Due date: February 22, 2018 at 5:00 PM in Bocar Ba's cubicle N2 (in front of office 148).

Group work: You may work in groups, but each person must submit individual answers. These answers must reflect the individual's own work and may not be copied from others.

Scratch work and code: Please show your work (where relevant) and append all code written for this assignment to the end of your submission. Please use brief but clear comments in the code to reference the applicable assignment section.

1. **True/False/Explain (14 points, 2 points each)**

   Be sure to explain your answer.

   (a) In a first differences regression, the residuals ($\widehat{\Delta u_{it}}$) will be uncorrelated with the regressors (ΔX_{it}).

   (b) A researcher is examining the impact of a large-scale, rural electrification program on economic activity. This electrification project employed an arbitrary population cutoff for determining eligibility: only villages with a population above 1000 were eligible for electrification. The researcher realizes there is substantial bunching in the data right above the 1000 person threshold; too many villages appear to be barely eligible for the program. This is a sign of manipulation of the running variable and could invalidate the RD design.

   (c) A researcher is interested in whether per-pupil spending on education improves student test scores, but she is worried about omitted-variable bias. She decides to use median income within a school district as an instrument for education spending because wealthier areas tend to invest more in education. She runs a regression of per-pupil education spending on median income in the district and finds that the t-statistic on the slope coefficient is 8. We can therefore conclude that median income is a good instrument for education spending.

   (d) First differencing by subtracting data for $t-1$ from t removes time-varying unobservables. This means that time-invariant unobservables could still cause bias.

   (e) Because the IV estimator uses only a subset of the variation in an explanatory variable, precision always increases when compared to OLS.

   (f) When using a regression discontinuity design, expanding the sample by increasing the bandwidth above and below the cutoff both improves precision and decreases bias.

   (g) When including fixed effect for entities (cross section units), we cannot estimate the coefficient of any regressor that does not vary over time within a cross-sectional unit.

2. **Guns and crime (12 points, 2 points each)**

School shootings have reignited the debate about gun control. Proponents of strict gun control argue that if many people are allowed to carry guns crime will increase, while opponents of strict gun control argue that crime will decrease if citizens are allowed to carry guns because criminals are deterred from attacking other people.

Some U.S. states have enacted laws that allow citizens to carry concealed weapons. These laws are known as shall-carry laws. In this exercise we will use various panel data methods to estimate the effect of shall-carry laws (`shall` = 1 if the state has a shall-carry law in effect in that year, 0 otherwise) on the violent crime rate (`vio` is violent incidents per 100,000 members of the population). We have a panel from 50 U.S. states plus the district of Columbia for the years 1990 through 1999.

(a) A researcher estimates the effect of having a shall-carry law in place on the violent crime rate by OLS. The regression output is given below.

```
. regr vio shall, vce(cluster stateid)

Linear regression                               Number of obs     =        510
                                                F(1, 50)          =       6.80
                                                Prob > F          =     0.0120
                                                R-squared         =     0.0903
                                                Root MSE          =     357.23

                                (Std. Err. adjusted for 51 clusters in stateid)
------------------------------------------------------------------------------
             |               Robust
         vio |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       shall |  -227.8783   87.36687    -2.61   0.012    -403.3598   -52.39674
       _cons |   652.6966   70.66633     9.24   0.000     510.7591    794.6341
------------------------------------------------------------------------------
```

Interpret the sign, magnitude and statistical significance of the estimated coefficient.

(b) What assumption(s) should hold for OLS to give a consistent estimate of the causal effect of a shall-carry law on the violent crime rates?

(c) The researcher is worried that the specification above contains omitted variable bias and modifies the specification by including state (entity) fixed effects.

```
. areg vio shall, absorb(stateid) vce(cluster stateid)

Linear regression, absorbing indicators         Number of obs     =        510
                                                F(   1,      50)  =      11.65
                                                Prob > F          =     0.0013
                                                R-squared         =     0.9388
                                                Adj R-squared     =     0.9320
                                                Root MSE          =    97.5947

                              (Std. Err. adjusted for 51 clusters in stateid)
------------------------------------------------------------------------------
             |               Robust
         vio |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       shall |  -49.48275   14.49558    -3.41   0.001    -78.59798   -20.36752
       _cons |   578.5401   6.025614    96.01   0.000     566.4373    590.6429
-------------+----------------------------------------------------------------
     stateid |   absorbed                                      (51 categories)
```

Interpret the sign, magnitude and statistical significance of the estimated coefficient.

(d) Suggest a possible reason why the conditional mean zero assumption might be violated.

(e) Lastly, the researcher estimates the effect of having a shall-carry law on the violent crime rate while including state (entity) fixed effects and year (time) fixed effects.

```
. areg vio shall i.year, absorb(stateid) vce(cluster stateid)

Linear regression, absorbing indicators         Number of obs     =        510
                                                F(  10,      50)  =       7.53
                                                Prob > F          =     0.0000
                                                R-squared         =     0.9558
                                                Adj R-squared     =     0.9499
                                                Root MSE          =    83.7600

                              (Std. Err. adjusted for 51 clusters in stateid)
------------------------------------------------------------------------------
             |               Robust
         vio |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       shall |   33.80662   23.33174     1.45   0.154    -13.05655    80.66979
             |
        year |
         91  |   21.40157   5.618125     3.81   0.000     10.11723    32.68591
         ...
         ...
         99  |  -124.5924   31.31439    -3.98   0.000    -187.4892   -61.69559
             |
       _cons |   563.8592   9.310376    60.56   0.000     545.1588    582.5597
-------------+----------------------------------------------------------------
     stateid |   absorbed                                      (51 categories)
```

What do you conclude about the effect of having a shall-carry law in place on violent crime rates?

(f) Which set of regression results do you think is most credible and why.

3. **Final Exam Prep: Class size and test scores [BONUS POINTS: 16 points, 4 each.]**

[Bonus points assigned for answers on these questions will be added the total problem set component of your overall grade.]

One particularly-creative scheme for estimating class size effects appears in Angrist and Lavy's 1999 paper in the Quarterly Journal of Economics titled: "Using Maimonides' rule to estimate the effect of class size on academic achievement"

Public schools in Israel generally follow a class size rule proposed by Moses Maimonides, a 12th-century rabbi and philosopher. Maimonides' rule is that a class should have no more than 40 students. If followed exactly, this leads to a systematic but discontinuous relationship between the number of students in a school and the average number of students in a class.

Below you will replicate some of the results in this paper. You can download the data from http://econ-www.mit.edu/files/1359

(a) Estimate an OLS regression of reading test scores (`avgverb`) on class size (`c_size`), first without and then with the percent of students that are disadvantaged (`tipuach`) as a control variable. Interpret your class size effect, do not forget to discuss the size of the effect. Do you think you can interpret your results as a causal effect? Explain. Can you say something about the omitted variable bias?

(b) Generate the class size predicted by Maimonides' rule using school size (c_size)

- g pclassize = c_size/(int((c_size-1)/40)+1) (Stata)
- pclassize = c_size/(floor((c_size - 1)/40) + 1) (R)

and draw a scatter plot of pclassize vs school size. Explain what you see. Why would pclassize be a valid instrument (in terms of the exclusion restriction) for class size?

(c) Use the data to estimate the first-stage regression. Does this instrument satisfy the relevance condition? What evidence, if any, supports your answer? The estimated coefficient is less than 1, what explains this?

(d) Estimate the effect of class size on reading test scores using the 2SLS approach and interpret your results. (Hint: you might want to have a look at the Stata command `ivreg` or the R command `ivreg` in the package AER.)