Econ 675: HW 4

Erin Markiewitz

November 9, 2018

Contents

1 Estimating Equations

This question considers identification under selection-on-observables for the generaic class of parameters

$$\theta_t(g) = \mathbb{E}[g(Y_i(t))], \ g \in G, \ t \in T$$

where G denotes a class of functions (e.g. $G = \{\mathbf{1}(. \leq y) : y \in \mathbb{R})\}$). Define the regression functions:

$$p_t(\mathbf{X_i} = \P[T_i = t|$$
$$\mathbf{X_i}]), e_t(g; \mathbf{X_i}) = \mathbb{E}[g(Y_i(t))|\mathbf{X_i}] == \mathbb{E}[g(Y_i(t))|\mathbf{X_i}, T_i = 1], \ g \in G, \ t \in T$$

Assume Ignorability: $Y_i(t) \perp D_i(t)|\mathbf{X_i}$ and $0 < x < p_t(\mathbf{X_i})$, for all $t \in T$ and for some fixed positive constant c.

1.1

In the following section, we prove the validity of three moment cofitions for the generic class of parameters. The first moment condition is the Inverse Probabiltiy Weighting (IPW) moment condition.

$$\psi_{IPW,t}(\mathbf{Z_i}; \theta_t(g)) = \frac{D_i(t) \cdot g(Y_i(t))}{p_t(\mathbf{X_i})} - \theta_t(g)$$

To begin, take the expectation of the moment condition.

$$\mathbb{E}\left[\frac{D_i(t) \cdot g(Y_i(t))}{p_t(\mathbf{X_i})}\right] - \theta_t(g)$$

By the law of iterative expectations

$$\mathbb{E}\left[\mathbb{E}\left[\frac{D_i(t) \cdot g(Y_i(t))}{p_t(\mathbf{X_i})}|\mathbf{X_i}\right]\right] - \theta_t(g)$$

$$\mathbb{E}\left[\frac{1}{p_t(\mathbf{X_i})}\mathbb{E}\left[D_i(t) \cdot g(Y_i(t))|\mathbf{X_i}\right]\right] - \theta_t(g)$$

$$\mathbb{E}\left[\frac{1}{p_t(\mathbf{X_i})}\mathbb{E}\left[D_i(t)|\mathbf{X_i}\right] \cdot \mathbb{E}\left[g(Y_i(t))|\mathbf{X_i}\right]\right] - \theta_t(g)$$

As $\mathbb{E}\left[D_i(t)|\mathbf{X_i}\right] = \Pr\left[D_i(t) = 1|\mathbf{X_i}\right] = \Pr\left[T_i = t|\mathbf{X_i}\right] = p_t(\mathbf{X_i})$

$$\mathbb{E}\left[\frac{p_t(\mathbf{X}_i)}{p_t(\mathbf{X}_i)} \cdot \mathbb{E}\left[g(Y_i(t))|\mathbf{X}_i\right]\right] - \theta_t(g)$$

which gives us our result,

$$\mathbb{E}\left[\mathbb{E}\left[g(Y_i(t))|\mathbf{X}_i\right]\right] - \theta_t(g) = \mathbb{E}\left[g(Y_i(t))\right] - \theta_t(g) = \theta_t(g) - \theta_t(g) = 0$$

The second moment condition of this exercise is the Regression Imputation (1) moment condition:

$$\psi_{RI1,t}(\mathbf{Z}_i; \theta_t(g)) = e_t(g; \mathbf{X}_i) - \theta_t(g)$$

Take the expectation

$$\mathbb{E}\left[e_t(g; \mathbf{X}_i)\right] - \theta_t(g)$$

$$\mathbb{E}\left[\mathbb{E}\left[g(Y_i(t))|\mathbf{X}_i\right]\right] - \theta_t(g)$$

$$\mathbb{E}\left[g(Y_i(t))\right] - \theta_t(g) = \theta_t(g) - \theta_t(g) = 0$$

The second moment condition of this exercise is the Regression Imputation (2) moment condition, which includes inverse probabilty weighting:

$$\psi_{RI2,t}(\mathbf{Z}_i; \theta_t(g)) = \frac{D_i(t) \cdot e_t(g; \mathbf{X}_i)}{p_t(\mathbf{X}_i)} - \theta_t(g)$$

Take the expectation

$$\mathbb{E}\left[\frac{D_i(t) \cdot e_t(g; \mathbf{X}_i)}{p_t(\mathbf{X}_i)}\right] - \theta_t(g)$$

iterate the expectation a bit

$$\mathbb{E}\left[\frac{1}{p_t(\mathbf{X}_i)} p_t(\mathbf{X}_i) \cdot \mathbb{E}\left[g(Y_i(t))|\mathbf{X}_i\right]\right] - \theta_t(g)$$

which gives the result

$$\mathbb{E}\left[g(Y_i(t))\right] - \theta_t(g) = \theta_t(g) - \theta_t(g) = 0$$

Last, we consider the doubly robust estimator's moment condition:

$$\psi_{DR,t}(\mathbf{Z}_i; \theta_t(g)) = \frac{D_i(t) \cdot g(Y_i(t))}{p_t(\mathbf{X}_i)} - \theta_t(g) - \frac{e_t(g; \mathbf{X}_i)}{p_t(\mathbf{X}_i)} \cdot (D_i(t) - p_t(\mathbf{X}_i))$$

Take expectations

$$\mathbb{E}\left[\frac{D_i(t) \cdot g(Y_i(t))}{p_t(\mathbf{X}_i)}\right] - \theta_t(g) - \mathbb{E}\left[\frac{e_t(g; \mathbf{X}_i)}{p_t(\mathbf{X}_i)} \cdot (D_i(t) - p_t(\mathbf{X}_i))\right]$$

From previous results the first two terms cancel,

$$-\mathbb{E}\left[\frac{D_i(t) \cdot e_t(g; \mathbf{X}_i)}{p_t(\mathbf{X}_i)}\right] + \mathbb{E}\left[e_t(g; \mathbf{X}_i)\right]$$

The result follows from the law of iterated expectations.

1.2

The IPW plug-in estimator:

$$\hat{\psi}_{IPW,t}(\mathbf{Z}_i; \theta_t(g)) = \frac{1}{n} \sum_{i=1}^{n} \frac{D_i(t) \cdot g(Y_i)}{\hat{p}_t(\mathbf{X}_i)}$$

Where $\hat{p}_t(\mathbf{X}_i)$ is the estimated propensity score from the first-stage regression of the treatment on the covariates.

To write down the RI1 plug-in estimator, start by putting a hat on it:

$$\hat{\psi}_{RI1,t}(\mathbf{Z}_i) = \frac{1}{n} \sum_{i=1}^{n} \hat{e}_t(\mathbf{X}_i)$$

where $\hat{e}_t(X_i) = \mathbb{E}[g(Y_i(t))|\mathbf{X}_i, T_i = t]$, the conditional expectation of the class of regression functions specified by G. We can rewrite the estimator above as:

$$\hat{\psi}_{RI1,t}(\mathbf{Z}_i) = \frac{1}{n} \sum_{i=1}^{n} \frac{D_i(t) \cdot \hat{e}_t(\mathbf{X}_i)}{\hat{p}_t(\mathbf{X}_i)}$$

To write down the RI1 plug-in estimator, just reweight using the estimated propensity score:

$$\hat{\psi}_{RI2,t}(\mathbf{Z}_i) = \frac{1}{n} \sum_{i=1}^{n} \frac{D_i(t) \cdot \hat{e}_t(\mathbf{X}_i)}{\hat{p}_t(\mathbf{X}_i)}$$

And the double robust plug in estimator

$$\hat{\psi}_{DR,t}(\mathbf{Z}_i) = \frac{1}{n} \sum_{i=1}^{n} \frac{D_i(t) \cdot g(Y_i)}{\hat{p}_t(\mathbf{X}_i)} - \frac{1}{n} \sum_{i=1}^{n} \frac{\hat{e}_t(\mathbf{X}_i)}{\hat{p}_t(\mathbf{X}_i)} (D_i(t) - \hat{p}_t(\mathbf{X}_i))$$

The relative performance of the estimators depends on the data generating process. As the IPW and R2 plug in estimators use the estimated propensity score reweight the treatment effects, both estimators will be inconsistent in finite samples when the propensity score is very close to either one or zero. (If you are only estimating the treatement effect on the treated, it is sufficient that the propensity score is not degenerative with respect to 1.) The double robust estimators includes further safeguards against bias induced by misspecification but at the cost of imposing additional specification choices.

Then again, transparency is a key feature of an estimator - especially in policy analysis. Conditioning on covariates allow for specification of the propensity score without prior knowledge of the outcome variable/equation.

1.3

(Just a hunch) The estimating equations in section 1.1 can be used to estimate the variance of the potential outcome variables. First, we specify the function

$$g(x) = (x - \mathbb{E}[x])^2, \ x \in \mathbb{R}$$

and its finite sample analogue

$$\hat{g}(x) = \frac{1}{n} \sum_{i=1}^{n} (x_i - \frac{1}{n} \sum_{i=1}^{n} x_i)^2, \ x_i \in \mathbf{X} \in \mathbb{R}^n$$

The validity of these moment conditions is established in section 1.1 more generally. Under this specification of g, $\theta_t(g) = \mathbb{V}[Y_i(t)]$ and $e_t(g; \mathbf{X}_i) = \mathbb{V}[Y_i(t)|\mathbf{X}_i]$, the unconditional and conditional variance of the potential outcome of treatement t, respectively. This gives us the moment conditions:

$$\psi_{IPW,t}(\mathbf{Z}_i; \sigma_t^2) = \frac{D_i(t) \cdot (Y_i(t) - \mathbb{E}[Y_i(t)])^2}{p_t(\mathbf{X}_i)} - \sigma_t^2$$

$$\psi_{RI1,t}(\mathbf{Z}_i; \sigma_t^2) = \mathbb{E}[(Y_i(t) - \mathbb{E}[Y_i(t)])^2|\mathbf{X}_i] - \sigma_t^2$$

$$\psi_{RI2,t}(\mathbf{Z}_i; \sigma_t^2) = \frac{D_i(t) \cdot \mathbb{E}[(Y_i(t) - \mathbb{E}[Y_i(t)])^2 | \mathbf{X}_i]}{p_t(\mathbf{X}_i)} - \sigma_t^2$$

$$\psi_{DR,t}(\mathbf{Z}_i; \sigma_t^2) = \frac{D_i(t) \cdot (Y_i(t) - \mathbb{E}[Y_i(t)])^2}{p_t(\mathbf{X}_i)} - \sigma_t^2 - \frac{\mathbb{E}[(Y_i(t) - \mathbb{E}[Y_i(t)])^2 | \mathbf{X}_i]}{p_t(\mathbf{X}_i)} \cdot (D_i(t) - p_t(\mathbf{X}_i))$$

Now in order to conduct the hypothesis test of $\mathbf{H}_0 : \sigma_t^2 = \sigma^2$ we need to use the finite sample analogue of the g function specified above. The variance of our moment conditions will be estimated using a simple GMM procedure. In this case, we use a two step procedure for the moment conditions that use IPW. In the first step, we estimate the propensity score and drop any observations with propensity scores sufficiently close to zero or one. For a given moment condition $M \in \{IPW, RI1, RI2, DR\}$ and treatment t we define the finite sample analogue as $\hat{\psi}_{M,t}$

So GMM is

$$\hat{\mathbf{\Omega}}_{M,t} = \frac{1}{n} \sum_{i=1}^{n} \hat{\psi}_{M,t} \hat{\psi}_{M,t}'$$

From Theorem 12.7.1 in Hansens's Econometrics text,

$$\hat{\mathbf{V}}_{\psi,M,t} = (\hat{\psi}_{M,t} \hat{\mathbf{\Omega}}_{M,t}^{-1} \hat{\psi}_{M,t}')^{-1}$$
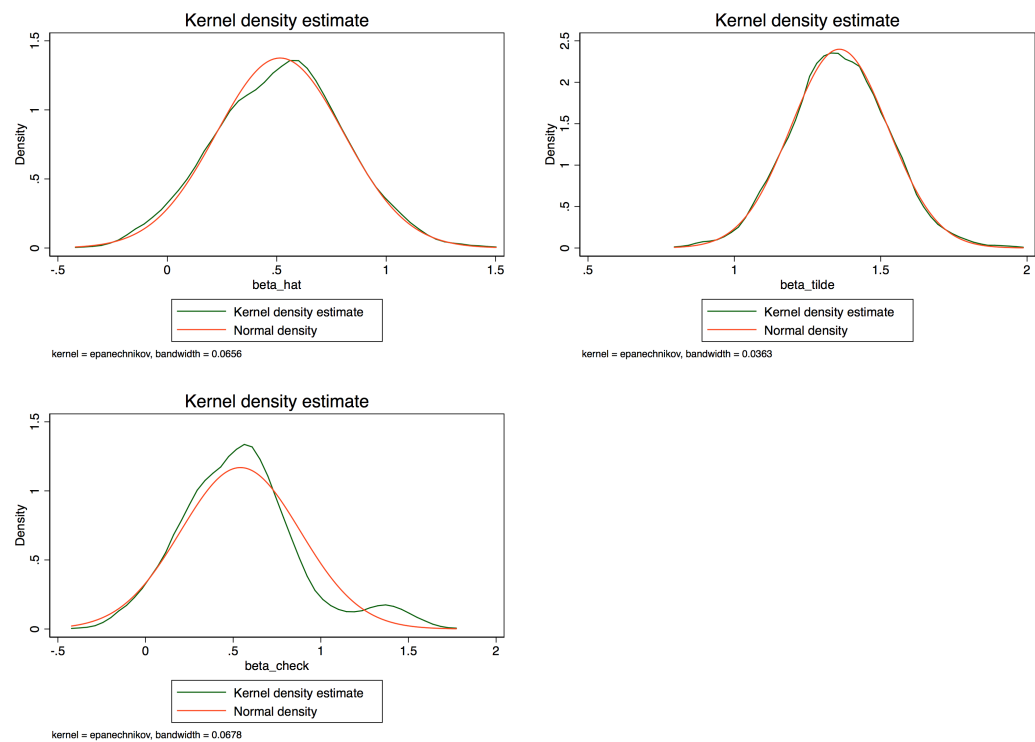
And so in order to conduct the hypothesis test we simply reject the null if $\hat{\sigma}_t^2$ is outside of the following confidence interval:

$$\mathbf{CI}_\alpha(\hat{\sigma}_t^2) = \left[\sigma^2 - \mathbf{\Phi}^{-1}(\frac{\alpha}{2})\sqrt{\frac{\hat{\mathbf{V}}_{\psi,M,t}}{n}}, \sigma^2 + \mathbf{\Phi}^{-1}(\frac{\alpha}{2})\sqrt{\frac{\hat{\mathbf{V}}_{\psi,M,t}}{n}} \right]$$

2

3

3.1

Stata Output:

	mean	sd	min	max
$\hat{\beta}$	.5153	.2900	-.3544	1.436
$\check{\beta}$	.5415	.3413	-.3544	1.705
$\tilde{\beta}$	1.358	.1663	.8311	1.950

3.2

Stata Output:

	Coverage Rate
$\hat{\beta}$	0.201
$\check{\beta}$	0.244
$\tilde{\beta}$	0

4 Code Appendix

Stata

R