

上海新金融研究院工作论文系列

No. SFIWP0049

方言，普通话与中国劳动力区域流动

李 秦 孟岭生

2015年5月4日

说明：上海新金融研究院（Shanghai Finance Institute, SFI）是一家非官方、非营利性的独立智库，致力于新金融领域的政策研究。研究院成立于2011年7月14日，由中国金融四十人论坛（China Finance 40 Forum, CF40）举办，与上海市黄浦区人民政府战略合作。

本工作论文是上海新金融研究院研究人员在工作期间形成的、尚未公开发表的研究成果，文中观点仅代表作者本人，不代表本研究院。未经书面同意，谢绝任何形式的转载和复制。

方言，普通话与中国劳动力区域流动¹

李秦² 孟岭生³

摘要：语言是沟通的媒介，也是地方文化的一种体现，劳动力在选择区域流动时除了经济原因外，语言沟通和相同的文化背景也是其重要的考虑因素。在过去关于中国劳动力区域转移的研究中，这两个因素往往被忽略了。用普通话普及率和方言作为语言沟通成本及共同文化背景的代理变量，在控制了经济因素和地理因素后，语言沟通和地区文化对中国劳动力区域流动造成的影响。结果表明在中国国内的确存在语言和文化障碍，劳动力更倾向于流动到普通话沟通障碍小，拥有共同文化背景的地方工作。此外方言差异和普通话水平对劳动力区域间流动的影响会随着性别、户口、教育水平和从事的产业的不同而不同，间接论证了推广普通话的经济意义。

关键词：方言；普通话；劳动力流动；引力模型

1. 引言

语言与经济现象息息相关。一方面语言是经济交流的载体，人们不仅用语言来沟通，同时用语言记录经济合同和交易契约。当两个经济主体使用的语言不一致时可能会带来非常高的交易成本，这些交易成本不仅包括翻译人员的工资等直接成本，还包括由于翻译不准确而导致交易双方误解产生的交易成本，第三方的参与而可能造成的泄密风险，以及因为语言不通而可能遭到不公正的待遇等间接成本（Deneckere and Severinov，2001）。另一方面对于劳动力来说语言也是重要的人力资本（Rubinstein，2000）。熟悉当地的语言能够提高其生产效率，更加容易找到工作、更方便生活和积累社会关系。有很多企业在招聘的时候会把中员工的语言技能作为一项重要考核指标，有时候甚至会把语种细分到方言。例如在高盛集团（Goldman Sachs）的职位申请页面上，就把中文分为客家话（Chinese Haka），粤语（Chinese Cantonese）、普通话（Chinese Mandarin）和其他方言（Chinese other）。语言也是地方文化的重要组成部分，能够反映语言使用者的特质，比如中国人使用的汉语和西方使用的英语反映两种截然不同的文化和习惯。即使在一个国家内部，各个地区也有不同的方言。拥有共同方言常常是地域归属感和认同感的重要部分。在现实生活中，如果一名外来者会使用当地的语言，那么他就更容易被当地居民所认同。

在经济学文献中，语言经济学（Economics of Language）的研究开始于20世纪80年代，当时出现在美国的大量非英语母语的移民引起了经济学家的兴趣，经济学学者开始把语言纳入到人力资本进行分析。早期的研究包括语言对降低交易成本的影响（Carliner，1981；McManus et al. 1983 和Tainer，1988）；语言对工作机会的影响（Kossoudji，

¹ 本文得到国家自然科学基金青年科学基金项目的资助（项目批准编号：71303131），作者在此表示感谢。

² 李秦，清华大学经济管理学院经济学博士研究生，上海新金融研究院青年研究员，E-mail:liq.10@sem.tsinghua.edu.cn。

³ 孟岭生，清华大学经济管理学院经济系副教授，E-mail:menglsh@sem.tsinghua.edu.cn。

1988) 以及语言能力对收入的影响 (Chiswick等人⁴的工作) 等。近些年关注点从语言对个人的影响转向了语言对地区经济的影响。有一大批文献是考察语言对贸易的影响, 例如 Rauch (1999) 和 Rauch and Trindade (2002) 发现移民的社会关系能够帮助他们克服迁入地和迁出地贸易时的语言交流障碍, 从而促进贸易; Melitz (2008) 用引力模型分析了使用同种语言对国家间贸易流量的影响。Falck et al. (2012) 用一个世纪前的方言作为区域文化的代理变量考察语言和文化对移民的影响。

以上这些文献都说明了语言其实是劳动力转移或者移民的影响因素之一, 但是在研究中国劳动力转移的研究中, 并没有把语言纳入考虑范畴。虽然中国在建国初期就已经开始推广普通话, 但是中国是一个方言大国, 全国超过 100 万人口使用的方言就有上千种。

《中国语言文字调查(2000)》的资料显示, 截至 2000 年全国仍然有 86.38% 的人使用汉语方言, 并且这一比例超过使用普通话交流的比例(53.06%)。虽然有部分方言与普通话较为接近, 不影响理解和交流, 但是还有部分方言与普通话差距很大, 会说方言成为工作的必要条件。例如很多到广州工作的外地人都认为, 在广州不会粤语很难找工作⁵; 上海的部分外地公务员被要求参加上海话培训班⁶。

在之前研究中国劳动力转移的文献大多集中在经济因素对劳动力转移影响的问题。例如田新民等(2009)建立了二元理论模型解释了城乡收入差波动对劳动力转移和城市经济效率的影响。蔡昉等(2001, 2003)以及孙文凯等(2011)等学者关注了户籍制度改革对中国农村劳动力流动的影响。邓曲恒和古斯塔夫森(2007)发现与留在农村的农民相比, 成为城市的永久移民的劳动力使自身经济条件得到很大改善。王德文等(2008)发现教育能提高农村迁移劳动力的就业和工资。

但是除了经济因素外, 文化和语言因素对中国劳动力流动是否有影响呢? 例如当北京和广州提供相差不大的就业机会、工资待遇等经济回报的时候, 劳动力是否会因为语言、文化等非经济原因而产生特定的偏好? 本文试图回答这一问题。

我们用普通话的普及水平作为语言沟通障碍的代理变量, 用是否在同一方言区作为地方文化的代理变量, 结果发现在控制了地理距离、经济因素(人均收入、基础建设、社会保障、政府教育投入等)后, 语言障碍和地方文化仍然对劳动力移民有显著的影响: 语言障碍越小(迁入地的普通话水平越高, 迁出地普通话水平越低), 移民数量越多; 地方文化越接近(迁入地和迁出地属于共同方言区), 移民数量越多。这说明除了经济因素外, 语言障碍和地方文化是劳动力选择移民时的重要考虑因素。进一步的研究表明地方文化因

⁴ Chiswick 在关于语言对收入的影响中做了大量的研究, 包括: Chiswick(1991); Chiswick et al.(1995, 1998, 2001, 2003, 2005a, 2005b, 2006, 2007a, 2007b, 2008a, 2008b), Chiswick et al. (2000), Chiswick and Repetto (2001), Dustmann (1994), Dustmann and Soest (2001)等。

⁵ 搜狐网: 《推广普通话有悖论: 会粤语普遍“占便宜”》<http://news.sohu.com/20040916/n222081437.shtml> (2004年09月16日)

⁶ 东方网: 《上海公务员进“沪语班” 基层工作要过语言关》<http://sh.eastday.com/m/20130601/u1a7430339.html> (2013年6月1日)

素对男性移民的影响较大，而语言因素对女性移民的影响更大。相较于第一和第二产业而言，第三产业的从业者受到这两种非经济因素的影响更大。随着教育程度的提高，地方文化和语言沟通对移民的影响逐渐增大，但这两种因素的影响随着年龄的提高却是降低的，即受到语言沟通和地方文化影响最大的人群是从事第三产业工作具有相对较高教育水平的年轻人。对此我们的解释是高教育水平的人从事第三产业的更多（样本中本科及以上学历的从业人员占三次产业就业人员比例分别为：农业 0.72%，工业 2.67%，服务业 7.13%），服务业的特点是“以人为本”——涉及到与人沟通的工作非常多，因此语言和文化因素对从业更为重要。

本文的一个贡献在于突破现有的关于中国劳动力区域流动的研究，把语言和文化纳入人力资本考虑的范畴，地域广阔，文化多元和人情社会是中国很重要的社会特色，在劳动力市场上的体现一个是供给方面，即劳动力的具有不同的禀赋——普通话水平和风土人情；一个是需求方面，即企业的人才需求不同——语言沟通和与本地的融合。当劳动力择业时，除了收入等经济因素外，非经济因素同样重要，有时甚至更为重要。本文第二个贡献是从劳动力自由流动的角度间接论证了推广普通话这一政策具有重要的经济意义。中国早在建国初期就推广普通话并卓有成效，普通话已经相当普及并已成为中国的官方语言，但是由于数据的限制，很难直接衡量推广普通话带来的经济效益，通过本文的研究发现，普通话普及率高的地方移民的数量也相对较多，说明普通话的推广促进了中国区域间劳动力的自由流动，具有重大的经济意义。

本文的结构安排如下。第二部分介绍中国的汉语方言使用情况和推广普通话的相关背景。第三部分详细介绍数据特征并给出一个简单的引力模型作为实证分析的框架。第四部分估计普通话普及水平和方言所代表的地区文化对劳动力转移的影响，并探讨了对不同性别、年龄、教育程度和行业的劳动力转移的影响。第五部分总结全文。

2. 背景介绍

中国的劳动力流动开始于 80 年代农村家庭承包责任制的实行，随着城市改革和对外开放的进程而达到高潮，流动方向上主要是从不发达地区流向发达地区，利益驱动是中国劳动力流动的主要原因。除此以外，中国的一大特色是人情社会，除了考虑经济因素外，文化融合和语言互通也是重要考虑因素。

广阔的国土面积和悠久的文明历史造就了中国方言众多，文化多样。全国超过 100 万人口使用的方言就有上千种。按照《中国语言地图集》⁷的分类，中国汉语方言分为九大方言区：官话区、晋语、赣语、徽语、吴语、湘语、客家话、粤语、平话。官话可细分为：东北官话、北方官话、胶辽官话、北京官话、中原官话、兰银官话、江淮官话、西南

⁷这是中国社会科学院语言研究所和澳洲人文科学院组织方言研究人员在方言调查、资料整理分析的基础上，于 1983 年开始编制，1987 年完成的中国各地区语言使用分布情况的地图集。该地图集是在全面的语言学调查的基础上，按古入声字、古浊声母字的演变规律对汉语方言进行分类，相比其他分类方法更为科学，已成为中国方言学界实际上的学科标准。

官话以及西南官话和土话并用区。方言区的划分与省份划分并不完全重合，而是长时期的山川阻隔，社会分化和历史积淀而成，因此往往存在一个省份存有不同的方言，不同省拥有共同方言的现象。可以说方言是地区文化的重要代表。

方言的存在对语言沟通也造成了影响，不同方言的语音、语汇和语法差异会非常大，会出现不同方言区的人互相听不懂的现象。新中国成立以后，为了统一语言，便于交流，从1956年开始推广普通话，1982年12月把“国家推广全国通用的普通话”写入了宪法。现在中国书面语统一为普通话（部分少数民族地区除外），口语方面是普通话为主，方言并用。《中国语言文字调查》的资料显示，截至2000年全国仍然有86.38%的人使用汉语方言，并且这一比例超过使用普通话交流的比例（53.06%），见图1。各个方言区能够用普通话交谈的比例平均仅为57.79%，这一比例最低为40.15%在西南官话区，最高85.73%在北京官话区。



图 1：中国汉语方言地图

数据来源：《中国语言地图集》A2 中国汉语方言图

3. 模型和数据

3.1 引力模型

在这一部分里我们引入一个引力模型来说明为什么语言沟通与地方文化会影响劳动力的区域转移，然后再在理论模型的基础上发展实证计量模型。

和 Anderson and Wincoop (2003) 和 Falck et al. (2012) 一致，我们引入一个离散区域选择模型。假设国家内部有 $r = 1, 2, \dots, R$ 个地区和许多不同的个体 h ，个体 h 在地区 r 的效用函数为：

$$V_r^h = u_r + \varepsilon_r^h \quad (1)$$

u_r 代表地区 r 具有的给生活在该地区的个体带来效用的因素，包括该地区的教育、医疗、卫生、住房、就业、基础设施建设等宏观经济因素，也包括该地区特有的历史习俗等文化因素。这些因素对每个生活在该地区的个体影响是相同的。 ε_r^h 表明的是该地区对每个个体不同的影响，比如该地区的产业构成对不同专业和不同技术水平的个体的影响，还包括该个体对这个地区的评价。同时我们假设不同个体的 ε_r^h 独立同分布，并服从二重指数分布 $F(x) = \Pr(\varepsilon_r^h \leq x) = \exp[-\exp(-x/\beta - \gamma)]$ 。其中 $\gamma (= 0.5772)$ 是欧拉常数，参数 $\beta > 0$ 。该分布的均值为零，方差为 $(\pi^2/6)\beta^2 = 1.6449\beta^2$ 。 β 表明个体之间异质性的程度。如果区域性流动没有成本，那么个体 h 移民到地区 r 的概率是：

$$P_r = \Pr\left[V_r^h > \max_{j \neq r} \{V_j^h\}\right] = \frac{\exp(u_r/\beta)}{\sum_{j=1}^R \exp(u_j/\beta)} \quad (2)$$

当 β 趋向于 0 时，地区对个体的影响只取决于地区的平均福利水平，即只和 u_r 有关，而当 β 趋向于无穷大时，个体差异非常大，这时候个体到任意一个地区的可能性相同，概率为 $1/R$ 。下面我们把该模型扩展到两期并引入流动成本。

在第一期，不同地区对不同个体的随机影响 ε_r^h 得到实现。第二期，个体在观察到自己的 ε_r^h 和不同地区的平均福利水平 u_r 以后选择他们愿意移民的地区。任意一个个体 h 从现在的居住地 r 移民到另一个地区 s （包括现在居住地 r ）的条件是他获得的效用增加要超过他付出的成本 c_{rs} （当选择不移民时， $c_{rs} = 0$ ），即

$$\begin{aligned} V_s^h - c_{rs} &> \max\{V_j^h - c_{rj}\} \\ c_{rr} &= 0, c_{rj} \geq 0 \end{aligned} \quad (3)$$

加总个体数据后，我们可以得到从地区 r 到地区 s 的地区的移民比例为 $P_{rs} = M_{rs}/L_r$ ，其中 L_r 是迁出地的人口数量， M_{rs} 为从 r 到 s 的移民的流量（gross migration flow）。两边取对数后得到引力方程：

$$\log(M_{rs}/L_r) = (u_s - c_{rs})/\beta - \log\left[\sum_{j=1}^R \exp(u_j - c_{rj})/\beta\right] \quad (4)$$

$$\text{即：} \quad \log(M_{rs}/L_r) = u_s/\beta - \log\left[\sum_{j=1}^R \exp(u_j - c_{rj})/\beta\right] - c_{rs}/\beta \quad (5)$$

令 $D_s = u_s/\beta$; $D_r = \log\left[\sum_{j=1}^R \exp(u_j - c_{rj})/\beta\right]$; $C_{rs} = -c_{rs}/\beta$ ，则

$$\log(M_{rs} / L_r) = D_s - D_r + C_{rs} \quad (6)$$

在地区水平上， D_s 包含了所有迁入地 s 的影响， D_r 包含了所有迁出地 r 的影响。在计量模型中我们用地区固定效应来控制，下面我们定义移民成本 C_{rs} 。

方言可以通过两个渠道影响劳动力移民，第一是语言上的沟通；第二是文化习俗的相容性，分别定义为沟通成本 $C_{rs_communicate}$ 和文化成本 $C_{rs_culture}$ 。除此以外，移民成本还包括物理上的交通成本 C_{rs_travel} ，因此， $C_{rs} = C_{rs_communicate} * C_{rs_culture} * C_{rs_travel}$ 。将该式代入方程（6），我们可以得到如下方程：

$$\begin{aligned} & \log(M_{rs} / L_r) \\ & = D_s - D_r + \alpha_1 \log(c_{rs_communicate}) + \alpha_2 \log(c_{rs_culture}) + \alpha_3 \log(c_{rs_travel}) + e_{rs} \end{aligned} \quad (7)$$

其中 e_{rs} 是误差项。地区控制变量 D_s 和 D_r 分别控制了所有与迁出地和迁入地相关的因素，包括工资水平，失业率，地区发展水平，福利情况等，两者的差值是地区间的差异。如果人们更愿意移民到那些语言障碍更小，文化更为近似，地理上离家更近的地方工作，那么 α_1 ， α_2 和 α_3 的符号均应为负。

3.2 数据

迁出地和迁入地的地区的控制变量 D_s 和 D_r 来源于《中国城市统计年鉴（2000年）》，包括各县级地区的经济数据，如 GDP，人口数量，人均工资，地方，政府基本建设支出，教育事业费，社会保障辅助支出。

语言沟通成本 $c_{rs_communicate}$ 是用迁入地和迁出地方言区的普通话普及率做代理变量，普通话普及率越高，说明方言的沟通作用越少，沟通成本越低。数据来自于 2000 年《中国语言文字调查》。正如前文所述，不同性别、年龄和方言的地区普通话水平不同，因此在文章第四部分的回归中，针对每一类特定的子样本使用了不同的普通话普及率。

文化成本 $C_{rs_culture}$ 是一个 0 和 1 的哑变量，如果迁出地和迁入地使用的是同一种方言则为 1，否则为 0。地区的方言数据来源于《中国语言地图集》的 A2“中国汉语方言图”。方言地图宏观地展示了汉语方言的分布。借助于地理信息系统（GIS），我们把该地图与 2000 年的中国行政区划匹配，得到 2000 年各县级行政区划的方言情况。

地理距离 c_{rs_travel} 是按照迁出地和迁入地所在省的省会城市间铁路距离作为两地间距离。当二站有两条以上径路时，选择最短的径路或直通快车运行的径路计算，本省之内的迁移距离设为该省会城市到最近省会城市铁路距离的 1/4⁸。

⁸ 这里我们采取了 Wei (1996) 对两地距离的测度方法。事实上，我们尝试过使用迁出地和迁入地所在省省会间直线距离，省内城市间距离为本省省会城市到最近省会城市直线距离的 1/4 的测度方法，以及省内城市距离为该省面积除以圆周率再开根号的方法 (Poncet, 2003)。回归结果相差不大。

被解释变量各地区的移民比例 $P_{rs} = M_{rs} / L_r$ 是用 2000 年中国第五次人口普查全国 1% 个体抽样数据构造的。移民的迁入地是指劳动力现在的居住地，迁出地是指五年前的居住地⁹，把现在的居住地作为迁入地，如果迁出地和迁入地不是同一地点，就认为该劳动力为“移民”。由于我们研究的对象是劳动力的区域转移，我们把年龄低于 16 周岁的样本去除。在这里我们不设置年龄的上线是因为有些特殊职业例如中医，存在退休后返聘的现象，从而可能仍然在劳动力市场。我们剔除了移民原因为非工作原因（包括学习培训、拆迁搬家、婚姻迁入、随迁家属、投靠亲友和其他）的样本点；没有迁出地记录，迁出地在外国以及户籍注册地与现在居住地一样的样本点；无法与上述方言地区数据匹配¹⁰的样本点后。然后把迁入地和迁出地两两配对，得到了 49326 对县级地区之间的移民流量数据 M_{rs} （大约占理论值的 2%）和迁出地的总样本数量 L_r ，即可得到 $P_{rs} = M_{rs} / L_r$ 。表 1 给出了描述性统计。

表 1 描述性统计

		个体数据				
变量	分类	样本量	比例	比例		
性别分布	男性	70488	59.09	59.09		
	女性	48801	40.91	40.91		
年龄分布	16-20 岁	17906	15.01	15.01		
	20-30 岁	58080	48.69	48.69		
	30-40 岁	25022	20.98	20.98		
	40-50 岁	7975	6.69	6.69		
	50 岁以上	10306	8.64	8.64		
户籍分布	农村户口	86459	72.87	72.87		
	城市户口	32165	27.11	27.11		
教育水平分布	未上过学	1799	1.51	1.51		
	扫盲班	292	0.24	0.24		
	小学	18,601	15.59	15.59		
	初中	70,060	58.73	58.73		
	高中	11,536	9.67	9.67		
	中专	6,961	5.84	5.84		
	大学专科	5,428	4.55	4.55		
产业分布	大学本科	4,185	3.51	3.51		
	研究生	427	0.36	0.36		
	第一产业	4,152	3.82	3.82		
	第二产业	63,057	57.96	57.96		
	第三产业	41,585	38.22	38.22		
		地区数据				
变量	分类	样本数量	平均值	标准差	最小值	最大值
普通话普及率 (百分比)	迁入地	41313	0.6512946	0.13551	0.4015	0.8573
	迁出地	41758	0.5629016	0.14077	0.4015	0.8573

⁹理想情况下最好使用劳动力的出生地作为迁出地，第一次外出的工作地点为迁入地，但由于数据的限制，只能把五年前居住地作为迁出地。

¹⁰第五次人口普查的编码和行政区划有一定的出入，我们做相应调整，此外我们去掉少数民族聚集区和同时讲多种方言的地区。

人均工资 (元/人, 年)	迁入地	40788	12055.11	4292.004	0.641336	19674.78
	迁出地	40920	8462.474	2920.208	0.641336	19674.78
GDP (万元)	迁入地	40914	717687.4	729.018	7120	3783100
	迁出地	40837	588277.8	1338866	5700	23830700
固定资产投资额 (万元)	迁入地	40788	3,615,793	4899673	104815	18696749
	迁出地	40920	1424922	2567160	75418	18696749
人口 (万人)	迁入地	41240	58.81437	36.04575	2	249
	迁出地	41207	83.84583	109.432	2	1057
政府教育支出 (万元)	迁入地	41516	11554.22	13749.74	189	95644
	迁出地	41435	8176.152	14528.74	21	251923
政府社会保障支出 (万元)	迁入地	41521	4342.853	5569.614	81	27566
	迁出地	41441	2771.467	8470.647	81	243408

移民流数据

变量	样本数量	平均值	标准差	最小值	最大值
移民数量	41758	2.856674	50.78957	1	7315
移民比例	41758	0.04107	0.0663219	0.000132	1
是否共同方言	42048	0.309622	0.4623434	0	1
地理距离(千米)	42048	866.0211	809.95	34.25	6313
是否同省	42048	0.307506	0.4614662	0	1
是否邻省	42048	0.591705	0.4915242	0	1

数据来源: 表 3 的数据根据第五次人口普查 1% 抽样和《2000 年中国城市统计年鉴》整理

4. 计量结果

基于方程 (7) 的理论模型我们进行计量回归。表 2 是基本回归结果。表 2 中第 1 到第 5 列是 OLS 估计结果, 第 6 列是泊松回归结果。在第 1 列的回归结果中, 解释变量只有共同方言、迁入地和迁出地的普通话水平和地理距离, 第 2 列控制了是否属于同一省份以及是否邻省, 第 3 列加入了迁入地和迁出地与方言的交叉项, 第 4 列加入了代表迁入地和迁出地的经济情况的控制变量, 第 5 列控制了迁出地的固定效应和迁入地的经济控制变量。

表 2 基本回归结果

变量	(1)	(2)	(3)	(4)	(5)	(6)
	移民比例 (取对数)	移民比例 (取对数)	移民比例 (取对数)	移民比例 (取对数)	移民比例 (取对数)	移民人数 (POISSON)
共同方言	0.338*** (0.0162)	0.261*** (0.0166)	0.223*** (0.0348)	0.222*** (0.0164)	0.179*** (0.0165)	0.343*** (0.00957)
迁入地普通话	0.0840*** (0.0257)	0.128*** (0.0259)	0.152*** (0.0309)	-0.0119 (0.0272)	0.0345* (0.0207)	0.173*** (0.0303)
迁出地普通话	-0.355*** (0.0234)	-0.325*** (0.0236)	-0.308*** (0.0271)	-0.558*** (0.0239)		-0.0952*** (0.0286)
距离	-0.0712*** (0.00585)	0.0522*** (0.0113)	0.0520*** (0.0113)	0.0415*** (0.0119)	0.0196 (0.0127)	8.99e-5*** (7.96e-06)
是否同省		0.454*** (0.0258)	0.453*** (0.0258)	0.456*** (0.0267)	0.247*** (0.0274)	0.345*** (0.0109)
是否邻省		-0.00408 (0.0170)	-0.00391 (0.0170)	0.0280 (0.0177)	0.223*** (0.0207)	0.475*** (0.0119)
共同方言与迁入地普 通话交叉项			-1.700** (0.716)			
共同方言与迁出地普 通话交叉项			1.631** (0.722)			

地区经济控制变量	否	否	否	是	是	是
迁出地固定效应	否	否	否	否	是	否
观测值个数	41,313	41,313	41,313	38,066	39,555	38,066
R ²	0.039	0.047	0.047	0.150	0.087	

注：①第（1）到（5）列是 OLS 估计，第（6）列是泊松回归；②“迁出地和迁入地控制变量”包括平均工资、GDP、固定资产投资、社会保障政府投入、教育投入、地区人口数量；③括号内为稳健标准差。*，**，***分别表示 10%，5%，1%显著水平。

基本回归结果表明即使在控制了地理距离、地区经济因素、迁出地固定效应以后，哑变量“共同方言”仍然十分显著。当迁入地和迁出地属于共同方言区时，两地区之间的移民数量占迁入地总人口的比例要正向变动近 0.2 个百分点。普通话普及率的影响较为复杂，迁入地的普通话水平越高，移民比例越高。但是迁出地的普通话水平则有两个影响，一方面普通话水平越高，该地区劳动力人力资本中语言沟通能力较高，有利于外出打工进而移民；另一方面普通话水平高也意味着本地区越开放，容易吸引外地区移民进来，使得本地区人口、工作机会更多，经济越发达，导致本地劳动力放弃到外地打工。因此迁出地普通话普及率对本地居民既有推力作用，又有拉力作用。从表 2 中可以看出，迁出地普通话普及率前面的系数变为负，说明迁出地普通话普及率的拉力作用更大。第 3 列中共同方言与迁入地和迁出地普通话水平的交叉项非常显著：如果迁入地和迁出地属于共同方言区，那么迁入地和迁出地的普通话水平对移民流都是负向影响；如果不属于共同方言区，迁入地和迁出地的普通话水平与移民流都是正向影响。可以看出，当无法用方言作为沟通语言时，普通话的普及促进了劳动力的流动。从这个意义上来看普通话的普及对中国劳动力的流动起到非常大的促进作用，在过去的劳动力转移的原因分析的文献中忽略了这一点。虽然从建国初期就已经开始推广普通话，但是正如《中国语言文字调查》发现的那样，不同地区的普通话普及率差异很大，特别是某些贫困地区和少数民族地区，普通话普及率非常低，这不仅制约着劳动力本身的人力资本，也影响了劳动力的引入。表 2 中另一个发现是地理距离对劳动力流动的影响。一般来说地理距离越远，劳动力移民应该越少。但表 2 中发现，劳动力会倾向于流动到同省和邻省，在控制了同省和邻省因素后，地理距离越远，劳动力移民越多。这可能与中国东西部经济发展不平衡有关系。

表 3 至表 6 是异质性回归的结果，回归计量模型与表 2 中的第（5）列一致。按照劳动力性别、从事的产业、教育水平、户口进行分类，比较不同特征的劳动力流动的特点。

表 3 分性别回归结果

变量	男性	女性
共同方言	0.188***	0.109***
	-0.0167	-0.0213
迁入地普通话	-0.0297	0.409***
	-0.0298	-0.035
迁出地普通话	-0.786***	0.0601*
	-0.0262	-0.031
地理控制变量	是	是

地区经济控制变量	是	是
观测值个数	30,522	19,505
R ²	0.187	0.161

首先来看男性劳动力与女性劳动力的差异。《中国语言文字使用调查》发现在各方言区男性和女性的普通话水平并不相同，在所有方言区男性的普通话水平都比女性要高。我们分为男性样本组和女性样本组分别回归，表6给出了回归结果。我们发现方言所代表的文化因素对男性的影响比女性略大，但是普通话普及率的影响反差很大。迁入地普通话的普及率对女性的影响要大的多，这可能是由于女性从事第三产业的居多（样本中，女性从事第三产业的比例是41.1%，男性的比例是36.8%），第三产业对语言沟通的要求更高一些，表4中的回归结果论证了这一点。迁出地的普通话普及率对迁入地男性移民比例的影响是负的，对女性是正的。可见迁出地的普通话水平对女性的人力资本积累作用更为明显，从而促进了移民。

从直觉上来说第三产业的从业者受到语言和文化的影响应该更为明显，我们分产业进行回归，发现结果和直觉一致。在表4中，“共同方言”、“迁入地普通话水平”和“迁出地普通话水平”前面的系数在第三产业的分组回归方程中绝对值教大，也最为显著。共同的文化背景和高水平的普通话都有利于劳动力的流动。

表4 分三次产业回归结果

变量	第一产业	第二产业	第三产业
共同方言	0.125***	0.146***	0.196***
	-0.048	-0.0218	-0.021
迁入地普通话	0.0213	-0.244***	0.219***
	-0.0878	-0.0386	-0.0368
迁出地普通话	-0.669***	-0.295***	-0.760***
	-0.0792	-0.0335	-0.0317
地理控制变量	是	是	是
地区经济控制变量	是	是	是
观测值个数	2,514	23,081	17,101
R ²	0.129	0.173	0.167

语言和文化对不同教育水平的劳动力移民的影响有何不同呢？把样本分为初中及以下，高中、中专和大专，大学本科以及研究生四组，分别考察语言和文化对劳动力转移的影响。表5汇报了结果。除了研究生学历的劳动力，方言和文化对不同教育水平的劳动力移民影响差不多。迁入地普通话对最低教育水平和最高教育水平的劳动力移民的影响都不显著，这可能是由于对于低技能劳动力来说，他们从事的职业要求的技能不高，流动性强，经济因素往往是流动的主要动因，体现在地区人均收入前的系数较大而且显著，语言沟通因素并不那么重要；对研究生以上学历的劳动力影响不显著的原因可能是由于样本量的限制，也有可能是由于高技能的劳动力本身普通话水平非常高，工作中使用的也多为普通话，因此语言不容易成为障碍。

表5 分教育水平回归结果

变量	初中及以下	高中、中专和大专	本科	研究生
----	-------	----------	----	-----

共同方言	0.174***	0.242***	0.191***	0.137
	-0.0207	-0.0198	-0.0498	-0.147
迁入地普通话	-0.018	0.119***	0.165*	0.234
	-0.0351	-0.0362	-0.0875	-0.261
迁出地普通话	-0.507***	-0.718***	-0.495***	-0.919***
	-0.0316	-0.0314	-0.0714	-0.182
地理控制变量	是	是	是	是
地区经济控制变量	是	是	是	是
观测值个数	27,844	14,572	3,198	343
R ²	0.183	0.131	0.363	0.357

为了进一步研究对农民工流动的影响，我们又按照户籍进行分类，分为农村户口样本组和城市户口样本组，结果发现地方文化（方言）对城市劳动力影响较大说明城市户口的劳动力可能更注重社会人文等非经济因素的影响。城市与农村劳动力流动的一个明显差异在于语言的影响。表6中迁入地普通话前的系数在农村样本组中为正而且绝对值较大；在城市样本组中为负且绝对值较小，迁出地普通话前的系数在城市组中更大，这些都表明普通话的普及对农民工的流动有更大的意义。普通话为他们外出务工提供了语言基础，内陆的农民工不用为了到广东打工而学习粤语，为了到福建打工学习闽南语，只要掌握了普通话便可以到任何地方工作和生活，节省了大量的时间和金钱成本。

表6 分户籍回归结果

变量	农村	城市
共同方言	0.211*** (0.0212)	0.308*** (0.0214)
迁入地普通话	0.132*** (0.0346)	-0.0815** (0.0361)
迁出地普通话	-0.312*** (0.0313)	-0.773*** (0.0322)
地理控制变量	是	是
地区经济控制变量	是	是
观测值个数	28,066	14,748
R ²	0.139	0.151

5. 结论与政策建议

地域广阔，方言众多和文化多元是中国很重要的社会特色，在劳动力市场上也有所体现：一方面体现在供给方面，即不同地区的劳动力的具有不同的文化禀赋——拥有不同的方言和地方文化并且普通话水平差异较大；另一方面体现在需求方面，即企业的人才需求不同——语言沟通能力和与本地化融合程度。当劳动力择业和移民时，除了考虑收入等经济因素外，非经济因素同样重要，在过去研究我国劳动力流动的文献中往往忽略了这一点，本文试图弥补这方面的空白：把语言和地方文化纳入劳动力流动的分析。我们用普通话的普及水平作为语言沟通障碍的代理变量，用是否在同一方言区作为地方文化的代理变量，结果发现在控制了地理距离、经济因素（人均收入、基础建设、社会保障、政府教育投入等）后，语言障碍和地方文化仍然对劳动力移民有显著的影响：语言障碍越小（迁入

地的普通话水平越高，迁出地普通话水平越低），地方文化越接近（迁入地和迁出地属于共同方言区），移民数量越多。这说明除了经济因素外，语言障碍和地方文化是劳动力选择移民时的重要考虑因素。进一步的研究还表明随着性别、教育水平、户口、产业分布的不同，语言和文化对劳动力流动的影响也会不同。

本文的研究不仅揭示中国劳动力市场的非经济特征，还间接地论证了推广普通话这一政策的经济意义。由于数据的限制，我们通常很难直接衡量推广普通话带来的经济效益，通过本文的研究发现，普通话普及率高的地方外来移民的数量也相对较多，同时本地居民稳定性更好（向外移民少），普通话的普及对农民工流动的促进作用更加明显，因此普通话的推广为中国区域间劳动力的自由流动打下了基础。普通话的普及的第二个意义在于促进中国第三产业的发展。较高水平的普通话水平不仅提升了服务业从业人员的服务质量（外地人不会因为口音问题而无法享受高质量服务），而且普通话普及率高的城市也更容易吸引更多的劳动力，相较于农业和制造业，从业于第三产业的劳动力的流动受到普通话普及率的影响更大。根据国家语言委员会的公开数据，截止 2010 年中国大陆普通话普及率达到 54%以上，这说明很多地方普通话仍然没有完全普及，根据前面的研究结论，国家应该加大力度普及普通话，特别是农村地区的普通话，这将有利于农民工的流动，也有利于我国产业结构调整 and 升级。

参考文献：

1. 蔡昉等，2003. 劳动力流动的政治经济学 [M] 上海，上海人民出版社
2. 蔡昉等，2001. 户籍制度与劳动力市场保护 [J]. 经济研究 12:41-49.
3. 邓曲恒和古斯塔夫森 2007. 中国的永久移民[J]. 经济研究 4:137-148
4. 孙文凯等，2011. 户籍制度改革对中国农村劳动力流动的影响[J]. 经济研究 1:28-41
5. 王德文等，2008. 农村迁移劳动力就业与工资决定：教育与培训的重要性[J]. 经济学（季刊），7(4)：1132-1148.
6. Anderson, J. E., and E. van Wincoop, 2003. Gravity with Gravitas: A Solution to the Border Puzzle [J]. American Economic Review 93(1): 170-192.
7. Carliner, G. 1981, Wage differences by language group and the market for language skills in Canada [J]. Journal of Human Resources 16 (3):384-399.
8. Chiswick, B.R. 1998 Hebrew language usage: Determinants and effects on earnings in Israel [J]. Journal of Population Economics 11 (2):253-271.
9. Chiswick, B.R. and Miller, P.W. 1995, The endogeneity between language and earnings: International analyses [J]. Journal of Labor Economics 13 (2):246-288
10. Chiswick, B.R. and Miller, P.W. 2001, A model of destination language acquisition: Application to male immigrants in Canada [J]. Demography 38 (3):391-409.
11. Chiswick, B.R. and Miller, P.W. 2003, The complementarity of language and other human capital: Immigrant earnings in Canada [J]. Economics of Education Review 22 (5):469-480.
12. Chiswick, B.R., Patrinos, H.A. and Hurst, M.E. 2000, Indigenous language skills and the labor market in a developing economy [J]. Economic Development and Cultural Change 48 (2):349-367.
13. Chiswick, B.R. 1991, Speaking, reading and earnings among low-skilled immigrants [J]. Journal of Labor Economics 9 (2): 149-170.
14. Dustmann, C. and Soest, A.V. 2001, Language fluency and earnings estimation with misclassified language indicators [J]. Review of Economics and Statistics 83 (4): 663-674.

15. Dustmann, C.1994 Speaking fluency, writing fluency and earnings of migrants [J]. *Journal of Population Economics* 7: 133-156.
16. Falck, O., Heblich, S., Lameli, A., and Südekum, 2012. Dialects, cultural identity, and economic exchange[J]. *Journal of Urban Economics*, 72(2):225-239.
17. Kossoudji, S.A. 1988, English language ability and the labor market opportunities of Hispanic and East Asian immigrant men [J]. *Journal of Labor Economics* 6(2): 205-228.
18. McManus, Walter, William Gould, and Finis Welch, 1983. Earnings of Hispanic Men: The Role of English Language Proficiency [J]. *Journal of Labor Economics* 1(2):101-130.
19. Melitz, J. 2008. Language and foreign trade. *European Economic Review* [J]. 52(4): 667-699.
20. Poncet Sandra, 2003. Measuring Chinese Home and International Integration [J], *China Economic Review*,14(1) : 122.
21. Rauch James E. and Trindade Vitor, 2002.Ethnic Chinese networks in international trade [J]. *The Review of Economics and Statistics*, 84(1): 116-130.
22. Rauch, James E., 1999. Networks versus markets in international trade [J]. *Journal of International Economics*, Elsevier, vol. 48(1): 7-35.
23. Rubinstein, A., *Economics and Language* [M]. Cambridge: Cambridge University Press,2000.
24. Tainer,E. 1988, English language proficiency and the determination of earnings among foreign-born men [J]. *The Journal of Human Resources*,23(1):108-122.
25. Wei ShangJin, 1996. Intranational versus International Trade: How Stubborn are Nations in Global Integration. National Bureau of Economic Research Working Paper No. 5531246

附录1 能用普通话与人交谈的比例（2000年）

能够使用普通话与人交谈的比例（百分比）		
按性别分类	男性	61.41
	女性	53.30
按居住地分类	城镇	67.36
	乡村	49.38
按年龄分类	15-29岁	70.12
	30-44岁	52.74
	45-59岁	40.59
	60-69岁	30.97
按教育程度分类	没上过学	10.36
	扫盲班	14.67
	小学	25.49
	初中	56.08
	高中	75.76
按方言地区分类	大专及以上	86.77
	官话区	49.92
	晋语区	43.61
	吴语区	69.40
	闽语区	80.28
	粤语区	61.00
	客家话区	64.36
	赣语区	60.86
	湘语区	54.80
	徽语区	56.13
平话区	51.66	
其他	46.19	

数据来源：根据《中国语言文字使用情况调查（2000）》整理

附录2 变量解释及数据来源

变量名	变量解释	数据来源
移民比例	每一对迁出地和迁入地的配对样本中，迁出地移民占迁出地的人口比例。	根据第五次人口普查的微观样本计算
普通话水平	即迁出地和迁入地的“能使用普通话交流的比例”，不同性别、户籍、年龄、教育水平的样本组该值均不同，异质性回归时将做相应调整	中国语言文字调查
共同方言	哑变量，移民迁出地和本地的方言如果相同则取值为1，否则取0。	中国语言地图集
距离	所在省省会间直线距离，其中省内城市间距离为本省省会城市到最近省会城市直线距离的1/4，与Wei（1996）等人方法一致。	作者自己测算
迁入地	根据普查问卷中“本户地址：__县（市、区）__乡（镇、街道）__普查区__调查小区”得到现居住地。	第五次人口普查样本
迁出地	根据普查问卷中“从何地来本乡镇街道居住？1. 本县、市、区以内；2. 本县、市、区以外：__省__地（市）__县（市、区）”得到迁出地。	第五次人口普查样本
地区经济情况	包括平均工资、地区GDP、人口数量、基础设施建设投资、社会保障政府投入、教育投入等。	中国统计年鉴
年龄	样本中只保留16岁以上的观测值	第五次人口普查样本
工作状态	样本中只保留正在工作以及未工作但正在找工作的观测值	第五次人口普查样本
个人教育水平	普查问卷中“受教育程度”从低到高分分为1—9档，分别代表未上过学，扫盲班，小学，初中，高中，中专，大学专科，大学本科，研究生。	第五次人口普查样本

Language, Culture and Labor Migration

Abstract: Language and culture background both play essential roles in labor migration. This paper tries to analyze how these two factors affect Chinese labor migration. Besides mandarin, people also use dialects to communicate in daily life. We use mandarin to measure the communication costs while use dialects to capture the culture's influence. Our findings imply that there are intangible linguistic and cultural borders within a country that impede labor transfer across its regions. People are more willing to move to culturally familiar environment with low communication cost. We also find that the effects of language and dialect will vary with the difference of gender, age, and education level and birth place.

Keywords: Dialect, culture, labor migration, gravity model