

Az egyéni munkaerő-piaci aktivitás becslése a MIDAS_HU mikroszimulációs nyugdíjmodellben

Vékás Péter¹

JEL kódok: H55, C15, C40, C53

A nyugdíj- és egyéb állami ellátórendszerek működésének elemzése területén a hagyományos makromodellek mellett Európában egyre inkább előtérbe kerülnek a *sztochasztikus-dinamikus mikroszimulációs módszerek*, melyek a makrogazdasági aggregátumok várható értékének időbeli előrejelzése mellett azok eloszlásának pontosabb vizsgálatát is lehetővé teszik, így eszközül szolgálhatnak a nyugdíjrendszert érintő politikai és nyugdíjszakmai döntések hosszú távú társadalmi-gazdasági hatásainak valósághűbb elemzésére. Segítségükkel vizsgálhatóvá válik pl. a különböző szocio-ökonómiai csoportok anyagi státuszának várható időbeli változása, az időskori szegénység várható alakulása stb.

Az Európában alkalmazott dinamikus nyugdíj-mikroszimulációs módszerek között az egyik legfejlettebb a belga Federal Planning Bureau intézményben Gijs Dekkers és munkatársai által fejlesztett *MIDAS_BE mikroszimulációs modell* és az annak támogatására szolgáló *LIAM2 mikroszimulációs szoftverkörnyezet*. Magyarországon az állami nyugdíjrendszer működéséért felelős Országos Nyugdíjbiztosítási Főigazgatóság folyamatosan fejleszti a MIDAS mikroszimulációs rendszer magyar körülményekre adaptált változatát, a *MIDAS_HU* modellt, melynek előkészítésében jelen cikk szerzője is részt vett.

Jelen tanulmány a magyarországi munkaerő-piaci aktivitás becslésével foglalkozik, mely a MIDAS modellek *munkaerő-piaci moduljának* meghatározó eleme. Az elemzés során többváltozós statisztikai technikákat (imputálás, szegmentálás, regresszió) alkalmaztunk egy óriási méretű, adminisztratív adatokat tartalmazó adatfájlon. A felhasznált adatfájl több mint négy évtized demográfiai, foglalkoztatottsági és járulékfizetési adatait tartalmazza egyénenkénti bontásban. Bár a becslést a MIDAS modellek logikájának megfelelően keresztmetszeti szemléletben hajtottuk végre, az egyenletekbe az eredeti belga modelltől eltérően hosszmetzeti adatokból nyert információkat is beépítettünk. Ezt egyrészt a rendelkezésünkre álló, jóval hosszabb idősorok információtartalmának optimális felhasználása, másrészt a hazai munkaerő-piacon – elsősorban az 1990-es évek elején – lezajlott jelentős strukturális átalakulások indokolták.

¹ Egyetemi tanársegéd, Budapesti Corvinus Egyetem, Operációkutatás és Aktuáriustudományok Tanszék, tudományos segédmunkatárs, MTA-BCE „Lendület” Stratégiai Interakciók Kutatócsoport, e-mail: peter.vekas@uni-corvinus.hu.

Bevezető, az előzmények áttekintése

A hazai nagy állami ellátórendszerek legjelentősebbike a minden munkavállalóra kötelezően kiterjedő állami öregségi, rokkantsági és hátramaradotti nyugdíjrendszer, melyet első, nem teljeskörű formájában az 1928. évi XL. törvény vezetett be, majd az 1951. évi törvénymódosítások tették teljessé (Szabó [2000]). A rendszer fedezetlen, *felosztó-kirovó* jellegű, vagyis a mindenkori munkavállalók és munkáltatók aktuális járulékbefizetései finanszírozzák az adott évben esedékes járadékszolgáltatásokat. Felosztó-kirovó jellegéből adódóan az állami nyugdíjbiztosítás finanszírozása és hosszú távú finanszírozhatósága erőteljesen függ a népesség kor szerinti megoszlásától, a keresőképes korúak munkaerő-piaci aktivitásától és a jövedelmek eloszlásától. Volumenének megfelelően a nyugdíjkassza pénzáramlásának alakulása a költségvetési, szocio-ökonómiai és általános nemzetgazdasági folyamatokra egyaránt jelentős befolyással bír. A jól megalapozott költségvetési, nemzetgazdasági és szociálpolitikai döntéshozatal érdekében kívánatos és elkerülhetetlen tehát a hazai állami nyugdíjrendszer folyamatainak rövid és hosszú távú, megfelelő informatikai háttérrel és korszerű módszertani eljárásokon alapuló modellezése.

Az Európai Unió országaiban a nyugdíjmodellezés területén a korábbi évtizedekben használt makro- és kohorsz-modelleket kiegészítve az ezredforduló óta eltelt években teret nyertek az ún. *stochasztikus-dinamikus mikroszimulációs modellek* (ld. Li [2011], Li–O’Donoghue [2013] és Zaidi–Rake [2001]), melyek a vizsgált makrogazdasági aggregátumok vagy születési év szerinti kohorszok modellezése helyett a rendszer által érintett járulékfizetők és járadékszolgáltatásban részesülők viselkedését és döntéseit egyéni szinten írják le és jelzik előre. A mikroszimulációs módszerek előnye, hogy a modellezni kívánt aggregátumok várható értékei mellett azok eloszlását is megfelelően képesek előrevetíteni, és az aggregátumokon belül végbemenő összetétel-változások más változókra gyakorolt hatását is statisztikai szempontból adekvát módon képesek kezelni. E modellek az egyéni döntéshozók viselkedését statisztikai-ökonometriai eljárások felhasználásával becsült *viselkedési egyenletekkel*, *intertemporális várható hasznossági függvények* maximalizálásával vagy e két módszercsalád ötvözésével is leírhatják.

Magyarországon a Nemzetgazdasági Minisztérium és elődintézményei által korábban alkalmazott makromodelleket kiegészítve először a Nyugdíj és Időskor Kerekasztal (NYIKA) hozott döntést egy átfogó mikroszimulációs nyugdíjmodell fejlesztéséről (Holtzer [2010]). Az első, részleges modelleket követően az Országos Nyugdíjbiztosítási Főigazgatóság (ONYF) a cél érdekében 2012-ben látott hozzá a Belgiumban használatos *MIDAS_BE*

(*Microsimulation for the Development of Adequacy and Sustainability*, ld. *Dekkers–Belloni* [2009], *Dekkers* [2010] és *Dekkers* [2013]) nagy méretű sztochasztikus–dinamikus mikroszimulációs modell hazai adaptációjához egy közel fél évszázadra kiterjedő, nagy részletességű, egyéni szintű adatokat tartalmazó informatikai adatbázis felhasználásával, 2013-tól az Európai Unió támogatásával.

A MIDAS modell számos egymással kölcsönhatásban álló *modulból és almodulból* (pl. munkaerőpiac, demográfia, foglalkoztatás, jövedelmek, születés, halálozás, házasságok stb.) áll, és az egyéni döntéseket többváltozós statisztikai eljárások segítségével becsült viselkedési egyenletek felhasználásával írja le. A brüsszeli székhelyű Federal Planning Bureau kimondottan a MIDAS modell támogatására fejlesztette ki a *LIAM2 mikroszimulációs modellkörnyezetet*, melyet az ONYF is alkalmaz. A MIDAS modell az egyéb előrejelzésekkel való konzisztencia érdekében számos ponton hazai és uniós intézmények aggregált előrejelzéseire történő *exogén kiigazításokat* tartalmaz.

A gazdasági környezethez és a rendelkezésre álló adatok struktúrájához alkalmazkodva a Belgiumban használatos modellhez képest a magyarországi *MIDAS_HU modell* fejlesztése érdekében számos módszertani változtatásra volt és van szükség. Míg Belgiumban a viselkedési egyenletek becsléséhez csak a 2002. és az azt megelőző évi foglalkoztatásra vonatkozó információkat használják fel, mivel ezekről az évekről rendelkeznek teljeskörű adatokkal, addig Magyarországon egy ennél jóval átfogóbb, közel fél évszázadot felölelő adatfájl szolgál a modell becslésének alapjául. A lehető legtöbb ismert információ felhasználásával nyerhető előnyök mellett erre már csak azért is szükség van, mert a hazai munkaerőpiac az 1990 körüli rendszerváltozást követően jelentős *strukturális átalakuláson* ment át, melynek következtében egy közel egymillió létszámú csoport hirtelen állástalanná vált, legnagyobb részük pedig végül a korhatár alatti öregségi nyugellátás valamely formáját, vagy rokkantsági nyugellátást vett igénybe. A hazai munkaerőpiac kínálati oldala belga megfelelőjénél kevésbé tekinthető tehát születési évek szerint homogén sokaságnak. Jelen tanulmány a MIDAS_HU modellbeli munkaerő-piaci aktivitást leíró *munkaerő-piaci modul* viselkedési egyenleteinek becslését mutatja be.

Az elmúlt évtizedekben rengeteg tanulmány született a hazai munkaerő-piacról kapcsolatban, melyekből a számunkra legfontosabbakat – kimondottan a teljesség igénye nélkül – igyekszünk összefoglalni. *Antal–Telegdy* [2003] makroökonómiai modell segítségével jelzi előre a foglalkoztatottság szerkezetének alakulását 2020-ig foglalkozási csoportok, nem és iskolai végzettség szerinti bontásban, a múltbeli trendek jövőbeli kivetítésével. A munkaerőpiac és a felsőoktatás kapcsolatát elemzi *Galasi–Varga* [2005].

A szerzők felhívják a figyelmet az iskolázottság és ezzel párhuzamosan a diplomások iránti munkaerő–kereslet jelentős emelkedésére, és a felsőfokú végzettségűek relatív foglalkoztatási és kereseti előnyének fokozatos csökkenésére. Szintén az iskolázottság és a foglalkoztatottság viszonyát elemzi *Kertesi–Varga* [2005] és *Kertesi–Köllő* [2006]. Az utóbbi tanulmány a diplomák piaci értékének 1995 és 2004 közötti változását a Foglalkoztatási Hivatal és a KSH mintavételes adatai alapján vizsgálva megállapítja, hogy bár a diplomák piaci értékének korábbi rendkívül gyors növekedése 2000 után lelassult, az adatok mégsem igazolják a diplomás munkanélküliséggel és a diplomák elértéktelenedésével kapcsolatos közkeletű aggodalmakat. *Augusztinovics* [2005] amellet érvel, hogy a nyugdíjrendszerek jövőjét fenyegető legnagyobb veszély valójában nem az öregedés, hanem a foglalkoztatottság kedvezőtlen alakulása. *Augusztinovics–Köllő* [2007] hosszmetzeti szemléletben, évjárat, nem és iskolai végzettség szerinti bontásban vizsgálják a keresőképes életpályaszakasz szerepét a nyugdíjas szakasz alakulásában, valamint felhívják a figyelmet arra, hogy bár 1993–2004 között több mint nyolc százalékponttal emelkedett a foglalkoztatási ráta, ez valójában az iskolázottság szerinti összetétel javulásának köszönhető, hiszen iskolázottsági csoportok szerinti bontásban a mutató emelkedése marginális volt. Tanulmányuk a járulékfizetéssel járó foglalkoztatás sűrűsége szerinti csökkenő sorrendben három csoportra osztja a népességet: ún. *alfákra*, *bétákra* és *gammákra*. *Augusztinovics–Gyombolai–Máté* [2008] adminisztratív adatok alapján, több mint hatmillió egyén tízévi járulékfizetési adatainak felhasználásával elemzik a járulékfizetést és a jövőbeli nyugdíjjogosultságokat, és felhívják a figyelmet az alfa, béta és gamma munkaerő-piaci státuszok közötti alacsony mobilitásra, valamint a jelen tanulmányhoz hasonló, hosszabb adminisztratív adatsorokon alapuló elemzéseket szorgalmazznak a jövőben. *Augusztinovics–Köllő* [2008] szintén a foglalkoztatottság és a nyugdíjak kapcsolatát elemzik. A Nyugdíj– és Időskori Kerekasztal kezdeményezésére a KSH és az ONYF együttműködésével létrehozott munkaerő-piaci adatbázist elemzi részletesen *Bálint–Köllő–Molnár* [2010], megvizsgálva a hosszú távú nyugdíjjogszerzést befolyásoló legfontosabb tényezőket. A szerzők kimutatják, hogy a jogszerző időszak hossza leginkább az iskolázottsággal és kisebb mértékben a területi elhelyezkedéssel és a településtípussal magyarázható, ugyanakkor a férfiak és nők között ceteris paribus csekély különbséget mutatnak ki. A makroszintű rövidtávú munkaerő-piaci kilátásokat elemzi *Czibik et al.* [2013]. A szerzők alapvetően pozitív, reményt keltő képet festenek a munkaerő-piac közeljövőbeli alakulásáról. A hazai munkaerőpiacon tapasztalható friss tendenciákat mutatja be *Bakó* [2013] és *Bakó* [2014].

A rendelkezésre álló adatok, adatkorrekció és szegmentálás

A munkaerő-piaci modul becslését csak a még élő, öregségi nyugdíjban még nem részesülő, ugyanakkor gazdasági szempontból már aktív korú népességen végeztük el. A becsléshez felhasznált anonim adatbázis évenkénti bontásban többek között a következő adatokat tartalmazza:

- születési dátumok,
- nemek,
- foglalkozási főcsoportok (FEOR=Foglalkozások Egységes Osztályozási Rendszere, 10 kategória),
- bejelentett lakóhelyhez tartozó irányítószámok,
- a nyugdíjjogosultság alapját képező ún. *osztónapok* száma (ezen belül az ún. *valódi, pszeudó* és *összes osztónapok száma*, ahol valódi = tényleges munkaviszony révén keletkezett, és pszeudó = pl. anyasági vagy egyéb nyugdíjjogosultságba beszámító ellátás révén keletkezett).

Az elemzés megkezdése előtt egyes hiányzó értékek pótlásához egy általunk kimondottan erre az adatbázisra kifejlesztett ún. *többszörös imputálási eljárást* (ld. *Carpenter-Kenward* [2013]) alkalmaztunk, melyet itt nem részletezünk, valamint eltávolítottuk azokat a rekordokat, melyekhez irányítószám-adatbázisunk alapján nem tartozott település (itt a megadott irányítószámok gyakran egyjegyű számok, máskor nem létező, vélhetően a rögzítéskor elgépeltek, esetleg időközben megváltozott értékek voltak). Ezáltal adatbázisunk minden egyes később felhasznált változóra kiterjedően teljessé vált, így a hiányzó értékek problémakörével a továbbiakban nem kellett foglalkoznunk. A szűrést követően megmaradó rekordok száma 5 394 648 volt. Ezt követően egy az utolsó évi ismert munkaerő-piaci aktivitást jelző bináris eredményváltozót képeztünk a 2012. évi, nyugdíjjogosultságba beszámító valódi osztónapok száma alapján annak lehetséges értéktartományát egy alkalmasan választott küszöbértéknél elvágva úgy, hogy a létrejövő változó számtani átlaga éppen megegyezzen a 2012. évi valódi osztónapok évesített értékének számtani átlagával.

A MIDAS modellek munkaerő-piaci egyenleteinek általános logikai felépítése a következő:

- Keresztmetszeti becslést alkalmaz bináris logisztikus regresszió felhasználásával az utolsó évi ismert adatokon, a közvetlenül megelőző évi aktivitásra vonatkozó adatot is felhasználva.

- A következő években a magyarázó változók előrevetített értéke alapján, a becsült logit értékeket egy megfelelően becsült varianciájú² véletlen hibataggal *perturbálva* egyénenként megállapítja a munkaerő-piaci aktivitás valószínűségét.
- Az egyéneket a becsült valószínűség szerint csökkenő sorrendbe rendezi, majd az exogén makroelőjelzések alapján becsült számú aktív egyént a rangsor elejéről kiválasztva határozza meg az adott évben aktív dolgozók csoportját (ún. *alignment*).

Kihívást jelentett számunkra, hogy a közel fél évszázadnyi adattömeg felhasználását, melyet a rendszerváltozás környékén bekövetkezett, korábban már említett strukturális törés is kívánatosá tesz, hogyan tudjuk összeegyeztetni a MIDAS modell fent felvázolt keresztmetszeti logikájával. Az életkorral, nemmel, foglalkozással és lakóhellyel kapcsolatos információk közül részben ezen adatok természeténél fogva, részben pedig a MIDAS modellek logikájához igazodva elegendőnek látszott a legutóbbi ismert értékek felhasználása. A historikus foglalkoztatási adatok ezzel szemben az adott egyén jövőbeli foglalkoztatási esélyeire nézve vélhetően rengeteg többletinformációt nyújtanak, így szinte tálcán kínálták magukat a munkaerő-piaci modulbeli hasznosításra. Ennek legmegfelelőbb módszertani implementációja azonban az ezen adatok hosszsmetszeti és a MIDAS modell viselkedési egyenleteinek keretsmetszeti jellege között feszülő ellentét miatt korántsem volt nyilvánvaló. Végül úgy döntöttünk, hogy a munkaerőpiacot a historikus foglalkoztatási adatok alapján *szegmentáljuk*, és a foglalkoztatási szegmenshez való tartozást használjuk fel prediktorként a keresztmetszeti egyenletek becslése során, valamint a szegmentálást az egyének évenkénti összes beszámítható osztónapja alapján végezzük el, mely a munkaerő-piaci aktivitás mérőszámának tekinthető.

A szegmentálás gyakorlati megvalósítására két lehetséges út kínálkozott: klaszteranalízis alkalmazása vagy *szakmai szempontok alapján definiált szegmensek képzése*. A klaszteranalízis azonban nehezen kivitelezhetőnek tűnt, mivel egyénenként eltérő a munkaerőpiacon eltöltött időszak hossza, és az egyes évekhez tartozó beszámítható osztónapok számai erősen korreláltak, továbbá interpretációs problémát okozott volna az elemzés során nyert, matematikai értelemben optimális klaszterek általános nyugdíjszakmai fogalmaknak való megfeleltetése. Ezért a szakmai szempontokat figyelembe véve, explicit módon definiált historikus munkaerő-piaci szegmensek képzése mellett döntöttünk.

² A normális eloszlású hibatagok varianciáját növelve nő a modellben az aktívak és inaktívak közötti mobilitás. A varianciát úgy kalibráljuk, hogy a 2011-ről 2012-re inaktív-aktív és aktív-inaktív irányban megváltozott státuszúak összesített aránya a modellben megegyezzen a ténylegesen tapasztalt aránnyal.

A teljes felhasznált állományról mindenekelőtt leválasztottuk a tartós (pl. rokkantsági) ellátásban lévőket és a pszeudó jogviszonnal (pl. anyasági ellátással) rendelkezőket. A maradék állományon a szegmensek képzése érdekében egyénenként az egyes évekhez tartozó összes beszámítható osztónapok számát az 1970-1991, 1992-2001 és 2002-2010 időszakokban átlagoltuk, amennyiben rendelkezésünkre állt az adott egyénre az adott időszakban figyelembe vehető adat. Az időszakok határainak megállapításánál a rendszerváltás környékén bekövetkezett strukturális törést is figyelembe vettük, mely 1992-ben teljesedett ki. Az ily módon nyert időszakonkénti átlagos értékeket a létrejövő kategóriák létszámának egyenletességét is figyelembe véve a következő három kategóriára osztottuk: alacsony (0-100 nap), közepes (100,01-300 nap) és magas (300,01-366 nap). Az üres értékek előfordulásának lehetőségét is figyelembe véve így elvben legfeljebb $4^3 = 64$ egyszerűsített *historikus foglalkoztatási pálya* keletkezett, melyeket szakmai szempontok alapján három *főszegmensbe* soroltuk: magasan, közepesen és alacsonyan foglalkoztatottak. Az összes lehetséges historikus foglalkoztatási pálya besorolását nem közöljük részletesen, de illusztratív céllal néhányat bemutatunk az 1. táblázatban:

1970-1991	1992-2001	2002-2011	Főszegmens
–	–	alacsony	alacsony
közepes	alacsony	alacsony	alacsony
–	közepes	közepes	közepes
–	magas	közepes	közepes
magas	közepes	magas	magas*
alacsony	magas	magas	magas*

1. táblázat: Példák egyes historikus foglalkoztatási pályák főszegmensekbe sorolására

A tisztán alacsony, közepes és magas pályák besorolása egyértelmű. A többi pálya besorolása egyfajta átlagos aktivitást tükröz, ahol az átlagolás során kiemelt hangsúlyt fektettünk a legfrissebb, legutolsó pályaszakaszra, abból kiindulva, hogy a közelmúltbeli aktivitás határozza meg legerősebben az egyén jövőbeli munkaerő-piaci részvételét. Az „alacsony” besorolást jellemzően a folyamatosan alacsony átlagos aktivitású vagy komolyabb státuszvesztéséget szenvedett egyének kapták. Az utóbbiak közt túlnyomórészt olyan egyéneket találunk, akik a 1991-'92 körüli nagy strukturális átalakulást követően kerültek ki a munkaerőpiac centrumából. A „magas” besorolású pályák minősítését utólag korrigáltuk: amennyiben 2002-2011 között előfordult olyan év, amikor egy egyén 100-nál kevesebb összes osztónappal rendelkezett, akkor besorolását „közepes”-re módosítottuk. Ennek oka az volt, hogy „magas” minősítést csak olyan olyan egyéneknek kívántunk adni,

akik folyamatosan valódi vagy pszeudó jogszerző státusszal rendelkeznek. Bár az így nyert besorolások hasonlóságot mutatnak *Augusztinovics–Köllő* [2007] alfa, béta és gamma kategóriáival, hangsúlyozottan nem azonosak azokkal.

Az alacsonyan, közepesen és magasan foglalkoztatottak az ellátottakkal és a speciális (pszeudó) jogviszonnal rendelkezőkkel együtt a modellben összesen öt főszegmenst képeznek. Az adott évbeli munkaerő-piaci aktivitást előzetes numerikus vizsgálataink alapján erőteljesen befolyásolja az azt közvetlenül megelőző évi aktivitás, mely a MIDAS_BE modell hasonló moduljában is a munkaerő-piaci részvétel legfontosabb prediktora. Az utolsó évi ismert aktivitás szerepét kihangsúlyozandó a 2011. évi valódi osztónapok számát a 2012. évi eredményváltozó képzésénél alkalmazott eljárással megegyező módon, átlagot megőrző *dichotóm vágással* binárisá alakítottuk, majd az öt főszegmenst a 2011. évi bináris foglalkoztatási változó értéke szerinti dichotóm felosztással végül a következő tíz *munkaerő-piaci szegmensre* bontottuk:

- ellátott, 2011-ben inaktív,
- ellátott, 2011-ben aktív,
- alacsonyan foglalkoztatott, 2011-ben inaktív,
- alacsonyan foglalkoztatott, 2011-ben aktív,
- speciális jogviszonyú, 2011-ben inaktív,
- speciális jogviszonyú, 2011-ben aktív,
- közepesen foglalkoztatott, 2011-ben inaktív,
- közepesen foglalkoztatott, 2011-ben aktív,
- magasan foglalkoztatott, 2011-ben inaktív,
- magasan foglalkoztatott, 2011-ben aktív.

Modellillesztés, modellszelekció

A jövőbeli munkaerő-piaci aktivitás becslése érdekében adatainkra *bináris logisztikus regressziós modellt* (ld. *Kleinbaum–Klein* [2010] és *Kovács–Szüle–Fliszár–Vékás* [2011]) illesztettünk, amelyben a 2012. évi munkaerő-piaci aktivitás valószínűségét a rendelkezésre álló potenciális magyarázó változók segítségével becsültük. Adatfájlunkat a módszertani korrektség érdekében egyszerű véletlen kiválasztás segítségével két diszjunkt részsokaságra: az egyének 90%-át tartalmazó *tanuló* és a maradék 10%-ot lefedő *tesztelő* állományra osztottuk (*osztott mintás validáció*, ld. pl. *Picard-Cook* [1984]). A modellek becslését minden esetben a *tanuló*, az illeszkedés vizsgálatát és a modellszelekciót pedig minden esetben a *tesztelő* állományon végeztük. Az óriási mintaméretből és a kiválasztás egyszerű véletlen mivoltából adódóan a létrejött két részsokaság minden kategorikus változó tekintetében reprezentatívnak tekinthető. Ezt numerikus vizsgálataink is igazolták.

Az általunk a későbbiekben végsősoron kiválasztott modellspecifikáció helyességének alátámasztására öt különböző, növekvő bonyolultságú modellt becsültünk, a modellek által becsült valószínűségeket elmentettük, és a *tesztelő* halmazra vonatkozó illeszkedési mutatókat mindegyik modell esetén külön-külön kiszámítottuk, majd a legegyszerűbb modellt leszámítva minden egyes modell esetén *likelihood-arány teszt* segítségével formálisan teszteltük, hogy az adott modell a hozzá képest egy szinttel egyszerűbb modellhez képest szignifikánsan nagyobb magyarázó erővel bír-e.

A vizsgált modellspecifikációk a következők voltak:

- 0. modell (egyetlen egyenlet): csak konstans.
- 1. modell (egyetlen egyenlet): 2011-ben aktív volt-e az adott egyén (2 kategória), konstans.
- 2. modell (egyetlen egyenlet): munkaerő-piaci szegmens (10 kategória), konstans.
- 3. modell (munkaerő-piaci szegmensenként 1-1, összesen 10 egyenlet, mely azonban formálisan egyetlen egyenletként is felfogható): nem (2 kategória), FEOR főcsoport (10 kategória), statisztikai régió (NUTS II, 7 kategória), településtípus (4 kategória), életkor és annak négyzete (folytonos változóként), konstans.
- 4. modell (munkaerő-piaci szegmensenként 1-1, összesen 10 egyenlet, mely azonban formálisan egyetlen egyenletként is felfogható): a 3. modellben szereplő valamennyi magyarázó változó, valamint az összes bináris kategóriaindikátor-változó és az életkor minden egyes lehetséges páronkénti keresztszorzata (*interakciója*).

A 3-4. modellspecifikációk felírásához azért választottuk a munkaerő-piaci szegmensek szerinti bontást, mert korábbi vizsgálataink igazolták, hogy ez a potenciális magyarázó változó minden egyéb szóba jöhető magyarázó változónál jóval jelentősebb magyarázó erővel bír. A változók beléptetését a 0-2. modellek esetén *Enter*, az ezeknél lényegesen több paramétert tartalmazó 3. modell esetén a magyarázó változók egyenkénti szignifikanciáját tesztelő *Wald-statisztikára épülő Forward*, a 4. modell esetén pedig az ilyen komplexitású modellek esetén fokozottan jelentkező számítási korlátokból adódóan ismét a számítási szempontból egyszerűbb *Enter* eljárással végeztük (ld. Kovács–Szüle–Fliszár–Vékás [2011]). Az egyénekhez tartozó *studentizált reziduumok* szignifikanciáját a sztenderd normális eloszlás alapján számított, a szimultán tesztelésnek megfelelően kiigazított p-értékek alapján vizsgáltuk. A kiigazítást a *Bonferroni*- és a kevésbé konzervatív *Benjamini-Hochberg eljárások* (*Benjamini-Hochberg* [1995]) alapján külön-külön egyaránt elvégeztük. Végül ezen eljárások egyik modellspecifikáció esetén sem jeleztek egyetlen szignifikáns outliert sem, mely vélhetően az óriási mintaméretnek köszönhető, hiszen egy többmillió rekordból álló adattömeg esetén nehezen elképzelhető, hogy egyetlen egyén adatai képesek lennének a becsült függvényegyütthatókat szignifikánsan megváltoztatni.

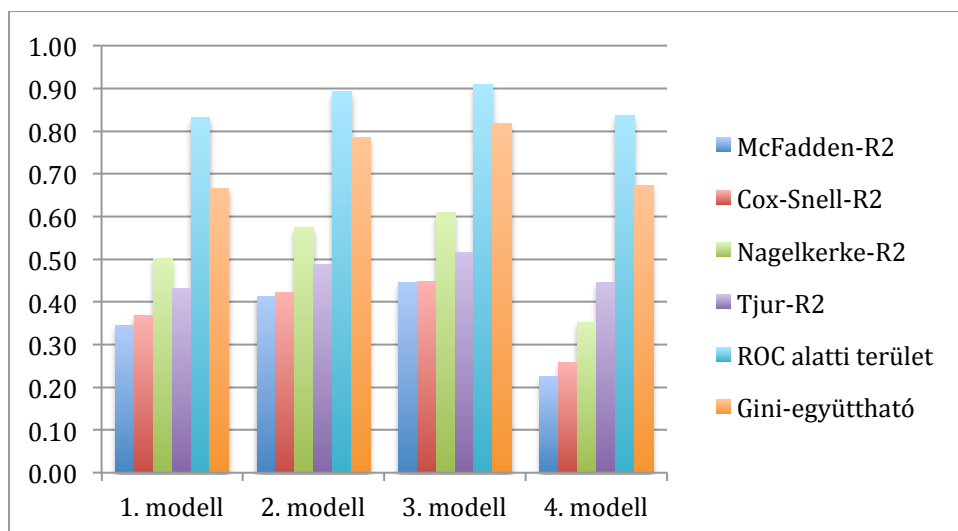
Az egyes modellekhez tartozó, a tesztelő állományon számított log-likelihood értékek és illeszkedési mutatók (McFadden-, Cox-Snell-, Nagelkerke- és Tjur-féle pszeudó-R² értékek, ROC=Receiving Operator Characteristic görbe alatti terület és Gini-együttható) illetve az egymásba ágyazott modellek közötti választást lehetővé tevő likelihood-arány tesztek p-értékei a 2. és 3. táblázatokban és az 1. ábrán láthatók.

Modell	Log-likelihood	McFadden-R ²	Cox-Snell-R ²	Nagelkerke-R ²
0. modell	- 358 640	0,0%	0,000	0,0%
1. modell	- 234 614	34,6%	0,369	50,1%
2. modell	- 210 400	41,3%	0,423	57,5%
3. modell	- 198 727	44,6%	0,447	60,8%
4. modell	- 277 876	22,5%	0,259	35,2%

2. táblázat: Az egymásba ágyazott modellek illeszkedési mutatói I.

Modell	Tjur-R ²	ROC alatti terület	Gini-együttható	LR-teszt p-értéke
0. modell	0,0%	0,500	0,0%	0,0%
1. modell	43,2%	0,833	66,6%	0,0%
2. modell	48,7%	0,893	78,5%	0,0%
3. modell	51,6%	0,908	81,7%	0,0%
4. modell	44,6%	0,837	67,3%	100,0%

3. táblázat: Az egymásba ágyazott modellek illeszkedési mutatói II.



1. ábra: Az egymásba ágyazott modellek illeszkedési mutatói

Eredményeink meggyőzően alátámasztják a 3. modell használatát: a többi modell minden illeszkedési mutató tekintetében gyengébbnek bizonyul, és az egymásba ágyazott likelihood-arány tesztek alapján is arra lehet következtetni, hogy a 3. modellig bezárólag eredményes a modell komplexitásának bővítése, azt követően azonban az interakciós tagok beléptetése kifejezetten károsnak tekinthető. Ez feltehetően a következő okokra vezethető vissza:

- nagy számú paramétert tartalmazó modellek esetén fellép a *túlillesztés* jelensége: addicionális magyarázó változók beléptetése ugyan elvben minden esetben javítja az illeszkedést a tanuló halmazon, a jövőbeli alkalmazás szempontjából relevánsabb tesztelő halmazon azonban már ronthatja és előbb-utóbb szükségszerűen rontja azt (amodell az érvényes statisztikai tendenciákon túl a tanuló halmazon fellelhető véletlen zajt reprodukálja, amely már nem általánosítható a tesztelő halmazra),
- a nagy számú magyarázó változó szerepeltetése *multikollinearitással* összefüggő numerikus nehézségeket okoz, a Hesse-mátrix közel szingulárisává válik, és a paraméterbecslő algoritmus gyakran szuboptimális megoldásra jut,
- a statisztikai elmélet és gyakorlat a *parszimónia elve* alapján előnyben részesíti az ugyanazon jelenségeket kevesebb paraméter segítségével leíró modelleket.

Eredményeink alapján összességében elmondható, hogy már a kizárólag a 2011. évi munkaerő-piaci aktivitást figyelembe vevő 1. modell is kifejezetten kedvező illeszkedést mutat, alátámasztva az öt főszegmens tíz szegmensre történő dichotóm felbontásának helyességét, melyet jelentősen fokoz a munkaerő-piaci szegmensek figyelembe vétele, és végül szignifikánsan javít az egyéb prediktorok beléptetése.

Klasszifikáció és a becült munkaerő-piaci egyenletek

A 2012. évi aktivitást modellünk a becült valószínűségekből képzett rangsor alapján, a rangsor elejéről megfelelő számú egyént kiválasztva úgy jelzi előre, hogy munkaerő-piaci szegmensenként az aktívnak besorolt egyének száma megegyezzen a ténylegesen aktív egyének számával. Az előző szakaszban végsősoron kiválasztott 3. modell besorolási pontosságát a tesztelő állományon vizsgáltuk. A 4. táblázat az itt tapasztalt *találati arányokat* foglalja össze tényleges aktivitás szerint bontva és összesen, és a tényleges találati arányok mellett referenciaként a véletlen besorolás segítségével elérhető arányokat is tartalmazza:

		Becült aktivitás			Találati pontosság	
		Nem aktív	Aktív	Összesen	Tényleges	Véletlen
Tényleges aktivitás	Nem aktív	152 222	53 921	206 143	73,8%	38,2%
	Aktív	53 921	279 079	333 000	83,8%	61,8%
	Összesen	206 143	333 000	539 143	80,0%	52,8%

4. táblázat: Találati arányok a tesztelő állományban tényleges aktivitás szerint és összesen

A 4. táblázat tanúsága szerint modellünk a nem aktív egyének 73,8%-át, az aktívak 83,8%-át és a teljes állomány 80%-át helyesen ismeri fel, szemben a véletlen besorolás segítségével elérhető 38,2, 61,8 és 52,8%-kal, ahol véletlen besorolás alatt azt értjük, hogy az egyéneket egyénekenként azonos valószínűséggel a ténylegesen tapasztalt számú aktív és nem aktív egyénekre osztjuk fel. Megállapítható, hogy modellünk a véletlenszerű besoroláshoz jóval megbízhatóbb: 2011-ről 2012-re a tesztelő állományban ötből négy egyént helyesen sorolt be.

A tíz becült egyenletet itt nem kívánjuk részletesen közzétenni, csak a legnépesebb, magasan foglalkoztatott, 2011-ben aktív szegmens egyenletének együtthatóit, az azokhoz tartozó Wald-tesztek p-értékeit és az esélyhányadosokat mutatjuk be illusztratív céllal:³

Változó	Együttható	Wald p-érték	Esélyhányados
nő	-0,309	0,0%	0,734
gazdasági, igazgatási, érdekképviseleti vezetők, törvényhozók	0,239	0,0%	1,270
felsőfokú képzettség önálló alkalmazását igénylő foglalkozások	0,547	0,0%	1,728
egyéb felsőfokú vagy középfokú képzettséget igénylő foglalkozások	0,289	0,0%	1,336
irodai és ügyviteli (üggyfélkapcsolati) foglalkozások	0,147	0,0%	1,158
kereskedelmi és szolgáltatási foglalkozások	-0,243	0,0%	0,785
mezőgazdasági és erdőgazdálkodási foglalkozások	-0,077	14,7%	0,926
ipari és építőipari foglalkozások	-0,111	0,0%	0,895
gépkézeltők, összeszerelők, járművezetők	-0,021	42,6%	0,979
szakképzettséget nem igénylő (egyszerű) foglalkozások	-0,366	0,0%	0,693

³ Referenciakategóriának a közép-magyarországi, egyéb (fel nem sorolt) foglalkozású férfiakat tekintettük.

Dél-Alföld	0,177	0,0%	1,194
Dél-Dunántúl	0,003	89,2%	1,003
Észak-Alföld	0,109	0,0%	1,115
Észak-Magyarország	0,053	0,5%	1,055
Közép-Dunántúl	0,188	0,0%	1,207
Nyugat-Dunántúl	0,271	0,0%	1,312
megyei központ vagy megyei jogú város	-0,005	80,0%	0,995
egyéb város	0,109	0,0%	1,115
község vagy nagyközség	0,069	0,0%	1,071
életkor	0,159	0,0%	1,172
életkor ²	-0,001	0,0%	0,999
konstans	-0,693	0,0%	0,500

5. táblázat: A magasan foglalkoztatott, 2011-ben aktív munkaerő-piaci szegmens egyenlete

Az 5. táblázat eredményei közül kiemelendő, hogy a vizsgált szegmensben nők esetén ceteris paribus a 2012. évi munkaerő-piaci aktivitás esélye 0,734-szerese a férfiak esetén érvényes esélynek, vagyis – az egyéb magyarázó változók hatását kontrollálva is – igen jelentős különbséget találtunk a férfiak és nők aktivitása között. Ugyanitt ceteris paribus felsőfokú képzettség önálló alkalmazását igénylő foglalkozások esetén 1,728-szerese, szakképzettséget nem igénylő (egyszerű) foglalkozások esetén pedig 0,693-szorosa az aktivitás esélye az egyéb (fel nem sorolt) foglalkozások esetén érvényes esélynek, valamint az egyes FEOR főcsoportok esélyhányadosai alapján általában is igen jelentős foglalkoztatási különbségeket találtunk a képzettebb és kevésbé képzett munkaerőt igénylő foglalkozások között, az előbbieket javára. Látszólag meglepő, hogy a fővároshoz és Közép-Magyarországhoz képest ceteris paribus szinte bármely településtípus és NUTS II régió esetén magasabb az aktivitás esélyhányadosa, mely azonban a fővárosra egyéb változók szerinti kedvezőbb összetételének tulajdonítható, mivel adataink tanúsága szerint Budapesten és Közép-Magyarországon magasabb az átlagos aktivitás szintje, mint az ország egyéb részein. A vizsgált szegmensben a 2012. évi munkaerő-piaci aktivitás logitja az életkor függvényében ceteris paribus konkáv parabolával modellezhető, mely a maximumát hozzávetőleg 48 éves korban veszi fel, összhangban azzal az általánosan megfigyelhető tapasztalattal, hogy a pályakezdést követően növekszik, a nyugdíjhoz közeledve pedig csökken az egyének munkaerő-piaci aktivitása. A nemre, életkorra és foglalkozási főcsoportokra vonatkozó eredmények valamennyi szegmensben túlnyomórészt megfelelnek a szakmai elvárásoknak illetve ahol ezekhez képest eltéréseket tapasztalunk, ott a különbségek jellemzően statisztikailag nem szignifikánsak.

Összefoglalás

Tanulmányunkban az egyéni munkaerő-piaci aktivitás becslését mutattuk be az egyének múltbeli munkaerő-piaci aktivitására, születési évére, nemére, lakóhelyére és foglalkozására vonatkozó információk felhasználásával, statisztikai viselkedési egyenletek segítségével. Az egyenletek becsléséhez a hiányzó adatok pótlását és az egyének szegmentálását követően bináris logisztikus regressziót alkalmaztunk, modellünket pedig körültekintően validáltuk. A modell illeszkedése megfelelőnek mondható: a tesztelő állományon 80% körüli találati pontossággal rendelkezik.

A felhasznált csoportképző és magyarázó változók értékeit előreszámítva a becslt egyenletek alapján, a tanulmányban ismertetett alignment eljárás figyelembe vételével egyénenként megbecsülhető a jövőbeli évekre vonatkozó munkaerő-piaci aktivitás, mely a MIDAS_HU modell egyéb moduljaival és almoduljaival kölcsönhatásban együttesen előrevetíti a nyugdíjassza jövőbeli bevételeinek és kiadásainak egyénenkénti alakulását.

Jelen tanulmány a szakirodalomban fellelhető korábbi, magyarországi munkaerő-piaccal kapcsolatos elemzések közül kiemelkedik abból a szempontból, hogy azoknál jóval szélesebb körű, egyesített adminisztratív adatbázisok felhasználásával készült. A MIDAS_BE modellhez képest is jelentős előrelépés, hogy két év adatai helyett több mint negyven év információit használtuk fel az egyenletek becslése során, így lehetőségünk nyílt a hazai munkaerő-piacon az 1990-es évek elején lezajlott strukturális átalakulás hatásainak figyelembe vételére is.

A becslt munkaerő-piaci egyenletek a MIDAS_HU modell keretein belül a jövőben lehetővé teszik a munkaerő-piaci aktivitás minden korábbinál pontosabb, egyéni szintű előrejelzését, melyet az Országos Nyugdíjbiztosítási Főigazgatóság és a nyugdíjszakma a nyugdíjrendszerrel kapcsolatos szakmai elemzések és hatástanulmányok készítése során reményeink szerint sikerrel fog tudni felhasználni a jövőben.

Hivatkozások

Antal, G., – Telegdy, Á. (2003). A foglalkoztatási szerkezet előrejelzése foglalkozási csoportok, nem és iskolai végzettség szerint. In: Fazekas, K. – Varga, J. (szerk.) *Trendek és előrejelzések. Munkaerő-piaci prognózisok készítése, szerkezetváltás a munkaerőpiacon*. MTA KRTK Közgazdaság–Tudományi Intézet, Budapest.

Augusztinovics, M. (2005): Népeség, foglalkoztatottság, nyugdíj. *Közgazdasági Szemle*, LII. évf., 5. sz., pp. 429–447.

Augusztinovics, M. – Gyombolai, M. – Máté, L. (2008). Járulékfizetés és nyugdíjjogosultság 1997–2006. *Közgazdasági Szemle*, 55. évf. 7–8. sz., pp. 665– 689.

Augusztinovics, M. – Köllő, J. (2007). Munkapiaci pálya és nyugdíj, 1970–2020. *Közgazdasági Szemle*, LIV. évf., 2007. június, pp. 529–559.

Augusztinovics, M. – Köllő, J. (2008). Decreased Employment and Pensions. In: Holzmann, R., L. Mackellar and J. Repansek (szerk.) *Pension Reform in South–Eastern Europe: Linking to Labor and Financial Market Reforms*. The World Bank, Washington D.C.

Bakó, T. (2013). A magyarországi munkapiac 2012–2013–ban. In: Fazekas, K. – Neumann, L. (szerk.): *Munkaerőpiaci Tükör 2013*. MTA Közgazdaság– és Regionális Tudományi Kutatóközpont, Közgazdaságtudományi Kutatóintézet, Országos Foglalkoztatási Nonprofit Kft., Budapest.

Bakó, T. (2014). A magyarországi munkapiac 2013–2014–ben. In: Fazekas, K. – Neumann, L. (szerk.): *Munkaerőpiaci Tükör 2014*. MTA Közgazdaság– és Regionális Tudományi Kutatóközpont, Közgazdaságtudományi Kutatóintézet, Országos Foglalkoztatási Nonprofit Kft., Budapest.

Bálint, M. – Köllő, J. – Molnár, Gy. (2010). Nyugdíj megszerzés és teljes aktív életpálya. *Statisztikai Szemle*, 88(6).

Benjamini, Y. – Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society, Series B*, 57(1): pp. 289–300.

Carpenter, J. – Kenward, M. (2013). *Multiple Imputation and its Applications*. John Wiley and Sons, New York.

Czibik, Á. – Makó, Á. – Máté, F. – Türei, G. – Várhalmi, Z. – Zichy, F. (2013). *Rövidtávú munkaerő-piaci prognózis 2014*. MKIK Gazdaság– és Vállalkozáskutató Intézet, Budapest.

- Dekkers, G. (2010). *The long-term adequacy of the Belgian public pension system: An analysis based on the MIDAS model* (working paper). Federaal Planbureau, Brussels.
- Dekkers, G. (2013). *An introduction to MIDAS_BE, the dynamic microsimulation model for Belgium* (working paper). Centre for Sociological Research, Brussels.
- Dekkers, G. – Belloni, M. (2009). *Micro simulation, pension adequacy and the dynamic model MIDAS: an introduction* (working paper). Federaal Planbureau, Brussels.
- Galasi, P. – Varga, J. (2005). *Munkaerőpiac és Oktatás*. MTA Közgazdaságtudományi Intézet, Budapest.
- Holtzer, P. (szerk., 2010). *Jelentés a Nyugdíj és Időskor Kerekasztal tevékenységéről*. Miniszterelnöki Hivatal, Budapest.
- Kertesi, G. – Köllő, J. (2006). Felsőoktatási expanzió, „diplomás munkanélküliség” és a diplomák piaci értéke. *Közgazdasági Szemle*, LIII. évf., 2006. március, pp. 201–225.
- Kertesi, G. – Varga, J. (2005). Foglalkoztatás és iskolázottság Magyarországon. *Közgazdasági Szemle*, LII. évf., 7–8. sz., pp. 633–662.
- 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.
- Li, J. (2011). *Dynamic Microsimulation for Public Policy Analysis*. Boekenplan Maastricht, Maastricht.
- Li, J. – O’Donoghue, C. (2013). A survey of dynamic microsimulation models: uses, model structure and methodology. *International Journal of Microsimulation*, 6(2): pp. 3-55.
- Picard, R. – Cook, D. (1984). Cross-Validation of Regression Models. *Journal of the American Statistical Association*, 79(387): pp. 575–583.
- Szabó, S. Cs. K. (2000): Nyugdíjrendszerünk 1929-től 1997-ig. In. Augusztinovics, M. (szerk.): *Körkép reform után. Tanulmányok a nyugdíjrendszerről*. Közgazdasági Szemle Alapítvány, Budapest.
- Zaidi, A. – Rake, K. (2001). Dynamic Microsimulation Models: A Review and Some Lessons for SAGE (p. 40). *SAGE Discussion Papers*. URL: www.lse.ac.uk/depts/sage, letöltés dátuma: 2015.05.14.