

国家通用语言能力对民族地区农村 劳动力非农就业的影响研究^{*}

丁 赛 阎 竣

本文基于“21世纪初中国少数民族地区经济社会发展综合调查”课题组的大样本调查数据,分析了民族地区农村劳动力掌握国家通用语言——普通话能力对非农就业的影响。调查数据显示,民族地区少数民族农村劳动力不能用普通话交流的占32.2%;从事非农就业的仅占36.9%,纯非农就业仅占10.8%,非农就业比重远低于全国农村平均水平。采用扩展回归模型,以首先学会的语言为工具变量,得到普通话水平对提升非农就业概率有显著影响的结论。如果农村劳动力从“无法用普通话交流”转变为“能流利运用普通话交流”,可使其非农就业概率提升52%,从事兼业和纯非农就业的概率分别提高24%和34%。如果从“无法用普通话交流”转变为“普通话虽不熟练但可以交流”,那么可使从事非农就业概率提升22%。普通话水平对非农就业的影响随少数民族聚居程度呈现出先下降后上升的U形,这意味着推广普通话以提升非农就业既要聚焦于少数民族聚居程度高的地区,同时也应重视少数民族聚居区域汉族的普通话推广和非农就业问题。

关键词:农村劳动力 国家通用语言 非农就业 民族地区

作者丁赛,女,中国社会科学院民族学与人类学研究所铸牢中华民族共同体意识研究基地研究员。地址:北京市中关村南大街27号6号楼,邮编100081。阎竣(通讯作者),女,首都经济贸易大学会计学院副教授。地址:北京市,邮编100070。

一、引 言

农村劳动力非农就业不仅是提高农民收入,摆脱贫困的重要方式,也是促进城乡要素自由流动和城乡融合发展的重要推动力。随着工业化和城镇化的持续推进,非农就业已经成为全国农村家庭收入的主要渠道。农村劳动力中,从事农业的劳动力比例已经低于45%,从事非农就业的比例已经上升至55%以上,^①非农收入比重上升至40%以上。^② 但我民族地区农

^{*} 本文系国家社科基金重点课题“少数民族传统文化在乡村振兴中的作用与创新转化研究”(项目编号:19AMZ015)、中国社会科学院创新工程重大科研规划项目“铸牢中华民族共同体意识重大问题研究”(项目编号:2019ZDGH017)的阶段性研究成果。

① 参见李实、Terry Sicular、Finn Tarp:《中国收入不平等:发展、转型和政策》,《北京工商大学学报》2020年第4期。

② 参见罗楚亮:《收入结构与居民收入差距变动——新中国成立以来收入差距的基本特征》,《北京工商大学学报》2020年第4期。

村劳动力非农就业的比例偏低,严重制约了民族地区的经济发展和农民收入提升。2018年,全国农村居民工资性收入位居四种收入来源的首位,占可支配收入的41%,^①而民族八省区的工资性收入占比仅为30.5%,^②比全国平均水平低10个百分点。在脱贫攻坚的主战场——“三区三州”,农村少数民族非农就业人数少、比例低,家庭收入来源单一的情况更加明显。究其原因,固然有少数民族农村劳动力平均受教育水平低,以及文化传统和生活习惯差异引起的新环境适应性不强等原因外,国家通用语言普通话能力弱也可能是导致其难以转移就业的重要原因。

语言不仅具有交际功能,也是经济优势的一种来源。^③语言表现为个体技能和知识的存量,是重要的人力资本,尽管获取语言技能需要花费成本,但一旦形成人力资本,人们就会有更多的交际可能,也会有可能得到较高的收入或更好的职位。^④语言的流利程度对劳动收入,尤其是对移民的劳动收入有着重要的影响。^⑤掌握国家通用语言的能力是语言资本的核心形式,直接关系到个人或群体所能达到的社会地位和获得的各种资源。^⑥反之,国家通用语言说得不好便成为阻碍个人获得经济社会资源的重要障碍。^⑦在城市劳动力市场中,少数民族普通话说得越标准,其收入就越高,相对于第一、第二产业,第三产业的从业者受语言的影响更大。^⑧

少数民族语言是各少数民族内部成员之间的交际语言,国家通用语即普通话是中华民族共同体的通用语言。^⑨我国除回族和满族通用汉语文外,其他53个少数民族都有本民族语

① 根据《中国统计年鉴2019年》(国家统计局编,中国统计出版社2019年版)表6-30计算得到。

② 根据《中国统计年鉴2019年》相关统计数据计算得到。

③ 参见 Albert Breton, “Economics of Nationalism,” *Journal of Political Economy*, Vol. 72, No. 4, 1964; Albert Breton, “Nationalism and Language Policies,” *Canadian Journal of Economics*, Vol. 11, No. 4, 1978。

④ 参见 Edward Lazear, “Culture and Language,” *Journal of Political Economy*, Vol. 107, No. 46, 1999; S. A. Kossoudji, “English Language Ability and the Labor Market Opportunities of Hispanic and East Asian Immigrant Men,” *Journal of Labor Economics*, Vol. 6, No. 2, 1988。

⑤ 参见 W. McManus, “Labor Market Cost of Language Disparity: An Interpretation of Hispanic Earnings Differences,” *American Economic Review*, Vol. 75, No. 4, 1985; D. Leslie and J. Lindley, “The Impact of Language Ability on Employment and Earnings of Britain’s Ethnic Communities,” *Economica*, Vol. 68, Issue 272, 2001; B. R. Chriswick and P. W. Mille, “The Complementarities of Language and Other Human Capital: Immigrant Earnings in Canada,” *Economics of Education Review*, Vol. 22, No. 5, 2003; Hoyt Bleakley and Aimee Chin, “Language Skills and Earnings: Evidence from Childhood Immigrants,” *Review of Economics and Statistics*, Vol. 86, No. 2, 2004; J. Kim and J. Zhou, “Language and Earnings Management,” *Journal of Accounting and Economics*, Vol. 63, Issues 2-3, 2017。

⑥ 参见 Pierre Bourdieu, “The Forms of Capital,” in John G. Richardson ed., *Handbook of Theory and Research for the Sociology of Education*, New York: Greenwood Press, 1992; Jeroen Smith and Aye Gunduz-Hogor, “Linguistic Capital: Language as a Socio-Economic Resource among Kurdish and Arabic Women in Turkey,” *Ethnic and Racial Studies*, Vol. 26, No. 5, 2003。

⑦ 参见 Christian Dustman and Arthur Van Soest, “Language and Earnings of Immigrants,” *Industrial and Labor Relations Review*, Vol. 55, No. 3, 2002; Joop Hartog and Rainer Winkelmann, “Comparing Migrants to Non-Migrants: The Case of Dutch Migrants to New Zealand,” *Journal of Population Economics*, Vol. 16, No. 4, 2003; Teresa Casey and Christian Dustmann, “Intergenerational Transmission of Language Capital and Economic Outcomes,” *The Journal of Human Resources*, Vol. 43, No. 3, 2008。

⑧ 参见李秦、孟岭生:《方言、普通话与中国劳动力区域流动》,《经济学报》2014年第4期; Zhao Chen, Ming Lu and Le Xu, “Returns to Dialect: Identity Exposure Through Language in the Chinese Labor Market,” *China Economic Review*, Vol. 30, 2014。

⑨ 参见苏德、张良、江涛:《新时代背景下的少数民族双语教育:机遇·挑战·策略》,《民族教育研究》2019年第4期。

言,其中22个少数民族共使用28种文字。^①中华人民共和国成立以来,我国民族地区有序实施了双语教育。^②改革开放后的20多年中,双语教育的特点是以民族语文学习为主,以汉语学习为辅。^③2002年第五次全国民族教育工作会议提出,要在认真做好学校教育中民族语文授课有关工作的同时,抓紧研究并采取有力措施加强和改进汉语教学,提高汉语教学质量。^④自此,双语教育模式也改变了原先双语教育过于偏重民族语言学习的缺陷。^⑤2015年颁布的《国务院关于加快发展民族教育的决定》规定,“坚定不移推行国家通用语言文字教育,确保少数民族学生基本掌握和使用国家通用语言文字,少数民族高校毕业生能够熟练掌握和使用国家通用语言文字。尊重和保障少数民族使用本民族语言文字接受教育的权利,不断提高少数民族语言文字教学水平”。^⑥

近年来民族地区推广国家通用语言——普通话的力度不断提升。民族地区,特别是深度贫困地区,由于教育条件落后,加上少数民族聚居地区日常习惯用民族语言沟通等因素,导致民族地区人口的国家通用语言能力较差。在这些地方,国家通用语言教育的条件不足,甚至缺失,^⑦与经济社会发展滞后和深度贫困问题交织,已经影响到劳动力转移就业和民族间的交往交流交融。全面建成小康社会“一个民族都不能少”的共同富裕目标,也包括了推广国家通用语言文字“一个民族都不能少”的共同语言目标。^⑧国家语言文字事业“十三五”发展规划确定了国家通用语言文字普及攻坚工程,到2020年,在全国范围内基本普及国家通用语言文字,特别是民族地区国家通用语言文字普及程度大幅度提高。^⑨教育部、国务院扶贫办和国家语委研究制定了《推普脱贫攻坚行动计划(2018—2020年)》,^⑩推普攻坚工程与国家脱贫攻坚工程衔接,全力推进。研究证实,国家通用语言能力是影响劳动者就业数量、质量及收入水平的重要因素。^⑪语言扶贫,也就是普通话的推广,有利于提升劳动者技能与素质,增强就业机会及收入水平,从而推动农村劳动力转移,而劳动力转移就业是脱贫最有效的方式。^⑫

语言作为身份认同的重要维度,人与人之间感情、信息交流不可或缺的工具和主要的文化载体,其对经济和社会的影响在国内外学术界均引起了越来越多的关注。我国民族地区农村少数民族国家通用语言能力与其经济活动的相关实证研究还不多见。本文基于“21世纪初中

① 参见《为人民谋幸福:新中国人权事业发展70年》,国家民委网站,2019年9月23日,<http://www.seac.gov.cn/seac/xwzx/201909/1136701.shtml>。

② 参见张学敏、石泽婷:《民族教育发展是中华民族共同体意识建设的内生逻辑》,《西南大学学报》2019年第4期。

③⑤ 参见褚慧:《改革开放40年我国少数民族双语教育研究热点、历史过程与展望》,《信阳师范学院学报》2020年第1期。

④ 参见《李岚清在第五次全国民族教育工作会议上强调加快民族教育事业发展 促进各民族共同繁荣》,《人民教育》2002年第9期。

⑥ 参见赵岩:《为推动和见证双语教育发展而自豪——访中央民族大学教授、中国少数民族双语教学研究会荣誉会长丁文楼》,《中国民族教育》2019年第9期。

⑦ 参见郝时远:《铸牢中华民族共同体意识必须推广国家通用语言文字》,《人民日报》2018年10月31日。

⑧ 参见张日培等:《语言产业发展的方略与措施》(笔谈),《河南师范大学学报》2019年第3期。

⑨ 参见《国家语言文字事业“十三五”发展规划》,百度文库,<https://wenku.baidu.com/view/3d5fb5e0227916888586d785.html>。

⑩ 《三部门关于印发〈推普脱贫攻坚行动计划(2018—2020年)〉的通知》,2018年2月27日,http://www.gov.cn/xinwen/2018-02/27/content_5269317.htm。

⑪ 参见黄少安、王麓淙:《民族地区语言扶贫的经济理论基础和实证分析》,《语言文字应用》2020年第4期。

⑫ 参见王麓淙、刘金林、马静、戴静超:《语言扶贫与农村劳动力转移——来自中国推广普通话的证据》,《制度经济学研究》2020年第2期。

国少数民族地区经济社会发展综合调查”课题组搜集的大样本数据(2013—2015年),以民族地区农村劳动力为研究对象,立足于交往交流交融及铸牢中华民族共同体意识的视角,着重分析国家通用语言(以下简称普通话)口语交流能力对农村劳动力非农就业的影响。

二、调查与数据

本文使用的数据来自中国社会科学院民族学与人类学研究所主持的国家社科基金特别委托暨中国社会科学院创新工程重大专项“21世纪初中国少数民族地区经济社会发展综合调查”课题组于2013—2016年在民族地区完成的家庭问卷调查数据(简称“大调查数据”)。本次调查由中国社会科学院民族学与人类学研究所主持,通过与新疆师范大学、广西民族大学、西藏大学、中央民族大学等高校的共同合作,以有调查经验的民族学或社会学专业的研究生和本科生为主体组建了调查队,具体采用分层随机抽样方法。首先,在民族自治地方选取30个调查县(市)/旗;其次,选取的县(市)/旗内的城乡调查社区或行政村以能代表当地城乡不同经济发展状况(高、中、低)和主要民族为依据;最后,每个县市的400—500份问卷根据当地的城镇化率进行城乡划分,在确定的社区或行政村层面对住户进行随机等距抽样。^① 调查问卷内容包括了经济发展、社会事业、民族文化、民族政策、民族关系、社会安全与社会和谐等方面,共获得样本13,393个。

基于本文分析目标,我们从中抽取了2013—2015年的数据,没有纳入2016年的2029个样本,主要是因为该年问卷微调时删除了一个我们需要的关键变量(您最先学会的语言)。由于我们分析的是就业问题,因此进一步删除了51个年龄在16岁以下的样本,以及仅从事家务劳动的样本1919个。此外,还删除了本文所用到的其他变量数据存在缺失的样本。最终纳入分析的样本为8857个。由于每次调查的区域和民族不同,故将2013—2015年数据合并分析,三年调查样本共包括了55个少数民族中的38个,^②来自14个省/区45个县,其中2013年3125人,2014年3421人,2015年2311人。

表1 民族地区少数民族和汉族劳动力的普通话口语交流能力 单位:%;个

普通话水平	少数民族	汉族	合计
无法交流	32.2	5.0	27.1
不熟练但可交流	33.9	37.8	34.6
熟练交流	33.9	57.2	38.3
合计	100.0	100.0	100.0
样本数	(7213)	(1644)	(8857)

如表1所示,民族地区16岁及以上劳动力中,只有38.3%能用普通话熟练交流,有27.1%无法用普通话交流。民族地区农村劳动力中的汉族和少数民族普通话口语交流能力存

^① 抽样方法参见王延中、丁赛主编:《中国民族地区经济社会调查报告:2013年调查问卷分析·综合卷》,中国社会科学出版社2015年版,第3—5页。

^② 调查样本在50个以上的少数民族共40个,包括蒙古族、回族、藏族、维吾尔族、苗族、彝族、壮族、布依族、朝鲜族、满族、侗族、瑶族、白族、土家族、哈尼族、哈萨克族、傣族、黎族、傈僳族、佤族、畲族、拉祜族、水族、纳西族、景颇族、柯尔克孜族、土族、达斡尔族、羌族、撒拉族、毛南族、锡伯族、普米族、塔吉克族、怒族、裕固族、独龙族、鄂伦春族、门巴族、珞巴族。

在明显差异。少数民族无法交流、不熟练但可以交流以及能熟练交流的各约占三分之一。汉族熟练使用普通话口语的比例为57.2%，仅5%的汉族无法用普通话交流，少数民族熟练使用普通话口语交流的比例比汉族低23个百分点。

部分民族的普通话交流水平如图1所示。位于西南的壮族、白族和苗族能熟练使用普通话的比例都超过了半数，蒙古族该比例接近一半，但撒拉族、彝族、藏族、维吾尔族该比例依次下降；藏族和维吾尔族该比例在调查样本的民族中是最低的。“其他少数民族”（不包括回族）细分为北方和南方后，北方“其他少数民族”普通话的熟练比例和无法交流比例都超过了南方其他少数民族。从推广国家通用语言文字的全局而言，民族地区的汉族不会讲普通话的人口比重仍相当大，同样是推广国家通用语言文字的攻坚对象。^①

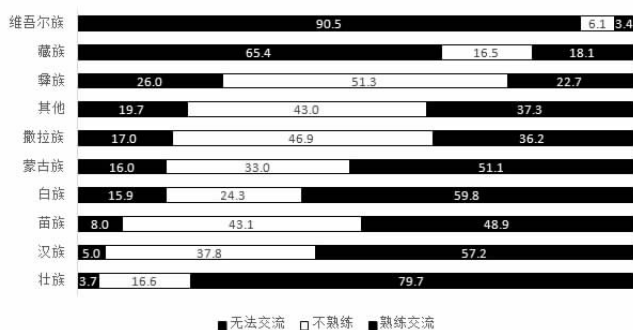


图1 主要民族的普通话水平

由表2可见，普通话水平在性别之间并无显著差异，但不同年龄和受教育程度的劳动力普通话水平差异显著。越年轻，能熟练使用普通话的比例越高，但在最年轻的16—25岁年龄段中，仍然有19%的人无法用普通话交流。个人受教育程度越高，相应的普通话水平也越高。

表2 民族地区少数民族和汉族劳动力的普通话口语交流能力 单位：%；个

	无法交流	不熟练但可交流	熟练交流	合计
性别：				
男	25.9	35.9	38.2	5589
女	29.2	32.4	28.4	3268
年龄：				
16—25岁	18.8	28.2	53.0	660
26—35岁	22.6	31.2	46.3	1778
36—45岁	26.6	33.0	40.4	2537
46—55岁	29.3	35.9	34.8	2167
55岁以上	33.2	41.3	25.5	1715
教育：				
小学以下	48.3	37.0	14.7	1732

^① 参见郝时远：《铸牢中华民族共同体意识必须推广国家通用语言文字》，《人民日报》2018年10月31日。

	无法交流	不熟练但可交流	熟练交流	合计
小学	33.3	38.7	28.1	3065
初中	15.6	33.7	50.7	2851
高中	9.5	25.2	65.3	888
大专及以上	5.9	16.5	77.6	321
少数民族人口占比:				
<30%	8.1	32.1	59.8	1456
[30%,65%]	8.4	52.1	39.5	1687
>65%	37.5	30.1	32.4	5714

民族地区中少数民族的聚居状况不尽相同,分析数据中县域少数民族人口比例在 6.22% 至 99.12% 之间波动。为更好地分析少数民族不同聚居程度下普通话口语能力对非农就业的影响,本文基于 2010 年人口普查数据,将少数民族人口比例在 30% 以下的县旗界定为一般聚居区,这一标准也是参照了国家对民族乡少数民族人口例(30%)的要求;中等聚居区的少数民族人口比例为 31%—65%;^①密集聚居区的少数民族人口比例为 66%—100%。大调查数据结果表明,三类地区随着少数民族人口比例的上升,能熟练使用普通话的比例也随之下降了 27 个百分点,无法交流的比例增加了近 30 个百分点。

如表 3 所示,民族地区农村劳动力中 60.3% 从事纯农业,26.1% 是兼业,13.6% 为纯非农就业(本文中的纯非农就业均为在县/旗内的纯非农就业)。根据第三次全国农业普查结果,全国农村住户就业人口的纯农业仅占 36%,兼业为 18%,纯非农占 45%。^② 民族地区的就业结构与全国平均水平相比,非农就业比重低 23 个百分点,纯非农就业差距更大,偏低 31 个百分点。

表 3 民族地区少数民族和汉族劳动力不同就业类型下的普通话口语交流能力 单位:%

	纯农业	兼业	纯非农
合计	60.3	26.1	13.6
民族:			
少数民族	63.1	26.1	10.8
汉族	47.8	26.3	25.9
性别:			
男	58.5	28.8	12.7
女	63.4	21.5	15.1
年龄:			
16—25 岁	43.5	31.1	25.5
26—35 岁	51.3	31.6	17.2

^① 65% 的比例是总样本中每个县少数民族人口比例的均值,因此将其作为分类标准之一。

^② 陈传波等:《关于农村劳动力就业情况的研究报告——基于全国农业普查资料的分析》,《第三次农业普查招标课题研究报告》,2018 年。

	纯农业	兼业	纯非农
36—45岁	56.0	29.2	14.8
46—55岁	63.6	25.2	11.2
55岁以上	78.1	15.2	6.7
教育:			
小学以下	76.2	17.3	6.5
小学	69.2	22.2	8.7
初中	52.8	32.6	14.7
高中	37.5	34.2	28.3
大专及以上	19.0	32.1	48.9

表4报告了民族地区的劳动力在纯农业、兼业和纯非农就业方面的差异。

表4 民族地区劳动力分普通话口语交流能力的就业分布 单位:%;个

	无法交流	不熟练但可交流	熟练交流
纯农业	79.5	63.5	43.8
兼业	14.7	26.9	33.5
纯非农	5.8	9.7	22.7
合计样本数	100.0 (2404)	100.1 (3064)	100.0 (3389)

由表4可见,民族地区农村劳动力的普通话水平与非农就业存在相关关系。在无法用普通话交流的劳动力中,有79.5%从事的是纯农业,纯非农就业仅占5.8%。普通话不熟练但可以交流的劳动力中,从事非农就业的比率明显高于不能用普通话交流的劳动力,分别有26.9%和9.7%的劳动力属于兼业和纯非农就业。而在能熟练用普通话交流的劳动力中,有56.2%为非农就业,包括33.5%兼业和22.7%的纯非农就业。

三、实证研究模型

上述描述分析中发现,普通话口语水平与非农就业之间存在显著的相关关系,但我们更感兴趣的是两者之间是否存在因果关系,提升普通话水平能否增加非农就业概率?如果能,能提升多少?然而,估计普通话水平对非农就业的影响,若采用普通最小二乘回归方法或者多元选择模型,均无法克服内生性问题。这里的内生性既源于反向因果关系,也源于内生选择。一方面,普通话水平会影响到非农就业选择;另一方面,非农就业机会也会影响普通话水平。即当人们预期有较多的非农就业机会,而这些机会又需要具备普通话能力时,自然就会更努力地提升其普通话水平。

为了应对上述内生性问题,我们采用扩展回归模型(ERM^①)中的二元选择模型

① 参见 W. W. Gould, "Ermistatas and Stata's New ERMs Commands," The Stata Blog: Not Elsewhere Classified, 2018, <https://blog.stata.com/2018/03/27/ermistatas-and-statas-new-erms-commands/>.

(EOPROBIT)来估计普通话水平对非农就业的影响程度,并将普通话水平设定为一个多水平内生处理效应,模型设定如下:

$$offfarm_j = 1(x'\gamma_j + u_j > 0) \text{ where } j = 0, 1, 2 \quad (1)$$

$$offfarm = \sum_{j=0}^2 1(\text{Mandarin} = j) offfarm_j \quad (2)$$

$$\text{Mandarin} = \begin{cases} 0 & \text{if } z'\alpha + v < cut_1 \\ 1 & \text{if } cut_1 \leq z'\alpha + v < cut_2 \\ 2 & \text{if } z'\alpha + v \geq cut_2 \end{cases} \quad (3)$$

$$\begin{bmatrix} u_j \\ v \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}\right) \quad (4)$$

其中

$$x'\gamma = \gamma_1 female + \gamma_2 age + \gamma_3 education + \gamma_4 health + \gamma_5 Han + \gamma_6 Minority + \gamma_7 Minority^2 + \gamma_8 hhsz + \gamma_0 + u \quad (5)$$

$$z'\alpha = \alpha_0 + \alpha_1 Born + \alpha_2 female + \alpha_3 edu + \alpha_4 age + \alpha_5 Minority \quad (6)$$

方程(1)是潜变量模型,实际上包含三个子方程,其中j取值0,1,2三个值,对应着三种不同的普通话水平,即Mandarin的取值为0,1,2,分别表示无法用普通话交流、能交流但不熟练以及能熟练交流。对每一个劳动者,对应于不同的普通话水平,假设其有三种潜在的不同的就业选择的净效用,如当他无法用普通话交流时(j=0),若他选择非农就业的净效用大于0时,他将选择非农就业,此时 $offfarm_0 = 1$,否则 $offfarm_j = 0$;当他能用普通话熟练交流时(j=2),若他选择非农就业的净效用大于0,他将选择非农就业,此时 $offfarm_2 = 1$ 。

现实中,我们都只能观察到一种状态,方程2将潜变量结果与实际观察到的处理联系起来,即如果实际观察到的是他能用普通话熟练交流,即j=2,那么是否从事非农就业 $offfarm$ 的取值就对应于 $offfarm_2$ 的取值,若为非农就业,则取值为1,否则为0,其中非农就业包含兼业和纯非农就业两类情形。

方程(3)为Mandarin的控制方程,z为工具变量集,其中一个外生工具变量Born是小时候最先会说的语言,有五个取值:普通话、汉语方言、本民族语言、其他少数民族语言和其他语言,若最先会说的是普通话和汉族方言,则Born取值为1,否则为0。从表5可知,最先会说的语言与普通话水平是相关的,因此满足工具变量的相关性条件;同时,因为当前的就业状态并不能影响儿童时期的普通话水平,因此,它也满足工具变量的外生性条件。

表5 最先会说的语言与普通话水平

单位:%;个

普通话水平	最先会说民族语言	最先会说普通话	合计
无法交流	37.8	4.1	27.1
不熟练但可交流	30.6	43.1	34.6
熟练交流	31.5	52.8	38.3
合计 (样本量)	100.0 (6051)	100.0 (2806)	100.0 (8857)

方程(3)中的 cut_1 和 cut_2 是一个待估参数,我们将普通话水平视为一个序次变量,当影响普通话水平的外生变量z的线性组合值小于 cut_1 时,我们将观察到其普通话水平为无法交流;

若线性组合值位于 cut_1 和 cut_2 之间,则它的普通话不熟练但可交流;若大于 cut_2 则能熟练交流。

方程(4)包含三组服从二元标准正态分布的随机向量。 u_j 和 v 为误差项,ERM 模型假定这两个误差项服从二元正态分布,两者的相关系数为 ρ_1 ,这个相关系数可以被估计出来,如果其估计值显著异于零,则意味着存在内生性。

方程(5)是方程(1)中的外生变量的线性组合,其中 x 为外生的控制变量,包括性别 female、年龄 age、受教育程度 education、家庭人口 hhsiz、健康状况 health、少数民族占比 minority 等。此外,性别、年龄、受教育程度和当地少数民族人口占比也会影响到普通话水平。

为了进一步分析普通话水平对兼业和纯非农就业的影响,我们进一步设定以下模型:

$$ordinalofffarm = \begin{cases} 0 & \text{if } \beta \times \text{Mandarin} + x'\gamma + u < c_1 \\ 1 & \text{if } c_1 \leq \beta \times \text{Mandarin} + x'\gamma + u \leq c_2 \\ 2 & \text{if } \beta \times \text{Mandarin} + x'\gamma + u > c_2 \end{cases} \quad (7)$$

其中 $ordinalofffarm$ 为序次变量,取值 0,1,2,分别为纯农业、兼业和纯非农就业。普通话水平 Mandarin 的控制方程与上述二元选择模型相同,不再列示。由于因变量为序次变量,因此我们采用 ERM 模型族中的有序多元选择模型(EOPROBIT)模型中的处理效应模型来估计。

四、分析结果

本文共估计了四个扩展回归模型,其中模型 I 的因变量为非农就业(0 和 1 选择),并假定普通话水平为外生变量,采用二元外生处理模型估计。模型 II 的因变量与模型 I 相同,但假定普通话水平为内生变量,采用二元内生处理模型估计。模型 III 的因变量为三类就业选择(纯农、兼业和非农),并假定普通话水平为外生处理,采用序次多元选择外生处理模型估计。模型 IV 的因变量与模型 III 相同,但假定普通话水平为内生变量,采用序次多元选择内生处理模型估计。估计结果见表 6,我们删除了基准组。

表 6 普通话对非农就业的影响估计

		I 外生处理	II 内生处理	III 外生处理	IV 内生处理
自变量	因变量	非农就业	非农就业	三类就业	三类就业
普通话水平	无法交流	-0.354		-1.054***	
	不熟练	0.279		-0.015	
	熟练交流	0.3		0.609***	
性别女	无法交流	-0.072	-0.077	-0.042	-0.05
	不熟练	-0.183***	-0.173***	-0.09	-0.084
女	熟练交流	-0.105*	-0.108*	0.001	-0.009
教育 小学	无法交流	-0.131	-0.213**	-0.126	-0.223***
	不熟练	0.097	-0.03	0.089	-0.074
	熟练交流	0.195*	0.073	0.190*	0.034

		I 外生处理	II 内生处理	III 外生处理	IV 内生处理
初中	无法交流	0.026	-0.196*	0.018	-0.251**
初中	不熟练	0.269***	-0.036	0.259***	-0.129
初中	熟练交流	0.445***	0.185	0.390***	0.054
高中	无法交流	0.178	-0.158	0.159	-0.253
高中	不熟练	0.627***	0.161	0.635***	0.034
高中	熟练交流	0.765***	0.407***	0.784***	0.308**
大专及以上	无法交流	1.226***	0.716*	1.027***	0.406
大专及以上	不熟练	0.884***	0.301	0.927***	0.179
大专及以上	熟练交流	1.090***	0.692***	1.121***	0.578***
民族汉族	无法交流	0.555***	0.411**	0.536***	0.352**
	不熟练	0.332***	0.165*	0.359***	0.141*
	熟练交流	0.175**	0.079	0.302***	0.169***
健康状况良好	无法交流	0.115	0.109	0.089	0.084
	不熟练	0.280***	0.259***	0.242**	0.212**
	熟练交流	0.358***	0.344***	0.354***	0.328***
年 龄	无法交流	-0.009***	-0.007**	-0.010***	-0.007**
	不熟练	-0.018***	-0.013***	-0.016***	-0.010***
	熟练交流	-0.016***	-0.013***	-0.015***	-0.010***
家庭人口数	无法交流	0.002	0.002	0	0
	不熟练	0	0.001	-0.008	-0.007
	熟练交流	0.001	0.004	-0.007	-0.003
少数民族占比	无法交流	0.006	0.011	0	0.006
	不熟练	-0.015***	-0.012**	-0.023***	-0.018***
	熟练交流	-0.018***	-0.016***	-0.025***	-0.021***
少数民族占比平方	无法交流	0	-0.000*	0	0
	不熟练	0.000***	0.000***	0.000***	0.000***
	熟练交流	0.000***	0.000***	0.000***	0.000***
控制方程因变量			普通话水平		普通话水平
性 别	女		0.032		0.031
首先学会的语言	汉语		0.541***		0.541***
少数民族占比			-0.012***		-0.012***
教 育	小学		0.379***		0.380***
	初中		0.832***		0.831***
	高中		1.182***		1.180***

		I 外生处理	II 内生处理	III 外生处理	IV 内生处理
	大专及以上		1.424***		1.419***
年 龄			-0.008***		-0.008***
分割点 1	无法交流			0.089	0.974***
	不熟练			-0.413*	-0.028
	熟练交流			-0.419*	-0.782***
分割点 2	无法交流			0.870***	1.693***
	不熟练			0.614***	0.877***
	熟练交流			0.580***	0.133
误差相关系数			-0.408***		-0.511***
Wald chi2		1675.63***	2040.46***	863.23***	488.82***
样本量		8857	8857	8857	8857

注：*、**和***分别表示在10%、5%和1%的水平下显著。

从表6可以总结出以下几点结论：

(1)模型I假定普通话水平为外生处理,但模型估计发现,当控制了教育、年龄、性别等因素后,甚至在10%的显著性水平上,普通话水平对非农就业也没有显著影响。

(2)模型II假定普通话水平为内生处理,主方程与处理方程的误差项的相关系数估计为-0.408,且在1%的统计显著性水平上显著,说明普通话水平是主方程中的内生处理变量,应该采用内生处理模型来分析。但内生处理模型无法直接给出处理效应,我们将在下文进一步估计。

(3)模型III表明,即使将普通话水平视为外生处理,它对三类就业的选择也具有显著影响,但模型IV中主方程与处理方程的误差项的相关系数估计达到-0.511,且在1%的统计显著性水平上显著,说明内生处理模型是更合适的选择。

(4)在外生控制变量中,性别和家庭人口对非农就业的影响不显著,其他变量均有显著的影响,但在不同的处理效应水平模型中,影响显著性和影响程度不同。

表7 普通话水平对非农就业的处理效应估计

处理效应	I 外生处理	II 内生处理
不熟练 vs 无法交流	0.105*** (0.017)	0.222*** (0.022)
熟练交流 vs 无法交流	0.218*** (0.017)	0.518*** (0.051)

注：*、**和***分别表示在10%、5%和1%的水平下显著。

表7报告了普通话水平对非农就业的处理效应的估计结果,采用工具变量考虑内生性偏误后的估计效应大约是外生处理效应的两倍。内生处理模型的结果表明,如果能将样本中所有人的普通话水平提升至不熟练但可以交流的程度,将能使非农就业概率提高22%,若能进一步提升至熟练交流水平,相对于普通话水平无法交流,可以使非农就业的概率提高至52%,这是非常显著的影响。

表 8 报告了普通话水平对三类就业的处理效应估计结果,如果样本中的农村劳动力的普通话水平均从无法交流提高到不熟练但可以交流的程度,可以使纯农就业概率下降 25%,相应使兼业和纯非农就业概率提升 16%和 8%。如果进一步将普通话水平都提升至熟练的程度,将使得纯农就业下降 58%,兼业提高 24%,纯非农就业提高 34%。

表 8 普通话水平对三类就业的处理效应估计

外生处理效应	内生处理效应			纯农业	兼业	纯非农
	纯农业	兼业	纯非农			
不熟练 vs 无法交流	-0.108*** (0.016)	0.097*** (0.012)	0.011 (0.012)	-0.249*** (0.017)	0.164*** (0.010)	0.084*** (0.012)
熟练交流 vs 无法交流	-0.214*** (0.017)	0.134*** (0.012)	0.080*** (0.012)	-0.580*** (0.382)	0.244*** (0.010)	0.336*** (0.037)

注: *、** 和 *** 分别表示在 10%、5% 和 1% 的水平下显著。

有研究证实,不同性别由于存在语言态度的不同从而掌握国家通用语言的能力也有所差异。^① 而且,同样的语言能力对不同性别经济影响也会有明显区别。^② 而女性由于从事第三产业的比例高,因而受到语言的影响就更大。^③ 但我们的研究没有发现语言限制对农村劳动力非农就业影响的程度存在性别差异。

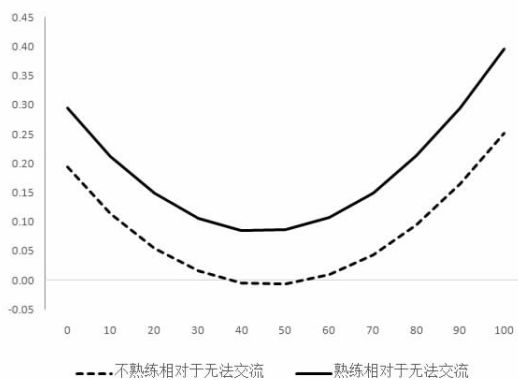


图 2 普通话水平对非农就业概率的影响

上图给出了随着少数民族聚居程度的变化,普通话水平从无法交流转变为不熟练交流和熟练交流后的非农就业概率变化情况,即随着县域少数民族人口占全县人口比重的提升,普通话水平对非农就业的影响程度呈现出先下降后上升的 U 形。

在少数民族人口比例占一半的县(旗),普通话水平对非农就业的影响最小;当少数民族人口占比超过一半之后,随着少数民族人口所占比重的提升,普通话水平对非农就业概率的影响

① 参见王桂荣:《通辽地区蒙古族学生的语言使用及语言态度调查——以舍伯吐镇为例》,内蒙古大学硕士论文,2013年;褚慧:《改革开放 40 年我国少数民族双语教育研究热点、历史过程与展望》,《信阳师范学院学报》2020 年第 1 期。

② 参见 Joanne Lindley, “The English Language Fluency and Earnings of Ethnic Minorities in Britain,” *Scottish Journal of Political Economy*, Vol. 49, No. 4, 2002; Teresa Casey and Christian Dustmann, “Intergenerational Transmission of Language Capital and Economic Outcomes,” *The Journal of Human Resources*, Vol. 43, No. 3, 2008。

③ 参见 Wenshu Gao and Russell Smyth, “Economic Returns to Speaking ‘Standard Mandarin’ among Migrants in China’s Urban Labour Market,” *Economics of Education Review*, Vol. 30, No. 2, 2011。

越来越大;当全部是少数民族时,不能用普通话交流的劳动力,假设都能熟练讲普通话,其非农就业概率将增加40%,即使普通话不熟练,也能使非农就业概率提升25%。

普通话水平对非农就业概率的影响在少数民族人口占比较少(低于一半)时,会随着少数民族人口占比的下降而呈上升趋势。对这一现象的解释,还需要更多的后续研究。需要指出的是,普通话水平和汉族方言是有差异的,随着讲地方方言的汉族人口比重上升,与少数民族通过普通话交流的需求可能会趋于下降,从而使得方言变得更难为外人所理解,进而对非农就业带来负面影响。

上述结果也证实了少数民族密集聚居区普通话水平对于非农就业的影响最大;而且随着少数民族人口密度的增加,普通话水平的正向影响越明显。结合在民族地区实地调研经验,可解释为少数民族密集聚居区普通话水平作为人力资本的重要方面在获得非农就业机会上具有非常明显的优势。首先,少数民族密集聚居区开展的精准扶贫和对口帮扶中的大型基础设施项目建设,主要是来自东中部地区的承建单位,为工作交流方便,也为了便于上岗培训,需要雇佣懂普通话的当地少数民族劳动力,这也成了当地获得非农就业机会的主渠道。其次,在少数民族人口比例低的一般聚居区,汉族人口比例高,普通话应是就业中最主要的交流语言,如果普通话水平低,相应地找到非农就业机会的概率也低。再次,在少数民族集中聚居区,由于少数民族和汉族人口均占有相当比例,非农就业中普通话的要求相对于少数民族一般聚居区要低,而且表明汉族农村劳动力的普通话水平也需提升。从现实经验出发,由于农村非农就业通常对劳动技能的学习要求不高,只要非农就业场所中有懂普通话的少数民族,其他不懂普通话的少数民族也可以完成工作。这意味着推广普通话以提升非农就业的重点是在少数民族聚居程度高的地区,但少数民族一般和集中聚居区的普通话推广和个人掌握程度的提升难度可能更大。我们不仅要重视少数民族的普通话水平提升,也需要高度重视民族地区的汉族的普通话水平提升,特别是在少数民族分布稀疏的汉族方言区。

五、结 论

众所周知,普通话水平偏低对个人而言会形成语言沟通障碍,在信息时代不仅难以获取相关信息,无法有效提升人力资本,对外交往圈狭小,而且非常不利于个人及其家庭的收入与发展。从宏观层面而言,民族地区各民族普通话掌握程度是民族地区加快全面发展,实现社会主义现代化和铸牢中华民族共同体意识所必需的前提条件,也直接影响到我国各个民族间的交往交流交融。本文基于中国社会科学院民族学与人类学研究所主持的国家社科基金特别委托暨中国社会科学院创新工程重大专项“21世纪初中国少数民族地区经济社会发展综合调查”课题组于2013—2015年的大样本调查数据,实证分析了民族地区农村劳动力普通话水平对非农就业的影响。主要结论如下:

首先,民族地区农村16—60岁的劳动力中无法用普通话交流的比例为27.1%,少数民族这一比例达到了32.2%。不同少数民族该比例存在差异,和其他少数民族相比,藏族和维吾尔族该比例最低。从南北地域看,与南方相比,北方少数民族无法用普通话交流的比例偏高。总体上,民族地区农村劳动力越年轻,能熟练使用普通话的比例越高,但在最年轻的16—25岁年龄段中,仍然有19%的人无法用普通话交流。劳动力受教育程度越高,相应的普通话水平也越高。以少数民族人口比例划分的一般、集中和密集三类聚居区中,少数民族人口比例越

高,其掌握普通话的比例越明显下降。

其次,普通话水平对非农就业的正向促进作用显著。在无法用普通话交流的劳动力中,有79.5%从事的是纯农业,纯非农就业仅占5.8%。普通话不熟练但可以交流的劳动力中,从事非农就业的比率提升至36.6%,其中9.7%的劳动力仅从事非农就业。而在能熟练用普通话交流的劳动力中,有56.2%为非农就业,包括33.5%的兼业和22.7%的纯非农就业。通过实证分析后发现,如果农村劳动力的普通话水平均从无法交流提高到不熟练但可以交流的程度,可以使纯农就业概率下降25%,非农就业概率提高22%,相应使兼业和纯非农就业概率提升16%和8%。如果进一步将普通话水平都提升至熟练的程度,将使得纯农就业下降58%,使非农就业的概率提高至52%,兼业提高24%,纯非农就业提高34%。

最后,本文证实,随着县域少数民族人口占全县人口比重的提升,普通话水平对非农就业的影响程度呈现出先下降后增加的U形,也就是说少数民族密集聚居区普通话水平对于非农就业的影响最大。而且随着少数民族人口密度的增加,普通话水平的正向影响越明显。在少数民族人口比例低的一般聚居区,汉族人口比例高,普通话应是就业中最主要的交流语言,如果普通话水平低,相应地找到非农就业机会的概率也低。而在少数民族集中聚居区的普通话水平对非农就业的影响程度相对最低,这也表明民族地区中的汉族普通话水平也需提升。

综上所述,本文的研究证实了在民族地区从人力资本角度看,无论是少数民族还是汉族学习并掌握普通话都是对其知识和技能的一种投资,能带来经济收益。换言之,农村少数民族和汉族劳动力的普通话能力强会使其更容易外出找到非农就业机会,更方便生活和积累社会关系。从宏观层面看,普通话能力强会增加非农就业的可能性,将有利于民族间交往交流交融,从而加强铸牢中华民族共同体意识。这也正是我国在民族地区推广国家通用语言——普通话的政治、社会和经济意义。

本文相应的政策建议为,在国家大力推广国家通用语言——普通话时,不仅要着重加强少数民族的普通话水平,同时也需重视民族地区的汉族普通话水平提升,特别是在少数民族人口分布较少的汉族方言区。在全国市场一体化加快形成的当下和今后,民族地区,乃至全国的少数民族和汉族应形成不断提高普通话交流能力的自觉意识,以提升自身的环境适应性和就业竞争能力。此外,熟练使用国家通用语言——普通话也是推进铸牢中华民族共同体意识工作的应有之义。

〔责任编辑 马 骅〕