

# Update 2

Alina Gafanova

July 14, 2021

## Done: Formal Stuff

Started a structured Box; shared Overleaf with everyone (thanks to Ishan); created a to-do list and meeting notes; stored all papers on Box.

## Done: Received Earthquakes Data from Stephanie

I talked to Stephanie and she guided me through the data on earthquakes. She also gave several references to better understand different earthquakes measures.

## Ongoing: Exploring Earthquakes Data

## Done: Primary Cross-Country Firm Microdata Search

I will have a short summary on firms microdata here in which I also implement comments by Ishan.

**Cross-country data** in theory could be the perfect source if it covered all countries and years with the same quality and was representative. However, COMPUSTAT has only publicly traded companies which is a severe limitation. And Worldbank Surveys' samples are tiny, also for some countries it's not even a panel but a cross-section.

The most promising one is still data by Bureau van Dijk (Orbis, Amadeus, Orbis Historical). It is primarily used for research of Europe or OECD, since these countries are better covered and have less bias. Ishan noted a problem with **different historical disks**<sup>1</sup>. Sebnem

---

<sup>1</sup>Data in Orbis depends a lot on the disk you were given or time that you downloaded from website, since Orbis would archive some data (eg if company went bankrupt) or add data (if they managed to find new data for earlier years).

Kalemli-Ozcan is the right person to contact about it since she knows how to overcome it (at least Kalemli-Ozcan et al. (2015) do it, but they definitely worked with Bureau van Dijk on it). Actually, I think her work with Bureau van Dijk led to creation of a newer database without this problem - Orbis Historical (quote from paper: “At the moment of writing of this paper, the Orbis Historical solution is still in the developing stage. Therefore, we focus on the steps required to work with the separate historic disks.”) So, if I am not mistaken, this issue should be already handled.

Moreover, I cannot assess Orbis quality for any of the countries we mostly care about because there are no thorough papers using it. I agree that coverage of our countries should be low quality in Orbis, since even European countries have problems. However, as noted above, Orbis Historical may be better. Unfortunately, Orbis Historical was still unavailable in 2015, when paper by Kalemli-Ozcan was published. I guess that’s why, due to novelty, I cannot find papers using Orbis Historical.

Adalet, McGowan, Andrews, and Millot paper that Ishan mentioned in his comments regarding Orbis is not particularly interesting since it only covers “best available data”, and there is really no intersection with countries we want (Europe and Korea). But I added its full description to Overleaf.

## Ongoing: Primary Country-Specific Firm Microdata Search

**China.** Ishan noted that he didn’t see Chinese firm-level data before 1998 used in research. I agree that people usually use data since 1998 (mentioned Kyle Meng; seminal paper by Klenow and Hsieh). Brandt et al. (2014) give an explanation: “(in 1998) ... it also changed the firm identifiers as part of a wider overhaul of the statistical apparatus. As a result, firms can only be easily linked from 1998 onwards and most studies start their sample in that year.” I changed available dates in excel and on overleaf.

**Country-specific Data** is documented in Excel and overleaf. Below is a small summary that groups first priority countries (we implement both EM-DAT and EQ list by Stephanie to prioritise this time). I covered all first-priority countries here.

**I. US, India, Indonesia, Mexico, China, Chile, Turkey, New Zealand:** we can access or know how to access this data.

**II. Japan, Philippines, Iran, Papua New Guinea, Peru, Colombia, Pakistan, Argentina:** data exists but access process is unclear. More research needed.

**III. Ethiopia, Tanzania, Europe (esp. Italy and Greece):** data exists and is potentially accessible but not first priority now.

**IV. Vanuatu, Solomon Islands, Guatemala, Afghanistan, Tonga, Fiji:** cannot find any systematic panel / no data with enough repeated observations / unsure about data

usefulness.

**V. Russia, South Africa, Kazakhstan, Ecuador, etc.:** countries that are not top priority, but are still worth checking, and I have not done it yet (or have not done in depth).

## Questions/tasks

- How to assess quality of Orbis Historical for our subset of countries since I don't have access to data nor did I find any relevant papers? And should I spend time on it at all?
- Ishan: could we briefly talk about how to write letters to other researchers about data? Like, should I add everyone to these letters/should I describe our project in detail/should I only contact people that are more or less in our network? Same question about writing to Statistical ministries/offices.
- Ishan: I have found at least 3 sources on Mexico: Mexican Economic Census, Annual Survey of Industries and Annual Survey of Manufacturing Industry. Could you please guide me which one is accessible to UChicago? I could not find it.
- Stephanie: what is the exact reason why we consider Italy and Greece irrelevant?