NBER Economics of Digitization
PhD Student Tutorial Application— Zhe Zhang

November 18, 2016

Zhe Zhang
zhezhang@cmu.edu; (510) 292-8232
Carnegie Mellon University, Heinz College
PhD student in Information Systems & Management (Spring 2018 Graduation)

Hello, I am a 4th Year PhD student in Information Systems at Carnegie Mellon University, researching the societal impact of information technology. I have an undergraduate background in economics and statistics from Stanford University, as well as prior academic research experience in environmental and energy industrial organization. At CMU, I have pursued coursework in machine learning & statistics, applied economics & econometrics, industrial organization, and seminars on the economics of information systems. Methodologically, my research tools are wide-ranging, from developing machine learning methods to analytical models to causal applied econometrics, but all towards the purpose of studying the societal impact of IT. Below, I will describe my current research interests and projects.

My research interest during my PhD and going forward is on the societal impact of information technology, particularly in metropolitan areas. I am currently in the midst of an applied economics project to study the impact of the entry of peer-to-peer transportation platforms (namely, UberX and Lyft) on various metropolitan areas. Compared to other studies on the entry of UberX and Lyft, we have unique longitudinal data over several years on the financial credit and debit card transactions of a national sample of consumers, including before and after the entry of such platforms. We are interested in their impact on consumer mobility and consumption patterns (e.g. frequency on consumption, distance traveled, total amount spent, and spatio-temporal diversity of consumption) as well as on local business performance (e.g. total revenues). With our individual panel data, we can not only use a difference-in-difference setup to estimate the adoption effect on adopters, but also how the adoption effect is moderated by the time of adoption (e.g. early vs late) and by usage patterns. Further, we are planning additional analyses at geographical and business levels, and utilizing a few instruments for adoption to estimate a local average adoption effect.

We believe this research can provide more general insights into how changes to urban transportation can influence the local economy and consumer behavior. This can provide insights for planners and government officials around future expected urban transportation changes, such as self-driving cars, increased transit-oriented development, or increased global urbanization. Along with my advisor on this project, Prof. Beibei Li at Carnegie Mellon, we will be presenting this work at the Workshop on Information Systems and Economics (WISE, co-located at ICIS 2016), and plan to submit the work to a journal in early 2017.

My two other research projects use different methodologies, but continue to be motivated by the emergence of technologies in the urban setting. In one project, we develop a simple analytical model to analyze how manufacturers are affected by the entry of a peer-to-peer (P2P) rental marketplace for their goods, and their optimal strategy in the presence of such

P2P markets. We find that in a large number of conditions, a firm benefits most from embracing P2P markets than other firm strategies. Compared to the existing literature on the impact of P2P markets on firms, we do not focus on the intuitive role of production costs, but contribute by highlighting the importance of consumer heterogeneity in usage rates. Particularly, we identify a novel effect that we believe is unique to P2P rental markets, an *equalizing effect* on the willingness-to-pay between heterogeneous consumers, which ends up benefiting the firm. This research is currently under journal submission and has been/will be presented at the Conference on Information Systems and Technology (CIST, co-located at INFORMS 2016) and WISE 2016. This is collaborative work with Prof. Vibhanshu Abhishek (Carnegie Mellon) and Prof. Jose Guajardo (UC Berkeley).

My other project developed due to my interactions with the great machine learning research community at Carnegie Mellon. I have been fortunate to work with students and faculty interested in the intersection of public policy and social sciences with machine learning. I believe this area, particularly as both data availability and the use of data in various sectors of society increases, will continue to be an important area of research. Specifically, I have focused on research to better assess the use of data in decision-making. There has been growth in the discussion of the use and fairness of data-driven decision-making in various sectors including criminal justice, health, public policy, and city management. This has been evidenced by several academic and press discussion around bias and fairness in the use of algorithms, including ProPublica's 2016 analysis of the COMPAS recidivism risk prediction and argument that African-American defendents are more likely to be mistakenly predicted as high-risk.

Originally motivated by these discussions and existing work on crime hotspot models, I worked with Prof. Daniel Neill, a joint faculty member of Heinz College and the Machine Learning Department at CMU, to develop methodology to detect predictive bias in probabilistic classifiers or risk assessment tools. Existing goodness-of-fit or model assessment methods usually focus on one dimension at a time or subpopulations of a priori interest. In our methodology, we are able to detect and identify where a classifier is statistically significantly biased, considering the space of all exponentially possible subpopulations. This simple methodology is valuable to practitioners, managers, and policymakers who are considering the use of a data-driven decision-aid — it's important to know the tool is biased for some subpopulation that we may not have considered checking for ahead of time. In fact, in our assessment of the COMPAS predictions, we find over- and under-estimating the risk for some subpopulations that have not been documented elsewhere. Additionally, we can build on our method's ability to detect poor performance in classifiers to contribute prediction-improving methods and more interpretable machine learning. This project is being presented the NIPS 2016 Workshop on Interpretable Machine Learning, and we plan to submit this work for publication in spring 2017.

Lastly, this past summer, I also pursued my interest in use of data and data science in various societal and governmental sectors by participating in the prestigious Data Science for Social Good Fellowship at Chicago. We worked with several governments and non-profits on data science tools, and there, I was also able to see more about the importance of model checking for bias and heterogeneity in models.