Többváltozós-többdimenziós egyenlőtlenség és a szegénység

Hajdu Ottó,
a Budapesti Corvinus Egyetem tanszékvezető egyetemi docense
E-mail: hajduotto@uni-corvinus.hu

A tanulmány célja egy új, terminológiánk szerint GVIP (generalized variance inequality and poverty) többváltozós módszer definiálása az egyenlőtlenség többdimenziós mérésében, majd szegmentált társadalomra megadni annak külső-belső csoportközi felbontását, és az elvet a szegénység mérésében is alkalmazni. A módszer a szóródás többváltozós, általánosított variancia mértékén alapul. A csoportközi dekompozíció a Wilks' lambda hányadost alkalmazza, lehetővé téve a numerikus számítások standard statisztikai programmal történő kalkulálását. A GVIP elv a szerző által definiált új mátrix, nevezetesen a Theil-kovarianciamátrix determinánsára épül, amit a tanulmány általánosított Theil-varianciaként nevez el. A szegénység mérésében transzformált eloszlásokra alkalmazva a GVIP mint szegénységi mérték adódik. A GVIP figyelembe veszi a dimenziók korrelációs rendszerét és aszimmetrikus eloszlását, és egydimenziós esetben is többváltozós technikát alkalmaz, kihasználva annak előnyeit.

TÁRGYSZÓ:
Általánosított entrópia és variancia.
Többdimenziós egyenlőtlenség és szegénység.
Diszkriminancia analízis, Wilks' lambda.

A tanulmány gazdasági-társadalmi jelenségek (dimenziók) egyenlőtlenségét vizsgálja, amit szóródásként értelmez és ennek megfelelően méri azt. Célunk több dimenzió tekintetében egyidejűleg és kompozit módon mérni az egyenlőtlenség fokát, figyelembe véve a dimenziók korrelációs kapcsolatait és aszimmetrikus eloszlását, majd szegmentált társadalomra megadni az egyenlőtlenség külső-belső arányát, végül pedig rangsorolni a csoporthatásokat a belső egyenlőtlenséghez való százalékos hozzájárulásuk alapján. A módszertani mondandó bemutatása, tárgyalása érdekében – illusztratív céllal – a tanulmány a jövedelem, fogyasztás, vagyoni helyzet dimenziókört alkalmazza.

A cikk egy új egyenlőtlenségi módszertant vezet tehát be, mely más területeken, például a szegénység mérésében (mint jelen tanulmányban is), tovább alkalmazható. A javasolt módszertan az egydimenziós, általánosított entrópia mértékből indul ki, mely vagy csak a jövedelem, vagy csak a fogyasztás, vagy csak a vagyon esetére, esetleg ezek valamely egydimenziós kombinációjára vonatkozik. Az entrópia azonban – statisztikai értelemben – egydimenziós esetben is kétváltozós számítás, mert formulája (például a jövedelem esetén) igényli magát a jövedelmet és a jövedelem logaritmusát is. Ebben a megközelítésben a jövedelem egy latens dimenzió, melynek két manifeszt változója valamely konkrét jövedelmi tétel és annak logaritmusa.

Az entrópia jellegű logaritmus-megalapozás figyelembe veszi a vizsgált eloszlás (jövedelem) aszimmetrikus voltát, közelebb hozva a szimmetrikus (normális) eloszlás esetét, lehetővé téve így statisztikai tesztek alkalmazását is. Különbséget teszünk tehát dimenziószám és változószám között: ha a vizsgált dimenziók száma p, az alkalmazott változószám $2p$. Ezért az egydimenziós vizsgálat is értelemszerűen kétváltozós.

A tanulmány új eredményként a dimenziókból és a logaritmusaikból képezi a Theil-féle kovarianciamátrixot, melynek determinánsa adja a Theil-variancia egyenlőtlenségi mértéket. A Theil-mátrix rendje $(2p, 2p)$, és elemeinek jelentését nevezetes (elsősorban információelméleti alapú) egyenlőtlenségi mértékek adják. A Theil-variancia figyelembe veszi mind a dimenziók, mind a változók korrelációs kapcsolatait az egyenlőtlenség fokában. Analógiaként hozva az euklideszi vs. Mahalanobis-távolságot, míg az előbbi korrelálatlan, addig az utóbbi korrelált koordinátatengelyeket feltételez. Jelen cikk ebben az értelemben a Mahalanobis-jellegű egyenlőtlenségi mértékek irányában lép tovább.

A társadalom csoportosítása esetén – alapvetően gazdasági-társadalmi jellegű csoportosításra gondolva (például település, régió, háztartástípus, szegény volt) – a kovarianciamátrix külső és belső komponensek összegére bontható, ahol a totális általánosított Theil-variancia mértékében a belső egyenlőtlenség arányát a Wilks' lambda jellemzi. Így a Wilks' hányados lehetővé teszi az alkalmazott csoportosítás százalékos hozzájárulásának az elemzését az egyenlőtlenség forrása tekintetében. Transzformált adatokra alkalmazva, az általánosított variancia mint általánosított szegénységi mérték is értelmezhető. Egyféle megfelelő transzformáció a cenzorálás, ahol a szegénységi küszöb feletti értékeket a küszöb szintje helyettesíti. Ezáltal a szegénységi mérték érzéketlen a küszöb fölötti átrendeződésekre.

A cikk felépítése a következő. Az *1*. fejezet röviden áttekinti az entrópia fogalmát, majd definiálja a Theil-kovariancia C_T mértéket, és megadja kapcsolatát a nevezetes Theil-féle T_1 és T_2 entrópia alapú indexekkel.[1] A *2*. fejezet definiálja a Theil-mátrixot, és javasolja a determinánsát mint új egyenlőtlenségi mértéket Theil általánosított variancia (röviden Theil-variancia) elnevezéssel, majd példán keresztül bemutatja számításának menetét. A *3*. fejezet ismerteti a Theil-variancia csoportközi felbontását, illusztrálja a számításokat, és értelmezi az eredményeket. A *4*. fejezet kiterjeszti a módszertant a többdimenziós esetre is. Az *5*. fejezet végül a Theil-varianciát cenzorált eloszlásra alkalmazva, értékét szegénységi mérőszámként értelmezi, és adott csoportosításra vonatkozóan dezaggregálja.

A tanulmány a számítások részleteinek bemutatásához, az eredmények könnyű ellenőrzése céljából egyfelől egy modellpélda jellegű illusztratív adatállományt használ, másfelől – az eredmények valós nagyságrendjének és a gyakorlati alkalmazás lehetőségeinek érzékeltetése érdekében – az új módszertan a magyar háztartásokat jellemző költségvetési felmérés – az ún. Háztartási Költségvetési Felvétel (HKF) – adatain is alkalmazásra kerül (*KSH* [2003]).

1. A Theil-kovariancia entrópia-felbontása

Mivel a központi mondanivaló a Theil-kovarianciamátrix elemeire épül, ezért a tárgyalást a Theil-kovariancia fogalmi bevezetésével és értelmezésével kezdjük.

1.1. Az entrópia jelölésrendszere és pszeudo-R^2 tartalma

Legyen az n tagú társadalom relatív jövedelmi eloszlása az átlagos jövedelem bázisában

$$r_i = \frac{Y_i}{\overline{Y}} \qquad \left(i = 1, 2, \ldots, n\right), \qquad /1/$$

[1] Innen ered a cikk által alkalmazott elnevezés, a Theil-kovariancia.

ahol Y_i az i egyén jövedelme. A relatív jövedelem az átlagos „1" jövedelemhez viszonyítva a

$$d_i = r_i - 1 \quad \left(\bar{d} = 0\right) \qquad /2/$$

hozamot eredményezi, melynek logaritmikus közelítését a relatív jövedelem logaritmusa adja[2]

$$D_i = ln\left(r_i\right) \approx d_i \quad \left(\text{ahol: } \bar{D} \leq 0\right) . \qquad /3/$$

Ez a közelítés annál pontosabb, minél közelebb áll a d_i tényleges hozam a zéróhoz, vagyis a jövedelmi átlagpont szűk környezetében. Totális egyenlőség esetén a *D-hozam* és a *d*-hozam egybe esik, növekvő egyenlőtlenség esetén pedig távolodnak egymástól.

A *Shannon-* [1948] entrópia eredetileg a kapott hír alapján történő előrejelzés bizonytalanságának a mértéke

$$H(r) = ln\left(n\right) - \frac{1}{n}\sum_{i=1}^{n} r_i D_i \quad . \qquad /4/$$

Ha a jövedelmek eloszlását magyarázó prediktor változók információja üres, akkor az információ zéró többletet ad az egyedi jövedelmek eloszlásáról, az információ hiányában az a priori előrejelzés egy mindenkire egyaránt vonatkozó konstans jövedelem, és a hír bizonytalansága maximális. Ha a hír információja nem üres, és ezt mindenki ismeri, akkor a jövedelmi modell előrejelzése maga az aktuális eloszlás, a posteriori csökkentve az előrejelzés bizonytalanságát az információ birtokában. Végül, ha létezik „a tökéletes információ", akkor ezt értelemszerűen csak egyvalaki ismerheti, így övé a totális jövedelem, a modell előrejelzése pedig egyértelmű, zéró bizonytalanság mellett. A klasszikus Shannon-entrópia növekvő értékkel egyre egyenletesebb eloszlást jelez, tehát egyenlőségi mutató.

A bizonytalanság csökkenését, közeledését a hír hatására a totális bizonyosság irányába – a lehetséges maximális és minimális szint közötti terjedelmen – az ún. pszeudo-R^2 illeszkedési mutató számszerűsíti relatív (százalékosan értelmezendő) formában: $R^2 = 1 - H\left(r\right)/ln\left(n\right)$.

[2] A közelítés a logaritmus függvény $r = 1$ pontban történő Taylor-sorának lineáris tagját használja: $ln\left(r\right) \approx \left(r-1\right) = d$. Ha például r = 1,01, akkor az egzakt hozam $d = r - 1 = 0,01$, azaz 1 százalék hozam, a közelítő hozam pedig D=ln(1,01)=0,00995.

Az egyenlőtlenség mérésére azonban – szemben a Shannon-entrópiával – célszerű olyan mutatót alkalmazni, mely növekvőleg az egyenlőtlenség emelkedését jelzi, és több hangsúlyt helyez az eloszlás szegényebb, alsó szegmensén történő változásokra. Ezeket a szempontokat teljesíti a jól ismert generalized entropy (GE) mérték, mely egy α paraméterrel figyelembe veszi az egyenlőtlenséggel szemben érzett averzió mértékét is. Ezek értelmében az α paraméterű általánosított entrópia (*Cowell* [1977]; *Bourguignon* [1979]; *Shorrocks* [1980]; *Cowell-Cuga* [1981a], [1981b]) formulája az /1/ és /3/ jelölésekkel

$$GE(\alpha) = \frac{1}{n\alpha(\alpha-1)}\sum_{i=1}^{n}\left[(r_i)^{\alpha} - 1\right], \quad \alpha \neq 0, \alpha \neq 1 \ , \qquad /5/$$

ahol L'Hospital-határértékben – és /4/ alapján

$$GE(1) = \frac{1}{n}\sum_{i=1}^{n} r_i D_i = ln(n) - H(r) \ , \qquad /6/$$

$$GE(0) = -\frac{1}{n}\sum_{i=1}^{n} D_i = -\bar{D} \ . \qquad /7/$$

A $GE(\alpha)$ index az egyenlőtlenség mutatója, növekvő értékkel az egyenletes eloszlástól való távolodást jelezve. Alacsonyabb α nagyobb súlyt ad az eloszlás alsóbb szegmensén, mint a felső szegmensen történő transzferváltozásra.[3]

Értelmüket tekintve a két α-specifikus GE-index a két nevezetes Theil-egyenlőtlenségi indexet jelenti, rendre:

– *GE*(1): a redundancia Theil_1-indexe (*Theil* [1967]) és
– *GE*(0): a mean logarithmic deviation (MLD), avagy Theil_2-index (*Theil* [1967] 125. old.).[4]

1.2. A Theil-kovariancia

A cikk a relatív jövedelem és a log-hozam közti kovarianciát Theil-kovariancia elnevezéssel vezeti be, és definiálja az alábbi módon

$$C_T = Cov(r, D). \qquad /8/$$

[3] A GE-index egy további paraméterrel történő kiterjesztését lásd *Cowell* [2005].
[4] *GE*(1) megszokott jelölése az irodalomban T_1, míg *GE*(0) jelölése T_2. Az *MLD* mérőszám magyar jövedelmi adatokra való egy alkalmazását lásd *Tóth István György* [2003].

A Theil-kovariancia tartalmát a *GE* felbontása adja /1/, /3/, /6/, /7/ alapján, mely (tekintve, hogy $\overline{r}=1$)[5]

$$C_T=\frac{1}{n}\sum_{i=1}^{n} r_i D_i-\overline{r}\cdot\overline{D}=GE(1)+GE(0)\ . \qquad /9/$$

A Theil-kovariancia jelentése: az MLD eltéréssel növelt Theil-redundancia index, és C_T eleget tesz mindazon kritériumoknak melyeket MLD és T_1 teljesít, így a Pigou–Dalton-transzfer érzékenységi kritériumnak is.[6] A C_T kovariancia növekvő értéke emelkedő korrelációt mérve azt jelzi, hogy a *D* log-hozam egyre redundánsabbá válik mert *d* egzakt ismerete önmagában tartalmazza a releváns szóródási információt.

2. A Theil-mátrix és a Theil-variancia definiálása

Az egyenlőtlenséget a tanulmány mint a szóródás egy megjelenési formáját tekinti, ezért fokát is mint a szóródás fokát méri. Többváltozós megközelítésben a szóródás klasszikus mértéke az ún. generalized variance (általánosított variancia – GV) mutató, formálisan az aktuális változók kovarianciamátrixának a determinánsa.[7] Kulcskérdés tehát a megfelelő kovarianciamátrix megadása.

2.1. A Theil-mátrix

A tanulmány által bevezetett új egyenlőtlenségi mátrix – terminológiánk szerint – a Theil-mátrix. Az egyszerűség kedvéért az egydimenziós-kétváltozós (*r*, *D*) esetben a Theil-(kovariancia) mátrix definíciója

$$\mathbf{C}_T=\begin{array}{c|cc}\text{Változó} & r & D\\ \hline r & Var_r & C_T\\ D & C_T & Var_D\end{array}\ , \qquad /10/$$

[5] Bár a $\left[GE(1)+GE(0)\right]/2$ átlagot az irodalom szimmetria tulajdonsága miatt használja, a *GE*(1)+*GE*(0) összeg kovarianciatartalmának felismerése és a kovarianciamátrix alkalmazásáig való továbbvezetése a szerző önálló eredménye. Szimmetria esetén a mutató invariáns az (*x*), vagy (1/*x*) argumentum használatára.

[6] Érzékeny a regresszív transzfer mértékére és helyzetére az eloszlásban: a regresszív transzfer egy adott jövedelmi tételt elvesz, és egy gazdagabbhoz csoportosítja át. További, az egyenlőtlenségi indexekkel szemben támasztott általános axiomatikus kritériumok tárgyalását lásd például *Cowell* [2009].

[7] Az általánosított variancia két- és többváltozós bemutatását lásd például *Hajdu* [2003] (59–60. old.).

ahol *Var* a relatív jövedelem és a log-hozam varianciákat jelöli és C_T a Theil-kovariancia.

2.2. A Theil-variancia

Az egyenlőtlenség mértéke definíciónk szerint a Theil-mátrix determinánsa, a cikk terminológiájában a Theil- (általánosított) variancia, melynek formulája kétváltozós esetben

$$T_{GV} = det\left(\mathbf{C}_T\right) = Var_r Var_D - C_T^2 \ . \qquad /11/$$

Ha r a D log-hozammal gyengén korrelál – vagyis a relatív jövedelem és a log-hozam között alacsony a redundancia –, akkor Var_r és Var_D együttes információja szükséges a szóródás méréséhez, ha viszont a redundancia jelentős, akkor a $Var_r Var_D$ felső korlát a redundancia mértékében redukálandó.

Lévén T_{GV} szóródást mér, így értéke az egyenlőtlenség növekedésével nő, alsó és felső korlátai pedig rendre

$$0 \le T_{GV} \le Var_r Var_D \ . \qquad /12/$$

Az egyenlőtlenség növekedésével a Var_r variancia és vele Var_D emelkedik, a T_{GV} rés tágul, de a tágulást a Theil-kovariancia (összetevőinek megfelelő növekedésén keresztül) mérsékli.

A felső korlátot az „1” értékhez igazítva, a pszeudo-R^2 tartalmú normalizált érték

$$R_{GV}^2 = \frac{T_{GV}}{Var_r Var_D} = 1 - \frac{C_T^2}{Var_r Var_D} = 1 - Corr^2\left(r, D\right) \ , \qquad /13/$$

mely a Pearson-determináció komplementereként az eloszlás egyenletességének a csökkenését adja az aktuális eloszlás ismeretében.

A Theil-mátrix elemeinek értelmezését segítik a felismerések, miszerint $Var_r = 2GE\left(2\right)$ az Y jövedelmek V_Y relatív szórásának a négyzete és $Var_D = Var_{lnY}$. Ezzel a Theil-mátrix tartalma

$$\mathbf{C}_T = \begin{array}{c|cc} \text{Változó} & r & D \\ \hline r & V_Y^2 = 2GE\left(2\right) & GE\left(1\right) + GE\left(0\right) \\ D & GE\left(1\right) + GE\left(0\right) & Var_{lnY} \end{array} \ . \qquad /14/$$

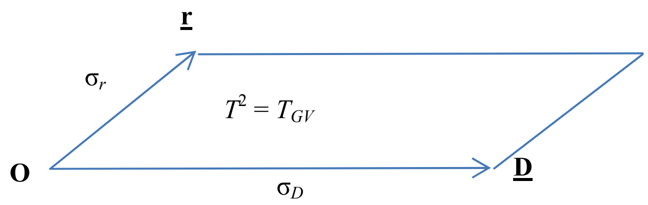
A Theil-variancia tehát egyidejűleg négy egyenlőtlenségi index hatását is magában foglalja, nevezetesen:[8]

1. V_Y : a jövedelem variációs koefficiense (relatív szórása),
2. Var_{lnY}: a logaritmikus jövedelmek varianciája,
3. $GE(1)$: a Theil-redundanciaindex,
4. $GE(0)$: a Theil-mean-logarithmic-deviation index.

Újra hangsúlyozzuk, hogy a *2.*, *3.*, *4.* indexek logaritmus alapúak, tehát hatásuk tompítja az aszimmetriából eredő torzítást statisztikai tesztek alkalmazásakor. Geometriai interpretációban T_{GV} annak a paralelogrammának a négyzetes területe, melyet az n-dimenziós $\underline{\mathbf{O}}$ origóból az $\underline{\mathbf{r}}$ és a $\underline{\mathbf{D}}$ pontokba mutató vektorok feszítenek ki, ahol σ_r és σ_D a megfelelő vektor hossza, γ pedig a két vektor hajlásszöge:

$$\begin{aligned} T_{GV} &= \left(\sigma_r \sigma_D \sin\gamma\right)^2 = Var_r Var_D \left(1 - cos^2\gamma\right) = \\ &= Var_r Var_D - Var_r Var_D Corr^2\left(r, D\right) = Var_r Var_D - Cov^2\left(r, D\right). \end{aligned} \qquad /15/$$

1. ábra. A Theil-variancia geometriai interpretációja



Az 1. ábra felhívja a figyelmet, hogy a többváltozós egyenlőtlenség mértékében a változók korrelációjának (ferde szögű „oblique” tengelyeknek) az alkalmazása – akár egy, akár több dimenzióban – elengedhetetlen. Az 1. ábra mutatja, hogy ha $Corr(r,D)$ emelkedik, vagyis a hajlásszög csökken, akkor a redundancia növekedésének hatására az általánosított variancia mértéke csökken és megfordítva. Egy dimenzióról kettőre térve az ábrázolás már nem lehetséges, mert a változók (tengelyek) száma négyre emelkedik.

Modellpélda

A kalkulációk könnyű ellenőrizhetősége érdekében tekintsünk egy száztagú társadalmat, ahol a rendezett jövedelmi konfiguráció: $\mathbf{Y} = [1,2,3,\ldots,98,99,100]$. Az át-

[8] Itt jegyezzük meg, hogy $GE(2)$ egyben a Hirschman–Herfindahl-indexet adja. A négy index tulajdonságainak részleteit lásd *Cowell* [2009].

lag és a medián egyaránt 50,5, $GE(1)$ = 0,18827, $GE(0)$ = 0,28458, V_Y^2 = 0,32673, $Var_{\ln Y}$ = 0,85267.[9]

A Theil-kovarianci a GE-felbontása az előbbi adatokkal

$$C_T = 0{,}18827 + 0{,}28458 = 0{,}47285 \ , \qquad /16/$$

majd a Theil-mátrix

$$\mathbf{C}_{\text{Theil}} = \begin{array}{c|cc} \text{Változó} & r & D \\ \hline r & 0{,}32673 & 0{,}47285 \\ D & 0{,}47285 & 0{,}85267 \end{array} \qquad /17/$$

és végül az általánosított Theil-variancia

$$T_{GV} = 0{,}32673 \cdot 0{,}85267 - 0{,}47285^2 = 0{,}05501 \ , \qquad /18/$$

melynek normált pszeudo-R^2 értéke

$$R_{GV}^2 = \frac{0{,}05501}{0{,}32673 \cdot 0{,}85267} = 0{,}19744 \ . \qquad /19/$$

Az egyenlőtlenség abszolút mértéke 0,05501, ami az aktuális eloszlás 19,744 százalékos távolságát jelzi a totális egyenlőség – a totális bizonytalanság – állapotától.

Mindezek alapján az egyenlőtlenség intenzitása mint pszeudo-R érték adódik

$$R_{GV} = \sqrt{0{,}19744} = 0{,}44434 \ . \qquad /20/$$

3. A Theil-variancia diszkriminanciaanalízise[10]

A kategória kimenetű, diszkriminátor változó egyenlőtlenségre gyakorolt diszkriminatív hatásának jellemzésére, a társadalmat $g = 1,2,\ldots,G$ csoportra bontjuk. A csoportok száma és kialakításuk módja tetszőleges lehet. Az egyszerűség kedvéért e

[9] A Gini-index értéke összehasonlításul: 0,33333.

[10] Összehasonlításul a Gini- és a generalized entropy (GE) -felbontások összefoglalását lásd például *Mussard–Seyte–Terraza* [2003].

fejezetben előbb csak két szomszédos csoportra, szegényekre és nem szegényekre, tehát egy alsó és egy felső szegmensre tagolunk, adott szegénységi küszöb alapján.

3.1. A Theil-mátrix külső-belső felbontása és a Wilks' lambda-hányados

A csoporthatások számításának alapja a Theil-mátrix külső-belső dekompozíciója (a variancia külső-belső felbontásának az analógiájára)

$$\mathbf{C}_{\text{Theil}} = \mathbf{C}_{\text{Külső}} + \mathbf{C}_{\text{Belső}} \,, \qquad /21/$$

ahol $\mathbf{C}_{\text{Külső}}$ a külső, $\mathbf{C}_{\text{Belső}}$ pedig a belső kovarianciamátrix jelölése. Tartalmilag a külső kovarianciamátrix a csoportátlagokkal kisimított változók kovarianciamátrixa, míg a belső kovarianciamátrix az átlagos csoporton belüli kovarianciamátrix.[11]

A belső kovarianciamátrix formálisan a

$$\mathbf{C}_{\text{Belső}} = \sum_{g=1}^{G} n_g \mathbf{C}_{T_g} \qquad /22/$$

súlyozott átlag, ahol n_g a g csoport népességaránya (összegük = 1), $\mathbf{C}_{Tg}$ pedig a g csoport átlagolandó Theil-mátrixa. A belső kovarianciamátrix elemei rendre a szegény és a nem szegény kovarianciamátrixok megfelelő elemeinek a súlyozott átlagai, súlyként a népességi arányokat használva.

Legyen a szegénységi küszöb a medián jövedelem 60 százaléka, ami az alsó 30 százalék népességet klasszifikálja szegényként. A küszöb tehát 30-70 százalék arányban bontja ketté a társadalmat, így a belső Theil-mátrix

$$\mathbf{C}_{\text{Belső}} = 0,3 \cdot \mathbf{C}_{\text{Szegény}} + 0,7 \cdot \mathbf{C}_{\text{Nemszegény}} = \begin{array}{c|cc} \text{Változó} & r & D \\ \hline r & 0,12087 & 0,13163 \\ D & 0,13163 & 0,28708 \end{array} \,, \qquad /23/$$

ahol az átlagolandó csoporton belüli Theil-mátrixok

$$\mathbf{C}_{\text{Szegény}} = \begin{array}{c|cc} \text{Változó} & r & D \\ \hline r & 0,02938 & 0,13194 \\ D & 0,13194 & 0,69913 \end{array} \text{ és} \qquad /24/$$

[11] A kovarianciamátrix külső-belső felbontására, és a Wilks' lambda származtatására egy számpéldát ad *Hajdu* ([2003] 101–105. old.).

$$\mathbf{C}_{\text{Nemszegény}} = \begin{array}{c|cc} \text{Változó} & r & D \\ \hline r & 0{,}16008 & 0{,}13150 \\ D & 0{,}13150 & 0{,}11049 \end{array} . \qquad /25/$$

A totális Theil-mátrix felbontása tehát (a külső Theil-mátrixot kivonással határozva meg):

$$\mathbf{C}_{\text{Theil}} = \begin{array}{c|cc} \text{Totális} & r & D \\ \hline r & 0{,}32673 & 0{,}47285 \\ D & 0{,}47285 & 0{,}85267 \end{array} = \begin{array}{c|cc} \text{Külső} & r & D \\ \hline r & 0{,}20586 & 0{,}34122 \\ D & 0{,}34122 & 0{,}56559 \end{array} + \begin{array}{c|cc} \text{Belső} & r & D \\ \hline r & 0{,}12087 & 0{,}13163 \\ D & 0{,}13163 & 0{,}28708 \end{array} . \qquad /26/$$

A csoportosítás irrelevanciáját a többváltozós statisztikai irodalom szerint a Wilks' lambda-hányados méri, mely a belső általánosított variancia és a totális általánosított variancia hányadosa, tehát a belső és a totális kovarianciamátrixok determinánsainak a hányadosa

$$\textit{Wilks' lambda} = \frac{\det \mathbf{C}_{\text{Belső}}}{\det \mathbf{C}_{\text{Theil}}} . \qquad /27/$$

A Wilks' lambda értelme a kategóriák által a totális varianciából meg nem magyarázott rész, ezért komplementere tartalmilag varianciahányados (variance explained) típusú VE mutató.

A belső Theil-mátrixból a Wilks' lambda értéke[12]

$$\textit{Wilks' lambda} = \frac{0{,}12087 \cdot 0{,}28708 - 0{,}13163^2}{0{,}05501} = 0{,}31585 , \qquad /28/$$

ahonnan a VE mutató alapján az alkalmazott szegmentáció (jelenleg a szegénységi küszöb) diszkrimináló ereje

$$VE = 1 - 0{,}31585 = 0{,}68415 . \qquad /29/$$

[12] A belső általánosított variancia pszeudo-R^2 normálása: $R^2 = \frac{0{,}13163^2}{0{,}12087 \cdot 0{,}28708} = 0{,}4993$. E szerint a belső variancia 49,93 százalékban közelítette meg a lehetséges maximális értékét.

A szegénységi küszöb 68,415 százalékban magyarázza az egyenlőtlenséget, és kapcsolata az egyenlőtlenséggel $\sqrt{VE} = 0,82714$ intenzitású.

3.2. A kanonikus korreláció megközelítés

Bár VE a külső egyenlőtlenség hatását méri, számításához a külső Theil-determináns értékén keresztül nem vezet út, mert

1. a Theil-mátrix additív felbontása a komponensek determinánsaira nem érvényes

$$det\,\mathbf{C}_{\text{Theil}} \neq det\,\mathbf{C}_{\text{Külső}} + det\,\mathbf{C}_{\text{Belső}} \ . \qquad /30/$$

2. Ha a csoportok száma megegyezik a változók számával (mint példánkban könnyen ellenőrizhető), akkor a külső determináns értéke mindig zéró, akkor is, ha a külső varianciák-kovarianciák nem zérók.[13]

Mindazonáltal a komplementer Wilks' hányados külső tartalma és külső számítási módja megadható a következő alternatív megközelítésben. Tekintsük az *r* relatív jövedelem és a *D* log-hozam lineáris kombinációját

$$\Delta = w_1 r + w_2 D \ . \qquad /31/$$

A *w* súlyokat úgy választjuk meg, hogy Δ belső varianciája legyen az „1” értéken normált és hozzá képest a külső variancia értéke maximált. A külső variancia maximált értéke nem más, mint a $\mathbf{C}_{\text{Belső}}^{-1}\mathbf{C}_{\text{Külső}}$ ANOVA-mátrix pozitív λ sajátértéke.[14]

Esetünkben λ = 2,16609, tehát Δ varianciája (1 + 2,16609), a külső variancia aránya pedig definíció szerint a négyzetes kanonikus korreláció[15]

$$Rho^2 = \frac{2,16609}{3,16609} = 0,68415 \ , \qquad /32/$$

ami a 0,31585 Wilks' hányados VE komplementere, aminek pozitív gyöke a Rho-kanonikus korreláció[16]

[13] Például két csoport két „külső” pontja maradék nélkül magyarázható két változóval, azaz két paraméterrel.

[14] A kanonikus korreláció és a diszkriminanciaanalízis kapcsolatának elméleti és számítási részleteit illetően lásd például *Hajdu* ([2003], [2010]). A maximált külső varianciát biztosító *w* súlyokkal Δ szokásos megnevezése: kanonikus diszkriminanciaváltozó.

[15] Ez az eredmény értelemszerűen megegyezik a /29/ szerint számítottal.

[16] E korreláció másik megközelítésben a szegény és nemszegény dummy változó, valamint az *r* és *D* változók által adott két változókör közötti kapcsolat szorosságát méri.

$$Rho = \sqrt{0,68415} = 0,82714 \quad /33/$$

Egydimenziós-kétváltozós esetben hangsúlyozzuk, hogy két olyan csoportra, melyek belül nem szóródnak, Rho értéke mindig 1.[17] Ez a helyzet akkor is, ha mindenki jövedelme egyenlő, kivéve egyetlen outliert. Így *Rho* = 1 akkor is, mikor az egyetlen outlier kap mindent.

3.3. Homogenitásvizsgálat

A klasszikus Box-M-statisztika alapján a sokasági kovarianciamátrixok egyezőségének a tesztelésére is lehetőség nyílik. Szegényekre és nem szegényekre bontva a társadalmat, a hipotézis

$$H_0 : \Sigma_{\text{Szegény}} = \Sigma_{\text{Nemszegény}} , \quad /34/$$

ahol Σ a sokasági csoporton belüli kovarianciamátrix jelölése. A hipotézis tesztelésére szolgáló Box-M-statisztika likelihood-arány (LR) próba, melynek formulája:[18]

$$Box\text{-}M = \sum_{g=1}^{G} \left(n_g - 1\right)\left[ln\, det\, \mathbf{C}_{Belső} - ln\, det\, \mathbf{C}_g \right], \quad /35/$$

ahol példánkban (a kovarianciákat itt korrigáltan számítva és egy tizedesre kerekítve[19]):

1. $ln\ det(\mathbf{C}_{\text{Belső}}) \approx -4,0,$
2. $ln\ det(\mathbf{C}_{\text{Szegény}}) \approx -5,7,$
3. $ln\ det(\mathbf{C}_{\text{Nemszegény}}) \approx -7,8.$

A homogenitás-tesztstatisztika értéke a fenti eredményekkel

$$M = 29 \cdot \left(-4,0 + 5,7\right) + 69 \cdot \left(-4,0 + 7,8\right) = 311,5 \ , \quad /36/$$

amelynek szignifikanciaértéke *F*-tesztet alkalmazva és kerekítve 0,000, tehát a csoporton belüli Theil-varianciák minden szokásos szignifikanciaszinten különböznek egymástól.

[17] Visszautalunk a 13. lábjegyzetre.

[18] Az M-statisztika tesztelését lásd *Michaleczky* ([1986] 67–68. old.), az LR-teszt elv leírását pedig *Hunyadi* ([2001] 369–376. old.).

[19] A kovariancia nevezőjében a mintaméretet itt 1-gyel csökkentve, a belső kovarianciamátrixban az ún. pooled kovarianciákat számítjuk.

Az M-statisztika additív struktúrája lehetővé teszi végül az egyes „kategóriák” százalékos hozzájárulásainak megadását is a belső egyenlőtlenség mértékén belül:

$$M_{\text{Szegény}} = \frac{29\cdot(-4.0+5.7)}{29\cdot(-4.0+5.7)+69\cdot(-4.0+7.8)} = 15.8\%\ , \qquad /37/$$

$$M_{\text{Nemszegény}} = \frac{69\cdot(-4{,}0+7{,}8)}{29\cdot(-4{,}0+5{,}7)+69\cdot(-4{,}0+7{,}8)} = 84{,}2\%\ . \qquad /38/$$

Ezek szerint a szegények köre 15,8 százalék arányban járul hozzá a belső egyenlőtlenség mértékéhez.

4. A Theil-mátrix többdimenziós kiterjesztése

Bővítsük a dimenziók számát háromra: jövedelem, kiadás, vagyon.[20] Jelölésünk szerint:

– a relatív jövedelmek rendre $j = r_{\text{jövedelem}}$, $k = r_{\text{kiadás}}$, $v = r_{\text{vagyon}}$,
– a log-hozamok pedig $J = D_{\text{jövedelem}}$, $K = D_{\text{kiadás}}$, $V = D_{\text{vagyon}}$.

A $p = 3$ dimenziós vizsgálat egy hatváltozós esethez vezet, ahol a $\mathbf{C}_{(6,6)}$ Theil-kovarianciamátrix:

$$\mathbf{C}_{T_{(6,6)}} = \begin{array}{c|cccccc} \text{Változó} & j & k & v & J & K & V \\ \hline j & C_{jj} & C_{jk} & C_{jv} & C_{jJ} & C_{jK} & C_{jV} \\ k & C_{kj} & C_{kk} & C_{kv} & C_{kJ} & C_{kK} & C_{kV} \\ v & C_{vj} & C_{vk} & C_{vv} & C_{vJ} & C_{vK} & C_{vV} \\ J & C_{Jj} & C_{Jk} & C_{Jv} & C_{JJ} & C_{JK} & C_{JV} \\ K & C_{Kj} & C_{Kk} & C_{Kv} & C_{KJ} & C_{KK} & C_{KV} \\ V & C_{Vj} & C_{Vk} & C_{Vv} & C_{VJ} & C_{VK} & C_{VV} \end{array}\ . \qquad /39/$$

[20] Irodalmi összehasonlításul: A GE-index többdimenziós kiterjesztéseinek különféle módjai olvashatók többek között: *Maasoumi* [1986], [1998] *Tsui* [1995], [1999], *Vega-Urrutia–Volij* [2011], *Lugo* [2005]. Egy másik, a Gini-index többdimenziós általánosítását adja *Gajdos–Weymark* [2003]. Egy új, ún. „hybrid” többdimenziós egyenlőtlenségi mértéket definiál *Araar* [2009], míg a többdimenziós egyenlőtlenségi összehasonlítások kérdését *Duclos–Sahn–Younger* [2009] tárgyalja.

Általában p-dimenziós esetben a Theil-mátrix $\mathbf{C}_{(2p,2p)}$, melynek determinánsa értelemszerűen a „kiterjesztett” Theil-variancia egyenlőtlenségi mérték

$$T_{GV} = det\left(\mathbf{C}_{T_{(2p,2p)}}\right) . \qquad /40/$$

Az elemi kovarianciák értelmezése (a többi kovariancia értelmezése analóg):

1. C_{jj}: a relatív jövedelem varianciája,
2. C_{kj} : a relatív kiadás és a relatív jövedelem kovarianciája,
3. C_{Jj} : a jövedelmi változók kovarianciája,
4. C_{Kj} : a kiadási log-hozam és a relatív jövedelem kovarianciája,
5. C_{JJ} : a jövedelmi log-hozam varianciája,
6. C_{KJ} : a kiadási és jövedelmi log-hozamok kovarianciája.

T_{GV} mint determináns maximális értékét korrelálatlanság esetén veszi fel, ekkor a kovarianciamátrix diagonális, determinánsa a diagonális varianciák szorzata. Ezzel a (0,1) intervallumra való normálása

$$R^2_{GV} = \frac{T_{GV}}{\prod_{t=1}^{p} Var_{r_t} \prod_{t=1}^{p} Var_{D_t}} . \qquad /41/$$

Kiemelendő, hogy többdimenziós esetben a $T_{GV(2p,\ 2p)}$ kiterjesztett Theil-variancia nemcsak a változóközi, hanem a dimenzióközi és a keresztkorrelációkat is figyelembe veszi.

Csoportosítás esetén a $\mathbf{C}^{-1}_{\text{Belső}}\mathbf{C}_{\text{Külső}}$ ANOVA-mátrixnak általánosságban $m = min\{2p, G-1\}$ pozitív λ_δ sajátértéke van, melyek a Δ_δ változók külső varianciái. A Δ_δ dimenziók relevanciája standard diszkriminanciaanalízis (discriminant analysis) eljárással tesztelhető, és a dimenziók stepwise algoritmussal szelektálhatók. A Wilks' lambda struktúrája a p-dimenziós esetben (lásd például *Hajdu* [2010]):[21]

$$\textit{Wilks' lambda} = \prod_{\delta=1}^{m} \frac{1}{1+\lambda_\delta} . \qquad /42/$$

[21] A többdimenziós egyenlőtlenségi dekompozíció módszertani kérdéseit lásd például *Zheng* [2005], *Cowell–Fiorio* [2010], *Kobus* [2011].

4.1. Háztartási költségvetési példa, településtípus szerinti dekompozícióval

Illusztratív céllal háztartásokat tekintünk, melyeknél most: *j* az évi nettó jövedelmet, *k* az évi kiadást, *v* pedig tulajdont (lakás+gépkocsi) jelent. Az alkalmazott csoportosítás példánkban a település típusa: *Budapest*, *Nagyváros*, *Többi város*, *Községek*.

A Theil-variancia értéke[22]

$$T_{GV} = det\left(\mathbf{C}_{T_{(6,6)}}\right) = 0,0000198526 \qquad /43/$$

és a településtípus szerint csoportosítva az ANOVA-mátrix három pozitív sajátértéke rendre

$$\lambda_1=0,13671, \quad \lambda_2=0,01977, \quad \lambda_3=0,00220.$$

A Wilks' lambda ebből

$$Wilks'\ lambda = \frac{1}{1,13671 \cdot 1,01977 \cdot 1,00220} = 0,8608\,. \qquad /44/$$

Így a településtípus a jövedelem, kiadás, vagyon együttes egyenlőtlenségből 13,92 százalékot magyaráz, kapcsolata pedig az „oblique pontfelhők” külső szóródásával a globális pontfelhő körül $Rho = \sqrt{0,1392} = 0,3731$ intenzitású.

4.2. Homogenitásvizsgálat

Kettőnél több (a példabeli négy) sokasági kovarianciamátrix azonosságát állító hipotézis:

$$H_0 : \Sigma_{Budapest} = \Sigma_{Nagyváros} = \Sigma_{Többiváros} = \Sigma_{Község}\ , \qquad /45/$$

ahol Σ a sokasági kovarianciamátrixot jelöli. A megfelelő mintabeli statisztikák (településtípus szerinti log-determinánsok) értékei következők:[23]

[22] Ennek normáló tényezője (a Theil-mátrix főátló elemek szorzata) 0,0123, de a csoporthatás elemzése szempontjából a nagyságrendnek nincs jelentősége.

[23] $ln\ det(\mathbf{C}_{Belső}) = -10,975$, ami az ún. „*pooled*” kovarianciamátrix log-determinánsa!

$$ln\ det(\mathbf{C}_{\text{Budapest}}) = -8{,}183\ ,$$
$$ln\ det(\mathbf{C}_{\text{Belső}}) - ln\ det(\mathbf{C}_{\text{Budapest}}) = -2{,}792\ ,$$

$$ln\ det(\mathbf{C}_{\text{Nagyváros}}) = -11{,}865\ ,$$
$$ln\ det(\mathbf{C}_{\text{Belső}}) - ln\ det(\mathbf{C}_{\text{Nagyváros}}) = 0{,}890\ ,$$

$$ln\ det(\mathbf{C}_{\text{Többiváros}}) = -12{,}355\ ,$$
$$ln\ det(\mathbf{C}_{\text{Belső}}) - ln\ det(\mathbf{C}_{\text{Többiváros}}) = 1{,}380\ ,$$

$$ln\ det(\mathbf{C}_{\text{Községek}}) = -12{,}616\ ,$$
$$ln\ det(\mathbf{C}_{\text{Belső}}) - ln\ det(\mathbf{C}_{\text{Községek}}) = 1{,}641\ .$$

A településtípusok népességarányai a mintában rendre 0,158, 0,228, 0,272, 0,342. $Box\text{-}M = 4907{,}486$, az F-teszten alapuló 0,000 szignifikanciaértékkel, tehát a csoporton belüli Theil-varianciák minden szokásos szignifikanciaszinten különböznek egymástól.

Egyedül Budapest hatása negatív –2,792, erősen átlag alatti. Valamennyi településtípus hatását pozitív értékű skálán rangsorolandó, Budapest hatását pozitív előjellel minimális értékként rögzítve, a távolságőrző transzformált skála értékei:

– Budapest = 2,792,
– Nagyváros = 6,474,
– Többi város = 6,964,
– Község = 7,225, ahol például 7,225 = 1,641 + 2abs(–2,792).

A településtípusok súlyozott, relatív hozzájárulásai a belső egyenlőtlenséghez rendre:

$$M_{\text{Budapest}} = \frac{0{,}158 \cdot 2{,}792}{0{,}158 \cdot 2{,}792 + 0{,}228 \cdot 6{,}474 + 0{,}272 \cdot 6{,}964 + 0{,}342 \cdot 7{,}225} = 11{,}9\%\ , \qquad /46/$$

$$M_{\text{Nagyváros}} = \frac{0{,}228 \cdot 6{,}474}{0{,}158 \cdot 2{,}792 + 0{,}228 \cdot 6{,}474 + 0{,}272 \cdot 6{,}964 + 0{,}342 \cdot 7{,}225} = 27{,}6\%\ , \qquad /47/$$

$$M_{\text{Többiváros}} = \frac{0{,}272 \cdot 6{,}964}{0{,}158 \cdot 2{,}792 + 0{,}228 \cdot 6{,}474 + 0{,}272 \cdot 6{,}964 + 0{,}342 \cdot 7{,}225} = 29{,}7\%\ , \qquad /48/$$

$$M_{\text{Község}} = \frac{0{,}342 \cdot 7{,}225}{0{,}158 \cdot 2{,}792 + 0{,}228 \cdot 6{,}474 + 0{,}272 \cdot 6{,}964 + 0{,}342 \cdot 7{,}225} = 30{,}8\%\ . \qquad /49/$$

A belső egyenlőtlenséghez Budapest járul hozzá a legkisebb (11,9%) mértékben, és a községek a legnagyobb (30,8%) mértékben.

5. Theil-variancia alapú szegénységi mértékek

A szegénységi P mérték általánosságban a szegénység kiterjedtségének, intenzitásának és az eloszlásának az eredője. A mérőszám formális „P” megadása – adott z küszöb mellett – vagy a *csonkolt*, vagy a *cenzorált* eloszlásra épül. Míg a csonkolt eloszlás elhagyja a küszöb fölötti tagokat, addig a cenzorált eloszlás megtartja, de értékeiket a küszöb szintjével helyettesíti. A Z_j dimenzió cenzorálása a z_j küszöb mellett:

$$y_{ji} = min\left\{Z_{ji}, z_j\right\}, \quad i = 1, 2, ..., n \ . \qquad /50/$$

Jelen cikk a cenzorált elvre építve definiál többváltozós-többdimenziós P mértéket:[24] A szegénység többváltozós, többdimenziós mértéke definíciónk szerint a cenzorált **y** eloszlások együttes, többdimenziós Theil-varianciája. Speciálisan egydimenziós-kétváltozós esetben az y eloszlás cenzorált Theil-mátrixa[25]

$$\mathbf{C}_T^c = \mathbf{C}_T\left(y\right) = \begin{array}{c|cc} \text{Változó} & r^c & D^c \\ \hline r^c & Var_r^c & C_T^c \\ D^c & C_T^C & Var_D^C \end{array}, \qquad /51/$$

melynek determinánsa

$$T_{GV}^c = det\left(\mathbf{C}_T^c\right) = Var_r^c \cdot Var_D^c - C_T^{c^2} \ , \qquad /52/$$

ami normált pszeudo-R^2 változatban

[24] A cenzorált eloszlás szegénységi alkalmazását *Hamada–Takayama* [1978] és *Takayama* [1979] vezette be az irodalomba. A Sen–Shorrocks–Thon (SST) (*Shorrocks* [1995] módon korrigált Sen-index) a legismertebb egydimenziós, és az *Alkire–Foster* [2009] módon korrigált Foster–Greer–Thorbecke- (FGT-) index), illetve a *Lugo–Maasoumi* [2008] -indexek a többdimenziós alkalmazások. A Lugo–Maasoumi-indexcsalád információelméleti megalapozottságú, mely speciális esetként tartalmazza a *Tsui* [2002] és a *Bourgougni–Chakravarty* [2003] indexeket is. A többdimenziós szegénységi indexek összefoglaló áttekintését egyébként lásd *Ravallion* [2011]. A többdimenziós témakörben meghatározó további tanulmányok: *Anand–Sen* [1997], *Chakravarty–Mukherjee–Renade* [1998], *Atkinson* [2003], *Thorbecke* [2008], *Chakravarty–Silber* [2008], *Kakwani–Silber* [2008].

[25] A klasszikus egydimenziós szegénységi index-elvek összefoglaló bemutatását lásd *Foster–Sen* [1997], *Zheng* [1997]. A módszertan fejlődését illetően a következő indexeket emeljük ki: *Watts* [1968], *Sen* [1976], *Anand* [1977], *Hamada–Takayama* [1978], *Thon* [1979], *Kakwani* [1980], *Takayama* [1979], *Clark–Hemming–Ulph* [1981], *Chakravarty* [1983], *Blackorby–Donaldson* [1980], *Foster–Greer–Thorbecke* [1984], *Hagenaars* [1987], *Atkinson* [1987], *Shorrocks* [1995].

$$R_P^2 = 1 - \frac{C_T^{c^2}}{Var_r^c \cdot Var_D^c} \quad . \qquad /53/$$

Mivel a cenzorált eloszlás csak a szegényjövedelmeket és a küszöb szintjét ismeri, így csak az aktuális küszöbalattiság információit tükrözi. A cenzorált eloszlás általánosított varianciája tehát növekvő értékkel a szegénység növekvő fokát jelzi.

Az R_P^2 szegénységi mérték örökli a Theil-variancia tulajdonságait, és a cenzorált eloszláson történő kalkulálásából eredő további jellemzői a következők:

1. Az R_P^2 szegénységi mérték eliminálja az index formulájából a szegények explicit létszámarányát, mivel ez a hatás implicit módon az értékében érvényesül.

2. A „z” küszöb emelése több szegénységet indukál, több szegénységi információval, melynek cenzorált $Var_r Var_D$ felső határa együtt nő a küszöbbel.

3. A cenzorálás a társadalmat küszöb alattiakra és éppen a küszöb szintjén levőkre bontja, így a külső-belső kovariancia-felbontás szegény vs. küszöb tekintetben értendő.

4. A belső varianciát csak a küszöb alattiak szóródása és létszámaránya mozgatja.

5. A külső variancia a küszöb alatti átlagos szegény és a küszöbszint varianciáját méri, így jelentése az átlagos szegény által a küszöb szintjén élőkkel szembeni depriváció mértéke.

6. A társadalmat exogén változók szerint csoportosítva (város-vidék, férfi-nő, aktív-inaktív) a szegénység foka külső-belső szempontból és a belső szegénységhez való hozzájárulás mértéke tekintetében is dezaggregálható, jellemezhető.

7. A szegénységi küszöb megadható dimenziókra szeparáltan, vagy az egyes dimenziók valamely súlyozott kombinációjára aggregáltan is.

8. Az R_P^2 szegénységi mérték figyeli a dimenziók korrelációit és aszimmetrikus voltát is.

Modellpélda

A százfős példában a cenzorált eloszlás: $y = \{1,2,\ldots,29,30|30,30,\ldots,30\}$, melyre a Theil-mátrix

$$\mathbf{C}_T^c = \begin{array}{c|cc} \text{Változó} & r^c & D^c \\ \hline r^c & 0{,}10127 & 0{,}18627 \\ D^c & 0{,}18627 & 0{,}38463 \end{array} \qquad /54/$$

a cenzorált Theil-variancia

$$T_{GV}^c = 0{,}10127 \cdot 0{,}38463 - 0{,}18627^2 = 0{,}004257 \ , \qquad /55/$$

melynek pszeudo-R^2 értéke

$$R_P^2 = 1 - \frac{0{,}18627^2}{0{,}10127 \cdot 0{,}38463} = 0{,}10929 \ , \qquad /56/$$

ahonnan

$$R_P = \sqrt{0{,}10929} = 0{,}3306 \ . \qquad /57/$$

Az alkalmazott szegénységi küszöb szintje által az aktuális eloszlás szegénységi mértékéből megmagyarázott hányad 10,929 százalék, és a szegénység 0,3306 intenzitással valósul meg.

A szegénységi mérték szegény vs. küszöb csoportközi elemzése a cenzorált Theil-mátrix

$$\mathbf{C}_{\text{Theil}}^c = \mathbf{C}_{\text{Külső}}^c + \mathbf{C}_{\text{Belső}}^c \qquad /58/$$

felbontásán alapulva, egy általánosított szegénységi arány és egy általánosított deprivációarány-mutatóhoz vezet el, a következő úton.

Példánkban a küszöb kettébontja a társadalmat 30-70 százalék arányban, ahol a belső Theil-mátrix

$$\mathbf{C}_{\text{Belső}}^c = 0{,}3 \cdot \mathbf{C}_{\text{Szegény}} + 0{,}7 \cdot \mathbf{0}_z = \begin{array}{c|cc} \text{Változó} & r^c & D^c \\ \hline r^c & 0{,}03416 & 0{,}07793 \\ D^c & 0{,}07793 & 0{,}20974 \end{array} \ , \qquad /59/$$

ahol az átlagolandó csoporton belüli Theil-mátrixok

$$\mathbf{C}_{\text{Szegény}} = \begin{array}{c|cc} \text{Változó} & r^c & D^c \\ \hline r^c & 0{,}11387 & 0{,}25976 \\ D^c & 0{,}25976 & 0{,}69913 \end{array} \text{ és} \qquad /60/$$

$$\mathbf{0}_z = \begin{array}{c|cc} \text{Változó} & r^c & D^c \\ \hline r^c & 0 & 0 \\ D^c & 0 & 0 \end{array} . \qquad /61/$$

Így a cenzorált belső Theil-variancia

$$T^c_{GV_{\text{Belső}}} = det\left(\mathbf{C}^c_{\text{Belső}}\right) = 0{,}03416 \cdot 0{,}20974 - 0{,}07793^2 = 0{,}001092 \ , \qquad /62/$$

mellyel a cenzorált belső pszeudo-R^2 értéke

$$R^2_{\text{Belső}} = \frac{0{,}001092}{0{,}03416 \cdot 0{,}20974} = 15{,}241\% \ . \qquad /63/$$

Értelmét tekintve a cenzorált belső R^2 általánosított szegénységi arány. Ugyanis a belső variancia átlagos csoporton belüli variancia, ahol a cenzorált eloszlásban a nem szegények kovarianciamátrixa definíció szerint zéró értéken rögzített ($\mathbf{0}_z$), ezért a belső variancia nő, ha *a*) emelkedik a szegények aránya, vagy ha *b*) nő a szegények körében a Theil-variancia. Az általánosított szegénységi arány példánkban 15,241 százalék, ami kisebb mint a H = 30 százalék standard „head-count-ratio" létszámarány.

Tekintsük most a szegénység harmadik tényezőjeként *c*) a szegények szegénységi küszöbbel szemben érzett deprivációjának a fokát azzal a követelménnyel, hogy legyen érzékeny a nem szegények népességi arányára is, akik nem depriváltak a küszöbbel szemben. Jelölje e hatást az IG (implicit gap, azaz implicít rés), és tételezzünk fel multiplikatív kapcsolatot a szegénységi komponensek között. Ekkor implicit módon a küszöbbel szembeni IG depriváció definíciónk szerint

$$R_P^2 = R^2_{\text{Belső}} \cdot IG_{\text{Külső}} \ . \qquad /64/$$

Az $IG_{\text{Külső}} = \frac{10{,}2}{15{,}9} = 72\%$ inflátor (deflátor) a szegények szegénységi küszöbbel szemben érzett deprivációjának a foka.

Többdimenziós szegénységi dekompozíció

A dimenziók számát többre – a háztartási költségvetési példánkban már alkalmazott háromra (jövedelem, kiadás, vagyon) – bővítve, a szegénységi mérték kalkulálása és településtípusok szerinti felbontása kiterjesztett cenzorált Theil-mátrix alkalmazásával a következő. A dimenziókat a mediánérték 60 százalékánál cenzorálva a *Budapest*, *Nagyváros*, *Többi város*, *Községek* csoportosítás hatváltozós diszkriminanciaanalízisének három sajátértéke rendre: 0,1090, 0,0067, 0,0006, amiből a Wilks' lambda = 0,8952, komplementerének a gyöke pedig a kanonikus korreláció: *Rho* = 0,32373. A településtípus tehát 89,52 százalékban nem magyarázza a háromdimenziós szegénységet és kapcsolata a szegénységi „kockával" *Rho* = 0,32373 intenzitású.

Az eddigiekben a háztartások éves összes kiadását és jövedelmét, valamint a teljes vagyonát tekintettük. Áttérve az egy fogyasztási egységre vetített szintre, az eredmények az alábbiak szerint módosulnak. Továbbra is a mediánérték 60 százalékánál cenzorálva a *Budapest*, *Nagyváros*, *Többi város*, *Községek* csoportok hatváltozós diszkriminanciaanalízisének három sajátértéke rendre: 0,1238, 0,0054, 0,0005, amiből a Wilks' lambda = 0,8847, a kanonikus korreláció pedig *Rho* = 0,33956. A településtípus tehát 88,47 százalékban nem magyarázza az egy fogyasztási egységre jutó háromdimenziós szegénységet, és kapcsolata e szegénységi „kockával" 0,33956 intenzitású.

*

A tanulmány egy új egyenlőtlenségi módszertant javasol, melynek alkalmazása más területeken (például a szegénységi elemzésekben, a relatív depriváció és a társadalmi kirekesztés mérésében, de az információelméletben, vagy az adatbányászatban) is új módszertant eredményezhet. Az eljárás lényegében egy sokdimenziós „oblique térben" húzódó pontfelhő varianciáját méri kompozit módon, egyenlőtlenségi tartalommal, entrópiaelméleti alapokon. A kulcsformula a többváltozós statisztika „generalized variance" mértéke, mely esetünkben egy speciális entrópia tartalmú kovarianciamátrixra vonatkozik. Mivel a kovarianciamátrix általánosságban csoportok esetén dezaggregálható belső és külső faktorok összegére, ezért a javasolt egyenlőtlenségi mutató is megadható külső és belső hatások eredőjeként. Így a javasolt módszerrel vizsgálható a különböző dimenziójú szegénységi küszöbök diszkriminatív hatása a szegénységi mérték tekintetében vagy az adott társadalmi-gazdasági csoportosítás prediktív ereje. Mindezen túl az egyes csoportok relatív hozzájárulása a belső egyenlőtlenséghez is elemezhető.

Irodalom

ALKIRE, S. – FOSTER, J. E. [2009]: *Counting and Multidimensional Poverty Measurement.* Working Paper 32. Oxford Poverty & Human Development Initiative. Oxford.

ANAND, S. [1977]: Aspects of Poverty in Malaysia. *Review of Income and Wealth*. Vol. 23. No. 1. pp. 1–16.

ANAND, S. – SEN, A. [1997]: *Concepts of Human Development and Poverty: A Multidimensional Perspective.* Human Development Papers. United Nation. New York.

ARAAR, A. [2009]: *The Hybrid Multidimensional Index of Inequality*. Centre interuniversitaire sur le risque, les polititiques économiques et l'emploi. Working Paper 09-45. October.

ARISTONDO, O. – VEGA, L. – URRUTIA, A. [2008]: A New Multiplicative Decomposition for the Foster-Greer-Thorbecke Poverty Indices. *Bulletin of Economic Research.* Vol. 62. No. 3. pp. 259–167.

ATKINSON, A. B. [1987]: On the Measurement of Poverty. *Econometrica*. Vol. 55. No. 3. pp. 749–764.

ATKINSON, A. B. [2003]: Multidimensional Deprivation: Contrasting Social Welfare and Counting Approaches. *Journal of Economic Inequality.* Vol. 1. No. 1. pp. 51–65.

BLACKORBY, C. – DONALDSON, D. [1980]: Ethical Indices for the Measurement of Poverty. *Econometrica*. Vol. 48. No. 4. pp. 1053–1060.

BLACKORBY, C. – DONALDSON, D. [1984]: Ethically Significant Ordinal Indexes of Relative Inequality. *Advances in Econometrics.* Vol. 3. No. 4. pp. 131–147.

BOSSERT,W. – CHAKRAVARTY, S. R. – D'AMBROSIO, C. [2009]: *Measuring Multidimensional Poverty: The Generalized Counting Approach*. www.ecineq.org/ecineq_ba/papers/Dambrosio.pdf

BOURGUIGNON, F. [1979]: Decomposable Income Inequality Measures. *Econometrica*. Vol. 47. No. 4. pp. 901–920.

BOURGUIGNON, F. – CHAKRAVARTY, S. R. [2003]: The Measurement of Multidimensional Poverty. *Journal of Economic Inequality*. Vol. 1. No. 1. pp. 25–49.

CHAKRAVARTY, S. R. [1983]: Ethically Flexible Measures of Poverty. *Canadian Journal of Economics*. Vol. 16. No. 1. pp. 74–85.

CHAKRAVARTY, S. R. [1997]: On Shorroks Reinvestigation of the Sen Poverty Index. *Econometrica*. Vol. 65. No. 5. pp. 1241–1242.

CHAKRAVARTY, S. R. – MUKHERJEE, D. – RENADE, R. R. [1998]: On the Family of Subgroup and Factor Decomposable Measures of Multidimensional Poverty. *Research on Economic Inequality*. Vol. 8. pp. 175–194.

CHAKRAVARTY, S. – DEUTSCH, J. – SILBER, J. [2008]: On the Watts Multidimensional Poverty Index and its Decomposition. *World Development.* Vol. 36. No. 6. pp. 1067–1078.

CLARK, S. – HEMMING, R. – ULPH, D. [1981]: On Indices or for the Measurement of Poverty. *The Economic Journal*. Vol. 91. No. 362. pp. 515–526.

COWELL, F. [2005]: *Theil, Inequality Indices and Decomposition*. ECINEQ 2005-1. Working Paper. ECINEQ Society for the Study of Ecomonic Inequality. London.

COWELL, F. A. [1977]: *Measuring Inequality*. Phillip Allan. Oxford.

COWELL, F. A. [2009]: *Measuring Inequality. Part of the series LSE Perspectives in Economic Analysis*. Oxford University Press. Oxford.

COWELL, F. A. – KUGA, K. [1981a]: Additivity and the Entropy Concept: An Axiomatic Approach to Inequality Measurement. *Journal of Economic Theory*. Vol. 25. No. 1. pp. 131–143.

COWELL, F. A. – KUGA, K. [1981b]: Inequality Measurement: An Axiomatic Approach. *European Economic Review*. Vol. 15. No. 3. pp. 287–305.

COWELL, F. – FIORIO, C. [2010]: *Inequality Decompositions*. Gini Discussion Paper 4, December. Growing Inequalities' Impacts. University of Amsterdam. Amsterdam.

DAGUM, C. [1997]: A New Approach to the Decomposition of the Gini Income Inequality Ratio. *Empirical Economics*. Vol. 22. No. 4. pp. 515–531.

DARDANONI, V. [1996]: On Multidimensional Inequality Measurement. In: *Dagum, C. – Lemmi, A*. (eds.): *Research on Economic Inequality: Income Distribution. Social Welfare, Inequality and Poverty*. Vol. 6 of Research on Economic Inequality. JAI Press Inc. pp. 201–205.

DUCLOS, J.-Y. – SAHN, D. E. – YOUNGER, S. D. [2006]: Robust Multidimensional Poverty Comparisons. *The Economic Journal*. Vol. 116. No. 514. pp. 943–968.

ÉLTETŐ, Ö. – FRIGYES, E. [1968]: New Inequality Measures as Efficient Tools for Causal Analysis and Planning. *Econometrica.* Vol. 36. No. 2. pp. 383–396.

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. old.

FERGE ZS. [1969]: *Társadalmunk rétegződése: elvek és tények.* Közgazdasági és Jogi Könyvkiadó. Budapest.

FOSTER, J. E. – SHNEYEROV, A. A. [1999]: A General Class of Additively Decomposable Inequality Measures. *Economic Theory.* Vol. 14. No. 1. pp. 89–111.

FOSTER, J. E. [2007]: *A Class of Chronic Poverty Measures.* Working Paper No. 07-W01. Vanderbilt University. Nashville.

FOSTER, J. E. – GREER, J. – THORBECKE, E. [1984]: A Class of Decomposable Poverty Measures. *Econometrica*. Vol. 52. No. 3. pp. 761–767.

FOSTER, J. E. – SEN, A. [1997]: *On Economic Inequality After a Quarter Century*. Calendron Press. Oxford.

GAJDOS,T. – WEYMARK, J. [2003]: *Multidimensional Generalized Gini Indices*. Working Paper No. 16. Applied Mathematics Working Paper Series. Vanderbilt University. Nashville.

HAGENAARS, A. [1987]: A Class of Poverty Indices. *International Economic Review*. Vol. 28. No. 3. pp. 583–607.

HAJDU O. [1997]: *A szegénység mérőszámai*. KSH Könyvtár és Dokumentációs Szolgálat. Budapest.

HAJDU, O. [1999]: On the Deprivation-Sensitive Measurement of Poverty. *Hungarian Statistical Review*. Special number 3. pp. 15–22.

HAJDU O. [2003]: *Többváltozós statisztikai számítások*. Központi Statisztikai Hivatal. Budapest.

HAJDU, O. [2009]: Poverty, Deprivation, Exclusion: A Structural Equations Modelling Approach. *Hungarian Statistical Review*. Special number 13. pp. 90–102.

HAJDU O. [2010]: Sajátértékek a statisztikában. *Statisztikai Szemle*. 88. évf. 7–8. sz. 773–788. old.

HAMADA, K. – TAKAYAMA, N. [1978]: Censored Income Distributions and the Measurement of Poverty. *Bulletin of the International Statistical Institute*. Vol. 47. No. 1. pp. 617–623.

HUNYADI L. [2001]: *Statisztikai következtetéselmélet közgazdászoknak.* Központi Statisztikai Hivatal. Budapest.

KAKWANI, N. C. [1980]: On a Class of Poverty Measures. *Econometrica*. Vol. 48. No. 2. pp. 437–446.

KAKWANI, N. C. – SILBER, J. [2008]: *Quantitative Approaches to Multidimensional Poverty Measurement*. Palgrave MacMillan. Basingstoke.

KOBUS, M. [2011]: *Attribute Decomposability of Inequality Indices via Copula*, http:/coin.wne.uw.edu.pl/mkobus/Attribute.pdf

LUGO, M. A. [2005]: *Comparing Multidimensional Indices of Inequality: Methods and Application*. ECINEQ WP 2005-14. ECINEQ Society for the Study of Ecomonic Inequality.

LUGO, M. A. – MAASOUMI, E. [2008]: *Multidimensional Poverty Measures from an Information Theory Perspective*. ECINEC 85. ECINEQ Society for the Study of Ecomonic Inequality.

MAASOUMI, E. [1986]: The Measurement and Decomposition of Multidimensional Inequality, *Econometrica*. Vol. 54. No. 4. pp. 991–997.

MIHALECZKY GY. [1986]: A többdimenziós normális eloszlás várhatóérték-vektorára és szórásmátrixára vonatkozó becslés és hipotézisvizsgálat. In: *Móri F. T. – Székely J. G.* (szerk.): *Többváltozós statisztikai analízis*. Műszaki Könyvkiadó. Budapest. pp. 49–69.

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.

RAVALLION, M. [2011]: *On Multidimensional Indices of Poverty.* Policy Research Working Paper 5580. The World Bank Development Research Group Director's Office. February.

SEN, A. K. [1976]: Poverty: An Ordinal Approach to Measurement. *Econometrica*. Vol. 44. No. 2. pp. 219–231.

SHANNON, C. E. [1948]: A Mathematical Theory of Communication. *The Bell System Technical Journal*. Vol. 27. July pp. 379–423; October pp. 623–656.

SHORROCKS, A. F. [1980]: The Class of Additively Decomposable Inequality Measures. *Econometrica.* Vol. 48. No. 3. pp. 613–625.

SHORROCKS, A. F. [1995]: Revisiting the Sen Poverty Index. *Econometrica*. Vol. 63. No. 5. pp. 1225–1230.

SPÉDER, ZS. [2002]: *A szegénység változó arcai*. Századvég Kiadó. Budapest.

SZIVÓS P. – TÓTH ISTVÁN GY. [2001]: *A jövedelmi szegénység: trend és profil 2000-ben*. *Statisztikai Szemle*. 79. évf. 10–11. sz. 848–861. old.

TAKAYAMA, N. [1979]: Poverty, Income Inequality and Their Measures: Professor Sen's Axiomatic Approach Reconsidered. *Econometrica*. Vol. 47. No. 3. pp. 747–759.

THEIL, H. [1967]: *Economics and Information Theory*. North-Holland Publishing Company. Amsterdam.

THON, D. [1979]: On Measuring Poverty. *Review of Income and Wealth*. Vol. 25. No. 4. pp 429–440.

THORBECKE, E. [2008]: Multidimensional Poverty: Conceptual and Measurement Issues. In: *Kakwani, N. – Silber, J.* (eds.): *The Many Dimensions of Poverty.* Palgrave Macmillan. New York.

TÓTH ISTVÁN GY. [2003]: Jövedelemegyenlőtlenségek – tényleg növekszenek, vagy csak úgy látjuk? *Közgazdasági Szemle.* 50. évf. 3. sz. 209–234. old.

TSUI, K. Y. [1995]: Multidimensional Generalizations of the Relative and Absolute Inequality Indices: The Atkinson-Kolm-Sen Approach. *Journal of Economic Theory*. Vol. 67. No. 1. pp. 251–265.

TSUI, K. Y. [1999]: Multidimensional Inequality and Multidimensional Generalized Entropy Measures: An Axiomatic Derivation. *Social Choice and Welfare*. Vol. 16. No. 1. pp. 145–157.

TSUI, K. Y. [2002]: Multidimensional Poverty Indices. *Social Choice and Welfare*. Vol. 19. No. 1. pp. 69–93.

VEGA, C. L. – URRUTIA, A. – DIEZ, H. [2009]: *The Bourguignon and Chakravarty Multidimensional Poverty Family: A Characterization.* ECINEQ WP 2009–109. ECINEQ Society for the Study of Ecomonic Inequality.

VEGA, C. L. – URRUTIA, A. – VOLIJ, O. [2011]: *An Axiomatic Characterization of the Theil Inequality Ordering.* Monaster Center for Economic Research. Ben-Gurion University of the Negev. Beer Sheva.

WATTS, H. W. [1968]: An Economic Definition of Poverty. In: *Moynihan, D. P.* (ed.): *On Understanding Poverty*. Basic Books. New York.

ZHENG, B. [1997]: Aggregate Poverty Measures. *Journal of Economic Surveys.* Vol. 11. No. 2. pp. 123–162.

ZHENG, B. [2005]: *Unit-Consistent Decomposable Inequality Measures*. Working Paper No. 05-02. University of Colorado. Denver.

Summary

The paper introduces a new multivariate methodology for measuring multidimensional inequality. The method proposed is based on the information theory generalized entropy indices and gives a composite inequality measure of a multivariate oblique space. The key formula is the so-called generalized variance metric applied to the special Theil covariance matrix yielding a between-within effects decomposable index of the total inequality. Even in the case of only one dimension, the new approach is multi (two) variate based. In addition, given a (socio-economic) segmentation of the population, the contribution of an individual group to the ”within-groups” inequality can also be quantified and ranked. Finally, the new inequality approach applied to a censored distribution yields a multivariate-multidimensional poverty measurement. Dimension-specific poverty lines or aggregate attribute poverty lines are also allowed.