

迁移、自选择与收入分配

——来自中国城乡的证据

邢春冰*

摘要 本文强调两个事实：第一，农村居民并非同质，他们根据自身的特征决定迁移与否；第二，那些成功地改变了户籍身份的永久移民与没能改变户籍身份的临时移民之间有着显著的差异。利用2002年的CHIP数据发现，永久移民的“正选择”效应非常明显，它使得农村地区教育水平较高、处于收入分布较高位置的样本减少。这导致农村的收入水平和农村内部的不平等程度降低，阻碍了城乡差距的降低。相反，临时移民的选择效应几乎可以忽略不计。

关键词 迁移，自选择，收入分配

一、引言

近年来，日益扩大的城乡收入差距和大规模的城乡移民已成为研究人员和政策决策者非常关注的问题。¹很显然，迁移是收入差距的结果。正因为收入差距的存在，才有了大规模的迁移；与此同时，迁移也被视为减少收入差距的重要途径。然而，可能存在的反向的因果关系往往被忽视了：迁移也可能是城乡收入差距扩大的原因。本文要强调的是，农村人口不是同质的，他们自我选择迁移与否。迁移是否缩小了收入差距，在很大程度上取决于谁在迁移以及迁移的模式。如果农村移民主要是由有能力和受过更多教育的人组成（“正选择”），同时这部分人能很好地融入城市劳动力市场不再返乡，城乡差距可能会增加（因为剩余的农村人口的收入水平降低了）。如果移民是随机地从农村人口中选择出来的，迁移则更有可能缩小城乡收入差距。

* 北京师范大学经济与工商管理学院、收入分配与贫困研究中心。通信地址：北京市新街口外大街19号，100875；电话：(010)58807847；E-mail: xingchunbing@gmail.com。作者感谢北京师范大学的李实教授、西安大略大学的Aldo Colussi教授、北京大学的陈玉宇教授的指导，感谢在西安大略大学、北京师范大学和OECD发展中心的研讨会上的与会者以及两位匿名审稿人的宝贵意见和建议。作者同时感谢张佳同学出色的翻译工作。本研究得到了教育部人文社科青年项目“城乡劳动力流动与收入差距”（编号08JC790008）、IDRC/CIGI中国青年学者贫困研究网络，以及北京师范大学劳动经济学创新群体项目的资助，在此一并致谢。文责自负。

¹ 1990年，我国城乡人均年收入的比例是2.2:1，到2004年该比例已上升至3.2:1(NBS, 2005；同时可参见李实和岳希明, 2004)。期间，农村向城市的迁移也大幅增加了。根据农业部的数据，农村的迁移人口（主要指农民工）已从1982年的200万上升到了2004年的1.02亿(Cai and Wang, 2007)。

迁移中自我选择的模式同样也是理解城乡内部收入差距变化的关键。如果移民是农村人口的随机子样本,那么在技能价格不变的假设下²,留在农村地区的人口的收入分配将保持不变。但如果移民并非随机产生,在相同的假设条件下(即技能价格不变),收入分布的形状就会发生改变。同样,移民的特征对城市居民的收入分配也有重要影响。如果迁移存在明显的“负选择”效应,伴随低技能劳动力的增加,城镇地区的不平等有可能会增加。³ 本文接下来的重点就是讨论移民的选择性及其与城乡收入差距以及农村地区收入分配间的关系。

尽管已有许多研究探讨了迁移对不平等和减少贫困的影响,⁴ 仍有一个基本问题尚待回答:“如果没有迁移,这些实际的移民在他的家乡会得到多少收入?”对于本文研究的主题来说,这个问题非常重要。例如,研究迁移对农村地区内部收入差距的影响,本质上是要作一项事实(factual)与反事实(counterfactual)的比较。事实就是我们观察到的农村留守者的收入差距;而要构造反事实的收入差距,农村样本中必须包括已迁移的工人——因为如果不迁移,他们将在农村地区工作。本文更感兴趣的是他们的反事实收入,而非他们在城市中的实际收入。同样的逻辑也适用于研究城乡收入差距问题。事实乃是农村留守者与包含移民尤其是永久移民在内的城市工人间的收入差距;而为了构造反事实,必须知道如果不迁移,这些移民在农村能够得到的收入。

本文利用了 DiNardo, Fortin, and Lemieux (1996) (后面简称为 DFL) 的框架来构造反事实收入分布,基本思路如下。农村留守者的收入分布由农村地区的收入结构(技能价格)以及留守者的特征分布决定。移民的反事实收入则对应农村地区的收入结构以及移民的特征。两者的区别仅在于特征分布的差异。为了得到移民的反事实分布,我们只需为农村留守者的收入样本重新设置权重,以反映移民与他们的特征差异,即可得到移民的反事实分布(具体内容见第四部分)。DFL 方法构造反事实的思路与 Oaxaca (1973) 提出的方法在概念上是一致的。Oaxaca 分解主要基于“具有平均特征水平的移民如果留在农村将会得到多少收入”这样的反事实。DFL 方法与 Oaxaca 方法的区别在于,后者仅集中研究平均水平,而前者研究整个分布。仅集中研究平均水平将掩盖许多信息:有可能分布改变了而均值保持不变。而 DFL 方法可

² 当然,这是一个很强的假定。当一些农村居民迁移出来,均衡的技术价格也许也会改变,特别是当技能分布在正选择的情况下改变时。但是,我们需要这个假设条件来构造反事实的收入分配。这个假定在 Dinardo, *et al.*, (1996) 以及 Chiquiar and Hanson (2005) 中也被使用过。

³ Borjas (1987, 1999) 讨论了国际迁移过程中的选择性问题,并作了全面的总结。

⁴ Du, *et al.* (2005) 发现迁移户的人均家庭收入比非迁移户平均高 8.5%—13.1%。由于大部分穷人都没有迁移,迁移对于贫困的影响很小。他们还发现,移民将很大一部分收入汇回家,满足其他家庭成员的需要。李实(1999)也发现汇款对于迁移家庭的人均收入起着重要的作用。相关的研究还有 Cai and Wang (2006)。

以确切地分析出到底是收入分布的哪个部分受迁移影响最大。⁵

本文利用了2002年的CHIP数据中的三个数据集（见数据介绍部分）来考察哪些人迁移到城市，以及这些人的技能和收入与那些留在农村地区的人相比有何不同。已经有很多研究考察了移民的特征，并将他们与城市居民以及农村留守者比较。⁶然而，现存的大多数文献的研究对象是那些迁移到城市但是没有获得城市户口的人，那些成功地获得了城市户口的移民很多时候被忽略了⁷。本文的特点之一就是同时考虑了两种类型的移民：临时移民和永久移民。我们将永久移民定义为那些获得了城市户口的移民，而临时移民则指没有获得城市户口的移民（即通常所说的农民工，下文中我们交替使用农民工和临时移民，两者是同一含义）。⁸描述性统计和计量分析结果都表明，这种区分对于了解迁移对城乡内部以及城乡之间的收入分配的影响都非常重要。结果表明永久移民比临时移民更具有选择性。他们受过更高的教育，迁移到城市是为了获得更高的工资和技能价格，如果留在农村他们将更集中地处于收入分配的中上部分。⁹此外，本文还发现永久移民的组成也有明显的改变。相比之下，临时移民的选择效应几乎可以忽略不计，它对农村收入水平和不平等水平都没有显著的影响。

本文由以下几部分组成。第二部分介绍和说明本文使用的数据；第三部分介绍了构造反事实收入分布的方法；第四部分给出了基本的实证结果，同时探讨了迁移对农村地区收入分配和城乡收入差距的影响；最后总结全文。

二、数据和描述统计

本文所使用的数据来自2002年中国家庭收入调查（CHIP），它由中国社

⁵ 另一个值得注意的方法是 Machado and Mata(2005)提出的，它也可以用来构造整个反事实分布。这种方法以刻画收入的条件分布（分位回归）为基础，为了构造反事实，他们使用重新采样程序得到与条件模型（conditional model）和协变量密度（covariate densities）一致的边缘分布（marginal distribution）。显然，因为使用了参数模型，MM方法必须受到限制，重新采样的程序也可能比较麻烦。相比之下，DFL方法采用的是非参数的核密度方法，所受限制较少，并且容易实现。

事实上，DFL和MM方法在概念上非常类似，使用这两种方法的研究者并没有发现其中任何一个方法显著优于对方。一个典型的研究是 Autor, *et al.* (2005)，他们扩展了MM方法来考察工资不平等变化的组成（composition）效应和价格效应。将其结果与 Lemieux(2005)比较，他们发现两种方法本身的差异并不足以导致不同的结论（他们的确得出了不同的结论，但并不是因为使用的方法不同）。但在本文中，DFL再加权（re-weighting）方法更容易用来计算反事实的不平等，而且很容易修改此方法来考察劳动参与的差别。因此，本文只使用DFL方法。

⁶ Zhao(2005)给出了很好的总结。总的来看，农村移民比非移民更年轻且受过更多的教育，他们中的大多数为中学或小学学历，女性和少数民族比例较低。但这些总结只适用于临时移民。

⁷ Deng and Gustafsson(2006)是少数几个例外之一。

⁸ 两种类型移民的划分与中国的户籍制度密切相关。一些文章对中国城乡迁移的制度背景尤其是户籍制度作了很好的介绍和分析，本文不再赘述。感兴趣的读者可参考 Zhao(2005)、Deng and Gustafsson(2006)、Wang and Cai(2006)等。

⁹ 李实和岳希明(2004)以及其他许多研究中国收入分配变化的文章都发现城乡收入差距扩大了，而农村内部的收入差距变化不大甚至微有降低，本文的结果与这些发现是一致的。

会科学院与国家统计局在2003年初合作完成。其中共有三个数据集,即城市住户调查、农村住户调查和城市农民工调查。城市住户调查涵盖了2个直辖市和10个省份中77个城市的6835个家庭户的信息。农村住户调查涵盖22个省份的122个县的9200户家庭。应当注意,农村和城市样本的抽样对象均为拥有本地户口的户籍人口。与此同时,无论是城市还是农村的数据都随机来自国家统计局调查的更大样本。因此,本文使用的数据具有全国代表性(见Deng and Gustafsson, 2006; Li, *et al.*, 2006)。城镇和农村调查数据分别涵盖了20632个和37969个个人信息,分别占整个样本的35.21%和64.79%,这与国家统计局2005年提供的数字39.09%和60.91%非常接近。¹⁰

(一) 两个临时移民(农民工)样本

为研究城乡迁移问题,CHIP数据中包含了一个城市农民工调查,它涵盖了与城市住户调查相同辖区内的2000个迁移家庭中的5327个人。然而,这个数据集对于本文的研究有两点不足。首先,由于调查的所有移民都是农村户口,它不包括永久移民。其次,样本可能更能代表迁移时间较长的农民工,这种代表性偏差缘于CHIP移民调查的抽样过程。正如Deng and Gustafsson (2006)评论该数据集时所言:“住在宿舍和工地的农民工很难找到,而那些有住房条件的农村移民与城市登记人口一样,更容易被取样。”

为了克服第二个缺点,我们从农村住户调查中识别出另一个农民工样本。2002年的农村调查显示,37969名农村居民中大约有3800个农村居民在外打工,我们也称其为临时移民或农民工。¹¹但是,这个样本也有代表性问题。首先,很难辨认这些移民是到其他农村地区还是城市地区。其次,来自农村住户调查的移民样本可能无法涵盖那些已经迁出去很长时间和那些举家搬迁的移民。幸运的是,虽然上述两个农民工样本都存在代表性问题,但误差是呈相反的方向。由于这两个数据集并非相互排斥的,我们可以分别考察这两个移民样本。如果两个样本得到的结果差异不大,那么样本的代表性问题就不严重。为方便起见,有时将来自城市农民工调查的临时移民记为第一类移民,来自农村住户调查中的记为第二类移民;而将接下来构造的永久移民样本记为第三类移民。

(二) 永久移民

在2002年的城市调查中,城镇居民会被问到:“你是什么时候获得城市

¹⁰ 农村和城市样本能够代表全国人口对于本文的分析非常重要。构造反事实工资密度时,很幸运本文使用的样本具有代表性。为了保证结果对此问题具有稳健性,我们尝试过调整样本大小,从城市调查中随机删除某些样本,以使两者的比例与统计局提供的数字一致,但结果并没有太大改变。

¹¹ 这里“外出”的判断标准是常住人口的外出务工收入是否大于零。

户口的？”那些回答了具体年份的样本就是之前为农村居民后来获得城市户口的人。我们利用这条信息来识别永久移民。从表 1 可以看到，约 20% 的城市居民不是与生俱来的城市户口，其中又有一半以上是近 20 年才获得城市户口，这与改革开放以来户籍制度逐渐放松的事实是一致的。

表 1 利用城市调查数据识别的永久移民

获得城市 户口时间		获得城市户口途径(%)							总计
		升学	提干	参军	土地被征	购房	其他	缺失	
城市本地人	16 278	—	—	—	—	—	—	—	—
—1950	133	18.05	8.27	15.79	2.26	9.02	37.59	9.02	100
1951—1960	509	28.29	4.91	13.36	0.59	3.93	40.86	8.06	100
1961—1970	484	18.18	3.93	31.82	1.45	2.27	33.88	8.47	100
1971—1980	839	23.6	4.53	14.18	2.86	1.91	49.7	3.22	100
1981—1990	1 312	33.16	1.45	5.34	8.08	3.96	43.45	4.57	100
1991—	884	16.52	0.68	2.71	12.44	7.81	53.05	6.79	100
缺失	193	—	—	—	—	—	—	—	—
总计	20 632								

根据获得城市户口的不同方式，我们将这些移民划分为两组：通过正规途径（升学、提干或者参军）获得城市户口的和通过非正规途径（买房或者“失地”¹²）获得城市户口的。¹³可以看到，升学是人们获得城市户口最重要的渠道，参军是第二大渠道，只有一小部分永久移民采用最后两种非正式渠道。不同渠道在不同时期发挥的作用亦有不同。首先，升学在大多数时期都起着最显著的作用。其次，改革开放以后，提干和参军都不再是成为永久移民的主要渠道。再次，到 20 世纪 90 年代，越来越多的农村居民通过在城镇地区购房而获得城镇户口；同时也有越来越多的农村居民在土地被占用以后获得城市户口。

永久移民的特征与他们是何时、以何种方式获得城市户口密切相关，本文的附录表 A1 给出了具体的统计描述。这些统计结果提醒我们，即使在永久移民组内，移民也不是同质的。永久移民由不同的力量驱动、面临不同的制度约束，这很大程度上也反映了中国发展和转型的特点。因此，在接下来的分析中考察不同的组别是非常重要的。在实证分析的主体部分，我们依据样本获得城市户口的年份，将样本划分为不同的小组。为了反映转型过程中一些重要的转折点，我们没有按照本部分采用的年份来分组，而是以一些关键年份来作为分组的依据。

¹² 那些因土地被征用而获得城市户口的人不一定非得迁移，我们将他们归类为永久移民是由于他们户籍身份的转变。后文将再回到这个问题。

¹³ 应该注意的是，CHIP 数据中关于户口途径的信息很不充分。近一半的永久移民没有回答获得城市户口的确切方式（见表 1 中的其他或缺失），本文只考察了回答了这个问题的样本。

另外值得一提的是,永久性迁移的过程是有选择性的,那些有高学历、高收入、高能力或那些住在近郊的人更有可能被选中(或自行选择)来获取城市户口。这种选择既可能发生在作为迁移起始地的农村,也可能发生在作为迁移目的地的城市。人们可能首先选择没有城市户口也迁往城市,之后再成功地获得城市户口,因此,永久移民与临时移民之间的区别并非十分清晰。然而,这一缺陷并不是致命的。首先,很大比例的永久移民是通过升学或者参军获得城市户口的,在这些情况下,很多人在迁移之前就已经作出了决策。其次,直到近期,临时移民获得城市户口才变得相对容易。对大多数临时移民而言,如果不是通过特定的正规渠道(上文中提到的)迁到市区,他们必须保持农村户口身份。最后,本文的重点是观察永久移民和临时移民的自我选择效应,我们的结论并不依赖于他们何时何地获得城市户口。

(三) 农村居民、城市居民和移民

1. 个人特征

在接下来的分析中,我们将样本限定在18至60周岁并且已经不在校就读的个体。作这样的处理后,农村和城市样本的比例基本保持不变。为避免教育的内生性问题,我们删除了不到16岁就获得城市户口的样本。接下来我们比较四个样本组,前两个样本组(市区当地人和永久移民)有城市户口,后两个样本组(农村居民和农民工)¹⁴为农村户口。

表2a报告了男性的简单统计描述。城市本地人和农村居民两个群体在几乎每项特征上都有显著差异。城市当地人比农村居民受过更多的教育;与农村居民相比,他们平均年龄大、少数民族比例低、党员比例高;城市当地人中有工资收入的比例也远高于农村居民。从各项特征上看,永久移民与城市当地人更为相似,而农民工与农村居民的更为相似。这些都表明,永久移民具有较强的“正选择”效应,其选择性大于城镇农民工调查中的农民工样本。对女性而言,此结论同样成立(见表2b)。

很明显,劳动参与率在不同的样本组中差别很大,因而有必要着重研究有工资收入的样本。对于农村住户调查中有工资收入的样本,可进一步将其分为两组,即本地的工资获得者(local wage earners)和外出务工者(第二类移民)。表3a和表3b分别对男性和女性工资获得者作了描述统计。一般来说,工资获得者的教育水平比整个样本略高,女性尤为如此。而城镇居民与农村居民、城市当地人与永久移民之间的比较与前面对所有居民的讨论类似。由于城镇居民和城镇调查的农民工(第一类移民)的劳动参与率相对较高,挣工资者的描述统计与表2a和表2b非常相似。因此,我们下面主要比较农

¹⁴ 农村居民也包括一些临时移民(第二类移民),后面的分析中将分离出来。

村调查的农民工（第二类移民）和农村当地工人这两组样本。教育方面，第二类移民比农村当地工人有更高的受教育年限，但差距不大。如果把教育水平划分为四档，初中毕业的工人所占比重最大。但是，第二类移民中初中毕业生的比例远高于在农村本地务工者。对于年龄，第二类移民中男性和女性的平均年龄分别为32和26岁，比农村本地工人年轻近十岁；女性工人的平均年龄比男性工人年轻。至于党员的比例，当地工人比临时移民要高出许多，男女均是如此。

表2a 描述统计,男性(18—60岁)

	城镇居民				农村居民		农民工 (城镇农民工调查)	
	本地人		移民		均值	标准差	均值	标准差
	均值	标准差	均值	标准差				
受教育年限	11.14	2.99	11.67	3.46	7.76	2.41	8.32	2.62
教育水平 ^a								
小学及以下	0.03	—	0.05	—	0.25	—	0.20	—
初中	0.29	0.45	0.20	0.40	0.56	0.50	0.56	0.50
高中	0.39	0.49	0.31	0.46	0.18	0.39	0.21	0.41
大学及以上	0.29	0.45	0.44	0.50	0.01	0.12	0.03	0.16
年龄	42.02	10.09	44.02	8.83	38.14	12.01	35.29	8.49
少数民族	0.04	0.21	0.03	0.18	0.13	0.34	0.09	0.28
党员	0.29	0.45	0.53	0.50	0.13	0.33	0.05	0.22
有工资收入	0.81	0.39	0.87	0.33	0.54	0.50	0.95	0.22
家庭人口数量	3.19	0.78	3.14	0.76	4.40	1.34	2.77	0.95
观测数	5 114		1 291		12 328		1 977	

注：a.“高中”组包括了技校。

表2b 描述统计,女性(18—60岁)

	城镇居民				农村居民		农民工 (城镇农民工调查)	
	本地人		移民		均值	标准差	均值	标准差
	均值	标准差	均值	标准差				
受教育年限	10.65	3.05	9.90	3.57	6.48	2.89	7.33	2.88
教育水平 ^a								
小学及以下	0.06	—	0.13	—	0.45	—	0.34	—
初中	0.30	0.46	0.33	0.47	0.43	0.50	0.52	0.50
高中	0.42	0.49	0.35	0.48	0.10	0.31	0.13	0.33
大学及以上	0.22	0.41	0.19	0.40	0.01	0.09	0.02	0.12
年龄	41.54	9.70	42.38	9.44	37.56	11.63	33.80	8.30
少数民族	0.04	0.20	0.04	0.20	0.13	0.33	0.09	0.28
党员	0.18	0.39	0.18	0.39	0.02	0.15	0.01	0.10
有工资收入	0.63	0.48	0.62	0.49	0.22	0.41	0.76	0.43
家庭人口数量	3.15	0.80	3.22	0.90	4.46	1.34	2.85	0.89
观测数	5 481		1 211		11 372		1 892	

注：a.“高中”组包括了技校。

表 3a 挣工资者的描述统计,男性(18—60岁)

	城镇居民				农村居民					
	本地人		永久移民		农村本地 就业工人		农民工 II (农村调查)		农民工 I (城镇调查)	
	均值	标准差	均值	标准差	均值	标准差	均值	标准差	均值	标准差
受教育年限	11.34	2.96	11.95	3.37	8.00	2.40	8.07	2.23	8.29	2.62
教育水平 ^a										
小学及以下	0.02	—	0.04	—	0.21	—	0.18	—	0.20	—
初中	0.27	0.44	0.17	0.38	0.54	0.50	0.63	0.48	0.56	0.50
高中	0.39	0.49	0.31	0.46	0.23	0.42	0.18	0.38	0.21	0.41
大学及以上	0.32	0.47	0.47	0.50	0.02	0.14	0.01	0.11	0.03	0.16
年龄	41.56	9.24	43.07	8.48	40.91	10.43	31.63	9.88	35.27	8.40
少数民族	0.04	0.20	0.03	0.18	0.08	0.27	0.11	0.31	0.09	0.29
党员	0.30	0.46	0.54	0.50	0.20	0.40	0.06	0.23	0.05	0.21
工作天数/年	274	49	275	47	168	107	207	97	326	58
工作小时数/天	8.31	1.50	8.22	1.22	7.75	1.91	8.57	1.50	10.29	2.51
观测数	4 152		1 129		4 122		2 543		1 874	

注:a.“高中”组包括了技校。

表 3b 挣工资者的描述统计,女性(18—60岁)

	城镇居民				农村居民					
	本地人		永久移民		农村本地 就业工人		农民工 II (农村调查)		农民工 I (城镇调查)	
	均值	标准差	均值	标准差	均值	标准差	均值	标准差	均值	标准差
受教育年限	11.42	2.78	10.97	3.29	7.60	3.01	7.85	2.42	7.48	2.87
教育水平 ^a										
小学及以下	0.02	—	0.05	—	0.31	—	0.25	—	0.32	—
初中	0.21	0.41	0.26	0.44	0.45	0.50	0.60	0.49	0.52	0.50
高中	0.46	0.50	0.40	0.49	0.21	0.41	0.13	0.34	0.14	0.34
大学及以上	0.30	0.46	0.29	0.45	0.02	0.16	0.01	0.11	0.02	0.13
年龄	38.99	8.43	39.19	7.95	35.88	10.20	25.53	7.75	33.46	7.90
少数民族	0.04	0.20	0.05	0.22	0.04	0.19	0.09	0.28	0.08	0.27
党员	0.21	0.40	0.22	0.41	0.07	0.26	0.01	0.10	0.01	0.11
工作天数/年	271	52	276	51	189	111	237	91	327	60
工作小时数/天	8.06	1.30	8.23	1.63	7.85	1.78	8.47	1.43	10.34	2.41
观测数	3 455		748		1 378		1 072		1 432	

注:a.“高中”组包括了技校。

我们还计算了所有组别的年工作天数和日工作时间。城市居民,无论是当地居民还是永久移民都比农村当地工人和第二类移民的工作天数多,但是比第一类移民少,农村当地工人的工作天数最少。至于日工作时间,临时移民,尤其是第一类移民每日工作时间最长。

在本部分,需要提及的最后一点是,永久移民和临时移民几乎在每一方面都是截然不同的。先前那些只是将重点放在后者的研究,很有可能误导我们的理解。接下来讨论他们的收入(小时工资和人均收入)的差别。

2. 收入分配和收入差距

图1给出了农村本地工人、城市本地工人和三种类型移民的对数小时工资密度，(左上和中上分别为男性和女性)。很明显，城市工人比农村当地工人有更高的工资水平，三种类型移民的工资分布各不相同。第一类移民，虽然他们取样于与城市工人相同的地区，他们的工资分布与城市工人并不相同，而与农村工人非常接近。第二类移民的工资水平甚至比农村工人还低些。相反，永久移民的工资分布几乎与城市当地工人的工资分布重叠。所有这些表明，不同类型的移民间有明显的异质性。用人均年收入的分布来代替工资分布也可以得到类似的结果(见图1的右上部分)¹⁵即使在永久移民内部，也有着明显的异质性。正如图1中下面的两张图显示的那样，不同组群的永久移民的工资分布差异很大。更近期的移民不仅工资水平较低，工资分布也更分散。以上描述对男性和女性都适用。

我们还计算了不同群体人均收入的均值以及不平等的各种指标(见表4)。第一行报告了各个群组的平均人均收入。前两列为农村居民和城市居民，直接计算城乡收入比为3.01:1。城市居民由城市当地人和永久移民两部分组成，这两组的平均收入是8290元和7678元。第一类移民既没有被包含在农村居民中，也不在城市居民之列，他们的平均收入是6552元。下面几行是衡量不平等的各项指标。农村居民的不平等程度最高，基尼系数为0.3683。第一类移民的基尼系数位居第二，为0.3484。城市居民、城市当地人和永久移民的基尼系数很相似，也是最低的(低于0.33)。

如果没有迁移，城乡收入差距和收入分配将会是怎样的？一个简单但也天真(naïve)的尝试是将迁移的样本从城市居民的样本中分离出来，合并到农村居民中去，然后再重新直接计算各种不平等指标。结果显示在表4的第6、7、8列。由于第一类移民和永久移民的平均收入水平均高于农村居民，不难看到合并后农村的收入水平和不平等指标都上升了。

说这种尝试天真，是因为农村居民、城市居民和移民面临不同的技能价格。如果移民没有迁移，他们就不会有作为移民时的收入水平。因此，需要回答的一个基本问题是：如果移民按照农村地区或者农村居民的标准获得收入，收入分布将会是怎样的？我们将在第四节中回到表4。

¹⁵ 在2002年的城市调查中，每个有工作的人都被问到：“你在2002年的总收入是多少？”收入包括基本工资、奖金、津贴等各种形式，我们将家庭中所有收入加总，用它除以家庭总人数就得到人均收入。农村家庭的收入数据较为简单，他们被问到：在2002年他们以家庭为基础的纯收入是多少。用纯收入除以家庭规模就可得到人均收入。另外需要注意的是第一类移民的收入分布，其年收入密度更接近于城市居民和永久移民，而不是农村居民。这在一定程度上是因为第一类移民的工作天数更多，并且每日的工作时间更长。

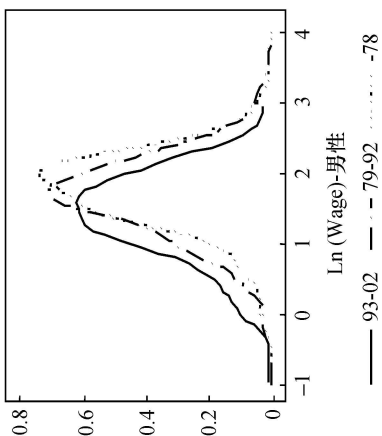
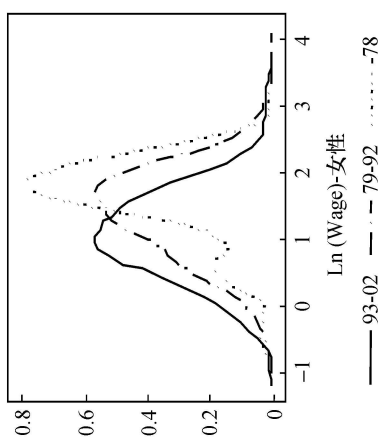
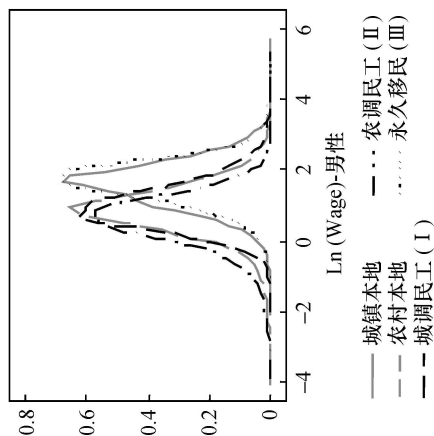
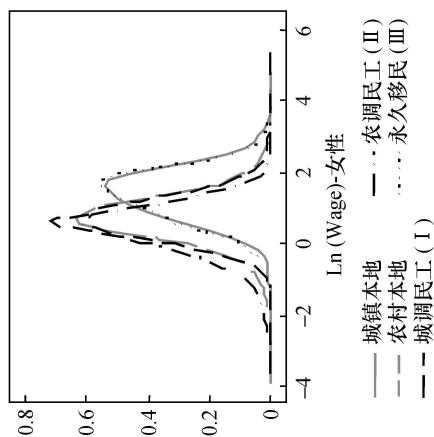
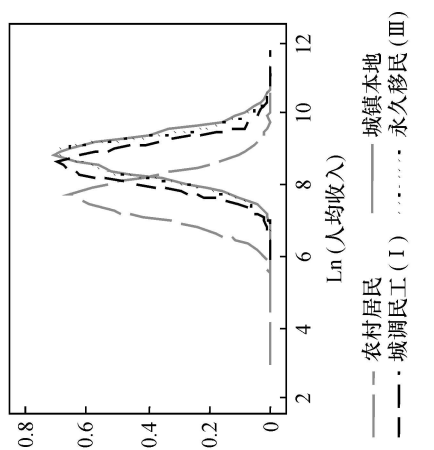


图1 工资(收入)分布

三、构造反事实收入分布：方法¹⁶

我们用 w 代表工资或收入, $f^i(w|x)$ 表示地区 i 在给定一系列特征 x 时的收入分布。 i 有两个可能的取值, Rural 和 Urban, 分别为农村本地居民和移民(农民工或永久移民)。 $f^{\text{Rural}}(w|x)$ 和 $f^{\text{Urban}}(w|x)$ 的差别反映了两个地区的技能价格的差异。一种了解这种差异的方法是估计各组的工资(收入)方程。从附录中的表 A2 可以看到, 城市居民(包括本地居民和永久移民)比农村居民和临时移民有更高的教育回报率, 经验、政治地位以及民族身份的回报率也不同。¹⁷ 接下来, 定义 $h(x|i=\text{Rural})$ 和 $h(x|i=\text{Urban})$ 分别代表农村居民和城市移民的特征分布。两者的差异反映了移民和农村留守者的特征分布差异。

农村地区工人的收入分布就可表示为

$$g(w | i = \text{Rural}) = \int f^{\text{Rural}}(w | x) h(x | i = \text{Rural}) dx.$$

同样, 城镇地区移民的收入分布可以表示为

$$g(w | i = \text{Urban}) = \int f^{\text{Urban}}(w | x) h(x | i = \text{Urban}) dx.$$

如果在城市工作的移民按照农村的技能价格获得收入, 则其收入分布为

$$g_{\text{Urban}}^{\text{Rural}}(w) = \int f^{\text{Rural}}(w | x) h(x | i = \text{Urban}) dx.$$

这个反事实收入分布与农村留守者的收入分布的区别在于特征分布的不同。由于该分布观测不到, 我们进一步将它改写为

$$\begin{aligned} g_{\text{Urban}}^{\text{Rural}}(w) &= \int f^{\text{Rural}}(w | x) h(x | i = \text{Urban}) \frac{h(x | i = \text{Rural})}{h(x | i = \text{Rural})} dx \\ &= \int \theta^M f^{\text{Rural}}(w | x) h(x | i = \text{Rural}) dx, \end{aligned}$$

其中 $\theta^M = \frac{h(x|i=\text{Urban})}{h(x|i=\text{Rural})}$ 。

DFL (1996) 指出, 通过引入一个可观测的收入样本(农村留守者的收入样本), 对它重新赋予权重(以反映移民和留守者在特征分布上的差异), 反事实的密度便可以估计出来。利用贝叶斯法则, 有

$$h(x) = \frac{h(x | i = \text{Urban}) \Pr(i = \text{Urban})}{\Pr(i = \text{Urban} | x)},$$

以及

$$h(x) = \frac{h(x | i = \text{Rural}) \Pr(i = \text{Rural})}{\Pr(i = \text{Rural} | x)}.$$

¹⁶ 下面的部分在很大程度上参考了 DFL(1996) 和 Chiquiar and Hanson(2005)。

¹⁷ 对条件分布的一种更完整的描述是利用分位回归, 但这对于 DFL 方法并不是不可或缺的。我们认为工资方程的 OLS 方程估计足以说明, 城市和农村地区的技能价格是不同的。

结合上述两个方程，可得到 θ^M 的表达式：

$$\theta^M = \frac{h(x | i = \text{Urban})}{h(x | i = \text{Rural})} = \frac{\Pr(i = \text{Rural})}{\Pr(i = \text{Urban})} \cdot \frac{\Pr(i = \text{Urban} | x)}{\Pr(i = \text{Rural} | x)}.$$

注意到式中的第一个比率 $\Pr(i = \text{Rural})/\Pr(i = \text{Urban})$ 由样本中移民和农村留守者的比率给定，第二个比率可以通过首先估计一个 Logit 或 Probit 模型，然后利用预测的概率计算，最终我们用 θ^M 的估计值作为权重进行核密度估计，从而得出反事实收入密度（本文后面的核估计均采用高斯核）：

$$\hat{g}(\omega) = \sum_{j=1}^n \frac{\hat{\theta}_j^M}{h} K\left(\frac{\omega - W_j}{h}\right).$$

利用农村本地工人的收入数据与 θ^M 的预测值，不仅可以计算出移民的反事实收入分布，还可以计算反事实收入不平等。假设永久移民和农村居民所占比例分别为 s 和 $(1-s)$ ，所有农村居民的 θ^M 之和为 1。¹⁸ 从而反事实的基尼系数通常可以利用农村居民的再加权样本估计得到。每个观测的 j 的权重是 $(1-s)/N + s \cdot \hat{\theta}_j^M$ ，其中 N 是农村留守居民的观测数量。

在给出实证结果之前，作两点说明。首先，收入 ω 可以采取多种形式。本文的第一个选择是对数小时工资。但是，并非每个人都有工资，而且那些有工资的人并不是从全体人口中随机抽取的。因此，在下面的部分中，我们也尝试用收入来规避使用工资数据造成的样本选择问题，在那些练习中， ω 代表人均收入的对数。为简便起见，我们暂时忽略劳动参与的问题。其次，对地区 i 的定义也需要进一步说明。这两个值（Rural 和 Urban）在不同的地方有不同的解释。当使用工资数据时，Rural 指的是农村本地工人，Urban 指的是不同类型的移民，如永久移民或两种类型的农民工。这意味着，当研究不同迁移样本的选择性时，农村本地工人总是被视为参照组。当使用人均收入数据时，Rural 指农村住户调查中的所有农村劳动力，包括外出打工者（第二类移民），因为他们也对其农村家庭收入有贡献。此时，Urban 只指永久移民或城镇调查农民工（第一类移民）。

四、构造反事实收入分布：实证结果

（一）基本的实证结果

首先对合并之后的农村本地工人样本和永久移民使用上述方法。为了构造永久移民的反事实收入密度，我们用这个合并的样本估计了一个 logit 模型 $\Pr(i = \text{Rural} | x)$ 。预测出的概率可以用来计算权重 $[1 - \Pr(i = \text{Rural} | x)]/$

¹⁸ 农村居民样本对应的 θ 不是一定总和为 1。如果不为 1，在直接使用 $(1-s)/N + s \cdot \hat{\theta}_j$ 作为权重之前首先应该作一些调整。当有不止两组时，可以直接进行扩展。

$\Pr(i=Rural|x)$ 。这个权重可以再用来构造永久移民的反事实工资密度。logit模型中使用的变量包括受教育年限、经验、经验的平方、少数民族虚拟变量、是否入党虚拟变量等。图2报告了由性别、是否为党员、年龄、教育这些变量预测的作为农村本地工人的概率。显然,教育发挥着非常重要的作用,对于具有初中及初中以下文化程度者,成为永久移民的可能性几乎为零,年轻

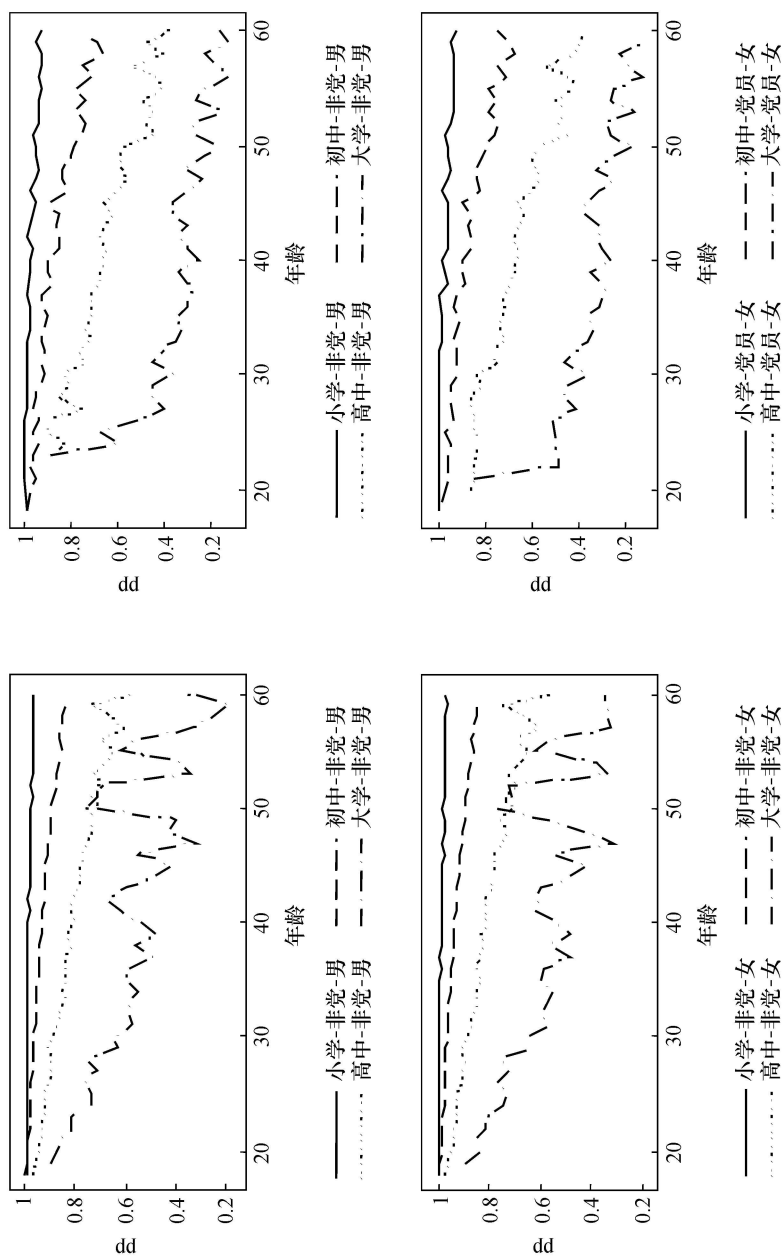


图2 成为农村居民的预测概率

人和男性尤为如此。大专（含）以上学历的人成为永久移民的概率最高，随着年龄的增加这个概率也在增加。此外，共产党员更有可能成为永久移民；在保持其他变量（如年龄、是否为党员、教育水平等）不变的情况下，女性成为永久移民的可能性更高。

本文最重要的实证研究结果见图3（上下两组图分别为男性和女性的情况）。左边和中间的图中黑色实线分别为永久移民（左）和农村本地工人（中）的工资核密度估计。同一性别的两组图中的虚线实际上是相同的，乃是永久移民的反事实核密度估计。可以清楚地看到，永久移民的反事实工资密度在它的实际密度曲线左方，这与农村地区的技能价格相对较低的事实相一致。然而，即使永久移民按照农村技能价格获得报酬，他们的反事实工资密度与农村居民的实际工资密度也相差很远，这意味着永久移民和农村居民存在较大的特征差异。在反事实工资密度达到峰值以前，它的分布密度均低于农村居民，而在高于其峰值的工资水平上，反事实分布的密度开始显著高于后者。为了便于比较，农村居民的工资密度与永久移民的反事实工资密度的差值显示在右侧的两张图上。结果显示，最有可能成为永久移民的那些人将拥有农村地区中等偏上的工资，而那些最不可能成为城市居民的人将拥有中等偏下的工资。如果城市中的男性永久移民回到农村地区，并且按照农村的技能价格获得工资，他们将更加集中地落在农村工资分布的上半部分。所有这些结果与“正选择”的假设是一致的。这一结论对女性也适用，并且农村女性居民的实际工资密度与女性永久移民的反事实密度间的差异似乎比男性更大。

所有这些结果都表明，永久移民永久地迁移到城市地区，有可能会降低农村地区的不平等，并有可能阻碍城乡之间收入差距的降低。

（二）稳健性检验

1. 按年龄组和迁移组

考虑到中国所经历的快速发展和转型，本文作了一些其他的尝试，其中的一些亦可被看做前面结果的稳健性检验。首先我们将男女的合并样本都按照年龄划分为三组：18—30岁、31—45岁和46—60岁，结果见图4。我们并没有给出农村居民的实际工资密度和永久移民的反事实工资密度，而是直接列出了两者的差值。在详述之前，我们可以先将总的结论给出：如果永久移民返回到农村，并按照农村的技能价格获得工资，他们的工资将会更为集中地处于农村收入分布的上半部分。有趣的是，选择效应的程度随年龄组的不同而不同。对男性而言（图4的左上图），随着年龄增加，永久移民的反事实密度越来越接近农村居民的实际工资密度。女性的这种趋势没有男性那样明显（右上图），这可能是由于女性样本较少的缘故。

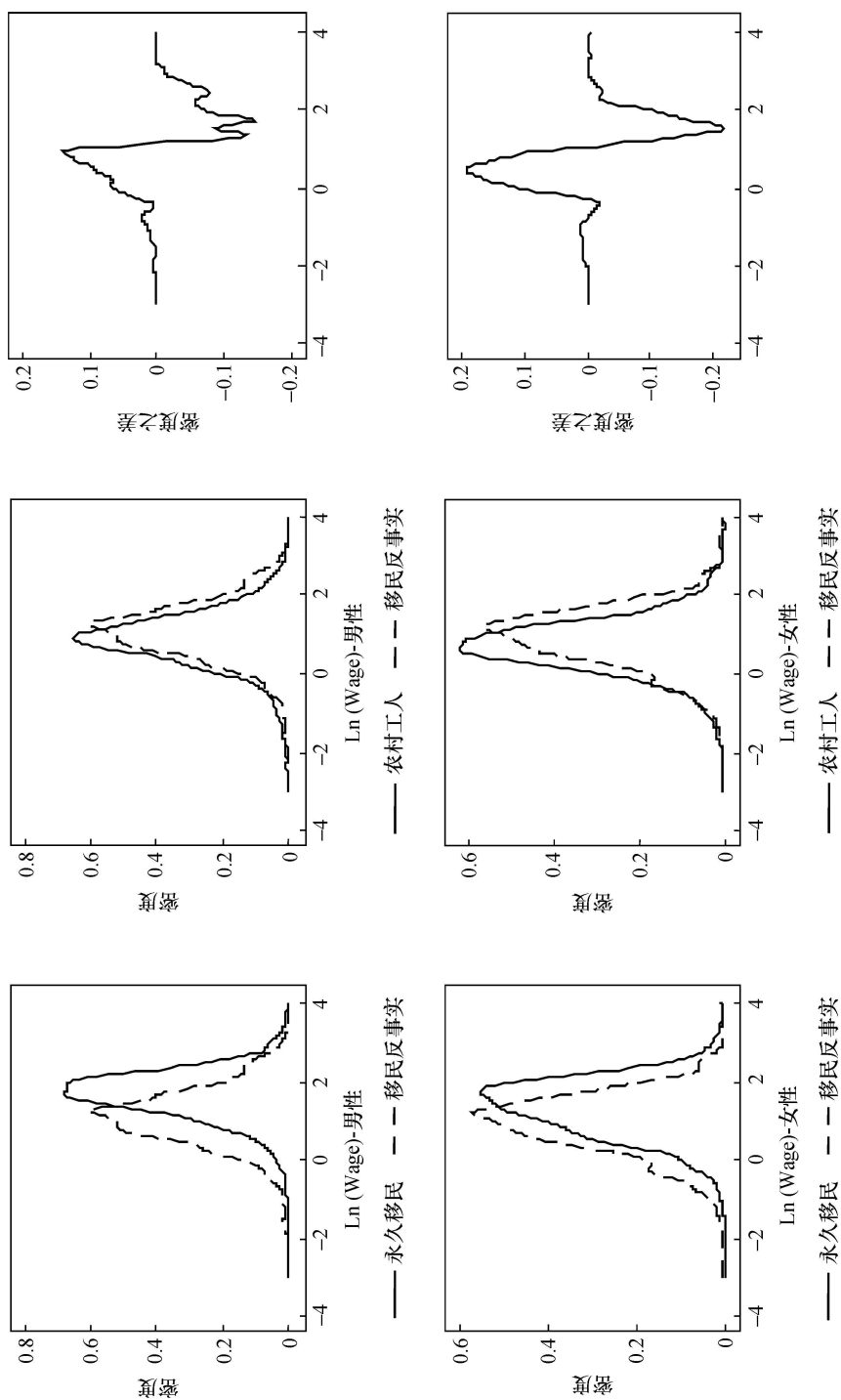


图3 永久移民的反事实收入分布

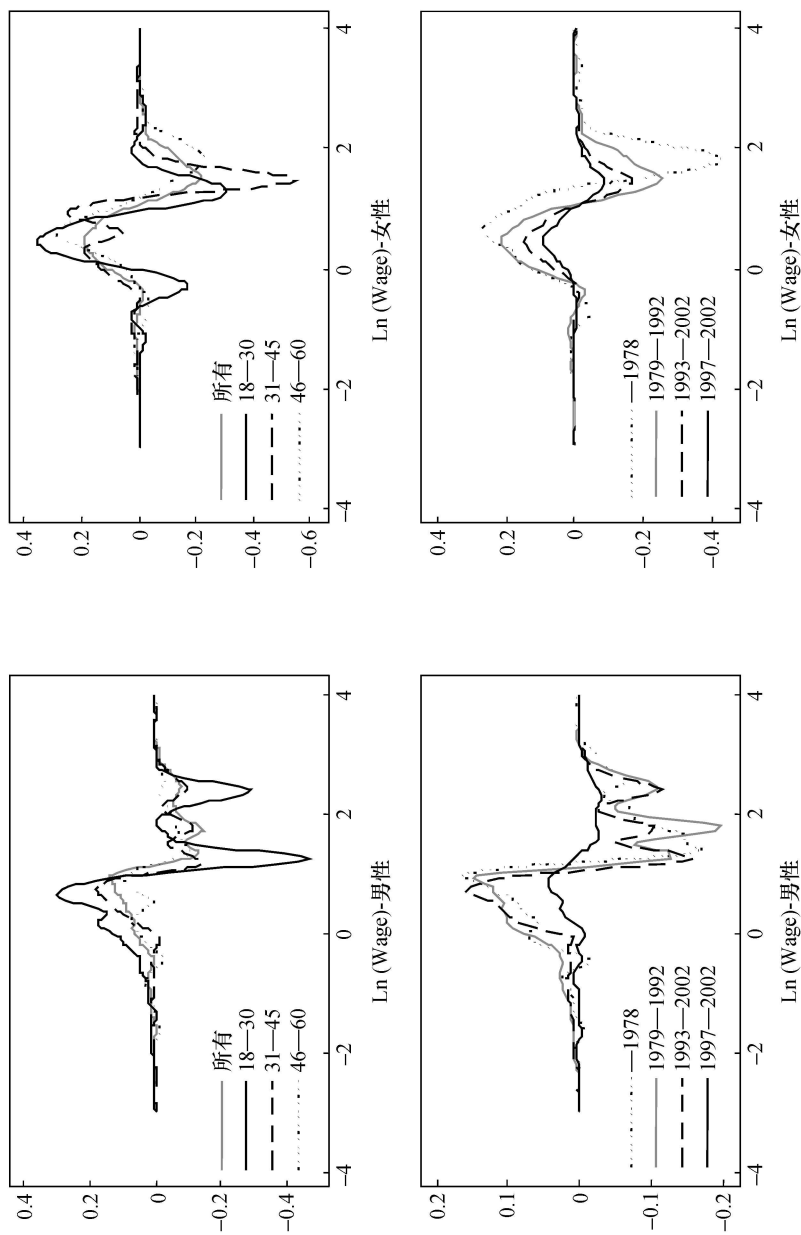


图4 农村事实分布与永久移民反事实分布的差值曲线
(按年龄分组和按移民年份分组)

年轻组别的结果更支持“正选择”的结论(至少男性)。这与如下事实是相符合的:永久移民样本中的年轻组别主要通过教育获得城市户口,而年长组除了包含那些很久以前通过教育获得城市户口的样本外,还包括通过其他方式(如买房和/或“失地”)而获得城市户口的样本。我们也可以将结果解释为,随着年龄增长,选择效应逐渐消失了。由于这两种解释有着不同的政策含义,我们做了以下工作,以确定哪一解释是最可能的。

正如前文提到的,永久迁移过程在不同时期受不同的体制因素影响。因此,接下来的一个工作是将样本划分为四个迁移组,即在1978年之前、1979—1992年间、1993—2002年间和1998—2002年间取得城市户口的永久移民,这样分组是因为1978年、1992年和1997年是中国转型进程中的三个关键年份。¹⁹可以清楚地看到,在工资分布的较高水平部分,永久移民的反事实密度依然高于农村留守者的实际工资密度,而在中间或者更低水平部分,却低于实际工资密度(见图4的下图),这说明“正选择”效应依然非常显著。然而,无论是男性还是女性,1979—1992年这一迁移组的正选择效应似乎更加显著,因为从图中看,对于大部分工资水平,零水平线与1979—1992年组的差值线之间的距离是所有分组中最大的(除了—1978组)。1993—2002年组与1998—2002年组的差值线位于1979—1992年组与零水平线之间。事实上,1998—2002年组的差值线与零线非常接近,说明农村居民的实际工资密度与反事实密度非常接近(尤其是男性)。

上述结果表明,随着中国劳动力市场改革的深化,特别是伴随着1998年的户籍改革,永久性迁移的过程变得越来越没有选择性。当然,这也可能是由于在近期,许多农村居民的土地被占用,他们因此而获得了城市户口有关。这种永久迁移过程既不是基于诸如教育水平等个人特征,也不是基于收入或财富水平等社会经济地位,而仅仅是由于这些人居住的地方离城市较近。

2. 考虑劳动参与差异

上述分析的主要缺点之一是忽视了城乡居民劳动参与率的差异,本节采用由Chiquiar and Hanson(2005)提出的方法来考虑城乡居民的劳动参与率相差很大的事实。我们仍然通过引入农村工人的观测密度,然后对它重新加权以反映城乡居民特征的差异以及劳动参与率的差异。需要修改权重,如下所示:

$$\theta = \frac{\Pr(i = \text{Rural}, D_i = 1)}{\Pr(i = \text{Urban}, D_i = 1)} \cdot \frac{\Pr(D_i = 1 \mid i = \text{Urban}, x)}{\Pr(D_i = 1 \mid i = \text{Rural}, x)} \cdot \frac{\Pr(i = \text{Urban} \mid x)}{\Pr(i = \text{Rural} \mid x)}$$

D_i 是一个虚拟变量,表示个人是否挣取工资(是=1/其他=0),右边第一项由样本固定确定,不会影响我们的分析,最后边的一项就是之前章节中所估计出的。中间一项是新出现的,

$$\theta^P = \Pr(D_i = 1 \mid i = \text{Urban}, x) / \Pr(D_i = 1 \mid i = \text{Rural}, x).$$

它被用来计算劳动参与的差异。²⁰分别用农村和城市的样本估计logit模

¹⁹ 见邢春冰(2008)。

²⁰ 更多细节可以参考Chiquiar and Hanson(2005)。

型，用估计出的概率构造 $\hat{\theta}^P$ ，然后用 $\hat{\theta} = \hat{\theta}^P \hat{\theta}^M$ 作为新的权重得出永久移民的反事实工资密度。

按年龄和迁移年份分组的结果显示在图5中，结论依然支持正选择的结论。将此结果与没有考虑劳动参与的结果相比较，差值曲线与零线的距离似乎更小。然而，模式很相似。对于男性的不同年龄组，18—30岁组的差值线离零线最远，46—60岁组最近。女性的结果与此类似，但不如男性清晰和明显。至于根据移民年份分组的结果，早期迁移组的结果更支持正选择的结论，迁移的年份越近，差值线离零线越近。这个结论对于男女均成立，尤其是男性。1998—2002年组男性的差值线与零线非常接近，这意味着如果回到农村并且按照农村居民的标准获得酬劳，这些移民的收入分布将和农村留守者非常相似。然而对于女性，即使是1998—2002年组，永久迁移过程似乎也有较强的选择性（见图5中下图）。

3. 教育的内生性和其他注意事项

在之前的分析中，我们删除了不到16岁就获得了城市户口的所有样本，目的是避免人们为了更好的教育机会而获得城市户口的情况。但是如前文所述，升学是农村居民获得城市户口的主要渠道。由于中国的高等教育资源几乎都集中在城镇地区，那些依靠升学获得城市户口的人几乎都是在城市完成最后的教育。因此上文“正选择”的发现可能只是简单地反映了教育是永久迁移的一种方式的事实。为了解决这个问题，我们删除了“升学”组再做验证。为了避免样本选择问题，进一步将分析限制在25—60岁的样本。男性和女性的结果分别显示在图5的右上图和右下图中。黑实线描绘了农村实际工资分布与25—60岁的永久移民的反事实工资密度的差异，差异的模式与先前章节的结果类似。较粗的灰色线描绘的是删除了升学组之后的永久移民的结果，“正选择”的结论仍是有效的，但程度较轻。男女均是如此。这个结果意味着即使不考虑升学这一获得城市户口的途径，永久迁移的过程仍具有正向选择性。

另一个尝试是同时删除“失地”和“升学”这两个组，结果由图中的虚线显示，几乎与灰线重合，这意味着我们的结果对“失地”组的删除并不敏感。

（三）临时移民的反事实工资密度

本小节考察一个稍有不同但很类似的问题，即如果临时移民没有迁移而是在本地挣取工资，他们的工资分布将会是怎样的？

男性和女性的基本结果都显示在图6中。为了便于比较，我们将三条差值曲线放在一起，每一条显示农村本地工人的实际工资密度与三组移民的反事实密度的差值。一个鲜明的特点是，两组农民工的差值曲线都比永久移民的更接近零线。永久移民与农民工的另一个区别是，当工资上升到整个分布的某一点处之前，两组农民工的差值曲线都在零线下方，在那一点之后，都在零线上方。这意味着，如果留在农村，这两组农民工将更集中地落入到农村工资分布的中

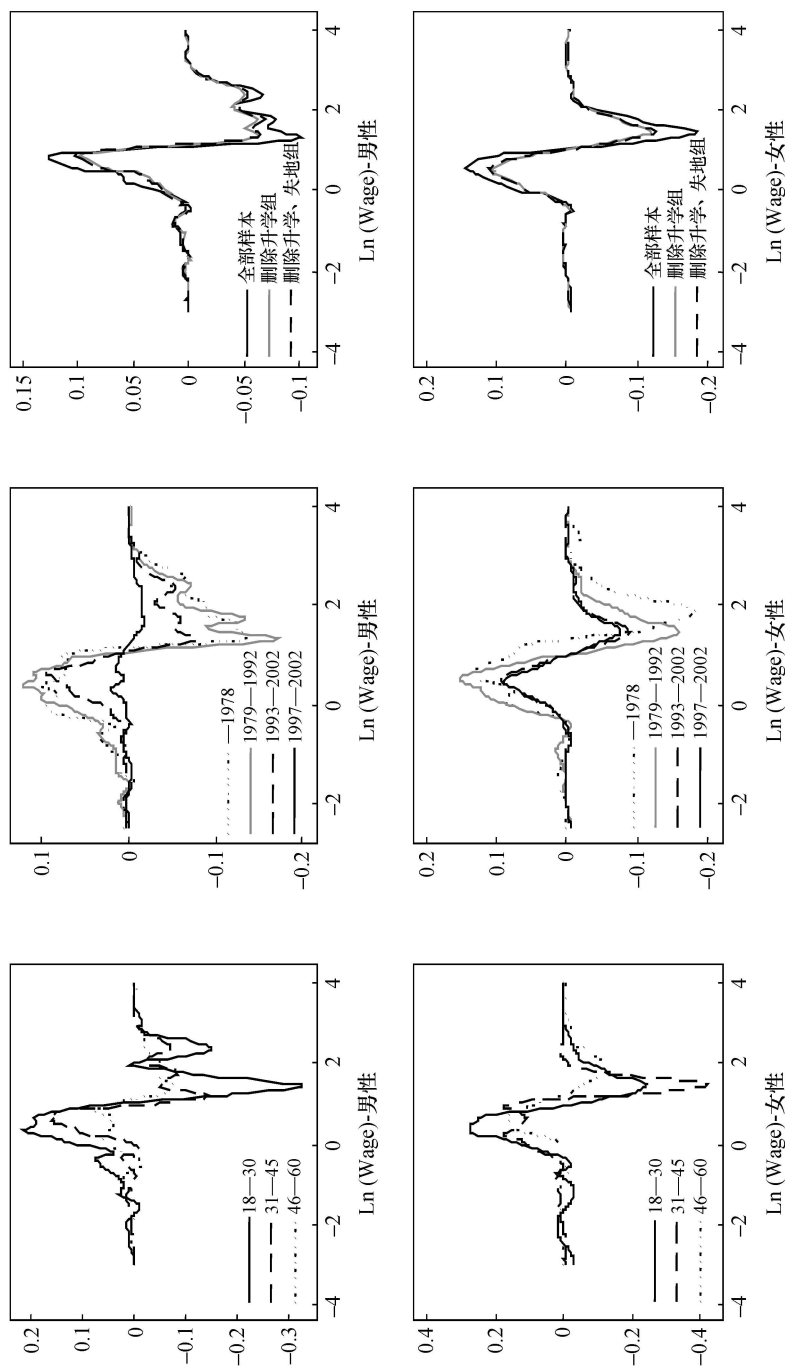


图5 稳健性检验

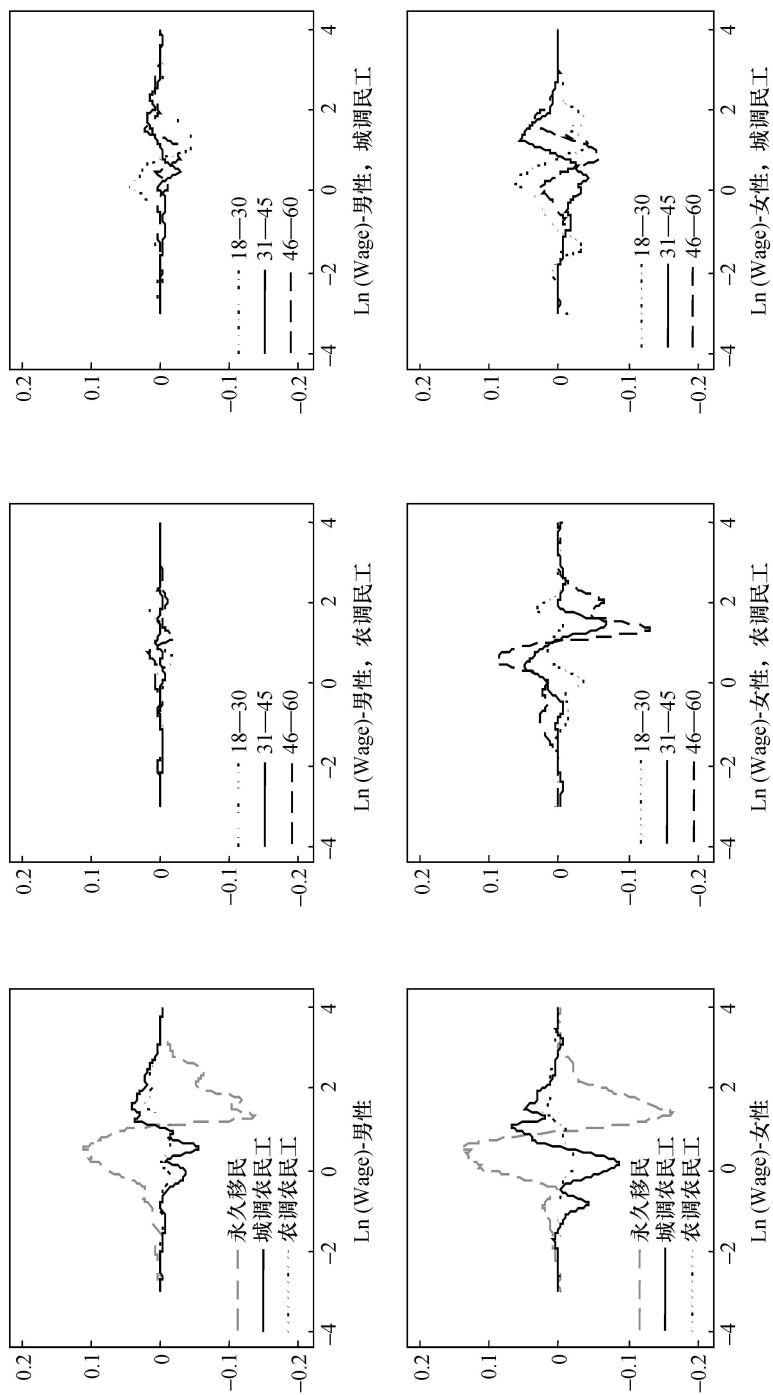


图6 农村工人的事实分布与农民工的反事实分布之差

下水平,而农村当地工人更集中地地位于分布的中部和上部。对比永久移民的“正选择”,这个结果更支持“负选择”。但如图6所示,农民工的负选择效应不像永久移民的正选择效应那么显著,尤其是对于男性和城镇调查的农民工(第一组移民)。同时,我们给出了两组农民工根据性别和年龄分组的差值曲线(中间和右边)。一致的结论是这些差值曲线与零水平线的距离非常接近。

应该牢记,本文使用的是农村本地工人作为参照组。由于参照组并不是从整个农村人口中随机选择出来的,临时移民的负效应并不一定就意味着临时移民是从整个农村人口中负向选择出来的。永久性迁移的选择效应也有同样的问题。然而,从描述部分可以清楚看到,参照组(农村本地工人)也具有一定的正向选择性,这会加强永久移民的“正选择”效应、同时减弱临时移民的“负选择”效应。

(四) 迁移对收入分配的影响

现在回到第3节提出的问题:如果没有迁移,城乡收入差距和农村的收入分配会是怎样的?为了克服使用工资数据的缺点,对于不平等的测量,在本节中我们使用人均收入数据。首先构造移民的反事实密度以考察他们的选择效应,参照组为农村住户调查中的劳动力。由于农村调查中的农民工(第二类移民)也为家庭收入作出贡献,很难将他们分离出去,因此他们也被包含在参照组样本中。图7分别列出了城镇调查农民工(第一类移民)和永久移民的反事实密度和差值曲线。尽管使用的测量收入的指标不同,结果却和前面的章节类似:永久迁移是正向的自我选择,而农民工却没有显著的自选择效应。

为考察选择效应对收入分布的影响,我们计算了反事实的收入水平和不平等指标。表4的第9列显示,如果永久移民回到农村,并且按照农村居民的标准获得报酬,农村地区的人均收入将会增加4%,从2715元上升到2814元。然而如果第一类移民回到农村,人均收入水平将会保持不变(第10列)。如果两种类型的移民都回到农村,收入水平将变为2798元(第11列)。由于在计算实际收入水平时,永久移民是被当做城市居民的,应将他们从城市样本中除去。而永久移民的人均收入(7678元)与城市本地人比相对较低,这就使得城市人均收入从8174元上升到8290元。因此,如果永久移民回到农村,城乡收入的比例将从3.01:1(8174:2715)下降到2.95:1(8290:2814);如果第一类移民返回到农村,这个比例将不变;如果两类移民都回到农村,这个比例将变为2.96:1。然而,绝对收入的反事实差距并没有缩小。

如果移民回到农村地区,不平等水平又将会如何呢?以基尼系数为例。如果永久移民回到农村,农村基尼系数将从0.3683上升到0.3755;如果第一类临时移民回到农村,基尼系数将保持不变;如果两者都回去,基尼系数将上升到0.3740。虽然反事实的不平等指标和实际指标相比变化不大,但改变的方向与不同类型移民的选择效应是一致的。

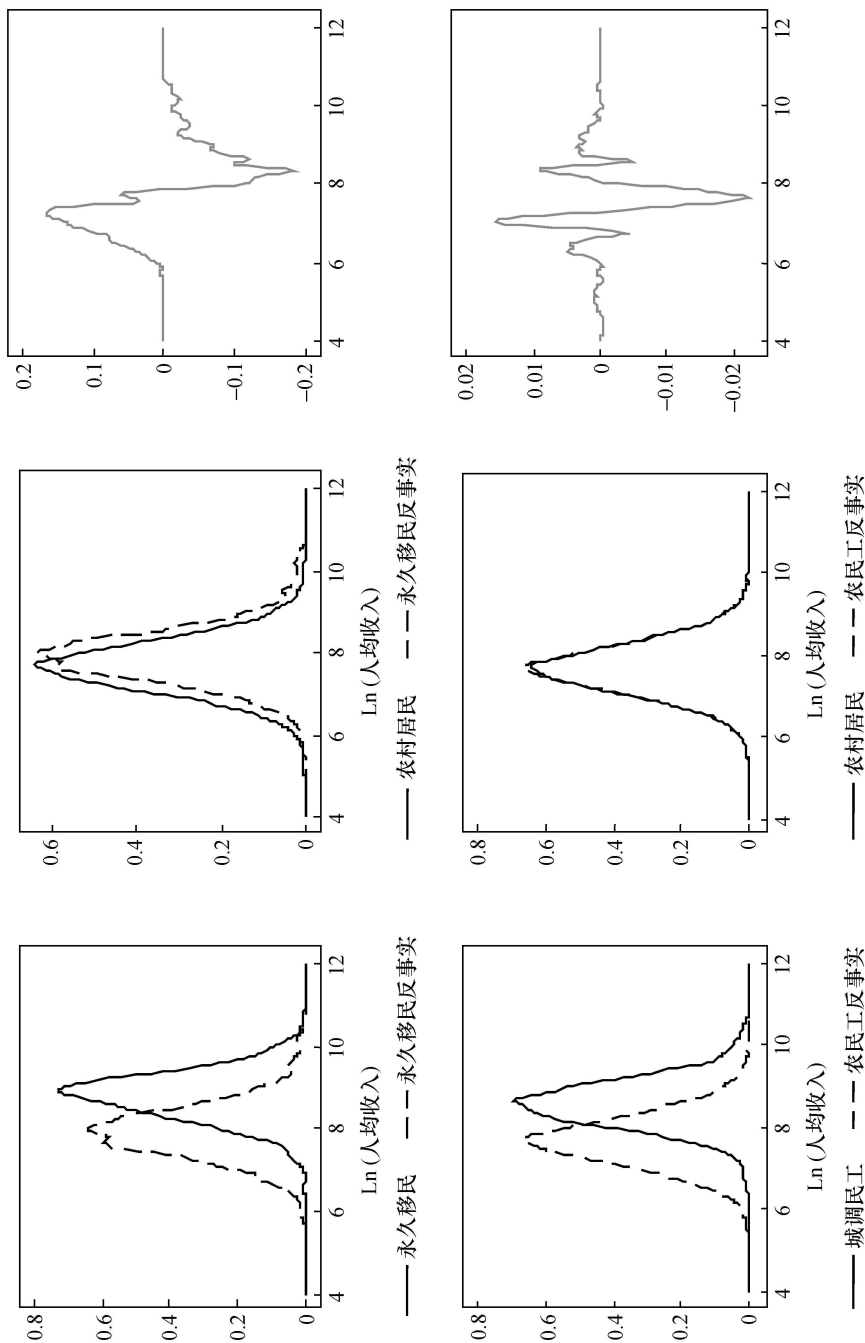


图7 利用人均收入计算事实分布与反事实分布
(上图：永久移民；下图：城镇调查的农民工)

五、结 论

由于大量农村居民为了更好的就业机会而离开农村, 迁移对城乡收入差距和城乡内部不平等正产生着深远的影响。这种影响的属性关键取决于谁在迁移以及他们的迁移模式。本文强调两个事实。首先, 农村居民并不是同质的, 他们自我选择迁移与否。第二, 那些成功改变了户籍身份的移民(永久移民)和没有获得城市户口的移民(临时移民)之间有显著的差异。采用 2002 年的 CHIP 数据集我们发现, 永久移民的正选择效应非常明显。同时, 由于永久移民在农村收入密度的上半部分占据更多的份额, 永久迁移使得城乡收入比上升。相反, 临时移民的选择效应几乎可以忽略不计, 因此它对农村收入水平和不平等都没有显著影响。

本文的政策含义是显而易见的。随着受过更多教育的农村居民离开, 农村地区面临着人才流失的问题; 移民的正选择性也同时意味着那些人力资本水平较低以及那些更需要帮助的人留在了农村。为了缩小城乡差距, 我们需要多维度的发展战略而不仅仅是促进迁移。为了提高那些留下的农民的收入水平和生活水准, 政府和相关组织机构应该为他们提供更多的公共服务。一方面, 要努力提高他们的人力资本水平, 进而提高他们的生产力。除了加强农村地区的教育以外, 还要进一步开展各种形式的培训。另一方面, 要进一步提高农村地区的社会保障水平和覆盖范围, 缩小农村居民在这方面与城镇居民的差别。

本文也存在一些不足之处。当把重点放在自我选择时, 我们没有强调迁移影响收入分布的其他几种重要渠道。当构造反事实收入密度时, 我们假设了收入的条件分布并不依赖于个人特征的分布, 这是一个很强的假设条件。在一般均衡框架下, 随着个人特征分布的改变, 它们的价格也随着变化, 本文忽略了这个影响。²¹ 汇款是迁移影响农村收入分布的另外一个重要途径, 许多研究者将重点放在其上(比如李实, 1999; Du *et al.*, 2005)。本文将它忽略是因为我们没有永久移民的汇款数据, 也因为不是本文的重点。另外一点不足在于, 我们只是考察了可观测的特征。然而, 如果可观测的技能与不可观察的技能间呈正相关并且足够强, 本文的结论对于不可观测的技能而言也同样适用。最后, 我们假定了教育水平在农村居民决定迁移之前已经确定。然而, 事实是更高的教育水平增加了迁移的可能性, 这意味着迁移可能对农村居民的人力资本投资决策有激励效应。如果那些有相对较高教育水平的人

²¹ 原因有两个。第一, 一般均衡效果很难考察, 许多研究都假设没有一般均衡效应(例如 Dinardo, *et al.*, 1996; Chiquiar and Hanson, 2005)。第二, 一般均衡效应可能不是一个严重的问题。直到近些年, 农民工的工资仍然很低, 这预示着在农村仍然有大量的剩余劳动力。由于有大量的过剩劳动力, 农村地区的技能价格改变得很小。这种情况在 2004 年左右开始转变, 一些沿海地区开始出现“民工荒”现象。人们才开始讨论刘易斯拐点是否到达。

没有完全成功迁出，迁移可能对农村的人力资本水平有积极的影响（如 Beine *et al.*，2001 和 Beine *et al.*，2006）。²² 尽管“智力外流”有可能被“智力获得（Brain-Gain）”抵消，在中国目前还没有这方面的研究。我们相信，“智力获得（Brain-Gain）”的问题是值得进一步研究的，但它不是本文的重点。

附 录

表 A1a 永久移民的描述统计

特征变量	获得城镇户口的途径					
	升学	提干	参军	失地	购房	其他
获得户口时的年龄	19.51 (5.52)	21.48 (6.98)	22.30 (7.05)	25.28 (12.07)	24.40 (13.47)	21.83 (11.17)
受教育年限	13.40 (2.78)	10.70 (3.37)	10.20 (3.50)	8.40 (3.29)	8.81 (3.88)	8.82 (3.73)
年龄	42.87 (13.06)	53.65 (12.09)	53.01 (11.16)	39.86 (13.87)	44.17 (17.78)	43.22 (15.73)
性别	0.39 (0.49)	0.40 (0.49)	0.05 (0.22)	0.61 (0.49)	0.59 (0.49)	0.61 (0.49)

注：通过升学获得城市户口的人教育水平最高，通过提干或参军获得城市户口的人教育水平相对较低些，通过非正规途径获得城市户口的人教育水平最低。那些通过升学、提干和参军而获得城市户口的人在迁移时相对比较年轻。通过失地或者买房而获得城市户口的人不仅年龄较大，并且年龄分布比较分散。2002 年（展开这个调查时），拥有最高平均年龄的组别是“提干”组和“参军”组（分别是 53.7 和 53.0），平均年龄最低的是“失地”组。这种反差反映了“提干”组和“参军”群体中的大多数都是在 20 世纪 80 年代以前获得的城市户口，而“失地”却是一个相对新的现象。括号中为标准差。

表 A1b 永久移民的描述统计

	获得城镇户口的时间					
	—1950	1950—60	1960—70	1970—80	1980—90	1990—
获得户口时的年龄	17.22 (8.53)	17.52 (7.09)	18.56 (7.56)	21.30 (7.68)	22.09 (10.32)	23.26 (13.05)
受教育年限	7.41 (4.46)	8.74 (4.26)	10.20 (3.78)	10.62 (3.62)	11.36 (3.68)	9.29 (4.01)
年龄	69.41 (8.54)	63.62 (7.46)	54.05 (7.85)	46.71 (7.97)	38.47 (10.61)	30.40 (13.23)
性别	0.35 (0.48)	0.44 (0.50)	0.31 (0.47)	0.40 (0.49)	0.51 (0.50)	0.62 (0.49)

注：移民受教育的平均年限从“—1950 组别”的 7.4 年稳步上升到“1980—90 组别”的 11.4 年，这种上升趋势也在一定程度上反映了教育扩张的事实。有趣的是，这一趋势在“1990—”这一组别停止了，这可能是由于永久移民的组成发生了改变。括号中为标准差。

表 A2a 男性工资方程(OLS)

	城镇居民		农村居民		
	本地人	永久移民	本地务工	农民工 II (农村调查)	农民工 I (城镇调查)
受教育年限	0.084*** (0.004)	0.059*** (0.007)	0.022*** (0.006)	0.023*** (0.008)	0.045*** (0.007)

²² “智力获得”可能还会通过移民返乡或者知识外溢等方式出现。

(续表)

	城镇居民		农村居民		
	本地人	永久移民	本地务工	农民工 II (农村调查)	农民工 I (城镇调查)
经验	0.030*** (0.004)	0.008 (0.010)	0.029*** (0.005)	0.048*** (0.005)	0.024*** (0.006)
经验平方/100	-0.032*** (0.008)	-0.013 (0.019)	-0.049*** (0.009)	-0.071*** (0.011)	-0.050*** (0.013)
少数民族	-0.058 (0.047)	0.014 (0.100)	0.011 (0.060)	-0.017 (0.059)	0.059 (0.057)
党员	0.152*** (0.022)	0.205*** (0.038)	0.166*** (0.034)	-0.052 (0.065)	0.045 (0.075)
移民年份					
1961—70		-0.476** (0.198)			
1971—80		-0.446** (0.202)			
1981—90		-0.528** (0.206)			
1990—		-0.637*** (0.212)			
常数项	0.541*** (0.074)	1.826*** (0.298)	0.620*** (0.117)	0.209 (0.161)	0.615*** (0.131)
R-squared	0.266	0.253	0.094	0.220	0.095
N	4 152	1 129	4 122	2 543	1 874

注:(1)控制了地区虚拟变量;(2)移民年份指获得城镇户口的时间;(3)括号中为标准误;(4)*、**、***分别代表在10%、5%和1%的水平上显著。

表 A2b 女性工资方程(OLS)

	城镇居民		农村居民		
	本地人	永久移民	本地务工	农民工 II (农村调查)	农民工 I (城镇调查)
受教育年限	0.109*** (0.005)	0.077*** (0.010)	0.025*** (0.008)	0.026** (0.011)	0.042*** (0.007)
经验	0.026*** (0.004)	0.029** (0.012)	0.013** (0.006)	0.007 (0.009)	0.021*** (0.006)
经验平方/100	-0.020* (0.011)	-0.067*** (0.025)	-0.015 (0.013)	0.009 (0.022)	-0.051*** (0.013)
少数民族	-0.012 (0.057)	0.008 (0.117)	-0.266** (0.134)	-0.125 (0.097)	-0.018 (0.060)
党员	0.135*** (0.029)	0.131** (0.060)	0.319*** (0.079)	0.624*** (0.223)	0.201 (0.138)
移民年份					
1961—70		0.080 (0.169)			
1971—80		-0.072 (0.169)			
1981—90		-0.228 (0.177)			

(续表)

	城镇居民		农村居民		
	本地人	永久移民	本地务工	农民工 II (农村调查)	农民工 I (城镇调查)
1990—	0.159*	0.817**	0.478***	0.095	0.409***
	(0.091)	(0.347)	(0.148)	(0.278)	(0.124)
常数项	0.295	0.253	0.131	0.268	0.126
	3455	748	1378	1072	1432

注:(1)控制了地区虚拟变量;(2)移民年份指获得城镇户口的时间;(3)括号中为标准误;
(4)*、**、***分别代表在10%、5%和1%的水平上显著。

参 考 文 献

- [1] Autor, D., L. Katz, and M. Kearney, "Trends in U. S. Wage Inequality: Re-Assessing the Revisionists", NBER Working Paper No. 11627, 2005.
- [2] Beine, M., F. Docquier, and H. Rapoport, "Brain Drain and Economic growth: Theory and Evidence", *Journal of Development Economics*, 2001, 64(1), 275—289.
- [3] Beine, M., F. Docquier, and H. Rapoport, "Brain Drain and Human Capital Formation in developing countries: Winners and Losers", Manuscript, IZA, Bonn, 2006.
- [4] Borjas, G., "Self-Selection and the Earnings of Immigrants", *American Economic Review*, 1987, 77(4), 531—553.
- [5] Borjas, G., "The Economic Analysis of Immigration", in Ashenfelter, O. and D. Card (eds.), *Handbook of Labor Economics*. Amsterdam: North-Holland, 1999, 1697—1760.
- [6] Chiquiar, D. and G. Hanson, "International Migration, Self-Selection, and the Distribution of Wages: Evidence from Mexico and the United States", *Journal of Political Economy*, 2005, 113(2), 239—281.
- [7] Cai, F., and D. Wang, "Impacts of Internal Migration on Economic Growth and Urban Development in China", Working Paper, Institute of Population and Labor Economics, CASS, 2007.
- [8] Deng, Q., and B. Gustafsson, "China's Lesser Known Migrants", IZA Discussion Paper No. 2152, 2006, <http://ssrn.com/abstract=908236>.
- [9] DiNardo, J., N. Fortin, and T. Lemieux, (1996), "Labor Market Institutions and the Distribution of Wages, 1973—1992: a Semi-parametric Approach", *Econometrica*, 64(5), 1001—1044.
- [10] Du, Yang, A. Park, and S. Wang, "Migration and Rural Poverty in China", *Journal of Comparative Economics*, 2005, 33(4), 688—709.
- [11] Lemieux, T., "Increasing Residual Wage Inequality: Composition Effects, Noisy Data, or Rising Demand for Skill?" *American Economic Review*, 2006, 96(3), 461—498.
- [12] 李实, "中国农村劳动力流动与收入分配", 《中国社会科学》, 1999年第4期, 第16—33页。
- [13] 李实、岳希明, "中国城乡收入差距调查", 《财经》, 2004年第3/4期合刊, 第30—38页。
- [14] Machado, J., and J. Mata, "Counterfactual Decomposition of Changes in Wage Distributions Using Quantile Regression", *Journal of Applied Econometrics*, 2005, 20(4), 445—465.
- [15] National Bureau of Statistics, *China Statistical Yearbook 2005*. Beijing: China Statistical Press, 2005.

- [16] Oaxaca, R., "Male-Female Wage Differentials in Urban Labor Markets", *International Economic Review*, 1973, 14(3), 693—709.
- [17] Park, A., D. Wang, and F. Cai, "Migration and Urban Poverty and Inequality in China", Working Paper, Institute of Population and Labor Economics, CASS, 2006.
- [18] Wang, D., and F. Cai, "Migration and Poverty Alleviation in China", Working Paper, Institute of Population and Labour Economics, CASS, 2006.
- [19] 邢春冰, "经济转型与不同所有制部门的工资分布, 从下海到下岗", 《管理世界》, 2007 年第 6 期, 第 23—37 页。
- [20] Zhao, Z., "Migration, Labour Market Flexibility, and Wage Determination in China: A Review", Labour and Demography series 0507009, EconWPA, 2005.

Migration, Self-Selection, and Income Distribution: Evidence from Rural and Urban China

CHUNBING XING

(Beijing Normal University)

Abstract In this paper, we emphasize two facts. First, rural residents are not homogeneous, they self-select to migrate or stay home. Second, there are significant differences between migrants who have successfully obtained urban hukou (permanent migrants) and those who have not (temporary migrants). Using the 2002 CHIP data, we find that permanent migrants are positively selected from rural population. As permanent migration occurs more in the upper half of rural income distribution, both average rural income and inequality decrease, and the urban-rural income gap increases. On the contrary, the selection effect of temporary migrants is almost negligible.

Key Words Migration, Self-Selection, Income Distribution

JEL Classification C14, O15, P25