

互联网如何推动劳动力大规模迁移?

——基于骨干网提速的证据

喻理 吕卓阳 王语柔 张云

目录

附录 I	背景事实	1
附录 II	描述性统计	2
附录 III	对工具变量外生性条件成立性的讨论	3
附录 IV	其他稳健性讨论	5
附录 V	异质性分析	10
附录 VI	理论分析	11
附录 VII	对双边宗族网络异质性的讨论	14
附录 VIII	移民网络的影响群体	15
附录 IX	排除替代性假说	16

附录 I 背景事实

(一) 宽带互联网出现的历史背景

在 20 世纪 90 年代中后期, 中国大陆开始接入互联网。1994 年, 高能物理研究所和斯坦福线性加速器中心之间建立了网络连接, 标志着中国大陆正式接入国际互联网。但当时的互联网使用饱受两个问题的困扰, 一是带宽不足和每秒 6-10kb 的拨号上网速度, 二是较为高昂的初装费和使用费用。前者使得互联网在 2000 年之前仅能以文本方式呈现信息, 后者则极大的限制了互联网的潜在用户数量^①。2000 年, 宽带互联网的出现极大地改善了上述状况。在国务院的督导下, 骨干网提速计划于 2000 年启动, 使中国互联网平均下载速度提升至 256-512kb/s, 网费也从 1999 年的数千元降至 2000 年的几百元。这些变化促使互联网拨号上网用户数从 1999 年的 256 万激增至 2001 年的 2650 万, 增长约 10 倍; 网络站点数则从 9906 个增加到 242739 个, 增长约 25 倍^②。

宽带的出现显著改善了劳动力市场的信息环境。首先, 改善了企业在 2000 年之前主要依赖报刊、电视广告或街边摊位进行招聘的方式。2000 年后, 借助网络基础设施的改善, 各地政府在大中城市全面推进劳动力市场信息化、规范化和现代化建设, 启动了全国劳动力市场信息骨干网建设, 整合各方就业信息, 向公众提供就业服务^③。其次, 促进了招聘网站的兴起^④。最后, 宽带互联网推动传统纸质媒体转型上网, 使得人们可以通过网络获取大量外地就业机会信息^⑤。

(二) 户籍制度的变革背景

中国计划经济体制的早期阶段, 政府主要通过户籍制度严格控制劳动力流动。这一制度根据职业将居民划分为农业户口与非农业户口, 同时根据其出生地区分农村人口与城市人口。彼时在农村和城市, 农业与非农业部门之间的流动均被禁止或受到严格管控。即使部分持有农业户籍的劳动力可以通过付出一定代价迁移至城市, 但由于户籍限制, 这部分迁移的劳动力也只能在指定的岗位就业 (Démurger *et al.*, 2009)。

直至 1980 年代末, 户籍制度才有所放松。主要表现为部分城市开始放松户籍制度, 但各省份可购买的城市户口数量仍然受到配额限制。在 1997 年的户籍试点改革中, 农村居民开始被允许迁往指定的乡镇和小城市, 但这些地方对农村居民的吸引力有限, 改革效果并不显著。随着中国加入世贸组织, 企业对劳动力的需求与劳动力的供给之间的矛盾日益突出, 导致地方政府限制人口流动的机会成本迅速提高 (Tian, 2022)。上述因素促使了中央政府在宏观上推动农村劳动力流动至城市的政策变革。与之伴随的是 2001 年以地级市为单位的渐进式户籍改革, 这次改革取消了农业与非农业户籍的区别, 用统一的入户条件替代原有的配额制度, 显著提升了城市劳动市场的劳动力供应增长。

^① 根据中国互联网络信息中心 (CNNIC) 1999 年的调查数据, 只有 2% 的受访者能够承担数千元的初装费用。

^② 信息来源: 2001 年中国互联网络发展状况统计报告。

^③ 信息来源: http://www.mohrss.gov.cn/SYrlzyhshbzb/zwgk/szrs/tjgb/200603/t20060301_69895.html。

^④ 我国三大招聘平台前程无忧、中华英才网和智联招聘分别成立于 1999 年、1997 年与 1997 年。三者均于 2000 年正式涉足互联网招聘。其中, 中华英才网和智联招聘在 2000 年之前通过公布黄页信息收费, 正式的互联网招聘业务始于 2000 年。此外, 部分地方主要的招聘网站也纷纷起于 2000 年。例如, 南方人才网自 2000 年正式开通, 是国家人事部与广州市政府合办的中国南方人才市场旗下的大型人力资源专业网站, 卓博人才网也起于 2000 年左右。信息来源: 中国对外服务工作行业协会。

^⑤ 仅 2000 年就有超过 20% 的新闻媒体实现信息互联网化。

附录 II 描述性统计

表 III 展示了本文关键变量的描述性统计, 从中可知, 1999-2004 年劳动力流动率为 0.8623%, 年平均劳动力流动数量为 3.3097 万人, 且省内劳动力流动数量要显著高于省外劳动力流动数量。其中, 劳动力流动率最高的城市主要为东莞、深圳及中山等珠三角城市。此外, 我们也根据 Chen *et al.* (2022) 的方法构造了 1990 年至 2010 年的劳动力流动率的长面板数据, 从 1990 年至 2010 年, 伴随着我国的经济的发展, 全国平均劳动力流动率由 2.8860% 增长至 14.1658%, 大量的流动人口从农村流入城市。根据本文的描述统计, 1999 年-2004 年全国劳动力流动总数量为 6692.1605 万人。

表 III 描述性统计

变量含义	样本数	均值	方差	最小值	最大值
1999—2004 年流动劳动力流动率 (%)	2022	0.8623	1.5431	0.0000	19.9505
1999—2004 年省内流动劳动力比例 (%)	2022	0.4786	0.5703	0.0000	5.3837
1999—2004 年跨省流动劳动力比例 (%)	2022	0.3837	1.1810	0.0000	17.5619
1999—2004 年流动劳动力人数 (万人)	2022	3.3097	8.6674	0.0000	130.8892
1990—2010 年流动劳动力流动率 (%)	1286	6.3058	8.8432	0.0000	76.8459
1990 年平均流动劳动力流动率 (%)	301	2.8860	5.3362	0.0000	69.8740
2000 年平均流动劳动力流动率 (%)	314	2.9230	4.8982	0.1578	46.1531
2005 年平均流动劳动力流动率 (%)	338	4.7502	7.3372	0.0501	66.3861
2010 年平均流动劳动力流动率 (%)	333	14.1658	10.6992	0.0000	76.8459

注: 表 III 前三行为使用 2005 年人口抽样调查数据构造的劳动力流动率面板数据, 第四行为使用 2005 年人口抽样调查数据构造的各地劳动力流动的平均人数。第五行为使用 1990-2010 年人口抽样调查数据构造的五年平均劳动力流动率面板数据, 第六至九行报告了利用逐轮人口抽样调查数据构造的劳动力流动率的描述性统计量。样本数变量维度为城市-年。

附录 III 对工具变量外生性条件成立性的讨论

本文在此采用两种方法对工具变量的外生性条件成立的成立性进行讨论。首先, 效仿 (Manacorda and Tesei, 2020) 的检验思路, 将工具变量在一阶段回归中的因变量替换为同各地经济发展情况有关的变量, 包括地区生产总值、地区生产总值增长率、人口密度等随时间变化的衡量地方经济发展程度的变量。如果工具变量在一阶段的回归中同上述变量相关, 则违背了外生性条件, 反之则为外生性条件的成立性提供了支持。两个工具变量的估计结果见表 III1 与表 III2, 均未发现各地经济发展情况同两个工具变量存在显著关联。

表 III1 对于闪电发生率外生性条件的检验

	(1)	(2)	(3)	(4)	(5)	(6)
	地区生产 总值	地区生产 总值增长率	人口 密度	行政区域 土地面积	规模以上 工业总产值	固定资产 年均余额
<i>Flash</i> × <i>Post</i>	-3.2844 (4.7251)	36.5122 (28.0526)	-4.2829 (6.7987)	-0.3720 (4.9268)	-4.2438 (3.9057)	2.0755 (3.6874)
城市固定效应	Y	Y	Y	Y	Y	Y
年份固定效应	Y	Y	Y	Y	Y	Y
R^2	0.973	0.414	0.939	0.957	0.987	0.963
N	1474	1198	1473	1474	970	1469

表 III2 对于各城市中心到虚拟骨干网最近福利外生性条件的检验

	(1)	(2)	(3)	(4)	(5)	(6)
	地区生产 总值	地区生产 总值增长率	人口 密度	行政区域 土地面积	规模以上 工业总产值	固定资产 年均余额
<i>Distance</i> × <i>Post</i>	-1.2311 (1.7251)	4.2331 (3.9871)	3.4711 (2.9901)	-0.1211 (1.1209)	1.5891 (2.2317)	2.0755 (3.6874)
城市固定效应	Y	Y	Y	Y	Y	Y
年份固定效应	Y	Y	Y	Y	Y	Y
R^2	0.973	0.414	0.939	0.957	0.987	0.963
N	1474	1198	1473	1474	970	1469

另外一种检验思路来自 Oster (2019)。在实证分析中, 由于无法将全部控制变量加入到回归方程中, 不可避免的会面临遗漏变量导致的内生性问题。正文中已考虑了诸多可能会影响工具变量外生性条件的混杂渠道, 但是可能还存在观测不到的混杂性因素^①。为了缓解这一问题, 本文借鉴 Oster (2019) 的思路, 在工具变量两阶段回归的过程中进行系数敏感性分析。即首先通过一阶段获取 DID 项的拟合值, 随后在第二阶段的回归中, 分别进行加入和不加入控制变量两类回归, 如果在加入正文的控制变量后第二阶段的估计系数仍然稳定, 那么就认为遗漏变量造成的偏差较小。Oster (2019) 提出的敏感性分析的步骤如下, 考虑如下一般性的回归方程:

$$Y = \alpha + M\beta + W_1 + W_2 + \varepsilon \quad (1)$$

其中, Y 是因变量, M 是处理变量 (在正文中为 DID 项), W_1 为可观测控制变量, 即原文工具变量回归设定中的控制变量, W_2 为不可观测的控制变量, 即我们担心的潜在影响工具变量外生性条件成立性的控制变量, ε 是回归误差项。Oster (2019) 的方法使用可观测变量 W_1 与 M 之间相关性信息来计算不可观测变量 W_2 与 M 之间的相关性, 以估计遗漏变量引起的偏误大小。Oster (2019) 证明, 处理效应的无偏估计为:

$$\beta^* \approx \tilde{\beta} - \delta(\tilde{\beta} - \tilde{\beta}) \frac{R_{\max} - \tilde{R}}{\tilde{R} - \tilde{R}} \quad (2)$$

^① 感谢审稿专家的宝贵建议。

其中, $\hat{\beta}$ 和 \hat{R} 由方程 $Y = \alpha + M\beta + \varepsilon$ 估计得到, 即只包含处理效应。 $\tilde{\beta}$ 和 \tilde{R} 由方程 $Y = \alpha + M\beta + W_1 + \varepsilon$ 估计得到, 即同时包含处理效应和可观测的控制变量 W_1 。 R_{\max} 则由方程 $Y = \alpha + M\beta + W_1 + W_2 + \varepsilon$ 估计得到, 即同时包含处理效应、可观测控制变量 W_1 以及不可观测控制变量 W_2 。

使用 Oster (2019) 的方法, 需要设定两个参数, 第一个参数是可观测的控制变量 W_1 与不可观测的控制变量 W_2 的相对重要性 δ 。其中, δ 为选择比例 (selection proportionality), 它衡量可观测变量与关注变量的相关关系相较于不可观测遗漏变量与关注变量的相关关系的强弱 (马双和赵方博, 2019)。当 $\delta = 1$ 时, 表明可观测的控制变量和不可观测的控制变量对系数大小的影响同等重要, 且对系数的影响方向相同; 而当 $0 < \delta < 1$ 时, 则假定认为不可观测的控制变量的重要性要小于可观测控制变量的重要性。第二个参数是模型最大的拟合优度 R_{\max} 。 R_{\max} 合计度量了可观测控制变量和不可观测控制变量对 Y 的解释能力, 也即是说, 若不可观测的遗漏变量能够被观测, 回归方程的最大拟合优度。

在本文中, 我们借鉴马双和赵方博 (2019) 的思路, 主要关注对参数 δ 的讨论。具体来说, 即通过给定 R_{\max} 的值, 计算使处理效应等于 0 时 δ 的取值, 若 δ 取值大于 1, 则通过了稳健性检验。另外, 考虑到如果 δ 的值小于 0, 则经过遗漏变量偏差调整后的系数会大于未经调整回归得到的系数, 鉴于正文的估计系数为正, 因此当 δ 的值小于 0 时, 忽略调整也不会影响实证结论的稳健性。

鉴于工具变量二阶段回归中的 R^2 没有实际意义, 本文同马双和赵方博 (2019) 的具体做法略有差异, 即我们假定 $R_{\max} = 1$, 而不再将其设定为其他值。

Oster (2019) 敏感性分析的估计结果见从表 III3, 从中可以看出对于第一个工具变量而言, 当令 $R_{\max} = 1$ 且处理效应为 0 时, $\delta = -0.0022$ 。根据 Satyanath *et al.* (2017) 的研究结论, 这表明忽略对未知控制变量的控制不会影响使用该工具变量时的实证结果的稳健性。对于第二个工具变量, 当令 $R_{\max} = 1$ 且处理效应为 0 时, $\delta = 0.6267$ 。该值小于 1, 显示不可观测的控制变量的重要性要小于可观测控制变量的重要性。该结果表明使用第二个工具变量时可能存在系数不够稳定的可能, 但是考虑到 Oster (2019) 的方法并不能完全适配工具变量法, 因此, 我们仅将这一结果视作第二个工具变量估计系数较为敏感的暗示性证据, 供读者参考。

表 III3 使用Oster (2019) 方法讨论工具变量的稳健性

检验的工具变量	δ
(1) 闪电发生率	-0.00224
(2) 各城市中心到虚拟骨干网的距离	0.62677

附录 IV 其他稳健性讨论

本文还针对基准回归的设定进行了其他方面的稳健性讨论,包括同期混杂政策对实证结果的影响、对移民率进行回迁调整后的估计结果、通过校准和拼接多次人口抽样调查数据获得的更长时段的估计结果、考虑宽带互联网的出现可能并非合理的准自然实验对估计的影响、考虑率高考移民以及基于不同年龄段劳动力定义下的实证结果,在各种设定下,本文的基本结论均保持稳健^①,细述如下。

(1) 更改测算因变量时对劳动力年龄段的定义。正文在界定劳动人口的年龄时间段时,参考了 Chen *et al.* (2022) 和 Guo *et al.* (2024) 的定义,他们在其论文中均将劳动人口年龄的定义为 15-60 岁。而至于为何选择 15 岁作为起点,本附录对这一问题进行说明。国际上通常将 15 岁定义为劳动年龄人口的起点^②,遵循这一惯例,不少使用中国数据来定义劳动人口的时候也将 15 岁作为年龄的起点 (Imbert *et al.*, 2022; Jin and Zhang, 2023)。中国义务教育阶段一般是到 15 岁,以 15 岁作为劳动力统计的起点,能够涵盖更多实际正在工作的人口。国家统计局在有关劳动就业统计中,也采用 15 岁以上的人口作为统计对象,譬如国家统计局将劳动人口定义为 15 岁及以上陆上非住院人口,并符合就业人口或失业人口定义的人士^③。

考虑到 15-18 岁之间的人口可能是因为教育等因素而流动(例如高考),纳入这部分人群可能会混杂本文的估计结果。为了讨论这一问题,我们首先展示了因高考或者非工作因素的移民比例。具体而言,我们使用 2005 年的人口抽样调查进行了一个简要的描述性统计,结果如表 IV1 所示,可以发现因教育而流动的人口仅占总流动人口的 1.1%,劳动力流动的主要原因还是出于工作需求,这一描述性统计也符合 Zhong (2018) 的发现。

表 IV1 基于 2005 年人口抽样调查的流动人口动机描述性统计

人口流动的理由	占比
工作或经商	82%
跟随亲戚	5.3%
婚姻目的	4.7%
教育目的	1.1%
其他	6.9%

进一步地,考虑到上述人群可能会影响估计的稳健性,我们补充了一系列关于年龄范围的稳健性检验,以保证宽带互联网对劳动力流动的影响不随不同劳动力年龄范围的定义改变而改变。具体而言,我们采取了五种定义方式测度各城市逐年的劳动力流动情况,一是效仿 Jin and Zhang (2023) 和 Imbert *et al.* (2022) 的做法,将劳动人口的年龄范围界定为 15-64 岁;二是考虑到法定劳动年龄的限制,界定为 16-60 岁;三是界定为 16-64 岁;四是考虑到高考移民问题,将劳动人口年龄限制在 18-60 岁;五是限制在 18-64 岁。

五种界定方式下的基准回归结果如表 IV2 所示。其中,五列中的 DID 项的系数均显著为正,且系数大小变化相比原文较为稳定,表明本文的估计结果在不同的对劳动人口年龄范围的定义下均是稳健的。

^① 限于篇幅,具体结果参加本文附录。

^② 来源: <https://zh.wikipedia.org/wiki/%E5%B0%B1%E4%B8%9A%E7%8E%87>。

^③ 来源: https://www.stats.gov.cn/zt_18555/ztsj/sjzsj/sjz2006/202303/t20230303_1924682.html。

表 IV2 改变年龄范围

	(1) 15-64 岁	(2) 16-60 岁	(3) 16-64 岁	(4) 18-60 岁	(5) 18-64 岁
<i>Intensity</i> × <i>Post</i>	0.0464*** (0.0040)	0.0459*** (0.0040)	0.0462*** (0.0041)	0.0425*** (0.0037)	0.0428*** (0.0037)
控制变量	Y	Y	Y	Y	Y
城市固定效应	Y	Y	Y	Y	Y
年份固定效应	Y	Y	Y	Y	Y
R^2	0.818	0.818	0.818	0.832	0.832
N	1393	1393	1393	1393	1393

我们也进一步估计了宽带互联网出现对因高考目的而流动的人口比率的影响, 回归结果如表 IV3 的第 (1) 列所示, 不难看出, 宽带互联网对高考移民确实存在显著正向影响, 但回归系数非常小。我们进一步在基准回归中加入 15-18 岁因学习培训而迁移的人数作为控制变量, 回归结果如第 (2) 列所示, 回归系数在 1% 的水平上显著, 且系数大小同基准回归的第 (3) 列相比仅略微下降 (原文为 0.0459)。综上所述, 忽略高考移民对于基准回归结果系数及稳健性的影响较小。

表 IV3 考虑高考移民的影响

	(1) 高考移民率	(2) 劳动力流动率
<i>Intensity</i> × <i>Post</i>	0.000368** (0.000145)	0.0454*** (0.0041)
高考移民	N	Y
控制变量	Y	Y
城市固定效应	Y	Y
年份固定效应	Y	Y
R^2	0.116	0.819
N	1393	1393

注: 该表第 (1) 列高考移民率的定义为 15-18 岁因学习培训原因迁移的劳动力流动率, 第 (2) 列加入了 15-18 岁因学习培训原因迁移的人数作为控制变量。

(2) 排除其他政策干扰。本文排除了四项混杂政策: 包括 1998 年至 2000 年的国有企业抓大放小改革, 1999 年的高等教育扩招, 2001 年中国加入世界贸易组织 (WTO), 1998-2007 年中国部分城市开始陆续放开户籍制度, 各城市的外资进入放宽政策以及各地开始加速的基础设施建设投资。上述政策或冲击都可能影响人口流动。为此, 我们在基准回归中分别加入了 (1) 1999 年国有企业产值在地市级工业总产值中的比例同 $Post_t$ 的交互项, (2) 1999 年城市级大学生的比例 $Post_t$ 的交互项, (3) 城市层面 2000 年平均关税同 $Post_t$ 的交互项, (4) 根据 Wang *et al.* (2021) 的研究, 在估计样本中剔除了 1999-2004 年陆续放宽户口制度的城市样本, 表 IV4 汇报了剔除上述同期政策后的估计结果, DID 项的系数均显著为正, 表明同期政策并不影响本文结论的稳健性。此外, 在囊括上述各项政策冲击下的 DID 项系数的大小相比基准回归并未大幅变化, 说明宽带出现这一冲击对劳动收入份额的正向作用存在相对的独立性。

表 IV4 排除同期政策的影响

	(1) 国企改革	(2) 高校扩招	(3) 加入 WTO	(4) 户籍制度放松
<i>Intensity</i> × <i>Post</i>	0.0459*** (0.0041)	0.0424*** (0.0040)	0.0373*** (0.0034)	0.0431*** (0.0049)
控制变量	Y	Y	Y	Y
城市固定效应	Y	Y	Y	Y
年份固定效应	Y	Y	Y	Y
R^2	0.850	0.852	0.796	0.837
N	1375	1323	1323	1183

(3) 忽略多次移民的影响。我们在此解释忽略多次移民对计算劳动力流动率的准确性影响较小的原因。根据 Imbert *et al.* (2022) 的讨论, 当识别多次移民时, 需要知道该个体在多个城市中转的信息。然而, 在 2005 年的人口抽样调查数据中并没有这么丰富的信息。事实上, 在 2005 年的人口抽样调查数据中, 仅包含受调查者在调查前一年与调查前五年的个人居住地信息, 且这两个问题只包含省份级别的信息。因此, Imbert *et al.* (2022) 仅能使用这部分信息采取折衷方案估计多次迁移的劳动力, 即他们定义当个体同时满足三个条件时, 可被认定为多次迁移的劳动力, 三个条件分别是: (1) 受调查者在调查前五年居住在户籍地所在省份; (2) 该调查者在调查时居住在其他省份; (3) 该调查者在一年前居住在另外的省份。

基于上述定义, Imbert *et al.* (2022) 发现多次移民的数量占总劳动力流动人数的比重较低 (1%-2%), 且主要集中在距离首次迁移后的第一年内, 这表明即便忽略多次流动的劳动力, 对以城市-年份为维度的劳动力流动情况的测量误差也较小。而如果选择不忽略这一问题而直接对多次流动的劳动力规模进行测度, 由于无法从抽样调查的问卷中获知多次迁徙者在中间年份的城市信息, 因此, 需要对这一缺失进行较强的假设, 产生更不可信的测算结果。基于以上原因, Imbert *et al.* (2022) 在计算劳动力流动率时选择不对多次流动的劳动力进行调整。我们也遵循他们的做法, 受限于数据限制, 不再对多次流动的劳动力进行矫正。

(4) 处理回迁的问题。本文定义劳动力流动的方法忽略了回迁劳动力流动, 从而造成了对劳动力流动数量的低估, 这可能会对估计系数的方向产生影响。为了讨论这一问题, 我们在此借鉴了 Imbert *et al.* (2022) 的方法, 以对 1999-2004 年的劳动力流动数量进行回迁调整。

2005 年的人口抽样调查数据汇报了每个受调查者当前居住地城市、五年前/一年前居住省份以及其户籍所在地省份。Imbert *et al.* (2022) 表明可以通过这几个变量来定义回迁劳动力。具体而言, 假定每个劳动力在其迁移的目的地的“生存”服从泊松分布, 劳动力流动的类型 (目的地省份与户口的类型) 服从指数分布, 即 $h(f) = \lambda_p e^{-\lambda_p f}$ 。其中, f 为劳动力流

动的类型, $1/\lambda_p$ 为目的 p 省的劳动力回迁率, 我们根据 2005 年人口抽样调查数据的受调查者户籍地信息、调查前五年居住地与当前居住地信息, 计算了 1999-2004 年每个省份每年的劳动力回迁率, 然后使用观测到的劳动力流动分布来推断真实的劳动力流动情况, 从而将回迁劳动力纳入对劳动力流动率的测算中, 获取校准后的劳动力流动率。

具体过程如下, T 为 2004 年, k 为离开户籍地的时间, 我们令 $M_{T,k}$ 为 $t=T-k$ 期到达当前居住地的劳动力流动数量, 那么经过校准后, 实际劳动力流动数量为 $\left[(\lambda_p + k)^2 / \lambda_p^2 \right] M_{T,k}$ 。

表 IV5 使用经过回迁校准后的劳动力流动率作为因变量, 估计结果如下:

表 IV5 回迁调整

	(1)	(2)
<i>Intensity</i> × <i>Post</i>	0.0481*** (0.0046)	0.0388*** (0.0035)
控制变量	N	Y
城市固定效应	Y	Y
年份固定效应	Y	Y
R^2	0.863	0.874
N	1404	1393

由第 (1) 列可知, 城市在冲击前的互联网基础设施每上升 10%, 劳动力流动率提高

0.48%，第(2)列加入控制变量后，回归系数依旧显著，城市冲击前的互联网基础设施每上升10%，劳动力流动率提高0.39%，这一结果小于基准回归的0.46%，表明忽略回迁移民会导致基准回归系数的高估，但纳入回迁移民后，本文的实证结论仍然成立。

(5) 拼接不同轮次的人口抽样调查数据。为了获得更长的劳动力流动情况的观察时段，需要对不同轮次的人口抽样调查数据进行拼接。考虑到不同轮次的人口抽样调查数据存在回忆偏差^①以及抽样标准不统一等问题，直接在端点年份拼接会造成“跳跃”的断点现象。为了缓解这一问题对估计的影响，我们采用“校准”的方法，以缓解样本抽样问题导致的同一城市不同年份之间劳动力流动人数以及调查人数不可比的问题。在此以2000年的人口抽样调查数据和2005年的人口抽样调查数据拼接时所产生的跳跃为例。

具体来说，我们首先根据这两轮人口抽样调查数据获取其在拼接处各自的劳动力流动总人数，然后将这两个劳动力流动人数取比值作为“校准”的系数，随后将1995-1998年每个城市逐年的劳动力流动人数乘以这一系数（后文称为系数A）。经过上述处理后，通过两次人口抽样调查数据构造的劳动力流动率将不会再发生跳跃。从形式上来看，我们使用如下公式对1995-1998年城市*i*在时间*t*的劳动力流动人数进行校准：

$$\text{校准后劳动力流动人数}_{i,t} = \frac{\text{劳动力流动人数}_{i,t}}{\text{Census 2000回溯1999年劳动力流动总人数}} \times \text{Census 2005回溯1999年劳动力流动总人数}$$

随后，我们也使用类似的方法拼接了利用2005年人口抽样调查构造的1999-2004年劳动力流动率面板数据与通过2010年人口抽样调查构造的2004-2009年劳动力流动率面板数据。并根据如下公式校准（对应系数在后文称为系数B）：

$$\text{校准后劳动力流动人数}_{i,t} = \frac{\text{劳动力流动人数}_{i,t}}{\text{Census 2010回溯2004年劳动力流动总人数}} \times \text{Census 2005回溯2004年劳动力流动总人数}$$

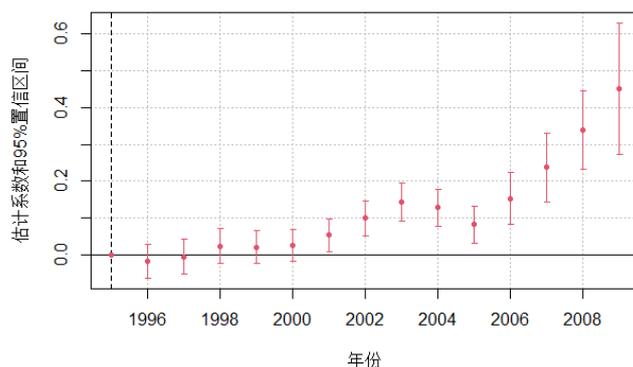
在上述调整后，最终得到了1995-2009时段的面板数据，这种方法下获得的平均劳动力流动率将不会存在上述“跳跃”的问题。但是值得注意的是，为了延长数据需要校准的过程去弥补跳跃，而校准是有隐含假设的。即为了构造1995-2004的区间需要假设存在一个合理的校准系数A，而为了将2010的人口抽样调查数据纳入，又需要假设存在一个合理的校准系数B，拼接的时间段越长，所需要纳入的人口抽样调查数据越多，所需要强加的假设也越多。

我们在此将上述延长的数据区间进行回归以作为稳健性检验，来尽可能缓解事前数据区间过少的问题。拓展时间区间的回归结果如表IV6所示，事件分析的结果见图IV1。两处的实证结果均表明，即便扩展数据区间，估计结果也仍然正向显著，并且未拒绝事前平行趋势成立的原假设。上述结果为本文的讨论增强了稳健性。

表IV6 拓展时间区间

	(1)	(2)	(3)
	1995-2009	1995-2004	1999-2009
<i>Intensity</i> × <i>Post</i>	0.1672*** (0.0269)	0.0893*** (0.0097)	0.1331*** (0.0245)
城市固定效应	Y	Y	Y
年份固定效应	Y	Y	Y
<i>R</i> ²	0.703	0.721	0.703
N	3369	2257	2465

^① 抽样调查问卷采用的是“回溯”的询问方法，即询问个体“在几年前迁移到本地”。由于回忆的越远越可能出现偏差，并且在不同轮次的调查中发生的程度也不一样。因此，这种异质性偏差导致不同的抽样调查数据计算出的劳动力流动率在拼接点处出现“跳跃”。



图IV1 1995-2009年劳动力流动率的事件分析估计结果(校准后)

(5) 处理城市基期差异存在异质性时间趋势的问题。鉴于不同城市之间在基期可能存在系统性差异, 这些差异在未来可能会出现随时间变化的异质性趋势, 导致基准回归的估计失效。为了缓解这一问题, 我们在此借鉴 Egger *et al.* (2022) 的思路, 选择了可能会造成各城市拥有不同固定电话的前置变量(各地的经济发展水平、邮局数、受教育程度等), 随后取这些变量在 1999 年的值同年份固定效应进行交互, 以控制不同城市之间的基期差异在未来的任何非参数形式的变异。估计结果见表 IV7, 其中, 不论是否增加各地在 1999 年的基期变量同时间的趋势, DID 项的估计系数均显著为正, 表明基准回归的结论仍然保持。

表IV7 缓解冲击的非随机性

	(1)	(2)
<i>Intensity</i> × <i>Post</i>	0.0472*** (0.0077)	0.0374*** (0.0053)
1999 年基期变量的时间趋势	Y	Y
控制变量	N	Y
城市固定效应	Y	Y
年份固定效应	Y	Y
R^2	0.8326	0.8480
N	1404	1393

(6) 其他稳健性检验。本文还进行了一系列其他稳健性检验。首先, 本文效仿 Chen and Liu (2022) 将 *Intensity* 更换为离散的测度, 即根据 1999 年全国人均固定电话密度的中位数对城市进行分组, 表 IV8 的第 (1) 列结果表明结果依旧稳健。其次, 前文估计是将 2005 年的人口普查回溯至 1999-2004 年, 这其中隐含的假设是将离开 1-2 年的人口归类为 2004 年的移民。我们也尝试放松这一设定, 即将 2005 年的人口抽样调查数据回溯至 2000-2005 年, 第 (2) 列的估计结果表明, 这一变化无损本文估计结果的稳健性。最后, 移民率是一个比例指标, 其增加可能来自本地移民的流出 (Wilson, 2022)。为避免这一问题, 我们直接将 (1) 式的因变量替换为各城市历年的移民总人数 (取对数), 估计结果见第 (3) 列。综上, 无论哪种设定, 结论均保持稳健。

表IV8 其他稳健性检验

	(1) 移民率	(2) 移民率	(3) 移民人数
<i>Intensity</i> × <i>Post</i>	0.0045*** (0.0008)	0.0524*** (0.0065)	0.6003** (0.2583)
控制变量	Y	Y	Y
城市固定效应	Y	Y	Y
年份固定效应	Y	Y	Y
R^2	0.842	0.815	0.933
N	1586	1562	1393

附录 V 异质性分析

从性别构成来看。由表 V1 可知,男性和女性劳动力的迁移决定都受到信息摩擦下降的影响。而从移民群体的技能水平分类来看。第(3)-(7)列的实证结果表明宽带出现后对具有初中学历的人的系数最大,其次是具有高中学历的人,而对学历在小学及以下或者学历为大学及以上的劳动力影响较小。对此可能的解释是,高技能的移民在没有宽带互联网的情况下便对工作市场的信息更加熟悉,故而对新信息的出现并不意外。而对仅具有小学文化及以下文化的劳动力而言,可能较难在城市中搜寻到合适的工作机会,故而宽带出现对其迁移决策的影响相对较小。

表V1 依据性别与学历划分

	(1) 女性	(2) 男性	(3) 未上学	(4) 小学	(5) 初中	(6) 高中	(7) 大学
<i>Intensity</i> × <i>Post</i>	0.0312*** (0.0027)	0.0251*** (0.0024)	0.0005** (0.0002)	0.0038*** (0.0008)	0.0258*** (0.0027)	0.0113*** (0.0012)	0.0045*** (0.0006)
控制变量	Y	Y	Y	Y	Y	Y	Y
城市固定效应	Y	Y	Y	Y	Y	Y	Y
年份固定效应	Y	Y	Y	Y	Y	Y	Y
R^2	0.814	0.828	0.589	0.782	0.818	0.867	0.807
N	1404	1404	1393	1393	1393	1393	1393

附录 VI 理论分析

在经典的迁移选择模型中，个体会在当其迁入地 (d) 的终生效用减去迁移的固定成本超过其在原籍地 (o) 的效用时选择迁移 (Jin and Zhang, 2023)。其中，现有文献提到了诸多能够影响个体迁移固定成本的因素，例如户籍制度、个人的社会网络关系、迁出地的交通条件等，而较少讨论迁入地的信息丰裕度也会影响个人的迁移选择。事实上，迁入地信息摩擦对个体迁徙的影响非常显著 (Schmutz and Sidibé, 2019; Porcher, 2022)，本文在此借鉴 Wilson (2022) 的框架，将信息因素纳入到人口流动的决策函数中。

当信息摩擦为零时，个体劳动力对目的地的工作信息及收入信息完全悉知，此时，只需要迁入地的效用 $u(y_d(t))$ 减去其迁移成本 c_{od} 大于其在迁出地的效用 $u(y_o(t))$ ，个体便会选择迁移。然而，当我们引入信息摩擦时，个体是否选择迁移的目标函数将从确定的效用函数转化为期望效用函数，这个期望值取决于个体对目的地信息集的先验分布。此时，个体在目的地的效用变为 $\sum_{t=0}^T \beta^t Eu(y_d(t))$ ，其中 β 为折现因子。类似地，当式 (3) 大于 0 时，个体会选择迁移，小于 0 则选择留在原籍，等于 0 时，个体无论迁移或留在原籍，在效用上都无差异。

$$\sum_{t=0}^T \beta^t \{Eu(y_d(t)) - u(y_o(t))\} - c_{od} \geq 0 \quad (1)$$

根据无差异条件，可由式 (3) 得到个体迁出的均衡成本 c_{od}^* ：

$$c_{od}^* = \sum_{t=0}^T \beta^t \{Eu(y_d(t)) - u(y_o(t))\} \quad (2)$$

不难看出， c_{od}^* 同个体对目的地收益的预期均值有关，预期均值越大，个体越可能选择迁移。假设原籍地的个体对目的地的预期收益 $y_d(t)$ 服从正态分布 $N(\mu_d, \sigma^2)$ 。此时个体迁出的均衡成本 c_{od}^* 会受到劳动力对目的地收入均值 μ_d 和收入方差 σ^2 的影响。当假设个体为风险回避者时 ($u' > 0$ 以及 $u'' < 0$)，容易得到：

$$\frac{\partial c_{od}^*}{\partial \mu_d} = \sum_{t=0}^T \beta^t \frac{\partial Eu(y_d(t))}{\partial \mu_d} > 0 \quad (3)$$

$$\frac{\partial c_{od}^*}{\partial \sigma^2} = \sum_{t=0}^T \beta^t \frac{\partial Eu(y_d(t))}{\partial \sigma^2} \leq 0 \quad (4)$$

式 (5) 和式 (6) 表明个体对目的地的收入均值预期越高，越可能选择迁移，而预期收益的方差越大，则表明其对迁徙收益的不确定性越高，则更倾向于留在原籍地。简而言之，预期收益均值的增加会增加个体迁出可接受的均衡成本 c_{od}^* ，而不确定性的增加则会降低该成本。更高的 c_{od}^* 表明该个体劳动力迁移的意愿更强烈。因此，如果个体在事前对迁移预估的收益较低或者对迁徙的风险偏好较为保守的话，则不太可能迁移。

接下来，本文在个体对迁入地的收益预期和风险中引入信息冲击。假设个体的信息集

为 $G_0 = (g_1, g_2, \dots, g_{n_0})$, 该信息集的变动同宽带互联网的出现相关。Kaplan and Schulhofer-Wohl (2017) 提出了一个结构模型指出, 信息能够帮助人们了解不同城市的基础设施和工作机会等信息。信息的更新重塑了人们对劳动力市场的预期, 进而影响其迁移决策, 而人们可接触的信息媒介差异, 可解释不同地区迁徙决策的巨大差别 (Farré and Fasani, 2013)。

在不完全信息的情形下, 个体对影响 $y_d(t)$ 分布的参数向量并不确信, 此时接收关于这些参数的额外信息有助于个体更新他们对 μ_d 和 σ^2 的信念, 进而影响其迁徙决策。在宽带互联网出现的情况下, 这是一个合理的情况。前文背景部分已经提到, 在宽带互联网出现之前, 纸质媒介、职业介绍所或以企业为主来分发工作机会信息的诸多渠道均未联网, 此时假设个体对于外地的信息知之甚少是合理的。为了简化分析, 本文假设预期收益 $y_d(t) = \mu_d + \varepsilon_d$, 给定个体事前对某地预期收益和预期风险的先验信息, 当宽带互联网出现后, 新信息将改变个体对 μ_d 和 σ^2 的信念。假设这一过程遵循贝叶斯方程, 具体如下:

$$P(\mu_d | G_0) = \frac{P(G_0 | \mu_d)P(\mu_d)}{P(G_0)} \quad (5)$$

显然, $P(\mu_d | G_0) \propto P(G_0 | \mu_d)P(\mu_d)$, 个体对 μ_d 的后验分布为:

$$P(\mu_d | G_0, \sigma^2) = N\left(\mu_d \left| \frac{\sigma^2}{n_0\lambda_0 + \sigma^2} \mu_{d0} + \frac{n_0\lambda_0}{n_0\lambda_0 + \sigma^2} \bar{x}, \frac{\sigma^2\lambda_0}{n_0\lambda_0 + \sigma^2} \right.\right) \quad (6)$$

其中, \bar{x} 为个体观察到的目的地的收入的样本均值。个体 μ_d 的后验分布为:

$$\mu_{d1} = \frac{\sigma^2}{n_0\lambda_0 + \sigma^2} \mu_{d0} + \frac{n_0\lambda_0}{n_0\lambda_0 + \sigma^2} \bar{x} \quad (7)$$

μ_d 的后验方差为:

$$\lambda_1 = \frac{\sigma^2\lambda_0}{n_0\lambda_0 + \sigma^2} \quad (8)$$

进一步地, 假设 μ_d 和 ε_d 互相独立, 并且 $\varepsilon_d \sim N(0, \sigma^2)$, 此时:

$$\begin{aligned} P(y | G_0) &= \int P(y | \mu_d)P(\mu_d | G_0) d\mu_d \\ &= \int N(y | \mu_d, \sigma^2) N(\mu_d | \mu_{d1}, \lambda_1) d\mu_d \\ &= N(y | \mu_{d1}, \lambda_1 + \sigma^2) \end{aligned} \quad (9)$$

式 (11) 表明个体劳动力对 $y_d(t)$ 的估计同其事前的信念和新增信息相关。如果个体劳动力在事前对目的地的期望收益较高, 而新信息中揭示的回报却低于其事前期望, 那么该个体会降低迁移的概率。反之, 其迁移的概率提高。因此, 仅从式 (11) 来看, 新信息究竟是促进还是抑制个体迁徙并无定论。例如在印度尼西亚, 在当地提供更公允信息的电视台出现后, 更真实的信息通过矫正过往人们对迁移过度乐观的期望, 反而降低了国内人口的流动 (Farré and Fasani, 2013)。但是鉴于中国劳动力市场的城乡二元分割特性, 农村部门的工资远远小于城市部门的工资, 例如, Imbert *et al.* (2022) 表明移民在城市的月收入

几乎是农村留守劳动力的两倍。因此,信息中揭示的回报低于其事前期望的可能性不大^①。新信息的出现应可预期能够显著地促进劳动力流动至工作机会地,加总到宏观层面表现为劳动力的流向与规模被信息变化重新塑造的过程。据此,引出本文的第一个假说:

假说 1: 宽带的出现能够通过增大迁入地的信息集,即降低迁入地的信息摩擦,来促进劳动力流入。

上述讨论仅聚焦于信息对迁移决策的影响,并未详述信息如何从城市传导至农村。考虑到 2000 年左右农村人口不太可能安装窄带或宽带互联网,且仅 4% 的互联网使用者通过网吧上网(CNNIC)。这些背景信息暗示农村人口通过互联网直接获取信息的可能性几乎为零。换言之,尽管宽带互联网的出现极大地推动了工作机会信息的流通,但它并不是主导信息传递的媒介。那么,工作信息如何从城市流动到农村呢? Porcher (2022) 表明信息会通过正式或非正式的网络传递,其中,非正式的信息传递网络遍布各类社会关系。例如,工人会利用从同事那里获得的信息(Dustmann and Glitz, 2015; Glitz and Vejlin, 2021; Caldwell and Harmon, 2019; Hjort and Poulson, 2019)、家庭成员(Kramarz and Skans, 2014)、邻居(Schmutte, 2015)、同学(Zimmerman, 2019)。这类建立在血缘、地理位置、共享的教育和工作经历等人际强关系基础上的社群网络不仅能实现求职者之间的信息低成本共享,还可以通过推荐求职者来缓解雇主和求职者间的信息不对称问题(左翔和李辉文, 2017)。以中国为例,有证据显示劳动力的就业决策受到移民社交网络的影响(郭云南和姚洋, 2013; 陈斌开和陈思宇, 2018),依靠熟人关系推荐或由熟人提供信息找工作的比例高达 60% (左翔和李辉文, 2017)。因此,本文推断社会关系网络中的信息分享可能是传递因宽带互联网出现而涌现的新信息流向农村的重要渠道。

据此,本文将 $y_d(t) = \mu_d + \varepsilon_d$ 修改为 $y_d(t) = f(X_i, X_d, News(S_{od})) + \varepsilon_d$, 表明个体对迁入地 d 的收益同其个人特质 X_i , 迁入地的特征 X_d 以及所获知的信息有关。其中,个体获知的信息集大小同目的地与原籍地的信息分享网络规模相关,用 S_{od} 表示,即 $News_{od} =$

$News(S_{od})$ 。例如,如果目的地 d 的早期移民来此找第一份工作时通过熟人介绍获取工作机会的比例较高,本文就认为迁入地 d 同原籍地 o 的信息分享网络联系较为紧密,尽管在人口普查数据中并不显示每个移民具体的原籍地,但是可以知道迁入地 d 中从农村到城市的初期移民通过关系获取工作机会的比例,这一加总指标可以理解为在各个原籍地上 o 上进行加权。假设 $\frac{\partial News(s_{od})}{\partial s_{od}} > 0$, 即信息分享网络越大,个体所能获知的信息也越多。由于:

$$\frac{\partial c_{od}^*}{\partial News_{od}} = \sum_{t=0}^T \beta^t \frac{\partial Eu(y_d(t))}{\partial f(X_i, X_{od}, News_{od})} \frac{\partial f(X_i, X_{od}, News_{od})}{\partial News_{od}} \quad (10)$$

此时易知 $\frac{\partial^2 c_{od}^*}{\partial News_{od} \partial s_{od}} > 0$, 即我们应该发现当某地信息分享网络越发达时,该地移入的人口

在宽带互联网出现后增加的数量也越显著,这一表达式将信息流向与劳动力流向之间的关系联系起来。综上,本文引出第二个假说:

假说 2: 迁入地城市的信息分享网络会影响个体的迁移选择,一个城市的信息分享网络越发达,个体越有可能知悉迁入地城市的信息变动,进而选择迁移。

^① Borjas and Bratsberg (1996) 发现,基于关于目的地工作机会的错误信息会导致发生迁徙行为的移民回迁。因而即便我们忽略了错误信息的存在,也仅仅会导致我们低估信息摩擦下降对移民流入的影响。

附录 VII 对双边宗族网络异质性的讨论

究竟是来源地还是目的地的宗族网络在促进人口流入目的地城市呢？为了讨论这一问题，我们测度了双边宗族网络的规模。具体来说，对于城市对 $A \rightarrow B$ ，容易得到来源地 A 的人均族谱密度 ($Zongzu_origin$) 和目的地 B 的人均族谱密度 ($Zongzu_des$)，随后将这两项分别与 DID 项交互，便可以查看究竟是来源地 A 的宗族网络在促进目的地人口流入还是目的地 B 的宗族网络在促进人口流入。值得一提的是，在加入目的地 B 的人均族谱密度与 DID 的交互项时，如果同时控制到达城市与年份的联合固定效应，会导致变异被固定效应吸收的太多而使得该系数的标准差无法估计，因此，我们在包含目的地 B 的人均族谱密度与 DID 的交互项时，转而控制到达城市与年份线性项的交互效应，特此说明。

回归结果如下表所示，可以看到在“城市对”的实证设定中，三列的估计结果均指向是来源地的宗族网络规模在目的地的人口移入中起到了正向调节作用，这一证据支持了应该关注来源地 A 的宗族网络规模，而非目的地 B 的宗族网络规模的观点。

表 VII 1 双边宗族网络的异质性作用

	(1)	(2)	(3)
		移民率	
$Intensity \times Post$	-0.0129*** (0.0029)	-0.0011 (0.0162)	-0.0141 (0.0167)
$Intensity \times Post \times Zongzu_origin$	0.0046*** (0.0010)		0.0047*** (0.0009)
$Intensity \times Post \times Zongzu_des$		0.0004 (0.0032)	0.0006 (0.0167)
出发城市×年份	Y	Y	Y
到达城市×年份	Y	N	N
到达城市×年份(Linear)	N	Y	Y
城市对固定效应	Y	Y	Y
年份固定效应	Y	Y	Y
R^2	0.5912	0.5956	0.5887
N	70632	71226	70632

附录 VIII 移民网络的影响群体

正文已表明, 过去的移民轨迹会影响到当下的劳动力流动轨迹, 那么移民的社会网络在宽带出现后主要影响的是哪一部分群体呢? 表 VIII1 的估计结果显示, 社会网络的影响主要发生在年轻的具有中低学历的群体中。

表 VIII1 信息流向与劳动力流向

	(1) 15-30	(2) 31-45	(3) 46-60	(4) 未上学	(5) 小学	(6) 初高中	(7) 大学
<i>Intensity</i> × <i>Post</i>	0.0198*** (0.0051)	0.0100*** (0.0028)	0.0022** (0.0011)	0.0012*** (0.0004)	0.0034*** (0.0013)	0.0144*** (0.0040)	0.0047*** (0.0012)
<i>Intensity</i> × <i>Post</i> × <i>Connection</i>	0.1019*** (0.0226)	-0.0071 (0.0151)	0.0007 (0.0047)	-0.0041** (0.0017)	0.0032 (0.0066)	0.0756*** (0.0186)	-0.0008 (0.0059)
控制变量	Y	Y	Y	Y	Y	Y	Y
城市固定效应	Y	Y	Y	Y	Y	Y	Y
年份固定效应	Y	Y	Y	Y	Y	Y	Y
R^2	0.822	0.847	0.691	0.591	0.782	0.820	0.808
N	1387	1387	1387	1387	1387	1387	1387

附录 IX 排除替代性假说

本文考虑了三个替代性假说：

1. 是工作岗位的涌现还是工作信息的涌现

Hjort and Poulson (2019) 表明更快速的互联网有助于带来更多的就业机会，也即所谓的工作创造。而城市就业机会的增加同样可能导致人口从农村涌向城市。为讨论这一替代性假说，本文采用两类方式来控制互联网的就业创造效应。第一种是控制各地新增企业的数量及其增速，如此处理的原因是因为新企业（尤其是中小企业）是工作创造的主要提供者（Adelino *et al.*, 2017）。基于此，我们依据历年中国工商企业注册数据库计算了城市层面新进入企业的数量及其增速，作为控制变量加入基准回归中。如果前文观察到的移民促进作用主要来自类似 Tian (2022) 提到的工作机会涌现而引发的拉动力，那么在控制新企业的数量及增速后，*Intensity*×*Post* 的估计系数应大幅下滑甚至不显著。然而表 VIII 第 (1) 列的估计结果表明，这一推论并不成立。在第二种方式中，本文则控制了各城市的就业总人数 (*EMP*) 及其未来一期的值 (*EMP_{t+1}*) 作为控制变量，以缓解宽带互联网的就业创造效应^①。其中，增加 *EMP_{t+1}* 是考虑到互联网的就业创造机制可能存在预期效应，提前吸引劳动力从农村流向城市。估计结果见第 (2) 至 (3) 列，不论增加当期 *EMP* 或 *EMP_{t+1}*，两列的估计系数均显著为正，且系数大小相比基准回归变动不大，表明互联网的工作创造效应解释力较小。

2. 是否为互联网的工资诱导效应

Hjort and Poulson (2019) 观察到更快速的互联网能够同时提高低技能劳动力与高技能劳动力的工资。因此，宽带互联网的出现可能对劳动力的流动存在工资诱导效应，即城市的平均工资因宽带互联网提高，进而导致人口从农村涌入城市的速度增加。为了缓解这一担忧，本文从两方面进行讨论，一是将城市统计年鉴获取的职工平均工资作为城市层面平均工资的代理变量（取对数），并将上述变量加入基准回归。二是从历年的《中国劳动统计年鉴》中获取了省级层面的逐年各行业的平均工资情况（包括采掘业，制造业，电力煤气供应业，批发和零售贸易餐饮业等行业），作为控制变量，以尽可能地控制工资诱导效应对本文机制的干扰。估计结果见表 IX 1 的第 (4) 至 (6) 列。其中，第 (4) 列的估计结果表明，在控制城市层面职工平均工资后，*Intensity*×*Post* 的系数仍然显著；第 (5) 列表明宽带互联网的出现对城市层面的职工平均工资无显著影响；最后，第 (6) 列的估计结果显示在控制省级层面各行业的平均工资后，*Intensity*×*Post* 的系数仍显著为正。上述三列的实证结果一定程度上支持了工资增长的诱导效应对本文机制影响较小的观点。

3. 考虑社会网络基期差异的时变影响。

另一个值得探讨的替代性假说是各地社会网络的异质性趋势在引发劳动力迁移，而宽带互联网的出现减少信息摩擦仅是对这一过程起到正向调节作用^②。在该假说下，社会网络的变动可能对劳动力的流动存在溢出效应 (*spillover effect*) 和集群效应 (*cluster effect*)。为了讨论该假说，我们采取了两种方法来控制这一效应。一是使用 1990 年各地移民占当地总人口的比例作为衡量各地在基期的社会网络规模，随后用该变量同年份固定效应交互，

^① 感谢审稿专家的宝贵建议。

^② 感谢审稿专家的宝贵建议。

以控制各地社会网络的随时间变化的异质性趋势。二是考虑到主要是来源地的宗族网络规模在起到信息分享的作用,我们还采取加权法来恢复目的地(A)的移民在其各自来源地(例如,B,C,D等地)的加权宗族网络(使用明朝的人均族谱数来测度)。具体算法为:首先测算A地在2000年这个时点上过去5年累计移民中来自B,C,D等地的人数,随后将这一向量相对总移民人数的值作为权重同B,C,D等地的人均族谱数做向量乘法,加总得到目的地A中移民的加权来源地宗族网络规模^①,随后将该变量同年份固定效应的交互项加入至基准回归中。第(7)(8)列显示,无论哪一种处理方法,在增加各地基期社会网络差异随时间变动的异质性趋势后,回归结果均显著为正,且DID项的估计系数相比基准回归变化较小,该结果一定程度上缓解了对上述替代性假说的担忧。

表IX1 排除替代性假说

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	控制 新增 企业	控制 EMP	控制 EMP _{t+1}	劳动力 流动率	平均 工资	劳动力 流动率	1990年 移民占比	加权来源 地宗族 比例
<i>Intensity</i> × <i>Post</i>	0.0413*** (0.0063)	0.0451*** (0.0039)	0.0395*** (0.0042)	0.0458*** (0.0040)	-0.1658 (0.2042)	0.0374*** (0.0044)	0.0428*** (0.0042)	0.0453** (0.0125)
职工平均工资	N	N	N	Y	N	N	N	N
各行平均工资	N	N	N	N	N	Y	N	N
控制变量	Y	Y	Y	Y	Y	Y	Y	Y
城市固定效应	Y	Y	Y	Y	Y	Y	Y	Y
年份固定效应	Y	Y	Y	Y	Y	Y	Y	Y
R ²	0.832	0.821	0.808	0.818	0.942	0.843	0.866	0.823
N	1387	1393	1166	1393	1393	1393	500	1393

参考文献

- [1] Adelino, M., S. Ma, and D. Robinson, "Firm Age, Investment Opportunities, and Job Creation". *The Journal of Finance*, 2017, 72(3), 999-1038.
- [2] Caldwell, S., and N. Harmon, "Outside Options, Bargaining, and Wages: Evidence from Coworker Networks". *Working paper*, 2019, 203-207.
- [3] Chen, L., and W. Liu, "The Effect of Internet Access on Body Weight: Evidence from China". *Journal of Health Economics*, 2022, 85, 102670.
- [4] Chen, S., P. Olivam, and P. Zhang, "The Effect of Air Pollution on Migration: Evidence from China". *Journal of Development Economics*, 2022, 156, 102833.
- [5] 陈斌开、陈思宇, "流动的社会资本——传统宗族文化是否影响移民就业?", 《经济研究》, 2018年第3期, 第35页—49页。
- [6] Démurger, S., M. Gurgand, S. Li, and X. Yue, "Migrants as Second-Class Workers in Urban China? A Decomposition Analysis". *Journal of Comparative Economics*, 2009, 37(4), 610-628.
- [7] Dustmann, C., and A. Glitz, "How Do Industries and Firms Respond to Changes in Local Labor Supply?" *Journal of Labor Economics*, 2015, 33(3), 711-750.
- [8] Egger, P. H., G. Loumeau, and N. Loumeau, "China's Dazzling Transport-Infrastructure Growth: Measurement and Effects". *Journal of International Economics*, 2023, 142, 103734.
- [9] Farré, L., and F. Fasani, "Media Exposure and Internal Migration—Evidence from Indonesia". *Journal of Development Economics*, 2013, 102, 48-61.
- [10] Glitz, A., and R. Vejlin, "Learning through Coworker Referrals". *Review of Economic Dynamics*, 2021, 42, 37-71.
- [11] Guo, R., J. Zhang, and M. Zhou, "The Demography of the Great Migration in China." *Journal of Development Economics*, 2024, 167, 103235.

^① 感谢审稿专家的宝贵建议。

- [12] 郭云南、姚洋,“宗族网络与农村劳动力流动”,《管理世界》,2013 第 3 期,第 69—81 页。
- [13] Hjort, J. and J. Poulsen, “The Arrival of Fast Internet and Employment in Africa.” *American Economic Review*, 2019, 109(3), 1032-1079.
- [14] Imbert, C., M. Seror, Y. Zhang, and Y. Zylberberg, “Migrants and Firms: Evidence from China”. *American Economic Review*, 2022, 112(6), 1885-1914.
- [15] Jin, Z. and J. Zhang, “Access to Local Citizenship and Internal Migration in a Developing Country: Evidence from a Hukou Reform in China”. *Journal of Comparative Economics*, 2023, 51(1), 181-215.
- [16] Kaplan, G., and S. Schulhofer-Wohl, “Inflation at the Household Level”. *Journal of Monetary Economics*, 2017, 91, 19-38.
- [17] Kramarz, F., and O. N. Skans, “When Strong Ties are Strong: Networks and Youth Labour Market Entry”. *Review of Economic Studies*, 2014, 81(3), 1164-1200.
- [18] Manacorda, M., and A. Tesei, “Liberation Technology: Mobile Phones and Political Mobilization in Africa”. *Econometrica*, 2020, 88(2), 533-567.
- [19] 马双、赵文博,“方言多样性与流动人口收入——基于 CHFS 的实证研究”,《经济学(季刊)》,2019 年第 1 期,第 393 页—414 页。
- [20] Oster, E, “Unobservable Selection and Coefficient Stability: Theory and Evidence.” *Journal of Business & Economic Statistics*, 2019, 37(2), 187-204.
- [21] Porcher, C. “Migration with Costly Information”. *Mimeo*, 2022, 1(3).
- [22] Satyanath, S., N. Voigtländer, and H. J. Voth, “Bowling for Fascism: Social Capital and The Rise of The Nazi Party”, *Journal of Political Economy*, 2017, 125(2), 478-526.
- [23] Schmutte, I. M. “Job Referral Networks and the Determination of Earnings in Local Labor Markets”. *Journal of Labor Economics*, 2015, 33(1), 1-32.
- [24] Schmutz, B., and M. Sidibé, “Frictional Labour Mobility”. *The Review of Economic Studies*, 2019, 86(4), 1779-1826.
- [25] Tian, Y., “International Trade Liberalization and Domestic Institutional Reform: Effects of WTO Accession on Chinese Internal Migration Policy”, *Review of Economics and Statistics*, 2022, 1-45.
- [26] Wang, F., C. Milner, and J. Scheffel, “Labour Market Reform and Firm-level Employment Adjustment: Evidence from the Hukou Reform in China”. *Journal of Development Economics*, 2021, 149, 102584.
- [27] Wilson, R. “Moving to Economic Opportunity the Migration Response to the Fracking Boom”. *Journal of Human Resources*, 2022, 57(3), 918-955.
- [28] Zhong, L. “Internal Migration and Extended Families in China”. *Working Paper*, 2018.
- [29] Zimmerman, S. D. “Elite Colleges and Upward Mobility to Top Jobs and Top Incomes”. *American Economic Review*, 2019, 109(1), 1-47.
- [30] 左翔、李辉文,“市场化进程中的劳动者社群网络与企业效率”,《经济研究》,2017 年第 3 期,第 106—121 页。

注：该附录是期刊所发表论文的组成部分，同样视为作者公开发表的内容。如研究中使用该附录中的内容，请务必在研究成果上注明附录下载出处。