Xinhui Sun

Urbana, IL | 217-721-9483 | xinhuis2@illinois.edu | LinkedIn
Personal Website: xinhuis.github.io

Education

Ph.D. in Economics, University of Illinois at Urbana-Champaign, May 2025 (expected)

M.S. in Policy Economics, University of Illinois at Urbana-Champaign, Dec 2019
Core Courses: Advanced Data Analysis, Machine Learning, Time Series Analysis, Industrial Organization

B.S. in Economics, University of International Business and Economics (UIBE), Jun 2018
Core Courses: Linear Algebra, Probability Theory, Investment Analysis, Financial Derivatives

Exchange student, University of California, Berkeley, Jan 2017 - May 2017
Core Courses: Stochastic Process, Statistical Methods of Data Science

Research Interests

Environmental Economics, Behavioral Economics, and Causal Inference

Working Experience

Gies College of Business, University of Illinois at Urbana-Champaign Illinois, United States
Research Assistant to Professor Tatyana Deryugina Aug 2022 – present

- Formatted and processed **272m+ geo-spatial data** and matched it with ZIP Code Tabulation Area.
- Performed InMAP simulations to distinguish PM2.5 that directly emitted by a source from PM2.5 that is formed from SO2.
- Cleaned weather and labor-related data from 1969-2020.
- Established panel fixed effect models to estimate the marginal effect of climate.

Department of Economics, University of Illinois at Urbana-Champaign Illinois, United States
Research Assistant to Professor Eunyi Chung June 2021- Aug 2022

- Assisted in developing two adjusted permutation tests, which are more robust than the traditional permutation test. Wrote corresponding **R codes** and Packages.
- Implemented a **Regression Discontinuity** application to this framework.
- Compared the performance of these two adjusted permutation methods in terms of type 1 error and power by running Monte Carlo Simulations using R.

Risk Management Department, Bank of Communications Shanghai, China
Data Analyst Intern Jun 2017 - Aug 2017

- Sorted and selected mortgage loan data using **SQL** and conducted data cleaning and feature selection.
- Established Logistic Regression to Predict the Probability of Default of each client.
- Utilized Decision Tree and Support Vector Machine to realize client classification and updated the database by red-flagging high-risk clients.
- Acquired up-to-date housing price data from public networks using a web crawler and analyzed the risks of housing mortgage loans via statistical changes in housing prices.

Work in Progress

Air Pollution, Avoidance Behaviors, and Daily Activities: Evidence from the U.S.
Paper in Progress

- Collected and formatted **26m+** mobile phone location data, and match each location with the nearest monitor group data using latitude and longitude.
- Performed **k-means** to cluster all air pollution data into spatial groups based on their location.

- Used change in wind direction as an **Instrumental Variable** to resolve the endogeneity problem of air pollution, and derived the causal effect of air pollution on time spent outdoors.

Go with the wind: Polluters' Strategic Response to Wind Directions
Paper in Progress

- Collected hourly emissions data from 1995 to 2022 using Clean Air Market **API**.
- Calculated the relative direction and distance of each plant to its nearest monitoring site.
- Found polluters emit more pollution on days when they are downwind of monitoring sties.
- Identified a potential mechanism: plants temporary turn off their emission control equipment to save cost.

Ensemble Machine Learning Model for Image Classification
Group project for STAT542 Statistical Learning

- Performed **unsupervised learning**, including k-means and density-based clustering, on the training data with 60k+ observations of the Fashion-MNIST dataset.
- Compared the performance of various multi-class **classifications**, including Linear and Quadratic Discriminant Analyses, mean-based classification, multi-class SVM, and random forest.
- Built the ensemble model by using the histogram-based **gradient boosting** and incorporating results from previous models and achieved 90.5% accuracy.

Teaching Experience

ECON 102: Microeconomic Principles Fall 2021, Spring 2022
Teaching Assistant

The Economic of the Firm (EMBA at the University of Warsaw) Summer 2019
Teaching Assistant to professor Hadi Esfahani

Econ 528: Microeconomics for Business Spring 2019, Summer 2019
Course Assistant

Awards

Cleo Fitzsimmons Award (awarded to the student with the highest GPA), University of Illinois 2022
Graduate Fellowship, University of Illinois 2020

Skills

Language: Mandarin (Native), English (Proficient), French (Basic)
Programming: R (dplyr, ggplot2, sparklyr), Python (numpy, pandas, scikit-learn), STATA, Git, Shell
Machine Learning: Linear/Logistic Regression, KNN, Support Vector Machine, Decision Tree, Random Forest, etc.