

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

作者：孙伟增 张晓楠 郑思齐

报告人：蒋舒颖

2020.10.29

目录

- 1 引言与文献评述
- 2 数据与实证方法
- 3 实证结果与讨论
- 4 异质性分析
- 5 机制探讨
- 6 结论性评述

一、引言与文献评述

实际问题



理论背景



研究意义与贡献

- 中国早期粗放的经济发展模式带来的环境隐患不断显现。—2013年极端雾霾天气
- 空气污染不仅直接危害人体生理健康，还会对人们的精神健康产生负面影响（降低人们的主观幸福感和心理健康程度）
- 出于身心健康的考虑，越来越多的家庭搬离了空气污染严重的城市。

空气污染已然成为影响劳动力就业选址的重要因素。

一、引言与文献评述

实际问题

理论背景

研究意义与贡献

- 根据Rosen-Roback的城市空间均衡理论，劳动力的空间流动的影响因素：**收入、房租(生活成本)、城市宜居性特征**
- 传统的发展经济学和新经济地理模型表明劳动力就业影响因素：**非农产业经济比重高、人口规模大、预期收入高**。（段成荣，2001; 王桂新等，2012）
- 城市经济学者基于Tiebout “用脚投票” 理论发现：具有**更高生活质量**的城市更容易吸引劳动力的进入(Whisler et al. ， 2008)。
 - 其中**公共服务水平，如基础教育和医疗水平**等，会对劳动力的选址行为产生显著影响（夏怡然和陆铭，2015）。
- 作为生活质量的重要组成因素，环境质量对劳动力选址决策的影响作用：
 - 人们确实会对环境质量 “用脚投票 “，并且存在规模效应和结构效应。（Banzhaf & Walsh, 2008)
 - 环境污染对人们的移民意愿会产生影响，推动着人群的迁移，其中对高技能劳动力的跨省流动存在显著的负向影响。（楚永生，2015；李明和张依然，2019）

一、引言与文献评述

实际问题

理论背景

研究意义与贡献

- 理论意义：丰富了环境污染与劳动力流动的相关研究成果。
- 现实意义：流动人口正成为中国区域经济增长的关键动力，充分理解流动人口的选址决策机制，将有助于城市政府制定相应的政策吸引流动人口，从而促进本地经济增长。
- 贡献：
 - 本文利用精确度高、覆盖范围广的PM2.5 卫星栅格数据和全国流动人口动态监测调查数据，准确地识别了空气污染对中国流动人口就业选址行为的影响效果，并得到了稳健的研究发现，为流动人口选址研究提供了新的经验证据。
 - 本文基于显示性偏好理论对流动人口选址过程中空气污染和收入之间替代关系的讨论，为环境污染的社会成本评估提供了新的视角。
 - 本文的异质性研究结论为地方政府通过改善环境质量吸引劳动力流入的政策制定提供了经验依据。

二、数据与实证方法

(一) 数据来源与处理

1、空气污染数据

PM2.5浓度数据来源于Van Donkelaar et al. (2016) 计算的全球年度PM2.5卫星栅格数据。我们首先利用地理信息系统将每个栅格定位到其空间位置所在的城市上，然后将落在每个城市内的所有栅格数据进行平均，即可得到各个城市在不同年份的PM2.5浓度水平。

2、流动人口调查数据

流动人口数据来自于2011—2015年国家卫生和计划生育委员会开展的全国流动人口动态监测调查。在实证研究中只保留了在调查时流入本地时间不超过一年的样本。经过样本筛选后，2011—2015年每年的样本量分别为2.7万、4.8万、4.1万、4.5万和4.5万。

二、数据与实证方法

(二) 计量模型——模型介绍

采用条件logit 模型来识别城市特征对于流动人口选址决策的影响。具体的实证模型设定如下：

$$\begin{aligned} choice_{ijt} = & \alpha_1 \cdot pm_{j,t-1} + \alpha_2 \cdot netincome_{ijt} + \alpha_3 \cdot distance_{ij} \\ & + \alpha_4 \cdot dial_distance_{ij} + \alpha_5 \cdot fellow_share_{ij,t-1} + \alpha_6 \cdot X_{j,t-1} + \varepsilon_{ijt} \end{aligned}$$

- 下标 i 代表流动人口个体， j 代表就业地， t 为年份。
- 被解释变量 $choice_{ijt}$ 是一个哑元变量， $choice_{ijt} = 1$ 表示在年份 t 时流动人口 i 选择到城市 j 就业； $choice_{ijt} = 0$ 表示没有选择到城市 j 就业。
- 解释变量 pm_{jt} 为城市 j 在年份 t 的平均 PM2.5 浓度。
- 其余为控制变量和随机扰动项。

二、数据与实证方法

(二) 计量模型——被解释变量的说明

- 每个流动人口 i 在进行就业地选择时都有多个备选城市(用 N_i 表示)
- 第一, 每个流动人口 i 和其对应的备选城市 j 是该数据的一个样本, 即总共有 $\sum_i^I N_i$ 样本, 其中 I 代表流动人口总数。
第二, 对于以流动人口 i 为核心的 N_i 个样本, 只有流动人口 i 的实际就业地城市 j 对应的choice 变量取值为 1,
第三, 解释变量 pm 和 X 均是随城市 j 和时间 t 变化的变量, 与 i 无关。

表 1 条件 logit 模型的数据结构

样本	流动人口 i	城市选择集 j	选择结果 $choice_{ij}$	空气污染 pm_{jt}	控制变量 X_{jt}
1	1	1	1		
2	1	2	0		
...	0		
N_1	1	N_1	0		
$N_1 + 1$	2	1	0		
$N_1 + 2$	2	2	1		
...	0		
$N_1 + N_2$	2	N_2	0		
...		

二、数据与实证方法

(二) 计量模型——备选城市集的说明

- 从理论上讲，对于每一个流动人口，所有城市都可以作为其就业地的备选方案
- 在现实中，由于地理、信息等条件的限制，流动人口很难同时掌握所有城市的信息。在面临更多的不确定性时，流动人口倾向于在信息比较充分的城市之间进行选择。
- 为此，首先计算来自每个省份的流动人口(按户籍所在地划分)在各个城市就业的人数占比，然后根据该比例从大到小对所选城市进行排序并累加，最后将**前 90%** 流动人口选择的所有城市作为该省份流动人口就业地选择的备选城市集。
- 通过上述处理，每个流动人口的备选城市数量为3—113 个，平均为 53 个城市。这样处理一方面保证了估计结果的准确性，另一方面也极大的降低了模型的运算量。

举例：户籍为福建省的流动人口

所在就业城市	占比
深圳	20%
广州	15%
福州	14%
.....
哈尔滨	10%

备选城市集：深圳、广州、福州.....

二、数据与实证方法

(二) 计量模型——四个维度控制变量的说明

- **维度一：**流动人口在就业城市的预期净收入 $netincome$ ，即家庭总收入与总支出之差。
 - 为了得到每个流动人口在各个备选城市的预期净收入，首先**分城市**估计家庭净收入的决定方程：

$$netincome_{it} = \beta_1 X_{it} + \beta_2 Z_{it} + \beta_3 year_t + \varepsilon_{it}$$

- 被解释变量 $netincome$ 是被调查者 i 在年份 t 实际获得的**家庭净收入**。
- X 代表被调查者 i 在年份 t 时**个体属性**，包括年龄、性别、民族、受教育水平、户口性质、婚姻状态以及户籍所在省份；
- Z 为个体 i 在年份 t 的**就业情况**，包含所在行业、职业、单位性质和就业身份(分为雇员、雇主、自营劳动者和家庭帮工 4 种类型)；
- $year$ 为**调查年份** t 。

这样对于每个城市都可以得到一组估计系数 β 。然后对于每个流动人口 i 和其对应的备选城市 j ，利用城市 j 的估计系数以及流动人口的个体特征和就业情况，就可以计算得到流动人口 i 到城市 j 就业的预期净收入。

二、数据与实证方法

(二) 计量模型——四个维度控制变量的说明

- **维度二：城市的经济、人口以及公共服务特征**，包括人均 GDP、年末总人口、第三产业占 GDP 的比重、每万人小学教师数、每百人公共图书馆藏书数、每万人医生数等变量。
- **维度三：从劳动力市场需求的角度，构造Bartik Index来反映各个城市对劳动力的需求水平**
 - Bartik Index 的计算方法：以本城市分行业就业人数比例作为权重，以全国分行业就业人数增长率作为劳动力需求的度量指标，通过求取加权平均计算得到的城市综合劳动力需求指数。
- **维度四：备选城市与流动人口家乡的地缘关系**，包括就业地与流动人口家乡的空间距离、就业地与流动人口家乡的方言距离、就业地的流动人口中与流动人口来自相同省份的比例，分别用来控制城际交通成本、文化差距、同乡关系对于流动人口就业地选择的影响。

二、数据与实证方法

(二) 计量模型

表 2 主要变量定义及描述性统计量

变量名	定义	样本量	均值	标准差	最小值	最大值
<i>choice</i>	被调查者是否流入该城市	8395081	0.02	0.14	0	1
<i>pm</i>	PM2.5 浓度 ($\mu\text{g}/\text{m}^3$)	8395081	43.40	15.86	6.72	94.48
<i>netincome</i>	预期家庭净收入(千元/月)	8395081	2.64	3.10	-60.49	261.27
<i>gdp</i>	人均 GDP(万元)	8395081	6.41	4.58	0.53	46.77
<i>tertiary_gdp</i>	第三产业占 GDP 比重(%)	8395081	42.78	11.08	9.76	77.95
<i>pop</i>	年末总人口(百万人)	8395081	5.74	4.50	0.2	33.75
<i>primary_teacher</i>	每百人小学教师数(人)	8395081	0.43	0.17	0.19	1.50
<i>book</i>	每人公共图书馆藏书(册)	8395081	1.16	1.62	0.02	14.39
<i>doctor</i>	每百人医生数(人)	8395081	0.28	0.15	0.06	0.87
<i>bartik_index</i>	劳动力需求指数	8395081	0.07	0.10	-0.10	0.32
<i>distance</i>	户籍所在省份与备选城市的直线距离(100km)	8395081	7.72	5.70	0.12	36.84
<i>fellow_share</i>	备选城市的流动人口中与该流动人口来自相同省份的比例	8395081	0.18	0.28	0	1
<i>dial_distance</i>	户籍所在省份与备选城市的方言距离指数	8395081	58.66	39.88	0	100

二、数据与实证方法

(二) 计量模型——工具变量

- 用各个城市分年度的**空气流动系数**(vc_{jt}) 作为空气污染的**工具变量**。该工具变量取决于城市当年风速和行星边界层高度，二者都是由气象状况和地理条件决定，在控制了城市固定效应后，与本地区的人口流入没有直接关系，可以满足工具变量的外生性假设。

基于该工具变量，本文分两阶段进行回归分析。

第一阶段：

$$pm_{jt} = \varphi_1 \cdot vc_{jt} + \varphi_2 \cdot X_{jt} + \omega_t + \rho_j + \varepsilon_{jt}$$

第二阶段：

$$\begin{aligned} choice_{ijt} = & \lambda_1 \cdot \widehat{pm}_{j,t-1} + \lambda_2 \cdot netincome_{ijt} + \lambda_3 \cdot distance_{ij} \\ & + \lambda_4 \cdot dial_distance_{ij} + \lambda_5 \cdot fellow_share_{ijt-1} + \lambda_6 \cdot X_{j,t-1} + \varepsilon_{ijt} \end{aligned}$$

- 其中，第一阶段在控制城市经济、人口和公共服务特征、控制年份固定效应 ω_t 和城市固定效应 ρ_j 的基础上，用工具变量**空气流动系数**解释就业地的空气污染水平，并基于回归系数得到当年 **PM2.5**浓度的预测值。
- 第二阶段中用一阶段得到的 **PM2.5 浓度预测值**替代实际 **PM2.5 浓度观测值**的作为**主要解释变量**，通过条件 logit 模型分析空气污染对流动人口选址决策的影响效果。

三、实证结果与讨论

(一) 基本结果

表 3 流动人口就业选址的条件 logit 模型估计结果

被解释变量: <i>choice</i>	条件 logit				条件 logit + IV
	(1)	(2)	(3)	(4)	(5)
<i>pm</i>	-0.00116 *** (-25.62)	-0.00118 *** (-25.98)	-0.000576 *** (-14.33)	-0.000281 *** (-10.55)	-0.00388 *** (-9.20)
<i>netincome</i>		0.00362 *** (27.02)	0.00174 *** (21.52)	0.00112 *** (23.89)	0.00205 *** (18.33)
<i>gdp</i>			0.00319 *** (23.41)	0.00179 *** (23.40)	0.00278 *** (19.03)
<i>tertiary_gdp</i>			0.00107 *** (22.60)	0.00197 *** (56.79)	0.00411 *** (20.69)
<i>pop</i>			0.00444 *** (65.25)	0.00405 *** (56.70)	0.00200 *** (6.95)
<i>primary_teacher</i>			0.0619 *** (19.21)	0.0113 *** (5.78)	0.0116 *** (3.20)
<i>book</i>			0.0112 *** (25.05)	0.00264 *** (10.91)	0.00240 *** (6.07)
<i>doctor</i>			0.223 *** (39.57)	0.182 *** (40.20)	0.335 *** (21.70)

续表 3

被解释变量: <i>choice</i>	条件 logit				条件 logit + IV
	(1)	(2)	(3)	(4)	(5)
<i>bartik_index</i>			0.773 *** (31.05)	0.398 *** (28.47)	0.770 *** (19.37)
<i>distance</i>				-0.00916 *** (-44.19)	-0.0167 *** (-23.27)
<i>fellow_share</i>				0.113 *** (56.58)	0.205 *** (25.22)
<i>dial_distance</i>				0.000955 *** (20.05)	0.00150 *** (14.91)
<i>dial_distance^2</i>				1.91E-7 (0.41)	2.36E-6 *** (2.75)
N	8395081	8395081	8395081	8395081	8290930
pseudo R ²	0.001	0.001	0.062	0.103	0.103

通过逐步加入控制变量的方式来考察空气污染变量对流动人口就业选址的影响。结果均显示城市 PM2.5 浓度与流动人口选址在该城市的概率呈现出显著的负相关关系。

三、实证结果与讨论

(一) 结果讨论

- 为了更加清晰的理解空气污染变化对中国流动人口空间流动影响效果的大小，分别计算了255个地级及以上城市2011—2015年间由于空气污染变化带来的流动人口流入数量的变化，并与收入变化带来的影响效果进行对比。
- 第一，大部分城市集中在45°线附近，表明2011年以来由于空气质量变化导致的流动人口流动已经逐渐接近收入变化带来的影响。
- 第二，在北京、广州、杭州等大城市，尽管收入增长对于流动人口的吸引力仍然占据主导地位，但是空气质量下降对流动人口就业选址的负向影响已经相当于收入效应的30%，成为限制劳动力供给的重要因素。
- 第三，在成都、重庆、西宁、济南等收入增幅较小或为负的城市，空气质量改善成为了它们进一步吸引劳动力的重要筹码。

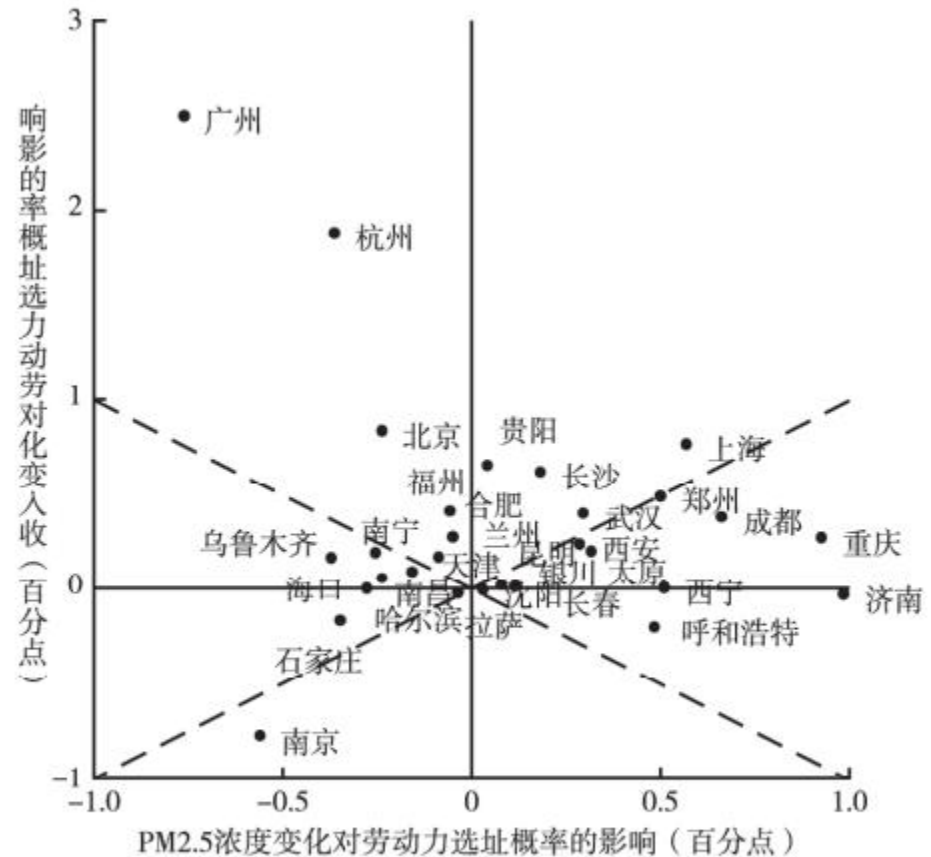


图1 2011—2015年PM2.5浓度和收入变化对流动人口选址概率的影响

四、异质性分析

（一）年龄和性别

15—29岁，30—44岁，45岁及以上；男性，女性

（二）受教育水平

高中及以下，大专及以上

（三）婚姻状态和生育状况

未婚，已婚；未生育，已生育

（四）来源地空气质量和户口性质

户籍地空气质量较差，户籍地空气质量较好；农业户口，非农户口

（五）职业类型

国家机关、党群组织、企事业单位负责人；专业技术人员；公务员、办事人员和有关人员；商业、服务业人员；农、林、牧、渔、水利业生产人员；生产、运输设备操作人员及有关人员；无固定职业

结果均稳健，且符合预期。

年龄越大、受教育水平越高，以及男性、已婚 / 已生育、非农业户口的流动人口在就业选址时对空气污染的敏感性要更高。

五、机制探讨

(一) 空气污染对流动人口健康状况的影响

空气污染 → 影响流动人口健康状况 → 就业选址决策

这一部分结合卫计委在2014年开展的“社会融合与心理健康”调查，该项调查主要针对北京市朝阳区、浙江省嘉兴市、福建省厦门市、山东省青岛市、河南省郑州市、广东省深圳市和中山市、四川省成都市八个城市（区）的流动人口样本进行访问。

表 9 专题调查中关于健康方面的问题

问题	回答
A. 我好像比别人容易生病	1. 绝对正确;2. 大部分正确;3. 不能肯定;4. 大部分错误;5. 绝对错误
B. 我认为我的健康状况在变坏	1. 绝对正确;2. 大部分正确;3. 不能肯定;4. 大部分错误;5. 绝对错误
C. 我的健康状况非常好	1. 绝对正确;2. 大部分正确;3. 不能肯定;4. 大部分错误;5. 绝对错误

取值为：1

表 10 空气污染对健康的影响的主要变量定义及描述性统计量

变量名	定义	样本量	均值	标准差
healthA	被调查者比别人不容易得病	14443	0.63	0.48
healthB	被调查者健康状况在改善	14443	0.64	0.48
healthC	被调查者健康状况较好	14443	0.80	0.40

五、机制探讨

(一) 空气污染对流动人口健康状况的影响

采用线性概率模型来估计空气污染对流动人口健康状况的影响，具体模型设定如下

$$health_{ijt} = \gamma_1 \log(pm_{jt}) + \gamma_2 X_{ijt} + \gamma_3 Y_{jt} + \rho + t + \varepsilon_{ijt}$$

- i 代表流动人口个体， j 代表城市， t 为调查年份；
- 被解释变量 $health$ 反映了在调查年份 t 时生活在城市 j 的流动人口 i 的身体健康状况。
- pm 为城市 j 在年份 t 时的平均 PM2.5 浓度。
- X 为流动人口的个体特征变量，包括年龄、性别、受教育水平、婚姻状态、家庭净收入、职业、就业身份。
- Y 为所在城市的经济人口和公共服务指标，包括人均 GDP、第二产业和第三产业占 GDP 的比值、总人口数、本地人均医院数量。
- 控制流动人口来源地的省份固定效应 ρ 和流动人口流入该城市的年份固定效应 t

五、机制探讨

(一) 空气污染对流动人口健康状况的影响

$\log(pm)$ 变量的系数在三个方程中都显著为负，表明空气污染对于流动人口的身体健具有显著的负面影响。

表 11 空气污染对健康的影响结果

	(1)	(2)	(3)
	<i>healthA</i>	<i>healthB</i>	<i>healthC</i>
$\log(pm)$	-0.153 ^{***} (-3.25)	-0.143 ^{***} (-3.05)	-0.0827 ^{**} (-2.03)
控制变量	是	是	是
N	14443	14443	14443
R ²	0.047	0.066	0.032

空气污染 → 影响流动人口健康状况 → 就业选址决策

六、结论性评述

（一）研究结论

- 第一，空气污染对于流动人口的就业选址具有显著的负向影响；
- 第二，通过对比空气污染和收入对流动人口就业选址概率的影响可以计算得到流动人口对于 $\text{PM}_{2.5}$ 浓度降低 $1 \mu\text{g} / \text{m}^3$ 的支付意愿约为326元/月。
- 第三，年龄越大、受教育水平越高，以及男性、已婚 / 已生育、非农业户口的流动人口在就业选址时对空气污染的敏感性要更高。
- 此外，研究还发现空气污染对于流动人口的健康状况具有显著的负向影响。

（二）政策意义

- 为中国城市与环境治理和劳动力流动相关的政策制定提供了经验依据。

感谢您的聆听！