



WIDER Working Paper 2020/28

Measuring social mobility rates in earlier and less-documented societies

Gregory Clark*

March 2020

Abstract: In societies where surnames are inherited from parents, we can use these names to estimate rates of intergenerational mobility. This paper explains how to make such estimates, and illustrates their use in pre-industrial England and modern Chile and India. These surname estimates have the advantage that they require much less data than traditional parent–child estimates. They are also more robust to errors in status data. Thus, they can be used to estimate social mobility rates in early societies such as England 1300–1800, or in less-developed societies now. Surnames measure a different aspect of social mobility than conventional measures, but this surname measure is the one that matters when we consider group-level convergence of social status, or the time needed, measured in generations, for any disadvantaged groups to attain at least average status. Surnames thus allow us to measure a key element in the multigenerational mobility process.

Key words: surnames, intergenerational mobility, long-run mobility, underlying mobility, group-level mobility

JEL classification: J15, J62

Acknowledgements: I thank Anirudh Krishna for extensive editorial suggestions, and Neil Cummins for allowing me to use the Families of England database.

* University of California, Davis, USA, and LSE, London, UK; gclark@ucdavis.edu.

This study has been prepared within the UNU-WIDER project [Social mobility in the Global South—concepts, measures, and determinants](#).

Copyright © UNU-WIDER 2020

Information and requests: publications@wider.unu.edu

ISSN 1798-7237 ISBN 978-92-9256-785-9

<https://doi.org/10.35188/UNU-WIDER/2020/785-9>

Typescript prepared by Luke Finley.

The United Nations University World Institute for Development Economics Research provides economic analysis and policy advice with the aim of promoting sustainable and equitable development. The Institute began operations in 1985 in Helsinki, Finland, as the first research and training centre of the United Nations University. Today it is a unique blend of think tank, research institute, and UN agency—providing a range of services from policy advice to governments as well as freely available original research.

The Institute is funded through income from an endowment fund with additional contributions to its work programme from Finland, Sweden, and the United Kingdom as well as earmarked contributions for specific projects from a variety of donors.

Katajanokanlaituri 6 B, 00160 Helsinki, Finland

The views expressed in this paper are those of the author(s), and do not necessarily reflect the views of the Institute or the United Nations University, nor the programme/project donors.

1 Introduction

Social mobility rates, as conventionally measured, require considerable amounts of information. We must link parents and children across generations, and link both parents and children to a common social or economic status scale. Such data are readily available in areas such as modern Nordic countries, where every individual is assigned by the government a unique identifier at birth and this identifier is used to track individuals in education, employment, medical care, and taxation.

But if we want to measure social mobility rates in less-developed societies, such as India since independence, or in any society before the last 50 years, we immediately run into data problems using conventional techniques. For the nineteenth and early twentieth centuries it is possible to link families using successive censuses, as for England 1841–1911 and the USA 1850–1940. But the linkage of individual parents and children through censuses, where spelling of surnames and first names is highly idiosyncratic, is a difficult and time-consuming process. Here there has been vigorous debate about the accuracy of matching algorithms, with claims that many parent–child matches are mistaken and also that matches are more likely when both parent and child are of higher social status, resulting in an overestimation of persistence in status across generations (see Ruggles et al. 2018).

We shall also see below that there are reasons to question whether the conventional estimates of social mobility, focusing just on parent and child, reveal its true rate for more generalized measures of status.

Another problem is that markers of social status can vary significantly across societies and time periods in how well they indicate underlying social status. In the nineteenth century in the USA, vast numbers of men were described as ‘farmers’. But farmers varied enormously in social status, from smallholders with a few rented acres to large-scale operators with many hired labourers.

The reported correlation of occupational status in nineteenth-century USA is around 0.3 between fathers and sons, implying high rates of social mobility. For England that correlation at the same time is around 0.45 (see Table 1). Is that because nineteenth-century England was a much less mobile society, or because as a much more urbanized and industrialized society it had occupational titles much more revealing of true social status?

What is proposed in this paper is another way of measuring social mobility in early or less institutionally developed societies, which uses the status of surnames. This measure has several advantages for measuring mobility in such societies. First, it can be done without having to link individuals across generations, so it is informationally less demanding. Without having to link individuals across generations this method can be employed using information on status from censuses, voter rolls, and probate records.

Second, this method is not affected by the degree of errors and noise in status measures across different societies. It will work just as well for relatively imprecise measures of status as for much more finely calibrated measures.

Further, it is possible to use surnames to estimate intergenerational mobility rates even when we have just three pieces of information:

1. the general frequency of surnames or surname types;

2. the frequency of these surname types among some elites or underclasses—university students, doctors, property holders, or convicted criminals, for example;
3. a measure of how elite or how disadvantaged the high-status or low-status group is.

The paper shows how to estimate intergenerational mobility rates using surnames. It discusses how to interpret these results compared with conventional estimates. Finally, it also outlines the elements that can frustrate such estimates.

2 Measuring social mobility rates in general

We assume social status can be measured by a cardinal number y which measures some aspect of social status such as income, wealth, occupational status, longevity, or height. Conventionally, social mobility rates have been estimated by economists from the estimated value of β in the equation

$$y_t = \alpha + \beta y_{t-1} + u_t \tag{1}$$

where y is the measure of social status, t indexes the generation, and u_t is a random shock. β will typically lie between 0 and 1, with lower values of β implying more social mobility. β is thus the persistence rate for status, and $1 - \beta$ the social mobility rate. If the variance of status on this measure is constant across generations then β is also the intergenerational correlation of status. And in this case β also estimates the share of the variance of status in each generation that is explicable from inheritance. This share then will be β^2 . The reason for this is that if σ^2 measures the variance of the status measure y , and σ_u^2 measures the variance of the random component in status, then, from Equation 1:

$$\begin{aligned} \text{var}(y_t) &= \beta^2 \text{var}(y_{t-1}) + \text{var}(u_t) \\ \sigma^2 &= \beta^2 \sigma^2 + \sigma_u^2 \end{aligned}$$

If Equation 1 is the correct description of the inheritance of social status in any society, then in steady state any measure of status such as the logarithm of income or wealth will show a normal distribution.

Equation 1 involves a number of strong simplifying assumptions. It assumes, for example, that social mobility rates are the same across the whole of the status distribution, from top to bottom. But we shall see that the empirical evidence is that this assumption is not too far from reality.

3 Measuring mobility rates from surnames

For the reason above, we have until recently had no idea of what social mobility rates were in pre-industrial societies. We have had no idea whether, for example, the Industrial Revolution in England was associated with a period of enhanced social mobility compared with what came before and what came after.¹

¹ See Clark and Cummins (2014a) for a review of the evidence on this.

However, in many societies people have surnames, and these surnames are inherited unchanged through the patriline. Men bearing the surname Boscawen born in England in 1900–30, for example, are descended from someone in the group of men bearing the surname Boscawen in 1870–1900. Thus, using surnames to group people, we can identify groups of sons who collectively are descended from a group of fathers, without knowing the exact descent relationships. The fact that surnames can proxy for the transmission of the Y chromosome between generations has long been of interest to geneticists (see, for example, King and Jobling 2009). However, only recently have there been attempts to utilize surnames to estimate social mobility rates.² Here we describe two methods of estimating intergenerational social mobility from surnames.

Instead of estimating β from

$$y_t = \alpha + \beta y_{t-1} + u_t \tag{1}$$

we can use

$$\bar{y}_{kt} = \alpha + \beta \bar{y}_{kt-1} + \bar{u}_{kt} \tag{2}$$

where k indexes surname groups and $\bar{}$ indicates averages. We can, for example, compare the average status of everyone born with the surname Boscawen in 1800–29 with that of those born with this surname in 1830–59, the 30-year interval between the time periods here representing the assumed average length of a generation.

This averaging across surnames would be expected to produce an attenuated estimate of the β linking fathers and sons for several reasons. First, we have to take all those born with a class of surnames in a time interval $(t, t + n)$ and compare them to those born in the time interval $(t + 30, t + n + 30)$, the 30 years representing the average interval between generations. This introduces error in that some children of the generation born in the interval $(t, t + n)$ will not be born in the interval $(t + 30, t + n + 30)$. And some of those born in the interval $(t + 30, t + n + 30)$ will have fathers not born in $(t, t + n)$. Second, the surname method counts those in $(t, t + n)$ who have no children equally with those who have large numbers of children. Third, the surname method includes wives of men bearing the surnames, who adopted those surnames on marriage. Fourth, there will potentially be some adopted children among the younger generation, as well as those who changed surnames from their birth surname. For all these reasons the surnames can only provide an imperfect estimate of the average of the actual parent–child status linkages. This imperfection should bias the surname estimates towards 0.

However, in practice these surname estimates of β are always much greater than the β estimated from individual family linkages. To take an example, I have assembled (along with Neil Cummins) a large genealogical database for families in England of 371,000 individuals born in the period 1750–2019, based around families with rarer surnames. Such surnames make it much easier to link individuals across generations. For these families we have multiple measures of social status. Looking just at men, these include wealth at death (for deaths 1858 and later), attainment of higher education (university or equivalent), and occupational rank at the age of 40. Table 1 shows the intergenerational correlation of status on these measures for the generations born in 1840–69,

² Weyl (1989) used surnames to identify social groups, and to measure their relative status in the modern US, but did not attempt to measure rates of regression to the mean.

1870–99, and 1900–29 compared with their fathers, where the lineages included are those which had average status in the nineteenth century. These correlations average around 0.4.

Table 1: Conventio intergenerational mobility estimates, England, births 1840–1929

| Birth period of sons | Ln wealth at death | Higher education | Occupational rank |
|----------------------|--------------------|------------------|-------------------|
| 1840–69 | 0.403 (0.020) | 0.458 (0.015) | 0.529 (0.015) |
| 1870–99 | 0.311 (0.018) | 0.353 (0.014) | 0.446 (0.013) |
| 1900–29 | 0.247 (0.022) | 0.246 (0.020) | 0.415 (0.019) |
| All | 0.352 (0.012) | 0.358 (0.009) | 0.465 (0.009) |

Note: standard errors in parentheses.

Source: Families of England database.

Table 2: Difference in status between elite and average surnames, men

| Birth period | Ln wealth at death | Higher education | Occupational rank |
|--------------|--------------------|------------------|-------------------|
| 1810–39 | 3.628 (0.102) | 0.328 (0.011) | 0.318 (0.007) |
| 1840–69 | 2.625 (0.079) | 0.250 (0.008) | 0.264 (0.005) |
| 1870–99 | 1.604 (0.064) | 0.166 (0.007) | 0.179 (0.005) |
| 1900–29 | 1.125 (0.069) | 0.146 (0.009) | 0.147 (0.006) |

Note: standard errors in parentheses.

Source: Families of England database.

Table 3: Intergenerational correlations of status revealed by surnames

| Birth period of sons | Ln wealth at death | Higher education | Occupational rank |
|----------------------|--------------------|------------------|-------------------|
| 1840–69 | 0.724 (0.038) | 0.762 (0.037) | 0.831 (0.025) |
| 1870–99 | 0.611 (0.038) | 0.664 (0.044) | 0.677 (0.027) |
| 1900–29 | 0.701 (0.053) | 0.877 (0.061) | 0.819 (0.036) |
| All | 0.677 (0.021) | 0.763 (0.032) | 0.772 (0.021) |

Note: standard errors in parentheses.

Source: Families of England database.

With the same database, we can implement an estimate of social mobility through surnames by instead looking at the average social status of all men with high-status surnames (measured by average wealth at death by surname for 1858–87) relative to men with average-status surnames.

Table 2 shows by period the difference in wealth, education, and occupational status between the high-status surnames and average surnames. What is surprising is how slowly the status of the elite surnames, on all dimensions, are regressing towards 0. Taking just the ratio status from one period to the next, we can derive an implied correlation of status across generations, as shown in Table 3.

As can be seen, these estimates in Table 3 show much greater persistence of status than the estimates for the individual father–son combinations, with correlations that average 0.74 as opposed to 0.4.

Why are these results so different? The reason is that social mobility in any society seems to be described by a process that is more complicated than would be suggested by Equation 1 above. At the family level, observed status y seems to be composed of both an underlying individual family status x and also a substantial transitory component u where

$$y_t = x_t + u_t \tag{3}$$

The underlying status is inherited strongly, so that

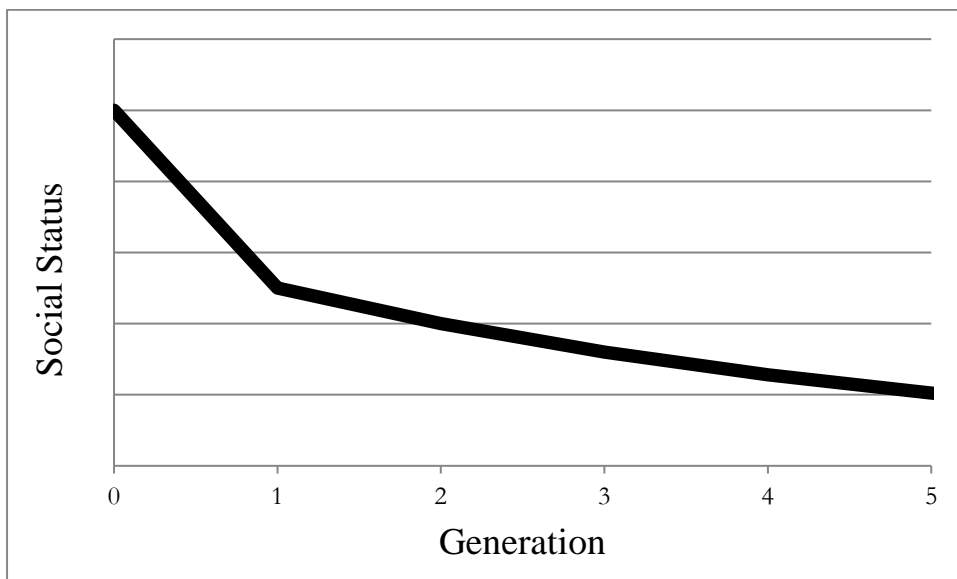
$$x_t = bx_{t-1} + e_t \tag{4}$$

where b is in the order of 0.7–0.8. In this case if we regress, as is conventionally done,

$$y_t = \beta_1 y_{t-1} + v_t \tag{5}$$

so that we are looking at the parent–child correlation, then $E(\hat{\beta}_1) = \frac{\sigma_x^2}{\sigma_y^2} b$. But if we look over n generations, where β_n is the correlation across n generations, then $E(\hat{\beta}_n) = \frac{\sigma_x^2}{\sigma_y^2} b^n$. Thus, if we observe someone with above mean status in period 0, as in Figure 1, the typical path of their descendants towards mean status across n generations will be one of fast regression to the mean in the first generation, followed by a much slower, constant regression in each of the subsequent generations.

Figure 1: Typical path of regression to the mean for an individual family



Source: author's illustration based on discussion above.

The transitory component in social status exists for two reasons. First, all measures of status are made with substantial amounts of error. That creates an appearance just of enhanced mobility across single generations. Second, there is an element of luck in the actual status attained by individuals.

This means that social mobility has two components, both of which are needed to describe the full process. There is the short-run parent–child mobility, the rate of which can vary substantially across aspects of status such as wealth, education, income, occupational status, and longevity. Then there is the underlying long-run persistence, which may be the same across all aspects of status.

If we have independent information on which of a set of surnames have on average high or low social status, then the intergenerational correlation of status observed for surnames will reveal the underlying long-run persistence rates within a society.

The surname studies identify this underlying persistence rate, because by averaging across people by surname we reduce the transitory component in Equation 3 above to 0, so that for each generation we now observe for each surname, or surname group or type, the underlying \bar{x} .

The underlying long-run persistence rate is what matters if we are looking at group social mobility rates within a society. For example, suppose we want to know how long it will take some elite or underclass group to come to average social status; b will give an indication of this. Note that the b estimated above for richer surnames in England is estimated in a society where most of the surname holders were white and had similar religious affiliations, and in a society without explicit barriers to intermarriage between social groups. That persistence rate, at 0.7–0.75, is still very strong. It implies that the holders of the wealthy surnames identified in Table 2 will only converge to within 10 per cent of average wealth at death after another seven generations from those born in 1900–29, or for those born in 2110–49 (assuming a persistence parameter of 0.7).

If we go to a society such as India, where there are strong barriers to intermarriage between such social groups as Hindus and Muslims, or between low-caste and high-caste Hindus, then we observe even lower rates of surname status mobility across generations when we look at surnames associated with specific social groups such as Muslims or Brahmins. There, in recent generations, despite the substantial system of reservations in higher education and in government employment, the persistence rate is more in the order of 0.9.

The underlying persistence rate can also be used to estimate the effects of educational and other reforms on the convergence of group social status.

4 Estimating long-run mobility rates from surnames, with direct measures of surname status

Where we have direct measures of social status by surname, implementing the estimation of the underlying social mobility rate b , as seen in Equation 4 above, is straightforward. We need only identify groups of surnames that are preselected as having high or low status and then examine what happens to the average status of these surnames over time. We need to make an assumption about what generation lengths are to get the intergenerational correlation b . But with surname averages it is possible to also estimate intergenerational correlations using periods shorter than a generation length, such as a decade. It is just a matter of the size of the data set.

One such source of status by surname and vintage even in poorer societies is electoral registers. These are often public documents that list, for the voters of a polity, their age and some measure of their social status, such as their address or their occupation. Thus the 2004 electoral register for Chile records, for 6 million voters, their name, age, location, and occupation. This allows people to be assigned a measured status in two ways.³ The first is based on the average earnings of their occupation. The second is based on the average earnings of people living in their municipality. This then allows estimates of average social status by surname type for those born between 1920 and 1979—two complete generations.⁴ Below is laid out, using the Chilean data, exactly how the procedure is implemented.

Similarly, the electoral registers for the UK for 2003–10 are publicly available. These give peoples' ages to within three years, as well as their exact address.⁵ Online measures are available of average house values specific to around 1.1 million postal codes in the UK—average values that vary between £40,000 and £59,000,000. Measures of social deprivation by postcode are also available online from the UK government, giving area averages of such measures as income, education, health, and crime rates (Ministry of Housing, Communities and Local Government 2019).

The entire electoral register for West Bengal in India is available online (Chief Electoral Officer, West Bengal 2019). This lists voters by age and by street address. As long as we can assign average social status to these addresses, we can again estimate social mobility rates by looking at the rate of convergence of surname status to the mean as we go from older to younger voters.

Even where electoral registers do not give voter ages, we can again measure the rate of social mobility as long as older registers are available and there is some indicator of voter status. Thus, the electoral registers from Australia for 1903–80, which give voter occupations, are available. We can then track the average status of high- or low-status surnames across multiple generations.⁶

In the case of Chile, to identify elite and underclass groups of surnames for the period 1920–49 we can use two procedures. First, surnames in an immigrant society like Chile can be classified by ethnic and national origin. Thus, there is a class of surnames associated with the Mapuche, the main surviving indigenous population of Chile (Galdames et al. 2008). There are also distinctive surnames associated with immigrant groups of Basque, German, French, and Italian origin. Basque settlers, for example, were an early elite in colonial Chile. In the nineteenth century, Chile attempted to recruit educated northern European immigrants, so modern-day Chileans with, for example, German surnames are the descendants of a nineteenth-century elite.

But, further, we can identify, as in the case of England, rare surnames associated with earlier wealth in Chile in the nineteenth and early twentieth centuries, including the wealthy from all ethnic groups. An agricultural rent report was compiled, for example, in 1853 to determine agricultural taxes. The average rental value of a parcel of land in the 1853 report was 379 pesos. We can thus classify holders of land parcels of rental value greater than 1,500 pesos as wealthy in 1853. From the list of surnames that show up among wealthy landowners in 1853, we selected those surnames that appeared less than 30 times in contemporary Chilean population censuses. Rarer surnames

³ For details on the sources for Chile see Clark et al. (2014: 199–211).

⁴ Since people only have occupational statuses once they complete school, this measure can only be computed for those aged 25 and above.

⁵ There is one drawback of the UK data, which is that people had to agree to the data being made public. About half of the electorate is covered by the public register.

⁶ Similarly, electoral censuses in Canada, 1935–80, and New Zealand, 1920–81, give occupations for voters.

were used since these are the ones that when found among landowners will on average have high status. There is a second list of large landholders in 1920, from which again we can select rare surnames.

Table 4 shows the numbers of people from each of four such surname groups listed with an occupation in the 2004 electoral register born in 1920–49 and 1950–79.⁷ For the country as a whole there are 2.3 times as many people recorded with an occupation in 1950–79 as earlier. But interestingly, for the low-status group, the Mapuche, the ratio in 1950–79 compared with 1920–49 is greater than average, at 2.47. For the high-status groups the ratio of 1950–79 to 1920–49 is lower than average.

The table also shows the average log occupational earnings of each group, relative to the average for all electors. Logarithms are used here, since occupational earnings are positively skewed. Thus Columns 5 and 6 show, for birth cohorts 1920–49 and 1950–79,

$$\frac{1}{N_{ik}} \sum_i \ln w_{ik} - \frac{1}{N} \sum_i \ln w_i \quad (6)$$

where $\ln w_i$ is the log occupational earnings for each elector, N is the total number of electors with occupations, N_{ik} is the number of electors with occupations in surname group k , and $\ln w_{ik}$ is the log occupational earnings of each member of group k .

Table 4: Estimated Chilean social mobility rates, births 1920–79

| Surname group | N 1920–49 | N 1950–79 | Ratio N | Ave. occupational earnings, 1920– 49 | Ave. occupational earnings, 1950– 79 | Implied b |
|------------------------------|----------------|----------------|-----------|-----------------------------------------------|-----------------------------------------------|-------------|
| Mapuche | 7,036 | 17,389 | 2.47 | -0.304 | -0.239 | 0.79 |
| Basque | 8,755 | 17,841 | 2.04 | 0.225 | 0.169 | 0.75 |
| Large landowners, 1853 | 2,731 | 5,201 | 1.90 | 0.396 | 0.371 | 0.94 |
| Large landowners, 1920 | 1,680 | 3,069 | 1.83 | 0.450 | 0.415 | 0.92 |
| All | 895,145 | 2,059,057 | 2.30 | 0.000 | 0.000 | - |

Note: the numbers reported in each period are of those whom the electoral register lists with an occupation.

Source: Clark et al. (2015): table 2.

For those with Mapuche surnames born in 1920–49, the value of -0.304 implies that their average occupational earnings are only 74 per cent of the overall average for this birth cohort. For those with the rare surnames of large landowners in 1920, the value of 0.450 for the 1920–49 birth cohort implies that their average occupational earnings are 57 per cent higher than the overall average for this birth cohort.

⁷ Most males of working age had listed occupations, so there is no reason to think that omitted occupations will bias the results.

The b estimate in the final column comes from the equation

$$\overline{\ln w_{k1}} = b \overline{\ln w_{k0}} \quad (7)$$

where the subscript 1 indicates the generation born in 1950–79 and the subscript 0 the generation born in 1920–49. As can be seen, these estimates suggest strong persistence of occupational status for both the high-status and the low-status groups. The estimates of b range from 0.75 to 0.94.

5 Estimating social mobility rates from surnames with even less information

A nice feature of using surnames to estimate social mobility rates is that we can derive the long-run underlying correlation of status with even less information than is used in the first example above. In particular, suppose that all we know about a society is the general distribution of surnames in the population, the distribution of surnames among an elite, and what percentage of the population that elite represents. We can still get a good estimate of social mobility rates.

In England, for example, we know the general distribution of surnames in 1538 and later.⁸ We know the distribution of surnames at Oxford and Cambridge universities, the only universities in England until 1836 and always the most prestigious universities, from 1200 onwards. And we also know what share of males attended these two elite universities from at least 1500 onwards. This allows an estimate of the persistence of educational status in England from 1500 to 2015.

Suppose, for example, the variance of status in an elite or underclass set of surnames can be assumed to be the same as that for the population as a whole. Then the situation is as in Figure 2. For names in general we will find that about 1 per cent are at Oxford or Cambridge. But for more elite surnames a higher fraction will be present at the university. Thus, for each period after 1500 we can estimate for each surname its relative status, the measure being:

$$\text{Relative representation} = \frac{\text{Share of surname } z \text{ at Oxbridge}}{\text{Share of surname } z \text{ in Oxbridge age cohort}} = RR_z$$

That is, we take the ratio of the share of people at Oxford or Cambridge with a given surname, compared with the share in the population as a whole aged 18–22 who have that surname. By definition, for the average surname in England in any period this number will be 1. But for high-status surnames the number will exceed 1 and for low-status surnames it will fall below 1.

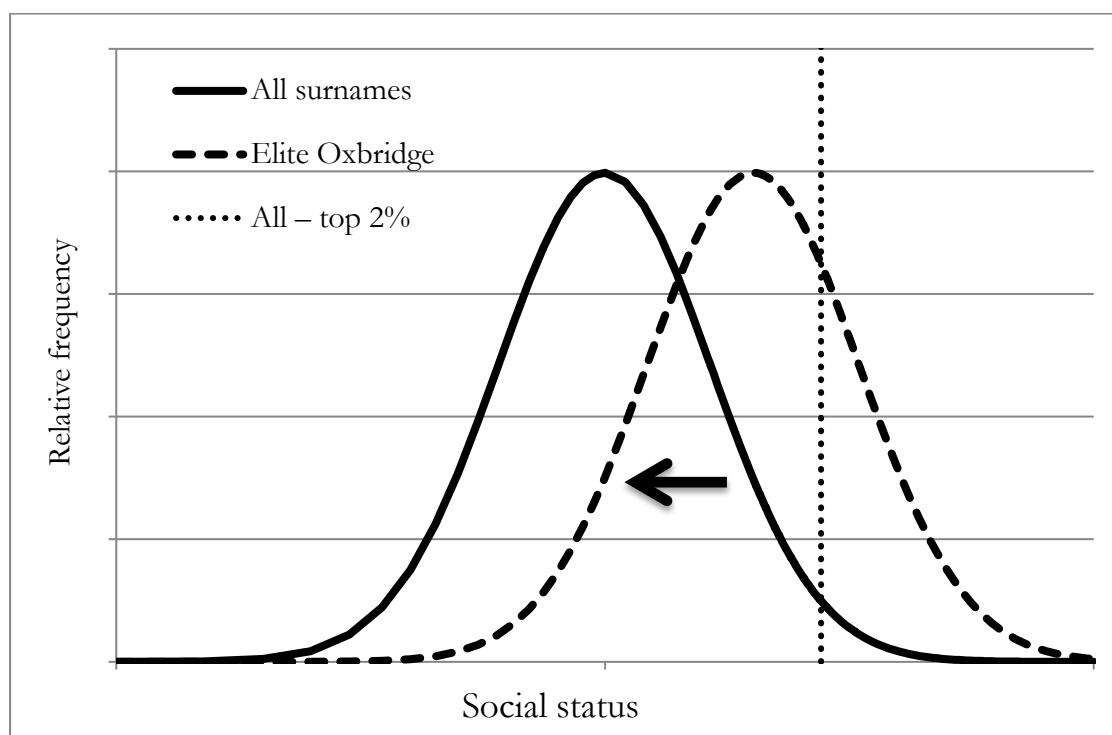
From the relative representation estimate for each surname we can derive an implied mean status of each surname, measured in standard deviation units.⁹ If, for example, the Oxbridge elite represents the top 1 per cent in educational status, and a surname is ten times more common among Oxbridge students than in the population as a whole, then its implied average status is 1.04 standard deviations above the mean. If the relative representation is 30 then the implied average status is 1.80 standard deviations above the mean. Then the implied intergeneration correlation of

⁸ There are extensive records of baptisms and marriages from parish records in England for 1538–1837, and then national registers of births, deaths, and marriages for 1837–2019.

⁹ The key assumption here is that the variance of status among holders of each surname is the same as the variance of status for society as a whole. We consider below how reasonable this assumption is.

status for these surnames across a generation (assumed, as before, to be 30 years) will just be the estimated mean status in generation $t + 1$ divided by that in generation t .

Figure 2: Regression to the mean of elite surnames



Source: Clark et al. (2014): figure 18.1.

To see how this works in practice, let us construct a set of surnames that was elite in England in terms of educational status in 1800–29. To do this we simply include all English surnames where less than 500 people held the surname in the census of 1881, but someone with that surname attended Oxford or Cambridge in 1800–29.¹⁰ This generates 2,354 individual surnames held by Oxbridge students in these years. These surnames were held by 277,247 people in 1881, and by 473,595 people in 2002. To estimate the population share with these rare surnames in each student cohort we use records of marriages in England for 1837–1915, and records of births for 1916–95. The share of the population with this sample of rare surnames in each generation of students, again taking a generation as 30 years, is shown in the second column of Table 5. This share was around 1.16 per cent of the population in 1800–29, but had fallen to 0.85 per cent by 2010–13. This reflects the substantial migration of people from Ireland and Scotland into England in the period 1800–1950, and then migrations from Europe and elsewhere into England in 1950 and later.

Table 5 shows the numbers of students with these surnames at Oxbridge in each 30-year period starting in 1800, as well as the total numbers of students observed in each period.¹¹ Column 5 shows the share of these surnames as a share of all Oxbridge students with English surnames. As can be seen, in 1800–29 these surnames represented more than 21 per cent of students despite being held by an estimated 1.2 per cent of the population. The last column shows the relative

¹⁰ We use the 1881 census to find the rarer surnames because this is one of the most carefully digitized nineteenth-century censuses.

¹¹ If a surname occurred multiple times that was counted.

representation of these surnames at Oxbridge from 1800 to 2013 by period. There is a steady decline in that relative representation across generations, though it is still around 1.5 in 2010–13.¹²

Table 5: Rare surnames at Oxbridge, 1800–29

| Generation | Share population rare Oxbridge surnames % | Rare surnames 1800–29 at Oxbridge | All Oxbridge attendees | Share rare surnames Oxbridge % | Relative representation |
|------------|-------------------------------------------|-----------------------------------|------------------------|--------------------------------|-------------------------|
| 1800–29 | 1.16 | 3,991 | 18,650 | 21.57 | 18.57 |
| 1830–59 | 1.16 | 2,856 | 24,415 | 11.82 | 10.17 |
| 1860–89 | 1.13 | 2,951 | 38,678 | 7.84 | 6.93 |
| 1890–1919 | 1.09 | 1,477 | 30,961 | 5.02 | 4.61 |
| 1920–49 | 1.04 | 1,917 | 67,927 | 3.08 | 2.96 |
| 1950–79 | 0.99 | 2,628 | 156,645 | 1.86 | 1.87 |
| 1980–2009 | 0.85 | 2,383 | 222,063 | 1.32 | 1.55 |
| 2010–13 | 0.85 | 437 | 49,243 | 1.28 | 1.51 |

Source: Clark and Cummins (2014b): table 3.

What is the persistence rate of educational status implied by the last column of Table 5? To calculate that, we translate the relative representation of the rare surnames into an implied deviation of mean educational status for this group from the social mean as in Figure 2. To do this we need to estimate for each generation what percentage of the population attend Oxford or Cambridge, to establish how elite this set of students is.

Column 3 of Table 6 shows this estimate. It is calculated, for example, that in 1830–59 only 0.6 per cent of each generation (of males in this case) attended Oxford or Cambridge. By 2010–13 the estimated share of the population cohort attending Oxford or Cambridge had risen to 1.24 per cent. This then yields the estimate in Column 4 of what the average educational status of the elite surnames was for each generation, measured as standard deviation units above the social mean. For 1830–59 the estimated mean deviation is 0.97 SD units, while by 2010–13 that had fallen to 0.16 SD units.

Note that in Table 6 we use only 1830–59 and later to estimate the intergenerational correlation of status. We do so because, as is portrayed in Figure 1, the regression towards the mean of surname status in the first generation will be faster than in later generations, because in that generation we observe both the more rapid short-run mobility and the slower underlying mobility. In later generations, where we have a pre-established set of elite surnames, the numbers of surname holders at Oxford and Cambridge will give an unbiased estimate of the average educational status of the target surnames, and of underlying social mobility rates.

¹² To calculate the relative representation in the periods after 1829, an allowance has to be made for the increasing share of foreign students at Oxbridge. The England and Wales surname share of students is calculated from 1830 on by period as 0.99, 0.97, 0.95, 0.92, 0.90, 0.82, and 0.69.

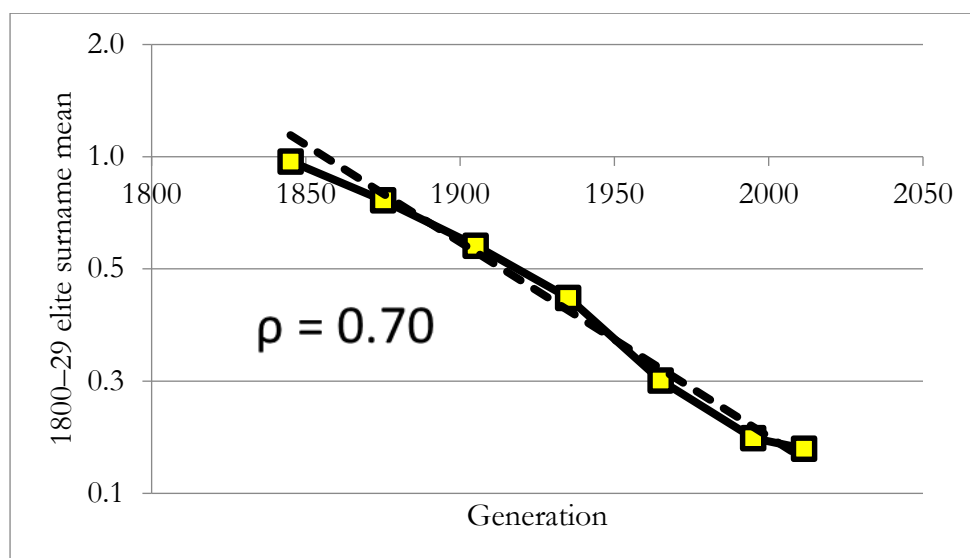
Table 6: Implied persistence rates for 1800–29 elite rare surnames

| Generation | Relative representation | Oxbridge elite share % | Implied mean status (standard deviation units) | Implied b |
|------------|-------------------------|------------------------|------------------------------------------------|-------------|
| 1830–59 | 10.17 | 0.62 | 0.97 | - |
| 1860–89 | 6.93 | 0.53 | 0.76 | 0.79 |
| 1890–1919 | 4.61 | 0.48 | 0.58 | 0.76 |
| 1920–49 | 2.96 | 0.70 | 0.42 | 0.72 |
| 1950–79 | 1.87 | 1.16 | 0.25 | 0.60 |
| 1980–2009 | 1.55 | 1.27 | 0.18 | 0.70 |
| 2010–13 | 1.51 | 1.24 | 0.16 | 0.89 |
| All | | | | 0.73 |

Source: Clark and Cummins (2014b): table 3.

These estimated mean statuses by generation are graphed in Figure 3, where the vertical axis is graphed as a log scale.

Figure 3: Mean status, rare elite surnames, Oxbridge, 1830–2013



Source: Table 6.

Once we know the implied mean of status for the 1800–29 elite rare surname group for 1830–2013, we can then calculate for each period the implied correlation of status b with the previous generation. From Equations 3 and 4, and assuming with averaging that $\bar{y}_t = \bar{x}_t$ —that is, that the average measured educational status of the surnames is the average actual status,

$$\bar{y}_{t+1} = b\bar{y}_t + \epsilon_{t+1} \quad (8)$$

where ϵ_{t+1} is an error term corresponding to various mismeasurements. These are errors in the measuring of the share of the surname population in each cohort, the share of these names at Oxbridge (in some periods we have just a sample of Oxbridge students, not the population), the

share of the domestic population among Oxbridge students, and the degree of eliteness that Oxbridge attendance implies.

The unbiased estimated value of b for each period is then

$$\frac{\bar{y}_{t+1}}{\bar{y}_t} \tag{9}$$

These estimates are shown in the final column of Table 6. The average is 0.74, though the individual b estimates range from 0.60 to 0.99.

Suppose we assume, however, that this variation is just the product of the aforementioned measurement errors, and fit one b value to the whole of the data. To do this, note that Equation 8 implies

$$\bar{y}_{t+n} = b^n \bar{y}_t + \epsilon_{t+n}^* \tag{10}$$

or

$$\ln \bar{y}_{t+n} = \ln \bar{y}_t + \ln(b) \cdot n + \ln \epsilon_{t+n}^* \tag{11}$$

So just by estimating the coefficient h in the OLS best-fitting relationship

$$\ln \bar{y}_{t+n} = g + h \cdot n$$

we can estimate the best-fitting b for the whole set of observations, assuming that this has a constant value. The b estimated in this way is 0.70, with 5 per cent confidence bounds of (0.67, 0.72). As Figure 3 shows, the R^2 of this fit is good, being 0.988.

Three things stand out in this estimate. First, there is a high degree of persistence of status implied in the estimates. Second the estimated persistence here is very similar to that found for wealth, education, and occupation in the high-status surnames in the lineage data set discussed above. But third is the seeming constancy of this strong persistence over generations.

Over the course of the generations entering college in 1830–2013, the social and institutional circumstances of England changed considerably. England came late to the idea of state support for education. Until late in the nineteenth century, education was largely organized through an ad hoc system of charity schools, religious schools, and local private provision. Thus, only with the Forster Act of 1870 was there any requirement of school attendance. And this requirement was only for the ages of 5–10, with exemptions for children who were sick, working, or living too far from a school. Also, before 1870 central government support for education was minimal. Not until 1833 did the central government direct any monies in support of constructing or maintaining schools.

Over the years 1880–1918 there were a series of parliamentary acts that expanded significantly both educational requirements and state support for education: the Elementary Education Act 1880, the Elementary Education (School Attendance) Act 1893, the Conservative Education Act 1902 (Balfour Act), and the Fisher Education Act 1918. Through these acts, required school attendance was extended eventually to the ages of 5–14. School fees were also abolished for all children in publicly supported schools.

The Education Act of 1944 extended compulsory schooling to the age of 15. It also codified a tripartite system of education. At 11, students were assigned, based on an exam, either to elite grammar schools or to more vocational secondary modern or technical schools.

Thus, we see families experience substantially different social and institutional regimes with respect to education across the course of their histories. Those cohorts born 1780–1869, and entering college 1800–90, mostly existed in the laissez-faire era, where there were no schooling requirements and there was only private and religious support for education for the poor. The cohort born 1870–99, and entering college 1890–1920, experienced the modest beginnings of the modern welfare state in education. Compulsory education was imposed for the first time, and state support to parents extended. Finally, the cohort born 1900–29, and entering college 1920–50, experienced for the first time substantial state-imposed educational requirements, with most children born in this cohort required to attend school to the age of 14 and with public funding of the costs of schooling. But, remarkably, the extension of state support for schooling seemingly had no impact on long-run mobility rates.

If we take the 2,354 rarer surnames which appear in the rolls of students at Oxford and Cambridge in 1800–29 then we can also look at how this group did on other measures of social status in England in 1830–2019. One of these is the political elite. We know the names of all of the 460–533 members of parliament (MPs) from England and Wales for every year in this interval. Table 7 shows the total numbers of MPs entering parliament by 30-year period from 1800 on, where we count each member just once, by their year of first entry to parliament. This is a much smaller group than that of students enrolling at Oxford and Cambridge, so the results are noisier. But as Columns 3 and 4 show, the rare surnames enrolling at Oxford and Cambridge in 1800–29 were also heavily over-represented among MPs, and continued to be over-represented even in the period 2010–19.

Table 7: Social status as measured by MPs, 1800–2019

| Generation | All MPs | Rare surname MPs | Share rare surnames MPs (%) | Relative representation among MPs |
|------------|---------|---------------------|--------------------------------------|-----------------------------------------|
| 1800–29 | 2,064 | 396 | 19.2 | 16.51 |
| 1830–59 | 2,473 | 417 | 16.9 | 14.51 |
| 1860–89 | 1,848 | 225 | 12.2 | 10.76 |
| 1890–1919 | 1,779 | 122 | 6.9 | 6.30 |
| 1920–49 | 1,914 | 75 | 3.9 | 3.77 |
| 1950–79 | 1,421 | 47 | 3.3 | 3.33 |
| 1980–2009 | 1,107 | 32 | 2.9 | 3.41 |
| 2010–19 | 488 | 9 | 1.8 | 2.18 |

Source: Clark et al. (2014): 102–05, with data updated from 2013 to 2019.

Table 8 shows the calculated persistence rates of these surnames among the political elite under alternative assumptions about how elite a class MPs were and are. We start with the generation of politicians entering parliament in 1830–59, since the typical age of entry to parliament would be 20–30 years later than the age of entry to college. The first assumption is that since the number of MPs increased little between 1800 and 2019 yet the population of England increased nearly eightfold, MPs represented an increasingly elite segment of society. Thus, MPs are assumed to now represent the top 0.1 per cent in terms of social status, compared with the top 0.4 per cent in

1830. Column 3 shows the implied average status of the rare Oxbridge surnames in terms of the political elite by generation and Column 4 the implied persistence rate of that status. Figure 4 shows the average status by generation and the fitted overall mobility rate, which is 0.78.

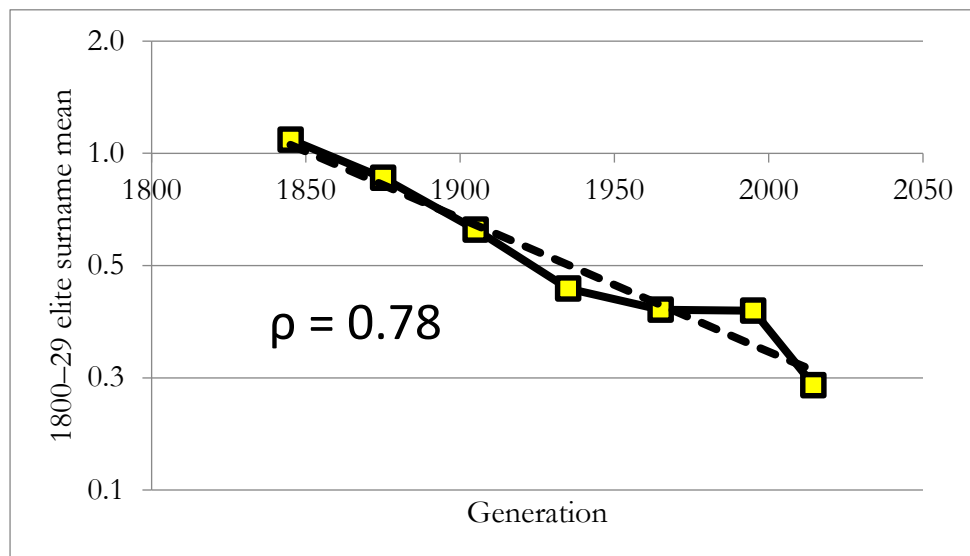
Table 8: Social mobility as measured by MPs, 1830–2019

| Generation | Assumed eliteness MPs (%) | Rare mean status | Implied b | Assumed eliteness MPs (%) | Rare mean status | Implied b |
|------------|---------------------------|------------------|-------------|---------------------------|------------------|-------------|
| 1830–59 | 0.4 | 1.09 | - | 0.1 | 0.96 | - |
| 1860–89 | 0.2 | 0.86 | 0.79 | 0.1 | 0.91 | 0.87 |
| 1890–1919 | 0.2 | 0.63 | 0.73 | 0.1 | 0.79 | 0.75 |
| 1920–49 | 0.1 | 0.43 | 0.69 | 0.1 | 0.60 | 0.70 |
| 1950–79 | 0.1 | 0.38 | 0.88 | 0.1 | 0.42 | 0.90 |
| 1980–2009 | 0.1 | 0.38 | 0.99 | 0.1 | 0.38 | 1.02 |
| 2010–13 | 0.1 | 0.24 | 0.48 | 0.1 | 0.39 | 0.47 |
| Average | | | 0.76 | | | 0.79 |

Source: Table 7.

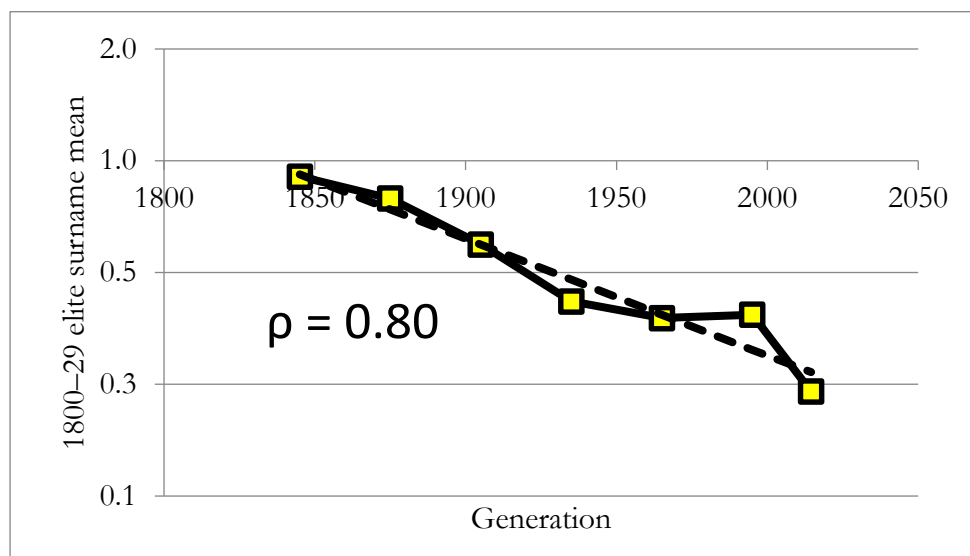
How sensitive is the estimated persistence rate ρ to assumptions about the eliteness of MPs in England? To test this, Table 8 also shows the calculation of ρ if we just assume that MPs represented a constant elite of the top 0.1 per cent of the population throughout the period. As can be seen in Figure 5, the results change very little with this change in assumptions. The estimated ρ changes from 0.78 to 0.80.

Figure 4: Social mobility rates, political elite, 1830–2019



Source: Table 8.

Figure 5: Social mobility rates, political elite, 1830–2019, alternative assumptions



Source: Table 8.

6 Mobile and immobile societies

Clark et al. (2014) apply the methods above to a variety of societies: the USA, England, Sweden, Chile, China, Japan, and India. In all cases and all periods, the rate of long-run social mobility is low, with the implied intergenerational correlation mostly in the range 0.7–0.8. What this means, however, is that surnames which were established more than three hundred years ago will not as a class exhibit much variation in average social status. This is true, for example, for two classes of elite surnames established long ago in England: the surnames of the Norman conquerors of 1066, and surnames of native English created around 1200–1300 that referred to places (Berkeley, Sussex, Rockingham, etc.). The processes of social mobility may be slow, but given enough time it will do its work. In most of these societies there are no persistent social classes.

However, there are societies where we can observe even slower rates of social mobility in surnames, and where there seem to be near-permanent social classes. One of these is India.

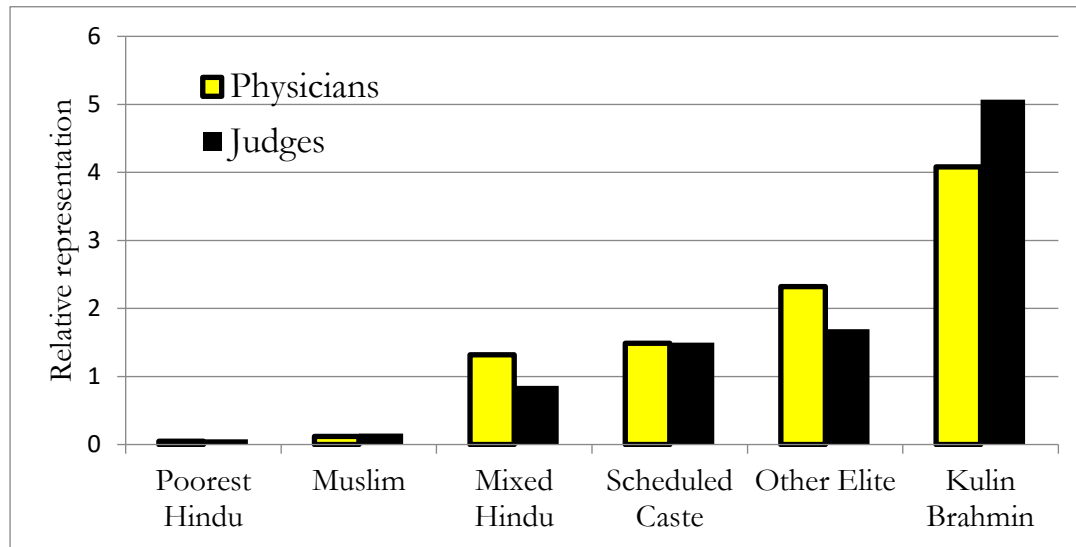
As in many societies, the Indian upper classes were the first to adopt surnames. In Bengal, where the East India Company established its rule in 1757, upper-class Hindus seem to have been already using surnames by the time of the British conquest. For the upper classes in Bengal, family surnames date from the arrival of the British in the eighteenth century or earlier. Petitioners to the East India Company courts in Bengal in the late eighteenth century typically have surnames, and these names are still common in Bengal: Banarji, Basu, Chattarji, Datta, Ghosh, Haldar, Khan, Mandal, Mitra, Sen (Government of Bengal Political Department 1930). Similarly, when the Hindoo College was established in Calcutta in 1817, its initial directors, governors, and secretary, upper-class Hindus, were all men with surnames: Roy, Bahadur, Thakoor, Deb, Sinha, Banerjee, Doss, Mukherjee.

Within Bengali surnames the most elite now are those belonging to the Kulin Brahmin group: Mukherjee, Banerjee, Chatterjee, Ganguly, Bhattacharjee, and Chakrabarti.¹³ Among judges and

¹³ ‘Kulin’ designates a superior Brahmin group.

registered doctors in West Bengal in 2011 these names are four to five times over-represented compared with their population shares, as can be seen in Figure 6. This implies extremely slow rates of social mobility at the group level over the years 1800–2011.

Figure 6: Representation of different surname types in Bengal elites, 2010–13



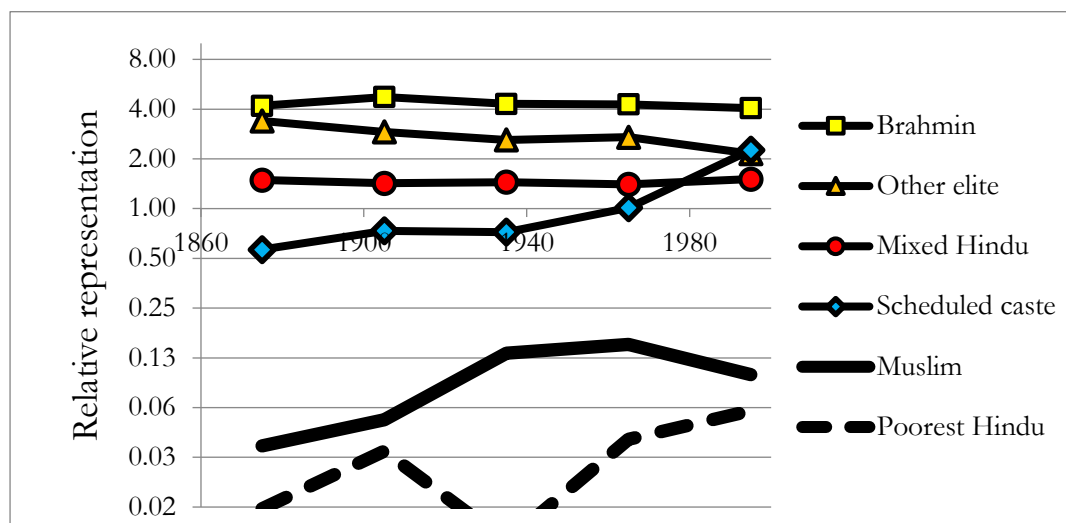
Source: Clark et al. (2014): figure 8.2.

The same figure shows the dramatic under-representation of two other sets of surnames. The first is surnames associated with the Muslim community. These surnames have a relative representation among judges and doctors that is 0.12.

The second set of surnames is those associated with lower-caste Hindu groups that had little or no representation among physicians before independence. The main one is Shaw/Show, held by 3.7 per cent of men on the Kolkata voting rolls. Others are Rauth/Routh, Paswan, Dhanuk, Balmiki, and Mahata/Mahato. Together these surnames are held by 7 per cent of the population of West Bengal. These surnames show a relative representation among elites in 2010–13 that is 0.05–0.08.

Since we can get records of doctors in the Province of Bengal under British rule in 1860–1947 and registered doctors in West Bengal after Indian independence, we can thus estimate social mobility rates by surname types in Bengal as long as we can estimate what the population frequencies of the surname types were over the period 1860–2013. Figure 7 and Table 9 show the relative representation of surname groups in Bengal among doctors for 1860–2011.

Figure 7: Relative representation of surname types among doctors in Bengal, 1860–2011



Source: Clark et al. (2014): figure 8.3.

Table 9: Relative representation of surname types among doctors in Bengal, 1860–2011

| Period | Muslim | Brahmin | Other elite | Poor Hindu | Scheduled caste | Mixed Hindu |
|-----------|--------|---------|-------------|------------|-----------------|-------------|
| 1860–89 | 0.04 | 4.19 | 3.39 | 0.02 | 0.57 | 1.49 |
| 1890–1919 | 0.05 | 4.73 | 2.92 | 0.03 | 0.73 | 1.42 |
| 1920–46 | 0.13 | 4.30 | 2.60 | 0.01 | 0.72 | 1.45 |
| 1947–79 | 0.15 | 4.27 | 2.71 | 0.04 | 1.01 | 1.40 |
| 1980–2011 | 0.10 | 4.05 | 2.15 | 0.06 | 2.26 | 1.51 |

Source: Clark et al. (2014): figure 8.3.

For the Muslim population, their representation is shown relative to the entire population and is always very low. Muslims always constituted a tiny share of doctors compared with their population share.¹⁴ The partition of Bengal in 1947 into largely Hindu West Bengal and mainly Muslim East Pakistan significantly reduced the Muslim population share in West Bengal relative to colonial Bengal. The removal of a large fraction of the population containing very few doctors has the effect of decreasing the relative representation of all the Hindu surname groups among physicians post 1947. Their share of doctors increased little as their population share increased. Since this partition-created decline gives a spurious impression of social mobility, for these other groups, relative representation is shown always with respect to the non-Muslim population only.

Census reports exist giving the Muslim share of the population in Bengal and West Bengal for each decade from 1871 on. Thus, there are good measures of the relative representation among physicians in Bengal from 1860 on. The striking feature is the very low representation of Muslims among physicians in all periods. Under British rule, Muslims experienced limited upward mobility. The implied persistence of status is high, with a calculated intergenerational correlation of 0.91.

However, from the 1970s until very recently, the Muslim community in West Bengal saw a further decline in representation among physicians, with no implied regression to the mean. Indeed,

¹⁴ Because Muslim and Hindu first names are also distinctive, the fraction of Muslim physicians in Bengal in the years 1860–2011 is easily estimated.

starting with the generation entering practice since independence in 1947, the implied persistence coefficient is 1.2, indicating that the Muslim community has been diverging further from the mean.

Bengal's system of reserving educational places and employment opportunities for disadvantaged castes and tribes explicitly excluded Muslims and Christians before 2014: only Hindus, Sikhs, and Buddhists were eligible.¹⁵ Thus, Muslims have been disadvantaged in admission to medical practice in West Bengal, compared with the Hindu, Sikh, and Buddhist populations, since independence. They could compete on equal terms for the unreserved positions in medical schools, but the advantages offered by the reservation system to other disadvantaged groups effectively penalized Muslims. This situation helps to explain the surprising negative social mobility implied for the Muslim community in recent generations. However, even absent the disadvantages imposed by the reservation system, there would be no group-level social mobility among Muslims in the period 1947–2011. Examination of the recent records of applicants to university in Bengal shows that a switch to a pure merit entry system would increase the numbers of people with Muslim surnames by very small amounts. The near absence of social mobility of the Muslim population cannot be attributed to the reservation system.

Even within the Hindu population, there has been very little social mobility among surname groups in Bengal from 1860 on. The Brahmin group of surnames is almost as heavily over-represented among the non-Muslim population in the period 1980–2011 as it was in the period 1860–89. Other elite Hindu surnames show a slow rate of decline in status. But the relative representation of mixed Hindu surnames—those which are held by both the upper castes and also the scheduled castes—does not change.¹⁶ And the relative representation of poor Hindu surnames of the nineteenth century—those with the highest potential for regression to the mean—also changes little. The only group showing a marked change in status is the group of surnames associated with scheduled caste lists for positions in universities and the police. This group went from being modestly disadvantaged among non-Muslim groups in 1860 to being one of the most elite surname groups, as measured by their relative representation among physicians now.

India here seems very distinct from England over the last 150 years. Note that India may well have similar rates of social mobility within such collections of families as the Kulin Brahmin surname group. People could be changing social position within the Brahmin or other social groups at much faster rates than the glacial pace of social mobility we observe for the group as a whole. The methods here are simply comparing Brahmin surnames as a group with those of, for example, the Muslim population.

Why is social mobility, at least at the group level, so low within India? One interesting difference between India and societies such as England is the high degree of group marital endogamy still found in India. As late as the 1960s, caste endogamy still seemed to be the rule for most marriages in Bengal, as seen in a detailed study of a modest-sized town in Bengal in the late 1960s (Corwin 1977). Another study, looking at marriages in rural villages in Karnataka and Uttar Pradesh in 1982–1995, found that of 905 marriages in the study, none involved couples who differed in their caste status (Dalmia and Lawrence 2001). In a high-caste group in Hyderabad, Kayasthas, only 5 per cent of marriages were outside the caste even by 1951–75 (Leonard and Weller 1980: tables 1–3). However, information on the degree of endogamy for marriages in Bengal in the 1970s and 1980s, which produced the most recent crop of physicians, is not readily available.

¹⁵ In 2013 a law was passed reserving 17 per cent of places in state-run universities for 'other backward classes'.

¹⁶ Such surnames include Das, Dasgupta, Majumdar, Ray, Roy, Saha, and Sarkar.

One source of information on the likely endogamy rate is the 2010 Kolkata voter roll, which gives surnames, first names, and ages of all voters. Many first names are highly specific to the Hindu, Muslim, and Christian/Jewish communities. Women who marry into one of these groups from another group will almost always have different first names from women born within the group. Also, if families with surnames associated with one group are assimilated into another group then, as a result of intermarriage and adoption of at least some elements of the culture of the wives, the children will again have different first names.

As Table 10 shows, the percentage of women in the Kulin Brahmin surname group with non-Hindu first names is extremely small. Because Muslims constitute nearly a quarter of the Kolkata population, this implies that intermarriage rates between Kulin Brahmin men and women of Muslim origin are extremely low, in the order of 0.1 per cent. A similar result holds for other high-caste Hindu surnames

Table 10: Female first name origins by surname group

| First-name type | Incidence in surname group (%) | | | |
|---------------------|--------------------------------|------------------------|--------|-----------|
| | Kulin Brahmin | Other high-caste Hindu | Muslim | Christian |
| Muslim | 0.1 | 0.1 | 98.9 | 0.4 |
| Christian | 0.3 | 0.6 | 0.2 | 57.4 |
| Hindu and Christian | 0.0 | 0.0 | 0.0 | 11.9 |

Source: Clark et al. (2014): table 8.5.

More women with Muslim surnames have Hindu first names: 0.9 per cent. But given the near-total absence of any sign of Muslim women's marriage into high-caste Hindu groups, if these findings are indicative of marriage alliances they are likely with lower-caste Hindus.

Intermarriage between Christians and high-caste Hindus appears to be substantially more common. Christian surnames account for a very small share of the surname stock in Kolkata, about 0.3 per cent, and are mainly Portuguese in origin. Given this small Christian population, the small share of women with high-caste surnames who have Christian surnames is nevertheless suggestive of significant intermarriage.

An alternative explanation for these female Christian first names may be that high-caste Hindu girls are given Christian first names at birth. The possibility of significant intermarriage between Christians and Hindus is, however, supported by the fact that just over 30 per cent of women with Christian surnames have first names that are Hindu. Also, almost 12 per cent of women with Christian surnames have a combination of Christian and Hindu first names.

The first name and surname evidence suggests almost no intermarriage between the largely poor Muslim community and either Hindus or Christians. Within the Hindu community, first name evidence does not allow us to determine the degree of marital endogamy within castes because many female first names are common to high- and low-caste groups.

If the marital endogamy of castes and religions in India explains low average social mobility for surname groups, we should find higher rates of social mobility for individual families within these groups. Families sharing the surname Banerjee, for example, will have the same rates of mobility as in any other society. It is just that the average status of the Banerjees will not converge towards that of the Shaws. We should also find that over time, all the major Kulin Brahmin surnames have the same average social status. This hypothesis is borne out by the incidence of these surnames among physicians.

Interestingly, other societies where there is evidence of an absence of normal rates of social mobility for some subgroups within the population also tend to be characterized by high degrees of marital endogamy within such groups. Thus, in Egypt, for example, the Christian Coptic population has remained an elite within a society that is more than 90 per cent Muslim for more than 1,000 years. But Christian and Muslim populations in Egypt show an almost complete absence of intermarriage.

7 Limitations of surname estimates

A key element of the surname estimates of mobility is that children inherit surnames strictly from one parent, typically the father. This condition can be shown to hold for England going back even as far as 1300. Thus, for the sample of families discussed above for England, when comparing individual estimates of social mobility with surname estimates, we have 72,853 men who were born and died in the period 1760–2019. Of these, 1,076 died with a different surname than they were given at birth, 1.5 per cent. But most of these changes were minor spelling variations: Skurr became Scurr, Beckerleg became Beckelegge. Another cause of name changes was the adoption of a hyphenated surname, sometimes adding the wife's name on marriage. Thus, Leschalles became Pige-Leschalles. Radical changes from one surname to another, such as when Twine became Methold, were rare—less than 0.2 per cent of men.

We can deal with the first source of changes, irregular spelling, by making sure to include all spelling variants of surnames. We can deal with hyphenation by counting all instances of the surname as including those where it is a component in a hyphenation. With that correction, surnames show very high fidelity across multiple generations.

However, there are many societies where surnames can change substantially across generations. Thus, in most of the Nordic countries surnames for the lower classes, until the end of the nineteenth century, were patronyms that changed each generation: Magnus Ollson's children would have the surname Magnusson. Also, the low status of names of this class, ending in -son, in societies such as Sweden has led in recent years to many people dropping such surnames and acquiring new ones, often at the time of marriage. It is possible to estimate social mobility rates for Sweden using surnames going all the way back to the eighteenth century, but only by looking at the small class of aristocrats and university graduates who had already adopted fixed, hereditary surnames by the eighteenth century.

In Japan there has been a long-standing practice of adult adoption, common among the samurai, whereby higher-status families without a son would adopt a 'surplus' son from another family, who would take the family name and ensure the continuity of the family. This is still an active practice, particularly among families running family business enterprises, and the great majority of adoptions in Japan are of adult males. Such a practice will lead surname estimates of mobility to overstate the persistence of status, since families will only adopt from a select group of candidates.

There are other societies where lower-class people have no surname, or surnames are just honorifics that change with each generation. Thus, in Muslim or lower-caste Indian communities women may adopt honorific surnames such as Begum, Bibi, or Devi.

In some societies, such as India, surnames carry such a strong signal of social status that there should be significant incentives for people, especially those upwardly mobile, to adopt higher-

status surnames.¹⁷ Such name-switching is limited, however, by the fact that people live in communities, and within extended families, where such opportunistic surname changes would attract social opprobrium. Thus, despite the known high status of such Brahmin surnames as Banerjee, Chatterjee, Ganguly, Goswami, and Mukherjee in West Bengal, the electoral register for Calcutta shows a strong decline in the percentage of Brahmin surnames at younger ages (reflecting both lower fertility among upper-class Indians and also greater adult longevity). If there were many people adopting such surnames in their 20s or 30s and passing them on to their children, we would not see such a pattern.

A more technical concern about the quality of surname estimates arises in the case where there is the most limited information—that is, where we know just the population shares of surnames in the society as a whole, and among some elite or underclass. To get intergenerational mobility estimates we have to make assumptions about exactly how elite or underclass the target group is. We also have to assume that status follows a normal distribution, and that this distribution has the same variance for the higher-status surnames as for the population as a whole.

How reasonable are these assumptions, and how sensitive to them are the estimates? We see above for English MPs that that the elite cut-off level does not seem to have much effect on the estimate. But the assumption that elite groups have the same variance of status as the population as a whole can be demonstrated to be incorrect for the English sample for 1810–1929 discussed above. For high-status surnames the standard deviation of wealth at death or occupational rank is higher than for the general population.¹⁸ How this affects the persistence estimates is hard to estimate theoretically. It implies that when we employ the assumption of constant variance we will initially overstate the mean status of the elite surname groups looking just at what fraction cross a given elite threshold. But as these surname elites converge towards average status, their variance should also change. So, the estimates of persistence rates might be lower than the complete measures show, or might be higher. However, with the English data for 1810–1929 discussed above we can compare the estimates of intergenerational correlations of status derived from complete information on status by surname in each generation, and those derived by observing just the fraction of persons above a cut-off level. For wealth, higher education, and occupational status in the period 1810–1929, the individual-level estimates of persistence average 0.74. The estimates using instead just some arbitrary cut-off of status average 0.83.

Thus, the methods of estimating intergenerational persistence of status that use the least information may tend to overestimate somewhat the levels of persistence. But any such bias is small relative to the difference between surname persistence rates and one-generation individual persistence rates. Surname studies clearly identify much more substantial long-run persistence of status than one-generation studies have been able to identify.

¹⁷ In England there is a class of names that now have unfortunate sexual connotations or other unfortunate connections: Cock, Prick, Bastard, Bomber, Nutter, Longbottom, etc. These surnames have been declining in relative frequency compared with the average surname.

¹⁸ This also true for occupational status in Australia during the period 1903–80: see Clark et al. (2020).

References

- Chief Electoral Officer, West Bengal (2019). 'Electoral Roll'. Available at: <http://ceowestbengal.nic.in/DistrictList> (accessed 10 February 2020).
- Clark, G., and N. Cummins (2014a). 'Inequality and Social Mobility in the Industrial Revolution Era'. In R. Floud, J. Humphries, and P. Johnson (eds), *The Cambridge Economic History of Modern Britain*. Cambridge: Cambridge University Press.
- Clark, G., and N. Cummins (2014b). 'Surnames and Social Mobility: England, 1170–2012'. *Human Nature*, 25(4): 517–37.
- Clark, G., N. Cummins, Y. Hao, D. Diaz Vidal, T. Ishii, Z. Landes, D. Marcin, F. Abu-Sneh, W. Chow, K.M. Jung, A.M. Marek, and K.M. Williams (2014). *The Son Also Rises: 1,000 Years of Social Mobility*. Princeton, NJ: Princeton University Press.
- Clark, G., N. Cummins, Y. Hao, and D. Diaz Vidal (2015). 'Surnames: A New Source for the History of Social Mobility'. *Explorations in Economic History*, 55(1): 3–24.
- Clark, G., A. Leigh, and M. Pottenger (2020). 'Frontiers of Mobility: Was Australia 1870–2017 a More Socially Mobile Society than England?' *Explorations in Economic History*, forthcoming.
- Corwin, L.A. (1977). 'Caste, Class and the Love-Marriage: Social Change in India'. *Journal of Marriage and the Family*, 39(4): 823–31.
- Dalmia, S., and P.G. Lawrence (2001) 'An Empirical Analysis of Assortative Mating in India and the U.S.' *International Advances in Economic Research*, 7(4): 443–58.
- Galdames, O.S., H. Amigo, and P. Bustos (eds) (2008). 'Apellidos mapuche: historia y significado'. Santiago: Department of Nutrition, Faculty of Medicine, University of Chile.
- Government of Bengal Political Department (1930). 'Series 2: Intermediate Revenue Authorities'. In *Press List of Ancient Documents relating to the Provincial Council of Revenue at Calcutta, Preserved in the Secretariat Room of the Government of Bengal*, 3(1): 1773–75.
- King, T.E., and M.A. Jobling (2009). 'What's in a Name? Y Chromosomes, Surnames and the Genetic Genealogy Revolution'. *Trends in Genetics*, 25(8): 351–60.
- Leonard, K., and S. Weller (1980). 'Declining Subcaste Endogamy in India: The Hyderabad Kayasths, 1900–75'. *American Ethnologist*, 1(3): 504–17.
- Ministry of Housing, Communities and Local Government (2019). 'English Indices of Deprivation 2019'. Available at: <http://imd-by-postcode.opendatacommunities.org/imd/2019> (accessed 10 February 2020).
- Ruggles, S., C.A. Fitch, and E. Roberts (2018). 'Historical Census Record Linkage'. *Annual Review of Sociology*, 44: 19–37.
- Weyl, N. (1989). *The Geography of American Achievement*. Washington, DC: Scott-Townsend.