

# Econ 2136

Lecture 1: Introduction

Edward Vytlacil  
Yale University

- **Staff**
- Econometrics vs. Statistics/Data Science
- Econ 2136

# 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](mailto:edward.vytacil@yale.edu)
  - Office Hours:
    - Location: 30 Hillhouse Ave, Room C149
    - Time: Tuesdays, 10:30am-12:00pm.

# Teaching Fellow: Wonwoo Bae

- Yale Economics Ph.d. student.
- Field: Econometric Theory
- Contact:
  - Email: [wonwoo.bae@yale.edu](mailto:wonwoo.bae@yale.edu)
  - Office Hours: TBA

- Staff
- **Econometrics vs. Statistics/Data Science**
- Econ 2136

# Course: Econ 2136, Econometrics

## What is econometrics?

- how is different from probability thoery? 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

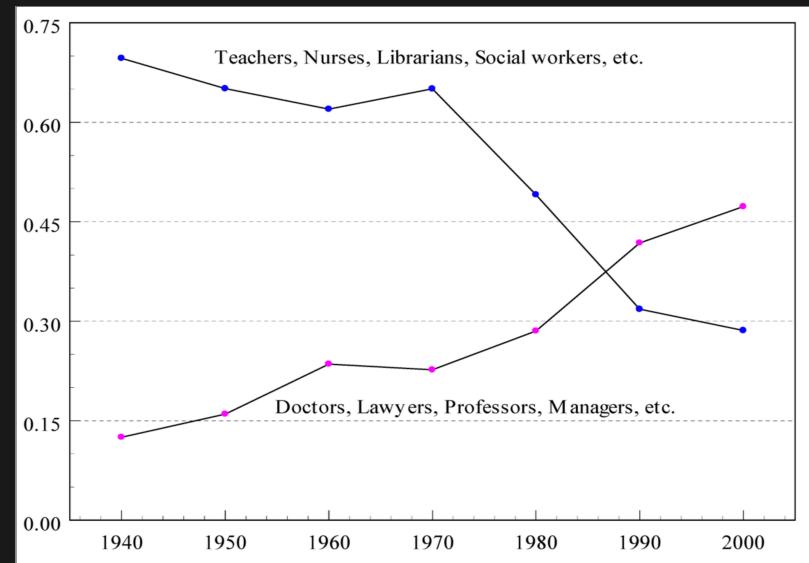
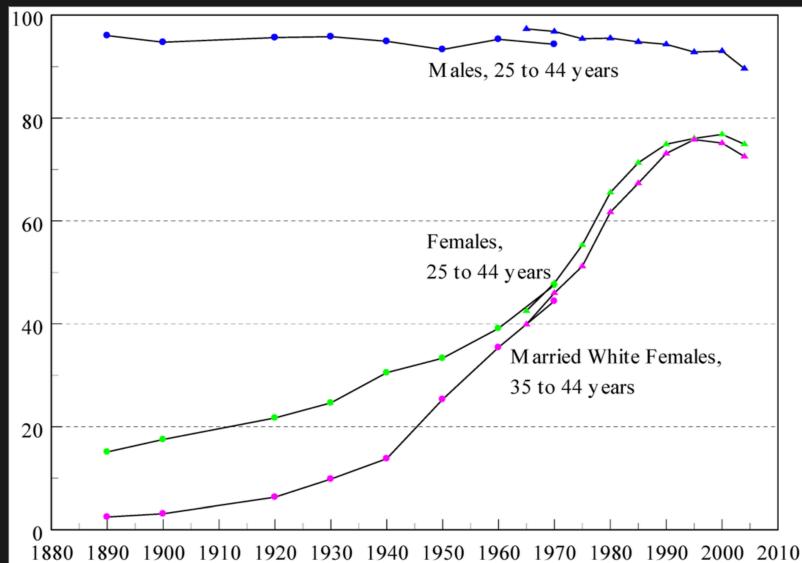
- Merging of statistics, economic theory, and data (and more recently data science)
- Econometrics origins in:
  - economic models,
  - not in experiments
    - though experimental analysis has become more important in economics/econometrics.

# 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

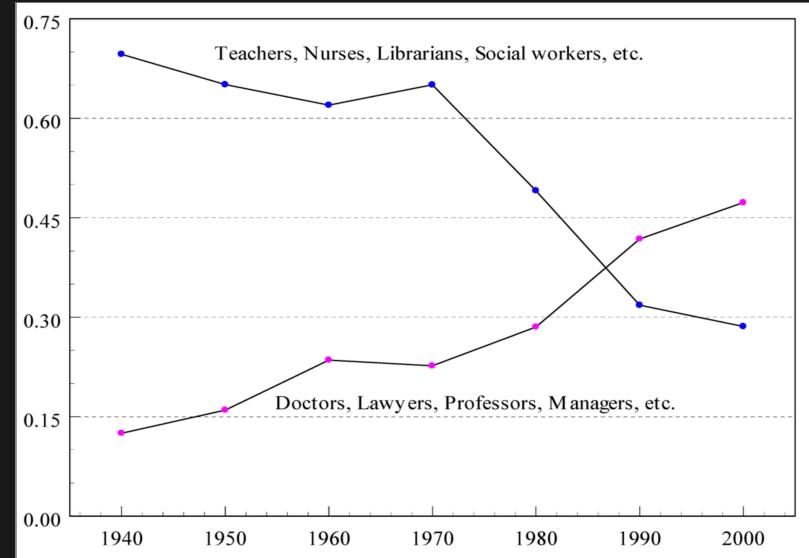
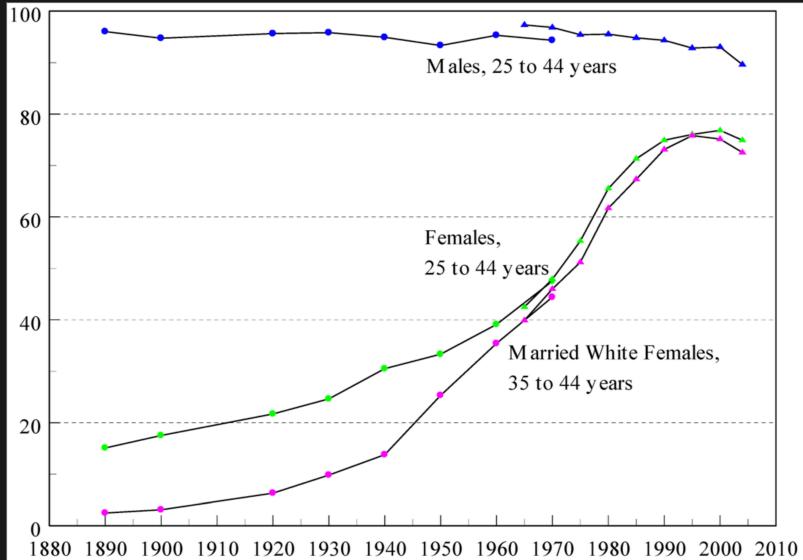
- Understanding economic phenomenon,
  - e.g., why did female labor force participation increase dramatically, especially in professional fields?



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

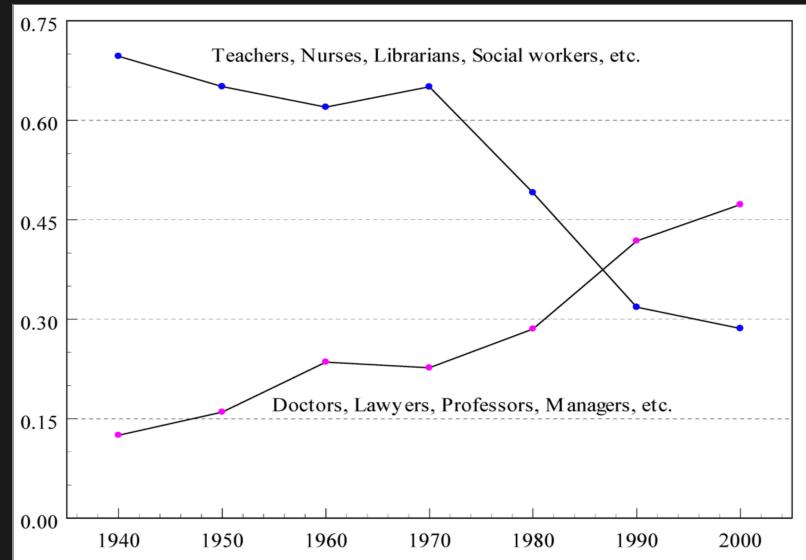
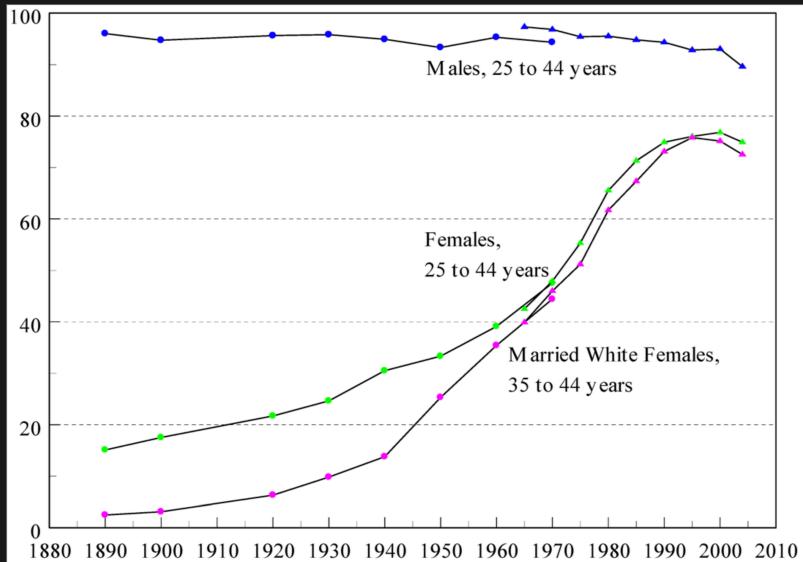
Econ 2136: Lecture 1

# 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 include 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 2136

# Econ 2136

- 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 1117/2123.

# Econ 2136

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

# Econ 2136

- develops econometric *theory*,
  - develops theory with mathematical rigor, building upon theoretical statistics;
  - far more theoretical than the Econ 1117/2123 sequence;
  - appropriate training to prepare for graduate level courses
  - why learn theory? why not just how to apply?

# Econ 2136

- 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 2136

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

# Goal of Econ 2136

- 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 2136

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

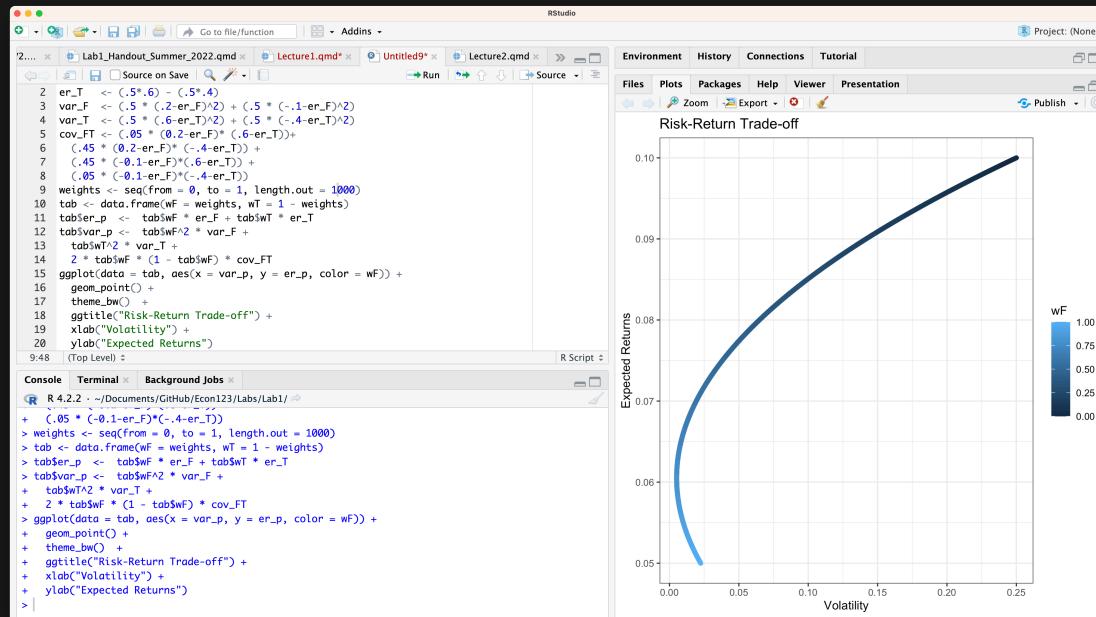
- Multivariate calculus
- Linear Algebra

# Econ 2136: Textbooks

- Required textbook: Hansen (2022a) *Econometrics*
  - Ph.D. level textbook for econometric theory . . .
- Optional textbooks:
  - Hansen (2022b) *Probability and Statistics for Economists*
    - optional for review of probability theory and statistics
  - James et al. (2013) *An introduction to statistical learning in R*
    - we will use for select topics.

# R in Econ 2136

- 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 2136

- 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 creating a table reporting baseline characteristics by treatment assignment. The code uses `tapply` to calculate means and variances across treatment groups (Control and Treated) and then calculates the difference between them.
- Search Bar:** Located at the top left, it includes fields for "balance at" and search options like "Next", "Prev", "All", "Replace", and "Replace All".
- Environment View:** Shows a preview of the generated table titled "Balance at Baseline". The table has columns for Control, Treated, Std. Diff., and Std. Diff. for variables like Girl, Age, Highest, Grade, and Enrolled.
- Console View:** Shows the R code being run, with the output indicating the code is running.
- Terminal View:** Shows the command used to generate the table.
- Project View:** Shows the current project is "Lecture1.qmd".

# R in Econ 2136

- 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 2136 topics will include:

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

# Econ 2136 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 2136 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 2136: Prerequisites

- Pre-Req:
  - Econ 2135: Introduction to Probability and Statistics, or S&DS 2410 and 2420.
  - What if you haven't taken Econ 2135 or S&DS 2410+2420?

# Econ 2136: Prerequisites

- **Math Pre-Req:**
  - multivariate calculus as acquired by having already taken MATH 1200 or equivalent course,
  - linear algebra, as acquired by having already taken MAT 2250 or 2260 or equivalent course.

# Lectures:

- Lectures will primarily use blackboard, will sometimes be accompanying handouts.
- We will sometimes live-code in R to analyze real data and help you with your problem sets.
- You are expected to attend lectures.  
I will call on students.

# Course Webpages

- **Canvas webpage**

<https://yale.instructure.com/courses/114504>

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

- **Ed Discussion webpage:**

<https://edstem.org/us/courses/90725/discussion/>

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

Assignments	Share of Course Grade
Online Quizzes	5%
Problem Sets	15%
Midterm	35%
Final	45%

# On-Line Quizzes:

- Posted on Fridays:
  - to the course webpage.
  - most weeks,
  - remain live for 48 hour once posted,
  - 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.
  - Due dates are strict.<sup>1</sup>
  - The lowest problem set score will be dropped.
1. Exceptions only with note from residential dean.

# Problem Sets

- You may work in groups of up to four students on the problem sets.<sup>1</sup>
  - However, you must turn in your own assignment and indicate on your submission the other members of the group.
1. If you don't have a group, but would like us to help match you to other students, please let the TA or me know.

# Exams

- Midterm:
  - in-class, Thursday February 26 (date tentative)
- Final:
  - Tuesday, May 5, 2026 at 7pm.
- 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 quiz goes live Friday January 23.
- First problem set due Thursday February 5 at 2:30pm.
  - Will include theoretical questions as well as use R.

# References

- Donoho, David. 2017. “50 Years of Data Science.” *Journal of Computational and Graphical Statistics* 26 (4): 745–66. <https://doi.org/10.1080/10618600.2017.138471>
- 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>.
- Samuelson, Paul A, Tjalling C Koopmans, and J RICHARD N Stone. 1954. “Report of the Evaluative Committee for Econometrica.” JSTOR.