Problemset (1), Advances in Causality and Foundations of Machine Learning, fall 2019

Maximilian Kasy

In this problemset you are asked to implement some simulations and estimators in R. Please submit complete commented code solving these problems, as well as the clearly labeled output of your programs, including brief discussions of your results. R-Markdown might be useful for writing up your answers. Please submit your answers via Canvas.

1. Write code to simulate i.i.d. draws (X_i, Y_i, Z_i) from the random coefficient model

$$
\begin{aligned}
Y &= U_1 + U_2 \cdot X \\
X &= V_1 + V_2 \cdot Z \\
Z &\sim N(1, 1) \\
(U_1, U_2, V_2, V_2)|Z &\sim N(\mu, \Sigma)
\end{aligned}
$$

   What is the average structural function in this model? What is the ATE, what is the LATE?

2. Write routines implementing the following five estimators:

   (a) Instrumental variables using the formula $\widehat{\beta} = (\boldsymbol{Z}'\boldsymbol{X})^{-1}\boldsymbol{Z}'\boldsymbol{Y}$.

   (b) Two-stage least squares, regressing Y on the first stage predicted values $\widehat{X}$.

   (c) Control function regression, regressing Y on X and the first stage residual.

   (d) The non-parametric conditional moment equality approach of Newey and Powell.

   (e) The non-parametric control function approach of Imbens and Newey.

3. Find and prove (minimal) restrictions on the covariance matrix Σ such that

   (a) The LATE equals the ATE.

   (b) The conditional moment equality approach is valid.

   (c) The control function approach is valid.

   Simulate 10.000 observations for each of these cases, and for a generic case where neither of the conditions holds. Run all 5 estimators for all 4 cases and discuss the results.