Rasim Alam, Leonie Beyrle and Taran Raghuram

MIT 17.835 – Project description – Irregularities in worldwide corporate ownership data

<u>Motivation and key question to be addressed</u> → *Great job providing a sharp motivation for the project.*

The global economy loses approximately $3.6 trillion to corruption annually, with most losses concentrated in developing countries[1]. About 70% of large-scale corruption cases involve the use of anonymous companies[2], because opaque corporate structures make it easy to hide stolen assets and proceeds of crime – often abroad. In recent years, a growing number of countries have sought to address this problem by requiring companies to disclose (in public registers) information about their beneficial owners, i.e. anyone who holds over 25% of the companies' shares or voting rights (or otherwise exercises substantial control). Are these new registers a useful tool for detecting irregularities indicative of corrupt behavior or largely ineffective token gestures?

<u>Target data</u> → *What do we know about the data generating process? Is compliance perfect in all countries? Any potential sources of bias to be aware of?*

The OpenOwnership Register is the world' largest open database of beneficial owners. It aggregates data from the national registers of the UK, Denmark, Slovakia and Ukraine, in a standardized format and is updated monthly. In February 2020, the Register was made available for bulk download for the first time (under an Open Data Commons Attribution Licence). Key characteristics of the data:

Do you want to focus on all 4 countries? It's worth considering focusing on one only if that facilitates analysis, matching, & interpretation.

- 10GB of data, provided as a JSONLines file, formatted using the Beneficial Ownership Data Standard (BODS), which allows for linking of entities through unique identifiers
- Over 20 million records, relating to 7 million beneficial owners from 170 jurisdictions
- Key variables for individuals: nationality, name, birthday, address, interest in company
- Key variables for corporates: type, name, founding date, jurisdiction of incorporation

<u>Tentative research plan</u>

Step 1 – Basic descriptive analysis of data and visualization

→ other aspects of the "geography of beneficiaries" may be interesting…

- Are there hundreds of companies registered to a single (address?)
- How many companies have a registered beneficial owner with an address in a secrecy jurisdiction (such as Delaware) - which is not allowed under most national regulations?
- How many companies said they had no beneficial owner or were unable to identify them?

Step 2 – Network analysis – Visualize connections between different individuals and entities

- Has a nominee director been used to "puppet master" a large network?
- Are there links with non-profits or trusts without any obvious reason for them?

Step 3 – Merging the Register with other datasets to identify red flags

Would you have to merge by name? If so, there are useful packages for probabilistic matching.

- Are any beneficial owners (i) on an official list of "Politically Exposed Persons", (ii) on a US/ EU sanctions list or (iii) subject to an Interpol red notice? Many sanctions lists are available as csv files and red notices are published (individually) as JSON files, so information is available but may require substantial cleaning and standardization.
- The Register uses the same identifiers as the Open Contracting Data Standard so it may be possible to link the beneficial ownership data to data about public contracting.

↳ This would be amazing.

---

[1] United Nations News. 2018. "The cost of corruption: values, economic development under assault, trillions lost", available at: https://news.un.org/en/story/2018/12/1027971.
[2] World Bank, Stolen Asset Recovery Initiative. 2011 "The Puppet Masters – How the Corrupt Use Legal Structures to Hide Stolen Assets and What to Do About It", available at: https://star.worldbank.org/sites/star/files/puppetmastersv1.pdf.