拼车中的规模效应：芝加哥市案例研究

Hao Liu[a,*], Saipraneeth Devunuri[b], Lewis Lehe[b], Vikash V. Gayah[a]

[a] *The Pennsylvania State University, University Park, PA, United States*
[b] *University of Illinois at Urbana-Champaign, Urbana, IL, United States*

arXiv:2303.03574v2 [stat.AP] 3 May 2023

摘要

拼车是一种网约车形式，乘客与其他乘客共乘车辆，如今已成为一些主要城市中常见的出行方式。目前，Uber、Lyft 和 Via 等交通网络公司（TNC）提供这种拼车服务，其用户数量不断增长，尤其是在COVID-19疫情之前。以往研究结果表明，拼车能够降低出行成本，甚至通过减少运送乘客所需的车辆数量来缓解拥堵。近期研究还提出，拼车应当具有正反馈机制，即服务质量会随着用户数量的增加而提升。具体而言，这类系统应能从规模经济和规模报酬递增中受益。本文利用交通网络公司向芝加哥市报告的2019年1月至9月行程数据，展示了这类机制存在的证据。具体来说，研究表明，在某一时段内请求或同意拼车行程的乘客数量增加，与行程绕行时间缩短、乘客成功匹配的比率提高、相较于非共享行程的成本降低，以及乘客拼车意愿增强相关。

关键词：拼车；交通网络公司（TNC）；规模效应；实证研究；共享意愿

1. 引言

拼车，也称为拼车服务，是一种网约车服务，乘客同意让驾驶员接载其他乘客并将其送达目的地；这些乘客针对同一服务分别提交了彼此不协调的请求（SAE International，2018；Shaheen 和 Cohen，2019）。UberPool 和 Lyft Line（现为 Lyft Shared Rides）于 2014 年 8 月推出（Shaheen，2018），是美国最常见的选择；而交通网络公司（TNC）Via 仅在少数城市提供拼车服务。滴滴出行的拼车选项名为*ExpressPool*（Li 等，2019）。拼车应与通过应用程序实现的合乘或共享出行区分开来，例如 Waze Carpool 和滴滴的 Hitch[1] 选项。前者的驾驶员虽不一定是专业或持证司机，但纯粹是为了报酬而驾驶；驾驶员与乘客之间并不共享出发地或目的地。

为了实现拼车，交通网络公司会利用路径规划与匹配算法以及出行时间数据，实时为乘客和驾驶员进行匹配。我们将通过拼车服务完成的行程称为拼车。普通的网约车行程则称为单人行程。我们还区分经授权的拼车行程（也称为请求共享行程），即用户授权共享、选择通过 UberPool、Lyft Shared Rides 等服务提交请求的行程；以及已匹配的拼车行程，即实际上与提出了不同请求的另一位乘客共同乘坐的行程。授权共享的乘客并不一定最终会匹配成功。

*通讯作者*Email addresses:* hfl5376@psu.edu (Hao Liu), sd37@illinois.edu (Saipraneeth Devunuri), lehe@illinois.edu (Lewis Lehe), vvg104@psu.edu (Vikash V. Gayah)[1]正如 Wang 等（2019年）所解释的，Hitch 面向希望分摊通勤出行和城内出行成本的驾驶员。

术语介绍到此为止，现在进入本文的主题。对于匹配的拼车行程，乘客几乎肯定会经历绕行，因为车辆必须绕路接送另一位乘客。因此，与单人乘车相比，拼车的预期出行时间更长，出行时间确定性也更低（因为可能在行程开始后才安排匹配）（Li 等，2019年）。拼车的好处是，交通网络公司收取的票价较低。美国的一些城市和州（如芝加哥、纽约市、佐治亚州、新泽西州）对拼车收取的消费税低于单人乘车，从而扩大了票价差额（Lehe 等，2021年）。因此，拼车是金钱成本与出行时间之间经典权衡的典型例子。Chen 等人（2017年）将出行时间列为决定是否选择拼车时最重要的行程"特征"，而接下来的两个特征都与行程价格有关。

在美国，交通网络公司仅在大型市场提供拼车服务。目前，$Uber\ Pool$ 仅在九个美国城市运营，$Lyft\ Line$ 则在约十五个城市提供服务。因此，显然，拼车服务要以财务上可持续的方式运营，潜在需求必须达到一定门槛。这一最低规模远高于维持其他网约车服务所需的规模，即便是相当小众的服务也不例外，例如豪华黑车服务（在数十个美国市场提供）。本文的论点是，作为一种经济活动，拼车具有两个相关特征：

- 第一，规模经济：当某种产出单位的平均生产成本随产量增加而下降时，就会产生规模经济（Silvestre，1987）。传统上，规模经济与钢铁等制造产品相关，关注的成本是生产者承担的成本。例如，钢铁企业建成一座高炉或电弧炉后，按吨计算，用它生产许多吨钢铁比只生产一吨更便宜。对于客运，学者发现用户一方也存在规模经济（Basso 和 Jara-Dı́az，2010）。以公交服务为例，假设产出单位是公交车次，并将乘客的出行时间价值（包括在站点等候的时间）计入"生产成本"。从这种视角得出的一个著名结论，就是所谓的莫林效应：Mohring（1972）指出，当乘客量增加带动服务频率提高（例如每小时增加公交车次）时，单个乘客在站点的等候时间会随着乘客总量增加而缩短。

  不难看出，拼车可能具有用户侧规模经济。如果拼车请求在空间和时间上的请求密度都非常高，那么交通网络公司就能将乘客匹配到一起，让他们拼乘绕行更短的行程。极端情况下，当某个市场中对拼车服务的请求密度趋于无穷大时，任何提出拼车请求的人都可以与其他出发地和目的地完全相同的乘客组成满载车辆；这样就无需绕行。

- 第二，规模报酬递增："如果所有投入要素按相同比例增加，产出却能以更大的比例增加，那么这项技术就表现出规模报酬递增"（Vassilakis，1987）。与规模经济一样，制造业是典型案例，但所谓的劳动力市场"搜寻理论"中也能观察到规模报酬递增（Diamond，1982；Mortensen 和 Pissarides，1994）。在这一理论中，产出单位是一次匹配，而匹配过程的投入要素是同时搜寻劳动者的雇主存量（即公开发布的职位空缺）以及搜寻雇主的劳动者存量（即失业人员）。只要职位空缺和失业人员这两类存量按相同比例增加（例如，职位空缺数和失业人员数均增加10%），由此带来的匹配总流量增幅超过该比例，就会产生匹配中的规模报酬递增。

  不难想象，匹配过程中的规模报酬递增也应当出现在拼车服务中。当市场中的拼车请求密度上升时，产生的影响不可能仅仅是匹配行程的绕行距离缩短，因为交通网络公司并不会匹配所有获准拼车的行程。如果匹配两趟行程所需的绕行距离过长，乘客就会被直接送往各自的目的地，就好像他们选择的是"单人行程"一样。因此，随着拼车请求增加，得到匹配的拼车请求占比也会上升。

总而言之，关于拼车请求数量增加时可能发生的情况，我们做出以下几项预测：

1. 绕行距离应当缩短

2. 最终成功匹配的拼车请求占比应当上升

此外，这些效应还可能产生另外两种潜在的连带效应。当拼车请求数量上升时：

1. 拼车行程票价与“单人乘车”票价之比应当下降（Zhu 等（2020）称之为“票价比”），因为票价部分取决于绕行距离，而且匹配成功的概率会上升。2. 所有行程请求中授权拼车的比例应当上升，因为拼车行程耗时更短，而且（可能）票价也会下降。

本文旨在确认实证数据中是否存在这种行为。本文其余部分安排如下：第 2 节介绍所用数据，第 3 节呈现实证研究结果。最后，第 4 节给出结论。

2. 数据

我们使用 Uber、Lyft 和 Via 于 2019 年向芝加哥市报告的“行程”数据集（芝加哥市将其称为“交通网络公司”），以确定真实拼车系统中是否存在规模经济和规模报酬递增。[2] 每家交通网络公司均提供了每趟出发地或目的地位于芝加哥市范围内行程的以下信息：

– 每次行程对应的起止时间戳，按最接近的 15 分钟间隔取整；– 行程的起点和终点所在的人口普查区；– 票价（美元），按最接近的 2.50 美元倍数取整；– 税费等附加费用（美元）；– 费用及其他收费（美元）；– 行驶距离（英里）；– 行程时长（分钟）；– 是否“获准”拼车；以及 – 拼车行程是否 “ 成功匹配 ”。

数据清理包括从 2019年 数据集中删除某些记录。清理过程以两个目标为指导。

1. 第一个目标几乎适用于任何网约车数据研究：删除可能有误或无关紧要的记录。无关紧要 是指那些虽不一定记录错误、但与了解网约车的性质无关的记录；例如，乘客上车后立即取消的行程。

2. 第二个目标是本文总体目的中特有的：比较相似时间的变量，以尽量减少匹配规模经济以外因素的影响。下面我们将依次介绍各个流程。

[2]该数据集可通过以下链接公开获取：https://data.cityofchicago.org/Transportation/Transportation-Network-Providers-Trips/m6dm-c72p

为剔除错误或无关紧要的记录，我们立即筛除了以下行程：

— 出行时间少于2分钟；— 行驶距离短于0.1英里；— 缺少上车点或下车点人口普查区信息；以及— 未获授权为拼车行程、但被编码为拼车行程的行程。

此外，由于我们需要将所有行程的出发地和目的地映射至芝加哥市的人口普查区，因此，上车点或下车点人口普查区位于芝加哥市以外的行程也被排除。文献中的其他研究，例如（Dean and Kockelman, 2021），也删除了此类行程。在该研究中，Dean and Kockelman（2021）关注促成拼车的因素，并使用社会经济变量、建成环境变量和时间变量构建了两个模型，用于估计交通网络公司行程中的拼车决策。此外，他们仅关注获准的共享行程，而我们同时使用授权共享出行和匹配的共享出行。

除筛选过程外，我们还发现9月之前和10月之后的月份之间存在系统性差异，因此将2019年数据集划分为两个时期：1月至9月和10月至12月。芝加哥市于4月 28, 2020[3]日发表声明：“由于数据处理错误，2019年前九个月的部分网约车服务提供商行程未向芝加哥市报告，因此最初未被纳入网约车服务提供商行程数据集。现已纠正这些错误。我们已收到相关行程，并将其添加至数据集。”尽管没有明确证据可以解释1月至9月期间观察到的拼车行程数量持续下降，但在10月至12月期间观察到的一些现象（例如短途行程出现极长的绕行距离）更加可疑且令人质疑。遗憾的是，目前没有确定的理论可以解释这些模式。因此，本文排除了10月至12月的数据。附录A提供了详细分析。由于2020年1月发生了网约车税上调，我们不考虑2020年1月及之后的行程。用于复现本文图表的处理后数据可在哈佛数据宇宙（Liu等人，2023）上获取。代码已开源，可在https://github.com/UTEL-UIUC/Ridesharing-Scale-Effects找到。

数据集中的出行分布如图1所示。热力图展示了每个人口普查区出发或到达的出行总次数。出行次数极高的人口普查区要么位于芝加哥市中心（下文称为中央商务区，即 CBD），要么位于芝加哥市内两座主要机场（奥黑尔机场和米德韦机场）。除非人口普查区属于旅游景点、大学、购物中心等特殊区域，否则距离市中心越远，出行集中度就越低。根据出行分布，我们沿图1中的虚线将芝加哥市划分为两个区域。我们采用这条边界，以确保市中心区域位于北侧，并尽量使两个区域的人口普查区数量相近。下一节将分别考察两个区域的 WTS 和匹配率。

我们按小时（h）、星期（d）和月份（m）对数据进行分组，并统计每个小时-星期-月份组合中的（已授权拼车/匹配的拼车/单人）出行总数或平均数。例如，3月所有星期一上午9点至10点之间，共有8,813次匹配的拼车出行。因此，3月星期一上午9点至10点之间，平均有2,203次匹配的拼车出行。部分指标基于这些聚合数据定义，将在使用时加以说明。某个分组的单位距离/时间票价，是通过对属于该组的所有出行的单位票价取平均值计算得出的。请注意，第3.1节采用了基于小时和星期的另一种聚合层级，以探究工作日的共享意愿（WTS）模式。绕行的定义将在第3.2节单独说明。表1列出了所用符号，供参考。

[3]本声明 可通过以下链接查看： http://dev.cityofchicago.org/open%20data/data%20portal/2020/04/28/tnp-trips-2019-additional.html

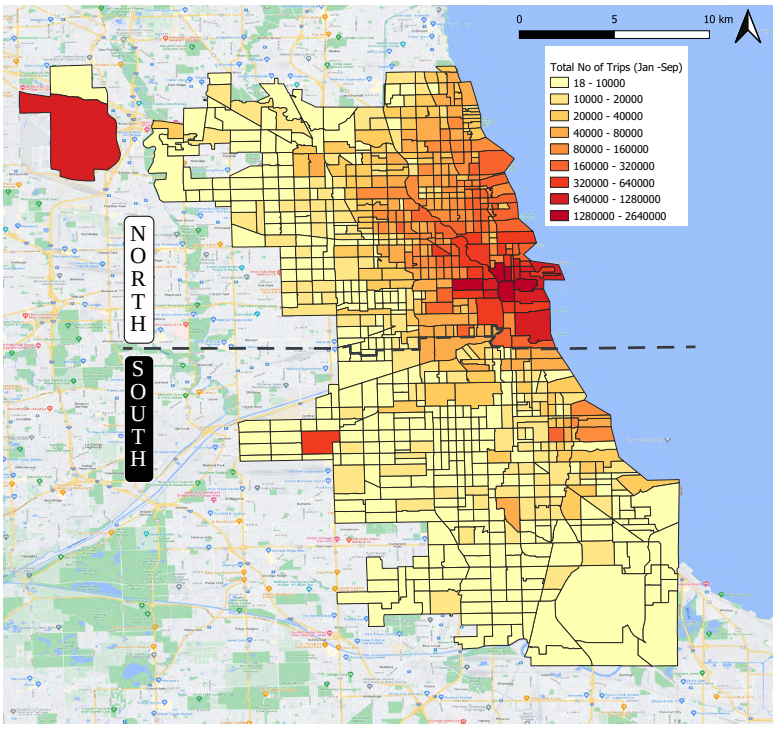


图1：2019年1月至9月的行程分布。

3. 研究结果

本节考察拼车中是否存在第1节所述的规模经济和规模报酬递增。我们展示了交通网络公司数据支持以下推测：

— 随着网络中的网约车行程总数增加，共享意愿随之提高；— 随着匹配的共享出行次数增加，共享带来的绕行距离缩短；— 随着获准共享行程的比例提高，共享行程的匹配率随之上升；— 随着网约车行程总数增加，共享行程与单人行程的单位票价比降低。

3.1. 共享意愿

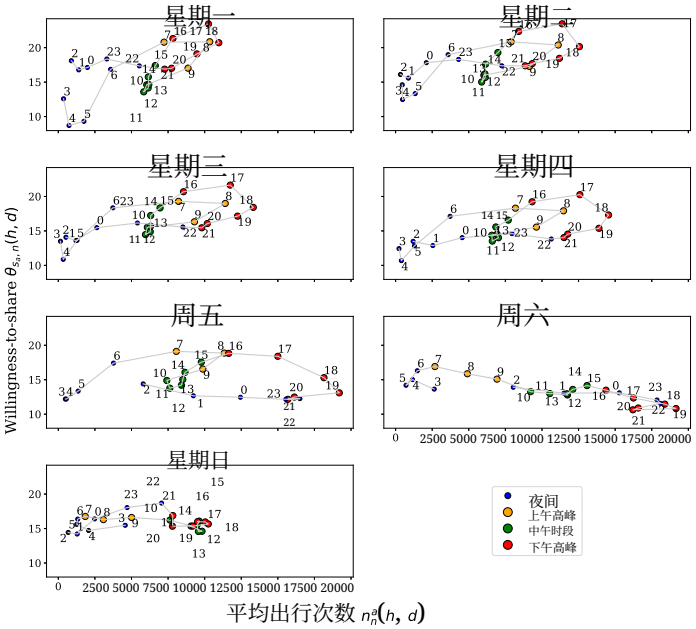
区域 i 的 WTS 定义为给定小时-日期组合中获准共享的行程所占比例：

$$\theta_{s_a,i}(h,d) = \frac{n^t_{s_a,i}(h,d)}{n^t_i(h,d)} \times 100 \quad (\% \text{ authorized shared trips}). \tag{1}$$

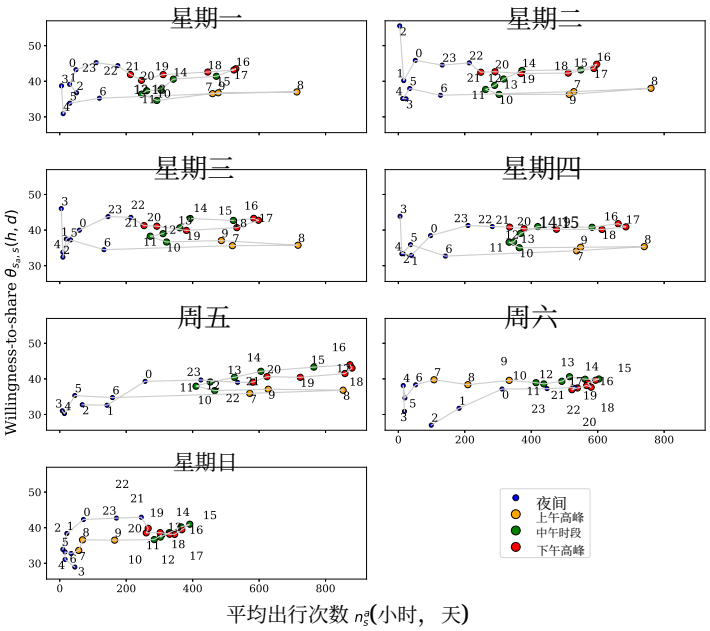
图2展示了两个区域的 WTS 随网约车行程总数变化的情况。横轴表示分析期内区域 i、n^a_i（h，d）在相同工作日中每小时两种模式的平均出行次数。每幅图的小时范围为凌晨3点至次日凌晨2点。如图2所示，两个区域的大多数工作日中，交通流量在凌晨3点达到最低，因此采用这一时段作为每个图表的起点。每幅图的小时范围从凌晨3点延伸至次日凌晨2点，这是因为如图2所示，两个区域的大多数工作日中，交通流量在凌晨3点达到最低。因此，如果将工作日的交通变化视为一个周期，我们便以交通流量最低的时刻作为周期起点。这仅为便于可视化，并不影响递增趋势这一节的重点。

表1：符号说明

符号	定义	表示
$n_{s_a,i}^{t(a)}(h,\ d,\ (m))$	区域 i 中的小时 h、星期 d（以及月份 m）中授权共享出行次数总数/平均数	共享授权数量
$n_{s_m}^{t(a)}(h,\ d,\ m)$	匹配的共享出行总数（/平均数），时间为小时 h、星期 d 和月份 m	共享匹配数量
$n_{s,i}^{t(a)}(h,\ d,\ m)$	h小时内单人行程的总数（/平均数），星期 d 和月份 m 的区域 i	单人行程数量
$n_i^{t(a)}(h,\ d,\ (m))$	中两种模式的行程总数（/平均数）h小时、星期 d （以及月份 m）的区域 i	总出行次数
$\theta_{s_a,i}(h,d,m)$	小时 h的获准拼车出行百分比，星期 d 和月份 m 的区域 i	共享意愿
$\theta_{s_m,\ i}(h,\ d,\ m)$	小时 h、周-的匹配的拼车出行百分比日 d 和月份 m 的区域 i	拼车匹配率
$n^{O,\ D}_{s_m}(h,\ d,\ m)$	地点对中匹配的拼车出行平均次数第 h小时、星期 d 和 m月份的(O,D)	共享匹配数量 每个起讫点
$d^{O,D}_{si(s_m)}(h,\ d,\ m)$	单人（/匹配共享）行程的平均行程距离第 h小时、星期 d 和月份中地点对(O,D)的行程 m	行驶距离
$de^{O,\ D}(h,\ d,\ m)$	地点对 (O,D) 在 h 小时、星期 d 的绕行距离以及 m 月份	由以下因素产生的绕行距离 共享出行
$c^{d(t)}_{si(s_m)}(h,d,m)$	在小时 h、星期 d 和月份 m 中基于距离的单人行程（/匹配拼车行程）平均单位票价（/时间）	单位票价
$e^{d(t)}(h,\ d,\ m)$	小时 h内拼车行程与单人行程的单位票价比，基于距离（/时间）的星期 d 和月份 m	共享出行带来的成本节省



（a）北侧



(b) 南侧

图2：获准拼车出行百分比。

在北侧区域，除周五夜间时段外，工作日的 WTS 会随着网络中的行程量增加而上升；周六则呈下降趋势，星期日则相对稳定。

随后，我们根据工作日的预期出行模式，将一天划分为4个时段：

- 7:00-9:00 — 上午高峰（黄色）
- 9:00-16:00 — 中午时段（绿色）
- 16:00-21:00 — 下午高峰（红色）
- 21:00-7:00 — 夜间（蓝色）

在北侧区域，周一至周四的模式相似。最高的WTS出现在16:00-18:00之间，而WTS在清晨时段（4:00-5:00）降至最低。除少数异常值外，例如周一3:00-4:00，其他时段中，WTS会随行程次数的增加/减少而上升/下降。高峰时段还观察到了滞后模式：高峰时段开始时的WTS显著高于行程量下降阶段的数值。周五在下午高峰时段之后呈现出与其他工作日不同的模式。与其他工作日相比，周五及周末这些时段的总出行频率更高，而WTS更低。具体而言，在北侧区域，平均WTS（定义为共享出行总次数除以总出行次数）从工作日的17.3%降至周末的13.6%，其中包括周五下午高峰时段之后的时段。南侧区域的相应变化从39.8%降至38.8%，幅度较不显著。行程量增加可能是由于周五下午至晚间的社交休闲出行增加所致。较低的WTS意味着出行者在这些活动中更可能选择独自乘车，以避免绕行。另一个可能的原因是，在周末以及周五夜间，更多人倾向于与朋友和家人共同使用私家车出行，这导致平均同行人数更高，而网约车行程的WTS更低。

尽管工作日北侧区域的WTS会随着交通流量增加，但南侧区域并未呈现这一增长趋势。工作日的大多数小时里，WTS维持在两个稳定水平：清晨时段的WTS相对较低且稳定，而下午时段和夜晚早期时段

的WTS相对较高且稳定。与北侧区域类似，周末的模式也不同于工作日。此外，尽管南侧区域的交通流量远低于北侧区域，但其平均WTS更高。不过，我们注意到，WTS取决于多种因素，例如与交通系统相关的属性和人口统计属性。因此，若不控制这些因素，较高的行程量并不一定会带来较高的WTS。不过，我们预计，当某区域的行程量足够高时，若其他属性不变，WTS会随行程量增加。如图2所示，南侧大多数小时的平均出行次数不到每小时800辆车，远低于北侧的数值。没有呈现增长趋势的一个可能原因是，行程量太低，不足以对WTS产生明显影响。

3.2. 绕行

本节探讨在给定时间内拼车行程数量增加后，拼车行程的绕行距离可能减少的幅度。对于给定的起讫点（OD）对，绕行距离定义为该时段内拼车行程的平均额外行驶距离，即与同一小时内该起讫点对之间的单人行程相比所增加的距离，可表示为：

$$de^{O,D}(h,d,m) = d_{s_m}^{O,D}(h,d,m) - d_{si}^{O,D}(h,d,m) \tag{2}$$

因此，如果某个小时内某一OD对之间只观察到一种出行方式，则该OD对之间的行程会从本分析中剔除，因为在这些情况下无法定义绕行距离。经过这一筛选过程后，保留了63,992个独特OD对中的11,299,984次行程。其中，8,680,382次为单人行程，2,619,602次为拼车行程。

我们考察了拼车行程数量最多的16个起讫点对中，绕行距离与匹配的拼车行程数量之间的关系。这些起讫点对总计包含1,261,302次单人行程和118,733次拼车行程。为简洁起见，图3仅展示前6个起讫点对的结果，其余10个起讫点对的结果见附录B。由于每个起讫点对的数据点数量都很大，而且大多数点对应的拼车行程数量相对较少，传统散点图中拼车行程数量较少区域内的点云非常密集，因此无法从图中直观地获取信息。为此，我们采用分箱散点图，以便更清晰地直观评估单调性等特征。图3使用R软件包“binsreg”绘制的分箱散点图展示了数据（Cattaneo等人，2019a,b）。对于每个起讫点对，我们采用包含8个分箱的分位数分箱，尽量将相同数量的观测值分配到8个分箱中。每个分箱中的点和线段分别表示该分箱内观测到的绕行距离在均值处计算的点估计，以及95%置信区间（p值为0.05）。阴影区域表示与各分箱相关的两个变量的置信带。图3显示，对于所有起讫点对，绕行距离总体上随拼车行程数量增加而减少。拼车行程数量较少时，这一趋势更为明显。相反，当拼车行程数量非常大时，下降趋势并不那么明显。

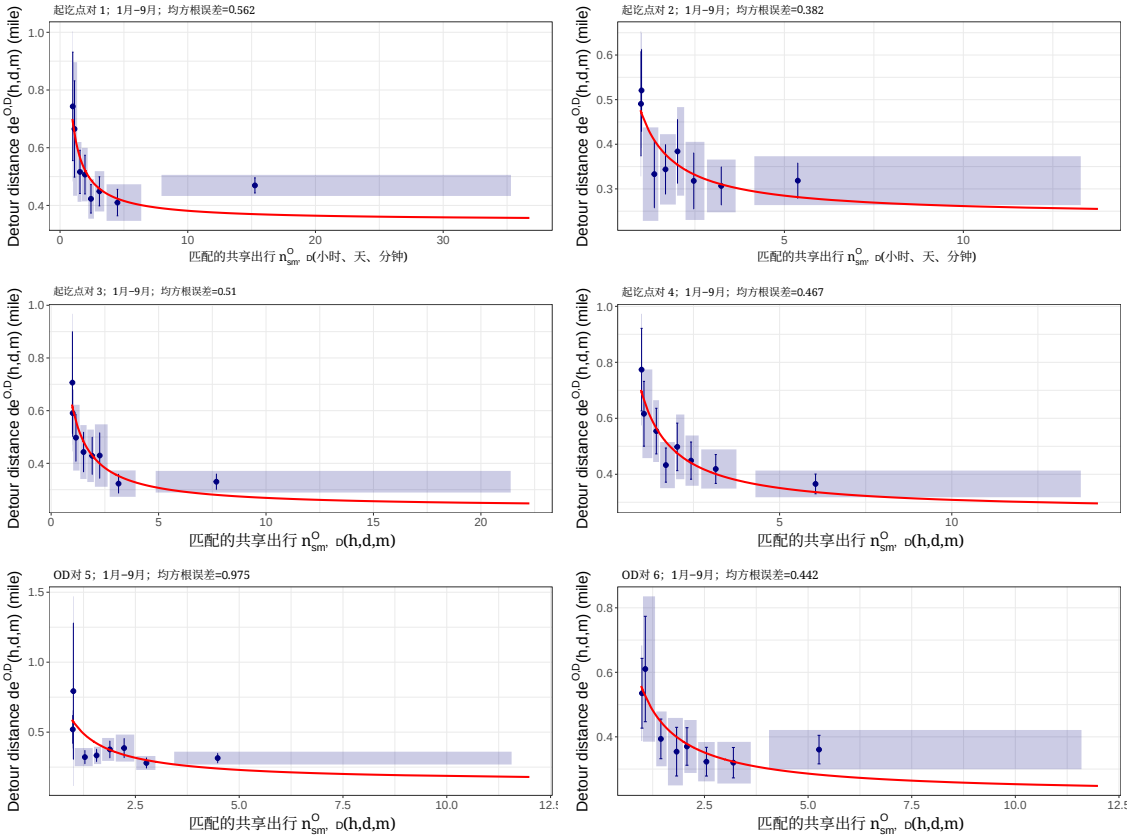


图3：六个起讫点对的绕行距离。

为检验　　为评估这一下降趋势的强度和统计显著性，我们还拟合了回归模型每个起讫点对　。根据观察到的趋势，采用了以下函数形式：

$$de^{O,D}(h,d,m) = \beta_0 + \beta_1 \frac{1}{n_{s_m}^{O,D}(h,d,m)}. \tag{3}$$

回归结果见表2。请注意，回归模型旨在帮助呈现这些点的趋势，而不是提出参数之间存在任何因果关系。此外，回归使用的是原始数据，而分箱仅用于更好地呈现数据。回归结果表明，在所有起讫点对和时间段内，下降趋势在统计上均显著。

除了表2所示的回归结果外，回归模型（3）的弹性可表示为：

$$\epsilon_{de}^{O,D}(h,d,m) = \frac{dy/y}{dx/x} = -\frac{\beta_1}{\beta_0 n_{s_m}^{O,D}(h,d,m) + \beta_1}, \tag{4}$$

该指标量化了因变量（即绕行距离）相对于自变量（即匹配的拼车行程数量）的相对变化。对于图3所示的6个起讫点对，当匹配的拼车行程数量为1时，匹配的拼车行程数量每增加x%，绕行距离就会分别减少0.499x%、0.496x%、0.624x%、0.613x%、0.754x%和0.600x%。当拼车行程数量较大时，这种影响会减弱。例如，当匹配的拼车行程数量为5时，匹配的拼车行程数量每增加x%，绕行距离就会分别减少0.165x%、0.163x%、0.248x%、0.242x%、0.371x%和0.231x%。这意味着绕行距离会变得

当拼车行程数量较大时，对其相对不敏感。不过，图中显示，随着拼车行程数量增加，绕行距离可以减少至原来的一半甚至更低。

请注意，与绕行距离不同，绕行时间更为复杂，因为它还取决于行驶速度。图2显示，拼车行程数量与总出行次数之间存在正相关关系。因此，较高的拼车行程数量意味着北侧区域的出行量较大，从而导致行驶速度降低。因此，绕行时间的变化取决于绕行距离和行驶速度的相对变化。此外，为保护个人隐私，交通网络公司会将所有行程的起止时间戳四舍五入至最近的15分钟时间间隔，这使得我们无法获得准确的绕行时间。由于缺乏行驶速度数据和准确的出行时间，这一问题未在本文中进行研究，但将作为我们未来的研究方向之一。

3.3. 共享行程匹配率

图2显示，愿意拼车的乘客比例会随着出行者数量的增加而上升，至少在芝加哥北侧区域是如此；大多数行程都发生在该区域。基于此，本节考察：随着更多乘客请求拼车，获准的共享行程被匹配的概率是否会增加。区域 i 中给定小时-星期-月份组合的匹配百分比定义为：

$$\theta_{s_m,i}(h,d,m) = \frac{n^t_{s_m,i}(h,d,m)}{n^t_{s_a,i}(h,d,m)} \times 100 \quad (\%\ \text{matched shared trips}). \tag{5}$$

图4—5展示了两个区域的匹配百分比与授权共享出行次数之间的关系。我们按照图2所定义的四个时段划分数据。图4—5表明，两个区域的匹配百分比均随授权共享出行次数增加而上升，这与我们对规模经济的预期一致。当网络中获准拼车出行的百分比上升时，更多获准行程可能会在彼此相近的位置起点和终点，因此会有更多请求得到匹配。然而，南侧区域的上升趋势不如北侧明显。这可能是因为夜间时段存在一些获准匹配行程次数极低的点，这些行程仍然得到了匹配。因此，如图5a所示，这些点的匹配百分比极高，例如达到100%。在剔除该区域中授权共享出行次数少于10次的点后（共筛除213个点），我们得到了图5b所示的结果；该图显示，匹配百分比随授权共享出行次数增加而上升的趋势更加显著。

根据图4—5中观察到的视觉模式，拟合了如下形式的分时段回归模型，

表2：绕行距离的回归结果（英里）

	因变量：					
	绕行距离 $de^{O,D}(h,d,m)$					
	(1)	(2)	(3)	(4)	(5)	(6)
$\frac{1}{n^{O,\ s_m\ D}(h,\ d,\ m)}$	0.341***	0.234***	0.380***	0.424***	0.431***	0.330***
	(0.060)	(0.047)	(0.058)	(0.057)	(0.122)	(0.056)
常数	0.348***	0.238***	0.231***	0.266***	0.145*	0.219***
	(0.036)	(0.031)	(0.040)	(0.037)	(0.083)	(0.036)
R^2	0.033	0.029	0.048	0.059	0.014	0.040
调整后的 R^2	0.032	0.028	0.047	0.058	0.013	0.039
显著性水平						*p<0.1; **p<0.05; ***p<0.01

$$\theta_{s_m,i}(h,d,m) = \beta_0 + \beta_1\sqrt{n^a_{s_a,i}(h,d,m)} + \delta^h_A\beta_{AMPeak} + \delta^h_M\beta_{Mid-day} + \delta^h_P\beta_{PMPeak}. \tag{6}$$

在公式（6）中，将时段视为分类参数并以“夜间”小时为参照水平，从而使系数 β_1 在所有时段保持固定； δ_i^h为二元变量，表示数据点所属的时段。回归结果见表3。请注意，所有模型的 $\sqrt{n_{s_a},i(h,d,m)}$ 系数均为正且显著，这表明匹配率与发起的拼车行程数量之间存在统计学上显著的正相关关系。此外，回归模型（6）的弹性可表示为，

$$\epsilon_{\theta_{s_m,i}}(h,d,m) = \frac{dy/y}{dx/x} = \frac{\beta_1}{2}\frac{\sqrt{x}}{\beta_0 + \beta_1\sqrt{x} + \delta_A^h\beta_{AMPeak} + \delta_M^h\beta_{Mid-day} + \delta_P^h\beta_{PMPeak}}, \tag{7}$$

与绕行距离不同，整体匹配百分比在数据集中获准的共享行程范围内并未呈现平稳趋势。例如，在下午高峰时段，当北部地区的 $n_{s_a}^a$ 从每小时1,000辆增至4,000辆时，弹性从0.128增至0.203；在南部地区，当筛选后的$n_{s_a}^a$ 从每小时100辆增至600辆时，弹性从0.122增至0.221。这表明，尽管南侧的拼车行程少于北侧，但获准的共享行程对南侧匹配百分比的相对影响比对北侧的影响更显著。

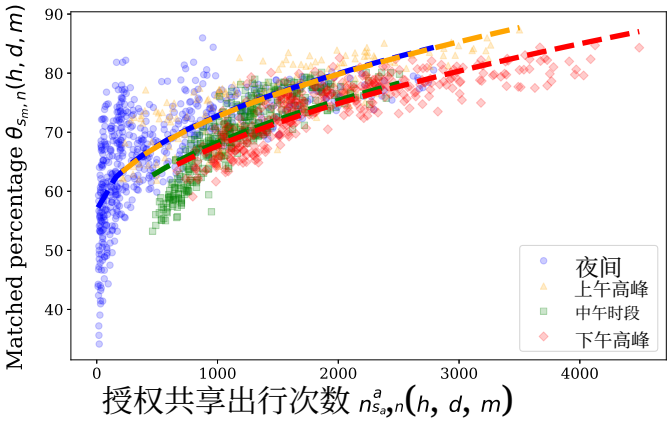
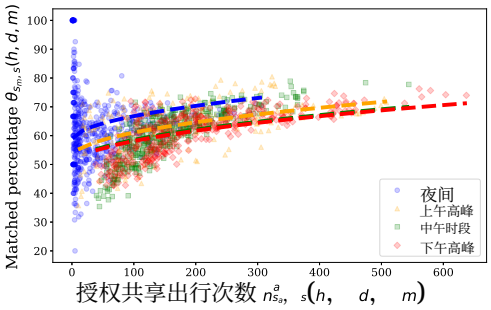
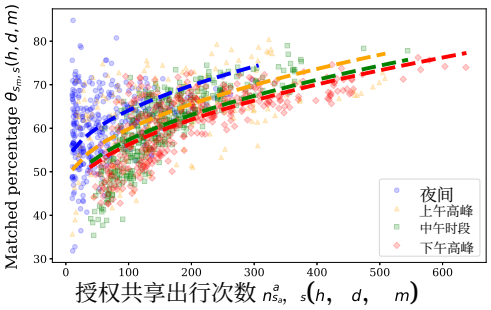


图4：匹　　配的北部地区拼车行程百分比，　　均方根误差=4.891



(a) 筛选前，均方根误差=10.199



(b) 筛选后，均方根误差=6.548

图5：南部地区匹配的拼车行程百分比。

按时间段进行的分析显示，上午高峰、中午时段和下午高峰时段的数据点比夜间时段的数据点更密集，也更接近回归线。这一现象的一个可能原因是，出于安全考虑，人们在夜间时段可能倾向于避免与陌生人同乘一辆车出行

表3：匹配的拼车出行百分比的回归结果

	因变量：		
	匹配百分比 $\theta_{s_m,\ i}(h,\ d,\ m)$		
	N 1月至9月	S 1月至9月之前	S 1月至9月之后
$\sqrt{n^a_{s_a,\ i}(h,\ d,\ m)}$	0.546***	0.859***	1.378***
	(0.013)	(0.071)	(0.051)
上午高峰	−0.120	−5.704***	−4.255***
	(0.455)	(0.996)	(0.646)
中午时段	−4.535***	−8.234***	−6.745***
	(0.362)	(0.843)	(0.549)
下午高峰	−5.083***	−8.621***	−7.831***
	(0.429)	(0.904)	(0.584)
常数	55.515***	58.172***	50.326***
	(0.335)	(0.573)	(0.512)
R^2	0.631	0.096	0.372
调整后的 R^2	0.631	0.094	0.370
显著性水平		*p<0.1; **p<0.05; ***p<0.01	

。另一个可能原因是，夜间时段的起讫点对分布可能比其他时段更加随机、分散。因此，匹配率也更加随机。

3.4. 单位票价比

现在我们考察行程货币成本的两个标准化指标。每次行程的单位距离票价（美元/英里）和单位时间票价（美元/分钟）分别通过总票价除以行程距离和出行时间得到。需要注意的是，为保护个人隐私，行程票价会四舍五入至最接近的2.50美元，小费会四舍五入至最接近的1.00美元，这对本分析构成了一项限制。拼车行程与单人行程之间的单位票价比称为票价比。图6和图 7分别展示了基于时间的票价比和基于距离的票价比随出行总次数变化的情况。正如预期，单位票价比始终小于1，这意味着拼车行程的成本始终低于单人行程。在所有时段，单位票价比均随着交通流量增加而下降，且所有系数都具有显著性。这些结果可能表明，随着交通流量增加，出行者会被鼓励选择拼车行程，以减轻交通负担。我们拟合了与公式（6）相同的回归模型，表4给出了回归结果，证明这些研究结果具有统计显著性。此外，当出行次数从10,000增加到20,000时，基于距离的票价比的弹性从-0.112降至-0.174，基于时间的票价比的弹性从-0.022降至-0.032。这表明，交通流量对基于距离的单位票价比的影响更为显著。如第3.2节所述，尽管假设交通流量不会影响行驶距离是合理的，但其对出行时间的影响却不容忽视。随着交通流量增加，预计出行时间会延长。因此，在假设总票价不变的情况下，两种模式的基于时间的单位票价都会下降。相应地，在总票价固定降幅的情况下，基于时间的单位票价比的降幅小于基于距离的单位票价比。

4. 结论

本文利用2019年向芝加哥市报告的实证数据，展示了拼车服务中存在规模经济和规模报酬递增的证据。由于数据中出现了可疑现象

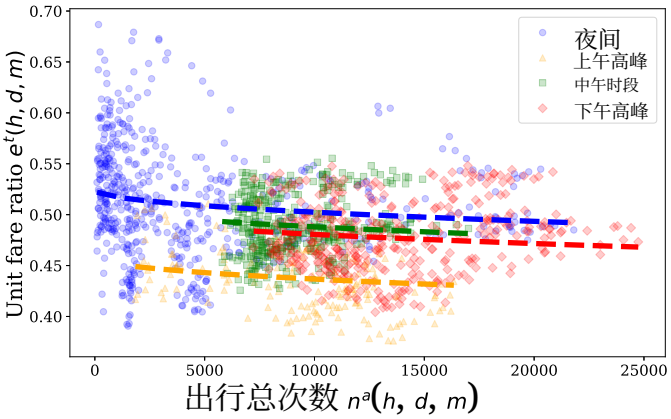


图6：基于时间的单位票价比，均方根误差=0.041

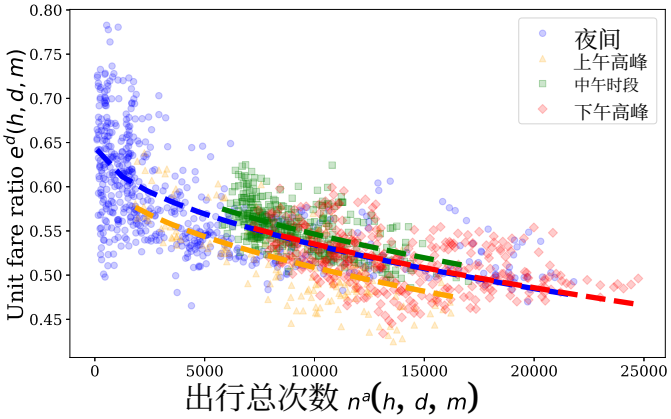


图 7：基于距离的单位票价比，均方根误差=0.036

，因此本研究排除了10月至12月这一时段。规模经济表现为：随着拼车请求数量增加，绕行距离缩短；规模报酬递增则表现为行程匹配率提高。具体而言，当某个OD对每小时的共享行程数量为1时，出行次数最多的16个OD对的绕行距离平均弹性为-0.584。随着拼车行程数量增加，其对绕行距离的相对影响会减弱；例如，当某个OD对每小时的共享行程数量增至5时，相应的平均弹性变为-0.226。此外，还观察到两种连带效应：随着授权共享出行次数增加，共享意愿提高，而拼车行程与单人行程之间的单位票价比下降。此外，尽管周末的出行总次数高于工作日，但周末的WTS低于工作日。例如，在北侧区域，平均共享意愿从工作日的17.3%降至周末的13.6%。由于交通流量增加会延长出行时间，基于距离的单位票价比的弹性高于基于时间的单位票价比。请注意，尽管上述所有理论均将授权共享出行次数作为自变量，但我们也采用其他替代指标，例如匹配的共享出行次数和共享行程总数，以使所表达的关系更加自然。例如，图3和图12将匹配的共享出行次数作为自变量，因为相比授权共享出行次数，用它来呈现绕行距离与匹配的共享出行次数之间的关系更为直观

表4：成本比率的回归结果

	因变量：	
	基于时间的 e^t	基于距离的 e^d
$\sqrt{n^a(h,d,m)}$	-0.0002^{***}	-0.001^{***}
	(0.00004)	(0.00004)
上午高峰	-0.066^{***}	-0.026^{***}
	(0.004)	(0.003)
中午时段	-0.014^{***}	0.012^{***}
	(0.003)	(0.003)
下午高峰	-0.022^{***}	0.001
	(0.004)	(0.003)
常数	0.524^{***}	0.654^{***}
	(0.003)	(0.003)
R^2	0.256	0.571
调整后 R^2	0.254	0.570

显著性水平 *p<0.1； **p<0.05； ***p<0.01

出行次数。不过，所有这些替代指标都与授权共享出行次数呈正相关，因此支持了我们得出的结论。总体而言，本文的研究结果表明，拼车服务使用量的增长能够带来益处。这可能会形成 Lehe and Pandey (2020) 所提出的正反馈循环：拼车使用量增加应能改善服务质量，从而进一步激励乘客使用该服务。因此，为了利用规模经济和规模报酬递增所带来的这一正反馈机制，应制定相关措施和策略，鼓励并提升拼车服务。

尽管观察到的经验趋势表明存在上述规模经济和规模报酬递增现象，但仅凭这些趋势无法确认它们确实存在于拼车领域。仍有可能是其他原因导致了这些趋势。尽管如此，这些观察到的趋势与这些现象若确实存在时的预期相符，也足以令人信服地将其归因于这些原因。未来的研究应努力分离这些影响、确定因果关系，并在其他数据集中确认它们的存在。

与绕行距离相比，绕行时间是一个更复杂的变量，因为它同时取决于绕行距离和行驶速度，而出行需求对二者的影响方向相反。拼车对绕行时间的影响可能是一个值得深入研究的方向。

尽管本文观察到拼车存在规模经济和规模报酬递增，但网约车行程会如何影响交通拥堵仍有争议。一些研究（Rayle 等（2016）；Clewlow 和 Mishra（2017）；Zimmer（2016）；Feigon 和 Murphy（2016，2018））认为，拼车可能会减少交通拥堵；然而，Roy 等（2020）和 Erhardt 等（2019）发现，在旧金山，交通网络公司与交通拥堵加剧有关。Erhardt 等（2022）指出，截至2018年，交通网络公司导致中等规模都市区的公交和铁路客运量减少约10%。Schaller（2021）还表明，网约车服务会因增加空驶里程，以及用户从公共交通转向网约车服务，而大幅增加车辆行驶里程（VMT）；因此，建议在密集的城市中心区优先发展公共交通，而非网约车服务。因此，研究拼车对芝加哥市交通运行的影响具有重要意义。

此外，由于新冠肺炎危机，拼车服务于2020年3月停止，并于2022年6月才重返芝加哥。Lyft 和 Uber 的拼车集约服务分别于2021年8月和2022年6月重返纽约市。重新推出拼车服务后是否会出现类似的规模效应，以及疫情对拼车模式有何影响，是值得探讨的课题。

致谢

本研究获得美国国家科学基金会资助，项目编号为 CMMI-2052337 和 CMMI-2052512。

附录

A. 2019年1月至9月与10月至12月之间的系统性差异

本附录展示了1月至9月与10月至12月2019年期间的系统性差异，以及两个时期中存在的意外模式。

拼车行程数量

图8展示了2019年每个月的单人行程、拼车行程和总出行次数。单人行程数量在全年相对稳定，而已授权且匹配的拼车行程数量从1月到9月持续下降，10月后则趋于稳定。Taiebat等人（2022）发现了相同的模式，并指出，9月前拼车行程数量持续下降的一个可能原因是拼车行程总成本不断上升。不过，他们还表明，2019年9月前拼车行程的单位成本（每单位距离）保持稳定，仅在2019年10月后有所下降。

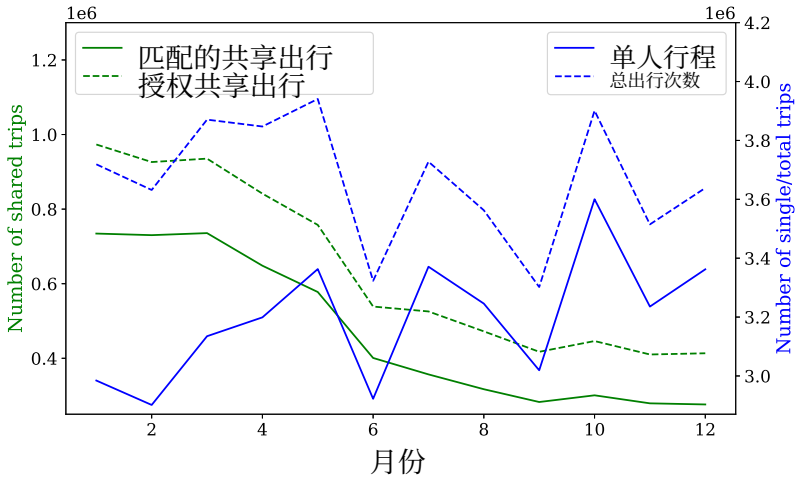
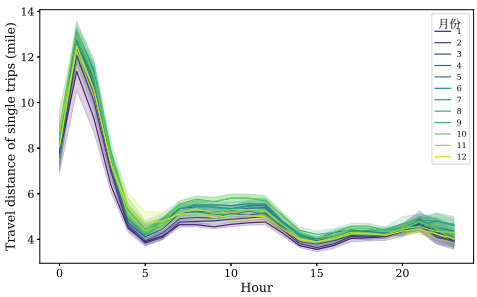


图8：每月出行次数的变化。

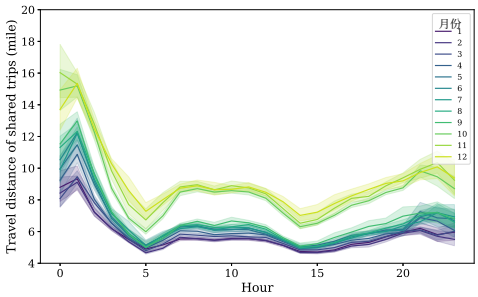
行驶距离与绕行距离

除了出行量之外，这两个时期拼车行程的行驶距离也有所不同。图9展示了2019年各月中，单人行程和拼车行程的平均行程长度如何随一天中的小时变化。请注意，单人行程的行驶距离在各个月份都保持一致，而拼车行程的行驶距离在这两个时期之间显著增加，尽管在每个时期内部相对稳定。

行驶距离的差异是否意味着10月之后人们倾向于拼车进行更长的行程？为检验这一点，图10展示了上车点与下车点OD对之间的绕行距离相对于实际行驶距离的情况。根据图9a，2019年单人行程的长度保持稳定，因此我们认为这可以合理估算OD对之间的距离。随后，我们将数据分为两组：1月至9月和10月至12月。图10显示，对于起讫点对之间的固定距离，1月至9月的绕行距离显著低于10月至12月。这表明，图9b中所示共享行程长度的增加，是由绕行距离增加而非



(a) 单人行程。



(b) 拼车行程。

图9：行程长度。

出发地与目的地之间的距离所致。此外，在一些城市，交通网络公司已不再在乘客的实际出发地和目的地接送乘客，而是让乘客步行至上车点，再从下车点步行离开，这种做法不应增加绕行距离。此外，图10显示，距离小于2英里的OD对，其绕行距离在1.5至2英里之间，高于预期。绕行距离如此大幅增加，使得10月至12月的出行距离数据令人怀疑。

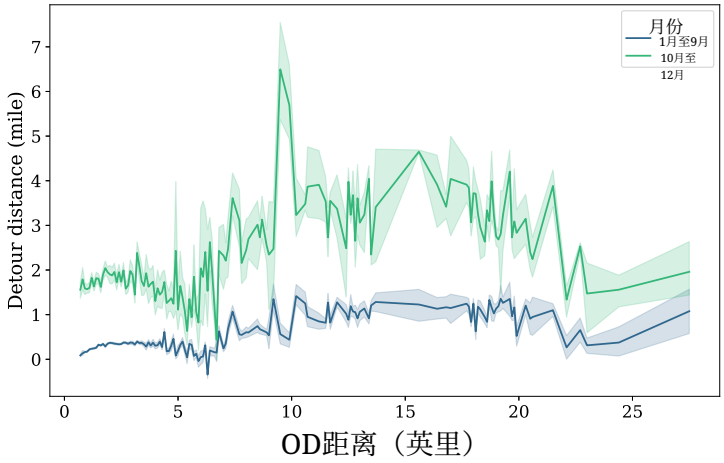


图10：绕行距离。

行驶速度

与图10类似，图11展示了与起讫点对之间距离相关的单人行程和拼车行程的平均行驶速度。这两个时期的单人行程平均速度非常相近。在1月至9月，拼车行程的行驶速度低于单人行程，这符合我们的预期。然而，在10月至12月，拼车行程的行驶速度反而更高，这是10月至12月期间存在的另一种异常模式。

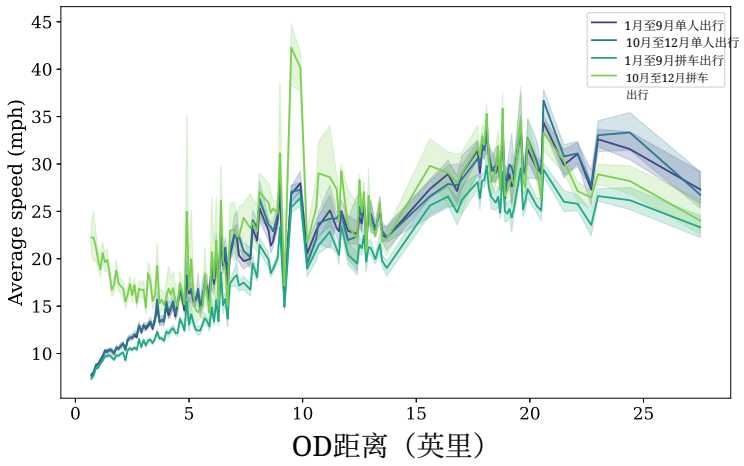
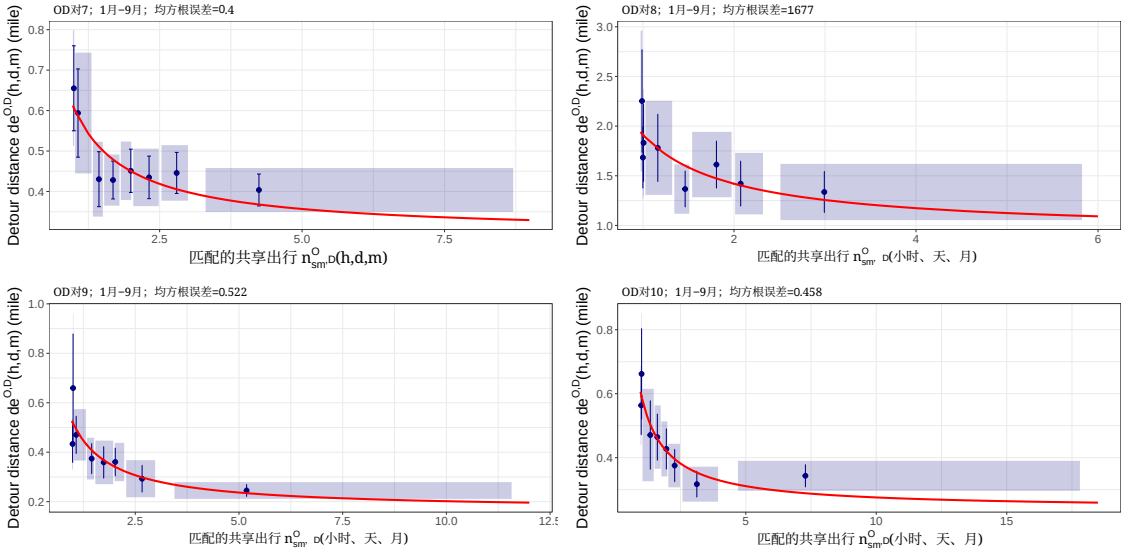


图11：速度比较。

遗憾的是，没有证据可以解释这些现象。10月至12月期间绕行距离和行驶速度中出现的异常模式，使得该时期的数据存在可疑之处。为避免混淆，本文排除了这一时期[4]。

B. 另外10个OD对的绕行距离

图12展示了 在出行次数最多的16个OD对中，其余10个起讫点对的绕行距离在1月至9月期间 2019年9月。



[4]尽管在2019年10月至12月期间的数据中观察到了意外现象，我们确认本文的所有研究结果在这一时期同样成立。

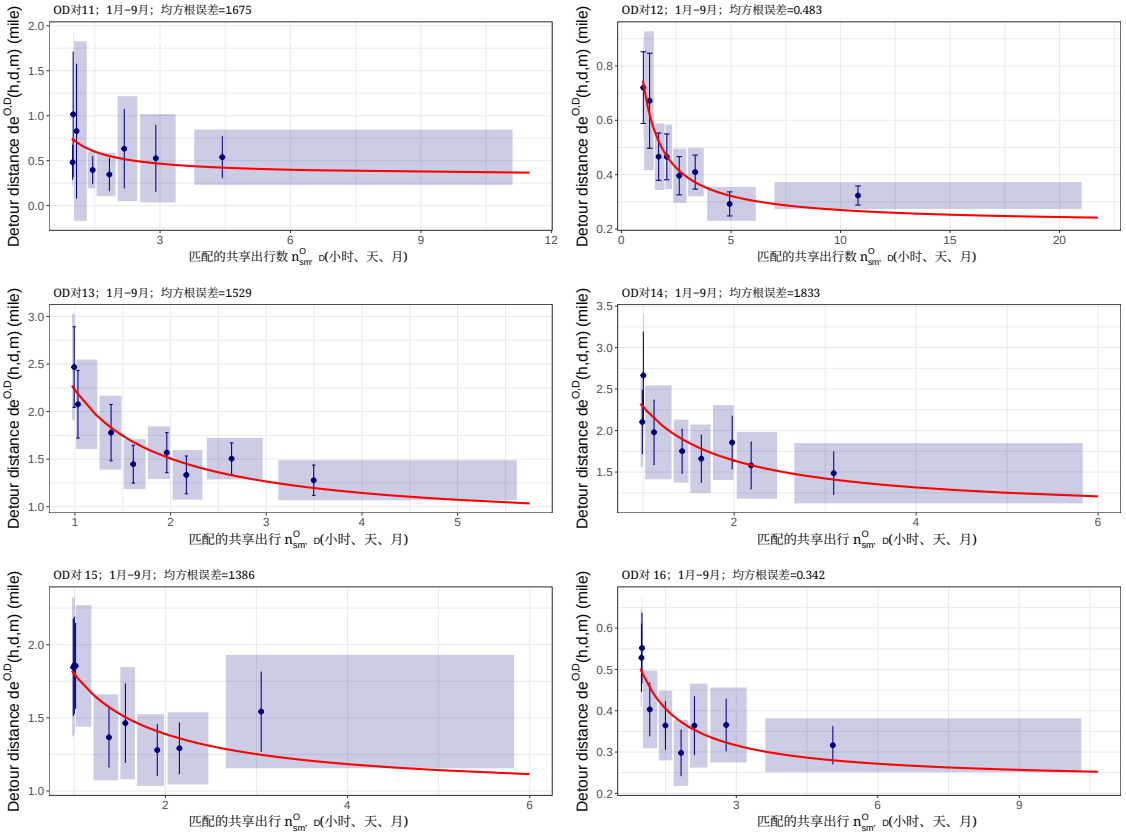


图12：十个OD对的绕行距离。

这10个起讫点对的回归结果如表5所示。当一个OD对每小时匹配的共享出行次数分别为1次和5次时，这10个起讫点对的绕行距离弹性分别介于（-0.704，-0.452）和（-0.321，-0.144）之间。

表5：绕行距离回归结果（英里）

	因变量：				
	绕行距离 $de^{O,\ D}(h,\ d,\ m)$				
	(7)	(8)	(9)	(10)	(11)
$\frac{1}{n^{O,s_m\ D}(h,d,m)}$	0.312***	0.985***	0.351***	0.355***	0.415
	(0.050)	(0.230)	(0.061)	(0.052)	(0.272)
常数	0.293***	0.927***	0.166***	0.240***	0.324*
	(0.032)	(0.181)	(0.044)	(0.035)	(0.195)
R^2	0.039	0.021	0.035	0.047	0.005
调整后 R^2	0.038	0.020	0.034	0.046	0.003
	(12)	(13)	(14)	(15)	(16)
$\frac{1}{n^{O,s_m\ D}(h,d,m)}$	0.519***	1.434***	1.298***	0.820***	0.266***
	(0.061)	(0.194)	(0.257)	(0.190)	(0.043)
常数	0.217***	0.790***	0.988***	0.979***	0.228***
	(0.034)	(0.130)	(0.188)	(0.144)	(0.031)
R^2	0.088	0.052	0.029	0.021	0.045
调整后的 R^2	0.087	0.051	0.028	0.020	0.044
显著性水平					*p<0.1；**p<0.05；***p<0.01

参考文献

Basso, L.J., Jara-D´ıaz, S.R., 2010. 城市公共交通补贴的理由与莫林效应。《运输经济学与政策杂志》44, 365–372. URL: https://ideas.repec.org/a/tpe/jtecpo/v44y2010i3p365-372.html. Cattaneo, M.D., Crump, R.K., Farrell, M.H., Feng, Y., 2019a. 分箱散点回归。arXiv 预印本 arXiv:1902.09615。Cattaneo, M.D., Crump, R.K., Farrell, M.H., Feng, Y., 2019b. 关于分箱散点图。arXiv 预印本 arXiv:1902.09608。Chen, X.M., Zahiri, M., Zhang, S., 2017. 理解按需出行服务的拼车行为：一种集成学习方法。《交通研究C辑：新兴技术》76, 51–70. doi:10.1016/j.trc.2016.12.018. Clewlow, R.R., Mishra, G.S., 2017. 颠覆性交通：网约车在美国的采用、使用及其影响。Dean, M.D., Kockelman, K.M., 2021. 共享网约车出行需求的空间差异及影响拼车的因素：来自芝加哥的启示。《交通地理学杂志》91, 102944. Diamond, P.A., 1982. 搜寻均衡中的总需求管理。《政治经济学杂志》90, 881–894. Erhardt, G.D., Hoque, J.M., Goyal, V., Berrebi, S., Brakewood, C., Watkins, K.E., 2022. 为什么美国公共交通乘客量下降？《交通研究A辑：政策与实践》161, 68–87. Erhardt, G.D., Roy, S., Cooper, D., Sana, B., Chen, M., Castiglione, J., 2019. 交通网络公司会减少还是加剧交通拥堵？《科学进展》5, eaau2670. Feigon, S., Murphy, C., 2016. 共享出行与公共交通的转型。项目 J-11，任务 21。Feigon, S., Murphy, C., 2018. 深入理解公共交通、共享出行与私人汽车之间的相互作用。技术报告。Lehe, L., Devunuri, S., Rondan, J., Pandey, A., 2021. 网约车税收。技术报告。伊利诺伊州交通中心。兰图尔。doi:10.36501/0197-9191/21-040. Lehe, L.J., Pandey, A., 2020. 超需求：一种具有后弯需求曲线的静态交通模型。《交通经济学》24. doi:10.1016/j.ecotra.2020.100182. Li, W., Pu, Z., Li, Y., (Jeff) Ban, X., 2019. 基于观测数据的拼车特征分析：以中国成都为例。《交通研究C辑：新兴技术》100, 330–353. doi:10.1016/j.trc.2019.01.030. Liu, H., Devunuri, S., Lehe, L., Gayah, V.V., 2023. 《拼车中的规模效应：以芝加哥市为例》的复制数据 doi:10.7910/DVN/GZMBJG. Mohring, H., 1972. 城市公交运输中的优化与规模经济。《美国经济评论》62, 591–604. doi:10.2307/1806101. Mortensen, D.T., Pissarides, C.A., 1994. 失业理论中的就业创造与就业破坏。《经济研究评论》61, 397–415. doi:10.2307/2297896. Rayle, L., Dai, D., Chan, N., Cervero, R., Shaheen, S., 2016. 只是更好的出租车吗？基于调查的旧金山出租车、公共交通和网约车服务比较。《交通政策》45, 168–178. Roy, S., Cooper, D., Mucci, A., Sana, B., Chen, M., Castiglione, J., Erhardt, G.D., 2020. 为什么交通拥堵日益严重？分解旧金山交通拥堵加剧的影响因素——确定交通网络公司的作用。《交通政策案例研究》8, 1371–1382. SAE International, 2018. 共享出行及相关赋能技术术语的分类与定义。技术报告。SAE International. Schaller, B., 2021. 拼车能否减少交通拥堵？来自 Uber 和 Lyft 的证据及其对城市的启示。《交通政策》102, 1–10. Shaheen, S., 2018. 共享出行：网约车与拼车集约的潜力。《三次革命：引导自动驾驶、共享和电动汽车迈向更美好的未来》，55–76doi:10.5822/978-1-61091-906-7. Shaheen, S., Cohen, A., 2019. 北美共享乘车服务：定义、影响与拼车集约的未来。《交通评论》39, 427–442. doi:10.1080/01441647.2018.1497728. Silvestre, J., 1987. 规模经济与规模不经济，载于：《新帕尔格雷夫经济学词典》。英国帕尔格雷夫·麦克米伦出版社，伦敦，第 1–7 页。Taiebat, M., Amini, E., Xu, M., 2022. 网约车出行中的拼车行为：一种机器学习推断方法。《交通研究D辑：交通与环境》103, 103166. Vassilakis, S., 1987. 规模报酬递增，载于：《新帕尔格雷夫经济学词典》。英国帕尔格雷夫·麦克米伦出版社，伦敦，第 1–8 页。doi:{10.1057/978-1-349-95121-5{_}746-1}. Wang, Z., Chen, X., Chen, X.M., 2019. 拼车正在塑造年轻人的出行行为：来自网约车平台比较调查的证据。《交通研究D辑：交通与环境》 75, 57 71 10.1016/j.–。trd.2019.08.017 doi:。Zhu, Z., Qin, X., Ke, J., Zheng, Z., Yang, H., 2020. 有接驳及竞争型拼车服务条件下的多模式通勤行为分析。《交通研究A辑：政策与实践》132, 713–727. doi:10.1016/j.tra.2019.12.018. Zimmer, J., 2016. 第三次交通革命。Medium，9月18日。