Cluster and network analysis for business models in the Mexican banking system

Luis Escobar-Farfan, Seran Martinez-Jaramillo, Manuel Ramos Francia, Santiago Garcia-Verdu
Banco de Mexico, Center for Latin American Monetary Studies

Discussion by Bernardus Van Doornik, Banco Central do Brasil
Summary

• Banks may engage in activities (business model) that are riskier than socially desirable

• The authors use
  • Factorial and reduced k-means cluster techniques (machine learning) to identify homogeneous groups of banks in terms of business models
  • Random forests technique to obtain the evolution of the importance for the balance sheet accounts
  • Stochastic Block Models to study the role played by the banks in the banking network

• Results
  • Relevant variables to define the business model: capital accounts, credit portfolio and other financial assets (securities and derivatives) important during the GFC
  • Banks’ business models change accordingly with the economic environment
  • Three different groups of banks (highly connected members, medium level of connectivity, and the periphery)

• Nice discussion of the methods and the importance of the analysis! Well written!
  • Maybe too long in the discussion of the techniques and short on the discussion of the results
  • “…first study in which both aspects of bank behavior (business models and interconnectedness) have been studied at the same time for such a long time span”
Comments

- There are papers that use machine learning with complex networks to study financial data

- Data:
  - The number of banks changed from 32 in 2007 to 48 in 2017
  - Did business models change because of the new competitors?
  - Accounting data: Does it deliver clear unbiased variables to classify business models?
  - Why not consider information on ownership, age, relative size, growth goals, operational plans, CEO changes, CEO interviews, M&A activities, ...?
• It is not clear whether the data authors use for clustering is a network or simply raw data
  • Community detection techniques are more robust to identify irregular cluster forms and do not have biases such as spherical clusters
  • In case data is a network type, community detections algorithms are preferred to traditional clustering techniques
  • Examples are Clauset et al. (2004)’s modularity optimization and Karypis (1999)’s CHAMELEON

• “Allowing the data to speak for themselves”: Absent on the economics behind the results. One example:
  • Capital is relevant to business models. But Capital is recommended by Basel and enforced by CBs. Thus Basel+CBs define business models?