Available online at www.sciencedirect.com

ScienceDirect

Explorations in Economic History 55 (2015) 3 – 24

www.elsevier.com/locate/eeh

Surnames: A new source for the history of social mobility

Gregory Clark[a], Neil Cummins[b], Yu Hao[c], Dan Diaz Vidal[d]

[a] Department of Economics, University of California, Davis, CA 95616, United States
[b] Department of Economic History, LSE, United Kingdom
[c] School of Economics, Peking University, Beijing, China
[d] Department of Economics, Wabash College, Crawfordsville, Indiana, USA

Available online 10 January 2015

Abstract

This paper explains how surname distributions can be used as a way to measure rates of social mobility in contemporary and historical societies. This allows for estimates of social mobility rates for any population for which we know just two facts: the distribution of surnames overall, and the distribution of surnames among some elite or underclass. Such information exists, for example, for England back to 1300, and for Sweden back to 1700. However surname distributions reveal a different, more fundamental type of mobility than that conventionally estimated. Thus surname estimates also allow for measuring a different aspect of social mobility, the underlying average social status of families. This is the aspect that matters for mobility of social groups, and for families across multiple generations.

Keywords: Social mobility; Intergenerational correlation; Status inheritance

1. Introduction

Why do we care about social mobility? At a descriptive level the social mobility rate is an important measure of the nature of social life. The lower the rate, the more children inherit the characteristics of their parents, and the more class-like is social structure. The higher the mobility rate, the less predictable are children's social outcomes at the moment of their birth. The social mobility rate indeed measures the degree to which history matters to current social outcomes. If it is high enough then every two or three generation societies are reborn in terms of the social hierarchy, and caste, class, and ethnicity are of marginal importance. If it is low enough then you cannot understand present social structure without an extensive investigation of lineage and history.

It is a significant and hereto unanswered question the degree to which the evolution of social structures and production technologies through history has influenced social mobility rates. Were all pre-industrial societies characterized by low social mobility rates? And did such low social mobility rates lead to a society where the talented were largely trapped at the bottom of societies in jobs that gave no avenue for their potential social contributions? Was modern economic growth accompanied by a significant increase in social mobility rates? Was modern economic growth created by a substantial increase in social mobility rates? Current methods of estimating social mobility rates do not allow

E-mail address: gclark@ucdavis.edu (G. Clark).

http://dx.doi.org/10.1016/j.eeh.2014.12.002
0014-4983

answers to these questions, but with surnames they can indeed be tackled.

Substantial social mobility has been taken by some also as, normatively, a requirement for the good society. In this view the great inequalities in income and wealth witnessed in most societies can only be justified if these prizes are available to all who work hard and take risks. The fact that social mobility rates, as conventionally measured, are high in modern successful societies such as the Nordic countries, serves as an indictment of social arrangements elsewhere where mobility rates are lower. There must be many societies where people derive unwarranted advantage, or suffer undeserved handicaps, as a consequence of the accidents of their birth.

2. Social mobility concepts

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

$$y_t = \alpha + \beta y_{t-1} + u_t \quad (1)$$

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 − β is the social mobility rate. Also 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 Eq. (1)

$$\mathrm{var}(y_t) = \beta^2 \mathrm{var}(y_{t-1}) + \mathrm{var}(u_t)\sigma^2 = \beta^2\sigma^2 + \sigma_u^2$$

If Eq. (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.

Eq. (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.

For example, a recent study of intergenerational wealth mobility in Sweden assembled data from Danish tax records that allows a comparison of the wealth of 1.2 million children with that of their parents.[1] The huge size of the Danish wealth data set means that the authors can divide the parents into percentiles and look at the average wealth of children for each parent percentile, measured again as a percentile of the child wealth distribution. Other than the top and bottom 3 or 4% of parental wealth, the picture has the linear character Eq. (1) assumes. One persistence rate, 0.20, describes inheritance across the middle 92% of the distribution (Fig. 1). The greatest deviation appears in the bottom 4% of parental wealth, where the children are much richer than we would expect. But the parents at the bottom of the distribution have negative wealth. This suggests not chronic, grinding poverty (the truly poor do not get to borrow much), but more likely indebtedness to finance a business venture or training. The fact that this is not truly the bottom of the wealth distribution explains the breakdown of the stable relationship. Children in the top 3% of the parental-wealth distribution also show slightly greater wealth inheritance. But though this effect is statistically significant, it represents only modest deviations from the single persistence rate in real terms: the persistence rates implied by the top three percentiles are 0.24, 0.23, and 0.22 respectively.

3. Estimating mobility rates from surnames

Conventional estimates of mobility rates require knowledge of the social status of parents and their children. Such data is publicly available on a systematic basis only in a few societies. In the contemporary world this requires long duration survey panels such as the US NLSY, or population registry data that assigns unique family identifiers, as in the modern Nordic countries. In the nineteenth and early twentieth century 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. And as we shall see below there are reasons to question if the conventional estimates of social mobility reveal its true rate for more generalized measures of status.

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

[1] Boserup et al., 2013.