

# The Quantitative Evolution of Entrepreneurs in the US

Sarada  
*University of Wisconsin-Madison*  
*sarada@bus.wisc.edu*

Nicolas L. Ziebarth  
*University of Iowa and NBER*  
*nicolas.lehmannziebarth@gmail.com*

January 23, 2015

What is the relationship between entrepreneurship and economic development? This is a question that has attracted the attention of economists since Schumpeter (1911). Recent empirical research such as Gollin (2008) drawing on cross-country evidence has found an inverse relationship between the level of development and rates of entrepreneurship as measured by self-employment. Fairlie and Woodruff (2007) confirm that this pattern prevails across the state-level in Mexico. These empirical findings are consistent with the model of Buera et al. (2009) that focuses on legal barriers to formal employment in developing countries. Besides driving this negative relationship with the level of entrepreneurship, these barriers create a tendency for lower skilled individuals to crowd into business ownership, driving down the productivity of these sectors. While suggestive, this cross-country evidence hardly seems sufficient to establish the institutional explanation for the patterns of entrepreneurship. Perhaps the differences are due to differences in demographics e.g. the paper by Lazear et al. (2014) or simply a reflection of the level of income.

We would argue that to understand this relationship it is necessary to follow a particular country over the course of its development in the face of a stable institutional environment. Work simply looking at time series variation within a country is quite scant. Gollin (2002) finds that this negative relationship between entrepreneurship and development holds within Japan comparing two different points in that country's development, 1930 and 1992. Even this comparison is clouded by not just large changes in living standards but the major changes in the Japanese economic institutional environment. We reconsider this relationship between entrepreneurship and development by studying the evolution of the American economy over the late 19th and early 20th century, a period of rapid growth and innovation.

We bring to bear on this question a new datasource of business-level records of a major credit rating agency, Dun & Bradstreet, scarcely used in economics. These credit reporting volumes contain information on broad swaths of the American economy at the time, going well beyond just manufacturing. For each individual business, the report lists the name of the business, which is often just the owner's name, location, industry, an estimate of net worth, and a credit score. We match the business owners' names to the Census of Population to obtain a rich set of individual and household demographic information, such as age, race, migration status, labor force participation, occupation, family size, and literacy. With this business-household matched data in place, we will

be able to study not just changes in the rate of entrepreneurship but also the changing demographics of entrepreneurs during this crucial period in the development of the American economy.

To understand the strengths and limitations of this source, it is useful to provide a short history of the company D&B and these data. Initially, the information in these reports was gathered by local agents not directly employed by the firm, often lawyers or postmasters who lived in the particular area. These correspondents would anonymously collect information (often skirting the law in the process) and report it back to the headquarters. The compensation for the information was sometimes monetary and, in other cases, the right to service the debt. Over time, D&B along with its competitors transitioned to a system of paid reporters who worked exclusively for a particular agency. Initially, D&B focused on collecting information on “credit constrained” businesses. Over time, D&B shifted towards a goal of being a “comprehensive” source covering all business. For example in the state of Mississippi in 1929, D&B successfully identified all of the manufacturing establishments as enumerated by the Census of Manufactures.

This grant would help fund the collection of these records for each of the census years between 1870 and 1930 for the following set of states: Connecticut, Colorado, Washington, Missouri, Maryland, and Wisconsin. Our sample would build on the existing samples: (1) a sample of 5 southern states in the late 19th century collected by Martin Ruef, (2) a sample for Mississippi for some years in the Great Depression collected by Mary Hansen and Nicolas Ziebarth, and (3) a sample from Maryland between 1860 and 1870 collected by Suresh Naidu and co-authors. Combining the existing samples with what we plan to collect would create if not a complete sample of all these records at least a sample broadly representative of the American economy.

We exploit the fact that many of the smaller businesses in the D&B data were simply named after their owner. Of course, not all businesses are associated with specific owner names and D&B provides no other information on the owners of these firms. So they must be excluded from the final matched dataset. While creating an obvious selection issue, we would argue that the set of businesses that are possibly matchable are *precisely* those of interest in the first place. In order to match the names of business owners D&B data to the Census, we have written a fuzzy matching script that keeps all matches that are “good enough” (we only search within the county the business is located) where “good enough” is defined using a support vector machine. We have applied this algorithm to the existing sample of late 19th century southern states from Ruef. Quite strikingly, the demographics of American entrepreneurs in the 19th century are remarkably similar to their counterparts in the late 20th and early 21st centuries. For example, in 1880, 76% of business owners were married while a hundred years later, 79% of self-employed in the 1980 PSID were also married. Furthermore, the average age of business owners in our 1880 sample and the self-employed in the 2005 PSID is almost identical around 43 years old. The matched dataset would also allow us to address the relationship between immigration and entrepreneurship. Specifically, in 1880, immigrants were more likely to be entrepreneurs and to operate more valuable businesses than native born business owners. These simple summary statistics offer suggestive evidence for the continuity of entrepreneurship in the US, but much more work remains.