



SIRE

FOCUS

2011-SIRE-03
www.sire.ac.uk

scottish institute for research in economics

The names in Spain are mainly not in vain:

Intergenerational mobility and the informational content of surnames

Maia Güell, José V. Rodríguez Mora and Chris Telmer

Almost all children inherit their surname from their parents. While it is unlikely that having any specific surname has much effect on the wellbeing of an individual, the surname is inherited along with other characteristics that actually do matter for the future welfare of the children; like wealth, or whether children were exposed to books; or beauty, or genes... Thus, surnames are informative about the wellbeing of individuals. Not because they matter in themselves, but because they travel across generations together with things that do matter.

This is not an inconsequential anecdote. On the contrary, it allows to measure in a novel and comprehensive manner the importance of background, and provides a way of viewing how the degree of intergenerational economic inheritance compares across countries and evolves across time. These are matters of obvious importance about which we know surprisingly little.

Intergenerational economic mobility (the probability that the child of a poor person ends up being rich, and vice versa) is notoriously difficult to measure. This is because the traditional procedure (correlating the lifetime income of parents with that of their children) requires very long panels of microeconomic data. To measure the mobility of one generation you need the lifetime income of parents and their children, a panel of at least 40 years. To measure mobility across two generations you need children, parents and grandparents: a minimum of a 70 year panel. Panels of this length are essentially unavailable. In a few cases (US, UK, Scandinavia) we have 30 or 40

year panels, no more. Even when the data are available, it is very difficult, almost impossible, to make comparisons across countries or across time. Consequently, we know very little about intergenerational mobility. We do not know whether it is larger in the US, in the UK, or in Europe; whether it is larger in growing or in stagnant societies, in richer or in poorer; in societies with a high degree of equality, or in very unequal ones. We do not know either how it has evolved over time; if it has increased or decreased.

Our research presents a novel and workable method of measuring intergenerational mobility which escapes from the slavery of panel data. This method consists in measuring the informational content of surnames: the more information a surname has on the economic welfare of its holder, the more that inheritance is determinant of the economic outcomes of the individuals. Thus, the more that surnames explain the wellbeing of individuals, the less mobility that there is. The data requirements are infinitely less demanding than in the traditional method, as it uses census data (a cross section of surnames and incomes or education) collected in practically all countries. Additionally, the study of the informational content of surnames allows to measure with ease the evolution of mobility over time, as every single census contains all the relevant information for many generations. It is possible to look at how informative the surnames are among the older cohorts, and among the younger ones. Thus, we can make an assessment of how mobility evolves over time.



This paper was presented at the launch of SIRE, November 2007 and has also been published as CEPR Discussion Paper Number 810

Contacts:

José V. Rodríguez Mora (University of Edinburgh)
Webpage: www.sevirodriguez mora.com
Email: Sevi.Rodriguez@ed.ac.uk

Maia Güell (University of Edinburgh)
Webpage: www.maiaguell.com
Email: Maia.Guell@ed.ac.uk

In the first stage, our paper develops a methodology to measure the degree of intergenerational mobility, as it is not obvious why surnames are informative, and even less so that the amount of information reveals the degree of mobility. Imagine that there were very few surnames, each shared by many individuals not necessarily family related. Surnames would not be informative, as they would not inform on family linkages. Two individuals called “Smith” are very unlikely to be family related, and thus any similarity between their incomes is a product of chance, and unrelated to the degree of inheritance. Fortunately, the distribution of surnames is always bound to be extremely skewed. That is, there are some very common surnames (their holders unlikely to be family related); but the huge majority of surnames are quite infrequent, accounting for a very large fraction of the population. Two holders of one of those unusual surnames are likely to be family related. These uncommon surnames are at the root of the mechanism, as the similarity of the incomes of their holders relates to how determinant background is; background being what they have in common. The reason for this skewness is that the process of generation and inheritance of surnames is akin to genetic process determinant of the distribution of DNA. It is a birth-death process by which surnames (lineages) die whenever the last male holder of a surname dies without male descendant (as in most western economies surnames are inherited across the male line). Lineages are also born; whenever somebody changes his surname (which happens occasionally), or when an immigrant arrives carrying a distinct surname. This process, being genetic is not “Darwinian”, as there is not necessarily differential survival in different mutations. ⁽¹⁾ The process is akin to the one generating the distribution of junk DNA (not coding for proteins) in the mitochondria or the Y chromosome (where there is no sexual mixing), and bound to generate very skewed distributions; thus, allowing to measure mobility. The paper presents a (genetic) model of the joint distribution of surnames and income. It shows that we can infer how important background is by looking at how informative surnames are. Extensions of the model allow for the possibility of assortative mating, and the introduction of ethnic differences in the income process (due to discrimination or any other reason).



The rationale for including assortative mating is that surnames are inherited only from the father, but background depends both on father and mother. More assortative mating means that agents are more likely to marry somebody of their own characteristics (the rich with the rich, etc.), resulting in a decrease of mobility and an increase of the informational content of surnames. This is because the characteristics of the father, explaining better the characteristics of the mother, also explain better the ones of the children.

The last issue that needs to be dealt with in order to have a comprehensive methodology is ethnicity. This is because surnames are not only informative about the family to which the individual belongs, but also about her ethnicity. Not controlling for ethnicity would bias the results, as the information on the surnames might be referring to the ethnicity of the holders, and not to their specific family background. The good news is that it is possible to use the surnames themselves in order to control for ethnicity.

The second stage of our paper, consists in using actual data to check that the methodology works. In this respect the Spanish naming convention comes in handy, as individuals hold two surnames (the first from the father, the second from the mother), pass only the first one to their kids (so the surname inheritance and survival works exactly as in the rest of the western world), and women never change surname upon marriage (allowing to use them in the study). Thus, the second surname can be used in order to identify ethnicity, while the first is used to identify family background. Also, the combination of the two surnames allows to identify siblings (as two individuals sharing two infrequent surnames in the same order are almost surely siblings) and to determine the degree of assortative mating among the parents (how much the surname of the father helps explain the surname of the mother).

The paper shows that in Catalonia surnames are informative, and in a manner that is perfectly coherent with the predictions of the model. Surnames contain information both about ethnicity (individuals with more “Catalan surnames” do better), but also about the family where the individual was brought up. Furthermore, the amount of information that surnames contain has increased steadily over time; indicating a decrease in mobility. A way to validate these results (and the methodology) is to use the second surname in order to identify siblings. Doing this, it is clear that effectively the correlation between siblings has increased over time, again indicative of a decrease in mobility. Finally, the decrease in mobility is explained by an increase in the degree of assortative mating that antedates the increase in information of surnames in one generation.

There are two readings of these results. One is the literal one: to show that in a modern society the degree of intergenerational mobility has decreased at the same time that the provision of public education has increased dramatically, and this is a consequence of an increase in assortative mating. The second reading is that surnames provide a viewpoint for looking at intergenerational mobility and the relevance of family background. This view is strongly reinforced by the fact that the results are identical whether using only one surname (which is all that it would be possible to do outside Spain) or whether using two and concentrating on siblings. This second approach (siblings) is devoid of any theoretical framework, obvious and self-explanatory. Thus, the new methodology proposed is strongly supported by the results: we can learn a lot by looking at how much surnames say.

⁽¹⁾ Actually it is “Darwinian” in the sense that the first person using surnames to study human populations was George Darwin (son of Charles) who used them in 1875 to study inbreeding in England. His parents being cousins, he was worried about the issue.