Microeconometrics and causal inference (Prof. Bastianin, Prof. Grembi)

Empirical project

A.Y. 2021/22

Aim

The aim is to let students develop their own short empirical project. While we provide a topic and some guidelines, **we also expect students to take an active part** and personalize the analysis with their intuitions and select additional empirical methods to enrich the analysis.

Comprehension of economic and policy issues should be the main factors guiding the empirical analyses and is assigned a large weight in the evaluation of empirical projects. Correct interpretation of statistical results and description of econometric methods contribute to the final grade.

Group

The project can be done in groups **not exceeding 4 students**.

Structure of the empirical project

The empirical project consists of a short essay **not exceeding 6 pages** excluding title page, figures and tables. Name, Surname, Student ID no. (matricola) and email address used in MS Teams, should appear in the title page. **Please include these information for each student in the group**.

Title page must also contain a **title** and a **short abstract of about 200 words** summarizing the project and main results. **See the attached word file as a title page**.

Divide the project into Sections: 1. Microeconometrics and 2. Causal Inference. In Section 1, also include the following subsections. (i) Data (describe data collection and dataset), (ii) Results. Clearly state to which part of the project you are answering.

Students will hand-in a single pdf file for each group. The pdf file will be uploaded in a folder that will shared via MS TEAMS in due time.

Topic

Analyse US presidential elections focusing on economic and incumbency variables. We use the research by Ray Fair as a starting point and try to enrich his model with a panel data analysis. https://fairmodel.econ.yale.edu/vote2020/index2.htm

Dataset

The first step in your analysis is to construct a panel dataset starting from the following information for the period 1976-2020

- Popular vote by US State: MIT Election Data and Science Lab, 2017, "U.S. President 1976–2020", https://doi.org/10.7910/DVN/42MVDX, Harvard Dataverse, V5 (Additional information here: https://electionlab.mit.edu/data)
- US Annual GDP by State in current and in chained dollars: https://apps.bea.gov/itable/iTable.cfm?ReqID=70&step=1#reqid=70&step=1&isuri=1
- US Population by State can be sourced from State Intercensal Datasets (https://www.census.gov/data/datasets.html)

These data can be used to construct the dependent variable of your regressions, that is the Democratic (D) share of the two party vote defined as 100*D/(D + R), where D and R are no. of votes for Democratic and Republican candidate.

You also need real per capita GDP by state and a price index based on GDP deflator defined as nominal (current dollar) GDP divided by real GDP. Then you can also try to compute a measure of strong growth following the approach of Fair (convert quarters into years). See the appendix for details on the construction of variables: https://fairmodel.econ.yale.edu/RAYFAIR/PDF/2018B.pdf

As for incumbency variables you can use the same variables used in the original paper.

Hint for extra points: if you wish to enrich the analysis you can add further variables such as state demographic characteristics from Census data collected before, additional economic variables such as State level unemployment from BLS, additional vote-related variables (e.g. the party of the Governor).

The dataset with the names of the students that have collaborated to building it must be attached to the pdf project. Please submit a compressed folder with both files.

Guidelines

Part 1. Microeconometrics (Bastianin)

Start from the estimation of a fixed effects model that contains the same variables used by Fair. Discuss coefficient estimates and inferences. Perform a test for the joint significance of economic variables in the model. Check if the panel dataset can be pooled by testing the joint significance of state-level fixed effects.

Now compute fitted values and assign the State to the candidate with the majority of votes. Translate popular vote into elector vote using the information here and compare your forecasts with the results for 2012 and 2020. (https://www.archives.gov/electoral-college/allocation).

Next, you need to apply the LASSO to the same regression, but add additional terms that you think can help forecasting presidential elections. Always keep real per capita GDP in the regression. Compare the performance of LASSO with that of the previous model. Provide valid inferences for the variables selected by the LASSO and comment in relation to the OLS estimates. Explain why a double-post-model selection approach is needed and a naïve post-model selection approach that simply excludes GDP from the LASSO penalty is not appropriate.

Part 2. Causal Inference (Grembi)

Think about the effect of the political colour of the winner on the public policies. What are the empirical challenges to disentangle causation from correlation, when a democrat is elected rather than a republican? What are the empirical challenges generally speaking? What are the models that can be applied to deal with these challenges? Why do we care about this? Why do you think that disentangling causation from correlation on this point is relevant policy-wise?