

# 铁路发展与移民研究

——来自 1891~1935 年中国东北的自然实验证据

李 楠

**【摘要】**文章以最新的移民重力模型为理论框架,利用 19 世纪中叶至 20 世纪初期东北地区移民和铁路发展的历史数据,通过构建双重差分模型对该时期铁路发展与移民之间的因果关系进行检验。研究发现,铁路发展对移民有显著的正向影响,特别是 1903 年以后,随着东北铁路网的形成,此影响格外显著。通过构建标准化回归发现,与其他影响移民的决定因素相比,交通的发展扮演着更加重要的角色。

**【关键词】**交通基础设施 铁路网络 移民 移民重力模型

**【作者】**李楠 香港科技大学社会科学部,博士候选人。

## 一、引言

19 世纪中叶至 20 世纪初期,大规模长距离人口迁移构成了当时世界经济发展的一个重要内容。这一人口流动不仅改变了人口地理分布,而且改变了新旧世界的财富分配,因此其备受社会科学研究者的关注,并且对其发生的原因、结果及影响进行了探讨。在解释大规模人口流动原因方面,多数学者主要从 Lewis(1954)和 Todaro(1969)的劳动力流动理论出发,认为近代工业革命引起的工业化和人口相对于土地的压力差异,导致了大规模的人口从工资率较低的地方流动到工资率较高的地方,从人口压力大的地方流动到人口压力相对较小的地方。McKeown(2004)和 Hatton 等(1998)认为,交通技术的革新与发展也是促进大规模移民的重要因素。然而,遗憾的是现有研究仅仅围绕前者,而对后者即交通技术、交通基础设施的改善与发展如何影响移民变化的研究仍然比较缺乏,特别是集中在两个重要问题上:一是交通基础设施的发展与移民的因果关系检验尚未建立,现有研究只讨论了两者之间的相关关系,没有检验交通基础设施的发展是否是促进移民增加的原因。二是在众多影响移民的因素中,交通设施改善与其他决定移民的因素相比,扮演着什么样的角色仍未能深入了解。因此,针对上述问题,本文将利用近代中国东北地区的铁路发展与人口流动的数据资料,提供一个自然实验证据来识别交通设施发展与移民的关系,探讨交通条件改善与其他影响移民因素相比究竟起着什么样的作用。

选择近代中国东北地区作为考察对象,主要原因在于:(1)自19世纪中叶以来,随着清政府修改东北移民政策,大量长期迫于生活压力的华北农民迁移到东北,不但成为19世纪末至20世纪初世界大移民的重要组成部分,而且以每年近50万人的移民规模超过了爱尔兰大移民和美国“西进运动”,成为20世纪早期世界上最大规模的移民运动(Gottschang, 1987)。(2)在近代中国东北地区,大规模、复杂的铁路网络陆续建立起来,其发展速度与规模完全可以同美国横跨东西部的铁路发展相提并论,而且在东北移民中起了巨大的作用(Gottschang, 2000; Lee, 1970; Ginsburg, 1949)。(3)中国东北移民路线地理因素上的差异为铁路与移民之间的因果关系识别提供了便利及自然实验的可能。

## 二、东北的移民与铁路发展

### (一) 近代东北移民

东北地区的开发始于19世纪中叶。第二次鸦片战争后,随着《天津条约》、《北京条约》的签订,西方势力进入中国东北的事实,让清政府感到移民实边的必要性<sup>①</sup>。为了解除边疆危机、增加税收用于支付庞大的战争赔款,清政府开始修改东北移民政策,允许汉人进入东北谋生,从此近代东北开发揭开序幕(孔经纬, 1986)。

17世纪,中国人口迅速增长,从万历十五年(1587年)的1.5亿人上升到乾隆五十九年(1794年)的3.1亿人(Ho, 1959)。华北地区也不例外,人口达到5000万,成为长江三角洲之后人口密度最大的地区(Skinner, 1977)。但用于耕种的土地面积却没有变化,人均土地面积迅速下降,从明代的人均15亩下降到清末的人均3.1亩(徐浩, 1999)。为了提高收入,农户将越来越多的剩余劳动力投入到有限的土地上<sup>②</sup>。此外,华北地区的农民也面临着自然灾害和社会冲突的威胁。在华北地区,特别是河南、河北、山东三省,自然灾害频繁发生。仅1912~1934年,三省中超过2000个县次遭受过自然灾害的袭击,平均每年有20个县次受到自然灾害的影响(夏明方, 2000)。社会动荡也是19世纪末和20世纪初中国社会的一个重要特征,如义和团运动、甲午中日战争和辛亥革命及其后的军阀混战。这些社会动荡都使华北地区成为主要战场,农民的财产及生产资料受到直接或间接的破坏(Myers, 1970)。

华北地区的人口压力、自然灾害及社会冲突与东北地区的开发形成了华北农民移民到东北的动力。前者形成了移民的推力,而后者成为移民的拉力,其结果是在19世纪中叶,巨大的东北移民潮发生在华北地区,特别是20世纪初,东北大豆贸易兴起,这种力量变得更加强烈(Kuang等, 2009)。清入关时,辽东半岛人口不超过30万人(曹树基, 1997),即使到了1850年,东北地区的人口规模也不超过300万人。但在此后的数年中,人口持续增长,

<sup>①</sup> 西方的威胁主要表现在:一是牛庄作为通商口岸被迫开放。牛庄位于辽河下游与渤海入海口处,地理位置优越,而且全年只有近4个月的结冰期,行船可以顺辽河而上进入东北地区腹地,由此西方势力可以向东北内陆渗透(Bank of Chosen, 1920)。二是沙皇俄国利用战争从东北地区掠夺100万平方公里土地。

<sup>②</sup> 这种现象被Huang(1985)称为“内卷化”现象,并且用其解释为什么中国近代经济停滞不前。

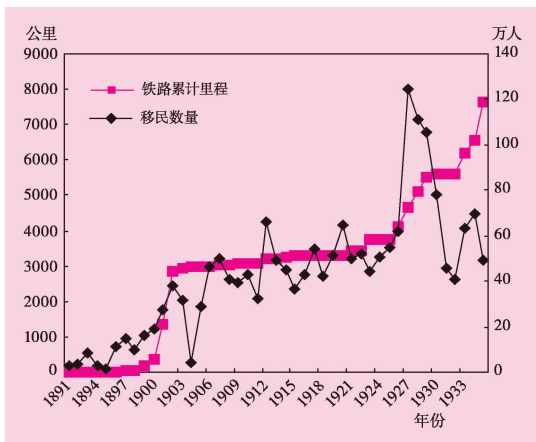


图1 1891~1935年东北地区移民及铁路累计里程变化趋势

资料来源:Gottschang等,2000;马千里等,1983。

末到20世纪初世界上规模最大的人口流动。

## (二) 近代东北铁路的发展

东北铁路修建始于19世纪末。最初的东清与京奉两条铁路的修建标志着东北地区铁路修建运动的开始。此后,东北地区的铁路网迅速形成。到1935年底,铁路累计里程从最初的65公里增加到7635公里,平均每年增加191公里。东北地区成为当时中国铁路里程最长的地区之一。

有关东北铁路发展的时期划分,政治学家和历史学家有着各自的标准。例如,日本满史学会基于铁路产权变化将东北铁路发展过程以1907、1921和1932年为时间点划分为4个时期(满洲历史学会,1959)。然而,Ginsburg(1949)则认为,应该按照铁路网络变化过程将铁路发展划分为3个阶段<sup>②</sup>。但这里将提供一个新的铁路发展阶段划分方法,即根据铁路累计里程变化来进行划分。图1、图2揭示了东北地区从1890~1935年铁路发展趋势。东北铁路发展可以划分为3个阶段:第一阶段(1896~1903年)为铁路网络发展初始阶段。东清、京奉铁路分别于1897和1903年建成。尽管他们不是中国最早的铁路,但这两条铁路的修建揭开了东北铁路发展新的一页。第二阶段(1904~1920年)为铁路稳定增长阶段。“T”字形铁路网络开始在东北地区形成,整个东北从南到北被铁路连接起来。此外,南满铁路公司的成立使大量从日本转移来的资本用于铁路建设。第三阶段(1921~1935年)为快速发展阶

据粗略估计,截至20世纪40年代末期,东北地区总人口超过4千万。因此,从1850~1940年,东北人口增长了近20倍,年平均增长速度超过3.4%。这一时期新增加的人口,仅1/3人口是自然增长,其余均为移民新增人口(Eckstein等,1974)。

从图1中可见,自19世纪末,东北移民数量稳定增长,特别是1927年,移民数量达到了最高点。尽管1928年以后,东北地区东北战争、自然灾害及经济危机的影响,但每年移民仍超过70万人<sup>①</sup>。因此,Gottschang(1987)认为,开始于19世纪的东北移民,是19世纪

① 1928年以来,一系列的水旱灾害在东北地区发生,由此成为东北地区该阶段移民下降的主要因素之一。并且1932年世界性的经济危机也对东北经济产生了一定影响,从而影响了东北地区的移民数量。但是由于存在时间滞后效应,尽管大危机发生在1929~1933年,但其影响并非在当时显现,而在其后危机影响逐渐显现出来。

② Ginsburg(1949)按照东北铁路网形成过程,将东北地区铁路发展过程划分为1896~1908、1908~1931、1931~1945年3个阶段。

段。随着“一战”后,中国国内民族主义上升,路权运动在东北地区兴起,要求采用自己的技术和资本兴建铁路。同时,东北交通委员会成立,制定了东北铁路网发展计划。该委员会拟修建东、西、南3条铁路线,并且基于这3条铁路设计,总共有35条铁路在东北地区修建起来<sup>①</sup>。到“二战”结束时,累计里程超过了1万多公里。铁路网络的形成,不仅促进了农业商品化和对外贸易的加速发展,同时也为移民创造了便利条件,特别是20世纪初期,移民可以通过铁路快速和低成本地到达满洲北部,使人口稀少的吉林、黑龙江两省在这一时期人口得到迅猛增加(王胜今,2005)。

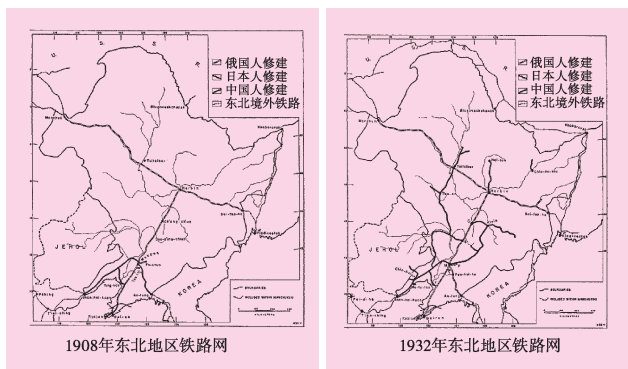


图2 1908和1932年中国东北铁路发展情况  
资料来源: Ginsburg, 1949。

通过铁路快速和低成本地到达满洲北部,使人口稀少的吉林、黑龙江两省在这一时期人口得到迅猛增加(王胜今,2005)。

### 三、模型及假说

#### (一) 移民重力模型及假说

本文的重点在于讨论铁路发展对移民的影响,因此,一个合适的理论框架是十分必要的。然而,现有研究均是基于 Sjaastad(1962), Todaro(1969)和 Harris 等(1970)发展起来的新古典移民决定模型作为基本理论框架。虽然传统的人口流动模型可以帮助我们考察交通发展对移民的影响,但该模型更适合微观层面的探讨而非宏观层面的分析。因此,最新发展起来的移民重力模型将作为本文基本理论模型框架。

重力模型一直是国际贸易理论中一个重要的分析工具。其基本含义为:国际贸易是两国经济总量和两国地理距离的函数,并假定两国的经济总量与贸易量有正向相关关系,理由是如果两国经济总量越大,则越容易发生贸易关系;但距离与贸易呈现负相关关系,因为空间距离的增加可能增加两国贸易的成本,从而减少贸易总量。凭借国际贸易重力模型设计理念, Lewer 等(2008)认为移民同国际贸易一样,可以由两国或地区对移民的吸引力及两地之间的距离所决定。因此,这里根据 Lewer 等(2008)的思路移民重力模型设定为:

$$mig_{ij} = \frac{f_i(x)f_j(x)}{dist_{ij}} \quad (1)$$

式中,  $mig_{ij}$  为从  $i$  国家或地区到  $j$  国家或地区的移民数量;  $f(x)$  分别为两个国家或地区决定移民数量的经济总量函数,其中  $x$  为一组决定移民数量的向量;  $dist_{ij}$  则是两个国家或地区之间的相对距离。

<sup>①</sup> 东线从葫芦岛到抚远(途经葫芦岛—奉天—海龙—吉林—海林—依兰—同江—抚远);西线从葫芦岛到黑河(途经葫芦岛—通辽—洮南—齐齐哈尔—嫩江—黑河);南线则是从朝阳到多伦(途经朝阳—赤峰—多伦)。

从模型 1 可以发现移民同两地对移民的吸引力具有正向相关关系,即  $\frac{\partial mig_{ij}}{\partial f_i(x)} > 0, \frac{\partial mig_{ij}}{\partial f_j(x)} > 0$ ;  
 而移民数量同两地之间相对距离呈负相关关系,即  $\frac{\partial mig_{ij}}{\partial dist_{ij}} < 0$ 。由于近代交通技术的进步,可  
 以影响移出地和定居地之间的相对距离及迁移成本,因此可以将交通设施视为两地距离的  
 函数,即  $dist_{ij} = g(rail_{ij})$ ,并随着交通发展,相对距离越短,即  $\frac{\partial dist_{ij}}{\partial rail_{ij}} < 0$ 。进而可知交通发展同  
 移民数量之间呈现出正向的函数关系,  $\frac{\partial mig_{ij}}{\partial rail_{ij}} > 0$ 。

基于移民重力模型及各变量之间的函数关系,本文需要验证的假说是,19 世纪末至 20  
 世纪初,随着华北与东北地区之间铁路的发展及东北铁路网络的形成,两地之间的相对距  
 离和交通成本大大减少,由此近代东北的铁路发展在人口流动的过程中扮演着重要的角  
 色。特别是新的交通方式(铁路)的发展要比传统的交通方式(轮船)更好地促进人口大规模  
 长距离的流动。

## (二) 移民重力模型的实证模型

尽管理论模型和假说已经给出,但想证明该假说则需要通过实证模型来检验。为方便  
 起见,假设各地区对移民的决定函数为  $f_l(x) = x^{\beta_l}$  且  $l = i, j$ ; 两地距离与铁路关系的函数假设  
 为:  $dist_{ij} = rail^{-\gamma}$ 。将各函数带入式(1)中则有:

$$mig_{ij} = \frac{x_i^{\beta_1} x_j^{\beta_2}}{rail^{-\gamma}} \quad (2)$$

将式(2)两边取自然对数,移民重力模型的实证模型为:

$$\ln mig_{ij} = \beta_0 + \gamma \ln rail + \beta_1 \ln x_i + \beta_2 \ln x_j + u_{ij} \quad (3)$$

式中,  $rail$  为铁路里程;  $x$  为度量各地区对移民影响的因素,如经济发展水平、社会冲突、  
 自然灾害等;  $u_{ij}$  为随机干扰项。但这里需要对系数进行说明,我们希望  $\gamma > 0$ , 因为随着铁路  
 里程的不断增加,移民成本不断减小,移入地和移出地相对距离逐渐缩小,因此铁路发展促  
 进移民的增加;但系数  $\beta_1$  和  $\beta_2$  则不确定,如果对移民影响的因素起到正向作用,则系数为  
 正,如果起到相反作用则系数为负<sup>①</sup>。下面将在式(3)的基础上利用近代东北铁路发展与移  
 民的历史资料来证明铁路发展对移民的影响。

## 四、数据及实证策略

### (一) 数据

为了检验铁路发展与移民之间的关系,笔者收集了 19 世纪末到 20 世纪初东北和华北  
 地区的社会经济数据(见表 1)。

<sup>①</sup> 这一假设与 Lewer 等(2008)的假设相反,在他们的研究中期望反映各地区对移民影响的系数均是  
 正相关。这一点似乎与实际经验不符,因此,本文给予修正,实际上该系数变化同新古典人口流动  
 决定模型相一致。

表 1 主要变量统计描述

变量名称	样本容量	均值	标准差	最小值	最大值
移民的对数值	90	11.879	1.101	8.290	13.736
累计铁路里程的对数值	90	6.820	2.873	0	8.940
移出省份对外贸易的对数值	82	4.326	0.843	3.184	5.813
东北地区对外贸易的对数值	90	5.377	1.377	2.674	7.070
移出省份自然灾害程度	90	0.609	0.498	0	2
东北地区自然灾害程度	90	0.825	0.690	0	2
移出省份人口密度	90	5.366	0.157	5.131	5.625
东北地区人口密度	90	2.607	0.533	1.791	3.303
移出省份社会冲突	90	0.133	0.341	0	1
东北地区社会冲突	90	0.133	0.341	0	1

1.移民数据。移民是本文重要变量,但数据却很难收集。因为,直到1927年,东北移民的信息才被满铁调查部报告。这里所采用的移民数据主要来自Gottschang(1987、2000)的早先工作,该数据记录了19世纪末到20世纪初经河北从陆路及经山东从海路前往东北移民的状况。这进而为本文采用自然实验方法识别铁路发展与移民之间因果关系提供了较好的机会。

2.铁路数据。铁路是本研究另一主要解释变量。铁路数据主要有两个来源:一是满洲历史学会(1988)所编写的《满洲开发四十年史》中的相关记录。在其文献中对东北铁路发展的具体过程进行了详细的描述。另一数据来自马里千等(1983)编写的《中国铁路建筑编年简史》。该书提供了较为详细的东北铁路修建数据。这里将两个数据合二为一,并相互佐证,最终构建出东北铁路发展数据,包括铁路名称、建设起止时间及铁路里程等。

3.贸易数据。由于民国时期国民经济核算体系尚未建立,不能提供人均GDP数据这样度量经济发展水平的数据。但是,根据国民经济核算法,对外贸易是国民经济总量的一个重要组成部分,贸易的变化在某种程度上也暗含着经济总产出的变化。故这里采用不同地区净贸易来近似捕捉各地的经济发展水平。本文所采用的贸易数据主要来自历年海关数据。

4.灾害及社会冲突数据。自然灾害和社会冲突是影响华北农户进入东北的主要因素之一。自然灾害数据来源于《中国近五百年旱涝分布图集》。在该图集中,自然灾害被分为1~5个数值。其中数值3为正常,小于3为水灾,而大于3为旱灾,离3的距离越远表明受害的程度越严重。在本研究中灾害种类并不重要,但其灾害的程度相对来说比较重要,因此灾害程度指数被构造,具体计算公式为:

$$dis-level_t = |disaster_t - 3| \quad (4)$$

式中, $dis-level_t$ 表示第 $t$ 年的灾害程度。其含义是如果数值离中值3越大,则表明灾害程度越大,越小则表明灾害程度越小。

社会冲突主要是指发生在东北和华北地区的战争、军事冲突及匪患等。1911年以前这方面的数据主要来自于中国人民解放军军事科学院编著的《中国军事史》,而1911年以后的数

据则来自 Gottschang (1987) 的记录。

5.人口压力。人口压力也是影响移民的一个重要的因素,但在民国时期没有相应人地比例的统计信息。这里使用不同省份的人口密度来近似度量人地比例之间的差异。不同时期的人口数据来源于赵文林、谢淑君(1988)编写的《中国人口史》,而各省地域面积则来自国民政府主计处编写的《中华民国统计提要》(1940)。然后根据不同时期的人口和土地面积计算出华北及东北各省人口密度。

(二) 实证策略

在现有实证研究中,如何识别铁路发展对移民之间的因果关系是非常困难的,但在本研究中双重差分模型可以帮助解决这一难题。如果一个突发事件仅仅影响一个经济体的一部分而对另一部分没有影响,就可以将其考虑为一个近似自然实验。经济体中受影响部分和没有受影响部分分别在事件冲击前后的差异就是事件对经济体的净影响。

图3揭示了山东、河北及东北地区在民国时期的地理位置。山东与东北隔海相望,而河北濒临东北,并且在19世纪末铁路将东北与河北连接起来。因此,只有从河北经山海关陆路进入东北的移民受铁路的直接影响,而从山东芝罘(烟台)渡海进入东北的移民则受传统交通方式(船运)的影响。从河北和山东两省移民到东北变化的趋势也提供了相应的证据。如图4所示,1903年以前从山东移民进入东北的移民多于从河北出关的移民。主要是因为,在地理位置上,山东离东北较近,可以用较低的成本从烟台和青岛渡海到达东北。然而,随着京奉铁路的修建,陆路进入东北的成本大为降低。1903年之后,从河北进入东北的移民超过从山东经海路进入东北的人数。特别是随着东北铁路网络的形成,经河北进入东北的移民可以迅速到达东北北部,更便捷地寻找生活的机会。因此,将经河北山海关进入东北地区的移民作为实验组,而将从山东经海路进入东北的移民作为对照组,采用双重差分模型作为识别铁路与移民因果影响的实证策略。

这里基于移民重力模型发展的双重差分模型设定为:

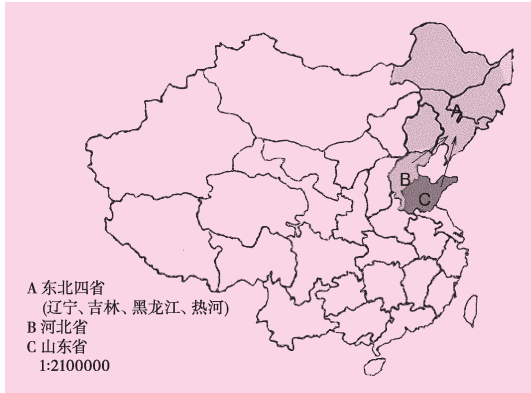


图3 民国时期东北、华北地理区位图

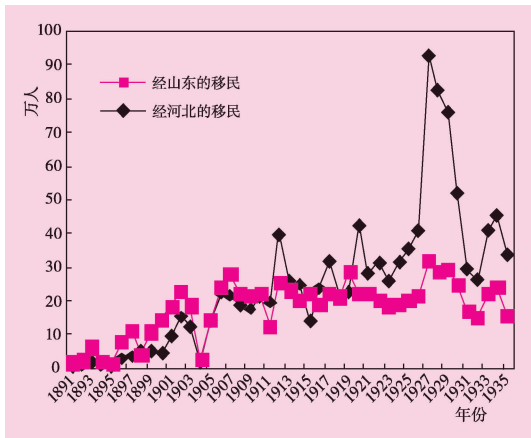


图4 1891~1935年经山东、河北迁入东北地区的移民变化趋势

资料来源:Gottschang等,2000。

$$mig_{i,t} = cont + \alpha \sum_{t=1896-1903}^{1921-1935} yearly_t + \beta region_i + \sum_{t=1896-1903}^{1921-1935} \gamma_t yearly_t \times region_i + \delta X_{i,t} + \varepsilon_{i,t} \quad (5)$$

式中,  $mig_{i,t}$  为在  $t$  时间迁移到地区  $i$  的移民数量;  $yearly_t$  则表明时间的虚拟变量, 本文通过时间上铁路里程变化差异, 用以衡量铁路发展对移民的影响;  $region_i$  为地区虚拟变量, 用以区分实验组与对照组, 其中经河北为 1, 经山东为 0;  $region_i \times yearly_t$  则为地区与时间虚拟变量的交互项;  $X_{i,t}$  则为控制变量, 主要衡量移入地和移出地经济社会对移民的影响;  $\alpha$ 、 $\beta$ 、 $\delta$ 、 $\varepsilon$  分别为时间虚拟变量、地区虚拟变量、控制变量的系数及随机干扰项; 这里最重要的为  $\gamma$ , 它为双重差分估计量, 用以衡量铁路发展对移民的因果影响。

## 五、实证结果及发现

为了得到最终的估计结果, 首先给出初步实证结果, 但鉴于可能存在内生性估计偏差, 估计结果可能被低估或高估。因此, 采用工具变量法对以上结果进行再次估计, 以确保结果的稳定与一致。最后为了比较铁路同其他移民决定因素的影响差异, 标准化回归方法被采用, 以考察铁路发展在移民中所扮演的角色。

### (一) 铁路发展对移民影响的初步估计

表 2 给出了利用双重差分模型识别铁路发展与移民之间因果关系的初步估计结果。在模型 1 和模型 2 中, 铁路里程的虚拟变量被用来构建双重差分估计量。由于数据为面板数据, 固定效应和随机效应估计方法被应用在本研究当中。模型 1 采用固定效应估计方法, 在控制移入地和移出地经济社会因素后, 双重差分估计量显示, 自 1903 年后, 铁路对移民有正向影响, 且显著水平达到 1%。模型 2 采用随机效应估计方法, 其结果与模型 1 相似, 即自 1903 年以后, 铁路对移民有正向影响且统计上显著。但哪个结果更可取则需要借助 Hausman 检验判定。Hausman 检验结果表明随机效应优于固定效应的估计结果, 故随机效应结果被采纳。

由于在度量的过程中, 虚拟变量可能捕捉到时间阶段相同的其他因素对移民的影响, 如同时期的大豆贸易等, 故采用连续变量进行再次度量。在表 2 的最后两列, 双重差分估计量由时间、地区虚拟变量同铁路累计里程的乘积构成<sup>①</sup>。经过 Hausman 检验, 随机效应估计结果被采纳。模型 3 和模型 4 的双重差分的估计结果与模型 1 和模型 2 相似, 自 1903 年以后, 系数为正且在统计上显著, 并且根据回归结果可以得知自 1903 年以来平均铁路累计里程每增加 1 公里经河北陆路进入东北的移民比经山东海路进入东北移民高出 12.7%  $[(0.082+0.173)/2=0.127]$ 。这表明铁路的修建特别是铁路网络的形成对移民具有正向影响, 并且铁路发展是促进移民增加的一个原因。

① 这里双重差分估计量由时间虚拟变量、地区虚拟变量同铁路累计里程的乘积构成, 所采用的模型

相应的变为:  $mig_{i,t} = cont + \alpha \sum_{t=1896-1903}^{1921-1935} yearly_t + \beta region_i + \gamma \sum_{t=1896-1903}^{1921-1935} yearly_t \times region_i \times length_{it} + \delta X_{i,t} + \varepsilon_{i,t}$ , 由此在这里双重差分变量为一个连续型变量, 用于度量铁路里程增加对移民的影响, 这比模型 1 和模型 2 中采用虚拟变量构建的双重差分估计量更有解释力, 直接体现了铁路发展与移民之间的关系。



表2 铁路对移民影响的初步估计结果

被解释变量(移民的对数值)	模型 1(FE)	模型 2(RE)	模型 3(FE)	模型 4(RE)
双重差分估计量				
地区虚拟变量×铁路里程				
1896~1903年	0.438(0.366)	-0.123(0.334)	0.088(0.081)	-0.017(0.075)
1904~1920年	1.011*** (0.299)	0.632** (0.287)	0.118*** (0.037)	0.082** (0.037)
1921~1935年	1.375*** (0.457)	1.433*** (0.483)	0.156*** (0.057)	0.173*** (0.059)
控制变量				
移出地区贸易对数值	0.356(0.270)	-0.284(0.176)	0.354(0.291)	-0.291(0.183)
东北地区贸易对数值	0.161** (0.070)	0.224*** (0.071)	0.122(0.077)	0.184** (0.078)
移出地区自然灾害程度	0.143(0.093)	0.090(0.096)	0.120(0.099)	0.063(0.101)
东北地区自然灾害程度	-0.175** (0.068)	-0.208*** (0.070)	-0.079(0.074)	-0.113(0.076)
移出地区社会冲突	0.131(0.138)	0.085(0.145)	0.098(0.146)	0.055(0.152)
东北地区社会冲突	-0.593*** (0.136)	-0.607*** (0.143)	-0.751*** (0.144)	-0.766*** (0.150)
移出地区人口密度	-0.012(0.488)	0.985*** (0.380)	0.062(0.517)	1.058*** (0.388)
观测值	82	82	82	82
F-statistic/Wald Chi-square	46.24	529.86(13)	40.91	478.17(13)
R-squared	0.899	0.886	0.888	0.875

注:(1)模型1和模型2,铁路虚拟变量捕捉铁路的发展过程。(2)模型3和模型4,铁路里程连续变量用以衡量铁路发展。(3)模型中均包括截距项及时间、地点虚拟变量,但在此表中未列出。(4)Hausman检验分别为9.11和7.66,且不显著。括号内为标准误差;\*、\*\*、\*\*\*分别表示在10%、5%、1%水平上显著。

## (二) 铁路发展对移民影响的稳定性检验(2SLS)

尽管双重差分模型有助于识别铁路发展与移民的因果关系,但不能剔除由于缺失变量、联立关系等所引起的内生性估计偏差。首先可能存在移民和铁路之间的联立问题。铁路发展可以降低移民成本,促使大量农民从华北移民到东北。但随着移民的增加,为东北地区提供了大量廉价的劳动力,又促进了铁路的发展,特别是Gottschang等(2000)指出在移民的职业分布上,除大部分移民从事农业生产、矿山开采以外,有相当一部分移民从事东北地区近代铁路的建设活动<sup>①</sup>。其次,缺失变量所引起的度量偏差也不容忽视,如东北地区的投资,一方面投资的增加创造出更多的就业机会从而促进移民,另一方面随着投资的增加铁路也不断发展。由于这些因素存在,会导致原有估计结果存在向上或者向下的偏差。

解决内生性度量偏差的一个可行性办法是采用工具变量法,然而寻找合适的工具变量是非常困难的,一个好的工具变量只能和内生变量有关而同被解释变量无关。在本文中,主要的内生变量是铁路里程,因此,所选取的工具变量要同铁路建设有关,而同移民变化无关。这里19世纪末至20世纪初的美国钢铁价格被用来作为工具变量以解决内生估计偏差问题。由于近代中国工业化水平较低,铁路修建需要高质量的钢材,而大部分钢材来自进口。故国际钢铁价格直接影响进口的数量,从而影响铁路的发展。而世界钢铁价格却不直接影响中国从华北

<sup>①</sup> 其中在1899年参加中东铁路建设的劳工数量就多达20万人(Gottschang等,2000:51)。

到东北的移民数量。由于当时美国是世界钢铁生产大国,在 19 世纪末到 20 世纪初,美国钢铁生产占世界钢铁生产的一半以上,其钢铁价格变化直接与国际钢铁价格有关(Allen, 1979)。

表 3 给出美国钢铁价格作为工具变量的有效性检验及第一阶段回归结果<sup>①</sup>。在第一列,对移民和滞后一期的美国钢铁价格做回归,并控制了滞后一期的移民、时间趋势及时间虚拟变量等。回归结果表明滞后一期的美国钢铁价格与东北的移民估计系数为 -0.153,虽呈现负相关关系但统计不显著,这表明两者之间没有直接影响。但在第二列,被解释变量与解释变量分别为东北地区的铁路里程和滞后一期的美国钢铁价格,估计结果显示,钢铁价格同铁路里程存在负相关关系且在 5% 统计水平下显著。这表明随着上一期美国钢铁价格上升 1%,东北铁路累计里程下降 16%。通过表 3 的估计,可知美国钢铁价格对于本研究而言是一个较好的工具变量。

表 3 美国钢铁价格与移民、铁路里程的回归结果

被解释变量	模型 1 (移民对数值)	模型 2 (铁路里程对数值)
解释变量		
滞后一期的钢铁价格对数值	-0.153(0.393)	-0.167**(0.076)
滞后一期移民对数值	0.098(0.189)	
滞后一期铁路里程对数值		0.393*** (0.047)
控制变量		
时间虚拟变量(1921~1935 年)	-0.070(0.341)	-0.050(0.074)
时间趋势	0.030(0.018)	0.017*** (0.006)
观测值	35	35
Adj. R-squared	0.154	0.979
F-statistic	2.55	400.17
Durbin-watson statistic	1.940	1.912

注:模型中包含常数项,但系数未列出。括号内为标准误差:\*、\*\*、\*\*\* 分别表示在 10%、5%、1% 水平上显著。

表 4 展示了第二阶段的回归结果。从中可以发现,采用工具变量同未采用工具变量的估计结果类似,显著程度未有变化,但系数存在差异,未采用工具变量估计系数偏小。这表明内生性问题产生了一个向下的偏差,表 2 中铁路发展对移民的影响被低估。在表 4 最后两列,铁路发展用铁路累计里程的连续变量进行度量,同样随机效应结果为理想的估计方法。估计结果与前两列相类似并同表 2 的回归结果比较,同样存在铁路发展对移民影响低估的问题,估计结果显示自 1903 年以来铁路累计里程平均每增加 1 公里经河北的移民比经山东的移民高出 13.7%。最终可以得知,尽管内生性问题对先前估计结果造成了低估,但通过比较发现双重差分估计量仅被低估 1%,这并未对先前得到的结论造成影响。

### (三) 铁路在移民过程中的作用

由于影响移民的因素众多,因此需要知道交通设施的改善同其他因素相比,在移民过程中起到什么样的作用。为了比较不同影响因素在移民过程中的作用,标准化回归方法被用来进行比较分析,估计结果在表 5 给出。表 5 模型 1 中的铁路建设采用铁路虚拟变量度量,而模型 2 采用铁路里程连续变量度量铁路的发展情况。同时,为了保证估计结果具有稳定性,工具变量法再次被采用,利用两阶段最小二乘法对各主要决定移民的因素进行标准化回

<sup>①</sup> 美国钢铁价格数据来源于 <http://minerals.usgs.gov/ds/2005/140/ironsteel.pdf>。

表4 铁路发展与移民估计结果的工具变量检验

被解释变量(移民的对数值)	模型 1(FE)	模型 2(RE)	模型 3(FE)	模型 4(RE)
双重差分估计量				
地区虚拟变量 × 铁路里程				
1896~1903 年	0.458(0.368)	-0.098(0.336)	0.129(0.098)	-0.030(0.086)
1904~1920 年	1.016*** (0.302)	0.657** (0.292)	0.129*** (0.037)	0.085** (0.036)
1921~1935 年	1.367*** (0.459)	1.457*** (0.488)	0.163*** (0.054)	0.178*** (0.058)
控制变量				
移出地区贸易对数值	0.366(0.270)	-0.296*(0.179)	0.379(0.281)	-0.316*(0.179)
东北地区贸易对数值	0.160** (0.070)	0.226*** (0.071)	0.144** (0.071)	0.223*** (0.070)
移出地区自然灾害程度	0.144(0.093)	0.091(0.096)	0.149(0.092)	0.090(0.096)
东北地区自然灾害程度	-0.174*** (0.068)	-0.208*** (0.071)	-0.147** (0.071)	-0.215*** (0.072)
移出地区社会冲突	0.132(0.138)	0.083(0.145)	0.109(0.139)	0.088(0.148)
东北地区社会冲突	-0.593*** (0.136)	-0.607*** (0.143)	-0.627*** (0.136)	-0.607*** (0.144)
移出地区人口密度	-0.015(0.488)	0.997*** (0.380)	-0.028(0.487)	0.971** (0.383)
观测值	82	82	82	82
F-statistic/Wald Chi-square	744403.59(13)	529.73(13)	75607.99(13)	534.47(14)
R-squared	0.899	0.886	0.901	0.887

注:(1)模型 1 和模型 2,铁路虚拟变量捕捉铁路的发展过程。(2)模型 3 和模型 4,铁路里程连续变量用以衡量铁路发展。(3)模型中均包括截距项及时间、地点虚拟变量,但在此表未列出。(4)Hausman 检验分别为 9.2 和 8.86,且不显著。括号内为标准误差;\*、\*\*、\*\*\* 分别表示在 10%、5%、1%水平上显著。

表5 铁路对移民影响的标准化回归结果

被解释变量(移民的对数值)	模型 1(OLS)	模型 2(OLS)	模型 3(2SLS)	模型 4(2SLS)
地区虚拟变量 × 铁路里程				
1896~1903 年	-0.026(0.334)	-0.016(0.075)	-0.020(0.335)	-0.028(0.086)
1904~1920 年	0.255** (0.287)	0.265** (0.037)	0.265** (0.292)	0.276** (0.036)
1921~1935 年	0.356*** (0.483)	0.365*** (0.059)	0.361*** (0.488)	0.374*** (0.057)
移出地区贸易对数值	-0.211(0.176)	-0.217(0.183)	-0.220(0.179)	-0.235*(0.178)
东北地区贸易对数值	0.267*** (0.071)	0.219** (0.078)	0.269*** (0.071)	0.266*** (0.070)
移出地区自然灾害程度	0.040(0.096)	0.028(0.101)	0.041(0.096)	0.040(0.095)
东北地区自然灾害程度	-0.127*** (0.070)	-0.069(0.076)	-0.128*** (0.071)	-0.132*** (0.071)
移出地区社会冲突	0.026(0.145)	0.017(0.152)	0.025(0.144)	0.027(0.147)
东北地区社会冲突	-0.190*** (0.143)	-0.241*** (0.150)	-0.190*** (0.143)	-0.190*** (0.143)
移出地区人口密度	0.142** (0.380)	0.153*** (0.388)	0.144** (0.380)	0.140** (0.383)
观测值	82	82	82	82
F-statistic	40.76	36.78	40.75	41.11
Adj. R-squared	0.864	0.851	0.865	0.866

注:(1)模型 1 和模型 2,铁路虚拟变量捕捉铁路的发展过程。(2)模型 3 和模型 4,铁路里程连续变量用以衡量铁路发展。(3)模型中均包括截距项及时间、地点虚拟变量,但在此表未列出。括号内为标准误差;\*、\*\*、\*\*\* 分别表示在 10%、5%、1%水平上显著。

归。回归后发现,模型 3 和模型 4 回归结果仍然同模型 1 和模型 2 一致,并无较大差异。

无论从以上哪个结果来看,铁路发展对移民的影响总是大于其他影响因素,而且从时间上来看,特别是 1920 年东北铁路网开始形成之后,对移民的影响进一步增大。从这一结果可知,一般传统移民理论均认为影响移民的主要动因在于移入地和移出地之间的边际工资率差异,如果移入地的预期收益高于移出地的预期收益,那么就会有大量移民从移出地流动出来。但从本研究的结果看,地区收入差异固然重要,但交通成本的降低及交通设施的改善与其他众多影响移民的因素相比,扮演着更加重要的角色。因为尽管两地边际工资率的差异是人口流动的动机,但在移民没有就业前,仅是收入预期而已。真正决定农户是否离开移出地则直接考虑的是迁移成本,如果迁移成本低则大量的农户走出移出地;如果移民成本高则农户将会持观望态度,尽管预期收益可能很高。由此可见交通基础设施的发展对移民起着巨大的作用,在整个移民的过程中扮演着重要的角色。19 世纪末 20 世纪初世界长距离人口大流动之所以能够实现也无不跟现代交通技术的革新及发展有关。东北的移民自 1903 年以后能够迅速增加,也正是这方面的体现。

## 六、结 语

19 世纪中叶至 20 世纪初,世界范围长距离人口流动成为当时经济发展的重要内容之一。大量的人口从劳动力富余的地方迁移到劳动力稀缺的地方,特别是新大陆边疆地区。有关为什么会出现这样大规模人口流动的问题上,众多学者虽然肯定现代交通技术的革新和发展对长距离移民有巨大作用,但在实证研究层面相关研究比较缺乏,特别是缺乏对移民与交通发展两者之间因果关系的检验。鉴于现有实证研究的不足,本文主要根据近代东北地区移民及铁路发展的史实为基础,利用移民路线上地理差异,通过构建双重差分模型,来弥补现有实证研究缺乏自然实验证据的不足。

本研究结果发现,铁路发展对移民具有正向影响,特别是 1903 年以后,随着中东铁路与南满铁路的对接及其他深入东北腹地铁路网络的形成,该作用更加明显,铁路累计里程平均每增加 1 公里经河北陆路进入东北的移民要比同期未受铁路影响从山东经海路进入东北的移民在数量上增加 13% 左右。此外,在比较影响移民的各因素过程中发现,虽然不同地区之间的边际工资率差异是导致移民的根本因素,但交通的改善同其他因素相比对移民有着更重要的作用。其原因在于,尽管预期收益等可以成为移民的主要动机,但是毕竟是预期,而交通设施的改善、交通成本的降低则是更为直接影响移民的因素。所有这些发现均表明,19 世纪中叶至 20 世纪初世界范围大规模长距离的移民不仅是由于人口压力和工资率差异的结果,交通技术的革新与发展也是实现这一巨大人口流动的主要因素。

### 参考文献:

1. 曹树基(1997):《中国移民史》,福建人民出版社。
2. 国民政府主计处编(1940):《中华民国统计提要》,国民政府主计处。
3. 孔经纬(1986):《东北经济史》,四川人民出版社。
4. 马里千等(1983):《中国铁路建筑编年简史(1881~1981)》,中国铁路出版社。

5. 满史学会(1988):《满洲开发四十年史》,东北沦陷四十年辽宁编写组译。
6. 王胜今(2005):《伪满时期中国东北地区移民研究:兼论日本帝国主义实施的移民侵略》,中国社会科学出版社。
7. 夏明方(2000):《民国时期自然灾害与乡村社会》,中华书局。
8. 徐浩(1999):《清代湖北农村人口和土地状况的考察》,《清史研究》,第2期。
9. 赵文林、谢淑君(1988):《中国人口史》,人民出版社。
10. 中国军事史编写组(2003):《中国历代战争年表》,解放军出版社。
11. 中央气象局气象科学研究院主编(1981):《中国近五百年旱涝分布图集》,地图出版社。
12. Allen, Robert(1979), International Competition in Iron and Steel, 1850-1913. *Journal of Economic History*. 39(4): 911-937.
13. Anderson, J. (1979), A Theoretical Foundation for the Gravity Equation. *American Economic Review*. 69, 106-116.
14. Bank of Chosen(1920), *Economic History of Manchuria*. Seoul: Bank of Chosen: 11-12, 16-17.
15. Eckstein, Alexander, Kang Chao, and John Chang(1974), The Economic Development of Manchuria: The Rise of A Frontier Economy. *Journal of Economic History*. 34(1): 239-260.
16. Gottschang, Thomas R. (1987), Economic Change, Disasters, and Migration: The Historical Case of Manchuria. *Economic Development and Cultural Change*. 461-490.
17. Gottschang, Thomas R. (2000), *Swallows and Settlers-The Great Migration from North China to Manchuria*. Ann Arbor: The University of Michigan.
18. Ginsburg, Norton(1949), Manchurian Railway Development. *Far Eastern Quarterly*. 8(4): 398-411.
19. Harris, John R. and Todaro, Michael P. (1970), Migration, Unemployment and Development: A Two-Sector Analysis. *American Economic Review*. 60(1): 126-142.
20. Hatton, T.J. and J.G. Williamson(1998), *The Age of Mass Migration: Causes and Economic Impact*. New York: Oxford University Press.
21. Ho, Pingti(1959), *Studies on the Population of China, 1368-1953*. Harvard University Press.
22. Huang, Philip C. C. (1985), *The Peasant Economy and Social Change in North China*. Stanford, California: Stanford University Press.
23. Lee, Robert H. G. (1970), *The Manchurian Frontier in Ch'ing History*. Cambridge, Mass. Harvard University Press.
24. Lewer, J.J. and Hendrik Van den Berg(2008), A Gravity Model of Immigration. *Economic Letters*. 99: 146-167.
25. Lewis, W. A. (1954), Economic Development with Unlimited Supplies of Labor. *Manchester School of Economic and Social Studies*. 22(5): 139-191.
26. McKeown, Adam. (2004), Global Migration, 1846-1940. *Journal of World History*. 15(2): 155-189.
27. Myers, Ramon. H. (1970), *The Chinese Peasant Economy: Agricultural Development in Hopei and ShanTung, 1890-1949*. Cambridge, Mass.: Harvard University Press.
28. Sjaastad, Larry A. (1962), The Costs and Returns of Human Migration. *Journal of Political Economy*. 70(5): 80-93.
29. Skinner, G.W. (1977), *The City in Late Imperial China*. Stanford University Press.
30. Todaro, Michael P. (1969), A Model of Labor Migration and Urban Employment for Less Development Countries. *American Economic Review*. 59(1): 138-148.

(责任编辑:朱犁)