How Strong is Assortative Mating? A Surname Analysis

Daniel Diaz-Vidal, Wabash College

Gregory Clark, University of California, Davis

November, 2015

A study of the status of surnames in Chile, for cohorts born 1920-1980, finds very slow underlying rates of social mobility, with an intergenerational correlation in social status of 0.83. This reflects a general finding internationally when social mobility rates are measured using surnames (Clark, Diaz-Vidal et al., 2014). But this high intergenerational correlation also creates a puzzle in terms of the mechanisms of inheritance of status. If status inheritance takes place at the level of the family then such strong persistence of status is only possible if marital matching is much more assortative than is conventionally measured. Using the structure of surname formation in Hispanic societies such as Chile we show that marriage is indeed significantly more assortative than conventionally measured. Thus the mechanism producing the observed high intergenerational persistence could still lie completely within families.

We first show the strong persistence of status across generations in Chile, as revealed by the movement in the average status of surnames. For births 1920-79 the intergenerational correlation of status at the level of surname groups is 0.83. This status persistence in Chile is typical of the patterns found for surnames in a wide range of countries, including England, the USA, Sweden, Japan, India, and China (Clark, Diaz-Vidal et al., 2014).

The strong persistence of status at the group level creates puzzles as to what mechanism can underlie this. Is this driven by inheritance of characteristics at the family level, so that the vehicle of inheritance might even be genetics, or does this imply that group level forces, true class effects, must be at work? For the inheritance of characteristics within families to explain this high level of persistence there has to be a higher degree of matching of spousal characteristics in marriage than is conventionally measured. We use surname evidence in Chile to show evidence of such high levels of assortment in marriage. We also show ancillary evidence of such high levels of assortment in underlying characteristics for English marriages 1800-1939.

Social Mobility in Chile from Surnames

Surnames allow us to measure rates of social mobility in societies such as Chile, because under naming conventions in these societies almost all the holders of a surname in one period will be the children of those who held the surnames in the period 30 years earlier. Thus by grouping people by surname we can parsimoniously measure rates of social mobility.

In Chile the source for this estimation is the Chilean Electoral Register of 2004, which records for six million voters their name, age, municipality, and occupation.[1] This allows people to be assigned a social status in two ways.[2] The first is based on the average earnings of their occupation.[3] The second on the average occupational earnings of all voters residing in their municipality. Since people only have an occupation on completing schooling, we look only at people born before 1980, who will be aged 25 by the time of the register. If we assume an average generation interval of 30 years we can then compare average occupational or locational status for those born 1920-49 compared to those born 1950-79. There are nearly 900,000 voters born in the first period, and over 2 million born in the second period.

To estimate social mobility rates we average earnings by surname groups, as described below. Then if $\overline{w}_{kt}$ is the average deviation of occupational or locational earnings in surname group k in period t (t = 0,1) from the population average, the intergenerational correlation of occupational status can be estimated as the b_i in

$$\overline{w}_{k1} = b_k \overline{w}_{k0} \qquad (1)$$

To get elite and underclass groups of surnames we can use a variety of procedures, detailed in Clark, Diaz-Vidal et al., 2014, 199-211. First surnames in an immigrant society like Chile can be classified by ethnic and national origin. Thus there are the surnames associated with the Mapuche, the main surviving indigenous population of Chile. There are also surnames associated with immigrant groups of Basque, German, French and Italian origin. Basque settlers, for example, were an early elite in colonial Chile. But further we can identify rare surnames associated with earlier wealth in Chile in the nineteenth and early twentieth century. An annual agricultural yield report was compiled, for example, in 1853 to

[1] Servicio Electoral Republica de Chile 2004. The electoral rolls listed 6,246,198 voters age 18 and above.

[2] For details on the sources for Chile see Clark et al., 2014, chapter 11, 199-211.

[3] This information is available for more than three hundred occupations. Chile, Ministerio del Trabajo y Prevision Social, 2008.

Table 1: Estimated Chilean Social Mobility Rates, births 1920-1979

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
Big Landowners, 1853	2,731	5,201	1.90	0.396	0.371	0.94
Big 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 those whom the electoral register lists with an occupation.

determine agricultural taxes.[4] This listed the value of land holdings of individuals in 1853. Another private survey of land holding was conducted in 1920. From these lists we selected those last names that appeared between 3 and 30 times in the Chilean Electoral Register, and were associated with various values of land held. Many holders of such rare surnames will be related, so if one holds substantial land wealth in 1853 or 1920 then the group as a whole is likely to be wealthy. The average yield value of a parcel of land in the 1853 report was 379 pesos. We take large holders as holding parcels of yield greater than 1,500 pesos.

Table 1 shows the numbers of people from each of four such surname groups found with an occupation in the 2004 electoral register born 1920-49 and 1950-79. 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 later is greater at 2.47, while for the high status groups it is lower.

[4] Chile, Estado que manifiesta la renta agricola. 1855.

The table also shows the average log occupational earnings of each group, relative to the average for all electors. Thus columns five and six show for each birth cohort

$$\frac{1}{N_{ik}}\sum_i \ln w_{ik} \; - \; \frac{1}{N}\sum_i \ln w_i \qquad (2)$$

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. For those with Mapuche surnames born 1920-49 the value of -0.304 implies that their average occupational earnings are only 74 percent 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 percent higher than the overall average for this birth cohort.

The b estimate in the final column comes just from the equation

$$\overline{\ln w_{k1}} \; = \; b_k \, \overline{\ln w_{k0}} \qquad (3)$$

where the subscript 1 indicates the generation born 1950-79, and the subscript 0 the generation born 1920-49. As can be seen, these estimates suggest strong persistence of occupational status for both the high status and low status groups. The average implied intergenerational correlation of status for an average of 11 such distinctive surname groups is 0.83-0.84 (Clark, Diaz-Vidal et al., 2014, table 11.2, 207). This finding of slow rates of social mobility for Chile echoes similar findings for other societies measuring social mobility through surnames: England, USA, Sweden, Japan, India, and China (Clark, Diaz-Vidal et al., 2014).

We posit that these slow surname estimates of mobility can be reconciled with more rapid mobility conventionally estimated from individual parent child correlations if social status has the following characteristic. Suppose that observed social status characteristics are a combination of an underlying inheritable status *x*, and some random component *u*.[5]

$$y_{it} \; = \; x_t \; + \; u_{it} \qquad (4)$$

Inheritance of y_i is driven by inheritance of x, where

$$x_{t+1} \; = \; bx_t \; + \; e_t \qquad (5)$$

This allows the intergenerational correlation on observed characteristics to be low, in the β in

[5] Where all variables here as measured as deviations from a population mean of 0.

$$y_{it+1} = \beta y_{it} + v_{it} \quad (6)$$

where $\beta = b\frac{\sigma_x^2}{\sigma_x^2 + \sigma_u^2}$, while the inheritance of underlying status can be strong. The surnames, because they aggregate across many individuals, are tracking the movement of the underlying status *x*, and thus revealing slow underlying mobility rates.

A puzzle, however, with this strong inheritance of underlying status is that the child would be expected to inherit the average of the underlying status of father and mother. Marriage in all societies is assortative. But the spousal correlation in characteristics like education, intelligence, health, income and occupation is typically modest, in the range 0.2-0.5.[6] If matching in marriage is based on observed social status, then the correlation of underlying inheritable status will be even lower – 0.1 to 0.25. On average a person with a high measure of underlying social status will marry someone much closer to the mean on the heritable component of status. But this implies that even if underlying status was inherited perfectly, the maximum correlation between any individual and their child in underlying status would be of the order of 0.5-0.6. Thus if we start with a group of people with high status surnames, then in the next generation we have to expect their children to have a correlation in underlying status of no more than 0.5-0.6. Social status would regress to the mean much more quickly than we observe, unless there are significant group effects that preserve social status.

In particular a low correlation in underlying status between marital partners will rule out genetics as the main transmitter of status. The reasoning is as follows. The correlation between one parent and descendants n generations later on any genetically determined trait is

$$h^2\left(\frac{1+m}{2}\right)^n \quad (7)$$

where h^2 is the hereditability of the trait, *n* is the number of generations, and *m* is the correlation in parent genotypes (Fisher, 1919, Nagylaki, 1978).[7] The intergenerational correlation of status for surnames, b, corresponds to the

$$\left(\frac{1+m}{2}\right) \quad (8)$$

[6]See, for example, Blossfeld, 2009, Breen and Salazar, 2011, Mascie-Taylor, 1988, 1989, Mascie-Taylor and Vandenberg, 1988, Schwartz, 2013.

[7] h^2 will be the observed correlation between the average value of the trait in the parents and the value in a child.

part of the above expression. Since the correlation of such phenotypes as IQ, education, and earnings between spouses is only typically in the range 0.2-0.6, it is expected that the correlation in genotypes will be much lower, assuming that the matching of spouses is based on phenotypes. m must lie in a range 0.1-0.3, so that a genetically determined trait can have at most an underlying intergenerational correlation of 0.55-0.65.

Below we use Spanish surname conventions that prevail in Chile to test whether the correlation in underlying status of marriage partners is as low as would be suggested from the observed correlations between marital partners in manifest status.

Assortative Mating and Surnames

Suppose, as above, that people have measures of manifest status y_i, but also a deeper latent status x, which is what gets transmitted to their children. Suppose also that the correlation across marital partners in the manifest status measures is ρ_y, and that the correlation in underlying hereditable status is ρ_x. If marital sorting is based on observed status elements such as education, then $\rho_y > \rho_x$. If, however, the matching is actually based on the deeper unobserved characteristic x, then $\rho_x > \rho_y$

Suppose also we know the frequency in the population, N_k, of each surname k, and the measured status of each holder. We can then calculate for each surname an average observed status on measure i which can be denoted

$$\bar{y}_{ijk} = \bar{x}_k + \bar{u}_{ijk} \tag{9}$$

where j indexes individuals. Each surname will reveal the average underlying status of the holders, plus a random element that diminishes as N_k gets larger.

Suppose also that average surname underlying status, at least for rare surnames, is correlated with individual status. This implies that for husband and wife the average status of each of their surnames can be represented as

$$\begin{aligned} \bar{y}_{ikM} &= \theta x_M + v_{iM} \\ \bar{y}_{ikF} &= \theta x_F + v_{iF} \end{aligned} \tag{10}$$

where θ is the correlation between underlying status of individuals and the average status of surnames, and M denotes the husband and F the wife. Now if we estimate the β in

$$\bar{y}_{ikM} = \beta \bar{y}_{ikF} + \eta \tag{11}$$

then

$$\hat{\beta} = \frac{\sum \bar{y}_{ikM}\bar{y}_{ikF}}{\sum \bar{y}_{ikF}\bar{y}_{ikF}} = \frac{\sum(\theta x_M + v_{iM})(\theta x_F + v_{iF})}{\sum(\theta x_F + v_{iF})(\theta x_F + v_{iF})}$$

And $$E(\hat{\beta}) = E\left(\frac{\sum(\theta x_M + v_{iM})(\theta x_F + v_{iF})}{\sum(\theta x_F + v_{iF})(\theta x_F + v_{iF})}\right) = \rho_x \frac{\theta^2\sigma_x^2}{\theta^2\sigma_x^2 + \sigma_v^2} = \rho_x \frac{\sigma_x^2}{\sigma_x^2 + \sigma_v^2/\theta^2} \tag{12}$$

where $\rho_x = \frac{\sum(x_M)(x_F)}{\sigma_{x_M}\sigma_{x_F}}$, and we assume $\sigma_{x_M} = \sigma_{x_F} = \sigma_x$. Thus $\hat{\beta}$ will provide a lower bound estimate of ρ_x. So the downward bias in $\hat{\beta}$ will depend on the size of

$$\frac{\sigma_v^2}{\theta^2}$$

The more frequent the surname, the larger is N_k, and the smaller will be σ_v^2. But correspondingly the correlation θ between individual family status and surname status will also typically be smaller. It is an empirical matter which scale of surname will give the least downward biased estimate of the correlation of underlying status between partners.

In the Spanish speaking world naming conventions allow us to measure the degree of assortative matching using the average measured status of surnames. If *Daniel Diaz Vidal* marries *Eva Abascal Aroyo* and produces a child Maria, then the child's name by convention is *Maria Diaz Aroyo*. Thus each person's name carries with it, potentially, information on the status match of the parents. We refer to the first of the surnames, inherited from the father, as the *paternal surname*, and the second, derived from the mother, as the *maternal surname*. Also in Spanish culture women after marriage retain their birth surname, so we can identify for both men and women the surname matching of their parents' fathers.

As noted above, we can divide the persons in the electoral register into two cohorts, born 1920-49, and 1950-79. Using the older generation we can attribute an average occupational status to each surname: this is the average of the logarithm of the average income associated with each occupation of the holders of the name born 1920-49. We can

attribute status to surnames in several possible ways. The first would be using only *paternal surnames,* the second using only *maternal surname*s, and the third using the average of both.

Figure 1 shows average attributed status (logarithm of occupational income) for individual surnames by surname frequency. As can be seen small size surnames vary much more in status than large ones. For the largest size surnames the correlation between individual underlying status and the surname status will be essentially 0. For the smallest size surnames the correlation between underlying status and average surname status will potentially be much larger. The random errors associated with average measured surname status for these names will, however, be larger.

Figure 2 shows the average status of all surnames by frequency groupings. There is a very slight tendency for rarer surnames to have higher status. But this effect is so weak that the correlation in average status between two randomly drawn surnames of the same frequency range will be close to 0.

Table 2 and figure 3 show the estimated correlations in occupational status of the paternal and maternal surnames for children in the second generation, those born 1950-79. As can be seen the correlation follows an inverted U with surname frequency. For the rarest surnames there is a low correlation in status, and for the most common surnames almost a zero correlation. The highest correlations occur in the frequency range 10-159, where the estimated correlation in surname occupational status across this entire range is 0.71 (standard error 0.003) based on 54,028 name pairings. Note that this will still be an underestimate of the true status correlation for underlying status between marital partners, because of the bias introduced by random elements in the measured average status of surnames.

This implies that marriage in Chile was highly assortative with respect to the underlying status of the parents, for people born 1920-49 at least, with an implied correlation in the underlying status of marital partners of more than 0.71. Note that in the estimates above paternal surnames pairing with identical maternal surnames are excluded, since in many cases this is because the father is unknown or undeclared, and it does not represent a true marital pairing. Since, however, some of these identical paternal/maternal surnames will be cases where the father and mother had the same paternal surnames, this omission implies a further downward bias in the estimates of the degree of assortative mating.

Figure 1: Surname Status Versus Surname Frequency, Chile

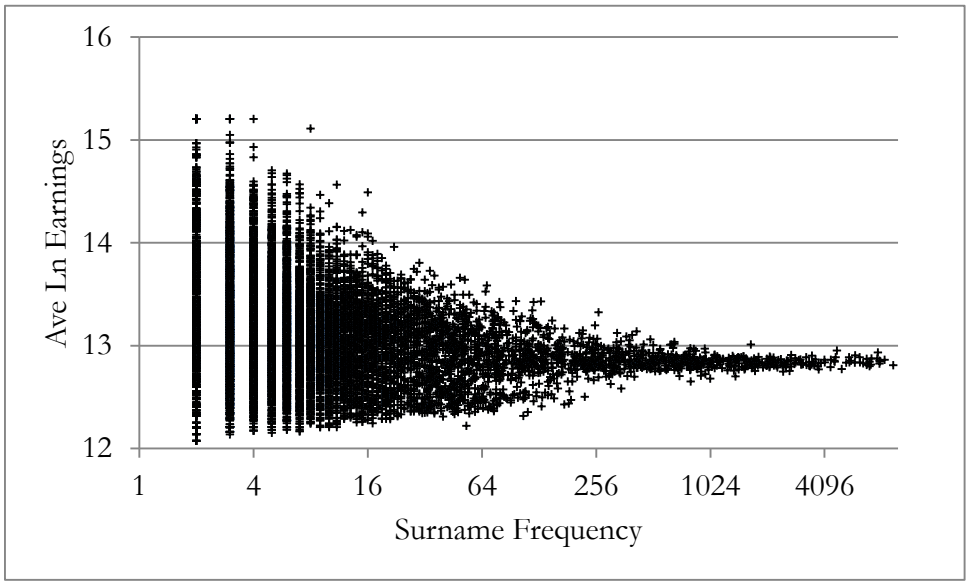


Figure 2: Average Surname Status Versus Surname Frequency, Chile

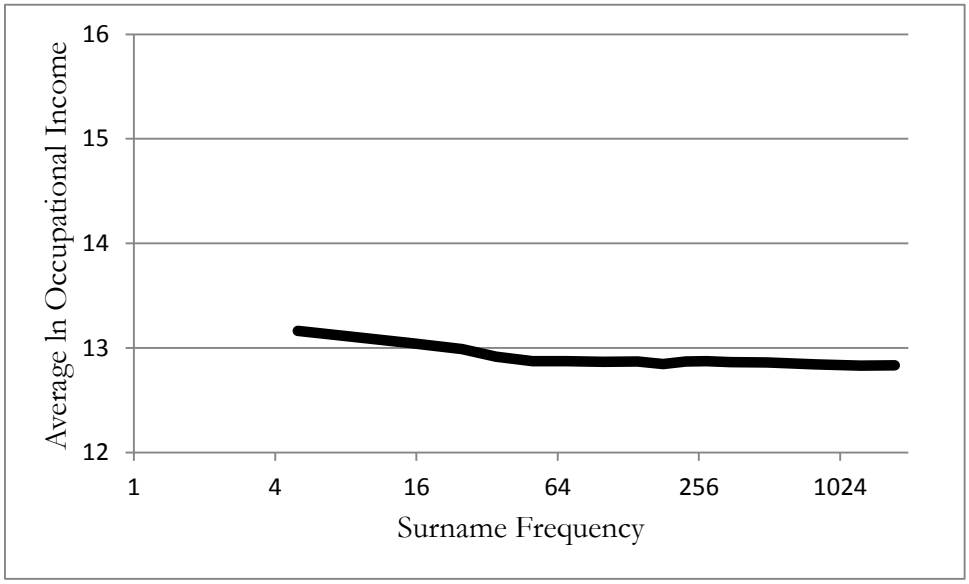
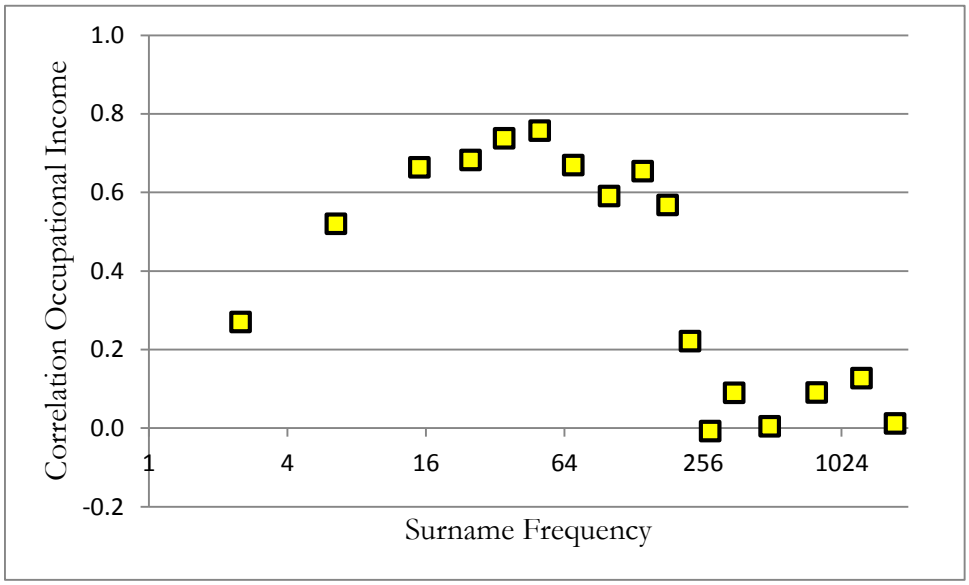


Table 2: Estimated Correlations of Surname Status, by Surname Frequency

Surname Frequency	Observations	Correlation	Standard Error
1-4	2,667	0.27	0.02
5-9	1,030	0.52	0.03
10-19	2,975	0.66	0.01
20-29	1,255	0.68	0.02
30-39	569	0.74	0.03
40-59	1,355	0.76	0.02
60-79	872	0.67	0.02
80-119	1,006	0.59	0.03
120-159	676	0.65	0.03
160-199	349	0.57	0.05
200-249	884	0.22	0.03
250-299	549	-0.01	0.05
300-399	2,526	0.09	0.02
400-599	7,077	0.01	0.01
600-999	28,729	0.09	0.01
1000-1499	15,221	0.13	0.01
1500-1999	13,140	0.01	0.01

Figure 3: Surname Status Correlations, Chile



We can confirm this result for Chile by using another method. In this we use the sets of surnames created to measure the rate of social mobility, discussed in the first section above, and in Clark, Diaz-Vidal et al., 2014. The one substantial low status group is the Mapuche, the descendants of the native Chilean population. The Mapuche have distinctive surnames, over four hundred of which were identified.[8] This cohort of surnames thus represents the lowest-skilled and most disadvantaged community in Chile. Although they used to be a rural population, confined to southern Chile, today nearly half the Mapuche reside in the largest city, Santiago.

There are a variety of high status surname groups we can identify. Some of these are from within ethnic communities such as the Italians, Germans and French. Italians as a whole, for example, were relatively elite migrants to Chile. Half of the Italian immigrants who entered Chile in 1853, for instance, were involved in commerce or were skilled professionals. More than 80 percent of Italian immigrants could read and write in 1865, compared to 35 percent of the host Chilean population. The electoral register of 2004 lists over six hundred thousand Chileans of Italian descent. A sample of surnames of the more successful Italian families was obtained from the *Commercial and Industrial Census of the Italian*

[8] Galdames, Amigo and Bustos, 2008.

Colony in Chile for 1926–27, from which were selected surnames held by dentists, doctors, jewelers, factory owners, and the like. Most of the surnames are rare.

But we can also find non-ethnic sets of rare surnames with unusually high status. One such surname group is the rare surnames of the the *encomenderos.* These were Europeans supposedly appointed as protectors of the indigenous population, but in reality grandees living of the labor of their charges. The tributes and benefits obtained by the *encomenderos* varied over time, but they undoubtedly represented the dominant socioeconomic elite of a polarized society in colonial times. The list of their surnames used here consists of rare surnames such as *Oyarzun, Ureta,* and *Iparaguirre,* all of which were held by at least one *encomendero,* but had fewer than fifty name holders in the general population in 1853 and 2004.

A second set of rare surname groups is that of landowners of various holding sizes identified in 1853 and 1920 land holding surveys. An agricultural yield report was compiled in 1853 to determine land taxes. This report includes the records of a large sample of landowners, who together possessed nearly fifteen thousand parcels of land. Owners with rare surnames (those appearing only three to thirty times in the 1853 census) were selected and divided into four wealth groups. These were based on the average annual value in pesos of the land owned: 350 or less (small); 350–1499 (medium); 1500–4999 (large); and 5,000 or more (very large). The last group represents the landowning elite in Chile in 1853.

In the 1920s, Juvenal Valenzuela created a detailed list of the major agricultural estates in Chile, and estimated the value of each estate. He found that in 1920, 10 percent of the landowners held 90 percent of the land, and he created a list of the top one thousand owners. From this list have been selected the surnames that had a frequency of three to thirty in the earliest-born cohorts in the 2004 electoral register. These are the large landowners of 1920.

Table 3 is constructed from cases where someone born 1950-79 had a paternal surname that belongs to one of the 12 distinctive surname groups, and a maternal surname that had a frequency 10-159. The second column shows the numbers of people included in each group. Column 3 shows the average occupational status of their father's surname in the generation of the parents (assumed to be those born 1920-49), measured as log occupational income. The groups are listed in order of average occupational income of the father's surnames.

Table 3: Observations and average status by surname group, born 1950-79

Surname Group	Persons (Gen 2) N	Average ln Occupational Income Paternal Surname (Gen 1)	Average ln Occupational Income Rare Surnames (10-159) Maternal Surname (Gen 1)
Rare, Large Landowners, 1920	1,237	13.35	13.22
Italian	643	13.25	13.09
Rare Encomienda	426	13.24	13.14
Rare, Largest Landowners, 1853	1,577	13.22	13.16
French	567	13.21	13.10
German	1,483	13.17	13.06
Basque	4,255	13.12	13.08
Rare, Large Landowners, 1853	1,457	13.09	13.08
Rare, Medium Landowners, 1853	2,119	13.02	13.00
Rare, Small Landowners, 1853	5,027	12.90	12.93
Mapuche	7,796	12.47	12.51
All	2,059,057	12.86	12.85

The status differences show clearly with the most elite group, those holding the rare surnames of large landowners in 1920, having average occupational incomes 141% higher than those in the least advantaged group, those with Mapuche surnames.

Column 4 of table 3 shows the average log occupational income of the maternal surnames – defined just as surnames held by 10-159 people in the electoral register in 2004. As can be seen, in general these rarer surnames have average occupational status. But as also

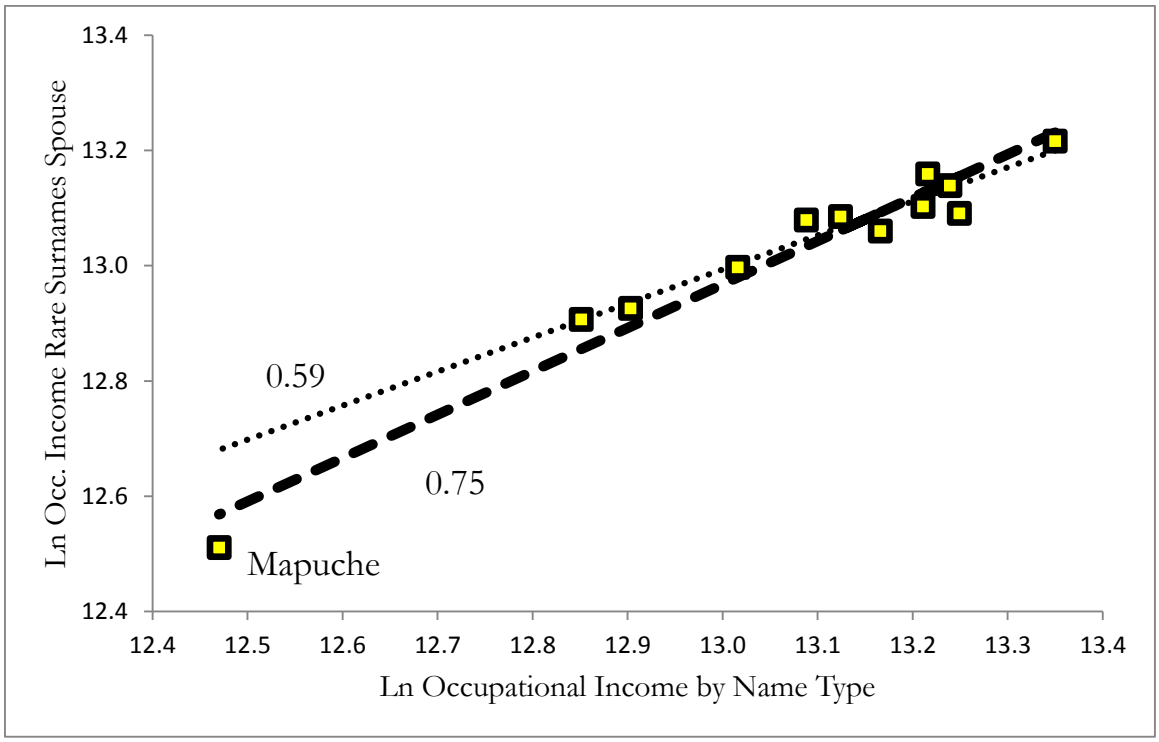
can be seen the maternal surname status correlates with the paternal surname status. Figure 4 shows the results of this exercise. There is a strong correlation between the average occupational status of the rare surnames and the status of the matched rare surnames. The regression coefficient on average occupational status by rare surnames (10-159) compared to the matched ethnic (or other) group surnames is 0.75 (standard error 0.05). Surnames are again matching very closely on status. Notice also that the R^2 of this regression is 0.95. The fit between status at the group level is very close. The high correlation of surname status across marital partners born 1920-1949 does not stem from marriage within ethnic groups such as Germans. The measure here is based on the correlation when someone with a German paternal surname marries someone with a rare paternal surname, most of which rare surnames which will not be German.

However, in the estimation above the Mapuche surname group is an outlier in terms of occupational status, at the low end of this group, and even though it is not a major outlier in terms of the regression line, it is significant in determining the slope of the line. It might be argued that the Mapuche, as a visible disadvantaged minority, are more likely to marry assortatively than the various European ethnic surname groups. If we omit the Mapuche, the regression slope becomes smaller at 0.59 (standard error 0.06, R^2=0.92), but still implies a high degree of assortment in underlying social status in marriage. However, if we think that these group level estimates of average occupational status are still subject to some measurement error, then the higher coefficient observed in the Mapuche case may stem just from the fact that with measurement error the estimated coefficient is the true coefficient multiplied by

$$\frac{\sigma_x^2}{\sigma_x^2 + \sigma_u^2}$$

where $\sigma_x^2$is the variance of the true underlying group occupational status, and σ_u^2 is the variance of the measurement error. By excluding the Mapuche we are reducing the true variance, and this will bias the coefficient further downwards if there is measurement error. Indeed the variance $\sigma_x^2 + \sigma_u^2$ (which is the variance of the observed group averages) drops from 0.057 to 0.023, so that if there is still measurement error in the group averages of occupational status, this will substantially increase any downward bias.

Figure 4: Average Occupational Status by Matched Ethnic and Rare Surnames, 1920-49



Other Tests of the Nature of Marital Matching

Above we see evidence from the correlation in surname status in Chile that marital assortative mating is much closer in the underlying social status of partners than in their status on any single aspect of measured status such as education. We can also test this proposition if we have data on more than one aspect of social status for marital partners.

Suppose as before that social status has manifest components, and an underlying heritable component that relates to the manifest components in the form

$$y_{iM} = x_M + u_{iM} \tag{13}$$

and

$$y_{iF} = x_F + u_{iF} \tag{14}$$

where M is the father, and F the mother, and now u_{iM} and u_{iF} are random errors with respect to the *x*'s. Let the correlation in marriage of manifest status i be ρ_{y_i} and the correlation of underlying status be ρ_x . If the matching is actually based on the underlying status *x* then the correlation between these values of *x* will have to be higher than for the observed elements of status. For then, based on (13) and (14),

$$corr(x_{Mt}, x_{Ft}) = \frac{\sum x_{Mt}x_{Ft}}{\sigma_{xM}\sigma_{xF}} = \frac{\sum x_{Mt}x_{Ft}}{\sigma_x^2} = \rho_x \tag{15}$$

and

$$corr(y_{iMt}, y_{iFt}) = \rho_y = \frac{\sum x_{Mt}x_{Ft}}{\sigma_x^2 + \sigma_{u_i}^2} = \rho_x \frac{\sigma_x^2}{\sigma_x^2 + \sigma_{ui}^2} \tag{16}$$

since we assume that u_{iM} and u_{iF} are uncorrelated across marital partners.

How can we detect if this is indeed the case that people are matching on underlying social status more closely than on any observed trait? Here the suggestion is that if we regress

$$y_{iM} = \lambda y_{iF} + \varepsilon_i,$$

but instrument for one observed status trait in wives, y_{iF}, with some other trait y_{jF}, then E($\hat{\lambda}$) = correlation of the *x*s. That is, by instrumenting for y_{iF} the estimated value of the marital correlation between partners will rise.

This follows from the fact that the IV estimate of λ is when y_{jF} is the instrument is

$$\hat{\lambda}_{IV} = \frac{\sum y_{jF} y_{iM}}{\sum y_{jF} y_{iF}} = \frac{\sum (x_F + u_{jF})(x_M + u_{iM})}{\sum (x_F + u_{jF})(x_F + u_{iF})}$$

But since all the u terms are assumed uncorrelated,

$$E(\hat{\lambda}_{IV}) = \frac{\sum x_F x_M}{\sum x_F x_F} = \rho_x$$

assuming that $\sum x_F^2 = \sum x_M^2$.[9]

Unfortunately, we do not have information in the Chile on multiple aspects of status for the individuals in the 2004 electoral register. However, such data is available for a sample of marriages in England 1800-1939. This sample derives from a large collection of data in England on wealth at death for people with rare surnames 1858-2012, a sample described in Clark and Cummins, 2015. For a large number of daughters in these families we have identified the husbands and have information on whether the husband was probated at death (an indicator of wealth), the husband's estimated wealth at death, their educational status, and their occupational status. Table 4 summarizes the data available on each characteristic for husbands. Women throughout most of this period did not have independent status measures. Few worked outside the home, and very few attended university. But we can measure their status in two indirect ways: through the status of their brothers, and through the status of their fathers. Table 4 shows the information for these women on the various measures of status for their brothers and their fathers.

The first measure of status is whether someone was probated at death (Dprob). Probate rates were high for high status individuals, and low for low status individuals. Those with wealth below a minimum threshold were not required to go through probate. The second measure is the logarithm of estimated wealth at death, where wealth for those not

[9] As before we assume here that all variables are normalized to mean 0 to make the exposition as simple as possible.

Table 4: Data on Husbands, Fathers and Brothers, England, 1800-1939

Status Measure	Group	Observations	Mean	Standard Deviation
DProb	Husbands	1,430	0.802	0.398
Lnwealth	Husbands	1,363	1.01	3.08
Doxb	Husbands	1,407	0.220	0.414
Occstat	Husbands	241	6.79	1.51
DProb	Brothers	4,252	0.693	0.461
Lnwealth	Brothers	3,932	0.39	3.26
Doxb	Brothers	4,370	0.170	0.376
Occstat	Brothers	1,373	6.43	1.42
DProb	Fathers	1,742	0.702	0.457
Lnwealth	Fathers	1,648	1.04	3.82
Doxb	Fathers	1,820	0.210	0.407
Occstat	Fathers	957	6.59	1.58

probated is estimated at half the threshold for probate at the time of their death (Lnwealth).[10] The third measure of status is whether someone attended Oxford or Cambridge, the two most elite universities in England (Doxb). Because our sample contains many relatively rich individuals this averages for the sample around 0.2, compared to 1-2% for the population as a whole. The final status measure is occupational status from the censuses of 1841-1911. The occupations are translated into a status score using a report from 1858 of the average wealth at death by occupation in England (Occstat). The measure here takes the natural logarithm of wealth at death for each occupation as the occupational status score. This score ranges from 4.19 (sawyers) to 9.53 (bankers). Since our sample

[10] Since people will dies in different periods and wealth is changing in both nominal and real terms over time we normalize this wealth measure by average estimated wealth at death in England for every decade from 1850 onwards.

skews towards high status families the mean of the husband, brothers and fathers in table 4 on this measure is always significantly above the entire sample mean for men pre-1880 of 5.99.

Table 5 shows the regression coefficient when we regress each of the four measures of status of husbands with that of their wife's brothers using OLS. The correlation is statistically highly significant but in the range 0.18-0.58. Marriage is assortative, but with the usual modest correlation in social attributes between the husband and the family of the wife. However, when we instrument for one social status measure of brothers with another status measure, in each case the coefficient increases substantially in value.[11] For the four measures of social status the average OLS regression coefficient for husbands' status estimated from brothers' is 0.35. When we instrument using another measure of status that average coefficient jumps dramatically to 0.87. Thus there is clear evidence here that the matching is based on some underlying characteristic of individuals that relates only with significant random error to individual measures of status.

Support for the idea that there is matching based on some unified underlying factor comes from the loose correlation we observe amount the four status characteristics of husbands. One correlation, that of *lnwealth* and *dprob*, is high at 0.78. But these two are both measures of the same thing, wealth. The correlation of *lnwealth* and *doxb* is only 0.31, and of *lnwealth* and *occstat* 0.51. And the correlation of *doxb* and *occstat* is only 0.27. Thus in any individual wealth, educational status and occupational status are loosely related. There is plenty of room for the randomness in the connection of individual status markers to underlying inheritable status incorporated in equations (13) and (14) above.

An alternative proxy for wife's status is the status of her father. Table 6 shows the OLS estimates of the connection between husband and father status. The average OLS regression coefficient for husbands' status estimated from fathers' is 0.37, slightly higher even than when we use brothers' status as a proxy. When we instrument for one social status measure of fathers with another status measure, in each case again the coefficient estimate increases substantially. The average estimated coefficient after instrumenting jumps dramatically again to 0.74.

[11] We do not instrument lnwealth using dprob and vice versa since lnwealth and dprob are alternate measures of the same thing, wealth at death.

Table 5: Correlations between Husbands and Wife's Brothers

Status Measure	N	OLS Coefficient	Instrument	N	IV Coefficient
Ln Wealth	2,886	0.378	Doxb	2,881	0.777
		(.022)			(0.075)
Ln Wealth			Occstat	1,398	0.861
					(0.053)
Dprob	3,252	0.251	Doxb	3,238	0.760
		(.024)			(.114)
Dprob			Occstat	1,463	0.739
					(.067)
Doxb	3,473	0.177	Lnwealth	2,959	0.767
		(.023)			(0.093)
Doxb			Dprob	3,184	0.839
					(0.137)
Doxb			Occstat	1,448	0.748
					(.084)
Occstat	311	0.582	Lnwealth	299	0.981
		(.063)			(.116)
Occstat			Dprob	301	1.040
					(.110)
Occstat			Doxb	311	1.365
					(.252)

Notes**:** Instrumental Variables regressions 2SLS, standard errors clustered by father. Because Doxb and Dprob are 0-1 variables, the estimation should properly be logit, but OLS is used here since the coefficients are then very simple to interpret.

Table 6: Correlations between Husbands and Wife's Fathers

Status Measure	N	OLS Coefficient	Instrument	N	IV Coefficient
Ln Wealth	1,246	0.433	Doxb	1,244	0.635
		(.020)			(0.075)
Ln Wealth			Occstat	686	0.616
					(0.041)
Dprob	1,377	0.367	Doxb	1,373	0.798
		(.029)			(.112)
Dprob			Occstat	719	0.688
					(.065)
Doxb	1,406	0.147	Lnwealth	1,291	0.927
		(.028)			(0.109)
Doxb			Dprob	1,365	0.759
					(0.113)
Doxb			Occstat	714	0.552
					(.097)
Occstat	117	0.533	Lnwealth	116	0.783
		(.078)			(.116)
Occstat			Dprob	112	0.861
					(.121)
Occstat			Doxb	117	0.872
					(.214)

Notes**:** Instrumental Variables regressions 2SLS, standard errors clustered by father. Because Doxb and Dprob are 0-1 variables, the estimation should properly be logit, but OLS is used here since the coefficients are then very simple to interpret.

Conclusions

Surname evidence suggests that conventional estimates of social mobility, looking just at one aspect of status across two generations, are misleading about the general rate of social mobility to be expected across multiple generations. Long run social mobility rates are much slower than conventionally estimated.

But this still leaves an interesting puzzle about what creates the strong intergenerational persistence of status in the long run. Can this be the result of the inheritance of characteristics within families? Or must it be the result of strong group or class effects on social outcomes? The crucial issue here is the degree to which marital partners match in terms of social status. The more closely people match in marriage, the more plausible it is that slow social mobility can be explained as a result just of effects within families.

Here we show that marital assortative mating is indeed very strong. We show this in two ways. First we examine the matching of surname status marriage in Chile in the twentieth century, and find surprisingly strong correlations. But second we look at marital matching in England 1800-1939, where we have individual evidence on the social status of the marital partners. Both these cases show a stronger underlying correlation in social status between marital partners than is shown in any manifest measure of status.

This data on the strength of matching in underlying status implies that strong status persistence could even be mainly the product of genetic inheritance of social abilities. If genetics controls social outcomes the long run intergenerational correlation of status would be, as noted above

$$\left(\frac{1+m}{2}\right)$$

where m is the genetic correlation between marital partners on the relevant characteristics. In this paper we find evidence of a match in underlying status between partners of 0.7-0.8. This would easily allow genetic inheritance alone to produce a long run intergenerational correlation of status in the order of 0.85-0.9, higher even than we typically observe using surname measures.

For England in the period 1858-2015 we have available comprehensive measures of the average status of surnames, since we have data on the probate frequency for each surname, and thus average wealth by surname, wealth at the surname level being strongly correlated

with educational and occupational status. We also have good measures of the average frequency of surnames from the birth, death and marriage registers 1837-2005. England has many rare surnames. Finally for 1911-2005 the index to the marriage register reveals the matching of surnames in marriage. Thus in future work we plan to replicate the surname matching exercise done above for Chile using the English data. When we look at marital matchings of rare surnames in the range of frequency 10-100 will we find the high correlation in social status across rare names observed above for Chile? And can we observe any trends in the strength of that matching over time with this longer panel of data.

References

Blossfeld HP. 2009. "Educational assortative marriage in comparative perspective." *Annual Review Sociology 35*: 513–530.

Breen R, Salazar L. 2011. "Educational assortative mating and earnings inequality in the United States." *American Journal Sociology 117*(3): 808–843.

Chile, Ministerio del Trabajo y Prevision Social, 2008. *Empleadores en la 2008, Servicio de Informacion para Estudios Secundarios and Trabajo.* Santiago.

Chile, Ministry of Planning and Cooperation. 2006. *Las trayectorias del desarrollo humano en las comunas de Chile, 1994–2003*. Santiago, Chile

Chile, Ofi cina del Censo. 1866. *Censo general de la Republica de Chile, 1865*. Santiago, Chile: Imprenta Nacional.

Chile, Estado que manifiesta la renta agricola. 1855. *Estado que manifi esta la renta Agricola de los fundos rusticos que comprende el impuesto anual establecido en la sustitucion del diezmo por la ley de 25 de Octubre de 1853*. Valparaiso, Chile: Imprenta del Diario.

Clark, Gregory and Neil Cummins. 2015. "Intergenerational Wealth Mobility in England, 1858-2012. Surnames and Social Mobility." *Economic Journal, 125*(582): 61-85.

Clark, Gregory, with Neil Cummins, Daniel Diaz Vidal, Yu Hao, et al. 2014. *The Son Also Rises: 1,000 Years of Social Mobility*. Princeton: Princeton University Press.

Clark, Gregory, with Neil Cummins, Daniel Diaz Vidal, Yu Hao. 2015. "Surnames: a New Source for the History of Social Mobility." *Explorations in Economic History, 55*(1), 3-24.

Epstein, Elizabeth and Ruth Guttman. 1984. "Mate selection in man: evidence, theory, and outcome." *Social Biology* 31(3-4): 243-78.

Fisher, R. A. 1918. "The Correlation between Relatives on the Supposition of Mendelian Inheritance." *Transactions of the Royal Society of Edinburgh, 52:* 399-433.

Galdames, Osvaldo Silva , and Hugo Amigo y Patricia Bustos, eds. 2008. *Apellidos Mapuche: Historia y signifi cado.* Departamento de Nutricion, Facultad de Medicina, Universidad de Chile, Santiago.

Goberno de Chile, Ministerio de Desarrollo Social. 2011. Encuesta Casen (National Socio-Economic Survey). Santagio: Chile. www.ministeriodesarrollosocial.gob.cl/observatorio/casen/.

Kamolnick, Paul. 2005. *The Just Meritocracy: IQ, Class Mobility, and American Social Policy*. Westport, Connecticut: Praeger.

Mascie-Taylor, C G N. 1988. "Assortative Mating for Psychometric Characters." In C. G. N. Mascie-Taylor and Anthony J Boyce, *Human Mating Patterns.* Cambridge: Cambridge University Press.

Mascie-Taylor, C G N. 1989. "Spouse Similarity for IQ and Personality and Convergence." *Behavior Genetics*, 19(2): 223-27.

Mascie-Taylor, C G N and S G Vandenberg. 1988. "Assortative Mating for IQ and Personality Due to Propinquity and Personal Preference." *Behavior Genetics*, 18(3): 339-

Nagylaki, Thomas. 1978. "The correlation between relatives with assortative mating." *Annals of Human Genetics, 42*: 131-137.

Nunez, Javier, and Cristina Risco. 2004. "Movilidad intergeneracional del Ingreso en un Pais en desarrollo: El caso de Chile." Working Paper no. 210, Department of Economics, Universidad de Chile.

Nunez, Javier, and Leslie Miranda. 2007 "Recent Findings on Intergenerational Income and Educational Mobility in Chile." Working Paper no. 244, Department of Economics, Universidad de Chile.

Schwartz CR. 2013. "Trends and variation in assortative mating: Causes and consequences." *Annual Review Sociology*, 39: 451–470.