NBER Economics of Digitization
PhD Student Tutorial Application— Zhe Zhang

November 16, 2017

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

CMU Coursework:

microeconomics, econometrics, analytical and structural marketing models, statistical inference, statistical machine learning, statistical regression, statistical causal inference, urban economics, behavioral economics, anomaly and pattern detection, seminars on: economics of information technology, data mining in business research, computational marketing methods

Research interests and projects:

I am PhD student in Information Systems at Carnegie Mellon University, researching the societal impact of information technology, using economics and machine learning methods to do so. I will going on the job market in Fall 2018. I have an undergraduate background in economics and statistics from Stanford University, as well as prior academic research experience in environmental policy and energy industrial organization. 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 planned dissertation is on "Essays on the societal impact of information technology", with an expected proposal date of early 2018 and an expected completion date of late 2018 / early 2019. The committee will include at least: Beibei Li and Vibhanshu Abhishek (professors in information systems and management at CMU) and Daniel Neill (professor in machine learning and public policy at NYU and CMU).

In one vein of research, we use economics and marketing methods to study how changes in information technology indirectly affects consumer behavior, urban outcomes, or other business sectors. One project that is furthest along is using developing an analytical model around the effect of peer-to-peer (P2P) rental markets on manufacturers. It is under 2nd round review at *Management Science*. In this work, we motivate that P2P rental markets represent a challenge for incumbent firms, but also an opportunity to adjust their own business models. We contribute research that identifies that consumer heterogeneity in usage rates as a driving factor of the mechanisms that explain the different market outcomes for a manufacturer—when they may be better or worse off because of the entry of P2P. In fact, we identify an *equalizing effect* of P2P rental markets, which we believe has been documented for the first time in the context of P2P in our work. Additionally, we investigate and identify

scenarios when alternative market structures for the OEM are optimal, including the OEM vertically integrating and offering their own P2P platform, or the OEM competing with P2P rentals with their own rentals.

Another applied microeconomics research project in this vein, which is more preliminary, but I anticipate using as my job market paper, is studying changes urban frictions and consumption patterns, using evidence from the entry and usage of Uber/Lyft. We use a unique, rich panel financial dataset to observe both urban consumption patterns and usage of Uber/Lyft. We will focus on three perspectives: consumer, merchant, and city/neighborhood, and use various casual identification analyses to study how the usage of Uber/Lyft affects those perspectives. Importantly, we will focus on not just overall changes, but heterogeneity in those changes. Initial consumer-level results—which were published in the Proceedings of KDD 2018 and also accepted presented for an oral presentation at KDD (9% rate) and *CIST*/*WISE* conferences—suggest overall increases in restaurant, food, and bar frequency, as a function of usage rather than as a function of adoption. Additional recent work suggests that changes are a function of Uber/Lyft usage, as the main driver of consumer change, and that its influence is stronger on some consumption categories than others. Further work will add matching and instrumental variable robustness, and explore geographic and demographic heterogeneity in these consumer results. We will also focus on the merchant and neighborhood/city geographic effects, since it's of managerial and urban interest to understand how Uber/Lyft may be affecting previously harder-to-reach neighborhoods.

Lastly, considering future projects in this vein, I would like to explore labor market outcomes and preferences surrounding the sharing economy.

In another vein of research, motivated by the increasing use of predictive analytics in various public and private sectors (e.g. criminal sentencing or stop-and-frisk decision-making), I study the issue of hidden bias in predictions. This falls within my dissertation because of the increasing role of machine learning in societal sectors and the potential important impact of biased or unfair predictions. This has gained popularity in the computer science community (see FAT* and FAT/ML conferences), so we contribute an important advance in auditing and interpreting for the presence of bias in predictions, novelly, without needing to specify subgroups of interest apriori. We develop an anomaly-detection statistical method to identify hidden subgroups where predictions are significantly over-estimating risk, under-estimating risk, or unusually high error rate (and also enable penalizing complexity and multiple methods for estimating significance). This has been accepted for presentation at workshops at KDD 2018 and NIPS 2017. In our applied results, as an example, we identify two subgroups of defendants that are over-estimated and under-estimated by COMPAS recidivism predictions, which we believe have not been documented elsewhere. In another example, we show the presence of subgroup bias in a published, proposed stop-and-frisk decision model. In future

work in this vein, we will develop further methodological extensions and use real datasets as case-studies to highlight the importance of auditing algorithms in multiple field journals. I have also done some preliminary behavioral economics work on the perceptions and preferences around potentially biased machine learning.

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.

I attended the NBER Digitization Tutorial last year (in 2017) and found it extremely valuable. I enjoyed meeting and staying in touch with several PhD students and found the discussions with faculty enriching and helpful for coming up with new research ideas. Additionally, it was encouraging to see the high-quality presentations on the last day, of the NBER Winter Meeting. Thus, I would quite like to attend again and both contribute and learn.