E31720 Problem Set 2

Franco Calle
(incidental discussions with Hugo Lopez, Phillip Monagan and Ed Jee)

University of Chicago

November 12, 2021

Problem 1[1]

Consider setting in section 5.2 with single endogenous variable $X_i = D_i$ and no covariates. Suppose $F_i = Z_i$ is a d_z dimensional vector, and homoskedasticity:

$$\mathbb{E}[U_i^2|Z_i] = \mathbb{E}[U_i^2] = \sigma_U^2$$
$$\mathbb{E}[V_i^2|Z_i] = \mathbb{E}[V_i^2] = \sigma_V^2$$
$$\mathbb{E}[U_iV_i|Z_i] = \mathbb{E}[U_iV_i] = \sigma_{UV}$$

Part A

Suppose:

$$D_i = Z_i'\gamma_n + V_i \tag{1}$$

Where $\gamma_n = \gamma/\sqrt{n}$. Find the limiting distribution of the two stage least squares estimator, $\hat{B}_{tsls}$.

Proof. First define the first stage estimation of Z on D: $\hat{\Gamma} = \left(\frac{1}{n}\sum_i Z_i'Z_i\right)^{-1}\left(\frac{1}{n}\sum_i Z_i'D_i\right)$. Now, when plugging this element into the second stage estimation we get:

$$\begin{aligned}
\hat{B}_{TSLS} &= \left(\frac{1}{n}\sum_i \hat{\Gamma}'Z_iD_i\right)^{-1}\left(\frac{1}{n}\sum_i \hat{\Gamma}'Z_iY_i\right) \\
&= \left(\frac{1}{n}\sum_i \hat{\Gamma}'Z_iD_i\right)^{-1}\left(\frac{1}{n}\sum_i \hat{\Gamma}'Z_i(\beta D_i + U_i)\right) \\
&= \beta\left(\frac{1}{n}\sum_i \hat{\Gamma}'Z_iD_i\right)^{-1}\left(\frac{1}{n}\sum_i \hat{\Gamma}'Z_iD_i\right) + \left(\frac{1}{n}\sum_i \hat{\Gamma}'Z_i(Z_i'\gamma_n + V_i)\right)^{-1}\left(\frac{1}{n}\sum_i \hat{\Gamma}'Z_iU_i\right) \\
&= \beta + \left(\frac{1}{n}\sum_i \hat{\Gamma}'Z_iZ_i'\gamma_n + \frac{1}{n}\sum_i \hat{\Gamma}'Z_iV_i\right)^{-1}\left(\frac{1}{n}\sum_i \hat{\Gamma}Z_iU_i\right) \\
\implies \hat{B}_{TSLS} - \beta &= \left(n\hat{\Gamma}'\left(\frac{1}{n}\sum_i Z_iZ_i'\right)\gamma_n + n\hat{\Gamma}'\left(\frac{1}{n}\sum_i Z_iV_i\right)\right)^{-1}\left(n\hat{\Gamma}'\left(\frac{1}{n}\sum_i Z_iU_i\right)\right)
\end{aligned}$$

Where the first equation is just the esimator of TSLS, then we replace Y in the equation, do some algebra and cancel out the component multiplying β. Then we pass β to LHS and multiply denominator and numerator by n. As can be seen the last line contains an additional term in the denominator:

[1] The code for this question can be found here: Question 1

$$n\hat{\Gamma}'\left(\frac{1}{n}\sum_i Z_iZ_i'\right)\gamma_n = \sqrt{n}\hat{\Gamma}'\left(\frac{1}{n}\sum_i Z_iZ_i'\right)\gamma$$

Note that by WLLN $\left(\frac{1}{n}\sum_i Z_i'D_i\right) \to^p \mathbb{E}[Z'D]$ and $\left(\frac{1}{n}\sum_i Z_i'Z_i\right) \to \mathbb{E}[Z'Z]$. Now, note that from equation 1 $\mathbb{E}[Z'Z]^{-1}\mathbb{E}[Z_iD_i] = \gamma/\sqrt{n}$, then we have that $\sqrt{n}\hat{\Gamma} \to^p \gamma$. We can use this result and apply CMT and slutsky in the previous equation to get the following conclusion:

$$\sqrt{n}\hat{\Gamma}'\left(\frac{1}{n}\sum_i Z_iZ_i'\right)\gamma \to^p \gamma'\,\mathbb{E}[Z_iZ_i']\gamma$$

Now

$$\hat{B}_{TSLS} - \beta = \left((\sqrt{n}\hat{\Gamma}')\left(\frac{1}{n}\sum_i Z_iZ_i'\right)\gamma + (\sqrt{n}\hat{\Gamma}')\sqrt{n}\left(\frac{1}{n}\sum_i Z_iV_i\right)\right)^{-1}\left((\sqrt{n}\hat{\Gamma}')\sqrt{n}\left(\frac{1}{n}\sum_i Z_iU_i\right)\right)$$

Then assuming that (U_i, V_i) comes from a bivariate standard normal with identity covariance matrix, and applying CLT the limiting distribution we get is:

$$\begin{aligned}\hat{B}_{TSLS} - \beta &\to^d \left(\gamma'\,\mathbb{E}[Z_iZ_i']\gamma + \gamma'\mathbb{V}(Z_iV_i)^{1/2}R_{fs}\right)^{-1}(\gamma'\mathbb{V}(Z_iU_i)^{1/2}R_{rf}) \\ &= \left(\gamma'\,\mathbb{E}[Z_iZ_i']\gamma + \gamma'\,\mathbb{E}[Z_iZ_i']^{1/2}\sigma_V R_{fs}\right)^{-1}(\gamma'\,\mathbb{E}[Z_iZ_i]^{1/2}\sigma_U R_{rf})\end{aligned}$$

For standard normals R_{rf} and R_{fs}. Where:

$$\begin{bmatrix}\sqrt{n}\frac{1}{n}\sum_i Z_iU_i \\ \sqrt{n}\frac{1}{n}\sum_i Z_iV_i\end{bmatrix} \to \begin{bmatrix}R_{rf}[\mathbb{E}[Z_iZ_i']\sigma_U^2]^{1/2} \\ R_{fs}[\mathbb{E}[Z_iZ_i']\sigma_V^2]^{1/2}\end{bmatrix}$$

Note That $\mathbb{V}(Z_iV_i) = \mathbb{E}[Z_iV_i^2Z_i] = \mathbb{E}[Z_i\,\mathbb{E}[V_i^2|Z_i]Z_i] = \mathbb{E}[Z_iZ_i]\sigma_V^2$, and the same for $\mathbb{V}(Z_iU_i) = \mathbb{E}[Z_iZ_i]\sigma_U^2$.

□

Part B

Consider the OLS estimator of Y_i on D_i with no constant. Show that the asymptoptic bias of this estimator is the same under 5.24 as it is under standard asymptotics with $\gamma_n = \gamma = 0$.

Proof. Recall the OLS distribution is:

$$
\begin{aligned}
\hat{B}_{OLS} &= \left(\frac{1}{n}\sum_i D_i^2\right)^{-1}\left(\frac{1}{n}\sum_i D_i Y_i\right) \\
&= \left(\frac{1}{n}\sum_i D_i^2\right)^{-1}\left(\frac{1}{n}\sum_i D_i\left(D_i\beta + U_i\right)\right) \\
&= \beta + \left(\frac{1}{n}\sum_i D_i^2\right)^{-1}\left(\frac{1}{n}\sum_i D_i U_i\right)
\end{aligned}
$$

Now we move β to LHS and replace D_i in the equation. Assuming $\gamma = \gamma/\sqrt{n}$.

$$
\begin{aligned}
\hat{B}_{OLS} - \beta &= \left(\frac{1}{\sqrt{n}}\sum_i (Z_i'\gamma_n + V_i)^2\right)^{-1}\left(\frac{1}{\sqrt{n}}\sum_i (Z_i'\gamma_n + V_i)U_i\right) \\
&= \left((\sqrt{n}\gamma_n')(\frac{1}{n}\sum_i Z_i'Z_i)(\sqrt{n}\gamma_n) + 2(\frac{1}{\sqrt{n}}\sum_i Z_iV_i)\gamma_n + n\left(\frac{1}{n}\sum_i V_i^2\right)\right)^{-1} \times \\
&\left(n(\frac{1}{n}\sum Z_iU_i)\gamma_n + n(\frac{1}{n}\sum_i V_iU_i)\right) \\
&\to^d \left(\gamma'\,\mathbb{E}[Z_iZ_i']\gamma + 2\gamma'\mathbb{V}(Z_iV_i)^{1/2}R_{fs} + n\mathbb{V}(V_i)\right)^{-1}\left(\gamma'\mathbb{V}(Z_iU_i)^{1/2}R_{rf} + n\mathbb{C}(V,U_i)\right)
\end{aligned}
$$

Now, assuming $\gamma = 0$ we get that distribution is degenerate in β plus a bias term:

$$
\hat{B}_{OLS} \to^p \beta + \underbrace{\mathbb{V}(V_i)^{-1}\mathbb{C}(V,U_i)}_{\text{abias}_{\text{ols}}}
$$

Under standard asymptotics, $\gamma_n = \gamma = 0$, so the result is straightforward, we get:

$$
\begin{aligned}
\hat{B}_{OLS} - \beta &= \left(\frac{1}{\sqrt{n}}\sum_i (Z_i'\gamma + V_i)^2\right)^{-1}\left(\frac{1}{\sqrt{n}}\sum_i (Z_i'\gamma + V_i)U_i\right) \\
&= \left(\frac{1}{\sqrt{n}}\sum_i (V_i)^2\right)^{-1}\left(\frac{1}{\sqrt{n}}\sum_i V_iU_i\right) \\
\implies \hat{B}_{OLS} &= \beta + \underbrace{\mathbb{V}(V_i)^{-1}\mathbb{C}(V,U_i)}_{\text{abias}_{\text{ols}}}
\end{aligned}
$$

Then under both asymptotics the asymptotic bias is the same.

□

Part C

Generalize the definition of the concentration parameters to:

$$\mu^2 = \frac{\gamma' \,\mathbb{E}[Z_i Z_i]\gamma}{\sigma_V^2} \text{ where } \mu = \frac{\mathbb{E}[Z_i Z_i']^{1/2}\gamma}{\sigma_V}$$

Show, asyvar of TSLS depends only on μ and d_z.

Proof. Now we can rewrite the asymptotic distribution in A:

$$\begin{aligned}\hat{B}_{TSLS} - \beta \to^d & \left(\mu^2 \cdot \sigma_V^2 + \mu' \cdot \sigma_V^2 R_{fs}\right)^{-1} (\mu' \cdot \sigma_V \cdot \sigma_U R_{rf}) \\ &= \left(\mu^2 \cdot \sigma_V + \mu' \cdot \sigma_V R_{fs}\right)^{-1} (\mu' \cdot \sigma_U R_{rf}) \\ &= \frac{\sigma_U}{\sigma_V}\left(\mu^2 + \mu' \cdot R_{fs}\right)^{-1} (\mu' \cdot R_{rf})\end{aligned}$$

Now we can take conditional expectations:

$$\begin{aligned}\mathbb{E}\left[\frac{\sigma_u}{\sigma_v}\frac{\mu R_{rf}}{\mu^2 + \mu R_{fs}} | R_{fs} = r\right] &= \frac{\sigma_u}{\sigma_v}\frac{\mu' \,\mathbb{E}[R_{rf}|R_{fs} = r]}{\mu^2 + \mu' r} \\ &= \rho\frac{\sigma_u}{\sigma_v}\frac{\mu' r}{\mu^2 + \mu' r}\end{aligned}$$

We can divide this expression with the OLS bias derived in B $\rho\frac{\sigma_u}{\sigma_v}$ and obtain the following:

$$\begin{aligned}\frac{\mu r}{\mu^2 + \mu' r} &= \frac{\mu' r}{\mu^2}\frac{1}{\left(1 + \frac{\mu' r}{\mu' \mu}\right)} \\ &\simeq \frac{\mu' r}{\mu^2}\left(1 - d_z\frac{\mu' r}{\mu^2}\right) \\ &= \frac{\mu' r}{\mu^2} - d_z\left(\frac{\mu' r}{\mu^2}\right)^2\end{aligned}$$

Where the first term just rearranges both numerator and denominator, the second line takes taylor expansion $1/(1 + f(x)) \sim 1 - h' Df(x)$ for the multivariate case, since the gradient has the same value in all entries, when multiplied by h we get d_z in our derivation. And the last line expands the product in parenthesis.

Now, we can do something analogous but with the unconditional expectation:

$$
\begin{aligned}
\mathbb{E}\left[\frac{\sigma_u}{\sigma_v}\frac{\mu' R_{rf}}{\mu^2+\mu' R_{fs}}\right] &\simeq \frac{\rho\sigma_u}{\sigma_v}\left(\frac{\mu E[R_{fs}]}{\mu^2} - d_z\,\mathbb{E}\left[\left(\frac{\mu' R_{fs}}{\mu^2}\right)^2\right]\right) \\
&= -\rho\frac{\sigma_u}{\sigma_v} d_z\,\mathbb{E}\left[\frac{\mu' R_{fs} R'_{fs}\mu}{\mu^4}\right] \\
&= -\rho\frac{\sigma_u}{\sigma_v} d_z \frac{\mu'\,\mathbb{E}[R_{fs}R'_{fs}]\mu}{\mu^4} \\
&= -\rho\frac{\sigma_u}{\sigma_v}\frac{d_z}{\mu^2} \\
\implies |\frac{abias_{iv}}{abias_{ols}}| &= \frac{d_z}{\mu^2}
\end{aligned}
$$

The first line used the approximation we derived previously with the taylor expansions, the second line eliminates the first term in parenthesis since E[R] = 0. The third line introduces the expectation within the numerator. The fourth uses the fact that variance of R_{fs} is identity matrix, then numerator resumes to: $\mu^2 d_z$. Finally we can conclude that the ratio of the biases is $\frac{d_z}{\mu^2}$ as shown in the last line.

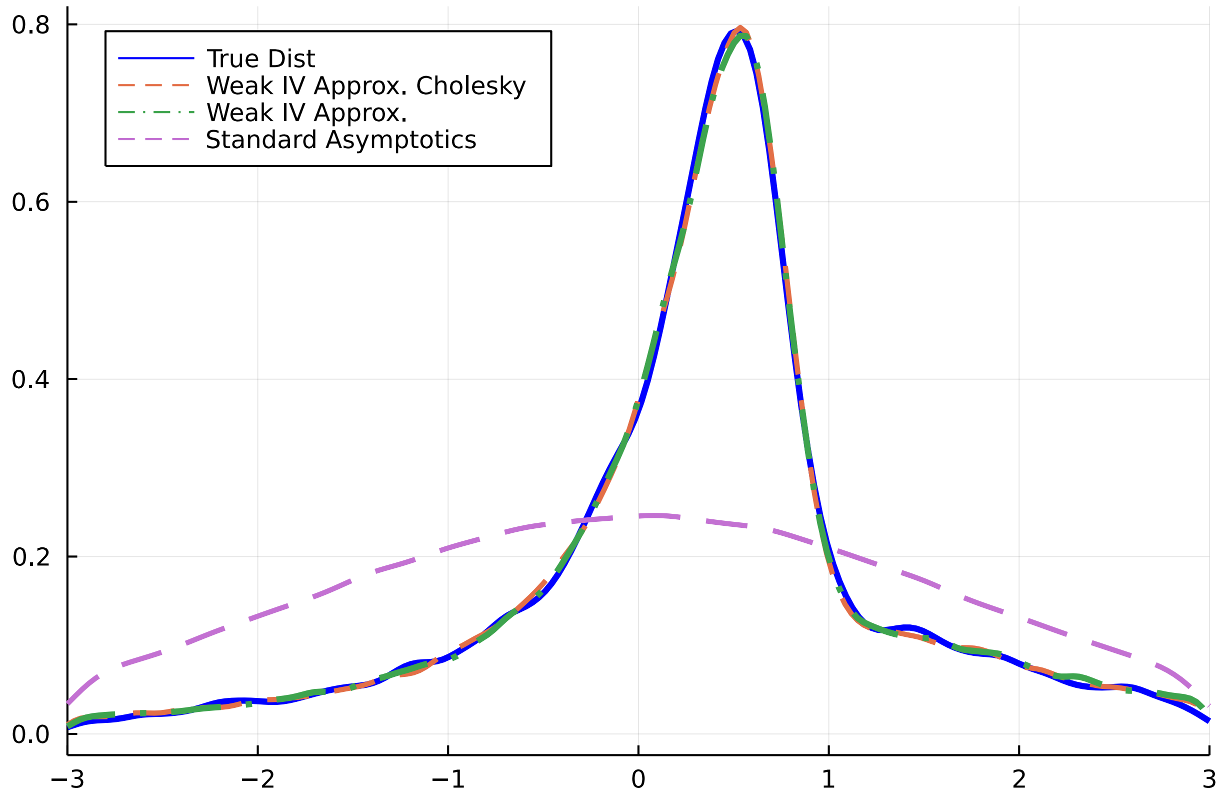
□

Part D

Use previous results in part (c) to simulate relative asymptotic bias of TSLS to OLS for many choices of μ. Provide numerical evidence that relative asymptotic bias depends only on d_z and μ^2 but not μ itself. Does relative bias increase or decrease with μ^2.

Proof. This proof is numerical, it is a plot that shows the weak iv asymptotic approximation using different ways to decompose μ. Cholesky decomposition vs non triangular square root of $E[Z_i Z_i']$. Since Cholesky$(E_{ZZ'})\gamma \neq \sqrt{(E_{ZZ'})}\gamma$, then $\mu_{\text{Cholesky}} \neq \mu$. Graphically we can see in figure 1 that there is no difference between the two ways to approximate μ.

Figure 1: Asymptotic approximation for weak IV



□

Part E

Use simulation to determine the value of μ^2 that corresponds to relative bias of $\alpha = .05, .10, .20, .30$ for $d_z = 3, 4, ..., 30$. This is just computing $\mu^2 = \frac{d_z}{\alpha}$ Here I show some estimates:

Table 1: Concentration term and critical values based on non-centered Chi squared $\chi^2(\frac{\mu^2}{d_z})$

	Concentration Term: μ^2				Critical Values $G^{-1}_{\mu^2,d_x}(1-p)$			
d_z	0.05	0.10	0.20	0.30	0.05	0.10	0.20	0.30
3	6.000	3.000	1.500	1.000	11.171	8.192	6.507	5.890
4	8.000	4.000	2.000	1.333	11.888	8.655	6.850	6.196
5	10.000	5.000	2.500	1.667	12.534	9.073	7.160	6.470
6	12.000	6.000	3.000	2.000	13.126	9.459	7.445	6.723
7	14.000	7.000	3.500	2.333	13.677	9.818	7.710	6.958
8	16.000	8.000	4.000	2.667	14.194	10.155	7.960	7.180
9	18.000	9.000	4.500	3.000	14.683	10.475	8.197	7.389
10	20.000	10.000	5.000	3.333	15.147	10.780	8.423	7.589
11	22.000	11.000	5.500	3.667	15.591	11.071	8.639	7.781
12	24.000	12.000	6.000	4.000	16.016	11.350	8.847	7.965
13	26.000	13.000	6.500	4.333	16.425	11.620	9.047	8.142
14	28.000	14.000	7.000	4.667	16.820	11.879	9.240	8.313
15	30.000	15.000	7.500	5.000	17.202	12.131	9.427	8.478
16	32.000	16.000	8.000	5.333	17.571	12.375	9.608	8.639
17	34.000	17.000	8.500	5.667	17.931	12.611	9.784	8.795
18	36.000	18.000	9.000	6.000	18.280	12.842	9.956	8.948
19	38.000	19.000	9.500	6.333	18.620	13.066	10.123	9.096
20	40.000	20.000	10.000	6.667	18.951	13.286	10.286	9.240
21	42.000	21.000	10.500	7.000	19.275	13.500	10.445	9.382
22	44.000	22.000	11.000	7.333	19.592	13.709	10.601	9.520
23	46.000	23.000	11.500	7.667	19.901	13.914	10.754	9.655
24	48.000	24.000	12.000	8.000	20.205	14.114	10.904	9.788
25	50.000	25.000	12.500	8.333	20.502	14.311	11.050	9.918
26	52.000	26.000	13.000	8.667	20.793	14.504	11.194	10.046
27	54.000	27.000	13.500	9.000	21.080	14.693	11.335	10.171
28	56.000	28.000	14.000	9.333	21.361	14.879	11.474	10.294
29	58.000	29.000	14.500	9.667	21.637	15.062	11.611	10.416
30	60.000	30.000	15.000	10.000	21.909	15.242	11.745	10.535

Part F

Let $\hat{F}_{\text{stat}}$ be the first stage F-statistic for the null hypothesis that the coefficient vector on Z_i is 0_dz. Show that the limiting distribution of $\hat{F_{stat}}$ under weak instrument asymptotics follows a non–central chi–squared distribution scaled by a constant. What are its degrees of freedom and non–centrality parameter?

Proof. Recall the estimator of γ_n for the first stage.

$$\begin{aligned}
\sqrt{n}\hat{\Gamma} &= \sqrt{n}\left(\frac{1}{n}\sum_i Z_i'Z_i\right)^{-1}\left(\frac{1}{n}\sum_i Z_i'D_i\right) \\
&= \sqrt{n}\gamma_n + \left(\frac{1}{n}\sum_i Z_i'Z_i\right)^{-1}\left(\frac{1}{n}\sum_i Z_i'V_i\right) \\
&\to^d \gamma + \mathbb{E}[Z_i'Z_i]^{-1/2}\sigma_v R_{fs}
\end{aligned}$$

And the homoskedastic standard error:

$$\hat{\Omega} = \left(\frac{1}{n}\sum_i Z_iZ_i'\right)^{-1}\left(\frac{1}{n-1}\sum_i (D_i - Z_i\hat{C})^2\right) \to^p \mathbb{E}[Z_iZ_i']^{-1}\sigma_V^2$$

Combining these two things we get the F statistic for our model:

$$\begin{aligned}
T_n = n\hat{\Gamma}\hat{\Omega}^{-1}\hat{\Gamma}' &\to^d \left(\gamma + \mathbb{E}[Z_i'Z_i]^{-1/2}\sigma_v R_{fs}\right)\mathbb{E}[Z_iZ_i']\sigma_V^{-2}\left(\gamma + \mathbb{E}[Z_i'Z_i]^{-1/2}\sigma_v R_{fs}\right)' \\
&= \left(\frac{\gamma\,\mathbb{E}[ZZ']^{1/2}}{\sigma_V} + R_{fs}\right)\left(\frac{\gamma\,\mathbb{E}[ZZ']^{1/2}}{\sigma_V} + R_{fs}\right)' \\
&= (\mu + R_{fs})(\mu + R_{fs})'
\end{aligned}$$

Now note that the distribution is a multiplication of two non centered multivariate normals, or in other words a non centered Chi squared with centrality value μ and d_z degrees of freedom.

□

Part G

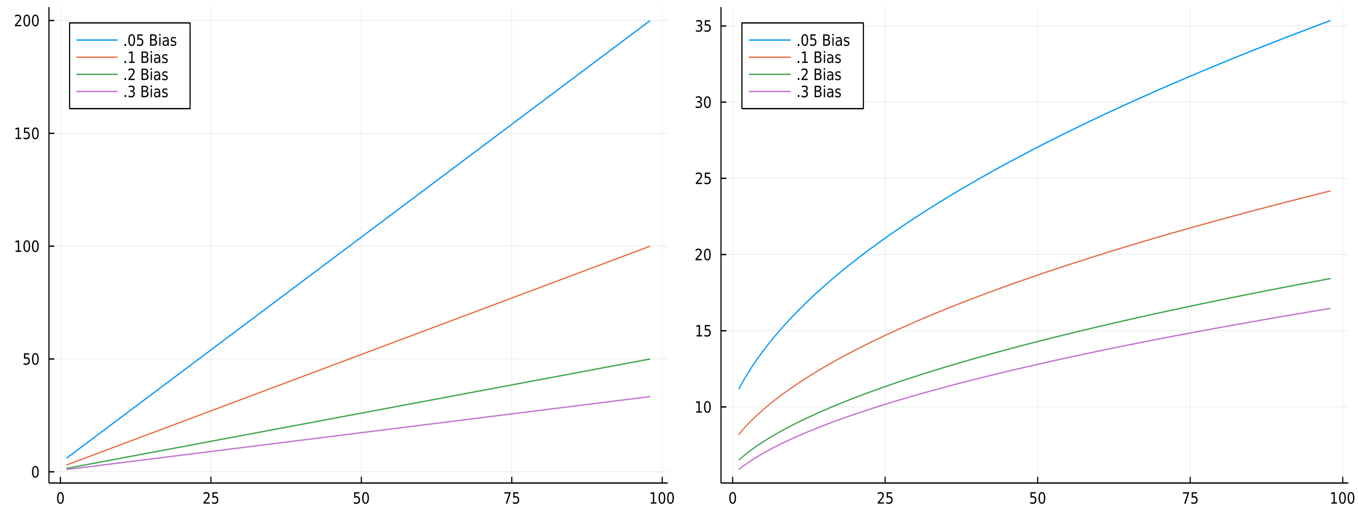
Define weak instruments as being when the relative absolute bias of TSLS to OLS is larger than α. Consider a test that rejects the null hypothesis of weak instruments when the first stage F–statistic is larger than a certain critical value. Use simulation to determine the appropriate level 5% critical values of this test for $\alpha = .05, .10, .20, .30$, and $d_z = 3, 4 \ldots, 30$. Develop a nice plot of the results. Check your results against Table 1 in Stock and Yogo (2005).

Solution

Here I use the values obtained in Part E and pass and divide each of them by their respective d_z. Then, I construct non centered χ^2 distribution with eigenvalue $\lambda = \frac{\mu^2}{d_z}$ and d_z degrees of freedom as in Stock and Yogo (2005) which would represent the asymptotic distribution of the first stage F stat. Then, for each case I get the quantile that represents the $c = 5\%$ critical values for each of the cases. In other words:

$$\begin{aligned}
Pr\left[\frac{1}{d_z}\chi^2_{d_z}(\mu^2 \cdot d_z) \geq x\right] &= 1 - p \\
G_{\mu^2, d_z}(x) &= 1 - p \\
\implies c_p &= G^{-1}_{\mu^2, d_z}(1 - p)
\end{aligned}$$

Where $G_{\mu^2,d_z}(x) = F_{\frac{1}{d_z}\chi^2_{d_z}(\mu^2 \cdot d_z)}(x)$. So c_p which in this exercise is $c_{0.05}$ is our critical value. My results look slightly similar to Table 1 in their paper and the plots they report. However, even when I get similar levels of μ^2 for each of the bias levels suggested and critical values that are concave over the number of degrees of freedom, I don't get that they converge to a constant as they show in figure 1.



(a) Eigenvalue μ^2 (d_x=1)

(b) Critical Value

Figure 2: Critical Values based on Size of TSLS Relative to OLS against d_z

Problem 2 [2]

It is often claimed that weak instruments cause the TSLS estimator to be biased. As usual, we should say "centered" not biased, since the mean of the TSLS estimator might not exist. towards the OLS estimator. We saw that this was true under homoskedasticity. Does it remain true under heteroskedasticity? If so, prove it. If not, provide a counterexample and verify your counterexample in a simulation.

Proof. I'll just prove this with the univariate version. Recall from the notes that:

$$\hat{B}_{iv} - \beta \rightarrow^d \underbrace{\frac{\mathbb{V}[Z_i U_i]^{1/2}}{\mathbb{V}[Z_i V_i]^{1/2}} \rho}_{\text{constant shift}} + \underbrace{\frac{\mathbb{V}[Z_i U_i]^{1/2} \tilde{R}_{rf}}{\mathbb{V}[Z_i V_i]^{1/2} R_{fs}}}_{\text{scaled Cauchy distribution}}$$

Where the constant shift has the form:

$$\frac{\mathbb{V}[Z_i U_i]^{1/2}}{\mathbb{V}[Z_i V_i]^{1/2}} \rho = \frac{\mathbb{E}[Z_i^2 U_i V_i]}{\mathbb{V}(Z_i V_i)} = \frac{\mathbb{E}[Z_i^2 \, \mathbb{E}[U_i V_i | Z]]}{\mathbb{E}[Z_i^2 \, \mathbb{E}[V_i^2 | Z]]}$$

Recall that the problem asks us to assume heteroskedasticity, therefore $E[V_i^2|Z] = \sigma_{v,i}$, $E[U_i^2|Z] = \sigma_{v,i}$ and $E[U_i^2 V_i^2 |Z] = \sigma_{vu,i}$.

Now, for the case of OLS, recall asymptotics of OLS under heteroskedasticity:

$$\begin{aligned}
\beta_{\text{OLS}} &= \frac{\mathbb{E}[Y_i D_i]}{\mathbb{E}[D_i^2]} \\
&= \beta + \underbrace{\frac{\mathbb{E}[U_i D_i]}{\mathbb{E}[D_i^2]}}_{\text{abias}_{\text{ols}}} \\
&= \beta + \underbrace{\frac{\mathbb{E}[U_i V_i]}{\mathbb{E}[V_i^2]}}_{\text{under } \gamma=0} \\
&= \beta + \underbrace{\frac{\mathbb{E}[\mathbb{E}[U_i V_i | Z]]}{\mathbb{E}[\mathbb{E}[V_i^2 | Z]]}}_{\text{LIE over Z}}
\end{aligned}$$

Note that the constant shift in the IV case is different than the OLS asymptotic bias ($\text{abias}_{\text{ols}}$) when assuming heteroskedasticity. Now we can not cancel out $E[Z_i^2]$ from the equation. So now depending on the interaction between Z_i and σ_i for each i and the probability density, those ratios will be different. A case in which both would be equal for instance is when Z is a constant so it cancels out. But in general it need not be.

To illustrate this, in the figure below I estimate the same model presented in class for the univariate version. I assume $\beta = 0$, $\gamma = 0.2$ and a fixed Z randomly generated from a non centered chi-sq

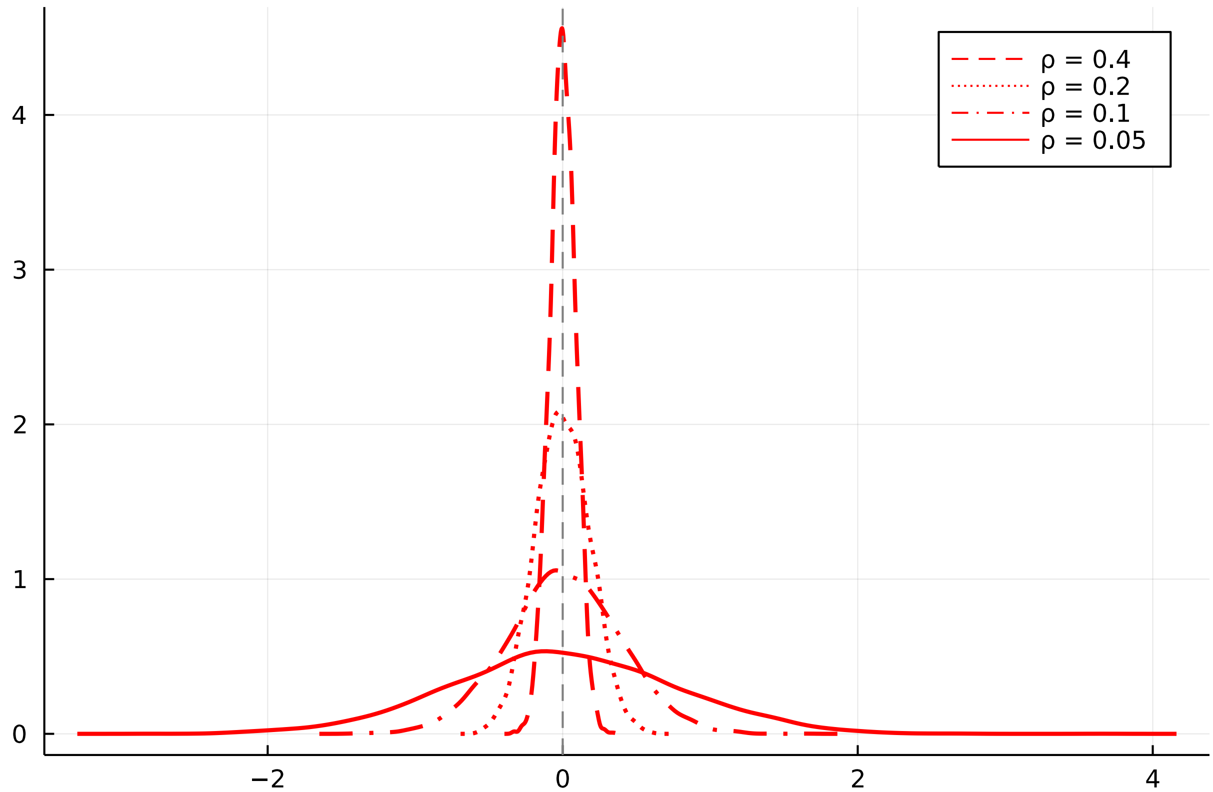
[2] The code for this question can be found here: Question 2

distribution, but varying ρ from strong covariance to weak covariance as suggested by Olea and Pflueger (2013).

What I've found after simulating the experiment 5000 times is that the normalized (by true ρ) distance of the biases under the context of heteroskedasticity is very variant, how much variation there is depends on (1) how strongly correlated are the error terms, and for a given value of ρ just depends on the draws of Z, which are random. In the case $\rho \to 1$ I have the distance of the biases becomes a degenerate distribution centered in 0 (meaning both IV and OLS biases are the same).

$$\delta = \left(\frac{\mathbb{E}[Z_i^2\,\mathbb{E}[U_iV_i|Z]]}{\mathbb{E}[Z_i^2\,\mathbb{E}[V_i^2|Z]]} - \frac{\mathbb{E}[\mathbb{E}[U_iV_i|Z]]}{\mathbb{E}[\mathbb{E}[V_i^2|Z]]} \right) / \rho$$

Figure 3: Bias distance under heteroskedasticity OLS vs IV. 5000 simulations



□

Problem 3

Suppose that $d_x = 1$. Let $\hat{a}(\beta_0)$ denote the realized Anderson-Rubin test statistic for the null hypothesis that $\beta = \beta_0$, with a homoskedastic estimate of the asymptotic variance matrix (for simplicity). Let $\hat{f}$ denote the realized homoskedastic F-statistic for testing the null hypothesis that all of the coefficients in the first stage are 0. Show that the difference between $\hat{a}(\beta_0)$ and $\hat{f}$ tends to 0 as $|\beta_0| \to \infty$. Explain why this means that an Anderson-Rubin confidence region is infinite if and only if the first-stage F-statistic is such that it would not reject the null hypothesis that all coefficients are zero.

Proof. Consider the second stage model: $Y = X\beta_0 + u$ and the first stage $X = F\gamma + v$. The Anderson Rubin test that $H_0 : \gamma = 0$ can be stated as follows.

$$\begin{aligned}
\alpha(\beta_0) &= \frac{(Y - X\beta_0)' P_F (Y - X\beta_0)/K}{(Y - X\beta_0)' M_F (Y - X\beta_0)/(N-K)} \\
&= \frac{(Y\frac{1}{\beta_0} - X)' P_F (Y\frac{1}{\beta_0} - X)/K}{(Y\frac{1}{\beta_0} - X)' M_F (Y\frac{1}{\beta_0} - X)(N-K)}
\end{aligned}$$

And when β_0 goes to infinity in absolute value we have the following:

$$\begin{aligned}
\implies \lim_{\beta_0 \to \infty} \alpha(\beta_0) &= \frac{X' P_F X}{X' M_F X} \frac{N-K}{K} \\
&= \frac{(F\hat{\gamma} + \hat{v})' P_F (F\hat{\gamma} + v)}{(F\hat{\gamma} + \hat{v})'\hat{v}} \frac{N-K}{K} \\
&= \frac{\hat{\gamma}' F' P_F F \hat{\gamma} + 2\hat{v}' P_F F\hat{\gamma} + \hat{v}' P_F \hat{v}}{\hat{\gamma}' F' \hat{v} + \hat{v}'\hat{v}} \frac{N-K}{K} \\
&= \frac{\hat{\gamma}' F' F \hat{\gamma}}{\hat{v}'\hat{v}} \frac{N-K}{K}
\end{aligned}$$

Where $P_F = F(F'F)^{-1}F'$ and $M_Z = I - P_F$. Also, $\hat{\gamma}$ is the estimator γ and $\hat{v}$ is the residual. The first line takes the limit when β_0 goes to infinity, the second line replaces the equation for X, note that $\hat{\gamma}$ and $\hat{v}$ are the sample estimation and residual respectively. The third line extends the expression. The fourth eliminates the two last components in the numerator since $\sum F_i \hat{v}_i = 0$ by construction. The final approximates the expression for a value of n large enough.

And the F stat for the null that $\gamma = 0$ is:

$$\begin{aligned}
\sqrt{n}(\hat{\gamma} - \gamma) &\to^d N(0, \mathbb{E}[FF']^{-1}\sigma_v^2) \\
T_n(\gamma_0) &= \frac{N-K}{K} \frac{(\hat{\gamma} - \gamma_0)'(F'F)(\hat{\gamma} - \gamma_0)}{\hat{v}'\hat{v}} \\
T_n(0) &= \frac{N-K}{K} \frac{\hat{\gamma}' F' F \hat{\gamma}}{\hat{v}'\hat{v}}
\end{aligned}$$

Note that both expressions $\lim_{\beta_0 \to \infty}$ and $T_n(0)$ are equal. Then $\lim_{\beta_0 \to \infty} \alpha(\beta_0) - T_n(0) = 0$. Which is the desired result.

Citing Kleibergen (2007) This implies that the level α AR confidence set has infinite range if and only if the F-test cannot reject $\gamma = 0$. Therefore, β is totally unidentified. Thus, infinite-length confidence sets arise exactly in those cases where the data do not allow us to conclude that β is identified at all.

$\square$

References: Kleibergen, F. (2007). Generalizing weak instrument robust IV statistics towards multiple parameters, unrestricted covariance matrices and identification statistics. Journal of Econometrics, 139(1), 181-216.

Problem 4 [3]

Part A

Replicate Tables 3, 4, and 5 of the paper. Note any discrepancies with the published results and discuss their likely causes. I find no discrepancies between my estimates and the tables reported in the paper.

Table 2: Assignment to Eskom Project: First Stage OLS Estimates

	Eksom Project = [1 to 0]			
	(1)	(2)	(3)	(4)
Gradient x 10	-0.083^{**}	-0.075^{**}	-0.078^{***}	-0.077^{***}
	(0.040)	(0.034)	(0.027)	(0.027)
Poverty Rate		0.023	0.019	0.017
		(0.069)	(0.070)	(0.069)
Female-headed HHs		0.393^{***}	0.165	0.155
		(0.120)	(0.107)	(0.107)
Adult sex ratio		-0.173^{***}	-0.130^{***}	-0.121^{***}
		(0.052)	(0.042)	(0.042)
Indian, white adults x 10		-12.357^{***}	-11.156^{**}	-11.050^{**}
		(4.007)	(4.588)	(4.522)
KMs to road x 10		0.003	-0.010	-0.010
		(0.009)	(0.010)	(0.010)
Kms to town x 10		0.016	0.008	0.008
		(0.015)	(0.015)	(0.016)
Women with high school		-0.269	-0.185	-0.152
		(0.500)	(0.411)	(0.417)
Men with high school		1.046^{**}	0.965^{**}	0.984^{**}
		(0.475)	(0.413)	(0.409)
Household Density x 10		0.017^{***}	0.012^{**}	0.013^{**}
		(0.004)	(0.006)	(0.006)
Kms to grid x 10		-0.040^{*}	-0.012	-0.011
		(0.021)	(0.023)	(0.023)
Change in water access				0.012
				(0.048)
Change in toilet access				0.155
				(0.104)
Constant	0.285^{***}	0.216^{**}		
	(0.052)	(0.102)		
District FE	N	N	Y	Y
Observations	1,816	1,816	1,816	1,816
R^2	0.010	0.074	0.177	0.178
Adjusted R^2	0.010	0.068	0.167	0.167
Residual Std. Error	0.399 (df = 1814)	0.387 (df = 1804)	0.366 (df = 1795)	0.366 (df = 1793)
Notes:	***Significant at the 1 percent level.			
	**Significant at the 5 percent level.			
	*Significant at the 10 percent level.			

[3]The code for this question can be found here: Question 4

Table 3: Effects of Elect. on Female Employment: Census Community Data

	Ordinart Least Squares				*Two Stage Least Squares*			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Eksom Project	−0.004	−0.001	0.000	−0.001	0.025	0.074	0.090*	0.095*
	(0.005)	(0.005)	(0.005)	(0.005)	(0.045)	(0.060)	(0.055)	(0.055)
Poverty Rate		0.029***	0.033***	0.031***		0.027**	0.032**	0.031**
		(0.011)	(0.010)	(0.010)		(0.012)	(0.013)	(0.013)
Female-headed HHs		0.042**	0.051***	0.047**		0.014	0.036	0.033
		(0.019)	(0.019)	(0.020)		(0.031)	(0.026)	(0.026)
Adult Sex Ratio		0.019**	0.017**	0.020***		0.033**	0.029**	0.032***
		(0.009)	(0.008)	(0.007)		(0.014)	(0.012)	(0.012)
Baseline Controls?	N	Y	Y	Y	N	Y	Y	Y
District FE?	N	N	Y	Y	N	N	Y	Y
Change: other services?	N	N	N	Y	N	N	N	Y

Notes: ***Significant at the 1 percent level.
**Significant at the 5 percent level.
*Significant at the 10 percent level.

Table 4: Effects of Elect. on Male Employment: Census Community Data

	Ordinart Least Squares				*Two Stage Least Squares*			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Eksom Project	−0.017**	−0.015***	−0.009	−0.010*	−0.063	0.069	0.033	0.035
	(0.007)	(0.006)	(0.006)	(0.006)	(0.073)	(0.082)	(0.064)	(0.066)
Poverty Rate		0.062***	0.064***	0.063***		0.059***	0.064***	0.062***
		(0.020)	(0.018)	(0.018)		(0.022)	(0.019)	(0.019)
Female-headed HHs		0.217***	0.233***	0.227***		0.187***	0.227***	0.220***
		(0.029)	(0.030)	(0.030)		(0.042)	(0.034)	(0.034)
Adult Sex Ratio		0.018*	0.012	0.017		0.034*	0.018	0.023
		(0.011)	(0.011)	(0.011)		(0.019)	(0.015)	(0.015)
Baseline Controls?	N	Y	Y	Y	N	Y	Y	Y
District FE?	N	N	Y	Y	N	N	Y	Y
Change: other services?	N	N	N	Y	N	N	N	Y

Notes: ***Significant at the 1 percent level.
**Significant at the 5 percent level.
*Significant at the 10 percent level.

Part B

What was the grid that Dinkelman used for the Anderson-Rubin test? What happens if you use a finer grid?

They used a 0.05 grid which started lower bound -.6 and ended at 1. It would have been better if they used a finer grid, say 0.005. In my code I did the experiment and found that the confidence set is [0.01-0.35] instead of [0.05-0.3] as reported in table 4 of the paper.

Part C

Use the jackknife instead of TSLS to estimate columns (5)–(8) of Tables 4 and 5. Discuss your findings.

Solution

Table 5 shows Jacknife estimates. It is interesting to find that the paper's results disappear, and coefficients become negative for the specifications that include fixed effects. What possibly happened was that the instrument was weak and idiosyncratic shocks in the first stage were correlated with shocks in the second stage U_i. According to the notes this issue should vanish asymptotically.

However, in Dinkelman's paper there are few observations and the fixed effects model requires many degrees of freedom, which exacerbates the problem. The fact that the coefficient flips is a clear sign that the error terms in both stages were correlated to begin with, and the usual IV method might recover biased estimates such as the one reported in table5 of the paper.

Jacknife in this case may be more reliable and actually brings us closer to OLS estimates, meaning that the bias in IV is similar to the bias of OLS, which suggests weak instruments.

Table 5: Jacknife IV estimation Results

	Female Employment				Male Employment			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Eksom Project	0.0274	0.284	-0.149	-0.119	-0.0693	0.338	-0.0474	-0.0443
Poverty Rate	-0.00562	0.0207	0.0345	0.0326	-0.0232	0.0519	0.0647	0.0628
Female-headed HHs		-0.0613	0.074	0.0637		0.0896	0.239	0.231
Adult Sex Ratio		0.0726	-0.005	0.004		0.0841	0.00631	0.0124
Baseline Controls	N	Y	Y	Y	N	Y	Y	Y
District Fixed Effects	N	N	Y	Y	N	N	Y	Y
Change: Other services	N	N	N	Y	N	N	N	Y

Problem 5

Consider the binary treatment potential outcomes model with $D \in \{0,1\}$ and $Y = DY(1) + (1-D)Y(0)$. For concreteness, suppose that D is whether one enrolls in a job training course, and Y is earnings at some point afterwards. Suppose that we also have a set of predetermined covariates, X. Our data consists of these variables for both workers who enrolled in the job training course, and those who did not.

Part A

Suppose job training experiment accepts all applicants into the course. Using the definitions in the lecture notes, show that selection on observables is not falsifiable.

Proof. Parameter space for $\theta = F$ can be represented by:

$$\Theta = \{F \in \mathcal{F} : (Y(0), Y(1)) \perp D|X \text{ under } F\}$$

And each θ implies a distribution for (Y,D,X) which we will call G_θ.

Let's define further our target parameter as the ATE:

$$\pi : \Theta \to R : \pi(F) = E_F[Y(1) - Y(0)|X]$$

And our identified set:

$$\Theta^*(G) = \{\theta \in \Theta : G_\theta = G\}$$
$$\Pi^*(G) = \{\pi(\theta) : \theta \in \Theta, G_\theta = G\}$$

Now, we can proof that the set is non empty for any $G \in \mathcal{G}$. Take any distribution $G \in \mathcal{G}$. By means of contradiction suppose there exists θ and $\theta^* \in \Theta$ such that $\theta \neq \theta^*$, but $F_\theta = F_{\theta^*} = F$. Then, we have that:

$$\begin{aligned}
\pi(\theta^*) &= E_{F_\theta^*}[Y(1) - Y(0)|X] \\
&= E_{F_\theta}[Y(1) - Y(0)|X] \\
&= \pi(\theta) \\
\implies \pi(\theta) &= \pi(\theta^*)
\end{aligned}$$

The first line just defines ATE under F_θ^*. The second line uses our assumption that $F_\theta = F_{\theta^*}$, the third line is basically the definition of the target parameter for F_θ, and the fourth line is the conclusion of our argument, which implies that the set is point identified.

Now that we have found that π is point identified for any $G \in \mathcal{G}$, assume by contradiction that $\exists\ \tau(G) : \mathcal{G} \to \{0,1\}$ so that $\tau(G) = 1$ for at least one $G \in \mathcal{G}$. Since we previously found that the model is identified for any $G \in \mathcal{G}$, that implies that there is no $G \in \mathcal{G}$ where $\tau(G) = 1$. This is a contradiction, therefore the model is not falsifiable.

□

Part B

Suppose the job training experiment has the following structure. First we open the program to everyone and collect a list of workers who apply to take the program. Then we offer the program to a random subset of these applicants, but not provide job training for any of the applicants not in this random subset. We collect data on the outcomes for workers who took the program, workers who applied to take the program but where not randomized in, and other workers who did not even apply to the program.

Explain how this structure could be used to falsify selection on observables.

Solution

Now we have information about people that applied $A = 1$ and did not get the training $A = 0$ because of random assignment. Now we have three groups D(1) & A(1), D(0) & A(1), D(0) & A(0). So what we have to do here is to compare two moments:

$$\phi_1(G) = \mathbb{E}[Y(D = 1) - Y(D = 0)|X]$$

ϕ_1 is the model with selection on observables, assuming that conditioning on X can control for all factors that induce people into selection, the difference in means should reveal the net ATE of the policy.

$$\begin{aligned}\phi_2(G) &= \mathbb{E}[Y(D, A)|D = 1, A = 1, X] - \mathbb{E}[Y(D, A)|D = 0, A = 1, X] \\ &= \mathbb{E}[Y(D, A)|D = 1, X] - \mathbb{E}[Y(D, A)|D = 0, X]\end{aligned}$$

ϕ_2 is again the selection on observables model but conditioned on applying. Since X controls for all the factors inducing people to apply, then X has enough information and A should be redundant which implies phi_2 should also give the same estimate as ϕ_1 if the model were true.

Finally we can use those two expressions to construct $\tau(G) = \mathbb{1}\{\phi_1(G) \neq \phi_1(G)\}$ and test for falsifiability. Note that the two expressions should be giving the same result if the model were true.