Dear Sir/Madam,

I am interested in the Digitization Tutorial as in the opportunity to meet other econometricians and game theorists. I believe my experience in estimating causal effects after model selection would match the topic of the meeting. Specifically, I worked on estimation of treatment effects after model selection in presence of partially missing response and/or covariates. The goal of the estimation was to insure that the treatment effect distribution is not affected by the estimation error of the propensity score and the errors in the model selection. Separately, I am also interested in parallelizing the computation of econometric estimators that allows fast implementation on distributed data sets and have experience in Hadoop.

The focus of my research is adaptive semiparametric estimation with applications to missing data and kidney allocation. In a semiparametric setup, the target parameter, such as treatment effect, satisfies some moment equations that depend on an unknown function or a high-dimensional parameter, such as the propensity score or the effect of the controls. Adaptivity of the moment conditions insures that the asymptotic distribution of the target parameter is not affected by the estimation error of the first stage parameter. This allows the researcher to avoid estimation of the asymptotic bias of the target parameter, coming from the first-stage estimation error.

In my co-authored project “Inference in Misspecified Linear Model with Partially Missing Response”, we are interested in the inference on the best linear predictor (BLP), in presence of partially missing response variable in high-dimensional context. We train the lasso based on the complete part of the data, by reweighting each observation by inverse propensity score. Then, we replace the response by a combination of weighted lasso prediction and the response itself. Here, the weight insures correct mean and adaptive standard errors of the estimator. The paper is an example of how the unlabeled data can be used for inference. I have extended this idea in my own project for generic moment conditions: logistic, instrumental variable, quantile regression.

Another project I have worked on was on kidney allocation. Jointly with Nikhil Agarwal and Victor Chernozhukov. I predicted conditional probability of the offer acceptance using patients’ and donors’ demographic and biological characteristics. We used logistic lasso, gradient boosting machine, and random forest to predict the outcome. The obtained MSE was of the order 0.001 with the of the methods in 12%-14%. This was an improvement compared to 3% regular logistic regression KPSAM model estimated by Organ Procurement and Transplantation Network. When working on this project, I developed extensive skills in R and Python.

I am looking forward to hearing from you. Thank you for your consideration.

Sincerely,

Vira