EC 587 HW3

Erik Andersen

October 2023

Question 1

a)

In this section, x is measured without measurement error. The mean and standard deviations are reported below.

Variable	Coefficient	SE
β_0	2.999582	0.05120259
β_1	1.000045	0.00502866

With no measurement error, the coefficients aren't biased, so we get the true value of $\beta = (3, 1)$

b)

To find the probability limit of $\hat{\beta}_1$ when there is measurement error, consider the formula for β_1: $\beta_1 = \frac{cov(x,y)}{var(x)}$. Now x is measured with measurement error, so $x = x + u$. Plugging this in, we get $\hat{\beta}_1 = \frac{cov(x+u,\beta x+\epsilon)}{var(x+u)}$. Taking the plim, we get:

$$plim\hat{\beta}_1 = \beta\frac{\sigma_x^2}{\sigma_x^2 + \sigma_u^2} \tag{1}$$

We defined the variance of x to be 4, and in this case, the variance of the measurement error is 1, so the plim of $\hat{\beta}_1$ is $\frac{4}{4+1} = 0.8\beta$. So we should expect to get a value of 0.8 for $\hat{\beta}_1$ in our simulations.

c)

The means are reported below.

Variable	Coefficient	SE
β_0	5.0001280	0.061912535
β_1	0.7999666	0.006039574

The estimates are exactly as predicted.

d)

Plugging the new measurement error variance into (1) gives us $\hat{\beta}_1 = \frac{4}{4+16} = 0.2\beta$. The estimated values are reported below.

Variable	Coefficient	SE
β_0	11.0002995	0.050121709
β_1	0.1999923	0.004614493

e)

Below are the estimated means and variances on the coefficients from the first stage regression $x^{(1)} = \gamma_0 + \gamma_1 x^{(2)} + u$.

Variable	Coefficient	SE
γ_0	7.9999113	0.1999772
γ_1	2.453540e-03	2.031207e-05

f)

Below are the estimates for $\hat{\beta}_1$ from the second stage regression.

Variable	Coefficient	SE
β_0	2.999744	0.0253136288
β_1	1.000025	0.0002505685

As expected, we recover the true values of the coefficients, but we have different standard errors.

g)

Below are the estimated means and standard errors if we reverse the order of the first stage regression

Variable	Coefficient	SE
β_0	2.0017287	0.0355442648
β_1	0.7998805	0.0003375255

Then here are the second stage results.

Variable	Coefficient	SE
β_0	2.999744	0.0253136288
β_1	1.000025	0.0002505685

The second stages are exactly the same in both regressions in both point estimates and standard errors.

h)

The average estimated across simulations is 0.007504754. The estimated standard error across betas is 0.01582936. So we estimated the parameter more precisely than indicated by the variance of the betas.

i)

The means for the ivreg are given below.

Variable	Coefficient	Variance
β_0	2.993525	0.0554598366
β_1	1.000582	0.0005364982

The mean of the variances across simulations is 0.0005332395, which is almost exactly the same. This is different than in h. In this case, the comparison was the same either way of calculating the se, where in h it gave us a different value. In h, we weren't taking into account the extra variance of running two regressions whereas we do in part i.

Question 2

a)

The table below gives the result of the ols regression of the endogenous treatment and controls on the proportion of women employed.

Variable	Coefficient	Variance
T	-5.169057e-04	0.0048657111

These match the rounded values from the paper.

b)

Below are the results from the first stage regression.

Variable	Coefficient	Variance
Gradient	-0.0077427714	0.0027770472

These match the paper again.

c)

The calculated F statistic is 8.211135, which basically matches the 8.26 from the paper. The p-value implied by this statistic is 8.084736e-20. So we strongly reject the null.

d)

If we calculate the f-statistic as the square of the T-stat from part by, we get 7.773675 which implies a p-value of 0.006595224 which also rejects the null. We get a smaller f stat, and so a larger p-value than from part c.

e)

Below are the estimate and se from the second stage regression.

Variable	Coefficient	Variance
T_hat	0.0950786454	0.0450853776

The coefficient matches, but the standard error is a bit smaller. This is likely because we did this with two stages, so we aren't taking into account the variation of our generated regressor from the first stage.

f)

The T-stat for $\hat{T}$ from part e is 2.1088577, which implies a p-value of 3.802573e-02. So the estimate is significant at the 5% level.

g)

The point estimate on T is 0.0950786 which is the same as e. The standard error is 0.0546590, which is the same as in the paper, but differs from part e. The larger standard error is due to the generated regressor in the first stage. The t-stat is 1.739 with implied p-value of 0.08212. So the estimate is only significant at the 10% level.

h)

The coefficient on gradient from the reduced form regression is -7.362e-04. This is how much the instrument effects the outcome directly. If the instrument is exogenous, this should be zero, which it basically is.

i)

The wald estimator give us 0.09507865, which is the same as the two previous results.

j)

The results from gmm and LIML are below.

Estimator	Coefficient	Variance
GMM	0.09507865	0.0542875361
LIML	0.09507865	0.05465903

These are basically exactly the same as previous results. They should be the same because IV is a special case of gmm and LIML.

k)

The coefficient and p-value are reported below.

Coefficient	Variance
-0.0963166181	0.0451813456

The t-stat and associated p-value are -2.1317784, 3.604326e-02. So we reject the null that there the residuals have no effect on the outcome variable. Thus, we cannot reject there being endogeneity in the instrument.