

Abstract

The main focus in econometrics is to provide an *explanation* of various observed outcomes. Structural econometricians obtain reliable estimates of parameters that describe an economic system to provide an understanding of the underlying processes that determine equilibrium outcomes. The estimation process is based on conditions implied by economic theory.

On the other hand, the main focus in machine learning is to provide accurate *predictions* of the variables of interest. While these techniques are extremely powerful for forecasting, it can be very hard to interpret the underlying structure implied by them.

As machine learning techniques become more popular and computers become capable of storing and processing large quantities of data, there have been some recent efforts to incorporate such techniques into structural econometric models. My research aims to extend this literature.

Chapter 1: Regularization Paths in Generalized Method of Moments

In the GMM framework, the objective function to be minimized is a weighted sum of squares of m moment conditions implied by economic theory. The derivative of the objective function with respect to the vector of parameters (θ) provides a system of k equations in k unknowns that is used to obtain parameter estimates. However if this matrix is nearly singular at the true parameter values, then the system of equations becomes highly unstable. This is analogous to the problem of multicollinearity in linear regression. In the linear regression framework the problem is addressed by regularization. However, due to the highly non-linear nature of the GMM objective function, techniques like ridge and spectral cut-off regularization are not readily generalizable to the GMM framework.

In the first chapter (co-authored with Fallaw Sowell), we re-interpret regularization as a set of possible solutions that lie along a *path* between the unconstrained minimum of the objective function and a pre-defined prior. Using this interpretation, we propose algorithms for finding the ‘regularized’ parameter estimates. We use the notion of *cross-validation* in GMM. We also show via simulations that our method performs very well when the system of equations is unstable. We discuss how to extend the techniques in higher dimensions and as an empirical application we employ this method on the Consumption based Capital Asset Pricing Model.

Chapter 2: Propensity Score Model Selection using Machine Learning Classifiers

The basic issue in estimating the effect of a particular treatment using observational data is that the data suffers from *selection bias*. In other words those who receive treatment (the treatment group) are inherently different from those who don’t (the control group). Heckman (in his seminal 1978 paper) shows that a naive estimate of the regression parameter on a treatment dummy suffers from an *omitted variable bias*. The problem arises because we only observe outcomes under a single state (either treatment or control)– thus we have to control for factors which simultaneously affect both outcome and selection into

the treatment group. Rubin and Rosenbaum (1983) pioneered the work on causal inference in the statistics literature. They suggest a two-step estimation procedure. In the first step the probability that an individual belongs to the treatment group is estimated (*Propensity Score of the individual*). The second step involves using the Propensity Score for pre-processing the data before estimating the Average Treatment Effect (ATE).

The use of Inverse Propensity Score Weighting (IPW) is now ubiquitous in the Causal Inference literature, however the estimation of propensity scores remains an open question. While many authors use logistic regression because of its interpretability, others argue in favor of non-parametric methods. We propose the use of machine learning classifiers (like Naive Bayes, Regression Trees and Support Vector Machines) for obtaining propensity scores. We show via theoretical arguments and simulation studies why its useful to consider a variety of propensity score models in the first step. We compare propensity scores estimates obtained from Linear Probit model as well as from semi-parametric classifiers like Naive Bayes, Random Forests and Support Vector Machines. In particular we find that propensity score estimates with Minimum Covariate Imbalance perform very well in terms of Mean Squared Error of ATE estimates across all simulations.

Chapter 3: Evaluating India's Safe Motherhood Scheme using Inverse Propensity Score Weighting

Conditional Cash Transfers (CCT) programs are becoming an increasingly popular policy tool in developing countries to incentivize certain behavior such as school enrollment, vaccination and health check ups amongst a targetted section of the population. The beneficiaries of CCTs are typically from poorer communities and the final aim of such programs is to help such communities get out of poverty. India's Safe Motherhood scheme or Janani Suraksha Yojana (JSY), launched in 2005, incentivizes eligible women to give birth in health care facilities. With more than 9 million beneficiaries, it is the world's largest CCT program in terms of the number of beneficiaries.

We use estimation techniques developed in Chapter 2 (IPW using Minimum Covariate Imbalance criteria) to evaluate the effectiveness of the scheme. In particular we estimate the ATE of receiving financial assistance via JSY on two health outcomes – number of stillbirths and infant mortality. We also estimate ATE on three behavioural outcomes – whether the mother had 3 or more ante-natal checkups, whether any post-natal check up was conducted within 2 weeks of delivery and the frequency of child check-ups within 10 days of delivery. We are not aware of any other paper that uses Propensity Score methods to evaluate JSY at the national level. Our results indicate that in certain geographical regions propensity scores obtained via machine learning techniques were picked leading to results that are qualitatively different from those obtained by the standard linear probit.