



西南财经大学

SOUTHWESTERN UNIVERSITY OF FINANCE AND ECONOMICS

硕士学位论文

MASTER'S DISSERTATION

户籍制度对我国流动人口工资收入的影响

The Influence of Household Registration System on the

Salary Income of Floating Population in China

学位申请人 _____ 苏英洁
指导教 师 _____ 吴开超
学科专业 _____ 西方经济学
学位类别 _____ 经济学

户籍制度对我国流动人口工资收入的影响

The Influence of Household Registration System on the
Salary Income of Floating Population in China

学位申请人： 苏英洁

学 号： 214020104009

学科专业： 西方经济学

研究方向： 家庭金融与资产配置

指导教师： 吴开超

定稿时间： 2017 年 3 月 15 日

西南财经大学

学位论文原创性及知识产权声明

本人郑重声明：所呈交的学位论文，是本人在导师的指导下独立进行研究工作所取得的成果。除文中已经注明引用的内容外，本论文不含任何其他个人或集体已经发表或撰写过的作品成果。对本文的研究做出重要贡献的个人和集体，均已在文中以明确方式标明，因本学位论文引起的法律结果完全由本人承担。

本人同意在校攻读学位期间论文工作的知识产权单位属西南财经大学。本人完全了解西南财经大学有关保留、使用学位论文的规定，即学校有权保留并向国家有关部门或机构送交论文的复印件和电子版，允许论文被查阅和借阅。本人授权西南财经大学可以将本学位论文的全部或部分内容编入有关数据库进行检索，可以采用影印、缩印、数字化或其他复制手段保存和汇编本学位论文。

本学位论文属于

1、☐ 保密，在_____年解密后适用本授权书。

2、☐ 不保密

特此声明。

学位申请人：苏英洁

2017 年 3 月 15 日

中文摘要

户籍长期以来被认为是影响工资性收入的一个重要因素，随着我国市场经济体系的不断完善，劳动力市场越来越趋向一体化。一体化要求劳动力、资本等生产要素可以自由流动以达到合理配置，户籍制度却人为地限制了城市与农村间人口的正常迁移，严重妨碍了人口的合理集聚，阻碍了中国城镇化进程。很多学者都批评户籍制度导致中国城乡二元劳动力市场的形成，认为相同条件下，外来农村户籍劳动力在就业机会、就业待遇以及就业保障方面均受到一定的限制。我国的户籍制度改革也在大刀阔斧推进，相关文件已经出台，“统一城乡户口登记制度”被提出，多个省份已按照当地实施意见划定新型户籍制度成型期限。

在目前依然拥有独特的城乡分割二元户籍制度的我国，已有不少学者研究出，户籍制度对外来务工人口的工资性收入产生了一定影响。本文致力于研究在证明了有影响的前提下，如何确定是户籍制度本身产生的，而非城市、农村成长环境不同造成的禀赋差异。进一步讨论如果没有户籍制度的限制，对于外来务工农业人口在城市工作的收入状况将会有多大程度的改善。并且细分到体制内与体制外单位，关注收入受户籍制度的影响是否会存在显著不同。

梳理前人研究发现，在对象选取上，已有的文献大都直接比较城市人口与农村人口间的收入差距，但这不可避免地忽略了成长氛围、文化观念、医疗条件、教育质量等一系列不可观测的因素带来的影响，将这些不可观测因素导致的差异归结于户籍歧视会产生估计偏误，本文则把研究对象锁定在城市外来人口——既包括通过农转非实现迁徙的“城市移民”，又包括以农业户口在城市从事非农工作的“流动人口”，最大范围控制成长环境；在方法选取上，大多数文献都是基于 Blinder 和 Oaxaca 的工资差异分解模型，但其中存在着“指数基准”问题，导致很多文献得出的回报结果差别甚远。本文则选取了倾向得分匹配法(PSM)来克服样本的选择性偏误，将总体外来人口分为两个样本组：处理组——具有城市户籍的外来人口，控制组——没有城市户籍的外来人口，两组进行匹配分析，估计户籍制度的平均处理效应(ATT)。

本文采用 CGSS2012 年全国微观调查数据，进行户籍制度收入效应的对比研究。样本显示，流动人口和城市移民工资性收入差异较大，为了证明户籍制度确实是形成差异的来源之一，本文首先用 Logit 模型估计出倾向性得分 $pscore$ ，即获得城市户籍的概率，作为不可观测的“能力”的代理变量，代入多元线性回归

中，户籍制度对工资性收入的影响是显著的。并且，加入“能力”代理变量后，户籍效应的影响略为降低，这便说明户籍制度本身便对流动人口有负向收入效应，而不是完全通过能力来反映。

为得到准确的户籍制度对工资性收入的平均处理效应，本文接下来使用 PSM 的三种匹配方法：最近邻匹配法、半径匹配法和核匹配法来进行估计。模型通过了匹配检验，运用每种匹配方法时都采用抽样法(Bootstrap)反复抽样 500 次来获取结果，以得到标准误下的 ATT。结果显示，无论使用哪一种方法，经匹配后的 ATT 均在 5%和 10%的水平下显著。以最近邻匹配法为例，该匹配方法得出的结论是，户籍制度的存在对于流动人口的收入有显著的负向影响，它使得流动人口的小时收入降低了近 17.8%。

进一步对外来人口样本按照体制内、体制外单位分类研究，发现在不同市场化程度的部门，ATT 表现出差异性。以最近邻匹配法为例，体制内的城市移民比流动人口小时工资收入仅高 21.7%，体制外的城市移民比流动人口小时工资收入仅高 12.7%，说明户籍制度的收入效应在市场化低的部门更高，并随市场化进程减弱。那么，若是消除户籍制度的影响，受益更多的是体制内单位的流动人口，对于体制外单位流动人口的收入提高不大。

针对实证回归的结果，本文提出了相关政策上的建议。首先，户籍制度的改革是有效的，改革的力度和进程应该依不同的城市和地区而有所差别。其次，针对户籍歧视在体制内、体制外单位的差别，应加大对国有、集体企业的户籍歧视管控，保障部门之间的自由竞争。此外，平衡城市和农村之间教育、医疗服务、福利、公共服务等资源分配的不均衡，解决历史遗留问题，是漫长而终将奏效的一条道路。

诚然，本文还存在研究的不足之处，一是样本量缺失问题，选取的 CGSS2012 年数据有一定缺失值，结论不可避免的可能存在一定偏差，较少的样本量也不利于更深层次的研究。二是仅仅使用了“农业”与“非农”这两个维度对样本进行划分，没有考虑户口所在地、交互作用等的影响。并且本文使用的是截面数据，没有纳入时间因素带来的波动性。针对以上问题，本文将在后续研究中继续改进。

【关键字】 户籍制度；收入；城乡差异；倾向得分匹配法；代理变量

Abstract

Household registration has long been regarded as an important factor affecting wage income. With the continuous improvement of China's market economy, the labor market is becoming more and more integrated. Integration of labor, capital and other factors of production can be free to flow to achieve a reasonable allocation, the household registration system has artificially restricted the normal migration of the population between cities and rural areas, seriously hampered the rational population of the population, hindering the process of urbanization in China. Many scholars have criticized the household registration system led to the formation of China's urban and rural dual labor market, that under the same conditions, foreign rural household labor force in employment opportunities, employment treatment and employment security are subject to certain restrictions. China's household registration system reform is also drastic, the relevant documents have been introduced, "unified urban and rural household registration system" was put forward, a number of provinces have been in accordance with the local implementation of the provisions of the new household registration system to create the deadline.

At present, there is still a unique urban and rural divide dual household registration system in China, whether the household registration system has an impact on the wage income of the migrant agricultural population? If there is an impact, how to determine the household registration system itself is generated, rather than due to urban and rural growth environment caused by different income differences? If there is no household registration system restrictions, for the migrant workers in the urban population of the income situation will be how much improvement? Is there a significant difference in the impact of household and institutional work on household registration systems? This article is devoted to these topics.

Combing predecessors study found that the object selection, the existing literature mostly directly compare the urban population and rural population income gap, but this inevitably ignores the growth atmosphere, cultural concepts, medical conditions, the quality of education and a series of The impact of these factors is attributed to the discrepancy caused by these unacceptable factors. This paper places the study object in the urban foreign population - both the "urban migration" The majority of

the literatures are based on the wage differential decomposition model of Blinder and Oaxaca, but there is a "floating population" in which the agricultural population is engaged in non-agricultural work in the city, and the maximum range controls the growth environment. Index basis "problem, resulting in a lot of literature, the results of the return is very different. In this paper, the preference score matching method (PSM) is selected to overcome the selective bias of the sample. The total population is divided into two groups: the treatment group - the foreign population with the household registration, the control group - no household registration Foreign population, the two groups to match the analysis, the estimated household registration system average treatment effect (ATT).

In this paper, the use of CGSS2012 years of the whole map of the data, the household income system income effect of comparative study. In order to prove that the household registration system is indeed one of the sources of differences, this paper first uses the Logit model to estimate the probability score pscore, that is, the probability of obtaining the urban household registration, as the unobservable Of the "ability" of the proxy variable, into the multiple linear regression, the household registration system on the impact of wage income is significant. Moreover, after adding the "capacity" proxy variable, the impact of the household registration effect is slightly reduced, which means that the household registration system itself has a negative income effect on the floating population, rather than through the ability to reflect.

In order to obtain the average effect of the household registration system on wage income, this paper uses the three matching methods of PSM: nearest neighbor matching method, radius matching method and kernel matching method. The model passed the matching test, and each sampling method was used to sample the bootstrap 500 times to obtain the standard ATT. The results showed that, regardless of which method was used, the matched ATT was significant at levels of 5% and 10%. In the case of the nearest neighbor matching method, the matching method concludes that the existence of the household registration system has a significant negative impact on the income of the floating population, which reduces the hourly income of the floating population by 17.8%.

Further studies on the classification of foreign population in accordance with the

system, the system outside the system, found in different degree of marketization of the sector, ATT showed differences. In the case of Nearest Neighborhood Matching, for example, the urban migrants in the system are only 21.7% higher than the migrant population's hourly wage income, and the urban migrants are only 12.7% higher than the migrant population's hourly wage income, indicating that the income effect of the household registration system is low Of the sector higher, and with the market process weakened. So, if the elimination of the impact of household registration system, benefit more is the unit within the system of floating population, for the system outside the unit population of the income increase is not big.

In view of the results of empirical regression, this paper puts forward some relevant policy suggestions. First of all, the reform of the household registration system is effective, the intensity of the reform and the process should be different according to different cities and regions. Second, the discrimination against the household registration in the system, the system outside the unit differences, should increase the state-owned, collective enterprise household registration discrimination, protection between the free competition. In addition, the balance between urban and rural education, medical services, welfare, public services and other resource allocation is not balanced, to solve the historical problems, is a long and end will work a path.

In addition, there are some shortcomings in this paper, one is the lack of sample size, selected CGSS2012 years of data have a balance of value, the conclusion may be a certain deviation, less sample size is not conducive to a deeper level of research. Second, only the use of "agriculture" and "non-agricultural" these two dimensions of the division of the sample, did not consider the account location, interaction and so on. And this article uses cross-section data, without the inclusion of time factors to bring volatility. In view of the above problems, this article will continue to improve in the follow-up study.

【KEY WORDS】 Household Registration System ; Income ; Differences between urban and rural areas ; Propensity Score Matching ; Proxy Variable

目录

1. 引言.....	11
1.1 研究背景	11
1.2 问题的提出	13
1.3 研究意义.....	14
2. 文献综述.....	15
2.1 工资收入的影响因素	15
2.1.1 先赋性人力资本因素	15
2.1.2 后赋性人力资本因素	16
2.1.3 非人力资本对工资的影响	16
2.2 户籍制度对工资收入的影响	17
2.2.1 城镇职工与农民工的收入差异	17
2.2.2 户籍制度影响工资收入	17
2.2.3 户籍制度影响就业机会	18
2.2.4 户籍制度影响社会保障	19
2.3 户籍制度工资效应的研究争议	19
2.3.1 针对外来人口与本地市民收入差异分解的争议	19
2.3.2 支持收入差异来自于户籍歧视的研究	19
2.3.3 支持收入差异来自于禀赋的研究	20
2.3.4 关于体制内、外部门与户籍制度关系的研究	20
2.4 小结	21
2.4.1 现有文献研究成果与存在的问题	21
2.4.2 本文的研究视角与创新	22
3. 理论基础与研究方法.....	23
3.1 相关理论.....	23
3.1.1 人力资本理论	23
3.1.2 二元劳动力市场理论	23
3.2 研究方法.....	24
3.2.1 代理变量方法	24
3.2.2 倾向得分匹配法 (Propensity Score Matching , PSM)	25

3.2.3 自抽样法(Bootstrap)	26
4. 模型设定、数据与变量.....	27
4.1 模型设定	27
4.1.1 代理变量的估计式	27
4.1.2 倾向得分的估计式	27
4.1.3 平均处理效应(ATT)估计式	28
4.1.4 模型假设与实现步骤	28
4.2 数据来源	29
4.3 指标选取与变量说明	30
4.3.1 因变量	30
4.3.2 自变量	31
4.4 描述性统计	32
4.4.1 整体与分样本描述性统计	33
4.4.2 按体制内外分组的描述性统计	33
5. 实证结果与分析.....	35
5.1 代理变量回归分析	35
5.1.1 获得代理变量的 Logit 回归结果	35
5.1.2 多元线性回归结果	36
5.2 倾向得分匹配回归分析	37
5.2.1 匹配检验	37
5.2.2 倾向得分匹配估计结果	40
5.3 体制内、外部门分组回归分析	41
5.4 小结	43
6. 结论、政策建议及本文的不足.....	43
6.1 主要结论.....	43
6.2 政策建议	46
6.3 本文的进步	47
参考文献.....	49
致 谢.....	52

1.引言

1.1 研究背景

2016年9月19日，北京市正式出台了《关于进一步推进户籍制度改革的实施意见》，提出取消农业户口和非农业户口的性质区分，建立起“统一城乡户口登记制度”——即原“农业”和“非农”户口均统一登记为居民户口，并建立与“统一城乡户口登记制度”相适应的教育、就业、卫生、住房、社保、土地和人口统计制度。户籍制度改革又一次被推向了风口浪尖，据不完全统计，截至目前，包括北京在内，已有上海、天津、广东、浙江、陕西、四川等30个省份出台户籍制度改革方案，各地普遍提出取消农业户口与非农业户口之分。那么，我国独有的城乡分割的二元户籍制度是如何演变和发展的？在高度市场化的今天，取消这一分割的呼声愈烈，是否真的是因为其阻碍了城乡间劳动力自由流动，导致城乡差距的扩大与阶层分化加剧？

户籍制度的制定离不开经济大环境，其用途在于管理住户和人口。1950年代中后期，新中国开始着手建立一个全新的户籍制度，用以加强对国内人口流动的行政控制，作为相互配套的几项基础制度其一（陆益龙，2003）。就这样，以户为单位，以人为对象的户口管理制度出现。在这一制度下，所有的家庭户都必须向居住地的公安机关登记，并被划分为农业户或非农户这两种类型。于是，每一位中国公民的户籍身份就被这样一种极具划分特性的户籍制度所决定（Chan & Zhang,1999）。户籍制度的核心除以上划分特性外，还依据户口籍地管理原则，对不同地区之间的户口迁移进行较为严格的行政控制，因此，二元户籍制度通过控制身份划分和自主迁徙，影响了中国城乡二分结构的形成，也加剧了城市等级差别现象的出现（陆益龙，2003）。

改革开放之前，不论是人口流动还是户口转变，二者都同时受到严格的限制。1978年改革开放以来，随着市场经济的初步发展，政府开始允许农村剩余劳动力进入城市寻找非农工作，劳动和资本开始在城乡间流动，户籍制度开始有了变革。1984年，我国推出一系列新基于打破计划经济体制下对农村户口迁入城镇的限制影响的户籍政策和管理制度，以放宽农业户口迁入城镇的政策。随后在1997年，又对城镇职工的未成年子女、投靠子女生活的老年人、分居夫妻、在城市投资兴办实业的公民等放松了农转非的限制。随着市场化发展进程加快，大

量的流动人口¹涌入到城市，他们虽在城市中工作，却普遍缺乏与当地居民同等的正当居民权益(Duckett J,1999)。结果，基于城乡户口划分的二元体制不仅继续存在，还进一步将这一隐形的身份界线从农村与城市间延伸到了城市内部(Chan,1994)。为了给予农民平等的身份、平等的待遇，2005 年以来，我国开始进行不分城乡的户籍制度改革，倡导取消农业户口和非农业户口之分，居民制度由此产生，但按地区不同，改革进程也不一致。

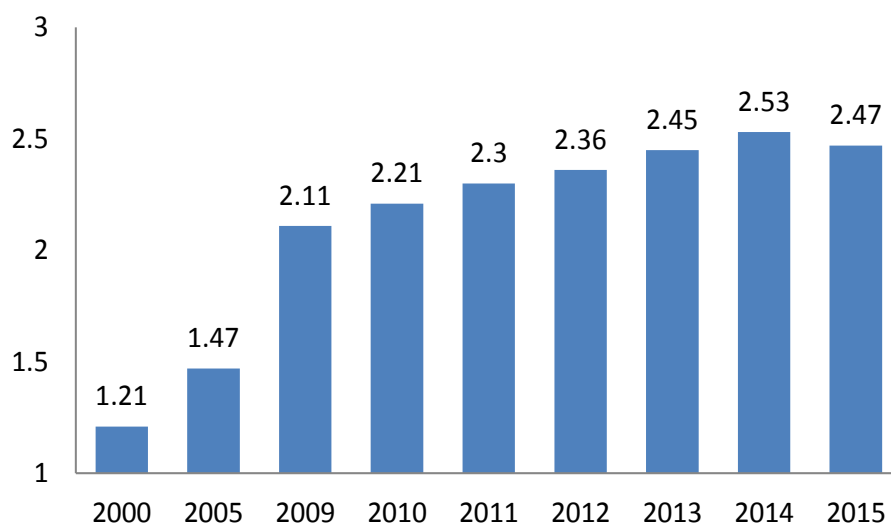
在户籍制度的演变过程中，有一项群体的生存与发展始终活跃在大众视野，那便是流动劳动力群体。随着我国市场化的不断加深，城市和农村经济发展差距逐步拉大，这种地区经济差异就使得劳动者追求更高收入成为劳动力流动的重要原因。自改革开放后的 30 多年来，我国城镇化进程不断加快，流动人口的规模持续增长，尤其是 20 世纪 90 年代以后，更多的农村劳动力涌入城市寻找生存与就业机会，流动人口增长明显加快。从图 1 可以看出，2005 年我国流动人口规模为 1.47 亿人，2010 年达到了前所未有的规模——2.2 亿人，约占我国总人口数量的 17%²。“十二五”期间，我国流动人口年均增长为 800 万人左右，截至 2015 年，这一规模达到 2.47 亿人，如图 2 所示，约占全国总人口的 18%，相当于每 6 个人中就有一个为流动人口。据预测，到 2020 年，我国流动人口将逐步增长至 2.91 亿³。既然劳动力流动起源于城乡间经济差距，那么这些由农村向城市流动的劳动者，收入增长合理化成为其自身最为关注的生计问题。数据²显示：从 1992 年到 2004 年，我国城镇职工人均工资收入由 2711 元增加到 16024 元，增长了近 6 倍，而国务院发展研究中心的报告显示，这十二年间，珠江三角洲外来农民工每月人均工资仅增长了 68 元。可以看出，流动人口的上升趋势并未减弱，而新生代正逐步成为流动人口的主体，这一群体不再是单单生存需要，还包含融入大城市并获得发展的强烈期望。但流动人口选择自主迁徙所获得的收入增长与城镇居民依然有较大差距。农村到城市流动劳动力收入的合理化成为亟待解决的问题，因为这既关系着流动人口自身经济状况的改善，也关系着城乡统筹发展和社会稳定（段成荣、孙磊，2011）。

¹流动人口是指离开了户籍所在地到其他地方居住的人口，本文主要研究由农村到城市的流动人口。

²本数据来自中华人民共和国国家统计局。

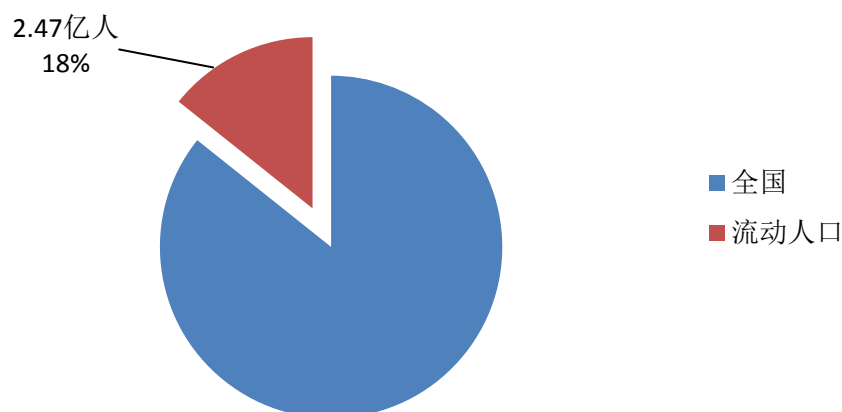
³本数据来自国家卫计委发布的《中国流动人口发展报告 2015》。

图 1 2000 年——2015 年我国流动人口数（单位：亿人）



数据来源：中华人民共和国国家统计局编《中国统计年鉴》

图 2 2015 年我国流动人口占比



数据来源：《中国流动人口发展报告 2016》

1.2 问题的提出

在当下中国现实中，户籍制度不仅仅是身份上的证明，同时也是一种经济地位和社会地位的象征，这与是否能够享受到福利有着密切相关（王美艳、蔡昉，2008）。越来越多的农村人口进入大城市，我们大致分为两类，第一类是选择自主迁徙并获得身份转换的，我们称为“城市移民”，第二类是真正意义上的“流动人口”，即选择自主迁徙但却没有获得身份转换，他们没有迁移户口却在城市

里定居下来，以农民工的身份工作和生活。虽然这些在城市工作的农村人口在很多方面与城市人口没有很大的差别（例如：拥有相同的医院和公园等公共设施），却未能享受到跟当地城市居民一样依附在户籍上的一些社会福利，包括教育、医疗、就业、保险、住房等方面。再者，涉及到后代上学、就业、晋升、劳资关系等，就更显示出户口不同所带来的差别。

因此，很多学者都批评户籍制度导致中国城乡二元劳动力市场的形成，相同条件下，外来农村户籍流动劳动力一般会从事最底层、最低级的工作，因其就业机会、就业待遇以及就业保障方面都受到一定的限制，而且获得较高层级工作岗位的机会少之又少（王美艳，2005）。有学者认为，雇主的歧视导致了这样一种现象，即许多拥有着高工作能力的外来流动人口获得的工作岗位，其工作要求比他们自身人力资本还要低，使他们的工资收入明显低于城市本地居民（田丰，2010）。同时还有学者认为，户籍制度与现代价值理念和现代社会发展方向相背离，在政治上表现为自由迁徙和定居的权利遭到损害，在经济上则表现为自身的知识与能力不能得到应有的回报（Solinger, 1999; Alexander & Chan, 2004; 王小章，2009）；而吴晓刚（2008）和陆益龙（2008）也指出：市场经济要求资本、劳动力等生产要素的自由流动，以达到合理配置的目的，而户籍制度却人为地限制了人口的流动，使得劳动力资源不能得到很好的合理配置。户籍制度这种长期对城市与农村间人口正常迁移的阻断，严重妨碍了人口的合理集聚，阻碍了中国城镇化进程，最终导致整个社会城乡差距的扩大与阶层分化加剧。段成荣、孙磊（2011）进一步指出体制内与体制外工作在教育回报、总收入等多个方面不同，而这恰是因为体制外市场化程度较高的原因。

如今我们大谈户籍改革，然而，改革的依据，改革的方向，都需要我们对户籍制度带来的影响进行清晰的佐证：第一，户籍制度是否对外来务工农业人口的工资性收入产生了影响？如果产生影响，如何确定是户籍制度本身产生的，而不是由于城市、农村成长环境不同造成的禀赋差异？第二、如果没有户籍制度的限制，对于外来务工农业人口在城市工作的收入状况将会有多大程度的改善？第三、体制内与体制外工作受户籍制度影响是否会有着显著不同？这将是本文全篇围绕探讨的议题。

1.3 研究意义

与以往直接比较城镇人口与农村人口收入差距的文献不同，本文把研究重点放在城市外来人口——既包括通过农转非实现迁徙的“城市移民”，又包括以

农业户口在城市进行非农工作的“流动人口”，从工资差异根源视角出发进行实证研究，具有重要的理论和现实意义。就理论意义来讲，可以为我国户籍制度改革提供相应借鉴，完善我国目前户籍制度可能存在的一些问题和缺陷，使我国户籍制度更加健全，此外，还能为缩小城乡间收入差距提供理论依据。就现实意义而言，从公民的角度，本文的实证研究将会有助于人们更深入、更准确认识户籍如何造成了工资差异，进而知道怎样从社会平等的视角出发看待国家出台的一系列改革政策；从社会的角度，基于工资差异根源的实证依据所进行的户籍改革，有利于流动人口在城市平等生活，有利于城市化的健康发展和社会的稳定和谐。

2.文献综述

本部分对相关文献进行综述。在实行市场经济的美国、西欧等国家政府实行人口统计的户籍制度，而城乡分割的二元户籍制度，为我国所特有。因而，关于国外户籍制度所产生的影响的研究相对较少，但仍有很多学者利用中国的数据针对我国的特有户籍制度现象做了很多研究。首先，本文从国内外学者的研究成果中归纳了影响工资性收入重要因素。其次，对工资性收入中的户籍因素进行了概括，包括户籍歧视影响了工资收入、就业机会、社会保障这几个方面。接着回顾了已有学者对户籍制度的研究的方法和总结。最后，对已经文献进行简要总结，确定本文的研究方向和思路。

2.1 工资收入的影响因素

2.1.1 先赋性人力资本因素

先赋性人力资本因素即劳动者的异质性，它会影响劳动者之间的工资差异，这些因素包括性别、健康等。

关于性别歧视所带来的影响，一向是研究工资性差异的重要来源。**Gustafsson and Li(2000)**在研究中国劳动力市场工资性别差异后得出：男女职工的教育回报率不同，男职工要高于女职工。同样针对中国劳动力市场，**王美艳（2005）**把行业分割纳入考虑，用实证方法研究了男女在行业获得和工资上的差异，研究表明差异主要是由同类行业内工资差异引起的，而非不同类行业，即歧视是造成差异的主要原因，女性劳动者处于相对不利的劳动力市场地位。

关于健康对工资影响的问题，是作为劳动者一种至关重要的人力资本因素，影响其工作质量，这便会影响其工资的高低。刘国恩(2004)在研究中国收入不平等问题中提出，健康对收入有显著的正向影响，这是健康不平等趋势，而若这一趋势逐步恶化，中国收入的不平等也会逐步加剧，最终将对我国造成不可挽回的损失。高梦滔、姚洋(2005)基于微观数据对大病风险进行研究，结果发现，大病冲击对于农业劳动者而言，人均纯收入将下降 5%-6%，并且这种负面影响大约要持续 15 年，而对于中下等收入居民影响更严重。

2.1.2 后赋性人力资本因素

后赋性人力资本因素，指劳动者的受教育程度、工作经验、在职培训等人力资本投资，这些都会对劳动者间的工资差异产生重要影响。

已有很多国内外学者研究劳动者的受教育程度对就业和工资性收入的影响。Spence and Solow 早在 20 世纪 70 年代便提出筛选理论，指出，教育只是作为一种信号，以便用人单位识别求职者的能力，劳动者受教育程度不同，获得的劳动报酬也不同。Victor(1979)在研究中证明了在非均衡环境条件下，教育程度会提高劳动者的“配置能力”，有较高受教育程度的劳动者会获得更好的就业机会和收入回报。Siherman and Galor(1990)则认为职业流动带来了职业上升，这是一种人力资本投资回报，不同的教育投资带来的将会是不同的职业发展路径，受教育程度较高的劳动者将会更有机会获得职业流动机会。

我国许多学者也支持教育促进就业，赖德胜(1998)同样研究了人力资本投资回报，结果证实了教育对收入分配有显著的正向影响，并且这种影响随社会的发展而日见明显。高梦滔、姚洋(2006)利用非参数回归方法对我国 8 个省份的 15 年面板数据进行统计分析，最后找出影响农民收入差距的原因——教育和在职培训，二者体现出来的人力资本造成收入差距扩大。

此外，工作经验也是影响人力资本的一个重要因素。Topel(1991)利用微观数据，将劳动者的工作年限作为工作经验的工具变量，发现工作年限与工资性报酬显著正相关。Zhang and Zhao(2002)按工作经验长短将 1988 年至 1999 年的样本观测值分成四个群组，研究发现，拥有不同工作经验群组的收入回报都随年份增加而增加，但相同工作经验下，年轻的劳动力群组会获得更好的收入回报。

2.1.3 非人力资本对工资的影响

非人力资本的影响因素主要有制度、文化、政治、法律等几个方面。已有文献从制度因素角度来解释对工资的影响，赵俊杰(2009)在对现实劳动力市场的研

究分析中指出，处于不完全竞争的状态下，劳动力市场呈现出二元分割的格局，在这种格局下，工资就不仅仅取决于供给与需求，也会取决于影响市场供需变化的制度因素。还有文献从城市偏向性政策角度论述了政治势力、经济政策等对城乡居民工资的影响。Krueger et al (1991)进行调查研究后得出，发展中国家在城镇化进程中，城市偏向的经济政策使得农业承担了大约 30%的总税率，将大量财富从农业中转移以补贴城市工业发展，解释了国家发展战略的所带来的城市 and 农村收入差距问题。

目前来看，大部分专家学者们认为这些非人力资本因素的确存在。赵俊杰（2009）进一步指出，制度因素分为户籍制度和其他制度因素，而由于文化、法律和政治因素的难以操作性，无法很好地测量，本文更多地关注制度因素层面中户籍这一层面的影响。

2.2 户籍制度对工资收入的影响

2.2.1 城镇职工与农民工的收入差异

首先，现有的实证研究证明我国城镇职工与农民工存在收入差异这一事实。刘安萍(2004)利用全国范围内数据，将城市最高收入的 10%的群体与农村最低收入的 10%的群体进行对比，发现前者的人均可支配收入是后者人均可支配收入的 8 倍以上，收入差异及其显著。王洪亮、徐翔(2006)对我国 1978 年至 2003 年城乡和地区间的收入差距进行统计分析，指出我国城乡收入差距在我国总体收入差距中占绝对主导作用，城乡收入差距要显著大于地区间收入差距，并且，在我国全体居民收入不均等中，城市劳动者与外来务工劳动者之间的收入不均等贡献率最高。严善平(2007)研究了上海市四次调查的微观数据，发现非农户口的工资水平比农民工高 15%，并且这一差距在四次调查过程中都没有发生变化，说明这种收入差距现象在上海市劳动力市场上长期存在。郭震(2013)通过对 2012 年一个全国性的抽样调查结果进行分析发现，城镇职工的年收入显著高于农民工，且随着分位数的提高，这种收入不均等的情况也越来越明显。以上研究结论不仅说明了我国城镇职工与农民工工资收入差异的存在，同时也强调了这种差异具有普遍性、长期存在性、分阶层不同、有扩大趋势这四大特点。

2.2.2 户籍制度影响工资收入

已有大量实证研究表明，户籍制度对于工资性收入差异具有一定的解释力度。Meng&Zhang (2001)使用 1995 年和 1996 年上海市的调查数据，研究发现户

籍制度的解释力度高达 110%。王美艳(2003)采用 2000 年北京大学社会学系等研究机构对一批企业和员工的调查数据的调查数据,计量研究结果表明,户籍制度可以解释 76%的农民工与城镇职工间工资差异,这是国内学者较早对户籍制度影响工资收入的实证研究结果。姚先国和谢嗣胜(2006)通过对浙江省 2003 年和 2004 年的调查数据进行分析发现,城市劳动者与外来务工劳动者平均工资水平差异巨大,其可以被户籍制度解释的力度有 55.2%。并且在研究户籍制度影响工资收入中,微观数据库 CHIPS、CHNS、CGSS 等都被学者们广泛运用。邓曲恒(2007)使用经典的 Oaxaca-Blinder 工资差异分解方法,利用 CHIPS2002 年的数据进行实证研究,发现户籍歧视可以解释城镇职工与农民工收入差异的 60%。余向华和陈雪娟(2012)使用了 1993 年-2009 年的 CHNS 数据,发现户籍制度的分割作用在城镇职工与农民工收入差异中一直保持在 40%以上。以上文献不仅说明户籍制度对工资收入在统计意义上的显著影响,也说明了这种劳动力市场不平等待遇一直存在并将长期存在的社会事实。

2.2.3 户籍制度影响就业机会

早期计划经济时期,在户籍制度的管控下,农村劳动力的身份转换和自由迁徙受限,基本只能在家务农。这时对于农村劳动力而言,“农转非”是获得进城就业机会的关键,然而实现农转非的途径有限,也有严格的指标控制。赵耀辉和刘启明(1997)在一项城镇人口迁徙抽样调查中提出,农民获得“农转非”有多项渠道,其中,城镇招工和家属随转为主要渠道,分别占从农村迁入城镇总人口的 46%和 35%。然而,在 1961 至 1963 年精减职工期间,通过城镇招工的农名工成为了主要的精简对象,可见精减职工实际上首要受害者是进城务工的农民工,尽管他们实现了自由迁徙,却不能保证与城镇职工同等的就业机会和就业地位,稳定性方面也大打折扣。汪和建(1998)把劳动力市场分为无组织或者只有简单组织的非正式部门和具有高度组织管理和社会认可的正式部门,研究发现,农民工大多数只能进入非正式部门,即使他们中的少部分能够进入正式部门,也只能以合同工的身份被雇佣,从事着城镇职工不愿从事的工作。

在市场经济形势大好的今天,户籍制度对就业机会影响也并没有消除。郭菲和张展新(2012)总结出农民工就业层次低的提点,发现国有部门受到地方政策保护,一没有本地户口的农民工大都很难进入国有部门,他们聚集最多的是股份制企业、私营企业或自营企业,占高达 78.5%的比例。此外,有学者还发现,户籍制度对就业机会的影响具有一定的代际效应。彭希哲等(2009)指出,拥有不同户籍的城镇和农村劳动力会因出生地、父母户口性质和父母职业不同而面临不平

等的就业机会，这种不平等是由户口因素“先天”决定，与个人能力和后天努力无关，彭希哲等强调这种先天歧视是户籍制度的最大弊端。

2.2.4 户籍制度影响社会保障

现有研究成果也确认了在社会保障的获得方面，城镇职工和农民工之间存在差异。姚先国和赖普清(2004)使用 2003 年至 2004 年公司和农民工调查数据，发现城镇职工与农民工在养老保险、失业保险、医疗保险方面均存在由户籍制度引起的差异，差异的程度分别为 31%、26%和 21%，这对户籍制度的影响研究又进了一步，因为如果仅关注工资收入差异，将低估城镇职工和农民工实际经济和福利总收益所带来的差异。成志刚和罗帅(2007)进一步对农民工相比城镇职工在获得社会保障上存在的问题进行了总结，主要存在参保率低、失业保险严重缺失以及养老保险难以企及等几个问题，他们强调，城乡二元分割体系亦造成了城乡二元社会保障体系，这在城镇职工与农民工不能获得同等条件的社会保障下首当其冲，农民工因为自身的户籍身份差异，无法与城镇职工享用同样的社保体系，获取跟城镇职工平等的福利。以上研究均披露出了当前农民工在获得社会保障上的劣势性。

2.3 户籍制度工资效应的研究争议

2.3.1 针对外来人口与本地市民收入差异分解的争议

国内外学者关于户籍产生的影响很多是基于理论分析，后来逐步出现很多关于户口的实证分析，而且这些研究基本上遵循了 Blinder 和 Oaxaca 的工资差异分解模型的分析思路（Blinder，Oaxaca，1973），把城市劳动者和外来劳动者工资收入差异分解为两部分：来自禀赋的收入差异和歧视造成的收入差异（谢嗣胜、姚先国，2006；王美艳，2005）。于是，研究歧视差异的思路便是，在控制前者禀赋差异的情况下，后者带来的差异水平越大，那么就意味着户籍改革给外来劳动者增进的福利越大。国内学者利用这种方法来测量歧视的原因，但是得出结论却有所不同，存有争议。一部分学者支持户籍歧视能解释一定力度的收入差异（姚先国、赖普清，2004；韦伟、傅勇，2004；邓曲恒，2007；陈维涛、彭小敏，2012 等），而另一部分学者却把外来人口与本地市民收入差异更多地归结于二者之间的禀赋（吴晓刚，2008；李培林、李炜，2007；邢春冰，2008 等）。

2.3.2 支持收入差异来自于户籍歧视的研究

很多学者将外来人口与本地市民收入差距主要归结为户籍歧视：邓曲恒（2007）利用 CHIPS2002 年的大样本数据进行研究，发现城镇居民和流动人口收入有着显著的差异，而户籍歧视对这些差异的解释力度为 60%。此外，韦伟、傅勇（2004）在研究人口流动模型中指出，中国的城乡收入差距问题在国家转型期出现得尤为严重，这是城乡分离制度安排所导致的，即存在户籍制度歧视。姚先国、赖普清（2004）利用我国浙江省流动劳动力的调查数据，研究表明本地市民与外来人口收入差异的 20%—30%是来自户籍歧视影响的。陈维涛、彭小敏（2012）利用 CHIP 数据得出，“非农户口”更有利于劳动者获得高收入的岗位，机会不平等导致了收入不平等，这就使得本地市民的收入、地位、福利等普遍比外来人口高。

2.3.3 支持收入差异来自于禀赋的研究

与上述得出结论不同，另一些研究者把本地市民与外来人口收入差距的主要差异来源，归于二者之间的禀赋，而非户籍歧视。李培林、李炜（2007）的研究发现，受教育程度和劳动技能所代表的人力资本是外来人口工资收入低于本地市民的主要受限原因，因户籍身份导致的收入水平受限并不显著。邢春冰（2008）也利用 2005 年人口普查数据的研究结果支持了这一点，同样认为受教育程度是造成城镇职工与农名工收入差距的最主要原因，数据显示，农名工的平均小时收入为城镇本地职工平均小时收入的 64%，但劳动者的禀赋特征就可以解释这其中约 90%左右的部分。吴晓刚（2008）利用一项全国性的抽样调查数据，在对城镇户口的影响因素实证分析时，发现了教育和政治条件虽然会增加农转非的机会，但是农村户口则明显减少了教育和政治地位获得的机会。肖文韬在其研究中指出，大量的农民工进城的事实可以看出户籍制度实际上没有成为农民工自由迁徙的阻碍，然后通过数据分析，他认为必要工业化人口与年度工业化人口的差异比率过高，就业形势严峻是阻碍农名工向城市自由转移的主要原因。

2.3.4 关于体制内、外部门与户籍制度关系的研究

有关体制与户籍制度关系的研究还很少，段成荣、孙磊（2011）利用 2005 年人口普查数据进行实证研究，得出体制内与体制外工作在性别、年龄教育回报、总收入等多个方面不同，而这恰是因为体制外拥有较高市场化程度的原因。而陆益龙（2008）也指出，虽然越来越多的流动人口进入城市，但是却是在工作上有着很多差别，而这种差别又是以一种稳定对不稳定、体制内对体制外的形式体现出来。魏万青（2012）研究了户籍身份在体制外和体制内部门各自作用的影响，

发现在市场化程度越高的部门，户籍身份的负向作用越为显著，这就说明随着市场化进程加快，并不能使外来人口的户籍身份限制得到有效缓解。进一步研究发现，将劳动者按受教育程度分为阶梯级，在收入前 10% 的高级人才群体中，市场化进程越高，户籍限制反而起到了更为明显的负向作用，户籍制度的弊端可见一斑。

2.4 小结

2.4.1 现有文献研究成果与存在的问题

综上，关于户籍制度造成的收入差异一直都是学术界研究的热点，研究主要聚焦于户籍人口与流动人口。对于影响因素的探究，主要包括户籍制度和人力资本两个因素，以及流动人口的社会网络、人口学特征等因素。大量文献已经实证得出户籍制度对工资性收入具有显著影响，这也是本文得以在前人成果上进一步研究的基础。

然而，在归纳前人成果的过程中也发现了一些问题。首先，在对象选取上，现有文献多数是直接比较城市本地户口居民和外来流动人口在同等条件下的收入差距，认为此差距即农村与城市户籍制度带来的差距，但这样直接比较是有问题存在的，因为这样只关注了外来流动人口与本地居民可观测到的因素，而忽略了他们之间不可观测到的因素。详细地说，前者是已经被国内外学者公认的——禀赋（主要为人力资本）和由户籍身份，但后者是大部分文献忽略了——成长氛围、文化观念、医疗条件、教育质量等一系列不可观测的因素，这些因素也会间接地影响个体工资性收入，将这种由不可观测因素导致的收入差异归结于户籍歧视显然过于勉强。举例来说，农民工与城市本地市民相比较，即使其他素质方面不存在劣势（如即使教育程度一致），但农民工由于从小在农村中所受的教育质量问题，也可能带来收入差异，这部分差异应归于教育质量资源配置的方面，而非户籍歧视。因此，如果直接将流动人口与本地人口直接比较，比较而来的差异就算做户籍身份不同带来的差异，得到的估计结果是有偏的，从统计意义上来说，为遗失关键变量（Brand & Xie, 2010）。

其次，在研究方法上，大多数文献都是基于 Blinder 和 Oaxaca 的工资差异分解模型，然而其中存在着“指数基准”问题（郭继强，2009；魏万青，2012），利用了不同的基准指数，会得到不同的回报结果——这就是大量学者利用这种方法测量“歧视”的来源，却得出不同结论的原因。

最后，大部分研究没有考虑到地域歧视的影响。国内外学界都有意识到对来

自其他地区的居民的歧视文化对其工资收入有一定的影响,但如果不区分户籍制度和地域歧视的影响效果,就等同于间接认为地域歧视完全是户籍制度造成的,这显然是牵强而不符合事实的。

2.4.2 本文的研究视角与创新

研究对象上,本文避开直接比较城市本地户口居民和外来流动人口的弊端,选取的均为外来人口,分为处理组与控制组:处理组为获得了城市户籍的外来人口——“城市移民”,即从小生活在农村,通过上学、入伍、招工招干、入伍、城市化过程等实现“农转非”的群体;控制组为没有获得户籍的外来人口——“流动人口”,即从小也是生活在农村,带着“农民工”的身份在城市工作和生存的群体。选取这两部分进行比较研究的优点是:两个群体绝大多数都是来自农村,不在流入地城市成长,这样就控制了成长环境一致,排除教育质量、成长环境、文化观念等不可观测的因素产生的差异,只单纯研究户籍制度所带来的差异。在研究对象选取上也控制了上文提到的“地域歧视”问题。进城务工的劳动者人生地不熟,在当地缺少血缘等关系带来的帮助,语言适应、文化适应都成为移居外地的障碍,收入也在一定时间处于相对低的水平 (Winkelman, 1999)。本文选取的均为外来人口,在当地的关系网等社会资源均处于比较匮乏的相似状态,从而排除了城市和农村间“地域歧视”的影响。

研究方法上,本文主要选取了由 Rosenbaum 和 Rubin(1983)提出的倾向得分匹配法(PSM)进行分析,并采取自抽样法(Bootstrap)进行实证回归,辅以代理变量方法,关于这几种方法将会在第三章详细阐述。本文的核心思想是用倾向得分匹配法(PSM)进行实证分析,因样本的选取本身就存在一定的选择性偏误,采用倾向得分匹配法(PSM)的最大优点是可以克服这种偏误。简单地说,想探讨户籍制度的影响,最理想的结果是选取同一群体没获得城市户籍的收入和获得了城市户籍的收入进行对比研究,但现实中只能观测到其中一种,不可能同时观测到这个群体拥有城市户籍和没有城市户籍两个结果,这是不现实的。因此,我们想研究这两种收入,只能选择两个群体——有城市户籍的外来人口(城市移民)和没有城市户籍的外来人口(流动人口)。但如果直接将二者进行比较,就未能控制“同一群体”这一条件,就产生了样本的选择性偏误。倾向得分匹配法的思路是:首先将整个群组分为处理组(城市移民)和控制组(流动人口),将两个组进行匹配,即为处理组找到尽可能相似的控制组,再来估计户籍制度的影响,这样便可以有效降低样本的选择性偏误。

研究数据上,本文基于数据的可得性与最新性,选取了中国人民大学社会学

系全国综合社会调查(CGSS)2012 年的微观数据。这样，本文在研究对象选取、克服样本的选择性偏误、新近经验数据上均有所创新与进步。

3. 理论基础与研究方法

3.1 相关理论

一直以来，学者们在解释户籍制度造成的收入差异方面存在着两种分析理论——微观角度的人力资本理论和宏观角度的二元劳动力市场理论，以下分别进行阐述。

3.1.1 人力资本理论

人力资本理论(Human Capital Theory)起源于 18 世纪，阐述的是教育程度、技能、工作经验等自身拥有的人力资本对劳动者收入起着决定作用，它的前提是在劳动力市场完全竞争的条件下。劳动者拥有更多人力资本（教育程度越高或工作经验越丰富），就拥有更强的生产能力，就能给企业带来更高的生产效率，创造更多的利润，劳动者个人也能得到更多的收入回报。不同的学者对人力资本有不同角度的界定，其中也有共同点，认为它是个人通过社会性的教育投资、健康投资、培训投资等方式，凝结在劳动者个体中专属的可以提高个人和社会预期货币性或非货币性收益的综合能力或素质的总称。人力资本理论对现代社会收入差异的解释是，不同的人对教育等不同方面的投入不同，所以每个劳动者自身形成的人力资本也是数量不一的，导致每个劳动者所具有的各方面的能力也不同，最终这种能力反映到每个人所获得的工资性收入亦是有差异的。

3.1.2 二元劳动力市场理论

二元劳动力市场理论(Dual Theory)出现在 20 世纪 70 年代初期，它强调劳动力市场具有比较明显的非竞争性。西方学术界在研究劳动力市场时，根据大的垄断企业和小的竞争性企业间市场势力差别，将劳动力市场区分为两大部门：一级劳动力市场（“主要部门”）和二级劳动力市场（“次要部门”）。这两大部门以迥然不同的市场机制运行，一级劳动力市场主要由拥有适当凭证和个人素质的劳动者构成，工资收入较高，稳定性有所保障，工作环境良好，而且拥有被提升的机会。二级劳动力市场的就业人员则是缺乏凭证和个人素质较差的群体，工资收入不高，稳定性和工作环境都较差，并且没有被提升的渠道和机会。在具体的实证

研究中，对一级和二级劳动力市场进行区分的指标有三个：就业稳定性、晋升机会以及福利的普及保障程度(李路路，边燕杰，2008)。

我国的劳动力市场同样存在着二元结构，与城市本地人口相比，外来人口无论是在受教育情况、文化背景、经济状况、关系网络等方面都存在一定的弱势，差距是不可避免的，从而影响到外来人口的职业选择，表现出来的现象便是与迁入地城市本地人口不同的职业结构。

结合以上两种理论，本文主要关注的是制度因素，然而即便流动人口和城市移民进入不同的劳动力部门，这种差异也不能说全是由制度因素造成的，要牢记前提是这两个群体在人力资本上的差异导致市场供需的自然选择。并且人力资本理论是建立在劳动力市场完全竞争条件上的，二元劳动力市场理论的建立基础是劳动力市场非完全竞争，这两种理论相辅，从微观、宏观视角共同构成本文研究的理论基础。

3.2 研究方法

为研究户籍制度对工资性收入的差异，本文首先拟采用代理变量方法，是考虑到遗漏变量问题，控制不可观测的“能力”因素造成的偏误。主要采用的方法为倾向性得分匹配法，目的是为控制选择性偏误，预测控制了选择性偏误后，得到的结果要比代理变量方法得出的更为显著。在具体的统计推断上，采用自抽样法，是为调整异方差，克服小样本偏误。三种方法的原理阐述如下。

3.2.1 代理变量方法

本文研究分析的第一步拟采用代理变量方法。在研究包含户籍制度在内的各项因素对收入影响时，我们并不能保证将所有影响因素都涵盖在内，故回归残差项往往与因变量收入有关，造成遗漏变量的偏误。这时寻求一个合适的代理变量，用以代表不可观测因素（如能力），可以有效控制偏误，使得剩余残差项与因变量收入无关。

需要指出的是，不可观测的“能力”一般会导致内生性问题，而内生性问题多由工具变量解决。具体是选择代理变量还是工具变量，要依模型设定和变量可获得性而定：工具变量起暂时替代的作用，替代的是模型中有的内生解释变量，相当于辅助工具，帮助内生解释变量能够更准确地解释回归结果；而代理变量起完全替代作用，替代的是无法直接获取到解释变量的观测值，代理变量本身就是具有统计意义的，可以用来解释回归结果。本文可能造成内生性问题的是无法直

接获取到解释变量的不可观测的“能力”，故使用代理变量方法。

3.2.2 倾向得分匹配法(Propensity Score Matching , PSM)

前文已提到，假设我们简单地将是否发生了某项事件(如是否获得非农户口)作为虚拟变量，而对总体进行 OLS 回归，所得到的参数估计会产生偏误，因为在这种情况下，我们只观测到了某一个群体因为发生某一事件后产生的表现，并且拿这种表现去和另一个没有发生这一事件的其他群体做比较，这样比较显然是不科学的，因为比较的基础不同。倾向得分匹配法（PSM）便可以解决这一问题，它是一种估计因果的处理效果的研究方法，被广泛用于统计分析，涵盖了医学、社会学、经济学等领域，多用于医学上的治疗、经济学上的政策或某项事件的影响效果研究。

从概念上来说，倾向得分匹配法（PSM）的定义为：在一定条件下，个体受到某种事件干预的可能性。具体地说，倾向得分匹配法（PSM）创造的是一个“准随机”试验，里面包含两组：处理组与控制组，它的目的是找到一群与处理组所有预处理相关特征类似的控制组，这样，足够的控制组产生的相较于处理组的不同结果就可以归因于该干预项目。倾向得分匹配法（PSM）把许多不同的特征因素浓缩成为一个指标——倾向得分值(propensity scores，简称 pscore)，它是所有协变量的一个函数，用来均衡两个试验组之间的协变量分布，是为克服样本选取带来的选择性偏误。在实证过程中，倾向得分值(pscore)一般运用 Logit 或 Probit 概率模型来估计。

结合本文来看，本文研究的是户籍制度对拥有不同户籍的外来人口的工资性收入的影响。自主迁徙是劳动者“自我选择”的权利，城市户籍身份是每个劳动者“自我选择”的结果，而不是随机获得的。是否拥有户籍身份是由每个劳动者各项特征变量所决定的，这是每个劳动者的个体差异，但这些特征变量既决定了劳动者是否拥有城市户籍，也决定了每个劳动者不同的工资收入水平。所以，即使我们用一般性方法证明了二者的正向影响效应，也不能说明是拥有城市户籍身份的人工工资收入更高，还是工资收入更高使这些外来人口更有可能获得城市户籍身份。那么，为了研究户籍制度的影响效应，尽可能地控制不可观测因素的干扰，本文把总体外来人口分为两个试验组：处理组——具有城市户籍的外来人口，控制组——没有城市户籍的外来人口。两组在很多方面相似，如成长环境都为农村，受到的教育质量相似等等。具体运用倾向得分匹配法（PSM）的步骤为：运用 Logit 模型获得倾向得分值(pscore)，找到与处理组样本特征变量类似的控制组，将二者进行收入差异对比。由于两组样本特征变量类似，便可以将他们之间的收入差

异归于户籍制度的影响，户籍制度的效果——平均处理效应(Average Treatment Effect,简称 ATT)也就可以较为准确地被估计出来了。

常用的估计 ATT 的匹配方法包括最近邻匹配法(nearest neighbor matching)、半径匹配法(radius matching)和核匹配法(kernel matching)。最近邻匹配法和半径匹配法本质上都是近邻匹配法，匹配结果是最近的部分个体。最近邻匹配法首先估计出倾向得分值(pscore)，并以此为依据，为处理组样本向前或向后找寻拥有最接近倾向得分值(pscore)的控制组样本。半径匹配法则是估计出倾向得分值(pscore)后，运用预先设好的常数，为处理组样本找到倾向得分值(pscore)之间的差值小于该常数的控制组样本。核匹配法是整体匹配法，根据个体距离不同赋予不同的权重，其权重使用核函数来实现，每位个体的匹配结果为不同组的全部个体。

3.2.3 自抽样法(Bootstrap)

自抽样法(Bootstrap)属于有放回的重重复均匀抽样方法，在现代统计学中，它是一种重要的估计统计量方差的方法，在小样本时效果较好。方差估计之后可进一步进行区间估计，构造置信区间等，由于倾向得分匹配法获得 ATT 的同方差假设不一定成立，所以采取自抽样法来估计平均处理效应(ATT)的标准误。运用自抽样法的核心思想和基本步骤如下：

- (1) 采用重抽样技术从原始样本中抽取一定数量的（自己给定的 N 个）样本，此过程允许重复抽样；
- (2) 根据抽出的样本计算给定的统计量 T；
- (3) 重复上述 M 次（一般大于 500），每次均抽取 N 个，得到 M 个统计量 T；
- (4) 综合计算上述 M 个统计量 T 的样本方差，得到原始统计量的方差。

本文拟采用 500 次的自抽样用以估计标准误，其实现步骤为：

- (1) 首先从原始样本中重复地随机抽取我们所需要的观测值作为经验样本；
- (2) 用上述匹配方法估计经验样本的平均处理效应（ATT）；
- (3) 将以上两步重复 500 次，便可以获得 500 个 ATT 的统计量；
- (4) 计算这些统计量的标准误，即可得到原始样本 ATT 的标准误。

这样，经由 500 次自抽样估计得到的 ATT 便是统计意义上有效的。

4.模型设定、数据与变量

4.1 模型设定

接下来是本文的实证模型设定。首先，本文把造成城乡收入差距原因分为与户籍制度有关的因素和与户籍无关的因素，无关的因素是可以由教育、经验等可以解释的部分，这部分与户籍制度没有关系，完全是个人的选择结果。进而，重点在于与户籍制度有关的因素，户籍制度的限制所产生的测量机制为：通过选择同是在农村长大同时也都在城市工作的被访者，考察是否转入城市户口对其工资收入的影响。

4.1.1 代理变量的估计式

第一步采用代理变量回归的方法。首先，遵循 Mincer 模型的基础，估计的工资方程形式如下：

$$W = \alpha + \beta registered + \sum_{j=0}^n r_j \overline{X_j} + \xi \quad (1)$$

(1)式中，因变量 W 为小时工资的对数， $\overline{X_j}$ 为影响工资决定的个人特征向量，它尽可能包含了关于影响个体收入的变量（如性别、教育、年龄等），这样便减少了遗漏变量的存在。 r_j 为各特征向量的估计系数，表示每个个体变量对工资的贡献程度。 $registered$ 为是否有城市户籍， ξ 为残差项， $j=0、1、2...n$ 。

然而，如果直接将两个外来人口群体通过 OLS 回归比较工资性收入，可能产生内生性干扰，因为残差 ξ 中有绝大多数可能包含了不可观测因素，如“能力”。由于不可直接获得，就没法直接加入 $\overline{X_j}$ 中。于是，本文采用“能力”的代理变量——每个外来人口拥有城市户籍身份的概率，加入 $\overline{X_j}$ 中进行估计，而这个概率便是倾向性得分 $pscore$ 。

在该模型的假设下，加入代理变量后，如果实证结果 $registered$ 的系数 β 得到为正，则说明拥有城市户口与工资之间存在正相关，因此便可验证户籍制度对工资差异有影响。

4.1.2 倾向得分的估计式

城市户籍身份不是随机分配给我们所选取的外来人口样本的，外来人口是否拥有城市户籍由他们的自身特征所决定。根据其理论定义，结合本文的实际研究问题，将倾向得分定义为：在外来人口的个人特征 \mathbf{x} 已知的情况下，外来人口获得城市户籍的条件概率是：

$$p(\mathbf{X}) = P[T = 1 | \mathbf{X} = [\mathbf{E} \ T]] \quad (2)$$

其中， $p(\mathbf{X})$ 就是需要估计的倾向得分值， T 是条件函数，如果外来人口是城市户籍，则 $T=1$ ，若不是城市户籍则 $T=0$ 。由于倾向得分值 $p(\mathbf{X})$ 的不可观测性，本文采用 Logit 模型来估计，选取的样本特征 \mathbf{x} 有年龄、性别、工作经验、受教育年限、婚姻状况等变量，用于给处理组（城市移民）匹配尽可能相配的控制组（流动人口）。

4.1.3 平均处理效应（ATT）估计式

根据上一节表达式估计出 $p(\mathbf{X})$ 之后，对于我们所选取的每个外来人口样本 i ，城市户籍身份对外来人口 i 工资性收入的平均处理效应（ATT）表达式如下：

$$\begin{aligned} ATT &= E[Y_{1i} - Y_{0i} | T_i = 1] \\ &= E\{E[Y_{1i} - Y_{0i} | T_i = 1, p(\mathbf{X}_i)]\} \\ &= E\{E[Y_{1i} | T_i = 1, p(\mathbf{X}_i)] - \{E[Y_{0i} | T_i = 0, p(\mathbf{X}_i)] | T_i = 1\}\} \end{aligned} \quad (3)$$

其中， Y_{1i} 表示个体 i 拥有城市户籍身份的工资性收入， Y_{0i} 表示个体 i 没有城市户籍身份的工资性收入。显然，现实中 Y_{1i} 和 Y_{0i} 只能有一个被我们观测到，同时得到这两种收入是“反事实”的。为此，用倾向得分匹配方法(PSM)创造“准随机”试验来估计 ATT，得到无偏的结果。

4.1.4 模型假设与实现步骤

本文的主要目的是运用倾向得分匹配方法(PSM)来探究户籍制度的影响差异，估计式(1)只是为(2)和 (3)形成参照对比。估计式(2)和 (3)的成立需要满足两个假设条件：

(1) 条件独立分布假设。在该假设下，控制外来人口的个人特征向量 \mathbf{x} ，个人的工资性收入独立于拥有城市户籍与否这个变量。并且，是否拥有城市户籍是随机分配给外来人口样本的，这样，当我们找到个人特征向量 \mathbf{x} 相近的样本，便可以比较拥有城市户籍与否对工资性收入带来的变化；

(2) 共同支撑假设。我们知道，倾向得分值 $p(X)$ 表示的是每个外来人口拥有城市户籍身份的概率，该假设确定了 $p(X)$ 的取值为 0 和 1 之间。这样，取值位于 $p(X)$ 两端的外来人口样本便被排除掉了，保证了匹配的准确和质量。

本文的实证阶段实现步骤如下：

首先，用 Logit 模型估计出倾向性得分 $pscore$ ：

$$p(X_i) = P(T_i | X_i) = \frac{\exp(\delta X_i)}{1 + \exp(\delta X_i)} \quad (4)$$

X_i 表示每个外来人口的个人特征变量，本文选取的特征变量有：年龄、年龄平方、性别、工作经验、受教育年限、婚姻状况、政治面貌、民族、健康水平、阶层水平。 δ 为估计得到的相应特征变量的系数。这样便估计出了每个外来人口获得城市户籍的概率 $pscore$ ，也即估计式(1)中所需的代理变量。

然后，将上式得到的倾向性得分 $pscore$ 代入(1)式，得到城市户籍的系数向量 β ，观测其统计意义上的显著性。

其次，根据得到倾向性得分 $pscore$ ，对处理组与控制组进行匹配。匹配的思想是，将样本等分为 n 个组，在每个小组中，分别对处理组和控制组算出其倾向性得分 $pscore$ ，进行比较，判断有无显著差异，若有则进一步分组，重复上述检验，直至每个小组中平均 $pscore$ 值在处理组和控制组之间都相等。再比较每个小组中处理组和控制组的各解释变量的均值，判断有无显著差异，如果有显著差异性则返回上述步骤，重新设定 $\exp(\delta X_i)$ 的形式，直至每个小组处理组和控制组的倾向性得分 $pscore$ 、各解释变量的均值均没有显著差异，匹配完成。

匹配的具体操作方法是最近邻匹配法(nearest neighbor matching)、半径匹配法(radius matching)和核匹配法(kernel matching)这三种，运用每种匹配方法的同时都采取自抽样法(Bootstrap)，也就是说，在用这些匹配方法估计平均处理效应 ATT 时，均重复 500 次，获得 500 个平均处理效应的统计量，以得到标准误下的 ATT。各匹配方法和自抽样的操作都可以通过 Stata 命令进行。

最后，比较匹配前后的 t 值和 p 值水平判断匹配结果是否显著，若显著，即可进一步解释估计出的 ATT 背后的涵义，并进行分组估计对比。

4.2 数据来源

本文所用数据来自于中国人民大学社会学系实施的中国综合社会调查 (Chinese General Social Survey, CGSS), 鉴于数据的可得性和最新性, 本文选择 2012 年的 CGSS 截面数据。CGSS 是我国最早的全国性和连续性调查项目, 调查点覆盖了中国大陆所有省级行政单位, 它始于 2003 年, 每年进行一次, 对全国 125 个县 (区), 500 个街道 (乡、镇), 1000 个居 (村) 委会、10000 户家庭中的个人进行调查, 定期、系统、全面地收集了社区、家庭、个人等不同层次的数据。到 2008 年, CGSS 完成了项目的第一期, 完成了五次调查并向社会公开数据。CGSS 于 2010 启动了第二期, 预计 2019 年结束, 期间每两年调查一次全国大调查。目前, CGSS 数据已成为研究中国社会现象最主要的数据来源之一, 广泛地应用于科研、报告及政府决策之中。

本次 CGSS2012 年的数据调查采取分层三阶段概率抽样的方法, 调查总体为全国城镇和农村居民, 被分成两个部分, 第一部分为必选层, 包含直辖区之内的城市的家庭户; 第二部分为抽选层, 包含市辖区之外的城市及农村家庭户。根据本文的研究目的, 我们选择的是 CGSS2012 中所有外来人口样本, 包括流动人口和城市移民。通过筛选和合并的方式, 首先去除了所有的农村样本, 保证个体的工作地点都在城市, 通过“您目前的户口登记状况”筛选出农业户口个体, 这部分为流动人口; 再通过“您的户口是哪一年迁到本地的”这一问题来判断城市移民, 剔除了自出生就是非农户口的个体, 最后将以上两部分样本合并, 就是本文用以研究的外来人口。经处理, 最后得到的总样本为 2676 人, 其中流动人口为 1439 人, 城市移民为 1237 人。

4.3 指标选取与变量说明

4.3.1 因变量

估计式 (1) 中的因变量代表的是工资性收入, 在 CGSS 数据中, 工资收入变量是被访个体在 2011 年全年的职业/劳动收入, 排除了个体的职业外收入和转移支付。同时, 本文还排除了个人全年工资性收入高于百万位数的情况, 是为剔除极值和异常值, 减少对回归结果的干扰。由于收入为全年的总收入, 为了排除因个体工作时间的长短而产生工资的差距, 我们利用全年总收入得到每周的总工资收入, 进而除以数据中个体每周工作时间而得到小时工资, 这样可以更好地衡量个体收入的差距。按照大多数学者运用工资分解方法的模型 (傅勇、韦伟, 2004; 谢嗣胜、姚先国, 2006; 邓曲恒, 2007; 陈维涛、彭小敏, 2012 等), 最终选取的变量是小时工资对数。

4.3.2 自变量

估计式（1）中所关注的变量为虚拟变量：是否拥有非农户口。其余表示外来人口样本特征的是控制变量，具体选取情况如下：

（1）年龄

CGSS 问卷中询问的是“您的出生日期是什么”，本文用此选项减去 2012 得到每个个体的年龄。原数据中个体年龄范围很广，而我们需要的时正处于工作阶段的人的工资水平，所以本文去掉了那些大于 60 周岁的个体。一般而言，个体都是 65 岁就基本已经停止工作，其次，即使仍在工作，那么也与没有退休时有一定差别，即此时工资已经不能很好地反映个体信息，加入后会对整个结果产生干扰。

（2）工作经验

已有文献经验的计算方法是按照个体最后工作单位所累积的工作年限作为个体的代理变量，因有数据表明：个体在之前的公司工作时间与在此公司工资关系微小，工资会随个体在最后公司（一般而言问卷中便是当前工作公司）工作年限的增加而增长，但随之增长的幅度会越来越小。而 CGSS 调查问卷中问的是“从第一份非农工作到目前的工作”总共的工作年限，问得更加严谨，是从非农工作算起，本文考察的便是对工资性收入影响，故直接运用这一问题数据作为工作年限。

（3）人力资本投资

其一，教育程度，问卷中对于教育调查的比较详细，本文进行了简化。所要提出的是，以往文献一般加入的是教育年限变量，而 CGSS2012 年数据并没有调查教育年限一项，只问了“目前的最高教育程度”其中包括正在就读的，并从“没有受过任何教育”到“研究生及以上”细分了 13 项。因所接受的教育年限也是作为教育程度的衡量，所以本文直接选取的变量是受教育程度，得分越高所受教育程度也就越高。

其二，健康，健康状况作为人力资本的重要内容，健康状况越佳，理应对收入有正向影响。本文中将健康分为很不健康、比较不健康、一般、比较健康、很健康五种情况，得分越高健康程度越佳。

（4）个体特征变量

其一，性别，在前文对先赋性人力资本因素的文献综述中已指出，性别歧视所一向是工资性收入差异的重要来源，一般来讲，使得女性的平均工资收入要低于男性的平均工资收入水平。

其二，婚姻，本文将婚姻状况分为两类：有配偶、其他（例如未婚、离异、丧偶等）。本文还将是否是党员、少数民族两个虚拟变量放入，其中党员作为政治身份的代名词，在当今中国的劳动力市场上已成为某些行业及工作的门槛，例如高等学校教育的思政类工作，国家公务员考试的某些岗位也有所需求等，党员身份成为不可忽视的一种政治资本。此外，本文还将自评阶级纳入个体特征变量中，CGSS 问卷中问道，“当今社会，有的人处在社会的上层，有的人则处于下层，您认为自己目前在哪个等级上”，该等级按照 1 到 10 打分。由于当今中国普遍存在的阶层代际流动性较差的现象，“代际传递效益”在中国很明显，该变量可以在一定程度上反映个体家庭背景、社会资源等，尤其是家庭背景。处在社会上层的家庭，父母学历高、职业好，会对子女在教育、就业和职业发展上提供更多的帮助，因此判断这一变量会有显著的正向影响效应。

（5）工作类型变量

关于职业类型，CGSS 问卷中提到了“您目前工作单位的所有制形式”，针对企业的性质进行了划分：国企、集体控股、私企、中外合资、港澳台资等。本文将国企和集体控股性质的企业作为基准类，为体制内单位，其他都是市场化程度较高类型的企业，为体制外单位。该变量不仅作为重要的因变量，还将在后文进行分体制部门进行对比研究。

自变量与因变量的具体涵义如下所示：

表 1 主要变量

变量名称	变量标示	变量涵义
lnwage	小时工资对数	月平均收入除以月工作时间，再取对数
age	年龄	年龄≤60 岁
age ²	年龄平方	年龄的平方
exp	工作经验	从第一份非农工作到目前工作的工作年限
edu	教育	最高教育程度（包括目前在读）
female	性别	女性=1，男性=0
married	婚姻状况	有配偶=1，无配偶=0
CCP	政治面貌	党员=1，非党员=0
minority	民族	少数民族=1，汉族=0
class	自评阶级	1-10 打分，10 代表最顶层，1 代表最底层
health	健康状况	1-5 打分，5 代表很健康，1 代表很不健康
registered	户籍身份	非农户口=1，农业户口=0
system	工作部门	体制内=1，市场部门=0

4.4 描述性统计

4.4.1 整体与分样本描述性统计

表 2 将全部样本、流动人口样本、城市移民样本进行了样本描述，针对该统计描述可以得出：

（1）流动人口和城市移民工资性收入差异较大，流动人口的人均小时工资为 11.26 元/小时，城市移民的人均小时工资为 16.73 元/小时，相差 5.74 元/小时。值得探讨的是，这二者都是成长在农村、工作在城市，但却依然存在如此大的差异，这对二者的不同点——户籍身份影响的探讨更具备可行性；

（2）城市移民人群的平均年龄是 42 岁，而流动人口群体为 38 岁，流动人口更趋于年轻，总体均值为 40 岁，说明样本平均年龄处于中年时期；

（3）工作经验上，城市移民的平均工作年限比流动人口高出 5 年有余，由于该项反映的是第一份非农工作到目前工作的年限，说明城市移民从事了更长时间的非农工作；

（4）从受教育程度来看，流动人口的平均受教育程度处于初中与职高之间，普遍偏低，城市移民则在中专与技校之间，二者差异较为明显；

（5）总样本、流动人口样本、城市移民样本的平均阶层水平趋于一致，说明本文选取的外来人口样本质量较好。对于同样生长在农村、工作在城市的群体来说，平均阶层水平差异不大，4.2 的均值说明都在中等稍偏下的水平。

表 2 主要变量的描述性统计

变量	全部样本	流动人口	城市移民
	均值（标准差）	均值（标准差）	均值（标准差）
wage	13.813（23.234）	11.259（21.710）	16.734（24.552）
lnwage	2.099（1.031）	1.876（1.024）	2.353（0.980）
age	40.104（10.971）	38.527（11.339）	41.939（10.231）
age ²	1728.678（890.433）	1612.840（901.617）	1863.431（858.138）
exp	14.065（9.641）	11.294（8.484）	16.798（9.934）
edu	5.572（3.020）	4.548（2.332）	6.762（3.283）
female	0.499（0.500）	0.511（0.500）	0.483（0.500）
married	0.842（0.365）	0.807（0.395）	0.882（0.323）
CCP	0.120（0.325）	0.053（0.224）	0.198（0.399）
minority	0.077（0.266）	0.081（0.273）	0.071（0.257）
class	4.210（1.728）	4.114（1.756）	4.322（1.688）
health	3.840（0.989）	3.864（1.008）	3.811（0.967）
system	0.183（0.387）	0.075（0.264）	0.309（0.462）
N	2676	1439	1237

4.4.2 按体制内外分组的描述性统计

表 3 是按照单位所有制类型：体制内和体制外分组进行的描述性统计，可见，大部分外来人口样本都工作于体制外单位，也就是市场化程度比较高的单位，为 2186 人，而体制内单位只有区区 490 人。进一步分析可以得出：

（1）体制内的人均小时工资为 16.19 元/小时，体制外单位的人均小时工资为 13.08 元/小时，所以体制内平均收入水平大于体制外单位；

（2）体制内单位的平均受教育程度远远高于体制外单位，其目前平均最高教育程度为技校至专科教育之间，而体制外的平均受教育程度仅为职高和普通高中，反映了体制内单位对劳动者教育水平的更高要求；

（3）从我们关注的户籍身份分布来看，体制外单位中拥有城市户籍的外来人口比例较总样本比例稍低，而体制内单位的外来人口中，拥有城市户籍身份的比例几乎是体制外单位的两倍之多。至于分体制内、外单位后，户籍制度对工资收入影响有没有不同、有何不同，将在下文实证部分进一步探讨；

（4）体制内、外单位的年龄、性别、阶层水平、健康状况等均没有显著差异，但党员比例一项，体制内单位约为体制外单位的三倍，表现出市场化程度更高的企业，党员的分布比例更少。

表 3 按体制内外分组的描述性统计

变量	全部样本	体制内	体制外
	均值（标准差）	均值（标准差）	均值（标准差）
wage	13.813（23.234）	16.186（19.802）	13.077（24.158）
lnwage	2.099（1.031）	2.465（0.798）	1.984（1.069）
age	40.104（10.971）	39.747（9.295）	40.184（11.313）
age ²	1728.678（890.433）	1666.045（754.985）	1742.717（917.625）
exp	14.065（9.641）	16.608（10.290）	13.118（9.214）
edu	5.572（3.020）	8.045（3.284）	5.017（2.660）
female	0.499（0.500）	0.418（0.494）	0.516（0.500）
married	0.842（0.365）	0.873（0.333）	0.834（0.372）
CCP	0.120（0.325）	0.267（0.443）	0.087（0.282）
minority	0.077（0.266）	0.080（0.271）	0.076（0.265）
class	4.210（1.728）	4.539（1.723）	4.136（1.721）
health	3.840（0.989）	3.994（0.873）	3.805（1.010）
registered	0.462（0.499）	0.780（0.415）	0.391（0.488）
N	2676	490	2186

5.实证结果与分析

正式回归之前，对于设定的模型，本文首先进行了多重共线性检验，即 Stata 里的 VIF 检验。结果显示，变量最大的 VIF 值为 2.20，平均 VIF 值为 1.38，故排除了选取的变量之间可能存在的严重多重共线性假设，对回归结果的有效性不造成影响。此外，由于本文使用的是截面数据，该数据类型可能出现的最大问题便是异方差，为了修正可能出现的异方差干扰因素，本文在 Stata 回归中采用异方差稳健标准误。

5.1 代理变量回归分析

5.1.1 获得代理变量的 Logit 回归结果

我们首先用多元线性回归模型考察户籍身份对工资性收入的影响，但在多元回归之前，我们需要得到模型所需的代理变量——倾向得分 `pscore`，也即获得城市户籍的概率。Logit 回归结果如表 4 所示。我们选取了年龄、年龄平方、性别、工作年限、受教育程度、婚姻状况、政治身份、民族、健康状况、阶层水平这 10 个维度来进行处理组与控制组之间的匹配，用 Stata 做出此回归结果后，原数据中自动生成倾向得分 `pscore`，只需在接下来的多元线性回归中加入估计式（1）作为代理变量即可。

表 4 获得倾向得分 `pscore` 的 Logit 回归结果

变量名	系数	标准差	发生比
age	0.175 ***	0.052	0.001
age2	-0.002 ***	0.001	0.007
exp	0.052 ***	0.008	0.000
edu	0.379 ***	0.025	0.000
female	0.475 ***	0.124	0.000
married	0.365 ***	0.194	0.060
CCP	0.449 ***	0.194	0.021
minority	0.122	0.243	0.614
health	-0.110 **	0.066	0.097
class	0.009 **	0.005	0.801
_cons	-7.283 ***	1.068	0.000
R-squared	0.235	Observations	2676

注：其中***、**、* 分别代表在 1%、5%、10%水平上显著

5.1.2 多元线性回归结果

多元线性回归结果如表 5 所示。因变量为小时工资对数，模型 1 为未加入代理变量获得城市户籍概率 **pscore** 的回归结果，模型 2 是加入了代理变量的回归结果。回归显示，年龄、工作经验、受教育水平、健康状况、阶层水平等都对外来人口工资性收入有正向影响效应，并在 1%的水平下显著，而性别变量由于代表的是“女性=1”的虚拟变量，对收入影响是显著的负效应。加入代理变量后发现，代理变量 **pscore** 在 5%的水平下显著，说明此“能力”代理变量是有效的。

模型 2 可以看出，年龄平方项在 1%的水平下显著为负，说明年龄对工资性收入呈倒 U 型的影响趋势。在外来人口群体中，受教育程度每提高一级，工资性收入将上涨 11.1%。工作年限对收入也有显著正向影响，外来人口的工作年限每增长 1 年，工资性收入将增加 1.3%。对健康状况而言，健康的收入回报率是 5.5%，并在 5%的水平下显著。而阶层水平对于外来人口收入的正向效应都大于前面三者，阶层水平从 1-10 每提高一级，对工资性收入的提高更是达到了 12.4%，统计意义上非常显著。这验证了我们在变量选取时的正向效应预期，阶层可以是家庭背景、社会资源的反映，阶层越高，越对阶层中个体的就业、收入有所帮助，可能是父母学历高、素质高，能为子女提供更好的教育条件，从而增长人力资本投资，也可能是父辈有能力、有资源，从而能为子女在职业发展上提供更多的帮助。可以想象，由于家庭环境好与阶层高导致的后代工作、收入有显著优势，又进一步造成后代步入好的阶层，后代便又有能力为他的下一代打造好的阶层条件，这映射出的便是“代际传递效益”。外来人口本就在“资源”、“关系”禀赋上有一定的匮乏，想要跳出阶层固化这一圈子便是相当有难度的。

再重点关注户籍身份变量的影响，加入获得城市户籍身份概率的代理变量后，结果显示，在 5%的水平下，控制其他因素不变，城市移民（处理组）比流动人口（控制组）的每小时工资性收入要高 4.1%，说明户籍效应较为显著。我们知道，获得城市户籍的概率 **pscore** 是作为“能力”的代理变量，加入“能力”代理变量后户籍效应的影响略为降低，**pscore** 自身的回归结果也表明，“能力”的确对工资性收入产生了正向影响。这个结果背后有着深层次的含义，通俗以为，户籍制度本身没有影响，只是不同户籍身份的个体成长环境不同，接受的教育质量等不一致，形成的能力也不一致，自然导致收入水平不同。但我们的回归结果表明，当控制了“能力”因素的影响后，户籍制度仍然在 5%的置信水平下显著，致使没获得城市户籍的外来人口工资要比获得城市户籍的外来人口工资低 4.1%，那么说明户籍制度本身便对流动人口群体有负向收入效应，而不是完全通过能力

来反映。这一结论具有一定的政策意义，让我们知道改革户籍制度将是对流动人口有益的。

表 5 ols 与代理变量回归结果

变量	模型1 lnwage	模型2 lnwage
registered	0.043** (0.021)	0.041** (0.019)
age	0.045*** (0.017)	0.042** (0.020)
age2	-0.001*** (0.000)	-0.001*** (0.000)
exp	0.014*** (0.003)	0.013** (0.005)
edu	0.119*** (0.008)	0.111*** (0.030)
female	-0.338*** (0.042)	-0.348*** (0.053)
married	-0.064 (0.072)	-0.072 (0.075)
CCP	0.003 (0.059)	0.005 (0.064)
minority	-0.372*** (0.090)	-0.374*** (0.090)
health	0.052** (0.023)	0.055** (0.025)
class	0.124*** (0.012)	0.124*** (0.012)
_pscore		0.121** (0.055)
Constant	0.158 (0.329)	0.237 (0.423)
Observations	2,676	2,676
R-squared	0.325	0.326

注：其中***、**、* 分别代表在 1%、5%、10%水平上显著，以上均为 Stata 消除异方差回归结果

5.2 倾向得分匹配回归分析

5.2.1 匹配检验

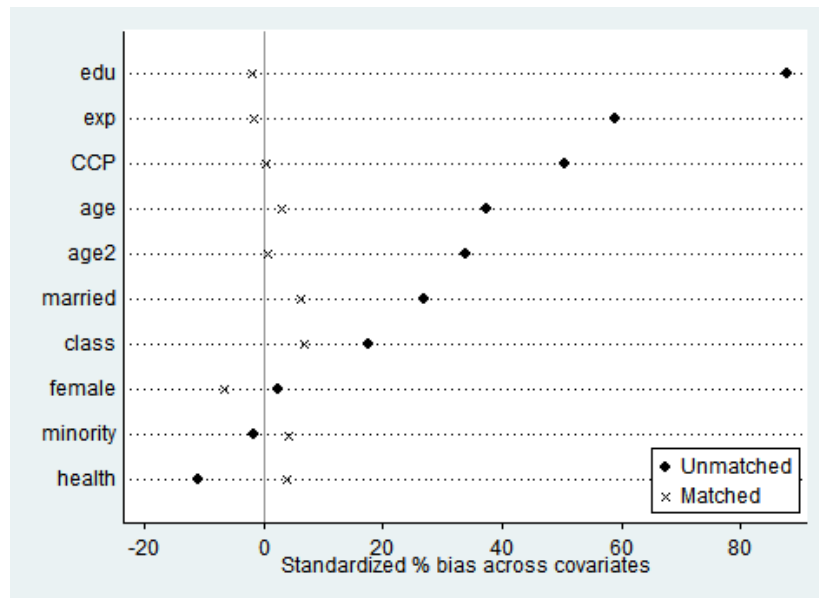
首先进行匹配检验，来考察其结果是否有效地平衡了我们的数据。表 6 反映的是匹配前后每个变量的标准误变化状况，可以看出，无论是对于处理组还是控

制组，匹配前绝大多数变量的均值和标准误都较大（female 和 minority 除外），从 t 值上看几乎都有统计意义上显著的差异，而匹配后上述均值和标准误却都有一定程度的缩小。从图 1 可以直观地看出来，除 minority 和 health 外，其他变量的标准化偏差都在匹配后缩小了，这就使得两组之间的均值更加接近。且匹配后大多数变量的标准化偏差都小于 5%（female、married 和 class 除外），所有变量的标准化偏差都小于 10%，说明处理组和控制组的个体特征差异得到了较好的控制。再从匹配后的 t 检验结果来看，所有的 t 检验结果都不拒绝原假设——处理组与控制组没有系统差异，那么就可以得出，模型通过匹配检验，处理组与控制组间数据有被较好地平衡。

表 6 变量误差削减结果

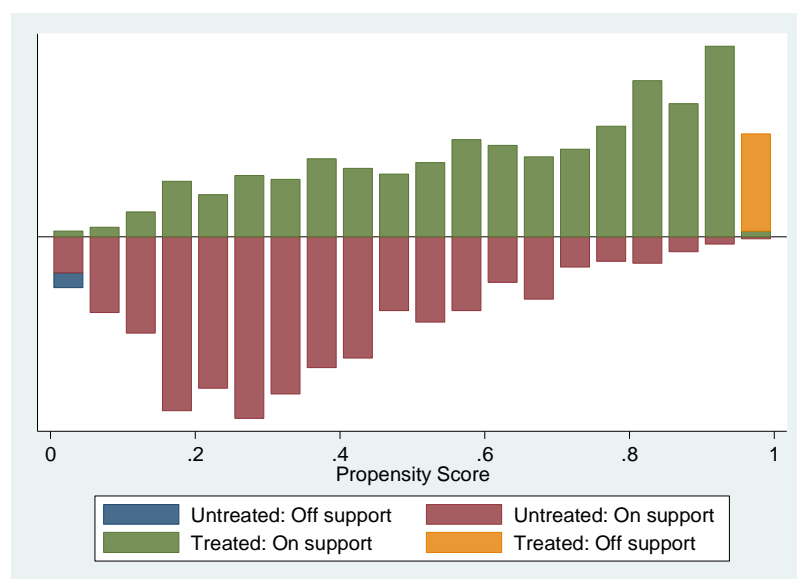
变量名	样本	均值		标准偏误 (%)	误差削减 (%)	t-test	
		处理组	控制组			t 值	P> t
age	匹配前	40.504	36.908	37.5	92.4	7.66	0
	匹配后	40.105	39.831	2.8		0.57	0.57
age2	匹配前	1721.800	1464.700	33.9	98	6.92	0
	匹配后	1690.400	1685.300	0.7		0.13	0.897
exp	匹配前	16.922	11.529	58.8	97.3	12.02	0
	匹配后	16.377	16.525	-1.6		-0.29	0.773
edu	匹配前	7.385	4.912	87.8	97.8	17.96	0
	匹配后	7.118	7.172	-1.9		-0.35	0.727
female	匹配前	0.400	0.389	2.3	-221.7	0.48	0.632
	匹配后	0.400	0.446	-7.4		-1.44	0.066
married	匹配前	0.900	0.807	26.8	76.9	5.47	0
	匹配后	0.894	0.872	6.2		1.34	0.181
CCP	匹配前	0.241	0.064	50.6	99.3	10.35	0
	匹配后	0.206	0.204	0.4		0.06	0.95
minority	匹配前	0.064	0.068	-1.8	-161.1	-0.37	0.713
	匹配后	0.063	0.043	4.7		1.09	0.091
health	匹配前	3.926	4.026	-11.1	67	-2.27	0.024
	匹配后	3.926	3.893	3.7		0.73	0.463
class	匹配前	4.442	4.149	17.4	61.5	3.57	0
	匹配后	4.391	4.052	6.7		1.14	0

图 1 各变量的标准化偏差图示



为了证明 PSM 在接下来实证回归中的可行性和合理性，我们不仅要检验处理组与控制组的匹配效果，还要考察匹配过程对样本容量的影响，看是否会造成样本大量损失。图 2 表示的是倾向得分的共同取值范围，反映的便是样本匹配后的遗失情况。图中红色与绿色条形图代表的是共同取值范围（on support），我们可以直观地看出，绝大多数的观测值都在共同取值范围内，只有少数样本不在其中。所以在进行倾向得分匹配时，仅仅会损失少量样本，也进一步说明了样本的匹配效果比较好。

图 2 倾向得分的共同取值范围



5.2.2 倾向得分匹配估计结果

通过第四部分的基本描述性统计分析,可知城市移民与流动人口之间确实存在工资性收入的差异,二者的人均小时工资相差 5.74 元/小时。接下来便进行 PSM 估计,具体地,分别使用最近邻匹配、卡尺范围为 0.01 的半径匹配、卡尺范围为 0.05 的半径匹配以及核匹配三种方法,并在使用每一种匹配方法同时利用自抽样法(Bootstrap)反复抽样估计样本的标准误,所得到的结果如表 7 所示。

表 7 样本总体的平均处理效应

PSM 匹配方法	变量名	处理组	控制组	ATT	标准误	t 值
最近邻匹配	lnwage	2.331	2.153	0.178**	0.086	2.05
半径匹配 (Caliper=0.01)	lnwage	2.319	2.183	0.135**	0.061	2.21
半径匹配 (Caliper=0.05)	lnwage	2.330	2.242	0.088*	0.046	1.89
核匹配	lnwage	2.331	2.242	0.089**	0.041	2.17

注: **、*、* 分别代表在 1%、5%、10%水平上显著;处理组为获得非农户籍的外来人口样本,控制组为农业户籍的外来人口样本;标准误采用自抽样法(Bootstrap)反复抽样 500 次估计所得;Stata 软件中,各 psm 方法依次可通过 neighbor、radius caliper、kernel 选择项实现。

上表显示的结果是分别对外来人口总体样本运用最近邻匹配法、卡尺范围为 0.01 的半径匹配法、卡尺范围为 0.05 的半径匹配法以及核匹配法进行匹配后,估计得出的平均处理效应 ATT 结果。可以看出,采用这三种方法,匹配后的 ATT 均在 5%和 10%的水平下显著,说明户籍制度对外来人口收入的影响确实是显著的。并且可以从表中得出,在我们用倾向匹配得分法(PSM)有效控制样本选择性差异后,户籍制度的作用仍然是统计意义上显著的,且三种匹配方法显著水平都为 5%。

以最近邻匹配法为例,它用已估计出的 pscore 值为基础,为处理组样本寻找到了 pscore 值最相近的控制组样本,匹配之后,处理组(城市移民)的工资性收入为 2.331,在 5%的显著性水平上高于控制组(流动人口)的工资性收入 2.153。ATT 的显著性程度较好,它表示的是城市户籍身份使得城市移民收入高出流动人口 17.8%,也就是户籍制度的作用结果。换言之,最近邻匹配方法得出的结论是,户籍制度的存在对于流动人口的收入有显著的负向影响,它使得流动人口的小时收入降低了近 17.8%。

由于不同的倾向得分匹配方法间存在差异,得到的实证结果也可能不尽相同。在卡尺范围为 0.05 的半径匹配法和核匹配法下,户籍制度使得获得了城市

户籍身份的外来人口小时工资收入分别提高了 13.5%和 8.9%，且统计意义上都在 5%的水平下显著。

最后，从每种方法的平均处理效应 ATT 值来看，都比前面的多元线性回归和代理变量回归得到的值要高，这说明样本的选择性偏误明显影响了估计结果，倾向得分匹配法（PSM）得到的回归结果更为可靠。这里得出的结论也与谢嗣胜及姚先国（2006）、魏万青（2012）等的研究发现一致。谢嗣胜、姚先国（2006）对户籍制度造成的城乡工人收入差距解释范围是 20%—30%，而魏万青（2012）研究得出，外来人口获得城市户籍身份要比同样外来务工的流动人口收入提高 39.8%。本文得到的为 17.8%，可能因为方法上的原因，或者本文使用的数据更新，处于户籍制度加大改革阶段，所以得到的对收入影响程度与已有文献有所不同。

5.3 体制内、外部门分组回归分析

以上讨论的是户籍制度本身对整体外来人口的工资性收入影响，为了探讨市场化进程下户籍制度对外来人口工资性收入的影响效果有何不同，本文将进一步将外来人口样本按照体制内、体制外单位分组，继续探讨户籍制度对不同工作性质人口工资性收入的影响，并分别用最近邻匹配、卡尺范围为 0.01 的半径匹配和核匹配三种方法估计 ATT。对比体制内和体制外部门，体制内部门显然拥有更多的资本流动和更市场化的竞争，也即市场化程度会更高。两个部门中户籍收入效应的估计结果如表 8 所示。

表 8 按体制内、外分组的样本平均处理效应

工作性质	PSM 匹配方法	变量名	处理组	控制组	ATT	标准误	t 值
体制内	最近邻匹配	lnwage	2.723	2.506	0.217*	0.115	1.89
	半径匹配 (Caliper=0.01)	lnwage	2.567	2.504	0.063	0.101	0.62
	核匹配	lnwage	2.583	2.504	0.079	0.106	0.75
体制外	最近邻匹配	lnwage	2.188	2.060	0.127**	0.061	2.08
	半径匹配 (Caliper=0.01)	lnwage	2.160	2.036	0.124*	0.067	1.86
	核匹配	lnwage	2.188	2.099	0.089*	0.046	1.93

注：**、*、* 分别代表在 1%、5%、10%水平上显著；处理组为获得非农户籍的外来人口样本，控制组为农业户籍的外来人口样本；标准误采用自抽样法（Bootstrap）反复抽样 500 次估计所得；Stata 软件中，各 psm 方法依次可通过 neighbor、radius caliper、kernel 选择项实现。

从表 8 可以看出，对于体制内单位，只有用最近邻匹配法估计后，城市户籍身份对外来人口的平均处理效应（ATT）在 10%的置信水平下显著，说明体制内单位户籍制度的对外来人口的收入负向影响作用较为显著。而用半径匹配和核匹配的结果不显著，分析可能存在的原因是体制内的样本量较少，导致样本标准误差偏大所造成。在最近邻匹配法下，体制内的城市移民比流动人口小时工资收入高 21.7%，这一结果比上一节总体回归结果中得平均处理效应 ATT 值更高，得出的结论是，在体制内单位中，若消除户籍制度的影响，可以使流动人口的小时收入提高近 21.7%，高于对整个外来人口的工资性收入拉动。

而对于体制外单位，可以看出，无论是采用哪种匹配方法，匹配后的 ATT 均在 5%和 10%的水平下显著，说明在市场化程度普遍比较高的体制外单位，户籍制度对外来人口工资性收入的负向效应也是有统计意义上显著影响的。也以最近邻匹配的结果为例，体制外样本的 ATT 要比总样本的 ATT 低，体制外的城市移民比流动人口小时工资收入仅高 12.7%，也就是说改革户籍制度可以使体制外从业的流动人口小时收入提高近 12.7%，小于对整个外来人口的工资性收入拉动。

对比体制内和体制外单位，可以得出，越是市场化程度高的企业，受到户籍制度的限制就越小。所以，从总体上看，户籍制度对外来人口收入的平均处理效应（ATT）是显著的，但是分不同市场化程度的部门，ATT 表现出差异性。对于体制内单位，户籍制度对外来人口的 ATT 只在最近邻匹配下显著，在另两种匹配方法下并不显著。而对于体制外单位，户籍制度对外来人口的 ATT 在三种匹配方法下都是显著的。随着市场化程度的提高，户籍制度对收入的限制呈减小趋势，若是消除户籍制度的影响，受益对多的是体制内单位的流动人口，其收入正向效应为 21.7%，其次是总体流动人口收入提高 17.8%，影响最小的是市场化程度高的体制外单位流动人口，收入提高仅 12.7%。

体制内单位（多为国企）就业歧视、户籍限制问题，也在社会上引起广泛关注。中国政法大学宪政研究所做过一项调查，结果显示 59.14%的用人单位对求职者，特别是刚毕业的大学生求职者提出了明确的户籍和地域要求⁴，这种不合理的要求严重限制人才的合理流动。该项调查还指出，存在户籍歧视的不同性质企业中，国企最严重，民企比国企稍好，合资企业和外资企业在所有单位中户籍歧视相对较轻，这也进一步印证了本文的观点。而针对此问题，我国在不断采取措施来改进，国资委办公厅于 2016 年 6 月印发了《关于做好 2013-2014 年国有

⁴ 本数据来自腾讯网 <http://news.qq.com/a/20100727/000080.htm>

企业招收高校毕业生工作有关事项的通知》，通知中便有要求国有企业积极履行社会责任，对不论何种来源和形式的就业歧视持坚决反对，严禁在就业市场上的招聘中违反相关规定对性别、户籍等采取准入门槛和差别对待，为吸纳更多就业者。诚然，高校大学生群体里的“外来人口”是能力、素质都较高的群体，正如吴晓刚（2008）在研究中指出的一样，那些从小生长于农村，通过受教育、升学等路径摆脱农村户口身份限制的外来人口，是流入地城市的“精英”。然而这些“精英”尚且受到户籍制度的限制，特别是在体制内单位，更别说那些未经过“极度正向选择”，永远带着“农民工”身份在城市从事着工作环境差、强度大、要求低的工作，也面临流动性差的现实，而这种流动性限制，不可忽视的源头便是户籍限制。

5.4 小结

本章首先使用 Logit 模型估计倾向得分值，再作为代理变量代入多元线性回归模型，得到多元线性回归下的户籍制度对收入的负向效应，结果显示，城市移民（处理组）比流动人口（控制组）的每小时工资性收入要高 4.1%。然而上述结果只是证明了户籍制度对外来人口的工资性收入有影响，不能克服样本的选择性偏误。为了得到更为准确的户籍制度平均处理效应（ATT），本文接下来用最近邻匹配法、卡尺范围为 0.01 的半径匹配法、卡尺范围为 0.05 的半径匹配法以及核匹配法估计，并用自抽样法反复抽样 500 次来估计样本的标准误。在匹配之前进行了匹配检验，证实了模型是有效的，处理组与控制组数据被有效地平衡了。匹配后发现，无论用哪种匹配方法，平均处理效应 ATT 均在 5%和 10%的水平下显著，以最近邻匹配方法得出，户籍制度的存在对于流动人口的收入有显著的负向影响，它使流动人口的小时收入降低了近 17.8%。进一步地，本文还分体制内、体制外单位探究了不同市场化程度下户籍的收入效应有何不同，发现越是市场化程度高的企业，受到户籍制度的限制就越小，国企等体制内单位中存在的户籍歧视最为严重。

6. 结论、政策建议及本文的不足

6.1 主要结论

随着我国市场经济体系的不断完善，劳动力市场越来越趋向一体化。传统计

划经济条件下遗留的户籍制度本身就存在着不合理性和不平等性,为实现劳动力市场一体化,我国从 20 世纪 80 年代中期便逐步实行户籍制度改革。我国的城镇化进程目前正是属于关键时期,户籍制度改革也被推向高潮。2014 年 7 月 30 日,备受关注的国务院《关于进一步推进户籍制度改革的意见》正式公布,随后两年内,根据各地实施意见,我国多个省份已划定新型户籍制度成型期限,甚至有的省份规定的新型户籍迁徙条件相较于国家出台的更为宽松。然而,户籍制度的推行、改革、创新演变,初衷是为外来人口捍卫福利,提高他们的就业流动性,至于起没起效果,有哪些效果,评价指标最终还是要落实到是否真正有助于他们的收入增长。所以,在评价户籍改革的政策效果时要有依据可循,本文便是用较新的具有权威性的我国全国性调查数据库,来实证探究户籍制度改革在统计意义上的有效程度。

本文使用中国综合社会调查(CGSS)2012 年数据,选取的全是外来人口样本,根据实证研究的需要对样本数据进行整理,最后使用了 2676 个样本观测值,其中处理组(城市移民)1237 个观测值,控制组(流动人口)1439 个样本观测值。首先,通过基本的描述性统计分析可以看出,拥有城市户籍身份的外来人口个体比没有城市户籍身份的外来人口个体收入高,并且对整个外来人口,体制内平均收入水平大于体制外单位。接着,本文运用多元线性回归、倾向得分匹配方法(PSM)来具体研究户籍收入效应。在使用多元线性回归时加入了倾向得分值 $pscore$,也即获得城市户籍的概率,作为“能力”的代理变量。为了排除因样本选择产生的偏误,本文最主要是运用倾向得分匹配方法(PSM),来获得平均处理效应(ATT)。在匹配前进行了匹配检验,通过检验后运用最近邻匹配、半径匹配和核匹配三种方法估计户籍制度对收入的 ATT,并且对每一种匹配方法都辅以自抽样法(Bootstrap)反复抽样 500 次,得到准确标准误下的户籍收入效应。最后,本文还按照市场化程度不同,对就业单位按分体制内、体制外分类,进一步市场化进程对户籍收入效应有何影响。所得到的结论如下:

(1) 户籍制度对外来人口工资性收入存在负向影响

户籍制度对外来人口工资性收入存在显著的负向效应,已成为无论用普通多元线性回归还是倾向得分匹配法(PSM)实证研究得出的不争事实。加入获得城市户籍身份概率($pscore$)的代理变量后多元线性回归表示,在 5%的显著水平下,城市移民(处理组)比流动人口(控制组)的每小时工资性收入更高。由于获得城市户籍的概率 $pscore$ 是作为“能力”的代理变量,而通常上,认为户籍制度不会直接影响收入的观点认为:户籍制度本身没有影响,只是不同户籍身份

的个体成长环境不同，接受的教育质量、文化观念等不一致，形成的能力也不一致，导致收入水平不同。但本文的回归结果表明，当控制了“能力”因素的影响后，户籍制度的收入效应仍然显著，说明户籍制度本身便对流动人口群体有负向收入效应，而非单纯以“能力”为介质来反映。

那么对外来人口群体来说，户籍制度具体对他们的工资性收入造成了多大的影响呢？这并不能直接使用多元线性回归的结果，因为直接利用 *ols* 进行比较会造成样本选择偏误，我们使用倾向匹配得分法（PSM）来克服这种偏误，得到所需的准确数值。将外来人口样本两组：处理组——具有城市户籍的外来人口，控制组——没有城市户籍的外来人口。运用倾向得分匹配法的思想是：通过找到与处理组样本特征变量类似的控制组，将二者进行收入差异对比。由于两组样本特征变量类似，便可以将他们之间的收入差异归于户籍制度的影响。分别通过最近邻匹配、半径匹配和核匹配的估计方法来得到户籍制度对收入的 *ATT* 后，发现无论哪种方法均在 5% 和 10% 的水平下显著，以最近邻匹配方法的结论，户籍制度使得城市移民收入高出流动人口 17.8%，也就是说，户籍制度的存在对于流动人口的收入有显著的负向影响，它使得流动人口的小时收入降低了近 17.8%。换言之，如果得以摒除户籍制度带来的影响，那么我们常见的“农民工”群体的人均小时收入将上涨 17.8%。可见户籍制度的限制是不容忽视的，从统计意义上证明了户籍制度改革是必要的。

（2）户籍制度的收入效应随市场化进程减弱

探究了户籍制度对整体外来人口的收入负向效应，本文又结合中国当前国情——市场经济体系不断完善，市场化程度不断提高，进一步按单位性质进行分类，用倾向得分匹配法（PSM）对比体制内和体制外单位户籍制度效应的差异。实证发现，首先，户籍制度对外来人口收入的负向效应，分不同市场化程度的部门是不同的，表现在平均处理效应（*ATT*）存在差异性。体制内单位的估计结果没有体制外单位显著，体制内单位只有在最近邻匹配下显著，相应结论为城市移民比流动人口小时工资收入高 21.7%；而对于体制外单位，户籍制度对外来人口的平均处理效应（*ATT*）在三种匹配方法下都是显著的，以最近邻匹配的结果来解释，此时城市移民比流动人口小时工资收入仅高 12.7%。两者的不同也可能侧面反映出我国体制内、体制外单位的收入本身就存在着一定的差异。

进一步分析得出，越是市场化程度高的企业，受到户籍制度的限制就越小。顺应着我国市场化进程，户籍制度对收入的限制呈减小趋势。总的来说，若是消除户籍制度的影响，受益对多的是体制内单位的流动人口，其收入正向效应为

21.7%，其次是总体流动人口收入提高 17.8%，影响最小的是体制外单位流动人口，收入提高 12.7%。这也反映出我国市场化进程本身便对户籍歧视有抑制作用，在市场化程度更高的部门，越存在公平竞争，看中的越为劳动者的能力、才干，个体可施展的空间大，机会多，流动性也更不受限制。反而在国企这样的体制内单位，存在更多的计划经济体制下遗留下来的户籍限制、户籍歧视问题。

（3）户籍歧视存在的原因值得关注

户籍制度为什么对外来人口收入的平均处理效应是显著的，其背后的作用机理值得关注。广大学者把研究所得的户籍制度效应称为户籍歧视(韦伟、傅勇，2004；王美艳，2005 谢嗣胜、姚先国，2006 等)，而歧视作为一种具有普遍性的社会问题，是社会经济、政治、文化、历史等各种因素综合作用而形成的结果，看待户籍歧视也要结合经济学角度和社会学视角。

本文并没有深层次挖掘来自户籍歧视的原因是什么，除了表面上制度方面本身的原因，控制了社会资源的配置之外，我国劳动力市场上供过于求的现象也不容忽视。供过于求使得就业机会有限，企业的用人标准相应被抬高。但需要注意的是，供过于求不应该成为户籍歧视的正当原因，就业岗位越有限，反而越应该强调获得就业岗位的竞争公平性。源于制度最基本的原因还是文化观念的歧视与不认同，社会上普遍对“农民工”身份蕴含着的不公正的价值观念，对“农民工”一直带有集体排他性的歧视与偏见。观念的根深蒂固，是最难突破的屏障。

诚然，户籍制度改革不是一朝一夕，而是一个复杂和艰巨的过程，但是目前户籍制度导致劳动力市场上外来人口（特别是农民工群体）依然遭受着就业和收入的差别待遇，既有损于效率，也更有损于社会公平，我们必须正视和重视这一制度带来的不公，明白只有消除户籍歧视，才能使我国更进一步接近劳动力市场一体化的实现目标。

6.2 政策建议

根据实证结果，本文出具以下政策建议：

首先，户籍制度确实不利于流动人口的待遇，实际的措施应当是改变户籍制度，打破户籍制度造成的人口之间的不流动产生的资源配置之不合理。如果不能改变户籍制度对资源的不公平配置，很难保证城市中的流动人口和迁入人口不受工资上或其它形式的歧视。针对改革，户籍制度改革的力度和进程理应依不同的城市和地区而有所差别。

其次，由于本文的结果是体制内的工作更易受到户籍的影响，而相比较却较

少看着个体的能力。因此政府更应该加大对国有、集体企业的改革管制，打破这种对个体的户籍歧视，使资源得到合理配置。随着我国市场进程加快，户籍歧视得到削弱，也说明我国应全面保障市场经济的健康发展和公平竞争。

最后，外来人口和城市本地居民会受教育、医疗、观念等影响。所以从根本上来讲，政府更应该注重城市和农村之间的教育、医疗服务、福利、公共服务等资源分配的不均衡，大力加强我国农村的基础设施建设，从根源上解决这种差距，来保证我国健康持续的发展。但是由于涉及到决策者对于公共物品市场认识的根本性转变，这些措施恐怕难以在短期内奏效。改革需要全社会的共同关注，而我们每个公民则可以从改变陈旧观念上来重新认识流动人口从业现象和问题。

6.3 本文的进步

关于户籍制度的研究，以往学者常常直接比较流动人口与城市居民，认为此差距即农村与城市户籍制度带来的差距，但这样往往导致样本选择性偏误。流动人口与城市居民存在的差异除了可观测到的禀赋差异（主要为人力资本差异）和由户籍制度导致的差异外，还存在着成长氛围、文化观念、医疗条件、教育质量等一系列不可观测的因素产生的影响。此外，大部分研究没有区分地域歧视的影响。国内外学界都有意识到对来自其他地区的居民的歧视文化对其工资收入有一定的影响，但如果不区分户籍制度和地域歧视的影响效果，就等同于间接认为地域歧视完全是户籍制度造成的，这显然是牵强而不符合事实的。为了控制这些不可观测因素的差异，并且控制地域歧视的影响，本文在城市居民的选择上只选取经历了“农转非”的样本，与流动人口对比，这样就保证了二者绝大部分都成长于农村、工作在城市，实证回归出来的结果依然是显著的，得到的户籍制度的收入效应数值也比以往的文献得出的数值要小。例如姚先国和谢嗣胜（2006）分析浙江省 2003 年和 2004 年的调查数据发现农民工与城镇职工工资水平中可以被户籍制度解释的力度有 55.2%；邓曲恒(2007)利用 CHIPs2002 年的数据，比较城市本地职工和农名工之间的收入，发现他们收入差异的 60%可以归结于歧视，以上这些文献都是直接比较流动劳动力与城市本地劳动力。本文即得出户籍制度使得城市移民收入高出流动人口 17.8%，是控制了样本选择偏误的结果。

其次，在研究方法上，大多数文献都是基于 Blinder 和 Oaxaca 的工资差异分解模型，该方法虽然可以有效控制部分变量，但是会产生“指数基准”问题（魏万青，2012），利用了不同的基准指数，会得到不同的回报结果——这就是大量学者利用这种方法测量“歧视”的来源，却得出不同结论的原因。并且，由于户

籍身份不是随机分配给外来人口的，而是样本个体自己选择的结果，因如果对这样的非随机数据资料进行简单地分组回归分析，会出现估计偏误问题。通常解决最小二乘法估计偏误问题有工具变量法和匹配法等，所以本文便运用 Rosenbaum 和 Rubin (1983)提出的倾向得分匹配方法(PSM)来研究。得到的估计结果效果也较好，用最近邻匹配法、卡尺范围为 0.01 的半径匹配法、卡尺范围为 0.05 的半径匹配法以及核匹配法均在统计意义上显著。

诚然，本文也有进一步需要改进的地方：首先，本文利用的 CGSS 数据有一定缺失值，因此结论会有一定的偏差。较少的样本量不利于更深层次的研究，比如，要想用倾向得分匹配进行分层回归变得困难。后续研究将寻找更新、更全面的数据来保障不受样本量的影响。第二，本文在研究外来人口劳动力市场的户籍工资差异时，仅在以“农业”和“非农”的户口类别作为划分标准，忽略了户口所在地（本地户口和外地户口）这一维度可能存在的影响，并且没有考虑户口的交互作用。进后续将考虑把样本分为本地城镇户口、本地农村户口，外地城镇户口、外地农村户口这四大群体，并检验户口类别和户口所在地这两种户籍属性的交互影响效应。最后，本文在研究过程中已经发现，不同年龄段人口的工资收入均值并非存在理论认为的变化关系，而是有一定的波动性。并且本文未考虑时间因素带来的变化。后续除了按年龄分层进行进一步研究外，还可以尝试加入一些历史、时间因素（如使用面板数据），以控制其对代际人口的总体影响。针对这些不足，笔者会继续探索方法，以期改进。

参考文献

- [1] Chan K W, Zhang L. The Hukou System and Rural-Urban Migration in China: Processes and Changes*[J]. J Child Adolesc Subst Abuse, 1999, 160(160):818.
- [2] Chan K W. Cities with Invisible Walls: Reinterpreting Urbanization in Post-1949 China[M]. 1994.
- [3] Duckett J. The Paradox of China's Post-Mao Reforms, by Merle Goldman; Roderick MacFarquhar[M]// The paradox of China's post-Mao reforms /. Harvard University Press, 1999.
- [4] Becker G S. The economics of discrimination / Gary S. Becker.[J]. 1971.
- [5] Meng X, Zhang J. Occupational Segregation and Wage Differentials between Urban Residents and Rural Migrants in Shanghai[J]. Journal of Comparative Economics, 2001, 29(3):485-504.
- [6] Becker S O, Ichino A. Estimation of average treatment effects based on propensity scores[J]. 2002, 2(4):358-377.
- [7] Blinder A S. Wage Discrimination: Reduced Form and Structural Estimates[J]. Journal of Human Resources, 1973, 8(4):436-455.
- [8] Cotton J. On the Decomposition of Wage Differentials[J]. Review of Economics & Statistics, 1988, 70(2):236-243.
- [9] Oaxaca R. Male-Female Wage Differentials in Urban Labor Markets[J]. International Economic Review, 1971, 14(3):693-709.
- [10] Solinger D J. Citizenship Issues in China's Internal Migration: Comparisons with Germany and Japan[J]. Political Science Quarterly, 1999, 114(3):455-478.
- [11] Zhigang L U, Song S. Rural-urban migration and wage determination: The case of Tianjin, China[J]. China Economic Review, 2006, 17(3):337-345.
- [12] Maurer-Fazio M. Differential rewards to, and contributions of, education in urban China's segmented labor markets[J]. Pacific Economic Review, 2004, 9(3):173-189.
- [13] Whalley J, Zhang S. A NUMERICAL SIMULATION ANALYSIS OF (HUKOU) LABOUR MOBILITY RESTRICTIONS IN CHINA[J]. Journal of Development Economics, 2007, 83(2):392-410.

- [14] Roberts K D. China's "tidal wave" of migrant labor: what can we learn from Mexican undocumented migration to the United States?[J]. [Unpublished] 1995, 1997, 31(2):249-293.
- [15] Cook R D, Weisberg S. The central role of the propensity score in observational studies for causal effects[J]. *Biometrika*, 1983, 70(1):41-55.
- [16] 郭震. 城镇居民和流动人口工资差距:户籍歧视还是性别歧视[J]. *南方经济*, 2013, V31(8):69-77.
- [17] 邓曲恒. 城镇居民与流动人口的收入差异——基于 Oaxaca-Blinder 和 Quantile 方法的分解[J]. *中国人口科学*, 2007, 2007(2):8-16.
- [18] 谢嗣胜, 姚先国. 农民工工资歧视的计量分析[J]. *中国农村经济*, 2006(4):49-55.
- [19] 余向华, 陈雪娟. 中国劳动力市场的户籍分割效应及其变迁——工资差异与机会差异双重视角下的实证研究[J]. *经济研究*, 2012(12):97-110.
- [20] 姚先国, 赖普清. 中国劳资关系的城乡户籍差异[J]. *经济研究*, 2004(7):82-90.
- [21] 都阳, 王美艳. 户籍制度与劳动力市场保护[J]. *经济研究*, 2001(12):41-49.
- [22] 严善平. 人力资本、制度与工资差别——对大城市二元劳动力市场的实证分析[J]. *管理世界*, 2007(6):4-13.
- [23] 刘安萍. 我国城乡收入差距问题探讨[J]. *统计与决策*, 2005(14):119-120.
- [24] 王洪亮, 徐翔. 收入不平等孰甚:地区间抑或城乡间[J]. *管理世界*, 2006(11):41-50.
- [25] 成志刚, 罗帅. 近十年我国农民工社会保障问题研究综述[J]. *湘潭大学学报(哲学社会科学版)*, 2007, 31(3):108-113.
- [26] 赵耀辉, 刘启明. 中国城乡迁移的历史研究:1949~1985[J]. *中国人口科学*, 1997(2):26-35.
- [27] 彭希哲, 郭秀云. 权利回归与制度重构——对城市流动人口管理模式创新的思考[J]. *人口研究*, 2007, 31(4):1-8.
- [28] 郭菲, 张展新. 流动人口在城市劳动力市场中的地位:三群体研究[J]. *人口研究*, 2012, 36(1):3-14.
- [29] 汪和建. 就业歧视与中国城市的非正式经济部门[J]. *南京大学学报(哲学·人文科学·社会科学)*, 1998(1):131-141.
- [30] 李路路. 透视不平等-国外社会阶层理论[M]. 社会科学文献出版社, 2002.

- [31] 李路路, 边燕杰. 制度转型与社会分层, 基于 2003 年全国综合社会调查[M]. 中国人民大学出版社, 2008.
- [32] 何芸. 二元分割与行业收入不平等——基于二元劳动力市场分割理论的分析[J]. 经济问题探索, 2015(1):179-185.
- [33] 陈祎, 刘阳阳. 劳动合同对于进城务工人员收入影响的有效性分析[J]. 经济学:季刊, 2010, 9(1):687-712.
- [34] 原新, 韩靓. 多重分割视角下外来人口就业与收入歧视分析[J]. 人口研究, 2009, 33(1):62-71.
- [35] 吴晓刚. 1993-2000 年中国城市的自愿与非自愿就业流动与收入不平等[J]. 社会学研究, 2008(6):33-57.
- [36] 蔡禾, 王进. “农民工”永久迁移意愿研究[J]. 社会学研究, 2007(6):86-113.
- [37] 陈祎, 刘阳阳. 劳动合同对于进城务工人员收入影响的有效性分析[J]. 经济学:季刊, 2010, 9(1):687-712.
- [38] 陈映芳. “农民工”:制度安排与身份认同[J]. 社会学研究, 2005(3):119-132.
- [39] 邓曲恒. 城镇居民与流动人口的收入差异——基于 Oaxaca-Blinder 和 Quantile 方法的分解[J]. 中国人口科学, 2007, 2007(2):8-16.
- [40] 陆益龙. 户口还起作用吗?——户籍制度与社会分层和流动(英文)[J]. 中国社会科学, 2008(2):149-162.
- [41] 王美艳. 城市劳动力市场上的就业机会与工资差异——外来劳动力就业与报酬研究[J]. 中国社会科学, 2005(5):36-46.
- [42] 王美艳. 转轨时期的工资差异:歧视的计量分析[J]. 数量经济技术经济研究, 2003, 20(5):94-98.
- [43] 魏万青. 户籍制度改革对流动人口收入的影响研究[J]. 社会学研究, 2012(1):152-173.
- [44] 肖文韬. 户籍制度保护了二元劳动力市场吗[J]. 中国农村经济, 2004(3):22-27.
- [45] 谢嗣胜, 姚先国. 农民工工资歧视的计量分析[J]. 中国农村经济, 2006(4):49-55.
- [46] 段成荣, 孙磊. 流动劳动力的收入状况及影响因素研究——基于 2005 年全国 1%人口抽样调查数据[J]. 中国青年研究, 2011(1):54-61.

致 谢

至此，本人已在西财度过了四年本科与三年研究生学习生活，回首既往，自己一生最宝贵的青春时光能于这样的校园之中，能在众多学富五车、富有人格魅力的老师们的熏陶下度过，实是荣幸之极。毕业论文完成之际，众多思绪一齐涌来，都化成了深深的感谢。

首先，我要感谢我的父母，感谢父母赋予我生命，没有您们便没有我今天体会到的这一切。谢谢您们给了我生命的体验，给了我认知快乐、痛苦、顺境、磨难的机会。儿女总是对父母做不到掩饰和同等宽容，我可能一直在您们面前都是任性的长不大的孩子。谢谢您们能以最大的爱来包容我，谢谢您们虽不善表达却一直默默地关注我，谢谢您们在我最痛苦的时候选择无条件支持我相信我，让我前所未有地感觉到我们一家人的心是那样紧紧联系在一起。不知为何，毕业之际的浮沉让我感受到了生命的力量——生生不息又无处不在的力量，也许这就是爸爸所说的起于“爱”的传承。又想起龙应台《目送》中的话：父母和孩子之间其实只有浅浅的缘分，今生今世不断在目送的背影中渐行渐远。感谢这“浅浅的缘分”，赐予我幸福的家庭、爱的能力、强大的力量，让我充满信心面对未知的道路。我想，我们和父母尽管渐行渐远，但心永远都是连在一起的。

其次，我要感谢我的老师们，特别是我敬爱的导师吴开超老师。感谢成长路上每位老师对我的谆谆教诲，是您们教会了我勤奋学习，诚实做人，踏实做事，以宽容之心面对生活，在点滴汇聚中使我逐渐形成比较成熟的人生观和价值观。特别感谢吴老师，遇到您和我们有爱的师门是我研究生生涯的最好礼物。您说，我就像您的女儿，您亲切地唤我们“丫头”，任我们蹭饭，带我们长见识，给我们建议又尊重我们的选择，在我无助的时候想到您就有许多安慰。感谢您教会我最重要的“用时间换空间”，教会我品味舍得、乐观面对。您和师母在我最失意的时候都牵挂着我，还有关心我的师姐们，都是吴老师带给我的宝贵财富。此外，还要感谢学院的两位辅导员蒋老师和王老师，在我心境状况不佳的情况下当了我的“垃圾回收站”，给我提供了很多帮助，谢谢您们。

再次，我要感谢整个大学阶段陪伴在我身边的同学、朋友，特别是在我遇到困难也对我不离不弃的朋友。我一直相信朋友之间是有磁场的，你是怎样的人，就会遇到怎样的人，不禁感叹，是有怎样的缘分才能遇到你们。你们也串联了我七年的校园生活，往事全部历历在目：大一的竞选班委，大二的校红十字会，大三的家庭金融调查，大四的保研，研一的校党建与研会，研二的大骨班，研三的

找工作波折之路……感谢所有或长或短的陪伴，感谢所有真诚的帮助。西财学子给我的印象都是朝气、勇敢、有想法的，庆幸与你们为伴，让我发现无数的闪光点。

最后，我还想感谢自己，感谢自己没有放弃，感谢自己重新认识了自己，感谢自己越来越敢于坚持内心的选择，感谢自己在无数次自我否定之后还是相信自己。都说当局者迷，旁观者清，自我认知是一条很长的道路，而我是个愿意折腾的探索者。以前常问生活的意义是什么，其实生活的意义就是生活本身吧，生活教会我们这么多，只需要平静地过好每一天，享受每一天带给我们的快乐和烦恼。常记：自我悦纳、心怀感恩。

诚然，由于时间的仓促及自身专业水平的不足，整篇论文肯定存在尚未发现的缺点和错误。恳请阅读此篇论文的老师、同学多予指正，不胜感激。

的确，生活总有许多不期而遇，让我充满希望！