10 Pooling Panel and Random Effects (Estimation: Micro Panel)

Model

$$y_{it} = a + bx_{it} + \varepsilon_{it}$$

1. Why pooling?
   (a) Economic theory must hold for all individuals.
   (b) More data: either more cross sectional or time series observations. Pooling means more 'efficient' and 'powerful' (will explain later)
2. Why not pooling?
   (a) Account for individual heterogeneity. So at least we have to allow some level heterogeneity such as
   $$y_{it} = a_i + bx_{it} + u_{it}$$
   (b) How to handle for a_i then? Either fixed or random effects.
   (c) What if a_i is observable? like gender, edu, age etc. You may want to include them. How?

10.1 Random Effects

Model:

$$y_{it} = a + bx_{it} + e_{it}, \quad e_{it} = a_i - a + u_{it} = \mu_i + u_{it}$$

Assumption:

A1 $\mathrm{E}(\mu_i x_{it}) = 0$ for all i

A2 $\mathrm{E}(\mu_i u_{it}) = 0$ for all i

Under A1 and A2, note that the pooled OLS becomes consistent but not efficient. The consistency (here we are assuming $N, T \to \infty$ or $N \to \infty$ for any T) requires that

$$\text{plim}_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} \frac{1}{T} \sum_{t=1}^{T} x_{it} e_{it} = 0.$$

Indeed under A1 and A2, we can prove that POLS estimator satisfies the above condition. However, the regression errors are not i.i.d. anymore.

$$e_{i1}e_{i2} = \mu_i^2 + u_{i1}u_{i2} + \mu_i u_{i1} + \mu_i u_{i2}$$

Taking expectation yields

$$\begin{aligned} \text{E}e_{i1}e_{i2} &= \text{E}\mu_i^2 + \text{E}u_{i1}u_{i2} + \text{E}\mu_i u_{i1} + \text{E}\mu_i u_{i2} \\ &= \sigma_\mu^2 \text{ if } \text{E}u_{i1}u_{i2} = 0 \text{ (no serial corr.)} \end{aligned}$$

where we assume $\text{E}(\mu_i u_{it}) = 0$. Also note that

$$\begin{aligned} \text{E}e_{i1}e_{i1} &= \text{E}\mu_i^2 + \text{E}u_{i1}u_{i1} + 2\text{E}\mu_i u_{i1} \\ &= \sigma_\mu^2 + \sigma_u^2 \end{aligned}$$

In this case, pooled GLS estimator becomes efficient and consistent. Here is how to obtain the feasible GLS estimator

1. Run

   $$y_{it} = a + bx_{it} + e_{it}$$

   and get the pooled OLS residuals $\hat{e}_{it}$. Let $\hat{b}_{\text{pols}}$ and $\hat{a}_{\text{pols}}$ be the POLS estimates for b and a.

2. Construct

$$\begin{aligned}
\hat{\sigma}_e^2 &= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(y_{it}-\hat{a}_{\text{pols}}-\hat{b}_{\text{pols}}x_{it}\right)^2=\frac{1}{NT}\sum_{i=1}^{T}\sum_{t=1}^{T}\hat{e}_{it}^2 \\
\hat{\mu}_i &= \frac{1}{T}\sum_{t=1}^{T}\hat{e}_{it}, \quad \hat{u}_{it}=\hat{e}_{it}-\hat{\mu}_i \\
\hat{\sigma}_\mu^2 &= \frac{1}{N}\sum_{i=1}^{N}\left(\hat{\mu}_i-\frac{1}{N}\sum_{i=1}^{N}\hat{\mu}_i\right)^2 \\
\hat{\sigma}_u^2 &= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\hat{u}_{it}-\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{u}_{it}\right)^2
\end{aligned}$$

Note if T is small, $(T-1)$ should be used for the above calculation.

3. Construct the sample covariance matrix

$$\hat{\Omega}=\begin{bmatrix} \hat{\sigma}_\mu^2+\hat{\sigma}_u^2 & \hat{\sigma}_\mu^2 & \cdots & \hat{\sigma}_\mu^2 \\ \hat{\sigma}_\mu^2 & \hat{\sigma}_\mu^2+\hat{\sigma}_u^2 & \cdots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ \hat{\sigma}_\mu^2 & \hat{\sigma}_\mu^2 & \dots & \hat{\sigma}_\mu^2+\hat{\sigma}_u^2 \end{bmatrix} \tag{23}$$

and then construct the feasible GLS estimator given by

$$\hat{b}_{\text{fgls}}=\left(\sum_{i=1}^{N}X_i'\hat{\Omega}^{-1}X_i\right)^{-1}\left(\sum_{i=1}^{N}X_i'\hat{\Omega}^{-1}Y_i\right)$$

where $X_i=[x_{1i},...,x_{Ti}]'$ and $Y_i=[y_{1i},...,y_{Ti}]'$

Remark 1: (Inconsistency relies on A1) If A1 does not hold (usually A1 does not hold), that is, if individual characteristics are correlated with regressors, then POLS estimator becomes inconsistent. Also the random effects estimator (FGLS) is also inconsistent. Because of this reason, many researchers in practice don't run the random effects model (or FGLS estimator). We will study the alternative estimation method in the below (fixed effects model).

Remark 2: (Including Observed Individual Effects) Even when A1 does not hold, if μ_i is observable, then the observed μ_i can be entered the regression as a regressor. That is,

$$y_{it} = a + \gamma_1 \mu_{1i} + \gamma_2 \mu_{2i} + ... + b x_{it} + u_{it}$$

We will study this model later (after studying fixed effects model) in detail.

11 Fixed Effects (Estimation: Micro Panel)

11.1 Eyeball Approach: Works well.

You need to draw some graphs (for your dissertation or journal article) why? looks good, and give more direct information. Try to draw one nice graph which explains main theme of the paper.

11.1.1 Single explanatory variables

Target: Want to explain the relationship between y_{it} and x_{it}. Plot y_{it} on x_{it}. Use **different** color for each i.

1. See if there is one unique relationship between y_{it} and x_{it} across i.
2. <insert a graph here> fixed effects (positive and positive)
3. <insert a graph here> fixed effects (positive but negative)
4. <insert a graph here> heterogeneity (positive but negative)
5. <insert a graph here> projected graph. (demean)

Demean:

$$
\begin{aligned}
y_{it} &= a_i + b x_{it} + u_{it} \\
\frac{1}{T}\sum_{t=1}^{T} y_{it} &= a_i + b\frac{1}{T}\sum_{t=1}^{T} x_{it} + \frac{1}{T}\sum_{t=1}^{T} u_{it} \\
y_{it} - \frac{1}{T}\sum_{t=1}^{T} y_{it} &= b\left(x_{it} - \frac{1}{T}\sum_{t=1}^{T} x_{it}\right) + u_{it} - \frac{1}{T}\sum_{t=1}^{T} u_{it} \\
\tilde{y}_{it} &= b\tilde{x}_{it} + \tilde{u}_{it}
\end{aligned}
$$

11.1.2 More than two variables

$$y_{it} = a_i + bx_{it} + cz_{it} + u_{it}$$

1. Don't plot either $\tilde{y}_{it}$ on $\tilde{x}_{it}$ or $\tilde{y}_{it}$ on $\tilde{z}_{it}$: Why?

2. running $\tilde{y}_{it}$ on $\tilde{x}_{it}$ implies

$$\tilde{y}_{it} = b\tilde{x}_{it} + \tilde{e}_{it},\ \ \tilde{e}_{it} = c\tilde{z}_{it} + \tilde{u}_{it}$$

   If $E(\tilde{x}_{it}\tilde{z}_{it}) \neq 0$, then $\hat{b}$ becomes inconsistent. Worst case: $b = 0$ but $E(\tilde{x}_{it}\tilde{z}_{it}) \neq 0$, then $\hat{b} \neq 0$.

3. Solution: Run

$$\tilde{y}_{it} = a_1\tilde{z}_{it} + \tilde{y}_{it}^{+},\quad \tilde{x}_{it} = a_2\tilde{z}_{it} + \tilde{x}_{it}^{+}$$

   and get residuals $\tilde{y}_{it}^{+}$ and $\tilde{x}_{it}^{+}$. Plot them. Similarly, Run

$$\tilde{y}_{it} = b_1\tilde{x}_{it} + \tilde{y}_{it}^{*},\quad \tilde{z}_{it} = b_2\tilde{x}_{it} + \tilde{z}_{it}^{*}$$

   and plot $\tilde{y}_{it}^{*}$ on $\tilde{z}_{it}^{*}$

4. Mathematically, it is a projection approach. $I - Z(Z'Z)^{-1}Z' = M_z$ or M_x matrix.

11.2 Common Time Effects:

$$y_{it} = a_i + \lambda_t + bx_{it} + u_{it}$$

Allows time dummies also. How to estimate $\hat{b}$?

1. Eliminate fixed effects by demeaning over t.

$$y_{it} - \frac{1}{T}\sum_{t=1}^{T} y_{it} = \lambda_t - \frac{1}{T}\sum_{t=1}^{T}\lambda_t + b\left(x_{it} - \frac{1}{T}\sum_{t=1}^{T}x_{it}\right) + u_{it} - \frac{1}{T}\sum_{t=1}^{T}u_{it}$$

   Still you have λ_t terms.

2. Rewrite this as

$$\tilde{y}_{it} = \tilde{\lambda}_t + b\tilde{x}_{it} + \tilde{u}_{it} \tag{24}$$

Take cross sectional mean

$$\frac{1}{N}\sum_{i=1}^{N}\tilde{y}_{it} = \tilde{\lambda}_t + b\frac{1}{N}\sum_{i=1}^{N}\tilde{x}_{it} + \frac{1}{N}\sum_{i=1}^{N}\tilde{u}_{it} \tag{25}$$

3. subtract (25) from (24).

$$\tilde{y}_{it} - \frac{1}{N}\sum_{i=1}^{N}\tilde{y}_{it} = b\left(\tilde{x}_{it} - \frac{1}{N}\sum_{i=1}^{N}\tilde{x}_{it}\right) + \left(\tilde{u}_{it} - \frac{1}{N}\sum_{i=1}^{N}\tilde{u}_{it}\right)$$

4. Finally evaluate

$$y_{it}^{\dagger} = \tilde{y}_{it} - \frac{1}{N}\sum_{i=1}^{N}\tilde{y}_{it} = y_{it} - \frac{1}{T}\sum_{t=1}^{T}y_{it} - \frac{1}{N}\sum_{i=1}^{N}y_{it} + \frac{1}{T}\sum_{t=1}^{T}\frac{1}{N}\sum_{i=1}^{N}y_{it} \tag{26}$$

we call it 'within transformation'.

Note: Fixed effects estimator is called either 'Least Squares Dummies Variable (LSDV)' estimator or 'Within Group' estimator.

Questions Consider the following data generating process

$$y_{it} = \mu_{y,i} + y_{it}^o, \quad x_{it} = \mu_{x,i} + x_{it}^o \tag{27}$$

where

$$\mu_{y,i} = a + b\mu_{x,i} + \epsilon_i \tag{28}$$

$$y_{it}^o = \alpha_i + \beta x_{it}^o + u_{it}^o \tag{29}$$

1. Suppose that you run the following cross section regression for $t = 1$.

$$y_{i1} = c_1 + \gamma_1 x_{i1} + \varepsilon_{i1} \tag{30}$$

   Prove that the OLS estimate becomes inconsistent generally. That is,

$$\text{plim}_{N\to\infty}\hat{\gamma}_1 \neq b$$

2. Rather than running (30), you run the following cross sectional regression with time series average.

$$\bar{y}_i = c + \gamma\bar{x}_i + \bar{\varepsilon}_i \tag{31}$$

   where

$$\bar{y}_i = \frac{1}{T}\sum_{t=1}^{T} y_{it}, \quad \bar{x}_i = \frac{1}{T}\sum_{t=1}^{T} x_{it}$$

   Derive the limiting distribution of $\hat{\gamma}$ in (31). Is the convergence rate equal to $\sqrt{NT}$ or $\sqrt{N}$?

Part II (POLS): Consider the following DGP

$$y_{it} = a_i + y_{it}^o, \quad y_{it}^o = \rho y_{it-1}^o + u_{it}, \quad u_{it} \sim iid\left(0, \sigma^2\right)$$

1. You run the POLS given by

$$y_{it} = a + \rho y_{it-1} + e_{it}$$

   Prove that when $\rho < 1$, the POLS estimator becomes inconsistent. Derive the exact bias.

Part III (Dynamic Panel Regression I) Consider the following DGP

$$y_{it} = a_i + y_{it}^o, \quad y_{it}^o = \rho y_{it-1}^o + u_{it}, \quad u_{it} \sim iid\left(0, \sigma^2\right)$$

Derive Nickell bias when $\rho = 1$

11.3 Dynamic Panel Regression

Read: Bertrand, M., E. Duflo and S. Mullainathan, 2004, How much should we trust differences-in-differences estimates?, *Quarterly Journal of Economics*, 249–275.

Model:

$$y_{it} = a_i + \lambda_t + bx_{it} + u_{it} \tag{32}$$

Now the regression error follows

$$u_{it} = \rho u_{it-1} + \varepsilon_{it}$$

Remark 1: As long as x_{it} is exogenous, the LSDV estimator in (32) becomes consistent. However, the statistical inference (in other words, $t-$value for $\hat{b}$) becomes an issue (in other words, the critical value for $\hat{t}_b$ must be different than the ordinary critical value). We will suggest the solution for the statistical inference later. (see section 3)

Remark 2: If T is large, then more efficient estimator can be obtain by running dynamic panel regression.

Let's transform (32) as

$$\rho y_{it-1} = a_i\rho + \rho\lambda_{t-1} + b\rho x_{it-1} + \rho u_{it-1} \tag{33}$$

and next subtract (33) from (32). Then we have

$$y_{it} = a_i (1 - \rho) + \rho\lambda_t + \rho\lambda_{t-1} + \rho y_{it-1} + bx_{it} - b\rho x_{it-1} + \varepsilon_{it}$$

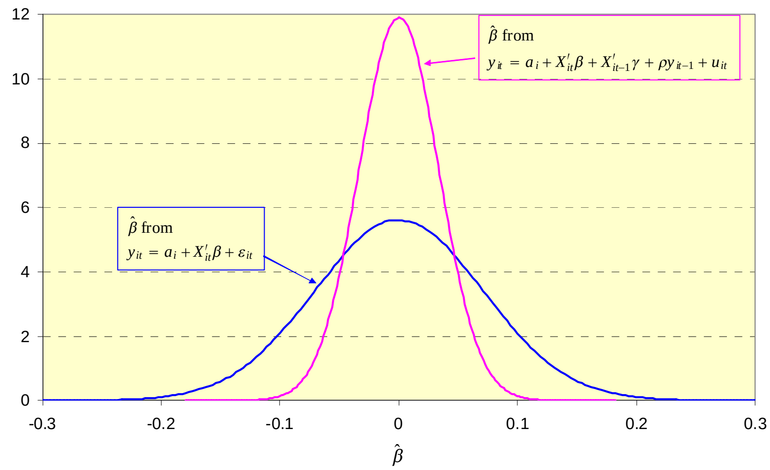
or

$$y_{it} = \alpha_i + \theta_t + \rho y_{it-1} + bx_{it} + \gamma x_{it-1} + \varepsilon_{it}. \tag{34}$$

By using within transformation, we can run

$$y_{it}^{\dagger} = \rho y_{it-1}^{\dagger} + bx_{it}^{\dagger} + \gamma x_{it-1}^{\dagger} + \varepsilon_{it}^{\dagger}$$

See (26) the definition of ‘†’.



Remark 3 (Consistency for ρ and γ): The LSDV estimators for ρ and γ are inconsistent but the LSDV estimator β becomes consistent. So the parameter of interest is here assumed to be β. Since the estimators for ρ and γ are inconsistent, the statistical inference for β should be carefully constructed. In the next section, we will study how to obtain robust statistical inference regardless of error term structures.

12 Pooling Panel and Random Effects (Testing: Micro Panel)

12.1 Bench Mark Model: Strongly Exogenous Single Regressor with Fixed Effects

Model

$$y_{it} = a_i + bx_{it} + u_{it} \tag{35}$$

Assumptions

1. $\mathrm{E}x_{it}u_{js} = 0$ for all i, j, s, t.

Here we are interested in testing the null hypothesis of $H_0 : b = 0$. To test this null hypothesis, we need a statistic. Usually we use a formal t statistic defined by

$$t_{\hat{b}} = \frac{\hat{b}}{\sqrt{Var\left(\hat{b}\right)}}$$

where $Var\left(\hat{b}\right)$ stands for the sample variance of the point estimate $\hat{b}$ which depends on the parametric assumptions for the regression errors.

In the below, we will study various hypotheses testings and statistics. Before that, I will address why the panel data is useful (and powerful) compared with either cross sectional or time series regressions.

12.1.1 More T or More N?

General statistical panel theory states that the panel gain comes from the use of more data. However, this statement is not quite right. One may have either a lengthy time series or cross section data. However whenever one uses a panel data, s/he can use either a short time series across some individuals, or a small individual over somewhat large time series data. For example, many empirical growth regressions have been based on cross sectional studies due to the data limitation. Even though PWT provides more than 150 countries panel data, it is often very hard to obtain a full set of panel data for all 150 countries. Here we consider which data sets (larger T or N) we should use to increase panel gain.

To attack this issue, we first consider the rate of convergence concept. Consider the following simple regression

$$y_s = bx_s + u_s, \text{ for } s = i \text{ or } t, \text{ and } s = 1,,,.S$$

where we assume the strong exogeneity of x_s. Typical limiting distribution theory says

$$\begin{aligned}
\hat{b} &= \frac{\frac{1}{S}\sum_{s=1}^{S} x_s y_s}{\frac{1}{S}\sum_{s=1}^{S} x_s^2} = b + \frac{\frac{1}{S}\sum_{s=1}^{S} x_s u_s}{\frac{1}{S}\sum_{s=1}^{S} x_s^2}, \\
\hat{b} - b &= \left(\frac{1}{\sqrt{S}}\right)\frac{\frac{1}{\sqrt{S}}\sum_{s=1}^{S} x_s u_s}{\frac{1}{S}\sum_{s=1}^{S} x_s^2} := \left(\frac{1}{\sqrt{S}}\right)\frac{A_S}{B_S}, \text{ let say}
\end{aligned}$$

We may assume that

$$A_S \Longrightarrow^d N\left(0, \Omega_A^2\right), \quad B_S \longrightarrow^p Q_B \text{ as } S \to \infty$$

where '$\Longrightarrow^d$' stands for convergence in distribution and '$\longrightarrow^p$' means convergence in probability. Then we finally have (following by Cramer's theorem)

$$\sqrt{S}\left(\hat{b} - b\right) \Longrightarrow^d N\left(0, Q_B^{-1}\Omega_A^2 Q_B^{-1}\right)$$

Alternatively

$$\frac{\sqrt{S}\left(\hat{b} - b\right)}{\sqrt{Q_B^{-1}\Omega_A^2 Q_B^{-1}}} \Longrightarrow^d N(0,1)$$

Meanwhile the testing hypothesis is given by

$$\begin{aligned}
H_0 &: b_s = 0, \text{ usually.} \\
H_A &: b_s \neq 0
\end{aligned}$$

Then we have

$$\frac{\sqrt{S}\hat{b}}{\sqrt{Q_B^{-1}\Omega_A^2 Q_B^{-1}}} \Longrightarrow^d N\left(\frac{\sqrt{S}b}{\sqrt{Q_B^{-1}\Omega_A^2 Q_B^{-1}}}, 1\right)$$

so that the power of the test (how frequently a test can reject the null hypothesis when the alternative is true) is getting larger if

1. true value of $|b|$ is getting larger,
2. Variance of b is getting smaller,
3. the number of observations, S, is getting larger.

Among them, the last item, 3, is only thing we can control for. We don't know the true value of b and the true variance of b either. However, we can increase the number of observations (by putting more labor hours for digging out the data).

Now, when we have both N and T dimensions, we can rewrite the pooled estimate of b as

$$\hat{b}_{\text{panel}} = \frac{\frac{1}{T}\sum_{t=1}^{T}\frac{1}{N}\sum_{i=1}^{N} x_{it}y_{it}}{\frac{1}{T}\sum_{t=1}^{T}\frac{1}{N}\sum_{i=1}^{N} x_{it}^2} = b + \frac{\frac{1}{T}\sum_{t=1}^{T}\frac{1}{N}\sum_{i=1}^{N} x_{it}u_{it}}{\frac{1}{T}\sum_{t=1}^{T}\frac{1}{N}\sum_{i=1}^{N} x_{it}^2},$$

and similarly

$$\hat{b}_{\text{panel}} - b = \left(\frac{1}{\sqrt{NT}}\right)\frac{A_{NT}}{B_{NT}}, \text{ let say}$$

and

$$A_{NT} \Longrightarrow^{d} N\left(0, \Omega_A^2\right), \quad B_{NT} \longrightarrow^{p} Q_B \text{ as } N, T \to \infty \text{ jointly.} \tag{36}$$

Then we have

$$\sqrt{NT}\left(\hat{b}_{\text{panel}} - b\right) \Longrightarrow^{d} N\left(0, Q_B^{-1}\Omega_A^2 Q_B^{-1}\right) \tag{37}$$

Now consider the above three criteria for the power of the test. Does panel data enable us to know either true value of variance of b? The answer is no. Then what about the last one? Does panel data enable us to use more observations? The answer is not straightforward. In practice, one often face the situation like this. When one use one dimensional data (for example time series), one may choose or select the longest time series data for y_t and x_t. Denote the size of the sample as T_s. Now if s/he has to use a panel data, usually s/he scarifies the lengthy time series in order to increase the cross section units. Denote the time series s/he will use for the panel data as T. From the direct calculation, we have the condition for the panel gain given by

$$N > T_s/T$$

That is, if you have 300 of T_s for one series but have to use only 30 of T in order to use the panel data, then the minimum number of the cross sections – you have to obtain – should be larger than 10.

However, we may need much larger cross sections if y_s and x_s are $I(1)$ (or in other words, nonstationary). In this case, the limiting distribution for $\hat{b}$ is different from a normal distribution (actually it becomes $\left(\int B_x du\right)\left(\int B_x^2 dr\right)^{-1}$) and also the convergence rate becomes T rather than $\sqrt{T}$. Hence the minimum condition for the panel gain changes as

$$\sqrt{N} > T_s/T.$$

In the above example, you need at least $\sqrt{N} > 10$ or $N > 100$.

Unfortunately, the most of macro data are nonstationary. So the important question becomes that how many observations should be scarified to use the panel data. Let k be the fraction of the sample you have to sacrifice to use additional N cross sections. Then we have

$$\sqrt{N} > \frac{T_s}{T}, \text{ or } \sqrt{N} > \frac{T_s}{(1-k)\,T_s} = \frac{1}{1-k},$$

so that

$$N > \left(\frac{1}{1-k}\right)^2.$$

To decode this formula, let say you have 120 monthly time series observations initially. In order to use the panel data, if you have to use 10 annual observations, then $T_s/T = 120/10 = 12$, so that the minimum N becomes 144. Remember that the power of a test with $N = 144$ and $T = 10$ will be exactly same as the power of the test with $T = 120$ and $N = 1$. However, if you can still use monthly observations but loose 2 years observations, then $T_s/T = 120/96 = 1.25$, so that the minimum N becomes 1.56 which is less than 2. Hence the power of a test with $N = 2$ and $T = 96$ will be larger than that with $N = 1$ and $T = 120$.

So the conclusion follows:

Recommendation (How to Construct a Panel Data)

1. When you are interested in the correlation among level variables, you should use the panel data set which contains more T, or the largest of $N \times T^2$ rather than $N \times T$.

2. When you are interested in the correlation among (quasi) difference variables (such as growth rates), you should use the panel data which total number of observations ($= N \times T$) is largest.

12.1.2 How to Calculate the Covariance Matrix

Here we are asking how to estimate Ω_A^2 and Q_B in (36) and (37). First consider Ω_A^2 which can be defined as

$$\begin{aligned}
\Omega_A^2 &= \frac{1}{NT} E\left(\sum_{i=1}^{N}\sum_{t=1}^{T} x_{it}u_{it}\right)^2 \\
&= \frac{1}{NT} E\left(x_{11}u_{11} + ... + x_{1T}u_{1T} + x_{21}u_{21} + ... + x_{NT}u_{NT}\right)^2 \\
&= \frac{1}{NT} E\left(x_{11}^2u_{11}^2 + ... + x_{1T}^2u_{1T}^2 + x_{21}^2u_{21}^2 + ... + x_{NT}^2u_{NT}^2\right) \\
&\quad + E\left(\text{cross products}\right)
\end{aligned} \tag{38}$$

If $E\left(u_{i1}u_{i2}\right) \neq 0$ due to serial correlation, then in general the expected values of the cross product terms are not equal to zero.

White (1980) suggests the use of the so called 'heteroskedasticity consistent estimator' which is given by

$$\hat{\Omega}_A^2 = \frac{1}{N}\sum_{i=1}^{N} \tilde{X}_i'\hat{u}_i\hat{u}_i'\tilde{X}_i$$

where $\hat{u}_i = \left(\hat{u}_{i1}, ..., \hat{u}_{iT}\right)'$, $\tilde{X}_i = \left(\tilde{x}_{i1}, ..., \tilde{x}_{iT}\right)'$. So that the sample covariance matrix becomes

$$V\left(\hat{b}\right) = \left(\sum_{i=1}^{N} \tilde{X}_i'\tilde{X}\right)^{-1}\left(\sum_{i=1}^{N} \tilde{X}_i'\hat{u}_i\hat{u}_i'\tilde{X}_i\right)\left(\sum_{i=1}^{N} \tilde{X}_i'\tilde{X}\right)^{-1} \tag{39}$$

and its associated t-statistic becomes

$$t_{\hat{b}} = \frac{\hat{b}}{\sqrt{\left(\sum_{i=1}^{N} \tilde{X}_i'\tilde{X}\right)^{-1}\left(\sum_{i=1}^{N} \tilde{X}_i'\hat{u}_i\hat{u}_i'\tilde{X}_i\right)\left(\sum_{i=1}^{N} \tilde{X}_i'\tilde{X}\right)^{-1}}} \tag{40}$$

Note that if x_{it} and u_{it} are *iid*, then the above formula can be simplified as

$$V\left(\hat{b}\right) = \hat{\sigma}_u^2\left(\sum_{i=1}^{N} \tilde{X}_i'\tilde{X}\right)^{-1} \tag{41}$$

which is the sample variance reported in canned statistical packages.

Here are a couple of very important facts:

Recommendation

1. Usually the sample variance in (2-5) is larger than that in (41). This implies that when there is either heteroskedasticity or autocorrelation, the standard t-ratio is much larger than its true value.

2. When T is fixed but N is large, the $t_{\hat{b}}$ in (40) is distributed as a normal. So the standard critical value can be used here. However when T is large but N is small, the t ratio asymptotically follows a $t-$distribution with $N-1$ degrees of freedom under homoskedasticity. (Hansen, 2007 Journal of Econometrics, 'Asymptotic Properties of a Robust Variance Matrix Estimator for Panel Data when T is large')

12.2 Testing

12.2.1 Some Basic Facts on Statistical Testing

Size and Power The size of a test stands for the rejection rate of the null when the null hypothesis is true, meanwhile the power of a test implies the rejection rate of the null when the alternative is true. Usually we set the size of a test at the significance level. For example, the critical value for the 5% significance level for a normal distribution (for two sides test) is 1.95. In other words, we permit ourselves that we would make a wrong decision at the 5% level. (5 out of 100 times). Setting a smaller size means that you want to be more conservative or don't want to make any mistake, but at the same time it also implies that the power of the test will be reduced.

Size Distortion You set the size at the 5% significance level. However (especially in the finite sample), a test does not produce exactly the 5% of the size. If a test over-rejects the null (when the null is true), then we say that the test suffers from oversize distortion. The opposite case is undersize distortion. Usually the undersized test is acceptable since it simply implies that you will make less mistake. The oversize problem becomes serious. The oversized test usually rejects the null very often even when the null is true.

Size Problem in Panel Data In univariate case, usually a well designed statistic does not suffer from the size distortion as n (the number of observations) goes to infinity. For example, the standard t-test for the univariate AR(1) regression produces somewhat serious size distortion with small T, but as $T \to \infty$, the size distortion goes away.

$$y_t = a + \rho y_{t-1} + u_t, \quad t_{\hat{\rho}} = \frac{\hat{\rho}}{\sqrt{V(\hat{\rho})}} \text{ for } \rho < 1 \tag{42}$$

It is because the asymptotic variance of $\hat{\rho}$ is designed in this way. However, in the panel data, the t-ratio produces more size distortion as $N \to \infty$ for fixed T.

$$y_{it} = a_i + \rho y_{it-1} + u_{it}, \quad t_{\hat{\rho}} = \frac{\hat{\rho}_{\text{lsdv}}}{\sqrt{V(\hat{\rho}_{\text{lsdv}})}} \text{ for } \rho < 1 \tag{43}$$

The underlying reason is simple. When T is small, the test statistic in (42) produces a small size distortion. In the panel data, the size distortion becomes cumulated as N increases. Similarly, as $T \to \infty$ for a fixed N, the usual panel statistic in (40) produces more size distortion if there is heteroskedasticity in the error terms.

12.2.2 Fixed versus Random Effects.

LSDV estimator is 'robust' and consistent whether or not the fixed effects a_i in (35) are correlated with x_{it}. Meanwhile the GLS (or random effects estimator) is 'efficient' and consistent only when a_i is not correlated with regressors. When the number of observations are small (such as moderately small N and T), the GLS becomes an attractive estimator if a_i is not correlated with regressors. Naturally econometricians have developed various test statistics to investigate if this condition holds or not.

There are broadly two ways to test the orthogonality between a_i and x_{it}. The first method is based on the pooled OLS regression residuals, and the second method is based on the difference between LSDV and GLS. We discuss the first method, first.

Breusch & Pagan (1980)'s LM Test BP tests if

$$\begin{aligned} H_0 &: a_i = a, \text{ for all } i, \\ H_A &: a_i \neq a \text{ for any } i \end{aligned} \tag{44}$$

When u_{it} in (35) is not serially correlated, these hypotheses can be rewritten as

$$\begin{aligned} H_0 &: E\left(\hat{e}_{it}\hat{e}_{is}\right)=0 \text{ for all } i, \\ H_A &: E\left(\hat{e}_{it}\hat{e}_{is}\right)\neq 0 \text{ for any } i \end{aligned}$$

where $\hat{e}_{it}$ is the pooled OLS regression residuals. That is,

$$\hat{e}_{it}=y_{it}-\hat{a}-\hat{b}_{\text{pols}}x_{it}.$$

The test statistic is given by

$$LM=\frac{NT}{2\left(T-1\right)}\left[\frac{\sum_{i=1}^{N}\left(\sum_{t=1}^{T}\hat{e}_{it}\right)^2}{\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{e}_{it}^2}-1\right]^2\Longrightarrow^d\chi_1^2$$

Note that

$$\mathrm{E}\left(\sum_{t=1}^{T}\hat{e}_{it}\right)^2=\mathrm{E}\left(\sum_{t=1}^{T}\hat{e}_{it}^2+\sum_{t=1}^{T}\sum_{s\neq t}^{T}e_{it}e_{is}\right)$$

and under H_0, we have

$$\mathrm{E}\left(\sum_{t=1}^{T}\hat{e}_{it}\right)^2=\mathrm{E}\left(\sum_{t=1}^{T}\hat{e}_{it}^2+\sum_{t=1}^{T}\sum_{s\neq t}^{T}e_{it}e_{is}\right)=\mathrm{E}\left(\sum_{t=1}^{T}\hat{e}_{it}^2\right)$$

since the expectation of the cross product terms become zero. For large T and N, also note that under the alternative and no serial correlation among u_{it}, we have

$$\mathrm{E}\left(\sum_{t=1}^{T}\hat{e}_{it}\right)^2\geq\mathrm{E}\left(\sum_{t=1}^{T}\hat{e}_{it}^2\right)$$

since

$$\mathrm{E}\left(e_{it}e_{is}\right)=\mathrm{E}\left(\mu_i^2+u_{it}u_{is}\right)=\sigma_\mu^2>0.$$

It is important to note that if u_{it} is serially correlated, then BP's LM test fails.

Hausman's Specification Test Hausman test is fairly a general test for misspecification, and can be applied to test the null hypothesis in (44). Under the null hypothesis

$$\text{plim}_{N,T\to\infty}\hat{b}_{\text{LSDV}}=\text{plim}_{N,T\to\infty}\hat{b}_{\text{GLS}}$$

since two estimators are both consistent. However, under the alternative, we have

$$\text{plim}_{N,T\to\infty}\hat{b}_{\text{LSDV}} = b \text{ but } \text{plim}_{N,T\to\infty}\hat{b}_{\text{GLS}} \neq b$$

so that

$$\text{plim}_{N,T\to\infty}\left(\hat{b}_{\text{GLS}} - \hat{b}_{\text{LSDV}}\right) \neq 0$$

Hence we can test H_0 by examining if the distance between $\hat{b}_{\text{GLS}}$ and $\hat{b}_{\text{LSDV}}$ is equal to zero or not. A typical test statistic in this case is given by

$$\left(\hat{b}_{\text{GLS}} - \hat{b}_{\text{LSDV}}\right)'\left[Var\left(\hat{b}_{\text{GLS}} - \hat{b}_{\text{LSDV}}\right)\right]^{-1}\left(\hat{b}_{\text{GLS}} - \hat{b}_{\text{LSDV}}\right)' \Longrightarrow^d \chi_k^2$$

when the dimension of $\hat{b}$ is k. For a single regressor case, we have simply

$$\frac{\left(\hat{b}_{\text{GLS}} - \hat{b}_{\text{LSDV}}\right)^2}{Var\left(\hat{b}_{\text{GLS}} - \hat{b}_{\text{LSDV}}\right)} \Longrightarrow^d \chi_1^2$$

Note that under H_0,

$$Var\left(\hat{b}_{\text{GLS}} - \hat{b}_{\text{LSDV}}\right) = \hat{\sigma}_u^2\left[\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{x}_{it}^2\right]^{-1} - \left[\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}\hat{\omega}_{ij}\tilde{x}_{it}\tilde{x}_{jt}\right]^{-1}$$

where $\hat{\omega}_{ij}$ is the ith and jth element of $\hat{\Omega}^{-1}$ and $\hat{\Omega}$ is defined at (23).

13 Dynamic Panel Regression I (Issues and Problems)

More than the quarter of theoretical studies on the panel data is focused on the dynamic panel regression. Modelling the 'dynamics' in the panel data is critically important. First we address where 'dynamic adjustment form' comes from.

13.1 Source of Serial Correlation

13.1.1 Univariate Series

Many economic variables such as income, consumption, wage, etc have the following transitional path.

$$y_{it} = y_i^* + (y_{i0} - y_i^*)\, e^{-\beta t}$$

where y_i^* is the steady state outcome. Note that all variables are in logarithm. Rewrite this model as

$$y_{it} = \left[y_i^* + (y_{i0} - y_i^*)\, e^{-\beta(t-1)}\right] e^{-\beta} + y_i^* \left(1 - e^{-\beta}\right) = y_i^* \left(1 - e^{-\beta}\right) + e^{-\beta} y_{it-1}$$

By letting $\rho = e^{-\beta}$, $a_i = y_i^*$ and adding a random error which can be exogeneous i.i.d. measurement errors, transitory shocks, etc, then we have

$$y_{it} = a_i (1 - \rho) + \rho y_{it-1} + u_{it}, \text{ for } t = 1, ..., T.$$

This simple growth model generates the time dependence between y_{it} and y_{it-1}.

In other words, all variables (growing variables such as wage, income, height etc) are serially correlated during transition periods.

13.1.2 General Regressions

In general regression models, the serial correlation occurs whenever the regression is not balanced. To understand the balancing concept, consider a simple regression model given by

$$y_{it} = \alpha_i + bx_{it} + u_{it} \tag{45}$$

Suppose that

1. y_{it} has a linear trend. $b \neq 0$. Regardless x_{it} has a linear trend or not, u_{it} contains a linear trend. So you have to include a trend in the regression. Why?

   (a) Let $y_{it} = a_{iy} + c_i t + y_{it}^o$, and $x_{it} = a_{ix} + d_i t + x_{it}^o$. You don't want to assume that the deterministic trend terms have a common relationshop since you can't write it as

   $$c_i t = a + b\,(d_i t) + e_{it}. \tag{46}$$

   Simply because the dependent variable is purely nonstochastic. Even when the dependent variable has a stochastic component (such as $\zeta_{it} = c_i t + \epsilon_{it}$, and $\zeta_{it} = a + b\,(d_i t) + e_{it}$, as long as $c_i \neq bd_i$ for any i, the error term includes a linear trend component.

   (b) The interest relation must be between y_{it}^o and x_{it}^o. In this case, you have to eliminate the trend term in the first place by including a linear trend component in the regression

   (c) If you are interested in analyzing growth rates in y_{it} and x_{it}, then you have to take the first difference to approximate the stochastic growth components. That is,

   $$\Delta y_{it} = \alpha_i + b\Delta x_{it} + \text{error}_{it} \tag{47}$$

2. y_{it} is serially correlated but x_{it} is not. Then u_{it} is serially correlated. (the opposite is not true) In this case, you may want to run the dynamic panel regression

   $$y_{it} = a_i + \rho y_{it-1} + bx_{it} + \gamma x_{it-1} + \varepsilon_{it} \tag{48}$$

   (a) From (45), you have

   $$u_{it} = \rho u_{it-1} + \varepsilon_{it}. \tag{49}$$

   Here I assume that the error term follows AR(1) structure for simplicity.

   (b) Then you have

   $$\rho y_{it-1} = \alpha_i \rho + b\rho x_{it-1} + \rho u_{it-1}. \tag{50}$$

   Subtracting (50) from (45) yields (48).

3. y_{it} is not serially correlated but x_{it} is very persistent (ρ is near unity). And more importantly $b \neq 0$. Then the regression in (45) is not well specified. Simply it becomes unbalanced regression. In this case, to balance out the serial correlation, u_{it} should be negatively correlated with x_{it}.

   (a) Example: Stock return predictability & UIP:

   $$y_{it} = a_i + bx_{it} + u_{it}$$

   where y_{it} is either stock return or depreciation rates, which are almost white noisy. x_{it} is either interest rate differential (for UIP), or dividend ratio (stock return). Both interest rate differential or dividend ratio is highly serially correlated. If $b \neq 0$, then x_{it} should be negatively correlated with u_{it}.

   (b) Hence u_{it} is serially correlated in this case also.

13.2 Modeling Dynamic Panel Regression

There are several types of dynamic panel regressions. Depending on the regression types, the properties of LSDV estimators are quite different. Hence modeling dynamic panel regression becomes very important.

$$\text{M1:} \qquad y_{it} = a_i + \beta x_{it} + u_{it}, \quad u_{it} = \rho u_{it-1} + \varepsilon_{it} \tag{51}$$

$$\text{M2:} \qquad y_{it} = a_i + \rho y_{it-1} + \beta x_{it} + \varepsilon_{it} \tag{52}$$

where I didn't include common time effects and linear trend components either. Note that M1 and M2 can be restated as

$$\text{M1:} \qquad z_{it} = \alpha_i + u_{it}, \quad z_{it} = y_{it} - \beta x_{it}, \quad u_{it} = \rho u_{it-1} + \varepsilon_{it} \tag{53}$$

$$\text{M2:} \qquad y_{it} = \alpha_i + u_{it}, \quad u_{it} = \rho u_{it-1} + e_{it}, \quad e_{it} = \beta x_{it} + \varepsilon_{it} \tag{54}$$

Note that in M1, x_{it} is correlated with y_{it} in level. Meanwhile in M2, x_{it} is correlated with the quasi-differenced y_{it}. Alternatively we can rewrite M1 as

$$\text{M1: } y_{it} = a_i + \rho y_{it-1} + \beta x_{it} + \gamma x_{it-1} + \varepsilon_{it}. \tag{55}$$

Hence if (52) is true, then (55) is not misspecified. Simply γ becomes zero if (52) is true. However, if (55) or M1 is true, then (52) becomes misspecified, which results in inconsistent estimator for β as well as ρ in (52). In this sense, (55) nests (52).

The economic interpretations are different across models. M1 states that the quasi-difference $(y_{it} - \rho y_{it-1})$ is explained by x_{it}. Meanwhile M2 implies that the level of y_{it} is explained by x_{it}. Hence usually x_{it} in (52) is assumed to follow a white noisy process (no serial correlation). Meanwhile x_{it} in (55) does not have such restriction.

13.3 Inconsistency of LSDV estimator

Here we analyze why the LSDV estimator under fixed effects becomes inconsistent as $N \to \infty$ but fixed T. The model we study is given by

$$y_{it} = a_i + \rho y_{it-1} + u_{it}, \quad u_{it} \sim iid\left(0, \sigma_u^2\right)$$

Nickell Bias (1981, Econometrica) Nickell extends the so-called 'Kendall' (1954, Biometrika) bias to the panel data setting.

1. To understand Kendall bias, we consider an univariate simple AR(1) model with constant

$$y_t = a + \rho y_{t-1} + u_t. \tag{56}$$

The OLS estimator is given by

$$\hat{\rho} = \frac{\sum_{t=2}^{T} \tilde{y}_{t-1} \tilde{y}_t}{\sum_{t=2}^{T} \tilde{y}_{t-1}^2},$$

and its expectation gives

$$E\hat{\rho} = E\left[\frac{\sum_{t=2}^{T} \tilde{y}_{t-1} \tilde{u}_t}{\sum_{t=2}^{T} \tilde{y}_{t-1}^2}\right] := E\frac{A_T}{B_T}$$

From Marriott and Pope (1954, Biometrika), we have

$$E\frac{A_T}{B_T} = \frac{EA_T}{EB_T}\left[1 - E\left(C_T\right)\right]$$

$$E(C_T) = \frac{Cov(A_T B_T)}{E(A_T)E(B_T)} + \frac{Var(B_T)}{[E(B_T)]^2}$$

Note that $E(C_T) \neq 0$ usually due to asymetric distribution of $\hat{\rho}$. In the finite sample, the empirical distribution of $\hat{\rho}$ is not a normal but skewed left a little bit. This asymetric distribution yields the small sample bias but usually it goes away quickly as T increases

2. The major bias arises from the first term EA_T/EB_T. To see this

$$\frac{EA_T}{EB_T} = \rho + \frac{E\sum_{t=2}^T \tilde{y}_{t-1}\tilde{u}_t}{E\sum_{t=2}^T \tilde{y}_{t-1}^2}$$

Note that

$$\begin{aligned} E\sum_{t=2}^T \tilde{y}_{t-1}\tilde{u}_t &= E\sum_{t=2}^T (y_{t-1}-\bar{y})(u_t-\bar{u}) = E\sum_{t=2}^T y_{t-1}u_t - \frac{1}{T}E\left(\sum_{t=2}^T y_{t-1}\right)\left(\sum_{t=2}^T u_t\right) \\ &= 0 - \frac{1}{T}E\left(\sum_{t=2}^T y_{t-1}\right)\left(\sum_{t=2}^T u_t\right) \end{aligned}$$

Since

$$y_t = a + \rho y_{t-1} + u_t = \frac{a}{1-\rho} + \sum_{j=0}^{\infty} \rho^j u_{t-j}$$

so that $Ey_t u_s = 0$ for all $t < s$. However

$$E\left(\sum_{t=2}^T y_{t-1}\right)\left(\sum_{t=2}^T u_t\right) = E(y_1 + \ldots + y_{T-1})(u_2 + \ldots + u_T)$$

and note that

$$Ey_1u_1 = \sigma_u^2, \quad Ey_2u_1 = E(\rho y_1 + u_2)u_1 = \rho\sigma_u^2, \ldots$$

hence this term is not eqaul to zero.

3. Finally we have

$$E\hat{\rho} = E\frac{A_T}{B_T} = \rho - b_1(T) - b_2(T)$$

where

$$\begin{aligned} b_1(T) &= \frac{E\sum_{t=2}^T \tilde{y}_{t-1}\tilde{u}_t}{E\sum_{t=2}^T \tilde{y}_{t-1}^2} = -\frac{1+\rho}{T} + O\left(T^{-2}\right) \\ b_2(T) &= -\frac{2\rho}{T} + O\left(T^{-2}\right) \end{aligned}$$

It is important to know that the first bias, $b_1(T)$, comes from the correlation between $\tilde{y}_{t-1}$ and $\tilde{u}_t$ (which are the regressor and the regression error after demeaning transformation), and the second bias, $b_2(T)$, comes from the asymmetric distribution of $\hat{\rho}$.

4. In panel regressions, this first part of the small time series bias remains permanently when $N \to \infty$. However the second part of the small bias goes away. The underlying reason is straightforward. As $N \to \infty$, the distribution of $\hat{\rho}_{\text{LSDV}}$ becomes symmetric. Hence the bias arised from asymmetric distribution goes away simply. However the first bias $b_1(T)$ does not go away since this bias is arised because of the time series correlation between the regressor, $\tilde{y}_{t-1}$, and the regression error, $\tilde{u}_t$. More formally, we states

$$\begin{aligned}
\text{plim}_{N\to\infty}\left(\hat{\rho}_{\text{LSDV}} - \rho\right) &= \text{plim}_{N\to\infty} \frac{\frac{1}{N}\sum_{i=1}^{N}\sum_{t=2}^{T}\tilde{y}_{it-1}\tilde{u}_{it}}{\frac{1}{N}\sum_{i=1}^{N}\sum_{t=2}^{T}\tilde{y}_{it-1}^2} \\
&= \frac{\text{plim}_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\sum_{t=2}^{T}\tilde{y}_{it-1}\tilde{u}_{it}}{\text{plim}_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\sum_{t=2}^{T}\tilde{y}_{it-1}^2} \\
&= \frac{E\frac{1}{N}\sum_{i=1}^{N}\sum_{t=2}^{T}\tilde{y}_{it-1}\tilde{u}_{it}}{E\frac{1}{N}\sum_{i=1}^{N}\sum_{t=2}^{T}\tilde{y}_{it-1}^2} \\
&= -\frac{1+\rho}{T} + O\left(T^{-2}\right)
\end{aligned}$$

Asymptotic Bias when $\rho = 1$ Nickell (1981) shows the asymptotic bias (or inconsistency of $\hat{\rho}_{\text{LSDV}}$) when $\rho < 1$. Here we study how the expression of the bias formula badly fails when $\rho = 1$.

1. Consider the following latent model

$$y_t = \alpha + y_t^o, \quad y_t^o = \rho y_{t-1}^o + u_t$$

then we have

$$y_t = \alpha(1-\rho) + \rho y_{t-1} + u_t$$

so that, if $\rho = 1$, then

$$y_t = y_{t-1} + u_t = \sum_{s=1}^{t} u_s = u_1 + ... + u_t$$

2. In the panel data, we have

$$Ey_{it}^2 = E\left(u_{i1} + ... + u_{it}\right)^2 = t\sigma_u^2 \text{ for } Eu_{it}^2 = \sigma_u^2 \text{ for all } i.$$

$$E\frac{1}{T-1}\sum_{t=2}^{T} y_{it-1}^2 = \frac{1}{T-1}\sum_{t=2}^{T} E\left(u_{i1} + ... + u_{it}\right)^2 = \sigma_u^2 \frac{1}{T-1}\sum_{t=1}^{T-1} t = \sigma_u^2 \frac{T}{2}.$$

3. Prove that

$$E\left(\hat{\rho}_{\text{LSDV}} - 1\right) = -\frac{3}{T} + O\left(T^{-2}\right) < -\frac{2}{T} + O\left(T^{-2}\right)$$

13.4 Inconsistency of the Pooled OLS Estimator

Derive the inconsistency of the pooled OLS estimator

$$E\left(\hat{\rho}_{\text{POLS}}\right) = ?$$

1. We are running

$$y_{it} = a + \rho y_{it-1} + e_{it}, \quad e_{it} = a_i - a + u_{it}$$

2. The POLS estimator is given by

$$\hat{\rho}_{\text{POLS}} = \rho + \frac{\sum_{i=1}^{N}\sum_{t=2}^{T}\left(y_{it-1} - \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=2}^{T} y_{it-1}\right)\left(e_{it-1} - \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=2}^{T} e_{it-1}\right)}{\sum_{i=1}^{N}\sum_{t=2}^{T}\left(y_{it-1} - \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=2}^{T} y_{it-1}\right)^2}$$

Note that

$$E\left(\left[\alpha_i - \alpha\right] + y_{it-1}^o\right)\left(\left[\alpha_i - \alpha\right](1-\rho) + u_{it}\right) = \sigma_\alpha^2 (1-\rho)$$

and

$$E\frac{1}{N}\sum_{i=1}^{N}\sum_{t=2}^{T}\left(y_{it-1} - \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=2}^{T} y_{it-1}\right)^2 = \sigma_\alpha^2 + \sigma_y^2$$

3. Let $\eta = \sigma_\alpha^2/\sigma_u^2$. And express the inconsistency in terms of η.

13.5 Asymptotic Distribution of LSDV estimator

$$\begin{aligned}\hat{\rho}_{\text{LSDV}} - \rho &= \frac{\sum_{i=1}^{N}\sum_{t=2}^{T}\tilde{y}_{it-1}\tilde{u}_{it}}{\sum_{i=1}^{N}\sum_{t=2}^{T}\tilde{y}_{it-1}^{2}} \\ &= -\frac{\sum_{i=1}^{N}\left(\sum_{t=2}^{T}y_{it-1}\right)\left(\sum_{t=2}^{T}u_{it}\right)}{\sum_{i=1}^{N}\sum_{t=2}^{T}\tilde{y}_{it-1}^{2}} + \frac{\sum_{i=1}^{N}\sum_{t=2}^{T}y_{it-1}u_{it}}{\sum_{i=1}^{N}\sum_{t=2}^{T}\tilde{y}_{it-1}^{2}}\end{aligned}$$

$$\frac{\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=2}^{T}y_{it-1}u_{it}}{\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=2}^{T}\tilde{y}_{it-1}^{2}} \Longrightarrow^{d} N\left(0, 1-\rho^{2}\right)$$

since

$$\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=2}^{T}y_{it-1}u_{it} \Longrightarrow^{d} N\left(0, \frac{\sigma_u^4}{1-\rho^2}\right)$$

Now we have

$$\begin{aligned}\sqrt{NT}\left(\hat{\rho}_{\text{LSDV}} - \rho\right) &= -\frac{1+\rho}{T}\sqrt{NT} + \frac{\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=2}^{T}y_{it-1}u_{it}}{\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=2}^{T}\tilde{y}_{it-1}^{2}} \\ &= -\left(1+\rho\right)\sqrt{\frac{N}{T}} + \frac{\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=2}^{T}y_{it-1}u_{it}}{\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=2}^{T}\tilde{y}_{it-1}^{2}}\end{aligned}$$

If $\frac{N}{T} \to c$ as $N, T \to \infty$,

$$\sqrt{NT}\left(\hat{\rho}_{\text{LSDV}} - \rho\right) \Longrightarrow^{d} -\left(1+\rho\right)c + N\left(0, 1-\rho^{2}\right)$$

If $\frac{N}{T} \to \infty$ as $N, T \to \infty$,

$$\sqrt{NT}\left(\hat{\rho}_{\text{LSDV}} - \rho\right) \to^{p} \infty$$

If $\frac{N}{T} \to 0$ as $N, T \to \infty$, then

$$\sqrt{NT}\left(\hat{\rho}_{\text{LSDV}} - \rho\right) \Longrightarrow^{d} N\left(0, 1-\rho^{2}\right)$$

Empirical Example Nominal wage $= y_{it}$, $S_i =$ treatment variable or dummy

$$y_{it} = \alpha + \beta S_i + u_{it}, \quad u_{it} = \rho u_{it-1} + \varepsilon_{it}, \quad \varepsilon_{it} \sim iidN\left(0, \sigma^2\right)$$

where S_i is a dummy variable. Suppose that u_{it} is serially correlated ($\rho \neq 0$).

Q1: Find the limiting distribution of $\hat{\beta}$ and $\hat{\alpha}$.

First transform the regression as

$$\begin{aligned} y_{it} - \frac{1}{N}\sum_{i=1}^{N} y_{it} &= \beta\left(S_i - \frac{1}{N}\sum_{i=1}^{N} S_i\right) + \left(u_{it} - \frac{1}{N}\sum_{i=1}^{N} u_{it}\right) \\ \tilde{y}_{it} &= \beta \tilde{S}_i + \tilde{u}_{it}, \text{ let say} \end{aligned}$$

Then

$$\hat{\beta} = \frac{\sum_{i=1}^{N} \tilde{S}_i \left(\sum_{t=1}^{T} \tilde{y}_{it}\right)}{\sum_{i=1}^{N} \tilde{S}_i^2} = \beta + \frac{\sum_{i=1}^{N} \tilde{S}_i \left(\sum_{t=1}^{T} \tilde{u}_{it}\right)}{\sum_{i=1}^{N} \tilde{S}_i^2}$$

Let

$$\hat{\beta} - \beta = \frac{\frac{1}{N}\sum_{i=1}^{N} \tilde{S}_i \left(\sum_{t=1}^{T} \tilde{u}_{it}\right)}{\frac{1}{N}\sum_{i=1}^{N} \tilde{S}_i^2}.$$

Assume that

$$S_i = \begin{cases} 0 \text{ if } i \in G_1 \text{ or } i = 1, ..., \frac{N}{2} \\ 1 \text{ if } i \notin G_1, \text{ or } i = \frac{N}{2}+1, ..., N \end{cases}.$$

$$E\left[\sum_{i=1}^{N} \tilde{S}_i \left(\sum_{t=1}^{T} \tilde{u}_{it}\right)\right]^2 = E\left[\sum_{i=1}^{N} \tilde{S}_i T \bar{u}_i\right]^2$$

where

$$\bar{u}_i = \frac{1}{T}\sum_{t=1}^{T} \tilde{u}_{it}.$$

Observe this

$$\begin{aligned} & E\left[\sum_{i=1}^{N} \tilde{S}_i \left(\sum_{t=1}^{T} \tilde{u}_{it}\right)\right]^2 \\ = \; & E\left[-\frac{1}{2}\sum_{i=1}^{N/2}\sum_{t=1}^{T} \tilde{u}_{it} + \frac{1}{2}\sum_{i=N/2+1}^{N}\sum_{t=1}^{T} \tilde{u}_{it}\right]^2 \\ = \; & E\left[\frac{1}{4}\left(\sum_{i=1}^{N/2}\sum_{t=1}^{T} \tilde{u}_{it}\right)^2 + \frac{1}{4}\left(\sum_{i=N/2+1}^{N}\sum_{t=1}^{T} \tilde{u}_{it}\right)^2 - \frac{1}{2}\left(\sum_{i=1}^{N/2}\sum_{t=1}^{T} \tilde{u}_{it}\right)\left(\sum_{i=N/2+1}^{N}\sum_{t=1}^{T} \tilde{u}_{it}\right)\right] \end{aligned}$$

Note that if there is no cross section dependence, then the last third term becomes zero. Hence we have

$$\begin{aligned}
E\left[\sum_{i=1}^{N}\tilde{S}_i\left(\sum_{t=1}^{T}\tilde{u}_{it}\right)\right]^2 &= \frac{1}{4}E\left(\sum_{i=1}^{N/2}\sum_{t=1}^{T}\tilde{u}_{it}\right)^2+\frac{1}{4}E\left(\sum_{i=N/2+1}^{N}\sum_{t=1}^{T}\tilde{u}_{it}\right)^2 \\
&= \frac{1}{4}\frac{N}{2}E\left(\frac{2}{N}\sum_{i=1}^{N/2}\sum_{t=1}^{T}\tilde{u}_{it}\right)^2+\frac{1}{4}\frac{N}{2}E\left(\frac{2}{N}\sum_{i=N/2+1}^{N}\sum_{t=1}^{T}\tilde{u}_{it}\right)^2 \\
&= \frac{NT}{8}\frac{\sigma^2}{(1-\rho)^2}+\frac{NT}{8}\frac{\sigma^2}{(1-\rho)^2}=\frac{NT}{4}\frac{\sigma^2}{(1-\rho)^2}
\end{aligned}$$

where we use the fact

$$E\left(\sum_{t=1}^{T}\tilde{u}_{it}\right)^2=\frac{\sigma^2}{(1-\rho)^2}.\quad \text{(To students: Prove this)}$$

Note that

$$E\left(\sum_{t=1}^{T}\tilde{u}_{it}\right)^2 > E\sum_{t=1}^{T}\tilde{u}_{it}^2=\frac{\sigma^2}{1-\rho^2}$$

Solution: Use panel robust HAC estimator. Prove this.

Next, Consider the convergence rate: => must be $\sqrt{NT}$. Why?

Limiting Distribution: Major (nice) term and nuisance term For LSDV.

Nice term:

$$G_{NT}=\frac{\sum^{NT}y_{it-1}u_{it}}{\sum^{NT}\tilde{y}_{it-1}^2}$$

Nuisance term:

$$N_{NT}=-\frac{1}{T}\frac{\sum^{N}\left(\sum^{T}y_{it-1}\right)\left(\sum^{T}u_{it}\right)}{\sum^{NT}\tilde{y}_{it-1}^2}$$

Note that

$$\hat{\rho}_{LSDV}-\rho=\frac{\sum^{NT}\tilde{y}_{it-1}\tilde{u}_{it}}{\sum^{NT}\tilde{y}_{it-1}^2}=G_{NT}+N_{NT}=G_{NT}+O_p\left(\frac{1}{?}\right),$$

and G_{NT} is $O_p\left(\frac{1}{\sqrt{NT}}\right)$.

$$\sqrt{NT}G_{NT}=\frac{\frac{1}{\sqrt{NT}}\sum^{NT}y_{it-1}u_{it}}{\frac{1}{NT}\sum^{NT}\tilde{y}_{it-1}^2}\rightarrow^d N\left(0,V^2\right)$$

but

$$
\begin{aligned}
N_{NT} &= -\frac{1}{T}\frac{\frac{1}{NT}\sum^{N}\left(\sum^{T} y_{it-1}\right)\left(\sum^{T} u_{it}\right)}{\frac{1}{NT}\sum^{NT}\tilde{y}_{it-1}^{2}} \\
&= -\frac{1}{T}\frac{\frac{1}{N}\sum^{N}\left(\frac{1}{\sqrt{T}}\sum^{T} y_{it-1}\right)\left(\frac{1}{\sqrt{T}}\sum^{T} u_{it}\right)}{\frac{1}{NT}\sum^{NT}\tilde{y}_{it-1}^{2}} = -\frac{1}{T}\frac{\frac{1}{N}\sum^{N} O_p(1)\, O_p(1)}{O_p(1)} \\
&= -\frac{1}{\sqrt{N}}\frac{1}{T}\frac{\frac{1}{\sqrt{N}}\sum^{N} O_p(1)\, O_p(1)}{O_p(1)} = \frac{O_p(1)}{\sqrt{N}T} = O_p\left(\frac{1}{\sqrt{N}T}\right)
\end{aligned}
$$

Hence

$$
\sqrt{NT}\left(\hat{\rho}_{LSDV} - \rho\right) = \sqrt{NT}G_{NT} + \sqrt{NT}N_{NT} = O_p(1) + O_p\left(\frac{1}{\sqrt{T}}\right)
$$

so that as $T \to \infty$, we can ignore the second term.

Sample Final Exam:

Part I: Definition and Explanation

Q1: Cointegration

Q2: Unit Root Test

Q3: Weakly Stationarity

Q4: Newey and West Estimator

Q5: Panel Robust Covariance Estimator

Q6: White Heteroskedasticity Consistent Estimator

Q7: Nickell Bias

Q8: Relationship among between, within and pooled estimators

Q9: First Difference GMM/IV estimator in Dynamic Panel Regression

Q10: Hausman Test for Fixed Effects

Q11: Granger Causality Test

Q12: Error Correction Model

Part II: Proof and Derivation

Consider the following DGP

$$y_{it} = a_i + y_{it}^o, \quad y_{it}^o = \rho y_{it-1}^o + u_{it}, \quad u_{it} \sim iid\,(0,1)\,, \quad y_{i0}^o = u_{i0}.$$

Q1: Assume $\rho = 1$. You run the following regression

$$y_{it} = \alpha y_{it-1} + e_{it} \tag{57}$$

(a) Show that the pooled OLS estimator in (57) becomes consistent for fixed T and large N. That is,

$$\text{plim}_{N\to\infty}\hat{\alpha}_{\text{pols}} = 1$$

(b) Derive the limiting distribution of $\hat{\alpha}_{\text{pols}}$ when $N, T \to \infty$ jointly.

Now you add fixed effects.

$$y_{it} = \beta_i + \alpha y_{it-1} + \varepsilon_{it} \tag{58}$$

(c) Show that the within group estimator in (58) becomes inconsistent. (for fixed T large N).

(d) Suppose that $N/T \to 0$ as $N, T \to \infty$. Derive the limiting distribution of $\hat{\alpha}_{\text{FE}}$.

Q2: Assume $|\rho| < 1$. You run (57).

(a) Find the moment conditions that the pooled OLS becomes consistent.

(b) Under the conditon of (a), derive the limiting distribution of $\hat{\alpha}_{\text{pols}}$.

14 Method of Moments (Chap 15)

Consider moment conditions such that

$$E(\xi_t - \mu) = 0$$

where ξ_t is a random variable and μ is the unknown mean of ξ_t. The parameter of interest, here, is μ. Consider the following minimum criteria given by

$$\arg\min_{\mu} V_T = \arg\underset{\mu}{\overset{2}{\min}} \frac{1}{T}\sum_{t=1}^{T} (\xi_t - \mu)$$

which becomes the minimum variance of ξ_t with respect to μ. Of course, the simple solution becomes the sample mean for μ since we have

$$\frac{\partial V_T}{\partial \mu} = -2\frac{1}{T}\sum_{t=1}^{T}(\xi_t - \mu) = 0, \quad \Longrightarrow \frac{1}{T}\sum_{t=1}^{T}\xi_t = \mu$$

The above case is the simple example of the method of moment(s).

Now consider more moments such that

$$\begin{aligned}
E(\xi_t - \mu) &= 0 \\
E\left[(\xi_t - \mu)^2 - \gamma_0\right] &= 0 \\
E\left[(\xi_t - \mu)\left(\xi_{t-1} - \mu\right) - \gamma_1\right] &= 0 \\
E\left[(\xi_t - \mu)\left(\xi_{t-2} - \mu\right) - \gamma_2\right] &= 0
\end{aligned}$$

Then we have the four unknowns: $\mu, \gamma_0, \gamma_1, \gamma_2$. We have four sample moments such that

$$\frac{1}{T}\sum_{t=1}^{T}\xi_t, \frac{1}{T}\sum_{t=1}^{T}\xi_t^2, \frac{1}{T}\sum_{t=1}^{T}\xi_t\xi_{t-1}, \frac{1}{T}\sum_{t=1}^{T}\xi_t\xi_{t-2}$$

so that we can solve this numerically.

However, we want to impose further restriction. Suppose that we assume ξ_t follows AR(1) process. Then we have

$$\gamma_1 = \rho\gamma_0, \quad \gamma_2 = \rho\gamma_0$$

so that the total number of unknowns is reducing to three (γ_0, ρ, μ). We can increase more cross moment conditions also. Let $\psi_T = \left(\frac{1}{T}\sum_{t=1}^{T}\xi_t, \frac{1}{T}\sum_{t=1}^{T}\xi_t^2, \frac{1}{T}\sum_{t=1}^{T}\xi_t\xi_{t-1}, \frac{1}{T}\sum_{t=1}^{T}\xi_t\xi_{t-2}\right)'$. Then we have

$$E\frac{1}{T}\sum_{t=1}^{T}(\xi_t - \mu)^2 = E\frac{1}{T}\sum_{t=1}^{T}\xi_t^2 - \mu^2 = \gamma_0$$

so that

$$E\frac{1}{T}\sum_{t=1}^{T}\xi_t^2 = \gamma_0 - \mu^2$$

Also note that

$$E\frac{1}{T}\sum_{t=1}^{T}\xi_t\xi_{t-1} = \rho\gamma_0 - \mu^2, \text{ and so on.}$$

Hence we may consider the following estimation

$$\arg\min_{\mu,\rho,\gamma_0} \left[\psi_T - \psi(\theta)\right]' \left[\psi_T - \psi(\theta)\right]. \tag{59}$$

where θ is the parameters of interest (true parameters, μ, γ_0, ρ). The resulting estimator is called 'method of moments estimator'. Note that MM estimator is a kind of minimum distance estimators.

In general, MM estimator can be used in many cases. However, this method has one weakness. Suppose that the second moment is relatively huge than the first moment. Since V_T function assigns the same weight across moments, the minimum problem in (59) tries to minimize the second moment rather than the first and second moment both. Hence we need to design the optimal weighted method of moments, which becomes generalized method of moments (GMM).

To understand the nature of GMM, we have to study the asymptotic properties of MM estimator. (in order to find the optimal weighting matrix). Now to get the asymptotic distribution of $\hat{\theta}$, we need a Taylor expansion.

$$\psi_T = \psi(\theta) + \frac{\partial \psi_T(\theta)}{\partial \theta'}\left(\hat{\theta} - \theta\right) + O_p\left(\frac{1}{T}\right)$$

so that we have

$$\sqrt{T}\left(\hat{\theta} - \theta\right) = \sqrt{T}\left[\psi_T - \psi(\theta)\right] G(\theta)^{-1} + O_p\left(\frac{1}{\sqrt{T}}\right)$$

where $G_T(\theta) = \frac{\partial \psi_T(\theta)}{\partial \theta'}$. Note that we know that

$$\sqrt{T}\left[\psi_T - \psi(\theta)\right] \to^d N(0, \Phi)$$

Hence we have

$$\sqrt{T}\left(\hat{\theta} - \theta\right) \to^d N\left(0, G(\theta)^{-1} \Phi G(\theta)'^{-1}\right)$$

where $G_T(\theta) \to^p G(\theta)$.

14.1 GMM

First consider infeasible generalized version of method of moments.

$$\arg\min_{\mu,\rho,\gamma_0} \left[\psi_T - \psi(\theta)\right]' \Phi^{-1} \left[\psi_T - \psi(\theta)\right].$$

where Φ is true unknown weighting matrix. Now feasible version becomes

$$\arg\min_{\mu,\rho,\gamma_0} \left[\psi_T - \psi(\theta)\right]' \mathbf{W}_T \left[\psi_T - \psi(\theta)\right] = \arg\min_{\mu,\rho,\gamma_0} G_T(\theta)' \mathbf{W}_T G_T(\theta)$$

where $\mathbf{W}_T$ is a consistent estimator of Φ^{-1}. Let

$$V_T = \left[\psi_T - \psi(\theta)\right]' \mathbf{W}_T \left[\psi_T - \psi(\theta)\right]$$

Then GMM estimator satisfies

$$\frac{\partial V_T\left(\hat{\theta}_{GMM}\right)}{\partial \hat{\theta}_{GMM}} = 2G_T\left(\hat{\theta}_{GMM}\right)' \mathbf{W}_T \left[\psi_T - \psi\left(\hat{\theta}_{GMM}\right)\right] = 0$$

so that we have

$$\psi\left(\hat{\theta}_{GMM}\right)=\psi_T(\theta)+G_T(\theta)\left(\hat{\theta}_{GMM}-\theta\right)+O_p\left(\frac{1}{T}\right)$$

Thus

$$\begin{aligned}&G_T\left(\hat{\theta}_{GMM}\right)'\mathbf{W}_T\left[\psi_T-\psi\left(\hat{\theta}_{GMM}\right)\right]\\=\;&G_T\left(\hat{\theta}_{GMM}\right)'\mathbf{W}_T\left[\psi_T-\psi\left(\hat{\theta}_{GMM}\right)\right]+G_T\left(\hat{\theta}_{GMM}\right)'\mathbf{W}_TG_T(\theta)\left(\hat{\theta}_{GMM}-\theta\right)=0\end{aligned}$$

Hence

$$\left(\hat{\theta}_{GMM}-\theta\right)=-\left\{G_T\left(\hat{\theta}_{GMM}\right)'\mathbf{W}_TG_T(\theta)\right\}^{-1}G_T\left(\hat{\theta}_{GMM}\right)'\mathbf{W}_T\left[\psi_T-\psi\left(\hat{\theta}_{GMM}\right)\right]$$

and

$$\sqrt{T}\left(\hat{\theta}_{GMM}-\theta\right)\to^d N(0,V)$$

where

$$V=\frac{1}{T}\left\{G'\mathbf{W}G\right\}^{-1}G'\mathbf{W}\mathbf{\Phi}\mathbf{W}G\left\{G'\mathbf{W}G\right\}^{-1}.$$

When $W=\Phi^{-1}$, then we have

$$V=\frac{1}{T}\left\{G'\mathbf{\Phi}^{-1}G\right\}^{-1}G'\mathbf{\Phi}^{-1}G\left\{G'\mathbf{\Phi}^{-1}G\right\}^{-1}=\frac{1}{T}\left\{G'\mathbf{\Phi}^{-1}G\right\}^{-1}.$$