

# Dingyi Li

Email: dl922@cornell.edu

Website: dingyili93.github.io

Last Update: April, 2026

## EDUCATION

---

- **Cornell University**  
Ph.D. in Applied Economics and Management 2019-2026 (Expected)  
M.S. in Applied Economics and Management 2017-2019
- **Renmin University of China**  
B.A. in Economics 2012-2016  
B.S. in Mathematics 2012-2016

## RESEARCH INTERESTS

---

Primary Field: Econometrics

Secondary Fields: Applied Econometrics, Machine Learning, Empirical Industrial Organization

## WORKING PAPERS

---

- “Efficient and Doubly Robust Estimation of the Average Treatment Effect with Confounders Missing Not at Random.” **Job Market Paper**

**Abstract:** When confounders are missing not at random, the usual conditional independence assumption based on the observed data can fail, and the causal parameter is no longer identifiable. We consider scenarios where confounder missingness depends on itself but is independent of the outcome, conditional on the treatment and confounders (Yang et al., 2019). We present an efficient and doubly robust estimator for the average treatment effect (ATE) that addresses missing confounders through novel propensity scores and conditional treatment effects, which account for patterns of missingness. This adjustment employs two integral equations: one has appeared in the literature, whereas the other appears to be novel. Solving integral equations relies on the completeness of the full-data distribution in the outcome. Moreover, we propose a low-rank assumption on the missingness mechanism, extending the analysis to empirical settings with limited outcome variation, such as binary outcomes. Our estimator is doubly robust, remaining consistent if either integral equation holds, and rate doubly robust, achieving  $\sqrt{n}$ -consistency when both propensity scores and conditional average treatment effects converge at the fourth-root rate. We then derive the semiparametric efficiency bound for the ATE under the outcome-independent missingness condition and show that our estimator achieves the bound. We evaluate our estimator with simulations and three empirical applications.

- “On the Testability of Anchor Words in Topic Models,” with Simon Freyaldenhoven, Shikun Ke, and José Luis Montiel Olea. **Conditionally Accepted at Quantitative Economics.**

**Abstract:** Topic models are a simple and popular tool for the statistical analysis of textual data. Their identification and estimation is typically enabled by assuming the existence of *anchor words*; that is, words that are exclusive to specific topics. In this paper we show that the existence of anchor words is statistically testable: There exists a hypothesis test with correct size that has nontrivial power. This means that the anchor-words assumption cannot be viewed simply as a convenient normalization. Central to our results is a simple characterization of when a column-stochastic matrix with known nonnegative rank admits a *separable* factorization. We test for the existence of anchor words in two different data sets derived from monetary policy discussions in the Federal Reserve and reject the null hypothesis that anchor words exist in one of them.

- “Identification and Estimation of Finite Mixtures of Multinomial Logit Models.”

**Abstract:** Finite mixtures of multinomial logit models can be used to capture consumer choice heterogeneity across multiple markets when only aggregate consumer choices per market are available. A motivating example is a nested logit where the composition of each mixture component (each nest of alternatives) is unknown a priori. We show that in order to identify these models, it suffices to require that each mixture component includes at least two component-exclusive alternatives. We refer to our assumption as the *pure-alternatives* condition, and we argue it is a natural extension of the *anchor-word* assumption used commonly in nonnegative matrix factorization problems in machine learning. Our identification result enables a consistent two-step estimator as the number of consumers, markets, and alternatives grow large. Applying this framework to the U.S. vehicle market, we find that consumer heterogeneity does not yield

substitution patterns between electric and internal combustion engine vehicles, suggesting consumer segments are distinctly aligned with specific vehicle types without crossover substitution.

## WORK IN PROGRESS

---

- “Systemic Risk, FOMC Statements, and Monetary Policy Shocks: A New Topic Model to Associate Text with Metadata,” with Shawn Mankad.
- “Pollution Avoidance and Willingness-to-Pay: Evidence from Travel Mode Choice in Beijing,” with Shanjun Li and C.-Y. Cynthia Lin Lawell.
- “Weak Sparse Models and Methods for Instrumental Variables.”
- “Casual Inference with Missing Not at Random and Unobserved Confounders under Multiple Outcomes,” with Peng Ding.

## TEACHING EXPERIENCE

---

- TA for AEM 2300 International Trade and Finance, Cornell University 2022-2024, 2026 Spring
- TA for AEM 3310 Introduction to Business Regulation, Cornell University 2023 Fall
- TA for AEM 4110 Introduction to Econometrics (4.9/5.0), Cornell University 2021 Fall
- TA for AEM 6120 Applied Econometrics (4.4/5.0), Cornell University 2020 Fall

## AWARDS

---

- Graduate Research Fellowship, Cornell University 2025
- Flaim and Neenan Family Fellowship, Cornell University 2025
- Graduate Research Fellowship, Cornell University 2024
- Ashley Graduate Fellowship, Cornell University 2023
- Transportation Networks and Smart Mobility Scholarship, Massachusetts Institute of Technology 2022
- Edward and Janet Heslop Fellowship, Cornell University 2021-2022
- Academic Excellence Scholarship (awarded annually to the top 3 students in the honors program), Renmin University of China 2012-2015

## PRESENTATIONS

---

- University of Mannheim, University of Macau, Cal Poly State University, University of Washington, Queen’s University 2026
- CES North American Conference (University of Michigan), CLIMB Retreat (UC Berkeley) 2025
- New York Camp Econometrics XVIII (Syracuse University), North American Summer Meeting of the Econometric Society (Vanderbilt University), Hong Kong University of Science and Technology, International Association for Applied Econometrics (Xiamen University; University of Macedonia), ESIF Economics and AI+ML Meeting (Cornell University) 2024
- Asia Meeting of the Econometric Society, EREE Student Workshop (Department of Agricultural and Resource Economics, UC Berkeley), Econometrics Student Conference (Department of Economics, UC Berkeley) 2023
- North American Summer Meeting of the Econometric Society (Tsinghua University) 2022
- World Conference of Spatial Econometrics Association (Tokyo, Japan) 2021