文化差异与移民：中国历史上文化因素对移民长期影响的考察（960-1982）*

李楠

上海财经大学 经济史学系

摘　要：文化作为影响移民活动的重要决定因素已不是一个新的视角，然而对两者关系进行定量分析并揭示其因果关系目前依然缺乏。本文将通过中国历史上的移民数据，采用新的文化差异度量指标，考察地区间文化差异对移民数量的影响。研究发现：采用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，即移民数量随着地区间文化差异的缩小而增加。并且这一结果具有一定稳定性，即使在控制其他社会、经济、地理因素以及采用地区间战争频率差异作为工具变量进行两阶段回归后估计结果依然稳定显著。本研究不仅为理解文化差异与移民之间关系提供了扎实的实证证据，同时对理解中国自十世纪以来移民特征及其决定因素有所帮助。

关键词：文化扩散、文化差异、移民、中国

一、引　言

大规模的人口迁移作为影响人类社会发展的重大现象已被社会科学家关注了很久。[①] 特别是近代工业革命以来，随着工业化和贸易的发展，大规模的移民运动不仅改变了劳动力和财富的地理分布，也促进了许多新兴大国的出现和多民族国家的形成。然而在现有有关移民决定因素的讨论中，多数学者（如 Carrington et al, 1996; Hatton and Williamson, 1994; Lewis,1954; Sjaastad, 1962; Todaro,1969 等）将经济因素作为解释移民的主要原因忽略了对他非经济因素的关注，而在这些被忽略的因素里文化便是其中之一。

文化作为影响移民行为的决定因素已不是一个新的话题,但无论在理论还是在实证研究方面，对两者关系的讨论依然相当脆弱（Epstein and Gang, 2010）。[②]这一方面固然有文化定义广泛、抽象且难以量化的原因（Guiso et al,2006）；另一方面则是缺乏可以反映文化变迁的长期数据（Grosjean, 2011; Roland, 2004）。因此，本文将采用遗传生物学的工具作为地区

* 在此非常感谢哈佛大学经济系 Jeffery Williamson 教授、伦敦经济学院马德斌教授、香港科技大学龚启圣（James Kung）教授、苏基朗（Billy So）教授、韩丽助理教授对本文具有建设性的建议和评论。当然问责自负。

① 正如Zimmermann and Bauer (2002）所说，人类历史就是一部移民史，移民对经济社会发展具有重要的影响。故在过去的二十年里，有关移民的研究不断涌现，其中具有代表性的如Hatton and Williamson (1994)、Lewis (1954)、O'Rourke and Williamson (1999)、Ranis and Fei (1961)、Sjaastad (1962)、Todaro (1969)等。

② Epstein and Gang (2010)在名为 "Migration and Culture"的文章中强调了文化对移民行为的重要性，但同时也指出在现有移民的研究中，文化同移民之间的关系依然是一个"黑箱"（black box），缺乏讨论。

间文化差异的指标,并结合中国历史上的长期移民数据来揭示文化差异与大规模人口迁移之间的因果关系。最终希望本研究可以丰富现有文献，同时克服现有研究的局限为揭示文化差异对移民影响提供扎实的实证证据。

选择中国作为研究对象的理由是：首先，中国幅员辽阔，地理环境复杂，地域文化的多元性为考察地区间文化差异对移民的影响提供了机会。其次，中国历史的连续性、长期性和稳定性可以对文化变迁过程更好的考察。再次，尽管中国历史上的移民活动多数是在省际之间进行，但无论从移民规模还是空间距离，这些移民活动都远远超过历史上欧洲和美洲的移民活动。中国历史上的大规模人口迁移不仅具有代表性,而且是移民研究中不能忽略的问题。[①]最后，对中国移民的研究对理解整个中国社会经济发展起到了非常重要的作用（葛剑雄等，1992)。中国历史上的移民活动不仅导致了人口重心和财富分布在空间上从北（黄河流域）向南（长江流域）的转换；而且随着移民活动（特别是边疆移民)，使中国的疆土得以开拓，最终导致了一个东方大国的产生。因此，对中国历史移民的分析，不仅可以揭示文化差异同移民的关系，也可以帮助我们进一步理解不同时期的移民特征。

文化差异对移民行为的作用主要体现在文化差异所产生的经济外部性对迁移预期的影响。一方面地区间文化差异可能造成移民的障碍。特别是迁出地同迁入地之间由于宗教矛盾与价值观差异可能阻碍两地间的人口流动,比如以色列与阿拉伯国家或者冷战时期西方世界同前东欧国家间的移民等（Dustmann and Preston, 2001; Dustmann and Preston, 2004; Dustmann et al., 2004 等)。而另一方面地区间文化趋同又会促进两地间人口的流动。这主要是因为相似的地域文化可以帮助迁移者降低由于语言、信息不对称等所造成的交易成本和迁移风险，从而增加迁移福利的预期(Akerlof , 1997; Stark, 1991; Taylor, 1986)。例如，Chiswick and Mill（1995）研究发现在迁入地有相似宗教、语言的移民在劳动力、房屋租赁等市场上获得较多的收益，更容易找到工作和住房。

尽管在理论方面有关于文化差异与移民关系的讨论,但是在实证方面却较少足够的实证证据加以证明，特别是可以揭示文化差异与移民之间的因果关系。故本文将采用 Carrington 等（1996)、Epstein and Gang (2010)等学者的理论框架，结合中国的历史移民经验来讨论文化差异对移民的影响。本文的假说认为：随着历史上文化在地域间不断扩散，地区间的文化差异被改变；而作为结果，当受到促使人口流动的外力（主要是北方游牧民族入侵）作用时，地区间文化差异对地区间移民呈现较为显著的负向影响,即人口将会流向具有相似文化的地

① 例如，仅在明朝初期的移民活动中，规模就超过了一千万人（参见，曹树基，1997)；而在十九世纪中叶到二十世纪初的满洲移民，规模更大，超过了欧洲爱尔兰大移民和美国“西进运动”（参见，Gottschange, 1987; Gottschang and Lary, 2000)。

区，地区间较小的文化差异将会导致较大规模的移民。

为检验该假说，中国科学院遗传与发育生物研究所根据不同时期中国人姓氏分布构造的古代移民率和地区间基因距离，分别用来衡量了中国古代地区间的移民程度和文化差异，并且通过移民双边模型来识别地区间文化差异对人口迁移的影响。估计结果表明：利用基因距离度量的文化差异同地区间移民率有显著的负向关系，即随着文化差异的缩小促进了人口迁移的增加。并且这一结果在控制地区间发展水平、地理因素等差异以及两阶段回归后依然稳定，而且与其他稳定性检验结果相一致。

本研究的实证结果不仅揭示了文化差异同移民之间关系，同时也对认识中国历史上长期移民特征有所帮助。首先，从移民的方向来看，在一千年的大规模人口迁移过程中，人口的主要流动方向是从北到南方，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，在移民规模方面，中国历史上分裂割据时期的移民规模要大于统一时期的移民规模。这些发现进一步支持了历史学家（如葛剑雄等，1992；曹树基，1997a, 1997b；吴松弟，1997）对中国历史上移民特征的论断，并为其提供了实证证据。

本研究的价值体现在以下几方面：首先，与以往只关注经济因素或较少采用实证方法的研究不同，这里不仅关注非经济因素的文化对移民的影响，同时提供了检验两者因果关系的实证证据。其次，本研究为理解中国历史上如何实现财富和人口重心从北向南转移以及为何中国可以成为一个幅员辽阔的国家提供了一个机会。最后，在实证研究方面，本文不仅提供了一个新的文化差异度量指标，同时这一尝试也为讨论历史上文化差异与经济发展之间关系的相关研究提供了可行的实证工具。

本文结构安排如下：在文章的第二部分，主要对历史背景及中国历史上文化扩散现象进行简要介绍，并且结合这些背景给出基本假说；在第三部分，主要对实证模型、变量设定及数据来源、实证策略等进行介绍；在第四部分，相应的估计结果及稳定性检验将被给出并讨论；最后是本文的总结。

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

大规模的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（公元前220年前），中国土地上就有了最早的移民活动[①]，但在中国历史上影响最重要的几次大规模人口迁移事

① 较有影响如殷商从商汤到盘庚时期一系列的迁都活动等。

件都发生在公元四世纪以后。[①]主要原因主要体现在：首先是人口压力迫使人口迁移的结果。随着北方经济中心地位的加强和政治中心的确立，北方中原地区成为经济政治中心，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，约五分之四的人口生活在秦岭淮河以北的地区，仅五分之一人口居住在南方地区（邹逸麟，2007）。即使到隋唐时期，北方人口的优势依然十分突出，北方地区人口约占全国总人口的 75%左右（冻国栋，2000）。人口密度的增加在古代社会以农业为中心的经济模式作用下，人口压力必然迫使北方农民外迁寻找更多生存机会。其次是持续不断的战争和社会动荡的结果。在中国历史上，汉代以前汉族农耕文明同草原文明虽然有冲突，但是大致双方可以达到平衡状态，但是在接下来的一千年里（从公元四世纪到十四世纪），中原王朝实力衰弱，双方进行了旷日持久的战争。而战争的主要发生地点在北方人口密集的地区。为躲避战祸，大量北方农民逃离家园选择南方定居，如江西、福建等。截止到十四世纪，仅 20%的人口生活在中国北方，而其余将近 80%人口生活在南方，这与秦汉乃至隋唐时期的人口分布形成了鲜明的对比（范玉春，2005；葛剑雄等，1993）。又如，由于元末明初，大部分战争发生在中国的北方和西南地区，这些地区人口迅速下降。在南宋嘉定十六年（1223 年），四川全境约有 259 万户，而到元末则不足 10 万户，下降了 96.1%（吴淞弟，1997）。

（二）移民中的文化扩散

人口的流动及交往的结果必然是地区间文化交流与融合出现，文化扩散现象产生。文化扩散一直被用来描述文化在地区或人群之间传播的现象，比如宗教、技术、语言的扩散等（Kroeber, 1940）。然而对历史上文化扩散现象本身是难以观察的，因为文化本身难以识别的原因。因此这里只能按照某些历史学家和生物学家（如葛剑雄等，1992；Cavalli-Sforza and Feldman, 1981 等）的建议，选择一些具有代表性的文化元素（如方言、宗教、价值观等）来观察地区间的文化交流。

例如，作为中国最基本文化特征的儒教，虽然起源于山东、河南一带并在秦汉以后进一步在北方地区发展，但是在宋代“靖康之难”以后，随着北方文人涌入南方，儒教的中心也从北方迁移到了南方，并在这里获得了新的发展，促进了新儒学的产生。在明代，这一变化更加突出，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，战争和社会动荡引起了文化的流动，从而改变了文化在地理空间的分布，地区间的文化差异也被改变。由于方言也是重要的文化元素，因此语言学家通常通过地区间语

① 这些大规模人口迁移主要包括西晋“八王之乱”、唐代“安史之乱”、宋代“靖康之难”的大规模人口流动，同时也包括明代洪武大移民以及近代东北大移民等。

言距离来展现文化差异的变化。在晋代以前，中国的汉语言边界以秦岭、淮河为主要分界线，但这一界线在经历了晋代“八王之乱”、唐代“安史之乱”、宋代“靖康之难”三次大规模移民后迁移到了长江中下游地区（邓晓华，王士元，2009）。以上这些均反映出在传统社会中，文化扩散对地区间文化差异的影响。

（三）文化扩散与移民的假说

地区间文化的交流必然改变地区间的文化差异，而这种新形成的文化差异又将对下一期的移民活动产生影响。例如，在明末清初四川饱受战争蹂躏，人口大幅度削减，因此四川成为江西、湖南、湖北移民的主要迁入地。明末清初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。而相似的文化为这些地区在清代中期继续向四川移民提供了的可能。

根据中国的历史的经验，本文力图揭示受文化扩散影响的地域间文化差异对移民行为的影响。本文的基本假说是：在移民的过程中文化因素对移民的决定扮演了重要的角色，文化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，两地区间较近的文化差异将导致更多的移民。

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，移民重力模型将作为实证模型的基本框架。然而在传统的移民重力模型中未考虑文化差异对移民的影响，因此新修正的移民重力模型设定为：

$$\ln mig_{ij} = \alpha \ln cult_{ij} + \beta X_{ij} + \varepsilon_{ij} \cdots\cdots (1)$$

这里 $\ln mig_{ij}$ 为从地区 i 到地区 j 的移民数量的对数值；$\ln cult_{ij}$ 为两个地区之间的文化差异；X 为一组与移民行为相关的控制变量，包括地区间的人口规模、经济发展水平的差异以及地理距离等。$\alpha, \beta, \varepsilon$ 为估计系数和随机扰动项。这里关心的是系数 α 的变化，该系数直接体现文化差异对移民的影响。根据假说该系数的符号预期为负，即 $\frac{\partial \ln mig_{ij}}{\partial \ln cult_{ij}} < 0$。

（二）变量及数据来源

为检验本文的假说，我们根据中国历史上的移民和社会经济信息构造了唯一一套跨时约一千年（从十到二十世纪，约 1000 年）的省际面板数据。选择这一时期作为主要观察时段

的原因主要是：首先，中国历史上大规模的移民都发生在这一时期里，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；最后则是在这一千年里中国经历了一个由分裂到统一的交替过程，这为讨论中国历史上的移民特征提供了较好的机会。

1、被解释变量

由于古代社会在人口流动方面的统计制度不完善和数据缺乏，地区间移民数量的直接数据不能获得。因此，通过历史文献中保存的大量姓氏资料构建的移民率来近似度量不同地区间的移民差异。① 这里使用的移民率数据来自于中国科学院遗传和发育生物研究所。该所科学家根据中国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 年地区间移民率。②

2、解释变量

由于文化非常难以度量，所以社会科学家经常通过不同的变量作为文化的代理变量。如多数经济学家和历史学家往往采用滞后一期的移民存量、方言以及宗教作为文化的度量。③ 虽然这些指标可以捕捉部分文化特征，但是这些指标依然存在一定的局限，例如不能对历史上长期文化变迁过程进行有效度量。

最近一部分经济学家开始注意基因距离在理解文化差异与经济发展之间关系的重要性（如 Ashraf and Galor, 2012; Giuliano 等，2006; Spolaore and Wacziarg, 2009 等）。因此，本文将采用基因距离作为地区间文化差异的度量。但与其他研究不同的是，本研究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用来度量地区间的文化差异。

姓氏群体遗传学主要是认为姓氏与 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的分布就相当于同一基因座上等位基因的分布。因此比较不同群体之间姓氏的分布可以知道不同群体之间的基因距离。本研究所采用的基因距离也来自中科院遗传与发育生物研究所。该所的生物学家利用保存在《宋人传记资料索引》和《明人传记资

① 最早强调用姓氏来研究移民的是 Charles R. Darwin 的儿子 George Darwin。他于 1857 年第一次强调姓氏可以作为一种基因符号可以被用来考察近亲结婚以及地区之间移民等问题(Darwin, 1875)。因此，自从十九世纪中叶以来，姓氏群体遗传学称为生物学的一个重要的分支，并且大量生物学家从事通过姓氏来研究近亲婚姻和移民问题。

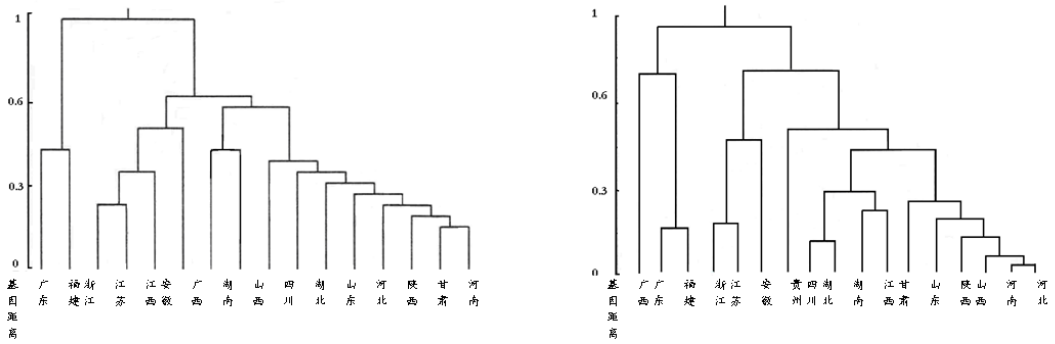
② 上世纪末，中科院遗传与发育生物研究所沿袭 Bodmer 和 Cavalli-Sforza (1968) 以及 Wijsman et al. (1984) 的方法，通过宋代到现今(1982)的人物传记和人口普查资料计算出从宋代至1982年的地区之间的迁移率（袁义达，张诚，2002)。

③ 例如 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(参见, Barro and McCleary, 2003, 2006; Falck et al., 2010; Guiso et al., 2003, 2006 and 2009; McCleary and Barro, 2006; Tabellini, 2007, 2010)采用宗教、方言及价值观和信仰作为文化的度量指标。

料索引》两个文献，通过计算不同时期不同省份之间同姓率的概率分布，分别得到了宋代和明代各省之间的基因距离。①

图 1 描述了宋代和明代地区间的基因距离。在宋代（见图 1.A)，河南与甘肃之间具有最短的基因距离（=0.06)，而山西同福建具有最长的基因距离（=0.93)。但是到了明代（见图 1.B)，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)。这种基因距离的改变不是外生的，正是长时期文化流动和扩散的结果（范玉春，2005)。

图1：宋代和明代各省之间的基因距离



A. 宋代各省际之间的基因距离　　B. 明代各省际之间的基因距离

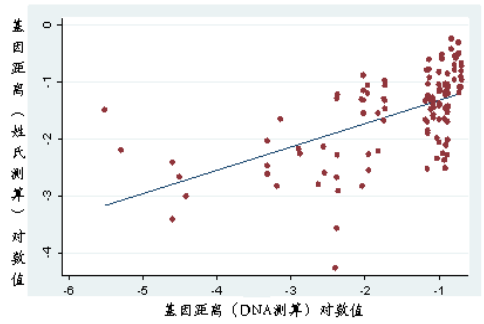
资料来源：袁义达等（1999）“宋朝中国人的姓氏分布与群体结构分布”,《遗传学报》，26（3)：187-197；袁义达等（2000）“中国人群体遗传II：姓氏传递的稳定性与地域人群的亲缘关系”,《遗传学报》，27（7)：565-572。

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；Sokal,1990；Wen, et al. 2004 等)，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，以证明利用姓氏基因距离的有效性；然后，再利用作为重要文化标志的方言构建的语言距离同利用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。②图 2.A 和 2.B 展示了基本分析结果，从中可以看到他们之间均表现出显著的正向相关关系。这些证据进一步证明通过姓氏基因距离可以作为文化差异的有效度量。

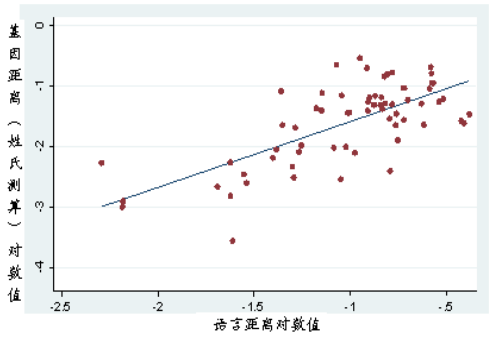
图2：基因距离（根据姓氏和DNA测算）与方言距离的相关分析

① 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)、袁义达等（2000)。

② 这里采用的姓氏和DNA测算的基因距离数据主要来自Du等（1991)；而语言距离数据则来源于陆致极（1992）中的（第四章）表 4和6。



A.采用DNA测算的基因距离与方言距离的相关分析



B. 采用姓氏测算基因距离与方言距离的相关关系

资料来源：A.有关采用1982年DNA和姓氏测算的基因距离数据主要来自Du, et al.（1991）.“Chinese Surnames and the Genetic Differences between North and South China.” Department of Genetics, Stanford University, School of Medicine 94305, Paper number 0027; B.方言距离数据主要来自于陆致极（1992）《汉语方言数量研究探索》，北京：语文出版社。

3、控制变量

除文化差异外，其他因素也可能影响移民行为，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中。这里人口和地区面积数据分别来自曹树基、谭其骧和吴松弟等历史人口地理学家的相关研究。[①] 根据这些人口和地区面积数据最终计算出不同省份在不同时期的人口密度。其次，除人口密度外，战争和自然灾害也是经常被历史学家经常考虑的移民决定因素。故地区间的战争和自然灾害发生频率差异也作为控制变量。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计》（宋正海，1992）。此外，由于文化变迁是一个缓慢的过程，而采用基因距离度量的文化差异可能相对于整个中国历史而言，仅仅是一个短时期的影响。因此为了考虑长时期文化差异对移民的影响，另外一个文化变量也被采用，即该地区是否具有文化遗址（如果有=1）作为控制变量加入模型中。文化遗址分布信息主要来自于《中国历史地图集》(Cheng and Hsu, 1980)。[②]最后，地理因素也被作为控制变量，特别是地区间地理距离在一个标准的移民重力模型中是非常重要。因此，本研究主要控制三个地理因素：第一个是移民方向虚拟变量，主要是移出地是否是北方地区（是=1）；其次是迁入地是否是边疆地区（是=1）；最后两个地区之间的欧几里德地理距离（Euclidian distance）被控制，地区之间的地理距离数据来自谭其骧主编的《中国历史地图集》。前两个虚拟变量主要考察中国历史上移民的主要特征，而空间地理距离则主要考察地区之间移民的相对成本。参与回归的主要变量统计描述在表 1 给出。

① 宋代的人口数据主要来自吴松弟（2000）；明清时期的人口数据则来自曹树基（2000a, 2000b）；采用的地理数据则出自谭其骧(1982)。

② 在本研究中，文化遗址仅仅包括新石器时代之前的文化遗址。

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
地区之间移民率 (x 10^{-4})	480	160.4	518.2	0	1692
地区之间基因距离(采用姓氏测算) (x 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

在估计结果报告前，可能影响估计结果的实证问题需要被讨论。特别是如何纠正由于缺失变量、联立关系以及度量误差所产生的内生性估计偏差。这里将从下三个方面来克服内生性估计偏差并建立文化差异与移民之间的因果关系识别。

首先，针对联立关系所产生影响动态移民重力模型被采用。由于当期的文化差异可能是当期的移民的结果，由此造成识别文化差异与移民之间因果关系的困扰。而若采用动态模型，那么当期的移民只能受上一期文化差异的影响，而不能去影响过去的文化差异，故新的动态重力模型设定为：

$$\ln mig_{ijt} = \alpha \ln cult_{ijt-1} + \beta X_{ijt} + \varepsilon_{ijt} \cdots\cdots (2)$$

这里 $\ln mig_{ijt}$ 表示在时间 t 从地区 i 到地区的 j 移民率的对数值； $\ln cult_{ijt-1}$ 是两地区之间滞后一期的文化差异的对数； X_{ijt} 为一组影响移民的社会、经济和地理变量作为控制变量； $\alpha, \beta, \varepsilon$ 为估计系数和随机扰动项。

其次，为剔除缺失变量对估计结果产生的偏差，工具变量法被采用。一个有效的工具变量应该同内生变量具有很强的相关关系，但是同被解释变量没有直接的影响。对于本研究而言，有效的工具变量应该对文化差异具有较强的相关关系，但是同移民率没有直接的影响。因此，地区间滞后两期的战争频率差异被用作工具变量识别由缺失变量所产生的估计偏差。其识别逻辑是：在上千年的社会发展过程中，相对于当期移民活动滞后两期（在本研究中滞后一期即滞后 600 年）的地区间战争频率差异对当前的移民率没有直接影响，但是却对滞后一期的文化差异产生影响。[①]如果这个假说成立，那么滞后两期的地区间的战争频率差异将

① 工具变量有效性考察以及第一阶段回归结果在表 A.1 (见附录)给出。在表 A.1 前两列中，工具变量对移民率没有直接的影响，但是在表 A.1 后两列（第一阶段回归）中，滞后两期的战争频率同滞后一期的文化

外生于当期的移民率。两阶段回归方程设定如下：

$$\ln mig_{ijt} = \alpha \ln \hat{cult}_{ijt-1} + \beta X_{ijt} + \varepsilon_{ijt} \cdots\cdots (3)$$

$$\ln cult_{ijt-1} = \gamma \ln warfreq_{ijt-2} + \varphi X_{ijt} + \sigma_{ijt-1} \cdots\cdots (4)$$

这里方程（3）同方程（2）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（4）获得的估计值。而在第一阶段方程（4）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$\ln warfreq_{ijt-2}$为相对于$t-1$期文化差异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。

最后，两方面的度量误差对估计结果的影响也被考虑。一方面度量误差是来自解释变量。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，但是同其他指标一样不能完全反应文化特征。因此另外两个同用姓氏基因距离具有高度相关性的变量被用来度量文化差异，即根据 DNA 计算的基因距离和语言距离。另外，由于缺乏历史移民数据，尽管本文使用移民率作为移民程度的度量，但是被解释变量依然可能存在一定的度量误差。因此，这里不仅采用全样本进行考察，同时也采用小样本进行估计一致性检验；与此同时，本文根据曹树基（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，利用地区间移民宗族和移民村庄数量差异作为新的被解释变量，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 2 报告了根据回归方程（2）采用最小二乘法（OLS）的估计结果。[①]在第一列对移民率同地区间文化差异进行回归，在没有控制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的对数值每增加一个单位，移民率将会减少 0.22%。而在接下来的第二和第三列，给出了添加控制变量的估计结果。其中在第二列分别控制住地区间的战争和自然灾害频率、人口密度、地理因素等的差异以及时间虚拟变量（用来度量统一和分裂时期移民的差异（统一时期=1））等。此时，基因距离的估计系数依然为负且统计上显著（显著水平小于 1%）。而在第三列，两个地理因素虚拟变量被加入模型，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（是

差异（内生变量）具有显著的正向相关关系 。所有的回归结果表明在本研究中战争频率是一个有效的工具变量。

① 由于本研究采用移民重力模型作为主要的实证框架，因此在模型估计中，带有多边聚类调整后的稳定性标准误差被给出（相关方法参见Cameron, Gelbach, and Miller, 2011）。

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 2 的最后两列中，作为控制长期文化扩散影响的虚拟变量被加入模型中，即迁出地是否有新旧石器文化遗址（有=1)。同时，为了提供更加可靠的估计结果，使用全样本和小样本（仅包括明代以后的数据）的估计结果分别给出。此时表 2 中第四和五列回归系数依旧表明姓氏基因距离对移民率有显著的负向影响。同时发现文化虚拟变量对移民也有较强的影响，即具有文化遗址的地方，越容易移民。这一点恰恰支持了 Zimmerman 和 Bauer（2002）有关文明通过移民扩散的观点。

表 2: 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)
移民目的地是否是边疆地区？（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

注释：1. 第一至第五列的被解释变量为通过姓氏计算的移民率；2. 第一到第五列解释变量为姓氏基因距离度量的文化差异；3. 第一到第四列给出了全样本的估计结果，即从宋代到 1982 年；4. 第五列 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；5. 根据 Cameron et al. (2011)的多边聚类方法调整的标准误差被给出。6. 括号内为稳定性标准误差。7. *** 显著水平为 1%; ** 显著水平为 5%; * 显著水平为 10%.

此外，从表 2 的回归结果也可以得到其他有效信息。首先，与传统有关移民决定因素认识一致的是地区间的人口密度和战争、自然灾害频率差异对移民有显著的影响，成为引起移

民的主要动力。其次，根据时间虚拟变量系数结果显示，分裂时期更容易导致移民，其原因在于分裂时期战争不断，这样会更加促进农民背井离乡躲避战争的威胁。最后，比较小样本（仅用明代后期数据）和全样本的估计结果时，地理虚拟变量表明：总体上看，在中国历史上移民的主要方向是从北向南，但是明代之后，特别是随着大一统国家的形成，边疆移民成为主要的移民方向。所有这些特征均与历史学家的考察基本一致（曹树基，1997a, 1997b）。

（二）工具变量回归结果

表 2 的回归结果仅展示了文化差异与移民之间的相关关系而非因果关系的考察，而且这些结果还可能受到潜在缺失变量所引起的内生性估计偏差的威胁。因此，地区间滞后两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用来识别文化差异同移民之间的因果关系。

带有工具变量的两阶段估计结果在表 3 给出。从表 3 第一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为负数并且统计显著，并且用姓氏基因距离度量的文化差异系数表明文化差异的对数值每增加一个单位，地区间的移民率将会减小 0.21%。而在第二和第三列控制住地区间人口密度、战争频率、自然灾害频率等差异以及地理因素后，该系数依然为负数且在统计上显著。未得到最终一致的估计结果，同表 2 识别策略一样，在表 3 最后两列分别检验了长期文化扩散对移民的影响以及分别用全样本和小样进行估计。估计结果均表明：在采用了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地负向相关关系，而且将表 3 同表 2 的回归系数进行对比发现，采用工具变量后基因距离的系数有所增加但估计系数无显著差异。这表明内生性估计偏差未对估计结果产生致命影响。所有证据都表明，地区间文化差异越小，移民数量越多，并且通过这一千年的数据分析，在中国历史上文化对移民具有持续的作用。

（三）稳定性检验

为进一步获得稳定的估计结果，这一部分对解释变量和被解释变量中存在的度量误差进行考察。为了消除解释变量中可能存在的度量误差，两个新的数据被用来替代姓氏基因距离作为文化差异新的代理变量。第一个是采用 DNA 计算的基因距离作为文化差异度量指标，数据来自 Du et al (1991)在上个世纪初的研究；另一个则是根据陆致极(1992)提供的中国地区间语言距离作为文化差异的度量。由于期望这些指标可以替换前面所采用明代的姓氏基因距离所测算的文化差异，因此在得到正式估计结果之前先将三个变量之间相关关系进行讨论，结果在附表 A.2 给出。附表 A.2 表明姓氏基因距离与 DNA 基因距离、语言距离之间存在较强的正向相关关系。因此，采用 DNA 基因距离和语言距离可以替代姓氏基因距离作为文化

差异的代理变量。

表 4 提供了仅采用明代以后数据分析的稳定性检验结果。在表 4 前两列，分别报告了仅控制了地区间地理距离情况下，采用 DNA 基因距离和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相关关系。在接下来的第三和第四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系数依然与前两列估计结果相一致。在最后的两列带有工具变量的两阶段回归结果被给出，估计结果也皆与前面的估计结果相同。

表 3: 文化差异与移民率：工具变量回归结果

	地区间移民率的对数值				
被解释变量	(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)
移民目的地是否是边疆地区？			0.033	0.009	0.414*
（是=1）			(0.149)	(0.150)	(0.226)
移民出发地是否为北方地区？			0.434***	0.480***	0.581***
（是=1）			(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

注释：1. 第一列至第五列的被解释变量为通过姓氏计算的移民率；2. 第一到第五列解释变量为姓氏基因距离度量的文化差异；3. 第一到第四列给出了全样本的估计结果，即从宋代到 1982 年；4. 第五列 仅给出了小样本的估计结果，即仅仅包括从明代到 1982 年；5. 工具变量为之后两期的各地战争频率的差异；6. 根据 Cameron et al. (2011)的多边聚类方法调整的标准误差被给出。6. 括号内为稳定性标准误差。7. *** 显著水平为 1%; ** 显著水平为 5%; * 显著水平为 10%.

此外，为了纠正被解释变量的误差，从地方志整理出来移民宗族和移民村落两个微观数据也被采用。这里分别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

息同采用姓氏和 DNA 测算的文化差异进行相关分析。[①]结果在附图 A.1 和 A.2 给出。所有的回归结果表明文化差异与移民之间存在显著的负向相关关系。因此，可知估计误差对先前估计结果没有影响。

表 4: 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)
移民目的地是否是边疆地区?			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

注释：1. 第一至第六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. 第一、第三和第五列解释变量为 DNA 基因距离度量的文化差异，而第二、第四和第六列解释变量为根据方言距离计算的文化差异；3. 第一、第四列给出了 OLS 估计结果，带有工具变量的 2SLS 估计结果在第五和第六列给出；4. 第五和第六列中的工具变量为滞后两期的各地战争频率的差异；5. 根据 Cameron et al. (2011)的多边聚类方法调整的标准误差被给出；6. 括号内为稳定性标准误差；7. *** 显著水平为 1%; ** 显著水平为 5%; * 显著水平为 10%.

五、结论

文化差异作为影响大规模移民的决定因素已经被给予了长期的关注。但由于文化本身难于识别和量化的原因，最终导致有关文化差异与移民之间关系的定量实证研究比较缺乏。因此，本研究使用唯一一套中国历史上的长时期人口迁移数据，通过构建移民重力模型并采用新的文化度量指标（即基因距离）检验文化差异对移民的因果影响。

① 由于微观移民族谱和移民村落信息获得有限，总共获得了不足 30 个样本，因此不能进行有效的回归分析，只能进行叫粗略的相关关系检验。

研究发现在控制住一系列社会、经济、地理等因素后，采用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，地区间的移民率越大。并且在通过工具变量法剔除了潜在内生性偏差对估计结果的影响后，该结果依然稳定。由此使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。

此外，本研究的发现不仅为识别文化对移民的影响提供了扎实的实证证据，同时也为理解中国历史上的大规模移民活动提供了一个量化的动态图景。实证结果揭示了中国历史移民的一下特征：首先，人口密度和战争对移民有显著的影响是移民的主要动力；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；最后，就移民方向来看，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,边疆移民成为主要的移民方向。所有这些均为今天中国的地理疆域以及人口分布格局的形成奠定了基础。

参考文献

Akerlof, G. A. 1997: "Social Distance and Social Decisions." *Econometrica,* 65 (5): pp. 1005-27.

Ashraf, Quamrul, and Oded Galor,"The 'Out of Africa' Hypothesis, Human Genetic Diversity, and Comparative Economic Development." *American Economic Review*, forthcoming.

Barro and McCleary, 2003: "Religion and Economic Growth." *American Sociological Review*. Oct., 68: pp. 760-81.

Battaglia, V., Fornarino, S. et al., 2009: "Y-chromosomal Evidence of the Cultural Diffusion of Agriculture in Southeast Europe." *European Journal of Human Genetics* 17: pp. 820-30.

Bodmer, W. F. and Cavalli-Sforza, L. L., 1968: "A Migration Matrix Model for the Study of Random Genetic Drift." *Genetics* 59: pp. 565-92.

Cameron, C., Gelbach, J. B., and Miller, D. L. 2011. "Robust Inference with Multi-way Clustering." *Journal of Business and Economic Statistics.*) 。

Carrington, William J., Detragiache, Enrica, and Tara Vishwanath., 1996: "Migration with Endogenous Moving Costs" *American Economic Review*, Sept.: 909-30.

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

Cheng, K. Y. and Hsu, S. M., 1980: *Historical Atlas of China*, Taiwan: Chinese Culture University Press.

Chiswick, B. C. and P. W. Miller., 1995: "The Endogeneity between Language and Earnings: International Analyses." *Journal of Labor Economics*, Vol. 13 (2): pp. 246-88.

Darwin, C. R.., 1875.*Insectivorous Plants*. London: John Murray.

Du, Roufu, Yuan Yida, Juliana Hwang, Joanna Mountain, and Li Luca 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, C. and Preston, I., 2001: "Attitudes to Ethnic Minorities, Ethnic Context and Location Decisions." *Economic Journal*, 111 (470): pp. 353-73.

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

Dustmann, C., Francesca, F. and Preston, I., 2004: "Ethnic Concentration, Prejudice and Racial Harassment of Minorities." CReAM Discussion Paper 05/04.

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

Falck et al., 2010: "Dialects, Cultural Identify, and Economic Exchange." IZA, working paper.

Giuliano, Paola, Spilimbergo, Antonio, and Tonon, Giovanni., 2006: "Genetic, Cultural and Geographical Distances." IZA working paper, DP No. 2229.

Gottschange, T., 1987: "Economic Change, Disasters and Migration: The Historical Case of Manchuria." *Economic Development and Cultural Change* 35 (3), pp. 461-90.

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

Grosjean, P. 2011: "The Weight of History on European Cultural Integration: A Gravity Approach." *American Economic Review: Papers & Proceedings*, 101 (3): 504-508.

Guiso, L., Sapienza, P., and L. Zingales., 2006: "Does Culture Affect Economic Outcomes?" *Journal of Economic Perspectives*, 20 (2): pp. 23-48.

Guiso et al., 2003: "People's Opium? Religion and Economic Attitudes." *Journal of Monetary Economics*, 50: pp. 225-82.

Guiso et al., 2009: Does Cultural Affect Economic Outcome?" *Journal of Economic Perspectives*, 20 (2): pp. 23-48.

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

Kroeber, Alfred L., 1940: "Stimulus diffusion." *American Anthropologist* 42 (1), Jan. - Mar.: pp. 1–20.

Lewis, W. A. 1954: "Economic Development with Unlimited Supplies of Labor." *The Manchester School of Economic and Social Sties* 22: pp. 139-91.

O'Rourke, K. H. and Williamson, J. 1999: *Globalization and History: The Evolution of A Nineteenth-Century Atlantic Economy*, Chapter 7, Cambridge, Massachusetts: The MIT Press.

Ranis, G. and J. C. H. Fei.. 1961; "A Theory of Economic Development." *American Economic Review* 51: pp. 533-565.

Roland, Gerard., 2004: "Understanding Institutional Change: Fast-Moving and Slow-Moving Institutions." *Studies in Comparative Institutional Development*, 38 (4): pp. 109-31.

Sjaastad, Larry A. , 1962; "The Costs and Returns of Human Migration." *Journal of Political Economy*, Vol. 70, No. 5: 80-93

Sokal, Robert R., 1988: "Genetic, Geographic and Linguistic Distances in Europe." *Proceedings of the Natural Academy of the United States of America*, Vol. 85: pp. 1722-6

Sokal, Robert R. et al., 1990: "Genetics and Language in European Populations." *American Naturalist*, Vol. 135: pp. 157-175.

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

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

Taylor, J. E., 1986: "Differential Migration, Networks, Information, and Risk." In Stark, O. (ed.) *Migration, Human Capital and Development*. Greenwich, CN: JAI Press, pp. 147-71.

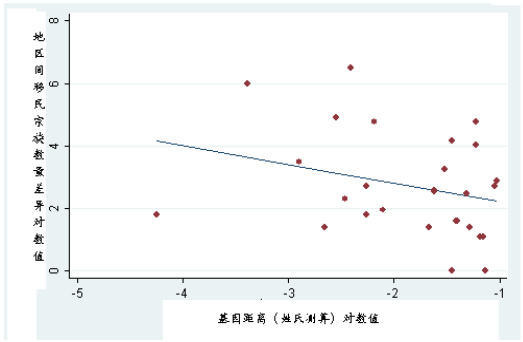
Tabellini, Guido. 2010: "Culture and Instiution: Economic Development in the Regions of Europe." *Journal of the European Economic Association*, June 8 (4): pp. 677-716.

Wen, Bo, Li, Hui, Lu, Daru, et al., 2004: "Genetic Evidence Supports Demic Diffusion of Han Culture." *Letters to*

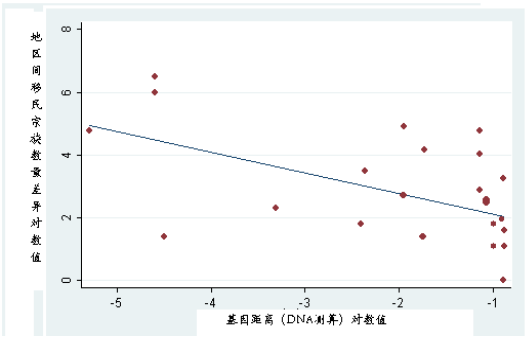
Nature 431(16): pp. 302-5.
Wijsman, E., Zei, G., Moroni, A., Cavalli-Sforza, L. L., 1984: "Surnames in Sardinia II. Computation of Migration Matrices from Surname Distributions in Different Periods." *Annual of Human Genetics* 48: pp. 65-78.
Todaro, M. P., 1969: "A Model of Labor Migration and Urban Unemployment in Less Development Countries." *American Economic Review* 59: pp. 138-48.
Zimmerman, K. F., and Bauer, T., 2002: *The Economics of Migration*, Mass. Northampton: Edward Elgar Publishing Inc..
曹树基，1997a:《中国移民史》（明時期），福州：福建人民出版社。
曹树基，1997b：《中国移民史》（清时期），福州：福建人民出版社。
曹树基，2000a:《中国人口史》（第四卷），上海：复旦大学出版社。
曹树基，2000b：《中国人口史》（第五卷），上海：复旦大学出版社。
邓晓华，王士元，2009：《中国的语言及方言的分类》，北京：中华书局。
冻国栋，2000：《中国人口史》（第二卷），上海：复旦大学出版社。
范玉春，2005：《移民与中国文化》，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》，北京：解放军出版社。
葛剑雄，曹树基，吴松弟，1992：《移民与中国》，香港：中华书局出版社。
葛剑雄，曹树基，吴松弟，1993：《简明中国移民史》，福州：福建人民出版社。
何炳棣，1969：《黄土与中国农业的起源》，香港：香港中文大学出版社。
陆致极，《汉语方言数量探索》，北京：语文出版社，1992。
宋正海，1992：《中国古代中大自然灾害和异常年表总集》，广州：广东教育出版社。
谭其骧，1982：《中国历史地图集》，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》，《长水集》，北京：人民出版社。
吴松弟，1997:《中国移民史》（宋辽金元时期），福州：福建人民出版社。
吴松弟，2000：《中国人口史》（第三卷），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》，《遗传学报》，26 (3): 187-97。
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》，《遗传学报》27 (7): 565-72。
袁义达，张诚，2002：《中国姓氏：群体遗传和人口分布》，上海：华东师范大学出版社。
赵文林，谢淑君，1988：《中国人口史》，北京：人民出版社。
邹逸麟，2007：《中国历史地理概述》，上海：上海教育出版社。

附　　录

图 A.1 地区间移民宗族数量差异同文化差异之间的相关分析



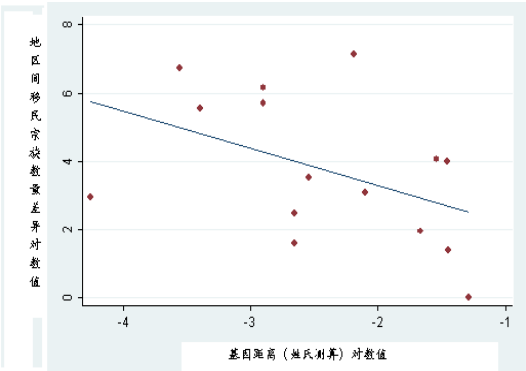
A. 基因距离（姓氏）同移民宗族数量的相关分析



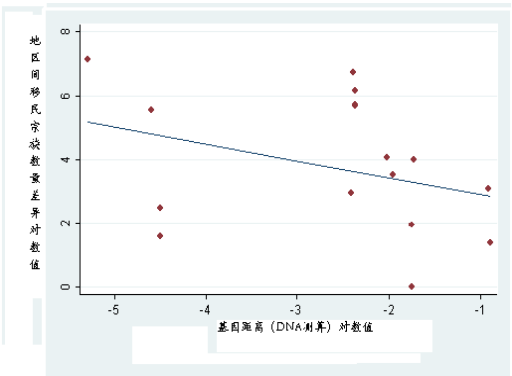
B. 基因距离（DNA）与移民宗族数量的相关分析

资料来源：A. 地区间移民宗族数量差异数据主要来自曹树基（1997a）《中国移民史》，福州：福建人民出版社；B. 姓氏测算和DNA测算的基因距离数据分别来自袁义达等（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》，《遗传学报》，27（7）：565-572和Du, et al. （1991）“Chinese Surnames and the Genetic Differences between North and South China.” Department of Genetics, Stanford University, School of Medicine 94305, Paper number 0027.

图 A.2 地区间移民村落数量差异同文化差异的相关分析



A. 基因距离（姓氏）与移民村数量的相关分析



B. 基因距离（DNA）与移民村数量的相关分析

资料来源：A. 地区间移民宗族数量差异数据主要来自曹树基（1997a）《中国移民史》，福州：福建人民出版社；B. 姓氏测算和DNA测算的基因距离数据分别来自袁义达等（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”，《遗传学报》，27（7）：565-572和 Du, et al.（1991）“Chinese Surnames and the Genetic Differences between North and South China.” Department of Genetics, Stanford University, School of Medicine 94305, Paper number 0027。

表 A.1 工具变量有效性检验及第一阶段回归结果

	移民率对数值	移民率对数值	第一阶段回归 基因距离
被解释变量	(1)	(2)	(3)
解释变量			
滞后两期的战争频率差异对数	-0.049	-0.041	0.237***
	(0.133)	(0.138)	(0.057)
控制变量			
移入地区是否为边疆（是=1）	0.408***	0.373***	-0.091
	(0.135)	(0.133)	(0.067)
移出地是否为北方地区（是=1）	0.274**	0.438***	0.427***
	(0.126)	(0.128)	(0.059)
地理距离对数值	-0.265	0.011	0.723***
	(0.182)	(0.187)	(0.083)
是否为统一时期（是=1）	-0.642***	-0.663***	-0.055
	(0.152)	(0.151)	(0.065)
基因距离对数值		-0.382***	
		(0.100)	
常数项	5.047***	7.593***	5.942***
	(0.284)	(0.782)	(0.129)
观测值	422	422	422
F-统计量	8.57	9.8	38.36
R-squared	0.096	0.123	0.346

注释：1. 第一和第二列被解释变量是从宋代到1982年用姓氏计算的移民率；2. 第三列被解释变量是宋代和明代姓氏基因距离；3. 第一至第三列估计方法是固定效应的OLS，但固定效应系数未报告；4. 根据Cameron et al. (2010)的多边聚类方法调整的标准误差被给出；5. 括号内为稳定性标准误差；6. *** 显著水平为 1%; ** 显著水平为 5%; * 显著水平为 10%。

表 A.2: 明代基因距离（姓氏）、基因距离（DNA）、语言距离的相关系数

	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	0.327*** (0.000)	1.000	
明代基因距离 (姓氏)	0.598*** (0.000)	0.535*** (0.000)	1.000

注释：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%.