PROPOSAL: "INEQUALITY AND VERTICAL MOBILITY OVER THE VERY LONG RUN."

José Antonio Espín Sánchez, Assistant Professor of Economics, Department of Economics (jose-antonio.espin-sanchez@yale.edu)

The recent increase in inequality in the U.S. and other developed countries has stimulated a debate among economists. This debate is best exemplified by Thomas Piketty's *Capital in the Twenty-first Century* and his critics. Taking aside the predictions for the future and the role that capital gains play in this debate, a deeper question is: what is a "normal" level of inequality in historical terms? In other words, if we take the estimates at face value, and inequality today is similar to that of the 19th century and higher than most of the middle of the 20th century, which one is the historical anomaly? Is it normal to have high levels of inequality, and the 20th century was an exception, or is it normal to have low levels of inequality and the 19th century was an exception? Without information on inequality before the 19th century, it is hard to answer these questions.

A second question, which might be even more important than the first one, is the role of vertical mobility, especially over the very long run. Recent work by Gregory Clark (*The Son Also Rises*) advanced the idea that there is a magic rule/number in all societies and at all times. This rule implies a vertical mobility that is neither especially low nor high, and is universal to the human race. This work and the notion of universality of this force are receiving fierce criticism. The main criticism is the source of the data and its inherent bias. Although the data in the book and related works comes from different sources, they share important shortcomings.

The data from the Cambridge Group (for the history of Population and Social Structure) has a survivor bias, that is, they look at (male) individuals alive today, and reconstruct their family tree backward. They do not have data on individuals without descendants, or whether any of their individuals have siblings without descendants. Data from censuses typically have information about the occupation, but not about income. Assigning average occupation-income to each individual in one category means that the changes are measuring at the extensive but not intensive margin, creating important biases. Finally, there is severe criticism on the use of "names" as a proxy for income or status.

I propose a project to reconstruct an individual level full population database from a region in southern Spain. The first part of the project is to collect and clean the demographic data from online sources (baptisms, marriage and death). The second part consists of collecting first proxies for income (wealth at death) and then actual income from personal tithe records.

The RA will help create the demographic dataset. The RA will expand it with auxiliary data from other sources. Knowledge of econometric software such as STATA or R is required. Knowledge of the Spanish language at a basic level is recommended but not required. The RA should also be familiar with OCR software or willing to learn it.