ECON675 – Assignment 4

Anirudh Yadav

November 5, 2018

Contents

1 Estimating equations

1.1 Moment conditions

The goal of this question is to show that the four given functions are valid moment conditions for the parameter $\theta_t(g)$. That is, we want to show that

$$\mathbb{E}[\psi_{\texttt{f},t}(\boldsymbol{Z}_i;\theta_t(g))] = 0,$$

for each $\texttt{f} \in \{\texttt{IPW}, \texttt{RI1}, \texttt{RI2}, \texttt{DR}\}$. Note that in the derivations below I invoke LIE a lot without specifically mentioning it.

Start with the inverse probability weighting function

$$\begin{aligned}
\mathbb{E}[\psi_{\texttt{IPW},t}(\boldsymbol{Z}_i;\theta_t(g))] &= \mathbb{E}\left[\frac{D_i(t)\cdot g(Y_i(t))}{p_t(\boldsymbol{X}_i)}\right] - \theta_t(g) \\
&= \mathbb{E}\left[\mathbb{E}\left[\frac{D_i(t)\cdot g(Y_i(t))}{p_t(\boldsymbol{X}_i)}|\boldsymbol{X}_i\right]\right] - \theta_t(g) \\
&= \mathbb{E}\left[\frac{1}{p_t(\boldsymbol{X}_i)}\mathbb{E}\left[D_i(t)|\boldsymbol{X}_i\right]\mathbb{E}\left[g(Y_i(t))|\boldsymbol{X}_i\right]\right] - \theta_t(g)
\end{aligned}$$

Now,

$$\mathbb{E}\left[D_i(t)|\boldsymbol{X}_i\right] = \Pr[D_i(t)=1|\boldsymbol{X}_i] = \Pr[T_i = t|\boldsymbol{X}_i] = p_t(\boldsymbol{X}_i).$$

Thus,

$$\begin{aligned}
\mathbb{E}[\psi_{\texttt{IPW},t}(\boldsymbol{Z}_i;\theta_t(g))] &= \mathbb{E}\left[\mathbb{E}\left[g(Y_i(t))|\boldsymbol{X}_i\right]\right] - \theta_t(g) \\
&= \mathbb{E}[g(Y_i(t))] - \theta_t(g) \\
&= 0.
\end{aligned}$$

Next, consider

$$\begin{aligned}
\mathbb{E}[\psi_{\texttt{RI1},t}(\boldsymbol{Z}_i;\theta_t(g))] &= \mathbb{E}[e_t(g;\boldsymbol{X}_i)] - \theta_t(g) \\
&= \mathbb{E}[\mathbb{E}[g(Y_i(t)|\boldsymbol{X}_i]] - \theta_t(g) \\
&= \mathbb{E}[g(Y_i(t)] - \theta_t(g) \\
&= 0.
\end{aligned}$$

And,

$$\begin{aligned}
\mathbb{E}[\psi_{\texttt{RI2},t}(\boldsymbol{Z}_i;\theta_t(g))] &= \mathbb{E}\left[\frac{D_i(t)\cdot e_t(g;\boldsymbol{X}_i)}{p_t(\boldsymbol{X}_i)}\right] - \theta_t(g) \\
&= \mathbb{E}\left[\mathbb{E}\left[\frac{D_i(t)\cdot e_t(g;\boldsymbol{X}_i)}{p_t(\boldsymbol{X}_i)}|\boldsymbol{X}_i\right]\right] - \theta_t(g) \\
&= \mathbb{E}\left[\mathbb{E}\left[e_t(g;\boldsymbol{X}_i)|\boldsymbol{X}_i\right]\right] - \theta_t(g) \\
&= \mathbb{E}[e_t(g;\boldsymbol{X}_i)] - \theta_t(g) \\
&= 0.
\end{aligned}$$

Finally, consider the doubly robust function

$$\mathbb{E}[\psi_{\texttt{DR},t}(\boldsymbol{Z}_i;\theta_t(g))] = \mathbb{E}\left[\frac{D_i(t)\cdot g(Y_i(t))}{p_t(\boldsymbol{X}_i)}\right] - \theta_t(g) - \mathbb{E}\left[\frac{e_t(g;\boldsymbol{X}_i)}{p_t(\boldsymbol{X}_i)}(D_i(t) - p_t(\boldsymbol{X}_i))\right].$$

Using the IPW result above, we know that the first two terms cancel each other out, so that

$$\begin{aligned}
\mathbb{E}[\psi_{\texttt{DR},t}(\boldsymbol{Z}_i;\theta_t(g))] &= -\mathbb{E}\left[\frac{e_t(g;\boldsymbol{X}_i)}{p_t(\boldsymbol{X}_i)}(D_i(t) - p_t(\boldsymbol{X}_i))\right] \\
&= -\mathbb{E}\left[\frac{e_t(g;\boldsymbol{X}_i)D_i(t)}{p_t(\boldsymbol{X}_i)}\right] + \mathbb{E}[e_t(g;\boldsymbol{X}_i)] \\
&= -\theta_t(g) + \theta_t(g) \\
&= 0.
\end{aligned}$$

So each of the four functions is a valid moment condition for $\theta_t(g)$.

1.2 Plug-in estimators

The plug-in IPW estimator is

$$\hat{\theta}_{\texttt{IPW},t}(g) = \frac{1}{n}\sum_{i=1}^{n}\frac{D_i(t)g(Y_i)}{\hat{p}_t(\boldsymbol{X}_i)},$$

where $\hat{p}_t(\boldsymbol{X}_i)$ is the estimated propensity score. Note that since there are multiple treatment levels, the estimated propensity score would have to be computed using a suitable discrete choice model. For instance, $\hat{p}_t(\boldsymbol{X}_i)$ could be estimated using a multinomial logit model.

The plug-in projection (or regression imputation) estimator is

$$\begin{aligned}
\hat{\theta}_{\texttt{RI1},t}(g) = \hat{\mathbb{E}}[e_t(g;\boldsymbol{X}_i)] = \frac{1}{n}\sum_{i=1}^{n}\hat{\mathbb{E}}[g(Y_i(t))|\boldsymbol{X}_i] &= \frac{1}{n}\sum_{i=1}^{n}\hat{\mathbb{E}}[g(Y_i(t))|\boldsymbol{X}_i, D_i(t)=1] \\
&= \frac{1}{n}\sum_{i=1}^{n}\hat{\mathbb{E}}[g(Y_i)|\boldsymbol{X}_i, D_i(t)=1],
\end{aligned}$$

where the second last equality uses the ignorability assumption. We need to make a choice about how to estimate the conditional expectation term. I think we could use NLS, or possibly a nonparametric method like kernel regression. To ease notation, let $\widehat{\mu}_t(\boldsymbol{X}_i)$ be the parametric or nonparametric estimate of $\mathbb{E}[g(Y_i)|\boldsymbol{X}_i, D_i(t)=1]$. Then, the projection estimator is

$$\hat{\theta}_{\texttt{RI1},t}(g) = \frac{1}{n}\sum_{i=1}^{n}\widehat{\mu}_t(\boldsymbol{X}_i)$$

The plug-in 'hybrid' imputation estimator

$$\hat{\theta}_{\texttt{RI2},t}(g) = \frac{1}{n}\sum_{i=1}^{n}\frac{D_i(t)\widehat{\mu}_t(\boldsymbol{X}_i)}{\hat{p}_t(\boldsymbol{X}_i)}.$$

Finally, the plug-in doubly robust estimator is given by

$$\begin{aligned}\hat{\theta}_{\texttt{DR},t}(g) &= \frac{1}{n}\sum_{i=1}^{n}\frac{D_i(t)g(Y_i)}{\hat{p}_t(\boldsymbol{X}_i)} - \frac{1}{n}\sum_{i=1}^{n}\frac{\widehat{\mu}_t(\boldsymbol{X}_i)}{\hat{p}_t(\boldsymbol{X}_i)}(D_i(t)-\hat{p}_t(\boldsymbol{X}_i)) \\ &= \frac{1}{n}\sum_{i=1}^{n}\left(\frac{D_i(t)(g(Y_i)-\widehat{\mu}_t(\boldsymbol{X}_i))}{\hat{p}_t(\boldsymbol{X}_i)} + \widehat{\mu}_t(\boldsymbol{X}_i)\right).\end{aligned}$$

As discussed in Abadie and Catteneo (2018), the relative performance of the above estimators depends on the features of the data generating process. In finite samples, IPW estimators become unstable when the propensity score approaches zero or one and regression imputation estimators may suffer from extrapolation biases. Doubly robust estimators include safeguards against bias caused by misspecification but impose additional specification choices that may affect the resulting estimate.

1.3 Estimating the variance of potential outcomes

2 Estimating average treatment effects