

San José State University (5/21/2020)

Economics 103a, Introduction to Econometrics & Research Methods, Section 80, SU 20

Instructor:	Matthew Holian, Ph.D.
Office Location:	DMH 166
Email / Phone:	Matthew.Holian@sjsu.edu / 408-924-1371
Office / Hours:	by appointment
Room/ Day/ Time:	Online
Prerequisites:	ECON 101 & intro statistics (e.g. STAT 95, UNVS 015S or equivalent).

Learning Management System

All students are assumed to be familiar with Canvas, the Learning Management System at SJSU.

Course Description

Research methods and core econometric techniques for analysis of causal effects, from difference-in-means tests of experimental data through multiple regression analysis of observational data. Topics include selecting an appropriate research question, reviewing the relevant literature, and obtaining data. Core econometric techniques used to analyze data in an original term paper.

Grading and Classroom Policy

This is a 4-unit course. The grading scale is: 60-62, D-, 62-68, D, 68-70, D+, 70-72, C-, 72-78, C, 78-80, C+, 80-82, B-, 82-88, B, 88-90, B+, 90-92, A-, 92-98, A, 98-100, A+. I generally don't allow makeup assignments unless there is sufficient notice and a well justified and documented reason. Disruptive use of laptops during class and food that emits externalities are not allowed.

Required Textbooks

- 1.) Angrist, J. D. and Pischke, J. 2014. *Mastering Metrics*, Princeton University Press, Princeton, N.J.

Resources: www.mastermetrics.com

- 2.) Bailey, M.W. 2016. *Real Econometrics: The Right Tools to Answer Important Questions*. Oxford University Press; 1st edition.

Resources: <https://global.oup.com/us/companion.websites/9780190296827/>

- 3.) Holian, M. J. Forthcoming. *Data and the American Dream: Contemporary Social Controversies and the American Community Survey*. Palgrave Macmillan.

Excerpts from this in-progress manuscript will be provided to students in class.

Recommended Textbooks

- 1.) Stock, J.H. and Watson, M.W. 2011. *Introduction to Econometrics*. Pearson, 3rd edition

This is the book we use in the graduate econometrics sequence, although it is perfectly accessible to undergraduates. Any recent edition is suitable. See also the publisher's Student Resources page for replication files for the book in Stata format: http://wps.aw.com/aw_stock_ie_3/178/45691/11696965.cw/

Required Computer Software

All students must have installed on their home machines free R and R Studio software. Students whose computers have limited memory are advised to create a free R Studio Cloud account.

Course and Program Learning Objectives (CLOs and PLOs)

This course reinforces PLO3: **research methods** and PLO5: **communication**, and introduces PLO4: **areas: quantitative methods**. Specific CLOs for this course include:

CLO 1.) Explain basic methods in econometrics and identify correct procedures

- a) Explain the difference between a variable and a statistic in the context of a regression equation.
- b) Define the terms "causal effect" and "ideal experiment". Explain the difference between descriptive statistics, inferential statistics, and causal inference.
- c) Give two examples of difference-in-means tests, using experimental and observational data, and explain when we can and cannot interpret a difference-in-means as an estimate of a causal effect.
- d) Describe how to use a simple (bivariate) regression model to carry out a difference in means test.
- e) Give an example of a regression coefficient estimate that suffers from omitted variable bias, and explain how the *regression control* technique could reduce bias in the example.
- f) Describe all the numbers in a Stargazer regression table in R; identify the main independent variable of interest, interpret the econometric models, test their statistical significance and evaluate them in terms of any potential bias.
- g) Discuss best practices in estimating standard errors.
- h) Discuss an example of a natural experiment, where: 1.) a *difference-in-means* is a plausible causal effect, and 2.) where a *difference-in-difference* (D-in-D) in means is a plausible causal effect. Finally, explain how an interaction model automates estimation of a D-in-D estimate.

CLO 2: Use technology to analyze data

- a) Create summary statistics for variables in a data set using the R software program.
- b) Estimate a regression model (coefficients and standard errors) and create a scatterplot with a regression line in R.
- c) Download data from the Internet and read it into a statistical software package
- d) Run an R script associated with a published research study by modifying the directory path, installing required packages, loading data, and obtaining results.
- e) Create a new script by modifying an existing script, and use your original results in a term paper

CLO 3: Prepare a scholarly research paper describing an original regression analysis:

- a) Formulate an interesting and important research question.
- b) Locate and describe data from Internet or other sources.
- c) Search and analyze scholarly literature related to research question.
- d) Write a review of econometric literature that is integrated and not merely an annotated bibliography; list and describe relevant studies and their research questions, the data and methods they used, and the results they found. Highlight any studies that provide compelling estimates of well-defined causal effects, or explain why a study does not.
- e) Develop, estimate and interpret a statistical model that can be used with the data to answer a question which is original and contributes to the literature.

The CLOs will be assessed through a term paper and oral examinations.

Assignments

<i>Assignment</i>	<i>Points</i>	<i>Due Dates</i>
Weekly assignments	10	Every Friday by 5p.m.
Weekly oral examinations	50	10 minutes, once each week, 5 points each
~2,000 word term paper (due in phases)	40	Abstract and Outline: 7/1 Rough Draft and Presentation: 8/5 Final Paper 8/10 @ 8a.m.

Grades will be based on the usual grading scale and **out of 100 points**.

Detailed rubrics for the term paper can be found at the end of this syllabus. We will have weekly check-in conversations, when I will evaluate your progress and provide assistance where needed. I will assess completion of learning objectives, and I will assign readings, software and writing assignments that will help you complete the learning objectives. Students will submit assignments on Canvas and we will communicate over Zoom video link.

Academic integrity

Cheating or plagiarism (presenting the work of another as your own) will result in a failing grade and sanctions by the University. Faculty members are required to report all infractions. **Note:** The term paper involves a replication and students will find references in the original studies they can use in their literature review sections. However, do not just paraphrase the description of this literature. I will consider too much paraphrasing to be unoriginal and it may result in a failing grade on the term paper, and reporting to the Student Conduct office.

University Policies

Per University Policy S16-9, university-wide policy information relevant to all courses, such as academic integrity, accommodations, etc. will be available on Office of Graduate and Undergraduate Programs' [web page](http://www.sjsu.edu/gup/syllabusinfo/) at <http://www.sjsu.edu/gup/syllabusinfo/>

Term Paper Components and Instructions

Students will write a term paper that analyzes micro data from www.ipums.usa.org and attempts to verify a previously published study. The studies listed below are examples of the types of studies students will locate, attempt to verify and then extend or reanalyze in some way.

Outline: By 7/1 students will write up 1.) an abstract describing their replication and extension study, and 2.) an outline that contains the title, written as a one sentence research question, five sections with section names, a complete bibliographic citation to the study you will attempt to verify, a regression equation you will estimate with a description of the variables, coefficients and estimation subsample. In the regression equation, index variables to make it clear to the reader what is the unit of observation (A household? A person?). Describe one or more original ideas for extending the model. Finally, include any results you may have already produced, even if they are preliminary.

At 8a.m. on 8/5 you will submit a rough draft of your paper. The paper must expand upon the outline, and include your best results to date. You'll make a short, seven-minute presentation of your rough draft, using slides, later that day. Both your presentation and your draft should

discuss best examples in the literature of estimating a well-defined causal effects, and your original contribution to this area. A rubric for presentations is on the last page of this syllabus.

After you receive feedback from the presentation, you will revise the rough draft and submit the final paper by 8a.m. on **8/10**.

All papers must have five numbered sections: 1.) Intro, 2.) Literature Review & Economic Theory, 3.) Data Description, 4.) Empirical Results, and 5.) Conclusion, as described in Stock and Watson's "Conducting a Regression Study Using Economic Data" (on Canvas, Files > PDFs). Sections will be about five paragraphs in length and each paragraph about five sentences. This "5x5" suggestion is a rule-of-thumb and need not be followed exactly. However, all papers *must* have these three tables: Variable Descriptions, Summary Statistics, and Regression Results. Original figures like maps created with GIS software are encouraged but not required; *all copied tables and figures are prohibited unless cleared with the instructor*. Original tables must be formatted exactly as described in class. Your paper should have seven or more references (listed in a Bibliography section) in Chicago format. All papers must list and describe an equation describing the empirical model, and must contain an abstract. As an example, a brief "letters" style article (like Holian 2020a, 2020b) is close to (though slightly shorter than) what is required here. In particular, compared to a letters article, your paper must have a whole literature review section and table of summary statistics.

Examples of Good Candidates for Replication in this Class

Bailey, James, and Dhaval Dave. "The effect of the Affordable Care Act on entrepreneurship among older adults." *Eastern Economic Journal* 45, no. 1 (2019): 141-159.

Comolli, Chiara Ludovica, and Fabrizio Bernardi. "The causal effect of the great recession on childlessness of white American women." *IZA Journal of Labor Economics* 4, no. 1 (2015): 21.

Costa, Dora L., and Matthew E. Kahn. "Electricity consumption and durable housing: understanding cohort effects." *American Economic Review*, 101, no. 3 (2011): 88-92.

Holian, Matthew J. "The impact of urban form on vehicle ownership." *Economics Letters* 186 (2020a): 108763.z

Holian, Matthew J. "The impact of building energy codes on household electricity expenditures." *Economics Letters* 186 (2020b): 108841.

Orrenius, Pia M., and Madeline Zavodny. "The Impact of Temporary Protected Status on Immigrants' Labor Market Outcomes." *American Economic Review* 105, no. 5 (2015): 576-80.

Winters, John V. "Is economics a good major for future lawyers? Evidence from earnings data." *The Journal of Economic Education* 47, no. 2 (2016): 187-191.

Finding Your Own Studies to Replicate

You may replicate one of the studies listed above, another study discussed in *Data and the American Dream*, or one you find on your own. Published studies that are good candidates for replication in this class, like most of the ones listed above, generally have two characteristics. First, they contain statistics that were estimated only with the micro data from IPUMS (they do

not require merging on data from other sources.) Second, they were estimated with the types of regression models that are the focus of this class.

I have several suggestions for finding studies to replicate. First, search the bibliography at: <https://bibliography.ipums.org/>. All of these studies use the IPUMS micro data in some way, but students will have to determine whether they use merged data or advanced methods.

Second, do a “cited reference search” using Google Scholar. For example, search for the following citation:

Ruggles, Steven, J. Trent Alexander, Katie Genadek, Ronald Goeken, Matthew B. Schroeder, and Matthew Sobek. "Integrated public use microdata series: Version 5.0 [Machine-readable database]." *Minneapolis: University of Minnesota* 42 (2010).

Authors who used this version of the IPUMS data are supposed to cite this study (although not all do.) In Google Scholar, you can see a list of all studies Google knows about that cited any reference by clicking the “ button beneath the reference. As I write this, 2,703 studies have cited Ruggles et al. (2010). Thousands of other studies have cited different versions of the IPUMS data. To find these, I suggest you click the link to Steven Ruggles’ Google Scholar author page, where you’ll see a list of his most cited studies. Do cited reference searches on these as well.

When you find articles, you will usually need to access them through our university’s subscriptions. Our library subscribes to most scholarly journals. Go to <http://library.sjsu.edu/> and click Journal Titles. Enter the journal title. You can do this faster; on Google Scholar, set up “library links” under “settings” to see direct links to articles.

Here are two other options for locating published econometric research: 1.) search the JSTOR database. <http://library.calstate.edu/sanjose/databases/alphabetical?alpha=J>. Limit your search to Economics journals in Advanced Search, and search using keywords in your area of interest; 2.) search the EconLit database (change end of link above from “J” to “E”). You can search by keyword; if you are interested in an area within economics, try also to search by subject (SU) using JEL code: <https://www.aeaweb.org/econlit/jelCodes.php>.

Term Paper Grading

For Rough Drafts and Term Papers, I’ll be looking for the things listed in the table below in each section. Of course following the guidelines of the assignment listed above is important; to repeat, all papers must have: 1.) five sections, 2.) about 25 paragraphs or 125 sentences, 3.) tables of Variable Descriptions, Summary Statistics, and Regression Results, 4.) no copied figures, 5.) References (at least seven) in Chicago format, 6.) an equation describing an empirical model, 7.) an abstract that summarizes the paper, and 8.) One or more JEL codes.

Content is most important. Remember this is a class about using observational data to estimate causal effects, so focus on discussing literature that strives to define and estimate causal effects, evaluate how credible the estimates in the study you replicate are as causal effects, and how your extension adds to our understanding, or improves upon the replication.

Grammar is also important. There are formal rules of grammar (Does each paragraph have a topic sentence? Do all sentences develop one controlling idea? Does paper feature appropriate punctuation, syntax, usage? Is the paper free of spelling errors? Are citations used appropriately?) There are also less formal “stylistic” elements of writing: Does paper avoid the

passive voice? Overall, is the writing style and voice appropriate? Does it appear the student read contemporary and seminal studies and is it written in the style of the profession?

The best way to learn the style of economic writing is to read economic writing; this suggests you should try to actually read the journal articles you review in your literature review section, not just the abstracts. You can also consult various guides to style and writing. In addition to the McCloskey (1985) article that will be discussed in class, a book-length treatment that is widely used across fields is Strunk (2007) *The elements of style*.

For outlines, rough drafts and final papers, the percentages listed below each criteria are the approximate weight of each section towards your final grade. The difference between the assignments, essentially, is just the level of completeness—an outline may not feature any results, a draft will have results but the writing will be incomplete, and a final paper will be a complete academic-style term paper.

Criteria	Description of Criteria for Rough Drafts and Term Papers
<i>Intro</i> 5%	Reveal the research question to the reader clearly and at the outset. Motivate interest. Describe what this paper does (i.e. how you answer the research question), the main findings, and how the paper is structured.
<i>Literature Review and Economic Theory</i> 40%	This section should describe the relevant theory that is behind the hypotheses tested in past econometric studies. You may be able to describe the theory so precisely that you include a formal mathematical model in this section, but a clear qualitative description of formal theory would be enough. Discuss at least seven econometric studies, noting which provides the most compelling estimate of the causal effect of interest. Explain why (i.e. discuss ways the state-of-the art study handles a well-defined econometric problem such as Omitted Variable Bias.) <i>The literature review must be integrated and not merely an annotated bibliography.</i>
<i>Data and Methods</i> 25%	Describe the data; even though I am grading it and I know where the data comes from, write it for a more general audience. Describe the estimation subsample, and present the tables of Variable Descriptions and Summary Statistics. Present the equation you will be estimating here.
<i>Results</i> 25%	Present regression results in a table with several models in different columns. Indicate in the text which of the models is your best attempt at replicating the previously published model. Also indicate which is the best attempt to estimate a causal effect (this may be your extension.) Interpret the results for the reader in terms of the sign, magnitude, economic and statistical significance, as well as the implications for the theory discussed in section 2.
<i>Conclusion</i> 5%	Overall, your paper should aim to discuss a specific causal research question that relates to a relevant economics literature. Acknowledge any limitations. Suggest directions for improvement in future research. Discuss policy implications and how your results can be used in a cost-benefit analysis, if applicable.

Course Schedule (To be updated periodically; note date on first page of this syllabus.)

Date / Location	Topic	Readings to be completed before class
1-Jun	Kick-off: <i>Intro to Econometrics and Research Methods</i>	Holian Ch 8 (Appendix) & Winters (2016)
3-Jun	Check in #1: <i>Descriptive Statistics, Causal Inference & Regression</i>	Holian Ch 1
8-Jun	Check in #2: <i>The Quest for Causality and Good Data Practices</i>	Bailey Ch 1-2, 3.1, 5.1, 5.2, 6.1 & 6.2
10-Jun	Group Discussion: <i>Bivariate & Multivariate OLS; Binary Variables</i>	
15-Jun	Check in #3: <i>Regression Control; Categorical Variables</i>	Holian Ch 2 & Bailey 6.3
17-Jun	Group Discussion: <i>Regression Control Case Study</i>	Costa and Kahn (2011) & Holian (2020)
22-Jun	Check in #4: <i>Difference-in-Differences; Interaction models</i>	Holian Ch 3 & Bailey 6.4 & 8.5
24-Jun	Group Discussion: <i>Difference-in-Differences Case Study</i>	Orrenius and Zavodny (2015)
29-Jun	Check in #5: <i>Randomized Trials and Review of Hypothesis Tests</i>	Angrist and Pischke, Ch 1 (including appendix)
1-Jul	Group Discussion: <i>Observational and Experimental Data; OUTLINE</i>	
6-Jul	Check in #6: <i>Regression Control Revisited</i>	Angrist and Pischke, Ch 2 (including appendix)
8-Jul	Group Discussion: <i>Omitted Variables Bias Revisited</i>	
13-Jul	Check in #7: <i>D-in-D Revisited</i>	Angrist and Pischke, Ch 5
15-Jul	Group Discussion: <i>D-in-D Case Studies</i>	Holian Ch 4-5
20-Jul	Check in #8: <i>Bivariate OLS and Hypothesis Testing, Revisited</i>	Bailey Ch 3, 4.1-4.3, 4.5-4.6, 5.4-5.5
22-Jul	Group Discussion: <i>Multivariate OLS Revisited</i>	
27-Jul	Check in #9: <i>Linear Probability Models and Nonlinear Models</i>	Bailey 7.1, 7.2 & 12.1
29-Jul	Group Discussion: <i>Review for Final Check-in Call</i>	
3-Aug	Check in #10	Bailey Ch 1-6 and sections 7.1, 7.2, 8.5 & 12.1
5-Aug	<i>Student Presentations</i> ROUGH DRAFT	
10-Aug	FINAL PAPER	

Additional References

McCloskey, D. (1985). Economical writing. *Economic Inquiry*, 23(2), 187-222.

Writing a Literature Review Section. (<https://library.sjsu.edu/video/writing-literature-review-section>)

Stock, J.H. & Watson, M.W. “Conducting a Regression Study Using Economic Data”
<https://wps.pearsoned.com/wps/media/objects/11422/11696965/stoc2517.ch10.343-350.C1.pdf>

Presentations:

This page contains more detailed guidelines for how to plan your presentation. We have a maximum of 10 minutes per speaker. After this amount of time passes, we absolutely have to move on to the next speaker. We want to allow a minute or two of discussion after each one. So, **plan for a seven minute presentation**. Please practice your presentations to make sure you don't exceed seven minutes by more than a minute at most.

I've found a good rule of thumb is to plan one slide for each minute of presentation, certainly not more. So say you **shoot for seven slides**: 1.) Title page, 2.) Literature Review, 3.) the regression model / empirical approach 4.) Table of Variable Descriptions 5.) Table of Summary Statistics 6.) Table of Regression Results 7.) Plan for future research.

Avoid distracting colors and themes, don't put too much text on a slide, don't make the numbers you want people to see in your tables tiny, don't read off your slides (the audience can read them). Do aim to teach the audience something about the econometric methods and data used in the literature you have reviewed, and explain your research findings clearly. In terms of what concepts you present, see the following rubric:

Criteria	Description of Criteria for Presentations
<i>Title slide</i> 5%	Slide 1: Reveal the research question to the audience clearly and at the outset. Motivate interest.
<i>Literature Review and Economic Theory</i> 40%	Slide 2: In this part, briefly discuss the relevant theory. A brief but clear qualitative description of the theory is adequate. On this slide, list up to three references, and discuss in some detail the one that provides the single best discussion of the estimation of the causal effect of interest. Explain why (i.e. discuss ways the state-of-the art study handles a well-defined econometric problem such as Omitted Variable Bias.) <i>The literature review must be integrated and not merely an annotated bibliography.</i>
<i>Data and Methods</i> 25%	Slides 3, 4 and 5: Present the equation you will be estimating here. Describe the data; even though I am grading it and I know where the data comes from, write it for a more general audience. Describe the estimation subsample, and present the tables of Variable Descriptions and Summary Statistics.
<i>Results</i> 25%	Slide 6: Present regression results in a table with several models in different columns. Indicate which of the models is your best attempt at replicating the previously published model, or indicate which is the best attempt to estimate a causal effect. Interpret the results for the audience in terms of the sign, magnitude, economic and statistical significance, as well as the implications for the theory discussed on slide 2.
<i>Conclusion</i> 5%	Slide 7: Acknowledge any limitations. Suggest directions for improvement in future research. Discuss policy implications and how your results can be used in a cost-benefit analysis, if applicable.