

## Empirical Project

As part of the course, you are required to write a project that demonstrate your understanding and ability to use econometric techniques in a practical context. The project will count for 30% of the course grade.

You will examine the data set provided and write up an empirical project based on this. *Maximum length of the project is three pages, excluding tables/graphs (use appendix section).* Below you will find a description, including description of data set and a brief discussion of the relevant research question(s).

### *Project Submission Instructions*

The due date is Friday, May 8 2020, 12:00 (noon). Works have to be submitted electronically on my e-mail ([ketevani.kapanadze@cerge-ei.cz](mailto:ketevani.kapanadze@cerge-ei.cz)). No late submissions are accepted!

You are encouraged to work in a group of up to 3 people. However, individual submissions are also possible. Only one submission from a group is required. The grade will be equally assigned to the group members.

### *Grading Criteria*

Grades will be allocated for:

- correct use of Gretl (or any other software of your choice) to implement the relevant econometrics;
- correct and clear motivation description of the steps leading up to your preferred model
- discussion of the results of your preferred model.

### **Organization and Description**

Two main skills are necessary to complete this assignment successfully. The first is an appropriate usage of econometric models and techniques. The second skill is an application of these methods in a sensible way and write-up the results in a clear and concise fashion.

In light of these considerations, I suggest that the project consists of two parts. Part 1 should be a short (approximately one quarter of the project) description of what econometric model (or models) and methods you propose to use and why you think they are appropriate. If you plan to do hypothesis testing, explain which tests you want to do and why.

Part 2 is the heart of the project. It should present empirical results and discuss them. It is a key part of the assessment for you to demonstrate the ability to estimate your econometric models and write up the results sensibly. The course lecture slides and textbooks give examples how empirical results are organized and discussed. The following are some general recommendations as to how you can format your empirical results section.

Your end goal will be a model that you must justify as being your preferred model. The main part of your empirical results section will involve presenting a table containing estimated coefficients (with its standard errors) for this preferred model and discussing the implications of the results. You may also wish to go beyond simply presenting a table of estimated coefficients and present tables or graphs of other features of interest (for example, summary statistics, correlation tables, etc.).

However, before you get to this main part of the empirical results section, *you should present evidence on why you choose your preferred model*. It will involve estimating some preliminary models, doing some hypothesis tests or model selection procedures. Your project should describe all these preliminary material in a clear and succinct fashion.

Since I am providing you with the data set and a brief introduction to the research question of interest, your project should not include a data description section nor an introduction.

The description of the data-set is presented in a separate file.

### ***What factors affect student performance at public schools?***

It is fairly easy to get school-level or at least district-level data in most states. Does spending per student matter? Do student-teacher ratios have any effects? It is difficult to estimate *ceteris paribus* effects because spending is related to other factors, such as family incomes or poverty rates. The data set MEAP93.RAW, for Michigan high schools, contains a measure of the poverty rates. Another possibility is to use panel data or to at least control for a previous year's performance measure (such as average test score or percentage of students passing an exam).

You can look at less obvious factors that affect student performance. For example, after controlling for income, does family structure matter? Perhaps families with two parents, but only one working for a wage, have a positive effect on performance. (There could be at least two channels: parents spend more time with the children, and they might volunteer at school.) What about the effect of single-parent households, controlling for income and other factors? You can merge census data for one or two years with school district data.

Do public schools with more private schools nearby better educate their students because of competition? There is a tricky simultaneity issue here because private schools are probably located in areas where the public schools are already poor. Hoxby (1994) used an instrumental variables approach, where population proportions of various religions were IVs for the number of private schools. Rouse (1998) studied a different question: Did students who were able to attend a private school due to the Milwaukee voucher program perform better than those who did not? She used panel data and was able to control for an unobserved student effect.<sup>1</sup>

---

<sup>1</sup> Note that the presented questions are just a subset of all available ones that one can explore using these datasets. Therefore, you are strongly encouraged to extend your analysis beyond the suggestions. Make sure that you present all your reasoning why you choose one model over another.