



浙江财经大学

硕 士 研 究 生 毕 业 论 文

题目：基于个人动态最优居住地选择视角的劳动力回流行为考察

| | |
|---------|--------------|
| 学生姓名： | 李晨逸 |
| 学 号： | 230507031002 |
| 指导教师： | 吴意云 |
| 所在学院： | 经济学院 |
| 专业名称： | 西方经济学 |

摘要

基于个人动态最优居住地选择视角的劳动力回流行为考察

摘要：随着经济的发展与社会工业化带来的社会关系剧变，劳动力迁移形成有规律的迁移模式是必然的。不同于我国研究中的传统城乡二元对立语境，本文基于理性预期构造了动态最优住址选择模型，基于 CLDS 数据对 2010 年至 2018 年间我国各省的人口流动进行实证检验，并得到的结论是（xxx、xxx、xxx）。本文的贡献在于（xxx、xxx、xxx）。

关键词：动态迁移决策模型、劳动力迁移、收入引致、迁移摩擦

Abstract

A Study of Return Migration Behavior Based on Dynamic Optimal Residential Location Decisions

Abstract: Here is the english version of abstract

keywords: a, b, c

目录

| | |
|-------------------------|----|
| 摘要 | i |
| Abstract | ii |
| 第一章 引言 | 1 |
| 第二章 文献综述 | 5 |
| 2.1 相关概念梳理 | 5 |
| 2.2 空间均衡方法 | 6 |
| 2.3 个体微观视角与动态选择方法 | 8 |
| 2.4 对回流迁移的研究 | 10 |
| 2.5 影响劳动力迁移的因素 | 12 |
| 第三章 理论模型 | 15 |
| 3.1 模型的基本设定 | 16 |
| 3.2 收入效用函数 | 18 |
| 3.3 户籍惩罚、家乡溢价与地区舒适度 | 21 |
| 3.4 迁移成本 | 22 |
| 3.5 状态转移概率 | 24 |
| 3.6 条件选择概率函数 | 25 |
| 3.7 劳动力迁移决策 | 26 |
| 第四章 实证处理 | 29 |
| 4.1 处理未观测到的异质性 | 29 |
| 4.1.1 离散化近似 | 29 |
| 4.1.2 有限混合 | 30 |
| 4.2 引入机器学习方法处理干扰函数 | 31 |
| 4.3 结合 EM 算法的 NFXP 估计策略 | 32 |
| 4.3.1 EM 框架 | 33 |
| 4.3.2 似然函数的构建 | 35 |

| | | |
|-------------|-----------------------|-----------|
| 4.4 | 数据与变量 | 36 |
| 4.5 | 初始值选择 | 41 |
| 第五章 | 估计结果 | 44 |
| 5.1 | 基准回归与子样本分析 | 44 |
| 5.1.1 | 从回归结果解读迁移摩擦 | 46 |
| 5.2 | 模型的拟合检验 | 46 |
| 5.3 | 机制分解 | 46 |
| 第六章 | 与宏观桥接 | 48 |
| 6.1 | 抽样映射与 | 48 |
| 6.2 | 外部性机制 | 48 |
| 6.3 | 模型校准 | 48 |
| 6.4 | ABM 模型的检验 | 49 |
| 6.5 | 反事实参数估计 | 50 |
| 第七章 | 结论与展望 | 52 |
| 附录 A | 详细的语言谱系树 | 53 |

1 引言

纵观古今内外，“人往高处走，水往低处流”的规律总是屡试皆准。早在上世纪末，以 Krugman (1991)和 Fujita et al. (1999)为代表的新经济地理学就指出经济活动的集中会产生规模经济和网络效应，促成产业集聚。经济的聚集会导致区域不平衡发展，人们倾向于从“边缘”地区流向产业集聚、薪资较高和就业机会多的“中心”地区，形成吸引效应，这使得劳动力迁移有了移动的规律。欧美国家作为世界上人均 GDP 最高的区域，吸引大量来自其他相对落后国家的居民。根据世界银行（World Bank Group）和美国移民委员会（American Immigration Council）公布的数据，2023 年美国的人均 GDP 为 82769 美元，外来人口达到了 4780 万，移民占美国人口的 14.3%，比 1970 年的 4.7% 增长了约三倍。这一由经济集聚驱动的中心-边缘迁移模式，在全球范围内得到印证，而在快速城市化的中国，其表现尤为深刻和复杂。改革开放以来，随着现代化交通工具的普及和人口流动政策的放宽，劳动力自由流动在我国成为可能。解放的经济活力逐渐形成了“东富西穷、南富北穷”的局面。经济发展的不平衡不仅表现在地区收入差距上，也体现在人口分布的变化中。富裕地区吸引了大量来自相对贫困地区的劳动力。国家统计局 2020 年发布的《第七次人口普查》显示，流动人口达到 3.76 亿，占全国总人口比重分别为 34.90% 和 26.62%，较 2010 年分别上涨 88.52% 和 69.72%。广东跨省流入高达 2962.21 万人，浙江也达到 1618.65 万人，上海跨省流入人口为 1047.97 万人。这三地的跨省流入人口数量位居前三。此外，北京流入 841.8 万人，位居第五。同时，我国也存在多个人口输出大省¹，例如安徽省净向外输出约 911 万人，占本省户籍人口的 13.29%；四川省净向外输出约 956 万人，占本省户籍人口的 10.63%；河南省净向外输出约 565 万人，约占本省户籍人口的 7%。

在我国激烈的劳动力迁移浪潮中大致存在以下规律。首先，劳动力净流入的区域符合常识中的迁移规律，劳动力从农村流向城市是迁移的主流趋势。大城市由于更高的工资水平、更丰富的就业资源、更高质量的基础服务设施，具有强大的虹吸效应，这一点与新经济地理学学者提出的观点相吻合。在可以预见的未来，这种劳动力向经济发达地区集中的趋势仍将持续。其次，永久迁移与暂时性迁移之间存在显著差异。第七次人口普查显示，2020 年时我国人户分离人口已达 4.93 亿，占总人口的 34.16%。（xxx）并且，尽管总体上劳动力向发达地区流动，仍有部分劳动力回流到欠发达地区。这种反直觉的

¹该数据来源于 2010 年的《第六次人口普查》。

现象已被部分学者注意到（石智雷等 (2012)，任远和施闻 (2017)），表明某些群体在外迁后因各种原因选择返回原居住地，这揭示了迁移决策背后更为复杂的动机。DaVanzo (1983)就指出过劳动力个体迁移行为的“反复无常”规律——尽管大多数个体从未迁徙，但迁徙的个体很可能会再次迁徙，通常会返回原籍地。这意味着迁徙决策应该被看作是一系列地点选择，个体知道有机会修改或逆转那些效果不佳的迁徙。

对于劳动力流动现象的研究，我国学术界长期围绕城乡二元分析框架与空间均衡模型展开。作为典型的发展中经济体，我国自计划经济时期形成的城乡二元结构构成了劳动力迁移的制度基础。改革开放后，工业化进程产生的劳动要素需求与农村剩余劳动力释放一拍即合，这一过程在学术研究领域直接映射为对传统二元经济理论的引用。其中，在 Lewis (1954)二元对立模型基础上，Todaro (1969)通过引入失业率与预期收入，突破了无限劳动供给假设的刚性约束，其“即使存在失业风险，人口仍会因预期收入差距迁移”的核心命题，恰与中国城市化进程中农民工“候鸟式迁移”的特征契合，成为解释中国农民工流动现象的核心理论工具。而后，Harris et al. (1970)进一步将城市正规部门与非正规部门纳入分析框架，通过工资刚性与就业概率的动态调整机制，构建起解释发展中国家城市失业与农村劳动力持续涌入并存现象的理论模型。这种强调制度分割与部门差异的分析视角，为中国学者解析户籍制度、土地制度等特殊约束条件下的劳动力流动提供了重要切入点。如熊彩云 (2007)构建的农民工定居决策模型，通过引入城市拉力（就业机会、公共服务）与农村推力（土地保障弱化、收入差距）的交互作用，拓展了传统二元模型的解释维度。黄忠华等 (2014)通过嵌入土地保险功能变量，钟水映等 (2015)纳入制度变迁因素。部分文献虽然淡化了城乡之间的对立，但依然在均衡状态下分析工资和租金的确定，同时考虑移民流动对这些结果的影响。踪家峰等 (2015)构建的三部门 Rosen-Roback 模型揭示，中国大城市存在显著工资溢价且技能异质性导致差异化集聚收益：高技能劳动力通过知识溢出获取短期增长红利，而低技能群体则需经历长期调整方能获益。王丽莉和乔雪 (2020)，王丽莉 (2023)，王丽莉和唐为 (2024)通过将土地供给管制、迁移壁垒等制度变量内生化的方法，定量识别出行政干预对劳动力空间配置的扭曲效应：建设用地指标的区域错配加剧东部大城市住房供给弹性不足，推高生活成本并阻碍生产率导向的人口集聚，导致 2010 年经济总产出损失达 3%-4%。刘华仁 (2024) 则构建包含人力资本溢出效应的量化空间均衡模型，证明高技能劳动力的区域再配置可通过生产率与公共服务双重渠道提升社会福利，这为人才政策优化提供了理论依据。

尽管现有研究取得显著进展，但传统劳动力迁移研究的局限性却愈发明显。首先，动态视角的缺位限制了理论的解释力。主流文献多依赖静态或比较静态分析，将迁移决

策简化为单期最优选择，忽视了个体在生命周期内因人力资本积累、预期调整和制度环境变迁而产生的动态交互。例如，韩其恒等(2018)虽尝试通过 OLG 模型捕捉迁移行为的代际动态，但其模型仍未能内生制度约束与技能积累的交互效应。其次，微观基础的薄弱削弱了研究的政策适用性。许多研究基于宏观数据分析迁移的总体趋势，未能深入探讨个体异质性因素（如风险偏好、社会网络或数字技能）对迁移路径分化的影响，导致模型缺乏行为依据，难以支持精准的政策仿真。最后，理论范式的滞后使得现有模型难以适应劳动力流动的新特征。随着城-城流动规模扩大和远程就业的兴起，传统城乡二元对立的分析框架无法解释迁移的可逆性、多向性和地理重构效应。这些局限使得学界难以回答两个核心问题：在面临户籍限制带来的未来不确定性与地区间显著的收入差距时，不同人力资本水平的个体是如何动态地权衡其迁移、定居或回流的决策路径的？制度变迁如何通过重塑迁移的成本-收益结构影响个体的跨期决策？

针对上述问题，本研究提出从静态均衡向动态演化、从宏观相关性向微观行为基础、从单向城乡迁移向多维空间重构的理论转向。为此，本文构建了一个动态离散选择模型，将劳动力回流建模为一个马尔可夫决策过程（Markov Decision Process, MDP）。与静态模型将迁移视为一次性选择不同，MDP 框架将迁移刻画为一个关于信息和预期的序列学习与适应过程，这更符合劳动力在真实世界中“走一步看一步”的决策本质。该模型假设个体在有限生命周期内基于效用最大化原则选择居住地，效用函数综合考虑了地区吸引力（如工资和公共服务）、制度摩擦（如户籍限制）以及个体异质性（如乡土黏性和人力资本水平）。通过非参数混合估计方法，本研究量化以下机制：（1）户籍制度改革和土地确权如何通过提升预期稳定性增强迁移的持久性；（2）人力资本积累速率的差异如何导致迁移路径的分化，例如高技能个体更倾向于城市间的多次迁移；（3）远程就业的普及如何通过降低物理迁移成本重塑迁移的地理指向和频率。模型不仅为解析中国劳动力流动的复杂性提供了新的分析工具，还通过揭示迁移决策的微观驱动因素，为优化区域发展和以人为核心的城镇化政策提供了实证依据。

本文的重要性在于，在当前工业化加速和经济转型的背景下，劳动力跨区域迁移问题愈发显得关键而复杂。人口流动不仅直接影响城市化进程、区域经济平衡和资源配置效率，还对社会福利和经济结构的长期调整产生深远影响。尤其在中国这一快速发展的经济体中，劳动力回流现象既是城市化进程的缩影，也是城乡协调发展的关键变量。深入理解和准确预测劳动力迁移趋势，不仅有助于把握经济发展的内在动力，还为政策制定提供了科学依据，以实现区域协调发展和社会整体进步。

本研究从动态决策的视角切入，系统分析劳动力回流的规律及其驱动机制，特别强

调收入差异作为核心经济因素的作用。与此同时，模型综合考虑了就业机会、生活成本、制度摩擦和文化因素的交互效应。例如，户籍制度作为中国特有的制度约束，可能通过限制公共服务获取提高迁移成本，从而影响个体是否选择永久定居城市。同样，地区间的文化差异（如方言）可能通过乡土黏性影响迁移决策的稳定性。通过构建理论模型并结合实证分析，本文旨在回答以下问题：地区间的收入差异如何激发劳动力迁移？收入差异如何与其他因素（如公共服务供给和户籍限制）共同作用于迁移决策？通过何种政策干预可以引导劳动力合理流动，从而优化资源配置、促进城乡协调发展并提升社会福利？

正如 Desmet et al. (2013) 在城市规模与福利分析中所指出的，理解决定人口分布的各种力量对于回答一系列关键问题至关重要：这些力量在决定城市规模分布中的相对重要性如何？如果各地区在公共服务、技术水平或制度摩擦方面趋于一致，人口将如何重新配置？这种重新配置对社会福利的总体影响是什么？这一论述启发本研究将收入差异置于更广泛的动态框架中，探讨其如何通过影响迁移决策重塑人口分布和区域经济格局。如果地区间的制度和经济差异缩小，劳动力流动和城市规模分布可能发生显著变化，进而影响整体社会福利。本研究通过量化这些机制，为劳动力迁移的理论和实证分析提供了一个综合框架。

为实现上述研究目标，本文结构安排如下：第二章对国内外相关文献进行系统综述，梳理劳动力迁移的理论演进和实证成果，分析现有研究的不足，并明确本文的创新点。第三章构建动态离散选择模型，详细阐述模型的基本假设、机制设计以及与现实经济现象的关联，推导劳动力迁移存在性的理论命题。第四章推导实证分析所需的似然函数，说明模型简化的必要假设（如支撑点离散化处理未知分布），并介绍数据来源（中国家庭追踪调查和地区年鉴）、变量构造和 Python 实现的计算方法。第五章展示实证分析的主要结论，通过对不同样本子集的检验和命题验证，评估模型在解释中国劳动力回流中的适用性和稳健性。第六章总结研究的主要发现，讨论模型的局限性（如理性人假设的适用性、函数形式的潜在任意性），并提出未来研究方向，例如引入收入分配结构或跨省迁移的网络分析。此外，更为详细的数学证明、高性能程序代码和数据处理流程均附在文末的附录中。

2 文献综述

劳动力流动作为经济社会发展的重要表征，其研究要点在于提供合适的方法和寻找符合现实逻辑的影响因素。本文通过系统梳理国内外近几十年来的经典文献与前沿成果，试图厘清学术脉络的演进逻辑，为后续实证研究提供标准。在研究方法层面，现有文献呈现出从静态分析向动态追踪的范式转变。就影响因素而言，学者们已逐步突破传统经济变量的单一解释，将制度约束、社会网络、文化认同等非经济维度纳入分析体系。

2.1 相关概念梳理

劳动力回流是指原先从一个地区迁出的劳动力再次返回原籍地的现象。这种迁移可被视为可撤回的 (retractable) 或短暂的 (temporary)，与传统静态迁移模型所假设的单向永久性迁移存在显著差异。

劳动力回流研究具有以下特点：首先，从历史维度看，它是一个相对较新的现象，在人类工业化历程中出现较晚；其次，它挑战了传统人口迁移理论中从欠发达地区向发达地区单向流动的基本假设；再次，对于后发工业化国家而言，这一现象具有重要的政策意义。

在传统经济理论框架下，当目标地经济衰退导致预期收益低于原籍地时，返迁决策符合理性选择理论；然而当目标地仍能提供显著收入溢价时，劳动力选择返迁则构成理论悖论。这一现象值得深入研究，特别是对于后发工业化国家而言。

蔡昉等 (2001) 等学者指出，后发国家通常采取非均衡发展战略，通过制度化的城乡二元结构形成工农产品价格剪刀差，实现资源从传统农业部门向现代工业部门的系统性转移。这种情况下，劳动力回流可能不利于资本深化与工业化进程。中国的工业化路径具有特殊性，融合了社会主义国家计划经济与市场经济转型的特点。20 世纪 50 至 80 年代，由于资源稀缺，我国建立了以户籍制度为核心的空间资源配置体系，将有限资本集中于少数城市，形成公共服务的高度集聚。改革开放后，我国借鉴东亚发展型国家经验，采取出口导向型发展战略，特别是 2001 年加入世贸组织后，充分发挥人口红利的比较优势。林毅夫等 (2003) 为这一发展路径提供了系统的理论支撑。在快速工业化进程中，大量人口涌入城市，形成了沿海发达地区持续吸纳内陆剩余劳动力的迁移格局。然而，我国劳动力市场中却出现明显的回流现象，即农村居民进城打工后又返回原籍，形成农民工返潮现象。这一现象与后发国家工业化进程中的劳动力配置目标存在理论上的

背离。

多项研究表明,户籍制度障碍可能是影响劳动力定居的首要因素。任远和邬民乐(2006)指出,以户籍制度为依托的流动人口管理制度及相关社会福利制度对流动人口的限制与排斥对其社会融合有根本性影响。陆益龙(2008)认为即使经历了深刻改革,户籍制度仍然对劳动力自由流动造成客观阻碍,制约经济高质量发展。另一种观点认为,源于1994年分税制改革的土地财政模式可能是劳动力回流的重要原因。陈映芳(2005)、聂辉华等(2013)及郁建兴等(2012)等研究认为,土地出让金作为地方政府重要财源,导致政府维持较高地价,推高房价,增加城市居民住房负担,从而影响劳动力的定居决策。

除户籍与土地财政外,影响劳动力回流的因素还可能包括原籍地经济发展与就业机会增加、工资差距缩小、生活成本差异变化;家庭纽带与社会融入因素;以及公共服务可及性、返乡创业政策等制度因素。综合来看,劳动力回流问题涉及多种制度与市场机制的交互作用。传统的线性迁移模型难以解释复杂现实情况,需引入更具动态性和多维视角的方法进行深入研究。接下来对劳动力流动研究的主要方法进行梳理,以期为本研究提供合适的方法论支持。

2.2 空间均衡方法

根据 Jia et al. (2023)的研究,劳动力迁移问题可采用两种研究方法,其中一种是空间均衡方法。该方法以迁移后形成的新空间一般均衡为核心,主要关注人口流动如何影响本地工资、区域房价等变量的市场出清。

空间均衡概念可追溯至20世纪中叶, Samuelson (1952)提出的“空间价格均衡”框架考虑了运输成本对不同地点价格的影响,奠定了空间经济学基础。Tiebout (1956)提出“用脚投票”理论,假设消费者在无迁移成本和信息完全的条件下,会自由选择最能满足其偏好的社区。Harris et al. (1970)则构建了研究农村到城市迁移的两部门内部贸易模型,探讨了城乡工资差异的均衡状态。

受城市经济学发展影响,20世纪60-70年代, Alonso (1964)、Muth (1969)和 Mills (1967)开创的单中心城市模型考察了城市内部空间结构,为城市间移民研究奠定基础。Rosen (1974)提出的享乐定价模型能够通过观察特定地点便利设施对工资和租金的影响来量化这些设施的价值,为移民研究提供了重要工具。基于 Rosen 模型, Roback (1982)正式确立了移民空间均衡模型,该模型假设个人会迁移至效用最高的地点,同时考虑金钱因素(工资、租金)和非金钱因素(便利设施)。在均衡状态下,工资和租金会调整以平衡不同地点的效用,使得没有人能通过迁移提高效用。这一模型成为理解便利设施如何影

响移民模式和城市增长的基础框架。

20 世纪 80-90 年代，空间均衡方法进一步扩展，纳入了住房市场动态、非贸易商品和异质性主体等因素。Glaeser, Gottlieb (2009)强调住房供应弹性在决定城市成功表现(人口增长或收入增长) 方面的关键作用。Moretti (2011)提出了假设工人流动性有限且住房供应非固定的一般均衡模型，更贴近现实情况。Diamond (2016)构建了结构性空间均衡模型，研究技能分类增加的原因及其福利影响，考虑了偏好和技能的异质性。Coen-Pirani et al. (2010)则开发了动态一般均衡模型，强调个体异质性在迁移决策中的作用。近二十年来，空间均衡方法整合到定量空间经济学中，Ahlfeldt et al. (2015)和 Redding et al. (2017)扩展了 Rosen-Roback 框架，引入市场准入项，提供了内生价格作为基本面和地理因素函数的对数线性方程，更好地捕捉现代经济复杂性。Glaeser, Gyourko et al. (2014)构建了动态的、线性的、理性的一般均衡模型，与住房市场典型事实相一致。Albert et al. (2022)则将该方法应用于国际移民研究，记录了移民如何受原籍国支出影响而选择生活成本高昂的城市。

表格2-1概述了空间均衡方法在移民研究中的发展历程。

表 2-1 使用空间均衡方法研究人口流动的历史

| 时期 | 关键发展 |
|--------------|--|
| 1950 年代 | Samuelson (1952)引入了空间价格均衡，通过运输成本平衡不同地点的供需。奠定了空间均衡经济学的基础，但最初侧重于商品市场。 |
| 1960-1970 年代 | Muth (1969)和 Mills (1967)开发了关注通勤和住房选择的城市内模型，影响了城际移民研究。为区位选择提供了空间框架，为移民应用奠定了基础。 |
| 1970 年代 | Rosen (1974)建立了享乐定价模型，通过工资和租金等观察到的价格来评估区位属性（例如便利设施），能够量化区位选择中的非金钱因素，为移民研究奠定了基础。 |
| 1980 年代 | Roback (1988)正式建立了一个一般均衡模型，其中工资、租金和便利设施使不同地点的效用均衡，从而解释了移民模式。成为研究移民的基石，将经济激励与区位决策联系起来。 |
| 1980-1990 年代 | Glaeser, Gottlieb (2009)将住房供应弹性、非贸易商品和异质性主体纳入模型。通过解决住房动态和不同的移民偏好增强了现实性。 |
| 2010 年代 | Redding et al. 2017整合市场准入和贸易成本，扩展 Rosen-Roback 模型，使用对数线性方程对工资和人口进行建模。改进了基础设施和贸易对移民影响的政策分析。 |
| 近期 | 拓展：国际移民、信息约束、多部门和异构代理模型 |

国内学者也运用空间均衡模型研究人口流动问题。梁若冰等 (2008) 基于 Tiebout 模型验证了中国城市住房价格与地方公共品供给关系。刘华等 (2024) 构建了包含异质性劳动力生产率溢出和公共服务溢出的空间一般均衡模型, 分析了人力资本空间配置优化的社会福利效应。王丽莉和乔雪 (2020) 结合空间均衡模型与地级市数据研究人口迁移对劳动力资源配置的影响。王丽莉 (2023) 分析了土地供给政策对劳动力空间配置效率的影响, 指出政府对建设用地指标的管控决定住房供给弹性, 东部大城市的严格规制阻碍了人口向高生产率地区转移。赵方等 (2017) 基于 Diamond (2016) 的模型研究中国劳动力迁移机制, 发现工资水平仍是主要影响因素, 但城市舒适度对流动人口也具有重要作用。

尽管空间均衡方法影响深远, 但其完全流动性和同质主体等假设过于简化了现实复杂性, 对均衡条件的依赖也难以充分捕捉短期迁移动态或社会网络和文化因素的作用。此外, 该方法在宏观层面有效, 但难以解释微观个体行为, 需要非均衡方法作为补充。这些局限性促使研究者发展更加细致的微观视角方法, 以更好地理解劳动力迁移决策过程。

2.3 个体微观视角与动态选择方法

与空间均衡方法所关注的宏观均衡不同, 个体微观视角研究方法着重于分析单个经济主体的决策过程。这种方法类似于交通道路模拟中的元胞自动机微观方法, 而空间均衡则对应于流体力学宏观方法。

个体视角的劳动力流动研究始于 Sjaastad (1962)。Sjaastad 提出了一个模型, 假设个人根据迁移的成本和预期回报做出地点决策。这种观点将移民视为投资行为, 强调了移民决策的动态特性。从生命周期角度看, 个人决策 (如储蓄、教育、婚姻) 与移民选择相互关联, 构成了一个统一的决策框架。Mincer (1978) 进一步发展了这一理论, 考虑了家庭联合决策的情况, 认识到家庭最优地点选择可能与个体配偶的最优选择存在差异。Kennan et al. (2011) 提出的个人移民选择模型是这一领域的重要发展。该模型允许个体技能、位置偏好和迁移成本的异质性, 并将移民决策作为最优搜索过程处理, 假设个人在扣除迁移成本后最大化预期终身收入。该研究的重要贡献在于将动态离散选择模型引入移民问题研究。

动态离散选择 (Dynamic Discrete Choice, DDC) 模型是研究个体或家庭如何在多个潜在居住地之间做出最优选择的理论框架, 适用于分析人口迁移、城市发展和住房市场动态。DDC 模型具有以下特征: 离散时间序列的动态性; 单个决策者面临有限且离散的选择集; 选择持续存在; 马尔可夫决策过程; 随机决策环境; 基于微观数据的显示偏

好。这些特征使 DDC 成为研究动态迁移决策过程的理想工具，特别适合分析可撤回的迁移行为。

离散选择模型的理论源头可追溯至 Thurstone (1927)提出的比较判定定律，该定律从心理激励角度解释了选择行为。Marschak (1960)将这种感知刺激解释为效用，提出了随机效用最大化（Random Utility Maximization, RUM）模型，奠定了现代离散选择模型的理论基础。McFadden (1973)则将二元逻辑模型扩展为条件逻辑模型，使其适用于多种选择情境。

动态离散选择模型的发展与动态规划理论密切相关。Bellman (1957)提出的动态规划为序列决策问题提供了数学框架，而 Blackwell (1962)则为马尔可夫决策过程（Markov Decision Process, MDP）建立了理论基础。经济学中的 DDC 模型研究始于 20 世纪 80 年代中期。Rust (1987)被视为该领域的开创性工作，提出了嵌套不动点（Nested Fixed Points, NFXP）算法进行最大似然估计。Eckstein et al. (1989)在此基础上构建了已婚女性劳动力参与率和生育率的动态模型，考虑了工作经验对工资的影响。Keane et al. (2011)系统介绍了离散选择动态规划（DCDP）模型的结构估计方法及其在劳动经济学中的应用。

由于 DDC 模型计算量巨大，Hotz et al. (1993)提出了条件选择概率方法，避免了反复求解完整动态规划问题的需要。近年来，Su et al. (2012)提出的带均衡约束的数学规划（MPEC）进一步减轻了计算负担。表格2-2总结了 DDC 模型方法论的关键发展。

表 2-2 不同作者对于动态离散选择模型的贡献

| 年份 | 作者 | 贡献 |
|--------|---------------------------------|-----------------------------|
| 1987 年 | John Rust | 开发了嵌套不动点算法，提出经典的公交车发动机更换模型。 |
| 1993 年 | V. Joseph Hotz、Robert A. Miller | 开发了条件选择概率，作为领先的非解估计方法。 |
| 2012 年 | Che-Lin Su、Kenneth Judd | 提出带均衡约束的数学规划方法进行估计。 |

近年来,研究重点转向融入更为现实的行为假设并拓展应用领域。Utara et al. (2024)探讨了动态不一致性问题，Heckman et al. (2007)将动态选择模型与处理效应估计相结合，增强了政策分析能力。计算能力的进步使高维状态和多智能体模型成为可能，推动了该方法在卫生经济学、环境经济学和市场营销等领域的应用。

如 Jia et al. (2023)所述，使用 DDC 模型探究劳动力迁移问题本质上是一种非均衡方

法，与空间均衡方法形成鲜明对比。两者代表了不同的分析范式：空间均衡方法基于新古典经济学的一般均衡理论，强调区域间要素调整与市场出清，其研究视角聚焦于城市或区域层面的整体均衡状态。该方法通常假设劳动力具有同质性或有限异质性，在计量分析中多采用结构方程估计或联立方程系统，依赖区域层面的汇总数据，典型应用包括区域经济收敛性研究以及税收与公共服务的政策分析。与之相对，基于 DDC 的最优居住地选择模型则以微观经济学的效用最大化理论为基础，关注个体或家庭在异质性偏好下的决策过程。该方法明确建模个体差异性和偏好多样性，在实证层面主要采用离散选择模型（如条件 Logit、嵌套 Logit 等），并需要微观个体数据作为支撑。其典型应用场景涵盖农民工市民化决策、高技能人才流动等侧重微观行为机制的研究领域。两种方法在理论基础、分析视角、异质性处理、计量工具及数据需求等方面均存在系统性差异，分别体现了宏观均衡框架与微观决策逻辑的学术分野。

对于劳动力回流这一具有暂时性迁移特征的研究问题，微观方法提供了更为适合的分析框架。通过动态离散选择模型构建最优居住地序列选择模型，可以有效捕捉个体在不同时期的迁移决策及其影响因素。

2.4 对回流迁移的研究

通常进行科学研究时会强调国内与国外文献的融合，但在劳动力回迁问题上，国内国外在具体议题上存在较大不同。具体而言，在我国独特国情下，劳动力回流问题指的都是农民工返潮现象，而国外研究大致都是研究在一个自由流动的背景下劳动力如何自由选择的。在这个前提下，我们有必要先对国内外的学术工作分开进行总结，接着再从国外文献的研究方法中汲取有用的共性以服务于国内研究课题。

国外关于劳动力回流（return migration）的研究重点关注自由流动背景下个体和家庭如何在不同地点间进行迁移决策。与我国不同，国外研究通常聚焦于国际移民的往返迁移、临时性迁移及其经济和社会驱动因素。

早期研究如 Gmelch (1980)综合了南欧、东欧和加勒比地区的案例，指出 20 世纪初欧洲移民中有四分之一返回原籍国。Wyman (1993)则通过历史记载分析了 1880-1930 年间从美国返回欧洲的移民，归因于怀旧情绪和本土主义怨恨等因素。这些研究奠定了对回流迁移现象的基本理解。

Sjaastad (1962)首次将移民定义为一种人力资本投资，强调个人在预期回报超过成本时选择迁移。这一静态框架为后续研究提供了基础，但其局限性在于无法处理多目的地选择问题。Tunali (2000)尝试扩展至动态框架，但仍局限于二元选择。Dierx (1988)则结

合面板数据，分析家庭人力资本分布对迁移决策的影响，弥补了单期模型的不足。Dahl (2002) 通过罗伊模型探讨了自我选择迁移机制，解释了美国劳动者高流动性与地区教育回报差异之间的关系。Gemici (2007) 进一步构建包含家庭内部协商的动态模型，发现家庭纽带会抑制人口流动性并减缓工资增长。在以上早期模型的基础上，Kennan et al. (2011) 他们利用全国青年纵向调查 (NLSY) 中针对高中教育白人男性的面板数据，开发了一个动态离散选择模型，该模型允许在多个备选方案中进行最优的地点决策序列。该模型通过考虑收入前景来捕捉回归移民，发现地域工资差异和对更佳地点匹配的追求会影响决策，包括在当前收入不利的情况下回归。这标志着一项重大进展。在这之后，Dustmann et al. (2016) 强调了临时迁移的重要性，提出一个通用理论框架，分析技能积累、技能回报率差异及消费偏好如何驱动跨国往返迁移。研究表明，即使缺乏外生冲击，这些内生机制仍能引致回流行为。Venator (2022) 整合家庭动态和空间因素，改进估计技术，推动了该领域方法论的进步。

国内关于劳动力回流的研究主要集中在农民工返潮现象，分析其影响因素及经济社会效应。然而，现有研究多停留在静态选择框架下，采用简单的 Logit 或 Probit 回归模型进行实证分析。王智强等 (2011) 利用中国健康与营养调查 (CHNS) 数据，通过 Probit 模型分析了农村劳动力迁移决策的影响因素。研究表明，婚姻状况、健康水平和娱乐偏好等变量显著影响迁移决策，而当前收入的作用不明显。此外，已婚女性倾向于与配偶共同迁移，这种行为加剧了农村空巢老人问题。研究还发现，提高教育水平是促进农村劳动力迁移的最有效手段，尤其是高等教育和职业教育。石智雷等 (2012) 基于湖北和河南两省的农户调查数据，利用 Multinomial Logistic 模型探讨了家庭禀赋对农村迁移劳动力回流的影响。研究发现，家庭人力资本和社会资本在不同阶段对迁移决策产生差异化影响：人力资本丰富的家庭更倾向于让劳动力留在或回流农村，但当人力资本达到一定水平后，劳动力又更愿意外出就业；社会资本则在初期促进外出务工，但在高水平时推动劳动力回流。此外，家庭经济资本兼具收入效应和替代效应，总体上更倾向于促使劳动力回流。任远和施闻 (2017) 以外出劳动力是否发生回流为因变量，构建 Logistic 模型分析回流迁移的影响因素。研究表明，城市就业排斥、经济收入不足以及社会保障缺失是推动劳动力回流的重要原因，同时家庭生活需求、农业活动和农地状况也对回流决策产生显著影响。研究进一步指出，回流迁移是“被动回流”与“主动回流”的结合，体现了个体决策与家庭决策的综合过程。回流不仅带来人力资本补偿，还促进了流出地非农经济发展和创业增长，成为城镇化过程中不可或缺的逆向迁移流。

如同先前指出，静态框架下理论模型无法涵盖迁移的暂时性和可撤回行，难以捕捉

迁移的动态特征和复杂性。相比之下,国外文献中广泛应用的动态离散选择模型能够更好地刻画迁移的暂时性和可撤回性。因此本文鉴国际经验,结合动态离散选择模型方法与回流问题探讨我国劳动力回流的驱动机制及其经济社会影响。

2.5 影响劳动力迁移的因素

为构建有效的动态离散选择 (DDC) 模型分析劳动力流动,需系统整合迁移决策理论的相关研究成果。Lee (1966)提出的“推-拉”理论构建了二元分析框架:推力因素促使个体离开原居地,拉力因素驱动个体流向特定目的地。DDC 模型的优势在于能够通过随机效用函数的结构化设定,将多维决策因素纳入迁移行为的分析框架。

房价作为影响劳动力流动的重要因素受到广泛关注。高波等 (2012)发现相对房价提高促使劳动力流出,并对低附加值产业产生挤出效应。王丽莉 (2023)指出,东部大城市的严格土地规制导致高房价阻碍劳动力向高生产率地区转移,加剧空间错配。周颖刚等 (2019)发现高房价增强了劳动力家庭的流动意愿,特别是对未购房的高技能劳动力具有显著挤出效应。Zhou et al. (2022)研究表明房价每上涨 1%,流动人口的平均受教育年限增加 0.297 年,显示高房价对低技能劳动力形成挤出效应。张莉等 (2017)论证了房价对劳动力流动的双重作用:一方面,高房价降低了未来收入的不确定性,吸引人才;另一方面,上涨的房价压缩了可支配收入,抑制劳动力流入,形成倒 U 型影响。

户籍制度是我国劳动力流动的重要制度性障碍。Ngai et al. (2019)指出户籍制度通过绑定公共福利限制了人口流动,同时土地政策导致农业就业过剩。李强 (2003)分析发现户籍制度使中国人口流动不再遵循传统的推拉规律,形成了独特的迁移模式。Tombe et al. (2019)量化了户籍政策带来的高迁移成本,发现 2000 年至 2005 年间,国内贸易和人口迁移成本的下降贡献了总劳动生产率增长的 36%。周文等 (2017)研究表明土地控制和户籍限制共同抑制了劳动力流动,改革这些制度有助于提高城市化水平并缩小城乡收入差距。安虎森等 (2011)探讨了城市高房价与户籍制度对城乡收入差距的“门槛效应”,发现市场开放度低时高房价扩大收入差距,户籍制度起抑制作用;但市场开放度高时,户籍制度反而加剧收入差距。

舒适度 (amenity) 包括空气质量、公共服务水平、城市天气等因素共同影响劳动力流动。杨曦 (2017)发现城镇化显著提升了实际 GDP 和城乡劳动力实际工资,但效应因城市规模而异。王丽娜 (2007)指出收入、生活成本和城市宜居性特征共同决定了劳动力的空间流动模式。夏怡然等 (2015)研究表明公共服务显著影响劳动力流向,长期流动人口更倾向于选择公共服务水平较高的城市。孙伟增等 (2019)分析发现 PM2.5 浓度每上

升 $1 \mu g m^{-3}$ ，流动人口到该城市就业的概率下降 0.39 个百分点，且年龄较大、受教育水平较高的人群对空气污染更敏感。

语言在迁移决策中具有重要影响。Adserà et al. (2015)发现语言接近性使移民率提高约 20%。Bauer et al. (2005)指出语言聚居区能通过减少沟通障碍鼓励移民。Isphording et al. (2014)研究发现较大的语言距离导致移民在经济社会融入方面面临显著困难。刘毓芸等 (2015)探讨了方言距离对劳动力流动的“倒 U 型”影响：当方言距离较小时促进流动，过大时阻碍流动。

年龄影响移民决策，年轻人通常更频繁迁移，这符合人力资本理论：移民是一种投资，对年轻工人的收益期更长。年轻人获得更多收益，老年人面临更高成本和更少收益，但某些老年人出于特定原因（如回国或与家人团聚）仍会迁移。**教育**作为人力资本核心要素，在人口迁移中扮演重要角色。Aydemir et al. (2022)发现完成中学教育可使青年男性移民概率增加约 50%。Wozniak (2010)表明受过高等教育的劳动者更倾向于迁往劳动力需求高的地区。Weber et al. (2023)揭示了经济发展与高等教育机会之间的正向关系，推动了国际学生流动。高等教育通常涉及正向选择机制，流动人口往往拥有更高的教育背景或技能水平。**迁移成本**影响个人决策和研究框架，包括财务成本、社会成本和机会成本。刘晨晖等 (2022)在效用函数中加入与迁移距离相关的迁移成本。以上讨论的劳动力迁移影响因素归纳如表2-3所示。

表 2-3 影响劳动力迁移因素的主要文献

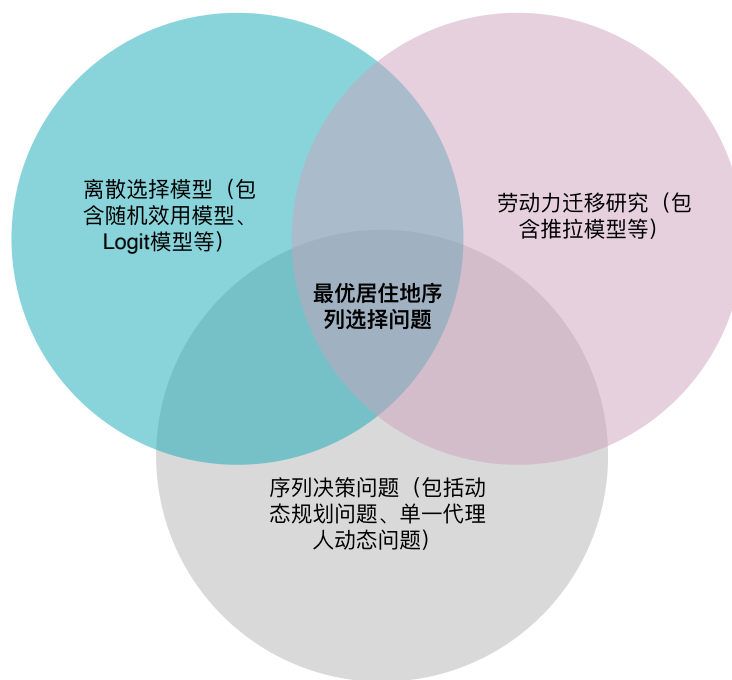
| 影响因素 | 主要讨论文献 |
|---------|--|
| 收入差异 | Kennan et al. (2011): 收入差距是造成劳动力流动的最显著因素 |
| 房价 | 张莉 等 (2017): 高房价导致居民永久性迁移比例下降 |
| 户籍 | Ngai et al. (2019): 户籍阻碍劳动力自由流动 |
| 公共服务 | 夏怡然 等 (2015)等文献均有涉及 |
| 气候等自然条件 | 洪大用 等 (2016): 空气质量对居民迁出意愿的影响 |
| 文化壁垒 | 刘毓芸 等 (2015): 方言的 U 型影响 |
| 年龄 | 多种文献均有涉及 |
| 迁移成本 | Todaro (1969): 探讨迁移成本的重要性 |

基于以上讨论，**迁移摩擦** (Migration Friction) 的概念逐渐凸显，指阻碍个体在不同地理区域间自由流动的各种制度性、经济性、社会性或文化性因素。蒋为 等 (2024)将跨

地区劳动力流动壁垒定义为阻碍个体自由流动的各类障碍总和,包括自然壁垒和制度壁垒。由于迁移摩擦具有不可观测性和微观不可加总的特征,其测度成为研究难点。王丽莉和乔雪 (2020)认为迁移摩擦由迁移成本、城镇失业工资以及工作搜寻匹配摩擦组成,这些因素共同塑造了城乡迁移的结构特征。刘修岩 等 (2017)强调迁移摩擦是中国城市体系扁平化的重要原因,消除迁移摩擦可带来人口重新配置和福利增进。

至此,我们梳理了基于 DDC 理论的动态最优居住地序列选择问题 (Dynamic Optimal Residential Sequence Decision Problem) 的所有相关文献。最优居住地序列选择模型的理论来源如图2-1所示。本文的下一章将构建一个适用于我国社会的理论模型。

图 2-1 最优居住地序列选择问题的理论来源



3 理论模型

劳动力流动问题本质上极为复杂。经过前述分析，我们认识到，劳动力流动和迁移不应被简单地视为一次性经济活动，而应被理解为贯穿个体整个生命周期的持续决策过程。假设决策者²是理性的，能够在多个地理位置之间进行自主选择。每个地区都对应一个独特的、排他性的效用流 (utility flow)。决策者在进行迁移时，需要承担相应的迁移成本。当决策者做出迁移决策后将承担相应的迁移成本。这意味着劳动力流动的本质是一个投资决策。决策者放弃当前的确定性（如本地工作），投资于一个充满不确定性但可能回报更高的未来（去大城市）。成本不仅是路费和短期收入损失，更是社会网络的断裂、制度性歧视以及巨大的心理压力。收益也不仅是更高的工资，更是子女的未来、更广阔的职业平台和个人价值的实现。决策者在不同阶段根据自身状况和外部环境变化，不断权衡是否迁移、迁移到何处，以及迁移后如何实现自身利益最大化。

本文通过 DDCM 框架构建了一个迁移决策模型，来刻画劳动力在不同阶段的定居与迁移选择，其必要性主要体现在以下三点。首先，静态模型通常假设个体基于当期特征作出一次性选择，忽视了人力资本积累、迁移经验、信息更新等动态因素对后续迁移意愿的影响。然而，在现实中，迁移意愿会随着年龄增长、技能提升、家庭结构变化而不断调整，这种时间路径依赖是静态框架无法捕捉的。其次，DDCM 可以更有效地处理不可观测的个体异质性。现实中的决策者在风险偏好、流动成本感知等方面存在差异，这些特质既难以直接观测，又会与迁移行为相互作用，形成复杂的内生机制。静态方法往往将这些差异视为随机误差处理，从而掩盖了其结构性影响；而 DDCM 可以在动态结构中引入状态变量与潜在类型 (latent types)，从而更系统地识别和解释这类异质性。再次，与空间均衡方法相比，DDCM 不依赖于市场已经达到长期均衡的假设，而是能够刻画非均衡过渡过程与前瞻性预期。空间均衡模型擅长解释长期劳动力配置与区域间工资、房价的均衡关系，但无法描述短期内政策冲击、经济波动或信息变化对个体迁移决策路径的影响。DDCM 则通过显式引入跨期预期效用函数，允许个体在决策中权衡即时收益与未来回报，从而能够在非均衡环境下分析政策效应与行为调整过程。为此，本章将以更贴近现实的建模方式分析迁移行为及其背后的逻辑。

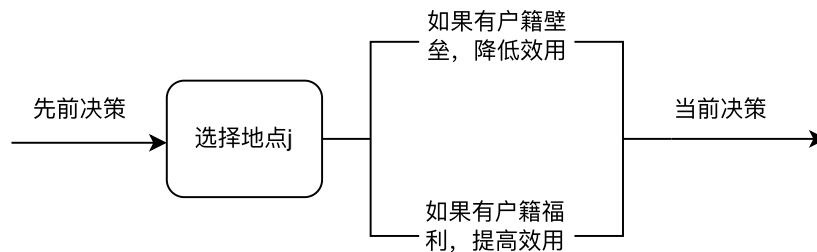
²文章中出现的个体指可以独立做出决策的经济单位，其形式可能为个人，也可能为家庭。为了消除歧义，后续将使用“决策者”一词替代。

3.1 模型的基本设定

本文考虑一个人口总量恒定的封闭经济体，人口流动仅发生在国内各地区之间。经济体由 m 个离散地区组成，形成迁移网络 $\mathcal{C} = \{1, 2, \dots, m\}$ ，这也是个体决策者的行动空间。各区域拥有固定的地理边界和空间关系，每个地区连通且可达，迁移连通性由 $m \times m$ 列的邻接矩阵 $Adj = [j_{xy} \in \{0, 1\}]$ 刻画。不同的区域都具有独特的地区特征，这包括经济特征与非经济特征等，每个地区的特征随时间动态变化，会影响个体的迁移决策。各地区的工资由本地劳动力市场决定。时间被划分为离散的时期 $t = \{0, 1, 2, \dots, t, t+1, \dots\}$ 。理性决策者 i 获取效用，在每个时期以效用最大化为目标做出迁移决策。地区特征（如工资、就业机会）、经济状况和政策随时间变化，影响个体的预期和决策。

本模型的一个核心特征是引入了中国的户籍制度。我们假设个体 i 的户口所在地 h_i 在其出生时确定且终身不变。当个体选择的居住地 j_t 与其户籍地 h_i 不一致时（即人户分离），其在获取公共服务、社会福利等方面会面临障碍，从而产生负效用。反之，居住在户籍地则无此惩罚。户籍制度对于决策者的影响效果如图3-1所示。³

图 3-1 户籍制度造成影响的途径



为了处理模型中的跨期决策问题，同时保持计算上的可行性，本文遵循了动态离散选择文献中的一个标准做法，假设存在一个完美的资本市场，个体可以在固定利率下自由借贷。此外，我们假设个体的效用函数对于由收入带来的消费是线性的，即收入的边际效用 α_w 是一个常数。在此假设下，家庭的资产持有对一次性迁移决策的边际影响可被忽略，因此模型中未将资产作为显式状态变量。

³ 一个自然的问题是：在存在诸多摩擦的经济体中，人口流动是否会最终停止？在经典的劳动力自由流动模型中，地区间工资差异会因套利行为而趋于收敛。然而，在本模型中，地区间的非经济舒适度差异、以及个体对不同地区的异质性偏好，是驱动持续流动的根本动力。更重要的是，我们引入的户籍制度、家乡溢价和多维度的迁移成本，恰恰是对劳动力自由流动的制约，这些摩擦的存在解释了为何现实中地区间的巨大差异能够长期持续存在，从而使模型的动态迁移过程更贴近现实。

一个理性的决策者代理人的决策遵循一个马尔可夫决策过程 (Markov Decision Process, MDP)。决策者在 T 期离散生命周期内选择最优居住地序列。令 $j_t \in \mathcal{C}$ 表示第 t 期的居住地选择，决策者的那么完整的选择序列可表示为 $\mathcal{J} = \{j_1, j_2, \dots, j_T\}$ 。典型的居住地选择模式如表格3-1所示。

| 个体流动形式 | 居住地选择序列 |
|--|---|
| 长期定居：决策者在特定时间段内总是停留在同一个地区 | $\mathcal{J} = \{\dots, 1, 1, 1, 1, 1, \dots\}$ |
| 单向迁移：如果 $j_p \neq j_q, \forall p \neq q$ ，那么个体就进行了迁移 | $\mathcal{J} = \{\dots, 1, 2, 3, 4, 5, \dots\}$ |
| 回流迁移：返回到先前居住过的地区 | $\mathcal{J} = \{\dots, 1, 2, 2, 1, 1, \dots\}$ |

表 3-1 居住地选择序列可能的形式

在每一时期 t 的决策时刻，一个理性代理人 i 已知其上一期所在地点 $l_{i,t-1}$ 、户籍 h_i 等信息，并据此选择本期的居住地 $j_t \in \mathcal{C}$ 。我们将个体在时期 t 决策前的全部相关信息定义为状态向量 x_{it} ：

$$x_{it} = (a_{it}, l_{i,t-1}, h_i, H_{it}, \{\nu_{ik}\}_{k \in H_{it}}, \eta_i, K_t), \quad (3.1)$$

其中 $a_{it} \in \{A_{min}, \dots, A_{max}\}$ 为代理人的年龄； $H_{it} = \{l_{i0}, \dots, l_{i,t-1}\}$ 为个体已到访的地点集合（历史记录）， $\{\nu_{ik}\}_{k \in H_{it}}$ 为个体已知的个体-地点匹配值， η_i 为个体固定效应， K_t 为宏观状态。在时期 t ，个体在观测到自身状态 x_{it} 以及当期的随机效用冲击 $\{\zeta_{itj}\}_{j \in \mathcal{C}}$ 后，做出居住地选择 j_t 。

在时期 t ，个体 i 若选择居住在潜在的地点 j 处，⁴ 其获得的总效用以 McFadden 效用函数形式给出：

$$\tilde{u}_{it}(j, x_{it}, \theta) = u_{it}(j, x_{it}, \theta) + \zeta_{itj} \quad (3.2)$$

其中 $u_{it}(j, x_{it}, \theta)$ 为可观测的效用流， θ 是需要估计的结构性参数； ζ_{itj} 为不可观测的效用冲击，是在任意时期和地点都独立同分布的随机扰动项，服从 Type I 极值分布（Gumbel 分布），且与状态变量 x_t 无关。

⁴ j 是一个泛指的、潜在的选项，是一个哑变量或者索引。而 j_t 一个已实现的、实际的选择，是一个随机变量，其值是在时期 t 做出的具体决策结果。此处切换为 j 是为了泛化模型。

居民的基础效用由货币性收入、城市宜居度、家乡溢价减去迁移成本组成。为了捕捉驱动迁移决策的关键因素，可观测效用流 u_t 函数的具体形式设定如下：

$$u_{it}(j, x_{it}; \theta) = \underbrace{U^{\text{income}}(w_{itj}|w_{it}^{\text{ref}})}_{\text{对经济收入的预期管理}} + \underbrace{\sum_s \alpha_s \cdot y_{sj}(K_t)}_{\text{地区舒适度}} + \underbrace{\alpha_{\text{home}} \cdot I(j = \text{home}_i)}_{\text{家乡溢价}} - \underbrace{\text{HukouPenalty}(j, \text{hukou}_i, K_t)}_{\text{户籍制度惩罚}} - \underbrace{\kappa_{it}(j, x_{it})}_{\text{迁移成本}} + \underbrace{\xi_{ij}}_{\text{偏好匹配}} \quad (3.3)$$

该函数由六个部分构成：

1. 对于经济收入的期望管理： $U^{\text{income}}(w_{itj}|w_{it}^{\text{ref}})$ ，当个体在地区 j 获得的工资 w_{itj} 与预期收入锚点进行比较，得到的一种满足感或者落差感。
2. 地区舒适度 (Amenities)：由气候、公共服务等一系列非经济因素 y_{sj} 决定的地区吸引力， α_s 是其边际效用。
3. 家乡溢价 (Home Premium)：当个体居住在户籍地 h_i 时，获得的额外心理效用 α_{home} 。这是一个只在家乡生效的拉力。
4. 户籍制度惩罚 (Hukou Penalty)：当个体居住在非户籍地时，由于制度壁垒产生的负效用。这是一个在所有非家乡地区生效的“推力”。
5. 迁移成本 (Migration Cost)：如果个体发生迁移（即 $j \neq l_{i,t-1}$ ），需要支付的一次性成本 κ_{it} 。
6. 非经济偏好匹配：个体 i 对地区 j 的一种持久的、非货币化的情感或文化契合度 ξ_{ij} 。与收入匹配项 ν_{ij} 类似，其真实值只有在个体访问过地区 j 后才能知晓。它捕捉了除收入外，个体与地区之间的特殊化学反应。

下文将对效用中货币收益项 w_{itj} 进行结构化建模，以便将个体工资形成机制与迁移选择共同识别。

3.2 收入效用函数

本模型引入了基于前景理论的损失厌恶机制，以提供了一个全新的、有别于传统迁移成本的机制来解释人们的迁移预期。假设个体的效用并非直接来源于其收入的绝对水平，而是来源于其实际收入与其参照点之间的比较，们使用对数的差值 $\ln(w_{itj}) - \ln(w_{it}^{\text{ref}})$

来近似工资的百分比变化。该设定不仅在数学上更为稳健，也更贴合经济学直觉，即决策者更关注收入的相对变化而非绝对差额。我们进一步引入损失厌恶系数 λ ，允许工资的百分比损失比等量的百分比收益产生更大的负效用……因此，我们将效用函数中的收入项从线性的 $\alpha_w \cdot w_{itj}$ 修正为 $U^{\text{income}}(\ln w_{itj} | \ln w_{it}^{\text{ref}})$:

$$U^{\text{income}}(\ln w_{itj} | \ln w_{it}^{\text{ref}}) = \begin{cases} \alpha_w \cdot (\ln w_{itj} - \ln w_{it}^{\text{ref}}) & , \ln w_{itj} \geq \ln w_{it}^{\text{ref}} \\ \alpha_w \cdot \lambda \cdot (\ln w_{itj} - \ln w_{it}^{\text{ref}}) & , \ln w_{itj} < \ln w_{it}^{\text{ref}} \end{cases} \quad (3.4)$$

这一设定意味着效用来自于实现的增益和损失。正的收入差距 $(\ln w_{itj} - \ln w_{it}^{\text{ref}})$ 代表了超出预期的成就感，而负的差距则代表了未达预期的失落感。损失厌恶系数 λ 则量化了人们“宁可少赚，不可赔”的心理，即损失带来的负效用要大于等量收益带来的正效用。 α_w 依然是收入的（收益域）边际效用，预期为正。

决策者的收入函数由多种因素组成。个体决策者 $i \in \mathcal{N}$ 在时期 t 、地点 j 的收入设定为以下函数形式：

$$w_{itj} = \mu_j(K_t) + G(X_i, a_{it}, t) + \eta_i + \nu_{ij} + \varepsilon_{itj} \quad (3.5)$$

1. **地区基础工资** $\mu_j(K_t)$ 为地点 j 的基准薪资水平，代表地区的经济基本面，如生活成本、区域产业聚集效应（如北京的传媒业、山西的采矿业）、政策红利（税收减免、人才补贴）等结构性特征，是所有在该地点个体的工资基准，其与个体无关，仅随宏观状态 K_t 变化。这是捕捉地区间核心经济差异的关键变量，是驱动劳动力进行跨区域套利、寻求更高经济回报的根本动力。在实证中，我们将通过一个独立的工资分解回归来预先估计每个地区的 μ_{jt} ，并将其作为可观测的地区特征代入主模型。
2. **个体可观测部分** $G(X_i, a_{it}, t)$ 捕捉由个体年龄（人力资本积累）决定的收入部分，是人力资本积累的代表。在传统的设定中， $G(\cdot)$ 通常被假设为一个关于年龄的二次多项式和线性控制变量的组合。然而，为了避免因函数形式误设而导致的偏误，我们在实证估计中将采用更灵活的方法。具体而言，我们将使用机器学习模型来非参数地估计 $G(\cdot)$ (详见 4.2 节)
3. **个体固定效应** η_i 代表个体不随地点改变的固有能力或特质，如先天能力、家庭背

景等。在我们的设定中，个体的固有能力 η_i （即不随地点改变的个体固定效应）被内含在了 $G(x_{it})$ 和 ε_{itj} 的组合中。在工资分解步骤中，个体固定效应将被吸收，而在 ML 插件的拟合中，其影响也将被非参数地捕捉。

4. **个体-地区匹配效应** ν_{ij} 捕捉了特定个体 i 在特定地区 j 的生产力特殊加成（如技能需求匹配、文化契合度），代表相同技能劳动者在不同地区的收入差异。该值在个体首次进入地区 j 时实现（从一个分布 $F_\nu(0, \sigma_{\nu,jt}^2)$ 中抽取），此后在该地区停留期间保持不变，但迁移到新地区 k 时会生成新的 ν_{ik} 。假设个体-地区匹配效应分布 $F_\nu(0, \sigma_{\nu,jt}^2)$ 的方差是随信息渠道变化的：

$$\nu_{ij} \sim F_\nu(0, \sigma_{\nu,jt}^2) \quad (3.6)$$

方差 $\sigma_{\nu,jt}^2$ 的函数形式为：

$$\sigma_{\nu,jt}^2 = \exp(\delta_0 - \delta_1 \cdot \text{Internet}_{j,t}) \quad (3.7)$$

其中 δ_0 和 δ_1 是待估参数。 δ_1 是核心参数，我们预期 $\delta_1 > 0$ 。这意味着，当一个地区的互联网普及率 $\text{Internet}_{j,t}$ 越高时，你去那里之前面临的收入不确定性 $\sigma_{\nu,jt}^2$ 就越低。这对模型意味着当个体考虑迁移到一个新的、互联网不发达的地区 j 时，他面临的 ν_{ij} 方差很大。这意味着工资可能是惊喜（抽到很高的 ν ），也可能是惊吓（抽到很低的 ν ）。风险厌恶的个体会因此不愿意去。当考虑迁移到一个新的、互联网发达的地区 k 时， ν_{ik} 的方差较小。个体对去那里能拿到的工资更有把握，决策的不确定性降低，从而更愿意迁移。

5. **暂时性随机效应** ε_{itj} 代表短期波动或随机扰动项，服从零均值正态分布的当期随机工资冲击，代表了不可预测的短期波动（如临时绩效波动、经济冲击）。因此若在数据中观测到个体 i 在时期 t 的写为 j_t ，其实际工资为 w_{itj}^{obs} ，可知其条件密度为

$$\mathcal{L}_{i,t}^{\text{wage}} = \frac{1}{\sigma_\varepsilon} \phi \left(\frac{w_{it,j_t}^{\text{obs}} - \mu_{j_t}(K_t) - G(X_i, a_{it}) - \eta_i - \nu_{ij_t}}{\sigma_\varepsilon} \right) \quad (3.8)$$

其中 $\phi(\cdot)$ 为标准正态密度，该方程为处于地点 j 的个体 i 在时期 t 的收入函数的密度函数 Ψ_{itj} 。这是个体 i 在时期 t 于地区 j 的工资观测值对总似然函数的贡献。

3.3 户籍惩罚、家乡溢价与地区舒适度

真实世界中户籍制度带来的效应是多维度的、有异质性的，为了细致刻画这种复杂影响，本文因此假设了一个分层的、与地区特征互动的户籍效应模型：

$$\begin{aligned} \text{HukouPenalty}(j, h_i, K_t) = & I(j \neq hukou_i) \cdot \left[\sum_{c \in C} \rho_{\text{base},c} \cdot \text{Tier}_{c(j)} + \rho_{\text{edu}} \cdot \text{Edu}(j) \right. \\ & \left. + \rho_{\text{health}} \cdot \text{Health}(j) + \rho_{\text{house}} \cdot \text{HousingPrice}(j) \right] \end{aligned} \quad (3.9)$$

其中：

1. 指示函数 $I(j \neq hukou_i)$ ：表示只有在人户分离时，惩罚才生效。
2. 基础惩罚， $\sum_{c \in C} \rho_{\text{base},c} \cdot \text{Tier}_{c(j)}$ ：不同等级的城市其户籍政策的宽松程度和隐性价值存在巨大差异，为此我们根据公认的城市经济地位，将全国所有地区划分为 C 个类别（例如：一线城市、新一线城市、其他城市）。其中 $\text{Tier}_{c(j)}$ 是一个虚拟变量，当地区 j 属于类别 c 时取值为 1，否则为 0。 $\rho_{\text{base},c}$ 则是对应类别 c 的待估参数。例如，我们预期 $\hat{\theta}_{\text{base}, \text{一线}}$ 将会显著为正且远大于其他类别的参数，这量化了进入顶级城市所需克服的最高制度性壁垒。
3. 与公共服务互动的惩罚， $\rho_{\text{edu}} \cdot \text{Edu}(j)$ ：户籍惩罚的大小与该地区的教育质量（如重点高中比例）成正比。教育资源越好的城市，没有本地户口的孩子上学越困难，因此惩罚越大。预期 $\rho_{\text{edu}} > 0$ 。 $\rho_{\text{health}} \cdot \text{Health}(j)$ 与该地区医疗水平（如三甲医院数量）成正比。医疗越好的地方，医保报销的地区差异带来的损失越大。预期 $\rho_{\text{health}} > 0$ 。
4. 与生活成本互动的惩罚， $\rho_{\text{house}} \cdot \text{HousingPrice}(j)$ ：与该地区的房价等生活成本指标互动。很多城市的购房资格与户口或社保年限挂钩，房价越高，这种限制带来的机会成本就越大。预期 $\rho_{\text{house}} > 0$ 。

户籍效应的机制扎根于制度性、政策性的进入壁垒。它源于公共服务（教育、医疗）的排他性、购房购车资格的限制等。这是一个推离异乡（Push Factor）的力。它是一种负向效用（成本或惩罚），在个体身处任何非户口所在地时都会承受。该变量对所有 $m-1$ 个非户口所在地都生效，且惩罚大小因地而异。户籍效应可以通过观察个体在不同城市间的选择来识别。例如，如果两个城市 A 和 B 工资和舒适度相似，但城市 A 的户籍管

制更严（公共服务更好），那么需要一个更高的工资差异才能吸引无户口者去 A 城。这个补偿性工资差异就反映了户籍惩罚的大小。

本文同时还在模型中假设了家乡溢价效应， $\alpha_{home} \cdot I(j = home_i)$ ，这个变量不同于户籍效应，是一个拉回家乡（Pull Factor）的力，是一种正向效用，只在个体身处家乡时才会获得。该变量刻画的是心理、社会、文化层面的情感依恋，是对家乡非理性偏好的量化。它源于家庭纽带、朋友圈、熟悉的文化环境等。只对家乡 h_i 生效。家乡溢价效应变量通过观察个体即使在面临更好的外地经济机会时，依然选择留在或返回家乡的行为来识别。

模型中的地区舒适度和户籍惩罚可能包含相同的地区特征变量，如教育和医疗资源。这并非重复计算，而是为了识别这些资源在效用层面上的两种不同渠道。

以教育变量为例，在地区舒适度中舒适度项 $\alpha_s^{Edu} \cdot Edu(j)$ 代表教育资源作为一种公共品带来的普遍效用。一个地区教育好，所有人（无论有无户口）都会觉得这个地方有文化氛围、重视下一代，从而增加其作为居住地的吸引力。这是一种对所有决策者生效的环境效应。而在户籍惩罚中， $\theta_{Edu} \cdot Edu(j)$ 代表教育资源作为一种排他性服务带来的效用损失。一个地区教育越好，没有户口的人在享受这种服务时面临的壁垒就越高（比如子女不能参加本地高考），因此他们感受到的惩罚就越大。这是一种只针对人户分离决策者的准入成本。

模型正是通过比较户籍人口与非户籍人口对这些资源的不同反应，来分别识别参数 α_s^{Edu} 和 θ_{Edu} 的。

以上效应结合在同一个模型中，是解释中国劳动力流动谜题（如大量回流、城乡差距长期存在）的关键且独立的机制。

3.4 迁移成本

决策者 i 的迁移成本 κ 设定为以下函数形式：

$$\kappa_{it}(j, x_{it}) = I(j \neq l_{i,t-1}) \cdot \left[\gamma_{0\tau} + \gamma_1 Dis(l_{i,t-1}, j) - \gamma_2 I(j \in Adj(l_{i,t-1})) - \gamma_3 I(j \in H_{it}) + \gamma_4 a_{it} - \gamma_5 n_{jt}(K_t) \right] \quad (3.10)$$

其中，指示函数 $I(j \neq l_{i,t-1})$ 确保了原地停留者成本为零。成本构成包括：

1. $\gamma_{0\tau}$: 与个体类型 τ 相关的固定成本，随个体类型哑变量 τ 变化，用以捕捉“恋家型”与“闯荡型”等未观测到的异质性。例如定居型个体的 $\gamma_{0\tau}$ 极高，几乎禁止迁

移；流动型个体的 $\gamma_{0\tau}$ 较低，更易迁移。不同类型的群体分布由对应的概率 π_τ 给出，并且满足 $\sum_{\tau} \pi_\tau = 1$ 。

2. $\gamma_1 Dis(l_{i,t-1}, j)$: 与物理距离相关的线性成本（以大圆距离衡量）。
3. $-\gamma_2 I(j \in Adj(l_{i,t-1}))$: 迁移到相邻地区的成本折扣，相邻地区迁移可能因文化相似、信息透明或交通便利而成本更低。
4. $-\gamma_3 I(j \in H_{it})$: 回流迁移的成本折扣，因为返回曾居住过的地区，社会网络和适应性成本较低。
5. $\gamma_4 a_{it}$: 年龄越大，迁移的心理或生理成本越高，这是随年龄增长而增加的迁移阻力。
6. $-\gamma_5 n_{jt}(K_t)$: 迁入地人口规模带来的成本减免效应，反映了更大城市网络带来的便利。目标地区人口规模 n_j 越大，迁移成本减少 $\gamma_5 n_j$ 。这是因为大规模地区可能有更多潜在支持（如亲友网络、服务机构），减少安家难度，从而降低迁移成本（ $\gamma_5 > 0$ ），符合引力模型中人口规模吸引迁移的经典结论。

迁移成本中参数的经济含义如表格3-2所展示。

表 3-2 迁移成本参数释义

| 参数符号 | 参数含义 | 参数意义 |
|------------------|--------|---|
| $\gamma_{0\tau}$ | 类型异质效应 | 捕捉不同类型的异质性，可以解释为何有些人从不迁移，识别“定居型”人群可针对性地设计激励措施（如搬迁补贴）。 |
| γ_1 | 物理迁移成本 | 个体与其携带物必然因为物理法则而产生运输成本。 |
| γ_2 | 相邻地区效应 | 反映地理邻近性的普遍优势。 |
| γ_3 | 历史地区效应 | 返回历史地区因熟悉度获得折扣，体现路径依赖。 |
| γ_4 | 年龄障碍效应 | 针对不同年龄群体制定差异化迁移政策（如青年人才引进计划）。 |
| γ_5 | 规模成本效应 | 大地区吸引力不仅体现在经济机会（如工资 μ_j ），也直接降低迁移成本。 |

迁移成本不仅影响当前选择，也会改变 H_{it} 与未来可获得的匹配信息，从而通过状态转移影响未来效用。

3.5 状态转移概率

状态转移概率 $p(x_{i,t+1}|x_{it}, j_t)$ 描述了系统如何从当前状态演化至下一期状态。值得强调的是，本模型通过将历史居住地集合 H_{it} 和已知的匹配值 $\{\nu_{ik}\}_{k \in H_{it}}$ 纳入状态向量 x_{it} ，从而内生化了路径依赖。这意味着，尽管状态转移本身遵循马尔可夫属性（即 $p(x_{t+1}|x_t, j_t, x_{t-1}, \dots)$ 简化为 $p(x_{t+1}|x_t, j_t)$ ），但由于过去已经被编码在现在 x_{it} 之中，个体的决策实质上是历史依赖的。这种设定既保持了马尔可夫决策过程的分析便利性，又能够捕捉学习效应和回流迁移等复杂的动态行为，是本模型的一个关键特征。

具体而言，由于状态变量是多维的，本文将其设置为：

- 年龄: $a_{i,t+1} = a_{it} + 1$ (确定性转移)
- 上一期位置: $l_{i,t} = j_t$ (确定性转移)
- 户口所在地: h_i (不变)
- 个体固定效应: η_i (不变)
- 历史居住地集合: $H_{i,t+1} = H_{it} \cup \{j_t\}$ (确定性转移)
- 宏观状态: $K_{t+1} \sim p(K_{t+1}|K_t)$ (随机转移，假设为一阶马尔可夫过程)
- 已知的匹配值:

$$p(\{\nu_{ik}\}_{k \in H_{i,t+1}} | \{\nu_{ik}\}_{k \in H_{it}}, j_t) = \begin{cases} \delta(\{\nu_{ik}\}_{k \in H_{i,t+1}} - \{\nu_{ik}\}_{k \in H_{it}}) & \text{if } j_t \in H_{it} \\ f_\nu(\nu_{ij_t}) & \text{if } j_t \notin H_{it} \end{cases}$$

这样的设定意味着，如果 j_t 是一个已经去过的地方 ($j_t \in H_{it}$)，那么个体不会学到新的信息，已知的匹配值集合不变。这里的 $\delta(\cdot)$ 是狄拉克函数，表示这是一个确定性的转移。如果 j_t 是一个新的地方 ($j_t \notin H_{it}$)，个体在选择 j_t 后，便会获知新的匹配值 ν_{ij_t} 。这个新值是从分布 F_ν 中抽取的，其概率密度为 $f_\nu(\cdot)$ 。而已知的其他匹配值保持不变。这个新信息将被纳入下一期的状态向量 $x_{i,t+1}$ 中，从而影响其未来的决策。

3.6 条件选择概率函数

个体的目标是选择一个居住地序列 $\mathcal{J}^* = \{j_1^*, \dots, j_T^*\}$ 以最大化其终身期望折现效用，即：

$$\max_{\mathcal{J}=\{j_1, j_2, \dots, j_T\}} \mathbb{E} \left[\sum_{t=1}^T \beta^{t-1} \tilde{u}_{it}(j_t, x_{it}) \right] \quad (3.11)$$

该问题可通过贝尔曼方程描述。令 $\beta \in (0, 1)$ 为折现因子。在第 t 期、状态为 x_t 时，选择潜在地区 j 的期望折现总效用（不含当期的随机冲击）可以用选择特定价值函数（Choice-Specific Value Function）表示：

$$v_{it}(x_{it}, j) = u_{it}(j, x_{it}; \theta) + \beta \cdot \mathbb{E} [\bar{V}_{i,t+1}(x_{i,t+1}) | x_{it}, j] \quad (3.12)$$

其中

$$\bar{V}_{it}(x_{it}) = \mathbb{E}_{\zeta} \left[\max_{k \in \mathcal{C}} \{v_{it}(x_{it}, k) + \zeta_{ik,t}\} \right] \quad (3.13)$$

为期望价值函数 (Expected Value Function)，即在状态 x_{t+1} 下，对所有可能随机扰动 ζ 取期望后的长期效用，是代理人知道 x_t 但在抽取当期随机冲击 ζ_t 之前的期望效用。

在每一期，当随机冲击向量 $\{\zeta_{ij,t}\}_{j \in \mathcal{C}}$ 实现后，理性代理人会选择使其总效用最大化的地区 j_t^* ：

$$j_{it}^* = \arg \max_{j \in \mathcal{C}} v_{it}(x_{it}, j) + \zeta_{ij,t} \quad (3.14)$$

代理人在选择集 \mathcal{C} 中选择行动 j 的概率为条件选择概率（Conditional Choice Probability, CCP）函数，可以表示为：

$$\begin{aligned} P(j_t | x_t, \theta) &= \text{Prob}(\tilde{u}_j > \tilde{u}_k, \forall k \neq j) \\ &= \text{Pr}(u_j + \zeta_j > u_k + \zeta_k, \forall k \neq j) \\ &= \text{Pr}(\zeta_k - \zeta_j < u_j - u_k, \forall k \neq j) \end{aligned} \quad (3.15)$$

根据 Rust (1987) 以及 Hotz et al. (1993) 的开创性工作，由于 ζ_j 独立同分布且服从 I 型极值分布（尺度为 1），CCP 可以写为 softmax 函数形式：

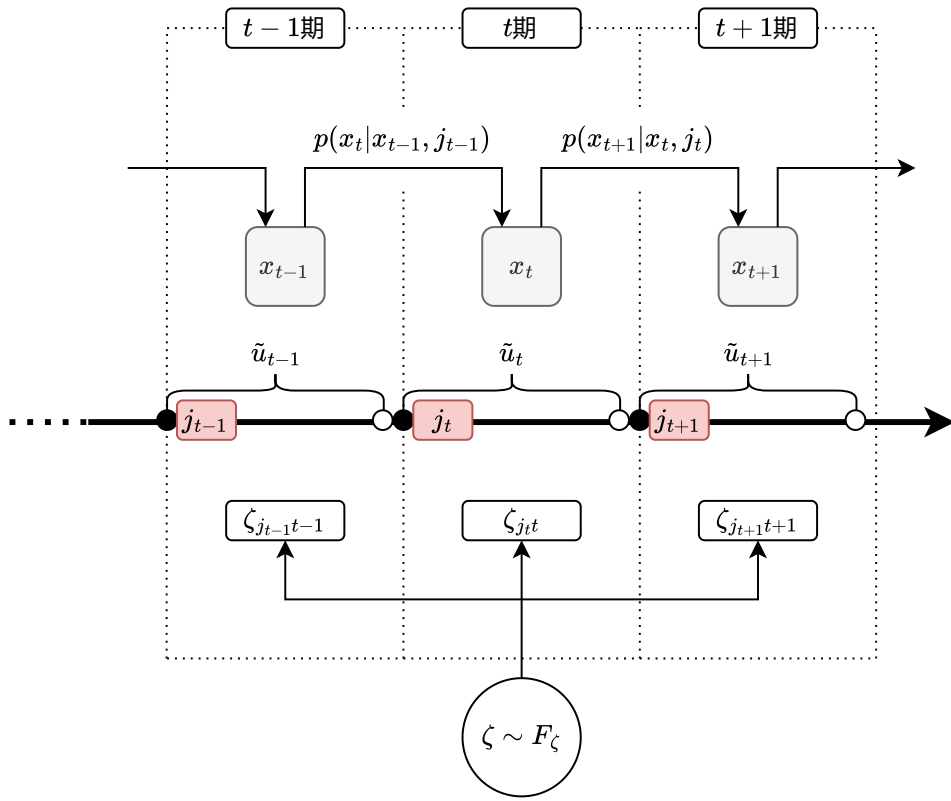
$$P(j_t = j | x_{it}; \theta) = \frac{\exp(v_{it}(x_{it}, j))}{\sum_{k \in \mathcal{C}} \exp(v_{it}(x_{it}, k))} \quad (3.16)$$

这极大地优化了动态规划问题中期望值的计算。并且由于 softmax 函数输出的概率分布

是平滑的，有利于策略迭代的稳定性。

图3-2 展示了一个动态规划决策过程。在离散周期 t 结束时或者下个周期 $t+1$ 即将开始时，决策者根据在抽取到的随机收益冲击 ζ 和状态转移概率 $p(x_{t+1}|x_t, j_t)$ 双重不确定性的影响下通过 Bellman 方程做出决策，进入下一个周期 $t+1$ 后获得该期的效用 \tilde{u}_{t+1} 。

图 3-2 动态离散选择模型下的劳动力迁移决策流程图



为保持可读性，文中所有符号均按表 3-3 给出含义。

3.7 劳动力迁移决策

综合以上设定，个体是否从地区 j 迁移到地区 k 的决策，本质上是一个对各项利弊的权衡。个体选择迁移到 k 的条件是 $v_{it}(x_{it}, k) + \zeta_{k,t} > v_{it}(x_{it}, j) + \zeta_{j,t}$ 。忽略未来期望

表 3-3 符号说明

| 符号 | 含义 |
|--|------------------------------|
| i, t, j | 个体、时间、地点索引 |
| a_{it} | 年龄 |
| $l_{i,t-1}$ | 时期 t 决策前的上一期地点 |
| h_i | 户籍地 |
| H_{it} | 已到访地点集合（历史） |
| w_{itj} | 工资（货币性收入） |
| $\mu_j(K_t)$ | 地点基准工资（随宏观态 K_t 变化） |
| ν_{ij} | 个体—地点匹配效应 |
| ε | 工资的暂态效应 |
| η_i | 个体固定效应 |
| $\kappa_{it}(j, \cdot)$ | 迁移成本 |
| $\text{HukouPenalty}(\cdot)$ | 户籍惩罚 |
| $\zeta_{ij,t}$ | 个体-地点-时间特定的随机效用冲击（Type I EV） |
| $\theta = (\alpha, \gamma, \rho, \delta, \lambda)$ | 待估结构参数向量 |
| β | 外生给定的折现因子 |

价值的差异，仅看当期效用，这一决策可以直观地理解为：

$$\begin{aligned}
& \underbrace{(\alpha_w(w_k - w_j) + (\text{Amenity}_k - \text{Amenity}_j) + \dots)}_{\text{地区 } k \text{ 相对于 } j \text{ 的吸引力增益}} \\
& > \underbrace{\kappa_{j \rightarrow k}}_{\text{迁移成本}} + \underbrace{(\text{HukouPenalty}_k - \text{HukouPenalty}_j)}_{\text{户籍惩罚变化}} + \underbrace{(\text{HomePremium}_k - \text{HomePremium}_j)}_{\text{家乡溢价变化}} \\
& + \underbrace{(\zeta_{j,t} - \zeta_{k,t})}_{\text{随机偏好冲击}}
\end{aligned} \tag{3.17}$$

目标地区带来的经济与非经济收益的提升，必须足以克服迁移的直接成本、可能的制度性惩罚增加、以及放弃家乡溢价的损失。

至此，模型完整地刻画了一个理性个体，在有限生命周期和多重约束下，如何动态地在迁移网络 \mathcal{C} 中选择其最优居住地序列 \mathcal{J}^* 。该框架允许个体进行多次、可逆的迁移，

从而能够灵活地解释现实世界中复杂的劳动力流动模式。这种对劳动力迁移的自由性的刻画，与现实世界中的行为是融洽的。

4 实证处理

将前述构建的动态迁移决策模型付诸实证检验，面临着重大的计算挑战。动态规划本身就面临着维度诅咒带来的计算资源稀缺问题，传统方法使用两步法来解决。但当模型中个体决策依赖于多维度的、持久性的未观测异质性（如个人能力 η_i 或潜在类别 τ ），在两步估计法中第一步估计的 CCPs 容易产生偏误，并且误差无法在第二步中完全纠正。这是因为从数据中估计的 CCP 是所有异质类型个体行为的加权平均，无法直接用于识别特定类型的结构参数。

为此本文将 CCPs 的计算内嵌在固定嵌套点（Nested Fixed Points, NFXP）的内循环中，并通过期望最大化（Expectation-Maximization, EM）算法机制动态更新未观测类型的后验概率，从而实现联合估计。这种方法能够避免由分步估计带来的信息丢失，并在有限样本下表现出更好的稳健性。

同时为了减轻这种全解法带来的计算资源压力，我们还设计了专门的多线程并行计算方法。

4.1 处理未观测到的异质性

在结构模型中，合理地刻画个体间无法被直接观测到的异质性，是模型识别和解释力的关键。如果不对此加以识别，模型可能会将个体持久性的偏好或能力差异错误地归因于状态依赖。为此，本研究采用了两种互补的策略来捕捉多维度的未观测异质性：离散因子近似 (Discrete Factor Approximation) 和有限混合模型 (Finite Mixture Model)。

4.1.1 离散化近似

传统做法通常会对关键随机变量或潜在变量作出明确的分布假设，例如，这类方法虽然具有一定的便利性，但也带来潜在的问题，若真实数据与假设分布存在系统性偏差，模型估计将产生偏误，且这种偏误可能无法通过增加样本量消除。同时，固定的分布假设限制了变量形态，在数据中存在厚尾、偏态、多峰等复杂特征时，传统方法难以捕捉其异质性。

本文不施加正态分布等强参数假设，而是采用灵活的非参数离散化方法进行近似未知分布的变量。我们将这些连续的未知分布用一组有限的支撑点（support points）及其对应的概率（mass points）来表示。对于任何不知道其分布的 F ，我们可以找到最佳近似分布 \hat{F} 来近似。对于连续分布的变量 F ，离散的 \hat{F} 是一组有限的支撑点组成，每个

支撑点都有对应的概率值, 即 $\{(x_i, p_i)\}, 1 \leq i \leq N$ 的组合, 其中 x_i 是离散点, p_i 是其概率, N 为给定的支持点数。对于每个 $q_r, (1 \leq r \leq n)$, 计算对应的效用值, 用这些有限的效用值来代替连续效用的积分。通过该处理后, 模型变为有限状态的动态规划问题, 可以直接构建似然函数并进行最大似然估计。支撑点位置和权重可以灵活调整, 以捕捉厚尾、多峰或非对称等特征。相比完全非参数估计, 离散化将连续分布压缩为有限维参数, 既降低计算负担、避免随机噪声, 又可通过支撑点直观解释经济含义。

在估计中, 我们假设每个支撑点的概率是均等的, 需要估计的是支撑点的位置 (即值的大小), 从而刻画分布的形状和范围。具体设定如表 4-1 所示。

表 4-1 未观测异质性属性的离散化设定

| 异质性来源 | 符号 | 设定 | 经济学含义与识别 |
|-------|--------------------------|-----------------|--|
| 个体能力 | η_i | N_η 个支撑点 | 捕捉不随时间地点改变的个体固有生产力 (如天赋、家庭背景)。通过个体在不同地点间持续的平均工资差异来识别。 |
| 收入匹配 | ν_{ij} | N_ν 个支撑点 | 捕捉特定个体 i 与特定地区 j 之间的特殊生产力协同效应。通过个体在某地异常高或低的工资, 以及试错式的迁移行为来识别。 |
| 偏好匹配 | ξ_{ij} | N_ξ 个支撑点 | 捕捉非经济的、持久的个体-地区情感契合度。识别于那些无法用经济回报解释的、持续的留存或回流选择。 |
| 工资波动性 | $\sigma_{\varepsilon,i}$ | N_σ 个支撑点 | 允许不同个体的暂时性工资冲击具有不同的方差, 捕捉个体面临的收入不确定性差异 (如行业风险)。通过个体工资的时间序列波动幅度来识别。 |

我们将所有通过离散化近似的未观测属性集合记为:

$$\omega_i = \{\eta_i, \{\nu_{ij}\}_{j \in \mathcal{C}}, \{\xi_{ij}\}_{j \in \mathcal{C}}, \sigma_{\varepsilon,i}\}$$

4.1.2 有限混合

除了上述个体属性的连续性异质性外, 本文还假设劳动力群体中存在离散的、具有异质行为模式的潜在类别, 避免了硬性分组 (如通过地区、行业、性别等显式变量进行分组) 的局限。

个体 i 属于类别 $\tau \in \{1, 2, \dots, K\}$ 的概率为 π_τ , 且 $\sum_{\tau=1}^K \pi_\tau = 1$ 。本文假设劳动力群体中存在三种潜在的决策类型, 分别表征了不同的迁移行为模式。

具体而言，稳定型（Type 0）个体具有较高的基础迁移成本（例如， $\gamma_0 = 5.0$ ），对迁移距离表现出高度敏感（例如， $\gamma_1 = -3.0$ ），同时拥有强烈的家乡情感依恋（例如， $\alpha_{\text{home}} = 2.0$ ），而损失厌恶系数相对较低（例如， $\lambda = 1.2$ ）。行为上，他们偏好稳定，迁移倾向很低，对距离变化反应强烈，且家乡溢价显著，损失厌恶程度适中。

机会型（Type 1）的特征为中等迁移成本（例如， $\gamma_0 = 1.0$ ），对距离的敏感度较低（例如， $\gamma_1 = -1.5$ ），家乡溢价处于中等水平（例如， $\alpha_{\text{home}} = 0.8$ ），损失厌恶系数中等（例如， $\lambda = 1.8$ ）。行为上，他们适度参与迁移，积极寻求外部机会，对距离的顾虑较少，家乡情结一般，损失厌恶程度居中。

冒险型（Type 2）个体的迁移成本极低（例如， $\gamma_0 = 0.1$ ），对距离几乎不敏感（例如， $\gamma_1 = -0.5$ ），家乡情感依恋很弱（例如， $\alpha_{\text{home}} = 0.1$ ），但损失厌恶系数较高（例如， $\lambda = 2.5$ ）。行为上，他们频繁迁移，积极追逐机会，对地理距离的约束忽略不计，家乡情结淡薄，但同时潜在损失表现出高度厌恶。

通过这种设定，模型能够内生地识别不同行为类型，并估计其规模比例 π_τ 及类型特定参数，从而更全面地捕捉劳动力迁移决策中的结构性异质性。

上述的两种方法在模型中的作用是互补的。非参数离散化侧重于对迁移决策中的随机冲击分布进行灵活刻画，增强模型在面对未知或复杂分布时的稳健性；有限混合方法则侧重于对劳动力个体层面的结构性异质性进行识别与建模，避免遗漏不可观测分组信息所带来的偏误。通过结合使用这两类方法，模型能够在不牺牲可计算性的前提下，更全面地反映劳动力迁移行为背后的复杂机制。

4.2 引入机器学习方法处理滋扰函数

本模型的似然函数依赖于对工资条件期望 $\mathbb{E}[w_{itj}|x_{it}]$ 和状态转移概率 $p(x_{i,t+1}|x_{it}, j_t)$ 的准确建模。这些函数在经济学上被称为滋扰函数 (Nuisance Functions)，因为它们对于模型是必需的，但其具体的参数并非我们最关心的研究对象。

传统做法通常对这些函数施加较强的参数化假设（如线性或二次型）。然而，函数形式的误设可能会传导至我们核心结构参数（如户籍惩罚、家乡溢价）的估计中，造成偏误。为了增强预测精度、减少函数形式误设，我们学习 Athey et al. (2019) 的做法，采用了近年来计量经济学与机器学习交叉领域的前沿方法。具体而言，我们使用 Ke et al. (2017) 提出的 LightGBM 梯度提升决策树作为插件，来非参数地估计这些滋扰函数。

我们使用梯度提升决策树作为插件，来非参数地估计这些滋扰函数。

我们将工资方程 w_{itj} 的确定性部分 $\mathbb{E}[w_{itj}|\dots]$ 视为一个复杂的函数，并通过训练一

个回归树模型 $\hat{g}(\cdot)$ 来拟合它：

$$\hat{\mathbb{E}}[w_{itj} | x_{it}, j] = \hat{g}(x_{it}, j) \quad (4.1)$$

该模型能够自动学习高维状态变量之间的非线性关系与交互效应。在结构模型的每一次似然计算中，我们都调用这个预先训练好的 $\hat{g}(\cdot)$ 来获得工资预测值。状态转移插件：类似地，对于状态向量中随机演化的部分（如宏观状态 K_t ），我们同样使用机器学习分类器/回归器来估计其转移概率 $\hat{p}(K_{t+1}|K_t)$ 。

我们通过 5 折交叉验证来稳健地评估两个模型的样本外预测性能。如表 4-2 所示，LightGBM 模型在所有关键指标上均优于传统的 OLS 模型。具体而言，ML 插件将预测的均方根误差从 2.241 降低至 2.225，并将 R^2 从 11.77% 提升至 13.07%。

表 4-2 工资滋扰函数预测性能对比 (5 折交叉验证)

| 模型 | RMSE (均值 \pm 标准差) | R^2 (均值 \pm 标准差) |
|-------------------------------|---------------------------------------|---------------------------------------|
| 基准模型 (OLS with quadratic age) | 2.2411 \pm 0.0590 | 0.1177 \pm 0.0079 |
| ML 插件 (LightGBM) | 2.2246 \pm 0.0613 | 0.1307 \pm 0.0090 |

注：RMSE 为均方根误差，值越小表示预测越精准。 R^2 为解释的方差比例，值越高表示拟合优度越好。结果表明，LightGBM 模型在样本外预测能力上显著优于传统参数化模型。

这一结果表明，LightGBM 模型能够更有效地从高维的个体状态变量中提取信息，捕捉到传统线性模型所忽略的复杂关系。通过为我们的主结构模型提供一个更为精确的工资预期基准，ML 插件有效地降低了因滋扰函数误设可能引发的估计偏误。

图 4-1 通过偏依赖图 (Partial Dependence Plots, PDP) 直观地对比了两个模型所学习到的年龄对工资的边际效应。

4.3 结合 EM 算法的 NFXP 估计策略

本文将未观测的异质性视为缺失数据，通过 EM 算法的迭代框架来处理，在 EM 算法的每一步最大化 (M 步) 中，将 NFXP 算法作为核心计算引擎，用于求解给定参数下的动态规划问题。这种方法能够在处理复杂异质性的同时，保证结构模型参数估计的一致性。整个估计过程在一个迭代循环中完成，每一次迭代包含 E 步和 M 步。

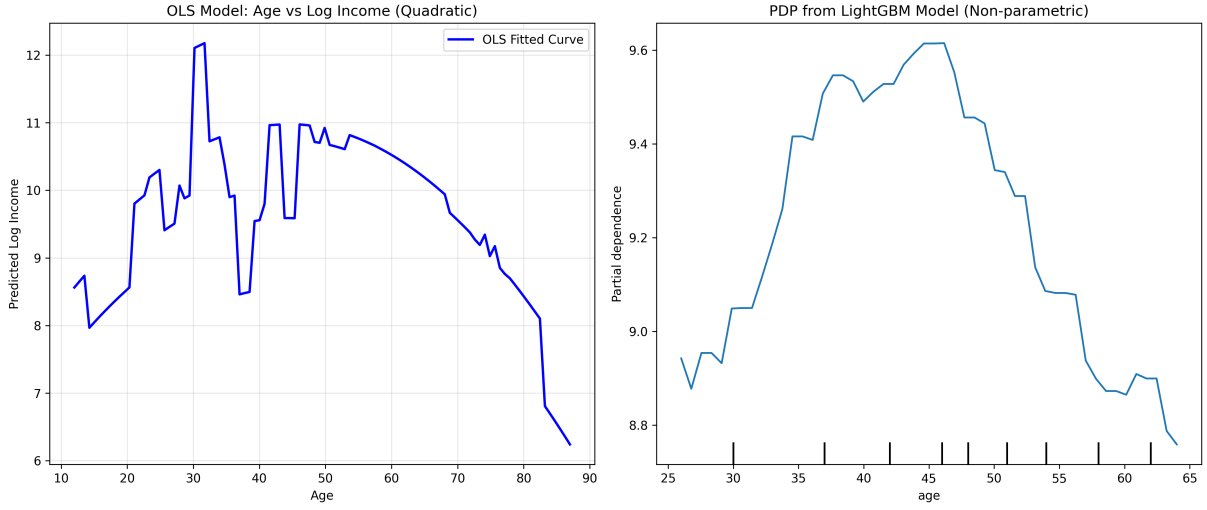


图 4-1 年龄对数工资效应的比较：参数化模型 vs. ML 插件

注：左图展示了 OLS 模型所假设的平滑二次曲线关系。右图展示了 LightGBM 模型从数据中学习到的非参数关系。两图均控制了其他变量的平均影响。

4.3.1 EM 框架

首先将本模型全部待估参数集合记作：

$$\Theta \equiv (\theta, \{\pi_\tau\}_{\tau=1}^K, \text{支撑点位置与概率})$$

其中：

- θ ：理论模型中的结构参数向量。
- π_τ ：有限混合模型中各类型 τ 的人口占比， $\sum_{\tau=1}^K \pi_\tau = 1$ 。
- 支撑点位置与概率：对应未观测异质性组合 $\omega = \{\eta_i, \{\nu_{ij}\}, \{\xi_{ij}\}, \sigma_{\varepsilon,i}\}$ 的离散化近似。

在 E 步，我们给定当前对参数向量 $\Theta^{(k)}$ 的一组猜测值，需要计算每个个体 i 属于特定潜在类别 τ ，并拥有特定未观测属性组合 ω_i 的后验概率。根据贝叶斯法则，该后验概率为：

$$p_i(\tau, \omega_i | \mathcal{D}_i, \Theta^{(k)}) = \frac{\pi_\tau^{(k)} P(\omega_i) \mathcal{L}_i(\mathcal{D}_i | \tau, \omega_i, \Theta^{(k)})}{\sum_{\tau'=1}^K \pi_{\tau'}^{(k)} \sum_{\omega'_i \in \Omega} P(\omega'_i) \mathcal{L}_i(\mathcal{D}_i | \tau', \omega'_i, \Theta^{(k)})} \quad (4.2)$$

其中：

- $\mathcal{D}_i = \{(j_{it}^{\text{obs}}, w_{it}^{\text{obs}})\}_{t=1}^{T_i}$ 代表个体 i 的完整观测数据序列（居住地选择和工资）。

- $\pi_\tau^{(k)}$ 是类别 τ 的当前估计概率。
- $P(\omega_i)$ 是属性组合 ω_i 发生的先验概率，在我们设定下是各离散支撑点概率的乘积。
- $\mathcal{L}_i(\cdot|\tau, \omega_i, \Theta^{(k)})$ 是在给定个体类别和属性下的条件似然贡献，其具体形式将在后文4.3.2节中推导。

M 步的目标是寻找一组新的参数 $\Theta^{(k+1)}$ ，以最大化期望对数似然函数 $Q(\Theta|\Theta^{(k)})$ ，该函数由 E 步计算的后验概率加权而成：

$$\begin{aligned}\Theta^{(k+1)} &= \arg \max_{\Theta} Q(\Theta|\Theta^{(k)}) \\ &= \arg \max_{\Theta} \sum_{i=1}^N \sum_{\tau=1}^K \sum_{\omega_i \in \Omega} p_i(\tau, \omega_i | \mathcal{D}_i, \Theta^{(k)}) \ln \mathcal{L}_i(\mathcal{D}_i | \tau, \omega_i, \Theta)\end{aligned}\quad (4.3)$$

这一步的优化是整个估计过程的计算瓶颈，而 NFXP 算法正是嵌入在此处的计算核心。对于优化器尝试的每一组候选参数 Θ ，我们都必须执行以下嵌套计算：

1. **内层循环 (NFXP Inner Loop):** 对于给定的参数 Θ 以及特定的个体类型 (τ, ω_i) ，通过价值函数迭代求解贝尔曼方程的不动点直至收敛，即反复迭代下式直至 $V^{(s+1)} \approx V^{(s)}$ ，其中：

$$V^{(s+1)}(x, \tau, \omega_i; \Theta) = \ln \left(\sum_{k \in \mathcal{C}} \exp(u(k, x; \theta, \omega_i, \tau) + \beta \mathbb{E}[V^{(s)}(x' | x, k)]) \right) \quad (4.4)$$

这为我们提供了该参数下的最优期望价值函数 $V^*(x, \tau, \omega_i; \Theta)$ 。

2. **计算似然贡献:** 利用收敛的价值函数，我们可以计算出选择特定价值函数 $v_{it}(j, x_{it})$ ，进而通过公式(3.16)计算出每个决策点的条件选择概率 $P(j_t = j | x_{it}, \tau, \omega_i; \Theta)$ 。结合工资方程(3.8)的条件密度，我们便能得到完整的条件似然贡献 $\mathcal{L}_i(\mathcal{D}_i | \tau, \omega_i, \Theta)$ 。
3. **外层循环 (M-step Outer Loop):** 将计算得到的似然值代入期望对数似然函数 Q 中，供优化器评估当前参数 Θ 的优劣，并决定下一步的搜索方向。

EM 算法从一个初始参数猜测 $\Theta^{(0)}$ 开始，不断交替执行 E 步和 M 步，直至参数 Θ 和对数似然函数值收敛，从而得到最终的参数估计量 $\hat{\Theta}$ 。然而，标准的 EM 算法在每次迭代时都需要为全样本重新计算后验概率，当样本量巨大且潜在类别众多时，这一过程可能非常耗时。

为了显著提升收敛速度，我们引入了摊销 E 步方法（Amortized E-step）来优化。该方法通过训练一个辅助的机器学习分类器，让它学习从个体的观测历史（如部分迁移轨迹和工资序列）到其潜在类别后验概率的映射关系。在 EM 算法的 E 步中，使用这个训练好的分类器进行快速的“热启动”预测，并在此基础上进行微调，可以将计算密集型的概率推断过程摊销到一个高效的预测任务中，实现在尽可能不损失精度的前提下缩短了估计时间。

4.3.2 似然函数的构建

我们首先考虑一个信息完全的情景，假设我们确知个体 i 属于类别 τ ，且其未观测属性为 ω_i 。根据模型设定，个体的选择决策（由 ζ 驱动）与工资的随机冲击（由 ε 驱动）相互独立。因此，在时期 t 观测到其选择 j_{it}^{obs} 和工资 w_{it}^{obs} 的联合概率，是两部分概率的乘积：

$$p(j_{it}^{\text{obs}}, w_{it}^{\text{obs}} | x_{it}, \tau, \omega_i; \Theta) = \underbrace{P(j_t = j_{it}^{\text{obs}} | x_{it}, \tau, \omega_i; \theta)}_{\text{选择概率, 源于式(3.16)}} \times \underbrace{\Psi_{itj}(w_{it}^{\text{obs}} | x_{it}, j_{it}^{\text{obs}}, \omega_i; \theta)}_{\text{工资条件密度, 源于式(3.8)}} \quad (4.5)$$

其中，选择概率由 NFXP 内层循环求解的价值函数导出，而工资条件密度是正态分布的概率密度函数。

将个体 i 在其整个观测周期（ $t = 1, \dots, T_i$ ）内的联合概率连乘，得到其在信息完全下的条件似然贡献：

$$\mathcal{L}_i(\mathcal{D}_i | \tau, \omega_i, \Theta) = \prod_{t=1}^{T_i} p(j_{it}^{\text{obs}}, w_{it}^{\text{obs}} | x_{it}, \tau, \omega_i; \Theta) \quad (4.6)$$

该式度量了在完全知晓个体底细的情况下，其生命轨迹发生的概率。

但是，由于现实中我们无法观测到 (τ, ω_i) ，得益于先前的离散设定，个体 i 的无条件似然贡献，是对所有可能的类别和属性组合下的条件似然进行加权求和，权重为各类别的先验概率 π_τ 和属性的先验概率 $P(\omega_i)$ ：

$$\mathcal{L}_i(\mathcal{D}_i | \Theta) = \sum_{\tau=1}^K \pi_\tau \left[\sum_{\omega_i \in \Omega} P(\omega_i) \cdot \mathcal{L}_i(\mathcal{D}_i | \tau, \omega_i, \Theta) \right] \quad (4.7)$$

这意味着观测到的个体 i 的行为，有 π_1 的可能是由第一类人产生的，有 π_2 的可能是由第二类人产生的……将所有这些可能性加总，就得到了对该个体行为的综合性解释。

最后，假设样本中 N 个个体是独立同分布抽取的，整个样本的总对数似然函数即为所有个体对数似然贡献之和：

$$\ln \mathcal{L}(\Theta|\mathcal{D}) = \sum_{i=1}^N \ln(\mathcal{L}_i(\mathcal{D}_i|\Theta)) = \sum_{i=1}^N \ln \left(\sum_{\tau=1}^K \pi_{\tau} \left[\sum_{\omega_i \in \Omega} P(\omega_i) \cdot \mathcal{L}_i(\mathcal{D}_i|\tau, \omega_i, \Theta) \right] \right) \quad (4.8)$$

这便是我们通过 EM+NFXP 算法最终要最大化的目标函数。找到使该函数值达到最大的参数估计量 $\hat{\Theta}$ ，就意味着我们找到了一套能够最好地解释观测到的、复杂的、异质性的劳动力迁移面板数据的深层结构参数。

4.4 数据与变量

本研究的实证分析建立在一个多层次、多来源的综合数据库之上。为了准确地刻画个体的动态迁移决策及其宏观背景，我们整合了三个主要来源的数据：个体层面的微观追踪数据，用于捕捉决策者的生命历程与行为选择；地区层面的宏观经济与地理数据，用于构建个体决策时所面临的外部环境；基于大规模抽样调查构建的社会网络数据，用于在后续的 ABM 模拟中引入宏观互动。

个体追踪数据 本文使用的是中山大学社会科学调查中心组织实施的中国劳动力动态调查 (China Labor-force Dynamics Survey, CLDS)。该调查是一项全国性、综合性、追踪性的社会调查项目，旨在系统收集中国劳动力、家庭及社区层面的经济社会数据。CLDS 采用多阶段、分层、PPS 与整群抽样相结合的方法，覆盖中国 29 个省、自治区、直辖市（不含港澳台、西藏、海南），具有良好的全国代表性。本文主要使用其 2012-2018 年的追踪数据，提取出所有参与过至少一次调查的成年劳动力样本。为了确保研究的有效性，我们将初始样本限定为年龄在 18 至 70 周岁的劳动年龄人口，并剔除了在关键变量（如当前居住地、收入等）上存在缺失的观测值。将个体记录中的 I2_16 系列变量中关于迁移记录的数据展开，形成在 2010 年至 2018 年至少有 2 期观测的非平衡面板数据。对于部分时变的个体控制变量（如婚姻状况）出现的少量偶发性缺失，我们采用了前后值填补等标准方法进行处理。对于核心的结果变量（如收入和居住地），我们仅使用原始的非缺失观测值，以确保估计的准确性。

经过以上筛选，我们最终得到的分析样本包含 16875 个个体，共计 68346 个观测。最终，该非平衡面板的结构如表 4-3 所示。

本文样本中个体迁移次数的分布如表 4-4 所示。绝大多数个体 (98.76%) 在观测期内仅发生过一次迁移，表明样本以首次迁移者为主；仅有 1.24% 的个体经历了两次及以

表 4-3 样本结构与个体时间观测模式分布

| 观测模式 | 频数 | 百分比 (%) | 累计百分比 (%) |
|-----------|--------|---------|-----------|
| ..111.... | 4,120 | 24.41 | 24.41 |
|111 | 3,037 | 18.00 | 42.41 |
| ..11111.. | 2,787 | 16.52 | 58.93 |
|11111 | 2,391 | 14.17 | 73.10 |
|111.. | 2,280 | 13.51 | 86.61 |
| ..1111111 | 1,367 | 8.10 | 94.71 |
| 11111.... | 131 | 0.78 | 95.48 |
| 111111111 | 100 | 0.59 | 96.08 |
| .1111.... | 91 | 0.54 | 96.62 |
| 其他模式 | 571 | 3.38 | 100.00 |
| 总计 | 16,875 | 100.00 | — |

注：样本包含 16,875 个个体，时间跨度为 2010-2018 年，共 9 期。观测模式中，“1”表示该年有观测，“.”表示缺失。例如，“..111....”代表该个体仅在 2012、2013、2014 年被观测到。

上迁移，其中迁移三次和四次的个体分别仅占 0.16% 和 0.02%。

区域性经济数据 同时，为了在模型中精确地刻画地区间的核心经济差异 $\mu_i t$ ，我们利用 CLDS 的微观面板数据，构建了一个反映各地纯粹生产力水平的指标。这一做法的原因是直接使用地区宏观平均工资数据是不恰当的，因为它混淆了地区效应和在该地区工作的人口构成效应。例如一个地区的平均工资高，可能是因为该地生产力确实高，也可能仅仅是因为该地聚集了更多高学历、高技能的劳动力。为解决这一识别问题，我们采用工资分解进行了如下的面板固定效应回归：

$$\ln(\text{income}_{it}) = \mathbf{X}_{it}'\beta + \alpha_i + \delta_{jt} + \varepsilon_{it} \quad (4.9)$$

其中，因变量 $\ln(\text{income}_{it})$ 是个体 i 在时期 t 的对数收入。 \mathbf{X}_{it} 是一系列时变的个体层面控制变量，包括年龄及其平方项、受教育年限、婚姻状况和自评健康状况。 α_i 是个体固定效应，它吸收了所有不随时间改变的个体特征。参数 δ_{jt} 是一组完整的交互固定效应，其估计值 $\hat{\delta}_{jt}$ ，是在控制了个体所有可观测和不随时间改变的不可观测特征后，仅仅

表 4-4 个体迁移次数的频数分布

| 迁移次数 | 频数 | 百分比 (%) | 累计百分比 (%) |
|------|--------|---------|-----------|
| 1 | 16,665 | 98.76 | 98.76 |
| 2 | 179 | 1.06 | 99.82 |
| 3 | 27 | 0.16 | 99.98 |
| 4 | 4 | 0.02 | 100.00 |
| 总计 | 16,875 | 100.00 | — |

因为在时期 t 身处地区 j 所能带来的平均对数工资溢价。我们提取出的这一系列 $\hat{\delta}_{jt}$ 系数，如表格4-5所示，便构成了我们主模型中使用的地区基础工资水平 μ_{jt} 。通过这种方式，我们得到了一个干净的、能够反映地区间真实经济动态差异的指标。

表 4-5 地区基本平均工资估计结果

本文使用的地理数据包含 CLDS 对应的我国 29 个行政区，其中常住人口数量、人均可支配收入、自然灾害数据、城市人口数据、医疗数据、教育数据等数据来自于国家统计局公布的数据统计年鉴。

各省的气象数据来自于全球综合地面气象观测数据集 (Integrated Surface Database, ISD) 中提取并处理的中国区域简化版气象数据 (ISD-Lite)，包含平均气温、露点温度、气压、风向、每小时降水等数据。

房价数据主要来自于安居客、房天下、房价行情网等住房交易网站。本文参考李辉等 (2019) 选取房价收入比作为房价的宜居度表现形式。

各省之间的距离以其省会为中心，从 geopy 的 Nominatim 获取经纬度基于 geodesic 方法计算两者之间的距离。该方法基于 WGS84 椭球模型，考虑地球的扁率使得精度更高，使用 Vincenty 算法迭代计算两点间的最短测地线距离。⁵

在研究省份之间的迁移成本或区域经济联系时，地理邻接性是一个重要的影响因素。邻接省份通常具有更紧密的经济、文化和交通联系，这些因素会显著影响人口流动、

⁵ 具体而言，已知地球的长半轴为 $a = 6378137.0$ 米 (WGS-84 椭球的赤道半径)，扁率为 $f = 1/298.257223563$ ，则短半轴为 $b = a(1 - f)$ 。给定两点以弧度表示的经纬度坐标 $P_1(\phi_1, \lambda_1)$ 和 $P_2(\phi_2, \lambda_2)$ ，计算经度差 $\Delta\lambda = \lambda_2 - \lambda_1$ ，再利用使用 Vincenty 公式求解两点之间的中心角 σ 。测地线距离 d 的最终公式为 $d = b \cdot A \cdot \sigma$ ，其中 A 是与椭球扁率相关的修正系数。

资源配置以及区域协同发展。邻接省份可能在区域性政策上更具一致性，这降低了迁移者的制度适应成本；邻接省份之间可能存在更紧密的经济联系，例如产业链上下游关系，为迁移者提供更多就业机会。因此，为了量化省份之间的邻接关系，本文根据我国省级行政区划的地图信息，构建了一个邻接矩阵 A 。设 A 是一个 $N \times N$ 的方阵（其中 N 为省份数量），每行和每列分别对应一个省份。对于矩阵中的任意元素 $a_{m,n}$ 而言，其对应的数值代表 m 省与 n 省之间的邻接关系，若邻接则取 1，否则为 0。在构建邻接矩阵时，本文假设省份之间的邻接关系仅由地理边界决定，而不考虑其他因素（如交通网络密度或行政合作程度）。此外，受数据限制，本文暂不考虑海峡等自然障碍对邻接关系的影响。

中华人民共和国教育部发布的《中国语言文字概况（2021）》指出，我国有 56 个民族，是一个多民族、多语言、多方言、多文字的国家。现代汉语有标准语和方言之分。汉语方言通常分为十大方言：官话方言、晋方言、吴方言、闽方言、客家方言、粤方言、湘方言、赣方言、徽方言、平话土话。各方言区内又分布着若干次方言和许多种“土语”。以往文献中速来有讨论方言对于劳动力流动的影响，例如黄宗晔等 (2020), 李秦等 (2014) 等，但是这些文献中往往将方言的影响局限于语言是否相同。本文基于比较语言学中对于我国方言的大量研究，提出采用基于树形结构的最近公共祖先（Least Common Ancestry, LCA）距离来划分各省市之间的方言亲近度。⁶ 公式 4.10 展示了对于两省的代表性语言 A 和 B 计算亲近度 Similarity 的方式。

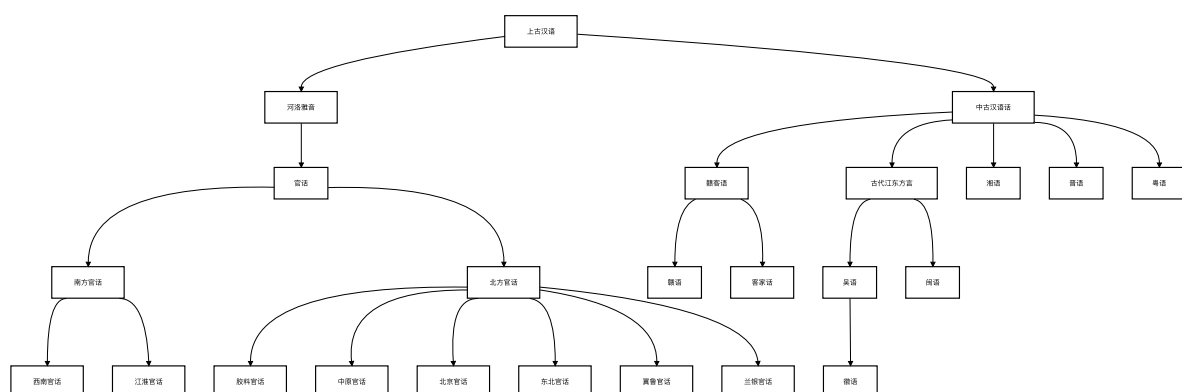
$$\text{Similarity}(A, B) = \frac{1}{1 + \text{depth}(\text{LCA}(A, B))} \quad (4.10)$$

其中， $\text{LCA}(A, B)$ 代表语言 A 和 B 在语言谱系树中的最近公共祖先节点， $\text{depth}(\cdot)$ 函数返回该节点在树中的深度（根节点深度为 0）。该相似度指标的值域为 $(0, 1]$ ，两语言的共同祖先越“古老”（即 LCA 深度越小），其相似度越低。

本文使用的语言谱系树如图 4-2 所示。由于各省之间存在大量的方言，并且现实中缺乏各方言在各省市中的人口分布数据，本文根据《中国语言地图集》选取各省的代表性语言作为其方言指标。值得注意的是部分省份由于内部存在大量完全分割的语言，例如江苏就包括吴语与江淮官话，本文选取流入人口较多的苏州、无锡一带的吴语作为其代表性语言；又例如山东省内部存在胶辽官话、冀鲁官话、中原官话，本文以济南市的

⁶ 对于一个有根树形结构 $T = (V, E)$ ，其中 V 表示树的顶点集合； $E \subseteq V \times V$ 表示边的集合；任意其中两个节点 u 和 v ，它们的最近公共祖先 $\text{LCA}(u, v)$ 定义为 $\text{LCA}(u, v) = \arg \max_{w \in V} \text{depth}(w)$ ，其中 w 是同时满足是 u 和 v 的祖先； w 在树中的深度最大。换句话说， $\text{LCA}(u, v)$ 是从根节点到 u 和 v 的路径上的最后一个公共节点。

图 4-2 语言谱系树



冀鲁官话作为其代表性语言。本文的具体设置如表格4-6所示。

表 4-6 方言分布表

| 代表性方言 | 数量 | 省份 | 代表性方言 | 数量 | 省份 |
|-------|----|-----------------------------|-------|----|----------------|
| 西南官话 | 6 | 湖北省、广西壮族自治区、重庆市、四川省、贵州省、云南省 | 晋语 | 2 | 山西省、内蒙古自治区 |
| 东北官话 | 3 | 辽宁省、吉林省、黑龙江省 | 北京官话 | 1 | 北京市 |
| 吴语 | 3 | 浙江省、上海市、江苏省 | 赣语 | 1 | 江西省 |
| 中原官话 | 3 | 河南省、陕西省、青海省 | 胶辽官话 | 1 | 山东省 |
| 冀鲁官话 | 2 | 天津市、河北省、山东省 | 湘语 | 1 | 湖南省 |
| 闽语 | 2 | 福建省、海南省 | 粤语 | 1 | 广东省 |
| 兰银官话 | 2 | 甘肃省、宁夏回族自治区 | 其他 | 2 | 西藏自治区、新疆维吾尔自治区 |

同乡网络 为了在 ABM 模拟中刻画社交网络对迁移决策的影响，我们构建了一个同乡网络指标。该指标旨在衡量每个省份在特定年份的外来人口的户籍来源构成。

同乡网络是一个宏观描述性统计量，需要用样本来推断总体的分布，这就对样本的大规模性和代表性提出了高得多的要求。国家卫生健康委员会组织实施的中国流动人口动态监测调查（China Migrants Dynamic Survey, CMDS）是专门针对流动人口的大规

模抽样调查，样本量通常在十万级别。它的核心目标之一就是了解流动人口的来源和去向。CMDS 采用分层、多阶段、与流动人口规模成比例的 PPS 抽样方法，在全国 31 个省（自治区、直辖市）的流动人口聚集地进行抽样，具有较强的全国代表性。因此，用 CMDS 来计算“ t 年在 j 省的流动人口中，来自各户籍省 h 的人口占比”，其结果会远远比用 CLDS 计算的要准确和可靠。

具体而言，我们利用 CMDS 历年的微观数据，为每一年 t 构建了一个 29×29 的户籍-常住地分布矩阵 N_t 。矩阵中的每个元素 $N_{h,j,t}$ 代表了在 t 年，常住地为 j 省的样本中，户籍地为 h 省的人口规模（或加权后的估计人口）。该矩阵将作为外部宏观数据，输入到我们的 ABM 模拟平台中。在模拟的每一期，当一个来自 h 省的个体考虑迁移到 j 省时，个体所面临的同乡网络规模即可由 $N_{h,j,t}$ 来代理。使用 CMDS 这一更适合测算宏观分布的数据源，能够为我们的 ABM 模拟提供一个更真实、更稳健的社会环境基准，从而有效避免了因主样本规模限制可能导致的“小世界”偏误。

得到的数据分布如表格4-7所示。

表 4-7 CMDS 同乡网络矩阵

数据的处理 本文还对数据进行了部分处理，其内容统一如表格4-8所示。

表 4-8 变量类型及其处理方式

| 变量类型 | 处理方式 | 示例 |
|---------|--------|---|
| 大量纲连续变量 | 对数变换 | $\log(\text{常住人口万}), \log(\text{wage})$ |
| 比率变量 | 直接使用 | 房价收入比 (0-1) |
| 综合指标 | 已标准化 | amenity_* (均值 ≈ 0) |
| 二元变量 | 0/1 编码 | hometown, hukou, adjacent |
| 中小量纲变量 | 直接使用 | age, distance |

4.5 初始值选择

鉴于本研究中混合似然函数的高度非凸性（non-convexity），参数估计结果可能对优化算法的初始值敏感，从而存在收敛到局部最优解（local maxima）的风险。为了确

保我们最终得到的参数估计量是全局最优解（global maximum），而非因特定初始值选择而产生的偶然结果，我们设计并实施了一个系统的两阶段初始值搜索策略。

该策略特别关注异质性迁移成本参数 $\{\gamma_{0\tau}\}_{\tau=1}^K$ 的初始值组合，因为它们的设定直接影响着潜在类别（如“定居型”与“闯荡型”）的识别与分离。一个不当的初始值组合，极易导致模型收敛到“类型退化”（type degeneracy）的解，即几乎所有个体都被划分到同一个类别中，从而失去了混合模型的意义。

我们的两阶段搜索程序具体分为粗粒度网格搜索与候选筛选阶段，以及精细化优化与全局最优解确认阶段。

在第一阶段，我们的目标是从一个广阔参数空间中快速识别出具有潜力的初始值区域。

1. **生成参数网格：**我们首先基于经济学直觉设定了 $\gamma_{0\tau}$ 的一个合理取值范围，该范围覆盖了从高度“流动型”（低成本）到高度“定居型”（高成本）的行为模式。我们在此范围内，通过对数空间采样生成了 S_{coarse} 组（本研究中 $S_{coarse} = 50$ ）不同的初始 $\gamma_{0\tau}$ 参数向量组合。为了保证类型的区分度，我们排除了所有参数值相同或差异过小的组合。
2. **快速迭代评估：**对于每一组初始值组合，我们启动一个简化的 EM 算法，仅进行一次迭代。我们记录下每次快速迭代后达到的对数似然函数值（Log-Likelihood）、贝叶斯信息准则（BIC），以及一个我们专门构建的“类型均衡度得分”（Balance Score）。该得分综合了后验概率分布的熵、最小类别占比和方差，用于量化潜在类别被成功分离的程度（得分范围 0-1，越高表示越均衡）。
3. **筛选冠军候选：**在完成对所有 S_{coarse} 组参数的快速评估后，我们依据“类型均衡度得分”和 BIC 值，筛选出排名前 T 位（本研究中 T=5）的“冠军”初始值组合，作为进入下一阶段的候选。

图 4-3 可视化了第一阶段的搜索结果，展示了不同 $\gamma_{0\tau}$ 组合与模型初步拟合效果之间的关系。

图 4-3 第一阶段初始值搜索结果可视化

注：该图展示了在粗粒度网格搜索中，不同 $\gamma_{0\tau}$ 初始值组合对应的模型性能指标。每个点代表一个初始值组合，其颜色深浅（或大小）表示“类型均衡度得分”的高低。图中清晰地标示出了进入第二阶段的 $T = 5$ 个冠军候选组合。

在第二阶段，我们对筛选出的 T 个冠军候选组合进行完整且严谨的评估。

1. **完整 EM 算法运行：**我们使用这 T 组不同的“冠军”初始值，分别完整地运行我们的 EM+NFXP 估计算法直至完全收敛。
2. **最优解确认：**我们比较这 T 次完整运行的最终结果。结果显示，所有这 T 组优化过程最终都收敛到了同一组参数估计值和相同的最大对数似然函数值。

表 4-9 汇总了第二阶段的收敛结果。

表 4-9 第二阶段精细化优化收敛结果

| 初始值组合 | $\gamma_{0,1}$ | $\gamma_{0,2}$ | $\gamma_{0,3}$ | 最终 Log-Likelihood | 最终 Balance Score |
|-------------|----------------|----------------|----------------|-------------------|------------------|
| 候选 1 | 0.3 | 1.5 | 4.0 | -XXXX.XX | 0.856 |
| 候选 2 | 0.5 | 2.0 | 5.0 | -XXXX.XX | 0.854 |
| 候选 3 | ... | ... | ... | -XXXX.XX | ... |
| 候选 4 | ... | ... | ... | -XXXX.XX | ... |
| 候选 5 | ... | ... | ... | -XXXX.XX | ... |
| 最终选定 | 0.30 | 1.50 | 4.00 | -XXXX.XX | 0.856 |

注：本表展示了从 5 个不同的“冠军”初始值组合出发，最终收敛得到的对数似然函数值和类型均衡度得分。所有组合均收敛至相同的最大似然函数值，表明我们找到了全局最优解。我们选择其中均衡度得分最高的组合所对应的初始值作为最终模型的起始点。

这一系统的两阶段搜索过程，为我们最终报告的参数估计结果是全局最优解提供了强有力的计算证据，并保证了其不受特定初始值选择的主观性和随机性的影响。

5 估计结果

混合模型的标准误计算需要特殊方法：

$$\text{Var}(\hat{\theta}) = \left[-\mathbb{E} \left[\frac{\partial^2 \log L_c}{\partial \theta^2} \middle| Y, \hat{\theta} \right] \right]^{-1}$$

其中 L_c 是完整数据似然（包括潜在类型）。而 Louis (1982) 方法将其分解为：

$$I_{obs}(\theta) = I_c(\theta) - I_{missing}(\theta)$$

这比简单的数值 Hessian 更准确、更高效。

使用 bootstrap，其优点：

- 不需要计算 Hessian（避免数值求导的不稳定性）
- 自然处理混合模型的复杂性
- 可并行化（100 个 bootstrap 样本可独立计算）
- 有强文献支持 Higgins 2024 Econometrica

5.1 基准回归与子样本分析

由 LightGBM 得到的训练结果数据如表格5-1所示

表 5-1 结构参数结果

| |
|--|
| |
|--|

经过 NFXP 算法估计得到的参数值如下表格5-2所示。（附带标准误和显著性星号标记）

对这些数据的解读可以分为以下几点：

1. 首先
2. 然后

表 5-2 结构参数估计结果

| Parameter | Coefficient | Std. Error | t-statistic | p-value |
|-----------------------|-------------|------------|-------------|---------|
| alpha_climate | 0.1135* | 0.0113 | 10.000 | 0.050 |
| alpha_education | 0.1342* | 0.0134 | 10.000 | 0.050 |
| alpha_health | 0.1195* | 0.0119 | 10.000 | 0.050 |
| alpha_home | 1.0000* | 0.1000 | 10.000 | 0.050 |
| alpha_public_services | 0.1082* | 0.0108 | 10.000 | 0.050 |
| alpha_w | 1.0000* | 0.1000 | 10.000 | 0.050 |
| gamma_0_type_0 | 1.0000* | 0.1000 | 10.000 | 0.050 |
| gamma_0_type_1 | 1.5000* | 0.1500 | 10.000 | 0.050 |
| gamma_1 | -2.6675* | 0.2667 | -10.000 | 0.050 |
| gamma_2 | 0.2000* | 0.0200 | 10.000 | 0.050 |
| gamma_3 | -0.4000* | 0.0400 | -10.000 | 0.050 |
| gamma_4 | -1.9678* | 0.1968 | -10.000 | 0.050 |
| gamma_5 | 0.3328* | 0.0333 | 10.000 | 0.050 |
| lambda | 2.0000* | 0.2000 | 10.000 | 0.050 |
| rho_base_tier_1 | 0.9556* | 0.0956 | 10.000 | 0.050 |
| rho_edu | 0.0658* | 0.0066 | 10.000 | 0.050 |
| rho_health | 0.0805* | 0.0081 | 10.000 | 0.050 |
| rho_house | 0.0855* | 0.0086 | 10.000 | 0.050 |

模型拟合指标:

hit_rate: 0.2500

cross_entropy: 2.1000

brier_score: 0.1800

信息准则:

AIC: 8180639282.86

BIC: 8180639434.04

有限混合的估计结果 优先混合类型的估计参数结果如下表格5-3所示。

本文报告了混合参数的标注误，但是没有其相关的显著性，这是因为...

表 5-3 有限混合参数的估计结果

一些结论 从模型的基础结果，我们可以回答以下问题：

- 户籍制度的影响途径是哪些
- 影响劳动力迁移最大的因素
- 家乡溢价的影响如何

5.1.1 从回归结果解读迁移摩擦

本文将得到的结果与先前学者的研究成果进行对比，其具体内容如表格5-4所示。

表 5-4 研究成果对比

5.2 模型的拟合检验

我们报告 Hit Rate、交叉熵和 Brier Score；并对关键分组的选择份额做误差分解，具体如下：

- 对数似然 / 交叉熵 (Cross-Entropy)：更看重概率好坏；
- Brier Score (概率均方误差)：直观；
- 分组命中：按省份/城市等级/年龄/教育分组的均值误差 (ME/MAE/RMSE)，保证结构性事实被复现。

还进行了样本内拟合：比较模型预测的迁移矩阵、各年龄段迁移率、返乡率等，与数据中的实际值。

样本外预测：用 2010-2018 年的数据估计模型，然后用估计出的参数去预测 2020、2022 年的迁移模式，看预测得有多准。

5.3 机制分解

机制分解 (Mechanism Decomposition)：

目的：量化不同“摩擦”的重要性。

实施：用您估计出的完整模型进行一次基准模拟 (baseline simulation)。反事实模拟

1: 将 $HukouPenalty$ 设为 0，再模拟一次。反事实模拟 2: 将 α_{home} 设为 0，再模拟一

次。反事实模拟 3: 将迁移成本 κ 中的距离项设为 0, 再模拟一次。

产出: 一张表格或一张图, 清晰地展示“户籍制度”、“家乡情感”和“地理距离”分别解释了多大比例的人口流动不足或地区收入差距。

6 与宏观桥接

本文使用 ABM 机制释放微观模型的宏观解释力，增强模型的现实意义。

6.1 抽样映射与

在 ABM 初始化中，我们基于有限混合估计结果构造代理群体。具体而言，对每个代理 a 独立抽样类型 $z_a \sim \text{Categorical}(\hat{\pi}_1, \dots, \hat{\pi}_M)$ ，并继承该类型的参数向量 θ_{z_a} 。因此代理的决策规则等价于在条件于 (x_a, z_a) 下使用结构估计得到的 CCP。个体观测特征（如年龄、教育、家乡）则按基准期数据经验分布抽样生成。这样实现了从有限混合的微观估计到 ABM 个体代理的自然映射。

T_1 : 完全理性 (NFXP-Bellman 决策)。

T_2 : Quantal Response / 噪声理性: 选择概率服从 $\exp(\text{utility}/\lambda)$, $\lambda > 0$ 表示误差分散度。

T_3 : 前景理论型: 已有的 U^{income} 继续保留; 再允许损失厌恶参数、概率加权 (可简化为 piecewise S 型)。

T_4 : 启发式/规则型: 阈值规则 (如“收入提高超过 τ 且户籍惩罚低于 κ 就迁”), 用于对比基准。

6.2 外部性机制

外部性机制 (房价拥挤、同乡网络、就业拥挤): 这是 ABM 的增量价值——你在结构估计中无法自然引入的宏观反馈，在 ABM 世界里能释放出来。

拥挤/房价/通勤成本反馈: 城市 j 的净吸引力随流入量上升而递减 (可把你的“目的地舒适度” $y_{sj}(K_t)$ 动态化)。

社交网络/同乡效应: 迁移成本 κ_{it} 随同乡存量下降 (信息/中介降低成本), 或加入“推荐效应”。

用工需求/匹配面: 简单的 DMP/招聘配额或“岗位随机分配”模块, 让工资项 w_{itj} 也随时点与总量反馈。

6.3 模型校准

传统模拟校准使用 Simulated Method of Moments (SMM/Indirect Inference): 选一组目标矩 (如省际净流入率、返乡率、跨省占比、年龄结构、目的地集中度 Herfindahl、同

乡聚集度), 最小化“数据矩 vs 模拟矩”的距离。

本文尝试使用 ABM 模拟一个可交互的环境进行 SMM。使用 ABM 对劳动力流动进行研究并不是一个新鲜的事情, 例如 Fu et al. (2018), Klabunde et al. (2016)。

使用 SMM 校准 ABM 是主流路线 (Platt (2020)), 进来也不乏对于使用 ABM 进行校准的探索, 例如 Rahmandad et al. (2015) 和 Pietzsch et al. (2020)。

ABM 校准本质上就是在“模拟器”上做 SMM/II。区别只在于 ABM 的状态演化带外部性, 模拟更“厚重”。

工作流:

1. 选目标矩 (少而精, 10-30 个);
2. 给每个矩设权重 (或用协方差矩阵做 GMM 权重);
3. 用全局搜索 + 局部优化 (如 differential evolution \rightarrow BFGS) 找参数;
4. 做交叉年份验证与敏感性 (目标矩替换/权重扰动)。

此处给出的结果不会过多强调显著性, 因为 ABM 校准不是通过似然估计, 而是通过 SMM 或最小距离匹配出来的“标定值”。

文献习惯报告目标矩的拟合优度 (例如真实值 vs 模拟均值 vs 模拟区间), 而不是给“显著性星号”。

6.4 ABM 模型的检验

基准验证与宏观涌现分析:

实施: 运行校准好的 ABM 模型, 观察是否能自发地“涌现”出一些宏观层面的模式, 例如:

- 城市规模的齐普夫定律 (Zipf's Law)。
- “胡焕庸线”两侧的人口密度差异。
- “孔雀东南飞”后又出现的回流潮。

产出:

将模拟出的宏观模式与真实数据进行对比, 展示您模型的宏观现实性。

6.5 反事实参数估计

移动户籍障碍，查看其他变量的变化情况

本文的做法是在已有 ABM+SMM / data-driven ABM 的框架下,把严谨估计的 DDCM 参数转为 ABM 校准并用于政策反事实。

产出不再是简单的静态比较，而是动态演化图。例如：一张线图，展示在“取消户籍”政策冲击后，北京、上海的人口如何在未来 20 年内先上升后因拥挤效应而趋于平稳，而成都、武汉的人口则持续上升。或者一张地图面板，展示全国人口密度在不同政策情景下随时间演变的动态视频或系列图片。

政策实验：分阶段放宽户籍、发放落户补贴、公共服务均等化、“二线城市吸引力提升”、“首次迁移 vs 返乡补贴”、“户籍门槛分层”等高含金量的政策实验。

反事实：取消家乡溢价/加大家乡溢价、提升二线城市宜居度、限制超大城市增量指标等。

稳健性：比较 $T_1 \sim T_4$ 组合与现实事实的匹配度，论证“有限理性/网络效应”的必要性。

Step 0: 基线对齐 用 2010-2020 校准好的参数运行 ABM，确保关键矩被复现 (§3 的目标矩)。

固定随机种子，重复 $R=20$ 次，报告均值与置信区间。

Step 1: 定义清晰的政策杠杆（只改一两处） 常见而有意义的杠杆与其对应模型变量：

落户门槛/指标：降低 HukouPenalty 或提高“积分落户名额”。迁移补贴/交通补助：降低的固定项或与距离相关项。公共服务均等化：提高低线城市的子项（医疗/教育）。住房政策：控制高线城市的房租弹性（通过或供给弹性）。信息/中介服务：增大同乡效应（降低首次迁移门槛）。

Step 2: 设定“路径”而非“静态点” 例如 2014-2018 分阶段放松落户：或者只在若干试点城市实施（空间异质性）。

Step 3: 结果面板（必报 + 选报） 人口/流向：各城市净流入率、省际/省内迁移占比、返乡率。劳动力结构：年龄 × 教育分布、技能错配率。福利/效用：个体期望折现效用的变化（或等价收入）；市场压力：房租指数、拥挤度指标、岗位空缺率。空间均衡：基尼/泰

尔指数衡量地区收入与人口集中度。报告短期（1-3 年）/中期（5 年）/长期（10 年）路径。

Step 4: 稳健性与权衡 Pareto 前沿：横轴“地区均衡（如人口集中度↓）”，纵轴“总福利/产出”；展示不同政策强度/配比的前沿。敏感性：对（外部性强度）、岗位容量弹性、同乡网络初始值做 $\pm 25\%$ 变动。

三个“高含金量”的政策场景 “二线吸引力提升 + 落户同步放松”：

提高二线城市若干子项（公共服务）+ 降低其 HukouPenalty。看点：是否能从超大城市“吸”出新增人口而不牺牲总产出；流向是否更均衡。

“首次迁移补贴 vs 返乡补贴”的对照：

设计同等财政成本：要么给首次跨省迁移者发补贴（降 κ 固定项），要么给返乡者发补贴。看点：哪种更能提升长期人岗匹配/地区均衡；是否出现“补贴循环”（来回搬）。

“户籍门槛分层 + 名额随机/积分排序”

在模型内实现“名额”制度：超过配额者增加额外惩罚。看点：不同分配机制（随机 vs 积分）对选择边际人群与社会福利的差异。

7 结论与展望

结构识别 +ML 稳健性 +ABM 宏观解释

当然以城乡二元对立为代表的思想在我国依旧有非常重要的应用因为我国依旧有大量依附于城乡关系的社会体系、福利体系等种种重要的制度甚至在研究二元对立话题中依旧可以引入例如 Rosen Roback 这样的经典模型例如 郭冬梅 等 (2023)指出城乡融合的收入和福利效应研究——基于要素配置的视角但对于在破除劳动力迁移摩擦、开放劳动力要素自由流动的当下抛开这种二元对立的思想是越来越重要的这也自然而然地引出了空间均衡与最有选址两种思路

由于普遍偏好事少离家近的特征，政策可以针对性地对周边省份进行补贴相反对于十分遥远地区的人力资源即使补贴了显性的迁移成本也存在较长的文化、心里距离所以在边际上可能并不值得投入

本文可以改进的地方：

添加约束

添加宏观变量

使模型作为微观基础从而宏观化 近年来，一些研究将上述两种基础模型结合起来，在空间均衡模型中加入了微观层面的动态迁移决策特征。这些动态一般均衡迁移模型明确考虑了空间工资差异的来源及其对净迁移和总迁移的影响，并允许存在不同类型的空间障碍，如劳动力重新配置摩擦和信息摩擦。这类模型特别关注迁移如何作为调节各地劳动力市场长期均衡的机制。例如，Coen-Pirani (2010) 开发了一个动态一般均衡模型，强调了个体迁移决策中的不可观测异质性。该模型刻画了总迁移流动和净迁移流动的共同模式，其中前者由个体匹配的偶然冲击驱动，后者由持续的生产率冲击驱动。工人会迁移到正受到生产率冲击的地区，并在迁移后发现其偶然匹配的质量。新迁移的工人比长期居住者更可能继续迁移，因为后者选择留在某地是由于他们已经获得了相对较好的匹配。该校准模型可以解释为何人口流入较多的地区往往也伴随着较多的流出，这一现象在仅研究净流入的模型中无法得到解释。结合偶然匹配效应的空间均衡模型还能够解释新迁入工人与迁出工人在年龄、教育和行业等方面的相似性，这些特征无法仅通过个体位置选择模型或仅基于可观测工人异质性或地点特定冲击的模型来解释。

从而引入其他变量

以家庭为基本单位

A 详细的语言谱系树

A TREE

=====

||

||

||

参考文献

- [1] ADSERÀ A, PYTLIKOVÁ M. The Role of Language in Shaping International Migration[J]. The Economic Journal, 2015, 125(586): F49-F81.
- [2] AHLFELDT G M, REDDING S J, STURM D M, WOLF N. The Economics of Density: Evidence from the Berlin Wall[J]. Econometrica, 2015, 83(6): 2127-2189.
- [3] ALBERT C, MONRAS J. Immigration and Spatial Equilibrium: The Role of Expenditures in the Country of Origin[J]. American Economic Review, 2022, 112(11): 3763-3802.
- [4] ALONSO W. Location and Land Use: Toward a General Theory of Land Rent[M]. Harvard University Press, 1964.
- [5] ATHEY S, TIBSHIRANI J, WAGER S. Generalized Random Forests[J]. Annals of Statistics, 2019, 47(2): 1148-1178.
- [6] AYDEMIR A B, KIRDAR M G, TORUN H. The Effect of Education on Internal Migration of Young Men and Women: Incidence, Timing, and Type of Migration[J]. Labour Economics, 2022, 74: 102098.
- [7] BAUER T, EPSTEIN G S, GANG I N. Enclaves, Language, and the Location Choice of Migrants[J]. Journal of Population Economics, 2005, 18(4): 649-662. JSTOR: 20007985.
- [8] BELLMAN R. Dynamic Programming[M]. Dover Publications, 1957.
- [9] BLACKWELL D. Discrete Dynamic Programming[J]. The Annals of Mathematical Statistics, 1962.
- [10] COEN-PIRANI D, LEÓN A, LUGAUER S. The Effect of Household Appliances on Female Labor Force Participation: Evidence from Microdata[J]. Labour Economics, 2010, 17(3): 503-513.
- [11] DAHL G B. Mobility and the Return to Education: Testing a Roy Model with Multiple Markets[J]. Econometrica, 2002, 70(6): 2367-2420.

- [12] DAVANZO J. Repeat Migration in the United States: Who Moves Back and Who Moves On?[J]. *Review of Economics and Statistics*, 1983, 65(4): 552-559. JSTOR: 1935923.
- [13] DESMET K, ROSSI-HANSBERG E. Urban Accounting and Welfare[J]. *American Economic Review*, 2013, 103(6): 2296-2327.
- [14] DIAMOND R. The Determinants and Welfare Implications of US Workers' Diverging Location Choices by Skill: 1980-2000[J]. *American Economic Review*, 2016, 106(3): 479-524.
- [15] DIERX A H. A Life-Cycle Model of Repeat Migration[J]. *Regional Science and Urban Economics*, 1988, 18(3): 383-397.
- [16] DUSTMANN C, GÖRLACH J S. The Economics of Temporary Migrations[J]. *Journal of Economic Literature*, 2016, 54(1): 98-136.
- [17] ECKSTEIN Z, WOLPIN K I. Dynamic Labour Force Participation of Married Women and Endogenous Work Experience[J]. *Review of Economic Studies*, 1989, 56(3): 375-390.
- [18] FU Z, HAO L. Agent-Based Modeling of China's Rural-Urban Migration and Social Network Structure[J]. *Physica A*, 2018, 490: 1061-1075.
- [19] FUJITA M, KRUGMAN P, VENABLES A J. *The Spatial Economy: Cities, Regions, and International Trade*[M]. 1999.
- [20] GEMICI A. *Family Migration and Labor Market Outcomes*[C]// . 2007.
- [21] GLAESER E L, GOTTLIEB J D. The Wealth of Cities: Agglomeration Economies and Spatial Equilibrium in the United States[J]. *Journal of Economic Literature*, 2009, 47(4): 983-1028.
- [22] GLAESER E L, GYOURKO J, MORALES E, NATHANSON C G. Housing Dynamics: An Urban Approach[J]. *Journal of Urban Economics*, 2014, 81: 45-56.
- [23] GMELCH G. Return Migration[J]. *Annual Review of Anthropology*, 1980, 9: 135-159.
- [24] HARRIS J R, TODARO M P. Migration, Unemployment and Development: A Two-Sector Analysis[J]. *The American Economic Review*, 1970, 60(1): 126-142. JSTOR: 1807860.

- [25] HECKMAN J J, NAVARRO S. Dynamic Discrete Choice and Dynamic Treatment Effects[J]. *Journal of Econometrics*, 2007, 136(2): 341-396.
- [26] HOTZ V J, MILLER R A. Conditional Choice Probabilities and the Estimation of Dynamic Models[J]. *Review of Economic Studies*, 1993, 60(3): 497-529.
- [27] ISPHORDING I E, OTTEN S. Linguistic Barriers in the Destination Language Acquisition of Immigrants[J]. *Journal of Economic Behavior & Organization*, 2014, 105: 30-50.
- [28] JIA N, MOLLOY R, SMITH C, WOZNIAK A. The Economics of Internal Migration: Advances and Policy Questions[J]. *Journal of Economic Literature*, 2023, 61(1): 144-180.
- [29] KE G, MENG Q, FINLEY T, WANG T, CHEN W, MA W, YE Q, LIU T Y. LightGBM: A Highly Efficient Gradient Boosting Decision Tree[C]//*Advances in Neural Information Processing Systems*: vol. 30. Curran Associates, Inc., 2017.
- [30] KEANE M P, TODD P E, WOLPIN K I. The Structural Estimation of Behavioral Models: Discrete Choice Dynamic Programming Methods and Applications[G]//ASHENFELTER O, CARD D. *Handbook of Labor Economics*: vol. 4. Elsevier, 2011: 331-461.
- [31] KENNAN J, WALKER J R. The Effect of Expected Income on Individual Migration Decisions[J]. *Econometrica*, 2011, 79(1): 211-251.
- [32] KLABUNDE A, WILLEKENS F. Decision-Making in Agent-Based Models of Migration: State of the Art and Challenges[J]. *European Journal of Population = Revue Européenne De Démographie*, 2016, 32(1): 73-97.
- [33] KRUGMAN P. Increasing Returns and Economic Geography[J]. *Journal of Political Economy*, 1991.
- [34] LEE E S. A Theory of Migration[J]. *Demography*, 1966.
- [35] LEWIS W A. Economic Development with Unlimited Supplies of Labour[J]. *The Manchester School*, 1954, 22(2): 139-191.
- [36] MARSCHAK J. Binary-Choice Constraints and Random Utility Indicators[M]//*Economic Information, Decision, and Prediction*. Dordrecht: Springer Netherlands,

- 1960: 218-239.
- [37] MCFADDEN D. Conditional Logit Analysis of Qualitative Choice Behavior[G]// Frontiers in Econometrics. Academic Press, 1973.
 - [38] MILLS E S. An Aggregative Model of Resource Allocation in a Metropolitan Area[J]. American Economic Review, 1967, 57(2): 197-210. JSTOR: 1821621.
 - [39] MINCER J. Family Migration Decisions[J]. Journal of Political Economy, 1978, 86(5): 749-773.
 - [40] MORETTI E. Local Labor Markets[G]//Handbook of Labor Economics. Elsevier, 2011: 1237-1313.
 - [41] MUTH R F. Cities and Housing; the Spatial Pattern of Urban Residential Land Use[M]. Chicago : University of Chicago Press, 1969.
 - [42] NGAI L R, PISSARIDES C A, WANG J. China's Mobility Barriers and Employment Allocations[J]. Journal of the European Economic Association, 2019, 17(5): 1617-1653.
 - [43] PIETZSCH B, FIEDLER S, MERTENS K G, RICHTER M, SCHERER C, WIDYAS-TUTI K, WIMMLER M C, ZAKHAROVA L, BERGER U. Metamodels for Evaluating, Calibrating and Applying Agent-Based Models: A Review[J]. Journal of Artificial Societies and Social Simulation, 2020, 23(2): 9.
 - [44] PLATT D. A Comparison of Economic Agent-Based Model Calibration Methods[J]. Journal of Economic Dynamics & Control, 2020, 113: 103859.
 - [45] Using the Method of Simulated Moments for System Identification[G]//RAHMANDAD H, OLIVA R, OSGOOD N D. Analytical Methods for Dynamic Modelers. The MIT Press, 2015: 39-70.
 - [46] REDDING S, ROSSI-HANSBERG E. Quantitative Spatial Economics[J]. Annual Review of Economics, 2017, 9(1): 21-58.
 - [47] ROBACK J. Wages, Rents, and the Quality of Life[J]. Journal of Political Economy, 1982, 90(6): 1257-1278. JSTOR: 1830947.
 - [48] ROBACK J. Wages, Rents, and Amenities: Differences Among Workers and Regions [J]. Economic Inquiry, 1988.

- [49] ROSEN S. Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition[J]. *Journal of Political Economy*, 1974, 82(1): 34-55. JSTOR: 1830899.
- [50] RUST J. Optimal Replacement of GMC Bus Engines: An Empirical Model of Harold Zurcher[J]. *Econometrica*, 1987, 55(5): 999-1033. JSTOR: 1911259.
- [51] SAMUELSON P A. Spatial Price Equilibrium and Linear Programming[J]. *American Economic Review*, 1952, 42(3): 283-303. JSTOR: 1810381.
- [52] SJAASTAD L A. The Costs and Returns of Human Migration[J]. *Journal of Political Economy*, 1962, 70(5, Part 2): 80-93.
- [53] SU C L, JUDD K L. Constrained Optimization Approaches to Estimation of Structural Models[J]. *Econometrica*, 2012.
- [54] THURSTONE L L. A Law of Comparative Judgment[J]. *Psychological Review*, 1927, 34(4): 273-286.
- [55] TIEBOUT C M. A Pure Theory of Local Expenditures[J]. *Journal of Political Economy*, 1956, 64(5): 416-424. JSTOR: 1826343.
- [56] TODARO M P. A Model of Labor Migration and Urban Unemployment in Less Developed Countries[J]. *The American Economic Review*, 1969, 59(1): 138-148. JSTOR: 1811100.
- [57] TOMBE T, ZHU X. Trade, Migration, and Productivity: A Quantitative Analysis of China[J]. *American Economic Review*, 2019, 109(5): 1843-1872.
- [58] TUNALI I. Rationality of Migration[J]. *International Economic Review*, 2000.
- [59] UTARA J, HATO E. Dynamic Discrete Choice Model and Its Estimation Algorithm with Dynamic Inconsistent Expected Utility in an Uncertain Situation[J]. *Transportation Research Part C: Emerging Technologies*, 2024, 158: 104408.
- [60] VENATOR J. Dual-Earner Migration Decisions, Earnings, and Unemployment[R]. Working Paper 1052. Boston College Working Papers in Economics, 2022.
- [61] WEBER T, VAN MOL C. The Student Migration Transition: An Empirical Investigation into the Nexus between Development and International Student Migration[J]. *Comparative Migration Studies*, 2023, 11(1): 5.

- [62] WOZNIAK A. Are College Graduates More Responsive to Distant Labor Market Opportunities?[J]. Journal of Human Resources, 2010, 45(4): 944-970.
- [63] WYMAN M. Round-Trip to America: The Immigrants Return to Europe, 1880-1930 [M]. Cornell University Press, 1993.
- [64] ZHOU J, HUI E C M. Housing Prices, Migration, and Self-Selection of Migrants in China[J]. Habitat International, 2022, 119: 102479.
- [65] 安虎森, 颜银根, 朴银哲. 城市高房价和户籍制度: 促进或抑制城乡收入差距扩大?——中国劳动力流动和收入差距扩大悖论的一个解释[J]. 世界经济文汇, 2011(4): 41-54.
- [66] 蔡昉, 都阳, 王美艳. 户籍制度与劳动力市场保护[J]. 经济研究, 2001, 2001(12): 41-49, 91.
- [67] 陈映芳. “农民工”: 制度安排与身份认同[J]. 社会学研究, 2005, 2005(3): 119-132, 244.
- [68] 高波, 陈健, 邹琳华. 区域房价差异、劳动力流动与产业升级[J]. 经济研究, 2012, 47(1): 66-79.
- [69] 郭冬梅, 陈斌开, 吴楠. 城乡融合的收入和福利效应研究——基于要素配置的视角[J]. 管理世界, 2023, 39(11): 22-46.
- [70] 韩其恒, 苗二森, 李俊青. 农村劳动力迁移摩擦影响农民工数量与工资结构吗?[J]. 管理科学学报, 2018, 21(1): 13-30.
- [71] 洪大用, 范叶超, 李佩繁. 地位差异、适应性与绩效期待——空气污染诱致的居民迁出意向分异研究[J]. 社会学研究, 2016.
- [72] 黄忠华, 杜雪君. 农村土地制度安排是否阻碍农民工市民化: 托达罗模型拓展和义乌市实证分析[J]. 中国土地科学, 2014, 28(7): 31-38.
- [73] 黄宗晔, 杨静. 方言对省际人口迁移的影响[J]. 人口研究, 2020, 44(4): 89-101.
- [74] 蒋为, 倪诗程, 吉萍. 中国跨地区劳动力流动壁垒: 测度方法、演进趋势与决定因素[J]. 经济研究, 2024, 59(4): 60-77.
- [75] 李辉, 王良健. 房价、房价收入比与流动人口长期居留意愿——来自流动人口的微观证据[J]. 经济地理, 2019, 39(6): 86-96.

- [76] 李强. 影响中国城乡流动人口的推力与拉力因素分析[J]. 中国社会科学, 2003, 2003(1): 125-136, 207.
- [77] 李秦, 孟岭生. 方言、普通话与中国劳动力区域流动[J]. 经济学报, 2014, 1(4): 68-84.
- [78] 梁若冰, 汤韵. 地方公共品供给中的 Tiebout 模型: 基于中国城市房价的经验研究[J]. 世界经济, 2008(10): 71-83.
- [79] 林毅夫, 刘培林. 中国的经济发展战略与地区收入差距[J]. 经济研究, 2003(3): 19-25+89.
- [80] 刘晨晖, 陈长石. 劳动力流动、技能匹配与地区经济差距[J]. 经济研究, 2022, 57(7): 45-63.
- [81] 刘华, 王姣, 陈力朋. 人力资本空间配置的社会福利效应研究——基于量化空间一般均衡模型的分析[J]. 经济研究, 2024, 59(10): 37-55.
- [82] 刘修岩, 李松林. 房价、迁移摩擦与中国城市的规模分布——理论模型与结构式估计[J]. 经济研究, 2017, 52(7): 65-78.
- [83] 刘毓芸, 徐现祥, 肖泽凯. 劳动力跨方言流动的倒 U 型模式[J]. 经济研究, 2015, 50(10): 134-146, 162.
- [84] 陆益龙. 户口还起作用吗——户籍制度与社会分层和流动[J]. 中国社会科学, 2008, 2008(1): 149-162, 207-208.
- [85] 聂辉华, 李翹楚. 中国高房价的新政治经济学解释——以“政企合谋”为视角[J]. 教学与研究, 2013(1): 50-62.
- [86] 任远, 施闻. 农村外出劳动力回流迁移的影响因素和回流效应[J]. 人口研究, 2017, 41(2): 71-83.
- [87] 任远, 邬民乐. 城市流动人口的社会融合: 文献述评[J]. 人口研究, 2006, 2006(3): 87-94.
- [88] 石智雷, 杨云彦. 家庭禀赋、家庭决策与农村迁移劳动力回流[J]. 社会学研究, 2012, 27(3): 157-181, 245.
- [89] 孙伟增, 张晓楠, 郑思齐. 空气污染与劳动力的空间流动——基于流动人口就业选址行为的研究[J]. 经济研究, 2019, 54(11): 102-117.

- [90] 王丽莉. 土地供给、房价与劳动力空间配置效率[J]. 经济学 (季刊), 2023, 23(2): 500-516.
- [91] 王丽莉, 乔雪. 我国人口迁移成本、城市规模与生产率[J]. 经济学 (季刊), 2020, 19(1): 165-188.
- [92] 王丽莉, 唐为. 劳动力流动对城市工资与福利的影响——基于空间均衡模型的分析[J]. 南开经济研究, 2024, 2024(7): 64-83, 123.
- [93] 王丽娜. 基于压力门槛假说的家庭迁居意向形成机制研究 ——以杭州市区为例 [D]. 2. 浙江大学, 2007: 94.
- [94] 王智强, 刘超. 中国农村劳动力迁移影响因素研究——基于 Probit 模型的实证分析[J]. 当代经济科学, 2011, 33(1): 56-61, 126.
- [95] 夏怡然, 陆铭. 城市间的“孟母三迁”——公共服务影响劳动力流向的经验研究[J]. 管理世界, 2015(10): 78-90.
- [96] 熊彩云. 农民工城市定居转移决策因素的推—拉模型及实证分析[J]. 农业经济问题, 2007(3): 74-81, 111.
- [97] 杨曦. 城市规模与城镇化、农民工市民化的经济效应——基于城市生产率与宜居度差异的定量分析[J]. 经济学 (季刊), 2017(4): 1601-1620.
- [98] 郁建兴, 高翔. 地方发展型政府的行为逻辑及制度基础[J]. 中国社会科学, 2012, 2012(5): 95-112, 206-207.
- [99] 张莉, 何晶, 马润泓. 房价如何影响劳动力流动?[J]. 经济研究, 2017, 52(8): 155-170.
- [100] 赵方, 袁超文. 中国城市化发展——基于空间均衡模型的研究[J]. 经济学 (季刊), 2017(4): 1643-1668.
- [101] 钟水映, 李春香. 乡城人口流动的理论解释: 农村人口退出视角——托达罗模型的再修正[J]. 人口研究, 2015, 39(6): 13-21.
- [102] 周文, 赵方, 杨飞, 李鲁. 土地流转、户籍制度改革与中国城市化: 理论与模拟[J]. 经济研究, 2017, 52(6): 183-197.
- [103] 周颖刚, 蒙莉娜, 卢琪. 高房价挤出了谁?——基于中国流动人口的微观视角[J]. 经济研究, 2019, 54(9): 106-122.

- [104] 踪家峰, 周亮. 大城市支付了更高的工资吗?[J]. 经济学 (季刊), 2015, 14(4): 1467-1496.