dc_239_11

Módszertani hozzájárulás
a szegénység többváltozós
statisztikai méréséhez

MTA doktori értekezés tézisei

Hajdu Ottó

Budapest, 2011.

Bevezetés

Történelmi alkalom volt, hogy az 1990. évben a Szerző összefoglalást adhatott – elsőként - a hazai tudományos *módszertani* érdeklődés számára a Statisztikai Szemlében – majd a kandidátusi értekezésében (Hajdu: 1991) - a nemzetközi szegénységi irodalom addigi statisztikai, módszertani eredményeiről.

Jelen értekezés a Szerző ezt követő új eredményeit, a területhez tett hozzájárulásait mutatja be. A disszertáció statisztikai témájú és módszertani érdeklődésű. A cél a szegénység és - hozzá kapcsolódóan - a depriváció és kirekesztés jelenségek társadalmi szintű mérésével kapcsolatos statisztikai módszertani kérdések felvetése, megválaszolása.

A szegénység számos vetületben van jelen, ezért az alkalmazott módszertani megközelítés *többdimenziós*. Bár a kérdések főként a *jövedelmi szegénység* tekintetében fogalmazódnak meg, de egyéb dimenziókkal való kapcsolatukban – pl. a *fogyasztás, kiadás, vagyoni helyzet* - is tárgyalásra kerülnek. Továbbá, tekintet nélkül a dimenziók számára – akár az egydimenziós esetben is – a mérés megvalósítása, technikája alapvetően többváltozós jellegű. Megfigyelési egységként az egyén, vagy a háztartás szerepel az éppen tárgyalt módszertani tartalomnak megfelelően.

A szegényvolt klasszifikálása alapvetően *objektív* probléma, de *szubjektív* kérdéseket is felvet, melyek más társadalmi jelenségek - így a *relatív depriváció* és a *társadalmi kirekesztés* – mérésének az igényéhez is elvezetnek, ahol a mérést célszerű az Igen/Nem elhatárolás helyett *fuzzy* jellegű alapokra helyezni.

Az értekezés kétféle tartalomban ad eredményeket:

1. *egyfelől új formulákat és elveket dolgoz ki, diszkutálja és javasolja alkalmazásra,*
2. *másfelől új statisztikai problémákat vet fel, melyek megoldását többváltozós statisztikai módszerek megfelelő alkalmazására vezeti vissza.* A tárgyalt módszertanok összefoglaló bemutatása az értekezés járulékos eredménye.

Az értekezés ezért két részre tagolódik. Az alapkutatások a 2. és 3., az alkalmazások pedig a 4-7. fejezetekben szerepelnek.

Az első fejezet tartalmazza azon fogalmak tömör áttekintését, melyek az alapkutatási eredmények érdekében szükségesek. Később, különösen a standard többváltozós statisztikai eljárások kapcsán

– diszkriminancia analízis, kvantilis és logisztikus regresszió, latens változók, korrespondencia analízis –az alapok ismeretét az értekezés feltételezi.

Az értekezés az újonnan javasolt mérőszámok viselkedésének a vizsgálatához *modell* példákat alkalmaz. Ahol az eredmények valósághű nagyságrendje érdekes, ott empirikus számítások szerepelnek.

Az egyes fejezetek tartalma rendre a következő.

1. *A szegénységmérés operacionalizálása*. Ismerteti azon fogalmakat, formulákat, melyek a mérés módszertanának megértését és továbbvitelét megalapozzák.
2. *A relatív depriváció mérése, és a mérés eredményének beépítése a szegénységi indexbe*. Ennek során az értekezés bevezet egy új relatív deprivációs elvet, diszkutálja, majd az új elvnek megfelelő relatív deprivációs mérőszámokat javasol, végül egy olyan új szegénységi index konstrukciót javasol, mely érzékeny a relatív deprivációban történt elmozdulásokra is. A fejezet a relatív deprivációval szembeni averzió mérése során felveti a szegénységi küszöb alatti sokaság aszimmetrikus volta értelmezésének és mérésének szükségét, amit jelen értekezés a gamma eloszlású változó hatványozására és a Lorenz-görbe alakjának a jellemzésére vezet vissza.
3. Az *egyenlőtlenség-mérés egy új*, *általánosított variancia* megalapozású módszerének a kidolgozása, definiálása, diszkutálása, majd annak a csoportközi dekompozícióban, végül a szegénység mérésében történő alkalmazása.
4. Kvantilis regresszió alkalmazása a *szegénységi küszöb rétegspecifikus* becslésében. Ennek során prediktor jellegű rétegképző ismérvek függvényében becsüljük az adott rétegben várható alsó és felső decilis értékét, mint szegénységi, vagy gazdagsági küszöböt. E relatív megközelítésben bármely rendű kvantilis értéke becsülhető.
5. A *logisztikus regresszió* alkalmazásának kismintás problémái. A szegény vagy nem szegény mivolt Igen/Nem jellegű klasszifikálása problematikus. Valaki kevésbé, más pedig inkább szegény, illetve kevésbé érzi magát szegénynek, más pedig inkább. E *fuzzy* megközelítés érvényesítését szolgálja a logisztikus regresszió módszere.
6. A SEM (*Structural Equation Modelling*) kauzalitási modell alkalmazása a *szegénység-depriváció-kirekesztés* strukturális rendszerben. A modell ezen latens változók között az ok-okozati irányultságokat hipotézisként kezeli, teszteli, majd becsli a strukturális koefficienseket. A hipotézisek tesztelésének kimenete a megfelelő empirikus manifeszt változók kiválasztásán alapul.
7. Korrespondencia analízis alkalmazása a *szegénységi kockázattal* (küszöb alá csúszással) asszociáló változók és a kategóriák egymással való megfeleléseinek a feltárásában, fuzzy szegény háztartások klasszifikálása érdekében.

Az értekezés önálló eredményei – a fejezetek szerint haladva - a következők.

1. <u>1.fejezet:</u> Az irodalomban ismert szegénységi mérőszámok bemutatása, a kompozit szegénységi index szerkesztési elveinek keretbe foglalása. (Hajdu: 1997a)

2. 2.fejezet: A regresszív jövedelmi transzfer hatásának vizsgálata a relatív depriváció egyéni és társadalmi szintű fokára, és erre érzékeny új deprivációs mutatók megadása, kidolgozása. (Hajdu: 1996, 1999)
3. 2.fejezet: Új szegénységi mutató konstrukció definiálása, mely érzékeny a szegények körében mért relatív deprivációs változásokra. (Hajdu: 1997b, 1999)
4. 2.fejezet: A deprivációs averzió becslése a gamma-eloszlású jövedelem hatványkitevő paramétere alapján. (Hajdu: 2008)
5. 2.fejezet: A Lorenz-görbe aszimmetriájának – mint a szegénységi mérték egy faktoraként való - értelmezése és mérése. (Hajdu: 2004a)
6. 3.fejezet: Egy új, csoporthatásokra dezaggregálható, információelméleti megalapozású egyenlőtlenségi *módszer* kidolgozása. *Elsőként a Pályázó MTA doktori értekezésében publikálva.*
7. 3.fejezet: Az új redundancia alapú egyenlőtlenségi módszer alkalmazása a szegénység mérésében. *Elsőként a Pályázó MTA doktori értekezésében publikálva.*
8. 4.fejezet: Rétegzett sokaság esetén a réteg-specifikus deprivációs küszöb rétegen belüli becslése kvantilis regresszió alkalmazásával, a rétegképző változók értékei függvényében. *Elsőként a Pályázó MTA doktori értekezésében publikálva*.
9. 5.fejezet: Rétegzés esetén a küszöbalatti kismintás esetből adódó becslési és tesztelési problémák egzakt mintavételi módszerrel való kezelése - a logisztikus regresszió prediktor változóinak a szelektálása során. (Hajdu: 2004c, 2006a)
10. 6.fejezet: A szegénység-depriváció-kirekesztettség kauzalitási rendszer tesztelése a SEM (Structural Equation Modelling) modell MIMIC (Multiple Indicator Multiple Cause) módszerének alkalmazásával. (Hajdu: 2009)
11. 6.fejezet: Háztartások jövedelmezőségének, kiadási hajlandóságának és fogyasztási színvonalának vizsgálata különböző méretdefiníciók mellett az MTMM (Multitrait-Multimethod) modell alapján, a CFA (konfirmatív faktoranalízis) alkalmazásával. (Hajdu: 2004d)
12. 6.fejezet: A SEM módszer ADF (Asymptotically Distribution Free) becslésének az alkalmazása a homoszkedaszticitási hipotézis tesztelésére a hibafaktorok kovariancia mátrixára vonatkozóan a faktormodellben. (Hajdu: 2004b)
13. 7.fejezet: Diszkretizált változók kategóriáinak az elhelyezése a prediktív térképen, a szegénységi küszöb alá *süllyedésre* utaló, azzal asszociáló kategóriákat kereső modellben, a többszörös korrespondencia analízis alkalmazásával. (Hajdu: 2002)

Az új formulák verifikálását szolgáló számítások a Szerző önálló programozási eredményei. Egyébként standard statisztikai programcsomagok kerültek alkalmazásra. Végül fölhívjuk a figyelmet, hogy bár a jelölésrendszer törekszik a konzisztenciára, az egyes fejezetek saját jelölésrendszert alkalmaznak.

1. A szegénység statisztikai mérésének operacionalizálása

A társadalmi szegénység fokának megadásakor alapvető mozzanat a szegények arányának, és a szegénység eloszlásának a vizsgálata. Az eloszlás egy jellemzője a centrális tendencia szegénységi küszöbhöz való viszonya, a *szegénység intenzitása*. Másfelől a szegénység mérésekor figyelmet kell szentelnünk a szegények körében a szóródás mértékének is. A szóródást első megközelítésben kézenfekvő egyenlőtlenségként kezelni és mérni.

A kompozit szemléletű egydimenziós szegénységi index

A szegénység társadalmi szintű mérésének első lépéseként elhatároljuk a szegények körét a nem szegényektől, majd a szegényekről rendelkezésre álló információt aggregáljuk egy P kompozit index értékében.

Tekintsük az n tagú társadalom *jövedelmi* vektorát, melyben a rögzített z szegénységi küszöb alá eső q számú személyt szegényként, a többi (n-q) számú egyént pedig mint nem szegényt klasszifikáljuk:

$$Z_1 \leq Z_2 \leq \ldots \leq Z_q < z \leq Z_{q+1} \leq Z_{q+2} \leq \ldots \leq Z_n . \qquad (1.1)$$

A szegénységi mérték kalkulálása előbb egy *identifikálási*, majd egy *aggregálási* lépést foglal magában.

Identifikálás

Az identifikálás alapvető kérdései - melyek eltérő klasszifikációra vezethetnek - az alábbiak.

- *A megfigyelés egysége:* A *megfigyelés egysége* lehet az *egyén*, a *háztartás*, vagy a *család*.
- *A hiány tárgya* tekintetében a társadalom egy tagja lehet jövedelmi szegény, fogyasztásában szegény, küszöb alatti kiadású, vagyontalan, kirekesztett, stb.
- *Az összehasonlíthatóság* miatt a háztartás létszámának a megadásakor az *ekvivalencia skála* a fogyasztási egység.
- A *szegényvolt elhatárolás* történhet *objektív küszöb* alapján. A küszöb rögzítésére *abszolút* és *relatív* módszerek is kínálkoznak. Relatív módszer pl. az alsó decilis értékének az alkalmazása, vagy a medián jövedelem adott százaléka. Ezzel szemben az abszolút módszer egy minimális fogyasztói kosár költségének a kalkulálásán alapul. Valamely módszer szerint a szegénység megszüntethető a társadalomban, míg más módszer szerint nem. *Szubjektív ítélet* szerint szegénynek érezheti magát valaki a környezete viszonylatában, ha relatív értelemben *deprivált* valamely jószág tekintetében. Ezen érzés jelentkezik, ha nem birtokolja a jószágot, de szeretné, és megvalósíthatónak is tartja a birtoklását. Végül egy kényszerű *kirekesztettség* elszenvedése valamely *társadalmi funkcionalitásból* is a szegénysorba jutás irányába mutathat.

Aggregálás

Követelmény, hogy a kompozit P index ne reagáljon a nem szegények körében történt jövedelmi változásokra. Ennek érdekében az aggregálás során a jövedelmeket a z küszöb szintjén *felülről cenzoráljuk*. A cenzorált jövedelmi eloszlás a társadalom minden tagjára értelmezett:

$$Z_i^c = \min\left\{Z_i, z\right\}, \quad \left(i = 1,2,\ldots,n\right) \qquad (1.2)$$

ahol a küszöb fölött lévők jövedelmeit magával a küszöbbel helyettesítjük.

Az aggregálási lépésben az egyedi szegénységi résekben lévő, az egyedi *depriváltság* fokát leíró

$$g_i = z - Z_i^c \tag{1.3}$$

„*poverty-gap*" információt tömörítjük a kompozit P indexbe. A küszöbszintet egységnyiként kezelve a relatív szegénységi rés: r_i (i=1,2,…,n).

Követelmény, hogy P *növekvően* reagáljon az alábbi faktorok *c.p.* növekményére:

1. „*Head-count" ratio*, mint a szegénység kiterjedtsége:

$$H = \frac{q}{n} \tag{1.4}$$

2. „*Income-gap ratio*", mint a szegénység intenzitása:

$$I_W = 1 - \frac{Z_W(.)}{z} \tag{1.5}$$

ahol $Z_W(.)$ a „*reprezentatív*" jövedelmi szint, mellyel mindenki jövedelmét helyettesítve az *aktuális* eloszlásban, az eloszlás *W társadalmi jóléti szintje* változatlan marad.

3. „*Welfare-gap ratio*", mint jóléti veszteség:

$$W_I = 1 - \frac{W(.)}{W(z)} \tag{1.6}$$

4. „*Inequality*", mint az E(.) egyenlőtlenség intenzitása:

$$0 \le E(.) \le 1 \tag{1.7}$$

Elvárás, hogy *Min* legyen P értéke, ha nincsenek szegények a társadalomban, és *Max* értékkel jelezze, ha mindenki jövedelme zéró. Ésszerű normalizálási mozzanat, hogy ha minden szegény jövedelme egyenlő, akkor a H arány és a küszöb alatti átlagos jövedelmi szint együtt elegendő információt tartalmaz:[1]

$$HI = \frac{q}{n}\frac{z - \bar{Z}_z}{z} \tag{1.8}$$

A HI mutató a „*normalizált szegénységi érték*", mely a küszöböt mindenkinek biztosító alap azon hányada, mellyel minden szegényt a küszöb szintjére lehetne emelni, tehát a szegénység (statikusan, *c.p.*) eliminálható lenne.

A fenti követelményeknek megfelelő kompozit szegénységi indexet a cenzorált, vagy a csonkolt eloszláson szerkeszthetünk. Mivel a teljes népesség cenzorált eloszlása implicite tartalmazza a szegények létszámarányát, ezért esetében a H mutató explicit beépítése a formulába nem szükségszerű. A *csonkolt* eloszlás a küszöb fölötti jövedelmeket elhagyja az eloszlásból, tehát a H mutató számítása és beépítése a P index formulájába ekkor szükséges.

A *csonkolt* eloszláson mérve a szegénységet - a Sen-index nyomán - a P index alábbi konstrukciós elvei rajzolódnak ki.

[1] A „*z*" alsó index arra utal, hogy az illető mértéket a küszöb alattiakra vonatkozóan számítottuk.

1. *A súlyozott szegénységi rések elve:* változtatható a súlyrendszer, és változtatható a normalizálási alap.
2. *Az átlagos egyenlőtlenségi rés elv*e: a HI_W mértékben I_W a totális egyenlőtlenség 1 felső határának, és a szegény jövedelmek $E(.)$ indexének a súlyozott átlaga az I és az (1-I) súlyok mellett. Cserélhető az $E(.)$ formula, és alkalmazható az abszolút és a relatív rések egyenlőtlenségére is.
3. *Az egyenlőtlenséggel inflált HI elve:* az (1+$E(.)$) inflátor faktorban változtatható az egyenlőtlenség mérőszáma, és annak argumentuma is.
4. *Az I_W jóléti rés HI_W etikai elve:* változtatási lehetőség a jóléti függvény w súlyának, és formulájának a cseréje.
5. *A H és HI határok átlaga:* cserélhető az $E(.)$ formula és az átlagolás módja.

A cenzorált eloszláson mérve a szegénységet, a következő elveket emeljük ki.

6. *Az átlagos relatív rés elve:* Az SST index a teljes cenzorált Z^c eloszláson a *relatív szegénységi rések súlyozott átlaga*, súlyként az első (2n-1) páratlan számokat alkalmazva. Az SST index felírható multiplikatív módon, az *inflált HI* elv alapján is: az inflátor faktorban a *cenzorált* jövedelmek *relatív szegénységi réseinek* Gini indexe kap helyet. A Gini-formula és a rés jellege változtatható.
7. *A Dalton-féle jóléti veszteség:* az átlagos cenzorált haszon relatív rése a küszöb hasznossági szintjének a viszonylatában. Additív módon csoporthatásokra bontható viszonyszám. A mutató konkrét értéke a hasznossági függvény alkalmas megválasztásán múlik.
8. *A cenzorált jövedelmek egyenlőtlenségi mértéke:* a szegénységi index Takayama-féle számítási módja a szegénység mértékét a *cenzorált* jövedelmek *Gini egyenlőtlenségeként* definiálja. A Gini-index helyett más formulát véve más P indexekhez jutunk.

A szegénységi mérőszámok konstrukciós elveit az 1.1 tábla foglalja össze.

1.1 tábla *Szegénységi mérőszámok konstruálása*

Definíció	*A szegénységi indexhez felhasznált eloszlás*	
	„Csonkolt”	*„Cenzorált”*
$\frac{1}{Norm}\sum_i súly_i \cdot f\left(rés_i\right)$	Sen, Thon, Kakwani, Foster et al., abszolút g réssel	Sen-Shorrocks-Thon (SST), relatív r réssel
$Átlag\left(H_{E_z}, \left[HI\right]_{1-E_z}\right)^*$	Robusztus (Gini, mértani) Blackorby et al. (Atkinson, számtani) Sen (Gini, számtani)	
Egyenlőtlenségi: $E\left(Z^c\right)$		Takayama (Gini)
Inflált: $HI \cdot \left[1 + E\left(rés\right)\right]$	Clark1 et al. (Atkinson), Sen (Gini) g abszolút rés	SST (Gini) r relatív rés
Etikai: HI_W	Blackorby et al., Sen	
Jövedelmi rés-arány: I_W		Clark2 et al., Hagenaars2
Jóléti veszteség: W_I		Chakravarty, Foster et al., Watts, Hagenaars1,3

* A jobb alsó index az átlagszámításban alkalmazott súlyt jelöli.

A Fuzzy szemléletű többdimenziós szegénységi index

A társadalom tagjai egyidejűleg több dimenzióban is lehetnek szegények, és szegénységük nem az Igen/Nem bináris módon klasszifikálandó, hanem egy latens, „szegénységi” tengelyen való helyzettel jellemzendő, ahol valaki inkább vagy kevésbé – tehát *fuzzy* (talán) szegény.

A vizsgált dimenziók legyenek az F_j (j=1,2,…,k) fogyasztások, i=1,2,…,n fogyasztó esetén, rendre w_j súlyokkal.

A μ *tagsági függvény* a j fogyasztás tekintetében három pozíciót különböztet meg az m_j küszöb rögzítésével:

1. Biztosan szegény j tekintetében: μ=1, ha zéró a fogyasztás,
2. Lehetséges (fuzzy) szegény j tekintetében: 0<μ<1, ha küszöb alatti a fogyasztás,
3. Biztosan nem szegény j tekintetében: μ=0, ha küszöb fölötti a fogyasztás.

Az *átfogó, többdimenziós, fuzzy* szegénységi index az *individuális tagsági-score* értékek átlaga, amely score értékekben az egyes fogyasztások rendre w_j súllyal szerepelnek.

A társadalmi kirekesztés mérése

Kirekesztés alatt most azt értjük, hogy valaki valamilyen társadalmi funkcionalitásban nem vehet részt. A többdimenziós fuzzy szemléletben itt a fogyasztás helyére a funkció lép, és a tagsági függvény bináris, 0/1 kimenetű.

Az egyéni *deprivációs score* így a súlyok összege minden dimenzióra, a társadalmi kirekesztés indexe pedig az *átlagos score* minden egyénre.

Az értekezés hozzájárulása a szegénység és egyenlőtlenség mérésének módszertanához a fenti keret tükrében látható.

2. A relatív deprivációs szegénységi index

A szegénység társadalmi fokának megadásakor a *szegényarány* melletti másik alapvető faktor a szegénység eloszlása. Az eloszlás helyzeti jellemzője a küszöb alatti centrális tendencia viszonya a küszöbhöz, a szóródási jellemző pedig rendszerint valamely egyenlőtlenségi mérték. Az eloszlás eredő jellemzője az I_W mérték. Mindazonáltal a szegények jövedelmi színvonalának és szóródásának a jellemzése az egyenlőtlenség mellett *a relatív depriváció mértékét* is igényli.

A fejezet egy olyan szegénységi mérőszám elvet javasol, mely a szegénységi mértékben a relatív depriváció fokát is figyelembe veszi.

A relatív depriváció tárgya bármely *jószág* lehet. A relatív depriváció két hatás eredőjeként alakul. Az egyik a depriváció *érzete*, a másik pedig annak *relatív* megítélése.

Adott jószág tekintetében deprivált személy *Runciman*-féle kritériumai:

> i) nem rendelkezik az illető jószággal, ii) más személyeket lát, akik e jószág birtokában vannak,
> iii) birtokolni akarja ezt a jószágot, iiii) megvalósíthatónak tartja, hogy e jószág birtokába jusson.

Szemben az egyenlőtlenséggel, a relatív depriváció magában foglalja, hogy az emberek inkább a társadalom adott csoportjaihoz, mintsem a társadalom egészéhez viszonyítják magukat. Azon egyének körét, akikhez „*i*” viszonyítja magát, „*i*” referencia csoportjának nevezzük. A relatív depriváció tárgya egy konkrét jövedelmi szint is lehet. Ekkor az Y_i jövedelemmel bíró személy deprivált mindazon referencia személyekkel szemben, akiknek a jövedelme nagyobb mint Y_i. E tényt az

$$i \prec j \in J_i \quad | \, Z_i < Z_j \tag{2.1}$$

reláció jelzi, ahol J_i az *i referencia csoportját* jelöli. A referencia csoportba nem tartozókkal szembeni depriváltság definíció szerint zérus.

Realitás, hogy az egyenlőtlenséget növelő regresszív transzfer nyomán a relatív depriváció egyéni mértéke (érzete) csökkenhet a referencia csoportokhoz való viszonyok módosulása szerint. Egy kompozit deprivációs metrika a csökkentő hatásokat is figyelembe veszi, melyek eredőjeként a globális depriváció kevésbé emelkedik, vagy - konstrukciójától függően, adott esetben – csökkenhet is. Célunk olyan *P* szegénységi index megadása, mely a relatív depriváció csökkenését a küszöb alatt - szegénységet csökkentő faktorként kezeli. A küszöb alatt kétféle deprivációt értelmezünk:

1. a *küszöbbel* szemben,
2. a *többi szegénnyel* szemben érzett depriváltságot.

A deprivációs függvény

Tekintsük a *z* szegénységi küszöb alatti *q* tagú társadalom egyedeinek a jövedelmeit, ahol az egyes jövedelmek nemcsökkenően sorba rendezve követik egymást

$$Z_1 \le Z_2 \le \ldots \le Z_q \tag{2.2}$$

Az „*i*” deprivációs függvényét a Z_j referencia jövedelem viszonylatában úgy definiáljuk, hogy

1. magasabb referencia jövedelemmel szemben magasabb legyen a depriváltság érzete,
2. a határdepriváltság a jövedelmi szint növekedésével csökkenjen,
3. a határdepriváció csökkenése a jövedelmi szint növekedésével egyre nagyobb legyen.

A relatív depriváció mérésére az alábbiakban két megközelítést alkalmazunk:

1. A „*Q*” deprivációs hányad, és
2. a „*Delta*” kommunalitás elvet.

A Q elv szerint az *i* szegény deprivációs hányada a Z_j referencia jövedelem tekintetében

$$\begin{aligned} Q_{ij}^r &= \left(1 - \frac{Z_i}{Z_j}\right)^r \quad | \, Z_i < Z_j \\ Q_{ij}^r &= 0 \quad \textit{egyébként} \end{aligned} \tag{2.3}$$

ahol *r* a *depriváció-averzió* paraméter, növekvő értékkel hangsúlyozva a szegényebb szegény depriváltságát. Pozitív depriváció csak gazdagabbal szemben jelentkezik, értéke egyébként zéró. A Q_{ij} típusú depriváció zéró *Z* jövedelmek esetén is értelmezett, számítható. „*Q*” jelentése r=1 esetén: a referencia jövedelem ekkora hányada (százaléka) hiányzik ahhoz, hogy a referencia személy tekintetében a depriváltság megszűnjön.

Másik megközelítésben a depriváció fokát a

$$\delta_{ij}^{r} = \left(1 - Q_{ij}\right)^{-r} \tag{2.4}$$

delta mutatóval is mérhetjük mindaddig, míg a tört értelmezve van. Delta jelentése az r=1 esetben a referencia jövedelem szintje, ha a deprivált jövedelem 1. Itt a deprivációs hányad

$$1 - Q_{ij} \tag{2.5}$$

komplementere *kommunalitás* értelmű, mert i ilyen hányadban eliminálta a j deprivációt.

A szegénység depriváció-érzékeny mértéke

Definiáljuk a $P=HI_W$ típusú szegénységi indexet úgy, hogy az I_W *küszöbalattiság* faktort a *küszöb* és a *reprezentatív deprivált* szegény jövedelme közötti *deprivációs réssel* mérjük, két lépésben:

1. előbb a *reprezentatív szegény* jövedelmi szintjét kalkuláljuk, mely a küszöbbel szembeni globális deprivációt reprezentálja (őrzi meg),
2. ebből jutunk el a *reprezentatív deprivált* szegény jövedelméhez, mely a szegények egymás közötti globális deprivációját reprezentálja (őrzi meg).

Az új szegénységi index konstrukciója

$$\boxed{P = H \cdot \left[1 - \hat{z}\left(1 - \Delta\right)\right]} \tag{2.6}$$

ahol

1. $\hat{z}$: *a reprezentatív szegény jövedelmi* szintje,
2. Δ: *a reprezentatív deprivált deprivációja*: akivel szemben már senki sem deprivált, reprezentálja a gazdagabb szegénnyel szembeni depriváció fokát egy virtuális, kétfős populációban,
3. a *reprezentatív deprivált szegény jövedelme:* a reprezentatív szegény jövedelméből való részesedése:

$$\hat{z}\left(1 - \Delta\right) \tag{2.7}$$

4. végül *a reprezentatív szegény deprivációs – szegénységi - rése*:

$$1 - \hat{z}\left(1 - \Delta\right) \tag{2.8}$$

Minél tágabb a deprivációs rés, annál magasabb a P szegénységi index, mely láthatóan

- értelmezve van a zéró jövedelmekre,
- normált: $0 \le P \le 1$,
- érzékeny a regresszív transzfer relatív deprivációt csökkentő faktoraira is,
- transzfer érzékenysége a $\hat{z}$ és a Δ metrikák megválasztásán múlik.

A reprezentatív szegény jövedelmének meghatározása

Deprivációs jövedelmi hányad megközelítésben – Q-elven - a Z jövedelmek r=1 *deprivációi* a z küszöb viszonylatában, rendre

$$\left(1-\frac{Z_1}{z}\right) \geq \left(1-\frac{Z_2}{z}\right) \geq \ldots \geq \left(1-\frac{Z_q}{z}\right) \tag{2.9}$$

ahol Z_i/z=1-Q_i a jövedelmi küszöbbel szembeni kommunalitását adja, *komplementere pedig azt mondja, hogy a küszöb hány százaléka hiányzik az i jövedelem küszöb szintre emeléséhez.* Alacsonyabb kommunalitás értelemszerűen magasabb deprivációt eredményez.

Legyen most az átlagos depriváció a küszöbbel szemben

$$Q_z^{(r)} = \left(\frac{1}{q}\sum_{i=1}^{q}\left(1-\frac{Z_i}{z}\right)^r\right)^{\frac{1}{r}} \tag{2.10}$$

Ebben a megközelítésben a reprezentatív szegény jövedelme

$$\boxed{\hat{z}_Q^{(r)} = 1 - Q_z^{(r)}} \tag{2.11}$$

Speciálisan $\hat{z}_Q^{(1)}$ a kommunalitások számtani átlaga.

Másik megközelítésben, a kommunalitások *r-rendű momentumára* alapozva a reprezentatív szegény-jövedelem

$$\boxed{\hat{z}_\delta^{(r)} = \left(\frac{1}{q}\sum_{i=1}^{q}\left(1-Q_i\right)^r\right)^{\frac{1}{r}}} \tag{2.12}$$

Ekkor $\hat{z}_\delta^{(1)}$ a kommunalitások számtani, $\hat{z}_\delta^{(-1)}$ pedig a harmonikus átlaga. Mivel a harmonikus átlag kisebb (nem nagyobb) mint az aritmetikai átlag, ezért c.p. $\hat{z}_\delta^{(-1)}$ magasabb szegénységet jelez mint $\hat{z}_\delta^{(1)}$ alkalmazása.

Relatív depriváció a szegények körében

Hajtsunk végre egy regresszív transzfert a *P* szegényebb szegénytől a gazdagabb *R* szegény felé úgy, hogy ami jövedelmet a szegényebb elveszt, azt a gazdagabb kapja, de mindketten a küszöb alatt maradnak. A populáció két csoportra bomlik: egyrészt a transzfer által nem érintett, másrészt a transzfert adó *P donor*, és a transzfert kapó *R címzett* egyének kétfős csoportjára.

A transzfer hatására mind a donor és a címzett által, mind a velük szemben érzett depriváltság mértéke változik. E változások eredőjére természetesen az is befolyással van, hogy a transzfer eredményeként megváltozik-e a referencia csoportok struktúrája.

Tekintsünk el a referencia csoportok megváltozásától: a donor jövedelme nem süllyed nála szegényebbé alá, a címzetté pedig nem emelkedik nála gazdagabbé fölé. *Ekkor a donorral szemben, és a címzett által érzett depriváltságok csökkennek, viszont a címzettel szemben, illetve a donor által érzett deprivációk nőnek. E hatások eredőjeként egy regresszív transzfer nyomán a relatív depriváció foka esetileg csökkenhet is.*

Vegyük végül, hogy a transzferrel vagy a donor referencia csoportja bővül, vagy a címzetté szűkül. Azoknak, akik bekerülnek a donor referencia csoportjába, megszűnik a donorral szembeni, és megjelenik a donor velük szembeni deprivációja. Másfelől megjelenik azoknak a címzettel szembeni deprivációja, akik kiesnek a címzett referencia csoportjából, miközben eltűnik R velük szemben érzett deprivációja. Látható tehát, hogy a jövedelmi transzfer deprivációt növelő vagy csökkentő hatása attól is függ, hogy a transzfer után a jövedelmi rangsorban P mennyivel kerül lejjebb, és R mennyivel feljebb.

Legyen az i szegény (átlagos) relatív depriváltsági értéke r averzió mellett

$$\begin{aligned} Q_{1_i}^{(r)} &= \frac{1}{q_i}\sum_{j=1}^{q} Q_{ij}^{r} \\ Q_{1_i}^{(r)} &= 0 \quad | \, q_i = 0 \end{aligned} \tag{2.13}$$

ahol q_i az i szegény által érzett *pozitív depriváció k száma*. Vezessük be a relatív depriváció társadalmi szintű mérésére a *globális deprivációs hányad* új mutatót, mely a Q_{ij} deprivációs hányadosokat minden pozitív deprivációs viszonyra átlagolja:

$$\boxed{Q^{(r)} = \left(\frac{1}{N_Q} \sum_{i=1}^{q} \sum_{j=1}^{q} Q_{ij}^{r} \right)^{\frac{1}{r}}} \tag{2.14}$$

ahol N_Q a *pozitív depriváció k száma*, minden páronkénti viszonylatot figyelembe véve. Q az egyéni Q_{1i} pozitív deprivációs hányadok súlyozott számtani átlaga. A deprivációs hányad értékét felhasználva, a deprivációs P szegénységi index a megfelelő behelyettesítéssel számítható.

Az alábbiakban azt vizsgáljuk, hogy a globális deprivációs hányad milyen körülmények között csökkenhet egy regresszív transzfer hatására - az egyszerűség kedvéért - r=1 averzió mellett.

Tekintsük a Z_1=[1;4;10;20;35] jövedelmi eloszlást, melyre $Q^{(1)} = 0.75$, és a Z_2=[1;4;10;34;35] eloszlást, melyre $Q^{(1)} = 0.7409$. Módosítsuk az eloszlásokat különböző d nagyságú transzferekkel, a tulajdonosok valamennyi párosítását tekintve. A transzferált eloszlások deprivációs hányadait Z_1 esetén a 2.1 tábla, Z_2 esetén pedig a 2.2 tábla közli. A táblákban a deprivációs hányadok d növelése mellett addig kerültek kiszámításra, míg a donor jövedelme nem vált negatívvá, és aláhúzás határolja el d azon tartományát, amely mellett a jövedelmi rangsor változatlan marad. Közben ***kiemelten*** jelenik meg azon *kritikus* d_c érték, mely a deprivációs hányadnak a d=0 eredeti állapothoz képest magasabb értékét az alacsonyabbtól elválasztja.

A 2.1 és 2.2 táblák eredményei alapján az alábbi *tendenciák* rajzolódnak ki. Mindaddig, míg a transzfer a jövedelmi *rangsort* változatlanul hagyja:

1. Ha a donornál nincsenek szegényebbek a jövedelmi rangsorban, akkor a deprivációs hányad értéke a transzfer növekedésével együtt minden esetben nő, de a növekedés annál csekélyebb mértékű, minél közelebb van a rangsorban a donor a címzetthez.
2. Ha a donor nem a legszegényebb, akkor a deprivációs hányad csökkenése is lehetséges, mégpedig kétféle módon. Egyrészt úgy, hogy d fokozatos növelésével először a deprivációs hányad is nő, majd elérve egy maximumot, attól kezdve csökken a d=0 eredeti állapothoz tartozó szintre, majd az eredeti állapothoz képest is csökken. Ez a helyzet pl. a 2.1. táblában (P,R)=(2,3),(2,4),(2,5) transzfer pozícióknál a d<=3 tartományon, a (P,R)=(3,4),(3,5) relációkban a d<=6 tartományon,

végül a $(P,R)=(4,5)$ párosításban a $d<=10$ tartományon. Másrészt a deprivációs hányad a $d=0$ szintről indulva, d növelésével rögtön elkezdhet csökkenni. Ez történik a 2.2. táblában a $(P,R)=(3,4)$, $d<=1$ tartományon.

3. Minél közelebb van a rangsorban a donor a címzetthez, annál szélesebb d azon tartománya, amelyre a deprivációs hányad értéke az eredeti állapothoz képest csökken.
4. Mikor a transzfer nagysága miatt a tulajdonosok *rangpozíciója megváltozik*, a deprivációs hányad növekedése és csökkenése d függvényében váltogathatja egymást, tehát a csökkenés a transzfer nagyságának több tartományán is bekövetkezhet. Ezt tapasztaljuk a 2.1. táblában a $(P,R)=(3,4)$ párosításban, a $d>=8$ tartományon.

Adott jövedelmi helyzetben egy d mértékű regresszív transzfer hatására a relatív depriváció globális mértéke csökkenhet, és a deprivációs hányad mutató ezt érzékeli.

Kiindulásként megállapítjuk, hogy ha a donornál nincsenek szegényebbek, akkor a donorral szemben nem léphet fel deprivációs csökkenés, miközben a címzett által érzett deprivációs csökkenést kiegyenlíti a donor által a címzett referencia csoportjával szembeni növekedés, tehát a deprivációs hányad értéke biztosan nő ez esetben, függetlenül a címzett pozíciójától.

Ha viszont vannak depriváltak a donorral szemben, akkor a deprivációs hányad változásának irányát és mértékét mozgató tényezők a következők:

1. a transzfer d mértéke,
2. P és R rangpozíciói a jövedelmi rangsorban,
3. P és R jövedelmeinek az egymáshoz való viszonya.

Részletesen (Hajdu: 1996):

1. Rögzített P és R szegények közötti d transzfer mellett, a deprivációs hányad változásának az irányát a Z_P és Z_R jövedelmek egymáshoz való viszonya szabályozza, mert adottnak véve a Z_P jövedelmi szintet, a Z_R jövedelem *kisebb/nagyobb* viszonya - a deprivációs hányad változás iránya szempontjából *kritikus* Z_{Rc} *érték* tekintetében - az irányt *esetlegesen* átválthatja.
2. Mikor a deprivációs jövedelmi hányad a transzfer nyomán alacsonyabb az eredeti szintjénél, d növelése a relatív depriváció fokát tovább csökkenti.
3. A transzfer d méretének a hatása a d_c kritikus értékhez való viszonyának a függvénye: $d<d_c$ esetén d növelése előbb növeli, majd csökkenti a relatív depriváció fokát, míg $d>=d_c$ esetén d növelése egyértelműen csökkenti azt.
4. A transzfer által nem érintettek oldaláról - rögzített d transzfer mellett – a deprivációs hányad csökkenését növeli, ha minél több olyan személy van, aki a donorral szemben deprivált, ha e személyek minél kevésbé depriváltak a donorral szemben, és ha a donor és címzett szomszédos helyet foglalnak el a jövedelmi rangsorban.
5. A transzfer által nem érintettek oldaláról - rögzített d transzfer mellett – a deprivációs hányad csökkenését növeli a donor jövedelmének csökkenése, és a címzett jövedelmének a növekedése, c.p.
6. Ha a transzfer nagysága a referencia csoportok struktúrájának a megváltozását eredményezi, akkor minél nagyobb transzfer mellett következik ez be, annál inkább a relatív depriváció fokának a növekedése várható.

Mindazonáltal, ha a regresszív transzfer deprivációt csökkentő faktorai esetleg enyhén túl is szárnyalják a növelő faktorokat, ez várhatóan csak tompítja a szegénységi index növekedését a transzfer nyomán, mert a reprezentatív küszöb a súlyozási mód megfelelő megválasztása mellett biztosan csökken.

2.1 tábla: A regresszív transzfer hatása a $Q^{(1)}$ indexre Z_1 esetén

Z_1 = [1, 4, 10, 20, 35]: $Q^{(1)}$ = 0.75										
	A transzfert adó és kapó személyek (*P_R*)									
d	*1_2*	*1_3*	*1_4*	*1_5*	*2_3*	*2_4*	*2_5*	*3_4*	*3_5*	*4_5*
0.2	0.7559	0.7579	0.7587	0.7591	0.7516	0.7524	0.7528	0.7507	0.7511	0.7504
0.4	0.7614	0.7658	0.7674	0.7683	0.7530	0.7546	0.7555	0.7513	0.7522	0.7507
0.6	0.7663	0.7735	0.7761	0.7774	0.7541	0.7567	0.7580	0.7519	0.7532	0.7510
0.8	0.7708	0.7811	0.7847	0.7865	0.7549	0.7585	0.7602	0.7524	0.7541	0.7513
1	0.7750	0.7886	0.7933	0.7955	0.7553	0.7600	0.7622	0.7528	0.7550	0.7506
1.2					0.7554	0.7612	0.7639	0.7531	0.7558	0.7518
1.4					0.7550	0.7620	0.7652	0.7533	0.7565	0.7520
1.6					0.7540	0.7623	0.7661	0.7534	0.7572	0.7522
1.8					0.7524	0.7620	0.7663	0.7535	0.7578	0.7524
2					***0.7500***	0.7609	0.7658	0.7534	0.7583	0.7525
2.2					0.7465	0.7588	0.7643	0.7532	0.7587	0.7526
2.4					0.7415	0.7552	0.7613	0.7530	0.7590	0.7526
2.591					0.7349	***0.7500***	0.7567	0.7526	0.7592	0.7526
2.6					0.7345	0.7497	0.7564	0.7526	0.7593	0.7526
2.768					0.7263	0.7428	***0.7500***	0.7521	0.7594	0.7526
2.8					0.7245	0.7412	0.7485	0.7520	0.7594	0.7526
3					0.7096	0.7278	0.7358	0.7514	0.7594	0.7525
3.2					0.7314	0.7511	0.7597	0.7506	0.7592	0.7525
3.333					0.7458	0.7667	0.7757	***0.7500***	0.7591	0.7524
3.4					0.7531	0.7744	0.7837	0.7497	0.7589	0.7523
3.6					0.7747	0.7977	0.8076	0.7486	0.7585	0.7522
3.8					0.7963	0.8209	0.8316	0.7473	0.7579	0.7520
4					0.8179	0.8442	0.8555	0.7458	0.7572	0.7518
5								0.7350	***0.7500***	***0.7500***
6								0.7154	0.7343	0.7471
6.5								0.7268	0.7477	0.7452
7								0.7370	0.7600	0.7429
7.5								0.7452	0.7703	0.7403
8								0.7500	0.7772	0.7372
8.5								0.7480	0.7774	0.7336
9								0.7293	0.7609	0.7295
9.5								0.7939	0.8277	0.7248
10								0.8583	0.8944	0.7194
15										0.7350
20										0.8977

2.2 tábla: A regresszív transzfer hatása a $Q^{(1)}$ indexre Z_2 esetén

$Z_2 = [1, 4, 10, 34, 35]$: $Q^{(1)} = 0.7409$										
	A transzfert adó és kapó személyek (*P_R*)									
d	*1_2*	*1_3*	*1_4*	*1_5*	*2_3*	*2_4*	*2_5*	*3_4*	*3_5*	*4_5*
0.2	0.7468	0.7488	0.7487	0.7498	0.7425	0.7424	0.7435	0.7407	0.7418	0.7420
0.4	0.7522	0.7566	0.7566	0.7588	0.7439	0.7438	0.7460	0.7405	0.7427	0.7431
0.6	0.7572	0.7644	0.7644	0.7677	0.7450	0.7450	0.7483	0.7402	0.7435	0.7441
0.8	0.7617	0.7720	0.7722	0.7766	0.7457	0.7459	0.7503	0.7398	0.7442	0.7452
1	0.7659	0.7795	0.7800	0.7855	0.7462	0.7467	0.7521	0.7394	0.7449	0.7462
1.2					0.7462	0.7482	0.7536	0.7401	0.7455	0.7472
1.4					0.7458	0.7494	0.7548	0.7407	0.7461	0.7482
1.6					0.7449	0.7500	0.7554	0.7412	0.7466	0.7492
1.8					0.7433	0.7502	0.7555	0.7417	0.7469	0.7501
2					***0.7409***	0.7495	0.7547	0.7420	0.7472	0.7511
2.2					0.7374	0.7478	0.7530	0.7423	0.7444	0.7520
2.4					0.7324	0.7447	0.7498	0.7424	0.7475	0.7529
2.6					0.7254	0.7396	0.7447	0.7425	0.7476	0.7538
2.705					0.7206	0.7358	***0.7409***	0.7425	0.7475	0.7543
2.8					0.7154	0.7315	0.7366	0.7424	0.7475	0.7547
3					0.7005	0.7187	0.7236	0.7423	0.7472	0.7556
3.2					0.7222	0.7425	0.7474	0.7420	0.7469	0.7564
3.4					0.7439	0.7663	0.7711	0.7415	0.7464	0.7572
3.6					0.7656	0.7900	0.7948	0.7409	0.7457	0.7580
3.8					0.7872	0.8138	0.8186	0.7401	0.7449	0.7588
4					0.8087	0.8375	0.8423	0.7392	0.7439	0.7596
4.459								0.7362	***0.7409***	0.7613
5								0.7310	0.7356	0.7633
6								0.7143	0.7186	0.7665
7								0.7389	0.7431	0.7694
8								0.7550	0.7591	0.7719
9								0.7375	0.7414	0.7741
10								0.8698	0.8736	0.7758
15										0.7780
20										0.7651
23.286										***0.7409***
24										0.7326
26										0.7448
28										0.7483
30										0.7308
32										0.7796
34										0.9032

Aszimmetria a szegények körében

A küszöb alatti eloszlásban nem csak a helyzet és a szóródás szegénységi jellemzők, hanem az eloszlás *aszimmetriája* is. Az alábbiakban az aszimmetria hatást jellemezzük kétféle irányból: i) előbb a szegények eloszlását közelítjük a balra modális gamma-eloszlás módosításával, ii) majd az átlagalattiság hatását vizsgáljuk a Lorenz-görbe aszimmetriájának a mérésével.

A depriváció-averzió gamma-eloszlású becslése

Az eloszlás alakja - *ferdesége, kurtozisa* – szegénységi jellemzők. Zéró közeli modális jövedelem értelemszerűen magasabb szegénységi averziót okoz a küszöb alatt, mint a küszöbhöz közelebbi módusz.

A jövedelmi struktúra matematikai eloszlással történő megadása az információsűrítés eszköze. Specifikálásakor becsüljük az eloszlás paramétereit, jellemezzük a mintához való illeszkedését, és kiválasztjuk a megfelelő függvénytípust. Bemutatjuk, hogy adott „jövedelmi” változó megfelelő transzformálása (esetünkben hatványozása) útján olyan mesterséges változó nyerhető, melynek valamely nevezetes elméleti valószínűség eloszlása - most a gamma eloszlás - alkalmazása révén az eloszlás lefutása (sűrűségfüggvénye) rugalmasan alakítható, így egy empirikus jövedelmi eloszláshoz „jobban” illeszthető. Továbblépési irányok tehát a transzformáció és az elméleti eloszlás alkalmas megválasztása.

A depriváció-averzió becslésére a gamma eloszlást alkalmazzuk, hatvány paraméterrel bővítve.

Legyen az X folytonos véletlen változó sűrűség függvénye $f(.)$, eloszlásfüggvénye pedig $F(.)$. Tekintsük az $Y = X^{1/r}\ \left(r > 0\right)$ változó $g(.)$ sűrűség és $G(.)$ eloszlás függvényét

$$G\left(x\right) = \Pr\left(Y < x\right) = \Pr\left(X < x^r\right) = F\left(x^r\right) \tag{2.15}$$

ahonnan

$$g\left(x\right) = G'\left(x\right) = F'\left(x^r\right) = r \cdot x^{r-1} \cdot f\left(x^r\right) \tag{2.16}$$

Legyen X>0 gamma-eloszlású $\alpha > 0$ alakparaméterrel és $\beta > 0$ helyzetparaméterrel

$$f(x) = \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-\frac{x}{\beta}} \tag{2.17}$$

Ekkor az Y transzformált változó sűrűségfüggvénye az x helyen (behelyettesítés és összevonás után)

$$\boxed{g\left(x\right) = \frac{r}{\beta^\alpha \Gamma\left(\alpha\right)} x^{r\alpha-1} e^{-\frac{x^r}{\beta}}} \tag{2.18}$$

ahol a teljes gamma-függvény

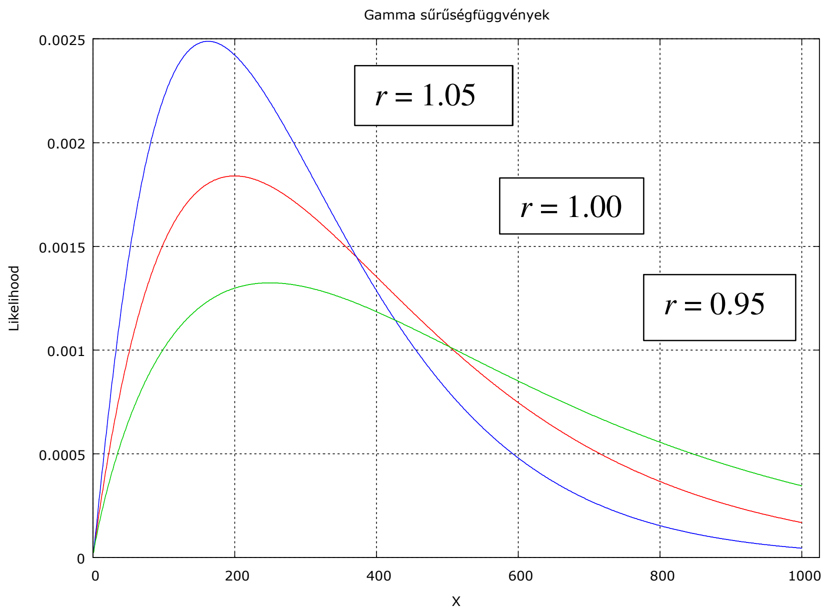
$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx \tag{2.19}$$

és ($\Gamma(\alpha)=(\alpha-1)!=(\alpha-1)\,\Gamma(\alpha-1)$ ha α integer.

A 2.1 ábrán három gamma-eloszlás szerepel, r=1.05, r=1 és r=0.95 kitevőkkel, Alfa=2 és Béta=200 paraméterek rögzítése mellett.

Az r-averzió növekedésével a módusz csökken, a csúcsosság nő, a szóródás pedig csökken. Az r=1 esethez képest az eloszlásfüggvény értéke az x jövedelmi szinten az r>1 vagy r<1 relációnak megfelelően nagyobb, vagy kisebb.

2.1 ábra *Gamma(r)-likelihoodok Alfa=2, Béta=200 mellett*



A momentumok tekintetében az X változó r. momentuma definíció szerint

$$E\left(X^r\right) = \beta^r \frac{\Gamma(\alpha + r)}{\Gamma(\alpha)} \tag{2.20}$$

A hatványparaméter értelmezése érdekében tekintsük a sűrűségfüggvény logaritmusát, az ún. *log-likelihood* kritériumot:

$$\ln L_x = \ln g\left(X\right) = \ln r - \alpha \ln \beta - \ln \Gamma\left(\alpha\right) + \left(r\alpha - 1\right) \ln X - \frac{1}{\beta} X^r \tag{2.21}$$

A log-likelihood konklúziók:

1. Mivel az L likelihood X tekintetében *log-log* modellt, X^r szerint pedig *log-lin* modellt követ, ezért $\left(r\alpha - 1\right)$ a likelihood x szerinti *rugalmasság*a, $100 / \beta$ pedig a *pillanatnyi növekedési ütem* a transzformált X^r tekintetében.
2. A módusz helyzete az X jövedelmi szint tekintetében:

$$\frac{\partial \ln L_x}{\partial X} = 0 \tag{2.22}$$

ahonnan

$$X = \left(\alpha\beta - \frac{\beta}{r}\right)^{1/r} = \left(E(X) - \frac{\beta}{r}\right)^{1/r} \tag{2.23}$$

Láthatóan a módusz értéke r tekintetében *reciprok* modell szerint alakul, a várható érték, és a helyzet paraméter függvényében. Ha r tart a végtelenbe, akkor a módusz a várható értékhez tart.

A paraméterek maximum likelihood (ML) módszerrel való becslése kézenfekvő. Az n-elemű $x_1, x_2, \dots, x_n$ független, véletlen minta együttes, maximálandó likelihoodja

$$L = \prod_{i=1}^{n} g\left(x_i\right) \rightarrow \max \tag{2.24}$$

a feltételes *log*-likelihood pedig *r feltétel mellett*:

$$\ln L = n \ln r - n\alpha \ln \beta - n \ln \Gamma\left(\alpha\right) + \left(r\alpha - 1\right) \sum_{i=1}^{n} \ln x_i - \frac{1}{\beta} \sum_{i=1}^{n} x_i^r \rightarrow \max \tag{2.25}$$

Így a maximum likelihood normálegyenletek:

$$\frac{\partial \ln L}{\partial \alpha} = 0 \Rightarrow r = \frac{\ln\left(\hat{\beta}\right) + \Psi\left(\hat{\alpha}\right)}{\frac{1}{n} \sum_{i=1}^{n} \ln x_i} \tag{2.26}$$

$$\frac{\partial \ln L}{\partial \beta} = 0 \Rightarrow \hat{\alpha}\hat{\beta} = \frac{1}{n} \sum_{i=1}^{n} x_i^r \tag{2.27}$$

ahol a *digamma*-függvény

$$\Psi\left(\alpha\right) = \frac{\partial \ln \Gamma\left(\alpha\right)}{\partial \alpha} = \frac{\Gamma'\left(\alpha\right)}{\Gamma\left(\alpha\right)} \tag{2.28}$$

A feltételes maximálás során az r tekintetében minimális negatív log-likelihood megoldást választjuk.

Lehet a célfüggvény továbbá a gamma-eloszlás *momentumainak* minél pontosabb közelítése a paraméterekkel, miszerint:

$$E\left(x^r \mid g\left(x, \hat{\alpha}, \hat{\beta}\right)\right) = \hat{\beta}^r \frac{\Gamma(\hat{\alpha} + r)}{\Gamma(\hat{\alpha})} \tag{2.29}$$

Ha a momentumok száma nagyobb a paraméterek számánál, akkor a minta momentumokat nem reprodukáljuk, hanem az elméleti momentumot lehető legjobban közelítő koefficienseket kalkulálunk. A paraméterbecslés az *általánosított momentum módszerrel* (GMM) megvalósítható. Mivel két szomszédos momentum hányadosa a momentum paraméter lineáris függvénye - a tengelymetszet a várható jövedelem, a meredekség pedig a helyzetparaméter:

$$R(r) = \frac{E\left(X^{r+1}\right)}{E\left(X^r\right)} = \beta(\alpha + r) = E(X) + \beta r \tag{2.30}$$

Ekkor a *feltételes* GMM pontbecslés *iteratív* meghatározása:

$$\left(\hat{R}(\hat{r}) - \hat{\beta}(\hat{\alpha} + \hat{r})\right)^2 \rightarrow \min \tag{2.31}$$

A Lorenz-görbe aszimmetriája

A Lorenz-görbe aszimmetriáját értelmezzük a szegények körében, és javaslunk három új mutatót a jelenség mérésére. Ha az aszimmetria szegénységet növelő faktor, akkor mértékét célszerű a szegénységi indexben figyelembe venni.

Definíció szerint a *Lorenz*-görbe a legszegényebb x népességi hányadhoz tartozó y jövedelmi részesedést ábrázolja. Az L-görbe a (0,1)-(1,0) szimmetria átlóval való „*M*” metszéspontjában

(*M* nincs föltüntetve az ábrán) a népességet két – „*alsó*” és „*felső*” – szegmensekre bontja. Az alsó szegmens részesedése az összjövedelemből éppen annyi, mint a felső szegmens aránya a népességen belül. Az ábrázolt 2.2 *L*-görbe láthatóan aszimmetrikus a metszéspontra. Kérdés, hogy mi a statisztikai tartalma ennek az aszimmetriának, és milyen mutatóval lehet tömören jellemezni.

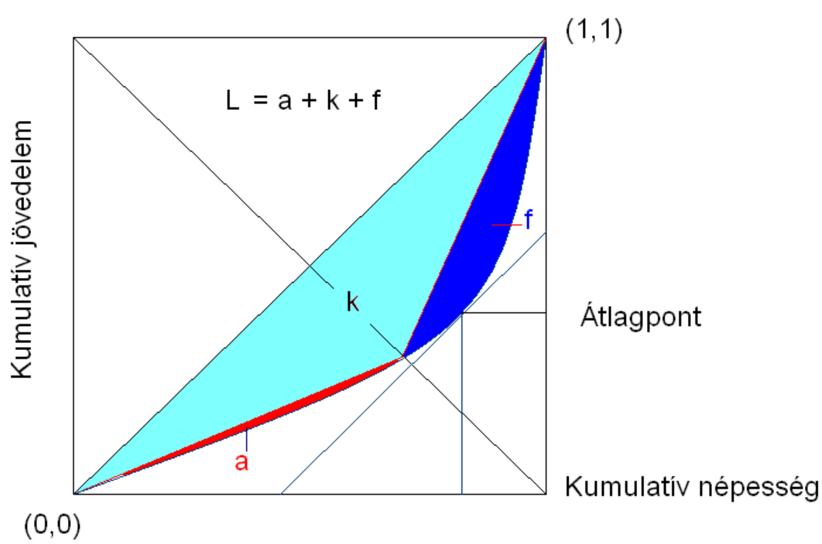
A jelenséget alapvetően az *átlag alatti versus átlag feletti deprivációként* értelmezzük. Részletesen kifejtve:

„M” egy *alsó* és egy *felső* szegmensre bontja a szegényeket, rendre *m* és 1-*m* arányban. Ekkor az egyes szegmensek egyenlőtlenségeit reprezentáló területek:

A küszöb alattiak egyenlőtlenségét a Lorenz-féle $L=a+k+f$ koncentrációs terület méri. Ebből az *alsó* szegmens egyenlőtlensége a „piros *a*” terület, a *felső* szegmens egyenlőtlenségét a „sötétkék *f*” terület méri. A két szegmens közti *külső* egyenlőtlenség fokát a - két szakaszra redukált - *L*-görbe és az átló által bezárt „világoskék *k”* terület méri.

A *Lorenz*-görbe *szimmetriájának* statisztikai tartalmát most az alsó és a felső szegmensen belüli azonos mértékű egyenlőtlenség nyújtja. Tehát az *L*-görbe aszimmetriája a *csoporton belüli egyenlőtlenség* struktúráját jellemzi, a *belső* egyenlőtlenségen is *belül* értelmezve.

2.2 ábra: Aszimmetrikus Lorenz-görbe: az alsó és a felső szegmens egyenlőtlensége



Másfelől, az aszimmetria a *Lorenz*-görbe *átlagpontja* oldaláról is megragadható. Az *L*-görbe nevezetes pontja az átlagpont, melyben a referencia átlóval húzott párhuzamos érinti a görbét. Az átlagpont első „*p*” koordinátája azt mutatja, hogy a népesség hány százalékának kisebb a jövedelme az átlagos jövedelemnél, második koordinátája pedig az általuk birtokolt relatív jövedelmi hányadot jelenti. Mint a 2.2 ábrán látható, aszimmetrikus *L*-görbe esetén a görbe átlagpontja eltolódik a szimmetria átlóról, és egy *másik struktúrában bontja* a társadalmat egy „*szegényebb*” és egy „*gazdagabb*” szegmensre, rendre *p* és (1-*p*) arányban.

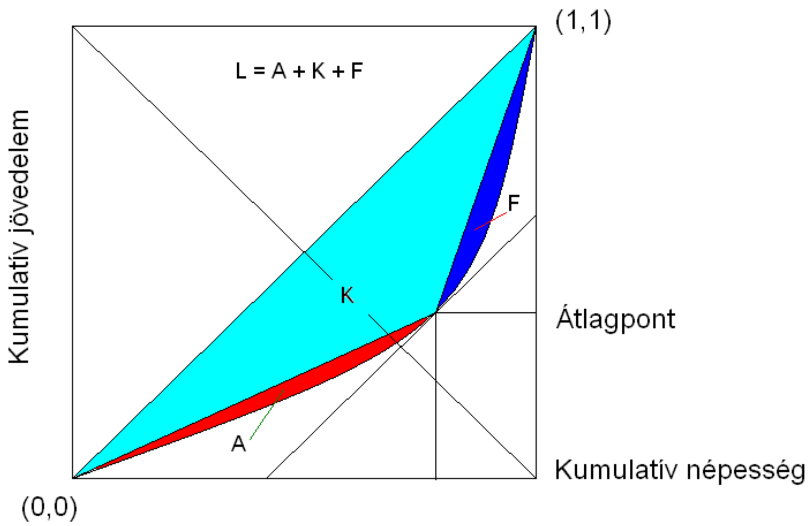
Az átlagpont a totális egyenlőtlenséget is másik struktúrában bontja *három* összetevőre, ha a referencia átló két végpontját a *P* átlagponttal kötjük össze. Ezt a 2.3 ábra szemlélteti (az ábrán a *P* hivatkozás nem szerepel):

A küszöb alatti népesség tényleges egyenlőtlenségének dekompozíciója: $L=A+K+F$. Ebben a *szegényebb* szegmens egyenlőtlenségét a „piros *A*" terület, a *gazdagabb* szegmens egyenlőtlenségét a „sötétkék *F*" terület, a két szegmens közti *külső* egyenlőtlenség fokát pedig most a „türkiz *K*" terület méri.

A Lorenz-görbe szimmetriáját ez esetben a szegényebb és gazdagabb szegmensek közötti, valamint az alsó és felső szegmensek közötti külső egyenlőtlenségek egyezőségében vagy különbözőségében ragadhatjuk meg. A $K>k$ reláció aszimmetrikus L-görbére utal.

<u>*Konklúzió:* Magasabb átlagpont magasabb szegénységi averziót tükröz.</u>

2.3 ábra: Aszimmetrikus Lorenz-görbe: a szegényebb és a gazdagabb szegmens külső egyenlőtlensége



A jelenség méréshez az alábbi ***új*** aszimmetria mértékeket javasoljuk.

A Lorenz-terület aszimmetria mutatók

Az *L*-terület mutatót úgy definiáljuk, hogy zéró értékkel szimmetriát, -1 és 1 értékkel pedig extrém aszimmetriát jelezzen, ha a belső egyenlőtlenség kizárólag az alsó, vagy kizárólag a felső szegmensen belüli egyenlőtlenségnek köszönhető. Mindeközben *abszolút értéke* a (0,1) intervallumra normált. Így a □-mutató:

$$\boxed{-1 \leq \quad \phi = \frac{f-a}{f+a} = \frac{f-a}{L-k} \quad \leq 1} \tag{2.32}$$

Másik aspektusból az átlagpont helyzetét is számításba vesszük. Ennek általunk alkalmazott módját szemlélteti a 2.4 ábra. A 2.4 ábrán olyan társadalom került ábrázolásra, ahol a szegényebbek részaránya alacsonyabb, mint az alsó szegmens részaránya: $p<m$. Az ábrán az árnyékolt területek összege „*a*", melyen belül az „*A*" jellegű terület belső összetevő: $a>A$, míg $f<F$.

Olyan társadalomban, ahol a szegényebb népesség aránya nagyobb, mint az alsó szegmens aránya, a megfelelő területek relációi megfordulnak: itt $a<A$ és $f>F$. Ezek alapján végül az L-terület φ-mutatót az A/a és az F/f törtekre alapozzuk:

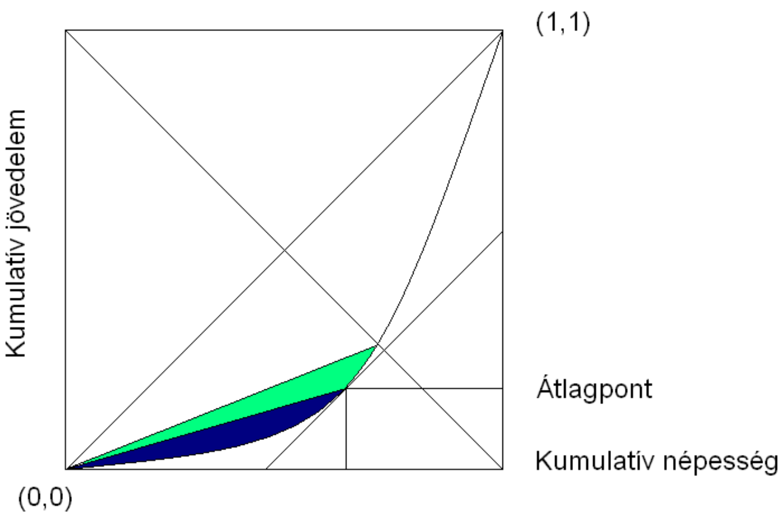
$$\boxed{-1 \leq \quad \varphi = \frac{A/a - F/f}{A/a + F/f} = \frac{fA - aF}{fA + aF} \quad \leq 1} \qquad (2.33)$$

Az A/a viszonyszám azt jellemzi, hogy a szegényebbek körében mért egyenlőtlenség hány százaléka az alsó jövedelmi szegmens egyenlőtlenségének. A tört értéke lehet kisebb és nagyobb is mint 1. Hasonlóan, az F/f viszonyszám azt jellemzi, hogy a gazdagabbak körében mért egyenlőtlenség hány százaléka a felső jövedelmi szegmens egyenlőtlenségének. A tört értéke lehet kisebb és nagyobb is mint 1, de ha $A/a>1$, akkor $F/f<1$, és megfordítva.

A mutató értékkészlete a következő nevezetes helyzeteket veszi fel:

- Ha a *Lorenz*-görbe szimmetrikus az átlóra, akkor $a=A=f=F$ és a mutató értéke zéró.
- Ha $F=A$, akkor a mutató egybeesik a ϕ mutatóval.
- Ha $F<f$, akkor $A>a$ és a mutató előjele pozitív.
- Ha $F>f$, akkor $A<a$ és a mutató előjele negatív.
- Ha $F=0 \mid F<f$, akkor a mutató értéke +1.
- Ha $A=0 \mid A<a$, akkor a mutató értéke -1.
- A mutató *abszolút* értéke nem kisebb, mint 0 és nem nagyobb, mint 1.

2.4 ábra: *Aszimmetrikus Lorenz-görbe: p<m, A<a, f<F*



A marginális hatás aszimmetria mutató

A Lorenz-görbe aszimmetriáját a meredekségén át is mérhetjük. Átló alatti átlagpont mellett a meredekség az átló metszéspontjában M>1, míg az ellenkező esetben M<1.

Tekintsük, és kössük össze az átlót alulról és felülről határoló L_{ower} és U_{pper} pontokat, melyek koordinátái (x_L,y_L) és (x_U,y_U). A differencia hányados e pontok tekintetében:

$$d_{U:L} = \frac{y_U - y_L}{x_U - x_L} \tag{2.34}$$

A differencia hányados jelentése, hogy ha egy ponttal nő a népességarány, akkor hány ponttal emelkedik a jövedelmi részesedése a metszéspont környezetében. Változatlan egyenlőtlenség mellett, c.p. a meredekség egynél magasabb értéke alacsonyabb szegénységet jelez.

A $d_{U:L}$ hányados *odds-típusú* információ, melynek értéke zérótól pozitív végtelenig mehet. A belőle származó aszimmetria mérték megfogalmazása:

$$\boxed{L_d = \frac{d_{U:L}}{1 + d_{U:L}}} \tag{2.35}$$

mely a (0;1) intervallumra normált, és szimmetria esetén az értéke 0.5. A szimmetria esetét az origóba helyezve, a korrigált mutató:

$$\boxed{L_M = \frac{d_{U:L}}{1 + d_{U:L}} - \frac{1}{2}} \tag{2.36}$$

3. Általánosított variancia egyenlőtlenség és szegénység

A fejezet célja egy *új*, GVIP[2] *többváltozós* módszer kidolgozása az *egyenlőtlenség többdimenziós* mérésében, szegmentált társadalomra megadni a csoportközi felbontását, majd a *szegénység* mérésében alkalmazni. GVIP az általánosított variancia mértéken alapul. A dekompozíció megvalósítása a Wilks-lambda alkalmazásához vezet, lehetőséget adva így a numerikus számítások standard statisztikai szoftverrel történő kalkulálására. A GVIP megközelítés a fejezetben definiált új, információ elméleti alapon kidolgozott indexekre épül.

A szegénység mérésében a cenzorált (csonkolt) eloszlásra alkalmazva GVIP mint szegénységi mérték adódik.

A fő mondanivaló tehát több dimenzió – *jövedelem, fogyasztás, vagyon* - tekintetében egyidejűleg mérni *kompozit* módon az egyenlőtlenség fokát, és megadni valamely szegmentáció által magyarázott variancia arányt. Ezzel együtt a szegénységre vonatkozó mérés többdimenziós felbontását is megadjuk.

Az egyenlőtlenséget egydimenziós (jövedelem) esetben is kétváltozós (*d:hozam*, *U:haszon*) varianciaként kezeljük és az általánosított varianciával mérjük.

$$\boxed{\mathbf{C} = \begin{array}{c|cc} \textit{Változó} & d & U \\ \hline d & V_d & T_C \\ U & T_C & V_U \end{array}} \tag{3.1}$$

Az egyenlőtlenség kétváltozós mértéke - definíciónk szerint - az *általánosított variancia*, ami a **C** mátrix determinánsa.

[2] Generalized Variance Inequality and Poverty

A *determináns* - bár most a jövedelmek *egydimenziós* egyenlőtlenségét méri - statisztikai értelemben mégis *kétváltozós* mérőszám. Mint *rés*, alsó és felső korlátai

$$0 \le \det(\mathbf{C}) \le V_d V_U \tag{3.2}$$

A V_d variancia emelkedésével a rés tágul, tehát az egyenlőtlenség mértéke emelkedik. Másképp, ha a hozam és a haszon gyengén korrelál egymással (alacsony a redundancia és vele a veszteség), akkor V_d és V_U minden információja szükséges a szóródás mérésében, ha viszont a redundancia (és vele a veszteség) jelentős, akkor a $V_d V_U$ felső korlát e redundancia mértékében csökkentendő.

Csoporthatás vizsgálata érdekében tekintsük a társadalom g=1,2,…,m csoportra bontását. A $\mathbf{C}$ mátrix csoportközi *külső-belső* dekompozíciója

$$\boxed{\mathbf{C} = \mathbf{C}_B + \mathbf{C}_W} \tag{3.3}$$

ahol $\mathbf{C}_{Between}$ a *külső*, $\mathbf{C}_{Within}$ pedig a *belső, átlagos csoporton belüli* kovariancia mátrix. Innen a *csoportosítás irrelevanciáját* mérő *Wilks-lambda* variancia hányados

$$Wilk's = \frac{\det \mathbf{C}_W}{\det \mathbf{C}} \tag{3.4}$$

melynek komplementere értelemszerűen „*Variance Explained*" típusú mutató

$$VE = 1 - Wilk's \tag{3.5}$$

A jövedelem mellett más dimenziókban is tekintve az egyenlőtlenséget, a $\mathbf{C}$ mátrix rendje a dimenziók száma szerint duplázódik, az általánosított variancia determináns elve változatlan.

A szegénység új mértéke - definíciónk szerint – az egyes dimenziók küszöbei szerint cenzorált eloszlások általánosított varianciája.

4. A relatív deprivációs szegénységi küszöb rétegspecifikus becslése

A *szegénységi küszöb* rögzítésének elterjedt *relatív* módjai a medián értékének bizonyos százalékát, vagy az alsó decilist, kvintilist, kvartilist, stb. adni meg, mint a küszöb értékét. A kvantilisek alkalmazásának indoka és előnye, hogy robusztusak az extrém értékek, „outlierek" tekintetében. Evidencia, hogy eltérő rétegekben a küszöb szintje is eltérő. Ha a módszer medián alapú, kézenfekvő annak értékét a rétegképző ismérvek szintjeivel magyarázni egy regressziós modellben. Ezáltal bármely kombinációval definiált rétegre önálló, specifikus medián-becslést kaphatunk.

Figyeljük emellett, hogy a szegénységi dimenzió akár a jövedelem, akár a fogyasztás, akár a kiadás, akár a vagyon, jellegéből adódóan – háztartásról háztartásra – *heteroszkedasztikusan* alakul. Ezért indokolt nem a medián valamely arányát, hanem egy alkalmas kvantilis réteg-specifikusan regresszált szintjét alkalmazni küszöbként. Ennek megvalósítását szolgálja a *kvantilis* regresszió.

Robusztus módszerként adódik a *medián* - a centrális tendencia robusztus paraméterének - a modellezése. Míg az OLS megközelítés a számtani átlag négyzetes minimum tulajdonságán alapul, a LAD (Least Absolute Deviation) medián regresszió a medián abszolút érték minimum tulajdonságát használja a regressziós koefficiensek robusztus becslése érdekében.

Mikor a függő változó empirikus értékei a LAD regresszióval nem párhuzamosan alakulnak az *X* prediktor változó tekintetében (szétnyílnak, zárulnak, stb.), akkor maga a centrális tendencia modell nem adekvát, és fölmerül az igény a függő változó eloszlásának valamely *Tau-rendű feltételes kvantilisét* prediktálni.

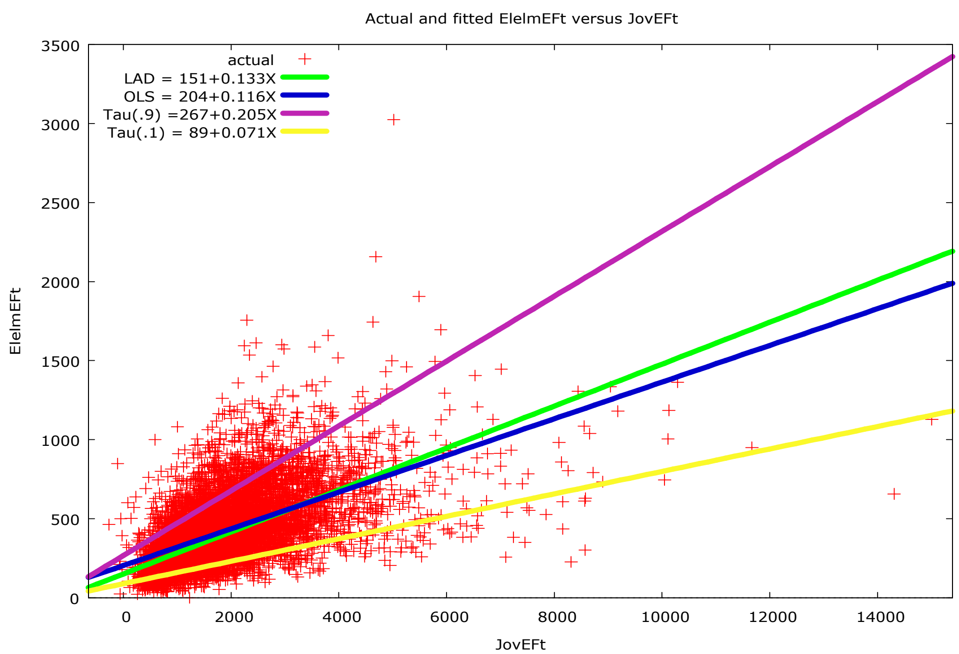
A mondottakat illusztrálandó, Magyar Háztartások (2003., *n*=8314) éves élelmiszer kiadásait ábrázoljuk az éves jövedelmeik (EFt) függvényében.

A diagram 4 regressziós egyenest ábrázol. A pontfelhő jellegzetességei:

1. Outlierek jelennek meg mind Jövedelem, mind Kiadás tekintetében,
2. A Kiadás terjedelme a jövedelmi szint emelkedésével tágul.

Látható, hogy egyetlen regressziós egyenessel nem lehet leírni a pontfelhőt, és ha a közepes kiadást modellezzük, akkor az OLS egyenes alkalmazása nem megfelelő, mert a feltételes átlag érzékeny az outlier értékekre.

A *centrális kiadás* leírására most célszerű a *feltételes mediánt* modellezni, míg a magas és az alacsony kiadások tekintetében a *feltételes felső és alsó decilis* modellezése indokolt.



Módszertanilag a medián LAD regresszió célfüggvénye:

$$\min_{\beta} \sum_{i=1}^{n} | Y_i - \underbrace{\beta X_i}_{Medián|X_i} | \qquad (4.1)$$

általában pedig a *Q*-kvantilis „*Tau*” rendje: 0<*Tau*<1. A *Tau*=0.5 medián esete kiterjeszthető bármilyen más 0<*Tau*<1 kvantilis esetére a *Tau* paraméter megfelelő megválasztásával.

Jelölje *diff* a regresszió eltérését az empirikus értéktől: regresszió fölötti megfigyelés *pozitív diff* értéket, regresszió alatti megfigyelés pedig *negatív diff* értéket eredményez:

$$diff_i = Y_i - \underbrace{Q_\tau \mid X_i}_{=\beta X_i} \tag{4.2}$$

Lévén távolságok összegét minimáljuk, pozitív *diff* értékeknek nagyobb mint 0.5 súlyt adva a regressziós egyenest a LAD egyenes fölé toljuk el, míg negatív *diff* értékeknek nagyobb mint 0.5 súlyt adva a regressziós egyenest a LAD egyenes alá húzzuk el.

Ha a kiadás nem volatilis, akkor az OLS lehet adekvát centrális, de a *nem medián kvantilis* értékek regresszálása ekkor is feladat.

A *Tau*-regresszió lényegében súlyozott „LAD" regresszió alkalmazását jelenti:

$$\sum_{i=1}^{n} \begin{cases} \tau * (diff > 0) \\ (\tau - 1) * (diff \le 0) \end{cases} \to \min \tag{4.3}$$

Az alsó decilis modelljében *Tau*=0.1 míg pl. a felső kvartilis modellezésekor *Tau*=0.75.

A szélesedő pontfelhőt érdemes kvantilisenként regresszálni, megőrizve az eloszlás széleinek az információit is.

A *kiadási határhajlandóságot* vizsgálva (jövedelem koefficiens) a medián LAD becslése 73 Ft. A magyarázó változók körét bővítve - a *specifikációs torzítás csökkentése* miatt - látható, hogy adott *X* prediktor szignifikánsan más eredményt mutat másik rendű regresszióval szemben.

*A becsült koefficiensekkel bármely **réteg deprivációs küszöbszintje** behelyettesítéssel kalkulálható.*

4.1. tábla A kvantilis regresszió becsült koefficiensei (Kimenetek rendre: Bp, Nagyváros, Többi város, Tagszám, Lakásérték, Gépkocsi futás, Üdülő van, Vállalkozók, Aktív keresők, Munkanélküliek, Eltartottak száma, Háztartásfő Neme, Iskolai végzettsége, Kora, a HT Jövedelme.

Quantile estimates, using observations 1-8314								
tau =	0.05	0.1	0.25	0.5	0.75	0.9	0.95	OLS
Coefficient	Dependent variable: ElelmEFt (medián=372.3)							
const	-28.17	-7.97	7.38	36.18	89.35	129.53	171.15	46.39
DBpNvTvKo_1	-22.44	-19.23	-30.52	-27.98	-28.03	-35.94	-14.68	-31.65
DBpNvTvKo_2	-1.27	-2.11	-12.42	-0.01	-1.48	-26.26	-39.10	-8.61
DBpNvTvKo_3	4.50	-1.32	-3.78	-2.53	-6.66	-16.31	-31.12	-7.02
TLetszam	32.99	34.20	44.40	51.85	65.15	79.08	97.66	52.15
LakasMFt	0.09	0.37	0.75	0.69	1.05	1.91	1.50	0.76
GepKoEKm	0.72	0.96	0.86	1.25	1.83	2.95	3.28	1.32
UduloVan	-5.77	-8.48	-2.21	-3.90	-6.39	-27.90	17.78	-1.37
Vallalk	-5.65	-9.50	0.24	-6.60	10.77	24.99	22.24	2.98
AKeres	7.85	7.41	8.30	5.59	-0.20	-7.72	-15.49	4.83
Mnelkuli	-14.13	-18.78	-16.73	-10.86	-18.47	-33.19	-51.01	-17.16
Eltartott	6.05	10.65	9.58	11.71	7.50	-2.52	-14.45	13.29
HFneme	7.64	19.49	24.03	27.43	32.41	31.36	39.14	31.54
HFiskv	2.91	2.80	3.27	3.03	2.46	2.28	1.59	3.79
HFkora	0.81	0.70	0.69	0.58	0.36	0.36	-0.09	0.77
JovEFt	0.035	0.038	0.050	0.073	0.090	0.119	0.137	0.069
Akaike criterion	109397	108308.1	107195.7	107618.8	110252.9	114461.6	117501.1	
Hannan-Quinn	109436	108346.5	107234.1	107657.2	110291.3	114500	117539.5	
Schwarz criterion	109509.8	108420.5	107308.1	107731.2	110365.3	114574.0	117613.5	

5. Egzakt logisztikus regresszió a szegénységi prediktorok szelektálásában

A logisztikus regresszió a klasszifikálás egy alapvető módszere, így alkalmazása *a szegénység mérésében* is kézenfekvő. Mikor a függő változó "Igen/Nem" kimenetű, a bináris regressziót alkalmazzuk. A függő változó eloszlása ismeretében a logisztikus regresszió paramétereinek a becslésére a maximum likelihood módszer adott, de az eljárás kedvező tulajdonságai (minimum variancia, konzisztencia) csak *nagymintás* esetben, aszimptotikusan érvényesek. A rétegzett szegénységi küszöb szerinti klasszifikálás azonban a kismintás, *ritka esemény* esete.

A fejezet célja, hogy a szegénységi kockázat előrejelzése kapcsán a logisztikus regresszió ML becslési problémáira fölhívja a figyelmet és a kezelésre módszertant javasoljon.

A feltétel nélküli maximum likelihood eljárás alkalmazása szempontjából alapvető probléma a *kiegyensúlyozatlan* minta, melyben relatíve alacsony az „Igen=1" esemény aránya, másfelől a *szeparált* minta, melyben az „Igen=1" esemény egyértelműen a magyarázó változó egy adott szegmenséhez, a „Nem=0" esemény pedig a komplementer szegmenshez tartozik. Míg az előbbi esetben van egyedi ML megoldás, de az torzított és magas mintavételi varianciával bír, addig az utóbbi esetben *nem is létezik* a ML.

A ritkaság kezelését az Igen/Nem események permutációin alapuló *egzakt logisztikus* regresszió (ELR) szolgálja. Az ELR eljárás a regressziós paraméterek *elégséges statisztikáinak* az egzakt, feltételes, permutációs eloszlásán alapuló módszere. Ha az aszimptotikus ML becslés nem létezik, az ELR módszer használatával akkor is következtetni tudunk a regressziós paraméterekre.

A fejezet a releváns szegénységi prediktor változók szelektálását tárgyalja, ha kiválasztásuk a *p*-value kritérium alapján történik. A modellépítés során a korrekt *p*-érték kalkulálása kulcskérdés.

Mivel a társadalmi-gazdasági jellemzők kombinációi alapján hasonló rétegű háztartások sokfélesége adódik, adott rétegben a mintavétel során *kicsiny méretű, vagy kiegyensúlyozatlan* csoportok kialakulása reális helyzet.

Ebben az esetben az egzakt következtetés nyújt korrekt p-értéket, és konfidencia intervallumot a kérdéses paraméterekre.[3]

Egy *n* elemű véletlen, $(n \times 1)$ rendű $\mathbf{y} = (y_1, y_2, ..., y_n)'$ mintában

$$t_j = \sum_{i=1}^{n} y_i x_{ij} \quad (j = 1, 2, ..., p) \tag{5.1}$$

a β_j paraméterre vonatkozó *elégséges* (*sufficient*) statisztika.

A mintavételi következtetés három módja áll rendelkezésre:

i) A feltétel nélküli likelihood, ii) a feltételes likelihood, és iii) a feltételes egzakt következtetés.

A $\mathbf{t}_2$ megszorítás kulcsszerep a következtetésben, mert így $\boldsymbol{\beta}_2$ eliminálódik a likelihoodból:

[3] A számítások a LogXact programmal készültek.

$$f_{\beta_1}(\mathbf{t}_1 \mid \mathbf{t}_2) = L(\beta_1 \mid \mathbf{t}_2) = \frac{c(\mathbf{t}_1, \mathbf{t}_2)e^{\beta_1 \mathbf{t}_1}}{\sum_{\mathbf{u}_1} c(\mathbf{u}_1, \mathbf{t}_2)e^{\beta_1 \mathbf{u}_1}} \to \max . \tag{5.2}$$

A fenti feladat maximálása $\boldsymbol{\beta}_1$ szerint adja a “conditional maximum likelihood estimates” (CMLE) becsléseket. Az alapvető különbség az UMLE és a CMLE következtetés között, hogy míg UMLE igényli a $\boldsymbol{\beta}_2$ zavaró paraméter becslését is, addig CMLE kontroll alatt tartja, és csak $\boldsymbol{\beta}_1$ becslésére koncentrál. Hipotézis tesztelésekor a

$$H_0 : \boldsymbol{\beta}_1 = \mathbf{0} \tag{5.3}$$

null-hipotézis tesztelésére a Chi2 *scores*, *likelihood ratio* és *Wald* statisztikák állnak rendelkezésre. Hangsúlyozzuk, hogy a *scores statisztika* nem igényli a *full* modell MLE becslését, csak a restriktív modell becslésén alapul. *Ez azt eredményezi, hogy a scores statisztika létezhet akkor is, mikor a full modell MLE becslése nem létezik.*

Mikor a CMLE módszerrel történik a paraméterek becslése, a fenti teszt statisztikák feltételes alkalmazása kézenfekvő. Ha a megfigyelés rétegzett, és a *rétegspecifikus tengelymetszetek* száma nagy a megfigyelések számához viszonyítva, a meredekség paraméterek MLE becslése inkonzisztens lehet. Ez esetben a $\boldsymbol{\beta}_1$ meredekség paraméterek CMLE becslése (a $\boldsymbol{\beta}_2$ stratum specifikus konstansokat eliminálva) konzisztens.

Kismintás, vagy kiegyensúlyozatlan esetben az *aszimptotikus* Chi2 eloszlás a *scores, likelihood ratio és Wald tesztekre* nem tartható. Ez esetben egzakt következtetéseket adhatunk a vizsgált paraméterekre: egzakt teszteket, és egzakt konfidencia intervallumokat.

Az *R* visszautasítási tartomány megválasztása az egzakt teszt típusának a megválasztásán múlik. Erre három módszert tekintünk.

i) Az *„exact conditional scores” teszt* esetén az *R* régiót a teszt statisztika (négyzetes, standardizált *u*-érték) mindazon értékei alkotják, melyek nagyobb-egyenlők, mint a teszt statisztika megfigyelt értéke. i)) Az *„exact conditional” probability teszt* esetén az *R* régiót a teszt statisztika mindazon értékei alkotják, melyek valószínűsége kisebb-egyenlő, mint a teszt statisztika megfigyelt értékének a valószínűsége. iii) Az *„exact likelihood ratio” teszt* esetén, az *R* régiót a teszt statisztika mindazon értékei alkotják, melyek LR értékei nagyobb-egyenlőek, mint a megfigyelt adat LR értéke. Az *egzakt* teszt véd az *első fajú hibával szemben*.

Tekintsük az exact conditional probability megközelítést egyváltozós esetben ($\mathbf{t}_1 = t_1$ skalár). Megtörténhet, hogy az $f_0(t_1|\mathbf{t}_2)$ feltételes valószínűségi eloszlás többmóduszú, és így *R* nem összefüggő intervallum. Ez nem eset az egzakt scores teszt esetén. A CMLE bizonyos helyzetekben nem létezik, ilyenkor a LR teszt alkalmazása lehetetlenül. A conditional scores teszt nem igényel becsült paramétereket.

Ha a CMLE nem létezik, akkor a „*median unbiased estimate*” (MUE) áll rendelkezésre egzakt feltételes becslésre, a következő egyenlet megoldásával:

$$f_{\hat{\beta}_1}(\mathbf{t}_1 \mid \mathbf{t}_2) = 0.5 . \tag{5.4}$$

Csak nyitott CI adható, mikor vagy $t_1 = t_{1\min}$ vagy $t_1 = t_{1\max}$ mert a kumulatív valószínűség $T_1|\mathbf{t}_2$ teljes terjedelmén mindig 1, és így független β_1 értékétől.

Ha az α szintű konfidencia intervallumot hipotézis tesztelésére használjuk, meg kell adnunk, hogy a döntés milyen *szignifikancia szint* alkalmazásával konzisztens. Jelen módszer biztosítja, hogy az egzakt *p*-value akkor, és csak akkor kisebb, mint α, ha az egzakt $(1-\alpha)$ CI nem tartalmazza a hipotetikus paramétert. E definíció szerint az egzakt kétoldali *p*-value kétszerese az egyoldali *p*-value értéknek: $p_2 = 2p_1$, ahol az egyoldali *p*-value a balszéli és jobbszéli valószínűségek közül a kisebbik.

Empirikus példák[4]

Tekintsük a legalább hattagú budapesti háztartásokat, adott évben.[5] A medián jövedelem 60 százaléka alatti háztartásokat kezeljük szegényként: *Poverty*=1 szegény háztartást jelöl.

Példa 1: A feltétel nélküli aszimptotikus MLE becslés nem létezik

5.1. tábla Paraméterbecslés, mikor az MLE nem létezik

Modell 1	Type	Beta	SE(Beta)	Type	95%CI Lower	95%CI Upper	2*1-sided=p_2
Const	MLE	?	?	Asymptotic	?	?	?
Nem	MLE	?	?	Asymptotic	?	?	?
	MUE	4.481	NA	Exact	2.804	+INF	1.094e-024
Modell 2							
Const	MLE	?	?	Asymptotic	?	?	?
Tartósan beteg	MLE	?	?	Asymptotic	?	?	?
	MUE	-5.29	NA	Exact	-INF	-3.614	5.809e-052
Modell 3							
Const	MLE	-8.522	0.3566	Asymptotic	-9.221	-7.823	2.493e-051
Iskola-score	MLE	0.5927	0.03053	Asymptotic	0.5328	0.6525	2.327e-043
	CMLE	0.5926	0.03053	Exact	0.534	0.6547	3.763e-202
Modell 4							
Const	MLE	?	?	Asymptotic	?	?	?
Iskolai végzettség	MLE	?	?	Asymptotic	?	?	?
	MUE	7.092	NA	Exact	5.418	+INF	6.977e-257

NA: not applicable, ?: does not exist, INF: infinite, e: exponent.

Elsőként a háztartásfő nemét véve mint egyedi prediktor változót (Modell 1), a "Nő" egy perfekt prediktor, így az MLE nem létezik (ezt jelzi a ? jel) miközben az MUE pontbecslés és az egyoldali CI elérhető. CI felső határa +INF, mert a zéró gyakoriság megjelenik a *Nem* terjedelmének *alsó* extrém értékénél, vagyis a Nőknél, mikor *Nem*=0 az 5.2 táblában.

Szemben ezzel, tekintsünk egy másik bináris prediktort, nevezetesen, hogy van-e tartósan beteg a háztartásban: "1:van", "0: nincs" (Modell 2). A konklúziók hasonlóak a fentiekhez azon kivétellel, hogy CI alsó határa (–INF), mivel a zéró frekvencia az 5.2 táblában megjelenik a *tartósan beteg jelenlét* terjedelmének *felső* extrém értékénél.

Kategóriák összevonása is befolyásolhatja az MLE létezését. Tekintsük ugyanis a háztartásfő iskolai végzettségét mint egyedi prediktort (Modell 3).[6] Látható, hogy mind az MLE mind a CMLE létezik, a tény ellenére, hogy zéró gyakoriságok csak az eloszlás alsó szélén jelennek meg

[4] A számítások a LogXact programmal készültek.
[5] KSH, Háztartási Költségvetési Felvétel, 2003.
[6] Az iskolai végzettség score (kód) teljes terjedelme: [1,2,...,13] ahol 13 PhD fokozatot jelöl.

az 5.2 táblában. Azonban, összevonva a végzettség szinteket három kategóriába (lásd 5.2 tábla) az MLE már nem létezik, ahogy ez az 5.1 táblában a Modell 4 alatt látható.

Példa 2: Az egzakt és aszimptotikus p-értékek különbözősége

A relatíve magas mintaméret ellenére – a minta kiegyensúlyozatlan volta (a szegény/nem szegény arány 642/6895) miatt – várható lenne, hogy az aszimptotikus és az egzakt *p*-értékek jelentősen különböznek.

Vegyük a *munkanélküli személyek számát* a háztartásban mint egyedüli prediktort (Modell 5). Az 5.3 tábla szerint esetünkben ez nem történik meg, mert a munkanélküliek száma minden szokásos szinten szignifikáns, és a pont és intervallum becslések értékei teljesen hasonlók.

Az *eltartott személyek száma* prediktort tekintve azonban az 5.3 táblában (Modell 6) mutatja, hogy az egzakt *p*-value jelentősen különbözhet a feltétel nélküli megfelelőjétől. Bár az eltartottak száma példánkban semmilyen megszokott szinten nem releváns, de extrém kritikus szintet alkalmazva a két módszer eltérő konklúzióra vezetne. E jelenség esetlegesen bármely prediktor esetén előállhat, a vizsgált réteg függvényében.

5.2 tábla Paraméterbecslés, mikor az MLE létezik

Modell 5	*Type*	*Beta*	*SE(Beta)*	*Type*	*95%CI Lower*	*95%CI Upper*	*2*1-sided=p_2*
Const	MLE	-2.642	0.04741	Asymptotic	-2.735	-2.549	3.92e-085
Munkanélküliek	MLE	1.491	0.07773	Asymptotic	1.339	1.644	6.443e-043
	CMLE	1.491	0.07772	Exact	1.336	1.647	3.333e-073
Modell 6							
Const	MLE	-1.459	0.4144	Asymptotic	-2.271	-0.6464	0.000432
Eltartottak	MLE	-0.0877	0.1012	Asymptotic	-0.286	0.1106	0.386
	CMLE	-0.0876	0.1011	Exact	-0.2912	0.1158	0.4143

Példa 3: A rétegspecifikus tengelymetszetek kiszűrése

Elemezzük újra a *munkanélküliek száma a háztartásban* prediktor hatását, de most úgy, hogy a háztartás gazdasági aktivitását – mint rétegképző változót – kontroll alatt tartjuk (Modell 7 az 5.4 táblában). Számos réteg képezhető a munkanélküliek számának és a háztartásfő gazdasági aktivitásának a kombinálásával. Kiemelendő, hogy az alkalmazott rétegzés után MLE nem adható, de az egzakt MUE létezik, és az egzakt *p*-érték az 5.4 táblában mutatja, hogy a "*Munkanélküliek száma*" továbbra is szignifikáns bármely szokásos szinten. Figyeljük meg, hogy mind a tengelymetszet, mind a réteg specifikus konstansok eliminálódtak a becslésből.

5.3 tábla Rétegzés a háztartásfő gazdasági aktivitása szerint

Modell 7	*Type*	*Beta*	*SE(Beta)*	*Type*	*95%CI Lower*	*95%CI Upper*	*2*1-sided=p_2*
Munkanélküliek	MLE	?	?	Asymptotic	?	?	?
	MUE	2.868	NA	Exact	2.023	+INF	9.471e-050
Modell 8							
Iskola-Score	MLE	-0.2139	0.09588	Asymptotic	-0.4018	-0.02596	0.0257
	CMLE	-0.2139	0.09588	Exact	-0.4065	-0.02154	0.02889

Végül az 5.4 tábla újra tekinti a háztartásfő iskolai végzettségének 13 fokozatú változóját, de most a rétegzett módon. Bár mind az MLE mind a CMLE létezik, de a prediktor 2% szinten már nem szignifikáns, sőt, a koefficiensek előjelei is megváltoztak. A tengelymetszet és a specifikus konstansok most is eliminálódtak a becslésből.

Példa 4: Ellentmondó teszt eredmények

Az eddigiekben csak a 2*1-sided típusú *p*-value került alkalmazásra, a döntési konzisztenciát biztosítandó a 95% CI határokkal. Azonban az egzakt *p*-érték változik a teszt statisztika speciális *scores*, *likelihood ratio* vagy *Wald* választásától függően is. Különösen akkor, ha a mintaméret extrém alacsony. Az alábbiakban ezt a problémát illusztráljuk.

Két prediktorra vonatkozóan az egzakt tesztek eredményeit az 5.6 tábla közli. Előbb a háztartásfő életkora szerepel, majd a válasz arra a kérdésre, hogy a háztartás korábban valaha elszenvedett-e szegénységet. A mintát leszűkítettük a *6 főnél több tagú, budapesti, férfi háztartásfős* háztartásokra.

Az 5.6 tábla mutatja, hogy az *életkor (Age) esetén* csak a score teszt létezik az aszimptotikus tesztek között, de a *p*-értéke 5% döntési szinten más döntésre vezet. Bár az egzakt teszt *p*-értékek most speciálisan azonosak (*p*=0.07143 egyaránt) ez általában nem szükségszerű. Míg a *p-mid* value az Exact Likelihood Ratio teszt esetén 5% szinten a null hipotézist elutasítja, addig a többi egzakt teszt elfogadja azt.

A “*Poverty Ever Before?*” kérdés esetén 5% döntési szinten az *Exact Probability teszt* döntése különbözik a többi típusú egzakt tesztétől, és mind a *p*-value mind a *p-mid* value értékek lényegesen eltérnek.

5.4 tábla *Egzakt teszt eredmények*

A teszt típusa	*Statistics*	*DF*	*p-value*	*p-mid*
	H_0: Beta_Age=0			
Score	4.317	1	0.03774	NA
Likelihood Ratio	?	?	?	?
Wald	?	?	?	?
Exact Score_asy	4.317	NA	0.07143	0.05357
Exact Score	3.777	NA	0.07143	0.05357
Exact Probability	0.03571	NA	0.07143	0.05357
Exact Likelihood Ratio	8.997	NA	0.07143	0.03571
	H_0: Beta_Poverty Ever Before=0			
Score	6.107	1	0.01347	NA
Likelihood Ratio	?	?	?	?
Wald	?	?	?	?
Exact Score	5.343	NA	0.03571	0.01786
Exact Probability	0.03571	NA	0.07143	0.05357
Exact Likelihood Ratio	8.997	NA	0.03571	0

6. A szegénységmérés SEM modelljei

Az alábbiakban - három alfejezetbe foglalva - a SEM[7] módszertant három problémakör elemzésére használjuk a szegénységmérés vonatkozásában.

A szegénység-depriváció-kirekesztés strukturális rendszere

A fejezet célja többváltozós statisztikai megközelítésben a szegénység kapcsolatát a relatív depriváltság és a társadalmi kirekesztés – mint latens változók – tükrében jellemezni. A cél a hipotetikus kapcsolatok előjelének és mértékének a becslése, és tesztelése. A konklúziók nem igénylik szegénységi küszöb definiálását, rögzítését, a szegények körének egzakt klasszifikálását. Az illusztratív célú számítások megfigyelési egységei háztartások, az eredmények a 2003. évi háztartási költségvetési felvétel (HKF) adataira épülnek.

Koncepciónk szerint a *szegénység-depriváltság-kirekesztettség* társadalmi jelenségek *latens*, közvetlenül nem megfigyelhető *endogén* változók egy strukturális körét alkotják, melyekre *exogén* tényezőként mind *latens*, mind *manifeszt* jellegű változók hatást gyakorolhatnak. E kapcsolatok hipotézisként kezelendők, melyeket a 6.1 ábra kapcsolásai – mint becsülendő paraméterek - reprezentálnak. Cél e paraméterek *becslése*, majd *tesztelése*. Ennek alapja a manifeszt típusú változók megfigyelése saját empirikus skálájukon, melyek egymással, és a latens kör tagjaival is kapcsolatban lehetnek. A manifeszt változók körében is megengedett az exogén és az endogén szerepkör.

Latens változó által okozott manifeszt változóra fenntartott az *indikátor* terminológia, mivel modell szerint a megfelelő latens változóban történt elmozdulás hatásaként jelentkezik a megfigyelhető változásuk. Az ábrákon *ovális* keret latens változót, míg *box* manifeszt változót jelöl. Irányított *nyíl* regressziós kapcsolatot, nem irányított *ív* pedig korrelációs kapcsolatot reprezentál. Endogén változóhoz – akár latens akár manifeszt – tartozik egy rámutató, *reziduális*, értelemszerűen latens változó, melyet mindig egy *üres-indítású*, egyirányú nyíl reprezentál.

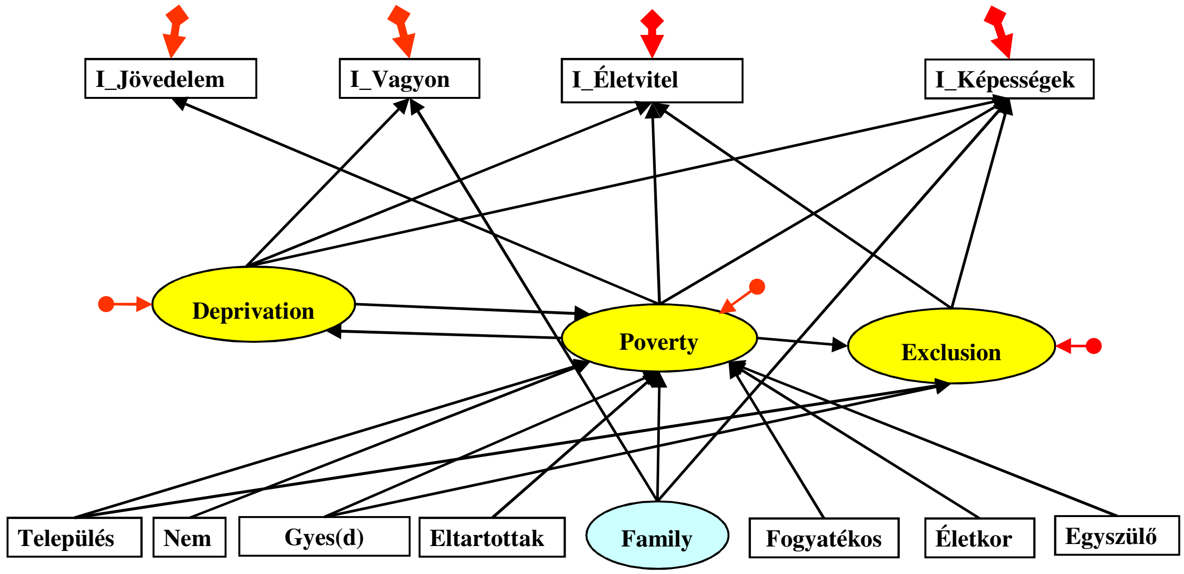
A reziduális változókat kivéve, valamennyi más változó a mérési skála tekintetében lehet *standardizált* is, ilyenkor a hozzájuk kapcsolódó *regressziós koefficiensek* standardizált jellegűek.

Az indikátorok hozzárendelése a látens változóhoz koncepcionálisan történik, de adott manifeszt változó több latens faktornak is lehet az indikátora. Minden latens változóból megengedünk nyilat minden indikátor irányában, de a hipotézis szerinti zéró nyilakat eleve elhagyjuk az ábrából.

“I_*argumentum*” típusú box a rámutató nyíllal azon latens változó indikátorait csoportosítja, amely latens változóból a nyíl indul. Adott indikátor egyidejűleg több I_box eleme is lehet.

[7] Structural Equation Modeling.

6.1. ábra *Induló koncepció: "Multiple Indicator Multiple Cause (MIMIC)" modell*



A modell latens blokkját tekintve, az alábbi konstrukciókat (faktorokat) definiáljuk. Az endogén változók:

1. *Poverty*: a háztartás szegénységben él, valamilyen változó tekintetében,
2. *Deprivation*: a háztartás szegénységben érzi magát a környezete tekintetében,
3. *Exclusion*: a háztartás nem vehet részt valamely gazdasági-társadalmi funkcionalitásban.

Az egyedüli exogén latens változó: „*Family*", a háztartás mögötti családi háttér.

Hipotézisünk szerint a szegénység és a depriváció kölcsönösen közvetlenül hat egymásra, de a kirekesztésre közvetlenül csak a szegénység bír hatással, bár a szegénységen át közvetetten a depriváció és a családi háttér is alakítja.

A manifeszt változók tekintetében csak az exogén változókat emeljük ki, speciális szerepük miatt:

> A település típusa, ahol a háztartás él, A háztartásfő neme, Kap-e a háztartás gyermek gondozási hozzájárulást, A25 évesnél fiatalabb eltartottak száma a háztartásban, Van-e fogyatékos személy a háztartásban, A háztartásfő életkori osztálya, Egyedüli szülő egy vagy több gyermekkel.

A háztartás településtípusa, vagy például a háztartásfő neme a modellből nem levezethető exogén jellemzők, viszont a latens konstrukciókat befolyásolják.

A *valamennyi előforduló* változó strukturális modellje általánosságban az alábbi formát ölti:

$$\mathbf{v}=\begin{bmatrix}\mathbf{y}_L\\ \mathbf{y}_M\\ \mathbf{x}_M\end{bmatrix}=\begin{bmatrix}\mathbf{B}_{LL} & \mathbf{B}_{LM} & \mathbf{0}\\ \mathbf{B}_{ML} & \mathbf{B}_{MM} & \mathbf{0}\\ \mathbf{0} & \mathbf{0} & \mathbf{0}\end{bmatrix}\times\begin{bmatrix}\mathbf{y}_L\\ \mathbf{y}_M\\ \mathbf{x}_M\end{bmatrix}+\begin{bmatrix}\mathbf{G}_{LM} & \mathbf{G}_{LL}\\ \mathbf{G}_{MM} & \mathbf{G}_{ML}\\ \mathbf{I}_{MM} & \mathbf{0}\end{bmatrix}\times\begin{bmatrix}\mathbf{x}_M\\ \mathbf{x}_L\end{bmatrix} \qquad (6.1)$$

ahol ($\mathbf{x}$,$\mathbf{y}$) rendre az *exogén* és *endogén* változók vektorát jelöli, (L,M) rendre *latens* vagy *manifeszt* változóra utal, **B** és **G** együttható-mátrixok, míg végül **I** az egységmátrix.

Következő lépésként a p-számú manifeszt változó szűrése és extrahálása a feladat az összes **v** változó közül - a mintabeli kovarianciáik modellezése érdekében – olymódon, hogy a manifeszt változók $\mathbf{m}_p$ vektorát csak az exogén változók függvényében fejezzük ki.

A modell valamennyi (becsülendő) paraméterét a $\boldsymbol{\theta}_q$ vektorba foglaljuk és úgy becsüljük, hogy minél közelebb legyen egymáshoz a megfigyelt mintabeli **S,** és a becsült paraméterek felhasználásával számított $\hat{\boldsymbol{\Sigma}} = \boldsymbol{\Sigma}\left(\hat{\boldsymbol{\theta}}\right)$ kovariancia mátrix. A q-számú szabad paraméter becslésére $p(p+1)/2$ számú nemlineáris egyenlet áll rendelkezésre:

$$Cov\left(m_j, m_t\right) = f_{jt}\left(\theta_1, \theta_2, \ldots, \theta_q\right) \tag{6.2}$$

Az identifikálhatósági követelménynek való megfelelést sok indikátor szerepeltetésével, másrészt a paraméterekre vonatkozó megszorítások számának a növelésével érhetjük el. A latens változók paraméterének megkötésekor tekintettel kell lennünk arra is, hogy a latens változó nem bír mértékegységgel, ezért skálát kell neki adni a paraméterbecslés során. Ha egy konstrukció esetében mind a regressziós koefficienseit, mind a varianciáját szabadon hagyjuk becsülni, akkor skálája meghatározatlan. Ezért vagy a varianciájára, vagy egyik indikátorának a koefficiensére megszorítást kell tenni. Ha valamely koefficienst 1 értéken rögzítünk, akkor a latens változónak a vonatkozó indikátor skáláját kölcsönözzük. Exogén latens változóknál szokás a loading koefficienst rögzíteni, endogén latens változónál pedig a strukturális koefficienst. Skála megadásának másik szokásos módja a változót egységnyi varianciához standardizálni. Az exogén latens változók varianciái paraméterek, ezek rögzítése 1 értéken kézenfekvő.
Az endogén latens változók (ko)varianciái nem paraméterek, értékük a modellből levezetett, közvetlenül nem rögzíthető. Érdemes tehát a standardizált paraméterek becslésére *iteratív* algoritmust alkalmazni.

Az *aszimptotikusan eloszlásfüggetlen* (ADF) esztimátorok alkalmazását az teszi szükségessé, hogy a háztartásokat jellemző indikátorok (jövedelem, demográfiai jellemzők, iskolázottság foka, stb.) nem normális eloszlásúak. A minimalizálandó diszkrepancia függvény a következő.

Legyen a sokasági kovariancia mátrix torzítatlan becslése az N elemű független, véletlen mintában **S**, amely kovariancia mátrix egyébként hipotézisünk szerint a $\boldsymbol{\theta}$ paraméter vektor valamely függvénye:

$$\boldsymbol{\Sigma} = \boldsymbol{\Sigma}(\boldsymbol{\theta}).$$

Az ADF súlyozott legkisebb négyzetek módszere (weighted least squares: WLS) a mintabeli **S** és a reprodukált $\hat{\boldsymbol{\Sigma}} = \boldsymbol{\Sigma}\left(\hat{\boldsymbol{\theta}}\right)$ kovariancia mátrixok különbözőségét mérő „*Fitting-Function*” diszkrepancia függvényt minimálja:

$$F(\mathrm{s}, \boldsymbol{\sigma}(\boldsymbol{\theta})) = (\mathrm{s} - \boldsymbol{\sigma}(\boldsymbol{\theta}))^T \mathbf{W}^{-1} (\mathrm{s} - \boldsymbol{\sigma}(\boldsymbol{\theta})) \rightarrow \min$$

ahol **s** és $\boldsymbol{\sigma}(\boldsymbol{\theta})$ $p^* = p(p+1)/2$ elemű vektorok, melyek rendre **S** majd $\boldsymbol{\Sigma}(\boldsymbol{\theta})$, nem-duplikatív elemeiből épülnek fel, **W** pedig (p^*, p^*) rendű pozitív definit súlymátrix. Az optimális súlymátrix a mintabeli kovarianciák mintavételi kovariancia mátrixa.

A *szegénység-depriváció-kirekesztés* példát folytatva, HKF adatokon, magyar háztartásokat jellemzően, a becslési eredményeket a 6.1 tábla közli. Mivel az indikátorok nem normalitása most adottság, ezért az ADF módszert alkalmazzuk, normalitásvizsgálat nélkül. Az eredmények alapján az *inszignifikáns* kapcsolat a *latens strukturális részben*: “Poverty” regressed on “Family background”,

6.1 tábla *Paraméterbecslések*

Paraméter	Koefficiens	St. hiba	*t*	Probability
A strukturális rész				
(POVERTY)-1->(DEPRIVATION)	–15.441	0.034	–452.015	0.000
(POVERTY)-2->(EXCLUSION)	–0.866	0.036	–23.870	0.000
(DEPRIVATION)-3->(POVERTY)	1.493	0.073	20.442	0.000
(FAMILY)-4->(POVERTY)	–0.071	0.044	–1.622	*0.105*

A latens strukturális részből kiemelt konklúziók: i) Ha valaki szegény, akkor következésképpen deprivált, és megfordítva, ii) Ha valaki szegény, akkor valamiből kirekesztett, iii) A családi háttér nincs befolyással a háztartás szegénységi statusára.

A 6.1 ábra kiinduló modellje szubjektív jellegű, hipotetikus, tehát vitatható. A modell módosítását, vagy megtartását illeszkedésvizsgálat támasztja alá.

A modellszelekció során a tárgyi és az alternatív *nested* modellek távolságát ítéljük meg. A modellszámítási eredmények:

> A minta méret: $N = 3571$, a manifeszt változók száma: $p = 21$, a szabadon becsülhető paraméterek száma: $q = 56$. A null modell goodness-of-fit Chi-square statisztikája 37577.82 df=20*21/2 = 210 szabadsági fokkal, a céfüggvény konvergált értéke: F = 5. A goodness-of-fit Chi-square távolság a tárgyi modellre: GoF_Chi2 = 17850 df = 231-56-3 = 172 szabadsági fokkal és 0.000 tail probability értékkel.

Az illeszkedés megítélésére a heurisztikus módszerek eredményeit adjuk meg. A Bentler-comparative-fit-index értékét kiragadva 52.69%, ami a modell leegyszerűsített voltát is figyelve, megfelelő illeszkedést mutat. A további heurisztikus jellegű illeszkedésvizsgálati indexek számított értékeit a 6.2 tábla közli (értelmezésüket lásd Hajdu:2003b).

6.2 tábla Heurisztikus goodness-of-fit indexek értékei

Index	Index formula
Population non-centrality index *	$NCI = \frac{\chi^2 - df}{N-1} = 4.9504$
Steiger–Lind root mean square error *	$RMSE = \sqrt{\frac{1}{df}\max\{NCI, 0\}} = 0.1697$
McDonald non-centrality index	$MDNI = \exp\{-0.5\max(NCI, 0)\} = 0.0841$
Population gamma index	$\Gamma_1 = \frac{p}{2NCI + p} = 0.6796$
Adjusted population gamma index	$\Gamma_2 = 1 - \frac{p(p+1)}{2df}(1-\Gamma_1) = 0.3621$
Jöreskog–Sörbom GFI	$GFI = 1 - \frac{2F}{tr\left(\left[\mathbf{S}\hat{\Sigma}^{-1}\right]^2\right)} = 0.5240$
Adjusted Jöreskog–Sörbom	$AGFI = 1 - \frac{p(p+1)}{2df}(1-GFI) = 0.3621$
Akaike information criterion *	$AC = F + \frac{2q}{N-1} = 5.0314$
Schwarz's Bayesian criterion *	$SC = F + \frac{q\ln(N)}{N-1} = 5.1283$

Browne–Cudeck cross validation index *	$CV = F + \frac{2q}{N-p-2} = 5.0316$
Bentler–Bonett, Tucker–Lewis non-normed fit index	$NNFI_{t/b} = 1 - \frac{df_b}{df_t}\frac{\chi_t^2 - df_t}{\chi_b^2 - df_b} = 0.4224$
Bentler comparative fit index	$BCFI_{t/b} = 1 - \frac{\chi_t^2 - df_t}{\chi_b^2 - df_b} = 0.5269$
Bollen's Rho	$\rho_{t/b} = 1 - \frac{df_b}{df_t}\frac{\chi_t^2}{\chi_b^2} = 0.4200$

* A csillaggal jelzett indexek a preferált modellt minimált értékükkel szelektálják.
Note. Mintaméret = N; p = a manifeszt változók száma; q = a szabad paraméterek száma. A t index a komplexebb tárgyi modellt, b pedig a *baseline modellt* jelzi; $F = \chi^2/(N-1)$ a "*fitting function*" célfüggvény konvergált értéke.

A multitrait - multimethod hipotézis tesztelése

A „multitrait-multimethod" (MTMM) modell akkor van érvényben, mikor latens *trait* változók mindegyikét különböző *method* módszerek mindegyikével mérjük. A „multitrait-multimethod" módszer a SEM modell egy speciális felírása. Ha ez a hipotetikus keret tartható, akkor a faktorok kétféle bontásban szeparálják az indikátorokat: egyfelől „*trait*", másfelől „*method*" faktorhoz rendelve azokat. Az érdeklődés általában a *trait* faktorokat övezi, míg a *method* faktorok nyújtják az alapját a változók közötti korrelációs rendszernek.

Az MTMM modell illeszkedése lehetőséget ad az indikátorok *konvergencia-diszkriminancia* tesztelésére, hogy megbízható mértékei-e az adott jelenségnek (*reliability*), az adott módszer érvényessége (*validity*) tekintetében.

A megbízhatóság nem más, mint két eljárás eredményei közötti megfelelés ugyanazon jelenség mérését célozva, maximálisan hasonló mérési módszerek alkalmazása mellett, míg az érvényességet a maximálisan különböző módszerek alkalmazása melletti megfelelés támasztja alá.

Egy 2-trait-2-method modellt reprezentál a 6.3 tábla korrelációs mátrixa.

6.3 tábla "MTMM" mátrix a 2-trait_2-method modell esetén

Method		M1		M2	
	Trait	T1	T2	T1	T2
M1	T1	1			
	T2	WMCT	1		
M2	T1	WTCM	CTCM	1	
	T2	CTCM	WTCM	WMCT	1

A nemdiagonális elemek három csoportba sorolhatók, nevezetesen: i) *within-method, cross-trait* (WMCT) korrelációk, ii) *within trait, cross-method* (WTCM) korrelációk, iii) *cross-trait, cross-method* (CTCM) korrelációk.

Statisztikailag szignifikáns zérótávoli *within-trait-cross-method* (WTCM) korrelációk az ún. *konvergencia validitást* támasztják alá, vagyis különböző módszerek egyetértését ugyanazon jelenség mérését illetően. Zéró közeli korrelációk máshol viszont a *diszkriminancia validitást* jelzik, miszerint a különböző jelenségek valóban megkülönböztethetők. A diszkriminancia-validitás egy további követelménye, hogy az *inter-trait* korrelációk szerkezete egyezzen meg tekintet nélkül arra, hogy adott indikátor melyik módszerből származik.

Az MTMM modell tesztelését kézenfekvő konfirmatív CFA faktormodellel végezni. Az alábbiakban az ún. *teljes* (*complete) modell* eredményeit mutatjuk be (lásd 6.4 és 6.5 ábrák).

A *complete* modell CFA specifikációjában minden egyes indikátor trait, method és unique faktorok függvénye. Legyen a modellben három, T1, T2, T3 trait, és három M1, M2, M3 method faktor, rendre I1,I2,…,I9 indikátort definiálva, U1,U2,…,U9 unique-faktoraikkal.

A *complete* modellben bármely indikátor a saját *trait*-faktor függvénye, és rögzített módszert használva, az adott *method* faktor függvénye. Ha egy indikátorra egy módszer nincs hatással, akkor az indikátorban csak a trait faktor szerepel nagy súllyal, a method faktor nem. Ha egy módszer hatása a mérésben jelentős, akkor valamennyi indikátorban nagy súllyal szerepel mind a vonatkozó trait, mind a vonatkozó method faktor.

A modell lehetővé teszi a trait és a method hatás mérését az által, hogy megengedi a trait faktorok egymás közti, a method faktorok egymás közötti, valamint az egyedi unique faktorok egymás közötti korreláltságát. Így tesztelhető a konvergencia, és a diszkriminancia validitás.

Modell vizsgálatunkban három trait faktort tekintünk, rendre: J: Jövedelemtermelés, K: Kiadási hajlandóság, F: Fogyasztási színvonal.

A method faktor szerepű módszerek a háztartás méretének különféle megjelenési formái. Három method faktort használva, a jelölések: M1: Méret 1, M2: Méret 2, M3: Méret 3.

A háztartás méretét *vetítési alapként* használva, az indikátorokat a T/M típusú – most szám szerint kilenc - T1/M1, T2/M1,…,T3/M3 viszonyszám definiálja. Az indikátorok korrelációs mátrixát a 6.4 tábla közli.

6.4 tábla Az MTMM indikátorok korrelációs mátrixa

Indikátor	**J/M1**	**K/M1**	**F/M1**	**J/M2**	**K/M2**	**F/M2**	**J/M3**	**K/M3**
1: J/M1	1							
2: K/M1	0.406	1						
3: F/M1	0.371	***0.689***	1					
4: J/M2	***0.641***	0.218	0.281	1				
5: K/M2	0.601	*0.405*	0.351	0.897	1			
6: F/M2	0.538	0.243	*0.371*	0.828	0.894	1		
7: J/M3	*-0.050*	-0.043	-0.121	*0.128*	0.066	-0.017	1	
8: K/M3	-0.100	*0.016*	-0.101	0.051	*0.072*	-0.017	0.971	1
9: F/M3	-0.155	-0.054	*-0.116*	-0.028	-0.022	*-0.033*	0.952	0.981

A WMCT korrelációra példa a 0.689 érték az F/M1 sorban az M1 módszer alapján, a WTCM korrelációra pedig a 0.641 érték a J/M2 sorban, a jövedelmet M1 és M2 bázisában véve. A korrelációs mátrix *distinct* elemeinek a száma 45 nemlineáris egyenlet felírását teszi lehetővé.

A paraméterek nagy száma miatt, identifikálhatósági szempontból elengedhetetlen bizonyos megszorításokat tenni a paraméterekre, aminek a mikéntje a teljes modell speciális eseteihez vezet.

A *complete* MTMM modell paramétereinek becslésére kétféle megközelítést alkalmazunk, melyek az alábbiak:

Included Method Factors with Uncorrelated Uniqueness (lásd:IMF_UU, 6.2 ábra)

A modell mind *korrelált trait*, mind *korrelált method* faktorokat feltételez, viszont nem enged meg korrelációkat trait és method faktor viszonylatban, és az egyedi faktorok körében sem. Ugyanakkor az egyedi faktorok varianciái becslésre kerülő paraméterek.

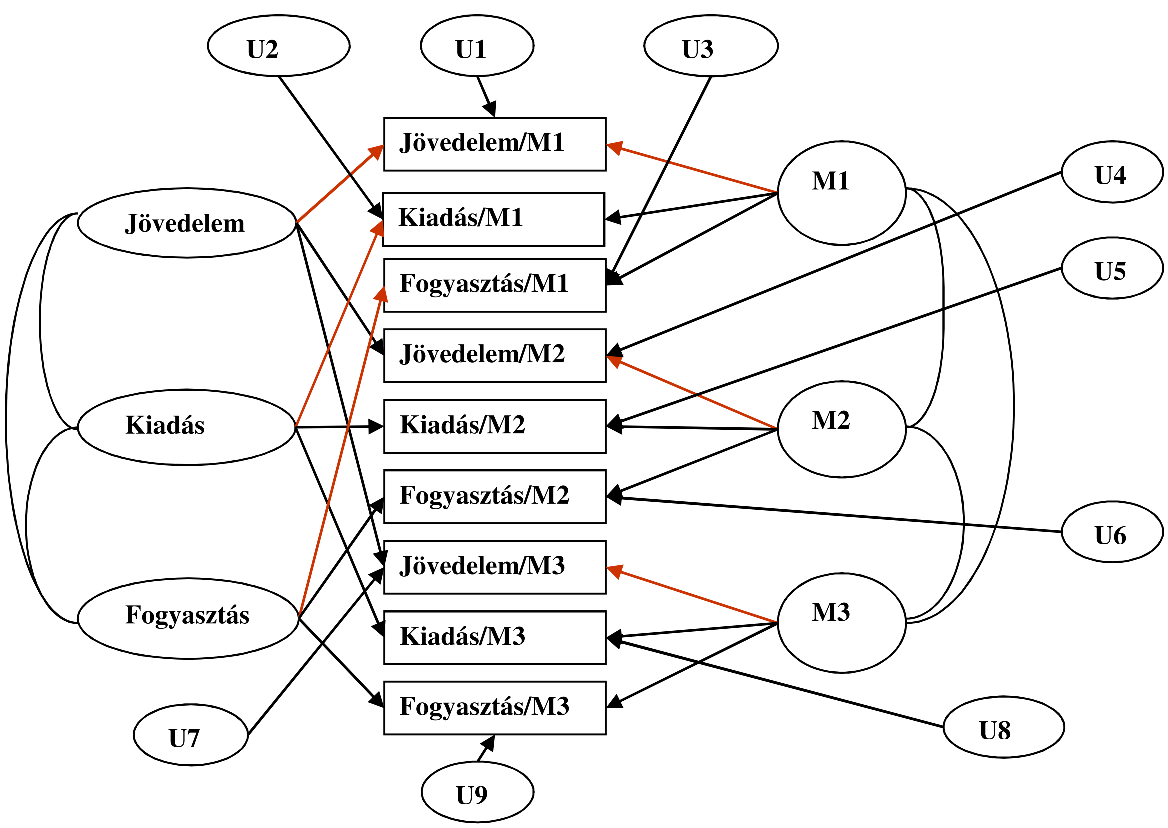
Jól illeszkedő modell mellett a konvergenciát a trait faktorok szignifikáns koefficiensei, a diszkriminanciát pedig alacsony korrelációk jelzik a trait és method faktorok körében, szignifikáns koefficiensekkel párosulva a method faktorok esetében.

Excluded Method Factors with Correlated Uniqueness (lásd:EMF_CU, 6.3 ábra)

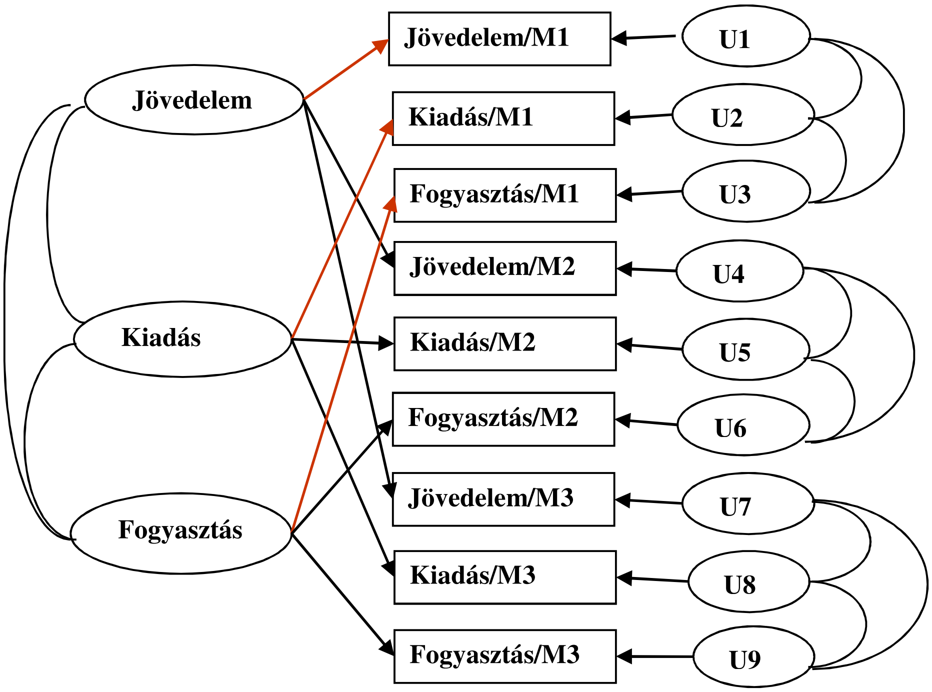
Ebben a modellben *method faktorokat explicit módon nem definiálunk*. E helyett a *standardizált* unique faktorok - azonos módszert használó indikátoraik körében- korrelálnak egymással. A szignifikáns *módszer-hatást* szignifikáns korrelációk támasztják alá az egyedi faktorok körében. Ez a modell *takarékosabb* a paraméterek számát illetően.

A goodness-of-fit indexek mindkét modell esetén kiváló illeszkedést mutatnak. Ez különösen jól látszik a Bentler-féle és a Bollen-féle normált mértékek tekintetében (97-98%). A magas *unique factor* korrelációk a *method* hatások evidenciáját támasztják alá.

6.2 ábra IMF_UU *Multitrait - Multimethod modell: 33 paraméter, 45 egyenlet*



6.3 ábra EMF_CU *Multitrait - Multimethod modell: 30 paraméter, 45 egyenlet*



A heteroszkedaszticitás SEM tesztelése a faktormodellben

Homogenitásvizsgálati problémánk, hogy tartható-e a feltevés, miszerint a hibafaktorok kovariancia mátrixának szerkezete a faktormodellben H_0: $\sigma^2\mathbf{I}$. Ha igen, akkor alkalmazható pl. az EPIC extrahálási módszer. A teszt a SEM illeszkedésének az alkalmazására vezet ADF, vagy iteratív WLS módszerrel végezve a koefficiensek becslését - attól függően, hogy az indikátorok normalitást mutatnak, vagy sem -, a becslést mind H_0, mind H_1 paramétereire elvégezve.

Mindkét becslési módszer eredményeit megadjuk az összehasonlítás érdekében. A pszeudo R^2 értéke rendre 98.83% és 74.24% az IWLS homogén és heterogén modellekre. E mértékek az ADF becslésekre rendre 78.16% és 34.95%. A két modell közötti differencia jelentősnek tűnik.

A *Bentler-indexek* számított értékei a - H_0 és H_1 modellek egymáshoz való viszonylatában a következők (az "*o*" alsó index a homogén, az "*e*" pedig a heterogén esetet jelzi, a *baseline* modell pedig a szűkebb, de nem a null modell):

$$NFI_{\text{homogeneous/heterogeneous}} \mid IWLS = 0.9546$$

$$NNFI_{\text{homogeneous/heterogeneous}} \mid IWLS = 0.9218$$

$$BCFI_{\text{homogeneous/heterogeneous}} \mid IWLS = 0.9573$$

$$NFI_{\text{homogeneous/heterogeneous}} \mid ADF = 0.6642$$

$$NNFI_{\text{homogeneous/heterogeneous}} \mid ADF = 0.408$$

$$BCFI_{\text{homogeneous/heterogeneous}} \mid ADF = 0.6771\,.$$

Minél *kisebb* az index értéke, annál *közelebb* van a két szóban forgó modell egymáshoz. Így az IWLS eredmények inkább H1 szerint a heterogén varianciák bővebb modelljének az elfogadását, míg az ADF eredmények a H0 szerinti szűkebb, homogén hibavarianciák modelljének az elfogadását javasolja. Itt emlékeztetünk arra, hogy az IWLS eredmények megkérdőjelezhetők, hiszen a normalitás, és a zéró kurtozis feltevése esettanulmányunkban logikailag nem vélelmezhető. Végső konklúzió tehát, hogy a $\sigma^2\mathbf{I}$ megszorítás érvénye – mikor alkalmazási előfeltétel - mindig tesztelendő.

7. A küszöb alá „*csúszás*" kockázatának korrespondencia ábrázolása

A korrespondencia analízis az asszociációs kapcsolat vizuális ábrázolása érdekében egy gyakorisági tábla adatait grafikus ábrává konvertálja.

Tekintsük az *n*-elemű sokaságot tagoló gyakoriságok p_{ij} (i=1,2,…,I, j=1,2,…,J) relatív gyakorisági tábláját, melyben az *egységnyi* gyakoriságra vetített *Pearson*-asszociációs mutató értéke – az *inercia*:

$$\frac{\chi^2}{n} = \sum_{i=1}^{I}\sum_{j=1}^{J}\frac{\left(p_{ij} - s_i o_j\right)^2}{s_i o_j}$$

ahol $s_i o_j$ az (i,j) cellának (a relatív peremmegoszlások alapján) az asszociáció teljes hiánya esetén várt relatív gyakorisága. A korrespondencia analízis az asszociáció *Pearson*-mértékét bontja komponensekre úgy, hogy a sorokat (oszlopokat) mint pontokat a *megoszlásaikból* képzett mesterséges *főtengelyek* redukált dimenziójú terében ábrázolja. A főtengelyek rendre csökkenő mértékben járulnak hozzá a *teljes inerciához.*

A sorkoordináta a standardizált oszlopkoordináták súlyozott átlaga, míg az oszlopkoordináta a standardizált sorkoordináták súlyozott átlaga, súlyként a megfelelő sor-, illetve oszlopprofilt alkalmazva. Az oszlopok és a sorok koordinátáinak egymásba való átvitele a koordináták *duális skálázását* jelenti. A skálázás révén egy oszlopprofil a pontok terében ahhoz a sorhoz húzódik közelebb, amelyiknek a súlya domináns az illető profilban. Közös koordináta rendszerben ábrázolva a sorok és az oszlopok pontfelhőjét, azon sorok és oszlopok kerülnek várhatóan közel egymáshoz, amelyek között szoros az asszociáció mértéke.

Mikor kettőnél több változó szerepel vizsgálatunkban, a változók sorrá, vagy oszloppá kombinálása helyett célszerű a korrespondencia analízis *többszörö*s változatát (MCA) alkalmazni. A többszörös analízis ekvivalens az ún. *indikátor* mátrix *egyszerű* analízisével. A $\mathbf{Z}_{(n,J)}$ indikátor mátrix sorait az i=1,2,...,n megfigyelési egységek alkotják, míg oszlopait a Z_q (q=1,2,...,Q) diszkrét változók kategóriái képezik, ahol a Z_q változónak J_q számú kategóriája van. Így a mátrix oszlopainak száma $J=J_1+J_2+...+J_Q$, és az oszlopok a Q számú csoport egyikének a tagjai. Az indikátor mátrix mindegyik sora Q számú „1" elemet tartalmaz attól függően, hogy az illető megfigyelési egység adott változó melyik kategóriájához tartozik. Egyébként a mátrix elemei zérók.

Az alábbiakban *háztartások* (HT) tekintetében - modell példában - vizsgáljuk, hogy prediktor kategóriák miként klasszifikálják MCA módszerrel a *fuzzy* szegény (a tagsági függvény értéke nagyobb, mint zéró, de kisebb, mint 1) háztartások lecsúszását, vagy a küszöb fölött maradását.

Az alkalmazott kategória-változók és kategóriáik:

- A statust tekintve S=O.K., ha a HT stabil (egyértelműen nem szegény), S=NEMO.K., ha a HT biztosan szegény és S=FUZZY, ha a HT szegénységi kockázatnak van kitéve, vagyis *fuzzy* (homályosan) szegény.
- A HT formációja F={F_1, F_2, F_3, F_4, F_5, F_6}.
- A háztartásfő iskolai végzettségének kategóriái: I={i1,i2,...,i15}.
- A HT településének régiója: R={CHU,CTD,WTD,STD,NHU,NGP,SGP} (Central Hungary, Central Transdanubia, Western Transdanubia, Southern Transdanubia, Northern Hungary, Northern Great Plain és Southern Great Plain).

Az alkalmazott, diszkretizált, folytonos prediktorok:

- P: Jövedelmezőség (HT méretarányos): Low, Moderate, Average, High, Extreme,
- L: Likviditás: Low, Moderate, Average, High,
- D: Eladósodottság: Low, Average, High,
- E: Vagyon: Low, Moderate, Average, High

Célunk, hogy az F, I, R, P, L, D, E prediktorok értékei alapján klasszifikáljunk egy FUZZY minősítésű *fuzzy szegény* háztartást *lecsúszás*, vagy *stabilitás* tekintetében.

Ez a korrespondenciák feltárását igényli a *dependent* és a *prediktor* kategóriák között. Másfelől eltérésnek kell látszani az O.K. és a NEMO.K. átlagos sorszerkezetek között a *prediktor* térben. Végül a háztartások ábra-legközelebbi szomszédait vizsgáljuk, hogy túlnyomóan szegények, vagy inkább stabilak. A feladat egyfelől *regressziós*, másfelől *diszkriminancia* jellegű.
A számítások az indikátor mátrix korrespondencia analízisére épülnek. i) A korrespondenciák feltárása valamennyi előforduló kategóriát érinti. ii) A diszkriminancia és az előrejelzés viszont egy redukált indikátor mátrixot alkalmaz, melynek oszlopait csak a prediktor kategóriák alkotják.

Mivel a FUZZY háztartások klasszifikálandók, ezért ők a számításból kimaradnak. Az indikátor mátrix oszlopainak száma az analízistől függően 45 vagy 43.

Asszociációk feltárása

A *45-oszlopú* indikátor mátrix totális inerciája (45/8-1)=4.625. Ez esetben 1/Q=1/8=0.125 tehát az ennél magasabb négyzetes szinguláris értékű CA tengelyek megtartandók. Az első három vezető tengely relatív részesedése rendre: μ_1^2=0.248 (5.4%), μ_2^2=0.184 (4%), μ_3^2=0.159 (3.4%). A kiszűrendő tengelyek száma most mindenképpen 3, tekintet nélkül a magyarázott variancia kumulált értékeire.

Mikor az indikátor mátrix két oszlopának a profiljai hasonlóak, akkor a megfelelő kategóriák pontjai a grafikus ábrán közel kerülnek egymáshoz. Így az azonos, vagy nagyon hasonló háztartások szomszédos pontok klasztereként jelentkeznek az ábrázoláskor.

Az extrahált tengelyek tartalma a következő.

1. Az első CA tengely jelentése: P_Low, E_Low,D_High, I-i8, F_3, F_4, P_Mod, P_Ext pozitív koordinátákkal, és S_NEM.OK., F_2, F_6, I_i1, I_i2, I_i5, L_Hi, negatív koordinátákkal.
2. A második tengely jelentése: I_i13, P_High, I_i14, I_i15, E_High, F_4, I_i11 pozitív koordinátákkal, és S_NEMO.K., F_COP, D_High, E_Low, I_i1, I_i2, P_LOw, I_i5, D_Av negatív koordinátákkal.
3. A harmadik tengely jelentése: S_NEMO.K., F_COP, I_i1, I_i2, E_High, pozitív koordinátákkal, és D_High, E_Low, I_i5, I_i3 negatív koordinátákkal.

Lévén az O.K. háztartások száma dominálja a NEMO.K. háztartások gyakoriságát, ezért az O.K. csoport az ábrán az origó által jól reprezentált. Így az origó körül elhelyezkedő, de ahhoz nem túl közeli pontok nem asszociálnak a szegénységgel.

A 7.1 ábrán az F_3, P_Ext, F_4, I_i15 kategóriák távol esnek a szegénység ponttól az Axis1, Axis2 síkban, míg az F_2 és I_i2 közel vannak a S_NEMO.K. kategóriához. Másfelől, a NEMO.K. kategória messze elhúzódik az origótól, magával húzva néhány prediktor kategóriát a megfelelő tengelyen ugyanabban az irányban, mint például I_i1, I_i3, I_i5, F_6 az első tengelyen és D_High, E_Low, P_Low, D_Av a második tengelyen. Az *átviteli formula* alapján a távolságok az ábra pontjai között, erős, vagy gyenge asszociációt jeleznek a pont és a NEMO.K. kategória között.

A prediktív térkép

Az MCA lehetővé teszi újabb pontok hozzáadását a meglévő ábrához, előrejelzési céllal. Az ún. *supplementary* változó kategóriái az ábrába beszúrandó pontok.

Elsőként az indikátor mátrix oszlopainak az ábrázolásával – két dimenzióban - megjelenítjük a prediktor kategóriák közötti kapcsolatokat. Így kapjuk az ún. *„prediktív"* térképet (7.2 ábra). Ezt követően a kiegészítő sorprofilt rávetítjük a prediktív térképre az *átviteli formulát* használva. A kiegészítő kategóriák helyzete a prediktor kategóriák helyzetében mutatja a függőségeket és a függetlenségeket.

Esetünkben az O.K. és a NEMO.K. kategóriák játsszák a prediktív térkép által diszkriminálandó kiegészítő kategóriák szerepét, és a FUZZY kiegészítő kategória pozíciója (az egyedi FUZZY háztartások átlaga) viszonyítandó az ő helyzetükhöz. A prediktív térkép számításából kizártuk a FUZZY háztartásokat.

A prediktív térképre egy kiegészítő *individuális* háztartás is rávetíthető, és vizsgálható, hogy a szomszédai túlnyomóan szegények, vagy stabilak.

Az O.K. kategóriával asszociáló prediktív kategóriák: F_5, I_i4, I_i6, I_i7, I_9, I_i10, R_CHU, R_CTD, L_Low, D_Low, E_Av. Ugyanakkor a NEMO.K. kategóriával asszociáló prediktív kategóriák az Axis1 és Axis 2 tengelyek mentén: I_i2, R_WTD, R_STD, R_NHU, R_NGP, R_SGP, az Axis1 és Axis 3 tengelyek mentén pedig: L_Mod, L_Av.

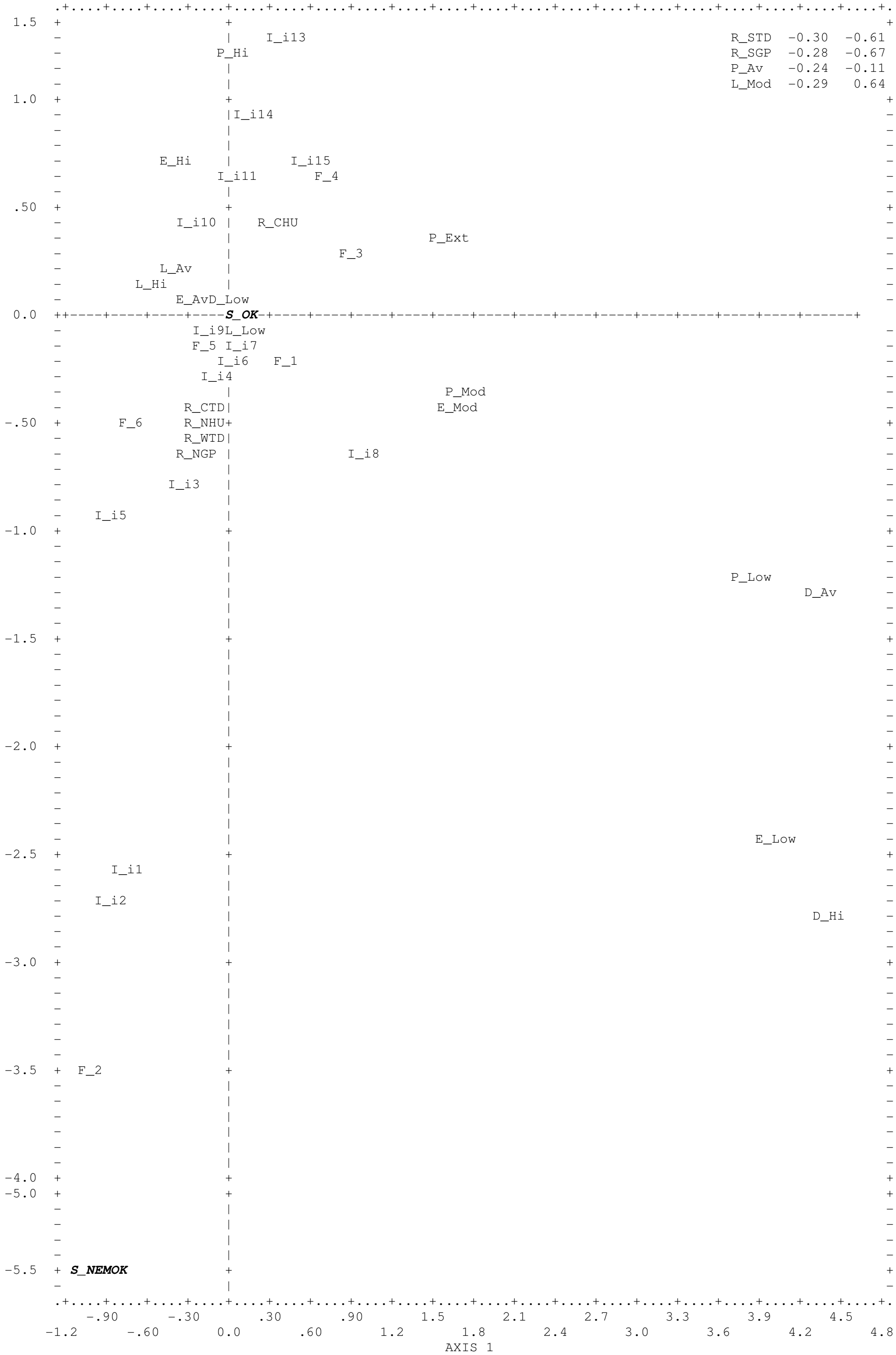
Adott háztartás is klasszifikálható a prediktív térképre vetítve.

A 7.2 ábra alapján egy O.K., FUZZY, NEMO.K. tengely rajzolódik ki, miszerint egy F_5 típusú, Central Hungary településű, 1. fokozatú iskolai végzettségű, alacsony jövedelmezőségű, alacsony likviditású, magas eladósodottságú, kisvagyonú háztartás (HT) koordinátái a szegény háztartás irányába mutatnak.

7.1 tábla Az OK, NEMOK, FUZZY csoportok kiegészítő profiljai (centroidjai)

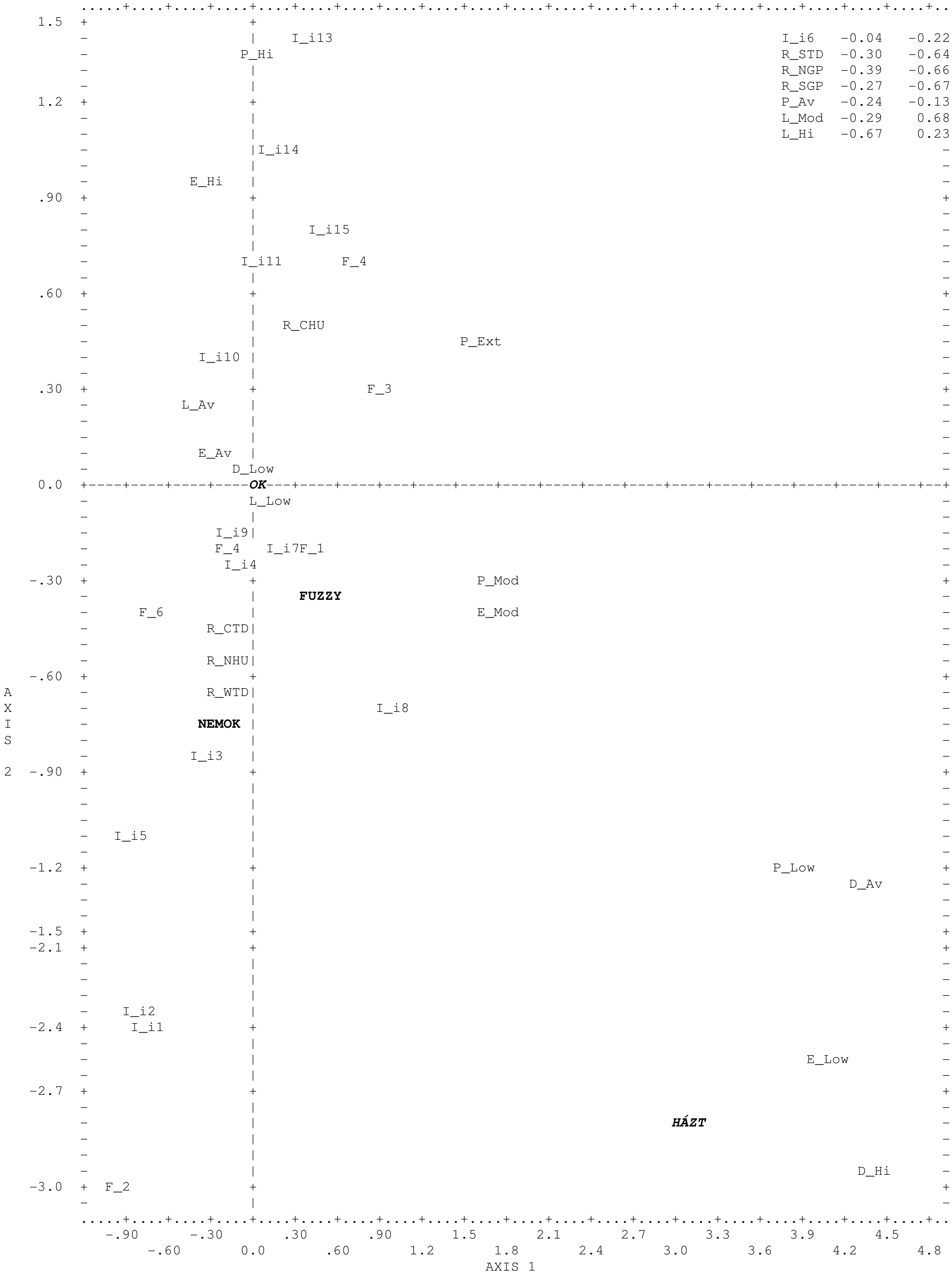
NAME	QLT	FACTOR AXIS 1	COR2	FACTOR AXIS 2	COR2	FACTOR AXIS 3	COR2
OK	0.721	0.001	0.050	0.001	0.321	-0.001	0.350
NEMOK	0.721	-0.296	0.050	-0.748	0.321	0.781	0.350
FUZZY	0.740	0.291	0.233	-0.426	0.499	-0.054	0.008

7.1 ábra Az oszlopkategóriák korrespondenciái, Axis 1_2, X=5.36%, Y=3.98%



* Az átfedések az ábra jobb felső sarkában koordinátáikkal felsorolásra kerültek.

7.2 ábra Prediktív térkép, Axes 1_2, X=5.50% Y=4.04%



Összefoglalás

Az értekezés bemutatta a Szerzőnek a szegénység statisztikai méréséhez történt módszertani hozzájárulását. Ennek eredményéül, az értekezés hozadékaként, a szakterület elmélete új formulákkal, elvekkel, szintézisekkel és alkalmazásokkal gazdagodott.

Összegezve, az alábbi eredmények hangsúlyozandók.

1. A relatív depriváció mértékének beépítése a szegénységi mértékbe,
2. A jövedelmi transzfer relatív depriváció szintjére gyakorolt hatásának új elvi megközelítése és elemzése,
3. Új relatív deprivációs mérőszámok megadása, definiálása,
4. Új, relatív depriváció-érzékeny szegénységi mérték konstrukció megadása, definiálása,
5. Új, többváltozós-többdimenziós egyenlőtlenségi módszertan kidolgozása, megadása
6. Az új egyenlőtlenségi módszertan dekompozíciós alkalmazása,
7. Az új egyenlőtlenségi módszertan szegénységi mérésben való alkalmazása,
8. A deprivációs küszöb fogalmának bevezetése, és kvantilis regressziós becslése,
9. A szegénységi küszöb alatti ritka esemény kismintás következtetési problémáinak egzakt jellegű kezelése,
10. A szegénység latens változóként kezelése, és strukturális összekapcsolása a relatív depriváció és társadalmi kirekesztés latens változókörrel,
11. A szegénységi küszöb alá csúszás prediktorainak detektálása diszkrét változós környezetben.

Az eredmények hasznosíthatók

- a szegénység fokának mérésében,
- a szegénység strukturális forrásának (területi, háztartási, stb.) felbontásában, feltárásában,
- szegénységi, relatív deprivációs küszöbértékek becslésében,
- szegénységi prediktorok detektálásában,
- végül releváns ok-okozati kapcsolatok környezeti tesztelésében.

A továbbfejlesztés lehetősége valamennyi problémakörben adott, de kiváltképp hangsúlyos az általánosított variancia alapú egyenlőtlenség- és szegénységmérés területén.

Hivatkozások

A szerzőnek az értekezés témakörében megjelent, kiemelt tudományos közleményei

Könyvek

1. Hajdu,O.(1997a): *A szegénység mérőszámai*. KSH, Könyvtár és Dokumentációs Szolgálat, Budapest, 99 oldal
2. Hajdu,O. (2003a): *Többváltozós statisztikai számítások*. Központi Statisztikai Hivatal, Budapest, 457 oldal

Tanulmánykötetek

1. Hajdu,O. (2004a): *A Lorenz-görbe aszimmetriája. in Egy reneszánsz statisztikus*. Sz. Vita László, KSH, Budapest, 36-42. oldal
2. Hajdu,O. (2008a): *Transzformált jövedelmek eloszlása. in A közgazdaságtudomány tisztessége*. Sz. Veress József, Műegyetemi Kiadó, 103-106. oldal

Folyóiratcikkek

1. Hajdu,O. (1996): *Relatív depriváció és szegénység: A jövedelmi transzfer deprivációs hatása*. Szigma, XXVII., 1-2.sz., 45-66. oldal
2. Hajdu,O. (1997b): *Relatív depriváció és szegénység: A szegénység depriváltság érzékeny mérése*. Szigma, XXVIII., 1-2.sz., 7-31. oldal
3. Hajdu,O.:(1999): *On the deprivation-sensitive measurement of poverty*. Hungarian Statistical Review, special number 3, pp. 15-22.
4. Hajdu,O. (2002): *Category selection and classification based on correspondence coordinates*. Hungarian Statistical Review, special number, Vol. 80., No. 7, pp. 103-126.
5. Hajdu,O. (2003b): *A kovariancia-struktúra modellek illeszkedésvizsgálata*. Statisztikai Szemle, 81. évf., 5-6. sz., 442-465. oldal
6. Hajdu,O. (2004b): *Diagnostics of the Error Factor Covariance Structure*. Hungarian Statistical Review, Vol. 82 Special Number 9, pp. 68-94
7. Hajdu,O. (2004c): *A csődesemény logit-regressziójának kismintás problémái*. Statisztikai Szemle, április, 82. évf., 4.sz., 390-422. oldal
8. Hajdu,O. (2004d): *Multitrait-Multimethod Models for Profitability Indicators*. Periodica-Politechnica, Vol.12. No.2. 211-222
9. Hajdu,O. (2004e): *Rotáció az egyszerű faktorstruktúráért*. Statisztikai Szemle, 82. évf., 10-11.sz. október-november 978-990. oldal
10. Hajdu,O. (2006a): *Exact inference on poverty predictors based on logistic regression approach*. Hungarian Statistical Review, special number 10. Vol.84 134-147
11. Hajdu,O. (2009): *Poverty, Deprivation, Exclusion: A Structural Equations Modelling Approach*. Hungarian Statistical Review, special number 13., Vol.87 90-102
12. Hajdu,O. (2010): *Sajátértékek a statisztikában*. Statisztikai Szemle, július-augusztus, 88. évf., 7-8.sz., 773-788. oldal

Konferencia előadások

1. Hajdu,O. (1992): *On a new poverty measure*. A Nemzetközi Ökonometriai Társaság konferenciáján elhangzott előadás, Tillburg
2. Hajdu,O. (2004f): *Strukturális egyenletek specifikálása latens változókkal*. A Tudomány Napján, a Magyar Tudományos Akadémián elhangzott előadás. Budapest, MTA

3. Hajdu,O. (2006b): *Szegénység, depriváltság, kirekesztettség: latens változók indikátoraik tükrében*. Az MTA Statisztikai Bizottsága ülésén elhangzott előadás, Budapest, KSH
4. Hajdu,O. (2008b): *A csődvalószínűség logit alapú becslésének mintavételi sajátosságai*. A X. Gazdaságmodellezési Szakértői Konferencián elhangzott előadás. Budatétény, Gazdaságmodellezési Társaság
5. Hajdu,O. (2011): *Egyenlőtlenségi mérőszámok alkalmazása az adatbányászatban*. Az MST: „Tematikus nap az egyenlőtlenség vizsgálatáról, méréséről" c. ülésén elhangzott előadás, KSH, Budapest

Disszertációk

1. Hajdu,O. (1984): *Adalékok a koncentráció statisztikai vizsgálatához, különös tekintettel a Lorenz-görbe aszimmetriájának mérésére és értelmezésére.* Egyetemi doktori értekezés, Pécs
2. Hajdu,O. (1991): *A jövedelemeloszlás. A szegénység mérése*. Kandidátusi Értekezés, Budapest

Meghatározó irodalmak

1. Agresti,A. (2002): *Categorical Data Analysis*, 2nd Edition, Wiley
2. Alkire,S.; Foster,J.E. (2007): *Counting and Multidimensional Poverty Measurement*. Oxford Poverty & Human Development Initiative OPHI Working Paper 7.
3. Bentler,P.M. (1990): *Comparative Fit Indexes in Structural Models*. Psychological Bulletin, Vol.107. No.2, 238-246.
4. Bollen,K.A. (1989): *Structural Equations with Latent Variables*. Wiley. New York
5. Chakravarty,S.R.; Mukherjee,D. (1999): *Measures of deprivation and their meaning in terms of social satisfaction*. Theory and Decision **47:** 89–100, Kluwer Academic Publishers. Printed in the Netherlands.
6. Éltető,Ö.; Frigyes,E. (*1968): New inequality measures as efficient tools for causal analysis and planning*. Econometrica, 36.k. 2.sz. 383-396.old.
7. Éltető,Ö.; Havasi,É. (2006): *Poverty in Hungary with special reference to child poverty*. Hungarian Statistical Review. Vol. 84. Special No. 10. pp. 3–17.
8. Éltető,Ö.; Havasi,É. (2009): *A hazai jövedelemegyenlőtlenség főbb jellemzői az elmúlt fél évszázad jövedelmi felvételei alapján*. Statisztikai Szemle, 87. évf. 1.sz. 5-40
9. Ferge,Zs. (1969) : *Társadalmunk rétegződése : elvek és tények. KJK,*
10. Foster,J.E.; Greer,J.; Thorbecke,E. (1984): *A Class of Decomposable Poverty Measures. Econometrica*, 52, 761–767;
11. Greenacre,M.J. (1984):*Theory and Application of Correspondence Analysis*. ACADEMIC PRESS, INC. London
12. Hirji,K.F. (1992): *Exact distributions for polytomous data*. JASA, 87, 487-492
13. Koenker,R.; Bassett,G.Jr. (1978): *Regression Quantiles*. Econometrica, 46, No.1. pp.33-50.
14. Mussard,S.; Seyte,F.; Terraza,M. (2003): *Decomposition of Gini and the generalized entropy inequality measures*. Economics Bulletin, Vol. 4, No. 7 pp. 1−6
15. Sen,A. (1976): *Poverty: An Ordinal Approach to Measurement*, Econometrica, 44, 219-231
16. Shorrocks,A.F. (1995): *Revisiting the Sen poverty index*. Econometrica 63: 1225 – 1230.
17. Spéder,Zs. (2002): *A szegénység változó arcai*. Századvég Kiadó 240 old.
18. Szalai,J. (2002): *A társadalmi kirekesztődés egyes kérdései az ezredforduló Magyarországán.* Szociológiai Szemle 2002/4. 34–50.
19. Szívós,P.; Tóth István,Gy. (2001): *A jövedelmi szegénység: trend és profil 2000-ben*. Statisztikai Szemle. 79. évf. 10-11. sz. 848-861
20. Takayama,N. (1979): *Poverty, Income Inequality and their Measures*: Professor Sen's Axiomatic Approach Reconsidered, Econometrica, 47, 747-759
21. Tóth István,Gy. (2003): *Jövedelemegyenlõtlenségek – tényleg növekszenek, vagy csak úgy látjuk*? Közgazdasági Szemle, L. évf., március (209–234. o.)
22. Zheng,B. (1997): *Aggregate Poverty Measures*. Journal of Economic Surveys, 11, 123-162