

DOI: 10.19361/j.er.2016.06.09

普通话能力对中国劳动者收入的影响

陈媛媛*

摘要: 语言作为一种人力资本,它对收入很可能会产生重要的作用。本文采用中国家庭动态跟踪调查(CFPS)、中国劳动力动态调查(CLDs)以及中国综合社会调查(CGSS)三套中国的综合数据,检验了普通话能力对劳动者收入的影响。结果发现,总体上,普通话以及单项能力(听和说)都会对劳动者收入产生比较显著的影响,在考虑了内生性之后,结论未变。进一步发现,普通话对服务业从业人员的工资有显著提升作用,但是对其他行业的影响则不显著。另外,普通话对城镇劳动者收入的影响比较显著,对农村劳动者收入的影响并不显著。此外,普通话能力在南方的作用相对北方要更大一些,且普通话对跨方言流动人口的影响要大于对方言区内人口的影响。

关键词: 普通话;收入;工具变量;PSM

一、引言

语言技能是一种重要的人力资本形式(Chiswick and Miller, 2003)。语言知识是知识资本的积累(Grenier and Vaillancourt, 1983),与教育和健康投资类似,它的获取是需要一定代价或者成本的,这些所花费的时间和金钱是对人力资本的一种特殊投资。这种投资一旦见效,人们就会有更多的交际可能,能够获得更好的求职信息,与潜在的招聘者更好地交流自己的特长和技能(Gao and Smyth, 2011),因此获得更高的收入或者更好的职位(Lazear, 1999; Chiswick and Miller, 2014)。

国内外很多实证研究发现语言对劳动者收入的提高会起到明显的作用。有些文献发现掌握一门外语会帮助劳动者提高收入(Di Paola and Tansel, 2015)。比如英语对于官方语言非英语的国家会有一个“收入溢价”的现象(Azam et al., 2013; Casale and Posel, 2011)。如果劳动者有双语能力,那么他获得的工资也会比不会双语的劳动者要高(Carlner, 1981)。单项语言能力(比如听、说、读、写)对收入也有一定的提高作用(Canevale et al., 2001)。还有些文献认为熟练掌握一国的流行语言(major language)也会显著提高劳动者的收入,比如中国的普通话就会提高迁移者的工作收入(Gao and Smyth, 2011)。

以往关于语言对收入(或劳动生产率)影响的研究很多以欧美跨国移民的经验研究为基

*陈媛媛,山东大学经济研究院,邮政编码:250101,电子信箱:whncy@163.com。

本文受到国家自然科学基金青年项目“特定制度安排下的FDI对我国环境质量的影响”(项目号:71403146)的资助。

础。但是,在中国同样存在着语言和文化的多样性。普通话和方言之间的关系与发达国家英语和其他语种之间的关系在一定程度上具有相似性^①。中国有55个少数民族,许多少数民族都有自己的语言,中国政府也鼓励在少数民族地区使用并发展这类语言(Zhou, 2003)。中国有七大方言区以及上百个次方言区(sub-dialects)(Leung and Ruan, 2012)^②。这些方言在语法、词汇、音韵和发音上都有所不同,因此不同方言区的人们在沟通上总是会有或多或少的困难。人们很可能听不懂离自己家乡100公里外的地区所使用的方言,甚至同一省份的县与县之间也存在不同的方言(Gong et al., 2011)。交流上存在障碍,对于找工作的效率、匹配程度、选择范围以及入职后的工作效率都可能存在负面影响,进而影响劳动效率和工资。尤其是随着中国流动人口的不断增加,不同方言背景的群体大量融入城市,如果仅靠方言,人们很难进行跨方言区或者跨省的沟通,而普通话作为通用语言,其重要性更加凸显。早在1956年,国务院就发布了《关于推广普通话的指示》来推广普通话的使用。这样,即使人们听不懂异于自己家乡的方言或少数民族语言,但是如果能听懂普通话,同样也可以交流。

普通话或者主要语言对于劳动者经济收入的影响主要包括以下几个方面:一是提高工作寻找效率(Chiswick and Miller, 2014)。熟练的普通话可以帮助受雇者更好地了解自己的各项技能,进而更好地找到其能力和技能与雇佣者要求相匹配的工作,而在很多招聘过程中,也要求受雇者使用普通话,这样更方便交流,方便信息更快更有效地传递。二是提高工作效率(Stöhr, 2015; Chiswick and Miller, 2014)。熟练的大众语言能够使受雇者与同事、领导以及客户之间更好地交流,进而提高工作效率。三是可以扩大社会网络(Wang et al., 2016),增加受雇者接触更多群体和阶层的机会,从而获得更多的经验、知识和技能的积累。对于外来移民者来说,另外一个好处是说普通话利于减少当地消费者和雇佣者对他们的歧视(Gao and Smyth, 2011)。

Gao 和 Smyth (2011) 经研究发现普通话对中国农民工工资有着积极影响,但他们没有考察普通话对本地居民或者外来市民是否也有相同的效应。事实上,这两类群体在就业中占据的份额也很重要。因此,本文打算在前文研究的基础上,考察普通话对总体劳动者的影响。为了验证估计结果的稳健性,我们采用三套微观数据(即CFPS、CGSS以及CLDS)分别研究,同时采用工具变量法和倾向得分匹配模型(PSM)来克服模型的内生性。

本文的主要安排如下:第二部分是文献综述,第三部分是数据介绍以及计量模型的选择,第四部分是计量结果及分析,第五部分是主要结论。

二、文献综述

关于语言对劳动者收入影响的文献大概可以分为三类:

一是研究外语对劳动者收入的影响,很多研究选择的外语种类都为英语(Azam et al.,

①虽然中国很多地区都存在方言,但很多人从小可能有机会接触到普通话,比如在学校里,即使普通话表达能力较差,但听写一般没有问题。然而,跨国移民“听说写”可能是一体的,若没有接触过流入地语言,可能任何一项语言能力都没有。所以,跨国移民中语言的影响很可能要大于国内人口流动中普通话的影响。

②主要包括北方方言以及南方地区的吴、湘、客家、闽、粤、赣方言。

2013; Casale and Posel, 2011; Donado, 2014; Grin, 2001; Isphording, 2013; Levinshon, 2007; Stöhr, 2015; Toomet, 2011; Williams, 2011)。且多数文献都发现在官方语言非英语的国家,相对不会说英语的劳动者,会说英语的劳动者工资水平要高。例如 Azam 等(2013)发现掌握高级英语技能的印度男性的“工资溢价”程度高达 35%。会说英语的南非非移民劳动者其工资回报要比不会说英语的劳动者高出 18%~44% (Levinsohn, 2007)。Toomet (2011)发现在爱沙尼亚,会说英语的就业者工资水平要比不会说英语的就业者高出 45%,而这一数字在拉脱维亚高达 62%。Guo 和 Sun(2014)考察了在中国会说英语的劳动者收入回报情况,其研究对象仅限于大学毕业生。刘泉(2014)也发现中国的外语熟练劳动者工资要比外语不熟练劳动者工资高出大约 69%。

二是研究双语或多语对收入的影响(K.Pendakur and R.Pendakur, 1998; Fry and Lowell, 2003; Saiz and Zoida, 2005)。例如 K.Pendakur 和 R.Pendakur (1998)就发现会双语的人能够更容易地获得其他技能,其就业率也相对较高。同时增加他们从事国际贸易和旅游工作的机会,提高他们的竞争优势。有些工作需要劳动者经常与他人(消费者、供应商等)进行口头交流,还有些工作要求劳动者具有一定的读写能力来安全、高效地完成工作。对于这两类工作,双语者比单语者更容易就业,调换工作也相对容易。

三是研究少数语言(方言)或者大众语言对劳动者收入的影响(Chiswick and Miller, 2007)。有些文献发现相对只会大众语言的就业者,那些会少数语言和大众语言的就业者,其经济回报并没有明显提高。可能的原因是他们使用的大众语言并不如只会大众语言就业者那么流利;或者他们说大众语言可能有一些口音,导致在就业中受到语言歧视。K.Pendakur 和 R.Pendakur (1998)研究了加拿大少数民族语言的经济回报,发现他们的工作回报率要低于使用大众语言的就业者。Chen 等(2014)研究了上海方言对外来移民的影响。结果发现,方言的使用会显著提高服务行业的收入,另外,方言口语的使用会显著影响就业者收入,但听力对收入的影响却不显著。Gao 和 Smyth (2011)研究了大众语言普通话对中国农村-城市的移民者的收入情况的影响,结果发现,普通话的熟练程度可以显著提高他们的收入,同时普通话对女性收入的影响显著,对男性则不显著。文献解释可能的原因是女性更多参与到了与本地人接触较多的职业中。换句话说,行业的差别导致普通话的重要性有所不同,有些行业对普通话的依赖性更强。该文献与本文所涉及的研究最为接近,但他们只是考察了移民者。本文则在这些文献的基础上,考察普通话对不同职业、不同地域的劳动者收入是否会产生不同的影响,另外我们采用三套微观数据进行稳健性分析。

三、研究设计及数据

(一) 数据介绍

本文使用三套比较权威的微观数据,即 2012 年“中国家庭动态跟踪调查”数据、2012 年“中国劳动力动态调查”数据以及 2013 年“中国综合社会调查”数据。

“中国家庭动态跟踪调查”(Chinese Family Panel Studies, CFPS)是北京大学中国社会科学调查中心(ISSS)实施的一项旨在通过跟踪搜集个体、家庭、社区三个层次的数据,反映中国社会、经济、人口、教育和健康的变迁。2010 年起在全国(西藏、青海、新疆、宁夏、内蒙古、海南、香港、澳门、台湾不在其列)正式实施,每年一次跟踪调查。

“中国劳动力动态调查”(China Labor-force Dynamics Survey, CLDS)是由中山大学实施的数据调查。CLDS 样本覆盖中国 29 个省份(除港澳台地区、西藏、海南外),调查对象为样本家庭户中的全部劳动力(年龄 15 至 64 岁的家庭成员)。目的是通过对中国城乡以村/居为追踪范围的家庭、劳动力个体开展每两年一次的动态追踪调查,系统地监测村/居社区的社会结构和家庭、劳动力个体的变化与相互影响。

“中国综合社会调查”(Chinese General Social Survey, CGSS)是中国第一个全国性、综合性、连续性的大型社会调查项目,在国际社会调查项目(ISSP)的 2007 年年会上,中国人民大学社会学系和香港科技大学社会调查中心作为 CGSS 的共同负责单位,被接纳为代表中国的会员。CGSS 系统、全面地收集社会、社区、家庭、个人多个层次的数据,2013 年调查在全国一共抽取了 100 个县(区),加上北京、上海、天津、广州、深圳 5 个大城市,作为初级抽样单元。

以上三套数据中,作者仅保留了有工作和普通话信息的群体,其中包括自我雇佣和受雇于人的样本(不包括退休人口)以及农业户口和非农业户口样本。同时还删除了 65 岁以上以及 15 岁以下的样本,他们不属于劳动力年龄人口。

(二) 研究设计

本文主要研究普通话能力对劳动者收入的影响,因此计量模型的选取是基于标准的明瑟(Mincer)半对数收入方程。

$$\log wage_{it} = c + \beta_1 \text{Mandarin}_{it} + \sum_l \beta_l \text{control}_{lit} + \mu_i + \delta_t + \varepsilon_{it}$$

其中,下标 i 指个体, t 指时间, μ_i 指不随时间变化的个体效应, δ_t 是不随个体变化的时间效应, ε_{it} 是扰动项。

这里的变量 $wage$ 是被解释变量,是每小时的工资收入。选取小时工资而非月工资或日工资可以排除因劳动者工作时间不同而造成的工资差异。 $Mandarin$ 是核心解释变量,表示受访者普通话的流利程度,在 CFPS 中,其取值范围是 0-5,不会说=0,说得最好=5。在 CLDS 中,其取值范围是 1-5。在 CGSS 中,考察了普通话的单项能力,其中普通话的听力能力,其取值范围是 1-5,不会=1,很好=5,普通话表达能力的度量方法类似。另外,在 CLDS 中,普通话能力是由访员根据受访者的表现对其进行评估的指标,较为客观。另外两个数据库的指标都是受访者的主观自评。 $control$ 表示其他控制变量^①,包括如下若干个影响收入水平的变量: age 是居民的年龄。 $gender$ 是性别,男性=1,女性=0。 edu 是受教育年限,取值范围是 0-22 年。 $mar1$ 、 $mar2$ 、 $mar3$ 是表示婚姻状况的虚拟变量,其中 $mar1$ 的取值是单身=1,其他=0; $mar2$ 的取值是结婚=1,其他=0; $mar3$ 的取值是离婚或丧偶=1,其他=0。 $hukou1$ 表示农业户口,其他=0; $hukou2$ 表示本地居民,其他=0。 $health$ 是健康状况,被访者认为自己非常健康=1,很健康=2,比较健康=3,一般=4,不健康=5。 $healthdif$ 表示健康变化,健康状况比一年前更好=1,没有变化=2,比一年前更差=3。 $fatheredu$ 表示父亲的受教育年限,

①本文中,控制变量我们主要介绍 CFPS 数据库里变量的选取度量方式。CLDS 和 CGSS 里选取方式类似,只是由于问卷设计的问题,其具体度量有细微差别。有些控制变量在某套数据库中有,在其他数据库中可能没有。比如 $contract$ (与用工企业是否签订合同,签订=1,没有签订=0)和 $union$ (所在单位是否有工会,有工会=1,无工会=0)这两个变量 CFPS 中就没有。

度量方法与 *edu* 一致。*sociallevel* 是受访者在本地的社会地位,取值范围是 1-5,1 是很低,5 是很高。*confidence* 是对自己未来信心程度,取值 1-5,5 是信心程度最高。*urban* 表示城乡变量,其中城、镇、县 = 1,农村、郊区 = 0。*jobtotal* 表示工作性质,其类别分为管理、技术、服务、生产、其他和无工作。*indtotal* 表示行业类别,包括制造、建筑、服务、科技和其他。*province* 是按照省国标码编码的省级变量。

(三)主要变量描述性统计

表 1 列出了变量的描述性统计。表 2 列出了三套数据中,按照普通话能力分组的平均工资情况。在 CFPS 数据库中,除了 0 和 4 组之外,其余组的平均工资是随着普通话能力提高而增加的。普通话能力最高的 5 组,其小时平均工资也最高,达到了 29.64 元。而对于 CLDS 数据库,趋势则很明显,即普通话越流利,其工资水平越高,且普通话指标是由访员根据被访者的表现做出的客观评估,因此该指标可信程度更高。在 CGSS 数据库,普通话能力细分为两组,一个是听力,一个是表达,我们发现随着听力和表达能力的提高,工资水平也是逐渐提高的,这个趋势比较明显。单纯从数据描述性分析上,我们似乎能得到普通话能力是与工资水平呈正向关系的,但由于工资的决定因素很多,所以需要通过计量回归来控制影响工资的其他变量,然后再看普通话和工资这两者的关系是否还是正向的。

表 1 三套数据变量描述性统计

变量	CFPS		CLDS		CGSS	
	均值	标准差	均值	标准差	均值	标准差
<i>logwage</i>	2.0512	1.0910	2.2248	1.0196	1.9789	1.1771
<i>mandarin</i>	2.7293	1.4477	3.5039	1.2679	3.7528	1.0015
					3.1444	1.1871
<i>gender</i>	0.5923	0.4914	0.4696	0.4991	0.5032	0.5000
<i>age</i>	41.2855	15.6104	41.7389	13.6712	43.2274	16.3882
<i>Mar1</i>	0.7638	0.4248			0.7846	0.4111
<i>Mar2</i>	0.0512	0.2204			0.1136	0.3173
<i>edu</i>	8.7874	4.1144	9.5674	3.0754	8.7481	4.7076
<i>hukou1</i>	0.6244	0.4843	0.2806	0.4493	0.5995	0.4900
<i>hukou2</i>	0.9334	0.2494	0.8949	0.3067	0.9013	0.2983
<i>health</i>	2.9798	1.1399	2.4018	0.9690	3.7114	1.0828
<i>healthdif</i>	2.1853	0.5889				
<i>fatheredu</i>	4.9566	4.5929	5.3824	4.5154	4.5428	4.6593
<i>sociallevel</i>	2.6425	0.9870	4.0657	1.9238	2.2962	0.5475
<i>contract</i>			0.1399	0.3469	0.4202	0.4936
<i>union</i>			0.1084	0.3109	0.0988	0.2984
<i>confidence</i>	3.7544	1.0541				
<i>urban</i>	0.5114	0.4999	0.6374	0.4808	0.4870	0.4999

注:限于表格的大小,只列出了变量的均值和标准差。

从这三套数据的描述性分析中可知,在普通话能力的测度上,三套数据各有优缺点,比如 CFPS 指标取值最细、最全面,有 6 个取值,但只测度了普通话整体能力,且是主观指标;CLDS 取值也比较细,虽然是客观指标,但也只是测度了普通话整体能力;CGSS 指标虽然测度了普通话的听力和表达这两个维度,但指标的主观性偏强。所以我们选择这三套数据同时进行实证分析,能够更全面细致地刻画普通话对工资的影响,使得回归结果的可信度大大提高。

表 2 按照普通话能力分组的平均工资统计

普通话流利程度	0	1	2	3	4	5
CFPS(主观指标)						
工资均值	18.7603	10.7511	11.4025	23.8822	18.3819	29.6409
样本量	959	815	1 281	3 577	1 910	1 022
CLDS(客观指标)						
工资均值	—	7.1524	10.0637	11.9627	20.5339	22.3607
样本量	—	239	798	780	2 184	1 975
CGSS(主观指标)						
听力能力						
工资均值	—	5.9312	6.2142	9.5224	15.2037	19.5821
样本量	—	74	507	1 876	2 278	1 925
表达能力						
工资均值	—	4.8803	7.6911	13.9526	17.5568	21.8105
样本量	—	591	1 243	2 190	1 579	1 058

四、计量结果

(一) 样本总体回归结果

表 3 的 2、4、6 和 7 列是三套数据 OLS 的回归结果。我们发现,普通话的回归结果均为正显著,这意味着普通话的确会提高劳动者的收入,这与表 1 的统计分析相吻合。进一步,普通话的听力和表达能力对收入的影响也显著为正。

在研究语言对收入影响的多数文献中,大都考虑了语言的内生性问题。根据以往语言对收入的研究,内生性问题主要来源于以下两个方面:第一,遗漏变量问题,如观测不到的个人、家庭以及地区等层面的异质性,会同时影响到个体的语言能力和收入,从而导致回归结果有偏(Gao and Smyth, 2011)。第二,反向因果问题,如高收入者有能力进行更多的人力资本投入来提高其语言能力,此外高收入者可以通过更广泛的社交活动来提高其语言能力(Wang et al., 2016)。一个解决的办法就是为语言寻找合适的工具变量,工具变量法基本上可以解决上述可能存在的两个问题。而选择工具变量的原则是工具变量应该与语言直接相关,但却不与收入直接相关。在 CFPS 中,选择的工具变量有三个:与家人日常交流的语言是不是普通话^①、受访时使用的语言是不是普通话以及出生地与现居住地的距离。在 CLDS 中,选择的工具变量有两个:受访时使用的语言是不是普通话以及上学孩子的个数。该套工具变量回归记为 IV1。CGSS 中没有普通话的合适的工具变量,因此没有考虑 IV 回归。

工具变量的回归结果显示,普通话对小时工资的影响仍然为正显著,与 OLS 一致。工具变量的弱工具变量和过度识别检验均通过,说明工具变量是有效的。进一步发现,IV 回归结果普通话的系数均要大于 OLS,在 CFPS 中,前者甚至是后者的 4 倍左右。这意味着 OLS 中,普通话能力对工资的影响被低估了,这与 Saiz 和 Zoida(2005)等的发现一致。OLS 和 IV 这两种方法的回归中,普通话对收入的影响是相同的,不同的是影响程度,IV 的影响程度更

①之所以用“受访时使用的语言是不是普通话”作为工具变量,参考 Shields 和 Price (2002)以及 Dustmann 和 Fabbri(2003)的研究,他们认为是否在受访时使用英语可以反映出受访者的语言能力,但其收入等没有直接影响。并且“受访时使用的语言是不是普通话”是相对客观的指标,作为工具变量可以有效地减轻分析中可能存在的度量误差问题。

高一些,更接近实际情况。

表 3 三套数据的普通话对收入影响的 OLS 和 IV1 回归结果

变量	OLS	IV1	OLS	IV1	OLS 听力	OLS 表达
	CFPS	CFPS	CLDS	CLDS	CGSS	CGSS
<i>mandarin</i>	0.0414 *** (4.79)	0.1659 *** (3.80)	0.1038 *** (7.24)	0.1422 *** (3.56)	0.0343 ** (2.48)	0.0322 *** (2.61)
<i>gender</i>	0.3958 *** (17.54)	0.3939 *** (17.34)	0.3090 *** (12.60)	0.3112 *** (12.71)	0.3053 *** (12.79)	0.3067 *** (12.82)
<i>age</i>	0.0286 *** (4.23)	0.0349 *** (4.92)	0.0490 *** (6.25)	0.0497 *** (6.37)	0.0409 *** (4.92)	0.0413 *** (4.98)
<i>age</i> ²	-0.0004 *** (-4.95)	-0.0004 *** (-5.39)	-0.0006 *** (-6.36)	-0.0006 *** (-6.36)	-0.0005 *** (-5.49)	-0.0005 *** (-5.53)
<i>mar1</i>	0.0706 * (1.74)	0.0749 * (1.84)			0.1325 *** (3.34)	0.1302 *** (3.28)
<i>mar2</i>	0.0546 (0.81)	0.0556 (0.82)			0.1416 ** (2.30)	0.1403 ** (2.27)
<i>edu</i>	0.0448 *** (11.22)	0.0345 *** (6.36)	0.0387 *** (7.16)	0.0371 *** (6.67)	0.0531 *** (10.98)	0.0528 *** (10.83)
<i>hukou1</i>	-0.1707 *** (-5.91)	-0.1360 *** (-4.39)	-0.1578 *** (-4.23)	-0.1453 *** (-3.70)	-0.0116 (-0.39)	-0.0099 (-0.33)
<i>hukou2</i>	-0.1983 *** (-5.29)	-0.1701 *** (-4.36)	-0.1728 *** (-4.81)	-0.1601 *** (-4.10)	-0.1173 *** (-3.48)	-0.1178 *** (-3.49)
<i>health</i>	-0.0509 *** (-4.79)	-0.0424 *** (-3.76)	-0.0313 ** (-2.11)	-0.0311 ** (-2.10)	-0.0267 * (-1.89)	-0.0274 * (-1.92)
<i>healthdif</i>	0.0203 (1.03)	0.0202 (1.01)				
<i>fatheredu</i>	0.0115 *** (4.23)	0.0093 *** (3.27)	0.0076 ** (2.55)	0.0073 ** (2.43)	0.0056 * (1.79)	0.0056 * (1.78)
<i>sociallevel</i>	0.0581 *** (5.22)	0.0472 *** (3.95)	0.0466 *** (7.29)	0.0469 *** (7.38)	-0.2657 *** (-11.56)	-0.2657 *** (-11.57)
<i>contract</i>			0.0319 (1.24)	0.0303 (1.18)	0.0682 *** (2.79)	0.0661 *** (2.70)
<i>union</i>			0.1168 *** (4.20)	0.1134 *** (4.06)	0.0844 *** (3.04)	0.0836 *** (3.01)
<i>confidence</i>	0.0179 * (1.67)	0.0079 (0.70)				
<i>urban</i>	0.0533 ** (2.30)	0.0585 ** (2.49)	0.1299 *** (4.04)	0.1242 *** (3.83)	0.1919 *** (6.64)	0.1891 *** (6.54)
工作性质	控制	控制	控制	控制	控制	控制
工作类型	控制	控制	控制	控制	控制	控制
地区	控制	控制	控制	控制	控制	控制
<i>N</i>	9 316	9 313	5 154	5 149	4 075	4 076
adj. <i>R</i> ²	0.2336	0.2168	0.3205	0.3199	0.4606	0.4604
<i>Cragg-Donald</i>		161.9122		283.1525		
<i>Wald F</i>						
<i>P</i>		0.0000		0.0000		
<i>Hansen J</i>		2.1132		0.1311		
<i>P</i>		0.3476		0.7173		

注:(1) * 表示在 10%水平上显著, ** 表示在 5%水平上显著, *** 表示在 1%水平上显著。括号里为系数的 *t* 统计值。下同。(2) 有些控制变量在某套数据库中有,在其他数据库中可能没有。比如 *contract* (与用工企业是否签订合同,签订=1,没有签订=0)和 *union* (所在单位是否有工会,有工会=1,无工会=0)这两个变量 CFPS 中就没有。而 CLDS 中没有婚姻状况,CLDS 和 CGSS 中没有 *confidence* 和 *healthdif* 这两个

变量。

但由于 CFPS 和 CLDS 的调查员一般是采用普通话进行访问,这个变量在一定程度上等同于“受访者的普通话水平是否熟练”,它与受访者的普通话水平变量含义非常接近。也就是说,“受访时使用的语言是不是普通话”作为工具变量可能并不合适^①。因此我们在 CFPS 中删除了“受访时使用的语言是不是普通话”工具变量,仅选取与家人日常交流的语言是不是普通话以及出生地与现居住地的距离;在 CLDS 中,删除“受访时使用的语言是不是普通话”工具变量后,存在弱识别,并且无法找到其他合适的工具变量。因此我们采用了 Lewbel (2012)所提出的工具变量回归方法进行了回归。其方法是不依靠外生工具变量而依靠创建内生工具变量来进行识别。该套工具变量回归记为 IV2。回归结果见表 4。可见,在去掉不合适的工具变量后,回归结果也比较稳健。

表 4 针对 CFPS 和 CLDS 两套数据的 IV2 回归结果

变量	CFPS	CLDS
<i>mandarin</i>	0.1058 * (1.81)	0.1173 *** (2.67)
控制变量	是	是
<i>N</i>	9316	5154
adj. <i>R</i> ²	0.2292	0.2970
<i>Cragg-Donald Wald F</i>	128.1046	35.2280
<i>P</i>	0.0000	0.0000
<i>Hansen J</i>	0.1549	29.3201
<i>P</i>	0.6939	0.2083

除了工具变量,还可以采取倾向得分匹配法 (Propensity Score Matching, PSM) 来克服内生性。我们采用此法来进一步验证回归结果的稳健性。与工具变量方法稍有不同,倾向匹配法主要是用来解决反向因果带来的内生性问题。具体说,我们研究普通话能力对收入的影响,如果仅简单地将会说普通话作为虚拟变量,进而对总体回归,参数估计就会产生偏误。因为在这样的情况下,我们只观察到了某一个样本因为发生了某一事件后(本文是会说普通话)产生的表现,并且拿这种表现去和另一些没有发生这件事情(不会说普通话)的其他样本去做比较。事实上,我们应该将某个样本会说普通话之前的工资与他会说普通话之后的样本作比较,来估计这个收入差是正还是负。于是,我们引入“倾向得分匹配”这一研究方法。这种方法能让我们从不会说普通话的样本(控制组)中,对每个人学会说普通话的概率进行估计,然后选出和某个会说普通话的样本(处理组)具有非常相似的学会普通话的概率,但是事实却不会说普通话的样本(控制组)——作为某个会说普通话样本的对照,然后再来看它们的区别。当样本中的每个会说普通话的样本(处理组)都找到了与之匹配的不会说普通话的样本(控制组),我们便能对这两组样本进行比较研究了。

在做 PSM 之前,我们需要将 CFPS、CLDS 和 CGSS 中的 *Mandarin* 变量合并为只有 0 和 1 这两个数值的虚拟变量,会说普通话=1,不会说=0 或者听懂普通话=1,听不懂=0。表 5 列出了 PSM 的回归结果。针对三套数据采用三种匹配方法下的回归结果仍显著为正,这意味

①感谢审稿人的指正。

着普通话能力会提高工资收入,这个结果应该还是比较稳健的。进一步,在 CGSS 中,听力能力和表达能力这两个单项能力对收入的影响也比较显著。

对于控制变量。我们发现男性工资要普遍高于女性。这与多数研究微观个体收入的发现一致。事实上,女性无论是在就业上还是工资待遇上可能都会遇到歧视,这与户籍歧视均是影响劳动者收入差距的重要因素。随着劳动者年龄的增长,工资收入呈现出先上升后下降的倒 U 型趋势,这符合生命周期理论,年轻劳动者无论在人力资本水平还是在劳动生产率上,都具有一定的优势,因此其工资水平也较高,而年龄过小者可能工作经验不足,年龄过大者可能体力不足,年龄对于他们反而会成为制约工资增长的因素。对于学历较低的外来务工人员来说,回归结果呈现出典型的以体力换收入的特点。相对于单身,结婚的劳动者收入较高,这是因为他们的家庭观念比较重,养家糊口的责任意识较强,努力工作的动机也越强烈。教育的回归结果为正,且显著性非常高,表明教育层次越高的劳动者,其工资水平也越高。这也符合人力资本理论预期。教育可以提升人力资本,也是提高潜在人力资本积累的重要因素。农民的收入相对非农民要低一些,同时本地人的收入要低于外来人口。健康以及健康状态的变化对工资的影响也比较显著,因为健康也是一种人力资本,身体越健康,越能全身心投入到工作中,工作回报也会越高。父亲受教育年限也会显著影响子辈收入,如果财富多和受过良好教育的父母可以为子女提供更多更好的学习机会,必然也会提高子女的教育质量和个人能力。事实上,如果父母的教育背景越好,则表明父母本身的能力也越高,这样子代也会遗传到这些高的能力和素质。社会地位越高的劳动者,其工资收入也越高。同样,对未来信心高,则可能是因为对目前的工作状态比较满意,工资比较高所致。城镇居民的收入普遍要比农村、郊区的高一些。除此之外,CLDS 和 CGSS 中一些控制变量在 CFPS 中没有,比如合同、工会等,它们对收入的影响也都比较显著。

表 5 三套数据 PSM 的 Mandarin(普通话能力)变量回归结果

变量	CFPS		CLDS	
<i>Nearest neighbor matching</i>	0.1030 ^{***}	(3.64)	0.1476 [*]	(1.77)
<i>Kernel matching</i>	0.1113 ^{***}	(3.88)	0.1536 ^{**}	(2.07)
<i>Radius matching</i>	0.1156 ^{***}	(3.96)	0.1667 ^{**}	(2.39)
控制变量	是	是	是	是
<i>N</i>	9 266		5 013	

变量	CGSS			
	听力		表达	
<i>Nearest neighbor matching</i>	0.1314 ^{***}	(2.85)	0.0878 [*]	(1.89)
<i>Kernel matching</i>	0.1178 ^{***}	(2.83)	0.0933 ^{**}	(2.09)
<i>Radius matching</i>	0.1168 ^{***}	(2.93)	0.0992 ^{**}	(2.18)
控制变量	是	是	是	是
<i>N</i>	4 053		3 980	

注:同表 3。表中列出的只是普通话能力 Mandarin 的回归结果。控制变量限于篇幅略去,同时下面所有表中的控制变量也都略去,如有需要可向作者索取。

(二) 样本分组回归结果

样本的总体结果表明,普通话(包括单项能力如听力和表达)对劳动者收入会产生一定的影响,且这种影响比较稳健。Chen 等(2014)发现服务业中上海话会显著影响外来农民工

收入,而对其他行业的影响不显著,主要是因为服务行业与本地人接触的机会多,而上海话便于他们之间的交流。表6列出了分行业的回归结果。

我们发现,按照服务业和非服务业划分,对于CFPS,采用OLS、IV(第二种工具变量选取方法)和PSM回归后,两组样本的结果好像没有太大差异,普通话能力对这两类行业的收入都有显著的影响。但是在CLDS中,我们看出了明显差异,三种回归方法都表明,普通话会显著提高服务行业的工资,但对非服务行业的影响却不显著。在CGSS中,结果也类似,普通话的听力能力和表达能力对服务业工资的影响也显著,而非服务业则不显著。

三套数据中有两套的回归结果表明相对其他行业,服务业中普通话对收入有着重要作用。主要原因是服务行业的性质决定的,服务业接触的各方面人群很多,如果善于使用普通话,与各地区、各层次的客户沟通更为方便,工作效率更高,因此收入也会高。根据《通用语言文字法》中有关持证上岗的相关规定,公共服务行业尤其是一些特定岗位人员比如广播员、解说员和话务员等,其普通话水平不低于二级,由此可见服务行业中普通话的重要程度。

表6 三套数据PSM的Mandarin(普通话能力)变量分组回归结果:按行业

数据库	方法	行业	系数	T(Z)值	控制变量	adj. R ²	Cragg-Donald Wald F P 值	Hansen J P 值	N
CFPS	OLS	服务业	0.061	3.27***	是	0.24			2 598
		非服务业	0.034	3.57***		0.24			6 718
	IV2	服务业	0.185	1.77*			0.00	0.36	2 597
		非服务业	0.143	2.68**			0.00	0.65	6 716
	PSM	服务业	0.215	4.78***					2 596
		非服务业	0.131	3.84***					6 718
CLDS	OLS	服务业	0.110	3.89***		0.31			1 198
		非服务业	0.047	1.30		0.34			3 956
	IV2	服务业	0.144	3.25***			0.00	0.39	1 197
		非服务业	0.466	0.53			0.00	0.51	3 952
	PSM	服务业	0.276	2.58***					1 050
		非服务业	0.333	1.50					3 956
CGSS	OLS 听	服务业	0.493	2.21**		0.48			1 641
		非服务业	0.130	1.52		0.47			2 434
	PSM 听	服务业	0.252	4.50***					1 622
		非服务业	0.253	1.55					2 414
	OLS 说	服务业	0.038	1.90*		0.48			1 642
		非服务业	0.025	1.62		0.47			2 434
	PSM 说	服务业	0.936	1.69*					1 642
		非服务业	0.857	1.10					2 434

注:限于篇幅,所有控制变量的回归结果均未列出,如有需要可向作者索取。PSM回归均采用的是核匹配方式。

语言对收入的影响除了上述提到的行业差异之外,是否还会有地域差异呢?表7列出了按照城乡分组的回归结果。与表4的结果类似,CFPS中普通话对城镇和乡村的收入都有显著影响,但在OLS和PSM中,普通话对城镇居民收入影响的显著性要更高一些。在CLDS和CGSS中,回归结果差异则较为明显,在城镇中,普通话对收入的影响比较显著,而在乡村则基本不显著。一般来说,乡村里的外来人口较少,主要是本地农民聚居地,他们从事的也主要是农业活动,彼此之间可能不用普通话而是使用方言就能够顺利地交流。因此普通话的作用和普及性要差。在城镇中,来自各个地区的外来人口相对较多,人员结构比较复杂,

这时候普通话更易于人们的交流,尤其是有着不同方言背景的人。另外,城镇中的许多工作比如服务业,对普通话的要求也更高,因此在城镇,普通话的重要性就要高于乡村,进而普通话对城镇居民收入的影响就会比较明显。

表 7 三套数据 PSM 的 *Mandarin* (普通话能力) 变量分组回归结果:按城乡

数据库	方法	地区	系数	$T(Z)$ 值	控制 变量	adj. R^2	Cragg-Donald Wald F P 值	Hansen J P 值	N
CFPS	OLS	城镇	0.052	4.32 ^{***}	是	0.26			5 185
		乡村	0.028	2.30 ^{**}		0.19			4 131
	IV2	城镇	0.088	1.70 [*]			0.00	0.51	5 184
		乡村	0.256	3.44 ^{***}			0.00	0.55	4 129
	PSM	城镇	0.107	3.22 ^{***}					5 185
		乡村	0.098	2.22 ^{**}					4 131
CLDS	OLS	城镇	0.103	5.55 ^{***}		0.32			2 735
		乡村	0.045	1.31		0.22			2 419
	IV2	城镇	0.158	3.50 ^{***}			0.00	0.17	2 734
		乡村	0.149	1.65 [*]			0.00	0.25	2 415
	PSM	城镇	0.198	2.37 ^{**}					2 628
		乡村	0.131	1.22					2 399
CGSS	OLS 听	城镇	0.036	2.30 ^{**}		0.48			2 651
		乡村	0.024	0.92		0.35			1 424
	PSM 听	城镇	0.118	2.27 ^{**}					2 623
		乡村	0.055	0.75					1 414
	OLS 说	城镇	0.025	1.73 [*]		0.48			2 651
		乡村	0.030	1.20		0.36			1 425
	PSM 说	城镇	0.084	2.05 ^{**}					2 651
		乡村	0.091	0.39					1 425

注:同表 3。

再次,考虑到中国的实际情况,中国方言分布的一个明显特征是:南方地区的方言差异极大,相隔数公里可能就会存在交流障碍;而北方地区以北方方言为主,虽有口音差别,但几乎不影响交流。因而,方言类研究可能对北方居民的影响不大,而对南方居民影响较大。因此,我们按照东北、北部沿海、东部沿海、南部沿海、黄河中游、长江中游、西南地区 and 西北地区将中国分为八大区域^①。由于子样本过多,有一些子样本的 IV 检验没有通过,因此我们只列出了 OLS 的回归结果,根据 Saiz 和 Zoida (2005),OLS 和工具变量回归结果基本一致,只是数值大小有所不同。因此,OLS 在此的回归结果还是有一定参考价值的。表 8 表明,相对于其他区域,南部沿海地区和西南地区的普通话能力对收入的影响要更明显一些,东北地区的普通话能力回归结果都不显著,北部沿海和东部沿海、黄河中游和长江中游地区、西北地区只有 CLDS 显著。而 CLDS 回归基本上在所有样本中差异不大,但在南部沿海地区显著性要更大。总的来说,普通话能力在南方的作用相对北方要更大一些。因为南方地区的方言差异较大,相隔数公里可能就会存在交流障碍;而北方地区以北方方言为主,虽有口音差别,但几乎不影响交流。

①东北综合经济区:辽宁、吉林、黑龙江;北部沿海综合经济区:北京、天津、河北、山东;东部沿海综合经济区:上海、江苏、浙江;南部沿海经济区:福建、广东、海南;黄河中游综合经济区:陕西、山西、河南、内蒙古;长江中游综合经济区:湖北、湖南、江西、安徽;大西南综合经济区:云南、贵州、四川、重庆、广西;大西北综合经济区:甘肃、青海、宁夏、西藏、新疆。

表 8 分区域回归结果

数据库	变量	东北	北部沿海	东部沿海	南部沿海	黄河中游	长江中游	西南地区	西北地区
CFPS	<i>mandarin</i>	0.0224 (0.75)	0.0358 (1.36)	0.0312 (0.99)	0.0828*** (3.66)	0.0108 (0.61)	0.0606** (2.12)	0.0646*** (2.62)	0.0402 (1.48)
	<i>N</i>	1 337	1 106	1 480	1 004	1 729	921	917	822
	adj. <i>R</i> ²	0.1866	0.2376	0.2938	0.1679	0.1283	0.2246	0.1846	0.1885
CLDS	<i>mandarin</i>	-0.1212 (-1.16)	0.1192*** (3.19)	0.0976** (2.25)	0.1069*** (4.24)	0.1026* (1.89)	0.1053** (2.57)	-0.0104 (-0.19)	0.1709*** (3.89)
	<i>N</i>	333	607	623	1 480	423	656	479	553
	adj. <i>R</i> ²	0.2527	0.3129	0.4211	0.3503	0.1803	0.2743	0.3356	0.3899
CGSS	听	0.0261 (0.61)	0.0287 (1.21)	-0.0051 (-0.15)	0.1683*** (2.90)	-0.0739 (-1.64)	0.0323 (0.74)	0.1000** (2.49)	-0.0571 (-0.68)
	<i>N</i>	485	957	724	274	335	549	609	142
	adj. <i>R</i> ²	0.3210	0.4891	0.5108	0.4538	0.3827	0.3049	0.3362	0.4811
	说	0.0028 (0.07)	0.0151 (0.80)	-0.0306 (-0.96)	0.1135* (1.87)	0.0393 (0.90)	0.0042 (0.10)	0.0943*** (2.62)	0.0373 (0.57)
	<i>N</i>	485	957	724	274	336	549	609	142
	adj. <i>R</i> ²	0.3205	0.4885	0.5115	0.4410	0.3766	0.3041	0.3352	0.4804

最后,如果存在同一方言区内的劳动力(包括流动和非流动人口),那么回归结果可能被低估,因为包含了大量不会受到语言影响的劳动力,语言的实际影响也可能会被削弱。因此我们将样本分为跨方言区和方言内区域^①两个子样本,并且以 CLDS 数据库为例,分别进行回归。^②

表 9 表明,普通话对跨方言区迁移人口收入的影响为 0.1146,要大于方言区内的影响(0.1033),表明总体样本确实存在低估(总体样本回归结果为 0.1038),但是低估的程度不大。其原因可能是受社会网络的影响,来自同一地区的迁移人口往往在迁入地城市聚集居住,形成了如浙江村、温州村等。这些人口主要用方言进行交流并从事相同或类似的职业,从而削弱了普通话能力对其职业和收入的影响。

表 9 跨方言区和方言区内的回归结果

变量	跨方言区	方言区内
<i>mandarin</i>	0.1146** (2.17)	0.1033*** (6.79)
<i>N</i>	756	4 398
adj. <i>R</i> ²	0.3616	0.3109

①我们主要以省为单位,考虑中国主要的七大方言区(见前文)。具体说,如果样本的出生地和现居住地不在同一个方言区,就将该样本归为跨方言区。但由于有些省份的不同区域属于不同方言区,而在数据库中又无法看到样本具体属于某省份的哪个县(因为数据库是以省为单位),因此我们认为这类省份的主要流行方言就是该省的方言区所在地。比如江西南部说客家话,但江西大部分地区主要说赣方言,因此我们认为来自江西的所有样本都说赣方言。

②由于 IV 回归跨方言区样本没有通过弱识别检验,故而采用了 OLS 回归。本文研究结果表明,OLS 回归结果与 IV 和 PSM 结果具有一致性。

总的来看,如果样本是跨方言区流动,且双方在交流中都使用本地方言,那么信息传递的流畅度和有效性就会大大降低。反之,如果说普通话则更利于来自不同方言区域的人们交流,也就是说,对于这类群体,普通话的作用更大一些。联系到本文,操持普通话使得这类人群在找工作中的优势也更加明显。

五、结语

语言作为一种人力资本形式,它对收入很可能产生重要影响。国内外很多学者对此进行了研究,比如他们中的很多研究了外语、少数民族语言、方言以及普通话对收入的影响,而研究结论也不一而足。本文采用了研究中常见的 CFPS、CLDS 和 CGSS 三套中国的综合数据,检验了普通话对劳动者收入的影响。

结论发现,总体上,普通话以及单项能力包括听力和表达都会对劳动者的收入产生比较显著的影响,在考虑了内生性之后,结论未变,且对于 CFPS、CLDS 和 CGSS 这三套数据都成立。进一步,与 Chen 等(2014)的研究结果类似,我们也发现了普通话对服务业从业人员的工资有显著提升作用,但是对其他行业的影响则不显著。这主要是由服务业的性质决定的,即需要大量地与各类人群交流。另外,普通话对城镇劳动者收入的影响比较显著,对农村劳动者收入的影响不显著。这主要是城镇相对农村,许多劳动者是来自五湖四海,语言背景不一而同,人群结构比较复杂,普通话更便于交流。同时城镇的许多工作也需要普通话,因此相对农村,普通话的作用更重要一些。以上两项结论主要是针对 CLDS 和 CGSS 数据成立。另外,普通话能力在南方的作用相对北方要更大一些。因为南方地区的方言差异较大,相隔数公里可能就会存在交流障碍;而北方地区以北方方言为主,虽有口音差别,但几乎不影响交流。最后,普通话对跨方言区流动人口的影响要大于对方言区内人口的影响。

因此,在中国方言种类如此之多的背景下,普及普通话的意义重大。本文发现,普通话能力对个人收入有重要影响。因此在宣传普及普通话的时候,适当地强调普通话对个人经济利益的影响,政策效果会更好一些。

参考文献:

- 1.刘泉,2014:《外语能力与收入——来自中国城市劳动力市场的证据》,《南开经济研究》第3期。
- 2.Azam, M., A. Chin, and N. Prakash. 2013. "The Returns to English Language Skills in India." *Economic Development and Cultural Change* 61(2):335-367.
- 3.Canevale, A., F. Richard, and L. Lowell. 2001. "Understanding, Speaking, Reading, Writing, and Earnings in the Immigrant Labor Market." *American Economic Review* 91(2):159-163.
- 4.Carliner, G. 1981. "Wage Differentials by Language Group and the Market for Language Skills in Canada." *Journal of Human Resources* 16(2):384-399.
- 5.Casale, D., and D. Posel. 2011. "English Language Proficiency and Earnings in a Developing Country: The Case of South Africa." *Journal of Socio-Economics* 40(10):385-393.
- 6.Chen, Z., M. Lu, and L. Xu. 2014. "Returns to Dialect: Identity Exposure through Language in the Chinese Labor Market." *China Economic Review* 30(9):27-43.
- 7.Chiswick, B. R., and P. W. Miller. 2003. "The Complementarity of Language and Other Human Capital: Immigrant Earnings in Canada." *Economics of Education Review* 22(3):469-480.
- 8.Chiswick, B. R., and P. W. Miller. 2007. *The Economics of Language: International Analyses*. London: Routledge.

9. Chiswick, B. R., and P. W. Miller. 2014. "International Migration and the Economics of Language." IZA Discussion Papers 7880, Institute for the Study of Labor (IZA). International Migration and the Economics of Language. http://samples.sainsburysebooks.co.uk/9780444537652_sample_757584.pdf.
10. Di Paola, A., and A. Tansel. 2015. "Returns to Foreign Language Skills in a Developing Country: The Case of Turkey." *Journal of Development Studies* 51(4):407-421.
11. Donado, A. 2014. "Foreign Languages and Their Impact on Income and Unemployment." Beitrage zur Jahrestagung des Vereins fur Socialpolitik 2014: Evidenzbasierte Wirtschaftspolitik - Session: Norms and Culture, No. F10-V3. https://www.econstor.eu/bitstream/10419/100288/1/VfS_2014_pid_123.pdf.
12. Dustmann, C., and F. Fabbri. 2003. "Language Proficiency and Labour Market Performance of Immigrants in the UK." *Economic Journal* 113(489):695-717.
13. Fry, R., and L. Lowell. 2003. "The Value of Bilingualism in the U.S. Labor Market." *Industrial and Labor Relations Review* 57(1):128-140.
14. Gao, W., and R. Smyth. 2011. "Economic Returns to Speaking 'Standard Mandarin' among Migrants in China's Urban Labour Market." *Economics of Education Review* 30(2):342-352.
15. Grin, F. 2001. "English as Economic Value: Facts and Fallacies." *World Englishes* 20(1):65-78.
16. Gong, Y., I. H.-S. Chow, and D. Ahlstrom. 2011. "Cultural Diversity in China: Dialect, Job Embeddedness, and Turnover." *Asia Pacific Journal of Management* 28(9):221-238.
17. Grenier, G., and F. Vaillancourt. 1983. "An Economic Perspective on Learning a Second Language." *Journal of Multilingual and Multicultural Development* 4(6):471-483.
18. Guo, Q., and W. Sun. 2014. "Economic Returns to English Proficiency for College Graduates in Mainland China." *China Economic Review* 30(7):290-300.
19. Lazear, E. 1999. "Culture and Language." *Journal of Political Economy* 107(6):95-126.
20. Leung, C., and J. Ruan. 2012. *Perspectives on Teaching and Learning Chinese Literacy in China*, Vol. 2. Berlin: Springer Netherlands.
21. Lewbel, A. 2012. "Using Heteroscedasticity to Identify and Estimate Mismeasured and Endogenous Regressor Models." *Journal of Business and Economic Statistics* 30(1):67-80.
22. Levinsohn, J. 2007. "Globalization and the Returns to Speaking English in South Africa." In *Globalization and Poverty*. Edited by J. Levinsohn, 629-646. Chicago: University of Chicago.
23. Isphording, I. E. 2013. "Returns to Foreign Language Skills of Immigrants in Spain." *Labour* 27(3):443-461.
24. Pendakur, K., and R. Pendakur. 1998. "Speak and Ye Shall Receive: Language Knowledge as Human Capital." In *Economic Approaches to Language and Bilingualism*. Edited by K. Pendakur and R. Pendakur, 89-121. Ottawa: Canadian Heritage.
25. Saiz, A., and E. Zoida. 2005. "Listening to What the World Says: Bilingualism and Earnings in the United States." *Review of Economics and Statistics* 87(3):523-538.
26. Shields, M. A., and S. Wheatley Price. 2002. "The English Language Fluency and Occupational Success of Ethnic Minority Immigrant Men Living in English Metropolitan Areas." *Journal of Population Economics* 15(1):137-160.
27. Stöhr, T. 2015. "The Returns to Occupational Foreign Language Use: Evidence from Germany." *Labour Economics* 32(1):86-98.
28. Pendakur, K., and R. Pendakur. 2002. "Language as both Human Capital and Ethnicity." *International Migration Review* 36(1):147-177.
29. Toomet, O. 2011. "Learn English, Not the Local Language: Ethnic Russians in the Baltic States." *American Economic Review* 101(3):526-531.
30. Wang, H., Z. Cheng, and R. Smyth. 2016. "Language and Consumption." *China Economic Review* 40(3):135-151.
31. Williams, D. 2011. "Multiple Language Usage and Earnings in Western Europe." *International Journal of*

Manpower 32(4):372–393.

32.Zhou, M. 2003. *Multilingualism in China: The Politics of Writing Reforms for Minority Languages* 1949–2002. Berlin: Mouton de Gruyter.

A Research of the Impact of Mandarin on Chinese Workers

Chen Yuanyuan

(The Center For Economic Research, Shangdong University)

Abstract: As a form of human capital, language, is likely to have an important effect on income. In this paper, we use three sets of Chinese comprehensive data; the CFPS, CLDS and CGSS and test the impact of the ability of mandarin on the income of workers. We find that on the whole, mandarin and individual ability (listening and speaking) will increase the workers' income significantly which is robust while considering endogenous variable. Further, it is found that mandarin has a significant effect on the wages of service workers, but not for other industries. In addition, the impact of mandarin on the income of urban workers is also significant, while the impact on rural labor income is not significant. In addition, the effect of Mandarin in the south is larger than that in the north, and the influence of Mandarin on the floating population from different dialect area is greater than that of the population in the same dialect area.

Keywords: Mandarin, Income, Instrumental Variables, PSM

JEL Classification: J31, I26

(责任编辑:赵锐、彭爽)

(上接第 97 页)

Firm Productivity and Firm Selection in the Presence of Agglomeration Economies: The Perspective of Heterogeneous Firm Productivity

Zhang Xin¹, Zhu Yingming¹ and Li Yujian²

(1; School of Economics and Management, Nanjing University of Science and Technology; 2; Rizhao Politechnic)

Abstract: Based on the analytical framework of heterogeneous firm trade theory, this paper estimates the total factor productivity of firms using a semi-parametric model integrated with agglomeration economies. Specifically, we study the influence of agglomeration externalities on firm selection. We have the following findings: the contribution of agglomeration economies on firm productivity varies by industry; The distribution of productivity exhibits a fat tail, in which firm selection exists; Agglomeration economies are not significant while firm selection is remarkable; Finally, the home market effect is the driving force of agglomeration, localization and urbanization, and it allows plants to spread.

Keywords: Agglomeration Economies, Heterogeneous Firm Productivity, Firm Selection, Sorting Effect

JEL Classification: C34, D24, R39

(责任编辑:彭爽、陈永清)