
Project instructions

Overview

The project is a group work that will count for the final mark of the course. It will require to apply different econometrics methodologies to real data using `Matlab`, `Python` or `R`.
The goal is to model and predict the real Gross Domestic Product (GDP) per capita of three countries of your choice using some of their respective economic data, and to compare the results. A written report and relevant code must be submitted prior to the deadline for the project to be graded.
This project description is available on the course web page. Please make sure to check for potential updates regularly.

Groups

You should work in groups of three people. This is an excellent opportunity to collaborate with your group colleagues, learn more by exchanging about your work and probably end up with a better project. Clearly, collaboration within groups is required and strongly encouraged but collaboration between groups is not allowed (at all), as group grades count for the final mark of the course.
Once you have formed a group—we recommend to do it as soon as possible if it's not already done—please email its composition to `suzanne.vissers@epfl.ch` who will assign you a group ID. If you have difficulties to form a group, please write a post on the Piazza forum of the course in the folder dedicated to the project.

Data

The Penn World Table provides purchasing power parity and national income accounts converted to international prices for 168 countries for some or all of the years 1950-2000. The data are available on http://datacentre.chass.utoronto.ca/pwt61/.
Download Penn World Table 6.1 (PWT 6.1) data for 3 countries of your choice (to make your project interesting, avoid the obvious country choices). Download all "PWT Topics" except those having a code starting with "RGDP" for the entire time period (1950 to 2000). Make sure that the countries you choose have data for the real GDP per capita and at least 12 "topics" for at least 40 years.

Analysis and modelling

The goal of the project is to model and predict the real Gross Domestic Product (GDP) per capita of three countries using some of their respective economic data, and to compare the results of competing/alternative models.
Use the first 80% or so of available data (typically years 1950 to 1990) in the dataset to fit a model, and perform predictions of the real GDP per capita for the remaining years. Build several competing

models for each country, test the significance of the regression model and the parameters/variables, and assess/compare prediction results. Make sure to assess and discuss the plausibility of the underlying assumptions of the models for the data at hand, use alternatives approaches as suitable, and compare/discuss results.

Code and programming

You might use `Matlab` or`Python` or `R` to perform the required analysis, modeling, and predictions (but pick only one of these programming languages for your project). Your `Matlab`/`Python`/`R` code should be directly executable, well organized, readable and abundantly commented. It will be used to check that your results are honest and correct, and that there wasn't code sharing between groups (which would be very strongly penalized).
One of the goals of this project is to help you understand the intricacies and details of the course material, and you'll have to demonstrate your understanding by implementing your own version of all formulas/methods covered in the course. You might use existing functions from `Matlab`/`Python`/`R` to ckeck/validate your code but this should not be part of the material you will submit at the end.
The data should NOT be hard coded in your code, but rather loaded dynamically from the project data file(s). Also, parameters such as names or numeric values should be stored in variables defined early and explicitly in your code.

Report

Each report should be typed in either English or French, and should not exceed 20 pages (including title page, figures, tables, and references). Unless explicitly requested, it is not necessary to provide the details of the methods or to include theoretical results; it suffice to briefly describe the methodology, to mention the approach/result used and to provide appropriate reference. Nevertheless, methodological choices should be justified appropriately and results should be explained clearly and commented appropriately. The report should include:

- a summary/abstract;
- an introduction presenting the problem and an outline of the report;
- a description of the data and an exploratory analysis (with appropriate tables/figures);
- a description of the methodology with relevant justifications, and the results (with appropriate tables/figures);
- an interpretation/discussion of the results/findings (with appropriate tables/figures);
- a conclusion and an outlook.

Usually, there is no need to include `Matlab`/`Python`/`R` code in the report. However, you might include some code if you (really) think it is relevant, in which case it should probably appear in an appendix. We strongly encourage you to typeset your report in LaTeX,[1].

[1] LaTeXcombined with `R` and the `knitr` package works like a charm to produce a dynamic report which automatically incorporates tables and graphics.

Support

If you need clarifications/help, you may ask questions during exercise sessions or on the Piazza forum in the folder dedicated to the project (but not to other group members).

Deadline

The deadline to submit your project is **Friday, 8 January 2021 by noon (11:59am)**.

Submission format[2]

Submit your project by e-mail to `suzanne.vissers@epfl.ch`. The submitted material should be bundled in a zipped file (naming convention: `Fin403_Group_XX.zip`, replace the `XX` with your group ID) containing the following files:

- a pdf file with your report;
- one or several CSV files with the required data to run your analysis;
- text files containing all your `Matlab`/`Python`/`R` code/commands to perform your analysis (including the commands needed to load your functions, add-on packages, the data, and to produce the tables and graphics)[3].

Make sure that you receive a feedback from us acknowledging receipt of your project. Assume that we did NOT receive your project submission if you don't receive a feedback from us (within a reasonable time frame), and resend it.

Marking scheme

All members of the group will get the same grade, which will be based on:

- The clarity and the quality of the presentation/exposition and the organization of the report (15%);
- The scope and the appropriateness of the methodologies and developments, and the relevance of analysis (75%);
- The quality of the `Matlab`/`Python`/`R` programming and compliance to submission format (10%);
- Originality and initiative (bonus up to 10%).

[2]Make sure to comply **exactly** with these instructions.

[3]You may submit "notebooks" (e.g., Live Editor file in `Matlab`, Jupyter Notebook in `Python`, or R Notebook in `R`), but they don't replace the report.