Homework #1

Professor: Pat Kline

Students: Christina Brown, Sam Leone, Peter McCrory, Preston Mui

Identification I (OLS)

Prove that $\theta \equiv \{\beta, F_X(\cdot), F_{u|X}(\cdot)\}$ is point identified in OLS:

First, $F_X(\cdot)$ is point identified from the joint distribution of $F_{Y,X}$ from $F_X(x) = \int_y \frac{\partial^2 F_{Y,X}}{\partial x \partial y}(y, x)dy$.
Then,

$$
\begin{aligned}
X_i Y_i &= X_i X_i' \beta + X_i u_i \\
E[X_i Y_i] &= E[X_i X_i']\beta + E[X_i u_i] \\
E[X_i Y_i] &= E[X_i X_i']\beta &&\text{(by A1)} \\
E[X_i X_i']^{-1} E[X_i Y_i] &= \beta &&\text{(by A2)}
\end{aligned}
$$

And so β is point identified. Finally, with β identified and using $F_{Y,X}$, one can derive the distribution of $F_{u|X}(\cdot)$ by

$$
\begin{aligned}
F_{u|X}(t) &= P(u_i < t | X_i) \\
&= P(Y_i - X_i'\beta < t | X_i) \\
&= P(Y_i < t + X_i'\beta | X_i) \\
&= \int_{y < t + X_i'\beta} dF_{Y|X}(y|x)dy
\end{aligned}
$$

where $F_{Y|X}$ is the conditional c.d.f. of Y given X implied by the j.d.f. of Y and X.

Identification II (A Structural Labor Supply Model)

a) Derive the payoffs to working 20 and 40 hours:

$$
\begin{aligned}
u_{20}(w_i; d_i, a) &= 20 w_i - \frac{d_i}{1+a} 20^{1+a} \\
u_{40}(w_i; d_i, a) &= 40 w_i - \frac{d_i}{1+a} 40^{1+a}
\end{aligned}
$$

b) Derive the probability of working 40 hours a week given wages:

Given wages, the worker will work 40 hours if $u_{40}(w_i; d_i, a) > u_{20}(w_i; d_i, a)$, which

holds when

$$40w_i - \frac{d_i}{1+a}40^{1+a} > 20w_i - \frac{d_i}{1+a}20^{1+a}$$
$$2w_i - \frac{2d_i}{1+a}40^a > w_i - \frac{d_i}{1+a}20^a$$
$$w_i(1+a) > 2d_i40^a - d_i20^a$$
$$\implies d_i < \frac{w_i(1+a)}{2\cdot40^a - 20^a}$$
$$\implies \log d_i < \log w_i + \log(1+a) - \log(2\cdot40^a - 20^a)$$

The probability of which is $\Phi\left(\frac{\log w_i + \log(1+a) - \log(2\cdot40^a - 20^a) - \mu}{\sigma}\right)$, where Φ is the c.d.f. of the standard normal.

c) Derive an individual's contribution to the likelihood of the observed data: Given an observation (h_i, w_i) under a given parameter a and parameters on the distribution of $\log d_i$, the likelihood of the observed data is $\Phi\left(\frac{\log w_i + \log(1+a) - \log(2\cdot40^a - 20^a) - \mu}{\sigma}\right)$ if $h_i = 40$ and $1 - \Phi\left(\frac{\log w_i + \log(1+a) - \log(2\cdot40^a - 20^a) - \mu}{\sigma}\right)$ if $h_i = 20$.

d) These likelihoods are, naturally, used in the maximum likelihood estimator.

e) The parameters (a, μ, σ) are set-identified (any two will imply the third).

f) Derive an expression for the probability of working 20 and 40 hours a week when the choice set for hours is $\{10, 20, 40\}$: The utility from working 20 and 40 hours is as above, and the utility of working 10 hours is

$$u_{10}(w_i; d_i, a) = 10w_i - \frac{d_i}{1+a}10^{1+a}$$

A worker works 40 hours when $u_{40} > u_{20}$ (as above) and $u_{40} > u_{10}$. The latter happens when

$$40w_i - \frac{d_i}{1+a}40^{1+a} > 10w_i - \frac{d_i}{1+a}10^{1+a}$$
$$4w_i - \frac{4d_i}{1+a}40^a > w_i - \frac{d_i}{1+a}10^a$$
$$3w_i(1+a) - 4d_i40^a > -d_i10^a$$
$$\implies d_i < \frac{3w_i(1+a)}{4\cdot40^a - 10^a}$$

However, this condition is met when $u_{40} > u_{20}$, unless $a < 0$ which is outside of the parameter space. So, the probability that the worker chooses 40 hours the same as in the $\{20, 40\}$ choice set.

Using a similar derivation to the above, the worker chooses 20 hours over 10 hours

when $d_i < \frac{w_i(1+a)}{2 \cdot 20^a - 10^a}$. So, the probability that the worker chooses 20 hours is

$$\Phi\left(\frac{\log w_i + \log(1+a) - \log(2 \cdot 20^a - 10^a) - \mu}{\sigma}\right)$$
$$-\Phi\left(\frac{\log w_i + \log(1+a) - \log(2 \cdot 40^a - 20^a) - \mu}{\sigma}\right)$$

g) With the addition of the option to work 10 hours, the parameters are still set-identified, in the sense that choosing one value would imply the others.

Identification III (Mixture of Normals)

We can write this problem as

$$Y_i = D_i X_i + (1 - D_i)W_i \text{ with } D_i \times X_i \times W_i \sim Bernoulli(\rho) \times N(0,1) \times N(0, \sigma^2).$$

The structure is thus pinned down by the parameters $\mu = (\sigma^2, \rho)$. A structure in this model space that is not identified–either globally or locally–is $\mu' = (1, \rho')$ for some $\rho' \in \mathbb{R}$.

By the law of total probability,

$$f_{\mu'}(y) = \rho'\phi(y) + (1 - \rho')\phi(y) = \phi(y) = \tilde{\rho}\phi(y) + (1 - \tilde{\rho})\phi(y) = f_{\tilde{\mu}}(y)$$

for some $\tilde{\rho} \in B_\epsilon(\rho')$, where $B_\epsilon(\rho')$ is an epsilon ball around the true parameter ρ'.

Thus, for any value of ρ, if the variance of W_i is equal to 1, we cannot identify the structure generating the observed data.

Quantile Treatment Effects

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

$$\begin{aligned}
F_Z(g(Q_\tau(Y_i))) &= Pr(Z_i < g(Q_\tau(Y_i))) \\
&= Pr(Y_i < Q_\tau(Y_i)) \quad \text{(Since } g(\cdot) \text{ is monotonically increasing)} \\
&= F_Y(Q_\tau(Y_i)) = \tau \\
\implies g(Q_\tau(Y_i)) &= F_Z^{-1}(\tau) = Q_\tau(Z_i) \quad\quad\quad \square
\end{aligned}$$

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 $\tau - th$ 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.

Proof.

Fix τ. Consider $Q_\tau(Y_i(1)|T_i = 1)$, the quantile of potential outcomes for the treatment, conditional on treatment. By definition (with t as the indicator variable for integration), we have

$$\tau = \int_{-\infty}^{Q_\tau(Y_i(1)|T_i=1)} t dF_{Y_i(1)|T_i=1}(t|T_i = 1)$$

$$\tau = \int_{-\infty}^{Q_\tau(Y_i(1)|T_i=1)} t dF_{Y_i(1)}(t) \qquad \text{By independence of the treatment}$$

$$\implies$$

$$Q_\tau(Y_i(1)|T_i = 1) = Q_\tau(Y_i(1))$$

A similar argument holds for $Q_\tau(Y_i(0))$. Thus, we have

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

$$\square$$

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 $Q_\tau(\delta_i)$.

Proof.

We can write the condition provided in this problem as

$$Pr(F_1(Y_i(1)) - F_0(Y_i(0)) < \epsilon) = 1 \; \forall \epsilon > 0$$

Let $Y_i^*(1)$ be an element in the support of $Y_i(1)$. Set $\tau = F_1(Y_i^*(1))$, so that $Q_\tau(Y_i(1)) = Y_i^*(1)$. Then, define $Y_i^*(0) \equiv Q_\tau(Y_i(0))$. Then, the condition above is equivalent to

$$\delta_i^* = Y_i^*(1) - Y_i^*(0)$$

holding with probability one.

Since $Y_i^*(0)$ is implicitly defined by our choice of $Y_i^*(1)$, we have that $\delta_i^* = h(Y_i^*(1))$ with probability one for some function $h : \mathbb{S}(Y_i(1)) \to \mathbb{R}$, where $\mathbb{S}(Y_i(1))$ is the support of the random variable $Y_i(1)$.

Denote $\tilde{\tau}(x) = F_{Y_i(1)}(x)$ as the the cdf of $Y_i(1)$ evaluated at x.

The cumulative density function of δ_i can thus be defined as

$$F_{\delta_i}(X) = \int_{-\infty}^{\infty} t \cdot \mathbb{1}(\underbrace{Q_{\tilde{\tau}(y_1)}(Y_i(1))}_{=y_1^*} - \underbrace{Q_{\tilde{\tau}(y_1)}(Y_i(0))}_{=y_0^*} < X) dF_{Y_i(1)}(t)$$

Then, define $Q_\tau(\delta_i) \equiv F_{\delta_i}^{-1}(\tau)$. The idea is that the density of observing δ_i^* equal to the density associated with $Y_i^*(1)$.

$$\square$$

4

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 $Q_\tau(\delta_i)$.

In the previous problem, observe that the term within the $\mathbb{1}(\cdot)$ expression corresponds to the QTE, evaluated at $\tau(y_1)$, which we can estimate by comparing $Q_t(Y_i(1)|T_i = 1)$ with $Q_t(Y_i(0)|T_i = 0)$.

Let J be the number of observations such that $T_i = 1$, so that the number of untreated observations is $N - J$. Without loss of generality, order $Y_i(1)$ such that $i = 1, 2, ..., J$ correspond to the treated group and $Y_{j-1} < Y_j$. Similarly order the untreated groups for $i > J$.[1]

Define the empirical estimator of the quantile function as

$$\hat{Q}_\tau(Y_i(1)) = \inf\left\{Y_i(1)|\tau \le \frac{i}{J}\right\}$$

Similarly, define

$$\hat{Q}_\tau(Y_i(0)) = \inf\left\{Y_i(0)|\tau \le \frac{i - J}{N - J}\right\}$$

One possible estimator of F_δ is as follows:

$$\hat{F}_\delta(X) = \sum_{i=1}^{J} \mathbb{1}\left(Y_i - \hat{Q}_{\frac{i}{J}}(Y_i(0)) < X\right)\frac{i}{J}$$

Then, we define the emprical estimator $\hat{Q}_\delta(\tau) = \inf\left\{X|\tau \le \hat{F}_\delta(X)\right\}$

Iterated Projections

Prove the law of iterated projections

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

Proof.
Define $W_i' = \begin{bmatrix} X_i' & Z_i' \end{bmatrix}$. Recall that the linear projection of W_i onto Y_i requires that, for $\beta = E[W_i W_i']^{-1}E[W_i Y_i]$, the following must hold:

$$E[W_i(Y_i - W_i'\beta)] = 0.$$

This further implies that

$$E[X_i \underbrace{(Y_i - W_i'\beta)}_{\equiv u_i}] = 0$$

---

[1] Here we are using $Y_i(1)$ as shorthand for $\{Y_i, T_i = 1\}$

Thus,

$$
\begin{aligned}
E^*[Y_i|X_i] &= X_i'E[X_iX_i']E[X_iY_i] \\
&= X_i'E[X_iX_i']^{-1}E[X_i(W_i'\beta + u_i)] \\
&= X_i'E[X_iX_i']^{-1}E[X_iW_i'\beta] + X_i'E[X_iX_i']^{-1}E[X_iu_i] \\
&= X_i'E[X_iX_i']^{-1}E[X_iE^*[Y_i|X_i,Z_i]] \\
&= E^*[E^*[Y_i|X_i,Z_i]|X_i]
\end{aligned}
$$

The third equality follows from the linearity of the expectation operator. The second term in the third equality is equal to zero by the observation made above. The fourth follows from the definition $E^*[Y_i|X_i,Z_i] = W_i\beta$. $\qquad\square$

FWL Theorem

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

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

Proof.
Observe that we can write $Y_i = X_i'\beta + Z_i'\gamma + \epsilon_i$ where ϵ_i is the conditional expectation error term from the projection of X_i and Z_i onto Y_i. Recall that by definition, $E[X_i\epsilon_i] = E[Z_i\epsilon_i] = 0$. Thus, we can write:

$$
\begin{aligned}
Y_i &= X_i'\beta + Z_i'\gamma + \epsilon_i \\
X_iY_i &= X_iX_i'\beta + X_iZ_i'\gamma + X_i\epsilon_i \\
E[X_iY_i] &= E[X_iX_i'\beta] + E[X_iZ_i'\gamma] + E[X_i\epsilon_i] \\
X_i'E[X_iX_i']^{-1}E[X_iY_i] &= X_i'E[X_iX_i']^{-1}E[X_iX_i']\beta + X_i'E[X_iX_i']^{-1}E[X_iZ_i']\gamma \\
E^*[Y_i|X_i] &= X_i'\beta + E^*[Z_i|X_i]'\gamma
\end{aligned}
$$

The third line follows from the first order condition of minimization. The fourth line premultiplies everything by $X_i'E[X_iX_i']^{-1}$. The final line simply replaces the expressions with their definition.

Now, subtract the final line from the first line:

$$
\underbrace{Y_i - E^*[Y_i|X_i]}_{\tilde{Y}_i} = X_i'\beta - X_i'\beta + \underbrace{Z_i'\gamma - E^*[Z_i|X_i]'\gamma}_{\tilde{Z}_i'\gamma} + \epsilon_i
$$

$$
\tilde{Z}_i\tilde{Y}_i = \tilde{Z}_i\tilde{Z}_i'\gamma + \tilde{Z}_i\epsilon_i
$$

$$
E[\tilde{Z}_i\tilde{Y}_i] = E[\tilde{Z}_i\tilde{Z}_i']\gamma + E[\tilde{Z}_i\epsilon_i]
$$

Observe that $E[\tilde{Z}_i\epsilon_i] = 0$ since $\tilde{Z}_i$ is a linear combination of Z_i and X_i, both of which have mean zero correlation with ϵ_i. More precisely, write $E^*[Z_i|X_i] = \Pi X_i$ for some matrix of coefficients of Π. So,

$$
E[\tilde{Z}_i\epsilon_i] = E[Z_i\epsilon_i] - E[\Pi X_i\epsilon_i] = 0 - \Pi 0 = 0.
$$

Thus, we have
$$E[\tilde{Z}_i \tilde{Y}_i] = E[\tilde{Z}_i \tilde{Z}_i']\gamma,$$
which after rearranging becomes $\gamma = E[\tilde{Z}_i \tilde{Z}_i']^{-1} E[\tilde{Z}_i \tilde{Y}_i]$, the desired result. □

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

Proof.
Observe

$$E^*[Y_i|X_i, Z_i] = X_i'\beta + Z_i'\gamma$$
$$E^*[E^*[Y_i|X_i, Z_i]|X_i] = X_i'\beta + E^*[Z_i|X_i]'\gamma$$
$$E^*[Y_i|X_i] = X_i'\beta + E^*[Z_i|X_i]'\gamma$$

Now, we have, after solving the above equation for $X_i'\beta$, the desired result:us

$$Y_i - X_i'\beta - Z_i'\gamma = Y_i - E^*[Y_i|X_i] + E^*[Z_i|X_i]'\gamma - Z_i'\gamma$$
$$= (Y_i - E^*[Y_i|X_i]) - (Z_i - E^*[Z_i|X_i])'\gamma$$

□

Weighted Average Derivative Properties

a) The covariance of X_i and Y_i is given by

$$\begin{aligned}
E[(Y_i - \bar{Y})(X_i - \bar{X})] &= E[Y_i(X_i - \bar{X}) - \bar{Y}(X_i - \bar{X})] \\
&= E[Y_i(X_i - \bar{X})] - E[\bar{Y}(X_i - \bar{X})] \\
&= E[Y_i(X_i - \bar{X})] \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} Y_i(X_i - \bar{X}) f_{y|x}(Y_i|X_i) dy f(x) dx \\
&= \int_{-\infty}^{\infty} \mu(X_i)(X_i - \bar{X}) f(x) dx
\end{aligned}$$

Where $f(x)$ is the p.d.f. of X_i. Since μ is continuous and differentiable, the fundamental theorem of calculus implies

$$\mu(X_i) = \kappa + \int_{-\infty}^{X_i} \mu'(t) dt$$

where $\kappa = \lim_{X_i \to \infty} \mu(X_i)$. Substituting this in for $\mu(X_i)$,

$$\begin{aligned}
&= \int_{-\infty}^{\infty} \left(\kappa + \int_{-\infty}^{X_i} \mu'(t) dt \right)(X_i - \bar{X}) f(x) dx \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{X_i} \mu'(t)(X_i - \bar{X}) dt f(x) dx \\
&= \int_{-\infty}^{\infty} \mu'(t) \int_{t}^{\infty} (X_i - \bar{X}) f(x) dx dt
\end{aligned}$$

7

The inner integral is

$$\int_t^\infty (X_i - \bar{X})dx = \int_t^\infty \left(X_i - \int_{-\infty}^\infty uf(u)du\right)f(x)dx$$

$$= \int_t^\infty X_i f(x)dx - \int_t^\infty \int_{-\infty}^\infty uf(u)duf(x)dx$$

$$= \int_t^\infty X_i f(x)dx - \int_t^\infty \left(\int_{-\infty}^t uf(u)du + \int_t^\infty uf(u)du\right)f(x)dx$$

$$= E[X_i|X_i > t]\rho_t - \rho_t\left(E[X_i|X_i < t](1-\rho_t) + E[X_i|X_i > t]\rho_t\right) \quad \rho_t = P(X_i > t)$$

$$= \rho_t(1-\rho_t)(E[X_i|X_i > t] - E[X_i|X_i < t]) \equiv \omega(t)$$

And so the covariance is

$$Cov(X_i, Y_i) = \int_{-\infty}^\infty \mu'(t)\omega(t)dt$$

and

$$Var(X_i) = Cov(X_i, X_i) = \int_{-\infty}^\infty 1 \cdot \omega(t)dt = \int_{-\infty}^\infty \omega(t)dt \qquad \square$$

b) Derive the expression for the weights when $\omega(X_i)$ is normally distributed.

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

$$= \left[\mu + \sigma\frac{\phi(\frac{x-\mu}{\sigma})}{1 - \Phi(\frac{x-\mu}{\sigma})} - \mu + \sigma\frac{\phi(\frac{x-\mu}{\sigma})}{\Phi(\frac{x-\mu}{\sigma})}\right]\left[(\Phi(\frac{x-\mu}{\sigma})(1 - \Phi(\frac{x-\mu}{\sigma}))\right]$$

$$= \sigma\left[\frac{\phi(\frac{x-\mu}{\sigma})}{\Phi(\frac{x-\mu}{\sigma})(1 - \Phi(\frac{x-\mu}{\sigma}))}\right]\left[\Phi(\frac{x-\mu}{\sigma})(1 - \Phi(\frac{x-\mu}{\sigma}))\right]$$

$$= \sigma\phi(\frac{x-\mu}{\sigma})$$

Then $\frac{Cov(Y_i, X_i)}{var(X_i)} = E[\mu'(x)]$