

# 中国的姓氏、籍贯和长期代际流动性 (1645—2012)

郝煜\*

**摘要** 本文拓展了 Clark *et al.* (2015) 提出的用姓氏测算社会流动性的方法, 以姓氏和籍贯的组合来识别精英, 测算了中国核心区域不同时期的社会流动性。本文的发现是: 第一, 清代大部分时期的代际相关性较高 (0.73—0.85), 即社会流动性较低。第二, 民国和新中国时期, 代际相关性有所下降, 但是幅度有限 (0.65—0.75)。第三, 现代中国的代际相关性, 显著高于常规方法基于个人和家庭水平数据测算的代际相关性 (0.3—0.6)。

**关键词** 姓氏, 籍贯, 代际流动性

**DOI:** 10.13821/j.cnki.ceq.2021.03.13

## 一、引言

民间智慧 (conventional wisdom) 对于社会流动有着生动却似乎矛盾的看法: 一方面, 人们说, “龙生龙, 凤生凤”; 另一方面, 又有所谓 “君子之泽, 五世而斩”, “富不过三代”。这些矛盾的看法, 是反映了时段社会流动的特点进而表明社会经济制度对社会流动的影响, 还是反映了社会流动的不同侧面和维度? 和大多数计量历史研究要解决的问题一样, 这是一个经验问题。

常规的社会学和经济学方法, 基于随机调查 (占此类研究的大部分) 和人口普查数据, 以个人和家庭的某个可观测指标 (如教育、职业、收入等) 为观测单位, 计算样本的代际相关性或弹性, 或代际转换矩阵 (transition matrix, 给出任意一个阶层的后代流动到另一个阶层的概率)。代际相关性也被称为代际弹性, 在本文中简称为  $\beta$ ,  $\beta$  值越大意味着越小的社会流动性, 或者社会阶层越固化。Corak (2013)、Chetty *et al.* (2014) 的发现是, 在不同的社会经济制度下,  $\beta$  的差异很大 (0.2—0.8), 且和静态的不平等程度高度相关。对此主流的解释是高度不平等的情况下, 低收入阶层在信贷约束下人力资本投资不足 (Becker and Tomes, 1979, 1986); 而经济不平等导致的政

\* 北京大学经济学院。通信地址: 北京大学经济学院 423 室, 100871; 电话: 18335139672; E-mail: maxhao1003@pku.edu.cn。作者感谢北京大学经济学院青年种子基金的资助, 感谢高岚清同学的助研工作, 以及匿名审稿人的宝贵意见和建议。文责自负。

治不平等,也通过对公共教育的投资不足导致较高的 $\beta$  (Ichino *et al.*, 2011)。此类研究的不足在于数据很少超过三代(跨多代研究见 Long and Ferrie, 2013)。

研究历史上的社会流动性则受到数据的局限,难以获得足够随机的样本。一个自然的妥协,就是研究社会金字塔的顶端,即“精英背景研究”。自潘光旦和费孝通(1947)、He (1962)对进士举人家世的研究,到梁晨等(2012)对近现代北京大学学生背景的考察,他们关心的问题是“布迪厄”假说:在一个注重选贤任能(meritocracy)的教育和选拔体制下,精英阶层是不断被洗牌,还是代代被“再生产”出来?但是,此类研究的致命伤在于跨期和跨区的可比性:当代北京大学学生30%来自干部家庭,明清进士60%来自三代以内有正途科举功名的家庭,这意味着当代的 $\beta$ 更低么?大学生和士绅如何可比?60%和 $\beta$ 如何可比?更一般地,金字塔顶端的流动是否可以代表整个社会的一般流动率。<sup>1</sup>

Clark (2015)和Clark *et al.* (2015)提出的姓氏方法在常规方法和精英研究之间架起了桥梁,为研究跨期跨区可比的长期社会流动提出了一种思路。具体来说,从文献和档案中找到长时期特定群体(通常是精英)的姓氏分布,比如某个时期全国进士的名单。如果能恰当地估计姓氏的人口分布,就可以计算一个衡量姓氏地位的指标“相对精英率”(relative representation among elites, 简称为 $RR_{kt}$ ):

$$RR_{kt}(\beta) = \frac{\text{姓氏 } k \text{ 在时期 } t \text{ 的精英样本所占比例}}{\text{姓氏 } k \text{ 在时期 } t \text{ 总人口所占比例}}$$

在一定的假设下(见第二部分),如果一个姓氏的 $RR=1$ ,意味着该姓氏的平均社会地位和总人口一致。如果某些姓氏在 $t=0$ 显示出很高的 $RR$ ,那么它们的 $RR$ 应该随着时间下降;反之,如果某些姓氏在 $t=0$ 显示出很低的 $RR$ ,那么它们的 $RR$ 应该随着时间上升。社会地位的代际相关性 $\beta$ 越弱,社会流动性越强,这些偏离均值的姓氏就更快地“回归均值”。

Hao and Clark (2016)提出了一种把姓氏方法应用于中国的方法,计算了一个“关于地域调整”的相对精英率,以江南地区的稀见精英姓氏的相对精英率为分子,以江南地区的大姓的相对精英率为分母,模拟了这些精英姓氏在地域内回归均值的过程。本文采用了一种不同的方法,不是把姓氏,而是把姓氏-籍贯组合作为观测对象。通过模拟从清代到当代江南和华北的精英姓氏-籍贯组合的升降过程,本文的发现是中国几个世纪以来的代际相关性不但在数值上(0.65—0.75)高于常规方法得到的结果(0.3—0.5),而且跨时期跨地区的差异也小于根据常规方法得到的 $\beta$ 的变动范围。这个发现和Clark

<sup>1</sup> 一个可能存在的样本选择问题是,进入金字塔顶端人群的家庭背景和遗漏变量(残差)存在负相关:越是家庭背景差的人,越需要较高的天赋和运气进入金字塔顶端;只研究金字塔顶端忽略了那些天赋和运气低于平均水平的人。因此基于精英样本这个截断样本(truncated sample)计算得到的 $\beta$ ,有可能低估代际相关。

等人的一系列发现是一致的。

如何解释这个违反直觉的发现呢？Clark *et al.* (2015) 提出了“真实社会地位”（true underlying social status, 简称 TSS）这个潜在变量（latent variable）。在个人或家庭水平上，TSS 可以看作是一个多维度的变量，收入、财富、教育、职业、官阶甚至族群，共同构成了一个个人或家庭的真实社会地位（图 1）。但是，常规方法往往计算的是某个单一维度变量（partial measure）的代际相关性，但是由于任何单一维度变量在“表达”TSS 时必然存在误差，比如暴发户的财富往往高估了其 TSS，姑且定义为“表达误差”（expression error）。即使那个变量的测量误差很小，所得的也仅仅是某单一维度变量的代际相关性，而这往往会低估 TSS 的代际相关性。附录 A 证明了 TSS 的代际相关性可以在群组（姓氏）水平上测算出来。

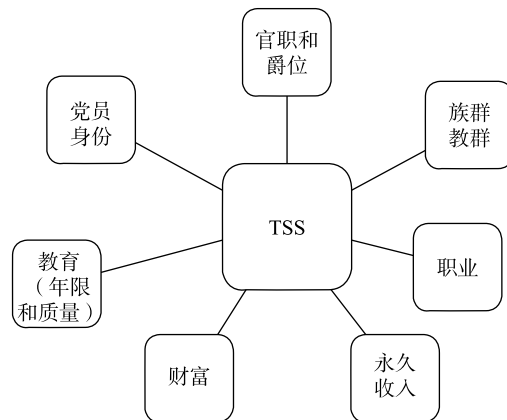


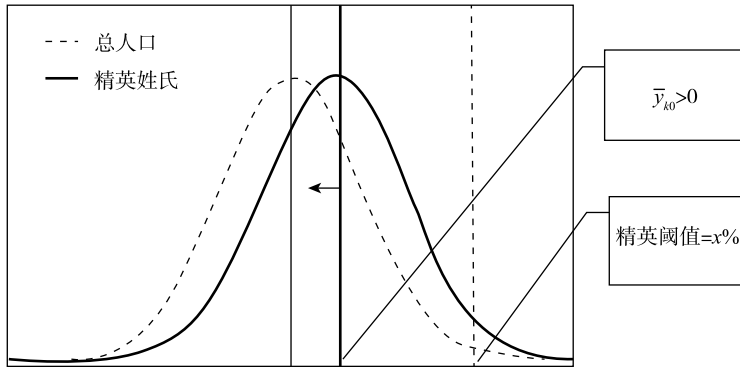
图 1 真实社会地位 (TSS) 的维度

## 二、姓氏方法

从历史数据通常只能看到**精英**（比如进士或举人）的姓氏分布随时间的变化。要从精英的姓氏分布的代际变动推断全体人口的代际相关性，需要做一些假设：

首先，总人口的某单一维度变量服从正态分布  $(0, \sigma^2)$ ，而“精英”姓氏  $k$  的该变量也服从同方差的正态分布，只不过在初始阶段，其均值高于零， $y_{k0} \sim N(\bar{y}_{k0}, \sigma_{k0}^2)^2$ ，如图 2 所示。在附录 B 中，我们基于清代同光时期地方志所载忠烈殉难录中的姓氏分布数据和 2005 年人口普查数据，对这两个假设做了一点初步的检验，发现正态分布假设和同方差假设是符合的。

<sup>2</sup> 有效的推断依赖于正态分布和同方差的假设。如果假定双头分布，或者初始方差小于人口方差，那么拟合的 *RR* 值会在第二期大幅下降或大幅上升，不符合观察到的 *RR* 值的时间趋势。Clark and Cummins (2014) 和 Clark (2015) 验证了这一点。

图2 变量  $y$  的分布

注：图中实心曲线为精英姓氏人口中变量  $y$  的分布，虚线曲线为总人口中变量  $y$  的分布，加粗直线为精英姓氏人口中变量  $y$  的初始均值，细直线为总人口中变量  $y$  的均值，虚线直线为总人口的精英阈值，即变量  $y$  取值最高的前  $x\%$ 。

其次，假定代际传递过程满足  $y_{it} = \beta y_{it-1} + v_{it}$ ，其中  $v_{it}$  是随机扰动项。只要  $0 < \beta < 1$ ，可以证明其分布逐渐逼近普通人口的分布：

$$y_{kt} \sim N(\bar{y}_{kt}, \sigma_{kt}^2) \xrightarrow{t \rightarrow \infty} N(0, \sigma^2),$$

即所谓“回归均值”。 $\beta$  越大，这个逼近过程越慢，该姓氏的相对精英率  $RR$  的下降过程越慢。其中，

$$\begin{aligned} RR_{kt}(\bar{y}_{k0}, \beta) &= \frac{\text{姓氏 } k \text{ 在精英样本所占比例}_t}{\text{姓氏 } k \text{ 在总人口所占比例}_t} \\ &= \frac{(\text{姓氏 } k \text{ 贡献的精英人数}_t / \text{精英总人数}_t)}{(\text{姓氏 } k \text{ 人口}_t / \text{总人口}_t)} \\ &= \frac{(\text{姓氏 } k \text{ 贡献的精英人数}_t / \text{姓氏 } k \text{ 人口}_t)}{(\text{精英总人数}_t / \text{总人口}_t)} \end{aligned}$$

也就是图2中姓氏  $k$  落在精英阈值（虚线直线）右侧的比例相对于总人口落在精英阈值右侧（即精英阈值 =  $x\%$ ）的比重。不失一般性，把  $\sigma^2$  标准化为1。给定一系列  $RR_{k1}, RR_{k2}, RR_{k3} \dots$ ；给定每一期的精英阈值  $x_1, x_2, x_3 \dots$ ，找到非线性最优解  $(\hat{\bar{y}}_{k0}, \hat{\beta})$  使得  $RR$  拟合路径和  $RR$  实际值的距离总和最短。<sup>3</sup>

精英在全体人口中的阈值（top  $x\%$ ）是一个可控参数。本文对不同历史时期讨论的精英样本是不同的，其在人口当中的比例也不同。中国进士和举人代表的上层绅士包括家庭成员约占人口的  $0.1\% - 0.2\%$ （Chang, 1955, p. 115）。民国的大学生群体及其家庭成员约占人口的  $1\%$ 。1949年以来的大

<sup>3</sup> 这个估计方法其实本质是一种最小二乘估计：如果我们假定  $\beta$  是不变的，时间序列一共有  $t$  期，那么相当于是从  $t$  个数据点估计两个参数  $(\hat{\bar{y}}_{k0}, \hat{\beta})$ 。

学生比例缓慢上升，而20世纪90年代后期的大学扩招之后，“985工程”院校的大学生比例达到了每年适龄人数的3%。在下文的实证模拟部分，本文将根据不同时期“精英”占人口的比重调整各期的精英阈值。最后，笔者把基于RR的姓氏方法的操作流程总结如下：

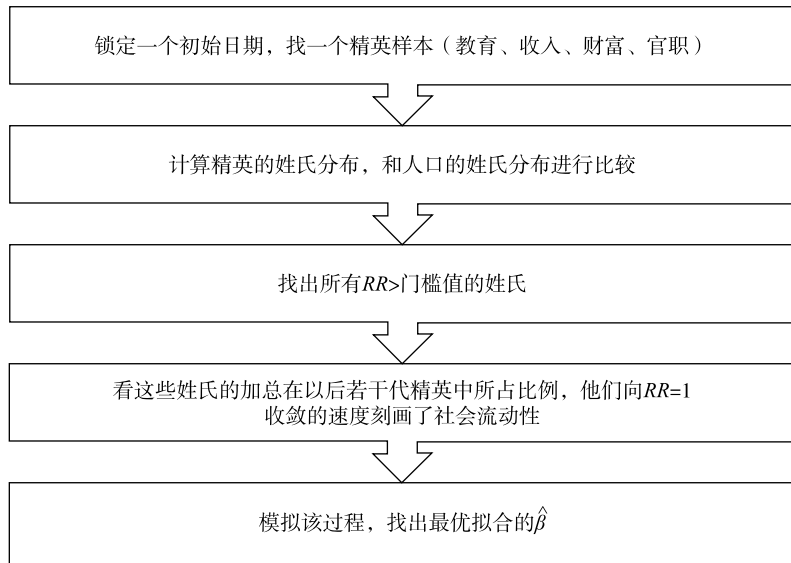


图3 基于RR的姓氏方法的操作流程

### 三、姓氏-籍贯方法

#### （一）姓氏方法应用于中国的困难

姓氏方法能够应用于中国么？在英国，类似 Darcy 这样人口不过数百人的贵族姓氏数以千计，很容易从初始阶段识别出 RR 值高于 100 的姓氏，其中很多是来自欧洲大陆的贵族姓氏。而中国的姓氏历史悠久，随着人口的繁衍和迁徙、民族的融合，源出多头，支派纷繁，早已失去了标示身份的功能。中国姓氏的另一特征是姓氏稀少，这是因为相比采用字母的语言，中国通常采用一到两个字作为姓氏，而常用的汉字不过几千个。当代人口排名前 100 的大姓占了当代汉族人口的 87%，明清进士的 86%；排名前 300 的大姓占总人口的 99%，占进士的 98%，所有这些姓氏的人口都在十五万以上。<sup>4</sup>

<sup>4</sup> 中国汉族人口中人数最多的 1500 个姓氏人口的全国数据来源是公安部公民身份信息服务中心；姓氏人口的省级地区分布来源是袁义达和张诚（2002）；姓氏人口的县级分布则根据地方志的各种名录估计。

笔者对在清代出过进士的所有姓氏计算了  $RR$ ，发现有相当多的姓氏的  $RR$  在全国层面上长期高于或低于全国平均水平，比如图 4 显示的江、浙、沪三大地域性姓氏沈、顾、钱的  $RR$  值在整个清代始终高于 2，而全国三大姓张、王、李的  $RR$  值则始终在 1 以下。这是否意味着有某种基于姓氏的歧视和制度障碍阻止了这两个姓氏组的  $TSS$  “向均值回归”？

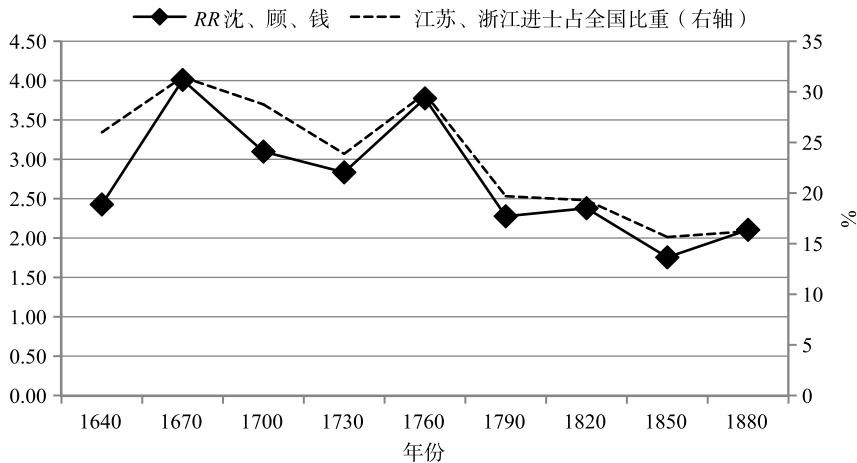


图 4 沈顾钱的相对精英率和江浙两省进士占全国比例

注：由于 1760 年以前各省人口不详，所以无法计算两省在进士的相对精英率。

对中国历史稍加考察，笔者就可以对此做一个初步解释。

第一，中国的姓氏地域集中度较高，比如王姓在全国汉族人口占 7.5%，但在山东省占 12%，在广东仅占 2%。而集中于江南地区的大姓顾、陆、沈、钱、俞等高达 50% 的人口居住在江、浙、沪地区，部分小姓甚至集中在几个地级甚至县级行政单位。

第二，进士在地区之间的名额分配是中央政府控制的，而非自由竞争的结果。在不同时期国家会根据政治需要调整进士名额在各省之间的分配（中央政府通常会给人口、钱粮较多和文风较盛的地区更多名额，但也会照顾京畿地区、偏远地区和少数民族地区）。总的来说，江浙地区获得了高于其人口比例的科举名额（包括生员、举人和进士名额），但相对于其文化教育水平，其科举名额又受到了较多的限制。在图 4 中，笔者看到，沈、顾、钱的  $RR$  值的时间趋势和清代江苏、浙江两省在全国进士比重的时间趋势高度一致。所以，沈、顾、钱的  $RR$  值相对于总人口的变动，只是反映了江浙两省的进士名额在全国总额的占比变动，而非代际流动。如果只看江浙地区，笔者发现沈顾钱的  $RR$  值和该地区平均水平一致。笔者搜集的 1905 年废除科举后的各种精英样本，就地域分配而言遵循不同的原则。例如民国大学生中江浙籍人士的比重高于明清进士（民国时期中央政府对大学考试没有名额分配），但 1949 年以后统一考试制度下大学生的地域来源更接近明清进士的情况。另外，

在当代知识精英和企业精英的样本中<sup>5</sup>，江浙籍人士的比重也高于其在大学生的比重（Hao and Clark, 2016）。

第三，由于各个地区的人口增长速度不同，比如东北和西南地区的人口自然增长率在整个清代快于江浙地区（此外，太平天国、捻军起义、回民起义等造成各地区人口消减的幅度不同），使得当代人口的姓氏分布不能反映历史上人口的姓氏分布，而且越是地域集中的姓氏偏差越大。总之，全国层面上各姓氏的RR值的时间趋势无法反映社会流动。

## （二）姓氏-籍贯方法

在Hao and Clark (2016)中，作者采用了把地域性大姓作为控制基准的方法。<sup>6</sup>本文则把特定的姓氏-籍贯组合，而不是姓氏，作为群组单位，并且把总样本限制在特定地区特定历史时段的精英。具体来说，RR的分子为识别出来的精英姓氏-籍贯所贡献的精英人数占所有来自该地区的精英的比例，而分母为该姓氏-籍贯人口占该地区总人口的比例。<sup>7</sup>

在传统中国社会，姓氏本身不能像在英国那样标示群组的社会地位，但在特定的历史时段加上籍贯就不同了。在魏晋南北朝时期，望族就把籍贯和姓氏结合起来，作为身份象征，如琅琊王氏和清河崔氏。江南地区自五代以降成为主要的移民迁入地，特别是靖康之变后迁来的北方官宦家族，明代以来迁来的徽商闽商家族，都给江南地区带来新的姓氏和宗族（吴仁安，2009），这些家族在经济文化上的优势直到明末清初还很显著。以姓氏-籍贯作为群组单位，笔者在每一期都可以识别出40—60个精英姓氏-籍贯组合，这也使得一些必要的稳健性检验成为可能。<sup>8</sup>

计算各期各姓氏-籍贯的相对精英率RR，面临的挑战分别来自分母和分子。分子方面，如何构建可以跨期比较的精英样本？如何确定模拟中的精英阈值？笔者从地方志搜集了清代江南全部两万多名举人的名单。中举者通过录取率极低的乡试进入上层绅士集团（Chang, 1955）。这些精英（包括其家

<sup>5</sup> 这两个精英样本分别来自排名前六位的大学的在职教师和2006年资产在一亿元以上的企业的法人代表。

<sup>6</sup> 作者首先识别出那些在某一时期全国进士样本中RR高于5，且50%人口居住在江浙沪的姓氏。然后计算不同时段和不同精英样本中这些姓氏的RR相对于沈顾钱的RR的比重： $RR_t(z/\text{沈顾钱}) = \frac{RR_{zt}}{RR_{\text{沈顾钱}t}}$ 。

<sup>7</sup> 本文考察了江南地区和山东省两个地区。其中江南地区涵盖的范围包括江苏的苏州、常州、镇江、太仓三府一州和浙江的杭州、嘉兴、湖州、绍兴、宁波五府。

<sup>8</sup> 这一方法的局限在于移民可能会导致籍贯的变动，既影响笔者对RR分子的估算，也影响对RR分母的估算。不过，至少在清代这个问题不算严重。个人一般隶属于宗族组织，科举参加者多报告祖籍和原籍。迁居以后，一般三代以后才报告现居地为籍贯。根据笔者对苏州清代20000名生员的统计，籍贯变动的情况只占0.2%。古代的地方志和科举名录、烈女节妇名录、近代的同学录、太平天国之后各州县编写的忠烈殉难录，以及国共内战和抗战死难烈士的名录，都报告了人物的县一级籍贯。这使得笔者能够用历史上的县级姓氏分布，而不是当代的人口普查数据的姓氏人口来估算RR的分母。

庭成员) 在清代的江南地区大约占总人口的 0.1% 到 0.2%。在举人样本中的  $RR$  值可以代表一个宗族的社会地位, 亦可以代表其文化水平和财富水平。<sup>9</sup>

另一个挑战是如何估计  $RR$  的分母, 即特定的姓氏-籍贯在江南人口中的比例。笔者对各姓氏-籍贯人口的基准估计来自晚清各县地方志的《太平天国忠烈殉难录》(30 000 多人), 得到了太平天国前姓氏-籍贯组合的人口分布。<sup>10</sup> 也就是说, 本文使用 1850 年左右人口存量的姓氏分布来估测 1645—1670 年的人口姓氏分布。但要根据此回溯清初的姓氏人口, 则需要考虑如下情况: 社会地位高的宗族可能平均出生率高于一般人口, 平均死亡率低于一般人口。笔者将在下一部分检验测算结果是否对改变这一假设稳健。表 1 总结了姓氏-籍贯方法的指标构建及数据来源:

表 1 姓氏-籍贯方法的指标构建及数据来源

	指标构建	数据来源		$RR$ 阈值
$RR$ 的分子	识别出来的精英姓氏-籍贯所贡献的精英人数占所有来自该地区的精英的比例	江苏南部 1645—1905	举人 《江苏选举志》	1645—1670 年=0.2% 1671—1880 年=0.1% 1880—1904 年=0.2% 1900—1930 年=0.5% 1930—1960 年=1% 1960—1990 年=1.5% 1990—2012 年=2%
		浙江北部 1645—1905	举人 民国《浙江通志: 选举志》	
		江南地区 1905—1949	各种民国大学同学录	
		江苏南部 1949—2011	南京大学本科生 南京大学学生信息查询系统	
		浙江北部 1949—2011	著名人物 《浙江历代名人录》	
		山东省 1645—1905	举人 民国《山东通志: 选举志》	
$RR$ 的分母	该姓氏-籍贯人口占该地区总人口的比例	江南地区	旧方志的《太平天国忠烈殉难录》	
		山东省	新方志的《民国忠烈将士姓名录》和《革命烈士录》	

<sup>9</sup> 虽然城市化和商品化使越来越多的精英选择经商置产捐纳监生(所谓异途), 但是考取举人进士仍是光耀门楣的正途。在宗族层面上, 经商可以视为为科举提供经济基础(教育和应考的费用早已不是一般家庭可以承担), 亦可以视为一种风险分散策略。笔者最近根据各地方志记载的捐纳官职的名录(例贡和仕籍), 发现姓氏在举人和捐纳学位者两个精英样本的  $RR$  值相关性很高。

<sup>10</sup> 对于各县地方志中的忠烈殉难录缺失或者样本较小的县市, 本文使用了基于当代人口普查的户籍人口姓氏分布。对于这部分县市, 一个棘手的情况是太平天国的人口损失和之后的移民降低了部分土著姓氏的人口比例(曹树基, 1998), 比如常州庄姓在常州的人口比例在同治和光绪年间下降了 25%。本文根据太平天国人口损失和移民数作了调整, 得到了太平天国前姓氏-籍贯组合的人口分布的估算。



## 四、实证结果

### （一）社会流动（1645—1911）

表2汇报了这一时段测算的代际相关性。笔者从清初和清中期的两个时段，以 $RR > 5$ 为标准，在两个地区分别识别出40—60个精英姓氏-籍贯，然后用第二部分提供的模型模拟了加总之后向均值回归的速度。总的来说，江南地区清中期以前的代际相关性为0.75—0.82，清中期以后的代际相关性为0.79—0.90。<sup>11</sup>

我们担心的问题是这些姓氏-籍贯之间向均值回归的速度可能有很大差异，那么模拟结果在多大程度上是被某几个特别成功或不成功的姓氏-籍贯组合所驱动的？比如常熟翁氏，武进恽氏和慈溪冯氏的 $RR$ 在整个清代并没有长期下降的趋势。为解决这个问题，笔者采用了bootstrap抽样方法，即从每一个精英姓氏-籍贯的样本，构建1000个bootstrap样本，再分别模拟每个样本的 $RR$ 向均值回归的速度，最后计算这些样本构成的95%置信区间。比如在表2第一行第一列，笔者从含有62个姓氏-籍贯组合的“浙北五府1645—1670精英”中，构建了1000个bootstrap样本（每个新样本是上述62个姓氏-籍贯组合的回置抽样，sampling with replacement），依次拟合每个新样本形成的 $RR$ 时间序列，最后计算这1000个拟合值 $\hat{\beta}$ 的95%置信区间。表2报告了每个精英样本通过bootstrap构建的95%置信区间，基准的模拟结果并不是被某几个特别成功或不成功的姓氏-籍贯组合所驱动的。

表2 对不同精英姓氏-籍贯组的拟合结果及95%置信区间

拟合的时间段	浙北五府	浙北五府	苏南四府	苏南四府
	1645—1670	1791—1820	1645—1670	1791—1820
	精英姓氏-籍贯 (62) <sup>1</sup>	精英姓氏-籍贯 (58) <sup>1</sup>	精英姓氏-籍贯 (48) <sup>1</sup>	精英姓氏-籍贯 (40) <sup>1</sup>
1645—1790	0.77 (0.75—0.80) <sup>2</sup>		0.79 (0.76—0.82) <sup>2</sup>	
1791—1905	0.84 (0.79—0.87) <sup>2</sup>	0.84 (0.80—0.87) <sup>2</sup>	0.87 (0.82—0.89) <sup>2</sup>	0.85 (0.83—0.87) <sup>2</sup>

注：（1）括号内是每个精英样本所含有的姓氏-籍贯组合个数；（2）括号内是每个精英样本通过bootstrap构建的95%置信区间。

<sup>11</sup> 清代的科举从1645年开始，直到1905年。因此我们追踪的初始阶段定为1645—1670年，此后每30年一个阶段，正好到1910年是十代。这样做的理由是Chang（1955）估算举人平均中举年龄为30岁左右，因此我们把30年作为一代。

下面,笔者检验测算结果是否对改变下述假设稳健:人口增长率在精英姓氏-籍贯和总人口之间无差异。首先笔者从浙北五府1645—1670年举人识别了59个的精英姓氏-籍贯。假设各姓氏人口自然增长速度一致,那么1850年的姓氏人口分布就可以代表这些1645—1670年的姓氏人口分布。笔者据此画出了加总后 $RR$ 值的时间趋势,并给出了 $\beta$ 的最优拟合值,如图5正方形节点的曲线所示( $\hat{\beta}=0.81$ )。

考虑两种情况,第一种情况假定清初精英姓氏-籍贯的人口增长率和全体人口一致,所以 $RR$ 每期的分母都等于这些精英姓氏-籍贯在近代人口中所占的比例。第二种情况采用每一代精英姓氏-籍贯人口增长快于一般人口10%这一假设<sup>12</sup>,对 $RR$ 每期的分母进行了回溯计算(backward projecting),这样在1645—1670年其人口比例是其近代人口比例的0.45, $RR$ 则是第一种情况得到的2倍以上。但 $\beta$ 的最优拟合值和原假设的结果相当接近( $\hat{\beta}=0.79$ )。可见在一定范围内精英人口增长速度对模拟结果的影响不大。<sup>13</sup>

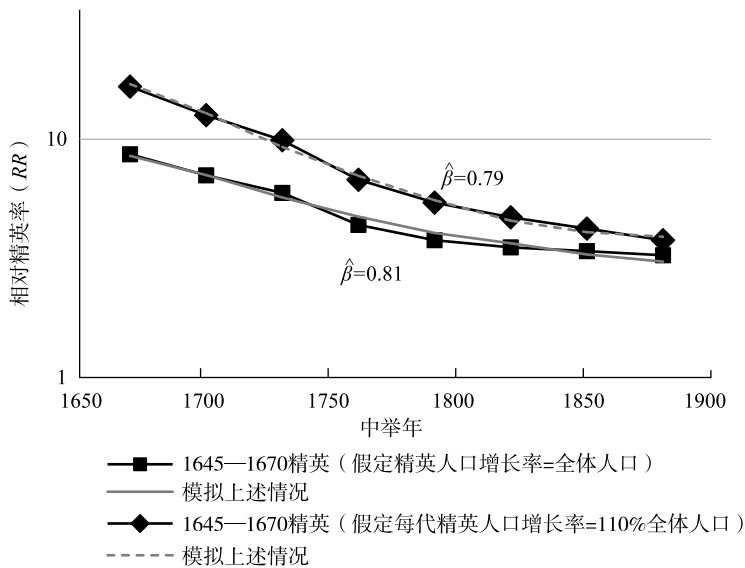


图5 精英姓氏-籍贯人口增长速度对模拟结果的影响

<sup>12</sup> Shiu (2017) 根据元代到清代的桐城族谱,发现士绅阶层的出生率和其他阶层相当接近。但是笔者考察了常州两大科举望族(恽和庄)相对于常州人口的自然增长速度,发现平均每30年,望族的人口增长率比一般人口高5%—10%。而根据海宁查氏和陈氏的族谱(见赖惠敏,2010),尽管士绅阶层的结婚比例和婚内生育率都显著高于平民阶层,但由于士绅阶层即使在那些科举望族族人所占的比例也低于10%,所以望族的总人口增长率不超过非望族的10%。详细的计算过程见 Hao and Clark (2016)。

<sup>13</sup> 具体来说,假定普通人口和精英姓氏-籍贯人口增长速度不同,比如1850年某个姓氏-籍贯在人口当中的比例是5%,但是其人口增长速度高于人口平均水平10%。以30年为一个世代,那么1820年这个姓氏-籍贯在人口当中的比例是5%/1.1,以此类推1790年的比例就是5%/(1.1<sup>2</sup>)。

最后，江南地区的测算结果能多大程度上代表中国的一般情况？Elman (1991) 和 Hymes (1986) 认为族内公产和公共教育导致了科举望族可以维持其优势地位，一定程度上可以抵消由于“诸子分家”造成的精英后代向下流动的趋势，形成某种“诗书世族”。如果这一假说成立，我们可能会看到宗族势力较强的华南和长江下游地区应该比北方显示出更高的代际相关性或者说更低的社会流动性。我们利用山东（1645—1905）和陕西（1645—1880）的举人数据和当代人口普查数据，用和上文同样的方法，测算了该地区相应时间段的社会流动性。我们发现代际相关性拟合值分别为 0.812 和 0.769（图 6），可见宗族对地域社会支配程度的差异，并没有造成社会流动的显著差异。我们的结果至少能很大程度上反映中国核心区域的社会流动状况。

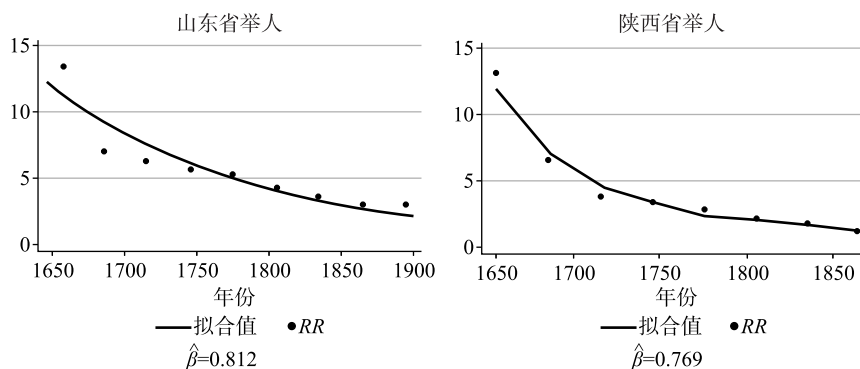


图 6 北方两省清代初期的科举精英 ( $RR_{\text{举人}} > 5$ ) 姓氏—籍贯的回归均值过程及模拟结果

总之，整个清代的代际相关性（0.77—0.87）接近英国中世纪到工业革命前的水平（Clark, 2015, p. 86），这意味着从精英选拔的角度看，选贤任能和贵族世袭之下的社会流动没有显著差异。清中期以前的代际相关性为 0.75—0.82，清中期以后的代际相关性为 0.79—0.90，这和 He (1962) 的结果是一致的。

## （二）社会流动（1905—2012）

废除科举是否造成科举时代精英的迅速没落？近现代一系列激进的再分配政策对社会流动有何影响？笔者从 1880—1905 年的江苏和浙江举人中，识别出了所有  $RR$  值大于 5 的精英姓氏—籍贯，其中浙北五府 62 个，苏南四府 41 个。在这一部分，我们把江苏南部和浙江北部分开处理，理由是江浙两省在中华人民共和国成立后的精英数据来源不同。<sup>14</sup> 所以本文分别对两组精英姓氏—籍贯做了加总，计算了  $RR$ ，并拟合了  $RR$  回归均值的过程，结果见图 7。其中括号内为通过 bootstrap 抽样得到的 95% 置信区间。民国时期的精英来自

<sup>14</sup> 浙江采用了《浙江历代名人录》，该书报告了 8 000 多名入选人的出生地和祖籍；江苏采用了南京大学学生信息查询系统，检索了该校 1952 年以来所有学生的姓名籍贯信息。

笔者搜集的多部大学毕业生名录。<sup>15</sup> 民国时期的江南地区, 大学生约占人口的 0.1%—0.5% (根据民国二十三年教育年鉴、民国三十五年江苏省和浙江省经济年鉴计算)。1949—1966 年和恢复高考后的 1977—1997 年, 大学生的人口比例仅仅提高到了 1.5%。1998 年之后大学扩招, 但精英大学的招生规模变化不大, 笔者的模拟根据不同时期样本的性质调整了精英阈值。<sup>16</sup>

总之, 对两个地区 1900—2012 年的模拟结果表明, 代际相关性  $\hat{\beta}$  比清代特别是晚清有一定程度的下降, 范围是 0.61—0.76。值得我们注意的是, 即便是中华人民共和国成立前后出生的两代人之间的代际相关性也没有如我们预测的那样, 被激进的再分配政策降到接近 0 的水平。而如果考虑到精英的改姓, 籍贯的错报, 向中国台湾、中国香港地区以及欧美的移民等因素, 实际的  $\beta$  甚至可能被低估了。<sup>17</sup>

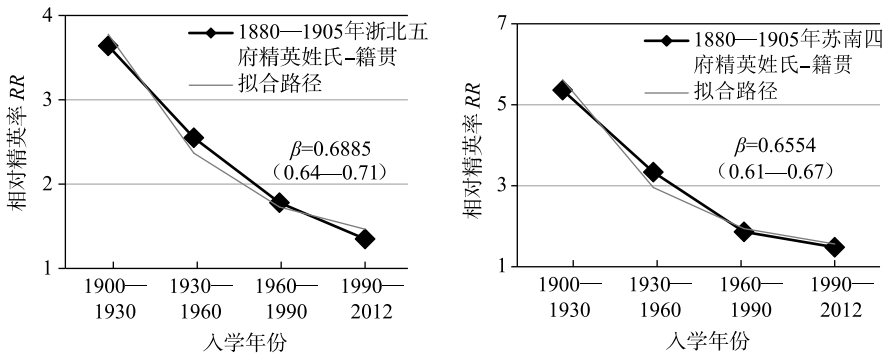


图 7 RR: 1880—1905 年浙北五府和苏南四府精英姓氏及其拟合路径

## 五、结论和比较

当代对于社会流动性的主要研究路径是通过研究代际相关性在地域之间、

<sup>15</sup> 民国时期江南的大学生样本来自笔者搜集到的留日学生、北京大学、清华大学、中央大学、南洋公学 (后为上海交通大学)、大同大学、武汉大学、浙江大学、金陵大学、复旦大学、燕京大学同学录中籍贯为江南地区的学生名单。1949 年以后, 虽然籍贯常常汇报为出生地或户籍所在地, 我们仍然可以找到比较可信的人物名录。

<sup>16</sup> 在相关的研究中, 笔者一直把获得某种学位的人数占总人口比例作为跨期可比的指标, 比如举人占总人口的比例是 0.05%, 民国大学生的占比是 0.1%—0.2%, 1950—1975 年的大学生占比是 1% 左右。这是长时段研究当中不可避免的问题, 因为很难找到科举时代精英在当代的完全对应。这种跨期比较隐含的假定是教育年限和学位在中国一直是获取社会地位的最直接途径 (敲门砖), 也是社会地位的最无争议性的衡量指标 (虽然和科举时代相比, 衡量的标准更加多元化了)。精英阈值分别设为: 1880—1905 年 = 0.2%, 1900—1930 年 = 0.5%, 1930—1960 年 = 1%, 1960—1990 年 = 1.5%, 1990—2012 年 = 2%。

<sup>17</sup> 姓氏方法的有效性依赖于地域流动很小, 且籍贯汇报原籍而不是现居住地这些假设。这个假设至少在太平天国以前的大部分地区是成立的 (湖广川陕的流动性较大), 但是对于太平天国之后可能适用性降低, 这可能导致了第四部分第 (二) 节的结果。如果落后地区的精英迁移到了发达地区, 那么我们观测到落后地区的社会流动率更高 (精英的后代消失), 而且我们观测到发达地区的社会流动率更高 (新的精英出现)。

时期之间和指标之间的差异性，识别差异性背后的决定机制。社会学文献更关注代际相关性的时间趋势变化和不同维度指标的差异。比如，Walder and Hu (2009) 发现：老精英（“地富反坏右”）的后代尽管在中华人民共和国成立后在入学、入党和提干上受到歧视，但平均受教育程度仍然高于贫下中农的后代，并且在医生和工程师等职业上保持优势；而在恢复高考后，随着教育重新成为社会流动的主要渠道，其社会地位得到一定程度的恢复。经济学文献则更关注收入的代际相关性的形成机制。其中微观层面的研究主要关注决定个人或家庭表现（比如在校学习成绩和劳动力市场表现）的先天因素和后天因素（nature vs. nurture）孰轻孰重，以及两者是如何互动的（见 Durlauf (2014) 的综述）。特别的，族群和性别如何影响社会流动具有重要的政策含义，也受到研究者的格外关注（Borjas, 1992）。宏观层面的研究则侧重研究静态的不平等是如何导致动态不平等的，比如 Chetty *et al.* (2014) 基于对全美全部人口的税收数据，绘制了全美流动性的县级地区差异地图，并发现收入的基尼系数和收入的代际流动性高度相关。Benhabib *et al.* (2019) 和 Piketty (2015) 则更强调财富不平等对于代际不平等的作用。Corak (2013) 综述了几十个国家的微观研究，进行了跨国比较，得出了类似的结论。

对于中国历史上的科举和社会流动的关系，量化历史研究者也试图识别其中的因果关系。Chen *et al.* (2020) 发现，科举在唐代并没有实现广泛的社会流动和选贤任能，而在宋代随着商业化的发展，特别是由此带来的出版书籍的普及和私人书院的兴起，科举精英的背景也日益草根化，这标准着门第社会的消亡和“平民社会”的兴起（钱穆，2001）。Bai *et al.* (2019) 也发现，到了南宋，在有大量北方移民迁来的南方地区，投资教育成为对于宗族生存发展至关重要的竞争策略。而到了明清，科举成为获取社会地位的唯一阶梯，大量人力资本和物质资本被投资于科举业。在激烈的竞争中，家庭背景、财富、个人天资以及社会网络都对科举事业的成功至关重要（Kung and Jiang, 2020; Huang, 2016; Zhang, 2013）。但是平民子弟是否仍然有机会通过科举进入精英阶层？有两个实证研究给予了侧面证明：一是 Bai and Jia (2016) 发现，废除科举导致了平民向上流动机会的阻塞和平民参与革命的增加；二是 Chen *et al.* (2020) 发现，明清进士人口比例越高，在控制了当代的收入水平和教育回报之后，当代的教育投资更丰厚，对教育的文化态度也更积极，这意味着科举塑造了当代人关于教育的文化价值观。总的来说，平民的向上流动机会虽然有限但并非渺茫，这和本文的发现，以及 He (1962) 的经典研究都是一致的。但是无论如何，这些研究的着眼点在于解释差异和变化，特别是在科举制下，为什么某些时期或者某些地区的社会流动性会高于另一些时期和另一些地区。

然而，本文的主旨是将一种新的社会流动测算方法应用于中国，并发现

不同社会、不同地区和不同时间段代际相关性的趋同而不是差异。比较姓氏方法和常规的经济学和社会学方法得到的代际流动性的结果<sup>18</sup>，本文发现，第一，代际相关性 $\hat{\beta}=0.6-0.9$ ，这意味着父辈能够解释子女36%—81%的变动幅度，祖父辈能解释孙辈20%—70%的变动幅度，也就是更接近“君子之泽，五世而斩”的情况；而常规方法的结果(0.3—0.6)更接近“富不过三代”的情况。第二，从清代早期到中后期，代际相关性从0.75左右上升到0.85左右，到民国和中华人民共和国成立，代际相关性下降到0.65—0.75。但变化不大，特别是中华人民共和国成立前后出生的两代人之间仍表现出了较强的代际相关性，和改革开放前后出生的两代人之间的代际相关性差别不大。这些发现首先打破了科举制度下社会高度流动的神话，其次显示革命并没有像想象的那样具有彻底的颠覆性(比如 $\hat{\beta}=0$ 或者 $\hat{\beta}<0$ )。

如表3所示，如果和其他国家横向比较的话，中国并不是一个特例。和贵族世袭制的中世纪英国相比(第一行第一列)，有教无类、不拘一格降人才的科举制度并没有带来更高的流动性。和初等教育高度均等化、收入分配高度平等的当代瑞典相比(第六行第三列)，当代中国并没有体现出更低的代际流动性。借用钱穆(2001, p. 54)的观察，“不论古今，世袭社会也有流动，选贤任能社会也有世袭。”

表3 用姓氏方法测算的基于不同精英样本度量的各国 $\hat{\beta}$ 值

时期	英国牛津 剑桥学生	中国中举者 或大学生	瑞典医生	美国律 师和医生	日本精英 大学生	印度非穆 斯林医生	中国台湾 台湾大学学生
1170—1590	0.83						
1645—1905		0.74—0.87					
1700—1900							
1830—2012	0.73—0.82						
1860—1947						0.91	
1900—2011		0.65—0.75	0.74		0.71—0.88		
1950—2011				0.74—0.84		0.88	0.69

资料来源：Hao and Clark (2016)；Clark (2015)，p. 34，p. 38，p. 57，p. 86，pp. 100—101，p. 151，p. 193；Hao (2017)。

<sup>18</sup> 国内的研究多采用 CHNS、CHIPS 等调查数据，在测算收入、教育、职业、健康等指标的代际相关性方面致力于解决数据的样本选择性问题 and 变量的测量误差问题，此外，研究者既关注最近 30 年各指标代际相关性的变化趋势，也关心指标之间的关系，还关心代际相关性水平和变化在城乡间的差异。总的来说，中国的收入代际相关性在 0.3—0.6，比如陈琳和袁志刚 (2012)，Gong *et al.* (2012)，郭丛斌和闵维方 (2007)，Zhang and Eriksson (2010)，Wu and Zhang (2010)，等等。本文的估计显著低于大多数文献基于家庭和个人数据计算的收入(教育)等相关性。

这种相似性意味着，姓氏方法测算的可能是一种“真实社会地位”（TSS）的代际相关性，而不是 TSS 某一个维度的代际相关性。根据第一部分和附录 A，尽管公共政策和社会经济制度对某些指标的代际流动有显著影响，它们对 TSS 的代际流动影响较小。对此一个可能的解释是，真实社会地位的代际变化，可能如 Clark（2015）中大胆假设的，是一种基因意义上的回归均值过程，因为只有基因过程背后的决定因素在社会和时间段之间差异很小，比如婚姻的选择性（assortative marriage）贯穿整个人类有文字记述的历史，并跨越了不同制度和文化。婚姻的选择性使得基因、家庭文化和社会资本在代际稳定地传承，而公共政策对这些因素的再分配非常有限。

## 附录 A：TSS 的代际相关性为什么可以在群组（姓氏）水平上测算出来

首先假定从  $N$  个家庭样本得到两代人的永久收入  $y$ ，并作如下回归。

$$y_{it} = \beta y_{it-1} + v_{it}, \quad (\text{A1})$$

$$x_{it+1} = \beta x_{it} + e_{it}.$$

其中  $y_{it-1}$  是父母的永久收入， $y_{it}$  是子女的永久收入。 $v_{it}$  是正态分布的干扰项。拟合的  $\beta$  越高，永久收入的代际相关性或弹性越高<sup>19</sup>，社会流动率越低。<sup>20</sup>

其次假定永久收入  $y$  和无法直接观测的真实社会地位 TSS 存在如下关系：

$$y_{it} = TSS_{it} + u_{it}, \quad (\text{A2})$$

其中  $u_{it}$  是永久收入对于 TSS 的表达误差；而 TSS 的代际动态由如下方程刻画：

$$TSS_{it} = bTSS_{it-1} + e_{it}, \quad (\text{A3})$$

其中  $u_{it}$  的存在，不在于研究者的收入数据不能够精确度量永久收入，而在于收入和教育不能完全表达 TSS，图 A1 阐明了这个观点。假定父母和子女 TSS 的相关性是  $0 < b < 1$ ，而笔者只能观测到父母和子女永久收入（或职业、教育）的相关性  $0 < \beta < 1$ 。对于有些人来讲，永久收入可能高估了其 TSS（比如“土豪”）；而对于另一些人来讲，永久收入可能低估了其 TSS（比如某

<sup>19</sup> 当两代之间的变量的标准差相等时，代际弹性=代际相关性；大多数社会流动研究通常考虑以收入或教育的排名或分位数代替收入或教育的数值，因此代际的变量标准差相等。

<sup>20</sup> 在数据允许的情况下，近来的研究采用了五年及以上收入的平均值以过滤收入的波动，或采用多个子女收入、职业和教育的平均值以控制子女间收入的取舍问题 trade-off，因而得到了更高的代际相关性，但这解决的是测量误差的问题，而不是表达误差的问题。

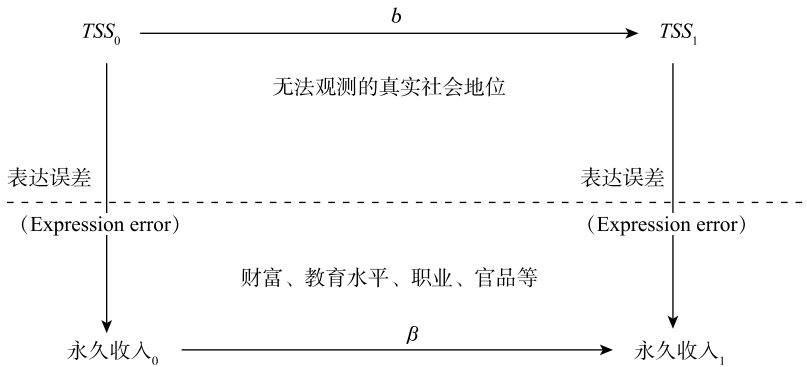


图 A1 真实社会地位 (TSS), 表达误差和单一维度变量

豪门公子去做非公益事业), 因此样本估计的  $\beta$  的期望值会低于  $b$  (attenuation bias), 因为从永久收入的角度看, 土豪相对于上一代实现了向上流动, 而做 NGO 的贵公子则看起来向下流动了, 总的来说, TSS 的流动性被高估了。<sup>21</sup>

从式 (A1)、(A2)、(A3) 可以证明  $E(\hat{\beta}) = b \frac{\sigma_{TSS}^2}{\sigma_u^2 + \sigma_{TSS}^2} < b$  (见 Hao and Clark, 2016, 附录 2)。在个人或家庭水平上, 永久收入的代际相关性将低估 TSS 的代际相关性。而低估的程度, 取决于扰动项 (噪音) 的变动相对于 TSS (信号) 的变动的比率。这样, 有两个推论, 第一, 收入/财富/教育较平等的社会, 低估程度更高。第二, 如果一个社会的收入分配变得更加平等, 低估程度更高。

一个自然的解决方案, 是以群组代替家庭或个人作为观测单位。直观来说, 如果表达误差只在个人水平上存在, 那么只要对该群组的人口加总平均, 那么个人水平上的表达误差, 在群组水平上就加总为零了。静态上, 不管用哪个指标 (收入、教育、职业), 都能准确在群组水平上量度 TSS。也就是说,

$$\overline{y_{kt}} = \overline{TSS_{kt}}. \quad (A4)$$

如果群组 A 的平均 TSS 高于群体 B 的平均 TSS, 那么 A 的平均收入、平均财富、平均受教育水平等都高于姓氏 B。其中  $k$  代表群组,  $\overline{y_{kt}}$  是群组  $k$  内某单一维度变量的平均值。因此动态上, 在群组水平上采用任意单一维度变量  $y$ ,

$$\overline{y_{kt}} = \beta \overline{y_{k, t-1}} + v_{kt}, \quad (A5)$$

其代际相关性  $\beta$  的最优估计值, 也是 TSS 代际相关性的无偏 (unbiased) 估

<sup>21</sup> 如果用教育呢? 也有类似的问题。比如比尔·盖茨, 因其从哈佛辍学, 那么用教育年限所量度的社会地位比他做大学教授的双亲都下降了。另外, 很多贫寒子弟读完了博士, 但社会地位的提升恐怕不如教育年限所显示的那样大, 不然“学术屌丝”“凤凰男”这些不雅的词汇也不会那么流行。职业也是如此, 特定职业的声望在不同时期和地区千差万别, 比如律师和医生在中国和美国社会是不可同日而语的。总之, 通过不同的单一维度变量所得到  $\beta$  对  $b$  的低估, 只有程度的差别 (种姓可能是低估程度最小的指标)。



计： $E(\hat{\beta}) = b$ 。当然，一个必须满足的条件是  $\bar{y}_{k,t-1}$  和  $v_{kt}$  不相关。用族群和教群作为分组依据，一般不满足这一条件。原因在于选择性的代际变动，地位较低的族群和教群后代更倾向于变动为地位较高的族群和教群。比如犹太人在最近两千年收入和教育代际的延续性至少有相当部分是由于贫穷和不识字的犹太人后代大部分被周边阿拉伯人同化了，而当代的犹太人则是一代一代被选择性筛选留下的精英（Clark, 2015）。如果笔者把族群和教群作为观测单位，那么会高估  $\beta$ ，低估社会流动性。

比较而言，姓氏则是一个选择性代际变动较小的群组。虽然，历史上既有社会下层的成功者改姓精英姓氏以提升社会地位，也有社会精英改普通姓氏消灾避祸（从土地改革到“文化大革命”），但是，这种改姓还是非常罕见的，通常不超过人口的1%（Clark and Cummins, 2014）。

## 附录 B：经验事实在多大程度上支持图 2 的关于正态分布的假设

根据 2005 年全国人口普查数据的一个抽样样本，笔者考察了江浙两省三组姓氏的出生于 1920—1949 年和 1950—1979 年两组人口受教育程度的分布状况（图 B1 和图 B2）。江浙三大地域性姓氏：沈顾钱（高度集中于江南地区，60% 的人口至今居住在江浙沪）和该地区全体人口的受教育程度分布基本一致。但是，从 1880—1910 年进士名单中识别出的江浙精英姓氏（ $RR > 5$ ，且 50% 人口居住于江浙沪），则看起来是全体人口分布的一个向右平移（rightward shift）。精英姓氏的中低文化水平比例低于沈顾钱和全体人口，但高中及以上文化程度者的比例则高于上述两者。表 B1 显示，这些姓氏人口受教育程度的标准差和总人口的标准差相差很小。上文所做的关于方差的假设是合理的。最后，尽管受教育年限的分布在代与代之间并不一致，但如果我们用教育的排名或者分位数来代替受教育年限，还是可以把该变量的分布看成是正态分布。

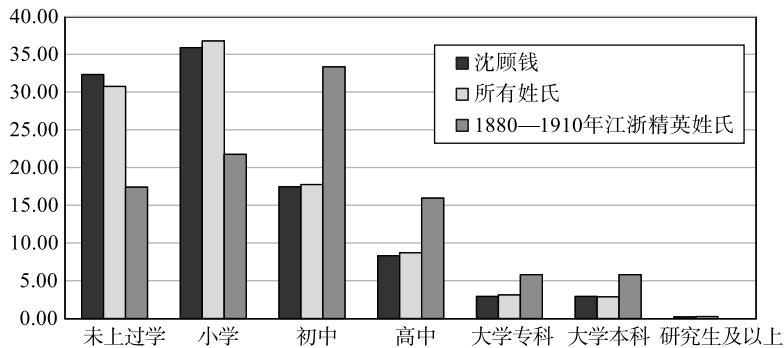


图 B1 三个姓氏组的 1920—1949 年出生人群受教育程度分布

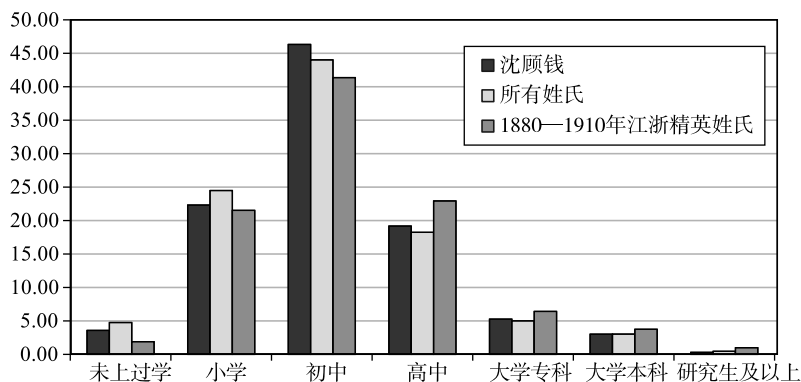


图 B2 三个姓氏组的 1950—1979 年出生人群受教育程度分布

表 B1 江浙两省三个姓氏组的受教育年限

		所有姓氏	沈顾钱	1880—1910 年精英姓氏
1920—1949 年 出生人口	样本量	43 876	2 229	69
	均值	5.69	5.66	8.01
	标准差	4.63	4.71	4.65
1950—1979 年 出生人口	样本量	98 134	4 270	119
	均值	8.96	8.94	9.63
	标准差	3.44	3.32	3.26

## 参考文献

- [1] Bai, Y., and R. Jia, "Elite Recruitment and Political Stability: The Impact of the Abolition of China's Civil Service Exam", *Econometrica*, 2016, 84 (2), 677-733.
- [2] Bai, Y., J. K. Kung, and C. Ma, "Endogenous Confucianism: The Path of Development in Historical China", *CUHK Working Paper*, 2019.
- [3] Becker, G. S., and N. Tomes, "An Equilibrium Theory of the Distribution of Income and Intergenerational Mobility", *Journal of Political Economy*, 1979, 87 (6), 1153-1189.
- [4] Becker, G. S., and N. Tomes, "Human Capital and the Rise and Fall of Families", *Journal of Labor Economics*, 1986, 4 (3, Part 2), S1-S39.
- [5] Benhabib, J., A. Bisin, and M. Luo, "Wealth Distribution and Social Mobility in the US: A Quantitative Approach", *American Economic Review*, 2019, 109 (5), 1623-47.
- [6] Borjas, G. J., "Ethnic Capital and Intergenerational Mobility", *The Quarterly Journal of Economics*, 1992, 107 (1), 123-150.
- [7] 曹树基, "太平天国战争对苏南人口的影响", 《历史研究》, 1998 年第 2 期, 第 64—74 页。
- [8] Campbell, C. and J. Z. Lee, "Kinship and the Long-term Persistence of Inequality in Liaoning, China, 1749-2005", *Chinese Sociological Review*, 2011, 44 (1), 71-103.
- [9] Chang, C. L., *The Chinese Gentry: Studies on Their Role in Nineteenth-Century Chinese Society*

- (Vol. 2). Seattle, U. of Washington P, 1955.
- [10] 陈琳、袁志刚，“中国代际收入流动性的趋势与内在传递机制”，《世界经济》，2012 年第 6 期，第 115—131 页。
- [11] Chen, T., J. K. S. Kung, and C. Ma, “Long Live Keju! The Persistent Effects of China’s Civil Examination System”, *The Economic Journal*, 2020, 130 (631), 2030-2064.
- [12] Chen, T., and J. Kung, “The Rise of Meritocracy”, *HKSTU Working Paper*, 2019.
- [13] Chetty, R., N. Hendren, P. Kline, and E. Saez, “Where Is the Land of Opportunity? The Geography of Intergenerational Mobility in the United States”, *The Quarterly Journal of Economics*, 2014, 129 (4), 1553-1623.
- [14] Clark, G., and N. Cummins, “Intergenerational Wealth Mobility in England, 1858 – 2012: Surnames and Social Mobility”, *The Economic Journal*, 2014, 125 (582), 61-85.
- [15] Clark, G., N. Cummins, Y. Hao, and D. D. Vidal, “Surnames: A New Source for the History of Social Mobility”, *Explorations in Economic History*, 2015, 55, 3-24.
- [16] Clark, G., *The Son Also Rises: Surnames and the History of Social Mobility*. Princeton University Press, 2015.
- [17] Corak, M., “Inequality from Generation to Generation: The United States in Comparison”, *The Economics of Inequality, Poverty, and Discrimination in the 21st Century*, 2013, 107-126.
- [18] Durlauf, S., *Intergenerational Mobility*. Lecture Notes, Beijing Summer Schools on Socioeconomic Inequality, 2014.
- [19] Elman, B. A., “Political, Social, and Cultural Reproduction via Civil Service Examinations in Late Imperial China”, *The Journal of Asian Studies*, 1991, 50 (1), 7-28.
- [20] 郭丛斌、闵维方，“中国城镇居民教育与收入代际流动的关系研究”，《教育研究》，2007 年第 5 期，第 3—14 页。
- [21] Gong, H., A. Leigh, and X. Meng, “Intergenerational Income Mobility in Urban China”, *Review of Income and Wealth*, 2012, 58 (3), 481-503.
- [22] Hao, Y., and G. Clark, “Social Mobility of China, 1645 – 2012: A Surname Study”, *UC Davis working paper*, 2016.
- [23] Hao, Y., “Converging Mainlander and Native Taiwanese, 1949 – 2012”, *Australian Economic History Review*, 2017, 57 (1), 84-107.
- [24] He, B., *The Ladder of Success in Imperial China: Aspects of Social Mobility, 1368-1911*. New York, Columbia U. P., 1962.
- [25] Huang, Y., *Essays in Economic History and Applied Microeconomics*. Doctoral dissertation, California Institute of Technology, 2016.
- [26] Hymes, R. P., *Statesmen and Gentlemen: The Elite of Fu-chou, Chiang-hsi, in Northern and Southern Sung*. Cambridge: Cambridge University Press, 1986.
- [27] Ichino, A., L. Karabarbounis, and E. Moretti, “The Political Economy of Intergenerational Income Mobility”, *Economic Inquiry*, 2011, 49 (1), 47-69.
- [28] Jiang, Q., and K. S. Kung, “Social Mobility in Late Imperial China: Reconsidering the ‘Ladder of Success’ Hypothesis”, *Modern China*, 2020: 009770042091452.
- [29] 赖慧敏，《清代的皇权与世家》。北京：北京大学出版社，2010 年。
- [30] 梁晨、李中清、张浩、李兰、阮丹青、康文林、杨善华，“无声的革命：北京大学与苏州大学学生社会来源研究，1952—2002”，《中国社会科学》，2012 年第 1 期，第 98—118 页。
- [31] Liu, Y., L. Chen, Y. Yuan, and J. Chen, “A Study of Surnames in China Through Isonymy”, *American Journal of Physical Anthropology*, 2012, 148 (3), 341—350.

- [32] 潘光旦、费孝通, “科举和社会流动”, 《社会科学》, 1947 年第 1 期, 第 1—21 页。
- [33] Piketty, T., “Putting Distribution Back at the Center of Economics: Reflections on Capital in the Twenty-First Century”, *Journal of Economic Perspectives*, 2015, 29 (1), 67-88.
- [34] 钱穆, 《中国历代政治得失》。北京: 生活·读书·新知三联书店, 2001 年。
- [35] Shiue, C. H., “Human Capital and Fertility in Chinese Clans before Modern Growth”, *Journal of Economic Growth*, 2017, 22 (4), 351-396.
- [36] Walder, A. G., and S. Hu, “Revolution, Reform, and Status Inheritance: Urban China, 1949–1996”, *American Journal of Sociology*, 2009, 114 (5), 1395-1427.
- [37] Wu, X., and Z. Zhang, “Changes in Educational Inequality in China, 1990-2005: Evidence from the Population Census Data”, *Research in the Sociology of Education*, 2010, 17, 123-152.
- [38] 吴仁安, 《明清江南著姓望族史》。上海: 上海人民出版社, 2009 年。
- [39] 袁义达、张诚, 《中国姓氏: 群体遗传和人口分布》。上海: 华东师范大学出版社, 2002 年。
- [40] Zhang, Y., and T. Eriksson, “Inequality of Opportunity and Income Inequality in Nine Chinese Provinces, 1989–2006”, *China Economic Review*, 2010, 21 (4), 607-616.
- [41] Zhang, L., “Legacy of Success: Office Purchase and State-Elite Relations in Qing China”, *Harvard Journal of Asiatic Studies*, 2013, 259-297.

## Surname, Place of Origin and Social Mobility in China (1645-2012)

YU HAO\*

(Peking University)

**Abstract** This study adapts surname methods proposed in Clark *et al.* (2014) to the case of China and estimates the rate of intergenerational correlation of status in Late Imperial, Republican and Communist China. Using surnames and places of origin to identify elite groups and examining the changing social status of originally elite groups over time, we find relatively high intergenerational correlation of social status in the Qing dynasty (0.73-0.85), and slightly lower correlation in the Republican and Communist eras (0.65-0.75). But even in the Communist era the correlation is much higher than are conventional estimates (0.3-0.6), where status is partially measured by income, wealth and education at individual or household level.

**Keywords** surname, places of origin, intergenerational correlation of status

**JEL Classification** N01, N31, N35

---

\* Corresponding Author: Yu Hao, the School of Economics, Peking University, Beijing, 100871, China; Tel: 86-18335139672; E-mail: maxhao1003@pku.edu.cn.