Cross-country differences in the size of venture capital financing rounds. A machine learning approach

by Marco Taboga

Motivation and aim of the paper

- Well-developed venture capital market widely considered important factor for economic growth
- Little money channelled by venture capital funds to Italian start-ups
  - In 2017, Italy 0.23 €bn, France 1.9, Germany 2.9
- Why? Less deals, but most importantly smaller deals
  - average deal size, Italy 2.4 €bn, France 6.1, Germany 11.6
- Firms that receive little funding are unlikely to become very big and give significant contribution to economic growth
- Aim: study determinants of deal size

BANCA D'ITALIA
EUROSISTEMA

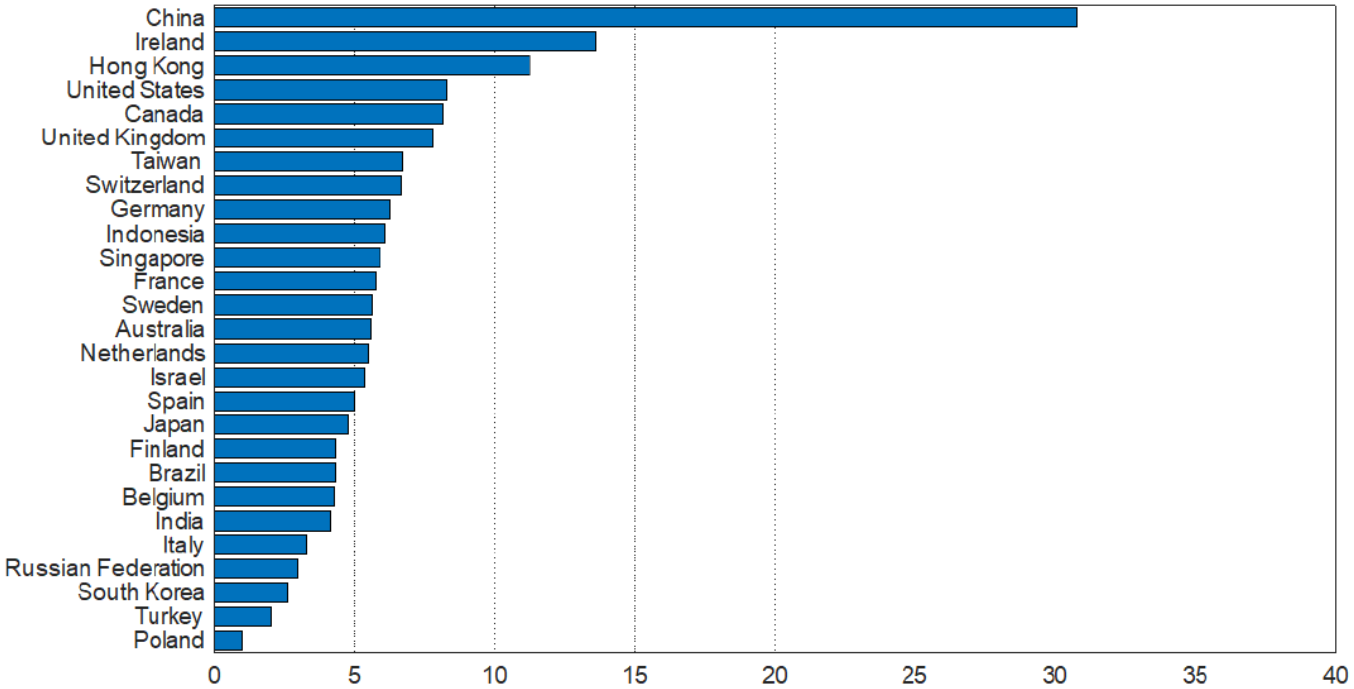
Determinants of deal size

- Dependent variable:
  - logarithm of funds (€) given by VC funds to a start-up in its Series A round
- Three blocks of independent variables:
  - Characteristics of start-up (755 variables)
  - Characteristics of VC industry in start-up's country (28 variables)
  - Other characteristics of start-up's country (1795 variables)
- Use ML to build one summary variable for each block

Data

- Crunchbase (info about start-up's, founders, deals, VC funds)
- World Bank data (info about countries)
  - World Development Indicators
  - Doing Business Indicators
- All series A rounds announced between Jan-2014 and Dec-2017
  - 7560 deals in more than 80 countries

Average size of series A funding rounds



BANCA D'ITALIA
EUROSISTEMA

Start-up's characteristics (sample)

- Firm age at time of Series A round
- Dummy =1 if certain words found in founders' bios
  - e.g., executive, experience, consulting, engineer, PhD, Fortune
- Dummy =1 if founders' alma maters are in top 10, 20, ..., 100 universities
- Dummy =1 if women in the group of founders
- Dummy for sector (e.g., software, telecom, biotech)
- Textual hints for smart branding
  - length of start-up name and web address, number of words in name, .com
- Dummy =1 if a given word included in start-up description
  - for all words that appear in more than one description

VC characteristics in start-up country (sample)

- Number of VC firms in country
- Number of investments made by each VC firm (quantiles of distr.)
- Number of investment exits (e.g., IPOs, mergers; quantiles)
- CrunchBase ranks of VC firms (quantiles)

Other characteristics of start-up country (sample)

- Population, its growth rate and composition, mortality and fertility rates
- GDP, its composition, distribution of income
- Trade (e.g., imports, exports of various goods)
- Taxation (e.g., incidence on GDP and different income categories)
- Characteristics of education system
- Characteristics of labor force
- Information about infrastructure (roads, electricity, telephone lines)
- Ease of doing business (various aspects)

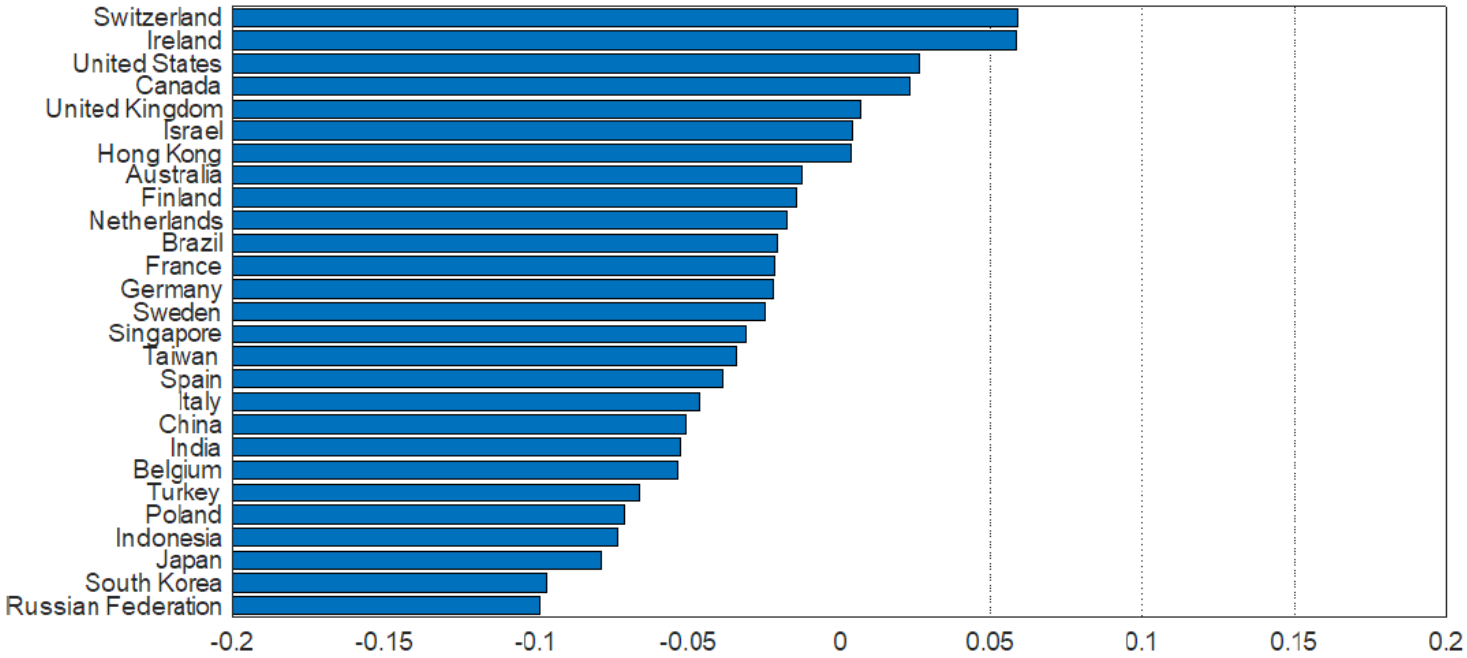
How to summarize variables in each block

- Stacked generalization + XGB (eXtreme Gradient Boosting) trees
  - Partition sample into K sub-samples
  - Use K-1 sub-samples to train and validate model
    - Target variable: log of deal size
    - Predictors: all the variables in the block
  - Use model prediction on excluded sub-sample as summary variable
  - Repeat K times so that summary variable is produced for whole sample

Why ML?

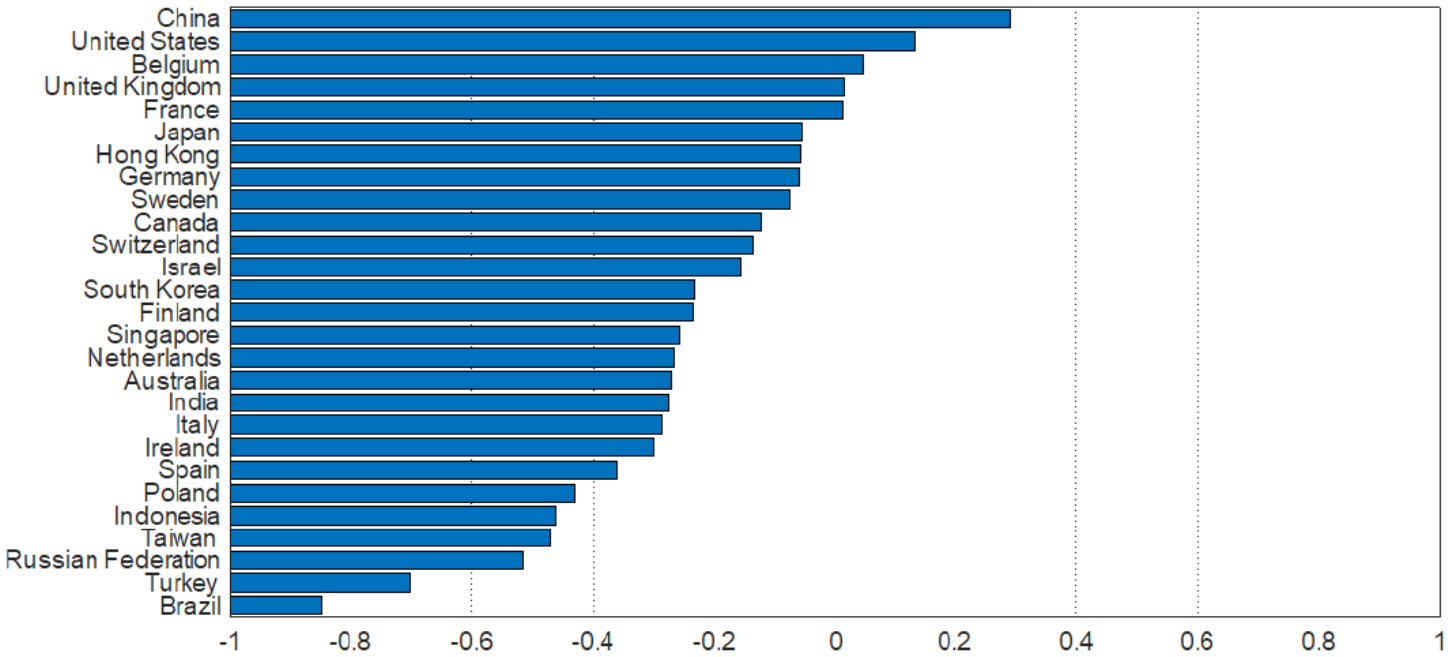
- XGB allows to deal with:
  - high-dimension of dataset
  - missing data
  - non-linearities
- Stacked generalization allows to re-use summary variable in second-order model to predict same target without introducing data leakages
- Summary variables have high predictive power for target and minimize over-fitting

Start-ups' characteristics (country averages)



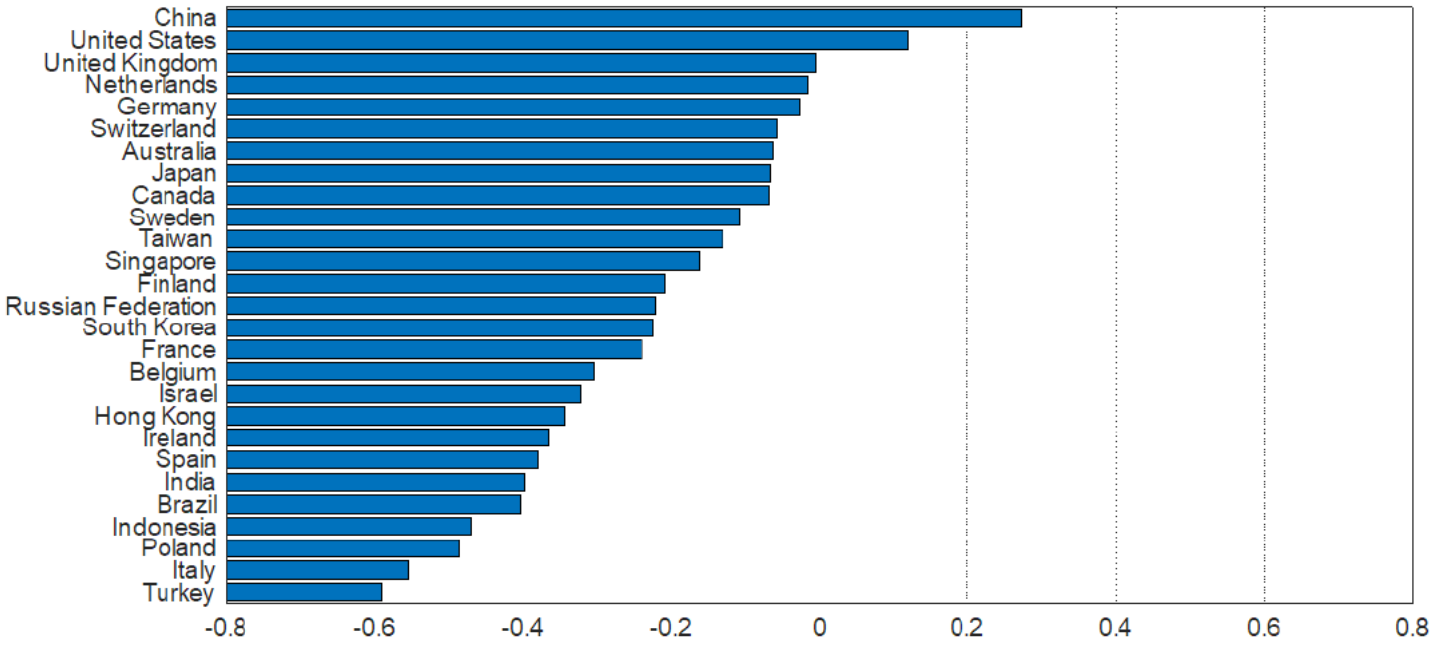
BANCA D'ITALIA
EUROSISTEMA

VC industries' characteristics



BANCA D'ITALIA
EUROSISTEMA

Other country-level characteristics



BANCA D'ITALIA
EUROSISTEMA

Second-order model

- Our second order-model is a linear regression:
  - Log of deal size as dependent variable
  - Three summary variables obtained from first-order model
  - Only two variables when start-up's characteristics are endogenized
- No exogenous shifter:
  - No clear-cut causal interpretation
  - Because we cannot completely rule out omitted variable bias
  - But we argue that omitted variable bias is unlikely for VC industry characteristics, which are the main area of policy intervention
    - Sign of confounder effects on regressor and regressand not economically plausible
    - Relevant country characteristics unlikely to be omitted (1795 variables in the dataset)

Main results

Regression models	1	2	3	4	5
Start-ups' characteristics	0.30***			0.27***	
	(9.92)			(11.8)	
Country-level VC characteristics		0.28***		-0.03	0.02
		(4.54)		(-0.89)	(0.76)
Other country-level characteristics			0.38***	0.37***	0.36***
			(8.71)	(7.62)	(8.15)
Constant	1.47***	1.47***	1.47***	1.47***	1.47***
	(19.3)	(41.0)	(38.4)	(26.3)	(37.5)
R^2	8%	6%	11%	17%	11%

BANCA D'ITALIA
EUROSISTEMA

Summary of main findings

- Size of financing rounds mainly associated with start-up's characteristics and the features of its home country (in particular, ease of doing business)
- Degree of development of VC industry highly correlated with deal size (and not redundant), but not significant once we control for other country-level characteristics
- Several robustness checks

Policy implications

- In the past, lots of government money worldwide poured into VC funds in the hope that a larger VC industry can spur more (and larger) deals
- Paper provides evidence that these policies are unlikely to be effective (domestic VC industry does not matter); resources would be better spent on improving ease of doing business
- Also theory (efficiency of financial markets) and high percentage of cross-border deals point to the fact that domestic VC industry should be irrelevant
- But analysis is non-causal and more evidence is needed

Thanks for your attention!

BANCA D'ITALIA
EUROSISTEMA