Ec. 2120, Fall 2020 December 7, 2020

FINAL Exam

Instructions

You have 105 minutes to finish this exam. You also have 15 minutes to scan and upload your answers. So, the total time you are allowed to keep this exam is 120 minutes. Please plan accordingly.

1. (25pt) Suppose you have an iid sample $W_i = (Y_i, X_i)$, $i = 1, ..., n$ where both Y_i and X_i are binary 0/1 variables. One is interested in the parameters (β_0, β_1) from a logit model. The logit model assumes that

$$P(Y = 1|X) = exp(\beta_0 + \beta_1 X)/(1 + exp(\beta_0 + \beta_1 X)).$$

(a) (7.5pts) Derive the MLE estimator of (β_0, β_1)? What is its asymptotic distribution? Be precise about the asymptotic variance.

(b) (7.5pts) Derive a GMM estimator of (β_0, β_1). What is its asymptotic distribution? Be precise about the asymptotic variance.

(c) (10pts) Derive a minimum distance estimator of (β_0, β_1). What is its asymptotic distribution? Be precise about the asymptotic variance.

2. (15pts) We have considered in class estimating the parameter β in R^k defined through

$$E^*[Y|X] = X'\beta$$

We can use an iid sample $W_i = (Y_i, X_i')'$ of $i = 1, ..., N$ to get, b, the least squares estimator of β, where we showed that $\sqrt{n}(b - \beta) \xrightarrow{d} \mathcal{N}(0, \alpha\Sigma\alpha')$, with

$$\alpha = [E(X_i X_i')]^{-1}, \quad \Sigma = E((Y_i - X_i'\beta)^2 X_i X_i').$$

Show how we can set up the above as a GMM problem to get a GMM estimator of β. What is the asymptotic distribution of this GMM estimator of β? Here, use the efficient weight matrix. Discuss.

3. (30pts) This question is related to the role the propensity score plays.

(a) (20pts) Define

$$Q_i = P(T_i = 1|Z_i) = E(T_i|Z_i)$$

Q_i is known as the propensity score. (Note: we have used the term propensity score before in a different context, to denote $g(s) = Pr(T_i|S_i = s)$ where S_i is a vector of randomly assigned IVs.) Assume also selection on observables, i.e., that in the context of two treatments indexed by $t = 0$ or 1, the potential outcomes $Y_i(0)$ and $Y_i(1)$ are such that $(Y_i(0), Y_i(1))$ is independent of T_i conditional on Z_i. Use iterated expectations to show that if T_i is randomly assigned conditional on Z_i (or selection on observables), then

$$P(T_i = 1|Y_i(0), Q_i) = E(T_i|Y_i(0), Q_i) = Q_i$$
$$P(T_i = 1|Y_i(1), Q_i) = E(T_i|Y_i(1), Q_i) = Q_i$$

and so

$$Y_i(0) \perp\!\!\!\perp T_i|Q_i \quad and \quad Y_i(1) \perp\!\!\!\perp T_i|Q_i.$$

(b) (10pts) We have considered a model in which treatment is randomly assigned conditional on some observed (baseline, pretreatment) covariates. We have also considered a model in which an instrumental variable is randomly assigned. In the case of a binary treatment, the propensity score plays a role in both models. With conditional random assignment of treatment, it is the probability of treatment conditional on the observed covariates. With random assignment of the instrumental variable, it is the probability of treatment conditional on the instrumental variable. Suppose that the propensity score is given and discuss its use in these two models. What are the similarities and differences?

4. (20pts) Let Y = duration measured by weeks of temporary total benefits paid; $Z_1 = 1$ if injured after the benefit increase, $Z_1 = 0$ otherwise; $Z_2 = 1$ if high earnings group, $Z_2 = 0$ otherwise. We shall only be using these three variables (and functions of them). Kentucky raised the benefit ceiling for high earners from \$131 to \$217 per week on July 15, 1980. We shall regard the data $\{(Y_i, Z_{i1}, Z_{i2})\}_{i=1}^n$ as i.i.d. from some unknown distribution F. Let (Y, Z_1, Z_2) be a random vector with distribution F, so that we can sometimes simplify notation by omitting the i subscript. Define the following functions of (Z_1, Z_2):

$$X_1 = Z_1 \cdot Z_2, \quad X_2 = Z_1, \quad X_3 = Z_2, \quad X_4 = 1.$$

Let g be a given function, which is strictly increasing. Examples include $g(Y) = Y$ and $g(Y) = \log(Y + .5)$, but for now we just want to be thinking of some given, strictly increasing function. We are going to define treatment effects on $g(Y)$ of the benefit ceiling increase.

Let r denote the regression function:

$$r(z_1, z_2) = E_F[g(Y) \,|\, Z_1 = z_1, Z_2 = z_2] \quad \text{for} \quad (z_1, z_2) \in \{(0,0), (0,1), (1,0), (1,1)\}.$$

The domain of the function r consists of four points. Let

$$E_F^*(g(Y) \,|\, X_1, \ldots, X_4) = \sum_{k=1}^{4} \beta_k X_k = X'\beta \quad \text{with} \quad X = \begin{pmatrix} X_1 \\ X_2 \\ X_3 \\ X_4 \end{pmatrix}, \quad \beta = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \end{pmatrix}$$

denote the (minimum mean-square-error) linear predictor. We know from Note 2 that the regression function can be expressed using the linear predictor coefficients β. This tells us that the data we have will allow us to estimate the 4 coefficients $\beta_1, \beta_2, \beta_3, \beta_4$ which correspond to estimating $r(1,1), r(0,1), r(1,0)$, and $r(0,0)$.

(a) (10pts) We want to define the effect of the ceiling increase on those in the high earnings group injured after the increase. This requires specifying a counterfactual $g(Y^*)$ corresponding to what would have been observed in the absence of the benefit increase. Here are two possible specifications:

(i) $E_F(g(Y^*) \,|\, 1,1) = r(0,1)+[r(1,0)-r(0,0)]$, (ii) $E_F(g(Y^*) \,|\, 1,1) = r(0,1)\cdot[r(1,0)/r(0,0)]$.

Briefly provide some interpretation and motivation for each of these specifications.

(b) (5pts) Given a specification for $E_F(g(Y^*) \,|\, 1,1)$, we can define an absolute treatment effect

$$\gamma_A = r(1,1) - E_F(g(Y^*) \,|\, 1,1).$$

Note that for specification (i) there is a "difference in differences" form:

$$\gamma_A = [r(1,1) - r(0,1)] - [r(1,0) - r(0,0)],$$

but this is not the case for specification (ii). Suppose that we know the joint distribution F. Is there anything about F that would indicate which of the two specifications for the counterfactual is more appropriate? Discuss briefly.

(c) (5pts) Given a specification for $E_F(g(Y^*) \,|\, 1,1)$, we can define a relative treatment effect

$$\gamma_R = r(1,1)/E_F(g(Y^*) \,|\, 1,1).$$

Note that for specification (ii) there is a "ratio of ratios" form:

$$\gamma_R = [r(1,1)/r(0,1)] \Big/ [r(1,0)/r(0,0)],$$

and so a "difference of differences" form in logs:

$$\log \gamma_R = [\log r(1,1) - \log r(0,1)] - [\log r(1,0) - \log r(0,0)],$$

but this is not the case for specification (i). Suppose that we know the joint distribution F. Is there anything about F that would indicate which of the two specifications for the counterfactual is more appropriate? Discuss briefly.

5. (10pts) The 2SLS is always efficient. Discuss.