NBER Digitization Tutorial: Student Application

CONTACT INFORMATION

Katherine Hoffmann Pham
44 West 4th Street, KMC 8-180B
New York, NY 10012
650-823-2853
khof@nyu.edu

COURSE OF STUDY

PhD Candidate, Information Systems (4th year)
Stern School of Business, NYU

- **Primary focus:** Economics track
- **Secondary focus:** Technical track
- **Supervisors:** Arun Sundararajan, Panos Ipeirotis

DISSERTATION

My dissertation proposal (in progress) focuses on the future of urban mobility. My first essay uses a natural experiment, subway disruptions in New York City, to explore substitution between public transportation, ridesharing, and other transportation modes. For my second essay, I hope to collaborate with a relatively new entrant to urban transport markets – an app-based electric scooter platform – to conduct a randomized experiment of customer incentives for pro-social and/or pro-platform behavior (better scooter parking and/or redistribution). For my final essay, I plan to take a deeper look at urban form as it relates to mode choice, with a focus on the spatial layout of cities and the graph structure of transportation networks.

- **Proposal committee**: Arun Sundararajan, Panos Ipeirotis, John Horton, Foster Provost
- **Proposal defense date:** December 2018 (planned)
- **Expected completion date:** May 2020

COURSEWORK

Economics and Statistics
Digital Economics Research Seminar
Microeconomics
Industrial Organization
Math for Economists
Choice Models in Operations
Causal Inference
Probability Theory
Statistical Inference & Regression Analysis
Econometrics

Behavioral
IT & Organizations Research Seminar
Behavioral Research Methods

Technical
Technical Foundations of Information Systems
Foundations of Machine Learning
Machine Learning & Computational Statistics
Data Science Research Seminar
Big Data Science
Fundamental Algorithms

RESEARCH INTERESTS AND PROJECTS

My ongoing projects span three core research areas: the future of urban mobility, machine learning and causal inference, and computational social science. In order of most to least advanced, my existing work has examined:

Ridesharing and Public Transportation, *with P. Ipeirotis and A. Sundararajan.*
This project uses open data from the city of New York to study the relationship between ridesharing and the use of public transportation. We treat subway service disruptions as a natural experiment and explore substitution to Uber, Lyft, taxis, and Citi Bike. We compile a rich and detailed dataset of trip counts by mode and four-hour intervals, disaggregated across 154 taxi zones and spanning a three and a half year period. This allows us to explore granular variations in substitution behavior by geography, time, demographic characteristics, and the level of subway service available.

Automatic Identification of Search Behavior in Trajectory Data, *with UN Global Pulse.*
This project focuses on automatically identifying maritime rescues off the coast of Libya. We use Automatic Identification System (AIS) trajectory data from NGO-run ships that aid refugee vessels in distress, and apply machine learning algorithms to segment and classify their behavior using characteristics such as latitude, longitude, course over ground, and speed. We have also supported efforts to use AIS data for a broader range of humanitarian purposes, such as modeling trade flows in Indonesia and identifying potential sea transmission routes for the 2017 plague outbreak in Madagascar.

Causal Inference and Machine Learning.
This project is a literature review of the emerging research at the nexus between causal inference and machine learning. In particular, I focus on how machine learning can be used to estimate heterogeneous treatment effects, address selection bias, and extend experimental estimates to new settings (e.g. different populations, different treatments). I am also interested in how observational data can be used to supplement limited experimental data.

Deep Learning Tools for Econometric Models, *with P. Ipeirotis.*
Deep learning toolkits such as Tensorflow are designed to fit complex neural network models that may have many layers of computation and vast numbers of parameters. However, we believe that they are also well-suited to estimating simpler structural econometric models. Advantages in this context include their implementation of state-of-the-art stochastic gradient descent algorithms, and their relative ease of model specification and programming. We demonstrate these benefits using the BLP method for estimating random coefficients logit models of consumer choice. With simulations, we aim to show that the BLP approach programmed in Tensorflow can recover the true parameters of the model.

Tracking Social Media Populations in Africa, *with F. Rampazzo and L. Rosenzweig.*
This project examines the Facebook advertising platform as a tool for collecting demographic data on hard-to-reach populations. While audience estimates generated by Facebook's advertising platform have been used to study expatriate populations in Europe and gender divides in global internet access, less is known about their potential in the developing world. We propose exploratory work along two dimensions. Using Kenya as a case study, we will compare subnational audience estimates with pre-existing population density data. Second, we will test the representativeness of surveys that use Facebook as a recruitment tool, comparing data collected via this channel to Afrobarometer and Demographic and Health Surveys (DHS) estimates.