megkapjuk egy mértékét a felfelé és a lefelé történő kategória-váltásnak. Ezek arányát tekintve külön férfiakra és nőkre, majd ezen arányokat egymáshoz viszonyítva 1,27-es értéket kapunk. Arányokról lévén szó ez szignifikánsan eltér az 1-től, méghozzá a férfiak javára. A nők magasabb mobilitása tehát valóban inkább lefelé irányul.

3.2.5. Előrejelzés

A modellt természetesen nem csak önmagában, hanem előrejelzésre is szeretnénk használni. Ehhez a modell bemutatásánál látott módon az átmenet-mátrixot hatványoznunk kell. Az előrejelzés jóságának ellenőrzéséhez kiszámoltuk az 1995-ről közvetlenül 2005-re történő besorolások közötti átmeneteket is, és összevetettük a megbecsült mátrix 10. hatványával: lényegében tehát a 10 év átmenetéből előre becsült és a 10 éves tapasztalati átmenet-mátrixot hasonlítottuk össze. A kvintilisekre elvégezve a fenti eljárást, láthatjuk, hogy a becsült mátrix 10. hatványának elemei elég közel esnek egymáshoz: a modell szerint tehát az 1995-ös jövedelmi helyzet igen kis mértékben determinálja csak a 2005-re várható jövedelmi csoportba történő besorolást (3.6 ábra). Lényegében tehát mindegy honnan indultunk, jó eséllyel 10 év elteltével akármelyik csoportba kerülhetünk. Érezhetően ez a gondolat kevésbé helytálló.

	1	2	3	4	5
1	13,1%	20,2%	23,9%	25,1%	17,7%
2	13,0%	20,1%	23,9%	25,2%	17,8%
3	12,4%	19,2%	23,6%	26,0%	18,9%
4	11,3%	17,8%	23,0%	27,2%	20,8%
5	10,6%	16,8%	22,4%	27,9%	22,3%

	1	2	3	4	5
1	25,7%	28,1%	21,4%	15,5%	9,4%
2	17,7%	28,6%	24,8%	18,4%	10,5%
3	12,5%	18,3%	29,8%	25,8%	13,6%
4	7,7%	11,0%	22,5%	35,3%	23,4%
5	5,7%	8,8%	11,0%	21,8%	52,6%

3.6. ábra. A Markov-lánc alapján becsült, illetve a valós átmenet-mátrix 1995 és 2005 összevetésében

A különbség valóban látványos: a tapasztalati átmenetmátrix szignifikánsan különbözik az előrejelzésből számítottól. Ennek magyarázatát az adja, hogy feltehetőleg olyan esetre alkalmaztuk a Markov-modellt, amelyben a modell feltevései jó eséllyel nem teljesülnek. Nevezetesen: a folyamat jövőbeli alakulása csak a jelenbeli állapotán keresztül függ a múlttól, tehát az, hogy melyik csoportban leszünk, az nem függ attól, hol voltunk korábban, csak attól, hol vagyunk most. Ez vélhetően hibás feltételezés.

A Markov-lánc modell tehát csak első közelítésnek jó, hosszútávon sajnos távol áll a valós helyzettől: a mobilitást szisztematikusan felülbecsüli. Ez például magyarázható lenne azzal, hogy a folyamat nem stacioner, tehát az i . állapot elhagyásának valószínűsége függ(het) attól, milyen régóta vagyunk az adott állapotban.

Ennek ellenére mégsem gondoljuk, hogy esetünkben a Markov-lánc modelljének a hosszútávon kevésbé jó becslési eredménye elsősorban a stacionaritás hiányára vezetődne vissza, mert ha az átmenet-mátrixokat minden évre külön-külön megbecsüljük, akkor sem kapunk szignifikánsan különböző mátrixokat.

A Markov-láncok modellje esetén a rossz becslési eredményt esetünkben feltételezhetőleg sokkal inkább egy más jellegű feltétel nem teljesülése okozza, nevezetesen: feltettük, hogy minden egyes kategóriában a csoportokra osztás után a csoporton belüli egyének már homogénnek tekinthetők, tehát akárhogy választunk reprezentánst közülük, ugyanolyan eséllyel lesz az elkövetkezendő időszakokban az adott állapotokban, mintha bármely más, vele egy kategóriában lévő egyént választottunk volna helyette.

Weber [2002] cikkében osztrák jövedelmi adatokra illesztett Markov-lánc modellt. Esetében a nemeken kívül más megfigyelhető, leíró statisztikai jellemzők is (pl: kor, iskolai végzettség, munkában eltöltött idő) rendelkezésre álltak. Ezt felhasználva megmutatta, hogy az egyént jellemző egyéb, számszerűen nem mérhető jellemzők (pl: képesség, szorgalom, szociális készség) legalább olyan fontosak, és meghatározóak lehetnek a jövedelmi mobilitás

szempontjából, mintha csak a „hagyományos” mutatókat tekintettük volna.

Éppen ezért a Markov-lánc modelljét nem időtől függő módon finomítjuk, hanem a populáció heterogenitását tételezzük fel, ezzel pedig eljutunk a mover-stayer modellig.

3.3. Mover-stayer modell

3.3.1. A modell felépítése

A mover-stayer modell a Markov-lánc modelljének egyfajta általánosítása: az egyének homogenitását oldjuk fel azért, hogy feltesszük: egyéneink mobilitás szempontjából kétfélek lehetnek, nevezetesen *moverek* (mozgók) és *stayerek* (maradók). A különböző mobilitás következménye, hogy nem lehet ugyanazt a modellt illeszteni a két almintára. Ahhoz, hogy a továbbiakban az egész mintát könnyen kezelhetően elemezzük, az almintákra feltevéseket teszünk.

A moverek-re úgy gondolunk, mint azok, akik a mozgást testesítik meg: dinamikájukat a fent bemutatott standard Markov-lánc modellel jellemezhetjük.

A stayerek, nevükből következően mindig ugyanabban az állapotban maradnak, tehát ők azok, akik egyáltalán nem váltanak állapotot.

A probléma az, hogy nem tudjuk az egyéneinkről eldönteni, hogy milyen típusúak, csupán a teljes minta jövedelmi dinamikáját ismerjük. A modell egyenlete így a következő alakban írható fel:

$$P = SI + (I - S)M$$

Itt P a teljes populáció átmenet-mátrixa, S egy olyan $n \times n$ -es diagonális mátrix, amelyben a főátló elemei az adott állapotban lévő stayerek aránya, n különböző állapot esetén. A stayerek egyáltalán nem váltanak állapotot, ezért átmenet-mátrixuk az I egységmátrix, M pedig a moverek átmenet-

mátrixa.

A fenti mátrixegyenlet tehát azt fejezi ki, hogy a teljes populáció dinamikáját a stayerek dinamikájának (I) és a moverek dinamikájának (M) súlyozott kombinációja adja.

A valóságban a bal oldalon lévő P elemeket tudjuk megfigyelni, S és M elemei ismeretlenek.

Ha nem pusztán egy periódust nézünk, akkor a fenti folyamatot úgy tehetjük többperiódusúvá, hogy a Markov-láncoknál látottak alapján a két mintára külön-külön hatványozzuk az átmenet-mátrixokat, ezzel megkapjuk a két kategóriára a megfelelő átmeneteket több periódus elteltével, majd a kapott mátrixok kombinációját tekintjük. Ez azért értelmes, mert a teljes mintát partícionálással osztottuk két részre. Így tehát P_T -vel jelölve a teljes populáció dinamikáját T átmenet után:

$$P_T = SI^T + (I - S)M^T$$

Mivel $I^T = I$ ezért a fenti egyenlet egyszerűbb felírása:

$$P_T = S + (I - S)M^T$$

Ezt felhasználva, vizsgáljuk meg mit mondhatunk a folyamat határeloszlásáról. Az egyszerű Markov-lánc esetében ennek nem szenteltünk kiemelt figyelmet, a mover-stayer modell esetében ezt vizsgálva azonban egy fontos következmény adódik.

Írjuk fel a modellünk alapján a T . átmenet utáni p_T eloszlást a kezdeti p_0 eloszlás segítségével:

$$p_T = p_0 P_T = p_0 (S + (I - S)M^T)$$

Számoljuk ki a határeloszlást, és az azt definiáló átmenet-mátrixot.

$T \rightarrow \infty$ esetén láttuk, hogy a stayerek I átmenet-mátrixának I^T hatványa is az egységmátrix, így semmit sem befolyásol, ezért elhagyható. A moverek

M átmenet-mátrixa hatványainak határértékét az A^* mátrix adta meg, ez most sincs másként, hiszen a moverek dinamikáját egyszerű Markov-lánc határozza meg:

$$\lim_{T \rightarrow \infty} M^T = A^*$$

P^* -al jelölve az egyensúlyi eloszláshoz tartozó átmenet-mátrixot, az egyensúlyi eloszlást felírva a következőket kapjuk:

$$p^* = p_0 P^* = p_0 (S + (I - S)A^*)$$

Már ebből az alakból is látszik a Markov-lánchoz képest a lényeges különbség: P^* egy olyan mátrix, amelyben minden sor különbözik, ezáltal a sorok különböző eloszlást is definiálnak. Esetünkben A^* sorait

$$p = (p_1; p_2; \dots; p_n)$$

vektorokkal jelölve:

$$P^* = \begin{bmatrix} s_1 + (1 - s_1)p_1 & (1 - s_1)p_2 & (1 - s_1)p_3 & \cdots & (1 - s_1)p_n \\ (1 - s_2)p_1 & s_2 + (1 - s_2)p_2 & (1 - s_2)p_3 & \cdots & (1 - s_2)p_n \\ (1 - s_3)p_1 & (1 - s_3)p_2 & s_3 + (1 - s_3)p_3 & \cdots & (1 - s_3)p_n \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ (1 - s_n)p_1 & (1 - s_n)p_2 & (1 - s_n)p_3 & \cdots & s_n + (1 - s_n)p_n \end{bmatrix}$$

A hosszútávú mobilitás megbecsléséhez kiszámoljuk P^* mobilitását a μ mobilitási mutatóval:

$$\mu(P^*) = \frac{n - \sum_{i=1}^n (s_i + (1 - s_i)p_i)}{n - 1} = \frac{n - \left(\sum_{i=1}^n s_i + \sum_{i=1}^n p_i - \sum_{i=1}^n s_i p_i \right)}{n - 1} =$$

$$\begin{aligned}
&= \frac{n-1 - \sum_{i=1}^n (s_i(1-p_i))}{n-1} = 1 - \frac{\sum_{i=1}^n (s_i(1-p_i))}{n-1} = 1 - \frac{\sum_{i=1}^n (s_i(1-\frac{1}{n}))}{n-1} = \\
&1 - \frac{\sum_{i=1}^n (s_i \frac{n-1}{n})}{n-1} = 1 - \frac{\frac{n-1}{n} \sum_{i=1}^n s_i}{n-1} = 1 - \frac{\sum_{i=1}^n s_i}{n} \leq 1
\end{aligned}$$

Egyenlőség pedig pontosan akkor teljesül, ha $\sum_{i=1}^n s_i = 0$, azaz egyetlen egy stayer sincs. Ekkor az S mátrix a nullmátrix, ezzel pedig a Markov-lánc modelljét kaptuk vissza.

Tehát amíg a Markov-lánc esetében a mobilitási mutató értéke 1-hez tart, hiszen az A^* mátrixunk nyoma 1, addig a mover-stayer modellben a mobilitási együttható határozottan 1-nél kisebb. Ezt egyébként onnan is tudhatjuk, hogy a modellünket eleve úgy alkottuk meg, hogy tartalmaz egy olyan alminitát, amely teljesen immobil.

Ezzel megkaptunk egy igen fontos következményt: míg a kezdeti állapottól való függés a Markov-lánc modelljénél teljesen eltűnik az idő múlásával, addig a mover-stayer modellre ez nem teljesül: a kezdeti csoportba történő besorolás mindig is meghatározó marad arra nézve, milyen valószínűségekkel kerülünk másik csoportba. A Markov-lánc esetében mindegy volt honnan indultunk, lényegében akárhová eljuthatunk, a mover-stayer modell esetében ugyanakkor a kezdeti besorolás némiképp determinálja a lehetőségeinket. Jobban belegondolva ez sokkal inkább megfelel a valós életben tapasztaltaknak.

3.3.2. A modell becslése

A mover-stayer modellnél is maximum-likelihood becslést alkalmazunk. A becslést *Frydman [1984]* alapján végezzük el. A log-likelihood függvény alakja

a Markov-láncoknál látottakhoz képest jóval bonyolultabb lesz: p_j -vel a kezdeti eloszlást, n_j^0 -al a kezdetben j . állapotban lévő egyéneink számát, a likelihood függvény:

$$\ln L = \sum_{j=1}^N n_j^0 \ln p_j + \sum_{j=1}^N \ln L_j$$

n_j -vel azon egyének számát jelölve, akik mindvégig a j . állapotban vannak, továbbá s_j -vel és m_{jk}^T -vel az S és az M^T mátrixok megfelelő elemeit, n_{jk} -val pedig a j . állapotból a k . állapotba történő átmenetek számát jelölve, L_j kifejtve a következő módon írható fel:

$$\begin{aligned} \ln L_j = & n_j \ln (s_j + (1 - s_j)m_{jj}^T) + (n_j^0 - n_j) \ln(1 - s_j) + \\ & + (n_{jj} - Tn_j) \ln m_{jj} + \sum_{k \neq j} n_{jk} \ln m_{jk} \end{aligned}$$

A likelihood függvény alakjából is látszik a modell ismertetésénél elmondottak igazolása: a kezdeti állapot kiemelt jelentőségű, a becslés a kezdeti állapotra, mint feltételre vonatkozik, így az alábbi feltételes becslés csak az adott kezdeti állapotra vonatkozóan érvényes.

A fenti egyenletet s_j szerint deriválva, majd 0-val egyenlővé téve kapjuk, hogy a maximalizálásnak eleget tevő s_j alakja az alábbi:

$$\hat{s}_j = \frac{n_j - n_j^0 m_{jj}^T}{n_j^0 (1 - m_{jj}^T)}$$

Ezt visszahelyettesítve a fenti egyenletbe, majd megfelelő deriválásokat és behelyettesítéseket végrehajtva megkapjuk a $\frac{\partial \ln L_j}{\partial m_{ji}} = 0$ feltételnek eleget tevő m_{ji} értékét:

$$\hat{m}_{ji} = n_{ji} \frac{1 - \hat{m}_{jj} - \sum_{\substack{k=i+1 \\ k \neq j}}^N \hat{m}_{jk}}{\sum_{\substack{k=i \\ k \neq j}}^N n_{jk}}$$

A fenti rekurziót minden i -re kiszámolva, n_j^* -al jelölve a j . állapot elhagyásának gyakoriságát $\left(n_j^* = \sum_{k=1}^N n_{jk}\right)$ majd az eredeti egyenletbe behelyettesítve kapjuk, hogy:

$$(n_j^* - Tn_j^0)\hat{m}_{jj}^{T+1} + (Tn_j^0 - n_{jj})\hat{m}_{jj}^T + (Tn_j - n_j^*)\hat{m}_{jj} + (n_{jj} - Tn_j) = 0$$

Kaptunk tehát egy $T + 1$ -ed fokú polinomot, amelyről megmutatható, hogy pontosan 1 gyöke esik a $[0, 1]$ intervallumba, így ezt numerikus úton meghatározva megkaphatjuk az \hat{m}_{ji} értékeinket.

Jól látszik azonban, hogy ez az eljárás eléggé bonyolult, azonban bizonyos esetekben némi egyszerűsítéssel is élhetünk. Ha kellően sok periódust tekintünk, akkor a fenti becslésekben az m_{jj}^T értékek a többi tagokhoz, illetve tényezőkhöz képest elhanyagolhatóan kicsivé válnak, hiszen $T \rightarrow \infty$ esetén $m_{jj}^T \rightarrow 0$ exponenciális sebességgel. Kellően nagy T esetén tehát nem okoz lényegi változást, ha az $m_{jj}^T = 0$ feltételezéssel élünk. Ezzel pedig becslőfüggvényeink jóval egyszerűbben írhatóak fel:

$$\hat{s}_j = \frac{n_j}{n_j^0}$$

$$(Tn_j - n_j^*)\hat{m}_{jj} + (n_{jj} - Tn_j) = 0$$

Ahonnán \hat{m}_{jj} -ot kifejezve:

$$\hat{m}_{jj} = \frac{n_{jj} - Tn_j}{n_j^* - Tn_j}$$

Az első becslés igen tömör és lényegre törő: azt állítja, hogy a stayerek arányát úgy becsülhetjük, hogy a mintában mindvégig az adott állapotban lévők számát osztjuk az adott állapotból induló összes egyén számával. Más-képp megfogalmazva mindezt: minden olyan egyénünk stayer, aki a megfigyelt időtartam alatt nem vált állapotot. A megfigyelési időszak növekedtével ez egyre inkább helytállónak bizonyul, hiszen aki mover, az 1 valószínűséggel előbb-utóbb kimozdul az adott állapotból.

A második becslés némiképpen „csalás”, hiszen \hat{m}_{jj} egyszerűsített értéke T -től is függ, azonban T -t növelve a konvergencia nagyságrendekkel lassabban, az exponenciálishoz képest pusztán csak hiperbolikus, (azaz $\sim \frac{1}{T}$) sebességgel történik, mint m_{jj}^T esetében, ezért van értelme a fenti kifejezésnek.

Itt mi tulajdonképpen nem is a precíz becslésre törekszünk, hiszen azt megtettük korábban, hanem valójában ötletet merítünk egy másik, algoritmikus becsléshez. Ez a Fuchs-Greenhouse által kimondottan a mover-stayer modellre kidolgozott *EM-algoritmus* (*Fuchs-Greenhouse [1988]*). Magát az EM-eljárást széles körben, egészen másféle becslésekhez is használják, mi most pusztán a számunkra adaptált változatának szükséges lépéseit ismergetjük. Az algoritmus 4 lépésből áll:

1. A modell paramétereire kezdeti becslést adunk.

χ -vel jelölve az indikátorfüggvényt, a fenti megjegyzéseink alapján legyenek ezek a következők:

$$s_j^{(0)} = \frac{n_j}{n_j^0}$$

$$m_{jk}^{(0)} = \frac{n_{jk} - \chi_{\{j=k\}} T n_j}{n_j^* - T n_j}$$

2. Megbecsüljük a stayerek számának várható értékét a paraméterek i . becslésének ismeretében.

Akik mindvégig a j . állapotban maradtak, azok egyrészt a stayerek (s_j),

másrészt azok a moverek, akik a T átmenet során (még) nem váltottak állapotot $((1 - s_j)m_{jj}^T)$.

Így azon stayerek várható számát, akik az s_j állapothoz tartoznak, n_{s_j} -vel jelölve:

$$n_{s_j}^{(i)} = n_j \frac{s_j^{(i)}}{s_j^{(i)}(1 - s_j^{(i)})(m_{jj}^{(i)})^T}$$

3. Az S és az M mátrix s_j illetve m_{jk} elemeit kiszámítjuk a stayerek számára vonatkozó fenti várható érték alapján:

$$s_j^{(i+1)} = \frac{n_{s_j}^{(i)}}{n_j^0}$$

$$m_{jk}^{(i+1)} = \frac{n_{jk} - \chi_{\{j=k\}} T n_{s_j}^{(i)}}{n_j^* - T n_{s_j}^{(i)}}$$

4. Az előző két lépést alkalmazzuk, amíg egy adott küszöbszint alá nem csökken a paraméter értékeinek változása, ami biztosan be fog következni, hiszen az eljárás konvergál (ld: *Fuchs-Greenhouse [1988]*).

3.3.3. Illesztés a jövedelmekre

A fenti algoritmust *Matlab*-ba beprogramozva (ismét deciliseket, illetve kvin-tiliseket tekintve) megkapjuk a szükséges átmenet-mátrixokat. Első ránézésre úgy tűnik, hogy a sima Markov-modellhez képest lényeges eltérés nincsen: az egy lépéses átmenet-mátrix elemei alig különböznek a Markov-lánc esetében becsült elemektől (3.7 ábra).

	1	2	3	4	5
1	56,49%	24,08%	12,59%	4,44%	2,39%
2	17,22%	58,77%	16,92%	4,54%	2,56%
3	5,81%	15,65%	59,15%	15,85%	3,53%
4	1,44%	2,63%	15,13%	68,16%	12,64%
5	1,08%	1,79%	2,97%	13,02%	81,14%

	1	2	3	4	5
1	56,27%	24,21%	12,66%	4,47%	2,40%
2	17,20%	58,82%	16,90%	4,53%	2,55%
3	5,80%	15,62%	59,22%	15,83%	3,53%
4	1,43%	2,62%	15,11%	68,20%	12,63%
5	1,09%	1,81%	3,00%	13,15%	80,95%

3.7. ábra. A Markov-lánc, illetve a mover-stayer modell által becsült átmenet-mátrix

Az igazi eltérést akkor láthatjuk, ha a 10 lépéses átmenetet számoljuk ki, amely esetén a különbség már jelentőssé válik: a Markov-moddellel becsült mátrixhoz képest a főátló elemei (néhol akár jelentősen is) nagyobbak lesznek. Kiszámolva a 10 lépéses átmenet-mátrixok μ mobilitását a Markov-lánc esetén 98,4% adódik. Ugyanez a mover-stayer modell esetén 90,0%. Összehasonlítva a fenti eredményeket a tapasztalt mobilitással, amelynek értéke 82,0%, azt tapasztaljuk, hogy a Markov-modell felül becsülését teljesen megszüntetni ugyan nem tudta a mover-stayer modell sem, de mindenképpen jelentősen mérsékelte a mobilitás valóságához képest túlzott mértékét (3.8 ábra).

	1	2	3	4	5
1	13,1%	20,2%	23,9%	25,1%	17,7%
2	13,0%	20,1%	23,9%	25,2%	17,8%
3	12,4%	19,2%	23,6%	26,0%	18,9%
4	11,3%	17,8%	23,0%	27,2%	20,8%
5	10,6%	16,8%	22,4%	27,9%	22,3%

	1	2	3	4	5
1	16,5%	19,4%	23,0%	24,1%	17,0%
2	12,5%	23,2%	23,0%	24,2%	17,1%
3	11,9%	18,5%	26,3%	25,0%	18,2%
4	10,7%	17,0%	21,9%	30,6%	19,8%
5	7,7%	12,3%	16,3%	20,3%	43,3%

3.8. ábra. A Markov-lánc, illetve a mover-stayer modell által becsült 10 lépéses átmenet-mátrix

Annak eldöntéséhez, hogy a mover-stayer modell a Markov-lánc modelljénél történő jobb illeszkedése vajon tényleg szignifikáns-e, a *likelihood-hányados tesztet* alkalmazzuk.

Ez abban az esetben hasznos, ha két olyan modellt vetünk össze, ahol az egyik modell a másik speciális esete. Ezalatt azt értjük, hogy az általános modell paraméterei megválaszthatóak úgy, hogy a speciális modellt kapjuk, ugyanazzal a struktúrával. Az általános modell tehát ugyanazt a struktúrát méri, de szofisztikáltabb módon, így az adatokra mindig jobban illeszkedik, ezért a log-likelihood értéke is nagyobb lesz.

A likelihood-hányados teszt azt vizsgálja, hogy a likelihood értékében bekövetkező növekedés számottevő-e, így a tesztstatisztika a következő:

$$LR = 2 \ln \frac{L_G}{L_S}$$

Itt L_G az általános, L_S a speciális modellt jelöli. Esetünkben a stayerek S mátrixát nullmátrixnak választva, a sima Markov-modellt kapjuk vissza, tehát valóban alkalmazhatjuk a fenti tesztet.

Felírjuk mindkét modell log-likelihood értékét. n -el jelölve az összes egyén, n_j^* -al pedig a j . állapotba lépések számát:

$$\begin{aligned} \ln L_{Markov-lánc} &= \sum_{j=1}^N n_j^0 \ln\left(\frac{n_j^0}{n}\right) + \sum_{(j;k)} n_{jk} \ln\left(\frac{n_{jk}}{n_j^*}\right) \\ \ln L_{Mover-Stayer} &= \sum_{j=1}^N \left[n_j^0 \ln\left(\frac{n_j^0}{n}\right) + (n_j^0 - n_j) \ln\left(\frac{n_j^0 - n_j}{n_j^0}\right) + \right. \\ &\quad \left. + (n_{jj} - Nn_j) \ln(\hat{m}_{jj}) + \sum_{j \neq k} n_{jk} \ln(\hat{m}_{jk}) \right] \end{aligned}$$

A Markov-lánc esetén a fenti érték decilisek esetén -1071832 , kvintilisek esetén pedig -704093 , míg a mover-stayer modell esetén ugyanezen esetekben -929633 , illetve -601574 . Ennek megfelelően az LR tesztstatisztika értéke 284398 , illetve 205039 .

LR aszimptotikusan χ^2 eloszlást követ, ahol a szabadságfok az általános modell speciális modellen felüli paramétereinek száma. A sima Markov-modell $(N-1)N$, a mover-stayer modell pedig N^2 paramétert becsül, tehát a szabadságfok (N) éppen az állapotok száma. A χ_N^2 eloszlás 99%-os szignifikanciaszinten számított kritikus értéke $N=5$ esetén $15,1$; $N=10$ esetén pedig $23,2$. Esetünkben ezek az értékek lényegesen magasabbak, így jól látható tehát, hogy a mover-stayer modell valóban lényegesen jobb illeszkedést mutat.

Összefoglalva láthatjuk tehát, hogy a Markov-láncok alkalmazása egyszerű, azonban sok esetben, például a jövedelmek dinamikáját tekintve nem teljesülnek a feltevései, amelyet a mover-stayer modellel némiképpen kiküszöböltünk. Természetesen tisztában vagyunk vele, hogy még a mover-stayer modell sem tökéletes megoldás, hiszen a movereket még tovább lehetne bontani aszerint, hogy mennyire mobilisek. A mover-stayer modellnek a további általánosításai (ld. például: *Spilerman S. [1978]*) azonban jóval bonyolultabbak (pl: kevert Markov-modellek), és egyáltalán nem biztos, hogy alkalmazásuk olyan mértékben hozzájárulna becsléseink javításához, amely miatt még szükséges lenne további általánosításokat tennünk. Ehelyett az alábbiakban megnézzük néhány egyszerű alkalmazását az eddigi eredményeinknek.

4. fejezet

Alkalmazások

4.1. Kockázati prémium a diákhitelezésben

A bevezetőben már említettük, hogy a magyar diákhitel törlesztése jövedelemarányos. A rendszer modelljéről részletes leírást találhatunk *Berlinger [2002]* cikkében. *Berlinger-Gerencsér-Mátyás-Rásonyi [2005]* tanulmánya szerint a hitelező szemszögéből rendszer hosszútávon fenntartható, zéró-profiton történő működtetését a modellben két kontroll-paraméter változtatásával érhetjük el. A két paraméter: a visszafizetési ráta és a kockázati prémium.

A visszafizetési rátát adottnak feltételezve vizsgálhatjuk külön a kockázati prémiumot. *Berlinger-Makara [2005]* cikkükben megmutatták, hogy a kockázati prémium hogyan függ a jövedelmek dinamikájától. Ehhez valós adatsorok csak korlátozott minőségben álltak rendelkezésre, ezért mikroszimulációs technikát alkalmazva építették fel a modellt.

A szimulációkat 4 különböző jövedelem-dinamika feltételezése mellett végezték el. Az első esetben csak a minimálbér, és az átlagos jövedelem alakulása szerepelt, változatlan jövedelem-dinamika mellett. A többi esetben jövedelem pályákat szimuláltak, és ezek alapján csoportokba sorolták a megfigyeléseket, mindezt három különböző scenárió mellett: változatlan, alacsony és magas jövedelem dinamika. A diákhitel kamatlábát zéró-profit feltételezése

mellett választották meg, továbbá feltételezték, hogy a kamatláb két tényező eredőjeként alakul ki: a hitel finanszírozási költségének (amely konstansnak tekinthető) és a kockázati prémiumnak az összege. A legfontosabb eredményük, hogy a jövedelem-pályák szerkezete jelentősen befolyásolja a kockázati prémiumot, ráadásul a korreláció negatív: determinisztikus jövedelemalakulás esetén a prémiumra 3,44%, alacsony mobilitást feltételezve 2,07%, magas mobilitás esetén pedig 1,51% adódott.

A fentiek fontos következménye, hogy a társadalmi mobilitás számszerűsítése fontos eszköze lehet például a kockázati prémium, és ezáltal a diákhitel kamatlábjának megválasztásához. Az átmenet-mátrixaink által képviselt mobilitást ezért a $\mu(A)$ illetve a $\nu(A)$ mobilitási mutatókkal is kifejezzük. A két mutató természetesen nem független egymástól, azonban a mobilitás pontosabb előrejelzéséhez érdemes mindkettőt figyelembe venni. (Esetleg tekinthetnénk alkalmasan megválasztott lineáris kombinációjukat.) Átmenet-mátrixunknak a mover-stayer modell által megbecsült mátrixot tekintjük, amely hosszútávon pontosabb képet ad a mobilitásról. A mobilitási értékeket a 4.1 ábra foglalja össze.

Mobilitás

$\mu(A)$	Teljes minta	Nők	Férfiak
Kvintilisek	44,13%	47,60%	41,33%
Decilisek	57,86%	63,54%	54,38%

$\nu(A)$	Teljes minta	Nők	Férfiak
Kvintilisek	29,96%	32,28%	28,00%
Decilisek	30,33%	35,12%	26,85%

4.1. ábra. Mobilitási mutatók összefoglalva

4.2. Foglalkoztatás

Napjaink Magyarországot tekintve ez lehet az egyik legfontosabb kérdés, amivel foglalkozni kell(ene). A hagyományos mutatókhoz képest *Augusztinovics Mária [2005]* művében részletesen foglalkozik a kérdéssel, nagy hangsúlyt fektetve a munkaerőpiac szegmentáltságára. Rámutat arra, hogy a társadalom előregedéséhez képest legalább akkora figyelmet kellene fordítani arra, hogy a keresőkorú népességen belül mekkora arányban vannak, akik valóban keresnek.

Egy egyszerű modell segítségével a keresőkorúakat két csoportja osztja: *alfákra* és *bétákra*. A felosztás alapja a vizsgált időszak alatt a munkában töltött idő aránya, ezt *foglalkoztatási sűrűségnek* definiálja: alfák azok, akik a vizsgált időtartam során végig dolgoznak, tehát esetükben ez az érték 1, a béták esetén értéke 0 és 1 közötti. Ezáltal foglalkoztatottnak tekinthetők az alfák, és azok a béták, akik az adott időszakban éppen dolgoznak. Fontos kérdés, főként a teljes keresőkorú népességre az alfák és a béták számának, illetve a béták foglalkoztatási sűrűségének a meghatározása. Egységnyinek feltételezve a keresőkorúak számát, az alábbi összefüggéseket írhatjuk fel:

$$m_\alpha + m_\beta = 1$$

$$m_\alpha + d_\beta m_\beta = E$$

Így ebből következően:

$$m_\alpha = \frac{E - d_\beta}{1 - d_\beta}$$

Itt m_α -val az alfák, m_β -val a béták arányát, d_β -val a béták foglalkoztatási sűrűségét, E -vel a foglalkoztatási hányadot jelöltük (ld: *Augusztinovics [2005]*).

A fenti vizsgálatot azonban, bár igen kezdetleges módon, megpróbálhatjuk az általunk vizsgált időszakban az adott generációra vonatkozóan elvé-

gezni. *APEH*-adatokról lévén szó, feltételezhetjük, hogy a teljes adatstruktúráinkban a hiányzó adatok azt jelölik, hogy az illetőnek az adott évben nem volt adóköteles jövedelme. Eltekintünk tehát az adatrögzítés hibáitól és hiányosságaitól, a szándékosan be nem vallóktól (vagyis a csalóktól), továbbá a halálesettől is (ami ezen korosztály esetén nem mérvadó) amellyel tehát feltételeztük a korfa adott ágának változatlanóságát. Arra vagyunk kíváncsiak tehát, hogy az eddig kevésbé vizsgált „hiányos adatokból” milyen következtetést vonhatunk le.

A fenti gondolat analógiájára alfának tekintjük azon egyéneket, akiknek az *APEH* felé bevallott éves bruttó jövedelme az 1995 – 2005-ös időszak során minden évben meghaladta az aktuális minimálbér értékét. Természetesen ez igen nagyvonalú gondolat, azonban a valóságtól feltehetőleg nem áll nagyon messze. Bizonyos értelemben alulbecsülünk, hiszen nem számoltunk az olyan jövedelmekkel, amelyekről nem kötelező bevallást leadni, de mégis a társadalmi szerepvállalás (pl. nyugdíj) során fontosak, ilyen például a *GYES*. Ugyanakkor a kapott értékeink más szemszögből tekintve felülbecsülik a valóságot, hiszen az éves bruttó jövedelem úgy is meghaladhatja az egy évre eső minimálbér összegét, hogy valójában mondjuk egész évben nem volt bevételünk, ám abban a 4 hónapban keresetünk igen jelentős volt.

Eredményeink az alábbiak lettek: az alfák becsült aránya 19%, a bétáké ennek megfelelően 81%, a béták átlagos foglalkoztatási sűrűsége pedig körülbelül 40%, mindez nagyjából 52%-os foglalkoztatási hányad mellett. Érdemes lehet mindezt összehasonlítani a teljes társadalomra Augusztinovics által becsült értékkel: felhasználva a foglalkoztatási hányad 60%-ra valószínűsített értékét, az alfák arányát kicsivel 40% fölé, a béták foglalkoztatási sűrűségét pedig 20 – 30%-ra becsülte. Jól láthatóan az alfák aránya jelentősen kisebb, ami nem meglepő egy olyan korszakban, amikor a karrier még kibontakozóban van.

Összefoglalás

Szakedolgozatom során igyekeztem a jövedelmek vizsgálatát a lehető legszélesebb módon bemutatni, és a gyakorlatban is hasznosítható módon, főként számszerű mutatókkal a jövedelmek eloszlását és belső struktúráját is jellemezni. Természetesen egy adott generáció vizsgálatát kiterjeszthetnénk szélesebb spektrumra is, illetve 1995 – 2005 adatait még tovább bővíthetnénk. Ahhoz azonban, hogy alapvető következtetéseket levonhassunk, véleményem szerint elegendőnek bizonyult az adatok vizsgálata.

Zárszóul elmondhatjuk, hogy a jövedelmek vizsgálata alapvető fontossággal bír, ugyanakkor az adatok és a módszerek különféle szemszögből történő elemzése szinte kimeríthetetlen, így a válság utáni (remélhetőleg elkövetkező) boldog békeidőben is bőven akadnak további lehetőségeink a jövedelmek szerkezetének feltárására.

Köszönetnyilvánítás

Ezúton szeretném megköszönni témavezetőmnek, *Berlinger Edinának* hasznos javaslatait, támogatását, amelyek nagyban hozzájárultak szakdolgozatom sikeres elkészítéséhez. Külön köszönetem fejezem ki a szakdolgozatom megírásához szükséges *APEH*-adatok rendelkezésemre bocsájtásáért.

Nagy köszönettel tartozom *Major Klárának*, aki további igen értékes tanácsaival látott el, és nagyfokú segítséget nyújtott a Markov-modellek Matlabba történő átültetéséhez.

Köszönöm továbbá *Makara Tamásnak* a kiváló ötleteit, amelyek motivációt nyújtottak szakdolgozatom elkészítéséhez.

Tartalomjegyzék

1. Bevezetés	1
2. A jövedelmek eloszlásvizsgálata	4
2.1. Bevezetés	4
2.2. A lognormalitás elvetése	5
2.3. További empirikus tények	14
2.4. A jövedelmek egyenlőtlensége	18
3. A jövedelmek belső struktúrája	26
3.1. Bevezetés	26
3.2. Markov-lánc modell	27
3.3. Mover-stayer modell	42
4. Alkalmazások	54
4.1. Kockázati prémium a diákhitelezésben	54
4.2. Foglalkoztatás	56

Irodalomjegyzék

- [1] Augusztinovics M. [2005]: *Népesség, foglalkoztatottság, nyugdíj*. Közgazdasági Szemle, LII. évfolyam, 2005. május, 429-447. oldal
- [2] Berlinger E. [2002]: *A jövedelemarányos törlesztésű diákhitel egyszerű modellje*. Közgazdasági Szemle, XLIX. évfolyam, 2002. december, 1042–1062. oldal
- [3] Berlinger E. - Gerencsér L.- Mátyás Z. - Rásonyi M. [2005]: *Optimal control of an income-contingent student loan system*. 21st European Conference on Modelling and Simulation
- [4] Berlinger E. - Makara T. [2005]: *Calculation of the risk premium in a self-sustaining, income-contingent student loan system*. European Simulation Symposium
- [5] Embrechts P. – Klüppelberg C. – Mikosch T. [1997]: *Modelling extremal events for insurance and finance*. Springer
- [6] Éltető Ö. – Vita L. [1982]: *Jövedelem-eloszlások közelítése és prognosztizálása*. Szigma 37. évfolyam, 1–2. szám, 15–39. oldal
- [7] Frydman H. [1984]: *Maximum likelihood estimation in the mover-stayer model*. Journal of the American Statistical Association, 79. évfolyam, 632-638. oldal
- [8] Fuchs C. – Greenhouse J. B. [1988]: *The EM algorithm for maximum likelihood estimation in the mover-stayer model*. Biometrics, 44. évfolyam, 605-613. oldal

- [9] Kovács I. [2010]: *A jövedelemeloszlás és jövedelem-egyenlőtlenség a személyi jövedelemadó-bevallási adatok tükrében*. MTA tanulmány
- [10] Major Klára (Szerk.) [2008]: *Markov-modellek*. BCE Makroökonómia Tanszék - ELTE Regionális Tudományi Tanszék
- [11] Marossy Z. [2010]: *A spot villamosenergia-árak elemzése statisztikai és ökonofizikai eszközökkel*. PhD értekezés
- [12] McNeil A. J. – Frey R. – Embrechts P. [2005]: *Quantitative risk management - Concepts techniques and tools*. Princeton University Press. Princeton series in finance.
- [13] Satya P. [2009]: *A measure of income mobility with an empirical application*. University of Western Sydney
- [14] Spilerman S. [1978]: *Extensions of the mover-stayer model*. American Journal of Sociology, 78. évfolyam, 559-626. oldal
- [15] Stokey N. L. – Lucas R. E. [1989]: *Recursive methods in economic dynamics*. Harvard University Press
- [16] Tóth I. GY. [2003] - *Jövedelemegyenlőtlenségek*. Közgazdasági Szemle, L. évfolyam, 2003. március, (209–234. oldal)
- [17] Weber A. [2002]: *State dependence and wage dynamics: a heterogeneous Markov chain model for wage mobility in Austria*. Working Paper IHS Economic Series 114
- [18] <http://de.wikipedia.org/wiki/Robin-Hood-Index>
- [19] http://geogr.elte.hu/REF/REF_Kiadvanyok/REF_RTT_11/RTT-11-03-teregyenlotlenseg.pdf
- [20] http://www.cs.elte.hu/~zempleni/gini_coef.pdf
- [21] http://www.ksh.hu/docs/hun/xstadat/xstadat_eves/i_qli041.html
- [22] http://www.ksh.hu/interaktiv_korfa