The due date for this assignment is Wednesday, December 13.

1. Consider a probit model $y^*_t = \beta_0 + \beta_1 x^*_t + \epsilon_t$, but you only observe $y_t = 1$ if $y^*_t > 0$ and $y_t = 0$ if $y^*_t < 0$. Suppose $x^*_t \sim N(0, 1)$, $\epsilon_t \sim N(0, 1)$ and is independent of $x^*_t$. Now suppose you don’t observe $x^*_t$ but you only observe $x_t = x^*_t + u_t$, where $u_t \sim N(0, \sigma^2_u)$ and is independent of both $x^*_t$ and $\epsilon_t$. Now suppose you obtain the probit maximum likelihood estimator of $\beta_0$ and $\beta_1$ assuming that $x_t$ is the true regressor $x^*_t$. Call these $\hat{\beta}_0$ and $\hat{\beta}_1$. Find the probability limits and the asymptotic distributions of $\hat{\beta}_0$ and $\hat{\beta}_1$.

2. Consider the following model:

(a) $y_{1i} = \beta x_{1i} + u_i$.
(b) $y_{2i} = \beta x_{2i} + u_i$.
(c) $U_{1i} = y_{1i} + v_{1i}$.
(d) $U_{2i} = y_{2i} + v_{2i}$.
(e) $u_i = \alpha_1 v_{1i} + \alpha_2 v_{2i} + w_i, i = 1, 2, \ldots, n$.

Define

$$y_i = \begin{cases} y_{1i} & \text{if } U_{1i} - U_{2i} > 0 \iff D_i = 1 \iff A_i \\ y_{2i} & \text{if } U_{1i} - U_{2i} < 0 \iff D_i = 0 \iff \bar{A}_i \end{cases}$$

Assume that $(v_{1i}, v_{2i})$ are i.i.d. $N(0, 0, 1, 1, \rho)$, $(w_i)$ are i.i.d. $N(0, \sigma^2)$ independent of $(v_{1i}, v_{2i})$, and that the statistician observes $D_i, y_i, x_{1i}, x_{2i}$ for every $i$. The unknown parameters of the model are $\alpha_1, \alpha_2, \beta, \rho$, and $\sigma^2$.

(a) Derive the likelihood function.
(b) Define a two step consistent estimator of $\beta$ simpler to compute than the maximum likelihood estimator.

3. Consider the strong ignorability (conditional independence) assumption:

$$Y_1, Y_0 \perp D | X.$$ 

You are interested in estimating a quantile treatment effect on the treated parameter $Q_r (Y_1 | D = 1) - Q_r (Y_0 | D = 1)$, where for $r = 0, 1, Q_r (Y_r | D = 1)$ denotes the $\tau$th quantile of $Y_r$ among the treated population of $D = 1$. You have an i.i.d set of observations of $Y_i, X_i, D_i$, where $Y = D Y_1 + (1 - D) Y_0$. 

1
(a) Suppose the propensity score function follows a (correctly specified) logic model:

\[ p(D = 1|X) = \frac{e^{X'y}}{1 + e^{X'y}}. \]

Describe a method of estimating \( Q_\tau(Y_1|D = 1) - Q_\tau(Y_0|D = 1) \) using inverse propensity weighting, where the propensity score function is estimated using the correctly specified logic model.

(b) (*) Derive the asymptotic distribution of your estimator.

(c) Show how you can estimate the asymptotic distribution consistently.

(d) (*) Suppose now \( p(D = 1|X) \) does not have a parametric functional form and is completely nonparametrically specified. Redo the previous three questions.

4. Consider the conditional version of the LATE model in Imbens and Angrist (1994):

(1) Monotonicity: \( D_1 \geq D_0 \); (2) Conditional Independence: \( Y_1, Y_0, D_1, D_0 \perp Z | X \); (3) Dummy instrument: \( Z = 0, 1 \). To simplify, suppose \( X \) is a scalar (one dimensional).

(a) Suppose you want to estimate a conditional local average treatment effect parameter

\[ \beta(x) \equiv E(Y_1 - Y_0 | D_1 > D_0, X = x). \]

Consider the following functional 2SLS regression estimators \( \hat{\alpha}(x) \) and \( \hat{\beta}(x) \) defined by the sample moment conditions

\[ \frac{1}{n} \sum_{i=1}^{n} \left( \frac{1}{Z_i} \right) \left( y_i - \hat{\alpha}(x) - \hat{\beta}(x) D_i \right) \kappa \left( \frac{x_i - x}{h} \right) = 0. \]

where \( \kappa(\cdot) \) is a symmetric density function. Specify conditions under which \( \hat{\beta}(x) \xrightarrow{p} \beta(x) \) for each \( x \).

(b) Derive the asymptotic distribution of \( \sqrt{n}h \left( \hat{\beta}(x) - \beta(x) \right) \) for each \( x \). Specify suitable regularity conditions that you need.

(c) Frolich (2006) considers the estimation of an unconditional local average treatment effect (ULATE) parameter, defined as

\[ E(Y_1 - Y_0 | D_1 > D_0), \]

without conditioning on \( X = x \), in this model under assumptions (1), (2) and (3). See


Define the propensity score function as \( Q(x) = P(Z = 1|X = x) \). Describe two methods of estimating ULATE, one using the propensity score and one without using the propensity score.
5. Jacknife: A brief account is given in Amemiya pp135. In general, let $\hat{\theta}$ be an estimator using all data. Let $\hat{\theta}_{-i}$ be the estimator obtained by omitting observation $i$. The $i$th jacknife pseudovalue is given as $\theta_i^* = n\hat{\theta} - (n - 1)\hat{\theta}_{-i}$. The Jacknife estimator is the average of these $n$ of $\theta_i^*$: $\hat{\theta}_J \equiv \frac{1}{n}\sum_{i=1}^{n}\theta_i^*$.

Suppose we estimate $\mu^2 = (EX)^2$ by the bootstrap principle, $\hat{\theta} = \bar{X}^2$. This is a biased estimate. For the Jacknife, $\hat{\theta}_{-i} = \left(\frac{1}{n-1}\sum_{j\neq i}X_j\right)^2$. So

$$\hat{\theta}_J = n\bar{X}^2 - \frac{(n - 1)}{n}\sum_{i=1}^{n}\left(\frac{1}{n-1}\sum_{j\neq i}X_j\right)^2$$

For this example, simplify $\hat{\theta}_J$ and show that $\hat{\theta}_J$ is unbiased.

6. (This is not an easy question) Consider a standard treatment effect model $Y_i = D_iY_{i1} + (1 - D_i)Y_{i0}$, with the usual notations from the lecture notes where $D_i$ is the dummy variable indicating treatment status. $D_i = 1$ if individual $i$ belongs to the treatment group, and $D_i = 0$ if individual $i$ belongs to the control group. You also have a set of covariates $X_i$, and an i.i.d. data set of $Y_i, D_i, X_i, i = 1, \ldots, n$.

You are willing to assume that the treatment assignment is completely random. In other words, $D_i$ is statistically independent of $Y_{i1}, Y_{i0}, X_i$: $(Y_{i1}, Y_{i0}, X_i)|D_i$. You are interested in estimating the average treatment effect $E(Y_{i1} - Y_{i0})$ and you consider three estimators.

(a) In the first estimator, you calculate the difference between the subsample mean of $Y$ when $D = 1$ and when $D = 0$. Define this as $\hat{\beta}_1$.

(b) In the second estimator, you regress $Y$ on a constant term and $D$. Call the estimate of the slope coefficient on $D$ $\hat{\beta}_2$.

(c) In the third estimator, you regress $Y$ on a constant term and $D - p$ and the covariates $X$, and the interaction between $D - p$ and $X$. Call the estimate of the slope coefficient on $D - p$ $\hat{\beta}_3$.

For convenience you may assume that $EX = 0$, and with no loss of generality you can replace the independent variable $D$ by $D - p$ where $p = P(D = 1)$ in formulating the regression function for $\hat{\beta}_2$ and $\hat{\beta}_3$ above. You can assume you know $p$ when analyzing these regressions. You can also assume that $X$ is a scalar although this is not necessary.

The questions are in the following.

(a) Show that all three estimators consistently estimate the average treatment effect.

(b) Derive the asymptotic distribution for $\hat{\beta}_1$ and provide a consistent estimate of the asymptotic variance.
(c) Derive the asymptotic distribution for $\hat{\beta}_2$ and provide a consistent estimate of the asymptotic variance.

(d) If you use Stata to run the regression to estimate $\hat{\beta}_2$, will either the usual homoscedastic standard errors or the White robust standard errors provide a consistent estimate of the asymptotic variance for $\hat{\beta}_2$?

(e) Derive the asymptotic distribution for $\hat{\beta}_3$ and provide a consistent estimate of the asymptotic variance.

(f) If you use Stata to run the regression to estimate $\hat{\beta}_3$, will either the usual homoscedastic standard errors or the White robust standard errors provide a consistent estimate of the asymptotic variance for $\hat{\beta}_3$?

(g) Can you rank the (correctly calculated) asymptotic variances between $\hat{\beta}_1$, $\hat{\beta}_2$ and $\hat{\beta}_3$? For example, can you identify which of the three has the smallest asymptotic variance?

References:

http://projecteuclid.org/DPubS?service=UI&version=1.0&verb=Display&handle=euclid.aos/1176347602