

KOVÁCS ERZSÉBET–RÉTALLÉR ORSOLYA–VÉKÁS PÉTER

Modellpontok szerepe a nyugdíj-hatásvizsgálatban

A mikroszimulációs nyugdíjmodell elkészítésekor az adatok körének meghatározása is lényeges kérdés. Rendelkezésünkre állt az Országos Nyugdíjbiztosítási Főigazgatóság teljes adminisztratív adatállománya. A nyugdíjmodellezés, a fenntarthatósági számítások elvégzése hosszú távra történik, és a teljes népességet érinti. A jogosultságszerzők és az ellátottak nagy száma, valamint a hosszú időtáv miatti jelentős számításigényt csökkenheti, ha a mikroszimuláció egyedi adatok helyett embercsoportokra, úgynevezett modellpontokra készül. A cikk a módszertani megfontolások ismertetése mellett bemutatja a modellpontok előállításának lépéseit, a népszámlálási adatokkal történő egybevetés nehézségeit, végül magukat az előállított modellpontokat.

Journal of Economic Literature (JEL) kód: C53, H55.

Mikroszimuláció a nyugdíjmodellezésben

A nagyszámú egyénre vonatkozó, részletes információkat felhasználó nyugdíjmodelleket *Gál és szerzőtársai* [2009] két kategóriába sorolja: a standard modellek, valamint a mikroszimulációs modellek bővebb családjába, s mindkettőn belül további alkategóriákat különböztet meg. Az Országos Nyugdíjbiztosítási Főigazgatóság (ONYF) MIDAS_HU nyugdíjmodellje a Belgiumban használatos MIDAS_BE modell (lásd *Dekkers* [2007]) magyarországi adaptációja, és ahhoz, valamint a korábban Magyarországon használatos Nyugdíj és Időskor Kerekasztal (NYIKA) modelljéhez (*Holtzer* [2010]) hasonlóan a dinamikus mikroszimulációs modellek (*Zaidi–Rake* [2001]) családjába tartozik. A Deloitte tanácsadó és könyvvizsgáló cég a Csehországban alkalmazott dinamikus mikroszimulációs modellel

Kovács Erzsébet, Budapesti Corvinus Egyetem Operációkutatás és Aktuáriustudományok Tanszék (e-mail: erzsebet.kovacs@uni-corvinus.hu).

Rétallér Orsolya, Rijksuniversiteit Groningen, Department of Economics, Econometrics and Finance (e-mail: o.retaller@rug.nl).

Vékás Péter, Budapesti Corvinus Egyetem Operációkutatás és Aktuáriustudományok Tanszék és MTA–BCE Lendület Stratégiai Interakciók Kutatócsoport (e-mail: peter.vekas@uni-corvinus.hu).

A kézirat első változata 2015. július 22-én érkezett szerkesztőségünkbe.

DOI: <http://dx.doi.org/10.18414/KSZ.2015.12.1328>

kapcsolatos tanulmányában (*Deloitte* [2011]) a mikroszimulációs nyugdíjmodellek számos előnyét és hátrányát sorolja fel. Az előnyök közül a következőket tartjuk fontosnak kiemelni:

- az egyének teljes életét modellezik, és figyelembe veszik az egyének életében bekövetkező eseményeket;
- elvben minden elérhető információ és egyéni adat felhasználható bennük;
- a jogszabályi környezet által diktált paramétereket képesek megjeleníteni;
- átfogó aggregált és egyéni szintű eredményeket produkálnak, eloszlás jellegű eredményeket szolgáltatnak például az időskori szegénység vizsgálatához;
- lehetővé teszik a nyugdíjrendszer aktuáriusi szempontból történő értékelését, azaz a befizetések és kifizetések ekvivalenciájának vizsgálatát;
- egyéb szociális ellátórendszerekre is kiterjeszthetők, így a szociálpolitika támogatására egységes eszközként szolgálhatnak.

A mikroszimulációs modellek alkalmazásának hátrányai a nyugdíjmodellezésben is tetten érhetők:

- magas implementációs (szoftver-, szakértői, személyi) költségekkel járnak;
- magas a számításigényük (mind szoftver-, mind hardverszempontból), és hosszú a számítási idejük;
- magasak a bemenő adatokkal szembeni követelményeik, és bonyolultak az alkalmazás előfeltevései;
- nehezebb biztosítani az egyéb (makroökonomiai, demográfiai stb.) feltevésekkel való konzisztenciájukat.

Az előnyök és hátrányok felsorolása alapján általánosságban megállapítható, hogy a mikroszimulációs megközelítéssel járó precizitás előnyeivel a megközelítés bonyolultságából adódó problémák állíthatók szembe.

A mikroszimulációs nyugdíjmodell elkészítése melletti szakmai döntés meghozatalakor az adatok körének meghatározását is mérlegelni kellett. A teljes ONYF adminisztratív adatállomány a rendelkezésünkre állt, és négy lehetőség közül választhattunk:

1. a teljes ONYF adminisztratív adatállomány felhasználása (ez ellen szólt azonban a hatalmas számításigény);
2. a jogosultságszerzőket reprezentáló véletlen minta (ez ellen szólt az eljárás költsége, valamint az új egyének besorolásának nehézsége – a későbbiekben ezt még részletezzük);
3. az adminisztratív adatállományból kiválasztott néhány kohorsz, amelyek tagjai már közel teljes jogosultságszerzési időszakkal rendelkeznek (az 1955–1959 között születettek elemzését tanulmányunk ismerteti);
4. az adatállományból kiindulva különböző számosságú embertípusokat képzünk, akik nemük, életkoruk és további jellemzőik alapján reprezentálják a jogosultságszerzőket, valamint a nyugdíjban részesülőket.

Az Országos Nyugdíjbiztosítási Főigazgatóság szakértőivel való egyeztetést követően a 4. utat választottuk, és a MIDAS_HU modell munkaerő-piaci moduljának egyéni

szereplői helyett szakértői szempontok alapján létrehozott homogén csoportokat, úgynevezett modellpontokat készítettünk. E homogén modellpontok kialakítása következtében ideális esetben az egyszerűsítés révén az egyénre vonatkozó információknak csak jelentéktelen része megy veszendőbe, így megőrzi a mikroszimulációs megközelítés előnyeit. Ezzel egyidejűleg – a modellpont-adatbázis teljes adatbázishoz képesti jóval kisebb mérete következtében – e módszer segítségével minimalizálhatók a mikroszimulációs modellek költségbeli, számítási és időbeli hátrányai.

A modellpontok előállítása

A nyugdíjmodellezés hosszú távú feladat, valamint a jogosultságszerzők és az ellátottak nagy száma miatt jelentős számítást igényel. Ez utóbbi szempontok figyelembevételével végül mintegy százezer körüli modellpont képzésére került sor, amelyeknek egyszerre a következő két célkitűzésnek kellett megfelelniük.

1. A lakosságot reprezentáló, a jogszerző időszak egészét jellemző demográfiai és jogosultsági adatok felhasználásával olyan homogén, a nyugdíjszámítás szempontjából releváns csoportok képzésére kellett törekedni, amelyek alapjául szolgálhatnak különböző nyugdíjszámítási variációk társadalmi csoportokra vonatkozó elemzésének, Augusztinovics Mária megfogalmazásában hatásvizsgálatának¹. Ehhez szükség volt a modellpontok képzéséhez felhasználandó ismérvek megfelelő kiválasztására is.

2. Lehetőleg a teljes lakosságot reprezentáló adatállomány kialakítására kellett törekedni, a teljesség érdekében statisztikai módszerekkel kiegészítve a rendelkezésre álló ONYF-adatállományt, hogy az kiterjedjen a teljes népességre, melyről rendelkezésünkre álltak a 2011. évi népszámlálás részletes adatai. Az összevetések elvégzése után gyakorisági korrekciókra volt szükség az egyezés biztosítása érdekében.

Módszertani szempontból nem volt egyértelmű feladat a homogén modellpontok kialakítására alkalmazandó eljárás kiválasztása. Ennek során a következő három eljáráscsaládot vettük alaposabban vizsgálat alá.

1. *Mintavételi módszerek.* A hagyományos – véletlen, egyszerű vagy rétegzett – statisztikai mintavételi módszerek (lásd Fuller [2009]) jól használhatók például a lakosság jövedelmének, fogyasztásának, életmódjának, utazási szokásainak stb. vizsgálatára. Ilyen megközelítést akkor érdemes követni, ha lehetőség van a nyugdíj-hatáselemzés szempontjait követő rétegzett mintavételre. A mintába kerülő személyek száma általában lényegesen kisebb, mint a modellpontok előírányzott száma, mert a kérdőíves megkeresés és adatgyűjtés időigényes és drága, így ez a megközelítés a teljes magyar népességet követő modellpontrendszer kialakítására nem alkalmas.

2. *Klaszterező eljárások.* A struktúrafeltáró módszerek közül a legalkalmasabb osztályozó eljárás a klaszterelemzés (hierarchikus, nem hierarchikus vagy

¹ Augusztinovics Mária nemcsak a Nyugdíj és Időskor Kerekasztal aktív résztvevője, a Holtzer [2010] kötete egyik szerzője volt, de számos kérdésben ötleteivel, javaslataival segítette, ösztönözte a többéves közös munkát. Ezúton is köszönettel emlékezünk rá és munkásságára.

kétlépcsős algoritmusokkal – lásd például *Everitt és szerzőtársai* [2010], *Füstös és szerzőtársai* [2007], *Kovács és szerzőtársai* [2011]). Ez a módszertan számítási igénye miatt többmillió sokaságra és ilyen nagyszámú klaszter keresésére az időbeli korlátok figyelembevételével nem használható. Mindkét eljárásváltozatnál gondot okozhat az is, hogy a rendelkezésre álló változók mérési skálái nagyon különbözők, vannak kategória-, és vannak értékváltozók is, amelyek együttes kezelése statisztikailag nehezen valósítható meg. Az egyes klaszterek mint modellpontok jellemzőihez nem adhatók meg kategóriahatárok, és nem biztosítható a megfigyelt jellemzőkkel le nem írt, további – jelenleg még esetleg meg sem született – személyek klasztereinek kialakítása.

3. *Többdimenziós kategóriakombinációk képzése.* Ezzel a szemlélettel változónként kategóriákat képzünk, és megfelelően alátámasztható kategóriaösszevonások után több változó kategóriáinak lehetséges kombinációiból alkotunk csoportokat, majd végül újabb kategória-összevonásokat vagy -szétválasztásokat hajtunk végre oly módon, hogy a modellpontok számossága megfelelő legyen. Külön figyelmet kell fordítani a többdimenziós kontingenciatábla egyes celláiban található hiányzó adatok feltöltésére. E megközelítést használta a Magyarországon korábban alkalmazott NYIKA-modell (*Holtzer* [2010]) is.

Mivel a jövőbeli állományalakulás modellezéshez új modellpontokat is létre kell hozni (például a később munkába álló jelenlegi iskolások számára), lényeges szempont, hogy ezeket könnyen be lehessen sorolni az ONYF adatállományából kialakított modellpontok rendszerébe. Az 1. és a 2. módszer családba tartozó eljárások esetében ez hasonlósági/távolsági mértékek vagy csoporthoz tartozást előre jelző, valószínűségi becslést is adó többváltozós statisztikai módszerek használatával érhető el. Hasonlósági mértékek segítségével azonban az egyenletes súlyozáson kívül más ésszerűen alátámasztható súlyrendszert nem lehet a változókhoz rendelni, a többváltozós klaszszifikációs eljárások (például logisztikus regresszió, diszkriminanciaelemzés stb. – lásd *Füstös és szerzőtársai* [2007], *Kleinbaum–Klein*, [2010], *Kovács és szerzőtársai* [2011]) használata pedig a nagy csoportszám miatt nem volt lehetséges. E szakmai indokok mérlegelésével végül a MIDAS_HU modell munkaerő-piaci moduljában a 3. módszer család alkalmazása mellett döntöttünk.

A teljes népesség és a modellpontok összevetése

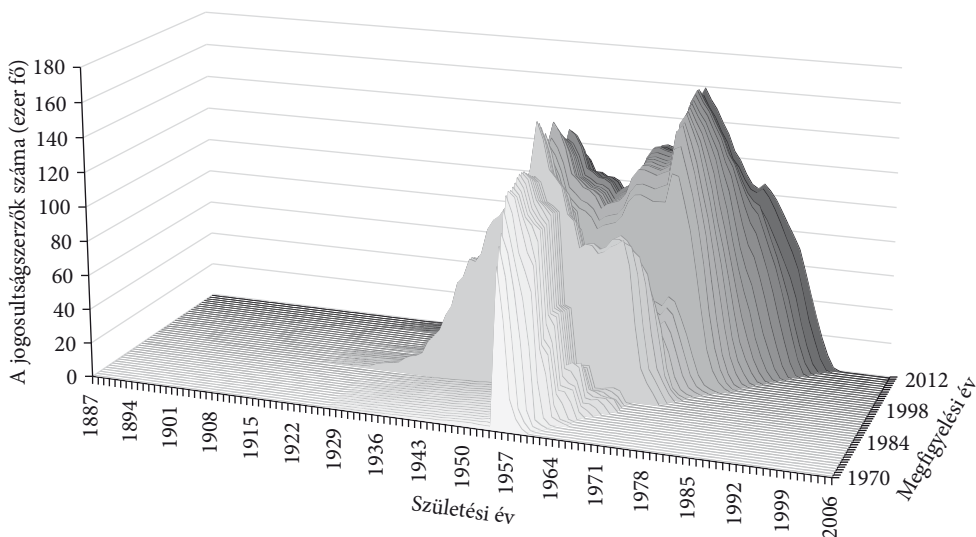
Minden statisztikai elemzésnek előfeltétele a megfelelő adatbázis. Ebben az esetben azonban egy meglehetősen speciális, némileg hiányosnak mondható adatbázissal rendelkezünk, amit a legkevésbé sem tekinthetünk véletlen mintának. Az ONYF-nél ugyanis eleve csak a friss jogosultságszerzési adatok kerültek számítógépes nyilvántartásba, a korábbi jogosultságszerzési adatok digitalizálása pedig még napjainkban is zajlik. Ez azt jelentette, hogy az idő előrehaladtával egyre teljesebb körű adatok álltak rendelkezésünkre. Vízválasztó évnek tekinthetjük ilyen szempontból az 1988-as évet, mert bár ekkor sem volt még százszázalékos

a feldolgozottság, mégis látványosan több adat áll már rendelkezésünkre, mint egy évvel korábbról.

Az adatok digitalizálása során kiemelt jelentőségűek az 1955–1959-es kohorszok. Ezek az évjáratok felülreprezentálnak mondhatók, de egyúttal értékes információt szolgáltatnak számunkra az adatok idősoros vizsgálatához. Az előbbieket illusztrálja az a háromdimenziós hisztogram, amely a születési évek és a megfigyelési évek bontásában ábrázolja a jogosultságot szerzők számát (1. ábra).

1. ábra

A jogosultságot szerzők száma (ezer fő)



Az adatok forrása: ONYF.

A teljes életpályáról, illetve a foglalkoztatottságról úgy készíthetünk becsléseket, hogy az említett kohorszok adatait alapul véve mind a naptári évek, mind az egyes életkorok sajátosságait figyelembe vesszük. Ehhez elsőként teljesebb körűvé kellett tennünk magát az adatállományt.

Természetesen a nemzetközi szakirodalomban is találhatunk példát olyan esetekre, amikor egy népességi adatokat tartalmazó adatbázis hiányosságait statisztikai módszerekkel igyekeznek kipótolni. Ilyen esetek például a fejlődő országokban a nem regisztrált születések és halálozások miatti utólagos népességszám-korrekciók, de gondolhatunk Kína speciális esetére is, ahol az „egygyermek-politikának” köszönhetően sokszor szándékosan titkolják a megszületett gyerekeket a hatóságok elől, így nem is regisztrálják őket. A téma fontosságát jelzi, hogy az ENSZ több kiadványt készített a demográfiai becslésekről,² de külön kézikönyv is készült a demográfiai mutatók hiányos adatbázisokból történő becslésére (*United Nations* [1967]).

² Az ENSZ kiadványai a <http://www.un.org/esa/population/techcoop/DemEst/DemEst.html> linkről letölthetők.

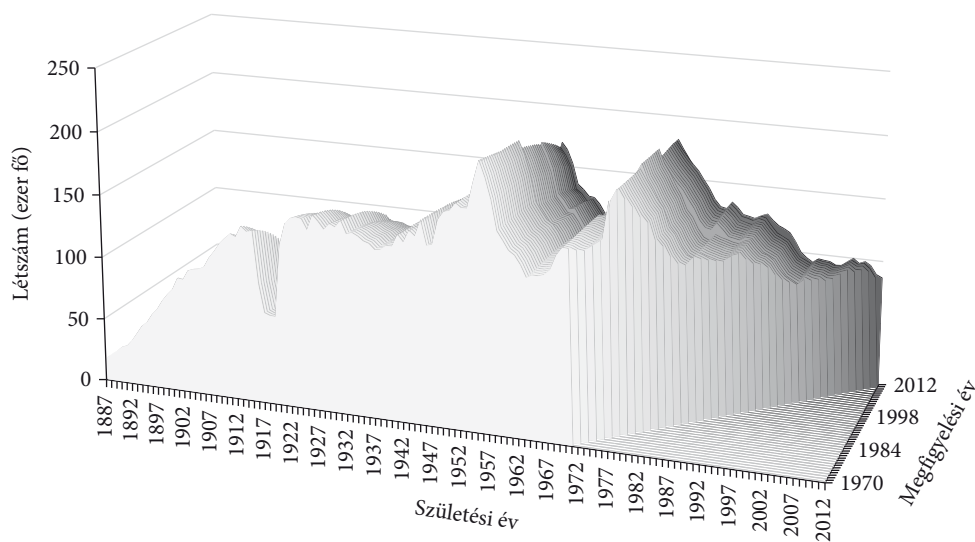
A foglalkoztatotti adatbázis vizsgálatához elsőként a teljes magyar lakosság vizsgálatát végeztük el. A teljes populáció behatóbb vizsgálata lehetővé tette, hogy az egyes kohorszok és megfigyelési évek jellemzőit jobban megismerjük, majd azok specifikumait az ONYF foglalkoztatotti adatbázisára is rávetítsük. A demográfiai vizsgálat után pedig újból a foglalkoztatási statisztikákra koncentráltunk: ennek érdekében adatbázisunkat a KSH foglalkoztatottsági adataival hasonlítottuk össze.

Demográfiai elemzés

A demográfiai adatokat a már korábban említett két szempontból vizsgáltuk meg: egyrészt 1970 és 2012 között minden megfigyelési évre elkészítettük a kohorszok szerinti megoszlást (mely keresztmetszeti elemzést tett lehetővé), másrészt a kohorszok létszámát ábrázoltuk az idő függvényében (ez pedig idősoros elemzésre adott lehetőséget). Az adatok nemek szerinti bontásban is rendelkezésünkre álltak, de terjedelmi korlátok miatt itt csak a teljes népességre vonatkozó 2. ábrát közöljük.

2. ábra

A magyar népesség születési évek szerinti megoszlása az 1970 és 2012 közötti évekre



Az adatok forrása: KSH.

A keresztmetszeti vizsgálat gyakorisági ábráján jól megfigyelhetők a magyar történelmi sajátosságok. Látható például, hogy az első világháború alatt született kohorsz létszáma lényegesen alacsonyabb, továbbá kiemelkednek a Ratkó-gyerekek³ és -unokák is, hiszen az 1954., majd az 1974–1975. években születettek létszáma jóval magasabb, mint más kohorszoké. Emellett megjegyezzük, hogy bár első pillantásra két

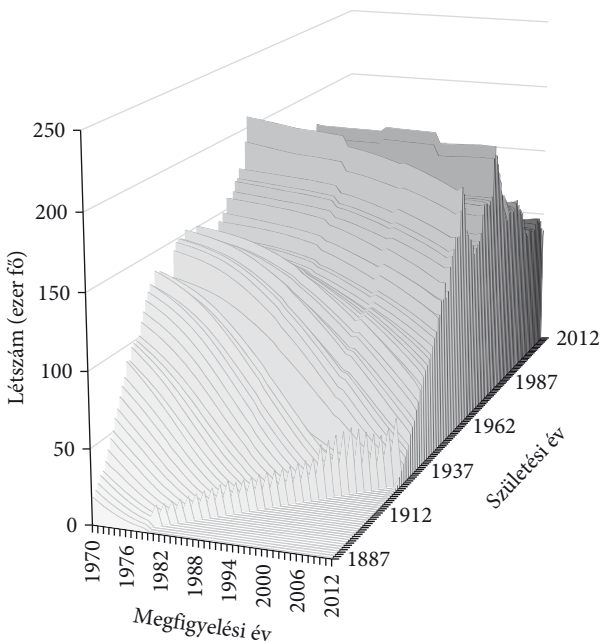
³ Ratkó Anna miniszter nevéhez kötődik az abortusztilalom az 1950-es évek első felében.

tetszőleges megfigyelési év között nem látszik szembeszökő változás a kohorszok megoszlásának tekintetében, de a későbbiekben bemutatásra kerülő statisztikai vizsgálat épp ennek ellenkezőjéről tett tanúbizonyságot.

A kohorszok *idősoros vizsgálata* megmutatja az azonos évben születettek létszámait a különböző megfigyelési években. A 3. ábrán továbbra is láthatjuk az előbb említett sajátosságokat, megfigyelhető továbbá a mérséklődő csökkenés minden kohorsz létszámában, amely a korábbi születési évek esetében jóval meredekebb. A csökkenés természetes: az életpálya vége felé a halálozási valószínűségek értéke nagymértékben emelkedik, az idő előrehaladtával javuló mortalitás pedig a mérséklődő csökkenést indokolja.

3. ábra

A teljes népesség kohorszok szerinti bontásban az 1970 és 2012 közötti évekre



Az adatok forrása: KSH.

Az egyes kohorszok életpályái nagyon hasonlóknak látszanak, ezek hasonlóságát homogenitásvizsgálattal is tetteltük. Eredményeink meglehetősen vegyes képet mutattak. Példaként említenénk, hogy 2001 és 2000 összehasonlításakor sokkal nagyobb különbségeket tapasztalunk a létszámok között, mint más évek összevetésekor. Ennek okát vélhetően a népszámlálási korrekció hiányában kell keresni.

A kohorszok vizsgálatát nemek szerint is elvégeztük, és mind ilyen bontásban, mind összességében azt tapasztaltuk, hogy bizonyos kohorszok jóval nagyobb eltérést mutatnak az előző kohorsszal való összehasonlításakor, mint mások. Nullhipotézisünk szerint a koréves megoszlások nem mutatnak szignifikáns eltérést. Feltételezésünket általában 40 év adatai alapján tetteltük, de természetesen a fiatalabb generációk esetében ennél kevesebb év adatai álltak rendelkezésünkre.

Így a homogenitásvizsgálathoz használt χ^2 -eloszlás szabadsági foka is egészen az 1964-es kohorszig 40 volt, ezután viszont (mivel 2012-ig álltak adatok rendelkezésünkre) értelemszerűen ez a szabadsági fok egyre csökkent. Vizsgálatunk során figyelmen kívül hagytuk a nagyon kis létszámú kategóriákat, így esetenként a szabadsági fok még ennél is kisebb volt.

A legnagyobb eltéréseket a nők esetében az 1955-ben, 1960-ban, illetve 1962-ben születettek kohorszainál tapasztaltuk (1. táblázat). Ugyanezen kohorszok a férfiak esetében is kiugróan magas χ^2 -értékeket produkálnak, de sajnos itt még több kiugrás is fellelhető: gyakorlatilag az 1950 és 1965 között születettek közül egyik kohorsz sem mutat statisztikailag azonos időbeli eloszlást az egy évvel korábbihoz képest.⁴

1. táblázat

A előző évben születettekhez képest legnagyobb eltéréseket mutató kohorszok homogenitásvizsgálata

Születési év	Nők		Férfiak	
	χ^2	p-érték	χ^2	p-érték
1955	679 163,84	0,00	776 662,26	0,00
1960	66 360,71	0,00	61 469,97	0,00
1962	58 318,11	0,00	60 924,96	0,00

Mindezekből látható, hogy még a demográfiai adatok ismeretében is nehéz egy-egy kohorsz életpályájára vonatkozóan előrejelzést készíteni, hiszen nem feltételezhetjük azt, hogy az egyes kohorszok életpályái minden esetben azonos képet mutatnak. Ezt tehát mindenképpen érdemes szem előtt tartanunk az ONYF jogosultsági adatbázisának elemzésekor.

Foglalkoztatottsági elemzés

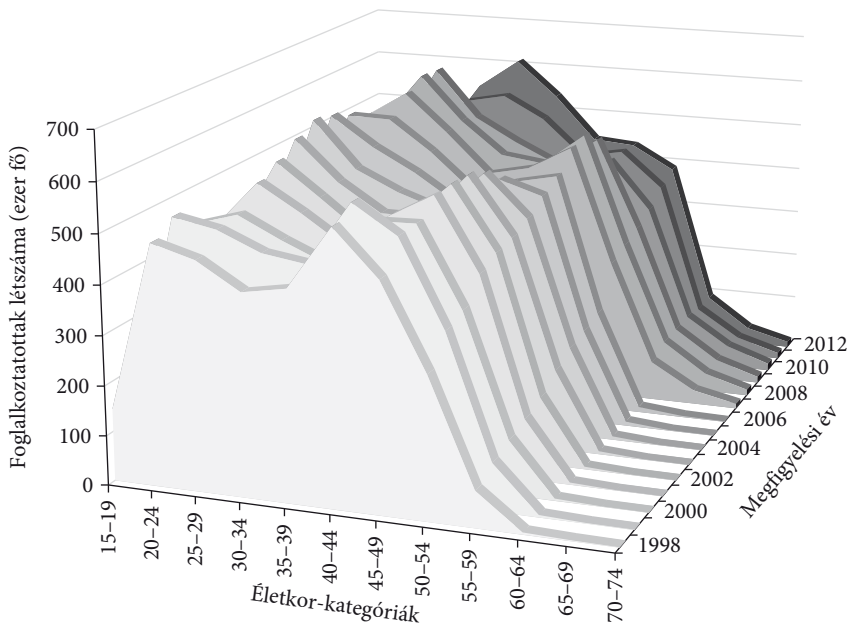
A következőkben ismét az ONYF jogosultsági adatbázisának elemzésére koncentrálnunk, pontosabban azokra, akik a megfigyelt időszak bármelyik negyedében jogosultságot szereztek. Az adatok megbízhatóságának vizsgálata érdekében más forrásokból is igyekeztünk azokat ellenőrizni. Az ONYF-adatbázisban szerepelnek olyanok, akik valós jogosultságot szereztek, és olyanok is, akik inaktív jogosultság (például gyes, gyed vagy más típusú segély) révén kerültek be. Jelen esetben az előbbiekre koncentrálnunk, vagyis azokra, akik a vizsgált időszak bármelyik negyedében fel tudnak mutatni legalább egy nap aktív jogosultságszerzést.

A negyedéves bontásnak egyszerű oka van: a foglalkoztatottsági adatok ellenőrzését a KSH korcsoportonként és nemenként bontott foglalkoztatottsági táblázatai

⁴ Megjegyeznénk továbbá, hogy a vizsgált folyamatok jellemzően lassan, több éven át szokták hatásukat kifejteni, így nem csupán az egymást követő évek összehasonlítása, de néhány éves eltéréssel való összevetés is indokolt lehetne. Ezt végül nem végeztük el, mivel legtöbbször már az egymást követő évek esetében is statisztikailag igazolt a különbözőség.

4. ábra

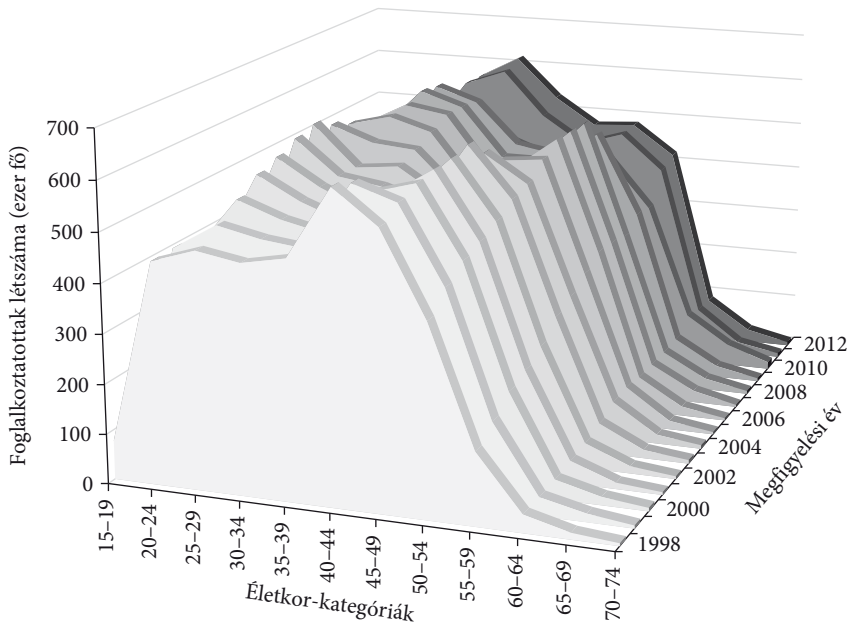
Az ONYF-adatbázis szerinti foglalkoztatottság



Az adatok forrása: ONYF.

5. ábra

A KSH adatai szerinti foglalkoztatottság



Az adatok forrása: ONYF.

segítségével kívántuk megvalósítani, így a lehetőségekhez mérten próbáltunk hasonló módszertannal dolgozni. Ez a törekvésünk eleve kudarcra volt ítélve, hiszen a két adatbázis eltérő módon definiálja a foglalkoztatást. A KSH felméréseiben foglalkoztatottnak számít az, aki az adott héten legalább egy órát dolgozott – erre vonatkozó felmérést pedig havonta készítenek. Az ONYF jogosultsági adatbázisa azonban ilyen részletességgel nem tartalmazott adatokat, így a legtöbb, amit tehattünk, hogy megbecsültük azok számát, akik az egyes negyedekben legalább egy napot munkában töltöttek, majd ezek átlagával becsültük a foglalkoztatottságot nemenként, korcsoportonként. Eredményeinket a 4. ábra tartalmazza, a KSH adatait pedig az 5. ábra mutatja a teljes népességre.

Nem csupán a létszámok tekintetében van különbség a két adatbázis között, de komoly különbségeket tapasztaltunk a foglalkoztatottak létszám szerinti megoszlásában is. Ez utóbbi állításunkat az ONYF- és a KSH-adatbázis létszámainak hányadosai is alátámasztják, amelyeket az 2. táblázatban közlünk.

2. táblázat

Az ONYF-foglalkoztatottak a KSH-adatbázis létszámainak arányában (százalék)

	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012
15–19	186	192	200	201	244	279	312	370	418	545	624	734	646	786	906
20–24	109	114	113	115	114	115	126	131	133	136	144	148	124	139	146
25–29	97	99	95	97	99	100	102	104	104	103	103	103	84	91	97
30–34	88	95	92	96	97	98	100	99	100	105	106	106	89	93	97
35–39	88	94	88	91	90	90	90	92	93	94	95	95	82	91	99
40–44	89	95	91	93	92	96	95	96	99	99	102	102	93	99	101
45–49	84	91	89	94	94	92	95	95	97	100	99	96	86	94	94
50–54	76	82	81	88	89	87	88	89	89	97	99	101	90	95	94
55–59	52	57	55	58	59	62	66	64	65	92	101	97	93	95	98
60–64	15	15	13	11	11	12	11	13	20	124	152	149	144	115	126
65–69	19	10	7	6	6	6	7	10	20	133	182	194	200	172	146
70–74	17	11	6	5	5	4	4	10	23	248	229	234	270	169	178

A két adatbázis hasonlóságát homogenitásvizsgálat segítségével is megvizsgáltuk, és várakozásainknak megfelelően minden évben el kellett utasítani a nullhipotézist, miszerint a két minta azonos eloszlásból származik. A 3. táblázat az egyes évekhez tartozó χ^2 -értékeket mutatja, amelyek mindegyikéhez 11-es szabadsági fok, valamint 0,000 p -érték tartozik.

Természetesen a korábban említett módszertani sajátosságok nagyban befolyásolhatták az eredményeket, de a rendelkezésünkre álló adatokból ezek korrekcióját számos nem tudtuk elvégezni.

Ezt követően a foglalkoztatás népességhez viszonyított arányát vizsgáltuk különböző megfigyelési évekre, kohorszonként és nemenként. Ennek illusztrálásához

a 6. ábrán idősoros bontást közlünk, ahol az 1955 előtt születetteket (a kevés rendelkezésünkre álló adat miatt) elhagytuk.

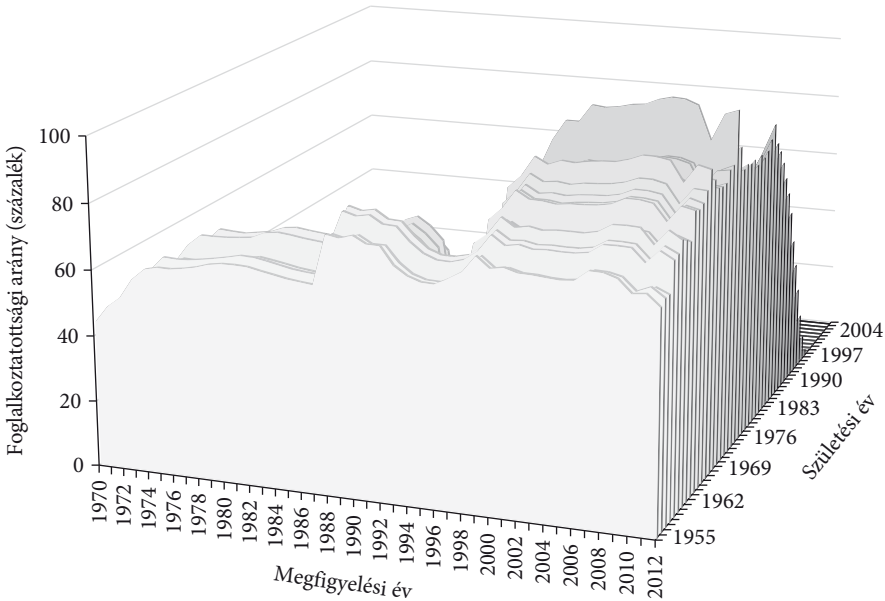
3. táblázat

Naptári évek szerinti homogenitásvizsgálat

Év	χ^2	Év	χ^2	Év	χ^2
1998	95 020	2003	118 102	2008	85 689
1999	95 495	2004	138 744	2009	85 098
2000	101 783	2005	148 434	2010	86 883
2001	96 947	2006	146 874	2011	83 535
2002	104 008	2007	74 726	2012	86 808

6. ábra

Kohorszok foglalkoztatási aránya idősoros bontásban (százalék)



Az adatok forrása: ONYF.

A foglalkoztatottsági arány első ránézésre meglehetősen stabilnak tűnik. Ezt ismételtelen homogenitásvizsgálattal teszteltük, eredményeink pedig igazolják a kohorszok közötti szemmel látható hasonlóságot. Minden egyes kohorszt az előzőhöz viszonyítva a 4. táblázatban szereplő eredményeket kaptuk.

A jogosultsági adatbázis önmagában is rengeteg érdekességet, magyar történelmi sajátosságot hordoz magában, de mint az a fentiekből jól látható, a KSH foglalkoztatotti adataival való statisztikai azonosságát nem állapíthatjuk meg. Természetesen érdemes lehet elgondolkozni azon, hogy miért ilyen kiemelkedően nagyok a

különbségek, de emellett megállapíthatjuk azt, hogy a 25–59 éves korcsoportokban az adatok meglehetősen megbízhatónak tűnnek, mivel ezekben a kategóriákban a KSH is hasonló foglalkoztatottsági adatokat tart nyilván.

4. táblázat

Homogenitásvizsgálat kohorszról kohorszra

Év	χ^2	p -érték	Év	χ^2	p -érték
1956	2,80	1,00	1964	7,69	1,00
1957	12,96	1,00	1965	48,40	0,23
1958	20,00	1,00	1966	3,23	1,00
1959	21,66	1,00	1967	3,43	1,00
1960	119,56	0,00	1968	3,62	1,00
1961	12,47	1,00	1969	4,05	1,00
1962	94,34	0,00	1970	12,88	1,00
1963	6,11	1,00	1971	547,12	0,00

A modellpontok bemutatása

A modellpontok kialakításához az ONYF 1970-től 2012-ig terjedő, 9 millió főre vonatkozó adminisztratív adatait kaptuk meg.⁵ A modellpontok felhasználása megkönnyíti az olyan nyugdíj-hatásvizsgálatok elvégzését, amelyek a társadalom egészét, a jövő generációt is érintik.

Mivel a nyugdíjrendszer fenntarthatósága szempontjából a bevételek és a kiadások együttes vizsgálata elengedhetetlen, ezért két állományból képeztünk modellpontokat.

1. A jogosultságszerzőkről sok egyedi adat áll az ONYF rendelkezésére, de olyan személyi jellemzők, mint iskolai végzettség, családi állapot, gyermekek száma,⁶ közvetlenül nem ismertek. További gondot okoz, hogy a járulékfizetők köre nem reprezentatív része a népességnek, mint a korábbiakban láttuk: életkor szerint nagy eltérések tapasztalhatók.

2. Az ellátottakról a korábbi személyi adatok (például foglalkozás, szolgálati idő hossza) nem ismertek. Részletes adatok érhetők el az ellátás típusáról és a járadék nagyságáról születési évekre és nemenként.

A modellpontok képzése során munkánk elengedhetetlen része volt a 2011. évi népszámlálás adataival való összevetés, amely viszonyítási pontot jelentett a jogosultságszerzők adatainak és megoszlásának ellenőrzésére, valamint az újonnan munkába állók modellpontjainak képzésében.

⁵ A munka lépéseit részletesen ismerteti Kovács [2014].

⁶ Gál Róbert Iván és Törzsök Árpád jelen számban írt arról, hogyan kezelhetők ezen hiányok.

A modellpontok kialakításakor az első döntési kritérium az életkor volt, hogy biztosítani tudjuk az átfedésmentes csoportosítást. Ezért a 16–61 éveseket tekintettük aktív korúaknak, és a 2012-ben érvényes nyugdíjkorhatár szerint a 62 éveseket vagy ennél idősebbeket már elkülönítve soroltuk modellpontokba. Ezzel a döntéssel leegyszerűsítettük a modellpontok képzését, mert kiküszöböltük a korhatár alatti nyugdíjasok és a korhatár felett dolgozók csoportosítását.

A jogosultságszerzők csoportosítása

A hatmillió 16–61 éves egyén adataiból hét jellemzőt használtunk fel a modellpontok képzése során.

I. *Életkor* (összesen 46 korév).

II. *Nem*.

III. *A lakóhely alapján* három régióba soroltuk a modellpontokat: 1. központi (Közép-Magyarország), 2. keleti (Észak és Alföld), 3. nyugati (Dunántúl).

IV. *A településtípus szerint* is három kategória maradt: főváros, város és község.

V. *A foglalkozás* (a legutolsó évi ismert FEOR-főcsoport) alapján négy csoportot képeztünk: 1. diplomás foglalkozások (1., 2. és 3. főcsoport), 2. kereskedelmi és irodai foglalkozások (4. és 5. főcsoport), 3. fizikai foglalkozások (6., 7. és 8. főcsoport), 4. egyéb foglalkozások (0. és ismeretlen főcsoportok).

VI. *A foglalkoztatás* öt kategóriája *Vékás Péter* e számban megjelent írásában szereplő csoportosításon alapul, amely a foglalkoztatottakat a következő öt főszegmensbe sorolja: 1. magasan foglalkoztatottak, 2. közepesen foglalkoztatottak, 3. alacsonyan foglalkoztatottak, 4. tartós (például rokkantsági) ellátásban lévők és 5. inaktív jogviszonnyal (például anyasági ellátással) rendelkezők.

VII. Végül a *jövedelemadatok kvintiliseit* képezve további öt osztályt alakítottunk ki.

A fenti hét változó alapján 165 600 modellpont képezhető, amelyek között számos ritka, kis gyakoriságú is előfordult. Ezért további összevonásokkal csökkentettük a modellpontok számát. Az összevonásokat követően a 16–61 évesek körében 92 720 modellpont áll rendelkezésünkre, ahol egy modellpont átlagosan 58 embert foglal magában, a legnépesebb modellpontot pedig 2983 egyén alkotja.

A modellpontok teljes körűvé tételéhez a 2011. évi népszámlálási adatok 10 százalékos véletlen mintáját használtuk fel. A KSH kutatószobájában a háztartás- és személyszintű anonimizált adatok érhetőek el. Az adatforrások közötti egyéves eltérés miatt az egyéneket egy évvel „megöregítettük”, majd az életkor mellett a nem, régió és aktivitás kategóriái ismeretében arányosítottuk. Így a 92 720 modellpont 6,3 millió embert reprezentál.

A népszámlálás 10 százalékos mintája alapján elmondható, hogy a törzsélekorokban és a tipikus foglalkoztatás mellett a KSH mintája és az ONYF adatbázisa jól megfelel egymásnak. Jelentősebb eltérések vannak a szélső életkorokban, a fiatalok és a nyugdíj előtt állók korcsoportjaiban, valamint a nem szabályos

foglalkoztatási kategóriákban. Ezért a minta alapján történő mérések és adatpótlások helyett a teljes népszámlálás adatait felhasználva érdemes elkészíteni a modellpont-gyakoriságok korrekcióját.

Az öregségi ellátottak modellpontjai

Az ONYF teljes adminisztratív adatállománya közel 2,7 millió főt tartalmazott, és ebből 2012-ben a legalább 62 éves, nyugdíjban részesülők száma 1,5 millió fő volt. Az életkor maximumát 104 évben⁷ határoztuk meg, így 43 korév, két nem és két szegmens (az utolsó évben aktív vagy inaktív) szerint vizsgálva az ellátottakat, 172 modellpontot készítettünk. A régiós és településtípus-adatok oly mértékben hiányosak voltak, hogy ezeket itt nem tudtuk kategóriaképzésre felhasználni. Az „üres”, kevés egyént tartalmazó kombinációkat összevonva, 156 modellpont maradt, átlagosan 9865 egyénnel. A legnépesebb modellpontot 57 823 egyén alkotja.

A modellpontok teljes körűvé tétele az ONYF 2013. januári állományi adataival készült. Az öregséginyugdíj-ellátásban részesülőket életkor és nem szerinti bontásban arányosan felszorozva, a 156 modellpont kétmillió ember mikroszimulációs vizsgálatára alkalmas.

Munkánk összegzésül elmondhatjuk, hogy a két részállományon előállított közel 93 ezer modellpont reprezentálja a magyar felnőtt népességet, és felhasználásuk hatékonyan segítheti a hatásvizsgálatokat.

Hivatkozások

- DEKKERS, G. J. M. [2007]: Construction of a dynamic cross-sectional micro simulation model: the state of affairs. AIM Interim Research Report, 4.10. task 4.1. Brüsszel.
- DELOITTE [2011]: The summary based on the Final Project Report of the Dynamic Microsimulation Model of the Czech Republic. Deloitte Summary Report.
- EVERITT, B. S.–LANDAU, S.–LEESE, M.–STAHL, D. [2010]: Cluster Analysis. 5. kiadás, John Wiley and Sons, New Jersey.
- FULLER, W. A. [2009]: Sampling statistics. John Wiley and Sons, New Jersey.
- FÜSTÖS LÁSZLÓ–KOVÁCS ERZSÉBET–MESZÉNA GYÖRGY–SIMONNÉ MOSOLYGÓ NÓRA [2007]: Alakfelismerés. Sokváltozós statisztikai modellezés a társadalomtudományokban. Új Mandátum Kiadó, Budapest.
- GÁL RÓBERT IVÁN–HORVÁTH ANDRÁS–ORBÁN GÁBOR–DEKKERS, GIJS [2009]: PENMI-CRO. Monitoring pension developments through micro socioeconomic instruments based on individual data sources. Feasibility study. Final Report for The European Commission. Társi, Budapest, <http://ec.europa.eu/social/BlobServlet?docId=4300&langId=en>.
- HOLTZER PÉTER (szerk.) [2010]: Jelentés a Nyugdíj és Időskor Kerekasztal tevékenységéről. Miniszterelnöki Hivatal, Budapest.

⁷ 100 éves vagy ennél idősebbek száma 2012-ben összesen 23 fő (köztük 6 férfi), ezek további bontása statisztikailag nem indokolt.

- HORVÁTH GYULA [2010]: A hatásvizsgálat mikroszimulációs modellje. Megjelent: *Holtzer* (szerk.) [2010] 7. melléklet 146–172. o.
- KLEINBAUM, D. G.–KLEIN, M. [2010]: *Logistic regression*. 3. kiadás, Springer-Verlag, New York.
- KOVÁCS ERZSÉBET [2014]: *Modellpont-csoportosítás az ONYF mikroszimulációs modelljéhez* (Tanulmány), Budapest.
- KOVÁCS ERZSÉBET–SZÜLE BORBÁLA–FLISZÁR VILMOS–VÉKÁS PÉTER [2011]: *Pénzügyi adatok statisztikai elemzése*. Tanszék Kft., Budapest.
- UNITED NATIONS [1967]: *Manual IV. Methods of estimating basic demographic measures from incomplete data*. United Nations Publications, Sales No. 67.XIII.2. <http://www.un.org/esa/population/techcoop/DemEst/manual4/manual4.html>.
- ZAIDI, A.–RAKE, K. [2001]: *Dynamic Microsimulation Models: A Review and Some Lessons for SAGE*. SAGE Discussion Papers, No. 2.