Keynote Speaker

Jianqing Fan, is a statistician, financial econometrician, and data scientist. He is Frederick L. Moore'18 Professor of Finance, Professor of Statistics, and Professor of Operations Research and Financial Engineering at the Princeton University where he chaired the department from 2012 to 2015. He is the winner of The 2000 COPSS Presidents' Award, Morningside Gold Medal for Applied Mathematics (2007), Guggenheim Fellow (2009), Pao-Lu Hsu Prize (2013) and Guy Medal in Silver (2014).

Title: Measuring Housing Activeness from Multi-source Big Data and Machine Learning

Abstract: Measuring timely high-resolution socioeconomic outcomes is critical for policy-making and evaluation, but hard to reliably obtain. With the help of machine learning and cheaply available data such as social media and nightlight, it is now possible to predict such indices in fine granularity. This paper demonstrates an adaptive way to measure the time trend and spatial distribution of housing activeness with the help of multiple easily-accessible datasets. We first identified the regional activeness status at the individual level from energy consumption data and then matched it with nightlight and land use data geographically. Then, we introduce the principle of robustification via truncation and factor-adjusted regularization methods for prediction (FarmPredict) to deal with two important stylized features in big data. The heterogeneity of big data is mitigated through the use of the government land planning data. FarmPredict effectively lifts the prediction space and solves the colinearity problem in high-dimensional data. It is applicable to all machine learning algorithms. FarmPredict allows us to extend the regional results to the city level, with a 75% out-of-sample explanation of the spatial and timeliness variation in the housing usage. FarmPredict is not only a model but an analytical framework of machine learning on high-dimensional data, showing broad potential applications to other social science problems. Since energy is indispensable for life, our method is highly transferable with the requirement of only public and accessible data. Our paper demonstrates the power of machine learning in understanding socioeconomic outcomes when the census and survey data is costly or unavailable.

(Joint work with Yang Zhou, Lirong Xue, Zhengyu Shi, Libo Wu)