

The Country-level Impact of Educational Features on the Rate of Entrepreneurship - Data Gathering and Analysis

Samuel Dundon and Thierry Racaud

November 14, 2014

Introduction to Collaborative Social Science Data Analysis

Research Question To what extent do the behavioural characteristics of education systems - individualistic, group-oriented, self-esteem reinforcing - affect the rate of entrepreneurship in OECD countries?

Causal Chain Education system / Teaching reinforcing self-esteem \rightarrow higher degree of individual self-esteem \rightarrow higher probability an individual will become an entrepreneur \rightarrow higher rate of entrepreneurship at a country-level.

Dependent Variables: Explaining the Country-level Rate of Entrepreneurship We have examined three dimensions of entrepreneurship to prepare for this stage of the project. We have analyzed established entrepreneurship in the country, nascent entrepreneurship (less than three months), and entrepreneurial intention. We will study them 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)

We chose the GEM measure of entrepreneurship, as it is often referred to, especially in OECD studies.

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. Of course, it is not easy to quantify how supportive of students' self-esteem a given education system can be and we are still looking for the most adequate measure.

To begin with, we decided to incorporate to GEM indicators that relate to self-esteem and confidence. First, included GEM survey data on the average "perceived capabilities" of an individual (18-64) of a given country to start a business. The data consists of average percentage at country-level. Second, we included additional GEM data on how the "fear of failure" could deter the average individual (18-64) of a country to become an entrepreneur. The data consists of average percentage at country-level.

Going forward, we are considering adding one of the six questions used by GEM to measure the importance of entrepreneurial education at country-level in its NES Survey. Item D01 states: "In my country, teaching in primary and secondary education encourages creativity, self-sufficiency, and personal initiative."

Control Variables: Controlling for Classic Determinants of Entrepreneurship To control for the general state of the economy globally and in each country, we included Annual GDP Growth from the World Bank as a control variable in our regression equations. We first considered including real GDP, but decided to include growth to more accurately capture fluxuations in the economy. In addition, we included a GEM indicator called “Perceived Opportunities.” For each country, it is the average percentage of individuals who believe there are opportunities to start a business in the area they live in. Since we look at average individual self-esteem in every OECD country, we believe it is appropriate to look at the opportunities perceived by the average 18-64-year-old individual of these countries.

In the future, we may include lagged GDP growth in recognition of the time taken for governments and individuals to react to the state of the economy, and are planning to include several control variables from GEM. Specifically, we intend to control for the classic determinants of entrepreneurship – economic and structural determinants. To do so, we will add nine key entrepreneurship conditions as defined in the GEM [National Experts Survey \(NES\)](#) that may affect the rate of entrepreneurship. These nine conditions are the following:

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

We are not yet fully sure on how best to control for all these variables, especially as GEM NES data does not seem to be available in dataset format as such. We may need to build up our own dataset, collecting data from the many national GEM NES reports. We will work on this in the coming weeks.

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

1. $PercentIntention = PercentPerceived + PercentFear * PercentOpportunities + AnnualGDPGrowth$
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 imbedded 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 datasets for each of these survey questions, which were called *Cdataset* (clean followed by the dataset name). This was done for each of the six GEM data sets before merging these sets two at a time to create the dataset *FinalGEM*.

World Bank Data

As mentioned above, one of our control variables, annual GDP growth, was obtained using the [World Development Indicators](#) API from the World Bank. Using this API, we accessed GDP data from all the OECD countries directly from R and R will pull this information from WDI whenever this code is run. To clean this data, we used plyr to change the variable names then recoded the U.S. and South Korea country names to maintain consistency with the GEM dataset and dropped an unnecessary column that was not included with the GEM data.

Finally, we merged the annual GDP growth data with the *FinalGEM* data set to create *CompleteDataset*. This is the data set used to run the regressions and generate the charts and figures outlined in this report. To reduce complexity in the environment we also included a command to remove unnecessary objects from the environment (all the individual original and cleaned GEM data, and the intermediate merged data sets used to create *FinalGEM*)

Data Analysis To review, the three regression equations run for this assignment are the following:

1. $PercentIntention = PercentPerceived + PercentFear * PercentOpportunities + AnnualGDPGrowth$
2. $PercentNascent = PercentPerceived + PercentFear * PercentOpportunities + AnnualGDPGrowth$
3. $PercentOwner = PercentPerceived + PercentFear * PercentOpportunities + AnnualGDPGrowth$

To examine the relationship between these variables, we used least square estimation, the results of which can be found in the following table.

Linear Regression Estimates

Dependent variable:

PercentIntention

PercentNascent

PercentOwner

(1)

(2)

(3)

Percent Perceived

0.31***

0.12***

0.09***

(0.04)

(0.01)

(0.02)

Percent Fear

0.45***

0.09***

0.18***

(0.10)

(0.03)

(0.04)

Annual GDP Growth

0.36***

0.12***

0.12***

(0.11)

(0.03)

(0.05)

Percent Opportunities

0.42***

0.09**

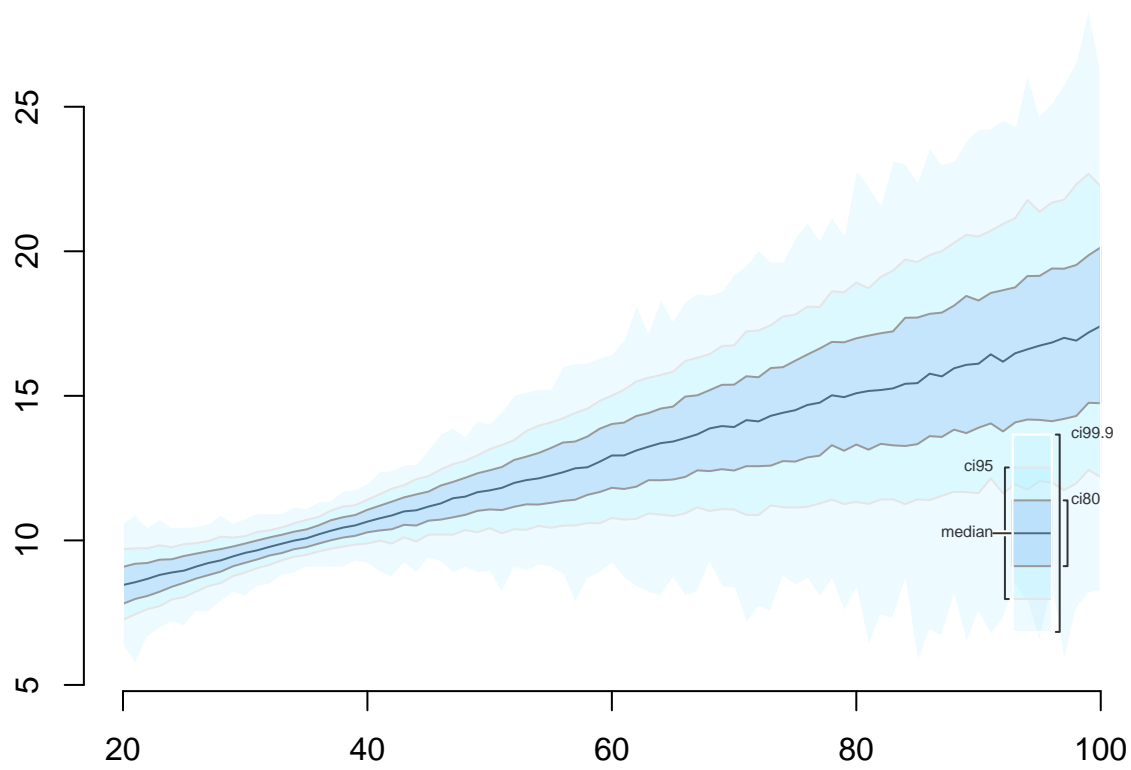
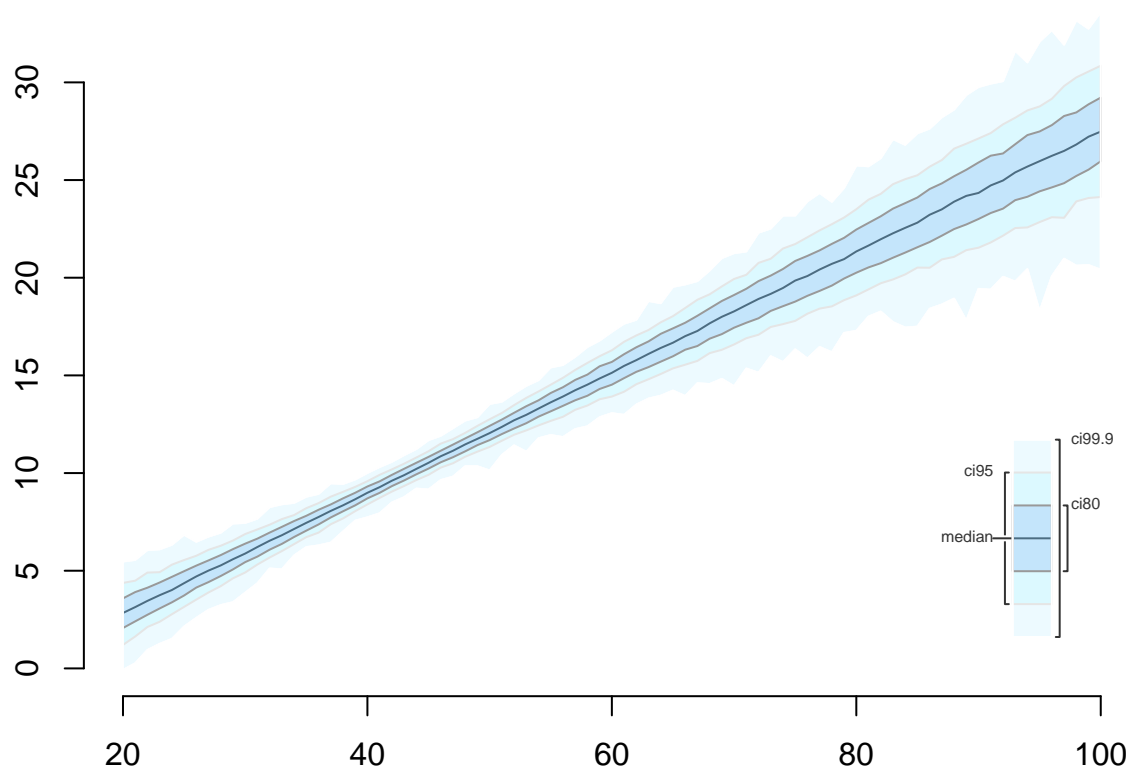
-0.03

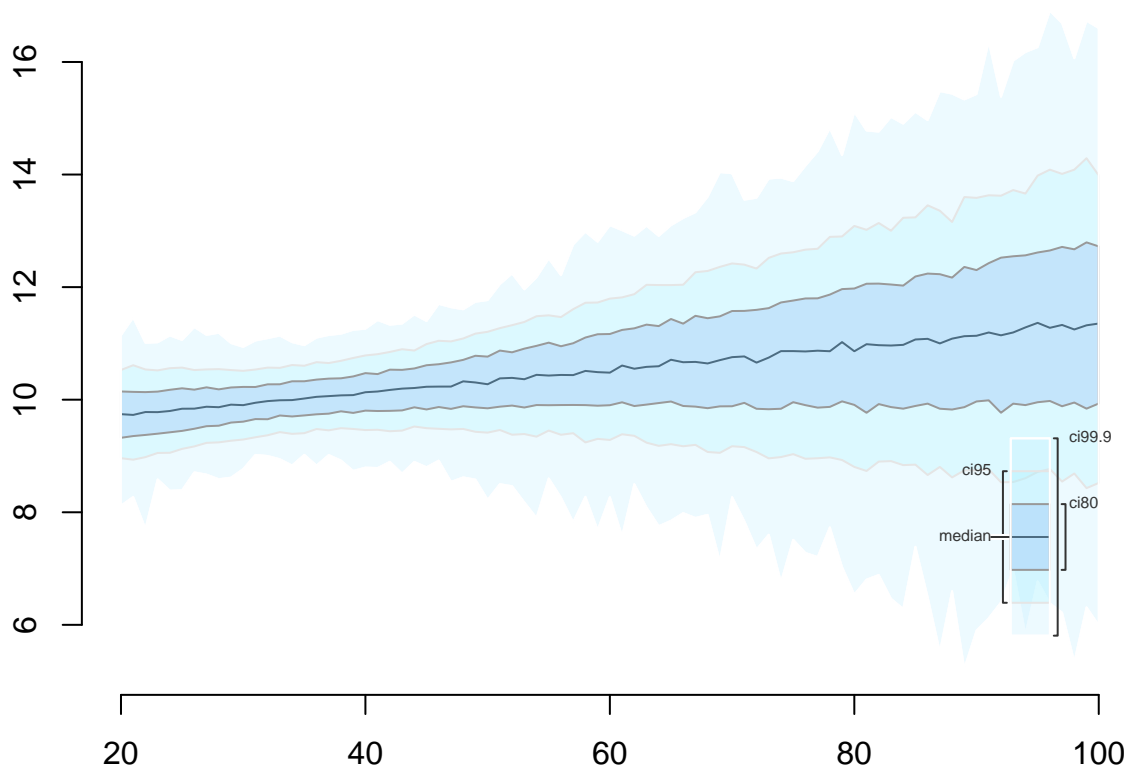
(0.13)

(0.04)

(0.06)
 Percent Fear x Percent Opportunities
 -0.01***
 -0.003***
 -0.004***
 (0.003)
 (0.001)
 (0.001)
 (Intercept)
 -20.01***
 -4.93***
 -3.74**
 (4.02)
 (1.23)
 (1.70)
 Observations
 295
 318
 317
 R2
 0.30
 0.39
 0.15
 Adjusted R2
 0.29
 0.38
 0.13
 Residual Std. Error
 5.88 (df = 289)
 1.97 (df = 312)
 2.73 (df = 311)
 F Statistic
 24.49*** (df = 5; 289)
 40.36*** (df = 5; 312)
 10.71*** (df = 5; 311)
 Note:

$p < 0.1$; $p < 0.05$; $p < 0.01$





Data Scraping and Cleaning In order to scrape our data, we used `repmis`, `tidyr`, `reshape2`, `plyr`, `WDI`, `stargazer` and `Zelig` packages before importing one dataset per variable mentioned above:

GEM Established Entrepreneurship Rate
 GEM Nascent Entrepreneurship Rate
 GEM Entrepreneurial Intention
 GEM Perceived Capabilities
 GEM Fear of Failure
 GEM Perceived Opportunities
 GDP Annual Growth Rate

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

First regressions R1

R2

R3

Summary

The impact of cultural factors in education on the rate of entrepreneurship in a given country is an important and under-researched topic that should be considered by policymakers constructing public education budgets. Through this project, we hope to encourage the consideration of these cultural factors by providing a qualitative and quantitative method for understanding the impact of these factors. Although several challenges remain to compiling the indexes discussed in this report, we are confident we can produce a helpful framework for examining and understanding these factors.

References