Application to NBER Digitization Tutorial

November 6, 2018

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

I am a fifth-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*, and *Machine Learning and Public Policy*.

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 dissertation [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. My theory shows that the coexistence of these two motives creates complex data sharing patterns. These patterns in turn have implications on the assessment of data quality collected by firms, and on incentive design to encourage data sharing. In particular, my theory offers the following predictions:

–Data shared by consumers are less representative when instrumental motive is stronger compared to the intrinsic motive. On the other hand, the representativeness of data changes non-monotonically with the correlation between these two motives.

–Holding fixed the willingness to protect privacy, incentives to encourage data sharing are less effective when intrinsic motives account for more variations in privacy preferences, and when the two motives are sufficiently negatively correlated.

I then design a field experiment that separately measures the intrinsic and instrumental preferences for privacy under a revealed preference framework. Using data generated from this experiment, I construct a structural model that allows me to perform counterfactual analyses. The first analysis compares the degree of sample selection under different incentive schemes. The second analysis calculates optimal incentives for data procurement, which takes into account the information value of nondisclosure.

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

[1] I expect to propose my dissertation in December 2018 and complete this project on or before March 2020.

the dollar value of privacy preferences among a large group of micro loan applicants. We 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 propose two solutions to the identity fragmentation bias: the first solution focuses on debiasing at the experiment design stage; the second solution focuses on data processing after it is collected.