Research Project Notebook

Wiktor Owczarz

June 22, 2018

1 Problem Statement

Rationale

Hypothesis: central bank researchers face higher **reputation risk**, compared to university researchers, when publishing banking competition studies. Economists employed by central banks are placed under **higher public scrutiny**, hence their publications can have stronger **signalling effect**.

As a result: central bank researchers are more prone to making **Type II Error** (failing to reject false null hypothesis). The cost of making **Type I Error** is high and might lead to financial instability.

Define reputation/signalling in context of economic theory

Define role of a central bank (supervisor/regulator) + literature

Objectives

Main Goal: Quantify if there are significant, measurable differences between banking competition papers published by the two groups of researchers.

Thesis Purpose: Create **methodology** that can be applied to analyse and measure the differences in competition papers.

Short: write methodology paper, define **heterogenity** and to measure it in a robust way

Possible *"sources"* of heterogeneity:

1. **Meta Study:** When looking at the same market, using the same *(similar?)* methodology two groups get either contrary or close but **different** (insignificant, smaller confidence, higher variability) outcomes.
2. **Text Analysis:** Looking at the two groups we find significant **differences in analysed texts**. As a result, we would be able to classify papers to either group based on the text structure.
   (a) Apply Latent Semantic Analysis to find **proximity** between different papers based on **bag of words approach**.
   (b)
   (c)

2 Methodology

Objective is to **create number of tools** that will allow us to analyse and compare competition texts and quantify if any differences.

1. **Initial Analysis**
   (a) Create corpus from the collected articles
   (b) Transform data into text vectors
   (c) Use tf-idf to weight the words
   (d) Use SVDTruncated to reduce dimensions
   (e) Run simple KMeans

   Very simplistic approach and does not fully capture the propose of the task

2. **Features engineering:**
   (a) word density: avg. length of words
   (b) punctuation cound
   (c) part of speech distributions
   (d) lexical diversity
   (e) **Uncertainty measure**: % share of the words that introduce ambiguity in the text

   Features add further complexity but ignore the data, calculations

References: (work-in-progress)

1. Stanley et. al. *Meta-analysis of economics research reporting guidelines*
2. A. P. Field, R. Gillett *How to do a meta-analysis*, British Journal of Mathematical and Statistical Psychology