CAUSAL INFERENCE IN NLP FOR ECONOMICS RESEARCH PROJECT

PITCH, PROPOSAL, FINAL PRESENTATION, PAPER & RESPONSE PRESENTATION

(MSc in Economics, WiSo Universität Hamburg)

Summer Semester 2022

Instructor : Dr. Huyen Nguyen
Website: http://huyenttnguyen.com
E-mail: huyen.nguyen-1@uni-hamburg.de

Task & percentage of course grade	Due date
Paper presentation (15%)	Chosen date, from Week 5 – Week 13 (detailed schedule below & in Syllabus)
Peer review presentation (15%)	Chosen date, from Week 5 – Week 13 (detailed schedule below & in Syllabus)
Midterm pitch recording (max 5'/team) (15%)	Due: 23.59 Tuesday 17.05 Upload to Open Olat/Pitches & Research Design Please upload the midterm pitch recordings and 1-page research design of your t following names and format: TeamNumber_TeamName_PITCH. mp4/wav TeamNumber_TeamName_DESIGN.pdf Each team must submit ONCE only for the pitch and the research design.
Midterm research proposal (max 1 page) (15%)	Same as the pitch recording Due: 23.59 Tuesday 17.05
Final oral presentation (max 25'/team) & slides (40%)	Presentation Date: Friday 15.07 Slides due on 23.59 Friday 08.07
BONUS: Proof-of-concept analysis plan (upgrade of max 0.7 on final grade)	Same date as final slides Due: 23.59 Friday 08.07

Paper Presentation (15%) and Peer Response Presentation (15%)

Throughout the course, you will learn to critically present research articles relevant to your research interests, as well as constructively evaluating the presentation of your classmates on a specific research articles to generate discussions in class. The most important criterion for both presentations is: *Be actionable. Give the author(s) a specific action they can do to implement your feedback.*

You will do these presentations in your chosen **groups of 3 students**, with a total of 8 to 9 groups for our entire class. Please make sure to get to know your classmates and their research interests on the Slack forum, and get yourself to a group by signing up on the following Google Sheet (also available on the #group-mix-and-match channel in Slack):

https://docs.google.com/spreadsheets/d/1EEUJp1Srz05BwEUH2u3CbYwMykmqEUhZdn2SqrmeaRs/edit?usp=sharing

These presentations will start from Week 5 onwards. Specifically, on the lecture days, assigned groups X needs to prepare a 30' mini lecture/presentation on the required readings (research paper(s) applying text in a causal context in real life); whereas assigned group Y will need to prepare a critique discussion of group X presentation (30'), along with relevant questions and discussions of the readings (30'). The group pair X-Y is as follows.

WEEK	Presenting Group X	Response Group Y
Week 5 (03.05)	G1	G8
Week 6 (10.05)	G2	G7
Week 7 (17.05)	G3	G6
Week 8 (31.05)	Group Research Project Consultation + (possible extra group presentations, in case we have > 24 registered students)	
Week 9 (07.06)	G4	G1
Week 10 (14.06)	G5	G2
Week 11 (21.06)	G6	G3
Week 12 (28.06)	G7	G4
Week 13 (05.07)	G8	G5

Some practical tips for you on how to read a research paper productively in social sciences: https://www.icpsr.umich.edu/files/instructors/How_to_Read_a_Journal_Article.pdf

(1) Paper presentation (30')

Imagine you are the authors of this paper and you are presenting this work in an international conference among peer scientists. Your task to is present this paper in an exciting, clear and concise way such that it sparks the interests of the audience to check out your paper. Therefore, you paper presentation should include:

(i) Motivation & research questions
(ii) Key finding Overview
(iii) Data & sources
(iv) Empirical strategy/Experiment design & assumptions
(v) Results
(vi) Conclusion & Future research avenues

(2) Peer Response presentation (30')

This presentation should consist of two parts:

(i) Constructive feedback on the presentation of your classmates
Make sure you find positive feedback you can give to your peers, with respect to their presentation styles (e.g. How did they motivate the paper? Are the slides structured clearly and cohesively? Are the graphics and slides supporting the key message they want to relay to the class?) and their analysis of the paper (e.g. Is the summary fair and sufficient? In what way? Did they explain the key theory/application well? Have they made any critical

(ii) Critical evaluation of the paper
Imagining you are the editor of a journal that the authors submit the paper to. Carefully summarize the paper's main contribution and evaluate the sensibility of the data, the empirical analysis and the main results.

Actionable points of to improve the paper should be provided, in relation to any other related scientific papers/ real-life applications that you can find to justify such suggestions.

Research Project: Pitch (15%), Proposal (15%) and Final Presentation (40%)

You research project can be anything from:
(1) a replication exercise of the latest research papers
(2) an improvement of the empirical techniques the papers use
(3) answering a new research question using existing/self-sourced datasets.

The main requirement is that the project should *apply at least two causal inference tools/framework* covered in class.

Overall, the research project presentation will be judged on the basis of its novelty (how new/exciting is it?) and applicability (how feasible is it in reality?).

Inspiration for interesting and relevant research projects in social sciences can be found in the course main reading list:
https://github.com/causaltext/causal-text-papers

Data source inspirations?

You can choose one of the following options (Whichever your option is, do check and decide EARLY ENOUGH IN ADVANCE)

1) Use your own procured data sources
2) Check one of the databases below:
   - List of interesting database
   - U.S. Congressional Record, https://data.stanford.edu/congress_text
   - CourtListener, https://github.com/idc9/law-net
   - Chris Bail's list of digital data sets
   - Data is plural spreadsheet
   - Google toolbox: search engine for data
   - Liste: https://github.com/awesomedata/awesome-public-datasets
   - Dataverse: data repository from Harvard
   - APIs list: https://github.com/public-apis/public-apis or https://apilist.fun/ or https://www.programmableweb.com/apis/directory
   - Datasets on Kaggle
   - Project ideas

3) Ask the instructor for sample data sources

Midterm research pitch (15%) & research design report (15%)

Preliminary topic choice & consultation session registration (for Week 8) should be done at the latest by **Week 5, Tuesday 03.05.**

By **Week 7**, 23.59 on Tuesday 17.05, in groups of three students (group matching coordination via Slack channel must be finished by Week 3, Tuesday 19.04), please submit the following items:

1) A 5-min pitch video recording of your research idea in .mp4 format

2) A 1-page detailed research design of your group (motivation, related literature, data, and approach)

Each task contributes equally to the 30% grade component you receive for the course i.e. 15% for the 5-min video recording, 15% for the research design.

1) Research pitch

The 5-min pitch video recording should serve as an inspiring pitch of ALL group members about your research project. Pitch it as if your audience is a to a non-technical decision-maker, such that he/she/they can understand the key ideas of and get excited about your research project.

NOTE: 5-minute means **exactly** 5 minutes, not one or a few seconds more.

Should you upload anything more than exactly 5 minutes, your team will be disqualified and notified, once the instructor has checked if your video meets the time limit after the deadline. You will then be required to reupload, which means that your team will incur a grade loss because of late submission in this task.

Regarding the technology you can use, some familiar options include PowerPoint (instructions here)
Zoom (instructions). Alternatively, you can play around with the professional pitch narrative and slide decks that are used to pitch to company investors.
https://www.pitchtape.com/
https://www.loom.com/blog/remote-pitch-deck-best-practices

Whichever technology you use, make sure your group names, individual names are visible on the first slide, while the rest of the slides, voices, faces are clear and visible throughout the entire 5 minutes.

2) Research design report

The 1-page causal inference research design report should explain the motivation, research question, related literature, data, hypotheses, expected results and desired contribution to a non-technical decision-maker.

Below is a detailed sample guideline:

Research Question

- *What is your research question? (i.e. what do you causally try to explain?)*
- *Discuss the motivation for the project -- why is it interesting or necessary?*
- *Briefly summarize at least 3 previous related research papers.*

Dataset/Corpus

- *Describe the corpus or dataset you will use for your project -- where you obtained it, or where you will obtain it from.*
- *How will the corpus help you answer your research question?*
- *Describe how you will organize and label the documents, and outline how you will merge them with associated metadata (if applicable).*
- *Perform a preliminary inspection of your corpus and explain what preprocessing/cleaning steps may be needed.*
- *If applicable, report some summary statistics about the data.*

Methods

- *Outline your approach. What tools will you use to solve your research problem?*
- *E.g., for machine learning: What model will you use? How will you evaluate and validate your model? How will you interpret or explain your model?*
- *Write out your regression equations.*
- *What are the empirical assumptions?*
- *How will you assess robustness of your results?*

Timeline

Provide a weekly/bi-weekly timeline for your research plan until the end of June 2022.

Final presentation (25') (40%)

The final oral presentation should be using slides and *last MAX 25 minutes*. Please submitted your presentation slides in the Open Olat folder **by 23.59 on Friday 08.07**.

One submission per group only, but do not forget to put the names of all group members in the first slide.

On your assigned oral exam date and time, you can present your work to the instructor (via Zoom/in person, depending on the situation we will be in at that point in time), followed by a 5-minute Q&A session.

Grading components:

- Individual:
  - Presentation skill: 10%
- Group
  - Quality of the material (slides): 10%
  - Summary: 10%
  - Critical assessment: 15% (i.e. uncertainty, implications, and limitations of your work)
  - Suggestion: 5% (e.g. way(s) to improve the methodology, extend data limitations, test another mechanism, address a complementary research question)

BONUS: Proof-of-concept analysis plan (upgrade of max 0.7 on the final grade)

In addition to the Final Presentation, you have the chance to obtain an upgrade of max 0.7 on your final grade by submitting a proof-of-concept analysis plan. This plan includes:

(1) a detailed econometric setting designs and assumptions (and any relevant justifications backed by other research papers)
(2) (sample) data analysis and preliminary results
(3) (if available)
(4) any coding files (in -py/.ipynb/.do files or .dta) you use to generate the results in your oral presentation
(5) a 1-page README file (in .txt) with any further step-by-step miscellaneous explanations of how to load the data, Python/Stata version and what packages to use to run your files.

Similar to the presentation slides, this *.zip file* also needs to be uploaded to the dedicated Open Olat folder **by 23.59 on Friday 08.07**.

Inspiration can be found here. As for the coding practice, please follow the below guidelines on writing efficient and clear codes in Python and Stata.

https://docs.python-guide.org/writing/style/

https://data-flair.training/blogs/python-best-practices/

https://michaelshill.net/2015/07/31/in-stata-coding-style-is-the-essential/