ØAMET4100 · Spring 2019
Lecture Note 6

Instructor: Fenella Carpena

February 14, 2019

This lecture note provides a review of regression with panel data (Stock & Watson, Chapter 10). This lecture note is not intended to be a comprehensive review of lecture or the textbook, since there is a lot more material than we have time to cover. However, I have tried to focus on the concepts which I believe are necessary to be successful in our class.

1 Example: Traffic Deaths and Beer Taxes

For this lecture, we will use the following example in the US context. Drunk driving is a major cause of fatal car crashes (e.g., as much as 25% of all fatal car crashes involve a driver who had been drinking). To address this issue, many states have implemented various government policies, such as taxing beer. Our goal is to investigate whether beer taxes—which is designed to discourage drunk driving—are indeed effective in reducing traffic deaths.

We have panel data on 48 states for the years 1982 to 1988. Two of the variables in the data are: (1) fatality rate, which is the number of traffic deaths per 10,000 people in each state and year; and (2) beer tax, which is the tax in dollars per case of beer (adjusted for inflation). How does this data look like? If we were to open it like a "spreadsheet," it would look like this:

State	Year	Fatality Rate	Beer Tax
Alabama	1982	2.128	1.539
Alabama	1983	2.348	1.789
Alabama	1984	2.336	1.714
Alabama	1985	2.193	1.653
Alabama	1986	2.669	1.610
Alabama	1987	2.719	1.560
Alabama	1988	2.494	1.501
Arizona	1982	2.499	0.215
Arizona	1983	2.267	0.206
Arizona	1984	2.829	0.297
Arizona	1985	2.802	0.381
Arizona	1986	3.071	0.372
Arizona	1987	2.767	0.360
Arizona	1988	2.706	0.347

As you can see in the above table, in panel data, each entity (e.g., Alabama, Arizona) is observed for multiple time periods. This means that we will need two subscripts to denote each observation, since **one observation is one particular entity at a particular point in time**. The subscript i is used for the entity that we are following over time (in our example, these are states), and $i = 1, ..., n$ where n denotes the total number of entities that we have. The subscript t is typically used to denote time periods, and $t = 1, .., T$ where T denotes the total number of time periods we observe. If we have a **balanced panel**, then the data will have $n \cdot T$ total observations. Note that in the dataset described above, $n = 48$ (because our panel data has 48 states) and $T = 7$ (because we observe seven years, from 1982 to 1988).

Why do we care about panel data? We care because it allows us to overcome omitted variable bias. It allows us to control for some types of omitted variables *even without actually observing them.*

2 Regressions with Entity Fixed Effects

2.1 Overview

Regressions with entity fixed effects allows us to control for omitted variables that vary across entities (e.g., states) but do not change over time. This type of model has n different intercepts, one for each entity. These intercepts can be represented by a set of binary variables. These binary variables absorb the influences of all omitted variables that differ from one entity to the next but are constant over time.

Consider the regression model

$$FatalityRate_{it} = \beta_0 + \beta_1 BeerTax_{it} + \beta_2 Culture_i + u_{it} \tag{1}$$

We want to estimate β_1, which is the effect of the beer tax on the traffic fatality rate, holding culture constant. In other words, in this regression, we would like to control for $Culture_i$, which represents the local cultural attitude in state i towards drinking and driving. If we did not include it in the regression, the regression would suffer from omitted variable bias. Why? Recall the two conditions for OVB: (1) cultural attitudes is correlated with the beer tax (e.g., some states in the US are more religious than others, and these states might want higher taxes on beer); and (2) cultural attitudes is correlated with the fatality rates (e.g., some states which are more rural may have a culture of driving big pick-up trucks). Hence, $Culture_i$ is an important omitted variable that should be included in the regression.

There are two key aspects here. First, cultural attitudes (unlike variables such as education and income) is an unobservable characteristic. So even if we tried very hard, we may not be able to find data that will allow us to include $Culture_i$ in the sample regression. We want to control for $Culture_i$ to avoid omitted variable bias, but $Culture_i$ cannot be measured. What can we do about this? As we will see, this is where panel data can help us.

Second, $Culture_i$ is constant over time for a given state: local cultural attitudes toward drinking and drive changes slowly, and thus could be considered constant between 1982 and 1988 in a particular state. This means regression equation (1) above can be interpreted as having n intercepts, one for each state. Specifically let $\alpha_i = \beta_0 + \beta_2 Culture_i$. Then, we can re-write equation (1) as

$$FatalityRate_{it} = \alpha_i + \beta_1 BeerTax_{it} + u_{it}. \tag{2}$$

Equation (2) is what is called the **fixed effects regression model**, in which the parameters α_1, α_2, …, α_n are treated as unknown intercepts to be estimated (i.e., one for each state).

The intercepts α_i's can be thought of as the "effect" of being in state i, so the terms α_1, α_2, …, α_n are called **entity fixed effects**. It does not have a subscript t because it does not vary of time. How do we interpret α_i? It contains **all** state-specific characteristics that affect fatality rate but are **not changing over time** (such as the local culture, political ideology, etc.).

What panel data allows us to do is the estimate the α_i's. This means that if we had panel data, we are able to control for $Culture_i$, as well as all other state-specific constant characteristics—even if those characteristics are unobservable!

2.2 Estimation

In practice, how can we estimate a regression with entity-specific intercepts, as in the regression specified in equation (2) above? We will discuss two potential approaches: first, using dummy variables, and second, using the "entity-demeaned" or "within" transformation.

Dummy Variables. The state-specific intercepts in the fixed effects regression model can also be expressed using binary variables to denote the individual states. Let $DState1_i$ be a binary variable that equals one when $i = 1$ (i.e., for the state of Alabama), $DState2_i$ be a binary variable that equals one when $i = 2$ (i.e., for the state of Arkansas), etc. The fixed effects regression in equation (2) can be written equivalently as

$$FatalityRate_{it} = \beta_0 + \beta_1 BeerTax_{it} + \gamma_2 DState2_i + \gamma_3 DState3_i + \ldots + \gamma_{48} DState48_i + u_{it}. \tag{3}$$

This procedure is also known as the **Least Squares Dummy Variable (LSDV) method.**

Note that because of the dummy variable trap, we cannot include the common intercept β_0 and all 48 dummy variables in the regression. Hence, the state dummies that we included in the regression equation above are only for State 2 to State 48 (i.e., we omitted the dummy variable for State 1). Equation (3) can then be estimated using the usual OLS approach as we have learned before. That is, OLS would choose $\widehat{\beta}_0$, $\widehat{\beta}_1$, $\widehat{\gamma}_2$, $\widehat{\gamma}_3$, ..., $\widehat{\gamma}_n$ to minimize the sum of squared residuals:

$$\sum_{i=1}^{n}\sum_{i=1}^{T}\left(Y_{it}-\widehat{\beta}_0-\widehat{\beta}_1 BeerTax_{it}-\widehat{\gamma}_2 DState2_i-\ldots-\widehat{\gamma}_{48}DState48_i\right)^2 \tag{4}$$

What is the relationship between equation (3) and equation (2)? In equation (2), we have n state-specific intercepts, but in Equation (3), we have a common intercept β_0 and $n-1$ coefficients on the dummy variables. In both equations, the slope for the effect of the beer tax, β_1, is the same. Further, note that α_1 in equation (2) maps to β_0 in equation (3), and α_2 in equation (2) maps to $\beta_0+\gamma_2$ in equation (3), etc. Hence, equations (2) and (3) are equivalent. The variation in the state-specific intercepts α_i's and the the coefficients of the dummy variables γ_i's have the same source: the unobservable characteristics (e.g., local culture) that differ across states but are constant over time.

Entity-Demeaned or Within Transformation. In the example we are using here, n is 48. But in many panel data applications, the number of entities n is very large. When n is large, the OLS regression with $n-1$ dummy variables as in equation (3) can be very tedious. Econometrics software such as Stata therefore use the "Entity-Demeaned" or "Within Transformation" for OLS estimation of fixed effects regression models. This approach is equivalent to putting dummy variables for fixed effects, but it is computationally faster because it employs some mathematical simplifications.

The Entity-Demeaned approach proceeds in two steps. In the first step, we "demean" Y_{it} (i.e., the variable $FatalityRate_{it}$) and X_{it} (i.e., the variable $BeerTax_{it}$). Then, in the second step, we use OLS on the demeaned variables.

Let's now consider the first step. In regression equation (2), we can take the average of both sides of the equation to obtain

$$\overline{FatalityRate}_i=\beta_1\overline{BeerTax}_i+\alpha_i+\overline{u}_i \tag{5}$$

where $\overline{FatalityRate}_i=\frac{1}{T}\sum_{t=1}^{T}FatalityRate_{it}$ is the entity mean (and similarly for $\overline{BeerTax}_i$ and $\overline{u}_i$).

If we subtract equation (5) from equation (2), we get

$$\widetilde{FatalityRate}_{it}=\beta_1\widetilde{BeerTax}_{it}+\widetilde{u}_{it} \tag{6}$$

where $\widetilde{FatalityRate}_{it}=FatalityRate_{it}-\overline{FatalityRate}_i$ is the demeaned variable. Similarly for $\widetilde{BeerTax}_{it}$ and $\widetilde{u}_{it}$ are the demeaned variables.

In the second step, we can use OLS to estimate equation (6).

What have we just done? Note that by demeaning the variables in equation (6), the term α_i dropped out from the regression. This makes the coefficients of the sample regression easier to compute because we do not have to calculate the value of α_i for each entity i. This is especially helpful when we have a large number of entities (think millions or billions). The punchline here is that by using the demeaned regression, we are still able to control for the fixed effects α_i without having to estimate these coefficients.

3 Regressions with Time Fixed Effects

In the previous section, we considered entity fixed effects that can control for characteristic that are constant over time, but differ across entities. Now, we will look time fixed effects, which control for **time fixed effects** that control for variables that are constant across entities, but change over time. One such variable is national car safety standards. These safety standards change over time, but it is constant

across states because it is a national standard.

For the moment, let's ignore the effect of $Culture_i$ on traffic deaths to simplify the explanation. A **time fixed effects regression model** takes the form

$$FatalityRate_{it} = \lambda_t + \beta_1 BeerTax_{it} + u_{it} \tag{7}$$

In the entity fixed effects model, each state had its own intercept. We have a similar case here with time fixed effects. Because λ_t varies over time but not over states, each time period has its own intercept.

How do we interpret the parameters λ_t? They can be thought of as the "effect" of year t on fatality rates. The terms $\lambda_1, \lambda_2, \ldots, \lambda_T$ are know as **time fixed effects**. Time fixed effects would contain time trends that influence traffic fatality rates, but are the same across states. Examples might include the national unemployment rate, national gross domestic product, and other national, US-wide economic trends.

As with the entity fixed effects model, we can use the LSDV and the within transformation (using the time-period means) to estimate a time fixed effects model.

For the LSDV, let $DYear1_t$ be a dummy variable equal to one for first year in the data (i.e., 1982), let $DYear2_t$ be a dummy variable equal to second year in the data (i.e., 1983), etc. The LSDV is then expressed as

$$FatalityRate_{it} = \beta_0 + \beta_1 BeerTax_{it} + \delta_2 DYear2_t + \delta_3 DYear3_t + \ldots + \delta_7 DYear7_t + u_{it} \tag{8}$$

As before, because a common intercept β_0 is present in the regression, we do not include all seven year dummies in the regression because of the dummy variable trap.

The main take-away here is that the time fixed effects regression model allows us to eliminate bias arising from omitted variables like national safety standards that change over time but are the same across states in a given year.

4 Regressions with both Entity and Time Fixed Effects

If some omitted variables are constant over time but vary across states (e.g., culture) while others are constant across states but vary over time (e.g., national safety standards), then it is appropriate to include **both** entity and time fixed effects in the regression.

The combined **entity and time fixed effects regression model** is written as

$$FatalityRate_{it} = \beta_1 BeerTax_{it} + \alpha_i + \lambda_t + u_{it} \tag{9}$$

where α_i is the entity (i.e., state) fixed effect and λ_t is the time fixed effect. Because we have both state and time fixed effects, we are controlling for both **state-specific characteristics that do not change over time** (which are captured in α_i) and **factors that vary over time but not across entities** (which are captured in λ_t).

Again as in previous sections, we can use the LSDV and the within transformation to estimate the model with both entity and time fixed effects. Note that for the within transformation, we would need to demean the variables from both their entity *and* the time-period means.

How will the LSDV look like? It will be the same as before, but now we have to include dummy variable for both entity and time. Hence, the specification would be

$$\begin{aligned} FatalityRate_{it} = \beta_0 + \beta_1 BeerTax_{it} & \\ + \gamma_2 DState2_i + \gamma_3 DState3_i + \ldots + \gamma_{48} DState48_i & \\ + \delta_2 DYear2_t + \delta_3 DYear3_t + \ldots + \delta_7 DYear7_t + u_{it} & \end{aligned} \tag{10}$$

where we again exclude the dummy variable for the first state as well as the dummy variable for the first year in the data because of the dummy variable trap.

An important point to remember is that even if we included both entity and time fixed effects, these fixed effects do not control for **state-specific characteristics** that **vary over time**, which can still be a source of omitted variable bias. These omitted variables will be absorbed in the population regression error u_{it}. In our example, this could be state gross domestic product (GDP) per capita (a measure of wealth, which varies over time across states). Note that state GDP per capita would be contained in u_{it}. For state GDP per capita to be cause omitted variable bias, we would need it to be correlated with both the beer tax (which is possible if in richer states, the state government taxes many products to fund social programs) and fatality rate (which is possible if in richer states, the state government spends more money on better roads, etc.).

5 Fixed Effects Regression Assumptions

The fixed effects regression assumptions extend the least squares assumptions to panel data. Under these assumptions, the fixed effects estimator is approximately normally distributed when n is large. To keep the notation simple as possible, we focus on the entity fixed effects regression model (i.e., the regression model without any time effects).

Consider the fixed effects regression

$$Y_{it} = \beta_1 X_{it} + \alpha_i + u_{it}, \qquad i = 1, \ldots, n, \qquad t = 1, \ldots, T$$

The four fixed effects regression assumptions are as follows.

1. $E(u_{it}|X_{i1}, X_{i2}, \ldots, X_{iT}, \alpha_i) = 0$. This means that u_{it} has conditional mean zero, given all values of X for all time periods for a given entity. This assumption plays the same role that we have seen before for cross-sectional data, namely, it is necessary for unbiasedness and implies that there is no omitted variable bias. The key aspect to note here is that for a given entity i, the conditional mean of u_{it} does not depend on past, present, and future values of X. This assumption is violated if, for example, the population error term in this period u_{it} is correlated with the value of X in the previous period, i.e., $X_{i,t-1}$.

2. $(X_{i1}, X_{i2}, \ldots, X_{iT}, u_{i1}, u_{i2}, \ldots, u_{it})$, $i = 1, \ldots, n$ are i.i.d draws from their joint distribution. This means that the variables for one entity are distributed identically to, but independently of, the variables for another entity. This assumption holds if entities are selected by simple random sampling from the population.

3. (X_{it}, u_{it}) have nonzero finite fourth moments. This means that large outliers are unlikely.

4. There is no perfect multicollinearity.

Why do we care about these assumptions? If they hold, then the fixed effects estimator is unbiased, consistent, and normally distributed when n is large. Why do we like this? We like it because hypothesis tests and confidence intervals can be computed using the usual method of using the normal distribution (as well as the F distribution).

6 Serial Correlation

There is an important difference between the fixed effects regression assumptions vs. the least squares assumptions (for cross-sectional data) that we have seen before. In the least squares assumptions, the i.i.d assumptions required that each observation is independent (e.g., person 1's X is independent of person 2's X). However, Assumption # 2 in panel data says something slightly different: it says that the variables for one entity is independent of another entity, but *within* a given entity, there is no requirement that the variables are independent.

Suppose that for a given state i, the beer tax in 1982 is correlated with the beer tax rate in 1983. Similarly, the beer tax in 1983 may be correlated with the beer tax in 1984. This correlation over time

may be because the state does not change the beer tax very often, so that if it is high in one year, it will also tend to be high the next year too. When such a correlation is present, we say that the beer tax is **autocorrelated** or **serially correlated** (i.e., it is correlated within itself, at different dates).

In a similar way, we can think of reasons why the population regression error u_{it} might be correlated over time. Recall that u_{it} consists of time-varying factors that affect Y (fatality rate), but is not captured by the X variables (beer tax, state and time fixed effects). What might be some factors that are contained in u_{it}? This could be factors like the quality of roads, which could also be correlated over time. The basic intuition is that what happens in one year tends to be correlated with what happens the next year, within a state: for example, a bad road that is not repaired this year will be in worse condition next year. As a result the population errors u_{it} will be serially correlated.

Why do we care about serial correlation? We care because if the population regression errors are serially correlated, then the heteroskedasticity-robust standard errors are no longer valid because these standard errors are obtained under the assumption of no serial correlation.

Standard errors that are valid if u_{it} is potentially heteroskedastic and potentially serially correlated over time within an entity are referred to as **heterskedasticity and auto-correlation consistent (HAC) standard erors**. In the context of panel data, one type of HAC standard errors is **clustered standard errors.** The term *clustered* comes from the fact that clustered standard errors allow for heteroskedasticity and serial correlation within an entity, but at the same time, we assume that errors are uncorrelated across clusters. Hence, clustered standard errors allow for hetorskedasticity and serial correlation in a way that is consistent with the second fixed effects regression assumption.

Like heteroskedasticity-robust standard errors, clustered standard errors are valid regardless of whether or not the true population regressions are heteroskedastic, serially correlated, or both. In Stata, we implement clustered standard errors by using the **cluster** option in a regression.