Application to NBER Digitization Tutorial

November 17, 2017

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 fourth-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 a general topic, which is characterizing the information environment in the digital space. Two themes spin from this general focus: one is how

economic agents react to and reshape the information structure; the other is how to develop the tools suitable for analyzing the digital market, in lieu of the challenges created by the growing data and the information infrastructure. Below provides a summary of the three projects that I am working on in these directions:

My first research project, *Review Manipulation in a Firm's Life-Cycle* (with Pradeep Chintagunta and Junhong Chu), describes review manipulation by sellers on *Taobao.com*. The motivation is to understand the following questions: how is the reputation mechanism in the online marketplace working? If it is not working well, what results in its problem? Will a platform's intervention suffice in reaching the first-best outcome, or will it backfire?

To answer these questions, we utilize a unique feature of *Taobao*'s review system: every review includes a rating for shipping quality which supposedly represents buyer evaluation of the shipping company's performance. Given that shipping companies are not part of a seller's internal service network, the short-term performance variation is exogenous to sellers once long-term contractual changes are controlled for. We find that despite the quality exogeneity, shipping rating rises with variables representing the net benefits of review manipulation: for example, rating rises when a seller's cumulative rating gets close to a threshold for ascending to a higher rank. We also find that shipping rating decreases as a store stays longer in the market, contradicting a "learning by doing" story while being consistent with the manipulation hypothesis. In particular, it resonates with Bresnahan, Davis and Yin (2014)'s anecdotal evidence that new entrants often have to hire people for posting reviews in order to overcome the "cold-start" problem. This finding suggests that a naive elimination of review frauds may stymie the platform growth, because it deprives new entrants of the only chance to overcome the cold-start problem.

My second project, *Consideration-Choice Model: A Double Machine Learning Approach* (with Brian Quistorff), tries to improve the limited consideration-choice model in IO and marketing by enhancing its computational performance using machine learning. We believe that this is an important problem to solve, because the computational demand has forced many existing papers to adopt restrictive assumptions on the size of product space, which is at odds with the original motivation for constructing a limited consideration model.

To improve computational performance while at the same time ensuring the inference estimate is causal, we apply double machine learning (DML) techniques developed by Chernozhukov et al. (2016) to a consideration-choice model. In particular, we derive the orthogonal moment conditions for a revised discrete choice model that allows both limited attention and consumer heterogeneity. We also present practical challenges when applying DML to estimate demand with long tails, and discuss possible solutions to solve this problem.

My third project, *Identity Fragmentation Bias* (with Sanjog Misra), documents the estimation bias when using a type of data commonly collected in the modern digital environment: data pulled from multiple devices, channels, or platforms that feature the same set of users, yet one where the identities from different sources cannot be matched, i.e. identities in the compiled data are "fragmented". Using a stylized econometric framework, we show that identity fragmentation in the dataset often leads to significantly biased estimates. Contrary to conventional wisdom, the bias does not always take the form of attenuation. Instead, this bias can be either upward or downward, sometimes even inducing a sign reversal, depending on a user's task specialization when using different platforms. This paper highlights a trade-off that researchers and practitioners face when they decide to collect and utilize digital data from different sources: mainly, the trade-off between a selected sample and a sample that may give misguided evidence due to poor matching. As a first step towards understanding this tradeoff, we discuss different possible solutions to overcoming the identity fragmentation bias.

Although these three papers use very different approaches, they all center around a common theme: a better understanding of the information structure in the digital space, be it centering on seller strategic behavior, consumer decision making, or research practices. Going forward, I will continue pursuing research topics related to this theme.