Recitation 6

Seung-hun Lee

Columbia University

November 15th, 2021

Motivation and advantages for panel estimation

- **Panel data**: We observe multiple individuals for multiple periods of time.

$$Y_{it} = \beta_0 + \beta_1 X_{1,it} + ... + \beta_k X_{k,it} + u_{it}$$

  $i = 1, 2, ..., N \rightarrow$ individuals, $t = 1, 2, ..., T \rightarrow$ time periods.
- Balanced: There are T datasets for each of the N individuals.
- Unbalalced: There are $t \leq T$ datapoints for some of the N individuals.

- Panel data allows us to use more datasets.
- Panel data allows us to control for **unobserved heterogeneity** that are
  1. different accross N entities but always remain same for T periods in a given entity (**cross section fixed effect**)
  2. different accross T times periods but remains the same for all N entities in a particular time period (**time fixed effects**)
  3. both of 1) and 2). (**two-way fixed effects**)

What OVB problems could we be dealing with?

- Suppose that $T = 2$ and we are interested in the relationship between vehicle related fatality rate (deaths per 10,000 people) and the beer tax. Suppose that we get these result for the two years

$$\begin{aligned}
\hat{Y}_{i1} =& 2.01 + \quad 0.15 X_{i1} \\
& (0.15) \quad\ (0.20) \\
\hat{Y}_{i2} =& 1.86 + \quad 0.44 X_{i2} \\
& (0.11) \quad\ (0.20)
\end{aligned}$$

- In such case, one might suspect that there is an omitted variable bias that affects these coefficients.
    - Omitted variable specific to the states (Strictness of the relevant law)
    - Time-trends? (Specific to each of years 1 and 2)

How can panel regression do better?

- Let Z_i denote the strictness of state laws on DUI that are unchanging.
- Now write
$$Y_{i1} = \beta_0 + \beta_1 X_{i1} + \beta_2 Z_i + u_{i1}$$
$$Y_{i2} = \beta_0 + \beta_1 X_{i2} + \beta_2 Z_i + u_{i2}$$
- Subtract the second equation from the first to get
$$(Y_{i2} - Y_{i1}) = \beta_1 (X_{i2} - X_{i1}) + \beta_2 (Z_i - Z_i) + u_{i2} - u_{i1}$$
  With Z_i being the same for all periods, the above equation is reduced to
$$(Y_{i2} - Y_{i1}) = \beta_1 (X_{i2} - X_{i1}) + (u_{i2} - u_{i1})$$
- The Z_i variable has no role in this equation - because it is now gone.
- If we estimate this particular β_1, we can obtain much more accurate estimates of the effect of beer tax on fatality rate.

Specific methodologies for cross-sectional FE

- There are two ways of estimating the data when $T \geq 3$
- Least square dummy variables (LSDV): Include $N - 1$ individual dummies
- Within estimation: Subtract "demeaned" equation from the original
- Use:
$$Y_{it} = \beta_0 + \beta_1 X_{it} + \beta_2 Z_i + u_{it} \tag{1}$$
where Z_i is the cross section fixed effect.
- Define $\alpha_i = \beta_0 + \beta_1 Z_i$. Then the above equation can be written as
$$Y_{it} = \beta_1 X_{it} + \alpha_i + u_{it} \tag{2}$$
- α_i term can be thought of as an effect of being an entity i, which is **correlated with** X_{it}

LSDV method

- Define a new variable D_{ki} as follows

$$D_{ki} = \begin{cases} 1 & \text{If } i = k \\ 0 & \text{Otherwise} \end{cases}, \quad k \in \{1, 2, ..., N\}$$

- Since we are going to include β_0, a common intercept, in our regression we need to remove one of the N (dummy variable trap)
- Then we can write

$$Y_{it} = \beta_0 + \beta_1 X_{it} + \delta_2 D_{2i} + ... + \delta_N D_{Ni} + u_{it} \tag{LSDV}$$

- This equation gives different intercepts for each i (can you see why?), while keeping the slope on X_{it} constant at β_1
- Control for unobserved cross section fixed effect by allowing the intercept to differ by each i

Within estimation methods

- Define $\bar{X}_i$, $\bar{Y}_i$ as sample mean of X_{it}, Y_{it} for given i over all possible t's.

$$\bar{X}_i = \frac{1}{T} \sum_{t=1}^{T} X_{it}$$

  Consequently, $\bar{Y}_i$ can be written as

$$\bar{Y}_i = \frac{1}{T} \sum_{t=1}^{T} Y_{it} = \frac{1}{T} \sum_{t=1}^{T} (\beta_1 X_{it} + \alpha_i + u_{it}) = \beta_1 \bar{X}_i + \alpha_i + \bar{u}_i$$

- Subtract Y_{it} by $\bar{Y}_i$ to get

$$Y_{it} - \bar{Y}_i = \beta_1 (X_{it} - \bar{X}_i) + (u_{it} - \bar{u}_i) \implies \tilde{Y}_{it} = \beta_1 \tilde{X}_{it} + \tilde{u}_{it}$$

- This process gets rid of α_i. Then, apply OLS estimation on this equation to get the within estimator

Having both FEs with two-way fixed effects

- We have a DGP

$$Y_{it} = \beta_1 X_{it} + \alpha_i + \lambda_t + u_{it}$$

- LSDV: With an overall constant β_0, we can put $N - 1$ individual and $T - 1$ time dummies
- WE: Demeaning should be done in the following method

$$Y_{it} - \bar{Y}_i - \bar{Y}_t + \bar{Y}$$

  This (and only this) would allow us to get rid of both the α_i individual fixed effect and the λ_t time effects

Least square assumptions for panels

P1 : $E[u_{it}|X_{i1}, .., X_{iT}, \alpha_i] = 0$. It means that the conditional mean of the u_{it} term does not depend on any of the X_{it} values for entity i, whether in the future or in the past.

P2 : $(X_{i1}, .., X_{iT}, u_{i1}, ...u_{iT})$ is IID across $i = 1, .., n$. **This does not rule out the correlation between u_{it}, u_{ij} within entity i for different j and t**, allowing serial correlation within the same entity

P3 : (X_{it}, u_{it}) have nonzero finite fourth moments (outliers are very unlikely) so that the panel estimators have a distribution

P4 : There is no perfect multicollinearity

$\rightarrow$ Because of P2, we need to use **clustered standard error** at a cross-sectional level.

Binary dependent variables: What do we do now?

- Y_i now takes either 0 or 1 (Think of yes-no questions)
- Assume that we are interested in how X_i affects responses to yes-no quesitons

$$Y_i = \beta_0 + \beta_1 X_i + u_i$$

- Non-regressional definition: We look at $E[Y_i|X_i]$, which can be broken into

$$E[Y_i|X_i] = 0 \times \Pr(Y_i = 0|X_i) + 1 \times \Pr(Y_i = 1|X_i)$$

- Or in the regression equation context,

$$\begin{aligned} E[Y_i|X_i] &= E[\beta_0 + \beta_1 X_i + u_i|X_i] \\ &= \beta_0 + \beta_1 X_i + E[u_i|X_i] \\ (\because E[u_i|X_i] = 0) &= \beta_0 + \beta_1 X_i \end{aligned}$$

or the probability of $Y_i = 1$ given X_i

Binary dependent variables: Interpreting main coefficient of interest

- Notice that $\beta_1 = \frac{\Delta Y_i}{\Delta X_i}$ and $\Delta Y_i =$ Change in $\Pr(Y_i = 1|X_i)$ with respect to change in X_i , or

$$\Delta Y_i = \Pr(Y_i = 1|X_i = x + \Delta X_i) - \Pr(Y_i = 1|X_i = x)$$

- Since $\Pr(Y_i = 1|X_i = x + \Delta X_i) - \Pr(Y_i = 1|X_i = x) = E[Y_i|X_i = x + \Delta X_i] - E[Y_i|X_i = x]$, we get

$$\beta_0 + \beta_1(x + \Delta X_i) - \beta_0 + \beta_1(x) = \beta_1 \Delta X_i$$

- So we get $\Delta Y_i = \beta_1 \Delta X_i \iff \beta_1 = \frac{\Delta Y_i}{\Delta X_i}$. Therefore, β_1 now measures how much the predicted probability of $Y_i = 1$ changes with respect to X_i (percentage points!)

Simplest approach: Linear probability models

- **Linear probability model** is the estimation in which you run an OLS on the type of regression equation where Y_i is a binary dependent variable.
- The advantage is that it is simple - there is no difference in terms of methods between this and the OLS methods we have learned so far.
- However, there are some critical disadvantages to this model.
  - By setting the regression model as above, we are assuming that the change of predicted probability of $Y_i = 1$ is constant for all values of X_i.
  - More critically, it is possible that the predicted probability $\hat{y}$ may be greater than 1 or strictly less than 0.
  - The distribution of the error term is no longer normal distribution, potentially affecting the asymptotic properties of the OLS estimators.

Setting up logit regression

- Logit regression: Let $Z_i = \beta_0 + \beta_1 X_i$.
- Logit regression assumes that $\Pr(Y_i = 1|X_i)$ is distributed as

$$\Pr(Y_i = 1|X_i) = F(Z_i) = \frac{1}{1 + e^{-Z_i}}$$

- Changes in X_i affect the probability $F(Z_i)$ in this manner

$$\frac{\partial F}{\partial X_i} = \frac{\partial F}{\partial Z_i}\frac{\partial Z_i}{\partial X_i}$$

  where $\frac{\partial Z_i}{\partial X_i} = \beta_1$
- Value of β_1 does not mean that much in. Its sign does, since

$$\frac{\partial F}{\partial Z_i} = \frac{e^{-\beta_0 - \beta_1 X_i}}{(1 + e^{-\beta_0 - \beta_1 X_i})^2}$$

  and its sign depends on that of β_1

Using normal CDF: Probit regression

- Probit regression: Let $Z_i = \beta_0 + \beta_1 X_i$.
- Probit regression assumes that $\Pr(Y_i = 1|X_i)$ is a standard normal distribution

$$\Pr(Y_i = 1|X_i) = F(Z_i) = \Phi(Z_i) = \Phi(\beta_0 + \beta_1 X_i)$$

  where $\Phi(v)$ means the cumulative normal function $\Pr(Z \leq v)$
- Again, taking the similar approach as before,

$$\frac{\partial F}{\partial X_i} = \frac{\partial F}{\partial Z_i}\frac{\partial Z_i}{\partial X_i}$$

  and $\frac{\partial F}{\partial Z_i}$ is the pdf of a standard normal distribution.
- Again, its sign depends on that of β_1

Different approach to regression: Maximum likelihood estimators

- Both probit and logit are nonlinear: β_0, β_1 parameters are no longer in linear relationship with the $X_i's$ and subsequently Y_i's
- A **likelihood function** is the conditional density of $Y_1, ..., Y_n$ given $X_1, ..., X_n$ that is treated as the function of the unknown parameters (β_0, β_1 in our case)
- What we are trying to do here is to find the values of β_i's that best matches the values of X_i's and Y_i's
- **Maximum likelihood estimators** is the value of β_i's that best describes the data and maximizes the value of the likelihood function

Maximum likelihood estimators in practice

- Assume Y_i's are IID normal with mean μ and standard error σ (both are unknown)
- The joint probability of Y_i's are (our likelihood function)

$$
\begin{aligned}
\Pr(Y_1 = y_1, ..., Y_n = y_n | \mu, \sigma) &= \Pr(Y_1 = y_1 | \mu, \sigma) \times .. \times \Pr(Y_n = y_n | \mu, \sigma) \\
&= \prod_{i=1}^{n} f(y_i | \mu, \sigma) \\
&= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(Y_i - \mu)^2}{2\sigma^2}} \\
&= (2\pi)^{-\frac{n}{2}} (\sigma^2)^{-\frac{n}{2}} e^{-\sum_{i=1}^{n} \frac{(Y_i - \mu)^2}{2\sigma^2}}
\end{aligned}
$$

Maximum likelihood estimators in practice

- Calculation is made easier by using log-likelihood functions (take logs to likelihood functions)

$$-\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln\sigma^2 - \sum_{i=1}^{n}\frac{(Y_i-\mu)^2}{2\sigma^2}$$

- We differentiate the above with respect to μ and σ to find the MLE of these parameters.
- This gets us

$$\mu_{MLE} = \frac{1}{n}\sum_{i=1}^{n} Y_i$$

$$\sigma^2_{MLE} = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \mu_{MLE})^2$$