ECONOMETRICS – TAKE-HOME #1

Context: Determinants of high-growth

Why do some firms grow much more than others? The data that you are going to use for this analysis comes from a random extraction of the Survey on Business Strategies (Encuesta Sobre Estrategias Empresariales – ESEE) in 2004 and 2005. This annual survey gathers exhaustive information on around 2000 manufacturing companies operating in Spain and employing at least ten workers.

With this type of information, a data analyst – in this case YOU! – can finally start looking for patterns regarding what determines high-growth. High-growth (HG) is thus a binary variable taking a value of 1 if the firm has a sales growth rate of more than 20% per year, 0 otherwise.

More precisely, your task is to:

- Provide descriptive statistics and figures of key innovation-related variables – i.e., R&D, product vs. product innovations, patents [2 points]
- Estimate a binary response model to explain the determinants of high growth [2 points]
- Be creative and find interesting patterns in the data [2 points]

This type of analysis is very important for managers and policy-maker, for example if they wish to identify promising companies for M&A and/or target financial support.

Here is the description of the variables at our disposal:

`id`: Company's ID [anonymized]
`year`: Year
`industry`: Company's main sector of activity [anonymized]
`yestab`: Year in which the company was established
`pertot`: Number of employees
`enggrad`: Share of engineers and graduates
`sales`: Total sales
`gom`: Gross Opearating Margins
`rdint`: Internal (in-house) R&D expenditures
`rdext`: External R&D expenditures
`patent`: Number of granted patents

You are required to prepare a 5-page presentation (PowerPoint or similar) with your main findings and attach the R code with comments.

ECONOMETRICS – TAKE-HOME #2

Context: Green Technologies and Greenhouse Gas Emissions

What is the relationship between green technologies and greenhouse gas emissions? The dataset to answer this question comes from an ongoing project with the JRC – European Commission. It contains regional-level information on patent applications labelled as 'green' (according to the OECD ENV-TECH classification), greenhouse gas (GHG) emissions, and a set of other variables:

`country`: Country code
`metro_code`: Metro-region code
`ghg_log`: Total amount of GHG emissions from industrial production, in logs[1]
`green_tech`: Number of green patents
`popul`: Population
`va_manuf`: Valude added of manufacturing
`urban`: urban region [dummy]
`intermediate`: intermediate region [dummy]
`rural`: rural region [dummy]
`rile`: Right-left indicator[2]
`qoi`: Quality of institutions[3]

With this type of information, a data analyst – in this case YOU! – can finally start looking for patterns regarding what determines regional GHG emissions. More precisely, your task is to:

- Provide descriptive statistics and figures of political (`rile`) and institutional variables (`qoi`), and green patents [2 points]
- Estimate an OLS model to explain the determinants of GHG emissions [2 points]
- Be creative and find interesting patterns in the data [2 points]

This type of analysis is very important for managers and policy-maker, for example if they wish to identify invest in specific technologies and/or propose solutions to cope with climate change.

You are required to prepare a 5-page presentation (PowerPoint or similar) with your main findings and attach the R code with comments.

[1] `https://ec.europa.eu/environment/industry/stationary/eper/legislation.htm`
[2] `https://manifesto-project.wzb.eu/down/tutorials/main-dataset.html`
[3] `https://www.sciencedirect.com/science/article/abs/pii/S0954349X18302066`