

Zipfov zakon in porazdelitev naselij v Sloveniji

Mitja Perko

Ekonomsko-poslovna fakulteta Maribor, Univerza v Mariboru

UDK 332.1(497.4)

JEL: R000, R120, R190

Povzetek

V tem članku preizkušamo veljavnost Zipfovega zakona za slovenska naselja. Zipfov zakon je pojav, ki govori o porazdelitvi naselij po velikosti v posamezni državi. Po Zipfovem zakonu naj bi porazdelitev naselij, ko jih razvrstimo po velikosti (populaciji), sledila Paretovi porazdelitvi, kjer je Paretoev eksponent enak 1. Naš članek se pridružuje skupini člankov s tega področja in se zgleduje po najnovejših metodah in merilih Terre (2009) za ocenitev Paretovega eksponenta in veljavnosti Zipfovega zakona: preizkus, ali je Paretoev eksponent na vzorcu slovenskih naselij značilno enak 1 in ali vzorec sledi Paretovi porazdelitvi. Ugotovimo, da je Paretoev eksponent pri vseh cenilkah in velikostih vzorca naselij, razen pri OLS cenilki in velikosti vzorca 200, statistično značilno enak 1. Preizkušanje, ali naš vzorec sploh sledi Paretovi porazdelitvi, z Anderson-Darlingovim in Cramer-von Misesovim preizkusom pa da dvopomenske rezultate. Na podlagi večje občutljivosti in rezultata Anderson-Darlingovega preizkusa Zipfov zakon za slovenska naselja zavrnemo.

Ključne besede: Zipfov zakon, Paretova porazdelitev, slovenska naselja

Abstract

In this paper, we test the validity of Zipf's Law for Slovenian settlements. Zipf's Law is a phenomenon that describes the size distribution of cities or settlements in a specific country. Zipf's Law states that the distribution of cities, once sorted according to size (population), should follow a Pareto distribution, in which the Pareto exponent is equal to 1. Our paper joins other papers in this research field and follows the example of Terra (2009) for determining the value of the Pareto exponent and the validity of Zipf's Law by testing whether the Pareto exponent is statistically equal to 1 and if our sample of Slovenian settlements follows a Pareto distribution. We find that the Pareto exponent is statistically equal to 1 with all of the estimators and with all sample sizes of settlements, except with the OLS estimator and sample size of 200; however, testing the validity of a Pareto distribution for Slovenian cities with the Anderson-Darling and Cramer-von Mises tests returns mixed results. Based on the results and increased sensitivity of the Anderson-Darling test, we reject Zipf's Law as applicable for Slovenian settlements.

Key words: Zipf's Law, Pareto distribution, Slovenian settlements

1. Uvod

Obstaja en izjemen empirični pojav, pri katerem se mora teorija ukloniti preverjenim dejstvom. To je Zipfov zakon.

– T. Knudsen (2001)

Zipfov zakon je empirični pojav, ki pojasnjuje, kakšna je porazdelitev mest v neki državi, ko mesta razvrstimo oziroma rangiramo po velikosti. Po Zipfovem zakonu naj bi porazdelitev mest sledila Paretovi porazdelitvi. Ali drugače: rang mesta je inverzno povezan z velikostjo mesta (Rose 2006). Zipfov zakon izhaja iz t. i. potenčnega zakona (power law), pri katerem je število mest s populacijo večjo od 5 proporcionalno S^{-a} , kjer je eksponent a enak 1. Če tako rangiramo slovenska mesta po velikosti, kjer Ljubljana zasede rang 1, Maribor rang 2 itd., ta rang in populacijo mesta logaritmiramo, bi morali dobiti padajočo premico z naklonom -1 . Če bi tak eksponent a tudi dobili, potem lahko trdimo, da je drugo največje mesto v državi za polovico manjše od največjega

mesta, tretje mesto po velikosti obsega le tretjino velikosti prvega mesta itd. Eksponent a imenujemo tudi Paretoev eksponent (oziroma koeficient).

V našem članku ocenjujemo, ali je omenjeni Paretoev eksponent za slovenska mesta oziroma naselja resnično enak 1, kakršnega predvideva Zipfov zakon. Prav tako ocenjujemo, ali porazdelitev slovenskih naselij sledi Paretovi porazdelitvi. To ocenjujemo s cenilko po metodi najmanjših kvadratov (OLS), popravljeno OLS cenilko, Hillovo cenilko (Hill 1975), cenilko največjega verjetja (ML) in z nepristransko cenilko z minimalno varianco (MVUE). Za izpeljavo te analize smo se odločili zaradi več razlogov: dosedanje raziskave niso vključevale analize na primeru Slovenije; razlog za to lahko morebiti iščemo v tem, da Slovenija ni izpolnjevala nekaterih meril (predvsem števila velikih mest) za vključitev v dosedanje analize. Poleg tega smo se želeli prepričati, ali tudi Slovenija spada v tisto skupino držav, za katere velja Zipfov zakon, ali ne.

Ta članek bo poskušal prispevati k razumevanju veljavnosti Zipfovega zakona za Slovenijo s pomočjo novejših metod in meril po vzoru Terre (2009). Z vidika novih metod to pomeni uporabo cenilke MVUE, ki je med drugimi omenjenimi cenilkami najmanj pristranska pri ocenjevanju Paretovega eksponenta. Z vidika meril pa smo veljavnost Zipfovega zakona za Slovenijo ugotavljali ne le na podlagi tega, kakšen je Pareto eksponent, temveč tudi na podlagi tega, ali porazdelitev naselij sploh sledi Paretovi porazdelitvi, kakor trdi Zipfov zakon.

Na podlagi omenjenih cenilk smo ocenili Pareto eksponent za različne velikosti vzorca, in sicer za vzorce 30, 50, 100 in 200 največjih slovenskih naselij. Ocene kažejo, da je Pareto eksponent statistično značilno enak 1 pri vseh cenilkah in velikostih vzorca, le pri vzorcu velikosti 200 je OLS cenilka edina, pri kateri ne moremo sprejeti hipoteze, da je eksponent značilno enak 1. Da bi ugotovili, ali porazdelitev slovenskih naselij sploh sledi Paretovi porazdelitvi, smo opravili Anderson-Darlingov in Cramer-von Misesov preizkus. Preizkusa se uporabljata za ugotavljanje, ali neki vzorec prihaja iz populacije z določeno porazdelitvijo. Ugotovimo, da s Cramer-von Misesovim preizkusom sprejememo hipotezo o Paretovi porazdelitvi, medtem ko Anderson-Darlingov preizkus Paretovo porazdelitev zavrne. Razlog za drugačno oceno Anderson-Darlingovega preizkusa je morebiti ta, da je ta preizkus dosti občutljivejši na deviacije repov porazdelitve v primerjavi s Cramer-von Misesovim preizkusom. Na podlagi tega se kompromisno odločimo, da Zipfov zakon za slovenska naselja zavrnemo.

Ta članek je razdeljen sledeče: v nadaljevanju predstavimo najpomembnejše raziskave empiričnega ocenjevanja Paretovega eksponenta. Nato opišemo, kako in s kakšnimi cenilkami je potekalo ocenjevanje Paretovega eksponenta v večini dosedanjih raziskav. Drugi del tega članka je namenjen naši analizi slovenskih naselij, rezultatom in sklepu o veljavnosti Zipfovega zakona.

2. Pregled literature

Razprava o Zipfovem zakonu je v literaturi obširna. Auerbach (1913) in Zipf (1949) sta bila med prvimi, ki sta prepoznala morebitno povezavo med velikostjo mesta in njegovim rangom (citirano v Terra 2009, 1). Prvo obsežnejšo analizo Zipfovega zakona sta naredila Rosen in Resnick (1980). Njuna študija je raziskovala Pareto eksponent za 44 držav sveta v letu 1970, rezultati pa so pokazali, da se ocenjen eksponent za večino držav giblje med 0,8 in 1,5, s povprečjem 1,13 in standardnim odklonom 0,19 (Gabaix in Ionides 2004). Prvi taki obširnejši analizi so sledile tudi druge (Alperovich 1984, 1988, Guerin-Pace 1995; citirano v Terra 2009, 1). Zanimanje za Zipfov zakon je znova obudil Gabaix (1999a, 1999b), pri čemer se je raziskovanje razdelilo na dve smeri: na raziskovanje o velikosti

Paretovega eksponenta ter na raziskovanje teoretične podlage za nastanek in obstoj tega empiričnega pojava.

Soo (2005) je naslednji, ki je opravil obširno analizo raziskovanja velikosti Paretovega eksponenta. Uporabil je OLS in Hillovo cenilko pri ocenjevanju veljavnosti Zipfovega zakona na vzorcu 75 držav. Z OLS cenilko Soo oceni povprečni Pareto eksponent v višini 1,11, s standardnim odklonom 0,2. Za države, ki sta jih ocenila Rosen in Resnick, Soo oceni povprečni Pareto eksponent v višini 1,179, s standardnim odklonom približno 0,2 (v primerjavi s povprečjem 1,13 pri Rosenu in Resnicku), 39 držav iz Soovega vzorca je imelo Pareto eksponent statistično večji kot 1, statistično manjši eksponent pa 14 držav. Pri oceni s Hillovo cenilko je povprečje vrednosti Paretovega eksponenta enako 1,167, s standardnim odklonom 0,25. S to cenilko 44 držav iz vzorca ni imelo Paretovega eksponenta statistično značilno različnega od 1.

Najobsežnejšo raziskavo velikosti Paretovega eksponenta na vzorcu 115 držav je opravil Terra (2009). Njegove ocene za povprečni Pareto eksponent so prikazane v tabeli 1. Naveden je tudi podatek, koliko držav od 115 je imelo eksponent statistično značilno enak 1 pri 5-odstotni stopnji značilnosti.

Tabela 1: Povprečni Pareto eksponent (α) z različnimi cenilkami, s standardnim odklonom in podatkom, za koliko od 115 držav je značilno enak 1 pri 5-odstotni stopnji značilnosti

Cenilka	Povprečje	Standardni odklon	$\alpha = 1$
OLS	1,041	0,226	20
popr. OLS	1,130	0,236	98
Hill	1,059	0,252	78
ML	1,080	0,255	76
MVUE	1,039	0,250	76

Vir: Terra (2009).

3. Ocenjevanje porazdelitve mest

Do zdaj smo besedno zvezo *potenčni zakon* omenili v povezavi s *porazdelitvijo* mest oziroma naselij. V tem poglavju opisujemo, kakšna je povezava med tema dvema pojmom. Teorija pravi, da lahko porazdelitev mest dobro opišemo s potenčnim zakonom. Poddomena potenčnega zakona pa je t. i. pravilo ranga-velikosti:

$$y_i = Ax_i^{-\alpha} \quad (1)$$

kjer je y_i rang i -tega mesta z x prebivalci, A je konstanta, α pa parameter, ki ju lahko ocenimo z ekonometričnimi metodami. Pravilo ranga-velikosti (oziroma Zipfov zakon) pa je pravzaprav Paretova porazdelitev, kjer je eksponent

$\alpha = 1$ (Capello in Nijkamp 2004, Guerin-Pace 1995, Kali 2003):

$$y_i = Ax_i^{-1} \quad (2)$$

Do enačbe 2 pridemo takole: ko integriramo Paretovo funkcijo gostote verjetnosti (enačba 3), dobimo kumulativno porazdelitveno funkcijo (enačba 4):

$$f(x) = \frac{\alpha k^\alpha}{x^{\alpha+1}}, x \geq k > 0 \quad (3)$$

$$F(x) = 1 - \left(\frac{x}{k}\right)^{-\alpha}, x \geq k > 0 \quad (4)$$

kjer je k minimalna vrednost v porazdelitvi, α pa je parameter, ki meri debelino repa Paretove porazdelitve. Iz tega oblikujemo t. i. funkcijo preživetja (oziroma komplementarno kumulativno porazdelitveno funkcijo):

$$S(x) = 1 - F(x) = \left(\frac{x}{k}\right)^{-\alpha} \quad (5)$$

Eeckhout (2004) pokaže, da lahko rang mesta v empirični porazdelitvi zapišemo tudi takole:

$$y_i = n S(x) \quad (6)$$

kjer je n število mest v vzorcu. Ko v enačbo 6 vstavimo enačbo 5, dobimo s preureditvijo:

$$y_i = \frac{nk^\alpha}{x_i^\alpha} \quad (7)$$

$$y_i = \frac{A}{x_i^\alpha} \quad (8)$$

kjer je $A = nk^\alpha$. Enačba 8 je pravzaprav enačba 1. Če drži Zipfov zakon, je Paretoev eksponent $\alpha = 1$.

3.1 OLS cenilka

Koliko pa sploh drži Zipfov zakon in Paretoev eksponent 1 za porazdelitev naselij? Da bi to preverili, je treba logaritmirati enačbo 8 (Mills in Hamilton 1989, Terra 2009, Knudsen 2001):

$$\ln y_i = \ln A - \alpha \ln x_i \quad (9)$$

To lahko zapišemo v obliki, ki omogoča ocenjevanje z ekonometričnimi metodami, in sicer:

$$\ln y_i = C - \alpha \ln x_i + \varepsilon_i, i = 1, \dots, n \quad (10)$$

kjer je $C (= \ln A)$ konstanta, α pa parameter, ki ju lahko ocenimo, ε_i pa je rezidual posamezne observacije. Če drži Zipfov zakon, je $\alpha = 1$ ter lahko med logaritmom ranga naselja in velikostjo naselja začrtamo (padajočo) premico z naklonom -1 . Zgornja enačba se ocenjuje z metodo najmanjših kvadratov oz. OLS – ordinary least squares (Gabaix in Ionnides 2004). Visok α koeficient bi za posamezno državo pomenil, da je velikost naselij enakomerneje porazdeljena, manjši koeficient pa bi nakazoval, da je populacija skoncentrirana v manjšem številu naselij (Soo 2005, Mu in Wang 2006, Terra 2009).

Metoda OLS je bila v preteklosti temeljna metoda ocenjevanja veljavnosti Zipfovega zakona. Toda Gabaix in Ionnides (2004) sta na podlagi Monte Carlo simulacij ugotovila, da ima metoda OLS številne pomanjkljivosti. OLS cenilka naj bi bila pristranska navzdol pri ocenjevanju Paretovega eksponenta v majhnih vzorcih. Težava nastane tudi pri oceni standardnih odklonov z metodo OLS, saj naj bi bili ti podcenjeni, kar pripelje do preveč zavrnitev Zipfovega zakona (Gabaix in Ionnides 2004, Soo 2005, Terra 2009). V literaturi so se zaradi tega v zadnjih letih začele razvijati in uporabljati drugačne metode za ocenjevanje Zipfovega zakona. Navajamo jih v nadaljevanju.

3.2 Popravljen OLS cenilka

Gabaix in Ibragimov (2007) trdita, da je OLS cenilka pristranska. Kot alternativo ponudita preprosto transformacijo omenjene cenilke, tako da se uporabi odvisna spremenljivka $y_i - 1/2$ namesto y_i . Ta preprosta transformacija naj bi minimizirala pristranskost OLS cenilke. Enačba, ki jo ocenimo, je:

$$\ln\left(y_i - \frac{1}{2}\right) = C - \alpha \ln x_i + \varepsilon_i, i = 1, \dots, n \quad (11)$$

s standardnim odklonom $\hat{\alpha} \sqrt{\frac{2}{n}}$.

Za preverjanje statistične značilnosti ne moremo uporabiti ocene OLS za standardni odklon, saj podcenjuje resnični standardni odklon, pri čemer bi njegova uporaba prevečkrat privedla do ocene, da je koeficient α statistično značilno različen od 1. Toda Terra (2009) z Monte Carlo simulacijami pokaže, da je ob uporabi popravljen OLS cenilke in standardnega odklona enako kot ob uporabi navadne OLS cenilke, verjetnost zagrašitve napake tipa I in tipa II pri preverjanju hipotez še vedno zelo velika.

3.3 Hillova cenilka

Naslednja priljubljena cenilka, ki se je v literaturi začela uporabljati v zadnjih letih, je Hillova cenilka (Hill 1975), ki sta jo kot dodatno alternativo OLS cenilki za ocenjevanje Paretovega eksponenta predlagala Gabaix in Ionnides

(2004). Hillova cenilka je pod ničelno hipotezo, da je resnična porazdelitev porazdelitev potenčnega zakona, cenilka največjega verjetja oziroma cenilka ML – maximum likelihood. Toda kot poudarjata avtorja, je Hillova cenilka manj uporabna v majhnih vzorcih, saj postane pristranska. Cenilka je taka:

$$\hat{\alpha}^{HILL} = \frac{n-1}{\left(\sum_{i=1}^{n-1} \ln x_i - \ln x_n\right)} \quad (12)$$

s standardnim odklonom

$$\sigma(\hat{\alpha}^{HILL}) = \hat{\alpha}^{HILL} \left(\frac{\sum_{i=1}^{n-1} \left(\tau_i - \left(\frac{1}{\hat{\alpha}^{HILL}} \right) \right)^2}{n-2} \right)^{\frac{1}{2}} (n-1)^{-\frac{1}{2}} \quad (13)$$

kjer je $\tau_i = i(\ln x_i - \ln x_{i+1})$.

3.4 Cenilka največjega verjetja in nepristranska cenilka z minimalno varianco

Za oceno Paretovega eksponenta se uporabljata tudi dve novejši cenilki – cenilka največjega verjetja (ML – maximum likelihood estimator) in nepristranska cenilka z minimalno varianco (MVUE – minimum variance unbiased estimator). Če predpostavljamo, da je porazdelitev naselij Paretova porazdelitev, je ML¹ cenilka taka:

$$\hat{\alpha}^{ML} = n \left(\sum_{i=1}^n \ln \left(\frac{x_i}{\hat{k}^{ML}} \right) \right)^{-1} \quad (14)$$

$$\hat{k}^{ML} = \min_i x_i \quad (15)$$

s standardnim odklonom

$$\sigma(\hat{\alpha}^{ML}) = \frac{\hat{\alpha}^{ML}}{\sqrt{n}} \quad (16)$$

95-odstotni interval zaupanja za cenilko α pa je

$$\left[\frac{\hat{\alpha}^{ML}}{2N} \chi^2_{2(n-1), 0,025}, \frac{\hat{\alpha}^{ML}}{2N} \chi^2_{2(n-1), 0,975} \right]$$

(Terra 2009).

Ker pa so ML cenilke velikokrat pristranske v majhnih vzorcih, sta Likeš (1969) in Baxter (1980; citirano v Terra 2009, 7) izpeljala še nepristranske cenilke z minimalno varianco (MVUE) za zgornja parametra, in sicer:

$$\hat{\alpha}^{MVUE} = \left(1 - \frac{2}{n} \right) \hat{\alpha}^{ML} \quad (17)$$

$$\hat{k}^{MVUE} = \left(1 - \frac{1}{(n-1)\hat{\alpha}^{ML}} \right) \hat{k}^{ML} \quad (18)$$

95-odstotni interval zaupanja za cenilko α je

$$\left[\frac{\hat{\alpha}^{MVUE}}{2N} \chi^2_{2(n-1), 0,025}, \frac{\hat{\alpha}^{MVUE}}{2N} \chi^2_{2(n-1), 0,975} \right]$$

(Terra 2009).

Če model ni napačno specificiran, sta zgornji cenilki asimptotično konsistentni in nepristranski. Cenilke pa so vendarle lahko pristranske v majhnih vzorcih, zato Terra (2009) z Monte Carlo simulacijami primerja lastnosti teh cenilk, kadar jih ocenjujemo na majhnem vzorcu. Ugotovi, da je pri majhnem vzorcu ($n = 20$) pristranskost nekaterih cenilk zelo velika: pri OLS cenilki (enačba 10) in ML cenilki (enačba 14) se pristranskost približuje 10 %. Hillova cenilka (enačba 12) in popravljena OLS cenilka (enačba 11) sta nekoliko manj pristranski (med 4 in 6 %). Najmanjša je pristranskost MVUE cenilke (manj kot odstotek), prav tako ima ta cenilka najmanjšo povprečno kvadratno napako (MSE – mean square error) med vsemi omenjenimi cenilkami. Za velike vzorce ($n = 200$) so skoraj vse cenilke zanemarljivo pristranske, razen OLS cenilke, ki je vztrajno pristranska navzdol. Monte Carlo simulacije kažejo, da je za ocenjevanje Paretovega eksponenta najprimernejša MVUE cenilka.

4. Empirična analiza porazdelitve naselij v Sloveniji

4.1 Uvod

Teoretičnemu opisu Zipfovega zakona in njegovi problematiki ocenjevanja sledi naša empirična analiza na primeru porazdelitve naselij v Sloveniji. Po naših informacijah Slovenija do zdaj ni bila vključena v nobeno analizo Zipfovega zakona na vzorcu večjega števila držav. Kot smo poskušali nakazati, je Zipfov zakon zanimiv empirični pojav, zato nas zanima tudi njegova morebitna veljavnost v Sloveniji.

¹ Za matematično izpeljavo cenilke glej Quandt (1964).

4.2 Podatki

Podatke, ki smo jih uporabili v analizi, smo pridobili s spletne strani SI-Stat podatkovnega portala Statističnega urada Republike Slovenije (SURS 2002). Podatki so bili pridobljeni pri popisu prebivalstva leta 2002, gre pa za število prebivalcev v naseljih v Sloveniji. Pri našem ocenjevanju porazdelitve naselij je bilo treba izbrati vzorec iz teh podatkov. Odločili smo se izbrati prvih 200 (urbanih) naselij po velikosti.² Omejevanje na določeno velikost vzorca je bilo izvedeno tudi pri drugih analizah ocenjevanja Zipfovega zakona. Povprečna velikost naselja v našem vzorcu je 5408 prebivalcev, največja vrednost v vzorcu pa ustreza Ljubljani z 258.873 prebivalci.

4.3 Metodologija

V naši empirični raziskavi se poskušamo zgledovati po vzoru Terre (2009). Njegov način preverjanja Zipfovega zakona je drugačen, kot so to počeli različni avtorji pred njim. Zipfov zakon Terra preveri v dveh korakih. Na začetku preveri, ali porazdelitev naselij v državi sploh sledi Paretovi porazdelitvi, pri drugem koraku pa poskuša oceniti Paretov eksponent in preizkusiti, ali je statistično značilno enak 1, kot narekuje Zipfov zakon. Taki metodologiji poskušamo slediti tudi mi.

Pri prvem koraku bomo preizkusili, ali porazdelitev naselij v Sloveniji sledi Paretovi porazdelitvi, in sicer z Anderson-Darlingovim in Cramer-von Misesovim preizkusom, ki preizkusita hipotezo, ali je porazdelitev vzorca podobna izbrani teoretični porazdelitvi, v našem primeru Paretovi porazdelitvi. Pri drugem koraku bomo ocenjevali cenilke, ki smo jih opisali v teoretičnem delu. Lahko bi ocenili samo cenilko MVUE, ki je po analizi Terre (2009) najmanj pristranska, vendar želimo, da bi bila naša analiza na primeru Slovenije dosledna in bi omogočila robustne rezultate ocene veljavnosti Zipfovega zakona.

V literaturi se pojavljajo tudi trditve in dokazi o tem, da naj bi bila vrednost Paretovega eksponenta v izračunu enačbe 10 (OLS) zelo odvisna od velikosti vzorca; večji kot je vzorec, nižja naj bi bila vrednost Paretovega eksponenta in obratno (Guerin-Pace 1995, Eeckhout 2004).³ Predlagano je bilo tudi, da naj bi v primerjavi s celotnim urbanim sistemom (Soo 2007) Zipfov zakon veljal samo v zgornjem delu porazdelitve naselij, tj. za večja naselja. Tako analizo občutljivosti bomo preizkusili tudi mi, in sicer na različnih velikostih vzorca – poleg ocene za izhodiščnih 200 naselij po velikosti še za prvih 30, 50 in 100. OLS in popravljeno OLS cenilko bomo ocenjevali v programu Eviews 6.

² Po slovenski zakonodaji naj bi sicer urbano naselje imelo vsaj 3000 prebivalcev. Za nas je to politična delitev, mi pa smo sledili merilu dovolj velikega števila pojavov. O problematiki definicije mesta/naselja in merjenja Paretovega eksponenta pišejo tudi Kosmopoulou et al. (2007).

³ Eeckhout (2004) namiguje tudi na to, da bi bila lahko naselja glede na velikost porazdeljena log-normalno. V literaturi večkrat zasledimo tako mnenje, vendar se do zdaj po našem opazovanju raziskave kljub nekaterim dokazom niso obrnile v smer log-normalne porazdelitve.

4.4 Rezultati

4.4.1 Ali porazdelitev naselij v Sloveniji sledi Pareto porazdelitvi?

Za pridobitev morebitnega odgovora na to vprašanje bomo uporabili Anderson-Darlingov in Cramer-von Misesov preizkus. Ničelna hipoteza teh dveh preizkusov je, da empirična porazdelitev vzorca velikosti n sledi določeni teoretični porazdelitvi. Alternativna hipoteza je, da empirična porazdelitev vzorca ne sledi izbrani teoretični porazdelitvi. Tabela 2 prikazuje rezultate Anderson-Darlingovega in Cramer-von Misesovega preizkusa, ki smo ju opravili v programu Eviews 6.

Tabela 2: Prikaz Anderson-Darlingovega in Cramer-von Misesovega preizkusa in preizkušanje hipoteze, ali empirična porazdelitev našega vzorca 200 naselij sledi teoretični Paretovi porazdelitvi

Metoda	Prilagojena vrednost	Verjetnost (p-vrednost)
Cramer-von Mises (W2)	0,06567	0,5458
Anderson-Darling (A2)	14,0901	0,0000

Vir: lastni izračun.

V tabeli 2 vidimo, da je p-vrednost za statistiko A2 (Anderson-Darlingov preizkus) manjša od 0,05, kar pomeni, da zavrnemo ničelno hipotezo pri 5-odstotni stopnji značilnosti, da je empirična porazdelitev našega vzorca enaka teoretični Paretovi porazdelitvi. Zanimivo pa je, da s statistiko W2 (Cramer-von Misesov preizkus) ničelne hipoteze ne moremo zavrniti pri isti stopnji značilnosti. Take ocene obeh preizkusov so nas nekoliko zbegale. Morebitni razlog za različni oceni preizkusov je lahko ta, da je Anderson-Darlingov preizkus veliko občutljivejši na deviacije repov empirične porazdelitve kot Cramer-von Misesov preizkus in zato ne sprejme ničelne hipoteze. Kaj ocena teh dveh statistik pomeni v okviru veljavnosti Zipfovega zakona, bomo razložili v sklepu tega članka.

4.4.2 Je Paretov eksponent za slovenska mesta enak 1?

V tem podpoglavju predstavljamo rezultate ocen različnih cenilk za Paretov eksponent za različno velike vzorce slovenskih naselij; te ocene so predstavljene v tabeli 3.

Tabela 3: Ocene Paretovega eksponenta z različnimi cenilkami in različnimi velikostmi vzorcev. Naveden je tudi podatek, ali je Paretov eksponent statistično značilno enak 1 pri 5-odstotni stopnji značilnosti

Vzorec (štev. naselij)	OLS	popr. OLS	Cenilka Hill	ML	MVUE
n = 30	0,963	1,100	1,205	1,247	1,163
N = 50	1,050	1,159	1,160	1,184	1,136
n = 100	1,068	1,136	1,092	1,103	1,081
n = 200	1,076*	1,118	1,059	1,065	1,054

Vir: lastni izračun.

* Statistično značilno različno od 1 pri 5-odstotni stopnji značilnosti.

Rezultati ocene OLS cenilke kažejo, da se ob večanju vzorca Paretov eksponent povečuje; ta ugotovitev ni v skladu z Eeckhoutovo ugotovitvijo (2004), da ocena eksponenta pada, ko se velikost vzorca povečuje. Ko preizkušamo ničelno hipotezo pri 5-odstotni stopnji značilnosti, če je Paretov eksponent enak 1, ugotovimo, da ničelne hipoteze ne sprejmemo le pri vzorcu velikosti 200. Pri ocenjevanju z OLS cenilko smo uporabili tudi Newey-Westovo metodo robustnih standardnih odklonov.

Ocene Paretovega eksponenta popravljene OLS cenilke so nekoliko višje kot navadne OLS cenilke, kar je skladno z rezultati Terrovih ocen (2009). Pri vseh navedenih ocenah Paretovega eksponenta ne moremo zavrniti ničelne hipoteze pri 5-odstotni stopnji značilnosti, da je Paretov eksponent enak 1.

Tudi pri Hillovi cenilki je podobno. Pri nobeni velikosti vzorca ne moremo zavrniti ničelne hipoteze, da je Paretov eksponent enak 1. Sicer smo v teoretičnem delu tega članka omenili, da je lahko Hillova cenilka pristranska v majhnih vzorcih, vendar je vzorec velikosti 200 tako velik, da je pristranskost cenilke, če sklepamo na podlagi Monte Carlo simulacij Terre (2009), zanemarljiva.

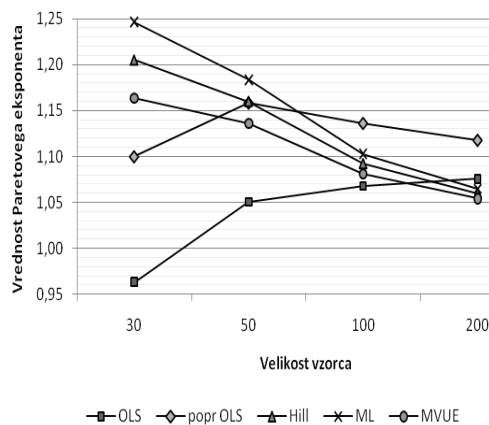
Kot pri Hillovi cenilki tudi pri ML cenilki velja, da je pristranska v majhnih vzorcih, vendar z naraščanjem velikosti vzorca tudi v tem primeru pristranskost postane zanemarljiva. Če preizkušamo hipotezo, da je Paretov eksponent enak 1, s 95-odstotnim intervalom zaupanja, ki ga poda Terra (2009), ugotovimo, da hipotetizirana vrednost pade v interval zaupanja in ničelne hipoteze ne moremo zavrniti.

MVUE cenilka naj bi bila najmanj pristranska med vsemi petimi cenilkami, ki smo jih ocenili. Z uporabo v teoretičnem delu omenjenega 95-odstotnega intervala zaupanja ugotovimo, da ne moremo zavrniti ničelne hipoteze, da je Paretov eksponent enak 1, za vse našete velikosti vzorcev.

Slika 1 nazorneje prikazuje gibanje ocen Paretovega eksponenta za različne velikosti vzorcev. Razberemo, da se ob večanju vzorca ocena Paretovega eksponenta za vse cenilke približuje posamezni določeni vrednosti, kar je razmeroma skladno s Soojevimi ugotovitvami (2007). Vrednost eksponenta za OLS cenilko z večanjem vzorca narašča, medtem ko za popravljeno OLS cenilko

sprva narašča, nato pa pada. Paretov eksponent pri Hillovi cenilki, ML in MVUE cenilkah pada z naraščanjem velikosti vzorca in se približuje okvirni vrednosti 1,06.

Slika 1: Gibanje ocenjene vrednosti Paretovega eksponenta za različne cenilke in različne velikosti vzorca



Vir: lastni izračuni.

4.5 Sklep in razprava

Tu poskušamo rezultate iz empirične analize na primeru Slovenije razložiti z vidika veljavnosti Zipfovega zakona. Novejši pristop, po katerem smo se zgledovali tudi mi, je potekal v dveh korakih; oceniti, ali porazdelitev mest sploh sledi Paretovi porazdelitvi, in kakšen je Paretov eksponent za to porazdelitev. Če bi upoštevali samo drugi korak, kot je to počela večina raziskav na tem področju še do pred nekaj let, bi to lahko privedlo do nepravilnih sklepov o veljavnosti Zipfovega zakona za slovenska mesta.

Pri prvem koraku smo ugotovili, da ne moremo zagotovo trditi, ali porazdelitev slovenskih naselij sledi teoretični Paretovi porazdelitvi. Na podlagi Anderson-Darlingovega preizkusa smo zavrnili ničelno hipotezo o tem, da empirična porazdelitev sledi teoretični, na podlagi Cramer-von Misesovega preizkusa pa ničelne hipoteze nismo mogli zavrniti. Taki dvopomenski rezultati so nas nekoliko zbegali. Rezultati testiranja porazdelitve vzbujajo morebitni dvom tudi o rezultatih drugega koraka, saj literatura s področja ekonometrije pravi, da je treba za ocenjevanje po metodi največjega verjetja

natančno specificirati resnično porazdelitev, ki generira podatke, sicer so lahko ocene ML cenilk nekonsistentne (Wooldridge 2010).

Če se osredotočamo na večino preteklih raziskav, pri katerih merilo veljavnosti Paretove porazdelitve ni bilo upoštevano, bi morali na podlagi rezultatov ocen cenilk trditi, da Zipfov zakon za slovenska naselja velja. Če pa upoštevamo Terrovo merilo (2009) in kot dejavnik veljavnosti Zipfovega zakona upoštevamo tudi potrebno veljavnost Paretove porazdelitve za izbran vzorec naselij, bi morali Zipfov zakon zavrniti kljub Zipfovemu zakonu naklonjenim ocenam Paretovega eksponenta.

Rezultati Anderson-Darlingovih in Cramer-von Misesovih statistik niso bili enopomenski, zato je bilo treba sprejeti kompromisno odločitev. Na podlagi večje občutljivosti Anderson-Darlingovega preizkusa in veljavnosti Paretove porazdelitve smo se odločili, da Zipfovega zakona za slovenska naselja ne sprejmemo.

Viri in literatura

Capello, R., in Nijkamp, P. 2004. *Urban dynamics and growth: advances in urban economics*. Amsterdam: Elsevier.

Eeckhout, Jan. 2004. Gibrat's law for all cities. *American Economic Review*. 94 (5), stran 1429–1451.

Gabaix, X., Ionnides, Y. M. 2004. The evolution of city size distributions. *Handbook of Regional and Urban Economics*. 4, stran 2341–2378.

Gabaix, Xavier, in Ibragimov, Rustam. (2007). *Rank-1/2: a simple way to improve the OLS estimation of tail exponents (working paper)*. Dostopno na: <http://www.nber.org/papers/t0342> [3. 1. 2011].

Gabaix, Xavier. 1999a. Zipf's law and the growth of cities. *American Economic Review*. 89 (2), stran 129–132.

Gabaix, Xavier. 1999b. Zipf's law for cities: an explanation. *Quarterly Journal of Economics*. 114 (3), stran 739–767.

Guerin-Pace, France. 1995. Rank-size distribution and the process of urban growth. *Urban Studies*. 32 (3), stran 551–562.

Hill, Bruce M. 1975. A simple general approach to inference about the tail of a distribution. *The Annals of Statistics*. 3 (5), stran 1163–1175.

Kali, Raja. 2003. The city as a giant component: a random graph approach to Zipf's law. *Applied Economics Letters*. 10 (11), stran 717.

Knudsen, Thorbjorn. 2001. Zipf's law for cities and beyond: the case of Denmark. *American Journal of Economics & Sociology*. 60 (1), stran 101.

Kosmopoulou, Georgia; Buttry, Nicholas; Johnson, Jeremiah; Kallsnick, Adam. 2007. Suburbanization and the rank-size rule. *Applied Economics Letters*. 14 (1–3), stran 1–4.

Mills, E. S., in Hamilton, B. W. 1989. *Urban economics*. Glenview (Illinois): Scott, Foresman and Co.

Mu, L., Wang, X. 2006. Population landscape: a geometric approach to studying spatial patterns of the US urban hierarchy. *International Journal of Geographical Information Science*. 20 (6).

Quandt, Richard E. 1966. Old and new methods of estimation and the Pareto distribution. *Metrika*. 10 (1), stran 55–82.

Rose, Andrew K. 2006. Cities and countries. *Journal of Money, Credit & Banking*. 38 (8), stran 2225–2245.

Soo, Kwok Tong. 2005. Zipf's law for cities: a cross-country investigation. *Regional Science and Urban Economics*. 35 (3), stran 239–263.

Soo, Kwok Tong. 2007. Zipf's law and urban growth in Malaysia. *Urban Studies*. 44 (1), stran 1–14.

Statistični urad Republike Slovenije, popis prebivalstva, gospodinjstev in stanovanj 2002 [online]. Dostopno na: [<http://www.stat.si/pxweb/Database/Popis2002/Naselja/Prebivalstvo/Demografske%20zna%C4%8Dilnosti/Demografske%20zna%C4%8Dilnosti.asp>]. [1. 1. 2011].

Terra, Sebastien. (2009). *Zipf's law for cities: on a new testing procedure (working paper)*. Dostopno na: [<http://econpapers.repec.org/paper/cdiwpaper/1102.htm>] [22. 12. 2010].

Wooldridge, Jeffrey M. 2010. *Econometric Analysis of Cross Section and Panel Data*. Cambridge (Mass, ZDA): MIT Press.