Application to NBER Digitization Tutorial

October 12, 2019

Applicant and Contact Information

Tesary Lin

University of Chicago, Booth School of Business

5807 South Woodlawn Avenue

Chicago, IL 60637

Email: tesarylin@chicagobooth.edu

Course of Study and Progress in the Program

I am a sixth-year Ph.D. student in Chicago Booth, with a major in quantitative marketing and a minor in economics. During the first two years of study, I have taken the *Microeconomics* and *Econometrics* sequences in the economics Ph.D. core curriculum, along with the *Advanced Industrial Organization* sequences, the *Economics of Information*, and the *Economics of Communication*. I have also taken courses in statistics and computation, including *Bayesian Inference* (I and II), *Computational Social Science*, *Machine Learning and Public Policy*, and *Applications of Machine Learning to the Empirical Sciences*.

My dissertation is titled *Measuring Intrinsic and Instrumental Preferences for Privacy*. My dissertation committee members include Pradeep Chintagunta (co-chair), Sanjog Misra (co-chair), Bradley Shapiro, and Oleg Urminsky.

Research Interests and Projects in Progress

My research interests center around information flow in the digital economy. Two themes emerge from this general focus: one is characterizing people's preferences for privacy, i.e., their inclination (not) to share personal data; the other is developing tools suitable for analyzing digital data, given challenges posed by the modern information infrastructure.

My job market paper [1], *Measuring Intrinsic and Instrumental Preferences for Privacy*, distinguishes between two types of motives for protecting one's privacy: the intrinsic motive, referring to one's pure "taste" for protecting their personal data; and the instrumental motive, referring to concerns about the consequences of revealing their "type" relevant to the specific transactional environment. Distinguishing between the two preference components not only improves the measurement of privacy preferences across contexts, but also plays a crucial role in developing inferences based on data voluntarily shared by consumers.

Combining a two-stage experiment and a structural model, I measure the dollar value of revealed preference corresponding to each motive, and examine how these two motives codetermine the composition of consumers choosing to protect their personal data. The compositional differences between consumers who withhold and who share their data strongly influence the quality of firms' inference on consumers and their subsequent managerial decisions. Counterfactual analysis investigates strategies firms and researchers can adopt to improve their inference, from two different angles. Ex ante, firms and researchers can allocate resources to collect personal data where their marginal value is the highest. Ex post, a consumer's data-sharing decision per se contains information that reflects how consumers self-select into data sharing, and improves aggregate-level managerial decisions. Firms and researchers can leverage this information instead of imposing arbitrary assumptions on consumers not in their dataset.

Another ongoing project of mine, *Privacy Preferences in Micro Loan Applications* (with Dang Chu, Mandy Hu and Pradeep Chintagunta), complements the first project by quantifying the dollar value of privacy preferences among a large group of micro loan applicants. We

[1] I completed my dissertation proposal in Dec. 12, 2018 and expect to complete this project before May 2020.

observe a promotion from the micro loan company that solicits personal data from applicants by offering them interest rate discounts. Upon seeing the promotion, applicants face a trade-off between privacy and monetary benefits. Although the discount rate offered is the same across applicants, the dollar value of discount varies with loan characteristics and with each applicant's expected repayment decision. We back out this expected dollar value for each applicant by integrating over their ex post repayment behaviors. Initial analysis reveals high preferences for protecting privacy among the applicants. Motivated by the first project, we plan to further examine the relationship between applicants' privacy decisions and their (ex ante) private information.

My third project, *Identity Fragmentation Bias* (with Sanjog Misra), shifts focus from data generation to data collection and analysis. It documents the estimation bias when researchers use data with "identity fragmentation": these are data pulled from multiple devices, channels or platforms that feature the same set of users, yet with identifiers that cannot be matched up across different sources. Using a stylized econometric model, we show that identity fragmentation generally leads to biased estimates. Contrary to conventional wisdom, the bias does not always take the form of attenuation; instead, the direction of bias depends on a user's task specialization when using different platforms. We compare three solutions to the identity fragmentation bias: identity stitching, debiasing at the experiment design stage, and stratified aggregation. Preliminary results show that stratified aggregation performs superior to alternative solutions in its robustness to underlying fragmentation patterns and in its ability to adapt to different modeling strategies.