

A Zipf-törvény érvényesülése a világ országaiban

A térbeliség felértékelődése a társadalomtudományokban számos régi földrajzi vonatkozású törvényszerűség és összefüggés újragondolását és a magyarázatok feltárásával kapcsolatos kutatások egész sorát hozta magával. A városok méret szerinti eloszlási viszonyainak alakulása, vagy szűkebb értelemben a Zipf-törvény is egyike azoknak az összefüggéseknek, amelyek a tér iránti megújult érdeklődés közepette az elmúlt két évtizedben számos tudományág képviselőinek érdeklődését vonta magára. A Zipf-törvénnyel kapcsolatos legfontosabb kérdés az lett, hogy a városok termelésben betöltött eltérő szerepe, méretheterogenitása és eloszlási viszonyai miért éppen a törvényben foglalt, a valóságban is meglepően gyakran érvényesülő valószínűségi eloszlás alapján írhatók le. Az élénk tudományos vitákban újabb és újabb megközelítések kerültek napvilágra, továbbá az adatok hozzáférhetőségének javulásával az addigiaknál jóval nagyobb mintákon és egyre kifinomultabb módszerekkel történő empirikus vizsgálatok is megjelentek.

A következőkben magam is a Zipf-törvény érvényesülésének empirikus tesztelését végzem el egy 102 országot tartalmazó sokaságon. Részletesen tárgyalom a szabállyal kapcsolatos, a hazai szakirodalomból mindeddig hiányzó elméleti kérdéseket, valamint az empirikus teszteléshez elengedhetetlen módszertani újításokat. A tanulmánynak nem célja, hogy országonként elemezze a városok eloszlási viszonyainak alakulását, csupán a Zipf-törvény globális érvényességének feltárására, illetve a városhálózatokat leíró paraméterek kitüntetett értékeinek vizsgálatára szorítkozik.

A Zipf-törvény matematikai interpretációja

A városok rangja és népességszáma közötti rang–nagyság összefüggés története csaknem 100 évre nyúlik vissza. A szabállyal kapcsolatos első felvetések Felix Auerbach német fizikushoz köthetők, aki 1913-ban publikált cikkében az Amerikai Egyesült Államok és további öt európai ország példáján megmutatta, hogy a városok lakosságszámának és méret szerinti rangjának szorzata országonként eltérő konstans ad eredményül. A rang és a város népességszám alapján meghatározott nagysága között tehát fordított arányosság van. Singer (1936) később a francia városhálózatot tanulmányozva arra a következtetésre jutott, hogy a városok népességszáma és rangja bizonyos hatványfüggvényszerű eloszlást, nevezetesen a Pareto-eloszlást követi (Berry–Okulicz-Kozaryn 2012, 1–2). A kérdés részletes továbbgondolására George Kingsley Zipf (1949) nyelvész vállalkozott, aki tézisében kifejtette, hogy ideális esetben adott ország i -edik városának népességszámmal jelzett nagysága (S_i) a legnagyobb népességszámú város (S_1) és az i -edik város rangjának (R_i) hányadosaként adható meg: $S_i = S_1 / R_i$. A Zipf-törvényként továbbélő szorosabb összefüggés a Pareto-eloszlás speciális eseteként fogható fel, ugyanis a Pareto-eloszlás tipikus sűrűségfüggvénye:

$$s_i = AR_i^{-\alpha} \quad (1),$$

vagy logaritmikusan transzformálva:

$$\ln s_i = \ln A - \alpha \ln R_i \quad (2).$$

A Zipf-törvény az (1) és (2) egyenletektől csupán annyiban módosul, hogy az α paraméter egyes értéket vesz fel. A Zipf-törvény tehát lényegében „korlátozott rang-nagyszábként” (Carroll 1982, 1) is értelmezhető.¹

A rang-nagyság, illetve a szorosabban vett Zipf-törvény formális valószínűségelméleti levezetése később született meg és lényegében a következő sémát követi; definíció szerint adott Pareto-eloszlást² követő X valószínűségi változó eloszlásfüggvénye a következő alakban írható fel:

$$F(x | \alpha, k) = P(X \leq x) = \int_{-\infty}^x f(x) dx = 1 - \left(\frac{k}{x}\right)^\alpha \quad (3),$$

ahol $x > 0$ a $k \leq x < \infty$ intervallumon értelmezett tetszőleges valós szám, $k > 0$ az x felvehető legkisebb értéke, α paraméter pedig a Pareto-kitevő. Az eloszlásfüggvényből származtatott sűrűségfüggvény:

$$f(x | \alpha, k) = \frac{\partial F(x)}{\partial x} = \frac{\alpha k^\alpha}{x^{\alpha+1}} \quad (4),$$

x és k azonos feltételei mellett. A városok eloszlási viszonyaira vonatkoztatva a fenti egyenleteket, a rang-nagyság összefüggés nehézség nélkül levezethető. A városok rangja (R) a megfigyelt városok számának (N) és a városok népességszámára (S) vonatkoztatott eloszlásfüggvény (3) összefüggéséből adódik:

$$R = N(1 - F(s)) = N \left(\frac{k}{s} \right)^\alpha \quad (5).$$

Annak ismeretében tehát, hogy S milyen $F(s)$ valószínűséggel vesz fel tetszőleges s -nél alacsonyabb értéket, egyszerűen megadható s népességű város rangja. Például abban az esetben, ha $k = s$, azaz s értékét a legkisebb felvehető k értékhez – vagyis momentán a városhálózatban előforduló legkisebb népességszámhoz – rögzítjük, akkor biztos, hogy $S \geq s$ értéket fog felvenni, így $F(s)$ valószínűség zérus. Ekkor s népességszámú város rangja a megfigyelt városok számával (N) lesz egyenlő. Ellenkező esetben, ha

¹ A rang-nagyság szabály elnevezése a nemzetközi és hazai szakirodalomban egyáltalán nem konzekvens. Az angol nyelvű regionális tudományi és közgazdasági folyóiratokban publikált tanulmányok döntő többsége a Zipf-törvényt a rang-nagyság összefüggés egységkitevős formájaként definiálja, azonban Gregory és szerzőtársai (2009, 220) által szerkesztett „*The Dictionary of Human Geography*” ötödik kiadásában a rang-nagyság szabály címszó alatt olvasható magyarázat csak a szoros értelemben vett Zipf-féle összefüggést fedi le. A magyar szakirodalomban az elnevezések terén hasonló fejtelenség uralkodik, a szerzők többsége a Zipf nevéhez köthető korlátozott összefüggést Auerbach–Zipf-szábként, esetleg törvényként emlegeti. Cséfalvay (1994, 307–314) „*A modern társadalomföldrajz kézikönyve*” című összefoglaló monográfiájában szisztematikusan az Auerbach-szábként megnevezést használja, a Bartke–Illés (1997, 103–108) szerzőpáros „*Telephelyelméletek*” című egyetemi jegyzetében Illés (1979) „*Regionális gazdaságtan*” című korábbi disszertációjához hasonlóan sorrend-nagyság szabályt emleget, a kapcsolódó kutatók közül egyedül Vilfredo Paretot említve, aki azonban közvetlenül nem köthető a koncepcióhoz.

A továbbiakban jómagam a nemzetközi szakirodalomban általánosan elterjedt elnevezésekhez igazodom, azaz rang-nagyság szabálynak a valamely hatványtörvénnyel (például Pareto-eloszlás) általánosan felírható összefüggést nevezem, a Zipf-törvény megnevezést pedig ennek korlátozott, egységkitevős alakjára alkalmazom, illetve ennek érvényesülését vizsgálom.

² Pareto-eloszlás és egyéb hatványfüggvényeszerű eloszlások igen gyakoriak a természeti és társadalmi rendszerekben; jól közelítik többek között a társadalmi jövedelemegyenlőtlenségeket, a tudományos publikációkra történő független hivatkozások számának, a telefonközpontokba érkező hívások számának, a Holdon azonosítható meteorokráterek méretének, vagy a Földön észlelt földrengések magnitúdójának eloszlásait. A hatványfüggvényeszerű eloszlásokkal jellemezhető természeti és társadalmi jelenségekről részletes összefoglalás olvasható Pinto et al. (2012) és Newman (2005) munkáiban.

s értékét megfelelően nagy népességszámban határozzuk meg, $F(s)$ valószínűség egységhez fog közelíteni, így a rang számértéke tetszőlegesen alacsony lesz.

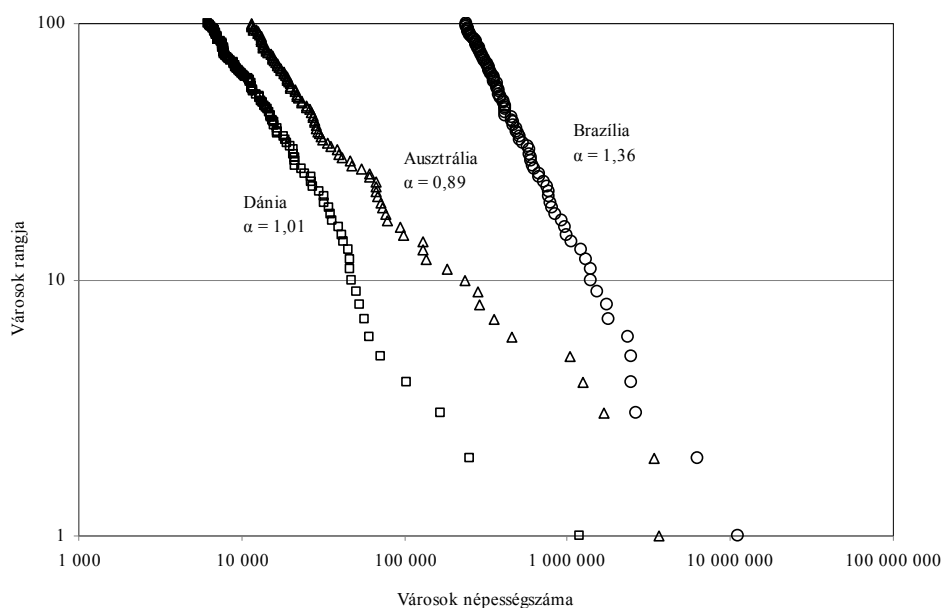
Ha a (5) egyenletet logaritmikusan transzformáljuk, az egyenlet rendezése után az ismerős

$$\ln R = K - \alpha \ln s \quad (6)$$

összefüggés adódik, ahol $K = \ln N + \alpha \ln k$ (Eeckhout 2004, 1441). Az összefüggés egyik jellegzetessége, hogy a városok rangját a méret függvényében logaritmikus skálázás mellett ábrázolva egyenes vonalat kapunk, amelynek meredeksége α értékétől függ (Mitzenmachel 2003, 228). Az 1. ábrán feltüntetett országok jól szemléltetik a logaritmikus skálázás mellett jellemző lineáris jelleget, valamint azt, hogy a meredekség a Pareto-kitevő értékének növekedésével együtt nő. A kis rangszámú városok esetén komoly eltérések tapasztalhatók a trendtől, ez azonban nem feltétlenül jelenti azt, hogy a becslt eloszlás nem illeszkedik elég jól a városhálózat egészére. A torzítás ugyanis annak tudható be, hogy a városok méretét növelve a megfigyelések száma csökken, tehát azok a kisméretű városok, amelyek sűrű elhelyezkedésük miatt a trend irányát intuitíve meghatározzák, túlreprezentáltak.

1. ábra

Néhány ország városainak eloszlása logaritmikus skálázás mellett



Megjegyzés: az országok mellett feltüntetett Pareto-kitevők maximum likelihood becslés (MLE) eredményei. A becslési módszert a későbbiekben részletezem.

A (6) egyenlet kapcsán szembetűnő továbbá az is, hogy nagyon hasonlít a (2) egyenletre, azonban mégsem tekinthető az inverzének, ugyanis $K \neq A$. Ennek ellenére mindkét egyenlet egyaránt alkalmas α paraméter becslésére, az eredmények interpretációja is

azonos.³ A Zipf-törvény érvényesülését mindkét megközelítés esetén $\alpha = 1$ jelzi, a teljes városhálózat szélsőséges koncentrációjára pedig az alacsony Pareto-kitevők utalnak. A Pareto-kitevő értéke azonban nem azonosítható egyedül a legnagyobb város és a városhálózat többi részének viszonyával, a városhálózat egészére értendő. Gyakran felbukkan az a hiba, miszerint a rang–nagyság viszonyok magyarázatában kizárólag a legnagyobb egy-két város méretével érvelnek.⁴ Az efféle elnagyolt magyarázatok ugyanis rendkívül félrevezetők lehetnek attól függően, hogy a kitevőket milyen eljárással becsüljük, illetve milyen megoldásokat választunk a koncepcióval kapcsolatos elméleti és módszertani nehézségekre.

A koncepció kapcsán felmerülő elméleti nehézségek

A Zipf-törvény, illetve a tágabban értelmezett rang–nagyság összefüggés kapcsán több olyan elméleti probléma is felmerül, ami jelentősen megnehezíti az empirikus vizsgálatok eredményeinek értelmezését, sőt egyes esetekben a vizsgálati eredményeket is befolyásolja.

Az első ilyen probléma a megfigyelés alapegységeinek megválasztása; a rang–nagyság szabály ugyanis alapvetően a városok, tágabb értelemben a városi terek eloszlási viszonyaira felírt összefüggés, de korántsem mindegy, hogy mit értünk „városi terek” alatt. Auerbach korai sejtése szerint például a közigazgatási határokat átlépő, valós gazdaságföldrajzi szerveződés eredményeként létrejövő városi agglomerációk a népesség tekintetében sokkal koncentráltabbak, mint a pusztán jogállás alapján besorolt, szuburbán térségeiktől leválasztott városok (Nitsch 2005, 91). A feltételezés első empirikus bizonyítására Rosen és Resnick (1980) nyomán került sor, akik úgy találták, hogy az agglomerációkra számolt Pareto-kitevők értékei alacsonyabbak a városokra számoltakhoz képest. Ezt az állítást negyed évszázaddal később Soo (2005) saját empirikus és Nitsch (2005) metaregressziós vizsgálata is megerősítették.

A közigazgatási egységek és a közigazgatási határokkal nem feltétlenül konzisztens nagyvárosi agglomerációk közti döntést legtöbbször a vizsgálni kívánt ország vagy térség jellemzői alapján hozzák meg a kutatók, ugyanis kisebb lakosságszámú országokban a kiterjedt agglomerációk száma a statisztikai becslések elvégzéséhez kevés vagy egyáltalán nincs is. Az elemzés alapegységeként tehát általában a mesterséges közigazgatási egységek kerülnek szóba, azonban a kérdés még korántsem dőlt el ezzel. Hasonlóan fontos ugyanis, hogy a város címet elfogadjuk-e mint egzakt városkritériumot, vagy esetleg adott település vizsgálatba történő bevonását egyéb funkciókhoz kötjük-e. Utóbbi esetben számos városi címet viselő település központi funkciók híján kiesik az elemzési körből, miközben több szolgáltatási szerepkörét tekintve fontos község bekerül. Kisterületű, alacsony népességű és egyedi természeti körülményekkel bíró országoknál ez a probléma egészen odáig fajulhat, hogy már eleve a rang–nagyság vizsgálat értelme is kétségsbe vonható. Izland esetében például a főváros és egy-két nagyobb, pár tízezer lelket számláló település mellett rengeteg párszáz fős település meglétéről beszélhetünk,

³ A Zipf-törvény empirikus vizsgálatában inkább az ímént levezetett (6) egyenlet terjedt el, a (2) egyenletet kevés tanulmányban alkalmazzák (Nitsch 2005, 89).

⁴ A magyar szakirodalomban sajnos fokozottan jellemző, hogy a szerzők a főváros „vízfej jellegét” emlegetik csupán, és megfélekeznek a kisebb városok méretbeli különbségeinek alaposabb vizsgálatot és széles rálátást kívánó vizsgálatáról. Példaként lásd Illés (1975), Cséfalvay (1994), Bartke–Illés (1997), valamint Kovács (2002) magyarázatait.

ahol a funkcionális városok száma éppen elég a vizsgálat lefolytatásához. Mérlegelve tehát a definíciós problémákat, Cheshire (1999) az alapegységek megválasztásának kérdésében arra a következtetésre jut, hogy a rang–nagyság szabály érvényesülésének vizsgálataiban amennyiben más korlátja nincs, szerencsésebb a funkcionális városhálózat elemeit szerepeltetni, ha a valódi – és nem csak nevében, jogcímében – városi terek eloszlási viszonyaira vagyunk kíváncsiak (Cheshire 1999, 1345).

Tisztázandó kérdésként merül fel továbbá az elemzési egységek megválasztásához szorosan kötődően a vizsgálni kívánt településhálózatok definiálása. Az elemzés ugyanis elvégezhető pusztán települések egy bizonyos halmazára, vagy egy jól definiált hálózatra, amelynek elemei szoros kapcsolatban állnak, szétválaszthatatlan, összefüggő rendszert alkotnak. A gyakorlatban ez a kérdés a területegység megválasztásához kapcsolódik, azaz, hogy elfogadjuk-e a közigazgatási határokat mint az egyes hálózatokat bizonyos mértékben szétválasztó térelemeket. Az adatok hozzáférhetősége és az egyszerű indoklás miatt a vizsgálatok színtere általában egy ország vagy régió, azonban a közigazgatási határokhoz való merev ragaszkodás egyes esetekben igencsak torzítja a valóságot. A rang–nagyság szabály idősoros vizsgálatainak interpretációja kapcsán mindig észben tartandó az ország- és az egyéb közigazgatási határok változása, ami a településhálózat alakításában, az egyes elemek közti kapcsolatok átszervezésében kulcsszerepet játszik (Cheshire 1999, 1344). A közelmúltból jó példaként szolgál az egykori Jugoszlávia szétválása, Bosznia-Hercegovina mesterséges, etnikai alapon történő széttagolása, vagy újabban Dél-Szudán megalakulása.

A megfigyelésbe vont elemi egységek kiválasztásának végiggondolásakor egy másik, hasonlóan fontos kérdés is felmerül, nevezetesen az elemszám megválasztása, illetve az ehhez kötődő méretkorlátok felállítása. A városok vizsgálatba történő bevonásakor népességhatárokat a legtöbbször azért van szükség, mert az egyes területegységek funkcionális városhálózatáról rendelkezésre álló információk híján a legkézenfekvőbb megoldás a városok és rurális terek elhatárolására a lakosságszám alapján történő tipizálás. Az elemszámok meghatározására és ezzel gyakorlatilag a térségtípusok elkülönítésére leggyakrabban alkalmazott két módszer egy fix elemszám meghatározása, illetve egy minimális népességszint rögzítése mint korlát. Azonban mindkét megoldásnak megvannak a hátrányai. Összehasonlító vizsgálatokban egyik megoldás sem adekvát, hiszen nincs két ország vagy egyéb elemezni kívánt területegység, amelynek funkcionális városhálózatát ilyen előre rögzített értékekkel összehasonlíthatóvá lehetne tenni.

A korlátok felállítása és az elemszám alakulása továbbá kölcsönösen meghatározza egymást, így bármelyik megoldást is alkalmazzuk, a becsülni kívánt kitevő értékei mindig függeni fognak az előre rögzített értékektől. Alacsony elemszám, illetve magas népességhatár esetén a kitevő értéke megnő, s vele nő a regressziós maradéktagok értéke is, ami a heteroszkedaszticitás problémáját veti fel (Eeckhout 2004, 1440–1443, Gonzalez-Val 2010, 961). Hasonló következtetésre jutott Guerin-Pace (1995) is, aki a francia városhálózatot vizsgálva az elemzésbe bevont városok alsó népességhatárainak variálásával próbálta elejét venni a problémának. Alperovich (1989) a városok lakosságából számolt megoszlási viszonyszámokkal becsülte, amivel az elemszám-érzékenységet csökkenteni tudta. Cheshire (1999, 1345) összefoglaló dolgozatában Wheaton és Shishado (1981) korábbi tanulmányát idézve pedig egy harmadik lehetőséget javasolt, ami a vizsgálatba bevont városi népesség az ország össznépességének bizonyos aránya-

ban történő meghatározását jelenti. Ezzel a megoldással kapcsolatban is felmerülnek a korábbiaknál már említett kritikák, valamint erősen megkérdőjelezhető egy olyan módszer alkalmazása, ami egy előre meghatározott, esetleg falusi népességet is magába foglaló bármilyen arányhoz, vagy például a városlakók arányához igazít. Ebben az esetben ugyanis visszatérünk az alapegységek már ismertett problémájához, nevezetesen, hogy elfogadjuk-e az adminisztratív városdefiníciót. Az elemszám meghatározása tehát az elkerülhetetlen statisztikai hatások miatt igen nehéz feladat, a választott megoldás viszont erősen függ a kutatási célkitűzéstől, illetve az elérhető adatoktól is. Összehasonlító elemzések esetén nem mindig van lehetőség minden vizsgálatba vont ország városhálózatának behatóbb előtanulmányozására, ilyenkor apróbb megkötések mellett főleg az adatelérhetőség határozza meg a vizsgálatot. Egyedi térségek vagy országok vizsgálatakor viszont már van lehetőség, sőt el is várható az adott városhálózat kielégítő ismerete, így ezekben az esetekben megfelelő indoklás mellett megadható a népességhatár, vagy az elemszámmra vonatkozó megkötés.

Már csak azért is elengedhetetlen a megfelelő indoklás, mert az elemszámok és korlátozások problémája nem csupán egy általános statisztikai nehézség, jóval tovább mutat annál. Eeckhout (2004) kifejtette, hogy a Zipf-törvényben foglalt Pareto-eloszlás csak a városhálózat felső szegmensére illeszkedik, a népességi korlátok egyre kisebb értékre történő lezorításával, azaz a kisebb települések modellbe való beillesztésével a Pareto-eloszlás erősen torzít. Az eloszlás a görbe farok részén nem illeszkedik, ellentétben a lognormális eloszlással, amely akár az egész településhálózatra alkalmazható. Gonzalez-Val (2010) vizsgálatai Eeckhout állítását támasztják alá, Schaffar (2004) a Pareto-eloszlást tartja hatásosabbnak, Warren (2007) azonban eltérő városi besorolás alkalmazása mellett mind a Zipf-törvényt, mind a lognormális eloszlást alkalmatlannak tartja a városhálózat eloszlási viszonyainak leírására.

A korlátozások problémájának felismerésével a pontosan illeszkedő eloszlások keresése a Zipf-törvénnyel kapcsolatos kutatások népszerű irányává vált, abban azonban a legtöbb szerző egyetért, hogy a városok eloszlásgörbéjének felső „*upper tail*” részében a legtöbb szóba került hatványfüggvényyszerű valószínűségi eloszlás szinte ugyanúgy viselkedik, gyakorlatilag lehetetlen megkülönböztetni őket. Duranton (2006) mindezeket mérlegelve úgy véli, hogy a Zipf-törvényt mint első megközelítést az illeszkedési problémák ellenére hasznos kiindulópontként szolgál a városhálózat kutatásához (Duranton 2006, 543).

A Pareto-kitevők becslésének módszertani kérdései

A következőket és leíró statisztika területi adatokra történő alkalmazásának több problémája is megjelenik a rang-nagyság szabály becslése kapcsán. Ezeknek egy része a megfelelő elemzési munkamenet ismeretében kiküszöbölhető, de akadnak olyan kérdések, amelyeknek megválaszolása általánosan elfogadott válasz híján egyéni kutatói döntést kíván.

Az első probléma a különböző becslési eljárások alkalmazásával nyert eltérő eredmények kezelése. Természetesen mindig törekedni kell a lehető leghatékonyabb, konzisztens és torzításmentes becslőeljárás alkalmazására, azonban a közel azonos becslőerővel bíró módszerek között is akadnak különbségek a kapott eredmények tekintetében. Ennek

kezelése igen egyszerű, érdemes egyszerre több eljárással megismételni a becslést ugyanarra a sokaságra. Ez a megoldás azonban nem teljesen adekvát a becsült paraméterek homogenizáló hatásának kiküszöbölésére. A területi egyenlőtlenségi mutatók esete jól példázza, hogy adott jelenség eltérő térbeli konstellációi miként vezetnek azonos számú eredményre. Mivel a Pareto-kitevő értelmezhető a városhálózat egyenlőtlenségeinek mértékét leíró mutatóként (Terra 2009, 9), ezért itt is felmerül az a kérdés, hogy vajon van-e értelme két térség alapvetően eltérő szerkezetű, de közel azonos Pareto-kitevővel leírható városhálózatát egy kalap alá venni. A válasz véleményem szerint egyértelmű: nincs. A kitűzött céltól függően tehát ezt a problémát is kezelni kell; amennyiben csak a Zipf-törvény általános érvényesülésére vonatkozik a kutatói kérdés, az egyedi hálózatok elemzésétől el lehet tekinteni, tipikus városföldrajzi vizsgálatok esetén azonban a Pareto-kitevő mindig a városhálózat képével együtt vizsgálandó.

Egyes szerzők a paraméterbecslés eredményeinek értelmezési nehézségeit úgy próbálták kiküszöbölni, hogy egyéb, nem paraméteres eljárásokkal becsülték a városok eloszlási viszonyait. Warren (2007) a momentumok módszerét alkalmazta, Ioannides és Overman (2003), Eeckhout (2004) és Gonzalez-Val (2010) ezzel szemben egy másfajta, bonyolultabb eljárást használtak, amely a városok hosszú távú növekedési üteméből következtet a Zipf-törvény (illetve rang–nagyság szabály) jelenlegi érvényesülésére.⁵ A városok méretének Pareto-eloszlása mindezek mellett különféle illeszkedésvizsgálatokkal is elemezhető, amelyeknek lényege az, hogy a szóban forgó térség városeloszlásának nevezetes értékeit (átlag, szórás, medián, szélsőértékek stb.) egy ismert, azonos típusú eloszlás értékeihez hasonlítjuk, és az eltérések mértékét vizsgáljuk. Ilyen eljárással végzett becsléseket például Terra (2009).

Bármilyen módszert is válasszunk, gyakran beleütközhetünk abba a kérdésbe, amely a következtető statisztika területi adatokon történő alkalmazásakor mindig megjelenik. Ez pedig a hipotézisvizsgálatok relevanciája. Dusek (2006) szerint a területi adatok nem tekinthetők véletlen mintának, se sztochasztikus folyamatok végeredményének, és mivel a területi statisztikai vizsgálatok döntő részében az alapsokaság egészére rendelkezünk adatokkal, kevés kivételtől eltekintve a szignifikanciavizsgálatok elvégzése felesleges, többletinformációval nem szolgál (Dusek 2006, 233–238). Gabaix és Ioannides (2004) a rang–nagyság szabály becslésével kapcsolatban osztja ezt a véleményt, a két szerző szerint ugyanis a városhálózat eloszlási viszonyainak vizsgálata inkább egy görbeillesztési feladat, mintsem egy alapsokaságra, azaz a városhálózat egészére végzendő hipotézisvizsgálat (Gabaix–Ioannides 2004, 2350–2351). Ezzel a karakterisztikus kijelentéssel jómagam nem feltétlenül értek egyet, hiszen a városok vizsgálatba történő bevonása bárhogy is legyen indokolva, a kapott sokaság nem biztos, hogy teljes mértékben meg egyezik azzal az ideális városhálózattal, amelyre a Zipf-törvény a valóságban esetlegesen érvényesül. Egy-két releváns település elhagyása, vagy néhány felesleges többlet bevonása alakíthat annyit az eredményeken, hogy az érvényesülő Zipf-törvény hipotézisét pusztán a Pareto-kitevő értéke alapján esetleg elvessük, elkövetve a statisztikai következtetés-elmélet elsőfajú hibáját.

⁵ Az eljárás a városok arányos, méretfüggetlen növekedéséből indul ki. A Gibrat-törvényként is ismert városnövekedési elképzelés tárgyalásától eltekintek, a koncepcióról részletesebben ld. Gabaix (1999), Gabaix–Ioannides (2004).

A Zipf-törvény érvényesülésének vizsgálata

A Zipf-törvény legfőbb elméleti és módszertani kérdéseinek felvázolása után a következő fejezetben saját elemzésem részleteit mutatom be, amit a felhasznált adatok bemutatásával kezdek, majd a korábban felvetett problémák lehetséges kezelésének részletezése után a paraméterbecslések eredményeit értékelem.

A felhasznált adatok a www.citypopulation.de című webhelyről, Thomas Brinkhoff gyűjtéséből származnak. Az adatbázist több szerző felhasználta tanulmányához, többek között Soo (2005), Nishiyama és szerzőtársai (2008), illetve Terra (2009). Brinkhoff adatai 228 ország statisztikai hivatalának adatközléséből, 15 megbízható nemzetközi webhelyről (például Statoids, CIA World Factbook, World Gazetteer stb.) és 12 nemzetközi szervezet évkönyveiből származnak. Soo a megbízhatóságot illetően előtanulmányt végzett a forrással kapcsolatban, eredményei szerint a közölt adatok kivétel nélkül megegyeztek a feltüntetett forrásokban foglaltakkal (Soo 2005, 248). A hazai adatokat ellenőriztem, a 2011-re vonatkozó népességi adatok megegyeznek a 2011-es Helységnevkönyvben (KSH 2011) közzétettekkel. A becslések megbízhatóságának érdekében a vizsgálatba nem vontam be olyan országokat, ahol 30-nál kevesebb városra voltak adatok, így összesen 102 országra végzem el a vizsgálatot, amelyeket kontinensenként csoportosítok.

Az országok Pareto-kitevőit három eljárással becslöm. Az első eljárás a Zipf-törvény érvényesülésének vizsgálataiban leginkább elterjedt, de korántsem a leghatékonyabb legkisebb négyzetek módszere (OLS), ahol (6) egyenlet alapján $x_i = \ln s_i$, $y_i = \ln R_i$, $\bar{x}$ és $\bar{y}$ pedig x_i és y_i számtani átlagai:

$$\alpha = \frac{\sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^N (x_i - \bar{x})^2} \quad (7).$$

OLS-becslést vagy annak némileg módosított változatát használta többek között Rosen és Resnick (1980), Alperovich (1989), Guerin-Pace (1995), Soo (2005), Nishiyama et al. (2008), Konishi és Nishiyama (2009), Terra (2009), Berry és Okulicz-Kozaryn (2012).

Az OLS-becsléssel kapcsolatos problémákra először Gabaix és Ioannides (2004) hívták fel a figyelmet, akik Monte Carlo-szimulációk segítségével kimutatták, hogy az elemszám csökkenésével az OLS erősen torzít. A probléma ezzel az, hogy a torzítás már abban az elemszám-tartományban is komoly problémát okoz, amellyel az empirikus vizsgálatok többsége operál. Gabaix és Ibragimov (2007) ezt a problémát egy elegáns megoldással hárították el, a városok rangjából rendre levontak $1/2$ -t, amellyel a torzítás mértéke jelentősen redukálódott. Nishiyama és szerzőtársai (2008), valamint Konishi és Nishiyama (2009) azonban két további problémát is azonosítottak, mégpedig a standard hibák alulbecslését és a regressziós együtthatók t -tesztjének rendellenes eloszlását. Az OLS-becslés kapcsán nyert paraméterek standardhibái alacsony elemszám mellett alulbecsültek, ami a hipotézisvizsgálatok elsőfajú hibájának veszélyét hordozza magában. Gabaix és Ibragimov (2007) módszere ezt a problémát nem tudja megfelelően kezelni, az eljárás ugyanis igen gyakran másodfajú hibához vezet (Terra 2009, 13). A t -teszt Student-féle t eloszlásának rendellenes torzulása továbbá alkalmatlanná teszi a standard

kvantiliseket a szignifikancia mérésére, ugyanis a t-teszt értéke az elemszám növekedésével aszimptotikusan „szétrobban”, azaz értéke meredeken végtelenbe tart (Nishiyama et al. 2008, 695). A hibát a szerzők úgy korrigálták, hogy a t-tesztet valós eloszlásának kritikus zónáit újrabecsülték, valamint az adatsorok metszését javasolták.⁶ A módszertani viták ellenére az OLS-beclést és a hozzá tartozó korrigálatlan t-tesztet népszerűsége és érthetősége miatt az eredmények fenntartásokkal történő kezelése mellett elvégzem mintegy referenciavizsgálatként, egyszersmind rámutatva a különböző becslési eljárásokkal kapott eredmények eltéréseinek mértékére.

A második alkalmazott módszer a maximum likelihood (MLE) becslés. A módszer lényege, hogy olyan paramétereket keresünk, amelyek adott eloszlású minta likelihood-függvényének értékét maximalizálják (Newman 2005, 350–351, Quandt 1966, 60). A független, Pareto-eloszlású $(x_1, x_2, \dots, x_N)$ minta likelihoodja α és k függvényében a következő alakot ölti:

$$P(x | \alpha, k) = \prod_{i=1}^N \frac{\alpha k^\alpha}{x_i^{\alpha+1}} \quad (8),$$

ahol k a korábbiakban megszokott módon x_i hipotetikusán felvehető legkisebb értéke: $x \geq k \geq 0$, és $k = \min(x_i)$, azaz k értékét a megfigyeléseink legalacsonyabb értékén rögzítjük. Mivel a logaritmusfüggvény szigorúan monoton növekvő, ezért a maximumhely meghatározása egyenértékű a loglikelihood-függvény maximumhelyének meghatározásával. Ennek érdekében a (8) egyenletet logaritmikusan transzformálva a

$$\begin{aligned} \ln L = \ln P(x | \alpha, k) &= \sum_{i=1}^N \ln \left(\frac{\alpha k^\alpha}{x_i^{\alpha+1}} \right) \quad (9), \\ &= N \ln \alpha + N \alpha \ln k - (\alpha + 1) \sum_{i=1}^N \ln x_i \end{aligned}$$

majd a (9) összefüggést α szerint differenciálva a

$$\frac{\partial \ln L}{\partial \alpha} = \frac{N}{\alpha} + N \ln k - \sum_{i=1}^N \ln x_i \quad (10)$$

(10) egyenlethez jutunk. A deriváltat 0-val egyenlővé téve, majd az egyenletet α -ra rendezve, egyszerűsítve adódik α becslőfüggvénye:

$$\alpha = N / \sum_{i=1}^N \ln \frac{x_i}{k} \quad (11).$$

Az α paraméter Zipf-törvényben foglalt egységnyi értéktől való eltéréseinek szignifikanciáját a $2n\bar{\alpha} / \alpha$ statisztikával mérjük, ahol esetünkben $\bar{\alpha} = 1$. A tesztstatisztika $2(n-1)$ szabadságfok mellett χ^2 eloszlást követ. Quandt (1966) megmutatta, hogy (11) konzisztens becslőfüggvény, valamint aszimptotikusan torzítatlan (Quandt 1966, 61–62). Ez a becslés megfelelő lehet, ugyanakkor nagymintás próbaként kis elemszámok esetén torzít (Terra 2009, 7). A problémát kiküszöbölendő, Terra ezért Baxter (1980) alapján az MLE-becslés egy némileg módosított változatát ajánlja a Zipf-törvény vizsgálatára, ami-

⁶ Az eljárás lényegéről és menetéről bővebben lásd Nishiyama et al. (2008) és Konishi–Nishiyama (2009).

nek neve: *minimum variance unbiased* (MVU) becslőfüggvény.⁷ Baxter (1980) alapján az MVU az MLE-becslésből eredeztethető a következőképpen:⁸

$$\alpha_{(MVU)} = \left(1 - \frac{2}{N}\right) \alpha_{(MLE)} \quad (12).$$

A (12) függvény konzisztens, hatékony és kis minták esetén is torzítatlan becslést ad, a módszer alkalmazását a kis elemszámú országok pontosabb becslhetősége miatt tehát indokoltnak tartom. Az MLE- és MVU-becslések Terra (2009) szimulációi szerint igen hasonló számszerű eredményeket produkálnak, megbízhatóságuk messze felülmúlja a klasszikus OLS-becslését (Terra 2009, 10).

Mindhárom becslés esetén azonos k értékekkel számolok, ezek azonban országonként változók; mivel nem rendelkezem kellő információval 102 ország városhálózatának jellegzetességeiről, a még vizsgálatba vont legalacsonyabb népességű városok meghatározását előzőleg nem elméleti, inkább matematikai úton határoztam meg. Clauset és szerzőtársai (2009) azt javasolják, hogy ott húzzuk meg k értékét, ahol a mintánk k -val jellemzett eloszlásfüggvénye a leginkább hasonlít egy ismert, szintén k és α paraméterekkel jellemezhető eloszlásfüggvényhez. A szerzők ajánlása szerint ez úgy kivitelezhető, hogy a rendelkezésre álló sokaságunk különböző x_i értékeit minimumnak (k) tekintve az afeletti értékekre MLE-becsléssel rendre kiszámoljuk α értékét. A feltételezett k , az elemszám, valamint az így kapott α segítségével szimulálható egy jól illeszkedő eloszlásfüggvény, ami referenciaként szolgál az összehasonlításhoz, amelyet kétmintás Kolmogorov–Smirnov-teszt segítségével végzünk. Ahol a mintánk eloszlásfüggvénye és a referenciaként használt mesterséges eloszlásfüggvényünk abszolút értékben vett különbségének szupréma minimális, ott kell meghatározni k értékét.

A népességeküszöb meghatározásánál magam is e módszer szerint jártam el, előnye, hogy az így nyert k -val a sokaságot legjobban jellemző Pareto-kitevőt becsülhetjük,⁹ komoly hátránya azonban, hogy a vizsgálatba vont elemek nem valószínű, hogy a funkcionális városok teljes körét lefednék. Ezért az eredmények értékelésénél mindenképp indokolt lehet a szignifikanciaszintek alkalmazása.

Az eredmények értékelése

Az OLS-becsléssel nyert Pareto-kitevők 102 országra számított hisztogramját, néhány kitüntetett sokasági értéket és eloszlásra vonatkozó tesztet a 2. ábra mutatja. A minta átlaga igen közel esik egyhez, értéke 1,01. Az átlag alatti Pareto-kitevővel rendelkező országok többségben (61-en) vannak, ebből 60 rendelkezik 1 alatti értékkel. Az OLS-becslés szerint tehát a relatíve koncentrált városhálózatú országok száma gyakoribb.

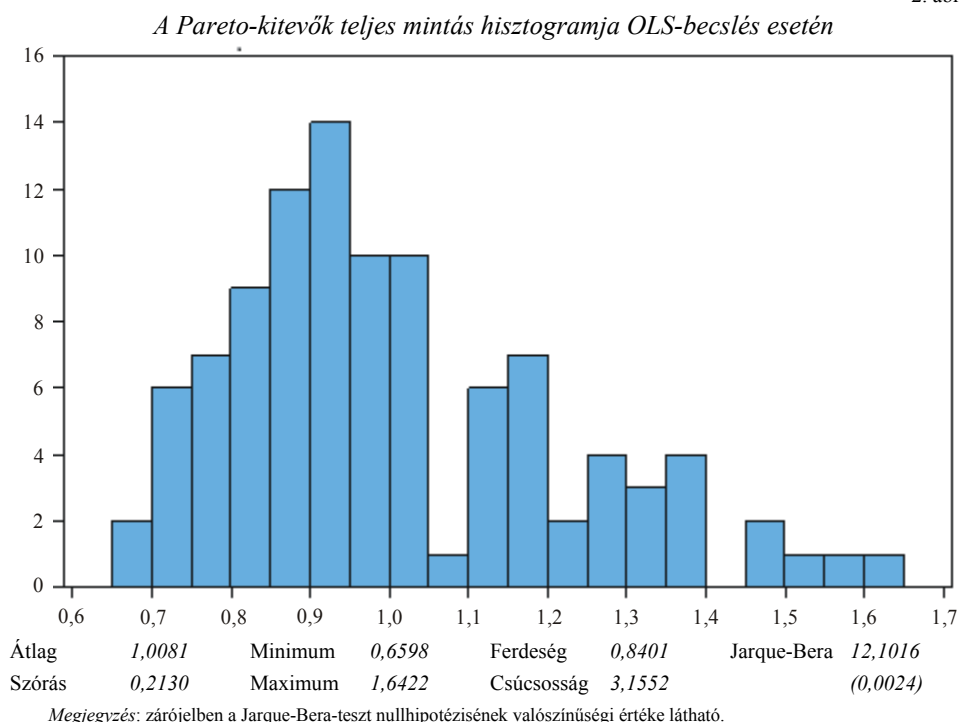
⁷ Valójában a MVU kifejezés általános gyűjtőnév, ami olyan becslőfüggvényekre vonatkozik, ami adott problémára a torzítatlan becslőfüggvények között a leghatékonyabb, azaz a legkisebb variációt produkálja, ezzel minimalizálva a becslés átlagos négyzetes hibáját.

⁸ A becslőfüggvények származtatásának részleteit lásd Baxter (1980).

⁹ Ez természetesen nem azt jelenti, hogy a városhálózat eloszlási viszonyainak leírására a Pareto-eloszlás a legalkalmasabb, létezhetnek egyéb eloszlástípusok is, amelyek esetenként jobban közelítik a sokaság eloszlását. Eeckhout (2004) és Gonzales-Val (2010) szerint például a lognormális eloszlás jobban illeszkedik a Pareto-eloszlásnál, Reed (2001) dupla Pareto-eloszlást, Cameron (1990) pedig Weibull-eloszlást javasol a városhálózatok eloszlási viszonyainak leírására. A különböző függvénytípusok illeszkedésének vizsgálata a Zipf-szabállyal kapcsolatos kutatások igen fontos szegmense.

A minta szórása 0,21, az átlag ± 1 szórásérték tartományon kívül 32 ország van, ebből 11 ország az átlag alatt. A legkisebb kitevő Bolíviáé (0,66), míg a legmagasabb, 1,64-os értéket Belgium veszi fel. Az OLS-becslés eredményeként kapott teljes mintás átlag egyedül Terra (2009) 1,04-os értékéhez hasonlít, Rosen és Resnick (1980), valamint Soo (2005) vizsgálatai jóval magasabb, 1,16 és 1,11-es értékeket adtak. A szélső értékek terén Terra és Soo tanulmányaihoz képest nincs jelentős eltérés.

2. ábra



A kitevők teljes mintabeli eloszlása igencsak aszimmetrikus, a gyakorisági görbe alakjára vonatkozó alakmutatók, a ferdeség (*skewness*) és csúcsosság (*kurtosis*) a normál eloszlástól jelentős eltérést mutatnak. A görbe unimodális, pozitív ferdeségű és a normál eloszláshoz képest csúcsosabb. Az alakmutatók együttes vizsgálatára alkalmas Jarque-Bera-féle normalitási teszt is elveti a minta standard normális eloszlásának nullhipotézisét. Ezek az eredmények némileg ellentmondanak Soo (2005) és Nitsch (2005) eredményeinek.

Az országok kontinensenként történő csoportosítása esetén némileg eltérő értékekkel találkozhatunk, amelyeket az 1. táblázat foglal össze. A változó elemszámok és a vizsgálati időpontok eltérései a kontinensek között nem biztosítanak összehasonlíthatóságot, de a bennük vizsgált országokat „mintaként” felfogva egyedileg elemezhetők.

1. táblázat

A Pareto-kitevő alakulása kontinensenként OLS-becslés esetén

Kontinens	n	$\alpha < 1$	$\alpha = 1$	$\alpha > 1$	Átlag	Szórás	Min	Max
Afrika	26	14	5	7	0,96	0,17	0,71	1,48
Ázsia	26	12	3	11	1,02	0,21	0,72	1,57
Ausztrália és Óceánia	2	2	0	0	0,72	0,02	0,69	0,74
Észak-Amerika	3	1	0	2	1,10	0,24	0,79	1,38
Európa	32	11	6	15	1,08	0,23	0,76	1,64
Közép- és Dél-Amerika	13	9	2	2	0,92	0,15	0,66	1,27
<i>Teljes minta</i>	<i>102</i>	<i>49</i>	<i>15</i>	<i>37</i>	<i>1,01</i>	<i>0,21</i>	<i>0,66</i>	<i>1,64</i>

Megjegyzés: a Zipf-törvény érvényesülése ($H_0: \alpha=1$) 5%-os szignifikanciaszint mellett értendő.

A kontinentális átlagokat tekintve Afrika és Ázsia mellett Európa és Észak-Amerika áll a legközelebb a Zipf-törvény érvényesüléséhez, ugyanakkor e két kontinensen van a legnagyobb szóródás az országos paraméterek között. Azonban Észak-Amerika egyik országában sem teljesül a Zipf-törvény, az Egyesült Államok 1,38-os értékével messze kiegyenlítettebb a 0,79-os értékkel bíró Kanadánál, illetve az 1,13-os kitevővel rendelkező Mexikónál. Európában a szóródás némileg kisebb, a legkiegyenlítettebb ország a világszinten is első Belgium, a legerősebben koncentrált szerkezetű ország 0,76-dal Észtország. A Zipf-törvény érvényesülése 6 esetben feltételezhető, ebből Szlovénia, Moldova, Románia és Bulgária kelet-európai átmeneti állam, a maradék kettő pedig Írország és Ausztria. A hat ország közül Moldova értéke áll a legközelebb 1-hez, esetében csupán egy 1,2 századnyi eltérés tapasztalható az egytől negatív irányban.

Az azonos elemszámokkal bíró Afrika és Ázsia átlaga jól közelíti az egyet, utóbbi felülről. Minimumaik nagyjából megegyeznek, Ázsia legkoncentráltabb városhálózati országa, azaz Afganisztán világszinten a negyedik, Afrika legkisebb Pareto-kitevőjű országa, Zimbabwe pedig kis eltéréssel a harmadik. A két kontinensen a legkiegyenlítettebbek Fülöp-szigetek és Tunézia, 1,57-os és 1,48-os értékekkel, a Zipf-törvény viszont csak kevés országban érvényesül. Afrikában Burkina Faso, Közép-afrikai Köztársaság, Ghána, Csád, Elefántcsontpart esetében, Ázsiában pedig Törökország, Tádzsikisztán és Szíria esetében beszélhetünk közel egységnyi kitevőkről. Törökország kitevője áll a legközelebb 1-hez a 102 országon belül, az eltérés mindössze 5 ezrelék. Figyelemre méltó India is, amelynek 1,02-os értéke érzés szerint igen közel áll egyhez, magas elemszáma miatt viszont elvetendő az egységkitevő nullhipotézise. India tehát mintapéldája a t-tesztrel kapcsolatos aggályoknak.

Közép- és Dél-Amerika átlaga az egytől negatívan csaknem 8 századdal tér el, valamint maximuma is a legkisebb (1,27), amit Brazília vesz fel. A többségében szignifikánsan egy alatti értékek dominálta Dél-Amerika adja a legkisebb kitevőjű országot, Bolíviát. Egységhez közeli kitevővel jellemezhető Uruguay és Nicaragua, ennél nagyobb értéket Brazília mellett pedig már csak Venezuela vesz fel.

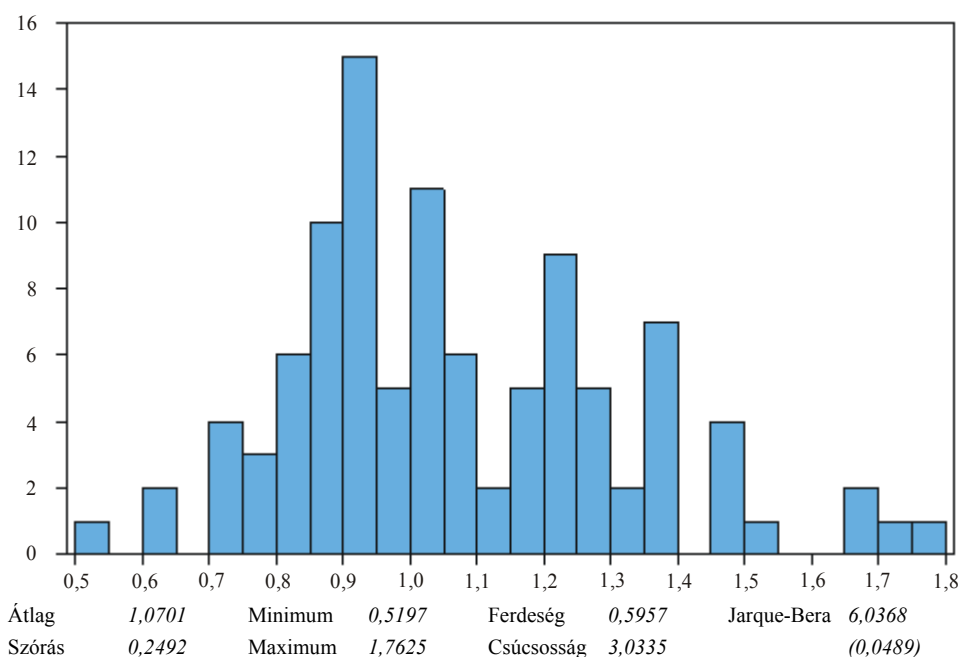
Ausztrália és Új-Zéland, a csendes-óceáni térség két képviselője igencsak különleges helyzetben van, hasonlóan alacsony, 0,7 körüli Pareto-kitevőik miatt átlagról és szórásról és szélsőértékekről nem nagyon van értelme beszélni.

Összességében OLS-beclsés használatával csak 15 esetben fogadható el a Zipf-törvény 5%-os szignifikanciaszint mellett, ami nagyságrendileg megegyezik Terra eredményeivel, de Soo-éval kevésbé. Esetemben azonban az adatok jócskán adnak példát az alacsony elemszám melletti lefelé torzítás, vagy a t-tesztek pontatlanságának eseteire. A korábban említett India mellett Dánia vagy Egyiptom esetében is elsőfajú hibára aszociálhatunk, ami Nishiyama és szerzőtársai (2008), valamint Konishi és Nishiyama (2009) bizonyítékai alapján a kiemelten magas elemszámukkal, s az ebből fakadóan alacsony standard hibákkal magyarázható. Mivel Soo több esetben nálam kisebb elemszámokkal dolgozott, ezért ott a t-tesztből fakadó kisebb torzulások eredményezhettek 73-ból 20, a Zipf-törvénynek megfelelő országot.

Az MLE-beclsés a minta egészének jellemzőit jelentősen megváltoztatta (3. ábra). Az α paraméter teljes mintás átlaga 1,07-ra, szórása pedig csaknem három tizeddel 0,25-ra nőtt az OLS-beclséshez képest. A ± 1 szórástávolságon kívül elhelyezkedő országok száma 28, s a szélsőértékek is kitolódtak mindkét irányban. Az átlaghoz legközelebb a Közép-afrikai Köztársaság áll, a minimumot Horvátország veszi fel (0,52), a maximumot (1,76) pedig ismét Belgium jegyzi.

3. ábra

A Pareto-kitevők teljes mintás hisztogramja MLE-beclsés esetén



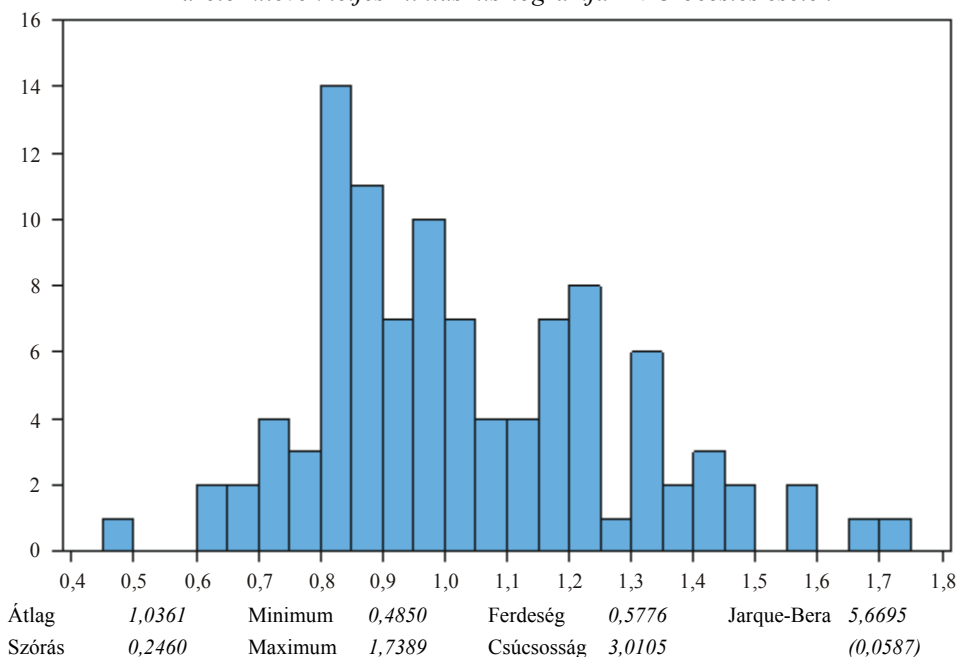
Az alakmutatók tekintetében az OLS-hez képest a legnagyobb változás a pozitív ferdeség mérséklődése és a csúcsosság kismértékű csökkenése. Az MLE-beclsések gyakorisági görbéje unimodális, a Jarque-Bera teszt normalitást feltételező nullhipotézise azonban 5%-os szignifikanciaszint mellett továbbra is elvetendő, a teljes minta eloszlása tehát – Soo és Nitsch eredményeivel ellentétben – nem normál eloszlású. Az MVU-beclsés

alkalmazásának esetén az alakra vonatkozó megállapítások nagyjából azonosak, a Jarque–Bera-teszt értéke a nullhipotézis elvetésének határán van, de nem vethető el. MVU-becslés esetén 1,03-ra csökken a sokasági átlag, a szórás (0,24) viszont az MLE értékével azonos. A minimum tovább csökken, amit változatlanul Horvátország vesz fel, a maximum pedig ismét Belgiumot illeti 1,74-os érték mellett.

A minta egészét jellemző kitüntetett értékek a maximum likelihood alapú becslések esetében nagyjából azonosak, országos értékeik pedig erős korrelációt mutatnak ($\rho = 0,99$). Az OLS és az MLE, valamint az OLS és az MVU között a kapcsolat intenzitása csökken, de még mindig erősnek mondható (0,89; 0,91). A becsléseket összehasonlítva megállapítható, hogy az országok többségében az MLE- és MVU-eljárással nyert Pareto-kitevők nagyobbak az OLS-nél látottaknál, azonban bőven akadnak kivételek. Az MLE és az MVU között pedig majdnem tökéletes egyenes arányosság tapasztalható, az MVU eredményei árnyalatnyival alacsonyabbak az MLE-vel becsült Pareto-kitevőknél.

4. ábra

A Pareto-kitevők teljes mintás hisztogramja MVU-becslés esetén



Az országokat kontinensenként csoportosítva az MLE és MVU között sok hasonlóság van, ahogy a 2. és 3. táblázat értékei is mutatják. A legnagyobb elemszámú Európa szélső értékei az OLS-hez képest kitolódtak, legkisebb értéke MLE és MVU esetén egyaránt Horvátországnak van (0,52; 0,49), a maximum változatlanul Belgiumot illeti. Európában az egységnél nagyobb és kisebb Pareto-kitevővel jellemezhető országok száma egyenlő, a Zipf-törvény mégis inkább az egységnél kisebb értékeket felvevő országokban fogadható el. A 21 országon belül Pareto-kitevőjét tekintve Görögország, Dánia, és Portugália áll

a legközelebb 1-hez. 5%-os szignifikanciaszint és adott elemszám mellett még éppen elfogadható Szlovénia 1,28-os és Észtország 0,87-os értéke. Magas elemszáma – és emiatt jóval nagyobb szabadságfoka – miatt nem fogadható el a szlovén értéknél kisebb német (1,22) és magyar (1,26) érték. Az MVU esetén annyi különbség adódik, hogy Zipf-törvény érvényesülésének valószínűségi intervallumán belül némileg változnak az értékek. Utóbbi esetben Portugália, Dánia és Litvánia veszi fel az egyhez legközelebbi értékeket.

2. táblázat

A Pareto-kitevő alakulása kontinensenként MLE-becslés esetén

Kontinens	n	$\alpha < 1$	$\alpha = 1$	$\alpha > 1$	Átlag	Szórás	Min	Max
Afrika	26	3	19	4	1,04	0,21	0,71	1,46
Ázsia	26	2	20	4	1,07	0,20	0,74	1,49
Ausztrália és Óceánia	2	1	1	0	0,77	0,12	0,65	0,89
Észak-Amerika	3	1	1	1	1,05	0,26	0,79	1,40
Európa	32	1	21	10	1,16	0,29	0,52	1,76
Közép- és Dél-Amerika	13	4	8	1	0,96	0,20	0,65	1,36
<i>Teljes minta</i>	<i>102</i>	<i>12</i>	<i>70</i>	<i>20</i>	<i>1,07</i>	<i>0,25</i>	<i>0,65</i>	<i>1,76</i>

Megjegyzés: ld. 1. táblázat.

Afrika esetében a legkisebb értéket Zambia veszi fel, a legnagyobb érték Tunéziát illeti meg becsléstől függetlenül. Az OLS-hez képest az 5%-os szignifikancia mellett létrehozott konfidenciaintervallumba 19 ország került be. Az egyhez Kamerun és Tanzánia áll a legközelebb, de közülük korábban egyik se lett elfogadva. A legnagyobb és legkisebb még elfogadott értékekkel Benin (1,30) és Namíbia (0,76) rendelkezik. MVU-becslést alkalmazva a Zipf-törvény érvényesülése csak 17 országban valószínűsíthető, a körből kiesik Kenya és Namíbia, egységhez legközelebb pedig Tanzánia, Kamerun, valamint a Közép-afrikai Köztársaság állnak.

3. táblázat

A Pareto-kitevő alakulása kontinensenként MVU-becslés esetén

Kontinens	n	$\alpha < 1$	$\alpha = 1$	$\alpha > 1$	Átlag	Szórás	Min	Max
Afrika	26	5	17	4	1,00	0,20	0,68	1,42
Ázsia	26	3	19	4	1,03	0,20	0,72	1,44
Ausztrália és Óceánia	2	1	1	0	0,75	0,13	0,62	0,87
Észak-Amerika	3	1	1	1	1,04	0,26	0,78	1,39
Európa	32	1	21	10	1,12	0,29	0,48	1,74
Közép- és Dél-Amerika	13	4	8	1	0,93	0,19	0,62	1,33
<i>Teljes minta</i>	<i>102</i>	<i>15</i>	<i>67</i>	<i>20</i>	<i>1,03</i>	<i>0,25</i>	<i>0,48</i>	<i>1,74</i>

Megjegyzés: ld. 1. táblázat.

Ázsiában a két becslés némileg azonos eredményeket ad, csupán Afganisztán besorolásában van eltérés. Az MLE-becslés eredményei alapján Afganisztán veszi fel a legkisebb (0,77) értéket, MVU esetén Szaúd-Arábia (0,71). A legnagyobb kitevővel mindkét esetben a Fülöp-szigetek rendelkezik (1,49; 1,44). A Zipf-törvény érvényesülése 20, illetve 19 országban valószínűsíthető, utóbbi esetben Afganisztán kiesik a körből. MLE-

becslés alapján Afganisztán és Oroszország a két szélső határon lévő ország, MVU esetében Grúzia és Azerbajdzsán veszi fel a szélső értékeket. Az egyhez legközelebb lévők pedig az agglomerációkkal számolt India, Pakisztán, Dél-Korea és Irán, ahol korábban az OLS-becslés t-tesztje elvetette az egységkitevő nullhipotézisét.

Ha kis elemszáma miatt Ausztráliát és Óceániát nem számítjuk, Közép- és Dél-Amerika átlaga és szórása a legkisebb a kontinensek között, továbbá a Pareto-kitevő tekintetében relative a leghomogénebb is. Az értékek 0,62 és 1,33 között szóródnak, a szélső értékek Paraguayt és Brazíliát illetik meg mindkét becslés esetén. Az országok többségében elfogadható a Zipf-törvény nullhipotézise, az egyhez rendre Nicaragua áll a legközelebb.

Észak-Amerika három országa igen szélsőséges értékeket vesz fel, minden csoportba jut egy ország. Az OLS-becslés eredményeihez hasonlóan a legkoncentráltabb Kanada, nagyfokú kiegyenlítettséget az Egyesült Államok mutat, a kettő között pedig Mexikó áll, melynek 0,96 körüli értéke a Zipf-törvény nullhipotézisének szempontjából elfogadható. Ausztrália és Új-Zéland értékei maradtak 1 alatt, azonban Ausztrália (0,89; 0,87) értéke MVU-becslés esetén a Zipf-törvény tekintetében elfogadható.

MLE- és MVU-becsléseket Sébastien Terra (2009) készített, 115 országos mintájában mindkét becsléssel 76 esetben valószínűsítette a Zipf-törvény érvényesülését. Saját vizsgálataim hasonló eredményt hoztak: az MLE eredményei alapján 70, MVU esetén 67 esetben azonosítható a szabály. A kontinensbeosztások némileg eltérőek,¹⁰ s mivel Terra tanulmánya nem részletezi, melyik csoportba melyik ország kerül, sajnos részletesebb összehasonlításra nincs lehetőség. Mindenesetre megállapítható, hogy az MLE- és MVU-becslések eredményei sem jelzik egyértelműen a Zipf-törvény szoros megkötéseinek érvényesülését, helytállóbb a rang–nagyság szabály érvényesüléséről beszélnünk a Pareto-kitevő egy szűk intervallumában. Az MVU eredményeit figyelembe véve például a minta elemeinek 90%-a a [0,6; 1,4] intervallumba esik, azonban nem szabad megfeledkezni arról, hogy az eredmények nagymértékben függenek a módszerektől, valamint számos egyéb eloszlástípus is létezik, amely a Pareto-eloszláshoz hasonló eredményességgel képes közelíteni a városhálózatok eloszlási viszonyait.

Összefoglalás

Az előzőekben a Zipf-törvény érvényességét vizsgáltam ökonometria módszerek segítségével. Tanulmányomban törekedtem arra, hogy a koncepcióval kapcsolatban az évtizedek alatt felvetődött elméleti és módszertani kérdéseket, problémákat mind a saját vizsgálataim megalapozása végett, mind pedig a szóban forgó ismeretanyag hiányos hazai megjelenése miatt tömören összefoglaljam. A koncepció tartalmának megismerése után 102 országból álló sokaságon teszteltem a Zipf-törvény általános érvényességét, amelynek eredményeként megállapítható, hogy a különféle becslési eljárások igen eltérő eredményekre vezethetnek, továbbá, hogy a Pareto-kitevő egységnyi értékének hipotézise a legmegengedőbb becslési eljárás szerint is csupán 70 esetben volt elfogadható. Ezek alapján leszögezhetjük, hogy a Zipf-törvény nem, ellenben a szabadabban értelmezett

¹⁰ Terra a Közel-Kelet térségét Ázsiától külön választja, Ausztráliát és Óceániát Ázsiához csatolja, valamint Dél-Amerika alatt Latin-Amerikát érti.

rang-nagyság szabály érvényesülhet a Pareto-kitevő viszonylag kicsi, 1-et is tartalmazó intervallumában. Az egységnyi kitevő átlagértékként való felfogása vizsgálataim alapján megállja a helyét, azonban fontos megjegyezni, hogy a becslési eljárásoktól, a rendelkezésre álló adatoktól, a vizsgálat időpontjától, illetve az alkalmazott küszöbértékektől függően ez kutatásonként nagymértékben eltérhet.

A Pareto-kitevők viszonylag kis intervallumban szóródó értékei miatt azonban továbbra is egyetérthetünk Fujita és szerzőtársainak (1999, 215) azon álláspontjával, miszerint a városok eloszlási viszonyainak alakulása egyszerűségénél fogva a társadalom térbeli szerveződésének egyik legfélelmetesebb jelensége.

IRODALOM

- Alperovich, G.* (1989): The distribution of city size: A sensitivity analysis. *Journal of Urban Economics*, 1.
- Bartke István – Illés Iván* (1997): Telephelyelméletek. Eötvös Kiadó, Budapest
- Baxter, M.* (1980): Minimum Variance Unbiased Estimation of the Parameters of the Pareto Distribution. *Metrika*, 1.
- Berry, B. – Okulicz-Kozaryn, A.* (2011): The city size distribution debate: Resolution for US urban regions and megalopolitan areas. *Cities*, 1.
- Cameron, T. A.* (1990): One-Stage Structural Models to Explain City Size. *Journal of Urban Economics*, 3.
- Carroll, G.* (1982): National city size distributions. What do we know after 67 years of research? *Progress in Human Geography*, 1.
- Cheshire, P.* (1999): Trends in sizes and structures of urban areas. In: Cheshire, Paul – Mills, Edward (eds.): *Handbook of Regional and Urban Economics*, Vol. 3, North-Holland, Amsterdam
- Clauset, A. – Shalizi, C. R. – Newman, M. E. J.* (2009): Power-law distributions in empirical data. *SIAM Review*, 4.
- Cséfalvay Zoltán* (1994): A modern társadalomföldrajz kézikönyve. Ikva, Budapest
- Dusek Tamás* (2006): Területi statisztika, valószínűségszámítás és statisztikai következtetésmélet. Területi Statisztika, 3.
- Duranton, G.* (2006): Some foundations for Zipf's law: Product proliferation and local spillovers. *Regional Science and Urban Economics*, 4.
- Eeckhout, J.* (2004): Gibrat's law for (All) cities. *American Economic Review*, 4.
- Fujita, M. – Krugman, P. – Venables, A. J.* (1999): *The Spatial Economy: Cities, Regions, and International Trade*. MIT Press, Cambridge
- Gabaix, X. – Ibragimov, R.* (2006): Log(Rank-1/2): A simple way to improve the OLS estimation of tail exponents. *NBER Papers*, 342.
- Gabaix, X.* (1999): Zipf's law for cities: An explanation. *Quarterly Journal of Economics*, 3.
- Gabaix, X. – Ioannides, Y. M.* (2004): The evolution of city size distributions. In: Henderson, V. – Thisse, J. F. (eds): *Handbook of Regional and Urban Economics*, 4., North-Holland, Amsterdam
- Gonzalez-Val, Rafael* (2010): The evolution of US city size distribution from a long term perspective (1900–2000). *Journal of Regional Science*, 5.
- Gregory, D. – Johnston, R. – Pratt, G. – Watts, M. – Whatmore, S.* (2009): *The Dictionary of Human Geography*. 5th Edition. Wiley-Blackwell, New York
- Guerin-Pace, F.* (1995): Rank size distribution and the process of urban growth. *Urban Studies*, 3.
- Illés Iván* (1979): Regionális gazdaságtan. Tankönyvkiadó, Budapest.
- Ioannides, Y. M. – Overman, H. G.* (2003): Zipf's Law for cities: An empirical examination. *Regional Science and Urban Economics*, 2.
- Kovács Zoltán* (2002): Népeség- és településföldrajz. Eötvös Kiadó, Budapest
- Konishi, Y. – Nishiyama, Y.* (2009): Hypothesis testing in rank-size rule regression. *Mathematics and Computers in Simulation*, 9.
- Kosmopoulou, G. – Buttry, N. – Johnson, J. – Kallsnick, A.* (2007): Suburbanization and the rank-size rule. *Applied Economics Letters*, 1.

- Mitzenmacher, M.* (2003): A brief history of generative models for Power-law and lognormal distributions. *Internet Mathematics*, 2.
- Newman, M. E. J.* (2005): Power laws, Pareto distributions and Zipf's law. *Contemporary Physics*, 5, pp.323–351.
- Nishiyama, Y. – Osada, S. – Sato, Y.* (2008): OLS estimation and the *t*-test revisited in rank–size regression. *Journal of Regional Science*, 4.
- Nitsch, V.* (2005): Zipf zipped. *Journal of Urban Economics*, 1.
- Pinto, C. – Lopes, M. – Machado, T.* (2012): A review of power laws in real life phenomena. *Communion in Nonlinear Science and Numerical Simulation*. Forthcoming
- Reed, W.* (2001): The Pareto, Zipf and other Power-laws. *Economic Letters*, 1.
- Rosen, K. T. – Resnick, M.* (1980): The size distribution of cities: An examination of the Pareto law and primacy. *Journal of Urban Economics*, 2.
- Quandt, R. E.* (1966): Old and New Methods of Estimation and the Pareto Distribution. *Metrika*, 1.
- Soo, K. T.* (2005): Zipf's Law for cities: A cross country investigation. *Regional Science and Urban Economics*, 3.
- Schaffar, A.* (2006): On Zipf's law: Testing over China's and India's city-size distribution. *Université de La Réunion and Université de Toulon*
- Terra, S.* (2009): Zipf's Law for cities: On a new testing procedure. *CERDI Working Papers no. 200920*.
- Warren, D.* (2007): Examining City Size Distributions using Urban Areas. 45th Annual Meeting of the Western Regional Science Association
- Wheaton, W. – Shishido, H.* (1981): Urban concentration, agglomeration economies and the level of economic development. *Economic Development and Cultural Change*, 5.
- Zipf, G. K.* (1949): *Human behavior and the principle of least effort*. Addison-Wesley, Oxford

Kulcsszavak: Zipf-törvény, rang–nagyság szabály, Pareto-eloszlás, város, empirikus becslés.

Resume

The size-distribution of cities follows a simple rule, called the Zipf's Law, which can be considered a special case of Pareto-distribution, where the shape parameter equals 1. In this study after a short summary of the theoretical and methodical background of the concept which has been so far missed from the Hungarian literature, the author tests for Zipf's Law in 102 countries. Using three parametric estimation methods, my results showed that, even using the most compliant method, the Zipf's Law only holds in 70 countries. Accordingly the Zipf's Law is not valid generally, but in a narrow interval containing 1, the rank-size rule does hold.