It is very difficult to measure changes in economic and social mobility across generations and to make comparisons across societies. Research by **Maia Güell**, **José V. Rodríguez Mora** and **Chris Telmer** finds that surnames can provide valuable insights into the importance of family background for people's outcomes in later life.

What's in a name?

Information on intergenerational mobility

Almost all children get their surname from their parents. While it is unlikely that a specific surname has much effect on its bearer's wellbeing, surnames are inherited together with other things that actually do affect a child's future life – such as genes, wealth, beauty and education. Surnames can therefore provide us with information about the wellbeing of individuals, not because they matter in themselves but because they travel across generations with things that do matter.

In this way, surnames offer a potential source of data for tracking the importance of family background for outcomes in later life and the degree to which people's economic and social status changes between generations. Surname data can show us how this 'intergenerational mobility' compares across countries and how it evolves over time.

It is notoriously difficult to measure the probability that the child of poor parents will become rich, and vice versa. This is because the traditional procedure (comparing the lifetime income of parents with that of their children) demands very long panels of data, which are hard to obtain. To measure mobility within one generation, you need the lifetime income of parents and children, a panel of at least 40 years. To measure mobility over two generations, you need lifetime income data for children, parents and grandparents – a minimum of a 70-year panel.

Even when the data are available, it is almost impossible to make comparisons across countries or over time. So we know very little about intergenerational mobility – whether it has risen or fallen over time; whether it is larger in the United States, the UK or continental Europe; or whether it is more prevalent in growing or stagnant societies, in richer or poorer societies or in societies where inequality is high or low.

Our research offers a new way of measuring intergenerational mobility that escapes from the slavery of panel data. Instead, our method measures the informational content of surnames: the more information a surname reveals about the economic welfare of its bearer, the more that inheritance determines people's economic outcomes. In other words, the more a surname tells us about the wellbeing of an individual, the less intergenerational mobility there is in that society.

The data requirements of this method are infinitely less demanding than the traditional method for measuring mobility, as we use census data (a cross-section of surnames as well as measures of income and education), which are collected in most countries. And by studying the informational content of surnames, it is

Gaudí Cornet
Vázquez Montalbán

easier to measure changes in mobility over time since every census contains all the relevant information for many generations. We can assess how mobility evolves over time by comparing the informational content of surnames among older and younger cohorts.

In the first stage of our analysis, we develop a methodology for measuring the degree of intergenerational mobility since it is not immediately obvious why surnames should be informative or that the amount of information they contain reveals the degree of intergenerational mobility. Imagine that there are very few surnames, each shared by many individuals who are not necessarily related. In this case, surnames would not be informative, since they would not indicate family linkages. For example, two individuals called Smith are very unlikely to be related, which makes any similarity between their incomes a product of chance and unrelated to the degree of inheritance.

Fortunately, the distribution of surnames is extremely skewed. In other words, there are some very common surnames (their bearers unlikely to be related), but the huge majority of surnames are quite infrequent, accounting for a very large fraction of the population. Two bearers of an uncommon surname are likely to be related. Uncommon surnames are central to our method, because income similarities between

The more a surname tells us about the wellbeing of an individual, the less intergenerational mobility there is in that society

individuals linked by an infrequent surname indicate the significance of background for economic outcomes.

The reason for this skewness is that the process of generation and inheritance of surnames is akin to the genetic process that determines the distribution of DNA. Surnames (lineages) die when the last male bearer of a surname dies without a male descendant (as surnames are inherited via the male line in most western societies). Lineages are born whenever a person changes his surname or an immigrant arrives carrying a distinct surname.

Our research develops a 'genetic' model of the joint distribution of surnames and income. It shows that by looking at the informational content of surnames, we can infer the importance of background. Extensions of the model allow for the possibility of 'assortative mating' (in which 'like marry like') and the introduction of ethnic differences in income (due to discrimination or any other reasons).

The rationale for including assortative mating is that surnames are inherited only from the father, but background depends on both parents. An increase in assortative mating – so that people are more likely to marry someone from a similar family background to themselves – results in a decrease in mobility and an increase in the informational content of surnames.

Surnames are not only informative about the family to which an individual belongs, but also about his or her ethnicity. Not controlling for ethnicity would bias the results, as the informational content of surnames might reflect ethnicity in addition to specific family background. Fortunately, we can use the surnames themselves to control for ethnicity.

The second stage of our analysis is to test the methodology against some data. We use data from Catalonia, where the Spanish naming convention comes in handy: individuals have two surnames (the first from the father and the second from the mother); they pass only the first of these to their children (the standard western tradition of inheriting the paternal

Dalí Domènech
Biedma Alba

surname); and women never change their surname when they get married.

Thus it is possible to identify family background through the first surname and ethnicity through the second surname. Our measure of the informational content of surnames captures how much first surnames explain the variation in individuals' outcomes beyond ethnicity and individual observable characteristics (gender, place of birth, etc.). We find that this is sizeable and, consistent with the model, larger when we focus on infrequent surnames, as these provide a closer approximation to family.

The combination of the two surnames also allows us to identify siblings (as two individuals who share two infrequent surnames in the same order are almost certainly siblings) and to determine the degree of assortative mating among the parents (how much the surname of the father helps to explain the surname of the mother).

Our analysis of the data shows that in Catalonia surnames are informative in a way that is perfectly coherent with the predictions of the model. Surnames contain information about both ethnicity (individuals with Catalan surnames do better) and family background. Furthermore, the amount of information that surnames contain has increased steadily over time, indicating a decrease in mobility.

One way to validate these results (and the methodology) is to identify siblings by using the two surnames. Doing this reveals that the correlation between siblings has increased over time, which also indicates decreased mobility. Moreover, this decrease in mobility is explained by an increase in assortative mating that predates the increase in the informational content of surnames by one generation.

There are two readings of our results. The first is literal: despite dramatic increases in the provision of public education in modern Spain, the degree of intergenerational mobility has decreased. This fall in mobility is a consequence of assortative mating.

Second, surnames enable us to look at the relevance of family background. This view is strongly reinforced by the fact that the results are identical whether using only one surname (for countries other than Spain) or using two and concentrating on siblings. Thus, the methodology is strongly supported by the results. We can learn a lot by looking at how much surnames say.

In Spain, the amount of information that surnames contain has increased steadily over time, indicating a decrease in mobility

This article summarises 'Intergenerational Mobility and the Informative Content of Surnames' by Maia Güell, José V. Rodríguez Mora and Chris Telmer, CEP Discussion Paper No. 810 (http://cep.lse.ac.uk/pubs/download/dp0810.pdf).

Maia Güell is professor of economics at the University of Edinburgh and a research associate in CEP's labour markets programme. **José V. Rodríguez Mora** is professor of economics at the University of Edinburgh. **Chris Telmer** is an associate professor at Carnegie-Mellon University.

Gasol Sáez
Miró Ferrà
Sánchez Vicario
Montalban