

中国县域流动人口分布样态及其演变*

——基于柯西分布的模拟及预测

曾永明 钟子康 刘厚莲

【摘 要】在区域人口增减分化的背景下,研究流动人口分布的统计特征及规律具有重要的现实意义。文章应用 2000~2020 年三次全国人口普查分县数据对中国县域流动人口的分布样态进行模拟和预测。研究发现,中国县域流动人口规模分布具有明显的非均衡性,存在极化分布特征。未来 30 年县域流动人口的活跃程度将持续提升,预计到 2050 年将有 1 824 个跨县流动活跃型县域,1 596 个县内流动活跃型县域。流动人口分布的不均衡性有可能加剧区域人口增减分化现象。一方面,跨县流出活跃型县域数量及流出人口规模将持续增加,加剧县域人口收缩;另一方面,跨县流入活跃型县域数量减少但流入强势县域的人数增加,越来越多的人口流入更少的县域单元,将加剧县域人口集聚。文章为评估与监测中国未来流动人口分布变化趋势,制定前瞻性的经济社会发展规划,引导流动人口有序流动提供了政策启示。

【关键词】流动人口规模分布 柯西分布 非均衡性 模拟预测

【作 者】曾永明 江西财经大学经济学院,副教授;钟子康 江西财经大学经济学院,硕士研究生;刘厚莲 中国人口与发展研究中心,副研究员。

一、引 言

改革开放 40 年来,中国流动人口规模不断扩大,已完成从“乡土中国”到“迁徙中国”的转变,高流动性成为中国人口形势中最显著特征之一(段成荣等,2022)。在普遍低生育率背景下,人口流动是区域人口增减分化的关键因素。第七次全国人口普查数据显示,中国人户分离人口高达 4.93 亿人,其中流动人口规模达 3.76 亿人,大约每 3 人中就有 1 人在流动。人口的强流动性成为影响社会经济发展的重要因素,一方面,人口集聚

* 本文为国家自然科学基金“邻里效应对流动人口相对贫困的影响机制与治理策略研究”(编号:72064018)、“数字乡村建设对脱贫地区农户收入增长与差距的双重效应研究”(编号:72263014)、国家社会科学基金“中国城市人口负增长对经济增长的影响及对策研究”(编号:20CRK003)的阶段性成果。

区将从人口要素集聚产生的“规模效应”中受益,为经济增长提供动力;另一方面,人口流失区则面临“人口收缩”带来的经济发展动能不足的问题(刘涛、卓云霞,2022)。当前人口增长模式发生转变,人口自然增长率持续走低,高度活跃的流动人口已成为各地区人口变动的主要因素,未来流动人口的分布格局将在很大程度上影响中国人口空间格局与社会经济的持续发展(刘厚莲、张刚,2023)。因此,有必要研究流动人口规模的分布及变化规律,面对人口负增长与流动人口规模扩大相互交织的新形势,把握未来流动人口的分布将会发生什么样的变化,对优化人口空间布局,促进人口与生产力及公共服务配置有效对接,实现社会经济高质量发展具有重大的现实意义。

流动人口分布研究历来是学术界关注的重点,目前已有大量学者从人口学、地理学、经济学、社会学等学科出发,对流动人口分布展开了不同视角、不同维度、不同空间尺度的立体研究,这些研究大致可以概括为以下几类。一是流动人口分布规律的研究。全球范围内大量经验证据表明,流动人口分布具有明显的集中化趋势,如日本东京、英国伦敦、中国上海等(Livi-Bacci,2017)。人口空间分布的内在逻辑是在流动中聚集,然后再流动、再聚集,直至集中到少数极小的地理空间(李鹏飞、陆铭,2022)。二是流动人口“来源地”与“目的地”的分布研究。流动人口“来源地”与“目的地”的分布和地区经济发展水平高度相关(王桂新,2022)。流动人口的“来源地”大多分布在经济欠发达地区,而“目的地”主要集中在经济发达地区(劳昕、沈体雁,2015)。三是流动人口的空间分布格局研究。空间属性是流动人口分布研究的重要内容,不少研究使用各种空间模型与可视化分析方法对流动人口的分布格局进行探究,如重心模型、空间 OD 模型、GIS 分析等(Zhao 等,2017;张建武等,2021)。随着“流动空间”不断被网络关系渗透,网络分析也逐渐成为流动人口分布研究的热点之一(张伟丽等,2021)。四是流动人口分布的影响因素研究。流动人口分布是多种因素共同作用的结果,其中包括不同地区自然、经济、社会等方面的差异,以及地理距离、社会关系网络、流动人口的主观意愿等(You 等,2018)。

从文献梳理来看,流动人口分布的空间属性与经济社会关联已得到较为系统、全面的探讨,研究成果也较为丰富;相对而言,目前探讨流动人口分布的内生统计特征及分布规律的研究较少。自然现象存在一些可被观察的统计规律,如 Paplexiou 等(2016)发现日降水量接近于广义伽马分布,Alavi 等(2016)发现风速更接近威布尔分布等。同样,看似充满随机性与不确定性的社会现象也具有一定的统计规律。当然,社会现象的统计规律较为复杂,难以通过简单观察分析得出,需要借助复杂的数理统计分析方法,从数据分布视角尝试构建一个最优的“分布函数”进行模拟,以探讨其潜在的规律。随着社会经济的不断发展和数据信息的日益丰富,借助分布函数描述社会经济现象的数理统计分析方法被应用于多个领域。流动人口作为新时代社会经济的活跃因子,其分布是否具备特定的统计规律?能否通过构建分布函数更好地进行模拟与预测?这对于丰富流动人

口分布研究内容,更加全面地掌握流动人口分布特征具有较高的探索价值。

传统的人口预测大多围绕人口总量及其变动展开,如出生、死亡、总人口及各年龄段人口数量预测。在中国人口负增长背景下,少子化、老龄化、劳动力数量减少这些人口总量变化趋势已基本达成共识,总量预测的现实价值因此有所下降(陈卫,2022)。少数研究对人口空间分布、人口迁移流等特征展开预测。例如,曾永明(2016)利用对数正态分布函数对中国人口空间分布不均衡状况进行了分布形态拟合及预测,为人口长期均衡发展提供了有益探索。流动人口作为影响未来中国人口分布的核心因素之一,其分布预测意义重大,但相关研究较为匮乏。戚伟与刘盛和(2015)拟合了中国城市流动人口位序规模分布,发现其符合幂函数规律;不过该研究并没有进行预测分析。同时,相较于省级与地级市尺度,县域作为中国人口流动的基础空间单元,县域尺度下的流动人口分布更具突出性与代表性。2022年5月中共中央办公厅、国务院办公厅印发《关于推进以县城为重要载体的城镇化建设的意见》指出,要“尊重县城发展规律”“顺应县城人口流动变化趋势”。可以说,流动人口在县域城镇化中发挥重要作用。

鉴于此,本文从县域尺度出发,探索流动人口分布的统计特征及其变化规律,尝试对县域流动人口的分布样态进行模拟,将复杂的社会现象具体化、模型化和规律化,并在此基础上对中国县域流动人口分布进行预测。进而,本文的边际贡献在于:一是基于人口统计视角,对中国县域流动人口分布的“统计特征及分布规律”进行拟合,丰富流动人口研究视角;二是跳出总量及趋势预测等传统视角,对流动人口进行中长期的分布预测,更加符合人口负增长背景下的现实需要,以期为促进人口与经济社会协调发展提供参考;三是探索县域尺度下人口流动规律和变化趋势,为推进以县城为重要载体的城镇化建设提供理论支撑。

二、概念界定与数据来源

(一) 概念界定

县级行政区域(简称“县域”)作为一个空间概念,具有广义与狭义之分。狭义上的县域不包含市辖区,特指县级市、县、自治县、旗、自治旗、林区和特区等县级行政单元。广义上的县域则泛指中国行政区划体系中级别为县的所有行政单元,市辖区也包含在内。市辖区作为地级市的核心组成部分,是城市经济发展的中心,城镇化水平普遍较高。相较于其他县级行政单元,市辖区具备更强的人口吸纳与集聚优势,是县级尺度下流动人口集聚的高密度区。为降低市辖区自身人口吸附力对研究结果的干扰,本文最终选择狭义县域作为研究对象,不将市辖区纳入县域范畴,以期精准地识别县域尺度下流动人口分布的规律与变动趋势。2000~2020年,中国县级行政区划发生了多次调整,主要体现

在行政区划名称调整(撤县设市或区、撤市设区等),新增或撤销、拆分或重组行政区划,导致研究期间县域单元数量不一致,本文的处理方式是以2020年县级行政区划为基准,调整、增添或删除2010和2000年的县域单元数量。基于此,本文在剔除973个市辖区后,最终保留1871个县域。

中国人口统计意义上的流动人口通常为“人户分离人口”(居住地与户籍地不一致),从第五次全国人口普查开始,流动人口统计标准为离开户口登记地所在乡镇街道半年以上的人户分离人口。根据流动人口跨越区域行政边界的不同,县域流动人口可以划分为三类:跨县流入人口、跨县流出人口和县内流动人口。其中,跨县流入人口是“居住在本县(市、区),户籍地在本县(市、区)以外”的人口;跨县流出人口是“户籍地在本县(市、区),居住地在本县(市、区)以外”的人口;县内流动人口是“户籍地与居住地均在本县(市、区),但户籍地与居住地所在的乡镇街道不一致”的人口。目前研究大多专注于跨县流动人口,忽略了县内流动人口这一群体(段成荣等,2022)。县内流动反映县域人口流动的不同侧面,并与跨县流动共同影响县域人口分布变动及社会经济发展。因此,本文关注的重点是跨县和县内流动人口的分布样态及其变动特征。

(二) 数据来源

本文县域流动人口数据主要来自第五、第六、第七次中国人口普查分县资料。其中,跨县流入人口根据“户口登记地在外乡镇街道”条目中的“本省及其他县(市、区)”和“省外”人口转换得到。县内流动人口根据“户口登记地在外乡镇街道”条目中的“本县(市、区)”人口得到。分县普查数据中没有直接公布跨县流出人口数据,但可以依据户籍人口与常住人口计算净流动人口。流入人口与本地未流动的常住人口之和等于所在县市的常住人口;而流出人口与本地未流动的常住人口之和等于所在县市的户籍人口。因此,常住人口与户籍人口之差等于净流动人口(戚伟等,2017)。据此,流出人口可根据以下公式计算:常住人口=户籍人口+流入人口-流出人口。但由于《中国人口普查分县资料—2020》中并没有直接公布县域尺度的“户籍人口”数据,本文利用《中国县域统计年鉴(2020)》中各县域单元公布的户籍人口数据代替。需要说明的是,分县普查资料与县域统计年鉴在户籍人口的统计口径上虽存在一定偏差,但县域统计年鉴是除人口普查数据外县级尺度下最准确的人口数据来源,且大部分统计部门发布的统计年鉴均会根据最新人口普查数据对以往年份的人口行政统计数据数据进行一定修正,因此本文暂时忽略这种与数据源相关的统计偏差。

为了更加详细地分析县域流动人口分布的规模与流动强度,本文利用各县“(跨县流入人口-跨县流出人口)/常住人口”表示跨县净流动率 I ,以综合反映各地区跨县流动人口的规模、流动强度及流动方向,当 $I>0$ 时,表明该地区以人口流入为主, I 越大,说明该地区人口流入的强度越高,流入人口的相对规模越大;当 $I<0$ 时,该地区以人口

流出为主, I 的绝对值越大,说明该地区人口流出的强度越高,流出人口的相对规模越大。利用各县的“县内流动人口/户籍人口”表示县内流动率 K ,代表县内流动人口的规模及流动强度; K 越大,说明该地区县内人口流动强度越高,流动人口的相对规模越大。

三、中国县域流动人口规模的分布样态描述

在改革开放以来较长一段时期内,中国流动人口分布呈较为明显的非均衡性特征,突出表现为人口从中西部地区大量流入东部沿海地区的不均衡态势。这种不均衡格局由地区间的流动人口规模差异导致,随着时间的推移,各地区流动人口规模会发生变化,区域间的差异也会随之改变。

为探讨这种由数量变化背后的统计规律,从而较为准确地估计流动人口规模变化的客观规律和发展趋势。本文从县域流动人口规模数据内生的统计分布特征出发,通过绘制跨县与县内流动人口规模的频率分布直方图,寻找最为适合的统计分布,以期发现其潜在分布规律。

如图1所示,以2020年跨县与县内流动人口规模为例,直观上看,跨县净流动率与县内流动率均表现为随着流动率的增大,频率先上升后下降的特征,符合这种变化态势的概率分布函数包括正态分布、泊松分布、威尔布分布、柯西分布等。不过,本文数据显示,净流动人口有正有负,这与威尔布和泊松分布要求变量值大于0的情形并不相符;正态分布又称“钟形曲线”,要求数据两端具有较为“对称”的分布形态,且偏度为0,这也与现实中的流动人口分布不完全吻合。相比之下,柯西分布存在“偏态”且具有“长尾”特征,尾部比正态分布更长更厚,可以更为准确地反映出分布两端“长尾”数值的特征。Rhee等(2011)利用GPS轨迹数据研究发现,人类日常流动存在边界效应,也即人们倾向于在预先规划的活动区域内移动,跨越这些活动区域边界流动的可能性很小。这一特征与柯西分布的“长尾”现象很吻合。因此,人口流动在统计分布上更符合具有“长尾”特征的柯西分布,而非正态分布。同理,宏观层次的人口流动与微观个体的流动具有相似的统计分布特征,人口在空间上的流动也具有“路径依赖”性,并且流动过程中存在趋利性行为,流入地往往是经济发展水平较高的地区,因此流动人口分布具有较为明显的极化特征,与柯西分布的“长尾”特征较为契合。图1显示,跨县流动人口规模的频率直方图呈现左侧偏长,右侧较短的特征,县内流动人口规模的频率直方图呈现出右侧偏长,左侧偏短特征,这说明流动人口规模分布具有明显偏态,这与正态分布所要求的偏度为0相悖,与具有“长尾”特性且存在偏态的柯西分布更为契合。

基于以上分析,本文选用柯西分布函数进行拟合。结果显示,跨县流动人口规模的拟合优度为0.9840,县内流动人口规模的拟合优度为0.9767,曲线拟合结果与跨县及县内流动人口规模分布的现实状况高度吻合。为验证柯西分布函数的有效性和适用性,本

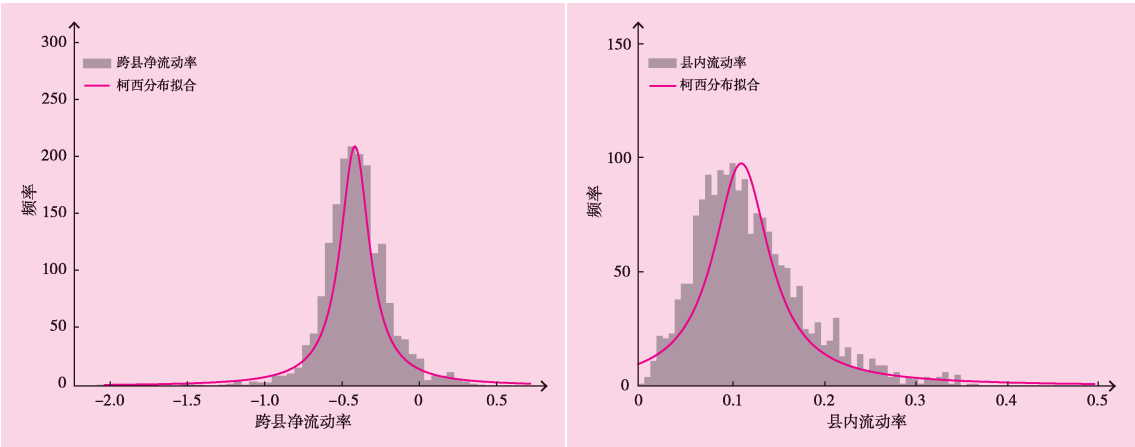


图1 2020年县域流动人口规模频率直方图及柯西分布拟合

文进一步利用分布拟合检验中常用的 Kolmogorov-Smirnov (KS) 检验,结果印证了柯西分布拟合的有效性。因此,本文的初步分析结果表明,县域流动人口规模分布服从柯西分布。本文以下分析将在此基础上展开。

四、基于柯西分布的流动人口规模概率分布函数拟合

(一) 柯西分布函数解析及其人口学涵义阐释

柯西分布 (Cauchy distribution) 又称柯西-洛伦兹分布或洛伦兹分布,作为概率分布时,常被称为柯西分布。该分布是一个连续型概率分布,其概率密度函数可以表示为:

$$F(X;\rho;\gamma)=\frac{1}{\pi}(\frac{\gamma}{(X-\rho)^2+\gamma^2}) \tag{1}$$

其中 $-\infty < X < \infty$ 。 ρ 为位置参数,表示分布峰值的位置,决定了概率分布的大致位置; γ 为尺度参数,表示最大值的 1/2 处宽度的一半,决定了概率分布的宽度;取值越大时分布越分散,越小则分布越集中。当 $\rho=0$ 且 $\gamma=1$ 时,式(1)为标准的柯西分布。

本文将县域流动人口规模拟合柯西分布函数,借助柯西分布的“长尾”与“尖峰”分布特性对流动人口分布的集中趋势与分化程度进行估计和讨论,赋予这两个参数人口学涵义。首先,位置参数 ρ 是流动人口分布的规模参数,代表了所有县域流动人口规模最集中的分布水平,反映县域流动人口规模的总体水平, ρ 越大,县域流动人口总规模越大;其次,尺度参数 γ 是流动人口分布的均衡参数,代表县域间流动人口规模分布的均衡程度, γ 越大,县域流动人口规模分布越不均衡,即地区间流动人口规模差异越大。非均衡性是中国人口空间分布研究持续关注的议题,在真实场景中,流动人口分布规律对人力要素的循环、物质资源的配置及区域社会经济发展格局均有重大影响,是宏观经济调控和社会发展规划需要考虑的因素。

根据不同参数的变化,可以衡量不同年份流动人口规模差异及变化规律。结合式(1),柯西分布函数的位置参数 ρ 与尺度参数 γ 采用极大似然估计法估计得到。

(二) 拟合结果与分析

本文使用 2000~2020 年跨县与县内流动人口规模数据分别拟合柯西分布函数,估计得出相应的位置参数与尺度参数,分别带入式(1),得到 2000、2010 和 2020 年全国跨县与县内流动人口规模的柯西分布函数。为直观起见,本文基于估计结果依次绘制成了概率密度函数曲线(见图 2)。

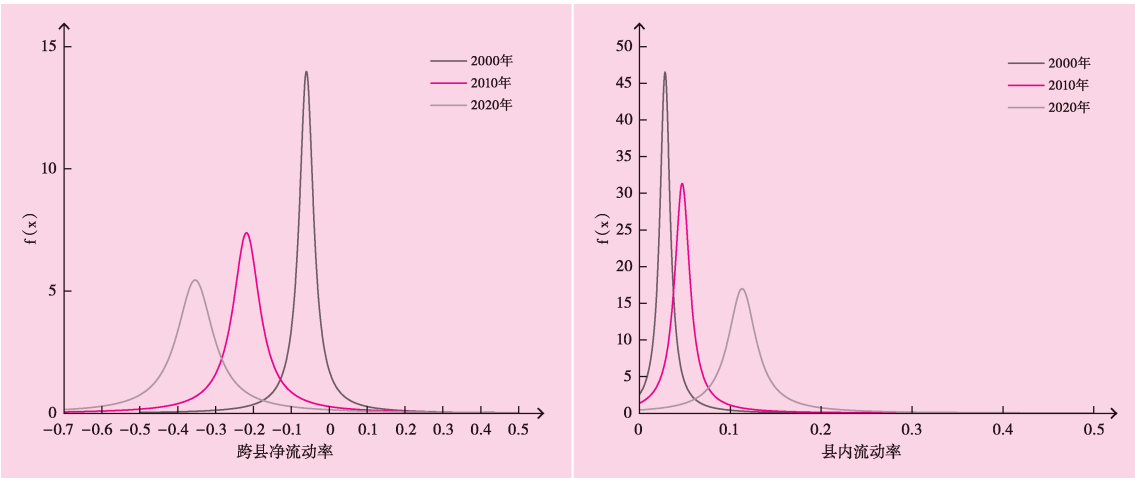


图 2 流动人口规模分布拟合曲线

如图 2 所示,3 个年份的拟合曲线呈不同形态,但也具有一定的规律性变化趋势。就跨县流动而言,2000~2020 年的位置参数估计值分别为 -0.0602、-0.2184 与 -0.3539,呈负向增加态势;尺度参数分别为 0.0455、0.0863 与 0.1168,呈正向增加态势;曲线整体逐步向左偏移,趋向扁平化。这说明,随着县域城镇化水平不断提高、基础设施日益完善,县域间人口流动越来越频繁,跨县流出人口规模不断扩大。与此同时,跨县流动人口分布的不均衡程度逐渐加深,区域规模差距不断扩大,这可能反映了人口在流动过程中的趋利性或选择性特征。这一现状与跨县流动人口以经济型流动为主有关,以务工经商为主要流动原因必然意味着流向选择的非随机性。不过,这在客观上也导致了县域流动人口的增减分化。从县内流动来看,2000~2020 年的位置参数估计值分别为 0.0282、0.0471 与 0.1130,呈正向增加态势,曲线整体向右偏移;尺度参数同样呈正向增加态势,分别为 0.0137、0.0203 与 0.0375,表明分布越来越分散。这些结果表明,以“乡城流动”为主的县内流动人口规模总体水平不断扩大,县域单元内部流动人口活跃程度不断提高。同时,由于各县域单元自身经济发展水平差异较大,县内流动人口规模也存在明显的区域异质性。

五、基于柯西分布的县域流动人口规模概率分布函数预测

(一) 流动人口规模分布预测的约束模型

根据前文分析可知,柯西分布函数的曲线形态由位置参数 ρ 与尺度参数 γ 共同决定。结合式(1),要确定预测时点流动人口规模的柯西分布函数,需要首先对预测时点的 ρ 与 γ 进行估计。

关于分布函数的参数估计,Flowerdew 等(1988)在利用泊松分布拟合英国城市间人口流动数据后,比较了无约束模型与约束模型的拟合优度与参数估计结果,发现利用约束模型进行参数估计能够改进拟合优度与估计结果。类似地,Asadoorian(2008)利用贝塔分布拟合全球人口空间时发现,宏观经济变量是影响空间分布函数的参数,利用这一经验关系可以拟合条件模型并改进参数估计。本文在借鉴这些文献的基础上,引入“约束模型”进行参数预测。事实上,不同水平的参数变化衡量了不同年份流动人口规模的分布差异及变化规律,是流动人口分布不均衡的表征。因此,影响流动人口分布的因素即决定参数大小的“约束条件”,将这些因素带入约束模型中,便能对未来中国流动人口规模柯西分布函数的参数进行有效估计,具体如下:

$$\rho_i = \alpha \ln Z_i + C_1 + \varepsilon_\rho \tag{2}$$

$$\gamma_i = \beta \ln Z_i + C_2 + \varepsilon_\gamma \tag{3}$$

其中 α 与 β 为“约束方程”的待估计系数, Z 为影响流动人口分布非均衡的因素, C 为常数项, ε 为误差项, i 为省份,本文主要考虑港澳台和北京、上海、天津以外的 28 个省份。

(二) 理论分析与变量选取

由 Bague 提出, Lee 进一步完善的“推拉理论”强调迁入地与迁出地之间的“条件”差距是诱发人口迁移流动的决定性因素(Ho 等,1970)。这种“条件”差距体现在多个维度,如气温、降水、地形、海拔等自然条件差距,经济发展水平、户籍制度约束、人口规模、公共服务供给等宏观经济社会方面的条件差距。尽管影响流动人口分布的因素非常多元,但多数因素难以进行长期预测,缺乏较为权威的预测数据。本文在参考已有文献的基础上(Asadoorian,2008),主要从自然条件与宏观经济社会因素出发,选取了气候、经济发展水平、人口规模 3 个具有较权威的预测数据的“条件变量”,对 2000~2050 年中国流动人口规模分布的参数进行模型检验与预测分析。

首先,基于人口迁移法则、推拉理论、经济增长理论等经典理论可知,无论是人口在农村与城市之间的流动,还是城市与城市之间的流动,抑或是跨区域跨国的流动均具有一定趋利性,人口迁移流动的最终目的往往是为了追求更美好的生活。一般来说,流入地与流出地的经济差距是产生人口流动“推拉作用”的关键因子,流动人口对于流入地的选择集中在经济发达地区(曾永明等,2023)。这一特征与中国流动人口的空间分布格

局高度契合,中国流动人口主要分布在经济条件较好的东部沿海地区,中西部地区则以人口流出为主。近年来,在区域均衡战略的持续作用下,区域经济差距逐渐缩小,人口流失严重的中西部地区也开始呈现“人口回流”趋势。因此,经济发展水平是影响流动人口分布的关键因素,本文采用人均 GDP 衡量地区经济发展水平。2000~2020 年 GDP 与人口总量数据来自《中国统计年鉴》,GDP 预测数据来自麻省理工学院综合全球系统建模(IGSM)中经济预测和政策分析模型(EPPA)对中国 GDP 年增长率的预测换算结果,人口总量预测数据来自联合国人口司《世界人口展望 2022》(下文简称 WPP2022)。

其次,在探究流动人口分布的影响因素时,流入地与流出地自身的人口规模也是不可忽视的重要因素,且二者有密切的关系。美国社会学家 Zipf(1946)从牛顿万有引力定律的角度扩展了人口迁移的“推拉理论”,并提出了引力模型这一分析框架,他认为人口迁移会受到“引力”的影响,两地间人口迁移流动的规模与两地人口规模正相关(流出地与流入地之间的人口规模越大,人口流动规模也会越大),与两地间距离负相关。美国经济学家 Krugman(1991)同样在新经济地理学中利用“中心—外围”理论强调了规模效应在吸引流动人口过程中的作用,他认为流动人口具有集聚效应,大城市或特大城市往往汇聚了一个区域绝大部分的流动人口。中国不仅人口规模巨大,人口空间分布也极不均衡,总体呈现出以“胡焕庸线”为界的东密西疏格局,人口规模的不均衡分布对流动人口分布产生了较强的影响。因此,本文采用衡量人口规模常用的指标人口密度表征地区人口规模,假定研究期内地区行政区划不变,人口密度为地区人口总量除以国土面积。2000~2020 年人口总量数据来自《中国统计年鉴》,预测数据延用 WPP2022 的预测数据。

最后,气候变化对人类的迁移流动行为具有极为重要的影响(Bittle, 2023)。根据联合国政府间气候变化专门委员会(IPCC)发布的报告《气候变化 2023》显示,2011~2020 年全球平均气温相较于 1850~1900 年已上升 1.1℃,气候变暖带来的危害远超预期,海平面上升、冰川消融、洪水、干旱等极端天气、城市热岛效应等现象均已达历史之最,产生的“气候变化移民”或“环境难民”数量剧增。由于区域气温差异诱发的季节性人口迁移流动现象更是屡见不鲜,如东北地区大规模向海南迁移的“候鸟”式人群。因此,长期来看,气候变化是预测中国未来流动人口分布不可或缺的重要因素。本文采用年平均气温指标衡量地区气候变化情况,2000~2020 年的数据来自国家气象科学数据中心,预测数据来自 IPCC 第六次评估报告(AR-6)^①基于中等二氧化碳排放水平对全球气温的中长期预测。需要特别说明的是,报告中预测全球年均气温将在 2021~2040 年上升 1.2~1.8℃,2041~2060 年上升 1.6~2.5℃,2061~2100 年上升 2.1~3.5℃。由于没有各年份的具体预测数值,本文利用插值处理等方式得到所需年份的预测数据。

^① <https://www.ipcc.ch/assessment-report/ar6>.

(三) 模型检验结果与分析

将 2000、2010、2020 年人均 GDP、人口密度和年平均气温 3 个变量带入“约束方程”进行模型拟合,结果如表 1 所示。

表 1 约束模型估计结果(N=84)

变 量	跨县 ρ	跨县 γ	县内 ρ	县内 γ
人均 GDP	-0.111***(-8.64)	0.037*** (6.85)	0.021*** (4.17)	0.005** (2.94)
人口密度	0.417** (3.34)	-0.094* (-1.80)	0.036 (0.73)	0.035* (1.96)
年平均气温	-0.074*** (-3.56)	0.001 (0.10)	0.032*** (3.88)	0.006** (2.11)
常数项	-0.341 (-0.55)	0.184 (0.71)	-0.696** (-2.84)	-0.282** (-3.17)
R ²	0.8344	0.6754	0.7391	0.6163

注:括号内数字为 t 值,*、**、*** 分别表示在 5%、1%、1%水平上显著。

人均 GDP、人口密度和年平均气温 3 个“条件变量”的系数大多通过了显著性检验,这说明所选“条件变量”对流动人口规模分布的参数有显著影响,是预测参数大小的重要条件。根据前文分析可知,影响流动人口分布的因素很多,但由于其他大部分“条件变量”缺乏可靠的长期预测数据,本文选择以上 3 个变量进行探索性分析。为了检验遗漏其他变量是否会影响模型估计结果,本文采用 Cinelli 等(2020)提出的敏感性分析方法检验模型结果的稳健性。检验结果表明,各参数拟合结果的决定性系数(R²)均在 0.6 以上,跨县 ρ 的 R² 达 0.8 以上,说明模型能较好地进行参数预测。后续研究若能收集到更多流动人口分布的“条件变量”加入到模型中,预期将能达到更好的拟合和预测效果。

结合式(2)和式(3),将各项指标的估计系数带入模型,可外推出 2000~2020 年各参数的预测值。通过与实际值相比较、计算二者之间的相对误差率及置信区间预测等方式来检验模型对参数预测的准确性。如表 2 所示,2000~2020 年各参数的实际值与预测值较为接近,二者之间相对误差率普遍较小,大多都在[-5%,15%]之间,最小误差率仅为-1%。同时,各参数的实际值均落在 95%置信水平的参数预测区间内,再次验证了预测结果的可靠性。相对误差率较大的年份,如 2000 年县内 γ 为-21%,虽是几个参数预测中的最大误差,但其实际值仍落在 95%置信水平的参数预测区间内,与预测值之间的差值也仅为 0.014。作为流动人口规模分布预测研究的创新性探索,这一误差率在可接受的范围内,在整体外推检验效果普遍较好的情况下,能够支持模型的预测结果具有一定的可靠性。

(四) 模型预测结果与分析

基于上述分析,将“条件变量”的预测数据带入模型中便可得出 2021~2050 年中国跨县与县内流动人口规模分布的位置参数与尺度参数。限于篇幅,本文仅报告和分析 2025、2030、2035、2040、2045 及 2050 年 6 个年份的预测结果,结果如表 3 所示。

参数预测结果显示,跨县流动人口规模分布的位置参数呈负向增大,尺度参数呈正向增大趋势,县内流动人口规模分布参数也呈类似发展趋势。这说明,在未来 30 年,县域流动人口规模总体水平将不断扩大,县域人口流动现象愈发频繁,这对于推进以县域为主体的新型城镇化具有积极作用。这一过程中,县域流动人口分布的不均衡性也将扩大,流动人口分布极化特征越发明显,由此可能加剧区域人口增减分化。

当然,预测的目的不仅仅是进行参数预测,更重要的是挖掘参数变化

背后所体现的流动人口规模分布特征及变化规律,以及所蕴含的政策启示。将 2030、2040、2050 年的参数预测值代入式(1),得出 3 个年份的流动人口规模分布概率密度函数曲线,结果如图 3 所示。图 3 同时展示了 2020 年的流动人口规模分布函数曲线,用于现状与未来的对比分析。

借鉴国际上判断流动人口规模常用的标准进行划分:当地区人口迁移量超过 10%时表明流动人口规模较大,具备高活跃性特征(Jones,1990)。如图 3 所示,2020~2050 年跨县流动人口规模分布曲线不断向左偏移,且持续扁平化;表明跨县净流动率小于 -10%的人口流出活跃型县域分布范围不断扩大,而净迁移率在 -10%~10%范围内的人口流动非活跃型县域分布范围不断缩小。2020~2050 年县内流动人口规模分布则呈现整体向右偏移的扁平化趋势,县内流动率小于 10%的人口流动非活跃型县域分布范围不断缩小,而县内流动率大于 10%的人口流动活跃型县域分布范围逐渐扩大。

总体而言,跨县与县内流动人口活跃程度不断提高,加速了以县域为载体的城镇化

表 2 2000~2020 年分布参数的实际值与预测值对比

	实际值	预测值	置信区间	误差率(%)
跨县 ρ				
2000 年	-0.0602	-0.0716	[-0.0935,-0.0498]	16
2010 年	-0.2184	-0.2286	[-0.2465,-0.2107]	4
2020 年	-0.3539	-0.3375	[-0.3600,-0.3150]	-5
跨县 γ				
2000 年	0.0455	0.0518	[0.0426,0.0608]	12
2010 年	0.0863	0.0940	[0.0873,0.1007]	8
2020 年	0.1168	0.1236	[0.1142,0.1330]	5
县内 ρ				
2000 年	0.0282	0.0286	[0.0199,0.0372]	-1
2010 年	0.0617	0.0695	[0.0632,0.0759]	11
2020 年	0.1130	0.1095	[0.1006,0.1184]	-3
县内 γ				
2000 年	0.0137	0.0113	[0.0081,0.0144]	-21
2010 年	0.0203	0.0229	[0.0206,0.0252]	11
2020 年	0.0375	0.0334	[0.0302,0.0366]	-12

表 3 2021~2050 年流动人口规模柯西分布参数预测值

	2025 年	2030 年	2035 年	2040 年	2045 年	2050 年
跨县 ρ	-0.4480	-0.4962	-0.5393	-0.5851	-0.6285	-0.6768
跨县 γ	0.1443	0.1523	0.1598	0.1693	0.1790	0.1916
县内 ρ	0.1272	0.1416	0.1543	0.1671	0.1789	0.1911
县内 γ	0.0342	0.0371	0.0394	0.0417	0.0435	0.0453

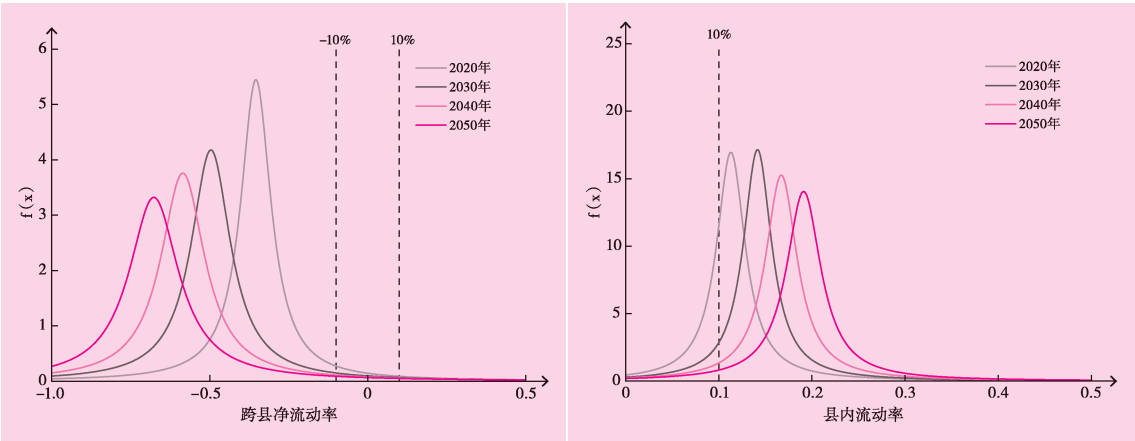


图3 2020~2050年流动人口规模柯西分布预测曲线

进程。城镇化发展会经历一个先加速、再减速的过程(简新华、黄锟,2010),其中2000~2030年跨县与县内流动人口活跃型县域分布范围增加幅度较大,此时正处于县域城镇化发展的加速期。在《关于推进以县城为重要载体的城镇化建设的意见》等多项政策的支持下,预计城镇化发展将持续补齐发展短板与弱项,不断放宽落户限制,促进县域流动人口活跃程度快速提高。2030~2050年增幅明显放缓,县域人口流动活跃程度增长放缓,此时县域城镇化发展进入减速期。

在对比预测曲线的基础上,通过计算流动人口规模分布范围的累计概率可以进一步识别流动人口规模分布范围的具体变化情况。参考戚伟等(2017)提出的原始复合指标法,从跨县和县内两个维度划分不同类型的流动人口规模。具体而言,本文将跨县流动人口规模分为7类,县内流动人口规模类型分为4类,分类标准和结果如表4所示。

首先,跨县流动人口规模类型中,以人口流出为主的活跃型县域单元($I < -10\%$)总体累计概率由2020年的86.28%上升至2050年的89.76%,若未来中国市辖区以外的县域数量保持1871个不变,到2050年将有1679个县域成为人口流出活跃县。其中,跨县流出强势区累计概率呈上升趋势,由2020年的21.47%上升至2050年的73.72%。这说明在人口流出活跃县数量增加的同时,人口流出规模也在持续扩大,到2050年将会大约有1379个县域单元人口净流出严重,导致人口收缩程度加剧。相反,人口流入活跃型县域单元($I > 10\%$)总体累计概率呈下降趋势,由2020年的8.02%下降到2050年的7.70%,下降幅度虽然不大,但在人口流出活跃县数量增加且人口流出规模大幅扩大的趋势下,以人口流入为主的县域单元数量下降意味着未来中国县域流动人口分布的“马太效应”更为明显,有更多的流动人口流入更少的县域。与此同时,跨县流入强势区的累计概率呈增加态势,由2020年的4.33%上升至2050年的5.14%,这同样说明未来中国县域流动人口的集聚特征将会更加明显,会产生一批新的“流动人口增长极”承接大量

流动人口。2020 年跨县流动平衡区的累计概率为 5.17%，到2050 年为 2.51%，呈持续下降趋势，2050 年将仅有 47 个县域处于人口流动不活跃状态。

其次，县内流动人口规模类型中，活跃型 ($K>10\%$) 县域数量持续增加，累计概率由 2020 年的 60.62% (1 134 个)

上升至 2050 年的 85.31% (1 596 个)，表明县内人口流动参与率将持续提高。2020~2050 年县内流动稳定区占比持续下降，累计概率由 39.38% (737 个) 下降至 14.69% (275 个)，而县内流动活跃区累计概率由 54.32% (1 016 个) 上升至 72.76% (1 361 个)，变化较为明显。同时，县内流动频繁区与强势区累计概率也呈缓慢上升趋势，2020 年的累计概率分别为 3.23% (60 个) 和 3.07% (54 个)，到 2050 年分别上升至 7.92% (148 个) 和 4.63% (87 个)。这说明未来城镇化水平不断提高的过程将最大限度地激活县域内部人口流动活力，本地农村人口将持续向县城流动，“乡城流动”人口规模不断扩大，进一步促进“乡土中国”向“城镇中国”转型。

综上，在“迁徙中国”形态下，人口大规模流动是社会发展的客观规律。从县域尺度来看，未来无论是跨县还是县内，都将会有更庞大规模的人口投入到流动的浪潮之中。与此同时，流动人口分布的“马太效应”将会加剧未来县域流动人口分布的不均衡性，从而凸显县域人口收缩与集聚的极化现象。一方面，人口流出活跃型县域的数量及流出人口规模将大幅增加，导致当地人口加剧收缩；另一方面，人口流入活跃型县域的数量有所减少。这意味着，有更多的流动人口流入更少的县域，新产生的人口流入强势区作为“流动人口增长极”将会吸收大量流动人口，加剧人口集聚。

六、结论与讨论

本文基于人口统计视角，构建了一个流动人口规模分布拟合及预测的研究框架，利用第五、第六、第七次全国人口普查分县数据对中国县域流动人口规模的分布样态与演进趋势进行了模拟及预测，主要研究结论包括以下几点。(1)中国县域流动人口规模的统计分布特征较为符合“长尾”“尖峰”的柯西分布，具有明显的非均衡性，且存在极化分

表 4 2020~2050 流动人口规模类型的累计概率					
分布范围	类型	累计概率(%)			
		2020 年	2030 年	2040 年	2050 年
$I<-50\%$	跨县流出强势区	21.47	49.21	64.83	73.72
$-50\%\leq I<-30\%$	跨县流出频繁区	41.99	29.78	18.11	11.31
$-30\%\leq I<-10\%$	跨县流出活跃区	22.82	9.33	6.37	4.76
$-10\%\leq I\leq 10\%$	跨县流动平衡区	5.70	3.72	2.98	2.51
$10\%< I\leq 30\%$	跨县流入活跃区	2.39	1.94	1.69	1.53
$30\%< I\leq 50\%$	跨县流入频繁区	1.30	1.19	1.09	1.03
$I>50\%$	跨县流入强势区	4.33	4.83	4.93	5.14
$0< K\leq 10\%$	县内流动稳定区	39.38	23.18	17.70	14.69
$10\%< K\leq 30\%$	县内流动活跃区	54.32	69.50	72.62	72.76
$30\%< K\leq 50\%$	县内流动频繁区	3.23	4.04	5.71	7.92
$K>50\%$	县内流动强势区	3.07	3.28	3.97	4.63

布特征。(2)本文的初步预测结果表明,未来30年中国县域流动人口的活跃程度将持续提升,无论是跨县流动还是县内流动人口规模都将明显增加。预计2050年将有1824个跨县流动活跃型县域($I>10\%$ 或 $I<-10\%$),1596个县内流动活跃型县域($K>10\%$)。(3)流动人口分布的“马太效应”将会加剧未来中国县域流动人口分布的不均衡性,凸显县域人口收缩与集聚的极化现象。一方面,跨县流出活跃型县域的数量及流出人口规模将持续增加,加剧流出极的县域人口收缩;另一方面,跨县流入活跃型县域数量减少但流入强势区数量增加,这意味着将会有更多的流动人口流入更少的县域,新产生的人口流入强势区作为“流动人口增长极”将会吸收大量流动人口,加剧流入极的县域人口集聚现象。

基于此,本文的政策启示主要有以下几点。(1)在中国已经步入且可能在一定时期内保持人口负增长的新时期,单纯依靠人口自然增长维持“人口红利”并不现实,除了人口高质量发展产生的人口素质红利,人口迁移流动表现出的强烈脉动或将为人口负增长时期激活经济发展潜力提供回旋空间(刘厚莲、原新,2020),因而需要思考如何发挥“流动人口红利”的潜在作用。县域单元作为经济发展的薄弱环节,人口、劳动力等生产要素相对稀缺,因而应当更加重视流动人口对县域经济社会发展的作用,通过人口跨区域流动、县内人口流动等途径,优化人口资源配置,减小人口空间错配,为进一步释放县域经济发展动能注入活力。(2)《关于推进以县城为重要载体的城镇化建设的意见》首次提出了“人口流失县城”这一概念,在城镇化发展进程中,区域人口增减分化是经济社会发展的客观规律。对此,在推进以县城为重要载体的城镇化建设过程中,应顺应流动人口分布变化规律及趋势,预先评估与监测未来哪些地区可能会成为人口流入密集区或人口流失严重区,前瞻性地制定各类经济社会发展规划,分类引导县城发展方向。(3)对于具有强人口流入趋势的县域,要加快改善本地人居环境,完善公共服务供给,提高县域人口承载能力。全面放开户籍制度约束,加快流动人口市民化进程,营造宽松的人口流动环境,提前做好应对流动人口集聚的准备。在此基础上,主动承接邻近地区的人口与产业转移,增强与周边城市的交通便捷性,打造成为与周边城市通勤便捷、功能互补、产业配套的卫星县城。(4)对于具有强人口流出趋势的县域,应立足自身区位条件与产业基础,通过县城转型发展,促进人口和公共服务资源适度集中,切实提高人民收入和生活水平,解决好民生问题。同时,严格控制城镇建设用地增量,有效盘活存量,打造一些具有特色优势的新产业,激活内生发展潜力,为转型发展提供支撑。

参考文献:

1. 陈卫(2022):《中国人口负增长与老龄化趋势预测》,《社会科学辑刊》,第5期。
2. 段成荣等(2022):《从657万到3.76亿:四论中国人口迁移转变》,《人口研究》,第6期。

3. 段成荣等(2022):《迁徙中国视野下我国县域人口流动状况与发展挑战》,《西北人口》,第6期。
4. 简新华、黄锬(2010):《中国城镇化水平和速度的实证分析与前景预测》,《经济研究》,第3期。
5. 劳昕、沈体雁(2015):《中国地级以上城市人口流动空间模式变化——基于2000和2010年人口普查数据的分析》,《中国人口科学》,第1期。
6. 李鹏飞、陆铭(2022):《大国空间治理的经济学分析》,《经济科学》,第6期。
7. 刘厚莲、原新(2020):《人口负增长时代还能实现经济持续增长吗?》,《人口研究》,第4期。
8. 刘厚莲、张刚(2023):《我国人口负增长态势:机遇、挑战与应对》,《行政管理改革》,第2期。
9. 刘涛、卓云霞(2022):《中国县级人口变动的空间格局及影响因素——基于第七次全国人口普查数据的新探索》,《人口研究》,第6期。
10. 戚伟、刘盛和(2015):《中国城市流动人口位序规模分布研究》,《地理研究》,第10期。
11. 戚伟等(2017):《1982~2010年中国县市尺度流动人口核算及地域类型演化》,《地理学报》,第12期。
12. 王桂新(2022):《中国省际人口迁移变化特征——基于第七次全国人口普查数据的分析》,《中国人口科学》,第3期。
13. 曾永明(2016):《中国人口空间分布形态模拟与预测——基于“五普”和“六普”的分县尺度人口密度研究》,《人口与经济》,第6期。
14. 曾永明等(2023):《网络视角下中国跨省人口流动格局跃迁及驱动机制:1991~2020年》,《中国人口·资源与环境》,第3期。
15. 张建武等(2021):《中国人口、经济、产业重心空间分布演变轨迹——基于1978~2019年省级数据的分析》,《中国人口科学》,第1期。
16. 张伟丽等(2021):《中国城市人口流动格局演变及影响因素分析》,《中国人口科学》,第2期。
17. Alavi O., Mohammadi K., Mostafaeipour A.(2016), Evaluating the Suitability of Wind Speed Probability Distribution Models: A Case of Study of East and Southeast Parts of Iran. *Energy Conversion and Management*. 119(7):101-108.
18. Asadoorian M.O.(2008), Simulating the Spatial Distribution of Population and Emissions to 2100. *Environmental and Resource Economics*. 39(4):199-221.
19. Bittle J.(2023), The Great Displacement: Climate Change and the Next American Migration. *Simon and Schuster Press*.
20. Cinelli C., Ferwerda J., Hazlett C.(2020), Sensitivity Analysis Tools for OLS in R and Stata. SSRN Working Papers No.3588978.
21. Flowerdew R., Lovett A.(1988), Fitting Constrained Poisson Regression Models to Interurban Migration Flows. *Geographical Analysis*. 20(4):297-307.
22. Ho J., Bogue D.J.(1970), Principles of Demography. *Population (French Edition)*. 25(3):412-412.
23. Jones H.R.(1990), Population Geography. New York, NY: Guilford Press.
24. Krugman P.(1991), Increasing Returns and Economic Geography. *Journal of Political Economy*. 99(3):483-499.
25. Livi-Bacci M.(2017), A Concise History of World Population. Hoboken, NJ: John Wiley & Sons Press.
26. Paplexiou S.M., Koutsoyiannis D.(2016), A Global Survey on the Seasonal Variation of the Marginal Distribution of Daily Precipitation. *Advances in Water Resources*. 94(8):131-145.
27. Rhee I., Shin M., Hong S., et al.(2011), On the Levy-walk Nature of Human Mobility. *IEEE/ACM Transactions on Networking (TON)*. 19(3):630-644.

28. You Z., Yang H., Fu M. (2018), Settlement Intention Characteristics and Determinants in Floating Populations in Chinese Border Cities. *Sustainable Cities and Society*. 39(5):476–486.
29. Zhao M., Liu S., Qi W. (2017), Exploring the Differential Impacts of Urban Transit System on the Spatial Distribution of Local and Floating Population in Beijing. *Journal of Geographical Sciences*. 27(6):731–751.
30. Zipf G.K. (1946), The P1P2/D Hypothesis: On the Intercity Movement of Persons. *American Sociological Review*. 11(6):677–686.

The Pattern and Dynamics of County-Level Population Flow in China: Simulation and Projection Using the Cauchy Distribution

Zeng Yongming Zhong Zikang Liu Houlian

Abstract: In the context of diverging regional population growth, it is vital to investigate whether the population flow follows some endogenous statistical distribution. The study analyzes the pattern and dynamics of the county-level population flow in China using data from the latest three population censuses. It finds that China's county-level population flow is regionally unbalanced, with features that meet the Cauchy distribution quite well. We model the parameters and use them to simulate and project the county-level population flow from 2000 to 2050. We find that, in the next 30 years, the county-level population flow is expected to be increasingly active. There will be 1824 counties active in inter-county population flow and 1596 counties active in intra-county population flow in 2050. By then, the unevenness of the county-level population flow will intensify, and regional population will further diverge. Over the study period, the number of counties with active inter-county outflow increases, so does the scale of outflow population, which aggravates the county population contraction. On the other hand, the number of counties with active inter-county inflow decreases, but the population inflows in the major hosting regions increase. Therefore, more and more people flow into fewer hosting counties, which will aggravate the population agglomeration of those counties. The article provides policy insights for assessing and monitoring the future trends of population flow in China, formulating prospective economic and social development plans, and guiding population flows in an orderly and efficient manner.

Keywords: Floating Population Distribution; Cauchy Distribution; Non-Equilibrium; Simulation and Prediction

(责任编辑:李玉柱)