Kuehn
Economics 440
Fall 2024

Final Paper Guidelines

DUE DATE: Friday, December 6, 5:00 PM

Objective: The objective of this final research paper is to apply the econometric skills and techniques learned in this course to a real-world dataset. You will work in **pairs** to select a dataset from the provided list and use one or more of the econometric techniques discussed in class to investigate a chosen question of interest. This project will allow you to demonstrate your understanding of econometric methods, data analysis, and empirical research.

Dataset Selection: Choose one of the datasets provided on the course website (listed below) or find a dataset on your own (a list of suggestions can be found below). Make sure the dataset is suitable for the question you are interested in and the approach you intend to use in answering that question.

Research Question: Formulate a clear and specific research question that you aim to answer using econometric analysis. Make sure your question is one that you can investigate with your chosen data and with the econometric techniques at your disposal.

Econometric Analysis: Analyze the data in order to answer your proposed research question. Your analysis should include at least one of the new techniques we learned in class this quarter:

- Propensity score matching
- Instrumental variables
- Difference in difference
- Regression discontinuity
- Binary logit model
- Multinomial logit model
- Nested logit model

Final Paper: Summarize your results in a research paper. The paper should be 8-12 pages long (size 12 font, double spaced, 1" margins). This is not strict and the length will depend on the topic you have chosen to write on, your data, and your own personal writing style. The paper must include:

- Introduction: Summarize the topic you are investigating, the specific research question that you are answering, the dataset you are using, and the technique you are using to answer the question. Don't include details here, but instead give a 3-6 sentence preview to the reader on what will be contained in the paper.
- Data Summary: Describe the data you are using including what variables you use and how they are measured. Summarize the data using either a table or a figure (e.g. scatterplot or histogram). Discuss any interesting patterns you see in the data that may be relevant to your topic.
- Model: Present the model that you are going to estimate including the econometric technique(s) you are using to answer the question. Explain the theoretical underpinnings of your chosen technique and its relevance to your research question.

- Results: Present and discuss the results. The results should be presented in a clearly readable table. The discussion should tell the reader what the results say about the question posed in the paper. This should include a statement of whether the results are statistically significant, an interpretation of the magnitude of the coefficients, and context for how these results answer the question. This section should also include the results of any robustness checks to test the sensitivity of your results to different model specifications or assumptions.
- Conclusion: Summarize the results, explore the implications of the results, and point to future research.
- References: Cite all references and relevant literature used in your paper

With the final paper I am NOT looking for a survey of existing research, a series of block quotes showing what some economist said about X, a regurgitation of my lectures, a list of pros and cons with a last minute conclusion, a book review, etc. I also do not want you to compress everything about econometrics your learned from this class into a single paper. You should only use those insights that are relevant to your topic and your argument

I am looking for a paper that attempts to answer a specific research question of interest to economists using data analysis. You should be able to summarize your paper in one or two sentences (the specific research question your paper addresses, your method for answering it, and your results). It should be a reasonably original paper (it is difficult with limited knowledge, time, and data to offer something truly original), and it should be your own take on the topic. It's also important to note that I do not expect or require you to find statistically significant results.

Data Sources Available on Class Website

- Telco Customer Churn (churn.csv): Customer data from a telecommunications provider on whether or not they left the company
- H-1B Firm Data (h1bfirmdata.xlsx): Annual data on publicly traded firms that hire foreign workers on H-1B visas (2010-2020)
- Heat Diseases Database (heartdiseas.csv): Database on whether or not patients have heart disease and their health-related characteristics
- Housing Sales Prices (houseprices.csv): Data on housing sales prices and home characteristics
- Movie Weekly Grosses (moviedata.xlsx): Data on weekly box office revenues for wide release movies (2009-2018)
- NBA Player Data (nbaplayerdata.xlsx): Annual data on NBA players and their statistics 2002-2019)
- NBA Team Data (nbateamdata.xlsx): Annual data on NBA teams and their total statistics (2002-2019)
- Over-the-Counter Medicine Data (OTCdata.xlsx): Weekly store-level data on demand for over-the-counter medicine

- NFL quarterback data (nflqbdata.csv): Advanced statistical data on NFL quarterbacks from 2018-2023
- NFL play by play data (nflpaybyplay_2023.csv): NFL play by play data from the 2023 season

Other Suggested Data Sources

- American Community Survey (ACS) (https://usa.ipums.org/usa): Monthly survey by U.S. Census Bureau on ancestry, educational attainment, income, language proficiency, migration, disability, employment, and housing characteristics
- American Time Use Survey (ATUS) (https://www.ipums.org/timeuse.shtml): Annual survey by Bureau of Labor statistics with nationally representative estimates of how, where, and with whom Americans spend their time
- Behavioral Risk Factor Surveillance System (BRFSS) (http://www.cdc.gov/BRFSS/): Annual data on U.S. residents' health-related risk behaviors, chronic health conditions, and use of preventive services
- Current Population Survey (CPS) (http://cps.ipums.org/cps/): Monthly survey by U.S. Census Bureau on labor force, employment, unemployment, hours of work, earnings, income, poverty, inequality, and other demographic and labor force characteristics
- Decennial Census (http://usa.ipums.org/usa/): Survey conduced every 10 years to determine the number of people living in the United States, the number of seats each state has in the House of Representatives, and demographic trends
- General Social Survey (GSS) (http://gss.norc.org/): Annual data on attitudes, behaviors, and beliefs on civil liberties, crime and violence, intergroup tolerance, morality, national spending priorities, psychological well-being, social mobility, and stress and traumatic events
- Integrated Postsecondary Education Data System (IPEDS) (https://nces.ed.gov/ipeds): Annual data from every university on enrollments, graduation rates, faculty and staff, finances, institutional prices, student financial aid, admissions, and academic libraries
- National Survey on Drug Use and Health (NSDUH) (https://nsduhweb.rti.org/respweb/about_nsduh.html): Annual data on tobacco, alcohol, and drug use, mental health, and other health-related issues
- National Health Interview Survey (NHIS) (https://nhis.ipums.org/nhis/): Annual survey on medical conditions, health insurance, doctor's office visits, physical activity, and other health behaviors
- National Longitudinal Study of Adolescent to Adult Health (Add Health) (http://www.cpc.unc.edu/projects/addhealth): Longitudinal sample of adolescents in grades 7-12 during the 1994-95 school year on respondents' social, economic, psychological, and physical well-being with contextual data on the family, neighborhood, community, school, friendships, peer groups, and romantic relationships
- National Sample Survey of Registered Nurses (NSSRN) (https://www.census.gov/programs-surveys/nssrn.html): Survey on supply and demand of nursing resources and characteristics of nurses such as education, training, employment, income, and demographics

- National Survey of College Graduates (NSCG) (https://highered.ipums.org/highered/): Data on college graduates' degree field, occupation, work activities, salary, and demographics