Regression II
Vanderbilt University
Prof. Sean P. Corcoran **72 total points**

Problem Set 5 *Solutions*

Question 1. This problem will use regression difference-in-differences to estimate the impact of a breakfast in the classroom (BIC) program on school meals program participation in New York City. BIC was not implemented under random assignment; rather, schools voluntarily adopted the program. We do, however, have data for these and other schools before and after adoption. (See Corcoran, Elbel, & Schwartz 2015 for details). (**40 points**)

(a) In Stata, open the panel dataset *NYCbkfastlunch.dta*. This file consists of school-level data in which the rows are elementary or middle schools observed in year t (t=2005 to 2012). The outcome variables of interest are *bkfast_part* and *lunch_part*, which are average daily participation rates in the school breakfast and lunch programs. Provide some descriptive statistics for these two variables. On what scale are they measured? (**2 points**)

See attached log. Each of the outcome variables ranges between 0-1 and can be interpreted as the proportion of students in attendance on the average school day in school *i* who received a breakfast or lunch. The average daily participation was 0.780 (or 78.0%) for lunch and only 0.258 (or 25.8%) for breakfast.

(b) Use `xtset` to declare the data as a panel. Which variable is the cross-sectional unit and which is the time dimension? Is this a balanced panel? (Show how you answer the last question). How many schools are observed in all 8 years? (**2 points**)

See attached log. *schoolid* is the cross-sectional unit, while *year* is the time dimension. This is a balanced panel, as easily seen in `xtdescribe` which shows the pattern of data availability. 907 schools have data in all 8 years.

(c) This dataset contains a time-varying treatment variable called *bicpost* that is equal to 1 in years in which the school offered BIC and 0 otherwise. For this part, limit the sample to schools that adopted BIC in 2010 and never-adopters. Estimate a simple "2x2" difference-in-differences regression comparing mean breakfast participation rates for two groups (treated and untreated schools) in two time periods (pre and post). Do the same for lunch participation rates, and interpret your results. (Note: ignore the panel nature of your data and just use the `regress` command for this part). Is the BIC effect statistically significant? Practically significant? What assumptions must be satisfied for this difference-in-differences to be considered a causal effect? (**6 points**)

See attached log. The do file first identifies the first year of treatment for each school and then flags schools that adopted BIC in 2010. The

2x2 DD coefficient is positive and statistically significant for breakfast, and statistically insignificant for lunch. In the case of breakfast, average daily participation increased 23.9 percentage points more in schools that adopted BIC in 2010 (the treated schools) relative to schools that never adopted BIC (the untreated schools).

For this estimate to be considered causal, the parallel trends assumption must be satisfied. That is, the pre-post change in breakfast participation among non-BIC schools must represent what would have occurred in BIC schools had they not been "treated" by adopting the program. As a panel model, the strict exogeneity assumption must also hold. This means the error term in every year t cannot be related to treatment adoption. This would be violated if there were anticipation effects, for example.

(d) Repeat part (c) but include year dummies in the regressions. How does this affect your impact estimates for BIC, if at all? Continue to use the `regress` command, and explain why one of the post-2010 year effects is not estimable. (**4 points**)

See attached log. There is only a trivial change to the estimated DD coefficient in both the breakfast and lunch regressions. One of the post-2010 year effects is omitted. This is because of perfect collinearity between the *post2010* variable and the three dummies for 2010, 2011, and 2012. (If you know three of these variables, the fourth is also known).

(e) In parts (c)-(d) you did not take advantage of the panel nature of your data. Re-estimate these regressions using `xtreg` and school fixed effects. How do your estimates compare? (**4 points**)

See attached log. Again there is only a trivial change to the estimated DD coefficient in both regressions.

(f) Compare the mean baseline characteristics of treated and untreated schools in parts (c)-(e). (Use 2009 as the baseline year). Look at the following: total enrollment, % ELL, % special education, % Asian, % black, % Hispanic, % female, % free lunch eligible (*free1*), % reduced price lunch eligible (*redu1*). How do schools that adopted BIC in 2010 compare to those that never adopted? (**4 points**)

See attached log. It is clear that schools which adopted BIC in 2010 differ systematically from those that didn't. They are more disadvantaged on a number of dimensions, including %ELL, %sped, and percent with family income low enough to be eligible for free meals. These differences suggest we should include these covariates in the regressions. Others—such as the race or gender composition of the school—may also help to reduce unexplained variation in the outcome.

(g) Now estimate the same regression models in part (e), but include the school covariates listed above in your regressions. How do your estimates of the "BIC effect" change, if at all? (And how are these covariates related to meal participation?) (**4 points**)

See attached log. It appears the inclusion of covariates has little effect on the difference-in-differences estimates for breakfast and lunch participation. Many of the included covariates are related to meals program participation. For example, *%ELL, %sped*, and percent eligible for free meals are all positively related to breakfast participation. *Total enrollment, Asian, Hispanic*, and *female* shares are all negatively related to breakfast participation.

(h) Repeat part (e)—the model without covariates—but try using the Stata command `xtdidregress`. Compare your results. (**4 points**)

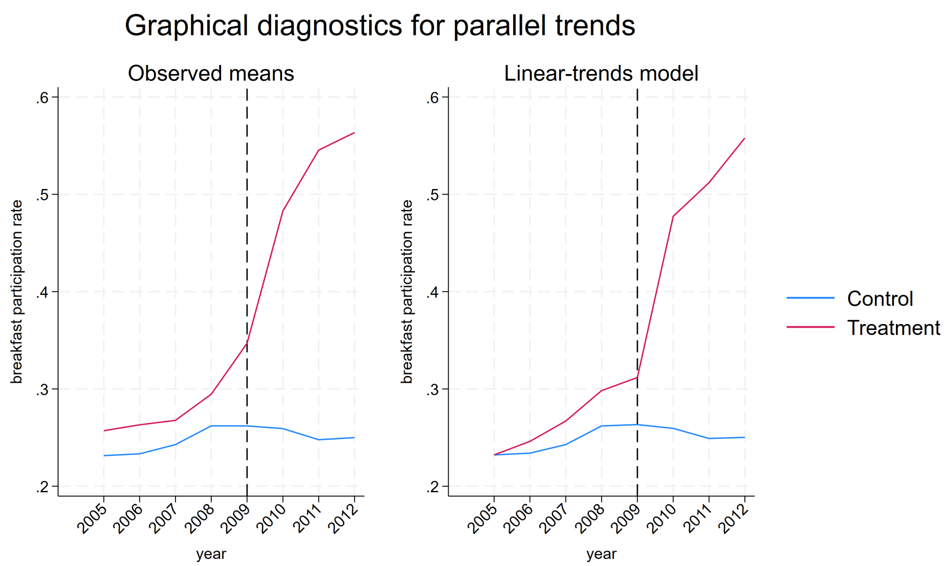
See attached log. The results are identical, although the standard errors are different (unless you—correctly—adjusted your standard errors for clustering by school).

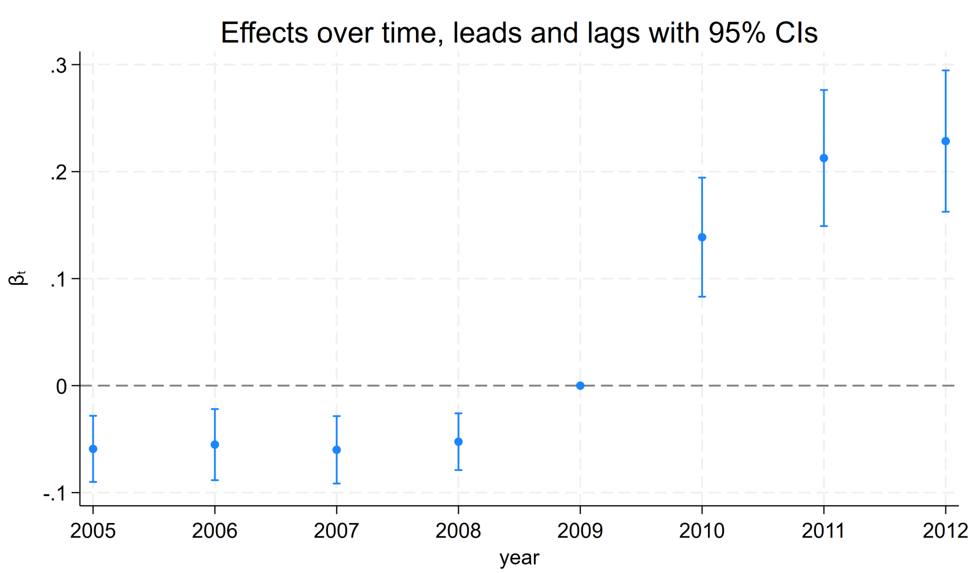
(i) State the assumption(s) that are required to hold in order to interpret the breakfast participation DD coefficient in part (h) as causal. Use the post-estimation commands `estat trendplots` and `estat granger` to probe these assumptions, and interpret the results. (**5 points**)

Again, the parallel trends and no anticipation assumptions must be satisfied (see part c). See attached log for the post-estimation commands (for the breakfast model only). The Granger causality test rejects the null hypothesis of "no anticipation." The graphical output is below. These results are *not* consistent with a common pre-trend for the treated and untreated schools. The second plot is a Granger plot (`estat grangerplot`) which shows more clearly that the trends are consistent until 2009—the year before treatment. This could be an anticipation effect, or an indication that our information about the timing of treatment is incorrect for one or more schools.

(j) Finally, incorporate <u>all</u> treated schools into the analysis regardless of when they adopted BIC. Estimate the two-way fixed effects model using `xtreg` (as in part e) and `xtdidregress` (as in part h). Try these without and with school covariates. How do your results compare? (**5 points**)

See attached log. The results are very similar to the previous ones, in all cases, when including the full set of treated schools. Note we have not addressed the potential problems introduced by staggered treatment timing.


Graphical diagnostics for parallel trends
Observed means
Linear-trends model
breakfast participation rate
.2
.3
.4
.5
.6
2005
2006
2007
2008
2009
2010
2011
2012
year
Control
Treatment


Effects over time, leads and lags with 95% CIs
β_t
-.1
0
.1
.2
.3
2005
2006
2007
2008
2009
2010
2011
2012
year

Question 2. For these questions, refer to the recent article by Cellini and Turner (2019), "Gainfully Employed? Assessing the Employment and Earnings of For-Profit College Students Using Administrative Data." You can find the article here: `http://jhr.uwpress.org/content/54/2/342.abstract`. **(32 points)**

(a) Cellini and Turner use a generalized difference-in-differences regression model to estimate the causal effect of attending a for-profit certificate program on labor market outcomes. What specific outcome variables do they examine, and what dataset(s) do they use? **(4 points)**

The main outcome variables include employment (0/1) and earnings (levels and logs). US DOE data identify all federally-aided students who exited a for-profit post-secondary institution or public community college certificate program between 2006 and 2008. Income and employment status are taken from tax data from the IRS (1999-2014).

(b) How is the "treatment" variable defined here and what are the possible "pre" and "post" years? How many potential pre and post years are there? **(4 points)**

"Treatment" is attendance at a for-profit post-secondary institution. The potential "pre-treatment" years are 1999-2007, depending on when students started and exited their program. The potential "post" years are 2007-2014, again depending on when students left their program. Thus, there are approximately 6 pre-treatment years and 5-6 post-treatment years available.

(c) The authors use three different groups of "untreated" individuals as comparison groups. What were they, and what was their rationale for looking at each? Which comparison group is their "preferred" one, and why? **(4 points)**

Their comparison groups include: (1) public community college students; (2) public community college students matched by demographics, prior earnings, field of study, geography, and age group; and (3) a matched sample of individuals with no post-secondary education. The first two comparison groups are used to estimate the effect of attending a for-profit institution relative to a public institution. The second is preferred by the authors since it accounts for differences in the composition of these two populations (for-profit and public). The third comparison group is used to estimate the effect of attending a for-profit institution relative to no college. One challenge with the "no college" comparison group is the lack of a defined pre- and post- period. Unlike the community college comparison group, the no college group also cannot be matched based on, say, field of study.

(d) Equation (1) on page 350 shows their regression specification. Carefully explain what each term represents, and how the causal effect of attending a for-profit certificate program is being identified. Why is there not a main effect for "For-Profit" in the model? **(5 points)**

$$y_{it} = \alpha_0 + \alpha_1(Post_{it}) + \alpha_2(Post_{it} * ForProfit_i) + d_t + d_a + d_i + \epsilon_i t$$

- **$Post_{it} = 1$ in all years following exit from the post-secondary program**
- **$ForProfit_i = 1$ for individuals *i* that attended a for-profit program**
- **d_t are year effects to capture mean differences in the outcome due to macroeconomic conditions**
- **d_a are age fixed effects to capture mean differences in the outcome due to, say, work experience**
- **d_i are individual fixed effects**
- **α_1 is the "first difference": the mean pre-to-post change in the outcome for individuals who attended public certificate programs**
- **α_2 is the "second difference" (or DD): the differential pre-to-post change in the outcome for individuals who attended a for-profit program**

There are individual fixed effects in this regression, which means coefficient estimates are identified using *within-person* changes over time. There is no $ForProfit_i$ main effect since it is collinear with the individual effect. (It does not vary over time, but rather just indicates whether or not the student ultimately enrolled in a for-profit institution).

(e) Carefully explain the main assumption necessary to interpret the difference-in-differences estimate here as a causal effect. What evidence do the authors provide that this assumption holds for their three different comparison groups? **(5 points)**

The assumption is that the pre-to-post change in the outcome for the comparison group (e.g., public college attendees) represents what would have occurred for the treatment group (for-profit attendees), had they not been treated. Figure 1 displays the trend in employment and earnings for each group, prior to their enrollment in a post-secondary program. While there are differences in the levels of earnings and employment, the time trends look very similar. This is particularly true for the matched sample. This provides some confidence that the trends would have remained the same in the absence of the treatment.

(f) The paper's main results are reported in Table 3. Carefully interpret the coefficients reported in Panel B. What additional evidence does Figure 2 provide? **(5 points)**

Panel B of Table 3 reports the estimated pre-post change for the matched public sample, and the DD. For individuals who enrolled in public community college programs, annual earnings increased by $1,069, on average, in years after exit from the program. For individuals in for-profit programs,

this change was $2,144 *lower*. Taken together, the annual earnings of the treatment group *declined* by an average of $1,075.

The log earnings column excludes individuals with zero earnings, and thus should be interpreted as conditional on working. For individuals in public community college programs, annual earnings increased by about 16.8%. For individuals in for-profit programs, this change was 11.3 percentage points *lower*. Taken together, their annual earnings—conditional on working—increased by 5.5%.

The first column shows the effect on employment. For individuals in public programs, there was no significant change in the probability of employment. However, for individuals in for-profit programs, the probability of employment *declined* by 1.5 percentage points after exiting their program.

Figure 2 is an event study—it shows the differential effect of attending a for-profit program separately for each year after exiting the program.

(g) Finally, Figure 4 shows the distribution of earnings effects *by school* for public and for-profit institutions. Cellini and Turner describe these as the result of "single-difference" regressions. Briefly explain what they mean by this, and why these should not be interpreted as the causal effects of attending specific institutions. **(5 points)**

It would be difficult to estimate a separate difference-in-differences model for each school since it is not obvious who the comparison group should be. (I.e., who are the students exiting public programs that might have attended *that specific for-profit program*?) Instead, Cellini & Turner estimate mean within-person, pre-post changes in earnings for each for-profit institution. These are single differences because they represent only the mean pre-post change of one group. As with any single difference, they should not be interpreted as causal, as it is difficult to separate the treatment effect from any change in the outcome over time that might have occurred in the absence of the treatment.

```
.
. // *********************************************************************
. // LPO-8852 Problem set 5 solutions
. // Last updated: October 7, 2024
. // *********************************************************************
.
. // ****
. // (a)
. // ****
. // Get data
.
.         use https://github.com/spcorcor18/LPO-8852/raw/main/data/NYCbkfastlunc
> h-v2.dta, clear

.         summ bkfast_part lunch_part

    Variable |        Obs        Mean    Std. dev.       Min        Max
-------------+---------------------------------------------------------
 bkfast_part |      7,255    .2580688    .1377646   .0054945      .9725
  lunch_part |      7,005    .7803097    .1838588   .0122249          1

.
.
. // ****
. // (b)
. // ****
. // Set panel
.
.         xtset schoolid year

Panel variable: schoolid (strongly balanced)
 Time variable: year, 2005 to 2012
         Delta: 1 unit

.         xtdescribe

schoolid:  1, 2, ..., 1179                                  n =        907
    year:  2005, 2006, ..., 2012                            T =          8
           Delta(year) = 1 unit
           Span(year)  = 8 periods
           (schoolid*year uniquely identifies each observation)

Distribution of T_i:   min      5%     25%       50%       75%     95%     max
                         8       8       8         8         8       8       8

     Freq.  Percent    Cum. |  Pattern
 ---------------------------+----------
      907    100.00  100.00 |  11111111
 ---------------------------+----------
      907    100.00         |  XXXXXXXX

.
.
. // ****
. // (c)
. // ****
. // Simple 2x2 DD for limited sample
.
.         tabulate year bicpost
```

```
           | =1 if post period for
           |      BIC school
      year |         0          1 |     Total
-----------+----------------------+----------
      2005 |       907          0 |       907 
      2006 |       907          0 |       907 
      2007 |       906          1 |       907 
      2008 |       902          5 |       907 
      2009 |       878         29 |       907 
      2010 |       848         59 |       907 
      2011 |       828         79 |       907 
      2012 |       800        107 |       907 
-----------+----------------------+----------
     Total |     6,976        280 |     7,256 

. 
.         // identify first year that school was treated
.         egen temp1=min(year) if bicpost==1, by(schoolid)
(6,976 missing values generated)

.         egen firstyear=max(temp1), by(schoolid)
(6,400 missing values generated)

.         drop temp1

.         table year firstyear

----------------------------------------------------------
        |                   firstyear                     
        |  2007   2008   2009   2010   2011   2012   Total
--------+-------------------------------------------------
year    |                                                 
  2005  |     1      4     24     30     20     28     107
  2006  |     1      4     24     30     20     28     107
  2007  |     1      4     24     30     20     28     107
  2008  |     1      4     24     30     20     28     107
  2009  |     1      4     24     30     20     28     107
  2010  |     1      4     24     30     20     28     107
  2011  |     1      4     24     30     20     28     107
  2012  |     1      4     24     30     20     28     107
  Total |     8     32    192    240    160    224     856
----------------------------------------------------------

. 
.         // simple DD: treated schools adopting in 2010 vs never adopted
.         // create "post" period variable
.         gen post2010 = (year>=2010 & year~=.)

. 
.         // create "ever treated" variable
.         gen everbic2010 = firstyear==2010

. 
.         _eststo bk1: reg bkfast_part i.everbic2010##i.post2010 if (firstyear==
> 2010 | firstyear==.)

      Source |       SS           df       MS      Number of obs   =     6,639
-------------+----------------------------------   F(3, 6635)      =    152.58
       Model |  7.27543865         3  2.42514622   Prob > F        =    0.0000
    Residual |  105.458173     6,635  .015894224   R-squared       =    0.0645
-------------+----------------------------------   Adj R-squared   =    0.0641
       Total |  112.733612     6,638  .016983069   Root MSE        =    .12607
```

```
------------------------------------------------------------------------------
  bkfast_part | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
--------------+----------------------------------------------------------------
1.everbic2010 |   .0395468    .010485     3.77   0.000     .0189927    .0601008
   1.post2010 |   .0060172   .0032553     1.85   0.065    -.0003643    .0123986
              |
  everbic2010#|
     post2010 |
         1 1  |   .2386908   .0171219    13.94   0.000     .2051262    .2722553
              |
        _cons |   .2463439   .0019936   123.57   0.000     .2424358    .2502521
------------------------------------------------------------------------------

.         _eststo lu1: reg lunch_part i.everbic2010##i.post2010 if (firstyear==2
> 010 | firstyear==.)

      Source |       SS           df       MS      Number of obs   =     6,423
-------------+----------------------------------   F(3, 6419)      =     10.27
       Model |  1.07347052         3  .357823508   Prob > F        =    0.0000
    Residual |  223.584325     6,419  .034831644   R-squared       =    0.0048
-------------+----------------------------------   Adj R-squared   =    0.0043
       Total |  224.657795     6,422  .034982528   Root MSE        =    .18663

------------------------------------------------------------------------------
   lunch_part | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
--------------+----------------------------------------------------------------
1.everbic2010 |   .0664884    .016285     4.08   0.000     .0345645    .0984124
   1.post2010 |   -.006335   .0048792    -1.30   0.194    -.0158999    .0032299
              |
  everbic2010#|
     post2010 |
         1 1  |   .0057016   .0260823     0.22   0.827    -.0454284    .0568315
              |
        _cons |   .7758114   .0030141   257.39   0.000     .7699027    .7817201
------------------------------------------------------------------------------

.
.
. // ****
. // (d)
. // ****
. // Add year effects
.
.         _eststo bk2: reg bkfast_part i.everbic2010##i.post2010 i.year if (firs
> tyear==2010 | firstyear==.)
note: 2012.year omitted because of collinearity.

      Source |       SS           df       MS      Number of obs   =     6,639
-------------+----------------------------------   F(9, 6629)      =     57.12
       Model |  8.11373426         9  .901526029   Prob > F        =    0.0000
    Residual |  104.619878     6,629  .015782151   R-squared       =    0.0720
-------------+----------------------------------   Adj R-squared   =    0.0707
       Total |  112.733612     6,638  .016983069   Root MSE        =    .12563

------------------------------------------------------------------------------
  bkfast_part | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
--------------+----------------------------------------------------------------
1.everbic2010 |   .0395506    .010448     3.79   0.000     .0190692    .0600321
   1.post2010 |   .0202802   .0061994     3.27   0.001     .0081274    .0324331
              |
  everbic2010#|
     post2010 |
         1 1  |   .2386869   .0170615    13.99   0.000     .2052409    .2721329
              |
         year |
        2006  |   .0019981   .0061686     0.32   0.746    -.0100944    .0140906
        2007  |    .011271   .0061686     1.83   0.068    -.0008216    .0233635
        2008  |   .0308351   .0061686     5.00   0.000     .0187425    .0429276
        2009  |   .0326727   .0061686     5.30   0.000     .0205802    .0447652
        2010  |   .0060077   .0061668     0.97   0.330    -.0060812    .0180966
        2011  |  -.0027192   .0061668    -0.44   0.659    -.0148081    .0093697
        2012  |          0  (omitted)
```

```
             |
       _cons |   .2309847   .0043796    52.74   0.000     .2223994    .2395701
------------------------------------------------------------------------------

.         _eststo lu2: reg lunch_part i.everbic2010##i.post2010 i.year if (first
> year==2010 | firstyear==.)
note: 2012.year omitted because of collinearity.

      Source |       SS           df       MS      Number of obs   =     6,423
-------------+----------------------------------   F(9, 6413)      =      4.98
       Model |  1.55895067         9  .173216741   Prob > F        =    0.0000
    Residual |  223.098845     6,413   .03478853   R-squared       =    0.0069
-------------+----------------------------------   Adj R-squared   =    0.0055
       Total |  224.657795     6,422  .034982528   Root MSE        =    .18652

------------------------------------------------------------------------------
  lunch_part | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
1.everbic2010 |   .0666754   .0162752     4.10   0.000     .0347706    .0985801
   1.post2010 |  -.0026473   .0093423    -0.28   0.777    -.0209614    .0156667
             |
  everbic2010#|
    post2010 |
        1 1  |   .0057818   .0260666     0.22   0.824    -.0453174    .0568811
             |
        year |
       2006  |   .0092882   .0093145     1.00   0.319    -.0089713    .0275477
       2007  |     .01739   .0093231     1.87   0.062    -.0008864    .0356663
       2008  |   .0228403    .009385     2.43   0.015     .0044426    .0412379
       2009  |   .0235812   .0094064     2.51   0.012     .0051416    .0420208
       2010  |   .0203108   .0092229     2.20   0.028     .0022309    .0383908
       2011  |   .0121711    .009234     1.32   0.188    -.0059306    .0302729
       2012  |          0  (omitted)
             |
       _cons |   .7612627   .0066357   114.72   0.000     .7482546    .7742708
------------------------------------------------------------------------------

.
.
. // ****
. // (e)
. // ****
. // Use xtreg and school fixed effects
.
.         _eststo bk3: xtreg bkfast_part i.bicpost i.year if (firstyear==2010 |
> firstyear==.), fe

Fixed-effects (within) regression               Number of obs     =      6,639
Group variable: schoolid                        Number of groups  =        830

R-squared:                                      Obs per group:
     Within  = 0.2143                                         min =          7
     Between = 0.0414                                         avg =        8.0
     Overall = 0.0697                                         max =          8

                                                F(8, 5801)        =     197.80
corr(u_i, Xb) = 0.0349                          Prob > F          =     0.0000

------------------------------------------------------------------------------
 bkfast_part | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
   1.bicpost |   .2386411   .0070527    33.84   0.000     .2248152    .2524671
             |
        year |
       2006  |   .0022187     .00255     0.87   0.384    -.0027804    .0072177
       2007  |   .0114915     .00255     4.51   0.000     .0064925    .0164905
       2008  |   .0310556     .00255    12.18   0.000     .0260566    .0360547
       2009  |   .0328933     .00255    12.90   0.000     .0278942    .0378923
       2010  |   .0265101   .0025628    10.34   0.000     .0214861    .0315341
       2011  |   .0177832   .0025628     6.94   0.000     .0127593    .0228072
       2012  |   .0205024   .0025628     8.00   0.000     .0154785    .0255264
             |
```

```
       _cons |   .2322215   .0018037   128.74   0.000     .2286855    .2357575
-------------+----------------------------------------------------------------
     sigma_u |  .11608572
     sigma_e |  .05193033
         rho |   .8332519   (fraction of variance due to u_i)
------------------------------------------------------------------------------
F test that all u_i=0: F(829, 5801) = 39.90                  Prob > F = 0.0000

.         _eststo lu3: xtreg lunch_part i.bicpost i.year if (firstyear==2010 | f
> irstyear==.), fe

Fixed-effects (within) regression               Number of obs     =      6,423
Group variable: schoolid                        Number of groups  =        830

R-squared:                                      Obs per group:
     Within  = 0.0324                                         min =          2
     Between = 0.0216                                         avg =        7.7
     Overall = 0.0023                                         max =          8

                                                F(8, 5585)        =      23.41
corr(u_i, Xb) = -0.0113                         Prob > F          =     0.0000

------------------------------------------------------------------------------
  lunch_part | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
   1.bicpost |   .0013516   .0090263     0.15   0.881    -.0163435    .0190467
             |
        year |
       2006  |   .0063772   .0031937     2.00   0.046     .0001163    .0126381
       2007  |   .0158261    .003197     4.95   0.000     .0095588    .0220935
       2008  |   .0252871   .0032223     7.85   0.000     .0189702     .031604
       2009  |   .0282942   .0032308     8.76   0.000     .0219605    .0346279
       2010  |   .0129657   .0031958     4.06   0.000     .0067006    .0192307
       2011  |   .0054245   .0032009     1.69   0.090    -.0008505    .0116995
       2012  |  -.0052814   .0032046    -1.65   0.099    -.0115636    .0010008
             |
       _cons |   .7647871   .0022679   337.22   0.000     .7603412    .7692331
-------------+----------------------------------------------------------------
     sigma_u |   .1764832
     sigma_e |  .06381114
         rho |  .88438184   (fraction of variance due to u_i)
------------------------------------------------------------------------------
F test that all u_i=0: F(829, 5585) = 59.53                  Prob > F = 0.0000

.
.
. // ****
. // (f)
. // ****
. // Descriptive statistics for BIC 2010 adopters and those who never adopt BIC
.
.         sum totalenrollment- pctfemale free1 redu1 if firstyear==2010 & year==
> 2009, sep(0)

    Variable |        Obs        Mean    Std. dev.       Min        Max
-------------+---------------------------------------------------------
totalenrol~t |         30       628.2    263.4843        246       1244
      pctell |         30       14.79    11.17585        1.1         44
     pctsped |         30       18.31    7.017483        5.7       36.5
    pctasian |         30         2.7    4.906294          0       20.8
    pctblack |         30       45.85    29.04892        1.7       95.2
     pcthisp |         30       46.34    27.65217        3.5       91.8
    pctwhite |         30    4.046667    11.81582         .2       64.6
   pctfemale |         30          49    3.054336       43.3       54.9
       free1 |         30    82.38852    12.33706   33.07393        100
       redu1 |         30    7.525835    3.735603   .8522727   15.38461
```

```
.         sum totalenrollment- pctfemale free1 redu1 if firstyear==. & year==200
> 9, sep(0)

    Variable |        Obs        Mean    Std. dev.       Min        Max
-------------+---------------------------------------------------------
totalenrol~t |        800    689.1938    321.8793        110       2145
      pctell |        800    12.53675    10.90083          0       77.3
     pctsped |        800    15.92213    5.769476          0         44
    pctasian |        800    14.09225    18.71298          0       91.8
    pctblack |        800    30.62975    29.66412          0       96.8
     pcthisp |        800     37.9295    26.02649        1.6       98.7
    pctwhite |        800    16.33788    22.33898          0       92.8
   pctfemale |        800    49.09438    3.938139       36.2        100
       free1 |        800    67.48113    22.88177    5.67201        100
       redu1 |        800    10.20999    5.023362          0   26.29717

.
.
. // ****
. // (g)
. // ****
. // fixed effects regression with covariates
.
.         global covars "pctell pctsped totalenrollment pctasian pctblack pcthis
> p"

.         global covars "$covars pctfemale free1 redu1"

.
.         _eststo bk4: xtreg bkfast_part i.bicpost $covars if (firstyear==2010 |
>  firstyear==.),fe

Fixed-effects (within) regression               Number of obs     =      6,639
Group variable: schoolid                        Number of groups  =        830

R-squared:                                      Obs per group:
     Within  = 0.2149                                         min =          7
     Between = 0.2035                                         avg =        8.0
     Overall = 0.2040                                         max =          8

                                                F(10, 5799)       =     158.69
corr(u_i, Xb) = 0.0871                          Prob > F          =     0.0000

------------------------------------------------------------------------------
  bkfast_part | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
    1.bicpost |   .2345413    .006981    33.60   0.000     .2208558    .2482268
       pctell |   .0005993   .0002543     2.36   0.018     .0001007    .0010978
      pctsped |   .0020669   .0003058     6.76   0.000     .0014673    .0026665
totalenroll~t |   -.000104   8.65e-06   -12.02   0.000    -.0001209   -.000087
     pctasian |   .0009075   .0003813     2.38   0.017       .00016    .0016549
     pctblack |  -.0000417   .0003147    -0.13   0.895    -.0006587    .0005753
      pcthisp |   .0003415    .000309     1.11   0.269    -.0002643    .0009473
    pctfemale |  -.0005469   .0003687    -1.48   0.138    -.0012697    .0001758
        free1 |   .0002421   .0000988     2.45   0.014     .0000484    .0004358
        redu1 |   .0004033   .0002282     1.77   0.077     -.000044    .0008507
        _cons |   .2661046   .0303249     8.78   0.000     .2066566    .3255527
-------------+----------------------------------------------------------------
      sigma_u |  .10614395
      sigma_e |  .05192171
          rho |  .80691974   (fraction of variance due to u_i)
------------------------------------------------------------------------------
F test that all u_i=0: F(829, 5799) = 26.88                 Prob > F = 0.0000
```

```
.         _eststo lu4: xtreg lunch_part i.bicpost $covars if (firstyear==2010 |
> firstyear==.),fe

Fixed-effects (within) regression               Number of obs     =      6,423
Group variable: schoolid                        Number of groups  =        830

R-squared:                                      Obs per group:
     Within  = 0.0777                                         min =          2
     Between = 0.3368                                         avg =        7.7
     Overall = 0.3055                                         max =          8

                                                F(10, 5583)       =      47.01
corr(u_i, Xb) = -0.0535                         Prob > F          =     0.0000

------------------------------------------------------------------------------
   lunch_part | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
    1.bicpost |  -.0060766   .0087263    -0.70   0.486    -.0231836    .0110304
       pctell |  -.0002715    .000313    -0.87   0.386    -.0008852    .0003421
      pctsped |    .000745   .0003748     1.99   0.047     .0000103    .0014797
totalenroll~t |  -.0000864   .0000106    -8.12   0.000    -.0001073   -.0000655
     pctasian |   .0042328   .0004605     9.19   0.000       .00333    .0051356
     pctblack |   .0043441   .0003814    11.39   0.000     .0035964    .0050918
      pcthisp |   .0033796    .000376     8.99   0.000     .0026425    .0041168
    pctfemale |  -.0008161   .0004515    -1.81   0.071    -.0017013     .000069
        free1 |   .0009428    .000121     7.79   0.000     .0007056    .0011801
        redu1 |    .001884   .0002761     6.82   0.000     .0013426    .0024253
        _cons |   .4647235   .0369153    12.59   0.000     .3923551     .537092
-------------+----------------------------------------------------------------
      sigma_u |   .1437498
      sigma_e |  .06231349
          rho |  .84181471   (fraction of variance due to u_i)
------------------------------------------------------------------------------
F test that all u_i=0: F(829, 5583) = 25.69                 Prob > F = 0.0000

.
.
. // ****
. // (h)
. // ****
. // try using xtdidregress
.
.         xtdidregress (bkfast_part) (bicpost) if (firstyear==2010 | firstyear==
> .), group(schoolid) time(year)

Treatment and time information

Time variable: year
Control:       bicpost = 0
Treatment:     bicpost = 1
-----------------------------------
             |   Control  Treatment
-------------+---------------------
Group        |
    schoolid |       800         30
-------------+---------------------
Time         |
     Minimum |      2005       2010
     Maximum |      2006       2010
-----------------------------------

Difference-in-differences regression                     Number of obs = 6,639
Data type: Longitudinal
```

```
                             (Std. err. adjusted for 830 clusters in schoolid)
------------------------------------------------------------------------------
             |               Robust
 bkfast_part | Coefficient  std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
ATET         |
     bicpost |
   (1 vs 0)  |   .2386411   .0265832     8.98   0.000     .1864628    .2908195
------------------------------------------------------------------------------
Note: ATET estimate adjusted for panel effects and time effects.
Note: Treatment occurs at different times.

.         xtdidregress (lunch_part) (bicpost) if (firstyear==2010 | firstyear==.
> ), group(schoolid) time(year)

Treatment and time information

Time variable: year
Control:       bicpost = 0
Treatment:     bicpost = 1
-----------------------------------
             |   Control  Treatment
-------------+---------------------
Group        |
    schoolid |       800         30
-------------+---------------------
Time         |
     Minimum |      2005       2010
     Maximum |      2008       2012
-----------------------------------

Difference-in-differences regression                     Number of obs = 6,423
Data type: Longitudinal

                             (Std. err. adjusted for 830 clusters in schoolid)
------------------------------------------------------------------------------
             |               Robust
  lunch_part | Coefficient  std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
ATET         |
     bicpost |
   (1 vs 0)  |   .0013516    .013313     0.10   0.919    -.0247795    .0274827
------------------------------------------------------------------------------
Note: ATET estimate adjusted for panel effects and time effects.
Note: Treatment occurs at different times and estimation sample contains units
      that switch in and out of treatment.

.
.
. // ****
. // (i)
. // ****
. // post-estimation commands
.
.         xtdidregress (bkfast_part) (bicpost) if (firstyear==2010 | firstyear==
> .), group(schoolid) time(year)

Treatment and time information

Time variable: year
Control:       bicpost = 0
Treatment:     bicpost = 1
-----------------------------------
             |   Control  Treatment
-------------+---------------------
Group        |
    schoolid |       800         30
-------------+---------------------
Time         |
     Minimum |      2005       2010
     Maximum |      2006       2010
-----------------------------------
```

```
Difference-in-differences regression                   Number of obs = 6,639
Data type: Longitudinal

                             (Std. err. adjusted for 830 clusters in schoolid)
------------------------------------------------------------------------------
             |               Robust
 bkfast_part | Coefficient  std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
ATET         |
     bicpost |
   (1 vs 0)  |   .2386411   .0265832     8.98   0.000     .1864628    .2908195
------------------------------------------------------------------------------
Note: ATET estimate adjusted for panel effects and time effects.
Note: Treatment occurs at different times.

.         estat trendplots

.         graph export "trendplots.png", as(png) name("Graph") replace
file trendplots.png saved as PNG format

.         estat granger

Granger causality test
H0: No effect in anticipation of treatment

F(4, 829) =   4.88
 Prob > F = 0.0007

.
.         // see also grangerplot command
.         estat grangerplot

.         graph export "grangerplot.png", as(png) name("Graph") replace
file grangerplot.png saved as PNG format

.
.
. // ****
. // (j)
. // ****
. // incorporate all schools into the analysis
.
.         // without covariates (xtreg)
.         _eststo bk5: xtreg bkfast_part i.bicpost i.year, fe

Fixed-effects (within) regression               Number of obs     =      7,255
Group variable: schoolid                        Number of groups  =        907

R-squared:                                      Obs per group:
     Within  = 0.3087                                         min =          7
     Between = 0.0667                                         avg =        8.0
     Overall = 0.1256                                         max =          8

                                                F(8, 6340)        =     353.93
corr(u_i, Xb) = 0.0196                          Prob > F          =     0.0000

------------------------------------------------------------------------------
 bkfast_part | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
   1.bicpost |   .2282953   .0048888    46.70   0.000     .2187115    .2378791
             |
        year |
       2006  |   .0028761   .0028522     1.01   0.313    -.0027151    .0084673
       2007  |   .0116797   .0028522     4.10   0.000     .0060884    .0172709
       2008  |   .0307599   .0028523    10.78   0.000     .0251684    .0363514
       2009  |   .0306009   .0028565    10.71   0.000     .0250012    .0362005
       2010  |   .0297214   .0028699    10.36   0.000     .0240955    .0353473
       2011  |    .020442   .0028838     7.09   0.000     .0147888    .0260953
       2012  |   .0219312   .0029099     7.54   0.000     .0162268    .0276357
             |
       _cons |    .230754   .0020174   114.38   0.000     .2267992    .2347088
```

```
-------------+----------------------------------------------------------------
     sigma_u |  .11573186
     sigma_e |  .06071925
         rho |  .78415233   (fraction of variance due to u_i)
------------------------------------------------------------------------------
F test that all u_i=0: F(906, 6340) = 29.04                  Prob > F = 0.0000

.         _eststo lu5: xtreg lunch_part i.bicpost i.year, fe

Fixed-effects (within) regression               Number of obs     =      7,005
Group variable: schoolid                        Number of groups  =        907

R-squared:                                      Obs per group:
     Within  = 0.0305                                         min =          2
     Between = 0.0357                                         avg =        7.7
     Overall = 0.0018                                         max =          8

                                                F(8, 6090)        =      23.95
corr(u_i, Xb) = -0.0168                         Prob > F          =     0.0000

------------------------------------------------------------------------------
  lunch_part | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
   1.bicpost |  -.0022084   .0053663    -0.41   0.681    -.0127282    .0083114
             |
        year |
       2006  |   .0060649   .0030923     1.96   0.050     2.94e-06    .0121268
       2007  |   .0150901   .0030987     4.87   0.000     .0090156    .0211645
       2008  |    .024707   .0031243     7.91   0.000     .0185822    .0308317
       2009  |   .0265117   .0031399     8.44   0.000     .0203565    .0326669
       2010  |   .0108692   .0031003     3.51   0.000     .0047915     .016947
       2011  |   .0040942   .0031218     1.31   0.190    -.0020256    .0102141
       2012  |  -.0060021   .0031511    -1.90   0.057    -.0121793    .0001751
             |
       _cons |   .7703875   .0021972   350.63   0.000     .7660803    .7746947
-------------+----------------------------------------------------------------
     sigma_u |  .17289228
     sigma_e |  .06459134
         rho |  .87752268   (fraction of variance due to u_i)
------------------------------------------------------------------------------
F test that all u_i=0: F(906, 6090) = 55.50                  Prob > F = 0.0000

.
.         // with covariates (xtreg)
.         _eststo bk6: xtreg bkfast_part i.bicpost i.year $covars, fe

Fixed-effects (within) regression               Number of obs     =      7,255
Group variable: schoolid                        Number of groups  =        907

R-squared:                                      Obs per group:
     Within  = 0.3318                                         min =          7
     Between = 0.2925                                         avg =        8.0
     Overall = 0.2963                                         max =          8

                                                F(17, 6331)       =     184.93
corr(u_i, Xb) = 0.1284                          Prob > F          =     0.0000

-------------------------------------------------------------------------------
  bkfast_part | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
--------------+----------------------------------------------------------------
    1.bicpost |   .2219539   .0048512    45.75   0.000     .2124438    .2314639
              |
         year |
        2006  |   .0015038   .0028377     0.53   0.596     -.004059    .0070666
        2007  |   .0084261   .0028691     2.94   0.003     .0028017    .0140505
        2008  |   .0249954   .0029865     8.37   0.000     .0191408    .0308499
        2009  |   .0248059   .0030573     8.11   0.000     .0188125    .0307993
        2010  |   .0234917    .003163     7.43   0.000     .0172912    .0296922
        2011  |   .0153152   .0031559     4.85   0.000     .0091286    .0215018
        2012  |   .0173182   .0031775     5.45   0.000     .0110892    .0235472
              |
       pctell |   .0003668   .0002937     1.25   0.212    -.0002089    .0009424
```

```
     pctsped |   .0009591   .0003656     2.62   0.009     .0002423    .0016758
totalenroll~t |  -.0000933   9.50e-06    -9.82   0.000    -.0001119   -.0000747
    pctasian |   .0003216   .0004358     0.74   0.461    -.0005328     .001176
    pctblack |   .0002158   .0003691     0.58   0.559    -.0005078    .0009394
     pcthisp |   .0004245   .0003504     1.21   0.226    -.0002624    .0011114
   pctfemale |  -.0009431   .0004012    -2.35   0.019    -.0017296   -.0001567
       free1 |   .0005054   .0001067     4.74   0.000     .0002963    .0007145
       redu1 |   .0009194   .0002621     3.51   0.000     .0004055    .0014332
       _cons |   .2549082    .034423     7.41   0.000     .1874275    .3223888
-------------+----------------------------------------------------------------
     sigma_u |  .10212137
     sigma_e |  .05973931
         rho |   .7450423   (fraction of variance due to u_i)
------------------------------------------------------------------------------
F test that all u_i=0: F(906, 6331) = 20.55                  Prob > F = 0.0000

.         _eststo lu6: xtreg lunch_part i.bicpost i.year $covars, fe

Fixed-effects (within) regression               Number of obs     =      7,005
Group variable: schoolid                        Number of groups  =        907

R-squared:                                      Obs per group:
     Within  = 0.1003                                         min =          2
     Between = 0.3706                                         avg =        7.7
     Overall = 0.3365                                         max =          8

                                                F(17, 6081)       =      39.88
corr(u_i, Xb) = -0.0196                         Prob > F          =     0.0000

------------------------------------------------------------------------------
  lunch_part | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
   1.bicpost |   -.004206   .0052159    -0.81   0.420    -.0144309     .006019
             |
        year |
       2006  |   .0094849   .0030108     3.15   0.002     .0035827    .0153872
       2007  |   .0174954   .0030516     5.73   0.000     .0115132    .0234777
       2008  |   .0281332    .003202     8.79   0.000     .0218561    .0344102
       2009  |   .0313211   .0032831     9.54   0.000     .0248852    .0377571
       2010  |   .0183397   .0033532     5.47   0.000     .0117662    .0249133
       2011  |   .0104505   .0033547     3.12   0.002     .0038741    .0170268
       2012  |   .0004976   .0033807     0.15   0.883    -.0061297    .0071249
             |
      pctell |  -.0003212   .0003142    -1.02   0.307    -.0009372    .0002948
     pctsped |  -.0005976   .0003886    -1.54   0.124    -.0013594    .0001643
totalenroll~t |  -.0000728   .0000101    -7.19   0.000    -.0000927    -.000053
    pctasian |   .0041843   .0004575     9.15   0.000     .0032874    .0050812
    pctblack |   .0042077   .0003889    10.82   0.000     .0034453    .0049701
     pcthisp |   .0036614   .0003702     9.89   0.000     .0029357    .0043871
   pctfemale |  -.0011776   .0004275    -2.75   0.006    -.0020156   -.0003395
       free1 |   .0011066   .0001135     9.75   0.000     .0008841    .0013291
       redu1 |   .0016304   .0002758     5.91   0.000     .0010897    .0021711
       _cons |   .4648011     .03636    12.78   0.000     .3935227    .5360796
-------------+----------------------------------------------------------------
     sigma_u |  .13699664
     sigma_e |  .06226828
         rho |  .82878041   (fraction of variance due to u_i)
------------------------------------------------------------------------------
F test that all u_i=0: F(906, 6081) = 25.38                  Prob > F = 0.0000
```

```
.
.         // without covariates (xtdidreg)
.         xtdidregress (bkfast_part) (bicpost) , group(schoolid) time(year)

Treatment and time information

Time variable: year
Control:       bicpost = 0
Treatment:     bicpost = 1
-----------------------------------
             |   Control  Treatment
-------------+---------------------
Group        |
    schoolid |       800        107
-------------+---------------------
Time         |
     Minimum |      2005       2007
     Maximum |      2006       2012
-----------------------------------

Difference-in-differences regression                     Number of obs = 7,255
Data type: Longitudinal

                             (Std. err. adjusted for 907 clusters in schoolid)
------------------------------------------------------------------------------
             |               Robust
 bkfast_part | Coefficient  std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
ATET         |
     bicpost |
   (1 vs 0)  |   .2282953   .0169869    13.44   0.000      .194957    .2616335
------------------------------------------------------------------------------
Note: ATET estimate adjusted for panel effects and time effects.
Note: Treatment occurs at different times.

.         xtdidregress (lunch_part) (bicpost) , group(schoolid) time(year)

Treatment and time information

Time variable: year
Control:       bicpost = 0
Treatment:     bicpost = 1
-----------------------------------
             |   Control  Treatment
-------------+---------------------
Group        |
    schoolid |       800        107
-------------+---------------------
Time         |
     Minimum |      2005       2007
     Maximum |      2008       2012
-----------------------------------

Difference-in-differences regression                     Number of obs = 7,005
Data type: Longitudinal

                             (Std. err. adjusted for 907 clusters in schoolid)
------------------------------------------------------------------------------
             |               Robust
  lunch_part | Coefficient  std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
ATET         |
     bicpost |
   (1 vs 0)  |  -.0022084   .0080415    -0.27   0.784    -.0179905    .0135736
------------------------------------------------------------------------------
Note: ATET estimate adjusted for panel effects and time effects.
Note: Treatment occurs at different times and estimation sample contains units
      that switch in and out of treatment.
```

```
.
.         // with covariates (xtdidreg)
.         xtdidregress (bkfast_part $covars) (bicpost) , group(schoolid) time(ye
> ar)

Treatment and time information

Time variable: year
Control:       bicpost = 0
Treatment:     bicpost = 1
-----------------------------------
             |   Control  Treatment
-------------+---------------------
Group        |
    schoolid |       800        107
-------------+---------------------
Time         |
     Minimum |      2005       2007
     Maximum |      2006       2012
-----------------------------------

Difference-in-differences regression                    Number of obs = 7,255
Data type: Longitudinal

                              (Std. err. adjusted for 907 clusters in schoolid)
--------------------------------------------------------------------------------
              |               Robust
  bkfast_part | Coefficient  std. err.      t    P>|t|     [95% conf. interval]
--------------+-----------------------------------------------------------------
ATET          |
      bicpost |
    (1 vs 0)  |   .2219539    .016661    13.32   0.000     .1892551    .2546526
--------------------------------------------------------------------------------
Note: ATET estimate adjusted for covariates, panel effects, and time effects.
Note: Treatment occurs at different times.

.         xtdidregress (lunch_part $covars) (bicpost) , group(schoolid) time(yea
> r)

Treatment and time information

Time variable: year
Control:       bicpost = 0
Treatment:     bicpost = 1
-----------------------------------
             |   Control  Treatment
-------------+---------------------
Group        |
    schoolid |       800        107
-------------+---------------------
Time         |
     Minimum |      2005       2007
     Maximum |      2008       2012
-----------------------------------

Difference-in-differences regression                    Number of obs = 7,005
Data type: Longitudinal

                              (Std. err. adjusted for 907 clusters in schoolid)
--------------------------------------------------------------------------------
              |               Robust
   lunch_part | Coefficient  std. err.      t    P>|t|     [95% conf. interval]
--------------+-----------------------------------------------------------------
ATET          |
      bicpost |
    (1 vs 0)  |   -.004206   .0079187    -0.53   0.595    -.0197472    .0113352
--------------------------------------------------------------------------------
Note: ATET estimate adjusted for covariates, panel effects, and time effects.
Note: Treatment occurs at different times and estimation sample contains units
      that switch in and out of treatment.
```

```
.
.
. // Close log and convert to PDF
.         capture log close
```