KÖZGAZDASÁGI SZEMLE, LXIX. ÉVF., 2022. NOVEMBER (1407–1432. o.)

NASZÓDI ANNA

Hogyan szálazzuk szét a megfigyelhető változások okait?

A tanulmány egy olyan dekompozíciós képlet alkalmazását mutatja be, amelylyel időbeli változásokat bonthatunk fel az azok hátterében álló tényezők *ceteris paribus,* valamint interakciós hatására. Ezt a dekompozíciós képletet *Biewen* [2014] javasolta. A jelentőségét abban látta, hogy – szemben az empirikus dekompozíciós irodalom néhány ágával – az ő megközelítése nem hagyja figyelmen kívül az interakciós tagokat. Az empirikus dekompozíciós irodalom azon ága is példa az interakciós tag gyakori kihagyására, amely aggregált szinten elemzi a párválasztási döntéseket. Ez a körülmény motiválja, hogy a tanulmány empirikus részében egy, az asszortatív házasságok irodalmában vizsgált probléma elemzésével illusztráljuk a Biewen-féle képlet alkalmazását. Az elemzést öt ország adatain végezzük el, ahol az adatok a párok iskolázottságáról szólnak. Az elemzés lehetővé teszi, hogy kvantitatív módon jellemezzünk egy közvetlenül nem megfigyelhető jelenséget, nevezetesen az iskolázottság szerinti homofíliának az egyik generációról a másikra történő megváltozását. Az interakciók figyelembevételével számolt hatások határozottan megerősítik *Naszódi–Mendonça* [2021] egyetlen ország – az Egyesült Államok – elemzésével kapott eredményét: a második világháborút követően volt egy olyan időszak, amikor a homofília gyengült – azaz az eltérő iskolázottságúak közötti házasságok elfogadottabbá váltak –, majd egy olyan, amelyet éppen az ellenkező irányú trend jellemzett. Ezt az eredményt azonban csak a dekompozíciós képlet gondos megválasztása mellett kapjuk robusztus formában.*

Journal of Economic Literature (JEL) kód: C02, D63, J12, I24.

* Kézdi Gábor szemléletet adott tanítványainak. Ebben a munkájában a CEU-n mint tanársegédje igyekeztem segíteni. A probit modell és a lineáris valószínűségi modell közötti választásról szóló gondolatai bátorságot adtak nekem ahhoz, hogy az elterjedt *Iterative Proportional Fitting* eljárás (más néven RAS algoritmus) rutinszerű alkalmazása helyett egy, a vizsgált kérdés megválaszolására alkalmasabb új eljárást fejlesszek és népszerűsítsek (lásd *Naszódi–Mendonça* [2021] és az NM-method című Wikipédia-szócikket: https://en.wikipedia.org/wiki/NM-method). A szerző ezúton köszöni meg *Lindner Attila, Telegdy Álmos,* valamint a két anonim bíráló hasznos megjegyzéseit.

Naszódi Anna, Európai Bizottság Közös Kutatóközpont *(Joint Research Centre, JRC),* Ispra, Olaszország, KRTK Közgazdaság-tudományi Intézet tiszteletbeli tagja (e-mail: anna.naszodi@gmail.com).

A kézirat első változata 2022. január 2-án érkezett szerkesztőségünkbe.
DOI: https://doi.org/10.18414/KSZ.2022.11.1407

Bevezetés

Dekompozíciók segítségével tipikusan arra a kérdésre keressük a választ, hogy egy megfigyelhető változó időbeli változásának mekkora része tudható be az azt mozgató egyes tényezők változásának. Egy klasszikus közgazdasági dekomponálási példa a következő: szeretnénk felbontani az egyének közötti béregyenlőtlenség Gini-mutatójának változását az eltérő végzettségűek relatív átlagbérében mért változás hatására – azaz az *árhatásra* – és az egyének végzettség szerinti eloszlásának változására, azaz az *összetételhatásra*.

Ez a tanulmány egy olyan dekompozíciós képlet alkalmazását mutatja be, amelylyel a megfigyelt változást nemcsak a tényezők (például végzettségspecifikus átlagbérek és végzettség szerinti eloszlás) *ceteris paribus hatására* bonthatjuk fel, hanem azok *interakciós hatására* is. Ezt a dekompozíciós képletet *Biewen* [2014] javasolta, és jelentőségét abban látta, hogy – ellentétben az empirikus dekompozíciós irodalom néhány ágával – az ő megközelítése nem hagyja figyelmen kívül az interakciós tagokat. Ha az interakciós tagok, azaz a tényezők együttes hatása nem elhanyagolható méretű, akkor a *ceteris paribus* hatások egy interakciós tagoktól eltekintő képlettel való mérése akár olyan mértékben is torzított lehet, ami már a kvalitatív következtetéseinket is megváltoztathatja.

Az utóbbi esetre mutatunk példát a tanulmány empirikus részében. Az ismertetésre kerülő példa nem a közgazdászok által gyakran elemzett bér- vagy jövedelemegyenlőtlenségről szól, hanem olyan egyenlőtlenségekről, amelyekről a házasságok, párválasztások árulkodnak.[1]

A házasságokat *Becker* [1973] vizsgálta először közgazdasági módszerekkel. Őt számtalan közgazdász követte, tovább erősítve a diszciplína hegemóniáját. Így az *asszortatív házasságok* irodalmához a demográfusok és szociológusok mellett egyre több közgazdász és ökonométer is hozzájárult.

Ez az irodalom többek között arra a kérdésre keresi a választ dekompozíciók segítségével, hogy az egymást követő *generációk házassági preferenciái* mennyire térnek el egymástól.[2] Így például, hogy hogyan változik egyik generációról a másikra az iskolázottság szerinti homofília,[3] azaz az azonos végzettségűekkel való házasodásra vonatkozó aggregált szintű preferencia, illetve az ezzel kapcsolatos társadalmi norma.

[1] Ebben a tanulmányban a házasság tágan definiált: házasságon mind a hivatalos egybekelést, mind pedig a párkapcsolatban való együttélést értjük. Ugyancsak tágan értelmezzük a férjek és feleségek csoportját, beleértve az együtt élő partnereket.

[2] A jövedelem- és vagyoni egyenlőtlenségre ható tényezők feltérképezését a közgazdászok egyre szélesebb köre végzi Thomas Piketty angol nyelvű könyvének 2014-es megjelenése óta (*Piketty* [2015]). Az asszortatív házasságok irodalmában vizsgált kérdések megválaszolása is hozzájárul ehhez a munkához, ugyanis az azonos végzettségűek – így a jó eséllyel azonos jövedelmi kategóriába tartozók – közötti házasságok gyakoribbá válása növeli a háztartások közötti gazdasági természetű egyenlőtlenséget (lásd *Mihályi–Szelényi* [2019]). Ebben a tanulmányban a homogám házasságok és az egyenlőtlenség egy, a fentitől eltérő kapcsolatára helyezzük a hangsúlyt: a homogám házasságok arányát meghatározó egyik tényezőt az egyenlőtlenség mutatójaként értelmezzük.

[3] A homofíliára és a homogámia gyakoriságára (azaz a homogám párok arányára) akár egymás szinonimájaként is gondolhatnánk, ám a kettő nem ugyanaz. Az utóbbi a különböző végzettségűek eloszlásától is függ a homofília, azaz a vágyak/preferenciák mellett.

Ennek a kérdésnek a vizsgálatát az motiválja – azontúl, hogy az egyes dekompozíciós képletek közötti különbséget illusztrálhatjuk vele –, hogy preferenciaváltozásokra vonatkozó mérésünk átfogó képet ad a különböző szintű iskolai végzettséggel rendelkezők közötti egyenlőtlenségek alakulásáról.[4]

Az egyenlőtlenség dinamikájáról a házasságpiac elemzése alapján kialakított képünk sajátosságai a következők. Egyrészt ezt a képet az ökonométer számára a házasságpiac egyensúlyának változása rajzolja ki. Másrészt, feltételezésünk szerint, ha az iskolázottság szerinti homofília erősödik (gyengül), akkor ez az eltérő végzettségűek közötti – a potenciális partnereket érdeklő – számtalan dimenzió szerinti egyenlőtlenség növekedésének (mérséklődésének) a jele. Harmadrészt, vélhetően minden olyan dimenzió (például jövedelem, vagyon, egészség, várható élettartam) aggregálásra kerül, amelyet a párválasztásnál fontosnak gondolnak az emberek. Így a homofília változásának számszerűsítésével az egyenlőtlenségeknek a gazdasági dimenzióin túli dimenzióit is figyelembe lehet venni. Végül, az egyenlőtlenség változásáról a homofíliaváltozás alapján készített kép esetében az egyes dimenziók súlyozását a házasságpiac végzi el.

Az aggregált preferenciák változását azok egy megfigyelhető, skalár eredményváltozóra gyakorolt hatásán keresztül mérjük a dekompozícióval. Az eredményváltozónak választott változónk ebben a tanulmányban *az iskolai végzettség szerinti homogám párok aránya az egyes generációkban*. A szociológus *Kalmijn* [1998] szerint a homogámia gyakoriságának alakulása a vizsgálatunk középpontjában álló házassági preferenciák *ceteris paribus* megváltozásán túl attól is függ, hogy hogyan változik a házasulandó férfiak és nők *végzettség szerinti eloszlása* a generációk között a preferenciák változatlansága mellett. Ezenkívül a homogám párok aránya függ a preferenciák, valamint az iskolázottság változásának *interakciós hatásától* is, mivel az iskolázottság változásához bizonyos mértékben igazodnak a preferenciák, amelyek pedig befolyásolják a homogámia gyakoriságát.[5] Tehát ha a preferenciaváltozás *ceteris paribus* hatását szeretnénk kimutatni, akkor mind az eloszlásváltozás hatására, mind az interakciós hatásra kontrollálnunk kell.

Érdemes áttekinteni, hogy a jellemzően közgazdászok által elemzett béregyenlőtlenség dekompozíciójához képest miben tér el az itt vizsgált kérdés.

– Egyrészt, a fiatal felnőttek közötti házasságok esetében különösen fontos lehet az *interakciós hatás*, hiszen elképzelhető, hogy a preferenciák érdemben igazodnak a piaci felhozatal változásához, és/vagy a magasabb iskolázottság elérésére (ellen)ösztönzőként hat a házassági preferenciák megváltozására. Az interakciós tag megmérése és az arra való kontrollálás empirikus feladat. Ezt a feladatot azonban csak egy interakciós tagot is tartalmazó dekompozíciós képlettel tudjuk elvégezni.

– Másrészt, a preferenciákon *aggregált szintű preferenciákat* kell érteni, amelyek generációkat jellemeznek, nem egyéneket. Amit a szociológusok a partnerek

[4] A homofília ilyen értelmezéséhez az a mögöttes feltevésünk, hogy az eltérő iskolázottság nem pusztán valamiféle másságot, hanem egyenlőtlenséget jelent.

[5] Egy további tényező a párosítási mechanizmus időbeli változása, amelyre az *online* párkeresés elterjedése is példa. Míg ebben a tanulmányban eltekintünk ennek a hatásnak a vizsgálatától, a *Naszódi–Mendonça* [2022] tanulmány diszkutálja azt.

iskolázottságára vonatkozó aggregált szintű házassági preferenciának, társadalmi normának vagy homofíliának hívnak, azt a közgazdaságtan nyelvére úgy fordíthatjuk le, mint a különböző végzettségeknek a házasságpiac által meghatározott értéke.

– Harmadrészt, ez alapján – a béregyenlőtlenség példájával analóg módon – valamiféle *árhatásként* gondolhatunk a preferenciatényező hatására. Bár ez a megközelítés hasznos, mégis megjegyzendő, hogy a piac értékítélete a piaci szereplők lecserélődésével, azaz a generációváltással változik leginkább: egy adott házassági preferenciákkal rendelkező generáció helyét átveszi a következő, másfajta preferenciákkal rendelkező – és egyben eltérő iskolázottságú – generáció.[6] Tehát a két feladat közötti analógia nem tökéletes, hiszen a házasságpiaci árhatás és az összetételhatás egyidejűleg jelentkezik.

A házasságokról szóló dekompozíciós probléma ezen sajátossága azért is fontos, mert – a béregyenlőtlenségről szóló problémával ellentétben – nincsenek olyan megfigyeléseink, amelyeken az árhatás már mutatkozik, de az összetételhatás még nem, és fordítva. Az adatokban látott, ilyen értelemben vett alacsony varianciát azonban ellensúlyozzák modellfeltevéseink, amelyekkel már lehetővé válik a *ceteris paribus* hatások és az interakciós hatások identifikációja.[7]

– Végül, míg a házasulandók iskolázottság szerinti eloszlását meg tudjuk figyelni – csakúgy, mint a béregyenlőtlenségről szóló példában az átlagos béreket és a végzettség szerinti eloszlást –, addig az aggregált házassági preferenciát (avagy a piac értékítéletét) még definiálni sem könnyű, nemhogy közvetlenül mérni. Ugyanakkor szerencsére a dekompozíciós módszer segítségével számszerűsíteni tudjuk a preferenciák változásának mértékét. Hangsúlyozzuk, hogy az *identifikációt* az biztosítja, hogy ki tudjuk számolni, hogy dekompozíciónk mekkora és milyen előjelű *ceteris paribus* hatást tulajdonít a preferenciák megváltozásának.

A dekompozíció elvégzéséhez csupán azt kell definiálnunk, hogy mit értünk az aggregált preferenciák változatlanságán. Ehhez a definícióhoz – *Naszódi–Mendonça* [2021]-et követve – a *Liu–Lu* [2006] által kifejlesztett ordinális mutatót fogjuk használni, amelyet a vizsgált generációkba tartozó férjek és feleségek végzettség szerinti együttes eloszlásából számolunk. A mutató elkészítéséhez sorba rendezzük az összes olyan lehetséges azonos peremeloszlású együttes eloszlást, amelyeknél a homogám párok aránya legalább akkora, mint annak várható értéke véletlenszerű párosítás mellett. Majd a percentilisértéküket rendeljük az eloszlásokhoz. Az együttes eloszlásokra

[6] Elméletileg egy csoport aggregált preferenciáinak változásához az egyéni preferenciák változása is hozzájárulhat. A preferenciák változásának korhatásra – tehát a házasságpiaci szereplők életkorának növekedésével együtt járó hatásra –, valamint generációváltási hatásra való szétbontásáról lásd *Naszódi* [2021*a*].

[7] Az asszortatív házasságoktól eltérő jelenségek dekompozícióval való vizsgálatára több olyan példa is van a közgazdaságtan irodalmában, ahol nagy varianciájú mikroadatokból történik a *ceteris paribus* hatások és az interakciós hatások identifikációja. Például *Foster és szerzőtársai* [2008] az iparágszintű termelékenységnövekedést bontja fel az egyes cégek termelékenységnövekedésére, másrészt az eltérő termelékenységű cégek piacirészesedés-változása okozta hatásra, végül pedig az előző két tényező interakciós hatására.

ható aggregált preferenciákat azonosnak tekintjük az együttes eloszlások Liu–Lu-féle mutatóinak azonossága esetén.

Az asszortatív házasságok empirikus irodalma meglehetősen megosztott azzal kapcsolatban, hogy a 20. századot milyen trendek jellemezték.[8] *Rosenfeld* [2008] megjegyzi, hogy az irodalom empirikus eredményeinek ellentmondásossága részben az eltérő adatokból ered, részben azonban az alkalmazott módszerek különbözőségéből is fakad. Például az eltérő eredmények más és más dekompozíciós képleten alapszanak.[9]

Biewen [2014] kétféle dekompozíciós képletet ismertet. Az egyik a szekvenciális dekompozíciós képlet. Ezt a képletet alkalmazza több tanulmány is az asszortatív házasságok irodalmában (lásd például *Dupuy–Weber* [2022]). A másik képlet egy additív interakciós dekompozíciós képlet, amelyet Biewen a szekvenciális dekompozíciós képlet alternatívájaként javasol.

A következőkben először összehasonlítjuk a Biewen által javasolt képletet a szekvenciális képlettel, majd bemutatjuk az interakciós képlet mellett szóló, *Biewen* [2014] által felhozott elméleti megfontolásokon alapuló érveket. Végül a tanulmány empirikus részében illusztráljuk a képletek alkalmazását öt ország három évtizedet átfogó házassági és együttélési adatain.

Kétféle dekompozíciós képlet

Dekompozíciós képletek kéttényezős esetre

A következőkben kétféle dekompozíciós képletet mutatunk be a kéttényezős esetre.[10] A megfigyelhető eredményváltozó értékét a valós értékű, determinisztikus f függvény, valamint az időben változó p és v tényezők határozzák meg. A dekompozíció segítségével szeretnénk megtudni, hogy az f függvény nulladik és első időszak közötti változása, azaz $f(p_1, v_1) - f(p_0, v_0)$ mennyiben tudható be külön-külön az egyes tényezők változásának.

A jelölések megértését egy konkrét példa segítheti. A bevezetőben tárgyalt, házasságokról szóló példában az f függvény a homogám párok arányát jelöli egy adott (megfigyelt vagy elképzelt) generációban. A nulladik időszakban és az első időszakban megfigyelt férjek ugyanahhoz a korosztályhoz tartoznak a különböző, egymást váltó generációkban. Míg v_0 a korábban született férjek és feleségeik végzettség szerinti eloszlását jelöli, addig v_1 a később született férjekét és feleségeikét. Hasonlóan, míg p_0

[8] *Rosenfeld* [2008] irodalomáttekintése szerint a tanulmányok egy része alapján az Egyesült Államokban a végzettség szerinti házassági endogámia az 1930-as éveket követően változatlan volt, vagy csökkenő, míg a tanulmányok egy másik csoportja szerint a késői 20. századot éppen az endogámia növekvő trendje jellemezte. Végül a Rosenfeld által áttekintett tanulmányok harmadik csoportja szerint az Egyesült Államokban az 1960-as éveket követően a végzettség szerinti endogámia mérséklődött.

[9] A különböző módszerekkel kapott eredmények eltérése abból is adódik, hogy eltérő mutatókkal jellemzik a házassági preferenciákat. Erről részletesebben lásd *Naszódi–Mendonça* [2021]-et, amely a *Liu–Lu* [2006] által kifejlesztett mutatót jobbnak találja mind analitikus, mind empirikus tulajdonságai alapján, mint az irodalomban elterjedt alternatív mutatókat.

[10] A képleteket tetszőleges számú tényező esetére kiterjesztve *Biewen* [2014] ismerteti.

az előbbi csoport aggregált házassági preferenciáját jelöli, addig p_1 az utóbbi csoportét. A preferenciákat a *Liu–Lu* [2006] által javasolt mutatóval, valamint a mutató egy általánosított verziójával jellemezzük (lásd a *Függelék A* részét).

A *szekvenciális dekompozíció* az (1) és a (2) képletek egyikén alapszik,[11]

$$f(p_1, v_1) - f(p_0, v_0) = \overbrace{\left[f(p_1, v_0) - f(p_0, v_0)\right]}^{\Delta p \text{ miatt}} + \overbrace{\left[f(p_1, v_1) - f(p_1, v_0)\right]}^{\Delta v \text{ miatt}}, \tag{1}$$

$$f(p_1, v_1) - f(p_0, v_0) = \overbrace{\left[f(p_1, v_1) - f(p_0, v_1)\right]}^{\Delta p \text{ miatt}} + \overbrace{\left[f(p_0, v_1) - f(p_0, v_0)\right]}^{\Delta v \text{ miatt}}, \tag{2}$$

attól függően, hogy mi a p és v tényezők megváltozásának feltételezett sorrendje. Pontosabban, az (1) egyenlet annak az esetnek felel meg, ahol p előbb változik meg, mint v. Míg a (2) egyenlet azt feltételezi, hogy v változása következik be előbb.

A *Biewen* [2014] által javasolt képlet a következő:

$$\begin{aligned} f(p_1, v_1) - f(p_0, v_0) = {} & \overbrace{\left[f(p_1, v_0) - f(p_0, v_0)\right]}^{\Delta p \text{ miatt}} + \overbrace{\left[f(p_0, v_1) - f(p_0, v_0)\right]}^{\Delta v \text{ miatt}} + \\ & + \underbrace{\left[f(p_1, v_1) - f(p_1, v_0) - f(p_0, v_1) + f(p_0, v_0)\right]}_{\text{interakciós tag}}. \end{aligned} \tag{3}$$

A (3) képlet $f(p_1, v_0) - f(p_0, v_0)$ tagját és $f(p_0, v_1) - f(p_0, v_0)$ tagját természetes rendre a p és v tényezők *ceteris paribus hatásának* nevezni, hiszen azok a másik tényező változatlansága melletti f-beli változások. Az azonban talán kevésbé magától értetődő, hogy a „maradéktagot” $[f(p_1, v_1) - f(p_1, v_0) - f(p_0, v_1) + f(p_0, v_0)]$ miért nevezhetjük *interakciós tagnak*, avagy a tényezők együttes megváltozása okozta hatásnak. A következő példa magyarázatot ad erre.

Tegyük fel, hogy az f egy $\mathbb{R} \times \mathbb{R} \to \mathbb{R}$ függvény, amely a (p_0, v_0) pontban végtelenszer deriválható, és amelyet a Taylor-sora előállít. Ekkor az $f(p_1, v_1)$ értéket kifejezhetjük az f függvény (p_0, v_0) pont körüli sorba fejtésével:

$$\begin{aligned} f(p_1, v_1) = {} & f(p_0, v_0) + \sum_{k=1}^{\infty} \frac{1}{k!} \frac{\partial^k f(p, v)}{\partial p^k}\bigg|_{p_0, v_0} (p_1 - p_0)^k + \sum_{k=1}^{\infty} \frac{1}{k!} \frac{\partial^k f(p, v)}{\partial v^k}\bigg|_{p_0, v_0} (v_1 - v_0)^k + \\ & + \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} \frac{1}{(i+j)!} \frac{\partial^{i+j} f(p, v)}{\partial p^i \partial v^j}\bigg|_{p_0, v_0} (p_1 - p_0)^i (v_1 - v_0)^j. \end{aligned} \tag{4}$$

A Biewen-féle dekompozíció által a p megváltozásának és a v megváltozásának tulajdonított *ceteris paribus* hatások a fenti Taylor-sor részei, mivel azok rendre

[11] Az irodalom nem egységes a dekompozíciós képletek elnevezésének tekintetében. Például *Biewen* [2012] a szekvenciális dekompozíciós képletet klasszikus Oaxaca–Blinder-dekompozícióként hivatkozza. *Blinder* [1973] és *Oaxaca* [1973] a diszkriminációt mint közvetlenül nem megfigyelhető jelenséget vizsgálják. Dekompozíciójuk során az eltérő csoportokba tartozó emberek (nők/férfiak, fekete/fehér bőrszínűek) bérkülönbségét bontják fel egyfelől megfigyelhető tényezők hatására (például végzettség, munkatapasztalat), másfelől az egyik csoporttal szembeni diszkrimináció hatására.

$$f(p_1, v_0) - f(p_0, v_0) = \sum_{k=1}^{\infty} \frac{1}{k!} \frac{\partial^k f(p, v)}{\partial p^k}\Big|_{p_0, v_0} (p_1 - p_0)^k \quad \text{és} \tag{5}$$

$$f(p_0, v_1) - f(p_0, v_0) = \sum_{k=1}^{\infty} \frac{1}{k!} \frac{\partial^k f(p, v)}{\partial v^k}\Big|_{p_0, v_0} (v_1 - v_0)^k. \tag{6}$$

Az (5) és a (6) egyenleteket behelyettesítve a (4) egyenletbe a következőt kapjuk:

$$f(p_1, v_1) = f(p_0, v_0) + \left[f(p_1, v_0) - f(p_0, v_0)\right] + \left[f(p_0, v_1) - f(p_0, v_0)\right] +$$

$$+ \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} \frac{1}{(i+j)!} \frac{\partial^{i+j} f(p, v)}{\partial p^i \partial v^j}\Big|_{p_0, v_0} (p_1 - p_0)^i (v_1 - v_0)^j. \tag{7}$$

Így a „maradéktagra" igaz, hogy

$$f(p_1, v_1) - f(p_1, v_0) - f(p_0, v_1) + f(p_0, v_0) =$$

$$= \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} \frac{1}{(i+j)!} \frac{\partial^{i+j} f(p, v)}{\partial p^i \partial v^j}\Big|_{p_0, v_0} (p_1 - p_0)^i (v_1 - v_0)^j. \tag{8}$$

A (8) egyenlet jobb oldalán szereplő $(p_1 - p_0)^i (v_1 - v_0)^j$ szorzat csak akkor vehet fel nullától eltérő értéket, ha $p_1 \neq p_0$ és $v_1 \neq v_0$. Azaz, ha a két tényező mindegyike megváltozik. Ezért nevezhetjük az $f(p_1, v_1) - f(p_1, v_0) - f(p_0, v_1) + f(p_0, v_0)$ „maradéktagot" együttes, avagy interakciós hatásnak.

A Taylor-soros példánkkal ellentétben a házasságokról szóló, empirikusan is elemzendő példa esetében a p preferenciaváltozó, akárcsak a v eloszlásváltozó nem skalárértékű; az f függvény pedig $[0, 1]^{2\times 2} \times \mathbb{N}^5 \to [0, 1]$ (lásd a *Függelék B* részét). Ugyanakkor, ha ezt az f függvényt is előállítja a Taylor-sora, akkor a sorba fejtéssel triviális megmutatni, hogy a teljes változás, azaz $f(p_1, v_1) - f(p_0, v_0)$ felbontható a *ceteris paribus* hatásokra, valamint egy olyan „maradéktagra", amely – éppúgy, mint a fenti $f\colon \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ függvény esetében – interakciós hatásként interpretálható. Tehát az e tanulmányban elemzett probléma esetében is megalapozott lehet a „maradéktagra" mint interakciós hatásra gondolni a sorbafejthetőségtől függően.

Egy szembetűnő különbség a Biewen-féle képlet és a szekvenciális képletek között az, hogy amíg a (3) képlet használatához mind az $f(p_1, v_0)$, mind az $f(p_0, v_1)$ értékeket ismernünk szükséges, addig a szekvenciális dekompozíció egyes verzióihoz elegendő csak az egyiket.

A házasságokról szóló példánál maradva, az $f(p_1, v_0)$ azt adja meg, hogy mennyi lenne a homogám párok aránya egy olyan képzeletbeli generáció esetében, amelynek preferenciái azonosak a házasságpiacon az első időszakban leginkább aktív generáció preferenciáival, míg az iskolázottsági eloszlásuk azonos a házasságpiacon a nulladik időszakban legaktívabb generáció iskolázottságával.

Az $f(p_0, v_1)$ a másik tényellentétes esetre adja meg a homogám párok arányát, ahol a preferenciákat a nulladik időszakból vesszük, míg az iskolázottsági eloszlásokat az első időszakból. Az *f* függvény definiálásáról lásd *Naszódi–Mendonça* [2021]-et vagy a *Függelék B* részét.

A dekompozíciós képletek közötti további különbségeket és a (3) képlet alkalmazása mellett szóló érveket az Elméleti megfontolások a dekompozíciós képlet megválasztásához című részben tárgyaljuk. Előtte azonban numerikus példákkal illusztráljuk a dekompozíciós képletek házasságokra való alkalmazását.

A dekompozíciós képletek alkalmazásának illusztrálása numerikus példákkal

A következőkben két dekompozíciós példát mutatunk be. Mindkét példában a házasságpiac tökéletes, és éppannyira egyszerű, mint a *Liu–Lu* [2006] modellben. Nincsenek keresési költségek, és minden szereplő tökéletesen informált a többi szereplő iskolai végzettségéről, amely az egyetlen releváns párválasztási tulajdonság a modellben. A szereplők iskolai végzettsége kétféle lehet: alacsony vagy magas. Csak két generáció van, és mindkét generációban azonos számú házasulandó férfi és nő van. Minden házasulandó heteroszexuális, monogám, és össze fog házasodni valakivel a saját generációjából.

Mint majd látni fogjuk, a két dekompozíciós példa abban különbözik egymástól, hogy míg az elsőben az interakciós tag nulla, addig a másodikban jelentősen különbözik nullától.[12] Emiatt az első példában a dekompozíció eredménye független attól, hogy melyik dekompozíciós képletet alkalmazzuk. Viszont a második példában a komponensek nagyon is érzékenyek a képletválasztásra.

Lássuk az első példát! Ebben a párok végzettség szerinti együttes eloszlását az egyik generációban a következő 2×2-es kontingenciatábla írja le: $\begin{bmatrix} 925 & 125 \\ 75 & 375 \end{bmatrix}$.

		A feleségek végzettsége	
		alacsony	magas
A férjek végzettsége	alacsony	925	125
	magas	75	375

Míg egy másik generációban az együttes eloszlás: $\begin{bmatrix} 500 & 100 \\ 0 & 400 \end{bmatrix}$. Ez utóbbi generáció tagjai később születettek, mint az előbbi generáció tagjai.

A homogám párok aránya az első, korábban születettek generációjában $f(p_0, v_0) = (925 + 375)/1500 = 86{,}\dot{6}\%$. A később születettek generációjában pedig $f(p_1, v_1) = (500 + 400)/1000 = 90\%$. Tehát a homogám párok aránya $3{,}\dot{3}$ százalékponttal nagyobb a később születettek generációjában.

A dekompozíciós képletekkel kiszámolhatjuk, hogy ez a növekmény mennyiben tudható be a két generáció eltérő *iskolázottságának*, az eltérő *preferenciáiknak*, valamint esetlegesen a preferenciák és az iskolázottság *együttes hatásának*.

[12] Könnyen belátható, hogy ha az interakciós tag nulla – ahogy az első példánkban is –, akkor az (1) és a (2) szekvenciális dekompozíciós képletek ugyanúgy bontják fel a teljes megfigyelt változást, ahogy a Biewen által javasolt (3) képlet. Ennek bizonyítását az olvasóra bízzuk.

A két kontingenciatábla összehasonlításából látható, hogy a később születettek sokkal *iskolázottabbak*, mint a korábban születettek. Az előbbiek körében majdnem minden második ember magas végzettségű (ez az arány 40 százalék a férfiak körében és 50 százalék a nők körében), míg a korábban születettek generációjában a férfiaknak csak a 30 százaléka, a nőknek csak az egyharmada magasan iskolázott.

De vajon hogyan változtak a *házassági preferenciák*? Ennek a kérdésnek a megválaszolásához kiszámoljuk a két kontingenciatáblának a *Függelék A* részének (F3) képletével megadott Liu–Lu-értékeit. Ezek rendre 3/4 és 1 [$(375-150)/(450-150)=3/4$ és $(400-200)/(400-200)=1$].

Modellünkben egy magasabb Liu–Lu-érték a homogámia iránt erősebb aggregált preferenciát jelez, amelyben a piaci egyensúlyt – súrlódások és információs tökéletlenségek hiányában – kizárólag a vágyak és a lehetőségek határozzák meg. Tehát nemcsak a homogámia gyakorisága nagyobb a később születettek generációjában, hanem a homofília is erősebb körükben.

Amit nem tudhatunk meg a két generáció Liu–Lu-értékeinek az összehasonlításából, az az, hogy a később születettek generációjában a homofília sokkal erősebb volt-e, vagy csak valamicskét volt erősebb, mint a korábban születettek körében. Ugyanis a Liu–Lu-mutató nem kardinális, hanem ordinális mutató. A dekompozíció segítségével azonban kardinális mutatóvá transzformáljuk a Liu–Lu-mutatók eltérését, amikor meghatározzuk, hogy a homogám párok arányában mért $3,\dot{3}$ százalékpontos különbség mennyiben tudható be pusztán a preferenciák változásának.

A dekompozícióhoz ki kell számolnunk, hogy milyen lett volna a piaci egyensúly, ha csak a preferenciák, illetve ha csak az iskolázottság változott volna meg az egyik generációról a másikra. Az előbbi esetben a férjek és feleségek iskolázottságának együttes eloszlását az $\begin{bmatrix} 1000 & 50 \\ 0 & 450 \end{bmatrix}$ kontingenciatábla adja meg. Ezt a táblát a *Függelék B* részében is ismertetett NM-módszerrel számolhatjuk ki.

A következőkben ellenőrizzük, hogy ez a tábla valóban a keresett tényellentétes tábla-e. A tábla Liu–Lu-értéke éppúgy 1 [$(450-150)/(450-150)=1$], mint a később születettek táblájának; valamint a tábla sor- és oszlopösszegei rendre azonosak a korábban születettek táblájának sor- és oszlopösszegeivel. Tehát ez a tábla valóban a keresett tényellentétes tábla. Ebben a táblában a homogám párok aránya $f(p_1, v_0) = (1000+450)/1500 = 96,\dot{6}\%$.

Ha azonban egy olyan generációt képzelünk el, amelynek az iskolai végzettsége a később születettek végzettségével egyezik meg, míg a házassági preferenciái a korábban születettekével, akkor annak kontingenciatáblája $\begin{bmatrix} 450 & 150 \\ 50 & 350 \end{bmatrix}$. Ellenőrzésképpen ennek a táblának is kiszámoljuk a Liu–Lu-értékét. Ez éppúgy 3/4 [$(350-200)/(400-200)=3/4$], mint a korábban születettek táblája esetében. Az is látható, hogy a sor- és oszlopösszegei ennek a táblának rendre azonosak a később születettek táblájának sor- és oszlopösszegeivel. Tehát ez a tábla az általunk keresett másik tényellentétes tábla. Ebben a homogám párok aránya $f(p_0, v_1) = (450+350)/1000 = 80\%$.

Ismerve a homogám párok arányát mindkét tényellentétes esetben, már el tudjuk végezni az (1), (2) és (3) képletekkel a dekompozíciókat.

Az *f* függvény változását a *Biewen által javasolt képlettel* a következőképpen bonthatjuk fel:

$$f(p_1, v_1) - f(p_0, v_0) = 3{,}\dot{3} \text{ %-pont} = \overbrace{\left[96{,}\dot{6}\% - 86{,}\dot{6}\%\right]}^{+10 \text{ %-pont } \Delta p \text{ miatt}} + \overbrace{\left[80\% - 86{,}\dot{6}\%\right]}^{-6{,}\dot{6} \text{ %-pont } \Delta v \text{ miatt}} +$$
$$+ \underbrace{\left[90\% - 96{,}\dot{6}\% - 80\% + 86{,}\dot{6}\%\right]}_{0 \text{ interakciós hatás}}.$$

Az *f* függvény változásának a *szekvenciális képletekkel* való felbontása pedig

$$f(p_1, v_1) - f(p_0, v_0) = 3{,}\dot{3} \text{ %-pont} = \overbrace{\left[96{,}\dot{6}\% - 86{,}\dot{6}\%\right]}^{+10 \text{ %-pont } \Delta p \text{ miatt}} + \overbrace{\left[90\% - 96{,}\dot{6}\%\right]}^{-6{,}\dot{6} \text{ %-pont } \Delta v \text{ miatt}},$$

illetve

$$f(p_1, v_1) - f(p_0, v_0) = 3{,}\dot{3} \text{ %-pont} = \overbrace{\left[90\% - 80\%\right]}^{+10 \text{ %-pont } \Delta p \text{ miatt}} + \overbrace{\left[80\% - 86{,}\dot{6}\%\right]}^{-6{,}\dot{6} \text{ %-pont } \Delta v \text{ miatt}}.$$

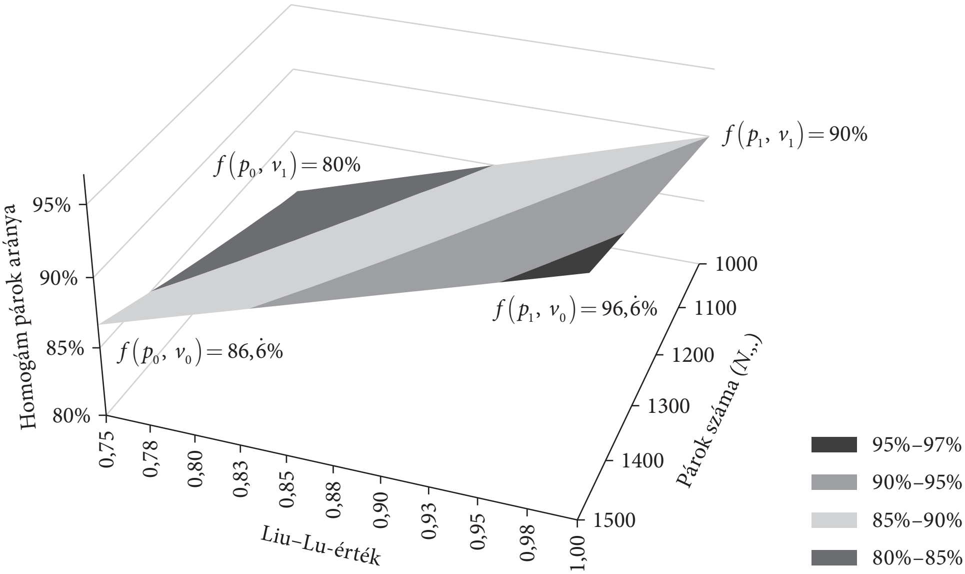
Tehát ebben a példánkban az alkalmazott dekompozíciós képlettől függetlenül arra jutunk, hogy ha csak a preferenciák változtak volna meg az egyik generációról a másikra, akkor 10 százalékponttal nőtt volna a homogám párok aránya. Ha viszont csak az iskolázottság változott volna, akkor $6{,}\dot{6}$ százalékponttal csökkent volna ugyanez az arány.

Az $f(p, v)$ függvény kiszámolható a megfigyelt két generációt jellemző tulajdonságok konvex kombinációira. Azaz bármely (p, v) párra, ahol $p \in [p_0, p_1]$ skalár – példánkban $p \in [3/4, 1]$ – és $v = \alpha v_0 + (1-\alpha) v_1$ vektor, ahol $\alpha \in [0, 1]$. Példánkban $v_0 = [1500, 1000, 500, 1050, 450]$ vektor, amely a korábban születettek között az összes pár számát (1500), valamint a kontingenciatáblájuk oszlopösszegét és sorösszegét tartalmazó vektor; valamint $v_1 = [1000, 500, 500, 600, 400]$ a később születettek között az összes pár számát (1000), valamint a kontingenciatáblájuk oszlopösszegét és sorösszegét tartalmazó vektor. Az *1. ábrán* láthatjuk az $f(p, v)$ függvényt a p skalár és a v vektor első elemének, azaz a párok számának függvényében.

Lássuk második példánkat! Ezt a példát rövidebben ismertetjük, mint az első példát, mivel itt már az olvasóra bízzuk a közölt eredmények ellenőrzését. Ebben a példában a párok végzettség szerinti együttes eloszlását megadó kontingenciatábla a korábban születettek generációjában éppúgy $\begin{bmatrix} 925 & 125 \\ 75 & 375 \end{bmatrix}$, mint az első példánkban, azonban a később születettek kontingenciatáblája az első példabelitől eltérő, mivel itt az $\begin{bmatrix} 150 & 45 \\ 5 & 600 \end{bmatrix}$.

Ha az $f(p, v)$ függvényt erre a példára is kiszámoljuk az ennek a példának megfelelő (p, v) párokra, majd a függvényt ábrázoljuk, akkor a *2. ábrát* kapjuk. Ez az ábra lényegesen különbözik az *1. ábrától:* míg az *1. ábrán* a függvény pontjai egy síkra esnek, addig a *2. ábrán* bemutatott függvény erősen nemlineáris.

1. ábra
A homogámia gyakorisága lineárisan függ a vágyaktól és lehetőségektől (első numerikus példa)



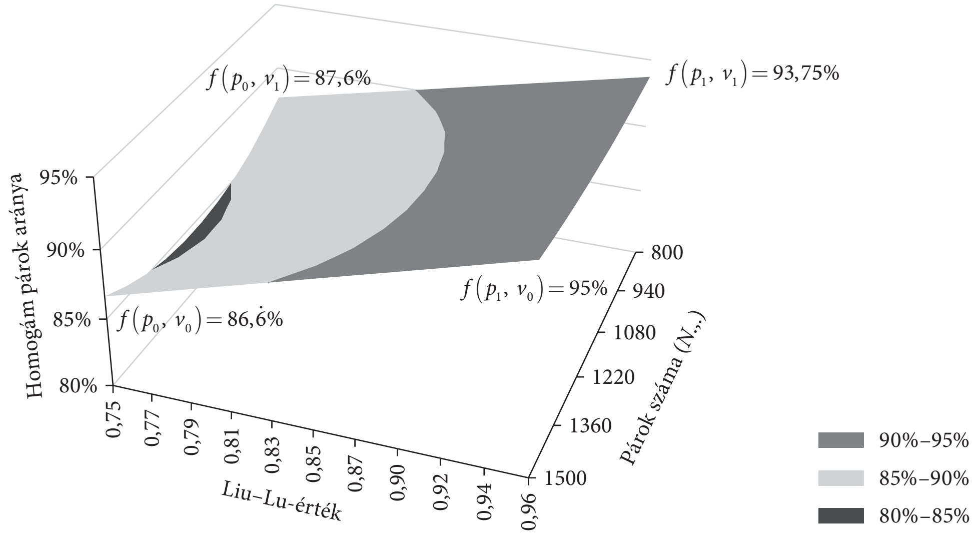
Megjegyzés: az első példában a korábban született férjek és feleségek megfigyelt végzettség szerinti együttes eloszlása $\begin{bmatrix} 925 & 125 \\ 75 & 375 \end{bmatrix}$, míg a később születetteknél ugyanez $\begin{bmatrix} 500 & 100 \\ 0 & 400 \end{bmatrix}$. Ezért a homogám párok aránya az előbbi generációban $f(p_0, v_0) = (925 + 375)/1500 = 86{,}\dot{6}\%$, míg a később születettek generációjában $f(p_1, v_1) = (500 + 400)/1000 = 90\%$. Az ábra az összes olyan lehetséges generáció homogám párarányát mutatja, amelynél a férjek és feleségek végzettség szerinti együttes eloszlásának Liu–Lu-értéke a két megfigyelt generációra jellemző Liu–Lu-érték konvex kombinációja. Valamint a párok száma, akárcsak a végzettség szerinti eloszlások is, konvex kombinációi a két megfigyelt generációra jellemzőnek. Látható, hogy a homogám párok aránya lineárisan függ mind a Liu–Lu-értéktől, azaz a vágyaktól, mind a lehetőségeket meghatározó eloszlásoktól.

A látott nemlinearitás összefügg azzal, hogy az *f* függvény Taylor-sorba fejtése során kapott interakciós tag eltér nullától, valamint azzal is, hogy a konkrét példánkban a Biewen-féle képlet szerint az interakciós hatás nem nulla, hanem –2,2 százalékpont.

A Biewen-féle képlet szerint a *v* megváltozásának *ceteris paribus* hatása közel +1 százalékpontos, a *p* megváltozásának pedig önmagában +8,3 százalékpontos hatása van a homogám párok arányára. A szekvenciális képletek alkalmazása azonban más felbontást eredményezne. Az (1) képlet szerint a *v* megváltozásának közel –1,2 százalékpontos, míg *p* megváltozásának +8,3 százalékpontos a *ceteris paribus* hatása. A (2) képlet ugyanezeket a hatásokat rendre +1 százalékpontnak és +6,1 százalékpontnak méri.

Az itt bemutatott két példából tehát látható, hogy van olyan probléma, amelynek a megoldásához nem kell különösebben törődnünk azzal, hogy milyen dekompozíciós képletet alkalmazunk. De olyan is akad, amihez gondosan kell képletet választanunk.

2. ábra
A homogámia gyakorisága nemlineárisan függ a vágyaktól és lehetőségektől (második numerikus példa)



Megjegyzés: az első példában a korábban született férjek és feleségek megfigyelt végzettség szerinti együttes eloszlása $\begin{bmatrix} 925 & 125 \\ 75 & 375 \end{bmatrix}$, míg a később születetteknél ugyanez $\begin{bmatrix} 150 & 45 \\ 5 & 600 \end{bmatrix}$. Ezért a homogám párok aránya az előbbi generációban $f(p_0, v_0) = (925 + 375)/1500 = 86{,}\dot{6}\%$, míg a később születettek generációjában $f(p_1, v_1) = (150 + 600)/800 = 93{,}75\%$. Az ábra az összes olyan lehetséges generáció homogám párarányát mutatja, amelynél a férjek és feleségek végzettség szerinti együttes eloszlásának Liu–Lu-értéke a két megfigyelt generációra jellemző Liu–Lu-érték konvex kombinációja. Valamint a párok száma, akárcsak a végzettség szerinti eloszlások is, konvex kombinációi a két megfigyelt generációra jellemzőnek. Látható, hogy a homogám párok aránya nemlineárisan függ a Liu–Lu-értéktől és a lehetőségeket meghatározó eloszlásoktól.

A későbbiekben vizsgált empirikus alkalmazások egyike rámutat majd arra, hogy az utóbbi típusú problémák csoportjába tartozó példát nemcsak mi tudunk konstruálni, hanem a társadalom is. Előtte azonban lássuk, hogy milyen elméleti érvek szólnak a Biewen-féle dekompozíciós képlet mellett!

Elméleti megfontolások a dekompozíciós képlet megválasztásához

Biewen [2014] két érvet ismertet, amelyek a (3) képlet szerinti dekompozíciós módszer mellett szólnak. Az egyik érv, hogy a (3) képlettel az interakciós tagot számszerűsíteni tudjuk.[13] Így *az interakciós tag nem torzítja a faktorok* ceteris paribus *hatá-*

[13] *Biewen* [2014] az interakciós képlet alkalmazását a német háztartások jövedelemegyenlőtlenségének elemzésével illusztrálja. Tanulmánya *DiNardo és szerzőtársai* [1996]-tal példázza, hogy a dekompozíciót alkalmazó empirikus munkák nem mindig különítik el az interakciós tagokat.

sait mérő tagokat. Ezzel szemben az (1) és (2) szekvenciális dekompozíciós képletek önkényesen hozzácsapják az interakciós tagot az egyik vagy a másik *ceteris paribus* hatást megjeleníteni hivatott taghoz.

Ha az interakciós tag nulla vagy elhanyagolható nagyságú, akkor a Biewen-féle képlet semmivel sem alkalmasabb a dekomponálásra, mint a szekvenciális képletek, hiszen szinte ugyanazokat a komponenspárokat eredményezi a három képlet. Ezt az esetet szemléltette az előbbiekben bemutatott első numerikus példa. Ha azonban az interakciós tag közgazdaságilag szignifikáns nagyságú, akkor a Biewen-féle képlet és a szekvenciális képletek akár egészen eltérő komponenspárokat is eredményezhetnek. Ezt az esetet szemléltette az előzőkben bemutatott második numerikus példa. Tanulmányunk empirikus része mindkét esetre ad további példákat.

A másik érv a Biewen-féle képlet mellett, hogy az azzal kapott komponensek nem függnek a tényezők megváltozásának feltételezett sorrendjétől, azaz a Biewen-féle dekompozíció *útfüggetlen*.[14] Az útfüggetlenség különösen akkor értékes tulajdonság, amikor semmilyen ismeretünk nincs a változások sorrendjéről, és csak tesztelhetetlen feltevéssel élhetünk róla. Ilyen feltevéssel kell élnünk, amikor az (1) vagy a (2) szekvenciális képletek egyikét választjuk.

Összefoglalva, a Biewen-féle képlet akkor jelent vonzó alternatívát a szekvenciális dekompozíciós képlettel szemben, ha *1.* meg tudjuk konstruálni a hozzá szükséges összes tényellentétes esetet, *2.* az interakciós tag nem elhanyagolható méretű, és *3.* nem ismerjük a tényezők megváltozásának sorrendjét.

A dekompozíciós képletek empirikus alkalmazása

A következőkben a végzettség szerint homogám párok részarányának változását fogjuk felbontani. Az elemzést azon fiatal párok körében végezzük, ahol a férfi partner kora 30 és 34 év közötti. Homogámnak akkor tekintünk egy párt, ha a férfi és a női partner azonos iskolai végzettségű. Három iskolázottsági szintet különböztetünk meg: *1.* az érettségivel nem rendelkezők, *2.* az érettségizettek és *3.* a felsőfokú diplomával rendelkezők szintjét.

Feltételezzük, hogy a végzettség szerint homogám párok részarányát csak a következő tényezők határozzák meg: *1.* a fiatal felnőttek házassági preferenciája, amelyet csak egy ordinális skálán tudunk közvetlenül mérni, *2.* a házasságpiaci felhozatal, azaz

[14] Az útfüggetlenséget *Shorrocks* [2013], valamint *Chantreuil–Trannoy* [2013] is a dekompozíciók elengedhetetlen tulajdonságának tartja. Ők egy olyan dekompozíciót javasolnak, amelyhez az összes, a tényezők különböző permutációjával kapható szekvenciális dekompozíciót el kell készíteni, majd komponensenként átlagolni kell őket. Például a kéttényezős esetben a bejárható utak száma kettő. Az ezeknek az útbejárásoknak megfelelő dekompozíciókat az (1) és a (2) képletek adják. A Δp-hez és Δv-hez tartozó átlagolt komponensek pedig rendre: $[f(p_1, v_0)-f(p_0, v_0)]/2+[f(p_1, v_1)+f(p_0, v_1)]/2$ és $[f(p_1, v_1)-f(p_1, v_0)]/2+[f(p_0, v_1)-f(p_0, v_0)]/2$. Az így kapott dekompozíció – amit Shapley-féle dekompozíciónak hívnak – útfüggetlen ugyan, de nem aggregációkonzisztens, azaz ha az egyik tényezőt további altényezőkre bontjuk – például ha az eloszlások megváltozását nemenként külön szeretnénk vizsgálni –, az sajnos befolyásolja a többi tényező mért hatását.

a különböző végzettségű házasulandó férfiak és nők eloszlása mint közvetlenül megfigyelhető tényező és végül *3.* az eloszlások és a preferenciák együttes hatása.

A dekompozíció során a homogám párok arányának megváltozását a fenti tényezők hatására bontjuk. Egy ilyen alkalmazást az motivál, hogy segítségével megismerhetjük a preferenciatényező változását, és azt egyetlen számmal jellemezhetjük. Ez a mérőszám a preferenciaváltozásnak a homogámpár-arány változásához való hozzájárulása. A mérőszám – kardinális jellegéből adódóan – nemcsak az iskolázottság szerinti homofília változásának irányát mutatja meg, hanem a mértékét is.

Az itt alkalmazott módszer teljesen megegyezik a *Naszódi–Mendonça* [2021]-ben kifejlesztett módszerrel. Ugyanakkor az empirikus alkalmazás eltérő: *Naszódi–Mendonça* [2021] empirikus része az 1980 és 2010 közötti időszakból származó, amerikai népszámlálások házassági és együttélési adatait elemzi. Ebben a tanulmányban nemcsak amerikai adatokat használunk, hanem négy európai ország adatait is. Nevezetesen Franciaországét, Magyarországét, Portugáliáét és Romániáét.

Naszódi–Mendonça [2021] empirikus eredményei a következők: egyrészt a korai *baby boomer* (az adatainkban az 1946 és 1950 között születettek) generációjába tartozók – fiatal felnőttként – jellemzően fontosabb párválasztási szempontnak tartották a végzettséget, mint a késői *baby boomerek* (1956 és 1960 között születettek). Másrészt a késői X-generációba (1976 és 1980 között születettek) tartozók körében jellemzően sokkal erősebb volt a homofília, mint a korai X-generáció tagjai körében (1966 és 1970 között születettek).

Többféleképpen reprodukáljuk *Naszódi–Mendonça* [2021] empirikus elemzését. Először ellenőrizzük a fenti eredmények robusztusságát a vizsgálandó ország megválasztására. Majd megnézzük, hogy ezek az eredmények érzékenyek-e a dekompozíciós képlet megválasztására.

Az érzékenységi vizsgálatokhoz *Naszódi–Mendonça* [2021]-gyel azonos módon választjuk meg az elemzett időszakot, azaz közel ugyanabból a három évtizedből származó népszámlálási adatokat használunk. Ugyancsak azonos az itt alkalmazott részpopuláció-választási kritérium a *Naszódi–Mendonça* [2021]-ben alkalmazott kritériummal: azon fiatal párok adatait elemezzük, ahol a férfi partnerek kora 30 és 34 év közötti a népszámlálások éveiben.[15]

Az Egyesült Államokban pontosan tízévente tartották a népszámlálásokat a vizsgált évtizedekben: 1980-ban, 1990-ben, 2000-ben és 2010-ben. A többi négy országban pedig közel tízévente került sor a cenzusra. 1980-ban a 30–34 év közöttiek a korai *baby boomerek* közül kerültek ki, 1990-ben pedig a késői *baby boomerek* generációjába tartoztak. 2000-ben ugyanezt a korosztályt a korai X-generáció tagjai képviselték, míg 2010-ben a késői X-generáció tagjai.

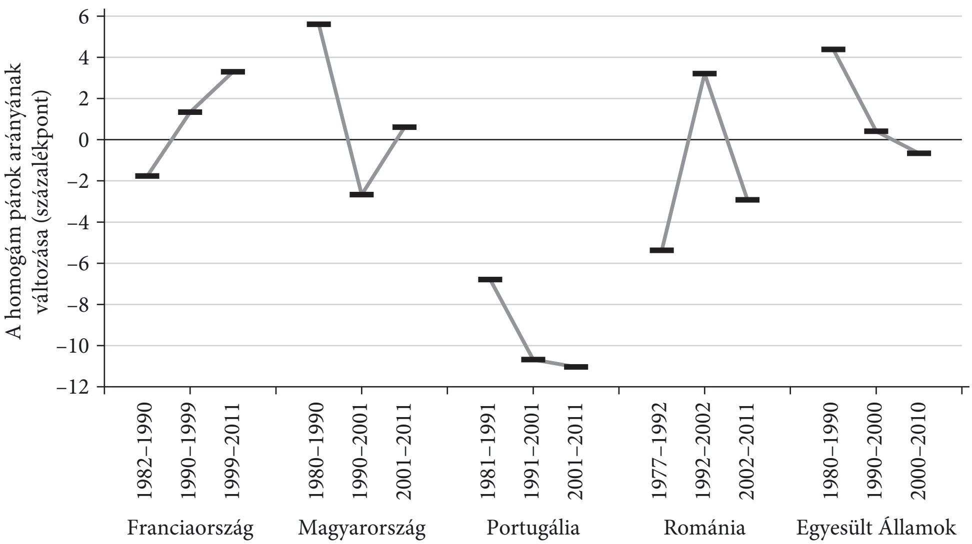
Az adatok forrása az Integrated Public Use Microdata Series (IPUMS). A négy európai ország kiválasztását az motiválta, hogy az IPUMS-ban ezekre az európai országokra érhető el a párok végzettség szerinti együttes eloszlása.

[15] Az érzékenységi vizsgálatok eredményei nem változnak, ha a fiatal párokat eltérő módon definiáljuk, azaz ha azon párokat tekintjük fiatalnak, ahol a férfi partnerek 25 és 39 év közöttiek. Ennek a vizsgálatnak az eredményei elkérhetőek a szerzőtől.

Robusztusak-e az eredmények a vizsgált ország megválasztására?

A *3. ábra* mutatja, hogy hogyan változott a homogám párok aránya a vizsgált korosztályban az egyes országokban és az egyes évtizedekben. Látható, hogy ez a mutató eltérően alakult a vizsgált öt országban. Például Magyarországon és az Egyesült Államokban az 1980-as években a homogám házasságok gyakoribbá váltak a fiatal felnőttek körében, míg a másik három elemzett országra ezzel ellentétes trend volt jellemző. Az ezredforduló után sincs közös trendje a homogámia gyakoriságának: csak Magyarországon és Franciaországban emelkedett a homogám párok aránya, míg a többi országban csökkent.

3. ábra
A végzettség szerint homogám párok részarányváltozása a vizsgált öt országban és három évtizedben (férfi partnerek kora: 30–34 év)

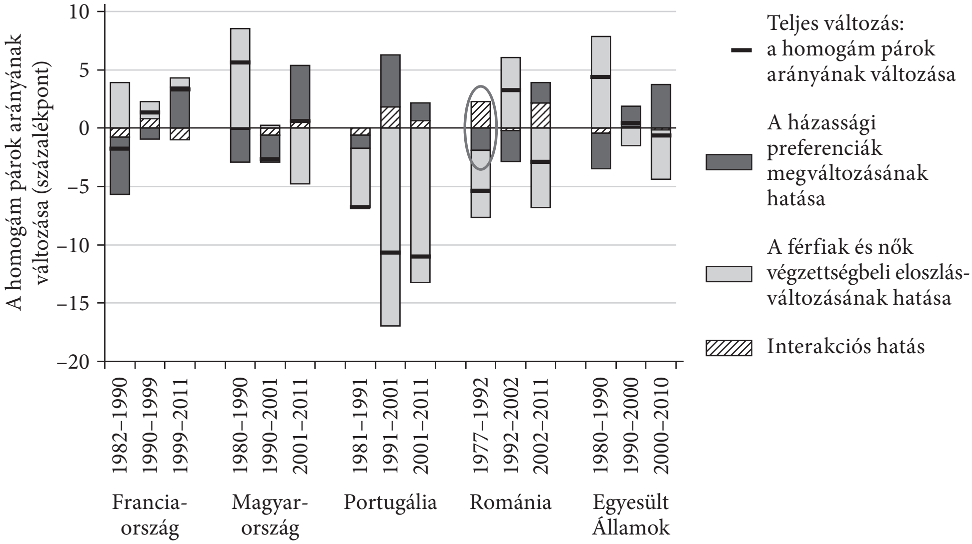


Forrás: IPUMS.

Lássuk, hogyan bonthatók fel a *3. ábrán* mutatott változások! A *4. ábra* mutatja az egyes évtizedekre és az egyes országokra külön-külön számolt dekompozíciók eredményét, amelyeket a (3) képletben szereplő Biewen-féle interakciós képlettel kaptunk.

A különböző oszlopok közül a fekete oszlopszakaszokat érdemes alaposan szemügyre venni, mivel ezek mutatják, hogy a házassági preferenciák generációk közötti megváltozása hogyan járult hozzá a homogám párok arányának megváltozásához. Egy pozitív (negatív) tartományban lévő oszlopszakaszt úgy kell értelmezni, hogy a két összehasonlított generáció közül a később születettek jellemzően jobban (kevésbé) vágytak az azonos végzettségűekkel való házasságra, mint a korábban születettek. De úgy is fogalmazhatnánk, hogy jellemzően elutasítóbbak (kevésbé elutasítóak) voltak az alacsonyabb végzettségűekkel való házasodással szemben.

4. ábra
A végzettség szerint homogám párok részarányváltozásának Biewen-féle interakciós képlettel végzett dekompozíciója a vizsgált öt országban és három évtizedben (férfi partnerek kora: 30–34 év)



Megjegyzés: a dekompozíciók a Biewen által javasolt (3) dekompozíciós képlettel készültek, amelyekhez a tényellentéteseket a *Naszódi–Mendonça* [2021] által kifejlesztett NM-módszerrel konstruáltuk. A bekarikázott komponensek például szolgálnak a dekompozíciós képlet gondos megválasztásának szükségességére: a bekarikázott interakciós tag megváltoztatná a preferenciaváltozás hatásának előjelét és mértékét, ha az utóbbit az előbbivel torzítottan identifikálnánk a (2) szekvenciális dekompozíciós képlettel.

Fontos eredmény, hogy mind az öt vizsgált ország esetében az 1980-as évekhez tartozó fekete oszlopszakasz a negatív tartományban van. Továbbá mind az öt országra igaz, hogy a 2000-es évekhez tartozó fekete oszlopszakasz a pozitív tartományban van, és viszonylag magas. Tehát nemcsak az Egyesült Államokra, hanem a vizsgált európai országokra is igaz, hogy a második világháborút követően volt egy olyan időszak, amikor a végzettség szerinti homofília gyengült a fiatal felnőttek körében, azaz a különböző iskolázottságú társadalmi csoportok közötti egyenlőtlenség csökkent.[16] Ezt az időszakot pedig egy olyan – a pénzügyi válság időszakát is magában foglaló – időszak követte, amelyet éppen az ellentétes irányú trend jellemzett mind az öt vizsgált országban.[17] Tehát a dekompozíciós eredmények robusztusak a vizsgált ország megválasztására.

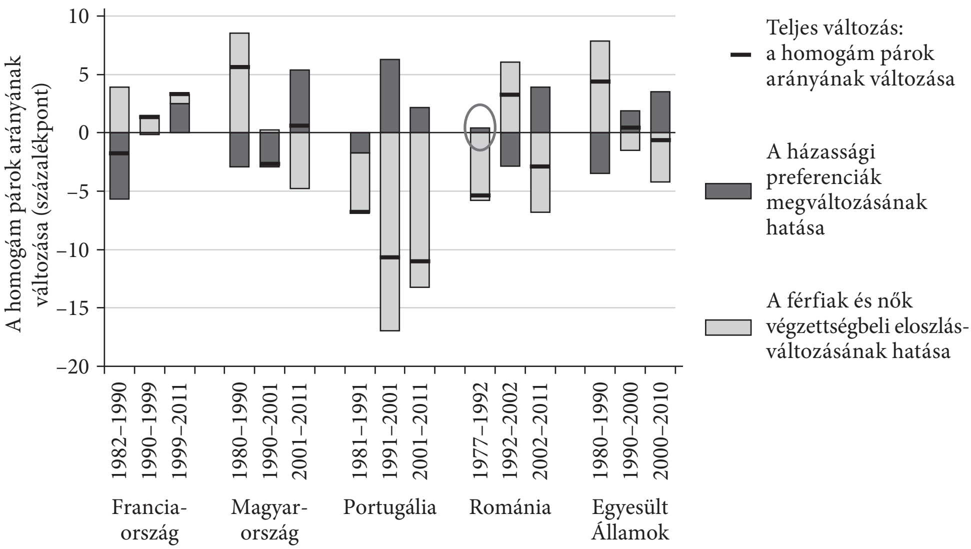
[16] Ennek az eredménynek a jelentősége a gazdaság- és társadalompolitikára vonatkozóan az, hogy a 21. század elejére extrém módon megnövekedett egyenlőtlenségek csökkentéséhez érdemes megfontolni az ezt a célt már sikeresen elérő időszakban alkalmazott szakpolitikák újrahasznosítását.

[17] A két időszak közötti trendforduló eltérő években volt az egyes országokban, ami miatt a *4. ábrán* az 1990-es évekhez tartozó fekete oszlopszakaszok némelyike még a negatív tartományban van, némelyike már a pozitívban.

Robusztusak-e az eredmények a dekompozíciós képlet megválasztására?

Lássuk, hogy vajon a képletválasztásra érzékenyek-e az eredmények! Ezt a *4.* és az *5. ábra* összevetéséből állapíthatjuk meg, ugyanis az *5. ábra* azoknak a szekvenciális képlettel kapott dekompozícióknak az eredményét mutatja, ahol a végzettség szerinti eloszlásról feltettük, hogy előbb változik, mint a preferenciák. A két ábra közötti fontos különbség, hogy a *4. ábrával* ellentétben az *5. ábrán* nem látunk interakciós hatásokat megjelenítő csíkos oszlopokat. Az interakciós hatásokat ugyanis a preferenciaváltozás *ceteris paribus* hatásához csoportosítja az *5. ábra* készítésénél alkalmazott (2) szekvenciális képlet.

5. ábra
A végzettség szerint homogám párok részarányváltozásának a szekvenciális képlettel végzett dekompozíciója a vizsgált öt országban és három évtizedben (férfi partnerek kora: 30–34 év)



Megjegyzés: a dekompozíciók a (2) szekvenciális dekompozíciós képlettel készültek, amelyekhez a tényellentéteseket a *Naszódi–Mendonça* [2021] által kifejlesztett NM-módszerrel konstruáltuk. A bekarikázott komponens szerint a preferenciaváltozás hatásának előjele és mértéke eltér a *4. ábrán* láthatótól.

Érdekes módon az Egyesült Államok esetében viszonylag alacsonyak a csíkos oszlopok a *4. ábrán*. Ez egyrészt azt jelenti, hogy az amerikaiak jellemzően nem nagyon szokták a vágyaikat/társadalmi normáikat a lehetőségeikhez igazítani – legalábbis ez látszik a házassági mintázatok generációk közötti összehasonlításából. Másrészt nincs lényeges különbség az interakciós és a szekvenciális dekompozíciós képlettel kapott eredmények között, ha csak az Egyesült Államokat vizsgáljuk. Az utóbbi megállapítás Portugáliára is igaz.[18]

[18] Megjegyzendő, hogy Portugáliában – az Egyesült Államokkal ellentétben – inkább a *ceteris paribus* hatások jelentős méretéhez képest tűnnek kicsinek az interakciós hatások, önmagukban nem azok.

Van azonban egy olyan ország, amely esetében a *4. ábra* csíkos oszlopai nem tekinthetők elhanyagolható méretűnek. Ez az ország Románia. Ezzel összefüggésben Románia esetében bizonyos eredmények nagyon is érzékenyek arra, hogy melyik képlettel végezzük a dekompozíciót. A robusztusság hiánya a vizsgált három időszak közül az 1980-as évekre vonatkozó eredményeket érinti a leginkább.

A Biewen-féle dekompozíció szerint Romániában ezt az időszakot a homofília jelentős gyengülése jellemezte – akárcsak a másik négy országban. A preferenciák generációk közötti – a vizsgált esetben a korai *baby boomerek* és a késői *boomerek* közötti – megváltozásának hatására a homogám párok arányának közel 2 százalékponttal kellett volna csökkennie a román férfiak és nők végzettség szerinti eloszlásának változatlansága mellett a Biewen-féle képlet szerint.

Ez a preferenciahatás azért nem elhanyagolható nagyságrendű, mert mindeközben a homogám párok aránya 5 százalékpontot csökkent, 74 százalékról 69 százalékra. Tehát a teljes változás több mint harmadát okozta önmagában a preferenciaváltozás. Figyelemre méltó, hogy az interakciós hatás ellentétes előjelű, és a preferenciahatásnál valamivel nagyobb abszolút értékű (lásd a bekarikázott oszlopokat a *4. ábrán*).

A Biewen-féle dekompozícióval szemben a (2) szekvenciális képlet ellentétes előjelű, gyenge hatást tulajdonít a preferenciák megváltozásának (lásd a bekarikázott oszlopot az *5. ábrán*). Ez a képlet 0,4 százalékpontos emelkedést tulajdonít a házassági preferenciák változásának, ami a Biewen-féle képlettel kapott *ceteris paribus* preferenciahatás és az interakciós hatás összegéből adódik.

Miért érdekes, hogy eltérő képletekkel ellentétes előjelű, illetve eltérő nagyságrendű hatásokat mérünk? Azért, mert a szekvenciális képlettel kapott eredmény megkérdőjelezi az interakciós képlettel kapott eredményünk általános jellegét. Nevezetesen azt az eredményt, hogy a késői *boomerek* körében a korai *boomerekhez* képest jelentősen gyengült a homofília a vizsgált társadalmak mindegyikében.

Összefoglalva, Románia példája rámutat arra, hogy bizonyos kvalitatív eredmények érzékenyek lehetnek a dekompozíciós képlet megválasztására.[19]

Összegzés

Ebben a tanulmányban összehasonlítottuk a Biewen által javasolt interakciós dekompozíciós képletet a szekvenciális dekompozíciós képlettel. Először *Biewen* [2014] nyomán ismertettünk néhány elméleti érvet, amelyek az előbbi képlet mellett szólnak. Így például azt, hogy a Biewen-féle képlet segítségével – szemben a szekvenciális dekompozíciós képlettel – a tényezők interakciós hatását is számszerűsíthetjük, és el tudjuk különíteni a tényezők *ceteris paribus* hatásától.

Majd a kétfajta képletet egy új empirikus alkalmazás segítségével is összevetettük. Az utóbbi gyakorlat tanulságai a következők. Először is, egy, a Biewen-féle képlettel

[19] Ha az (1) szekvenciális képletet alkalmazzuk, akkor további példákat is találunk a robusztusság hiányára. Például Franciaország esetében a változó iskolázottsági szint hatásának előjelét befolyásolja, hogy ennek tulajdonítjuk-e az 1999 és 2011 közötti interakciós hatást, vagy sem. Ugyanezt találjuk Magyarország esetében az 1990 és 2001 közötti időszakra vonatkozóan.

kapott kvalitatív eredmény robusztus arra nézve, hogy melyik ország adatain végezzük a dekompozíciót. Másodszor, ugyanez az eredmény érzékeny a képletválasztásra, ami abból adódik, hogy az interakciós tag az egyik elemzett ország esetében nem elhanyagolható méretű.

Ha evidensnek gondoljuk, hogy az elemzett társadalmakban az iskolázottság szerinti homofília erősségének azonos irányba kellett volna változnia az elemzett egyes évtizedekben, akkor az országválasztásra vonatkozó robusztusságot az egyes dekompozíciós képletekkel szembeni empirikus kritériumnak is tekinthetjük. Ebben az esetben nem csupán teoretikus megfontolások alapján javasolható a Biewen-féle interakciós dekompozíciós képlet alkalmazása, hanem egy empirikus kritérium alapján is.

Biewen [2014] a német háztartások jövedelemegyenlőtlenségének empirikus elemzése alapján javasolja dekompozíciós képlete alkalmazását. Ennek során a jelen tanulmányban ismertetettel hasonló eredményre jut: az általa vizsgált interakciós hatás (például az adórendszer és a háztartások bizonyos szocioökonómiai tulajdonságait honoráló munkaerőpiaci hozadék változásának együttes hatása), éppúgy, ahogy az általunk vizsgált interakciós hatás (például az iskolázottság és az aggregált házassági preferenciák változásának együttes hatása), közgazdaságilag szignifikáns. Tehát *Biewen* [2014] és a jelen tanulmány empirikus eredményei együttesen azt támasztják alá, hogy a dekompozíciós képlet választására vonatkozó érzékenység nem korlátozódik sem a béregyenlőtlenség-elemzésre, sem az asszortatív házasságok elemzésére. Ezért a dekompozíciós képlet körültekintő megválasztása fontos lehet további alkalmazások esetében is.

Hivatkozások

Becker, G. S. [1973]: A theory of marriage: Part I. Journal of Political Economy, Vol. 81. No. 4. 813–846. o. https://doi.org/10.1086/260084.

Biewen, M. [2012]: Additive decompositions with interaction effects. IZA Discussion Papers, No. 6730. https://doi.org/10.2139/ssrn.2114910.

Biewen, M. [2014]: A general decomposition formula with interaction effects. Applied Economics Letters, Vol. 21. No. 9. 636–642. o. https://doi.org/10.1080/13504851.2013.879280.

Blinder, A. S. [1973]: Wage discrimination: Reduced form and structural estimates. The Journal of Human Resources, Vol. 8. No. 4. 436–455. o. https://doi.org/10.2307/144855.

Chantreuil, F.–Trannoy, A. [2013]: Inequality decomposition values: the trade-off between marginality and consistency. Journal of Economic Inequality, Vol. 11. No. 1. 83–98. o. https://doi.org/10.1007/s10888-011-9207-y.

DiNardo, J.–Fortin, N. M.–Lemieux, T. [1996]: Labor market institutions and the distribution of wages, 1973–1992: a semiparametric approach. Econometrica, Vol. 64. No. 5. 1001–1044. o. https://doi.org/10.2307/2171954.

Dupuy, A.–Weber, S. [2022]: Marriage market counterfactuals using matching models. Economica, Vol. 89. 23–49. o. https://doi.org/10.1111/ecca.12386.

Foster, L.–Haltiwanger, J.–Syverson, C. [2008]: Reallocation, firm turnover, and efficiency: Selection on productivity or profitability? American Economic Review, Vol. 98. No. 1. 394–425. o. https://doi.org/10.1257/aer.98.1.394.
Kalmijn, M. [1998]: Intermarriage and homogamy: causes, patterns, trends. Annual Review of Sociology, Vol. 24. No. 1. 395–421. o. https://doi.org/10.1146/annurev.soc.24.1.395.
Liu, H.–Lu, J. [2006]: Measuring the degree of assortative mating. Economics Letters, Vol. 92. No. 3. 317–322. o. https://doi.org/10.1016/j.econlet.2006.03.010.
Mihályi Péter–Szelényi Iván [2019]: Rent-Seekers, Profits, Wages and Inequality. The Top 20%. Springer International Publishing, https://doi.org/10.1007/978-3-030-03846-5.
Naszódi Anna [2021*a*]: A note on what surveys say about the applicability of the Iterative Proportional Fitting algorithm. Unpublished manuscript, under review.
Naszódi Anna [2021*b*]: A new method for identifying changing marital preferences for race and education level. https://doi.org/10.48550/arXiv.2103.06991.
Naszódi Anna–Mendonça, F. [2021]: A new method for identifying the role of marital preferences at shaping marriage patterns. Journal of Demographic Economics, First View, 1–27. o. https://doi.org/10.1017/dem.2021.1.
Naszódi Anna–Mendonça, F. [2022]: Changing educational homogamy: Shifting preferences or evolving educational distribution? Journal of Demographic Economics, 1–29. o. http://dx.doi.org/10.1017/dem.2022.21.
Oaxaca, R. [1973]: Male-female wage differentials in urban labor markets. International Economic Review, Vol. 14. No. 3. 693–709. o. https://doi.org/10.2307/2525981.
Piketty, T. [2015]: A tőke a 21. században. Kossuth, Budapest.
Rosenfeld, M. J. [2008]: Racial, educational, and religious endogamy in the United States: A comparative historical perspective. Social Forces, Vol. 87. No. 1. 1–32. o. https://doi.org/10.1353/sof.0.0077.
Shorrocks, A. F. [2013]: Decomposition Procedures for Distributional Analysis: A Unified Framework Based on the Shapley Value. The Journal of Economic Inequality, Vol. 11. No. 1. 99–126. o. https://doi.org/10.1007/s10888-011-9214-z.
Stephan, F. F.–Deming, W. E. [1940]: On a least squares adjustment of a sampled frequency table when the expected marginal totals are known. The Annals of Mathematical Statistics, Vol. 11. No. 4. 427–444. o. https://doi.org/10.1214/aoms/1177731829.

Függelék

A) A Liu–Lu-mutató

A *Függeléknek* ebben a részében először ismertetjük a *Liu–Lu* [2006] által kifejlesztett skalárértékű, ordinális mutatót. Ez a mutató akkor alkalmazható, amikor az asszortatív tulajdonságot egy egydimenziós, bináris változó reprezentálja. Ezután a Liu–Lu-mutató *Naszódi–Mendonça* [2021] által javasolt általánosítását ismertetjük arra az esetre, amikor az asszortatív tulajdonság egy egydimenziós, trinomiális, rendezett értékű változó (például alacsony, közepes, magas végzettség/jövedelem/sítudás).

Liu–Lu [2006] felteszi, hogy a gazdaságban azonos számú házasulandó férfi és nő van. A számukat *N*-nel jelöljük. Továbbá felteszi, hogy minden házasulandó heteroszexuális, monogám, és meg fog házasodni.[20] Ezért N azonos a megkötendő házasságok számával is.

A férfiak és nők vagy magas, vagy alacsony végzettségűek. Így a Liu–Lu-modellben négyféle pár van. Ennek megfelelően a párok végzettség szerinti eloszlását egy 2×2-es kontingenciatábla írja le:

$$K = \begin{bmatrix} N_{L,L} & N_{L,H} \\ N_{H,L} & N_{H,H} \end{bmatrix}. \tag{F1}$$

A jelölés értelmezése a következő: az $N_{L,L}$ azon párok száma, ahol a férj és a feleség is alacsony végzettségű. Az $N_{L,H}$ azon párok száma, ahol a férj alacsony végzettségű, míg a feleség magas végzettségű. Az $N_{H,L}$ azon párok száma, ahol a férj magas végzettségű, a feleség alacsony végzettségű. Végül az $N_{H,H}$ jelöli azon párok számát, ahol mindkét fél magas végzettségű.

A férfiak és nők végzettség szerinti eloszlásairól, azaz a tábla peremeloszlásairól feltesszük, hogy nem degeneráltak. Tehát legalább egy férfi és legalább egy nő van mindkét iskolázottsági kategóriában ($N_{L,.} \geq 1$, $N_{H,.} \geq 1$, $N_{.,L} \geq 1$, $N_{.,H} \geq 1$).

A Liu–Lu-mutatót a következő képlet adja meg:

$$\mathrm{LL}(K) = \left\{ \begin{array}{ll} \dfrac{N_{H,H} - Q^-}{\min\left(N_{H,.},\, N_{.,H}\right) - Q^-}, & \text{ha} \quad N_{H,H} \geq Q \\ \dfrac{N_{H,H} - Q^+}{Q^+ - \max\left(0,\, N_{H,.} - N_{.,L}\right)}, & \text{ha} \quad N_{H,H} < Q \end{array} \right\}, \tag{F2}$$

ahol N (párok száma), $N_{H,.}$ (magas végzettségű férjek száma), $N_{.,H}$ (magas végzettségű feleségek száma) értelmezése ugyanaz, mint eddig. Míg $Q = N_{H,.}N_{.,H}/N$ a (H, H)-típusú párok számának várható értéke azon feltevés mellett, hogy a párosítás véletlenszerű. Továbbá Q^- a Q egészrészét jelöli. Tehát Q^- az a legnagyobb egész szám, amelyik kisebb Q-nál, vagy egyenlő vele. Végül Q^+az a legkisebb egész szám, amelyik nagyobb Q-nál, vagy egyenlő vele.

Mivel a valóságban a magas végzettségűek nagyobb arányban választanak maguknak magas iskolázottságú párt, mint ha véletlenszerű lenne a párosítás, ezért az empirikusan releváns kontingenciatáblákra teljesül, hogy $N_{H,H} \geq Q$. Ha ezen táblákra leszűkítjük a vizsgálatunkat, akkor a Liu–Lu-mutató képlete a következőképpen egyszerűsödik:

$$\mathrm{LL}^{\mathrm{sim}}(K) = \frac{N_{H,H} - Q^-}{\min\left(N_{H,.},\, N_{.,H}\right) - Q^-}. \tag{F3}$$

[20] A valóságban vannak egyedülállók, sőt a 2000-es években az arányuk mind az öt vizsgált országban megugrott. Ahogy a munkaerőpiac bizonyos kérdéseinek vizsgálatánál félrevezető lehet, ha a munkanélkülieket kihagyjuk az elemzésből, éppúgy félrevezető lehet, ha a házasságok elemzésénél eltekintünk az egyedülállóság lehetőségétől. Ez motiválta, hogy *Naszódi–Mendonça* [2022] az egyedülállókat is bevette az elemzésébe. A szerzőpáros a gazdagabb modellt gazdagabb adatokon becsülte: nemcsak népszámlálási adatokat használt, hanem egy magyar *online dating* szolgáltatótól származó adatokat is. A cikk megmutatta, hogy a jelen tanulmányban is alkalmazott módszer eredményeitől nem kapott lényegesen különböző eredményeket az egyedülállók figyelembevételével.

Az egyszerűsített Liu–Lu-képlettel definiált mutató tulajdonságai a következők. Egyrészt ennek a mutatónak az értékkészlete a [0, 1] intervallum. A mutató értéke akkor 0, amikor – a *K* tábla adott peremeloszlásai mellett – a párosítás eredményeként kapott *K* kontingenciatábla a véletlenszerű párosítás várható értékével egyezik meg. A maximális értékét, azaz 1-et pedig akkor veszi fel a mutató, amikor a *K* táblában a homogám párok aránya a lehető legnagyobb az adott peremeloszlások mellett. Utóbbi annak az esetnek felel meg, amikor a magas végzettségűek csak akkor választhatnak maguknak alacsony végzettségű párt, ha már nem maradt magas végzettségű ellenkező nemű házasulandó.

A mutató 0 és 1 közötti értéke úgy értelmezhető, hogy a házasulandók jellemzően nem annyira közömbösek a leendő partnerük végzettségével szemben, hogy ezen dimenzió szerint véletlenszerű lenne közöttük a párválasztás, de nem is annyira jellemző rájuk a homofília, hogy a homogám párok arányát maximalizálnák.

Az (egyszerűsített) Liu–Lu-mutató magasabb értékét úgy értelmezhetjük, hogy az a homogámia iránt jellemzően erősebb vágyat jelent a házasodók körében (avagy a heterogámia kevésbé elfogadott a körükben). Ezen a ponton meg kell jegyeznünk, hogy ezt az interpretációt *Naszódi–Mendonça* [2021] javasolja, és nem a Liu–Lu szerzőpárostól származik. *Liu–Lu* [2006] azonban hangsúlyozza mutatójuk következő tulajdonságát: a mutató független a *K* tábla peremeloszlásaitól. Ez a tulajdonság azért lényeges, mert nyilvánvaló, hogy egy kéttényezős modellben – ahol a piaci egyensúly csak a preferenciáktól és a piaci felhozataltól (azaz a peremeloszlásoktól) függ – egy olyan mutató, amely független a peremeloszlás tényezőtől, csak a másik tényezőt reprezentálhatja. Tehát a *Liu–Lu* [2006] által diszkutált tulajdonságból egyenesen következik a mutatójuk preferenciamutatóként való interpretálása.[21]

Nagy előnye a Liu–Lu-mutatónak – az irodalomban elterjedt asszortativitásmutatókkal szemben –, hogy eltérő peremeloszlású táblákat is plauzibilis módon rangsorol a homofília erőssége alapján. A plauzibilitást *Naszódi–Mendonça* [2021] vizsgálja. Először megállapítja, hogy van egy olyan évtized a mintában, amelyre a Liu–Lu-mutató az irodalomban elterjedt alternatív asszortativitásmutatókkal ellentétes irányú trendet azonosít. Majd a Pew Research Center kérdőíves felméréséből származó adatok segítségével végzi a modellszelekciót. A modellszelekció eredményeként a Liu–Lu-mutató alkalmazását validálják. A validáló adatok a különböző generációkba tartozó megkérdezettek házassági preferenciáiról szólnak. A felmérés szerint a kérdéses trend határozottan a Liu–Lu-mutató trendjével egyezik meg, nem pedig az alternatív mutatók trendjével.

Ugyanez a validáció azt a – sok kutató számára meglepő – eredményt is hozta, hogy az irodalomban elterjedt iteratív arányos illesztés *(iterative proportional fitting, IPF)* algoritmus nem alkalmas tényellentétes konstruálására.[22]

[21] Ezen a ponton érdemes megjegyezni, hogy a preferenciatényezőként való interpretációt azért nem érezhetjük természetesnek, mert hozzá vagyunk szokva, hogy a valóságban a piaci egyensúlyt nem csak két tényező határozza meg. A valóságban a párválasztásunkat az is befolyásolhatja, hogy kikkel találkozunk és kikkel nem, valamint mennyi cipőtalpköltséget szánunk a párkeresésre. A Liu–Lu-modell azonban eltekint ezektől a tényezőktől, mivel tökéletesen informált szereplőket feltételez.

[22] Az IPF algoritmust szokták RAS algoritmusnak is nevezni (lásd https://en.wikipedia.org/wiki/Iterative_proportional_fitting). Az IPF-módszer kifejlesztői, a *Stephan–Deming* [1940] tanulmány

Lássuk hogyan általánosította *Naszódi–Mendonça* [2021] a Liu–Lu-mutatót arra az esetre, amikor az iskolai végzettség három lehetséges értéket vehet fel: alacsony, közepes és magas (*L, M, H*) értékeket! Az ennek megfelelő kontingenciatábla egy 3×3-as tábla:

$$\begin{bmatrix} N_{L,L} & N_{L,M} & N_{L,H} \\ N_{M,L} & N_{M,M} & N_{M,H} \\ N_{H,L} & N_{H,M} & N_{H,H} \end{bmatrix}, \qquad (F4)$$

ahol $N_{m,w}$ azon párok száma, ahol a férfi partner végzettsége $m \in \{L, M, H\}$, míg a feleségek végzettsége $w \in \{L, M, H\}$. Az általánosított Liu–Lu-mutató nem skalár, hanem mátrixértékű. Az alábbiakban röviden összefoglaljuk az általánosított, mátrixértékű mutató kiszámításának lépéseit.

Először az iskolázottsági változót kell dichotomizálni, hogy visszavezethessük a problémát a már *Liu–Lu* [2006] által megoldott 2×2-es esetre, melyben $m, w \in \{L, H\}$. Ezt négy különböző módon tehetjük meg, attól függően, hogy az *M*-típusú férfiakat és az *M*-típusú nőket az (*F*4) képletbeli táblában alacsony vagy magas iskolai végzettségűnek minősítjük-e át a dichotóm világban. A kontingenciatábla négy eltérő dichotomizációja a táblák következő halmazát adja: $\{K^{H,H}, K^{H,L}, K^{L,H}, K^{L,L}\}$, ahol a $K^{m,w}$ jelölésű 2×2-es kontingenciatáblát úgy kapjuk, hogy az *M*-típusú férjeket az *m*-típusba, míg az *M*-típusú feleségeket a *w*-típusba soroljuk át ($m, w \in \{L, H\}$).

A második lépés során az eredeti, 2×2-es kontingenciatáblára kifejlesztett Liu–Lu-mutatót kell kiszámítani az (*F*2) képletet alkalmazva az egyes dichotomizációkkal kapott kontingenciatáblákra. Ennek eredményeként megkapjuk az alábbi, úgynevezett Liu–Lu-mátrixot:

$$\mathrm{LL}^{\mathrm{gen}} = \begin{bmatrix} \mathrm{LL}\left(K^{H,H}\right) & \mathrm{LL}\left(K^{H,L}\right) \\ \mathrm{LL}\left(K^{L,H}\right) & \mathrm{LL}\left(K^{L,L}\right) \end{bmatrix}. \qquad (F5)$$

Tehát így általánosítjuk az (*F*2) képlettel adott, *Liu–Lu* [2006] által javasolt mutatót arra az esetre, amikor az asszortatív tulajdonság – példánkban az iskolai végzettség – egy egydimenziós, trinomiális, rendezett értékű változó.

B) Az $f\left(p_{t_p}, v_{t_v}\right)$ függvény definíciója az NM-módszer alapján

A *Függeléknek* ebben a részében definiáljuk az $f\left(p_{t_p}, v_{t_v}\right)$ függvényt, amely a t_p időszakból származó preferenciák és a t_v időszakból származó peremeloszlások mellett megadja a homogám párok arányát az NM-módszer alapján.

szerzői tudtak a módszerük alkalmazhatóságának ilyen jellegű korlátozottságáról. Ugyanis ők – azontúl, hogy más alkalmazásra fejlesztették ki a módszerüket – hangsúlyozták is, hogy az IPF nem alkalmas predikcióra. Így tényellentétes konstruálására sem. Sajnos mégis elterjedt az IPF erre a célra való használata a nemzetközi irodalomban (lásd *Naszódi–Mendonça* [2021]).

Ha $t_p = t_v = t$, akkor a p_t preferenciák és v_t peremeloszlások mellett kialakuló párosítás eredményét meg tudjuk figyelni a t-edik időszakban. Jelöljük K_t-vel a t-edik időszakban megfigyelhető kontingenciatáblát, amely a férjek és a feleségek végzettség szerinti együttes eloszlását adja meg, valamint N_t-vel a párok számát ebben az időszakban. Ezekkel a jelölésekkel $f(p_t, v_t) = \text{sum}\,[\text{diag}\,(K_t)]/N_t$.

Ha azonban $t_p \neq t_v$, akkor az f függvény két függvény kompozíciójaként áll elő. Azaz $f\left(p_{t_p}, v_{t_v}\right) = g \circ h\left(p_{t_p}, v_{t_v}\right)$, ahol a h függvény megadja, hogy a p_{t_p} preferenciák és v_{t_v} peremeloszlások mellett milyen lenne a férjek és feleségek végzettségének együttes eloszlását leíró kontingenciatábla. Ezt a táblát jelöljük K^*-gal. A K^*-ot tényellentétes kontingenciatáblának hívjuk a továbbiakban. A h függvény definiálása nem triviális feladat. Viszont amint a K^* táblát megkonstruáltuk a h függvénnyel, az f függvény értékét már triviális kiszámolni a $g(K^*) = \text{sum}\,[\text{diag}\,(K^*)]/N^*$ függvénnyel, ahol N^*a K^* táblában lévő párok számát jelöli.

A továbbiakban röviden ismertetjük a K^* tényellentétes kontingenciatábla konstruálásának *Naszódi–Mendonça* [2021]-ben javasolt módját, azaz az NM-módszert.[23] Ez a tényellentétes konstruálás azon a feltevésen alapszik, hogy az aggregált preferenciák akkor azonosak, ha a(z általánosított) Liu–Lu-mutatók megegyeznek.[24] Először azt az esetet tárgyaljuk, ahol az asszortatív tulajdonság bináris, majd a trinomiális esetet.

Ha az asszortatív tulajdonság egy bináris változó, akkor a férjek és feleségek végzettség szerinti együttes eloszlását leíró kontingenciatábla 2 × 2-es. Jelöljük K_{t_p}-vel a t_p időszakban megfigyelt kontingenciatáblát:

$$K_{t_p} = \begin{bmatrix} N_{L,L} & N_{L,H} \\ N_{H,L} & N_{H,H} \end{bmatrix}. \qquad (F6)$$

Ezt a táblát az irodalomban használt elnevezést követve magtáblának *(seed table)* hívjuk. Továbbá jelöljük K_{t_v}-vel a t_v időszakban megfigyelt kontingenciatáblát, amelynek a sorösszegét adó 2 × 1-es dimenziójú vektort R_{t_v}-vel jelöljük, míg az oszlopösszegét adó 1 × 2 dimenziójú vektort C_{t_v}-vel. Ezeket a vektorokat célperemeloszlás- *(target marginals)* vektoroknak hívjuk.

A célunk a 2 × 2-es tényellentétes kontingenciatábla, azaz

$$K^* = \begin{bmatrix} N^*_{L,L} & N^*_{L,H} \\ N^*_{H,L} & N^*_{H,H} \end{bmatrix} \qquad (F7)$$

elemeinek a meghatározása az alábbi megkötések mellett:

$$R_{t_v} = \begin{bmatrix} N^*_{L,\cdot} \\ N^*_{H,\cdot} \end{bmatrix} \quad \text{és} \quad C_{t_v} = \begin{bmatrix} N^*_{\cdot,L} & N^*_{\cdot,H} \end{bmatrix}, \quad \text{valamint} \qquad (F8)$$

[23] Az NM-módszer Excel, Visual Basic és R-ben implementált kódját lásd: https://data.mendeley.com/datasets/x2ry7bcm95/2.

[24] Az asszortatív házasságok irodalmában egy elterjedt – de semmiképpen sem követendő – gyakorlat, hogy a tényellentétes táblát az IPF-eljárással konstruálják. Ezen eljáráshoz kapcsolódó mögöttes feltevés az, hogy az aggregált házassági preferenciákat nem a Liu-Lu-mutatóval, hanem az esélyhányadossal *(odds ratio)* lehet reprezentálni.

$$\mathrm{LL}\left(K_{t_p}\right) = \mathrm{LL}\left(K^*\right). \qquad (F9)$$

Az ($F8$) egyenlet teljesülése esetén a K_{t_v} táblával jellemzett egyensúly és a K^* táblával jellemzett egyensúly azonos végzettségbeli eloszlások mellett alakul ki. Az ($F9$) egyenlet teljesülése esetén a K_{t_p} táblával jellemzett egyensúly és a K^* táblával jellemzett egyensúly azonos preferenciák mellett alakul ki, mivel a táblák Liu–Lu-mutatói megegyeznek.

A probléma megoldásához az NM-módszert használjuk. Ehhez feltesszük, hogy $N_{H,H} \geq Q$. Az utóbbi feltevés lehetővé teszi, hogy a Liu–Lu-mutató ($F3$) egyenletben szereplő egyszerűsített változatát használjuk. Ha behelyettesítjük az ($F9$) egyenletbe az ($F3$) egyenletet, akkor a következő összefüggést kapjuk:

$$\frac{N_{H,H} - \mathrm{int}\left(N_{H,\cdot}N_{\cdot,H}/N\right)}{\min\left(N_{H,\cdot},\, N_{\cdot,H}\right) - \mathrm{int}\left(N_{H,\cdot}N_{\cdot,H}/N\right)} = \frac{N^*_{H,H} - \mathrm{int}\left(N^*_{H,\cdot}N^*_{\cdot,H}/N^*\right)}{\min\left(N^*_{H,\cdot},\, N^*_{\cdot,H}\right) - \mathrm{int}\left(N^*_{H,\cdot}N^*_{\cdot,H}/N^*\right)}. \qquad (F10)$$

A megoldást az ($F10$) egyenlet átrendezésével kapjuk:

$$N^*_{H,H} = \frac{\left[N_{H,H} - \mathrm{int}\left(N_{H,\cdot}N_{\cdot,H}/N\right)\right]\left[\min\left(N^*_{H,\cdot},\, N^*_{\cdot,H}\right) - \mathrm{int}\left(N^*_{H,\cdot}N^*_{\cdot,H}/N^*\right)\right]}{\min\left(N_{H,\cdot},\, N_{\cdot,H}\right) - \mathrm{int}\left(N_{H,\cdot}N_{\cdot,H}/N\right)} +$$

$$+ \mathrm{int}\left(N^*_{H,\cdot}N^*_{\cdot,H}/N^*\right). \qquad (F11)$$

Az ($F11$) egyenlet jobb oldala az ($F7$) és az ($F8$) képletekből ismert értékű változók függvényében fejezi ki $N^*_{H,H}$-t. A K^* tábla többi három elemét pedig triviális feladat meghatározni az $N^*_{H,H}$ és a célperemeloszlások ismeretében.

Végül nézzük azt a problémát, ahol az asszortatív tulajdonság az $\{L, M, H\}$ rendezett, háromelemű halmazból veszi fel az értékét!

Bár a jelölésen nem változtatunk a korábbiakhoz képest (továbbra is K_{t_p} jelöli a magtáblát, a K^* a meghatározandó tényellentétes táblát, R_{t_v} és C_{t_v} jelölik a célperemeloszlás-vektorokat, és $f = g \circ h$), de a táblák, vektorok és függvények mérete, dimenziói mások, mint eddig voltak. Ebben az esetben a K_{t_p} magtábla és a K^* meghatározandó tényellentétes tábla 3×3-as. A h függvény egy $[0, 1]^{2\times 2} \times \mathbb{N}^5 \to \mathbb{N}^{3\times 3}$ függvény, mivel az a preferenciákat reprezentáló 2×2-es Liu–Lu-mátrixból és az öt független elemet tartalmazó célperemeloszlás-vektorokból egy 3×3-as tényellentétes táblát képez. A g függvény $\mathbb{N}^{3\times 3} \to [0, 1]$, mivel az a 3×3-as tényellentétes táblából egy $[0, 1]$ intervallumba eső skalárt képez. Az f függvény pedig $[0, 1]^{2\times 2} \times \mathbb{N}^5 \to [0, 1]$.

A K^* sorösszegének meg kell egyeznie a 3×1 dimenziójú R_{t_v} vektorral, míg az oszlopösszegének meg kell egyeznie az 1×3-as dimenziójú C_{t_v} vektorral. Valamint az ($F9$) egyenletbeli preferenciákra vonatkozó megkötésnek is teljesülnie kell.

Ennek a 3×3-as problémának a megoldásához négy 2×2-es problémát kell megoldanunk, amelyeket különböző dichotomizálásokkal kapunk. A *Függelék A* részének jelölését követjük, tehát a $K^{m,w}$ egy 2×2-es kontingenciatáblát jelöl, amelyet úgy kapunk a 3×3-as K táblából, hogy az M-típusú férjeket az m-típusba, míg az M-típusú feleségeket w-típusba soroljuk át $(m, w \in \{L, H\})$.

Ezt a jelölést alkalmazva a négy 2×2-es probléma a következőképpen írható fel. Egyrészt, meg kell határozzuk a $(K^{H,H})^*$ tényellentétes táblát, ahol a célperemeloszlások $R^H = \left[R_{t_v,1} \;\; R_{t_v,2} + R_{t_v,3}\right]^{\mathrm{T}}$ és $C^H = \left[C_{t_v,1} \;\; C_{t_v,2} + C_{t_v,3}\right]$, valamint $\mathrm{LL}\left(K_{t_p}^{H,H}\right) = \mathrm{LL}\left[\left(K^{H,H}\right)^*\right]$.

Másrészt meg kell határoznunk a $(K^{L,L})^*$ tényellentétes táblát, ahol a célperemeloszlások $R^L = \left[R_{t_v,1} + R_{t_v,2} \;\; R_{t_v,3}\right]^{\mathrm{T}}$ és $C^L = \left[C_{t_v,1} + C_{t_v,2} \;\; C_{t_v,3}\right]$, valamint $\mathrm{LL}\left(K_{t_p}^{L,L}\right) = \mathrm{LL}\left[\left(K^{L,L}\right)^*\right]$; harmadrészt a $(K^{H,L})^*$ tényellentétes táblát, ahol a célperemeloszlások $R^H = \left[R_{t_v,1} \;\; R_{t_v,2} + R_{t_v,3}\right]^{\mathrm{T}}$ és $C^L = \left[C_{t_v,1} + C_{t_v,2} \;\; C_{t_v,3}\right]$, valamint $\mathrm{LL}\left(K_{t_p}^{H,L}\right) = \mathrm{LL}\left[\left(K^{H,L}\right)^*\right]$; és végül a $(K^{L,H})^*$ tényellentétes táblát, ahol a célperemeloszlások $R^L = \left[R_{t_v,1} + R_{t_v,2} \;\; R_{t_v,3}\right]^{\mathrm{T}}$ és $C^H = \left[C_{t_v,1} \;\; C_{t_v,2} + C_{t_v,3}\right]$, valamint $\mathrm{LL}\left(K_{t_p}^{L,\mathrm{H}}\right) = \mathrm{LL}\left[\left(K^{L,H}\right)^*\right]$.

Az így kapott négy 2×2-es tényellentétes tábla megadja a 3×3-as K^* tényellentétes tábla négy elemét: $N_{L,L}^* = \left(K^{H,H}\right)_{1,1}^*$; $N_{H,H}^* = \left(K^{L,L}\right)_{2,2}^*$; $N_{L,H}^* = \left(K^{H,L}\right)_{1,2}^*$; és $N_{H,L}^* = \left(K^{L,H}\right)_{2,1}^*$. A K^* tábla további öt elemét pedig kifejezhetjük a fenti négy elem és a célperemeloszlások ismeretében.[25]

Tehát így definiáljuk a h függvényt az NM-módszerrel arra az esetre, amikor az asszortatív tulajdonság három különböző értéket vehet fel. A g és f függvények ebben az esetben is rendre $g(K^*) = \mathrm{sum}\,[\mathrm{diag}\,(K^*)]/N$ és $f = g \circ h$.

[25] *Naszódi–Mendonça* [2021] megadja a tényellentétes konstruálási probléma megoldását az általános esetre is, ahol a magtábla tetszőleges méretű. *Naszódi* [2021*b*] pedig két dimenzió (iskolázottság és rassz) szerinti párválasztás esetére terjeszti ki a modellt.