

Thesis Draft

Suicidal Bumblebee

January 2025

- 证明工资分布和地区分布有直接关系而不是只和行业工资否则无法使用地区平均工资作为 μ_j
- 在原模型基础上引入了新的特点，例如流动壁垒（迁移摩擦）、地区异质性
- 要用 MLE，要对抗动项的条件概率分布进行假设
- 由于模型迭代极为复杂没有解析解，所以我们退而求其次求解析解
- 本文主线：传统的城乡二元对立模型难以满足研究多层次的劳动力的动态住选址选择问题，故而采用最有居住地选择模型
- 乡土文化中分出两种类型的人，一种是对离家近的地方有执念的，一种是没有的
- 给出城市的文化亲近度
- 用 samuelson 冰山理论应用了摩擦系数
- 在原先的纯空间距离上还加入了空间经济距离

1 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最优住址选择模型，更适用于当下城市化背景中城乡间隔逐渐消除的大背景，并检验（xxx 年）间我国各省的人口流动，并基于（xxx 数据）进行实证检验，得到的结论是（xxx、xxx、xxx）

2 引言

世界人口分布的特点 -> 我国收入的地理分布 -> 劳动力的流动符合直觉 -> 直觉中仍有劳动力回流这种反直觉的现象 -> 为了分析这点必须要先了解劳动力为何流动 -> 世界上的普遍研究做法 -> 我国的研究方法 -> 提出其中的不适合于缺陷 -> 引出我们的方法 -> 在原作者的基础上改进 -> 介绍论文各章节

世界人口分布和移动的共同点（简略）Krugman (1991); Fujita et al. (1999); Puga (1999) 的新经济地理学指出劳动力从“边缘”地区向商品多样化、工资高和就业机会多的“中心”地区集聚

经济活动集中在某个地区时，能够产生规模经济和网络效应，这导致了产业集聚。尤其是在制造业、服务业等领域，企业集聚在一起可以获得更多的商业机会，减少运输成本，并从市场规模中获益。

新经济地理学强调地区之间的竞争和合作机制，导致区域内的资源和生产活动会相互强化。例如，Krugman 通过模型证明，虽然初期经济活动可能会在某些地区零散分布，但由于市场和生产的反馈机制，某些地区会因为集聚效应逐渐吸引更多的资源，形成经济集中，最终导致区域的不平衡发展。

根据新经济地理学的理论，劳动力的迁移并不是随机的。人们往往会选择进入那些产业集聚、机会多、工资水平较高的地区。因此，地区的经济集聚会吸引更多的劳动力流入，形成所谓的“吸引效应”。

随着劳动力和产业的集中，特定地区会形成自我强化的迁移机制。例如，企业聚集会促进更多的就业机会，而就业机会的增多又会吸引更多的劳动力流入，这一过程可能形成区域性“良性循环”。

通过分析劳动力迁移的影响，新经济地理学表明，劳动者在选择迁移时会受到工资差异、就业机会、生活成本等多重因素的影响。尤其是在低成本地区和高薪地区之间，劳动力的流动往往导致区域收入差距的加大。

我国收入分布和地理分布之间的关系东富西穷南富北穷【表格汇总居民可支配收入】相邻的城市形成产业聚集相邻的城市形成城市圈自改开以来所有城市的人均可观增长（尽管有资源耗尽城市的例子）

我国人口分布和劳动力移动的特殊点阐述我国劳动力流动的特点劳动力净流入符合常识中的劳动力迁移规律即我国劳动力流动的巨大特点之一就是由农村流向城市是绝对的主流但是依旧有回流（李芝倩 2007）xxx 通过数据举例

但富的城市仍有虹吸效应包括财富虹吸和人口虹吸按照 Krugman 1991 的观点这种贫富差距并不会收敛这种大城市与小县城之间长久的对立（城乡之间的对立）是我国长久持续且深刻的工业化带来的必然结果我们可以断言劳动力还会持续地、有规律地进行流动（引用适当文

献，例如 Lucas 2004 指出即使城市存在大量失业，当劳动力仍会大量前往...)

为了理解劳动力迁移为什么会有这样的规律我们需要了解迁移决策背后的原因从古至今全球各地的 people 总是在迁移古时候迁移的原因多种多样但自从进入了定居文明社会之后人口的迁移规律必然受到了经济因素的制约这其中不外乎天灾政策按照 xxx 的观点迁移成本的下降必然是劳动力迁移现象上升的决定性因素之一进入大航海时代之后人类的交通能力更是得到质的飞跃迁移的成本变得更加简单殖民就是这一时刻最形象的劳动力迁移对于我国而言自从 xx 时代 xx 政策后劳动市场自由流动人们可以选择定居的地方简直目不暇接制约劳动力迁移的客观因素逐渐被移除那么剩下的就是主观因素了需要说明的是本文关注的仅仅是劳动力的迁移而不是人口的迁移因为导致后者的原因是极为复杂的甚至可能有超出经济学范畴的因素更具体的说本文关注的是受到收入影响的劳动力迁移 (income-induced migration) 经济学中对于劳动力迁移的研究已经有丰富的成果传统经济学基于功利主义认为影响劳动力迁移的最主要因素是预期收入的变化(引用适当文献)人口迁移往往并不意味着经济的正增长在历史中从台湾岛出发的南岛语系人群因为所居住的岛屿能提供的资源与人口增加之间的压力面临马尔萨斯人口陷阱而前往其他岛屿这与预期收入的变化导致迁移的规律相吻合但他们冒然的前往往往并不一定能带来技术的传播或经济的发展在工业化快速崛起传统农业文化分崩离析之际人口的迁移必然会变得更自由(引用新法律新法规和文献)但是在很大程度上这种迁移的模式仍然是未知的因此预测人口移动是必要的同时在人才战争中具有政策意义

劳动力主观选择定居的过程主要受到以下因素影响孟母三迁为代表的 amenities 工资差为代表的 wage income xxx 因素

对于以上因素世界上各类经济学方法百花齐放 xxx 历史 xxx 举例 xxx 优劣

但相较于外国各类方法齐出的繁荣景象我国由于历史原因依旧采用比较原始的方法我国传统劳动力流动分析依旧大量采用了城乡对立的基本思想 xxx 举例即(解释传统的 Todaro 二元对立到)

但是这背后隐藏的方法危机已经悄然显现 1 多次选择(先落地城市 A 再搬去城市 B) 2 劳动力的流动中往往存在着可逆的迁移选择(去大城市打工最后又回到小县城不管是出于各种原因)这实际上也是一种将多次选择问题石智雷 2012 劳动力回流问题任远 2017 但是在关注回流问题的时候，同时也有大量进城人们也必须先进城才能再回流传统的二元对立模型就此陷入了困境学者们无体系地应用计量经济学来解决问题但是治标不治本劳动力迁移问题需要一个可撤回的基础模型 3 当个体对城市与城市之间出现有不同的偏好时且他们都有异质性偏好就会多种目的地的选择(例如浙江人选择去杭州还是上海或者是北京)劳动力在前往城市时要理性地思考去哪个城市而不是机械地蜂拥而至劳动力不仅在城市与乡村之间流动也在城市与城市之间流动低端城市的劳动力会前往更高端的城市那么一个城市也能成为传统二元对立语境下的农村这时候就会陷入城市之间的华夷之辨即使中国有特色的户籍制度再善于分辨的人恐怕也无法厘清这之间的关系接下来的问题就是如果城乡之间已经

但在当今人口快速流动的现代化社会中传统的城乡对立劳动力流动模型已经难以胜任研究多变的劳动力迁移现象

首先需要将劳动力的迁移行为视作一个在连续时间上的选择行为序列看待 xxx 引用例证同时由于对于个体的迁移行为之后的效用提升可能存在逆向因果效应和自选择效应问题例如我们不知道是因为迁移导致的工资提升还是其他因素为了克服这种情况需要采取一个在 horizon 里每期都进行迁移选择的效用最大化模型来分析从而控制各种 causality problems 正因为如此本文在 Kennan and Walker 2010 的动态最优搜寻工作模型基础上引入了流动摩擦、居民异质性、乡土文化来分析劳动力迁移作为经济现象背后的原因相较于原作者的种种缺点本文通过采取了几种不同的 xxx 方法优化了原作者的局限性

本文第二章阐述了一个理论模型第三部分对于实证模型中需要的似然函数给出了详细推导并对变量进行必要的简单化假设第四部分给出了实证的结论并通过对样本的截取形成了不同的研究命题的结论第五部分提出文章可以改进的地方和未来可以应用的方向更具体的证明、代码放在文末的附录中

3 文献综述

3.1 模型

对于劳动力流动有多种解释按照 Jia et al. 2023 的分类可以将其划分一种是空间均衡模型，另一种是最优居住地选择模型。

3.1.1 空间均衡模型

中国是典型的城乡二元经济体，改革开放后工业化进程的快速推进，增加了对劳动要素的需求，农村劳动力流动的限制也逐渐得到放松。大量农村劳动力流入城市，对推动中国的工业化、城市化和经济发展做出了重大贡献（李扬和殷剑锋，2005）。

我国在有大量的基于以 Todaro 模型为代表的城乡二元对立模型进行讨论的论文

Lewis 1954, Ranis and Fei 1961 首先提出了发展经济学的二元对立模型，用于解释发展中国家在发展中城市部门吸收农村人口实现现代化的过程，他们的结论是现代经济部门工资高于传统农业部门，劳动力将从农业向工业部门转移。

Todaro 1969 建立了一个城乡二元对立的模型，其中劳动力受到预期城乡收入差距前往现代工业部门工作，即使这会导致失业率的上升从而下降自身的预期收益。他的做法有多种优点，例如加入了一次性的迁移成本、将失业率代入了模型。Todaro 指出即使面临失业风险，人口也会因为预期城乡收入差距而迁移。Harris and Todaro 1970 在其原先基础上加入了 (xxx) 对

于我国的劳动力迁移研究而言，Todaro 最重要的假设是其假设现代工业部门规模经济，而农业部门 CRTS，这代表了城乡二元对立的思想，适合我国发展中国家的国情。

Rosen, Roback 1982 国内也开始慢慢有 Rosen Roback 模型作为基础的论文发表王丽莉 2020、2023

以上为模型及其衍生在本质上属于空间均衡模型，尽管广泛被使用，但是他们将劳动力的迁移视作独立的事件，无法描绘出可逆的迁移选择

3.1.2 最优居住地选择模型

Tiebout 1956 提出用脚投票，第一次用模型将 amenities 这个概念引入对劳动人口的流动进行考察

从 Sjaastad 1962 开始，开始重视劳动者本身的主观意愿 *Sjaastad's (1962) treatment of migration as an investment emphasizes the dynamic aspect of migration – expected costs and payoffs to migration change over time. Viewed within a life cycle perspective, individuals (or families) decide whether and when to move. Allowing households to make multiple migration decisions substantially increases the model's complexity. Decisions made in previous periods (e.g., savings, education, marriage, fertility) determine choices available in the current period, and expectations of future events also influence current decisions. Within this perspective migration and fertility choices are connected through the “primitives” of the decision making process: current opportunities (determined in part by decisions made in the past), expectations (anticipated future events and outcomes), and preferences (values assigned to different outcomes). The economic perspective thus provides a unified framework connecting several important demographic behaviors.*

Everett S. Lee 1966, Bagne 1969 推出了推拉模型，系统阐述推拉理论，指出流入地的有利生活条件为拉力，流出地的不利生活条件为推力，劳动力流动受这两股力量的影响。可以将其视作计量经济学线形多元回归，通过添加多种推拉因素作为控制变量，消除

从 (xx 年代) 开始，经济学家们将效用最大化融入了，由于效用不可测量，最好的替代品就是消费（类似于 GDP 的概念，效用不可测量，但带来效用的消费可测量），如果引入理性的概念居民必定会在既定收入的限制下实现收入越高、效用越大的效用最大化，那么测量效用的替代品就转而成为了收入。这即是收入成为劳动力迁移研究中决定性因素的历史脉络。

DaVanzo (1983), who documented the richness of individual migration histories, pointing out that although most individuals never move, those who do are likely to move again, often returning to a home location. This means that migration decisions should be viewed as a sequence of location choices, where the individual knows that there will be opportunities to modify or reverse moves that do not work out well.

Kennan-Walker 糅合了多年来劳动力迁移经济学的发展，并凝结在了 KW 模型中。

3.1.3 基于匹配理论

彭国华 2015 刘晨晖 2022 将匹配理论与劳动力迁移结合

3.1.4 劳动力回流问题

3.2 对于影响劳动力流动的因素

3.2.1 劳动力壁垒

3.2.2 居住成本与迁移成本

Helpman (1998) 在 Krugman 1991 的基础上引入住房市场因素，指出房价会影响劳动者相对效用，从而抑制劳动力在高房价地区的集聚；还提出经济集聚导致的劳动力涌入也会推高房价 Brakman et al. (2002); Saks (2004); Rabe & Taylor (2012) 实证检验了 Helpman 的理论。Hanson et al. (1999, 2005) 实证分析支持 Helpman (1998) 的结论。

Dohmen (2005); Meen & Nygaard (2010) 指出尽管高房价地区会抑制劳动力流入，但预期套利机会会促使劳动力流入。

张莉 2017 经济研究

周颖刚 2019 经济研究根据 Roback(1982) 和 Diamond(2016) , 一方面劳动力从低工资地区向高工资地区流动, 以提高劳动力家庭的效用水平; 另一方面, 房价是劳动力在城市居住的主要成本, 高房价水平降低了劳动力在居住城市的效用水平

Diamond 2016

Todaro 在模型中加入了一次性的迁移成本，

3.2.3 人力资本

Moretti (2004); Fu and Liao (2012) 指出人口密度高的地方有利于技能匹配和获得学习机会，劳动力更倾向于流向教育水平高、人口密度强的城市，这实际也是侧面证明了预期收入

刘毓芸等 (2015): 基于文化经济学理论，指出方言距离较小时促进劳动力流动，反之则阻碍劳动力流动。

Schultz (1961), who considered migration as a form of investment in human capital

3.2.4 舒适度、公共服务与

Tiebout 1956 夏怡然 2015，城市间的“孟母三迁”——公共服务影响劳动力流向的经验研究

在社会学的研究中，舒适度这个概念是人口迁移的主要因素，这一概念涵盖了空气质量、公共服务、城市天气等多种变量。以王丽娜 2007 为例

孙伟增 2019 基于 Rosen-Roback 模型得出空气污染对于流动人口的就业选址具有显著的负向影响

Dorfman 2016 *This paper extends the literature on amenity migration by focusing on healthcare access for later-life migrants.*

3.2.5 具有中国特色的因素：户籍、乡土观念等

李强 2003 认为由于户籍制度，中国的人口流动将不再遵循一般的推拉规律陆益龙 2008 用 OLS 分析户籍的社会流动影响

乡土观念文化特征是否能融入

同学聚集亲戚介绍离得近

3.2.6 其他

信息的沟通过程想象在古代大概只有自然灾害、战争等一些有限的因素导致能导致劳动力有规律迁移但在当下同村人之间相互对远方陌生城市中的高工资的推崇就会导致劳动力的迁移如今东南亚的一些电信诈骗产业往往通过宣传虚假的高工资

4 理论模型

本文借鉴 Kennan and Walker 2011，构建包含 xxx 的最优住址模型。与其不同的是，首先，本文详细刻画 xxx；其次，本文在模型中引入迁移摩擦，从而更加符合中国人口无法自由流动的现实制度背景。

4.1 假设

和所有的劳动力模型一样这个模型 imposes some strong assumptions 如下

- 劳动力自由流动
- 本文的重要假设是工资是个人技能集的本地价格 $w^D = w^S = w^S(\sum_i Skill_i) \Rightarrow w^*$
- 资产不影响居民的选择（成吗?）
- 个人之间不相互影响（iid 抽样）

4.2 初步模型

劳动力的流动是一个极其复杂的问题经过以上的分析劳动力的流动和迁移问题不应该被视作一次性的经济活动相反这应该是在个人的 **life span** 里不断进行考量的问题同时劳动力迁移之后得到的境遇提升可能是因为未观测到的个人特质难过必须要控制各种交错变量中的数量经济学问题为了完成 **xx** 任务必须构建一个动态劳动力迁移模型 **The intuition is like this** 动态迁移模型的基本思想是个人可以自由选择存在多个基于不同位置而不同的 (相互排斥的) 收益流 **payoff flow** 居民可依靠自己的偏好等个人信息选择在这些备选方案之间切换但在迁移时需支付迁移成本在任何给定时间, 个人都处于特定位置, 必须选择是留在那里还是去其他地方, 将当前收益流换成其中一个备选方案。当前位置的未收益通常不确定, 备选收益更是如此。个人的行动应使已实现收益的预期现值 (扣除迁移成本) 最大化

在任意地点 j 使用状态变量 $x = (\ell, \omega, a)$ 表达去过的地区序列 ℓ 、地区包含的工资 w 、地区包含的偏好信息、年龄 a 用 v 表示稳定效用用 ζ 表示随机效用那么居民的基本问题是在每个给定状态 x 和随机项 ζ_j 的情况下, 通过选择 j 来实现使得个体在生命周期中的总效用最大化本质上反映了一个动态离散选择问题 (**Dynamic Discrete Choice Problem**) 即

$$V(x, \zeta) = \max_j [v(x, j) + \zeta_j] \quad (4.1)$$

其中

- $v(x, j) = u(x, j) + \beta \sum_{x'} p(x'|x, j) \bar{v}(x')$
- $\bar{v}(x) = E_{\zeta} V(x, \zeta)$
- β is the discount factor
- E_{ζ} denotes the expectation with respect to the distribution of the J -vector with components ζ_j
- $p(x'|x, j)$ 是状态转移概率

使得在状态 x 下选择地点 j 的概率为 $\rho(x, j) = \exp(\bar{\gamma} + v(x, j) - \bar{v}(x))$, 之中的 $\bar{\gamma}$ 为欧拉常数

其中基础效用 $v(x, j) = u(x, j) + \beta \sum_{x'} p(x'|x, j) \bar{v}(x')$ 分为两部分一是选择 j 带来的直接效用二是折现的通过转移概率加权得到的未来期望效用, $\bar{v}(x) = E_{\zeta} V(x, \zeta)$ 表明在状态 x 下去掉随机项影响得到的期望效用

公式中的随机效用项假设服从于一类极值分布 We assume that ζ_j is drawn from the Type I extreme value distribution. In this case, using arguments due to McFadden (1973) and Rust (1987), we have

$$\exp(\bar{v}(x)) = \sum_{k=1}^J \exp(v(x, k))$$

这表示如果变量服从一类极值分布，那么 $\exp(\bar{v}(x))$ 可以表示为所有 $v(x, k)$ 的指数和它意味着，在状态 x 下，选择某一选项 j 的概率与效用的指数值成比例。这个性质广泛用于预测个体选择的分布。

那么以上的地点选择概率也就是 $\rho(x, j) = \frac{\exp(v(x, j))}{\sum_{k=1}^J \exp(v(x, k))}$ (需要证明吗?)

假设收入的边际效用是固定的，居民可以在固定利率下自由借贷，那么居民的期望效用最大化问题就可以转化为预期永久收入，除去迁移成本（需要证明吗?）

居民在地点为 h 的效用流等于基础效用 + 随机效用 $\tilde{u}_h(x, j) = u_h(x, j) + \zeta_j$

在 h 地点的居民的基础效用组成部分如下 $u_h(x, j) = w_{ij}(a) + \sum \alpha_k Y_k(\ell^0) + \sum h + \xi(\ell^0, \omega) - \text{MovingCost}$ 下面逐步将其拆分讲解

4.3 工资性收入（货币收入）

其中工资收入的组成是

$$w_{ij}(a) = \mu_j + \nu_{ij} + G(X_i, a, t) + \eta_i + \varepsilon_{ij}(a)$$

地区平均收入 μ_j 个人与地区的匹配效应 ν_{ij} 时间效应 $G(X_i, a, t)$, where X_i is characteristics, a is age and t is a time effect 个体固定效应 η_i 暂态效应 $\varepsilon_{ij}(a)$ 地区固定效应 ρ_j 暂态效应 ε 的实现会影响当前时期的收入，但对任何地方的未来工资增长都没有影响，因此对移民决策没有影响。另一方面，单个效应 η_i 是永久性的，位置匹配效应 ν_{ij} 对于位置 j 是永久性的，因此这两个组成部分都会影响迁移决策，因此必须将其视为状态变量。

假设 η, ν, ε 都是 iid 分布于各个个体和地区同时前面两个的实现能被居民观测到

地区工资之间的差异可能会抹平其他因素的差距例如现实生活中很常见的就是某些工资高工资但具备极差的生活条件的生活方式但是具备异质性的劳动者仍然会转移到具备例如刘晨晖 2022 > 研究发现：技能匹配度是劳动力流动所形成的地区经济差距是否稳定的关键因素；当不考虑高技能劳动者时，劳动力流动能够推动地区经济差距收敛，但考虑高技能劳动者集聚效应与技术溢出后，劳动力流动不再必然推动地区经济差距收敛；迁移成本在不同情形下对地区经济差距的影响是异质性的：当不考虑高技能劳动者时，迁移成本是区域均衡发展的阻力；而考虑高技能劳动者之后，集聚效应、技术溢出以及技能匹配均可能成为地区经济差距扩大的原因，迁移成本有时反而会起到缩小地区经济差距的作用。

年龄 a 是讨论个体迁移问题中很重要的一环年龄与个人的工资直接挂钩，例如 $G(a, X_i, t)$ ，它代表了收入年龄分布，本质上是时间效应年轻人缺少家庭因素更容易定居在外地年轻人对于未来容易抱有较高幻想，即倾向于认为自己有更高的预期收入 *A standard human capital explanation for this age effect is that migration is an investment: if a higher income stream is available elsewhere, then the sooner a move is made, the sooner the gain is realized. Moreover, since the worklife*

is finite, a move that is worthwhile for a young worker might not be worthwhile for an older worker, since there is less time for the higher income stream to offset the moving cost (Sjaastad

1962

). In other words, migrants are more likely to be young for the same reason that students are more likely to be young. KW 模型原来适用于年轻劳动力的迁移 this paper focuses on the relationship between income prospects and migration decisions at the start of the life cycle

ν_{ij} is the location match effect This means that a worker might earn more or less in one location compared to another, even if their skills and experience are the same This effect is considered permanent for the duration of the time an individual works in a location. Once an individual experiences a location match effect in a particular location, it does not change unless the worker moves to a different location. The location match effect is treated as a random variable¹². This means that for each individual-location pairing, there is a draw from a distribution of possible match effects. Individuals only learn the realization of this match effect after they visit the location

个体固定效应 η_i 个体不随着地区变化的工资收入

ρ_j == is the location effect, 这会捕捉一些受地区影响的经济变量 == - 某些产业在特定地区呈现区域产业集聚 (Regional Economic Clustering), 例如北京相较于其他城市有极强的新闻媒体产业, 山西某些城市有丰厚的煤矿资源带来的矿物资源产业繁荣。- 特定城市会受到政策照顾, 使得所在地区的劳动力税收减免 - 地区通过人才政策进行额外补贴等 - 特定的政策效应多地采用了人才补贴的做法, 以杭州市为例。。此种做法即地方的政策性福利

4.4 宜居度

== 舒适度由各种变量组成 == 对于住房成本而言分成短期成本 (例如租房成本), 长期成本 (例如住房成本和维护成本) 要依据当前住房成本和当地区平均住房成本之间的差值进行舒适度的调整增减 $\sum_k s_{ij}^k + \Delta_r(r_j) + \Delta_h(h_j)$, where $\sum_k s_{ij}^k$ is the total amenity value; Δ_r is a positive function, r_j is the weighted rent cost in city j ; and Δ_h is a negative function, h_j is the house price index in city j . 这会将租房和住房市场同时纳入模型中考察其对劳动力流入与流出的影响

4.5 乡土文化

== 家乡溢价 == $\sum_i \alpha_i^H$ 个体可能出于恋家、乡土文化等各种原因更偏好家乡, 从而宁愿降低一些收入而选择家乡, 这种溢价用参数 α^H 捕捉

同时如果不同地区具有相似的文化特征, 这种文化相似性有助于减少劳动力迁移中的适应成本, 从而使得劳动力的迁移决策变得更加容易; 同时, 这种相似性往往呈现出地域上的相近,

这种文化的相似性可以用参数捕捉刘毓芸 2015 提出劳动力倾向于在方言区内进行流动这种观点实则属于文化特征的一部分是否能融入是劳动力是否回流问题中非常重要的因素这与文化特征相关 (reference some 回流文章其中特别喜欢讲劳动力能否融入)

包含家乡这一因素是极具理论进步性的一步对于劳动力回流的研究可以从这个角度突破

4.6 迁移成本

劳动力迁移成本

$$\Delta_{\tau}(x, j) = [\gamma_{0\tau} + \gamma_1 \cdot D(\ell^0, j) - \gamma_2 \cdot \chi(j \in A(\ell^0)) - \gamma_3 \cdot \chi(j = \ell^1) + \gamma_4 \cdot \alpha - \gamma_5 \cdot n_j] \chi(j \neq \ell^0)$$

First term $\gamma_{0\tau}$ is an intercept denoting unobserved heterogeneity in moving cost

Second term $\gamma_1 \cdot D(\ell^0, j)$ means that moving cost is an affine function of distance, where $D(\ell^0, j)$ is the distance between locations ℓ^0 and j

Third term $\gamma_2 \cdot \chi(j \in A(\ell^0))$ means moving to an adjacent location may be less costly, as $\chi_{A(l)}(j)$ is an indicator function equal to 1 if j is adjacent to ℓ , and 0 otherwise

Fourth term $\gamma_3 \cdot \chi(j = \ell^1)$ means a move to a previous location may also be less costly, relative to moving to a new location.

Fifth term $\gamma_4 \cdot \alpha$ means the cost of moving is allowed to depend on age, a

Sixth term $\gamma_5 \cdot n_j$ means that it is possible to be cheaper to move to a large location, as measured by population size n_j

讨论劳动力流动壁垒问题蒋为 2024 程名望 2007 王丽莉 2020、2023 提出了迁移摩擦这一概念, 可以将迁移成本认为是 (引用) 进一步降低人口流动壁垒将有利于我国城市规模的扩张与劳动力资源配置效率的改进。所以可以采用 Samuelson 的 iceberg 理论流动壁垒可以被认为是一次性付清的迁移成本令 $MovingCost = \frac{SomeMoney}{CertainCoefficient}$

4.7 状态转移概率

先前我们使用了状态变量 x 囊括了 M 维最近地区向量 ℓ , M 维地区工资与地区效用信息 ω 和年龄 a , 此状态变量有改变的可能, The state x' is the new state after a move. 用以下式子来描

述 The transition probability is

$$p(x'|x, j) = \begin{cases} 1, & \text{if } j = \ell^0, \tilde{x}' = \tilde{x}, a' = a + 1 \\ 1, & \text{if } j = \ell^1, \tilde{x} = (\ell^0, \ell^1, x_\nu^0, x_\nu^1, x_\xi^0, x_\xi^1), a' = a + 1 \\ \frac{1}{n^2}, & \text{if } j \notin \{\ell^0, \ell^1\} \tilde{x}' = (j, \ell^1, s_\nu, x_\nu^0, s_\xi, x_\xi^0), 1 \leq s_\nu \leq n_\nu, 1 \leq s_\xi \leq n_\xi, a' = a + 1 \\ 0, & \text{otherwise} \end{cases} \quad (4.2)$$

This probability describes the likelihood of moving from one state x to a new state x' given that a specific location j has been chosen. It's a conditional probability, and it focuses on the mechanics of how the state changes based on a location decision. It encapsulates both deterministic changes (like an increase in age) and random changes (like the draw of a new location match component).

4.8 劳动力迁移决策

移动的 motivation: 基于对于工资的假设不同地区之间 j 和 k 之间只有平均工资 μ 、地区匹配效应 ϵ 和地区固定效应 η 是随着地区之间变化的不同地区的选择就会对收入产生影响劳动力会因为更好的劳动力市场更好的地区匹配更好的政策福利而选择迁移

Probability that a person in state x will choose location j can then be written as

$$\rho(x, j) = \exp[v(x, j) - \bar{v}(x)]$$

方程 $v(x, j) = u(x, j) + \beta \sum_{x'} p(x'|x, j) \bar{v}(x')$ 可以解释为我们通过值函数迭代计算 v , 假设有限时间 T 。我们将年龄作为状态变量, 在年龄 $T + 1$ 时 $v = 0$, 这样连续迭代就会产生一个越来越年轻的人的价值函数。

至此, 模型刻画出了在 finite horizon 内, 个体可以在不同城市 $j \in J$ 中自由选择最优住址, 这意味着迁移可以是多次的, 可撤回的该模型存在一定缺陷例如在模型中的第 1 期往往受到未被数据记录从而未被模型捕捉到的第 0 期及更久之前的影响从而导致有 sample selection 的嫌疑这点对于年岁较大的人更甚

5 实证检验

5.1 实证模型

5.1.1 年龄效应

出于简化参照以往的计量经济学经验将年龄效应设置为年龄的有待估参数的一次项和二次项的和

5.1.2 location match effect

由于该模型存在大量的 0 所以出于计算上的简便我们使用离散估计替代一些变量
通过离散估计法估计地区匹配效应的分布

选择一些 ** 支持点 (support points) **, 将原本连续的分布 F 用离散分布 \hat{F} 代替。支持点被选为分布的某些分位点 (比如中位数或四分位点), 并将每个分位点分配一个权重 (概率)。对于每个 q_r (分位点位置), 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。

假设迁移到某地会带来不同的匹配效应 (比如迁移到某城市带来的工资或福利变化), 这些匹配效应具有随机性且可以用分布来表示。但在现实中, 不可能完整地计算整个分布下的期望值, 因此用离散点近似表示。本质上是一种简化模型, 减少计算成本, 同时保留分布的主要信息。

5.1.3 固定效应

** 数学上 **: η 表示每个个体工资的固定成分 (例如长期能力或个体特质)。假设其分布是一个对称的离散分布, 具有 7 个支持点。每个支持点有相等的权重。因为是离散分布, 只需要估计 3 个参数: 分布的中间点 (中心趋势) 和离散范围。 ** 经济意义 **: 固定效应代表长期的工资差异来源, 例如教育水平、经验等。选择离散分布是为了简化复杂的连续分布, 同时通过支持点捕捉关键的分布特性 (如中心趋势和差异性)。

η 的离散分布提供了一种对长期工资特质的简化建模方式, 减少计算复杂度, 同时保留重要信息。

5.1.4 短期波动

数学上 ε 表示工资的短期波动成分, 比如由于经济周期、随机冲击等导致的变化假设 ε_{it} (某人在某个时间的短期波动) 来源于一个均值为零的正态分布, 其方差 σ_ε 随人而异对于每个人, $\sigma_\varepsilon(i)$ 来自一个离散分布 (4 个支持点, 均匀分布), 因此需要估计 4 个参数 ε 反映工资中不可预测的临时变化, 例如经济不确定性或工作绩效波动允许 $\sigma_\varepsilon(i)$ 因人而异捕捉到工资短期波动的个体差异性 (例如高风险行业可能有更大波动)

对 ε 的建模允许方差随个体变化, 进一步增加模型的灵活性, 使其能够更好地拟合数据。

5.1.5 工资的影响

由于个体效应固定、暂态效应无影响影响工资差距的就是地点平均工资 μ 、地点匹配工资效应 ν 、地方福利 ρ

5.1.6 似然函数

**** 未观测变量 ****

在每个地点中个体都有从地区匹配效应的分布中抽取一个值，假设这个分布是一个有限集合上的均匀分布， $Y = \{\nu(1), \nu(2) \dots \nu(n_\nu)\}$ 其结果是 ω_ν^i ， $\omega_\nu^i(j)$ 代表在地点 j 的匹配效应， $1 \leq j \leq N_i$ ， N_i 是 i 到达过的地方的数量总共有 $\{\omega_\nu^i(1), \omega_\nu^i(2) \dots \omega_\nu^i(j) \dots \omega_\nu^i(N_i)\}$

地区匹配偏好也是如此来自于 $\Xi = \{\xi(1), \xi(2) \dots \xi(n_\xi)\}$ 其结果是 ω_ξ^i

固定效应也是如此来自于 $H = \{\eta(1), \eta(2) \dots \eta(n_\eta)\}$ 其结果是 ω_η^i

暂态效应也同样来自来自 $\varsigma = \{\sigma_\epsilon(1), \sigma_\epsilon(2) \dots \sigma_\epsilon(n_\epsilon)\}$ 其结果是 ω_ϵ^i

未观测到的因素组成一个 $N_i + 3$ 维的参数向量 $\omega^i = (\omega_\xi^i, \omega_\eta^i, \omega_\epsilon^i, \omega_\nu^i(1), \omega_\nu^i(2) \dots \omega_\nu^i(N_i))$

参数向量的可能实现组成一个集合，用 $\Omega(N_i)$ 表示

**** 似然贡献 ****

$\mathcal{K}_{it} = (\mathcal{K}_{it}^0, \mathcal{K}_{it}^1)$ 表示当前位置和上一个位置

个体 i 在 t 时期选择目的地的似然概率是 $\lambda_{it}(\omega^i, \theta_\tau)$ 同时在 $\lambda_{it}(\omega^i, \theta_\tau) = \rho(x, j) = \rho_{h(i)}(\ell(i, t), \omega_\nu^i(\mathcal{K}_{it}^0), \omega_\nu^i(\mathcal{K}_{it}^1), 1), \theta_\tau)$ 公式表明个体的选择不仅受到过去历史（如上一个地点和上一个选择）和随机效应（如匹配效应和暂态效应）的影响，还考虑了个体的长期偏好和未来选择的影响。通过对这个似然函数进行建模，我们可以估计个体在特定情况下做出选择的概率，从而理解个体在复杂环境下如何做出决策。

由于暂态效应服从期望为 0 的正态分布 $\varepsilon_{ij}(a) = w_{ij}(a) - \mu_j - \nu_{ij} - G(X_i, a, t) - \eta_i$ 也服从该分布从而令 Ψ 表示标准正态分布的概率累积函数那么收入函数的密度函数，

$$\Psi_{it}(\omega^i, \theta) = \phi\left(\frac{w_{it} - \mu_{\ell^0(i,t)} - G(X_i, a_{it}, \theta) - \nu(\omega_\nu^i(\mathcal{K}_{it}^0)) - \eta(\omega_\eta^i)}{\sigma_\epsilon(\omega_\epsilon^i)}\right)$$

$\mu_{\ell^0(i,t)}$ ：是个体 i 在位置 ℓ 基本工资水平 $G(X_i, a_{it}, \theta)$ ：代表个体 i 在时间 t 受到的外部影响（如经济环境、工作特征等 $\nu(\omega_\nu^i(\mathcal{K}_{it}^0))$ 、 $\eta(\omega_\eta^i)$ ：分别是地区匹配效应和固定效应，它们是收入的随机组成部分，影响个体的收入 $\sigma_\epsilon(\omega_\epsilon^i)$ ：表示暂态效应的标准差，它反映了短期收入波动的程度这是在时期 t 给定个体 i 的选择和随机效应（如匹配效应、固定效应等），观察到收入 w_{it} 的概率概率，即被观测收入的似然概率。

由此通过个体在其历史轨迹中给出的两种似然贡献我们得到了类型为 τ 的个体 i 的似然函数即在所有可能的不可观测变量的组合下，个体在所有时期中的观测收入 w_{it} 上选择地点的概率 ρ_{it} 的乘积：

$$L_i(\theta_\tau) = \frac{1}{n_\nu n_\epsilon n_\xi (n_\nu)^{N_i}} \sum_{\omega^i \in \Omega(N_i)} \left(\prod_{i=1}^{T_i} \psi_{it}() \lambda_{it}() \right)$$

由于模型允许存在异质性让 $L_i(\theta_\tau)$ 代表 τ 类型个体的似然函数，其中 θ 是该个体的待估参数

向量样本的似然函数是一个混合类型的联合对数似然函数，把每个观测 i 的贡献相加

$$\Lambda(\theta) = \sum_{i=1}^N \log\left(\sum_{\tau=1}^K \pi_{\tau} L_i(\theta_{\tau})\right) \quad (5.1)$$

其中混合比例由 π_{τ} 给出，且 $\sum_{\tau=1}^K \pi_{\tau} = 1$ 每个个体 i 都做出了贡献这是在给定参数 θ_{τ} 的条件下，个体 i 在类型 τ 下的似然。即，个体 i 可能属于某种类型，似然函数 $L_i(\theta_{\tau})$ 捕捉了该个体数据与类型。混合似然 $\sum_{\tau=1}^K \pi_{\tau} L_i(\theta_{\tau})$ 这表示对所有类型的加权平均，权重是 π_{τ} ，即每种类型的概率。通过这种加权求和，模型允许每个个体属于不同的类型，并通过类型的权重（概率）来加权它们的贡献。每个个体可能属于不同的类型，这些类型有不同的收入和行为模式。每个数据点 i 来自某个成分，但成分的归属是未知的，因此需要将每个数据点的似然表示为各成分似然的加权和，然后取对数并求和得到整个样本的对数似然。因此通过混合模型，能够捕捉到个体之间的异质性。

如此使用混合似然估计通过计算就可以得到各种参数的似然估计

5.2 数据上

地区数据的界限由于劳动力的迁移在不同省之间呈现趋势但在迁入大省中仍然存在人口净流出市这说明劳动力移动的法律边界和现实边界存在重合但是最好应该以市为标准当然这一准则不包括某些政策带来的效应上面说了为什么界限要往下卡在市，下面说一下为什么界限不继续往下到区或者村。首先是政策往往以市为最小执行单位；其次是在市内迁移中个人会为了出于固定资产投资等非模型考量的原因，这些原因并不是因为在迁居的地区能提供更好的个人预期工资市级包含了县、乡、村等行政级别每个市都包含了城市与农村（虽然城市化率略有不同）避免了城乡之间的划分对立

人口数据的界限许召元 2007

在我国存在广泛的农民工进城现象大多数农民工只是暂时居住在

蒋琪 2018 劳动经济研究：<u> 本研究使用 CFPS2010 年和 2014 年两期数据构成面板进行影响评估。CFPS 采用的是内隐分层、多阶段、多层次、与人口规模成比例的概率抽样方式，样本覆盖除中国港澳台、新疆、西藏、青海、内蒙古、宁夏和海南之外的其他省/市/自治区。这些地区的人口占全国的 95<u> 关键结果变量为个人的年总收入 (PIncome)。本研究选取的关于个人总收入的变量为“qk601”(2010 年) 和“p_iincome”(2014)

“PIncome” CFPS “K :
” “ () ?” CFPS (2011; 2017; 2017; 2017) CFPS
/u>

基于使用了 CFPS2010 到 2022 年间的数据 CFPS 数据记录了个人的迁移轨迹（将访问当年的地址视作居住地址）年龄收入

气候数据来自国家气象局

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

地方教育经费、地方医疗经费来自财政局

房价数据来自于 xxx.com 网站

距离数据各省市以其省会为经纬度源自 geopy 的 Nominatim 广义距离代表了各种形式的距离比如经济距离

对于文化亲近度本文将距离、方言分布、文化分布作 PCA 分析得到文化亲近度地理亲近度使用 ** 指数衰减函数 ** 或 ** 反距离加权 ** $Closeness = e^{-\lambda \cdot Distance}$, (λ 为衰减参数) 饮食文化相似度菜系分类树 (如八大菜系及其子类), 省份所属菜系标签, 基于树形结构的 ** 最近公共祖先 (LCA) 距离 **, $Similarity = \frac{1}{1+LCA \text{ depth}}$ 方言相似度基于各省方言人口比例 (如吴语、粤语、官话等) 采用余弦相似度或 Jensen-Shannon 相似度比较两省方言分布向量最后使用熵值法得到文化亲近度指标其步骤如下计算指标比重 $p_{ij} = \frac{x_{ij}}{\sum_{i=1}^n x_{ij}}$ (对第 j 个指标, 第 i 个省份) 计算信息熵 $e_j = -k \sum_{i=1}^n p_{ij} \ln p_{ij}$, 其中 $k = 1/\ln(n)$ 权重 $w_j = 1 - e_j$ $\sum_{j=1}^m (1 - e_j) w_j = 1$ 指数 = $\sum (\text{标准化指标值} \times w_j)$

由于教育的变量存在高度相关性所以使用主成分分析法将教育变量整合为一个指标同样的对于医疗资源使用熵值法合成一个指标

5.3 模型上

变量表

5.4 方法上

基于以上的推导在代码中需要先求出个体的似然而个体的总似然是 迁移选择概率和工资观测概率的乘积那么集体的似然函数同时对未观测的随机效应 (固定效应、地区匹配效应等) 进行积分 (离散求和):

$$L_i = \sum_{\text{所有随机效应组合}} \left(\prod_t \rho(x_t, j_t) \cdot P(w_t | \text{随机效应}) \right)$$

本文采用混合似然估计法进行参数求取由于计算的复杂性传统的 stata 难以解决如此复杂的计算因此本文基于 python 代码通过逆向归纳 (从最大年龄向最小年龄迭代) 计算每个状态下的期望价值函数 (EV 矩阵), 并得到选择概率

v

(通过 Logit 公式) $\rho(x, j) = \frac{\exp(v(x, j))}{\sum_{k=1}^J \exp(v(x, k))}$

对于求取参数的方法，本文采取以自动微分 (PyTorch) 为核心，结合 L-BFGS 优化器，并利用离散化与并行处理提升效率传统的 Newton 线搜索方法上求解似然函数最大值，Newton 法收敛速度快，但计算海森矩阵需要消耗大量资源虽然可以通过对似然函数进行 LU 分解，在计算海森矩阵的逆时提供可靠帮助，降低数值不稳定性，提高效率但计算海森矩阵的代价仍然较高，尤其是在参数较多时。本文做出改进，使用现代优化算法 Quasi-Newton 方法中的 L-BFGS 方法可以避免收敛困难和数值不稳定。BFGS 是一种拟牛顿法，通过迭代逼近目标函数的海森矩阵 (Hessian 矩阵) 的逆矩阵，从而避免直接计算二阶导数。它只需要利用梯度信息来更新海森矩阵的近似，使得每次迭代都能更准确地找到下降方向。在高维参数空间中，避免了直接计算海森矩阵的高计算成本。

在具体的梯度计算中自动微分通过计算图追踪运算过程，利用链式法则自动计算导数避免数值微分的截断误差。计算复杂度与原始函数同阶，适合高维参数。当模型复杂、参数众多，且需要频繁计算梯度时 (如神经网络训练)，自动微分显著优于手动编码梯度。对于需要高阶导数或大规模并行计算的情况尤为适合。

对固定效应和地区匹配效应 `upsilon` 离散为有限点，通过网格遍历求和这点与原作者的做法保持一致

同时基于更现代的方法本文用

`joblib.Parallel`

并行计算个体似然

本文使用类方法封装待估参数类继承

`torch.nn.Module`

，所有参数为

`torch.nn.Parameter`

，支持自动梯度计算

直接计算大量概率的乘积可能导致数值下溢，取对数可以避免这个问题。但需要确保每个个体的似然计算已经处理了对数转换，或者在汇总时处理。

与传统的计量方法不同本文的优化方法基于深度学习两者的参数初始化往往不同在深度学习中，参数通常随机初始化 (如 Xavier 初始化)，但在结构方程模型或动态离散选择模型中，初始值的选择更为关键，因为模型可能存在多个局部最优，良好的初始值有助于找到全局最优。优化算法 (如牛顿法、L-BFGS) 需要一个初始猜测点开始迭代。初始值的选择可能会影响收敛速度和结果。例如，如果真实值接近 -0.1，好的初始值能加快收敛。根据经济学理论进行初始值赋予能促进优化方法的应用，例如距离增加可能降低迁移概率，因此

gamma_distance

预期为负，将其初始值设为负符合理论预期，有助于引导优化方向。对于这种方法的引入，存在可能的担心，即初始值的设定会引入主观偏差。但实际上，只要优化过程充分收敛，初始值的影响会减小。不过，若初始值离真实值太远，可能导致优化失败，因此基于领域知识的合理初始值是有必要的。

为了避免陷入局部最大化最后用 Nelder-mead 检查局部最大值比对检查以上两种方法

6 估计结果

使用

7 结论

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 2023 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

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从而引入其他变量

8 附录：证明

Rust 极值分布公式中的随机效用项假设服从于一类极值分布 We assume that ζ_j is drawn from the Type I extreme value distribution. In this case, using arguments due to McFadden (1973) and Rust (1987), we have

$$\exp(\bar{v}(x)) = \sum_{k=1}^J \exp(v(x, k))$$

这表示如果变量服从一类极值分布，那么 $\exp(\bar{v}(x))$ 可以表示为所有 $v(x, k)$ 的指数和它意味着，在状态 x 下，选择某一选项 j 的概率与效用的指数值成比例。这个性质广泛用于预测个体选择的分布。

$$\text{地点选择概率 } \rho(x, j) = \frac{\exp(v(x, j))}{\sum_{k=1}^J \exp(v(x, k))}$$

Probability that a person in state x will choose location j can then be written as

$$\rho(x, j) = \exp[v(x, j) - \bar{v}(x)]$$