

Econometrics PS6

Jabez Koh

29 April 2023

1. The Canvas Module for Pset 6 contains a version of the Krueger (1999) data set with information on randomly assigned class size, student and teacher characteristics, and school identifiers.
 - (a) Table VII in Krueger (1999) reports OLS and 2SLS estimates of class size effects. What are the instrument(s)? What's the motivation for 2SLS in this application?

The author used the initial assignment to a class type as an instrument for actual class size. Specifically, two instruments were used - (1) S_{ios} , which is a dummy variable indicating assignment to a small class in the first year the student is observed in the experiment; and (2) R_{ios} , which is a dummy variable indicating assignment to a regular class in the first year the student is observed in the experiment.

The motivation for using 2SLS is the following:

- (a) The appropriate model of achievements is to use actual class size.
- (b) The actual class size is not random as student mobility and enrollment differs across schools.
- (c) Using the random initial assignment to a class type, the authors could get the causal impact of class size on achievement.

- (b) Replicate the estimates in the first row of this table (for kindergarteners). Using the estimated first stage that goes with the 2SLS estimates in the replication to explain why 2SLS estimates are close to the corresponding OLS estimates.

TABLE VII
OLS and 2SLS Estimates of Effect of Class Size on Achievement
Dependent Variable: Average Percentile Score on SAT

| Grade | OLS | 2SLS | Sample Size |
|----------------------|--------------|--------------|--------------------|
| | (1) | (2) | (3) |
| K | -.62 | -.71 | 5,861 |
| (Paper) | (.14) | (.14) | |
| K | -.62 | -.71 | 5,683 |
| (Replication) | (.14) | (.15) | |

The estimated first stage is appended below:

| | | | | | | |
|----------------------|------------------|-----------------|---------------|--------------|------------------|------------------|
| cltypek | | | | | | |
| small class | -7.527766 | .1744252 | -43.16 | 0.000 | -7.869707 | -7.185825 |
| regular class | -.2536257 | .1627716 | -1.56 | 0.119 | -.5727213 | .0654698 |

The goal of Project STAR is to reduce class size from 22 to 15 students, a 7 students decrease. From the first stage equation, being in a small class indeed reduced the actual class size by around 7 students, while being in a regular class has no statistical significant impact (at all 90%, 95% and 99% confidence) on the actual class size. This shows that the instruments are highly correlated with the actual class size, which results in very similar effect size for OLS vs 2SLS.

```

.
.
. * Load data
.
. use k99.dta, clear

.
. *****Question 1b*****
.
. * Replicate Table VII
.
.       reg pscorek cs i.schidkn srace ssex sesk tracek totexpk hdegk if
stark == 1, cluster(classid)

```

```

Linear regression          Number of obs      =
5,683                     F(85, 315)        =
75.50                     Prob > F          =
0.0000                    R-squared         =
0.3100                    Root MSE       =
22.671

```

(Std. err. adjusted for 316 clusters in classid)

| ----- | | | | | | |
|-----------|-------------|-----------|-------|-------|------------|---|
| ----- | | | | | | |
| | | Robust | | | | |
| pscorek | Coefficient | std. err. | t | P> t | [95% conf. | |
| interval] | | | | | | |
| ----- | | | | | | |
| ----- | | | | | | |
| cs | -.6191919 | .141179 | -4.39 | 0.000 | -.8969648 | |
| -.3414189 | | | | | | |
| | | | | | | |
| schidkn | | | | | | |
| 2 | -28.31833 | 6.446872 | -4.39 | 0.000 | -41.0027 | - |
| 15.63396 | | | | | | |
| 3 | .0531128 | 6.294534 | 0.01 | 0.993 | -12.33153 | |
| 12.43776 | | | | | | |
| 4 | -19.87166 | 5.729031 | -3.47 | 0.001 | -31.14366 | - |
| 8.599654 | | | | | | |
| 5 | -19.84574 | 9.133431 | -2.17 | 0.031 | -37.81598 | - |
| 1.875498 | | | | | | |
| 6 | -12.52528 | 5.795064 | -2.16 | 0.031 | -23.9272 | - |
| 1.123355 | | | | | | |
| 7 | 8.835789 | 6.513257 | 1.36 | 0.176 | -3.979197 | |
| 21.65078 | | | | | | |
| 8 | -10.78434 | 6.246069 | -1.73 | 0.085 | -23.07363 | |
| 1.504943 | | | | | | |
| 9 | 2.929278 | 6.026757 | 0.49 | 0.627 | -8.928509 | |
| 14.78707 | | | | | | |
| 10 | 9.169704 | 8.06569 | 1.14 | 0.256 | -6.69973 | |
| 25.03914 | | | | | | |

| | | | | | | | | |
|----------|----|--|-----------|----------|-------|-------|-----------|---|
| 29.6015 | 11 | | 17.31602 | 6.244136 | 2.77 | 0.006 | 5.030535 | |
| 18.33002 | 12 | | 5.221786 | 6.662302 | 0.78 | 0.434 | -7.886449 | |
| 27.49103 | 13 | | 15.73002 | 5.977568 | 2.63 | 0.009 | 3.969017 | |
| 10.13074 | 14 | | -3.547827 | 6.952176 | -0.51 | 0.610 | -17.2264 | |
| 2.864711 | 15 | | -10.36924 | 6.726199 | -1.54 | 0.124 | -23.6032 | |
| 1.646128 | 16 | | -15.40002 | 6.990459 | -2.20 | 0.028 | -29.15391 | - |
| 7.597293 | 17 | | -6.802139 | 7.318556 | -0.93 | 0.353 | -21.20157 | |
| -.526273 | 18 | | -12.8082 | 6.242327 | -2.05 | 0.041 | -25.09012 | |
| 16.44173 | 19 | | 2.204592 | 7.236069 | 0.30 | 0.761 | -12.03254 | |
| 12.49144 | 20 | | -3.474395 | 8.114686 | -0.43 | 0.669 | -19.44023 | |
| 25.00994 | 21 | | 12.3755 | 6.421495 | 1.93 | 0.055 | -.2589465 | |
| 17.20805 | 22 | | 2.309253 | 7.572359 | 0.30 | 0.761 | -12.58954 | |
| 26.80765 | 23 | | 11.19661 | 7.934362 | 1.41 | 0.159 | -4.414437 | |
| 2.171538 | 24 | | -10.28383 | 6.33048 | -1.62 | 0.105 | -22.7392 | |
| 12.73839 | 25 | | -6.710685 | 9.885051 | -0.68 | 0.498 | -26.15976 | |
| 3.040806 | 26 | | -15.96063 | 6.56654 | -2.43 | 0.016 | -28.88045 | - |
| 36.34423 | 27 | | 21.66963 | 7.458409 | 2.91 | 0.004 | 6.995036 | |
| 10.63022 | 28 | | -3.594464 | 7.229741 | -0.50 | 0.619 | -17.81915 | |
| 23.82272 | 29 | | 5.290205 | 9.419209 | 0.56 | 0.575 | -13.24231 | |
| 50.0045 | 30 | | 32.5435 | 8.874607 | 3.67 | 0.000 | 15.0825 | |
| 29.07661 | 31 | | 16.13666 | 6.576767 | 2.45 | 0.015 | 3.196721 | |
| 4.92598 | 32 | | -16.94212 | 6.10724 | -2.77 | 0.006 | -28.95826 | - |
| 3.248761 | 33 | | -11.13779 | 7.312011 | -1.52 | 0.129 | -25.52435 | |
| 5.044605 | 34 | | -18.09872 | 6.634797 | -2.73 | 0.007 | -31.15284 | - |
| 3.802104 | 35 | | -15.87468 | 6.135923 | -2.59 | 0.010 | -27.94725 | - |
| 10.00856 | 36 | | -4.947029 | 7.601223 | -0.65 | 0.516 | -19.90261 | |
| 13.9522 | 37 | | -.2295561 | 7.207924 | -0.03 | 0.975 | -14.41132 | |
| 12.26287 | 38 | | -1.07009 | 6.776521 | -0.16 | 0.875 | -14.40305 | |
| 6.921639 | 39 | | -7.707228 | 7.435168 | -1.04 | 0.301 | -22.3361 | |

| | | | | | | | | |
|-----------|----|--|-----------|----------|-------|-------|-----------|---|
| 29.24323 | 40 | | 10.73197 | 9.408406 | 1.14 | 0.255 | -7.779294 | |
| 23.07131 | 41 | | 10.7444 | 6.26519 | 1.71 | 0.087 | -1.58251 | |
| 3.756429 | 42 | | -7.500579 | 5.72141 | -1.31 | 0.191 | -18.75759 | |
| 3.28129 | 43 | | -11.45673 | 7.490643 | -1.53 | 0.127 | -26.19474 | |
| 31.71669 | 44 | | 13.27225 | 9.374445 | 1.42 | 0.158 | -5.172192 | |
| 14.38327 | 45 | | -25.60206 | 5.701987 | -4.49 | 0.000 | -36.82086 | - |
| 9.728513 | 46 | | -4.444781 | 7.203621 | -0.62 | 0.538 | -18.61808 | |
| 5.824034 | 47 | | -7.931812 | 6.991451 | -1.13 | 0.257 | -21.68766 | |
| 15.04021 | 48 | | 1.457707 | 6.903348 | 0.21 | 0.833 | -12.12479 | |
| 17.48025 | 49 | | 5.859892 | 5.906085 | 0.99 | 0.322 | -5.760469 | |
| 15.03167 | 50 | | 2.194327 | 6.524617 | 0.34 | 0.737 | -10.64301 | |
| 17.23486 | 51 | | 4.254529 | 6.597292 | 0.64 | 0.519 | -8.725797 | |
| 16.72229 | 52 | | 5.410203 | 5.749402 | 0.94 | 0.347 | -5.901881 | |
| 10.40882 | 53 | | -23.38012 | 6.592704 | -3.55 | 0.000 | -36.35142 | - |
| 13.93139 | 54 | | .5405915 | 6.805913 | 0.08 | 0.937 | -12.8502 | |
| 1.257024 | 55 | | -12.5635 | 7.024323 | -1.79 | 0.075 | -26.38402 | |
| 20.44364 | 56 | | -33.55838 | 6.665607 | -5.03 | 0.000 | -46.67312 | - |
| 8.461751 | 57 | | -20.06135 | 5.895531 | -3.40 | 0.001 | -31.66094 | - |
| 18.18195 | 58 | | 6.500361 | 5.937206 | 1.09 | 0.274 | -5.181231 | |
| 18.5875 | 59 | | 2.44302 | 8.205482 | 0.30 | 0.766 | -13.70146 | |
| 26.55596 | 60 | | -12.88273 | 6.949461 | -1.85 | 0.065 | - | |
| 1.614908 | 61 | | -13.05466 | 5.814291 | -2.25 | 0.025 | -24.49442 | - |
| 2.001329 | 62 | | -10.01472 | 6.107193 | -1.64 | 0.102 | -22.03076 | |
| 23.69685 | 63 | | 8.765589 | 7.58886 | 1.16 | 0.249 | -6.165671 | |
| 1.290844 | 64 | | -10.58186 | 6.034341 | -1.75 | 0.080 | -22.45457 | |
| 18.58937 | 65 | | 5.579085 | 6.612517 | 0.84 | 0.399 | -7.431199 | |
| -.8017748 | 66 | | -14.71061 | 7.069207 | -2.08 | 0.038 | -28.61944 | |
| 7.394343 | 67 | | -5.299642 | 6.451758 | -0.82 | 0.412 | -17.99363 | |
| 28.88983 | 68 | | 15.49485 | 6.80804 | 2.28 | 0.024 | 2.099872 | |

| | | | | | | | | |
|----------|---------|--|-----------|----------|-------|-------|-----------|---|
| 20.24134 | 69 | | 8.982388 | 5.722398 | 1.57 | 0.117 | -2.276564 | |
| 9.728558 | 70 | | -4.107226 | 7.03208 | -0.58 | 0.560 | -17.94301 | |
| 3.735803 | 71 | | -15.21295 | 5.833295 | -2.61 | 0.010 | -26.69009 | - |
| 23.59664 | 72 | | 10.20317 | 6.807277 | 1.50 | 0.135 | -3.190312 | |
| 19.19481 | 73 | | 1.06372 | 9.215182 | 0.12 | 0.908 | -17.06737 | |
| 17.52679 | 74 | | .72788 | 8.538101 | 0.09 | 0.932 | -16.07103 | |
| 10.56249 | 75 | | -2.345154 | 6.560352 | -0.36 | 0.721 | -15.2528 | |
| 6.419606 | 76 | | -6.072909 | 6.34936 | -0.96 | 0.340 | -18.56542 | |
| 7.335483 | 78 | | -19.82213 | 6.346375 | -3.12 | 0.002 | -32.30877 | - |
| 7.276054 | 79 | | -5.690083 | 6.590079 | -0.86 | 0.389 | -18.65622 | |
| 11.81017 | 80 | | -3.578305 | 7.821239 | -0.46 | 0.648 | -18.96678 | |
| | | | | | | | | |
| 3.309827 | srace | | -5.411083 | 1.06797 | -5.07 | 0.000 | -7.512338 | - |
| 5.698751 | ssex | | 4.443381 | .6380457 | 6.96 | 0.000 | 3.188011 | |
| 15.1406 | sesk | | 13.59772 | .7841743 | 17.34 | 0.000 | 12.05484 | |
| 5.751215 | tracek | | 1.263755 | 2.280766 | 0.55 | 0.580 | -3.223706 | |
| 0.015 | totexpk | | .2653425 | .1088921 | 2.44 | | | |
| | | | .0510946 | .4795903 | | | | |
| 1.367389 | hdegk | | -.5096388 | .9540055 | -0.53 | 0.594 | -2.386667 | |
| 57.72318 | _cons | | 44.04913 | 6.949877 | 6.34 | 0.000 | 30.37508 | |


```
.
.       ivregress 2sls pscorek (cs = ib3.cltypek) i.schidk srace ssex
sesk tracek totexpk hdegk if stark ==
> 1, cluster(classid) first
```

First-stage regressions

| | | |
|--------|-----------------|---|
| 5,683 | Number of obs | = |
| 316 | No. of clusters | = |
| 63.72 | F(86, 5596) | = |
| 0.0000 | Prob > F | = |
| 0.9140 | R-squared | = |

0.9127 Adj R-squared =

1.1734 Root MSE =

| ----- | | | | | | |
|-------------|---------|-------------|---------------------|-------|-------|-------------------------|
| ----- | | | | | | |
| | cs | Coefficient | Robust std. err. | t | P> t | [95% conf. interval] |
| -----+----- | | | | | | |
| ----- | | | | | | |
| | schidkn | | | | | |
| 2.223358 | 2 | -.1052606 | 1.187836 | -0.09 | 0.929 | -2.433879 |
| 3.581466 | 3 | 2.645784 | .4772942 | 5.54 | 0.000 | 1.710102 |
| 3.91705 | 4 | 2.411308 | .7680836 | 3.14 | 0.002 | .9055659 |
| 4.218586 | 5 | 2.374901 | .9404692 | 2.53 | 0.012 | .5312164 |
| 2.861223 | 6 | .3214846 | 1.295528 | 0.25 | 0.804 | -2.218254 |
| 3.958689 | 7 | 3.051411 | .4628054 | 6.59 | 0.000 | 2.144133 |
| 4.724751 | 8 | 3.731752 | .5065318 | 7.37 | 0.000 | 2.738753 |
| 1.331865 | 9 | .4776551 | .435735 | 1.10 | 0.273 | -.3765544 |
| 1.90327 | 10 | .2634962 | .8364538 | 0.32 | 0.753 | -1.376278 |
| 2.197263 | 11 | 1.459609 | .3762793 | 3.88 | 0.000 | .7219561 |
| 2.642258 | 12 | 1.481424 | .5921449 | 2.50 | 0.012 | .3205901 |
| 3.446013 | 13 | 2.423109 | .5217864 | 4.64 | 0.000 | 1.400205 |
| 1.464466 | 14 | .5100729 | .4868387 | 1.05 | 0.295 | -.4443198 |
| 4.220595 | 15 | -1.919385 | 1.173854 | -1.64 | 0.102 | - |
| 5.875612 | 16 | 4.938643 | .4779509 | 10.33 | 0.000 | 4.001674 |
| 4.129381 | 17 | 3.198205 | .4749955 | 6.73 | 0.000 | 2.26703 |
| 5.061462 | 18 | 3.424953 | .8347884 | 4.10 | 0.000 | 1.788444 |
| 5.785612 | 19 | 4.110654 | .8544016 | 4.81 | 0.000 | 2.435695 |
| 2.838748 | 20 | 1.346882 | .7610055 | 1.77 | 0.077 | -.1449846 |
| 6.380459 | 21 | 4.052704 | 1.187395 | 3.41 | 0.001 | 1.724948 |
| 2.479857 | 22 | 1.279547 | .6122816 | 2.09 | 0.037 | .079238 |
| 3.341514 | 23 | 2.20807 | .5781732 | 3.82 | 0.000 | 1.074626 |

| | | | | | | | |
|-----------|----|--|-----------------------|----------|-------|-------|-----------|
| 1.651971 | 24 | | -.0225226 | .8541642 | -0.03 | 0.979 | -1.697016 |
| 6.648187 | 25 | | 4.209577 | 1.243943 | 3.38 | 0.001 | 1.770966 |
| 3.327471 | 26 | | .0335601 | 1.680234 | 0.02 | 0.984 | -3.260351 |
| 6.277735 | 27 | | 4.631441 | .8397797 | 5.52 | 0.000 | 2.985147 |
| 3.623241 | 28 | | 2.704557 | .4686238 | 5.77 | 0.000 | 1.785872 |
| 2.752243 | 29 | | .6143801 | 1.09053 | 0.56 | 0.573 | -1.523483 |
| 3.943776 | 30 | | 2.99517 | .4838872 | 6.19 | 0.000 | 2.046563 |
| 1.92785 | 31 | | .1865444 | .8882456 | 0.21 | 0.834 | -1.554762 |
| 3.805607 | 32 | | 2.681042 | .5736439 | 4.67 | 0.000 | 1.556478 |
| 6.75303 | 33 | | 5.268065 | .7574856 | 6.95 | 0.000 | 3.783099 |
| 1.957522 | 34 | | .4318151 | .778268 | 0.55 | 0.579 | -1.093892 |
| 4.843125 | 35 | | 4.0911 | .3836099 | 10.66 | 0.000 | 3.339076 |
| 3.677892 | 36 | | 2.31467 | .6953835 | 3.33 | 0.001 | .9514488 |
| 5.866744 | 37 | | 4.432265 | .7317326 | 6.06 | 0.000 | 2.997785 |
| 3.272339 | 38 | | -1.576915 | .8648412 | -1.82 | 0.068 | - |
| -.0255684 | 39 | | .1185096 -.8382542 | .4145536 | -2.02 | 0.043 | -1.65094 |
| 2.678676 | 40 | | 1.431541 | .6361677 | 2.25 | 0.024 | .1844049 |
| 1.679174 | 41 | | .3511365 | .6774362 | 0.52 | 0.604 | -.9769013 |
| 4.111989 | 42 | | 3.283597 | .4225652 | 7.77 | 0.000 | 2.455205 |
| 2.691628 | 43 | | 1.882152 | .4129162 | 4.56 | 0.000 | 1.072676 |
| 3.800887 | 44 | | 2.95755 | .4301891 | 6.87 | 0.000 | 2.114212 |
| 4.095092 | 45 | | 3.140261 | .4870626 | 6.45 | 0.000 | 2.185429 |
| 4.371411 | 46 | | 3.385047 | .5031474 | 6.73 | 0.000 | 2.398682 |
| 5.676562 | 47 | | 3.815804 | .9491785 | 4.02 | 0.000 | 1.955045 |
| 2.624577 | 48 | | 1.21444 | .7193151 | 1.69 | 0.091 | -.1956968 |
| 2.990165 | 49 | | 1.278702 | .8730228 | 1.46 | 0.143 | -.4327616 |
| 4.24386 | 50 | | 3.015756 | .6264596 | 4.81 | 0.000 | 1.787652 |
| 4.862561 | 51 | | 3.838165 | .5225472 | 7.35 | 0.000 | 2.81377 |
| 1.620077 | 52 | | .4900421 | .5764344 | 0.85 | 0.395 | -.639993 |

| | | | | | | | |
|-----------|-------|--|-----------|----------|-------|-------|-----------|
| 3.243525 | 53 | | 2.172783 | .5461891 | 3.98 | 0.000 | 1.10204 |
| 2.24946 | 54 | | 1.401862 | .4323622 | 3.24 | 0.001 | .5542643 |
| 2.881036 | 55 | | 1.302008 | .8054668 | 1.62 | 0.106 | -.2770192 |
| 3.138964 | 56 | | 2.214347 | .4716497 | 4.69 | 0.000 | 1.289731 |
| 1.477681 | 57 | | -.6714942 | .4112385 | -1.63 | 0.103 | - |
| 3.383799 | 58 | | 2.40911 | .4971918 | 4.85 | 0.000 | 1.434421 |
| 3.693032 | 59 | | -1.716201 | 1.008388 | -1.70 | 0.089 | - |
| 4.729469 | 60 | | 2.705213 | 1.032579 | 2.62 | 0.009 | .6809561 |
| 3.073973 | 61 | | 1.392444 | .8577531 | 1.62 | 0.105 | -.2890847 |
| 4.40535 | 62 | | 3.376281 | .5249315 | 6.43 | 0.000 | 2.347211 |
| 2.123135 | 63 | | 1.334339 | .4023672 | 3.32 | 0.001 | .5455436 |
| 2.591187 | 64 | | .7364347 | .9461147 | 0.78 | 0.436 | -1.118317 |
| -.5816623 | 65 | | -2.332973 | .8933488 | -2.61 | 0.009 | -4.084283 |
| 2.99587 | 66 | | 1.761967 | .6294177 | 2.80 | 0.005 | .5280643 |
| -.5096984 | 67 | | -2.073952 | .7979304 | -2.60 | 0.009 | -3.638205 |
| 5.656746 | 68 | | 3.943481 | .8739417 | 4.51 | 0.000 | 2.230216 |
| 2.89023 | 69 | | 1.912126 | .4989337 | 3.83 | 0.000 | .9340226 |
| 3.9434 | 70 | | 3.131901 | .4139479 | 7.57 | 0.000 | 2.320403 |
| 4.306959 | 71 | | 3.581128 | .3702487 | 9.67 | 0.000 | 2.855297 |
| 2.738949 | 72 | | 1.818284 | .4696342 | 3.87 | 0.000 | .8976187 |
| 2.252592 | 73 | | 1.382822 | .4436723 | 3.12 | 0.002 | .5130524 |
| 4.647712 | 74 | | 3.689496 | .4887889 | 7.55 | 0.000 | 2.73128 |
| 3.916496 | 75 | | 2.844933 | .5466079 | 5.20 | 0.000 | 1.773369 |
| 4.166558 | 76 | | 3.121991 | .5328366 | 5.86 | 0.000 | 2.077425 |
| 4.067287 | 78 | | 2.781439 | .6559146 | 4.24 | 0.000 | 1.495592 |
| 2.313026 | 79 | | 1.610773 | .3582212 | 4.50 | 0.000 | .9085207 |
| 1.292103 | 80 | | .5729106 | .3668621 | 1.56 | 0.118 | -.1462814 |
| -.0056967 | srace | | .0700143 | .0386204 | 1.81 | 0.070 | |
| | | | .1457253 | | | | |

| | | | | | | | |
|-----------|---------------|--|-----------|----------|--------|-------|-------------|
| | ssex | | -.0325505 | .0285803 | -1.14 | 0.255 | |
| -.088579 | | | .0234781 | | | | |
| | sesk | | .0065429 | .0352071 | 0.19 | 0.853 | |
| -.0624766 | | | .0755625 | | | | |
| | tracek | | -.226096 | .3385281 | -0.67 | 0.504 | |
| -.8897424 | | | .4375503 | | | | |
| | totexpk | | .0048919 | .0134125 | 0.36 | 0.715 | |
| -.0214017 | | | .0311855 | | | | |
| | hdegk | | -.1378125 | .1430075 | -0.96 | 0.335 | |
| -.4181626 | | | .1425376 | | | | |
| | | | | | | | |
| | cltypek | | | | | | |
| | small class | | -7.527766 | .1744252 | -43.16 | 0.000 | -7.869707 - |
| 7.185825 | | | | | | | |
| | regular class | | -.2536257 | .1627716 | -1.56 | 0.119 | |
| -.5727213 | | | .0654698 | | | | |
| | | | | | | | |
| | _cons | | 20.96402 | .6589286 | 31.82 | 0.000 | 19.67227 |
| 22.25578 | | | | | | | |

| | | |
|--|---------------|---|
| Instrumental variables 2SLS regression | Number of obs | = |
| 5,683 | | |
| | Wald chi2(85) | = |
| 7513.30 | | |
| | Prob > chi2 | = |
| 0.0000 | | |
| | R-squared | = |
| 0.3098 | | |
| | Root MSE | = |
| 22.501 | | |

(Std. err. adjusted for 316 clusters in classid)

| | | | Robust | | | |
|-----------|---------|-------------|-----------|-------|-------|----------------------|
| | pscorek | Coefficient | std. err. | z | P> z | [95% conf. interval] |
| | cs | -.7122104 | .1452407 | -4.90 | 0.000 | -.996877 |
| -.4275439 | | | | | | |
| | schidkn | | | | | |
| | 2 | -28.39978 | 6.347465 | -4.47 | 0.000 | -40.84058 - |
| 15.95898 | | | | | | |
| | 3 | .2223749 | 6.167717 | 0.04 | 0.971 | -11.86613 |
| 12.31088 | | | | | | |
| | 4 | -19.65124 | 5.63126 | -3.49 | 0.000 | -30.6883 - |
| 8.614168 | | | | | | |
| | 5 | -19.65929 | 8.887111 | -2.21 | 0.027 | -37.07771 - |
| 2.240874 | | | | | | |
| | 6 | -12.51921 | 5.724397 | -2.19 | 0.029 | -23.73882 - |
| 1.299594 | | | | | | |

| | | | | | | | | |
|----------|----|--|-----------|----------|-------|-------|-----------|---|
| 21.64972 | 7 | | 9.090623 | 6.40782 | 1.42 | 0.156 | -3.468473 | |
| 1.547849 | 8 | | -10.53132 | 6.162957 | -1.71 | 0.087 | -22.6105 | |
| 14.43292 | 9 | | 2.888608 | 5.890065 | 0.49 | 0.624 | -8.655706 | |
| 24.53685 | 10 | | 9.173745 | 7.838463 | 1.17 | 0.242 | -6.18936 | |
| 29.31924 | 11 | | 17.40924 | 6.076647 | 2.86 | 0.004 | 5.499228 | |
| 18.16374 | 12 | | 5.306639 | 6.559864 | 0.81 | 0.419 | -7.550458 | |
| 27.34742 | 13 | | 15.92706 | 5.826825 | 2.73 | 0.006 | 4.50669 | |
| 9.982887 | 14 | | -3.635632 | 6.948352 | -0.52 | 0.601 | -17.25415 | |
| 2.267183 | 15 | | -10.72736 | 6.629991 | -1.62 | 0.106 | -23.7219 | |
| 1.747674 | 16 | | -15.02315 | 6.773329 | -2.22 | 0.027 | -28.29864 | - |
| 7.629477 | 17 | | -6.655096 | 7.288182 | -0.91 | 0.361 | -20.93967 | |
| -.515138 | 18 | | -12.4915 | 6.1105 | -2.04 | 0.041 | -24.46786 | |
| 16.45261 | 19 | | 2.504579 | 7.116473 | 0.35 | 0.725 | -11.44345 | |
| 12.244 | 20 | | -3.372526 | 7.967759 | -0.42 | 0.672 | -18.98905 | |
| 25.1387 | 21 | | 12.69526 | 6.348809 | 2.00 | 0.046 | .2518263 | |
| 16.94621 | 22 | | 2.389842 | 7.426854 | 0.32 | 0.748 | -12.16652 | |
| 26.63052 | 23 | | 11.29846 | 7.822623 | 1.44 | 0.149 | -4.033596 | |
| 1.628059 | 24 | | -10.44664 | 6.160674 | -1.70 | 0.090 | -22.52134 | |
| 12.90301 | 25 | | -6.406429 | 9.851934 | -0.65 | 0.516 | -25.71586 | |
| 3.382351 | 26 | | -16.04177 | 6.459003 | -2.48 | 0.013 | -28.70118 | - |
| 36.56056 | 27 | | 22.08967 | 7.383244 | 2.99 | 0.003 | 7.618782 | |
| 10.46799 | 28 | | -3.504424 | 7.128915 | -0.49 | 0.623 | -17.47684 | |
| 23.31045 | 29 | | 5.26973 | 9.20462 | 0.57 | 0.567 | -12.77099 | |
| 49.84951 | 30 | | 32.75819 | 8.720219 | 3.76 | 0.000 | 15.66688 | |
| 28.65332 | 31 | | 16.12063 | 6.394347 | 2.52 | 0.012 | 3.587943 | |
| 5.174513 | 32 | | -16.80994 | 5.936553 | -2.83 | 0.005 | -28.44537 | - |
| 3.355276 | 33 | | -10.70486 | 7.173672 | -1.49 | 0.136 | -24.765 | |
| 5.449882 | 34 | | -18.21896 | 6.514958 | -2.80 | 0.005 | -30.98805 | - |
| 3.833506 | 35 | | -15.49994 | 5.952372 | -2.60 | 0.009 | -27.16638 | - |

| | | | | | | | | |
|----------|----|--|-----------|----------|-------|-------|-----------|---|
| 9.979865 | 36 | | -4.833483 | 7.55797 | -0.64 | 0.522 | -19.64683 | |
| 14.13331 | 37 | | .118197 | 7.150701 | 0.02 | 0.987 | -13.89692 | |
| 11.76362 | 38 | | -1.261611 | 6.64565 | -0.19 | 0.849 | -14.28685 | |
| 6.507303 | 39 | | -7.908499 | 7.355136 | -1.08 | 0.282 | -22.3243 | |
| 28.97412 | 40 | | 10.73855 | 9.304032 | 1.15 | 0.248 | -7.497015 | |
| 22.76267 | 41 | | 10.639 | 6.185663 | 1.72 | 0.085 | -1.48468 | |
| 3.686987 | 42 | | -7.327549 | 5.619764 | -1.30 | 0.192 | -18.34208 | |
| 2.996442 | 43 | | -11.39235 | 7.341354 | -1.55 | 0.121 | -25.78114 | |
| 31.35856 | 44 | | 13.40051 | 9.162439 | 1.46 | 0.144 | -4.557536 | |
| 14.45626 | 45 | | -25.33831 | 5.552166 | -4.56 | 0.000 | -36.22036 | - |
| 9.738261 | 46 | | -4.143715 | 7.082771 | -0.59 | 0.559 | -18.02569 | |
| 6.092404 | 47 | | -7.604828 | 6.988512 | -1.09 | 0.277 | -21.30206 | |
| 14.78752 | 48 | | 1.447366 | 6.806326 | 0.21 | 0.832 | -11.89279 | |
| 17.32235 | 49 | | 5.924641 | 5.815264 | 1.02 | 0.308 | -5.473066 | |
| 14.80551 | 50 | | 2.333032 | 6.363624 | 0.37 | 0.714 | -10.13944 | |
| 17.12045 | 51 | | 4.486874 | 6.44582 | 0.70 | 0.486 | -8.146701 | |
| 16.48083 | 52 | | 5.447348 | 5.629431 | 0.97 | 0.333 | -5.586134 | |
| 10.51865 | 53 | | -23.23771 | 6.489437 | -3.58 | 0.000 | -35.95677 | - |
| 13.87211 | 54 | | .6245316 | 6.759092 | 0.09 | 0.926 | -12.62304 | |
| 1.143668 | 55 | | -12.48478 | 6.953415 | -1.80 | 0.073 | -26.11322 | |
| 20.70366 | 56 | | -33.43903 | 6.497758 | -5.15 | 0.000 | -46.1744 | - |
| 8.844192 | 57 | | -20.24063 | 5.814614 | -3.48 | 0.000 | -31.63706 | - |
| 17.97233 | 58 | | 6.65879 | 5.772319 | 1.15 | 0.249 | -4.654748 | |
| 17.9263 | 59 | | 2.139663 | 8.054554 | 0.27 | 0.791 | -13.64697 | |
| 26.19096 | 60 | | -12.76734 | 6.848915 | -1.86 | 0.062 | - | |
| 1.970316 | 61 | | -13.0831 | 5.669891 | -2.31 | 0.021 | -24.19588 | - |
| 1.961533 | 62 | | -9.757334 | 5.979124 | -1.63 | 0.103 | -21.4762 | |
| 23.39219 | 63 | | 8.85445 | 7.417351 | 1.19 | 0.233 | -5.68329 | |
| 1.122293 | 64 | | -10.38728 | 5.872338 | -1.77 | 0.077 | -21.89685 | |

| | | | | | | | | |
|----------|---------|--|-----------|----------|-------|-------|-----------|---|
| 18.12085 | 65 | | 5.342274 | 6.519803 | 0.82 | 0.413 | -7.436306 | |
| 1.045397 | 66 | | -14.62036 | 6.926128 | -2.11 | 0.035 | -28.19532 | - |
| 6.900444 | 67 | | -5.537856 | 6.346188 | -0.87 | 0.383 | -17.97616 | |
| 28.77307 | 68 | | 15.76544 | 6.636667 | 2.38 | 0.018 | 2.757817 | |
| 20.1097 | 69 | | 9.136061 | 5.598898 | 1.63 | 0.103 | -1.837577 | |
| 9.763312 | 70 | | -3.902741 | 6.972604 | -0.56 | 0.576 | -17.56879 | |
| 3.779686 | 71 | | -15.00005 | 5.724782 | -2.62 | 0.009 | -26.22042 | - |
| 23.17815 | 72 | | 10.24744 | 6.597419 | 1.55 | 0.120 | -2.683262 | |
| 19.09604 | 73 | | 1.066103 | 9.199115 | 0.12 | 0.908 | -16.96383 | |
| 17.32945 | 74 | | .9355834 | 8.364371 | 0.11 | 0.911 | -15.45828 | |
| 10.48278 | 75 | | -2.081419 | 6.410423 | -0.32 | 0.745 | -14.64562 | |
| 6.410942 | 76 | | -5.871676 | 6.266757 | -0.94 | 0.349 | -18.1543 | |
| 7.550566 | 78 | | -19.57742 | 6.136263 | -3.19 | 0.001 | -31.60427 | - |
| 7.050925 | 79 | | -5.673903 | 6.492379 | -0.87 | 0.382 | -18.39873 | |
| 11.305 | 80 | | -3.552521 | 7.580506 | -0.47 | 0.639 | -18.41004 | |
| | | | | | | | | |
| 3.33349 | srace | | -5.404714 | 1.056766 | -5.11 | 0.000 | -7.475938 | - |
| 5.67896 | ssex | | 4.439537 | .6323706 | 7.02 | 0.000 | 3.200113 | |
| 15.11455 | sesk | | 13.59159 | .7770375 | 17.49 | 0.000 | 12.06862 | |
| 5.701163 | tracek | | 1.268832 | 2.261435 | 0.56 | 0.575 | -3.163498 | |
| 0.013 | totexpk | | .2692334 | .1083076 | 2.49 | | | |
| | | | .0569545 | .4815123 | | | | |
| 1.325605 | hdegk | | -.5234827 | .9434295 | -0.55 | 0.579 | -2.372571 | |
| 59.39485 | _cons | | 45.80676 | 6.932827 | 6.61 | 0.000 | 32.21867 | |

Instrumented: cs

Instruments: 2.schidkn 3.schidkn 4.schidkn 5.schidkn 6.schidkn 7.schidkn
 8.schidkn 9.schidkn 10.schidkn 11.schidkn 12.schidkn
 13.schidkn 14.schidkn 15.schidkn 16.schidkn 17.schidkn
 18.schidkn 19.schidkn 20.schidkn 21.schidkn 22.schidkn
 23.schidkn 24.schidkn 25.schidkn 26.schidkn 27.schidkn
 28.schidkn 29.schidkn 30.schidkn 31.schidkn 32.schidkn
 33.schidkn 34.schidkn 35.schidkn 36.schidkn 37.schidkn
 38.schidkn 39.schidkn 40.schidkn 41.schidkn 42.schidkn
 43.schidkn 44.schidkn 45.schidkn 46.schidkn 47.schidkn
 48.schidkn 49.schidkn 50.schidkn 51.schidkn 52.schidkn

53.schidkn 54.schidkn 55.schidkn 56.schidkn 57.schidkn
58.schidkn 59.schidkn 60.schidkn 61.schidkn 62.schidkn
63.schidkn 64.schidkn 65.schidkn 66.schidkn 67.schidkn
68.schidkn 69.schidkn 70.schidkn 71.schidkn 72.schidkn
73.schidkn 74.schidkn 75.schidkn 76.schidkn 78.schidkn
79.schidkn 80.schidkn srace ssex sesk tracek totexpk hdegk
1.cltypek 2.cltypek

.
.
.
. cap log close

2. The Catholic Church runs a large network of American parochial schools offering a low-cost private alternative to traditional public schools. Economists and educators have long debated the relative merits of Catholic schools and public schools. This debate motivates Evans and Schwab (1995) to estimate the effects of attending a Catholic high school on high school graduation and college attendance. Let Y_i be a dummy variable indicating college attendance and let CHS_i be a dummy variable indicating Catholic high school attendance. A linear model for the causal effect of Catholic school attendance on college-going is:

$$Y_i = \beta_0 + \beta_1 CHS_i + \beta_2 X_i + u_i, \quad (1)$$

where the controls, X_i , include variables like gender, race, family income, and parental education.

- (a) Equations like (1), with a dummy dependent variable, are called *linear probability models*. Why does this name make sense?

This name makes sense for two reasons:

- (a) It is a linear model where both the dependent variable (college attendance) and independent variable (Catholic high school attendance) are both linear as opposed to logs.
- (b) The dependent variable (college attendance) is a dummy variable, which means that the coefficient on Catholic high school attendance (β_1) is interpreted as the effect on the probability of attending college.

- (b) Evans and Schwab (1995) reports 2SLS estimates of this model using a dummy for being Catholic as an instrument for CHS_i . Call the Catholic instrument Z_i and use this notation to describe the first and second stage equations that generate the 2SLS estimates reported in the last column of Table VI in Evans and Schwab (1995). Specifically, write out the two equations, using Greek for parameters, and briefly explain which variables appear in each one and the roles they play.

First stage: $CHS_i = \pi_0 + \pi_1 Z_i + \pi_2 \mathbf{X}_i + \mu_i$

CHS_i is the dummy indicating high school attendance. The purpose of having this as the outcome variable is to use the fitted values for the second stage.

Z_i is the instrument (in this case, being Catholic). The purpose of this is to remove the omitted variable bias.

\mathbf{X}_i is the vector of exogenous variables (e.g. gender, race, family income etc) that are the covariates. These are included as the causal model of interest includes the covariates.

Second stage: $Y_i = \beta_0 + \beta_1 \hat{CHS}_i + \beta_2 \mathbf{X}_i + \epsilon_i$

Y_i is the outcome variable, which in the case of Table VII, is either a dummy indicating high school graduate or a dummy indicating college entrant.

\hat{CHS}_i is the fitted value of dummy indicating high school attendance. Given that the fitted value is a linear function of the instrument, it is uncorrelated with things that are uncorrelated to the instrument. Hence, the OVB is removed.

\mathbf{X}_i is the vector of exogenous variables (e.g. gender, race, family income etc) that are the covariates which are part of the causal model of interest.

- (c) Under what assumptions does this 2SLS procedure capture the causal effects of Catholic high school attendance on college-going? Are these plausible?

First, the instrument (being Catholic) must be correlated to the endogenous variable (Catholic high school attendance). The authors were able to ascertain this assumption, stating that "In a probit model that explains the probability a student will attend a Catholic school, the t-statistic on the CATHOLIC RELIGION variable is 36.3".

Second, being Catholic does not lead to differences in high school graduation and college attendance other than through the channel of going to a Catholic high school, or exclusion restriction. This doesn't seem very likely as perhaps Catholics have different values and upbringing that might affect their high school graduation and college attendance, such that increased self-belief as a result of believing in divine intervention. Nonetheless, the authors quoted other papers that found evidence that Catholics and non-Catholics have similar level of education.

3. This question explores the IV strategy used in Angrist & Krueger (1991) to estimate the economic returns to schooling.

In most US states, children enter kindergarten in the calendar year in which they turn 5 years old: a child who is born in January 2015 and another who is born in December 2015 will both enter kindergarten in September 2020. States also enforce compulsory attendance laws. If the state dropout age is 16, for example, students can drop once they turn 16, whether or not they've finished the school year.

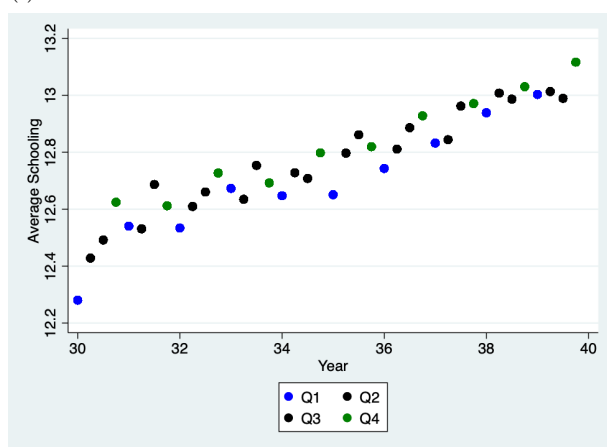
- (a) Consider young workers Sanjay and Emma, both living in North Carolina, where students enter kindergarten in the year they turn 5 and the dropout age is 16. Sanjay was born in January 2000, while Emma was born in December 2000.
- i. Which of these two children was older when they entered kindergarten, and by how much?
 - ii. Suppose Sanjay and Emma both left school in the quarter of the year they turned 16 (the first quarter is Jan-March, the second is April-June, etc). Assuming no one repeats a grade, how many years of schooling will each have completed upon leaving school (excluding their year in kindergarten)?

(i) In Jan 1 2005, Sanjay and Emma would have turn 5 in the calendar year and can enter kindergarten. Sanjay would be exactly 5 (assume born in Jan 1), while Emma would be 4 years and 1 month old (assume born in Dec 1). Hence, Sanjay would be older when she enter kindergarten, and by 11 months.

(ii) Sanjay would be able to leave school in Jan 2016, and would have had 10 years of schooling (11 years - 1 year of kindergarten). Emma would be able to leave school in Oct 2016, and would have had 10 years and 9 months of schooling (11 years 9 months - 1 year of kindergarten).

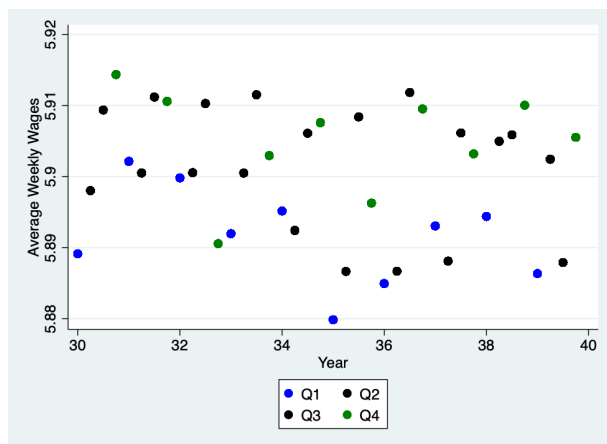
- (b) Download the AK91 data from the MM resources page under the section for Chapter 6. This file contains over 300,000 observations on men born 1930-39 in the 1980 Census. Variables include quarter-of-birth (QOB), year of birth (YOB), years of schooling, and log weekly wages.
- Construct a single variable combining year and quarter of birth (e.g., men born in the first quarter of 1930 can be coded 30.00, while those born QOB II can be coded 30.25 etc.). Plot average schooling against this variable. How many points does your plot have? Highlight cohorts born in quarter 1 in one color and cohorts born in quarter 4 in another color (consider using Stata's `graph twoway` command to accomplish this).
 - Make a similar plot for average weekly wages, again highlighting cohorts born in quarter 1 and quarter 4. In what sense do the average schooling and average wage plots move together?

(i)



The plot has 40 points. 4 points for each of the 10 years.

(ii)



In general, within the same year, the average schooling and average weekly wages of those born in Q4 is greater than those born in Q1.

```

.
.
. * Load data
.
. use ak91.dta, clear

. *****Question 3b*****
.
. * Construct variable for year & quarter
.
.     gen yqob = yob + (qob-1)*0.25

.
. * Plot average schooling against yqob
.
.     egen ms = mean(s), by(yqob)

.     graph twoway ///
>         (scatter ms yqob if qob == 1, mcolor(blue)) ///
>         (scatter ms yqob if qob == 2, mcolor(black)) ///
>         (scatter ms yqob if qob == 3, mcolor(black)) ///
>         (scatter ms yqob if qob == 4, mcolor(green)) ///
>         , legend(order(1 "Q1" 2 "Q2" 3 "Q3" 4 "Q4")) ///
>         xtitle("Year") ytitle("Average Schooling")

.
.     graph export "Q3bi_plot.png", replace
file /Users/clairchew/Documents/CW/Masters/DEDP Masters/14.320
Econometric/Problem Sets/Pset6/Q3bi_plot.png
    saved as PNG format

.
. * Plot average weekly wages against yqob
.
.     egen mlnw = mean(lnw), by(yqob)

.     graph twoway ///
>         (scatter mlnw yqob if qob == 1, mcolor(blue)) ///
>         (scatter mlnw yqob if qob == 2, mcolor(black)) ///
>         (scatter mlnw yqob if qob == 3, mcolor(black)) ///
>         (scatter mlnw yqob if qob == 4, mcolor(green)) ///
>         , legend(order(1 "Q1" 2 "Q2" 3 "Q3" 4 "Q4")) ///
>         xtitle("Year") ytitle("Average Weekly Wages")

.
.     graph export "Q3bii_plot.png", replace
file /Users/clairchew/Documents/CW/Masters/DEDP Masters/14.320
Econometric/Problem Sets/Pset6/Q3bii_plot.png
    saved as PNG format

```

4. Continue working with the Angrist & Krueger (1991) data.

- (a) Replicate MM Table 6.5, which shows OLS and IV estimates of the returns to schooling using alternative quarter-of-birth instruments, with and without year-of-birth controls. Report your replication results in a table format similar to Table 6.5.

| Table 6.5 | | | | | | |
|---|---------|-----------|---------|-----------|-------------------|---------------------------------------|
| Returns to Schooling Using Alternate Quarter of Birth Instruments | | | | | | |
| | OLS | 2SLS | OLS | 2SLS | 2SLS | 2SLS |
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Original | | | | | | |
| Years of education | .071 | .074 | .071 | .075 | .105 | |
| | (.0004) | (.028) | (.0004) | (.028) | (.020) | |
| First-stage <i>F</i> -statistic | | 48 | | 47 | 33 | |
| Replication | | | | | | |
| Years of education | .071 | .074 | .071 | .075 | .105 | .089 |
| | (.0004) | (.028) | (.0004) | (.028) | (.020) | (.016) |
| First-stage <i>F</i> -statistic | | 49 | | 48 | 32 | 5 |
| Instruments | None | Quarter 4 | None | Quarter 4 | 3 quarter dummies | 3 quarter & 27 quarter x year dummies |
| Year of birth controls | No | No | Yes | Yes | Yes | Yes |

- (b) Add to this a set of 2SLS estimates using 3 QOB plus 3 QOB * 9 YOB dummies as instruments.
- What controls are necessary here?
 - What do the extra instruments buy you?
 - How do the plots in Q3b relate to this 2SLS strategy?

Table 6.5
Returns to Schooling Using Alternate Quarter of Birth Instruments

| | | OLS | 2SLS | OLS | 2SLS | 2SLS | 2SLS |
|------------------------|---------------------------------|-----------------|----------------|-----------------|----------------|-------------------|---------------------------------------|
| | | (1) | (2) | (3) | (4) | (5) | (6) |
| Original | Years of education | .071 (.0004) | .074 (.028) | .071 (.0004) | .075 (.028) | .105 (.020) | |
| | First-stage <i>F</i> -statistic | | 48 | | 47 | 33 | |
| Replication | Years of education | .071 (.0004) | .074 (.028) | .071 (.0004) | .075 (.028) | .105 (.020) | .089 (.016) |
| | First-stage <i>F</i> -statistic | | 49 | | 48 | 32 | 5 |
| Instruments | | None | Quarter 4 | None | Quarter 4 | 3 quarter dummies | 3 quarter & 27 quarter x year dummies |
| Year of birth controls | | No | No | Yes | Yes | Yes | Yes |

- (i) The year of birth controls are necessary here so that we can examine the within-year impact, which is what we are interested in.
- (ii) The extra instruments buy us precision, as the standard error is the lowest among the rest of the 2SLS regressions.
- (ii) The plot in Q3bi is essentially the first stage of this 2SLS strategy, while the plot in Q3bii is essentially the reduced form of this 2SLS strategy.

```

. *****Question 4a*****
.
.      * Creating the table
.      matrix table1 = J(3,6,.)
.
.      matrix rownames table1 = "Years of education" "" "First-stage
F-statistic"
.
.      matrix colnames table1 = "OLS" "2SLS_Q4" "OLS_yob"
"2SLS_Q4_yob" "2SLS_3Q_yob" "2SLS_3Q*9Y_yob"
.
.      * Filling the table
.
.      quietly reg lnw s, robust
.
.      matrix table1 [1,1] = _b[s]
.
.      matrix table1 [2,1] = _se[s]
.
.      quietly ivregress 2sls lnw (s = 4.qob), robust first
.
.      matrix table1 [1,2] = _b[s]
.
.      matrix table1 [2,2] = _se[s]
.
.      quietly estat firststage
.
.      mat fstat = r(singleresults)
.
.      matrix table1 [3,2] = fstat[1,4]
.
.      quietly reg lnw s i.yob, robust
.
.      matrix table1 [1,3] = _b[s]
.
.      matrix table1 [2,3] = _se[s]
.
.      quietly ivregress 2sls lnw (s = 4.qob) i.yob, robust first
.
.      matrix table1 [1,4] = _b[s]
.
.      matrix table1 [2,4] = _se[s]
.
.      quietly estat firststage
.
.      mat fstat = r(singleresults)
.
.      matrix table1 [3,4] = fstat[1,4]
.
.      quietly ivregress 2sls lnw (s = i.qob) i.yob, robust first

```

```

.           matrix table1 [1,5] = _b[s]
.           matrix table1 [2,5] = _se[s]
.
.           quietly estat firststage
.           mat fstat = r(singleresults)
.           matrix table1 [3,5] = fstat[1,4]
.
.           quietly ivregress 2sls lnw (s = i.qob##i.yob) i.yob, robust
first
.           matrix table1 [1,6] = _b[s]
.           matrix table1 [2,6] = _se[s]
.
.           quietly estat firststage
.           mat fstat = r(singleresults)
.           matrix table1 [3,6] = fstat[1,4]
.
.           matrix list table1, format(%8.4f)
table1[3,6]

2SLS_3Q*
               OLS      2SLS_Q4      OLS_yob      2SLS_Q4_yob
2SLS_3Q_yob      9Y_yob
Years of e~n      0.0709      0.0740      0.0711      0.0752
0.1053      0.0891
               r2      0.0004      0.0280      0.0004      0.0284
0.0201      0.0162
First-stag~c      .      48.9906      .      47.7309
32.3233      4.8008

.
.           * Output Table
.           putexcel set Q4_Table1, modify
.
.           putexcel A1 = matrix(table1), names
nformat(number_d3)
file Q4_Table1.xlsx saved

.
. cap log close

```

5. Review the Ashenfelter-Rouse (1998) study summarized in LN12 and posted in Module E. As a reminder, this study uses data collected from identical twins to estimate the economic returns to schooling. The idea here is to compare education and income within pairs of twins, thereby controlling for their shared family background and similar (perhaps identical) genetic heritage.
- (a) Download data file ar98.dta from the Pset6 Canvas page. This file has data on wages and schooling for 340 twin pairs (680 total observations). Regress log wages (*lwage*) on years of schooling (*educ*), age (*age*), age-squared (*age2*), and dummies for female (*female*) and white (*white*). Interpret the estimated schooling and age coefficients.

```
. reg lwage educ age age2 female white, robust
```

```
Linear regression               Number of obs   =      680
                               F(5, 674)         =      59.28
                               Prob > F           =      0.0000
                               R-squared           =      0.3369
                               Root MSE        =      .50744
```

| lwage | Coefficient | Robust std. err. | t | P> t | [95% conf. interval] | |
|--------|-------------|---------------------|-------|-------|----------------------|-----------|
| educ | .1097706 | .0105169 | 10.44 | 0.000 | .0891209 | .1304204 |
| age | .1032049 | .0118986 | 8.67 | 0.000 | .0798422 | .1265677 |
| age2 | -.0010548 | .0001458 | -7.23 | 0.000 | -.0013411 | -.0007685 |
| female | -.3143924 | .0401221 | -7.84 | 0.000 | -.3931718 | -.235613 |
| white | -.1081378 | .0755145 | -1.43 | 0.153 | -.2564097 | .0401341 |
| _cons | -1.069035 | .2959249 | -3.61 | 0.000 | -1.65008 | -.487989 |

Schooling coefficient: An additional year of schooling increases wages by about 11 log points, holding all other things constant.

Age: Being older by a year increase your wages by $(.10 - 2 * .001 * age)$ log points, holding all other things constant. Qualitatively, wages increase as you get older, but the marginal increase decreases as you get older.

```
.
.
. * Load data
.
. use ar98.dta, clear
```

```
. *****Question 5a*****
```

```
. * Conduct regression
```

```
. reg lwage educ age age2 female white, robust
```

```
Linear regression          Number of obs    =
680                        F(5, 674)        =
59.28                      Prob > F         =
0.0000                     R-squared        =
0.3369                     Root MSE
= .50744
```

```
-----
-----
            |               Robust
            | lwage | Coefficient std. err.      t    P>|t|    [95% conf.
interval]
-----+-----
-----
educ |   .1097706   .0105169    10.44
0.000   .0891209   .1304204
age |   .1032049   .0118986     8.67
0.000   .0798422   .1265677
age2 |  -.0010548   .0001458    -7.23   0.000   -.0013411
-.0007685
female |  -.3143924   .0401221    -7.84   0.000   -.3931718
-.235613
white |  -.1081378   .0755145    -1.43   0.153
-.2564097   .0401341
_cons |  -1.069035   .2959249    -3.61   0.000   -1.65008
-.487989
-----
-----
```

(b) Consider the regression model

$$\ln Y_{if} = \alpha' X_{if} + \beta S_{if} + \gamma A_f + \epsilon_{if},$$

where f stands for family and subscript $i = 1, 2$ indexes twins in family f . The vector X_{if} includes the covariates from part (a), including a constant, while A_f is an *unobserved* ability variable assumed to be fixed within families.

- i. Explain why the regression of log wages on X_{if} and S_{if} without control for A_{if} is likely to generate a biased estimate of β .
- ii. Show (mathematically) that a regression of the within-family difference in log wages on the corresponding difference in schooling eliminates ability bias. What's the key assumption that makes this work?

(i) The exclusion of ability would likely result in omitted variable bias, since ability is likely to be positively correlated with years of schooling, and ability is also likely to be positively correlated with earnings even for those with the same amount of schooling. Hence, the exclusion of ability will likely cause an inflated schooling effect (β).

(ii) Main assumption - assume that the ability of the twins within the family is the same, we can drop the i subscript of the ability variable:

$$\ln Y_{if} = \alpha' X_{if} + \beta S_{if} + \gamma A_f + \epsilon_{if}$$

Hence, for each twin:

$$\begin{aligned}\ln Y_{1f} &= \alpha' X_{1f} + \beta S_{1f} + \gamma A_f + \epsilon_{1f} \\ \ln Y_{2f} &= \alpha' X_{2f} + \beta S_{2f} + \gamma A_f + \epsilon_{2f}\end{aligned}$$

Taking the within family difference:

$$\ln Y_{1f} - \ln Y_{2f} = \alpha' (X_{1f} - X_{2f}) + \beta (S_{1f} - S_{2f}) + (\epsilon_{1f} - \epsilon_{2f})$$

The ability term (or the ability bias) has been differenced away (proved).

- (c) Run the regression suggested by 4b(ii). Note that there are 340 unique twin differences, and that these are already computed for you (be sure to delete the redundant differences). Check your results by comparing them with those reported in the first two columns of MM Table 6.2. Do the first-differenced results align with your expectations about the direction of ability bias in the undifferenced model?

```
. reg lwage1 educ1 if twin_no > 1, robust
```

```
Linear regression          Number of obs   =       340
                          F(1, 338)        =       9.72
                          Prob > F         =    0.0020
                          R-squared        =    0.0310
                          Root MSE      =    .50773
```

| lwage1 | Coefficient | Robust std. err. | t | P> t | [95% conf. interval] | |
|--------|-------------|---------------------|------|-------|----------------------|----------|
| educ1 | .0615309 | .0197365 | 3.12 | 0.002 | .0227091 | .1003527 |
| _cons | .0220385 | .0275553 | 0.80 | 0.424 | -.032163 | .07624 |

The first-differenced results shows a much smaller return to education (reduced from .11 to .06). This coincides with my expectation that omitting ability will cause an inflated schooling effect, as explained in Q5bii.

```

.
.
. *****Question 5c*****
.
. * Set the differences
.
.     foreach var in lwage educ {
. 2.         gen `var'1 = `var' - `var'[_n-1] if twin_no > 1
. 3.         }
(340 missing values generated)
(340 missing values generated)

```

```

.
. * Run the regression
.
.     reg lwage1 educ1 if twin_no > 1, robust

```

| | | |
|-------------------|---------------|---|
| Linear regression | Number of obs | = |
| 340 | | |
| | F(1, 338) | = |
| 9.72 | | |
| | Prob > F | = |
| 0.0020 | | |
| | R-squared | = |
| 0.0310 | | |
| | Root MSE | |
| = .50773 | | |

```

-----
-----
      |
      |               Robust
lwage1 | Coefficient  std. err.      t    P>|t|    [95% conf.
interval]
-----+-----
-----
educ1 |   .0615309   .0197365     3.12
0.002   .0227091   .1003527
_cons |   .0220385   .0275553     0.80   0.424
-.032163   .07624
-----
-----

```

- (d) Using the difference in cross-sibling reports of educational attainment as an instrument for the difference in own reports of educational attainment, construct 2SLS estimates of the returns to schooling for both models, thereby completing your replication of MM Table 6.2. What do these results suggest about the relative importance of measurement error and ability bias in OLS estimates of the economic returns to schooling?

Log wages with IV:

```
.          ivregress 2sls lwage (educ = educt_own) age age2 female white, robust
```

Instrumental variables 2SLS regression

| | | |
|---------------|---|--------|
| Number of obs | = | 680 |
| Wald chi2(5) | = | 285.48 |
| Prob > chi2 | = | 0.0000 |
| R-squared | = | 0.3365 |
| Root MSE | = | .50534 |

| lwage | Coefficient | Robust std. err. | z | P> z | [95% conf. interval] | |
|--------|-------------|---------------------|-------|-------|----------------------|-----------|
| educ | .1157409 | .011401 | 10.15 | 0.000 | .0933953 | .1380865 |
| age | .1032894 | .0118698 | 8.70 | 0.000 | .080025 | .1265538 |
| age2 | -.0010536 | .0001452 | -7.25 | 0.000 | -.0013382 | -.0007689 |
| female | -.3120337 | .0399475 | -7.81 | 0.000 | -.3903295 | -.233738 |
| white | -.1049618 | .075274 | -1.39 | 0.163 | -.2524962 | .0425726 |
| _cons | -1.162293 | .3081491 | -3.77 | 0.000 | -1.766254 | -.5583318 |

Instrumented: **educ**
Instruments: **age age2 female white educt_own**

Differenced regression with IV:

```
.          ivregress 2sls lwage1 (educ1 = deduct) if twin_no > 1, robust
```

Instrumental variables 2SLS regression

| | | |
|---------------|---|--------|
| Number of obs | = | 340 |
| Wald chi2(1) | = | 9.77 |
| Prob > chi2 | = | 0.0018 |
| R-squared | = | 0.0143 |
| Root MSE | = | .51058 |

| lwage1 | Coefficient | Robust std. err. | z | P> z | [95% conf. interval] | |
|--------|-------------|---------------------|------|-------|----------------------|----------|
| educ1 | .1066801 | .0341354 | 3.13 | 0.002 | .039776 | .1735843 |
| _cons | .021286 | .0277251 | 0.77 | 0.443 | -.0330542 | .0756263 |

Instrumented: **educ1**
Instruments: **deduct**

Using the twin's report as an IV was to eliminate the measurement error (assuming that the error is random and uncorrelated with the twin's report).

With measurement error removed, we see that the return to education for the simple regression and differenced regression are very similar. Suggesting that there is no or very minimal ability bias.

When we compare between the simple regression with or without the IV, we find that the return to education increased slightly for the regression with IV. Also, when we compare between the differenced regression with or without the IV, we find that the return to education increased significantly for the regression with IV. This seems to suggest that there is measurement error in the years of schooling, which was made worse by the differenced regression.

```
. *****Question 5d*****
.
.       gen educt_own = educt[_n+1] if twin_no == 1
(340 missing values generated)
.
.       replace educt_own = educt[_n-1] if twin_no == 2
(340 real changes made)
.
.       ivregress 2sls lwage (educ = educt_own) age age2 female white,
robust
```

| | lwage | Coefficient | Robust std. err. | z | P> z | [95% conf. interval] |
|-----------|--------|-------------|---------------------|-------|-------|-------------------------|
| | educ | .1157409 | .011401 | 10.15 | | |
| 0.000 | | .0933953 | .1380865 | | | |
| | age | .1032894 | .0118698 | 8.70 | | |
| 0.000 | | .080025 | .1265538 | | | |
| | age2 | -.0010536 | .0001452 | -7.25 | 0.000 | -.0013382 |
| -.0007689 | | | | | | |
| | female | -.3120337 | .0399475 | -7.81 | 0.000 | -.3903295 |
| -.233738 | | | | | | |
| | white | -.1049618 | .075274 | -1.39 | 0.163 | |
| -.2524962 | | .0425726 | | | | |
| | _cons | -1.162293 | .3081491 | -3.77 | 0.000 | -1.766254 |
| -.5583318 | | | | | | |

```
.       ivregress 2sls lwage1 (educ1 = deduct) if twin no > 1, robust
```

```

= .51058
Root MSE

-----
-----
      |               Robust
      |               std. err.      z      P>|z|      [95% conf.
lwage1 | Coefficient
interval]
-----+-----
-----
      educ1 | .1066801 .0341354 3.13
0.002      .039776 .1735843
      _cons | .021286 .0277251 0.77 0.443
-.0330542 .0756263
-----
-----
Instrumented: educ1
Instruments: deduct

.
.
. cap log close

```

6. The MM website contains the youth mortality data used to make MM Table 5.2. Data sources and methods are given in the Empirical Notes section of the book.
- (a) Use these data to replicate results in the table for estimated MLDA effects on all-cause and motor vehicle accident (MVA) mortality with and without state-specific trends and with and without weighting by state population (hint: replication code is archived on the book website). You should limit the sample to years before 1983 (included) throughout this problem. Report these in a table formatted like the original.

Table 5.2
Regression DD estimates of MLDA effects on death rates

| Dependant variable | (1) | (2) | (3) | (4) |
|---------------------------------|-----------------|----------------|-----------------|----------------|
| All death | 10.80 (4.59) | 8.47 (5.10) | 12.41 (4.60) | 9.65 (4.64) |
| Motor vechicle accidents | 7.59 (2.50) | 6.64 (2.66) | 7.50 (2.27) | 6.46 (2.24) |
| State trends | No | Yes | No | Yes |
| Weights | No | No | Yes | Yes |

```

.
.
. * Load data
.
. use deaths.dta, clear

.
. *****Question 6a*****
.
. * Limit sample to on or before 1983 for all questions
.
.         drop if year > 1983
(11,934 observations deleted)

.
. * Construct table 5.2 in 'Metrics
. * Regression DD Estimates of MLDA-Induced Deaths among 18-20 Year Olds,
from 1970-1983
.
.
. * death cause: 1=all, 2=MVA, 3=suicide, 6=internal
. foreach i in 1 2{
.   2.
.   * no trends, no weights
.   qui xi: reg mrate legal i.state i.year if agegr == 2 & dtype == `i',
cluster(state)
.   3. outreg2 legal using "/Users/clairchew/Documents/CW/Masters/DEDP
Masters/14.320 Econometric/Problem Sets/P
> set6/table52.xls", append bdec(2) sdec(2) excel noaster cttop("`i'")
cttop(" no tr, no w")
.   4.
.   * time trends, no weights
.   qui xi: reg mrate legal i.state*year i.year if agegr == 2 & dtype ==
`i', cluster(state)
.   5. outreg2 legal using "/Users/clairchew/Documents/CW/Masters/DEDP
Masters/14.320 Econometric/Problem Sets/P
> set6/table52.xls", append bdec(2) sdec(2) excel noaster cttop("`i'")
cttop(" tr, no w")
.   6.
.   * no trends, weights
.   qui xi: reg mrate legal i.state i.year if agegr == 2 & dtype == `i'
[aw=pop], cluster(state)
.   7. outreg2 legal using "/Users/clairchew/Documents/CW/Masters/DEDP
Masters/14.320 Econometric/Problem Sets/P
> set6/table52.xls", append bdec(2) sdec(2) excel noaster cttop("`i'")
cttop(" no tr, w")
.   8.
.   * time trends, weights
.   qui xi: reg mrate legal i.state*year i.year if agegr == 2 & dtype ==
`i' [aw=pop], cluster(state)
.   9. outreg2 legal using "/Users/clairchew/Documents/CW/Masters/DEDP
Masters/14.320 Econometric/Problem Sets/P
> set6/table52.xls", append bdec(2) sdec(2) excel noaster cttop("`i'")
cttop(" tr, w")
.   10. }
/Users/clairchew/Documents/CW/Masters/DEDP Masters/14.320
Econometric/Problem Sets/Pset6/table52.xls

```

[illegible]

- (b) The posted data include mortality for 15-17 year olds and 21-24 year olds. Estimate the effects of the fraction legal ($LEGAL_{1820_{st}}$) as defined in MM on these death rates (this variable is the fraction legal among 18-20 year olds, so here you are asking how fraction legal drinking among 18-20 year olds affects mortality in other age groups). What do you expect here and how does this work out?

Table for 6b

Regression DD estimates of MLDA effects on death rates

| Dependant variable | (1) | (2) | (3) | (4) |
|---|----------------|-----------------|----------------|-----------------|
| Panel A: Impact of 15-17 Years Old | | | | |
| All death | 4.40 (2.78) | 5.67 (3.59) | 1.28 (2.08) | 2.46 (2.55) |
| Motor vechicle accidents | 1.87 (1.43) | 0.79 (1.92) | 1.19 (1.03) | 0.74 (1.33) |
| Panel A: Impact of 21-24 Years Old | | | | |
| All death | 2.13 (5.48) | -1.27 (4.40) | 7.00 (4.91) | 0.69 (5.03) |
| Motor vechicle accidents | 0.66 (3.06) | -1.68 (2.56) | 2.08 (2.28) | -0.54 (2.04) |
| State trends | No | Yes | No | Yes |
| Weights | No | No | Yes | Yes |

I would expect that the fraction of those between aged 18-20 that are above the legal drinking age to only have an impact on their own mortality, except for rare occasions where they cause the death of others e.g. through accidents or murder.

The regression outcomes provides evidence of this, where none of the coefficients are statistically significant at the 95% confidence level.

```
. *****Question 6b*****
.
. * death cause: 1=all, 2=MVA
. foreach i in 1 2{
.   2.
.     foreach j in 1 3{
.       3.
.       * no trends, no weights
.       qui xi: reg mrate legal1820 i.state i.year if agegr == `j' & dtype ==
`i', cluster(state)
.       4. outreg2 legal using "/Users/clairchew/Documents/CW/Masters/DEDP
Masters/14.320 Econometric/Problem Sets/P
> set6/table2.xls", append bdec(2) sdec(2) excel noaster cttop("`i' Age
`j'") cttop(" no tr, no w")
.       5.
.       * time trends, no weights
.       qui xi: reg mrate legal1820 i.state*year i.year if agegr == `j' & dtype
== `i', cluster(state)
.       6. outreg2 legal using "/Users/clairchew/Documents/CW/Masters/DEDP
Masters/14.320 Econometric/Problem Sets/P
> set6/table2.xls", append bdec(2) sdec(2) excel noaster cttop("`i' Age
`j'") cttop(" tr, no w")
.       7.
.       * no trends, weights
.       qui xi: reg mrate legal1820 i.state i.year if agegr == `j' & dtype ==
`i' [aw=pop], cluster(state)
.       8. outreg2 legal using "/Users/clairchew/Documents/CW/Masters/DEDP
Masters/14.320 Econometric/Problem Sets/P
> set6/table2.xls", append bdec(2) sdec(2) excel noaster cttop("`i' Age
`j'") cttop(" no tr, w")
.       9.
.       * time trends, weights
.       qui xi: reg mrate legal1820 i.state*year i.year if agegr == `j' & dtype
== `i' [aw=pop], cluster(state)
.       10. outreg2 legal using "/Users/clairchew/Documents/CW/Masters/DEDP
Masters/14.320 Econometric/Problem Sets/P
> set6/table2.xls", append bdec(2) sdec(2) excel noaster cttop("`i' Age
`j'") cttop(" tr, w")
.       11. }
.       12. }
/Users/clairchew/Documents/CW/Masters/DEDP Masters/14.320
Econometric/Problem Sets/Pset6/table2.xls
dir : seeout
/Users/clairchew/Documents/CW/Masters/DEDP Masters/14.320
Econometric/Problem Sets/Pset6/table2.xls
dir : seeout
/Users/clairchew/Documents/CW/Masters/DEDP Masters/14.320
Econometric/Problem Sets/Pset6/table2.xls
dir : seeout
/Users/clairchew/Documents/CW/Masters/DEDP Masters/14.320
Econometric/Problem Sets/Pset6/table2.xls
dir : seeout
/Users/clairchew/Documents/CW/Masters/DEDP Masters/14.320
Econometric/Problem Sets/Pset6/table2.xls
```

[illegible]

- (c) As an alternative to the fraction legal ($LEGAL_{st}$) defined in the book, code a dummy variable indicating states and years allowing any under-21s to drink. Re-estimate the Table 5.2 models with this dummy variable replacing $LEGAL_{st}$ and report these new results as additional rows below your replication results. Compare results from the two specifications: are they consistent in magnitude and direction?

| Table for 6c | | | | |
|--|-----------------|-----------------|-----------------|-----------------|
| Regression DD estimates of MLDA effects on death rates | | | | |
| Dependant variable | (1) | (2) | (3) | (4) |
| Panel A: Using Fraction Legal | | | | |
| All death | 10.80 (4.59) | 8.47 (5.10) | 12.41 (4.60) | 9.65 (4.64) |
| Motor vehicle accidents | 7.59 (2.50) | 6.64 (2.66) | 7.50 (2.27) | 6.46 (2.24) |
| Panel B: Using Legal below 21 | | | | |
| All death | 11.90 (3.29) | 11.30 (4.73) | 12.33 (3.62) | 11.94 (3.48) |
| Motor vehicle accidents | 9.46 (2.15) | 9.20 (3.09) | 7.18 (1.93) | 6.91 (1.95) |
| State trends | No | Yes | No | Yes |
| Weights | No | No | Yes | Yes |

Using the alternate dummy variable, directions of the coefficients are still positive, and they are generally bigger. The standard errors are also generally smaller.

```

.
. *****Question 6c*****
.
. * Create dummy
. gen alllegal1820 = 0

. replace alllegal1820 = 1 if legal1820 > 0
(8,514 real changes made)

.
. ** Creating the table
.
.      * death cause: 1=all, 2=MVA, 3=suicide, 6=internal
.      foreach i in 1 2{
2.
.      * no trends, no weights
.      qui xi: reg mrate alllegal1820 i.state i.year if agegr == 2 &
dtype == `i', cluster(state)
3.      outreg2 legal using
"/Users/clairchew/Documents/CW/Masters/DEDP Masters/14.320
Econometric/Proble
> m Sets/Pset6/table3.xls", append bdec(2) sdec(2) excel noaster
cttop("`i'") cttop(" no tr, no w")
4.
.      * time trends, no weights
.      qui xi: reg mrate alllegal1820 i.state*year i.year if agegr ==
2 & dtype == `i', cluster(state)
5.      outreg2 legal using
"/Users/clairchew/Documents/CW/Masters/DEDP Masters/14.320
Econometric/Proble
> m Sets/Pset6/table3.xls", append bdec(2) sdec(2) excel noaster
cttop("`i'") cttop(" tr, no w")
6.
.      * no trends, weights
.      qui xi: reg mrate alllegal1820 i.state i.year if agegr == 2 &
dtype == `i' [aw=pop], cluster(state)
7.      outreg2 legal using
"/Users/clairchew/Documents/CW/Masters/DEDP Masters/14.320
Econometric/Proble
> m Sets/Pset6/table3.xls", append bdec(2) sdec(2) excel noaster
cttop("`i'") cttop(" no tr, w")
8.
.      * time trends, weights
.      qui xi: reg mrate alllegal1820 i.state*year i.year if agegr ==
2 & dtype == `i' [aw=pop], cluster(st
> ate)
9.      outreg2 legal using
"/Users/clairchew/Documents/CW/Masters/DEDP Masters/14.320
Econometric/Proble
> m Sets/Pset6/table3.xls", append bdec(2) sdec(2) excel noaster
cttop("`i'") cttop(" tr, w")
10.      }
/Users/clairchew/Documents/CW/Masters/DEDP Masters/14.320
Econometric/Problem Sets/Pset6/table3.xls
dir : seeout
/Users/clairchew/Documents/CW/Masters/DEDP Masters/14.320
Econometric/Problem Sets/Pset6/table3.xls
dir : seeout

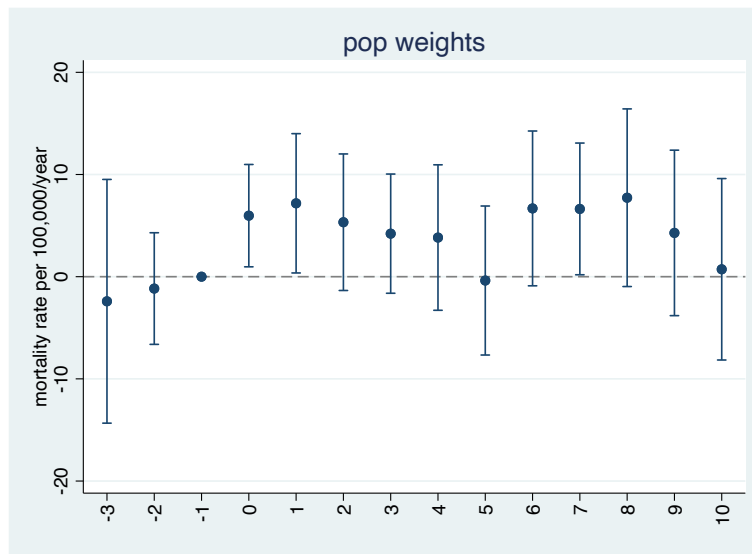
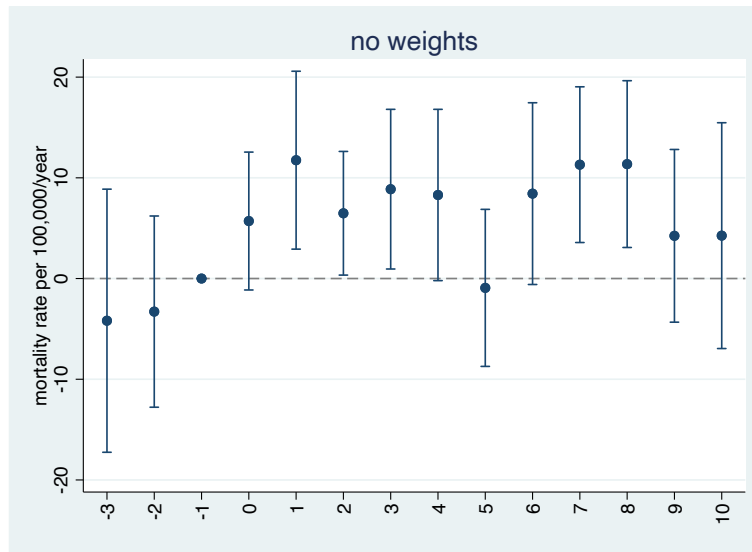
```

```

/Users/clairchew/Documents/CW/Masters/DEDP Masters/14.320
Econometric/Problem Sets/Pset6/table3.xls
dir : seeout
/Users/clairchew/Documents/CW/Masters/DEDP Masters/14.320
Econometric/Problem Sets/Pset6/table3.xls
dir : seeout
/Users/clairchew/Documents/CW/Masters/DEDP Masters/14.320
Econometric/Problem Sets/Pset6/table3.xls
dir : seeout
/Users/clairchew/Documents/CW/Masters/DEDP Masters/14.320
Econometric/Problem Sets/Pset6/table3.xls
dir : seeout
/Users/clairchew/Documents/CW/Masters/DEDP Masters/14.320
Econometric/Problem Sets/Pset6/table3.xls
dir : seeout

```

- (d) (more challenging) Set the MLDA problem up as a staggered-adoption event-study design, with 3 leads and 10 lags around the adoption date. Here adoption is defined as allowing any under-21s to drink, as in part (c), and the 3rd lead captures treatment effects 3 or more years before adoption, while the 10th lag captures treatment effects 10 or more years after adoption. Focus on age 18-20 MVA mortality in 1970-1983, and drop drop Illinois (FIPS code 17) and Michigan (FIPS code 26), yielding a sample that has only adoptions.
- Plot unweighted and population-weighted event-study estimates with confidence bands including a zero for the reference year. Do the event-study estimates show evidence of confounding trends? Does weighting matter?
 - Comment on the dynamics seen in your estimated treatment effects. Do mortality declines induced by outlawing youth drinking appear to be lasting or transitory?



(i) The event-study estimates suggest some evidence of confounding trends. Prior to the policy adoption (i.e. from the lead 3 to lead 1 years), we see an increasing trend in mortality rate, although the slope is quite gentle and the standard errors are large. Hence, this may violate the parallel trends assumption.

In this example, it seems that weighting reduces the standard errors, making the point estimates more precise. However, it does not change the general direction and trend of the estimates.

(ii) Here, adoption is defined as allowing any under-21s to drink. Hence, the estimated treatment effects are for allowing youth drinking. We see from the event-study that after the adoption, the mortality rate increases in year 0, and slowly reduce back to the level of year -1 by year 5. We noticed another jump at year 6, altho it reduced again by year 10. Hence, the effect seems to be transitory. If we can assume the same for outlawing youth drinking, it would also be transitory.

```

.
.
. *****Question 6d*****
.
. * Remove unnecessary data
. drop if state == 17 | state == 26
(504 observations deleted)

. drop if agegr != 2
(8,232 observations deleted)

. drop if dtype !=2
(3,430 observations deleted)

.
. * Inform stata that this is a panel data
. tsset state year

Panel variable: state (strongly balanced)
Time variable: year, 1970 to 1983
Delta: 1 unit

.
. * Generate years of lag / lead
. gen rt_0 = alllegal1820 - L.alllegal1820
(49 missing values generated)

. label variable rt_0 0

. gen year_rt0 = year if rt_0 == 1
(665 missing values generated)

. bys state (year) : egen year_switch = max(year_rt0)
(392 missing values generated)

. gen relative_year = year - year_switch
(392 missing values generated)

.
. * Generate dummies for which lead and lag they are
. local max_lead 3

. local max_lag 10

.
. * Lag until max lag - 1
. forvalues lag = 1/`=max_lag'-1' {
2.     gen rt_lag`lag' = (relative_year == `lag')
3.     label variable rt_lag`lag' `lag'
4.     }

.
. * Max lag
. gen rt_lag`max_lag'_plus = (relative_year >= `max_lag')

. label variable rt_lag`max_lag'_plus "`max_lag'"
.

```

```

.      * Note: lead variables are created in this descending order
because that's the order
.      * in which we want coefplot to show them later (but it's no
necessary for estimation)
.      gen rt_lead`max_lead'_minus = (relative_year <= -`max_lead')

.      label variable rt_lead`max_lead'_minus "-`max_lead'"

.      forvalues lead = `=`max_lead'-1'(-1)1 {
2.          gen rt_lead`lead' = (relative_year == -`lead')
3.          label variable rt_lead`lead' "-`lead'"
4.      }

.
. * normalize everything relative to -1
. * Note: Technically we should drop rt_lead1. Setting it equal to 0 is a
trick for plotting
. * because stata drops it (estimates a coeff of 0 and no standard
errors) but then this
. * zero without SE shows up in coefplot
. replace rt_lead1 = 0
(21 real changes made)

.
. local y mrate

. local varl: var label `y'

.
. * no trends, no weights
. reg `y' rt_lead* rt_0 rt_lag* i.state i.year, cluster(state)
note: rt_lead1 omitted because of collinearity.

```

```

Linear regression              Number of obs      =
637                           F(24, 48)
=                               Prob > F
=                               R-squared          =
0.7988                        Root MSE         =
11.08

```

(Std. err. adjusted for 49 clusters in state)

```

-----
-----
               |
           mrate | Coefficient   Robust      t      P>|t|      [95% conf.
interval]       |
-----+-----
rt_lead3_minus | -4.187978    6.496685    -0.64   0.522    -17.25044
8.874483
      rt_lead2 | -3.283729    4.722168    -0.70   0.490    -12.77828
6.210827
      rt_lead1 |           0    (omitted)

```

| | | | | | | | | |
|----------|---------------|--|-----------|----------|--------|-------|-----------|---|
| 12.55453 | rt_0 | | 5.709093 | 3.404616 | 1.68 | 0.100 | -1.136345 | |
| 20.5839 | rt_lag1 | | 11.75115 | 4.393016 | 2.67 | 0.010 | 2.918396 | |
| 12.62024 | rt_lag2 | | 6.480627 | 3.05357 | 2.12 | 0.039 | .3410129 | |
| 16.80132 | rt_lag3 | | 8.874203 | 3.942592 | 2.25 | 0.029 | .9470904 | |
| 16.80152 | rt_lag4 | | 8.2969 | 4.229816 | 1.96 | 0.056 | -.2077144 | |
| 6.869584 | rt_lag5 | | -.9277394 | 3.878041 | -0.24 | 0.812 | -8.725062 | |
| 17.45551 | rt_lag6 | | 8.431159 | 4.488309 | 1.88 | 0.066 | -.5931918 | |
| 19.03783 | rt_lag7 | | 11.30643 | 3.845257 | 2.94 | 0.005 | 3.57502 | |
| 19.64227 | rt_lag8 | | 11.36276 | 4.117862 | 2.76 | 0.008 | 3.083239 | |
| 12.81461 | rt_lag9 | | 4.240908 | 4.264178 | 0.99 | 0.325 | -4.332796 | |
| 15.46943 | rt_lag10_plus | | 4.259085 | 5.575527 | 0.76 | 0.449 | -6.951264 | |
| | state | | | | | | | |
| 8.698607 | 2 | | .3113188 | 4.171463 | 0.07 | 0.941 | -8.075969 | |
| 22.94452 | 4 | | 20.00728 | 1.460854 | 13.70 | 0.000 | 17.07003 | |
| 9.674058 | 5 | | 1.28677 | 4.171463 | 0.31 | 0.759 | -7.100518 | |
| 1.823184 | 6 | | -6.564104 | 4.171463 | -1.57 | 0.122 | -14.95139 | |
| 3.571431 | 8 | | -4.815857 | 4.171463 | -1.15 | 0.254 | -13.20315 | |
| 10.1179 | 9 | | -13.05515 | 1.460854 | -8.94 | 0.000 | -15.99239 | - |
| -10.259 | 10 | | -13.19624 | 1.460854 | -9.03 | 0.000 | -16.13349 | |
| 30.11331 | 11 | | -38.50059 | 4.171463 | -9.23 | 0.000 | -46.88788 | - |
| 3.469985 | 12 | | 1.036053 | 1.210529 | 0.86 | 0.396 | -1.397879 | |
| 3.920936 | 13 | | .9836918 | 1.460854 | 0.67 | 0.504 | -1.953553 | |
| 19.91054 | 15 | | -22.84778 | 1.460854 | -15.64 | 0.000 | -25.78502 | - |
| 27.9764 | 16 | | 19.58911 | 4.171463 | 4.70 | 0.000 | 11.20183 | |
| 2.715185 | 18 | | -5.672103 | 4.171463 | -1.36 | 0.180 | -14.05939 | |
| 7.344258 | 19 | | 4.407013 | 1.460854 | 3.02 | 0.004 | 1.469768 | |
| 8.610168 | 20 | | .2228801 | 4.171463 | 0.05 | 0.958 | -8.164408 | |
| 7.841915 | 21 | | -.5453735 | 4.171463 | -0.13 | 0.897 | -8.932662 | |
| 7.397236 | 22 | | -.9900519 | 4.171463 | -0.24 | 0.813 | -9.37734 | |

| | | | | | | | | |
|----------|----------|--|-----------|----------|--------|-------|-----------|---|
| 20.49338 | 23 | | 12.10609 | 4.171463 | 2.90 | 0.006 | 3.718801 | |
| 15.95365 | 24 | | -17.13186 | .5859908 | -29.24 | 0.000 | -18.31007 | - |
| 13.98086 | 25 | | -16.41479 | 1.210529 | -13.56 | 0.000 | -18.84872 | - |
| -.338711 | 27 | | -2.772643 | 1.210529 | -2.29 | 0.026 | -5.206575 | |
| 10.85797 | 28 | | 2.470681 | 4.171463 | 0.59 | 0.556 | -5.916607 | |
| 7.689458 | 29 | | -.6978303 | 4.171463 | -0.17 | 0.868 | -9.085118 | |
| 54.07837 | 30 | | 51.64444 | 1.210529 | 42.66 | 0.000 | 49.21051 | |
| 15.84128 | 31 | | 7.453992 | 4.171463 | 1.79 | 0.080 | -.9332958 | |
| 56.05227 | 32 | | 47.66498 | 4.171463 | 11.43 | 0.000 | 39.2777 | |
| 4.488325 | 33 | | -2.054393 | 1.210529 | -1.70 | 0.096 | - | |
| 16.90143 | .3795391 | | | | | | | |
| 49.28727 | 34 | | -19.33536 | 1.210529 | -15.97 | 0.000 | -21.76929 | - |
| 14.83926 | 35 | | 40.89998 | 4.171463 | 9.80 | 0.000 | 32.51269 | |
| 13.02679 | 36 | | -23.22655 | 4.171463 | -5.57 | 0.000 | -31.61384 | - |
| 19.97302 | 37 | | 4.6395 | 4.171463 | 1.11 | 0.272 | -3.747788 | |
| 7.971712 | 38 | | 11.58573 | 4.171463 | 2.78 | 0.008 | 3.198442 | |
| 7.695414 | 39 | | -16.359 | 4.171463 | -3.92 | 0.000 | -24.74629 | - |
| 12.35356 | 40 | | 6.559811 | .5647986 | 11.61 | 0.000 | 5.424207 | |
| 6.97927 | 41 | | 3.966269 | 4.171463 | 0.95 | 0.346 | -4.42102 | |
| 20.19556 | 42 | | -15.36656 | 4.171463 | -3.68 | 0.001 | -23.75385 | - |
| 4.771212 | 44 | | -23.13281 | 1.460854 | -15.84 | 0.000 | -26.07005 | - |
| 23.17566 | 45 | | -3.616076 | 4.171463 | -0.87 | 0.390 | -12.00336 | |
| 13.68961 | 46 | | 14.78837 | 4.171463 | 3.55 | 0.001 | 6.401083 | |
| 4.039726 | 47 | | 10.16297 | 1.753991 | 5.79 | 0.000 | 6.636337 | |
| 2.464914 | 48 | | 1.605794 | 1.210529 | 1.33 | 0.191 | -.828138 | |
| 14.57534 | 49 | | -5.922374 | 4.171463 | -1.42 | 0.162 | -14.30966 | |
| 7.922901 | 50 | | 11.04871 | 1.753991 | 6.30 | 0.000 | 7.522072 | |
| 10.00371 | 51 | | -9.101115 | .5859908 | -15.53 | 0.000 | -10.27933 | - |
| 17.63813 | 53 | | 1.616418 | 4.171463 | 0.39 | 0.700 | -6.77087 | |
| | 54 | | 9.250839 | 4.171463 | 2.22 | 0.031 | .863551 | |

| | | | | | | | | |
|-----------|------|--|-----------|----------|-------|-------|-----------|---|
| 13.25665 | 55 | | 4.869358 | 4.171463 | 1.17 | 0.249 | -3.51793 | |
| 79.20336 | 56 | | 76.76943 | 1.210529 | 63.42 | 0.000 | 74.3355 | |
| | | | | | | | | |
| | year | | | | | | | |
| 7.803971 | 1972 | | 2.133887 | 2.820047 | 0.76 | 0.453 | -3.536198 | |
| 5.244057 | 1973 | | .5748324 | 2.322264 | 0.25 | 0.806 | -4.094392 | |
| 3.248379 | 1974 | | -8.626896 | 2.675034 | -3.22 | 0.002 | -14.00541 | - |
| 6.603417 | 1975 | | -11.18414 | 2.278249 | -4.91 | 0.000 | -15.76487 | - |
| -.3887153 | 1976 | | -6.748739 | 3.163192 | -2.13 | 0.038 | -13.10876 | |
| 1.661632 | 1977 | | -4.283769 | 2.956977 | -1.45 | 0.154 | -10.22917 | |
| 2.492504 | 1978 | | -3.957673 | 3.20803 | -1.23 | 0.223 | -10.40785 | |
| 2.238743 | 1979 | | -4.424407 | 3.313953 | -1.34 | 0.188 | -11.08756 | |
| -.7603725 | 1980 | | -6.112507 | 2.661913 | -2.30 | 0.026 | -11.46464 | |
| 3.702512 | 1981 | | -10.3467 | 3.304521 | -3.13 | 0.003 | -16.99088 | - |
| 9.291665 | 1982 | | -16.10615 | 3.389222 | -4.75 | 0.000 | -22.92064 | - |
| 10.80413 | 1983 | | -18.3128 | 3.734478 | -4.90 | 0.000 | -25.82147 | - |

| | | | | | | | | |
|----------|-------|--|----------|----------|-------|-------|---------|--|
| 65.47594 | _cons | | 59.43777 | 3.003116 | 19.79 | 0.000 | 53.3996 | |
|----------|-------|--|----------|----------|-------|-------|---------|--|

```
. coefplot, keep(rt_lead* rt_0 rt_lag*) omit vertical      xlabel(,
angle(90)) ytitle("`varl'") ///
>      title("no weights") ciopts(recast(rcap)) yline(0, lcolor(gray)
lpattern(dash)) // the options in thi
> s line are not necessary but make the graph prettier
```

```
. graph export answer_6d_event_study_`y'_nt_nw.pdf, replace
>
file /Users/clairchew/Documents/CW/Masters/DEDP Masters/14.320
Econometric/Problem
Sets/Pset6/answer_6d_event_study_mrate_nt_nw.pdf saved as PDF format
```

```
.
. * no trends, weights
. reg `y' rt_lead* rt_0 rt_lag* i.state i.year[aw=pop], cluster(state)
(sum of wgt is 148,014,284)
note: rt_lead1 omitted because of collinearity.
```

```
Linear regression                                Number of obs      =
637
```

```
F(24, 48)
```

```
=
```

| | | | |
|--------|---|-----------|---|
| = | . | Prob > F | |
| 0.8042 | | R-squared | = |
| 7.0936 | | Root MSE | = |

(Std. err. adjusted for 49 clusters in state)

| ----- | | | | | | |
|---------------------|-----------|-----------|-------|------------|-----------|---|
| ----- | | | | | | |
| | Robust | | | | | |
| mrate Coefficient | std. err. | t | P> t | [95% conf. | | |
| interval] | | | | | | |
| ----- | | | | | | |
| ----- | | | | | | |
| rt_lead3_minus | -2.40797 | 5.930959 | -0.41 | 0.687 | -14.33296 | |
| 9.517022 | | | | | | |
| rt_lead2 | -1.162201 | 2.716178 | -0.43 | 0.671 | -6.623442 | |
| 4.29904 | | | | | | |
| rt_lead1 | 0 | (omitted) | | | | |
| rt_0 | 5.973717 | 2.490231 | 2.40 | 0.020 | .9667715 | |
| 10.98066 | | | | | | |
| rt_lag1 | 7.181863 | 3.389837 | 2.12 | 0.039 | .3661391 | |
| 13.99759 | | | | | | |
| rt_lag2 | 5.33144 | 3.322924 | 1.60 | 0.115 | -1.349746 | |
| 12.01263 | | | | | | |
| rt_lag3 | 4.212088 | 2.900942 | 1.45 | 0.153 | -1.620647 | |
| 10.04482 | | | | | | |
| rt_lag4 | 3.830512 | 3.540769 | 1.08 | 0.285 | -3.288681 | |
| 10.9497 | | | | | | |
| rt_lag5 | -.3746967 | 3.626509 | -0.10 | 0.918 | -7.666282 | |
| 6.916888 | | | | | | |
| rt_lag6 | 6.68289 | 3.767492 | 1.77 | 0.082 | -.8921597 | |
| 14.25794 | | | | | | |
| rt_lag7 | 6.631844 | 3.205785 | 2.07 | 0.044 | .1861809 | |
| 13.07751 | | | | | | |
| rt_lag8 | 7.729073 | 4.32362 | 1.79 | 0.080 | -.9641486 | |
| 16.42229 | | | | | | |
| rt_lag9 | 4.285245 | 4.027069 | 1.06 | 0.293 | -3.81172 | |
| 12.38221 | | | | | | |
| rt_lag10_plus | .7256365 | 4.415915 | 0.16 | 0.870 | -8.153155 | |
| 9.604428 | | | | | | |
| state | | | | | | |
| 2 | 3.544522 | 2.924275 | 1.21 | 0.231 | -2.335127 | |
| 9.424172 | | | | | | |
| 4 | 20.55172 | 1.238177 | 16.60 | 0.000 | 18.0622 | |
| 23.04124 | | | | | | |
| 5 | 3.30884 | 2.924489 | 1.13 | 0.263 | -2.57124 | |
| 9.188919 | | | | | | |
| 6 | -4.074013 | 2.926888 | -1.39 | 0.170 | -9.958915 | |
| 1.810889 | | | | | | |
| 8 | -2.226398 | 2.928869 | -0.76 | 0.451 | -8.115284 | |
| 3.662488 | | | | | | |
| 9 | -11.6851 | 1.182038 | -9.89 | 0.000 | -14.06174 | - |
| 9.308451 | | | | | | |

| | | | | | | | | |
|-----------|----|--|-----------|----------|--------|-------|-----------|---|
| 9.339143 | 10 | | -11.68938 | 1.168905 | -10.00 | 0.000 | -14.03963 | - |
| 30.78515 | 11 | | -36.72167 | 2.95256 | -12.44 | 0.000 | -42.65819 | - |
| 3.927341 | 12 | | 1.893742 | 1.011421 | 1.87 | 0.067 | -.1398566 | |
| 4.289923 | 13 | | 1.894992 | 1.191132 | 1.59 | 0.118 | -.4999388 | |
| 19.15502 | 15 | | -21.4992 | 1.165887 | -18.44 | 0.000 | -23.84337 | - |
| 27.63292 | 16 | | 21.74762 | 2.927084 | 7.43 | 0.000 | 15.86233 | |
| 2.556623 | 18 | | -3.346457 | 2.935928 | -1.14 | 0.260 | -9.249537 | |
| 7.904379 | 19 | | 5.554251 | 1.168849 | 4.75 | 0.000 | 3.204123 | |
| 8.431505 | 20 | | 2.521625 | 2.939311 | 0.86 | 0.395 | -3.388254 | |
| 7.660271 | 21 | | 1.767955 | 2.930575 | 0.60 | 0.549 | -4.124361 | |
| 7.583844 | 22 | | 1.700066 | 2.926328 | 0.58 | 0.564 | -4.183711 | |
| 20.52901 | 23 | | 14.64099 | 2.928441 | 5.00 | 0.000 | 8.752963 | |
| 15.71666 | 24 | | -16.67075 | .4745222 | -35.13 | 0.000 | -17.62484 | - |
| 13.7063 | 25 | | -15.61393 | .9487737 | -16.46 | 0.000 | -17.52157 | - |
| -.0062119 | 27 | | -1.926028 | .9548309 | -2.02 | 0.049 | -3.845844 | |
| 10.59239 | 28 | | 4.703979 | 2.928632 | 1.61 | 0.115 | -1.18443 | |
| 7.508919 | 29 | | 1.609376 | 2.93417 | 0.55 | 0.586 | -4.290168 | |
| 54.56852 | 30 | | 52.63237 | .9629561 | 54.66 | 0.000 | 50.69621 | |
| 15.73308 | 31 | | 9.827779 | 2.937035 | 3.35 | 0.002 | 3.922474 | |
| 53.72082 | 32 | | 47.88681 | 2.901573 | 16.50 | 0.000 | 42.05281 | |
| 3.318705 | 33 | | -1.30953 | .9992739 | -1.31 | 0.196 | - | |
| 16.5107 | 34 | | -18.44889 | .9639722 | -19.14 | 0.000 | -20.38709 | - |
| 48.97621 | 35 | | 43.10235 | 2.921399 | 14.75 | 0.000 | 37.22848 | |
| 14.91309 | 36 | | -20.81463 | 2.935162 | -7.09 | 0.000 | -26.71617 | - |
| 12.79077 | 37 | | 6.891642 | 2.933964 | 2.35 | 0.023 | .9925121 | |
| 19.92423 | 38 | | 14.02307 | 2.934974 | 4.78 | 0.000 | 8.121914 | |
| 8.033353 | 39 | | -13.94356 | 2.939474 | -4.74 | 0.000 | -19.85377 | - |
| 7.824748 | 40 | | 6.887276 | .4662568 | 14.77 | 0.000 | 5.949804 | |
| 12.23194 | 41 | | 6.333534 | 2.933606 | 2.16 | 0.036 | .4351225 | |

[illegible]

```

. coefplot, keep(rt_lead* rt_0 rt_lag*) omit vertical      xlabel(,
angle(90)) ylabel("\varl'") ///
>      title("pop weights") ciopts(recast(rcap)) yline(0, lcolor(gray)
lpattern(dash))

. graph export answer_6d_event_study_y'_nt_w.pdf, replace
file /Users/clairchew/Documents/CW/Masters/DEDP Masters/14.320
Econometric/Problem
    Sets/Pset6/answer_6d_event_study_mrate_nt_w.pdf saved as PDF format

.
. cap log close

```

7. (extra credit) So far, we've discussed two scenarios where IV methods may be helpful: OVB problems and possible attenuation bias from measurement error. A third type of IV application uses instrumental variables methods to estimate supply and demand elasticities in simultaneous equations models (SEMs).
- (a) Read Lecture Note 15 and the Angrist, Graddy, and Imbens (2000) paper on the Fulton Fish Market (posted in Module F on Canvas). Focus on Section 5 of the paper, which uses data on offshore weather conditions to construct 2SLS estimates of a linear equation describing the daily demand for fish. Using SEM theory, explain why weather instruments identify demand elasticities rather than supply elasticities in the market for fish.

Whether instruments are likely to cause a shift in the supply curve and not the demand curve. With a shift in the supply curve, the equilibrium price and quantity would change through a movement along the demand curve. Hence, we can obtain the demand elasticities by comparing the change in price over the change in quantity demanded.

- (b) Use the fish.dta set (used in Pset 5) to replicate the demand elasticity estimates reported at the end of LN15.¹ Present your results with a brief explanation: what is the unit of observation? What are the endogenous variables, instrumental variables, and exogenous covariates? Label the first-stage, reduced form, and 2SLS estimates in your output. How are these different types of parameter estimates related? How and why do 2SLS and OLS estimates of demand elasticities differ?

Table for 7b

| | Speed 3 as instrument | | | | | Mixed 3 & Stormy 3 as instruments | | | |
|----------|-----------------------|------------|----------|-----------|----------|-----------------------------------|----------|-----------|----------|
| | OLS | FS | RF | 2nd Stage | 2SLS | FS | RF | 2nd Stage | 2SLS |
| | (ln qty) | (ln price) | (ln qty) | (ln qty) | (ln qty) | (ln price) | (ln qty) | (ln qty) | (ln qty) |
| Ln Price | -.525 | | | -1.57 | -1.57 | | | -1.27 | -1.27 |
| | (.176) | | | (.564) | (.640) | | | (.420) | (.445) |
| Speed 3 | | .020 | -.032 | | | | | | |
| | | (.006) | (.011) | | | | | | |
| Mixed 3 | | | | | | .281 | -.262 | | |
| | | | | | | (.097) | (.178) | | |
| Stormy 3 | | | | | | .387 | -.522 | | |
| | | | | | | (.092) | (.169) | | |

(see log for full equations)

The unit of observation here is days, i.e. the number of days whereby the researcher observe the prices.

The endogenous variable is log of the sum of the quantities sold to Asian and White. For the first-stage, the endogenous variable is the log of the weighted average price sold to Asian and White.

There are two sets of instrumental variables, as seen in the table. (i) Speed 3, which is the 3 day lagged maximum wind speed; (ii) Mixed 3 & Stormy 3, where Stormy 3 is $(\text{speed3} > 18) * (\text{wave3} > 4.5)$ [wave 3 is the average max wave height of 3 and 4 day lagged], Mixed 3 is $(1 - \text{stormy3}) * (\text{speed3} > 15) * (\text{wave3} > 3)$.

The exogenous covariates are dummies for when the day is a Monday, Tuesday, Wednesday or Thursday.

The second stage is related to the first stage, as we use the fitted values of ln price from the first stage, for the second stage.

The 2SLS estimate is essentially the same as the second stage, with fitted values from the first stage. The only difference is that it provides a more accurate standard error.

Comparing the OLS and 2SLS estimates, we find that the magnitude of 2SLS estimates are a lot bigger (i.e. more negative). This is structural equations in a simultaneous equations model are not regressions, so OLS does not reliably estimate the coefficients in them.

```

.
.
. * Load data
.
. use fish.dta, clear

. *****Question 7b*****
.
. * Create the dummy variables for instruments
.
.     gen stormy3 = (speed3>18)*(wave3>4.5)
.
.     gen mixed3 = (1-stormy3)*(speed3>15)*(wave3>3)
.
.
.     gen stormy2 = (speed2>12)*(wave2>5.5)
.
.     gen mixed2 = (1-stormy2)*(speed2>10)*(wave2>3)
.
.
. * Create log price and quantity
.
.     gen lnprice = log(avgprice)
.
.     gen lnqty = log(totalqty)
.
.
. * OLS estimate
.
.     reg lnqty day1 day2 day3 day4 lnprice

```

| | Source | SS | df | MS | Number of obs | = |
|-------------|--------|------------|----|------------|---------------|---|
| 97 | | | | | | |
| -----+----- | | | | | F(5, 91) | = |
| 5.04 | | | | | | |
| Model | | 12.1722085 | 5 | 2.4344417 | Prob > F | = |
| 0.0004 | | | | | | |
| Residual | | 43.960225 | 91 | .483079395 | R-squared | = |
| 0.2168 | | | | | | |
| -----+----- | | | | | Adj R-squared | = |
| 0.1738 | | | | | | |
| Total | | 56.1324335 | 96 | .584712849 | Root MSE | |
| = .69504 | | | | | | |

| | lnqty | Coefficient | Std. err. | t | P> t | [95% conf. interval] |
|-------------|-------|-------------|-----------|-------|-------|----------------------|
| -----+----- | | | | | | |
| ----- | | | | | | |
| day1 | | -.3109272 | .2258233 | -1.38 | 0.172 | |
| -.7594975 | | .1376431 | | | | |
| day2 | | -.6827901 | .222667 | -3.07 | 0.003 | -1.125091 |
| -.2404896 | | | | | | |
| day3 | | -.5338939 | .2199374 | -2.43 | 0.017 | -.9707725 |
| -.0970152 | | | | | | |

```

      day4 |   .0672273   .2204205    0.30   0.761
-.370611   .5050656
      lnprice |  -.5246553   .1761115   -2.98   0.004   -.8744792
-.1748314
      _cons |    8.244317   .1628134   50.64   0.000    7.920909
8.567726

```

```

.
. * Speed 3 as instrument
.
. * First stage
. reg lnprice day1 day2 day3 day4 speed3

```

```

      Source |           SS           df           MS       Number of obs   =
97
-----+-----
2.17
      Model |    1.6717106             5    .33434212       Prob > F           =
0.0646
      Residual |   14.0413996           91    .154301095       R-squared           =
0.1064
-----+-----
0.0573
      Total |   15.7131102           96    .163678231       Root MSE           =
= .39281

```

```

      lnprice | Coefficient   Std. err.      t    P>|t|     [95% conf.
interval]
-----+-----

```

```

      day1 |   -.020083   .1280266    -0.16   0.876
-.2743922   .2342262
      day2 |  -.1004775   .1293114    -0.78   0.439
-.3573387   .1563837
      day3 |  -.0038505   .1252044    -0.03   0.976
-.2525537   .2448527
      day4 |   .1026251   .1242444     0.83   0.411
-.1441711   .3494214
      speed3 |   .0201874   .0064022     3.15
0.002   .0074701   .0329046
      _cons |  -.6651257   .1516013   -4.39   0.000    -1.966263
-.3639884

```

```

. predict phat1, xb
.
. * Reduced form
. reg lnqty day1 day2 day3 day4 speed3

```

```

      Source |           SS           df           MS       Number of obs   =
97

```

```

-----+-----
4.77                                F(5, 91)                                =
      Model | 11.6510004                5  2.33020008  Prob > F                                =
0.0006
      Residual | 44.4814331                91  .488806957  R-squared                                =
0.2076
-----+-----                                Adj R-squared                                =
0.1640
      Total | 56.1324335                96  .584712849  Root MSE
= .69915

```

```

-----
-----
      lnqty | Coefficient  Std. err.      t    P>|t|    [95% conf.
interval]
-----+-----
-----
      day1 | -.2669628    .2278686    -1.17   0.244
- .7195958    .1856702
      day2 | -.5323006    .2301553    -2.31   0.023   - .9894759
- .0751253
      day3 | -.4803295    .2228455    -2.16   0.034   - .9229846
- .0376743
      day4 | .0049691     .2211368     0.02   0.982
- .434292     .4442301
      speed3 | -.0316299    .0113951    -2.78   0.007   - .0542647
- .008995
      _cons | 8.999321     .269828     33.35   0.000    8.463341
9.535302
-----
-----

```

```

.
.      * Second stage
.      reg lnqty phat1 day1 day2 day3 day4

```

```

      Source |          SS            df          MS       Number of obs   =
97
-----+-----                                F(5, 91)                                =
4.77
      Model | 11.6510004                5  2.33020007  Prob > F                                =
0.0006
      Residual | 44.4814331                91  .488806957  R-squared                                =
0.2076
-----+-----                                Adj R-squared                                =
0.1640
      Total | 56.1324335                96  .584712849  Root MSE
= .69915

```

```

-----
-----
      lnqty | Coefficient  Std. err.      t    P>|t|    [95% conf.
interval]
-----+-----
-----
      phat1 | -1.566815     .5644648    -2.78   0.007   -2.688055
- .445575

```

```

      day1 |  -.2984291   .227249   -1.31   0.192
-1.7498314   .1529731
      day2 |  -.6897303   .2240115   -3.08   0.003   -1.134702
-1.2447589
      day3 |  -.4863624   .2225836   -2.19   0.031   -.9284975
-1.0442274
      day4 |   .1657637   .2274403    0.73   0.468
-1.2860185   .6175459
      _cons |   7.957192   .2205117   36.09   0.000    7.519173
8.395212

```

```

.
.      * 2SLS
.      ivregress 2sls lnqty day1 day2 day3 day4 (lnprice=speed3)

Instrumental variables 2SLS regression          Number of obs   =
97                                              Wald chi2(5)       =
18.56                                         Prob > chi2        =
0.0023                                       R-squared
=                                           Root MSE
= .79221

```

```

-----
-----
      lnqty | Coefficient  Std. err.      z    P>|z|    [95% conf.
interval]
-----+-----
-----
      lnprice |  -1.566815   .6395994   -2.45   0.014   -2.820407
-1.3132232
      day1 |  -.2984291   .2574976   -1.16   0.246
-1.8031152   .2062569
      day2 |  -.6897303   .2538292   -2.72   0.007   -1.187226
-1.1922342
      day3 |  -.4863624   .2522112   -1.93   0.054
-1.9806874   .0079625
      day4 |   .1657637   .2577143    0.64   0.520
-1.3393472   .6708745
      _cons |   7.957192   .2498635   31.85   0.000    7.467469
8.446916

```

```

-----
-----
Instrumented: lnprice
Instruments: day1 day2 day3 day4 speed3

```

```

.
. * Mixed 3 and Stormy 3 as instruments
.
.      * First stage
.      reg lnprice day1 day2 day3 day4 mixed3 stormy3

```

```

Source |      SS      df      MS      Number of obs      =
-----+-----
3.36
Model |  2.87271871      6  .478786451  Prob > F      =
0.0050
Residual | 12.8403915      90  .142671017  R-squared      =
0.1828
-----+-----
0.1283
Total | 15.7131102      96  .163678231  Root MSE
= .37772

```

```

-----
-----
lnprice | Coefficient Std. err.      t      P>|t|      [95% conf.
interval]
-----+-----
day1 |  -.0094215   .1245151   -0.08   0.940
-.2567924   .2379495
day2 |  -.0177709   .1210333   -0.15   0.884
-.2582247   .2226828
day3 |   .0368483   .1197558    0.31   0.759
-.2010675   .2747642
day4 |   .1245099   .1201232    1.04   0.303
-.1141358   .3631555
mixed3 |  .2812021   .0965125    2.91
0.005   .0894632   .4729409
stormy3 |  .3871992   .0915453    4.23
0.000   .2053286   .5690699
_cons |  -.4866297   .0973803   -5.00   0.000   -.6800926
-.2931668
-----
-----

```

```

.      predict phat2, xb

.
.      * Reduced form
.      reg lnqty day1 day2 day3 day4 mixed3 stormy3

```

```

Source |      SS      df      MS      Number of obs      =
-----+-----
4.31
Model | 12.5323209      6  2.08872015  Prob > F      =
0.0007
Residual | 43.6001126      90  .484445695  R-squared      =
0.2233
-----+-----
0.1715
Total | 56.1324335      96  .584712849  Root MSE
= .69602

```

| lnqty | Coefficient | Std. err. | t | P> t | [95% conf. interval] |
|---------|-------------|-----------|-------|-------|----------------------|
| day1 | -.3171298 | .229444 | -1.38 | 0.170 | |
| day2 | -.6655481 | .2230281 | -2.98 | 0.004 | -1.108632 |
| day3 | -.5577861 | .2206741 | -2.53 | 0.013 | -.9961936 |
| day4 | -.0345707 | .221351 | -0.16 | 0.876 | |
| mixed3 | -.2615067 | .1778435 | -1.47 | 0.145 | |
| stormy3 | -.5222825 | .1686905 | -3.10 | 0.003 | -.8574155 |
| _cons | 8.650079 | .1794426 | 48.21 | 0.000 | 8.293585 |

```

.
.      * Second stage
.      reg lnqty phat2 day1 day2 day3 day4

```

| Source | SS | df | MS | Number of obs | = |
|----------|------------|----|------------|---------------|---|
| Model | 12.2836698 | 5 | 2.45673397 | Prob > F | = |
| Residual | 43.8487636 | 91 | .481854546 | R-squared | = |
| Total | 56.1324335 | 96 | .584712849 | Root MSE | = |

| lnqty | Coefficient | Std. err. | t | P> t | [95% conf. interval] |
|-------|-------------|-----------|-------|-------|----------------------|
| phat2 | -1.268173 | .419728 | -3.02 | 0.003 | -2.101911 |
| day1 | -.3020106 | .2255832 | -1.34 | 0.184 | |
| day2 | -.6877415 | .222399 | -3.09 | 0.003 | -1.12951 |
| day3 | -.499983 | .220345 | -2.27 | 0.026 | -.9376713 |
| day4 | .1375271 | .2230704 | 0.62 | 0.539 | |
| _cons | 8.039471 | .1935591 | 41.53 | 0.000 | 7.65499 |

```
.
.      * 2SLS
.      ivregress 2sls lnqty day1 day2 day3 day4 (lnprice=mixed3
stormy3)
```

```
Instrumental variables 2SLS regression          Number of obs   =
97                                              Wald chi2(5)      =
22.67                                         Prob > chi2       =
0.0004                                       R-squared         =
0.0635                                       Root MSE         =
=      .73618
```

```
-----
-----
      lnqty | Coefficient  Std. err.      z    P>|z|      [95% conf.
interval]
-----+-----
-----
      lnprice |   -1.268173   .4451392    -2.85   0.004    -2.14063
      -.3957166
      day1    |   -.3020106   .2392405    -1.26   0.207
      -.7709133   .1668921
      day2    |   -.6877415   .2358635    -2.92   0.004    -1.150025
      -.2254575
      day3    |   -.4999831   .2336852    -2.14   0.032    -.9579976
      -.0419686
      day4    |    .1375271   .2365755     0.58   0.561
      -.3261524   .6012066
      _cons   |    8.039471   .2052776    39.16   0.000     7.637134
      8.441808
-----
```

```
-----
Instrumented: lnprice
Instruments: day1 day2 day3 day4 mixed3 stormy3
```

- (c) Use Stata's `reshape` command to change the data structure from time series of prices and quantities for Asians and whites to panel data that stacks the ethnic groups in a format where ethnicity is identified by a dummy variable rather than by distinct variables for Asians and Whites. Use this stacked data set to compute separate demand elasticities for Asians and Whites and test whether they differ. Discuss your results in light of the fact that Asian buyers appear to pay less than others for the same fish. (Stata hint: this requires a 2SLS specification that includes ethnicity interactions in both the first and second stages.)

2SLS using `speed3` as instrument:

```
. * 2SLS using speed3
. ivregress 2sls lnqty_ day1##asian day2##asian day3##asian day4##asian (lnpri
> ed3_a)
```

Instrumental variables 2SLS regression

| | | |
|---------------|---|--------|
| Number of obs | = | 194 |
| Wald chi2(11) | = | 55.13 |
| Prob > chi2 | = | 0.0000 |
| R-squared | = | 0.0397 |
| Root MSE | = | .90031 |

| lnqty_ | Coefficient | Std. err. | z | P> z | [95% conf. interval] |
|------------|-------------|-----------|-------|-------|----------------------|
| lnprice_ | -1.029889 | .6432916 | -1.60 | 0.109 | -2.290718 .2309389 |
| lnprice_a | -.8635829 | 1.009463 | -0.86 | 0.392 | -2.842095 1.114929 |
| 1.day1 | .1358369 | .2937555 | 0.46 | 0.644 | -.4399133 .7115871 |
| 1.asian | .4488928 | .39016 | 1.15 | 0.250 | -.3158068 1.213592 |
| day1#asian | | | | | |
| 1 1 | -.7761388 | .4164982 | -1.86 | 0.062 | -1.59246 .0401826 |
| 1.day2 | -.6979756 | .2885945 | -2.42 | 0.016 | -1.26361 -.1323407 |
| day2#asian | | | | | |
| 1 1 | -.1414003 | .4085262 | -0.35 | 0.729 | -.9420969 .6592963 |
| 1.day3 | -.5011328 | .2910387 | -1.72 | 0.085 | -1.071558 .0692927 |
| day3#asian | | | | | |
| 1 1 | .0243863 | .4071425 | 0.06 | 0.952 | -.7735983 .8223709 |
| 1.day4 | .5185477 | .2911541 | 1.78 | 0.075 | -.0521039 1.089199 |
| day4#asian | | | | | |
| 1 1 | -.5777739 | .4101497 | -1.41 | 0.159 | -1.381652 .2261047 |
| _cons | 6.941718 | .2497298 | 27.80 | 0.000 | 6.452257 7.43118 |

Instrumented: lnprice_ lnprice_a
Instruments: 1.day1 1.asian 1.day1#1.asian 1.day2 1.day2#1.asian 1.day3
1.day3#1.asian 1.day4 1.day4#1.asian speed3 speed3_a

2SLS using mixed3 and stormy3 as instruments:

```
. * 2SLS using mixed3 and stormy3
.       ivregress 2sls lnqty_ day1##asian day2##asian day3##asian day4##asian (lnpri
> rmy3 mixed3_a stormy3_a)
```

```
Instrumental variables 2SLS regression           Number of obs   =       194
                                                Wald chi2(11)    =       63.20
                                                Prob > chi2      =       0.0000
                                                R-squared        =       0.1223
                                                Root MSE        =       .86074
```

| lnqty_ | Coefficient | Std. err. | z | P> z | [95% conf. interval] | |
|------------|-------------|-----------|-------|-------|----------------------|-----------|
| lnprice_ | -1.007577 | .5331203 | -1.89 | 0.059 | -2.052473 | .0373199 |
| lnprice_a | -.5574151 | .7545696 | -0.74 | 0.460 | -2.036344 | .9215141 |
| 1.day1 | .1348981 | .2805454 | 0.48 | 0.631 | -.4149607 | .684757 |
| 1.asian | .5375505 | .3351546 | 1.60 | 0.109 | -.1193404 | 1.194441 |
| day1#asian | | | | | | |
| 1 1 | -.7582123 | .3970781 | -1.91 | 0.056 | -1.536471 | .0200465 |
| 1.day2 | -.6976338 | .2758678 | -2.53 | 0.011 | -1.238325 | -.1569429 |
| day2#asian | | | | | | |
| 1 1 | -.1331186 | .3903023 | -0.34 | 0.733 | -.8980971 | .6318599 |
| 1.day3 | -.5032274 | .2767515 | -1.82 | 0.069 | -1.04565 | .0391956 |
| day3#asian | | | | | | |
| 1 1 | .0255568 | .3881757 | 0.07 | 0.948 | -.7352536 | .7863672 |
| 1.day4 | .5164339 | .2768349 | 1.87 | 0.062 | -.0261525 | 1.05902 |
| day4#asian | | | | | | |
| 1 1 | -.5963256 | .3896831 | -1.53 | 0.126 | -1.36009 | .1674392 |
| _cons | 6.946844 | .2281248 | 30.45 | 0.000 | 6.499728 | 7.39396 |

Instrumented: lnprice_ lnprice_a

Instruments: 1.day1 1.asian 1.day1#1.asian 1.day2 1.day2#1.asian 1.day3
1.day3#1.asian 1.day4 1.day4#1.asian mixed3 stormy3 mixed3_a
stormy3_a

To test whether the elasticities differs, I will look at the coefficient for $\ln price_a$.

Whether using $speed3$, or $mixed3$ and $stormy3$, as instruments, the magnitude of the coefficients are smaller than the standard error. The magnitudes of the z-scores are less than .9, which the p-values are higher than .39. Hence, we are unable to conclude that the elasticities for White and Asian are statistically different.

From the two sets of results we see that the coefficients for $\ln price$ are around 1, i.e. the percentage change in quantity that they buy is the same as the percentage change in prices. However, the elasticity for Asians are more negative (by .56 to .86), although the standard errors are large. Nonetheless, the magnitude is suggestive that Asians are more price elastic, i.e. a one percentage reduction in prices would increase the quantity that they buy by more than one percent (around 1.6 to 1.9%). Hence, the fishmonger would reduce the price for Asians to increase the revenue.

```

.
.
. *****Question 7c*****
.
. * Reshape data
.
.     * Generate unique id
.     gen t = _n
.
.
.     * Resahpe to panel format
.     reshape long price_ qty_, i(t) j(race) string
(j = a w)

```

| Data | Wide | -> | Long |
|------------------------|-----------------|----|--------|
| Number of observations | 97 | -> | 194 |
| Number of variables | 23 | -> | 22 |
| j variable (2 values) | | -> | race |
| xij variables: | | | |
| | price_a price_w | -> | price_ |
| | qty_a qty_w | -> | qty_ |

```

.
.
. * Generate dummies for Asian
.     gen asian = 0
.
.     replace asian = 1 if race == "a"
(97 real changes made)
.
.
. * Create log price and quantity
.     gen lnprice_ = log(price_)
.
.     gen lnqty_ = log(qty_)
.
.
. * Generate interaction terms for Asian
.     gen lnprice_a = lnprice_ * asian
.
.     gen lnqty_a = lnqty_ * asian
.
.     gen speed3_a = speed3 * asian
.
.     gen mixed3_a = mixed3 * asian
.
.     gen stormy3_a = stormy3 * asian
.
.
.
. * 2SLS using speed3
.     ivregress 2sls lnqty_ day1##asian day2##asian day3##asian
day4##asian (lnprice_ lnprice_a=speed3 spe
> ed3_a)

```

| | | |
|--|---------------|---|
| Instrumental variables 2SLS regression | Number of obs | = |
| 194 | | |
| | Wald chi2(11) | = |
| 55.13 | | |
| | Prob > chi2 | = |
| 0.0000 | | |
| | R-squared | = |
| 0.0397 | | |
| | Root MSE | |
| = | .90031 | |

```

-----
-----
      lnqty_ | Coefficient   Std. err.      z    P>|z|      [95% conf.
interval]
-----+-----
-----
      lnprice_ |   -1.029889   .6432916    -1.60   0.109    -
2.290718      .2309389
      lnprice_a |   -.8635829   1.009463    -0.86   0.392    -2.842095
1.114929
      1.day1 |    .1358369   .2937555     0.46   0.644
-.4399133      .7115871
      1.asian |    .4488928    .39016     1.15   0.250    -.3158068
1.213592
      |
      day1#asian |
      1 1 |   -.7761388   .4164982    -1.86   0.062    -
1.59246      .0401826
      |
      1.day2 |   -.6979756   .2885945    -2.42   0.016    -1.26361
-.1323407
      |
      day2#asian |
      1 1 |   -.1414003   .4085262    -0.35   0.729
-.9420969      .6592963
      |
      1.day3 |   -.5011328   .2910387    -1.72   0.085    -
1.071558      .0692927
      |
      day3#asian |
      1 1 |    .0243863   .4071425     0.06   0.952
-.7735983      .8223709
      |
      1.day4 |    .5185477   .2911541     1.78   0.075    -.0521039
1.089199
      |
      day4#asian |
      1 1 |   -.5777739   .4101497    -1.41   0.159    -
1.381652      .2261047
      |
      _cons |    6.941718   .2497298    27.80   0.000    6.452257
7.43118
-----
-----

```

```

Instrumented: lnprice_ lnprice_a
Instruments: 1.day1 1.asian 1.day1#1.asian 1.day2 1.day2#1.asian 1.day3
              1.day3#1.asian 1.day4 1.day4#1.asian speed3 speed3_a

```

```
.      test lnprice_ lnprice_a
```

```
( 1)  lnprice_ = 0
( 2)  lnprice_a = 0
```

```
      chi2( 2) =      8.49
Prob > chi2 =      0.0144
```

```
.
. * 2SLS using mixed3 and stormy3
.      ivregress 2sls lnqty_ day1##asian day2##asian day3##asian
day4##asian (lnprice_ lnprice_a=mixed3 sto
> rmy3 mixed3_a stormy3_a)
```

```
Instrumental variables 2SLS regression      Number of obs   =
194                                           Wald chi2(11)      =
63.20                                         Prob > chi2        =
0.0000                                       R-squared          =
0.1223                                       Root MSE
=      .86074
```

```
-----
-----
      lnqty_ | Coefficient  Std. err.      z    P>|z|      [95% conf.
interval]
-----+-----
-----
      lnprice_ | -1.007577   .5331203    -1.89   0.059   -
2.052473   .0373199
      lnprice_a | -.5574151   .7545696    -0.74   0.460   -
2.036344   .9215141
      1.day1 | .1348981    .2805454     0.48   0.631
-.4149607   .684757
      1.asian | .5375505    .3351546     1.60   0.109   -.1193404
1.194441
      |
      day1#asian |
      1 1 | -.7582123   .3970781    -1.91   0.056   -
1.536471   .0200465
      |
      1.day2 | -.6976338   .2758678    -2.53   0.011   -1.238325
-.1569429
      |
      day2#asian |
      1 1 | -.1331186   .3903023    -0.34   0.733
-.8980971   .6318599
      |
      1.day3 | -.5032274   .2767515    -1.82   0.069   -
1.04565    .0391956
      |
      day3#asian |
      1 1 | .0255568    .3881757     0.07   0.948
-.7352536   .7863672
```

| | | | | | | |
|------------|--|-----------|----------|-------|-------|-----------|
| 1.day4 | | .5164339 | .2768349 | 1.87 | 0.062 | -.0261525 |
| 1.05902 | | | | | | |
| day4#asian | | | | | | |
| 1 1 | | -.5963256 | .3896831 | -1.53 | 0.126 | - |
| 1.36009 | | .1674392 | | | | |
| _cons | | 6.946844 | .2281248 | 30.45 | 0.000 | 6.499728 |
| 7.39396 | | | | | | |

Instrumented: lnprice_ lnprice_a
Instruments: 1.day1 1.asian 1.day1#1.asian 1.day2 1.day2#1.asian 1.day3
1.day3#1.asian 1.day4 1.day4#1.asian mixed3 stormy3
mixed3_a
stormy3_a

. test lnprice_ lnprice_a

(1) lnprice_ = 0
(2) lnprice_a = 0

chi2(2) = 12.16
Prob > chi2 = 0.0023

.
.
. cap log close