

中国城市人口规模分布规律研究*

邓智团 樊豪斌

【摘要】文章利用中国1995、2000、2005、2010和2015年城市人口规模数据,通过多参数模型的比较与检验,对中国城市人口规模分布与变化规律进行实证分析。研究结果显示,中国城市人口规模分布并不遵循幂律分布,不适合采用帕累托分布或齐夫定律分析;与简单的对数正态分布相比,中国城市人口规模分布更符合双帕累托对数正态分布,且拥有对数正态分布的中间主体,上下尾部更接近幂律分布,从2000年开始上尾比下尾更厚重;从模型拟合变化来看,过去20年里的5个年份不同人口规模城市的增长速度不同,发散式增长显著,与双帕累托对数正态分布模型拟合的最佳分布或适度分布相比,2000年后处于上尾的大中城市的数量和人口规模增长迅速,处于下尾的中小城市数量多但人口规模增长动力相对不足。这些研究结论有助于检视当前中国的城市人口规模政策,并对新政策的制定提供决策支持。

【关键词】城市人口规模 位序—规模分布 双重帕累托对数正态分布 发散式增长

【作者】邓智团 上海社会科学院城市与人口发展研究所,副研究员;樊豪斌 美国克莱姆森大学商学院,博士研究生。

一、引言

改革开放以来,中国经历了世界历史上规模最大的城镇化进程,城镇化水平从1978年的17.9%迅速提升为2015年的56.1%,城市的数量和规模发生显著变化。但在这个快速变化过程中,中国城市人口规模政策却相对稳定,从《全国城市规划工作会议纪要》(1980)、《城市规划法》(1989)到《国家新型城镇化规划》(2014),“控制大城市发展”是一贯的政策基调。2015年时隔37年后召开的“中央城市工作会议”,明确提出要认识、尊重、顺应城市发展规律,端正城市发展指导思想。而城市规模^①政策一直是中国城市政策制定与研究的讨论焦

* 本文为国家社科基金重大项目“城乡协调发展与我国包容性城镇化新战略研究”(编号:15ZDC015)的阶段性成果。

① “城市规模”通常可指城市人口规模、城市经济规模、城市土地规模和城市建设用地规模等。除特别说明外,本文中“城市规模”与“城市人口规模”内涵一致,在行文中交替使用。

点之一。因此,深入探究城市人口规模增长与分布的规律,有助于检视中国的城市规模政策,对完善中国城市发展制度顶层设计和推动城市有序发展具有重要的现实意义。

在理论研究中,城市人口规模分布被视为有机的整体(尹文耀,1988),关于不同规模的城市如何演进、生长与分布一直是城市发展研究探寻的重要规律之一。自从1931年吉布拉定律和1935年齐夫定律出现后,关于城市规模分布规律的研究争论一直没有停止过,近些年还出现一些重要研究成果,如2013年《科学》发表关于城市规模起源的研究(Bettencourt, 2013),2004和2009年《美国经济评论》发表多篇关于城市规模分布规律的研究与评论文章(Eeckhout, 2004; Eeckhout, 2009; Levy, 2009)。早期的文献认为城市规模分布遵从两大定律:帕累托分布和吉布拉定律,认为帕累托分布是城市规模分布普遍遵循的幂律分布,齐夫定律是帕累托分布中指数为1的一种特殊形式;吉布拉定律最早是说明一个公司的均衡增长率独立于绝对规模,后来应用到城市规模分布研究^①。且所有这些研究,几乎都是采用截尾的城市样本数据,研究结论值得商榷。

新近研究中,有学者基于非截尾数据,确实发现单帕累托分布并不成立(Eeckhout, 2004),这对城市规模增长与分布研究提出了新的问题:什么是城市规模分布的合适参数估计模型?从城市规模分布中潜在的城市增长过程中能够得到什么新启发?Eeckhout(2004)基于城市随机增长模式进行建模,发现对数正态模型(LN)可以较好地拟合美国城市分布的实际情况,且是单纯地由吉布拉定律渐进形成,即城市随机增长,但该模型的上尾并不具备幂律分布的特点。所以严格地讲,对数正态分布(LN)并不与帕累托分布或齐夫定律一致。随后,Giesen等(2010)采用非截尾的8个国家(德国、美国、法国、巴西、捷克、匈牙利、意大利、瑞士)的城市规模数据,证实采用双重帕累托对数正态分布(DPLN)模型拟合实际城市分布比对数正态分布(LN)更好,理由是双重帕累托对数正态分布(DPLN)拥有对数正态分布的中间主体,并能在上尾和下尾均捕捉到了幂律分布特性,这是基于随机城市增长过程的均衡状态。由于对一个国家城市人口规模分布规律的探究,可以找到该国城市人口规模适度分布与最佳分布,再通过与实际城市人口规模分布的比照,能提出城市人口规模分布的优化方向,进而制定出差异化的城市规模政策(尹文耀,1988)。因此,研究已表明,不能简单地使用帕累托分布对一个国家城市人口规模分布进行研究,亟需借助新的方法展开深入研究,以免与潜在规律相去甚远而导致基于此制定的城市人口规模政策出现误导。

中国是世界上人口规模最大的国家,也是世界上内部差异最大的国家之一,在经济快

^① 帕累托(Pareto)分布和齐夫(Zipf)分布是常用的幂律分布,帕累托(Pareto)分布最初主要是用来描述人类社会中财富的分布规律,也被称作80~20规律;齐夫定律(Zipf's Law)在描述城市人口规模分布时,认为最大城市的规模是第二大城市的2倍,是第三大城市的3倍,以此类推等。吉布拉定律(Gibrat's Law)不是幂律分布,是指在一个特定的时期所有不同规模的企业预期增长幅度的概率是一样的,即增长的概率与规模无关,属于均衡增长,在城市人口规模分布的模拟研究中应用有限(Reed, 2001; 方正、王杰, 2011)。

速发展的同时,城市人口规模和城市数量也在发生急剧的变化。上述研究方法对中国城市人口规模增长与分布是否同样适用?目前关于中国城市人口规模分布的实证研究,主要是用幂律分布(如帕累托分布和齐夫定律)来分析中国城市(人口或建设用地)规模,在这些分析方法中的具体参数表现(Soo, 2005;高鸿鹰、武康平, 2007;Ye等, 2012;沈体雁、劳昕, 2012;Huang等, 2015),也有基于吉布拉对数分布定律来分析中国城市人口规模分布与理想分布状态的偏差及其影响因素(魏守华等, 2015),得出中国城市人口规模分布服从幂律分布和吉布拉分布的结论,据此模拟出中国城市人口规模的适度分布或最佳分布,并作为中国城市人口规模分布优化调整方向的依据。这些研究明显的不足之处在于:一是模型相对单一,仅限于帕累托模型,缺少对其他模型的拟合、检验与比较;二是采用截尾数据,多以地级以上城市为对象,城市规模相对较大,缺少中小规模的县级城市,以及人口规模较大的能进行“镇改市”的大镇,这实质是人为地将样本数据的下尾剔除,无法观察到中小城市在我国整体城市人口规模分布格局中的状态与地位,导致模型拟合与研究结论出现偏差。为此,本文将基于 Giesen 等(2010)与 Eeckhout(2010)的研究,以中国城市人口规模分布为对象,采用 5 个时间节点(1995、2000、2005、2010 和 2015 年)近似非截尾的城市人口规模数据,通过对不同分布模型的检验与比较,研究中国城市人口规模分布和变化规律,检验双重帕累托对数正态分布(DPLN)模型是否更好地拟合中国实际,并通过对双重帕累托对数正态分布参数的考察分析,对当前城市人口规模政策的有效性和科学性进行检视,为城市人口规模政策的调整与优化提供参考。

二、模型设计与数据

(一) 城市增长理论模型

Eeckhout(2004)的模型是基于假设有固定数量的城市,劳动者可以自由迁徙。空间均衡来自于正负规模外部性的权衡,而这种每个城市内部所累积的规模外部性是无法溢出到其他城市的。在每一个时期,每一个城市据点都承受了特定且随机的生产力冲击。最终所有的城市都是根据纯粹的吉布拉定律增长,可以表示为:

$$\frac{dPop_i}{Pop_i} = \mu dt + \sigma dB_i \quad (1)$$

其中, $\frac{dPop_i}{Pop_i}$ 是 i 城市在 t 时期的人口变动百分比。参数 μ 表示增长趋势, B_i 是均值为 0 方差为 σ^2 的独立冲击。早在 1931 年,吉布拉就预期存在一个随机的成比例的增长过程伴随着附加的随机冲击会渐进地指向一个对数正态分布(LN)。Eeckhout 的对数正态分布模型(LN)可以很好地拟合美国实际城市规模数据。对数正态规模分布是取 $\ln S$ 为正态分布的随机变量,该分布的理论密度函数为 $\mathcal{O}(\hat{\mu}, \hat{\sigma})$, 即:

$$\varnothing(\hat{\mu}, \hat{\sigma}) = \frac{1}{\hat{\sigma}\sqrt{2\pi}} e^{-\frac{(\ln S - \hat{\mu})^2}{2\hat{\sigma}^2}} \quad (2)$$

其中, S 为城市规模随机变量。

Reed(2002)构建了一个更具统计意义的模型,即虽然城市遵循吉布拉定律随机地增长,但在每一个时间间隔 dt 存在新城市出现的概率 λ_{dt} ,该新城市是已有城市的卫星城市。新城市的初始规模从一个均值为 μ_0 、方差为 σ_0^2 的对数正态分布中随机获得。这些新城市也会按比例增长。在时刻 t 总共有 $e^{\lambda t}$ 个城市,其中一些比另一些要老。Reed(2002)证明这一增长过程,类似于在生物学中所描绘的尤尔(Yule)过程^①,会逐渐趋近一个双重帕累托对数正态分布(DPLN),密度函数为:

$$f(x) = \frac{\alpha\beta}{\alpha+\beta} \left[x^{-\alpha-1} e^{(\alpha\mu_0 + \frac{\alpha^2\sigma_0^2}{2})} \Phi\left(\frac{\log(x) - \mu_0 - \alpha\sigma_0^2}{\sigma_0}\right) + x^{\beta-1} e^{(-\beta\mu_0 + \frac{\beta^2\sigma_0^2}{2})} \Phi^c\left(\frac{\log(x) - \mu_0 + \beta\sigma_0^2}{\sigma_0}\right) \right] \quad (3)$$

其中, α 和 β 分布是上尾和下尾的帕累托系数, μ_0 和 σ_0 是对数正态中间主体的参数, Φ 代表了标准正态分布的累计概率密度函数, $\Phi^c = 1 - \Phi$ 表示标准正态分布的互补累计概率密度函数。值得注意的是,如果 $\alpha \rightarrow \infty$ 和 $\beta \rightarrow \infty$,则简单的对数正态分布嵌套于 DPLN 中。此后,Reed 等(2004)发现了 DPLN 分布的诸多性质特点,随机变量 X 可以表示为 $u \times v_1 / v_2$,其中 u 、 v_1 和 v_2 是独立的,且 u 服从以参数 μ_0 与 σ_0 的对数正态分布,且服从帕累托分布。如果 $\beta > 1$, DPLN 是单峰分布;如果 $\alpha \rightarrow \infty$ 或 $\beta \rightarrow \infty$, DPLN 可以被表示为一个右侧或左侧混合的有限帕累托对数正态分布。但是,该模型无法分离出该分布的对数正态中间主体部分和帕累托左右尾部。也就是说,尽管不正式或是不完善的估计是可以得到的,还是无法用参数表示出哪个城市规模是 DPLN 上尾开始点,或哪个城市规模是下尾结束点。

基于此,本文做一个区分不同城市增长理论的研究尝试,通过对 Eeckhout(2004)的单纯吉布拉定律与 Reed(2002)广义城市增长理论进行比较,实证分析哪一个理论模型的稳定状态分布与中国城市实际规模分布相契合,从而尝试寻找到所寓意的城市增长过程。

(二) 数据来源与描述统计

样本的不完全性往往是这类研究的一个普遍缺陷,即采用截尾数据,如 Soo(2005)使用截尾跨国数据考察齐夫定律。为减少截尾数据对研究带来的偏差,本文采用由联合国人口司经济与社会事务部编撰的《世界城市化展望(2014年修订本)》数据,涵盖中国所有市域城镇人口达 30 万(以 2014 年为基准)的城市(见表 1)。在这些城市样本中,城市规模是指年末常住人口数量,其中 2015 年数据为基于 2014 年的预期值,数据的门槛值是在 2015 年市域城镇人口在 30 万以上的城市(1995 年时最小城市规模仅为 6 万人)。表 1 显示了从 1995 到 2015 年间的 5 个年份观测样本的描述统计信息。从表 1 中可以看出样本覆盖人口占历年全国城镇总人口的 63% 左右。1995~2015 年,最大城市的绝对规模增加额是最小城市绝

^① 尤尔过程是指一个群体内的成员通过类似分裂的方式产生新成员,成员没有死亡,且无相互作用,产生新成员的概率相同。

表 1 描述统计

	年 份				
	1995	2000	2005	2010	2015
样本数(个)	398	398	398	398	398
覆盖率(%)	62.43	65.94	64.08	63.42	64.42
最小值	60.89	109.07	173.05	238.27	308.30
最大值	10449.53	13958.98	16763.46	19979.98	23740.78
平均值	551.71	760.55	904.99	1067.33	1246.94
中位数	265.65	381.91	444.88	524.10	610.35
方差	922573	1646489	2394906	3452985	4871985

资料来源:United Nations, 2014。

对规模增加额的 53 倍,中位数与平均值均不断提升,但平均值的绝对增长幅度约为中位数绝对增长幅度的 2 倍,且方差增加了 5 倍,由此可见,中国大城市的人口规模增长速度要远高于小城市的人口规模增长速度。

三、模型估计与检验

(一) 模型估计

本文回归方法为最大似然估计法,需要分别找到拟合对数正态分布(LN)与双重帕累托对数正态(DPLN)分布的参数。对于 LN 来说,只需按照密度函数连乘的常规方法就可推导出其对数似然函数,参数系数值可以直接用最大似然方法回归分析得出。对于 DPLN 分布,由于有 4 个参数需要估计,通常不容易找到可行的初值赋予 α, β, μ_0 和 σ_0 。Reed(2002)提出一个赋予 α 和 β 初值的方法是:对大城市样本进行(对数 - 对数水平上)其降序排列对其规模的回归;对小城市样本进行(对数 - 对数水平上)其升序排列对其规模的回归,运用所获得的 α 与 β 初值通过矩估计法可以找到 μ_0 和 σ_0 的初值。但该方法不太适用于本研究,为此本文采用 R 软件中的 NLM 函数,这是一个使用牛顿范式算法寻找到特定函数最小值的函数,进而可获得进行最大似然估计的初始值。使用 NLM 函数可先求解目标函数的非线性最小二乘问题,拟合出参数的系数值,用该系数值作为初始值,再利用 R 软件中的最大似然估计函数 MLE 函数结合前一步得到的初始值进行最大似然回归。两个分布的最大似然回归参数系数值及各自的对数似然值如表 2 所示。

表 2 LN 与 DPLN 分布的参数系数值

	1995 年		2000 年		2005 年		2010 年		2015 年	
	LN	DPLN	LN	DPLN	LN	DPLN	LN	DPLN	LN	DPLN
$\hat{\mu}$	5.769	-	6.149	-	6.326	-	6.488	-	6.640	-
$\hat{\sigma}$	0.902	-	0.825	-	0.810	-	0.803	-	0.803	-
$\hat{\alpha}$	-	8.834	-	12.071	-	12.807	-	18.093	-	14.816
$\hat{\beta}$	-	45.283	-	45.067	-	45.000	-	53.848	-	54.490
$\hat{\mu}_0$	-	1.650	-	1.747	-	1.782	-	1.825	-	1.837
$\hat{\sigma}_0$	-	0.100	-	0.100	-	0.100	-	0.100	-	0.100
$\ln L_j^i$	-523.7	-491.9	-488.3	-445.1	-480.7	-445.1	-477.7	-432.5	-477.6	-432.2

注: $\ln L_j^i$ 是分布 j 在 i 年的对数似然值, $j=LN, DPLN$ 。

依据表 2 中估计出来的 LN 与 DPLN 分布的参数系数值,对中国城市人口规模分布的实际分布和模型拟合分布进行图形化分析(见图 1)。图 1 描绘了样本中 5 个年份中国城市人口规模分布的情况,其中实线为样本中的中国城市对数化后的核密度估计分布,即实际分布,是假定正态分布并采用 Silverman 的最优带宽测度中国城市对数化后实际规模分布的非参数核密度估计(KDE)。

值得注意的是,估计的 DPLN 分布的上尾参数系数($\hat{\alpha}$)在任何年份都比 1 大,表明中国城市规模分布规律与齐夫定律吻合程度不高。但需要谨慎比较以上的估计结果与文献中的齐夫系数,即与拟合的帕累托分布的形状参数比较。为此,通过对 1995 到 2015 年的 5 个年份观测值做标准齐夫定律的位序—规模回归 $\ln \text{rank} = \ln c - \eta \ln \text{size}$,具体估计值如表 3 所示。

从表 3 的齐夫模型估计结果来看,1995 到 2015 年间的 5 个年份中,中国城市人口规模采用近似截尾数据估计的标准位序—规模回归的系数绝对值($\hat{\eta}$)均大于 1。而且 Wald 检验在 10% 的统计显著性水平上也拒绝 $\hat{\eta}$ 等于 1。由此可见, $\hat{\alpha}$ 偏离 1 属于正常情况,这意味着,从中国当前的城市人口规模分布现状来看,并不适合采用齐夫定律进行分析,典型的幂律分布在大城市间并不存在的。不过,对于与上尾参数不同的下尾参数 $\hat{\beta}$,已有文献中并没有提供一个可以进行对照分析的基准点,而且在下尾处也没有一个幂律分布的理论说明一个国家规模第二小的城市是规模最小的城市的两倍。但从表 2 中可以看出, $\hat{\beta}$ 回归的结果均远大于 1 的。虽然之前的文献均显示齐夫系数($\hat{\eta}$)是高度敏感于所选择的城

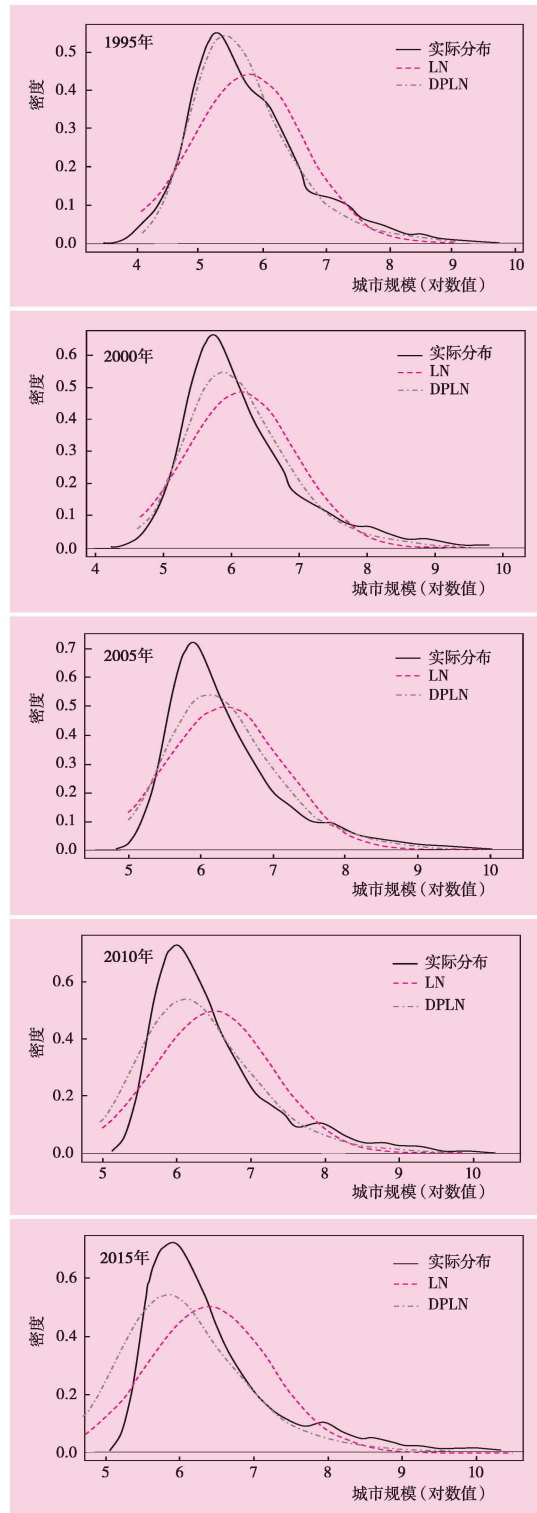


图 1 中国城市人口规模的实际分布与拟合分布

表3 标准齐夫定律的位序—规模回归

	年 份				
	1995	2000	2005	2010	2015
$\ln \hat{c}$	11.083	12.164	12.542	12.806	12.992
$\hat{\eta}$	-1.055	-1.166	-1.193	-1.204	-1.204

注:Wald 检验的统计显著性水平为 10%。Wald 检验 $H_0: \hat{\eta}=1$, 拒绝 H_0 。

(二) 模型检验

1. 非正式检验

为了能够比较两个拟合分布与实际分布的偏离情况,在图 2 左侧画出了经验分布与拟合分布的分布函数值(CDF)的逐点垂直偏差;右侧则为该偏差的累计值。如果以城市规模对数值 7 为上尾和中间部分的分界点,则可以发现在城市规模值大于 7 的上尾的逐点垂直偏差(左列图)均是 DPLN 比 LN 更靠近 0,即 DPLN 的偏差更小,从而说明在上尾部分 DPLN 比 LN 更好地拟合实际情况。标准的 Kolmogorov-Smirnov 检验(KS-test)可以判断整个分布逐点差异的最小上界或上确界。在本文样本数据的 5 个年份里,这个上确界对于 LN 显然大于 DPLN。因此,KS 检验会更早地拒绝 LN 的参数化结果。同时,历年的 DPLN 的累计偏离值均一致地落在 LN 的下方,也喻示 DPLN 实现了比 LN 更好的参数化结果,即使图 1 显示的 DPLN 拟合效果欠佳。

2. 正式检验

正式检验通过将相对应的理论分布与经验分布的垂直逐点偏差加总,得到累计偏离值(见表 3)。例如,在 2010 年累计偏离值对 LN 为 15.494,对 DPLN 为 7.952(见图 2 右侧图形)。所有 5 个年份的累计偏差值均显示了同一结果,即 LN 大于 DPLN。意味着 DPLN 模型比 LN 模型能够更好地拟合实际分布。另外,从统计的观点来看,DPLN 由于具备更加弹性的函数形式从而具有一定的优势。本文利用在表 2 中给出的对数似然值来计算赤池信息量准则(AIC)和施瓦茨准则(BIC)。具体检验结果如表 3 所示。依照 AIC 和 BIC 选择标准,二者均支持产生更小数值结果的分布。AIC 和 BIC 的结果均显示,连续 3 年 DPLN 一致小于 LN。由于 BIC 准则在对添加变量的惩罚度上强于 AIC 准则,所以从 BIC 结果可知,具有简单模型的对数正态分布(LN)不足以完整描绘实际情况,从而更偏向于较为复杂与丰富的模型 DPLN。

在给定 LN 是嵌套于 DPLN 中的情况下,用标准的似然比检验(LR)来比较这两个模型。 $\ln L_{DPLN}^i$ 和 $\ln L_{LN}^i$ 分别表示 i 年份两个拟合分布的对数似然值,该检验统计量 $LR^i=2(\ln L_{DPLN}^i - \ln L_{LN}^i)$,由于 DPLN 比 LN 多出两个参数,所以服从 $\chi^2(2)$,该检验的原假设是 DPLN 并没有带来模型拟合的显著改进,从表 4 中 LR 行的结果可以发现,可以拒绝该原假设。最后,贝叶斯因子是另外一个进行模型比较的准则。该准则没有模型嵌套的前提要求,从而比似然比检验更具灵活性。Kass 等(1995)指出,贝叶斯因子可以很容易采用施瓦茨准则进行估计。所

市规模的门阙值,但本文所采用截尾样本数据的门阙值较小,且确保了所选 5 个年份的城市相对稳定,从而在一定程度上降低了截尾数据的影响。下面本文将对拟合的 LN 与 DPLN 进行“非正式”(图形)及正式(参数)的检验。

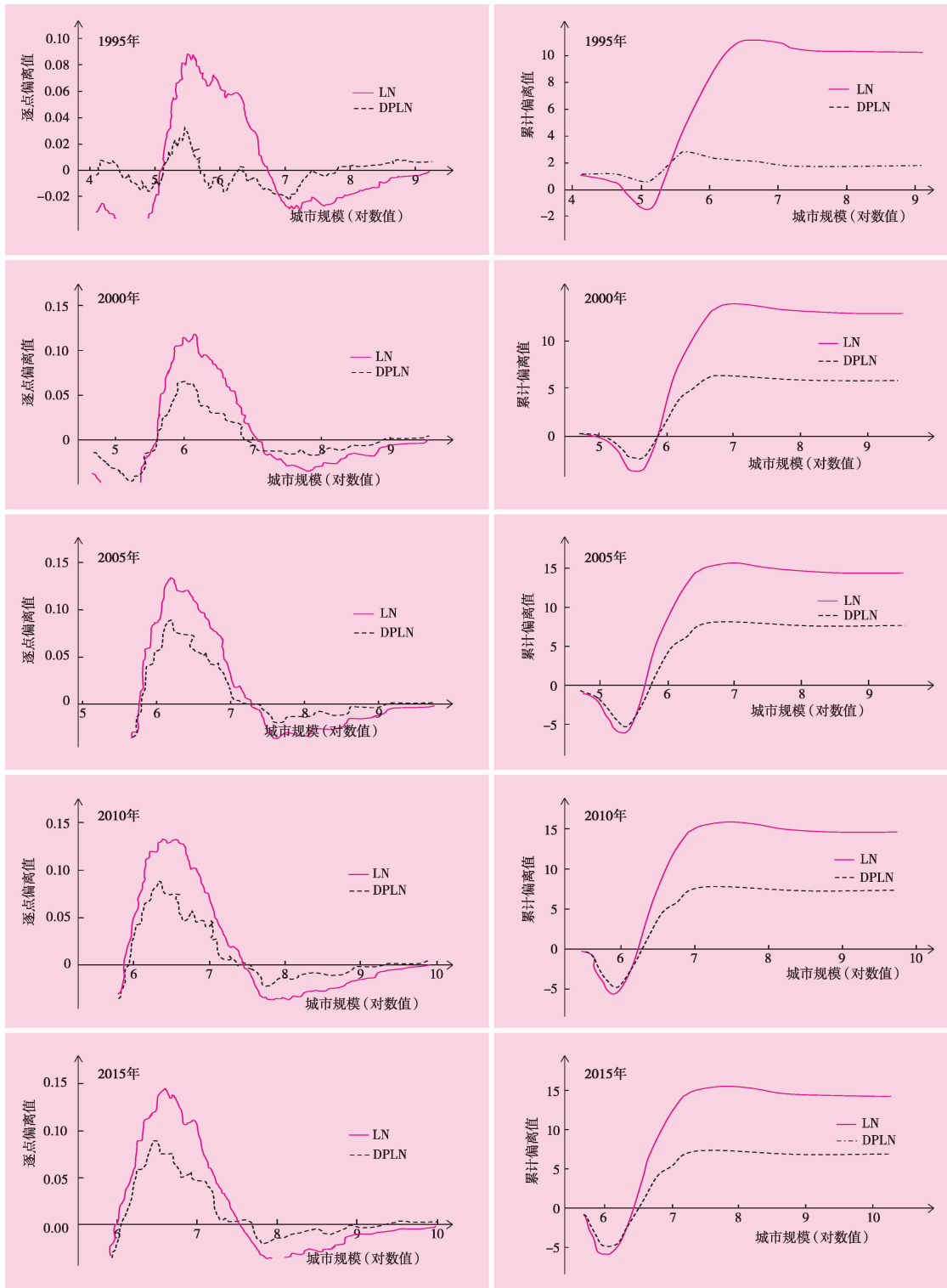


图 2 中国城市人口规模分布拟合与实际的偏离测度

表4 模型比较

	1995年		2000年		2005年		2010年		2015年	
	LN	DPLN	LN	DPLN	LN	DPLN	LN	DPLN	LN	DPLN
累计偏离值	10.302	0.742	13.034	5.811	14.505	7.952	15.494	7.952	15.666	7.952
AIC	1051.381	991.897	980.607	898.124	965.328	877.289	959.349	873.063	959.150	872.385
BIC	1059.355	1007.842	988.579	914.070	973.301	914.070	967.321	889.008	967.123	888.332
LR	63.486		86.482		71.204		90.286		90.764	
贝叶斯因子	6.517e-12		6.615e-17		1.374e-13		9.875e-18		7.776e-18	

注:第一行数据报告了对应的3个回归的理论分布与经验分布的累计垂直偏差。*AIC*是赤池信息量准则,按照公式 $AIC_j = 2k_j - 2 \ln L_j^i$ 计算出 i 年和 j 分布的值;*BIC*为施瓦茨准则以公式 $BIC_j = k_j \ln N^i - 2 \ln L_j^i$ 计算得出;其中 k_j 为 j 分布中自由参数或变量的个数, N^i 为 i 年观测变量(城市)的个数, $\ln L_j^i$ 为表2中报告的对数似然值。这两个模型选择准则均支持产生更小数值结果的分布。似然比检验的统计量,在5%显著水平上的假设检验的临界值为5.99。即如果 $B^i < 1/10$ 就是明显证据支持DPLN,如果 $1/10 < B^i < 1/3$ 就是适度证据支持DPLN,如果 $1/3 < B^i < 1$ 就是弱证据支持DPLN,当 B^i 大于1时则意味着支持LN(Kass等,1995)。

以本文定义所计算的贝叶斯因子的统计量为: $B^i \approx \exp(S^i)$, $S^i = \frac{1}{2} (BIC_{DPLN}^i - BIC_{LN}^i)$ 以比较DPLN和LN的拟合优良性。 B^i 计算数值可以用Jeffrey比例解释,表4中结果显示有明显证据支持DPLN。

(三) 分析讨论

为探寻城市人口规模分布与演化背后的潜在机制,本文的研究与大多经济学模型的建立与运用类似,包括理论假设、数学模型、实证分析和模型检验等4个过程,用模型检验来判断模型设计的科学性,进而判断潜在机制假设的合理性。在研究过程中,本文主要对两个有不同理论假设的模型进行对比分析:对数正态分布(LN)模型是基于固定数量城市据点及纯粹吉布拉定理的假设,即规模外部性无溢出,无新城市出现;双帕累托对数正态分布(DPLN)是基于一个类似于生物学中尤尔过程的理论假设,即城市遵循吉布拉定理随机增长,并有新城市出现。过去20年中国城市人口规模一直是单向变化,城市数量与规模均增加,所蕴含的城市增长理论说明中国城市人口规模变化存在外溢效应,这与双帕累托对数正态分布(DPLN)的假设相吻合。从模型拟合结果来看,本文5个单一年份的静态结果(5个年份可视为动态演变)均指向双帕累托对数正态分布(DPLN),加上前文对两个模型的正式与非正式检验,表明中国城市人口规模分布与变化更符合双帕累托对数正态分布(DPLN)。从本文双帕累托对数正态分布的具体拟合参数来看,双帕累托对数正态分布拟合更适合中国城市人口规模分布的实际情况。通过对样本期间每个年份的双帕累托对数正态分布参数的比较,可以发现上尾参数($\hat{\alpha}$)在任何年份都比1大,说明了中国城市实际规模分布并不是显著的幂律分布,也说明了对中国城市人口规模分布规律进行研究不能简单地使用齐夫定律或帕累托分布。而且前文的非正式检验也表明,双帕累托对数正态分布的累计偏离值相

对较小,拟合效果更好。因此,综合非正式检验和正式检验的结果,中国城市人口规模更符合双帕累托对数正态分布。虽然所拟合的对数正态分布与双帕累托对数正态分布除 1995 年外,其他 4 个年份的拟合情况并不是很理想^①,但是可以明显地发现 5 个年份的城市人口规模分布并非帕累托分布或齐夫分布所应呈现出幂律分布的单边拖尾状态,而是在上尾部分和下尾部分均可能存在幂律分布的可能。不过,虽然双帕累托对数正态分布对实际数据的拟合有优势,但该模型 4 个参数的估计略显复杂,这给双帕累托对数正态分布模型在城市人口规模分布规律研究中的广泛应用带来困难。因此,在理论研究层面,基于双帕累托对数正态分布的城市人口规模分布研究仍对当前大多采用单尾幂律分布来研究中国城市人口规模分布的理论方法及据此形成的结论形成巨大挑战。

由于城市规模分布变化不外乎 3 种情景:较小规模城市增长速度相对较快的收敛增长、不同规模城市增长速度基本相同的均衡增长和较大规模城市增长速度较快的发散增长(江曼琦等,2006),这 3 种情景大致对应 3 个发展阶段。当进入城市发展相对成熟的发达国家阶段后,城市人口规模的增长和分布变化将以均衡增长为主,表现为大、中、小城市的规模增长速度基本一致,即城市规模增长速度与城市规模大小并无直接关系(Eaton 等,1997; Black 等,1999)。从中国实际城市人口规模分布与双帕累托对数正态分布对比结果来看,实际城市人口规模分布并不对称,并显示了右尾比左尾要厚重的特征,与上尾参数($\hat{\alpha}$)一样,下尾参数($\hat{\beta}$)也显示了非一致的参数估计结果。这一发现表明,在过去 20 年里,中国城市规模变化呈现出明显的发散式增长特征,较小规模城市的生长与较大规模城市的生长相对较快,特别是 2000 年后,大城市的规模增长速度相对更快、更稳定,意味着中国城市人口规模分布正越来越远离双帕累托对数正态分布。进一步分析双帕累托对数正态分布拟合结果,可以发现中国城市规模的格局是人口规模相近的中小城市数量巨大(双帕累托对数正态分布的分布密度明显偏左和相对陡峭的下尾),规模较大的城市成长迅速且数量较多(平缓厚重的上尾)。简单而言,当前中国人口规模相对较大的城市处于增长快速通道中,而规模相对较小的城市增长动力不足^②。由于中国城市的发展一直受市场选择和政府干预的共同

① 模型拟合的方法存在优化的空间,如采用 Bootstrap 或从已获得参数的模型随机生成的新数据中重复抽样,再利用多次迭代拟合出上尾的位序—规模曲线,从而与实际的位序—规模分布曲线进行比较。但由于在当前现实操作中,该分析还存在诸多有关于代码与数据选取的问题,需要在后续的研究中重点关注。

② 据《国家新型城镇化报告 2015》公布的数据显示,中国农民工的流向以地级及以上城市为主,占 70% 以上,而流向小城镇的低于 10%(徐绍史、胡祖才,2016)。另据国家发改委城市和小城镇改革发展中心的研究,截至 2016 年 5 月,全国县级以上新城新区超过 3 500 个,若以平均每个新城新区规划人口 10 万、50 万或 100 万来估算,这些新城新区规划人口分别达 3.5 亿、17.5 亿或 35 亿,而《国家新型城镇化规划(2014~2020 年)》仅提出到 2020 年再推进 1 亿人口在城镇落户的目标。因此,即使扣除新城新区原有人口(假设原有人口和新增人口各占 50%),需要转移迁入的人口也远超能够转移迁移的人口。大量缺少吸引力的中小城市将面临发展困境。详细情况可参见新华网“新华视点”,《规划 3 500 个新城容纳 34 亿人口,谁来住?》(http://news.xinhuanet.com/2016-07/13/c_1119214468.htm)。

作用,城市人口规模增长和分布变化是这一作用过程的结果。从市场选择的角度看,当前城市人口规模的扩大可能带来城市病问题,但也能因城市人口规模增长带来城市就业机会的增加(陆铭等,2012),在一定范围内的城市人口规模增长还有利于提升农民工的城市融入程度(朱明宝、杨云彦,2016)^①,甚至适度的城市人口规模增长能带来要素集聚和技术进步进而有利于发挥城市发展的规模效率(邓智团、宁越敏,2011)。因此,这些发现可能意味着中国城市规模增长与分布变化正处于必须要经历的发散式增长阶段,这一结论在现实实践上将有助于促进对中国城市规模政策的深入讨论,以形成差异化的城市发展政策,释放不同人口规模城市发展的自身潜力和客观需求。

四、结论和政策建议

本文的主要目的是探究中国城市人口规模分布规律,并尝试对其隐含的发展机制与变化原因进行分析。相对已有研究,本文采用了5个不同年份的近似非截尾城市样本数据,基于对不同类型模型比较,最后采用多参数模型对城市规模分布进行拟合、筛选检验,发现中国城市人口规模分布与双重帕累托对数正态分布(DPLN)模型相契合。虽然本文的处理避开了已有城市人口规模分布规律研究在数据和方法上的两大陷阱:截尾数据和预设模型,但由于城市数量多和数据可得性的原因,本文的数据并非完全意义上的非截尾数据,因此研究结论将受到一定影响,但与幂律分布的相关研究比较而言,仍更具科学性和准确性。研究结果表明,(1)尽管对数正态分布(LN)在拟合一个国家所有城市实际人口规模数据上有一定的优势,但是在考虑更多参数需要估计的情况下,双重帕累托对数正态分布更符合中国城市人口规模分布的实质,表明中国城市的城市增长模式宜采用广义城市增长理论进行描述,而不是简单的单帕累托分布,即对数正态分布和齐夫定律不能准确地解释中国人口规模分布的发展与变化;(2)从1995年到2015年的5个时间点,中国城市人口规模分布符合双重帕累托对数正态分布,并且拥有对数正态分布的中间主体,在上尾展现出幂律分布的特性;(3)双帕累托对数正态分布参数变化表明中国城市规模分布变化还是发散式增长,特别是2000年以后,大中型城市发展速度较快,小城市增长相对乏力,不同于发达国家城市规模的均衡增长。

从分布模型的实质来看,城市人口规模分布变化主要是数学机制,但在一个各种要素可以平等、自由流动的世界里,在更为宏观的空间和更长尺度的时间上,各种要素及城市间的相互作用在本质上可能是随机的(Reed,2004)。因此,要解释某一个特定城市的发展与演化,可能需要借助对经济和地理等要素的综合分析,而要解释整个国家城市人口规模分布

^① 根据朱明宝、杨云彦(2016)的研究,农民工城市融入程度与城市规模间高度相关,随着城市人口规模的不断增长其融入度不断提升,市辖区人口规模在600万人左右时,农民工的城市融入最高,之后城市规模越大,农民工的城市融入度反而逐步降低,表现为“倒U”形关系。

及其变化,则可能不能局限于特性因素。从中国城市发展的现实实践来看,由于户籍制度对人口自由流动的限制和背后潜在的社会福利差异,城市规模变化呈现出显著的发散式增长特征,大城市规模增长速度快于中小城市,而且基于双重帕累托对数正态分布拟合表明,中国规模较大城市的数量和规模都偏小。或许意味着,中国规模较大城市还处于要素集聚与技术进步带来的城市规模效率上升阶段,即城市规模扩大带来的社会收益仍大于规模扩大所产生的社会成本。因此,这些研究发现可以用来检视中国已有的城市规模政策,也可为制定新的城市发展政策提供决策支持。

一是城市人口规模政策的制定应基于对当前中国城市整体发展阶段的准确判断。改革开放前,中国城市规模增长直接受到国家政策计划控制,一定程度上实现了收敛式增长,但并非市场选择的结果。随着改革开放的推进,沿海地区经济的迅速发展及人口流动的放开,虽然推行明确的“控制大城市、合理发展中等城市、积极发展小城市”的城市发展政策,但大城市出现了“补偿式”发展。而根据双帕累托对数正态分布模型所揭示的规律,如果将城市人口规模的的增长与分布变化视为一个随机的过程,在较长的时期内,人口规模分布的变化将逐步回归到均衡增长模式。当前中国继续推行“升级版”的限制大城市、积极发展中小城市的城市发展方针,在一定意义上是希望中小城市的发展可以缓解大城市和特大城市出现的城市病问题,而且其目标也与国外发达国家普遍存在的均衡增长相呼应,即最终实现均衡增长。但中国城市的整体发展是否已进入向均衡增长转变的阶段,是需要深入研究、准确判断以制定针对性城市规模政策的关键命题。

二是城市人口规模政策的制定应基于中国“城市”内涵实质的准确认识。与欧美发达国家城市相比,中国城市更多的是一个行政概念,地域上实质包含两大战略空间:“城市”和“农村”,即包括一个欧美国家意义上“城市”^①的城市化区域及“城市”外围包含大量中小城镇的农村区域。因此,不能用针对城市单一战略空间的政策管理两大战略空间,而应针对两大战略空间制定差异化的大城市规模发展政策。限制第一大战略空间大城市中心城区的人口规模,同时发挥第二大战略空间大城市“中心城区”外围包含城镇的农村区域的区位优势,在分流大城市中心城区人口的同时,也可吸引外部流入大城市的人口向这些区域集聚,最大限度地释放大城市或特大城市的发展优势。

因此,虽然在市场失灵的情况下,政策干预有利于降低成本提升效率,但城市规模政策的制定应充分考虑城市与区域的发展阶段与自身特质,并努力遵循城市规模发展变化的潜在规律,避免人为、失当的城市发展政策导致城市发展运行“走弯路”,以提升中国城市化整体推进的质量和效率。

^① 欧美国家的城市与中国有较大不同,真正城市化的区域才能称之为城市,不包括城市化外围的农村地区。欧美国家城市是城市化区域,而一个大的城市化区域可能包含多个城市,并将这个大的城市化区域称为(大)都市区或都市圈。

参考文献:

1. 邓智团、宁越敏(2011):《要素集聚、技术进步与城市生产率——基于长三角 16 城市的实证研究(1978~2008)》,《南京社会科学》,第 2 期。
2. 高鸿鹰、武康平(2007):《我国城市规模分布 Pareto 指数测算及影响因素分析》,《数量经济技术经济研究》,第 4 期。
3. 江曼琦等(2006):《中国城市规模分布演进的实证研究及对城市发展方针的反思》,《上海经济研究》,第 6 期。
4. 陆铭等(2012):《城市规模与包容性就业》,《中国社会科学》,第 10 期。
5. 沈体雁、劳昕(2012):《国外城市规模分布研究进展及理论前瞻——基于齐普夫定律的分析》,《世界经济文汇》,第 5 期。
6. 魏守华等(2015):《中国城市规模偏差研究》,《中国工业经济》,第 4 期。
7. 尹文耀(1988):《论城市人口规模适度分布与最佳分布》,《中国人口科学》,第 4 期。
8. Bettencourt, L.M.(2013), The Origins of Scaling in Cities. *Science*. 340(139):1438-1442.
9. Black, D., and Henderson, V. (1999), A Theory of Urban Growth. *Social Science Electronic Publishing*. 107(2):252-284.
10. Eaton, J. and Eckstein, Z. (1994), Cities and Growth: Theory and Evidence from France and Japan. *Regional Science and Urban Economics*. 27(4): 443-474.
11. Eeckhout, J.(2004), Gibrat's Law for (all) Cities. *The American Economic Review*. 94(5):1429-1451.
12. Eeckhout, J.(2009), Gibrat's Law for (all) Cities: Reply. *The American Economic Review*. 99(4):1676-1683.
13. Giesen, K., Zimmermann, A., and Suedekum, J. (2010), The Size Distribution across all Cities—Double Pareto Lognormal Strikes. *Journal of Urban Economics*. 68(2):129-137.
14. Huang, Q., He, C., Gao, B., Yang, Y., Liu, Z., Zhao, Y. and Dou Y. (2015), Detecting the 20 year City-Size Dynamics in China with a Rank Clock Approach and DMSP/OLS Nighttime Data. *Landscape and Urban Planning*. 137(5):138-148.
15. Kass, R., and Raferty, A. (1995), Bayes Factors. *Journal of the American Statistical Association*. 90(430):773-795.
16. Levy, M.(2009), Gibrat's Law for (all) Cities: Comment. *The American Economic Review*. 99(4):1672-1675.
17. Reed, W.(2002), On the Rank-Size Distribution for Human Settlements. *Journal of Regional Science*. 42(1):1-17.
18. Reed, W.J., and Jorgensen, M.(2004), The Double Pareto-Lognormal Distribution——A New Parametric Model for Size Distributions. *Communications in Statistics Theory and Methods*. 33(8):1733-1753.
19. Soo, K.(2005), Zipf's Law for Cities: a Cross-Country Investigation. *Regional Science and Urban Economics*. 35(3):239-263.
20. Ye X. and Xie Y.(2012), Re-Examination of Zipf's Law and Urban Dynamics in China: A Regional Approach. *The Annals of Regional Science*. 49(1):135-156.
21. United Nations(2015), *World Urbanization Prospects: The 2014 Revision*, CD-ROM Edition.

(责任编辑:朱 犁)