Draft1

Sam

December 11, 2014

Quantitative Analysis of the Impact of Education on Entrepreneurship

Data Sources

To complete a quantitative analysis of the impact of educational factors – specifically the encouragement of creatitivity, self-sufficiency, and personal initiative – on the country-level rate of entrepreneurship, we have drawn on publicly available data from the Global Entrepreneurship Monitor (GEM) National Expert Survey (NES) and Adult Population Survey (APS).

The Global Entrepreneurship Monitor project is an annual assessment of the entrepreneurial activity, aspirations, and attitudes of individuals across a wide range of countries. The countries covered have grown from a low of ten to covering 75% of world population and 89% of world GDP by 2013. The 2001-2010 multi-year APS data base, for example, contains over 1.3 million observations of a total of 85 countries. This data is often cited in academic research, including reports by the Organisation for Economic Co-operation and Development.

The NES is structured around nine entrepreneurial framework contitions evaluated by thirty-six experts in each surveyed country. These nine entrepreneurial framework conditions are included below (a minimum of four experts must be interviewed on each of the framework conditions).

1. Finance
2. Government policies
3. Government programs
4. Entrepreneurial Education and Training
5. Research and Development Transfer (R & D)
6. Commercial and professional infrastructure
7. Internal Market openness
8. Physical infrastructure and services
9. Cultural and social norms

The APS is a questionnaire distrubuted to a minimum of 2,000 adults in each participating country. The survey is created by a central GEM team, but the implementation is managed by national teams that compete for the role of implementing the survey. Experts are chosen based on experience and specialization and are expected to represent the entire country (including urban and rural areas).

Data Selection

After reviewing the available data, we decided to select the following variables for our analysis.

Source	Indicator
APS	Rate of Entrepreneurship (Intention, Nascent, Established)

Source	Indicator
NES	Way of Teaching
APS	Perceived Capabilities
APS	Fear of Failure / Perceived Opportunities
APS	Perceived Opportunities

The survey questions corresponding to these variables are as follows:

1. **Entrepreneurial Intention**: Percentage of 18-64 population (individuals involved in any stage of entrepreneurial activity excluded) who intend to start a business within three years.
2. **Nascent Entrepreneurship**: Percentage of 18-64 population who are currently a nascent entrepreneur, i.e., actively involved in setting up a business they will own or co-own; this business has not paid salaries, wages, or any other payments to the owners for more than three months.
3. **Established Entrepreneur**: Percentage of 18-64 population who are currently owner-manager of an established business, i.e., owning and managing a running business that has paid salaries, wages, or any other payments to the owners for more than 42 months.
4. **Way of Teaching**: In my country, teaching in primary and secondary education encourages creativity, self-sufficiency, and personal initiative.
5. **Perceived Capabilities**: Percentage of 18-64 population who believe to have the required skills and knowledge to start a business
6. **Fear of Failure**: Percentage of 18-64 population with positive perceived opportunities who indicate that fear of failure would prevent them from setting up a business.
7. **Perceived Opportunities**: Percentage of 18-64 population who see good opportunities to start a firm in the area where they live

The APS data represent the national-level average of values between 1 and 100 that were gathered as responses to surveys within each country. The NES data is calculated based on the average answer to a survey to the national experts. For each question, the experts surveyed had to state whether it is:

1. Completely false
2. Somewhat false
3. Neither true or false
4. Somewhat true

(97. Do not know)
(98. Not applicable)

Analytical Framework

To examine the specific effect of the Way of Teaching on the rate of entrepreneurship we decided to select the following dependent, control, and explanatory variables.

Variable	Indicator
Y Dependent Variable	Rate of Entrepreneurship (Intention, Nascent, Established)
X1 Explanatory Variable	Way of Teaching
X2 Control Variable	Perceived Capabilities
X3 Control Variable	Fear of Failure / Perceived Opportunities
X4 Control Variable	Perceived Opportunities

The

In order to trace specifically how supportive of one's self-confidence a given education system is, we chose to use one specific item of the GEM National Experts Survey (NES) as a proxy, namely D01.

"In my country, teaching in primary and secondary education encourages creativity, self-sufficiency, and personal initiative"

Dependent Variables: Explaining the Country-level Rate of Entrepreneurship We have examined three dimensions of entrepreneurship to prepare for this stage of the project, established entrepreneurship in the country, nascent entrepreneurship (less than three months), and entrepreneurial intention. We will study each of these successively as we may well observe different effects.

To measure these three dimensions of entrepreneurship, we have used three Global Entrepreneurship Monitor (GEM) indicators:

- Established Business Ownership Rate (Percentage)
- Nascent Entrepreneurship Rate (Percentage)
- Entrepreneurial intention Rate (Percentage)

Key Explanatory Variable: Degree of Self-esteem Reinforcement within the Education System As Inglehart has shown how personal values are shaped before the age of 20, we will try to show how compulsory education (from 6 to 16 years of age) shapes individuals, especially in terms of self-esteem. To do so, we decided to incorporate GEM indicators that relate to self-esteem and confidence. First, we included GEM survey data on the average "perceived capabilities" of an individual (18-64) in a given country to start a business. Second, we included additional GEM data on how the "fear of failure" could deter an average individual (18-64) in a country to become an entrepreneur. Both of these data sets indicate the average percentage for each characteristic at the country-level.

Regressions The following equations were used to determine the relationship between our dependent and explanatory variables:

1. $PercentIntention = PercentPerceived(PerceivedCapabilities) + WayofTeaching + PercentOpportunities(Perceived$
2. $PercentNascent = PercentPerceived + PercentFear * PercentOpportunities + AnnualGDPGrowth$
3. $PercentOwner = PercentPerceived + PercentFear * PercentOpportunities + AnnualGDPGrowth$

The variables *Percent Fear* and *Percent Opportunities* were included as an interaction term $PercentFear * PercentOpportunities$ because of the design of the survey. PercentFear is: "the Percentage of 18-64 population with positive perceived opportunities who indicate that fear of failure would prevent them from setting up a business."

Data Collection and Cleaning To collect the data necessary to run these regression equations, we downloaded *.csv* data from GEM and used the WDI API to obtain the necessary annual GDP growth of OECD Countries.

Necessary R Packages

The following packages were needed to complete the data collection portion of this assignment:

- repmis

- tidyr
- reshape
- plyr
- WDI

GEM Data

GEM data was embedded in an interactive database and was difficult to extract automatically without writing a program. To simplify this process, we downloaded the necessary *.csv* files and saved them in our GitHub repository (to reduce complexity, we deleted the first three unnecessary rows before saving). We then linked to the RAW version of the files and added the code required to load them from *R*. The repmis package is required to load these documents from the repository.

Cleaning the GEM data required deleting unnecessary characters in the year columns, changing the years from columns to rows (changing the data from wide to long), and changing the column names to describe which data was being analyzed. In the process of doing this we created new data sets for each of these survey questions, which were called C*dataset* (clean followed by the data set name). This was done for each of the six GEM data sets before merging these sets two at a time to create the data set *FinalGEM*.

For all 33 OECD countries we should have 13 observations, one per year from 2001 to 2013, for each variable. However, there are a few missing values, especially regarding the perceived opportunities for some countries, namely Estonia, Luxembourg and Slovak Republic.