

Válasz

Sipos Béla opponensi bírálatára

„Módszertani hozzájárulás a szegénység többváltozós statisztikai méréséhez”

című MTA Doktori értekezésemre

Megköszönöm Sipos Bélának, az MTA doktorának, hogy értekezésemre értékelte. Köszönöm észrevételeit, támogató véleményét, és külön köszönet illeti a dolgozat minden részletére kiterő, gondos olvasást.

A megjegyzésekre, kérdésekre reagálva, válaszaim a következők.

1. fejezethez tett észrevételek, kérdések megválaszolása

Az egydimenziós klasszikus szegénységi indexek tekintetében - bemutatásuk érdekében - az értekezés áttekintése „teljeskörűnek” mondható, mivel megad egy olyan általános index-szerkesztési keretet, melyben adott faktort, ható tényezőt kicserélve, illetve módosítva, egy új indexhez jutunk, tartalmazva értelemszerűen az eddig, jelenleg elérhetők mindegyikét is.

2. fejezethez tett észrevételek, kérdések megválaszolása

Az értekezés egy új, individuális meglátása, elve, hogy környezetünkben csak „fölfelé” - a jobb módúak felé - tekintve és figyelmen kívül hagyva az általunk már meghaladt jóléti szinteket, a társadalmi szintű relatív depriváltság foka akár csökkenhet is. Egyetértek az Opponens véleményével, miszerint érdemes lenne figyelembe venni a „fölről lefelé” értelmezett elégedettségi érzeteket is, ellensúlyozandó a „kisebb mint” irányú deprivációs relációkat.

A regresszív transzfer deprivációs hatásával kapcsolatban feltett kérdésre, hogy vannak-e olyan transzfer szituációk, mikor egy deprivációs mutatótól biztosan a növekedés, vagy biztosan a csökkenés jelzését várjuk, a válasz igen, pl. ha a legszegényebb a jövedelmi tétel átadását elszenvető donor, bárki is a címzett személy. Más transzfer szituációk is adhatók a transzfer mértéke, a donor és a címzett személyek jövedelmei és rangpozíciói, továbbá az összes többi szereplők jövedelmei és rangpozíciói, vagyis a referencia csoportok struktúrái függvényében.

Az értekezésben javasolt egydimenziós P szegénységi index létrehozásának elvi alapját azon tulajdonsága adja, miszerint a relatív deprivációs szint küszöb alatti **csökkenését** szegénységet **csökkentő** faktorként kezeli, melynek hatására a P index *kevésbé emelkedik, vagy akár*

csökkenhet is. Ezen a ponton az Opponensnek igaza van, hogy a reprezentatív szegény, és a reprezentatív deprivált személy fogalmának bővebb kifejtése az újonnan bevezetett *P* szegénységi index működésének megértését segítené, bár ez a későbbiekben számpéldán keresztül megtörténik.

A „*pszeudo egyenlőség-egyenlőtlenség*” fogalmának bevezetésével kapcsolatban köszönöm az Opponens egyetértését azon kérdésben, hogy a bevezetett fogalom bár *szokatlan*, de *kézenfekvő* irány az *egyenlőtlenség* fokának a mérése során.

A deprivációs függvények tekintetében a tartalmi különbség a Q-deprivációs hányad elv, és a Delta-kommunalitás elv között alapvetően az, hogy míg Q az *elszenvedett* átlagos depriváció mértéke, addig Delta az *eliminált* depriváció átlagos mértéke.

A reprezentatív szegény jövedelmének az értéke a depriváció-érzékeny P-indexben a (2.12) formula szerint az átlagos Q-depriváció komplementere, míg a (2.13) formula szerint a komplementer deprivációk átlaga, melyek az alkalmazott depriváció-averzió paraméter értéke függvényében különböznek egymástól. A deprivációs hányad Q-mértékek *zéró*-, *pozitív*- és *globális*-típusú változatai az átlagolás súlyozásának a megválasztásában különböznek, tehát az alkalmazási előnyeik és hátrányaik is a súlyozás mikéntjén alapulnak. Igény szerint azon változat alkalmazása indokolt, mely inkább, vagy kevésbé érzékeny egy regresszív transzfer megvalósulására.

A relatív deprivációs hatványparaméter becslését és elemzését - egyetértve az Opponenssel – természetesen érdemes lenne elvégezni más nevezetes jövedelmi eloszlásokra is, mint például a log-normális, a Pareto, Champernown, stb.

A gamma-eloszlás hatványkitevő harmadik paramétere statisztikai tartalmát illetően a relatív deprivációval szembeni averzió egyféle mértéke. E hatványparaméter becslési problémája inkább kétparaméteres - miközben a harmadik paramétert iteratív keressük - de alkalmazható a három paraméter szimultán módon való becslési módszere is. Léván ismert típusú sűrűségfüggvény paramétereit becsüljük, ezért a javasolt becslési módszer a Maximum Likelihood.

A Lorenz-görbe aszimmetriáját illetően az Opponensnek igaza van, miszerint a (2.55) formula speciális esete a (2.56) általánosabb formulának, ezért a publikációból elhagyható lett volna. Az értekezés valóban nem adja meg továbbá, hogy a Lorenz-görbe aszimmetriájának foka - ha már rendelkezésre áll az értéke - milyen módszertan mentén építhető be a *P* szegénységi indexbe. Ez további kutatás témája. A szimmetria átló fölötti átlagpont növeli a szegények szegénységét, mert ekkor túlnyomó létszamarányban szerepelnek az átlag alatti szegények,

miközben az átlag fölötti jövedelmek az összes jövedelemnek csak csekély hányadát adják. Ezt az esetet pozitív aszimmetria jelzi. A (2.55) vagy a (2.59) mutatók bármelyike beépíthető a szegénységi mértékbe, mintegy egyenlőtlenség - inflátor faktor.

3. fejezethez tett észrevételek, kérdések megválaszolása

A GVI mint entrópia alapú módszer - egyetértve az Opponens észrevételével –minden olyan területen alkalmazható, ahol a rendezettség és a rendezetlenség mérése a kérdés, mint például a neurális hálózatok „cross-entropy” hibaszámítási módszere kategória változók előrejelzése során.

A GVI módszer szempontjából a *többsváltozós* és a *többsdimenziós* megközelítések között alapvető különbség, hogy a mindenkori p -dimenziós megközelítés $2p$ -változós alkalmazást indukál.

A Theil-kovariancia (3.13) közgazdasági-statisztikai tartalma: a relatív jövedelem és a log-hozam (vagyis a relatív jövedelem hasznossága) közötti kovariancia, mely a $GE(1)$ és a $GE(0)$ paraméterű általánosított entrópiák összege, ahol $GE(0)$ az értekezés szóhasználatában az $L=$ Loss mérték. Itt, ahogy $G(0)$ abszolút mértékben növeli a kovarianciát, a CIF szorzó relatív értelemben inflálja azt. A (3.20) Theil-kovariancia mátrix elemei megfelelően 5 nevezetes egyenlőtlenségi mutatók valamelyikeinek a függvényei. Ezek rendre: a $GE(0)$, $GE(1)$, $GE(2)$, paraméterű általánosított entrópiák, továbbá a jövedelem relatív szórása, és a logaritmusos jövedelmek varianciája. Hozzáteszem, hogy $GE(0)$ a nevezetes Theil-féle Mean-Logarithmic Deviation index, $GE(1)$ a nevezetes Theil-féle redundancia index és $GE(2)$ a nevezetes Hirschman-Herfindahl index.

Aláhúzendó, hogy a *Theil-kovariancia és a Theil-mátrix fogalmi bevezetése az értekezés önálló eredménye, a Theil szóhasználat az alkalmazott terminológiában csupán a mátrix elemeinek a tartalmára utal.*

A Wilks' Lambda dekompozícióját illető kérdésre a válasz igen, az értekezés megoldotta az éppen alkalmazott kategóriák (most Budapest, Nagyváros, Többsváros, Községek) százalékos hozzájárulásának a számszerűsítését a belső egyenlőtlenséghez. Itt a (négyzetes) kanonikus korreláció (3.41) szerinti számításának numerikus oka van a javasolt GVI módszertanban.

Összefoglalóan a GVI egyenlőtlenségi módszertannak a meglévő mutatókkal szembeni előnye, hogy figyelembe veszi a dimenziók korrelációs rendszerét és aszimmetrikus eloszlását is, egydimenziós esetben is többsváltozós megközelítést alkalmaz, szolgáltatja a többsváltozós külső-belső százalékos felbontási arányt, és a belső többsváltozós egyenlőtlenségen belül megadja a kategóriák százalékos struktúráját.

A cenzorált eloszlás Wilks' Lambda hányadosa egydimenziós-kétváltozós jövedelmi esetben - adott szegénységi küszöb mellett - „általánosított szegénységi arányként” működik (52-53. old.), mert a Lambdát, mint súlyozott átlagos csoporton belüli hatást itt kifejezetten a szegények létszamaránya, és körükben a küszöb alatti Theil-variancia mozgatja. Ugyanakkor a szegények küszöbrel szemben érzett globális relatív deprivációjának foka - a Wilks' Lambda komplementereként - mint általánosított relatív jövedelmi rés a küszöbrel szemben értelmezhető.

Többdimenziós esetben az új GVIP módszer az egyes dimenziók eltérő szegénységi küszöb-szintjei mellett külön-külön cenzorál, majd képezve a cenzorált relatív jövedelmeket és a cenzorált log-hozamokat, e rendszerben a GVI egyenlőtlenség alkalmazásával a GVIP szegénység automatikusan adódik.

Többdimenziós esetben - másik technikával - az új GVIP módszer az egyes dimenziók eltérő szegénységi küszöb-szintjei mellett számítja a dimenzió-specifikus általánosított szegénységi arányokat, melyeket leginkább mértanilag átlagol, elvezetve így az implicit relatív jövedelmi réshez.

Fontos kérdés a többdimenziós megközelítésben, hogy milyen más többdimenziós mértékek állnak rendelkezésre jelenleg az irodalomban, és az értekezésben javasolt GVI(P) elv milyen tulajdonságaiban haladja meg azokat, és ad új ismeretet? Nos, a meglévő algoritmusok az egyedi, egyváltozós dimenziókat lineárisan kombinálják, és a kombinációt mérik valamely egydimenziós metrikával (egyenlőtlenségi, szegénységi). A dezaggregáció kérdésében vagy csak a belső-külső, vagy csak a főátlaghoz való additív hozzájárulást kalkulálják.

4. fejezethez tett észrevételek, kérdések megválaszolása

A szegénységi (gazdagsági) küszöbök – lényegében a relatív deprivációs küszöbök - kvantilis regresszió alkalmazásával történő *rétegspecifikus* meghatározását kézenfekvően a nevezetes modell-specifikációs eszközök, mint a változók transzformálása, a prediktorok megfelelő interakcióinak és kvadratikus hatásainak szelektálása, a heteroszkedaszticitással korrigált GLS standard hibák számítása minden bizonnyal tovább finomítja, csökkentve ezzel a modell-specifikációs torzítás mértékét.

5. fejezethez tett észrevételek, kérdések megválaszolása

A szegénységi küszöb alá kerülés kockázatát az egzakt logisztikus regresszióval magyarázó modellt érintően egyetértek az Opponens azon észrevételével, miszerint az ún. *elégséges* statisztika árnyaltabb tárgyalása, bemutatása több területet érdemelt volna az értekezésben, lévén kulcs-fogalom az eredmények értelmezéséhez. Mindazonáltal az elégséges statisztika

fogalma a statisztikai irodalomban közismert, bár a hazai irodalomban történő tárgyalása szűknek mondható. A fejezet a *kisminta*, a *ritkaság* és a *szeparáltság* kezelését a korrekt, értsd *egzakt p-value* értékek a meghatározására és alkalmazására fűzi fel. A *feltételes egzakt likelihood becslés és az egzakt medián következtetés* lényegi különbségének bemutatása (60.old.) *megtörténik* a fejezetben, de a matematikai részletek ennél mélyebb tárgyalása jelentősen több terjedelmet igényelne, az áttekinthetőség rovására menne.

A zavaró, azaz „nuisance” paraméter abban az értelemben zavaró, hogy értékét akkor is becsülni kell az összes paraméterek között, ha az értéke nem érdekes a következtetések szempontjából, viszont jelenléte, azaz becslési kényszere az érdeklődés homlokterében álló paraméter becslését inkonzisztenssé teheti, teszi.

Az ún. „Scores” teszt előnye a „Likelihood-Ratio” teszt ellenében, hogy végrehajtása nem igényli a kérdéses paraméter becsült értékének az ismeretét, így akkor is végrehajtható, mikor a ML becslés nem létezik. Ezért az elutasítási „Rejection” tartomány kijelölésére a *score* elv preferált a LR elvvel szemben, míg a harmadik, a *probability* elv alapján megtörténhet, hogy az egzakt permutációs mintavételi eloszlás többmódusú lesz, ami nem illeszkedő, szabdalt elutasítási tartományokat eredményezhet. A *korrigált, ún. mid-p-value* alkalmazása a diszkrét mintavételi eloszlás folytonossági korrekciója, a szignifikancia szint korrekciója nélkül.

Ha az elégséges t-statisztika megfigyelt értéke egybeesik a terjedelmének valamelyik extrém szélső (min, max) értékével, akkor a Feltételes Likelihood nem maximálható a Béta paraméter tekintetében, mert a megfigyelt t-érték *nem belső pontja a skálának*. Ilyenkor a t-skála minden lehetséges kimenetének a Likelihoodja egyirányban változik Béta növelésével, a ML becslés tehát ez esetben nem létezik.

6. fejezethez tett észrevételek, kérdések megválaszolása

A szegénységmérés SEM modelljeit érintő bírálattal kapcsolatban külön köszönöm az Opponensnek, hogy észrevételei között kiemeli, hangsúlyosnak tartja, hogy a SEM modell alkalmazása lehetővé teszi a szegénység *olyan úton* történő többdimenziós mérését, ami nem igényli a szegénységi indikátorok szegénységi küszöbeinek a rögzítését.

A (6.4) egyenletnél valóban fölcserélődött a **v** vektorban a manifeszt **m** és a latens **l** változók sorrendje. Igen precíz meglátás.

A 6.1 ábrán publikált induló koncepcióhoz, hipotézishez képest, a 6.3 ábrán kiterjesztett hipotézis a korábbi manifeszt indikátorok körében újabb latens köröket fogalmaz meg. A modell fokozatos bővítésének megfelelő eszköze a ML elven alapuló Lagrange-multiplikátor teszt alkalmazása, mert nem igényli a bővebb modell ML becslését. A paraméterbecslési és a

modell-identifikáció során az *exogén manifest* változók közötti kovarianciák számossága *növeli* a modellezhető egyenletek számát, így a szabadsági fokot, míg az *endogén latens* változók száma *csökkenti* a szabadsági fokot.

A SEM modell környezetben az ún. szaturált modell maga a manifest változók kovariancia mátrixa, míg az ún. null-modell csak az egyediségeket, tehát a hibafaktorokat tartalmazza.

Bár a 6.5.2 fejezet felsorolja a Trait és a Method faktorok indikátorainak egyféle lehetséges listáját, itt nem volt cél rögzíteni, hogy végül mely konkrét indikátorokba mutatnak a Trait és a Method faktorokból irányított nyilak a 6.4 és a 6.5 ábrákon, mert a mondanivaló az elvi keret, mely bármely alkalmas indikátorral feltölthető, felparamétrezhető! A becslési eljárást illetően, mivel aszimmetrikus háztartásstatisztikai változók kerültek alkalmazásra, a 6.5.2 alfejezet az ADF (Asymptotically Distribution Free) módszert alkalmazta.

A Within-Trait - Cross-Method korrelációk – módszerről-módszerre haladva a 6.7 táblában - rendre azért növekszenek, mert a háztartás létszámát használva vetítési alapként, az értekezés egyfelől a tagok számát, majd az ennél alacsonyabb fogyasztási egységet alkalmazza, és a zsugorított skálák között szorosabb a korreláció.

A 6.10 táblázatában a megszorított, tehát szabadon nem becsült CFA paraméterek standard hibái értelemszerűen nincsenek értelmezve, ezért hiányzó, missing p-értékeket produkálnak! Az EMF_CU modell általánosságban is takarékosabban bánik a becsülendő paraméterekkel, mint az IMF_UU modell. A 6.6 és 6.7 ábrák modelljeiben az egyes indikátorok definíciójában az 1.decilis változó értéke 1, ha az illető háztartás az alsó decilisben van, egyébként zéró, Q az illető háztartás deprivációs hányada és a [10 mínusz aktuális decilissorszám] változó a felső decilistől vett elmaradást méri.

A 109.oldalon a GoF (Goodness of Fit) heurisztikus mutatók értékéből az ADF és az IWLS becslések által adott homogenitás/heterogenitásra vonatkozó konklúzió – megköszönve az Opponens észrevételét - természetesen fordított megfogalmazásban helyes, mint ahogy az értekezés Téziseiben már így, helyesen szerepel. A Goodness of Fit heurisztikus mutatók heterogén és homogén modellek illeszkedését összevető, növekmény jellegű *Bentler*-indexek értékei az IWLS eredmények szerint inkább a heterogén, míg az ADF eredmények inkább a homogén variancia elfogadása irányába mutatnak.

Véleményem szerint az Orthosim rotációs eljárásnak – pl. a klasszikus Varimax eljárással szemben – abban rejlik az előnye, hogy a faktorsúly mátrix négyzetes elemeinek az általánosított varianciáját maximálja. Ha az egyedi (hiba) faktorok kovariancia mátrixának

struktúrája éppen $\sigma^2 \mathbf{I}$, akkor az egyedi faktorok homoszkedasztikusak, és jogosulttá válik a Kaiser-féle EPIC faktor-extrahálási módszer alkalmazása.

7. fejezethez tett észrevételek, kérdések megválaszolása

Az ún. Burt-mátrix, vagy az Indikátor Mátrix egyszerű Correspondencia Analízise ugyanazon sajátértékekre (értsd négyzetes szinguláris értékekre), tehát ugyanazon *főinerciákra* vezet.

A (7.16) átviteli formula alapján a nagyobb távolságok a 7.1, 7.2, 7.3 ábrák pontjai között erősebb, az alacsonyabb távolságok pedig gyengébb asszociációt jeleznek az illető pontok és a NEMOK kategória között!

Budapest, 2013. április 07.

Hajdu Ottó