

# 城市轨道交通通勤与职住平衡状况的关系研究 ——基于大数据方法的北京实证分析

申犁帆<sup>1</sup>, 张纯<sup>2\*</sup>, 李赫<sup>3</sup>, 王焱<sup>4</sup>, 王子甲<sup>5</sup>

(1. 武汉大学城市设计学院, 武汉 430072; 2. 北京交通大学建筑与艺术学院, 北京 100044;  
3. 中国银行国际金融研究所, 北京 100818; 4. 广州市城市规划勘测设计研究院, 广州 510030;  
5. 北京交通大学土木建筑工程学院, 北京 100044)

**摘要:**城市轨道交通网络的发展在提高居民通勤效率的同时也对其职住平衡状况产生了一定影响。论文以北京市206个轨道站点为例, 基于高斯混合模型(Gaussian mixture model, GMM)和一卡通刷卡数据将轨道站点按职住功能进行分类, 利用腾讯“宜出行”定位数据考察轨道站点周边的动态人口分布并计算就业居住比。研究发现: ①中心城区的职住状况明显优于中心城区以外区域; ②轨道交通线网末端区域的职住平衡程度较差, 仅有少数成规模的高端服务产业集中分布的轨道站点周边形成了区域性就业中心; ③部分就业-居住较为均衡的城郊地区仍存在一定的职住不匹配现象。随后, 通过计算一卡通进出站比和“宜出行”职住比得到进出站均衡度和职住平衡度, 利用广义自回归条件异方差(generalized autoregressive conditional heteroskedasticity, GARCH)模型对轨道交通通勤和职住平衡程度进行相关性分析, 研究结果表明: ①进出站均衡度与职住平衡度具有非常显著的正向关系, 即站点进出站人数越接近, 站点周边区域的职住状况越好; ②典型就业地站点与站点周边区域的职住平衡程度显著正相关, 而典型居住地站点与站点周边区域的职住状况存在显著的负相关性。这表明, 人口稠密的聚居区无法带动同样数量就业岗位的产生, 而完善的就业中心能够吸引一定数量的人口在附近居住; ③轨道站点的区位条件与职住平衡状况存在一定正向关系; ④GMM能够对属性复杂模糊的轨道站点进行有效的聚类分析; ⑤具有实时性强、精确度高、覆盖度广、获取难度低等优点的“宜出行”数据能够在微观空间尺度下弥补其他捕捉和分析实时人口时空分布特征方法的局限性。

**关键词:**城市轨道交通; 通勤行为; 职住平衡; 大数据; 高斯混合模型; GARCH模型; 北京

近些年, 中国城市轨道交通建设进入到快速发展的阶段。在有轨道交通运营的城市, 越来越多的就业者选择利用轨道交通进行通勤。伴随着轨道交通带来的通勤方式的变化和通勤效率的提高, 城市居民的职住状况也在发生改变。目前, 相关研究主要关注城市轨道交通对站点周边土地价值的影响以及建成环境、传统出行方式、居民社会经济属

性与职住平衡的关联性, 而对于轨道交通通勤和居民职住状况之间关系的研究较少。在研究策略方面, 基于传统问卷调查方法获得的人口数据、出行数据和社会经济数据已经越来越难以准确高效地支撑居民职住关系方面的研究。尽管大数据的出现为研究城市人口活动的时空分布提供了全新的视角和海量的数据样本, 但对于微观空间层面上居

收稿日期: 2018-07-10; 修订日期: 2018-12-10。

基金项目: 国家自然科学基金项目(51678029, 51778039); 中国城市轨道交通协会专项研究项目(A17M00080)。[Foundation: National Natural Science Foundation of China, No. 51678029 and 51778039; Specific Research Project of China Urban Rail Transit Association, No. A17M00080.]

第一作者简介: 申犁帆(1987—), 男, 北京人, 博士, 研究方向为城市轨道交通与空间规划、城市居民职住关系、城市可持续发展。E-mail: 495062785@qq.com

\*通信作者简介: 张纯(1983—), 女, 北京人, 副教授, 博士生导师, 研究方向为转型期城市规划、城市形态与可达性、社区演变与规划。E-mail: zhangc@bjtu.edu.cn

引用格式: 申犁帆, 张纯, 李赫, 等. 2019. 城市轨道交通通勤与职住平衡状况的关系研究: 基于大数据方法的北京实证分析 [J]. 地理科学进展, 38(6): 791-806. [Shen L F, Zhang C, Li H, et al. 2019. Relationship between urban rail transit commuting and jobs-housing balance: An empirical analysis from Beijing based on big data methods. Progress in Geography, 38(6): 791-806.] DOI: 10.18306/dlkxjz.2019.06.001

民职住状况的研究,现有的一些主流大数据方法皆存在一定的局限性。

随着居民日常出行活动日益依赖于城市轨道交通,一些学者开始关注轨道交通对于城市人口分布和日常活动的影响。尽管城市轨道交通的发展能够提升轨道沿线的土地价值,但高效的轨道交通在降低人们的时间成本的同时也提高了人们的经济成本,轨道交通带动了站点周边住宅价格和租金的上涨,进而影响到不同偏好居民的居住选择(Ho et al, 2002)。例如,Cervero等(2006)在对美国旧金山湾区的研究中发现,轨道交通通勤行为与以公共交通为导向的紧凑混合型社区居住选择之间有着内在联系。在自选择(self selection)因素的作用下,一些居民会搬到轨道站点附近居住并放弃私家车的通勤方式。在中心城区的轨道交通沿线,级差地租下土地利用类型被改变,老旧破败的居住区逐渐被高端的住宅、商业和服务业所取代,城市中心区的居住人口逐渐减少(陆化普等, 2013)。与此同时,就业中心的集聚会导致其就业吸引范围持续扩大,从而造成就业者的通勤距离和通勤时间不断增加(孟斌等, 2017)。在城市郊区的轨道交通沿线,由于便捷的交通和相对低廉的生活成本,轨道交通沿线的末端地区会呈现出人口集聚和产业集中的双核空间集聚发展趋势,形成沿轨道交通网络分布的放射状点轴格局,使该区域的居住人口不断增加。因此,轨道交通的发展会促使城市中心区的人口向城市郊区迁移(曾智超等, 2005; 孟庆艳等, 2008; 赵晖等, 2011)。尽管轨道交通的发展扩大了人们的居住空间范围,为人们提供了更多的居住地选择,然而城市中心区更多的就业机会、更高的薪酬以及更加丰富完善的生活服务设施,使得人们除居住外的大部分日常活动(如工作、购物、娱乐、就医等)仍然保留在城市中心区内。因此城市轨道交通的发展在一定程度上加剧了城市内部区域性的职住不平衡(赵晖等, 2011; 孔令斌, 2013)。此外,申犁帆、张纯等(2018)通过对北京近郊区的实证研究发现,在就业可达性和用地混合度差异下,城市轨道交通通勤与职住平衡状况之间的关系会产生一定变化。城市轨道交通网络主要是通过居民的通勤出行活动来对其职住状况产生影响。但是,目前的研究还没有涉及到从整个城市的空间尺度研究城市轨道交通的通勤行为和居民职住关系之间的关联性。

对职住平衡程度的测算分为平衡度测算和自

足性测算(Cervero, 1989; Ewing et al, 2004; 孟晓晨等, 2009)。尽管计算自足性比平衡度得到的职住平衡程度更加准确,但由于自足性的测算主要是依靠传统调查问卷收集居民的就业地和居住地信息,存在数据采集周期长、数据采集范围小、样本量有限、数据准确度不稳定等缺陷,因此大多数研究仍利用平衡度的方法(职住比)来测算职住平衡(刘志林等, 2011; 张纯等, 2016; 程鹏等, 2017)。目前,主流方法是以经济普查数据中的就业岗位数计算职住比。但由于经济普查数据为每5 a统计一次(最新的第三次全国经济普查数据统计于2013年),在数据时效性方面具有较大的局限性。另外,尽管官方经济普查数据提供了街道层面的就业岗位信息,但该数据仅统计注册在某街道的企业员工总数,这与该街道内实际的就业岗位数可能存在偏差。除此之外,本文是针对轨道站点周边区域的职住关系研究,而站点的影响范围无法与街道的行政范围有较好的重叠,因此按街道统计的数据不能有效地反映站点周边区域的实际职住状况。

近年来,随着大数据技术的日趋成熟,全球定位(GPS)数据、自动售检票系统(AFC)数据、手机信令等移动定位数据采集方法开始广泛应用于城市就业-居住时空关系的相关研究中。在GPS数据方面,一些研究者利用出租车GPS、兴趣点、出行调查等多源数据分析居民出行特征,并通过聚类算法识别居住、就业、娱乐等功能区(Liu et al, 2009; 陈泽东等, 2018);申悦等(2013)通过GPS数据对北京市郊区大型社区居民的就业等日常活动空间特征进行了分析。此外,还有研究者利用互联网地图上的人口热力图反映出的特定时段活动人口强度均值比作为职住关系指数,以此测算样本地区的职住平衡状况(冷炳荣等, 2015; 谭欣等, 2016)。在AFC数据方面,一些学者利用公交智能卡刷卡数据对北京市域内的居住、就业等主要功能区进行了识别(龙瀛等, 2012; 韩昊英等, 2016)。还有研究者利用轨道交通一卡通起止点数据识别了基于轨道交通出行的上海市典型就业地和居住地,并对居民的轨道交通通勤特征进行了分析(许志榕, 2016; 许园园等, 2017)。Medina等(2013)结合公交刷卡数据、居民出行调查数据和建筑信息数据对新加坡的主要就业地空间分布进行了识别和分析。马亮(2017)通过比较不同年份的轨道交通刷卡数据,分析了典型就业地站点通勤圈范围的变化特征,以此评价典型就业



地的职住平衡状况。在手机信令数据方面,许宁等(2014)利用大规模短期规则采样的个人通信移动终端定位数据,对居民的就业地和居住地进行了识别。张天然(2016)基于手机信令数据分别考察了上海中心城区和郊区新城的职住状况,并对居民的通勤行为特征进行了分析。丁亮等(2016)利用手机信令数据识别了上海的就业地和居住地,并对中心城区的就业中心进行了测度。

尽管如此,上述常见的大数据方法也存在一些缺陷。GPS所获得的数据为非抽样数据,如基于定位设备的GPS数据只能严格受制于GPS设备的载体(如出租车等),而基于百度热力图的GPS数据来源于用户手机端的百度产品,这对互联网服务商的用户覆盖度提出了较高的要求(2018年初手机app月活跃用户数排名中百度仅为第10位)。同样,AFC刷卡数据也属于非抽样数据,依赖于出行者所选择的交通方式(地面公交、轨道交通),并且其非连续性、非实时性的数据属性无法对进站或上车之前以及出站或下车之后的用户活动进行空间定位。尽管手机信令数据被研究者们广泛认为能够较好地支持海量居民活动时空信息分析(柴彦威等,2010;刘瑜等,2011),且具有覆盖人群广、采样周期长、实时性和连续性强等优势,但是,手机信令数据可能会因通信基站分布密度较小且不均(市区覆盖半径通常为100~500 m)(李祖芬等,2016)以及通信基站超负荷或信号不稳定等问题造成手机定位偏移,最大误差甚至能达到800 m(丁亮等,2016)。因此,手机信令数据无法准确有效地分析微观空间尺度下的人口分布及活动状况。

因此,本文以北京市为例,利用工作日通勤时段的轨道交通一卡通刷卡数据代入高斯混合模型(Gaussian mixture model, GMM),对轨道站点进行职住功能的聚类分析,同时整合早高峰时段出站和进站的累计人数、出进站比等轨道交通通勤行为数据。然后,利用同一时段腾讯“宜出行”程序的人口定位数据计算站点周边区域的职住比。通过调整一卡通进出站比和“宜出行”职住比的平衡偏离值得到进出站均衡度和职住平衡度。最后,构建广义自回归条件异方差(generalized autoregressive conditional heteroskedasticity, GARCH)模型对相关变量进行回归分析。通过对轨道交通通勤与职住状况之间关系的阐述和分析,希望能够为完善城市空间功能结构、优化轨道交通线网和站点布局、协调居民职住关系提供实证案例参考和数据分析支持。

## 1 数据收集及处理

### 1.1 数据来源和说明

本文所使用的数据主要包括:① 2015年9月份连续10个工作日的北京市轨道交通站点刷卡数据。其中,为使数据更有针对性地反映本文的研究内容,剔除了单程票、员工卡等刷卡数据,只统计一卡通充值卡的刷卡数据,数据为每30 min统计的刷卡次数;② 通过腾讯“宜出行”程序抓取的与站点刷卡数据时间相对应的实时定位数据,数据抓取时段间隔为30 min;③ 利用大数据挖掘技术获取的轨道站点POI及其属性数据;④ 从北京地铁官方网站上获取的轨道站点信息(出入口数量和站间距)数据。

为了能够更准确地提取城市轨道交通的通勤数据,本文统计了工作日早高峰时段(7:30~9:30)和晚高峰时段(17:00~19:00)轨道站点的一卡通进出站刷卡数据。其中,考虑到西二旗、上地、五道口等3个站点周边的互联网相关企业较多,基于互联网企业惯常的工作时间,我们将上述3个站点的早高峰时段和晚高峰时段分别顺延30 min至8:00~10:00和17:30~19:30。刷卡数据采集时间为2015年9月14—18日和9月21—25日。

### 1.2 轨道站点合理步行范围的划定

不少已有研究的问卷调查数据显示,大多数轨道交通通勤者通过步行方式到达轨道站点乘车,且出行者步行接驳轨道站点的平均时间为10 min左右(岳芳等,2007; Zielstra et al, 2011)。也就是说,受心理因素(情绪)和物理因素(体力)的限制,通勤者采用步行方式接驳轨道站点的平均时耗约为10 min。由此,本文基于已有研究(申犁帆,王烨等,2018),将从轨道站点出入口步行10 min所能够到达的全部区域看作该站点的合理步行范围(图1)。

### 1.3 相关兴趣点的抓取

高德地图平台提供了千万级别的兴趣点(Point of Interest, POI),并对这些兴趣点进行了详细分类。同时,平台提供了应用程序的编程接口(Application Programming Interface, API)以便调用数据。我们通过自编写的Python程序,批量抓取了研究范围内的大型公园、三甲医院、旅游景点等重要兴趣点。

### 1.4 “宜出行”数据的采集

“宜出行”是腾讯公司开发的基于地图显示当前选定区域人流分布的手机端小程序,能够有效分析既定区域内的人流热度和分布变化。相比传统的静态人口普查数据、出行调查数据和其他动态的



图1 轨道站点的合理步行范围示意

Fig.1 Reasonable pedestrian catchment areas of urban rail transit (URT) stations

人口定位数据,“宜出行”位置数据具有强实时性、高精度性、广覆盖性、易获得性等优点。

如图2所示,由于“宜出行”程序地图上显示的热力图不能直接转化为实际人口数,需要利用互联网爬虫技术获取原始数据。“宜出行”的原始数据为25 m×25 m网格的点数据,每个点都有一个基于定

位数据赋予的值来表示相对人口数(图3)。为了隐藏原始数据,地图供应商会对每个时刻的人口数据随机乘以一个系数后再对外发布。经过长时间的数据比对,笔者发现地图公开数据的最大公约数就是该系数。因此,将公开数据除以其最大公约数便可得到原始人口数据值。具体操作过程为:①通过

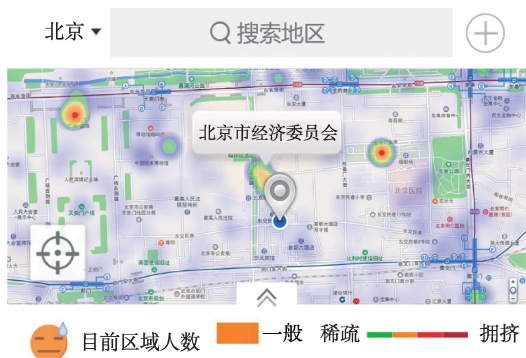


图2 “宜出行”热力图展示

Fig.2 Thermodynamic diagram by "Yichuxing"

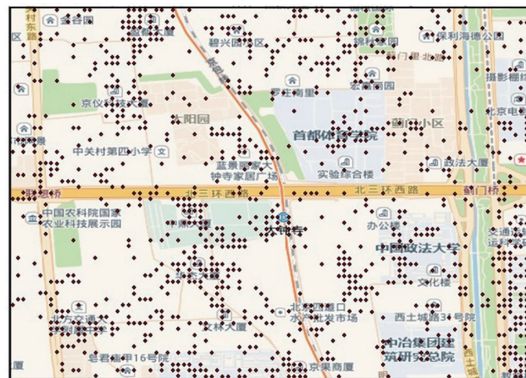


图3 “宜出行”点数据展示

Fig.3 Point data of "Yichuxing"



Python编程,抓取10个工作日白天以及前一天夜间的“宜出行”数据作为本研究的原始数据;②利用ArcGIS对在地图上表现为非均匀分布的原始点数据进行空间化表达;③将采集到的“宜出行”原始数据值乘以相对系数来进行标准化赋值处理(申犁帆,王烨等,2018)。

虽然“宜出行”无法采集到不使用智能手机人群的数据,考虑到本文主要针对北京市辖区范围内的通勤人群,该群体的智能手机使用率高,因此具有良好的数据采集覆盖度。另外,本文中的“宜出行”数据仅用于比较不同轨道站点周边区域之间的白天人口数(就业岗位数)和夜间人口数(居住人口数)的比值,故该数据在可靠性、适用性等方面都具有较高的利用价值。

## 2 研究设计与分析

### 2.1 研究范围

本文的研究范围为截至2015年9月北京市范围内所有已开通运营的城市轨道交通(包括地铁和轻轨)站点。由于数据源的不完整性,本文中的样本站点不包括由京港地铁运营的4号线、14号线和16号线沿线站点。为了保证数据分析的准确性,同时剔除了与上述3条线路换乘的所有站点。此外,本文还舍去了机场线沿线、北京西站、天安门东、天安门西等在数据采集时段可能包含大量非通勤客流的站点,筛选后的样本站点总数为206个(表1和图4)。所有样本站点及其合理步行范围分布于北京11个市辖区的129个乡镇、街道和地区。

表1 研究范围内的轨道站点  
Tab.1 URT stations in the study area

所属区		站点名称						
海淀区	健德门	知春路	巴沟	车道沟	公主坟	白堆子	西小口	军事博物馆
	牡丹园	知春里	火器营	慈寿寺	莲花桥	五棵松	育新	海淀五路居
	西土城	苏州街	长春桥	西二旗	花园桥	玉泉路	上地	西钓鱼台
	五道口	大钟寺	北沙滩	万寿路	六道口	永泰庄	白石桥南	清华东路西口
西城区	复兴门	积水潭	木樨地	长椿街	阜成门	湾子	鼓楼大街	广安门内
	车公庄	北海北	和平门	虎坊桥	达官营	车公庄西	南礼士路	什刹海
东城区	建国门	王府井	安定门	雍和宫	东直门	前门	天坛东门	东四十条
	北京站	珠市口	北新桥	灯市口	崇文门	东单	广渠门内	张自忠路
	朝阳门	磁器口	桥湾	东四	和平里北街	南锣鼓巷		
朝阳区	四惠东	褡裢坡	光熙门	永安里	四惠	黄渠	北苑路北	传媒大学
	望京西	立水桥	常营	青年路	国贸	管庄	立水桥南	金台夕照
	呼家楼	东大桥	焦化厂	安贞门	双井	化工	南楼梓庄	和平西桥
	大郊亭	高碑店	林翠桥	北土城	肖村	双合	奥体中心	广渠门外
	芍药居	太阳宫	三元桥	亮马桥	草房	关庄	百子湾	惠新西街北口
	劲松	潘家园	成寿寺	团结湖	北苑	双桥	安华桥	惠新西街南口
	马泉营	崔各庄	八里桥	安立路	孙河	柳芳	小红门	奥林匹克公园
	十里堡	农业展览馆	欢乐谷景区	大屯路东				
石景山	八宝山	古城路	苹果园	八角游乐园				
丰台区	刘家窑	六里桥	科怡路	分钟寺	宋家庄	郭公庄	大葆台	六里桥东
	石榴庄	大红门	角门东	草桥	纪家庙	首经贸	丰台站	丰台南路
	泥洼	丰台科技园	丰台东大街					
大兴区	旧宫	亦庄桥	万源街	荣昌东街	荣京东街	亦庄文化园		
昌平区	天通苑	育知路	平西府	霍营	龙泽	天通苑南	生命科学园	沙河高教园
	回龙观	巩华城	朱辛庄	沙河	南邵	天通苑北	回龙观东大街	
房山区	篱笆房	广阳城	稻田	长阳	苏庄	良乡南关	良乡大学城	良乡大学城北
	良乡大学城西							
通州区	东夏园	郝家府	经海路	潞城	果园	北运河西	通州北关	物资学院路
	次渠南	九棵树	临河里	次渠	梨园	同济南路	通州北苑	土桥
顺义区	南法信	后沙峪	花梨坎	顺义	石门	俸伯	国展	

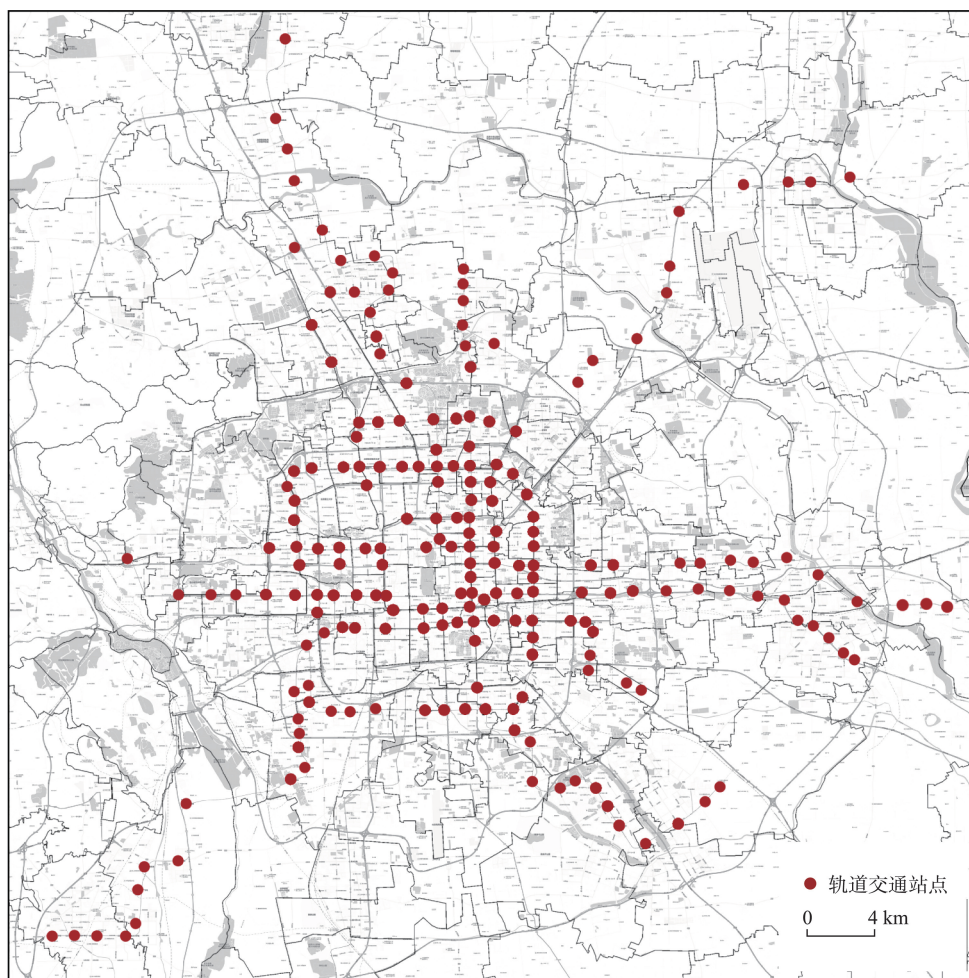


图4 样本轨道站点的空间分布

Fig.4 Spatial distribution of URT stations

## 2.2 变量选取

如表2所示,为了将轨道站点按职住功能进行合理分类,进而考察轨道交通通勤行为与轨道站点周边职住状况之间的关联性,本文选取了涉及轨道站点通勤特征(eiratio、jobs、housing、egrass、ingrass)、属性特征(initial、transfer、exit)和区位特征(location、distance、facility)等若干因素代入高斯混合模型(代入部分通勤特征变量)和广义自回归条件异方差模型(代入属性特征、区位特征变量以及部分通勤特征变量)进行分析。

## 2.3 基于高斯混合模型的站点职住功能分类

目前,对于轨道站点的分类主要基于以节点和场所为导向的定性分类方法(金磊等, 2010; 邵滢宇等, 2015),分类结果通常只能按照大类对站点进行模糊分类,不能有效反映具有混合功能的站点属性特征,而少数定量分类方法(如K-均值聚类法)也存在较大的局限性(李向楠, 2015; 罗芳等, 2015)。与

传统定量分类的迭代算法不同,GMM是基于后验概率对轨道站点进行聚类分析,特别是对识别混合类型的站点具有较好的效果(岳真宏等, 2017)。因此,本文选择使用GMM对样本站点进行职住聚类分析。

首先,为考察城市轨道交通站点周边区域的职住属性,根据已有研究(岳真宏等, 2017),通过计算轨道站点的刷卡数据得到aeratio、airatio、peratio、piratio等4类质心附近的站点初始变量指标(表3),然后将该指标代入高斯混合模型。在此,假设每个站点分别由2个单高斯分布的混合模型表示,即轨道站点由高斯混合模型生成的概率密度函数为:

$$P(x) = \sum_{k=1}^2 p(k) p(x|k) = \sum_{k=1}^2 \varphi_k \eta(x|\mu_k, \sigma_k) \quad (1)$$

式中: $x$ 为维度为 $d$ 的向量; $k$ 为单高斯函数的数量; $\varphi_k$ 为第 $k$ 个单高斯分布被选中的概率; $\mu_k$ 、 $\sigma_k$ 为第 $k$ 个单高斯分布; $\eta(x|\mu_k, \sigma_k)$ 为第 $k$ 个单高斯分布的



表2 变量定义

Tab.2 Definition of variables

变量	定 义
housing	轨道站点是否为典型居住地站点。用虚拟变量表示,其中:1表示是典型居住地站点,0表示不是典型居住地站点。用式(1)计算得到
jobs	轨道站点是否为典型就业地站点。用虚拟变量表示,其中:1表示是典型就业地站点,0表示不是典型就业地站点。用式(1)计算得到
jhratio	“宜出行”职住比。用站点步行可达范围内工作日的夜间“宜出行”点数据和日间“宜出行”点数据之比的均值表示
eiratio	轨道站点进出站比。用站点工作日早高峰时段出站刷卡数和进站刷卡数之比的均值表示
jhbal	职住平衡度。用“宜出行”职住比减1后得到
eibal	一卡通进出站均衡度。用轨道站点进出站比减1后得到
aeratio	轨道站点早高峰进站比。用站点工作日早高峰时段进站刷卡数和全天进站刷卡数之比的均值表示
airatio	轨道站点早高峰出站比。用站点工作日早高峰时段出站刷卡数和全天出站刷卡数之比的均值表示
peratio	轨道站点晚高峰进站比。用站点工作日晚高峰时段进站刷卡数和全天进站刷卡数之比的均值表示
piratio	轨道站点晚高峰出站比。用站点工作日晚高峰时段出站刷卡数和全天出站刷卡数之比的均值表示
egrass	轨道站点出站数。用站点工作日早高峰时段出站累计刷卡数表示
ingrass	轨道站点进站数。用站点工作日早高峰时段进站累计刷卡数表示
initial	轨道站点是否为始发站。用虚拟变量表示,其中:1表示是始发站,0表示不是始发站
transfer	轨道站点是否为换乘站。用虚拟变量表示,其中:1表示是换乘站,0表示不是换乘站。当轨道站点同属于换乘站和始发站时(如西二旗站),则将该站归为换乘站
exit	轨道站点的出入口数(个)。经修正后统计得到
location	轨道站点的区位条件。用站点距离天安门的直线距离(km)表示
distance	轨道站点的站间距。轨道站点与相邻所有站点的平均站间距离(km)
facility	轨道站点周边是否有主要公共服务设施(包括三甲医院、大型公园、著名景点等)。用虚拟变量表示,其中,1表示有,0表示没有

表3 质心附近典型站点初始变量指标

Tab.3 Initial variable values of typical stations near the center of mass

站点名称	概率	早高峰 进站比	早高峰 出站比	晚高峰 进站比	晚高峰 出站比
天通苑	1	0.50	0.04	0.05	0.31
回龙观	1	0.58	0.05	0.05	0.33
北苑	1	0.78	0.13	0.13	0.70
永安里	1	0.06	0.56	0.48	0.09
朝阳门	1	0.06	0.53	0.43	0.10
复兴门	1	0.07	0.61	0.49	0.08
四惠	1	0.25	0.38	0.31	0.20
花园桥	1	0.21	0.40	0.36	0.19
和平西桥	1	0.19	0.44	0.37	0.18

概率密度函数,具体表达式如下:

$$\eta(x|\mu_k,\sigma_k)=\frac{1}{\sqrt{2\varphi|\sigma|}}e^{-\frac{1}{2}(x-\mu)^T\sigma^{-1}(x-\mu)} \quad (2)$$

利用高斯混合模型进行聚类分析的一般步骤为:

首先,估算数据由每一个单高斯分布所生成的概率。对于每一个数据 $x_i$ 而言,它由第 $k$ 个单高斯分布生成的概率为:

$$\gamma(i,k)=\frac{\varphi_k\eta(x_i|\mu_k,\sigma_k)}{\sum_{j=1}^2\varphi_j\eta(x_i|\mu_j,\sigma_j)} \quad (3)$$

然后,估算每一个单高斯分布的参数。在此,利用最大似然估计(Maximum Likelihood Estimate, MLE)得到对应的参数值:

$$\mu_k=\frac{1}{N_k}\sum_{i=1}^N\gamma(i,k)x_i \quad (4)$$

$$\sigma_k=\frac{1}{N_k}\sum_{i=1}^N\gamma(i,k)(x_i-\mu_k)(x_i-\mu_k)^T \quad (5)$$

式中:  $N_k=\sum_{i=1}^N\gamma(i,k)$ ,同时  $\varphi_k=\frac{N_k}{N}$ 。

最后,重复迭代上面2步,直到似然函数值收敛为止。

将高斯混合模型筛选后的206个样本站点按照轨道交通通勤特征分为3类:典型就业地站点、典型居住地站点、职住混合地站点。其中,典型就业地站点和典型居住地站点是由一个单高斯分布生成,但具体某个站点属于哪个单高斯分布未知。职住混合地站点由GMM计算出的后验概率得到,当某站点从属于上述2种功能类型的概率差距相对不大时,则视该站点属于职住混合地站点。在此,利用

双机制阈值模型(Two-regime Threshold Model)来确定后验概率差距的阈值(Hansen, 2000)。具体表达式如下:

$$r_t = \left( c_{10} + \sum_{k=1}^p \theta_k r_{t-k} \right) I(r_{t-d^*} \leq \gamma) + \left( c_{20} + \sum_{k=1}^p \theta_k r_{t-k} \right) I(r_{t-d^*} > \gamma) + \mu_t \tag{6}$$

式中: $r_t$ 为轨道站点的通勤特征数据; $c_{10}$ 、 $c_{20}$ 为常数; $\theta_k$ 为系数; $I(\cdot)$ 表示一个指数函数; $\gamma$ 为需要求证的阈值。当指数函数的值取1时,括号内的条件会被满足;如果指数函数的值取0时,则条件不满足。

如表4所示,通过确定GMM功能聚类后验概率的阈值便可得到轨道站点的职住类别属性。

图5展示了利用GMM得到的轨道站点职住功能聚类情况。其中,共识别出典型就业地站点19个,典型居住地站点86个,职住混合地站点101

个。从图5中可以看出,典型就业地站点主要位于长安街沿线和城六区(西城、东城、朝阳、海淀、丰台、石景山)东部的CBD区域,四环外与外延轨道线路相交汇的个别站点(如上地、丰台科技园等)也形

表4 部分轨道站点的GMM聚类后验概率  
Tab.4 Posteriori probability of GMM clustering of some URT stations

站点名称	站点类别	典型就业	典型居住
天通苑	典型居住	0.071	0.929
回龙观	典型居住	0.062	0.938
北苑	典型居住	0.033	0.967
永安里	典型就业	0.940	0.060
朝阳门	典型就业	0.933	0.067
复兴门	典型就业	0.945	0.055
四惠	职住混合	0.423	0.577
花园桥	职住混合	0.394	0.606
和平西桥	职住混合	0.381	0.619

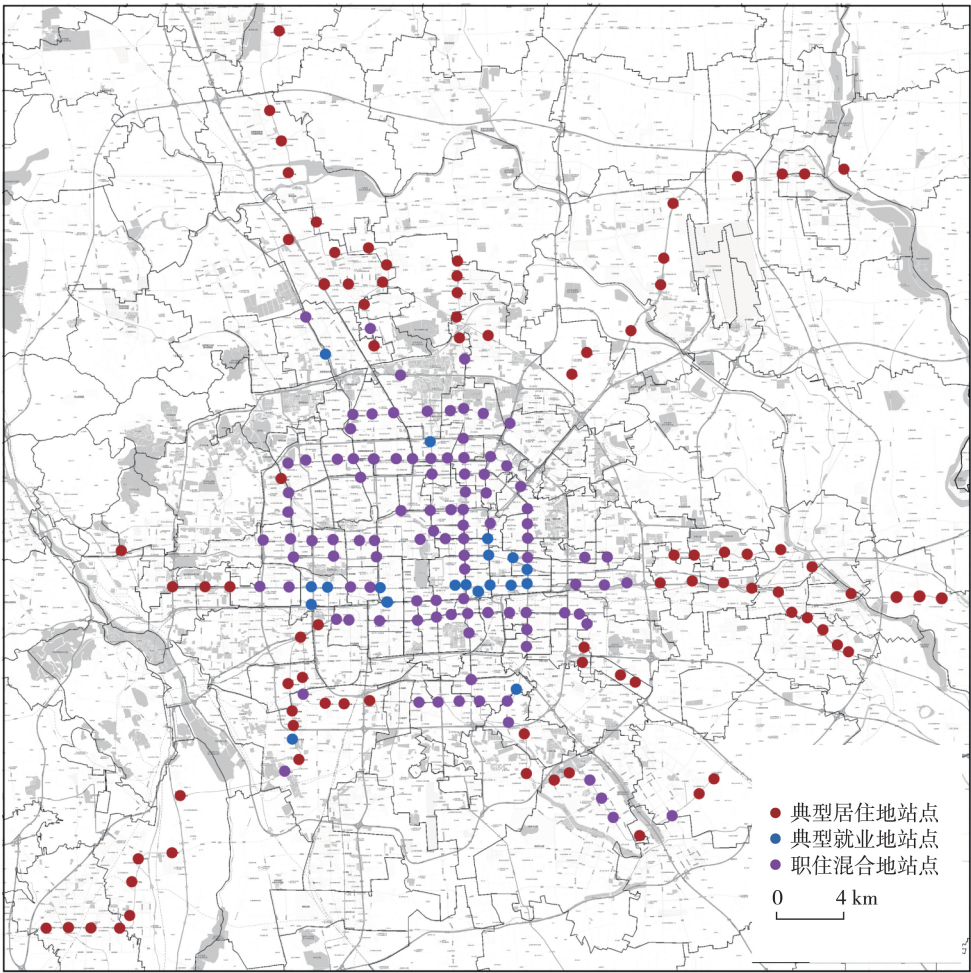


图5 轨道站点的聚类结果展示  
Fig.5 URT stations based on GMM clustering



成了区域性的就业集中地。相比之下,典型居住地站点大多位于四环以外区域,主要集中在城六区的丰台、石景山以及远郊区的昌平、顺义、通州、房山等。在中心城区内,就业中心地站点多位于经济发展成熟的朝阳区和海淀区,居住中心地站点多位于经济发展相对落后的丰台区和石景山区。另外,职住混合地站点主要集中在城六区特别是四环路以内的中心城区,同时在东南部的亦庄以及东部的高碑店地区也有一定分布。

2.4 “宜出行”职住比和一卡通出进站比的计算

由于数据的局限性,我们无法确定所有站点周边区域居住者的就业地以及站点周边区域就业者的居住地。因此,本文使用平衡度(职住比)的方法来测算轨道站点周边区域的职住平衡状况,具体表达式如下:

$$JHR_n = \frac{J_n}{H_n} \tag{7}$$

式中:JHR<sub>n</sub>为轨道站点*n*合理步行范围内的职住比;J<sub>n</sub>为轨道站点*n*合理步行范围内的就业岗位数;H<sub>n</sub>为轨道站点*n*合理步行范围内的常住人口数。

通过连续多个工作日对“宜出行”数据的比对,笔者发现工作日14:00~16:00和21:00~23:00的数据最稳定。因此,为了能够较为准确地测算轨道站点周边的职住状况,本文用工作日日间(14:00~16:00)站点合理步行范围内的“宜出行”定位数据表示该范围内的就业岗位数,用工作日前一天夜间(21:00~23:00)站点合理步行范围内的“宜出行”定位数据表示该范围内的居住人口数,以此来计算轨道站点周边区域的职住比:

$$jhratio_n = \frac{j_n}{h_n} \tag{8}$$

式中:jhratio<sub>n</sub>为通过对“宜出行”定位数据计算得到

的轨道站点*n*周边区域的职住比;j<sub>n</sub>为轨道站点*n*的合理步行范围内就业岗位数相对值;h<sub>n</sub>为轨道站点*n*的合理步行范围内居住人口数相对值。

相应地,为了分析轨道交通通勤行为与站点周边职住平衡状况之间的关系,利用工作日早高峰时段轨道站点的一卡通出站和进站刷卡数之比计算该站点的出进站比:

$$eiratio_n = \frac{egrass_n}{ingrass_n} \tag{9}$$

式中:eiratio<sub>n</sub>为工作日早高峰时段轨道站点*n*的一卡通出进站比;egrass<sub>n</sub>为早高峰时段轨道站点*n*的一卡通累计出站数;ingrass<sub>n</sub>为早高峰时段轨道站点*n*的一卡通累计进站数。

轨道站点周边区域人口空间分布的昼夜对比,在一定程度上能够反映出研究范围内的就业和居住状况(表5、图6)。在人口就业空间分布特征方面,就业岗位多集中在四环路沿线及以内的中心城区。其中,以国贸为中心向外辐射的地铁1、2、10号线东段部分站点所在的中央商务区就业岗位最为稠密。此外,西南部的丰台科技园、西北方向的中关村、五道口以及北部的西二旗也有不少就业岗位集聚并形成了区域性的就业中心。然而,昌平北部、顺义、通州、亦庄、良乡等城郊区域轨道交通沿线的就业岗位不足。由此可知,就业地的空间分布显著不均,第三产业的金融、商务、信息技术、教育科研等领域的空间集聚效应明显。在人口居住空间分布特征方面,中心城区的居住地分布相对比较均衡,而中心城区以外的居住地则呈现出较大的空间分布差异。不少研究发现,城市轨道交通的开发和利用能够带动其沿线土地及物业的增值(谷一桢等,2010;刘康等,2015),而住房价格随空间区位条件的渐弱而递减(张文忠,2001;汤庆园等,2012)。

表5 部分轨道站点的“宜出行”职住比值  
Tab.5 Jobs-housing ratio around some URT stations by "Yichuxing"

站点名称	职住比	站点名称	职住比	站点名称	职住比	站点名称	职住比
四惠东	2.147	安定门	1.120	灯市口	1.486	裕褆坡	0.672
四惠	1.872	雍和宫	2.267	东单	1.877	青年路	1.315
国贸	3.515	东直门	1.529	崇文门	1.719	十里堡	0.788
永安里	2.187	朝阳门	2.305	磁器口	1.117	呼家楼	1.406
建国门	2.417	北京站	1.466	刘家窑	0.956	东大桥	1.599
王府井	2.098	前门	1.461	潞城	1.811	东四	1.160
复兴门	2.916	和平门	0.971	东夏园	1.584	北海北	1.272
木樨地	1.399	长椿街	1.091	郝家府	1.974	车公庄	1.495
万寿路	1.167	阜成门	1.400	草房	0.760	百子湾	1.503
五棵松	1.549	天通苑	0.569	常营	1.799	大郊亭	1.205

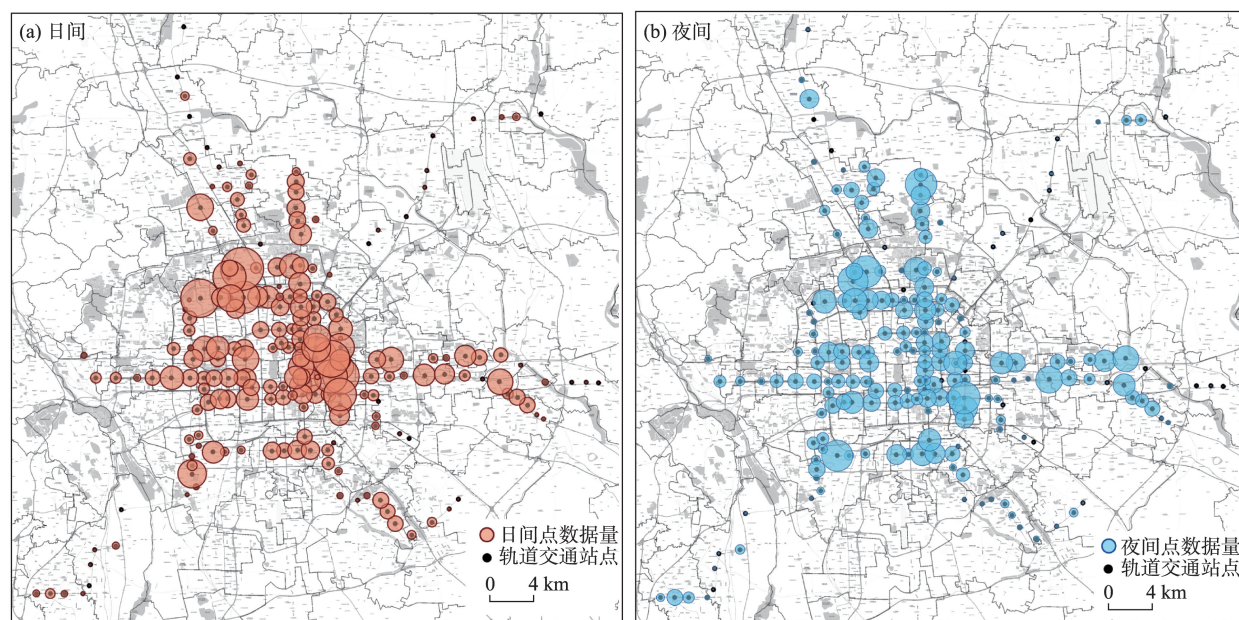


图6 轨道站点合理步行范围内的“宜出行”人口昼夜空间分布  
Fig.6 Spatial distribution of daytime and nighttime population within reasonable pedestrian catchment area of URT stations by "Yichuxing"

因此,基于出行效率和居住成本的双重考虑,大量就业居民选择在既能满足日常通勤需求、居住成本的压力又相对较小的轨道交通线网末端周边区域居住,进而在中心城区以外形成了若干人口聚居区(如沙河、天通苑、物资学院路、通州北苑等)。这些居住片区与在同一条轨道线路上的相邻站点周边居住人口稀少的区域形成了明显对比。

如图7所示,从整体上看,中心城区的职住平衡程度要优于中心城区以外的地区,并且东部要好于西部。从空间功能特征看,规模成片的大型住区所在区域的职住状况较差(如天通苑、回龙观等)。另外,与赵晖等(2011)的研究结果有所不同,笔者发现大多数轨道交通线路末端区域的职住状况相对较差(如5号线、7号线、15号线、昌平线、亦庄线等)。近年来北京快速增长的房价大大增加了居民的居住成本,而城市轨道交通的拓展也会助推轨道沿线住房租售价格的上涨,且居民收入的涨幅远不及居住成本的增速。以2008—2017年为例(图8),链家数据表明北京市二手住房销售均价从12125元上涨到62608元,10 a间增长了约4.2倍,轨道交通覆盖的城市中心区域其增长幅度则更高。然而同一时期北京市城镇居民可支配收入仅增长了2.5倍(北京市统计局,2009—2018)。由此可以看出,大部分就业者很难有能力在与其就业地同样区位条件的地方择居。同样地,传统制造业和低端服务业通常

也难以承受轨道交通沿线地区较高的租金成本,只有规模化的产品附加值和劳动回报率较高的高端服务产业聚集的地区和强力宏观政策支持下的地区能够在中心城区以外的轨道交通沿线形成城市次中心。具体来说,受北京副中心建设与发展的影响,位于通州区的6号线和八通线东段站点周边区域的职住平衡程度相对较好。此外,位于中心城区以外的亦庄、西二旗、生命科学园等区域依托于互联网高科技产业相关发展政策的长期支持,其职住状况比其他城郊地区更好。

通过对比轨道站点职住功能聚类(图5)和站点周边的职住状况(图7),还能够发现尽管位于石景山、通州、昌平、顺义的部分站点周边区域的职住比相对均衡,但仍体现出本地居住外出就业的潮汐式通勤客流特征。也就是说,城郊部分地区看似充分的工作岗位并不能满足本地居民的就业需求,存在一定的职住不匹配现象。在这种情况下,选择外出就业的郊区居民会比中心城区居民承受更高的时间成本(远距离通勤),而选择在居住地周边就业的郊区居民也不得不面临就业选择不足、薪酬较低等问题,工作回报率和满意度方面的差异会加剧城市内部区域间的社会经济发展失衡。

## 2.5 基于GARCH模型的轨道交通通勤与职住平衡的相关性分析

基于数据变量中可能存在的非正态分布属性,



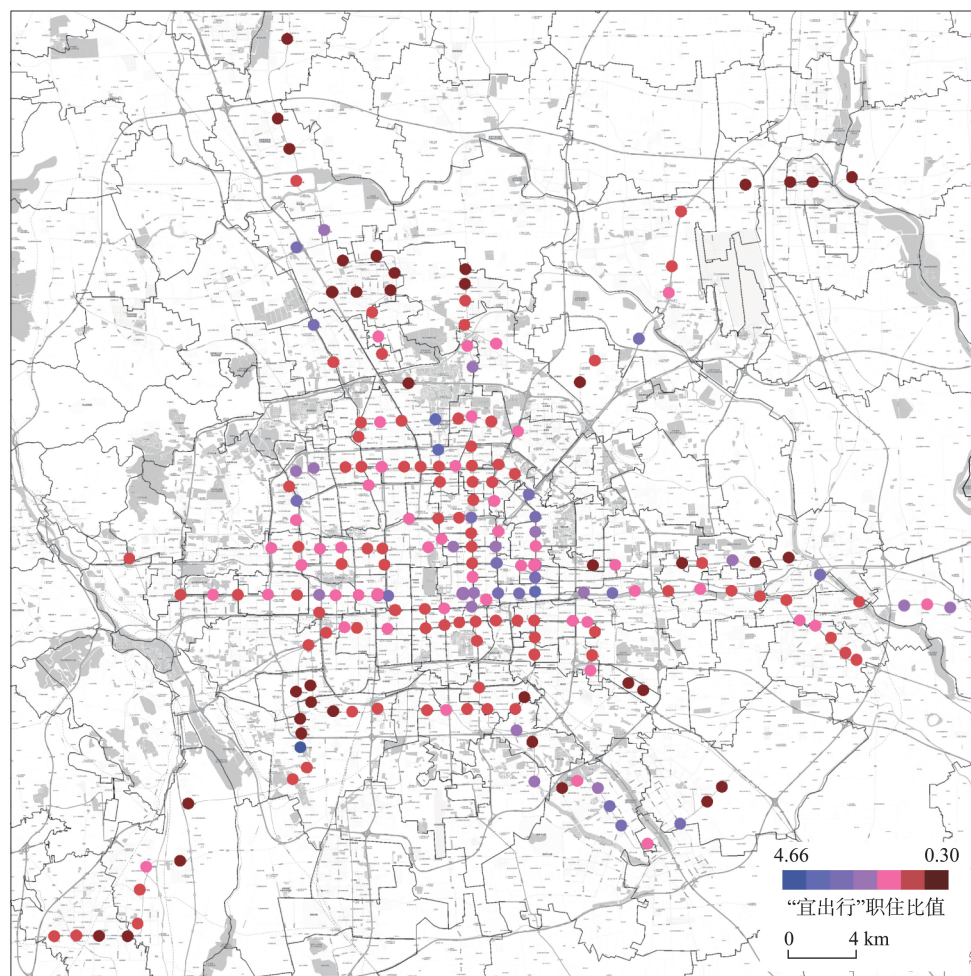


图7 轨道站点合理步行范围内的“宜出行”职住比

Fig.7 Jobs-housing ratio within reasonable pedestrian catchment area of URT stations by "Yichuxing"

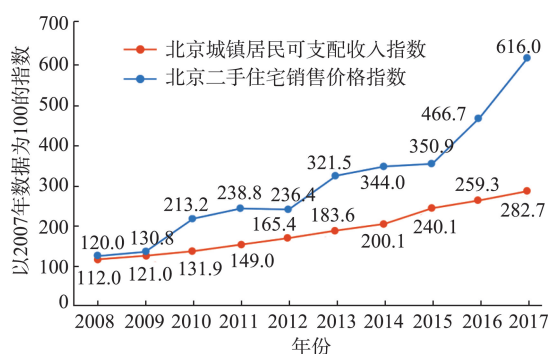


图8 北京市居民收入与住宅价格的指数增长比较

Fig.8 Comparison of exponential growth between resident income and housing price in Beijing

并且为解决数据的残差存在异方差而导致分析结果不稳定的问题,本文使用GARCH模型对数据进行回归分析(Bollerslev, 1986; Lamoureux et al, 1990; Nelson, 1990)。

首先,使用Jarque-Bera方法对自变量和因变量

进行正态分布检验。其中,若偏度和峭度分别为0和3,则表明数据符合正态分布;反之,则表示数据为非正态分布。如表6所示,几乎所有变量的偏度值和峭度值都远离0和3,故不符合正态分布规则。此外,根据Jarque-Bera检测,若拒绝原假设(原假设为接受正态分布),则表明数据不属于正态分布。由检验结果可知(表6),所有数据都显著拒绝了原假设,再次证明各变量不符合正态分布。所以用传统的普通最小二乘法模型无法得到稳健的结果。

由职住平衡的平衡度测算可知,职住比值越接近于1,表示职住关系越平衡;职住比值越远离于1,则表示职住关系越失衡。然而,若将“宜出行”职住比值和站点一卡通出进站比值直接代入模型并不能直观地反映出职住和出进站的平衡状况。在这里,用上述两个比值分别减1后得到与1的偏离程度,以此来反映职住平衡度和一卡通出进站均衡度,然后利用GARCH模型对轨道站点一卡通出进

表6 误差正态性检验结果

Tab.6 Normal distribution test results of Jarque-Bera

参数	jhbal	eibal	initial	transfer	exit	location	distance	facility	ingrass	egrass	jobs	housing
平均值	0.27	0.62	0.05	0.17	3.47	12.45	1.62	0.19	50283.89	52567.85	0.09	0.42
中位值	0.11	-0.21	0	0	3.00	10.41	1.40	0.00	35777.00	30156.00	0	0
最大值	3.66	16.74	1.00	1.00	9.00	34.95	5.25	1.00	231610.00	370124.00	1.00	1.00
最小值	-0.70	-0.92	0	0	1.00	0.95	0.78	0.00	960.00	382.00	0	0
标准偏差	0.61	2.35	0.22	0.38	1.40	7.96	0.75	0.39	42650.71	61332.01	0.29	0.49
偏度	1.85	3.54	4.20	1.76	1.13	0.83	1.96	1.59	1.89	2.27	2.82	0.33
峭度	8.32	19.47	18.65	4.09	5.03	2.91	7.38	3.52	7.25	9.27	8.94	1.11
JB 检测	360.57	2756.99	2708.54	116.31	78.99	23.77	297.46	88.65	277.33	513.91	575.96	34.44
或然率	0	0	0	0	0	0	0	0	0	0	0	0
总和	56.30	127.91	10.00	35.00	715.00	2565.41	334.73	39.00	10358481.00	10828978.00	19.00	86.00
总平方和	76.67	1128.15	9.51	29.05	401.33	12999.49	115.07	31.62	3.73E+11	7.71E+11	17.25	50.10
样本量	206	206	206	206	206	206	206	206	206	206	206	206

站均衡度与站点周边职住平衡度进行回归分析:

$$\begin{aligned} \text{jhbal} = & c_1 + c_2\text{eibal} + c_3\text{initial} + c_4\text{transfer} + \\ & c_5\text{exit} + c_6\text{location} + c_7\text{distance} + \\ & c_8\text{facility} + c_9\text{ingrass} + c_{10}\text{egrass} + \\ & c_{11}\text{jobs} + c_{12}\text{housing} \end{aligned} \quad (10)$$

约束条件为:

$$\varepsilon_t = u_t \sqrt{h_t} \quad (11)$$

$$h_t = b_1 + b_2 h_{t-1} + b_3 \varepsilon_{t-1}^2 \quad (12)$$

式中: $\varepsilon_t$ 为残差值; $h_t$ 为条件方差。

在这里,利用最大似然法则(Maximum Likelihood, ML)对GARCH模型进行计算。

如表7所示,将自变量轨道站点的一卡通出进站均衡度和因变量“宜出行”职住平衡度进行回归后,结果显示:出进站均衡度与职住平衡度之间存在非常显著的正向关系,即通勤时段轨道站点进站和出站的人数越接近,站点周边区域的职住状况越好;相反地,站点进出站人数差距越大,站点周边区域的职住状况越差。另外,在轨道站点的职住功能属性与站点周边职住平衡状况之间的关系方面,典型就业地站点与站点周边区域的职住平衡程度显著正相关,而典型居住地站点与站点周边区域的职住状况存在显著的负相关性。同样地,早高峰时段累计出站刷卡数与职住平衡度在1%的显著性水平上正相关,而累计进站刷卡数与职住平衡度存在一定负向关系。上述关系表明,人口居住越密集,职住关系越趋于失衡。既定区域内居住人口的增长并不会带来同等数量就业岗位的增加。相反,就业岗位越多,职住关系越趋于平衡。通常情况下,购物和餐饮等商业设施多集中分布于就业充足的区域,出于通勤、购物、就餐、社交等日常活动出行效率和便捷度的考虑,即便居住成本较高,不少就业

表7 出进站比对职住比的回归结果

Tab.7 Regression result of jobs-housing ratio and station's egrass-ingrass ratio ratio

变量	相关系数	标准误差	Z检验统计量	概率
eibal	0.1097	0.0149	7.3770	<0.0001
initial	-0.4282	0.0663	-0.6458	0.5184
transfer	0.0171	0.0501	0.3412	0.7330
exit	0.0003	0.0129	0.0216	0.9828
location	-0.0047	0.0026	-1.8174	0.0692
distance	0.0900	0.0283	3.1866	0.0014
facility	0.0726	0.0407	1.7830	0.0746
ingrass	<0.0001	<0.0001	-1.8911	0.0586
egrass	<0.0001	<0.0001	4.2743	<0.0001
jobs	14.1867	2.0788	6.8246	<0.0001
housing	-6.0645	2.1116	-2.8720	0.0041

者仍会选择靠近就业中心——特别是设施齐全、环境优越的就业中心居住。也就是说,完善成熟的就业中心会吸引一定数量的就业者就近居住。相对而言,就业岗位集中的区域通常比居住单元集中的区域职住状况更好。

此外,从加入的控制变量与职住平衡的回归结果可知,轨道站点的始发站属性、换乘站属性、出入口数量与站点周边的职住平衡程度没有联系。轨道站点的区位与职住平衡状况存在一定正相关性,即站点距离城市中心越近,其职住状况越好;相反,则其职住状况越差。另外,轨道站点的站间距越近,职住平衡状况越好。相比其他轨道站点,周边有三甲医院、大型公园、旅游景点的站点职住状况更好。根据前文的描述,总体上北京中心城区的职住状况要优于其他区域,而站间距离相对较近的站点多位于站点密度较大的中心城区。同样地,对于北京而言,重点医院以及可供人们日常锻炼游览的



大型公园和旅游景点主要集中在中心城区。另外,良好的公共服务设施也会对相关产业起到一定的吸引和带动作用。

### 3 结论与讨论

本文以北京市206个轨道站点为例,利用工作日早晚高峰时段的站点一卡通出进站刷卡数据构建高斯混合模型,将轨道站点按照职住属性分为典型就业地站点、典型居住地站点、职住混合地站点等3类,从轨道交通通勤的视角分析居民职住空间关系。然后,以从“宜出行”程序上获取的白天和夜间移动定位数据计算轨道站点合理步行范围内的职住比。由站点周边的就业-居住特征可知,北京就业岗位多位于中心城区特别是四环路沿线及以内的区域,中心城区以外区域的居住空间分异显著。因此,整体上看中心城区的职住状况要好于其他区域。与一些已有研究的结论不同,轨道交通线网末端区域的职住平衡程度通常较差,仅有少数成规模的高端服务业集中分布的轨道站点所在地能够形成区域性的就业中心。另外,尽管职住比数据显示部分城郊地区的就业岗位和居住人口在数量上较为均衡,但仍存在比较突出的职住不匹配问题。

由上文中的回归结果,可以得到以下结论:

首先,轨道站点早高峰出进站比与“宜出行”职住比之间存在的显著正向关系表明,轨道交通通勤行为与职住平衡状况具有高度一致性,通勤时段出进站人数均衡的站点,其周边区域的职住关系也比较平衡。因此,从整体上看轨道站点的通勤出行特征能够有效反映出当地的职住平衡状况。相比实际就业岗位和居住人口数据的难获取性,利用轨道站点刷卡数据测算该地区的职住状况更加便捷可靠。从另一个角度来看,职住失衡可能会导致在特定的时间截面上轨道站点出现客运资源浪费、出行效率低下、安全隐患增加等问题,而职住平衡则有利于改善轨道站点的客流疏导和接运环境。此外,相比传统的人口调查统计方法,“宜出行”数据能够更加便捷有效地反映出既定区域内的职住时空特征。

第二,由典型居住地站点和职住平衡度之间的负向关系以及典型就业地站点和职住平衡度之间的正向关系可知,以居住功能为核心的区域很难保持可持续的发展活力,典型居住区的职住状况通常较差并且难以有效改善,该结论支持了已有的研究(陈蕾等,2011)。相比之下,交通便利、设施完善、环

境优越的就业中心能够吸附大量与之相匹配的商业和其他服务业部门,进而对居民产生足够的吸引力。为改善既定区域内的职住平衡状况,在就业中心周边开发居住单元比在大尺度的居住区发展就业岗位具有更强的可行性。

第三,与之前相关研究的结论有所不同,本文结合大数据方法和数学建模进行分析后发现,在缺乏强有力政策支持和引导的情况下,轨道站点末端地区很难自发形成区域性的商业和就业中心。除少数高端第三产业聚集区和行政职能外迁区外,大多数位于北京近郊区的轨道交通网络末端站点周边的职住状况都比较差。在北京等居住成本高企的中国大都市,对于大多数就业者来说,只有在就业待遇(高收入)和居住水平(低支出)形成错位的情况下,才能维持其较为合理的生活水平。

最后,通过大数据方法对职住平衡状况的分析发现,北京仍属于单中心的城市空间结构。居民主要的就业机会和日常活动场所大多集中在中心城区(四环内),因而居民的职住平衡状况普遍较差且难以得到有效改善。尽管部分城郊地区就业岗位和居住人口在数量上比例均衡,但在很大程度上存在就业-居住关系错位的现象。职住关系的失衡不仅会影响到城市交通的日常运营,还会加剧城市内部区域间的社会经济不平衡,进而影响到整个城市的可持续发展。因此,致力于在交通可达性较好的城郊区域发展良好的公共服务设施(教育、医疗、体育等)在一定程度上能够支撑和带动相关产业的发展,促进区域性中心的形成与拓展,进而改善整个城市的职住平衡状况。

### 参考文献(References)

- 北京市统计局. 2009—2018. 北京统计年鉴 2009—2018 [M]. 北京: 中国统计出版社. [Beijing Municipal Bureau of Statistics. 2009—2018. Beijing statistical yearbook 2009—2018. Beijing, China: China Statistics Press. ]
- 柴彦威, 赵莹, 马修军, 等. 2010. 基于移动定位的行为数据采集与地理应用研究 [J]. 地域研究与开发, 29(6): 1-7. [Chai Y W, Zhao Y, Ma X J, et al. 2010. Mobile positioning method for spatial temporal behavioral data collection and its geographical applications. Areal Research and Development, 29(6): 1-7. ]
- 陈蕾, 孟晓晨. 2011. 北京市居住-就业空间结构及影响因素分析 [J]. 地理科学进展, 30(10): 1210-1217. [Chen L, Meng X C. 2011. A study on the job-housing spatial balance of Beijing. Progress in Geography, 30(10): 1210-1217. ]

- 陈泽东, 谯博文, 张晶. 2018. 基于居民出行特征的北京城市功能区识别与空间交互研究 [J]. 地球信息科学学报, 20(3): 291-301. [Chen Z D, Qiao B W, Zhang J. 2018. Identification and spatial interaction of urban functional regions in Beijing based on the characteristics of residents' traveling. *Journal of Geo-information Science*, 20(3): 291-301. ]
- 程鹏, 唐子来. 2017. 上海中心城区的职住空间匹配及其演化特征研究 [J]. 城市规划学刊, (3): 62-69. [Cheng P, Tang Z L. 2017. A study on the spatial match of jobs-housing balance and its change over time in the central city of Shanghai. *Urban Planning Forum*, (3): 62-69. ]
- 丁亮, 钮心毅, 宋小冬. 2016. 上海中心城就业中心体系测度: 基于手机信令数据的研究 [J]. 地理学报, 71(3): 484-499. [Ding L, Niu X Y, Song X D. 2016. Measuring the employment center system in Shanghai central city: A study using mobile phone signaling data. *Acta Geographica Sinica*, 71(3): 484-499. ]
- 谷一桢, 郑思齐. 2010. 轨道交通对住宅价格和土地开发强度的影响: 以北京市13号线为例 [J]. 地理学报, 65(2): 213-223. [Gu Y Z, Zheng S Q. 2010. The impacts of rail transit on property values and land development intensity: The case of no.13 line in Beijing. *Acta Geographica Sinica*, 65(2): 213-223. ]
- 韩昊英, 于翔, 龙瀛. 2016. 基于北京公交刷卡数据和兴趣点的功能区识别 [J]. 城市规划, 40(6): 52-60. [Han H Y, Yu X, Long Y. 2016. Identifying functional zones using bus smart card data and points of interest in Beijing. *City Planning Review*, 40(6): 52-60. ]
- 金磊, 彭建, 柳昆, 等. 2010. 城市地铁车站分类理论及方法研究 [J]. 地下空间与工程学报, 6(S1): 1339-1342. [Jin L, Peng J, Liu K, et al. 2010. Sorting theory and investigation of urban metro station. *Chinese Journal of Underground Space and Engineering*, 6(S1): 1339-1342. ]
- 孔令斌. 2013. 城市平衡的影响因素及改善对策 [J]. 城市交通, 11(6): 1-4. [Kong L B. 2013. Influential factors and improvement measures for job-housing balance. *Urban Transport of China*, 11(6): 1-4. ]
- 冷炳荣, 余颖, 黄大全, 等. 2015. 大数据视野下的重庆主城区职住关系剖析 [J]. 规划师, 31(5): 92-96. [Leng B R, Yu Y, Huang D Q, et al. 2015. Big data based job-residence relation in Chongqing metropolitan area. *Planners*, 31(5): 92-96. ]
- 李向楠. 2015. 城市轨道交通站点分类的聚类方法研究 [J]. 铁道标准设计, 35(4): 19-23. [Li X N. 2015. Classifying urban rail transit stations using cluster analysis. *Railway Standard Design*, 35(4): 19-23. ]
- 李祖芬, 于雷, 高勇, 等. 2016. 基于手机信令定位数据的居民出行时空分布特征提取方法 [J]. 交通运输研究, 2(1): 51-57. [Li Z F, Yu L, Gao Y. 2016. Residents' Trips based on Cellular Signaling Data. *Transport Research*, 2(1): 51-57. ]
- 刘康, 吴群, 王佩. 2015. 城市轨道交通对住房价格影响的计量分析: 以南京市地铁1、2号线为例 [J]. 资源科学, 37(1): 133-141. [Liu K, Wu Q, Wang P. 2015. Econometric analysis of the impacts of rail transit on property value: The number 1 and 2 lines in Nanjing. *Resources Science*, 37(1): 133-141. ]
- 刘瑜, 肖昱, 高松, 等. 2011. 基于位置感知设备的人类移动性研究综述 [J]. 地理与地理信息科学, 27(4): 8-13. [Liu Y, Xiao Y, Gao S, et al. 2011. A review of human mobility research based on location aware devices. *Geography and Geo-information Science*, 27(4): 8-13. ]
- 刘志林, 王茂军. 2011. 北京市职住空间错位对居民通勤行为的影响分析: 基于就业可达性与通勤时间的讨论 [J]. 地理学报, 66(4): 457-467. [Liu Z L, Wang M J. 2011. Job accessibility and its impacts on commuting time of urban residents in Beijing: From a spatial mismatch perspective. *Acta Geographica Sinica*, 66(4): 457-467. ]
- 龙瀛, 张宇, 崔承印. 2012. 利用公交刷卡数据分析北京职住关系和通勤出行 [J]. 地理学报, 67(10): 1339-1352. [Long Y, Zhang Y, Cui C Y. 2012. Identifying commuting pattern of Beijing using bus smart card data. *Acta Geographica Sinica*, 67(10): 1339-1352. ]
- 陆化普, 丁宇, 张永波. 2013. 中国城市职住均衡实证分析与关键对策 [J]. 城市交通, 11(3): 1-6. [Lu H P, Ding Y, Zhang Y B. 2013. Urban job-housing balance in China: Empirical analysis and countermeasures. *Urban Transport of China*, 11(3): 1-6. ]
- 罗芳, 柴蕾, 邱星. 2015. 基于交通换乘的城市轨道交通车站分类研究 [J]. 洛阳理工学院学报(自然科学版), 25(3): 45-48. [Luo F, Chai L, Qiu X. 2015. Classification of urban rail transit station based on traffic transfer. *Journal of Luoyang Institute of Science and Technology (Natural Science Edition)*, 25(3): 45-48. ]
- 马亮. 2017. 基于轨道交通刷卡数据的城市通勤圈范围研究 [J]. 城市轨道交通研究, 20(8): 80-84. [Ma L. 2017. Study on urban commuting circle based on rail card data. *Urban Mass Transit*, 20(8): 80-84. ]
- 孟斌, 高丽萍, 黄松, 等. 2017. 北京市典型就业中心职住关系考察 [J]. 城市问题, (12): 86-94. [Meng B, Gao L P, Huang S, et al. 2017. Explore the home-job relationship of typical employment centers in Beijing. *Urban Problems*, (12): 86-94. ]
- 孟庆艳, 徐浩澜. 2008. 上海轨道交通1号线对人口再分布的作用 [J]. 城市轨道交通研究, 11(3): 7-12. [Meng Q Y, Xu H L. 2008. The impact of UMT on population redistribution: A case study of Shanghai UMT line 1. *Urban Mass Transit*, 11(3): 7-12. ]
- 孟晓晨, 吴静, 沈凡卜. 2009. 职住平衡的研究回顾及观点综述 [J]. 城市发展研究, 16(6): 23-35. [Meng X C, Wu J, Shen F B. 2009. The studies review of urban jobs-housing balance. *Urban Studies*, 16(6): 23-35. ]
- 邵滢宇, 丁柏群. 2015. 基于聚类分析的地铁站点分类: 以哈



- 尔滨地铁1号线为例[J]. 森林工程, 31(3): 106-111. [Shao Y Y, Ding B Q. 2015. Metro stations classification based on clustering analysis: A case study of Harbin metro line 1. Forest Engineering, 31(3): 106-111. ]
- 申犁帆, 王烨, 姜冬睿, 等. 2018. 轨道站点合理步行可达范围建成环境与轨道通勤的关系研究: 以北京市44个轨道站点为例[J]. 地理学报, 73(12): 2423-2439. [Shen L F, Wang Y, Jiang D R, et al. 2018. Relationship between built environment of rational pedestrian catchment areas and URT commuting ridership: Evidence from 44 URT Stations in Beijing. Acta Geographica Sinica, 73(12): 2423-2439. ]
- 申犁帆, 张纯, 李赫, 等. 2018. 大城市通勤方式与职住失衡的相互关系[J]. 地理科学进展, 37(9): 1277-1290. [Shen L F, Zhang C, Li H, et al. 2018. Interaction between commuting modes and jobs-housing imbalance in metropolis: An empirical study by Bayesian-Tobit analysis in Beijing. Progress in Geography, 2018, 37(9): 1277-1290. ]
- 申悦, 柴彦威. 2013. 基于GPS数据的北京市郊区巨型社区居民日常活动空间[J]. 地理学报, 68(4): 506-516. [Shen Y, Chai Y W. 2013. Daily activity space of suburban mega-community residents in Beijing based on GPS data. Acta Geographica Sinica, 68(4): 506-516. ]
- 谭欣, 黄大全, 赵星烁, 等. 2016. 基于百度热力图的职住平衡度量研究[J]. 北京师范大学学报(自然科学版), 52(5): 622-628. [Tan X, Huang D Q, Zhao X S, et al. 2016. Jobs housing balance based on Baidu thermodynamic diagram. Journal of Beijing Normal University (Natural Science), 52(5): 622-628. ]
- 汤庆园, 徐伟, 艾福利. 2012. 基于地理加权回归的上海市房价空间分异及其影响因子研究[J]. 经济地理, 32(2): 52-58. [Tang Q Y, Xu W, Ai F L. 2012. A GWR-based study on spatial pattern and structural determinants of Shanghai's housing price. Economic Geography, 32(2): 52-58. ]
- 许宁, 尹凌, 胡金星. 2014. 从大规模短期规则采样的手机定位数据中识别居民职住地[J]. 武汉大学学报(信息科学版), 39(6): 750-756. [Xu N, Yin L, Hu J X. 2014. Identifying home-work locations from short-term, large-scale, and regularly sampled mobile phone tracking data. Geomatics and Information Science of Wuhan University, 39(6): 750-756. ]
- 许园园, 塔娜, 李响. 2017. 基于地铁刷卡数据的城市通勤与就业中心吸引范围研究[J]. 人文地理, 32(3): 93-101. [Xu Y Y, Ta N, Li X, et al. 2017. Analysis of commuting behavior and employment center using subway smart card data. Human Geography, 32(3): 93-101. ]
- 许志榕. 2016. 上海市职住关系和通勤特征分析研究: 基于轨道交通客流数据视角[J]. 上海城市规划, (2): 114-121. [Xu Z R. 2016. Study on job-housing relationship and characteristics of commuting in Shanghai: Based on the perspective of rail transit passenger flow data. Shanghai Urban Planning Review, (2): 114-121. ]
- 岳芳, 毛保华, 陈团生. 2007. 城市轨道交通接驳方式的选择[J]. 都市轨道交通, 20(4): 36-39. [Yue F, Mao B H, Chen T S. 2007. Study on the choice of access modes to urban rail transit. Urban Rapid Rail Transit, 20(4): 36-39. ]
- 岳真宏, 陈峰, 王子甲, 等. 2017. 基于刷卡数据和高斯混合聚类的地铁车站分类[J]. 都市快轨交通, 30(2): 48-51, 107. [Yue Z H, Chen F, Wang Z J, et al. 2017. Classification of metro stations by clustering smart card data using the Gaussian mixture model. Urban Rapid Rail Transit, 30(2): 48-51, 107. ]
- 曾智超, 林逢春. 2005. 城市轨道交通对城市人口迁移的作用[J]. 城市轨道交通研究, 8(2): 19-22. [Zeng Z C, Lin F C. 2005. The effect of urban rail transit on population migration. Urban Mass Transit, 8(2): 19-22. ]
- 张纯, 易成栋, 宋彦. 2016. 北京市职住空间关系特征及变化研究: 基于第五、六次人口普查和2001、2008年经济普查数据的实证分析[J]. 城市规划, 40(10): 59-64. [Zhang C, Yi C D, Song Y. 2016. Characteristics of job-housing spatial relationship and changes in Beijing: An empirical study based on data from the 5th, 6th population census and economy census in 2001 and 2008. City Planning Review, 40(10): 59-64. ]
- 张天然. 2016. 基于手机信令数据的上海市域职住空间分析[J]. 城市交通, 14(1): 15-23. [Zhang T R. 2016. Job-housing spatial distribution analysis in Shanghai metropolitan area based on cellular signaling data. Urban Transport of China, 14(1): 15-23. ]
- 张文忠. 2001. 城市居民住宅区位选择的因子分析[J]. 地理科学进展, 20(3): 268-275. [Zhang W Z. 2001. An analysis of the factors that influence the urban residential location selection. Progress in Geography, 20(3): 268-275. ]
- 赵晖, 杨军, 刘常平. 2011. 轨道沿线居民职住分布及通勤空间组织特征研究: 以北京为例[J]. 经济地理, 31(9): 1445-1451. [Zhao H, Yang J, Liu C P. 2011. Research on characteristics of the job/housing distribution and the changing commuting along subway lines: A case study from Beijing. Economic Geography, 31(9): 1445-1451. ]
- Bollerslev T. 1986. Generalized autoregressive conditional heteroskedasticity [J]. Journal of Econometrics, 31(3): 307-327.
- Cervero R. 1989. Jobs-housing balancing and regional mobility [J]. Journal of the American Planning Association, 55(2): 136-150.
- Cervero R, Duncan M. 2006. Which reduces vehicle travel more: Jobs-housing balance or retail-housing mixing [J]. Journal of American Planning Association, 72(4): 475-490.
- Ewing R, Pendall R, Don C. 2004. Measuring sprawl and its impact [M]. Washington, DC: Smart Growth America.
- Hansen B. 2000. Sample splitting and threshold estimation [J]. Econometrica, 68(3): 575-603.
- Ho M, Chiu R. 2002. Impact of accessibility on housing expenditure and affordability in Hong Kong's private rental sec-

- tor [J]. *Journal of Housing & the Built Environment*, 17 (4): 363-383.
- Lamoureux C, Lastrapes W. 1990. Heteroskedasticity in stock return data: Volume versus GARCH effects [J]. *The Journal of Finance*, 45(1): 221-229.
- Liu L, Biderman A, Ratti C. 2009. Urban mobility landscape: Real time monitoring of urban mobility patterns [C]// *The 11th International Conference on Computers in Urban Planning and Urban Management*. Hong Kong, China: The University of Hong Kong.
- Medina S, Erath A. 2013. Estimating dynamic workplace capacities by means of public transport smart card data and household travel survey in Singapore [J]. *Transportation Research Record*, 2344: 20-30.
- Nelson D. 1990. Stationarity and persistence in the GARCH(1, 1) model [J]. *Econometric Theory*, 6(3): 318-334.
- Zielstra D, Hartwig H. 2011. A comparative study of pedestrian accessibility to transit stations using free and proprietary network data [J]. *Transportation Research Record Journal*, 2217: 145-152.

## Relationship between urban rail transit commuting and jobs-housing balance: An empirical analysis from Beijing based on big data methods

SHEN Lifan<sup>1</sup>, ZHANG Chun<sup>2\*</sup>, LI He<sup>3</sup>, WANG Ye<sup>4</sup>, WANG Zijia<sup>5</sup>

(1. School of Urban Design, Wuhan University, Wuhan 430072, China; 2. School of Architecture and Design, Beijing Jiaotong University, Beijing 100044, China; 3. International Finance Institute, Bank of China, Beijing 100818, China;  
4. Guangzhou Planning & Design Survey Research Institute, Guangzhou 510030, China;  
5. School of Civil Engineering, Beijing Jiaotong University, Beijing 100044, China)

**Abstract:** The development of urban rail transit (URT) network improves the commuting efficiency of residents while it has a certain impact on their jobs-housing balance. This study took 206 URT stations in Beijing as an example and classified them according to their jobs-housing functions based on the Gaussian mixture model (GMM) and smart card data. The dynamic population distribution characteristics around URT station were explored and jobs-housing ratio was calculated by "Yichuxing" position data. The study found that: 1) The jobs-housing balance in the central city is obviously better than that outside of the central city. 2) At the ends of the URT network, the jobs-housing balance is worse while only a few stations with concentrated distribution of top service industries have formed regional employment centers. 3) There still exists a certain degree of jobs-housing mismatch in the areas around some suburban stations where employment and residential functions are relatively equal. Station outflow-inflow and jobs-housing balances were calculated by the station egress-ingress ratio and the jobs-housing ratio, and the correlation between URT commuting behavior and jobs-housing balance was analyzed by generalized autoregressive conditional heteroskedasticity (GARCH) model. The results of this study indicate that: 1) There is a very strong positive relationship between URT station egress-ingress balance and jobs-housing balance. The closer the numbers of URT station outflow and inflow population, the better the jobs-housing balance around the URT station is. 2) There is a strong positive relationship between employment opportunity and jobs-housing balance around a URT station; and there is a strong negative relationship between residential function and jobs-housing balance around a URT station. This suggests that dense settlement will not generate the same quantity of jobs while well-developed employment hubs can attract a certain number of residents to live nearby. 3) There is a positive correlation between locational conditions of URT stations and jobs-housing balance. 4) The GMM can effectively cluster URT stations with complex and unclear attributes. 5) With its advantages of real-time data capturing, high precision, wide coverage, and great accessibility, "Yichuxing" position data can effectively compensate for the limitations of other methods on collecting and analyzing spatial-temporal characteristics of real-time population distribution at the microscopic scale.

**Keywords:** urban rail transit; commuting behaviour; jobs-housing balance; big data; Gaussian mixture model; generalized autoregressive conditional heteroskedasticity (GARCH) model; Beijing