Shaoyu Liu

(+1) 5105992806 | sl4640@columbia.edu

421 West 118th Street, Room 34B, New York, NY 10027 USA

Education

Columbia University in the City of New York **2019-Present**

Ph.D. Candidate in Economics (Applied Microeconomics; Econometrics; Machine Learning)

Key Coursework: Labor Economics; Public Economics; Micro-econometrics; Industrial Organization; Natural Language Processing; Topics on Machine Learning; Advanced Algorithms

University of California, Berkeley **2015-2019**

B.A in Economics and Statistics (*magna cum laude*, GPA 3.87/4.0)

Project Experience

The Impact of High-level Scientists Recruitment Program on Research Productivity *Working Paper*

with Elliott Ash, David Cai and Mirko Draca

• Automated scraping and cleaning of 300 GB JSON data from Scopus publication database using Python on computer cluster; developed a PostgreSQL database to facilitate efficient data storage and analysis

• Performed data wrangling, data preprocessing and measured research similarity of authors using k-nearest neighbors algorithm on high dimensional large publication datasets

• Showed positive spillover effects of scientists re-emigration on publications and citations through event study and difference-in-difference methods

Estimating Policy Impact on Economics Publications in China *Work in Progress*

• Detected latent topic distributions and ideologies of economists from text data by applying topic modeling algorithms (Latent Dirichlet Allocation and Ideal Points models) on a corpus of 100,000 economics papers

• Measure heterogeneous impact of government policy shocks on economists productivity and research topic distribution through a difference-in-difference design

Basic Search Engine for Academic Articles *Project*

• Created a search engine of Scopus database article abstracts that enable fast similarity search by computing pretrained sentence embeddings based on SBERT

• Implemented state-of-the-art ANN algorithms, e.g. locally sensitive hashing (LSH), neural LSH and Hierarchical Navigable Small World algorithms that speed up similarity search on high dimensional space

UC Berkeley, Columbia University

Research Assistant to Profs. Gerard Roland, Guo Xu, and Miguel Urquiola **2016-2019; 2019; 2021**

• Automated web scrapping on 12 university websites and analyzed faculty salary and department data

• Collected, cleaned and merged over 100,000 rows of precipitation GIS data with household survey data using R; performed exploratory data analysis and visualization to identify key metrics for further analysis

• Estimated demand elasticities in U.S. automobile market using nested logit and random coefficient logit models

• Conducted detailed literature review on identification strategies of recent relevant papers, and used instrumental variable approach to identify the causal relationship between income and survey attitudes

Awards and Fellowships

Dean's Fellowship, Columbia University, 2019

Best Paper Award, Carroll Round Conference for International Economics, Georgetown University, 2019

Institute of International Studies Undergraduate Travel Grant, UC Berkeley, 2019

Dean's List, UC Berkeley, 2015

Presentations and Publications

2021 China Meeting of the Econometric Society, China, 2021 (scheduled); *Economics of Science and Innovation Workshop*, Barcelona Graduate School of Economics, Spain, 2021 (scheduled); *Applied Micro Method Colloquium*, Columbia University, 2021; *18th Carroll Round Conference for International Economics*, Georgetown University, 2019

Institutions and Missing Women: Evidence from China, *Journal of Politics and Society*, 15(3), 107-133, 2019

Skills and Personal

Skills: Python(Pandas, Numpy, SKlearn, PyTorch), SQL(proficient), STATA(familiar), R(familiar)