Economics 696F, Causal Inference and Program Evaluation

Problem Set 2: due Friday, March 2

1. Write a program (in any language you like) which takes observations from two samples, and calculates the Mann-Whitney/Wilcoxon rank test for equality of two distributions, **and** generates a one-sided p-value by simulation.

2. Revisit the data in `ps1a.txt`, and use the Mann-Whitney program you calculated above to test the hypothesis that the distribution of $Y(1)$ and $Y(0)$ are the same (vs. the alternative that the distribution of $Y(1)$ is a positive shift of the distribution of $Y(0)$).

3. Revisit the data in `ps1b.txt` and the setup in HW1, Question 3.

   (a) Assume the propensity score has the form

   $$Pr(T = 1|X = x) = \frac{\exp(\gamma_1 + x'\gamma_2)}{1 + \exp(\gamma_1 + x'\gamma_2)}$$

   Estimate the propensity score using MLE, and then estimate the treatment effect on the treated using k-nearest-neighbor matching. Use $k = 1$, and $k = 5$, and discuss the results. You do not need to provide standard errors, just point estimates.

   (b) Repeat (c), but in the logit model use x along with all pairwise interactions of the x variables and squared age, education, and previous earnings. Discuss your results.

4. Consider the standard causal model with binary treatment, and assume unconfoundedness (and the support condition). Suppose we know the propensity score

   $$p(x) := Pr(T_i = 1|X_i = x).$$

   Consider the propensity score weighting estimator

   $$\hat{\tau}_w := \frac{1}{n}\sum_{i=1}^{n} \frac{T_i Y_i}{p(X_i)} - \frac{1}{n}\sum_{i=1}^{n} \frac{(1 - T_i)Y_i}{1 - p(X_i)}.$$

   Derive the asymptotic distribution of this estimator; in particular, provide the asymptotic variance as explicitly as possible.

5. Suppose (y_i, x_{i1}, x_{i2}) are i.i.d. with

   $$y_i = x_{i1}'\beta_1 + x_{i2}\beta_2 + \epsilon_i,$$

where x_{i1} is $k \times 1$ and x_{i2} is a scalar, for $i = 1, \ldots, n$. Let $x_i = (x'_{i1}, x_{i2})'$, and $\beta = (\beta'_1, \beta_2)'$. Assume $E(\epsilon_i | x_i) = 0$, so that the least squares regression of y_i on x_i would be consistent for β. In answering the following questions, state any additional assumptions needed.

(a) Let $w_i = (x'_{i1}, x^2_{i2})'$. Suppose we estimate β using TSLS, with instruments w_i. Give an explicit expression for $\hat{\beta}_{TSLS}$.

(b) Is $\hat{\beta}_{TSLS}$ consistent? (Prove your answer.)

(c) What is the limiting distribution of $\sqrt{n}(\hat{\beta}_{TSLS} - \beta)$, allowing for possible heteroskedasticity?

(d) Suppose that instead of using the instruments w_i, we estimate β using TSLS with instrument $z_i = (x'_{i1}, -x_{i2})'$. Provide an expression for the TSLS estimator, simplifying it until it is easily interpretable.