

FÜGGELÉK

MOLNÁR GYÖRGY

AZ ADATÁLLOMÁNY ÉS A ROTÁCIÓS PANEL



## 1. A háztartási költségvetési felvétel néhány jellegzetessége

A KSH minden évben valamivel több mint tízezer háztartás részvételével készíti el a háztartási költségvetési felvételt (HKF), amelynek összefoglaló eredményeit a *Családi költségvetés adattár* (2000-től Háztartás-statisztikai Évkönyv) című, évente megjelenő kiadványsorozatban publikálja. Az adatfelvétel módszertanát ezek a kötetek tartalmazzák, itt nem térünk ki részletes ismertetésükre. A teljes HKF adatbázisán alapuló számításokra a kötet mindkét tanulmányában a *teljes minta* kifejezéssel hivatkozunk, megkülönböztetve ezeket a panelrészmintán alapuló számításoktól.

Az adatfelvétel részben havi naplóvezetésen, részben éves kikérdezésen alapul. Az éves kikérdezést évente változóan egy vagy két alkalommal, a következő év első negyedében, az adóbevallások idején bonyolítják le. Ennek során valamennyi jövedelmi kategóriában, valamint az élelmiszerek és az élvezeti cikkek kivételével a fontosabb és/vagy nagyobb kiadási tételek esetében is megkérdezik az egész évre vonatkozó értékeket. Ha az éves kikérdezést két részletben bonyolítják le, akkor különválasztják a kiadásra és a jövedelemre vonatkozó kérdéseket. Amennyiben a háztartás részt vett a havi naplóvezetésben, de egyik vagy mindkét éves kikérdezés valamilyen okból meghíúsult, akkor a KSH munkatársai a havi adatok és a többi háztartás átlagának figyelembevételével imputálják a megfelelő éves értékeket. A teljes mintában általában 6-6,5 százalék, 2001-ben valamivel több, mint 8 százalék volt az ilyen módon imputált háztartások aránya.

Ez az eljárás véleményünk szerint erősen vitatható. Elemzésünk azt mutatta, hogy az ilyen módon imputált háztartások döntő többsége vagy az első, vagy a tizedik jövedelmi tizedbe került. Nagyon sok esetben olyan havi adatfelvétel adatait szorozták fel 12-vel, amelyben lényegében nem volt jövedelem (ami erősen valószínűsíti, hogy már a havi adatközlés, naplózás sem volt megfelelő). Más esetekben viszont egy-egy kiugró jövedelmi tétel (például a decemberi jutalmak) lett 12-vel szorozva, ami kiugróan magas jövedelmet eredményezett.

Elképzelhető, hogy ez az eljárás a publikált átlagos jövedelmi és kiadási értékeket talán kevésbé torzítja. A meghíúsult megfigyelések figyelembevétele azonban a mi elemzési céljainknak egyáltalán nem felel meg (különösen a panel esetében), ezért a teljes mintából is elhagytuk azokat. A meghíúsult megfigyelések elhagyása után 1998-tól kezdődően minden évben mintegy 9500 háztartás található a teljes mintákban.

A szokásos adatimputálás egy másik esete, amikor az éves kikérdezés nem hiúsult meg, de a jövedelemként megjelölt összeg nem felel meg a törvényi szabályozásnak (például a nyugdíj alacsonyabb a nyugdíjminimumnál). A KSH imputálási gyakorlata ebben az esetben az, hogy a törvényi minimumot nem elérő jövedelmi tételeket a minimális szintre emelik. A nyugdíjak esetében ez az imputálási gyakor-

lat egyes években azt jelentette, hogy a rokkantsági és özvegyi nyugdíjakat is az öregségi nyugdíjminimum szintjére emelték, holott azok értéke nem feltétlenül kell, hogy elérje a nyugdíjminimumot.

Részletesen megvizsgáltuk azoknak a jövedelmi tételeknek az esetét, amikor a törvényben meghatározott érték bizonyos paraméterek függvényében egyértelműen rögzített (például családi pótlék), vagy a minimális mellett maximális értéke is van (például a munkanélküli-támogatások különböző formái). Egyértelműen megállapítható volt, hogy ezekben az esetekben egyformán valószínű mindkét irányú eltérés. Emiatt a szokásos imputálási gyakorlat torzítja az eredeti adatfelvételt, és csökkenti az egyenlőtlenségeket.

Az éves kikérdezéskor meghiúsult megfigyelések elhagyásából és a KSH gyakorlatához képest sokkal kisebb mértékű és óvatosabb imputálásból adódik, hogy a teljes minta alapján elvégzett számításaink kismértékben különböznek a KSH által publikált adatoktól.

A tanulmányok során végig a háztartások *nettó jövedelmével* számolunk. Az általunk alkalmazott jövedelemfogalom tartalmazza a saját termelésből történő fogyasztás értékét és a mezőgazdasági bevételek és kiadások egyenlegét is. A jövedelem képzésekor figyelembe vettük a háztartáson kívülről kapott, illetve háztartáson kívülre adott támogatások egyenlegét is. A KSH szokásos fogalomhasználatától eltérően jövedelemként vettük számba az ingó és ingatlan vagyontárgyak eladásából, a szociálpolitikai kedvezményből, a kárpótlási jegy hasznosításából, valamint a kapott vagyon eladásából származó jövedelmet. Ennek fő oka az volt, hogy lehetőség szerint elkerüljük a negatív jövedelmeket. Ezt a célt néhány mezőgazdasági vállalkozó esetében nem sikerült megvalósítani. Évenként egy-három háztartás esetében előfordult, hogy a mezőgazdasági kiadások meghaladták nemcsak a mezőgazdasági bevételeket, hanem a háztartás összes jövedelmét is. Ezekben az esetekben a háztartások hitelből vagy korábbi megtakarításaikból finanszírozták megélhetésüket, ezek a tételek azonban nem részei a jövedelemnek. Azoknak az egyenlőtlenségi mutatóknak az esetében, amikor nem lehet negatív jövedelemmel számolni, az érintett háztartásokat elhagytuk a mintából.

A kiadási adatok értelmezésekor, különösen a kiadási mobilitás esetében, azt is figyelembe kell vennünk, hogy számos kiadási tételre vonatkozóan csak a havi értékeket ismerjük. Az éves adat előállításakor ezt szorozzuk 12-vel. Ez az eljárás alkalmanként jelentős, a valóságot nem tükröző ingadozást vihet az adatokba. Példaként említhető, hogy december hónapban mind az élelmiszer-, mind pedig az élvezetecikk-kiadások jellemzően magasabbak. A panelben szereplő háztartásokat viszont nem feltétlenül ugyanabban a hónapban kérdezték meg minden évben.

1993 előtt a HKF előállítása során a jövedelmeket csak a havi (illetve akkor többnyire két hónapig tartó) naplóvezetés során kérdezték, az éves kikérdezés csak a nagyobb kiadási tételekre vonatkozott. Ennek a ténynek fontos következménye van az 1993 előtti és utáni HKF-adatok összehasonlíthatóságát illetően. Egy rövidebb időszakot tartalmazó hosszabb időszakon ugyanis a szokásosan alkalmazott egyenlőtlenségi mérőszámok szükségképpen alacsonyabb értéket szolgáltatnak, mint a rövidebb időszakon. Tehát az éves adatokon alapuló egyenlőtlenség

mindig kisebb, mint az adott éven belüli havi adatokon alapuló. (Elméletileg egyenlőség is fennállhat, mégpedig abban az esetben, ha az éves és a havi jövedelmek hányadosa minden egyes háztartás esetében pontosan ugyanakkora. A valóságban ez nyilván nem fordul elő.) A tétel bizonyítása megtalálható például *Shorrocks* [1978]-ban.

Ez azt jelenti, hogy *az 1993 előtti és az 1993-tól kezdődő állományok jövedelmi adatai csak erős fenntartásokkal hasonlíthatók össze* (mi nem is tesszük ebben a kötetben). Ez a megjegyzés különösen fontos a tranzíció előtti és utáni jövedelemegyenlőtlenségek esetleges összehasonlítása esetén. *Az 1993 előtti jövedelemegyenlőtlenségek felülbecsültek a későbbiekhez képest, ennek következtében az egyenlőtlenség növekedése alábecsült.*

## 2. A rotációs panelek előállítása

A KSH mintavételi eljárásából adódóan elvileg minden háztartás három évig marad a HKF-mintában. Ebből az következik, hogy a mintából bármelyik három évre vonatkozóan előállítható egy, a teljes minta harmadát kitevő rotációs panel. A rotációs elnevezés abból ered, hogy minden évben a minta harmada kikerül, tehát két egymást követő – két közös évet tartalmazó – hároméves panel kétharmada ugyanazokból a háztartásokból áll. A gyakorlatban – az adatfelvételben való ismételt részvétel nagyarányú megtagadása miatt – a panelek mérete jóval kisebb az elméleti értéknél. A HKF-ből nyerhető panelekkel a KSH sohasem foglalkozott, mi állítottuk elő azokat.

A rotációs panelek értelmezéséhez még egy megjegyzést kell tenni. A HKF mintavételének alapja a lakás. Ha tehát valamelyik háztartás – vagy annak egy része – időközben elköltözik, akkor az automatikusan kikerül a mintából, nem követik a velük történeteket. Ebben az értelemben tehát csak kvázipanelről beszélhetünk.

A kötet első tanulmányában felhasználtuk az 1993–1995 és az 1996–1998 közötti időszakokra vonatkozóan korábban előállított rotációs paneleket is.<sup>1</sup> 1995 és 1996 között a teljes mintakeretet kicserélték, ezért nincs a két hároméves panelciklust átfedő panel.

Eredeti elképzelésünk az volt, hogy ezt követően három rotációs panelt állítunk: az 1997–1999, 1998–2000 és 1999–2001 évi három-három éves időszakokra vonatkozóan. Sajnos, az 1997–1999 közötti időszakban olyan nagy mértékű volt az ismétlődő lekérdezések megtagadása, hogy a mindhárom évi HKF háztartásai között szereplő, egyszer sem megghiúsult háztartási kikérdezések száma mindössze 1236 volt. Ezen belül mindössze 64 darab (5,2 százalék) volt a budapesti háztartások száma. A panel kis mérete, ezen belül különösen Budapest rendkívül alacsony aránya lehetlenné tette a minta reprezentatívá tételét. Ezért erre az időszakra vonatkozóan nem végeztük el a panelvizsgálatokat. Szerencsére a korábbi, 1996–1998 közötti panel lefedte ennek az időszaknak az elejét, így nem képződött „folytonossági hiány” az egyes évek között, aminek a mobilitási elemzések szempontjából van nagy jelentősége.

Az ismétlődő lekérdezések sikertelenségének oka kettős. Az egyik egy speciális, 1997-re vonatkozó helyzet. Ebben az évben a szokásosnál kisebb volt a teljes minta mérete: 7560 háztartás, amelyből mintegy 7000 volt a nem megghiúsultak száma. A mintán belül mindössze 9 százalék volt Budapest aránya.

A másik ok általános, ami ismételten felveti a KSH által alkalmazott mintavételi eljárás problémáit, amelyekről az előbbiekben hivatkoztam tanulmányokban is

<sup>1</sup> Ezek előállításáról lásd *Kapitány-Keszthelyiné-Molnár* [1999], *Kapitány-Molnár* [2001], [2002], *Molnár* [2002].

írtunk. A teljes minta méretének biztosítása érdekében a KSH természetesen alkalmazza a pótcímek rendszerét, nincs viszont külön pótcímlista az előző vagy előző két évben, az adatfelvételben részt vett háztartások körére. Vagyis először eleve elhagyják a minta harmadát, és csak utána derül ki, hogy a mintában bentmaradók vállalják-e az adatszolgáltatást a következő évben is. Ez azt jelenti, hogy ha egy ilyen háztartás megtagadja a részvételt a további lekérdezésben (vagy egyszerűen csak nem találták otthon), akkor helyére olyan háztartás kerül, amelyik első ízben került be a mintába. Így fordulhat elő, hogy az elvileg egyharmados panelminta általában csak egyhatod körül mozog.

Ezen a helyzeten könnyen lehetne segíteni, ha volna két speciális, rétegzett pótcímlista a második, illetve a harmadik éve a mintában lévő háztartások körére is, vagyis nem hagynák el eleve a minta harmadát. (A rétegzésre azért van szükség, mert egyes csoportok – például a viszonylag frissen nyugdíjazott tagot tartalmazó háztartások – egyébként még a jelenleginél is túlreprezentáltabbá válnának a mintában.) Ez a mintavételi korrekció nem csupán a panel előállítására, hanem a mintavételi hiba csökkentése miatt is fontos volna. Az egyharmados rotációs arány ugyanis ideálisnak volna tekinthető a statisztikai hiba csökkentése, az egymást követő évek adatainak összhangja szempontjából is.

Külön ki kell emelni a Budapesttel kapcsolatos nehézségeket is. 1998-tól kezdve a teljes mintában Budapest részaránya többé-kevésbé megfelel a lakosságarányának, a panelben azonban már jóval kisebb a főváros súlya. A budapesti adatok minősége is jóval alatta marad az átlagosnak. Itt sokkal nagyobb az adattisztítási szükséglet, jóval több az adatokon belüli belső ellentmondás. Ez a panel összeállítása során válik különösen egyértelművé, amikor egyes adatoknak (például a születés időpontja) értelemszerűen változatlanoknak kellene maradniuk az időben, vagy nem csökkenhetnének (mint a legmagasabb iskolai végzettség).

A vizsgált időszak másik két rotációs paneljének mérete elfogadható volt. Az 1998–2000-es panel 1891 háztartásból áll, melyek taglétszáma 1998-ban 4887, 1999-ben 4776, 2000-ben pedig 4673 személy volt. Az 1999–2001-es panelben 1732 háztartás van, közülük 1611 töltötte ki a kiegészítő kérdőívet is. A személyek száma évenként rendre: 4421, 4334 és 4285.

A konzisztencia érdekében a *súlyozást* az eredetileg tervezett három panelciklus közös évére, 1999-re végeztük el. Ezt a megoldást az 1997–1999-es panel kiesése után is megtartottuk. A súlyozáshoz az iteratív skálázás módszerét alkalmaztuk (lásd *Darroch–Ratcliff* [1972]). Az eljárás legfontosabb előnyei közé tartozik, hogy megőrzi az eredeti mintasúlyoknak azt a tulajdonságát, hogy az egyes háztartások minden tagjának azonos a súlya. További előnye, hogy eleve biztosítja a súlyok nem negativitását, és lehetőséget ad azok ésszerű keretek között tartására is. Az eljárás a következő feladatot oldja meg:

$$\min \sum_{j=1}^n [w_j (\log w_j - \log w_j^0) - (w_j - w_j^0)],$$

feltéve a  $\sum_{j=1}^n w_j q_{ij} = c_i$ ,  $i = 1, 2, \dots, m$  kalibrálási egyenletek és az esetleges

$l \leq w_j \leq u$ ,  $j = 1, 2, \dots, n$  korlátok teljesülését, ahol

- $n$  a háztartások száma a mintában;
- $w_j^0$  és  $w_j$  az eredeti mintasúlyok, illetve a kalibrált súlyok;
- $q_{ij}$  az  $i$ -edik kontrollváltozó értéke a  $j$ -edik háztartásban;
- $c_i$  az  $i$ -edik kontrollváltozó értéke a teljes népességben;
- $l$  és  $u$  a kalibrált súlyokra adott alsó, illetve felső korlát.

Tekintettel a panelek viszonylag kis méretére és Budapest alacsony arányára a korábbiaktól eltérően nem alkalmaztunk Magyarország régiói szerinti területi kalibrálást.

A következő tényadatokra kalibráltunk:

- nőkre és férfiakra külön-külön a [0–19], [20–34], [35–54] és [55–] éves korosztályok létszáma;
- alkalmazottak (beleértve az alkalmi munkából élőket) száma;
- vállalkozók és segítő családtagjaik száma;
- nyugdíjasok száma;
- gyesen, gyeden, gyeten, szülési szabadságon lévők száma;
- tanulók száma;
- nem tanuló gyerekek száma;
- a teljes népesség létszáma megyei jogú városban, illetve egyéb településen élők bontásában;
- az aktív keresők (alkalmazottak és vállalkozók együtt) száma megyei jogú városban, illetve egyéb településen élők, valamint háromfokozatú iskolai végzettség szerinti bontásban;
- az összes, az egyszemélyes és a gyermekes háztartások száma.

A fenti felsorolásban megyei jogú városon Budapestet, a megyeszékhelyeket és Dunaújvárost, Hódmezővásárhelyt, Nagykanizsát és Sopront értjük. A háromfokozatú iskolai végzettség: középfokú végzettség alattiak, középfokú és felsőfokú végzettségűek.

A súlyozási eljárás során előre megadott külső korlátokat használtunk annak érdekében, hogy néhány extrém esetben ne kapjunk túlságosan kicsiny vagy túlságosan nagy súlyokat. Mindkét panel esetében 600 volt az alsó és 12000 a felső korlát. Mindkettőt csak néhány háztartás éri el.

### *A súlyozás során alkalmazott kontrollváltozók<sup>2</sup>*

Háztartásokban élő népesség: férfi–nő, négy korcsoport bontásban

Férfi	0–19 éves	1 235 324
Férfi	20–34 éves	1 140 755
Férfi	35–54 éves	1 381 483
Férfi	55– éves	1 022 797

<sup>2</sup> A sarokszámok előállításáért *Keszthelyiné Rédei Mária*nak és *Soukup István*nak, a KSH munkatársaik tartozunk köszönettel.



Nő	0–19 éves	1 178 601
Nő	20–34 éves	1 093 996
Nő	35–54 éves	1 442 111
Nő	55– éves	1 518 790
Összesen		10 013 857

Háztartásokban élő népesség: nagyváros, egyéb település bontásban

Nagyváros	3 871 814
Egyéb település	6 142 043
Összesen	10 013 857

Aktív keresők: iskolai végzettség és településtípus szerinti bontásban

Megnevezés	Nagyváros	Egyéb település	Összesen
Legfeljebb alapfokú végz.	548 154	1 270 207	1 818 361
Legfeljebb középfokú végz.	572 614	630 364	1 202 977
Magasabb végzettség	398 092	238 334	636 426
Összesen	1 518 850	2 138 905	3 657 765

#### Aktivitás

Alkalmazásban álló + szövetkezeti tag + alkalmi munkás	3 157 798
Vállalkozó + segítő családtag	499 967
Nyugdíjas	2 859 813
Gyes, gyed, gyet, szülési szabadságon lévő	289 981
Tanuló	1 658 892
Nem tanuló gyerek	766 950

#### Háztartásadatok

Egytagú háztartások	900 303
Gyerekes háztartások	1 340 451
Összes háztartás	3 766 109