

结伴出行推动消费升级：基于铁路大数据的分析^①

曹炜威^{1,5,6}，刘 圣²，李宜威²，贾建民^{3,4*}

- (1. 西南交通大学经济管理学院, 成都 610031; 2. 岭南大学商学院, 中国香港 999077;
3. 香港中文大学(深圳)经管学院, 深圳 518172; 4. 深圳市大数据研究院, 深圳 518172;
5. 中国民用航空飞行学院, 广汉 618307; 6. 民航飞行技术与飞行安全重点实验室, 广汉 618307)

摘要：结伴行为日益普遍，其在出行、购物、餐饮、休闲及其他商业领域对消费者的购买决策有着重要的影响。随着大数据时代的到来，人们的行为能够被更加精准、系统地记录，为研究结伴行为提供了有力的工具。本文认为，结伴行为背后蕴含了消费者对高品质生活的追求，有助于推动产品消费升级。本文借助铁路出行的视角，基于成渝铁路的客票销售数据，探究消费者结伴出行对铁路客运产品选择的影响。通过应用赫克曼二阶段模型并控制超过130万名乘客的个体固定效应，本文揭示了结伴出行显著正向影响消费者对高级列车和高等座位的选择。另外，人口统计学特征起着显著的调节作用，女性以及年长者结伴出行更倾向于选择高级列车和高等座位。这些发现不仅为铁路管理提供了新思路，更重要的是揭示了结伴行为对消费升级的正向影响作用，为商业服务和公共管理等广泛领域都提供了重要的参考。企业、单位及政府通过鼓励人们的结伴行为，有利于刺激消费者对优质产品和服务的选择。

关键词：结伴；消费升级；大数据；选择行为；高铁

中图分类号：C931.6 **文献标识码：**A **文章编号：**1007-9807(2020)02-0017-21

0 引言

结伴，一个基于社会互动、情感交流、熟人同行、从众等动机的行为，在当今的社会活动中日益普遍，如结伴出行、结伴购物、结伴旅游、结伴休闲等^[1]。结伴在许多方面产生着重大的影响：如结伴群体的目标制定和分工较为合理，而目标是刺激和改善行为的一个重要因素，清晰明确的目标使群体的行动进展更加顺利^[2]。在决策过程中，个体的决策行动会受到结伴群体的影响从而调整自己的行为^[3]。在实现目标和行动的过程中，结伴的个体之间可以互相沟通和交流，为其提供了良好的人际关系交往机会，这既满足了个体的情感需求，也使其更容易产生积极的团队意

识^[4]。结伴情景下所取得的成就往往也比不结伴要好，积极的情感和显著的成就均会使个体的满足感更高^[5]。

结伴活动中，消费者不免面临各种决策问题，而当前学术研究多把目光集中在个体如何根据其信念、偏好和心理动机等做出决策上，对于涉及到其他人的决策研究略显不足。实际上，生活以及商业消费中很多的决策都会以不同的形式涉及到其他人^[6]，直接地表现之一为结伴情况下的消费行为。消费者的决策过程受到群体规范的影响，尤其是当结伴群体中存在一些代表性人物的时候^[7]；而有时仅是他人的“存在”就能影响消费行为，不管他人是朋友或是家人^[8]。结伴消费时，群体中的个人会更倾向调整自己的购买行为从而与群体

① 收稿日期：2019-05-30；修订日期：2019-12-27。

基金项目：国家自然科学基金资助重大项目(71490722)。

通讯作者：贾建民，男，四川邻水人，教授，博士生导师。Email: jmjia@cuhk.edu.cn

协调,在维持自己良好的情绪和体验的同时也兼顾团队其他成员的意见和利益.结伴对购买行为的影响还体现在推动对产品或服务的消费升级上(消费需求由生存型向享受型再向发展型转变),促进对中高端产品的需求.当前我国经济已经由高速增长转向高质量发展,消费对经济增长的拉动作用日益突出,居民个人的消费需求也在不断升级^[9].一方面,结伴群体往往追求产品结构的丰富性,并伴随着高频率的购买行为^[8];另一方面,消费者结伴购物时更加看重产品的质量维度,会更愿意选择高质量的产品和服务.优质的中高端产品和服务有助于提高生活品质,能够更好地满足消费者的行为偏好和心理需求^[10].因此,结伴行为的背后有可能意味着人们对于高品质生活的追求,有助于推动消费升级,这也是本文的研究重点.

如今,大数据时代的到来,消费者在很多社会经济系统中留下了丰富的行为痕迹.利用先进的技术手段能够获取顾客的多维度行为特征数据,借助强大的计算工具可以对这些数据进行实时性的分析^[11].这一方面有助于了解市场的动态变化,另一方面也可以加强对企业优势和劣势的认知,为管理的改进方向提供参考和建议^[12].当前快速发展的信息技术为研究消费者行为(如结伴)提供了有力的工具.本文将借助铁路客运产品的海量在线销售数据(超过130万人的3 939 487条出行记录),以铁路出行中客运列车和座位的选择为例,检验结伴出行是否影响人们对高品质产品的选择,以此揭示结伴行为对消费升级的影响.

本文在回顾前人研究的基础上,首次在铁路研究中使用赫克曼二阶段模型(Heckman Two-Stage Model)分步对乘客的出行行为和铁路产品的选择行为进行建模,检验结伴出行这一核心变量对出行列车和座位选择的影响.通过分析2017年1月1日~2017年3月31日期间成都铁路局辖内销售的客票记录数据并控制超过130万乘客的个体固定效应,结果发现:结伴出行显著正向影响高级列车和高等座位的选择(中高端产品和服务);女性结伴显著促进对高级列车和高等座位的选择;年长者结伴显著促进对高级列车和高等座位的选择.这些发现反映了结伴行为背后人们

对高品质生活的追求,揭示了结伴行为对消费升级的推动作用.虽然铁路产品是一个比较特殊的领域,但是在本文使用的两个因变量中(列车和座位),座位涉及到广泛的商业领域,比如电影院、飞机、高档餐厅等.所以通过结伴出行对于铁路列车和座位选择的显著正向影响可以反映结伴者对中高端产品的消费.结伴的商业价值不仅为铁路部门的管理提供了思路,也为整个商业和公共管理领域提供了有价值的参考.

1 文献与假设

结伴出行通常发生在关系亲密的熟人之间,例如家人、亲戚、朋友或同事^[13].和关系亲密的人在一起时,人们的价格敏感度会降低,心理上对价格较贵的产品和服务的接受度会更大,表现出对中高端产品更强的消费欲望,从这点来看结伴出行将有助于推动消费行为升级^[14].结伴出行有助于同行者产生积极的情感体验,激发享乐的动机^[15],从而导致出行者偏向选择更高品质的产品.Garcia-Rada等在其研究中指出,决策中涉及到与自身具有浪漫关系的他人时,相对于实用型产品消费者更倾向于选择享乐型产品^[16].在铁路出行客运产品中,高级的列车和座位具有更舒适的乘坐服务和乘车环境(比如车内清洁、工作人员服务态度好、出发到达准时等),代表出行环境下的中高端消费品,能够满足结伴出行时的享乐动机,更符合人们结伴情境下的选择倾向.因此,本文提出以下假设:

H1a 结伴出行正向影响对高级列车的选择.

H1b 结伴出行正向影响对高等座位的选择.

结伴出行和个性特征差异的协同效应也会对铁路产品选择产生影响.出行者个体特征的差异会导致其在心理和行为的诸多方面都大相径庭,如交通出行、情感交流、互联网使用等^[17].性别是个体特征差异的主要表现之一,不同性别的出行者在出行链条与出行方式选择上都有较大不同,这既有生理层面的影响也有心理和行为层面的原因^[18].相比男性,女性在消费决策中容易受到周边环境的影响,这可能使得女性会倾向于价格更

高且品质也更好的中高端产品^[19]。同时,女性相比男性更加体贴和关怀他人^[20],所以女性在结伴出行中也可能因顾及他人的感受而选择高品质的客运产品。代表中高端出行消费品的高级列车和高等座位往往让人感到更舒适,能满足享乐动机和女性体贴关怀的想法。所以本文提出以下假设:

H2a 女性结伴出行正向影响对高级列车的选择。

H2b 女性结伴出行正向影响对高等座位的选择。

除了性别,年龄是个体特征差异的另一个表现。不同年龄的个体在消费过程中的目标选择上存在较大差异。已有的证据表明,消费者的年龄会影响个体的消费行为^[21]。平均而言,年长者的健康状态相对较差,相对年轻者而言身体和心理更需要被照顾^[22-23],出行中客观上更需要环境舒适、服务优质的乘车环境(高级列车和高等座位可以提供)。与之相应,人们理解长者的需求,也会给予长者较多的关心和帮助^[24],所以与年长者同行,会更加倾向选择环境、服务和安全性更好的中高端客运产品满足长者的客观需要。而且,年长者在社会交往中获取社会认同感的意愿较为强烈^[25]。在城际铁路出行中,高级列车和高等座位还具有符号价值,一定程度上能够体现身份地位,有助于出行者获取社会认同。因此,年长者结伴可能对消费升级具有促进作用,正向影响对高级列车和高等座位的选择。所以,本文提出以下假设:

H3a 年长者结伴出行正向影响对高级列车的选择。

H3b 年长者结伴出行正向影响对高等座位的选择。

2 数据

2.1 数据总览

本文研究的目的是借助铁路出行视角检验结伴对消费升级的促进作用。客票数据中包含丰富的个人信息和出行属性信息,例如旅客身份唯一标识符、性别、年龄、证件地址、所乘列车类型(高铁、动车、普速列车)、座位席别(商务座、一等座、二等座、硬座、硬卧、软座、软卧)、购票时间、购票

渠道(窗口、自助购票机、互联网、移动端)、出行距离、车票费用、出发站、到达站、订单序列号等。根据研究目的,借助相关字段信息能够获取研究所需变量。在本次研究中以成渝铁路为例,收集了2017年1月1日~2017年3月31日连续三个月内成渝间直达旅客的出行数据,共计3 939 487条,作为观测样本实证分析结伴出行对客运产品选择的影响。

2.2 成渝间列车及座位描述

随着经济的发展、技术的进步以及在国家经济发展蓝图中战略地位的提升,成渝经济区铁路运输系统客运列车产品结构日趋完善,形成了包括G字头高铁、D字头动车和K字头列车为主的普速列车在内的多层次客运市场结构。高铁平均运营时速300km/h,成渝间耗时约1.5h;动车平均运营速度200km/h,全程耗时约2h10min;普速列车平均运营时速约100km/h,全程平均耗时约3h40min。高铁、动车和普速列车路线走向和途径站点不尽相同,但均连接了成都和重庆两大区域核心城市,为两地人口流动提供服务。K字头为主的普速列车运营时间较早,是沿线城市间人口流动的基础交通工具。动车于2009年7月7日全线投入运营,成渝高铁于2016年1月1日正式开通运营,自此成渝间形成了三种类型列车共存的客运格局。成渝间开行的列车中,不同类型的列车可选择的座位及相应的票价如表1所示。

表1 成渝间不同等级列车及座位描述

Table 1 Description of grades of trains and seats

列车类型	座位类型	票价(元)
普速列车 (K字头为主)	硬座	46.5
	硬卧	92.5
	软卧	140.5
动车(D字头)	二等座	96.5
	一等座	155.5
高铁(G字头)	二等座	146.0
	一等座	233.5
	商务座	438.0

综合考虑列车运营速度、车厢环境、座位舒适度、票价等因素,结合对相关领域专家访问对列车和座位进行如下等级排序。列车等级排序:高铁>动车>普通列车。座位等级排序:高铁商务座、高铁一等座>高铁二等座、动车一等座、普速列车软卧>动车二等座、普速列车硬卧>普速列车硬座。

由于商务座乘客比例不足 1%, 所以将其与一等座合并归为一类。

2.3 变量描述

2.3.1 列车

列车等级信息是假设检验所需的信息之一。列车的等级(高铁、动车、普通列车)由销售数据中的列车车次名称获得(G字头高铁, D字头动车和K字头普通列车)。钟业喜等^[26]在其对于高铁影响的研究中根据高铁的定义将G开头的高速动车组、D开头的动车组和C开头的城际高速均视为高速列车, 其他视为普通列车。本文所用数据为成渝铁路段, 不存在C字头的城际高速, 所以此研究将列车等级分为高铁、动车和普通列车。根据表2中的票价信息, 高铁大于动车、动车大于普通列车, 故本研究按列车等级高低分别将高铁、动车和普通列车取值为3、2和1。

2.3.2 座位

座位等级信息是假设检验所需的另一信息。座位的等级信息由客票记录数据直接获得。过去的研究一般将座位等级信息编译为分类变量, 如王文宪等^[27]在其对于铁路产品选择研究中, 把客运产品分为硬座、硬卧及软卧。结合过去研究的度量方法和表1中的票价信息, 本文将高铁商务座、高铁一等座取值4, 高铁二等座、动车一等座和普通列车软卧取值3, 动车二等座和普通列车硬卧取值2, 普通列车硬座取值为1。

2.3.3 是否结伴出行

结伴出行会影响消费产品的选择, 该变量为本文研究的核心信息。结伴信息根据乘车日期、车次和订单号信息来判断。过去对于结伴的测度常用的方式为是否结伴, 例如Zhang等^[28]在对结伴进行研究时把结伴的存在与否作为变量对其影响进行分析。本文对于结伴与否信息的提取方式是: 如果某几个出行者购买同一天的同一车次且购票记录的订单号相同, 表示本次出行为结伴出行, 变量取值为1; 否则取值为0。

2.3.4 其他信息

根据客票字段信息还可以提炼多种其他的个体特征信息, 包括性别、年龄、是否城市人口、是否移动端购票、出行平均花费、出行平均里程, 以此控制人口统计学特征、购票特征、社会阶层与地位因素对列车和座位选择的影响。

1) 人口统计学特征

人口统计学特征对于出行产品的选择有显著影响^[29-30], 如性别、年龄和是否城市人口变量。性别在研究中一般作为二分变量进行测度, 即男性或者女性。如果是女性, 取值1; 若为男性, 取值0。年龄在研究中一般被当作连续变量^[31], 本文年龄变量根据客票数据中包含的旅客出生日期计算而来。是否城市人口变量根据客票数据中包含的旅客地址信息判断, 地址一定程度上可视为常住地址的代理变量。如果常住地为城市地址, 变量值为1; 否则为0。还有一些其他的人口统计学信息在已有文献中被指出有显著影响如出行者个人收入、行程目的、行程前时间等, 这些信息无法从客票记录数据中获得, 所以在结伴出行对铁路产品的分析模型中(下一节), 每个出行者的个体固定效应将被加入分析, 从而控制由于个体差异(如个人收入、行程目的等)带来的影响, 弥补了铁路人口统计学特征变量信息的限制。

2) 购票渠道变量

购票方式对消费者的选择行为会产生影响, 例如Hemamalini^[32]指出消费者认为在线订票更加方便。由于其他购票方式如传统窗口购票和自助端购票数据无法观测旅客是否结伴出行, 故本文研究所用的数据仅包括互联网购票和移动端购票数据。如果采用移动端方式购票, 是否移动端购票变量取值1, 否则取值0。

3) 平均花费和平均里程

消费者过去的经历会影响未来的选择, 过去出行的平均花费和平均里程属于历史铁路出行的两个直接表现。价格会影响消费者的感受, 如满意度和公平感^[33], 从而影响出行者的选择。出行者的历史平均花费反映了出行者乘车的平均价格高低, 将会对未来的乘车行为产生影响。另一方面, 出行距离也会对出行的选择产生显著影响^[34]。根据2014年1月1日~2016年12月31日旅客的出行记录对历史平均花费和平均里程进行计算。由该时间段出行的票价总花费除以这段时间段出行总次数, 从而得到每个出行者过去出行中每次铁路出行的花费。由该时间段出行的总里程除以这个时间段出行总次数, 从而得到每个出行者过去出行中每次铁路出行的里程。表2是变量的描述性统计。

表2 变量的描述性统计

Table 2 Descriptive statistics of variables

变量	描述	均值	标准差	最小值	最大值
(1) 列车	此次出行的列车等级(高铁、动车和普速列车).	2.688	0.615	1.000	3.000
(2) 座位	此次出行的座位等级(商务座、一等座、二等座、硬座、硬卧、软座、软卧).	2.773	0.647	1.000	4.000
(3) 是否结伴	当次出行是否与人结伴.	0.858	0.348	0.000	1.000
(4) 是否女性	出行者本人的性别是否为女性.	0.435	0.495	0.000	1.000
(5) 年龄	出行者的年龄.	34.010	12.117	2.000	102.000
(6) 是否移动端购票	当次出行是否使用了移动端进行购票.	0.732	0.442	0.000	1.000
(7) 平均花费	2014年~2016年旅客铁路总花费(角)除以总次数.	1143.800	650.509	20.000	26725.000
(8) 平均里程	2014年~2016年旅客铁路总里程(公里)除以总次数.	399.100	310.665	3.000	4551.000
(9) 是否城市	出行者常住地址是否为城市.	0.541	0.498	0.000	1.000

2.4 数据描述性分析

为了初步探究结伴对列车选择的影响,首先将总体数据分为有结伴和无结伴两个组,分别计算两组(结伴与无结伴)中每位出行者选择高铁的次数在该出行者所有出行中的占比.通过对结伴出行的高铁选择占比与无结伴出行的高铁选择占比的T检验,结果显示结伴出行的高铁选择比例均值显著高于个体(无结伴)出行的高铁比例均值($t=4.22, p<0.001$).图1(左)展示了个体出行与结伴出行选择高铁的均值.同样,为了初步探究结伴对于座位选择的影响,分别计算两个组(个体出行与结伴出行)中每位出行者选择三等级座位和四等级座位(高铁商务座、高铁一等座

取值4,高铁二等座、动车一等座和普速列车软卧取值3)的次数在该出行者所有出行次数中的占比.通过对于结伴出行的选择高等座位占比与个体(无结伴)出行的高等座位占比的T检验,结果显示结伴出行选择高等座位比例均值显著高于无结伴出行选择高等座位比例均值($t=3.40, p<0.001$).图1(右)展示了选择高等座位的均值及95%置信区间下均值的范围.初步描述性统计分析的检验结果支持本文的核心假设,即结伴出行对出行者高级列车和高等座位的选择有潜在正向的影响,而在更加严密的模型设定下,假设是否依旧成立,本文在下一部分对此做进一步验证.

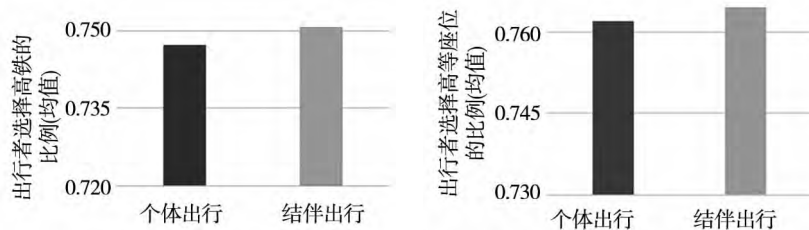


图1 结伴出行与否对于列车和座位选择的描述性统计

Fig. 1 Descriptive analysis of travelling companion's impact on train and seat choice

旅客选择铁路产品出行可能会受前期乘坐体验的作用而影响后续选择行为,因此本文抽取了

总计出行两次且均为单独出行(103490人,206980次出行)的观测样本,检验多次出行的旅

客在列车和座位选择上是否具有 consistency, 结果如图 2. 结果显示, 旅客单独出行两次, 其第二次出行会更多地选择高铁 ($t = 2.30, p < 0.05$) 和高等座位 ($t = 2.32, p < 0.05$). 也就是

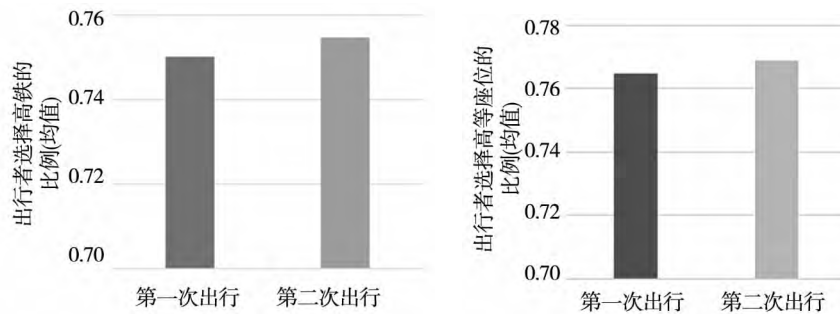


图 2 旅客多次出行中列车和座位选择一致性分析

Fig. 2 Consistency analysis of multi-trips impact on train and seat choice

以上分析通过比较结伴人群与非结伴人群的铁路产品消费, 初步证实了结伴行为对于消费的刺激作用. 为了更加直观地检验结伴行为对于消费的影响, 本文直接比较了同一乘客在非结伴与结伴下的铁路产品消费. 抽取了出行两次的旅客

说, 多次出行会一定程度上刺激旅客选择高级铁路客运产品. 故本文在后续模型分析中将控制出行次数的固定效应, 这在一定程度上可以缓解多次出行对铁路客运产品选择的影响.

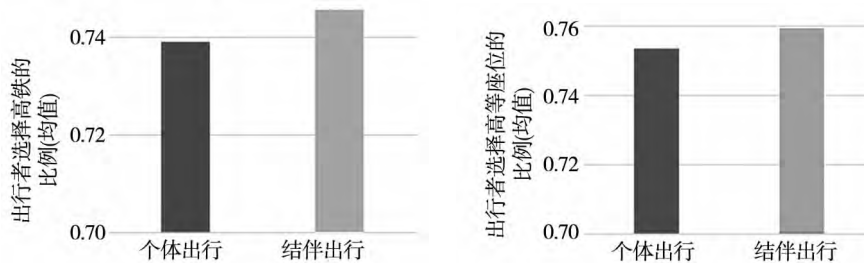


图 3 同一乘客结伴出行与非结伴出行选择的差异性分析

Fig. 3 Consumption difference between individual travel and travelling companion of the same passenger

3 模型介绍

模型选择方面, 过去对于出行产品和方式选择的研究往往通过单一 Logit 模型定量分析出行属性特征(程前、程后接驳时间、在途时间、服务频次、票价花费等)和个体属性特征(性别、年龄、教育、收入等)对出行方式选择行为的影响^[35-37]. 例如 Behrens 等^[38]基于混合 Logit 模型研究伦敦至巴黎通道内人们在高铁和民航两种客运产品之间的选择行为; Park 等^[39]使用了多项 Logit 模型研究韩国高铁的开通对民航出行的影响, 发现票价费用和服务频次显著影响旅客选择交通方式的效用感知. Lee 等^[40]应用混合 Logit 模型研究首尔至济州岛通道上出行者在多种交通方

作为观测样本, 一次为单独出行, 一次为结伴出行 (39 467 人, 78 934 次出行). 描述性统计如图 3 所示, 结果表明: 若一个人仅出行两次, 相对于单独出行而言结伴出行时会更多地选择高铁 ($t = 2.08, p < 0.05$) 和高等座位 ($t = 1.93, p = 0.05$).

式之间的选择行为, 指出除在途时间、出行花费和服务频次外, 交通方式的安全性也是影响旅客选择的重要因素.

已有的铁路文献对乘客出行的选择问题的研究, 往往建立在乘客一定会出行这一假设上, 忽略了乘客出行行为本身和具体客运产品选择之间的关系. 实际上, 影响乘客出行的因素往往也会影响到乘客客运产品的选择. 对这些因素的忽略可能带来潜在的内生性 (Endogeneity) 问题. 本文建议使用赫克曼二阶段模型 (Heckman two-stage model), 分步对乘客的出行行为和铁路产品的选择行为进行建模. 本文在第一阶段使用 logit 回归来探究乘客是否出行, 然后在第二阶段(基于乘客会出行的情况下)刻画结伴出行对于乘客的产品和服务选择的影响. 为了加强模型识别的准确度, 加

入了三个维度的固定效应,包括 1 317 854 个个体固定效应,3 个月的时间固定效应,以及 213 个出行次数的固定效应. 具体来说,本文建立以下的第一阶段模型

$$\text{logit}(\text{travel}_{it}) = \mu_i + \mu_t + \varepsilon_{it} \quad (1)$$

第一阶段使用了 Logit 模型分析乘客是否出行, i 和 t 分别代表第 i 个乘客和第 t 个月, travel_{it} 表示乘客是否出行(1 表示出行,0 表示不出行), logit 为逻辑连接函数(*logit link*)^②, μ_i 和 μ_t 分别表示乘客个体和时间的固定效应, ε_{it} 是一个随机误差项. 值得注意的是,此阶段无法加入出行次数 j 的固定效应,因为无论乘客 i 是第几次出行均代表出行行为. 此处估参的意义在于计算出逆米尔斯比率^③($\text{inverse_mills_ratio}_{it}$)^[43], 该比率类似于线性回归里的残差,是一种对可能影响乘客出行行为的潜在因素的衡量指标. 本文将逆米尔斯比率作为一个控制变量进入第二阶段的模型,有效地降低了第二阶段中可能出现的遗漏变量(*omitted variable*)问题:即影响乘客出行的潜在因素也有可能影响乘客对铁路产品的选择. 另外,如铁路数据的变量种类往往不如其他商业领域(如网络购物等)丰富,第一阶段的建模在变量种类不丰富的情况下(缺少同时随个体和时间变化的变量),依旧可以很好地衡量影响乘客客运产品选择的潜在因素. 接下来,为了探究核心变量结伴出行对乘客客运产品选择的影响,本文建立以下的第二阶段模型

$$\begin{aligned} \text{Product}_{ijt} = & \beta_1 \text{group}_{ijt} + \beta_2 \text{mobile_ticketing}_{ijt} + \\ & \beta_3 \text{inverse_mills_ratio}_{it} + \mu_i + \mu_t + \\ & \mu_j + \varepsilon_{ijt} \end{aligned} \quad (2)$$

第二阶段使用了普通线性回归^④,在稳健性检验中此处会被换成 Logit 回归和有序多项 Logit 回

归来检验结论的一致性. Product_{ijt} 为列车 train_{ijt} (1—3 表示从普通列车到高铁)或者座位 seat_{ijt} (1—4 表示从普速列车硬坐到高铁商务座). group_{ijt} 是本研究的核心变量,表示是否结伴出行(1 为结伴出行,0 为非结伴出行), $\text{mobile_ticketing}_{ijt}$ 表示移动端购票的购票渠道控制变量. 因为同时随个体和时间维度变化的控制变量较少,第一阶段模型计算的 $\text{inverse_mills_ratio}$ 能够有效地控制此类潜在遗漏变量带来的内生性问题. μ_i 、 μ_t 和 μ_j 分别表示乘客个体、时间和出行次数的固定效应. ε_{ijt} 是随机误差. 其中, β_1 为核心待估参数,表示结伴出行对列车等级和座位选择的影响. 为了探究性别与年龄因素对这一影响的调节作用,本文建立以下的第二阶段模型

$$\begin{aligned} \text{Product}_{ijt} = & \beta_1 \text{group}_{ijt} + \beta_4 \text{group}_{ijt} \times \\ & \text{gender}_i + \beta_5 \text{group}_{ijt} \times \text{age}_i + \\ & \beta_2 \text{mobile_ticketing}_{ijt} + \\ & \beta_3 \text{inverse_mills_ratio}_{it} + \\ & \mu_i + \mu_t + \mu_j + \varepsilon_{ijt} \end{aligned} \quad (3)$$

$\text{group}_{ijt} \times \text{gender}_i$ 表示结伴出行和是否为女性的交叉项,用以探究女性结伴出行对于铁路产品和服务选择的影响, $\text{group}_{ijt} \times \text{age}_i$ 表示结伴出行与年龄的交叉项,用以探究年长者结伴出行对于铁路产品和服务选择的影响, β_4 和 β_5 为核心待估参数,其他设定均与方程(2)相同^⑤.

4 实证分析结果

4.1 假设检验

表 3 中的(1)栏和表 4 中的(1)栏分别展示了方程(2)的估参结果. 结果显示,结伴出行显著正向影响乘客对高级列车($\beta = 0.0126$, $p < 0.001$)和

② 早期的赫克曼二阶段模型中的第一阶段使用的是 Probit 模型,因为计算逆米尔斯比率是基于第一阶段的误差项服从正态分布的假设,而 Probit 模型对于误差项的假设就是服从一个标准的正态分布. 但是 Dubin 和 Rivers^[41] 在其研究中提出赫克曼模型的中第一阶段的误差项不必完全依赖正态分布的假设,也可以是逻辑斯谛分布. 并且, Bushway 等^[42] 也提出 Logit 模型和 Probit 模型在因变量的概率不接近 0 或者 1 的情况下,两者差距是极小的. 只有在预测概率接近 0 或者 1 的情况下,Logit 模型才会显著不同于 Probit 模型. 本文在第一阶段选择了 Logit 模型,一方面因为第一阶段的预测概率的 80% 都集中于 0.3 到 0.4(并未接近 0 或者 1),所以 Logit 和 Probit 的差异是非常小. 同时也是为了计算的方便,并和过往的铁路研究中使用 Logit 模型的设置保持一致.

③ 逆米尔斯比率的计算方式是首先计算出 Logit 模型的线性预测结果,然后用这个结果的正态概率密度函数除以这个结果的概率密度函数的积分(累积分布)^[42].

④ 第二阶段使用普通线性回归一方面是因为赫克曼二阶段模型的第二阶段一般是使用普通线性回归^[44],同时也考虑到普通线性回归能够减小超过 130 万的固定效应(个体,时间,次数)所带来的计算量.

⑤ gender_i 和 age_i 的简单主效应(*simple main effect*)被个体的固定效应所吸收.

高等座位($\beta = 0.0130, p < 0.001$) 的选择, 假设 H1a 和 H1b 得到了证实. 同时可以发现移动端购票也对高级列车($\beta = 0.0309, p < 0.001$) 和高等座位($\beta = 0.0334, p < 0.001$) 的选择产生了显著正向的影响. 这个发现也可以为结伴出行的商业应用提供组合参考, 比如在鼓励人们结伴出行的同时改善和优化移动端购票体验, 使人们更多地使

用移动端购票, 提升消费者选择高品质产品的概率. 另外, 可以从表 3 中的(1) 栏和表 4 中的(1) 栏中发现逆米尔斯比率显著影响了出行者对于列车($\beta = -0.2085, p < 0.001$) 和座位($\beta = -0.2968, p < 0.001$) 的选择. 说明第二阶段确实存在潜在遗漏变量的问题, 赫克曼的两阶段矫正法利用第一阶段的逆米尔斯比率对估参偏差进行了有效的纠正.

表 3 结伴对列车选择的影响

Table 3 The impact of travelling companions on train choice

变量	(1)			(2)		
	均值	标准差	t 值	均值	标准差	t 值
是否结伴	0.0126***	0.0012	10.2210	-0.0021	0.0031	-0.6710
性别				NA	NA	NA
年龄				NA	NA	NA
结伴 × 女性				0.0114***	0.0017	6.4860
结伴 × 年龄				0.0003***	<0.0001	3.8650
移动端购票	0.0309***	0.0008	36.4560	0.0309***	0.0008	36.4680
逆米尔斯比率	-0.2085***	0.0409	-5.0890	-0.0205***	0.0410	-5.0050
个体固定效应	是			是		
时间固定效应	是			是		
次数固定效应	是			是		
观测数	3 939 487			3 939 487		

注: 1. ***、**、* 分别表示在 0.1%、1%、5% 的置信水平下显著。
 2. 截距项被个体差异的固定效应所吸收。
 3. NA 表示该效应被个体差异的固定效应所吸收。

表 4 结伴对座位选择的影响

Table 4 The impact of travelling companions on seat choice

变量	(1)			(2)		
	均值	标准差	t 值	均值	标准差	t 值
是否结伴	0.0130***	0.0012	10.5830	-0.00397	0.0031	-1.2190
性别				NA	NA	NA
年龄				NA	NA	NA
结伴 × 女性				0.0128***	0.0018	7.2910
结伴 × 年龄				0.0004***	<0.0001	4.4750
移动端购票	0.0334***	0.0008	39.4870	0.0335***	0.0008	39.5000
逆米尔斯比率	-0.2968***	0.0409	-7.2490	-0.2930***	0.0410	-7.1570
个体固定效应	是			是		
时间固定效应	是			是		
次数固定效应	是			是		
观测数	3 939 487			3 939 487		

注: 1. ***、**、* 分别表示在 0.1%、1%、5% 的置信水平下显著。
 2. 截距项被个体差异的固定效应所吸收。
 3. NA 表示该效应被个体差异的固定效应所吸收。

从表3中的(2)栏和表4中的(2)栏可以看出,在分别加入了结伴出行与性别和年龄的交叉项后,交叉项显著正向影响对高级列车和高等座位的选择.首先,从结伴出行与性别的交叉项结果可以看出,女性结伴出行显著正向影响出行者选择高级列车($\beta=0.0114$, $p<0.001$)和高等座位($\beta=0.0128$, $p<0.001$),假设H2a和H2b得到证实.从结伴出行与年龄的交叉项结果可以看出,年长者结伴出行显著正向影响出行者选择高级列车($\beta=0.0003$, $p<0.001$)和高等座位($\beta=0.0004$, $p<0.001$),假设H3a和H3b得到证实.除此之外,可以发现移动端购票对列车($\beta=0.0309$, $p<0.001$)和座位($\beta=0.0335$, $p<0.001$)选择依旧有显著正向的影响,逆米尔斯比率对于第二阶段的样本偏差的纠正作用依然显著.

需要说明的是,在主分析中并未使用乘客居住地是否为城市、平均花费和平均里程这三个对消费者选择有潜在影响的变量,因为这些个体差异性变量的效应已经被个体的固定效应所吸收.同时,列车出发时间、到达时间可能也会影响到旅客出行中的铁路客运产品选择行为.在接下来的稳健性分析部分中,在适当的模型中将会使用这些变量.

4.2 稳健性检验

本文从四个方面对研究结论进行了稳健性分析.

4.2.1 替代性的操作化

首先对因变量采取另一种量度方式,即检验本文的发现在替代性的操作化(alternative operationalization)下是否成立.在主分析中列车等级被分为1到3(普通列车、动车、高铁分别取值为1、2和3),将原本的2(动车)和3(高铁)合并为2(较高等级列车),而1维持不变(普通列车).主分析中列车座位等级被分为1到4(普速列车硬卧取值为1,动车二等座和普速列车硬卧取值为2,高铁二等座、动车一等座和普速列车软卧取值为3,高铁商务座、高铁一等座取值为4),将原本的座位等级2和3合并为2(中等座位),4变为3(高等座位),1维持不变(普通座位).在模型的其他设置不变的情况下,重新对模型进行估参.表5和表6展示了对应的分析结果,新结果与主分析结果在方向与显著性上保持一致:结伴出行显著正向影响对高级列车和高等座位的选择;结伴和女性交叉项显著正向影响对高级列车和高等座位的选择;结伴和年龄交叉项显著正向影响对高级列车和高等座位的选择.

表5 结伴对列车选择的影响(等级1~等级2的列车)

Table 5 The impact of travelling companions on train choice (grade1 and grade2)

变量	(1)			(2)		
	均值	标准差	t 值	均值	标准差	t 值
是否结伴	0.0017***	0.0005	3.3190	-0.0066***	0.0014	-4.8320
性别				NA	NA	NA
年龄				NA	NA	NA
结伴×女性				0.0013	0.0007	1.7440
结伴×年龄				0.0002***	0.0002	6.6180
移动端购票	0.0008*	0.0004	2.2630	0.0008*	0.0004	2.2650
逆米尔斯比率	-0.1262***	0.0175	-7.1880	-0.1271***	0.0175	-7.2400
个体固定效应	是			是		
时间固定效应	是			是		
次数固定效应	是			是		
观测数	3 939 487			3 939 487		

注:1.***、**、* 分别表示在0.1%、1%、5%的置信水平下显著.

2.截距项被个体差异的固定效应所吸收.

3.NA表示该效应被个体差异的固定效应所吸收.

表6 结伴对座位选择的影响(等级1~等级3的座位)

Table 6 The impact of travelling companions on seat choice (grade1 ~ grade3)

变量	(1)			(2)		
	均值	标准差	t 值	均值	标准差	t 值
是否结伴	0.003 5***	0.003 5	5.375 0	-0.009 2***	0.001 7	-5.556 0
性别				NA	NA	NA
年龄				NA	NA	NA
结伴×女性				0.002 5**	0.000 9	2.749 0
结伴×年龄				2.749 0***	2.749 0	8.136 0
移动端购票	0.004 0***	0.000 4	9.034 0	0.004 0***	0.000 4	9.038 0
逆米尔斯比率	-0.189 1***	0.021 5	-8.808 0	-0.190 2***	0.021 5	-8.860 0
个体固定效应	是			是		
时间固定效应	是			是		
次数固定效应	是			是		
观测数	3 939 487			3 939 487		

注: 1. ***、**、* 分别表示在 0.1%、1%、5% 的置信水平下显著。

2. 截距项被个体差异的固定效应所吸收。

3. NA 表示该效应被个体差异的固定效应所吸收。

4.2.2 有序多分类回归

除了改变对因变量的测量方式, 本文还尝试了改变模型设置. 主分析将列车和座位当做连续变量, 使用了普通线性回归. 在稳健性检验中, 将列车和座位当做有序变量和分类变量, 分别使用有序多分类回归(ordinal regression, OR) 和多项 Logit 回归进行估参. 但是如果在这两种回归中依旧加入个体、时间和次数的固定效应, 会产生上百万的虚拟变量, 计算量非常大. 由于有序多分类回归或多项 Logit 回归中添加随机效应是相对合理且显著降低计算量的手段, 所以第二种检验首先将主分析模型里所有的固定效应替换为随机效应, 检验结论的稳健性. 如果将固定效应换成随机效应后, 结论仍然稳健, 此时则将普通的线性回归换成有序多分类回归. 主分析模型将列车和座位视作数值型连续变量, 采用普通线性回归方法进行参数估计可能存在一定的不合理性. 事实上, 不同等级列车(或座位)的差距可能是有较大的不同, 从速度、票价花费、舒适度等多个客观因素将

列车和座位划分等级, 不同分类之间仅保持相对顺序(而非等距)是一种相对保守的做法. 另一方面, 由于带有三个随机效应的多项 Logit 回归计算量过大, 本文继而在保持固定效应的设定下, 将普通的线性回归换成多个 Logit 模型, 把列车和座位等级当成分类变量来验证结论的稳健性, 对不同的列车或者座位进行两两比较.

1) 随机效应模型

需要说明的是, 此时的模型中可以加入在数据说明部分提到的三个重要的控制变量: 平均花费、平均里程和是否城市. 主分析模型中的三个固定效应(个体、时间、次数)的控制很严格, 所以在去除固定效应后应该加入控制变量对此缺陷进行弥补. 此外, 这一改变也有助于看到这三个在文献中提到的显著影响变量在本文数据中的表现. 由于加入的三个控制变量(平均花费、平均里程、是否城市)中含有缺失值, 模型剔除缺失值后最终剩下的观测数为 2 721 052. 随机效应模型的分析结果如表 7 与表 8 所示.

表 7 结伴对列车选择的影响(随机效应)

Table 7 The impact of travelling companions on train choice (random effect)

变量	(1)			(2)		
	均值	标准差	t 值	均值	标准差	t 值
截距	2.440 0***	0.011 6	211.390 0	2.450 0***	0.012 2	200.800 0
是否结伴	0.011 4***	0.001 2	9.330 0	-0.006 1*	0.003 1	-1.968 0
性别				-0.058 5***	0.001 9	-30.104 0
年龄				0.000 8***	<0.000 1	9.508 0
结伴×女性				0.012 1***	0.001 8	6.855 0
结伴×年龄				0.000 4***	<0.000 1	4.965 0
移动端购票	0.041 8***	0.000 8	49.760 0	0.041 6***	0.000 84	49.472 0
逆米尔斯比率	-0.016 0***	0.001 6	-10.160 0	-0.013 0***	0.001 6	-8.288 0
平均花费	0.000 3***	<0.000 1	216.470 0	0.000 3***	<0.000 1	208.019 0
平均里程	-0.000 6***	<0.000 1	-206.450 0	-0.000 6***	<0.000 1	-202.374 0
是否在城市	0.113 0***	0.001 3	86.680 0	0.109 0***	0.001 3	82.231 0
个体随机效应	是			是		
时间随机效应	是			是		
次数随机效应	是			是		
观测数	2 721 052			2 721 052		

注: ***、**、* 分别表示在 0.1%、1%、5% 的置信水平下显著。

表 8 结伴对座位选择的影响(随机效应)

Table 8 The impact of travelling companions on seat choice (random effect)

变量	(1)			(2)		
	均值	标准差	t 值	均值	标准差	t 值
截距	2.460 0***	0.015 5	158.181 0	2.430 0***	0.016 1	150.965 0
是否结伴	0.013 5***	0.001 2	11.041 0	-0.001 0	0.003 1	-0.309 0
性别				-0.077 5***	0.002 0	-39.546 0
年龄				0.001 9***	<0.000 1	21.806 0
结伴×女性				0.013 0***	0.001 8	7.375 0
结伴×年龄				0.000 3***	<0.000 1	3.579 0
移动端购票	0.044 1***	0.000 8	52.304 0	0.043 7***	0.000 8	51.878 0
逆米尔斯比率	-0.008 8***	0.001 6	-5.492 0	-0.004 7**	0.001 6	-2.965 0
平均花费	0.000 4***	<0.000 1	272.027 0	0.000 4***	<0.000 1	259.568 0
平均里程	-0.000 8***	<0.000 1	-251.085 0	-0.000 8***	<0.000 1	-244.848 0
是否在城市	0.133 0***	0.001 3	99.752 0	0.124 0***	0.001 3	92.438 0
个体随机效应	是			是		
时间随机效应	是			是		
次数随机效应	是			是		
观测数	2 721 052			2 721 052		

注: ***、**、* 分别表示在 0.1%、1%、5% 的置信水平下显著。

从表7和表8可以看出,当把固定效应换成随机效应并加入三个控制变量后,如表7(1)栏和表8(1)栏所示,结伴出行依旧显著正向显著对高级列车($\beta=0.0114, p<0.001$)和高等座位($\beta=0.0135, p<0.001$)的选择.从表7(1)和表8(1)可以看出,在加入结伴出行与性别、年龄的交叉项后,交叉项依旧正向显著:结伴与性别交叉项显著正向影响高级列车($\beta=0.0121, p<0.001$)和高等座位($\beta=0.0130, p<0.001$)的选择;结伴与年龄交叉项显著正向影响对高级列车($\beta=0.0004, p<0.001$)和高等座位($\beta=0.0003, p<0.001$)的选择.

基于上述的替换的可行,接下来将分别使用带有随机效应的有序多分类回归和带有固定效应

的Logit模型来检验主分析结论的稳健性.

2) 随机效应下有序多分类回归

表9和表10展示了随机效应下有序多分类回归分析的结果.表9(1)栏和表10(1)栏显示,在未加入结伴出行与性别和年龄的交叉时,结伴出行对于高级列车($\beta=0.0041^a$)^⑥和高等座位($\beta=0.0037^a$)的影响是正向的.表9(2)栏表明女性($\beta=0.7251^a$)或长者($\beta=0.0274^a$)结伴更倾向选择高级列车.表10(2)栏中,虽然女性对座位等级的影响的后验可信区间包含了0,但是整个区间更加偏向为正,且该参数也为正向($\beta=0.0469$),与主分析结果方向一致.结伴与年龄交叉项($\beta=0.0074^a$)正向影响出行者对高等座位的选择,与主分析结果保持一致.

表9 结伴对列车选择的影响(随机效应+有序多分类回归)

Table 9 The impact of travelling companions on train choice (random effect + OR)

变量	(1)				(2)			
	均值	标准差	置信区间(95%)		均值	标准差	置信区间(95%)	
截距[1]	3.5472	2.0448	-0.3732	7.6595	21.6664 ^a	1.3759	18.9993	24.5223
截距[2]	12.8209 ^a	2.0434	8.9104	17.0179	23.3877 ^a	1.3756	20.7155	26.2479
是否结伴	0.0041 ^a	0.0011	0.0019	0.0061	-0.0831 ^a	0.0125	-0.108	-0.0592
性别					-0.8847 ^a	0.0487	-0.9814	-0.7914
年龄					-0.0461 ^a	0.0028	-0.0515	-0.0403
结伴×女性					0.7251 ^a	0.0328	0.6643	0.7884
结伴×年龄					0.0274 ^a	0.002	0.0233	0.0313
移动端购票	1.2021 ^a	0.0076	1.1879	1.2171	1.1169 ^a	0.0322	1.0524	1.1779
逆米尔斯比率	5.2539 ^a	0.0136	5.2275	5.2794	2.4214 ^a	0.0159	2.39	2.4523
平均花费(每次)	0.0199 ^a	0.0018	0.0165	0.0235	0.018 ^a	0.0011	0.0159	0.0203
平均里程(每次)	-0.0486 ^a	0.0006	-0.0499	-0.0474	0.009 ^a	0.0009	0.0073	0.0108
是否在城市	6.0302 ^a	0.0207	5.9877	6.0689	1.2893 ^a	0.0131	1.2623	1.3152
个体随机效应	是				是			
时间随机效应	是				是			
次数随机效应	是				是			
观测数	2 721 052				2 721 052			

注: 1. 因为传统的点估计(point estimation)方法难以估计海量数据下的3维随机项(个体、时间、次数),该分析使用了变分贝叶斯估参.

2. “a”表示该参数后验的均值在95%的可信区间内不包括0.

⑥ “a”表示该参数后验的均值在95%的可信区间内不包括0.

表 10 结伴对座位选择的影响(随机效应+有序多分类回归)
Table 10 The impact of travelling companions on seat choice (random effect + OR)

变量	(1)				(2)			
	均值	标准差	置信区间(95%)		均值	标准差	置信区间(95%)	
截距 [1]	5.562 5 ^a	0.729 3	4.123 3	7.055 0	16.046 3 ^a	0.370 3	15.363 2	16.840 1
截距 [2]	37.054 2 ^a	0.738 8	35.694 1	38.574 4	225.108 0 ^a	0.371 2	224.418 0	225.883 2
是否结伴	0.003 7 ^a	0.001	0.001 7	0.005 8	0.574 0 ^a	0.032 8	0.510 0	0.639 4
性别					-3.848 0 ^a	0.007 7	-3.862 5	-3.832 6
年龄					0.201 6 ^a	0.002 6	0.196 7	0.206 7
结伴 × 女性					0.046 9	0.040 5	-0.029 6	0.125 1
结伴 × 年龄					0.007 4 ^a	0.002 4	0.002 4	0.012
移动端购票	1.124 0 ^a	0.024 4	1.077 5	1.171 8	0.684 7 ^a	0.023 8	0.637 8	0.732 4
逆米尔斯比率	0.803 8 ^a	0.02	0.765 6	0.842 7	7.147 4 ^a	0.009 5	7.128 7	7.166 0
平均花费(每次)	0.022 3 ^a	0.000 6	0.021 2	0.023 6	0.044 2 ^a	0.000 2	0.043 8	0.044 5
平均里程(每次)	-0.016 3 ^a	0.000 4	-0.017 1	-0.015 4	-0.080 6 ^a	0.000 7	-0.082 0	-0.079 2
是否在城市	3.878 6 ^a	0.027 4	3.823 9	3.932 3	7.471 8 ^a	0.046 2	7.376 7	7.566 1
个体随机效应	是				是			
时间随机效应	是				是			
次数随机效应	是				是			
观测数	2 721 052				2 721 052			

注: 1. 因为传统的点估计(point estimation)方法难以估计海量数据下的3维随机项(个体、时间、次数),该分析使用了变分贝叶斯估参。
2. “a”表示该参数后验的均值在95%的可信区间内不包括于0。
3. 为降低计算量,基于变换座位等级分类方法不影响估参结果,此处使用三个等级的座位分类作为因变量(将主分析的因变量从1-4变为1-3)。

3) 固定效应下的多个二项 Logit 模型回归

本节的稳健性检验将因变量视为多个二分变量,采用二项 Logit 模型进行两两对比检验.对于列车的选择,因为在列车的三个等级中,高铁的数据量在总数据量中超过七成,所以这里将分别比较3(高铁)和2(动车)、3(高铁)和1(普通列车).表11和表12分别展示了这两组二项 Logit 回归的结果,从表11(1)栏和表12(1)栏可以看出,在两种情况下结伴出行显著正向影响对高级列车—高铁($\beta = 0.116 9, p < 0.001$)和($\beta =$

$0.143 7, p < 0.001$)的选择.从表11(2)栏可以看出,高铁和动车相比,女性($\beta = 0.057 1, p < 0.001$)结伴和长者($\beta = 0.000 9$)结伴正向影响对高铁的选择.虽然此处结伴出行和年龄的交叉项系数不显著,但其方向保持正向,与主分析结果比较一致.表12(2)栏可以看出,高铁与普通列车相比,女性($\beta = 0.116 7, p < 0.001$)结伴和长者($\beta = 0.005 6, p < 0.001$)结伴显著正向对高铁的选择.所以在多个二项 Logit 模型下,结伴出行对于列车选择的分析结果依旧和主分析保持一致.

表 11 结伴对列车选择的影响(高铁与动车)

Table 11 The impact of travelling companions on train choice (G-train and D-train)

变量	(1)			(2)		
	均值	标准差	z 值	均值	标准差	z 值
是否结伴	0.116 9***	0.010 4	11.255 8	0.055 7*	0.027 1	2.053 1
性别				NA	NA	NA
年龄				NA	NA	NA
结伴 × 女性				0.057 1***	0.014 7	3.873 9
结伴 × 年龄				0.000 9	0.000 7	1.351 4
移动端购票	0.295 6***	0.007 3	40.268 5	0.298 4***	0.007 3	40.611 8
逆米尔斯比率	-1.088 1**	0.331 2	-3.284 9	-0.346 1***	0.047 8	-7.236 3
个体固定效应	是			是		
时间固定效应	是			是		
次数固定效应	是			是		
观测数	3 615 960			3 615 960		

注: 1. ***、**、* 分别表示在 0.1%、1%、5% 的置信水平下显著。

2. 截距项被个体差异的固定效应所吸收。

3. NA 表示该效应被个体差异的固定效应所吸收。

表 12 结伴对列车选择的影响(高铁与普通列车)

Table 12 The impact of travelling companions on train choice (G-train and K-train)

变量	(1)			(2)		
	均值	标准差	z 值	均值	标准差	z 值
是否结伴	0.143 7***	0.019 0	7.546 4	-0.084 5	0.048 2	-1.751
性别				NA	NA	NA
年龄				NA	NA	NA
结伴 × 女性				0.116 7***	0.027 4	4.264 5
结伴 × 年龄				0.005 6***	0.001 2	4.467 4
移动端购票	0.142 1***	0.013 1	10.819 1	0.142 2***	0.013 1	10.827 4
逆米尔斯比率	-4.454 6***	0.602 5	-7.393 8	-4.421 2***	0.602 8	-7.334 2
个体固定效应	是			是		
时间固定效应	是			是		
次数固定效应	是			是		
观测数	3 356 524			3 356 524		

注: 1. ***、**、* 分别表示在 0.1%、1%、5% 的置信水平下显著。

2. 截距项被个体差异的固定效应所吸收。

3. NA 表示该效应被个体差异的固定效应所吸收。

对于座位等级的选择,为降低计算量和比较组数,将 2(动车二等座和普速列车硬卧)和 3(高铁二等座、动车一等座和普速列车软卧)合并为 2(中等座位),4(高铁商务座、高铁一等座)变成了 3(高等座位),1(普速列车硬坐)维持不变(普通座位)^⑦。由于变换后的等级 2 的座位占比超过八

成,故这里只比较等级为 3 和 2 的座位、等级 2 和 1 的座位,来检验结伴出行对于座位选择的作用。

表 13 和表 14 分别展示了这两组 Logit 回归的结果。从表 13(1) 栏和表 14(1) 栏可以看出,结伴出行者在两种情况下都显著正向影响对高等座位的选择:在三等座位和二等座位的选择中显著正向

⑦ 第一个稳健性检验表明该变换不影响主分析结果的方向和显著性。

影响对三等座位($\beta=0.0530$, $p<0.01$)的选择,在二等座位和一等座位的选择中显著影响对二等座位($\beta=0.1374$, $p<0.001$)的选择.从表13(2)栏可以看出,在三等座位和二等座位之间选择时,女性($\beta=0.0763$, $p<0.01$)结伴和长者($\beta=0.0034$, $p<0.01$)结伴显著正向影响三等座位(更

高等的座位)的选择.从表14(2)栏可以看出,在二等座位和一等座位之间选择时,女性($\beta=0.0655$, $p<0.01$)结伴和与长者($\beta=0.0081$, $p<0.001$)结伴显著正向影响二等座位(更高等的座位)的选择.所以在Logit模型的设置下,结伴出行对于座位选择的分析结果和主分析依旧保持高度一致.

表 13 结伴对座位选择的影响(三等座位与二等座位)

Table 13 The impact of travelling companions on seat choice (grade 3 and grade 2)

变量	(1)			(2)		
	均值	标准差	z 值	均值	标准差	z 值
是否结伴	0.0530**	0.0173	3.0677	-0.0943*	0.0477	-1.9768
性别				NA	NA	NA
年龄				NA	NA	NA
结伴×女性				0.0763**	0.0259	2.9484
结伴×年龄				0.0034**	0.0012	2.88
移动端购票	0.0615***	0.0124	4.9431	0.0616***	0.0124	4.9513
逆米尔斯比率	-3.2073***	0.5593	-5.7347	-3.1590***	0.5596	-5.6447
个体固定效应	是			是		
时间固定效应	是			是		
次数固定效应	是			是		
观测数	3 676 914			3 676 914		

注: 1. **、*、* 分别表示在 0.1%、1%、5% 的置信水平下显著.

2. 截距项被个体差异的固定效应所吸收.

3. NA 表示该效应被个体差异的固定效应所吸收.

4. 为降低计算量,基于变换座位等级分类方法不影响估参结果,此处使用三个等级的座位分类作为因变量(将主分析的因变量从 1-4 变为 1-3).

表 14 结伴对座位选择的影响(二等座位与一等座位)

Table 14 The impact of travelling companions on seat choice (grade 2 and grade 1)

变量	(1)			(2)		
	均值	标准差	z 值	均值	标准差	z 值
是否结伴	0.1374***	0.0173	7.9454	-0.1538***	0.0425	-3.618
性别				NA	NA	NA
年龄				NA	NA	NA
结伴×女性				0.0655**	0.0246	2.6667
结伴×年龄				0.0081***	0.0011	7.3304
移动端购票	0.0744***	0.0117	6.3368	0.0744***	0.0117	6.3399
逆米尔斯比率	-5.2056***	0.5547	-9.3843	-5.2550***	0.5552	-9.4649
个体固定效应	是			是		
时间固定效应	是			是		
次数固定效应	是			是		
观测数	3 722 407			3 722 407		

注: 1. **、*、* 分别表示在 0.1%、1%、5% 的置信水平下显著.

2. 截距项被个体差异的固定效应所吸收.

3. NA 表示该效应被个体差异的固定效应所吸收.

4. 为降低计算量,基于变换座位等级分类方法不影响估参结果,此处使用三个等级的座位分类作为因变量(将主分析的因变量从 1-4 变为 1-3).

4.2.3 考虑列车发车/到达时间与是否结伴
交叉影响的稳健性分析

不同等级列车发车和到达时间上存在显著差异,就发车时间而言,普速列车(K字头快速列车和Z字头直达列车)发车时间早于D字头动车发车时间($t = -346.36, p < 0.001$);D字头动车发车时间早于G字头高铁($t = -301.65, p < 0.001$).就发车时间的均值来讲,普速列车(11时左右)最早,其次是D字头动车(14时左右),最后是G字头高铁(15时左右).就到达时间而言,普速列车到达时间稍早于D字头动车到达时间($t = -9.06, p < 0.001$);D字头动车到达时间早于G字头高铁($t = -411.65, p < 0.001$).就到达时间的均值来讲,普速列车和D字头动车都是15时左右,G字头高铁为18时左右.不同客运产品发车/到达时间段与是否结伴对旅客出行选择可能存在交叉影响,从而影响主效应结果.为了验证这一假设,本文根据当期列车时刻表信息提取了列车发车/

到达时间,将列车发车/到达时间归类为四个时间段:上午、中午、下午和夜晚^⑧.接着在模型中分别加入四个时间段变量与是否结伴的交叉项来检验列车时间与是否结伴对于铁路客运产品选择的交叉影响.就发车时间来讲,不同时间段与是否结伴对于铁路产品选择的交叉影响是不同的.从表15可以看出:旅客选择上午发车的铁路产品时,交叉项显著正向影响选择高级列车($\beta = 0.0152, p < 0.001$)和高等座位($\beta = 0.0144, p < 0.001$);旅客选择中午发车的铁路产品时,交叉项显著正向影响选择高级列车($\beta = 0.0084, p < 0.001$)和高等座位($\beta = 0.0154, p < 0.001$);旅客选择下午发车的铁路产品时,交叉项显著负向影响选择高级列车($\beta = -0.0064, p < 0.001$)和高等座位($\beta = -0.0040, p < 0.001$);旅客选择夜晚发车的铁路产品时,交叉项显著负向影响选择高级列车($\beta = -0.0159, p < 0.001$)和高等座位($\beta = -0.0263, p < 0.001$).

表 15 发车时间与结伴对于铁路产品选择的交叉影响

Table 15 The cross impact of start time and travelling companions on train and seat choice

变量	列车选择				座位选择			
	上午发车	中午发车	下午发车	夜晚发车	上午发车	中午发车	下午发车	夜晚发车
是否结伴	0.009 8*** (0.001 3)	0.011 9*** (0.001 3)	0.016 7*** (0.001 4)	0.015 5*** (0.001 3)	0.010 2*** (0.001 3)	0.011 2*** (0.001 3)	0.016 1*** (0.001 4)	0.017 5*** (0.001 3)
发车时间	0.017 3*** (0.001 4)	0.066 5*** (0.001 7)	-0.044 1*** (0.001 2)	-0.010 7*** (0.001 7)	0.012 5*** (0.001 4)	0.060 9*** (0.001 7)	-0.045 8*** (0.001 2)	0.005 2** (0.001 7)
是否结伴 × 发车时间	0.015 2*** (0.001 5)	0.008 4*** (0.001 8)	-0.006 4*** (0.001 3)	-0.015 9*** (0.001 8)	0.014 4*** (0.001 5)	0.015 4*** (0.001 8)	-0.004 0** (0.001 3)	-0.026 3*** (0.001 8)
移动端购票	0.031 2*** (0.000 8)	0.031 0*** (0.000 8)	0.031 0*** (0.000 8)	0.031 2*** (0.000 8)	0.033 7*** (0.000 8)	0.033 6*** (0.000 8)	0.033 5*** (0.000 8)	0.033 6*** (0.000 8)
逆米尔斯比率	-0.242 7*** (0.041 0)	-0.226 8*** (0.040 9)	-0.255 6*** (0.040 9)	-0.218 8*** (0.041 0)	-0.324 2*** (-0.324 2)	-0.314 3*** (0.040 9)	-0.343 9*** (0.040 9)	-0.301 1*** (0.040 9)
个体固定效应	是	是	是	是	是	是	是	是
时间固定效应	是	是	是	是	是	是	是	是
次数固定效应	是	是	是	是	是	是	是	是
出发时间	是	是	是	是	是	是	是	是
观测数	3 939 487	3 939 487	3 939 487	3 939 487	3 939 487	3 939 487	3 939 487	3 939 487

注: 1. ***, **, * 分别表示在 0.1%、1%、5% 的置信水平下显著。

2. 截距项被个体差异的固定效应所吸收。

3. 均值下方的括号内为标准差。

⑧ 若发车/到达时间为 6 点至 10 点则该时间点被归为上午;若发车/到达时间为 11 点至 13 点则该时间点被归为中午;若发车/到达时间为 14 点至 18 点则该时间点被归为下午;其余时间点被归为夜晚。

对于列车的到达时间来讲,从表 16 可以看出:旅客选择上午到达的铁路客运产品时,交叉项显著负向影响对高级列车($\beta = -0.0316, p < 0.001$)和高等座位($\beta = -0.0418, p < 0.001$)的选择;旅客选择中午到达的铁路产品时,交叉项显著正向影响对高级列车($\beta = 0.0217, p < 0.001$)和高等座位($\beta = 0.0260, p < 0.001$)的选择;旅客选择下午到达的铁路产品时,交叉项显著

负向影响对高级列车($\beta = -0.0078, p < 0.001$)和高等座位($\beta = -0.0031, p < 0.05$)的选择;旅客在选择夜晚到达的铁路产品时,交叉项正向影响对高级列车($\beta = 0.0032, p < 0.05$)和高等座位($\beta = 0.0001, p = 0.9320$)的选择,虽然夜晚到达与是否结伴对座位选择的交叉影响不显著,但是其影响方向与对列车选择的影响是一致的。

表 16 到达时间与结伴对于铁路产品选择的交叉影响

Table 16 The cross impact of arrival time and companions on train and seat choice

变量	列车选择				座位选择			
	上午到达	中午到达	下午到达	夜晚到达	上午到达	中午到达	下午到达	夜晚到达
是否结伴	0.015 0*** (0.001 3)	0.009 8*** (0.001 3)	0.013 5*** (0.001 3)	0.015 9*** (0.001 3)	0.016 6*** (0.001 3)	0.009 3*** (0.001 3)	0.012 5*** (0.001 3)	0.017 1*** (0.001 3)
到达时间	-0.040 8*** (0.001 9)	0.066 0*** (0.001 5)	0.160 5*** (0.001 3)	-0.196 7*** (0.001 3)	-0.021 8*** (0.001 9)	0.057 0*** (0.001 5)	0.143 1*** (0.001 3)	-0.180 9*** (0.001 3)
是否结伴 × 到达时间	-0.031 6*** (0.002 1)	0.021 7*** (0.001 6)	-0.007 8*** (0.001 4)	0.003 2* (0.001 4)	-0.041 8*** (0.002 1)	0.026 0*** (0.001 6)	-0.003 1* (0.001 4)	0.000 1 (0.001 4)
移动端购票	0.030 3*** (0.000 8)	0.030 6*** (0.000 8)	0.031 6*** (0.031 6)	0.032 7*** (0.000 8)	0.033 0*** (0.000 8)	0.033 2*** (0.000 8)	0.034 1*** (0.000 8)	0.035 2*** (0.000 8)
逆米尔斯比率	-0.176 5*** (0.040 9)	-0.282 0*** (0.040 8)	-0.162 9*** (0.040 4)	-0.431 2*** (0.040 1)	-0.271 5*** (0.040 9)	-0.364 6*** (0.040 8)	-0.254 5*** (0.040 5)	-0.503 1*** (0.040 2)
个体固定效应	是	是	是	是	是	是	是	是
时间固定效应	是	是	是	是	是	是	是	是
次数固定效应	是	是	是	是	是	是	是	是
出发时间	是	是	是	是	是	是	是	是
观测数	3 939 487	3 939 487	3 939 487	3 939 487	3 939 487	3 939 487	3 939 487	3 939 487

注: 1. ***、**、* 分别表示在 0.1%、1%、5% 的置信水平下显著。

2. 截距项被个体差异的固定效应所吸收。

3. 均值下方的括号内为标准差。

5 结束语

本文从结伴出行对于铁路客运产品选择影响的视角实证分析结伴对消费升级的促进作用. 本文利用成渝直达旅客的出行记录作为观测样本, 除了抽取现有文献中提到的潜在显著因素, 如人口统计学特征、平均花费、列车出发达到时刻等, 首次提取了一个重要的信息: 消费者结伴出行. 为了准确地研究结伴出行对于铁路产品(不同等级列车和座位)的影响, 首次在铁路研究领域采用了赫克曼二阶段模型, 并严格控制了 130 多万个个体差异固定效应、3 个时间固定效应和 213 个出行次数的固定效应. 主分析和一系列的稳健性

分析均显示, 人们结伴出行背后体现的是对于更加舒适的生活和中高端消费品的追求, 即结伴有助于推动消费升级. 具体来说: 结伴出行显著正向影响出行者对高级列车和高等座位的选择(中高端的出行产品). 同时, 结伴出行的影响受到性别与年龄的调节作用, 女性结伴出行更倾向于选择高级列车和高等座位; 年长者结伴更倾向于选择高级列车和高等座位.

本文的发现对包括大数据、铁路管理、商业运营和公共管理在内的多个领域有一定的启发性意义. (1) 首先, 本文的分析基于海量数据, 充分证实大数据对于理解消费者行为所起到的突出性作用. 如今, 消费者产生的数据具有实时、集中、多样、细微、海量的特点, 这是一个蕴含丰富商业价

值的资源,对竞争日趋激烈的市场来说,大数据能够让企业更加了解消费者的偏好和动机,更好更快地满足消费者的需求。(2)同时,本文关于结伴出行对于铁路产品选择的影响发现有利于铁路客运管理部门更好地理解顾客的需求,从而帮助铁路部门更好地按需服务和精准营销。例如,对于结伴出行的顾客,特别是女性或长者出行的顾客,铁路部门可以提供特定的配套优质产品和服务,开通相关的绿色通道,使其更加享受结伴出行的乐趣,在提高顾客满意度的同时也增强了铁路产品的竞争力。从更广的商业角度来说,本文从铁路产品这个视角出发真正想看到的是结伴对消费升级的推动作用。当前我国社会主要矛盾已经转化为人民日益增长的美好生活需要和不平衡不充分的发展之间的矛盾,消费升级作为国家当前的发展战略,能够带动产业结构迈向中高端,有助于缓解这一矛盾的发展。如果结伴出行正向显著影响出行者对于代表中高端产品的高级列车和高等座位的选择,则有理由相信结伴行为本身会对广阔的商业领域,如航空、电影、餐饮、购物等带来消费结构的提升,促进消费升级。企业的管理者以各种形式推动人们结伴消费,比如会员俱乐部、结伴定制化服务与结伴回馈,可以显著刺激人们对于高品质服务和产品的消费,进而增大企业的利润。对于消费者来说,这些结伴定制化产品和服务也能更好地满足消费者的心理需求,提高消费者生活品质,实现企业消费者的共赢^⑨。本文也建议使用二阶段模型来纠正以铁路数据为代表的量级巨大但信息维度相对单薄的数据潜在遗漏变量的问题。第一阶段计算出的逆米尔斯比率作为控制变量参与第二个阶段的主模型估计,能在一定程度上弥补第二阶段样本偏差的缺陷,进而降低传统铁路文献中直接研究第二阶段所带来的对于隐含假设的依赖(如消费者一定会出行)。配合有力的

海量个体固定效应(如本文超过130万个个体固定效应),能在铁路数据信息有限的情况下,尽可能准确地对模型和核心变量进行估参。(3)另外,从公共管理的角度来讲,本文从动态的角度(乘客连续三个月的出行记录)去研究人们的流动性(空间变化),并结合出行者之间的关联(结伴行为),用一个综合的视角,即时间-空间-关联的角度,来审视消费者行为背后的动机和商业价值,政府和管理者可以更加快速、准确、全面地掌握当下的市场趋势和需求,更好更快地满足消费者日益丰富的需求。

本文研究也有一定的局限性。第一,文中铁路数据只用到成渝的铁路数据,此处假设了该数据可以代表广泛的产品和服务。虽然成渝铁路的发展比较成熟和完善,具有代表性,但是如果可以获得更加一般性的产品消费数据或者更具代表性的区域铁路数据,那么结论的普适性会更有基础。在未来的研究中,一方面可以将中国的铁路数据和外国的铁路数据进行对比,在发现普遍性趋势的情况下探究地域差异性。另一方面,还可以结合铁路数据和其他数据,比如出行者对于移动APP的使用情况、出行者对于航空产品的选择等等,通过融合数据对更多的消费者行为进行关联。第二,此次数据只涉及了移动端和互联网订票的数据。因为对于其他的购票方式的数据,比如传统窗口的购票数据,无法准确记录结伴出行行为的发生,所以本文只使用了移动端和互联网订票的数据。最后,本文研究欲透过铁路数据的分析折射出结伴行为对消费升级的推动作用。在未来的研究中,可以探究在其他的领域里结伴行为对高品质产品和服务选择的影响。本文抛砖引玉,希望不仅能阐释结伴行为与消费升级间的关系,更能激发未来的一系列以消费者结伴行为为主题的研究。

参 考 文 献:

- [1] Social Support Measurement and Intervention: A Guide for Health and Social Scientists [M]. New York: Oxford University Press, 2000.
- [2] O'Leary-Kelly A M, Martocchio J J, Frink D D. A review of the influence of group goals on group performance [J]. Acade-

^⑨ 以铁路为载体研究结伴对消费升级的促进作用可能具有一定的局限性,这里要感谢匿名审稿专家对这一问题所提改进意见。期望这一研究能够激发更多学者在消费者结伴行为为主题的研究上有更多创新性的发现。

- my of Management Journal, 1994, 37(5): 1285 – 1301.
- [3] Ariely D, Levav J. Sequential choice in group settings: Taking the road less traveled and less enjoyed [J]. Journal of Consumer Research, 2000, 27(3): 279 – 290.
- [4] Greguras G J, Diefendorff J M. Different fits satisfy different needs: Linking person-environment fit to employee commitment and performance using self-determination theory [J]. Journal of Applied Psychology, 2009, 94(2): 465 – 477.
- [5] Sparrowe R T, Liden R C, Wayne S J, et al. Social networks and the performance of individuals and groups [J]. Academy of Management Journal, 2001, 44(2): 316 – 325.
- [6] Simpson J A, Griskevicius V, Rothman A J. Consumer decisions in relationships [J]. Journal of Consumer Psychology, 2012, 22(3): 304 – 314.
- [7] Triandis H C. Individualism-collectivism and personality [J]. Journal of Personality, 2001, 69(6): 907 – 924.
- [8] Nicholls J A F, Roslow S, Dublsh S. Time and companionship: Key factors in hispanic shopping behavior [J]. Journal of Consumer Marketing, 1997, 14(3): 194 – 205.
- [9] 石明明, 江舟, 周小焱. 消费升级还是消费降级 [J]. 中国工业经济, 2019, (7): 42 – 60.
Shi Mingming, Jiang Zhou, Zhou Xiaoyan. Consumption upgrading or downgrading [J]. China Industrial Economics, 2019, (7): 42 – 60. (in Chinese)
- [10] Sørensen F, Jensen J F. Value creation and knowledge development in tourism experience encounters [J]. Tourism Management, 2015, 46: 336 – 346.
- [11] Sohail M S, Shanmugham B. E-banking and customer preferences in Malaysia: An empirical investigation [J]. Information Sciences, 2003, 150(3 – 4): 207 – 217.
- [12] 孙书娜, 孙谦. 投资者关注和股市表现——基于雪球关注度的研究 [J]. 管理科学学报, 2018, 21(6): 60 – 71.
Sun Shuna, Sun Qian. Investor attention and market performance: Evidence based on “Xueqiu attention” [J]. Journal of Management Sciences in China, 2018, 21(6): 60 – 71. (in Chinese)
- [13] Rempel J K, Holmes J G, Zanna M P. Trust in close relationships [J]. Journal of Personality and Social Psychology, 1985, 49(1): 95 – 112.
- [14] Goldsmith R E, Flynn L R, Kim D. Status consumption and price sensitivity [J]. Journal of Marketing Theory and Practice, 2010, 18(4): 323 – 338.
- [15] Rook K S, Ituarte P H G. Social control, social support, and companionship in older adults’ family relationships and friendships [J]. Personal Relationships, 1999, 6(2): 199 – 211.
- [16] Garcia-Rada X, Anik L, Ariely D. Consuming together (versus separately) makes the heart grow fonder [J]. Marketing Letters, 2019, 30(1): 27 – 43.
- [17] Yeh J C, Hsiao K L, Yang W N. A study of purchasing behavior in Taiwan’s online auction websites [J]. Internet Research, 2012, 22(1): 98 – 115.
- [18] 刁雅静, 何有世, 王念新, 等. 朋友圈社交行为对购买意愿的影响研究: 认同与内化的中介作用及性别的调节作用 [J]. 管理评论, 2019, 31(1): 136 – 146.
Diao Yajing, He Youshi, Wang Nianxin, et al. Influence of friends’ social interaction on their purchase intention: The mediating role of identification and internalization and the moderating role of gender [J]. Management Review, 2019, 31(1): 136 – 146. (in Chinese)
- [19] Djafarova E, Rushworth C. Exploring the credibility of online celebrities’ instagram profiles in influencing the purchase decisions of young female users [J]. Computers in Human Behavior, 2017, 68: 1 – 7.
- [20] Paris L D, Howell J P, Dorfman P W, et al. Preferred leadership prototypes of male and female leaders in 27 countries [J]. Journal of International Business Studies, 2009, 40(8): 1396 – 1405.
- [21] 黄燕芬, 张超, 田盛丹. 人口年龄结构和住房价格对城镇居民家庭消费的影响机理 [J]. 人口研究, 2019, 43(4): 17 – 35.
Huang Yanfen, Zhang Chao, Tian Shengdan. The influencing mechanisms of population age structure and housing price on household consumption in urban China [J]. Population Research, 2019, 43(4): 17 – 35. (in Chinese)
- [22] Evashwick C, Rowe G, Diehr P, et al. Factors explaining the use of health care services by the elderly [J]. Health Services Research, 1984, 19(3): 357 – 382.

- [23] Thorslund M, Norström T, Wernberg K. The utilization of home help in Sweden: A multivariate analysis [J]. *The Gerontologist*, 1991, 31(1): 116–119.
- [24] Eng C, Pedulla J, Eleazer G P, et al. Program of all-inclusive care for the elderly (PACE): An innovative model of integrated geriatric care and financing [J]. *Journal of the American Geriatrics Society*, 1997, 45(2): 223–232.
- [25] Kim H S, Sherman D K, Taylor S E. Culture and social support [J]. *American Psychologist*, 2008, 63(6): 518–526.
- [26] 钟业喜, 黄洁, 文玉钊. 高铁对中国城市可达性格局的影响分析 [J]. *地理科学*, 2015, 35(4): 387–395.
Zhong Yexi, Huang Jie, Wen Yuzhao. Impact of high-speed railway on spatial pattern of Chinese cities' accessibility [J]. *Scientia Geographica Sinica*, 2015, 35(4): 387–395. (in Chinese)
- [27] 王文宪, 陈钉钧, 吕红霞, 等. 基于 Nested Logit 模型的铁路旅客客运产品选择研究 [J]. *交通运输系统工程与信息*, 2015, 15(4): 140–146.
Wang Wenxian, Chen Dingjun, Lü Hongxia, et al. Research of transport product choice for railway passengers based on nested logit model [J]. *Journal of Transportation Systems Engineering and Information Technology*, 2015, 15(4): 140–146. (in Chinese)
- [28] Zhang C B, Wei Y Y, Cheng F. Influences of mobile phone use of pedestrians on street-crossing behaviors and safety at signalized intersections [J]. *Journal of Transport Information and Safety*, 2019, 37(1): 20–34.
- [29] Muro-Rodríguez A I, Perez-Jiménez I R, Gutiérrez-Broncano S. Consumer behavior in the choice of mode of transport: A case study in the Toledo-Madrid corridor [J]. *Frontiers in Psychology*, 2017, 8: 1011.
- [30] 张睿, 马瑜, 赵冰茹, 等. 京沪线高铁与民航旅客出行选择行为研究 [J]. *交通运输系统工程与信息*, 2016, 16(1): 223–228.
Zhang Rui, Ma Yu, Zhao Bingru, et al. Passenger choice behavior of high-speed rail and airline between Beijing and Shanghai [J]. *Journal of Transportation Systems Engineering and Information Technology*, 2016, 16(1): 223–228. (in Chinese)
- [31] Borle S, Dholakia U M, Singh S S, et al. The impact of survey participation on subsequent customer behavior: An empirical investigation [J]. *Marketing Science*, 2007, 26(5): 711–726.
- [32] Hemamalini K. Influence of product types on consumer's attitude towards online shopping: An empirical study in the Indian context [J]. *International Journal of Marketing Studies*, 2013, 5(5): 41–52.
- [33] Herrmann A, Xia L, Monroe K B, et al. The influence of price fairness on customer satisfaction: An empirical test in the context of automobile purchases [J]. *Journal of Product & Brand Management*, 2007, 16(1): 49–58.
- [34] Munshi T. Built environment and mode choice relationship for commute travel in the city of Rajkot, India [J]. *Transportation Research Part D: Transport and Environment*, 2016, 44: 239–253.
- [35] Bhat C R, Gossen R. A mixed multinomial logit model analysis of weekend recreational episode type choice [J]. *Transportation Research Part B: Methodological*, 2004, 38(9): 767–787.
- [36] Jou R C, Hensher D A, Hsu T L. Airport ground access mode choice behavior after the introduction of a new mode: A case study of Taoyuan international airport in Taiwan [J]. *Transportation Research Part E*, 2011, 47(3): 371–381.
- [37] 芮海田, 吴群琪. 高铁运输与民航运输选择下的中长距离出行决策行为 [J]. *中国公路学报*, 2016, 29(3): 134–141.
Rui Haitian, Wu Qunqi. Medium-and-long distance travel mode decision between high-speed rail and civil aviation [J]. *China Journal of Highway and Transport*, 2016, 29(3): 134–141. (in Chinese)
- [38] Behrens C, Pels E. Intermodal competition in the London-Paris passenger market: High-speed rail and air transport [J]. *Journal of Urban Economics*, 2012, 71(3): 278–288.
- [39] Park Y, Ha H K. Analysis of the impact of high-speed railroad service on air transport demand [J]. *Transportation Research Part E: Logistics and Transportation Review*, 2006, 42(2): 95–104.
- [40] Lee J K, Yoo K E, Song K H. A study on travelers' transport mode choice behavior using the mixed logit model: A case study of the Seoul-Jeju route [J]. *Journal of Air Transport Management*, 2016, 56: 131–137.
- [41] Dubin J A, Rivers D. Selection bias in linear regression, logit and probit models [J]. *Sociological Methods & Research*, 1989, 18(2–3): 360–390.
- [42] Bushway S, Johnson B D, Slocum L A. Is the magic still there? The use of the Heckman two-step correction for selection

- bias in criminology [J]. *Journal of Quantitative Criminology*, 2007, 23(2): 151–178.
- [43] Poppo L, Zenger T. Do formal contracts and relational governance function as substitutes or complements? [J]. *Strategic Management Journal*, 2002, 23(8): 707–725.
- [44] Puhani P. The Heckman correction for sample selection and its critique [J]. *Journal of Economic Surveys*, 2000, 14(1): 53–68.

Travelling companion stimulates consumption upgrade: An analysis based on railway big data

CAO Wei-wei^{1,5,6}, LIU Sheng², LI Yi-wei², JIA Jian-min^{3,4*}

1. School of Economics and Management, Southwest Jiaotong University, Chengdu 610031, China;
2. Faculty of Business, Lingnan University, Hong Kong 999077, China;
3. School of Management and Economics, The Chinese University of Hong Kong (Shenzhen), Shenzhen 518172, China;
4. Shenzhen Research Institute of Big Data, Shenzhen 518172, China;
5. Civil Aviation Flight University of China, Guanghan 618307, China;
6. Key Laboratory of Flight Techniques and Flight Safety, Guanghan 618307, China

Abstract: The prevalence of group-based behavior has gradually increased in many business scenarios, including travelling, shopping, catering, and leisure activities. Such group-based behavior has the potential to influence consumers' purchasing decision. In the era of big data, human activities are recorded in ways that are far more detailed and systematic than before, which allows us to more accurately investigate the effects of group-based behaviors. We believe that the desire for a high-quality life fundamentally and consumption upgrade motivates group-based behaviors. Leveraging the sales data from Chengdu-Chongqing Railway, we investigate the effect of group travel on the choice of railway products and services. Based on Heckman's two-stage model with controlling for more than 1.3 million passenger fixed effects, our results show that travel with companions will enhance the possibility of choosing high-speed trains and premium seats. Further, such an effect is moderated by gender and age. Specifically, when travelling with female passengers and the elderly, one is more likely to choose high-speed trains and premium seats. These findings not only provide new insights into the railway management, but also suggest a general desire of human being for high-quality life and consumption upgrade beneath the group-based behaviors, which shed light on a wide range of sectors including business services and public administration. Through the encouragement of group-based behaviors, companies, organizations, and governments can stimulate consumers' and citizen's choices for high-quality product and services.

Key words: companions; consumption upgrade; big data; choice behavior; high speed rail