Suggested Answers Problem set 4
ECON 60303

Bill Evans
Spring 2014

1. A program that answers part A) is on the web page and it is named psid_iv_comparison.do. Below are some key results and a summary table is included. The standard errors are so much smaller in the fixed-effect estimate because the RMSE is ½ the size – the fixed-effects explain a large fraction of the variance in outcomes.

```
. ************************************;
. * get fixed effects estimates;
. ************************************;
. areg wagel union tenure, absorb(id);

Linear regression, absorbing indicators                Number of obs =    3945
                                                       F(  2,  3154) =   36.31
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.8486
                                                       Adj R-squared =  0.8106
                                                       Root MSE      =  .23987

------------------------------------------------------------------------------
       wagel |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       union |   .0358574   .0240811     1.49   0.137    -.0113588    .0830735
      tenure |   .0121499   .0014669     8.28   0.000     .0092738    .0150261
       _cons |   2.273801    .013821   164.52   0.000     2.246702      2.3009
-------------+----------------------------------------------------------------
          id |      F(788, 3154) =     20.308   0.000         (789 categories)

. ************************************;
. * check that within panel means have zero mean;
. ************************************;
. sum uniond tenured;

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
      uniond |      3945   -8.91e-10    .1589137        -.8         .8
     tenured |      3945   -7.44e-09     2.60879     -23.72       20.7


. ************************************;
. * run 2sls;
. ************************************;
. reg wagel union tenure (uniond tenured);

Instrumental variables (2SLS) regression

      Source |       SS       df       MS              Number of obs =    3945
-------------+------------------------------           F(  2,  3942) =    7.37
       Model |  80.8876107     2  40.4438054           Prob > F      =  0.0006
    Residual |  1117.58494  3942  .283507089           R-squared     =  0.0675
-------------+------------------------------           Adj R-squared =  0.0670
       Total |  1198.47255  3944  .303872352           Root MSE      =  .53245

------------------------------------------------------------------------------
       wagel |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       union |   .0358574   .0534535     0.67   0.502    -.0689417    .1406564
      tenure |   .0121499   .0032561     3.73   0.000     .0057661    .0185337
       _cons |   2.273801   .0306789    74.12   0.000     2.213653    2.333949
```

Parameter Estimates and Standard errors

Covariates	Fixed Effect	2SLS
Union	0.0359	0.0359
	(0.0240)	(0.0534)
Tenure	0.0121	0.0121
	(0.0015)	(0.0032)

2. a) It is easy to show that the instrument is uncorrelated with $u_{i.}$. The correlation coefficient between u_i and $\tilde{X}_{it}$ for the sample is by definition $\hat{\sigma}_{\tilde{x}u} = \frac{1}{n-1}\sum_{i=1}^{n}\sum_{t=1}^{T}(\tilde{X}_{it} - \bar{\tilde{X}})(u_i - \bar{u})$. Working with the summation terms, note that we can drop one of the means so drop $\bar{u}$ and note that by construction $\bar{\tilde{X}}$ =0 so $\hat{\sigma}_{\tilde{x}u} = \frac{1}{n-1}\sum_{i=1}^{n}\sum_{t=1}^{T}\tilde{X}_{it}u_i$. Because u_i does not vary within t, we can write this as $\hat{\sigma}_{\tilde{x}u} = \frac{1}{n-1}\sum_{i=1}^{n}u_i\sum_{t=1}^{T}\tilde{X}_{it}$. Note that by construction, $\sum_{t=1}^{T}\tilde{X}_{it} = 0$ so by construction $\hat{\sigma} = 0$.

b) Sort the data by person then year. Let $Z = \tilde{X}$ and because the data is sorted by person then year $Z = \tilde{X} = MX$ where $M = i_n \otimes M_t$ and $M_t = I_t - \frac{1}{t}i_t i_t'$. The IV estimate for β is $\hat{\beta} = (Z'X)^{-1}Z'Y = [(MX)'X]^{-1}[(MX)'Y] = [X'MX]^{-1}[X'MY]$ which is by definition the fixed effect estimate.

3. Let $\text{var}(\hat{\theta}_1) = \sigma_\theta^2$ and $\text{var}(\hat{\pi}_1) = \sigma_\pi^2$.

$$\text{var}\begin{bmatrix}\hat{\theta}_1 \\ \hat{\pi}_1\end{bmatrix} = \begin{bmatrix}\sigma_\theta^2 & 0 \\ 0 & \sigma_\pi^2\end{bmatrix} = \Sigma$$

$Var\left(\hat{\beta}_1\right) = g'\Sigma g$ were $g = \begin{bmatrix}\frac{1}{\hat{\pi}_1} \\ \frac{-\hat{\theta}_1}{\hat{\pi}_1^2}\end{bmatrix}$ and therefore $Var\left(\hat{\beta}_1\right) = \frac{\sigma_\theta^2}{\hat{\pi}_1^2} + \frac{\sigma_\pi^2\hat{\theta}_1^2}{\hat{\pi}_1^4}$

$$t(\hat{\beta}_1)^2 = \frac{\frac{\hat{\theta}_1^2}{\hat{\pi}_1^2}}{\frac{\sigma_\theta^2}{\hat{\pi}_1^2} + \frac{\sigma_\pi^2\hat{\theta}_1^2}{\hat{\pi}_1^4}} = \frac{1}{\frac{\sigma_\theta^2}{\hat{\pi}_1^2}\frac{\hat{\pi}_1^2}{\hat{\theta}_1^2} + \frac{\sigma_\pi^2\hat{\theta}_1^2}{\hat{\pi}_1^4}\frac{\hat{\pi}_1^2}{\hat{\theta}_1^2}} = \frac{1}{\frac{\sigma_\theta^2}{\hat{\theta}_1^2} + \frac{\sigma_\pi^2}{\hat{\pi}_1^2}} = \frac{1}{\frac{1}{t(\hat{\theta}_1)^2} + \frac{1}{t(\hat{\pi}_1)^2}}$$

The reason why this result is important is that consider a well-specified 2SLS model with a strong first stage. This means that if $t(\hat{\pi}_1)^2$ is large then $\frac{1}{t(\hat{\pi}_1)^2}$ →0. We can then see that

$t(\hat{\beta}_1)^2 \approx \frac{1}{\frac{1}{t(\hat{\theta}_1)^2}} = t(\hat{\theta}_1)^2$. Therefore, with a strong first stage, the t-statistic on the 2SLS estimate will be functionally the same as the t-statistic on the reduced form.

4. We answered this in class. They key point of the monotonicity assumption is that the upper right corner box in the 2x2 box is the empty set. Suppose there are two types of families: those that prefer a mix so they try for more kids when endowed with two boys or two girls, and families that prefer more boys than girls, so when they are endowed with a mix of boys and girls, they try for a third. In this case, the numerator and denominator in the Wald estimate will contain a mixture of two groups of people and as a result, it is not clear for what group the IV estimate is measuring.

5. It is easiest to first show why the model works when we use $(z_i - z_0)$ in $h^m(z_i)$. To make things easy, assume that ρ=1 and hence, in the first stage, $h^2(z_i) = D_i\delta_1^{2+}(z_i - z_0) + (1-D_i)\delta_1^{2-}(z_i - z_0)$.

In this case $x_i = \pi_0 + D_i\pi_1 + w_i\pi_2 + D_i\delta_1^{2+}(z_i - z_0) + (1-D_i)\delta_1^{2-}(z_i - z_0) + v_i$ and

$$E[x_i \mid D_i = 1] = \pi_0 + \pi_1 + w_i\pi_2 + D_i\delta_1^{2+}(z_i - z_0)$$
$$E[x_i \mid D_i = 0] = \pi_0 + w_i\pi_2 + (1-D_i)\delta_1^{2-}(z_i - z_0)$$
$$\text{ATT } = E[x_i \mid D_i = 1] - E[x_i \mid D_i = 0] = \pi_1 + D_i\delta_1^{2+}(z_i - z_0) - (1-D_i)\delta_1^{2-}(z_i - z_0)$$

For people that are just below the cutoff, regardless of the δ_1^{2-}, $(1-D_i)\delta_1^{2-}(z_i - z_0) \approx 0$ because $(z_i - z_0) \approx 0$. For people right above the cutoff, $D_i\delta_1^{2+}(z_i - z_0) \approx 0$ again because $(z_i - z_0) \approx 0$. Therefore *ATT* collapses to $ATT = E[x_i \mid D_i = 1] - E[x_i \mid D_i = 0] = \pi_1$

When we so not rescale the running variable $x_i = \pi_0 + D_i\pi_1 + w_i\pi_2 + D_i\delta_1^{2+}z_i + (1-D_i)\delta_1^{2-}z_i + v_i$ and hence $E[x_i \mid D_i = 1] - E[x_i \mid D_i = 0] = \pi_1 + D_i\delta_1^{2+}z_i - (1-D_i)\delta_1^{2-}z_i$. Now, the running variable terms do not drop out even though they are equal in value (z).

Empirical portion

The program that answers questions 1-7 is called twin1sta.do and the program for question 8 is called twin1st_random_instrument.do.

1. Although twins by construction increase the number of children in the house by 1 when they are first born, the coefficient on twin1st in the first stage is only 0.27. The coefficient is < 1 because the constraint

generated by twins may not bind over time. Suppose a woman was planning on having two kids: one at age 22 and another at age 25. If she had a twin at age 22, the coefficient on win1st in the 1st stage at that age would be 1 – however, by age 26, the coefficient would have fallen to zero because the mother would have had the second child anyway. Note the R^2 in the first stage is 0.152 and the ratio $Var(\hat{\beta}_1^{OLS})/Var(\hat{\beta}_1^{2SLS})$ for weeks worked is $(0.5745)^2/(1.4756)^2=0.152$

2. Note that in 5 of 6 cases, there is a statistically significant in the exogenous covariates difference between mothers with and without twins. Although twins are thought to be random, the correlation between twin status and observed characteristics gives one pause for concern that the results are subjected to omitted variables bias.

3. Note that there is a slight different in the 2SLS estimates in the case of the Wald estimate and the one controlling for covariates. This is because the control variables are correlated with the twin instrument.

4. The correlaton coefficient is 0.39, so he correlation between z and ε must be about 0.4 the rate of x and ε in order for the 2SLS estimate to be more consistent that OLS.

5. Note that the 1st stage coefficients increase as we interact age at first birth with twin. In general, the twin birth should be more of a shock to fertility the older the mother which is exactly the case. The first-stage F statistic indicates there is no concern about finite sample bias.

6. The p-value on the test of over-identifying restrictions indicates that we cannot reject the null the model is appropriately specified. Note the similarity of the estimates between the over-identified model and the just identified model in the previous table.

7. Results from the simulation are in the table below. The key aspect of the simulation is that as you add junk to the model in the form of 5, 10 and 30 instruments, the first stage F plummets. Notice as well that as this F falls, the 2SLS estimate systematically moves towards the OLS estimate. Therefore, adding useless instruments to the model does come at a cost in that the model approaches the OLS estimates.

Answers for Question 2
Parameter estimates and (standard errors)

	1st stage	Reduced-form		Wald	
Outcome:	Second	weeks	worked	Weeks	Worked
Second		-0.990	-0.249	-3.605	-0.091
		(0.407)	(0.009)	(1.476)	(0.032)
Twin1st	-0.274				
	(0.006)				
R^2	0.1519	0.0005	0.0006	0.0087	0.0084

Answers for Problem 3
Coefficient on twin1st and (standard error)

educ	agefts	agem	White	black	other_race
0.127	0.749	0.521	-0.034	0.033	0.0005
(0.045)	(0.064)	(0.087)	(0.006)	(0.006)	(0.0029)

Answers for Question 4
Parameter estimates and (standard errors)
Controlling for covariates

	1st stage	OLS		2SLS	
Outcome:	Second	weeks	worked	Weeks	Worked
Second		-9.801	-0.180	-3.555	-0.077
		(0.577)	(0.013)	(1.402)	(0.030)
Twin1st	0.283				
	(0.006)				
1st stage F	2572				
R^2	0.1519	0.226	0.054	0.061	0.049

Correlation coefficient (second, twin1st)=0.39

Answers for question 6/7
Parameter estimates and (standard errors)
Controlling for covariates

	1st stage	OLS		Wald	
Outcome:	Second	Weeks	worked	Weeks	worked
Second		-9.801	-0.180	-3.143	-0.077
		(0.577)	(0.013)	(1.370)	(0.030)
Twin1st x agefst<20	0.234				
	(0.008)				
Twin1st x 20≤agefst<25	0.283				
	(0.008)				
Twin1st x Agefst>24	0.383				
	(0.011)				
1st stage F	907.4				
Overid test (p-value)				1.90	1.92
				(0.39)	(0.38)
R^2	0.233	0.054	0.061	0.060	0.049

Answers for question 8

Estimation Method	Instruments	1st stage F (actual or average)	OLS or 2SLS (actual or average)
OLS			-9.80
2SLS	Mysteryz	45.9	-4.97
2SLS	1000 replications, mysterz and 5 random instruments	8.45	-5.49
2SLS	1000 replications, mysterz and 10 random instruments	5.06	-5.82
2SLS	1000 replications, mysterz and 30 random instruments	2.44	-7.01
2SLS	1000 replications, 10 random instruments	0.98	-9.98