Tér és Társadalom 21. évf. 2007/1. 53-67. p.

MARKOV LÁNCOK HASZNÁLATA A REGIONÁLIS JÖVEDELEMEGYENLŐTLENSÉGEK ELŐREJELZÉSÉBEN[1]

(Forecasting Regional Income Inequalities Based on Markov Models)

MAJOR KLÁRA

Kulcsszavak:

területi jövedelemegyenlőtlenségek Markov láncok Mover–Stayer modell

A jövedelemegyenlőtlenségek változásának vizsgálatára használt módszertanok egyike a Markov láncok modelljének illesztése. A Markov modell azonban túlbecsüli a hosszú távú mobilitást, ezért hosszabb távú felzárkózási folyamatok előrejelzésére nem alkalmas. A Mover–Stayer modell, mint a Markov modell általánosítása alkalmas arra, hogy jelentősen javítson az alapmodell hosszú távú előrejelzési képességén.
A tanulmányban Frydman (1984) módszertanát alkalmazva numerikus úton kiszámítottuk mind az alap Markov modell, mind a Mover–Stayer modellek paramétereit a magyarországi kistérségek jövedelmi adatai esetében. A két modell összehasonlításával megállapíthatjuk, hogy a Mover–Stayer modellből következő hosszú távú mobilitás közel hasonló lett a megfigyelt értékhez. A Mover–Stayer modell jobb illeszkedését likelihood-arány teszt alkalmazásával vizsgáltuk.

Bevezetés

Az országok, régiók, területegységek gazdagságának, szegénységének kérdése régóta a közgazdaságtan alapkérdései közé tartozik. A relatív jövedelmi pozíciók magyarázatára, változásának előrejelzésére több különböző megközelítés, modellezési gyakorlat, tudományos irány született. A kilencvenes években kiteljesedett, ún. konvergencia vita ehhez az ághoz az empirikus módszertan hihetetlen felfutásával járult hozzá. Ez volt az az évtized, amelyben a Penn World Table adatbázisra építve a kutatók egyre intenzívebben foglalkoztak azzal a kérdéssel, hogy empirikus alapokon prognosztizálják a világméretű jövedelmi különbségek változásának tendenciáját. Ebben az igen termékeny évtizedben több régi, „elfeledett" modellcsalád is újra feléledt és számos esetben alkalmazásra került. Ebbe a sorba tehetjük a jövedelem-eloszlások változásának előrejelzésére alkalmas Markov-modellek családját is, amelyet a kérdéses kutatási iránytól függetlenül is előszeretettel alkalmaznak a szociológiai kutatásokban a társadalmi státuszban végbemenő generációs mobilitások vizsgálatára, vagy például a munkapiaci folyamatok leírása során a munkapiaci státusz változásának modellezésére.

A jövedelmi különbségek vizsgálatára történő felelevenítése elsőként talán Quah nevéhez fűzhető. 1993-as tanulmányában (*Quah* 1993) az európai régiók egy főre jutó jövedelmének eloszlásában végbement változás vizsgálatára alkalmazta a

Major Klára : Markov láncok használata a regionális jövedelemegyenlőtlenségek előrejelzésében.
Tér és Társadalom 21. évf. 2007/1. 53-67. p.

Markov láncok modelljét. Az alkalmazás során talált empirikus eredményei hasonlóak voltak a társadalomtudományok más területein is talált eredményekhez: magasfokú perzisztencia, igen alacsony jövedelmi mobilitás. Mindemellett Quah már ebben a cikkében is megemlíti, hogy a Markov láncok modellje túlzott mértékű leegyszerűsítés abban az értelemben, hogy a hosszú távú mobilitást szisztematikusan felülbecsüli, amely általánosan megfigyelhető a modell társadalomtudományi alkalmazásai során. Ezek az észrevételek azonban nem újkeletűek, a már említett szociológiai alkalmazásokban már mintegy két évtizeddel korábban is feljegyezték őket (pl. *Spilerman* 1978 vagy *Singer–Spilerman* 1976).

Ennek ellenére a Markov modell – talán egyszerűségénél, könnyen interpretálhatóságánál fogva – megőrizte népszerűségét és mind a mai napig számtalan publikációban képezi a jövedelmi dinamikai vizsgálatok alapját (ld. például *leGallo* 2001 műhelytanulmányát). *Bickenback* és *Bode* 2001-es műhelytanulmányában ezért kifejezetten azt a kérdést vizsgálja, hogy mennyire alkalmas a Markov modell ezen területi jövedelmi folyamatok leírására. Az USA államainak adatain végzett empirikus vizsgálatai során nem csak a Markov modell paramétereinek számszerűsítését végezték el, de tesztelték ezek szignifikanciáját is. Megállapításaik szerint az adatokból nyerhető becslések nem felelnek a modell feltevéseinek[2], azaz a Markov modellek illesztése téves következtetések levonásához vezethet.

Ahogyan a probléma, úgy néhány megoldási út is ismert volt. Az alap-Markov modell (hívjuk a továbbiakban így a Markov láncok modelljét) általánosításai, bővítései alkalmasak arra, hogy kezeljék ezt a problémát. Többféle úton is el lehet indulni az általánosítás felé, akár az időtől függő átmenetek bevezetésével vagy heterogén populáció feltevésével. A jelen tanulmányban az utóbbi úton kívánunk egy lehetséges alkalmazást bemutatni. A legegyszerűbb, heterogén populációra épülő Markov modell általánosítás talán az ún. Mover–Stayer modell, amelyben a populációt mindössze két alcsoportra bontjuk: a mozgókra (mover), illetve maradók (stayer). A modell így az alap Markov modell általánosításának tekinthető, hiszen a mozgók csoportjának jövedelmi dinamikáját egy hagyományos Markov lánc modell írja le, míg a maradók jövedelmi dinamikáját pedig egy igen speciális Markov lánc modell, az egész populációban megfigyelt mobilitás pedig a két részfolyamat összegeként áll elő.

Mivel a területi jövedelmi folyamatokat igen magas perzisztencia, alacsony mobilitás és rövid (10–30 év alatt nem számottevő) jövedelmi változások jellemzik, ezért feltehető, hogy a Mover–Stayer modell jobban illeszkedik a megfigyelt változásokra, és így várhatóan pontosabb becslést ad a hosszabb távú mobilitásra. Jelen tanulmányban az a célunk, hogy ezt megmutassuk Magyarország kistérségeinek jövedelmi adatain.

A tanulmány felépítse a következő. Az első fejezetben röviden összefoglaljuk a Markov és a Mover–Stayer modellek alapvető koncepcióját és becslésének kérdéseit. A második fejezetben bemutatjuk mindkét modell becsléséből kapott értékeket, kivetítjük ezeket a minta teljes hosszára (13 év). A Mover–Stayer statisztikai értelemben vett jobb illeszkedését likelihood-arány teszttel mérjük a második alfejezetben. A tanulmányt az összefoglalás fejezi be.

Major Klára : Markov láncok használata a regionális jövedelemegyenlőtlenségek előrejelzésében.
Tér és Társadalom 21. évf. 2007/1. 53-67. p.

Elméleti alapok

A tanulmány első felében összefoglaljuk az alap-Markov modell és a Mover–Stayer modell leírását és becslési technikáit. A kifejtés során elsősorban arra fogunk törekedni, hogy az egy lépéses átmenetekből hogyan kapunk több lépéses átmenetet, mivel vizsgálatunk célja a modellek hosszú távú előrejelzési képességének tesztelése lesz.

A Markov láncok alapmodellje és az átmenetmobilitás

Tegyük fel, hogy a vizsgálati egységek egyes jövedelmeit besoroltuk jövedelmi kategóriákba, azaz véges számú *állapot* valamelyikébe. Az állapotok számát már előre meghatároztuk, a továbbiakban jelöljük ezt J-vel.

A Markov lánc modellje szerint egy vizsgálati egység (itt: kistérség) jövőbeni jövedelmi pozícióját jelen pozíciója (állapota) és a változás valószínűsége határozza meg. Ez utóbbi kizárólag az állapotok függvénye. Ez azt jelenti, hogy a jövőben várható állapot nem függ egyéb tényezőktől, például attól, hogy az adott egység mióta tartózkodik a jelen állapotban, vagy milyen úton jutott el a jelen állapotba. Formálisan ezt az összefüggést az alábbiakban tudjuk felírni:

$$\pi_{t+1} = \pi_t M \qquad (1)$$

ahol π_t jelöli a t-ik időpontban az egyes elemek eloszlásvektorát, azaz elemei rendre annak valószínűségét adják meg, hogy az egyes egyedek milyen valószínűséggel találhatók az egyes állapotokban. Az M mátrix adja meg az átmenetvalószínűségek JxJ-s mátrixát.[3]

Az M mátrix elemeit az egyik állapotból a másik állapotba történő elmozdulás feltételes valószínűségeiként értelmezhetjük. A mátrix főátlója ennek megfelelően a helyben maradás, azaz a nem mozgás valószínűségét mutatja. Ez az értelmezés segít megérteni, hogy miért alkalmas az alábbi mutató az általánosan vett mobilitás mérésére (*Shorrocks* 1978):

$$mobilitás(M) = \frac{J - trace(M)}{J - 1} \qquad (2)$$

Ebben a kifejezésben *trace(M)* a mátrix nyomát, azaz főátlóiban szereplő elemeinek összegét adja meg. A fentebb definiált mobilitási mutató értéke a gyakorlati alkalmazások esetében általában 0 és 1 közé esik, értéke minél kisebb, annál kisebb a vizsgált jelenség általános mobilitása.

A Markov lánc modell empirikus becslésére maximum likelihood becslési technikát alkalmazunk, az egyes átmenetvalószínűségek becslőfüggvényét a relatív gyakoriságok számításával nyerjük (pl. *Frydman* 1984).

Amennyiben több periódus alatti mobilitást szeretnénk vizsgálni, úgy az (1) alatti képlet iterálásával kapjuk, hogy T periódus alatt az eloszlás változását leíró összefüggés az M mátrix hatványaival adható meg:

$$\pi_{t+T} = \pi_t M^T \qquad (3)$$

Major Klára : Markov láncok használata a regionális jövedelemegyenlőtlenségek előrejelzésében.
Tér és Társadalom 21. évf. 2007/1. 53-67. p.

Amiből az is látható, hogy a több periódus alatt összességében megtett mobilitás az M^T mátrix nyomának ismeretében számítható (2) képlet segítségével, formálisan:

$$mobilitás(M^T) = \frac{J - trace(M^T)}{J-1}. \qquad (4)$$

Miért lehet érdekes a több periódus alatti mobilitás, illetve annak becslése? A Markov modell becsléséhez szükséges adatok többnyire *panel* szerkezetűek, azaz több egyed több peridódusban megfigyelt állapotának táblázatszerű (J x $T+1$) elrendezése. Ez számtalan információt tartalmaz. Többek között lehetővé teszi azt, hogy egyfelől az egymást követő periódusokban megfigyelt átmenetekből közvetve következtessünk a T időszak alatti összes mobilitásra ((3)-as képlet segítségével); másfelől azt is, hogy az adatokból közvetlenül tegyük meg ezt, összehasonlítva a kiinduló időszak és a záró időszaki értékeket. A modell jóságának egyfajta kritériuma, hogy a két megközelítés mennyire vezet hasonló eredményekre. Más szavakkal a mintából nyert becslés képes-e a mintán belüli folyamatok „előrejelzésére".

A jövedelmi folyamatok vizsgálata során, hasonlóképpen a szociológiai alkalmazásokhoz, igen gyakori, hogy a fenti két megközelítés ellentmondó eredményekre vezet (pl. *Quah* 1993). Az ellentmondás oka a társadalmi-gazdasági folyamatok magas perzisztenciájában keresendő. A modellezési gyakorlatban ennek kezelésére a Markov lánc alapmodell különböző kiterjesztéseit alkalmazzák, ezek közé sorolható a Mover–Stayer modell is.

A Mover–Stayer modell és becslése

A Mover–Stayer modell a Markov lánc modell kiterjesztése heterogén populáció esetére. Tegyük fel, hogy a vizsgálati egyedek nem egyformák jövedelmi mobilitásuk szempontjából, azaz nem lehet mindegyikükre ugyanazt a Markov modellt illeszteni, nem lehet egyetlen „közös" M mátrixszal leírni mindegyikük várható jövedelmi pályáját. Az egyszerűség kedvéért tegyük fel, hogy a heterogén populáció valójában két, különböző típusú egyedből áll, egyikük mobilitását leírhatjuk egy szokásos Markov lánccal. Őket hívjuk mobiloknak (mover). A populáció többi egyede pedig feltevés szerint egyáltalán nem mobil, az ő jövedelmi pozíciójuk tehát változatlan. Őket hívjuk maradóknak (stayer). A modellezés problémája, hogy nem ismerjük az egyes egyedek típusát, nem tudjuk megmondani, hogy ki melyik kategóriába tartozik. Amit ismerünk, az a teljes populáció által megvalósított jövedemi mobilitás, formálisan

$$P_1 = S \cdot I + (I - S)M \qquad (5)$$

mátrix, ahol M továbbra is a mobilis (rész)populáció Markov mátrixa, S a maradók arányát adja meg az egyes állapotokban (J x J diagonális mátrix), I pedig az egységmátrix. Mivel nem ismerjük az egyes részpopulációk arányát, ezért (5) jobb oldalán M és S is ismeretlen, egyedül P_1-et tudjuk megfigyelni, ami a teljes populáció által produkált jövedelmi átmenetvalószínűségeket tartalmazza. Az (5) képlet felírásából láthatjuk, hogy a modellt felfoghatjuk úgy is, amelyben a két részpopuláció Markov-

Major Klára : Markov láncok használata a regionális jövedelemegyenlőtlenségek előrejelzésében.
Tér és Társadalom 21. évf. 2007/1. 53-67. p.

mátrixa eltérő, a mozgóké M, a maradóké I, a megfigyelhető átmenetmártrix (P_1) pedig ezen két átmenetmátrix súlyozott átlaga.

Több periódus alatti átmenetvalószínűségek kiszámításához az (5)-ös képlet alapján most a következőképpen gondolkodhatunk: mindkét részpopuláció átmenetvalószínűségeit saját Markov-mátrixának hatványai írják le, így a teljes populáció megfigyelhető, $T+1$ időszak alatti átmenetmátrixát a

$$P_T = S + (I - S)M^T \qquad (6)$$

kifejezéssel kapjuk[4].

A Mover–Stayer modell becsléséhez szintén a maximum-likelihood eljárást alkalmazzuk. Ennek alkalmazásakor azonban némileg nehezebb dolgunk van, mint a Markov láncok alapmodellje esetében, mert nem ismerjük, hogy az egyes egyedek melyik részpopulációhoz tartoznak, ezért nem tudjuk S és M mátrixokat közvetlenül a relatív gyakorisággal becsülni. A modell becsléséhez direkt és indirekt módszereket egyaránt kifejlesztettek, ezeket összegezzük röviden a következő alpontban.

A Mover–Stayer modell becslési módszerei

A Mover–Stayer modell log-likelihood-függvénye az alábbi (ld. pl. *Frydman* 1984):

$$\log L = \sum_{j=1}^{J} n_j(0)\log\eta_j + \sum_{j=1}^{J}\log L_j \qquad (7)$$

ahol $\{\eta_j\}_{j=1,\dots,J}$ a kezdeti időpontbeli eloszlás, $n_j(0)$ a kezdeti időpontban a j állapotban lévő egyedek száma és

$$\log L_j = n_j \log\left(s_j + (1-s_j)m_{jj}^T\right) + \left(n_j(0) - n_j\right)\log(1-s_j) + \left(n_{jj} - Tn_j\right)\log m_{jj} + \sum_{k\neq j} n_{jk}\log m_{jk}$$

ahol s_j és m^T_{jk} az S és M^T mátrixok megfelelő elemei, n_j azon egyek száma, amelyek mindvégig a j állapotban vannak, n_{jk} a $j \to k$ átmenetek száma.

A modell változóinak a likelihood függvény alapján történő közvetlen meghatározása (*Frydman* 1984)

A loglikelihood függvénynek a modellváltozók (s_j, m_{jk}) szerinti differenciálásával megkapjuk az elsőrendű feltételeket. Az ismeretlenek kifejezésével és a maradék egyenletekbe történő helyettesítésével végül egyismeretlenes egyenlethez jutunk, amelyből numerikus módszerekkel m_{jj} értéke meghatározható (külön, minden j-re). A numerikus módszerek alkalmazása elkerülhetetlen: az m_{jj} értékét meghatározó egyenlet $T+1$-ed fokú polinom, amelyről megmutatható, hogy pontosan egy gyöke esik 0 és 1 közé. Frydman útmutatásait követve beprogramoztuk a megoldóalgoritmust Matlab programmal és meghatároztuk a Mover–Stayer modell változóinak értékét.

A modell változóinak EM algoritmussal történő meghatározása (*Fuchs–Greenhouse* 1988)

A teljesség kedvéért megemlítjük, hogy Fuchs és Greenhouse szerzőpáros egy közelítő algoritmust dolgozott ki a modell változóinak meghatározására. Az általuk kifejlesztett, ún. EM algoritmus alapgondolata szerint a becslési nehézséget az adja, hogy az adathalmaz *hiányos*: nem tartalmazza azt az információt, hogy az egyes egyedek melyik típusba tartoznak. Bontsuk tehát két lépésre a becslési eljárást, az ún. E-lépésben becsüljük meg a hiányzó adatokat, majd az M-lépésben e becslésre támaszkodva, azaz a plusz információk birtokában könnyűszerrel becsülhetőek a modell paraméterei. Az M-lépést követően azonban újra kell számítani az E-lépést, hogy konzisztensek-e a kapott paraméterértékek a hiányzó adatokkal. Az algoritmust addig ismételjük, amíg konvergál. Az EM algoritmus előnye a Frydman féle direkt technikához képest, hogy az egyes lépésekben megoldandó egyenletek mind lineárisak, így könnyebb programozni, és gyorsabb is lesz a kód. Az EM algoritmus további előnye, hogy alkalmas a továbbfejlesztésre, ahogyan ezt a következő módszertani publikációból láthatjuk.

Az EM algoritmus továbbfejlesztése kevert Markov-modellekre (*Frydman* 2005)

Kevert Markov lánc-modelleknek (Mixed Markov chains) nevezzük azokat a modelleket, ahol a részpopulációk száma nem feltétlen 2, ettől különböző egész szám is lehet. Az egyes részpopulációk különböznek mobilitási sebességükben, ennek megfelelően a mozgást leíró Markov mátrixukban. A modell becslése a Fuchs–Greenhouse által kifejlesztett EM algoritmus továbbfejlesztésén alapszik. Ennek részletes tárgyalásától most eltekintünk.

Emprikus eredmények

Az alábbiakban mind az alap-Markov modellre, mind a Mover–Stayer modellre közzétesszük a számítási eredményeket. Megmutatjuk, hogy az általunk végzett vizsgálat során a hosszabb távú mobilitás előrejelzésében a Mover–Stayer modell szignifikánsan jobban illeszkedett az adatokra, mint az alap-Markov modell.

Az adatbázis

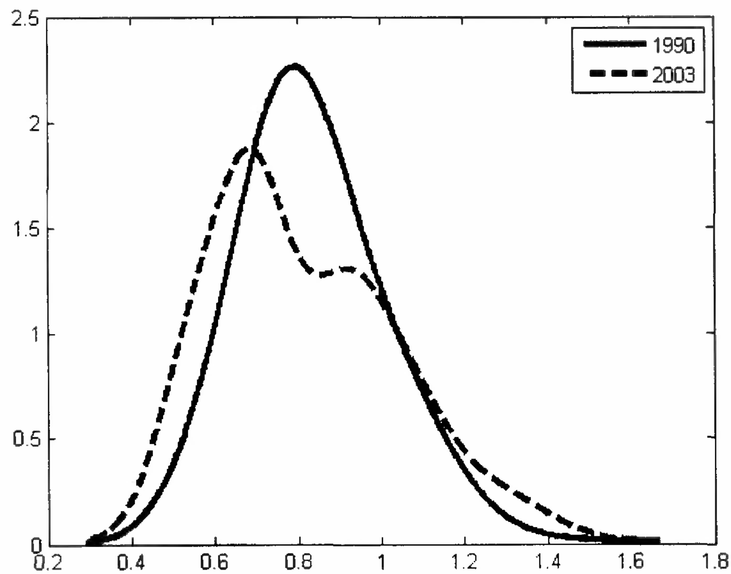
A vizsgálat során használt adatok Magyarország kistérségeinek (az új felosztás szerinti, 168 kistérségre vonatkozó) egy főre jutó személyi jövedelemadó alapját képező adózás előtti jövedelme képezte az 1990–2003 közötti időszakban. A rendelkezésre álló 14x168-as méretű adattáblába rendezett adatok nem csak az átmenetek megfigyelését, de az egyes kistérségek jövedelmi pozíciójának nyomonkövetését is lehetővé teszik. Az egy főre jutó jövedelmeket az országos (súlyozott) átlag százalékában fejeztük ki, ezzel az adatok nagyságrendileg a (0,5; 1,6) intervallumba kerültek.

Major Klára : Markov láncok használata a regionális jövedelemegyenlőtlenségek előrejelzésében.
Tér és Társadalom 21. évf. 2007/1. 53-67. p.

A jövedelemeloszlás vizsgálatához kernel becslés módszerével megbecsültük az egy főre jutó relatív jövedelmek (folytonos) eloszlásfüggvényét *(1. ábra)*. A jövedelmek eloszlása a jelen alkalmazásban is közel log-normális alakú, bár lokális tulajdonságaiban jelentős változás következett be a vizsgálati periódusban (kétmóduszúság). Ennek vizsgálata nem képezi jelen tanulmány tárgyát, ugyanakkor az egyértelműen elmondható, hogy a vizsgált periódusban polarizáció volt megfigyelhető: csökkent az átlagoshoz hasonló, „közepes" jövedelmű kistérségek száma, és növekedett az átlaghoz képest magas vagy alacsony jövedelmű kistérségek gyakorisága. Mindezen fontos információk mellett a sűrűségfüggvény nem mond semmit az egyes kistérségek felzárkózási esélyeiről, valamint a mobilitásról, ezért lépünk tovább Markov modell alkalmazása felé.

1. ÁBRA

Az egy főre jutó relatív jövedelmek kernel becslése
A sávszélességi paraméter plug-in eljárással becsülve, értéke
1990-re: 0,739; 2003-ra: 0,0746.
(Kernel Estimation of the Income Distribution)



Forrás: Saját számítások.

A kistérségeket ezt követően kellett jövedelmi kategóriákba sorolni, azaz az egyes relatív jövedelmi pozíciókat állapotoknak megfeleltetni. Ehhez a leíró statisztikákat hívtuk segítségül *(1. táblázat)*.

1. TÁBLÁZAT
Az egy főre jutó relatív jövedelmek leíró statisztikái Magyarország kistérségei esetében
(Microregional Income Distribution in Hungary – Basic Statistics)

	1990	*2003*	*1990–2003*
Minimum	0,50	0,43	0,39
1. kvartilis	0,71	0,65	0,66
2. kvartilis	0,82	0,77	0,78
3. kvartilis	0,94	0,97	0,96
Maximum	1,46	1,51	1,56
Variancia	0,03	0,05	0,04

Forrás: Saját szerkesztés.

Az egyes években megfigyelt relatív jövedelmi pozíciók eloszlása nagyon hasonló volt, terjedelmük, szórásuk, különböző percentiliseik igen stabilnak mutatkoztak. Ezt mutatja szúrópróbaszerűen az első és utolsó évre az *1. táblázat.* (Természetesen a minimum és maximum értékek ingadozhatnak.) Ezen jelentős stabilitás miatt és a szakirodalomban bevett módszernek megfelelően a jövedelmi kategóriák képzése során a kvartilisekből indultunk ki, ezzel 4 jövedelmi kategóriát hoztunk létre. Annak érdekében, hogy egyik év se kapjon kitüntett szerepet a teljes minta (1990–2003) alapján megállapított kvartilis-értékek képezték az osztályközöket *(1. táblázat 4. oszlop),* amelynek segítségével az egyes kistérségek jövedelmi állapotának megállapítása történt.

A jövedelmi kategóriákba osztás révén jellemezni tudjuk a relatív jövedelmek területi megoszlását grafikusan is *(2. ábra).* Az országos térkép kiszínezéséhez most ritkán alkalmazott, némileg talán elsőre bonyolultnak tűnő beosztást választottunk. Azon kistérségeket, amelyek a vizsgált 14 éves időszak alatt végig azonos kategóriában voltak tömör színnel, míg a pozíciót váltókat pöttyös háttérrel szineztük ki. Eközben igyekeztünk az árnyalásnak is szerepet adni: minél sötétebb egy kistérség színe, annál nagyobb jövedelmi kategóriát jelöl. A *2. ábrán* látható az eredmény.

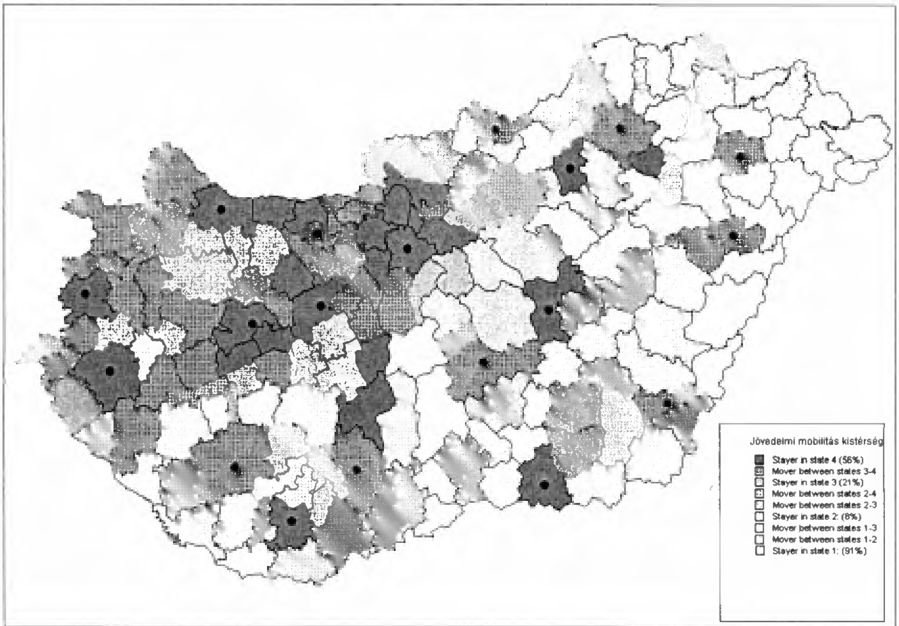
A térképből a magyarországi jövedelmek jól ismert térszerkezete tűnik elénk, a „globálisan”, azaz országos méretekben megfigyelhető nyugat–kelet lejtő. A nyugati kistérségek általában sötétebb színűek, a keleti országszélen pedig nem csak hogy nagyon világos, de tartósan ott ragadó kistérségeket találunk.

Másrészt szembeötlő a helybenmaradó kistérségek magas száma. Valójában ez az a tulajdonsága az adathalmaznak, amely már a kutatás elején is sugallja a Mover–Stayer modell alkalmazásának az igényét: számtalan olyan kistérség van, amelyek egyáltalán nem mutattak fel semmilyen mobilitást. Az alap-Markov modell esetében ez csak egyféleképpen magyarázható: a helyben maradás valószínűségével, amely értékeket a Markov mátrix főátlói tartalmaznak. 14 év, azaz 13 átmenet alatt a helyben maradás valószínűsége a főátlóban szereplő érték 13-ik hatványa, amely még 95%-os egylépéses helybenmaradás esetén is alig több mint 50%. Az adatok-

Major Klára : Markov láncok használata a regionális jövedelemegyenlőtlenségek előrejelzésében.
Tér és Társadalom 21. évf. 2007/1. 53-67. p.

ból ennél azonban sokkal magasabb 13 átmenet alatti helybenmaradási arány látható. Ezért a Markov és a Mover–Stayer modellek illesztésével és az illeszkedés jóságának megvizsgálásával megbizonyosodhatunk sejtésünk helyességéről.

2. ÁBRA
Az egyes kistérségek jövedelmi állapotai 1990–2003 között
(Changes in the Relative Income Positions of the Hungarian Microregions (1990–2003, Two-dimensional Categorisation)



Jelmagyarázat: Tömör színűek a maradók, satírozott hátterűek mozgók. A sötétebb szín magasabb jövedelmi kategóriát jelöl.

Forrás: Saját számítások.

Globális egyenlőtlenségek változása

Az országos átlagjövedelem százalékában kifejezett egy főre jutó jövedelmek alapján képzett jövedelmi kategóriák vizsgálatával képet nyerhetünk a globális egyenlőtlenségek változásáról. Itt a globális jelzőt országos viszonylatban kell értelmezni: amikor az egyes kistérségek alacsony vagy magas jövedelmét említjük ezt nem abszolút értékben, hanem az országos átlag százalékában kell érteni.

Egy lépéses átmenetek

Az alap Markov modell számításához mindössze a megfigyelt 13x168=2184 átmenetből kellett relatív gyakoriságot számítanunk. Ezeket az eredményeket tartalmazza a 2. *táblázat.*

2. TÁBLÁZAT
Egy lépéses átmenetmátrix, mobilitási mutató = 15,51%
(One-step Mobility Matrix, Mobility Index: 15,51%)

	1	2	3	4
1	0,92	0,08	0,00	0,00
2	0,12	0,83	0,06	0,00
3	0,00	0,08	0,85	0,07
4	0,00	0,00	0,06	0,94

Forrás: Saját számítás.

A Markov modell becslésével nyert egy lépéses átmenetvalószínűségek igen hasonlatosak a szakirodalomban található, más területegységekre, időszakokra, állapottérfelosztásra készült átmenetmátrixokhoz. A jövedelmi folyamatokra egyaránt jellemző, hogy a magas és alacsony jövedelmi kategóriákban a nem-mozgás valószínűsége magasabb, mint a közepes jövedelmi kategóriákban. Ezt a megfigyelést a korábban a térkép kapcsán említett polarizációs jelenséggel lehet összefüggésbe hozni: a középső jövedelmi kategóriákban sokkal magasabb a megfigyelt jövedelmi mobilitás, mint a szélsőséges állapotokban. Ennek lehet eredménye a közepes jövedelmi kategória „szűkülése", amely a folytonos sűrűségfüggvénybecslésből annyira jól kivehető.

Frydman (1984) módszertanát alkalmazva kiszámítottuk a Mover–Stayer modell ismeretlen együtthatóit, az S és M mátrixokat egyaránt. A becsült értékeket mutatja a *3. táblázat.* A fentebb elmondottakon túlmenően még azt is megállapíthatjuk, hogy a szélsőséges jövedelmi kategóriákban kiemelkedően magas a maradók (stayerek) aránya, az 1-es kategóriában egyenesen 92%. Az adattáblában 24 olyan kistérséget találtunk, amelyek mind a 13 év alatt végig az 1-es kategóriában tartózkodtak. Ezen kistérségek 92%-át, azaz kb. 22-t lehet a maradók kategóriájába sorolni! Ezen kistérségek esetében nem pusztán „véletlen" de strukturális tényezőkkel kell magyaráznunk a mobilitás elmaradását – még egy ilyen nagyon egyszerű megközelítésben is, mint a Mover–Stayer modell, ahol a strukturális tényezők explicit nem jelennek meg.

A Mover–Stayer modell paramétereinek számításával (az alap-Markov modellhez képest) alternatív módon is kiszámítottuk az átmenet mátrixát és mobilitási mutató értékét. Láthatjuk, hogy egyikben sem hozott látványos változást a bonyolult módszertan alkalmazása, ami egyáltalán nem meglepő. Az egy lépéses átmenetek tanulmányozására nincsen szükség ilyen kifinomult módszertan alkalmazására, a két modell közötti különbség elsősorban akkor látszik, amikor a hosszabb távú, jelen esetben mondjuk 13 éves mobilitás előrejelzésére kívánjuk felhasználni őket.

Major Klára : Markov láncok használata a regionális jövedelemegyenlőtlenségek előrejelzésében.
Tér és Társadalom 21. évf. 2007/1. 53-67. p.

3. TÁBLÁZAT

Mover–Stayer modell becslésének eredménye a magyarországi kistérségek relatív egy főre jutó jövedelmi pozíciói alapján, 1990–2003.

(Mover–Stayer Mobility Estimation of the Microregional Incomes in Hungary, 1990–2003)

Álla-pot	*Maradók aránya (S)*	*Mozgók Markov mátrixa (M)*				*Megfigyelhető átmenet-mátrix (P_1)*			
1	0,92	0,83	0,17	0	0	0,99	0,01	0,00	0
2	0,08	0,13	0,81	0,06	0	0,12	0,82	0,06	0
3	0,21	0	0,11	0,79	0,10	0	0,09	0,83	0,08
4	0,57	0	0	0,13	0,87	0	0	0,06	0,94

Jelmagyarázat: Az utolsó négy oszlopban szerepelnek a megfigyelhető, az alap-Markov modellel közvetlenül összevethető átmenetvalószínűségek. Mobilitási mutató P_1 mátrixra = 13,84%.

Forrás: Saját szerkesztés.

13 éves átmenetvalószínűségek

A 13 éves átmenetvalószínűségek számításával most a modellek hosszabb távú előrejelző képességét fogjuk „tesztelni". Természetesen nem szükséges mindenképpen 13 éves horizontot választani, ez azonban a jelen esetben igen praktikus: az adataink is pontosan ilyen hosszúságú intervallumra állnak rendelkezésre, így mérni tudjuk a modellek előrejelző képességét, ha összehasonlítjuk az általuk adott 13-éves mobilitásra vonatkozó előrejelzést azzal, amit magukból az adatokból nyerhetünk.

A 13 éves átmenetvalószínűségek számításához a megbecsült modellből az alap-Markov modell esetében az átmenetmátrix hatványozásával (M^T) jutunk el, míg a Mover–Stayer modell esetében a (6) képletben szereplő $P_T = S + (I - S)M^T$ kifejezést kell kiszámítanunk. A számítási eredményeket és az adatokból nyert közvetlen becslést tartalmazza a *4. táblázat.*

4. TÁBLÁZAT

13 éves átmenetvalószínűségek a Markov modell, a Mover–Stayer modell alapján, valamint közvetlenül az adatokból becsülve

(13 years Mobility Matrices on the Base of Markov-model, the Mover–Stayer Model and Direct Estimation)

	Markov modell becslése (M^T)				*Mover-Stayer becslése (P_T)*				*Adatokból közvetlenül becsült, megfigyelt átmentek*			
Álla-pot	*1*	*2*	*3*	*4*	*1*	*2*	*3*	*4*	*1*	*2*	*3*	*4*
1	0,59	0,30	0,09	0,03	0,95	0,04	0,01	0	1,00	0	0	0
2	0,45	0,31	0,15	0,09	0,30	0,46	0,15	0,08	0,42	0,49	0,09	0
3	0,19	0,22	0,28	0,31	0,15	0,24	0,42	0,19	0,02	0,31	0,46	0,21
4	0,06	0,11	0,27	0,56	0,04	0,10	0,14	0,72	0	0	0,16	0,84
	Mobilitási m. = 75,40%				Mobilitási m. = 48,23%				Mobilitási m. = 40,35%			

Forrás: Saját szerkesztés.

A 13 éves horizontra számított mobilitás mindkét modell esetében magasabb, mint az empirikusan megfigyelt érték, a „túlbecslés" mértéke azonban jelentősen különbözik! Míg a Markov modell esetében a becsült 13 éves mobilitási mutató értéke 75%, addíg a Mover–Stayer modell esetében csak 48%. Ez utóbbi lényegesen közelebb van az empirikusan megfigyelhető 40%-os értékhez, mint az előző.

További lényeges különbség, hogy az empirikus 13-lépéses átmenetmátrixnak számtalan zérus értéke van: ezek olyan átmeneteket mutatnak, amelyekre nem volt példa a mintában. Például nem volt egyetlen olyan kistérség sem, amelyik a 13 év alatt az 1-es kategóriából a 3-as kategóriába került volna. Ennek valószínűsége a Markov modell alapján azonban 9%-ra tehető, míg a Mover–Stayer modell esetében 1%. Másként is megfogalmazhatjuk ugyanezt a különbséget: az empirikus átmenetmátrix főátlóinak értékeit az alap-Markov modell szisztematikusan és jelentősen *alulbecsüli*. Ez az alulbecslés azt jelenti, hogy míg a mintából közvetlenül megfigyelve 84% volt a relatív gyakorisága annak, hogy egy kistérség a legnagyobb jövedelmi kategóriából indulva (4) ott is marad, addig a Markov modell szerint ez pusztán 56%. Az alacsonyabb helybenmaradási valószínűség nagyobb mobilitást jelent, hiszen ha 13 év alatt mindössze 56% a helybenmaradás valószínűsége akkor 44% a mozgás, változás valószínűsége (szemben az empirikus 16%-kal).

A táblázatból és a számokkal való játékból látszik: a Markov modell számottevően felülbecsüli a hosszú távú mobilitást, és ehhez képest a Mover–Stayer modell felülbecslése kisebbnek látszik. Ahhoz persze, hogy a két modell előrejelző képességében lévő különbségről egyértelmű kijelentést tudjunk tenni számszerűen is meg kell vizsgálnunk, hogy az eltérés jelentős, azaz szignifikáns-e. Ezt méri a következő alpontban bemutatásra kerülő illeszkedés jósága teszt.

Az illeszkedés jósága

A két modell illeszkedésének jóságát ún. likelihood-arány teszttel fogjuk mérni. A likelihood-arány teszt akkor alkalmas két modell jóságának összehasonlítására, ha az egyik modell a másik általánosításának tekinthető. Jelen esetben erről van szó: a Mover–Stayer modell speciális esetének tekinthető a Markov modell, hiszen ha az S mátrix a zéró mátrix, akkor a Mover–Stayer modell alapegyenlete $P_1 = S + (I\text{-}S)M = M$ alakban lesz felírható, azaz visszakapjuk a kiinduló Markov modell.

Ezt az összefüggést nem csak formálisan lehet indokolni. A Mover–Stayer modellben abból a feltevésből indultunk ki, hogy a populáció heterogén, méghozzá két típusú egyedből áll: mozgókból és maradókból. Amennyiben a maradók aránya minden állapotban zérus (S=0), akkor az azt jelenti, hogy a populációban csak egyféle egyed van, és mindegyik egyed mozgását egy Markov mátrixszal lehet leírni, tehát visszakaptuk az alap-Markov modellt.

A teszt elvégzéséhez mindkét modell likelihood függvényének értékét ki kell számítanunk. A Mover–Stayer modell esetében ezt már megmutattuk a (7) képlet alatt, a teljesség kedvéért azonban mindkét modell *log*likelihood függvény képletét felírjuk *Frydman* (1984) alapján. A Markov modell log-likelihood függvénye a korábbi jelölések felhasználásával

$$\log L_{Markov} = \sum_{j=1}^{J} n_j(0)\log\left(\frac{n_j(0)}{n}\right) + \sum_{j,k} n_{jk}\log\left(\frac{n_{jk}}{n_j^*}\right)$$

ahol n_j* a j-állapotba lépések összes számát jelenti az utolsó előtti periódusig. A becsült paraméterek száma pedig $(J-1)\times J$. A Mover–Stayer modell log-likelihood függvénye hasonlóan

$$\log L_{Mover-Stayer} = \sum_{j=1}^{J}\left\{ n_j(0)\log\left(\frac{n_j}{n}\right) + (n_j(0)-n_j)\log\left(\frac{n_j(0)-n_j}{n_j(0)}\right) + \right.$$
$$\left. + (n_{jj} - Jn_j)\log(\hat{m}_{jj}) + \sum_{j\neq k} n_{jk}\log(\hat{m}_{jk})\right\}$$

ahol n a vizsgálatban szereplő egyedek számát jelöli. A Mover–Stayer modell ismeretlen paramétereinek száma J^2. Az általánosabb modell tehát pontosan *J* változóval tartalmaz többet, mint az alap-Markov modell.

A likelihood arány teszt alkalmazásakor azt nézzük, hogy a log-likelihood függvény értékében bekövetkező növekedés „megéri-e azt az áldozatot, amit a több becsülendő paraméter jelent". Ehhez vizsgálnunk kell a log-likelihood függvény értékében bekövetkező javulást, amit a specifikusról az általános modellre való áttéréssel nyerünk, azaz a tesztstatisztikát az

$$LR = 2(\log L_{Mover-Stayer} - \log L_{Markov})$$

kifejezésnek megfelelően számítjuk. A tesztstatisztika aszimptotikusan χ^2 eloszlást követ, az eloszlás szabadságfokát az általánosabb modell többletparamétereinek száma jelenti, ami a jelen alkalmazásban *J*.

A modell paramétereinek számításával egyidejűleg kiszámítottuk az egyes modell log-likelihood függvényének értékeit is. A Markov modell log-likelihood értékére $\log L_{Markov} = -1103{,}6$ adódott, míg a Mover–Stayer modell esetében ennek értéke $\log L_{Mover-Stayer} = -830{,}2$. A tesztstatisztika értéke ennek megfelelően LR=546,8, ami magasan szignifikáns, tekintettel arra, hogy a χ^2 eloszlás küszöbértéke 4 szabadságfog és 99,9%-os szignifikanciaszinten $\chi^2_{0.999}(4) = 18{,}5$. A teszt alkalmazásával megbizonyosodhattunk arról, hogy a két modell által adott hosszú távú előrejelzés nem csak szemmel láthatóan tér el egymástól, hanem statisztikai értelemben is különböző. Egyúttal arról is bizonyosságot szerezhettünk, hogy a Mover–Stayer modell valóban annyival jobban illeszkedik az adatokhoz, hogy megéri az a többletráfordítás, amely a bonyolultabb modell megoldása, a több becsülendő paraméter száma jelent.

Major Klára : Markov láncok használata a regionális jövedelemegyenlőtlenségek előrejelzésében.
Tér és Társadalom 21. évf. 2007/1. 53-67. p.

Összegzés, további kutatás irányok

A jövedelemi különbségek kutatásának fontos ágát képezi a Markov modellcsalád. Ezen modellek illesztésével a kutatók közvetlenül az eloszlás változását vizsgálják, ennek sajátosságait igyekeznek feltárni. A társadalmi folyamatokban meglévő magas perzisztencia, azaz alacsony jövedelmi mobilitás azonban a Markov lánc modellekhez képest bonyolultabb struktúrát igényel, „indokolatlan" illesztése téves következtetések levonásához vezet (magasabb mobilitás).

A Markov lánc modell alapfeltevése szerint a vizsgált jövedelmi folyamat stacioner, ami talán értelmezhető úgy, hogy változás *üteme* időben állandó, független attól, hogy milyen az egyedek állapotok közötti eloszlása, az egyes állapotokban tartózkodás hossza. Mindezek feltehetően túl erős feltevések a társadalmi folyamatok esetében.

A jelen alkalmazásban empirikusan megmutattuk, hogy Mover–Stayer modell statisztikai értelemben szignifikánsan jobban illeszkedik az adatokra, mint a Markov-lánc modell. Érthetően lassítja a jövedelmi dinamikát egy teljesen állandó, nem mozgó, helybenmaradó rész-populáció bevezetése. Mindezt azonban úgy teszi, hogy miközben a bonyolultabb struktúra jobban visszaadja a megfigyelési időszakban tapasztalt hosszabb távú mobilitást, a rövid távú, azaz 1-lépéses átmenetek becslését is megközelítően ugyanolyan pontossággal elvégzi.

Mindazonáltal a Mover–Stayer modell továbblépésén is el kell gondolkodnunk. Valójában a populáció kettéosztása két alpopulációra igen önkényes lépés. Arra vonatkozóan sem találunk semmilyen megbízható döntési kritériumot, hogy az egyébként folytonos jövedelmi adatokat hogyan osszuk be diszkrét jövedelmi állapotokba (kategóriákba). A kategóriák számának függvényében természetesen a Mover–Stayer modell becslési eredményei is változni fognak. Könnyen belátható, hogy amint növeljük a kategóriák számát, úgy csökken az egyes jövedelmi kategóriák terjedelme, emiatt értelemszerűen egyre csökkenni fog a maradók (stayerek) aránya az egyes kategóriákban. Természetesen létezik az a finomságú felosztás, amikor a maradók, mint alpopuláció egyszerűen eltűnik. A Mover–Stayer modell tehát várhatóan igen érzékeny ezen paraméterére. Mindezek miatt a modellből levonható kvantatív eredményeket mindenképpen fenntartással kell kezelni (pl. maradók aránya az 1-es jövedelmi kategóriában). A modell kvalitatív tulajdonságai, a vizsgálat üzenete azonban egyértelmű: a jövedelmi dinamika vizsgálata során az alacsony hosszú távú mobilitást figyelembe kell venni és a modellstruktúra kiválasztása során az illeszkedés jóságát feltétlenül meg kell vizsgálni. A jelen alkalmazás kiterjesztése a kevert Markov modellek családjára (több alpopuláció) folyamatban van.

Jegyzetek

[1] A tanulmány a „Jövedelmi differenciálódás szimulációs vizsgálata magyarországi kistérségek esetében" c. OTKA-60771 sz. kutatás keretében készült, elhagzott a Regionális modellek c. konferencián.

[2] A Markov modell alapfeltevését, azaz a vizsgált folyamat stacionaritását vetették el az alkalmazott khinégyzet teszt segítségével. A stacionaritás feltevése alapvető, ennek hiányában a Markov modell nem illeszthető, illetve ha mégis, a becslőfüggvények mechanikus alkalmazása torzított becslésekhez vezet.

Major Klára : Markov láncok használata a regionális jövedelemegyenlőtlenségek előrejelzésében.
Tér és Társadalom 21. évf. 2007/1. 53-67. p.

[3] A Markov láncok modell részleteiben nem járatos olvasónak javasoljuk Major, 2005 fejezetet, ahol az alapmodell részletesebb kifejtése található.

[4] Felhasználtuk, hogy nyilván $I^T = I$, másfelől $SI = SI^T = S$.

Irodalom

Bickenbach, F.–Bode, E. (2001) *Markor or not Markov – this should be the question.* Working Paper 1086, Kiel Institute of World Economics, Kiel.

Frydman, H. (1984) Maximum likelihood estimation in the mover-stayer model. – *Journal of the Amercian Statistical Association.* 79. 632–638. o.

Frydman, H. (2005) Estimation in the mixture of markov chains moving with different speeds. – *Journal of the American Statistical Association.* 100. 1046–1053. o.

Fuchs, C. – Greenhouse, J. B. (1988) The EM algorithm for maximum likelihood estimation in the mover-stayer model. – *Biometrics.* 44. 605–613. o.

leGallo, J.L. (2001) *Space-time analysis of gdp disparities among european regions: A markov chains approach.* Technical Report 2001–06. Laboratoire d'Analyse et de Techniques Economiques, Bourgogne.

Major K. (2005) Időbeli átmenetek: a Markov láncok. – Nemes Nagy J. (szerk.) *Regionális elemzési módszerek.* ELTE Regionális Földrajzi Tanszék – MTA–ELTE Regionális Tudományi Kutatócsoport, Regionális Tudományi Tanulmányok. 11. 124–135. o.

Quah, D.T. (1993) Empirical cross-section dynamics in economic growth. – *European Economic Review.* 37. 951–958. o.

Shorrocks, A.F. (1978) The measurement of mobility. – *Econometrica.* 46. 1013–1024. o.

Singer, B.–Spilerman, S. (1976) Some methodological issues in the analysis of longitudinal surveys. – *Annals of Economic and Social Management.* 5. 447–474. o.

Spilerman, S. (1978) Extensions of the mover-stayer model. – *American Journal of Sociology.* 78. 559–626. o.

FORECASTING REGIONAL INCOME INEQUALITIES BASED ON MARKOV MODELS

KLÁRA MAJOR

It is known that the simple Markov model overpredicts the long run horizon mobility of the income distribution process. Dissolving the homogeneity assumption of the Markov model we can have better forecasts. One generalization of the Markov model, the Mover–Stayer model assumes heterogenous population: some units are moving according to a common Markov chain but there are some (unknown) units whose are not moving at all. They are called stayers.

Based on Frydman, 1984 methodology we compute both the Markov and Mover–Stayer models for Hungarian micro-regions income data and find that the Mover–Stayer model fits better the regional relative income data than the simple Markov model. Using likelihood ratio test statistics we show that the difference is highly significant.