Panel Data Models - part I

(Wooldridge, chapters 10)

1 What is Panel Data

- A time series of cross sections where the same individual units are followed over a number of time periods - that is, a collection of N time series.
- Two sample dimensions: Cross-sectional (N, indexed by $i = 1, ...N$) and Time-series (T, indexed by $t = 1, ...T$).
- Two processes can be used,
  - Individual units observed until "lost".
  - Individual units observed for a finite number of time periods and then dropped. If there is no attrition, this is a *balanced panel*.
- Unbalanced panels usually result from *attrition*, whereby individual units can be lost at some point.
  - If the attrition process is independent of the dependent variable, then attrition is *exogenous* and balanced and unbalanced panels share the same properties. This is the maintained assumption in this course and we assume $T_i = T$, $\forall i$, to simplify the notation.
  - Otherwise, attrition is *endogenous* (for example, the dependent variable is firm profit and firm failure results from negative profits). As time passes the sample of individuals becomes less and less representative of the population. This selection process *must* be modelled.

- The observable information in a balanced panel with K explanatory variables is,

$$\mathbf{y} = \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \vdots \\ \mathbf{y}_N \end{bmatrix} = \begin{bmatrix} y_{11} \\ \vdots \\ y_{1T} \\ y_{21} \\ \vdots \\ y_{2T} \\ \vdots \\ y_{NT} \end{bmatrix} \quad \text{and} \quad X = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_N \end{bmatrix} = \begin{bmatrix} x_{111} & x_{112} & \cdots & x_{11K} \\ \vdots & \vdots & & \vdots \\ x_{1T1} & x_{1T2} & \cdots & x_{1TK} \\ x_{211} & x_{212} & \cdots & x_{21K} \\ \vdots & \vdots & & \vdots \\ x_{2T1} & x_{2T2} & \cdots & x_{2TK} \\ \vdots & \vdots & & \vdots \\ x_{NT1} & x_{NT2} & \cdots & x_{NTK} \end{bmatrix}$$

  where we use small letters for variables, small bold letter for vectors and capital letters for matrices. Above, $\mathbf{y}$ is a column of dimension NT, $\mathbf{y}_i$ is a column of dimension T for each $i = 1, \ldots, N$, X is a matrix $NT \times K$ and X_i is a matrix $T \times K$ for $i = 1, \ldots, N$.

- Different types of panel data,

  1. Household panels.
  2. Individual level panels.
  3. Firm level panels.
  4. Countries followed over time.
  5. Industries followed over time.

2 Why/When do we need Panel Data

1. Endogeneity: panel data may offer the solution to deal with *unobserved heterogeneity* across individuals when IV is not adequate.

   - Omitted variable problem: if an important explanatory variable is unobservable and related with the observables, then the fundamental OLS assumption of no-correlation between the error term and the regressors is violated. As a consequence, OLS is inconsistent.
   - The cross-sectional solution is to use IV.
   - When Panel Data is available, we may have other alternatives.

- To exemplify, suppose the omitted variable f is constant over time. The model is,

$$y_{it} = \mathbf{x}_{it}\beta + f_i + u_{it}$$

- Taking first differences eliminates f,

$$\Delta y_{it} = \Delta \mathbf{x}_{it}\beta + \Delta u_{it}$$

- OLS can be applied to consistently estimate β for as long as,

$$E(\Delta \mathbf{x}'\Delta u) = 0$$

2. Dynamics: panel data is required in the estimation of dynamic economic models in the presence of individual level heterogeneity.

   - Most economic decisions are dynamic in nature. Consider the following examples,
     - Labour supply and human capital formation: an agent deciding about labour supply takes into account his previous labour market experience (which affects human capital) and expectations of future gains from present work (which again affects human capital). So past earnings (reflecting human capital) affect future earnings.
     - Habit formation: past consumption patterns affect the utility of consumption in the future. So past levels of consumption affect future levels of consumption.
   - In these examples, present decisions are a function of non-contemporaneous variables as well as of contemporaneous variables.

3 Three types of estimators

1. Consistent for fixed T (time dimension) as $N \to \infty$ (cross section dimension). Suitable for cases where N is large and T is relatively small.

2. Consistent as $T \to \infty$ and N is fixed. Suitable for cases where T is large and N is relatively small.

3. Consistent as both N and $T \to \infty$. Suitable for cases where both N and T are large.

We will study case 1. One must be careful when applying these estimators to the other cases since some of the arguments do not hold in the other cases. In particular when T is small we do not generally need to worry about non-stationarity of the regressors.

4 The observation sample

- Sample of observations $\left\{(y_{it}, \mathbf{x}_{it}) \in \mathbb{R} \times \mathbb{R}^K, i = 1, ..., N, t = 1, ..., T\right\}$.
- We assume **Random sampling** - sample is iid across individuals:
  1. $(y_{it}, \mathbf{x}_{it})_{t=1,...,T} \perp\!\!\!\perp (y_{jt}, \mathbf{x}_{jt})_{t=1,...,T}, \forall_{i \neq j}$;
  2. $(y_{it}, \mathbf{x}_{it})_{t=1,...,T}$ and $(y_{jt}, \mathbf{x}_{jt})_{t=1,...,T}$ have the same distribution, $\forall_{i \neq j}$.

5 The model

The basic model we consider is:

$$\begin{aligned} y_{it} &= \mathbf{x}_{it}\beta + e_{it} \\ &= \mathbf{x}_{it}\beta + f_i + u_{it} \end{aligned}$$

for $i = 1, ..., N$ and $t = 1, ..., T$ and where

- $e_{it} = f_i + u_{it}$ is the unobservable component. f_i is the unobserved effect and u_{it} is the idiosyncratic time-varying shock.
- The absence of a t subscript from f_i implies that it does not vary over time.
- The regressors $\mathbf{x}_{it}$ may or may not vary over time.
- $\mathbf{x}_{it}$ is $1 \times K$ and β is $K \times 1$.

6 Alternative assumptions on the error components

6.1 Strict versus weak exogeneity

Or how u is related with $\mathbf{x}$.

- *Strict exogeneity*

$$\begin{aligned} E\left(u_{it} \,\middle|\, \{\mathbf{x}_{is}\}_{s=1,...,T}\right) &= E\left(u_{it} | X_i\right) \\ &= E\left(u_{it}\right) = 0 \end{aligned}$$

Notice that $E(\mathbf{x}_{it}u_{is}) = 0$ is implied by this assumption.

- *Weak exogeneity* or predetermined regressors

$$E\left(u_{it}\left|\{\mathbf{x}_{is}\}_{s\leqslant t}\right.\right) = 0$$

Notice that we are always assuming that $E\left(u_{it}\left|\{X_j\}_{j=1,\ldots,N}\right.\right) = E\left(u_{it}\left|X_i\right.\right)$ - random sampling.

6.2 Random effects versus fixed effects

Or how f is related with $\mathbf{x}$.

- *Random effects*

$$E(f_i|\mathbf{x}_{i1},...\mathbf{x}_{iT}) = E(f_i) = 0$$

The implication is that

$$\text{corr}(\mathbf{x}_{it}, f_i) = 0$$

- *Fixed effects*

$$E(f_i|\mathbf{x}_{i1},...\mathbf{x}_{iT}) = E(f_i|X_i) = g(X_i)$$

where g is a non-constant function of X_i. The implication is that generally,

$$\text{corr}(\mathbf{x}_{it}, f_i) \neq 0$$

7 The random effects model

7.1 Assumptions

1. *Random individual effects:* $E\left(f_i|\mathbf{x}_{i1},...,\mathbf{x}_{iT}\right) = E\left(f_i\right) = 0$.

2. *Strictly exogenous regressors:* $E\left(u_{it}|\mathbf{x}_{i1},...,\mathbf{x}_{iT}\right) = E\left(u_{it}\right) = 0$.

3. *Homoscedasticity:*

$$\begin{aligned} \operatorname{var}(f_i|\mathbf{x}_{i1},...,\mathbf{x}_{iT}) &= \sigma_f^2 \\ \operatorname{var}(u_{it}|\mathbf{x}_{i1},...,\mathbf{x}_{iT}) &= \sigma_u^2 \\ \operatorname{cov}(u_{it},u_{is}|\mathbf{x}_{i1},...,\mathbf{x}_{iT}) &= 0 \quad (t \neq s) \end{aligned}$$

4. *Linearly independent regressors:*

$$\text{plim}_{N\to\infty}\frac{1}{N}X'X = E\left(X_i'X_i\right) = M_{XX}$$

is positive definite for all T.

Assumptions 1 to 3 imply that,

$$\begin{aligned} E(e_{it}|\mathbf{x}_{i1},...,\mathbf{x}_{iT}) &= 0 \\ \operatorname{cov}(e_{it},e_{is}|\mathbf{x}_{i1},...,\mathbf{x}_{iT}) &= \sigma_f^2 + \delta_{ts}\sigma_u^2 \end{aligned}$$

where $\delta_{ts} = 1$ if $t = s$ and 0 otherwise (Kronecker delta).

7.2 The OLS estimator

The OLS estimator is,

$$\begin{aligned} \beta^{OLS} &= (X'X)^{-1}X'\mathbf{y} \\ &= (X'X)^{-1}X'(X\beta + \mathbf{e}) \\ &= \beta + (X'X)^{-1}X'\mathbf{e} \end{aligned}$$

where $\mathbf{e} = [\mathbf{e}_1', \ldots, \mathbf{e}_N']'$ and $\mathbf{e}_i = [e_{i1}, \ldots, e_{iT}]'$ for $i = 1, \ldots, N$.

Properties of OLS under assumptions 1 to 4

- *It is unbiased*: $E\left[\beta^{OLS}|X\right] = \beta$ since $E[\mathbf{e}|X] = 0$
- *It is consistent*:

$$\underset{N\to\infty}{\text{plim}}\,\beta^{OLS} = \beta + \left(\underset{N\to\infty}{\text{plim}}\frac{1}{N}X'X\right)^{-1}\left(\underset{N\to\infty}{\text{plim}}\frac{1}{N}X'\mathbf{e}\right)$$

By the LLN,

$$\begin{aligned}\operatorname*{plim}_{N\to\infty}\frac{1}{N}X'X &= \operatorname*{plim}_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}X_i'X_i \\ &= E\left(X_i'X_i\right)=M_{XX}\end{aligned}$$

and by assumption 4, M_{XX} is pd, and hence invertible.

On the other hand, assumptions 1 and 2 ensure that,

$$\begin{aligned}\operatorname*{plim}_{N\to\infty}\frac{1}{N}X'\mathbf{e} &= \operatorname*{plim}_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}X_i'\mathbf{e}_i \\ &= E\left(X_i'\mathbf{e}_i\right)=0\end{aligned}$$

which proves consistency.

Remark: OLS requires less than strict exogeneity for consistency: only contemporaneous correlation needs to be excluded.

- *But it is not efficient*: since it does not explore the structure of the error term.

7.3 Variance of the error

- Under assumption 3, the elements of the covariance matrix for $\mathbf{e}_i$ are,

$$\begin{aligned}E\left(e_{it}^2|X_i\right) &= \sigma_f^2+\sigma_u^2 \\ E\left(e_{it}e_{is}|X_i\right) &= \sigma_f^2 \quad \text{for } t\neq s\end{aligned}$$

- Thus for each individual i we can write the $T\times T$ covariance matrix as

$$\begin{aligned}V &= E(\mathbf{e}_i\mathbf{e}_i'|X_i) \\ &= \sigma_u^2 I_T+\sigma_f^2 J_T\end{aligned}$$

where I_T is the identity matrix of size $T\times T$ and J_T is a $T\times T$ matrix of 1's.

7.4 The Generalised Least Squares Estimator (GLS)

- The GLS estimator can be applied under assumptions 1-3 and the following alternative to assumption 4:

4'. *Linearly independent regressors:*

$$\begin{aligned}\operatorname*{plim}_{N\to\infty} \frac{1}{N} X'\mathbb{V}^{-1}X &= \frac{1}{N}\operatorname*{plim}_{N\to\infty} \sum_{i=1}^{N} X_i' V^{-1} X_i \\ &= E\left(X_i' V^{-1} X_i\right) = M_{XVX}\end{aligned}$$

is positive definite for all T.

where $\mathbb{V} = \operatorname{diag}\{V_1, V_2, ..., V_N\} = \operatorname{diag}\{V, V, ..., V\}$ is $NT \times NT$.

- The GLS estimator is obtained from the transformation of the model,

$$V^{-\frac{1}{2}}\mathbf{y}_i = V^{-\frac{1}{2}} X_i \beta + V^{-\frac{1}{2}}\mathbf{e}_i$$

which if estimated by OLS yields,

$$\beta^{GLS} = \left(\sum_{i=1}^{N} X_i' V^{-1} X_i\right)^{-1} \left(\sum_{i=1}^{N} X_i' V^{-1}\mathbf{y}_i\right)$$

Properties of GLS under assumptions 1 to 3 and 4'

- *It is consistent*

$$\operatorname*{plim}_{N\to\infty} \beta^{GLS} = \beta + \left(\operatorname*{plim}_{N\to\infty} \frac{1}{N} X'\mathbb{V}^{-1}X\right)^{-1} \left(\operatorname*{plim}_{N\to\infty} \frac{1}{N} X'\mathbb{V}^{-1}\mathbf{e}\right)$$

By the LLN,

$$\begin{aligned}\operatorname*{plim}_{N\to\infty} \frac{1}{N} X'\mathbb{V}^{-1}X &= \operatorname*{plim}_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} X_i' V^{-1} X_i \\ &= E\left(X_i' V^{-1} X_i\right) = M_{XVX}\end{aligned}$$

which by assumption 4' is pd.

On the other hand, the LLN and the LIE ensure that

$$\begin{aligned}\operatorname*{plim}_{N\to\infty} \frac{1}{N} X'\mathbb{V}^{-1}\mathbf{e} &= \operatorname*{plim}_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} X_i' V^{-1}\mathbf{e}_i \\ &= E\left(X_i' V^{-1}\mathbf{e}_i\right) \\ &= E\left(X_i' V^{-1} E\left(\mathbf{e}_i | X_i\right)\right) = 0\end{aligned}$$

which proves consistency.

- *It is asymptotically normally distributed*

$$\sqrt{N}\left(\beta^{GLS}-\beta\right) = \left(\frac{1}{N}\sum_{i=1}^{N}X_i'V^{-1}X_i\right)^{-1}\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N}X_i'V^{-1}\mathbf{e}_i\right)$$

By the CLT

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}X_i'V^{-1}\mathbf{e}_i \overset{d}{\to} \mathcal{N}\left(0,\text{var}\left(X_i'V^{-1}\mathbf{e}_i\right)\right)$$

where

$$\begin{aligned}\text{var}\left(X_i'V^{-1}\mathbf{e}_i\right) &= E\left[X_i'V^{-1}\mathbf{e}_i\mathbf{e}_i'V^{-1}X_i\right]\\ &= E\left[X_i'V^{-1}E\left(\mathbf{e}_i\mathbf{e}_i'|X_i\right)V^{-1}X_i\right]\\ &= E\left[X_i'V^{-1}VV^{-1}X_i\right]\\ &= E\left[X_i'V^{-1}X_i\right]\end{aligned}$$

But then,

$$\sqrt{N}\left(\beta^{GLS}-\beta\right) \overset{d}{\to} \mathcal{N}\left(0,E\left[X_i'V^{-1}X_i\right]^{-1}\right)$$

7.5 Feasible GLS - FGLS

- The GLS estimator above is not implementable because the σ_u and σ_f are not known. We develop a feasible procedure.
- The resulting estimator will no longer be unbiased, but in general it will be consistent and asymptotically efficient.
- Using the fact that OLS is consistent under assumptions 1 to 4, we can use its predicted residuals to estimate the variance of e_{it}:

$$\begin{aligned}E\widehat{\left(e_{it}^2|X_i\right)} &= \widehat{\sigma_u^2+\sigma_f^2}\\ &= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\widehat{e}_{it}^2\end{aligned}$$

and the covariance between e_{it} and e_{is} for any $t\neq s$:

$$\begin{aligned}E\left(\widehat{e_{it}e_{is}|}X_i\right) &= \widehat{\sigma}_f^2\\ &= \frac{1}{N(T-1)}\sum_{i=1}^{N}\sum_{t=1,t\neq s}^{T}\widehat{e}_{it}\widehat{e}_{is}\end{aligned}$$

- Then

$$\widehat{\sigma}_u^2 = \widehat{\sigma_f^2 + \sigma_u^2} - \widehat{\sigma}_f^2$$

- We can now construct $\widehat{V}$ and substitute it into the GLS formula.
- This "works" because OLS is consistent. The feasible estimator is neither linear nor unbiased. This is because the estimated quantity $\widehat{V}$ is a function of the dependent variable $\mathbf{y}$.

8 The fixed effects model with strict exogeneity

- Consider the model,

$$y_{it} = \mathbf{x}_{it}\beta + \mathbf{z}_i\gamma + f_i + u_{it} \quad i = 1, ..., N \text{ and } t = 1, ..., T$$

where we have distinguished between those explanatory variables that vary with time ($\mathbf{x}_{it}$) and those that do not vary over time ($\mathbf{z}_i$).

- We now consider the case where some of the x's are endogenous but the endogeneity can be modelled as a dependence between the regressors and an unobserved component that is fixed over time. This is of course a modelling assumption and in practice would have to be justified by an economic model.
- We continue considering assumption 2 but relax assumption 1 by considering,

$$E(f_i|X_i) \neq 0$$

where f is constant over time.

8.1 The Within Groups estimator

- To tackle the endogeneity problem, we define the individual-specific means,

$$\overline{y}_i = \frac{1}{T}\sum_{t=1}^{T} y_{it}, \qquad \overline{\mathbf{x}}_i = \frac{1}{T}\sum_{t=1}^{T} \mathbf{x}_{it}, \qquad \overline{\mathbf{z}}_i = \frac{1}{T}\sum_{t=1}^{T} \mathbf{z}_i = \mathbf{z}_i$$

- The Within Groups (WG) estimator uses centered observations,

$$\widetilde{y}_{it} = y_{it} - \overline{y}_i, \qquad \widetilde{\mathbf{x}}_{it} = \mathbf{x}_{it} - \overline{\mathbf{x}}_i, \qquad \widetilde{\mathbf{z}}_i = \mathbf{z}_i - \overline{\mathbf{z}}_i = \mathbf{0}$$

- The average model is

$$\overline{y}_i = \overline{\mathbf{x}}_i\beta + \mathbf{z}_i\gamma + f_i + \overline{u}_i$$

  and the centered model is,

$$\widetilde{\mathbf{y}}_i = \widetilde{X}_i\beta + \widetilde{\mathbf{u}}_i$$

- The WG estimator of β is the OLS estimator applied to the centered model:

$$\beta^{WG} = \left(\sum_{i=1}^N \widetilde{X}_i'\widetilde{X}_i\right)^{-1}\left(\sum_{i=1}^N \widetilde{X}_i'\widetilde{\mathbf{y}}_i\right)$$

- Note that all covariates that are time-independent disappear from the centered model, so we will not be able to identify their impact.

8.2 The WG matrix operator

- The centered observations are the residuals of the regressions of the variable on a constant only:

$$\widetilde{y}_{it} = y_{it} - \widehat{y}_{it} \quad \text{where} \quad \widehat{\mathbf{y}}_i = \mathbf{1}_T\left(\mathbf{1}_T'\mathbf{1}_T\right)^{-1}\mathbf{1}_T'\mathbf{y}_i$$
$$\widetilde{x}_{itk} = x_{itk} - \widehat{x}_{itk} \quad \text{where} \quad \widehat{\mathbf{x}}_{ik} = \mathbf{1}_T\left(\mathbf{1}_T'\mathbf{1}_T\right)^{-1}\mathbf{1}_T'\mathbf{x}_{ik} \quad \text{for } k = 1, ..., K$$

  and where $\mathbf{1}_T$ is a column vector of size T filled with 1's.

- Define

$$P_T = \mathbf{1}_T\left(\mathbf{1}_T'\mathbf{1}_T\right)^{-1}\mathbf{1}_T'$$
$$Q_T = I_T - P_T$$

  We can write, in matrix notation,

$$\widetilde{\mathbf{y}}_i = Q_T\mathbf{y}_i$$
$$\widetilde{X}_i = Q_T X_i$$

- Since Q_T is symmetric and idempotent ($Q_TQ_T = Q_T$), the WG estimator can be re-written as

$$\begin{aligned}\beta^{WG} &= \left(\sum_{i=1}^N \widetilde{X}_i'\widetilde{X}_i\right)^{-1}\left(\sum_{i=1}^N \widetilde{X}_i'\widetilde{\mathbf{y}}_i\right) \\ &= \left(\sum_{i=1}^N X_i'Q_TX_i\right)^{-1}\left(\sum_{i=1}^N X_i'Q_T\mathbf{y}_i\right)\end{aligned}$$

8.3 Properties of the WG estimator

WG assumptions

2. *Strictly exogenous regressors:* $E(u_{it}|\mathbf{x}_{i1},...,\mathbf{x}_{iT}) = E(u_{it}) = 0$.

3". *Homoscedasticity:*

$$\begin{aligned} \operatorname{var}(u_{it}|\mathbf{x}_{i1},...,\mathbf{x}_{iT}) &= \sigma_u^2 \\ \operatorname{cov}(u_{it}, u_{is}|\mathbf{x}_{i1},...,\mathbf{x}_{iT}) &= 0 \quad (t \neq s) \end{aligned}$$

4". *Linearly independent regressors:*

$$\operatorname*{plim}_{N\to\infty} \frac{1}{N}\widetilde{X}'\widetilde{X} = E\left(\widetilde{X}_i'\widetilde{X}_i\right) = M_{\widetilde{X}\widetilde{X}}$$

is positive definite for all T. Note that $\widetilde{X}$ can only include the time-varying regressors for this assumption to hold.

Properties of the WG estimator under assumptions 2, 3" and 4"

- *It is unbiased*: $E\left[\beta^{WG}|X\right] = \beta$ since $E[\mathbf{u}|X] = \mathbf{0}$.
- *It is consistent*:

$$\begin{aligned} \operatorname*{plim}_{N\to\infty} \beta^{WG} &= \beta + \left(\operatorname*{plim}_{N\to\infty}\frac{1}{N}\widetilde{X}'\widetilde{X}\right)^{-1}\left(\operatorname*{plim}_{N\to\infty}\frac{1}{N}\widetilde{X}'\widetilde{\mathbf{u}}\right) \\ &= \beta + \left(\operatorname*{plim}_{N\to\infty}\frac{1}{N}\widetilde{X}'\widetilde{X}\right)^{-1}\left(\operatorname*{plim}_{N\to\infty}\frac{1}{N}\widetilde{X}'\mathbf{u}\right) \end{aligned}$$

By the LLN,

$$\begin{aligned} \operatorname*{plim}_{N\to\infty}\frac{1}{N}\widetilde{X}'\widetilde{X} &= \operatorname*{plim}_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\widetilde{X}_i'\widetilde{X}_i \\ &= E\left(\widetilde{X}_i'\widetilde{X}_i\right) = M_{\widetilde{X}\widetilde{X}} \end{aligned}$$

which is pd by assumption 4".

On the other hand, assumption 2" ensures that,

$$\begin{aligned} \operatorname*{plim}_{N\to\infty}\frac{1}{N}\widetilde{X}'\mathbf{u} &= \operatorname*{plim}_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\widetilde{X}_i'\mathbf{u}_i \\ &= E\left(\widetilde{X}_i'\mathbf{u}_i\right) = 0 \end{aligned}$$

which proves consistency.

- *It is asymptotically normally distributed*

$$\sqrt{N}\left(\beta^{WG}-\beta\right) = \left(\frac{1}{N}\sum_{i=1}^{N}\widetilde{X}_i'\widetilde{X}_i\right)^{-1}\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\widetilde{X}_i'u_i\right)$$

By the CLT

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\widetilde{X}_i'u_i \xrightarrow{d} \mathcal{N}\left(0,\mathrm{var}\left(\widetilde{X}_i'u_i\right)\right)$$

where

$$\begin{aligned}\mathrm{var}\left(\widetilde{X}_i'u_i\right) &= E\left[\widetilde{X}_i'u_iu_i'\widetilde{X}_i\right]\\ &= E\left[\widetilde{X}_i'E\left(u_iu_i'|X_i\right)\widetilde{X}_i\right]\\ &= E\left[\widetilde{X}_i'\left(\sigma_u^2 I_T\right)\widetilde{X}_i\right]\\ &= \sigma_u^2 E\left[\widetilde{X}_i'\widetilde{X}_i\right]\end{aligned}$$

And then,

$$\sqrt{N}\left(\beta^{WG}-\beta\right) \xrightarrow{d} \mathcal{N}\left(0,\sigma^2 E\left[\widetilde{X}_i'\widetilde{X}_i\right]^{-1}\right)$$

9 Comparing the GLS and WG estimators

- Start by considering the GLS estimator,

$$\beta^{GLS} = \left(\sum_{i=1}^{N}X_i'V^{-1}X_i\right)^{-1}\left(\sum_{i=1}^{N}X_i'V^{-1}\mathbf{y}_i\right)$$

where V is the covariance matrix of the error term and we have seen that,

$$V = \sigma_u^2 I_T + \sigma_f^2 J_T$$

- But then,

$$\begin{aligned}V &= \sigma_u^2 I_T + \sigma_f^2 J_T\\ &= \sigma_u^2 I_T + T\sigma_f^2 P_T\\ &= \sigma_u^2(P_T+Q_T) + T\sigma_f^2 P_T\\ &= \sigma_u^2(Q_T+\eta P_T)\end{aligned}$$

where $\eta = \frac{\sigma_u^2+T\sigma_f^2}{\sigma_u^2}$ and $Q_T = I_T - P_T$.

- You can check that,

$$V^{-1} = \frac{1}{\sigma_u^2}\left(Q_T + \frac{1}{\eta}P_T\right)$$

- The GLS estimator can be re-written as,

$$\beta^{GLS} = \left(\underbrace{\sum_{i=1}^{N} X_i' Q_T X_i}_{\text{within group}} + \underbrace{\frac{1}{\eta}\sum_{i=1}^{N} X_i' P_T X_i}_{\text{between group}}\right)^{-1}\left(\underbrace{\sum_{i=1}^{N} X_i' Q_T \mathbf{y}_i}_{\text{within group}} + \underbrace{\frac{1}{\eta}\sum_{i=1}^{N} X_i' P_T \mathbf{y}_i}_{\text{between group}}\right)$$

- And the estimator for the asymptotic covariance matrix of the GLS estimator is,

$$\begin{aligned}\widehat{\text{var}}\left(\beta^{GLS}\right) &= \left(\sum_{i=1}^{N} X_i' V X_i\right)^{-1} \\ &= \sigma_u^2\left(\underbrace{\sum_{i=1}^{N} X_i' Q_T X_i}_{\text{within group}} + \underbrace{\frac{1}{\eta}\sum_{i=1}^{N} X_i' P_T X_i}_{\text{between group}}\right)^{-1}\end{aligned}$$

- To compare with the variance of the WG estimator suppose for simplicity that all regressors are time varying. The estimator for the asymptotic covariance matrix of the WG estimator is,

$$\widehat{\text{var}}\left(\beta^{WG}\right) = \sigma_u^2\left(\sum_{i=1}^{N} X_i' Q_T X_i\right)^{-1}$$

- Now note that for any two positive definite matrices of the same dimensions, A and B, the matrix $A-B$ is positive semi definite if and only if $B^{-1} - A^{-1}$ is positive semi-definite.

- We apply this to compare the covariance matrices for the two estimators:

$$\sigma_u^2\left[\widehat{\text{var}}\left(\beta^{GLS}\right)^{-1} - \widehat{\text{var}}\left(\beta^{WG}\right)^{-1}\right] = \frac{1}{\eta}\sum_{i=1}^{N} X_i' P_T X_i$$

  which is positive semi-definite since P_T is symmetric and idempotent and $\eta > 0$

- Thus, GLS is at least as efficient as WG and identifies all the parameters of the model while WG only identifies the parameters attached to time-varying regressors.

- But WG is more robust as it does not require mean independence between f and the regressors.