Research Statement

Zhonghui Zhang

My research fields of interest are in econometrics, machine learning, and financial econometrics. My mathematics background from college and master's program provides me an excellent intuition to understand the most complicated theory in econometrics. Combining Ph.D. training gives me the ability to dig deeper and make breakthroughs in the existing literature. My preference in research is to study econometrics theories that can be widely used in practice and to improve them. Machine learning is perfectly matching my interest. To my understanding, machine learning is to let the computer select the optimal econometrics/statistics models to achieve a specific purpose based on some criteria. About the difference between econometrics and statistics is an open question that everyone has a different answer. My opinion is that they have a lot of overlapping, but at the same time have their emphasis. The specific topic I'm working on is combining clustering methods with the economic model. Because we are now in the age of big data, it is particularly important to select useful signals from massive noise, and it is what clustering is concerning. Using the tool of machine learning to study economic problem is an inevitable trend, but there are still a lot of research questions that need to solve. Machine learning is not merely econometrics or statistics plus programming. It also requires lots of domain knowledge. My economics background provides me such domain knowledge, so I can have a better sense of how agents behave given different setups. Thus I can limit the scope of searching optimal model to a small area and significantly improve the efficiency of the algorithm.

During my time in the Ph.D. program at the University of Connecticut, I have been fortunate to work with my advisor Prof. Chihwa Kao. Under his guidance, I have been exposed to the mainstream of econometrics research and received world-class training. I am also fortunate to have the opportunity to have Prof. Min Seong Kim and Prof. Jungbin Hwang, who have published many papers in the top field journal such as the Journal of Econometrics, as my committee members. Below I will discuss my job market paper, several working papers, and my future research plan.

Job market paper: "Mahalanobis Metric Based Clustering for Fixed Effects Model", with C. Kao and M. S. Kim, accepted by Sankhya Series B

This paper proposes a Mahalanobis metric based k-means algorithm (KMM) for group membership estimation in linear panel data models with time-varying grouped fixed-effects by Bonhomme and Man-

resa (2015). The proposed method improves the accuracy of estimates by taking serial correlation and heteroscedasticity into account. We also derive the optimal β for group membership estimation and show that it may be different from the true coefficient parameter. Since the optimal β is not feasible in practice, we propose the data driven selection method for its implementation.

"Is the recursive preference asset pricing model more flexible? Evidence."

This paper investigates why there exists considerable variation in estimates of the coefficient of relative risk aversion (CRRA) and the elasticity of intertemporal substitution(EIS) in the consumption-based asset pricing model with Epstein and Zin (1989) preferences. Using the estimation method developed by Chen et al. (2013), we show the Epstein and Zin (1989) structure collapses to the time-separable structure. This result is consistent with the argument in Kocherlakota (1990) saying that the recursive preference-based utility function does not have more explanatory power than the time-separable one. We also show the choice of parameters might lead to "ill-behaved"conditional moment, which might cause either GMM method to get "stuck", or the estimates do not move much from the starting points. Lastly, our result is robust to the choice of instruments for computing the conditional moment function in the GMM method.

"Nonlinear models with latent grouping and groupedfixed effect."

We extend the linear panel data model with grouped fixed effect and unknown group membership in Bonhomme and Manresa (2015) to nonlinear. Unlike Bonhomme et al.(2017, Working Paper), we assume the unobservable heterogeneities are from the mixture of a certain number of group-specific distributions. Our method provides information on higher moments than k-means which only offer the first moment. We also relax the restrictions on the group-specific distribution and leave it to be estimated nonparametrically, and provide a guide on selecting bandwidth. We show that the coefficient of interest covariate in the grouping object or "moment" in Bonhomme et al. (2017, Working Paper) is a "nuisance" concerning grouping. This implies, in the grouping step, we can choose some particular value for the latent common parameter just for grouping purpose. Lastly, we study the distribution of unobservable heterogeneities given different value of common parameter and T through Monte Carlo simulation.

Future Work:

In the age of big data, the blessing is that we have a larger sample size to obtain more precise estimation and forecasting results. However, we also have to suffer from the curse of high-dimensionality. For the coming years, I will continue pursuing research on the overlapping area of machine learning and classical economics. For example, one research question on my list is how to deal with the large-dimensional

covariance matrix. In my job market paper, I use the covariance matrix to reduce the serial correlation and heteroscedasticity issue and thus improve the performance of the K-means algorithm in the panel data model setup. But the question I didn't touch is that what if the dimension of covariance goes to infinity? There is literature in statistics trying to solve this problem by either assuming sparsity or imposing specific structure on the covariance. I will follow this topic and try to make contributions to the literature in econometrics.

Another idea in my mind is introducing spectral clustering to the panel data model. Spectral clustering is a technique with roots in graph theory, where the approach is used to identify communities of nodes in a graph based on the edges connecting them. The method is flexible for non-graph data as well. With the help of modern social media or K Nearest Neighbour (KNN), it is possible to get the network (graph) of individuals. Hence, spectral clustering provides economists an alternative to identify grouping and studying the corresponding group fixed effect rather than rely on prior information such as race, state, and so on.