

Econ 136

Lecture 1: Introduction

Edward Vytlacil
Yale University

- **Staff**
- Econometrics vs. Statistics/Data Science
- Econ 136

Professor: Edward Vytlacil

- Economics Ph.d., University of Chicago,
 - studied under James Heckman.
- Professor of Economics, Yale University,
 - previously on faculty at Stanford, Columbia, NYU.

Professor: Edward Vytlacil

- Field: Microeconomics, Applied-Microeconomics;
 - Subfield: Causal Inference/treatment effects,
 - Applications: education, health, labor, corporate finance
- (former) Co-Editor of [Journal of Applied Econometrics](#).
- Fellow of the Econometric Society.
- Founding member and former Director of the [International Association of Applied Econometrics](#).

Professor: Edward Vytacil

- Contact:
 - Email: edward.vytacil@yale.edu
 - Office Hours:
 - Location: [30 Hillhouse Ave, Room C149](#)
 - Time: Tuesdays, 9:30-11:00am.

Teaching Fellow: Kan Yao

- Yale Economics Ph.d. student.
- Field: Econometric Theory
- Contact:
 - Email: kan.yao@yale.edu
 - Office Hours: TBA

- Staff
- **Econometrics vs. Statistics/Data Science**
- Econ 136

Course: Econ 136, Econometrics

What is econometrics?

- how is different from probability theory? statistics?
data science?

Probability theory

- Branch of mathematics,
- Derive implications of known probabilistic model.

Statistics builds upon prob. theory

- “*The science of learning from data*”
(estimation and inference of probabilistic model),
- Origins in:
 - math/applied math,
 - experiments,
 - data scarcity.

Data Science

- If statistics is the “*science of learning from data*,” then what is data science?
- Is data science another name for applied statistics?

Data Science builds upon stats & CS

- Data science origins in:
 - CS/engineering (not math),
 - environment of data abundance.
- Data science overlaps with, but is different from stat:
 - different focus, perspective,

Data Science builds upon stats & CS

- Data science different focus from stat, more focus on:
 - computation,
 - algorithms,
 - data visualization,
 - prediction,
 - domain expertise,
 - work flow...

Econometrics

- Is econometrics the application of statistics to economic data? of data science to economic data?
- **No!!**
 - Does not *just* apply statistics or data science ...
 - Does not *just* consider “economic data.”

Econometrics

Samuelson, Koopmans, and Stone (1954):

the quantitative analysis of actual economic phenomena based on the concurrent development of theory and observation, related by appropriate methods of inference

- For contemporary economics:
 - “economic phenomena” should be broadly construed,
 - econometrics sometimes tightly connected to econ theory, sometimes not.

Econometrics builds upon stats and econ theory

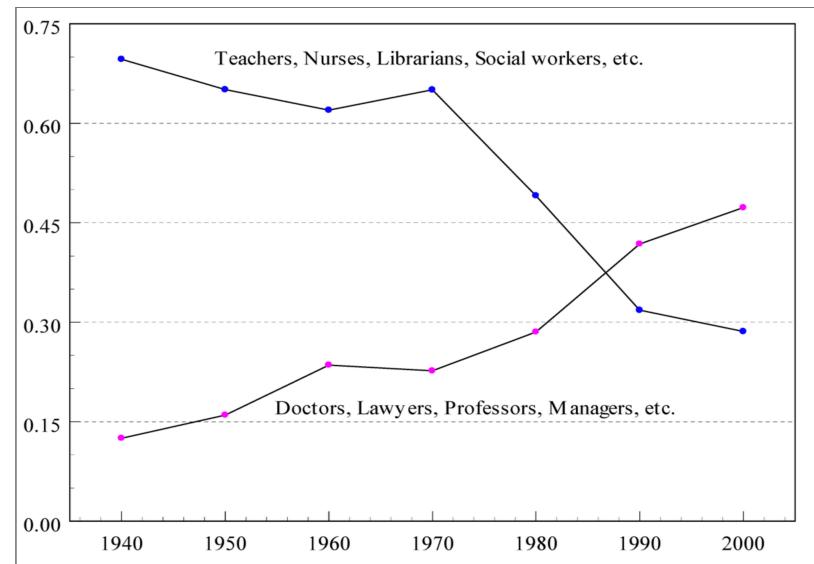
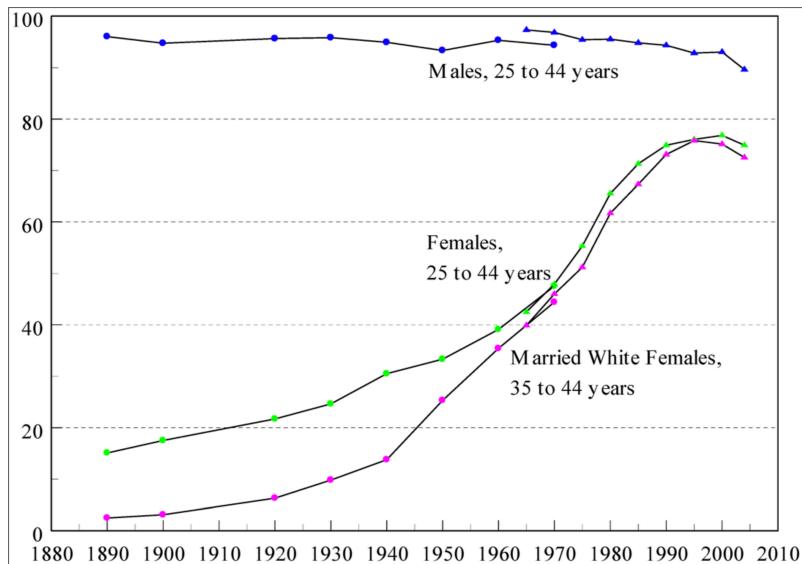
- Merging of statistics, economic theory, and data (and more recently data science)
- Econometrics origins in:
 - economic models,
 - not in experiments
 - though experimental analysis has become more important in economics/econometrics.

Econometrics builds upon stats and econ theory

- Focus on counterfactual prediction,
 - sometimes answering “why” and often “what if”?
- Different perspective on models and causality from stats.
 - *ceterus paribus* paradigm in economics,
 - though some partial convergence across fields.

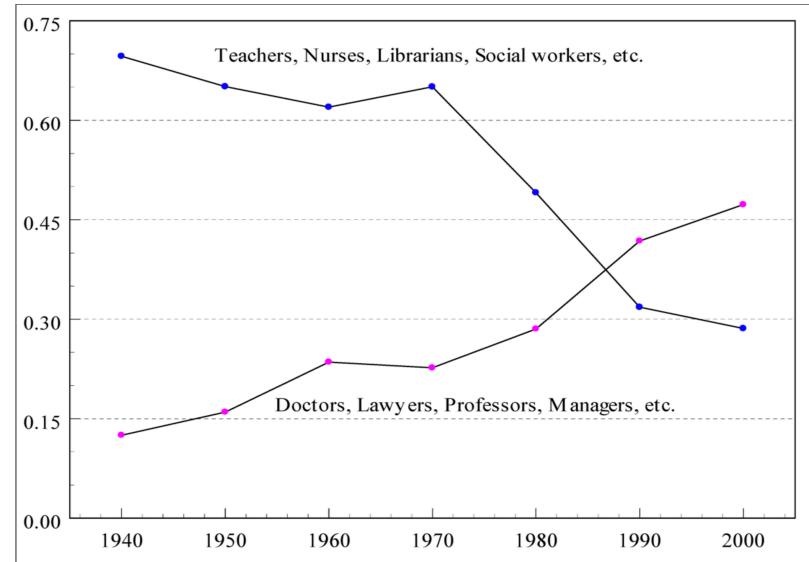
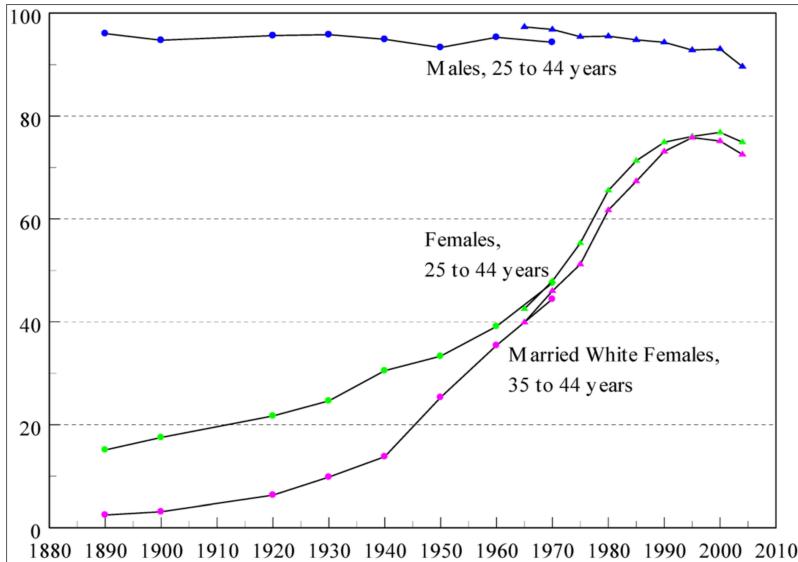
Goals of econometrics include

- Understanding economic phenomenon,
 - e.g., why did female labor force participation increase dramatically, especially in professional fields?



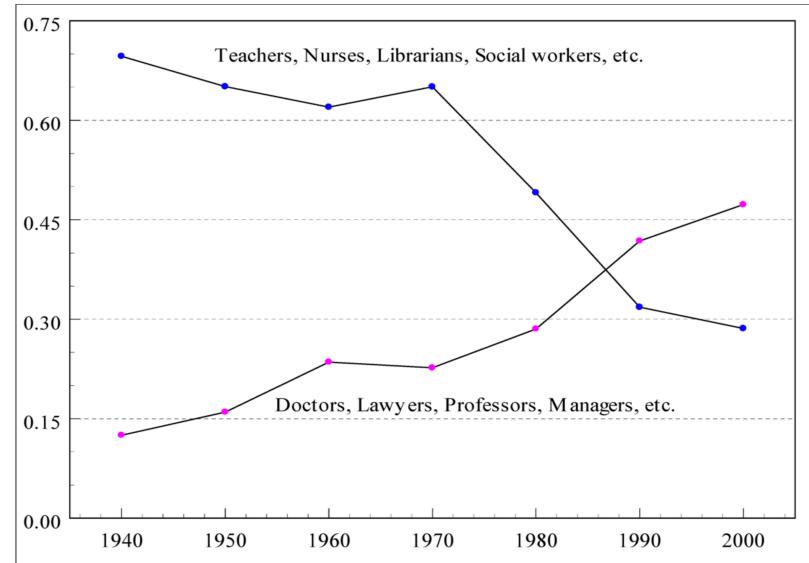
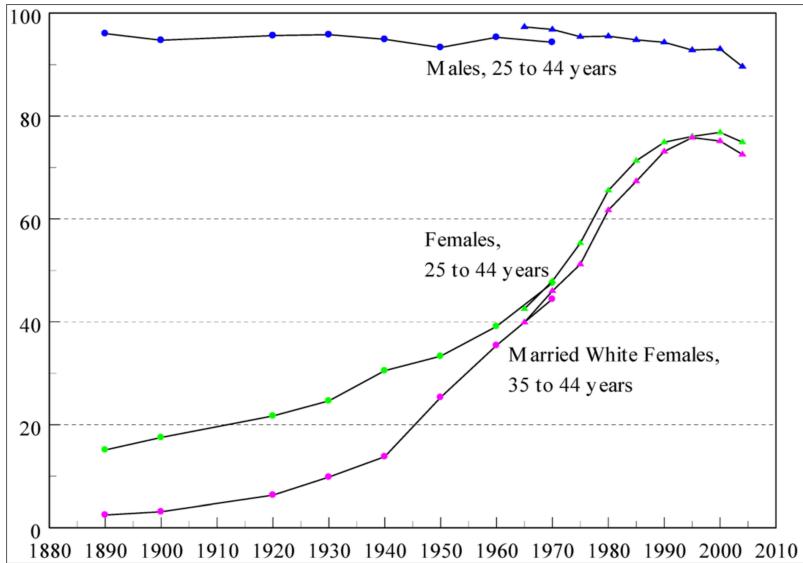
Labor force participation (left) and occupations of college educated women (right).

Goals of econometrics include



- Distinguish correlation vs causality,
 - e.g., connection between access to birth control/abortion and female labor force participation?
 - answer “what if” questions.

Goals of econometrics include



- Important date for access to birth control/abortion includes 1965 ([Griswold v. Connecticut](#)), 1971 ([Eisenstadt v. Baird](#)), 1973 ([Roe v. Wade](#)).

Goals of econometrics include

- Estimate economically meaningful quantities,
 - e.g., supply and demand functions, hedonic equations.
- Test economic theory, inform economic theory,
 - e.g., distinguish taste-based vs statistical discrimination

Influence of Econometrics

- Econometrics grew out of econ, but has been influential in
 - other social sciences
(political science, law, ...),
 - policy,
 - industry.
- Has influenced statistics and CS/AI,
 - though relationship often contentious.

- Staff
- Econometrics vs. Statistics/Data Science
- **Econ 136**

Econ 136

- Is designed for students in:
 - Economics & Mathematics joint major
 - Computer Science & Economics joint major.
- Is designed to prepare students for graduate level courses.
- Very different from Econ 117/123.

Econ 136

- Is an *Econometrics course*,
 - Considers empirical applications of interest to economists;
 - Teaches how economists think about data, connecting data to economic models.

Econ 136

- develops econometric *theory*,
 - develops theory with mathematical rigor, building upon theoretical statistics;
 - far more theoretical than the Econ 117/123 sequence;
 - appropriate training to prepare for graduate level courses
 - why learn theory? why not just how to apply?

Econ 136

- develops econometric *practice*,
 - will develop when to use alternative econometric methods, how to justify their use, and interpret the results,
 - will use **R** to analyze real data of economic interest, including from recent economics publications.

Econ 136

- incorporates computationally intensive methods such as bootstrap and cross-validation,
- incorporates aspects of data science.

Goal of Econ 136

- By the time you complete this course, for you to:
 - have a rigorous training in econometric theory, preparing you for future graduate-level work in econometrics or related fields;
 - have the strong basis for conducting original empirical research in economics, other social science, policy, or industry.

Math in Econ 136

The course is substantially more mathematically rigorous than typical undergraduate-level econometrics courses, and will make extensive use of:

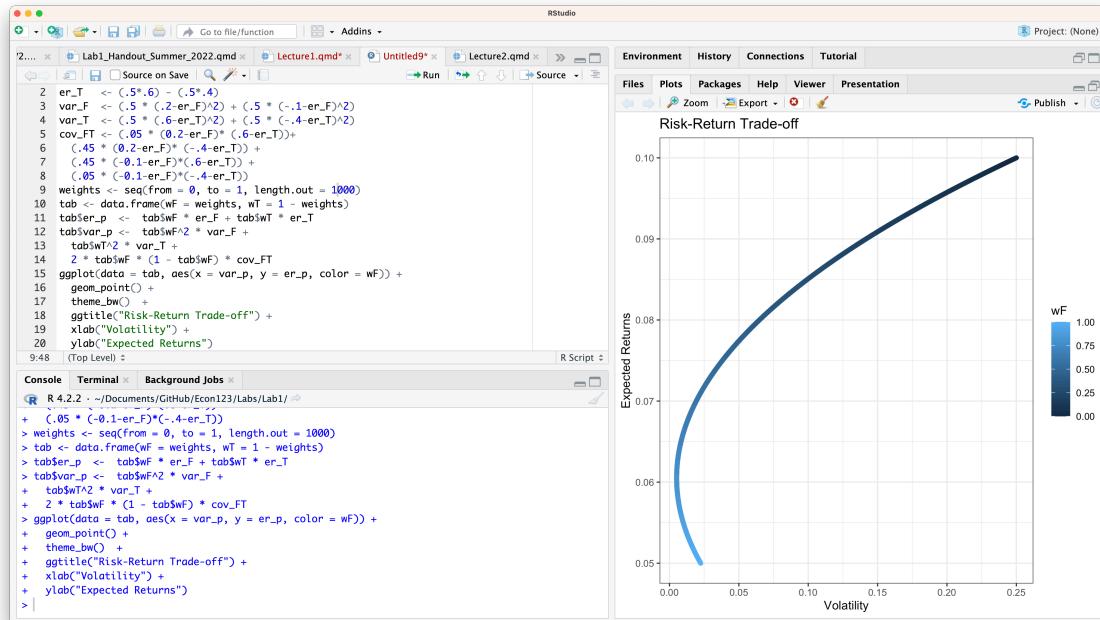
- Multivariate calculus
- Linear Algebra

Econ 136: Textbooks

- Required textbook: Hansen (2022a) *Econometrics*
 - Ph.D. level textbook for econometric theory . . .
- Optional textbook:
 - Hansen (2022b) *Probability and Statistics for Economists*
 - optional for review of probability theory and statistics

R in Econ 136

- The course will use more advanced coding than it typical for an undergraduate economics course, using R,
 - in R-Studio IDE, to



R in Econ 136

- create *reproducible documents* with R-Markdown/Quarto
 - combining text, code and results.

The screenshot shows the RStudio interface with the following components:

- Source View:** Displays R code for generating a table. The code includes functions like `tapply` and `sum` to calculate means and standard deviations by treatment group.
- Text View:** Contains explanatory text about the table's purpose and a citation to Negi and Wooldridge (2021).
- Table View:** A "Balance at Baseline" table comparing Control and Treated groups across various characteristics.
- Preview View:** Shows a rendered version of the document with the generated table and citation.
- Console View:** Displays the command `Running 3:15`.
- Quarto View:** Shows the document's location and bibliography information.

	Balance at Baseline			
	Control	Treated	Diff.	Std Diff
Girl	0.50	0.48	-0.02	-0.04
Age	10.51	10.49	-0.02	-0.01
Highest	3.26	3.28	0.02	0.01
Grade				
Enrolled	0.87	0.87	0.002	0.01

```
1 return_mean_by_treatment <- function(x){  
2   means.t<-tapply(x,dfPre$treat,mean)  
3   var.t<-tapply(x,dfPre$treat,var)  
4   return(c(means.t,var.t))  
5 }  
6  
7 vars <- c("girl","age1","hgc1","school")
```

R in Econ 136

- In this course, we will use R,
 - what if you want to use STATA, python, ...
- Advantages of R over other options? Disadvantages?

If you are not already comfortable with R, I recommend reading James et al. ([2013](#)) *An introduction to statistical learning in R* Chapter 2.3, and working through [Project 1](#), and [Project 2](#) of [Hands-On Programming with R](#) by Garrett Grolemund. If you are new to *ggplot*, see [ggplot tutorial](#)

Econ 136 topics will include:

- Conditional expectations and linear projections,
- Causal Analysis,
- Asymptotic Analysis,
- Linear regression analysis,
- Bootstrap,
- Instrumental Variables,
- Limited Dependent Variable models.

Econ 136 Applications include:

- **Finance:**
 1. Asset diversification,
 2. Capital Asset Pricing Model,
- **Labor and education economics:**
 1. Returns to schooling,
 2. Labor supply
 3. Effect of early childhood interventions.

Econ 136 Applications include:

- Discrimination, including in
 1. loans,
 2. job market,
 3. police force.

Will relate to economic models of discrimination:
statistical- vs taste-based discrimination.

Econ 136: Prerequisites

- Pre-Req:
 - Econ 135: Introduction to Probability and Statistics, or S&DS 241 and 242.
 - What if you haven't taken Econ 135 or S&DS 241+242?

Econ 136: Prerequisites

- Math Pre-Req:
 - multivariate calculus as acquired by having already taken MATH 120 or equivalent course,
 - linear algebra, as acquired by having already taken MAT 225 or 226 or equivalent course.

Classes alternate lectures and labs:

- **Lectures:** Lectures will primarily use blackboard, will sometimes be accompanying handouts.
- **Labs:** In-class labs where we live-code in R to analyze real data. The labs will be designed to directly help you with your problem sets.
- You are expected to attend lectures and labs.
I will call on students.

Course Webpages

- **Canvas webpage**

<https://yale.instructure.com/courses/103938>

- Includes schedule, handouts, lab materials, problem sets, readings and more.
- Updated syllabus as we go.

- **Ed Discussion webpage:**

- place to ask questions about the course, get help from staff and peers.

Assignments	Share of Course Grade
Online Quizzes	10%
Problem Sets	30%
Midterm	25%
Final	35%

On-Line Quizzes:

- Posted on Fridays:
 - to the [course webpage](#).
 - approximately every other week,
 - available on-line starting on Fridays at 2:30pm, due following Sunday at 11:59pm
 - you have one hour to complete the quiz once you start it
- Quizzes will focus on theoretical questions, some questions on R coding.

On-Line Quizzes:

- Open book/open notes,
 - but you cannot collaborate with, or discuss with, other students until the solutions are posted.
- Lowest quiz score will be dropped.

Problem Sets

- Will include primarily theoretical questions but also computational/empirical work.
- Approximately every other week, due on Thursdays at 2:30pm.
- Due dates are strict.¹
- The lowest problem set score will be dropped.

Problem Sets

- You may work in groups of up to four students on the problem sets.¹
- However, you must turn in your own assignment and indicate on your submission the other members of the group.

Exams

- **Midterm:**
 - in-class, Tuesday February 25 (date tentative)
- **Final:**
 - Saturday, May 3, 2025 at 9am.
- Exams will focus on theory, with some R related questions including interpreting empirical output from R.

What's next

- On Thursday, lecture will review rules for expected value and variance of random vectors.
 - Handouts:
 - [Handout 1: Rules for Expected Value and Variance](#)
 - [Handout 2: Binomial, Normal and t- Distributions](#)

What's next

- On Thursday, lecture will review rules for expected value and variance of random vectors.
 - *Optional Reading: Review Expectations and Variance*
 - Hansen (2022b) [Probability and Statistics for Economists](#), chapters 2.1-2.3, 2.5-2.8, 2.10, 2.13-2.14, 2.18-2.19, 3.1-3.2, 3.4, 3.8, 3.12-3.14, 4.1-4.7, 4.11-4.12, 5.1-5.2, 5.7-5.8

What's next

- On Thursday, lecture will review rules for expected value and variance of random vectors.
 - *Optional Reading: Review Matrix Algebra*
 - Hansen (2022b) [Probability and Statistics for Economists](#) Appendix A.11, or...
 - Hansen (2022a) [Econometrics](#) Appendix A.1, A.3, A.4

What's next

- To prepare for first problem set, if you are not already comfortable with R,
 - James et al. (2013) An introduction to statistical learning in R Chapter 2.3.
 - if you are new to *ggplot*, see [ggplot tutorial](#)
- To prepare for use of R later in course, work through [Project 1](#), and [Project 2](#) of [Hands-On Programming with R](#) by Garrett Grolemund.

What's next

- First problem set due Thursday February 6 at 2:30pm.
 - Will include theoretical questions as well as use R.
- First quiz goes live Friday January 24 at 2:30pm, closes Sunday January 26 at 11:59pm.

References

- Donoho, David. 2017. “50 Years of Data Science.” *Journal of Computational and Graphical Statistics* 26 (4): 745–66. <https://doi.org/10.1080/10618600.2017.138477>
- Goldin, Claudia. 2006a. “The Quiet Revolution That Transformed Women’s Employment, Education, and Family.” National Bureau of Economic Research Cambridge, Mass., USA.
- . 2006b. “The Quiet Revolution That Transformed Women’s Employment, Education, and Family.” *American Economic Review* 96 (2): 1–21.
- Hansen, Bruce. 2022a. *Econometrics*. Princeton University Press.
<https://press.princeton.edu/books/hardcover/9780691235899/econometrics>.
- . 2022b. *Probability and Statistics for Economists*. Princeton University Press.
<https://press.princeton.edu/books/hardcover/9780691235943/probability-and-statistics-for-economists>.
- James, Gareth, Daniela Witten, Trevor Hastie, Robert Tibshirani, et al. 2013. *An Introduction to Statistical Learning in r*. Vol. 112. Springer.
<https://www.statlearning.com>.

