

LEMORZSOLÓDÁS ÉS SÚLYOZÁS AZ ÉLETÜNK FORDULÓPONTJAI PANELFELVÉTEL BEN

Bartus Tamás

ÖSSZEFOGLALÓ

A 2001-ben újtára indított Életünk fordulópontjai panelvizsgálat mind ez idáig utolsó, negyedik hullámában már csak a kiinduló minta felével készült interjú. Jelen tanulmány (1) elemzi az Életünk fordulópontjai felvételnél tapasztalt lemorzsolódás társadalmi-demográfiai aspektusait, okait; (2) a lemorzsolódás-elemzésre támaszkodva kidolgoz egy súlyozási eljárást, melynek célja a szelektív panelkopásból fakadó torzítások minimalizálása; végül (3) megvizsgálja, a súly csökkenti-e bizonyos demográfiai eseményekre vonatkozó becslések torzítását. Az eredmények szerint a lemorzsolódás szelektív: valószínűbb a fiatalok, a férfiak, a nem házasok, az iskolázatlanok és a Közép-Magyarországon élők között. A szelektív lemorzsolódást kompenzáló súlyok valamelyest javítanak a születések számára vonatkozó becsléseken, ugyanakkor nem mérséklük a házasságkötések és a válások számára vonatkozó becslések megbízhatatlanságát. A társadalmi-demográfiai ismérvek válaszadásra gyakorolt hatását tükröző súlyok tehát nem szavatolják a tetszőleges súlyozott becslések megbízhatóságát.

Tárgyszavak: Életünk fordulópontjai, lemorzsolódás, súlyozás, adatminőség

Bartus Tamás

Budapesti Corvinus Egyetem, Szociológia és Társadalompolitika Intézet

E-mail: tamas.bartus@uni-corvinus.hu

BEVEZETÉS*

A 2001-ben útjára indított Életünk fordulópontjai panelvizsgálat célja azt volt, hogy lehetővé váljon a demográfiai viselkedés – a gyerekvállalás, a párkapcsolat-formálódás és a nyugdíjba vonulás – okainak és következményeinek részletes vizsgálata (Spéder 2003). Az első hullám 16 363, 1926. január 1. és 1983. december 31. között született, Magyarországon magánháztartásban élő egyénre terjedt ki. Az adatfelvételt 2004-ben, 2008-ban és 2012-ben ismételték meg. A mind ez idáig utolsó, negyedik hullámában már „csak” 8103 egyénnel készült interjú, ami a kiinduló minta felét teszi ki. Az első hullám során adatot szolgáltató, de a 2–4. hullám alatt lemorzsolódott személyek közül 1735 meghalt, 2921 elköltözött vagy elérhetetlen volt, 3604 pedig megtagadta az adatszolgáltatást vagy válaszára képtelennek bizonyult.

A lemorzsolódásnak két, teljesen eltérő következménye van. Az egyik a becslések pontatlanabbá válása. Közismert, hogy a becslések mintavételi ingadozását mérő standard hibák fordítottan arányosak a mintanagyság négyzetgyökével. A kiinduló minta megfelezősége tehát azzal jár, hogy a 4. hullám adataiból készített becslések standard hibái durván 1,4-szeresei azoknak a standard hibáknak, amelyeket akkor kapnánk, ha a 4. hullám és az 1. hullám mintanagysága azonos lenne. A másik következmény a becslések esetleges torzulása. A 4. hullámból számított becslések torzítottak lesznek, ha a kutatás során vizsgált (függő) változó korrelál a 4. hullámba kerülés valószínűségével (Holt – Elliot 1991, Bethlehem 2002). A regresszióelemzés kontextusában ez a szelekciós torzítás problémája (Heckman 1979).

A torzítás irányának és nagyságának pontos megismerése lehetetlen: a függő változó és a mintába kerülés korrelációját nem tudjuk kiszámolni, hiszen a függő változót nem figyelhetjük meg a következő hullámból kimaradó egyéneknél. A torzítás legfeljebb a panelfolytatást befolyásoló tényezők elemzésével becsülhető, feltéve, hogy ezek a tényezők erősen korrelálnak a függő változóval. A panelkopással és a nem válaszolással kapcsolatos korábbi kutatások alapján a lemorzsolódás az olyan társadalmi-demográfiai ismérvek függvénye, mint a nem, az etnikai hovatartozás, az életkor, az iskolázottság, a jövedelem, a családi állapot és a lakóhely jellege (Beckett et al. 1988, Fitzgerald et al. 1998, Zabel 1998, Alderman et al. 2001, Baigrie – Eyal 2014, Buber-Ennsner 2014, Vergauwen et al. 2015, Watson – Wooden 2009). Ezek az eredmények kétféleképpen használhatók a szelektív panelkopásból faka-

* A tanulmány az MTA Bolyai János kutatási ösztöndíj támogatásával készült. Az elemzés a KSH Népeségtudományi Kutatóintézetrel való együttműködés keretében valósult meg.

dó torzítások minimalizálására. Az egyik módszer az adatbázis súlyozása és a – panelkopást korrigáló – súlyok használata a becslések során (Little – Rubin 1987, Fitzgerald et al. 1998). A másik – a szelekciós torzítás Heckman (1979) által kidolgozott elemzésén nyugvó – módszer a kutatás középpontjában álló regressziós modell és a továbblépést magyarázó regressziós modell együttes becslése (Beckett et al. 1988, Lillard – Panis 1998, Fitzgerald et al. 1998, Behr – Bellgardt – Rendtel 2003, Zabel 1998, Baigrie – Eyal 2014). Az utóbbi eljárást alkalmazó kutatások – különösen a Panel Study of Income Dynamics adatfelvétel elemzésekor – gyakran arra a következtetésre jutottak, hogy a panelkopás nem torzítja a tartalmi szempontból fontos regressziós becsléseket, ez tehát véletlenszerűnek tekinthető (Beckett et al. 1988, Lillard – Panis 1998, Fitzgerald et al. 1998, Behr – Bellgardt – Rendtel 2003, Zabel 1998).

Jelen tanulmány (1) elemzi az Életünk fordulópontjai felvételnél tapasztalt lemorzsolódás társadalmi-demográfiai aspektusait, okait; (2) a lemorzsolódás-elemzésre támaszkodva kidolgoz egy súlyozási eljárást, amelynek célja a szelektív panelkopásból fakadó torzítások minimalizálása; végül (3) megvizsgálja, hogy a súly csökkenti-e bizonyos demográfiai eseményekre vonatkozó becslések torzítását. A lemorzsolódás elemzése – a statisztikai módszer, a függő és magyarázó változók specifikálása – a nemzetközi gyakorlatot követi. Az elemzés során a sikertelen kapcsolat-felvételre és a válasz-megtagadásra fókuszálunk: az időközben elhalálozott egyénekre – a Brit Háztartáspanel és a Német Gazdasági-Társadalmi Panel gyakorlatát követve (Taylor et al. 2010, Kroh 2009) – úgy tekintünk, mint akik nem részei a célsokaságnak. A j -edik hullámban való részvételt a $(j-1)$ hullámból származó változókkal magyarázzuk. A magyarázó változók kiválasztása a korábbi kutatási eredményeket veszi figyelembe: eszerint a lemorzsolódás inkább a férfiakra, a fiatalokra, az egyedülállókra, az iskolázatlanokra és a városokban élőkre jellemző, míg a munkaerőpiaci aktivitás hatásával kapcsolatos eredmények vegyesek (Watson – Wooden 2009). A lemorzsolódás elemzése lehetővé teszi a szelektív panelkopást kompenzáló súlyok kidolgozását is.

A súlyok érvényességét egyrészt az veszélyeztetheti, hogy a mintába kerülés nem véletlenszerű az adott társadalmi-demográfiai ismérvekkel rendelkezők között (Little – Rubin 1987); másrészt az, hogy hibával mérjük a mintába kerülést befolyásoló tényezőket. A sokaságra vetített súlyozott becslések és a sokasági adatok összehasonlítása lehetővé teszi a súlyozási eljárás mögött meghúzódó feltevések felülvizsgálatát. Az Életünk fordulópontjai adatbázis 2. hullámából származó, házasságkötésekre és születésekre vonatkozó becslések érvényességét Vergauwen és munkatársai (2015) vizsgálták, és az 1970–2000

közötti évekre vonatkozó retrospektív becsléseket megbízhatónak találták. Jelen tanulmány a 4. hullám adataiból tesz retrospektív, a 2001 és 2008 közötti évekre vonatkozó becsléseket a házasságkötések, a válások és a születések számára. A becslések és az ismert sokasági adatok összehasonlításával azt mutatjuk be, hogy az Életünk fordulópontjai kutatás számára fontos demográfiai események számát nem tudjuk jól becsülni a 4. hullámból, és a becslés hibáit a súlyozás sem tudja teljes mértékben orvosolni. A tanulmány e probléma elemzésével zárul.

TOVÁBBLÉPÉS AZ ÉLETÜNK FORDULÓPONTJAI PANELFELVÉTELBEN

Továbblépés és nem válaszolás a 2-4. hullámoknál

A továbblépés hiányának, azaz a lemorzsolódásnak három különböző oka van: a célsokaság elhagyása (elhalálozás), a kapcsolatfelvétel meghiúsulása (külföldre költözés, ismeretlen címre költözés, az interjúalany távolléte a kérdezőbiztosi megkeresések során, a válaszadás meghiúsulásának egyéb oka), illetve az interjú meghiúsulása (válaszképtelenség vagy a válaszadás elutasítása, miután a potenciális válaszadóval felvették a kapcsolatot). A nemzetközi gyakorlat alapján különbséget teszünk a nyers továbblépési, illetve a válaszadási valószínűség között. A nyers továbblépési valószínűség az előző hullámban részt vettek számára vetíti a továbblépők számát. A válaszadási valószínűség ezzel szemben az időközben elhunytakra (és a külföldre távozottakra) úgy tekint, mintha nem vettek volna részt a megelőző felvételben, tehát a célsokaságban maradókra vetíti a válaszadók számát. A nagy hagyományra visszatekintő panelfelvételek – például az említett Brit Háztartáspanel és a Német Gazdasági-Társadalmi Panel – súlyozási gyakorlatát követve elemzésünk középpontjában a nem válaszolás, ill. a válaszadási valószínűség áll. A nemzetközi gyakorlattól eltérően a külföldre költözést nem tekintjük a célsokaság elhagyásának. Egyrészt a külföldre költözöttek bizonyos eséllyel hazatérhetnek. Másrészt, ha az ismeretlen címre költözők között volnának külföldre távozottak, akkor nem lenne éles határvonal a célsokaság elhagyása és a sikertelen kapcsolatfelvétel között.¹

A lemorzsolódás okait, valamint a nyers továbblépési és a válaszadási valószínűségeket az *1. táblázat* foglalja össze. A nyers folytatási arány és a válaszadási

¹ A gondolatért az egyik anonim bírálónak tartozom köszönettel.

arány egyaránt csökkent a panelfelvétel előrehaladtával. A nyers folytatási arány csökkenése részben a minta öregedésének és a halálozási arány növekedésének tudható be: a kiinduló mintanagyság zsugorodásával párhuzamosan nőtt az elhunytak abszolút száma, arányuk pedig 3%-ról 5,6%-ra nőtt. Másik oka – és egyben a válaszadási arány csökkenésének fő oka – pedig a válasz-megtagadások növekedése. A válassz megtagadók részaránya a 2. hullámbeli 6%-ról 10%-ra ugrott a 3. hullámra, majd a 4. hullám idején megközelítette a 11%-os mértéket. A kapcsolat-felvételi arány ezzel szemben némileg javult a 3. hullámra.

1. táblázat: A lemorzsolódás okai, valamint a nyers folytatási és a válaszadási arányok
Causes of panel attrition, raw panel continuation rates and response rates

	1-->2. hullám	2-->3. hullám	3-->4. hullám
Kiinduló minta	16363	13540	10641
Megvalósult minta	13540	10641	8103
Lemorzsolódók száma:			
válassz megtagadás	1041	1400	1163
elérhetetlen	1282	861	778
meghalt	500	638	597
Nyers folytatási arány (%)	82,75	78,59	76,15
Válassz megtagadás (%)	6,36	10,34	10,93
Elérhetetlen (%)	7,83	6,36	7,31
Meghalt (%)	3,06	4,71	5,61
Válasszadási arány (%)	85,36	82,48	80,68

Megjegyzések: Nyers folytatási arány = a megvalósult minta és a kiinduló minta esetszámának hányadosa.

Válasszadási arány = a megvalósult minta nagysága osztva a kiinduló minta és a meghaltak különbségével.

Forrás: saját számítások az Életünk fordulópontjai adatfelvétel 1–4. hullámának adataiból.

Érdemes megjegyezni, hogy a válassz megtagadók száma a 3. és a 4. hullám lebonyolításakor haladta meg az elérhetetlenek számát. Ezzel szemben az 1. és a 2. hullám lekérdezésekor a válassz megtagadók voltak valamivel kevesebben (az 1. hullámra vonatkozó adatokat lásd: Gyabronkáné 2003). Ez a tény egybecseng a nemzetközi szakirodalomból is ismert tapasztalattal, miszerint a válassz megtagadási arányok folyamatosan növekednek az utóbbi évtizedekben (lásd például: Meyer – Mok – Sullivan 2015).

A válaszadást befolyásoló változók kiválasztása

A panelkopással kapcsolatos nemzetközi tapasztalatok alapján a válaszadási hajlandóság társadalmi és demográfiai változók függvénye. Ebben az alfejezetben a j -edik ($j=\{1,2,3\}$) hullámból vett társadalmi-demográfiai ismérvek szerint vizsgáljuk a $(j+1)$ -edik hullámra vonatkozó válaszadási arány eloszlását. A magyarázó változók kiválasztásának fő szempontja az volt, hogy az elemzés összehasonlítható legyen a nemzetközi szakirodalommal. A lemorzsolódással kapcsolatos korábbi kutatásokban a nem, az etnikai hovatartozás, az életkor, az iskolai végzettség, a családi állapot és a gyerekek száma vagy jelenléte, a munkaerő-piaci státusz és a lakóhely jellemzői a tipikus magyarázó változók. A változók kiválasztása mögött világos elméleti megfontolások húzódnak meg. Egyrészt a válaszadás már feltételezi a kapcsolatfelvételt. Könnyebb kapcsolatot létesíteni olyan személyekkel – például nőekkel, kisebb gyermekeket nevelőkkel, házassággal, inaktívakkal és kisebb települések lakóival –, akik a háztartásban betöltött szerepük és a foglalkozásuk miatt eleve gyakrabban vannak otthon. Másrészt a válaszadásnak időkölségei vannak: a kérdőív megválaszolásával a kérdezett egyéb értékes tevékenységeiről mond le. Az időallokáció közgazdasági elmélete (Becker 1965) alapján ez a dolgozók, és különösen a magas jövedelemmel rendelkezők számára költséges. Harmadrészt a kérdezettek tarthatnak attól, hogy adataikat mégsem kezelik bizalmasan (Groves – Couper 1998). A kérdezettekbe, a kutatóintézetekbe és általában a tudományos kutatásba vetett bizalom szintje valószínűleg magasabb az iskolázottabbak körében. Végül az interjú témája is befolyásolja a válaszadást (Groves – Singer – Corning 2000): egy, a demográfiai eseményekkel kapcsolatos kérdőívre kisebb eséllyel válaszolnak a téma iránt közömbös egyének, valamint akik személyes kudarcként éltek át olyan eseményeket, amelyek – a kutatás témája alapján – szóba fognak kerülni a kérdés során. Az elméleti megfontolások alapján inkább a férfiak, a fiatalok, az egyedülállók, az iskolázatlanok és a városokban élők között magas a lemorzsolódás; a munkaerő-piaci aktivitás hatását azonban nehéz megjósolni (Watson – Wooden 2009, Vergauwen et al. 2015).

E szempontok alapján az alábbi változókat használtuk: a születési kohorsz kategóriákat (az 1926–1941, 1942–1951, 1952–1961, 1964–1971 és 1972–1983 között születetteket); a nemzet; a település régióját, illetve jogállását; a legmagasabb iskolai végzettséget (0–8 általános, szakmunkásképző, érettségi és diploma); az aktivitást (tanuló, alkalmazott, vállalkozó, egyéb/alkalmi munkavégzés, munkanélküli, GYES/GYED, nyugdíjas és egyéb inaktív); a családi állapotot (nőtlen/hajadon, házasság, elvált, özvegy); és a háztartásban élő, 0–3 és 4–6 éves gyermek

jelenlétét.² A kiválasztott változóknál előfordulnak adathiányok: ezeket nem pótoltam, az elemzési minták tehát kisebbek lesznek az 1. táblázat „kiinduló minta” sorában feltüntetett számoknál.

A válaszadást befolyásoló tényezők I. Leíró elemzés

A válaszadási arányok időbeli alakulását és a kiválasztott társadalmi-demográfiai ismérvek szerinti eloszlását a 2. táblázat mutatja. A *nemek* szerinti különbség megfelel a várakozásoknak: a nők 2–4 százalékponttal nagyobb valószínűséggel válaszolnak. A *születési kohorsz* szerinti bontás azt mutatja, hogy a válaszadási hajlandóság az idősebb korosztályoknál nagyobb: a továbblépés valószínűsége az életkor monoton csökkenő függvénye. A legfiatalabb és a legidősebb kohorsz közti különbség a kezdeti 7 százalékponttól 10 százalékpontra nőtt. Az idősebbek magasabb válaszadási hajlandósága nem meglepő annak fényében, hogy már az első hullám lekérdezésekor is ők választottak nagyobb eséllyel (Gyabronkáné 2003).

Az *iskolázottság* szerinti különbségek mérsékeltek és időben változnak. A második adatfelvétel idején az érettségizettek és a diplomások kisebb eséllyel választottak, mint az iskolázatlan egyének. A harmadik hullámra az összefüggés U alakúvá vált: a legkisebb mértékben a középfokú végzettséggel rendelkezők választottak. A negyedik hullámra viszont lényegében eltűntek az iskolázottság szerinti különbségek.

Meglepő módon nincs komoly különbség a gyermeket nevelő és a gyermektelen személyek válaszadási hajlandósága között. A *családi állapot* szerinti különbségek ezzel szemben jelentősek, és megfelelnek a várakozásoknak: az egyedülálló és az elváltak a házasoknál és az özvegyeknél kisebb eséllyel válaszolnak. A válaszmegtagadás mértéke a házastársukkal együtt élők között a legkisebb, míg a párkapcsolattal rendelkező, de partnerüktől külön élők között a legnagyobb. Ez utóbbiak között a válaszadási hajlandóság még az állandó partnerrel nem rendelkezőknél is kisebb. Az élettárral élők válaszadási hajlandósága alacsonyabb a házastárral élőkénél, a különbség különösen a 2. és a 4. hullámnál volt jelentős.

² A 6 évnél idősebb gyermek jelenlétére nem hoztunk létre változókat. A magyarázó változók időbeli késleltetése és a szomszédos hullámok közti minimális 3 éves távolság miatt az előző hullámban legalább 7 éves gyermek már legalább 10 éves abban az időpontban, amikor a kérdézbiztos felkeresni igyekszik az egyik szülőt. Feltételezésünk szerint a legalább 10 éves gyermeket (vagy gyermekeket) nevelő felnőttek a kapcsolatfelvétel és a válaszadás tekintetében ugyanúgy viselkednek, mint a gyermektelenek, mivel az ilyen idős gyerekek elég nagyok ahhoz, hogy szülői felügyelet nélkül is otthon legyenek, akár a fiatalabb testvérekkel is.

2. táblázat: Válaszadási arányok a panelkopást magyarázó változók szerint (%)
 Response rates by selected explanatory variables (%)

	1-->2. hullám	2-->3. hullám	3-->4. hullám
Összesen	85,36	82,48	80,68
Nem			
férfi	84,46	80,02	78,72
nő	86,09	84,40	82,11
Születési kohorsz			
1926–1941	88,48	86,82	84,46
1942–1951	88,53	86,74	85,63
1952–1961	86,29	82,95	83,85
1962–1971	83,53	80,84	77,24
1972–1983	81,45	77,32	74,60
Iskolai végzettség			
0–8 általános	87,13	84,34	81,33
szakmunkásképző	86,08	80,47	79,74
érettségi	83,96	81,63	80,91
diploma	83,72	84,81	80,77
0–3 éves gyerekek a háztartásban			
nincs	85,34	82,55	80,84
van	85,51	81,70	78,97
4–6 éves gyerekek a háztartásban			
nincs	85,43	82,37	80,87
van	84,23	84,02	77,78
Hivatalos családi állapot			
nőtlen/hajadon	80,93	76,14	72,55
házas	87,86	84,32	83,21
elvált	81,11	80,77	79,70
özvegy	86,42	87,66	81,15
Partnerkapcsolati helyzet			
házastársával él	88,22	84,40	83,46
élettársával él	81,12	82,55	76,08
különélő partnere van	76,91	75,00	70,93
nincs állandó partnere	83,21	79,87	78,12
Aktivitás			
tanuló	82,45	72,36	62,07
alkalmazott	83,93	81,64	78,96
vállalkozó	81,56	75,38	78,22
egyéb/alkalmi munka	83,01	76,40	81,25
munkanélküli	86,14	78,83	74,00
GYES/GYED	87,92	83,05	81,52
nyugdíjas	88,76	86,16	84,47
eltartott, egyéb inaktív	86,10	82,47	77,89
Településtípus			
Budapest	71,29	76,43	74,82
megyei jogú város	85,87	82,74	82,94
egyéb város	86,94	82,72	81,33
község	91,03	84,85	80,95
Régió			
Közép-Magyarország	74,65	77,07	73,96
Közép-Dunántúl	88,79	83,23	80,30
Nyugat-Dunántúl	89,36	80,13	82,04
Dél-Dunántúl	87,17	85,28	78,11
Észak-Magyarország	90,38	86,17	83,51
Észak-Alföld	90,04	84,49	85,03
Dél-Alföld	89,73	85,09	84,54

Forrás: saját számítások az Életünk fordulópontjai adatfelvétel 1–4. hullámának adataiból.

Az *aktivitás* kategóriák szerinti bontás időben stabil mintát mutat: a tanuló és a munkával rendelkezők kisebb eséllyel válaszolnak, mint az inaktívak. A munkanélküliek válaszadási hajlandósága a második hullámnál az inaktívakéhoz hasonlított, a harmadik vagy negyedik hullámnál viszont a dolgozókénál is alacsonyabb lett. A legkisebb eséllyel a tanulók válaszolnak: lemorzsolódásuk különösen a negyedik hullámban vált számottevővé.

A *regionális és a település típusa* szerinti vizsgálat egyrészt alátámasztja azt a jól ismert ténytet, hogy Közép-Magyarországon, illetve Budapesten lényegesen kisebb a válaszadási arány. Ezt a különbséget már az első hullám lekérdezésekor is megfigyelték (Gyabronkáné 2003). Ami viszont meglepő, hogy bár a válaszadási arány más régiókban, illetve településeken is romló tendenciát mutat, Közép-Magyarországon és Budapesten időben állandó maradt – sőt, a 3. hullámnál még javult is. A romlás különösen a községekben feltűnő: a falusiak kezdeti 90%-os válaszadási aránya a negyedik hullámra 81%-ra csökkent.

A válaszadást befolyásoló tényezők II. Regressziós elemzés

A lemorzsolódást befolyásoló tényezők önálló hatásáról pontosabb képet kaphatunk, ha regresszióelemzéssel kiszűrjük az egyes változók közötti összefüggéseket. Három probit regresszióval modellezzük az első, a második és a harmadik hullámból a rá következő hullámban való részvétel valószínűségét. A $(j+1)$ -edik hullámban ($j=\{1,2,3\}$) való részvétel magyarázatához a j -edik hullámból vesszük a magyarázó változókat. A becslési minta most sem tartalmazza azokat, akik halálozás miatt nem vettek részt a következő felvételen.

A becslés eredményeit a 3. táblázat tartalmazza. Az első alfejezetben felsorolt várakozások és a korábbi kutatási eredmények alapján magasabb válaszadási hajlandóságot várunk a nők, az idősebbek, az iskolázottabbak, a házasok, a gyerekesek és a kisebb településeken élők esetében (Watson – Wooden 2009, Vergauwen et al. 2015). Az eredmények többé-kevésbé megfelelnek ezeknek a tapasztalatoknak. A férfiak nagyobb eséllyel morzsolódnak le, mint az amúgy hasonló ismérvekkel rendelkező nők. A válaszadási arány az életkor monoton növekvő függvénye: minél idősebb valaki, annál nagyobb eséllyel vesz részt a következő adatfelvételen. A leíró elemzésekkel ellentétben az iskolai végzettség javít a válaszadási hajlandóságon: minél iskolázottabb valaki, annál nagyobb eséllyel válaszol. A házasságban élők nagyobb eséllyel válaszolnak, mint a nőtlenek és a hajadonok; ez utóbbiakéhoz hasonlít az elváltak és az özvegyek válaszadási hajlandósága is. A tényleges párkapcsolati státusz csak a 2. és a 4. hullámnál befolyásolta szignifikánsan a válaszadást: a házastársukkal élőkhez viszonyítva a párkapcsolattal rendelkező, de partnerüktől külön élők hajlandósága volt a legalacsonyabb.

3. táblázat: A választást magyarázó probit regressziós modellek együtthatói
 Probit regression estimates of response

	1-->2. hullám		2-->3. hullám		3-->4. hullám	
Nem						
férfi	0		0		0	
nő	0,067*	(2,426)	0,135***	(4,475)	0,107**	(3,281)
Születési kohorsz						
1926–1941	0		0		0	
1942–1951	-0,012	(0,238)	0,001	(0,021)	-0,022	(0,402)
1952–1961	-0,166**	(2,831)	-0,163*	(2,544)	-0,128	(1,888)
1962–1971	-0,312***	(4,832)	-0,261***	(3,736)	-0,399***	(5,390)
1972–1983	-0,382***	(5,484)	-0,343***	(4,553)	-0,442***	(5,651)
Iskolai végzettség						
0–8 általános	0		0		0	
szakmunkásképző	0,119**	(3,131)	-0,015	(0,368)	0,059	(1,295)
érettségi	0,150***	(4,030)	0,108*	(2,566)	0,129**	(2,840)
diploma	0,183***	(3,930)	0,214***	(4,220)	0,190***	(3,629)
0–3 éves gyerekek a háztartásban						
nincs	0		0		0	
van	0,039	(0,791)	0,021	(0,389)	0,052	(0,922)
4–6 éves gyerekek a háztartásban						
nincs	0		0		0	
van	-0,089	(1,344)	0,201**	(2,802)	-0,047	(0,643)
Hivatalos családi állapot						
nőtlen/hajadon	0		0		0	
házas	-0,337**	(3,217)	-0,004	(0,035)	-0,124	(1,028)
elvált	-0,117*	(2,081)	-0,020	(0,341)	0,041	(0,686)
özvegy	-0,085	(1,278)	0,163*	(2,275)	-0,065	(0,928)
Partnerkapcsolati helyzet						
házastársával él	0		0		0	
élettársával él	-0,434***	(4,093)	0,017	(0,135)	-0,249*	(2,012)
különélő partnere van	-0,560***	(5,142)	-0,162	(1,271)	-0,389**	(2,924)
nincs állandó partnere	-0,426***	(4,254)	-0,159	(1,327)	-0,308**	(2,602)
Aktivitás						
tanuló	0		0		0	
alkalmazott	-0,143*	(2,423)	0,126	(1,596)	0,289	(1,188)
vállalkozó	-0,230**	(3,033)	-0,107	(1,144)	0,220	(0,876)
egyéb/alkalmi munka	-0,139	(1,623)	0,082	(0,602)	0,314	(1,227)
munkanélküli	-0,092	(1,175)	0,037	(0,379)	0,131	(0,523)
GYES/GYED	0,080	(0,788)	-0,011	(0,097)	0,403	(1,539)
nyugdíjas	-0,119	(1,587)	0,092	(0,987)	0,258	(1,037)
eltartott, egyéb inaktív	-0,205*	(2,453)	0,040	(0,376)	0,178	(0,691)
Településtípus						
Budapest	0		0		0	
megyei jogú város	0,067	(1,184)	0,003	(0,042)	-0,050	(0,728)
egyéb város	0,237***	(4,931)	0,078	(1,398)	-0,029	(0,483)
község	0,460***	(9,148)	0,179**	(3,037)	-0,027	(0,427)
Régió						
Közép-Magyarország	0		0		0	
Közép-Dunántúl	0,422***	(7,596)	0,211***	(1,601)	0,268***	(4,339)
Nyugat-Dunántúl	0,481***	(7,889)	0,088	(4,446)	0,344***	(5,139)
Dél-Dunántúl	0,327***	(5,692)	0,287***	(5,534)	0,213***	(3,299)
Észak-Magyarország	0,464***	(8,413)	0,322***	(4,520)	0,412***	(6,670)
Észak-Alföld	0,500***	(9,538)	0,274***	(5,980)	0,483***	(8,128)
Dél-Alföld	0,502***	(9,366)	0,303***	(4,364)	0,440***	(7,324)
Konstans	1,134***	(8,477)	0,650***	(3,998)	0,633*	(2,223)
Megfigyelések száma	15750		12033		9983	

Forrás: saját számítások az Életünk fordulópontjai adatfelvétel 1–4. hullámának adataiból.

Megjegyzések: A zárójelben szereplő számok t-statisztikák. *: $p < 0,05$; **: $p < 0,01$; ***: $p < 0,001$

A nemzetközi tapasztalatokkal ellentétben nincs bizonyítékunk arra, hogy a városokban élők a falvakban élőkénél nagyobb eséllyel morzsolódnának le. A budapestiekre jellemző válaszadási arány csak a második hullámban tér el szignifikánsan a kisebb városokra és a falvakra jellemző arányoktól. A harmadik és a negyedik hullámban viszont a településtípus változók együttthatói – egy kivételtől eltekintve – nem szignifikánsak. A regionális eltérésekre viszont van bizonyítékunk, a közép-magyarországi régióhoz viszonyítva minden más régióban magasabb a válaszadási arány.

Arra sincs komoly bizonyíték, hogy a háztartásban élő, nem iskoláskorú gyermekek, illetve az aktivitás befolyásolná a válaszadást. A gyerekek jelenlétével kapcsolatos hat együtttható közül csak egy szignifikáns, az aktivitást mérő 21 együtttható közül pedig csak kettő.

A 2. táblázatban közölt eredményekhez képest a regressziós elemzés azt sugallja, hogy a budapestiekre, illetve a tanulókra jellemző magas lemorzsolódási arány nem értelmezhető oksági hatásként. Lehetséges, hogy a budapestiek, illetve a tanulók viselkedése mögött egy összetételhatás húzódik meg: például hogy tagjai átlagosan fiatalabbak, akik eleve kisebb eséllyel válaszolnak. Természetesen más magyarázatok is elképzelhetők: például hogy a budapestiek (és a fiatalok) elfoglaltabbak, vagy bizalmatlanabbak a megkeresésekkel szemben. A napi időbeosztással és a bizalommal kapcsolatos, viszonylag megbízható adatok sajnos hiányoznak az Életünk fordulópontjai felvételből, így ezek a magyarázatok nem tesztelhetők.

AZ ÉLETÜNK FORDULÓPONTJAI ADATBÁZIS SÚLYOZÁSA

Az előző regressziós becslések lehetővé teszik az Életünk fordulópontjai adatbázis súlyozását. A módszer és az eljárás a nemzetközi gyakorlatot követi. A longitudinális súlyozásnak két fő lépése van: a szelektív nem válaszolásokat korrigáló nyers longitudinális súly elkészítése, majd a súlyozott minta kalibrálása, azaz a minta és a sokaság valamely ismérvek szerinti eloszlásában mutatkozó különbségek eltüntetése (Rendtel – Harms 2009).

A nyers longitudinális súly

A nyers longitudinális súly közvetlenül a panelkopást korrigálja. Pontosabban: a t -edik ($t=\{2,3,4\}$) hullám adataihoz rendelt nyers longitudinális súly a t -edik hullámba kerülés valószínűségének reciproka. A szóban forgó valószínűség két

másik valószínűség szorzata: (1) annak valószínűsége, hogy az egyén a $(t-1)$ hullám után részt vesz a t -edik hullámban (P_{it}); és (2) az egyén részt vesz a $t-1$ hullámban. A második valószínűség a $t-1$ hullámhoz rendelt $W_{i(t-1)}$ súly reciproka. A t -edik hullámban szereplő i egyénhez rendelendő V_{it} nyers longitudinális súlyt tehát az alábbi képlet definiálja:

$$V_{it} = W_{i(t-1)} / P_{it} \quad [1]$$

Az egyenlet nevezőjében szereplő feltételes valószínűség a következő hullámba lépést magyarázó, a 3. táblázatban feltüntetett probit becslésekből számolható ki. Ha $t=2$, akkor a $t-1$ hullámhoz rendelt súly a Szelényi (2003) által kidolgozott keresztmetszeti alapsúly. Ha viszont $t>2$, akkor a $t-1$ hullámhoz rendelt súly a kalibrált longitudinális súly.

A kalibrált longitudinális (vagy keresztmetszeti) súly

A súlyozási eljárás második lépése a kalibrálás – azaz a nyers longitudinális súly olyan módosítása, hogy a súlyozott minta bizonyos ismérvek szerinti eloszlása megegyezzen az ismert sokasági eloszlással (Rendtel – Harms 2009, Watson – Wooden 2009). A kalibrálás lényegében mátrixsúlyozás. A t -edik hullámban szereplő i egyénhez rendelendő W_{irt} kalibrált longitudinális súlyt az alábbi képlet definiálja:

$$W_{irt} = \frac{N_{rt} n_t}{N_t \sum_r (V_t)} \frac{W_{i(t-1)}}{P_{it}} \quad [2]$$

ahol r a sokaságban és a mintában egyaránt megfigyelt ismérvek kombinációi által definiált rétegeket indexeli, N_{rt} a sokasági létszám az r rétegben a t -edik hullám idején, n_t a mintanagyság (illetve a nyers longitudinális súlyok mintabeli összege), pedig a t hullámra vonatkozó nyers longitudinális súlyok összege az r rétegben (lásd például Moore et al. 2000).

A mátrixsúlyozási módszer használata miatt a kalibrált longitudinális súlyt keresztmetszeti súlynak is lehet nevezni. A kizárólag mátrixsúlyozással számolt keresztmetszeti súlyok és a kalibrált longitudinális súlyok között azonban van egy fontos különbség: adott r rétegen belül a mátrixsúlyozással készített keresztmetszeti súlyok azonosak, a longitudinális súlyok viszont eltérnek, hiszen az r rétegen belül eltérő válaszadási hajlandósággal rendelkező egyéneket találhatunk.

A kalibráláshoz szükséges adatok a népesség nem, születési év és a lakóhely régiója szerinti (háromdimenziós) eloszlása a t -edik hullám eszmei időpontját követő év (rendre 2005, 2009 és 2013) első napjára vonatkozóan. A KSH adataiból csak az 1926–1983-ban születettek relevánsak, mivel a demográfiai panel első hullámába csak ilyen születési dátumú egyének kerültek. A sokasági és a mintabeli eloszlásokat elvileg 812 rétegben kell összehasonlítani, mivel a születési év (1926–1983), a régió (7 kategória) és a nem változók kombinációinak száma: $58 \cdot 7 \cdot 2 = 812$. A szóban forgó ismérvek mindegyik kombinációjánál kiszámoltuk a nyers longitudinális súlyok összegét. Ezután a [2]-es egyenlet segítségével kiszámoltuk a kalibrált longitudinális súlyt.

A súlyozási eljárás

Észre kell vennünk, hogy az [1]-es és a [2]-es képletek rekurzívak: a t -edik hullám súlyozása megköveteli a $(t-1)$ hullám súlyozását, sőt, a $(t-1)$ hullám súlyozása a $(t-2)$ hullám súlyozását, és így tovább. Az eljárás „nulladik” lépése tehát az adatfelvétel 1. hullámának súlyozása. Ez a súly már ismert: az 1. hullám ún. keresztmetszeti alapsúlyát Szelényi (2003) készítette el. Az első tényleges feladat a 2. hullám súlyozása.

A 2. hullám súlyozása négy lépésből áll: (1) a panelkopást korrigáló tényező kiszámítása probit becslésekből; (2) a nyers longitudinális súly kiszámítása az alapsúly és az előző korrekciós tényező hányadosaként (lásd az [1]-es egyenletet); (3) a populációtól való eltérést korrigáló tényező kiszámítása; (4) a kalibrált longitudinális súly kiszámítása a szóban forgó korrekciós tényező és a nyers longitudinális súly szorzataként (lásd a [2]-es egyenletet). A (2)-es és a (4)-es lépéseknél kiszámolt nyers és kalibrált súlyokat imputáltuk, majd egy iteratív eljárás keretében normalizáltuk és cenzoráltuk. Az imputálásra azért volt szükség, mert a válaszadás előrejelzéséhez használt egyes magyarázó változóknál előfordulnak adathiányok, és ezért a súlyváltozó sem lesz adathiányoktól mentes. Az imputálás során a súlyváltozó hiányzó értékeit a számított súlyok átlagával pótoltuk. A normalizálás célja az, hogy a súlyok összege megegyezzen a megfigyelések számával. A [2]-es egyenlet eleve feltételezi, hogy a súlyok normalizáltak: hiszen a képletben a súlyok mintabeli összege helyett a mintanagyság szerepel. A cenzorálás célja annak elkerülése, hogy a súlyozott varianciabecslések túlságosan nagyok legyenek. A nemzetközi gyakorlatból ismert eljárás lényege, hogy a súlyváltozó egy adott c számnál nagyobb értékeit a c értékre kódolják át. Például a Brit Háztartáspanel súlyozásakor a küszöbérték 1,75 (Taylor et al. 2010). Vannak azonban olyan panelfelvételek is – például a National Longitudinal

Survey of Youth és a Német Társadalom-Gazdasági Panelfelvétel –, ahol nem használják a cenzorálást. Az Életünk fordulópontjai adatfelvétel 2–4. hullámaira számolt kalibrált longitudinális súlyokat a 2,5-ös küszöbértéknél cenzoráltuk.³ A cenzorálás mellékesen megszünteti a súlyok normalizáltságát: 2,5-nél nagyobb értékek 2,5-re állításával ugyanis csökken a normalizált súlyok összege. Emiatt a cenzorálást és a normalizálást egy iteratív eljárás keretében végeztük el: a tulajdonképpeni cenzorálás után a súlyt normalizáltuk, ami azzal járhat, hogy a 2,5-ös és ahhoz közeli súlyok 2,5-nél nagyobbak lesznek. A cenzorálást és a rá következő normalizálást addig folytattuk, amíg a normalizált súlyok maximuma pontosan 2,5 lett.

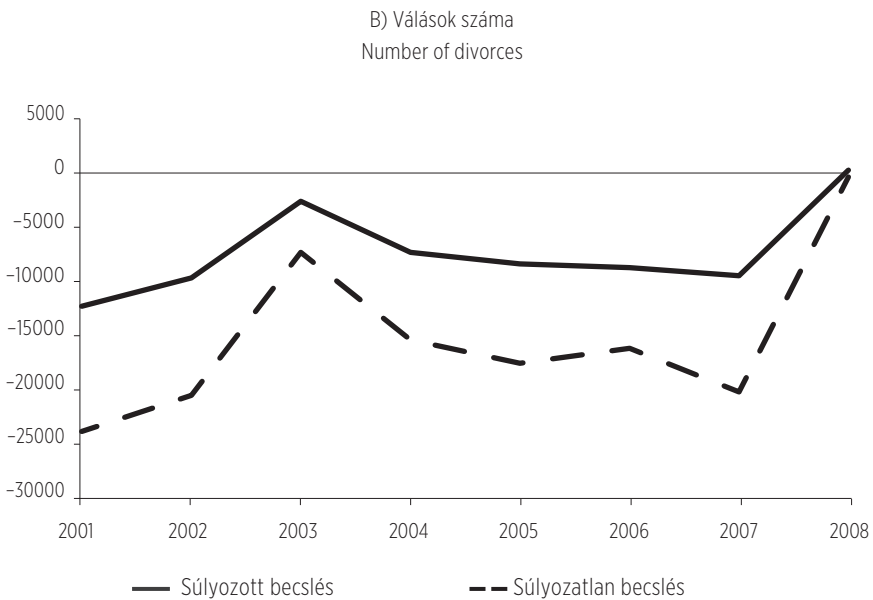
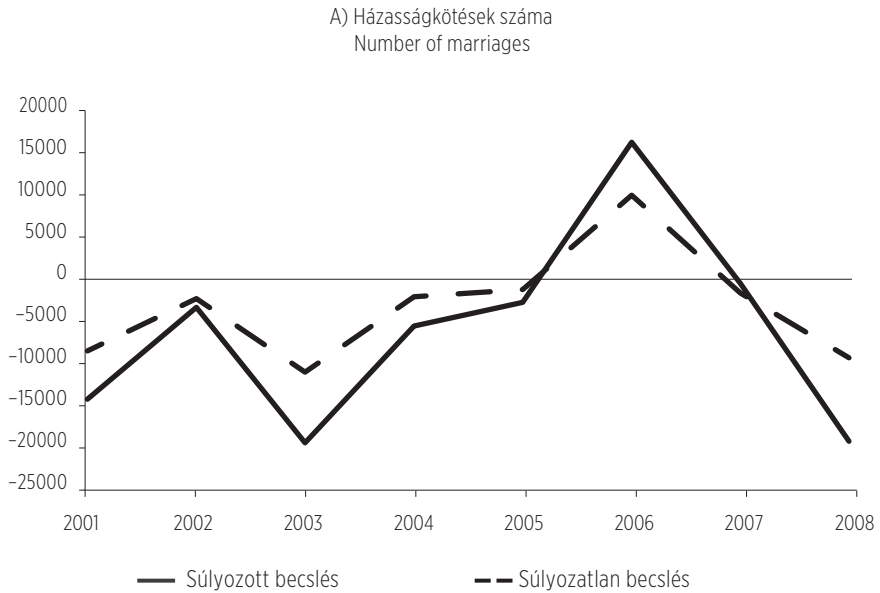
A 2. hullámra vonatkozó nyers és kalibrált longitudinális súlyok kiszámítása után az előző két bekezdésben bemutatott módszerrel készítettem el a 3. hullámra vonatkozó súlyokat. A súlyozás által megkövetelt alapsúly a 2. hullámhoz rendelt kalibrált longitudinális súly. Végül a 4. hullám adatait súlyozzuk úgy, hogy a 3. hullámhoz rendelt kalibrált súlyt tekintjük alapsúlynak.

A SÚLYOK MEGBÍZHATÓSÁGA

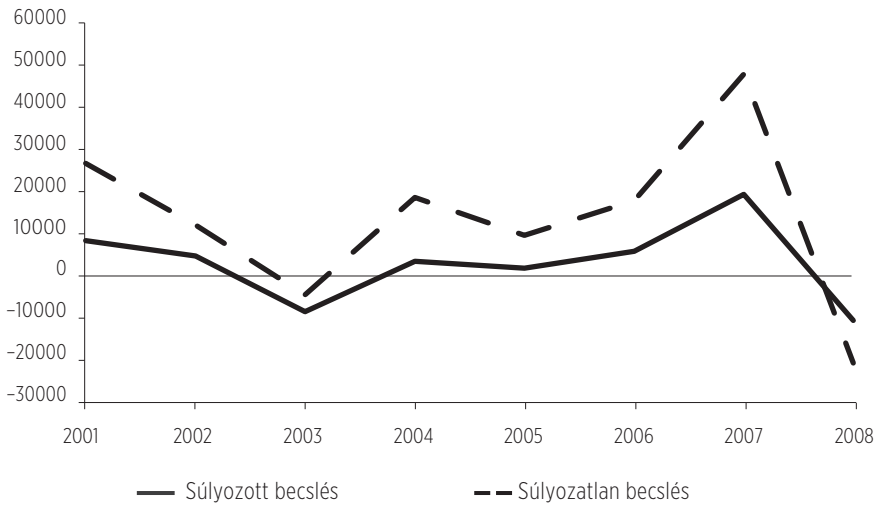
A kalibrált longitudinális súlyok használatával úgy tehetünk, mintha az Életünk fordulópontjai felvétel reprezentatív minta lenne. A reprezentativitás önmagában azonban nem szavatolja a súlyozott becslések helyességét – hiszen a súly nem biztos, hogy tökéletesen méri a mintába kerülés feltételes valószínűségét. Emlékeztetőül: a súlyozás azt feltételezi, hogy a mintába kerülést ténylegesen befolyásoló tényezők pótolhatók, felcserélhetők a nem válaszolás modellezéséhez használt változókkal. Az adatok érvényességének gyakori ellenőrzési módja a (súlyozott) becslések és a – teljes körű számbavételek vagy a nagymintás felvételekből számolt – publikált statisztikák összehasonlítása. Az Életünk fordulópontjai felvétel 2. hullámának nemzetközi kontextusa a Generations and Gender Survey (GGS) első hulláma. A nemzetközi adatfelvétel megbízhatóságát Vergauwen és munkatársai (2015) elemezték, a német mintáét Kreyenfeld és munkatársai (2010), a bolgár, grúz és magyar mintáét pedig Burkimsher (2009). A magyar adatokra is kiterjedő elemzések azt találták, hogy az Életünk fordulópontjai felvétel 2. hullámából pontos retrospektív becslések tehetők a termékenységi arányra (Burkimsher 2009, Vergauwen et al. 2015).

³ A kalibrált súly a 2–4. hullámoknál rendre 41, 233 és 404 esetben haladta meg a küszöbértéket.

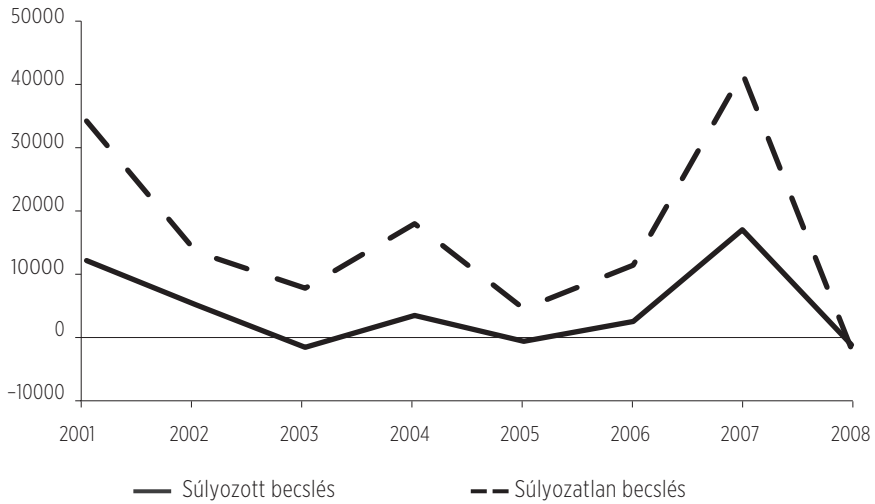
1. ábra: Demográfiai eseményekre vonatkozó retrospektív becslések torzítása
 Biases of retrospective estimates of demographic events



C) Élveszületések száma
Number of live births



D) Házasságon belüli élveszületések száma
Live births within marriage



Jelen elemzésben az Életünk fordulópontjai felvétel 4. hullámában résztvevő egyének mintáját használjuk, és a 2001–2008 közötti évekre vonatkozó retrospektív becslések megbízhatóságát vizsgáljuk. Konkrétan: a kutatás fő célkitűzéseivel összhangban a házasságkötések, a válások, a születések és a házasságon belüli születések számára vonatkozó súlyozatlan és súlyozott becsléseket hasonlítjuk össze a KSH *Demográfiai évkönyveiben* közölt adatokkal.

Az Életünk fordulópontjai 1–3. hullámának felhasználásával egy olyan eseménytörténeti adatbázist készítettünk el, amely a megtapasztalt demográfiai eseményeket és ezek dátumát tartalmazza.⁴ Az eseménytörténeti és a súlyokat tartalmazó adatokból aggregálással létrehoztam egy olyan „makroszintű” adatbázist, amely – a *Demográfiai évkönyvhöz* hasonlóan – naptári év szerinti bontásban tartalmazza egyrészt a szóban forgó demográfiai események súlyozatlan, illetve súlyozott számát, a releváns népesség létszámát, valamint a 3. hullámba került emberek súlyozatlan, illetve súlyozott létszámát. Ezután mindegyik évre először súlyozatlan, majd súlyozott becslést tettem a vizsgált demográfiai események számára. Annak érdekében, hogy a becslések összevethetők legyenek az ismert statisztikákkal, a becsléseket megszoroztam a – demográfiai évkönyvekből ismert – sokasági létszám és a mintanagyság hányadosával. A becslések és a valós adatok különbségeit – a torzításokat – az *1. ábra* mutatja.

A házasságkötésekre vonatkozó súlyozatlan és súlyozott becslések nem térnek el számottevően. A 2006-os év kivételével a becslések alacsonyabbak a valóságos adatoknál. A torzítás abszolút mértéke – ami a 10 ezres nagyságrendet csak 2001-ben és 2008-ban haladja meg valamivel – jelentősnek mondható, mivel a vizsgált időszakban évente körülbelül 40 ezer házasságkötés történt.

Még ennél is nagyobb mértékben torzítanak a válásokra vonatkozó becslések, különösen a kétezres évek elején. A torzítás a 2001-es mínusz 10 ezres nagyságrendről 2008-ra plusz ezres nagyságrendűre olvadt. A torzítás relatív mértéke azonban komoly, mivel ebben az időszakban évente durván 24 ezer volt a válások száma. A problémát a súlyok nem tudják orvosolni, a súlyozott becslések nem térnek el számottevően a súlyozatlan becslésektől. Az elváltak lemorzsolódását tehát a súlyozás nem tudja kompenzálni. Ennek valószínűleg az a legfontosabb oka, hogy a válaszdadás függ a lemorzsolódás-elemzéseknél

⁴ Az adatbázis nem tartalmazza azokat az eseményeket, amelyek bekövetkezési dátumát nem mondták meg a kérdeztettek. A 2001–2008-as évekre vonatkozóan a szülés események 5%-ánál hiányzik a dátum. A párhuzamosított adatokban bekövetkező változásoknál az adathiány becslését megnehezíti, hogy a 3. hullámban dátumokra kérdeztek rá, és emiatt nem lehet a dátumoktól függetlenül számba venni az eseményeket. Mindenesetre az adathiányok miatt a későbbi becslések nem teljesen megbízhatóak, az eseménytörténeti adatbázisból számolt becslések valamelyest alábecslik a valós demográfiai adatokat.

használt változókkal gyengén korreláló egyéb tényezőktől is. Azok a személyek például, akik két hullám között válnak el, költöznek el, vagy nem szeretnének részt venni egy olyan kutatásban, melynek egyik témája – a család – számukra kellemetlen, kényes beszédtema. Lehetséges, hogy a két hullám között bekövetkező válásokat és elköltözéseket csak gyengén magyarázzák a lemorzsolódás-elemzés során használt változók, és emiatt a súlyozott becslések nem tudnak érdemben javítani a becsléseken.

A súlyozás a születések számára vonatkozó becsléseknél bizonyult hatásosnak. A születésekre és a házasságon belüli születésekre vonatkozó súlyozatlan becslések – a 2008-as év kivételével – jelentősen felülbecslik a valóságos trendet. A 3. táblázatban feltüntetett regressziós becslések alapján a felülbecslés annak tulajdonítható, hogy a folyamatos részvétel esélye kisebb az inkább gyermektelen fiatalok között. Szintén felülbecsléshez vezet, hogy a válaszadás gyakoribb a falusiak körében.

ÖSSZEZÉS

A tanulmány az Életünk fordulópontjai követéses vizsgálatra jellemző lemorzsolódás mértékét, okait és néhány következményét vizsgálja. Az adatfelvétel 4. hullámában már csak a kiinduló minta fele vett részt. A kutatás előre haladásával a lemorzsolódás legfontosabb okává – nem meglepő módon – a nem válaszolás vált: míg a 2. hullám idején a kiinduló minta 6%-a tagadta meg az együttműködést, addig ez az arány a 4. hullámra 11%-ra emelkedett. Az adatfelvétel erőssége az, hogy az ismeretlen címre költözött vagy az otthon nem tartózkodó egyének arányát sikerült 8%-os szint alatt tartani.

A lemorzsolódásból fakadó talán legfontosabb veszélyforrás a becslések torzulása. A lehetséges torzítások számbavételét segíti a lemorzsolódást befolyásoló tényezők feltárása. A nemzetközi gyakorlatot követve a j -edik ($j=\{1,2,3\}$) hullámból vett társadalmi-demográfiai ismérvek szerint vizsgáltuk a $(j+1)$ -edik hullámra vonatkozó válaszadási arány eloszlását. Az elemzés egyrészt megerősítette azt a korábbi tapasztalatot, miszerint a fiatalabbak és a Budapesten, illetve Közép-Magyarországon élők válaszolnak kisebb eséllyel. Újszerű az az eredmény, hogy a budapestieknél tapasztalható alacsony együttműködési hajlandóság nem lakóhelyi jellemző, hanem inkább valamilyen összetétel-hatással magyarázható. Az elemzés a „középosztályosodás” tézist is megerősíti: az iskolázottabb egyének nagyobb eséllyel válaszolnak.

A lemorzsolódás feltételes függetlensége esetén az esetleges torzítások súlyozott becslésekkel minimalizálhatók. A súlyok a megfigyelt társadalmi-demográfiai ismérvek válaszadásra gyakorolt hatásait tükrözik. A tanulmány harmadik fejezete a longitudinális súlyozási eljárást dokumentálja. A súlyok megbízhatóságát a demográfiai eseményekre vonatkozó becslések sokasági adatokkal történt összevetésével teszteltük.

A megbízhatósági elemzések alapján a súlyozás sikere viszonylagos: a súlyozott becslések nem felelnek meg tökéletesen a valóságos adatoknak. Különösen a válásokra, valamint a házasságkötésekre vonatkozó becslések térnek el számottevően a valós adatokból. A súlyozás kudarca arra utal, hogy téves lehet az elemzés mögött meghúzódó feltevés, miszerint adott társadalmi-demográfiai ismérvekkel rendelkező egyének között véletlenszerű a panelkopás. A kudarc másik magyarázata, hogy a megelőző hullámból vett változókkal nem lehet pontosan megjósolni a válaszadást ténylegesen befolyásoló tényezőket. Eredményeink alátámasztják azt az eredményt, miszerint nem véletlenszerű lemorzsolódás (ill. adathiány) esetén a súlyozás képtelen javítani a becsléseken – sőt, még rosszabb becslésekhez is vezethet (Holt – Elliot 1991).

Az itt kidolgozott súlyok emiatt nem tekinthetők „omnibusz” – azaz bármilyen kutatási kérdés vizsgálatára alkalmas – súlyoknak. Egyrészt a lemorzsolódást kompenzáló súlyoknak elvileg a függő változó mintába kerülésre gyakorolt hatását kell modellezniük: tehát a súlynak igazodnia kell az elemzés céljához (Fitzgerald et al. 1998). Az osztrák GGS adatok súlyozásakor például figyelembe vették, hogy a nők – és csak a nők – között a paritás is befolyásolja a válaszadást (Buber 2010). Szintén az osztrák minta lemorzsolódás-elemzése mutatja, hogy a terhesség és a családdal kapcsolatos attitűd is befolyásolja a panelfolytatást (Buber-Ennsner 2014). Másrészt a longitudinális súly a modellezéshez használt változók és a regressziós modell specifikációjának függvénye. Alternatív modell-specifikációk értelemszerűen módosítani fogják a súlyozott becsléseket (Ayala et al. 2006). Az itt kidolgozott eljárás replikálható, a felhasználók a probit modell specifikációjának módosításával a saját elemzéseikhez jobban illő súlyokat is előállíthatnak. A súlyok alkalmazása tehát nem lehet mechanikus; a súlyozott becslések mellett a kutatóknak kísérletet kell tenniük a tartalmi regressziós modellek és a mintába kerülést modellező probit modellek együttes becslésére.

IRODALOM

- Alderman, Harold – Behrman, Jere R. – Kohler, Hans-Peter – Maluccio John A. – Watkins, Susan Cotts 2001: Attrition in Longitudinal Household Survey Data. *Demographic Research*, 5(4), 79–124.
- Ayala, Luis – Navarro, Carolina – Sastre, Mercedes 2011: Cross-country Income Mobility Comparisons under Panel Attrition. The Relevance of Weighting Schemes. *Applied Economics*, 43(25), 3459–3521.
- Baigrie, Nic – Eyal, Katherine 2014: An Evaluation of the Determinants and Implications of Panel Attrition in the National Income Dynamics Survey (2008–2010). *South African Journal of Economics*, 82(1), 39–65.
- Becker, Gary S. 1965: A theory of the allocation of time. *The Economic Journal*, 75(299), 493–517.
- Beckett, Sean – Gould, William – Lillard, Lee – Welch, Finis 1988: The Panel Study of Income Dynamics after Fourteen Years. An Evaluation. *Journal of Labor Economics*, 6(4), 472–492.
- Behr, Andreas – Bellgardt, Egon – Rendtel, Ulrich 2003: *The estimation of male earnings under panel attrition. A cross country comparison based on the European Community Household Panel*. CHINTEX Working Paper 11. <http://www.bellgardt-behr.de/paper11.pdf>
Letöltve: 2016.05.16.
- Bethlehem, Jelke G. 2002: Weighting Nonresponse Adjustments Based on Auxiliary Information. In Groves, Robert M. – Dillman, Don A. – Eltinge, John L. – Little, Roderick J. A. (eds.): *Survey Nonresponse*. Wiley, New York, 265–287.
- Buber, Isabella 2010: *Parity-Specific Weights for the Austrian Generations and Gender Survey*. VID Working Papers 4/2010. Vienna Institute of Demography, Vienna.
- Buber-Ennser, Isabella 2014: Attrition in the Austrian Generations and Gender Survey. Is there a bias by fertility-relevant aspects? *Demographic Research*, 31(16), 459–496.
- Burkimsher, Marion 2009: *Assessment of fertility indicators derived from GGP samples. Bulgaria, Hungary and Georgia*. [Unpublished manuscript]. http://drmarionb.free.fr/UnpublishedPapers/GGP_FertilityIndicatorAssessment.pdf
Letöltve: 2016. 05. 16.
- Fitzgerald, John – Gottschalk, Peter – Moffitt, Robert 1998: An Analysis of Sample Attrition in Panel Data. The Michigan Panel Study of Income Dynamics. *The Journal of Human Resources*, 33(2), 251–299.
- Groves, Robert M. – Couper, Mick P. 1998: *Nonresponse in Household Interview Surveys*. John Wiley & Sons, New York.
- Groves, Robert M. – Singer, Eleanor – Corning, Amy 2000: Leverage-Saliency Theory of Survey Participation. Description and an Illustration. *The Public Opinion Quarterly*, 64(3), 299–308.
- Gyabronkáné Németh Zsuzsanna 2003: A megíúsult kérdőívfelvételek: mértékek és okok. In Kapitány Balázs (szerk): *Módszertan és dokumentáció. Az adatfelvétel ismertetése*. Műhelytanulmányok 2. KSH Népszégtudományi Kutatóintézet, Budapest, 51–63.
- Heckman, James J. 1979: Sample Selection Bias as a Specification Error. *Econometrica*, 47(1), 153–162.

- Holt, David – Elliot, David 1991: Methods of Weighting for Unit Non-Response. *Journal of the Royal Statistical Society. Series D (The Statistician)*, 40(3), 333–342.
- Kreyenfeld, Michaela – Hornung, Anne – Kubisch, Karolin – Jaschinski, Ina 2010: *Fertility and Union Histories from German GGS Data. Some Critical Reflections*. MPIDR Working Paper 2010-023. Max Planck Institute for Demographic Research, Rostock.
- Kroh, Martin 2009: *Documentation of Sample Sizes and Panel Attrition in the German Socio Economic Panel (SOEP) (1984 until 2008)*. Data Documentation 47. Deutsches Institut für Wirtschaftsforschung, Berlin.
- KSH 2011: *Demográfiai évkönyv 2010*. Központi Statisztikai Hivatal, Budapest.
- Lillard, Lee A. – Panis, Constantijn W. A. 1998: Panel Attrition from the Panel Study of Income Dynamics. Household Income, Marital Status, and Mortality. *The Journal of Human Resources*, 33(2), 437–457.
- Little, Roderick J. A. – Rubin, Donald B. 1987: *Statistical Analysis with Missing Data*. John Wiley & Sons, New York.
- Meyer, Bruce D. – Mok, Wallace K. C. – Sullivan, James X. 2015: Household Surveys in Crisis. *The Journal of Economic Perspectives*, 29(4), 199–226.
- Moore, Whitney – Pedlow, Steven – Krishnamurty, Parvati – Wolter, Kirk 2000: *National Longitudinal Survey of Youth 1997 (NLSY97). Technical Sampling Report*. National Opinion Research Center, Chicago IL.
- Rendtel, Ulrich – Harms, Torsten 2009: Weighting and Calibration for Household Surveys. In Lynn, Peter (ed.): *Methodology of Longitudinal Surveys*. John Wiley & Sons, New York, 265–286.
- Spéder Zsolt (2003): Kutatási koncepció. In Kapitány Balázs (szerk.): *Módszertan és dokumentáció. Az adatfelvétel ismertetése*. Műhelytanulmányok 2. KSH Népszégtudományi Kutatóintézet, Budapest, 11–28.
- Szelényi Barbara 2003: Az adatbázis súlyozása. In Kapitány Balázs (szerk.): *Módszertan és dokumentáció. Az adatfelvétel ismertetése*. Műhelytanulmányok 2. KSH Népszégtudományi Kutatóintézet, Budapest, 51–63.
- Taylor, Marcia Freed – Brice, John – Buck, Nick – Prentice-Lane, Elaine (eds.) 2010: *British Household Panel Survey User Manual. Volume A. Introduction, Technical Report and Appendices*. University of Essex, Colchester.
- Vergauwen, Jorik – Wood, Jonas – De Wachter, David – Neels, Karel 2015: Quality of demographic data in GGS Wave 1. *Demographic Research*, 32(24), 723–774.
- Watson, Nicole – Wooden, Mark 2009: Identifying Factors Affecting Longitudinal Survey Response. In Lynn, Peter (ed.): *Methodology of Longitudinal Surveys*. John Wiley & Sons, New York, 155–183.
- Zabel, Jeffrey E. 1998: An Analysis of Attrition in the Panel Study of Income Dynamics and the Survey of Income and Program Participation with an Application to a Model of Labor Market Behavior. *Journal of Human Resources*, 33(2), 479–506.

ATTRITION IN AND WEIGHTS FOR THE HUNGARIAN GENERATIONS AND GENDER SURVEY

ABSTRACT

The Hungarian Generations and Gender Survey (Turning Points of the Life Course) experienced 50 percent of sample loss between its initial 2001 wave and 2014, when the fourth wave was registered. The present paper (1) examines the socio-demographic correlates of panel attrition; (2) develops weights which can be used to minimize potential attrition bias, and (3) tests the reliability of the weights by comparing weighted estimates on demographic events to known population quantities. Panel attrition is found to be correlated with socio-demographic characteristics including age, sex, marital status, education and place of residence. The weights have only a mild effect on improving sample estimates on the number of live births, but cannot reduce the bias in the estimated number of marriages and divorces. It is concluded that the weights reflecting the impact of socio-demographic characteristics on response do not guarantee the reliability of sample estimates.