Ecmt 675: Econometrics I

Assignment 8

Problem 1 (10.4)

a.

Including the aggregate time effect, $d2_t$, can be very important. Without it, we must assume that any change in average y over the two time periods is due to the program, and not to external, secular factors.

b.

The presence of c_i allows program participation to be correlated with unobserved individual heterogeneity. This is crucial in contexts where the experimental group is not randomly assigned.

c.

$$y_{i1} = \theta_1 + c_i + u_{i1}$$
$$y_{i2} = \theta_1 + \theta_2 + \delta_1 prog_{i2} + c_i + u_{i2}$$

$$\therefore \Delta y_i = \theta_2 + \delta_1 prog_{i2} + \Delta u_i$$

$$\text{If } prog_{i2} = 1, \text{ then } \quad \bar{\Delta y}_{treat} = \hat{\theta}_2 + \hat{\delta}_1 \quad \text{(Treat Group)}$$
$$\text{If } prog_{i2} = 0, \text{ then } \quad \bar{\Delta y}_{control} = \hat{\theta}_2 \quad \text{(Control Group)}$$

$$\therefore \hat{\delta}_1 = \bar{\Delta y}_{treat} - \bar{\Delta y}_{control}$$

d.

In general, for T time periods we have

$$y_{it} = \theta_1 + \theta_2 d2_t + \theta_3 d3_t + ... + \theta_T dT_t + \delta_1 prog_{it} + c_i + u_{it}.$$

We have separate year intercepts, an unobserved effect c_i , and the program indicator.

e.

First, the model from part d is more flexible because it allows any sequence of program participation. In addition, equation (10.81) is restrictive in terms of aggregate time effects: it assumes that any aggregate time effects correspond to the start of the program only. It is better to use the unobserved effects model from part d, and estimate it using either FE or FD.

Problem 2 (10.8)

a.

```
. reg lcrime d78 clrprc1 clrprc2, r

Linear regression                               Number of obs =     106
                                                F(  3,   102) =   24.01
                                                Prob > F      =  0.0000
                                                R-squared     =  0.4710
                                                Root MSE      =  .45495

------------------------------------------------------------------------------
             |               Robust
      lcrime |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         d78 |  -.0547246   .0883541    -0.62   0.537    -.2299747    .1205254
     clrprc1 |  -.0184955   .0047622    -3.88   0.000    -.0279413   -.0090497
     clrprc2 |  -.0173881   .0045592    -3.81   0.000    -.0264311    -.008345
       _cons |    4.18122   .1934741    21.61   0.000     3.797465    4.564975
------------------------------------------------------------------------------

. predict ehat, res

. gen ehat_1=ehat[_n-1] if d78
(53 missing values generated)

. reg ehat ehat_1, r
```

```
Linear regression                                       Number of obs =      53
                                                        F(  1,    51) =   23.74
                                                        Prob > F      =  0.0000
                                                        R-squared     =  0.3990
                                                        Root MSE      =  .33545

------------------------------------------------------------------------------
             |               Robust
        ehat |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      ehat_1 |   .5739582   .1178004     4.87   0.000     .3374639    .8104525
       _cons |  -3.01e-09   .0460779    -0.00   1.000    -.0925053    .0925053
------------------------------------------------------------------------------
```

Strong evidence of positive serial correlation in the composite error.

b.

```
. xtreg  lcrime d78 clrprc1 clrprc2, fe r

Fixed-effects (within) regression               Number of obs      =       106
Group variable (i): district                    Number of groups   =        53

R-sq:  within  = 0.4209                         Obs per group: min =         2
       between = 0.4798                                        avg =       2.0
       overall = 0.4234                                        max =         2

                                                F(3,50)            =      8.84
corr(u_i, Xb)  = 0.3645                         Prob > F           =    0.0001

------------------------------------------------------------------------------
             |               Robust
      lcrime |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         d78 |   .0856556   .0554876     1.54   0.129    -.0257945    .1971057
     clrprc1 |  -.0040475   .0042659    -0.95   0.347    -.0126158    .0045207
     clrprc2 |  -.0131966   .0047286    -2.79   0.007    -.0226942    -.003699
       _cons |   3.350995   .2622724    12.78   0.000     2.824205    3.877785
-------------+----------------------------------------------------------------
     sigma_u |  .47140473
     sigma_e |   .2436645
         rho |  .78915666   (fraction of variance due to u_i)
------------------------------------------------------------------------------
```

No, we don't need to test for serial correlation.

c.

```
. test  clrprc1= clrprc2

 ( 1)  clrprc1 - clrprc2 = 0
```

```
       F(  1,    50) =    1.82
            Prob > F =    0.1828
```

We can not reject the null hypothesis that $\beta_1 = \beta_2$. We can combine two variables together, using the average of $clrprc_{i,t-1}$ and $clrprc_{i,t-2}$. And then do again. We can find the coefficient is more significant.

```
. xtreg  lcrime d78 avgclr, fe

Fixed-effects (within) regression               Number of obs      =       106
Group variable (i): district                    Number of groups   =        53

R-sq:  within  = 0.4076                         Obs per group: min =         2
       between = 0.5101                                        avg =       2.0
       overall = 0.4360                                        max =         2

                                                F(2,51)            =     17.54
corr(u_i, Xb)  = 0.3947                         Prob > F           =    0.0000

------------------------------------------------------------------------------
      lcrime |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         d78 |   .0993289   .0625916     1.59   0.119    -.0263289    .2249867
      avgclr |  -.0166511   .0050672    -3.29   0.002    -.0268239   -.0064783
       _cons |   3.316358   .2305685    14.38   0.000     2.853472    3.779243
-------------+----------------------------------------------------------------
     sigma_u |  .47091962
     sigma_e |  .24402899
         rho |   .7883154   (fraction of variance due to u_i)
------------------------------------------------------------------------------
F test that all u_i=0:     F(52, 51) =     5.84                 Prob > F = 0.0000
```

Problem 3 (10.10)

a.

```
. reg  cmrdrte d93 cexec cunem, r

Linear regression                               Number of obs =     102
                                                F(  3,    98) =    4.77
                                                Prob > F      =  0.0038
                                                R-squared     =  0.0252
                                                Root MSE      =  4.3003

------------------------------------------------------------------------------
             |               Robust
     cmrdrte |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         d93 |  -1.296717   1.296942    -1.00   0.320    -3.870457    1.277023
       cexec |  -.1150682   .0334867    -3.44   0.001    -.1815215   -.0486149
       cunem |   .1630854   .3001721     0.54   0.588    -.4325963    .7587671
```

```
       _cons |    1.51099   1.049792     1.44   0.153    -.5722869    3.594267
------------------------------------------------------------------------------

. reg ehat ehat_1, r

Linear regression                                      Number of obs =      51
                                                       F(  1,    49) =    0.43
                                                       Prob > F      =  0.5146
                                                       R-squared     =  0.0013
                                                       Root MSE      =  1.0908

------------------------------------------------------------------------------
             |               Robust
        ehat |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      ehat_1 |   .0065807   .0100236     0.66   0.515    -.0135626    .0267239
       _cons |  -9.10e-10   .1527453    -0.00   1.000    -.3069532    .3069532
------------------------------------------------------------------------------
```

There is no serial correlation problem in the error term.

b.

```
. xtreg  cmrdrte d93 exec unem, fe

Fixed-effects (within) regression               Number of obs      =       102
Group variable (i): id                          Number of groups   =        51

R-sq:  within  = 0.0694                         Obs per group: min =         2
       between = 0.0381                                        avg =       2.0
       overall = 0.0010                                        max =         2

                                                F(3,48)            =      1.19
corr(u_i, Xb)  = -0.4963                        Prob > F           =    0.3221

------------------------------------------------------------------------------
     cmrdrte |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         d93 |  -.0347133   1.165336    -0.03   0.976    -2.377778    2.308352
        exec |  -.1132541   .2416214    -0.47   0.641    -.5990664    .3725582
        unem |  -1.117735   .8831703    -1.27   0.212    -2.893467    .6579985
       _cons |   7.514069   4.790136     1.57   0.123    -2.117144    17.14528
-------------+----------------------------------------------------------------
     sigma_u |  3.6278788
     sigma_e |  4.2458968
         rho |  .42198992   (fraction of variance due to u_i)
------------------------------------------------------------------------------
F test that all u_i=0:     F(50, 48) =     1.04                 Prob > F = 0.4458
```

The FE estimates are similar to the FD estimates. While executions have an estimated negative effect, and the unemployment rate a positive effect, neither is close to being significant at the usual significance levels.

c.

The explanatory variable $exec_{it}$ might fail strict exogeneity if states increase future executions in response to current positive shocks to the murder rate. Given only a three year window, this is perhaps unlikely, as the judicial process in capital cases tends to move slowly. Nevertheless, with a longer time series, we could add $exec_{i,t+1}$ in the equation and estimate it by fixed effects.

Problem 4 (10.12)

a.

```
. reg lwage educ black hisp exper expersq married union d81-d87, r cluster(nr)

Linear regression                                      Number of obs =    4360
                                                       F( 14,   544) =   47.10
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.1893
Number of clusters (nr) = 545                          Root MSE      =  .48033

------------------------------------------------------------------------------
             |               Robust
       lwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        educ |   .0913498   .0110822     8.24   0.000     .0695807    .1131189
       black |  -.1392342   .0505238    -2.76   0.006    -.2384798   -.0399887
        hisp |   .0160195   .0390781     0.41   0.682     -.060743     .092782
       exper |   .0672345   .0195958     3.43   0.001     .0287417    .1057273
     expersq |  -.0024117   .0010252    -2.35   0.019    -.0044255   -.0003979
     married |   .1082529    .026034     4.16   0.000     .0571135    .1593924
       union |   .1824613   .0274435     6.65   0.000     .1285531    .2363695
         d81 |     .05832   .028228     2.07   0.039     .0028707    .1137693
         d82 |   .0627744   .0369735     1.70   0.090    -.0098538    .1354027
         d83 |   .0620117    .046248     1.34   0.181    -.0288348    .1528583
         d84 |   .0904672    .057988     1.56   0.119    -.0234407     .204375
         d85 |   .1092463   .0668474     1.63   0.103    -.0220644     .240557
         d86 |   .1419596   .0762348     1.86   0.063     -.007791    .2917102
         d87 |   .1738334   .0852056     2.04   0.042     .0064611    .3412057
       _cons |   .0920558   .1609365     0.57   0.568    -.2240773    .4081888
------------------------------------------------------------------------------

. reg lwage educ black hisp exper expersq married union d81-d87

      Source |       SS       df       MS              Number of obs =    4360
-------------+------------------------------           F( 14,  4345) =   72.46
       Model |  234.048277    14  16.7177341           Prob > F      =  0.0000
    Residual |  1002.48136  4345  .230720682           R-squared     =  0.1893
-------------+------------------------------           Adj R-squared =  0.1867
```

```
       Total |  1236.52964  4359  .283672779          Root MSE      =  .48033

------------------------------------------------------------------------------
       lwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        educ |   .0913498   .0052374    17.44   0.000     .0810819    .1016177
       black |  -.1392342   .0235796    -5.90   0.000    -.1854622   -.0930062
        hisp |   .0160195   .0207971     0.77   0.441    -.0247535    .0567925
       exper |   .0672345   .0136948     4.91   0.000     .0403856    .0940834
     expersq |  -.0024117     .00082    -2.94   0.003    -.0040192   -.0008042
     married |   .1082529   .0156894     6.90   0.000     .0774937    .1390122
       union |   .1824613   .0171568    10.63   0.000     .1488253    .2160973
          d81 |     .05832   .0303536     1.92   0.055    -.0011886    .1178286
          d82 |   .0627744   .0332141     1.89   0.059    -.0023421    .1278909
          d83 |   .0620117   .0366601     1.69   0.091    -.0098608    .1338843
          d84 |   .0904672   .0400907     2.26   0.024     .011869     .1690654
          d85 |   .1092463   .0433525     2.52   0.012     .0242533    .1942393
          d86 |   .1419596    .046423     3.06   0.002     .0509469    .2329723
          d87 |   .1738334    .049433     3.52   0.000     .0769194    .2707474
        _cons |   .0920558   .0782701     1.18   0.240    -.0613935    .2455051
------------------------------------------------------------------------------
```

Even if c_i is uncorrelated with all explanatory variables, there still exists the serial correlation problem in the error term. Therefore, we need the robust standard error to adjust it. You can see that the robust standard errors are substantially larger than the usual ones, in some cases more than double.

b.

```
. xtreg lwage educ black hisp exper expersq married union d81-d87, re

Random-effects GLS regression                   Number of obs      =      4360
Group variable (i): nr                          Number of groups   =       545

R-sq:  within  = 0.1799                         Obs per group: min =         8
       between = 0.1860                                        avg =       8.0
       overall = 0.1830                                        max =         8

Random effects u_i ~ Gaussian                   Wald chi2(14)      =    957.77
corr(u_i, X)       = 0 (assumed)                Prob > chi2        =    0.0000

------------------------------------------------------------------------------
       lwage |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        educ |   .0918763   .0106597     8.62   0.000     .0709836    .1127689
       black |  -.1393767   .0477228    -2.92   0.003    -.2329117   -.0458417
        hisp |   .0217317   .0426063     0.51   0.610    -.0617751    .1052385
       exper |   .1057545   .0153668     6.88   0.000     .0756361    .1358729
     expersq |  -.0047239   .0006895    -6.85   0.000    -.0060753   -.0033726
     married |    .063986   .0167742     3.81   0.000     .0311091    .0968629
       union |   .1061344   .0178539     5.94   0.000     .0711415    .1411273
          d81 |    .040462   .0246946     1.64   0.101    -.0079385    .0888626
```

```
         d82 |   .0309212   .0323416     0.96   0.339    -.0324672    .0943096
         d83 |   .0202806    .041582     0.49   0.626    -.0612186    .1017798
         d84 |   .0431187   .0513163     0.84   0.401    -.0574595    .1436969
         d85 |   .0578155   .0612323     0.94   0.345    -.0621977    .1778286
         d86 |   .0919476   .0712293     1.29   0.197    -.0476592    .2315544
         d87 |   .1349289   .0813135     1.66   0.097    -.0244427    .2943005
       _cons |   .0235864   .1506683     0.16   0.876     -.271718    .3188907
-------------+----------------------------------------------------------------
     sigma_u |  .32460315
     sigma_e |  .35099001
         rho |  .46100216   (fraction of variance due to u_i)
------------------------------------------------------------------------------
```

The random effects estimates on the time-constant variables are similar to the pooled OLS estimates. The effect of experience is initially large for the random effects estimates.

c.

The variable $exper_{it}$ is redundant because everyone in the sample works every year.

```
. xtreg lwage expersq married union d81-d87, fe

Fixed-effects (within) regression               Number of obs      =      4360
Group variable (i): nr                          Number of groups   =       545

R-sq:  within  = 0.1806                         Obs per group: min =         8
       between = 0.0286                                        avg =       8.0
       overall = 0.0888                                        max =         8

                                                F(10,3805)         =     83.85
corr(u_i, Xb)  = -0.1222                        Prob > F           =    0.0000

------------------------------------------------------------------------------
       lwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     expersq |  -.0051855   .0007044    -7.36   0.000    -.0065666   -.0038044
     married |   .0466804   .0183104     2.55   0.011     .0107811    .0825796
       union |   .0800019   .0193103     4.14   0.000     .0421423    .1178614
         d81 |   .1511912   .0219489     6.89   0.000     .1081584     .194224
         d82 |   .2529709   .0244185    10.36   0.000     .2050963    .3008454
         d83 |   .3544437   .0292419    12.12   0.000     .2971125    .4117749
         d84 |   .4901148   .0362266    13.53   0.000     .4190894    .5611402
         d85 |   .6174823   .0452435    13.65   0.000     .5287784    .7061861
         d86 |   .7654966   .0561277    13.64   0.000     .6554532    .8755399
         d87 |   .9250249   .0687731    13.45   0.000     .7901893    1.059861
       _cons |   1.426019   .0183415    77.75   0.000     1.390058    1.461979
-------------+----------------------------------------------------------------
     sigma_u |  .39176195
     sigma_e |  .35099001
         rho |  .55472817   (fraction of variance due to u_i)
------------------------------------------------------------------------------
F test that all u_i=0:     F(544, 3805) =     9.16           Prob > F = 0.0000
```

The marriage and union premiums fall even more, although both are still statistically significant and economically relevant.

d.

```
. xtreg lwage expersq married union d81-d87 d81educ-d87educ, fe

Fixed-effects (within) regression               Number of obs      =      4360
Group variable (i): nr                          Number of groups   =       545

R-sq:  within  = 0.1814                         Obs per group: min =         8
       between = 0.0211                                        avg =       8.0
       overall = 0.0784                                        max =         8

                                                F(17,3798)         =     49.49
corr(u_i, Xb)  = -0.1732                        Prob > F           =    0.0000

------------------------------------------------------------------------------
       lwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     expersq |  -.0060437   .0008633    -7.00   0.000    -.0077362   -.0043512
     married |   .0474337   .0183277     2.59   0.010     .0115006    .0833668
       union |   .0789759   .0193328     4.09   0.000     .0410722    .1168796
         d81 |   .0984201    .145999     0.67   0.500    -.1878239    .3846641
         d82 |   .2472016   .1493785     1.65   0.098    -.0456682    .5400714
         d83 |    .408813   .1557146     2.63   0.009     .1035207    .7141053
         d84 |   .6399247   .1652396     3.87   0.000     .3159578    .9638917
         d85 |   .7729397   .1779911     4.34   0.000     .4239723    1.121907
         d86 |   .9699322   .1941747     5.00   0.000     .5892354    1.350629
         d87 |   1.188777   .2135856     5.57   0.000     .7700231     1.60753
     d81educ |   .0049906    .012222     0.41   0.683    -.0189718     .028953
     d82educ |    .001651   .0123304     0.13   0.893    -.0225239    .0258259
     d83educ |  -.0026621   .0125098    -0.21   0.831    -.0271887    .0218644
     d84educ |  -.0098257   .0127593    -0.77   0.441    -.0348414      .01519
     d85educ |  -.0092145   .0130721    -0.70   0.481    -.0348436    .0164146
     d86educ |  -.0121382   .0134419    -0.90   0.367    -.0384922    .0142159
     d87educ |  -.0157892    .013868    -1.14   0.255    -.0429785    .0114002
       _cons |   1.436283   .0192766    74.51   0.000     1.398489    1.474076
-------------+----------------------------------------------------------------
     sigma_u |  .39876325
     sigma_e |   .3511451
         rho |  .56324361   (fraction of variance due to u_i)
------------------------------------------------------------------------------
F test that all u_i=0:     F(544, 3798) =     8.25                 Prob > F = 0.0000

. testparm d81educ-d87educ

 ( 1)  d81educ = 0
 ( 2)  d82educ = 0
 ( 3)  d83educ = 0
 ( 4)  d84educ = 0
 ( 5)  d85educ = 0
 ( 6)  d86educ = 0
 ( 7)  d87educ = 0

       F(  7,  3798) =    0.52
```

```
            Prob > F =    0.8202
```

There is no evidence that the return to education has changed over time for the population represented by these men.

e.

```
. xtreg lwage expersq married union unionp1 d81-d86, fe

Fixed-effects (within) regression               Number of obs      =      3815
Group variable (i): nr                          Number of groups   =       545

R-sq:  within  = 0.1474                         Obs per group: min =         7
       between = 0.0305                                        avg =       7.0
       overall = 0.0744                                        max =         7

                                                F(10,3260)         =     56.34
corr(u_i, Xb)  = -0.1262                        Prob > F           =    0.0000

------------------------------------------------------------------------------
       lwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     expersq |  -.0054448   .0008771    -6.21   0.000    -.0071646    -.003725
     married |   .0448778   .0208817     2.15   0.032     .0039353    .0858203
       union |   .0763554   .0217672     3.51   0.000     .0336766    .1190342
     unionp1 |   .0497356   .0223618     2.22   0.026      .005891    .0935802
         d81 |   .1528275   .0226526     6.75   0.000     .1084128    .1972422
         d82 |   .2576486    .026211     9.83   0.000     .2062568    .3090403
         d83 |   .3618296   .0328716    11.01   0.000     .2973786    .4262806
         d84 |   .5023642   .0422128    11.90   0.000     .4195979    .5851305
         d85 |   .6342402   .0539623    11.75   0.000     .5284368    .7400435
         d86 |   .7841312   .0679011    11.55   0.000     .6509981    .9172642
       _cons |   1.417924   .0204562    69.32   0.000     1.377815    1.458032
-------------+----------------------------------------------------------------
     sigma_u |  .39716048
     sigma_e |  .35740734
         rho |   .5525375   (fraction of variance due to u_i)
------------------------------------------------------------------------------
F test that all u_i=0:     F(544, 3260) =     7.93           Prob > F = 0.0000
```

Yes, it is significant. It means that union fails the strictly exogenous assumptions.

Problem 5 (10.14)

a.

$$c_i = \alpha + z_i\gamma + h_i$$

$$\text{Var}(c_i) = \text{Var}(z_i\gamma) + \sigma_h^2 = \gamma'\text{Var}(z_i)\gamma + \sigma_h^2 \geq \sigma_h^2$$

$\because$ $\text{Var}(z_i)$ is positive definite.

Also, if $\gamma \neq 0$, then $\text{Var}(c_i) > \sigma_h^2$

b.

If we estimate the model by fixed effects, the associated estimate of the variance of the unobserved effect is σ_c^2. If we estimate the model by random effects, the variance component is σ_h^2. This makes intuitive sense: with random effects, we are able to explicitly control for time-constant variances, and so the z_i are effectively taken out of the error term.