

13.4 Policy Analysis with Two-Period Panel Data

1. **Michigan Job Training Program**: recall the data set which we used in Stats 2. This was a model in which we explain the effect of job training on scrap rate – the proportion of products which were deemed unfit for sales.

If there were grants which improved worker performance and reduced scrap, we could examine the effect of such grants. Grant program started in 1988, so we could use 1987 as control group and estimate the following model.

$$scrap_{it} = \beta_0 + \delta_0 y88_t + \beta_1 grant_{it} + a_i + u_{it}$$

- (a) a_i contains the **unobserved firm effect** or the firm fixed effect such as average employee ability, capital, managerial skills etc. If these a_i is correlated with the ability to avail grant, then we need to be careful. What if grants were given to firms which had less trained workforce.

- (b) **Selection Bias:** These grants were applied on a first come first serve basis. What if the firms which applied early could have more productive managers? Again a_i correlated with $grant_{it}$. We could take a first difference:

$$\Delta scrap_i = \delta_0 + \beta_1 grant_i + \Delta u_i$$

When we estimate this model (13.31) we know that the grant lowers scrap rate on an average by -0.739, but we obtain a large standard error. We take log of the scrap rate to estimate the percentage effect. We obtain the following

$$\begin{aligned}\widehat{\Delta \log(scrap)} &= -.057 - .317 \Delta grant \\ & (.097) \quad (.164) \\ n &= 54, R^2 = .067.\end{aligned}$$

This implies that the grant reduced a scrap rate by 27.2 %. This statistic is marginally significant. Also note that comparing pooled model and FD model tells us something about the bias.

2. **Program Evaluation:** Let y_{it} denote any outcome variable which is affected by $prog_{it}$, a participation dummy variable. The simplest unobservable model is

$$y_{it} = \beta_0 + \delta_0 d_{2t} + \beta_1 prog_{it} + a_i + u_{it}$$

When we remove the unobserved effect through differencing, the estimate of the program has a simple representation:

$$\hat{\beta}_1 = \overline{\Delta y_{treat}} - \overline{\Delta y_{control}}$$

This means that the difference in average change in y over two time periods for the two groups gives us a version of the difference in difference estimator. Now, instead of program which varies across two time periods, we could have program vary across two state or cities too.

3. **Effect of DRunk Driving LaWs on TrAffic FaTaLitiEs:** There are two types of drunken driving laws: (a) Open container laws in which it is illegal to have open containers of alcoholic beverages in vehicles and (b) Administrative Laws: which allows arresting for drunk driving before a formal conviction.

There is a possibility to test these programs because different states implemented these laws in different years. How do these laws affect traffic fatalities? Lets consider a model:

$$\Delta dthrte = \beta_0 + \delta_0 \Delta open + \delta_1 \Delta admn + \Delta u_{it}$$

where *dthrte* is the number of traffic deaths per 100 milion miles driven. Note that in 1985, 19 states had open container laws and 21 had admin laws. This increased to 22 and 29 respectively.

Let us estimate this model:

$$\widehat{\Delta dthrte} = -.497 - .420 \Delta open - .151 \Delta admn$$

$$(.052) \quad (.206) \quad (.117)$$

$$n = 51, R^2 = .119.$$

We obtain that open container laws reduced traffic fatalities by 0.42 when the average death rate was 2.7. That is a statistically significant lowering of traffic fatalities, significant at 5%.

13.5 Differencing with More Than Two Time Periods

1. **General Model for Policy Change:** Consider a model where i = individuals and $T = 3$.

$$y_{it} = \delta_1 + \delta_2 d2_t + \delta_3 d3_t + \beta_1 x_{it1} + \beta_2 x_{it2} + \cdots + \beta_k x_{itk} + a_i + u_{it}$$

Notice that we have T-1 time dummies for T time periods. If your data permits you to have this, then time dummies for all time periods is effective.

2. **Consistent estimation:** requires that the errors are uncorrelated with all the x 's i.e. explanatory variables are strictly exogenous after individual level unobservable is accounted for:

$$Cov(x_{itj}, u_{is}) = 0 \quad \forall t, s \text{ and } j$$

Note that this tells us that x_{it+1} should not react to u_{it} too. Omitting important time varying explanatory variable leads to violation of the strict exogeneity.

3. **First differencing:** We can eliminate a_i if we suspect it is causing

$$Cov(x_{itj}, v_{is}) \neq 0$$

where v_{is} is a composite error and $t = 2, 3$

$$\Delta y_{it} = \delta_2 \Delta d2_t + \delta_3 \Delta d3_t + \beta_1 \Delta x_{it1} + \cdots + \beta_k \Delta x_{itk} + \Delta u_{it}$$

The only requirement for consistent OLS estimation is that Δu_{it} is uncorrelated with Δx_{itj} . Note that with 3 time periods, after differencing we have same individual with data on 2 time periods after differencing.

- (a) Notice that the first difference model does not have an intercept. Some regression packages automatically add an intercept giving you correct R-squared.
- (b) Another solution is to compute the following model. This would give correct estimates for β and R-squared.

$$\Delta y_{it} = \alpha_0 + \alpha_3 d3_t + \beta_1 \Delta x_{it1} + \cdots + \beta_k \Delta x_{itk} + \Delta u_{it}$$

4. We can extend this idea to larger time periods. When we have same T time periods for each N cross sectional units, we say that the data set is a **balanced panel**. When T is relatively small to N, include a dummy for T-1 time periods and then do the first differencing.

$$\Delta y_{it} = \delta_2 \Delta d_{2t} + \delta_3 \Delta d_{3t} + \cdots + \delta_T \Delta d_{Tt} + \beta_1 \Delta x_{it1} + \cdots + \beta_k \Delta x_{itk} + \Delta u_{it}$$

- (a) To estimate this model using pooled OLS, arrange data in wide form and carefully difference the data.
- (b) We assume that Δu_{it} is uncorrelated overtime for inference to be valid. We could test this by regressing $\Delta u_{it} = \rho \Delta u_{it-1} + w_{it}$ and testing for $H_0 : \rho = 0$.
- (c) Even if there were problems of serial correlation and heteroskedasticity, we could use cluster robust standard errors. This could solve inference problem if we have small T.
- (d) Alternatively, we could use FGLS to incorporate serial correlation and make AR(1) corrections to the original regression.

13.5.1 An Example

1. **Effect of Enterprise Zones on Unemployment Claims** as in **Boarnet and Bogart (1996)** and **Papke (1993)**
2. Enterprise Zones are areas where tax benefits are given to firms so that they employ more people. In Indiana, not all cities were enterprise zones and hence served as control groups to this natural experiment. A policy evaluation model was:

$$\log(uclms)_{it} = \theta_t + \beta_1 ez_{it} + a_i + u_{it}$$

where $uclms$ is the number of unemployment claims filed during year t in city i . θ_t denoted time dummies and ez_{it} is a dummy for enterprise zone.

3. The unobserved effect a_i represents fixed factors that affect the economic climate in city i i.e. city fixed effects.

4. Note that enterprise zones are not allotted randomly since they are likely to be in economically depressed areas, a_i is correlated with ez_{it} . Thus we difference the equation to eliminate it:

$$\Delta \log(uclms_{it}) = \delta_1 \Delta d81_t + \delta_2 \Delta d82_t + \cdots + \delta_8 \Delta d88_t + \beta_1 \Delta ez_{it} + \Delta u_{it}$$

5. Estimates from the data EZUNEM tell us that $\hat{\beta}_1 = -0.182$ with a t-stat of -2.33.
- (a) However there is serial correlation in the OLS residuals. This implies that we need to use cluster-robust standard errors.

13.5.2 A Second Example

1. County Crime Rates in North Carolina: following Cornwell and Trumbull (1994) examine crimes in 90 counties in North Carolina.

$$\begin{aligned}
 \widehat{\Delta \log(\text{crmrte})} = & .008 - .100 \, d83 - .048 \, d84 - .005 \, d85 \\
 & (.017) \, (.024) \quad (.024) \quad (.023) \\
 & [.014] \, [.022] \quad [.020] \quad [.025] \\
 & + .028 \, d86 + .041 \, d87 - .327 \, \Delta \log(\text{prbarr}) \\
 & (.024) \quad (.024) \quad (.030) \\
 & [.021] \quad [.024] \quad [.056] \\
 & - .238 \, \Delta \log(\text{prbconv}) - .165 \, \Delta \log(\text{prbpris}) \quad [13.40] \\
 & (.018) \quad (.026) \\
 & [.040] \quad [.046] \\
 & - .022 \, \Delta \log(\text{avgssen}) + .398 \, \Delta \log(\text{polpc}) \\
 & (.022) \quad (.027) \\
 & [.026] \quad [.103] \\
 n = 540, R^2 = .433, \bar{R}^2 = .422.
 \end{aligned}$$

We difference the data to eliminate county level fixed effects a_i . *crmrte* is number of crimes per capita, *prbarr* is the estimated probability of arrest, *prbconv* is the estimated probability of conviction, *prbpris* is probability of serving time in prison, *avglen* is the average length of sentence served and *polpc* is the number of police officers per capita. Since this is a log-log model, the estimated β are elasticities.

2. **Interpretation:** An 1% increase in number of police officers per capita increases crime by 3.98%. There are two ways to think about it. First, this is true and we need to understand why. Second, there is some issue with the regression.
 - (a) When there are more police officers then there is less measurement error in crime i.e. more crimes are reported.
 - (b) Police force might be endogenous – counties with higher police presence are historically counties with more force. In this case the estimates are unreliable.

3. Even after using cluster robust standard errors, the variables are significant but they defy economic intuition. This might be harder to detect in other examples.
4. [Tangent about interactions and first difference](#): The idea is that with slope interactions, first differencing becomes tricky and not very straight forward. (page 450-451)

13.5.3 Potential Pitfalls in First Differencing Panel Data

1. Strict Exogeneity: is a very strong assumption and having more time periods does not overcome endogeneity.
2. If one or more variables are subject to measurement errors, then their difference would be subject to a more serious measurement error.