



**北京师范大学经济与工商管理学院
工作论文（working paper）系列
经济类 No. 50**

**邢春冰、贾淑艳、李实：教育回报率的地区
差异及其对劳动力流动的影响**

2013 年 9 月

教育回报率的地区差异及其对劳动力流动的影响*

邢春冰 贾淑艳 李实

内容提要：使用 2005 年 1%人口抽样调查数据以及 2011 年流动人口动态监测调查数据，我们发现农民工教育回报率存在明显的地区差异：东部地区和发达城市的教育回报率明显高于中西部地区。比较两个年份的结果可以看出，农民工的总体教育回报率有所降低，地区间回报率的差异也有所降低，并呈现出向均值收敛的特征。控制福利状况和迁移时间以及修正样本选择偏差对文章的结果影响不大。教育回报率的地区差异对不同教育水平劳动力的相对迁移概率产生了显著的影响。进一步减少流动障碍，促进农民工合理流动将有助于提升我国的经济增长潜力。

关键词：农民工 教育回报率 地区差异

一、引言

人力资本理论认为，教育作为人力资本投资的主要形式，可以提高劳动力的生产率，进而提高其收入。而教育回报率——多接受一年教育所带来的收入增加——则不仅影响家庭和个人的教育投资决策，也影响到教育政策的制定。此外，在一个多个部门（如地区）的经济中，不同部门之间的教育回报率差异要么反映了部门分割状况，要么反映了不同部门间的特征差异，对了解劳动力市场的状况和制定相关政策具有重要意义。研究表明，我国不同部门、不同群体的教育回报率存在显著差异。^①本文以城镇劳动力市场中的农民工为研究对象，考察他们的教育回报率，并着重分析这一群体教育回报率的地区差异及其变化趋势。

农民工是我国劳动力市场中一个庞大而特殊的群体。我国巨大的地区和城乡收入差距使大量的劳动力从欠发达的农村地区流向发达的城市地区。尽管户籍制度不再将农民限制在土地上，但这部分劳动力通常无法获得城镇地区的本地户口，形成了庞大的“农民工”群体。2011 年，我国外出农民工的数量为 1.58 亿。^②由于教育水平低，加之歧视和信息不对称等原因，农民工的工资水平显著低于城镇本地职工。这使得研究这一群体的工资决定和教育回报尤其重要。

农民工群体是推进我国城镇化的一支重要力量。他们在决定是否迁移的同时，还需要决定迁移的目的地。在一系列影响劳动力区位选择的因素中，（预期的）工资水平是一个非常重要的因素。如果各地区的教育回报率不同，就意味着不同教育水平劳动力的相对工资水平存在差异，而这个差异会影响到劳动力迁移的模式。给定相同的平均收入，回报率高的地区会吸引更多高教育水平的劳动力；而低教育水平的劳动力则更倾向于迁往低回报率的地区。可见，教育回报率的地区差异将对我国的城镇化进程产生重要的影响。而通过考察教育回报率地区差异的变化，我们可以推断劳动力市场对地区差异反应和调整的过程。

本文使用 2005 年 1%人口抽样调查数据以及 2011 年流动人口动态监测调查数据来估计农民工的教育回报率及其地区差异。其中，2005 年的人口抽样调查是一份具有全国代表性的数据，利用它不仅可以描绘农民工群体教育回报率的地区差异，而且可以利用农民工的来源地等信息进行样本选择修正。利用 OLS 回归，我们发现农民工教育回报率存在明显的地区差异：东部地区和发达城市的教育回报率明显高于中西部地区。利用 2011 年的流动人口监测数据，我们发现了同样的规律。比较两个年份的结果可以看出，农民工的总体教育回报率有所降低，地区间回报率的差异也有所降低，并呈现出向均值收敛的特征。

* 邢春冰，北京师范大学经济管理学院副教授，邮箱：xingchb@bnu.edu.cn；贾淑艳，北京师范大学经济管理学院西方经济学硕士研究生，邮箱：201121030009@mail.bnu.edu.cn；李实，北京师范大学经济管理学院教授，邮箱：lishi@bnu.edu.cn。感谢匿名审稿人的宝贵意见。文责自负。

^① 如教育回报率的性别差异研究（Beaudry and Lewis, 2012），所有制差异（邢春冰，2005），城乡差异（姚先国、张海峰，2004），行业差异（高梦滔、张颖，2007）等。

^② <http://www.askci.com/news/201204/27/2715551256411.shtml>。

我们还考察了地区间工资以外福利水平的差异和迁移时间差异对结果可能造成的影响。结果表明，控制上述两方面因素之后各地区的教育回报率及其差异变化不大。而为了克服选择性偏差，我们采用 Dahl（2002）提出的半参数多部门样本选择模型。该模型与 Heckman 的两步法类似，但是通过采用非参数方法估计迁移概率和控制样本选择纠正项，它可以解决多部门的样本选择问题，而且克服了参数模型结果不稳健的弱点。利用 2005 年的数据进行样本选择修正的结果表明，样本选择偏差并不严重。这些练习说明了本文结果的稳健性。

最后，我们考察了地区间的人口流动与教育回报率的地区差异之间的关系。结果表明，相对于低教育水平的农民工而言，高教育水平的农民工更倾向于流动到教育回报率比较高的地区。这表明教育回报率的地区差异对我国劳动力流动的模式产生了显著影响。这也解释了为什么教育回报率的地区差异会逐渐缩小。

本文的结构安排如下。第二节简要回顾了与农民工教育回报率相关的文献，第三节介绍数据，第四节报告了 OLS 回归得到的教育回报率估计结果，第五节是稳健性检验，并报告了进行样本选择修正之后 2005 年各地区的教育回报率，最后一节总结全文。

二、文献综述

近年来，农民工群体的收入和福利状况备受关注，对其教育回报率的研究也逐渐增多。一些研究以（来自）某个地区的农民工为样本考察其教育回报率。例如，罗忠勇（2010）利用 2008 年珠三角地区农民工数据估计的教育回报率为 4.4%；张世伟等（2009）利用 2005 年吉林省进城务工人员抽样调查数据和 2006 年吉林省农村抽样调查数据估计的农民工教育回报率为 4.55%。还有一些研究使用了更大范围的样本。比如，张泓骏、施晓霞（2006）利用 1999 年农业部在河北、陕西、安徽、湖南、四川、浙江六个省农村地区的住户调查数据，估计得到外出打工人口的教育回报率为 5.36%；戎建（2008）利用 2004 年复旦大学中国社会主义市场经济研究中心的来自 22 个省份的农村调查数据，得到迁移者的教育回报率为 4.77%。对农民工教育回报率的估计还出现在一些其他研究中。比如，人们在考察城镇职工与农民工的收入差距时，Oaxaca 分解的第一步是估计两类人群的工资方程，教育回报率是不可或缺的内容。多数的研究结果表明，农民工的教育回报率低于城镇职工（相关的研究见 Meng and Zhang, 2001；邢春冰，2008；Maurer-Fazio and Dinh, 2004；邓曲恒，2007）。

总结上述研究，农民工在我国城镇劳动力市场中的教育回报率相对较低，不同时间和不同地区的教育回报率存在显著差异。但是，由于数据缺乏可比性，要从已有的研究中得到教育回报率的地区分布和跨期变化规律非常困难。一些研究在使用多个地区的数据时，又往往只估计一个总体教育回报率——这实际上假设不同地区的教育回报率相同。

如果我们关注一个随机抽取的农民工在某个城镇劳动力市场上所面对的教育回报率，OLS 回归往往给出的是有偏的结果，这是一个典型的样本选择问题。但由于农民工面临多个地区选项，问题处理并不像两个选项的样本选择问题那么简单（Heckman, 1979）。^①在较早的研究中，Lee（1983）假设当劳动力面临多个选择时，真正起作用的选项是他的最优也是其实际做出的选择，这使得样本选择纠正变得简单。但是 Lee（1983）属于参数模型，对扰动项的实际分布比较敏感。Ahn and Powell（1993）提出一个半参数方法，通过控制个体选择概率的未知函数来纠正样本选择偏差。Dahl（2002）借鉴上述研究提出一套简单易行的修正方法，研究了美国教育回报率的州际差异。他利用非参数方法估计人们选择不同的州的概率，并通过控制少数几个概率（如选择其实际迁移的州的概率以及选择留在出生地的概率）的多项式函数来修正样本选择偏差。他发现自我选择对各州的教育回报率有显著影响。本文在进行样本选择修正时将主要借鉴 Dahl（2002）的研究。

涉及到农民工的流动，以往的研究更多强调地区间平均工资水平差异的影响（如 Lin et al.,

^① 一些研究使用了 Heckman 的两步法来处理样本选择问题（如张泓骏和施晓霞，2006），但这只是处理了有工作和没工作的样本间的选择问题。本文则假定农民工在迁移之前已经找好工作，重点考察他们在地区间的选择。

2004)。而教育回报率的地区差异则反映出工资水平的地区差异将因劳动力的教育水平而不同，进而对于理解不同教育水平农民工的流动模式具有重要意义。

与已有研究相比，本文的贡献体现在如下三个方面。首先，我们的数据覆盖范围广，包含全国 31 个省、自治区和直辖市的 300 多个城市，数据（特别是 2005 年的人口抽样调查数据）的代表性也比较好。其次，我们放松了不同地区农民工教育回报率相等的假设，并考虑了样本选择问题。另外，我们使用了 2005 和 2011 两个年份的数据，既能反映农民工教育回报率的变化趋势，又具有较强的时效性。

三、数据描述

我们使用的第一套数据是 2005 年 1%人口抽样调查数据的五分之一随机样本。该数据涵盖了全国 31 个省、自治区和直辖市的 2585481 个人。数据中包含了农业和非农户籍人口，调查了他们的性别、年龄和教育水平等个人特征信息以及就业身份、职业、行业、工作时间和收入等与就业有关的信息。对于流动人口，还调查了他们的户口所在地、离开户口所在地的时间和迁移目的等信息。第二套数据是国家人口计生委的 2011 年流动人口动态监测调查数据。该调查采取分层、多阶段、与规模成比例的 PPS 抽样，在全国范围进行，按照随机原则在 31 个省（区、市）和新疆生产建设兵团抽取样本点。调查的主要内容包括流动人口基本信息、就业、居住、计划生育和生殖健康等公共服务情况以及社会参与和社会融合状况等。抽样过程中，各省样本量分 4 个等级，最多的为 10000 人，其次为 6000 人，再次为 4000 人，最少为 2000 人，最终全国总样本量为 128000 人。

在个人层面，我们保留了两个年份里户主的相关信息，包括性别、年龄、教育程度、户口性质、户口登记地情况、流动范围、离开户口所在地时间（2005）或本次来本市时间（2011）、行业、职业、就业身份、上月收入和每周工作时间等。由于本文主要关注农民工的教育回报率，我们将样本限制在有工资收入的跨县级及县级以上的地区流动、流动时间在半年及以上的农业户籍人口。按照通常的做法，我们在样本中排除了雇主、自营劳动者、退休者、学生以及家庭帮工（Coleman, 1993; Mwabu and Schultz, 1996）。考虑到我国的相关规定（如最低工作年龄要求和退休年龄要求），样本年龄被限定在 16 至 59 岁。此外，我们根据数据中的月平均工资和每月工作小时数计算得到小时工资，利用年龄和教育年限计算得到潜在经验：年龄-教育年限-6。

在进行上述限定并删除缺失信息的样本后，2005 年的最终有效样本量为 30176，其中男性 22702 人，女性 7474 人；2011 年的有效样本量为 38609，其中男性 22743 人，女性 15866 人。2005 年，男性和女性的平均年龄分别为 31.87 和 26.28 岁；2011 年则分别为 32.78 和 30.34 岁。

表 1 对两个年份的样本做了更加细致的描述。从受教育水平的统计结果可以看出，具有初中以下学历的样本在两个年份里分别占 79%和 74%，其中，初中学历的样本在两个年份里的比例都是最高的。两年间，高中及以上学历的农民工相对数量显著上升，从 2005 的 21%上升到 2011 年的 26%。这在一定程度上表明，上世纪 90 年代末以来的高等教育扩招对提高农业户籍人口的教育水平发挥了积极作用。

表 1 还报告了农民工的行业和职业分布，与流动人口的低教育水平相对应，两个年份的样本中从事对能力要求比较高的产业如科学研究、技术服务和地质勘查业的比例相对较低，样本主要集中在制造业和建筑业。总的来说，2011 年流动人口职业和行业的分布与 2005 年相比更加分散。从表中我们可以看出 2011 年从事制造业和生产的流动人口比例明显下降，而从事服务行业和其他行业以及商业服务的流动人口比例明显增多。

表 1 数据描述

变量	2005 年		2011 年	
	人数	占比 (%)	人数	占比 (%)
性别				
男性	22702	75.23	22743	58.91
教育水平				
文盲	488	1.62	611	1.58
小学	4822	15.98	5671	14.69
初中	18433	61.08	22126	57.31
高中	5570	18.46	8132	21.06
大专	722	2.39	1641	4.25
大学及以上	141	0.46	428	1.10
行业				
农林牧渔业	516	1.71	728	1.88
采矿业	622	2.06	900	2.33
制造业	17107	56.69	13988	36.23
电力燃气及水的生产和供应业	144	0.48	292	0.76
建筑业	3134	10.39	4987	12.92
交通运输、仓储和邮政业	1340	4.44	1849	4.79
批发和零售业	2544	8.43	2646	6.85
住宿和餐饮业	1680	5.57	4385	11.36
金融保险房地产	444	1.47	485	1.26
科学研究、技术服务和地质勘查业	86	0.28	361	0.93
居民服务和其他行业、社会服务	1150	3.81	4542	11.76
卫生社会保障社会福利业	131	0.43	404	1.05
教育/文化及广播电影电视	435	1.44	329	0.85
公共管理和社会组织	187	0.62	119	0.31
其他	656	2.17	2594	6.72
职业				
机关、党群组织、企事业单位负责人	374	1.24	1480	3.83
专业技术人员	1001	3.32	4069	10.54
办事人员和有关人员	1571	5.20	1484	3.84
商业服务业人员	6587	21.83	13097	33.92
农、林、牧、渔、水利业生产人员	536	1.77	523	1.35
生产、运输设备操作人员及有关人员	20107	66.63	16848	43.64
无固定职业	0	0	1108	2.87
样本量	30176		38609	

四、基本模型及计量分析结果

我们采用经典的明瑟工资方程来估计教育回报率，模型的基本形式如下：

$$\ln w = \beta_0 + \beta_1 * schyear + \beta_2 * exper + \beta_3 * expersq + \beta_4 * gender + \varepsilon \quad (1)$$

$\ln w$ 为小时工资收入的自然对数， $schyear$ 代表受教育年限，因为数据中只有受教育水平，我们根据中国的实际情况将教育水平换算成教育年限。^① $exper$ 和 $expersq$ 代表经验和经验平方。 $gender$ 是虚拟变量，男性为 0，女性为 1。式中 $\beta_0 \sim \beta_4$ 为待估参数，其中 β_1 衡量了在其他条件不变的情况下，教育年限对工资水平的边际影响。我们首先利用这个模型对总体或不同地区的样本做回归。

表 2 给出了总体的和分区域的教育回报率估计结果。从中可以看出，不管是 2005 年还是 2011

^① 本文为不同教育水平赋予教育年限的方式如下：文盲、小学、初中、高中、大专、大学和研究生的教育年限分别为 0、6、9、12、15、16 和 19 年。

年,流动人口教育回报率的地区差异都十分明显,其中东部和发达城市的教育回报率明显高于其他地区。从回归的 R^2 可以看出,发达地区教育和经验对收入的解释力更强。同时,随着时间的变化,所有地区的教育回报率均有所下降。不过在表 2 的回归中,我们没有控制大区域内的省份虚拟变量。因此,教育回报率的大小可能反映了地区内部不同省份间的教育和工资水平的差异。

表 2 总体和分区域的教育回报率

区域	2005				2011			
	Coef.	S.E.	Adj R2	Obs	Coef.	S.E.	Adj R2	Obs
北京和上海	0.0722	(0.0030)	0.1250	5976	0.0699	(0.0038)	0.1876	2762
东部	0.0814	(0.0017)	0.1373	20111	0.0480	(0.0015)	0.1369	18390
中部	0.0423	(0.0055)	0.1085	2067	0.0416	(0.0026)	0.1556	8602
西部	0.0558	(0.0047)	0.1118	2022	0.0397	(0.0023)	0.1223	8855
整体	0.0751	(0.0014)	0.1248	30176	0.0466	(0.0011)	0.1245	38609

注：西部省份包括重庆市、四川省、云南省、贵州省、西藏自治区、陕西省、甘肃省、青海省、宁夏回族自治区、新疆维吾尔自治区；东部包括天津市、河北省、辽宁省、江苏省、浙江省、福建省、广东省、海南省、山东省；中部包括山西省、内蒙古自治区、吉林省、黑龙江省、安徽省、江西省、河南省、湖北省、湖南省、广西壮族自治区。回归方程中还控制了经验、经验平方和性别虚拟变量。所有系数均在1%的水平上显著；括号中为标准误。

表 3 给出了分省的估计结果。从 1-4 列可以看出,不同省份农民工的教育回报率存在巨大的差异。2005 年,教育回报率最高的省份为广东,基本上达到 10%,最低的省份为吉林,其教育回报率甚至为负(不显著)。总体来看,东部省份的教育回报率相对较高。到 2011 年,不同省份的教育回报率差异仍然十分显著,其中北京的教育回报率最高,为 8.3%(标准误为 0.006);黑龙江的教育回报率为-0.2%。总体上仍然是东部地区的教育回报率高于中西部地区。

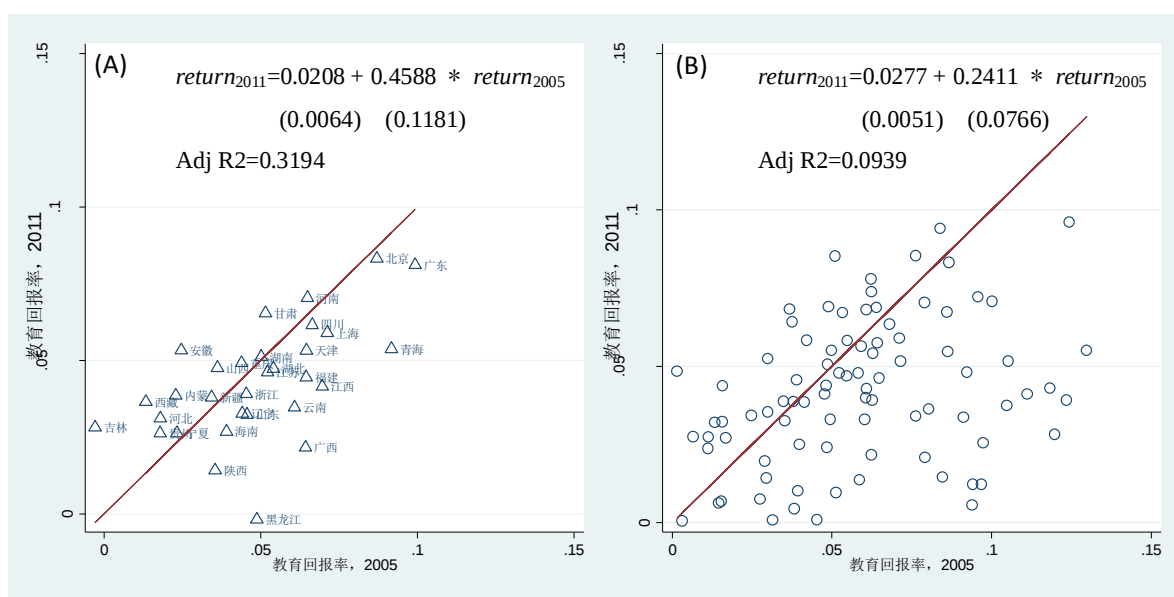


图 1 省级层面 (A) 和城市层面 (B) 的教育回报率, 2005 和 2011

比较 2005 和 2011 年的结果可以看出，教育回报率的地区差异有所降低。所有省份教育回报率的标准差由 2005 年的 0.024 下降到 2011 年的 0.019。图 1-A 则更形象的说明了教育回报率地区差异的变化模式。图中的实线为 45 度线，横轴和纵轴则分别代表 2005 和 2011 年的教育回报率。可以看出，两个年份不同省份的教育回报率之间存在显著的正相关关系，大部分省份教育回报率的相对位置变化不大。而大部分省份位于 45 度线的下方则意味着总体教育回报率有所下降。同时，一些 2005 年教育回报率很低的省份的教育回报率到 2011 年有所上升，这说明，在总体教育回报率下降的同时，不同地区的技能价格有一个向均值回归的过程。反映到两个年份教育回报率的回归中，2005 年教育

回报率的系数显著为正，但小于 1（为 0.4588）。

表 3 样本修正前后各省的教育回报率

省份	2005 年		2011 年		2005 年		2005 年			
	未进行样本选择修正				未进行样本选择修正		经过样本选择修正			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
北京	0.087	(0.007)	0.083	(0.006)	0.077	(0.007)	0.088	(0.007)	0.077	(0.007)
天津	0.065	(0.007)	0.053	(0.006)	0.054	(0.007)	0.067	(0.007)	0.057	(0.007)
河北	0.018	(0.020)	0.031	(0.006)	0.016	(0.020)	0.042	(0.022)	0.037	(0.023)
山西	0.036	(0.016)	0.048	(0.010)	0.032	(0.016)	0.036	(0.016)	0.031	(0.016)
内蒙古	0.023	(0.012)	0.039	(0.007)	0.019	(0.012)	0.021	(0.012)	0.017	(0.012)
辽宁	0.044	(0.014)	0.033	(0.005)	0.032	(0.014)	0.038	(0.014)	0.027	(0.014)
吉林	-0.003	(0.021)	0.028	(0.012)	-0.023	(0.022)	0.004	(0.024)	-0.013	(0.025)
黑龙江	0.049	(0.014)	-0.002	(0.007)	0.047	(0.014)	0.053	(0.014)	0.051	(0.014)
上海	0.071	(0.003)	0.059	(0.005)	0.058	(0.003)	0.064	(0.003)	0.053	(0.003)
江苏	0.052	(0.005)	0.046	(0.004)	0.041	(0.005)	0.051	(0.005)	0.040	(0.005)
浙江	0.045	(0.004)	0.039	(0.003)	0.038	(0.004)	0.043	(0.004)	0.036	(0.004)
安徽	0.025	(0.019)	0.053	(0.008)	0.003	(0.018)	0.009	(0.020)	-0.004	(0.019)
福建	0.065	(0.006)	0.045	(0.004)	0.056	(0.006)	0.065	(0.006)	0.057	(0.006)
江西	0.070	(0.021)	0.042	(0.007)	0.064	(0.022)	0.069	(0.022)	0.067	(0.023)
山东	0.046	(0.009)	0.032	(0.006)	0.028	(0.009)	0.046	(0.009)	0.028	(0.009)
河南	0.065	(0.027)	0.070	(0.006)	0.048	(0.026)	0.071	(0.031)	0.051	(0.029)
湖北	0.054	(0.016)	0.047	(0.011)	0.043	(0.016)	0.079	(0.019)	0.068	(0.019)
湖南	0.050	(0.019)	0.051	(0.009)	0.035	(0.018)	0.050	(0.022)	0.032	(0.021)
广东	0.099	(0.002)	0.081	(0.003)	0.081	(0.002)	0.101	(0.002)	0.083	(0.002)
广西	0.064	(0.020)	0.022	(0.008)	0.050	(0.022)	0.065	(0.021)	0.050	(0.022)
海南	0.039	(0.017)	0.027	(0.007)	0.033	(0.017)	0.053	(0.019)	0.047	(0.019)
重庆	0.044	(0.019)	0.049	(0.006)	0.034	(0.019)	0.027	(0.023)	0.012	(0.023)
四川	0.066	(0.020)	0.062	(0.007)	0.042	(0.021)	0.072	(0.024)	0.049	(0.025)
贵州	0.018	(0.016)	0.026	(0.009)	0.017	(0.016)	0.012	(0.017)	0.011	(0.017)
云南	0.061	(0.008)	0.035	(0.007)	0.054	(0.008)	0.049	(0.009)	0.042	(0.010)
西藏	0.013	(0.109)	0.037	(0.009)	0.079	(0.139)	0.071	(0.099)	0.088	(0.132)
陕西	0.036	(0.013)	0.014	(0.007)	0.028	(0.014)	0.030	(0.014)	0.021	(0.015)
甘肃	0.052	(0.019)	0.065	(0.008)	0.044	(0.019)	0.050	(0.019)	0.043	(0.020)
青海	0.092	(0.026)	0.054	(0.012)	0.078	(0.027)	0.093	(0.027)	0.081	(0.028)
宁夏	0.023	(0.020)	0.026	(0.012)	0.005	(0.022)	0.023	(0.021)	0.002	(0.023)
新疆	0.034	(0.011)	0.038	(0.007)	0.029	(0.011)	0.033	(0.011)	0.027	(0.011)
福利和迁移	no		no		yes		no		yes	
时间										
标准差	0.024		0.019							

注：回归方程中还控制了经验、经验平方、性别虚拟变量和常数项；括号中为标准误。

最后，我们还估计了市级层面的教育回报率。但是由于很多城市的农民工样本较小，导致教育回报率的估计不准确。图 1-B 是利用两年教育回报率均为正且剔除异常值之后的 87 个市的数据画出的散点图，其中横轴和纵轴分别代表 2005 和 2011 年的教育回报率，图中的直线是 45 度线。图 1-B 表现出的规律与图 1-A 基本一致。我们可以看出教育回报率有一个趋同的趋势：2005 年教育回报率比较低的城市到 2011 年教育回报率下降程度相对较小，甚至有一部分城市的教育回报率有所上升；而 2005 年教育回报率较高的市到 2011 年教育回报率下降幅度相对比较大。两方面的原因可能导致了教育回报率的上述变化。首先，技能价格（教育回报率）的地区差异激励高技能人口流动到回报率比较高的地方，劳动力供给结构的变化使得低回报地区的教育回报率相对上升，高回报地区的教育回报率相对降低；另一方面，也可能是地方政府教育政策的相应调整使得本地劳动力市场的

供给结构发生变化，进而使农民工的教育回报率发生不同程度的变化^①。在后面的分析中，我们将对第一个假说进行验证。不过在此之前，我们首先考察所得到的教育回报率地区差异是否稳健，以及它是否受到样本选择偏差的影响。

五、稳健性检验和样本选择修正

如果上节中教育回报率的地区差异是对人力资本价格的真实反映，它将对不同技能水平劳动力的区域选择产生重要影响。不过，这种差异可能只是个人或工作特征的地区差异的反映，或者是劳动力自选择的结果。这种情况下，观测到的教育回报率地区差异不会为劳动力的跨区流动提供激励。本节将通过一些稳健性检验以及样本选择修正来说明这些因素对教育回报率地区差异的解释力。

（一）控制福利状况和迁移时间

首先，我们考察社会福利状况的地区差异可能对结果造成的影响。不妨想象两个城市 A 和 B，并假设两个城市中低技能的农民工收入和福利状况相同。对于高技能的劳动力而言，A 地区提供的工资高但是社会福利较差，B 地区的工资低但是社会福利较好。那么在均衡的时候，我们完全可能观测到 A 地区的教育回报率高于 B 地区。为了考察这种可能性，我们在工资方程中加入了反映农民工参加社会保险状况的控制变量，包括：是否参加失业保险、是否参加养老保险和是否参加基本医疗保险（三者均为虚拟变量）。

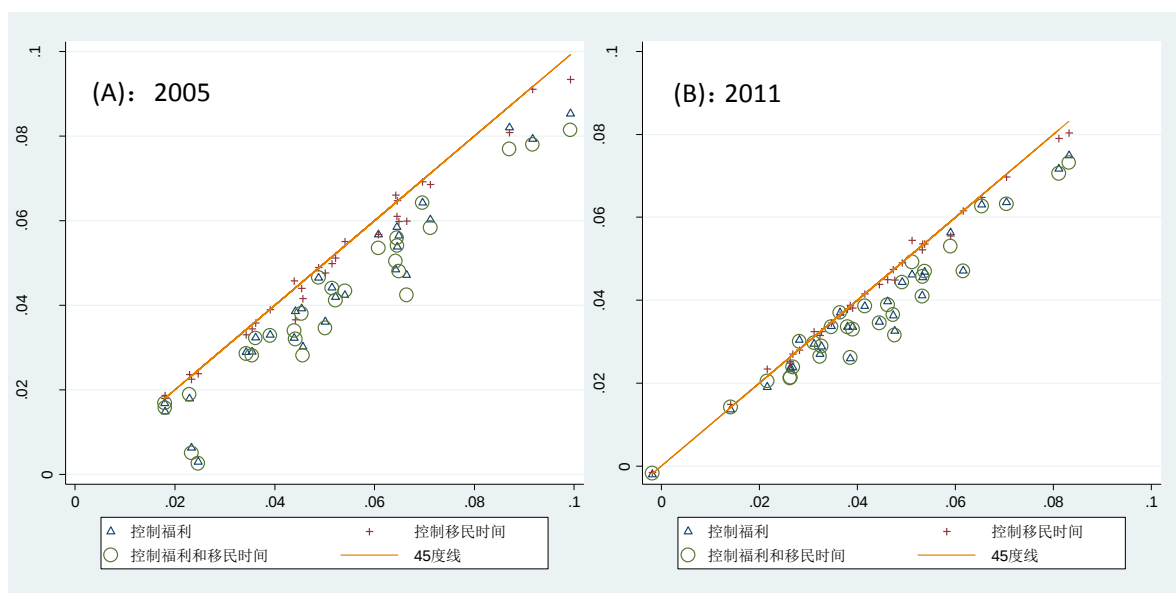


图 2 控制变量对教育回报的影响

图 2A 和 B 分别给出了 2005 和 2011 年控制这些变量之后的估计结果。横轴代表未加这些控制变量时的教育回报率，纵轴则代表加了不同的控制变量之后的教育回报率。图中大部分省份在控制了社会保险状况后的教育回报率都低于 45 度线（图中的实线），这说明高教育水平农民工的福利状况要好于低教育水平的农民工，这使得控制福利状况后的教育回报率下降。但是，2005 年各省教育回报率下降的幅度没有系统的差异；2011 年高教育回报率地区的教育回报下降幅度略高于低回报地区，但这没有显著降低教育回报率的地区差异，也基本未改变不同地区在教育回报率上的排序。

接下来，我们考察劳动力流动时间差异可能带来的影响。同样考虑 A 和 B 两个城市，并假设两

^① 由于来自不同的调查，将 2005 年和 2011 年各地区的教育回报率进行比较多少有些令人担忧。因此，我们同时利用相同来源（国家统计局）的两个年份的数据考察了地区间教育回报率趋同的特征。结果显示，尽管是两个相邻年份（2011 和 2012）的数据，地区间的教育回报率仍然表现出趋同的趋势。两年间，教育回报率出现下降的主要是 2011 年教育回报率偏高的省份，而那些上升的省份则主要是教育回报率偏低的省份。如果拿 2012 年的教育回报率对 2011 年的做回归，后者的回归系数为 0.43，说明在较短的时间里，流动人口教育回报率的地区差异明显下降。尽管如此，到 2012 年，流动人口教育回报率的地区差异仍然十分明显。一些沿海省份的教育回报率仍然接近 7%，而一些内陆省份的教育回报率则仅为 3%。需要指出的是，由于数据代表性的问题，这里教育回报率均没有进行样本选择修正。

个城市中低技能的农民工收入和迁移时间相同^①。如果 A 城市高技能劳动力的迁移时间较长（比如经济发展较早），而同类型的劳动力只是近期才迁移到 B 城市，我们就可能观察到 A 城市的教育回报率高于 B 城市——因为在城市地区生活的时间越长，劳动力对城市环境越熟悉，越可能赚更高的工资。为了排除这种可能，我们在工资方程回归中控制了相关的因素。对于 2005 年的数据，我们控制了劳动力离开户口所在地的时间；而对于 2011 年的数据，我们控制了农民工来“本市”的时间。无论是 2005 年还是 2011 年，控制移民时间后的各省教育回报率都在 45 度线附近，表明这对各省教育回报率的估计基本没有影响。因此也不难想象，当我们同时控制社会福利和迁移时间后，估计结果与单独控制社会福利时的结果相差不大。

（二）修正样本选择偏差

以上我们使用迁移到各地区的流动人口样本通过 OLS 方法得到各地区的教育回报率。由于劳动力往往根据自身的比较优势选择地区，这可能导致 OLS 估计结果有偏。我们参考 Dahl（2002）的方法，利用 2005 年的数据分别对省和市层面的估计进行样本选择修正^②。下面我们通过一个简化的模型来说明样本选择偏差存在的原因，进而提出修正的办法。

我们假设农民工在选择地区时遵循效用最大化原则。如果出生在 j 地区的个人 i 迁移到地区 k，则其效用为：

$$V_{ijk} = V_{jk} + e_{ijk} \quad (k=1, \dots, N) \quad (2)$$

其中， V_{jk} 是由可观测的特征决定的效用， e_{ijk} 是由（对研究人员）不可观测的特征决定的效用，其中既包括影响收入的不可观测因素也包括影响个人喜好的不可观测因素。只有农民工在 k 地区的效用最大时，他才会选择这个地区：

$$M_{ijk} = 1, \text{ 当且仅当 } V_{jk} + e_{ijk} \geq V_{jm} + e_{ijm}, \forall m, \text{ 否则 } M_{ijk} = 0. \quad (3)$$

我们假设 k 地区的工资由下式决定：

$$y_{ik} = \alpha_k + x_i' \delta_k + s_i \beta_k + u_{ik} \quad (k=1, \dots, N) \quad (4)$$

其中， α_k 代表特定地区的常数项， x_i' 代表个人特征向量，包括经验、经验平方以及性别， s_i 表示受教育年限， u_{ik} 代表误差项。当利用迁移到 k 地区的样本进行回归时，

$$E[u_{ik} | M_{ijk} = 1] = E[u_{ik} | e_{ijm} - e_{ijk} \leq V_{jk} - V_{jm}, \forall m] \neq 0 \quad (5)$$

如果上述条件期望与 x_i' 或 s_i 相关，OLS 结果就是有偏的。由上式也可以看出，样本选择偏差是由不可观测的收入决定因素在地区间的相关性以及不可观测的收入决定因素与不可观测的农民工地区偏好之间的相关性决定的。由于有 N 个选项，（5）式的大小取决于 $(u_{ik}, e_{ij1} - e_{ijk}, \dots, e_{ijN} - e_{ijk})$ 的联合分布，这使样本选择修正非常困难。Dahl（2002）受到 Lee（1983）的启发，将（3）式改写为：

$$M_{ijk} = 1, \text{ 当且仅当 } \max_m (V_{jm} - V_{jk} + e_{ijm} - e_{ijk}) \leq 0. \quad (3')$$

相应的，（5）式的大小取决于 u_{ik} 和 $\max_m (V_{jm} - V_{jk} + e_{ijm} - e_{ijk})$ （最大统计量，maximum order statistic）的联合分布，而这个联合分布又受到可观测特征决定的地区间的效用差异 $(V_{jm} - V_{jk})$ 的

^① 我们也可以假设迁移时间对于低技能劳动力的人力资本及其收入没有影响。

^② 之所以没有对 2011 年样本数据进行修正分析是因为 2011 年的样本数据是对全国 31 个省份采用 PPS 抽样，在修正过程中计算迁移概率时，会与人口流动的真实情况存在较大偏差。

影响。Dahl（2002）假设 $V_{jm} - V_{jk}, \forall m$ 对于联合分布的影响可以通过农民工选择某些地区的概率来反映。其中一个较强的假设是：

$$\begin{aligned} & g_{jk} \left(u_{ik}, \max_m (V_{jm} - V_{jk} + e_{ijm} - e_{ijk}) | V_{j1} - V_{jk}, \dots, V_{jN} - V_{jk} \right) \\ & = g_{jk} \left(u_{ik}, \max_m (V_{jm} - V_{jk} + e_{ijm} - e_{ijk}) | p_{ijk} \right) \end{aligned} \quad (6)$$

p_{ijk} 代表个体 i 从 j 省（市）转移到 k 省（市）的概率。由此，我们可以得到样本选择修正模型：

$$y_{ik} = \alpha_k + x_i' \delta_k + s_i \beta_k + \sum_{j=1}^N \left[M_{ijk} \times \lambda_k(p_{ijk}) \right] + \omega_{ijk}, \quad (k=1, \dots, N) \quad (7)$$

$$E[\omega_{ik} | x_i, s_i, p_{ijk}, M_{ijk} = 1] = 0 \quad (k=1, \dots, N) \quad (8)$$

（7）式中新加入的部分为样本修正函数，我们这里采用迁移概率的二项式形式。^① M_{ijk} 为虚拟变量，代表在 k 地区工作的个体 i 来自地区 j 。

为得到迁移概率，我们将所有样本按地域、年龄和教育水平分为 28 个特征组。其中地域有华北、东北、华东、华中、华南、西北以及西南等七个组^②，教育水平分为初中及以下和高中及以上两组，年龄则分为 30 岁以下和 30 岁以上两组。迁移概率则定义为所有具有相同特征的人群中（在一个特征组）迁移到某个地区（ k ）的比例，采用这种非参的方式估计样本迁移的概率避免了多项选择模型（如 mlogit 模型）中的独立于无关选项（IIA）假设，也避免了在嵌套模型中进行人为分组。

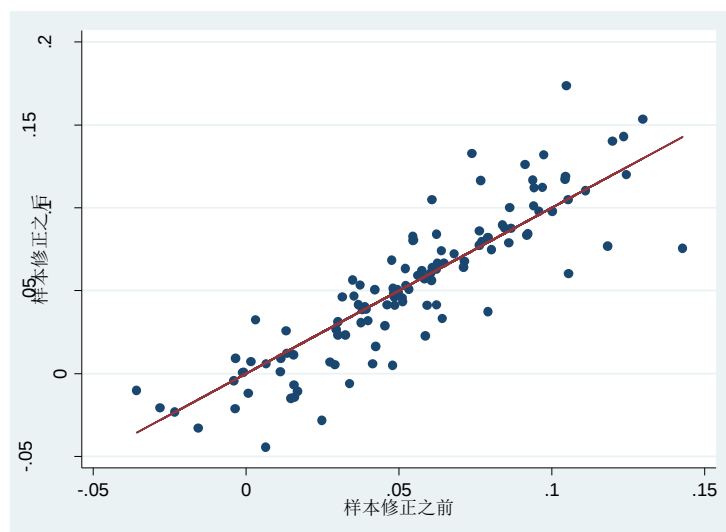


图 3 样本修正前后各市的教育回报率，2005

表 3 给出了 2005 年样本选择修正前后各省的教育回报率，其中 1、7 两列是没有控制福利状况和迁移时间的结果。在对每个省份的样本进行样本修正之后，有些省份的教育回报率有所上升，如河北、海南和西藏；有些省份的教育回报率则有所下降，如安徽和重庆。然而，大多数省份在进行样本选择修正之后教育回报率的变化很小。此外，教育回报率在样本选择修正前后的变化也没有系

^① 我们还假设从不同地区（ j ）迁移到地区 k 的样本对应的修正项函数是相同的： $\lambda_{jk}(\bullet) = \lambda_k(\bullet)$ 。

^② 华北（北京、天津、河北、山西、山东）、东北（内蒙古、辽宁、吉林、黑龙江）、华东（上海、江苏、浙江、福建）、华中（河南、湖北、湖南、安徽、江西）、华南（广东、广西、海南）、西北（陕西、甘肃、宁夏、新疆）以及西南（重庆、四川、贵州、云南、西藏、青海）。

统性的规律。5、9 两列报告了控制福利状况和迁移时间后的估计结果，总体来看，样本选择修正的影响也比较小。

图 3 利用市级层面的数据给出了样本选择修正前后的结果。该图为 2005 年 120 个（剔除了异常值）样本城市修正前后流动人口教育回报率的散点图，图中的直线为 45 度线。我们可以通过数据点偏离 45 度线的模式来考察样本修正前后教育回报率的变化。图 3 中的大部分城市主要分布在 45 度线附近，只有少数城市偏离 45 度线的程度较大。总体来看，不同城市间的农民工教育回报率仍然存在显著差异，而且大部分城市教育回报率大小的排序变化不大。这些结果均说明，样本选择偏差对估计结果的影响不大，不影响本文的结论。

六、教育回报率的地区差异与劳动力的流动

如果各省之间真实的教育回报率确实存在差异且个人根据比较优势选择迁移目的地，迁移概率则会随着教育回报率的差异而变化。我们假设地区间的收入差距和不同地区的吸引力差距会影响劳动力的迁移概率，具体的我们有以下等式：

$$\ln(p_{jk}^N) = \theta_0^N + \theta_1(y_k^N - y_j^N) + \theta_2(A_k - A_j) + v_{jk}^N \quad (9)$$

其中，上标 N 代表教育水平， p_{jk}^N 为教育水平为 N 的人当中从地区 j 流动到地区 k 的人数比例，

$\ln(\bullet)$ 代表迁移概率的自然对数， y_k^N 代表教育水平为 N 的个体在第 k 个省份的平均收入。 A 代表某地一系列特征变量，本文变量的选取包括城镇登记失业率、人口密度（人每平方千米）、人均资本存量、医疗卫生机构床位数（每千人口）和气候变量（包括各地主要城市年平均气温、降水量、相对湿度），还包括一些政府公共支出变量（即：人均财政教育费用支出、人均科学事业费、人均卫生经费支出、人均社会保障补助支出），以及用以考察距离的位置变量（主要城市经度、纬度）， v_{jk}^N 代表误差项。与以往考察收入差距对劳动力流动的影响的研究（如 Lin et al., 2004）不同，等式（9）允许不同教育水平的劳动力的地区收入差距和迁移概率存在差异，它还允许不同教育水平的劳动力对城市特征及其差异拥有不同的评价。不过，我们假定不同教育水平的劳动力对收入差异反应的敏感程度相同。

我们将农民工的教育水平分为两组，高中及以上（ $N=HS$ ）和初中及以下（ $N=MS$ ）。对这两组人群，（9）式均成立。通过将两组人群相减可得下式：

$$\ln(p_{jk}^{HS}) - \ln(p_{jk}^{MS}) = \theta_0 + \theta_1 \Delta \beta^{HS} + \theta_2 \Delta A + v_{jk} \quad (10)$$

其中 $\theta_0 = \theta_0^{HS} - \theta_0^{MS}$ ， $\theta_2 = \theta_2^{HS} - \theta_2^{MS}$ ， $v_{jk} = v_{jk}^{HS} - v_{jk}^{MS}$ ， $\Delta \beta^{HS} = \beta_k^{HS} - \beta_j^{HS}$ 。为了估计（10）

式，我们生成一个虚拟变量（教育水平为高中及以上等于 1，否则等于 0），并以此来替代每个省份的工资方程中的教育年限。同样，我们可以得到样本选择修正之后的教育回报率。表 4 给出了利用 2005 年的数据估计的结果。^①

由表 4 可以看出，结果里的截距项均显著为正，说明高教育水平人群比低教育水平人群的迁移概率更高。其他回归系数反映了地区收入差距和特征差异对不同教育水平相对迁移概率的影响。表 4 的 1-2 列验证了迁移概率对修正和未修正教育回报率的反应，两列结果都显示地区之间教育回报

^① 对于 2011 年的数据我们做了相同的工作，结果除了截距项的系数显著为正之外，修正教育回报率差异的系数基本上都不显著，虽然观察值有所增加，但是解释变量的解释能力明显下降（Adj R-squared 明显很小），我们认为主要原因在于 2011 年的数据由于抽样的原因不能很好的反映流动人口在地区间的流动情况——主要表现在无法准确估计 p_{jk}^{HS} 和 p_{jk}^{MS} 。

率的差异对不同教育水平人口流动概率的差异有显著影响。其中，第 1 列的结果表明，如果 k 地区的教育回报率比 j 地区高 1%，高教育人群从 j 地迁移到 k 地的概率就比低教育人群高 1.33%。第 2 列中的回归系数为 0.74，低于第 1 列的结果，这大概是由于未经样本选择修正的教育回报率已经在一定程度上反映了劳动力流动的结果。表 4 的 3-8 列利用修正后的教育回报率，同时控制其他变量，如失业率、社会福利状况、气候、政府支出、距离等，其结果与未加控制变量时的结果很接近：系数变化不大而且都十分显著。总体而言，表 4 的结果说明不同教育水平的人口在流动的过程中会根据自身的比较优势进行区域选择。而教育回报率的地区差异和相对迁移概率间的显著关系说明我国的劳动力资源在随着外部需求的变动而进行重新配置，这是我国劳动力市场日趋灵活和不断完善的表现。近些年的户籍制度改革可能是劳动力市场得以改善的原因之一。

表 4 流动概率与教育回报率的地区差异，2005 年

$\ln(p_{jk}^{HS}) - \ln(p_{jk}^{MS})$	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
修正教育回报率的地区差异	1.3255*** (0.3496)		1.2519*** (0.3533)	1.3710*** (0.3702)	1.3038*** (0.3729)	1.1136** (0.3750)	1.3260*** (0.3525)	1.2948** (0.4809)
未修正教育回报率地区差异		0.7366* (0.3422)						
失业率			-0.0723 (0.0536)					-0.1671 (0.0857)
人口密度				-0.00002 (0.0001)				0.0002 (0.0002)
资本存量				0.1304 (0.0740)				-0.1532 (0.1519)
各地区医疗卫生机构床位数				-0.1505 (0.0883)				-0.1199 (0.1552)
气候					Y			Y
政府支出						Y		Y
距离							Y	Y
截距	0.1139* (0.0497)	0.1274* (0.0512)	0.0957 (0.0515)	0.1358* (0.0601)	0.1195* (0.0522)	0.1785** (0.0610)	0.1404 (0.0969)	0.1498 (0.1019)
Adj R-squared	0.0444	0.0125	0.0471	0.0567	0.0710	0.0648	0.0409	0.0906
Obs	289	289	289	271	289	289	286	268

注：*、**、***分别表示 10%、5%、1%的显著性水平；括号中为标准误。各地区失业率、医疗卫生机构床位数以及气候、政府支出（除

财政教育费用支出）数据来自《中国统计年鉴（2006）》；资本存量数据来自张军的研究；财政教育费用支出来自

<http://wenku.baidu.com/view/7be01d81ec3a87c24028c4f7.html>；距离变量根据主要城市经纬度换算而来，主要城市经度、纬度数据来自

<http://wenku.baidu.com/view/1d307558312b3169a451a4eb.html>。

七、结论

农民工的教育回报率及其地区差异既影响到这一群体的收入水平也影响到不同教育水平农民工的区域选择，对理解中国的劳动力市场和制定相关政策具有重要意义。本文利用 2005 年 1%人口抽样调查数据和 2011 年流动人口动态监测调查数据，估计了不同地区农民工的教育回报率。结果表明，农民工的总体教育回报率存在下降的趋势。更为重要的是，我们发现 2005 年和 2011 年农民工的教育回报率都存在显著的地区差异，两年里各地区的教育回报率存在显著的正相关性。我们也观察到农民工的区域间流动使得教育回报率的地区差异有所下降并呈现出向均值收敛的特征。

近年来，伴随着教育扩展，我国劳动力（包括农民工）的教育水平显著提高。如何有效的利用劳动力的人力资本也成为全社会关注的问题。本文的结果表明，教育回报率的地区差异已经对我国不同教育水平劳动力的区域选择产生了显著的影响。在一个可以自由流动的劳动力市场中，如果劳动力的偏好相同，区域间的流动将促使地区间的教育回报率趋同（Black et al., 2009）。由于农民工

是一个流动性和同质性较强的群体，教育回报率仍存在显著的地区差异说明他们在流动过程中面临较高的成本。当然，即便不存在流动成本，如果劳动力的偏好存在差异，地区的产业结构、生活便利设施的差异和劳动力的自选择过程也会导致教育回报率的地区差异。本文的结果对于控制福利状况、迁移时间以及样本选择偏差的修正非常稳健，同时这种地区差异有随时间下降的趋势，这说明教育回报率的地区差异不大可能是农民工在区域选择过程中的偏好异质性所导致的。因此，政府部门可以通过致力于减少劳动力流动的阻力（如取消或进一步改革现在的户籍制度）来促进劳动力的流动。

对于那些具有较高教育回报率的地方政府而言，一方面要努力吸引更多高教育水平的劳动力，另一方面也可以通过培训项目来增加现有劳动力的技能。而一旦劳动力（特别是高教育水平的劳动力）可以自由流动，一些地区的高教育回报率也可以为低教育回报地区的人力资本投资提供激励，从而提高我国的总体教育水平。对这一论断的实证检验超出了本文的范围，但是现有的很多研究都支持这一结论（如 Beine et al., 2008；邢春冰，2013）。^①

参考文献

- 邓曲恒，2007：《城镇居民与流动人口的收入差异》，《中国人口科学》第 2 期。
- 高梦滔、张颖，2007：《教育收益率、行业与工资的性别差异：基于西部三个城市的经验研究》，《南方经济》第 9 期。
- 罗忠勇，2010：《农民工教育投资的个人收益率研究——基于珠三角农民工的实证调查》，《教育与经济》第 1 期。
- 戎建，2008：《迁移回报率与中国农村劳动力流动》，《中国经济问题》第 5 期。
- 邢春冰，2005：《不同所有制企业的工资决定机制考察》，《经济研究》第 6 期。
- 邢春冰，2008：《农民工与城镇职工的收入差距》，《管理世界》第 5 期。
- 邢春冰，2013，《教育扩展、迁移与城乡教育差距——以大学扩招为例》，《经济学（季刊）》，已接收。
- 姚先国、张海峰，2004：《中国教育回报率估计及其城乡差异分析》，《财经论丛》第 11 期。
- 张泓骏、施晓霞，2006：《教育、经验和农民工的收入》，《世界经济文汇》第 1 期。
- 张世伟、赵亮、万相昱，2009：《人力资本对农村迁移劳动力收入的影响研究》，《重庆大学学报（社会科学版）》第 15 期。
- Ahn, H., J. L. Powell, 1993, "Semiparametric Estimation of Censored Selection Models with a Nonparametric Selection Mechanism", *Journal of Econometrics*, 58 (1): 3-29.
- Beaudry, P., E. Lewis, 2012, "Do Male-Female Wage Differentials Reflect Differences in the Return to Skill? Cross-City Evidence from 1980-2000", *NBER Working Paper*, No. 18159.
- Beine, M., Frederic D., and Rapoport H., 2008, "Brain Drain and Human Capital Formation in Developing Countries: Winners and Losers", *Economic Journal*, 118(528): 631-652.
- Black, D., N. Kolesnikova, and L. Taylor, 2009, "Earnings Functions When Wages and Prices Vary by Location", *Journal of Labor Economics*, 27(1): 21-47.
- Coleman, M. T., 1993, "Movements in the Earnings-Schooling Relationship, 1940-88", *Journal of Human Resources*, 28(3): 660-680.
- Dahl, G. M., 2002, "Mobility and the Return of Education: Testing a Roy Model with Multiple Markets", *Econometrica*, 70(6): 2367.
- Heckman, J. J., 1979, "Sample Selection Bias as a Specification Error", *Econometrica*, 47(1):153-162.
- Lee, L. F., 1983, "Generalized Econometric Models with Selectivity", *Econometrica*, 51(2): 507-512.
- Lin, J.Y., G. W. Wang, Y. H. Zhao, 2004, "Regional Inequality and Labor Transfers in China", *Economic Development*

^① Beine et al. (2008) 指出跨国移民中高教育水平劳动力迁移概率增加会激励输出国居民进行更多的人力资本投资；邢春冰（2013）将大学扩招当做一次自然实验，其结果表明，成功迁移到城市的概率增加将激励农村居民进行更多的人力资本投资。

and Cultural Change, 52(3):587-603.

Maurer-Fazio, M., N. Dinh, 2004, “Differential Rewards to, and Contributions of, Education in Urban China's Segmented Labor Markets”, *Pacific Economic Review*, 9(3): 173-189.

Meng, X., J. Zhang, 2001, “The Two-Tier Labor Market in Urban China: Occupational Segregation and Wage Differentials between Urban Residents and Rural Migrants in Shanghai”, *Journal of Comparative Economics*, 29(3): 485-504.

Mwabu, G., T. P. Schultz, 1996, “Education Returns across Quantiles of the Wage Function: Alternative Explanations for Returns to Education by Race in South Africa”, *American Economic Review*, 86(2): 335-339.

Regional Distribution of the Return to Education for Rural-to-Urban Migrants and Its Impact on Migration

Xing Chunbing Jia Shuyan Li Shi

(School of Economics and Management, Beijing Normal University)

Abstract: Using two large survey data in 2005 and 2011, this paper finds a significant regional difference in the return to education for rural-to-urban migrants: the east and the rich cities have higher returns than the central and west regions. Between 2005 and 2011, the overall return to education of migrant workers declined, and the regional difference shows the characteristics of convergence to the mean. These results are robust when we control for the welfare condition and migration time, and after we correct for the sample selection bias. We also find that migration flows were responsive to the regional difference in the returns to education. Further reducing movement barriers will help sustain China's economic growth.

Key Words: Migrant Workers; the Return to Education; Regional Difference

JEL Classification: C13, I21, J31