Applied Microeconomics - Pset 2

Jeanne Sorin *

November 10, 2020

Problem 1

Let $\mathcal{H}$ denote the set of all proper distribution functions for a scalar random variable. Suppose that we observe a non-negative random variable Y with distribution $G \in \mathcal{G}$, where $\mathcal{G}$ is the subset of $\mathcal{H}$ such that $P_G(Y \geq 0) = 1$. Assume that Y is determined by

$$Y = \max\{U, 0\}$$

where U is an unobserved random variable with distribution function F. Assume that $F \in \mathcal{F}$, where $\mathcal{F}$ is some subset of $\mathcal{H}$ specified below. Notice that any $F \in \mathcal{F}$implies a distribution for Y called G_F , defined as

$$G_F(y) = P_F[max\{U, 0\} \leq y]$$

Let $\mathcal{F}^*(G)$ denote the sharp identified set for F, that is

$$F^*(G) = \{F \in \mathcal{F} : G_F(y) = G(y) \forall y \in R\}$$

Consider the target parameter $\pi : \mathcal{F} \to R$ defined as $\pi(F) = E_F[U]$. Let $\Pi^*(G)$ denote the sharp identified set for π.

As a reminder, from the supp. notes, *the definition of falsifiability is that there exists a known function $\tau : \mathcal{G} \to \{0, 1\}$ such that whenever $\tau(G) = 1, F^*(G) = \emptyset$ and $\tau(G) = 1$ for at least one $G \in \mathcal{G}$.*
G denotes the distribution of the non-random variable Y in the data. F is the distribution of U.

(a) *Suppose that $\mathcal{F} = \mathcal{H}$. Determine $\Pi^*(G)$ for any $G \in \mathcal{G}$. Is the model falsifiable?*

$$\Pi^*(G) = \{E_F(U) : F \in \mathcal{H}, G_F(y) = G(y)\}$$

This model is not falsifiable because for each $G \neq \emptyset$, we can't find a $\Pi^*(G)$ that is $\emptyset$ (needed for the definition of falsifiability):

- if $G(y) = 0$, then $\Pi^*(G)$ could be any $g <= 0$ (non-empty) and vice-versa.
- if $G(y) > 0$, the $\Pi^*(G)$ count be any $g \in (-\infty, g]$ (non-empty) and vice-versa.

(b) *Suppose that $\mathcal{F} = \{F \in \mathcal{H} : F(-1) = 0, F(2) = 1\}$. Determine $\Pi^*(G)$ for and $G \in \mathcal{G}$. Is the model falsifiable?*
We are given that

$$F(-1) = P(U \leq -1) = 0$$
$$F(2) = P(U \leq 2) = 1$$

*I thank Tom Hierons, Nadia Lucas, Tanya Rajan, Sun Yixin, Camilla Schneier and George Vojta, for comments and discussions.

The corresponding $\Pi^*(G)$ is such that

$$\Pi^*(G) = \{E_F(U) : F \in \mathcal{H}, F(-1) = 0, F(2) = 1, G_F(y) = G(y)\}$$

The model is falsifiable because

- $G(2) < 1 \Rightarrow \Pi^*(G) = \emptyset$

Note that the above assumes that $F(2) = 1, F(x) < 1$ for $x < 2$, and more precisely for $x < 0$. Otherwise, if we also had, for example $F(0) = 1$, then this model would not be falsifiable.

(c) *Suppose that $\mathcal{F} = \{F \in \mathcal{H} : F(0) = 1/2\}$. Determine $\Pi^*(G)$ for and $G \in \mathcal{G}$. Is the model falsifiable?*
We are given that

$$P(U \leq 0) = 1/2$$

The corresponding $\Pi^*(G)$ is such that

$$\Pi^*(G) = \{E_F(U) : \underbrace{F \in \mathcal{H}, F(0) = 1/2}_{F \in \mathcal{F}}, G_F(y) = G(y)\}$$

This model is falsifiable as, according to it, $F(0) = 1/2 \Rightarrow G(0) = 1/2$. So one could can try to falsify the model by showing that $G(0) \neq 1/2$.

(d) *Suppose that $\mathcal{F} = \{F \in \mathcal{H} : E_F[U] = 0\}$. Determine $\Pi^*(G)$ for and $G \in \mathcal{G}$. Is the model falsifiable?*

$$\Pi^*(G) = \{E_F(U) : \underbrace{F \in \mathcal{H}, E_F(U) = 0}_{F \in \mathcal{F}}, G_F(y) = G(y)\}$$

The model is falsifiable in the case where $P(U = 0) < 1$: $E_F(U) = 0 \Rightarrow \{P(U > 0) > 0; P(U < 0) > 0\} \Rightarrow \{P(Y > 0) > 0; P(Y = 0) > 0\} \Rightarrow \{P(E(Y) > 0) = 1; G(0) > 0\}$.
Thus, one could try to falsify the model by testing whether $E_F(Y) = 0$, or $G(0) = 0$. Either equality would imply that the model is wrong.

(e) *Now change the target parameter to $\pi(F) = med_F(U) = inf\{u : F(u) \geq 1/2\}$ (the median of U when it is distributed like F). Suppose that $\mathcal{F} = \mathcal{H}$. Determine $\Pi^*(G)$ for and $G \in \mathcal{G}$. Is the model falsifiable?*

$$\Pi^*(G) = \{med_F(U); F \in \mathcal{H}, G_F(y) = G(y)\}$$

The model is not falsifiable (because too general):

- If $G(0) = 0$ then $\Pi^*(G) = med(Y)$ and vice-versa (non-empty)
- If $G(0) > 0$ then $\Pi^*(G) \in (-\infty, med(Y)]$ and vice-versa (non-empty)

Problem 2

Consider the simple instrumental variables model

$$Y = \alpha X + U$$

where X is scalar. Suppose that the instrument, Z is also scalar, and that there is homoskedasticity with respect to Z in both the reduced form and first stages. (This is the case used to discuss weak instruments in the supplemental notes.) Let F denote the sample first stage F–statistic. Determine the asymptotic bias of F as an estimator of the concentration parameter ("μ^2") under weak instrument asymptotics. Construct an alternative estimator that is asymptotically unbiased.
Hint: No need to reinvent what I did in the supplemental notes. You may use any of the derivations there. Constructing the alternative estimator should be easy.

In the supplement, the sample first stage F-statistic is defined as

$$\hat{F} \equiv (\sqrt{n}\frac{\hat{\pi}}{\hat{\sigma}_\pi})^2$$

Besides, also from the supplement we have the numerator to be equal to

$$\begin{aligned}
\sqrt{n}\hat{\pi} &=\equiv \frac{\sqrt{n}\frac{1}{n}\sum_i Z_i(Z_i\pi_n + V_i)}{\frac{1}{n}\sum_i Z_i^2} \\
&= \sqrt{n}\pi_n + \frac{\sqrt{n}\frac{1}{n}\sum_i Z_iV_i)}{\frac{1}{n}\sum_i Z_i^2} && \text{rearranging} \\
&\to^d \pi + \frac{R_{fs}}{E[Z^2]} && \text{LLN, CMT, Slutsky} \\
&= \frac{\pi E[Z^2] + R_{fs}}{E[Z^2]}
\end{aligned}$$

So

$$\begin{aligned}
\hat{F} &\equiv (\sqrt{n}\hat{\pi}\frac{1}{\hat{\sigma}_\pi})^2 \\
&\to^d (\frac{\pi E[Z^2] + R_{fs}}{E[Z^2]}\frac{1}{\sigma_\pi})^2 && \text{LLN, CLT, CMT, Slutsky}
\end{aligned}$$

Using the definition os $\sigma_\pi^2 = \frac{\sigma_v^2}{E(Z^2)}$[1].

$$\begin{aligned}
\hat{F} &\to^d (\frac{\pi E[Z^2] + R_{fs}}{E[Z^2]}\frac{E[Z^2]^{1/2}}{\sigma_v})^2 \\
&= (\underbrace{\frac{\pi E[Z^2]^{1/2}}{\sigma_V}}_{\mu} + \underbrace{\frac{R_{fs}E[Z^2]^{-1/2}}{\sigma_V}}_{\bar{R}_{fs}})^2
\end{aligned}$$

Thus, the asymptotic bias for $\hat{F}$ as an estimator for μ^2 :

$$\begin{aligned}
|A_{sy}B_{ias}(\hat{F})| &= \lim_{n\to\infty} |E[\mu^2 - \hat{F}]| \\
&= \lim_{n\to\infty} |E[\frac{2\pi R_{fs}}{\sigma_v^2} + \frac{R_{fs}^2}{\sigma_v^2}]|
\end{aligned}$$

[1] A typo in the notes?

A consistent unbiased estimator of $\hat{F}$ would get rid of the asymptotic bias. A consistent estimator of the bias, using LLN and CMT can be

$$\tilde{\hat{F}} = (\frac{\sqrt{n}\hat{\pi} - \frac{\frac{1}{n}\sum_i Z_i V_i}{(\frac{1}{n}\sum_i Z_i^2)}}{\hat{\sigma}_\pi})^2$$

Using the fact that

- $\hat{\pi}$ is a consistent estimator of π
- $\frac{1}{n}\sum_i Z_i \hat{V}_i \to^p E[ZV] = 0$ by WWLN, iid distribution and CMT
- $\frac{1}{n}\sum_i Z_i^2 \to^p E[Z^2]$

We can use Slutsky's Lemma and the CMT to show that

$$\tilde{\hat{F}} \to^d (\frac{\pi + \frac{R_{fs}}{E(Z^2)} - \frac{R_{fs}}{E(Z^2}}{\sigma_p i}) = \mu^2$$

Problem 3

Consider the binary treatment potential ouctomes model $Y = DY(1) + (1-D)Y(0)$. Suppose we have a binary treatment Z that is independent of $(Y(0), Y(1), D(0), D(1))$ where $D = ZD(1) + (1-Z)D(0)$. Maintain the Imbens Angrist (1994) monotonicity condition that $D(1) \geq D(0)$, and assume that $P(D=1|Z=1) - P(D=1|Z=0) > 0$. Let $G \in \{c, a, n\}$ denote the types. Let $F_d(y|g)$ denote the distribution function for $Y(d)$ conditional on $G = g$.

Before answering the questions, let's recall what we observe in the data

$$\begin{aligned} F_{zd}(y) &= P(Y \leq y|D=d, Z=z) \\ G = a &: D(1) = 1, D(0) = 1 \\ G = n &: D(0) = 0, D(0) = 1 \\ G = c &: D(1) = 1, D(0) = 0 \end{aligned}$$

a) *Show that $F_1(y|a)$ is point identified for any y.*

$$\begin{aligned} F_1(y|a) &= P(Y \leq y|G=a, D=1) \\ &= P(Y(1) \leq y|D(1)=1, D(0)=1) \\ &= P(Y \leq y|D(0)=1) && \text{using monotonocity} \\ &= P(Y \leq y|D=1, Z=0) \end{aligned}$$

So $F_1(y|a)$ is point identified for any y, as it is observed in the data. We use the fact that $D = 1, Z = 0$ necessarily corresponds to always-takers given the monotonicity assumption, and the independence assumption allows us to condition on whichever Z we want.

b) *Show that $F_0(y|a)$ is point identified for any y.* Similarly, as $F_0(y\nu)$ is the probability distribution of never takers and is observed in the data.
Using, as above, monotonicity and independence:

$$\begin{aligned} F_1(y|a) &= P(Y \leq y|G=n, D=0) \\ &= P(Y \leq y|D=0, Z=1) \end{aligned}$$

c) *Show that both $F_1(y|c)$ and $F_0(y|c)$ are point identified for any y.*
From the data and the independent assumption we directly have

$$\begin{aligned} P(G=a) &= P(D=1|Z=0) \\ P(G=n) &= P(D=0|Z=1) \end{aligned}$$

Besides, from the monotonicity assumption

$$P(G=c) = 1 - P(G=a) - P(G=n) = 1 - P(D=1|Z=0) - P(D=0|Z=1) > 0$$

We are interested $F_1(y|c)$.
Notice that in the data, we observe $P(Y \leq y|D=1, Z=1)$ which is a combination of always takers and

compliers:

$$P(Y \leq y|D=1,Z=1) = P(Y \leq y|D=1,Z=1,G=c)\frac{p_c}{p_c+p_a} + P(Y \leq y|D=1,Z=1,G=a)\frac{p_a}{p_c+p_a} \quad \text{using LTP}$$

$$\begin{aligned}\underbrace{P(Y \leq y|D=1,Z=1)}_{\text{observed}} &= \underbrace{P(Y \leq y|D=1,Z=1,G=c)}_{F_1(y|c) \text{ under independence of Z}} \quad \underbrace{\frac{p_c}{p_c+p_a}}_{\text{known given monotonicity}} \\ &+ \underbrace{P(Y \leq y|D=1,Z=1,G=a)}_{F_1(y|a) \text{ identified}} \quad \underbrace{\frac{p_a}{p_c+p_a}}_{\text{known given monotonicity}}\end{aligned}$$

Rearranging, we can solve for $F_1(y|c)$

$$F_1(y|c) = [P(Y \leq y|D=1,Z=1) - F_1(y|a)\frac{p_a}{p_c+p_a}]\frac{p_a+p_c}{p_c}$$

So $F_1(y|c)$ is point identified.

Similarly for $F_1(y|c)$ using the fact that $P(Y \leq y|D=0,Z=0)$ is a weighted average of never takers and compliers.

d) *Suppose that both Y(0) and Y(1) are continuously distributed scalar random variables with support over the entirety of the real line. Let F1 and F0 denote the unconditional distribution functions of Y(1) and Y(0). Define the random variables $U(0) = F_0(Y(0))$ and $U(1) = F_1(Y(1))$. Assume that $U(0) = U(1) = U$ with probability 1. Show that $F_0(y)$ and $F_1(y)$ are point identified for any y.*

From the assumptions we have that the distributions of $Y(1)$ and $Y(0)$ are the same:

$$\begin{aligned}F_0(Y(0)) &= P(Y \leq Y(0)|D=0) \\ F_1(Y(1)) &= P(Y \leq Y(1)|D=1) \\ F_0(Y(0)) &= F_1(Y(1))\end{aligned}$$

Together with

$$\begin{aligned}F_1(y) &= P(Y(1) \leq y) = P(Y \leq y|D=1) \\ F_0(y) &= P(Y(0) \leq y) = P(Y \leq y|D=0)\end{aligned}$$

Setting $y = Y(1)$ for $F_1(y)$ and $y = Y(0)$ for $F_0(y)$ we see that the distributions of $Y(1)$ and $Y(0)$ are identical. Thus $F_0(y)$ and $F_1(y)$ are point identified as they are the same for both distribution and can be observed.

$$\begin{aligned}F_1(y) &= P(Y \leq y|D=1) \\ &= P(Y \leq y|G=a,D=1)p_a + P(Y \leq y|G=c,D=1)p_c + P(Y \leq y|G=n,D=1)p_n \\ &= \underbrace{P(Y \leq y|Z=0,D=1)}_{\text{observed}F_1(y|a)}p_a + \underbrace{P(Y \leq y|G=c,D=1)}_{F_1(y|c) \text{ from above}} + \underbrace{P(Y \leq y|Z=1,D=0)}_{\text{using the assumption}F_0(y|n)}p_n\end{aligned}$$

Similarly for $F_0(y)$ where the assumption allows us to replace $F_0(y|a)$ by $F_1(y|a)$.

e) *Maintain the assumptions from the previous part. Let G denote the unconditional distribution function of $Y(1) - Y(0)$. Show that $G(y)$ is point identified for any y.*

$$\begin{aligned}G(y) &= P(Y(1) - Y(0) \leq y) \\ &= P(F_1^{-1}(U) - F_0^{-1}(U) \leq y) \qquad \text{assuming CDF strictly increasing}\end{aligned}$$

From above we have shown that $F_1(y)$ and $F_0(y)$ are point identified, so using $U \sim U[0,1]$ WLOG, we can recover $F_1^{-1}(U)$ and $F_0^{-1}(U)$. We can recover $G(y)$ by integrating the above over the domain of U.

Problem 4

This problem is about the paper "Children and Their Parents' Labor Supply: Evidence from Exogenous Variation in Family Size" by Angrist and Evans (1998, The American Economic Review). The authors define their endogenous variable as a binary indicator for whether a family has exactly two or more than two children. They use two instru- ments: the same-sex instrument and the twin birth instrument, both separately and together.

a) *Discuss the interpretation and credibility of the monotonicity condition when using the same-sex instrument separately.*

The monotonicity condition implies that all parents would be more likely to have a 3rd child if they have 2 children of the same sex rather than 2 children of different sex. While this statement seems to be the case **on average**, whether it applies for all families may be questionable. As a toy example, some parents may be really happy with 2 children of the same sex that get along and may be discouraged to have a third child in order not to disrupt family equilibrium.
Besides, notice that the "evidence" presented in Table 3 are only fractions, so talk to "the average" but do not prove that the monotonicity assumption holds across all families.

b) *Discuss the interpretation and credibility of the monotonicity condition when using the twin birth instrument separately.*

If the twin-birth instrument means that parents with twins as their second birth (and de facto 3rd too), then monotonicity holds trivially as the third child comes mechanically with the 2nd child. This is an example of one-sided full compliance.

c) *Discuss the interpretation and credibility of the monotonicity condition when using the same-sex and twin birth instruments together.*

Moreover, in addition to monotonicity not necessarily holding for the same sex instrument in itself, another problem may be that monotonicity is usually thought as holding for a given instrument.
The monotonicity condition would mean here not only that parents are more likely to have a third child if they have two children of the same sex AND / OR twins as second births. While this may be argued for, it is unclear how to rank these instrument vectors in terms of monotonicity here: whether $(Z_{ss} = 1, Z_t = 0) \geq (Z_{ss} = 0, Z_t = 1)$ or $(Z_{ss} = 1, Z_t = 0) \leq (Z_{ss} = 0, Z_t = 1)$ holds is not obvious.
The monotonicity assumption therefore seems hard to defend in the case of two instruments.

Problem 5

Consider an instrumental variables model with a binary treatment $D \in \{0,1\}$ and binary instrument $Z \in \{0,1\}$. The usual potential outcomes are Y(0) and Y(1) and the usual potential treatment choices are D(0) and D(1). Let $X \in \{1, ..., K\}$ be a vector of covariates that has a discrete distribution with K points of support. Let β_{tsls} denote the population coefficient on D in the following two stage least squares specification:

$$\text{reduced form: Y on Z and } \{\mathbf{1}[X=k]\}_{k=1}^{K}$$
$$\text{first stage: D on Z and } \{\mathbf{1}[X=k]\}_{k=1}^{K}$$

and assume that β_{tsls} exists. Note that, relative to the case discussed in the slides and notes, the first stage in this specification does not contain interactions between Z and X. Maintain the usual exogeneity condition that Z is independent of (Y(0), Y(1), D(0), D(1)) conditional on X.
Assume that $P[D(1) \geq D(0)|X=k] = 1 \forall k = 1, ..., K$.

a) *Explain how the monotonicity condition given here differs from the one discussed in the slides.*

Here the monotonicity condition is conditioning on X. This is different than the general monotonicity condition $D(1) \geq D(0)$, but simply a special case of the monotonicity condition slide 10 (conditional monotonicity condition) with discrete $X = k$. So the monotonicity condition must hold for each discrete k.
This has implications on the first stage: because we do not "condition" on X, but "fix" X, we do not interact the dummies for X with the instrument.

b) *Show that*

$$\beta_{tsls} = E[\frac{cov(D,Z|X)}{E(Cov(D,Z|X))} E[Y(1)-Y(0)|D(1)=1, D(0)=0, X]]$$

where $p(z,x) = P(D=1|Z=z, X=x)$ is the propensity score.

Let's start from the model

$$Y = \alpha + \beta D + \gamma X + \epsilon$$

β can be written in a switching regression framework as it is the expected difference in potential outcomes for compliers, i.e. for those with $D(1)=1, D(0)=0$: $\beta_X = E(Y(1)-Y(0)|D(1)=1, D(0)=0, X)$.

Now lets move to the TSLS.

$$\begin{aligned}\beta_{tsls} &= \frac{\beta_{\text{reduced form}}}{\beta_{\text{first stage}}} \\ &= \frac{cov(\tilde{Y}, \tilde{Z})}{cov(\tilde{D}, \tilde{Z})}\end{aligned}$$

Where, ala FWL

$$\begin{aligned}\tilde{Y} &= Y - E(Y|X) \\ \tilde{D} &= D - E(D|X) \\ \tilde{Z} &= Z - E(Z|X)\end{aligned}$$

Dissecting the denominator

$$\begin{aligned}cov(\tilde{D}, \tilde{Z}) =& cov(D - E(D|X), Z - E(Z|X)) \\ =& E[(D - E(D|X)(Z - E(Z|X))] \\ &= E[E[(D - E(D|X)(Z - E(Z|X))]|X] && \text{using LIE} \\ &= E[E(DZ|X) - E(Z|X)E(D|X)] \\ &= E[cov(D,Z|X)]\end{aligned}$$

Similarly for $cov(\tilde{Y}, \tilde{Z})$. Plugging these into the definition for β_{tsls}:

$$\beta_{tsls} = \frac{E[cov(Y, Z|X)]}{E[cov(D, Z|X)]}$$

Finally, we use the definition of LATE conditional on X:

$$\begin{aligned}
LATE(X) &= E(Y(1) - Y(0)|D(1) = 1, D(0) = 0, X) \\
&= \frac{cov(Y, Z|X)}{cov(D, Z|X)} \\
\leftrightarrow cov(Y, Z|X) &= LATE(X) \cdot cov(D, Z|X)
\end{aligned}$$

Back to β_{tsls}

$$\begin{aligned}
\beta_{tsls} &= \frac{E[cov(Y, Z|X)]}{E[cov(D, Z|X)]} \\
&= \frac{E[\cdot cov(D, Z|X)]}{E[cov(D, Z|X)]} \\
&= E[LATE(X)\frac{cov(D, Z|X)}{E[cov(D, Z|X)]}] && \text{using conditional indep}
\end{aligned}$$

c) *Explain what this result shows and provide some intuition.*

This result shows that in this context, β_{tsls} can be interepreted as a weighted average of conditional LATEs (compliers' average treatment effects conditional on X) Weights (all positive) are the relative propensity scores (propensity score in bin $X = k$ divided by average propensity score). More weights is put on bins that have a higher propensity score (conditional on X), i.e. where the instrument induces more compliers.

Problem 6

The DGP is such that

$$\begin{aligned} Y &= X + U \\ X &= 0.3Z_1 + V \\ (U,V) &\sim \mathcal{N}(0,\Sigma) \\ Cov(U,V)\Sigma &= \begin{bmatrix} 0.25 & 0.20 \\ 0.20 & 0.25 \end{bmatrix} \\ Z_1, ...Z_{20} &\sim \mathcal{N}(0,1) \end{aligned}$$

As it is obvious from the DGP we are in the case where the X is correlated with U through V, so OLS will give us biased results. This is confirmed by the first four rows of the table. Importantly, note that the bias is so large that the true coefficient (1) is not even contained in the 95 % confidence interval.

Unsurprisingly, the TSLS approach using the only Z1 as instrument gives unbiased results, as it maps the DGP. Adding irrelevant instruments lead to biased estimates (TSLS All Zs).

The Jackknife for the TSLS Z1[2] is, as the standard TSLS Z1 approach, unbiased because we are looking at the true DGP. Interestingly, using the Jackknife with irrelevant instruments (Jackknife IV All Zs), with large N, almost always contains the true estimate in its 95% interval and has a bias close to 0. This is in line with theory, as, from the supplementary notes, the jackknife IV removes the finite sample endogeneity issue generated in the second stage, by eliminating the role that X_i had in predicting its own first stage fitted values.

[2] I believe there's a typo in the Jackknife formula in the notes and it should be $\hat{\beta}_{jack} = (\tilde{X}'_{jack,i}X)^{-1}(\tilde{X}'_{jack,i}Y)$

Table 1: Problem 6's beautiful table

Estimator	N	Bias	Median	Std	CI 95%
OLS					
	100	0.581	1.580	0.059	0
	200	0.586	1.590	0.038	0
	400	0.586	1.590	0.031	0
	800	0.589	1.590	0.023	0
TSLS Z1					
	100	-0.002	1.010	0.185	0.920
	200	-0.018	0.981	0.125	0.920
	400	-0.001	1.010	0.095	0.940
	800	-0.008	0.994	0.065	0.960
TSLS All Zs					
	100	-0.368	0.524	0.645	0.390
	200	-0.419	0.528	0.591	0.600
	400	-0.456	0.457	0.545	0.750
	800	-0.473	0.485	0.531	0.840
Jackknife IV Z1					
	100	-0.116	0.934	0.217	0.990
	200	-0.019	1.010	0.127	0.950
	400	-0.021	0.979	0.087	0.960
	800	-0.012	0.995	0.070	0.910
Jackknife IV All Zs					
	100	-0.166	0.898	0.806	0.680
	200	-0.051	0.981	0.198	0.850
	400	-0.028	0.972	0.099	0.920
	800	-0.0002	1.010	0.062	0.920

Problem 7

Table 2: Reproduction of Table 2 from Abadie 2003

	OLS 1	First Stage 2	Second Stage 3	Least Squares Treated 4
Participation	13527.05 (1810.27)		9418.83 (2095.1)	9556.44 (1912.58)
Constant	-23549 (2178.08)	-0.0306 (0.0087)	-23298.74 (2163.92)	-25396.87 (2916.81)
FamilyIncome	976.93 (83.37)	0.0013 (0.0001)	997.19 (83.02)	1011.52 (84.47)
Age	-376.17 (236.98)	-0.0022 (0.001)	-345.95 (236.95)	56.47 (302.77)
Age2	38.7 (7.67)	0.0001 (0)	37.85 (7.68)	30.83 (9.92)
Married	-8369.47 (1829.93)	-0.0005 (0.0079)	-8355.87 (1829.86)	-7477.75 (2434.64)
FamilySize	-785.65 (410.78)	0.0001 (0.0024)	-818.96 (410.09)	-1056.82 (452.75)
Eligibility		0.6883 (0.008)		

Robust Standard Errors in Parenthesis for col 1, 2 and 3. Bootstraped SE in Parenthesis for column 4.

As one can see in Table 2, I match coefficients and standard errors for columns 1 to 3 (column 3's standard errors are really close to Abadie's). Column 4's coefficients and standard errors are slightly different than Abadie's but within one standard errors of his estimates. The discrepancies potentially arise from different ways of estimating the propensity score (used in computing the κ).

As you can check in the code, I find that the AR 95% Confidence Interval for the coefficient on participation is $[5740, 13087]$, which is slightly tighter than the tsls 95% confidence interval one would get using the standard error from column 3 : $[5312, 13525]$[3]. This makes sense since the AR test sacrifices.

Finally, the jackknife IV estimator with bootstrapped standard errors on the participation regressor is $9,418.5$, with bootstrapped standard error $2,194.92$ [4]. This is very close to coefficients and standard errors obtained in column 3. The reason is that we are not dealing with a "multiple instrument" potential problem here (as it was the case in problem 6), so the Jackknife does not make much of a difference.

[3] $CI(95\%) = [coef - 1.95 * se; coef + 1.96 * se]$

[4] The enclosed code computes the coefficients for all regressors, but only the coefficient of interest is presented here for clarity.