

KÉZDI GÁBOR

Az önbevalláson alapuló kereseti adatok érvényessége

A tanulmány a kérdőíves felvételekben a megkérdezett személyek által megjelölt és a valóságos keresetek közötti összefüggéseket vizsgálja. A téma igen fontossá vált napjainkra, amikor nem egy háztartási felvétel kizárólag az aktív keresők közvetlen megkeresésével próbálja a kereseteket felmérni. A rendszerváltással az információk elaprózottsága és adatvédelmi problémák miatt csak nagy költségekkel és magas meghíúsulási arányokkal lehet olyan felvételt készíteni, amelynek céljai között együtt szerepel a családi háttér és a megbízható, munkáltatói információkon nyugvó keresetek felmérése.*

A problémát a KSH 1988. évi jövedelemfelvétele alapján vizsgálom,¹ amelyben megfelelő részletességgel megtalálhatók mind a vállalattól, mind a megkérdezettől származó kereseti adatok. Minden kereset 1987-re vonatkozik. Az elemzés csak a főmunkaidőben foglalkoztatott alkalmazottakra terjed ki. Nyilvánvaló, hogy a rendszerváltás előtti adatok alapján csak igen korlátozottan érvényes válaszokat adhatunk a jelenre vonatkozóan, frissebb adatforrás azonban nem áll rendelkezésre. A minden bizonnyal érdekes összehasonlításokkal szolgáló korábbi jövedelemfelvételeket pedig azért nem használtam e célra, mert a rendelkezésre álló állományok részletes kereseti komponensek helyett összevont – vállalati adatok alapján számított, de egyes esetekben önbevallás alapján korrigált – adatok szerepeltek.

Részben az időbeli távolság, részben a problémának a mindenkori kérdezési szituációtól függő természete miatt elkerülöm a direkt becslés kérdését. Nem azt vizsgálom tehát, hogy hogyan lehet közvetlenül megbecsülni az egyes személyek valós keresetét az általuk megjelölt kereset alapján, hiszen ez az út nyilvánvalóan nagymértékű bizonytalanságokkal és torzító tényezőkkel lenne terhelt. A megközelítésem középpontjában ezért sokkal inkább hosszabb távú megállapítások állnak, vagyis megkísérlem feltárni a kétféle kereseti adat összefüggéseinek részleteit is. Két, egymással összefüggő kérdést vizsgállok. 1. Vajon miben tér el a kétféle forrásból származó keresetek eloszlása? 2. Vajon mitől függ, hogy a kérdőíves adatfelvétel szituációjában az emberek keresetük mekkora részét jelölik meg?

A továbbiakban *önbevalláson* alapuló kereseteknek a megkérdezett által megjelölt komponensek alapján számított kereseteket, *valós* kereseteknek pedig a vállalati információkon nyugvó kereseteket nevezem. Az előbbi tartalmazza a főmunkahelyi átlagos havi

* A tanulmány A kereseti egyenlőtlenségek alakulása Magyarországon 19972–1996 című F 20686 OTKA-kutatás keretében készült. Köszönetet illeti *Kőrösi Gábort* segítő megjegyzéseiért, az esetleges hibák természetesen a szerzőt terhelik.

¹ Lásd KSH kiadvány. A KSH jövedelemfelvételeit a Tárki-adatbank bocsátotta rendelkezésünkre.

keresetet, a munkáltatótól kapott éves jövedelem ledolgozott hónapokra vetített részét, valamint az egyéb tevékenységből, másodállásból stb. származó bevallott havi jövedelmeket. A vállalati keresetek az 1987 folyamán folyósított kereseti jellegű tételek ledolgozott hónapokra vetített hányadának összegeként számíthatók.²

Önbevallás és valós keresetek

Az 1. táblázat tartalmazza az önbevalláson alapuló és a valós keresetek eloszlásainak főbb paramétereit.

1. táblázat

A valós és az önbevallás útján kapott havi kereset eloszlásának főbb paraméterei*
(teljes munkaidőben foglalkoztatott alkalmazottak, mintanagyság: 17 687 fő)

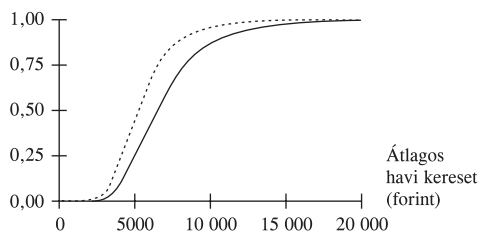
Megnevezés	Átlag	Szórás	Relatív szórás	Medián	Percentilis arányok				Gini
					75/25	90/10	95/5	99/1	
Valós keresetek	7360	3504	0,476	6651	1,68	2,75	3,76	7,13	0,233
Önbevallás	5780	2474	0,428	5300	1,60	2,43	3,28	5,60	0,207
Önbevallás/valós	0,79	0,71	0,90	0,80	0,95	0,88	0,87	0,79	0,89

*Az egyén munka jellegű fizetési komponensekből számított éves keresetének a ledolgozott hónapokra vetített átlaga. A KSH mintavételi súlyaival súlyozott számítások.

1987-ben az önbevallással kapott keresetek átlagosan a vállalati információkon alapuló keresetek 79 százalékát tették ki. Az alacsonyabb átlag mellett az egyenlőtlenség is kisebb volt: az önbevallás esetében 48 helyett 43 százalékos volt a relatív szórás, s a 0,233-es Gini-féle együttható értéke is 0,207-re csökkent. Különösen az eloszlás szélein jelentős az önbevallás "összehúzó" hatása: míg a harmadik és az első kvartilis keresetek aránya mindössze 5 százalékkal marad el a valós eloszlás megfelelő hányadosától, a tizedik és az első decilis hányadosánál az eltérés már 12, a kilencvenkilencedik és első percentilisek esetében pedig 21 százalékos. Amint az 1. és 2. ábrán látható eloszlás- és sűrűségfüggvények tanúsítják, az önbevallás nemcsak csúcsosabb, de jobbra kevésbé elnyúló eloszlást eredményez. Az említett "összehúzó" hatás tehát elsősorban a magas keresetűek alacsonyabb bevallási arányából adódik.

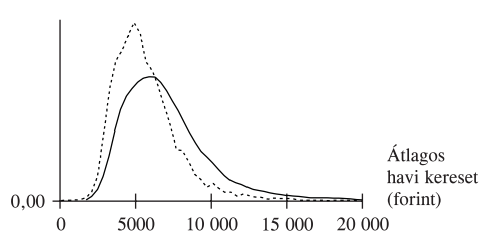
1. ábra

Az önbevalláson alapuló (- -) és a valós (-) keresetek eloszlásfüggvénye



2. ábra

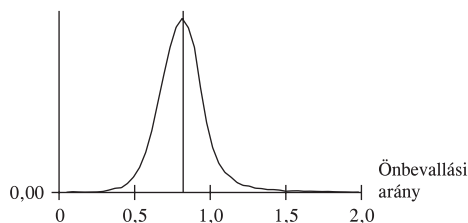
Az önbevalláson alapuló (- -) és a valós (-) keresetek sűrűségfüggvénye



² A keresetek számításáról részletesebben lásd Ábrahám Árpád-Kézdi Gábor: Foglalkoztatás és kereset panelszerkezetben. Munkaerőpiaci műhelytanulmányok, Budapest, MTA KTI és BKE Emberi erőforrás tan-
szék, 1998, megjelenés alatt.

A kétféle kereseti adat közvetlen összehasonlításához definiáljuk az *önbevallási arányt*, vagyis a megkérdezett személy és a vállalat által szolgáltatott információkon alapuló kereseti adatok hányadosát. Sűrűségfüggvényét a 3. ábra mutatja.

3. ábra
Az önbevallási arány sűrűségfüggvénye



A teljes mintában az önbevallási arány szórása 0,23, relatív szórása 0,28. Igen érdekes, hogy az eloszlás viszonylag szimmetrikus és meglehetősen csúcsos (kurtózisa 400). A változó szinte kizárólag a 0,5–1,5 intervallumra korlátozódik, ez az intervallum már a 2. és a 99. percentiliseket is tartalmazza.

Mint az ábrán jól látható, egyes esetekben teljesen abszurd önbevallási arányok fordulnak elő. Bár ezek az adatok is hordoznak információt, a további elemzések szempontjából hihetetlennek (*outlierek*) minősítettem minden olyan esetet, ahol az önbevallás meghaladta a vállalati adat kétszeresét, vagy nem érte el annak felét. Ezeket az eseteket (424 egyén, a minta 2,4 százaléka) tehát kizártam a mintából.³

A keresetek önbevallási hányadának meghatározói

Ebben a fejezetben az önbevallási arány eltérései mögötti tényezőket igyekszem feltárni. Megközelítésem alapvetően empirikus: a rendelkezésre álló változók magyarázó erejének maximalizálására törekszem. A feltárt kapcsolatok irányán és hozzávetőleges erősségén túl a hatások konkrét mértékét és a függvények formáit nem magyarázom. A predikciós erő és belső konzisztencia alapján legeredményesebbnek ítélt regressziós modellben az önbevallási hányadot magyarázó változóként a valós (vállalati forrású) kereset és a munkaerő-piaci tapasztalat szerepel, a nem és az iskolai végzettség kategóriái szerinti teljes interakcióban. A valós keresetet hatását negyedfokú polinommal specifikálom. A durván az életkor és a legutolsó iskola lehetséges legkorábbi befejezése közti idővel közelíthető munkaerő-piaci tapasztalat hatását másodfokú (férfiaknál harmadfokú) polinom segítségével lehetett optimálisan becsülni. Egyetlen más háttérváltozóként a főmunkahelyen kívüli keresetek meglétét jelző dummy változó szerepel a modellben, külön férfiakra és nőkre.

A modellilleszkedés kielégítő, a 27,3 százalékos R^2 17 ezres egyéni keresztmetszeti mintában inkább jónak mondható. A kihagyott változóval kapcsolatban elhanyagolható problémát jelez a Ramsey-féle RESET teszt 3,7-es F értéke ($p=0,025$). A becslés hibatagjai heteroszkedasztikusak voltak, a Cook-Weisberg-féle χ^2 próba robusztus, 1861-es értéket ($p=0,000$) eredményezett, s a heteroszkedaszticitás ténye jól kimutatható. A pa-

³ A vállalati források alapján számított keresetek közötti outlierek kiszűrését *Ábrahám-Kézdi* i.m. részletezi.

2. táblázat
Az önbevallási arányt magyarázó modell becslése

Megnevezés	Férfiak			Nők		
	paraméter	SE*	p*	paraméter	SE*	p*
0–7 általános konstans	2,16849	0,41119	0,000	2,28466	0,41331	0,000
8 általános konstans	1,51854	0,07766	0,000	1,62554	0,13343	0,000
Szaktanulmányok konstans	1,87811	0,08473	0,000	2,55944	0,30846	0,000
Középfok konstans	1,62453	0,12361	0,000	1,53855	0,07682	0,000
Felsőfok konstans	1,28592	0,07425	0,000	1,66743	0,16844	0,000
Kereset 0–7 általános	–0,00065	0,00022	0,003	–0,00083	0,00026	0,002
Kereset 8 általános	–0,00022	0,00003	0,000	–0,00036	0,00008	0,000
Kereset szaktanulmányok	–0,00035	0,00003	0,000	–0,00091	0,00020	0,000
Kereset középfok	–0,00025	0,00005	0,000	–0,00025	0,00003	0,000
Kereset felsőfok	–0,00011	0,00002	0,000	–0,00027	0,00006	0,000
Kereset ² 0–7 általános	1,05E–07	0,00000	0,012	1,66E–07	0,00000	0,004
Kereset ² 8 általános	1,78E–08	0,00000	0,000	5,62E–08	0,00000	0,000
Kereset ² szaktanulmányok	3,42E–08	0,00000	0,000	1,71E–07	0,00000	0,000
Kereset ² középfok	2,12E–08	0,00000	0,000	2,62E–08	0,00000	0,000
Kereset ² felsőfok	6,13E–09	0,00000	0,000	2,82E–08	0,00000	0,000
Kereset ³ 0–7 általános	–7,53E–12	0,00000	0,024	–1,47E–11	0,00000	0,005
Kereset ³ 8 általános	–6,05E–13	0,00000	0,000	–3,99E–12	0,00000	0,000
Kereset ³ szaktanulmányok	–1,39E–12	0,00000	0,000	–1,44E–11	0,00000	0,002
Kereset ³ középfok	–7,78E–13	0,00000	0,006	–1,18E–12	0,00000	0,000
Kereset ³ felsőfok	–1,42E–13	0,00000	0,002	–1,31E–12	0,00000	0,004
Kereset ⁴ 0–7 általános	1,97E–16	0,00000	0,036	4,65E–16	0,00000	0,006
Kereset ⁴ 8 általános	7,01E–18	0,00000	0,000	1,03E–16	0,00000	0,001
Kereset ⁴ szaktanulmányok	1,97E–17	0,00000	0,000	4,43E–16	0,00000	0,005
Kereset ⁴ középfok	1,03E–17	0,00000	0,037	1,83E–17	0,00000	0,000
Kereset ⁴ felsőfok	1,17E–18	0,00000	0,012	2,20E–17	0,00000	0,008
Tapasztalat 0–7 általános	0,01375	0,00830	0,097	0,00486	0,00307	0,114
Tapasztalat 8 általános	0,01655	0,00384	0,000	0,00183	0,00113	0,107
Tapasztalat szaktanulmányok	0,02561	0,00291	0,000	0,00559	0,00189	0,003
Tapasztalat középfok	0,02857	0,00393	0,000	0,00823	0,00131	0,000
Tapasztalat felsőfok	0,02430	0,00556	0,000	0,00950	0,00202	0,000
Tapasztalat ² 0–7 általános	–0,00036	0,00037	0,323	–0,00008	0,00006	0,179
Tapasztalat ² 8 általános	–0,00052	0,00017	0,002	–0,00001	0,00002	0,667
Tapasztalat ² szaktanulmányok	–0,00111	0,00017	0,000	–0,00012	0,00005	0,026
Tapasztalat ² középfok	–0,00117	0,00020	0,000	–0,00013	0,00003	0,000
Tapasztalat ² felsőfok	–0,00097	0,00029	0,001	–0,00021	0,00006	0,000
Tapasztalat ³ 0–7 általános	2,84E–06	0,00000	0,540	.	.	.
Tapasztalat ³ 8 általános	5,02E–06	0,00000	0,022	.	.	.
Tapasztalat ³ szaktanulmányok	1,42E–05	0,00000	0,000	.	.	.
Tapasztalat ³ középfok	1,45E–05	0,00000	0,000	.	.	.
Tapasztalat ³ felsőfok	1,21E–05	0,00000	0,009	.	.	.
Van-e mellékkeresete	0,01415	0,00492	0,004	0,00979	0,00509	0,054

n 17263

R^2 0,2735

Varianciaanalízis, F -próba 85,13 $p = 0,000$

Root MSE 0,1434

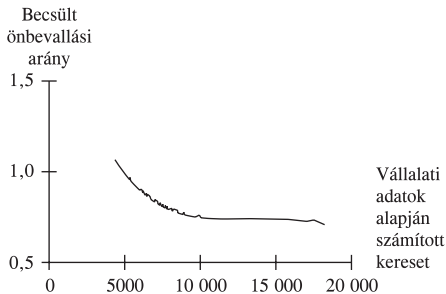
Cook–Weisberg-féle heteroszkedaszticitás-próba $\chi^2 = 1861$ $p = 0,000$

Ramsey-féle RESET teszt kihagyott változókra $F = 3,69$ $p = 0,025$

* A paraméterek t -próbái Huber-féle heteroszkedaszticitás-konzisztens standard hiba becslésen alapulnak.

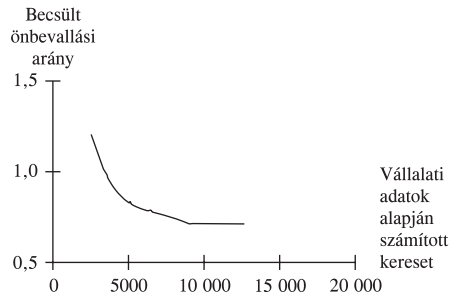
4. ábra

Becsült önbevallási arány és kereset,
medián iskolai végzettséggel
és tapasztalattal rendelkező férfiak



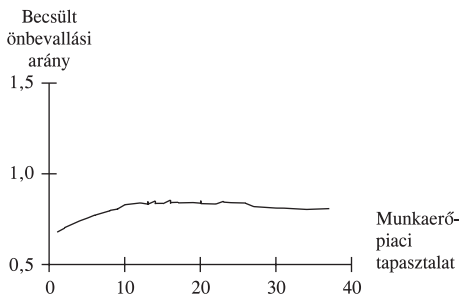
5. ábra

Becsült önbevallási arány és kereset,
medián iskolai végzettséggel
és tapasztalattal rendelkező nők



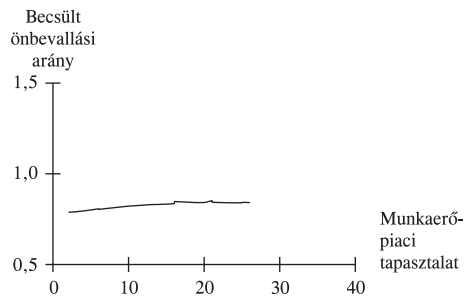
6. ábra

Becsült önbevallási arány és tapasztalat,
medián iskolai végzettségű és keresetű férfiak



7. ábra

Becsült önbevallási arány és tapasztalat,
medián iskolai végzettségű és keresetű nők



raméterek t -próbáit ezért Huber-féle (vagy ami ezzel ekvivalens: White-féle) heteroszkedaszticitás-konzisztens standard hiba becslés alapján újra elvégeztem. A becslés és a főbb próbák eredményeit a 2. táblázat tartalmazza nemek szerinti blokkokban.

Az interakciós hatások és a magasabb hatványok miatt igen bonyolult a regressziós táblázat interpretációja, és a konstansok elvi értelmezhetetlensége miatt is (önbevallási arány 0 keresetnél) célszerűbb a becslési eredmények grafikus vizsgálatából következtetéseket levonni. A 4–7. ábrák nemek szerinti bontásban mutatják a becslült önbevallási arány és a valós kereset, illetve a munkaerő-piaci tapasztalat kapcsolatát. A két magyarázó változó becslült hatását mindig a medián iskolai végzettség kategóriában (szakmunkásképző), a másik változó medián értékénél (a valós keresetek esetében a 49–51. percentilisek közötti értékeknél) láthatjuk.

Már ezen ábrák alapján is levonhatjuk elemzésünk legfőbb következtetését: az önbevallási arányt elsősorban a valós keresetek határozzák meg, mégpedig igen meredek negatív összefüggés alapján. Az életkor hatása igen kicsi, leginkább a fiatalabb férfiak többiekénél alacsonyabb bevallási arányára korlátozódik. A mellékkereseteket kontrolláló változó hatása nagyon kicsi, a nőknél nem különböztethető meg nullától, és – meglepő módon – egyebek változatlansága mellett a férfiaknál viszont 1 százalékkal magasabb önbevallási arányt jelez.

A 8–17. ábrák a nem és iskolai végzettség tíz kategóriájában mutatják a becslült önbe-

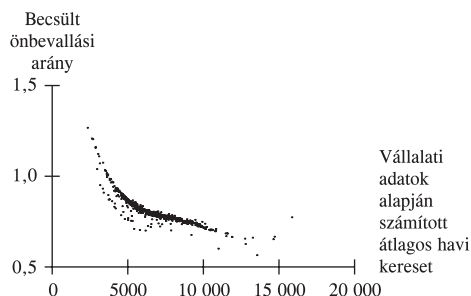
vallási arány és a valós kereset kapcsolatát. A többi magyarázó változó igen kicsi hatását az alakzatok "vékonysága", vagyis a becslésnek rögzített valós keresethez tartozó szóródása jelzi. Az összefüggések erejének összehasonlításán túl ezek az ábrák a becslés megfelelő parametrizálását is igazolják.

A legfontosabb eredményt, az önbevallási arány és a kereset erős negatív kapcsolatát mindenütt megfigyelhetjük. A nem és iskolai végzettség kategóriákban mindenütt körülbelül az első decilis a választóvonal: alatta a megkérdezettek inkább magasabb, felette inkább alacsonyabb összeget jelöltek meg. A leginkább iskolázott kategórián kívül a nők becsült "görbéje" enyhén balra helyezkedik el a férfiakétól, ami egyebek változatlansága mellett némileg alacsonyabb önbevallási arányt jelez. Azonos keresetek mellett, úgy tűnik, a magasabb végzettségűek általában magasabb önbevallási aránnyal rendelkeznek, ám egyrészt kicsik ezek az eltérések, másrészt az iskolai végzettséggel a keresetek eloszlása is felfelé tolódik el. Érdekes módon a férfiak esetében (a legiskolázatlanabbakat leszámítva) az adott iskolai kategória felső decilisében a keresetek növekedése már nem csökkenti a bevallási arányt, ám ez az összefüggés a nőket nem jellemzi.

A feltárt összefüggések alapján úgy tűnik, hogy az adott – nem és iskolai végzettség által meghatározott – kategórián belüli relatív kereseti helyzet határozza meg az önbevallási arány várható értékét. Az ez alapján nem megmagyarázott variancia a teljes variancia csaknem háromnegyedét teszi ki, ám e szóródás mögött minden valószínűség szerint olyan tényezők állnak, amelyek nem kapcsolhatók sem a keresetek, sem a munkaerőpiaci tapasztalat, sem a mellékkeresetek különbségeihez.

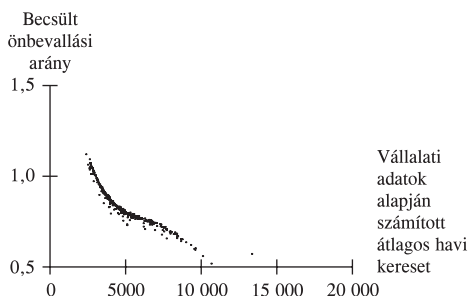
8. ábra

Becsült önbevallási arány és kereset
– férfiak, 0–7 általános



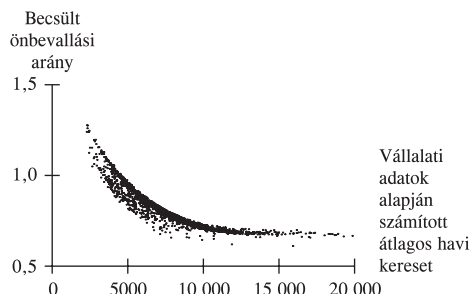
9. ábra

Becsült önbevallási arány és kereset
– nők, 0–7 általános



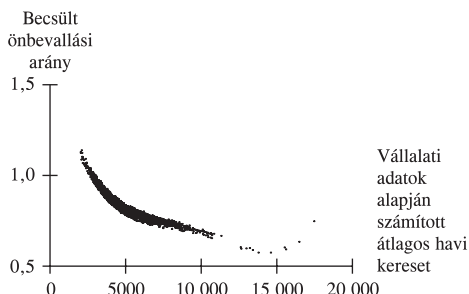
10. ábra

Becsült önbevallási arány és kereset
– férfiak, 8 általános



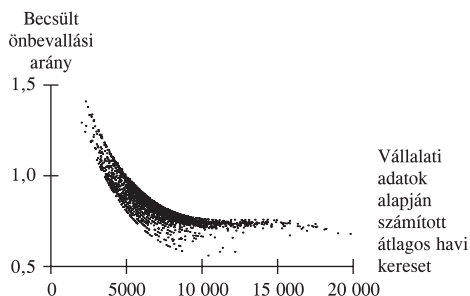
11. ábra

Becsült önbevallási arány és kereset –
nők, 8 általános



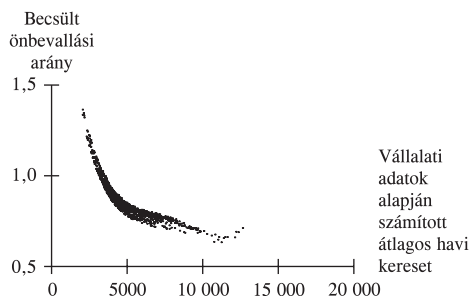
12. ábra

Becsült önbevallási arány és kereset
– férfiak, szakmunkásképző



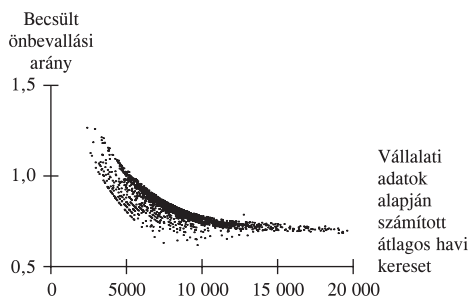
13. ábra

Becsült önbevallási arány és kereset
– nők, szakmunkásképző



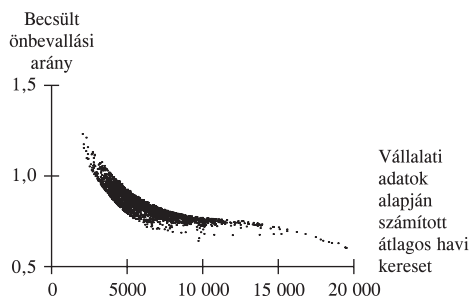
14. ábra

Becsült önbevallási arány és kereset
– férfiak, középfok



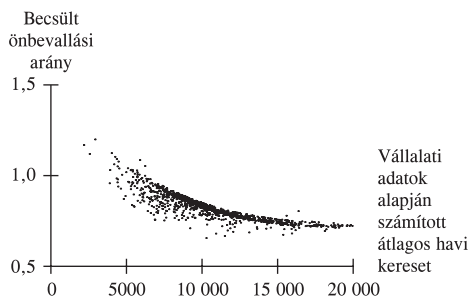
15. ábra

Becsült önbevallási arány és kereset
– nők, középfok



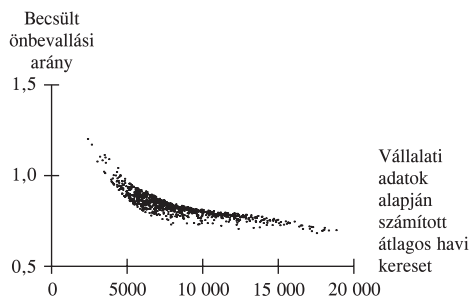
16. ábra

Becsült önbevallási arány és kereset
– férfiak, felsőfok



17. ábra

Becsült önbevallási arány és kereset
– nők, felsőfok



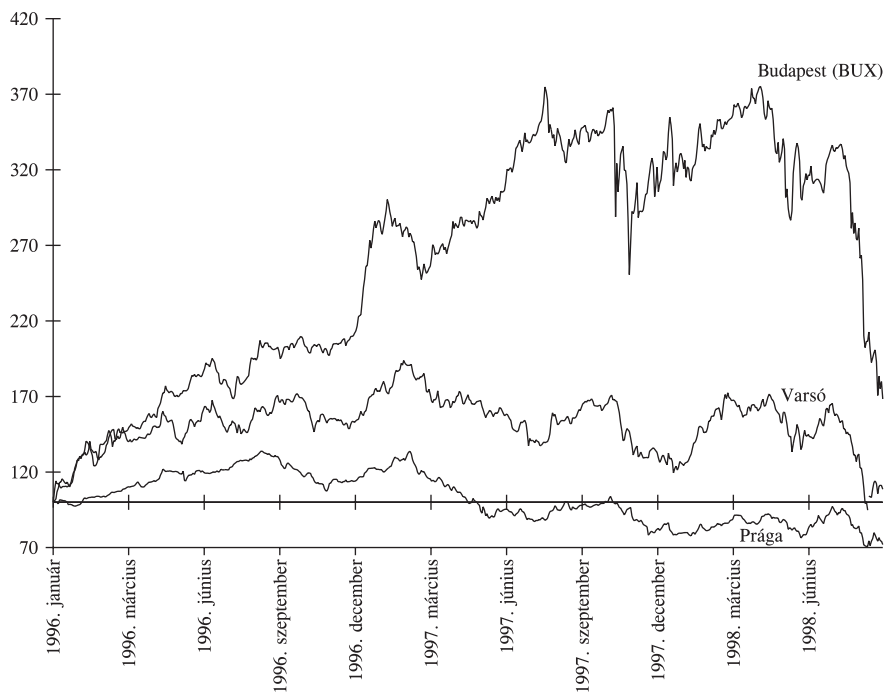
Rövid összegezés

A tanulmány megállapította, hogy az önbevallás átlagosan körülbelül két tizeddel alacsonyabb keresetet eredményez, s ilyen esetben a keresetek szóródása is kisebb. A részletesebb vizsgálat kimutatta, hogy az önbevallási arány legfontosabb meghatározója maga a kereset: magasabb kereset egyebek változatlansága mellett várható értékben mindig alacsonyabb bevallási arányt eredményez. Az összefüggés ereje iskolai végzettség és nem szerint némiképpen eltér, jelentős kvalitatív különbségek azonban nem fedezhetők fel.

A feltárt összefüggések alapján a kizárólag önbevalláson alapuló kereseti adatokat igen körültekintően szabad csak elemezni. Napjainkban a feltárt eltérések valószínűleg inkább növekedtek, részben a személyi jövedelemadó megjelenésével felmerülő "bruttó" és "nettó" terminusokban gondolkodás szükségszerű hibaforrásai miatt, részben az – anonim felvételekkel kapcsolatban valójában irreleváns – állampolgári öntudat megnövekedése miatt. Amennyiben figyelmen kívül hagyjuk a problémát, olyan torzítást vihetünk az elemzésünkbe, ami tartalmi kérdésekben is hamis megállapításokat eredményezhet. Ha azonban tisztában vagyunk az önbevallási adatok korlátaival, ezek az adatok is hasznos információkkal szolgálhatnak.

9. ábra

Részvényárak Budapesten és néhány kiemelt kelet-közép-európai tőzsdén, 1996–1998
(dollarban számítva, 1996. január = 100)



Forrás: a szerzők számításai.