

Modellpontok szerepe a nyugdíj-hatásvizsgálatban

Kovács Erzsébet-Rétallér Orsolya-Vékás Péter

JEL kódok: C53, H55

A mikroszimulációs nyugdíjmodell elkészítésekor az adatok körének meghatározása is megfontolásra került. Rendelkezésünkre állt az Országos Nyugdíjbiztosítási Főigazgatóság (ONYF) 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 miatt jelentős számításigényt csökkenheti az, ha az egyedi adatok helyett embercsoportokra, ún. modellpontokra készül a mikroszimuláció. A cikk a módszertani megfontolások ismertetése mellett bemutatja a modellpontok előállításának lépéseit, a népszámlálási adatokkal történő egybevetés nehézségeit, végül magukat az előállított modellpontokat.

1. Mikroszimuláció a nyugdíjmodellezésben

A nagy mennyiségű egyénre vonatkozó, részletes információkat felhasználó, ún. *mikroszimulációs nyugdíjmodelleket* Gál-Horváth-Orbán-Dekkers (2009) két kategóriába: a *standard* és a *mikroszimulációs modellek* bővebb családjába sorolja, és mindkettőn belül további alkategóriákat különböztet meg. Az Országos Nyugdíjbiztosítási Főigazgatóság MIDAS_HU nyugdíjmodellje a Belgiumban használatos MIDAS_BE modell (ld. Dekkers, 2007) magyarországi adaptációja, és ahhoz valamint a korábban Magyarországon használatos NYIKA modellhez (ld. Holczer, 2010) hasonlóan a dinamikus mikroszimulációs modellek (ld. Zaidi-Rake, 2001) családjába tartozik. A Deloitte tanácsadó és könyvvizsgáló cég a Csehországban alkalmazott dinamikus mikroszimulációs modellel 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 modellezzik 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, eloszlásjellegű eredményeket szolgáltatnak pl. 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 érhetőek:

- Magas implementációs (szoftver-, szakértői, humán-) költségekkel járnak.
- Magas a számításigényük (szoftver és hardver szempontjából egyaránt), és hosszú a számítási idejük.
- Magasak a bemenő adatokkal szemben támasztott követelményeik, és bonyolultak az alkalmazás előfeltételei.
- Nehezebb biztosítani az egyéb (makroökonómiai, demográfiai, stb.) feltevésekkel való konzisztenciájukat.

A két felsorolás 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 meghatározása is megfontolásra került. A teljes ONYF adminisztratív adatállomány a rendelkezésünkre állt, és négy lehetőség közül választhattunk:

- A teljes ONYF adminisztratív adatállomány felhasználása ellen szólt a hatalmas számításigény.
- A jogosultságszerzőket reprezentáló véletlen minta ellen szólt az eljárás költsége valamint az új egyének besorolásának nehézsége (a későbbiekben ezt még részletezzük).

- Az adminisztratív adatállományból kiválasztott néhány kohorsz, akik már közel teljes jogosultságszerzési időszakokkal rendelkeznek. (Az 1955-1959 között születettek elemzését tanulmányunk 4. alfejezete ismerteti.)
- Az adatállományból kiindulva különböző számosságú ember-típusokat képzünk, akik nemük, életkoruk és további jellemzőik alapján reprezentálják a jogosultságszerzőket és a nyugdíjban részesülőket is.

Az Országos Nyugdíjbiztosítási Főigazgatóság szakértői úgy döntöttek, hogy a MIDAS_HU modell munkaerőpiaci moduljában egyéni szereplők helyett szakértői szempontok alapján létrehozott, homogén csoportokat, ún. *modellpontokat* szerepeltetnek. E homogén modellpontok kialakítása következtében ideális esetben az egyszerűsítés révén az egyénre vonatkozó információknak csak jelentéktelen része megy veszendőbe, megőrizve így a mikroszimulációs megközelítés előnyeit. Ezzel egyidejűleg a modellpont-adatbázis teljes adatbázishoz képesti jóval kisebb mérete következtében e módszer segítségével minimalizálhatók a mikroszimulációs modellek fent említett költségbeli, számítási és időbeli hátrányai.

2. Modellpontok előállítás

A nyugdíjmodellezés hosszú távú feladat, valamint a jogosultságszerzők és az ellátottak nagy száma miatt jelentős számítást igényel. Ez utóbbi szempontok figyelembe vételével végül mintegy százezer körüli modellpont képzésére került sor, melyeknek egyszerre a következő két célkizűzésnek 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 elemzéseknek, Augusztinovics Mária megfogalmazásában hatásvizsgálatoknak¹. Ehhez

¹ Augusztinovics Mária nemcsak a Jelentés (Holczer et al., 2010) egyik szerzője, hanem a több éves munka során számos kérdésben ötletadó, inspiráló volt. Ezúton is köszönettel emlékezünk rá és munkásságára.

szükséges volt a modellpontok képzéséhez felhasználandó ismérvek megfelelő kiválasztása 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-es népszámlálás részletes adatai. Az összevetések elvégzése után gyakorlati 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 eljárás családot vettük alaposabban vizsgálat alá.

a) Mintavételi módszerek

A hagyományos statisztikai mintavételi módszerek (véletlen, egyszerű vagy rétegzett mintavételi eljárások, ld. *Fuller, 2009*) jól használhatók pl. 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.

b) Klaszterező eljárások

A struktúrafeltáró módszerek közül a legalkalmasabb osztályozó eljárás a klaszterelemzés (hierarchikus, nemhierarchikus vagy kétlépcsős algoritmusokkal, ld. *Everitt et al., 2010, Füstös et al., 2007, Kovács et al., 2011*). Ez a módszertan számítási igénye miatt többmillió sokaságra és ilyen nagyszámú klaszter keresésére az időbeli korlátok figyelembe vételével nem használható. Mindkét eljárás-változatnál gondot okozhat az is, hogy a rendelkezésre álló változók mérési skálája nagyon különböző, vannak kategória-, és vannak érték-változók is, amelyek együttes kezelése statisztikailag nehezen valósítható meg. Az egyes klaszterek, mint modellpontok jellemzőihez nem adhatók meg kategória-határok, és nem

biztosítható a megfigyelt jellemzőkkel le nem írt további – jelenleg még esetleg meg sem született – újabb személyek számára klaszterek kialakítása.

c) Többdimenziós kategória-kombinációk képzése

Ezzel a szemlélettel változónként kategóriákat képezünk, és megfelelően alátámasztható kategória-összevonások 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 hangsúlyt kell fordítani a többdimenziós kontingenciatábla egyes celláiban található hiányzó adatok feltöltésére. E megközelítést alkalmazta a Magyarországon korábban alkalmazott NYIKA modell (*Holczer et al.*, 2010) is.

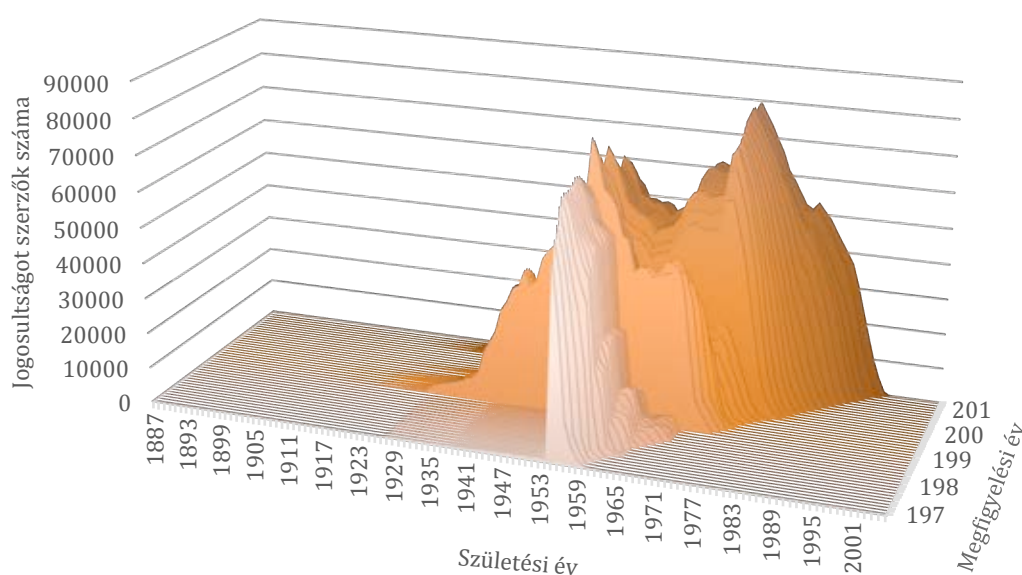
Mivel a jövőbeli állományalakulás modellezése céljából új modellpontokat is szükséges létrehozni (pl. a később munkába álló jelenlegi iskolások számára), ezért lényeges szempont, hogy ezek az ONYF adatállományából kialakított modellpontok rendszerébe könnyedén besorolhatók legyenek. Az a) és b) módszercsaládba tartozó eljárások esetén ez hasonlósági/távolsági mértékek vagy csoporthoz tartozást előrejelző, valószínűségi becslést is adó többváltozós statisztikai módszerek használatával érhető el. Hasonlósági mértékek segítségével azonban nem rendelhető az egyenletes súlyozáson kívül más ésszerűen alátámasztható súlyrendszer a változókhoz, a többváltozós klasszifikációs eljárások (pl. logisztikus regresszió, diszkriminancia-elemzés stb., ld. *Füstös et al.*, 2007, *Kleinbaum-Klein*, 2010, *Kovács et al.*, 2011) használata pedig a nagy csoportszám miatt nem volt lehetséges. E szakmai indokok mérlegelésével végül a MIDAS_HU modell munkaerő-piaci moduljában a c) módszercsalád használata mellett döntöttünk.

3. 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

körü adatok álltak rendelkezésünkre. Vízvásztó évnek tekinthetjük ilyen szempontból az 1988-as évet, mert bár ekkor sem volt még 100%-os a feldolgozottság, mégis látványosan több adat áll már rendelkezésünkre, mint egy évvel korábbról.

Az adatok digitalizálása során magasabb prioritást kaptak az 1955-1959-es kohorszok. Ezek az évjáratok felülreprezentáltak mondhatók, de egyúttal értékes információt szolgáltatnak számunkra az adatok idősoros vizsgálatához. Az előbbieket illusztrálja az 1. ábra, amely egy háromdimenziós hisztogram: a születési évek és a megfigyelési évek bontásában ábrázolja a jogosultságot szerzők számát.



1. ábra: Jogosultságot szerzők száma (az adatok forrása: ONYF)

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

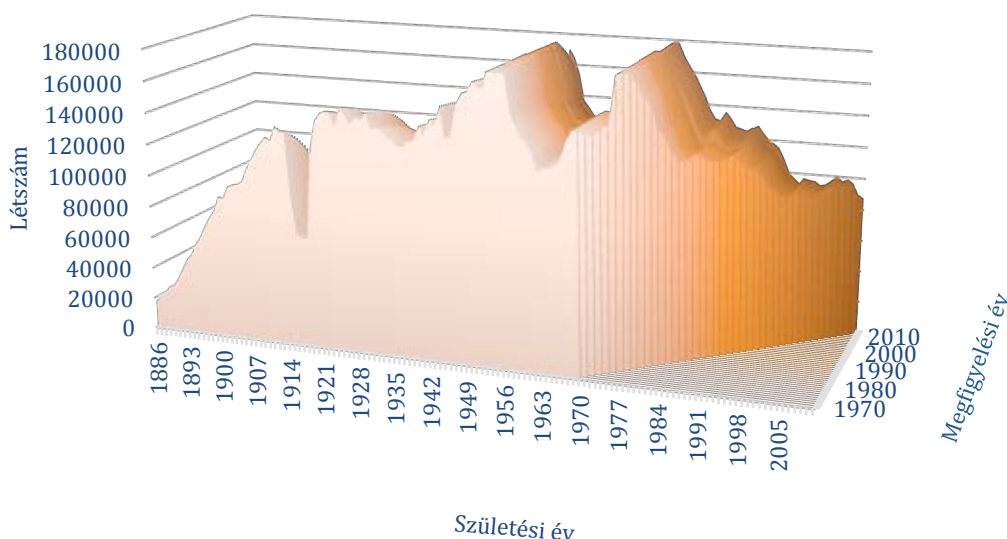
Természetesen a nemzetközi szakirodalomban is találhatunk példát olyan esetekre, amikor egy népességi adatokat tartalmazó adatbázis hiányosságait statisztikai módszerekkel igyekeznek kipótolni. Ilyen esetek például a fejlődő országokban a nem regisztrált születések és halálozások miatti utólagos népességszám korrekciók, de gondolhatunk Kína speciális esetére is, ahol az egy gyermek politikának köszönhetően sokszor szándékosan titkolják a hatóságok elől, így nem is regisztrálják

a megszületett gyerekeket. 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).

A foglalkoztatotti adatbázis vizsgálatához elsőként a teljes magyar lakosság vizsgálatát végeztük el. A teljes populáció behatóbb vizsgálata lehetővé tette számunkra, 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.

3.1. 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.

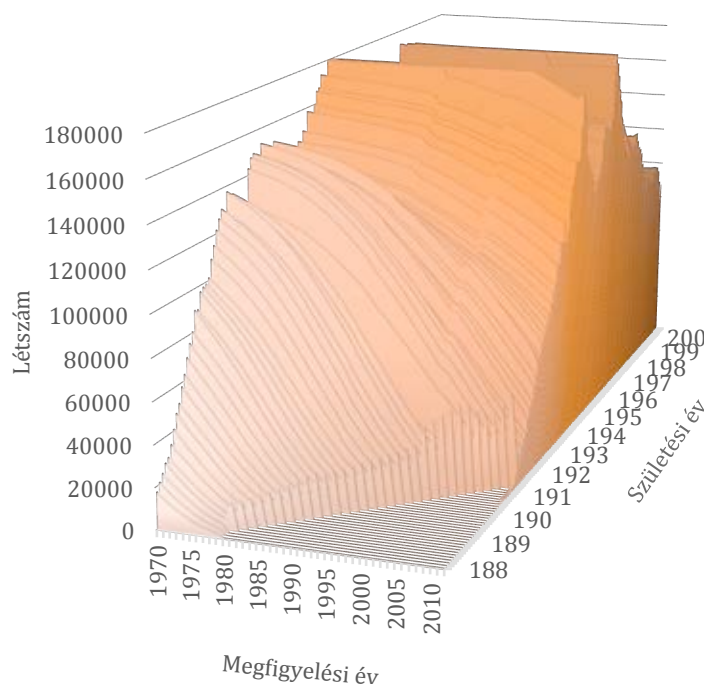


2. ábra: A magyar népesség megoszlása (az adatok forrása: KSH)

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

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-es, majd az 1974-1975-ös években születettek létszáma jóval magasabb, mint más kohorszoké. Emellett megjegyeznénk, hogy bár első pillantásra két 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 monoton csökkenés minden kohorsz létszámában, amely a korábbi születési évek esetében jóval meredekebb. A csökkenés természetesen érthető is, hiszen az életpálya vége felé a halálozási valószínűségek értéke nagymértékben emelkedik, az idő előrehaladtával javuló mortalitás pedig a mérséklődő csökkenést indokolja.



2. ábra: A teljes népesség kohorszok szerinti bontásban (az adatok forrása: KSH)

Az egyes kohorszok életpályái nagyon hasonlóknak látszanak, ezek hasonlóságát homogenitás-vizsgálattal is teszteltük. Eredményeink meglehetősen vegyes képet

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

mutattak. Példaként említenénk, hogy a 2001-es és 2000-es évek összehasonlításakor sokkal nagyobb különbségeket tapasztalunk a létszámok között, mint más évek összehasonlításakor. Ennek oka vélhetőleg a népszámlálási korrekció hiányában keresendő.

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 teszteltük, de természetesen a fiatalabb generációk esetében ennél kevesebb év adatai álltak rendelkezésünkre. Így a homogenitás-vizsgálathoz 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 a következő kohorszoknál tapasztaltuk (az előző évben születettekhez viszonyítva): az 1955-ös kohorsz ($\chi^2=679163,84$; p-érték: 0,00), az 1960-as születésűek ($\chi^2=66360,71$; p-érték: 0,00), illetve az 1962-es kohorsz ($\chi^2=58318,11$; p-érték: 0,00). Ugyanezen kohorszok a férfiak esetében is kiugróan magas χ^2 -értékeket produkálnak (a χ^2 -értékek rendre: 776662,26; 61469,97; 60924,96; az ezekhez tartozó p-értékek mindegyike 0,00), 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.

Megjegyeznénk továbbá, hogy a vizsgált folyamatok jellemzően lassan, több éven át szokták hatásukat kifejteni, így nem csupán az egymást követő évek összehasonlítása, de pár 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.

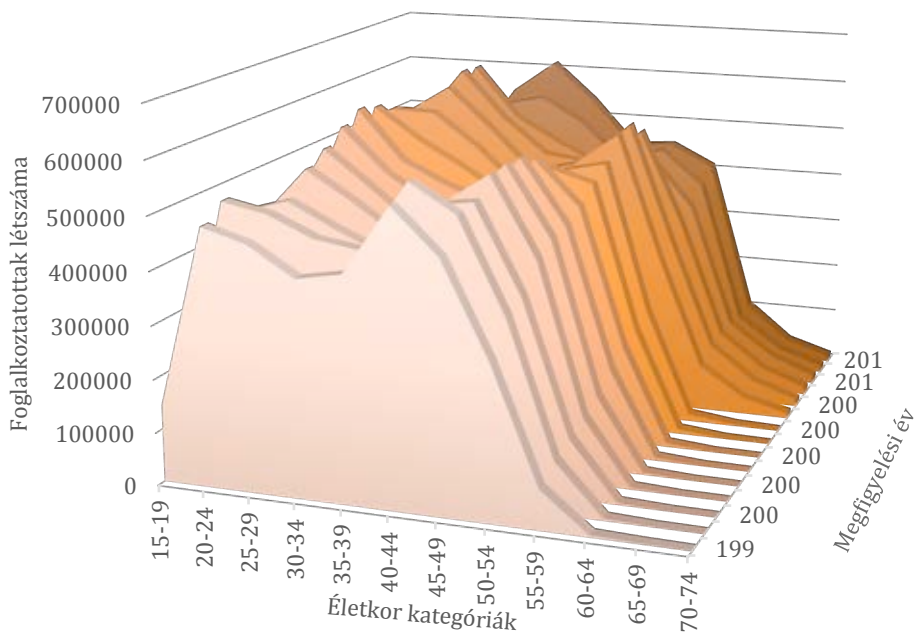
Mindezekből látható, hogy még a demográfiai adatok ismeretében is nehéz egy-egy kohorsz életpályájára vonatkozóan előrejelzést készíteni, hiszen nem feltételezhetjük azt, hogy az egyes kohorszok életpályái minden esetben azonos képet festenek. Ezt

tehát mindenképpen érdemes szem előtt tartanunk az ONYF jogosultsági adatbázisának elemzésekor.

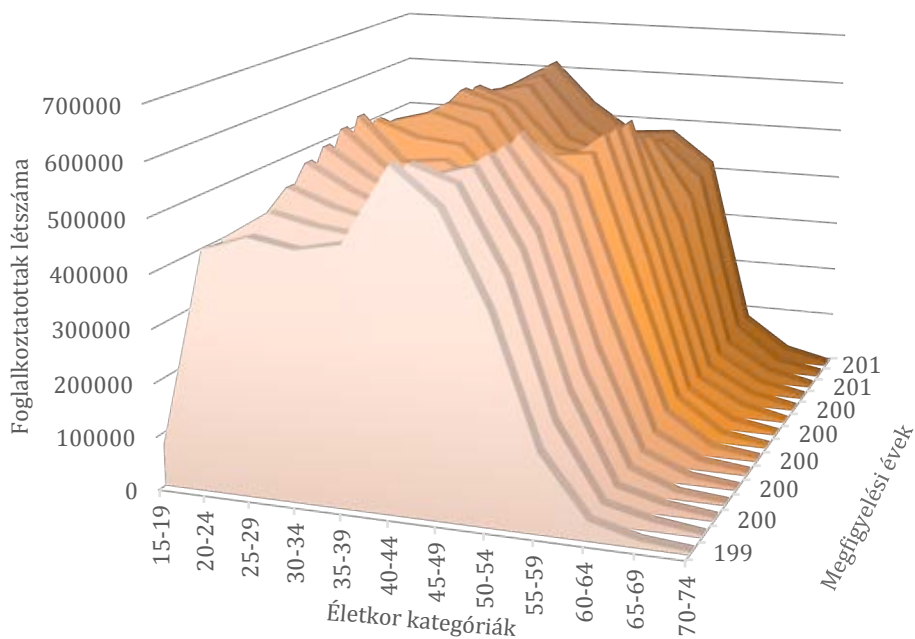
3.2. Foglalkoztatottsági elemzés

A következőkben ismét a jogosultsági adatbázis elemzésére koncentrálunk, pontosabban azokra, akik a megfigyelt időszak alatt bármelyik negyedévben jogosultságot szereztek. Az adatok megbízhatóságának vizsgálata érdekében más forrásokból is igyekszünk azokat ellenőrizni. A jogosultsági adatbázisban szerepelnek olyanok, akik valós jogosultságot szereztek, és olyanok is, akik pszeudó jogosultság (például GYES, GYED, vagy más típusú segély) révén szerepelnek benne. Jelen esetben az előbbiekre koncentrálunk, vagyis azokra, akik a vizsgált időszak bármelyik negyedévében fel tudnak mutatni legalább egy napnyi jogosultságszerzést.

A negyedéves bontásnak egyszerű oka van: a foglalkoztatottsági adatok ellenőrzését a KSH korcsoportonként és nemenként bontott foglalkoztatottsági táblái 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 vonatkozólag felmérést pedig havonta készítenek. A jogosultsági adatbázis azonban ilyen részletességgel nem tartalmazott adatokat, így a legtöbb, amit tehattünk, hogy megbecsültük azok számát, akik az egyes negyedévekben legalább egy napot munkával töltöttek, majd ezek átlagával becsültük a foglalkoztatottságot nemenként, korcsoportonként. Eredményeinket a 4. ábra tartalmazza, a KSH adatait pedig az 5. ábrán közöljük a teljes népességre.



3. ábra: Foglalkoztatottság az ONYF adatbázisa szerint (az adatok forrása: ONYF)



4. ábra: Foglalkoztatottság a KSH adatai szerint (az adatok forrása: ONYF)

Nem csupán a létszámok tekintetében van különbség a két adatbázis között, de komoly különbségeket tapasztaltunk a foglalkoztatottak létszám szerinti megoszlásában is. Ez utóbbi állításunkat a jogosultsági és a KSH adatbázisok létszámainak hányadosai is alátámasztják, melyeket az 1. táblázatban közlünk.

ONYF /KSH	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%

1. táblázat: A KSH és az ONYF foglalkoztatottak arányszámai

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

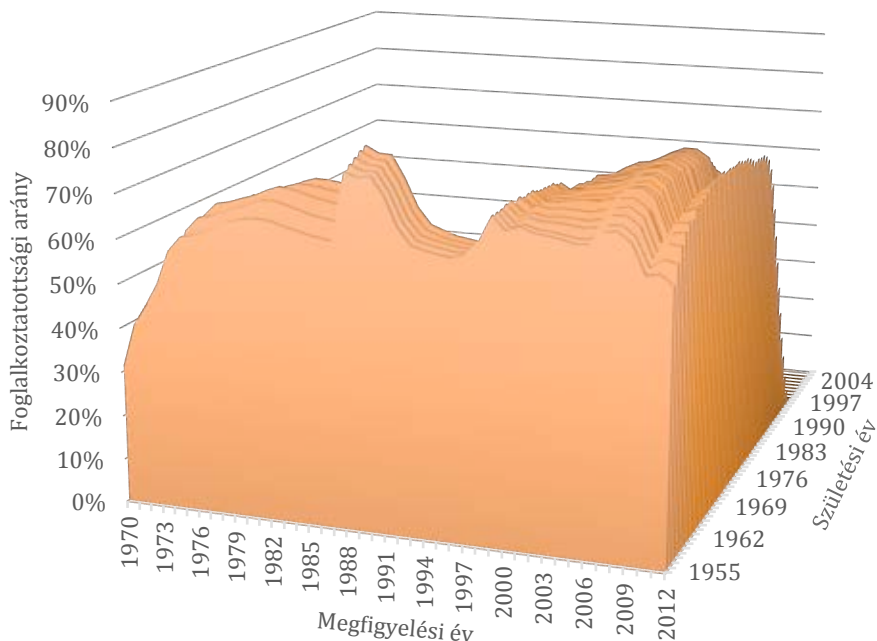
Év	χ^2
1998	95 020
1999	95 495
2000	101 783
2001	96 947
2002	104 008
2003	118 102
2004	138 744
2005	148 434
2006	146 874
2007	74 726
2008	85 689
2009	85 098
2010	86 883

2011	83 535
2012	86 808

2. táblázat: Naptári évek szerinti tesztek

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

Ezt követően a foglalkoztatás népességhez viszonyított arányát vizsgáltuk meg különböző megfigyelési évekre, kohorszokként és nemenként. Ennek illusztrálásához a 6. ábrán idősoros bontást közlünk, ahol az 1955 előtt születetteket (a kevés rendelkezésünkre álló adat miatt) elhagytuk.



5. ábra: Kohorszok foglalkoztatási aránya idősoros bontásban (az adatok forrása: ONYF)

A foglalkoztatottsági arány első ránézésre meglehetősen stabilnak tűnik. Ezt ismételt homogenitás-vizsgálattal teszteltük, eredményeink pedig igazolják a kohorszok közötti szemmel látható hasonlóságot. Minden egyes kohorszt az előzőhöz viszonyítva a 3. táblázatban szereplő eredményeket kaptuk.

Év	χ^2	p-érték
1956	2,80	1,00
1957	12,96	1,00
1958	20,00	1,00

Év	χ^2	p-érték
1964	7,69	1,00
1965	48,40	0,23
1966	3,23	1,00

1959	21,66	1,00
1960	119,56	0,00
1961	12,47	1,00
1962	94,34	0,00
1963	6,11	1,00

1967	3,43	1,00
1968	3,62	1,00
1969	4,05	1,00
1970	12,88	1,00
1971	547,12	0,00

3. táblázat: Homogenitásvizsgálat kohorszról kohorszra

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. A modellpontok bemutatása

A modellpontok kialakításához az ONYF 1970-től 2012-ig terjedő, 9 millió főre vonatkozó adminisztratív adatait kaptuk meg.⁴ A modellpontok felhasználása megkönnyíti az olyan nyugdíjhatásvizsgálatok 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:

- a) A jogosultság-szerzőkről sok egyedi adat áll az ONYF rendelkezésére, de olyan személyi jellemzők, mint iskolai végzettség, családi állapot, gyermekek száma⁵ közvetlenül nem ismertek. További gondot okoz, hogy a járulékfizetők köre nem reprezentatív része a népességnek, mint a 3. fejezetben láttuk: életkor szerint nagy eltérések tapasztalhatóak.
- b) Az ellátottakról a korábbi személyi adatok (pl. foglalkozás, szolgálati idő hossza) nem ismertek. Részletes adatok érhetőek el az ellátás típusáról és a járadék nagyságáról születési évekre és nemenként.

⁴ A munka lépéseit részletesen ismerteti Kovács (2014).

⁵ Gál Róbert Iván és Törzsök Árpád írt arról, hogy kezelhetők ezen hiányok (Gál-Törzsök, 2015).

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

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.

4.1. A jogosultságszerzők csoportosítása

A hatmilliónyi 16-61 éves egyén adataiból 7 jellemzőt használtunk fel a modellpontok képzése során. Az életkor (46 korév) és a nemek összevonás nélkül egyedi azonosítóként szerepeltek.

A lakóhely alapján 3 régióba soroltuk a modellpontokat:

- központi (Közép-Magyarország),
- keleti (Észak és Alföld),
- nyugati (Dunántúl).

A település típus szerint is 3 kategória maradt: főváros, város és község.

A foglalkozás (a legutolsó évi ismert FEOR főcsoport) alapján 4 csoportot képeztünk:

- diplomás foglalkozások (1-es, 2-es és 3-as főcsoport),
- kereskedelmi és irodai foglalkozások (4-es és 5-ös főcsoport),
- fizikai foglalkozások (6-os, 7-es és 8-as főcsoport),
- egyéb foglalkozások (0-ás és ismeretlen főcsoportok).

A foglalkoztatás 5 kategóriája *Vékás* (2015) csoportosításán alapul, mely a foglalkoztatottakat a következő öt főszegmensbe sorolja:

- magasan foglalkoztatottak,
- közepesen foglalkoztatottak,

- alacsonyán foglalkoztatottak,
- tartós (pl. rokkantsági) ellátásban lévők és
- pszeudó jogviszonnyal (pl. anyasági ellátással) rendelkezők.

Végül a jövedelem adatok kvintiliseit képezve további 5 osztályt alakítottunk ki.

A fenti 7 változó alapján 165 600 modellpont képezhető, melyek 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 magába, a legnépesebb modellpontot pedig 2983 egyén alkotja.

A modellpontok teljeskörűsítéséhez a 2011-es Népszámlálási adatok 10%-os véletlen mintáját használtuk fel. A KSH Kutatószobában a háztartás és személy szintű anonimizált adatok érhetőek el. Az adatforrások közötti egy évnnyi eltérés miatt az egyéneket egy évvel „megöregítettük”, majd az életkor mellett a nem, régió és aktivitás kategóriák ismeretében arányosítottuk. Így a 92 720 modellpont 6,3 millió embert reprezentál.

A népszámlálás 10%-os mintája alapján elmondható, hogy a törzs-életkorokban és a tipikus foglalkoztatás mellett jól megfelel egymásnak a KSH minta és az ONYF adatbázis. 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 érdemes a teljes népszámlálás adatait felhasználva elkészíteni a modellpontokra a gyakoriságok korrekcióját.

4.2. 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 évben⁶ 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 település típus adatok oly mértékben

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

hiányosak voltak, hogy ezeket itt nem tudtuk kategóriaképzésre felhasználni. Az „üres”, kevés egyént tartalmazó kombinációkat összevonva 156 modellpont maradt, átlagosan 9 865 egyénnel. A legnépesebb modellpontot 57 823 egyén alkotja.

A modellpontok teljeskörűsítése az ONYF 2013. januári állományi adataival készült, az öregségi nyugdíj ellátásban részesülőket életkor és nem szerinti bontásban arányosan felszorozva a 156 modellpont 2 millió 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, Brussels.

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* (5th edition). John Wiley & Sons, New Jersey.

Fuller, W.A. (2009). *Sampling statistics*. John Wiley & Sons, New Jersey.

Füstös, L. – Kovács, E. – Meszéna, Gy. – Simonné Mosolygó, N. (2007): *Alakfelismerés. Sokváltozós statisztikai modellezés a társadalomtudományokban*. Új Mandátum Kiadó, Budapest.

Gál, R.I. – Horváth, A. – Orbán, G. – Dekkers, G. (2009). *PENMICRO: Monitoring pension developments through micro socioeconomic instruments based on individual data sources: feasibility study* (p. 67.). TÁRKI, Budapest.

Gál, R.I. – Törzsök, Á. (2015). *Háztartásformálódás a MIDAS-HU modellben*. Közgazdasági Szemle évf. sz. old.

Holczer Péter (szerk.) (2010). *Jelentés a Nyugdíj és Időskor Kerekasztal tevékenységéről*. Miniszterelnöki Hivatal, Budapest.

Kleinbaum, D.G. – Klein, M. (2010). *Logistic regression*, 3rd edition. Springer-Verlag, New York.

Kovács, E. – Szüle, B. – Fliszár, V. – Vékás, P. (2011). *Pénzügyi adatok statisztikai elemzése* (egyetemi tankönyv). Tanszék Kft., Budapest.

Kovács Erzsébet (2014). *Modellpont-csoportosítás az ONYF mikroszimulációs modelljéhez* (Tanulmány), Budapest.

United Nations (1967). Manual IV. *Methods of estimating basic demographic measures from incomplete data* (United Nations Publications, Sales No. 67.XIII.2)

Vékás Péter (2015). *Az egyéni munkaerő-piaci aktivitás becslése a MIDAS_HU mikroszimulációs nyugdíjmodellben*. Közgazdasági Szemle évf. sz. old.

Zaidi, A. – Rake, K. (2001). *Dynamic Microsimulation Models: A Review and Some Lessons for SAGE* (p. 40). SAGE Discussion Papers.