High-Dimensional Econometrics
Computer Session 1: Using ML Tools for Inference

Jérémy L'Hour
jeremy.l.hour@ensae.fr

February 19, 2019

1. Orthogonal Score for the ATET (Examen 2018)

We observe an iid sample of the random vector $W_i = (Y_i^{obs}, D_i, X_i')'$ for $i = 1, ..., n$. Y_i^{obs} is the outcome variable which value depends on whether the unit i is treated or not. If unit i is treated ($D_i = 1$), then $Y_i^{obs} = Y_{1i}$. If unit i is not treated ($D_i = 0$), then $Y_i^{obs} = Y_{0i}$. X_i is a vector of covariates of dimension $p > 1$ which includes an intercept. The index i is dropped when unnecessary. The quantity of interest is the Average Treatment on the Treated defined as:

$$\tau_0 = \mathbb{E}\left[Y_1 - Y_0 | D = 1\right]. \tag{ATET}$$

Define $\pi = \mathbb{P}(D = 1)$ and the propensity score $p(X) = \mathbb{P}(D = 1|X)$. We make the following two assumptions. The Conditional Independence Assumption:

$$Y_0 \perp\!\!\!\perp D | X, \tag{CIA}$$

and the Common Support Assumption:

$$0 < p(X) < 1. \tag{CommonSup}$$

Assume the propensity score is given by a Logit, *i.e.* $p(X) = \exp(X'\beta_0)/(1 + \exp(X'\beta_0))$.

1. (a) Define:

$$m(W_i, \tau, \beta) = \left(D_i - \exp(X'\beta)(1 - D_i)\right) Y_i^{obs} - D_i \tau.$$

Verify that $\mathbb{E}m(W_i, \tau_0, \beta_0) = 0$.

(b) Propose an estimator of the propensity score $\hat{p}$. Suggest an estimator $\hat{\tau}$ of τ_0.

(c) What is the problem with this estimator in high-dimension?

2. Assume that the outcome under no treatment is given by $Y_0 = X^\top \gamma_0 + \varepsilon$ with $\varepsilon \perp\!\!\!\perp X$ and $\mathbb{E}\varepsilon = 0$.

   (a) Show that $\mathbb{E}[DX(Y_0 - X'\gamma_0)] = 0$.

   (b) Suggest an orthogonal moment condition ψ. Prove that it is orthogonal.

2. Post-Selection Inference: Simulation Study

All the necessary files can be found at `https://github.com/jlhourENSAE/hdmetrics`.

Consider the data-generating process contained in the file `DataSim.R`. The outcome equation is linear and given by: $Y_i = D_i\tau_0 + X_i^T\beta_0 + \varepsilon_i$, where $\tau_0 = .5$, $\varepsilon_i \perp\!\!\!\perp X_i$, and $\varepsilon_i \sim \mathcal{N}(0,1)$. The treatment equation follows a Probit model, $D_i|X_i \sim$ Probit $\left(X_i^T\delta_0\right)$. The covariates are simulated as $X_i \sim \mathcal{N}(0,\Sigma)$, where each entry of the variance-covariance matrix is set as follows: $\Sigma_{j,k} = .5^{|j-k|}$. Every other element of X_i is replaced by 1 if $X_{i,j} > 0$ and 0 otherwise. The most interesting part of the DGP is the form of the coefficients δ_0 and β_0:

$$\beta_{0j} = \begin{cases} \rho_d(-1)^j/j^2, \; j < p/2 \\ 0, \text{ elsewhere} \end{cases}, \; \delta_{0j} = \begin{cases} \rho_y(-1)^j/j^2, \; j < p/2 \\ \rho_y(-1)^{j+1}/(p-j+1)^2, \text{ elsewhere} \end{cases}$$

We are interested in comparing the performance of the naive and the double selection procedures when estimating τ_0. The file `LassoFISTA.R` contains a function to compute the Lasso solution.

1. Comment the form of the coefficients β_0 and δ_0. What do you think will happen with the naive selection procedure?

2. The code `DoubleML_SimExo.R` contains the code for the naive selection procedure. Try and understand the code. Run it. What do you observe?

3. Write the code for the double selection procedure without sample-splitting under "METHOD 2". Optional: write the code to compute standard error.

4. Write the code for the double selection procedure with sample-splitting under "METHOD 3".

5. Run the simulation. What are your conclusions?