Gentry Johnson | Application
NBER Digitization Tutorial 2018

1133 East-West Highway Apt. 928W – Silver Spring, MD 20910 – USA
+1 (920) 371 2082 • johnsong@econ.umd.edu
www.econ.umd.edu/gradprofile/Johnson/Gentry
PhD Candidate in Economics, University of Maryland

NBER – Economics of Digitization November 9, 2017
Stanford University

Statement of Interest

My name is Gentry Johnson and I am applying for admission to the 2018 NBER Digitization Tutorial. I am a third-year PhD candidate in economics at the University of Maryland where I study applications of machine learning to the estimation of structural models in industrial organization. I expect to graduate in the spring of 2020. My formal major field is industrial organization and my minor field is econometrics. I am also developing a course in advanced computing and data methods for undergraduates.

At the core of my work is the question: how is the increasingly massive amount of data available to economists best leveraged without discarding our elegant set of causal econometric models? The NBER Digitization Tutorial would be an excellent opportunity for me to engage with the tools and researchers which can help further this agenda.

Description of Coursework and Research

In addition to the standard first-year economic theory sequence, I have completed courses in advanced econometric theory, computational methods and structural modeling, theoretical industrial organization, advanced microeconomic theory, decision theory, and labor economics. I will also complete coursework in signal processing and the mathematics of machine learning.

As I am only beginning my third year of study, I have not yet formally proposed a thesis nor formed a committee. Nonetheless, I will briefly describe my research agenda. I am currently being advised in this course of study by Andrew Sweeting, Ginger Jin, and Guido Kuersteiner.

My research focuses on the use of supervised and unsupervised machine learning techniques, not as a replacement for the theoretically based structural models familiar to IO economists, but rather as a tool to improve computational efficiency, increase estimate precision, and build-out deep data sets. All of this is done such that structural models continue to function as the core of any econometric

work–but with inputs that are richer in an informational sense and, at the same time, leaner in a dimensionality sense.

This work utilizes the Nielsen Scanner and Homescan data sets from the University of Chicago. I am exploring the use of deep learning to discover structure in demographic and product characteristic data, and then to engineer compact vectors of composite features which represent important underlying interactions and non-linearities in the data while minimizing information loss. The value of this process as an input into random coefficient demand models grows as the availability of high-dimensional data grows.

To that end, I plan to use the “feature extraction" functionality of, for instance, image recognition algorithms in order to turn existing public domain information into usable economic data. This can include representing product images through meaningful continuous sets of features which capture some of the “unobserved" product-level variation that is currently pushed into a structural error term in random coefficient demand models.

Understanding how these types of data are being used currently by economics researchers will help me to imagine new avenues for this program of melding machine learning methods for “knowledge discovery" with econometric tools. Using modern methods such as Natural Language Processing to assemble economic data sets from previously untapped resources could very well lead to a tidal wave of new, robust empirical research fit that reflects the economy's movement towards the digital–both in terms of methodology and policy relevance.

These methods can be expanded to address other aspects of both structural and reduced form models. For instance, I have plans to apply these methods to questions such as market definition and broadly, to instrumental variables estimation. The resources and connections at the NBER Digitization Tutorial will be invaluable to me, and I look forward to the opportunity to expand my knowledge and skills as a researcher.