Homework #1

Due in class 9/9/16
Please work in groups of 2-4
Answers must be typed

1. **Identification I (OLS)**. Consider the linear regression model:

$$Y_i = X_i'\beta + u_i$$

where Y_i and u_i are scalar random variables and X_i is a $K \times 1$ random vector. The structures in this model can be indexed by $\theta \equiv \left\{\beta, F_X(.), F_{u|X}(.)\right\}$ where $F_X(.)$ is the distribution function of X_i and $F_{u|X}(.)$ is the distribution function of u_i given X_i. Denote the joint distribution of the observed variables (Y_i, X_i) by $F_{Y,X}(.,.)$ and the corresponding joint distribution implied by a structure θ as $F_\theta(.,.)$. Suppose the following two conditions hold:

A1) $E_\theta[u_i|X_i] = 0$

A2) $E_\theta[X_i X_i']$ is nonsingular

Prove that θ is point identified.

2. **Identification II (A Structural Labor Supply Model).** Consider a population of agents with the following utility function:

$$\begin{aligned} u_i(h_i) &= e_i - \frac{d_i}{1+a} h_i^{1+a} \\ &= w_i h_i - \frac{d_i}{1+a} h_i^{1+a} \end{aligned}$$

where $e_i = w_i h_i$ is weekly earnings, w_i is the wage, h_i is hours worked per week, $a \in (0, \infty)$ is a parameter governing curvature in the utility function, and d_i is a random variable representing heterogeneity in tastes for work.

Suppose $\log(d_i) \overset{iid}{\sim} N(\mu, \sigma)$ and that the choice set for hours per week is $\{20, 40\}$. Agents are assumed to maximize utility. We observe hours worked and earnings, but not d_i.

a) Derive the payoffs (u_{20}, u_{40}) to working 20 and 40 hours per week respectively.

b) Derive the probability of working 40 hours per week given wages.

c) Derive an individual's contribution to the likelihood of the observed data

d) To which estimator does this likelihood correspond?

e) Which parameters (or combinations of parameters) are identified?

Suppose now that the choice set for hours is $\{10, 20, 40\}$,

f) Derive an expression for the probability of working 20 hours per week and the probability of working 40 hours per week.

g) Which parameters (or combinations of parameters) are identified?

3. Identification III (Mixture of Normals). Consider the following mixture of normals model:

$$Y_i \sim \begin{cases} N(0,1) & w/ \text{ probability } p \\ N(0,\sigma^2) & w/ \text{ probability } 1-p \end{cases} \quad \text{where } p \in (0,1), \sigma^2 > 0$$

List a structure in the model space that is not identified (globally or locally).

4. **Quantile Treatment Effects**. The τ-th quantile of a continuous random variable Y_i is defined

$$Q_\tau(Y_i) \equiv F_Y^{-1}(\tau),$$

where $F_Y(y) = Pr[Y_i \leq y]$ is the CDF of Y_i.

a) Let $Z_i = g(Y_i)$, where $g(\cdot)$ is a monotonically increasing function. Prove that $Q_\tau(Z_i) = g(Q_\tau(Y_i))$.

b) Consider an experiment with a treatment variable $T_i \in \{0,1\}$ and an observed outcome Y_i. Let $Y_i(1)$ and $Y_i(0)$ denote potential values of Y_i with and without treatment. Define the τ-th quantile treatment effect

$$QTE(\tau) = Q_\tau(Y_i(1)) - Q_\tau(Y_i(0)).$$

Assume that potential outcomes are independent of T_i. Show that $QTE(\tau)$ is identified for every τ.

c) Suppose the following condition holds with probability one:

$$F_1(Y_i(1)) = F_0(Y_i(0)),$$

where $F_t(y)$ is the CDF of $Y_i(t)$. Let $\delta_i = Y_i(1) - Y_i(0)$. Show that $Q_\tau(\delta_i)$ is identified for every τ. Comment on the interpretation of $Q_\tau(\delta_i)$.

d) You have an *iid* sample $\{Y_i, T_i\}_{i=1}^N$ that satisfies the assumptions in parts b) and c). Propose an estimator of $Q_\tau(\delta_i)$.

5. **Iterated Projections.** Prove the law of iterated projections

$$E^*[Y_i|X_i] = E^*[E^*[Y_i|X_i, Z_i]|X_i]$$

6. **FWL theorem**. Suppose the population projection $E^*[Y_i|X_i, Z_i] = X_i'\beta + Z_i'\gamma$. Let $\widetilde{Z}_i = Z_i - E^*[Z_i|X_i]$ and $\widetilde{Y}_i = Y_i - E^*[Y_i|X_i]$:

a) Show that $\gamma = E\left[\widetilde{Z}_i \widetilde{Z}_i'\right]^{-1} E\left[\widetilde{Z}_i \widetilde{Y}_i\right]$

b) Show that $Y_i - E^*[Y_i|X_i, Z_i] = (Y_i - E^*[Y_i|X_i]) - (Z_i - E^*[Z_i|X_i])'\gamma$

7. **Weighted average derivative properties**

a) Show that for scalar random variables Y and X and $E[Y|X=x] = \mu(x)$ continuously differentiable:

$$\frac{cov(Y_i, X_i)}{var(X_i)} = \frac{\int \mu'(x)\,\omega(x)\,dx}{\int \omega(x)\,dx}$$

where $\omega(x) = \{E[X|X \geq x] - E[X|X < x]\}\{P(X_i \geq x)[1 - P(X_i \geq x)]\}$.

b) Derive the expression for the weights $\omega(X_i)$ when X_i is normally distributed (hint: you may want to look up the formula for the conditional expectation of a truncated normal. This is available in most textbooks, but also on Wikipedia). What does OLS estimate in this situation?

c) Prove that for scalar random variables Y and S and a vector random variable X_i

$$\frac{E[Y_i(S_i - E[S_i|X_i])]}{E[S_i(S_i - E[S_i|X_i])]} = \frac{E\left[\int \mu_X'(t)\,\omega_X(t)\,dt\right]}{E\left[\int \omega_X(t)\,dt\right]}$$

where $\mu_X(t) = E[Y_i|X_i, S_i = t]$ and $\omega_X(t)$ is a weighting function. What are the expressions for $\mu_X(t)$ and $\omega_X(t)$?

d) When does the ratio $\frac{E[Y_i(S_i - E[S_i|X_i])]}{E[S_i(S_i - E[S_i|X_i])]}$ coincide with the population regression coefficient on S in a regression of Y_i on X_i and S_i?