

Core Econometrics - OLS

Table of Contents

1. [MT Part II: OLS](#)
 1. [2.1 OLS](#)
 2. [2.2 Partialling Out](#)
 3. [N 2.4 Finite Sample Inference](#)
 4. [N 2.5 Finite Sample Inference](#)
 1. [Derive Finite-sample Distribution of OLS Estimator](#)
 2. [Testing Linear Restrictions](#)
 5. [N 2.6 Large Sample Properties](#)
 6. [N 2.7 Large Sample Inference](#)
 7. [N 2.8 Heteroskedasticity](#)
 8. [N 2.9 Cluster Robust Inference](#)

MT Part II: OLS

2.1 OLS

[OLS Estimator](#) #flashcard

$$\hat{\beta}_{OLS} = \underbrace{(X^T X)^{-1} X^T y}_{A_{k \times n}}$$

where $A_{k \times n}$ is known as the pseudoinverse of X

[Perfect Multicollinearity](#) #flashcard

$\text{rank}(X) < k \implies (X^T X)^{-1}$ does not exist

[Classical Linear Regression Model with Stochastic Regressors](#) #flashcard

$$\begin{cases} y = X\beta + u \\ \mathbb{E}[u|X] = 0 \\ \text{Var}[u|X] = \sigma^2 I \\ X \text{ has full rank (with probability 1)} \end{cases} \quad \begin{array}{l} \text{linear in parameter} \\ \text{zero conditional mean} \\ \text{conditional homoskedasticity} \\ \text{no perfect multicollinearity} \end{array}$$

[\(Conditional\) Gauss-Markov Theorem](#) #flashcard

In classical linear regression model:

$$\text{Var}[\hat{\beta}_{OLS}|X] \leq \text{Var}[\tilde{\beta}|X]$$

for any other unbiased estimator $\tilde{\beta}$ that is a linear function of the random vector y conditional on X (BLUE).

[Projection Matrix](#) #flashcard

The $n \times n$ matrix:

$$\hat{y} = \underbrace{X(X^T X)^{-1} X^T y}_{P_{n \times n}}$$

with property:

$$PX = X$$

and symmetric and idempotent:

$$P^T = P, P^2 = P$$

[Annihilator Matrix](#) #flashcard

The $n \times n$ matrix:

$$\begin{aligned} \hat{u} &= (I - X(X^T X)^{-1} X^T) y \\ &= \underbrace{(I - P)y}_{M_{n \times n}} \end{aligned}$$

with the property:

$$MX = 0$$

Normal Equations #flashcard

$$X^T \hat{\mu} = 0$$

Estimator of Standard Error #flashcard

$$\hat{\sigma}_{OLS}^2 = \frac{\hat{u}^T \hat{u}}{n - K} = \frac{\sum_{i=1}^n \hat{u}_i^2}{n - K}$$

2 2 Partialling Out

General Partialling Out Algorithm (Frisch-Waugh-Lovell Theorem) #flashcard

1. Regress y on all regressors except x_k
2. Regress x_k on all regressors except x_k
3. Regress the residual from 1 on 2

Partial R^2 #flashcard

The R^2 in the final step of partialling out is an estimator for the **partial correlation** between y and x_k ($\rho_{Y,X_k| \{X_i\}_{i \neq k}}$)

N 2 4 Finite Sample Inference

Properties of OLS under CRM Assumptions #flashcard

Given the 4 assumptions of CRM, we have the following theorems:

- **Conditional/Unconditional Unbiasedness + Consistency:**

$$\mathbb{E} [\hat{\beta}|X] = \beta, \mathbb{E} [\hat{\beta}] = \beta$$

- **Gauss-Markov Theorem:**

$$Var[\tilde{\beta}|X] - Var[\hat{\beta}|X] \text{ is PSD } \forall \tilde{\beta} \text{ that is unbiased}$$

N 2 5 Finite Sample Inference

Inference in CRM with Normal Errors #flashcard

- Assumptions:

$$\begin{cases} y &= X\beta + u \\ u|X &\sim N(0, \sigma^2 I) \\ X &\text{has full rank with probability 1} \end{cases}$$

- Result:

$$\begin{cases} \hat{\beta}|X &\sim N(\beta, \sigma^2(X^T X)^{-1}) \\ t_k = \frac{\hat{\beta}_k - \beta_k}{se_k} &\sim t_{n-K} \end{cases}$$

where se_k is the (k, k) entry of $\hat{\sigma}^2 I = \frac{\hat{u}^T \hat{u}}{n - K} I$

Derive Finite-sample Distribution of OLS Estimator

Summing up Standard Normal Variables Yields χ^2 #flashcard

Let $z \sim N(0, I)$ be a standard normal vector with independent elements, then:

$$w = z^T z = \sum_{i=1}^n z_i^2 \sim \chi_n^2$$

Combine 2 χ^2 RVs $\rightsquigarrow F$ #flashcard

Let 2 random scalar RVs: $w_1 \sim \chi_n^2, w_2 \sim \chi_m^2$ be independent of each other, then:

$$v = \frac{\frac{w_1}{w_2}}{m} \sim F_{n,m}$$

Combine N and $\chi^2 \rightsquigarrow$ Student t-distribution #flashcard

Let 2 scalar random variables $z \sim N(0, 1)$ and $w \sim \chi_n^2$ to be independent. Then:

$$\frac{z}{\sqrt{\frac{w}{n}}} \sim t_n$$

2 Ways to Convert a Normal Vector into a χ^2 Scalar #flashcard

- Both ways convert a Normal vector into a χ^2 scalar:
 - If $y \sim N(\mu, \Sigma)$ is an $n \times 1$ vector, then the scalar:

$$w = (y - \mu)^T \Sigma^{-1} (y - \mu) \sim \chi_n^2$$

- If $z \sim N(0, 1)$ is an $n \times 1$ vector and G is an $n \times n$ non-stochastic, symmetric, and idempotent matrix with $\text{rank}(G) = r < n$, then the scalar:

$$w = z^T G z \sim \chi_r^2$$

Show $\hat{\beta}$ Follows a t-Distribution (I typed this throughout only for fun -- probably not the most important derivation)

- Setup:

$$\begin{cases} y &= X\beta + u \\ u|X &\sim N(0, \sigma^2 I) \\ X &\text{has full rank} \end{cases}$$

- beware of the additional Normality assumption (required for inference in finite samples)

- Objective (want to show that):

$$\underbrace{t_k}_{\text{our test stat}} \sim t_{n-k} \text{ (t distribution with DoF)=n-k}$$

- Main idea: manipulate our expression so that we can use the definition of t-distribution:

Combine N and $\chi^2 \rightsquigarrow$ Student t-distribution #flashcard

Let 2 scalar random variables $z \sim N(0, 1)$ and $w \sim \chi_n^2$ to be independent. Then:

$$\frac{z}{\sqrt{\frac{w}{n}}} \sim t_n$$

- Step 1: Derive an basic expression for test statistic:

$$\begin{aligned} t_k &= \frac{\hat{\beta}_k - \beta_k}{\sqrt{\hat{v}_{kk}}} \\ &= \frac{\hat{\beta}_k - \beta_k}{\sqrt{v_{kk}}} \cdot \frac{\sqrt{v_{kk}}}{\sqrt{\hat{v}_{kk}}} \\ &= \frac{\frac{\hat{\beta}_k - \beta_k}{\sqrt{v_{kk}}}}{\frac{\sqrt{\hat{v}_{kk}}}{\sqrt{v_{kk}}}} \end{aligned}$$

- v_{kk} is the variance of $\hat{\beta}$ and \hat{v}_{kk} is its estimator. They are the (k, k) element of the var-cov matrix $\sigma^2 I$ and $\hat{\sigma}^2 I$

- Step 2: Derive the distribution of numerator $z_k = \frac{\hat{\beta}_k - \beta_k}{\sqrt{v_{kk}}} \sim N(0, 1)$:

- Show $\hat{\beta}_k \sim N(\beta_k, v_{kk})$

$$\begin{aligned} \hat{\beta}_k &= \beta + (X^T X)^{-1} X^T u \text{ and } u \sim N(0, \sigma^2 I) \\ \implies \hat{\beta}_k &\sim N\left(\beta, (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \sigma^2 ((\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T)^T\right) \\ &\sim N\left(\beta, \sigma^2 (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T ((\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T)^T\right) \\ &\sim N\left(\beta, \sigma^2 (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1}\right) \\ &\sim N\left(\beta, \sigma^2 (\mathbf{X}^T \mathbf{X})^{-1}\right) \end{aligned}$$

- Derive the distribution for z_k :

$$v_{kk} = \text{Var}(\hat{\beta}_k) = (\text{k,k}) \text{ element of } \sigma^2(X^T X)^{-1}$$

$$\implies z_k = \frac{\hat{\beta}_k - \beta_k}{\sqrt{v_{kk}}} \sim N(0, 1) \quad (\text{Standardisation})$$

- Step 3: Derive the distribution of the denominator $\frac{\sqrt{\hat{v}_{kk}}}{\sqrt{v_{kk}}}$:

- Get a more comfortable expression:
- As we defined before: v_{kk} and \hat{v}_{kk} are the (k, k) element of var-cov matrix $\sigma^2 I$ and $\hat{\sigma}^2 I$ corresponding
- \implies Only difference between those 2 matrices is: $\sigma^2 I$ has σ^2 as its diagonal elements while $\hat{\sigma}^2 I$ has $\hat{\sigma}^2$ as its diagonal elements
- \implies

$$\begin{aligned} \frac{\sqrt{\hat{v}_{kk}}}{\sqrt{v_{kk}}} &= \sqrt{\frac{\hat{v}_{kk}}{v_{kk}}} \\ &= \sqrt{\frac{\hat{\sigma}^2}{\sigma^2}} \\ &= \sqrt{\frac{(n - K)\hat{\sigma}^2}{n - K}} \end{aligned} \quad (\text{argued above})$$

- Show $(n - K)\frac{\hat{\sigma}^2}{\sigma^2} \sim \chi^2_{n-K}$:

- Start from the expression of \hat{u} :

$$\begin{aligned} \hat{u} &= y - X(X^T X)^{-1} X^T y \\ &= \underbrace{(I - X(X^T X)^{-1} X^T)}_M y \\ &= M(X\beta + u) \\ &= M\left(X\beta + \sigma \frac{u}{\sigma}\right) \\ &= M\cancel{X}^0 \beta + M\sigma \frac{u}{\sigma} \\ &= M\sigma \underbrace{\frac{u}{\sigma}}_{\equiv z \sim N(0, 1)} \\ &= M\sigma z \end{aligned}$$

- Then, we can derive a more comfortable expression for $(n - K)\frac{\hat{\sigma}^2}{\sigma^2}$:

$$\begin{aligned} (n - K)\frac{\hat{\sigma}^2}{\sigma^2} &= (n - K) \frac{\left(\frac{\hat{u}^T \hat{u}}{n - K}\right)^2}{\sigma^2} \\ &= (n - K) \frac{(M\sigma z)^T (M\sigma z)}{(n - K)^2 \sigma^2} \\ &= \frac{z^T \cancel{M}^T \cancel{M}^M z \sigma^2}{(n - K)\sigma^2} \\ &= \frac{z^T M z}{n - K} \end{aligned}$$

- Recall our method to convert normal variables to χ^2 variables:

- Both ways convert a Normal vector into a χ^2 scalar:
- If $y \sim N(\mu, \Sigma)$ is an $n \times 1$ vector, then the scalar:

$$w = (y - \mu)^T \Sigma^{-1} (y - \mu) \sim \chi_n^2$$

- If $z \sim N(0, 1)$ is an $n \times 1$ vector and G is an $n \times n$ non-stochastic, symmetric, and idempotent matrix with $\text{rank}(G) = r < n$, then the scalar:

$$w = z^T G z \sim \chi_r^2$$

- Apply the above theorem:

$$(n - K)\frac{\hat{\sigma}^2}{\sigma^2} = \frac{z^T M z}{n - K} \sim \chi_{n-K}^2$$

- Show the numerator and denominator are independent:
- Combine our results to get the final conclusion:

Testing Linear Restrictions

3 Elements of a Hypothesis Test #flashcard

1. Null and alternative hypothesis
2. Test statistic and its distribution under the null
3. Decision rule at a significant level α

Simple Significance Test for CRM with Normal Errors #flashcard

1. Null and alternative hypothesis:

$$\begin{aligned} H_0 : \beta_k &= 0 \\ H_1 : \beta_k &\neq 0 \end{aligned}$$

2. Test statistic and its distribution under the null

$$t_k = \frac{\hat{\beta}_k}{se_k} \sim t_{n-K}$$

3. Decision rule at a significant level α : reject H_0 if $|t_k| > c_{\frac{\alpha}{2}}(n - K)$ where $c_{\frac{\alpha}{2}}(n - K)$ is the corresponding critical value

General Linear Hypothesis Test of CRM with Normal Errors (Standard) #flashcard

1. Null and alternative hypothesis:

$$\begin{aligned} H_0 : H\beta &= \theta \\ H_1 : H\beta &\neq \theta \end{aligned}$$

where $H_{p \times K}$ is the restriction matrix

2. Test statistic and its distribution under the null

$$v = \frac{1}{p} w = \frac{1}{p} (H\hat{\beta} - \theta)^T \left[\widehat{Var}(H\hat{\beta}|X) \right]^{-1} (H\hat{\beta} - \theta) \sim F_{p,n-K}$$

3. Decision rule at a significant level α :

Reject H_0 if $v > c_\alpha(n - K)$

where $c_\alpha(n - K)$ is the corresponding critical value

(There is a **proof** of distribution in slides)

General Linear Hypothesis Test of CRM with Normal Errors (Using R-Squared) #flashcard

1. Null and alternative hypothesis:

$$\begin{aligned} H_0 : H\beta &= \theta \\ H_1 : H\beta &\neq \theta \end{aligned}$$

where $H_{p \times K}$ is the restriction matrix

2. Test statistic and its distribution under the null

$$v = \frac{n - K}{p} \times \frac{R_{\text{Unrestricted}}^2 - R_{\text{Restricted}}^2}{1 - R_{\text{Unrestricted}}^2} \sim F_{p,n-K}$$

3. Decision rule at a significant level α :

Reject H_0 if $v > c_\alpha(n - K)$

where $c_\alpha(n - K)$ is the corresponding critical value

- Note that **Global Significance Test** is a special case:

- $H_0 : \beta_2 = \beta_3 = \dots = \beta_K = 0$
- $v = \frac{n - K}{p} \times \frac{R^2}{1 - R^2} \sim F_{K-1, n-K}$

N 2 6 Large Sample Properties

Consistency and Asymptotic Normality Requirement on OLS #flashcard

- $y = XB + u$
- (u_i, x_i^T) are iid with $\mathbb{E}[x_i u_i] = 0 \forall i = 1, \dots, n$
- The $K \times K$ matrix $M_{XX} = \mathbb{E}[x_i x_i^T]$ exists and is non-singular
 - $\hat{\beta} \xrightarrow{p} \beta$
- The $K \times K$ matrix $M_{X\Omega X} = \mathbb{E}[u^2 x_i x_i^T]$ exists and is non-singular
 - $\hat{\beta} \xrightarrow{D} N(0, M_{XX}^{-1} M_{X\Omega X} M_{XX}^{-1})$

Proof of Consistency and Normality of OLS #flashcard

- Requirements:
 - $y = XB + u$
 - (u_i, x_i^T) are iid with $\mathbb{E}[x_i u_i] = 0 \forall i = 1, \dots, n$
 - The $K \times K$ matrix $M_{XX} = \mathbb{E}[x_i x_i^T]$ exists and is non-singular (for consistency)
 - The $K \times K$ matrix $M_{X\Omega X} = \mathbb{E}[u^2 x_i x_i^T]$ exists and is non-singular (for normality)
- Proof of consistency:
 - Decompose y in the expression of $\hat{\beta}$:

$$\hat{\beta} = \beta + (X^T X)^{-1} X^T u$$

- Manipulate formula:

$$\hat{\beta} = \beta + \left(\frac{X^T X}{n} \right)^{-1} \left(\frac{X^T u}{n} \right)$$

- Then, apply LLN and then continuous mapping theorem:

$$\begin{aligned} \hat{\beta} &= \beta + \underbrace{\left(\frac{X^T X}{n} \right)^{-1}}_{\xrightarrow{p} M_{XX}^{-1}} \underbrace{\left(\frac{X^T u}{n} \right)}_{\xrightarrow{p} 0} \\ &\xrightarrow{p} \beta + M_{XX}^{-1} \times 0 \\ &\xrightarrow{p} \beta \end{aligned}$$

- Proof of normality:
 - Decompose y in the expression of $\hat{\beta}$:

$$\hat{\beta} = \beta + (X^T X)^{-1} X^T u$$

- Manipulate formula:

$$\hat{\beta} = \beta + \left(\frac{X^T X}{n} \right)^{-1} \left(\frac{X^T u}{n} \right)$$

- Then apply LLN, CLT, and then Slutsky Theorem:

$$\begin{aligned} \hat{\beta} &= \beta + \underbrace{\left(\frac{X^T X}{n} \right)^{-1}}_{\xrightarrow{p} M_{XX}^{-1}} \underbrace{\left(\frac{X^T u}{n} \right)}_{\xrightarrow{D} N(0, M_{X\Omega X})} \\ &\xrightarrow{D} N(\beta, M_{XX}^{-1} M_{X\Omega X} M_{XX}^{-1}) \end{aligned}$$

N 2 7 Large Sample Inference

OLS Large Sample Inference Under Homoskedasticity #flashcard

- Assumptions
 - $y = XB + u$
 - (u_i, x_i^T) are iid with $\mathbb{E}[x_i u_i] = 0 \forall i = 1, \dots, n$
 - The $K \times K$ matrix $M_{XX} = \mathbb{E}[x_i x_i^T]$ exists and is non-singular
 - The $K \times K$ matrix $M_{X\Omega X} = \mathbb{E}[u^2 x_i x_i^T]$ exists and is non-singular
 - $\hat{\beta} \xrightarrow{D} N(0, M_{XX}^{-1} M_{X\Omega X} M_{XX}^{-1})$
- Homoskedasticity:
 - $M_{X\Omega X} = \mathbb{E}[u^2 x_i x_i^T] = \sigma^2 \mathbb{E}[x_i x_i^T] = \sigma^2 M_{XX}$
 - $\hat{\beta} \xrightarrow{D} N(0, \sigma^2 M_{XX}^{-1})$

- Using consistent estimators:

- $\hat{\sigma}^2 = \frac{1}{n-K} \sum_{i=1}^n \hat{u}_i^2$
- $\hat{M}_{XX} = \frac{1}{n} \sum_{i=1}^n x_i x_i^T = \frac{X^T X}{n}$

- We can get the final result for large-sample inference:

$$\hat{\beta} \sim^a N\left(\beta, \hat{\sigma}^2(X^T X)^{-1}\right)$$

Simple Significance Test Large-sample #flashcard

- Null and alternative hypothesis:

$$\begin{aligned} H_0 : \beta_k &= 0 \\ H_1 : \beta_k &\neq 0 \end{aligned}$$

- Test statistic and its distribution under the null

$$t_k = \frac{\hat{\beta}_k}{se_k} \sim N(0, 1)$$

where se_k is the (k, k) element of $\hat{\sigma}^2(X^T X)^{-1}$

- Decision rule at a significant level α : reject H_0 if $|t_k| > c_{\frac{\alpha}{2}}(n - K)$ where $c_{\frac{\alpha}{2}}(n - K)$ is the corresponding critical value

Asymptotic General Linear Hypothesis Test #flashcard

- Null and alternative hypothesis:

$$\begin{aligned} H_0 : H\beta &= \theta \\ H_1 : H\beta &\neq \theta \end{aligned}$$

where $H_{p \times K}$ is the restriction matrix

- Test statistic and its distribution under the null

$$w = (H\hat{\beta} - \theta)^T \left[\widehat{Var}(\hat{\beta}|X) H^T \right]^{-1} (H\hat{\beta} - \theta) \sim^a \chi_p^2$$

where $\widehat{Var}(\hat{\beta}|X) = \hat{\sigma}^2(X^T X)^{-1}$

- Decision rule at a significant level α :

Reject H_0 if $w > c_\alpha(n - K)$

where $c_\alpha(n - K)$ is the corresponding critical value

- Proof of distribution

More precisely, recall the asymptotic approximation

$$\hat{\beta} \sim_a N\{\beta, \hat{\sigma}^2(X^T X)^{-1}\}$$

Then,

$$H\hat{\beta} \sim_a N\{H\beta, H\hat{\sigma}^2(X^T X)^{-1} H^T\}$$

Property 1 of quadratic forms (slide 15 of Topic 3)

$$w = (H\hat{\beta} - \theta)^T \{H\hat{\sigma}^2(X^T X)^{-1} H^T\}^{-1} (H\hat{\beta} - \theta) \sim_a \chi_p^2$$

Demand Var → Σ^{-1} (const) Demand Var
→ χ_p^2 if the other X 's knew β , so we will
not end up with a χ^2 -distribution

N 2.8 Heteroskedasticity

Inference Under Heteroskedasticity #flashcard

- We keep all the same assumptions except homoskedasticity

- $y = XB + u$
- (u_i, x_i^T) are iid with $\mathbb{E}[x_i u_i] = 0 \forall i = 1, \dots, n$
- The $K \times K$ matrix $M_{XX} = \mathbb{E}[x_i x_i^T]$ exists and is non-singular
- The $K \times K$ matrix $M_{X\Omega X} = \mathbb{E}[u^2 x_i x_i^T]$ exists and is non-singular
- \Rightarrow Normality: $\sqrt{n}(\hat{\beta} - \beta) \xrightarrow{D} N(0, M_{XX}^{-1} M_{X\Omega X} M_{XX}^{-1})$

- However, now $M_{XX}^{-1}M_{X\Omega X}M_{XX}^{-1}$ does not simplify to $\sigma^2(X^T X)^{-1}$, we need to estimate it directly using the **White/Eicker-Huber-White Estimator**:

$$\begin{aligned} & (X^T X)^{-1} \left(\sum_{i=1}^n \hat{u}_i^2 x_i x_i^T \right) (X^T X)^{-1} \\ &= \underbrace{\frac{1}{n} \left(\frac{X^T X}{n} \right)^{-1}}_{\rightarrow^p M_{XX}^{-1}} \underbrace{\left(\frac{1}{n} \sum_{i=1}^n \hat{u}_i^2 x_i x_i^T \right)}_{\rightarrow^p M_{X\Omega X}} \underbrace{\left(\frac{X^T X}{n} \right)^{-1}}_{\rightarrow^p M_{XX}^{-1}} \\ &\rightarrow^p \frac{1}{n} M_{XX}^{-1} M_{X\Omega X} M_{XX}^{-1} \end{aligned}$$

- Therefore:

$$\hat{\beta} \sim^a N \left(\beta, (X^T X)^{-1} \left(\sum_{i=1}^n \hat{u}_i^2 x_i x_i^T \right) (X^T X)^{-1} \right)$$

White Test for Heteroskedasticity #flashcard

1. Estimate the model using OLS and compute residuals \hat{u}_i
2. Run **auxiliary regression** of squared residual \hat{u}_i^2 on all regressors and cross terms
3. Run a global significance test for the auxiliary regression (H_0 : all coefficient = 0)

Generalised Least Squares (GLS) #flashcard

- GLS Assumptions
 - $y = X\beta + u$
 - $\mathbb{E}[u|X] = 0$
 - $Var[u|X] = \Omega$ is a **known** positive semi-definite matrix
 - $\implies \exists B$ s.t. $\Omega^{-1} = B^T B$ and $B\Omega B^T = I$
 - X has full rank
- We can implement GLS as following:
 - Let $y^* = By$, $X^* = BX$, $u^* = Bu$
 - Then:
 - $y^* = X^*\beta + u^*$
 - $\mathbb{E}[u^*|X] = 0$
 - $Var[u^*|X] = Var[Bu|X] = BVar[u|X]B^T = I$ is positive definite
 - This satisfies all assumptions of Gauss-Markov Theorem $\hat{\beta}^*$ is BLUE (**Aitken's Theorem**) with specific formula:

$$\begin{aligned} \hat{\beta}^* &= \hat{\beta}_{GLS} = (X^{*T} X^*)^{-1} (X^{*T} y^*) \\ &= (X^T B^T B X)^{-1} (X^T B^T B y) \\ &= (X^T \Omega^{-1} X)^{-1} (X^T \Omega^{-1} y) \\ &\text{equivalently } = \arg \min_{\beta} (y - X\beta)^T \Omega^{-1} (y - X\beta) \end{aligned}$$

- We can show that:
 - $\mathbb{E}[\hat{\beta}_{GLS}|X] = \beta$
 - $Var[\hat{\beta}_{GLS}|X] = (X^T \Omega X)^{-1}$

Feasible Generalised Least Squares (FGLS) #flashcard

- The cov-var matrix Ω is impossible to be estimated:
 - Completely unrestricted $\implies \frac{n(n+1)}{2}$ variables
 - Assume no autocorrel
 - ation $\implies n$ variables
- A feasible method is to impose a parametric form
- Example:
 - Assume $Var[u_i|X] = \sigma^2 x_{iK}^2$
 - Estimate the model: $y_i^* = x_i^{*T} \beta + u_i^*$ where $y_i^* = \frac{y_i}{x_{iK}}$, $x_i^* = \frac{x_i}{x_{iK}}$, $u_i^* = \frac{u_i}{x_{iK}}$
- Another flexible example: $Var[u|X] = \sigma^2 \exp(\delta_0) + \delta_1 x_{1i} + \cdots + \delta_K x_{iK}$

N 2 9 Cluster Robust Inference

Cluster Robust Inference #flashcard

- Assumptions:

$$\underbrace{\begin{bmatrix} y_1 \\ y_2 \\ \dots \\ y_C \\ y \end{bmatrix}}_{X} = \underbrace{\begin{bmatrix} X_1 \\ X_2 \\ \dots \\ X_C \end{bmatrix}}_{X} \beta + \underbrace{\begin{bmatrix} u_1 \\ u_2 \\ \dots \\ u_C \\ u \end{bmatrix}}_{u}$$

where (y_{ic}, x_{ic}^T) are independent across clusters but not within clusters (u_{ic}, u_{jc} can be correlated with general parameters).

- Key assumption: $\mathbb{E}[u_{ic}|x_{ic}] = 0$
- The var-cov matrix is:

$$\mathbb{E}[uu^T|X] = \Omega = \begin{bmatrix} \Omega_1 & 0 & \dots & 0 \\ 0 & \Omega_2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \Omega_C \end{bmatrix}$$

where each Ω_c is a symmetric, but otherwise unrestricted, $n_c \times n_c$ matrix

- Cluster-robust Inference: As the number of clusters $C \rightarrow \infty$:

$$\sqrt{C}(\hat{\beta} - \beta) \xrightarrow{D} N(0, V)$$

where the asymptotic variance V can be consistently estimated by:

$$\hat{V} = C(X^T X)^{-1} \left(\sum_{c=1}^C X_c^T \hat{u}_c \hat{u}_c^T X_c \right) (X^T X)^{-1}$$

or

$$\hat{\beta} \sim^a N \left(\beta, (X^T X)^{-1} \left(\sum_{c=1}^C X_c^T \hat{u}_c \hat{u}_c^T X_c \right) (X^T X)^{-1} \right)$$

and $\hat{u}_c = y_c - X_c \hat{\beta}$ is the OLS residual for the n_c observations in cluster c