Social Mobility in China, 1645-2012: A Surname Study

Yu (Max) Hao and Gregory Clark, University of California, Davis
maxhao@ucdavis.edu, gclark@ucdavis.edu

11/10/2013

The dragon begets dragon, the phoenix begets phoenix, and the son of the rat digs holes in the ground (traditional saying).

This paper estimates the rate of intergenerational social mobility of status in Late Imperial, Republican and Communist China by examining the changing social status of originally elite surnames over time. It finds much lower rates of mobility in all eras than previous studies have suggested, though there is some increase in mobility in the Republican and Communist eras. But even in the Communist era social mobility rates are much lower than are conventionally estimated for China, Scandinavia, the UK or USA. These findings are consistent with the hypotheses of Campbell and Lee (2011) of the importance of kin networks in the intergenerational transmission of status. But we argue it more likely reflects mainly a systematic tendency of conventional mobility studies to overestimate rates of social mobility of status, where status is partially measured by income, wealth and education at individual or household level.

本文以姓氏和籍贯来识别精英，用一个简单的跨期社会地位传递模型来模拟“精英姓氏”走向平庸的历史过程，从而测算不同时期的社会流动性。具体地说，在某一初始阶段，精英阶层的姓氏分布和全体人口的姓氏分布存在差异，一些姓氏更集中于精英阶层。而时间流逝，社会流动将使得这两个分布趋向一致，那些“精英姓氏”在精英阶层的比重将趋近于其在全体人口中所占的比重。这个趋近速度，就刻画了社会地位代际流动性的高低。

我们的发现是：第一，科举时代，社会地位的代际相关性很高（0.8-0.95），高于现代，也高于同时期的英国。第二，及废科举，入民国，土改兴，文革终，代际相关性有所下降，但是幅度有限（0.63-0.74）。晚清的精英姓氏在当代出现在精英层的概率仍是平均人口的 1.2-1.6 倍。第三，现代中国的代际相关性高于最近经济学和社会学得到的收入和教育的代际相关性（0.3-0.4）。总的来说，“龙生龙、凤生凤”更接近我们的观察。

一般而言，我们无法直接观测和量度“真实社会地位”。当真实社会地位被教育、收入、职业、寿命和官位量度的时候，我们往往会高估社会流动速度。在一个既定的时刻，在个人/家庭水平上，上述指标只是“真实社会地位”和“误差项”的和（信号+噪音）。比方说，在收入分配非常平等的社会，收入的高低无法精确量度社会地位的高低（信号/噪音比率较低），所以收入的代际相关性也就无法测度社会地位的代际相关性。姓氏方法可以部分地解决这个问题：误差项在群体水平上和较长时段加总为零。Gregory Clark 的最近一系列研究表明，英国、美国、日本和瑞典现代的社会流动性和中国现代差距不大。 特别是厉行“社会主义”的瑞典，“真实社会地位”的代际相关性远高于收入或教育的代际相关性所显示的情况。一个可能的解释是收入再分配政策可以拉近财富，但是社会地位、社会关系和社会资本难以无法用社会政策再分配，而更多的经家庭和婚姻传承。

本文的精英数据来自作者搜集的近十万个各种科举功名获得者，现当代大学生，官员，富人，教授等各种名录；人口姓氏分布则来自近代历来各县志的忠烈录（太平天国殉难）、烈女姓氏录、抗战和革命英烈名录。为保证不同时期数据的可比性，模型中一个重要的参数是精英阶层所代表的的 top x%， 可根据不同时期精英样本的性质调整，。由于姓氏的地域集中性，现代姓氏人口分布和历史姓氏人口分布可能存在系统性偏差，本文采用了两种方法来控制地域性。

1.1 Introduction

This paper estimates intergenerational social mobility rates in China across three eras: the Late Imperial Era, 1644-1911, the Republican Era, 1912-49 and the Communist Era, 1949-2012. Was the economic stagnation of the late Qing era associated with low intergenerational mobility rates? Did the short lived Republicachieve greater social mobility after the demise of the centuries long Imperial exam system, and the creation of modern Westernized education? Theexam system was abolished in 1905, just before the advent of the Republic. Exam titles brought high status, but taking the traditional exams required huge investment in a form of “human capital” that wasunsuitable to modern growth (Yuchtman 2010). Did the end of the exam system result in a period of rapid social mobility with the rise of a new western-educated elite? And did the turmoil of the Communist era radically disrupt the old social order, andbring an entirely new class of people to positions of

education, prosperity, and power? In particular was there a comeback of the old elites in the post-Mao era, 1978-2012?

Table 1.1: Estimates of b for China, by Period and Group

PERIOD	North Zhejiang 1645-90 Exam Elite	North Zhejiang 1781-1810 Exam Elite	North Zhejiang 1870-1900 Exam Elite	South Jiangsu 1645-90 Exam Elite	South Jiangsu 1781-1810 Exam Elite	South Jiangsu 1870-1900 Exam Elite
1645-1810[a]	0.83	0.89	0.81	0.85	0.89	0.86
1781-1900[a]	0.92	0.87	0.81	0.95	0.92	0.86
1871-1949[b]	0.84	0.84	0.78	0.85	0.86	0.81
1930-1990[c]	0.66	0.70	0.74	-	-	-
1950-2010[d]				0.63	0.66	0.66

Data for elites: [a] *juren* (provincial exam passers), [b] college students from lower Yangzi, [c] notable people from North Zhejiang, [d] Nanjing University incoming students from South Jiangsu

We measure intergenerational mobility rates through b, the implied intergenerational correlation of social status. Table 1.1 summarizes our estimates of b for initial elite surname groups in two provinces of China starting in 1645-90. These estimates suggest a strong persistence of status over generations, even in the Communist era, with the fall of the Empire in 1912 producing very modest gains in social mobility rates. This persistence is so strong that we shall see that the elite surnames of the Imperial Era are still slightly overrepresented among modern elites in Communist China. These persistence rates for status are much higher than has been estimated by most other scholars for China (Gong, Leigh and Meng, 2010). We detail below why we believe the surname estimates correctly show such low rates of social mobility.

1.2 Concepts and methodologies

1.2.1 Surnames and Social Mobility

We measure social mobility in China over these 360 years by observing the speed with which surnames concentrated among elites achieve a distribution across the ranks of society which is the same as the overall distribution of status. Despite the small number of surnames among the Han Chinese, a mere 4,820 in all (Yuan and Qiu, 2010, p1-3), we are able to identify some rarer surnames which had on average high occupational status in Imperial China, and then track the status of these surnames over generations from then to the present.

We assume in this paper that if x_t is a latent variable of true social status of families in generation t then

$$x_{t+1} = bx_t + e_t$$

where x_t and x_{t+1} are assumed to have a mean of 0, and a constant variance σ^2, and x is normally distributed. The error term e_t is the random shock that occurs when parents transit social status to children. It is not related to the social status and is not auto-correlated. However, we typically do not directly observe the complete social status of families, but some partial measure, y_t, where such measures would be earnings, wealth, years of education, or occupational status. For each generation t

$$y_t = x_t + u_t$$

where u_t is a random component linking the underlying status of the family to the particular observed measure of status. This implies that the conventional studies of social mobility, based on estimating the β in the relationship,

$$y_{t+1} = \beta y_t + v_t$$

where v_t is the random shock that occurs when parents transit their earnings, wealth, year of education, or occupational status (again, not correlated to y_t and not auto-correlated), will underestimate the true b linking social mobility across generations. In particular the expected value of β will be

$$E(\beta) = b\frac{\sigma_x^2}{\sigma_x^2 + \sigma_u^2} \quad \text{(See appendix for proof)}$$

Thus conventional estimates of social mobility, based as they are on one generation studies, and on partial measures of overall social status, will systematically tend to over-estimate social mobility rates (Solon, 1989).[1] However, the surname measures that we use here over multiple generations, even when they are based on partial measures of social mobility such as educational or occupational status, will closely approximate to the true underlying b. This is because by aggregating over groups of individuals with the same surname we can make the error component linking observed status *y* and underlying status *x* go to zero.

What we will observe using surnames is that in the initial period or generation some surname group has a higher average social status on such measures, as is portrayed in figure 1.1. However, we typically do not observe the whole distribution of status across a surname group, but just what fraction of that group lies in some upper tail of the distribution, and

[1] Clark and Cummins (2012) summarize the sources of u_t, the random component linking underlying status to the various observed aspects . First there is an element of luck in the status attained by individuals given their underlying aptitudes. People happen to choose a successful field to work in, or firm to work for. They just succeed in being admitted to Oxbridge, as opposed to just failing. Second people make tradeoffs between income, education, occupational prestige, and other aspects of status. They choose to be philosophy professors as opposed to finance executives.So his social status is understated if income is used to measure social status.

how this is changing over time. Our measure for China before 1905, for the exam era, is the fraction of a surname group that attains various degrees on the Imperial Exam system.

To extract implied intergeneration's bs we proceed as follows. Define the *relative representation* of each surname or surname type, z, in an elite group such as doctors as

$$relative\ representation\ of\ z = \frac{Share\ of\ z\ in\ elite\ group}{Share\ of\ z\ in\ general\ population}$$

With social mobility any surname which in an initial period has a relative representation differing from 1 should tend towards 1, and the rate at which it tends to 1 is determined by the rate of social mobility.

The overrepresentation of the surname in this elite could be produced by a range of values for the initial mean status, $\bar{y}_{z0}$, and the initial variance of status, σ_{z0}^2, for this particular surname z. But for any assumption about ($\bar{y}_{z0}$, σ_{z0}^2) there will be an implied path of relative representation of the surname over generations for each possible *b*. This is because

$$\bar{y}_{zt} = \bar{y}_{z0} b^t \rightarrow 0, \text{ as } t \rightarrow \infty$$

Also since $\sigma_{zt}^2 = b^2 \sigma_{zt-1}^2 + (1 - b^2)\sigma^2$,

$$\sigma_{zt}^2 = b^{2t} \sigma_{z0}^2 + (1 - b^{2t})\sigma^2 \quad \rightarrow \sigma^2, \text{ as } t \rightarrow \infty$$

If we can observe relative representation over multiple periods we can determine empirically what the best fitting values of b and σ_{z0}^2 are.

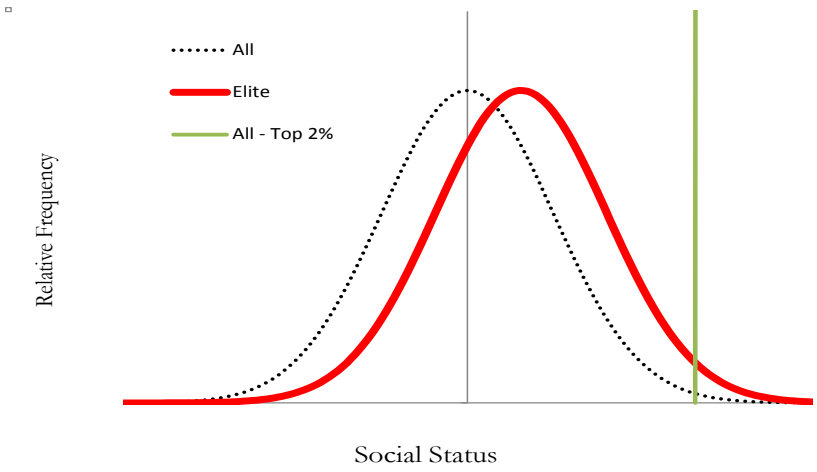
It is not possible to identify$\bar{y}_{z0}$ and σ_{z0}^2 for the elite surname just by observing its overrepresentation among an elite in the first period. However, by observing overrepresentation over multiple generations, we can fix these by choosing them along with *b* to best fit the relative representation of the elite surname z in the social elite. Intuitively, we try to find a combination of b,$\bar{y}_{z0}$ and σ_{z0}^2 so that the derived path of regression is the most close to the actual path of regression that we observed. More specific, we minimize the sum of squared deviations (in logs) of the simulated value of overrepresentation and the actual value of that for each period (Since we are solving a set of non-linear equations, there is no closed form solution to b,$\bar{y}_{z0}$ and σ_{z0}^2).

While we can in general expect that

$$0 < \sigma_{z0}^2 < \sigma^2$$

it turns out to matter little to the estimated size of b in later generations what specific initial variance is assumed (See Clark and Cummins, 2012, for more examples). Below we assume that the initial variance of the elite surname status is the same as the overall variance, since this assumption fits the observed time path of relative representation well in most cases.

Figure 1.1: Initial Position of an Elite



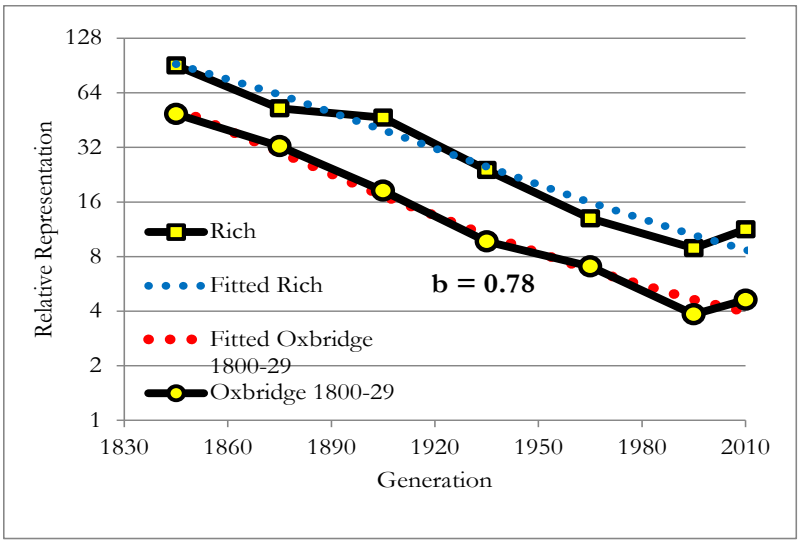
To illustrate how this estimate works in practice consider the data in table 1.2, which is drawn from a well-observed English elite. This shows the relative representation at Oxford and Cambridge Universities of high average wealth rare surnames, based on the wealth at death of those born 1780-1809 who died 1858 and later. In 1800-29 the high surnames of the wealthiest show up at 94 times their share in the population among entrants to Oxford and Cambridge. Relative representation for this elite group declines very little in the years 1830-59, for the children of the first generation. We thus take this second generation as the baseline, and ask what the subsequent decline implies about the rate of social mobility.

Table 1.2: Relative Representation of Rare Surnames at Oxbridge, 1800-2010

Period	All Oxbridge Attendees	N Wealthy Surnames	Relative Representation Wealthy Surnames	Relative Representation Any Rare Surnames 1800-29
1800-29	18,651	169	94	117
1830-59	24,418	210	91	49
1860-89	35,503	184	55	34
1890-1919	22,005	77	43	19
1920-49	44,231	73	25	9.8
1950-79	95,792	67	9.1	6.3
1980-2010	213,303	65	9.2	4.0

Source: Clark and Cummins, 2012.

Figure 1.2: Relative Representation at Oxbridge, 1830-2010



Source: Clark and Cummins, 2012.

The table shows that the rich rare surnames steadily converge in relative representation towards 1. However, the rate of convergence is slow. Even for the cohort entering Oxbridge 1980-2010 the rich rare surnames are still 9 times more frequent relative to the stock of 18 year olds with that name than are common indigenous English names such as *Brown* or *Clark*.

What does the pattern in decline of relative representation shown in table 1.2 imply about the b for education? If we assume a normal distribution of status, and that all those of high status had the same variance as the general population, then we can estimate what the b for educational status 1830-2010 was. Since the high status surnames had a relative representation of 91 among the top 0.7% of the educational hierarchy in 1830-59, this fixes what the mean status of those names had to be, relative to the social mean, assuming the variance of their status was the same as that of the general population.[2] For each possible b their relative representation would decline generation by generation in a predictable manner.

[2] The share of population that attend Oxbridge was 0.7% 1830-2010

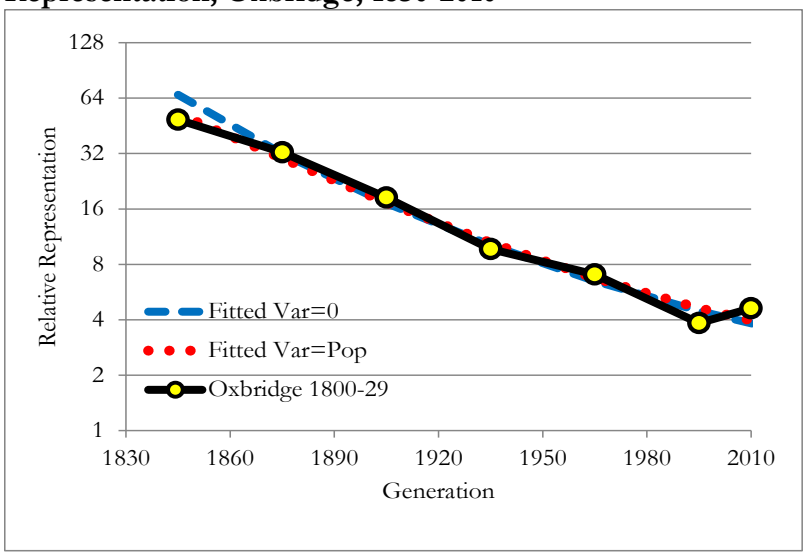
Figure 2 shows the actual pattern, as well as the single b that best fits the data.[3] For the wealthy group that is b = 0.79. Notice also that there is no sign that educational mobility has speeded up in the last few generations. The single b of 0.79 fits the pattern well in all generations.

The rare surnames in this English sample are all associated with wealth. We can form from the Oxbridge records another larger rare surname group which consists just of rare surnames that show up as entrants to Oxbridge 1800-29. Table 1.4 also shows the relative representation of these surnames at Oxbridge to 2010. Here there is a large decline between 1800-29 and 1830-59. But to measure the true implied b it is necessary to start with the generation 1830-59, where the elite surnames were selected based on their occurrence earlier, and so the data is not contaminated by positive errors in the first period. As can be seen this group also remains an elite even to 1980-2010. We can also simulate for the implied b for the regression to the mean of this group 1830-59 to 1980-2010, assuming as before that the initial variance in status was the same as for the population. It is 0.78, as is shown in figure 1.5. As before there is no sign of any speeding up of the process in the most recent generations.

Suppose we instead assume that the status variance of the rare surname group observed at Oxbridge in 1800-29 is instead 0 in 1830-59. How would that change the estimated b to best fit the observed pattern of relative representation? Figure 1.3 shows the fitted path in this case that minimizes the sum of squared deviations. Here the fit is less good. Such an assumption about initial variance implied a much more rapid initial decline in relative

representation, which is not consistent with the data. However, the implied b that best fits the observed pattern changes hardly at all.

Figure 1.3: Assumed Elite Status Variance and the Implied Path of Relative Representation, Oxbridge, 1830-2010



Source: Clark and Cummins, 2012.

So if we use the pattern of relative representation over many generations to estimate the implied b, even though we have to make an assumption about the initial variance in status of the elite, that will have little effect on the estimated value of b. In the results below we thus assumed that the variance in status of elite groups always equals that of the population.

In this Oxbridge illustration the cutoff for entry into this elite remained at roughly 0.7% of each generation from 1800 to 2012. If the population share in the observed elite changes over time, as we will find in China, then we have to adjust the b estimates for such changes. But this adjustment is simple to accomplish within this basic framework for measuring intergenerational mobility.

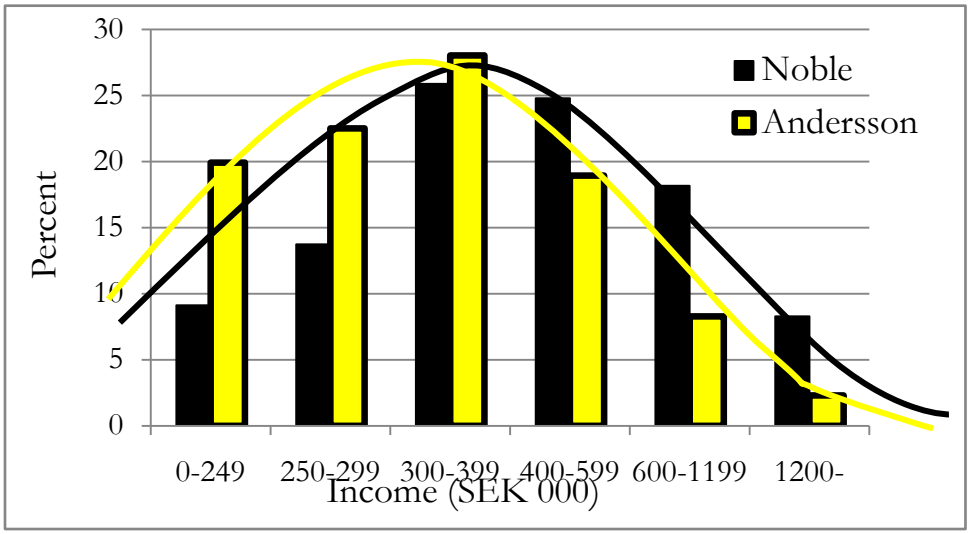
More importantly this measure of intergenerational mobility is premised on the idea that even though we observe only the extremes of the distribution, social mobility rates are all across the distribution of status governed by the simple equation

$$x_{t+1} = bx_t + e_t$$

This has the implication that even social elites will tend to have the same variance of status as the population as a whole, as illustrated in figure 1.1. For Sweden we can illustrate that this assumption does indeed seem reasonable.

In Sweden the surname Andersson, for example, is a relatively lower class name, while there are a set of surnames that belong to aristocrats, whose titles of nobility were mainly conferred before 1700, and which remain concentrated in elites. If we look at taxable income in Sweden in 2008, and take a sample of each type of surname, we find that the distribution of income among the elite looks like it has been shifted to the right compared to the Anderssons, as in figure 1.4.

Figure 1.4: Taxable Income, Noble Surnames v. Anderson, 2008, Sweden



Source: Clark, 2012.

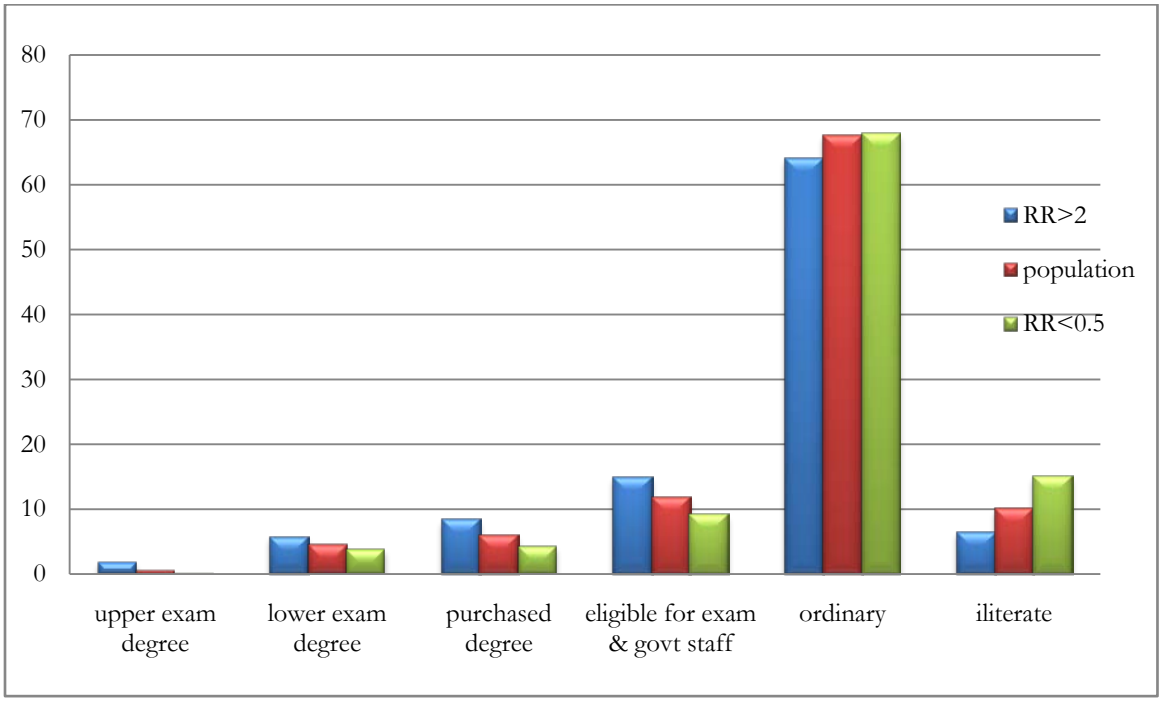
We do not have such detailed data on the complete distribution of outcomes by surname groups in China to test this idea. But by exploring dead men list of Suzhou Countyduring the Taiping Rebellion 1851-1865 (*咸同忠烈姓名录*), we can compare status distribution across surnames. The dataset contains 6400 dead men of SuzhouCounty (苏州, located in the core of lower Yangzi and the biggest commercial city in the late Qing). The name list well documents each individual's exam degree, purchased degree, if qualified for attending exam, if government staff or ordinary people. Besides, among ordinary people, I also distinguish those whose given names are numbers and diminutive (so they are very likely illiterate)[4]. So I can divide people into six stratifications, where 68% can be labeled as ordinary people.

In figure 1.5, we show status distribution of two surnames groups as well as the entire population. RR>2 is the group of surnames whose Relative representation among juren (provincial exam passers) during 1800-60 is larger than 2. And RR<0.5, similarly, is the group of surnames whose Relative representation among *juren* (provincial exam passers) is smaller than 0.5. As is shown, 68% of the dead men can be labeled as ordinary people (*平民*), 4.5% are lower exam holders (*shengyuan,生员和贡生*, the perfecture exam passers), 6% are purchased degree holders or position purchasers (*监生和捐虚衔*), 12% are primitive exam passers *(童生)* who are eligible for prefecture exams) and government staff (*职员胥吏*) and 10% are illiterate. This sample is slightly biased toward upper class but inclusive enough to study difference in status distribution across surnames.

[4]Examples are surnames+阿+x，surname + big/small + animal，surname + number+官/，surnames + number, etc

What is reassuring is that group "RR>2"(7% of Suzhou population) not only show up more frequently among juren, but also more frequently among lower exam passers, purchased degree holders and government staff, and importantly, less frequently among the illiterate. Only 6% of its population areilliterate. On the contrary, "RR<0.5" (20% of the population) show up less frequently among upper class above ordinary people, but also more frequently among the illiterate: more than 15% of them are illiterate. To sum up, the variance of both surname groups have the same variance of status as the population as a whole, and the status distribution of "RR>2" is a leftward shift to that of "RR<0.5".[5]

Figure 1.5: status distribution across surname groups (SUZHOU county, 1851-1865) %



[5]We also find similar pattern in Wuxi (无锡)，Wujin(武进)，YuYao (余姚), Haining(海宁), Xiaoshan(萧山) , where such dead men lists are available.

1.2.2 Surnames and Social Mobility in China

Initially it seemed as though surnames would not be useful in China as a tracer of social mobility rates, because of the small numbers of surnames, and their long antiquity. As noted, there are only around 4,000 surnames in use by the Han Chinese. The 100 most common of these accounts for nearly 85% of the population (see appendix). In contrast in England and Wales in 2002 there were 270,000 surnames shared by 5 or more people (Clark and Cummins, 2012).

The great antiquity of Chinese surnames also meant that there were many generations over which elite surnames could gradually revert to average status. Though surnames first appeared among the nobility and high civil servants, the introduction of civil service exams in 600 AD, by which commoners could obtain official positions by taking exams on the Confucian classics, was associated with a decline in average status for the original noble surnames. By the Qing era (1644-1911), all the aristocratic surnamesof great antiquity -王, 谢, 陆, 顾, 孙, 崔, 卢, 李, 郑, 裴, 韦, 杜, 朱 – appear amongthe national exam passers, *jinshi*,at no greater a rate than the average Chinese surname.The common surnames in China are all, once regional distribution is controlled for, of average social status.

However, we can find surnames that even in recent generations were of high average status by using surname – place of origin combinations. Given how numerous most surnames are in China, families and lineages would distinguish themselves in China in the Imperial and Republican eras by noting their county of origin – the place where the family originated. Such Surname-Place of Origin combinations (S-POs) includes the Fan Family of Ningbo (宁波范氏), the Zha family of Haining(海宁查氏), and the Weng family of

Changshu (常熟翁氏).These surnames themselves - Fan, Zha, Weng - are common and average in social status at the national level. But such a name combined with a place of origin can have a relative representation of much greater than 1 among exam passers in the Qing era, and thus allow us to track rates of social mobility from 1645 on (see appendix for the full list of lower Yangzi elite S-POs).

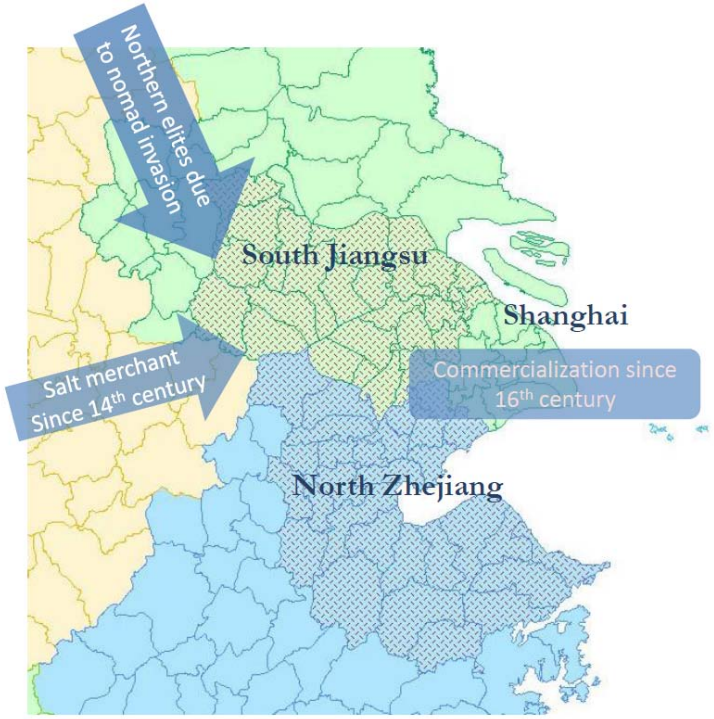
To measure social mobility in this way we concentrate on two regions in the lower Yangzi, South Jiangsu and North Zhejiang, and on surname lineages identified by the combination of a surname and a lineage place of origin. We have very rich source of exam passers' names and modern college students' names for 370 years. There are also voluminous studies of kinships in this area based on genealogies and gazetteers. We find S-POs of initially high status by looking at the relative representation of these surnames among passers of Provincial Exams (juren) in the Qing era.

It wascommon practice that exam passers, on their published exam paper, reported their remote ancestors, their mother's remote ancestors, and any exam passers within five generations in the entire kinship. Editors of local chronicles of the origin counties also reportedeven exam passers whose great-grand-fathers had emigrated from the county 100 years before. Hence it is possibleto trace such S-POs of high status,even inclusive of migration outside their county of origin, at least within the exam era.

Such elite S-POs are particularly concentrated in the lower Yangzi delta, the most developed region of pre-industrial China. During the Mongol invasions of 1205-79, many elite kinships from Northern China migrated south of the Yangzi River, settling in the cities and countryside around Hangzhou, the capitalof the southern Song dynasty (Wu, 2010). Another source of these regional elite kinships wereSalt Monopoly merchant families

whomigrated from Anhui province fromthe 15th century on (Ho, 1999).Moreover, the 15th to 19th centuries witnessed the emergence of various *nouveau riche*, who profited from rural industrialization and commercialization, and achieved high social status by producing exam passers (Li, 2000). The geographic coverage of my study is as depicted in figure 1.5.

Figure 1.5: lower Yangzi delta and sources of elite kinships



To trace these groups into the Republican era and further, we need to know the lineage place of origin of people listed among various elites. We can do this in the Republican era because university rosters and alumni records still reported places of origin. Under Communism, however, such practices became less common. Fortunately, two sources are available with places of origin reported at the county level even in the Communist era: acatalogue of notable peoplefor modern Zhejiang, andthe records of incoming students of Nanjing University (ranked 3rd-10th among universities in China). With these sources we can get sufficiently large samples of elite surname frequencies to get estimates of b across the three regimes of Chinese History: Empire, Republic, and Communist.

1.3 Data and Results: Lower Yangzi S-POs

1.3.1 Lower Yangzi local surname elites

For the lower Yangzi, we track the status over time of elite S-POs, such as the Yun family from Wujin, for two subareas: South Jiangsu and North Zhejiang. We establish the share of surnames in each district by county of origin from the records of 21,644 soldiers from the lower Yangzi dying in the civil wars of 1927-36, and 1945-49, and the external wars of 1937-45 and 1950-3 (12,737 from South Jiangsu, 8,907 from North Zhejiang). For the Yun(恽) from Wujin(武进) in South Jiangsu, for example, their population share in South Jiangsu is 9 out of 12,737 dead soldiers, or 0.071%. It might appear more natural to use modern population censuses, or the like, to estimate these population frequencies. But such S-PO censuses exist for only a few counties[6]. We thus assume that the share of surnames found among soldiers dying 1933-53 is representative of the population, and that the population share of surnames is constant over time. Below we test the validity of these assumptions.

We can similarly calculate the share of name-place combinations like Yun-Wujin among the juren of South Jiangsu in 1645-90. The juren lists are collected from local Chronicles. These local documents record not only places of birth of exam passers but also their places of origin (ancestors) based on relevant genealogies. In this case it is 5 out of 1,045 juren, or 0.48%. Dividing this by the population share shows that the relative representation of this S-PO was 6.8 among juren in 1645-90. All S-POs from South Jiangsu with a relative

[6]Some counties report surname population by residency rather than places of origin. For example, the population of a surname may exclude emigrants who hold the surname and include immigrants who hold the same surname from another origin. For small native surnames, surname population by modern residency will understate historical surname population share in that county. For Zhang, Wang, Li, the bias is in the other direction.

representation of 5or more were included in the 1645-90 elite surname sample, which gave 48 surname-county pairs[7]. Setting it to be Table 1.3 shows that in South Jiangsu these 48 S-POs constituted 2.04% of the sample of dead soldiers, but 21.7% of juren from 1645-90.

Table 1.3: Summary Statistics for six Elite Surname Groups

	North Zhejiang 1645-90 elites	**North Zhejiang 1781-1810 elites**	**North Zhejiang 1870-1900 elites**	**South Jiangsu 1645-90 elites**	**South Jiangsu 1781-1810 elites**	**South Jiangsu 1870-1900 elites**
Population share - 1933-53 (%)	1.90	1.77	1.86	2.04	1.98	1.78
Share of 1645-90 juren (%)	**15.92**	7.69	3.89	**21.70**	6.91	6.13
Share of 1781-1810 juren (%)	5.68	**12.69**	5.99	8.48	**16.08**	7.68
Share of 1871-1900 juren (%)	4.78	6.33	**12.92**	7.60	9.9	**19.58**
Share of 1973-89 elites (%)	2.21	2.46	3.02	2.11	2.87	2.58
Summary statistics of S-POs of each group						
Number	62	58	62	48	40	41
Max	23,000	16,000	25,000	24,000	30,000	23,000
Min	800	500	400	600	800	300

Note: for each column, the number in bold give the "peak" of these elites.

A similar procedure, with a cutoff of at least 5 times relative representation for an S-PO to be included was used to construct the other 5 samples of surnames from each subarea for elite surnames 1645-90, 1781-1810, and 1870-1900. Table 1.3 summarizes the share of the

[7] Setting it to be 5 will mostly rule out those very big surnames such as 王，张，李，刘，杨，顾，沈，朱，etc. These surnames typically have multiple kinships in a given county and very are unlikely to be highly over-representative at county level.

elites of four periods, 1645-90, 1781-1810, 1871-1900, and 1973-89 that these surnames constituted. See the appendix for the full list of these S-POs.

1.3.2 Regression to the mean for S-PO elites

Figures 1.6 and 1.7 show, for South Jiangsu and North Zhejiang, the share of common surnames, and of the two earlier S-PO elite surname groups, among various elite populations 1645-2012.

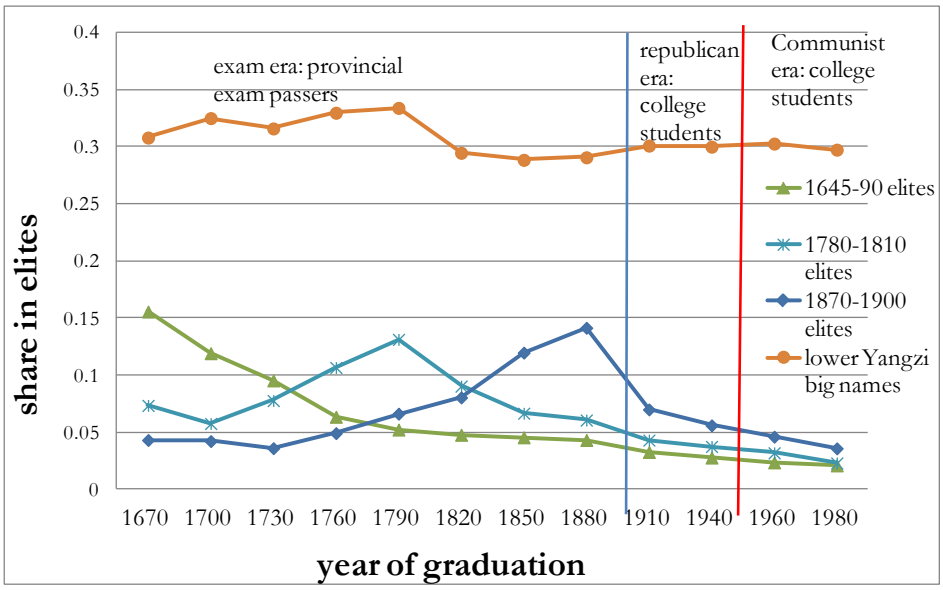
Up to 1905 the elite are the juren, the provincial degree holders, under the Imperial Exam. For the Republican era, we use as a measure of the elite a sample of 11,000 university students who graduated 1900-50. After 1949 we have different elite samples for the two subareas.

For north Zhejiang there is the Zhejiang Provincial Archive, which lists notable Zhejiang people born 1930-70, covering a wider spectrum of occupations, such as artists, professors, and high-rank officials. This population includes people who had migrated out of the region.

For South Jiangsu the Nanjing university archive allows searching for students by names, places of origin and year of graduation after 1952. Nanjing University, located in South Jiangsu, ranks 3rd-10thamong Chinese universities, and enrolls disproportionally from Jiangsu (40% of incoming students). To enroll in Nanjing University, students have to achieve scores in the highest 3-5% of those taking the College Entrance Exam. Those taking the entrance exam constitute only 20-50% of each population cohort. Thus Nanjing students represent an upper 0.6-2.5% of the South Jiangsu distribution of scholastic ability. For North Zhejiang

after 1949, the elite sample (*notable people*) only covers people graduating until 1990, whereas for South Jiangsu the dataset covers college students from Nanjing University until 2010.

Figure 1.6: Share of north Zhejiang elites of 1645-90 and 1780-1810 in the elites of other generations



Note: Elites S-POs are picked up by calculating over-representation ratio of each S-PO from 1645-1690, which are shares of S-POs among *juren* over shars of S-POs among soilders.

1. Exam elite surnames are collected from historical county-level chronicles, which keep the records of their names, places of origin and years of passing.
2. College students' surnames are collected from various university alumni books, which include places of origin and years of admission of these students.
3. Surnames of soldiers are well recorded by modern county chronicles in the chapters of "honoredsoldiers"

In figures 1.6 and 1.7 the local dominant surnames account for around 30% of elites for each period from 1645-2012, suggesting that relative name frequencies were not changing over time even over this long era. The various elite surname groups are 1.5-2% of the population, but their shares in elites differ greatly from their population shares.Figures 1.8 and 1.9 show, based on the shares of names in soldier deaths from these two areas 1933-53, the Relative Representation over the generations of the 1645-90elites, the 1781-1810 elites, and the reference lower Yangzi big surnames. See appendix for the sources of surnames.

Figures 1.7: Share of south Jiangsu elites of 1645-90 and 1780-1810 in the elites of other generations

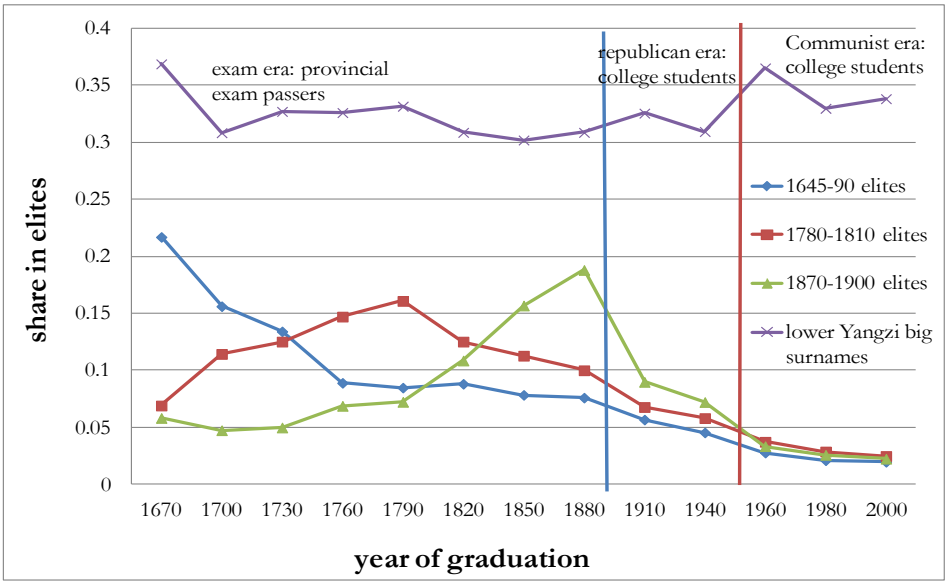


Figure 1.8: Relative Representation of north Zhejiang Surname Elites

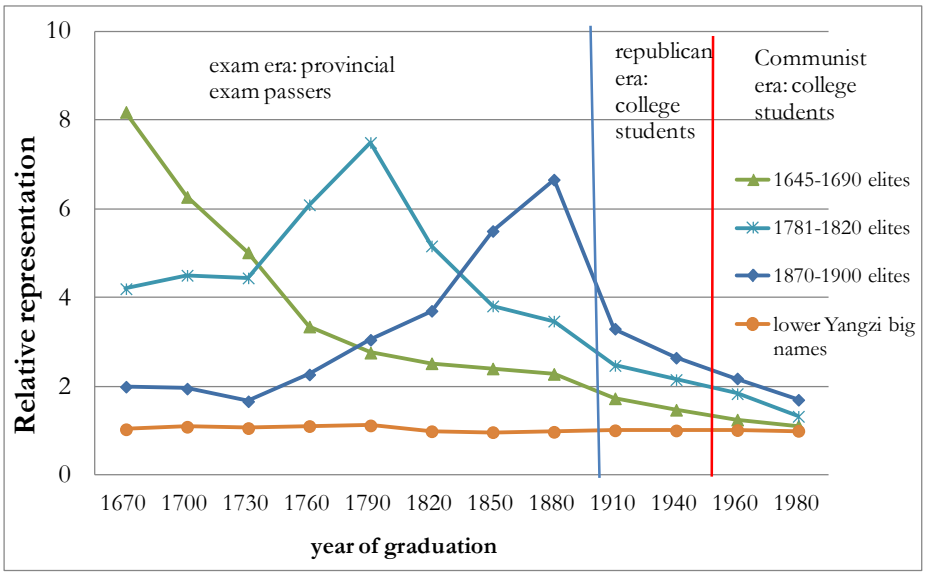
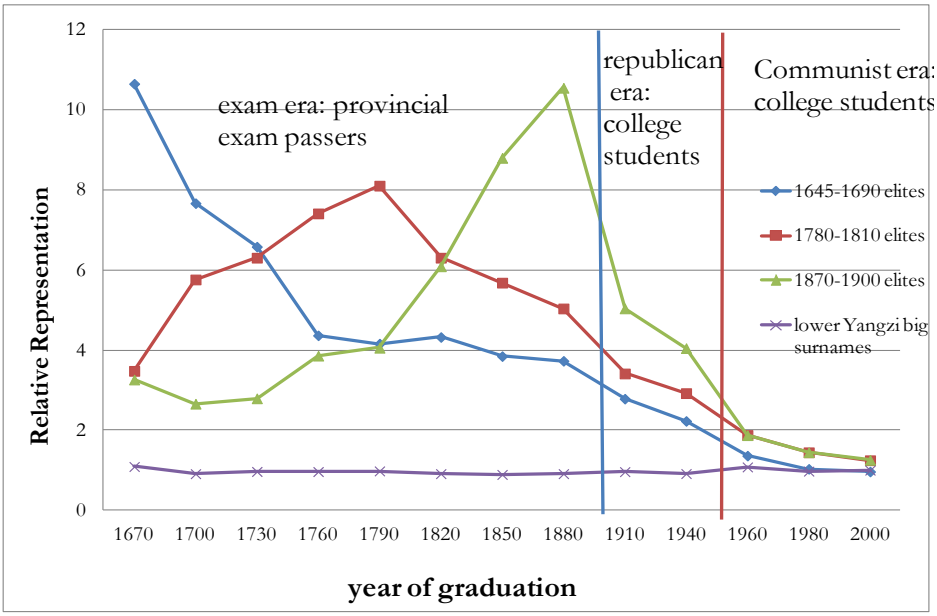


Figure 1.9: Relative Representation of south Jiangsu Surname Elites



1.3.3 Estimating b by period

We estimated the b's by period by the simulation described above for these six elite S-PO groups. To do that we need to estimate for each period what share of the population our observed elite groups constituted. Table 1.4 contains the information needed for this calculation. We take the elite share of the population as that listed in the last column of table 4.

In the exam era, juren were a very small share of the male population. Based on the numbers observed in each district, and the populations of the districts, typically only 0.03% of males would attain this status. To get the shares of the population before 1900 that belong to the provincial exam elite (juren), we assume an average year of passing at 30-year

old, and that half the population was female.[8] The juren were thus a very exclusive elite in Imperial society.

The university students of the Republican era that we employ as the measure of the elite of this era were a much larger slice of the society. In 1949 an estimated 0.34% of males in recent cohorts in the lower Yangzi attained a college degree, so we assume this elite was the top 0.2-0.5% of the society. College students before 1949 were largely from families of wealth and high status. Universities had their own admission exams which favored students of decent backgrounds (for example, many universities had their exams in English). After graduation, they were placed as government officials, managers, schoolmasters, and engineers. Since 1949 there has been a progressive expansion of university education. Students from families of factory workers and farmers were given more opportunities. By 1950-66, 2% of people aged 18-21 entered college. For 1966-76 the college entrance exam, the CEE, was suspended. But in 1977-90 college graduates represented 2-4% of the cohort, and by 1991-2000 college students represent 5% of the cohort.

Our observed elites before 1900 are the complete sample of the high gentry class, whereas after 1900 our observed elites are only part of those with the highest educational attainment. As shown in the last column of table 1.4, we estimate the 1901-49 elite share,to be the share of each cohort entering college. For 1950-2000 we set the cutoff to be the college student share times 0.2, because Nanjing University typically enrolled only the top 20% of incoming college students from South Jiangsu. Knowing the relative representation of each surname group in the elites of each generation 1645-2012, and how exclusive these

[8] Because of female infanticide the female population would be a smaller share than a half, but we do not try and estimate the strength of this effect.

elites were, we can proceed to estimate the best fitting bs for the Imperial, Republican and Communist eras.

Table 1.4: The lower Yangzi elite, 1645-2000

Period (yearof graduation)	Size of Elite	Population (millions)	Elite Share (males) %	Elite share used for simulation %
	North Zhejiang			
1645-1690	1,156	5	0.048	0.05
1690-1720	1,040	8	0.026	0.03
1721-50	1,247	9.5	0.026	0.03
1751-80	1,462	11	0.027	0.03
1781-1810	1,584	12.5	0.025	0.03
1811-40	1,714	14	0.024	0.03
1841-70	1,754	15.3	0.023	0.03
1871-1900	1,548	5.9	0.052	0.06
1901-33	2,313	9.4[a]	0.049	0.2
1934-49	3,017	8.4[b]	0.072	0.5
1950-66	2,526	10.3[c]	0.049	0.5
1977-90	1,948	12.3[d]	0.032	0.7
	South Jiangsu			
1645-1690	1,025	4.8	0.044	0.05
1690-1720	917	7.5	0.024	0.03
1721-50	883	8.5	0.021	0.03
1751-80	1,043	9.8	0.021	0.03
1781-1810	1,102	11.2	0.020	0.03
1811-40	1,283	12.4	0.021	0.03
1841-70	1,427	13.8	0.021	0.03
1871-1900	1,550	4.8	0.065	0.06
1901-33	2,449	7.8	0.063	0.2
1934-49	2,619	7.6	0.069	0.5
1950-66	2,843	10.5	0.054	0.5
1977-90	4,437	14.7	0.060	0.7
1991-2000	4,034	17.6	0.052	1

Notes: [a]1928, [b]1947, [c]1957, [d]1980, from population census.
Source: Pre-modern data is from Ho, 1962, and Cao, 1998.

Figure 1.10: Estimating b by Period for South Jiangsu1645-90Surname Elite

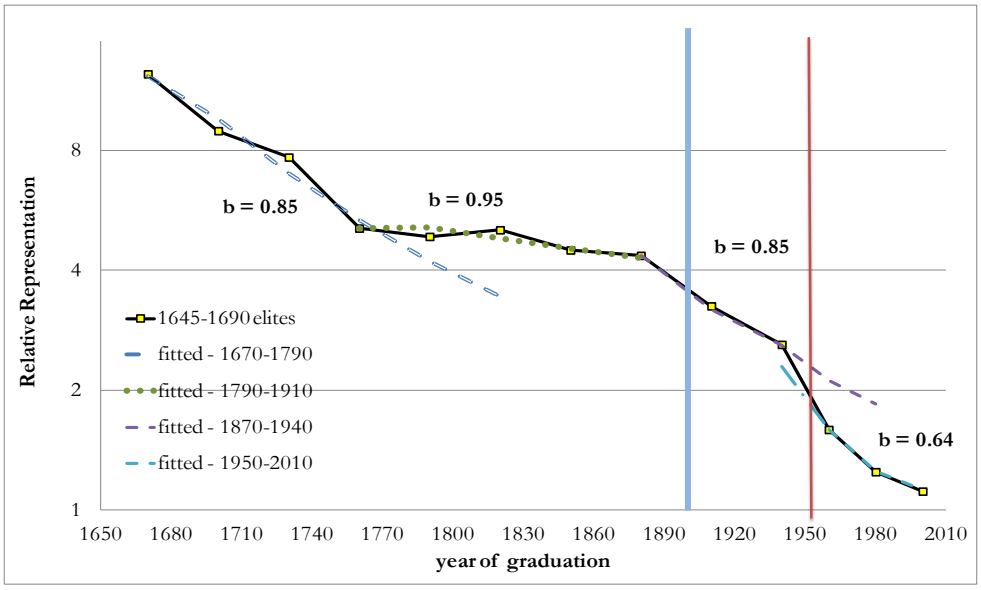
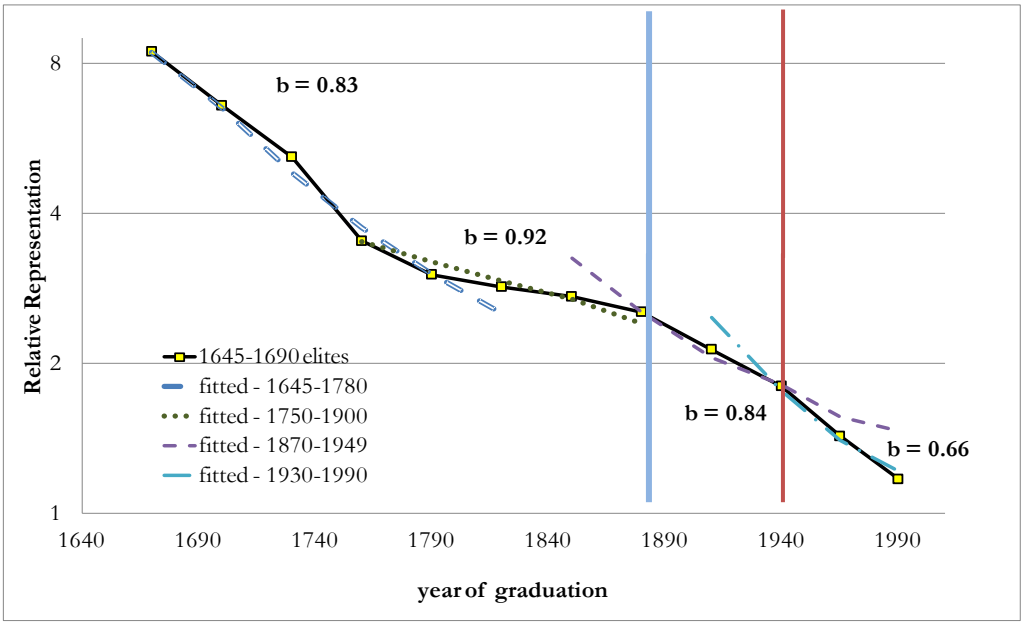


Figure 1.11: Estimating b for North Zhejiang 1645-90 elite, b varies



However, it is evident from the data that even within the Imperial era, there was a period of faster downwards mobility in the early Qing, and of slower mobility in the late Qing.Figure 1.10, for example, shows the relative representation of the 1645-90 surname elite by generation in South Jiangsu in logarithms. A constant b roughly corresponds to the downward slope of this line plotting relative representation being constant. As can be see the slope changes over the generations. The best fitting b for each sub-period is shown on the figure (instead of fitting b for all 12 generations, we fit b separately, for 1645-1780, 1750-1900, 1870-1950, and 1950-2010)

Interestingly, North Zhejiang shows a very similar pattern over time, and produces similar estimates for the various sub-period bs. The message is that social mobility rates as revealed by the surnames were extremely slow in the Imperial Era, and continued to be slow even in very different regime of the Republic. There was an increase in mobility rates under Communism, but even in this era the implied b is 0.64-0.66, which implies a high degree of persistence, explaining why the relative representation of these surnames —identified by their overrepresentation among the elite in 1645-90, continues to be greater than one even among the most recent elites. Sociological studies based on household intergenerational surveys typically give a b of 0.3-0.6 for income, education and party membership (Gong, Lei, and Meng, 2010; Walder and Hu, 2009).

1.3.4 Upward mobility

The pattern of mobility for the middle Qing 1781-1810 surname elites is very similar for the years after 1810 to that of their 1645-90 counterparts as shown in figures 1.12 and 13.

This is shown in the bs estimated in table 1.1. But here we can also estimate the rate of social mobility by looking at the upward path of relative representation as a surname becomes elite. The model posited above that underlying status is linked across generations by the formula

$$x_{t+1} = bx_t + e_t$$

also has implications about what the path of relative representation will be for surnames observed to be elite in any specific generation in the periods before that observation. If we observe any group of families that now deviate from average social status, then they will have deviated on average by a lesser amount, determined by *b,* in the previous generation. For these groups, future and past trajectories are precisely symmetrical.

The intuition is as follows. A group of families now of high social status is identified from an elite sample of a given period. These families are not escapable of the law of regression to mean. Toarrive at this status, they must have beenhit over many generations by a series of positive random shocks, e_t.Since random shock is normally distributed, the greater this persistence rate, these families on average need more of positive shocks, hence longer periods, and flatter implied path to the elite.

Table 1.1, and figures 1.12 and 1.13 show the implied bs also for the rise of these mid Qing elites. These b values are remarkably similar to those for the downwards movement of the 1645-90 elites over the same period, as can be seen in table 1.1. The implication is that along with slow downward mobility in the Qing we see the upward movement of much less elite lineages over many generations to positions of high average status.

Figure 1.12: South Jiangsu 1781-1810 Surname Elite

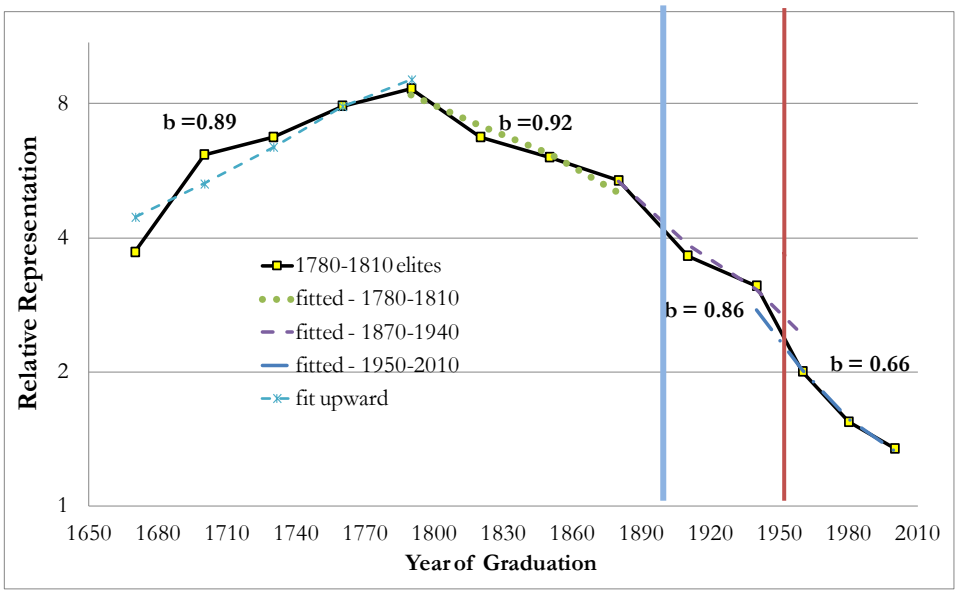
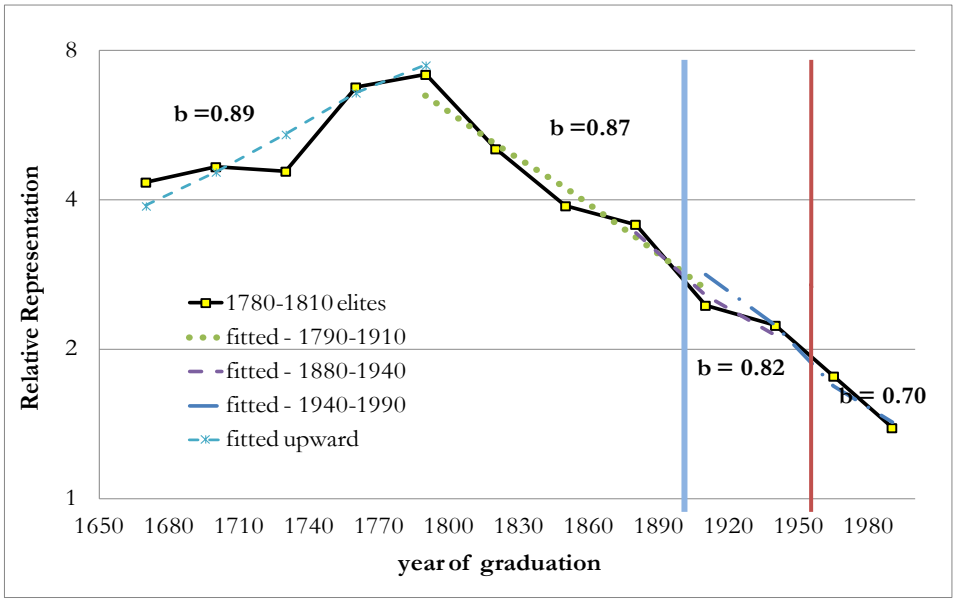


Figure 1.13: North Zhejiang 1781-1810 Surname Elite



Finally figures 1.14 and 1.15 show the rise and decline of the 1871-1900 exam era surname elites, measured by the relative representation by generation. The declines post 1900 again is similar to that for the earlier samples of elite S-POs, as table 1.1 shows. And again the rates of rise of these surnames look similar to the rates of decline of the earlier

elites 1645-1900. Except, however, both sets of surnames show no decline in relative representation as we move backwards from 1710-40 to 1680-1710 and then 1645-80.

Figure 1.14: South Jiangsu, 1870-1900 Surname Elite

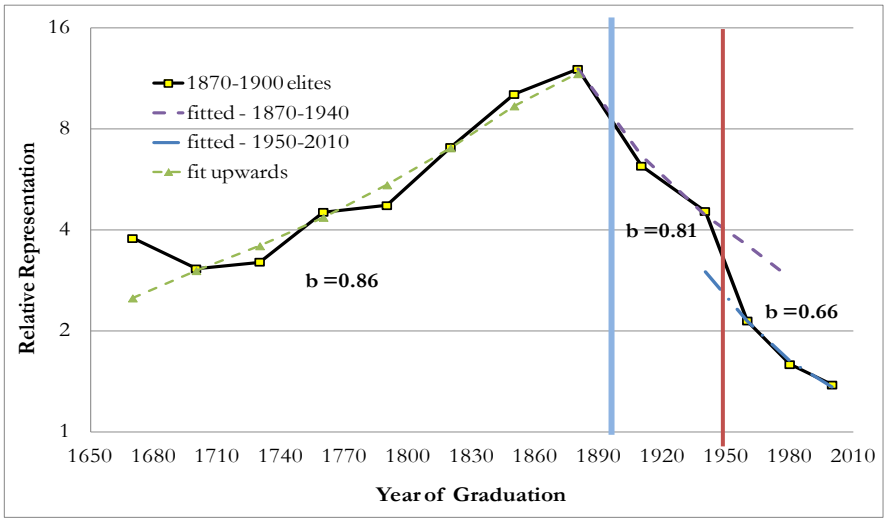
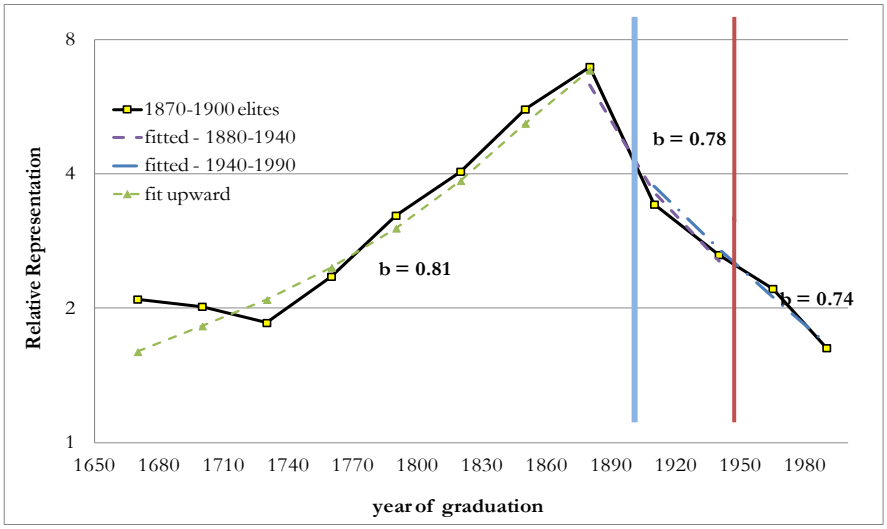


Figure 1.15: North Zhejiang, 1870-1900 Surname Elite



Averaging across the various periods in table 1.1 gives the following period estimates for b:

Early Qing (1645-1810)	0.86
Late Qing (1781-1900)	0.88
Republic (1871-1949)	0.83
Communism (1930-2010)	0.68

Social mobility rates are thus extremely slow in the Imperial and Republican eras, and still very modest in the Communist era, though faster than before.

This implies that even with the disruptions of the end of the of the Imperial Exam system in 1905, of the Imperial Era itself in 1912, of the Japanese war and Civil War of the 1930s and 1940s, the high status surname groups of the Qing saw only a modest decline in their social position as measured by graduates of the new modern universities.

The Communist era brought higher implied social mobility rates. But that is still a rate of social mobility that is surprisingly slow, slower than has traditionally been estimated for most modern countries.[9] Despite successive waves of property confiscations and even of executions, the old elite survived to a degree even to the present day. Under the Agrarian Reform Law of 1950, for example, rural landowners saw their property seized and redistributed, and the landlords themselves were executed or sent for re-education. One potential interpretation is that property confiscations and executions under Communism were targeted to the "exploiting class", in essence the wealthy, rather than to kinships of high social status through occupation and education (Campbell and Lee, 2010).

The persistence of the old elite under Communism has a number of possible sources. They may have benefited from being more urban in their residence, and hence having better access to education in the 1950s and again in 1977 and later when the college entrance exam was fully restored. The advantages in income and educationof cities compared tothe countryside was institutionalized under Communism by settinglow agricultural prices, and by the system of household registration (*hukou* system) (Bian, 2001, Wu and Trieman, 2007).

[9]In the future we will test if our finding is robust to using other measures of status, say, and party membership. And we will test the elite S-POs from 1930-1949 to better test the effect of communism on social mobility.

Note also that North Zhejiang shows less increase in social mobility than South Jiangsu after the arrival of Communism. One potential reason is that for the North Zhejiang elite sample after 1949, we used "notable people of Zhejiang."This includes those who migrated to Taiwan, Hong Kong, and elsewhere overseas, and their descendants. The South Jiangsu elite sample, in contrast, uses incoming students to Nanjing University. Nanjing after 1949 only enrolled those residing in mainland China. Nevertheless, the difference is too modest to alter our results.

1.3.5 Robustness checks

These results are derived using a number of assumptions.These assumptions include that the elite S-PO share of the population all the way from 1645 to 2012 is given by their share among the dead soldiers of the lower Yangzi 1933-53. They are also based on sampling of surnames. How sensitive to the results are these assumptions, and how important could be sampling errors?Could the soldier lists for example, misrepresent the population shares of these S-POs? Or could their population share have been much lower in the past, because they had faster population growth as elite than the general population?

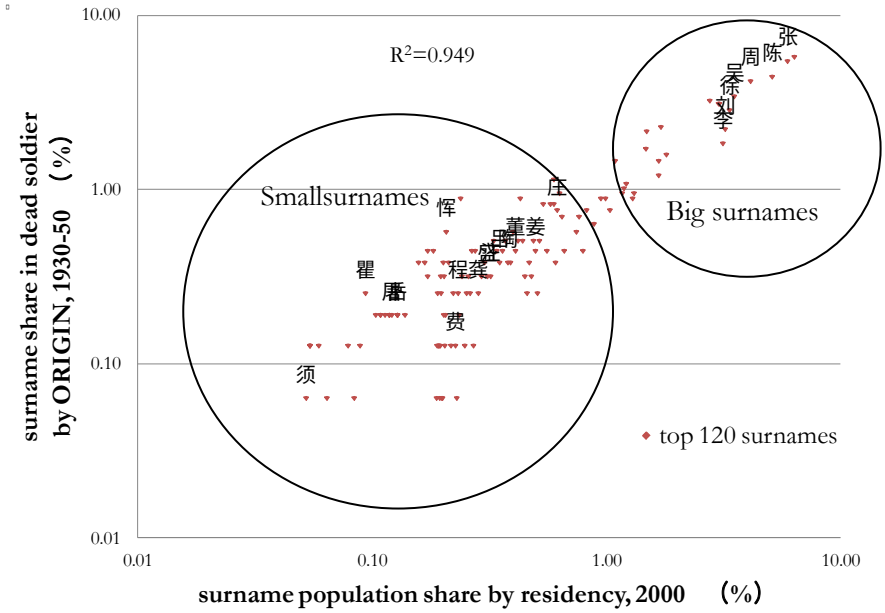
1.3.5 (a)Is population share of S-POs stable 1950-2000?

The first check we can employ is to ask whether the frequency of surname-place of origin combinations among the dead soldiers is indicative of the current shares of these S-POs. As noted above, modern censuses do not give the place of origin of surnames. But for some counties we do get surname frequencies in recent years, so that we do get surname—place of registered residency (户籍人口). We can compare this to the earlier surname—place of origin frequencies.

Thus for Wujin county (containing the city of Changzhou) we have the name distribution of the modern population of 2 m in 2000. We can compare the surname shares of residents in Wujin in 2000 (for the 120 most common surnames) with the share of surnames among the 1543 dead soldiers 1933-53 who have Wujin listed as the place of origin. If out and in migration rates do not differ systematically for the various surnames these two measures should be correlated. Figure 1.16 shows this comparison, where the share is measured in log form to give more weight to the smaller share surnames. There is a strong correlation between the 1933-53 measures and the modern residency surname shares. Indeed if we regress for Wujin,

$$LN(Share_{SOLDIERS}) = \alpha + \beta LN(Share_{2000})$$

Figure 1.16: Measuring surname population share in places of origin Wujin county (武进县，包括今常州市区和武进县)



Sources:"*Surnames in Changzhou*" (常州姓氏，2000); " *Gazetteers of Wujin county: name list of revolutionarysoldiers*" (武进县志：革命烈士姓名录，1992)

the estimated value of β is 0.92, and the R^2 is 0.95. If we further include in this regression an indicator for the surnames of place of origin Wujin that are in the elite surname samples from the Qing era, then that indicator is insignificantly different from 0. These 16 earlier elite surnames (labeled in the left circle in figure 1.16) have the expected frequency among the soldiers. Thus the relative shares of these surnames did not seem to change 1953-2000, and the shares in 1933-53 are a good indication of their later population shares.

1.3.5 (b)Is population share of S-POs stable 1645-1949?

Another issue is whether the population growth rate of elite surname groups was more rapid than for surnames as a whole 1645-1949? If this was the case then the actual social mobility rates would be higher than estimated since the original elite S-POs would have an even higher relative representation in the Imperial era.

The evidence on the relative fertility of Qing elites is conflicting, however. A study of 7 lineage groups in Tongcheng(桐城)county in the Middle Yangzi, for example, finds that fertility among elite families in the eighteenth and nineteenth centuries was not much higher than for more average ones (Shiue, 2012). The males of the upper classes were more likely to marry, and to have multiple wives, but they had about the same number of children as lower class males. Campbell and Lee (2010), in contrast, find that for the same period the upper class in the Northeast of China (the frontier of China) was more successful in reproducing themselves. Thus it is uncertain if the Chinese upper classes had significantly higher fertility. The lower Yangzi delta resembles Tongcheng more than the Northeast in terms of culture, population density, literacy rate and incomes. So it seems reasonable to assume that these elite surnames were likely to have had a stable share of the population. Besides, we are tracking S-POs that contain a higher fraction of members in the elite class, but not the elite

class itself. The majority of the population even with these surnames will have been the illiterate lower class. So even if the members of the elite within this surname group did have a higher natural population growth rate, there should be little systematic difference between the elite S-PO's population growth rate and that of the general population.

To obtain an upper limit of how much faster an elite S-PO grows than the general population, we calculate the number of sons given birth to by an elite S-POs as well as that of the general population based on Shuie (2012). Her sample contains 8,000 male head of household born 1300-1850, and she reports average number of wives and sons by status, as in column 4 of table 1.5. By her criteria, exam degree holders qualify for "gentry" class, who account for 3.32% of all males. "Near gentry" class includes purchased title holders, educated scholars without degrees, and others with indicator of wealth (make large contributions to clan assets or having 2 or more wives or concubines at the same time). "Moderate wealth" includes those village head, or those who received some honorary or posthumous title because of their contributions to the lineage, and Men with no other evidence of status other than having multiple sequential marriages. There was non-linear relationship between class and fertility: near gentry and moderate wealth gave more birth to sons (2.17) than "no indication of wealth or title" (1.95); but the gentry class (2.04) were almost as fertile as the lowest class. On average, 2.02 sons were given birth to by each male in her sample.

Table 1.5: Differential fertility rates across classes

Number of wives and total number of sons born to each male head of household, by status of the male (dropping non-household-heads and dropping ones with missing year of death)

1300-1850 Middle Yangzi (Shuie, 2012)[a]	(1) observation	(2) share %	(4) number of sons	(5) benchmark population	(6) Haining-Chen 1600-1850[b]	(7) an very elite kinship

no indication of wealth or title	5793	71.24	*1.95*	60.00	31.50	20.00
moderate wealth	1421	17.47	*2.17*	15.00	20.00	30.00
near gentry	648	7.97	*2.33*	4.00	22.00	25.00
gentry	270	3.32	*2.04*	1.00	10.00	15.00
never married	**omitted**	**NA**	**0.00**	20.00	16.50	10.00
average			2.02	1.61	1.76	1.93
relative to benchmark			1.26	1.00	1.10	1.20

Source: a, calculated from Shuie (2012), table 2
b, calculated from Lai (2010), page 87-91, p94-96

However, this sample is selective and hence not an appropriate benchmark of the general population. First, her sample dropped people who never became a household head (not married the whole life for whatever reasons) and those who are missing year of death.[10] Second, genealogies have incomplete or no records of those from lower class and women. Last, these genealogies were relatively more "elite" kinships of Tongcheng that had higher fraction of people being degree holders. In some sense, genealogy itself was an indicator of high status of a group of people at certain point of time. The people whose descendants died out, dispersed, or were too poor to construct a genealogy were largely truncated from the survived records of history. This is reflected by the unusually high fraction of gentry (3.32%) and near-gentry class (7.97%), as in column 2 of table 1.5.

To fix this problem we created a benchmark population and applied the fertility rates across classes from column 4. This is shown as column 5. Based on our knowledge of civil degree holders, "gentry" class should account for roughly 1% of male population and 4% for

[10]The average wives were 1.15 which was too high given our knowledge that in pre-modern China the sex ratio was 110:100 or even higher.

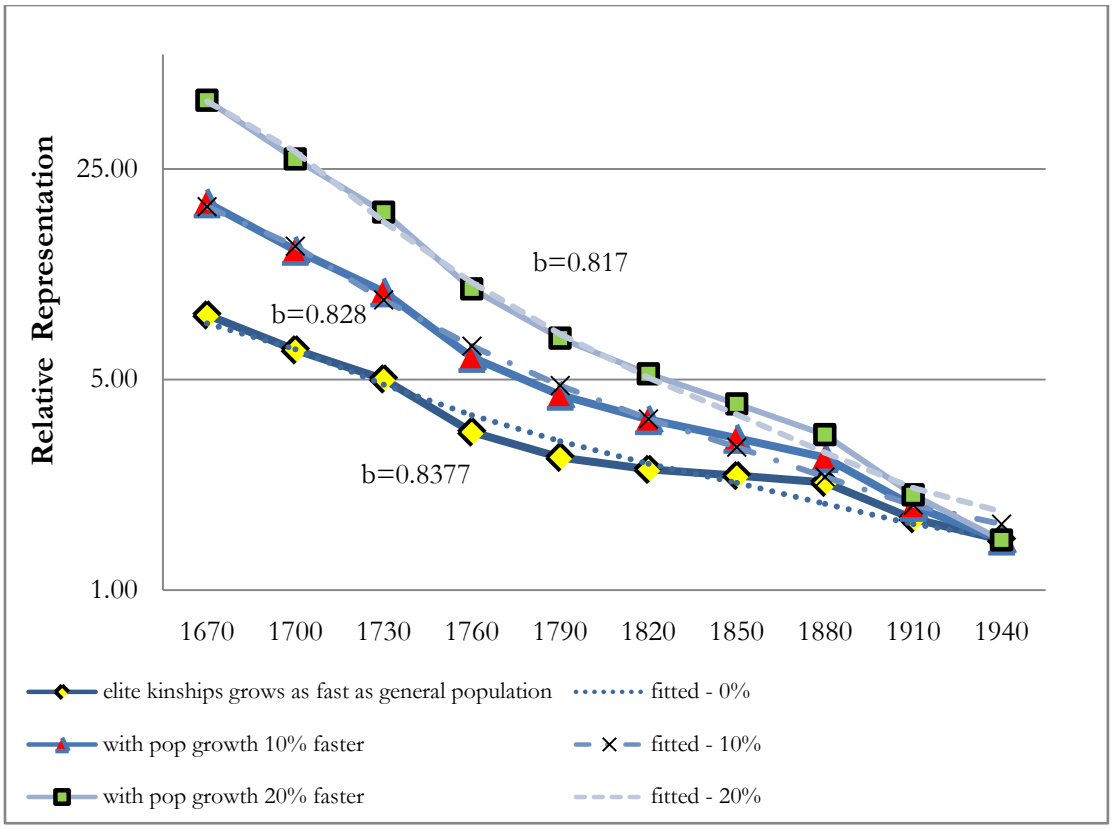
"near gentry". There should be on average 20% of males never got married because the male/female ratio in pre-modern China was 113-119 (Jiang, 1994, p300), and the males of upper class have multiple wives. "Moderate wealth" and others account for 15% and 60% respectively. The average sons by a male is 1.61, 25% lower than the average in column 4.

The next step is to estimate that number by an elite kinship (column 6). According to Lai (2010)'s study on the clan of Chen in Hai-Ning county (海宁-陈) 1600-1850, a very clan with great exam achievement in the lower Yangzi, as many as 10% of males qualify for "gentry" and 22% qualify for "near gentry" class. In another word, the clan's relative representation among "gentry" was about 10 for 1600-1850, and could be even higher for the more exclusive group of juren, who only accounts for 0.03%-0.06% of all males. Throughout this period 16.5% of males never get married. I break down the other 51.5% of population into 20% of moderate wealth and 31.5% of "no indicator of tile or wealth". Applying the numbers in column 4, the average sons by a male is 1.76, 10% higher than the benchmark population.[11]

In column 7, I created an even more elite kinship, with higher share of males being gentry and having wealth, and lower share of males never being married. The average sons by a male is 1.93, 20% higher than the benchmark population.

Figure 1.17: Persistence Rates with Faster Population Growth: North Zhejiang 1645-90 Elite

[11]Lai (2010, p82-83) also reports sons per male by two categories: title and no title. On average, title holders have 14.6% more sons than non-title holders (for those who are ever married). This result is very similar to Shuie's results (12.9%) if I combine her category 3 and 4 into group of "with title", as and combine her category 1 and 2 into group of "no title".



To test how sensitive our b estimates are to differential rates of population growth in figure 1.17, we assume that the population growth rate of the North Zhejiang elite S-POs was 10% or 20% higher than general population per generation 1645-1949. If 10% greater growth rate is applied to the elite S-POs, population share of this elite group in 1940 should be 2.4 times that in 1650, 9 generations before. The initial relative representation should be 2.4 times greater than we calculated. But this assumption, as can be seen in the figure reduces the average b estimated 1645-1949 of 0.844 only to 0.828. Similar, if 20% of greater growth rate is applied to the elite S-POs, the b estimated 1645-1949 reduced to 0.817. Thus our b estimates are not particularly sensitive to differential population growth rates among the elite S-POs.[12]

1.3.5 (c)Sampling error

The third issue is sampling error. Each of the Qing elite groups contains 40-62 individual S-POs. But these S-POs have very different levels of relative representation, and they may have regressed towards the mean of 1 at different individual rates. To get an idea of the likely sampling error we constructed 1000 new samples of S-POs for each of the original samples, selecting the S-POs with replacement from the original sample (bootstrap). For example, as shown in figure 1.18 we resample the 1870-1900 North Zhejiang elites (with 62 S-POs) for 1000 times, from which we can obtain the upper and lower bounds of 95% confidence interval of relative representation at each period of time.

Figure 1.18: resample the 1870-1900 Zhejiang elites and 95% conf. interval

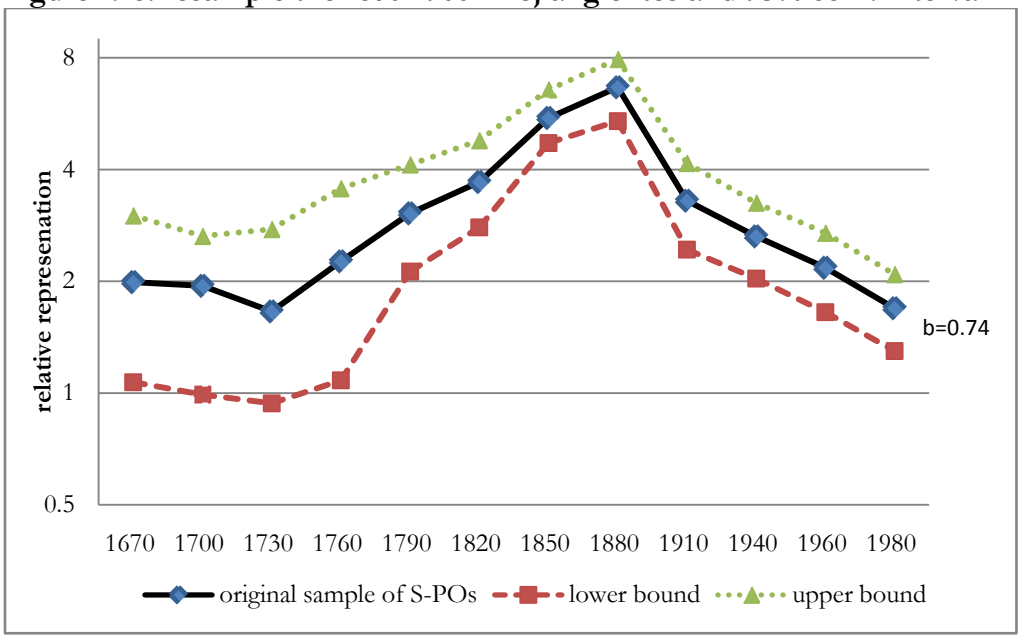


Table 1.6 shows the 95% confidential intervals ofbs for the periods 1645-1810, 1781-1900, 1881-1950, and 1930-90 (or 1950-2010). For a given sample of S-PO elites, we

constructed 1000 new samples of S-POs for each of the original samples, selecting the names with replacement from the original sample, and then estimated b for each of these new samples for the period of time that we are interested in. Then we calculate the standard error and confidence interval of these bs for each sample of S-POs.

We only report results from selective samples of S-POs. For example, the 1645-1690 elites' relative representative is much greater than 1 before 1950, and bootstrap give a relatively smaller range of bs. But after 1950, those S-POs are close to mediocrity and boostrap gives a much larger range of bs. (The standard error, obtained from the 1000 bootstrapped samples, is large after 1950.) Even the 1870-1900 elites become close to mediocrity in recent decades. Thus to get a better estimate of the b for the Communist era we will in the future derive a new S-PO sample of elite surnames specifically for the period 1930-49, just before the Communist takeover.

Table 1.6: b Estimates from Resampling

	North Zhejiang			**South Jiangsu**		
Interval	1645-1690 elite	1781-1810 elite	1870-1900 elite	1645-1690 elite	1781-1810 elite	1870-1900 elite
1645-1810	0.75-0.86 (0.83)			0.81-0.88 (0.85)		
1781-1900	0.87-0.97 (0.92)	0.79-0.88 (0.87)		0.93-0.98 (0.95)	0.87-0.95 (0.92)	
1881-1950		0.76-0.88 (0.84)	0.73-0.84 (0.78)		0.82-0.91 (0.86)	0.79-0.88 (0.81)
1930-1990			0.68-0.80 (0.74)			
1950-2010						0.59-0.74 (0.66)

Note: in ( ) are βs obtained from original samples

1.4 Data and Results: National Surname elites

1.4.1 National surnames and regional controls

For the Communist era most of our information on the status of surnames comes from sources at the national level. As confirmation on these results for the lower Yangzi, we can also look at individual surnames which at a national level were of average high status in the Imperial era and measure what their current status is.

We identify our elite sample as surnames which showed an unusually high frequency among *Jinshi*(进士), the national exam passers at the highest level of the exam system in the late Qing (1820-1905). Surnames were included in this sample if they showed a relative representation among *Jinshi* at least 4 times that of the three most common Chinese surnames, 王 (Wang),李 (Li)，and 张 (Zhang), constituting 21.4% of the modern Han population. There were 13 such national elite surnames.

However, these 13 surnames are all concentrated in the lower Yangzi among the modern population, and all lower Yangzi surnames tend to be overrepresented both among exam passers in the Imperial era and in modern elites. Geography still matters to social status in China. This will make rates of social mobility lower, but to exclude this geography effect and measure just social mobility within an area like the lower Yangzi we calculate the relative representation of these 13 Qing elite surnames in modern elites by comparing them to common surnames concentrated now in the lower Yangzi. We thus take their relative

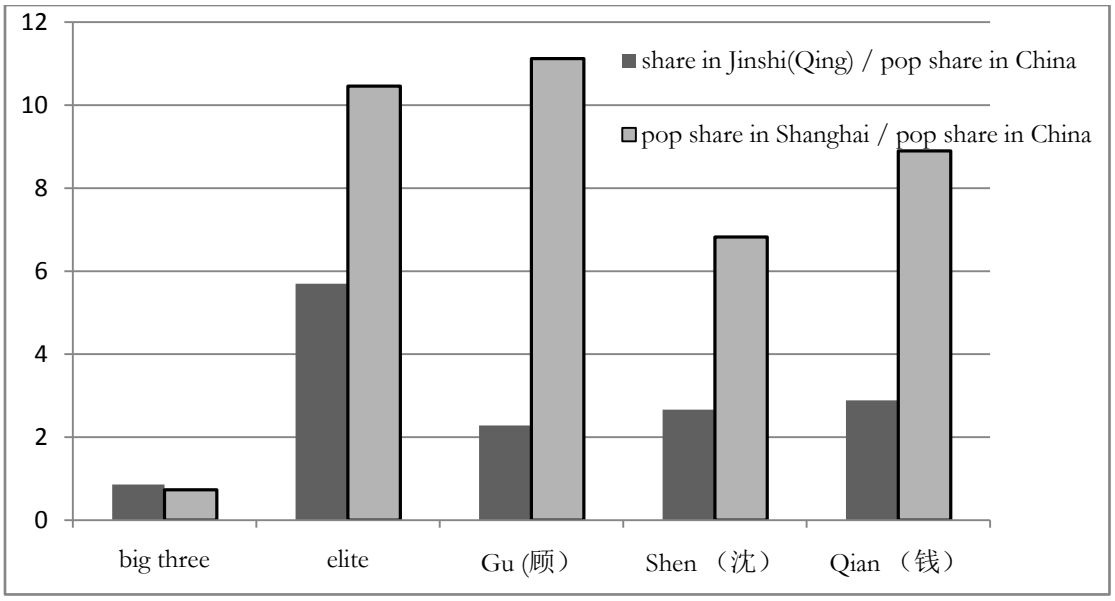
representation of elite surnames[13]compared to three equally regionally favored surnames, Gu, Shen and Qian, the "regional 3", that have only an average status at the regional level.

To establish that the Qing elite surnames are as regionally concentrated as the "regional 3", figure 1.19 displays the relative frequency of the surname groups among those withhousehold registration in Shanghai in 1997. The late Qing elite surnames and the regional surnames are equally concentrated in Shanghai relative to the big three, which are modestlyunderrepresented in Shanghai. Although the "Regional 3"are all overrepresented by 2 to 3 in late Qing elite,this is due to regionaladvantages and also perhaps lower population growth in the lower Yangzi, so that the population shares of this area were higher 1820-1905.

Figure 1.20 portrays the relative frequency of the different surname groups among elite groups over time. Even using the "regional 3"as the standard, for example, the late Qing elite surnames had a relative representation in the Qing era of 3.37 among *Jinshi* 1870-1905.By 1912-49 under the Nationalists that relative representation had fallen to 2.31 for government officials. By 1949-88 under the Communists it fell further to 1.81 for members of the National People's Congress (NPC, 全国人大) and the National Committee of the Chinese People's Political Consultative Council (CPPCC, 政协). Even in the most recent

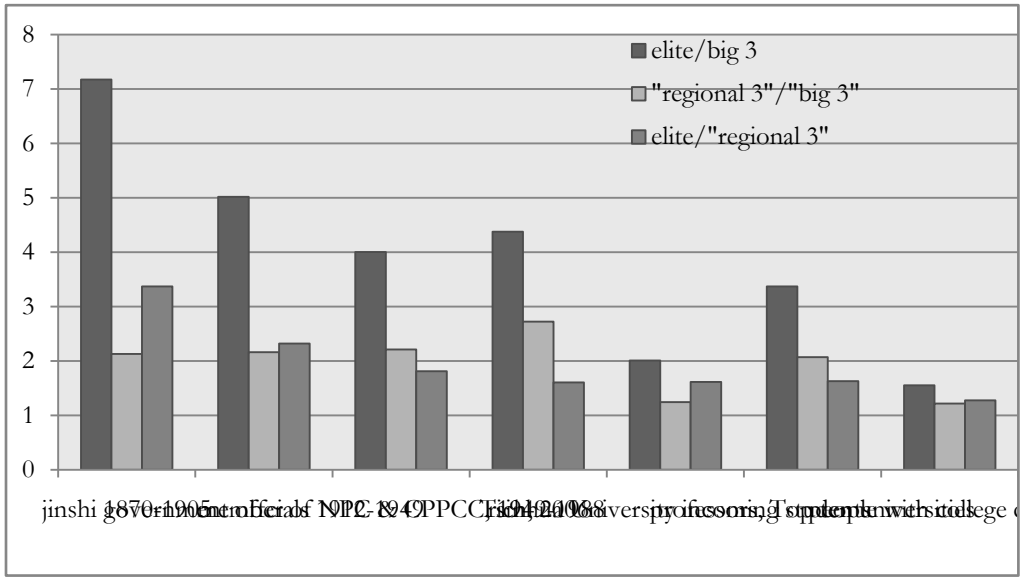
Figure 1.19: late Qing elite surnames and regional control surnames

[13] The late Qing elite surnames are 13 lower Yangzi concentrated surnames that are still more representative than "regional 3" in 1820-1905. They are 诸，竺，茅，濮，裘，巢，恽，端木，钮，忻，萨，笪，宓, totaling to be 0.058% of population



Source: *jinshi* names from *Name List of jinshi in Ming and Qing (明清进士题名录)*
Modern population share from CNIIC, see appendix
Shanghai population share is in 1992 by residency (with Shanghai *hukou*)

Figure 1.20: Relative Representation of Surname Groups among Recent Elites (RR=elite share/pop share)



elites, these surnames remain overrepresented relative to common lower Yangzi surnames.Their relative representation among the rich of 2006 was 1.61, among professors at the top 7 universities in China 2012 1.63, among incoming students to Tsinghua University 2003-11 1.62, and among college degree holders in 2010 1.28. This implies a b of 0.69-0.70 for the last four generations, assuming these elites are on average the top 0.1% of the population. So at these higher levels of status social mobility in the Communist era is estimated to still be very slow.

Note that if we did not control for geography, and took the relative representation of the late Qing elite as that relative to the national "big 3" surnames, then estimated social mobility rates would be much lower, and indeed just as low as in the Imperial era. Below we estimatemobility rates from two sequences of modern elites: educational elites measured by incoming college students, educationalelites measured by publications, and political elites measured by government officials. For each category, we do this by tracking relatively representation of elite surnames relative to that of "regional 3".

1.4.2 Educational elites

One modern indicator we have of high social status, in terms of an educational elite, is publications by surname. Table 1.7 shows the sources available on publications at different periods. Publications represent, however, a shifting level of eliteness in society over time. Before 1949, for example, less than 0.17% of the population had college degrees, and likely to be publishing. Post 1978 those with college degrees increased to 2.23% of the population. To try to keep the selectivity of this elite constant for 1977-83, and 1998-2000 we used only Engineering Publications from the CNKI dataset. And for 2010-12 we used PhD dissertations, doctoral degree recipientsthen accounting for less than 0.1% of the entire

population. However, the most elite students in China in this period would complete PhDs abroad.

With publications it is also not generally possible to divide authors into discrete birth cohorts.Authors at any time will mostly be in the age range 30-60, but some will be younger and older. So we can only relate these publication records to rough birth cohorts. The special case of the Cultural Revolution makes it easier. The 1930-1948 birth cohorts should have entered college 1948-1966, and the later birth cohort could not enter college until 1977 when the college entrance examination was restored. So we can be sure that the vast majority of publications 1977-1983 were by the 1930-1948 birth cohorts, and the vast majority of those publishing in China in 1998-2000 completed their college education after 1981.

Table 1.7: Publications Sources, 1900-2010

Period	Publications Source	Years of publication	Indicated birth cohort	Number of observations
1900-49	*Dachenglaojiu*(大成老旧) dataset, publications of the Republican Era: www.dachengdata.com	1910-1930	1870-1890	41,112
		1940-1950	1890-1920	103,217
1977-2000	CNKI dataset (first author only, EI source, only mainland China)	1977-1983	1930-1948	46,780
		1998-2000	1949-1970	96,069
2000-10	Dataset of PhD dissertation from Chinese universities	2010-2012	1975-1985	37,588

Figure 1.21: Publication Rates by Surname and Implied bs

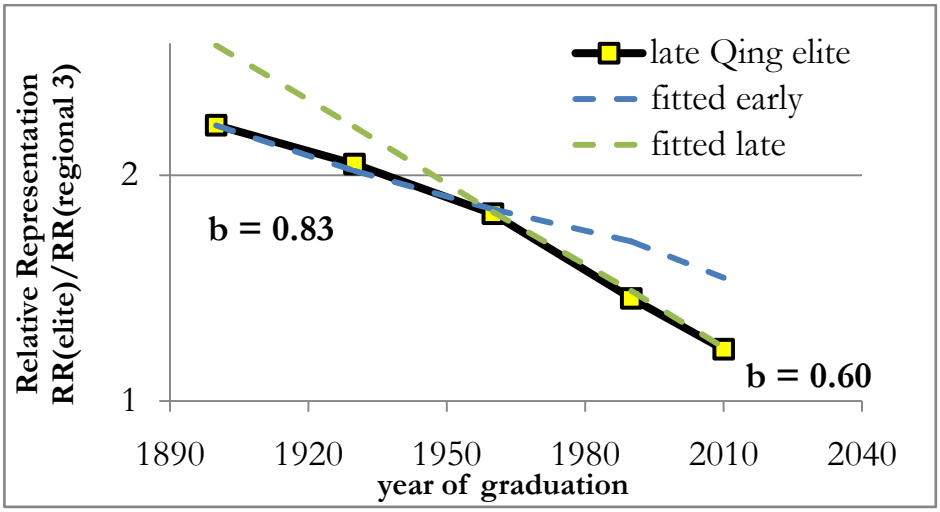


Figure 1.21 shows the relative representation of the late Qing elite surnames relative to the "regional 3" by estimated graduation period. For the first three cohorts, those who were educated under the Imperial and Republican eras, the implied rate of social mobility is very low.The estimated b for these observations is 0.83, implying a very strong inheritance of status. This is true for publications in Communist China, 1977-83, whose authors received high education before the Cultural Revolution.

But if we instead estimate the generational b to fit the last four cohorts the implied b is much lower, 0.60. In the Communist era, measured by publications, there has been a much more rapid downward mobility of the late Qing elite surnames. The rates of educational and occupational mobility implied by b of 0.60, however, are not low by the standards of such measures for parents and children in various modern societies. However, as noted, a problem with this measure in later years is that they do not include the highest status academics. The publications include only those in Chinese journals, not the highest prestige foreign journals. And the measure of PhDs excludes the substantial number of the most

high status Chinese students who earned their PhDs abroad in the US and other developed countries.

1.4.3 Political elites

Another measure of status is participation in political elites. For the Republican era we have lists of civil and military officials in the years 1912-49. For the Communist era we have first members of NPC and CPPCC in the years 1949-88, then central government officials 1992 and 2010. Table 1.8 summarizes this data.

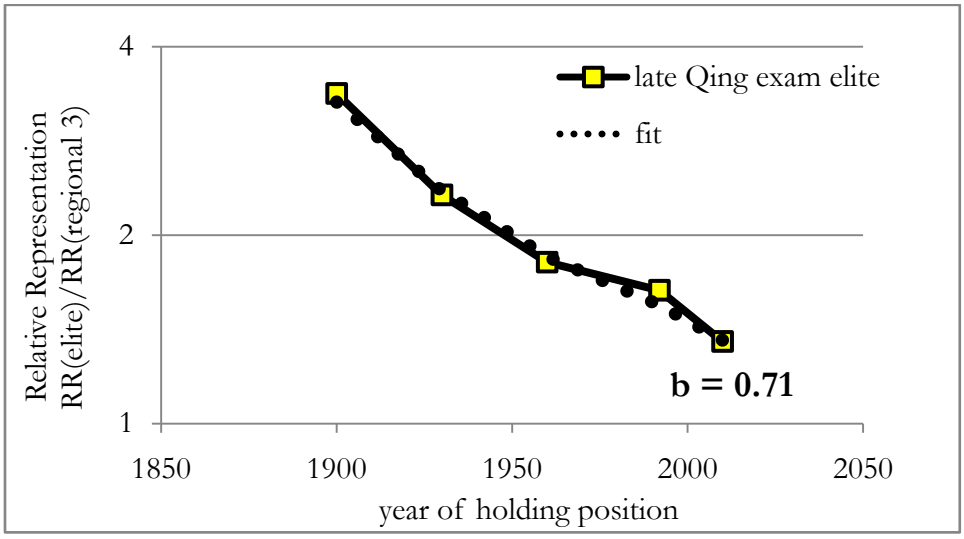
Figure 1.23 shows the relative representation of the late Qing Exam Elite surnames, relative to the "regional 3". This goes from 3.26 in late Qing, to 2.3 in the Republican era, and to 1.37 by 2010. The figure also shows the best fitting b for this elite, which fits well both for the Republican and Communist eras, is 0.71. Thus even in the Communist era there has been high persistence of the exam elite surnames among the modern political elite, even slightly more persistence than for education. This is surprising given that coming from an established elite background was a great political disability through much of the Communist period. The ability of this group to retain an above average representation in spite of the handicap many possessed in this dimension testifies to the strength of persistence in social position, even with such dramatic regime changes.

Table 1.8: The Qing Exam Surnames among Political Elites

	Civil and military officials (1912-49)	Members of the NPC and CNPPC (1949-88)	Central officials (1990)	Central officials (2010)	Population Han only (2010)

	Share %	Share %	Share %	Share %	Share %
Elite Sample size	31,194	16,320	10,962	11,636	-
BIG 3	17.7	19.1	22.7	23.5	22.0
Gu/Shen/ Qian	1.40	1.59	1.71	1.71	0.85
Late Qing Elite	0.236	0.202	0.197	0.172	0.059

Figure 1.23: Qing Elite Representation among Modern Political Elites



1.5 Interpretation

How do our results for China compare to conventional intergenerational studies? There is debate over the rate of intergenerational mobility within the late Imperial period. Early studies of the exam system suggested substantial mobility, even within the Imperial era, when looking at the fathers and grandfathers of degree winners (Ho, 1964). In the early Qing only 50-60% of*juren* and *jinshi*had fathers or grandfathers who had attained that status,though

by the middle and late Qing thishad risen to 60-70%. If even half of all sons of *juren* attained *juren* status, and the process was Markov, then this would imply that the relative representation of elite surnames would drop by a half in each generation. We see above in figures 1.11-1.14 that the rate of decline of relative representation is much slower than this.The Ho study somehow is failing to predict the long run downward mobility rates of these elites.

However, as we explain in the second part, looking at the correlation of fathers and sons on any single dimension of social status will produce estimates of the persistence of status that greatly underestimate the true persistence rates of more general measures of status, or the persistence of status even on a single dimension over many generations. While the son of a *juren* may be unlikely to attain *juren* status, he may with high probability inherit many other aspects of status such as wealth or occupation, so that his general social rank is similar to that of his father. In all, social status is difficult to measure at individual level.

When kinship background is taking into account, scholars of China do find lower social mobility than when only parental status is considered. An important source here is the published*Collection of exam essays (*清代朱卷集成), which contains details of the kinship backgrounds of 8,000 exam passers. 87% of *juren* came from families containing *juren* or *jinshi* within the previous five generations (Zheng, 2006). This result suggests a very low mobility rate, consistent with our surname estimates.

The established estimates of b for Communist China vary greatly depending on the measure, the data and the methodology used. The urban intergenerational correlation of income is reported to be only 0.32-0.63 for the post-reform era(Gong, Leigh, and

Meng,2010, Guo and Ming, 2008).[14]But these estimates are based on one dimension of status, and only two generations, and as argued above (and in Clark and Cummins, 2012) they will underestimate the persistence of status on more global measures.

Others report lower mobility rates in last ten years relative to 1977-1990 due to greater inequality in higher education and job attainment (Chen, Naidu and Yuchtman, 2010, Wu and Zhang, 2010). Walder and Hu (2009) report a different pattern of status inheritance among the old elites and the new elites. The descendants of the old elites(地富反坏右)—landlords and businessmen before 1949—had to pursue professional careersinstead of politics after 1949because they were discriminated against in the Communist Party. Nonetheless, they maintained advantages in education attainment. Once the discrimination inside the Party was lifted after 1976, they performed much better in all occupations relative to descendants of Old poor (贫下中农), those whose forbears were tenants and workers before 1949 (and were given priority earlier in becoming party members). Liang and Lee (2012) study the occupational source of college students of Peking University and Suzhou University, and find that the share of students from lower class families increased greatly 1952-1966, but changed little thereafter.

In sum, tracking relative status of surname cohorts can allow us to acess social mobility over multiple generations and evaluate the impacts of policy changes. The Qing Exam elite survived through the Republican Era (1912-1949) and Communist Era (1949-). The Qing elite surnames are still overrepresented in modern elite populations. Rates of social mobility did seem to increase with the advent of the Communist era. But mobility rates since 1949

[14]Wu and Trieman (2007) arguethat taking into account urban-rural inequality, the correlation should be even higher.

estimated for surname groupings remain low compared to most modern estimates of individual social mobility rates.

Some scholars have interpreted this slow mobility and persistence of elites into the present as revealing the importance of kin networks, and extended family strategies (Campbell and Lee, 2011, 2012). Given the nature of the provincial exam (the pass rate is 0.33-0.5%), it was unlikely that a nuclear family could produce *Juren* or *Jinshi* over consecutive generations. For any individual child luck was as important as talent in determining success in the exams (Elman, 1992). The best strategy, then, for the heads of kinships was to pool resources and create public goods within kingships, or club goods (exclusive of other kingships).

First, kinships allotted a portion of kin land as "land for education" (学田), with the rent going to special funds for education and exam preparation. Children from poor families could get financial support from other families within the kinship. Second, they picked a select fewof the most talented children within the kinship, hired them the best teachers, built lineage-based schools (族学), and hadthese chosen children take the exams as often as possible. Other children were allocated to variousalternative occupations for risk diversification.

Success in the exams did not only bring glory to the chosen candidates their common ancestor, but also protection of property rights of the entire kinship. As they obtained more land and wealth, they had more resource to invest in education. In all, the relative status of kinships was more stable over generations than that of individual families (Hymes, 1986). For "outsider" kinships, it may take one or two generations to become rich and literate, and

get intothe lower gentry class (*Shengyuan*, 生员/秀才, the lowest degree, 0.4-0.7% of male population in the late Qing). But it took several generations for a commoner family to produce *Juren* and *Jinshi* (5% and 1% of *shengyuan*, respectively). That is why it always took hundreds of years for an elite kinship to regress to mediocrity.

When the exam elite had to abandon traditional education in the last decade of the Qing, they did not hesitate to go abroad to acquire modern education. Children of elite kinships were privately tutored, and then sent to expensive middle schools and universities to take modern education. Elites were more inclined to live in cities (absentee landlords) and had less incentive to invest in kinship exclusive public goods in their villages of origin. Nonetheless, wealth and talent became more important to decide one's success than kinship backgrounds. So we expect that exam elites persist into Republican era, but modern elites came from various kinship backgrounds (Ye, 2007). That means a modest increase in social mobility. Our empirical findings support this view.

Another possible interpretation, however, is that the persistence of kin groups among the elite is exactly the same phenomena that is found in other societies, and has little to do with the peculiar importance of kin networks in Chinese society. Table 1.9, for example, shows the b values estimated for wealth, education and occupation in England and Sweden over the last 160 years. These average b values are high, in the order of 0.7-0.8. Yet this not because of the importance of kin networks in English and Swedish society, but because the b estimated from surnames measures a different thing than the b estimated from individuals. The surname b is a measure of the rate of long run social mobility of overall social status for families, while is slower than the one-generation rate of mobility on individual aspects of status that the traditional mobility studies measure. That is why surname mobility rates in

Sweden or England, where nuclear families dominated and kin networks beyond these were unimportant, are almost as slow as those of Imperial and Republican China. It thus remains to be determined to what extent it was kin networksthat produced the strong persistence of surname elites in China.

Table 1.9: b Values for England and Sweden from Surnames

Period	England Wealth at death	England Education	Sweden Education	Sweden Doctors
1850-99	0.71	0.78-0.81	0.75-0.82	-
1900-49	0.71-0.86	0.78-0.81	0.85	0.70
1950-2012	0.61-0.68	0.78-0.81	0.66	0.70

Sources: Clark and Cummins, 2012, Clark, 2012.

References:

Bian, Y. (2002). "Chinese Social Stratification and Social Mobility." *Annual Review ofSociology*, 28, 91–116.

Campbell, C., and Lee, J. Z.(2011). "Kinship and the Long-Term Persistence of Inequality in Liaoning, China, 1749-2005."*Chinese Sociological Review* 44 (1), 71-103.

Campbell, C., and Lee, J. Z.(2010). "Social, Economic, and Demographic Determinants of Descent Line Growth and Extinction over the Long Term in Historical China."*California Center for Population Research working paper*

Cao, S. J. (1998),"population loss in lower Yangzi due to Taiping Rebellion"*, History Studies.* 曹树基，"太平天国对江南人口的影响"，《历史研究》，1998年第二期.

Cao, S. J.(1997), *the history of migration in China, volume 6*, The People's Press of Fujian.曹树基，《中国移民史，第六卷》.福州：福建人民出版社.

Chang, C. L. (1955). *The Chinese Gentry: Studies on Their Role in Nineteenth Century Chinese Society*.Seattle: University of Washington Press.

Chen, Y. Y., Naidu, S., and Yuchtman, N.(2010). "Intergenerational Mobility and Institutional Change in 20th Century China."*UC Berkeley working paper.*

Clark, G. (2012). "Swedish Social Mobility from Surnames, 1700-2012." Working Paper, UC Davis.

Clark, G., and Cummins, N. (2012)."What is the True Rate of Social Mobility? England, 1800-2012." Working Paper, University of California, Davis.

Corak, M. (2012). "Inequality from generation to generation: the United States in comparison."*The Economics of Inequality, Poverty, and Discrimination in the 21st Century, ABC-CLIO, forthcoming.*

Du, R., Yuan, Y., Hwang, J., Mountain, J., and Cavalli-Sforza, L. L. (1991)."Chinese Surnames and the Genetic Differences between North and South China." Department of Genetics, Stanford University, School of Medicine.

Elman, B. A. (1992). "Political, Social, and Cultural Reproduction via Civil Service Examination in Late Imperial China." *Journal of Asian Studies* 50(1):7-28.

Guo, C. B. and W. F. Min, (2008). "Education and Intergenerational Income Mobility in Urban China,"*Front Education China*, 3, 22–4.

Gong, C. H., Leigh, A. and Meng, X. (2010). "Intergenerational income mobility in urban China", IZA Discussion Paper No. 4811, Institute for the Study of Labor, Bonn.

*Ho, P. T. (*1959). "*Studies of Population of China, 1368-1953",* Cambridge, Mass: Harvard University Press.

Ho, P. T. (1964). *"The Ladder of Success in Imperial China: aspects of Social Mobility (1368-1911)."* New York: Wiley.

Ho, P. T.(1999). "The Salt Merchants of Yang Chou: A Study of Commercial Capitalism in Eighteenth Century China" *Research in Chinese Social Economy*

Hymes, R. P.(1986).*"Statesmen and Gentlemen: The Elites of Fu-chou, Chiang-hsi in Northern and Southern Sung."* Cambridge: Cambridge University Press.

Jiang, T.(1994).*Population in modern China.*The People's Press of Zhejiang.姜涛,《中国近代人口史》.杭州：浙江人民出版社.

Lai,H. M. (2010). *Elite lineages and imperial families in Qing time.* Peking University Press,.赖慧敏，《清代的皇权与世家》.北京：北京大学出版社.

Li, B. Z. (2000). *The early industrialization in Jiangnan1550–1850."*Beijing: Publishing Press of Social Science Materials. 李伯重，《江南的早期工业化》，北京：中国社会科学文献出版社.

Li, N.(2011). "Essays on Cultural Diffusion, Migration, and Human Capital: Investigation from China's Historical Experience." Department of Social Science, Hong Kong University of Science and Technology, Dissertation.

Liang, C., and Lee,J. Z. (2012). "Revolution in silence: social source of university students of China, 1952-2002." *Social science in China, 2012-1, 98-118.*梁晨，李中青，"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年第一期.

Rawski, E.(1979). *Education and Popular Literacy in Ch'ing China.*AnnArbor: University of Michigan Press.

Shiue, C., (2012). "Human Capital and Fertility in Chinese Clans, 1300-1850", NBER working paper

Solon, G., (1989). "Biases in the Estimation of Intergenerational Earnings Correlations," *Review of Economics and Statistics,*February 1989, 71, 172-4.

Walder, R., and Hu, S. H.(2009). "Revolution, Reform, and StatusInheritance: Urban China, 1949-1996." *American Journal of Sociology* 114:5 (March): 1395-1427.

Wang, Y. J., (1960). "Western Impact and Social Mobility in China." *American Sociological Review,* **25**(6): 843-855.

Wu, X. G., andTreiman, D. J. (2007). "Inequality and equality underChinese socialism: The *Hukou*system and intergenerationaloccupational mobility". *American Journal of Sociology*, *103*(2), 415–445.

Wu, X. G., and Zhang, Z. N., (2010). "Changes in Educational Inequality in China, 1990-2005: Evidence from the Population Census Data." *Research in Sociology of Education*, Volume 17, 123-152.

Wu, R. A. (2009). *A brief history of lower-Yangzi decent kinships, 1368-1911*. Shanghai: ShanghaiRenmin Press.吴仁安，《明清江南著姓望族史》，上海人民出版社.

Ye, F. G.(2007). "Higher education and social mobility in China, 1840-1922" in Zhang and Chen, eds, *"Universities: social stratification and social mobility."* Beijing: Beijing Normal University Press. 叶赋贵，"中国高等教育和社会流动"，王晨，张斌贤编，《大学：社会分层和社会流动》，北京师范大学出版社.

Yuan, Y. D., and Qiu, J. R. (2010). *The encyclopedia of Chinese surnames*. Jiangxi: Jiangxi Renmin Press. 袁义达，邱家儒，《中国姓氏大辞典》，江西人民出版社.

Yuan, Y. D., and Zhang, C. (2002). "Surname of China: Population Genetics and Distribution of Population" Shanghai:Huadong Normal Uni. Press. 袁义达，张诚，《中国姓氏，群体遗传和人口分布》. 上海: 华东师大出版社.

Yuchtman, N. (2010). "An Economic Analysis of Traditional and Modern Education in Late Imperial and Republican China." Working paper, Haas School of Business, UC Berkeley.

Zheng, R. L.(2006)."Research on the Relationship between Examination and Society", Xiamen University PhD dissertation.郑若玲．" 考试与社会关系研究 "，厦门大学博士论文.

1.A.1. Appendix:Surname sources:

Exam elite names (national): (1) *Jinshi roster of Ming and Qing Dynasty* (book) 明清进士题名录;

Exam elite names and places of origin (lower Yangzi): (1) *Zhejiang juren list obtained from Chronicle of Zhejiang Jiang, Chapter of who's who (*民国浙江通志-选举志*); (2) Jiangsu juren list obtained from Chronicle of Jiangsu, Chapter of who's who* (江苏省志-选举志).

Republican era elite names (national): (1)*High Ranked civil and military leaders: Republican era* (published book) 中华民国职官志; (2) *Name list of college faculties (1941-1944).*

Republican era elite names and places of origin (lower Yangzi) (1900-1949): alumni books and university yearbooks, University students in Japan (1906), Peking University (1905-1948) , Tsinghua University (1911-1937), Nanyang University (1898-1925), Central University (Nanjing University) (1916-1936, 1945-1947), Zhejiang University (1918-1947), Wuhan University (1922-1935), Yanjing University (1924-1928), Datong University (1923-1935, 1940-1948), (1906 年留日学生，京师大学堂/北京大学，清华大学，南洋公学/交通大学，南高/东南/中央大学，求是高等/浙江大学，武汉大学，燕京大学，大同大学).

Communist era elite (national): (1) government official*High Ranked civil and military leaders: Communist era (1949-1988)* 中华人民共和国职官志**(2)***1992 and 2006 yellow book of China government agencies* (published reference book); (3) 2006 census on 1,400,000 China enterprises (from which we selected those with assets of ￥100 millions and above, 130,000 chairman of board).

Communist era elite (lower Yangzi):

Notable People of Zhejiang Province(1900-1970)
http://www.zjda.gov.cn/zjda/module/search/SearchAction.do?method=search&strTypeCode=0008;

Search Engine for Nanjing University incoming students (1952-2011),
http://dawww.nju.edu.cn/pub/?id=1

Dead soldiers' surnames, 1933-49: county chronicles published in 1980s-1990s, chapter of who's who: honored soldiers (各县县志，人物志，革命烈士英名录).

Population, ethnicity, and Educational attainment of 1500 most common surnames (national); regional distribution of selected (successful) surnames: purchased from China's National Identity Information Center （CNIIC, 全国公民身份证信息服务中心）, whose data came from China's household registration (*hukou*) which covers the entire population.

1.A.2. Appendix: Rare Elite Surnames, National

	Population 2010	Lower Yangzi prefecture of *jinshi*/other provinces	Qing *jinshi* 1820 to 1905	Officials1912 -49	Top rich 2006	People with graduate degree per 10,000	Professors top 10 universities 2012
big three	270.5 m.	All over China	1,702	4.737	22,934	26.9	4,924
			Lower Yangzi Regional three				
Gu 顾	2,675,692	苏锡常, 杭州,南通, 嘉兴, 扬州/ 江西, 河南, 河北, 贵州,广西,云南	41	88	643	29.0	112
Shen 沈	5,305,442	杭嘉湖, 苏锡常, 南通, 扬州,宁绍, 上海/河南, 河北, 广东, 福建	85	201	1186	26.8	184
Qian 钱	2,446,177	杭嘉湖, 苏锡常, 南通, 扬州, 上海/ 云南,河南,贵州	36	106	580	29.1	97
			Lower Yangzi Elite surnames				
裘	135,594	宁波, 绍兴/江西	4	14	51	33.8	6
诸	118,161	杭嘉湖，绍兴	2	4	51	32.0	6
茅	99,073	镇江， 扬州	2	7	39	36.2	7

竺	67,163	宁波	2	6	26	33.1	8
濮	61,612	南京/安徽	8	4	22	30.8	5
巢	50,917	常州	1	1	12	30.2	2
钮	75,654	湖州, 嘉兴, 上海	3	8	12	29.7	6
笪	23,054	镇江	1	1	6	37.3	1
宓	22,668	宁波	1	3	11	45.0	1
恽	17,476	常州/北京	7	5	15	50.4	2
端木	13,303	南京	2	2	8	38.3	1
萨(闽)	9,048	福州	5	6	1	47.7	2
忻	25,432	宁波	1	1	15	57.8	4

Note: 1. I reported Han population for*all* surnames*Except for* **萨(闽)** for which I report population whose ethnicity is Mongolia.

2. Top universities that I collected faculty surnames for are Peking University, Tsinghua university, Renmin University of China, Shanghai Jiaotong University, Zhejiang University, Fudan University, Nanjing University

1.A.3 Appendix: Lower Yangzi Elite Surname-Place of Origin

area	prefecture	county	1645-1690 elite	1780-1810 elite	1870-1900 elite
江南	苏州	吴县长洲元和	韩，申，缪，管，彭，宋，范，汪	潘，石，戴，毛，席，陶，郭，宋，彭，程	潘，彭，程，曹，范，汪，查，陶，孔，贝
		常熟昭文	翁，蒋，曾，汪，归，赵	屈，瞿，言，席	庞，翁，曾，归，浦，季
		吴江震泽	叶，汝	费	庞，汝，柳
		太仓		毕	
	常州	武进阳湖	庄，龚，恽，屠，须，瞿，岳，董	盛，卜，吕，陶，庄，龚，瞿	程，费，瞿，董，恽，姜，汪
		宜兴荆溪	路，万，储	潘，路，任，崔	潘，任，崔
		江阴	章，季，金，惠	缪	章，沙，祝
		无锡金匮	鲍，安，侯，施，秦，嵇	邹，华，侯，秦	孙，薛，陶，裘，过
	镇江	丹徒	欧阳，何，卞，符，荆，樊	茅，柳，戴	茅，鲍，唐
		丹阳	贺，荆	吉	
		溧阳	费，任，潘	狄，彭，强	狄
		金坛	冯，于，史，曹	虞	
浙北	杭州	钱塘仁和	汪，翁，查	许，汪，翁，翟，项，阎，成，戴	许，汪，濮，樊，夏，武，钟
		富阳	裘		
		萧山	任，来	瞿，钟，蔡	施，韩，林，汤
		海宁	查，郭，羊，祝	许，祝，查，	查，都，蒋
	绍兴	山阴会稽	阮，茹，祁，姜，鲁	童，商，茹，平，姜，杜，史，樊，秋	马，何，冯，杜，石
		诸暨	余，钱		斯
		上虞			经，钱
		余姚	史，诸	吕，邵	翁
	嘉兴	嘉兴秀水	屠，谭，盛，吕，范，虞，曹	汪，虞，何，怀	郑，殷，褚
		嘉善	丁，孙，支，魏	周，孙，浦	郁，夏
		海盐	俞，萧，曹	顾，吾	富，任
		平湖	屠，胡，冯	屈，奚	戈，屈，奚
		桐乡石门	钟，唐，劳	冯，皇甫，蔡	周，蔡
	湖州	乌程归安	闵，凌，戴	章，闵，凌，严，崔	钮，梁
		德清	蔡，车，戴，嵇	谈，冯	俞，傅
		长兴	臧，叶，丁，蒋	臧，孙	
	宁波	鄞县	左，万，戎，倪，管，范，	卢，范，郭，竺	袁，范，童，忻，水
		镇海	谢	谢	谢，郑，盛，方，虞
		慈溪	秦，冯，姜，向	冯，任，董	叶，宓，魏，
		奉化	邬		孙

1.A.4 Appendix: proof of proposition 1

The OLS estimate from the following equation:$y_{t+1} = \beta y_t + v_t$, where y is the partial measure of status, then

$E(\beta)$will be$b\frac{\sigma_x^2}{\sigma_x^2+\sigma_u^2} > b$, where x is the true underlying social status, linked to partial measure as below,

$$y_t = x_t + u_t, \quad \text{and } x_{t+1} = bx_t + e_t$$

$x$is assumed to have a mean of 0, and a constant variance σ^2, and is normally distributed.
e_t is the random shock that occurs when parents transit social status to children. It is not related to the social status of parents, x_t, and is not auto-correlated.
u_t is a random component linking the underlying status of the family to the particular observed measure of status. It is not related to the social status, x, and is not auto-correlated.
$v_t$is the random shock that occurs when parents transit their earnings, wealth, year of education, or occupational status (again, not correlated to y_t and not auto-correlated).

Proof:$E(\beta) = \frac{\sum_i (y_{t+1} y_t)}{\sum_i y_t^2}$, wherei is individual family line

$$= \frac{\sum_i (x_{t+1} + u_{t+1})(x_t + u_t)}{\sum_i (x_t + u_t)^2}$$

$$= \frac{\sum_i x_{t+1} x_t + \sum_i x_{t+1} u_t + \sum_i x_t u_{t+1} + \sum_i u_{t+1} u_t}{\sum_i x_t^2 + \sum_i x_t u_t + \sum_i u_t^2}$$

$$= \frac{\sum_i (x_{t+1} x_t)}{\sum_i x_t^2 + \sum_i u_t^2}$$

Since x_t and $u_t$are not correlated (including their lagged terms), and $u_t$is not auto-correlated.

$$= \frac{\sum_i b(x_t + e_t) x_t}{\sigma_x^2 + \sigma_u^2}$$

$= b\frac{\sigma_x^2}{\sigma_x^2+\sigma_u^2}$Because$x_t$ and $e_t$are not correlated.#

So if we don't observe the complete social status of families x, but some partial measure y, the intergenerational correlation of y that derived from OLS estimation,β, will systematically underestimate the intergenerational correlation of x, b(or overstate the mobility rate). The bias is greater when y is a poor measurement of x (for example, the sample contains a relatively homogenous group of people in terms of y).