2. Static Linear Models

Linear (or log linear) static model.

$y_{it} = \mathbf{x}_{it}\beta + c_i + u_{it}$ [$\mathbf{x}_{it}$ is 1xK vector; β is Kx1 vector]

$\mathbf{y}_i = \mathbf{X_i}\beta + c_i + u_\mathbf{i}$ [$\mathbf{y}_i$ is T x 1; $\mathbf{X_i}$ is T x K]

Dependent variable y_{it} i=1,.....,N t=1,...,T.

- No restriction on the values the dependent variable can take. [**v.imp**]

Explanatory variables x_{it} a vector of K variables.

(i) can vary over i,t.

(ii) can vary over i only.

(iii) can vary over t only.

TYPES OF **MODELS**

1. Fixed Effects (FE) model.

2. Random Effects (RE) model.

- Problem with the interpretation – a lot of confusion
- To do with the estimation……
- Mainly dependent on whether $E(c_i \mid \mathbf{x}_{i1},\ldots,\mathbf{x}_{iT}) = E(c_i)$
- Should treat the c as a random draw and then choose the estimation method depending on the assumption you make regarding various exogeneity statuses.

2.1. Fixed Effects (FE) Model

$y_{it} = \mathbf{x}_{it}\beta + c_i + u_{it}$ (1)

What does FE mean here? Estimation does not depend on the specification of c.

Assumptions

FE1a. $E(u_{it} \mid \mathbf{x_{i1},\ldots,x_{iT}}, c_i) = 0$ [st exog cond on c] t=1,…,T.

Implies $E(y_{it} \mid \mathbf{x_{i1},\ldots,x_{iT}}, c_i) = \mathbf{x}_{it}\beta + c_i$

FE1b. $E(c_i \mid \mathbf{x_{i1},\ldots,x_{iT}}) \neq E(c_i)$ **Important assumption here.**

[allows for arbitrary corr]

Problem: Can't distinguish time-invariant covariates from c here.

Estimation

Basic principle – find a transformation that will eliminate the c_i.

One such transformation is **within transformation** – takes within group mean deviations.

Notation: $\mathbf{x}_{i.} = \dfrac{\sum_t \mathbf{x}_{it}}{T}$; $\mathbf{x}_{..} = \dfrac{\sum_i \mathbf{x}_{i.}}{N}$; $\tilde{\mathbf{x}}_{it} = \mathbf{x}_{it} - \mathbf{x}_{i.}$;

Averaging (1) over *i* gives

$$y_{i.} = \mathbf{x}_{i.}\beta + c_i + u_{i.} \qquad (2)$$

Subtracting (2) from (1) now gives

$$\tilde{y}_{it} = \tilde{\mathbf{x}}_{it}\boldsymbol{\beta} + \tilde{\mathrm{u}}_{it} \qquad (3)$$

Apply OLS to (3) to get $\hat{\boldsymbol{\beta}}_w$.

$$\hat{\boldsymbol{\beta}}_w = \left[\sum_i \tilde{\mathbf{X}}_{\mathbf{i}}'\tilde{\mathbf{X}}_{\mathbf{i}}\right]^{-1}\left[\sum_i \tilde{\mathbf{X}}_{\mathbf{i}}'\tilde{\mathbf{y}}_{\mathbf{i}}\right] = \mathbf{W}_{\mathbf{xx}}^{\mathbf{-1}}\mathbf{w}_{\mathbf{xy}} \qquad \{ = [\mathbf{X}'\mathbf{X}]^{-1}[\mathbf{X}'\mathbf{y}] \}$$

$\mathbf{x_{it}}$ is 1 x K; $\mathbf{X}_{\mathrm{i}}$ is T x K; $\mathbf{X}$ is NT x K;

$\mathbf{y_i}$ is T x 1

NOTES:

(i) Under strict exogeneity assumption for the **x,** this would give you unbiased (conditional on **x)** and consistent estimator for **β** for large N <u>or</u> large T <u>or</u> large N & T. **WHY?**

(ii) Time invariant variable effects cannot be estimated. **WHY?**

(iii) The above is also known as **within-group estimator (WG).**

(iv) If we assume that

FE3: $E(\mathbf{u_i}\, \mathbf{u_i}' \mid \mathbf{x}_i, c_i) = \sigma_u^2\, \mathbf{I}_T$, [cond covar = uncod covar]

then the WG estimator is also efficient (since u_{it} have constant variance and serially uncorrelated).

(v) Can show that $\sqrt{N}\left(\hat{\boldsymbol{\beta}}_w - \boldsymbol{\beta}\right) \sim Normal\left(0, \sigma_u^2 [\mathrm{E}(\tilde{\mathbf{X}}_i' \tilde{\mathbf{X}}_i)]^{-1}\right)$ [need the rank condition **FE2:** $rank\,\mathrm{E}(.)$ is K].

Thus, $\text{Avar}\,(\hat{\boldsymbol{\beta}}_w) = \sigma_u^2 [\mathrm{E}(\tilde{\mathbf{X}}_i' \tilde{\mathbf{X}}_i)]^{-1} / N$.

Sample analogue would be $\hat{\sigma}_u^2 [\sum_i (\tilde{\mathbf{X}}_i' \tilde{\mathbf{X}}_i)]^{-1}$

with $\hat{\sigma}_u^2 = \dfrac{WGres\ \mathrm{SS}}{NT - N - K} = \dfrac{SSR_w}{N(T-1) - K}$. (4)

(vi) If you assume c_i is a parameter to estimate, you can create N-1 dummies and use OLS. **PROPERTIES OF THE ESTIMATOR?**

(vii) Same $\hat{\boldsymbol{\beta}}_W$ as before – apply Frisch-Waugh theorem - LSDV model.

$\hat{c}_\text{i} = y_{i.} - \mathbf{x}_{i.} \hat{\boldsymbol{\beta}}_w$ based on T observations.

$\hat{c}_\text{i}$ is unbiased (conditional on the **x** s) but **inconsistent** for c_i under fixed T and large N. But consistent as $T \rightarrow \infty$.

(**incidental parameter problem**)

(viii) If T=2, $y_{it} - y_{i.} \equiv y_{i2} - y_{i1}$ (eq. in first differences)

2.2. Random Effects (RE) Model (Variance Components or Error Components)

$y_{it} = \mathbf{x}_{it}\,\beta + c_i + u_{it}$ (1)

Assumptions

RE1a. $E(u_{it} \mid \mathbf{x}_{i1},\ldots,\mathbf{x}_{iT}, c_i) = 0$ [had before – st exog]

RE1b. $E(c_i \mid \mathbf{x}_{i1},\ldots,\mathbf{x}_{iT}) = E(c_i) = 0$ **Important assumption here.** (orthogonality)

Write (1) as

$y_{it} = \mathbf{x}_{it}\,\beta + v_{it}$ (5)

Assumptions

RE2: rank[E($\mathbf{X}_i'\Omega^{-1}\mathbf{X}_i$)] = K [**OR** rank[E($\mathbf{X}_i'$ **[plim $\hat{\Omega}^{-1}$]** $\mathbf{X}_i$)] = K]

RE3: (a) E($u_i u_i' | \mathbf{x}_i, c_i) = \sigma^2_u \mathbf{I}_T$; (b) E($c_i^2 | \mathbf{x}_i) = \sigma^2_c$

$$E(\mathbf{v}_i \mathbf{v}_i') \equiv \mathbf{\Omega} = \begin{bmatrix} \sigma_c^2 + \sigma_u^2 & \sigma_c^2 & \cdots & \sigma_c^2 \\ \sigma_c^2 & \sigma_c^2 + \sigma_u^2 & \sigma_c^2 \cdots & \sigma_c^2 \\ . & & & . \\ \sigma_c^2 & & & \sigma_c^2 + \sigma_u^2 \end{bmatrix} = \sigma_u^2 \mathbf{I_T} + \sigma_c^2 \mathbf{J_T} \mathbf{J_T}'$$

$\mathbf{J_T}$ is a T x 1 vector of ones. (6)

Estimation

Basic principle – use feasible GLS.

NOTES

(i) Under the assumptions, FGLS is consistent and efficient. The general form is

$$\hat{\boldsymbol{\beta}}_{\mathrm{RE}} = \left(\sum_i \mathbf{X}_i' \hat{\boldsymbol{\Omega}}^{-1} \mathbf{X}_i \right)^{-1} \left(\sum_i \mathbf{X}_i' \hat{\boldsymbol{\Omega}}^{\mathbf{-1}} \mathbf{y}_i \right) \qquad (7)$$

Consistency does not rely on $E(\mathbf{v}_i\ \mathbf{v}_i')=\boldsymbol{\Omega}$. If the prob limit of $\hat{\boldsymbol{\Omega}}$ is not the same as $E(\mathbf{v}_i\ \mathbf{v}_i')$ does not matter. GLS is still consistent.

$$A\hat{\mathrm{v}}\mathrm{ar}(\hat{\boldsymbol{\beta}}_{RE}) = \left(\sum_i \mathbf{X}_i' \hat{\boldsymbol{\Omega}}^{-1} \mathbf{X}_i \right)^{-1}$$

Efficiency requires the assumption that the unconditional variance is equal to the conditional variance and the error structure is used in the estimation.

(ii) $Covar(v_{it},v_{is}) = Covar(c_i+u_{it},c_i+u_{is}) = Var(c_i) = \sigma^2_c \neq 0$ AND

does not go to 0 as N→ ∞ or as T→ ∞.

$Corr(v_{it},v_{is}) = \sigma^2_c / (\sigma^2_c + \sigma^2_u)$; **interpretation?**

All observed correlation between two periods is due to heterogeneity.

(iii) For FGLS we require $\hat{\sigma}_c^2$ and $\hat{\sigma}_u^2$.

(iv) **Estimation of σ_c^2 and σ_u^2.**

Take (5) $$y_{it} = \mathbf{x}_{it}\,\boldsymbol{\beta} + v_{it} \tag{5}$$

OLS consistent but not efficient under our assumptions. Use the OLS residuals to get a consistent estimator of σ^2_v .

$$\hat{\sigma}_v^2 = \frac{SSR_{(pooled)}}{NT-K} \tag{8}$$

Can still use WG. The SSR from this can be used to estimate σ_u^2 as

$$\hat{\sigma}_u^2 = \frac{SSR_{WG}}{(NT-N-K)} \tag{9}$$

Hence $\hat{\sigma}_c^2 = \hat{\sigma}_v^2 - \hat{\sigma}_u^2$ (10)

<u>PROBLEM</u>: Might get a negative value for $\hat{\sigma}_c^2$. Set $\hat{\sigma}_c^2$=0 and proceed. Your model is probably **misspecified**.

There are other ways to estimate these error components.

v)Consider $y_{it} = \mathbf{x}_{it}\boldsymbol{\beta} + c_i + u_{it}$ (11)

Ignore the randomness of c. Estimate (11) by OLS [**pooled model**]

$$\hat{\boldsymbol{\beta}}_p = \left[\sum_i (\mathbf{X}_i'\mathbf{X}_i)\right]^{-1}\left[\sum_i (\mathbf{X}_i'\mathbf{y}_i)\right] = \mathbf{T}_{xx}^{-1}\,\mathbf{t}_{xy}$$

(using NT obs. df=NT-K in general). T_{xx} is the total sum of squares

Now consider $y_{i.} = \mathbf{X}_{i.}\boldsymbol{\beta} + c_i + u_{i.}$ (7)

OLS on (7) gives (called **Between Group (BG) estimator**);

$$\hat{\boldsymbol{\beta}}_b = \left[T\sum_{i}\left(\mathbf{X}_{i.}'\mathbf{X}_{i.}\right)\right]^{-1}\left[T\sum_{i}\left(\mathbf{X}_{i.}'\mathbf{y}_{i.}\right)\right] = \mathbf{B}_{xx}^{-1}\mathbf{b}_{xy} \qquad \text{(df=N-K)}$$

$\mathbf{B}_{xx}$ the between group sum of squares.

$$\hat{\boldsymbol{\beta}}_w = \left(\sum_{i}\tilde{\mathbf{X}}_i'\tilde{\mathbf{X}}_i\right)^{-1}\left(\sum_{i}\tilde{\mathbf{X}}_i'\tilde{\mathbf{y}}_i\right) = \mathbf{W}_{xx}^{-1}\,\mathbf{w}_{xy} \qquad \text{(NT obs. df=NT-N-K).}$$

$\mathbf{W}_{xx}$ is the within group sum of squares

You can show

$$\hat{\boldsymbol{\beta}}_{\mathbf{p}} = \left[\left(\mathbf{W}_{\mathbf{xx}} + \mathbf{B}_{\mathbf{xx}} \right)^{\mathbf{-1}} \mathbf{W}_{\mathbf{xx}} \right] \hat{\boldsymbol{\beta}}_{\mathbf{w}} + \left[\left(\mathbf{W}_{\mathbf{xx}} + \mathbf{B}_{\mathbf{xx}} \right)^{\mathbf{-1}} \mathbf{B}_{\mathbf{xx}} \right] \hat{\boldsymbol{\beta}}_{\mathbf{b}}$$

(weights add up to 1)

Therefore, $\hat{\boldsymbol{\beta}}_p$ is a weighted average of $\hat{\boldsymbol{\beta}}_w$ and $\hat{\boldsymbol{\beta}}_{\mathrm{b}}$.

vii)Similarly we can show that the $\hat{\boldsymbol{\beta}}_{gls}$ is a weighted average of $\hat{\boldsymbol{\beta}}_w$ and $\hat{\boldsymbol{\beta}}_{\mathrm{b}}$ but with different set of weights.

$$\hat{\boldsymbol{\beta}}_{gls} = \left[\left(\mathbf{W}_{\mathbf{xx}} + \theta \mathbf{B}_{\mathbf{xx}} \right)^{\mathbf{-1}} \mathbf{W}_{\mathbf{xx}} \right] \hat{\boldsymbol{\beta}}_{\mathbf{w}} + \left[\left(\mathbf{W}_{\mathbf{xx}} + \theta \mathbf{B}_{\mathbf{xx}} \right)^{\mathbf{-1}} \theta \mathbf{B}_{\mathbf{xx}} \right] \hat{\boldsymbol{\beta}}_{\mathbf{b}} = \kappa \hat{\boldsymbol{\beta}}_{\mathbf{w}} + (1 - \kappa) \hat{\boldsymbol{\beta}}_{\mathbf{b}}$$

where, $\theta = \dfrac{\sigma_u^2}{\sigma_u^2 + T\sigma_c^2}$.

viii)When θ=1 (when $\sigma^2{}_c = 0$), $\hat{\boldsymbol{\beta}}_{gls} = \hat{\boldsymbol{\beta}}_p$. i.e. classical regression model under this assumption $\sigma^2{}_c = 0$, OLS is BLUE.

ix)When θ=0 (when T$\rightarrow \infty$ or when $\sigma^2{}_u = 0$), $\hat{\boldsymbol{\beta}}_{gls} = \hat{\boldsymbol{\beta}}_W$.

When T is large, we get more efficient estimators by using the within group variations. Since the BG estimator ignores this variation, we set the weight attached to this to zero.

x) **Feasible GLS - 2 step estimation**

θ is unknown. Estimate θ. Use a two-step method.

Step 1: Est the variance components which will let you estimate θ.

Step 2: Transform the variables and regress y_{it}^* on x_{it}^* s.

$$y_{it}^* = y_{it} - (1-\sqrt{\theta})\, y_{i.} \text{ and } x_{it}^* = x_{it} - (1-\sqrt{\theta})\, x_{i.} \quad (12)$$

2-step GLS is more efficient than the WG estimator even for moderate sample sizes such as T>= 3 when N-K >= 9 or T=2 when N-K >= 10.

Can continue with the iterations until convergence. But this is not necessary.

Estimation of σ_u^2 and σ_α^2 - see above………

xi) Maximum Likelihood Estimation: Under the assumption of Normally distributed errors, with $N \rightarrow \infty$ and fixed T, $\hat{\boldsymbol{\beta}}_{MLE}$ is consistent. But when N is fixed and $T \rightarrow \infty$, $\hat{\boldsymbol{\beta}}_{MLE} \rightarrow \hat{\boldsymbol{\beta}}_w$ but $\hat{\sigma}_c^2$ will be inconsistent since when N is fixed, there is not enough variation in the c_i s to give consistent parameter estimator however large the T is!

1.3 First Differencing vs WG

Another way to eliminate c_i is to take first differences.

Estimation equation: $\Delta y_{it} = \Delta \mathbf{x}_{it} \boldsymbol{\beta} + \Delta u_{it}$ (13)

- lose one observation at the beginning
- can't estimate coefficients on time-invariant variables
- need strict exog assumption in terms of differences

  i.e. $E(\Delta u_{it} | \Delta \mathbf{x}_{i2}, \ldots, \Delta \mathbf{x}_{iT})=0$ **FD1**

- $\text{rank}(\sum_{t=2}^{T} E(\Delta \mathbf{x}_{it}' \Delta \mathbf{x}_{it}))=K$ **FD2**

- $E(\Delta u_{it}\ \Delta u_{it}' \mid \mathbf{x_{i1}, \ldots, x_{iT}}, c_i) = \sigma^2_e\, \mathbf{I_{T-1}}$ where $e_{it} = \Delta u_{it}$

  No serial correlation implies u is a random walk!

- $A\hat{v}ar(\hat{\boldsymbol{\beta}}_{FD}) = \hat{\sigma}_e^2 (\Delta \mathbf{X}'\Delta \mathbf{X})^{-1}$ **with** $\hat{\sigma}_e^2 = SRR_{FD}/[N(T-1)-K]$

- A powerful method used in program evaluation

  Let T=2 $\Delta y_{i2} = \alpha_2 + \Delta \mathbf{z}_{i2}\, \boldsymbol{\beta} + \delta\, \mathbf{prog_{i2}} + \Delta u_{i2}$

  when $\boldsymbol{\beta}=0$ estimate of $\delta = \Delta\bar{y}_{treated} - \Delta\bar{y}_{control}$ **(DID estimator)**

- **WG** more efficient when u_{it} are serially uncorrelated and **FD** more efficient when u_{it} is a random walk. Very different est can imply failure of strict exog assumption.

2.4 Summary on Estimators

$$y_{it} = \mathbf{x}_{it}\boldsymbol{\beta} + c_i + u_{it} \qquad (1)$$

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + c_i\mathbf{J}_T + \mathbf{u}_i$$

Estimating equation: $\widehat{y}_{it} = \widehat{\mathbf{w}}_{it}\boldsymbol{\gamma} + \widehat{u}_{it}$

All equations estimated by OLS

Pooled: $\widehat{y}_{it} = y_{it}$; Same for the **w**.

- No transformation needed;
- All variables in the model.

WG: $\widehat{y}_{it} = y_{it} - y_{i.}$

- Variables in mean deviations
- Only time varying variables are in the equation

GLS $\widehat{y}_{it} = y_{it} - (1 - \sqrt{\hat{\theta}})\, y_{i.}$ $\left(\theta = \dfrac{\sigma_u^2}{\sigma_u^2 + T\sigma_c^2}\right)$

- Variables in special deviations
- All variables in the equation

<u>FD</u> $\overset{\frown}{y}_{it} = y_{it} - y_{it-1}$

- Variables in first differences
- Only time varying variables are in the equation
- Also lose the first observations. So have only T-1 obs per i.

Now, write $\overset{\frown}{y}_{it} = \overset{\frown}{\mathbf{w}}_{it}\boldsymbol{\gamma} + \overset{\frown}{u}_{it}$ as

$$\overset{\frown}{\mathbf{y}}_i = \overset{\frown}{\mathbf{W}}_i\boldsymbol{\gamma} + \overset{\frown}{\mathbf{u}}_i$$

OLS est $\hat{\boldsymbol{\gamma}} = (\sum_i \overset{\frown}{\mathbf{W}}_i{}'\overset{\frown}{\mathbf{W}}_i)^{-1}(\sum_i \overset{\frown}{\mathbf{W}}_i{}'\overset{\frown}{\mathbf{y}}_i)$

Asy Covar matrix $\hat{\mathbf{V}} = \hat{\sigma}_{\hat{u}}^2 (\sum_i \widehat{\mathbf{W}}_i ' \widehat{\mathbf{W}}_i)^{-1}$

Or in the general case:

$$\hat{\mathbf{V}} = (\sum_i \widehat{\mathbf{W}}_i ' \widehat{\mathbf{W}}_i)^{-1} \left[\sum_i \widehat{\mathbf{W}}_i ' \, \mathrm{E}(\widehat{\mathbf{u}}_i \widehat{\mathbf{u}}_i ' \,|\, \widehat{\mathbf{W}}_i) \widehat{\mathbf{W}}_i \right] (\sum_i \widehat{\mathbf{W}}_i ' \widehat{\mathbf{W}}_i)^{-1}$$

2.5 <u>Robust Covariances</u>

- Issue: Robust covariances vs GLS
- What do we want to use for $\left[\sum_i \widehat{\mathbf{W}}_i ' \, \mathrm{E}(\widehat{\mathbf{u}}_i \widehat{\mathbf{u}}_i ' \,|\, \widehat{\mathbf{W}}_i) \widehat{\mathbf{W}}_i \right]$?

- Use $\left[\sum_i \widehat{\mathbf{W}}_i{}' \, (\hat{\hat{\mathbf{u}}}_{\mathrm{i}} \hat{\hat{\mathbf{u}}}_{\mathrm{i}}{}') \widehat{\mathbf{W}}_i \right] = \sum_i \sum_t \sum_s \mathbf{w}_{it}^{'} \mathbf{w}_{is} \hat{\hat{\mathbf{u}}}_{\mathrm{it}} \hat{\hat{\mathbf{u}}}_{\mathrm{is}}$

- allows for arbitrary serial correlation and time-varying variances

- but note, the above is calculated under large N and fixed T asymptotics as well as independence over i

- can be calculated using cluster-robust command using clustering at the individual level.

Also note

- **<u>GLS (RE)</u>**: $A\hat{\text{v}}\text{ar}(\hat{\boldsymbol{\beta}}_{gls}) = \left(\sum_i \mathbf{X}_i' \hat{\boldsymbol{\Omega}}^{-1} \mathbf{X}_i\right)^{-1}$ ordinary case

  Here: $\left(\sum_i \mathbf{X}_i' \hat{\boldsymbol{\Omega}}^{-1} \mathbf{X}_i\right)^{-1} \left(\sum_i \mathbf{X}_i' \hat{\boldsymbol{\Omega}}^{-1} \hat{\mathbf{u}}_i \hat{\mathbf{u}}_i' \hat{\boldsymbol{\Omega}}^{-1} \mathbf{X}_i\right) \left(\sum_i \mathbf{X}_i' \hat{\boldsymbol{\Omega}}^{-1} \mathbf{X}_i\right)^{-1}$

- Can't use residual sums of squares to test restrictions re coefficients.

  Need to use the general Wald test: H_0: $\mathbf{R}\boldsymbol{\beta}=\mathbf{r}$

  $$\text{W} = \left(\mathbf{R}\hat{\boldsymbol{\beta}} - r\right)'\left(\mathbf{R}\hat{\mathbf{V}}\mathbf{R}'\right)^{-1}\left(\mathbf{R}\hat{\boldsymbol{\beta}} - r\right) \sim asy \;\; \chi^2(q)$$

- For large T and small N, you will have to specify a particular form of correlation such as ARMA for example.

2.6 Other considerations

Mundlak (1978) E'trica

Approximate $c_i = \mathbf{x}_{i.}\boldsymbol{\pi} + v_i$ $\quad v_i \sim$ iid

Including this in the equation and estimating by **GLS** gives you the same WG estimators for β.

i.e. OLS estimation of (see (12))

$$y_{it}^{*} = \mathbf{x}_{it}^{*}\,\boldsymbol{\beta} + \mathbf{x}_{i.}\boldsymbol{\pi} + \text{error}$$

$$\hat{\boldsymbol{\beta}}_{gls} = \hat{\boldsymbol{\beta}}_{w} \quad \text{and} \quad \hat{\boldsymbol{\pi}} = \hat{\boldsymbol{\beta}}_{b} - \hat{\boldsymbol{\beta}}_{w}$$

(what does this mean? BLUE of the RE model when the correlation is allowed for gives you the WG est.)

Can test for $\pi=0$ **Wu-Hausman test for strict exog.**

<u>Chamberlain (1982, 1984)</u>

$c_i = \mathbf{x_{1i}}\,\boldsymbol{\pi_1} + \ldots\ldots + \mathbf{x_{Ti}} + v_i$ $v_i \sim$ iid (more general than Mundlak)

Use Minimum Distance methods. This gives WG for $\boldsymbol{\beta}$.

2.7 Tests

2.7.1 Wu-Hausman Test

This is a test for strict exogeneity.

First consider the general test.

Consider two estimators $\hat{\boldsymbol{\beta}}_1$ and $\hat{\boldsymbol{\beta}}_2$ say, such that

Under H_0	Under H_1
$\hat{\boldsymbol{\beta}}_1$ is consistent and efficient	$\hat{\boldsymbol{\beta}}_1$ is inconsistent
$\hat{\boldsymbol{\beta}}_2$ is consistent	$\hat{\boldsymbol{\beta}}_2$ is consistent

Let $\hat{\mathbf{q}} = \hat{\boldsymbol{\beta}}_2 - \hat{\boldsymbol{\beta}}_1$

then, $m = \hat{\mathbf{q}}' \left[\mathbf{V\hat{a}r\hat{q}} \right]^{-1} \hat{\mathbf{q}}$ ~ asy $\chi^2_{(K)}$ on H_0. (K restrictions)

It is easily shown that ***Var*** $[\hat{\mathbf{q}}]$ **=** ***Var***$[\hat{\boldsymbol{\beta}}_2]$ **-** ***Var***$[\hat{\boldsymbol{\beta}}_1]$**.**

Var $[\hat{\mathbf{q}}]$ is positive definite (i.e will have an inverse which is needed) but the estimated ***Var*** $[\hat{\mathbf{q}}]$ need not be!

This is a problem with this test. But there are regression based versions.

Example 1 – **WG vs GLS**

Maintained Hypotheses:

RE1a. $E(u_{it} \mid \mathbf{x}_{i1},\ldots,\mathbf{x}_{iT}, c_i) = 0$ [st exog]

RE3: (a) $E(u_i u_i' \mid \mathbf{x}_i, c_i)=\sigma^2_u \mathbf{I}_T$; (b) $E(c_i^2 \mid \mathbf{x}_i)=\sigma^2_c$

NOTE: RE3 gives GLS more efficient than WG

Want to test: RE1b. $E(c_i \mid \mathbf{x}_{i1},\ldots,\mathbf{x}_{iT}) = E(c_i)$

For the test $\hat{\boldsymbol{\beta}}_2$ is our $\hat{\boldsymbol{\beta}}_w$ and $\hat{\boldsymbol{\beta}}_1$ is our $\hat{\boldsymbol{\beta}}_{RE}$.

Notes:

- The null can fail for many reasons such as simultaneity, measurement errors, omitted variables etc.
- When $T \rightarrow \infty$ and N is fixed, $\hat{\beta}_{GLS} \cong \hat{\beta}_{w}$ and hence the test will have low power.
- Regression based test (no problem with –ve variance):

$$y_{it}^{*} = y_{it} - (1-\sqrt{\theta})\, y_{i.} \quad \text{and} \quad \mathbf{x}_{it}^{*} = \mathbf{x}_{it} - (1-\sqrt{\theta}\,)\, \mathbf{x}_{i.}$$

$$y_{it}^{*} = \ldots\ldots\ldots + (\mathbf{x}_{it} - \mathbf{x}_{i.})\, \delta + \text{error} \quad \text{and test for } H_0\text{: } \boldsymbol{\delta = 0}$$

Can have individual specific variables in the RE part (assuming that they are not correlated with the unobs heterogeneity) and include the time varying covariates in mean deviations as extra variables and do the same test.

It can be shown that in the above regression,

$\hat{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}}_b$ (between group) and $\hat{\boldsymbol{\delta}} = \hat{\boldsymbol{\beta}}_{WG} - \hat{\boldsymbol{\beta}}_b$

Since $\hat{\boldsymbol{\beta}}_{gls} = \kappa\hat{\boldsymbol{\beta}}_{\mathbf{w}} + (1-\kappa)\hat{\boldsymbol{\beta}}_{\mathbf{b}}$ we can do the same test by comparing different estimators as follows:

$$\hat{\boldsymbol{\beta}}_{gls} - \hat{\boldsymbol{\beta}}_{wg} \quad ; \quad \hat{\boldsymbol{\beta}}_{gls} - \hat{\boldsymbol{\beta}}_{b} \; ; \qquad \hat{\boldsymbol{\beta}}_{wg} - \hat{\boldsymbol{\beta}}_{b} \quad ; \quad \hat{\boldsymbol{\beta}}_{gls} - \hat{\boldsymbol{\beta}}_{p}$$

All estimators are consistent under the null and therefore should converge under the null.

If there is heteroskedasticity and serial correlation, then WG or GLS are not optimal under H_0 or H_1 and we cannot rank these in terms of asymptotic efficiency. Can use White's robust covar – Arellano.

2.7.2 Test for Random Effects

Breusch & Pagan Test (1980 R E Studies)

H_0: $\sigma_c^2 = 0$ in the RE Model. **LM test.**

Then,
$$m = \frac{NT}{2(T-1)} \left[\frac{\sum_i \left(\sum_t \hat{u}_{it} \right)^2}{\sum_i \sum_i \hat{u}_{it}^2} - 1 \right]^2 \text{ asy} \sim \chi^2(1) \text{ on } H_0.$$

where, $\hat{u}_{it}$= residuals from the pooled model regression (OLS). This is so since under H_0 the RE model collapses to the pooled model.

Unfortunately this does not account for the fact that $\sigma_c^2 > 0$ under $H_{1.}$

Improved one sided test:

$$\text{HONDA}=\sqrt{\frac{\text{NT}}{2(\text{T}-1)}}\left[\frac{\sum_i\left(\sum_t \hat{u}_{it}\right)^2}{\sum_i\sum_i \hat{u}_{it}^2} - 1\right] \text{asy}\sim N(0,1) \text{ on } \text{H}_0$$

Generally has low power. Better to use the CHOW test for this…WG residuals vs pooled OLS residuals.

2.8 Model Specification and Estimation – extension

RE Model

Some regressors correlated with the c_i.

Important model: For eg. in earnings equations.

Consider $y_{it} = \boldsymbol{\beta}'\mathbf{x}_{it} + \boldsymbol{\gamma}'\mathbf{z}_i + c_i + u_{it}$

Remember! No correlation between the regressors and c_i

The efficient estimation is the GLS.

Some time varying regressors correlated with the c_i

Can do WG estimation but this will not let us estimate γ.

Use a two-step method as follows:

<u>Step 1</u>: Get the $\hat{\boldsymbol{\beta}}_w$.

<u>Step 2:</u> Estimate the γ by doing an OLS of

$$y_{i\cdot} - \hat{\boldsymbol{\beta}}_w \; x_{i\cdot} = \gamma \, z_i + \text{error}.$$

As $N \to \infty$ $\hat{\beta}_w$ and $\hat{\gamma}$ are consistent.

But for a fixed N and $T \to \infty$, $\hat{\beta}_w$ is consistent but not $\hat{\gamma}$ - because equation uses N observations.

Some x_{it} and some z_i are correlated with c_i (Hausman & Taylor 1981 Econometrica)

Write equation as $y_{it} = \beta_1'\mathbf{x_{1it}} + \beta_2'\mathbf{x_{2it}} + \gamma_1'\mathbf{z_{1i}} + \gamma_2'\mathbf{z_{2i}} + c_i + u_{it}$ ()

Assume there are

- k_1 variables in $\mathbf{x_1}$; k_2 variables in $\mathbf{x_2}$;
- g_1 variables in $\mathbf{z_1}$; g_2 variables in $\mathbf{z_2}$.
- $\mathbf{x_1}$ and $\mathbf{z_1}$ are uncorrelated with the c_i and $\mathbf{x_2}$ and $\mathbf{z_2}$ are correlated with the c_i.
- $E(u_{it} \mid \mathbf{x_{i1},..,x_{iT}}, c_i)=0$ [strict exog conditional on c]

GLS inconsistent.

Use instrumental variable estimation (IV) along with GLS

Need instruments.

Step 1: Obtain $\hat{\boldsymbol{\beta}}_{\mathbf{1w}}$ and $\hat{\boldsymbol{\beta}}_{\mathbf{2w}}$

Step 2: Write equation as $y_{i\cdot} - \hat{\boldsymbol{\beta}}_{\mathbf{w}} \; \mathbf{x}_{\mathbf{i\cdot}} = \gamma_1 \, \mathbf{z}_{\mathbf{1i}} + \gamma_1 \, \mathbf{z}_{\mathbf{2i}}$ +error.

Use IV to estimate the γ coefficients (OLS inconsistent).

Step 3: Obtain the variance components as before

$\hat{\sigma}_u^2$ and $\hat{\sigma}_u^2 + T\hat{\sigma}_c^2$ (assumes RE spec correct!)

Step 4: Use the step 3 ests to transform the variables to y_{it}* etc.

Step 5: apply IV - HT showed that $\mathbf{x}_{\mathbf{1i\cdot}}$ are valid instruments for $\mathbf{z}_{\mathbf{2i\cdot}}$

Must have $\mathbf{k_1} \geq \mathbf{g_2}$.

One big advantage: instruments are from within the model…….

STATA – use *xthtaylor*

STATA – *xtreg* with the RE option

STATA – *xtreg* with the FE option

Test for over-identification

$\hat{\mathbf{q}} = \hat{\boldsymbol{\beta}}_{HT} - \hat{\boldsymbol{\beta}}_{w}$; then $\hat{\sigma}_u^2\ [\hat{\mathbf{q}}'[\hat{\mathrm{var}}(\hat{\boldsymbol{\beta}}_w) - \hat{\mathrm{var}}(\hat{\boldsymbol{\beta}}_{HT})]^{-1}\hat{\mathbf{q}} \sim asy\ \chi^2(k)$

where $k=\min[k_1-g_2, NT-k_1-k_2]$

Notes: Found to be sensitive to the assumptions of exogeneity of the different sets of x and z variables. Chowdhury & Nickell, 1985, Journal of Labor Economics is a very good example of this.