

KÖZGAZDASÁGI SZEMLE, LXII. ÉVF., 2015. DECEMBER (1291–1308. o.)

VÉKÁS PÉTER

# Az egyéni munkaerő-piaci aktivitás becslése mikroszimulációs modellkeretben

A tanulmány a magyarországi egyéni munkaerő-piaci aktivitás becslésével foglalkozik, amely a MIDAS-HU nyugdíj-mikroszimulációs modell munkaerő-piaci moduljának kulcsfontosságú építőeleme. Az elemzés érdekében többváltozós statisztikai módszereket alkalmaztunk egy óriás méretű, adminisztratív adatokat tartalmazó adathalmazon. A felhasznált adatok több mint négy évtized teljes körű demográfiai, foglalkoztatottsági és járulékfizetési adatait tartalmazzák 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 munkaerőpiacon – elsősorban az 1990-es évek elején – lezajlott jelentős strukturális átalakulások indokolták.\*  
Journal of Economic Literature (JEL) kód: H55, C15, C40, C53.

## Előzmények

A hazai nagy állami ellátórendszerek legjelentősebbike a minden munkavállalóra kötelezően kiterjedő állami öregségi, rokkantsági és hátramaradotti nyugdíjrendszer, amelyet első, nem teljes körű 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 fenntarthatósága erőteljesen függ a népesség

\* Köszönet illeti Rézmovits Ádámot és Tóth Krisztiánt (Országos Nyugdíjbiztosítási Főigazgatóság) a jelen cikk elkészültéhez nélkülözhetetlen ötleteikért, szakmai segítségükért és tanácsaikért, valamint a névtelen lektort a tanulmány bírálatában szereplő számos előremutató javaslatáért.

Vékás Péter, BCE Operációkutatás és Aktuáriustudományok Tanszék, MTA–BCE Lendület Stratégiai Interakciók Kutatócsoport (e-mail: [peter.vekas@uni-corvinus.hu](mailto:peter.vekas@uni-corvinus.hu)).  
A kézirat első változata 2015. július 22-én érkezett szerkesztőségünkbe.  
DOI: <http://dx.doi.org/10.18414/KSZ.2015.12.1291>

kor szerinti megoszlásától, a keresőképes korúak munkaerő-piaci aktivitásától és a jövedelmek eloszlásától. Volumenének megfelelően a nyugdíjassza pénz-áramlásának alakulása a költségvetési, társadalmi-gazdasági és általános nemzetgazdasági folyamatokra egyaránt jelentős befolyást gyakorol. A jól megalapozott költségvetési, gazdaságpolitikai és szociálpolitikai döntéshozatal érdekében tehát kívánatos és elkerülhetetlen 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 kohorszmodelleket kiegészítve az ezredforduló óta eltelt években teret nyertek az úgynevezett sztochasztikus-dinamikus mikroszimulációs modellek (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 előre vetíteni, és az aggregátumokon belül végbemenő összetétel-változások más változókra gyakorolt hatását is statisztikai szempontból adekvát módon képesek kezelni. Segítségükkel elemezhető például olyan jövedelem-újraelosztási hatások, amelyeket a makromodellek nem képesek megragadni. E modellek az egyéni döntéshozók viselkedését statisztikai-ökonometriai eljárások felhasználásával becsült *viselkedési egyenletekkel, intertemporális várható hasznossági függvények* maximalizálásával vagy e két módszerrel kombinálva is leírhatják.

Magyarországon a Nemzetgazdasági Minisztérium (NGM) és elődintézményei, valamint a Magyar Nemzeti Bank (MNB) által korábban alkalmazott makromodelleket kiegészítve először a Nyugdíj és Időskor Keresasztal (NYIKA) hozott döntést egy átfogó mikroszimulációs modellkeret és azon belül több mikroszimulációs nyugdíjmodell létrehozásáról (Holtzer [2010]). Az első, részleges modelleket követően az Országos Nyugdíjbiztosítási Főigazgatóság (ONYF) – az Európai Unió támogatásával – 2012-ben látott hozzá a Belgiumban alkalmazott MIDAS\_BE (*Microsimulation for the Development of Adequacy and Sustainability*) (leírását lásd Dekkers–Belloni [2009], Dekkers [2010], [2013]) nagyméretű sztochasztikus-dinamikus mikroszimulációs modell hazai adaptációjához. Ehhez közel fél évszázadra kiterjedő, nagy részletességű, egyéni szintű adatokat tartalmazó informatikai adatbázist használt fel.

A MIDAS modell számos egymással kölcsönhatásban álló modulból és almodulból (például 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, amelyet az ONYF is alkalmaz. A MIDAS modell az egyéb előrejelzésekkel való konzisztencia érdekében számos ponton hazai és uniós intézmények aggregált előrejelzéseire támaszkodik.

történő *exogén kiigazításokat* tartalmaz. A MIDAS modell munkaerő-piaci egyenleteinek általános felépítése a következő:

- keresztmetszeti becslést alkalmaz bináris logisztikus regresszió felhasználásával az utolsó évi ismert adatokon, a közvetlenül megelőző évi aktivitásra vonatkozó adatot is felhasználva;
- az elkövetkező években a magyarázó változók előrevetített értéke alapján, a becsült logitértékeket egy megfelelően becsült varianciájú<sup>1</sup> véletlen hibataggal perturbálva, egyénenként megállapítja a munkaerő-piaci aktivitás valószínűségét;
- az egyéneket a becsült valószínűség szerint csökkenő sorrendbe rendezi, majd az exogén makro-előrejelzések alapján becsült számú aktív egyént a rangsor elejéről kiválasztva határozza meg az adott évben aktív dolgozók csoportját (úgynevezett kiigazítás).

A gazdasági környezethez és a rendelkezésre álló adatok struktúrájához alkalmazkodva a Belgiumban használatos modellhez képest a magyarországi MIDAS 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 teljes körű 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 magyarországi munkaerőpiac az 1990 körüli rendszerváltozást követően jelentős *strukturális átalakuláson* ment át, aminek 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. Ezért a magyarországi munkaerőpiac kínálati oldala belga megfelelőjénél kevésbé tekinthető születési évek szerint homogén sokaságnak. Tanulmányunk a magyar MIDAS modellbeli munkaerő-piaci aktivitást leíró *munkaerő-piaci modul* viselkedési egyenleteinek becslését mutatja be.

## A hazai szakirodalom áttekintése

Az elmúlt évtizedekben számtalan tanulmány született a hazai munkaerő-piaci aktivitással kapcsolatban, amelyek közül néhány elemzésünk témája szempontjából is jelentős publikációt – kimondottan a teljesség igénye nélkül – igyekszünk itt összefoglalni.

*Bakó–Cseres–Gergely–Galasi* [2013] tanulmányukban az MTA KRTK KTI munkaerő-piaci előrejelző rendszerét ismertetik. Ez a gazdaságpolitikai döntés-előkészítésre is alkalmas makroszemléletű modell Magyarországon kiemelkedő abból a szempontból, hogy tizenkét különböző teljes körű adminisztratív és mintavételes adatforrás

<sup>1</sup> A normális eloszlású hibatagok varianciáját növelve, a modellben emelkedik az aktívák és inaktívák közötti mobilitás. A varianciát úgy kalibráljuk, hogy a 2011-ről 2012-re inaktív-aktív és aktív-inaktív irányban megváltozott státusúak összesített aránya a modellben megegyezzen a ténylegesen tapasztalt aránnyal.

egyesítésére épül, és a munkakeresletet és munkakínálatot, valamint a két oldal eltérését egyaránt explicit módon modellezi, továbbá egy tíz ágazatból álló többszektoros makrogazdasági előrejelzést is felhasznál.

*Antal-Telegdy* [2013] szintén 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.

*Cseres-Gergely-Hámori* [2009] az alacsony iskolázottságú, 25–54 éves férfiak iskolázottsága és foglalkoztatottsága közötti kapcsolatot elemzi az Európai Munkaerő-felvétel (*European Labour Force Survey*, EU LFS) hosszmetzeti adatain felépített regressziós modell segítségével, nemzetközi összehasonlításban.

*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 vizsgálja a keresőképes életpályaszakasz szerepét a nyugdíjas szakasz alakulásában, valamint felhívja a figyelmet arra, hogy bár 1993–2004 között jelentősen emelkedett a foglalkoztatási ráta, ez valójában az iskolázottság szerinti összetétel javulásának köszönhető, hiszen iskolázottsági csoportok szerinti bontásban a mutató emelkedése marginális volt. A tanulmány az egyéni munkaerő-piaci aktivitás szerinti csökkenő sorrendben három csoportra osztja a népességet: úgynevezett alfákra, bétákra és gammákra, amelyekhez hasonló, historikus foglalkoztatássűrűségekre vonatkozó kategóriák az itt ismertetendő modellben is megjelennek.

*Augusztinovics-Gyombolai-Máté* [2008] adminisztratív adatok alapján, több mint hatmillió egyén tízévi járulékfizetési adatainak felhasználásával elemzi a járulékfizetést és a jövőbeli nyugdíj jogosultságokat, és felhívja a figyelmet az alfa, béta és gamma munkaerő-piaci státusok 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 elemzi. A Nyugdíj és Időskor Kerekasztal kezdeményezésére a KSH és az ONYF együttműködésével létrehozott munkaerő-piaci adatbázist elemzi részletesen.

*Bálint-Köllő-Molnár* [2010] a hosszú távú nyugdíjgörszerzést befolyásoló legfontosabb tényezőket vizsgálja. 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.

*Cseres-Gergely-Scharle* [2005] az időskorú népesség alacsony aktivitását elemzi, és arra a következtetésre jut, hogy elsősorban a hazai öregségi és rokkantnyugdíj-rendszerben

kell keresni az alacsony időskorú foglalkoztatottság intézményi okait. A vizsgált jelenség az itt ismertetett modellben is statisztikailag szignifikánsan megjelenik.

Jelen cikkhez hasonlóan foglalkoztatásra vonatkozó bináris logisztikus regressziós modellt mutat be csehországi adatokon *Fialová–Mysíková* [2009], bár modelljük az itt szereplő modellel ellentétben jóval kisebb létszámú mintából becsüli az egyenleteket. A magyarázó változók hatásainak előjelei és szignifikanciái nagyon közeli egyezést mutatnak az itt bemutatandó eredményekkel.

A hazai munkaerőpiacon tapasztalható friss tendenciákat ismerteti *Bakó* [2013] és [2014], valamint (*Fazekas–Varga* [2015]).

## Felhasznált adatok, adatkorrekciós lépések

A magyar MIDAS modell munkaerő-piaci moduljának becslését az ONYF nem nyilvános, teljes körű, az 1970–2012 közötti évekre vonatkozó adminisztratív adatállományán végeztük, csak az időszak végén még élő, öregségi nyugdíjban még nem részesülő, ugyanakkor gazdasági szempontból aktív korú népességre leszűkítve az elemzést. Az elemzésben felhasznált változók köre egyénenként, évenkénti bontásban a következő volt:

- születési dátum,
- nem,
- foglalkozási főcsoport (FEOR = Foglalkozások Egységes Osztályozási Rendszere, 10 kategória),
- bejelentett lakóhelyhez tartozó irányítószám,
- a nyugdíj jogosultság alapját képező úgynevezett *osztónapok*<sup>2</sup> száma (ezen belül az úgynevezett valódi, pszeudó és összes osztónapok száma, ahol valódi = tényleges munkaviszony révén keletkezett, és pszeudo, például anyasági vagy egyéb nyugdíj-jogosultságba beszámító ellátás révén keletkezett).

Az elemzés megkezdése előtt egyes hiányzó értékek pótlásához egy általunk kimondottan erre az adatbázisra kifejlesztett úgynevezett többszörös imputálási eljárást (*Carpenter–Kenward* [2013]) alkalmaztunk, amelyet itt nem részletezünk. Eltávolítottuk azokat a rekordokat, amelyekhez 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, nyugdíj jogosultságba beszámító valódi osztónapok száma alapján. Ennek lehetséges értéktartományát egy alkalmasan választott küszöbértéknél elvágtuk úgy, hogy a létrejövő bináris változó átlaga éppen megegyezzen a 2012. évi valódi osztónapok évesített értékének átlagával.

<sup>2</sup> Azon napok száma az adott naptári évben, amelyekre vonatkozóan az egyén a nyugdíj-megállapítás során figyelembe vehető jövedelemmel rendelkezik.

## Szegmentálás

Kihívást jelentett számunkra, hogy a közel fél évszázadnyi adattömeg felhasználását, amelyet egyebek mellett a rendszerváltozás környékén bekövetkezett, korábban már említett strukturális törés is megkövetel, hogyan tudjuk összeegyeztetni a MIDAS modell fent felvázolt keresztmetszeti jellegével.

Az életkorral, nemmel, foglalkozással és lakóhellyel kapcsolatos információk közül részben ezen adatok természeténél fogva, részben pedig a MIDAS modellek felépítéséhez igazodva elegendőnek látszott a legutóbbi ismert értékek felhasználása. A historikus foglalkoztatási adatok ezzel szemben az adott egyén jövőbeli foglalkoztatási esélyeire nézve vélhetően jelentős többletinformációt nyújtanak, így szinte tálcán kínálták magukat a munkaerő-piaci modulbeli hasznosításra. Ennek legmegfelelőbb módszertani implementációja azonban az ezen adatok hosszmetzeti és a MIDAS modell viselkedési egyenleteinek keresztmetszeti jellege közötti 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, amely a munkaerő-piaci aktivitás mérőszámának tekinthető.

A szegmentálás gyakorlati megvalósítására két lehetséges út kínálkozott: klaszteranalízis (*Kovács és szerzőtársai* [2011]) alkalmazása vagy szakmai szempontok alapján definiált szegmensok képzése. A klaszteranalízis nehezen kivitelezhetőnek tűnt, mivel az egyes évekhez tartozó beszámítható osztónapok számai erősen korrelálnak, 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 szegmensok képzése mellett döntöttünk.

A teljes felhasznált állományról mindenekelőtt leválasztottuk a tartós (például rokkantsági) ellátásban lévőket és a pszeudó jogviszonnyal (például anyasági ellátással) rendelkezőket. A maradék állományon a szegmensok 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, amely 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, amelyeket szakmai szempontok alapján három *fő szegmensbe* soroltuk: magasan, közepesen és alacsonyan foglalkoztatottak. Az összes lehetséges historikus foglalkoztatási pálya besorolását nem közöljük részletesen, de illusztratív céllal néhányat bemutatunk az 1. táblázatban.

## 1. táblázat

Példák egyes historikus foglalkoztatási pályák fő szegmensekbe sorolására

1970–1991	1992–2001	2002–2011	Fő szegmens
–	–	alacsony	alacsony
közepes	alacsony	alacsony	alacsony
–	közepes	közepes	közepes
–	magas	közepes	közepes
magas	közepes	magas	magas*
alacsony	magas	magas	magas*

\* A besorolás a legfrissebb, legutolsó pályaszakasz alapján történt.

A tisztán alacsony, közepes és magas pályák besorolása egyértelmű. A többi pálya besorolása egyfajta átlagos aktivitást tükröz, ahol az átlagolás során kiemelt hangsúlyt helyeztünk 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 státusvesztést szenvedett egyének kapták. Az utóbbiak közt túlnyomórészt olyan egyéneket találunk, akik az 1991–1992 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, besorolását „közepesre” módosítottuk. Ennek oka az volt, hogy „magas” minősítést csak olyan egyéneknek kívántunk adni, akik folyamatosan valódi vagy pszeudó jogszerző státussal rendelkeznek. Az így nyert besorolások sok hasonlóságot mutatnak *Augusztinovics–Köllő* [2007] alfa, béta és gamma kategóriáival – bár hangsúlyozottan nem azonosak azokkal –: ebben a sorrendben nagyjából megfeleltethetők a modellben szereplő „alacsony”, „közepes” és „magas” besorolásoknak, ugyanakkor azokkal ellentétben nem veszik figyelembe az egyéni foglalkoztatottság hosszú távú trendjét, és a dolgozó tömegektől nem különítik el explicit módon a foglalkoztatási szempontból különleges státusú, ellátott egyéneket, így a modellbeli besorolásnál kevésbé informatívak.

Az alacsonyan, közepesen és magasán foglalkoztatottak az ellátottakkal és a speciális (pszeudó) jogviszonnyal rendelkezőkkel együtt a modellben összesen öt fő szegmenst képeznek. Az adott évbéli 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 kiemelve a 2011. évi valódi osztónapok számát a 2012. évi eredményváltozó képzésénél alkalmazott eljárással megegyező módon, átlagot megőrző *dichotóm vágással* binárisra alakítottuk, majd az öt fő szegmenst a 2011. évi bináris foglalkoztatási változó értéke szerinti *dichotóm felosztással* végül a következő tíz *munkaerő-piaci szegmensre* bontottuk: 1. ellátott, 2011-ben inaktív; 2. ellátott, 2011-ben aktív; 3. alacsonyan foglalkoztatott,

2011-ben inaktív; 4. alacsonyán foglalkoztatott, 2011-ben aktív; 5. speciális jogviszonyú, 2011-ben inaktív; 6. speciális jogviszonyú, 2011-ben aktív; 7. közepesen foglalkoztatott, 2011-ben inaktív; 8. közepesen foglalkoztatott, 2011-ben aktív; 9. magasán foglalkoztatott, 2011-ben inaktív; 10. magasán foglalkoztatott, 2011-ben aktív.

## Modellillesztés, modellszelekció

### *Az alkalmazható többváltozós statisztikai módszerek*

A foglalkoztatást reprezentáló bináris változó becslésére a többváltozós statisztikai módszertan számos keresztmetszeti eljárást ismer. Ezek közül a legelterjedtebbek:

- bináris logisztikus regresszió (logit modell, lásd például *Kleinbaum–Klein* [2010] és *Kovács és szerzőtársai* [2011]),
- probit modell (*Bliss* [1934]),
- diszkriminanciaelemzés (lásd például *Kovács és szerzőtársai* [2011]),
- döntési fák (CHAID, CART stb., lásd például *Rokach–Maimon* [2008]),
- Support Vector Machine (SVM, *Cortes–Vapnik* [1995]),
- neurális háló (lásd például *Anderson* [1995]).

Ezek közül a bináris logisztikus regresszió alkalmazása mellett döntöttünk, figyelembe véve, hogy az adott alkalmazásban szükséges a klasszifikáció egyszerű matematikai reprezentációja és az eredményváltozóra ható tényezők hatásának explicit és szövegesen értelmezhető számszerűsítése. Döntésünk során a robusztusság követelménye is a választott módszer használata melletti érv volt.

Meg kell jegyeznünk még, hogy az adatok hosszmetzeti jellegéből adódóan a panelökonometriai eljárások (*Fitzmaurice és szerzőtársai* [2011]) megfelelők lehetnek a foglalkoztatás előrejelzésére, ugyanakkor – célunkra tekintettel szükségszerűen – figyelembe kellett vennünk azt a korlátot, hogy a MIDAS mikroszimulációs modellcsalád és a támogatására készült, az ONYF által is használt LIAM2 szoftverkörnyezet felépítése szerint a modellben tisztán keresztmetszeti egyenletek segítségével zajlik a becslés, és a munkaerő-piaci előrejelzések a keresztmetszeti egyenletek előregörgetett, ismételt alkalmazásával nyerhetők.

Az ismertetett eljárások előnyeit és hátrányait figyelembe véve a jövőbeli munkaerő-piaci aktivitás becslése érdekében a bináris logisztikus regressziós modell alkalmazása mellett döntöttünk. Ebben a modellben a 2012. évi munkaerő-piaci aktivitás valószínűségét a rendelkezésre álló potenciális magyarázó változók segítségével becsültük. Adatfájlunkat az egyszerű véletlen kiválasztás segítségével két diszjunkt részsokaságra osztottuk: az egyének 90 százalékát tartalmazó *tanuló* és a maradék 10 százalékot lefedő *tesztelő* állományra (osztott mintás validáció, lásd például *Picard–Cook* [1984]). A modellek becslését a *tanuló*, az illeszkedés vizsgálatát és a modellszelekciót pedig 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 végső soron 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 próba segítségével formálisan megvizsgáltuk, hogy az adott modell magyarázó ereje az egy szinttel egyszerűbb modellhez képest szignifikánsan nagyobb-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 (két kategória), konstans,
2. MODELL (egyetlen egyenlet): munkaerő-piaci szegmens (tíz kategória), konstans,
3. MODELL (munkaerő-piaci szegmensenként egy-egy, összesen tíz egyenlet, amely azonban formálisan egyetlen egyenletként is felfogható): nem (két kategória), FEOR főcsoport (tíz kategória), statisztikai régió (NUTS II, hét kategória), településtípus (négy kategória), életkor és annak négyzete (folytonos változók), konstans,
4. MODELL (munkaerő-piaci szegmensenként egy-egy, összesen tíz egyenlet, amely 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 ennek a potenciális magyarázó változónak a magyarázó ereje minden más szóba jöhető magyarázó változónál jóval jelentősebb. A változók beléptetését a 0–2. modellek esetén az egyszerre beléptető Enter-eljárással, az ezeknél lényegesen több paramétert tartalmazó 3. modell esetén a magyarázó változók egyenkénti szignifikanciáját tesztelő Wald-statisztikára épülő, lépésenként beléptető Forward-eljárással, a 4. modell esetén pedig az ilyen komplexitású modelleknél fokozottan jelentkező számítási korlátokból adódóan ismét a számítási szempontból egyszerűbb, egyszerre beléptető Enter-eljárással végeztük (*Kovács és szerzőtársai* [2011]). Az egyénekhez tartozó studentizált reziduumok (becsült szórással osztott hibatagok) szignifikanciáját a standard 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 is elvégeztük. Végül ezen eljárások egyik modellspecifikáció esetén sem jeleztek egyetlen szignifikáns kiugró értéket (*outlier*) sem, ami vélhetően az óriási mintaméretnek köszönhető, hiszen egy több millió 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 loglikelihood-értékek és illeszkedési mutatók [McFadden-, Cox–Snell-, Nagelkerke- és Tjör-féle pszeudó- $R^2$  értékek, a ROC (*Receiving Operator Characteristic*) görbe alatti terület

és az ebből számítható Gini-együttható], illetve az egymásba ágyazott modellek közötti választást lehetővé tevő *likelihood*-arány tesztek *p*-értékei a 2. táblázatban és az 1. ábrán láthatók.

## 2. táblázat

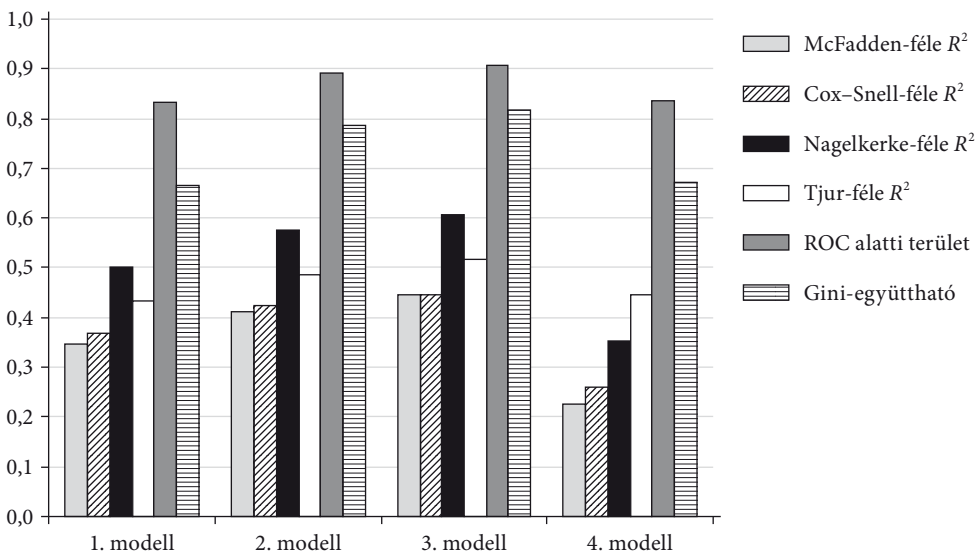
Az egymásba ágyazott modellek illeszkedési mutatói\*

Modell	Log-likelihood	McFadden-féle $R^2$ (százalék)	Cox-Snell-féle $R^2$ (százalék)	Nagelkerke-féle $R^2$ (százalék)
0. modell	-358 640	0,0	0,0	0,0
1. modell	-234 614	34,6	36,9	50,1
2. modell	-210 400	41,3	42,3	57,5
3. modell	-198 727	44,6	44,7	60,8
4. modell	-77 876	22,5	25,9	35,2
	Tjur-féle $R^2$ (százalék)	ROC alatti terület	Gini-együttható (százalék)	LR-teszt <i>p</i> -értéke (százalék)
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

\* A *p*-értékek kivételével valamennyi mutató esetében a magasabb érték utal pontosabb illeszkedésre.

## 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 likelihoodarány-próbák alapján is arra lehet következtetni, hogy a 3. modellig érdemes a modell komplexitását bővíteni, azt követően azonban az interakciós tagok beléptetése kifejezetten károsnak tekinthető. Ez feltehetően a következő okokra vezethető vissza:

- nagyszámú paramétert tartalmazó modellek esetén fellép a *túllillesztés* 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 (a modell az érvényes statisztikai tendenciákon túl a tanuló halmazon fellelhető véletlen zajt reprodukálja, amely már nem általánosítható a tesztelő halmazra);

- a nagyszámú magyarázó változó szerepeltetése *multikollinearitással* összefüggő numerikus nehézségeket okoz, a Hesse-mátrix közel szingulárisává válik, és a paraméterbecslő algoritmus gyakran az optimálisnál rosszabb 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 fő szegmens tíz szegmensre történő dichotóm felbontásának helyességét, amit jelentősen fokoz a munkaerő-piaci szegmensek figyelembevétele, és végül szignifikánsan javít más prediktorok beléptetése.

## Klasszifikáció és a becsült munkaerő-piaci egyenletek

A 2012. évi aktivitást modellünk – a becsült valószínűségekből képzett rangsor alapján, a rangsor elejéről megfelelő számú egyént kiválasztva – úgy jelzi előre, hogy munkaerő-piaci szegmensenként az aktívnek besorolt egyének száma megegyezzen a ténylegesen aktív egyének számával. Az általunk kiválasztott 3. modell besorolási pontosságát a tesztelő állományon vizsgáltuk. A 3. 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.

A 3. táblázatban szereplő tényleges találati pontosságok alapján modellünk a nem aktív egyének 73,8 százalékát, az aktívak 83,8 százalékát és a teljes állomány 80 százalékát helyesen ismeri fel, szemben a véletlen besorolás segítségével elérhető 38,2, 61,8 és 52,8 százalékkal, ahol véletlen besoroláson azt értjük, hogy az egyéneket egyénenként azonos valószínűséggel a ténylegesen tapasztalt számú aktív és nem aktív egyénre osztjuk fel. Megállapítható, hogy modellünk a véletlenszerű besorolásnál jóval megbízhatóbb: 2011-ről 2012-re a tesztelő állományban ötből négy egyént helyesen sorolt be.

## 3. táblázat

Találati arányok a tesztelő állományban tényleges aktivitás szerint és összesen

Tényleges aktivitás	Becsült aktivitás			Találati pontosság (százalék)	
	nem aktív	aktív	összesen	tényleges	véletlen
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

A logit modell segítségével becsült egyenletek a következő alakban írhatók fel:

$$\ln\left(\frac{p}{1-p}\right) = b_{S_0} + b_{S_1}K + b_{S_2}K^2 + b_{S_N}N + \sum_{k=1}^9 b_{S_{F_k}}F_k + \sum_{k=1}^6 b_{S_{R_k}}R_k + \sum_{k=1}^3 b_{S_{T_k}}T_k,$$

ahol

$p = \mathbb{P}(Y = 1 | S, K, N, \{F_k\}_{k=1}^9, \{R_k\}_{k=1}^6, \{T_k\}_{k=1}^3)$  az  $S \in \{1, 2, \dots, 10\}$  szegmenstől és a magyarázó változóktól függő becsült valószínűség. Ugyanitt  $Y$  a foglalkoztatást jelző bináris változó,  $K$  a vizsgált egyén életkora, továbbá  $N$  a női nem,  $F_k$  ( $k = 1, 2, \dots, 9$ ) a foglalkozási kategóriák,  $R_k$  ( $k = 1, 2, \dots, 6$ ) a régiók,  $T_k$  ( $k = 1, 2, 3$ ) pedig a településtípusok bináris indikátorváltozói. Az  $S$  szegmensbeli egyenletben  $b_{S_1}$  és  $b_{S_2}$  az életkor első és második hatványainak logit együtthatói,  $b_{S_{(\cdot)}}$  a  $(\cdot)$ -tal jelölt, tetszőleges egyéb magyarázó változó logit együtthatója,  $b_{S_0}$  pedig az ugyanitt szereplő additív konstans.<sup>3</sup>

A tíz becsült egyenletet itt nem kívánjuk részletesen közzétenni, a 4. táblázatban csak a legnépesebb, magasan foglalkoztatott, 2011-ben aktív szegmens egyenletének együtthatóit, az azokhoz tartozó, az együtthatók szignifikanciájára vonatkozó Wald-próbák  $p$ -értékeit és az esélyhányadosokat (*odds ratio*) mutatjuk be illusztratív céllal.

A 4. táblázat eredményei közül kiemelendő, hogy a vizsgált szegmensben nők esetében *ceteris paribus* a 2012. évi munkaerő-piaci aktivitás esélye 0,734-szerese a férfiak esetében 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 esetében 1,728-szerese, szakképzettséget nem igénylő (egyszerű) foglalkozásokéban pedig 0,693-szorosa az aktivitás esélye az egyéb (fel nem sorolt) foglalkozások esetében 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 előbbieik javára. Látszólag meglepő, hogy a fővároshoz és Közép-Magyarországhoz képest *ceteris paribus* szinte bármely településtípus és NUTS II régió esetén magasabb az aktivitás esélyhányadosa, ami

<sup>3</sup> Az egyes munkaerő-piaci szegmensek kódszámai a szegmentálásról szóló rész végén bemutatott felsorolásban szereplő sorszámok. A foglalkozási kategóriák, a régiók és a településtípusok kódszámai az 4. táblázatban szerepelnek. Kérés esetén kutatási célra a szerző e-mailben rendelkezésre bocsátja a további egyenleteket is.

## 4. táblázat

A magasan foglalkoztatott, 2011-ben aktív munkaerő-piaci szegmens logit egyenlete\*

Változó	Együttható	Wald-féle <i>p</i> -érték (százalék)	Esély- hányados
Konstans	-0,693	0,0	0,500
Életkor	0,159	0,0	1,172
Életkor × életkor	-0,001	0,0	0,999
Nő	-0,309	0,0	0,734
1. Gazdasági, igazgatási, érdek-képviselési vezetők, törvényhozók	0,239	0,0	1,270
2. Felsőfokú képzettség önálló alkalmazását igénylő foglalkozások	0,547	0,0	1,728
3. Egyéb felsőfokú vagy középfokú képzettséget igénylő foglalkozások	0,289	0,0	1,336
4. Irodai és ügyviteli (ügyfélkapcsolati) foglalkozások	0,147	0,0	1,158
5. Kereskedelmi és szolgáltatási foglalkozások	-0,243	0,0	0,785
6. Mezőgazdasági és erdőgazdálkodási foglalkozások	-0,077	14,7	0,926
7. Ipari és építőipari foglalkozások	-0,111	0,0	0,895
8. Gépközelők, összeszerelők, járművezetők	-0,021	42,6	0,979
9. Szakképzettséget nem igénylő (egyszerű) foglalkozások	-0,366	0,0	0,693
1. Dél-Alföld	0,177	0,0	1,194
2. Dél-Dunántúl	0,003	89,2	1,003
3. Észak-Alföld	0,109	0,0	1,115
4. Észak-Magyarország	0,053	0,5	1,055
5. Közép-Dunántúl	0,188	0,0	1,207
6. Nyugat-Dunántúl	0,271	0,0	1,312
1. Megyeszékhely vagy megyei jogú város	-0,005	80,0	0,995
2. Egyéb város	0,109	0,0	1,115
3. Község vagy nagyközség	0,069	0,0	1,071

\* Referenciakategória: a közép-magyarországi, egyéb (fel nem sorolt) foglalkozású férfiak.

azonban a főváros egyéb változók szerinti kedvezőbb összetételének tulajdonítható, mivel adataink tanúsága szerint Budapesten és Közép-Magyarországon magasabb az átlagos aktivitás szintje, mint az ország többi részén. A vizsgált szegmensben a 2012. évi munkaerő-piaci aktivitás logitja (az esélyhányados logaritmus) az életkor függvényében *ceteris paribus* konkáv parabolával modellezhető, amely globális 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 az eddigi tapasztalatoknak, illetve ahol ezekhez képest eltéréseket tapasztalunk, ott a különbségek jellemzően statisztikailag nem szignifikánsak.

Meg kell még jegyeznünk, hogy a magyar MIDAS modellben a tényleges előrejelzést a számítások egyszerűsítése érdekében egyének helyett a magyarázó változók szerint képzett homogén csoportokra, úgynevezett *modellpontokra* végezzük el (lásd Kovács Eszebet, Rétaillé Orsolya és Vékás Péter jelen számunkban közölt cikkét), ahol a csoportokra jellemző értékeket helyettesítjük be a becsült egyenletekbe.

## Közgazdasági értelmezés

A becsült regressziós egyenleteknek természetesen közgazdasági értelmezés is adható. Ennek során a klasszikus munkaerő-piaci modellre szorítkozunk (lásd például *Ehrenberg–Smith* [2015]), amelyben az egyéni munkaerő-kínálat a munka és a szabadidő közötti választásként írható le, a munkaerő-keresletet pedig az egyén munkájának határterméke determinálja. Megjegyezzük ugyanakkor, hogy eredményeink fejlettebb közgazdasági modellkeretben is jól interpretálhatók.

Az életkor függvényében fiatal korban a munkaerő-kínálatot csökkentő tényező lehet az iskolába járásra, egyéb emberitőke-beruházásra és a diákmunkára fordított idő, melyek felértékelik a szabadidőt a munkával szemben, illetve időskorban a hazai öregségi és rokkantnyugdíj-rendszerbe beépített, minél későbbi nyugdíjba vonulást és az idősödő korosztályok munkavállalását elősegítő pénzügyi ösztönzők hiánya (*Cseres-Gergely–Scharle* [2005], valamint *Cseres-Gergely–Hámori* [2009]). Emellett fiatal korban a tapasztalat hiánya, az életkor előrehaladtával pedig a fizikai és szellemi teljesítmény csökkenése és az esetleges betegségek növekvő kockázata egyaránt csökkentik a munka határtermékét. A becsült modellben az itt felsorolt hatások statisztikailag szignifikánsan megjelennek az életkor négyzetének szerepeltetése révén.

A férfiakhoz képest a nők esetében a munkaerő-kínálatot csökkentheti, hogy társadalmi szerepükből adódóan jellemzően több időt fordítanak a hagyományos női szerepkörrel járó feladatok (például gyermeknevelés, háztartásvezetés) ellátására (*Jaumotte* [2013]). Ugyanakkor azonos körülmények között jellemzően a kereslet is alacsonyabb a nők munkája iránt (például esetleges diszkrimináció, mélyen gyökerező kulturális attitűdök és a gyermekvállalás lehetőségéből fakadó munkáltatói bizonytalanság következtében). A becsült egyenletekben ezek a hatások a nőket jelző bináris változó negatív együtthatójában is jelentkeznek, amely valamennyi munkaerő-piaci szegmensben statisztikailag szignifikáns.

A közgazdasági elmélet szerint az iskolába járás mint emberitőke-beruházás hatékony képzés esetén növeli az egyének határtermékét, így a képzetlen munkavállalók határterméke a legalacsonyabb. A modellben ez a hatás szignifikánsan jelentkezik szinte valamennyi munkaerő-piaci szegmensben: a felsőfokú végzettséget igénylő 1–3. foglalkozási csoportok logit együtthatói szinte minden szegmensben szignifikánsak, pozitívak és nagyobbak a többi csoport együtthatóinál, míg a 9. foglalkozási csoportbeli képzetlen munkások logit együtthatója a legalacsonyabb a szegmensek többségében.

A régiók és településtípusok hatásának magyarázata már kevésbé egyértelmű: az együttthatók gyakran nem szignifikánsak, az azonos változók együttthatóinak előjele pedig munkaerő-piaci szegmensenként változó. Bár a községben lakók esélyei több munkaerő-piaci szegmensben *ceteris paribus* kedvezőbbek a többi településtípushoz képest, összességében esélyeik mégis kedvezőtlenebbek a községben élők más magyarázó változók szerinti kedvezőtlenebb összetétele miatt.<sup>4</sup>

## Összefoglalás

Tanulmányunk az egyéni munkaerő-piaci aktivitás becslését mutatta be az egyének múltbeli munkaerő-piaci aktivitására, születési évrre, nemére, foglalkozására és lakóhelyére vonatkozó információk felhasználásával, statisztikai viselkedési egyenletek segítségével. Az egyenletek becsléséhez a hiányzó adatok pótlását és az egyének foglalkoztatástörténete szerinti szegmentálását követően bináris logisztikus regressziót alkalmaztunk, figyelembe véve a módszer más eljárásokkal szembeni előnyeit és a MIDAS modellek szerkezetéből adódó korlátokat. A becslést szegmensenként külön-külön végeztük el, modellünket pedig körültekintően validáltuk. A modell illeszkedése megfelelőnek mondható: a tesztelő állományon 80 százalékos körüli találati pontossággal rendelkezik.

A felhasznált csoportképző és magyarázó változók értékeit előre számítva a becslést egyenletek alapján, a tanulmányban ismertetett kiigazítási eljárás (exogén előrejelzésekhez történő kiigazítás) figyelembevételével egyénenként megbecsülhető a jövőbeli évekre vonatkozó munkaerő-piaci aktivitás ténye, amely a magyar MIDAS modell többi moduljával és almoduljával kölcsönhatásban együttesen előrevetheti például a nyugdíjassza jövőbeli bevételeinek és kiadásainak egyénenkénti alakulását, egyes társadalmi csoportok anyagi helyzetének jövőbeli változásait stb.

Mivel az ismertetett modell elkészültének időpontjában a teljes mikroszimulációs modell még félkész állapotban volt, így az itt szereplő eredmények szükségszerűen részlegesek: a modulok közötti kölcsönhatások miatt a félkész modell teljes körű nyugdíj-hatásvizsgálat elvégzésére és munkaerő-piaci forgatókönyvek, illetve paraméteres nyugdíjreformok vizsgálatára nem volt alkalmas. A teljes modell elkészültét követően már ilyen jellegű elemzésekre és azok eredményeit bemutató tanulmányok közzétételére is lehetőség nyílik, mellyel a szerző a jövőben élni kíván.

A magyar MIDAS mikroszimulációs modell egyik lényeges korlátja, hogy míg az összetétel-változásokat viszonylag pontosan vetíti előre, a foglalkoztatás volumenét nem képes és nem is kívánja előre jelezni: azt külső előrejelzések alapján exogén paraméterként kezeli, és arra vonatkozóan készít részletes, egyéni szintű prognózist, hogy várhatóan kik lesznek aktívak és inaktívak a jövőben a munkaerőpiacon. A modellben alkalmazott statisztikai megközelítés további lényeges korlátja, hogy bár az

<sup>4</sup> A regionális hatások tekintetében hasonlóan vegyes képet és gyenge szignifikanciaszinteket mutat be csehországi adatokon *Fialová–Mysíková* [2009], és a kor és nem hatásának mértéke is hasonló modellünk eredményeihez.

egyéni foglalkoztatás tényét kielégítő pontossággal képes legalább rövid távon előre jelezni, a munkaerőpiac keresleti és kínálati oldalán jelentkező hatásokat – például a *Bakó és szerzőtársai* [2013] által bemutatott előrejelző rendszerrel szemben – nem képes elválasztani egymástól, és explicit módon nem jelenik meg benne a kereslet és kínálat között jelentkező mennyiségi eltérés. Továbbá optimalizáló viselkedés hiányában az exogén paraméterek (például nyugdíjkorhatár) változása esetén nem veszi figyelembe az egyéni viselkedés esetleges változásait.

Fontos még megjegyezni, hogy a felhasznált adminisztratív adatállomány természeténél fogva nem terjed ki a feketegazdaságban foglalkoztatottakra: ez csökkentheti bizonyos, a modell segítségével készített jövőbeli elemzések megbízhatóságát, ugyanakkor az ONYF számára legfontosabb nyugdíjárulék-befizetések és nyugdíjkifizetések szempontjából a gazdaság e szegmense kevésbé releváns, így például a nyugdíjkassza egyenlegének előrejelzése során e tényező hatása nem számottevő.

Tanulmányunk a szakirodalomban fellelhető korábbi, magyarországi munkaerőpiacra kapcsolatos elemzéseknél jóval szélesebb körű, egyesített adminisztratív adatbázisok felhasználásával készült, ami jelentős előrelépésnek számít. A MIDAS\_BE modellhez képest is jelentős haladás, hogy két év adatai helyett több mint negyven év információit használtuk fel az egyenletek becslése során, így lehetőségünk nyílt a magyarországi munkaerőpiacon az 1990-es évek elején lezajlott strukturális átalakulás hatásainak figyelembevételére is.

A becsült munkaerő-piaci egyenletek a MIDAS modell keretein belül a jövőben lehetővé teszik a munkaerő-piaci aktivitás minden korábbinál pontosabb, egyéni szintű előrejelzését, amelyet az Országos Nyugdíjbiztosítási Főigazgatóság és a nyugdíjszakma a nyugdíjrendszerrel kapcsolatos elemzések és hatástanulmányok készítése során – reményeink szerint – sikerrel használhat a jövőben.

### *Hivatkozások*

- ANDERSON, J. A. [1995]: *An Introduction To Neural Networks*. MIT Press, Cambridge, Massachusetts.
- ANTAL GÁBOR–TELEGDY ÁLMOS [2013]: A foglalkoztatási szerkezet előrejelzése foglalkozási csoportok, nem és iskolai végzettség szerint. Megjelent: *Fazekas Károly–Varga Júlia* (szerk.): *Trendek és előrejelzések. Munkaerő-piaci prognózisok készítése, szerkezetváltás a munkaerőpiacon*. MTA KTI, Budapest.
- AUGUSZTINOVICS MÁRIA [2005]: Népeség, foglalkoztatottság, nyugdíj. *Közgazdasági Szemle*, 52. évf. 5. sz. 429–447. o.
- AUGUSZTINOVICS MÁRIA–GYOMBOLAI MÁRTON–MÁTÉ LEVENTE [2008]: Járulékfizetés és nyugdíjjogosultság 1997–2006. *Közgazdasági Szemle*, 55. évf. 7–8. sz. 665–689. o.
- AUGUSZTINOVICS MÁRIA–KÖLLŐ JÁNOS [2007]: Munkapiaci pálya és nyugdíj, 1970–2020. *Közgazdasági Szemle*, 54. évf. 6. sz. 529–559. o. <http://epa.oszk.hu/00000/00017/00138/pdf/2augusztinovics-kollo.pdf>.
- AUGUSZTINOVICS MÁRIA–KÖLLŐ JÁNOS [2008]: *Decreased Employment and Pensions*. Megjelent: *Holzmann, R.–Mackellar, L.–Repensek, J.* (szerk.): *Pension Reform in South-Eastern Europe: Linking to Labor and Financial Market Reforms*. The World Bank, Washington D.C.



- BAKÓ TAMÁS [2013]: A magyarországi munkapiac 2012–2013–ban. Megjelent: *Fazekas Károly–Neumann László* (szerk.): Munkaerőpiaci Tükör 2013. MTA KRTK, Közgazdaságtudományi Kutatóintézet–Országos Foglalkoztatási Nonprofit Kft., Budapest, [http://econ.core.hu/file/download/mt\\_2013\\_hun/egyben.pdf](http://econ.core.hu/file/download/mt_2013_hun/egyben.pdf).
- BAKÓ TAMÁS [2014]: A magyarországi munkapiac 2013–2014–ben. Megjelent: *Fazekas Károly–Neumann László* (szerk.): Munkaerőpiaci Tükör 2014. MTA KRTK, Közgazdaságtudományi Kutatóintézet–Országos Foglalkoztatási Nonprofit Kft., Budapest, [http://econ.core.hu/file/download/mt\\_2014\\_hun/egyben.pdf](http://econ.core.hu/file/download/mt_2014_hun/egyben.pdf).
- BAKÓ TAMÁS–CSERES-GERGELY ZSOMBOR–GALASI PÉTER [2013]: Az MTA KRTK KTI munkaerő-piaci előrejelző rendszere. *Közgazdasági Szemle*, 60. évf. 2. sz. 117–133. o.
- BÁLINT MÓNIKA–KÖLLŐ JÁNOS–MOLNÁR GYÖRGY [2010]: Nyugdíj megszerzés és teljes aktív életút. *Statisztikai Szemle*, 88. évf. 6. 623–647. o.
- 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*, Vol. 57. No. 1. 289–300. o. [http://engr.case.edu/ray\\_soumya/mlrg/controlling\\_fdr\\_benjamini95.pdf](http://engr.case.edu/ray_soumya/mlrg/controlling_fdr_benjamini95.pdf).
- BLISS, C. I. [1934]: The Method of Probits. *Science*, Vol. 79. No. 2037. 38–39. o. <http://dx.doi.org/10.1126/science.79.2037.38>.
- CARPENTER, J.–KENWARD, M. [2013]: *Multiple Imputation and its Applications*. John Wiley and Sons, New York, <http://dx.doi.org/10.1002/9781119942283>.
- CORTES, C.–VAPNIK, V. [1995]: Support-vector networks. *Machine Learning*, Vol. 20. No. 3. 273. <http://dx.doi.org/10.1007/bf00994018>.
- CSERES-GERGELY ZSOMBOR–HÁMORI SZILVIA [2009]: Az iskolázottság és a foglalkoztatás kapcsolata nemzetközi összehasonlításban (1999–2005). Megjelent: *Fazekas Károly* (szerk.): *Oktatás és foglalkoztatás*. KTI Könyvek, 12. sz. MTA KTI, Budapest, 43–61. o.
- CSERES-GERGELY ZSOMBOR–SCHARLE ÁGOTA [2005]: Az idősök alacsony aktivitásáról és ennek intézményi okairól. Megjelent: *Hárs Ágnes–Landau Edit–Nagy Katalin* (szerk.): *Európai foglalkoztatási stratégia: Lehetőségek és korlátok az új tagállamok számára*. Kopint–Datorg Konjunktúra Kutatási Alapítvány, Budapest, 129–142. o.
- DEKKERS, G. [2010]: The long-term adequacy of the Belgian public pension system: An analysis based on the MIDAS model. Working Paper, No. 10-10. Federal Planning Bureau, Brüsszel, <http://www.plan.be/admin/uploaded/201005110837500.wp201010.pdf>.
- DEKKERS, G. [2013]: An introduction to MIDAS\_BE, the dynamic microsimulation model for Belgium. Using process produced data in pension microsimulation models, Budapest, július 8–9. Centre for Sociological Research, CANPI, Brüsszel, [https://lirias.kuleuven.be/bitstream/123456789/441042/1/2013\\_MIDAS\\_HU\\_an+introduction+to+midas.pdf](https://lirias.kuleuven.be/bitstream/123456789/441042/1/2013_MIDAS_HU_an+introduction+to+midas.pdf).
- DEKKERS, G.–BELLONI, M. [2009]: Micro simulation, pension adequacy and the dynamic model MIDAS: an introduction. Working Paper, Federal Planbureau, Brüsszel, [http://www.ief.es/documentos/investigacion/seminarios/economia\\_publica/2009\\_21Enero.pdf](http://www.ief.es/documentos/investigacion/seminarios/economia_publica/2009_21Enero.pdf).
- EHRENBERG, R. G.–SMITH, R. S. [2015]: *Modern Labor Economics: Theory and Public Policy* [12<sup>th</sup> edition]: Prentice Hall, Upper Saddle River, New Jersey.
- FAZEKAS KÁROLY–VARGA JÚLIA (szerk.) [2015]: *Munkaerőpiaci Tükör 2014*. MTA KTI, Budapest. [http://econ.core.hu/file/download/mt\\_2014\\_hun/egyben.pdf](http://econ.core.hu/file/download/mt_2014_hun/egyben.pdf).
- FIALOVÁ, K.–MYSÍKOVÁ, M. [2009]: Labor Market Participation: The Impact of Social Benefits in the Czech Republic. IES Working Paper: 4/2009. Institute of Economic Studies, Faculty of Social Sciences, Charles University, Prága. <http://ies.fsv.cuni.cz/default/file/download/id/9986>.

- FITZMAURICE, G. M.–LAIRD, N. M.–WARE, J. H. [2011]: *Applied Longitudinal Analysis*. 2. kiadás, John Wiley and Sons, Hoboken, New Jersey.
- GALASI PÉTER–VARGA JÚLIA [2005]: *Munkaerőpiac és oktatás*. KTI Könyvek, 1. sz. MTA Közgazdaságtudományi Intézet, Budapest.
- HOLTZER PÉTER (szerk.) [2010]: *Jelentés a Nyugdíj és Időskor Kerekasztal tevékenységéről*. Miniszterelnöki Hivatal, Budapest.
- JAUMOTTE, F. [2003]: *Female Labour Force Participation: Past Trends and Main Determinants in OECD Countries*. OECD Economics Department Working Papers, No. 376., [http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?doclanguage=en&ote=eco/wkp\(2003\)30](http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?doclanguage=en&ote=eco/wkp(2003)30).
- KERTESI GÁBOR–KÖLLŐ JÁNOS [2006]: Felsőoktatási expanzió, „diplomás munkanélküliség” és a diplomák piaci értéke. *Közgazdasági Szemle*, 53. évf. 3. sz. 201–225. o.
- KERTESI GÁBOR–VARGA JÚLIA [2005]: Foglalkoztatás és iskolázottság Magyarországon. *Közgazdasági Szemle*, 52. évf. 7–8. sz. 633–662. o.
- KLEINBAUM, D. G.–KLEIN, M. [2010]: *Logistic regression*. 3. kiadás, Springer-Verlag, New York, <http://dx.doi.org/10.1007/978-1-4419-1742-3>.
- KOVÁCS ERZSÉBET–SZÜLE BORBÁLA–FLISZÁR VILMOS–VÉKÁS PÉTER [2011]: *Pénzügyi adatok statisztikai elemzése*. 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*, Vol. 6. No. 2. 3–55. o. [http://www.microsimulation.org/IJM/V6\\_2/2\\_IJM\\_6\\_2\\_2013\\_Li\\_Odonoghue.pdf](http://www.microsimulation.org/IJM/V6_2/2_IJM_6_2_2013_Li_Odonoghue.pdf).
- PICARD, R.–COOK, D. [1984]: Cross-Validation of Regression Models. *Journal of the American Statistical Association*, Vol. 79. No. 387. 575–583. o. <http://dx.doi.org/10.2307/2288403>.
- ROKACH, L.–MAIMON, O. [2008]: *Data mining with decision trees: theory and applications*. World Scientific Publishing Corporation, Szingapúr, <http://dx.doi.org/10.1142/6604>.
- SZABÓ SÁNDORNÉ–CSEMNICZKI KATALIN [2000]: *Nyugdíjrendszerünk 1929-től 1997-ig*. Megjelent: *Augusztinovics Mária* (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*. SAGE Discussion Papers, No. 2.