Econ 883: Problem Set 3

Federico A. Bugni*

November 17, 2020

Due date: Monday, November 30, 4 p.m., as an upload in Sakai

1. Consider the partially linear model:

$$Y \;=\; X'\beta + g(Z) + U,$$

where $(Y, X, Z) : \Omega \to \mathbb{R} \times \mathbb{R}^{d_X} \times \mathbb{R}^{d_Z}$ are the observables, $\beta \in \mathbb{R}^{d_X}$ is an unknown vector, $g : \mathbb{R}^{d_Z} \to \mathbb{R}$ is an unknown function, and U is an unobserved error term that satisfies:

$$E[U|(X,Z) = (x,z)] \;=\; 0 \;\forall (x,z) \in S_{(X,Z)},$$

and S_W denotes the support of W. The unknown parameters of the model are $\theta = (\beta, g)$.

We have established the identification of θ under assumptions on $\Sigma_X \equiv E[(X - E[X|Z])(X - E[X|Z])']$. As explained in the class, identification is a critical condition for inference. The goal of this problem is to further explore the identification of this model.

(a) Show that $\Sigma_X = E[V(X|Z)]$.

(b) For each of the following scenarios, investigate whether θ is identified or not.

i. X is mean independent of Z and $V(X)$ is non-singular.

ii. X is perfectly determined by Z, i.e., $X = h(Z)$ for a non-stochastic function $h : \mathbb{R}^{d_Z} \to \mathbb{R}^{d_X}$.

iii. Z is perfectly determined by X, i.e., $Z = h(X)$ for a non-stochastic function $h : \mathbb{R}^{d_X} \to \mathbb{R}^{d_Z}$.

iv. X includes a constant explanatory variable, i.e., $X_{1,i} = 1$ for all $i = 1, \ldots, n$.

v. Z contains a constant explanatory variable, i.e., $Z_{1,i} = 1$ for all $i = 1, \ldots, n$.

vi. $(x_j, z_j) \in S_{(X,Z)}$ for $j = 1, \ldots, d_X$ and $M \equiv \{(x_j - E[X|Z = z_j])\}_{j=1}^{d_X}$ is a full-rank matrix.

vii. Support of X includes $\mathbf{0}_{d_X \times 1}$ and Z has full support, i.e., $S_Z = \mathbb{R}^{d_Z}$.

2. Consider the partially linear model:

$$Y \;=\; X'\beta + g(Z) + U, \tag{1}$$

where $(Y, X, Z) : \Omega \to \mathbb{R} \times \mathbb{R}^{d_X} \times \mathbb{R}^{d_Z}$ are the observables, $\beta \in \mathbb{R}^{d_X}$ is an unknown vector, $g : \mathbb{R}^{d_Z} \to \mathbb{R}$ is an unknown function, U is an unobserved error term, $U \perp (X, Z)$, $E(U) = 0$, and $E(U^2) = \sigma^2 < \infty$.

*Email: federico.bugni@duke.edu. Please watch out for typos and mistakes.

Let $\hat{\beta}$ be the estimator of β defined in Robinson (1988, Eq. (2.3)) and, so,

$$\sqrt{n}(\hat{\beta} - \beta) \xrightarrow{d} N(0, \sigma^2 \Sigma_X^{-1}) \tag{2}$$

with $\Sigma_X \equiv E[(X - E[X|Z])(X - E[X|Z])']$.

Our objective in this problem is to estimate $g(z)$ for some $z \in S_X \subseteq \mathbb{R}^{dz}$.

(a) Show that $g(z) = E[Y - X'\beta | Z = z]$.

(b) Based on (a), it is natural to estimate $g(z)$ using a non-parametric mean regression of $Y - X'\hat{\beta}$ on $Z = z$, i.e.,

$$\hat{g}(z) \equiv \frac{1}{nh_n^{dz} \hat{f}_Z(z)} \sum_{i=1}^{n} (Y_i - X_i'\hat{\beta}) K\left(\frac{z - Z_i}{h_n}\right)$$

with

$$\hat{f}_Z(z) \equiv \frac{1}{nh_n^{dz}} \sum_{i=1}^{n} K\left(\frac{z - Z_i}{h_n}\right).$$

Is this estimator consistent? Impose conditions as needed to establish your answer.

3. The goal of this problem is to explore the performance of Robinson (1988)'s semiparametric estimator in finite samples via simulations. Assume that $\{(Y_i, X_i, Z_i)\}_{i=1}^n$ is an i.i.d. sample that satisfies:

$$Y = X'\beta + g(Z) + U,$$

where $(Y, X, Z) : \Omega \to \mathbb{R} \times \mathbb{R} \times \mathbb{R}$ are the observables (i.e. $d_X = d_Z = 1$), $\beta \in \mathbb{R}$ is an unknown scalar, $g : \mathbb{R} \to \mathbb{R}$ is an unknown function, and U is an unobserved error term satisfies:

$$E[U|(X, Z) = (x, z)] = 0 \ \forall (x, z) \in S_{(X,Z)},$$

and $S_{(X,Z)}$ denotes the support of (X, Z). Furthermore, assume that $Z = X^2$.

Simulate $S = 10{,}000$ independent datasets with $n = 100$ observations each and the following true parameter values: $\beta = 0$, $g(z) = z$, and $(X', U)' \sim N(\mathbf{0}_2, \mathbf{I}_{2\times 2})$.

(a) For each dataset, compute the simulated bias, variance, and MSE of the following estimators:

i. Robinson (1988)'s estimator with trimming according to $c_n \in \{0, 0.001, 0.01, 1\}$,

ii. OLS coefficient of associated to X from running a regression of Y on X and a constant.

iii. OLS coefficient of associated to X from running a regression of Y on X, X^2, and a constant.

(b) Comment on your general findings.

(c) For each dataset, use Robinson (1988)'s estimator with trimming according to $c_n \in \{0, 0.001, 0.01, 1\}$ to test $H_0 : \beta = b$ vs $H_1 : \beta \neq b$ for $b \in \{-2, -1, -0.5, 0, 0.5, 1, 2\}$. Report the empirical rejection rate for each combination of (b, c_n). Comment on your findings.

References

Robinson, P. M. (1988): "Root-N-Consistent Semiparametric Regression," *Econometrica*, 56, 931–954.