1. Fixed Effects vs. Cross-Sectional Regression

When we compare the results of two regressions, one with fixed effects and one without, how do we decide which is better?

First, here's what you shouldn't do:
DO NOT JUST COMPARE THE R^2 BETWEEN THE TWO MODELS.

Of course the R^2 will go up, because you just added $(m-1)$ dummy variables (fixed effects), where *m* is the number of individuals in the dataset. In any case, `xtreg` does not report the relevant R^2 for comparison. But that doesn't get at the real question we want to answer.

What should you do?
1. *(Purely statistical argument that a robot could make)* To find out if adding the fixed effects added explanatory power to the model, look at the results from the F-test of joint significance of the fixed effects. This is found at the very bottom of the `xtreg` output. If you reject the null hypothesis (e.g. p-value below 0.05), then including the fixed effects increased the explanatory power of the regression.

2. (*Intuitive argument that only a person can make)* Think of how adding the fixed effects reduces the chance of omitted variables bias. Remember that in the cross-sectional regression, OVB results from violating MLR.4:

MLR.4 for cross-sectional regression:$E(u_{it}|x_{it}) = 0$

Adding the fixed effects, OVB still results from violating MLR.4, which now looks like this:

MLR.4 for fixed effects regression: $E(u_{it}|x_{it}, a_i) = 0$

In the context of the specific research question at hand, think of why MLR.4 could be violated in the cross-sectional regression. Then see *which of these violations is mitigated when you add the fixed effects*. This is the most important benefit of the fixed effects regression over the cross-sectional one.

2. Practice with Panel Data and Fixed Effects

Here is a practice problem from the 2008 final exam. There is a very good chance that you will see something quite similar this year. It is worth spending the time to go through this entire problem.

Using state level data on murder rates (***mrdrte***) and unemployment rates (***unem***) in 1987, 1990, and 1993, we want to estimate the effect of unemployment on murder rate. Two estimations are reported below in which ***d90*** and ***d93*** represent dummy variables for the years 1990 and 1993, and ***state*** is a variable that takes the values 1,2, …, 51 for the states (plus D.C.).

a) Write the equation of the model corresponding to the first estimation [be very careful with indices].

b) Does the coefficient on the unemployment variable correctly identify the effect of unemployment on crime? If yes, justify. If not, give a concrete example that illustrates the reason for a biased estimate and the direction of the bias.

c) Write the equation corresponding to the second estimation [be careful with indices]. Describe precisely each new variable you introduce and explain what it represents.

d) Interpret the coefficient on the year 1990 dummy variable in the second estimation.

e) From the second estimation, what do you conclude about the effect of unemployment on crime? Are there still potential sources of bias in this estimation? Justify your response.

REGRESSION 1

```
. reg mrdrte unem d90 d93

      Source |       SS       df       MS              Number of obs =     153
-------------+------------------------------           F(  3,   149) =    3.84
       Model |  920.910876     3  306.970292            Prob > F      =  0.0111
    Residual |  11924.4272   149   80.029713           R-squared     =  0.0717
-------------+------------------------------           Adj R-squared =  0.0530
       Total |  12845.3381   152  84.5088034           Root MSE      =  8.9459

------------------------------------------------------------------------------
      mrdrte |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        unem |   1.443404   .4460414     3.24   0.001     .5620206    2.324788
         d90 |   2.677292   1.815129     1.47   0.142    -.9094277    6.264011
         d93 |   1.667332   1.771566     0.94   0.348    -1.833306    5.167971
       _cons |  -1.999366   3.062258    -0.65   0.515    -8.050428    4.051696
------------------------------------------------------------------------------
```

REGRESSION 2

```
. xtreg mrdrte unem d90 d93, fe i(state)

Fixed-effects (within) regression               Number of obs      =       153
Group variable (i): state                       Number of groups   =        51

R-sq:  within  = 0.0676                         Obs per group: min =         3
       between = 0.1015                                        avg =       3.0
       overall = 0.0314                                        max =         3

                                                F(3,99)            =      2.39
corr(u_i, Xb)  = 0.0951                         Prob > F           =    0.0731

------------------------------------------------------------------------------
      mrdrte |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        unem |   .2019432   .2947557     0.69   0.495    -.3829162    .7868025
         d90 |   1.577016   .7433858     2.12   0.036     .1019775    3.052055
         d93 |   1.681938   .6959821     2.42   0.017     .3009584    3.062917
       _cons |   5.778023   1.911012     3.02   0.003     1.986161    9.569885
-------------+----------------------------------------------------------------
     sigma_u |  8.6877605
     sigma_e |  3.5144936
         rho |  .85936665   (fraction of variance due to u_i)
------------------------------------------------------------------------------
F test that all u_i=0:     F(50, 99) =    17.33              Prob > F = 0.0000
```