

Causal Machine Learning for Economic Analysis

Department of Economics

ECON 4450 / ECON 5450 - Winter 2026/2027

Instructor: Prof. Mutlu Yuksel, Dalhousie University, Department of Economics

Class Location: Studley LSC – Common Area C210

Class Time: Monday and Wednesday, 10:05 AM – 11:25 AM

Office hours: To be announced on Brightspace, or by appointment

Office: 6220 University Avenue, B24

E-mail: mutlu@dal.ca

Is This Course for You?

Machine learning is often associated with prediction, but machine learning methods are increasingly being used to answer causal questions as well. **Does a policy, program, or intervention work? Who benefits from it, and who does not? Which workers, patients, customers, or communities are most likely to respond to a particular policy, program, or treatment? How can we use data to improve targeting and decision-making?** This course is designed for students who want to expand their existing applied economics toolkit by learning how contemporary statistical and machine learning methods can complement traditional econometric approaches to answer these questions.

You do not need previous experience with machine learning, and you do not need to be a computer programmer. You should, however, be comfortable with the basic ideas of regression and be willing to work with data and code. If you have taken an introductory econometrics course and understand how to interpret a regression, you have the background needed to start. We will build the machine-learning tools step by step, with an emphasis on intuition, application, and interpretation rather than mathematical proofs or computer-science theory.

The course combines causal inference and machine learning using applications from economics, public policy, business, health, and the social sciences. You will learn how methods such as regularization, double machine learning, and causal forests can complement the econometric tools you already know. By the end of the course, you should be able to understand when and why these methods are useful, implement several of them in applied settings, and interpret their results critically. R will be used in class and assignments, but prior expertise in R is not required. The course may be particularly useful for students interested in applied empirical research, causal inference, policy evaluation, business analytics, or related quantitative work.

This is a hands-on course with no midterm or final exam. Assessment is based primarily on applied assignments in which you use the methods to answer research questions. Undergraduate and graduate students attend the same classes; graduate students complete additional work connecting the methods to recent research.

If machine learning sounds interesting but somewhat intimidating, the course is designed to make these methods accessible without removing the analytical and empirical work required to use them well.

Course Description

This course bridges machine learning (ML) techniques and causal inference methods to address primarily economic questions. Students will explore key ML concepts such as penalized regression, regularization, cross-validation, and hyperparameter tuning, along with tree-based methods, and study their implementation within the context of economic research. The course integrates causal inference strategies grounded in the Rubin Causal Framework, focusing on ML-based approaches for both selection on observables and unobservable confounding variables. Emphasis is placed on estimating both conditional average treatment effects (CATE) and heterogeneous treatment effects (HTE) using advanced techniques like double/debiased ML, causal forests, and meta-learners. Practical skills are developed through R-based labs, equipping students to combine ML with standard econometric tools for robust causal analysis and informed decision-making in economic applications.

Course Prerequisites: For ECON 4450: ECON 3339.03 with a minimum grade of C; or ECON 3338.03 or STAT 3340.03 with a minimum grade of B+ and permission of the instructor. For ECON 5450: Good standing graduate student in Economics.

Learning Objectives

- **Familiarize with ML concepts** by implementing supervised ML methods alongside traditional econometric techniques to enhance predictive accuracy and causal analysis.
- **Estimate and interpret treatment effects** by applying penalized regression, double/debiased ML, causal forests, and meta-learners to derive both average and heterogeneous (conditional) treatment effects.
- **Address selection bias effectively** using ML-based approaches for both observable and unobservable confounders, ensuring credible causal inferences.
- **Apply covered methods to real-world datasets** to answer practical economic, social, and health science questions, connecting theoretical concepts with applied analysis

Student Resources

Causal Inference and Machine Learning in Economics, Social, and Health Sciences, Mutlu Yuksel and Yigit Aydede, Routledge, 2025. Freely available online at www.causalmlbook.com

Course Format

The course is in-person and meets twice a week, on Mondays and Wednesdays from 10:05 AM to 11:25 AM in Studley LSC Common Area C210. Class sessions combine lectures on theoretical concepts and methods with hands-on implementation using R (or optionally Stata or Python; instructor support is provided for R). Regular attendance is recommended to build both theoretical understanding and the practical skills essential for the assignments (and the review papers in ECON 5450). ECON 4450 (undergraduate) and ECON 5450 (graduate) are taught together in the same class; ECON 5450 students have additional requirements (see Assessment).

Lab Sessions: Details of the lab component (time, location, and support) will be announced on Brightspace.

Course Materials: All course material will be available online on Brightspace.

Course Content

Week 1: Introduction & Context

- **Machine Learning vs. Traditional Econometrics:** Contrasting objectives—prediction vs. structural inference.
- **Correlation vs. Causation:** Policy examples, program evaluation, business analytics.
- **Course Overview:** Structure, software tools (primarily R, optional Python/Stata).
- **Lab (R):** Simple correlation/regression exercise (e.g., using `lm()`), illustrating naive predictive modeling vs. causal interpretation.

Week 2: Potential Outcomes & Causal Concepts + Estimation vs. Prediction Error

- **Rubin Causal Model (RCM):** Potential outcomes, SUTVA, unconfoundedness.
- **Parameters of Interest:** ATE, ATT, CATE, why economists need them.
- **Estimation vs. Prediction Error:** The difference between minimizing MSE for forecasting vs. unbiased effect estimation.
- **Lab (R):** Simulate a simple RCM scenario; compare “predicted outcome error” vs. “estimation bias” for a treatment effect.

Week 3: ML Fundamentals (Part I) — Bias-Variance, Parametric vs. Nonparametric

- **Bias-Variance Tradeoff:** Overfitting, underfitting in ML/econometrics.
- **Parametric vs. Nonparametric:** Linear models vs. flexible ML (kernels, trees).
- **Regularization:** The intuition behind penalizing model complexity.
- **Lab (R):** Fit polynomial regressions or kernel regressions; examine in-sample vs. out-of-sample error via `caret` or `tidymodels`.

Week 4: ML Fundamentals (Part II) — Cross-Validation & Model Selection

- **Train/Test Splits & K-Fold Cross-Validation:** Minimizing overfitting, ensuring generalizability.
 - **Hyperparameter Tuning:** Grid search, random search, cross-validation scores.
 - **Model Selection Criteria:** AIC/BIC vs. out-of-sample MSE, economic interpretability.
 - **Lab (R):** Use `caret` or `tidymodels` to systematically search hyperparameters.
-

Week 5: Linear Penalized Models

- **Multiple Linear Regression Recap:** Interpretation, limitations.
- **Ridge, LASSO, Elastic Net:** Motivations, penalties, L1 vs. L2.
- **High-Dimensional Covariates:** Why LASSO aids variable selection.
- **Lab (R):** Fit LASSO (glmnet) on an economic dataset; interpret selected features.

Winter Break: February 15–19 (No Classes)

Week 6: Matching, Propensity Score, Doubly Robust Estimators

- **Propensity Score Estimation:** Logistic regression vs. random forest/GBM-based propensity.
- **Matching & IPTW:** Diagnostics (balance checks, overlap).
- **Augmented Inverse Probability Weighting (AIPW) & Doubly Robust Estimators:** Combining outcome regression with propensity weighting for robustness.
- **Lab (R):** Use MatchIt or twang for ML-based propensity estimation and AIPW. Demonstrate matching/weighting on real or synthetic data.
- **Objective:** Show how flexible ML can enhance confounder control and robustness in observational data for causal analysis.

Week 7: Double/Debiased Machine Learning (DML)

- **Orthogonal Moments & Cross-Fitting:** Importance for high-dimensional settings.
- **DML for ATE:** ML-based nuisance function estimates (propensity, outcome).
- **Common Pitfalls:** Hyperparameter tuning, sample splitting, finite-sample issues.
- **Lab (R):** Implement double ML using the DoubleML package.
- **Objective:** Show how combining ML-based nuisance estimation with orthogonal moment conditions yields robust causal estimates.

Week 8: Nonlinear & Tree-Based Methods

- **Decision Trees (CART):** Splitting criteria (MSE, Gini), pruning.
- **Random Forests:** Bagging, feature subsampling, interpretability measures.
- **Gradient Boosting:** Ensemble approach, common hyperparameters.
- **Lab (R):** Fit random forest or XGBoost; compare performance vs. parametric models using randomForest, ranger, or xgboost.

Week 9: ML Methods in Selection on Unobservable Confounding Factors

- **Classic IV Setup:** Relevance, exogeneity, 2SLS.
- **ML-based IV:** LASSO IV, random forests for the first stage, double ML for IV.
- **Applications:** Education and labor market examples.
- **Lab (R):** Implement partial LASSO or random forest first-stage in R (glmnet or ranger) combined with IV estimation (AER or ivreg).

- **Objective:** Demonstrate how ML can flexibly handle the first stage in an IV context, improving instrument relevance.

Week 10: Heterogeneous Treatment Effects (HTE)

- **Why HTE Matters:** Policy targeting, subgroups, microfinance/labor examples.
- **Estimating HTE:** Interactions in linear models vs. tree-based approaches.
- **Challenges:** Overfitting, sample size limitations, interpretability.
- **Lab (R):** Estimate HTE using flexible ML methods; interpret subgroup effects.
- **Objective:** Focus on detecting and interpreting treatment effect heterogeneity for targeted policy insights.

Week 11: Causal Forests

- **Causal Trees vs. Causal Forests:** Honest splitting, ensemble approach for effect heterogeneity.
- **Interpretability:** Feature importance, partial dependence plots.
- **Lab (R):** Fit a causal forest (using the grf package) to detect subpopulations with different treatment effects.
- **Objective:** Highlight tree-based algorithms for uncovering nuanced causal relationships.

Week 12: Personal Causal Effects — Meta-Learners & Personalized Treatment Effects

- **Meta-Learners (S, T, X) Overview:** Estimation logic for personalized effects.
- **Personalized Treatment Effects:** Applications in microfinance, healthcare, labor markets.
- **Challenges:** Overfitting, sample size limitations, interpretability.
- **Lab (R):** Implement meta-learners (S-, T-, and X-learners built from grf, glmnet, and ranger) and estimate personalized treatment effects.
- **Objective:** Conclude the course by highlighting methods for estimating individual-level causal effects, focusing on actionable insights for policy.

Assessment

The course includes 5 bi-weekly assignments, each focused on applying methods covered in class to real-world economic data. Datasets will be provided, and students will implement analyses using R (or optionally Stata or Python) to answer specific research questions. Assignments will combine hands-on coding with sections requiring interpretation and assessment of results. Many assignments will be based on methodologies used in published papers from top economics journals, allowing students to connect theory with practice. Assignments are due at 11:59 PM on the dates listed below. Assignment weights differ between the two courses and are shown as ECON 4450 / ECON 5450.

- **Assignment 1: ML Fundamentals, Bias-Variance Tradeoff & Cross-Validation**
 - **Due Date:** End of Week 4 (Thursday, February 4, 11:59 PM)
 - **Weight:** 20% (ECON 4450) / 12% (ECON 5450)

- **Description:** Implement and interpret the bias-variance tradeoff and use cross-validation for model selection using R.
-

- **Assignment 2: Penalized Regression & Regularization**

- **Due Date:** End of Week 6 (Friday, February 26, 11:59 PM)
 - **Weight:** 20% (ECON 4450) / 12% (ECON 5450)
 - **Description:** Apply LASSO and Ridge regression for variable selection.
-

- **Assignment 3: Matching, Propensity Scores, Doubly Robust Estimators & Double/Debiased ML**

- **Due Date:** End of Week 8 (Friday, March 12, 11:59 PM)
 - **Weight:** 20% (ECON 4450) / 12% (ECON 5450)
 - **Description:** Estimate treatment effects using matching and AIPW, and implement DML for ATE estimation with R.
-

- **Assignment 4: Tree-Based Methods, Selection on Unobservables & ML-based IV**

- **Due Date:** End of Week 10 (Thursday, March 25, 11:59 PM)
 - **Weight:** 20% (ECON 4450) / 12% (ECON 5450)
 - **Description:** Fit and compare tree-based models with parametric benchmarks, and use ML techniques for IV estimation in a selection-on-unobservables context.
-

- **Assignment 5: Causal Forests & Personalized Causal Effects**

- **Due Date:** End of Week 12 (Friday, April 9, 11:59 PM)
 - **Weight:** 20% (ECON 4450) / 12% (ECON 5450)
 - **Description:** Estimate heterogeneous treatment effects using causal forests and meta-learners.
-

ECON 5450 Only: Review Papers

In addition to the five assignments, ECON 5450 students write 2 review papers, due at the end of Weeks 9 and 11 and worth 20% each. For each review paper, students will consult recent articles from top economics journals or NBER papers, review the methods used, and write a 2-page discussion paper comparing these methods. The review papers aim to deepen students' understanding of how machine learning and causal inference methods are applied in contemporary economic research. Approval of article choices by the instructor is required before proceeding.

- **Review Paper 1: Method Comparison in Recent Economic Literature (Selection on Observables)**

- **Due Date:** End of Week 9 (Friday, March 19, 11:59 PM)
- **Weight:** 20%
- **Description:**
 - Identify recent articles from top economics journals or NBER papers that use penalized regression, matching/AIPW, or DML for causal inference.
 - Review and compare the methods used, discussing their strengths and limitations.
 - Write a 2-page discussion paper; article choices must be approved by the instructor before you begin
- **Review Paper 2: Method Comparison in Recent Economic Literature (Heterogeneous Treatment Effects)**
 - **Due Date:** End of Week 11 (Friday, April 2, 11:59 PM)
 - **Weight:** 20%
 - **Description:**
 - Identify recent articles that estimate heterogeneous treatment effects using causal forests, meta-learners, or related methods.
 - Compare how the papers estimate and interpret heterogeneity, discussing strengths and limitations.
 - Write a 2-page discussion paper; article choices must be approved by the instructor before you begin.

Total Weight

Component	ECON 4450	ECON 5450
Assignments (5)	100% (5 × 20%)	60% (5 × 12%)
Review Papers (2)	None	40% (2 × 20%)
Total	100%	100%

ECON 4450: Conversion of numerical grades to final letter grades follows the

[Dalhousie Grade Scale](#)

C+ (65-69)	D (50-54)
C (60-64)	F (0-49)
C- (55-59)	

ECON 5450: Conversion of numerical grades to final letter grades follows the

[Dalhousie Grade Scale](#)

A+ (90-100)	B+ (77-79)	
A (85-89)	B (73-76)	F (0-69)
A- (80-84)	B- (70-72)	

Term Calendar and Key Dates

Classes meet on Mondays and Wednesdays. Weeks are numbered 1 to 12 consecutively; the Winter Break week is not numbered. Assignment due dates are shown in the schedule below.

Week	Class dates	Notes
1	Mon, Jan 11, Wed, Jan 13	Classes begin (Monday, January 11)
2	Mon, Jan 18, Wed, Jan 20	
3	Mon, Jan 25, Wed, Jan 27	
4	Mon, Feb 1, Wed, Feb 3	Assignment 1 due Thu, Feb 4. Munro Day, Friday, February 5: University closed
5	Mon, Feb 8, Wed, Feb 10	
—	Feb 15 – Feb 19	WINTER BREAK: no classes
6	Mon, Feb 22, Wed, Feb 24	Assignment 2 due Fri, Feb 26
7	Mon, Mar 1, Wed, Mar 3	
8	Mon, Mar 8, Wed, Mar 10	Assignment 3 due Fri, Mar 12
9	Mon, Mar 15, Wed, Mar 17	Review Paper 1 (ECON 5450) due Fri, Mar 19
10	Mon, Mar 22, Wed, Mar 24	Assignment 4 due Thu, Mar 25. Good Friday, March 26: University closed
11	Mon, Mar 29, Wed, Mar 31	Review Paper 2 (ECON 5450) due Fri, Apr 2
12	Mon, Apr 5, Wed, Apr 7	Last class Wednesday, April 7. Assignment 5 due Fri, Apr 9

Please note: there is no class on April 13. Friday, February 5 (Munro Day) and Friday, March 26 (Good Friday) are University closures and do not affect Monday/Wednesday classes.

Course Policies on Missed or Late Academic Requirements

Students are strongly encouraged to submit assignments and review papers on time to keep up with the course's pace. Late submissions will be accepted up to 3 days past the deadline with a penalty of 10% per day. After 3 days, submissions will not be accepted without prior approval from the instructor.

If serious issues prevent timely submission, students can discuss their situation with the instructor. In such cases, with approval, a missed assignment may be substituted with a replication study based on an article from the course book, which contains numerous papers with available datasets. The replacement task will involve implementing the missing assignment's methods on a dataset from an agreed-upon paper aligned with the student's research interests.

Course Policies related to Academic Integrity

Students may work in pairs for individual assignments but must clearly state their contributions and sign a statement confirming them. Each student must submit their own write-up, even if analysis was done collaboratively.

All sources, including large language models (LLMs) like ChatGPT, are allowed but must be cited explicitly. Failure to acknowledge sources or contributions will be treated as a breach of academic integrity.

Rubric for Review Papers (ECON 5450 Only; Total: 100 Points)

Note: The review papers should be written in an essay style, presenting a clear and cohesive assessment and discussion rather than a simple bullet-point response.

1. Introduction and Article Selection (20 Points)

- **Clarity and Relevance (10 Points):**
 - Introduces the selected articles clearly, explaining why they are relevant to the methods covered in the course.
 - Provides context for the main research questions each article addresses.
- **Research Questions (10 Points):**
 - Clearly presents the main research questions explored in each article.
 - Explains the significance of these questions in the broader economic context.

2. Methodology (30 Points)

- **Description of Methods (15 Points):**
 - Provides a detailed explanation of the econometric and ML methods used in each paper, such as LASSO, causal forests, and double/debiased ML.
 - Clearly distinguishes between traditional econometric methods and ML-based approaches, explaining the rationale for using ML methods.
- **Comparison with Base Models (15 Points):**
 - Discusses how ML methods improved the estimation of coefficients compared to baseline models like OLS or standard IV.
 - Provides specific examples of how ML methods addressed limitations of traditional approaches, such as multicollinearity, high dimensional data or selection bias.

3. Findings and Results (30 Points)

- **Summary of Results (15 Points):**
 - Summarizes the main findings of each paper, focusing on key coefficients and causal effects.
 - Clearly explains the implications of these results in the context of the research questions.
 - **Impact of ML Methods (15 Points):**
 - Discusses how ML methods improved the precision, robustness, or interpretability of the results.
 - Uses specific examples to show how ML methods refined coefficient estimates or provided more reliable causal inferences.
-

4. Discussion and Critical Analysis (20 Points)

- **Strengths and Limitations (10 Points):**
 - Critically assesses the strengths and weaknesses of the methods used in each paper.
 - Considers issues like overfitting, interpretability, and potential biases in the context of the results.
- **Suggestions and Future Directions (10 Points):**
 - Suggests other ML methods that could potentially improve the analysis.
 - Discusses the limitations of the ML methods used and possible ways to address them in future research.

Total: 100 Points

Syllabus Appendix

(Formerly University Policies and Statements)

Here is the link to the new “Syllabus Appendix” to guide students towards supports, university policies, and guidelines.

[Syllabus Appendix: Student Supports & University Statements, Policies, and Guidelines](#)