

EÖTVÖS LORÁND TUDOMÁNYEGYETEM TERMÉSZETTUDOMÁNYI KAR

FÖLDRAJZ- ÉS FÖLDTUDOMÁNYI INTÉZET

Földrajztudományi Központ

Regionális Tudományi Tanszék

A településméretek eloszlási viszonya; hamis hatványtörvény?

CZALLER LÁSZLÓ

Tudományos Diákköri Dolgozat

Témavezető:

FARKAS MÁTÉ, PhD-hallgató



BUDAPEST, 2013

A településméretek eloszlási viszonya; hamis hatványtörvény?

(Size distribution of settlements; A false power-law?)

CZALLER LÁSZLÓ

Absztrakt

A településméretek eloszlásának kutatása számos olyan kérdést vet fel, melynek egyaránt léteznek elméleti és statisztikai-módszertani vonzatai. Az illeszkedésvizsgálatok kapcsán felmerülő problémák gyakorlati kezelésére is többféle megoldás ismert, viszont ezek függvényében az eredmények is különféleképpen interpretálандók. Ebben a tanulmányban két olyan problémát járok körbe, melyek precíz megoldása a rang-nagyság koncepciót, illetve a településméretek hatványtörvény-jellegét is megkérdőjelezi. Az első komoly problémát a vizsgálati alapegységek megválasztása jelenti, illetve annak eldöntése, hogy mely egységeket vonjuk be az elemzésbe. Az utóbbi kérdés különösen fontos, hiszen a településméretek eloszlásának vizsgálata tradicionálisan csak az eloszlás „felső farok” részére, azaz a városokra terjed ki. A települések teljes körét lefedő vizsgálatok azonban már egyre inkább elvetik a Pareto-eloszlást, ami arra enged következtetni, hogy a településméretek eloszlása csupán egy hamis hatványtörvény, ami sok esetben más eloszlásokat takar.

Ennek felderítésére először a Pareto-eloszlás illeszkedését vizsgálom Magyarország és az Amerikai Egyesült Államok településeinek teljes körén, illetve különböző népességeküszöbök alkalmazásával kialakított változó elemszámú mintákon nem-parametrikus tesztek segítségével. Eredményeim szerint a Pareto-eloszlás alkalmatlan a városméretek eloszlási viszonyainak modellezésére, legjobb illeszkedését olyan népességeküszöb mellett mutatja, melybe már számos kisebb község is beleesik. A vizsgálat második felében a Pareto-eloszlást egy bizonyos körülmények között igen hasonló lefutást eredményező eloszlástípussal, a lognormális eloszlással hasonlítom össze. A vizsgálat alapját képező likelihood-arány tesztek eredményei mindkét ország esetén arra engednek következtetni, hogy a Pareto-eloszlás és a lognormális eloszlás magas népességeküszöbök mellett hasonlóan közelíti a településméretek eloszlását, egyre alacsonyabb népességeküszöbök alkalmazásával, illetve a településhálózat egészen viszont már a lognormális modell alkalmasabb a jelenség közelítésére.

Kulcsszavak: *település, város, méreteloszlás, Pareto-eloszlás, lognormális eloszlás*

Tartalomjegyzék

Bevezetés.....	4
1. A településméretetek eloszlása.....	6
2. A méreteloszlások vizsgálatának főbb elméleti problémái	7
2.1. A vizsgálati alapegységek meghatározása.....	8
2.2. Az eloszlás csonkolása	10
3. A településméretetek eloszlásának empirikus vizsgálata	12
3.1. Adatok és definíciók.....	13
3.2. A Pareto-eloszlás illeszkedésének vizsgálata	14
3.3. A Pareto- és lognormális modellek összehasonlítása	20
4. Elméleti következtetések.....	24
Összefoglalás.....	25
Függelék	27
Irodalomjegyzék.....	28

Bevezetés

A települések lakosságszámmal közelített méretének nagymértékű eltérései régóta foglalkoztatják a társadalom térbeli szerveződésével foglalkozó tudományok képviselőit, melyben kétség kívül nagy szerepe van annak is, hogy a térbeli kérdések közgazdasági főáramba történő integrálódása az elmúlt évtizedekben igencsak felgyorsult. Ismét előtérbe kerültek a teoretikus városgazdaságtan azon kérdései, melyek a települések fejlődésének, és méretbeli sokszínűségének kutatásához kötődnek.

Az egyik legtöbbet vizsgált kérdés a települések méret szerinti eloszlásához kötődik, melynek kapcsán mindmáig igen elterjedt az a megközelítés, miszerint bizonyos jól körülhatárolt térbeli keretek között a települések méretének eloszlása Pareto-eloszlással közelíthető, azaz egyfajta hatványtörvény szerint alakul. A mára klasszikussá vált koncepció annak ellenére uralja a földrajzi és közgazdasági irodalmat, hogy az elmúlt bő két évtizedben számos olyan empirikus vizsgálat látott napvilágot, ami megkérdőjelezi a Pareto-eloszlás alkalmasságát, nem beszélve azokról az illeszkedésvizsgálatok lefolytatásához köthető alapvető problémákról, melyek igencsak megnehezítik az eredmények értelmezését, összehasonlítását. A településméretekre történő függvényillesztés két legnehezebben kezelhető problémája a vizsgálati alapegységek megválasztása, illetve annak meghatározása, hogy hol csonkoljuk az eloszlást, azaz milyen népességeküszöb felett vagy alatt vizsgáljuk a megfigyelési egységeket. Mindezen nehézségek gyakorlati kezelése statisztikai módszerekkel lehetséges, azonban a nehézséget éppen az okozza, hogy ezek a problémák nem csupán módszertani jellegűek, hanem komoly elméleti vonzatai is vannak. Szorosan köthető hozzájuk például a városok definíciós problémája, melynek feloldása azért lenne különösen fontos, mert a függvényillesztés során az eloszlást a rendszerint csonkolják, így csupán az eloszlás „felső farok” része kerül a vizsgálatba. Sok esetben pedig ennek eredményeiből következtetnek a funkcionális városok eloszlási viszonyaira, holott a városok népességeküszöbökkel történő lehatárolása meglehetősen pontatlan és önkényes, ráadásul a vizsgálatok igen érzékenyek a csonkolási értékek meghatározására.

A településméretek eloszlásának hatványtörvény-jellegét kritizálók szerint a klasszikus modell a csupán a legnagyobb városokra illeszkedik, amint lejjebb haladunk a településlejtőn, a Pareto-eloszlás alakparaméterének értéke csökken, és az eloszlás illeszkedése romlik. A települések szélesebb körére kiterjesztett vizsgálatok kapcsán tehát más eloszlások is szóba jöhetnek, mint a különböző típusú exponenciális eloszlások (Laherrere–Sornette 1998, Malacarne et al 2001), a lognormális eloszlás (Eeckhout 2004), vagy az előzőek

kombinálásával létrehozott összetett eloszlások (Reed – Jorgensen 2004, Giesen et al. 2010, Ioannides – Skouras 2012). A városméreték eloszlásának témáját érintő eddigi legélénkebb tudományos vita a Pareto- és a lognormális eloszlások illeszkedése vel kapcsolatos (Eeckhout 2004, 2009, Levy 2009, Malevergne et al. 2011, Bee et al. 2011), melynek a fent említett problémák fontos sarokkövei.

A Pareto- és a különféle konkurens eloszlások nagyon hasonló viselkedése, illetve az illeszkedés jóságának érzékenysége a csonkolási értékre azt sugallja, hogy a települések méret szerinti eloszlása – Perline (2005) terminológiáját felhasználva – a klasszikus felfogással ellentétben nem nevezhető „erős”, azaz az eloszlás egészére érvényes hatványtörvénynek, legfeljebb „gyenge”, vagy „hamis” hatványtörvénynek. Az utóbbi két jelző arra utal, hogy a vizsgált jelenség lefutását tekintve igen hasonló a hatványtörvényekhez, de más eloszlást, például nyújtott exponenciális, lognormális vagy egyéb összetett eloszlást takar (Perline 2005).

Jelen dolgozatomban azt a kérdést járom körbe, hogy milyen következtetések vonhatók le a függvényillesztések eredményeiből annak függvényében, hogy a legnagyobb településeknek (városoknak), avagy a településhálózat egészének eloszlását szeretnénk közelíteni, illetve hogy a megfigyelési egységek és a csonkolás problémáját elméleti, avagy pusztán statisztikai szempontok alapján kívánjuk megoldani. Ennek érdekében két eltérő méretű ország, Magyarország és az Amerikai Egyesült Államok példáján vizsgálom a Pareto-eloszlás illeszkedését különféle népességszűbök beiktatásával, majd a teljes településhálózaton. Ezt követően a klasszikus modellt a lognormális eloszlással hasonlítom össze. Az illeszkedési vizsgálatokat és a modellek közvetlen összehasonlítását egy olyan nem-parametrikus és parametrikus tesztekkel álló algoritmus szerint végzem el, melyet Clauset és szerzőtársai (2009) javasoltak a természetben és társadalomban előforduló vélt vagy valós hatványtörvények tesztelésére, s melyet mindeddig tudomásom szerint még nem használtak a településméreték eloszlásának vizsgálatára.

A dolgozat felépítése mindezek alapján a következő logikát követi; az első és második fejezetben a településméreték eloszlásának klasszikus hatványtörvény-megközelítését, illetve azon elvi és módszertani problémákat foglalom össze, melyek az egész településhálózat vizsgálatának szükségességét, illetve más eloszlások megfontolását pedzegetik. A harmadik fejezetben az empirikus vizsgálatok elvégzésére kerül sor: először a Pareto-eloszlás tesztelését, majd a rivális eloszlások összehasonlító vizsgálatát végzem el, melyet követően a főbb következtetéseket, illetve a vizsgálatom elméleti hozzájárulásait foglalom össze.

1. A településméretek eloszlása

A településméretek eloszlási viszonyainak leírására vonatkozó első törekvések Auerbach (1913) és Singer (1936) nevéhez köthetők, akik szerint adott ország településeinek mérete egyfajta hatványtörvény, a Pareto-eloszlás szerint alakul, melynek eloszlás- és sűrűségfüggvénye az (1) és (2) egyenletekkel írható fel,

$$F_p(x | \alpha, k) = 1 - \left(\frac{k}{x}\right)^\alpha \quad (1),$$

$$f_p(x | \alpha, k) = \frac{\alpha k^\alpha}{x^{\alpha+1}} \quad (2),$$

ahol x az egyes települések népességét jelöli, α a Pareto-kitevőnek is nevezett alakparaméter, k pedig a vizsgálatba bevont legkisebb település népessége, azaz egy olyan skálaparaméter, amely alatt a valószínűségi változó nem vehet fel értékeket. A Pareto-eloszlás egyik jellegzetessége, hogy a települések méretét és rangját, azaz a méret szerint vett rangsorban elfoglalt helyét log-log ábrán feltüntetve közel egyenes vonalat kapunk.

A Pareto-eloszlás egyik kitüntetett formája a Zipf-törvény (vagy rang-nagyság szabály), melynek érvényesülése esetén a kitevő értéke $\alpha = 1$. Ez a szűken értelmezett törvényszerűség azonban korántsem nevezhető általánosnak a világ országaiban. A Zipf-törvény érvényesülésével foglalkozó több országra kiterjesztett tanulmányok eredményei mind azt bizonyítják, hogy bár az egységnyi Pareto-kitevő számos országban alkalmas a legnagyobb települések eloszlásának leírására, azonban a kivételek száma túl magas ahhoz, hogy a Zipf-törvényt egy általános szabályszerűségnek tekinthessük (Rosen – Resnick 1980, Soo 2005, Czaller 2012). Mindazonáltal ezek az eredmények is szigorú fenntartásokkal kezelendők, hiszen ahogy a későbbiekben látni fogjuk, a Pareto-kitevők értéke igen érzékeny arra, hogy a skálaparamétert milyen népességszámban határozzuk meg.

Az empirikus tények és a paraméterek becslésénél felmerülő módszertani nehézségek ellenére a földrajzi és közgazdasági irodalom mindmáig paradigmaként tekint a Zipf-törvényre, mint a (legnagyobb) települések eloszlását igen jól közelítő modellre. A népszerűség oka abban keresendő, hogy a Zipf-törvény egy matematikailag egyszerűen felírható és interpretálható függvényformát takar, ráadásul mindeddig nem sikerült kielégítő magyarázatot találni arra, hogy több országban és eltérő időpontokban a városok méret szerinti rangsora miért hasonlít kísértetiesen az egységnyi kitevőjű Pareto-eloszlás lefutására. Az egyszerűség és a magyarázatok hiánya pedig együtt kivételes figyelemfelkeltő hatással bír, s már önmagában érdekessé teszi a városok eloszlásának alakulását (Benguigui – Blumenfeld-

Liberthal 2011). Krugman (1996) például a Zipf-törvényt olyan misztikus szabályszerűségnek nevezi, melynek kapcsán minden eddigi jól felépített közgazdasági modell csődöt mondott, Fujita és szerzőtársai (1999) valamint Gabaix (1999) pedig nemes egyszerűséggel megdöbbentő szabályszerűségnek nevezik a koncepciót. Tovább árnyalja a képet az is, hogy a különböző hatványtörvények, s köztük a Pareto-eloszlás igen gyakran megfigyelt eloszlástípusok mind a természetben, mind a társadalom komplex rendszereiben (Newman 2005). Azok a társadalmi törvényszerűségek pedig, melyek a természetben tapasztalt jelenségekkel valamilyen módon analógiát mutatnak, általában nagy figyelmet kapnak, miközben a tudományos identitás erősítését is szolgálják. A Pareto-eloszlás népszerűsége mögött megbúvó másik ok a praktikussági tényező. Mivel az eloszlásfüggvény kezelhető számú paramétert tartalmaz, képlete egyszerű és jól interpretálható, bizonyos megkötések mellett előáll sztochasztikus folyamatok végeredményeképpen (Gabaix 1999), így tehát alkalmassá válik arra, hogy a térbeliséget vizsgáló közgazdasági modellek alapjaként szolgáljon. Nem beszélve arról, hogy a Pareto-kitevő értéke egyfajta urbanizációs mérőszámnak is felfogható, hiszen adott ország településhálózatának koncentráltóságáról is biztosít némi információt (Cheshire 1999). Minél alacsonyabb a kitevő értéke, a vizsgálatba vont településhálózat a népesség tekintetében annál koncentráltabb, illetve minél magasabb az érték, a hálózat annál kiegyenlítettebb.

Mielőtt azonban a Pareto-kitevők értékéből bármilyen következtetést levonnánk, figyelembe kell venni azt, hogy az egyes országokban a közigazgatási térfelosztás nem mindenhol azonos kritériumok alapján van megszervezve, így egyrészt az összehasonlíthatóság is nehézkessé válik, másrészt a vizsgálatok a legtöbbször nem terjednek ki a településhálózat egészére, hanem rendszerint csak a városokra, ami számos problémát generál.

2. A méreteloszlások vizsgálatának főbb elméleti problémái

A települések méreteloszlásának elemzése összetettebb megközelítést igényel, mint a természetben megfigyelhető legtöbb jelenségek azonos jellegű vizsgálata. A társadalmi jelenségek kapcsán ugyanis folyton belebotlunk a definíciós problémákba, melyek sok esetben nehezen átléphető korlátokat állítanak a vizsgálati lehetőségek elé. Ahogy a jövedelemegyenlőtlenségek kapcsán jogos problémaként merül fel az, hogy az alapegységek az egyének, vagy a háztartások legyenek-e, vagy hogyan értelmezzük a jövedelem fogalmát,

úgy a településméretek kapcsán is megjelennek olyan kérdések, minthogy milyen térbeli keretek között vizsgálódunk, vagy egyáltalán mi az elemzésünk tárgya, alapeleme.

A településhálózatok kapcsán 3+1 némileg könnyen elhatárolható problémakört nevezhetünk meg. Az első annak a térrésznek a meghatározásához kötődik, melyen belül vizsgáljuk a településként vagy városként definiált területi egységeinket. A második az alapegységek megválasztásával, ehhez szorosan kapcsolódva a harmadik pedig az eloszlás csonkolásával kapcsolatos. A negyedik probléma az időbeliséghez fűződik, mely főként akkor áll fent, amikor időben dinamikus vizsgálatokat végzünk. Ekkor ugyanis egyaránt figyelmi kell a térbeli keretek esetleges módosulására éppúgy, ahogy az alapegységek számának és kiterjedésének változásaira. A következőkben részletesebben a második és a harmadik problémát bontom ki, hiszen ahogy a tapasztalat mutatja, az illeszkedésvizsgálatok kapcsán ezek tűnnek a legnehezebben kezelhető kérdéseknek.

2.1. A vizsgálati alapegységek meghatározása

A térbeli folyamatok elemzésének végkimenetelére minden esetben rányomják bélyegüket a térfelosztás sajátosságai, illetve a vizsgálat megfigyelési egységeinek megválasztása. Ez a két gyakorlati nehézség a településméretek eloszlása kapcsán is felmerül, ugyanakkor a térfelosztási és a módosító területi egység problémáknál jóval tovább mutat.¹ A vizsgálatok célja egészen pontosan a földrajzi térben megjelenő népességgkoncentrációk eloszlási viszonyainak felderítése, amelyhez nem elég pusztán felosztanunk a teret, azt is meg kell határoznunk, hogy mely egységek nevezhetők ténylegesen koncentrációnak.

A népességi góccokat reprezentáló alapegységek megválasztása korántsem egyszerű feladat, ugyanis az egyes közigazgatási egységek, illetve a közigazgatási határokkal nem feltétlenül konzisztens nagyvárosi agglomerációk vizsgálata kapcsán eltérő eredmények várhatók. Rosen és Resnick (1980), majd később Soo (2005) és Nitsch (2005) több országra kiterjedő elemzésükben rávilágítottak arra, hogy a több település összevonásával létrehozott agglomerációk méret szerinti eloszlása jóval koncentráltabb, mint a közigazgatási egységeké, tehát az egyes metropoliszok vonzáskörzetébe tartozó szuburbán térségek nagysága erőteljesen befolyásolja a Pareto-kitevő értékét. Brakman et al. (2001) az Amerikai Egyesült Államok városi térségeinek eltérő beosztása kapcsán hasonló eredményre jutott.

¹ A társadalmi tér vizsgálata kapcsán felmerülő statisztikai dilemmákról lásd Dusek (2004) tanulmányát, a módosítható területi egység problémájáról részletesebben Openshaw (1984) munkáját.

A városi agglomerációk létrehozása azonban nem mindenhol lehetséges, illetve sok helyen értelmetlen. Tipikusan ilyenek a kisebb területű, illetve a ritkán lakott országok, ahol csak néhány nagyobb agglomeráció hozható létre kisebb egységek összevonásával, így ezekben az esetekben a megfigyelések elemszáma sokszor nem elégséges a vizsgálatok lefolytatására. Mindezek alapján a közigazgatási határokra való hagyatkozás általános a méreteloszlások vizsgálatában, ami sok esetben megfelelő kereteket biztosít, azonban a városfejlődés előrehaladott szakaszában lévő fejlett országok esetén torz képet mutathat. Kosmopoulou és szerzőtársai (2007) rámutattak arra, hogy évtizedekkel ezelőtt az adminisztratív településhatárok még alkalmasak voltak a legnagyobb városok eloszlási viszonyainak vizsgálatára, azonban napjainkban ez a beosztás már nem állja meg a helyét. A vándorlás költségeinek csökkenésével ugyanis a szuburbanizáció és az 'urban sprawl' jelensége jelentősen megnövelte a népességagglomerációk valódi földrajzi kiterjedését, melyek már jócskán átlépik az adminisztratív határokat (Kosmopoulou et al. 2007). A méreteloszlások elemzésékor tehát mindenképp szerencsésebb, ha olyan lehatárolást választunk, ami funkcionális egységek, illetve valós gazdaságföldrajzi szerveződés alapján létrejött nagy agglomerációk méretét és kiterjedését jól közelítik (Cheshire 1999).²

Ennek érdekében több kísérlet is született arra, hogy a települési szintűnél részletesebb bontású adatok aggregálásával hozzanak létre olyan magas népsűrűségű körzeteket,³ melyek függetlenek a közigazgatási határoktól (Duranton 2005, Rozenfeld et al 2011). Az egyik legkomolyabb ilyen jellegű munka Rozenfeld és munkacsoportjának (2011) nevéhez köthető, akik az Amerikai Egyesült Államok és Nagy-Britannia egészét lefedő nagy térbeli részletességű adatait úgy aggregálták, hogy figyelembe vették a szomszédos egységek népsűrűségét, ami egy bizonyos szintet meg kellett, hogy haladjon, illetve az összevonás után számolt közös népsűrűséget soha nem engedték egy bizonyos előre meghatározott érték alá csökkenni. Az így kapott kompakt népességklaszterek legnagyobb előnye, hogy tetszőlegesen jól közelíti a népességkoncentrációk valódi földrajzi kiterjedését, azonban attól függően, hogy a népsűrűségi paramétert milyen értékben határozzuk meg, egyfelől komoly teret enged a szubjektív megítélésnek, másrészt nem fedi le feltétlenül a teljes teret. A szerzők a klasszikus modell szerint megbecsülték a Pareto-kitevő értékét az így kapott klaszterekre, ami a Zipf-törvénynek megfelelően egyhez közeli értéket vett fel.

² Ezen megállapítás kapcsán széles a szakmai konszenzus, ami az eloszlások vizsgálatában meglehetősen ritka jelenségnek számít.

³ Ilyen nagy felbontású területi beosztások például a népszámlálási körzetek, a telektömbök, vagy a postai körzetek.

A megfigyelési egységek képzésének egy másik érdekes megközelítése Holmes és Lee (2010) munkájának gyümölcse. A két szerző úgy függetlenítette az elemzést a közigazgatási határok károsan befolyásoló hatásától, hogy az Egyesült Államok területét 6x6 kilométeres cellákra osztotta fel, majd ezeknek vizsgálta az eloszlását. Eredményeik azonban nem meglepő módon nem voltak alkalmasak a Zipf-törvény kimutatására, hiszen 36 négyzetkilométeren fizikai képtelenség annyi lakost összezsúfolni, amennyi kellene ahhoz, hogy a legnagyobb egységek megfelelően simuljanak az egységnyi kitevőjű Pareto-eloszláshoz.

Ezek az eredmények arra engednek következtetni, miszerint a megfigyelési egységek megválasztása jelentős hatást gyakorol a becslések végkimenetére, továbbá azt sugallják, hogy a Zipf-törvény azon felosztások esetén érvényesül, melyek a lehető legjobban közelítik a népesség valós földrajzi koncentrációjának kiterjedését. Az ilyen megállapítások azonban mégsem fogadhatóak el általános érvényű igazsággként, ugyanis az is fontos kérdésként merül fel, hogy milyen szelekciós kritériumok alapján határozzuk meg azt, hogy mi számít népességgel koncentrációnak. Jogos kérdést vet fel például, hogy egy 100 fős községet érdemes-e egy vizsgálatba vonni egy milliós metropolisszal, mely a népességszámbeli eltérés mellett gazdasági szerkezetében, a földhasznosítás vagy a kulturális tényezők jellegében, illetve a közigazgatási funkciók meglétében is lényeges különbségeket mutat.

2.2. Az eloszlás csonkolása

Az elemzés alanyainak megválasztásához szorosan kötődő csonkolási probléma az eloszlásvizsgálatok neuralgikus pontja, ahogy arra a fenti példa is egyértelműen rámutat. Ha nem a teljes sokaságot kívánjuk vizsgálni, minden esetben meg kell határoznunk egy olyan alsó értékhatárt, mely felett valóban beszélhetünk a népesség „koncentrációjáról”. Cheshire (1999) három olyan egymással összefüggésben lévő szelekciós kritériumot nevez meg, ami alkalmas lehet a probléma megoldására. Az első és második állandó népességgküszöbök, vagy elemszámok meghatározását, a harmadik pedig az össznépesség bizonyos hányadának bevonását jelenti. Az első két módszer szoros kapcsolatban áll egymással, hiszen minél magasabb értékben határozzuk meg a népességgküszöböt, az elemszám annál kisebb lesz, illetve minél több települést vonunk be az elemzésbe, értelemszerűen az alsó népességkorlát értéke csökken. Hasonló logikával a harmadik megoldás is összefüggésbe hozható az első kettővel, éppen ezért bármely módszert is alkalmazzuk, az eloszlást valamilyen értéknél csonkolni kell.

Az eloszlások csonkolásának gyakorlata a kezdetek óta fontos eleme a méreteloszlások vizsgálatának, ami az adatok elérhetőségével, illetve a kutatói kérdésfeltevessel is összefüggésbe hozható. Auerbach (1913) például az 1910-es németországi népszámlálási adatok alapján számolt, de csak a 100.000 lakosnál népesebb 94 város példáján mutatta be az eloszlás hatványtörvény-jellegét. Mondanivalója megerősítése végett bár megismételte a vizsgálatot 481 településre is, de így is elhagyta a megfigyelések több mint 90%-át. Singer (1936) kifejezetten a városok eloszlási viszonyainak alakulását tanulmányozta változó népességszűbök mellett, melynek eredményeiből hét ország urbanizáltságának fokáról vont le következtetéseket.⁴ A későbbiekben általánossá vált az a gyakorlat, hogy a vizsgálatokat csak a városokra végezték el, azaz a teljes településhálózat eloszlásának legfelső tartománya került csak górcső alá. Ennek ellenére nem létezik konszenzus még egy-egy ország kapcsán se arról, hogy milyen népességi érték alatt kell csonkolni az adatsort (Gabaix – Ioannides 2004).

A küszöbértékek megválasztásának kérdése azért kritikus jelentőségű, mert egyrészt szoros kapcsolatban áll a népességkoncentrációk – vagy városok – méret szerint történő meghatározásával is, ráadásul elég erőteljes hatással van az illesztett függvények paraméterértékeinek alakulására. A Pareto-eloszlás alakparaméterére irányuló vizsgálatok tanulsága szerint például minél alacsonyabb értéknél történik az eloszlás csonkolása, annál alacsonyabb Pareto-kitevőket kapunk (Parr 1976, Guérin-Pace 1995, Eeckhout 2004), továbbá alacsony vizsgálati elemszámnál egy-egy megfigyelési egység elhagyása komoly változásokat idéz elő a kitevő értékében, ami megbízhatatlanná teszi a becslést. Ez a probléma egyes szerzők szerint a Zipf-törvény szigorú megköötéseinek indokoltságát is kétségbe vonja (Benguigui – Blumenfeld-Liberthal 2011). Számos tanulmány emellett rámutat arra is, hogy a Pareto-eloszlás illeszkedése a csonkolási érték csökkentésével együtt romlik (Parr – Suzuki 1973, Malecki 1980, Guérin-Pace 1995). Malecki (1980) amerikai példán megmutatta, hogy a települések népességének és rangjának logaritmusára felírt regressziók magyarázott varianciája egyre csökken, ahogy a népességszűbök értékét csökkentjük. Hasonló módszert alkalmazva Guérin-Pace (1995) arra a következtetésre jutott, hogy alacsony küszöbértékeknél a Pareto-eloszlásnál megfigyelhető lineáris kapcsolat a méret és rang között torzul, a függvény lefutásában némi konkavitás is megfigyelhető.⁵

⁴ Valószínűsíthető, hogy H. W. Singer volt az első kutató, aki a magyar városhálózat Pareto-eloszlását vizsgálta 1933-as népességadatok alapján!

⁵ Ezek az eredmények jól szemléltetik a csonkolással kapcsolatos problémát, azonban meg kell jegyezni, hogy az OLS-becsléssel történő kitevőbecslés, illetve az illeszkedés jóságának determinációs együtthatóval történő mérése komoly módszertani hibákat hordoz magában. Erről lásd részletesebben Gabaix – Ioannides (2004), és Nishiyama et al. (2008) tanulmányait.

Mivel a Pareto-eloszlás megfelelő illeszkedése az elméleti megfontolások alapján rögzített csonkolási értékek mellett nem biztosítható, valamint az alakparaméter értékének bizonytalansága is elég nagy, ezért érdemes lehet a települések teljes körére elvégezni a függvényillesztést, vagy olyan statisztikai kritériumokat keresni a csonkolási érték megállapítására, mely alapján biztosítható a jó illeszkedés. Eeckhout (2004) azonos megfontolások alapján szakítva az addigi gyakorlattal, az Amerikai Egyesült Államok összes településére elvégezte az illeszkedési vizsgálatot, melynek eredményei a Pareto-eloszlás helyett lognormális eloszlást támasztottak alá, ami felveti annak a gyanúját, miszerint a településméretek eloszlásának felső tartományában megfigyelt Pareto-eloszlás csak látszat.

A csonkolási probléma tehát nem pusztán elméleti gyökerű, hanem komoly statisztikai vonzatai is vannak. A hatványtörvények vizsgálatakor ugyanis a csonkolás bevett szokás, ami más jelenségek esetén is sokszor ad hoc jelleggel, vagy log-log ábráról történő vizuális meggyőződéssel történik. Perline (2005) ezzel kapcsolatban felhívja a figyelmet arra, hogy az eloszlások csonkolása komoly információvesztéssel jár, és magában hordozza annak a veszélyét, hogy bizonyos jelenségek kapcsán rögtön hatványtörvényekről kezdünk beszélni, holott más értéknél történő csonkolás esetén, vagy az eloszlás teljes egészén vizsgálódva a hatványtörvények hipotézise elvetendő más eloszlástípusok javára. Perline ezeket a hatványtörvényeket a „gyenge”, vagy „hamis” jelzőkkel illeti, utalva arra, hogy a hatványtörvény-jelleg érzékeny a csonkolásra, illetve más típusú, de hasonló lefutású eloszlásokat takar. Ez a települések vizsgálata kapcsán is így van, függetlenül a definíciós és egyéb elméleti problémáktól. Ha tehát a települések eloszlását akarjuk vizsgálni, nem feltétlenül hagyatkozhatunk a kis településeket szisztematikusan kihagyó, kisebb mintákon végzett elemzésekre, ha viszont kifejezetten a legnagyobb városok eloszlására vagyunk kíváncsiak, a vizsgálatba vonandó egységeket a lehető legnagyobb körütekintéssel kell indokolni, hogy a kapott eloszlás valóban arra vonatkozzon, amire szeretnénk vonatkoztatni.

3. A településméretek eloszlásának empirikus vizsgálata

Az empirikus elemzések alkalmával mindegyik fent említett problémát fontos szem előtt tartani, illetve a vizsgálat kérdésfeltevéseivel összhangban kezelni. A következőkben Magyarország és a mindeddig legtöbbet elemzett ország, az Amerikai Egyesült Államok településhálózatát vizsgálom különféle népességszűbök mentén történő csonkolás mellett. Vizsgálatom kettős célt szolgál: Az egyik cél a két ország település-méreteloszlásának deskriptív jellemzése, illetve annak a hipotézisnek az igazolása, hogy a Pareto-eloszlás nem

feltétlenül alkalmas a teljes településhálózat modellezésére, de még a városok eloszlásának jellemzésére sem.

Az országok megválasztása teljesen tudatos. Magyarország egy kis területű, kevés valódi városi agglomerációval jellemezhető ország, mely történelmi fejlődése során több határváltozást és államberendezkedést megtapasztalt, míg az Amerikai Egyesült Államok egy nagy területű, többpólusú ország, ahol a mai piacgazdasági szerkezet több mint egy évszázados zavartalan fejlődés eredménye. A két ország tehát több tekintetben egymás ellentéte, ami a településhálózat eloszlási viszonyaira is hatással van, ugyanakkor leszögezem, hogy a vizsgálat nem összehasonlító jellegű. az eredmények a két ország saját kontextusában értelmezendők.

Az empirikus elemzés során főként a Clauset és szerzőtársai (2009) által kifejezetten a hatványtörvények tesztelésére javasolt becslési és modellszelekciós eljárásokra hagyatkozom, melyek matematikai részleteit csak a módszerek megértéséhez szükséges mértékig közlöm a tanulmányon belül, a mélyebb részleteket a függelék tartalmazza.

3.1. Adatok és definíciók

A vizsgálat térbeli keretei értelemszerűen a két ország jelenlegi határai,⁶ a megfigyelési egységeket pedig a két ország közigazgatási térfelosztásának legkisebb szintjei. Magyarország esetében ez 3152 darab településeket jelenti, az USA esetében pedig a „*places*”, névvel ellátott 29313 lakott helységet, melyek lehetnek közigazgatási jogállással, és helyi önkormányzattal rendelkező területi egységek, vagy statisztikai célból létrehozott, úgynevezett CDP-k, (*Census Designated Places*)⁷ (US Census Bureau 2010).

Az adatok mindkét esetben 2010-re vonatkoznak, így az időbeliség problémájától eltekinthetünk. Az évszámválasztásból fakadóan az Egyesült Államokra vonatkozó adatok a

⁶ Esetünkben az országhatárok, mint térbeli keretek azért tűnnek reálisnak, mert ezeken belül mind Magyarország, mind pedig az Amerikai Egyesült Államok tekintetében egységes, funkcionális településrendszerekről beszélhetünk. Az országhatárok területi keretként való felfogása azokban az esetekben viszont semmiképp sem megfelelő, ahol a vizsgálat tárgyát képező ország települései és más országok települései között a vándorlásban, kereskedelemben megnyilvánuló interakciók intenzitása az országon belüli szinthez hasonló. Ezekben az esetekben az országhatároknál tágabb keretet érdemes megszabni hasonlóan azokhoz az esetekhez, amikor egy adott ország nemrég szakadt le az anyaországról. Ilyenkor ugyanis a településrendszer kevésbé tekinthető fejlődésében kompaktnak, több szálon kötődik az anyaország hálózatához. Példaként szolgál erre Koszovó, vagy Montenegró esete. Bár a szakirodalomban főként az országokra történnek becslések, de az utóbbi időben több tanulmány is megjelent, ami szubnacionális szinteken (Giesen – Suedekum 2011, Garmestani et al 2008), vagy az imént vázolt megfontolásokból az országosnál tágabb keretek között vizsgálódik (Dimou – Schaffar 2009).

⁷ A *Census Designated Place* nevű egységek olyan jogállás nélküli területeknek számítanak, melyek nem tartoznak egyik önkormányzathoz sem. A CDP-k létrehozása és megszüntetése államonként van szabályozva. A 2010-ben számuk 9794 volt, azaz majdnem az összes helység egyharmada. A legutóbbi népszámlálás alapján ezek lakosságszáma 38,5 millió fő.

legutóbbi népszámlálás eredménytábláiból származnak,⁸ a hazai településhálózat adatai viszont a 2001-es népszámlálásból továbbvezetett, a KSH Helységnévtárában közzétett adatok.

A magyar települések és amerikai helységek használatát az indokolja, hogy egyfelől tartalmilag analógnak mondhatók, illetve ez volt az egyetlen olyan vizsgálati szempontból racionálisnak mondható felosztás, amihez hozzá tudtam jutni, illetve ami lefedi az országok teljes területét.

3.2. A Pareto-eloszlás illeszkedésének vizsgálata

Vizsgálatunk első lépéseként a Pareto-kitevő értékét becsüljük meg a két országra változó népességeküszöbök mentén történő csonkítással. A becsléshez a maximum likelihood módszert (MLE) alkalmazom, melynek képlete

$$\hat{\alpha} = n / \sum_{i=1}^n \ln \left(\frac{x_i}{k} \right) \quad (3),$$

ahol $\hat{\alpha}$ a kitevő becsült értéke, x_i az i -edik település népessége, k pedig a vizsgálatba bevont legkisebb település népessége, n pedig az elemszám.⁹ Az MLE legfőbb előnye, hogy egyszerű, s emellett az alakparaméterre aszimptotikusan torzítatlan és konzisztens becslést ad (Newman 2005, Clauset et al. 2009). A különböző népességi küszöbök mellett számolt Pareto-kitevők értékét az 1. ábra mutatja. Magyarország esetében az látható, hogy a Pareto-kitevők értéke a 10.000 főnél nagyobb küszöbértékek esetén körülbelül 1,2 és 1,4 között ingadozik, 6000 főig kismértékben csökken, 6000 és 1400 fő között lényegében 1,1-es érték körül stagnál, majd az 1400 főnél kisebb küszöbértékek esetén a kitevő meredeken kezd csökkenni, az egész településhálózat bevonásával az érték 0,23. Az Egyesült Államok esetén 50.000 fő feletti küszöböknel a kitevő 1,5 és 1,25 között fluktuál, majd szintén meredek csökkenésbe kezd.

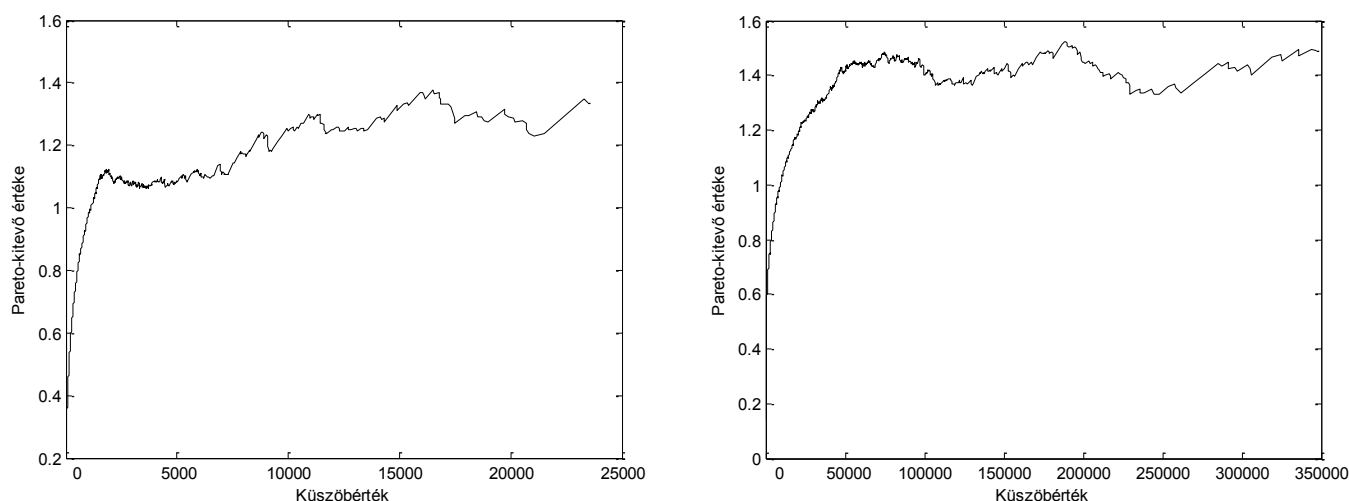
Ha tovább vizsgáljuk az ábrákat, feltűnhet, hogy mindkét ország esetén elvileg minden további nélkül megoldható úgy a csonkítás, hogy közel egységnyi Pareto-kitevőt tudjunk becsülni. Magyarország esetében például, ha a népességeküszöböt $k = 1139$ főnél, azaz Homokszentgyörgy község népességénél rögzítjük, és a becslést az első 1261 településre végezzük el, MLE módszert alkalmazva a kitevőre $\alpha = 0,999995$ értéket kapunk, amivel lényegében teljesítjük a Zipf-törvényben foglalt kritériumot, bár kétséges, hogy a küszöbérték

⁸ Elérhető: <http://2010.census.gov/2010census/data/>

⁹ A Pareto-kitevő becslőfüggvényének a levezetését a függelék „A” része tartalmazza.

meghatározására tudnánk-e érdemi indokkal szolgálni. Ugyanez eljátszható az Egyesült Államok esetében is: ha a küszöbérték a rangsor 4116. helyén álló Pennsylvania állambeli Lititz nevű település 9369 fős népességével azonos, a becsült értékünk $\alpha = 0,9997$. Természetesen ismét felmerül a kielégítő indoklás kérdése.

1. ábra: A Pareto-kitevők változása különböző küszöbértékek mellett Magyarországon és az Amerikai Egyesült Államokban



Megjegyzés: A bal oldali ábrán Magyarország, a jobb oldalin pedig az USA látható.

Az eredményekből látható, hogy a Pareto-eloszlás alakparamétereinek becslése körül igen nagy a bizonytalanság, azonban ez a függvény illeszkedéséről még nem mond semmit. Ennek federítését Kolmogorov-Smirnov tesztek segítségével végezzük el, mely elég közkedvelt próbának számít az illeszkedésvizsgálatokban. Az eljárás lényege, hogy az MLE módszerrel nyert alakparaméter segítségével megadunk az adatsorra egy olyan elméleti Pareto-eloszlásfüggvényt, mely elvileg a legjobban illeszkedik az adatokra, majd ezt az elméleti függvényt hasonlítjuk az adatsor tapasztalati eloszlásfüggvényéhez. Képletszerűen a következőképpen fest:

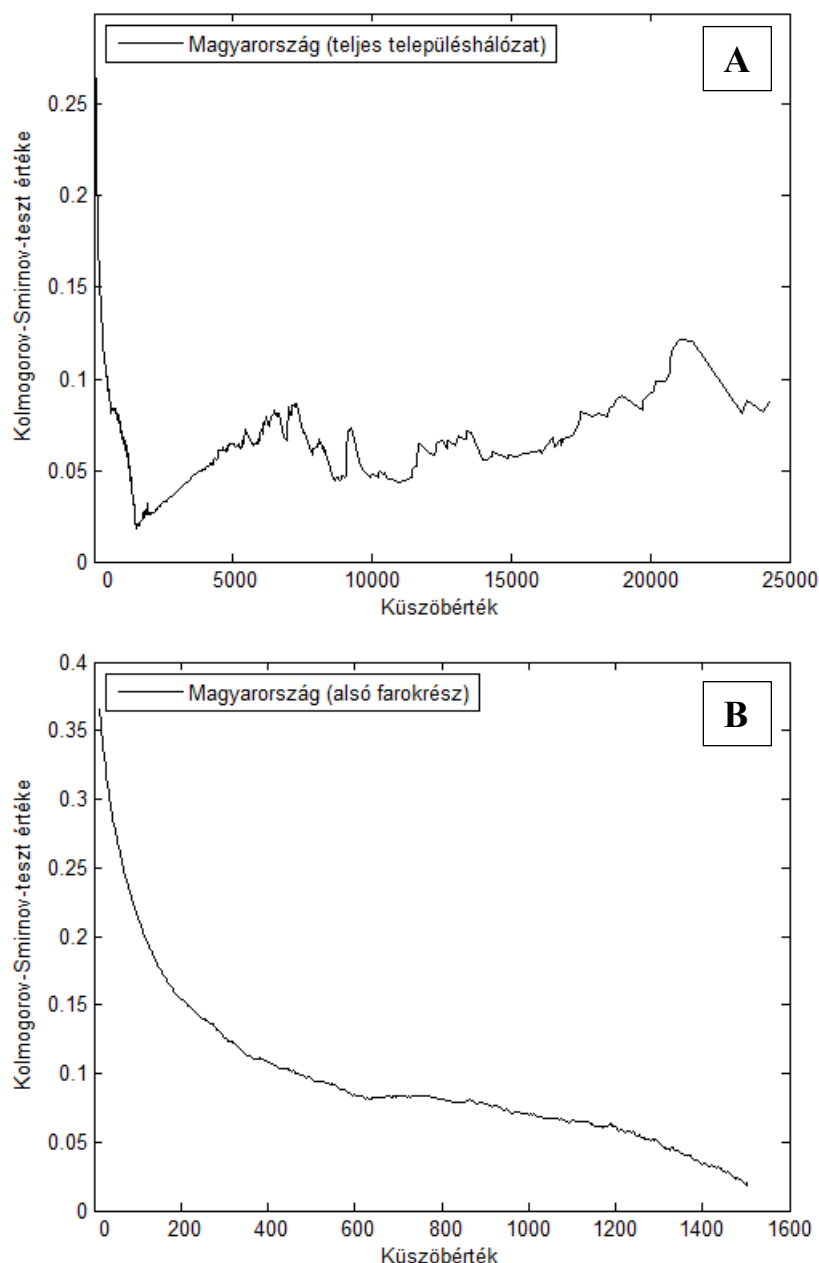
$$D = \sup_{X \geq k} |F(x) - P(x)| \quad (4)$$

A (4) egyenletben $F(x)$ jelöli a k értékét meghaladó megfigyelési egységek tapasztalati eloszlását, $P(x)$ pedig a legjobban illeszkedő Pareto-eloszlást jelenti, ugyanezen megfigyelési egységekre. A tesztstatisztika értéke a két eloszlásfüggvény közti eltérések abszolút értékének szupréuma, tehát lényegében a legnagyobb eltérés számértéke. Értelmszerűen minél kisebb a tesztstatisztika értéke, annál jobb az elméleti eloszlás illeszkedése az adatsorra.

A két országra eltérő népességekűsűbök mellett kiszámolt Kolmogorov-Smirnov tesztstatisztikákat a 2. és 3. ábrák foglalják össze. A 2. ábra „A” részében látható, hogy

Magyarország esetében a legjobb illeszkedés nem a magas küszöbértékeknél várható, hanem 1501-nél, ami egész pontosan Acsa község népessége. A legjobb illeszkedés tehát az elvártaakkal ellentétben nem a legnagyobb városokra a legjobb, de nem is a települések teljes körére, hiszen ha 1500-nál kisebb értéken csonkítunk, az illeszkedés először moderáltan, majd 600 alatt meredeken kezd romlani.

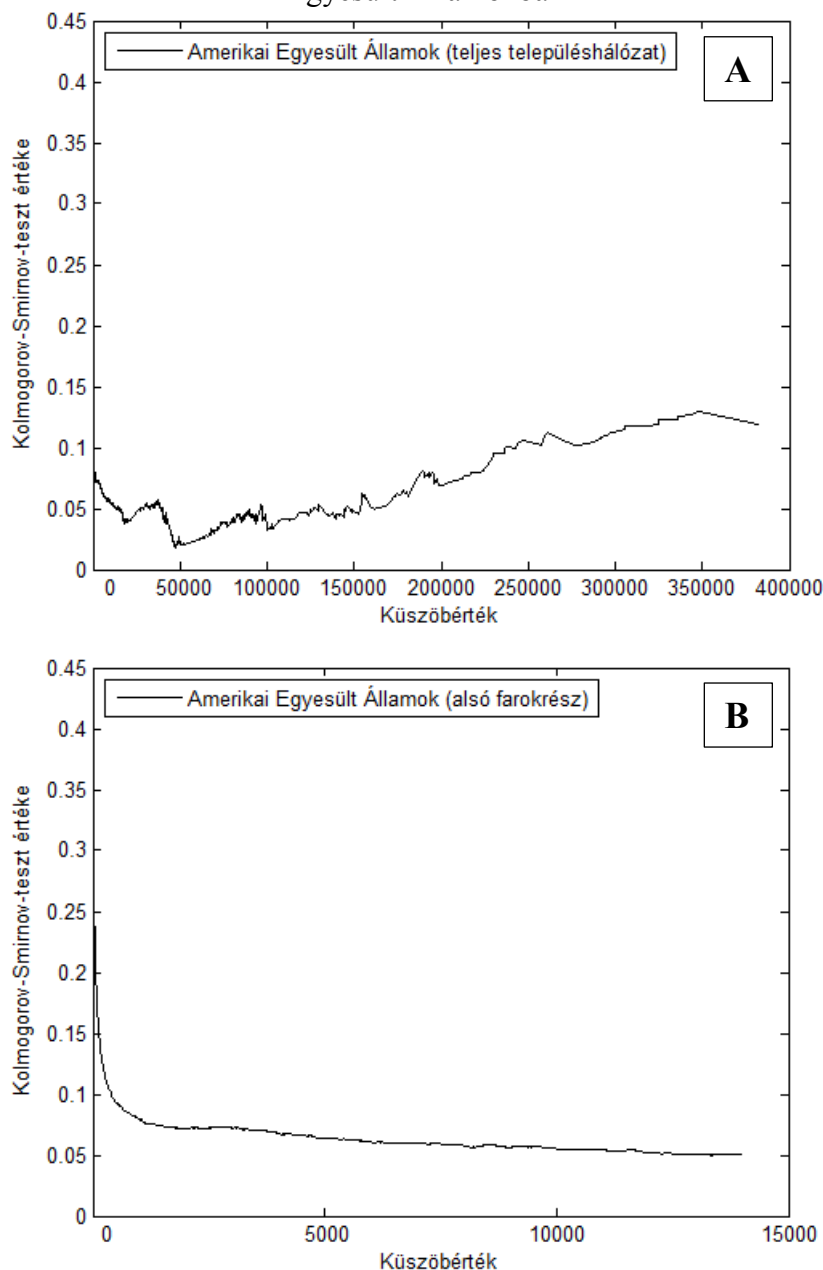
2. ábra: A Kolmogorov-Smirnov tesztek értékei változó küszöbértékek mellett Magyarországon



Az Egyesült Államok hasonló képet mutat, kisebb ingadozások mellett az elemszám csökkentésével az illeszkedés javul. A tesztstatisztika minimumát akkor kapjuk, ha a küszöbértéket 47299 főben, azaz a Michigan állambeli Roseville város lakosságszámában

határozzuk meg. Ennél alacsonyabb minimumértékeknél az illeszkedés tovább romlik, 1000 fő alatt rohamos mértékben. Az USA esetében tehát a Pareto-eloszlás legjobb illeszkedése az első 849 város esetén várható.

3. ábra: A Kolmogorov-Smirnov tesztek értékei változó küszöbértékek mellett az Amerikai Egyesült Államokban



A változó népességeküszöbök mellett kalkulált tesztstatisztikákból levonható legfőbb következtetés az, hogy minél több elemet vonunk be a vizsgálatba, a Pareto-eloszlás illeszkedése először javul, majd egyre rosszabb lesz. Ez összhangban van az eddigi empirikus tapasztalatokkal, azonban a tesztek önmagukban nem tudósítanak arról, hogy a Pareto-eloszlás valóban leírja-e az adataink eloszlását. A Kolmogorov-Smirnov teszt egyik

legnagyobb technikai előnye azonban az, hogy értékeinek segítségével ellenőrizhető a Pareto-eloszlás hipotézise is. A tesztstatisztika eloszlásának kvantilisei erre lehetőséget kínálnak, azonban ezek meghatározásánál fokozott figyelemmel kell eljárni. A Kolmogorov-Smirnov-próba eloszlására készített eredeti táblázatok ugyanis csak abban az esetben alkalmazhatók, ha a két vizsgált eloszlás független egymástól (Goldstein et al 2004). Esetünkben ez a helyzet nem áll fenn, ugyanis az eljárás folyamán az adatsorból becsüljük az elméleti eloszlásfüggvényünk paraméterét, így a becsült paraméter értéke, s ebből fakadóan az elméleti eloszlásfüggvény alakja függ az adatsortól. Ha ebben a helyzetben az eredeti táblázatok értékeire hagyatkozánk, a hipotézisvizsgálatok másodfajú hibájának valószínűsége megnőne, hiszen a Pareto-eloszlás nullhipotézisét esetleg olyan esetekben is elfogadnánk, amikor nem lenne indokolt. Ennek elkerülése végett a teszt kvantiliseinek újrabecslése javasolt, melynek több módja is létezik (Goldstein et al. 2004, Clauset et al. 2009, van Zyl 2011).

Az egyik legkézenfekvőbb megoldás a p-értékek meghatározására a Monte Carlo szimuláció alkalmazása, melynek lényege, hogy nagyszámú szimulált Pareto-eloszlású mintára egyenként alkalmazzuk a tesztstatisztikát, majd megvizsgáljuk, hogy azok milyen arányban haladják meg az adatsorunkra számolt értéket. Ez az arány pedig megadja az általunk keresett p-értéket, melynek 0,05-nél (0,1-nél) magasabb értékei esetén 5%-os (10%-os) szignifikanciaszint mellett elfogadhatjuk a Pareto-eloszlás nullhipotézisét (Clauset et al 2009).

A Monte Carlo módszerrel kalkulált p-értékeket néhány elméleti és statisztikai szempontból kiemelt küszöbérték mellett számoltam ki, ezeknek összefoglalása az 1. és 2. táblázatban található. Mindkét ország esetén megpróbáltam több olyan mesterséges népességi küszöböt meghatározni, melyek változó definíciók mellett alkalmasak lehetnek a városok körének lefedésére. Magyarország esetében a központi funkciók megléte alapján a 20.000, vagy 10.000 fős alsó népességhatárt állja meg a helyét leginkább, ennél kisebb küszöbök megadásának indoka az, hogy így számos kisebb városi jogállású település is bekerülhet a vizsgálatba. Az Egyesült Államok esetén a megfigyelési egységek értékei között több nagyságrendnyi eltérés tapasztalható, így itt értelemszerűen más küszöböket alkalmaztam. A 200.000 és 100.000 fős népességhatárok lefedik a legnagyobb városok körét. Az ennél kisebb küszöbök pedig a közép és kisvárosok bevonását szolgálják. Mindkét ország esetében mellékeltem továbbá azt a küszöbértéket, amely mellett a Pareto-kitevő becsült értéke a legközelebb esik 1-hez, illetve ahol a legkisebb a Kolmogorov-Smirnov tesztstatisztika értéke minimális. Ezzel a két népességhatárral kiválóan tudjuk tesztelni, hogy ott, ahol elméletileg

érvényesül a Zipf-törvény, illetve ahol az illeszkedés a legjobb, valóban beszélhetünk-e Pareto-eloszlásról.

1. táblázat: A Pareto-eloszlás illeszkedése változó küszöbértékek mellett a magyar településhálózatra

Népességeküszöb (fő)	Elemsszám (db)	Pareto-kitevő	KS-teszt	p-érték
20000	61	1,279	0,096	0,000
10000	144	1,241	0,049	0,000
5000	277	1,082	0,065	0,000
2000	759	1,102	0,026	0,042
1000	1394	0,971	0,071	0,000
1501 (Legjobb illeszkedés)	1037	1,095	0,018	0,389
1139 (Pareto-kitevő értéke = 1)	1262	1,000	0,065	0,000
12 (Teljes településhálózat)	3152	0,233	0,366	0,000

Magyarország esetén csak egy küszöbérték mellett fogadható el a Pareto-eloszlás nullhipotézise, ez pedig a Kolmogorov-Smirnov teszt alapján legjobb illeszkedést biztosító érték.¹⁰ Itt a Pareto-kitevő értéke 1-hez közeli, de a vizsgálatba vont 1394 település nem csupán városokat, hanem számos községet is magába foglal, tehát nem állíthatjuk, hogy ez az illeszkedés csak a legnagyobb agglomerációkra lenne érvényes. Ugyanakkor a teljes településhálózatra egyáltalán nem illeszkedik a Pareto-eloszlás, ezért az egészen biztosan alkalmatlan a teljes hazai hálózat, illetve a városok méreteloszlásának modellezésére. Mi több, a Zipf-törvény érvényesülését sem támaszthatjuk alá, hiszen annál a csonkolási értéknél sem fogadhatjuk el a Pareto-eloszlást, amely mellett a kitevő értékét 1-nek becsültük. Statisztikai értelemben ez azt jelenti, hogy az eloszlás illeszkedése a csonkolásra szélsőségesen érzékeny, ami elveti annak lehetőségét, hogy a hazai településhálózatot valódi hatványtörvényként definiáljuk.

Az Egyesült Államok településhálózata a magyarhoz képest eltérő eredményeket produkál, ami nem meglepő, tekintettel a hazainál kilencszer nagyobb elemszáma és a tekintélyes méretbeli különbségekre. Az első és legfontosabb levonható következtetés azonban mégis nagyon hasonló. Az amerikai településméretek eloszlása is érzékeny a csonkolásra, az egész hálózat nem írható le Pareto-eloszlással, illetve a három legkisebb mesterségesen megadott küszöbérték esetén sem támaszthatjuk alá a hatványtörvényt,

¹⁰ Értelemszerűen a legjobb illeszkedés sem jelenti minden esetben, hogy a Pareto-eloszlás hipotézise elfogadható. Az 1990-es népszámlálás adatain megismételve az elemzést például azt kapjuk, hogy a legjobb illeszkedés mellett a p-érték 0,05-höz igen közeli értéket vesz, tehát 5%-os szignifikanciaszint mellett a hipotézis elvetésének éppen a határán van. Ez ismét előtérbe hozza az időbeliség problémáját, tudniillik nem mindegy, melyik évre végezzük el az elemzést.

ugyanúgy, ahogy a legnagyobb településeket tömörítő 200.000 fős népességekűs mellett sem. A Zipf-törvény szintén nem igazolható, ugyanakkor a legjobb illeszkedést nyújtó, és az ahhoz igen közel lévő 50.000 fős küszöbértékek esetén a hipotézis elfogadható, tehát a nagy és közepes városok bizonyos körére illeszkedik a Pareto-modell. Ezzel párhuzamosan viszont annak a kérdése továbbra is fennáll, hogy beszélhetünk-e valódi, erős hatványtörvényről. Mivel a legkülönbözőbb csonkolási paraméterek közül csak kettő (egymáshoz igen közeli) esetén sikerült igazolni a hatványtörvény-jelleget, ezért itt is erős a gyanú arra, hogy az amerikai települések eloszlása kapcsán is inkább hamis hatványtörvényt találunk.

A hamis hatványtörvények valójában olyan egyéb eloszlásokat takarnak, melyek lefutása bizonyos tartományban nagyon hasonló a hatványtörvényekére. Ilyen eloszlások lehetnek többek között a nyújtott exponenciális, lognormális, „Pareto-végződésű” exponenciális, vagy egyéb összetett eloszlások.

2. táblázat: A Pareto-eloszlás illeszkedése változó küszöbértékek mellett az amerikai településhálózatra

Népességekűs (fő)	Elemsszám (db)	Pareto-kitevő	KS-teszt	p-érték
200000	109	1,442	0,069	0,000
100000	282	1,397	0,034	0,030
50000	779	1,416	0,021	0,321
20000	2123	1,182	0,040	0,002
10000	3937	1,021	0,055	0,000
5000	6455	0,870	0,064	0,000
2000	10976	0,718	0,072	0,000
47299 (Legjobb illeszkedés)	849	1,427	0,018	0,601
9369 (Pareto-kitevő értéke = 1)	4115	1,000	0,064	0,000
1 (Teljes településhálózat)	29312	1,408	0,408	0,000

3.3. A Pareto- és lognormális modellek összehasonlítása

Az illeszkedés érzékenysége és az alakparaméterek becslése körüli bizonytalanság kapcsán az elmúlt évtizedekben több függvényforma is terítékre került, mint a településméret eloszlását legjobban közelítő modell. Parr és Suzuki (1973) már egészen korán felhívta a figyelmet arra, hogy az eloszlások felső tartományában észlelhető Pareto-végződés nem zárja ki más eloszlások, például lognormális eloszlás hasonló illeszkedését. Hsing (1990) a Pareto-eloszlás tesztelésére a Box-Cox transzformáció alkalmazását ajánlotta, mellyel elvetette a szóban forgó eloszlás hipotézisét. Alperovich és Deutsch (1995) nemzetközi adatokon alkalmazta a módszert, mely szintén megcáfolta a Pareto-eloszlás

alkalmasságát. Cameron (1990) alternatív modellként a Weibull-eloszlást ajánlotta, Laherrere és Sornette (1998) pedig több természeti és társadalmi jelenség, köztük a települések méreteloszlása kapcsán rámutatott, hogy a nyújtott exponenciális eloszlás sok esetben jobban simul az adatokra, mint a hatványtörvények. Eeckhout (2004, 2009) úgy találta, hogy a települések teljes körére a lognormális modell a Paretonál jobban illeszkedik, azonban később több szerző is rámutatott, hogy a lognormális modell a legnagyobb megfigyelési egységekre kevésbé simul rá, az eloszlás legfelső tartományában a Pareto-modell jobban alkalmazható (Levy 2009, Bee et al. 2010, Malevergne et al. 2011, Ioannides – Overman 2012). Ugyanakkor ezek a vizsgálatok az eloszlás csonkolása nélkül, a településhálózat egészére vonatkoztak. Ahhoz viszont, hogy a települések hamis hatványtörvény-jellegének gyanúját cáfolhassuk, vagy megerősíthessük, meg kell győződnünk egyéb modellek illeszkedéséről is. Mivel a hatványtörvények lefutása bizonyos körülmények között sokszor összetéveszthető a lognormális eloszlásával, ezért ezt a két modellt hasonlítom össze a két ország településeinek példáján. A lognormális eloszlás sűrűségfüggvényét az (6) egyenlet mutatja,

$$f_L(x | \mu, \sigma) = \frac{1}{x\sigma\sqrt{2\pi}} \exp\left(-\frac{(\ln x - \mu)^2}{2\sigma^2}\right) \quad (6),$$

melynek μ és σ paraméterei az átlagot és a szórást mutatják. A Pareto- és a lognormális sűrűségfüggvények alapvetően elég különbözőek, viszont hasonlónak válnak, ha a logaritmikusan transzformált adatsor szórása kellően nagy értéket vesz fel. Ennek szemléltetésére tekintsük a Pareto- (2) és lognormális (6) sűrűségfüggvények logaritmusait:

$$\ln f_P(x | \alpha, k) = \ln \alpha + \alpha \ln k - (\alpha + 1) \ln x \quad (7)$$

$$\begin{aligned} \ln f_L(x | \mu, \sigma) &= -\ln x + \ln \sqrt{2\pi}\sigma - \frac{(\ln x - \mu)^2}{2\sigma^2} = \\ &= -\frac{(\ln x)^2}{2\sigma^2} + \left(\frac{\mu}{\sigma^2} - 1\right) \ln x - \ln \sqrt{2\pi}\sigma - \frac{\mu^2}{2\sigma^2} \end{aligned} \quad (8)$$

Ha a szórás (σ) értéke elég magas a lognormális sűrűségfüggvényben, akkor azon tagok értéke, melyek a szórásnégyzetet tartalmazzák a nevezőben, elég kicsivé válnak ahhoz, hogy az eloszlás nagy terjedelmében a (7) egyenlethez hasonlóan logaritmikus skálázású ábrán lineárisnak hassanak (Mitzenmacher 2003). Ez a hasonlóság azonban nem feltétlenül jelenti, hogy a két eloszlás azonos módon illeszkedik az adatokra. Ennek illusztrálására az ismertetett két ország településhálózatát használok fel, de az összehasonlító vizsgálatot nem csupán az összes településre, hanem különböző küszöbértékek segítségével kialakított kisebb

sokaságokra is elvégzem. Ehhez azonban a lognormális eloszlásfüggvény némi átalakításra szorul, ugyanis a széles körben alkalmazott kétparaméteres lognormális eloszlásfüggvény a Pareto-eloszlással ellentétben nem operál küszöbértékkel, ami a csonkolt kisebb elemszámú minták esetén jelentős torzulást eredményezhet a becsült átlagban és szórásban (Warren 2007).

Mivel jelen esetben az adatsorunk kis értékeit távolítjuk el, azaz balról csonkítunk, a (6) számú sűrűségfüggvény a következőképpen módosul:

$$f_L(x | \mu, \sigma, k) = \frac{1}{x\sqrt{2\pi\sigma^2} \left(1 - \Phi\left(\frac{\ln k - \mu}{\sigma}\right)\right)} \exp\left(-\frac{(\ln x - \mu)^2}{2\sigma^2}\right), k \leq x \leq \infty \quad (7)$$

ahol $\Phi(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^t \exp\left(-\frac{t^2}{2}\right) dt$ a standard normális eloszlás eloszlásfüggvénye. Ezzel az átalakítással a Pareto-eloszlás és a lognormális eloszlás összehasonlíthatóvá válik, akár csonkoljuk az adatsort valamilyen k népességeküszöbnél, akár nem. A csonkolt lognormális eloszlás MLE paraméterei nem adhatók meg analitikusan, azokat a likelihood-függvény numerikus maximalizálásával számíthatjuk ki. Az így kapott paraméterek segítségével pedig már megoldhatóvá vált a csonkolt lognormális görbék illesztése, illetve a modellek közvetlen összehasonlítása.

A Pareto- és lognormális eloszlások összehasonlítására kézenfekvő módszer a likelihood-arány tesztek alkalmazása (Clauset et al. 2009). Ehhez először az egyes modellek log-likelihood függvényének az értékeit ($\ln L_1$ és $\ln L_2$) számítjuk ki, melyeket aztán kivonunk egymásból $\mathcal{R} = (\ln L_1 - \ln L_2)$ ¹¹. A különbség előjele azt adja meg, hogy melyik modell alkalmasabb a vizsgált jelenség leírására. Ha az érték pozitív, akkor az első modellt preferálhatjuk, ha negatív, akkor a másodikat. Ha a különbség 0, vagy ahhoz nagyon közeli érték, akkor nincs különbség a két modell között. Mivel a log-likelihood értékek különbségének az elemszám (n) és szórás (σ) segítségével normalizált értéke, azaz $n^{-1/2}(\ln L_1 - \ln L_2)/\sigma$ normál eloszlást követ, könnyedén tesztelhető az a nullhipotézis, miszerint a log-likelihood értékek különbsége 0, azaz a két modell hasonlóan teljesít a jelenség leírásában (Vuong 1989).¹²

A korábban ismertetett küszöbértékek mellett számolt log-likelihood értékeket, ezek különbségét, valamint a megfelelő p-értékeket a 3. és 4. táblázat foglalja össze a két országra külön-külön. A magyar települések esetén a lognormális és Pareto-eloszlások egyedül a

¹¹ A logaritmikus transzformáció tulajdonságai miatt ez a módszer ekvivalens azzal, mintha a likelihood-függvény értékek arányát számoltuk volna ki.

¹² A teszt matematikai részleteit lásd a függelék „B” részében.

megfelelően alacsonyan meghúzott küszöbértékek mellett különböztethető meg, magas népességeküszöbök mentén történő csonkolás esetén a log-likelihood értékekben észlelhető minimális eltérések statisztikai fluktuációk hatásait tükrözik. Alacsony küszöbértékek mellett, illetve azok nélkül a log-likelihood értékek negatív előjele a második, azaz a lognormális modellt hozza ki győztesen. Az eredmények kapcsán talán a leginkább meglepő az, hogy azzal a küszöbértékkel számolva, amellyel korábban a Kolmogorov-Smirnov teszt alapján elfogadtuk a Pareto-eloszlást, mint potenciális modellt, sem tudtunk a lognormális alternatívánál hatásosabbnak lenni. A hazai településhálózat eloszlása ezek alapján tehát semmiképp sem tekinthető Paretonak, inkább hamis hatványtörvényként definiálható, hiszen létezik olyan egyéb modell is, ami nagyon hasonlóan, sőt, egyes esetekben jobban közelíti a méreteloszlást.

3. táblázat: A Pareto- és lognormális eloszlások összehasonlító vizsgálatának eredményei Magyarország esetén

Népességeküszöb (fő)	Log likelihood (Pareto)	Log likelihood (Lognormális)	Log-likelihood különbség	p-érték
20000	-697,776	-697,770	-0,001	0,965
15000	-1011,084	-1010,735	-0,035	0,112
10000	-1555,143	-1555,050	-0,090	0,839
5000	-2870,475	-2868,717	-0,276	0,381
2000	-7143,583	-7143,450	-0,130	0,793
1501 (Legjobb Pareto illeszkedés)	-9473,699	-9473,477	-0,220	0,721
1139 (Pareto-kitevő értéke = 1)	-11405,848	-11399,214	-6,634	0,059
12 (Teljes településhálózat)	-29069,505	-26724,907	-2344,598	0,000

4. táblázat: A Pareto- és lognormális eloszlások összehasonlító vizsgálatának eredményei az Amerikai Egyesült Államok esetén

Népességeküszöb (fő)	Log likelihood (Pareto)	Log likelihood (Lognormális)	Log-likelihood különbség	p-érték
200000	-1475,127	-1475,087	-0,040	0,872
100000	-3636,058	-3635,959	-0,099	0,803
50000	-9486,390	-9486,268	-0,121	0,807
20000	-24589,499	-24573,461	-16,038	0,005
10000	-43972,966	-43911,347	-61,619	0,000
5000	-69748,235	-69575,777	-172,458	0,000
2000	-113321,600	-112915,672	-405,928	0,000
47299 (Legjobb Pareto illeszkedés)	-10280,877	-10280,878	0,002	0,997
9369 (Pareto-kitevő értéke = 1)	-45887,394	-45812,784	-74,609	0,000
1 (Teljes településhálózat)	-294963,016	-267425,186	-27537,831	0,000

Az Egyesült Államok helységeit vizsgálva is hasonló következtetésekre juthatunk, mint a hazai települések esetében. Az eredmények összességében itt is azt mutatják, a települések eloszlási viszonyait magas küszöbértékek mellett a csonkolt lognormális eloszlás hasonlóan hatékonyan képes közelíteni, mint a Pareto-eloszlás, a 20.000 főnél alacsonyabb népességek mellett pedig jóval akkurátusabb. Ez természetesen nem jelenti azt, hogy a lognormális eloszlás közelítené a legjobban a jelenséget, azonban egyértelműen rávilágít arra, hogy a településméret eloszlása egy hamis hatványtörvény, mely valójában valamilyen egyéb összetett eloszlást követ.

4. Elméleti következtetések

A városfejlődéssel kapcsolatos kutatások mindeddig nem tudták egyértelműen azonosítani azokat a tényezőket, melyek a településhálózatok egészének eloszlási viszonyait meghatároznák számos olyan sztochasztikus modell került kidolgozásra (Gabaix 1999, Blank – Solomon 2000, Benguigui – Blumenfeld-Liberthal 2007), mely a települések fejlődésének folyamatát az eloszlás felső tartományában bizonyos hányadára bizonyos megkötések mellett képes magyarázni.

Ioannides és Skouras (2012) ugyanakkor megjegyzi, hogy ezek a modellek nem veszik számításba annak lehetőségét, hogy a települések fejlődését méretkategóriánként más és más, településenként eltérő jelentőségű tényezők befolyásolják. Eltérő fejlődési pályákat generálhat például szektoriális specializáció, mely a települések méretét is meghatározza (Black – Henderson 2003, Abdel-Rahman – Anas 2004). Henderson (1974) szerint a növekedést a tudás és technológia hatékonyabb ágazaton belüli áramlásából, a munkaerőpiac egyesítéséből, illetve a méretgazdaságosságból fakadó előnyök serkentik, melyek jellemzően ágazatspecifikusak. Az empirikus vizsgálatok emellett rávilágítanak arra is, hogy azok a városok, melyeknek ágazati szerkezetére a specializáció jellemző, rendszerint kisebbek és lassabb ütemben növekednek, mint azok, melyeknek szolgáltató és ipari szektora is egyaránt diverzifikált (Glaeser et al. 1992). Abdel-Rahman és Fujita (1990) szerint a külső méretgazdaságosság mértéke, mely az emelt szintű szolgáltatások iránti ipari keresletből, illetve a szolgáltató szektor növekvő mérethozadékából származik, az egyik legfontosabb tényezője az ipari koncentrálnak, melynek városnövekedést serkentő dinamikája városonként eltérő. Megemlítendő továbbá a vállalatok menedzsment és termelő funkcióinak

egyre erőteljesebben érzékelhető területi szétválása, mely szintén rányomja bélyegét a városfejlődés ütemére (Duranton – Puga 2005).

Ezek a folyamatok Ioannides és Skouras (2012) állítása szerint olyan bonyolult összetett eloszlásokat eredményeznek, melyek lefutása bizonyos küszöbértékeknél változik. Az érvelés alapján a településhálózat eloszlásának felső, kifejezetten városokat tartalmazó tartományában a Pareto-eloszlás jól illeszkedik, míg az eloszlás teste, mely kisvárosokat és községeket, azaz főleg rurális tereket foglal magába, egyéb eloszlással írható le. Jelen tanulmány eredményei azonban kétségbe vonják, hogy a gazdasági tényezők hatásai egyértelműen köthetők lennének egy-egy jól meghatározott eloszláshoz. A legnagyobb hazai városokra ugyanis rendre el kellett vetni a Pareto-eloszlás nullhipotézisét az illeszkedésvizsgálatok alkalmával, továbbá a települések azon köre, amire érvényes lett volna a szóban forgó eloszlás, már nem kizárólag városokat, hanem községeket is tartalmazott, melyeknél kétséges, hogy ugyanazok a tényezők jelölnék ki a fejlődési pályákat, mint a főváros, vagy a nagyobb regionális központok esetében.

Az Egyesült Államok kapcsán hasonló gondolatok fogalmazódhatnak meg bennünk a vizsgálat eredményeit látva, ugyanis ebben az esetben az 50.000 főnél népesebb városokra illeszkedik jól a Pareto-eloszlás, ugyanakkor ennél a küszöbértéknél számos több tízezres lakosságszámú középváros indokolatlanul ki van zárva a vizsgálatból. Ez ismét elveti annak lehetőségét, hogy a Pareto-eloszlást kifejezetten a városi terekre alkalmazzuk, hiszen ha csak a metropoliszokra vizsgálódunk, nem megfelelő az illeszkedés, ha a középvárosokat is beemeljük a mintába, akkor sem. Az eloszlások településtípusokhoz való hozzákapcsolásának koncepciója tehát átgondolandó.

Vizsgálataink egy másik fontos elméleti következtetése az, hogy erősen kétségbe vonható mindenfajta több országot magába foglaló empirikus vizsgálat eredménye, ugyanis azon felül, hogy a Pareto-kitevők értéke számos külső hatásra érzékeny, egyáltalán nem biztos, hogy a becsült kitevővel egy megfelelően jól illeszkedő Pareto-eloszlást definiálunk. A kitevők becslése – illetve azok „városodási indexként” történő alkalmazása – érvényét veszti, ha a települések méretének eloszlását megfelelően közelíteni képtelen modellt definiálnak.

Összefoglalás

A településméretek eloszlásának vizsgálata több olyan probléma is felszínre kerül, mely közvetlenül befolyásolja a vizsgálatok végeredményeit. Elméleti szempontból komoly kérdést

vet fel az, hogy a vizsgálatok megfigyelési egységeit miben állapítjuk meg, illetve, hogy milyen népességi küszöbérték mentén csonkoljuk az eloszlást. Magyarország és az Amerikai Egyesült Államok településhálózatának példáján végzett vizsgálatok számos szerző korábbi eredményeivel összhangban azt támasztják alá, miszerint a Pareto-kitevők értéke érzékeny a népességeküszöbök megválasztására, ez pedig lehetőséget kínál arra, hogy az eredményeket a kívánt cél érdekében a kutatók szabadon módosítsák. Egy másik kardinális – habár az adatok elérhetőségének szempontjából talán könnyebben megválaszolható – kérdés a vizsgálati időpontok megválasztásával kapcsolatos. Ekkor ugyanis nem csak a közigazgatási határok változására, hanem a vizsgálatba kerülő települések körére is figyelni kell. A településhálózat időben folyamatosan átalakul, így akármilyen eljárással számolunk, a kapott eredmények mindenképpen eltérőek lesznek. Az időbeliség és a vizsgált sokaság megválasztása tehát alapvetően aláaknázza a Zipf-törvény „egyetemes” érvényesülésének lehetőségét az egyes országokban.

Technikai szempontból közelítve a hatványtörvény-koncepció szintén több ponton támadható, hiszen ahogy a két ország közigazgatási térfelosztásának legalacsonyabb szintjén végzett elemzésünk eredményei egyértelműen rámutatnak, a Pareto-eloszlás illeszkedése nem feltétlenül a legnagyobb városokra a legjobb, hanem a települések egy olyan szélesebb körére, ami már községeket is magába foglal, vagy nem fedi le kielégítő mértékben a városok körét. Az eloszlásokkal kapcsolatos országos vizsgálódások kapcsán tehát a funkcionális nagyvárosok kitüntetett kezelése nem feltétlenül indokolt. Másrészt megmutatható az is, hogy a különböző küszöbértékek mentén csonkolt lognormális eloszlás hasonlóan, bizonyos körülmények között pedig kifejezetten jobban modellezi a méreteloszlásokat, mint a klasszikus Pareto. Így a települések kapcsán Perline (2005) terminusát használva hamis Pareto-eloszlásról beszélhetünk csupán, ami cáfolja a városméret eloszlásának hatványtörvény-jellegét, legalábbis országos léptékekben, az általunk választott megfigyelési egységekre vonatkoztatva.

A vizsgálat eredményei megkérdőjelezzik a Pareto-eloszlás egyeduralmát a városgazdasági szakirodalomban, valamint további kérdéseket vetnek fel azzal kapcsolatban is, hogy a több országra kiterjedő nemzetközi elemzések eredményei mennyire tekinthetők relevánsnak. A településméret eloszlására megfelelően illeszkedő – valószínűsíthetően valamilyen összetett – modell, valamint az eloszlás lefutását befolyásoló gazdasági tényezők kilétét továbbra is a homály fedi, azonban jelen tanulmány a meglepő eredmények miatt komoly adalékkul szolgálhat a téma további kutatásához.

Függelék

A.)

A Pareto-kitevő alakparaméterének maximum likelihood becslésőfüggvénye Czaller (2012) alapján a következőképpen adható meg: a független, $X \sim P(I)(\alpha, k)$ ($x_1, x_2, \dots, x_N$) minta likelihoodja a következő alakot ölti:

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

ahol k a korábbiakban megszokott módon x_i hipotetikusan felvehető legkisebb értéke: $x \geq k \geq 0$, és $k = \min(x_i)$. Mivel a logaritmusfüggvény szigorúan monoton növekvő, ezért a maximumhely meghatározása egyenértékű a loglikelihood-függvény maximumhelyének meghatározásával. Ennek érdekében a (F.1) egyenletet logaritmikusan transzformálva

$$\begin{aligned} \ln L = \ln P(x | \alpha, k) &= \sum_{i=1}^N \ln \left(\frac{\alpha k^\alpha}{x_i^{\alpha+1}} \right) \\ &= n \ln \alpha + n \alpha \ln k - (\alpha + 1) \sum_{i=1}^N \ln x_i \end{aligned} \quad (\text{F.2})$$

majd α szerint differenciálva

$$\frac{\partial \ln L}{\partial \alpha} = \frac{n}{\alpha} + n \ln k - \sum_{i=1}^N \ln x_i \quad (\text{F.3})$$

adódik. A deriváltat 0-val egyenlővé téve, majd az egyenletet α -ra rendezve, egyszerűsítve adódik α becslőfüggvénye:

$$\alpha = n / \sum_{i=1}^N \ln \left(\frac{x_i}{k} \right) \quad (\text{F.4.})$$

B.)

Vuong (1989) és Clauset et al. (2009) alapján a Vuong-féle likelihood-arány teszt a következőképpen írható fel: legyen a két összehasonlítandó modellünk likelihoodja

$$L_1 = \prod_{i=1}^n p_1(x_i) \quad L_2 = \prod_{i=1}^n p_2(x_i) \quad (\text{F.5}),$$

illetve, ezek arányának logaritmus

$$R = \ln \left(\frac{L_1}{L_2} \right) = \sum_{i=1}^n [\ln p_1(x_i) - \ln p_2(x_i)] \quad (\text{F.6}),$$

ahol $\ln p_j(x_i)$ az egyes x_i értékekre számolt log-likelihood a j -edik modell szerint.

Ahogy $n \rightarrow \infty$, $R \sim N(\mu, \sigma)$ $n\sigma^2$ varianciával, ahol σ^2 az egyes $[\ln p_1(x_i) - \ln p_2(x_i)]$ kifejezések varianciája, amely a mintából becsülhető:

$$\sigma^2 = \frac{1}{n} \sum_{i=1}^n [\ln p_1(x_i) - \ln p_2(x_i)]^2 - \left[\frac{1}{n} \sum_{i=1}^n (\ln p_1(x_i) - \ln p_2(x_i)) \right]^2 \quad (\text{F.7}).$$

Annak valószínűsége pedig, hogy a log-likelihood értékek különbsége egyenlő, vagy nagyobb az általunk megfigyelt R abszolút értékénél, megadható a következő egyenlettel:

$$P(t \geq |R|) = \frac{1}{\sqrt{2\pi n \sigma^2}} \left(\int_{-\infty}^{-|R|} e^{-\frac{t^2}{2n\sigma^2}} dt + \int_{|R|}^{\infty} e^{-\frac{t^2}{2n\sigma^2}} dt \right) = 1 - \text{erf}\left(\frac{|R|}{\sigma\sqrt{2n}}\right) \quad (\text{F.8})$$

ahol $\text{erf}\left(\frac{|R|}{\sigma\sqrt{2n}}\right) = \frac{2}{\sqrt{\pi}} \int_{|R|/\sigma\sqrt{2n}}^{\infty} e^{-t^2} dt$ Az $\text{erf}(z)$ függvény a Gauss-féle hibafüggvényre utal.

Irodalomjegyzék

- Abdel-Rahman, H. – Anas, A. (2004): Theories of sytems of cities. In: Henderson, J. V. – Thisse, J. F. (eds): *Handbook of Regional and Urban Economics Vol. 4*, North-Holland, pp.2293-2340.
- Abdel-Rahman, H. – Fujita, M. (1990): Product variety, Marshallian externalities, and city sizes. *Journal of Regional Science*, 2, pp.165-183.
- Alperovich, G. – Deutsch, J. (1995): The size distribution of urban areas: Testing for the appropriateness of the Pareto distribution using generalized Boks-Cox transformation function. *Journal of Regional Science*, 2, pp.267–276.
- Auerbach, F. (1913): Das Gesetz der Bevölkerungskonzentration. *Petermanns Geographische Mitteilungen*, 59, pp.74–76.
- Bee, M. – Riccaboni, M. – Schiavo, S. (2011): Pareto versus lognormal: A maximum entropy test. *Working Paper, Department of Economics*, University of Trento, 2, p.30.
- Benguigui, L – Blumenfeld-Lieberthal, E. (2007): A dynamic model for city size distribution beyond Zipf 's law. *Physica A*, 2, pp.613–627.
- Benguigui, L – Blumenfeld-Lieberthal, E. (2011): The end of a paradigm: is Zipf's law universal? *Journal of Geographical Systems*, 1, pp. 87-100.
- Black, D. – Henderson, J. V. (2003): Urban evolution in the USA. *Journal of Economic Geography*, 3, 343-372.
- Blank, A. – Solomon, S. (2000): Power laws in cities population, financial markets and internet sites (scaling in systems with a variable number of components). *Physica A*, 1-2, pp.279-288.
- Brakman, S. – Garretsen, H. – van Marrewijk, C. (2001): *An Introduction to Geographical Economics*. Cambridge University Press, Cambridge. p.350.+
- Cameron, T. A. (1990): One-Stage Structural Models to Explain City Size. *Journal of Urban Economics*, 3, pp.294-307.

- Cheshire, P. (1999): Trends in sizes and structures of urban areas. In: Cheshire, P – Mills, E. (eds.): *Handbook of Regional and Urban Economics*, Vol. 3, North-Holland, pp.1339-1373.
- Clauset, A. – Shalizi, C. R. – Newman M. E. J. (2009): Power-law distributions in empirical Data. *SIAM Review*, 4, pp.661-703.
- Czaller, L. (2012): A Zipf-törvény érvényesülése a világ országaiban. *Területi Statisztika*, 5, pp.461-478.
- Dimou, M.– Scaffar, A. (2009): Urban Hierarchies and City Growth in the Balkans. *Urban Studies*, 13, pp.2891-2906.
- Duranton, G. – Puga, D. (2005): From Sectoral to Functional Urban Specialization. *Journal of Urban Economics*, 2, pp.343-370.
- Duranton, G. (2005): Testing for Localization Using Micro-Geographic Data. *Review of Economic Studies*, 4, pp.1077-1106.
- Dusek, T. (2004): *A területi elemzések alapja*. Regionális Tudományi Tanulmányok 10, ELTE Regionális Földrajzi Tanszék, Budapest, p.245.
- Eeckhout, J. (2004): Gibrat's law for (all) cities. *American Economic Review*, 4, pp.1429-1451.
- Eeckhout, J. (2009): Gibrat's law for (all) cities: A Reply. *American Economic Review*, 6, pp.1676-1683.
- Fujita, M. – Krugman, P. – Venables, A. J. (1999): *The Spatial Economy: Cities, Regions, and International Trade*. MIT Press, Cambridge, p.367.
- Gabaix, X. (1999): Zipf's law for cities: An explanation, *Quarterly Journal of Economics*, 3, pp.739-767.
- Gabaix, X. and Ioannides, Y. (2004), The Evolution of City Size Distributions. In: Henderson, J. V. – Thisse, J. F. (eds): *Handbook of Regional and Urban Economics* Vol. 4, North-Holland, pp.2341-2378.
- Garmestani, A. S. – Allen, C. R. – Gallagher, C. M. (2008): Power laws, discontinuities and regional city size distributions. *Journal of Economic Behavior and Organization*, 1, pp.209-216.
- Giesen, K – Zimmermann, A. – Suedekum, J. (2010): The size distribution across all cities – Double Pareto Lognormal Strikes. *Journal of Urban Economics*, 2, pp.129-137.
- Glaeser, E. – Kallal, H. – Scheinkman, J. – Schleifer, A. (1992): Growth in cities. *Journal of Political Economy*, 6, pp.1126-1152.
- Goldstein, M.L. – Morris, S.A. – Yen, G.G. (2004): Problems with fitting to the Power-law distribution. *European Physical Journal B*, 2, pp.255–258.
- Guerin–Pace, France (1995): Rank-size distribution and the process of urban growth. *Urban Studies*, 3, pp.
- Henderson, J. V. (1974): The Size, and Types of Cities. *American Economic Review*, 4, pp.640-656.
- Holmes, T. J. – Lee, S. (2010): Cities as six-by-six squares: Zipf's law? In: Glaeser, E. (ed): *Agglomeration Economics*. University of Chicago Press, Chicago, pp.105-132.
- Ioannides, Y. – Skouras, S. (2012): US city size distribution: Robustly Pareto, but only in the upper tail. *Journal of Urban Economics*, forthcoming.
- Kosmopoulou, G. – Buttry, N. – Johnson, J. – Kallsnick, A. (2007): Suburbanization and the rank-size rule. *Applied Economics Letters*, 1, pp.1-4.
- Krugman, P. (1996): Confronting the mystery of urban hierarchy. *Journal of the Japanese and International Economies*, 4, pp.399-418.
- Laherrere, J. – Sornette, D. (1998): Stretched exponential distributions in nature and economy: 'Fat tails' with characteristic scales. *European Physics Journal B*, 2, pp.525-529.

- Levy, M. (2009): Gibrat's Law for (all) cities: a comment. *American Economic Review*, 4, pp.1672-1675.
- Malacarne, L. C. – Mendes, R. S. – Lenzi, E. K. (2001). q -exponential distribution in urban agglomeration. *Physical Review E*, 5, 017106.
- Malecki, E. J. (1980): Growth and change in the analysis of rank - size distributions: empirical findings. *Environment and Planning A*, 1 pp.41-52.
- Malevergne, Y., Pisarenko, V., and Sornette, D. (2011). Testing the Pareto against the Lognormal Distributions with the Uniformly Most Powerful Unbiased Test Applied to the Distribution of Cities. *Physical Review E*, 83 (3), 036111
- Mitzenmachel, M. (2003): A brief history of generative models for Power-law and lognormal distributions. *Internet Mathematics*, 2. pp.226-251.
- Newman, M. (2005): Power laws, Pareto distributions and Zipf's law. *Contemporary Physics*, 5, pp.323-351.
- Nishiyama, Yoshihiko – Osada, Susumu – Sato, Yasuhiro (2008) OLS estimation and the t-test revisited in rank–size regression. *Journal of Regional Science*, 4, pp.691-716.
- Nitsch, V. (2005): Zipf zipped. *Journal of Urban Economics*, 1. pp.86-100.
- Openshaw, S. (1984): *The modifiable areal unit problem*. GeoBooks, Norwich, p.41.
- Parr, J. (1976): “A class of deviations from rank-size regularity: Three interpretations. *Regional Studies*, 10 (3), pp.285-292.
- Parr, J. – Suzuki, K. (1973): Settlement popultion and the lognormal distribution. *Urban Studies*, 3, pp.336-352.
- Perline, R. (2005): Strong, weak, and false inverse power laws. *Statistical Science*, 1, pp.68-88.
- Reed, W. J. – Jorgensen, M. A. (2004): The double Pareto-lognormal distribution - A new parametric model for size distributions. *Communications in Statistics: Theory and Methods*, 8, pp.1733-1753.
- Rosen, K. T. – Resnick, M. (1980): The size distribution of cities: An examination of the Pareto law and primacy. *Journal of Urban Economics*, 2.pp.165-186.
- Rossi-Hansberg, E. – Wright, M. L. J. (2007). Urban structure and growth. *Review of Economic Studies*, 74, pp.597–624.
- Rozenfeld, H. D. – Rygski, D. – Gabaix, X. – Makse, H. A. (2011): The area and population of cities: New insights from a different perspective on cities. *American Econimc Review*, 5, pp.2205-2225.
- Singer, H. W. (1936): The ‘courbe des populations.’A parallel to Pareto's Law. *The Economic Journal*, 46, (182), pp.254-263.
- Soo, K, T (2005): Zipf's Law for cities: A cross country investigation. *Regional Science and Urban Economics*, 3, pp.239-263.
- US Census Bureau (2010): *Geographic Areas Reference Manual*. United States Department of Commerce. <http://www.census.gov/geo/www/garm.html>
- Vuong, Q. H. (1989): Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica*, 2, pp.307-333.
- Warren, Drake (2007): Examining City Size Distributions using Urban Areas. *45th Annual Meeting of the Western Regional Science Association*.
- van Zyl, M. (2011): Application of the Kolmogorov-Smirnov test to estimate the threshold when estimating the extreme value index. *Communications in Statistics – Simulation and Computation*, 40 (2), pp.199-207.