

# ECON7225: Machine Learning and Econometrics

Jui-Chung Yang

Spring, 2026

E-mail: [jcyang1225@ntu.edu.tw](mailto:jcyang1225@ntu.edu.tw)

Class Hours: Th 9:10-12:10

Class Room: College of Social Sciences 605

Web: [sites.google.com/site/juichungyang](https://sites.google.com/site/juichungyang)

Office Hours: Th 12:20-13:10

Office: College of Social Sciences 755

---

## Course Description

This graduate-level course explores the intersection of **machine learning** and **econometrics**, bridging predictive analytics with statistical inference and causal modeling. Since Varian (2014), the conventional distinction has been that **machine learning** focuses on using data to predict variables, while **econometrics** applies statistical methods for prediction, inference, and causal analysis in economic relationships. However, more recent perspectives suggest that machine learning represents a new wave of **nonparametric** statistical and econometric techniques (Chernozhukov, 2016). Moreover, machine learning has introduced exciting advancements in econometric applications, such as estimating **conditional average treatment effects** (Athey and Imbens, 2016) and serving as the first stage in **two-stage estimation** (Chernozhukov et al., 2018).

In this course, we – **a group of well-trained econometricians** – will explore machine learning together. On the **estimation and prediction** side, we will cover techniques like **Lasso, neural networks, and random forests**. On the **inference and causal modeling** side, we will discuss inference methods after applying machine learning, as well as approaches inspired by computational statistics, including **post-model selection inference** and **large-scale hypothesis testing**. More importantly, we will explore open research questions in inference, and you are encouraged to engage with these challenges as potential research projects.

Since machine learning relies on computational methods, proficiency in at least one **statistical programming language** – such as **Python or R** – is required. You are welcome to use any other language of your choice, provided it allows you to implement the techniques covered in this course.

## Materials

- Hansen, Bruce (2022). *Econometrics*, Princeton University Press.  
[press.princeton.edu/books/hardcover/9780691235899/econometrics](https://press.princeton.edu/books/hardcover/9780691235899/econometrics)
- Efron, B., and Hastie T. (2016). *Computer Age Statistical Inference: Algorithms, Evidence and Data Science*, Cambridge.  
[hastie.su.domains/CASI\\_files/PDF/casi.pdf](https://hastie.su.domains/CASI_files/PDF/casi.pdf)
- Gareth, J., Witten, D., Hastie, T., and Tibshirani, R. (2021). *An Introduction to Statistical Learning: With Applications in R*, Second Edition, Springer.  
[www.statlearning.com/](https://www.statlearning.com/)
- Chan, F., and Matyas, L. (2022). *Econometrics with Machine Learning*, Springer.  
[link.springer.com/book/10.1007/978-3-031-15149-1](https://link.springer.com/book/10.1007/978-3-031-15149-1)

## Grading

One midterm exam (April 14, 40%). One final presentation (June 9, 60%).

## Schedule

Week 1 Resampling Methods

Week 2 Resampling Methods

Week 3 Nonparametric Methods

Week 4 Nonparametric Methods

Week 5 Model Selection and Stein Shrinkage

Week 6 Model Selection and Stein Shrinkage

Week 7 Ridge Regression and Lasso

Week 8 Midterm

Week 9 Regression Trees, Random Forests, and GRF

Week 10 Regression Trees, Random Forests, and GRF

Week 11 Double/Debiased Machine Learning

Week 12 Double/Debiased Machine Learning

Week 13 Financial Machine Learning

Week 14 Financial Machine Learning

Week 15 Special Topics

Week 16 Final Presentation