文化因素对人口流动的长期影响：基于中国历史经验的实证分析

社会
2015·4
CJS
第35卷

李　楠

摘　要：文化差异作为解释移民活动的决定因素已不是一个新的视角，然而截至目前，学界尚未对两者之间的量化因果关系进行过讨论。本文利用中国历史上的移民数据，采用新的文化差异度量指标（姓氏基因距离）来揭示地区间文化差异对移民活动的因果影响。研究发现：采用姓氏基因距离度量的地区间文化差异对移民行为具有较强的负向影响，即移民数量随着文化差异的缩小而增加，并且即使控制其他经济、地理等因素以及采用工具变量法进行两阶段回归，该结果依然稳健。本文不仅为理解文化差异与移民行为之间的关系提供了扎实的实证证据，同时也对理解中国自10世纪以来的移民特征及其决定因素有所贡献。

关键词：文化扩散　文化差异　移民

DOI:10.15992/j.cnki.31-1123/c.2015.04.006

Longstanding Cultural Impact on Population Migration in Chinese History

LI Nan

Abstract: Although social scientists and practitioners have long agreed that culture is an important determinant in the migration of populations, so far there has been no study, which provides empirical evidence of a causal relationship between culture and migration. Two things may have contributed to the lacking of empirical research on the subject. Culture is both a tangle and intangible concept, for which testable measures are difficult to design. Also, cultural

---

*作者：李　楠　上海财经大学经济史学系（Author: LI Nan, Department of Economic History, Shanghai University of Finance and Economics）E-mail: li.nan@mail.sufe.edu.cn

** 感谢上海财经大学创新团队项目“中国经济转型的历史与思想研究”（项目编号2014110309）以及2014年度上海市浦江人才计划项目（项目编号14PJC042）对本研究的资助。[This study was supported by the Project of Innovation Team of Shanghai University of Finance and Economic (No. 2014110309) and 2014 Shanghai Pujiang Project Grant (No. 14PJC042).]

change are incremental and subtle, and occur over a huge time. This makes it almost impossible to collect consistent historical data. In an attempt to overcome these shortcomings, this paper examines the impact of long term cultural change on inter-regional population migration by using generic distance from surnames as a measurable variable for cultural variation. The author compiled a database of the historical migration data in the last 1000 years. The finding indicates that the higher the cultural variation, the lower the migration activities. In other words, homogeneity of culture encourages migration while differences of cultures discourage migration. This finding stands the test of controlling variables such as socio-economic and geographic elements. This study has established solid empirical evidence on the casual relation between culture and migration. Furthermore it contributes to the understanding of the characteristics and determinants of Chinese internal migration since the tenth century.

Keywords: cultural diffusion, cultural difference, population migration

一、引　言

文化因素作为理解移民行为的决定因素已不是一个新的话题。早在20世纪,经济学家就发现,除直接经济因素外,文化会通过自身所产生的经济外部性而对移民迁移预期产生影响。一方面,迁入地与迁出地之间的文化差异(如宗教、价值观等)可能增加迁移预期成本,阻碍两地间人口流动,如以色列与阿拉伯国家或者冷战时期东西方国家间的移民等(Dustmann and Preston, 2004);另一方面,迁入地与迁出地之间的文化趋同又会帮助迁移者降低因语言、习俗等差异造成的迁移风险(Akerlof, 1997; Taylor, 1986)。尽管在理论方面已有相关文献进行过讨论,但在实证研究层面至今依然没有揭示两者因果关系的实证证据(Epstein and Gang, 2010)。这既是因为文化的定义广泛抽象且难以量化(Guiso, *et al.*, 2006),也是因为缺乏反映文化长期变化的数据(Roland, 2004)。[1]本文利用姓氏遗传学方法,通过构建地区姓氏基因距离作为地区间文化差异的度量指标,同时结合中国历史上长时期

1. 文化的改变是一个长期的历史过程,绝非短期可以实现。现有实证证据表明,文化改变对经济社会因素产生影响至少需要400年以上的时间(Grosjean,2011)。这也是本文选择中国历史数据,采用大跨度、长时段研究的原因。

移民数据,采用双边模型框架,就文化差异与大规模人口迁移之间的因果关系进行识别,提供实证证据。

将中国作为研究对象主要有以下原因:一是其幅员辽阔、地理环境复杂,地域文化具有多样性;二是其历史具有连续性、长期性、稳定性,为考察文化变迁提供了较好的条件;三是尽管中国历史上的移民活动多发生在省际之间,但从移民规模来看,远远超过近代欧洲和美洲的移民活动。这使得中国历史上大规模人口迁移不仅具有代表性,而且成为世界移民史中不能忽略的问题。[2]最后,研究中国历史上的移民活动对于理解整个中国社会经济发展过程也将起到非常重要的作用(葛剑雄等,1992),例如,如何理解唐宋之际人口与经济重心在空间上从黄河流域向长江流域的转换,边疆移民与中国大一统之间的关系等问题。[3]

本文的假设是:随着地域间文化的不断扩散,文化差异逐渐被改变;当某一地区人口受到移民压力时(如北方游牧民族入侵等),文化差异对移民呈现出较为显著的负向影响,即人口将会流向具有相似文化的地区,较小的文化差异将会导致较大规模的人口流动。为检验该假设,根据现有保存下来的中国历史上的移民和社会经济信息,本文构建了唯一的一个时间跨度达 1200 年、包含中国核心区域 18 个省的省际面板数据。通过采用双边移民模型进行考察,本文发现:在中国历史上无论哪一时期,利用姓氏基因距离度量的文化差异同地区间移民率均有显著的负向关系,即地区间较小的文化差异带来了两地间较大的人口流动。这一结果即使在控制地区间发展水平、地理因素等差异后依然稳定。这也为理解中国历史上长期移民特征提供了帮助。首先,从移民的方向来看,在一千多年的人口迁移过程中,人口主要流动方向是从北向南,但是随着统一帝国的出现(特别是明代之后),边疆移民成为移民的主要内容。其次,从移民规模来看,分裂割据时期的移民规模大于统一时期。这些发现进一步支持了部分移民史学者关于历史移民特征的推断(如葛剑雄等,1993;曹树基,1997a, 1997b;吴松弟,1997)。本文的以上分析结

2. 例如,仅明朝初期的移民活动,规模就超过了 1000 万人(参见曹树基,1997a),而近代东北大移民规模更大,甚至超过了欧洲爱尔兰大移民和美国“西进运动”(参见 Gottschang and Lary, 2000)。

3. 虽然人口史学家(如曹树基,1997a, 1997b;吴松弟, 1997)早已对唐宋时期中国人口重心转移、边疆移民等问题进行了讨论,但这些讨论一方面缺乏可以量化的实证证据支持,另一方面缺乏内在机制的讨论,而本文恰恰提供了一个地区文化差异与移民行为间关系的研究视角。

果，不仅为揭示文化差异与移民之间的因果关系提供了相应的实证证据，也对中国历史上长期移民特征进行了量化分析与总结。

二、历史背景：中国历史上的移民与文化扩散

（一）“流动”的国家：中国历史上的移民

大规模人口流动在中国历史发展过程中扮演着十分重要的角色，甚至可以说中国的历史就是一部移民史（范玉春，2005；葛剑雄等，1992）。尽管在先秦时期，中国就有了最早的移民活动，[4]但对中国历史进程具有最重要影响的几次人口迁移都发生在公元4世纪以后，如西晋“八王之乱”、唐代“安史之乱”、宋代“靖康之难”之后的大规模人口流动以及明清时期的洪武大移民、东北大移民等。这些大规模人口迁移的主要原因是人口压力和战乱。秦汉时期，随着北方经济政治中心地位的加强与确立，北方中原地区在汉代成为经济政治中心，人口密度超过其他地区。据汉代人口记录，在全国近3000万人口中，约4/5人口生活在秦岭淮河以北，仅1/5人口居住在南方地区（邹逸麟，2007）。即使到隋唐时期，北方人口依然约占全国总人口的75%左右（冻国栋，2000）。人口压力的增加必然迫使北方农民外迁，寻找更多生存机会。其次，大规模的人口迁移也是中国历史上持续不断的战争和社会动荡的结果。汉代以前，农耕文明同草原文明虽有冲突，但双方大致可以达到平衡状态，但是在接下来的一千多年里（从公元4世纪到14世纪），随着中原王朝在汉唐末期实力衰弱，出现分裂割据状态，草原民族不断南侵，双方进行了旷日持久的战争。这些战争爆发的地点恰恰是北方人口密集地区。为躲避战祸，大批北方农民逃离家园，选择南方定居。截至14世纪，仅20%的人口生活在中国北方，其余近80%的人口生活在南方地区（范玉春，2005；葛剑雄等，1993）。这一人口分布格局与秦汉乃至隋唐时期的人口分布形成了鲜明对比。

（二）人口流动中的文化扩散

在缺乏文化传播工具的传统社会，人口流动是地区间文化交流的唯一途径（葛剑雄等，1992），而人口流动的结果必然出现文化扩散，使

4. 如殷商时期从商汤到盘庚一系列的迁都活动等。

地区间文化得以交流与融合。文化扩散这一概念最早由克鲁伯(Kroeber,1940)提出,用来描述文化在地区或人群之间的传播,如宗教、语言的扩散等。但由于文化难以识别,我们对历史上的文化扩散现象难以观察。因此,本文只能借鉴历史学家和生物学家(如葛剑雄等,1992;Cavalli-Sforza and Feldman,1981)的做法,选择宗教、方言、价值观等文化元素来观察人口流动背景下的地区间文化扩散。

例如,作为中国基本文化特征的儒教,虽然起源于山东、河南一带,但是在宋代"靖康之难"后,随着大量北方文人南迁,不仅儒教的中心从北方转移到了南方,而且在南方得到了新的发展,产生了以"程朱理学"为代表的新儒家。到明代,这一变化更加突出,全国有近76%的新儒家思想家来自东南四省(包括福建、江苏、江西、浙江),其余约16%的思想家从北方四省(包括河北、河南、山东、山西)产生(谭其骧,1994;邹逸麟,2007)。可见,由于人口压力、战争以及社会动荡所导致的移民行为促进了文化的流动。这不仅对中国传统社会的文化地理分布产生了重大影响,地区间的文化差异也因此而改变。

这种通过文化扩散影响地区间文化距离的现象也可以从方言的演化看出端倪。方言是重要的文化元素,语言学家通常利用地区间方言的语音差异来探索文化差异的变化。邓晓华等(2009)对我国古代地区间语音差异的研究发现,在晋代以前,汉语言边界以秦岭、淮河为主要分界线,但这一界线在经历了"八王之乱"、"安史之乱"、"靖康之难"三次大规模移民后便推移到了长江中下游地区。此外,同一地区内的语音也呈现出多样化和复杂化的特征。以上这些间接证据均反映出,在传统社会中,伴随着人口流动,文化扩散对地区间的文化差异产生了较大影响。

(三)文化扩散与移民的假说

借助人口流动实现的文化交流必然对地区间的文化差异产生影响,缩小或者扩大地区之间的文化差异,而这种新形成的文化差异又将对下一期的人口迁移活动产生影响。例如,在明末清初,四川饱受战争蹂躏,人口大幅度减少,四川成为江西、湖南、湖北移民的主要迁入地。连续的大规模人口迁移使四川与这些地区的生活习惯、语言、宗教信仰等非常接近,而相似的文化又为这些地区在清代中期继续向四川移民提供了可能,形成了中国移民史上所谓的"湖广填四川"。

通过考察中国的历史经验,本文力图揭示地域间文化差异对移民

行为的影响。本文的假设是:在移民的过程中,地区间文化差异对人口流动具有重要影响,文化差异的外部性对移民决定具有很强的负向影响,在其他条件不变的情况下,地区间较小的文化差异将会导致更多的人口流动。

三、模型设定及实证策略

(一)双边模型的设定

为考察文化差异对移民的影响,动态双边移民模型将作为本文的基本实证框架。采用动态双边模型的优势在于,不仅可以考察移民活动的方向性,更重要的是可以消除文化差异与人口流动之间可能存在的相互联立关系。动态双边移民模型设定为:

$$\ln mig_{ijt} = \alpha + \beta \ln cult_{ijt-1} + \gamma X_{ijt} + \varepsilon_{ijt} \quad (1)$$

这里,$\ln mig_{ijt}$ 表示在时间 t 从地区 i 到地区的 j 移民数量的对数值,$\ln cult_{ijt-1}$ 是两地区之间滞后一期的文化差异的对数,X_{ijt} 为一组与移民相关的社会、经济和地理因素控制变量,α、β、γ、ε 为估计系数和随机扰动项。其中,β 为文化差异与移民规模的弹性,该系数直接体现文化差异对移民的影响。如果 β 符号为负,则意味着随着地区间文化差异的缩小,两地间移民数量不断增加。

(二)变量及数据来源

为考察地区间文化差异对移民的影响,本文根据中国历史上的移民和经济社会信息构建了一套跨时一千多年(从公元 10—20 世纪,约 1200 年)、包含中国核心地区 18 个省的省际面板数据。由于数据自身的限制,该数据仅由三个截面、两个时段构成:一个时段是从宋代到明代,另一个时段是从明代到 1982 年,每个时段覆盖约 600 年的移民、文化和经济社会变化的信息。[5]选择这一时期作为观察时段的原因是:首先,中国历史上大规模的移民活动都发生在这一时间段内,例如宋代"靖康之难"后的移民、明清时期的边疆移民等;其次,该时期也是中国历史上文化扩散最活跃的时期,特别是北方知识分子和农民将文化和农业技术向南方地区进行传播;最后,在这一千多年里,中国社会经历了一个由分裂到统一的交

5. 之所以选择较长时间跨度,一方面是缺乏史料,无法进行微观考察;另一方面则是大量实证证据表明,文化改变对经济社会产生影响至少需要 400 年以上的时间(参见 Roland,2004;Grosjean,2011)。

替过程,这为讨论不同历史时期人口流动特征提供了较好的机会。

1. 被解释变量

由于历史上地区间移民数量的直接数据无法获得,因此,通过历史文献中保存的大量姓氏资料而构建的移民率被用来近似地度量不同地区间移民规模差异。[6]本文使用的移民率数据来自中国科学院遗传与发育生物研究所。该所的科学家根据历史文献所记录的姓氏信息和现代人口普查资料分别构建了从宋代到明代和从明代到 1982 年的地区间移民率(袁义达、张诚,2002)。[7]

2. 解释变量

由于文化本身难以度量,社会科学家多采用与文化因素具有较强相关性的变量作为文化的代理变量,如多数经济学家和历史学家采用方言、宗教以及滞后一期的移民存量作为文化的代理度量指标。[8]这些指标可以捕捉文化的部分特征,但依然存在一定局限性,例如不能对历史上长期的文化变迁过程进行有效度量。

近期一些经济学家开始注意到人群间基因距离对理解文化差异与经济发展之间关系的重要性(如 Giuliano, *et al*., 2014; Spolaore and Wacziarg, 2009 等)。因此,本文也将尝试采用基因距离作为地区间文化差异的主要度量指标。但与其他研究不同的是,本文将根据遗传学的一个分支即姓氏群体遗传学的相关理论,通过姓氏在不同历史时期的地域分布差异构建姓氏基因距离,进而度量地区间的文化差异。

姓氏群体遗传学认为姓氏与 Y 染色体遗传方式一致,特别是在父姓社会中,一个群体或地区的不同姓氏分布就相当于同一基因座上等位基因分布。因此,比较不同群体之间的姓氏分布可以知道不同群体之间的

---

6. 最早用姓氏来研究移民的是英国生物学家查尔斯·达尔文的儿子乔治·达尔文。他于 1875 年首次提出,姓氏可以作为一种基因符号用来考察近亲结婚以及地区间移民等问题(Darwin, 1875)。因此,自 19 世纪中叶以来,姓氏群体遗传学成为生物学的一个重要分支,大量生物学家通过姓氏来研究近亲婚姻、移民等问题。

7. 20 世纪末,中科院遗传与发育生物研究所沿袭波德莫和卡瓦利—斯福扎(Bodmer and Cavalli-Sforza,1968)以及韦斯曼等人(Wijsman, *et al*., 1984)的方法,通过宋代到现今(1982 年)的人物传记和人口普查资料,计算出了这一历史时段地区之间的迁移率(参见袁义达、张诚,2002)。

8. 例如,哈顿和威廉姆森(Hatton and Williamson, 1994)利用滞后一期的移民存量作为文化因素的度量指标,而其他经济学家(如 Barro and McCleary, 2003; Falck, *et al*., 2012; Guiso, *et al*., 2006)采用宗教、方言及价值观等作为文化的度量指标。

基因距离。中科院遗传与发育生物研究所的生物学家利用《宋人传记资料索引》(昌彼得等,1988)和《明人传记资料索引》(昌彼得等,1987)两个文献,通过计算不同时期各省之间同姓率的概率分布,得到宋代和明代各省间的姓氏基因距离。[9]研究发现:在宋代,河南与甘肃之间具有最短的基因距离(=0.06),而山西与福建具有最长的基因距离(=0.93);但是到了明代,河北和河南具有最近的基因距离(=0.01),而广西和陕西具有最远的基因距离(=0.79)。

尽管许多实证研究表明,根据Y染色体(姓氏)构建的基因距离同文化差异具有较强的相关性(Battaglia, *et al.*, 2009; Sokal, 1988; Wen, *et al.*, 2004),但是这里仍然要提供相应的证据证明姓氏基因距离是文化差异的有效代理变量。首先,采用姓氏基因距离同DNA基因距离进行相关性分析,以证明姓氏基因距离作为基因距离的有效性;然后,再利用作为重要文化标志的方言构建的语言距离同姓氏基因距离进行相关分析,用来证明姓氏基因距离可以作为文化差异的有效代理变量。[10]表1面板A给出了基本分析结果,从中可以看到它们之间均表现出显著的正相关。这进一步证明,姓氏基因距离可以作为文化差异的有效度量。

表1:基因距离(姓氏)、基因距离(DNA)、语言距离的相关系数

面板A			
	基因距离1982(DNA)	语言距离1982	基因距离1982(姓氏)
基因距离1982(DNA)	1.000		
语言距离	0.607*** (0.000)	1.000	
基因距离1982(姓氏)	0.643*** (0.000)	0.691*** (0.000)	1.000
面板B			
	基因距离1982 (DNA)	语言距离1982	明代基因距离(姓氏)
基因距离1982(DNA)	1.000		
语言距离	0.607*** (0.000)	1.000	
明代基因距离(姓氏)	0.598*** (0.000)	0.535*** (0.000)	1.000

注:1. 括号内为p值;
2. *** 显著水平为1%;** 显著水平为5%;* 显著水平为10%。

3. 控制变量

除文化因素外,其他因素也可能影响地区间移民数量的变化,因此

9. 有关宋、明时期各省份之间基因距离的计算,参见袁义达等,1999,2000。

10. 姓氏和DNA测算的基因距离数据来自Du, *et al.*, 1991;语言距离来源于陆致极,1992:第4章,表4、表6。

需要对这些因素进行控制。首先,人口压力是导致移民的重要因素,因此,地区间的人口密度差异被作为控制变量放入模型。本研究中,宋代与明清时期的人口分别来自吴松弟(2000)与曹树基(2000a, 2000b)编著的《中国人口史》,各地区面积数据来自《中国历史地图集》(谭其骧,1982),根据这些数据最终计算出不同省份在不同历史时期的人口密度。其次,战争与自然灾害也是决定移民的重要因素,因而这些变量也作为控制变量被纳入考虑。有关战争和自然灾害的数据分别来自《中国历代战争年表》(傅仲侠等,2003)和《中国古代重大自然灾害和异常年表总集》(宋正海,1992)。此外,地区间开发早晚差异也是影响移民方向的重要因素,在此也将考虑。本研究采用该地区是否具有新石器文化遗址(如果有=1)控制地区间开发早晚差异对移民数量的影响。新石器文化遗址分布信息来自于《中国历史地图集》(Cheng and Hsu, 1980)。最后,有三类地理因素也被作为控制变量:一是移民方向虚拟变量,主要是移出地是否是北方地区(是=1);二是迁入地是否是边疆地区(是=1);三是各地间的欧几里德地理距离也被考虑,其数据来自《中国历史地图集》(谭其骧,1982)。前两个虚拟变量主要识别历史上移民的主要特征,而地理距离则衡量地区间移民的相对成本。主要变量统计描述在表 2 中给出。

表 2: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
地区之间移民率($\times 10^{-4}$)	480	160.4	518.2	0	1692
地区之间基因距离(采用姓氏测算)($\times 10^{-4}$)	480	2865.9	1716.4	142	9314
该地区是否有文化遗址?(有=1)	480	0.758	0.429	0	1
移民目的地是否为边疆?(是=1)	480	0.717	0.451	0	1
移民移出地是否为北方?(是=1)	480	0.594	0.492	0	1
是否为统一时期?(是=1)	480	0.5	0.501	0	1
人口密度差异的对数值	312	1.134	0.841	0.026	3.947
战争频率差异的对数值	422	0.549	0413	0	2.028
自然灾害(水、旱灾)发生频率的对数值	480	1.027	0.703	0	2.779
地区间地理距离(单位:千米)	480	1508.9	682.9	252	3240

(三)实证策略

在报告估计结果前,需要进一步讨论实证策略,特别是如何纠正模

型中由于缺失变量、度量误差所产生的内生估计偏差。本文将从两个方面来解决这一问题,进而识别文化差异与移民规模之间的因果关系。

首先,为剔除缺失变量对估计结果产生的偏差,本文采用工具变量法。一个有效的工具变量应该同内生变量具有较强的相关关系,但对被解释变量没有直接的影响。对本研究而言,有效的工具变量应与文化差异有较强的相关关系,但对移民率没有直接影响。因此,地区间滞后一期的战争频率差异可作为工具变量识别缺失变量所产生的估计偏差。其逻辑是:滞后一期的地区间战争频率差异对当前的移民率没有直接影响,但对滞后一期的文化差异产生影响。如果这个假说成立,那么滞后一期的地区间战争频率差异将外生于当期的移民率。两阶段回归方程设定如下:

$$\ln mig_{ijt} = \alpha \ln \hat{cult}_{ijt-1} + \beta X_{ijt} + \varepsilon_{ijt} \quad (2)$$

$$\ln cult_{ijt-1} = \gamma \ln warfreq_{ijt-1} + \varphi X_{ijt} + \sigma_{ijt-1} \quad (3)$$

这里,方程(2)同方程(1)基本相同,所不同的是作为核心被解释变量的 $\ln cult_{ijt-1}$ 是通过第一阶段回归方程(3)获得的估计值。而在第一阶段方程(3)中,$\ln cult_{ijt-1}$ 是两地区之间 $t-1$ 期的文化差异的对数,即所需处理的内生变量;而 $\ln warfreq_{ijt-1}$ 为滞后一期的地区间战争频率差异,是识别文化差异内生变量的工具变量。

其次,还要考虑可能对估计结果产生影响的度量误差。一方面,尽管姓氏基因距离是一个衡量历史上地区间文化差异的有效变量,但同其他指标一样,它并不能完全反应文化特征,会产生度量误差。因此,可根据 DNA 计算的基因距离和语言距离替代姓氏基因距离,用来度量文化差异。另一方面,由于缺乏历史移民数据,尽管本文使用移民率作为移民规模的度量,但是被解释变量依然可能存在度量误差。因此,在采用小样本进行一致性估计检验的同时,本文也根据曹树基(1997a)整理的移民宗族和移民村庄的迁移信息重新构建两个数据库,考察被解释变量的度量误差是否对估计结果产生影响。

四、实证结果

(一) OLS 估计结果

表 3 报告了根据回归方程(2)采用最小二乘法(OLS)的估计结

表 3:文化差异对移民的影响:OLS 回归结果

被解释变量	地区间移民率的对数值				
	(1)	(2)	(3)	(4)	(5)
解释变量					
基因距离对数值	−0.220***	−0.338***	−0.436***	−0.357***	−0.658***
	(0.081)	(0.120)	(0.115)	(0.122)	(0.140)
控制变量					
人口密度差异的对数值		0.206***	0.210***	0.214***	0.138**
		(0.047)	(0.049)	(0.049)	(0.067)
战争频率差异的对数值		0.306*	0.377**	0.324*	0.840**
		(0.157)	(0.162)	(0.165)	(0.358)
自然灾害差异的对数值		−0.001	0.017	0.024	0.118
		(0.096)	(0.097)	(0.097)	(0.146)
移民目的地是否为边疆地区?(是=1)			0.033	0.008	0.418*
			(0.150)	(0.150)	(0.226)
移民出发地是否为北方地区?(是=1)			0.433***	0.487***	0.559***
			(0.142)	(0.147)	(0.204)
地理距离的对数值		−0.115	−0.083	−0.309	0.006
		(0.232)	(0.228)	(0.245)	(0.338)
是否为统一时期(是=1)		−0.732***	−0.728***	−0.729***	
		(0.151)	(0.148)	(0.148)	
是否有新石器文化遗址(有=1)				0.349*	0.853***
				(0.187)	(0.262)
常数项	6.098***	7.500***	7.870***	7.340***	8.024***
	(0.630)	(0.860)	(0.823)	(0.876)	(1.005)
观测值	480	312	312	312	156
F-统计量	7.36	17.7	10.32	9.68	9.65
R-squared	0.013	0.177	0.202	0.210	0.288

注:1. 第 1—5 列的被解释变量为通过姓氏计算的移民率,解释变量为以姓氏基因距离度量的文化差异;

2. 第 1—4 列给出了全样本的估计结果,即从宋代到 1982 年;第 5 列仅仅给出了小样本的估计结果,即仅包括从明代到 1982 年的样本;

3. 标准误差根据卡梅伦(Cameron, *et al.*, 2011) 的多边聚类方法调整给出;

4. 括号内为稳定性标准误差;

5. *** 显著性水平为 1%; ** 显著性水平为 5%; * 显著性水平为 10%。

果。[11]在第 1 列未加入任何控制变量的情况下,回归结果显示,用姓氏基因距离度量的文化差异对移民率有较强负向影响。回归系数表明,基因距离每增加 1%,移民率将会减少 0.22%。在接下来的第 2 列,分别控制住地区间战争和自然灾害频率、人口密度、地理因素等差异以及时间虚拟变量[度量统一和分裂时期移民差异(统一时期=1)]后,基因

11. 由于采用双边移民模型作为实证框架,因此,稳定性标准误差根据调整后的多边聚类方法调整给出(参见 Cameron, *et al.*, 2011)。

距离的估计系数依然为负且统计上显著(显著水平小于1%)。而在第3列,两个地理因素虚拟变量加入模型,即迁入地是否为边疆地区(是=1)和迁出地区是南方还是北方(北方=1)。控制这些变量后,回归结果依然同前两列一致,估计系数无显著差异。

在表3的最后两列,有关地区开发早晚的虚拟变量[即迁出地是否有新旧石器文化遗址(有=1)]被加入模型中。同时,为了提供更可靠的估计结果,分别使用全样本和小样本(仅包括明代以后的数据)进行分析。此时,表3中第4列和第5列的回归系数依然表明,姓氏基因距离对移民率有显著的负向影响。同时,还可发现,文化遗址虚拟变量对移民也有较强的影响,即越具有文化遗址的地方越容易移民。这恰恰支持了齐默曼和鲍尔(Zimmerman and Bauer,2002)有关文明通过移民扩散的观点。

此外,从表3的回归结果也可发现:首先,与之前有关移民决定因素认识一致的是,地区间的人口密度、战争、自然灾害频率差异对移民有显著影响,是引起移民的主要动力。其次,根据时间虚拟变量估计结果,分裂比统一时期更容易导致移民,因为分裂时期战争不断,人们为了躲避战争而不得不大量迁移。最后,比较明代以后小样本和全样本的估计结果时,地理虚拟变量表明:总体上看,中国历史上移民方向是从北向南,但是明代之后,特别是随着大一统国家的形成,边疆移民成为移民的主要方向。以上这些发现进一步验证了之前人口史学家的判断(曹树基,1997a, 1997b)。

(二)工具变量回归结果

尽管表3的回归结果给出了文化差异与移民之间关系的考察,但这些结果还可能受到潜在内生性估计偏差的影响。因此,地区间滞后一期的战争频率差异作为滞后一期的地区间文化差异的工具变量可用来解决内生性估计偏差。

带有工具变量的两阶段估计结果在表4给出。从表4第1列未加入任何控制变量的估计结果看,使用工具变量后,文化差异对移民影响的系数依然显著为负,且用姓氏基因距离度量的文化差异系数每增加1%,地区间移民率将减小0.21%。而在第2列和第3列控制住地区间人口密度、战争频率、自然灾害频率、地理等因素后,该系数依然为负且统计显著。最后,为得到最终一致性的估计结果,同表3识别策略一样,在最后两列分别给出了长期文化扩散对移民的影响以及分别采用全样本和小样本得出的

回归结果。估计结果表明:在采用带有工具变量的两阶段回归后,地区间基因距离依然同移民率呈现显著的负向关系,而且对比表4和表3的回归系数可发现,采用工具变量后,基因距离的系数虽然有所增加但无显著差异。这表明内生性估计偏差未对估计结果产生致命影响。

表4:文化差异与移民率:工具变量回归结果

被解释变量	地区间移民率的对数值				
	(1)	(2)	(3)	(4)	(5)
解释变量					
基因距离对数值	−0.207*	−0.286*	−0.441***	−0.319*	−0.748***
	(0.114)	(0.170)	(0.168)	(0.184)	(0.166)
控制变量					
人口密度差异的对数值		0.206***	0.210***	0.213***	0.137**
		(0.047)	(0.049)	(0.049)	(0.067)
战争频率差异的对数值		0.298*	0.378**	0.315*	0.860**
		(0.157)	(0.163)	(0.167)	(0.360)
自然灾害差异的对数值		0.002	0.017	0.025	0.121
		(0.097)	(0.097)	(0.097)	(0.146)
移民目的地是否为边疆地区?(是=1)			0.033	0.009	0.414*
			(0.149)	(0.150)	(0.226)
移民出发地是否为北方地区?(是=1)			0.434***	0.480***	0.581***
			(0.143)	(0.147)	(0.205)
地理距离的对数值		−0.168	−0.078	−0.356	0.140
		(0.265)	(0.257)	(0.289)	(0.371)
是否为统一时期(是=1)		−0.726***	−0.728***	−0.725***	
		(0.151)	(0.149)	(0.148)	
是否有新石器文化遗址(有=1)				0.370*	0.803***
				(0.196)	(0.270)
常数项	6.339***	7.181***	7.898***	7.106***	8.535***
	(0.909)	(1.136)	(1.109)	(1.231)	(1.144)
观测值	480	312	312	312	156
F-统计量	13.64	9.61	9.67	9.16	10.09
R-squared	0.076	0.177	0.202	0.209	0.286

注:1.第1—5列的被解释变量为通过姓氏计算的移民率,解释变量为以姓氏基因距离度量的文化差异;

2.第1—4列给出了全样本的估计结果,即从宋代到1982年;第5列仅给出了小样本的估计结果,即仅仅包括从明代到1982年;

3.工具变量为之后两期的各地战争频率的差异;

4.标准误差根据卡梅伦(Cameron, *et al.*, 2011)的多边聚类方法调整给出;

5.括号内为稳定性标准误差;

6. ***显著性水平为1%;**显著性水平为5%;*显著性水平为10%。

(三)稳健性检验

为进一步获得稳健性的估计结果,还需要对存在于解释变量和被解释变量中的度量误差进行考察,识别这些潜在的度量误差是否对估

计结果产生影响。为了识别解释变量中可能存在的度量误差,分别采用 DNA 计算的基因距离以及地区间语言距离替代姓氏基因距离,作为文化差异的新的代理变量。基因距离与语言距离数据分别来自杜若甫 Du, *et al*., 1991)和陆致极(1992)的研究。由于期望这些指标可以替代明代姓氏基因距离,因此在得到正式估计结果之前对三个变量之间的相关关系进行讨论,结果在表 1 面板 B 给出。表 1 面板 B 表明,姓氏基因距离与 1982 年的 DNA 基因距离、语言距离之间存在较强的正相关。因此,采用 DNA 基因距离和语言距离可以替代明代姓氏基因距离,作为文化差异的代理变量。

表 5 提供了仅采用明代以后数据分析的稳定性检验结果。表 5 前两列分别报告了控制地区间地理距离情况下, DNA 基因距离、方言距离对地区间移民率的影响。估计系数进一步确认了文化差异与移民具有显著的负向关系。在接下来的第 3 列和第 4 列控制了一系列社会、经济以及地理因素后,采用 DNA 基因距离与方言距离度量的文化差异依然与前两列估计结果一致。最后两列给出的是带有工具变量的两阶段回归结果,估计结果也皆与先前的估计结果相同。

此外,为了识别被解释变量的度量误差,从地方志整理出来的移民宗族和移民村落两个数据也被采用,即,将采用两地之间移民村落数量差异以及移民族谱中的迁出地和迁入地信息测算的文化差异同采用姓氏和 DNA 测算的文化差异进行相关分析。所有结果均表明,文化差异与移民之间存在显著的负向相关性。[12] 由此可知,估计误差对先前估计结果没有影响。

五、结论

文化差异作为影响大规模移民的决定因素已受到较长时间的关注,但由于文化本身难以量化以及缺乏长时段数据,最终导致有关文化差异与移民关系的实证研究比较缺乏。因此,本文使用唯一一套中国历史上长时段文化差异与人口迁移的省际面板数据,通过构建动态双边移民模

12. 由于移民群体族谱和村落观测值有限(仅 30 个),不能进行有效的回归分析,只能进行粗略的相关关系检验。其中采用族谱分析得到的族谱与姓氏基因距离、DNA 距离的相关系数分别为 −0.288 和 −0.529;而根据移民村落信息与姓氏基因距离、DNA 距离得到的相关系数分别为 −0.447 和 −0.343。

型,采用新的文化度量指标检验文化差异对移民的因果影响。

表 5:文化差异与移民:稳定性检验结果

被解释变量	地区间移民率的对数值(明代到 1982 年)					
	(1)	(2)	(3)	(4)	(5)	(6)
解释变量						
基因距离对数值(DNA)	−0.163***		−0.170*		−0.430**	
	(0.094)		(0.078)		(0.206)	
语言距离对数值		−1.151*		−1.284**		−1.105*
		(0.581)		(0.594)		(0.694)
控制变量						
人口密度差异的对数值			0.137**	0.133*	0.131*	0.136**
			(0.062)	(0.081)	(0.072)	(0.059)
战争频率差异的对数值			0.727***	0.967*	0.779**	0.950**
			(0.197)	(0.499)	(0.351)	(0.330)
自然灾害差异的对数值			0.064	0.078	0.008	0.091
			(0.130)	(0.199)	(0.161)	(0.203)
移民目的地是否为边疆地区?			0.479**	0.670**	0.522**	0.662**
(**是**=1)			(0.209)	(0.303)	(0.243)	(0.294)
移民出发地是否为北方地区?			0.412*	0.630**	0.434**	0.606**
(**是**=1)			(0.292)	(0.257)	(0.216)	(0.242)
是否有新石器文化遗址?			1.091***	0.883**	0.886***	0.901**
(**有**=1)			(0.272)	(0.346)	(0.311)	(0.297)
地理距离的对数值	0.174	−0.735**	−0.652**	−0.201	−0.153	−0.250
	(0.258)	(0.334)	(0.298)	(0.442)	(0.516)	(0.295)
常数项	4.298***	6.871***	4.386***	3.725***	4.581***	3.697***
	(0.393)	(0.688)	(0.514)	(0.693)	(0.542)	(0.711)
观测值	240	132	156	110	156	110
F-statistic	5.12	8.16	12.78	5.56	8.51	18.53
R-squared	0.013	0.559	0.217	0.257	0.190	0.257

注:1. 第 1—6 列的被解释变量为通过姓氏计算的从明代到 1982 年的移民率;
2. 第 1、3、5 列解释变量为 DNA 基因距离度量的文化差异,而第 2、4、6 列解释变量为根据方言距离计算的文化差异;
3. 第 1、4 列给出了 OLS 估计结果,带有工具变量的 2SLS 估计结果在第 5、6 列给出;
4. 第 5、6 列中的工具变量为滞后两期的各地战争频率的差异;
5. 标准误差根据卡梅伦(Cameron, *et al.*, 2011)的多边聚类方法调整给出;
6. 括号内为稳定性标准误差;
7. ***显著性水平为 1%;**显著性水平为 5%;*显著性水平为 10%。

研究发现,在控制住一系列社会、经济、地理等因素后,采用姓氏基因距离度量的文化差异对移民有显著的负向影响,即文化差异越小,地

区间的移民规模越大。在采用工具变量法剔除了潜在内生性偏差后，该结果依然稳定。由此我们相信，文化差异是影响地区间移民的一个重要决定因素，这为识别文化对移民的影响提供了扎实的实证证据。

与此同时，本文也为理解中国历史上较长时段的大规模移民活动提供了一个量化的动态考察。实证结果揭示了中国历史上移民活动的以下几个特征：首先，人口密度和战争对移民活动有显著影响，是中国历史上移民的主要动力；其次，分裂时期与统一时期相比更容易导致大规模的移民活动；再次，从移民方向上看，总体上中国的移民方向是从北向南，但是在明代之后随着大一统国家的形成，边疆成为移民的主要方向；最后，通过对迁入地与迁出地开发历史早晚的考察，发现移民主要是从农耕文明的成熟地区向不成熟地区流动。所有这些不仅为理解中国历史上的移民活动、地理疆域变化提供了量化的数据，同时也为理解当前中国地理疆域的形成以及人口分布格局奠定了基础。

参考文献(References)

曹树基. 1997a. 中国移民史(明时期)[M]. 福州：福建人民出版社.
曹树基. 1997b. 中国移民史(清时期)[M]. 福州：福建人民出版社.
曹树基. 2000a. 中国人口史(第四卷)[M]. 上海：复旦大学出版社.
曹树基. 2000b. 中国人口史(第五卷)[M]. 上海：复旦大学出版社.
昌彼得、王德毅. 1988. 宋人传记资料索引[M]. 北京：中华书局.
昌彼得、乔衍琯、宋常廉. 1987. 明人传记资料索引[M]. 北京：中华书局.
邓晓华、王士元. 2009. 中国的语言及方言的分类 [M]. 北京：中华书局.
冻国栋. 2000. 中国人口史(第二卷)[M]. 上海：复旦大学出版社.
范玉春. 2005. 移民与中国文化 [M]. 南宁：广西师范大学出版社.
傅仲侠、田昭林、张醒、杨伯时. 2003. 中国历代战争年表 [M]. 北京：解放军出版社.
葛剑雄、曹树基、吴松弟. 1992. 移民与中国 [M]. 香港：中华书局.
葛剑雄、曹树基、吴松弟. 1993. 简明中国移民史 [M]. 福州：福建人民出版社.
陆致极. 1992. 汉语方言数量探索 [M]. 北京：语文出版社.
宋正海. 1992. 中国古代重大自然灾害和异常年表总集 [M]. 广州：广东教育出版社.
谭其骧. 1982. 中国历史地图集 [M]. 上海：地图出版社.
谭其骧. 1994. 中国文化的时代差异和地域差异 [G]//长水集. 北京：人民出版社：171－188.
吴松弟. 1997. 中国移民史 (宋辽金元时期)[M]. 福州：福建人民出版社.
吴松弟. 2000. 中国人口史 (第三卷)[M]. 上海：复旦大学出版社.
袁义达、金锋、张诚、斋藤成也. 1999. 宋朝中国人的姓氏分布与群体结构分化 [J]. 遗传学报(3)：187－197.
袁义达、张诚. 2002. 中国姓氏：群体遗传和人口分布 [M]. 上海：华东师范大学出版社.
袁义达、张诚、杨焕明. 2000. 中国人群体遗传 II：姓氏传递的稳定性与地域人群的亲缘关系 [J]. 遗传学报(7)：565－572.
邹逸麟. 2007. 中国历史地理概述 [M]. 上海教育出版社.

Akerlof, George. A. 1997. "Social Distance and Social Decisions." *Econometrica* 65 (5): 1005—1027.

Barro, Robert J., and Rachel M. McCleary. 2003. "Religion and Economic Growth." *American Sociological Review* 68: 760—781.

Battaglia, Vincenza, Simona Fornarino, Nadia Al-Zahery, Anna Olivieri, Maria Pala, Natalie M. Myres, Roy King, Siiri Rootsi, Damir Marjanovic, Dragan Primorac, Rifat Hadziselimovic, Stojko Vidovic, Katia Drobnic, Naser Durmish, Antonio Torroni, Silvana Santachiara-Benerecetti, Peter Underhill, and Ornella Semino. 2009. "Y-Chromosomal Evidence of the Cultural Diffusion of Agriculture in Southeast Europe." *European Journal of Human Genetics* 17: 820—830.

Bodmer, Walter F., and Luigi L. Cavalli-Sforza. 1968. "A Migration Matrix Model for the Study of Random Genetic Drift." *Genetics* 59: 565—592.

Cameron, Colin, Jonah B. Gelbach, and Douglas L. Miller. 2011. "Robust Inference with Multi-Way Clustering." *Journal of Business and Economic Statistics* 29 (2): 238—249.

Cavalli-Sforza, Luigi L. 1981. *Cultural Transmission and Evolution: A Quantitative Approach*. Princeton, NJ: Princeton University Press.

Cheng, Kuangyu, and Shengmo Hsu. 1980. *Historical Atlas of China*. Taiwan: Chinese Culture University Press.

Darwin, Ceorge H. 1875. "Marriages between First Cousins in England and Their Effects." *Journal of Statistical Society* 38:153—184.

Du, Roufu, Yida Yuan, Juliana Hwang, Joanna Mountain, and Luigi L. Cavalli-Sforza. 1991. *Chinese Surnames and the Genetic Differences between North and South China*. Department of Genetics, Stanford University, School of Medicine 94305, Paper number 0027.

Dustmann, Christian, and Ian Preston. 2004. *Racial and Economic Factors in Attitudes towards Migration*. London: University College London, Center for Research and Analysis of Migration, Discussion Paper 0104.

Epstein, Gil S., and Ira N. Gang. 2010. "Migration and Culture." IZA working paper, DP No. 5123.

Falck, Oliver, Stephan Heblich, Lameli Alfred, and Jens Südekum. 2012. "Dialects, Cultural Identify, and Economic Exchange." *Journal of Urban Economics* 72 (2-3): 225—239.

Giuliano, Paola, Antonio Spilimbergo, and Giovanni Tonon. 2014. "Genetic Distance, Transportation Cost, and Trade." *Journal of Economic Geography* 14 (1): 178—198.

Gottschang, Tthomas, and Diana Lary. 2000. *Swallows and Settlers: The Great Migration from North China to Manchuria*. Ann Arbor: Center for Chinese Studies, the University of Michigan.

Grosjean, Pauline. 2011. "The Weight of History on European Cultural Integration: A Gravity Approach." *American Economic Review* 101(3): 504 — 508.

Guiso, Luigi, Paola Sapienza, and Luigi Zingales. 2006. "Does Culture Affect Economic Outcomes?" *Journal of Economic Perspectives* 20 (2): 23—48.

Hatton, Timothy J., and Jeffery G. Williamson. 1994. "What Drove the Mass Migrations from Europe in the Late Nineteenth Century?" *Population and Development Review* 20: 1—27.

Kroeber, Alfred L. 1940. "Stimulus Diffusion." *American Anthropologist* 42 (1): 1—20.

Roland, Gerard. 2004. "Understanding Institutional Change: Fast-Moving and Slow-Moving Institutions." *Studies in Comparative Institutional Development* 38 (4): 109—131.

Sokal, Robert R. 1988, "Genetic, Geographic and Linguistic Distances in Europe." *Proceedings of the Natural Academy of the United States of America* 85: 1722—1726.

Spolaore, Enrico, and Romain Wacziarg. 2009. "The Diffusion of Development." *Quarterly Journal of Economics* 124 (2): 469—529.

Stark, Oded. 1991. *The Migration of Labor*. Oxford: Basil Blackwell.

Taylor, J. Edward. 1986. "Differential Migration, Networks, Information, and Risk." In Oded Stark (ed.), *Migration, Human Capital and Development*. Greenwich, CN: JAI Press: 147—171.

Wen, Bo, Hui Li, Daru Lu, Xiufeng Song, Feng Zhang, Yungang He, Feng Li, Yang Gao, Xianyun Mao, Liang Zhang, Ji Qian, Jingze Tan, Jianzhong Jin, Wei Huang, Ranjan Deka, Bing Su, Ranajit Chakraborty, and Li Jin. 2004, "Genetic Evidence Supports Demic Diffusion of Han Culture." *Letters to Nature* 431(16): 302—305.

Wijsman, Ellen, Gianna Zei, Antonio Moroni, and Luigi L. Cavalli-Sforza. 1984. "Surnames in Sardinia II. Computation of Migration Matrices from Surname Distributions in Different Periods." *Annual of Human Genetics* 48: 65—78.

Zimmerman, Klaus F., and Thomas K. Bauer. 2002. *The Economics of Migration*. Mass. Northampton: Edward Elgar Publishing Inc.

责任编辑:田　青