EC327: Financial Econometrics, Spring 2014

Wooldridge, Introductory Econometrics (5th ed, 2012)

Chapter 13:
Pooling cross sections over time

In EC228, we have discussed regressions estimated from the two basic types of economic datasets: cross sections and time series. Empirical research is making broader use of richer forms of data that possess both cross-sectional and time dimensions. In this section of the course, we will discuss some simple econometric models which allow for pooling of cross sections over time. Data with both cross-sectional and time series characteristics can be usefully employed to answer many questions that we cannot address with data of one sort or the other.

There are two types of data that have both cross-sectional and time dimensions: *independently pooled cross sections* (IPCS) and *panel*, or *longitudinal* data. IPCS data represent the results of, e.g., a repeated survey. Consider a case where 5,000 likely voters are sampled each week by CNN to determine their satisfaction with the Bush administration's policies on Iraq, and we have 20 weeks of those survey data. We have 100,000 observations, each identified by a respondent code and the date on which it was collected. Although the respondent codes run from 1 to 5,000 each week, it is important to note that Mr. 1,234 this week has no relation to Mr. 1,234 next week or the week after. These are independent surveys, each representing a random sample from the population, which rules out correlation in the error terms within each survey's observations. Only by chance will the same individual appear more than once in the pooled dataset, and even if she does, we will not know it. Thus, we have 20 independent cross sections and may pool them into a single dataset of 100,000 observations.

In contrast, a *panel* dataset differs meaningfully from IPCS. When panel data are collected, information from the same individual units are recorded at each point in time. The original framework of this sort was a panel of forecasters who were surveyed each quarter. Today, panel data refer to not only individual-level surveys but any dataset with the characteristic of having t observations for each of i specific units. For instance, we could consider the quarterly GDP growth rates of each of the G-8 countries for the last 10 years, or annual financial data for each of the Dow Jones Industrials firms for each of the last 20 years. Panel data are quite easily assembled for many economic units: countries, states, cities, counties, school districts, firms, or their officers (e.g., CEOs) from readily available sources (see the *Guide to Economic and Financial Data at Boston College* link on the course home page). There are also a number of sizable panel surveys available from ICPSR the most celebrated of which is the Panel Study of Income Dynamics, a household survey that has been carried out for over 25 years.

Although panel datasets are much more useful in several ways than IPCS, they also bring complexity from an econometric standpoint. In a IPCS, each cross-section contains randomly selected individuals from the population at each point in time. Pooling those cross-sections does not lead to any correlation of observations' errors over time. However, when we work with panel data, we cannot assume that observations are independently distributed it over time. In individual-level data, the collection of unobservable factors that affect an individual's wage will be present at each point in time, leading to correlations across time that we call *unobserved heterogeneity*. Consequently, a number of econometric methods have been developed to deal with these features of the data.

Pooling independent cross sections over time

Surveys such as the *Current Population Survey* represent ICPS data, in the sense that a random sample of U.S. households is drawn at each time period. There are no links between households appearing in the sample in 2004 and those appearing in 2005. What are the advantages of pooling? We gain sample size, of course, which will increase the precision of estimators if the relationships being estimated are temporally stable. With that caveat, we can use IPCS to draw inferences about the population at more than a single point in time, and make inferences about how U.S. households behaved during the 1990s rather than just in 1995.

Temporal stability of any relationship may not be reasonable, so we often allow for some variation across time periods: most commonly, in the intercept term of the relationship for each time period, which can readily be accomplished with indicator variables. The coefficients of those indicator variables themselves may be of interest. Imagine that we had data comprised of random samples of 200 BC seniors from class years 2000, 2001, . . . , 2005. We know their graduating GPA their college, age, gender, first year GPA and SAT score on admission to Boston College. We can fit the equation

$$gradGPA_i = \beta_0 + \sum_{j=2001}^{2005} \beta_j Y_j + \beta_1 A\&S_i + \beta_2 Age + \beta_3 M_i + \beta_4 fyGPA_i + \beta_5 SAT_i + u_i$$

where the variable Y_{2001} equals 1 for those graduating in 2001, zero otherwise, and the variable $A\&S$ equals 1 for Arts & Sciences graduates and 0 for professional school graduates.

We are assuming that the effects of college, age, gender, first year GPA and SAT score on graduating GPA are constant over the six-year interval. How would we interpret the coefficient on $A\&S$? The indicator variables $Y_{2001} \ldots Y_{2005}$ allow the intercept in this relationship to shift over time. The intercept for the class of 2000 is β_0; the intercepts for each of the other years add their indicators' coefficients to that value. How do we interpret these intercept terms in the context of this equation? What does it mean to say that the intercept of the relationship is higher or lower in 2005 than it was in 2000?

The joint test of those indicator coefficients equalling zero considers the hypothesis that the intercept of this function is temporally stable. We perform that test conditional on the assumption that the other coefficients do not shift over time, which may be erroneous.

What would we conclude if some of the coefficients on the year indicator variables were significantly *positive*? What would this represent?

How could we test that the benefit (or burden) of being an $A\&S$ student varied over time? If we found that it did, how should we respecify the equation to take that variation into account? What would we conclude if a time-varying $A\&S$ coefficient was increasing between 2001 and 2005?

We could, of course, interact *all* of the regressors (college—SAT) with the year indicator variables. If we run the regression in this form, we get a single set of estimates, but each year has its own regression coefficients. This is quite similar to the notion of estimating the equation separately for each year. Why, then, would we run this regression? For one thing, we probably want to *test* whether the coefficients on a certain variable are time-varying. The only way to do that is in the context of this interacted regression. We might find, for example, that the effect of being in $A\&S$ differs significantly over time, but the effects of age or gender do not. If that is the case, then we should apply the constraints of constant coefficients (and drop the related interactions) from the model to gain efficiency. An F-test of all interaction terms being jointly zero will test the fully interacted model against the special case of that model in which all coefficients are temporally stable. If that F-test rejects its null, we should allow for some (or all) of the coefficients to vary over time.

The fully interacted model differs from a set of separate regressions in one important aspect: it assumes homoskedasticity throughout the IPCS. A set of separate regressions would generate a set of σ^2 estimates which would numerically differ, and might differ statistically. If they did, the pooled regression would suffer from groupwise heteroskedasticity, the groups being years. In the presence of groupwise heteroskedasticity, t- and F-tests based on the pooled regression will be invalid. We should test for that (e.g., `robvar` in Stata) and correct for it with feasible GLS or with robust standard errors to ensure that the tests mentioned above will be valid.

Policy analysis with pooled cross sections

IPCS can be very useful in analyzing the effects of policy changes or events. For instance, the construction of an expressway or a commuter rail line that reduces travel time to a distant suburb is likely to affect property values. Likewise, the establishment of a Wal-Mart nearby may have clear effects on local merchants' revenues and residents' wages. If we have a cross-section measure of these data collected prior to a change and another collected following a change, we may use regression techniques to disentangle the effects properly attributable to the change. We need not have the same units in those cross-sections; for instance, a house sale may be recorded in only one of the cross-sections, a store may exist in one sample but not the other, or a worker may relocate from or into the area between measurements.

Summary statistics in this context are dangerous. Property values may have increased for many reasons, as may the retail sales of local merchants or the local unemployment rate (for instance, a local manufacturing plant may have closed in the same year that a new Wal-Mart was hiring many local workers). Proponents of a given strategy (pro or con) may readily "lie with statistics" by using summary statistics or aggregate measures. How can regression on IPCS solve this problem?

Wooldridge provides the example of the construction of an incinerator in a neighborhood of North Andover, Mass. in 1981–1985. Data on housing values are available for 1978 (before planning for the incinerator was initiated) and for 1981 (after its location and likely effects were known). We would expect that house prices closer to the incinerator might be depressed. If an indicator variable `nearinc` measures proximity (e.g., if you live this close, you might smell something or have ash falling on your yard), we might naïvely regress

$$rprice_{1981} = \gamma_0 + \gamma_1 nearinc + u \qquad (1)$$

using the 1981 real prices of houses which sold that year. This regression is essentially the test for the difference of two means: those closer to and farther away from the incinerator. Does it establish that the incinerator reduced property values? By no means. It is likely that the incinerator was built in the less desirable section of North Andover, in which case we would expect that the preexisting homes would sell for less in that neighborhood in any event. Indeed, if we reestimate equation (1) using $\texttt{rprice}_{1978}$ as the response variable, we find that proximity to the incinerator lowered values in that year—even though there was no plan or rumor to build the unit at that time! Clearly, this strategy is not appropriate because it fails to take into account that housing prices are not randomly distributed through the town.

We have established that in both 1978 and 1981 the γ_1 coefficient is significantly negative. The marginal effect of proximity to the establishment of the incinerator during that interval is measured by the *change* in the coefficient between 1978 and 1981. So we want to allow γ_1 to change over the interval. We pool the two cross-sections and run a single regression:

$$rprice_{pooled} = \beta_0 + \gamma_0 d_{1981} + \beta_1 nearinc + \gamma_1 nearinc \cdot d_{1981} + u \quad (2)$$

with d_{1981} as an indicator variable set to 1 for the 1981 observations and 0 for the 1978 observations.

This equation essentially allows both intercept and slope of the housing price equation to shift during that time. It implements the *difference-in-differences* (DID) estimator:

$$\gamma_1 = (\bar{p}_{1981,n} - \bar{p}_{1981,f}) - (\bar{p}_{1978,n} - \bar{p}_{1978,f}) \quad (3)$$

where p is $rprice$, the real price of housing, and the subscripts n and f represent houses *near* and *far* from the incinerator, respectively.

At this point, we are estimating the means of four groups of houses from our pooled sample: each of the terms in equation (3). We would expect each parenthesized expression to be negative since the neighborhood in which the incinerator was sited appears to have lower property values, *cet. par.*, than the rest of the town. But the marginal effect of the incinerator is captured in the *difference* of these measured *differences*: the widening (or narrowing) of the gap between those means. The coefficient γ_1 in equation (2) computes that DID measure. If the incinerator depressed nearby property values, then we would expect γ_1 to be negative.

When Kiel and McClain (*JEEM*, 1995) ran this regression on 321 observations of the pooled sample, they found a negative but not significant coefficient. When they added additional characteristics, such as the age of the house, or the age and other characteristics (size, number of rooms and baths, etc.) they found a clearly significant negative coefficient, indicating that the establishment of incinerator did indeed depress housing values in the neighborhood even after controlling for a number of other factors.

If this equation was estimated with $\log(rprice_{pooled})$ as the dependent variable, the coefficient γ_1 becomes an estimate of the percentage effect. From this sample, the estimate (including other housing characteristics) becomes -0.132, or approximately 13%, with a significant t-ratio.

Control vs. treatment groups

The DID estimator is applicable to any situation where we can view the outcome of a *natural experiment*: the effect of an exogenous event, such as a policy change, on some economic variables that measure individuals' responses to that event. A natural experiment is characterized by a *control group*, not affected by the change, and a *treatment group* whose members are affected by the change.

Unlike a controlled experiment, in which the researcher may design the experiment to include randomly selected members of each group, a natural experiment requires the researcher to work with those observations generated by economic processes. For instance, the researchers of the North Andover incinerator study had no control over which houses would sell in 1978 or in 1981 and generate price data for the samples. To the extent that some families may have chosen to sell their houses in 1981 *due to* the negative amenity now in their neighborhood, we cannot consider housing transactions as random events within the town.

The control group–treatment group setup leads to a 2×2 table of categories: each group before and after the policy change. If we define indicator dT as defining membership in the treatment group and indicator $d2$ as defining membership in the after-treatment cross-section, we have the equation

$$y = \beta_0 + \gamma_0 d2 + \beta_1 dT + \gamma_1 dT \cdot d2 + u \qquad (4)$$

where we are likely to augment the equation with other explanatory factors which we may observe for each unit in the sample (for instance, the size or number of rooms in the house in the incinerator example).

This equation gives rise to the DID estimator with the coefficients characterized as γ_0 as the change in the mean of the control group and $(\gamma_0 + \gamma_1)$ as the change in the mean of the treatment group. These changes in the means of the two groups are the differences. Their difference—leading to DID—is γ_1. In application, we consider these changes to be in the conditional means of y, conditioned on the various other explanatory factors that we include in the equation.

We might, for instance, want to calculate the impact of an increase on the cigarette tax on consumption in one state. Smokers from that state are the treatment group. In an adjoining state, no change to cigarette taxes was implemented during that period; smokers from that state form the control group. We would include a number of demographic factors in each random sample to control for the possibly different composition of the population of smokers in each state. For instance, it might be the case that median incomes in the treatment group are lower than those in the control group, so that the effect of the tax might be a greater impact on their budgets. We also would want to assume that cross-border sales are not important, as they are in some cases where tax treatments differ sizably across state borders.

We now turn from the analysis of IPCS to the simplest kind of panel data: the case in which we have two successive measurements on a cross-section of individual units.

Two-period panel data analysis

In its simplest form, panel data refers to measurements of $y_{i,t},\ t = 1, 2$: two cross-sections on the same units, $i = 1, \ldots, N$. Say that we run a regression on one of the cross-sections. Any regression may well suffer from omitted variables bias: there are a number of factors that may influence the cross-sectional outcome, beyond the included regressors. One approach would be to try to capture as many of those factors as possible by measuring them and including them in the analysis. For instance, a city-level analysis of crime rates versus the city's level of unemployment might be augmented with control variables such as city size, age distribution, gender distribution and ethnic makeup, education levels, historical crime rates and so on.

An alternative approach would consider many of these city-specific factors as *unobserved heterogeneity*, and use panel data from repeated measurements on the same city to capture their net effects. Some of those factors are time-invariant (or approximately so), while some city-specific factors will change over time. We can deal with all of the time-invariant factors if we have at least two measurements per city by considering them as *individual fixed effects*. Likewise, the net effect of all time-varying factors can be dealt with by a *time fixed effect*.

For a model (such as crime rate vs. unemployment rate) with a single explanatory variable,

$$y_{it} = \beta_0 + \gamma_0 d2_t + \beta_1 X_{it} + a_i + u_{it} \qquad (5)$$

where the indicator variable $d2_t$ is zero for period one and one for period two, not varying over i. Both the y variable and the X variable have both i and t subscripts, varying across (e.g.) cities and the two time periods, as does the error process u.

For the crime example, coefficient γ_0 picks up a macro effect: for instance, crime rates across the U.S. may have varied, on average, between the two time periods. The individual time effect picks that up.

The term a_i is an individual fixed effect, with a different value for each unit (city) but not varying over time. It picks up the effect of everything beyond X that makes a particular city unique, without our having to specify what those factors might be.

How might we estimate equation (5)? If we merely pool the two years' data and run OLS we can derive estimates of the β and γ parameters, but are ignoring the a_i term, which is being included in the composite error term $v_{it} = a_i + u_{it}$. Unless we can be certain that $E(v_{it}|X_{it}) = 0$, the pooled approach will lead to biased and inconsistent estimates. This zero conditional mean assumption states that the unobserved city-specific heterogeneity must not be correlated with the X variable: in the example, with the unemployment rate. But if a city traditionally has suffered high unemployment and a shortage of good jobs, it may also have historically high crime rates. This will imply that this correlation is very likely to be nonzero, and OLS will be biased.

This same argument applies if we use a single cross section; as we described above, we would be likely to ignore a number of important quantifiable factors in estimating the simple regression from a cross-section of cities at one point in time.

Therefore, we apply a strategy that will allow for the presence of unobserved heterogeneity and deal with it appropriately. There are two approaches which we might follow, as we now develop.

The first difference model

If we take the first difference of equation (5), we arrive at

$$\Delta y_{it} = \gamma_0 + \beta_1 \Delta X_{it} + \Delta u_{it} \tag{6}$$

where Δ refers to the first difference operator, $z_t - z_{t-1}$. When we difference the units vector multiplying β_0, we get a vector of zeroes.

When we difference the vector $d2_t$ for each city, we get 1, so that γ_0 now becomes the intercept for this equation. When we difference the units vector multiplying a_i, we get a vector of zeroes—so that the unobserved heterogeneity term disappears in the differencing process, solving the problem.

This first difference equation may be consistently estimated with OLS given the usual zero conditional mean assumption: in this context, that $E(\Delta u_i | \Delta X_i) = 0$. If X is strictly exogenous, this assumption will be satisfied. If it is weakly exogenous, it may be harder to establish this assumption. In particular, this assumption rules out the case where a set of Xs includes a lagged dependent variable, y_{t-1}.

A second condition must be satisfied for equation (6) to be estimated: there must be time variation in each of the X variables. Any time-invariant effect will be captured by the a_i term, and differenced out. We can only include a single time-invariant term for each unit in equation (5), in the form of a_i. If we consider a panel dataset of individual-level data, this implies that time-invariant characteristics such as gender or race cannot be included among the regressors in X, since when differenced they will disappear. As a corollary, regressors with minimal time variation will be problematic. We may include them in the regression, but they are likely to have little explanatory power, since their differences will have a small variance.

It is important to note that equation (6) refers to a model in which the objective is no longer the explanation of the variation in y_{it} across units and time, but rather the explanation of the variation in Δy_i across units: that is, why did some cities experience a large increase in the crime rate between the two periods, while others enjoyed a decline? Likewise, the regressors only provide an explanation of that phenomenon in terms of their changes over time. Although coefficient $\beta_1 = \partial \Delta y / \partial \Delta X$, it also equals the original $\partial y / \partial X$ from equation (5). X will only play an important role if its *changes* are systematically related to *changes* in y.

The model generalizes to the case where we have multiple time-varying explanatory factors in X, including the case where there may be measurements for several past periods. For instance, we might include the current unemployment rate and two of its own lags in X. This would require that we gather city-specific data for this explanatory variable for four periods, since we would model $crime_{it}$ as a function of $unemp_{it}, unemp_{i,t-1}, unemp_{t-2}$ for $t = 1, 2$.

We may estimate this model, in its general form, with panel data that have been identified as such by `tsset`, using the `D.` operator to specify the first differences of the variables in the `regress` command. Alternatively, we may use the user-written Stata command

`xtivreg2` *depvar indepvars* , `fd`

with the `fd` option specifying the first-differenced model. We need not have an instrumental variables problem in order to use this command.

Organization of panel data

For most panel data applications, we want to organize the data in what Stata calls the *long* format, in which data are stacked by panel. There are two ways in which data indexed by both i and t subscripts can be organized: in the *wide* format, where variables from different units are stored next to each other and named by the unit to which they belong: e.g., `GDPGermany`, `GDPFrance`, etc.; or the *long* format, in which a single variable `GDP` is stored as the timeseries for Germany, followed by that for France, and so on. The long format will naturally arise if you have data organized on different spreadsheets, one for each unit, and combine them vertically. However, there are a number of instances where the data are provided in wide format, but you want to make use of them with panel data techniques in long format. In this case, you should use Stata's `reshape` command, which can either `reshape long` or `reshape wide`, depending on the original form of the data.

To reshape wide-format data into the long format, you must identify the time-series calendar variable and have variable names for each unit in some systematic form. For instance, if you have variables `GDPGermany, GDPFrance` you can specify that all variables named `GDP...` are to be reshaped. But if you have named the variables `GermanGDP` and `GDPFrance`, they will have to be renamed in order to use `reshape`.

To reshape data from long into wide format (for instance, to produce a comparison table or certain graphs) you must identify both the time-series calendar variable and the panel unit identifier. These two variables are those used in the `tsset` command to instruct Stata that this is panel data: e.g., `tsset` *panelvar datevar*. If you are manually combining data from different panel units (for instance, in a spreadsheet environment) be sure to create the panel variable before performing the combination. You can also use Stata's `append` command to combine Stata-format data files for different units, but again it is important to have a panel identifier variable in each Stata-format data file before doing the `append`.

The graphics command `xtline` is useful in that it can produce line graphs of time series for different panels with the data organized in long format. For instance:

```
webuse grunfeld,clear
generate km = kstock/mvalue
xtline km if company < 5
```

A number of Stata's graphics commands (including `tsline` will work with data in wide format.

Policy analysis with two periods of panel data

Panel data sets—even with only two observation periods per panel unit—are very useful for policy analysis and program evaluation. They differ from IPCS in that we observe the same individuals in both (or several) periods. Some of these individuals are not affected by a particular program; they are the control group. Other individuals are affected: the treatment group. Wooldridge uses the example of a job training program's effect on worker productivity. The unit of observation is not the worker but the firm, since job training grants were given to specific firms. Productivity is measured by the scrap or defect rate. The more productive are workers, the fewer defective products come off the assembly line. We measure firms' attributes in 1987 and 1988, leading to the equation

$$scrap_{it} = \beta_0 + \gamma_0 d_{1988} + \beta_1 grant_{it} + a_i + u_{it} \quad (7)$$

where d_{1988} is an indicator variable for 1988 observations, and $grant_{it}$ is an indicator variable for those firms who received grants in 1988 when the government program was initiated.

The difficulty here is the *unobserved heterogeneity* at the firm level. Some firms will be likely to have a higher or lower defect rate in both years, as captured by the parameter a_i. We can remove this effect by differencing, yielding

$$\Delta scrap_{it} = \gamma_0 + \beta_1 \Delta grant_{it} + \Delta u_{it} \qquad (8)$$

The original intercept is removed by differencing, and since the difference of d_{1988} is 1 for all differenced observations, the coefficient γ_0 becomes the constant term. The term $\Delta grant_{it}$ is 1 for those firms in the treatment group and 0 for those firms in the control group. The coefficient β_1 is negative (as theory predicts) but not statistically significant. If the dependent variable is expressed as $\log(scrap_{it})$, it becomes statistically significant and implies an approximately 27.2% reduction in the scrap rate.

If we ignore the issue of unobserved heterogeneity and estimate equation (7) without the a_i term using pooled OLS, we find a positive and insignificant effect of the job training program. Since this differs so meaningfully from the first-difference estimates, it seems clear that firms are not randomly selected into the treatment group. As one would hope, firms with lower-ability workers are more likely to receive a job training grant.

In general terms, the strategy of program evaluation can be written as

$$y_{it} = \beta_0 + \gamma_1 d2_t + \beta_1 prog_{it} + a_i + u_{it} \quad (9)$$

where $d2$ is an indicator for the post-treatment period and $prog_{it}$ is an indicator of participation in the program under study. If units only participated in the post-treatment period, we find

$$b_1 = \overline{\Delta y_{treat}} - \overline{\Delta y_{control}} \quad (10)$$

where we calculate the average change in y over the two time periods for the treatment and control groups; the treatment effect is the difference between those differences. Thus, we have a panel version of the DID estimator, with the important advantage that we can difference y values for the same individuals over the time periods. If there are additional time-varying factors for which we want to control, we merely difference them and include them in the estimated equation. All time-invariant factors are captured in the a_i term and differenced out.

The first difference model with more than two time periods

We can also apply the FD model for three or more time periods; if we have T observations per individual, we will have $N(T-1)$ observations in the differenced data set, assuming a balanced panel. We would include a separate intercept for each time period. If the unobserved heterogeneity term a_i is correlated with any of the explanatory variables, pooled OLS on the levels dataset will yield biased and inconsistent results.

The key assumption in applying the FD model in this context is that $cov(x_{itj}, u_{is}) = 0 \; \forall \; t, s, j$. That is, the idiosyncratic errors attached to each observation (unit and time period) are uncorrelated with past, present and future values of the set of explanatory variables x, which are assumed to be strictly exogenous. This rules out the case where one of the explanatory variables is lagged y, since it cannot be strictly exogenous with respect to u_{it}.

Why is pooled OLS, ignoring a_i, inconsistent? Because a_i is likely to be correlated with one or more elements of x, which implies that x will be correlated with the *composite error* $(a_i + u_{it})$. We eliminate this correlation in the FD model by differencing a_i out of the equation, which can then be estimated by pooled OLS. If we start with the levels equation for three periods

$$y_{it} = \gamma_1 + \gamma_2 d2_t + \gamma_3 d3_t + \beta_1 x_{it1} + \ldots + \beta_k x_{itk} + a_i + u_{it} \tag{11}$$

where γ_1 is the intercept for period 1, $\gamma_1 + \gamma_2$ is the intercept for period 2, and $\gamma_1 + \gamma_3$ is the last period's intercept. When differenced, this equation yields

$$\Delta y_{it} = \gamma_2 \Delta d2_t + \gamma_3 \Delta d3_t + \beta_1 \Delta x_{it1} + \ldots + \beta_k \Delta x_{itk} + \Delta u_{it} \tag{12}$$

Note that $\Delta d2_t$ takes on a value of 1 for period 2 and -1 for period 3, while $\Delta d3_t$ is 0 for period 2 and 1 for period 3. There is no intercept in this formulation, but since we can write $\gamma_2 \Delta d2_t + \gamma_3 \Delta d3_t$ in terms of an intercept α_0 and a term $\alpha_1 d3_t$, we may prefer this form so that an intercept can be included in the pooled OLS regression. With more than three time periods, we would include an intercept and $T - 2$ time indicators (for all but the first two periods).

Although the FD model may be quite useful in evaluating program outcomes in multi-period panel datasets, we should be aware that there are pitfalls as well. Variables with little time variation will not be useful in such a model, since their differences may be quite small. If explanatory variables are measured with error, their differences may be overpowered by *differences* in measurement error over time. In that respect, the FD model may be worse than pooled OLS in terms of bias. Nevertheless, the model has been applied in many contexts where it has proven to be quite successful: see the textbook examples (13.8, 13.9) of the effects of enterprise zones on local unemployment rates, and the effects of various policy changes on local crime rates.