Table of Contents

TABLE of FLASHCARDS

Section (31)	Flashcard
N 2 8 Heteroskedasticity	**Feasible Generalised Least Squares (FGLS)** #flashcard
N 2 9 Cluster Robust Inference	**Cluster Robust Inference** #flashcard

MT Part II: OLS

2 1 OLS

OLS Estimator #flashcard

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

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

Perfect Multicollinearity #flashcard

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

Classical Linear Regression Model with Stochastic Regressors #flashcard

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

(Conditional) Gauss-Markov Theorem #flashcard

In classical linear regression model:

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

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:

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

with property:

$$PX = X$$

and symetric and indopotent:

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

Annihilator Matrix #flashcard

The $n \times n$ matrix:

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

with the property:

$$X = 0$$

Normal Equations #flashcard

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

Estimator of Standard Error #flashcard

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

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 ($_{Y,X_k|\{X_i\}_{i\neq k}}$)

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**:

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

- **Gauss-Markov Theorem**:

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

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:

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

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 is an $n \times n$ non-stochastic, symmetric, and idempotent matrix with $rank() = r < n$, then the scalar:

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

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):

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

- 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:

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

- ${}_{kk}$ is the variance of $\hat{\beta}$ and $\hat{}_{kk}$ 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 $z_k = \frac{\hat{\beta}_k - \beta_k}{\sqrt{{}_{kk}}} \sim N(0,1)$:

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

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

- Derive the distribution for z_k:

$$
\begin{aligned}
{}_{kk} &= Var(\hat{\beta}_k) = \text{(k,k) element of } \sigma^2 (X^T X)^{-1} \\
\implies z_k &= \frac{\hat{\beta}_k - \beta_k}{\sqrt{{}_{kk}}} \sim N(0,1) \qquad \text{(tandardisation)}
\end{aligned}
$$

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

- Get a more comfortable expression:
- As we defined before: ${}_{kk}$ and $\hat{}_{kk}$ 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
- $\implies$

$$
\begin{aligned}
\frac{\sqrt{\hat{}_{kk}}}{\sqrt{{}_{kk}}} &= \sqrt{\frac{\hat{}_{kk}}{{}_{kk}}} \\
&= \sqrt{\frac{\hat{\sigma}^2}{\sigma^2}} \qquad \text{(arued aboe)} \\
&= \sqrt{\frac{{\color{red}(n-K)} \frac{\hat{\sigma}^2}{\sigma^2}}{\color{red}n-K}}
\end{aligned}
$$

- Show $(n-K)\frac{\sigma^2}{\hat{\sigma}^2} \sim \chi^2_{n-K}$:
- Start from the expression of $\hat{u}$:

$$
\begin{aligned}
\hat{u} &= y - X(X^T X)^{-1} X^T y \\
&= \underbrace{\Big(I - X(X^T X)^{-1} X^T\Big) y}_{} \\
&= (X\beta + u) \\
&= \Big(X\beta + \sigma \frac{u}{\sigma}\Big) \\
&= \; X^{\;0} \beta + \sigma \frac{u}{\sigma} \\
&= \sigma \quad \underbrace{\frac{u}{\sigma}}_{z \sim N(0,1)} \\
&= \sigma z
\end{aligned}
$$

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

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

- 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:
>
> $$w = (y-\mu)^T\Sigma^{-1}(y-\mu) \sim \chi^2_n$$
>
>   - If $z \sim N(0,1)$ is an $n \times 1$ vector and is an $n \times n$ non-stochastic, symmetric, and idempotent matrix with $rank() = r < n$, then the scalar:
>
> $$w = z^T z \sim \chi^2_r$$

- Apply the above theorem:

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

- 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

Simple Significance Test for CRM with Normal Errors #flashcard

1. Null and alternative hypothesis:

$$\begin{aligned}{}_0 :&\beta_k = 0 \\ {}_1 :&\beta_k \neq 0\end{aligned}$$

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 : reject $_0$ if $|t_k| > c_{\frac{}{2}}(n-K)$ where $c_{\frac{}{2}}(n-K)$ is the corresponding critical value

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

1. Null and alternative hypothesis:

$$\begin{aligned} {}_0 &: \beta = \theta \\ {}_1 &: \beta \neq \theta \end{aligned}$$

   where ${}_{p\times K}$ is the restriction matrix
2. Test statistic and its distribution under the null

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

3. Decision rule at a significant level :

$$\text{Reect}_0 \text{ if } > c(n-K)$$

   where $c(n-K)$ 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:

$$\begin{aligned} {}_0 &: \beta = \theta \\ {}_1 &: \beta \neq \theta \end{aligned}$$

   where ${}_{p\times K}$ is the restriction matrix
2. Test statistic and its distribution under the null

$$= \frac{n-K}{p} \times \frac{R^2_{\text{nrestricted}} - R^2_{\text{Restricted}}}{1 - R^2_{\text{nrestricted}}} \sim F_{p,n-K}$$

3. Decision rule at a significant level :

$$\text{Reect}_0 \text{ if } > c(n-K)$$

   where $c(n-K)$ is the corresponding critical value

- Note that **Global Significance Test** is a special case:
  - ${}_0 : \beta_2 = \beta_3 = \ldots \beta_K = 0$
  - $= \frac{n-K}{p} \times \frac{R^2}{1-R^2} \sim F_{K-1,n-K}$

N 2 6 Large Sample Properties

Consistency and Asymptotic Normality Requirement on OLS #flashcard

- $y = X + u$
- (u_i, x_i^T) are iid with $\mathbb{E}[x_i u_i] = 0 \; \forall \; i = 1, \ldots, n$
- The $K \times K$ matrix ${}_{XX} = \mathbb{E}\left[x_i x_i^T\right]$ exists and is non-singular
  - $\implies$ Consistency: $\hat{\beta} \to^p \beta$
- The $K \times K$ matrix ${}_{XX} = \mathbb{E}\left[u^2 x_i x_i^T\right]$ exists and is non-singular
  - $\implies$ Normality: $\sqrt{n}(\hat{\beta} - \beta) \to N\left(0, {}^{-1}_{XX} {}_{XX} {}^{-1}_{XX}\right)$

Proof of Consistency and Normality of OLS #flashcard

- Requirements:
  -
    - $y = X + u$

- (u_i, x_i^T) are iid with $\mathbb{E}\left[x_i u_i\right] = 0 \, \forall \, i = 1, \ldots, n$
- The $K \times K$ matrix $_{XX} = \mathbb{E}\left[x_i x_i^T\right]$ exists and is non-singular (for consistency)
- The $K \times K$ matrix $_{XX} = \mathbb{E}\left[u^2 x_i x_i^T\right]$ exists and is non-singular (for noramlity)

- 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:

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

- 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:

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

N 2 7 Large Sample Inference

OLS Large Sample Inference Under Homoskedasticity #flashcard

- **Assumptions**
  - $y = X + u$
  - (u_i, x_i^T) are iid with $\mathbb{E}\left[x_i u_i\right] = 0 \, \forall \, i = 1, \ldots, n$
  - The $K \times K$ matrix $_{XX} = \mathbb{E}\left[x_i x_i^T\right]$ exists and is non-singular
  - The $K \times K$ matrix $_{XX} = \mathbb{E}\left[u^2 x_i x_i^T\right]$ exists and is non-singular
    - $\implies$ Normality: $\sqrt{n}(\hat{\beta} - \beta) \to N\left(0, {}_{XX}^{-1} {}_{XX} {}_{XX}^{-1}\right)$
  - **Homoskedasticity**:
    - $\implies {}_{XX} = \mathbb{E}\left[u^2 x_i x_i^T\right] = \sigma^2 \mathbb{E}\left[x_i x_i^T\right] = \sigma^2 {}_{XX}$
    - $\implies \sqrt{n}(\hat{\beta} - \beta) \to N\left(0, \sigma^2 {}_{XX}^{-1}\right)$
- Using consistent estimators:
  - $\hat{\sigma}^2 = \frac{1}{n-K} \sum_{i=1}^n \hat{u}_i^2$
  - $\hat{}_{XX} = \frac{1}{n} \sum_{i=1}^n x_i x_i^T = \frac{X^T X}{n}$

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

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

Simple Significance Test Large-sample #flashcard

1. Null and alternative hypothesis:

$$\begin{aligned} {}_0 &: \beta_k = 0 \\ {}_1 &: \beta_k \neq 0 \end{aligned}$$

2. Test statistic and its distribution under the null

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

   where se_k is the (k,k) element of $\hat{\sigma}^2 (X^T X)^{-1}$
3. Decision rule at a significant level : reject ${}_0$ if $|t_k| > c_{\frac{}{2}}(n-K)$ where $c_{\frac{}{2}}(n-K)$ is the corresponding critical value

Asymptotic General Linear Hypothesis Test #flashcard

1. Null and alternative hypothesis:

$$\begin{aligned} {}_0 &: \beta = \theta \\ {}_1 &: \beta \neq \theta \end{aligned}$$

   where ${}_{p \times K}$ is the restriction matrix
2. Test statistic and its distribution under the null

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

   where $\widehat{Var}(\hat{\beta}|X) = \hat{\sigma}^2 (X^T X)^{-1}$
3. Decision rule at a significant level :

$$\text{Reect }{}_0 \text{ if } > c(n-K)$$

   where $c(n-K)$ is the corresponding critical value

- Proof of distribution
  -

More precisely, recall the asymptotic approximation

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

Then,

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

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

$$w = \underbrace{\underbrace{(H\hat{\beta} - \theta)'}_{\text{Demeaned Vec}^T}\{\underbrace{H\hat{\sigma}^2 (X'X)^{-1} H'}_{\to \Sigma^{-1} \text{(const)}}\}^{-1} \underbrace{(H\hat{\beta} - \theta)}_{\text{Demeaned Vec}}}_{\to \chi^2_p} \sim_a \chi^2_p$$

$\nexists$ the other χ^2 from $\hat{\sigma}^2$, so we will not end up with a $\mathcal{F}$-distribution

N 2 8 Heteroskedasticity

Inference Under Heteroskedasticity #flashcard

- We keep all the same assumptions except homoskedasticity

- $y = X + u$
- (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 ${}_{XX} = \mathbb{E}\left[x_i x_i^T\right]$ exists and is non-singular
- The $K \times K$ matrix ${}_{XX} = \mathbb{E}\left[u^2 x_i x_i^T\right]$ exists and is non-singular
    - $\implies$ Normality: $\sqrt{n}(\hat{\beta} - \beta) \to N\left(0, {}^{-1}_{XX} {}_{XX} {}^{-1}_{XX}\right)$

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

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

- 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 ($_0$: all coefficient = 0)

Generalised Least Squares (GLS) #flashcard

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

$$
\begin{aligned}
\hat{\beta} = \hat{\beta}_{LS} &= (X^T X)^{-1} (X^T y) \\
&= (X^{TT} X)^{-1} (X^{TT} y) \\
&= (X^{T-1} X)^{-1} (X^{T-1} y) \\
\text{euialently} &= arg \min_{\beta} (y - X\beta)^{T-1} (y - X\beta)
\end{aligned}
$$

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

Feasible Generalised Least Squares (FGLS) #flashcard

- The cov-var matrix 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: $Var[u|X] = \sigma^2 \exp(\delta_0) + \delta_1 x_{1i} + \ + \delta_K x_{iK}$

N 2 9 Cluster Robust Inference

Cluster Robust Inference #flashcard

- **Assuptions**:
- $\underbrace{\begin{matrix} y_1 \\ y_2 \\ \dots \\ y \end{matrix}}_{y} = \underbrace{\begin{matrix} X_1 \\ X_2 \\ \dots \\ X \end{matrix}}_{X} \beta + \underbrace{\begin{matrix} u_1 \\ u_2 \\ \dots \\ u \end{matrix}}_{u} \quad where (y_{ic}, x_{ic}^T) are independent acrss custers but nt within custers (u_{ic}, u_{jc} can be crreated with genera patterns.$
- Key assumption: $\mathbb{E}\left[u_{ic}|x_{ic}\right] = 0$
- The var-cov matrix is:

$$\mathbb{E}\left[uu^T|X\right] = \ = \begin{matrix} {}_1 & 0 & \dots & 0 \\ 0 & {}_2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \end{matrix}$$

  where each ${}_c$ is a symmetric, but otherwise unrestricted, $n_c \times n_c$ matrix
- **Cluster-robust Inference**: As the number of clusters $\to \infty$:

$$\sqrt{}(\hat{\beta} - \beta) \to N(0, V)$$

  where the asymptotic variance V can be consistently estimated by:

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

  or

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

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