

Project Guidelines: Second Submission

Applied Causal Machine Learning

I have discussed the details of Project 2 before and after each class and in our office hour discussions. Additionally, the feedback provided on your Project 1 is there to support and guide you as you work on this assignment. There should be nothing new in this PDF; it is a recap of what was discussed before and acts as a reference. If you have any questions or need clarification, please don't hesitate to ask! Make sure to bring up any concerns well in advance of the deadline. I'm happy to help, and office hours are a great time to address any uncertainties. If we don't get questions from you, we assume you are clear about expectations!

I want to take a moment to highlight how impactful this paper can be for your future. You're working with real-world data and **cutting-edge causal ML models** *that are still rare in economics publications*. This makes you pioneers in this space, and the work you're doing now will likely be seen as groundbreaking when you reflect on it in the future. Trust me, potential employers love seeing this kind of innovative and practical work—so your paper isn't just an academic exercise; it's a career booster. Think of it as building a treasure chest of skills—you may not realize its full value yet, but future you will be saying, 'Wow, I was awesome back then too!'

Also, remember: you are awesome now. The effort, creativity, and dedication you're putting into this project are inspiring. And if this feels challenging, just remember: you are creating something brand new! Keep up the amazing work—you've got this!

Below, I added a lot of detailed explanations for each section with some boxes containing the summaries for each part. Feel free to refer to each section when you need clarifications.

1 House Keeping

Word Limit and Quality of Writing

- We have observed that some submissions are too lengthy. Please make sure your writing does not exceed a maximum of 7000 words.
- **Using AI for text generation is not allowed for your ‘final project’** and is considered an academic integrity issue. You can use it to get help with grammar, idea generation, code, etc., however generating blocks of code is not allowed. If you have GPT generated code in previous projects and reports, please edit them for your final project.
- If further explanation is necessary, use the appendix. However, grading will primarily focus on the content within your main paper, not the appendix. Do the cost benefit!
- Aim for concise and efficient writing. Re-read your paper multiple times to make sure it is clear, well-organized, and free of unnecessary sentences.
- Eliminate redundant content to enhance readability and maintain focus on your *core arguments and contributions*.
- There should not be any bullet points in your paper, no tables that span more than a page, no screenshot outputs.
- Please submit your PDF and your code to Crowdmark by the deadline.

Overview

This project builds upon the work completed in Project 1 and all the reports submitted so far. You will expand the depth and complexity of your analysis by incorporating new models and addressing prior feedback that we gave to you in your project 1 and during office hours. In Project 2, you are expected to enhance your initial work thorough in depth data exploration, a well-written literature review, and clear interpretation of results. ***This project is to be formatted as a professional journal paper in LaTeX.*** Your submission should adhere to a structured, formal academic tone with clear, meaningful insights derived from a variety of models discussed in class.

Key requirements include running and interpreting the minimum specified models, revising based on prior feedback, and following academic standards for format, tone, and rigor.

While these guidelines outline the minimum requirements for an acceptable paper, you are expected to **go beyond** these and tailor their work to fully capture the unique potential of their individual research questions and datasets. Each paper will differ based on its scope, data, and specific research angle, so you should consider how to extend your analysis, interpret null (or unexpected) results, and add additional techniques as relevant to your study. Thoughtful expansions— through a deeper literature review, new methodological approaches (not covered in class), or innovative visualizations—will

enhance the depth of your work. In short, these guidelines are a starting point; high-quality papers will build upon them to create original research.

You are encouraged to create your own section titles that best fit your paper's content and structure, rather than strictly following the ones suggested here. Tailoring section titles to your paper's unique focus can make it clearer and more engaging for readers. **It's also important that you thoroughly read any paper you decide to cite.** *Citing sources without understanding them not only risks academic integrity but also weakens your analysis.* By reading each source, you write your literature review in an accurate and meaningful way and you ground your work in the right branch of research.

Important Notes

1. Make sure all tables and figures are embedded directly from your code. Screenshots are not acceptable.
2. Incorporate all models discussed up to the week before the deadline. Extensions of the deadline may require additional models covered in the new time frame. For instance, if we extend the deadline by one week, the models covered in that week will be required. Deep IV is not required but encouraged if you have an instrument.
3. Cite at least 10-20 relevant studies in your literature review, and ensure citations are accurate and follow a consistent style.
4. **Important:** all papers will be run through one or more plagiarism checker platforms as part of the department's policy.

1. Introduction

Key Points: Introduction Section

- Clearly define the research question or hypothesis in an economically relevant context. Explain its significance and originality.
- Connect the research question to existing literature, identifying gaps or unresolved issues to justify the study's contribution.
- Conclude with a structured outline of the paper, offering a preview of key findings or anticipated contributions.
- Maintain a logical flow to set the tone for the analysis. Keep in mind the word limit!

The introduction section of an economics paper serves as the most important part of the paper that introduces the reader to the research question, its importance, and the paper's overarching goals. It should begin with a clear statement of the research question or hypothesis, positioned in an economically relevant context that establishes why the question matters. This section should capture the reader's interest by framing the question as addressing a critical issue or gap within the field. By doing so, the introduction sets the tone for the analysis and clarifies what the study aims to achieve.

You should also establish a logical pathway for the arguments and findings that will follow in the next sections.

In addition to presenting the research question, the introduction should provide a brief overview of the relevant literature. This means connecting the research question to existing studies, theories, or frameworks, demonstrating both the context of the question and its novelty within the field. A well-rounded literature review highlights influential works, identifies gaps or unresolved issues, and justifies why this study is necessary and unique. The goal is not to conduct an exhaustive review here but to strategically situate the study within the broader discourse. This contextualization strengthens the study's credibility and positions it as a meaningful contribution to the economic literature.

Finally, the introduction should conclude with a concise outline of the paper's structure and a preview of key findings or anticipated contributions. This brief road map guides the reader on what to expect and reinforces the central themes that the paper will explore. A high-quality introduction is structured yet open-ended enough to prompt critical thought, with clear, logical flow and academically grounded language. To achieve full marks, an introduction should address several important points:

- Defining a relevant research question
- Emphasizing its economic significance (motivation)
- Contextualizing it within the literature
- Offering a well-organized preview of the paper's approach, data, and findings

Important: the following are what goes into an introduction, they are not subsections.

Research Question

Clearly state the specific research question or hypothesis your analysis will address. This question should be focused and suitable for causal analysis methods. Make sure that the question is framed in an economically relevant context and addresses a meaningful economic issue. Sometimes you just need to rephrase your question for an economic paper. Feel free to talk to me if you need help with writing this part.

A strong research question is at the heart of any good economics paper! It is the focal point that guides the entire analysis. In economics, a research question should address a specific (not too wide, not too narrow), relevant issue within the field, ideally one that contributes new insights to existing knowledge. It should be precise and well-defined, yet open enough to allow for meaningful exploration and analysis. Avoid questions that are too broad or vague, as these can make it difficult to conduct focused research and provide clear answers. If you frame your question too broadly, your paper may not be able to answer it. **The reader should feel like she has a pretty good idea about the answer to your research question after reading your paper!**

When presenting the research question in a paper, it's important to communicate why it is worth investigating. This involves explaining the broader economic or social significance of the question, helping readers understand its importance and relevance. For instance, the introduction should frame the question within current economic challenges, policy debates, or theoretical gaps, making it clear why this particular topic matters. A well-presented research question often addresses **the “why” and “how”** of the issue at

hand. Why does this topic matter in today's economic landscape, and how does it affect real-world outcomes? By addressing these points, you build a compelling case for the research's value.

Finally, a good research question should connect smoothly with the methods and analysis sections of the paper, acting as a road map for what's to come. The question should suggest an analytical approach that will yield actionable insights or contribute to economic understanding. In other words, a well-formulated question naturally leads to a logical analysis framework, whether through econometric methods, causal inference models, or theoretical simulations. In presenting the question, consider stating it clearly at the end of the introduction, setting up the reader's expectations for the study's direction and objectives. This clarity helps ensure that every part of the paper—from literature review to conclusion—is aligned with and contributes to answering the central research question.

Importance and Motivation

Discuss why your research question is important within the broader context of economics. Explain how it relates to existing economic challenges, trends, or theoretical issues. This section should clarify why your research question is relevant for current economic research and how answering it contributes to a better understanding of economic phenomena.

When writing about the importance of a research question within the broader context of economics, it's important to start by situating the question within current economic challenges or significant issues. This could involve **discussing the practical, policy, or societal impacts that the topic touches upon**. For instance, if the research question addresses labor markets, it may be valuable to frame the question in the context of rising wage inequality or employment instability. Similarly, if the focus is on housing markets, relate the question to pressing issues like housing affordability, urbanization, or housing bubbles. By connecting the question to a specific, tangible problem or trend, you demonstrate the topic's urgency and why it deserves attention from an economic perspective. This is an easy yet important way to motivate your paper. I should also say that you don't have to necessarily have policy implications but having them makes your paper more valuable in applied topics (non-theoretical).

Next, linking the research question to relevant economic theories, models, or frameworks provides an academic anchor. This involves showing how the question aligns with, challenges, or extends existing theories in the field. For example, a question about income distribution could draw on classical theories of inequality or modern perspectives on economic mobility. Such connections not only clarify the intellectual foundations of the question but also highlight the paper's potential to contribute new insights. Relating the question to established economic theories creates a bridge between the research and a wider scholarly conversation, demonstrating how the paper might refine or even re-frame economic understanding in that area.

Finally, it is important to emphasize the broader implications of answering this question and how it contributes to a more nuanced understanding of economic phenomena. Clarify how insights from this research could influence policy, improve economic forecasting, or offer guidance on economic planning and interventions. When this section clearly communicates the potential outcomes and contributions of the research, it provides readers with a sense of the study's purpose and relevance. Ultimately, the discussion should leave readers understanding not only why the question matters but also how the paper's

findings could deepen or broaden the field's knowledge on complex economic issues.

Literature Review

Conduct a thorough literature review that synthesizes at least ten to twenty relevant studies. Discuss how each study informs your approach and analysis. Identify gaps or limitations in prior research that your project will address. This section should justify your methodology, models, and approach within the context of existing literature, highlighting how your work expands or improves upon past studies. **Your lit review should not sound like a list of papers but more like a story about past work that justifies your choices.** You can choose whether you want to cite all papers in the intro or spread them in the paper in different discussions, this depends on your writing style. Again, citing a paper that you have not read poses a plagiarism risk.

One more time: a literature review in economics provides an organized summary of past research that is directly relevant to the paper's research question. This section should not merely list studies; instead, it should synthesize findings, methods, and gaps within the existing body of work. By presenting a well-structured narrative, a good literature review helps readers understand the foundation of the research question and its place within the broader economic field. This review should outline the key contributions and limitations of past studies, emphasizing what remains unresolved or open to further exploration.

An effective literature review goes beyond summarizing each paper individually. It identifies common themes, theoretical debates, or methodological approaches shared across studies, explaining how these relate to the current research question. *It's equally important to highlight any differences or controversies in the literature, as this can help clarify the unique contribution that the current paper seeks to make.* **Students should connect each cited study directly to their own research**, showing why it was relevant to include and how it informs the approach they will take. The literature review also justifies the choice of methods, data, or theoretical perspectives by demonstrating how these elements are informed by previous work.

In grading a literature review, we look for depth, clarity, and coherence. A high-quality review will be well-organized, connecting papers in a logical flow that provides context for the research question. We expect students to have thoroughly read and understood each cited paper; citing a study without reading it or relying solely on summaries found in other papers is not acceptable and violates academic integrity. Marks will be deducted for superficial summaries, disconnected studies, or signs that the student has not genuinely engaged with the literature. A strong literature review is one that builds a clear, reasoned case for the research question, demonstrating both the relevance and originality of the student's own work within the economic discipline.

Contribution and Paper Structure

Using your results and approach, provide a concise overview of how your paper contributes to the field of economics and applied causal machine learning. Outline the structure of the paper, giving readers a road map of what to expect. This section should give readers a clear sense of the paper's focus, approach, and anticipated findings.

In the contribution section, you should clearly and concisely explain how your paper's findings add value to the existing body of economic research. This section should communicate what is unique or innovative about your approach, data, or findings and clarify

how these aspects address gaps identified in the literature review. Contributions can range from offering new insights (finding) into an established debate, or presenting fresh data that provides a different perspective on the topic. A strong contribution section helps readers understand why this research is needed and how it advances the field.

An effective contribution section will go beyond general statements by specifying the distinct elements of the research that make it valuable. If the research uses unique or updated data, the contribution section should highlight how this new information enhances the robustness or relevance of the findings. Each point made in the contribution section should directly support the originality and impact of the research, linking back to the gaps or limitations in the literature discussed earlier.

When grading the contribution section, we look for clarity, specificity, and relevance. A high-quality contribution section will be detailed and convincing, showing a clear understanding of how the research fits into the broader economic conversation. *General statements without supporting details or claims that are not fully explained will lead to mark deductions.* We also look for alignment between the contribution and the paper's actual findings and methods; contributions that don't match the research's content may indicate a lack of understanding or weak analysis. Ultimately, the contribution section should convey a well-reasoned and genuine advancement of knowledge in the economic field.

2. Data and Summary Statistics

Key Points: Data and Summary Statistics

- Clearly describe the dataset, including source, scope, sample size, time period, and geographic coverage. Address any data limitations and their implications.
- Define and explain key variables (outcome, explanatory, control) and their relevance to the research question. Writing should *not* be in list format.
- Present summary statistics with well-formatted tables, figures, map, etc that provide clear insights into the data structure and relevance to your research question.
- Your sum stat section should be enough to understand the paper. Each table and figure should add a new insight about your research question.
- Use visualizations (e.g., histograms, scatter plots, maps) to illustrate key patterns, make sure they are well-labeled and directly tied to the research question.
- Results with no interpretation/explanations should not be included in the paper.
- Optionally discuss data limitations; focus on how they affect the interpretation or validity of results, if relevant to your study.

The data section should provide a comprehensive description of your dataset, define key variables, and present summary statistics through meaningful tables, figures, and plots. This section establishes the foundation for your analysis by helping the reader

understand the data structure and limitations. *The story you tell with your summary statistics tables and figures in this section should be enough for the reader to understand the paper!*

Data Source and Description

Describe the source and scope of your dataset. Include details such as the dataset's level of observation, sample size, time period, and geographic coverage. Discuss any potential data limitations, such as missing data or potential biases, and their implications for your analysis. Are you overestimating or underestimating as a result of this limitation?

Key Variables and Definitions

In a story format define and describe your key variables, including outcome variables, explanatory variables, and control variables. Explain why each variable is important to your research question and how it contributes to the analysis. Make sure that variable definitions are precise and that their relevance to the study is clear. Do not make lists. This should be embedded in the paper; doesn't have to be a separate subsection if you don't see it fit.

Summary Statistics and Visualizations

Provide summary statistics for your key variables. Include tables **AND** figures that illustrate the distribution and characteristics of these variables. Graphical representations such as histograms, bar plots, and scatter plots should be well-labeled and help the reader understand the data's structure and relevance to the research question. Every table and figure should be accompanied by an explanatory narrative that guides the reader through the insights these visualizations provide.

- As always, tables and figures should be formatted as a journal paper and not in an arbitrary style.
- A table or visualization that is not relevant to the research question and/or is not well-explained will not receive any points.
- You should tell me what I learn from each table and plot. One easy way to make this section more intuitive and helpful is to think about the heterogeneities you want to talk about in the results section and visualize them in your summary statistics.

A summary statistics section in an economics paper provides an organized overview of the dataset and offers the readers a clear understanding of the variables, distributions, and **relationships that will be analyzed in the study**. This section typically begins with descriptive statistics, such as means, medians, standard deviations, and ranges for key variables, presented in well-formatted tables (not too long, not too short, focusing on more relevant variables).

Each table should be thoughtfully designed to emphasize important aspects of the data that are **relevant to the research question**, ensuring that readers can understand the dataset's scope and any important patterns. The goal is to set a foundation for the analysis, so **tables should not only be comprehensive but also purposeful**, focusing on the variables that drive the research question.

In addition to tables, a variety of visualizations—such as maps, histograms, scatter plots, box plots, and density plots—should be included to get deeper insights into the

data. These visualizations are most effective when they directly relate to the research question, showing relationships, trends, or distributions that add value to the analysis.

For instance, if the research question involves examining income distribution across regions, relevant visualizations might include histograms of income levels by region or density plots showing income disparities, or a heat-map depending on the geographical details. We will look for clarity, readability, and interpretability in all visualizations, as well as for their relevance to the research question. Random or poorly chosen visuals that do not add clear insights or support the research question will lead to mark deductions.

I recommend starting with pen and paper before coding. Draw the graph you have in mind and ask yourself: does this graph describe the point that I want to emphasize? Can this point be visualized better using a map? Do I need to complement this with a table (write the columns on the paper)? How can I visualize the subgroups?

Data Limitations

You can choose to have this discussion or not depending on how relevant the data limitations are to your research question and the validity of your results. Identify any limitations in the data that may affect the interpretation of your results. Discuss the potential impact of these limitations and how they may influence the validity or generalizability of your findings. Transparency about data quality and limitations is critical in academic research.

3. Model Analysis and Implementation

Key Points: Model Analysis and Implementation

- Clearly specify models, including assumptions, variables, and econometric structures. Articulate the purpose of each model and its contribution to addressing the research question. Keep in mind the word limit, be very concise.
- Detail the identification strategy for your causal models (e.g., IV, DID) and justify its use. Run the necessary robustness checks to validate findings with alternative specifications.
- Present results in clean, journal-quality tables (refer to resources we provided for the previous project). Interpret key coefficients, discuss statistical significance, and explain economic relevance. Avoid redundancy and make sure tables are concise and informative.
- **Provide a comparative narrative that connects results across models**, highlighting meaningful differences or similarities. Build a cohesive analysis that deepens understanding of the research question.

This section is the core of your analysis and should include multiple modeling approaches to address your research question.

Model Specification

Clearly specify each model you are implementing, including assumptions, variables, and the econometric structure in the context of your research question. **Important: don't write the models in a lecture note style, your models should be explained in the context of your paper and with your variables.**

Make sure each model's purpose and role in your analysis is clearly articulated, detailing how it contributes to answering your research question. Specify key parameters and settings used for each model, including tuning parameters if applicable. Be mindful of the word count. You can skip the parameter tuning if necessary.

Identification Strategy and Robustness Checks

Discuss the identification strategy you are employing for each causal model. For methods like Instrumental Variables (IV), Difference-in-Differences (DID), or other causal techniques, explain the underlying assumptions and justify their use in your context. For robustness, run alternative specifications and discuss how these additional tests validate the robustness of your findings. Most of robustness checks can go to appendix, however, testing the main assumptions such as parallel trends should be reported in the paper.

Presentation of Results and Interpretation

You are expected to run and interpret the following models as a minimum: OLS, IV or DID or another causal inference method used in the main paper, LASSO, RIDGE, all Tree-based models, Matching, IPW (Inverse Probability Weighting), DAGs (Directed Acyclic Graphs), Meta Learners, Doubly Robust methods, Double Machine Learning, Causal Forest, Deep IV (optional), and other models covered up to the project deadline.

For each model, present results in clean, well-formatted tables or graphs, ensuring they meet journal-quality standards (no screenshots, no python tables). Interpretation is key: provide a detailed explanation of important coefficient/results, its statistical significance if applicable, and its economic meaning. Discuss any patterns or insights revealed by the models and describe how each model helps answer the research question. Avoid redundancy; each result should add new information to your analysis. Long tables that span across multiple pages receive a deduction. Draw your tables using pen and paper before coding them and make sure the organization makes sense.

Important: when working on your literature review, you should study the formatting of the regression, summary statistics tables, and figures in published journal papers and use the same formatting in your paper. Poorly formatted tables (not in journal format) that present irrelevant information or lack important information such as number of observations will not receive marks.

Comparative Narrative and Analytical Insights

Provide a structured narrative that connects the findings from each model. Discuss how each model's results (or null results) build upon or contrast with other models in your analysis. Interpret any meaningful differences or similarities across models, explaining how they contribute to a cohesive understanding of your research question. Make sure that your analysis is comprehensive and logically structured to help readers through your reasoning and findings.

Pro tip: print your paper and grade it as a referee or read it as if you are not familiar with this topic, use a red pen to make the parts that are not clear or are not coherent. Edit. Repeat.

4. Discussion of Results

Key Points: Discussion of Results

- Summarize key findings across all models, connecting them to the research question and theoretical context. Interpret null or surprising results with explanations of model suitability.
- Compare findings to existing literature, highlighting how your results confirm, challenge, or extend current knowledge in the field.
- Discuss practical and policy implications where applicable, offering recommendations or insights that arise from your analysis.
- Provide a reflective and critical narrative to tie your findings to broader economic discussions and contributions. Do you have any policy recommendations? What did you learn in your detective work?

The discussion section should provide interpretation of your results, connecting them to the initial research question and the broader literature.

Important: you may not get significant results in all of the models you work with and that is expected since not all models are suitable for all causal settings. If you believe a particular model is not suitable for your setting, please interpret your results in that light. You should explain why you are getting surprising results or no results at all. Why is this model not appropriate for your data or causal structure?

Key Findings

Summarize the main findings across all models. Discuss how these results answer the research question and relate back to the theoretical context established in the introduction. If null results or opposite signs observed, interpret the why.

Implications for Literature and Theory

Compare your findings with those in the literature. Discuss how your results confirm, challenge, or extend existing knowledge on the topic. This section should demonstrate a thorough understanding of the theoretical and empirical landscape surrounding your research question.

Practical and Policy Implications (if Applicable)

Discuss the practical implications of your findings (if any). Highlight any potential policy recommendations, economic insights, or areas for applied work that emerge from your analysis. If you do have a policy recommendation in your paper, it should be thoughtful. Random recommendations that are not feasible or not rooted in your results reduces the paper's credibility.

5. Conclusion and Future Directions

Key Points: Conclusion

- Restate the research question and summarize key findings, emphasizing how they contribute to understanding the question and broader economic field.
- Avoid introducing new data or analysis; focus on reflecting on insights already discussed in the paper.
- Discuss limitations of the study and provide specific, relevant suggestions for future research directions.
- Make sure clarity and coherence, offering a concise yet impactful summary that reinforces the paper's contributions and encourages further inquiry.

A conclusion in an economics paper serves as the final summary of the study, reiterating key findings and reinforcing the research question's relevance. Unlike the introduction, which introduces the research question and builds a foundation, the conclusion should focus on closing the discussion by concisely summarizing the main insights gained. This section should begin by restating the research question and summarizing how the study's findings contribute to our understanding of it. In doing so, the conclusion helps the reader see the study's impact in a broader context, tying back to both the research question and the contributions discussed earlier. A strong conclusion offers a clear sense of closure, grounding the results within the broader field of economics.

An effective conclusion is concise but comprehensive, capturing the essence of the study without introducing new analysis, data, or information. This is crucial for maintaining the section's focus and purpose, as the conclusion is meant to reflect on findings rather than extend them. You should avoid adding additional insights, untested claims, or complex details that were not previously discussed in the analysis section. Instead, the goal is to reinforce the paper's primary takeaways, clarifying how they align with or differ from existing literature and theories. In this way, the conclusion solidifies the value of the paper's contributions and allows the readers to leave with a clear understanding of the research's core impact.

To strengthen the paper's relevance, the conclusion can also suggest directions for future research. Here, students can briefly discuss potential limitations of their study, such as data constraints, methodological trade-offs, or the scope of the research question. You can (don't have to) then suggest specific ways in which future work could address these limitations or expand upon the findings. This could include exploring alternative data sources, using different methodological approaches, or applying the research question to different economic contexts. In grading the conclusion, we will look for coherence, clarity, and adherence to the paper's established scope. Marks will be deducted for unfocused or overly detailed conclusions, vague future research suggestions (if this is added), or new information that doesn't belong in this section. A well-written conclusion should briefly reinforce the study's contributions, reflect on its impact, and encourage further inquiry in a relevant and meaningful way.

References

List all studies, datasets, and sources cited in your paper, following consistent citation formatting (APA or Chicago).

Appendix (Optional)

If needed, include supplementary materials that support your analysis but are not important to the main narrative.

Formatting

Your submission must adhere to professional academic standards for formatting and style. Pay attention to details in each of the following aspects:

- **Tables and Figures:** All tables and figures must be embedded directly into the LaTeX document and formatted according to journal standards. You might find this latex template helpful: <https://www.aeaweb.org/journals/templates>
- **Writing Style:** Maintain a formal, academic tone throughout your paper. Avoid informal language and make sure clarity and conciseness.
- **Citations and References:** Use a consistent citation style (APA or Chicago) and list all references at the end of the paper.

Grading Expectations and Requirements for Model Presentation and Interpretation

For each of the following models, students are expected to provide clear and thorough interpretations, showing effort and a solid understanding of each model's purpose and output. If models are presented without sufficient interpretation or appear to be run without understanding, points will be deducted. Each model is an opportunity to explore various aspects of the research question, and students should use these methods thoughtfully to demonstrate a comprehensive economic analysis.

You do not need to explain the theories in a lecture style. Your readers can search and learn about each model if needed, however, explaining your model in the context of the method used is necessary. For instance if you are talking about the moment function or an objective function of a method, you can talk about it in terms of your variables. Avoid writing in lecture style. Also, there is a 7000 word limit that comes with a cost benefit analysis!

1. Ordinary Least Squares (OLS)

We usually report the simple OLS model before diving into our causal models. This OLS serves as a benchmark to help us see the direction of the bias due to confounders which will be fixed in the causal models.

- **Presentation Expectations:** Results from OLS should be displayed in a well-organized table, including coefficients, standard errors, and p-values. Interpretation should cover direction, magnitude, and significance of each key variable and relate findings to the research question.
- **Grading Criteria:** Tables must be clear, and interpretations should connect findings to the research question. Marks will be deducted for superficial interpretation or lack of relevance.

Note: you can choose to present your OLS results as one or more columns in your IV or DID tables. It is a design choice but they should be reported before any other model. If you do IV or 2SLS, you must report the first stage (in any paper).

2. Instrumental Variables (IV), or Difference-in-Differences (DID), or Another Causal Method

Present traditional OLS/IV/matching estimators first and then precisely specify your treatment effect estimand and your identifying assumptions.

- **Presentation Expectations:** For IV, describe and justify the instrument's relevance and validity. For DID, define treatment and control groups and discuss assumptions like common trends. Results should clearly reflect causal impacts. Use tables or figures to provide evidence for assumptions.
- **Grading Criteria:** Understanding of causal inference is key. Points will be deducted if causal methods or assumptions are not clearly explained or tables are not thoughtfully created. I should understand your paper by looking at this table without needing any explanations about the variables or the story.

3. LASSO

- **Presentation Expectations:** Present selected variables and their coefficients after running LASSO, explaining variable selection and relevance. Adjustments in tuning parameters (e.g., cross-validation) should be noted. You can present these results in tables or figures, your choice.
- **Grading Criteria:** Proper use of LASSO to select variables is important, with deductions for lack of justification or inadequate model tuning discussion (we talked about lambda in class). What are you learning from your shrinkage model?

4. Ridge Regression

- **Presentation Expectations:** Report coefficients, regularization parameter values, and discuss multicollinearity issues if relevant.
- **Grading Criteria:** Thoughtful interpretation of how Ridge affects estimates relative to OLS is required. Points will be deducted for poor explanation of Ridge's purpose. Compare your findings with LASSO and OLS.

5. Tree-Based Models (Decision Trees, Random Forests, Gradient Boosting)

- **Presentation Expectations:** Include variable importance scores, accuracy/error metrics, and visual aids like feature importance plots and tree presentations.
- **Grading Criteria:** Clear interpretation, use of visuals, and comparisons of model findings are required. Points will be deducted if explanations are not clear or visuals are missing. Keep concise, there is a word limit.

6. Matching Methods

- **Presentation Expectations:** Show balance tables pre- and post-matching, and discuss the matched sample's results in relation to the research question. Test your assumptions and provide tables/figures.
- **Grading Criteria:** Quality of balance tables and matched sample interpretation are critical. Deductions for poorly explained matching assumptions or lack of comparison with other findings..

7. Inverse Probability Weighting (IPW)

- **Presentation Expectations:** Present results showing IPW's impact on covariate balancing and treatment effect estimation.
- **Grading Criteria:** Correct application and understanding of IPW are important. Points deducted for insufficient IPW interpretation or misuse.

8. Directed Acyclic Graphs (DAGs)

You can decide to present your DAG model in the identification section after your OLS section. In general, you can choose to move these models around or group them as you see fit for your paper, but all of the models that we covered in class should exist in your paper unless discussed with me for cases where it is not possible to do so given data and scope.

- **Presentation Expectations:** Present a DAG modeling assumed causal relationships, explaining connections. What's the identification strategy?
- **Grading Criteria:** Accurate DAG representation and reasoning for each relationship. Deductions for unclear or unsupported DAGs.

9. Meta Learners (T-learner, S-learner, X-learner)

- **Presentation Expectations:** Present results showing treatment effect heterogeneity across subgroups. Explain the choice of meta-learner.
- **Grading Criteria:** Quality of subgroup analysis and interpretation of heterogeneous effects are key. Deductions for lack of explanation, lack of visualization of the heterogeneities, or misinterpretation of the models.

10. Doubly Robust Methods

- **Presentation Expectations:** Present doubly robust estimators, discussing outcome regression and propensity score effects.
- **Grading Criteria:** Clear understanding of doubly robust benefits and reliability. Points deducted for superficial explanation or misuse. Please compare your results with the appropriate models.

11. Double Machine Learning

- **Presentation Expectations:** Report results from double machine learning, specifying confounder control in high-dimensional settings. Are you running it linearly or non-linearly? Why?
- **Grading Criteria:** Clear interpretation and application are required. Deductions for glossing over complex assumptions, lack of comparison, or missing results.

12. Causal Forest

- **Presentation Expectations:** Present treatment effect heterogeneity and variable importance, with visualizations like treatment effect plots or reporting your first tree.
- **Grading Criteria:** Interpretation of heterogeneity and quality of visuals are key. Deductions for inadequate explanations or missing visuals. Compare the appropriate causal ML models with each other.

13. Deep IV (optional)

This section is optional especially since it requires an IV.

- **Presentation Expectations:** Present estimated treatment effects from Deep IV and discuss how it addresses endogeneity in non-linear data.
- **Grading Criteria:** Understanding of Deep IV's benefits and assumptions is critical.

Closing

I have truly enjoyed teaching this course and working with each of you. It has been a privilege to witness your creativity, hard work, and intellectual curiosity throughout the semester. You are a group of incredibly bright and talented students, and it has been inspiring to see how you tackle challenges with determination and ingenuity. Each of you brings a unique perspective to the table, and that has made this experience both enriching and rewarding for me as your instructor.

As you wrap up this project, I hope you recognize the immense progress you've made—not just in mastering complex concepts but in pushing the boundaries of your own potential. Remember, every great economist, researcher, or leader started with moments just like this: putting in the effort, learning from mistakes, and building skills

that will shape their future. Whatever path you choose after this course, know that you are capable of achieving incredible things. Believe in yourself, keep growing, and never hesitate to take on challenges that push you out of your comfort zone. You are resilient, talented, and, most importantly, you have what it takes to make a meaningful impact in the world. I am so proud of each and every one of you, and I can't wait to see all that you will accomplish in the years ahead.

Stay connected!

Cheers,

Nazanin