

Panel Data Models

Table of contents

1 Overview	2
2 Model Specification	2
2.1 Unobserved heterogeneity	3
3 Exogeneity	3
3.1 Strict exogeneity	4
3.2 Weak exogeneity	4
3.3 Contemporaneous exogeneity	5
4 Basic model	5
5 Pooled OLS	6
6 Between-Group estimator	7
7 Generalized Least Squares	8
7.1 Feasible GLS	9
8 Within-Group Estimator	10
9 Conditional variance	12
9.1 Consistency and asymptotic distribution	13
9.2 Fixed-effects estimator	14
10 First-Difference Estimator	15
11 Wu-Hausman Test	16

1 Overview

In this handout we will see how to test **static linear** panel data models. We will review a number of estimators for these models, including:

- within-group;
- first differences;
- between-group;
- and feasible GLS.

Further reading can be found in:

- Section 21 of Cameron and Trivedi (2005)
- Section 10.1-10.3 of Verbeek (2017)

2 Model Specification

A static linear model has the form,

$$Y_{it} = X'_{it}\beta + \alpha_i + \varepsilon_{it}$$

for $i = 1, \dots, n$ and $t = 1, \dots, T$. We will restrict this discussion to models where T is fixed. Asymptotically, this means that as n increases T remains fixed.

If collect all T observations of unit i , we can describe them by the model,

$$Y_i = X'_i\beta + \alpha_i\ell + \varepsilon_i$$

where Y_i is a $T \times 1$ random vector; X_i a $T \times k$ random matrix; and ℓ a $T \times 1$ vector of 1's.

This model has placed no restriction on the values of the outcome variable, Y_i , and regressors, X_i . In particular, the regressors may include time-varying as well as time-invariant variables.

We can extend this specification to include a linear or non-linear trend in time; for example, ϕt . However, the more common option is to include a very flexible time trend using fixed effects:

$$\delta_t = \sum_{j=1}^T \delta_j \mathbf{1}\{t = j\}$$

Time fixed-effects are essentially a dummy variable for each time-period. This flexible function - sometimes referred to as saturated - can approximate any functional form of the underlying time trend. Models that include both α_i and δ_t are referred to as **two-way fixed-effects** models.

2.1 Unobserved heterogeneity

The term α_i is particularly important. It represents time-invariant unobservables or **unobserved heterogeneity** across i . I strongly recommend you read the discussion on pages 285-286 of Wooldridge (2010).

The α_i is unobserved, and therefore is a component of the error term $v_{it} = \alpha_i + \varepsilon_{it}$. The time-invariant component of the error is *permanent*; often referred to as unobserved heterogeneity or individual effect. The time-varying component is *transient*; sometimes referred to as the idiosyncratic error.

As with the time-varying part of the error term, α_i is *random*. We should not think of α_i as a unit-specific intercept, since then it is just a non-random parameter. This would imply that the correlation between α_i and X_i is by definition 0.

Given the model specification, we must consider all three components - $\{X_i, \alpha_i, \varepsilon_{it}\}$ - when making assumptions regarding exogeneity. **We will make all assumptions conditional on α_i .** This is consistent with the idea that α_i is a permanent shock, and is therefore realized before ε_{it} .

If we do not condition on α_i , then when we evaluate $E[Y_i|X_i]$, we also need to consider both $E[\varepsilon_i|X_i]$ and $E[\alpha_i|X_i]$; not just $E[\varepsilon_i|X_i, \alpha_i]$. Moreover, in general it will be that case that $E[\alpha_i|X_i] \neq E[\alpha_i]$ (i.e., not mean independent). For example, α_i may represent unobserved ability in a wage equation. This will correlated with regressors like education.

3 Exogeneity

Having assumed that the samples are independent across i , we can define mean independence of the transient error term component for unit i . There are three potential assumptions we can make:

1. Strict exogeneity:

$$E[\varepsilon_i|X_i, \alpha_i] = 0$$

or,

$$E[\varepsilon_{it}|X_{i1}, X_{i2}, \dots, X_{iT}, \alpha_i] = 0 \quad \forall t$$

2. Weak or sequential exogeneity:

$$E[\varepsilon_{it}|X_{i1}, X_{i2}, \dots, X_{it}, \alpha_i] = 0 \quad \forall t$$

Exogeneity with respect to the past sequence of regressors (or predetermined regressors).

3. Contemporaneous exogeneity:

$$E[\varepsilon_{it}|X_{it}, \alpha_i] = 0 \quad \forall t$$

Exogeneity only with respect to the contemporaneous value of X_i .

3.1 Strict exogeneity

Strict exogeneity is a very strong assumption. It implies that X is uncorrelated with past, current, and future values of the transient error term (conditional on α_i): $E[X_{it}, \varepsilon_{is}|\alpha_i] = 0 \quad \forall t, s$. Crucially, X_{it} cannot respond to the history of idiosyncratic shocks $\varepsilon_{i1}, \varepsilon_{i2}, \dots, \varepsilon_{it}$.

This assumption is required for unbiasedness of the linear estimator, but not for consistency. Under strict exogeneity,

$$E[Y_{it}|X_i, \alpha_i] = E[Y_{it}|X_{it}, \alpha_i] = X'_{it}\beta + \alpha_i$$

The first equality implies that once you control for X_{it} , there is no additional partial effect of X_{is} ($\forall s \neq t$) on (the mean of) Y_{it} . This assumption relates to the assumed *static* nature of the model: the model includes not lags or leads among the regressors.

It would be violated if the set of regressors included a lagged dependent variable. The assumption is also violated if the regressors are endogenous.

3.2 Weak exogeneity

Weak exogeneity, also referred to as sequential exogeneity, implies that the error term is uncorrelated with past and contemporaneous values of the regressors:

$$E[X_{is}\varepsilon_{it}] = 0 \quad \forall s = 1, \dots, t$$

In the static linear model, it also implies that,

$$E[Y_{it}|X_i, \alpha_i] = E[Y_{it}|X_{it}, \alpha_i] = X'_{it}\beta + \alpha_i$$

This structure can permit a lagged dependent variable amongst the regressors (assuming there is no serial correlation of the error term). However, the assumption remains violated if the regressors are endogenous.

3.3 Contemporaneous exogeneity

This assumption implies that the error term is only uncorrelated with regressors in the same time period,

$$E[X_{it}\varepsilon_{it}] = 0$$

Regardless, it still implies that,

$$E[Y_{it}|X_i, \alpha_i] = E[Y_{it}|X_{it}, \alpha_i] = X'_{it}\beta + \alpha_i$$

4 Basic model

We begin by describe a basic model under the Static Linear Panel Model (SLPM) assumptions,

SLPM 1: The model is static and linear in parameters:

$$Y_{it} = X'_{it}\beta + \alpha_i + \varepsilon_{it}$$

SLPM 2: Balanced panel: we observe each i in all T time periods.

SLPM 3: Independent sampling along the cross-sectional dimension (i).

SLPM 4: Strict exogeneity: $E[\varepsilon_{it}|X_i, \alpha_i] = 0 \forall t$

SLPM 5: Conditional homoskedasticity and serial uncorrelatedness of the transient error term component.

$$Var(\varepsilon_i|X_i, \alpha_i) = \begin{bmatrix} \sigma_\varepsilon^2 & 0 & \cdots & 0 \\ 0 & \sigma_\varepsilon^2 & & \\ \vdots & & \ddots & \\ 0 & & & \sigma_\varepsilon^2 \end{bmatrix} = \sigma_\varepsilon^2 I_T$$

Assumptions 4 & 5 are referred to as a ‘classical error term’ structure.

Under assumptions 1-5:

- $E[Y_i|X_i, \alpha_i] = X'_i\beta + \alpha_i$
- $Var(Y_i|X_i, \alpha_i) = \sigma_\varepsilon^2 I_T$

At this stage, the unanswered question is how to deal with the unobservables α_i in the equation. Generally, there are two approaches:

1. Assume away the relationship between α_i and X_i .
2. Remove α_i from the equation prior to estimator; for example, within-group estimator and first-difference.

Adopting approach (1), we will review the pooled OLS, between-group, and (feasible) Generalized Least Squares (GLS) estimators. Under approach (2), will review the within-group and first-difference estimators.

5 Pooled OLS

We can ‘assume away’ the relationship between α_i and X_i . Specifically, we will assume conditional mean independence:

- **CLPM 6** $E[\alpha_i|X_i] = E[\alpha_i|X_{i1}, X_{i2}, \dots, X_{iT}] = E[\alpha_i] = 0$

As in the CLRM, the assumption that the unconditional mean of α_i is zero is not binding given the inclusion of a constant among the regressors. We will of course, need to assume that there is no perfect collinearity among the regressors.

- **CLPM 7** $\text{rank}(X) = k$

In addition, we will need to make assumptions concerning α_i . First, mean independence,

This implies uncorrelatedness. Under this assumptions, the model can be written as,

$$Y_{it} = X'_{it}\beta + v_{it}$$

where $E[v_i|X_i] = 0$. Alternatively, we can stack all nT observations together,

$$Y = X\beta + v$$

Second, we need to make an assumption regarding the variance of the unobserved heterogeneity:

- **CLPM 8** $\text{Var}(\alpha_i|X_i) = \sigma_\alpha^2$

Under these assumptions, the error term $v_i = \alpha_i\ell + \varepsilon_i$ has the variance,

$$\text{Var}(v_i|X_i) = E[v_i v'_i|X_i] = \sigma_\alpha^2 \ell \ell' + \sigma_\varepsilon^2 I_T = \Omega$$

For all $s \neq t$ the $\text{Cov}(v_{it}, v_{is}) = \sigma_\alpha^2$. And the diagonal elements are given by $\sigma_\alpha^2 + \sigma_\varepsilon^2$.

Under these assumptions, the OLS estimator,

$$\begin{aligned}
\hat{\beta}^{OLS} &= (X'X)^{-1}X'Y \\
&= \beta + (X'X)^{-1}X'v \\
&= \beta + \left(\sum_i X'_i X_i \right)^{-1} \sum_i X'_i v_i
\end{aligned}$$

is both unbiased and consistent. This is because,

$$p \lim \frac{1}{n} \sum_i X'_i v_i = \sum_{t=1}^T p \lim \frac{1}{n} \sum_{i=1}^n X_{it} v_{it} = \sum_{t=1}^T E[X_{it} v_{it}] = 0$$

The asymptotic distribution is given by,

$$\begin{aligned}
\sqrt{n}(\hat{\beta}^{OLS} - \beta) &= \left(\frac{1}{n} \sum_i X'_i X_i \right)^{-1} \frac{1}{\sqrt{n}} \sum_i X'_i v_i \\
&\rightarrow_d N(0, V^{-1} \Sigma V^{-1})
\end{aligned}$$

where,

- $V = E[X'_i X_i]$
- $\Sigma = E[X'_i \Omega X_i]$

We say that the approximate distribution of the pooled OLS estimator is given by,

$$\hat{\beta}^{OLS} \overset{a}{\sim} N\left(\beta, \left(\sum_i X'_i X_i \right)^{-1} \sum_i X'_i \Omega X_i \left(\sum_i X'_i X_i \right)^{-1}\right)$$

However, it is NOT efficient. Moreover, the usual homoskedastic estimator for the variance will be biased and inconsistent. Unobserved heterogeneity will result in a serial correlation that is not accounted for by the standard estimator.

6 Between-Group estimator

An alternative to pooled OLS, is to collapse the multiple observations of unit i into a single cross-section aggregate. This transforms the model into,

$$\bar{Y}_i = \bar{X}'_i \beta + \bar{v}_i$$

where $\bar{v}_i = \alpha_i + \bar{\varepsilon}_i$. The variance of this error term is,

$$E[\bar{v}_i^2|X_i] = \sigma_\alpha^2 + \frac{\sigma_\varepsilon^2}{T}$$

The OLS estimator for β is given by,

$$\hat{\beta}^{BG} = (\bar{X}'\bar{X})^{-1}\bar{X}'\bar{Y} = \left(\sum_i \bar{X}_i\bar{X}_i'\right)^{-1}\sum_i \bar{X}_i\bar{Y}_i$$

Since the variance term is homoskedastic, the standard homoskedastic variance estimator will be unbiased and consistent. This approach removes the problem of serially correlated error terms across repeated observations of i in pooled OLS by collapsing all observations to a single observation. However, it also reduces the information in the data and is therefore less efficient.

7 Generalized Least Squares

This final approach takes seriously the structure of the composite error-term variance-covariance matrix,

$$E[v_i v_i' | X_i] = \begin{bmatrix} \sigma_\alpha^2 + \sigma_\varepsilon^2 & \sigma_\alpha^2 & \cdots & \sigma_\alpha^2 \\ \sigma_\alpha^2 & \sigma_\alpha^2 + \sigma_\varepsilon^2 & & \\ \vdots & & \ddots & \\ \sigma_\alpha^2 & & & \sigma_\alpha^2 + \sigma_\varepsilon^2 \end{bmatrix}$$

The $nT \times nT$ matrix, $E[v v' | X]$ is a block-diagonal matrix in which the off-diagonal values are $E[v_{it} v_{js} | X] = \sigma_\alpha^2$ only for $i = j$ and $s \neq t$; and zero otherwise.

The Generalized Least Squares solution says that if we transform the model in the following way, the resulting error term will be classical.

$$\underbrace{Y_{it} - \theta \bar{Y}_i}_{Y_{it}^+} = \underbrace{(X_{it} - \theta \bar{X}_i)'}_{X_{it}^+} \beta + \nu_{it}$$

where,

$$\theta = 1 - \frac{\sigma_\varepsilon}{\sqrt{T\sigma_\alpha^2 + \sigma_\varepsilon^2}}$$

Consider the transformed error term ν ,

$$\nu_{it} = v_{it} - \theta \bar{v}_i = (1 - \theta)\alpha_i + \varepsilon_{it} - \frac{\theta}{T} \sum_t \varepsilon_{it}$$

The serial correlation of this error term is 0. Consider, for $t \neq s$

$$\begin{aligned}
E[\nu_{it}\nu_{is}|X_i] &= E\left[\left((1-\theta)\alpha_i + \varepsilon_{it} - \frac{\theta}{T} \sum_{t'} \varepsilon_{it'}\right)\left((1-\theta)\alpha_i + \varepsilon_{is} - \frac{\theta}{T} \sum_{t'} \varepsilon_{it'}\right) | X_i\right] \\
&= (1-\theta)^2 \sigma_\alpha^2 - 2\frac{\theta}{T} \sigma_\varepsilon^2 + \frac{\theta^2}{T^2} \sum_{t'} \sigma_\varepsilon^2 \\
&= \frac{\sigma_\varepsilon^2 \sigma_\alpha^2}{T \sigma_\alpha^2 + \sigma_\varepsilon^2} + \frac{\theta(\theta-2)}{T} \sigma_\varepsilon^2 \\
&= \frac{\sigma_\varepsilon^2 \sigma_\alpha^2}{T \sigma_\alpha^2 + \sigma_\varepsilon^2} - \frac{\sigma_\varepsilon^2}{T} \left(1 - \frac{\sigma_\varepsilon}{\sqrt{T \sigma_\alpha^2 + \sigma_\varepsilon^2}}\right) \left(1 + \frac{\sigma_\varepsilon}{\sqrt{T \sigma_\alpha^2 + \sigma_\varepsilon^2}}\right) \\
&= \frac{\sigma_\varepsilon^2 \sigma_\alpha^2}{T \sigma_\alpha^2 + \sigma_\varepsilon^2} - \frac{\sigma_\varepsilon^2}{T} \left(1 - \frac{\sigma_\varepsilon^2}{T \sigma_\alpha^2 + \sigma_\varepsilon^2}\right) \\
&= \frac{\sigma_\varepsilon^2 \sigma_\alpha^2}{T \sigma_\alpha^2 + \sigma_\varepsilon^2} - \frac{\sigma_\varepsilon^2 \sigma_\alpha^2}{T \sigma_\alpha^2 + \sigma_\varepsilon^2} \\
&= 0
\end{aligned}$$

The GLS estimator is then given by,

$$\hat{\beta}^{GLS} = \left[\sum_i X_i^{+'} X_i^+ \right]^{-1} \sum_i X_i^{+'} Y_i^+$$

You can show that $\hat{\beta}^{GLS}$ is a weighted average of $\hat{\beta}^{WG}$ and $\hat{\beta}^{BG}$. Also, take note of the fact that as $T \rightarrow \infty$, $\theta \rightarrow 1$. Thus, $\hat{\beta}_{GLS} \rightarrow \hat{\beta}^{WG}$ as $T \rightarrow \infty$. In addition, if $\sigma_\alpha^2 = 0$, then $\theta = 0$ and $\hat{\beta}^{GLS} = \hat{\beta}^{OLS}$, the pooled OLS estimator.

However, this estimator is NOT feasible. This is because we do not observe $\{\sigma_\alpha^2, \sigma_\varepsilon^2\}$.

7.1 Feasible GLS

A feasible version of the GLS estimator is given by the following steps:

1. Estimate σ_ε^2 using the WG estimator.
2. Use the pooled OLS or BG estimator then estimate σ_α^2 , using the value of σ_ε^2 from step 1. For example, the RSS from pooled OLS (divided by $nT - k$) is a consistent estimator for $\sigma_\varepsilon^2 + \sigma_\alpha^2$. Similar, the RSS from BG estimator (divided by $n - k$) is a consistent estimator for $\sigma_\varepsilon^2/T + \sigma_\alpha^2$.
3. Using the estimated $\{\hat{\sigma}_\alpha^2, \hat{\sigma}_\varepsilon^2\}$, compute the transformed model and estimate using $\hat{\beta}^{FGLS}$ using OLS.

8 Within-Group Estimator

The approach taken here is to transform the linear model such that α_i is eliminate from the equation. Having done so, we do not need to make any assumption regarding $E[\alpha_i|X_i]$.

Here we will exploit the fact that α_i is time-invariant. We begin by computing the unit-level average of the model. For the left-hand side,

$$\bar{Y}_i = \frac{1}{T} \sum_{t=1}^T Y_{it}$$

and the right-hand side,

$$\frac{1}{T} \sum_{t=1}^T (X'_{it}\beta + \alpha_i + \varepsilon_{it}) = \bar{X}'_i\beta + \alpha_i + \bar{\varepsilon}_i$$

Next, subtract this from each value, to create a demeaned expression

$$\underbrace{Y_{it} - \bar{Y}_i}_{\tilde{Y}_{it}} = \underbrace{(X_{it} - \bar{X}_i)'\beta}_{\tilde{X}'_{it}\beta} + \underbrace{\alpha_i - \alpha_i}_{=0} + \underbrace{\varepsilon_{it} - \bar{\varepsilon}_i}_{\tilde{\varepsilon}_{it}}$$

The permanent error-term component drops out precisely because it is time-invariant. The transformed model is given by,

$$\tilde{Y}_{it} = \tilde{X}'_{it}\beta + \tilde{\varepsilon}_{it}$$

This model can be estimated by OLS. The solution is given by,

$$\hat{\beta}^{WG} = (\tilde{X}'\tilde{X})^{-1}\tilde{X}'\tilde{Y} = \left(\sum_{i=1}^n \tilde{X}'_i\tilde{X}_i\right)^{-1} \sum_{i=1}^n \tilde{X}'_i\tilde{Y}_i$$

where \tilde{X}_i is a $T \times k$ matrix. **Note, this is different to the standard cross-section expression\$.**

This solution assumes that the $k \times k$ matrix $\sum_{i=1}^n \tilde{X}'_i\tilde{X}_i$ is invertible. We must therefore assume,

- **SLPM 7^b:** $rank(\tilde{X}) = k$

Given that \tilde{X} is the within-group demeaned value of X , this implies the regressors must all be time-varying. In addition, if the model includes a time trend (including time-FEs), the included variables cannot vary uniformly with time. An example of this is age. If all units' age values increase by the same amount each period, then

$$a\tilde{g}e_{it} = age_{it} - a\bar{g}e_i = t - \bar{t}$$

This is because an individual's age can be expressed as a time-invariant value of their year of birth yob_i plus a linear time-trend that has a unit-specific intercept. The demeaned value of age is perfectly colinear with a demeaned linear time-trend. It would also be perfectly colinear with a higher-order polynomial time-trend or year FEs.

$$\tilde{X}_i = \begin{bmatrix} \tilde{X}_{i11} & \tilde{X}_{i12} & \cdots & \tilde{X}_{i1k} \\ \tilde{X}_{i21} & \tilde{X}_{i22} & & \\ \vdots & & \ddots & \\ \tilde{X}_{iT1} & & & \tilde{X}_{iT k} \end{bmatrix} = \begin{bmatrix} \tilde{X}'_{i1} \\ \tilde{X}'_{i2} \\ \vdots \\ \tilde{X}'_{iT} \end{bmatrix}$$

$\tilde{X}'_i \tilde{X}_i$ is therefore a $k \times k$ matrix, which can be expressed as,

$$\tilde{X}'_i \tilde{X}_i = \sum_{t=1}^T \tilde{X}_{it} \tilde{X}'_{it} = \begin{bmatrix} \sum_t \tilde{X}_{it1}^2 & \sum_t \tilde{X}_{it1} \tilde{X}_{it2} & \cdots & \sum_t \tilde{X}_{it1} \tilde{X}_{itk} \\ \sum_t \tilde{X}_{it2} \tilde{X}_{it1} & \sum_t \tilde{X}_{it2}^2 & & \\ \vdots & & \ddots & \\ \sum_t \tilde{X}_{itk} \tilde{X}_{it1} & & & \sum_t \tilde{X}_{itk}^2 \end{bmatrix}$$

Finally, we can express $\tilde{X}' \tilde{X}$ as,

$$\tilde{X}' \tilde{X} = \sum_{i=1}^n \tilde{X}'_i \tilde{X}_i = \begin{bmatrix} \sum_i (\sum_t \tilde{X}_{it1}^2) & \sum_i (\sum_t \tilde{X}_{it1} \tilde{X}_{it2}) & \cdots & \sum_i (\sum_t \tilde{X}_{it1} \tilde{X}_{itk}) \\ \sum_i (\sum_t \tilde{X}_{it2} \tilde{X}_{it1}) & \sum_i (\sum_t \tilde{X}_{it2}^2) & & \\ \vdots & & \ddots & \\ \sum_i (\sum_t \tilde{X}_{itk} \tilde{X}_{it1}) & & & \sum_i (\sum_t \tilde{X}_{itk}^2) \end{bmatrix}$$

We can use the same method to describe $\sum_{i=1}^n \tilde{X}'_i \tilde{Y}_i$, a $k \times 1$ vector. Substituting in the definition of \tilde{Y}_i from the transformed model, we get that,

$$\hat{\beta}^{WG} = \beta + \left(\sum_{i=1}^n \tilde{X}'_i \tilde{X}_i \right)^{-1} \sum_{i=1}^n \tilde{X}'_i \tilde{\varepsilon}_i$$

9 Conditional variance

Given the demeaning of the model, the error term is no longer uncorrelated across t for the same i :

$$\begin{aligned} E[\tilde{\varepsilon}_{is}\tilde{\varepsilon}_{it}|X] &= E[(\varepsilon_{is} - \bar{\varepsilon}_i)(\varepsilon_{it} - \bar{\varepsilon}_i)|X] \\ &= E\left[\left(\varepsilon_{is} - \frac{1}{T} \sum_{s'} \varepsilon_{is'}\right)\left(\varepsilon_{it} - \frac{1}{T} \sum_{s'} \varepsilon_{it'}\right) \middle| X_i\right] \\ &= \begin{cases} \sigma_\varepsilon^2(1 - 1/T) & \text{for } s = t \\ -\sigma_\varepsilon^2/T & \text{for } s \neq t \end{cases} \end{aligned}$$

The off-diagonal elements for the same i are the same for all time-periods. Collected together, this creates a block diagonal matrix, where covariance terms are zero for across different units.

$$E[\tilde{\varepsilon}_{is}\tilde{\varepsilon}_{it}|X] = \begin{bmatrix} \Omega & 0 & \cdots & 0 \\ 0 & \Omega & & \\ \vdots & & \ddots & \\ 0 & & & \Omega \end{bmatrix}$$

with each $T \times T$ Ω matrix given by,

$$\begin{aligned} \Omega &= E[\tilde{\varepsilon}_i \tilde{\varepsilon}_i' | X_i] \\ &= \begin{bmatrix} \sigma_\varepsilon^2(1 - 1/T) & -\sigma_\varepsilon^2/T & \cdots & -\sigma_\varepsilon^2/T \\ -\sigma_\varepsilon^2/T & \sigma_\varepsilon^2(1 - 1/T) & & \\ \vdots & & \ddots & \\ -\sigma_\varepsilon^2/T & & & \sigma_\varepsilon^2(1 - 1/T) \end{bmatrix} \\ &= \sigma_\varepsilon^2 M_\ell \end{aligned}$$

where $M_\ell = I_T - \frac{\ell\ell'}{T}$ for the $T \times 1$ vector of ones ℓ . This follows from the fact that $\tilde{\varepsilon}_i = M_\ell \varepsilon_i$. Which implies that,

$$\begin{aligned} E[\tilde{\varepsilon}_i \tilde{\varepsilon}_i' | X_i] &= E[M_\ell \varepsilon_i \varepsilon_i' M_\ell' | X_i] \\ &= M_\ell E[\varepsilon_i \varepsilon_i' | X_i] M_\ell \\ &= M_\ell \sigma_\varepsilon^2 I_T M_\ell \\ &= \sigma_\varepsilon^2 M_\ell \end{aligned}$$

Thus,

$$E[\tilde{\varepsilon}_{is}\tilde{\varepsilon}_{it}|X] = \sigma_\varepsilon^2 \begin{bmatrix} M_\ell & 0 & \cdots & 0 \\ 0 & M_\ell & & \\ \vdots & & \ddots & \\ 0 & & & M_\ell \end{bmatrix} = \sigma_\varepsilon^2 M_{n,\ell}$$

where the $rank(M_{n\cdot\ell}) = nT - n$. This follows from the fact that each M_ℓ has $rank(M_\ell) = T - 1$. The matrix $M_{n\cdot\ell}$ is a $nT \times nT$ matrix: a block diagonal matrix of n M_ℓ matrices. With this we can now solve for the conditional variance of the WG estimator.

$$\begin{aligned} Var(\hat{\beta}^{WG}|X) &= (\tilde{X}'\tilde{X})^{-1}\tilde{X}'E[\tilde{\varepsilon}_{is}\tilde{\varepsilon}_{it}|X]\tilde{X}(\tilde{X}'\tilde{X})^{-1} \\ &= \sigma_\varepsilon^2(\tilde{X}'\tilde{X})^{-1}\tilde{X}'M_{n\cdot\ell}\tilde{X}(\tilde{X}'\tilde{X})^{-1} \\ &= \sigma_\varepsilon^2(\tilde{X}'\tilde{X})^{-1} \end{aligned}$$

The final line follows from the fact that $\tilde{X} = M_{n\cdot\ell}X$ and $M_{n\cdot\ell}$ is an idempotent (orthogonal) projection matrix.

An unbiased and consistent estimator for σ_ε^2 is given by,

$$\hat{\sigma}_\varepsilon^2 = \frac{RSS}{dof} = \frac{\sum_{i=1}^n \sum_{t=1}^T (Y_{it} - \bar{Y}_i - (X_{it} - \bar{X}_i)' \hat{\beta}^{WG})^2}{nT - n - k}$$

The residual degrees of freedom is equal to $nT - n - k$, not $nT - k$. While there are nT observations and k parameters, we must deduct the n unit-level means computed. Another way to see this is to use the above projection matrix decomposition:

$$\begin{aligned} RSS &= \hat{\varepsilon}'\hat{\varepsilon} \\ &= \tilde{Y}'(I_{nT} - \tilde{X}(\tilde{X}'\tilde{X})^{-1}\tilde{X}')\tilde{Y} \\ &= Y'M_{n\ell}(I_{nT} - M_{n\ell}X(X'M_{n\ell}X)^{-1}X'M_{n\ell})M_{n\ell}Y \\ &= Y'\underbrace{(M_{n\ell} - M_{n\ell}X(X'M_{n\ell}X)^{-1}X'M_{n\ell})}_{M_+}Y \end{aligned}$$

This new matrix M_+ is also an idempotent projection matrix, with rank $rank(M_+) = rank(M_{n\ell}) - minrank(M_{n\ell}), rank(X) = nT - n - k$. As we saw with the CLRM, this term will have a χ^2 distribution with dof equal to the rank of the orthogonal projection matrix.

9.1 Consistency and asymptotic distribution

Under assumptions SLPM 1-6; especially 4 & 5, $\hat{\beta}^{WG}$ is consistent,

$$\hat{\beta}^{WG} \rightarrow_p \beta \quad \text{as } n \rightarrow \infty$$

and asymptotically normal,

$$\hat{\beta}^{WG} \overset{a}{\sim} N(\beta, \sigma_\varepsilon^2(\sum_i \tilde{X}'_i \tilde{X}_i)^{-1})$$

Both results require that $E[\tilde{X}_i, \tilde{\varepsilon}_i] = 0$. This is maintained by strict exogeneity (CLPM 4), which states that in addition to $E[X_{is}, \varepsilon_{it}] = 0 \forall s, t$,

$$E[\bar{X}_i, \bar{\varepsilon}_i] = 0$$

This would not be true under the weak exogeneity assumption.

9.2 Fixed-effects estimator

The above estimator is sometimes referred to as the (unit) Fixed-Effects estimator. This not mean that α_i is non-random. It simply means that the within-group estimator is also given by a fixed-effects model. Consider the model,

$$Y_{it} = \sum_{j=1}^n \phi_j \mathbf{1}\{i = j\} + X'_{it}\beta + v_{it}$$

This model includes a dummy variable for each unit. For each unit, only one dummy variable can be =1, the dummy variable with parameter ϕ_i . Thus, the expression $\sum_{j=1}^n \phi_j \mathbf{1}\{i = j\} = \phi_i$ for any i . The model can therefore be written as,

$$Y_{it} = \phi_i + X'_{it}\beta + v_{it}$$

This looks very similar to our SLPM, but with the distinguishing feature that ϕ_i is taken as a population parameter. This is sometimes referred as a unit-specific constant.

Including a constant for each i has the same effect as demeaning the model. This alternative approach is referred to as the Least Squares Dummy Variable (LSDV) method.

$$\hat{\beta}^{WG} = \hat{\beta}^{LSDV}$$

This equivalence can be shown using Frisch Waugh Lovell Theorem (or partitioned regression). The LSDV estimator can be computed in two steps. First, regress each regressor and outcome on a setting of unit level dummies.

$$X_k = \sum_{j=1}^n \phi_j \mathbf{1}\{i = j\} + \xi$$

Each $\mathbf{1}\{i = j\}$ corresponds to a dummy variable where T values are =1 (for unit i), and the remainder 0. Next, use the residuals in the main equation. The residual from the regression is given by,

$$M_{n \cdot \ell} X_k = \tilde{X}_k$$

Thus, we regression is given by,

$$M_{n,\ell}Y = \tilde{Y} = M_{n,\ell}X\beta + \varepsilon = \tilde{X}\beta + \varepsilon$$

Employing the LSDV approach, one can estimated the unit-FE as,

$$\hat{\phi}_i = \bar{Y}_i - \bar{X}_i' \hat{\beta}^{LSDV}$$

While this estimator is unbiased, $E[\hat{\phi}_i|X_i] = \phi$, it is NOT consistent for fixed T . It is only consistent if $T \rightarrow \infty$ as $n \rightarrow \infty$.

The equivalene of these approaches also explains why the degrees of freedom in the residual is $nT - n - k$. By including n dummy variables, the number of parameters we need to estimate is $n + k$.

10 First-Difference Estimator

As with the WG estimator, the first-difference (FD) estimator removes α_i from the model through differencing. However, this time the transformation is just difference over time:

$$\underbrace{Y_{it} - Y_{it-1}}_{\tilde{Y}_{it}} = \underbrace{(X_{it} - X_{it-1})' \beta}_{\tilde{X}_{it}' \beta} + \underbrace{\alpha_i - \alpha_i}_{=0} + \underbrace{\varepsilon_{it} - \varepsilon_{it-1}}_{\tilde{\varepsilon}_{it}}$$

The estimation sample will include 1 less period: $t = 2, \dots, T$ for each i .

Under CLPM 1-6, $\hat{\beta}^{FD}$ is consistent and asymptotically normal. However, we need to account for the error term structure. This is because,

$$E[\tilde{\varepsilon}_{is} \tilde{\varepsilon}_{it}] = \begin{cases} E[(\varepsilon_{is} - \varepsilon_{is-1})(\varepsilon_{it} - \varepsilon_{it-1})] = 2\sigma_\varepsilon^2 & \text{for } s = t \\ E[(\varepsilon_{is} - \varepsilon_{is-1})(\varepsilon_{it} - \varepsilon_{it-1})] = -\sigma_\varepsilon^2 & \text{for } s = t - 1 \\ 0 & \text{otherwise} \end{cases}$$

This is a MA(1) error term structure, in which the first-order correlation is non-zero.

For $T = 2$, you can show that,

$$\hat{\beta}^{WG} = \hat{\beta}^{FD}$$

You can also show that if you apply GLS to the first-differenced equation, you get the WG estimator (for $T \geq 2$).

11 Wu-Hausman Test

The Wu-Hausman test is used to test the exogeneity assumption underlying a particular estimator. You need two estimators: $\{\hat{\beta}_1, \hat{\beta}_2\}$ such that,

- Under $H_0 : \beta_1 = \beta_2$
- $\hat{\beta}_1$ is consistent
- $\hat{\beta}_2$ is consistent
- $Var(\hat{\beta}_1|X) < Var(\hat{\beta}_2|X)$: the former is more efficient
- Under $H_1 : \beta_1 \neq \beta_2$
- $\hat{\beta}_1$ is inconsistent
- $\hat{\beta}_2$ is consistent

The test statistic is given by,

$$\text{Stat} = (\hat{\beta}_2 - \hat{\beta}_1)' (Var(\hat{\beta}_2 - \hat{\beta}_1|X))^{-1} (\hat{\beta}_2 - \hat{\beta}_1)$$

Under H_0 , this statistic converges in distribution to χ^2_k , where k is the number of regressors. The inner matrix is the inverse of the variance-covariance matrix.

$$\begin{aligned} Var(\hat{\beta}_2 - \hat{\beta}_1|X) &= Var(\hat{\beta}_2|X) + Var(\hat{\beta}_1|X) - 2Cov(\hat{\beta}_2, \hat{\beta}_1|X) \\ &= Var(\hat{\beta}_2|X) - Var(\hat{\beta}_1|X) \end{aligned}$$

Line 2 follows from line 1, because of a result demonstrated by Hausman (1978):

$$0 = Cov(\hat{\beta}_1, \hat{\beta}_1 - \hat{\beta}_2) = Var(\hat{\beta}_1) - Cov(\hat{\beta}_1, \hat{\beta}_2)$$

This result holds in cases where the variances of the respective estimators can be ranked: i.e. one estimator is more efficient than the other. As a result, the $Var(\hat{\beta}_2 - \hat{\beta}_1|X)$ matrix is positive definite. This implies that its inverse exists.

This test can be applied in this setting to test the null: $H_0 : E[X_{it}\alpha_i] = 0$ (i.e. uncorrelatedness). This condition must hold for the (F)GLS estimator to be consistent. If H_0 is false, then we know that the WG estimator is consistent, but that it is less efficient. To test this hypothesis we use the coefficients from the FGLS and WG estimators. Of course, this means that you can only test for restrictions on time-varying regressors (a restriction of WG estimator).¹

Cameron, A Colin, and Pravin K Trivedi. 2005. *Microeconometrics: Methods and Applications*. Cambridge university press.

¹This restriction can be dropped using Mundlak's correction.

- Verbeek, Marno. 2017. *A Guide to Modern Econometrics*. John Wiley & Sons.
- Wooldridge, Jeffrey M. 2010. *Econometric Analysis of Cross Section and Panel Data*. MIT press.