NHH

HOME EXAMINATION FIE401

Fall, 2022

Start: *07.12.22, Start: 09:00*

End: *09.12.22, End: 14:00*

THE HOME EXAMINATION SHOULD BE SUBMITTED IN WISEFLOW

You can find information on how to submit your paper here:
https://www.nhh.no/en/for-students/examinations/home-exams-and-assignments/

Your candidate number will be announced on StudentWeb. The candidate number should be noted on all pages (not your name or student number). In case of group examinations, the candidate numbers of all group members should be noted.

SUPPLEMENTARY REGULATIONS FOR HOME EXAMINATIONS

You can find supplementary regulations under the headline “Regulations”

https://www.nhh.no/en/for-students/regulations/

Find more information under chapter 4.0 in the Supplementary provisions to the regulations for fulltime study programs

Number of pages, including front page: 5

Number of attachments: 3

FIE401 - Final Project

Formalities

This final project will be handed out on 7.12.2022 at 9:00 and has to be submitted no later than the 9.12.2022 at 14:00.

In addition to the required hand-in files, there will be an oral exam on the 12.12.2022 and 14.12.2020. The oral exam will be held in Auditorium L. Please sign up for time slots on CANVAS as a group. You do not need to be present at the presentations of your peers other than your own group. You will have strictly 3 minutes to present your solution which will be followed by a question and answer session lasting up to 15 minutes. The latter will refer to both your analysis of the final project as well as all subjects covered in class.

This exam is group-based. All group members will receive the same grade for the written submission while grades on the oral exam will be individual.

Please submit the following documents

- **Coding file (.R)** Please comment your general coding approach for each task. You do not need to explain the used functions.
- **Report (.pdf)** In the report, you should present, analyse, and interpret your results. You should provide numeric and written answers to all questions asked in the exam. Do not discuss your coding in the pdf as this should be found in the R file. Please keep your answers short and precise and focus on the questions asked in the outline of the project.
- **Presentation (.pdf)** It is not allowed to change the slides between your submission and the actual presentation. Make sure that you use your 3 minutes presentation time wisely.

Formal requirements

- **Report**
  - Motivate all your choices.
  - Interpret your findings from a statistical and economic perspective.
  - Your report should read like a mini thesis/article having following structure:
    1. Abstract - summarize your findings (word limit = 100 words)
    2. Intro - outline the research objective and findings (max 3 paragraphs, appr. one page)
    3. Data - summarize the data and variable used (Table 1)
    4. Analysis - your analysis (Table 2, 3 & 4)
    5. Conclusion - your findings distilled
  - The report should include four tables.
  - The maximum word count of the report without title, tables, abstract, and header is 3000.
  - Make sure that tables are self-contained and as described in the last lecture.
  - Present a maximum of 3 models each in Table 2-4.
  - Use fairly plain formatting ensuring readability.
  - No appendix.
  - You do not need to explain the empirical models used as this will be part of the oral exam.
- **Presentation**
  - The main purpose of the presentation is to be a starting point for the oral exam.
  - The time limit of 3 minutes will be strictly enforced.

- Discuss your choices of how you set up the econometric analysis.
- All four tables of the report have to be in the presentation.
- You do not need to prepare an appendix for the presentation.
- Do not write your names or any identifying information on the slides.
- The easiest set up for this presentation is four slides each presenting one of your four tables.
- It is your responsibility to bring your presentation on a USB drive when coming to the oral exam.

Your task

In this final project, you will investigate the following research question:

Does political representation affect firms' decisions to pollute?

Do politicians' beliefs and preferences affect firm decisions? Anecdotal evidence suggests that it does. Amazon Inc. cited opposition of local politicians such as Representative Alexandria Ocasio-Cortez when announcing its decision to cancel its planned second headquarter location in New York.[1]

In this exam, we will investigate the causal link between political partisanship (Republican versus Democrat) and industrial pollution.[2] It is well known, that Republicans and Democrats have different preferences regarding pollution. For example, Republicans tend to vote against environmental bills, while Democrats tend to support them.

Some information about U.S. House of Representatives elections: Every even-numbered year on the first Tuesday after the first Monday in November, there are U.S. House of Representatives elections in the 435 congressional districts across each of the 50 U.S. states. The elected officials represent their congressional district from early January the subsequent year for a two year terms. For example, the election held on the 3th of November 2020 determined the Representatives that serve from 3rd of January 2021 to 2023, referred to as the 117th congress. In this exam, please only use the 107th to 116th congress. Also, we disregard special elections which are held off-cycle if a Representative unexpectedly leaves office.

For more information on elections in the U.S House of Representatives, please consider reading point 2.4 of following Wikipedia page LINK. See this LINK for a map pf congressional districts in 2020.

Some information on pollution data: Data on industrial pollution in the U.S. is obtained through the Toxic Release Inventory (TRI) database produced by the U.S. Environmental Protection Agency (EPA). For each chemical covered by the TRI program, facilities have to report the annual number of pounds released, as well as other plant information, such as location and a measure of output. For this exam, please focus only on chemicals that are regulated under the Clean Air Act.

Please match elections and pollution so that pollution reported for year t is associated with the political representative of the congressional district in power at time t (not elected in year t). An easy way to implement this is by constructing a variable that captures the congress in a given reporting year and use this variable to join the two data sets.

Table 1: Summary Stats

- Report statistics for all variables that are relevant when assessing the regression output presented later.

Table 2: Does political partisanship affect pollution

- Please regress pollution on partisanship.
- Please address following questions in your model as well as writing:

[1] For more informatio see this [LINK](https://www.cnbc.com/2019/02/14/ocasio-cortez-amazon-cancelling-nyc-hq-proves-anything-is-possible.html)

[2] There are more parties than just Republican and Demcrat in the data provided for this exam. For the purpose of this exam, you can study political partisanship from a narrow angle, namely as whether a Republican or a Democrat is in power.

  - Are detected changes because firms change output and hence pollution, or do firms change pollution for a given output level?
  - You want to study variation of pollution within facility-chemical level. You do not want to focus on variation in pollution between facilities.
  - Specific states experience unobserved changes around the same period. Address in the model and/or writing.
  - You would like to include facility-time fixed effects. Please include them or explain why you cannot/would not do this?

  - Different firms might take the same operations decisions for a given chemical. In other words, errors would cluster on the dimension of chemicals. Correct using clustered standard errors.

- Please describe as precisely as possible whether you are concerned about a bias.

Table 3: Causality

- Here is an argument on how to deal with a potential biases: If one limits the sample to elections that were just won marginally (only a difference of 5% between the votes for winning candidate versus runner-up[3]), the regression outcome could not have been predicted. Hence the winner and his/her party can be considered as if random.

Argue why this approach can help you in identifying an unbiased effect. Implement it.

Table 4: Are firms shifting pollution from Democratic to Republican districts, or vice versa?

- Until now, you have studied pollution levels. However, firms might decide to shift pollution from one area to another, rather than increasing the level of overall pollution.
- This argument would predict that the change in pollution explained by political partisanship depends on how many facilities a firm has in Democratic/Republic districts. For example, a firm that has many facilities in Republican congressional districts might not increase its pollution too much in a district with a newly elected Republican representative. In contrast, a firm that has most facilities in Democratic districts might increase the pollution substantially in a district with a newly elected Republican representative.

Please construct a regression model that addresses whether firms shift pollution as outlined above.

Data

- Data on elections in the House of Representative: Each observation is the election outcome for a candidate running in an election for a seat in the House of Representatives. The file is of the type `.tab` which indicates tab delimited. You can read the file with `read_delim()` from the `tidyverse`-package. Following variables are included:
  - year: election year
  - state: name of the state
  - state_po: abbreviation of the state
  - state_fips: abbreviation of the state
  - state_cen: (yet another) abbreviation of the state
  - state_ic: (yet another) abbreviation of the state
  - office: political office
  - district: congressional district

[3]For example, in the 2016 election (115th congress), candidate Ruben Kihuen, a Democrat, received 128.985 votes, 48.5%. Runner-up was candidate Cresent Hardy, a Republican, receiving 118.328 votes, 44.5%. Hence the vote margin is 48.5% - 44.5% = 4%.

  - stage: primary of general election
  - runoff: runoff election
  - special: special elections
  - candidate: candidate name
  - party: party of the candidate
  - writein: write in candidate
  - candidatevotes: votes the candidate receives
  - totalvotes: total amount of votes
  - Please not consider the remaining variables
- Data on pollution: Each observation is pollution (into the air) per facility and chemical. Following variables are included:
  - reportingYear: calender year
  - facilityID: ID of the facility (e.g. factory)
  - facilityCounty: county where the facility is
  - facilityState: state where the facility is
  - facilityZIP: zip code where the facility is
  - dbnrParent: firm identifier
  - chemicalID: chemical name
  - chemicalName: chemidal name
  - unitOfMeasurement: unit output is measured
  - cleanAirActIndicator: chemical is regulated under the Clean Air Act
  - emissionAir: emissions released into the air
  - prodIndexNormLevel: measure of production. This variable is normalized. So you cannot compare production levels to other facilities. The measure is only interpretable in the time series of a given facility.
  - prodIndexChange: change relative to last year in the output
- Data on facility location: This data allows you to map facilities to elections. Following variables are included:
  - facilityID: ID of the facility (e.g. factory)
  - statename: name of the U.S. state
  - district: congressional district (note that each state has a district 001)
  - congress: congress. Due to gerrymandering, a facility might be in a congressional district for a given year and another in a different year.
  - stateAbb: abbreviation of U.S. state

Additional guidance

- Do not be discouraged if data is fairly messy, it usually is.
- Try to understand the data before you decide on an econometric approach.
- You need to make informed and motivated choices for variables and models used. Use economic intuition rather than maximizing statistical significance or R2.
- You do not need to any gather additional data.
- Using code found on the internet is allowed, as long as you reference the sources correctly.
- Plagiarism in any way is not allowed.
- If you have questions during the exam, do not visit my office but write an email to me. If I answer your request, I will publish my comment on Canvas so that information is public and distributed fairly.
- Good luck!