From conceptual ideas to data analysis

Javier Garcia-Clemente

Department of Economics
University of Huelva

8th Doctoral Workshop Seminar, June 2023

Table of Contents

At this point I already...

1. know the state of the art
2. have identified the gap in the literature
3. have a research question and hypothesis on that

Problem

I need data to empirically test my hypothesis

Data typology...

We can work with cross sectional data, time series or panel data.
And could be:

1. macro data (aggregated observational units)
2. micro data (individual observational units)

It will depend on the research question and the methods applied to address it

Disclaimer

I will leave aside unstructured data

Where to find data at the macro level...

1. Official statistical organisms: e.g. World Bank, OECD, Eurostat...
2. Alternative sources: e.g. corporate data, Google...

Pros

Easy access and use

Cons

Too aggregated?

Where to find data at the micro level...

1. Survey data: e.g. LFS, CPS, SHARE, EWCS...
2. Administrative sources: e.g. MCVL...

Pros
Granularity and potential

Cons
Hard access and sunk costs

In my case...

1. Available macro data was so aggregated that it does not account for that heterogeneity within the entrepreneurs
2. But literature states that entrepreneurs are too heterogeneous to be analysed as a whole
3. In this particular situation, we needed the use of micro data
4. Next part shows an example of how to manage survey micro data and turn it into macro if needed

Table of Contents

First steps...

1. Consult user guide, questionnaires, how to apply for access and any official information on the survey (usually accessible on the website)

2. Import data to be ready to use with your software. It might be very time consuming so try to find help or some aid from other users

3. Identify your variables of interest

In my case (LFS)...

1. EU-LFS information link
2. For EU-LFS, you can find user routines to read the raw files in Stata or SPSS, which helps a lot: GESIS link
3. Let´s assume at this point data is ready to use in Stata
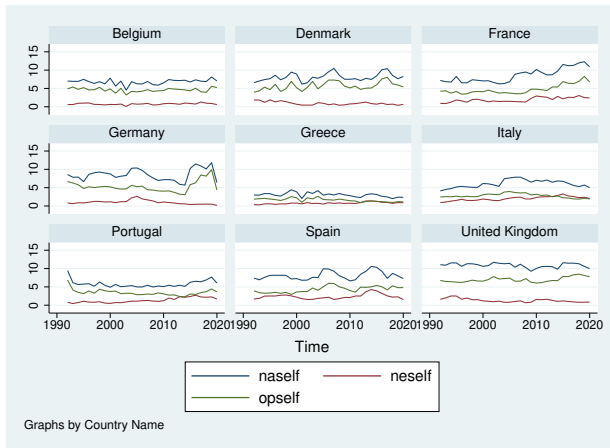
Some coding...

Go to do file 'DoFile8thDocWorkshop'

Some coding...

1. Same program can be 'easily' optimized and looped over multiple target variables and multiple waves, years and countries. Same applies to quarterly data. Appending those resulting files we can obtain aggregate time series or panel data from the micro data

2. Just be careful with the low representativeness of the sample when disaggregating too much

Work in progress...

1. We aimed to identify this source of heterogeneity: opportunity vs. necessity entrepreneurs
2. Following previous literature we used a proxy based on past labour state
3. Then, we combine the variables from past and current labour states for our nascent entrepreneurs
4. Finally extending a similar routine to obtain an European country panel for the rates of nascent opportunity and necessity self-employed over the stock of self-employed
5. Now we can run an independent macro analysis for each category

Figure: Percentage of overall nascent over stock of self-employed (naself), by necessity (neself) and by opportunity (opself).

Work in progress...

1. An interesting but controversial topic is the dynamics of entrepreneurship

2. Our idea is to apply a panel VAR on these data and other macroeconomic variables (GDP growth, unemployment rates and institutions)

3. While there is an unclear effect of those variables on the net generation of entrepreneurs (overall nascents)

4. Preliminary results point at significant positive responses of the opportunity to GDP growth shocks and of the necessity to unemployment shocks, which supports the pull-push hypothesis in entrepreneurship dynamics

Extra: another data application for Spanish-LFS...

1. Spanish LFS stock micro data is freely available on INE website

2. Download my Stata codes to work with it on my website

Thank you

javier.garcia@dege.uhu.es