การทดสอบสมมติฐานสำหรับความต่างของความชันสำหรับตัวแบบข้อมูลพาเนลสามมิติ

นางสาวศศิธร กนกหิรัญญากร

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต
สาขาวิชาคณิตศาสตร์ประยุกต์และวิทยาการคณนา
ภาควิชาคณิตศาสตร์และวิทยาการคอมพิวเตอร์
คณะวิทยาศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย
ปีการศึกษา 2559

HYPOTHESIS TESTING FOR SLOPE HETEROGENEITY FOR THREE DIMENSIONAL PANEL DATA MODEL

Miss Sasitorn Kanokhilunyakorn

A Thesis Submitted in Partial Fulfillment of the Requirements
for the Degree of Master of Science in Applied Mathematics and
Computational Science
Department of Mathematics and Computer Science
Faculty of Science
Chulalongkorn University
Academic Year 2016

Thesis Title	HYPOTHESIS TESTING FOR SLOPE HETEROGENEITY FOR THREE DIMENSIONAL PANEL DATA MODEL
By	Miss Sasitorn Kanokhilunyakorn
Field of Study	Applied Mathematics and Computational Science
Thesis Advisor	Jiraphan Suntornchost, Ph.D.

Accepted by the Faculty of Science, Chulalongkorn University in Partial Fulfillment of the Requirements for the Master Degree

. . . . Dean of the Faculty of Science

(Associate Professor Polkit Sangvanich, Ph.D.)

THESIS COMMITTEE

. . . . Chairman

(Assistant Professor Petarpa Boonserm, Ph.D.)

. . . . Thesis Advisor

(Jiraphan Suntornchost, Ph.D.)

. . . . Examiner

(Kitiporn Plaimas, Ph.D.)

. . . . External Examiner

(Monthira Duangsaphon, Ph.D.)

ศศิธร กนกหิรัญญากร : การทดสอบสมมติฐานสำหรับความต่างของความชันสำหรับตัวแบบข้อมูลพาเนลสามมิติ. (HYPOTHESIS TESTING FOR SLOPE HETEROGENEITY FOR THREE DIMENSIONAL PANEL DATA MODEL) อ.ที่ปรึกษาวิทยานิพนธ์หลัก : อาจารย์ ดร. จิราพรรณ สุนทรโชติ, 77 หน้า.

ในหลายทศวรรษที่ผ่านมา เราได้เห็นหลายการทดสอบสำหรับความต่างของความชันที่มีบทบาทสำคัญในการวิเคราะห์ข้อมูลพาเนลในปัจจุบัน อย่างไรก็ตามการทดสอบเหล่านั้นถูกนำเสนอสำหรับตัวแบบข้อมูลพาเนลแบบฉบับ ในขณะที่ข้อมูลพาเนลแบบหลายมิติมีความพร้อมและมีความสำคัญในการใช้งานมากขึ้น ดังนั้นเป้าหมายหลักของงานวิจัยคือการหาค่าสถิติสำหรับการทดสอบความต่างที่ใช้สำหรับตัวแบบข้อมูลพาเนลสามมิติ การทดสอบของเราคือค่าสถิติการทดสอบตัวคูณลากรานจ์ขึ้นอยู่กับฟังก์ชันควรจะเป็นปรกติตามเงื่อนไขของค่าสถิติที่เพียงพอ การทดสอบสามารถประมาณได้ด้วยการแจกแจงปรกติเมื่อจำนวนของระดับของทิศทางที่พิจารณามีแนวโน้มอนันต์และจำนวนของช่วงเวลานั้นคงที่ ผลการจำลองแสดงว่าการทดสอบที่ได้เสนอไปมีผลที่ดีตามข้อกำหนด

ภาควิชา..........คณิตศาสตร์และ.........
.....วิทยาการคอมพิวเตอร์......
สาขาวิชา.......คณิตศาสตร์ประยุกต์.....
.......และวิทยาการคณนา......
ปีการศึกษา ....2559................

ลายมือชื่อนิสิต ....
ลายมือชื่อ อ.ที่ปรึกษาหลัก ....

5772161223 : MAJOR APPLIED MATHEMATICS AND COMPUTATIONAL SCIENCE KEYWORDS : HYPOTHESIS TEST/ HETEROGENEITY/ THREE DIMENSIONAL DATA/ PANEL DATA ANALYSIS

SASITORN KANOKHILUNYAKORN : HYPOTHESIS TESTING FOR SLOPE HETEROGENEITY FOR THREE DIMENSIONAL PANEL DATA MODEL ADVISOR : JIRAPHAN SUNTORNCHOST, Ph.D. 77 pp.

Tests for slope heterogeneity play important roles in panel data analysis nowadays as we can see several tests have been proposed in the last decades. However, those tests were proposed for classical panel models while multi-dimensional panel data have been becoming more available and are important in applications. Therefore, the main goal of this thesis is to derive a heterogeneity test statistic that is applicable for three-dimensional panel data model. Our test is the Lagrange multiplier test statistic based on the conditional Gaussian likelihood function of a sufficient statistic. The test is asymptotically normally distributed when the number of levels of considered direction tends to infinity and the number of time periods is fixed. Simulation results show that the proposed test has good size under considered specifications.

Department : ....Mathematics and.........
.......Computer Science........
Field of Study : Applied Mathematics and
...Computational Science...
Academic Year : ....2016................

Student's Signature : ....
Advisor's Signature : ....

ACKNOWLEDGEMENTS

First, I would like to express my deep gratitude to my thesis advisor, Dr. Jiraphan Suntornchost, for insightful suggestions on my work. She encouraged and advised me thorough the thesis process. I also would like to thanks and deep appreciation to my thesis committees, Assistant Professor Dr. Petarpa Boonserm, Dr. Kitiporn Plaimas and Dr. Monthira Duangsaphon, for their comments and suggestions. Moreover, I would like to thank all teachers who have instructed and taught me for valuable knowledge.

In particular, thank to my dear friends for giving me good advices for my thesis and experiences at Chulalongkorn university. Finally, I would like to express my gratitude to my beloved family for their love and financial support throughout my undergraduate and graduate study.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

In statistics, classical panel data refer to two-dimensional data involving measurements over time. Panel data methodology within economics had been largely developed through economic applications. Economic applications of panel data methods are not confined to survey or economic problems. For example, Accuracy and rationality of state General Fund Revenue forecasts: Evidence from panel data [H. N. Mocan, Sam Azad, 1995], Deterioration of Firm Balance Sheet and Investment Behavior: Evidence from Panel Data on Thai Firms [S. Rungsomboon, 2005], Dynamics of social health insurance development: Examining the determinants of Chinese basic health insurance coverage with panel data [J. Q. Liu, 2011], et cetera. Most panel data models assume that the slope coefficients are the same for all cross sectional units. However, the intercept parameter is allowed to have some variations over individuals. In general, fixed effects estimation may not estimate the cross sectional regression parameters when there are different regression parameters. Therefore, it is possible to test if the regression slope coefficients vary over the cross sectional unit. The test is valid where both time (T) and the number of cross sectional units (N) tend to infinity. In 1970, Swamy [26] proposed a test, where the null hypothesis is to have a constant slope for each cross-sectional unit, but each unit is allowed to have a different error variance. The asymptotic distribution of the test is valid when $T \to \infty$ and N is fixed. In 2008, Pesaran and Yamagata [21] proposed a test which is valid when $N, T \to \infty$. Recently, in 2014, Juhl and Lugovskyy [16] proposed a test which is valid when $N \to \infty$ and T is fixed. In Monte Carlo experiments, they showed that their test performs very well relative to existing tests. In particular, the Juhl and Lugovskyy test was derived from the conditional Lagrange multiplier test has excellent size and power.

In applications, there are some situations where the classical panel data model is not optimal. For example, data from complex survey having multi-dimensional strata. For example, international tourism data [30], bilateral donors' aid allocation [15], etc. Since complex assumptions can be made on the precise structure of the correlations among errors in this model. In this study, we extend the heterogeneity hypothesis test to handle three dimensional panel data models. A multidimensional setup in panel data model was proposed by Matyas [18] and then Matyas and Balazsi [19] represented parameter estimations of three dimensional fixed effects panel data models. In general, fixed effects estimation may not estimate the cross sectional regression parameters where there are different regression parameters. Also, it is possible to test if the regression slope coefficients vary over the cross sectional unit. In this work, three dimensional panel data represents time series of two way data. We wish to test if the regression slope coefficients vary over the two factors in the regression models for fixed time periods by extending the conditional Lagrange multiplier test for classical panel data model proposed in Juhl and Lugovskyy to three dimensional panel data. Moreover, we evaluate the size of the statistic using a simulation experiment based on data generated from a variety of distributions.

This thesis is organized as follows. In Chapter 2, we provide the basic knowledge of statistics, matrix calculus, linear regression model and panel data models which will be used in this thesis. In Chapter 3, we construct a test statistic for slope heterogeneity for three dimensional panel data model by conditional Lagrange multiplier. In Chapter 4, an empirical study of the performance of our proposed test statistic based on simulations and discussions are provided. Conclusions are provided in Chapter 5.

CHAPTER II
PRELIMINARIES

In this chapter, we provide basic knowledge of statistics, matrix calculus, linear regression model and panel data models which will be used in this thesis.

2.1 Basic Knowledge of Statistics

In this section, we provide basic knowledge of statistics such as a statistical distribution, principal of data reduction, method of parameter estimators and hypothesis tests. We also give brief introductions of the conditional likelihood and Lagrange multiplier test that are the main ideas to construct our test statistic in this thesis. In addition, we give a review of limit theorems such as Chebychev's weak law of large numbers and Linderberg-Feller central limit theorem which will be used in studying asymptotic properties of our statistic.

2.1.1 Statistical Distribution

Statistical distributions are used to model populations. There are two categories of statistical distributions. The first category is the discrete distributions where a random variable X is said to have a discrete distribution if the range of X is countable. In many situations, discrete random variables have integer-valued outcomes. The second category is the continuous distributions where a random variable X has its range as an interval of possible outcomes. Some well-know continuous distributions which will be used in this thesis are discussed in Definitions 2.1 - 2.4 as follows.

Definition 2.1. [8] The *uniform distribution* is defined by spreading mass uniformly over an interval say, $[a, b]$, denoted by $U[a, b]$. Its probability density function is given by

$$f(x|a,b) = \begin{cases} \dfrac{1}{b-a} & \text{if } x \in [a,b] \\ 0 & otherwise. \end{cases}$$

The uniform distribution $U[a,b]$ has expectation and variance given by

$$E(X) = \frac{b+a}{2} \quad \text{and} \quad Var(X) = \frac{(b-a)^2}{12}.$$

Definition 2.2. [8] The *normal distribution* (sometimes called the *Gaussian distribution*) has two parameters, usually denoted by μ and σ^2, which are its mean and variance. The probability density function of the normal distribution with mean μ and variance σ^2 (usually denoted by $N(\mu, \sigma^2)$) is given by

$$f(x|\mu,\sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/2\sigma^2}, \qquad -\infty < x < \infty.$$

If a normal distribution has zero mean and variance one, it is called the *standard normal distribution*.

Definition 2.3. [8] The *multivariate normal distribution* (sometimes called the *multivariate Gaussian distribution*) of a k-dimensional random vector has two parameters, usually denoted by $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$, which are its k-dimensional mean vector and $k \times k$ covariance matrix. The probability density function of the multivariate normal distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$ (usually denoted by $N_k(\boldsymbol{\mu}, \boldsymbol{\Sigma})$) is given by

$$f(\boldsymbol{x}|\boldsymbol{\mu},\boldsymbol{\Sigma}) = (\sqrt{2\pi})^{-\frac{1}{2}k} |\boldsymbol{\Sigma}|^{-\frac{1}{2}} e^{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})}.$$

Definition 2.4. [8] The *chi-squared distribution with k degrees of freedom*, denoted by $\chi^2(k)$, is the sum of the squares of k independent standard normal random variables. Its probability density function is given by

$$f(x|k) = \frac{1}{2^{\frac{k}{2}}\Gamma(\frac{k}{2})} x^{\frac{k}{2}-1} e^{-\frac{x}{2}}, \quad x \geq 0.$$

The chi-square distribution $\chi^2(k)$ has expectation and variance given by

$$E(X) = k \quad \text{and} \quad Var(X) = 2k.$$

Definition 2.5. [8] The *t-distribution with k degrees of freedom*, denoted by $t_{(k)}$, is in the form $\dfrac{Z}{\sqrt{U/k}}$ where Z is distributed as a standard normal, U is distributed as a $\chi^2(k)$, and Z and U are independent. Its probability density function is given by

$$f(x|k) = \frac{\Gamma(\dfrac{k+1}{2})}{\sqrt{k\pi}\Gamma(\dfrac{k}{2})} \left(1 + \frac{x^2}{k}\right)^{-\dfrac{k+1}{2}}, \quad -\infty < x < \infty.$$

The t-distribution $t_{(k)}$ has expectation and variance given by

$$E(X) = 0 \text{ for } k > 1 \quad \text{and} \quad Var(X) = \frac{k}{k-2} \text{ for } k > 2.$$

2.1.2 Principles of Data Reduction

In this part, we study in principles of data reduction. Data reduction is a function of the data called *statistics* that may summarize all the information in the sample about parameters. We introduce two principles which are the sufficient principle and the likelihood principle.

The Sufficient Principle

If $T(X)$ is a sufficient statistic for θ, then any inference about θ should depend on the sample X only through the value of $T(X)$. That is, if x and y are two sample points such that $T(x) = T(y)$, then the inference about θ should be the same whether $X = x$ or $X = y$ is observed. The formal definition of a sufficient statistic is given as follows.

Definition 2.6. [8] A statistic $T(X)$ is a *sufficient statistic* for θ if the conditional distribution of the sample X given the value of $T(X)$ does not depend on θ.

We introduce the following theorem to find a sufficient statistic by easier way.

Theorem 2.7 (Factorization theorem)**.** [8] *Let $f(x|\theta)$ denote the joint probability density function or probability mass function of a sample X. A statistic $T(X)$ is a sufficient statistics for θ if and only if there exist functions $g(t|\theta)$ and $h(x)$ such that, for all sample points x and all parameter points θ,*

$$f(x|\theta) = g(T(x)|\theta)h(x).$$

We can use the Factorization theorem to find a sufficient statistic by factoring the joint probability density function or probability mass function of the sample into two parts, with one part not depend on θ that is the $h(x)$ function. The another part, the one that depends on θ, usually depends on the sample x only through some function $T(x)$ and this function is a sufficient statistic for θ.

The Likelihood Principle

In this part, we study an important statistic called the likelihood function that also can be used to summarize data.

Definition 2.8. [8] Let $f(\boldsymbol{x}|\theta)$ denote the joint probability density function or probability mass function of the sample $\boldsymbol{X} = (X_1, X_2, ...X_n)$. Then, given that $\boldsymbol{X} = \boldsymbol{x}$ is observed, the function of θ defined by

$$L(\theta|\boldsymbol{x}) = f(\boldsymbol{x}|\theta)$$

is called the *likelihood function*.

The likelihood principle specifies how the likelihood function should be used as a data reduction device.

Definition 2.9. [8] Likelihood principle is defined as if $\boldsymbol{x}$ and $\boldsymbol{y}$ are two sample points such that $L(\theta|\boldsymbol{x})$ is proportional to $L(\theta|\boldsymbol{y})$, that is, there exists a constant $C(\boldsymbol{x}, \boldsymbol{y})$ such that

$$L(\theta|\boldsymbol{x}) = C(\boldsymbol{x}, \boldsymbol{y})L(\theta|\boldsymbol{y}) \quad \text{for all } \theta,$$

then the conclusions drawn from $\boldsymbol{x}$ and $\boldsymbol{y}$ should be identical.

2.1.3 Methods of Parameter Estimators

In this part, we study how to estimate parameters. For example, estimating a parameter with its sample parallel is usually reasonable. Particularly, the sample mean is a good estimate for the population mean. In more complicated models, we need a method of estimating parameters. In this thesis, we introduce two methods of parameter estimators such as method of moments and maximum likelihood estimators.

Method of Moments

Let $X_1, X_2, ..., X_n$ be a sample from a population with probability density function or probability mass function. Method of moments are found by equating the first k sample moments to the corresponding k population moments, and solving the resulting system of simultaneous equations. Define

$$\begin{aligned}
m_1 &= \frac{1}{n}\sum_{i=1}^{n} X_i^1, \qquad \mu_1^{'} = EX^1, \\
m_2 &= \frac{1}{n}\sum_{i=1}^{n} X_i^2, \qquad \mu_2^{'} = EX^2, \\
&\vdots \\
m_k &= \frac{1}{n}\sum_{i=1}^{n} X_i^k, \qquad \mu_k^{'} = EX^k.
\end{aligned}$$

The population moment $\mu_j^{'}$ will typically be a function of $\theta_1, ..., \theta_k$, say $\mu_j^{'}(\theta_1, ..., \theta_k)$. The estimators $(\tilde{\theta}_1, ..., \tilde{\theta}_k)$ of $(\theta_1, ..., \theta_k)$ are obtained by solving the following system of equations for $(\tilde{\theta}_1, ..., \tilde{\theta}_k)$ in terms of $(m_1, ..., m_k)$:

$$\begin{aligned} m_1 &= \mu_1^{'}(\theta_1, ..., \theta_k), \\ m_2 &= \mu_2^{'}(\theta_1, ..., \theta_k), \\ &\vdots \\ m_k &= \mu_k^{'}(\theta_1, ..., \theta_k). \end{aligned}$$

Maximum Likelihood Estimators

The method of maximum likelihood is one of the most popular techniques for deriving estimators. If $X_1, ..., X_n$ are independent and identically distributed with probability density function or probability mass function $f(\boldsymbol{x}|\theta_1, ..., \theta_k)$, the likelihood function is defined by

$$\begin{aligned} L(\boldsymbol{\theta}|\boldsymbol{x}) &= L(\theta_1, .., \theta_k | x_1, ..., x_n) \\ &= \prod_{i=1}^{n} f(x_i|\theta_1, .., \theta_k). \end{aligned}$$

Definition 2.10. [8] For each sample point $\boldsymbol{x}$, let $\hat{\boldsymbol{\theta}}(\boldsymbol{x})$ be a parameter estimator at which $L(\boldsymbol{\theta}|\boldsymbol{x})$ attains its maximum as a function of $\boldsymbol{\theta}$, with x held fixed. A *maximum likelihood estimator* (MLE) of the parameter $\boldsymbol{\theta}$ based on a sample X is $\hat{\boldsymbol{\theta}}(X)$.

Conditional Likelihood

In this part, we introduce conditional likelihood function and the conditional maximum likelihood estimator of the parameter vector.

Let $\boldsymbol{\phi} = (\boldsymbol{\theta}, \boldsymbol{\lambda})$, where $\boldsymbol{\theta}$ is the parameter of interest and $\boldsymbol{\lambda}$ is a vector of nuisance parameters. If there is a sufficient statistic $S_{\boldsymbol{\lambda}}$ of $\boldsymbol{\lambda}$ then the nuisance parameters can be eliminated from the likelihood by conditioning on it. The conditional log-likelihood can be obtained

$$l_c(\boldsymbol{\theta}) = \log(f_{\boldsymbol{Y}|S_{\boldsymbol{\lambda}}}(\boldsymbol{y})) = \log[f_{\boldsymbol{Y}}(\boldsymbol{y})] - \log[f_{\boldsymbol{\lambda}}(\boldsymbol{y})], \tag{2.1}$$

where $f_{\boldsymbol{Y}|S_{\boldsymbol{\lambda}}}(\boldsymbol{y})$ is the conditional distribution of the response $\boldsymbol{Y} = [Y_1\ Y_2\ \cdots\ Y_n]^\top$ given $S_{\boldsymbol{\lambda}}$. Then $\hat{\boldsymbol{\theta}}$ maximizing $l_c(\boldsymbol{\theta})$ is the maximum conditional likelihood estimator. If $S_{\boldsymbol{\lambda}}$ depends on $\boldsymbol{\theta}_0$ where $\boldsymbol{\theta}_0$ is the parameter of interest, the conditional log-likelihood can be obtained,

$$l_c(\boldsymbol{\theta}, \boldsymbol{\lambda}, \boldsymbol{\theta}_0) = \log\left(f_{\boldsymbol{Y}|S_{\boldsymbol{\lambda}(\boldsymbol{\theta}_0)}}\right) = \log\left[f_{\boldsymbol{Y}}\left(\boldsymbol{y}\right)\right] - \log[f_{S_{\boldsymbol{\lambda}(\boldsymbol{\theta}_0)}}(\boldsymbol{y})]. \tag{2.2}$$

Then, $\hat{\boldsymbol{\theta}}$ is the solution of the equation

$$\left[\frac{\partial l_c(\boldsymbol{\theta}, \boldsymbol{\lambda}, \boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}}\right]_{\boldsymbol{\theta}=\boldsymbol{\theta}_0, \boldsymbol{\lambda}=\hat{\boldsymbol{\lambda}}(\boldsymbol{\theta})} = 0, \tag{2.3}$$

where ∂ denotes for a partial derivative.

Mean Squared Error

We study finite sample measures of the quality of an estimator.

Definition 2.11. [8] The mean squared error (MSE) of an estimator W of a parameter a is the function of a parameter θ is the function of θ defined by $E_\theta(\hat{\theta} - \theta)^2$.

Definition 2.12. [8] The *bias* of a point estimator $\hat{\theta}$ of a parameter θ a is the difference between the expected value of $\hat{\theta}$ and θ; that is, $Bias_\theta(\hat{\theta}) = E_\theta(\hat{\theta}) - \theta$. An estimator whose bias is identically (in θ) equal to 0 is called an *unbiased* and satisfies $E_\theta(\hat{\theta}) = \tau(\theta)$ for all θ.

Note that from the two previous definitions, we have

$$E_\theta(\hat{\theta} - \theta)^2 = Var_\theta\hat{\theta} + (E_\theta(\hat{\theta}) - \theta)^2 = Var_\theta\hat{\theta} + (Bias_\theta(\hat{\theta}))^2.$$

Therefore, if $\hat{\theta}$ is an unbiased estimator of θ,

$$E_\theta(\hat{\theta} - \theta)^2 = Var_\theta \hat{\theta}.$$

2.1.4 Hypothesis Testing

In this part, we introduce basic knowledge of hypothesis testing.

Definition 2.13. [8] A *hypothesis* is a statement about a population parameter.

The goal of a hypothesis testing is to decide, based on a sample from the population, which of two complementary hypotheses is accepted.

Definition 2.14. [8] The two complementary hypotheses in a hypothesis testing problem are called the *null hypothesis* and the *alternative hypothesis*. They are denoted by H_0 and H_A, respectively.

In a hypothesis testing problem, after observing the sample the experimenter must decide either to accept H_0 or to reject H_0.

Definition 2.15. [8] A *hypothesis testing procedure* or *hypothesis test* is a rule that specifies:

i. For which sample values the decision is made to accept H_0 as true.

ii. For which sample values H_0 is rejected and H_A is accepted as true.

The subset of the sample space for which H_0 will be rejected is called the *rejection region* or *critical region*. The complement of the rejection region is called the *acceptance region*.

Definition 2.16. Type I error is rejection H_0 when it is true.

Definition 2.17. Type II error is failing to reject H_0 when it is false.

Definition 2.18. The *size* of a test is the probability of falsely rejecting the null hypothesis. That is, it is the probability of making a Type I error. It is denoted by α. For a simple hypothesis,

$$\alpha = P(\text{rejects } H_0 | H_0 \text{ is true}).$$

Definition 2.19. The *power*, $1-\beta$, is the probability of correctly rejecting H_0 when it is false where β is the probability of a type II error.

Lagrange Multiplier Test

In this research, we apply a Lagrange multiplier test to our model. Therefore, in this section, we introduce the Lagrange multiplier test presented by Breusch and Pagan [6] as follows.

Consider a sample of size N from a distribution which is known apart from a finite number K of unknown parameters $\boldsymbol{\theta} = (\theta_1, ..., \theta_K)^\top$ giving a log-likelihood $L(\boldsymbol{\theta})$. The hypothesis to be tested is specified as $p < K$ restrictions on $\boldsymbol{\theta}$,

$$H_0 : h_j(\boldsymbol{\theta}) = 0 \qquad j = 1, ..., p.$$

Aitchison and Silver [1,2] approached this problem by setting up the Lagrangean function

$$L(\boldsymbol{\theta}) + \sum_{j=1}^{p} \boldsymbol{\lambda}_j h_j(\boldsymbol{\theta}) \tag{2.4}$$

and derivative of (2.4) with respect to the unknown parameters $\boldsymbol{\theta}$ and the Lagrange multipliers $\boldsymbol{\lambda}_j$ to yield $\tilde{\boldsymbol{\theta}}$ and $\tilde{\boldsymbol{\lambda}}$ as the solution of the first order conditions

$$\tilde{\boldsymbol{D}} + \tilde{\boldsymbol{H}}\tilde{\boldsymbol{\lambda}} = \boldsymbol{0},$$
$$h_j(\tilde{\boldsymbol{\theta}}) = 0 \qquad j = 1, ..., p,$$

where

$$\tilde{\boldsymbol{D}} \text{ is the } (K \times 1) \text{ vector } \left[\frac{\partial L}{\partial \theta_i}(\tilde{\boldsymbol{\theta}})\right],$$
$$\tilde{\boldsymbol{H}} \text{ is the } (K \times p) \text{ matrix } \left[\frac{\partial h_j}{\partial \theta_i}(\tilde{\boldsymbol{\theta}})\right],$$

and

$$\tilde{\boldsymbol{\lambda}} \text{ is a } (p \times 1) \text{ Lagrange multipliers vector.}$$

When the null hypothesis is correct, the restricted estimate $\tilde{\boldsymbol{\theta}}$ will tend to be near the unrestricted maximum likelihood estimate so that $\tilde{\boldsymbol{D}}$ will be close to the zero vector. Under the regularity condition that the order of differentiation and integration can be interchanged, Feigin (1976) showed that $\boldsymbol{D}$ will be a zero mean martingale, and under conditions presented in Crowder (1976)

$$\boldsymbol{c}_N^{-1/2}\boldsymbol{D} \to N(0, \lim_{N\to\infty} \boldsymbol{c}_N^{-1/2}\boldsymbol{F}\boldsymbol{c}_N^{-1/2}) \text{ as } N \to \infty,$$

where $\boldsymbol{F} = E(-\partial^2 L/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top)$ is called Fisher's information matrix and $\boldsymbol{c}_N$ is some suitably chosen norming matrix. This leads to the LM test based upon the statistic

$$LM = \tilde{\boldsymbol{D}}^\top \tilde{\boldsymbol{F}}^{-1} \tilde{\boldsymbol{D}} = \tilde{\boldsymbol{\lambda}}^\top \tilde{\boldsymbol{H}}^\top \tilde{\boldsymbol{F}}^{-1} \tilde{\boldsymbol{H}} \tilde{\boldsymbol{\lambda}}, \tag{2.5}$$

where $\tilde{\boldsymbol{F}}$ is the information matrix when the null hypothesis is true, evaluated at the restricted estimates $\tilde{\boldsymbol{\theta}}$. The term $\tilde{\boldsymbol{D}}^\top \tilde{\boldsymbol{F}}^{-1} \tilde{\boldsymbol{D}}$ is the "score" statistic (see Rao [22]) while $\tilde{\boldsymbol{\lambda}}^\top \tilde{\boldsymbol{H}}^\top \tilde{\boldsymbol{F}}^{-1} \tilde{\boldsymbol{H}} \tilde{\boldsymbol{\lambda}}$ is the Lagrangean multiplier statistic (see Byron [7]), making it clear that the two test statistics are identical so that the choice of which form to use is based on convenience. Under the usual maximum likelihood regularity conditions, the LM statistic is asymptotically equivalent to the Wald and Likelihood Ratio statistics. That is, when H_0 is true it is asymptotically distributed as a chi-square distribution χ^2. Alternative, one-side score test $\sqrt{LM}$ is asymptotically distributed as the standard normal distribution when H_0 is true.

2.1.5 Convergence of Random Variables

In this part, we will discuss convergence modes and limit theorem of random variables. We study the behavior of certain sample quantities as the sample size approach to infinity.

Convergence in Probability

Definition 2.20. [8] A sequence of random variables X_n $(n \geq 1)$ *converges in probability* to a random variable X, denoted by $X_n \xrightarrow{p} X$, if

$$\lim_{n\to\infty} P(|X_n - X| < \epsilon) = 1,$$

for all $\epsilon > 0$.

Theorem 2.21. [8] *Suppose that X_n $(n \geq 1)$ are random variables and c is a constant such that*

$$E(X_n) \to c \quad and \quad Var(X_n) \xrightarrow{p} 0,$$

then

$$X_n \xrightarrow{p} c.$$

Theorem 2.22 (Chebychev's Weak Law of Large Numbers)**.** [12] *Let $X_1, X_2, ...$ be a sample of observations with $E[X_i] = \mu_i$ and $Var[X_i] = \sigma_i^2 < \infty$. Define $\bar{\sigma}_n^2/n = (1/n^2)\sum_{i=1}^{n} \sigma_i^2 \to 0$ as $n \to \infty$. Then, for every $\epsilon > 0$,*

$$\lim_{n\to\infty} P(\bar{X}_n - \bar{\mu}_n < \epsilon) = 1,$$

that is, $\bar{X}_n$ converges in probability to $\bar{\mu}_n$.

Convergence in Distribution

Definition 2.23. [8] A sequence of random variables X_n with distribution function F_{X_n} $(n \geq 1)$ *converges in distribution* to a random variable X, denoted by $X_n \xrightarrow{d} X$, if

$$\lim_{n\to\infty} F_{X_n}(x) = F_X(x),$$

at all point x where $F_X(x)$ is continuous. Here, F_{X_n} and F_X are the cumulative distribution functions of random variables X_n and X, respectively.

Theorem 2.24 (Lindeberg-Feller Central Limit Theorem)**.** [12] *Suppose that* $\{\boldsymbol{X}_i\}$, $i = 1, 2, ..., n$, *are sample of random vectors with* $E[\boldsymbol{X}_i] = \boldsymbol{\mu}_i$ *and* $Var[\boldsymbol{X}_i] = \boldsymbol{Q}_i$, *and all mixed third moments of the multivariate distribution are finite. Let*

$$\bar{\boldsymbol{\mu}}_n = \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\mu}_i, \quad \text{and} \quad \bar{\boldsymbol{Q}}_n = \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{Q}_i.$$

We assume that

$$\lim_{n\to\infty}\bar{\boldsymbol{Q}}_n = \boldsymbol{Q},$$

where $\bar{\boldsymbol{Q}}$ *is a finite, positive definite matrix, and that for every* i,

$$\lim_{n\to\infty}(n\bar{\boldsymbol{Q}}_n)^{-1}\boldsymbol{Q}_i = \lim_{n\to\infty}\left(\sum_{i=1}^{n}\boldsymbol{Q}_i\right)^{-1}\boldsymbol{Q}_i = \boldsymbol{0}.$$

We allow the means of the random vectors to differ, although in the cases that we will analyze, they will generally be identical. The second assumption states that individual components of the sum must be finite and diminish in significance. There is also an implicit assumption that the sum of matrices is nonsingular. Because the limiting matrix is nonsingular, the assumption must hold for large enough n, *which is all that concerns us here. With these in place, the result is*

$$\sqrt{n}(\bar{\boldsymbol{X}}_n - \bar{\boldsymbol{\mu}}_n) \xrightarrow{d} N(0, \boldsymbol{Q}).$$

2.1.6 Big O and Little o Notation

A sequence X_n of random vectors is said to be $O_p(1)$ if it is stochastically bounded and $o_p(1)$ if it converges in probability to zero. The notations gain power when we consider pairs of sequences. Suppose X_n and Y_n are random sequences taking values in any normed vector space, then

$$X_n = O_p(Y_n) \tag{2.6}$$

means $X_n/||Y_n||$ is stochastically bounded and

$$X_n = o_p(Y_n) \tag{2.7}$$

means $X_n/||Y_n||$ converges in probability to zero.

Definition 2.25. An estimator $\hat{\theta}_n$ of a parameter θ is a *consistent* estimator of θ if and only if

$$\lim_{n\to\infty} P(|\hat{\theta}_n - \theta| < \epsilon) = 1.$$

Definition 2.26. The *root* $-$ *N consistency* means that the sampling error is $O_p(N^{-1/2})$.

Theorem 2.27. *Let* x_i $(i = 1, 2, ..., N)$ *be* $O_p(1)$*, then* $\sum_{i=1}^{N} x_i = O_p(N)$*.*

Properties 2.28.

1. $O_p(N^a) + O_p(N^b)$ *is* $O_p(N^{max(a,b)})$*.*

2. $O_p(N^a) \times O_p(N^b)$ *is* $O_p(N^{a+b})$*.*

3. $O_p(N^a) + o_p(N^b)$ *is* $O_p(N^a)$ *if* $a \geq b$ *and is* $o_p(N^b)$ *if* $a < b$*.*

4. $O_p(N^{-1})$ *is* $o_p(1)$*.*

2.2 Basic Knowledge of Matrix Algebra

In this section, we give basic knowledge of matrix algebra which are rank, inverse, generalized inverse, positive definite, trace, kronecker product, vec operator and vector and matrix differential with their properties.

2.2.1 Rank

Before defining the rank of a matrix, we introduce the concept of linear independence and dependence. A set of vectors $\boldsymbol{a}_1, \boldsymbol{a}_2, ..., \boldsymbol{a}_n$ is said to be *linearly dependent* if scalars $c_1, c_2, ..., c_n$ (not all zero) can be found such that

$$c_1\boldsymbol{a}_1 + c_2\boldsymbol{a}_2 + \cdots + c_n\boldsymbol{a}_n = \mathbf{0}. \tag{2.8}$$

If no coefficients $c_1, c_2, ..., c_n$ can be found to satisfy (2.8), the set of vectors $\boldsymbol{a}_1, \boldsymbol{a}_2, ..., \boldsymbol{a}_n$ is said to be *linearly independent*.

Definition 2.29. [23] The *rank* of any square or rectangular matrix $\boldsymbol{A}$ is defined as

$$\begin{aligned} \operatorname{rank}(\boldsymbol{A}) &= \text{number of linearly independent columns of } \boldsymbol{A} \\ &= \text{number of linearly independent rows of } \boldsymbol{A}. \end{aligned}$$

Theorem 2.30. [23] *Suppose that a matrix $\boldsymbol{A}$ is $n \times p$ of rank p, where $p < n$. Then $\boldsymbol{A}$ has the maximum possible rank and is said to be of full rank.*

2.2.2 Inverse

Definition 2.31. [23] A full rank square matrix is said to be *nonsingular*. A nonsingular matrix $\boldsymbol{A}$ has a unique *inverse*, denoted by $\boldsymbol{A}^{-1}$, with the property that

$$\boldsymbol{A}\boldsymbol{A}^{-1} = \boldsymbol{A}^{-1}\boldsymbol{A} = \mathbb{I},$$

where $\mathbb{I}$ is the identity matrix.

Theorem 2.32. [23] *If $\boldsymbol{A}$ is nonsingular, then $\boldsymbol{A}^\top$ is nonsingular and its inverse is defined as*

$$(\boldsymbol{A}^\top)^{-1} = (\boldsymbol{A}^{-1})^\top.$$

Theorem 2.33. [23] *If $\boldsymbol{A}$ and $\boldsymbol{B}$ are nonsingular matrices of the same size, then $\boldsymbol{AB}$ is nonsingular and*

$$(\boldsymbol{AB})^{-1} = \boldsymbol{B}^{-1}\boldsymbol{A}^{-1}.$$

2.2.3 Generalized Inverse

We consider a general matrix $\boldsymbol{A}$. A solution of a consistent system of equations $\boldsymbol{A}x = \boldsymbol{c}$ can be expressed in terms of a generalized inverse of $\boldsymbol{A}$.

Definition 2.34. [23] A *generalized inverse* of an $n \times p$ matrix $\boldsymbol{A}$ is any matrix $\boldsymbol{A}^{-}$ that satisfies

$$\boldsymbol{A}\boldsymbol{A}^{-}\boldsymbol{A} = \boldsymbol{A}.$$

If $\boldsymbol{A}$ is invertible, $\boldsymbol{A}^{-} = \boldsymbol{A}^{-1}$. A generalized inverse is also called a *conditional inverse.*

2.2.4 Positive Definite Matrices

In this part, we introduce the knowledge of positive definite that we use in this thesis.

Definition 2.35. [23] If the symmetric matrix $\boldsymbol{A}$ has the property $\boldsymbol{y}^{\top}\boldsymbol{A}\boldsymbol{y} > 0$ for all possible $\boldsymbol{y}$ except $\boldsymbol{y} = \boldsymbol{0}$, then the quadratic form $\boldsymbol{y}^{\top}\boldsymbol{A}\boldsymbol{y}$ is said to be *positive definite*, and $\boldsymbol{A}$ is said to be a *positive definite* matrix. Similarly, if $\boldsymbol{y}^{\top}\boldsymbol{A}\boldsymbol{y} \geq 0$ for all $\boldsymbol{y}$ and there is at least one $\boldsymbol{y} \neq 0$ such that $\boldsymbol{y}^{\top}\boldsymbol{A}\boldsymbol{y} = 0$, then $\boldsymbol{y}^{\top}\boldsymbol{A}\boldsymbol{y}$ and $\boldsymbol{A}$ are said to be *positive semidefinite.*

Theorem 2.36. [23]

1. *If $\boldsymbol{A}$ is positive definite, then all its diagonal element a_{ii} are positive.*

2. *If $\boldsymbol{A}$ is positive semidefinite, then all $a_{ii} \geq 0$.*

Theorem 2.37. [23] *Let $\boldsymbol{P}$ be a nonsingular matrix*

1. *If* $\boldsymbol{A}$ *is positive definite, then* $\boldsymbol{P}^\top \boldsymbol{A}\boldsymbol{P}$ *is positive definite.*

2. *If* $\boldsymbol{A}$ *is positive semidefinite, then* $\boldsymbol{P}^\top \boldsymbol{A}\boldsymbol{P}$ *is positive semidefinite.*

Corollary 2.38. [23] *A positive definite matrix is nonsingular.*

Theorem 2.39. [23] *If* $\boldsymbol{A}$ *is positive definite, then* $\boldsymbol{A}^{-1}$ *is positive definite.*

2.2.5 Trace

Definition 2.40. [23] The trace of an $n \times n$ matrix $\boldsymbol{A} = (a_{ij})$ is a scalar function defined as the sum of the diagonal elements of $\boldsymbol{A}$; that is, $\text{trace}(\boldsymbol{A}) = \sum_{i=1}^{n} a_{ii}$.

Theorem 2.41 below lists some important properties of trace which will be used in this thesis.

Theorem 2.41. 1. *If* $\boldsymbol{A}$ *and* $\boldsymbol{B}$ *are* $n \times n$ *matrix, then*

$$\text{trace}(\boldsymbol{A} \pm \boldsymbol{B}) = \text{trace}(\boldsymbol{A}) \pm \text{trace}(\boldsymbol{B}).$$

2. *If* $\boldsymbol{A}$ *is an* $n \times p$ *matrix and* $\boldsymbol{B}$ *is a* $p \times n$ *matrix, then*

$$\text{trace}(\boldsymbol{A}\boldsymbol{B}) = \text{trace}(\boldsymbol{B}\boldsymbol{A}).$$

3. *If* $\boldsymbol{A}$ *and* $\boldsymbol{B}$ *are the same size matrix, then*

$$[\text{trace}(\boldsymbol{A}\boldsymbol{B})]^2 \leq [\text{trace}(\boldsymbol{A})^2][\text{trace}(\boldsymbol{B})^2].$$

4. *If* $\boldsymbol{A}$ *is any* $n \times n$ *matrix and* $\boldsymbol{P}$ *is any* $n \times n$ *nonsingular matrix, then*

$$\text{trace}(\boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P}) = \text{trace}(\boldsymbol{A}).$$

5. *If* $\boldsymbol{A}$ *is any* $n \times n$ *matrix and* $\boldsymbol{C}$ *is any* $n \times n$ *orthogonal matrix, then*

$$\text{trace}(\boldsymbol{C}^\top \boldsymbol{A}\boldsymbol{C}) = \text{trace}(\boldsymbol{A}).$$

6. *If $\boldsymbol{A}$ is an $n \times p$ matrix of rank r and $\boldsymbol{A}^-$ is a generalized inverse of $\boldsymbol{A}$, then*

$$\text{trace}(\boldsymbol{A}^-\boldsymbol{A}) = \text{trace}(\boldsymbol{A}\boldsymbol{A}^-) = r.$$

2.2.6 Kronecker Product

The kronecker product transforms two matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ into a matrix.

Definition 2.42. [17] If $\boldsymbol{A}$ is an $m \times m$ matrix, $\boldsymbol{B}$ is a $p \times q$ matrix, then the knorecker product of $\boldsymbol{A}$ and $\boldsymbol{B}$ is defined as the matrix

$$\boldsymbol{A} \otimes \boldsymbol{B} = \begin{bmatrix} a_{11}\boldsymbol{B} & a_{12}\boldsymbol{B} & \cdots & a_{1n}\boldsymbol{B} \\ a_{21}\boldsymbol{B} & a_{22}\boldsymbol{B} & \cdots & a_{2n}\boldsymbol{B} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1}\boldsymbol{B} & a_{m2}\boldsymbol{B} & \cdots & a_{mn}\boldsymbol{B} \end{bmatrix}$$

which is an $mp \times nq$ matrix.

Example 2.43. Let $\boldsymbol{A} = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}$ and $\boldsymbol{B} = \begin{bmatrix} 1 & 2 & 1 \\ 2 & 1 & 2 \end{bmatrix}$. Then

$$\boldsymbol{A} \otimes \boldsymbol{B} = \begin{bmatrix} 1\boldsymbol{B} & 2\boldsymbol{B} & 3\boldsymbol{B} \\ 4\boldsymbol{B} & 5\boldsymbol{B} & 6\boldsymbol{B} \end{bmatrix} = \begin{bmatrix} 1 & 2 & 1 & 2 & 4 & 2 & 3 & 6 & 3 \\ 2 & 1 & 2 & 4 & 2 & 4 & 6 & 3 & 6 \\ 4 & 8 & 4 & 5 & 10 & 5 & 6 & 12 & 6 \\ 8 & 4 & 8 & 10 & 5 & 10 & 12 & 6 & 12 \end{bmatrix}.$$

Note that it is not true in general that $\boldsymbol{A} \otimes \boldsymbol{B} \neq \boldsymbol{B} \otimes \boldsymbol{A}$.

Theorem 2.44. [17] *If $\boldsymbol{A}$ is an $m \times n$ matrix, $\boldsymbol{B}$ is an $r \times s$ matrix, $\boldsymbol{C}$ is an $n \times p$ matrix, $\boldsymbol{D}$ is an $s \times t$ matrix and α is a scalar, then*

1. $\boldsymbol{A} \otimes \boldsymbol{B} \otimes \boldsymbol{C} = (\boldsymbol{A} \otimes \boldsymbol{B}) \otimes \boldsymbol{C} = \boldsymbol{A} \otimes (\boldsymbol{B} \otimes \boldsymbol{C})$,

2. $(\boldsymbol{A} + \boldsymbol{B}) \otimes (\boldsymbol{C} + \boldsymbol{D}) = \boldsymbol{A} \otimes \boldsymbol{C} + \boldsymbol{A} \otimes \boldsymbol{D} + \boldsymbol{B} \otimes \boldsymbol{C} + \boldsymbol{B} \otimes \boldsymbol{D}$,

3. $(\boldsymbol{A} \otimes \boldsymbol{B})(\boldsymbol{C} \otimes \boldsymbol{D}) = (\boldsymbol{AC} \otimes \boldsymbol{BD})$,

4. $\alpha \otimes \boldsymbol{A} = \alpha\boldsymbol{A} = \boldsymbol{A}\alpha = \boldsymbol{A} \otimes \alpha$.

Theorem 2.45. [17] *For all* $\boldsymbol{A}$ *and* $\boldsymbol{B}$, $(\boldsymbol{A} \otimes \boldsymbol{B})^\top = \boldsymbol{A}^\top \otimes \boldsymbol{B}^\top$.

Corollary 2.46. [17] *If square matrices* $\boldsymbol{A}$ *and* $\boldsymbol{B}$ *are symmetric, then* $\boldsymbol{A} \otimes \boldsymbol{B}$ *is symmetric.*

Theorem 2.47. [17] *If* $\boldsymbol{A}$ *and* $\boldsymbol{B}$ *are square matrices,* $\text{trace}(\boldsymbol{A}\otimes\boldsymbol{B}) = (\text{trace}\boldsymbol{A})(\text{trace}\boldsymbol{B})$.

Theorem 2.48. [17] *Let* $\boldsymbol{A}$ *and* $\boldsymbol{B}$ *are nonsingular,* $(\boldsymbol{A} \otimes \boldsymbol{B})^{-1} = \boldsymbol{A}^{-1} \otimes \boldsymbol{B}^{-1}$.

2.2.7 Vec Operator

The vec operator transforms a matrix into a vector by stacking its columns one below the other.

Definition 2.49. [17] Let $\boldsymbol{A}$ be an $m \times n$ matrix,

$$\boldsymbol{A} = \begin{bmatrix} \vdots & \vdots & & \vdots \\ \mathbf{a}_1 & \mathbf{a}_2 & \cdots & \mathbf{a}_n \\ \vdots & \vdots & & \vdots \end{bmatrix},$$

where $\mathbf{a}_i;\ i = 1, 2, ..., n$ is an $m \times 1$ vector, then vec $\boldsymbol{A}$ is the $mn \times 1$ vector,

$$\text{vec } \boldsymbol{A} = \begin{bmatrix} \mathbf{a}_1 \\ \mathbf{a}_2 \\ \vdots \\ \mathbf{a}_n \end{bmatrix}.$$

Note that vec $\boldsymbol{A}$ = vec $\boldsymbol{B}$ does not imply $\boldsymbol{A} = \boldsymbol{B}$.

Theorem 2.50. [17] *For any column vector* $\boldsymbol{a}$, $\text{vec } \boldsymbol{a}^\top = \text{vec } \boldsymbol{a} = \boldsymbol{a}$.

Theorem 2.51. [17] *For any two column vectors* $\boldsymbol{a}$ *and* $\boldsymbol{b}$, $\text{vec } \boldsymbol{a}\boldsymbol{b}^\top = \boldsymbol{b} \otimes \boldsymbol{a}$.

Theorem 2.52. [17] *For* $\boldsymbol{A}$ *and* $\boldsymbol{B}$ *are matrices of the same order,* $(\text{vec } \boldsymbol{A})^\top \text{vec } \boldsymbol{B} = \text{trace}(\boldsymbol{A}^\top \boldsymbol{B})$.

Theorem 2.53. [17] *Let* $\boldsymbol{A}$, $\boldsymbol{B}$ *and* $\boldsymbol{C}$ *be three metrics such that the matrix product* $\boldsymbol{ABC}$ *is defined. Then*

$$\text{vec } \boldsymbol{ABC} = (\boldsymbol{C}^\top \otimes \boldsymbol{A}) \text{ vec } \boldsymbol{B}. \tag{2.9}$$

Theorem 2.54. [17] *Let* $\boldsymbol{A}$, $\boldsymbol{B}$, $\boldsymbol{C}$ *and* $\boldsymbol{D}$ *be four matrices such that the matrix product* $\boldsymbol{ABCD}$ *is defined and square. Then*

$$\text{trace } (\boldsymbol{ABCD}) = (\text{vec } \boldsymbol{D}^\top)^\top (\boldsymbol{C}^\top \otimes \boldsymbol{A}) \text{ vec } \boldsymbol{B} \tag{2.10}$$

$$= (\text{vec } \boldsymbol{D})^\top (\boldsymbol{A} \otimes \boldsymbol{C}^\top) \text{ vec } \boldsymbol{B}^\top. \tag{2.11}$$

2.2.8 Vector and Matrix Differentials

In this part, let $\boldsymbol{x}$ and $\boldsymbol{y}$ be vectors of orders m and n, respectively,

$$\boldsymbol{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{bmatrix}, \quad \boldsymbol{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix},$$

where each component y_i may be a function of the x_j, or $y = y(x)$.

The derivative of the vector $\boldsymbol{y}$ with respect to vector $\boldsymbol{x}$ is the $m \times n$ matrix,

$$\frac{d\boldsymbol{y}}{d\boldsymbol{x}} = \begin{bmatrix} \frac{dy_1}{dx_1} & \frac{dy_2}{dx_1} & \cdots & \frac{dy_n}{dx_1} \\ \frac{dy_1}{dx_2} & \frac{dy_2}{dx_2} & \cdots & \frac{dy_n}{dx_2} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{dy_1}{dx_m} & \frac{dy_2}{dx_m} & \cdots & \frac{dy_n}{dx_m} \end{bmatrix}.$$

The derivative of a scalar c with respect to vector is a vector,

$$\frac{dc}{d\boldsymbol{x}} = \begin{bmatrix} \frac{dc}{dx_1} \\ \frac{dc}{dx_2} \\ \vdots \\ \frac{dc}{dx_m} \end{bmatrix}.$$

The derivative of a vector with respect to scalar c is a vector,

$$\frac{d\boldsymbol{y}}{dc} = \begin{bmatrix} \frac{dy_1}{dc} & \frac{dy_2}{dc} & \cdots & \frac{dy_n}{dc} \end{bmatrix}.$$

The summary of differentiation results is presented in Facker [11]. We state those properties as follows. Let $\boldsymbol{A}$ be an $n \times m$, $\boldsymbol{B}$ be a $p \times q$, $\boldsymbol{x}$ be an $m \times 1$, and $\boldsymbol{X}$ be defined comformably. The result shows as following

1. $$[Df]_{ij} = \frac{df_i(\boldsymbol{x})}{d\boldsymbol{x}_j}. \tag{2.12}$$

2. $$\frac{d\boldsymbol{A}\boldsymbol{x}}{d\boldsymbol{x}} = \boldsymbol{A}. \tag{2.13}$$

3. $$D[\alpha f(\boldsymbol{x}) + \beta g(\boldsymbol{x})] = \alpha Df(\boldsymbol{x}) + \beta Dg(\boldsymbol{x}). \tag{2.14}$$

4. $$D[f(g(\boldsymbol{x}))] = f'(g(\boldsymbol{x}))g'(\boldsymbol{x}). \tag{2.15}$$

5. $$D[f(\boldsymbol{x})g(\boldsymbol{x})] = (g(\boldsymbol{x})^\top \otimes \mathbb{I}_m)f'(\boldsymbol{x}) + (\mathbb{I}_p \otimes f(\boldsymbol{x}))g'(\boldsymbol{x}). \tag{2.16}$$

6. $$\frac{d\boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x}}{d\boldsymbol{x}} = \boldsymbol{x}^\top(\boldsymbol{A} + \boldsymbol{A}^\top). \tag{2.17}$$

7.
$$\frac{d\text{vec}(\boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x})}{d\text{vec}(\boldsymbol{A})} = (\boldsymbol{x}^\top \otimes \boldsymbol{x}^\top). \tag{2.18}$$

8.
$$\frac{d\boldsymbol{x}^\top \boldsymbol{A}^\top \boldsymbol{A}\boldsymbol{x}}{d\boldsymbol{A}} = 2\boldsymbol{x}^\top \otimes \boldsymbol{x}^\top \boldsymbol{A}^\top. \tag{2.19}$$

9.
$$\frac{d\boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{A}^\top \boldsymbol{x}}{d\boldsymbol{A}} = 2\boldsymbol{x}^\top \boldsymbol{A} \otimes \boldsymbol{x}^\top. \tag{2.20}$$

10.
$$\frac{d\boldsymbol{A}\boldsymbol{X}\boldsymbol{B}}{d\boldsymbol{X}} = \boldsymbol{B}^\top \otimes \boldsymbol{A}. \tag{2.21}$$

11.
$$\frac{d\boldsymbol{A}^{-1}}{d\boldsymbol{A}} = -(\boldsymbol{A}^{-\top} \otimes \boldsymbol{A}^{-1}). \tag{2.22}$$

12.
$$\frac{d\ln|\boldsymbol{A}|}{d\boldsymbol{A}} = \text{vec}(\boldsymbol{A}^{-\top})^\top. \tag{2.23}$$

13.
$$\frac{d\text{trace}(\boldsymbol{A}\boldsymbol{X})}{d\boldsymbol{X}} = \text{vec}(\boldsymbol{A}^\top)^\top. \tag{2.24}$$

2.2.9 Random Vectors and Matrices

In this thesis, we are interested in linear models, it is convenient to express the observed data in form a vector or a matrix. A *random vector* or a *random matrix* is a vector or matrix whose elements are random variables.

Mean Vectors

The expected value of a $p \times 1$ random vector $\boldsymbol{y}$ is defined as the vector of expected value of the p random variables $y_1, y_2, ..., y_p$ in $\boldsymbol{y}$:

$$E(\boldsymbol{y}) = E \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_p \end{bmatrix} = \begin{bmatrix} E(y_1) \\ E(y_2) \\ \vdots \\ E(y_p) \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_p \end{bmatrix} = \boldsymbol{\mu}, \tag{2.25}$$

where $E(y_i) = \mu_i$ $(i = 1, 2, ..., p)$.

Theorem 2.55. [23] *If $\boldsymbol{x}$ and $\boldsymbol{y}$ are $p \times 1$ random vectors, the expected value of their sum is the sum of their expected values*

$$E(\boldsymbol{x} + \boldsymbol{y}) = E(\boldsymbol{x}) + E(\boldsymbol{y}). \tag{2.26}$$

Covariance Matrix

The variances $\sigma_1^2, \sigma_2^2, ..., \sigma_p^2$ of $y_1, y_2, ..., y_p$ and the covariances σ_{ij} for all $i \neq j$ can be conveniently displayed in the *covariance matrix*, which is denoted by $\boldsymbol{\Sigma}$,

$$\boldsymbol{\Sigma} = cov(\boldsymbol{y}) = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2p} \\ \vdots & \vdots & & \vdots \\ \sigma_{p1} & \sigma_{p2} & \cdots & \sigma_{pp} \end{bmatrix}.$$

Theorem 2.56. [23] *If $\boldsymbol{a}$ is a $p \times 1$ vector of constants and $\boldsymbol{y}$ is a $p \times 1$ random vector with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$, then*

1. $E(\boldsymbol{a}^\top \boldsymbol{y}) = \boldsymbol{a}^\top \boldsymbol{\mu}$, *and*

2. $Var(\boldsymbol{a}^\top \boldsymbol{y}) = \boldsymbol{a}^\top \boldsymbol{\Sigma} \boldsymbol{a}$.

2.3 Basic Knowledge of Linear Regression Model

In this section, we are interested in multiple linear regression model, we attempt to predict a dependent variable or response variable y on the basis of an assumed linear relationship with several independent or predictor variables $x_1, x_2, ..., x_K$.

2.3.1 The Model

The multiple linear regression model can be expressed as

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_K x_K + \epsilon. \tag{2.27}$$

To estimate the β's in (2.27), we will use a sample of N observations on y and the associated x variables. The model for the ith observation is

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_K x_{iK} + \epsilon_i, \quad i = 1, 2, ..., N. \tag{2.28}$$

To complete the model in (2.27), we make the following additional assumptions:

1. $E(\epsilon_i) = 0$ for $i = 1, 2, ..., N$.

2. $Var(\epsilon_i) = \sigma^2$ for $i = 1, 2, ..., N$.

3. $Cov(\epsilon_i, \epsilon_j) = 0$ for all $i \neq j$.

For the N observations, we can be written in matrix form as

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1K} \\ 1 & x_{21} & x_{22} & \cdots & x_{2K} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_{N1} & x_{N2} & \cdots & x_{NK} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_K \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_N \end{bmatrix}$$

or

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}. \tag{2.29}$$

The proceeding two assumptions on ϵ_i or y_i can be expressed in terms of the model in (2.29):

1. $E(\boldsymbol{\epsilon}) = 0$ or $E(\boldsymbol{y}) = \boldsymbol{X\beta}$.

2. $Cov(\boldsymbol{\epsilon}) = \sigma^2\mathbb{I}$ or $Cov(\boldsymbol{y}) = \sigma^2\mathbb{I}$.

2.3.2 Estimations of $\boldsymbol{\beta}$ and σ^2

We study the least squares approach to estimation of the β's in the model (2.29). For parameters $\beta_0, \beta_1, ..., \beta_K$, we seek $\hat{\beta}_0, \hat{\beta}_1, ..., \hat{\beta}_K$ that minimize

$$\begin{aligned} \sum_{i=1}^{N} \hat{\epsilon}_i^2 &= \sum_{i=1}^{N} (y_i - \hat{y}_i)^2. \\ &= \sum_{i=1}^{N} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \hat{\beta}_2 x_{i2} \cdots \hat{\beta_K} x_{iK})^2. \end{aligned} \tag{2.30}$$

To find the values of $\hat{\beta}_0, \hat{\beta}_1, ..., \hat{\beta}_K$ that minimize (2.30), we could differentiate $\sum_{i=1}^{N} \hat{\epsilon}_i^2$ with respect to each $\hat{\beta}_j$ and set results equal to zero. The result is given in the following theorem.

Theorem 2.57. [23] *If $\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{\epsilon}$, where $\boldsymbol{X}$ is $N \times (K+1)$ of rank $(K+1) < N$, then the value of $\hat{\boldsymbol{\beta}} = (\hat{\beta}_0, \hat{\beta}_1, ..., \hat{\beta}_K)^\top$ that minimizes* (2.30) *is*

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{y}.$$

Properties 2.58. [23] *If $E(\boldsymbol{y}) = \boldsymbol{X\beta}$, then $\hat{\boldsymbol{\beta}}$ is an unbiased estimator for $\boldsymbol{\beta}$.*

The method of least square does not yield a function of the $\boldsymbol{y}$ and $\boldsymbol{X}$ values in the sample that we can minimize to obtain an estimator of σ^2. However, we can devise an unbiased estimator for σ^2 based on the least-square estimator $\hat{\boldsymbol{\beta}}$. By assumption $Cov(\boldsymbol{y}) = \sigma^2\mathbb{I}$, σ^2 is the same for each y_i, $(i = 1, 2, ..., n)$. σ^2 is defined by $\sigma^2 = E[y_i - E(y_i)]^2$, and by assumption, $E(y_i) = \boldsymbol{x}_i^\top \boldsymbol{\beta}$, we obtain

$$\sigma^2 = E[y_i - \boldsymbol{x}_i^\top \boldsymbol{\beta}]^2,$$

where $\boldsymbol{x}_i^\top$ is the ith row of $\boldsymbol{X}$.

We estimate σ^2 by a corresponding average from the sample

$$s^2 = \frac{1}{N-K-1}\sum_{i=1}^{N}(y_i - \boldsymbol{x}_i^\top\hat{\boldsymbol{\beta}})^2. \tag{2.31}$$

$$= \frac{SSE}{N-K-1}, \tag{2.32}$$

where N is the sample size and K is the number of x's.

Theorem 2.59. [23] *If s^2 is defined by* (2.31) *or* (2.32) *and if $E(\boldsymbol{y}) = \boldsymbol{X\beta}$ and $Cov(\boldsymbol{y}) = \sigma^2\mathbb{I}$, then s^2 is an unbiased estimator of σ^2.*

2.4 Basic Knowledge of Panel Data Model

In this section, we are interested in panel data. Classical panel data refer to two-dimensional data. The data have both cross-sectional and time-series directions. We provide multiple observations on each individual in the sample.

In this section, we introduce a linear panel data model. Commonly, the model used to assess the effects of both time and individual factors is written as

$$y_{it} = \alpha_{it} + \boldsymbol{x}_{it}^\top\boldsymbol{\beta}_{it} + \epsilon_{it} \qquad i = 1, ..., N, \quad t = 1, ..., T,$$

where N is the number of cross-sectional unit, T is the number of time unit, α_{it} is an 1×1 vector of intercept, $\boldsymbol{\beta}_{it} = (\beta_{1it}, \beta_{2it}, ..., \beta_{Kit})^\top$ is a $K \times 1$ vector of slope coefficient parameters that vary across i and t, $\boldsymbol{x}_{it} = (x_{1it}, x_{2it}, ..., x_{Kit})$ is an $1\times K$ vector of covariate variables and ϵ_{it} is the error term.

In this research, we assume that regression parameters are constant overtime, but can vary across individuals. The model of interest is written as

$$y_{it} = \alpha_i + \boldsymbol{x}_{it}^\top\boldsymbol{\beta}_i + \epsilon_{it}. \tag{2.33}$$

Three types of restrictions can be imposed on the model in (2.33):

1. Regression slope coefficients are constant but the intercepts varies over individuals,

$$y_{it} = \alpha_i + \boldsymbol{x}_{it}^\top \boldsymbol{\beta} + \epsilon_{it}. \tag{2.34}$$

2. The intercepts are constant but regression slope coefficients varies over individuals,

$$y_{it} = \alpha + \boldsymbol{x}_{it}^\top \boldsymbol{\beta}_i + \epsilon_{it}.$$

3. Both slope and intercept coefficients are constant,

$$y_{it} = \alpha + \boldsymbol{x}_{it}^\top \boldsymbol{\beta} + \epsilon_{it}.$$

There are two types of panel data model as follows.

Definition 2.60. [13] A *homogeneous* panel data model is a model in which both intercept and slope coefficients are constant.

Definition 2.61. [13] A *heterogeneous* panel data model is a model in which both intercept and slope coefficients vary across individuals.

In this research, we study in model (2.34), the slope heterogeneity in panel data model, and write in vector form as following:

$$\begin{bmatrix} \boldsymbol{y}_1 \\ \boldsymbol{y}_2 \\ \vdots \\ \boldsymbol{y}_N \end{bmatrix} = \begin{bmatrix} \boldsymbol{\iota} \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \end{bmatrix} \alpha_1 + \begin{bmatrix} \mathbf{0} \\ \boldsymbol{\iota} \\ \vdots \\ \mathbf{0} \end{bmatrix} \alpha_2 + \cdots + \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \\ \vdots \\ \boldsymbol{\iota} \end{bmatrix} \alpha_N + \begin{bmatrix} \boldsymbol{X}_1 \\ \boldsymbol{X}_2 \\ \vdots \\ \boldsymbol{X}_N \end{bmatrix} \boldsymbol{\beta} + \begin{bmatrix} \boldsymbol{\epsilon}_1 \\ \boldsymbol{\epsilon}_2 \\ \vdots \\ \boldsymbol{\epsilon}_N \end{bmatrix} \tag{2.35}$$

or

$$\boldsymbol{y}_i = \alpha_i \boldsymbol{\iota}_T + \boldsymbol{X}_i \boldsymbol{\beta} + \boldsymbol{\epsilon}_i \quad i = 1, 2, ..., N \tag{2.36}$$

where

$$\boldsymbol{y}_i = \begin{bmatrix} \boldsymbol{y}_{i1} \\ \boldsymbol{y}_{i2} \\ \vdots \\ \boldsymbol{y}_{iT} \end{bmatrix}, \quad \boldsymbol{X}_i = \begin{bmatrix} x_{i11} & x_{i12} & \cdots & x_{i1K} \\ x_{i21} & x_{i22} & \cdots & x_{i2K} \\ \vdots & \vdots & & \vdots \\ x_{iT1} & x_{iT2} & \cdots & x_{iTK} \end{bmatrix},$$

$$\boldsymbol{\beta}^\top = (\beta_1, \beta_2, ..., \beta_K), \quad , \boldsymbol{\epsilon}_i^\top = (\epsilon_{i1}, ..., \epsilon_{iT}),$$

$\boldsymbol{\iota}_T$ denotes the $T \times 1$ vector of ones and $\mathbb{I}_T$ denotes the $T \times T$ identity matrix. In this work, we assume that all $\boldsymbol{X}_i$ are full-rank.

The proceeding three assumptions on ϵ_{it} can be expressed in terms of the panel data model in (2.36):

1. $E(\boldsymbol{\epsilon}_i) = \mathbf{0}$ or $E(\boldsymbol{y}_i) = \alpha_i \boldsymbol{\iota}_T + \boldsymbol{X}_i \boldsymbol{\beta}$.

2. $E(\boldsymbol{\epsilon}_i \boldsymbol{\epsilon}_i^\top) = \sigma_\epsilon^2 \mathbb{I}_T$.

3. $E(\boldsymbol{\epsilon}_i \boldsymbol{\epsilon}_j^\top) = \mathbf{0}$ if $i \neq j$.

2.4.1 Estimations of $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, and σ_ϵ^2

We know that the least square estimator of (2.36) is the best linear unbiased estimator (BLUE). To find the regression coefficient estimators in panel data model, define $\boldsymbol{N}_0 = \mathbb{I}_T - \dfrac{\boldsymbol{\iota}_T \boldsymbol{\iota}_T^\top}{T}$, where properties of $\boldsymbol{N}_0$ are listed in the following proposition.

Proposition 2.62. *1)* $\boldsymbol{N}_0 \boldsymbol{\iota}_T = 0$.
2) $\boldsymbol{N}_0^\top \boldsymbol{N}_0 = \boldsymbol{N}_0$.

Proof. 1)

$$\begin{aligned} \boldsymbol{N}_0 \boldsymbol{\iota}_T &= \left(\mathbb{I}_T - \frac{\boldsymbol{\iota}_T \boldsymbol{\iota}_T^\top}{T} \right) \boldsymbol{\iota}_T \\ &= \boldsymbol{\iota}_T - \frac{\boldsymbol{\iota}_T \boldsymbol{\iota}_T^\top \boldsymbol{\iota}_T}{T} \end{aligned}$$

$$= \boldsymbol{\iota}_T - \boldsymbol{\iota}_T$$
$$= 0,$$

where we use the fact that $\boldsymbol{\iota}_T^\top \boldsymbol{\iota}_T = T$.

2)

$$\begin{aligned}
\boldsymbol{N}_0^\top \boldsymbol{N}_0 &= \left(\mathbb{I}_T - \frac{\boldsymbol{\iota}_T \boldsymbol{\iota}_T^\top}{T}\right)^\top \left(\mathbb{I}_T - \frac{\boldsymbol{\iota}_T \boldsymbol{\iota}_T^\top}{T}\right) \\
&= \left(\mathbb{I}_T - \frac{\boldsymbol{\iota}_T \boldsymbol{\iota}_T^\top}{T}\right)\left(\mathbb{I}_T - \frac{\boldsymbol{\iota}_T \boldsymbol{\iota}_T^\top}{T}\right) \\
&= \mathbb{I}_T - \frac{\boldsymbol{\iota}_T \boldsymbol{\iota}_T^\top}{T} - \frac{\boldsymbol{\iota}_T \boldsymbol{\iota}_T^\top}{T} + \frac{\boldsymbol{\iota}_T \boldsymbol{\iota}_T^\top}{T}\frac{\boldsymbol{\iota}_T \boldsymbol{\iota}_T^\top}{T} \\
&= \mathbb{I}_T - \frac{\boldsymbol{\iota}_T \boldsymbol{\iota}_T^\top}{T} - \frac{\boldsymbol{\iota}_T \boldsymbol{\iota}_T^\top}{T} + \frac{\boldsymbol{\iota}_T T \boldsymbol{\iota}_T^\top}{TT} \\
&= \mathbb{I}_T - \frac{\boldsymbol{\iota}_T \boldsymbol{\iota}_T^\top}{T} - \frac{\boldsymbol{\iota}_T \boldsymbol{\iota}_T^\top}{T} + \frac{\boldsymbol{\iota}_T \boldsymbol{\iota}_T^\top}{T} \\
&= \mathbb{I}_T - \frac{\boldsymbol{\iota}_T \boldsymbol{\iota}_T^\top}{T} \\
&= \boldsymbol{N}_0,
\end{aligned}$$

where, similar to 1), we use the fact that $\boldsymbol{\iota}_T^\top \boldsymbol{\iota}_T = T$.

□

Multiplying $\boldsymbol{N}_0$ to (2.36), from Proposition 2.62(1), we obtain

$$\begin{aligned}
\boldsymbol{N}_0 \boldsymbol{y}_i &= \boldsymbol{N}_0 \alpha_i \boldsymbol{\iota}_T + \boldsymbol{N}_0 \boldsymbol{X}_i \boldsymbol{\beta} + \boldsymbol{N}_0 \boldsymbol{\epsilon}_i \\
&= \boldsymbol{M}_0 \boldsymbol{X}_i \boldsymbol{\beta} + \boldsymbol{N}_0 \boldsymbol{\epsilon}_i. \qquad (2.37)
\end{aligned}$$

The ordinary least square estimator $\hat{\boldsymbol{\beta}}$ of $\boldsymbol{\beta}$ is obtained by minimizing the sum of squared errors as the following

$$\begin{aligned}
SSE_\beta &= \sum_{i=1}^{N} (\boldsymbol{N}_0 \boldsymbol{\epsilon}_i)^\top (\boldsymbol{N}_0 \boldsymbol{\epsilon}_i) \\
&= \sum_{i=1}^{N} (\boldsymbol{N}_0 \boldsymbol{y}_i - \boldsymbol{N}_0 \boldsymbol{X}_i \boldsymbol{\beta})^\top (\boldsymbol{N}_0 \boldsymbol{y}_i - \boldsymbol{N}_0 \boldsymbol{X}_i \boldsymbol{\beta})
\end{aligned}$$

$$
\begin{aligned}
&= \sum_{i=1}^{N} (\boldsymbol{N}_0\boldsymbol{y}_i)^\top \boldsymbol{N}_0\boldsymbol{y}_i - (\boldsymbol{N}_0\boldsymbol{y}_i)^\top \boldsymbol{N}_0\boldsymbol{X}_i\boldsymbol{\beta} - (\boldsymbol{N}_0\boldsymbol{X}_i\boldsymbol{\beta})^\top \boldsymbol{N}_0\boldsymbol{y}_i + (\boldsymbol{N}_0\boldsymbol{X}_i\boldsymbol{\beta})^\top \boldsymbol{N}_0\boldsymbol{X}_i\boldsymbol{\beta} \\
&= \sum_{i=1}^{N} \boldsymbol{y}_i^\top \boldsymbol{N}_0\boldsymbol{y}_i - \boldsymbol{y}_i^\top \boldsymbol{N}_0(\boldsymbol{X}_i\boldsymbol{\beta}) - (\boldsymbol{X}_i\boldsymbol{\beta})^\top \boldsymbol{N}_0\boldsymbol{y}_i + (\boldsymbol{X}_i\boldsymbol{\beta})^\top \boldsymbol{N}_0(\boldsymbol{X}_i\boldsymbol{\beta}) \qquad (2.38) \\
&= \sum_{i=1}^{N} \boldsymbol{y}_i^\top \boldsymbol{N}_0\boldsymbol{y}_i - 2\boldsymbol{y}_i^\top \boldsymbol{N}_0(\boldsymbol{X}_i\boldsymbol{\beta}) + (\boldsymbol{X}_i\boldsymbol{\beta})^\top \boldsymbol{N}_0(\boldsymbol{X}_i\boldsymbol{\beta}),
\end{aligned}
$$

where we use Proposition 2.62(2) to obtain (2.38).

Taking partial derivative of SSE_β with respect to $\boldsymbol{\beta}$,

$$\frac{\partial SSE_\beta}{\partial \boldsymbol{\beta}} = \sum_{i=1}^{N} -2\boldsymbol{X}_i^\top \boldsymbol{N}_0\boldsymbol{y}_i + 2\boldsymbol{X}_i^\top \boldsymbol{N}_0\boldsymbol{X}_i\boldsymbol{\beta}.$$

Since $\left.\frac{\partial SSE_\beta}{\partial \boldsymbol{\beta}}\right|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}} = 0, \; -2\sum_{i=1}^{N} \boldsymbol{X}_i^\top \boldsymbol{N}_0\boldsymbol{y}_i + 2\hat{\boldsymbol{\beta}}\sum_{i=1}^{N} \boldsymbol{X}_i^\top \boldsymbol{N}_0\boldsymbol{X}_i = 0.$

Therefore, the least square estimator $\hat{\boldsymbol{\beta}}$ is

$$\hat{\boldsymbol{\beta}} = \left(\sum_{i=1}^{N} \boldsymbol{X}_i^\top \boldsymbol{N}_0\boldsymbol{X}_i\right)^{-1} \sum_{i=1}^{N} \boldsymbol{X}_i^\top \boldsymbol{N}_0\boldsymbol{y}_i.$$

Theorem 2.63. [10] *If the matrix*

$$\frac{1}{N}\sum_{i=1}^{N} \boldsymbol{X}_i^\top \boldsymbol{N}_0\boldsymbol{X}_i$$

is positive definite and converges to a non-stochastic positive definite matrix in the limit, then the least squares estimator $\hat{\boldsymbol{\beta}}$ for $\boldsymbol{\beta}$ in the model (2.36) is consistent.

Proposition 2.64. *The least square estimator $\hat{\boldsymbol{\beta}}$ is an unbiased estimator for $\boldsymbol{\beta}$.*

Proof.

$$
\begin{aligned}
E[\hat{\boldsymbol{\beta}}] &= E\left[\left(\sum_{i=1}^{N} \boldsymbol{X}_i^\top \boldsymbol{N}_0\boldsymbol{X}_i\right)^{-1} \sum_{i=1}^{N} \boldsymbol{X}_i^\top \boldsymbol{N}_0\boldsymbol{y}_i\right] \\
&= \left(\sum_{i=1}^{N} \boldsymbol{X}_i^\top \boldsymbol{N}_0\boldsymbol{X}_i\right)^{-1} \sum_{i=1}^{N} \boldsymbol{X}_i^\top \boldsymbol{N}_0 E[\boldsymbol{y}_i]
\end{aligned}
$$

$$= \left(\sum_{i=1}^{N} \boldsymbol{X}_i^\top \boldsymbol{N}_0 \boldsymbol{X}_i\right)^{-1} \sum_{i=1}^{N} \boldsymbol{X}_i^\top \boldsymbol{N}_0 [\alpha_i \boldsymbol{\iota}_T + \boldsymbol{X}_i \boldsymbol{\beta}]$$

$$= \left(\sum_{i=1}^{N} \boldsymbol{X}_i^\top \boldsymbol{N}_0 \boldsymbol{X}_i\right)^{-1} \sum_{i=1}^{N} \boldsymbol{X}_i^\top \boldsymbol{N}_0 \boldsymbol{X}_i \boldsymbol{\beta}$$

$$= \boldsymbol{\beta}.$$

□

The ordinary least square estimator of α_i is obtained by minimizing the sum of squared errors as the following

$$\begin{aligned} SSE_\alpha &= \sum_{i=1}^{N} \boldsymbol{\epsilon}_i^\top \boldsymbol{\epsilon}_i \\ &= \sum_{i=1}^{N} (\boldsymbol{y}_i - \alpha_i \boldsymbol{\iota}_T - \boldsymbol{X}_i \boldsymbol{\beta})^\top (\boldsymbol{y}_i - \alpha_i \boldsymbol{\iota}_T - \boldsymbol{X}_i \boldsymbol{\beta}) \\ &= \sum_{i=1}^{N} (\boldsymbol{y}_i - \boldsymbol{X}_i \boldsymbol{\beta})^\top (\boldsymbol{y}_i - \boldsymbol{X}_i \boldsymbol{\beta}) - (\boldsymbol{y}_i - \boldsymbol{X}_i \boldsymbol{\beta})^\top (\alpha_i \boldsymbol{\iota}_T) - (\alpha_i \boldsymbol{\iota}_T)^\top (\boldsymbol{y}_i - \boldsymbol{X}_i \boldsymbol{\beta}) \\ &\quad + (\alpha_i \boldsymbol{\iota}_T)^\top (\alpha_i \boldsymbol{\iota}_T) \\ &= \sum_{i=1}^{N} (\boldsymbol{y}_i - \boldsymbol{X}_i \boldsymbol{\beta})^\top (\boldsymbol{y}_i - \boldsymbol{X}_i \boldsymbol{\beta}) - 2(\boldsymbol{y}_i - \boldsymbol{X}_i \boldsymbol{\beta})^\top (\alpha_i \boldsymbol{\iota}_T) + (\alpha_i \boldsymbol{\iota}_T)^\top (\alpha_i \boldsymbol{\iota}_T) \\ &= \sum_{i=1}^{N} (\boldsymbol{y}_i - \boldsymbol{X}_i \boldsymbol{\beta})^\top (\boldsymbol{y}_i - \boldsymbol{X}_i \boldsymbol{\beta}) - 2\alpha_i (\boldsymbol{y}_i - \boldsymbol{X}_i \boldsymbol{\beta})^\top \boldsymbol{\iota}_T + \alpha_i^2 \boldsymbol{\iota}_T^\top \boldsymbol{\iota}_T. \end{aligned}$$

Taking partial derivatives of SSE_α with respect to α_i,

$$\frac{\partial SSE_\alpha}{\partial \alpha_i} = -2(\boldsymbol{y}_i - \boldsymbol{X}_i \boldsymbol{\beta})^\top \boldsymbol{\iota}_T + 2\alpha_i \boldsymbol{\iota}_T^\top \boldsymbol{\iota}_T.$$

Since $\left.\dfrac{\partial SSE_\alpha}{\partial \alpha_i}\right|_{\alpha_i = \hat{\alpha}_i} = 0, -2(\boldsymbol{y}_i - \boldsymbol{X}_i \hat{\boldsymbol{\beta}})^\top \boldsymbol{\iota}_T + 2\hat{\alpha}_i \boldsymbol{\iota}_T^\top \boldsymbol{\iota}_T = 0.$

Therefore, the least square estimator $\hat{\alpha}_i$ $(i = 1, 2, ..., N)$ is

$$\hat{\alpha}_i = \frac{(\boldsymbol{y}_i - \boldsymbol{X}_i \hat{\boldsymbol{\beta}})^\top \boldsymbol{\iota}_T}{T}$$

$$= \frac{\boldsymbol{\iota}_T^\top(\boldsymbol{y}_i - \boldsymbol{X}_i\hat{\boldsymbol{\beta}})}{T}.$$

Proposition 2.65. *The least square estimator $\hat{\alpha}_i$ $(i = 1, 2, ..., N)$ is an unbiased estimator for α.*

Proof.

$$\begin{aligned} E[\hat{\alpha}_i] &= E\left[\frac{\boldsymbol{\iota}_T^\top(\boldsymbol{y}_i - \boldsymbol{X}_i\hat{\boldsymbol{\beta}})}{T}\right] \\ &= \frac{\boldsymbol{\iota}_T^\top(E[\boldsymbol{y}_i] - \boldsymbol{X}_i E[\hat{\boldsymbol{\beta}}])}{T} \\ &= \frac{\boldsymbol{\iota}_T^\top(\alpha_i\boldsymbol{\iota}_T + \boldsymbol{X}_i\boldsymbol{\beta} - \boldsymbol{X}_i\boldsymbol{\beta})}{T} \\ &= \alpha_i\frac{\boldsymbol{\iota}_T^\top\boldsymbol{\iota}_T}{T} \\ &= \alpha_i\frac{T}{T} \\ &= \alpha_i. \end{aligned}$$

□

To find an estimator for σ_ϵ^2 based on the unbiased least square estimators $\hat{\alpha}_i$ and $\hat{\boldsymbol{\beta}}$. We estimate σ_ϵ^2 by corresponding average from the sample

$$\begin{aligned} \hat{\sigma_\epsilon}^2 =& \frac{1}{N(T-1)-K}\sum_{i=1}^{N}(\boldsymbol{N}_0\boldsymbol{y}_i - \boldsymbol{N}_0\hat{\alpha}_i\boldsymbol{\iota}_T - \boldsymbol{N}_0\boldsymbol{X}_i\hat{\boldsymbol{\beta}})^\top(\boldsymbol{N}_0\boldsymbol{y}_i - \boldsymbol{N}_0\boldsymbol{X}_i\hat{\boldsymbol{\beta}}) \\ =& \frac{1}{N(T-1)-K}\sum_{i=1}^{N}(\boldsymbol{N}_0\boldsymbol{y}_i - \boldsymbol{N}_0\boldsymbol{X}_i\hat{\boldsymbol{\beta}})^\top(\boldsymbol{N}_0\boldsymbol{y}_i - \boldsymbol{N}_0\boldsymbol{X}_i\hat{\boldsymbol{\beta}}) \\ =& \frac{SSE_\beta}{N(T-1)-K}, \end{aligned}$$

where N is the sample size, T is the time size and K is the number of x's.

CHAPTER III
HYPOTHESIS TESTING FOR SLOPE HETEROGENEITY FOR THREE DIMENSIONAL PANEL DATA

In this chapter, we construct a test statistic for slope heterogeneity for three dimensional panel data models by conditional Lagrange multiplier.

3.1 Model of Slope Heterogeneity

Consider the three dimensional panel data model

$$y_{ijt} = \alpha_i + \gamma_j + \boldsymbol{x}_{ijt}^\top \boldsymbol{\beta}_{ij} + \epsilon_{ijt}, \tag{3.1}$$

where y_{ijt} is the response variable for factor 1 level i $(i = 1, ..., I)$, factor 2 level j $(j = 1, ..., J)$ at time t $(t = 1, ..., T)$, $\boldsymbol{x}_{ijt}$ is a $K \times 1$ vector of covariates, α_i is the effect of factor 1 level i, γ_j is the effect of factor 2 level j, ϵ_{ijt} is an error with zero mean and variance σ_ϵ^2 and $\boldsymbol{\beta}_{ij}$ represents the regression slope coefficients for the (i, j) unit. To allow for potential slope heterogeneity, we suppose that $\boldsymbol{\beta}_{ij}$, a $K \times 1$ vector, is represented by

$$\boldsymbol{\beta}_{ij} = \boldsymbol{\beta} + \boldsymbol{\eta}_i + \boldsymbol{\omega}_j, \tag{3.2}$$

where $\mathrm{E}(\boldsymbol{\eta}_i|\boldsymbol{x}) = \mathbf{0}$ and $\mathrm{E}(\boldsymbol{\omega}_j|\boldsymbol{x}) = \mathbf{0}$. Let the covariance matrices of parameter heterogeneity be given as

$$\mathrm{E}(\boldsymbol{\eta}_i\boldsymbol{\eta}_i^\top|\boldsymbol{x}) = \boldsymbol{\Sigma}_\eta \quad \text{and} \quad \mathrm{E}(\boldsymbol{\omega}_j\boldsymbol{\omega}_j^\top|\boldsymbol{x}) = \boldsymbol{\Sigma}_\omega.$$

3.2 Testing Slope Heterogeneity

In this study, we test for slope heterogeneity of β_{ij} presented in (3.2) where the regression coefficients depend on either one of the two factors of interest. The two tests can be performed in a similar manner. Therefore, in this thesis, we derive a test statistic on the effect of one factor while the test for another factor can be similarly constructed. The hypothesis of interest is

$$H_0: \ \boldsymbol{\Sigma}_\omega = 0$$
$$v.s.$$
$$H_A: \ \boldsymbol{\Sigma}_\omega \neq 0.$$

Under the null hypothesis, the model in (3.1) can be written in a matrix form as follows, for $j = 1, 2, ..., J$,

$$\begin{bmatrix} y_{11}^{(j)} \\ y_{12}^{(j)} \\ \vdots \\ y_{1T}^{(j)} \\ \vdots \\ y_{I1}^{(j)} \\ y_{I2}^{(j)} \\ \vdots \\ y_{IT}^{(j)} \end{bmatrix} = \gamma_j \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \\ \vdots \\ 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} + \begin{bmatrix} 1 & 0 & \cdots & 0 & x_{111}^{(j)} & x_{112}^{(j)} & \cdots & x_{11k}^{(j)} \\ 1 & 0 & \cdots & 0 & x_{121}^{(j)} & x_{122}^{(j)} & \cdots & x_{12k}^{(j)} \\ & & \vdots & & & & \vdots & \\ 1 & 0 & \cdots & 0 & x_{1T1}^{(j)} & x_{1T2}^{(j)} & \cdots & x_{1Tk}^{(j)} \\ & & \ddots & & & & \vdots & \\ 0 & 0 & \cdots & 1 & x_{I11}^{(j)} & x_{I12}^{(j)} & \cdots & x_{I1k}^{(j)} \\ 0 & 0 & \cdots & 1 & x_{I21}^{(j)} & x_{I22}^{(j)} & \cdots & x_{I2k}^{(j)} \\ & & \vdots & & & & \vdots & \\ 0 & 0 & \cdots & 1 & x_{IT1}^{(j)} & x_{IT2}^{(j)} & \cdots & x_{ITk}^{(j)} \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_I \\ \beta_1^{(j)} \\ \beta_2^{(j)} \\ \vdots \\ \beta_k^{(j)} \end{bmatrix} + \begin{bmatrix} \epsilon_{11}^{(j)} \\ \epsilon_{12}^{(j)} \\ \vdots \\ \epsilon_{1T}^{(j)} \\ \vdots \\ \epsilon_{I1}^{(j)} \\ \epsilon_{I2}^{(j)} \\ \vdots \\ \epsilon_{IT}^{(j)} \end{bmatrix}$$

or

$$\boldsymbol{y}_j = \gamma_j \boldsymbol{\iota}_{IT} + \boldsymbol{X}_j \boldsymbol{\beta}_j + \boldsymbol{\epsilon}_j, \tag{3.3}$$

where $\boldsymbol{y}_j$ is an $IT \times 1$ vector of responses $(y_{it}^{(j)})$, $\boldsymbol{X}_j$ is a covariate matrix with dimension $IT \times (K + I)$, $\boldsymbol{\epsilon}_j$ is an $IT \times 1$ vector of error terms $(\epsilon_{it}^{(j)})$ with a multivariate normal distribution with mean zero and covariance matrix $\sigma_\epsilon^2 \mathbb{I}_{IT}$ where $\mathbb{I}_{IT}$ is an identity matrix of size $IT \times IT$, $\boldsymbol{\iota}_{IT}$ is a $IT \times 1$ vector of ones and

$\boldsymbol{\beta}_j = (\alpha_1, \cdots, \alpha_I, \beta_1, \cdots, \beta_K)^\top$ is a $(K+I)$ vector of slope coefficients.

To find parameter estimations of the model, define $\boldsymbol{M}_0 = \mathbb{I}_{IT} - \frac{\boldsymbol{\iota}_{IT}\boldsymbol{\iota}_{IT}^\top}{IT}$, where properties of $\boldsymbol{M}_0$ are listed in the following proposition.

Proposition 3.1.

1) $\boldsymbol{M}_0\boldsymbol{\iota}_{IT} = 0$.

2) $\boldsymbol{M}_0^\top = \boldsymbol{M}_0$.

3) $\boldsymbol{M}_0^\top\boldsymbol{M}_0 = \boldsymbol{M}_0$.

Proof. 1)

$$\begin{aligned}
\boldsymbol{M}_0\boldsymbol{\iota}_{IT} &= \left(\mathbb{I}_{IT} - \frac{\boldsymbol{\iota}_{IT}\boldsymbol{\iota}_{IT}^\top}{IT}\right)\boldsymbol{\iota}_{IT} \\
&= \boldsymbol{\iota}_{IT} - \frac{\boldsymbol{\iota}_{IT}\boldsymbol{\iota}_{IT}^\top\boldsymbol{\iota}_{IT}}{IT} \\
&= \boldsymbol{\iota}_{IT} - \boldsymbol{\iota}_{IT} \\
&= 0,
\end{aligned}$$

where we use the fact that $\boldsymbol{\iota}_{IT}^\top\boldsymbol{\iota}_{IT} = IT$.

2)

$$\begin{aligned}
\boldsymbol{M}_0^\top &= \left(\mathbb{I}_{IT} - \frac{\boldsymbol{\iota}_{IT}\boldsymbol{\iota}_{IT}^\top}{IT}\right)^\top \\
&= \mathbb{I}_{IT} - \frac{\boldsymbol{\iota}_{IT}\boldsymbol{\iota}_{IT}^\top}{IT} \\
&= \boldsymbol{M}_0.
\end{aligned}$$

3)

$$\begin{aligned}
\boldsymbol{M}_0^\top\boldsymbol{M}_0 &= \left(\mathbb{I}_{IT} - \frac{\boldsymbol{\iota}_{IT}\boldsymbol{\iota}_{IT}^\top}{IT}\right)^\top\left(\mathbb{I}_{IT} - \frac{\boldsymbol{\iota}_{IT}\boldsymbol{\iota}_{IT}^\top}{IT}\right) \\
&= \left(\mathbb{I}_{IT} - \frac{\boldsymbol{\iota}_{IT}\boldsymbol{\iota}_{IT}^\top}{IT}\right)\left(\mathbb{I}_{IT} - \frac{\boldsymbol{\iota}_{IT}\boldsymbol{\iota}_{IT}^\top}{IT}\right) \\
&= \mathbb{I}_{IT} - \frac{\boldsymbol{\iota}_{IT}\boldsymbol{\iota}_{IT}^\top}{IT} - \frac{\boldsymbol{\iota}_{IT}\boldsymbol{\iota}_{IT}^\top}{IT} + \frac{\boldsymbol{\iota}_{IT}\boldsymbol{\iota}_{IT}^\top}{IT}\frac{\boldsymbol{\iota}_{IT}\boldsymbol{\iota}_{IT}^\top}{IT} \\
&= \mathbb{I}_{IT} - \frac{\boldsymbol{\iota}_{IT}\boldsymbol{\iota}_{IT}^\top}{IT} - \frac{\boldsymbol{\iota}_{IT}\boldsymbol{\iota}_{IT}^\top}{IT} + \frac{\boldsymbol{\iota}_{IT}IT\boldsymbol{\iota}_{IT}^\top}{ITIT}
\end{aligned}$$

$$
\begin{aligned}
&= \mathbb{I}_{IT} - \frac{\boldsymbol{\iota}_{IT}\boldsymbol{\iota}_{IT}^\top}{IT} - \frac{\boldsymbol{\iota}_{IT}\boldsymbol{\iota}_{IT}^\top}{IT} + \frac{\boldsymbol{\iota}_{IT}\boldsymbol{\iota}_{IT}^\top}{IT} \\
&= \mathbb{I}_{IT} - \frac{\boldsymbol{\iota}_{IT}\boldsymbol{\iota}_{IT}^\top}{IT} \\
&= \boldsymbol{M}_0,
\end{aligned}
$$

where, similar to 1), we use the fact that $\boldsymbol{\iota}_{IT}^\top\boldsymbol{\iota}_{IT} = IT$. □

Multiplying $\boldsymbol{M}_0$ to (3.3), from Proposition 3.1(1), we obtain

$$
\begin{aligned}
\boldsymbol{M}_0\boldsymbol{y}_j &= \boldsymbol{M}_0\gamma_j\boldsymbol{\iota}_{IT} + \boldsymbol{M}_0\boldsymbol{X}_j\boldsymbol{\beta}_j + \boldsymbol{M}_0\boldsymbol{\epsilon}_j \\
&= \boldsymbol{M}_0\boldsymbol{X}_j\boldsymbol{\beta}_j + \boldsymbol{M}_0\boldsymbol{\epsilon}_j. \qquad (3.4)
\end{aligned}
$$

The ordinary least square estimator $\hat{\boldsymbol{\beta}}$ of $\boldsymbol{\beta}$ is obtained by minimizing the sum of squared errors as the following

$$
\begin{aligned}
SSE_\beta &= \sum_{j=1}^{J} (\boldsymbol{M}_0\boldsymbol{\epsilon}_j)^\top(\boldsymbol{M}_0\boldsymbol{\epsilon}_j) \\
&= \sum_{j=1}^{J} (\boldsymbol{M}_0\boldsymbol{y}_j - \boldsymbol{M}_0\boldsymbol{X}_j\boldsymbol{\beta})^\top(\boldsymbol{M}_0\boldsymbol{y}_j - \boldsymbol{M}_0\boldsymbol{X}_j\boldsymbol{\beta}) \\
&= \sum_{j=1}^{J} (\boldsymbol{M}_0\boldsymbol{y}_j)^\top\boldsymbol{M}_0\boldsymbol{y}_j - (\boldsymbol{M}_0\boldsymbol{y}_j)^\top\boldsymbol{M}_0\boldsymbol{X}_j\boldsymbol{\beta} - (\boldsymbol{M}_0\boldsymbol{X}_j\boldsymbol{\beta})^\top\boldsymbol{M}_0\boldsymbol{y}_j \\
&\quad + (\boldsymbol{M}_0\boldsymbol{X}_j\boldsymbol{\beta})^\top\boldsymbol{M}_0\boldsymbol{X}_j\boldsymbol{\beta} \\
&= \sum_{j=1}^{J} \boldsymbol{y}_j^\top\boldsymbol{M}_0\boldsymbol{y}_j - \boldsymbol{y}_j^\top\boldsymbol{M}_0(\boldsymbol{X}_j\boldsymbol{\beta}) - (\boldsymbol{X}_j\boldsymbol{\beta})^\top\boldsymbol{M}_0\boldsymbol{y}_j + (\boldsymbol{X}_j\boldsymbol{\beta})^\top\boldsymbol{M}_0(\boldsymbol{X}_j\boldsymbol{\beta}) \qquad (3.5) \\
&= \sum_{j=1}^{J} \boldsymbol{y}_j^\top\boldsymbol{M}_0\boldsymbol{y}_j - 2\boldsymbol{y}_j^\top\boldsymbol{M}_0(\boldsymbol{X}_j\boldsymbol{\beta}) + (\boldsymbol{X}_j\boldsymbol{\beta})^\top\boldsymbol{M}_0(\boldsymbol{X}_j\boldsymbol{\beta}),
\end{aligned}
$$

where we use Proposition 3.1(2) to obtain (3.5).
Taking partial derivative of SSE_β with respect to $\boldsymbol{\beta}$,

$$
\frac{\partial SSE_\beta}{\partial \boldsymbol{\beta}} = \sum_{j=1}^{J} -2\boldsymbol{X}_j^\top\boldsymbol{M}_0\boldsymbol{y}_j + 2\boldsymbol{X}_j^\top\boldsymbol{M}_0\boldsymbol{X}_j\boldsymbol{\beta}.
$$

Since $\left.\dfrac{\partial SSE_\beta}{\partial \boldsymbol{\beta}}\right|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}} = 0, \quad -2\sum_{j=1}^{J} \boldsymbol{X}_j^\top \boldsymbol{M}_0 \boldsymbol{y}_j + 2\hat{\boldsymbol{\beta}} \sum_{j=1}^{J} \boldsymbol{X}_j^\top \boldsymbol{M}_0 \boldsymbol{X}_j = 0.$

Therefore, the least square estimator $\hat{\boldsymbol{\beta}}$ is

$$\hat{\boldsymbol{\beta}} = \left(\sum_{j=1}^{J} \boldsymbol{X}_j^\top \boldsymbol{M}_0 \boldsymbol{X}_j\right)^{-} \sum_{j=1}^{J} \boldsymbol{X}_j^\top \boldsymbol{M}_0 \boldsymbol{y}_j.$$

Proposition 3.2. *The least square estimator $\hat{\boldsymbol{\beta}}$ has the following property*

$$E[\hat{\boldsymbol{\beta}}] = \left(\sum_{j=1}^{J} \boldsymbol{X}_j^\top \boldsymbol{M}_0 \boldsymbol{X}_j\right)^{-} \sum_{j=1}^{J} \boldsymbol{X}_j^\top \boldsymbol{M}_0 \boldsymbol{X}_j \boldsymbol{\beta}.$$

Proof.

$$\begin{aligned}
E[\hat{\boldsymbol{\beta}}] &= E\left[\left(\sum_{j=1}^{J} \boldsymbol{X}_j^\top \boldsymbol{M}_0 \boldsymbol{X}_j\right)^{-} \sum_{j=1}^{J} \boldsymbol{X}_j^\top \boldsymbol{M}_0 \boldsymbol{y}_j\right] \\
&= \left(\sum_{j=1}^{J} \boldsymbol{X}_j^\top \boldsymbol{M}_0 \boldsymbol{X}_j\right)^{-} \sum_{j=1}^{J} \boldsymbol{X}_j^\top \boldsymbol{M}_0 \; E[\boldsymbol{y}_j] \\
&= \left(\sum_{j=1}^{J} \boldsymbol{X}_j^\top \boldsymbol{M}_0 \boldsymbol{X}_j\right)^{-} \sum_{j=1}^{J} \boldsymbol{X}_j^\top \boldsymbol{M}_0 (\gamma_j \boldsymbol{\iota}_{IT} + \boldsymbol{X}_j \boldsymbol{\beta}) \\
&= \left(\sum_{j=1}^{J} \boldsymbol{X}_j^\top \boldsymbol{M}_0 \boldsymbol{X}_j\right)^{-} \sum_{j=1}^{J} \boldsymbol{X}_j^\top \boldsymbol{M}_0 \boldsymbol{X}_j \boldsymbol{\beta},
\end{aligned}$$

where we use Proposition 3.1(1) to obtain the last equation.
Note that if $\sum_{j=1}^{J} \boldsymbol{X}_j^\top \boldsymbol{M}_0 \boldsymbol{X}_j$ is invertible, $\hat{\boldsymbol{\beta}}$ is an unbiased estimator for $\boldsymbol{\beta}$. □

The ordinary least square estimator $\hat{\gamma}_j$ of γ_j is obtained by minimizing the sum of squared errors as the following

$$\begin{aligned}
SSE_\gamma &= \sum_{j=1}^{J} \boldsymbol{\epsilon}_j^\top \boldsymbol{\epsilon}_j \\
&= \sum_{j=1}^{J} (\boldsymbol{y}_j - \gamma_j \boldsymbol{\iota}_{IT} - \boldsymbol{X}_j \boldsymbol{\beta})^\top (\boldsymbol{y}_j - \gamma_j \boldsymbol{\iota}_{IT} - \boldsymbol{X}_j \boldsymbol{\beta})
\end{aligned}$$

$$
\begin{aligned}
&= \sum_{j=1}^{J} (\boldsymbol{y}_j - \boldsymbol{X}_j\boldsymbol{\beta})^\top (\boldsymbol{y}_j - \boldsymbol{X}_j\boldsymbol{\beta}) - (\boldsymbol{y}_j - \boldsymbol{X}_j\boldsymbol{\beta})^\top (\gamma_j \boldsymbol{\iota}_{IT}) - (\gamma_j \boldsymbol{\iota}_{IT})^\top (\boldsymbol{y}_j - \boldsymbol{X}_j\boldsymbol{\beta}) \\
&\quad + (\gamma_j \boldsymbol{\iota}_{IT})^\top (\gamma_j \boldsymbol{\iota}_{IT}) \\
&= \sum_{j=1}^{J} (\boldsymbol{y}_j - \boldsymbol{X}_j\boldsymbol{\beta})^\top (\boldsymbol{y}_j - \boldsymbol{X}_j\boldsymbol{\beta}) - 2(\boldsymbol{y}_j - \boldsymbol{X}_j\boldsymbol{\beta})^\top (\gamma_j \boldsymbol{\iota}_{IT}) + (\gamma_j \boldsymbol{\iota}_{IT})^\top (\gamma_j \boldsymbol{\iota}_{IT}) \\
&= \sum_{j=1}^{J} (\boldsymbol{y}_j - \boldsymbol{X}_j\boldsymbol{\beta})^\top (\boldsymbol{y}_j - \boldsymbol{X}_j\boldsymbol{\beta}) - 2\gamma_j (\boldsymbol{y}_j - \boldsymbol{X}_j\boldsymbol{\beta})^\top \boldsymbol{\iota}_{IT} + \gamma_j^2 \boldsymbol{\iota}_{IT}^\top \boldsymbol{\iota}_{IT}.
\end{aligned}
$$

Taking partial derivative of SSE_γ with respect to γ_j,

$$
\frac{\partial SSE_\gamma}{\partial \gamma_j} = -2(\boldsymbol{y}_j - \boldsymbol{X}_j\boldsymbol{\beta})^\top \boldsymbol{\iota}_{IT} + 2\gamma_j \boldsymbol{\iota}_{IT}^\top \boldsymbol{\iota}_{IT}.
$$

Since $\left.\dfrac{\partial SSE_\gamma}{\partial \gamma_j}\right|_{\gamma_j=\hat{\gamma}_j, \boldsymbol{\beta}=\hat{\boldsymbol{\beta}}} = 0, \ -2(\boldsymbol{y}_j - \boldsymbol{X}_j\hat{\boldsymbol{\beta}})^\top \boldsymbol{\iota}_{IT} + 2\hat{\gamma}_j \boldsymbol{\iota}_{IT}^\top \boldsymbol{\iota}_{IT} = 0.$

Therefore, the least square estimator $\hat{\gamma}_j$ $(j = 1, 2, ..., J)$ is

$$
\begin{aligned}
\hat{\gamma}_j &= \frac{(\boldsymbol{y}_j - \boldsymbol{X}_j\hat{\boldsymbol{\beta}})^\top \boldsymbol{\iota}_{IT}}{IT} \\
&= \frac{\boldsymbol{\iota}_{IT}^\top (\boldsymbol{y}_j - \boldsymbol{X}_j\hat{\boldsymbol{\beta}})}{IT}.
\end{aligned}
$$

Proposition 3.3. *The least square estimator* $\hat{\gamma}_j$ $(j = 1, 2, ..., J)$ *has the following property*

$$
E[\hat{\gamma}_j] = \gamma_j + \frac{\boldsymbol{X}_j \left(\boldsymbol{\beta} - \left(\sum_{j=1}^{J} \boldsymbol{X}_j^\top \boldsymbol{M}_0 \boldsymbol{X}_j \right)^{-} \sum_{j=1}^{J} \boldsymbol{X}_j^\top \boldsymbol{M}_0 \boldsymbol{X}_j \boldsymbol{\beta} \right)}{IT}.
$$

Proof.

$$
\begin{aligned}
E[\hat{\gamma}_j] &= E\left[\frac{\boldsymbol{\iota}_{IT}^\top (\boldsymbol{y}_j - \boldsymbol{X}_j\hat{\boldsymbol{\beta}})}{IT} \right] \\
&= \frac{\boldsymbol{\iota}_{IT}^\top (E[\boldsymbol{y}_j] - \boldsymbol{X}_j E[\hat{\boldsymbol{\beta}}])}{IT} \\
&= \frac{\boldsymbol{\iota}_{IT}^\top (\gamma_j \boldsymbol{\iota}_{IT} + \boldsymbol{X}_j\boldsymbol{\beta} - \boldsymbol{X}_j E[\hat{\boldsymbol{\beta}}])}{IT}
\end{aligned}
$$

$$= \gamma_j + \frac{\boldsymbol{X}_j(\boldsymbol{\beta} - E[\hat{\boldsymbol{\beta}}])}{IT}$$

$$= \gamma_j + \frac{\boldsymbol{X}_j\left(\boldsymbol{\beta} - \left(\sum_{j=1}^{J} \boldsymbol{X}_j^\top \boldsymbol{M}_0 \boldsymbol{X}_j\right)^- \sum_{j=1}^{J} \boldsymbol{X}_j^\top \boldsymbol{M}_0 \boldsymbol{X}_j \boldsymbol{\beta}\right)}{IT}.$$

Note that if $\hat{\boldsymbol{\beta}}$ is an unbiased estimator for $\boldsymbol{\beta}$, then $\hat{\gamma}_j$ is an unbiased estimator for γ_j. □

The likelihood function of $\boldsymbol{\theta} = (\gamma_j, \boldsymbol{\beta}, \sigma_\epsilon^2, \boldsymbol{\Sigma}_\omega)$ is given by

$$\begin{aligned} L(\boldsymbol{\theta}|\boldsymbol{y}) &= f(y_1, y_2, ..., y_J|\boldsymbol{\theta}) = \prod_{j=1}^{J} f(\boldsymbol{y}_j|\boldsymbol{\theta}) \\ &= \prod_{j=1}^{J} (2\pi|\boldsymbol{\Sigma}_j|)^{-1/2} \exp\left(-\frac{1}{2}(\boldsymbol{y}_j - \gamma_j \boldsymbol{\iota}_{IT} - \boldsymbol{X}_j\boldsymbol{\beta})^\top \boldsymbol{\Sigma}_j^{-1}(\boldsymbol{y}_j - \gamma_j \boldsymbol{\iota}_{IT} - \boldsymbol{X}_j\boldsymbol{\beta})\right) \\ &= \left[\prod_{j=1}^{J} (2\pi|\boldsymbol{\Sigma}_j|)^{-1/2}\right] \\ &\quad \exp\left(-\frac{1}{2}\sum_{j=1}^{J}(\boldsymbol{y}_j - \gamma_j \boldsymbol{\iota}_{IT} - \boldsymbol{X}_j\boldsymbol{\beta})^\top \boldsymbol{\Sigma}_j^{-1}(\boldsymbol{y}_j - \gamma_j \boldsymbol{\iota}_{IT} - \boldsymbol{X}_j\boldsymbol{\beta})\right). \end{aligned} \tag{3.6}$$

The log-likelihood function is

$$\begin{aligned} \ln L(\boldsymbol{\theta}|\boldsymbol{y}) &= \ln\Bigg[\Big[\prod_{j=1}^{J} (2\pi|\boldsymbol{\Sigma}_j|)^{-1/2}\Big] \exp\Big(-\frac{1}{2}\sum_{j=1}^{J}(\boldsymbol{y}_j - \gamma_j \boldsymbol{\iota}_{IT} - \boldsymbol{X}_j\boldsymbol{\beta})^\top \boldsymbol{\Sigma}_j^{-1} \\ &\quad (\boldsymbol{y}_j - \gamma_j \boldsymbol{\iota}_{IT} - \boldsymbol{X}_j\boldsymbol{\beta})\Big)\Bigg] \\ &= \sum_{j=1}^{J} \ln(2\pi|\boldsymbol{\Sigma}_j|)^{-1/2} - \frac{1}{2}(\boldsymbol{y}_j - \gamma_j \boldsymbol{\iota}_{IT} - \boldsymbol{X}_j\boldsymbol{\beta})^\top \boldsymbol{\Sigma}_j^{-1}(\boldsymbol{y}_j - \gamma_j \boldsymbol{\iota}_{IT} - \boldsymbol{X}_j\boldsymbol{\beta}) \\ &= \sum_{j=1}^{J} -\frac{1}{2}\ln(2\pi) - \frac{1}{2}\ln|\boldsymbol{\Sigma}_j| - \frac{1}{2}(\boldsymbol{y}_j - \gamma_j \boldsymbol{\iota}_{IT} - \boldsymbol{X}_j\boldsymbol{\beta})^\top \Sigma_j^{-1} \\ &\quad (\boldsymbol{y}_j - \gamma_j \boldsymbol{\iota}_{IT} - \boldsymbol{X}_j\boldsymbol{\beta}) \\ &= -\frac{J}{2}\ln(2\pi) - \frac{1}{2}\sum_{j=1}^{J}\ln|\boldsymbol{\Sigma}_j| - \frac{1}{2}\sum_{j=1}^{J}(\boldsymbol{y}_j - \gamma_j \boldsymbol{\iota}_{IT} - \boldsymbol{X}_j\boldsymbol{\beta})^\top \Sigma_j^{-1} \\ &\quad (\boldsymbol{y}_j - \gamma_j \boldsymbol{\iota}_{IT} - \boldsymbol{X}_j\boldsymbol{\beta}). \end{aligned}$$

Ignoring the constant term, the log-likelihood function is written as

$$l(\gamma_j, \boldsymbol{\beta}, \sigma_\epsilon^2, \boldsymbol{\Sigma}_\omega) = -\frac{1}{2}\sum_{j=1}^{J} \ \ln|\boldsymbol{\Sigma}_j| - \frac{1}{2}\sum_{j=1}^{J} \ \boldsymbol{e}_j^\top \boldsymbol{\Sigma}_j^{-1} \boldsymbol{e}_j,$$

$$\begin{aligned} \boldsymbol{e}_j &= \boldsymbol{y}_j - \gamma_j \boldsymbol{\iota}_{IT} - \boldsymbol{X}_j \boldsymbol{\beta} \\ &= \boldsymbol{X}_j \boldsymbol{\omega}_j + \boldsymbol{\epsilon}_j, \end{aligned}$$

and

$$\begin{aligned} \boldsymbol{\Sigma}_j &= E[\boldsymbol{e}_j \boldsymbol{e}_j^\top | \boldsymbol{X}_j] \\ &= \boldsymbol{X}_j \boldsymbol{\Sigma}_\omega \boldsymbol{X}_j^\top + \sigma_\epsilon^2 \mathbb{I}_{IT}. \end{aligned}$$

Therefore, under null hypothesis, $\boldsymbol{\Sigma}_j = \sigma_\epsilon^2 \mathbb{I}_{IT}$.

The likelihood function has J parameters associated with γ_j $(j = 1, 2, ..., J)$, but we wish to consider when $J \to \infty$. Therefore, to avoid the problem of estimation of γ_j, we study the likelihood conditional on the sufficient statistic S_{γ_j} for γ_j given as follows.

Proposition 3.4. *A sufficient statistic for* γ_j *is* $S_{\gamma_j} = \dfrac{\boldsymbol{y}_j^\top \boldsymbol{\Sigma}_j^{-1} \boldsymbol{\iota}_{IT}}{\boldsymbol{\iota}_{IT}^\top \boldsymbol{\Sigma}_j^{-1} \boldsymbol{\iota}_{IT}}$.

Proof. From (3.6), we can write the likelihood function as

$$\begin{aligned} L(\boldsymbol{\theta}|\boldsymbol{y}) = & \left[\prod_{j=1}^{J} (2\pi|\boldsymbol{\Sigma}_j|)^{-1/2}\right] \exp\left(-\frac{1}{2}\sum_{j=1}^{J} (\boldsymbol{y}_j - \boldsymbol{X}_j\boldsymbol{\beta})^\top \boldsymbol{\Sigma}_j^{-1} (\boldsymbol{y}_j - \boldsymbol{X}_j\boldsymbol{\beta})\right) \\ & \exp\left(-\frac{1}{2}\sum_{j=1}^{J} -2(\boldsymbol{y}_j - \boldsymbol{X}_j\boldsymbol{\beta})^\top \boldsymbol{\Sigma}_j^{-1} \gamma_j \boldsymbol{\iota}_{IT} + \gamma_j^2 \boldsymbol{\iota}_{IT}^\top \boldsymbol{\Sigma}_j^{-1} \boldsymbol{\iota}_{IT}\right) \\ = & \left[\prod_{j=1}^{J} (2\pi|\boldsymbol{\Sigma}_j|)^{-1/2}\right] \exp\left(-\frac{1}{2}\sum_{j=1}^{J} (\boldsymbol{y}_j - \boldsymbol{X}_j\boldsymbol{\beta})^\top \boldsymbol{\Sigma}_j^{-1} (\boldsymbol{y}_j - \boldsymbol{X}_j\boldsymbol{\beta})\right) \\ & \exp\left(\sum_{j=1}^{J} \gamma_j \boldsymbol{\iota}_{IT}^\top \boldsymbol{\Sigma}_j^{-1} \boldsymbol{\iota}_{IT} \left[\frac{(\boldsymbol{y}_j - \boldsymbol{X}_j\boldsymbol{\beta})^\top \boldsymbol{\Sigma}_j^{-1} \boldsymbol{\iota}_{IT}}{\boldsymbol{\iota}_{IT}^\top \boldsymbol{\Sigma}_j^{-1} \boldsymbol{\iota}_{IT}} - \gamma_j\right]\right) \\ = & \left[\prod_{j=1}^{J} (2\pi|\boldsymbol{\Sigma}_j|)^{-1/2}\right] \exp\left(-\frac{1}{2}\sum_{j=1}^{J} (\boldsymbol{y}_j - \boldsymbol{X}_j\boldsymbol{\beta})^\top \boldsymbol{\Sigma}_j^{-1} (\boldsymbol{y}_j - \boldsymbol{X}_j\boldsymbol{\beta})\right) \\ & \exp\left(\sum_{j=1}^{J} \gamma_j \boldsymbol{\iota}_{IT}^\top \boldsymbol{\Sigma}_j^{-1} \boldsymbol{\iota}_{IT} \left[\frac{\boldsymbol{y}_j^\top \boldsymbol{\Sigma}_j^{-1} \boldsymbol{\iota}_{IT}}{\boldsymbol{\iota}_{IT}^\top \boldsymbol{\Sigma}_j^{-1} \boldsymbol{\iota}_{IT}} - \frac{(\boldsymbol{X}_j\boldsymbol{\beta})^\top \boldsymbol{\Sigma}_j^{-1} \boldsymbol{\iota}_{IT}}{\boldsymbol{\iota}_{IT}^\top \boldsymbol{\Sigma}_j^{-1} \boldsymbol{\iota}_{IT}} - \gamma_j\right]\right) \end{aligned}$$

$$= \left[\prod_{j=1}^{J}(2\pi|\mathbf{\Sigma}_j|)^{-1/2}\right]\exp\Bigg(-\frac{1}{2}\sum_{j=1}^{J}(\boldsymbol{y}_j - \boldsymbol{X}_j\boldsymbol{\beta})^\top\mathbf{\Sigma}_j^{-1}(\boldsymbol{y}_j - \boldsymbol{X}_j\boldsymbol{\beta})$$

$$-\sum_{j=1}^{J}\gamma_j\boldsymbol{\iota}_{IT}^\top\mathbf{\Sigma}_j^{-1}\boldsymbol{\iota}_{IT}\left[\frac{(\boldsymbol{X}_j\boldsymbol{\beta})^\top\mathbf{\Sigma}_j^{-1}\boldsymbol{\iota}_{IT}}{\boldsymbol{\iota}_{IT}^\top\mathbf{\Sigma}_j^{-1}\boldsymbol{\iota}_{IT}} + \gamma_j\right]\Bigg)$$

$$\exp\left(\sum_{j=1}^{J}\gamma_j\boldsymbol{\iota}_{IT}^\top\mathbf{\Sigma}_j^{-1}\boldsymbol{\iota}_{IT}\left[\frac{\boldsymbol{y}_j^\top\mathbf{\Sigma}_j^{-1}\boldsymbol{\iota}_{IT}}{\boldsymbol{\iota}_{IT}^\top\mathbf{\Sigma}_j^{-1}\boldsymbol{\iota}_{IT}}\right]\right)$$

$$= h(\boldsymbol{y})\exp\left(\sum_{j=1}^{J}\gamma_j\boldsymbol{\iota}_{IT}^\top\mathbf{\Sigma}_j^{-1}\boldsymbol{\iota}_{IT}\left[\frac{\boldsymbol{y}_j^\top\mathbf{\Sigma}_j^{-1}\boldsymbol{\iota}_{IT}}{\boldsymbol{\iota}_{IT}^\top\mathbf{\Sigma}_j^{-1}\boldsymbol{\iota}_{IT}}\right]\right).$$

From the Factorization theorem, $S_{\gamma_j} = \dfrac{\boldsymbol{y}_j^\top\mathbf{\Sigma}_j^{-1}\boldsymbol{\iota}_{IT}}{\boldsymbol{\iota}_{IT}^\top\mathbf{\Sigma}_j^{-1}\boldsymbol{\iota}_{IT}}$ is a sufficient statistic for γ_j. □

3.2.1 Conditional Likelihood

To find the conditional likelihood function, we condition on the sufficient statistic S_{γ_j} as follows.

The conditional likelihood function of $\boldsymbol{\theta} = (\boldsymbol{\beta}, \sigma_\epsilon^2, \mathbf{\Sigma}_\omega)$ is

$$\begin{aligned} l_c &= \ln(f_{\boldsymbol{y}|S_\gamma}) \\ &= \ln\left(\prod_{j=1}^{J} f(\boldsymbol{y}_j)\right) - \ln\left(\prod_{j=1}^{J} f(S_{\gamma_j})\right) \\ &= \sum_{j=1}^{J}\ln f(\boldsymbol{y}_j) - \sum_{j=1}^{J}\ln f(\boldsymbol{S}_{\gamma_j}), \end{aligned} \tag{3.7}$$

where $f(\boldsymbol{y}_j)$ and $f(\boldsymbol{S}_{\gamma_j})$ are the probability density functions of $\boldsymbol{y}_j$ and S_{γ_j}, respectively. Since $\boldsymbol{y}_j \sim N(\boldsymbol{\mu}_j, \mathbf{\Sigma}_j)$,

$$f(\boldsymbol{y}_j) = (2\pi|\mathbf{\Sigma}_j|)^{-1/2}\exp\left(-\frac{1}{2}(\boldsymbol{y}_j - \gamma_j\boldsymbol{\iota}_{IT} - \boldsymbol{X}_j\boldsymbol{\beta})^\top\mathbf{\Sigma}_j^{-1}(\boldsymbol{y}_j - \gamma_j\boldsymbol{\iota}_{IT} - \boldsymbol{X}_j\boldsymbol{\beta})\right)$$

$$\ln f(\boldsymbol{y}_j) = -\frac{1}{2}\ln(2\pi) - \frac{1}{2}\ln|\mathbf{\Sigma}_j| - \frac{1}{2}\boldsymbol{e}_j^\top\mathbf{\Sigma}_j^{-1}\boldsymbol{e}_j, \tag{3.8}$$

where $\boldsymbol{e}_j = \boldsymbol{y}_j - \gamma_j\boldsymbol{\iota}_{IT} - \boldsymbol{X}_j\boldsymbol{\beta}$.

Using Theorem 2.56 and $\boldsymbol{y}_j \sim N(\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j)$, then

$$S_{\gamma_j} = \boldsymbol{a}_j \boldsymbol{y}_j \sim N(\boldsymbol{a}_j \boldsymbol{\mu}_j, \boldsymbol{a}_j \boldsymbol{\Sigma}_j \boldsymbol{a}_j^\top),$$

where $\boldsymbol{a}_j = (\boldsymbol{\iota}_{IT}^\top \boldsymbol{\Sigma}_j^{-1} \boldsymbol{\iota}_{IT})^{-1} \boldsymbol{\iota}_{IT}^\top \boldsymbol{\Sigma}_j^{-1}$.
Therefore, the probability density function of S_{γ_j} is

$$f(S_{\gamma_j}) = (2\pi |\boldsymbol{a}_j \boldsymbol{\Sigma}_j \boldsymbol{a}_j^\top|)^{-1/2} \exp\left(-\frac{1}{2} (\boldsymbol{a}_j \boldsymbol{y}_j - \boldsymbol{a}_j \boldsymbol{\mu}_j)^\top (\boldsymbol{a}_j \boldsymbol{\Sigma}_j \boldsymbol{a}_j^\top)^{-1} (\boldsymbol{a}_j \boldsymbol{y}_j - \boldsymbol{a}_j \boldsymbol{\mu}_j) \right),$$

then

$$\begin{aligned} \ln f(S_{\gamma_j}) &= -\frac{1}{2} \ln(2\pi) - \frac{1}{2} \ln |\boldsymbol{a}_j \boldsymbol{\Sigma}_j \boldsymbol{a}_j^\top| - \frac{1}{2} (\boldsymbol{y}_j - \boldsymbol{\mu}_j)^\top \boldsymbol{a}_j^\top (\boldsymbol{a}_j \boldsymbol{\Sigma}_j \boldsymbol{a}_j^\top)^{-1} \boldsymbol{a}_j (\boldsymbol{y}_j - \boldsymbol{\mu}_j) \\ &= -\frac{1}{2} \ln(2\pi) - \frac{1}{2} \ln |\boldsymbol{a}_j \boldsymbol{\Sigma}_j \boldsymbol{a}_j^\top| - \frac{1}{2} \boldsymbol{e}_j^\top \boldsymbol{a}_j^\top (\boldsymbol{a}_j \boldsymbol{\Sigma}_j \boldsymbol{a}_j^\top)^{-1} \boldsymbol{a}_j \boldsymbol{e}_j. \end{aligned} \tag{3.9}$$

Note that

$$\begin{aligned} \boldsymbol{a}_j \boldsymbol{\Sigma}_j \boldsymbol{a}_j^\top &= (\boldsymbol{\iota}_{IT}^\top \boldsymbol{\Sigma}_j^{-1} \boldsymbol{\iota}_{IT})^{-1} \boldsymbol{\iota}_{IT}^\top \boldsymbol{\Sigma}_j^{-1} \boldsymbol{\Sigma}_j \boldsymbol{\Sigma}_j^{-1} \boldsymbol{\iota}_{IT} (\boldsymbol{\iota}_{IT}^\top \boldsymbol{\Sigma}_j^{-1} \boldsymbol{\iota}_{IT})^{-1} \\ &= (\boldsymbol{\iota}_{IT}^\top \boldsymbol{\Sigma}_j^{-1} \boldsymbol{\iota}_{IT})^{-1} \boldsymbol{\iota}_{IT}^\top \boldsymbol{\Sigma}_j^{-1} \boldsymbol{\iota}_{IT} (\boldsymbol{\iota}_{IT}^\top \boldsymbol{\Sigma}_j^{-1} \boldsymbol{\iota}_{IT})^{-1} \\ &= (\boldsymbol{\iota}_{IT}^\top \boldsymbol{\Sigma}_j^{-1} \boldsymbol{\iota}_{IT})^{-1}, \end{aligned} \tag{3.10}$$

therefore,

$$\begin{aligned} \boldsymbol{a}_j^\top (\boldsymbol{a}_j \boldsymbol{\Sigma}_j \boldsymbol{a}_j^\top)^{-1} \boldsymbol{a}_j &= \boldsymbol{\Sigma}_j^{-1} \boldsymbol{\iota}_{IT} (\boldsymbol{\iota}_{IT}^\top \boldsymbol{\Sigma}_j^{-1} \boldsymbol{\iota}_{IT})^{-1} (\boldsymbol{\iota}_{IT}^\top \boldsymbol{\Sigma}_j^{-1} \boldsymbol{\iota}_{IT}) (\boldsymbol{\iota}_{IT}^\top \boldsymbol{\Sigma}_j^{-1} \boldsymbol{\iota}_{IT})^{-1} \boldsymbol{\iota}_{IT}^\top \boldsymbol{\Sigma}_j^{-1} \\ &= \boldsymbol{\Sigma}_j^{-1} \boldsymbol{\iota}_{IT} (\boldsymbol{\iota}_{IT}^\top \boldsymbol{\Sigma}_j^{-1} \boldsymbol{\iota}_{IT})^{-1} \boldsymbol{\iota}_{IT}^\top \boldsymbol{\Sigma}_j^{-1}. \end{aligned} \tag{3.11}$$

From (3.9), (3.10), and (3.11),

$$\ln f(S_{\gamma_j}) = -\frac{1}{2} \ln(2\pi) + \frac{1}{2} \ln(\boldsymbol{\iota}_{IT}^\top \boldsymbol{\Sigma}_j^{-1} \boldsymbol{\iota}_{IT}) - \frac{1}{2} \boldsymbol{e}_j^\top \left[\frac{\boldsymbol{\Sigma}_j^{-1} \boldsymbol{\iota}_{IT} \boldsymbol{\iota}_{IT}^\top \boldsymbol{\Sigma}_j^{-1}}{\boldsymbol{\iota}_{IT}^\top \boldsymbol{\Sigma}_j^{-1} \boldsymbol{\iota}_{IT}} \right] \boldsymbol{e}_j. \tag{3.12}$$

Substitute (3.8) and (3.12) into (3.7),

$$l_c = \sum_{j=1}^{J} \left(-\frac{1}{2} \ln |\boldsymbol{\Sigma}_j| - \frac{1}{2} \ln(\boldsymbol{\iota}_{IT}^\top \boldsymbol{\Sigma}_j^{-1} \boldsymbol{\iota}_{IT}) - \frac{1}{2} \boldsymbol{e}_j^\top \boldsymbol{\Sigma}_j^{-1} \boldsymbol{e}_j + \frac{1}{2} \boldsymbol{e}_j^\top \left[\frac{\boldsymbol{\Sigma}_j^{-1} \boldsymbol{\iota}_{IT} \boldsymbol{\iota}_{IT}^\top \boldsymbol{\Sigma}_j^{-1}}{\boldsymbol{\iota}_{IT}^\top \boldsymbol{\Sigma}_j^{-1} \boldsymbol{\iota}_{IT}} \right] \boldsymbol{e}_j \right). \tag{3.13}$$

Set $\boldsymbol{\zeta}_j = \boldsymbol{y}_j - \boldsymbol{X}_j \boldsymbol{\beta} = \boldsymbol{e}_j + \gamma_j \boldsymbol{\iota}_{IT}$, so, $\boldsymbol{e}_j = \boldsymbol{\zeta}_j - \gamma_j \boldsymbol{\iota}_{IT}$. Therefore,

$$\begin{aligned} \boldsymbol{e}_j^\top \boldsymbol{\Sigma}_j^{-1} \boldsymbol{e}_j &= (\boldsymbol{\zeta}_j - \gamma_j \boldsymbol{\iota}_{IT})^\top \boldsymbol{\Sigma}_j^{-1} (\boldsymbol{\zeta}_j - \gamma_j \boldsymbol{\iota}_{IT}) \\ &= \boldsymbol{\zeta}_j^\top \boldsymbol{\Sigma}_j^{-1} \boldsymbol{\zeta}_j - 2\gamma_j \boldsymbol{\zeta}_j^\top \boldsymbol{\Sigma}_j^{-1} \boldsymbol{\iota}_{IT} + \gamma_j^2 \boldsymbol{\iota}_{IT}^\top \boldsymbol{\Sigma}_j^{-1} \boldsymbol{\iota}_{IT}, \end{aligned} \tag{3.14}$$

and

$$\begin{aligned} \boldsymbol{e}_j^\top \left[\frac{\boldsymbol{\Sigma}_j^{-1} \boldsymbol{\iota}_{IT} \boldsymbol{\iota}_{IT}^\top \boldsymbol{\Sigma}_j^{-1}}{\boldsymbol{\iota}_{IT}^\top \boldsymbol{\Sigma}_j^{-1} \boldsymbol{\iota}_{IT}} \right] \boldsymbol{e}_j &= (\boldsymbol{\zeta}_j - \gamma_j \boldsymbol{\iota}_{IT})^\top \left[\frac{\boldsymbol{\Sigma}_j^{-1} \boldsymbol{\iota}_{IT} \boldsymbol{\iota}_{IT}^\top \boldsymbol{\Sigma}_j^{-1}}{\boldsymbol{\iota}_{IT}^\top \boldsymbol{\Sigma}_j^{-1} \boldsymbol{\iota}_{IT}} \right] (\boldsymbol{\zeta}_j - \gamma_j \boldsymbol{\iota}_{IT}) \\ &= \boldsymbol{\zeta}_j^\top \left[\frac{\boldsymbol{\Sigma}_j^{-1} \boldsymbol{\iota}_{IT} \boldsymbol{\iota}_{IT}^\top \boldsymbol{\Sigma}_j^{-1}}{\boldsymbol{\iota}_{IT}^\top \boldsymbol{\Sigma}_j^{-1} \boldsymbol{\iota}_{IT}} \right] \boldsymbol{\zeta}_j - 2\gamma_j \frac{(\boldsymbol{\iota}_{IT}^\top \boldsymbol{\Sigma}_j^{-1} \boldsymbol{\iota}_{IT}) \boldsymbol{\iota}^\top \boldsymbol{\Sigma}_j^{-1}}{\boldsymbol{\iota}_{IT}^\top \boldsymbol{\Sigma}_j^{-1} \boldsymbol{\iota}_{IT}} \boldsymbol{\zeta}_j \\ &\quad + \gamma_j^2 \frac{(\boldsymbol{\iota}_{IT}^\top \boldsymbol{\Sigma}_j^{-1} \boldsymbol{\iota}_{IT}) \boldsymbol{\iota}_{IT}^\top \boldsymbol{\Sigma}_j^{-1} \boldsymbol{\iota}_{IT}}{\boldsymbol{\iota}_{IT}^\top \boldsymbol{\Sigma}_j^{-1} \boldsymbol{\iota}_{IT}} \\ &= \boldsymbol{\zeta}_j^\top \left[\frac{\boldsymbol{\Sigma}_j^{-1} \boldsymbol{\iota}_{IT} \boldsymbol{\iota}_{IT}^\top \boldsymbol{\Sigma}_j^{-1}}{\boldsymbol{\iota}_{IT}^\top \boldsymbol{\Sigma}_j^{-1} \boldsymbol{\iota}_{IT}} \right] \boldsymbol{\zeta}_j - 2\gamma_j \boldsymbol{\iota}_{IT}^\top \boldsymbol{\Sigma}_j^{-1} \boldsymbol{\zeta}_j + \gamma_j^2 \boldsymbol{\iota}_{IT}^\top \boldsymbol{\Sigma}_j^{-1} \boldsymbol{\iota}_{IT}. \end{aligned} \tag{3.15}$$

Substitute (3.14) and (3.15) into (3.13), the conditional likelihood function is

$$\begin{aligned} l_c = &-\frac{1}{2} \sum_{j=1}^{J} \left[\ln |\boldsymbol{\Sigma}_j| + \ln(\boldsymbol{\iota}_{IT}^\top \boldsymbol{\Sigma}_j^{-1} \boldsymbol{\iota}_{IT}) \right] \\ &- \frac{1}{2} \sum_{j=1}^{J} \boldsymbol{\zeta}_j^\top \left[\boldsymbol{\Sigma}_j^{-1} - \frac{\boldsymbol{\Sigma}_j^{-1} \boldsymbol{\iota}_{IT} \boldsymbol{\iota}_{IT}^\top \boldsymbol{\Sigma}_j^{-1}}{\boldsymbol{\iota}_{IT}^\top \boldsymbol{\Sigma}_j^{-1} \boldsymbol{\iota}_{IT}} \right] \boldsymbol{\zeta}_j, \end{aligned} \tag{3.16}$$

where $\boldsymbol{\zeta}_j = \boldsymbol{y}_j - \boldsymbol{X}_j \boldsymbol{\beta}$.

To derive a test statistic, we examine the score evaluated at the true parameters under the null hypothesis.

Theorem 3.5. *The derivative of the conditional likelihood evaluated at* $\boldsymbol{\theta}_0 =$

$(\boldsymbol{\beta}, \sigma_\epsilon^2, \boldsymbol{\Sigma}_\omega = \boldsymbol{0})$ *is*

$$\begin{aligned}\frac{\partial l_c}{\partial \omega} &= \frac{1}{2\sigma^4}\sum_{j=1}^{J}\left[\boldsymbol{\epsilon}_j^\top \boldsymbol{M}_0 \boldsymbol{X}_j \boldsymbol{X}_j^\top \boldsymbol{M}_0 \boldsymbol{\epsilon}_j - \sigma_\epsilon^2 \text{trace}(\boldsymbol{X}_j^\top \boldsymbol{M}_0 \boldsymbol{X}_j)\right] \\ &\sim \sum_{j=1}^{J}\left[\boldsymbol{\epsilon}_j^\top \boldsymbol{M}_0 \boldsymbol{X}_j \boldsymbol{X}_j^\top \boldsymbol{M}_0 \boldsymbol{\epsilon}_j - \sigma_\epsilon^2 \text{trace}(\boldsymbol{X}_j^\top \boldsymbol{M}_0 \boldsymbol{X}_j)\right].\end{aligned}$$

Proof. From (3.16),

$$l_c = -\frac{1}{2}\sum_{j=1}^{J}(A_j + B_j + C_j - D_j),$$

where, for each $j = 1, 2, ..., J$,

$$\begin{aligned} A_j &= \ln |\boldsymbol{\Sigma}_j|, \\ B_j &= \ln(\boldsymbol{\iota}_{IT}^\top \boldsymbol{\Sigma}_j^{-1} \boldsymbol{\iota}_{IT}), \\ C_j &= \boldsymbol{\zeta}_j^\top \boldsymbol{\Sigma}_j^{-1} \boldsymbol{\zeta}_j,\end{aligned}$$

and

$$D_j = \boldsymbol{\zeta}_j^\top \left[\frac{\boldsymbol{\Sigma}_j^{-1}\boldsymbol{\iota}_{IT}\boldsymbol{\iota}_{IT}^\top\boldsymbol{\Sigma}_j^{-1}}{\boldsymbol{\iota}_{IT}^\top\boldsymbol{\Sigma}_j^{-1}\boldsymbol{\iota}_{IT}}\right]\boldsymbol{\zeta}_j.$$

For all terms of A_j, B_j, C_j, and D_j, the partial derivatives with respect to ω are obtained as follows.

To obtain $\frac{\partial A_j}{\partial \omega}$, using (2.13), (2.15), (2.18), and (2.23), we obtain

$$\begin{aligned}\frac{\partial A_j}{\partial \omega} &= \frac{\partial \ln|\boldsymbol{\Sigma}_j|}{\partial \text{vec}(\boldsymbol{\Sigma}_j)^\top}\frac{\partial \text{vec}(\boldsymbol{\Sigma}_j)}{\partial \text{vec}(\boldsymbol{\Sigma}_\omega)^\top}\frac{\partial \text{vec}(\boldsymbol{\Sigma}_\omega)}{\partial \omega} \\ &= \text{vec}(\boldsymbol{\Sigma}_j^{-1})^\top(\boldsymbol{X}_j \otimes \boldsymbol{X}_j)\text{vec}(\mathbb{I}_{I+K}). \end{aligned} \tag{3.17}$$

To obtain $\frac{\partial B_j}{\partial \omega}$, using (2.13), (2.15), (2.18), (2.22), and (2.23), we obtain

$$\begin{aligned}\frac{\partial B_j}{\partial \omega} &= \frac{\partial \ln(\boldsymbol{\iota}_{IT}^\top\boldsymbol{\Sigma}_j^{-1}\boldsymbol{\iota}_{IT})}{\partial \text{vec}(\boldsymbol{\Sigma}_j^{-1})^\top}\frac{\partial \text{vec}(\boldsymbol{\Sigma}_j^{-1})}{\partial \text{vec}(\boldsymbol{\Sigma}_j)^\top}\frac{\partial \text{vec}(\boldsymbol{\Sigma}_j)}{\partial \text{vec}(\boldsymbol{\Sigma}_\omega)^\top}\frac{\partial \text{vec}(\boldsymbol{\Sigma}_\omega)}{\partial \omega} \\ &= -\text{vec}((\boldsymbol{\iota}_{IT}^\top\boldsymbol{\Sigma}_j^{-1}\boldsymbol{\iota}_{IT})^{-1})^\top(\boldsymbol{\iota}_{IT}^\top\boldsymbol{\Sigma}_j^{-1} \otimes \boldsymbol{\iota}_{IT}^\top\boldsymbol{\Sigma}_j^{-1})(\boldsymbol{X}_j \otimes \boldsymbol{X}_j)\text{vec}(\mathbb{I}_{I+K})\end{aligned}$$

$$= -\frac{1}{\boldsymbol{\iota}_{IT}^\top \boldsymbol{\Sigma}_j^{-1} \boldsymbol{\iota}_{IT}} (\boldsymbol{\iota}_{IT}^\top \boldsymbol{\Sigma}_j^{-1} \otimes \boldsymbol{\iota}_{IT}^\top \boldsymbol{\Sigma}_j^{-1})(\boldsymbol{X}_j \otimes \boldsymbol{X}_j)\mathrm{vec}(\mathbb{I}_{I+K}). \tag{3.18}$$

To obtain $\frac{\partial C_j}{\partial \omega}$, using (2.13), (2.18), and (2.22), we obtain

$$\begin{aligned}\frac{\partial C_j}{\partial \omega} &= \frac{\partial \boldsymbol{\zeta}_j^\top \boldsymbol{\Sigma}_j^{-1} \boldsymbol{\zeta}_j}{\partial \mathrm{vec}(\boldsymbol{\Sigma}_j^{-1})^\top} \frac{\partial \mathrm{vec}(\boldsymbol{\Sigma}_j^{-1})}{\partial \mathrm{vec}(\boldsymbol{\Sigma}_j)^\top} \frac{\partial \mathrm{vec}(\boldsymbol{\Sigma}_j)}{\partial \mathrm{vec}(\boldsymbol{\Sigma}_\omega)^\top} \frac{\partial \mathrm{vec}(\boldsymbol{\Sigma}_\omega)}{\partial \omega} \\ &= -(\boldsymbol{\zeta}_j^\top \otimes \boldsymbol{\zeta}_j^\top)(\boldsymbol{\Sigma}_j^{-1} \otimes \boldsymbol{\Sigma}_j^{-1})(\boldsymbol{X}_j \otimes \boldsymbol{X}_j)\mathrm{vec}(\mathbb{I}_{I+K}). \end{aligned} \tag{3.19}$$

To obtain $\frac{\partial D_j}{\partial \omega}$, let

$$\begin{aligned} F(\omega) &= \frac{\mathbb{I}_{IT}}{\boldsymbol{\iota}_{IT}^\top \boldsymbol{\Sigma}_j^{-1} \boldsymbol{\iota}_{IT}} \\ G(\omega) &= \boldsymbol{\Sigma}_j^{-1} \boldsymbol{\iota}_{IT} \boldsymbol{\iota}_{IT}^\top \boldsymbol{\Sigma}_j^{-1}, \end{aligned}$$

then $D_j = \boldsymbol{\zeta}_j^\top F(\omega) G(\omega) \boldsymbol{\zeta}_j$.

Using (2.16), we have

$$\frac{\partial \mathrm{vec}(F(\omega)G(\omega))}{\partial \omega} = (G(\omega)^\top \otimes \mathbb{I}_{IT}) \frac{\partial \mathrm{vec}(F(\omega))}{\partial \omega} + (\mathbb{I}_{IT} \otimes F(\omega)) \frac{\partial \mathrm{vec}(G(\omega))}{\partial \omega}. \tag{3.20}$$

Using (2.13), (2.15), (2.18), and (2.22),

$$\begin{aligned}\frac{\partial \mathrm{vec}(F(\omega))}{\partial \omega} &= \mathrm{vec}(\mathbb{I}_{IT}) \frac{\partial (\boldsymbol{\iota}_{IT}^\top \boldsymbol{\Sigma}_j^{-1} \boldsymbol{\iota}_{IT})^{-1}}{\partial \mathrm{vec}(\boldsymbol{\Sigma}_j^{-1})^\top} \frac{\partial \mathrm{vec}(\boldsymbol{\Sigma}_j^{-1})}{\partial \mathrm{vec}(\boldsymbol{\Sigma}_j)^\top} \frac{\partial \mathrm{vec}(\boldsymbol{\Sigma}_j)}{\partial \mathrm{vec}(\boldsymbol{\Sigma}_\omega)^\top} \frac{\partial \mathrm{vec}(\boldsymbol{\Sigma}_\omega)}{\partial \omega} \\ &= -(\boldsymbol{\iota}_{IT}^\top \boldsymbol{\Sigma}_j^{-1} \boldsymbol{\iota}_{IT})^{-2} (\boldsymbol{\iota}_{IT}^\top \boldsymbol{\Sigma}_j^{-1} \otimes \boldsymbol{\iota}_{IT}^\top \boldsymbol{\Sigma}_j^{-1})(\boldsymbol{X}_j \otimes \boldsymbol{X}_j)\mathrm{vec}(\mathbb{I}_{I+K}) \\ &= -\frac{\boldsymbol{\iota}_{IT}^\top \boldsymbol{\Sigma}_j^{-1} \otimes \boldsymbol{\iota}_{IT}^\top \boldsymbol{\Sigma}_j^{-1}}{(\boldsymbol{\iota}_{IT}^\top \boldsymbol{\Sigma}_j^{-1} \boldsymbol{\iota}_{IT})^2} (\boldsymbol{X}_j \otimes \boldsymbol{X}_j)\mathrm{vec}(\mathbb{I}_{I+K}). \end{aligned} \tag{3.21}$$

Using (2.13), (2.16), (2.18), and (2.22),

$$\begin{aligned}\frac{\partial \mathrm{vec}(G(\omega))}{\partial \omega} &= \frac{\partial \mathrm{vec}(\boldsymbol{\Sigma}_j^{-1} \boldsymbol{\iota}_{IT} \boldsymbol{\iota}_{IT}^\top \boldsymbol{\Sigma}_j^{-1})}{\partial \mathrm{vec}(\boldsymbol{\Sigma}_j^{-1})^\top} \frac{\partial \mathrm{vec}(\boldsymbol{\Sigma}_j^{-1})}{\partial \mathrm{vec}(\boldsymbol{\Sigma}_j)^\top} \frac{\partial \mathrm{vec}(\boldsymbol{\Sigma}_j)}{\partial \mathrm{vec}(\boldsymbol{\Sigma}_\omega)^\top} \frac{\partial \mathrm{vec}(\boldsymbol{\Sigma}_\omega)}{\partial \omega} \\ &= -(\boldsymbol{\iota}_{IT} \boldsymbol{\iota}_{IT}^\top \boldsymbol{\Sigma}_j^{-1} \otimes \mathbb{I}_{IT} + \mathbb{I}_{IT} \otimes \boldsymbol{\Sigma}_j^{-1} \boldsymbol{\iota}_{IT} \boldsymbol{\iota}_{IT}^\top)(\boldsymbol{\Sigma}_j^{-1} \otimes \boldsymbol{\Sigma}_j^{-1})(\boldsymbol{X}_j \otimes \boldsymbol{X}_j)\mathrm{vec}(\mathbb{I}_{I+K}) \end{aligned}$$

$$
\begin{aligned}
&= -(\boldsymbol{\iota}_{IT}\boldsymbol{\iota}_{IT}^{\top}\boldsymbol{\Sigma}_j^{-1} \otimes \mathbb{I}_{IT})(\boldsymbol{\Sigma}_j^{-1} \otimes \boldsymbol{\Sigma}_j^{-1})(\boldsymbol{X}_j \otimes \boldsymbol{X}_j)\mathrm{vec}(\mathbb{I}_{I+K}) \\
&\quad - (\mathbb{I}_{IT} \otimes \boldsymbol{\Sigma}_j^{-1}\boldsymbol{\iota}_{IT}\boldsymbol{\iota}_{IT}^{\top})(\boldsymbol{\Sigma}_j^{-1} \otimes \boldsymbol{\Sigma}_j^{-1})(\boldsymbol{X}_j \otimes \boldsymbol{X}_j)\mathrm{vec}(\mathbb{I}_{I+K}) \\
&= -(\boldsymbol{\Sigma}_j^{-1}\boldsymbol{\iota}_{IT}\boldsymbol{\iota}_{IT}^{\top}\boldsymbol{\Sigma}_j^{-1} \otimes \boldsymbol{\Sigma}_j^{-1})(\boldsymbol{X}_j \otimes \boldsymbol{X}_j)\mathrm{vec}(\mathbb{I}_{I+K}) \\
&\quad - (\boldsymbol{\Sigma}_j^{-1} \otimes \boldsymbol{\Sigma}_j^{-1}\boldsymbol{\iota}_{IT}\boldsymbol{\iota}_{IT}^{\top}\boldsymbol{\Sigma}_j^{-1})(\boldsymbol{X}_j \otimes \boldsymbol{X}_j)\mathrm{vec}(\mathbb{I}_{I+K}). \qquad (3.22)
\end{aligned}
$$

Substitute (3.21) and (3.22) into (3.20), we have

$$
\begin{aligned}
\frac{\partial D_j}{\partial \omega} &= \frac{\partial \boldsymbol{\zeta}_j^{\top} F(\omega) G(\omega) \boldsymbol{\zeta}_j}{\partial \omega} \\
&= -\left[(\boldsymbol{\zeta}_j^{\top}\boldsymbol{\Sigma}_j^{-1}\boldsymbol{\iota}_{IT}\boldsymbol{\iota}_{IT}^{\top}\boldsymbol{\Sigma}_j^{-1}\boldsymbol{\zeta}_j) \frac{\boldsymbol{\iota}_{IT}^{\top}\boldsymbol{\Sigma}_j^{-1} \otimes \boldsymbol{\iota}_{IT}^{\top}\boldsymbol{\Sigma}_j^{-1}}{(\boldsymbol{\iota}_{IT}^{\top}\boldsymbol{\Sigma}_j^{-1}\boldsymbol{\iota}_{IT})^2} \right] (\boldsymbol{X}_j \otimes \boldsymbol{X}_j)\mathrm{vec}(\mathbb{I}_{I+K}) \\
&\quad + \left(\frac{\boldsymbol{\zeta}_j^{\top}\boldsymbol{\Sigma}_j^{-1}\boldsymbol{\iota}_{IT}\boldsymbol{\iota}_{IT}^{\top}\boldsymbol{\Sigma}_j^{-1} \otimes \boldsymbol{\zeta}_j^{\top}\boldsymbol{\Sigma}_j^{-1}}{\boldsymbol{\iota}_{IT}^{\top}\boldsymbol{\Sigma}_j^{-1}\boldsymbol{\iota}_{IT}} (\boldsymbol{X}_j \otimes \boldsymbol{X}_j)\mathrm{vec}(\mathbb{I}_{I+K}) \right) \\
&\quad + \left(\frac{\boldsymbol{\zeta}_j^{\top}\boldsymbol{\Sigma}_j^{-1} \otimes \boldsymbol{\zeta}_j^{\top}\boldsymbol{\Sigma}_j^{-1}\boldsymbol{\iota}_{IT}\boldsymbol{\iota}_{IT}^{\top}\boldsymbol{\Sigma}_j^{-1}}{\boldsymbol{\iota}_{IT}^{\top}\boldsymbol{\Sigma}_j^{-1}\boldsymbol{\iota}_{IT}} (\boldsymbol{X}_j \otimes \boldsymbol{X}_j)\mathrm{vec}(\mathbb{I}_{I+K}) \right). \qquad (3.23)
\end{aligned}
$$

Combining the terms (3.17), (3.18), (3.19), and (3.23), therefore

$$
\begin{aligned}
\frac{\partial l_c}{\partial \omega} &= -\frac{1}{2}\sum_{j=1}^{J} \mathrm{vec}(\boldsymbol{\Sigma}_j^{-1})^{\top}(\boldsymbol{X}_j \otimes \boldsymbol{X}_j)\mathrm{vec}(\mathbb{I}_{I+K}) && (A1) \\
&+ \frac{1}{2}\sum_{j=1}^{J} \frac{1}{\boldsymbol{\iota}_{IT}^{\top}\boldsymbol{\Sigma}_j^{-1}\boldsymbol{\iota}_{IT}} (\boldsymbol{\iota}_{IT}^{\top}\boldsymbol{\Sigma}_j^{-1} \otimes \boldsymbol{\iota}_{IT}^{\top}\boldsymbol{\Sigma}_j^{-1})(\boldsymbol{X}_j \otimes \boldsymbol{X}_j)\mathrm{vec}(\mathbb{I}_{I+K}) && (A2) \\
&+ \frac{1}{2}\sum_{j=1}^{J} (\boldsymbol{\zeta}_j^{\top} \otimes \boldsymbol{\zeta}_j^{\top})(\boldsymbol{\Sigma}_j^{-1} \otimes \boldsymbol{\Sigma}_j^{-1})(\boldsymbol{X}_j \otimes \boldsymbol{X}_j)\mathrm{vec}(\mathbb{I}_{I+K}) && (A3) \\
&+ \frac{1}{2}\sum_{j=1}^{J} \left[(\boldsymbol{\zeta}_j^{\top}\boldsymbol{\Sigma}_j^{-1}\boldsymbol{\iota}_{IT}\boldsymbol{\iota}_{IT}^{\top}\boldsymbol{\Sigma}_j^{-1}\boldsymbol{\zeta}_j) \frac{\boldsymbol{\iota}_{IT}^{\top}\boldsymbol{\Sigma}_j^{-1} \otimes \boldsymbol{\iota}_{IT}^{\top}\boldsymbol{\Sigma}_j^{-1}}{(\boldsymbol{\iota}_{IT}^{\top}\boldsymbol{\Sigma}_j^{-1}\boldsymbol{\iota}_{IT})^2} \right] (\boldsymbol{X}_j \otimes \boldsymbol{X}_j)\mathrm{vec}(\mathbb{I}_{I+K}) && (A4) \\
&- \frac{1}{2}\sum_{j=1}^{J} \left(\frac{\boldsymbol{\zeta}_j^{\top}\boldsymbol{\Sigma}_j^{-1}\boldsymbol{\iota}_{IT}\boldsymbol{\iota}_{IT}^{\top}\boldsymbol{\Sigma}_j^{-1} \otimes \boldsymbol{\zeta}_j^{\top}\boldsymbol{\Sigma}_j^{-1}}{\boldsymbol{\iota}_{IT}^{\top}\boldsymbol{\Sigma}_j^{-1}\boldsymbol{\iota}_{IT}} (\boldsymbol{X}_j \otimes \boldsymbol{X}_j)\mathrm{vec}(\mathbb{I}_{I+K}) \right) && (A5) \\
&- \frac{1}{2}\sum_{j=1}^{J} \left(\frac{\boldsymbol{\zeta}_j^{\top}\boldsymbol{\Sigma}_j^{-1} \otimes \boldsymbol{\zeta}_j^{\top}\boldsymbol{\Sigma}_j^{-1}\boldsymbol{\iota}_{IT}\boldsymbol{\iota}_{IT}^{\top}\boldsymbol{\Sigma}_j^{-1}}{\boldsymbol{\iota}_{IT}^{\top}\boldsymbol{\Sigma}_j^{-1}\boldsymbol{\iota}_{IT}} (\boldsymbol{X}_j \otimes \boldsymbol{X}_j)\mathrm{vec}(\mathbb{I}_{I+K}) \right). && (A6)
\end{aligned}
$$

Next, we compute $(A1) - (A6)$ under the null hypothesis. Throughout the calculation, we use the notation $\overset{H_0}{=}$ to refer to equal under the null hypothesis H_0.

Applying Theorem 2.54 and Theorem 2.41(2) to $(A1)$, respectively, then $(A1)$ under the null hypothesis is

$$
\begin{aligned}
-\frac{1}{2}\sum_{j=1}^{J}&\mathrm{vec}(\boldsymbol{\Sigma}_j^{-1})^\top(\boldsymbol{X}_j\otimes\boldsymbol{X}_j)\mathrm{vec}(\mathbb{I}_{I+K})\\
&=-\frac{1}{2}\sum_{j=1}^{J}\mathrm{trace}[\boldsymbol{X}_j\boldsymbol{X}_j^\top\boldsymbol{\Sigma}_j^{-1}]\\
&=-\frac{1}{2}\sum_{j=1}^{J}\mathrm{trace}[\boldsymbol{X}_j^\top\boldsymbol{\Sigma}_j^{-1}\boldsymbol{X}_j]\\
&\overset{H_0}{=}-\frac{1}{2}\sum_{j=1}^{J}\mathrm{trace}[\boldsymbol{X}_j^\top(\sigma_\epsilon^2\mathbb{I}_{IT})^{-1}\boldsymbol{X}_j]\\
&=-\frac{1}{2\sigma_\epsilon^2}\sum_{j=1}^{J}\mathrm{trace}[\boldsymbol{X}_j^\top\mathbb{I}_{IT}\boldsymbol{X}_j].
\end{aligned}
$$

Applying Theorem 2.44(3), Theorem 2.53, and Theorem 2.41(2) to $(A2)$, respectively, then $(A2)$ under the null hypothesis is

$$
\begin{aligned}
\frac{1}{2}\sum_{j=1}^{J}&\frac{1}{\boldsymbol{\iota}_{IT}^\top\boldsymbol{\Sigma}_j^{-1}\boldsymbol{\iota}_{IT}}(\boldsymbol{\iota}_{IT}^\top\boldsymbol{\Sigma}_j^{-1}\otimes\boldsymbol{\iota}_{IT}^\top\boldsymbol{\Sigma}_j^{-1})(\boldsymbol{X}_j\otimes\boldsymbol{X}_j)\mathrm{vec}(\mathbb{I}_{I+K})\\
&=\frac{1}{2}\sum_{j=1}^{J}\frac{(\boldsymbol{\iota}_{IT}^\top\boldsymbol{\Sigma}_j^{-1}\boldsymbol{X}_j\otimes\boldsymbol{\iota}_{IT}^\top\boldsymbol{\Sigma}_j^{-1}\boldsymbol{X}_j)}{\boldsymbol{\iota}_{IT}^\top\boldsymbol{\Sigma}_j^{-1}\boldsymbol{\iota}_{IT}}\mathrm{vec}(\mathbb{I}_{I+K})\\
&=\frac{1}{2}\sum_{j=1}^{J}\mathrm{vec}\left[\frac{\boldsymbol{\iota}_{IT}^\top\boldsymbol{\Sigma}_j^{-1}\boldsymbol{X}_j\boldsymbol{X}_j^\top\boldsymbol{\Sigma}_j^{-1}\boldsymbol{\iota}_{IT}}{\boldsymbol{\iota}_{IT}^\top\boldsymbol{\Sigma}_j^{-1}\boldsymbol{\iota}_{IT}}\right]\\
&=\frac{1}{2}\sum_{j=1}^{J}\left[\frac{\boldsymbol{\iota}_{IT}^\top\boldsymbol{\Sigma}_j^{-1}\boldsymbol{X}_j\boldsymbol{X}_j^\top\boldsymbol{\Sigma}_j^{-1}\boldsymbol{\iota}_{IT}}{\boldsymbol{\iota}_{IT}^\top\boldsymbol{\Sigma}_j^{-1}\boldsymbol{\iota}_{IT}}\right]\\
&=\frac{1}{2}\sum_{j=1}^{J}\mathrm{trace}\left[\frac{\boldsymbol{\iota}_{IT}^\top\boldsymbol{\Sigma}_j^{-1}\boldsymbol{X}_j\boldsymbol{X}_j^\top\boldsymbol{\Sigma}_j^{-1}\boldsymbol{\iota}_{IT}}{\boldsymbol{\iota}_{IT}^\top\boldsymbol{\Sigma}_j^{-1}\boldsymbol{\iota}_{IT}}\right]\\
&=\frac{1}{2}\sum_{j=1}^{J}\mathrm{trace}\left[\frac{\boldsymbol{X}_j^\top\boldsymbol{\Sigma}_j^{-1}\boldsymbol{\iota}_{IT}\boldsymbol{\iota}_{IT}^\top\boldsymbol{\Sigma}_j^{-1}\boldsymbol{X}_j}{\boldsymbol{\iota}_{IT}^\top\boldsymbol{\Sigma}_j^{-1}\boldsymbol{\iota}_{IT}}\right]\\
&\overset{H_0}{=}\frac{1}{2\sigma_\epsilon^2}\sum_{j=1}^{J}\mathrm{trace}\left[\frac{\boldsymbol{X}_j^\top\boldsymbol{\iota}_{IT}\boldsymbol{\iota}_{IT}^\top\boldsymbol{X}_j}{\boldsymbol{\iota}_{IT}^\top\boldsymbol{\iota}_{IT}}\right]
\end{aligned}
$$

$$= \frac{1}{2\sigma_\epsilon^2} \sum_{j=1}^{J} \text{trace}\left[\boldsymbol{X}_j^\top \frac{\boldsymbol{\iota}_{IT}\boldsymbol{\iota}_{IT}^\top}{IT} \boldsymbol{X}_j \right].$$

The sum of $(A1)$ and $(A2)$ is

$$-\frac{1}{2\sigma_\epsilon^2} \sum_{j=1}^{J} \text{trace}[\boldsymbol{X}_j^\top \boldsymbol{M}_0 \boldsymbol{X}_j]. \tag{3.24}$$

Applying Theorem 2.44(3) and Theorem 2.53 to $(A3)$, respectively, then $(A3)$ under the null hypothesis is

$$\begin{aligned}
&\frac{1}{2}\sum_{j=1}^{J}(\boldsymbol{\zeta}_j^\top \otimes \boldsymbol{\zeta}_j^\top)(\boldsymbol{\Sigma}_j^{-1} \otimes \boldsymbol{\Sigma}_j^{-1})(\boldsymbol{X}_j \otimes \boldsymbol{X}_j)\text{vec}(\mathbb{I}_{I+K}) \\
&\quad = \frac{1}{2}\sum_{j=1}^{J}(\boldsymbol{\zeta}_j^\top \boldsymbol{\Sigma}_j^{-1}\boldsymbol{X}_j \otimes \boldsymbol{\zeta}_j^\top \boldsymbol{\Sigma}_j^{-1}\boldsymbol{X}_j)\text{vec}(\mathbb{I}_{I+K}) \\
&\quad = \frac{1}{2}\sum_{j=1}^{J}\text{vec}[\boldsymbol{\zeta}_j^\top \boldsymbol{\Sigma}_j^{-1}\boldsymbol{X}_j\boldsymbol{X}_j^\top \boldsymbol{\Sigma}_j^{-1}\boldsymbol{\zeta}_j] \\
&\quad = \frac{1}{2}\sum_{j=1}^{J}[\boldsymbol{\zeta}_j^\top \boldsymbol{\Sigma}_j^{-1}\boldsymbol{X}_j\boldsymbol{X}_j^\top \boldsymbol{\Sigma}_j^{-1}\boldsymbol{\zeta}_j] \\
&\quad = \frac{1}{2}\sum_{j=1}^{J}\text{trace}[\boldsymbol{\zeta}_j^\top \boldsymbol{\Sigma}_j^{-1}\boldsymbol{X}_j\boldsymbol{X}_j^\top \boldsymbol{\Sigma}_j^{-1}\boldsymbol{\zeta}_j] \\
&\quad \overset{H_0}{=} \frac{1}{2\sigma_\epsilon^4}\sum_{j=1}^{J}\text{trace}[\boldsymbol{\zeta}_j^\top \boldsymbol{X}_j\boldsymbol{X}_j^\top \boldsymbol{\zeta}_j].
\end{aligned}$$

Applying Theorem 2.53, Theorem 2.41(2), and $\boldsymbol{P}_0 = \boldsymbol{\iota}_{IT}(\boldsymbol{\iota}_{IT}^\top\boldsymbol{\iota}_{IT})^{-1}\boldsymbol{\iota}_{IT}^\top$ to $(A4)$, respectively, then $(A4)$ under the null hypothesis is

$$\begin{aligned}
&\frac{1}{2}\sum_{j=1}^{J}\left[(\boldsymbol{\zeta}_j^\top\boldsymbol{\Sigma}_j^{-1}\boldsymbol{\iota}_{IT}\boldsymbol{\iota}_{IT}^\top\boldsymbol{\Sigma}_j^{-1}\boldsymbol{\zeta}_j)\frac{\boldsymbol{\iota}_{IT}^\top\boldsymbol{\Sigma}_j^{-1}\boldsymbol{X}_j \otimes \boldsymbol{\iota}_{IT}^\top\boldsymbol{\Sigma}_j^{-1}\boldsymbol{X}_j}{(\boldsymbol{\iota}_{IT}^\top\boldsymbol{\Sigma}_j^{-1}\boldsymbol{\iota}_{IT})^2}\right]\text{vec}(\mathbb{I}_{I+K}) \\
&= \frac{1}{2}\sum_{j=1}^{J}\left[\frac{\boldsymbol{\zeta}_j^\top\boldsymbol{\Sigma}_j^{-1}\boldsymbol{\iota}_{IT}\boldsymbol{\iota}_{IT}^\top\boldsymbol{\Sigma}_j^{-1}\boldsymbol{\zeta}_j\boldsymbol{\iota}_{IT}^\top\boldsymbol{\Sigma}_j^{-1}\boldsymbol{X}_j \otimes \boldsymbol{\iota}_{IT}^\top\boldsymbol{\Sigma}_j^{-1}\boldsymbol{X}_j}{(\boldsymbol{\iota}_{IT}^\top\boldsymbol{\Sigma}_j^{-1}\boldsymbol{\iota}_{IT})^2}\right]\text{vec}(\mathbb{I}_{I+K}) \\
&= \frac{1}{2}\sum_{j=1}^{J}\text{vec}\left[\frac{\boldsymbol{\iota}_{IT}^\top\boldsymbol{\Sigma}_j^{-1}\boldsymbol{X}_j\boldsymbol{X}_j^\top\boldsymbol{\Sigma}_j^{-1}\boldsymbol{\iota}_{IT}\boldsymbol{\zeta}_j^\top\boldsymbol{\Sigma}_j^{-1}\boldsymbol{\iota}_{IT}\boldsymbol{\iota}_{IT}^\top\boldsymbol{\Sigma}_j^{-1}\boldsymbol{\zeta}_j}{(\boldsymbol{\iota}_{IT}^\top\boldsymbol{\Sigma}_j^{-1}\boldsymbol{\iota}_{IT})^2}\right]
\end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{2}\sum_{j=1}^{J}\left[\frac{\boldsymbol{\iota}_{IT}^{\top}\boldsymbol{\Sigma}_j^{-1}\boldsymbol{X}_j\boldsymbol{X}_j^{\top}\boldsymbol{\Sigma}_j^{-1}\boldsymbol{\iota}_{IT}^{\top}\boldsymbol{\zeta}_j^{\top}\boldsymbol{\Sigma}_j^{-1}\boldsymbol{\iota}_{IT}\boldsymbol{\iota}_{IT}^{\top}\boldsymbol{\Sigma}_j^{-1}\boldsymbol{\zeta}_j}{(\boldsymbol{\iota}_{IT}^{\top}\boldsymbol{\Sigma}_j^{-1}\boldsymbol{\iota}_{IT})^2}\right]\\
&= \frac{1}{2}\sum_{j=1}^{J}\operatorname{trace}\left[\frac{\boldsymbol{\iota}_{IT}^{\top}\boldsymbol{\Sigma}_j^{-1}\boldsymbol{X}_j\boldsymbol{X}_j^{\top}\boldsymbol{\Sigma}_j^{-1}\boldsymbol{\iota}_{IT}^{\top}\boldsymbol{\zeta}_j^{\top}\boldsymbol{\Sigma}_j^{-1}\boldsymbol{\iota}_{IT}\boldsymbol{\iota}_{IT}^{\top}\boldsymbol{\Sigma}_j^{-1}\boldsymbol{\zeta}_j}{(\boldsymbol{\iota}_{IT}^{\top}\boldsymbol{\Sigma}_j^{-1}\boldsymbol{\iota}_{IT})^2}\right]\\
&= \frac{1}{2}\sum_{j=1}^{J}\operatorname{trace}\left[\frac{\boldsymbol{X}_j^{\top}\boldsymbol{\Sigma}_j^{-1}\boldsymbol{\iota}_{IT}^{\top}\boldsymbol{\zeta}_j^{\top}\boldsymbol{\Sigma}_j^{-1}\boldsymbol{\iota}_{IT}\boldsymbol{\iota}_{IT}^{\top}\boldsymbol{\Sigma}_j^{-1}\boldsymbol{\zeta}_j\boldsymbol{\iota}_{IT}^{\top}\boldsymbol{\Sigma}_j^{-1}\boldsymbol{X}_j}{(\boldsymbol{\iota}_{IT}^{\top}\boldsymbol{\Sigma}_j^{-1}\boldsymbol{\iota}_{IT})^2}\right]\\
&\overset{H_0}{=} \frac{1}{2\sigma_\epsilon^4}\sum_{j=1}^{J}\operatorname{trace}\left[\frac{\boldsymbol{X}_j^{\top}\boldsymbol{\iota}_{IT}^{\top}\boldsymbol{\zeta}_j^{\top}\boldsymbol{\iota}_{IT}\boldsymbol{\iota}_{IT}^{\top}\boldsymbol{\zeta}_j\boldsymbol{\iota}_{IT}^{\top}\boldsymbol{X}_j}{(\boldsymbol{\iota}_{IT}^{\top}\boldsymbol{\iota}_{IT})^2}\right]\\
&= \frac{1}{2\sigma_\epsilon^4}\sum_{j=1}^{J}\operatorname{trace}\left[\frac{\boldsymbol{\zeta}_j^{\top}\boldsymbol{\iota}_{IT}\boldsymbol{\iota}_{IT}^{\top}\boldsymbol{\zeta}_j\boldsymbol{\iota}_{IT}^{\top}\boldsymbol{X}_j\boldsymbol{X}_j^{\top}\boldsymbol{\iota}_{IT}^{\top}}{(\boldsymbol{\iota}_{IT}^{\top}\boldsymbol{\iota}_{IT})^2}\right]\\
&= \frac{1}{2\sigma_\epsilon^4}\sum_{j=1}^{J}\operatorname{trace}\left[\left(\frac{\boldsymbol{\zeta}_j^{\top}\boldsymbol{\iota}_{IT}}{\boldsymbol{\iota}_{IT}^{\top}\boldsymbol{\iota}_{IT}}\right)^2\boldsymbol{\iota}_{IT}^{\top}\boldsymbol{X}_j\boldsymbol{X}_j^{\top}\boldsymbol{\iota}_{IT}^{\top}\right]\\
&= \frac{1}{2\sigma_\epsilon^4}\sum_{j=1}^{J}\operatorname{trace}\left[[\boldsymbol{\zeta}_j^{\top}\boldsymbol{\iota}_{IT}(\boldsymbol{\iota}_{IT}^{\top}\boldsymbol{\iota}_{IT})^{-1}]^{\top}[\boldsymbol{\zeta}_j^{\top}\boldsymbol{\iota}_{IT}(\boldsymbol{\iota}_{IT}^{\top}\boldsymbol{\iota}_{IT})^{-1}]\boldsymbol{\iota}_{IT}^{\top}\boldsymbol{X}_j\boldsymbol{X}_j^{\top}\boldsymbol{\iota}_{IT}\right]\\
&= \frac{1}{2\sigma_\epsilon^4}\sum_{j=1}^{J}\operatorname{trace}\left[[\boldsymbol{\zeta}_j^{\top}\boldsymbol{\iota}_{IT}(\boldsymbol{\iota}_{IT}^{\top}\boldsymbol{\iota}_{IT})^{-1}]\boldsymbol{\iota}_{IT}^{\top}\boldsymbol{X}_j\boldsymbol{X}_j^{\top}\boldsymbol{\iota}_{IT}[\boldsymbol{\zeta}_j^{\top}\boldsymbol{\iota}_{IT}(\boldsymbol{\iota}_{IT}^{\top}\boldsymbol{\iota}_{IT})^{-1}]^{\top}\right]\\
&= \frac{1}{2\sigma_\epsilon^4}\sum_{j=1}^{J}\operatorname{trace}\left[\boldsymbol{\zeta}_j^{\top}\boldsymbol{\iota}_{IT}(\boldsymbol{\iota}_{IT}^{\top}\boldsymbol{\iota}_{IT})^{-1}\boldsymbol{\iota}_{IT}^{\top}\boldsymbol{X}_j\boldsymbol{X}_j^{\top}\boldsymbol{\iota}_{IT}(\boldsymbol{\iota}_{IT}^{\top}\boldsymbol{\iota}_{IT})^{-1}\boldsymbol{\iota}_{IT}^{\top}\boldsymbol{\zeta}_j\right]\\
&= \frac{1}{2\sigma_\epsilon^4}\sum_{j=1}^{J}\operatorname{trace}\left[\boldsymbol{\zeta}_j^{\top}\boldsymbol{P}_0\boldsymbol{X}_j\boldsymbol{X}_j^{\top}\boldsymbol{P}_0\boldsymbol{\zeta}_j\right].
\end{aligned}
$$

Applying Theorem 2.44(3), Theorem 2.53, and $\boldsymbol{P}_0 = \boldsymbol{\iota}_{IT}(\boldsymbol{\iota}_{IT}^{\top}\boldsymbol{\iota}_{IT})^{-1}\boldsymbol{\iota}_{IT}^{\top}$ to $(A5)$, respectively, then $(A5)$ under the null hypothesis is

$$
\begin{aligned}
&-\frac{1}{2}\sum_{j=1}^{J}\left(\frac{\boldsymbol{\zeta}_j^{\top}\boldsymbol{\Sigma}_j^{-1}\boldsymbol{\iota}_{IT}\boldsymbol{\iota}_{IT}^{\top}\boldsymbol{\Sigma}_j^{-1}\otimes\boldsymbol{\zeta}_j^{\top}\boldsymbol{\Sigma}_j^{-1}}{\boldsymbol{\iota}_{IT}^{\top}\boldsymbol{\Sigma}_j^{-1}\boldsymbol{\iota}_{IT}}(\boldsymbol{X}_j\otimes\boldsymbol{X}_j)\operatorname{vec}(\mathbb{I}_{I+K})\right)\\
&\qquad = -\frac{1}{2}\sum_{j=1}^{J}\left[\frac{\boldsymbol{\zeta}_j^{\top}\boldsymbol{\Sigma}_j^{-1}\boldsymbol{\iota}_{IT}\boldsymbol{\iota}_{IT}^{\top}\boldsymbol{\Sigma}_j^{-1}\boldsymbol{X}_j\otimes\boldsymbol{\zeta}_j^{\top}\boldsymbol{\Sigma}_j^{-1}\boldsymbol{X}_j}{\boldsymbol{\iota}_{IT}^{\top}\boldsymbol{\Sigma}_j^{-1}\boldsymbol{\iota}_{IT}}\right]\operatorname{vec}(\mathbb{I}_{I+K})\\
&\qquad = -\frac{1}{2}\sum_{j=1}^{J}\operatorname{vec}\left[\frac{\boldsymbol{\zeta}_j^{\top}\boldsymbol{\Sigma}_j^{-1}\boldsymbol{X}_j\boldsymbol{X}_j^{\top}\boldsymbol{\Sigma}_j^{-1}\boldsymbol{\iota}_{IT}\boldsymbol{\iota}_{IT}^{\top}\boldsymbol{\Sigma}_j^{-1}\boldsymbol{\zeta}_j}{\boldsymbol{\iota}_{IT}^{\top}\boldsymbol{\Sigma}_j^{-1}\boldsymbol{\iota}_{IT}}\right]
\end{aligned}
$$

$$= -\frac{1}{2}\sum_{j=1}^{J}\left[\frac{\boldsymbol{\zeta}_j^\top\boldsymbol{\Sigma}_j^{-1}\boldsymbol{X}_j\boldsymbol{X}_j^\top\boldsymbol{\Sigma}_j^{-1}\boldsymbol{\iota}_{IT}\boldsymbol{\iota}_{IT}^\top\boldsymbol{\Sigma}_j^{-1}\boldsymbol{\zeta}_j}{\boldsymbol{\iota}_{IT}^\top\boldsymbol{\Sigma}_j^{-1}\boldsymbol{\iota}_{IT}}\right]$$

$$= -\frac{1}{2}\sum_{j=1}^{J}\text{trace}\left[\frac{\boldsymbol{\zeta}_j^\top\boldsymbol{\Sigma}_j^{-1}\boldsymbol{X}_j\boldsymbol{X}_j^\top\boldsymbol{\Sigma}_j^{-1}\boldsymbol{\iota}_{IT}\boldsymbol{\iota}_{IT}^\top\boldsymbol{\Sigma}_j^{-1}\boldsymbol{\zeta}_j}{\boldsymbol{\iota}_{IT}^\top\boldsymbol{\Sigma}_j^{-1}\boldsymbol{\iota}_{IT}}\right]$$

$$\stackrel{H_0}{=} -\frac{1}{2\sigma_\epsilon^4}\sum_{j=1}^{J}\text{trace}\left[\boldsymbol{\zeta}_j^\top\boldsymbol{X}_j\boldsymbol{X}_j^\top\boldsymbol{\iota}_{IT}(\boldsymbol{\iota}_{IT}^\top\boldsymbol{\iota}_{IT})^{-1}\boldsymbol{\iota}_{IT}^\top\boldsymbol{\zeta}_j\right]$$

$$= -\frac{1}{2\sigma_\epsilon^4}\sum_{j=1}^{J}\text{trace}\left[\boldsymbol{\zeta}_j^\top\boldsymbol{X}_j\boldsymbol{X}_j^\top\boldsymbol{P}_0\boldsymbol{\zeta}_j\right].$$

Applying Theorem 2.44(3), Theorem 2.53, and $\boldsymbol{P}_0 = \boldsymbol{\iota}_{IT}(\boldsymbol{\iota}_{IT}^\top\boldsymbol{\iota}_{IT})^{-1}\boldsymbol{\iota}_{IT}^\top$ to $(A6)$, respectively, then $(A6)$ under the null hypothesis is

$$-\frac{1}{2}\sum_{j=1}^{J}\left(\frac{\boldsymbol{\zeta}_j^\top\boldsymbol{\Sigma}_j^{-1}\otimes\boldsymbol{\zeta}_j^\top\boldsymbol{\Sigma}_j^{-1}\boldsymbol{\iota}_{IT}\boldsymbol{\iota}_{IT}^\top\boldsymbol{\Sigma}_j^{-1}}{\boldsymbol{\iota}_{IT}^\top\boldsymbol{\Sigma}_j^{-1}\boldsymbol{\iota}_{IT}}(\boldsymbol{X}_j\otimes\boldsymbol{X}_j)\text{vec}(\mathbb{I}_{I+K})\right)$$

$$= -\frac{1}{2}\sum_{j=1}^{J}\left[\frac{\boldsymbol{\zeta}_j^\top\boldsymbol{\Sigma}_j^{-1}\boldsymbol{X}_j\otimes\boldsymbol{\zeta}_j^\top\boldsymbol{\Sigma}_j^{-1}\boldsymbol{\iota}_{IT}\boldsymbol{\iota}_{IT}^\top\boldsymbol{\Sigma}_j^{-1}\boldsymbol{X}_j}{\boldsymbol{\iota}_{IT}^\top\boldsymbol{\Sigma}_j^{-1}\boldsymbol{\iota}_{IT}}\right]\text{vec}(\mathbb{I}_{I+K})$$

$$= -\frac{1}{2}\sum_{j=1}^{J}\text{vec}\left[\frac{\boldsymbol{\zeta}_j^\top\boldsymbol{\Sigma}_j^{-1}\boldsymbol{\iota}_{IT}\boldsymbol{\iota}_{IT}^\top\boldsymbol{\Sigma}_j^{-1}\boldsymbol{X}_j\boldsymbol{X}_j^\top\boldsymbol{\Sigma}_j^{-1}\boldsymbol{\zeta}_j}{\boldsymbol{\iota}_{IT}^\top\boldsymbol{\Sigma}_j^{-1}\boldsymbol{\iota}_{IT}}\right]$$

$$= -\frac{1}{2}\sum_{j=1}^{J}\left[\frac{\boldsymbol{\zeta}_j^\top\boldsymbol{\Sigma}_j^{-1}\boldsymbol{\iota}_{IT}\boldsymbol{\iota}_{IT}^\top\boldsymbol{\Sigma}_j^{-1}\boldsymbol{X}_j\boldsymbol{X}_j^\top\boldsymbol{\Sigma}_j^{-1}\boldsymbol{\zeta}_j}{\boldsymbol{\iota}_{IT}^\top\boldsymbol{\Sigma}_j^{-1}\boldsymbol{\iota}_{IT}}\right]$$

$$= -\frac{1}{2}\sum_{j=1}^{J}\text{trace}\left[\frac{\boldsymbol{\zeta}_j^\top\boldsymbol{\Sigma}_j^{-1}\boldsymbol{\iota}_{IT}\boldsymbol{\iota}_{IT}^\top\boldsymbol{\Sigma}_j^{-1}\boldsymbol{X}_j\boldsymbol{X}_j^\top\boldsymbol{\Sigma}_j^{-1}\boldsymbol{\zeta}_j}{\boldsymbol{\iota}_{IT}^\top\boldsymbol{\Sigma}_j^{-1}\boldsymbol{\iota}_{IT}}\right]$$

$$\stackrel{H_0}{=} -\frac{1}{2\sigma_\epsilon^4}\sum_{j=1}^{J}\text{trace}\left[\boldsymbol{\zeta}_j^\top\boldsymbol{\iota}_{IT}(\boldsymbol{\iota}_{IT}^\top\boldsymbol{\iota}_{IT})^{-1}\boldsymbol{\iota}_{IT}^\top\boldsymbol{X}_j\boldsymbol{X}_j^\top\boldsymbol{\zeta}_j\right]$$

$$= -\frac{1}{2\sigma_\epsilon^4}\sum_{j=1}^{J}\text{trace}\left[\boldsymbol{\zeta}_j^\top\boldsymbol{P}_0\boldsymbol{X}_j\boldsymbol{X}_j^\top\boldsymbol{\zeta}_j\right].$$

The sum of $(A3), (A4), (A5)$, and $(A6)$ is

$$\frac{1}{2\sigma_\epsilon^4}\sum_{j=1}^{J}\text{trace}[\boldsymbol{\zeta}_j^\top\boldsymbol{X}_j\boldsymbol{X}_j^\top\boldsymbol{\zeta}_j] + \text{trace}\left[\boldsymbol{\zeta}_j^\top\boldsymbol{P}_0\boldsymbol{X}_j\boldsymbol{X}_j^\top\boldsymbol{P}_0\boldsymbol{\zeta}_j\right]$$

$$- \text{trace}\left[\boldsymbol{\zeta}_j^\top\boldsymbol{X}_j\boldsymbol{X}_j^\top\boldsymbol{P}_0\boldsymbol{\zeta}_j\right] - \text{trace}\left[\boldsymbol{\zeta}_j^\top\boldsymbol{P}_0\boldsymbol{X}_j\boldsymbol{X}_j^\top\boldsymbol{\zeta}_j\right]$$

$$
\begin{aligned}
&= \frac{1}{2\sigma_\epsilon^4}\sum_{j=1}^{J} \text{trace}[\boldsymbol{\zeta}_j^\top \boldsymbol{X}_j \boldsymbol{X}_j^\top (\mathbb{I}_{IT} - \boldsymbol{P}_0)\boldsymbol{\zeta}_j] - \text{trace}\left[\boldsymbol{\zeta}_j^\top \boldsymbol{P}_0 \boldsymbol{X}_j \boldsymbol{X}_j^\top (\mathbb{I}_{IT} - \boldsymbol{P}_0)\boldsymbol{\zeta}_j\right] \\
&= \frac{1}{2\sigma_\epsilon^4}\sum_{j=1}^{J} \text{trace}[\boldsymbol{\zeta}_j^\top \boldsymbol{X}_j \boldsymbol{X}_j^\top (\mathbb{I}_{IT} - \boldsymbol{\iota}_{IT}(\boldsymbol{\iota}_{IT}^\top\boldsymbol{\iota}_{IT})^{-1}\boldsymbol{\iota}_{IT}^\top)\boldsymbol{\zeta}_j] \\
&\quad - \text{trace}\left[\boldsymbol{\zeta}_j^\top \boldsymbol{P}_0 \boldsymbol{X}_j \boldsymbol{X}_j^\top (\mathbb{I}_{IT} - \boldsymbol{\iota}_{IT}(\boldsymbol{\iota}_{IT}^\top\boldsymbol{\iota}_{IT})^{-1}\boldsymbol{\iota}_{IT}^\top)\boldsymbol{\zeta}_j\right] \\
&= \frac{1}{2\sigma_\epsilon^4}\sum_{j=1}^{J} \text{trace}[\boldsymbol{\zeta}_j^\top \boldsymbol{X}_j \boldsymbol{X}_j^\top \boldsymbol{M}_0\boldsymbol{\zeta}_j] - \text{trace}\left[\boldsymbol{\zeta}_j^\top \boldsymbol{P}_0 \boldsymbol{X}_j \boldsymbol{X}_j^\top \boldsymbol{M}_0\boldsymbol{\zeta}_j\right] \\
&= \frac{1}{2\sigma_\epsilon^4}\sum_{j=1}^{J} \text{trace}[\boldsymbol{\zeta}_j^\top (\mathbb{I}_{IT} - \boldsymbol{P}_0)\boldsymbol{X}_j \boldsymbol{X}_j^\top \boldsymbol{M}_0\boldsymbol{\zeta}_j] \\
&= \frac{1}{2\sigma_\epsilon^4}\sum_{j=1}^{J} \text{trace}[\boldsymbol{\zeta}_j^\top (\mathbb{I}_{IT} - \boldsymbol{\iota}_{IT}(\boldsymbol{\iota}_{IT}^\top\boldsymbol{\iota}_{IT})^{-1}\boldsymbol{\iota}_{IT}^\top)\boldsymbol{X}_j \boldsymbol{X}_j^\top \boldsymbol{M}_0\boldsymbol{\zeta}_j] \\
&= \frac{1}{2\sigma_\epsilon^4}\sum_{j=1}^{J} \text{trace}[\boldsymbol{\zeta}_j^\top \boldsymbol{M}_0\boldsymbol{X}_j \boldsymbol{X}_j^\top \boldsymbol{M}_0\boldsymbol{\zeta}_j] \\
&= \frac{1}{2\sigma_\epsilon^4}\sum_{j=1}^{J} \boldsymbol{\zeta}_j^\top \boldsymbol{M}_0\boldsymbol{X}_j \boldsymbol{X}_j^\top \boldsymbol{M}_0\boldsymbol{\zeta}_j. \qquad (3.25)
\end{aligned}
$$

The sum of (3.24) and (3.25) is

$$
\frac{1}{2\sigma^4}\sum_{j=1}^{J}\left[\boldsymbol{\zeta}_j^\top \boldsymbol{M}_0\boldsymbol{X}_j \boldsymbol{X}_j^\top \boldsymbol{M}_0\boldsymbol{\zeta}_j - \sigma_\epsilon^2 \text{trace}(\boldsymbol{X}_j^\top \boldsymbol{M}_0 \boldsymbol{X}_j)\right]. \qquad (3.26)
$$

Since $\boldsymbol{M}_0\boldsymbol{\iota}_{IT} = 0$,

$$
\begin{aligned}
\boldsymbol{M}_0\boldsymbol{\zeta}_j &= \boldsymbol{M}_0(\boldsymbol{y}_j - \boldsymbol{X}_j\boldsymbol{\beta}) \\
&= \boldsymbol{M}_0(\boldsymbol{y}_j - \gamma_j\boldsymbol{\iota}_{IT} - \boldsymbol{X}_j\boldsymbol{\beta}) \\
&= \boldsymbol{M}_0\boldsymbol{\epsilon}_j. \qquad (3.27)
\end{aligned}
$$

Therefore, substituting (3.27) into (3.26), under the null hypothesis, the score is

$$
\frac{1}{2\sigma^4}\sum_{j=1}^{J}\left[\boldsymbol{\epsilon}_j^\top \boldsymbol{M}_0\boldsymbol{X}_j \boldsymbol{X}_j^\top \boldsymbol{M}_0\boldsymbol{\epsilon}_j - \sigma_\epsilon^2 \text{trace}(\boldsymbol{X}_j^\top \boldsymbol{M}_0 \boldsymbol{X}_j)\right]
$$

which is proportional to

$$\sum_{j=1}^{J} \left[\boldsymbol{\epsilon}_j^\top \boldsymbol{M}_0 \boldsymbol{X}_j \boldsymbol{X}_j^\top \boldsymbol{M}_0 \boldsymbol{\epsilon}_j - \sigma_\epsilon^2 \text{trace}(\boldsymbol{X}_j^\top \boldsymbol{M}_0 \boldsymbol{X}_j) \right]. \tag{3.28}$$

□

The score function in Theorem 3.5 suggests a statistic that is valid under $E[\boldsymbol{\epsilon}_j \boldsymbol{\epsilon}_j^\top | \boldsymbol{X}_j] = \sigma_\epsilon^2 \mathbb{I}_{IT}$. Under the null hypothesis and the fact that $\boldsymbol{M}_0 \boldsymbol{\zeta}_j = \boldsymbol{M}_0 \boldsymbol{\epsilon}_j$, we have

$$\begin{aligned}
E[\boldsymbol{\zeta}_j^\top \boldsymbol{M}_0 \boldsymbol{X}_j \boldsymbol{X}_j^\top \boldsymbol{M}_0 \boldsymbol{\zeta}_j | \boldsymbol{X}_j] &= \text{trace} E[\boldsymbol{\epsilon}_j^\top \boldsymbol{M}_0 \boldsymbol{X}_j \boldsymbol{X}_j^\top \boldsymbol{M}_0 \boldsymbol{\epsilon}_j | \boldsymbol{X}_j] \\
&= \text{trace} E[\boldsymbol{\epsilon}_j \boldsymbol{\epsilon}_j^\top \boldsymbol{M}_0 \boldsymbol{X}_j \boldsymbol{X}_j^\top \boldsymbol{M}_0 | \boldsymbol{X}_j] \\
&= \text{trace}(\sigma_\epsilon^2 \boldsymbol{M}_0 \boldsymbol{X}_j \boldsymbol{X}_j^\top \boldsymbol{M}_0) \\
&= \text{trace}(\sigma_\epsilon^2 \boldsymbol{X}_j^\top \boldsymbol{M}_0 \boldsymbol{M}_0 \boldsymbol{X}_j) \\
&= \text{trace}(\sigma_\epsilon^2 \boldsymbol{X}_j^\top \boldsymbol{M}_0^\top \boldsymbol{M}_0 \boldsymbol{X}_j) \\
&= \text{trace}(\sigma_\epsilon^2 \boldsymbol{X}_j^\top \boldsymbol{M}_0 \boldsymbol{X}_j) \\
&= \sigma_\epsilon^2 \text{trace}(\boldsymbol{X}_j^\top \boldsymbol{M}_0 \boldsymbol{X}_j),
\end{aligned}$$

where we have used $E[\boldsymbol{\epsilon}_j \boldsymbol{\epsilon}_j^\top | \boldsymbol{X}_j] = \sigma_\epsilon^2 \mathbb{I}_{IT}$, Theorem 2.41(2), Proposition 3.1(2) and Proposition 3.1(3).

Hence, from the above,

$$\begin{aligned}
&E\left[\boldsymbol{\epsilon}_j^\top \boldsymbol{M}_0 \boldsymbol{X}_j \boldsymbol{X}_j^\top \boldsymbol{M}_0 \boldsymbol{\epsilon}_j - \sigma_\epsilon^2 \text{trace}(\boldsymbol{X}_j^\top \boldsymbol{M}_0 \boldsymbol{X}_j) | \boldsymbol{X}_j\right] \\
&= E\left[\boldsymbol{\epsilon}_j^\top \boldsymbol{M}_0 \boldsymbol{X}_j \boldsymbol{X}_j^\top \boldsymbol{M}_0 \boldsymbol{\epsilon}_j | \boldsymbol{X}_j\right] - \sigma_\epsilon^2 \text{trace}(\boldsymbol{X}_j^\top \boldsymbol{M}_0 \boldsymbol{X}_j) \\
&= \sigma_\epsilon^2 \text{trace}(\boldsymbol{X}_j^\top \boldsymbol{M}_0 \boldsymbol{X}_j) - \sigma_\epsilon^2 \text{trace}(\boldsymbol{X}_j^\top \boldsymbol{M}_0 \boldsymbol{X}_j) \\
&= 0,
\end{aligned} \tag{3.29}$$

that is each term in the sum of (3.28) has mean zero, and we can apply a central limit theorem. In this discussion, we propose a statistic, CLM, when $IT - 1 > I + K$ for which the parameter estimates $\hat{\gamma}_j$ and $\hat{\beta}_j$ are well-defined. The statistic is defined as

$$CLM = \frac{\frac{1}{\sqrt{J}} \sum_{j=1}^{J} \left[\tilde{\boldsymbol{\epsilon}}_j^\top \boldsymbol{M}_0 \boldsymbol{X}_j \boldsymbol{X}_j^\top \boldsymbol{M}_0 \tilde{\boldsymbol{\epsilon}}_j - \tilde{\sigma}_j^2 \text{trace}(\boldsymbol{X}_j^\top \boldsymbol{M}_0 \boldsymbol{X}_j) \right]}{\tilde{\sigma}_{clm}},$$

where $\tilde{\boldsymbol{\epsilon}}_j = \boldsymbol{y}_j - \hat{\gamma}_j \boldsymbol{\iota}_{IT} - \boldsymbol{X}_j \hat{\boldsymbol{\beta}}_j,$

$$\tilde{\sigma}_j^2 = \frac{\tilde{\boldsymbol{\epsilon}}_j^\top \tilde{\boldsymbol{\epsilon}}_j}{IT - I - K - 1},$$

and

$$\tilde{\sigma}_{clm}^2 = \frac{1}{J} \sum_{j=1}^{J} \left[\tilde{\boldsymbol{\epsilon}}_j^\top \boldsymbol{M}_0 \boldsymbol{X}_j \boldsymbol{X}_j^\top \boldsymbol{M}_0 \tilde{\boldsymbol{\epsilon}}_j - \tilde{\sigma}_j^2 \text{trace}(\boldsymbol{X}_j^\top \boldsymbol{M}_0 \boldsymbol{X}_j) \right]^2.$$

In order to study asymptotic properties of the statistic CLM, we require the following assumptions

Assumption 1. *Conditional on* $\boldsymbol{X}$, ϵ_{ijt} *are independent across* i, j *and* t, *and* $E(\epsilon_{ijt}|\boldsymbol{X}) = 0$ *where* $\boldsymbol{X} = (\boldsymbol{X}_1^\top, ..., \boldsymbol{X}_J^\top)^\top$.

Assumption 2. *The matrix*

$$\frac{1}{J} \sum_{j=1}^{J} \boldsymbol{X}_j^\top \boldsymbol{M}_0 \boldsymbol{X}_j$$

is positive definite and converges to a non-stochastic positive definite matrix in the limit.

Assumption 3. $0 < E[(\boldsymbol{\epsilon}_j^\top \boldsymbol{\epsilon}_j)^4 | \boldsymbol{X}_j] < J_1 < \infty$ *and* $E[\text{trace}[(\boldsymbol{X}_j^\top \boldsymbol{M}_0 \boldsymbol{X}_j)^4]] < J_2 < \infty$ *for each* j.

Under assumptions 1-3, we obtain the limiting distribution of CLM as the following theorem.

Theorem 3.6. *Suppose that assumptions 1-3 hold,* $E[\boldsymbol{\epsilon}_j \boldsymbol{\epsilon}_j^\top | \boldsymbol{X}_j] = \sigma_\epsilon^2 \mathbb{I}_{IT}$ *and that* $J \to \infty$. *Then under the null hypothesis of* $\boldsymbol{\Sigma}_\omega = 0$,

$$CLM \xrightarrow{d} N(0, 1).$$

Proof. The fixed effect residuals with group means subtracted are given by

$$\begin{aligned}\tilde{\boldsymbol{\epsilon}}_j &= \boldsymbol{y}_j - \hat{\gamma}_j \boldsymbol{\iota}_{IT} - \boldsymbol{X}_j \hat{\boldsymbol{\beta}}_j \\ &= \gamma_j \boldsymbol{\iota}_{IT} + \boldsymbol{X}_j \boldsymbol{\beta} + \boldsymbol{\epsilon}_j - \hat{\gamma}_j \boldsymbol{\iota}_{IT} - \boldsymbol{X}_j \hat{\boldsymbol{\beta}}_j \\ &= \boldsymbol{\epsilon}_j + (\gamma_j - \hat{\gamma}_j)\boldsymbol{\iota}_{IT} + \boldsymbol{X}_j(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}_j). \end{aligned} \tag{3.30}$$

Using (3.30), we consider

$$\begin{aligned}&\frac{1}{\sqrt{J}}\sum_{j=1}^{J} \tilde{\boldsymbol{\epsilon}}_j^\top \boldsymbol{M}_0 \boldsymbol{X}_j \boldsymbol{X}_j^\top \boldsymbol{M}_0 \tilde{\boldsymbol{\epsilon}}_j \\ &= \frac{1}{\sqrt{J}}\sum_{j=1}^{J} (\boldsymbol{\epsilon}_j + (\gamma_j - \hat{\gamma}_j)\boldsymbol{\iota}_{IT} + \boldsymbol{X}_j(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}_j))^\top \boldsymbol{M}_0 \boldsymbol{X}_j \boldsymbol{X}_j^\top \boldsymbol{M}_0 (\boldsymbol{\epsilon}_j + (\gamma_j - \hat{\gamma}_j)\boldsymbol{\iota}_{IT} \\ &\quad + \boldsymbol{X}_j(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}_j)) \\ &= \frac{1}{\sqrt{J}}\sum_{j=1}^{J} \boldsymbol{\epsilon}_j^\top \boldsymbol{M}_0 \boldsymbol{X}_j \boldsymbol{X}_j^\top \boldsymbol{M}_0 \boldsymbol{\epsilon}_j && (3.31) \\ &\quad + \frac{1}{\sqrt{J}}\sum_{j=1}^{J} 2\boldsymbol{\epsilon}_j^\top \boldsymbol{M}_0 \boldsymbol{X}_j \boldsymbol{X}_j^\top \boldsymbol{M}_0 ((\gamma_j - \hat{\gamma}_j)\boldsymbol{\iota}_{IT} + \boldsymbol{X}_j(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}_j)) && (3.32) \\ &\quad + \frac{1}{\sqrt{J}}\sum_{j=1}^{J} (((\gamma_j - \hat{\gamma}_j)\boldsymbol{\iota}_{IT} + \boldsymbol{X}_j(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}_j)))^\top \boldsymbol{M}_0 \boldsymbol{X}_j \boldsymbol{X}_j^\top \boldsymbol{M}_0 ((\gamma_j - \hat{\gamma}_j)\boldsymbol{\iota}_{IT} + \boldsymbol{X}_j(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}_j)). \\ &&& (3.33)\end{aligned}$$

From Assumption 2 and Theorem 2.63, $(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}_j)$ and $(\gamma_j - \hat{\gamma}_j)$ are $O_p(J^{-1/2})$. Finally, we use Theorem 2.27 in (3.32) and (3.33), then we can see that (3.32) and (3.33) are stochastically bounded $O_p(J^{-1/2})$. Hence, the term

$$\frac{1}{\sqrt{J}}\sum_{j=1}^{J} \tilde{\boldsymbol{\epsilon}}_j^\top \boldsymbol{M}_0 \boldsymbol{X}_j \boldsymbol{X}_j^\top \boldsymbol{M}_0 \tilde{\boldsymbol{\epsilon}}_j = \frac{1}{\sqrt{J}}\sum_{j=1}^{J} \boldsymbol{\epsilon}_j^\top \boldsymbol{M}_0 \boldsymbol{X}_j \boldsymbol{X}_j^\top \boldsymbol{M}_0 \boldsymbol{\epsilon}_j + O_p(J^{-1/2}). \tag{3.34}$$

Similarly, from the definition of $\tilde{\sigma}_j^2$ and (3.30), we have

$$\frac{1}{\sqrt{J}}\sum_{j=1}^{J} \tilde{\sigma}_j^2 \text{trace}(\boldsymbol{X}_j^\top \boldsymbol{M}_0 \boldsymbol{X}_j)$$

$$= \frac{1}{\sqrt{J}} \sum_{j=1}^{J} \frac{\tilde{\boldsymbol{\epsilon}}_j^\top \tilde{\boldsymbol{\epsilon}}_j}{IT - I - K - 1} \text{trace}(\boldsymbol{X}_j^\top \boldsymbol{M}_0 \boldsymbol{X}_j)$$

$$= \frac{1}{\sqrt{J}} \sum_{j=1}^{J} \frac{[\boldsymbol{\epsilon}_j + (\gamma_j - \hat{\gamma}_j)\boldsymbol{\iota}_{IT} + \boldsymbol{X}_j(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}_j)]^\top [\boldsymbol{\epsilon}_j + (\gamma_j - \hat{\gamma}_j)\boldsymbol{\iota}_{IT} + \boldsymbol{X}_j(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}_j)]}{IT - I - K - 1}$$

$$\text{trace}(\boldsymbol{X}_j^\top \boldsymbol{M}_0 \boldsymbol{X}_j)$$

$$= \frac{1}{\sqrt{J}} \sum_{j=1}^{J} \frac{\boldsymbol{\epsilon}_j^\top \boldsymbol{\epsilon}_j}{IT - I - K - 1} \text{trace}(\boldsymbol{X}_j^\top \boldsymbol{M}_0 \boldsymbol{X}_j)$$

$$+ \frac{1}{\sqrt{J}} \sum_{j=1}^{J} 2\boldsymbol{\epsilon}_j^\top [(\gamma_j - \hat{\gamma}_j)\boldsymbol{\iota}_{IT} + \boldsymbol{X}_j(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}_j)] \text{trace}(\boldsymbol{X}_j^\top \boldsymbol{M}_0 \boldsymbol{X}_j)$$

$$+ \frac{1}{\sqrt{J}} \sum_{j=1}^{J} [(\gamma_j - \hat{\gamma}_j)\boldsymbol{\iota}_{IT} + \boldsymbol{X}_j(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}_j)]^\top [(\gamma_j - \hat{\gamma}_j)\boldsymbol{\iota}_{IT} + \boldsymbol{X}_j(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}_j)] \text{trace}(\boldsymbol{X}_j^\top \boldsymbol{M}_0 \boldsymbol{X}_j)$$

$$= \frac{1}{\sqrt{J}} \sum_{j=1}^{J} \frac{\boldsymbol{\epsilon}_j^\top \boldsymbol{\epsilon}_j}{IT - I - K - 1} \text{trace}(\boldsymbol{X}_j^\top \boldsymbol{M}_0 \boldsymbol{X}_j) + O_p(J^{-1/2}). \tag{3.35}$$

Combining the term (3.34) and (3.35) gives us

$$\frac{1}{\sqrt{J}} \sum_{j=1}^{J} \left[\tilde{\boldsymbol{\epsilon}}_j^\top \boldsymbol{M}_0 \boldsymbol{X}_j \boldsymbol{X}_j^\top \boldsymbol{M}_0 \tilde{\boldsymbol{\epsilon}}_j - \tilde{\sigma}_j^2 \text{trace}(\boldsymbol{X}_j^\top \boldsymbol{M}_0 \boldsymbol{X}_j)\right]$$

$$= \frac{1}{\sqrt{J}} \sum_{j=1}^{J} \left[\boldsymbol{\epsilon}_j^\top \boldsymbol{M}_0 \boldsymbol{X}_j \boldsymbol{X}_j^\top \boldsymbol{M}_0 \boldsymbol{\epsilon}_j - \frac{\boldsymbol{\epsilon}_j^\top \boldsymbol{\epsilon}_j}{IT - I - K - 1} \text{trace}(\boldsymbol{X}_j^\top \boldsymbol{M}_0 \boldsymbol{X}_j)\right] + O_p(J^{-1/2}). \tag{3.36}$$

Let

$$\lim_{J \to \infty} \frac{1}{J} \sum_{j=1}^{J} Var \left[\boldsymbol{\epsilon}_j^\top \boldsymbol{M}_0 \boldsymbol{X}_j \boldsymbol{X}_j^\top \boldsymbol{M}_0 \boldsymbol{\epsilon}_j - \frac{\boldsymbol{\epsilon}_j^\top \boldsymbol{\epsilon}_j}{IT - I - K - 1} \text{trace}(\boldsymbol{X}_j^\top \boldsymbol{M}_0 \boldsymbol{X}_j)\right] = \sigma_{clm}^2.$$

Since

$$E \left[\boldsymbol{\epsilon}_j^\top \boldsymbol{M}_0 \boldsymbol{X}_j \boldsymbol{X}_j^\top \boldsymbol{M}_0 \boldsymbol{\epsilon}_j - \frac{\boldsymbol{\epsilon}_j^\top \boldsymbol{\epsilon}_j}{IT - I - K - 1} \text{trace}(\boldsymbol{X}_j^\top \boldsymbol{M}_0 \boldsymbol{X}_j)\right] = 0, \tag{3.37}$$

then

$$\lim_{J\to\infty} \frac{1}{J}\sum_{j=1}^{J} E\left[\boldsymbol{\epsilon}_j^\top \boldsymbol{M}_0 \boldsymbol{X}_j \boldsymbol{X}_j^\top \boldsymbol{M}_0 \boldsymbol{\epsilon}_j - \frac{\boldsymbol{\epsilon}_j^\top \boldsymbol{\epsilon}_j}{IT-I-K-1}\text{trace}(\boldsymbol{X}_j^\top \boldsymbol{M}_0 \boldsymbol{X}_j)\right]^2 = \sigma_{clm}^2.$$

We use the fact that

$$\frac{1}{\sqrt{J}}\sum_{j=1}^{J}\left[\boldsymbol{\epsilon}_j^\top \boldsymbol{M}_0 \boldsymbol{X}_j \boldsymbol{X}_j^\top \boldsymbol{M}_0 \boldsymbol{\epsilon}_j - \frac{\boldsymbol{\epsilon}_j^\top \boldsymbol{\epsilon}_j}{IT-I-K-1}\text{trace}(\boldsymbol{X}_j^\top \boldsymbol{M}_0 \boldsymbol{X}_j)\right]$$
$$= \sqrt{J}\left[\frac{1}{J}\sum_{j=1}^{J}\left[\boldsymbol{\epsilon}_j^\top \boldsymbol{M}_0 \boldsymbol{X}_j \boldsymbol{X}_j^\top \boldsymbol{M}_0 \boldsymbol{\epsilon}_j - \frac{\boldsymbol{\epsilon}_j^\top \boldsymbol{\epsilon}_j}{IT-I-K-1}\text{trace}(\boldsymbol{X}_j^\top \boldsymbol{M}_0 \boldsymbol{X}_j)\right]\right].$$

Therefore, we can apply the multivariate Linderberg-Feller Central Limit Theorem, then (3.36) is $N(0,\sigma_{clm}^2) + O_p(J^{-1/2})$.

It remains to show that $\tilde{\sigma}_{clm}^2 \xrightarrow{p} \sigma_{clm}^2$. From definition of $\tilde{\sigma}_{clm}^2$, we have

$$\begin{aligned}\tilde{\sigma}_{clm}^2 &= \frac{1}{J}\sum_{j=1}^{J}\left[\tilde{\boldsymbol{\epsilon}}_j^\top \boldsymbol{M}_0 \boldsymbol{X}_j \boldsymbol{X}_j^\top \boldsymbol{M}_0 \tilde{\boldsymbol{\epsilon}}_j - \tilde{\sigma}_j^2 \text{trace}(\boldsymbol{X}_j^\top \boldsymbol{M}_0 \boldsymbol{X}_j)\right]^2 \\ &= \frac{1}{J}\sum_{j=1}^{J}(\tilde{\boldsymbol{\epsilon}}_j^\top \boldsymbol{M}_0 \boldsymbol{X}_j \boldsymbol{X}_j^\top \boldsymbol{M}_0 \tilde{\boldsymbol{\epsilon}}_j)^2 \\ &\quad - \frac{2}{J}\sum_{j=1}^{J}(\tilde{\boldsymbol{\epsilon}}_j^\top \boldsymbol{M}_0 \boldsymbol{X}_j \boldsymbol{X}_j^\top \boldsymbol{M}_0 \tilde{\boldsymbol{\epsilon}}_j)\tilde{\sigma}_j^2 \text{trace}(\boldsymbol{X}_j^\top \boldsymbol{M}_0 \boldsymbol{X}_j) \\ &\quad + \frac{1}{J}\sum_{j=1}^{J}\left(\tilde{\sigma}_j^2 \text{trace}(\boldsymbol{X}_j^\top \boldsymbol{M}_0 \boldsymbol{X}_j)\right)^2. \end{aligned} \tag{3.38}$$

Using (3.30), we consider

$$\frac{1}{J}\sum_{j=1}^{J}(\tilde{\boldsymbol{\epsilon}}_j^\top \boldsymbol{M}_0 \boldsymbol{X}_j \boldsymbol{X}_j^\top \boldsymbol{M}_0 \tilde{\boldsymbol{\epsilon}}_j)^2$$
$$= \frac{1}{J}\sum_{j=1}^{J}((\boldsymbol{\epsilon}_j + (\gamma_j - \hat{\gamma}_j)\boldsymbol{\iota}_{IT} + \boldsymbol{X}_j(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}_j))^\top \boldsymbol{M}_0 \boldsymbol{X}_j \boldsymbol{X}_j^\top \boldsymbol{M}_0$$
$$(\boldsymbol{\epsilon}_j + (\gamma_j - \hat{\gamma}_j)\boldsymbol{\iota}_{IT} + \boldsymbol{X}_j(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}_j)))^2$$
$$= \frac{1}{J}\sum_{j=1}^{J}\Bigg(\boldsymbol{\epsilon}_j^\top \boldsymbol{M}_0 \boldsymbol{X}_j \boldsymbol{X}_j^\top \boldsymbol{M}_0 \boldsymbol{\epsilon}_j$$

$$
\begin{aligned}
&+ 2\boldsymbol{\epsilon}_j^\top \boldsymbol{M}_0 \boldsymbol{X}_j \boldsymbol{X}_j^\top \boldsymbol{M}_0((\gamma_j - \hat{\gamma}_j)\boldsymbol{\iota}_{IT} + \boldsymbol{X}_j(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}_j)) \\
&+ (((\gamma_j - \hat{\gamma}_j)\boldsymbol{\iota}_{IT} + \boldsymbol{X}_j(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}_j)))^\top \boldsymbol{M}_0 \boldsymbol{X}_j \boldsymbol{X}_j^\top \boldsymbol{M}_0 \\
&((\gamma_j - \hat{\gamma}_j)\boldsymbol{\iota}_{IT} + \boldsymbol{X}_j(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}_j))\Big)^2 \\
= &\frac{1}{J}\sum_{j=1}^{J} \left(\boldsymbol{\epsilon}_j^\top \boldsymbol{M}_0 \boldsymbol{X}_j \boldsymbol{X}_j^\top \boldsymbol{M}_0 \boldsymbol{\epsilon}_j\right)^2 \\
&+ 2(\boldsymbol{\epsilon}_j^\top \boldsymbol{M}_0 \boldsymbol{X}_j \boldsymbol{X}_j^\top \boldsymbol{M}_0 \boldsymbol{\epsilon}_j)\big(2\boldsymbol{\epsilon}_j^\top \boldsymbol{M}_0 \boldsymbol{X}_j \boldsymbol{X}_j^\top \boldsymbol{M}_0((\gamma_j - \hat{\gamma}_j)\boldsymbol{\iota}_{IT} + \boldsymbol{X}_j(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}_j)) \\
&+ (((\gamma_j - \hat{\gamma}_j)\boldsymbol{\iota}_{IT} + \boldsymbol{X}_j(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}_j)))^\top \boldsymbol{M}_0 \boldsymbol{X}_j \boldsymbol{X}_j^\top \boldsymbol{M}_0((\gamma_j - \hat{\gamma}_j)\boldsymbol{\iota}_{IT} \\
&+ \boldsymbol{X}_j(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}_j)))\big) \\
&+ \big(2\boldsymbol{\epsilon}_j^\top \boldsymbol{M}_0 \boldsymbol{X}_j \boldsymbol{X}_j^\top \boldsymbol{M}_0((\gamma_j - \hat{\gamma}_j)\boldsymbol{\iota}_{IT} + \boldsymbol{X}_j(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}_j)) \\
&+ (((\gamma_j - \hat{\gamma}_j)\boldsymbol{\iota}_{IT} + \boldsymbol{X}_j(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}_j)))^\top \boldsymbol{M}_0 \boldsymbol{X}_j \boldsymbol{X}_j^\top \boldsymbol{M}_0 \\
&((\gamma_j - \hat{\gamma}_j)\boldsymbol{\iota}_{IT} + \boldsymbol{X}_j(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}_j))\big)^2 \\
= &\frac{1}{J}\sum_{j=1}^{J} \left(\boldsymbol{\epsilon}_j^\top \boldsymbol{M}_0 \boldsymbol{X}_j \boldsymbol{X}_j^\top \boldsymbol{M}_0 \boldsymbol{\epsilon}_j\right)^2 + o_p(1), \qquad (3.39)
\end{aligned}
$$

where we have used the Properties 2.28.

Similarly, from (3.30) and definition of $\tilde{\sigma}_j^2$, we have

$$
\begin{aligned}
&-\frac{2}{J}\sum_{j=1}^{J}(\tilde{\boldsymbol{\epsilon}}_j^\top \boldsymbol{M}_0 \boldsymbol{X}_j \boldsymbol{X}_j^\top \boldsymbol{M}_0 \tilde{\boldsymbol{\epsilon}}_j)\tilde{\sigma}_j^2 \mathrm{trace}(\boldsymbol{X}_j^\top \boldsymbol{M}_0 \boldsymbol{X}_j) \\
&= -\frac{2}{J}\sum_{j=1}^{J}\Bigg((\boldsymbol{\epsilon}_j + (\gamma_j - \hat{\gamma}_j)\boldsymbol{\iota}_{IT} + \boldsymbol{X}_j(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}_j))^\top \boldsymbol{M}_0 \boldsymbol{X}_j \boldsymbol{X}_j^\top \boldsymbol{M}_0 \\
&(\boldsymbol{\epsilon}_j + (\gamma_j - \hat{\gamma}_j)\boldsymbol{\iota}_{IT} + \boldsymbol{X}_j(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}_j))\Bigg) \\
&\Bigg(\frac{(\boldsymbol{\epsilon}_j + (\gamma_j - \hat{\gamma}_j)\boldsymbol{\iota}_{IT} + \boldsymbol{X}_j(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}_j))^\top(\boldsymbol{\epsilon}_j + (\gamma_j - \hat{\gamma}_j)\boldsymbol{\iota}_{IT} + \boldsymbol{X}_j(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}_j))}{IT - I - K - 1} \\
&\mathrm{trace}(\boldsymbol{X}_j^\top \boldsymbol{M}_0 \boldsymbol{X}_j)\Bigg) \\
&= -\frac{2}{J}\sum_{j=1}^{J}\Bigg(\boldsymbol{\epsilon}_j^\top \boldsymbol{M}_0 \boldsymbol{X}_j \boldsymbol{X}_j^\top \boldsymbol{M}_0 \boldsymbol{\epsilon}_j\Bigg)\Bigg(\frac{\boldsymbol{\epsilon}_j^\top \boldsymbol{\epsilon}_j}{IT - I - K - 1}\mathrm{trace}(\boldsymbol{X}_j^\top \boldsymbol{M}_0 \boldsymbol{X}_j)\Bigg) + o_p(1). \qquad (3.40)
\end{aligned}
$$

Similarly, from the definition of $\tilde{\sigma}_j^2$ and (3.30), we have

$$\begin{aligned}
&\frac{1}{J}\sum_{j=1}^{J}\left(\tilde{\sigma}_j^2 \text{trace}(\boldsymbol{X}_j^\top \boldsymbol{M}_0 \boldsymbol{X}_j)\right)^2 \\
&= \frac{1}{J}\sum_{j=1}^{J}\Bigg(\frac{(\boldsymbol{\epsilon}_j + (\gamma_j - \hat{\gamma}_j)\boldsymbol{\iota}_{IT} + \boldsymbol{X}_j(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}_j))^\top(\boldsymbol{\epsilon}_j + (\gamma_j - \hat{\gamma}_j)\boldsymbol{\iota}_{IT} + \boldsymbol{X}_j(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}_j))}{IT - I - K - 1} \\
&\qquad \text{trace}(\boldsymbol{X}_j^\top \boldsymbol{M}_0 \boldsymbol{X}_j)\Bigg)^2 \\
&= \frac{1}{J}\sum_{j=1}^{J}\left(\frac{\boldsymbol{\epsilon}_j^\top \boldsymbol{\epsilon}_j}{IT - I - K - 1}\text{trace}(\boldsymbol{X}_j^\top \boldsymbol{M}_0 \boldsymbol{X}_j)\right)^2 + o_p(1). \qquad (3.41)
\end{aligned}$$

Combining the term (3.39), (3.40), and (3.41) into (3.38) gives us

$$\tilde{\sigma}_{clm}^2 = \frac{1}{J}\sum_{j=1}^{J}\left[\boldsymbol{\epsilon}_j^\top \boldsymbol{M}_0 \boldsymbol{X}_j \boldsymbol{X}_j^\top \boldsymbol{M}_0 \boldsymbol{\epsilon}_j - \frac{\boldsymbol{\epsilon}_j^\top \boldsymbol{\epsilon}_j}{IT - I - K - 1}\text{trace}(\boldsymbol{X}_j^\top \boldsymbol{M}_0 \boldsymbol{X}_j)\right]^2 + o_p(1).$$

Let

$$\begin{aligned}
z_j &= \left[\boldsymbol{\epsilon}_j^\top \boldsymbol{M}_0 \boldsymbol{X}_j \boldsymbol{X}_j^\top \boldsymbol{M}_0 \boldsymbol{\epsilon}_j - \frac{\boldsymbol{\epsilon}_j^\top \boldsymbol{\epsilon}_j}{IT - I - K - 1}\text{trace}(\boldsymbol{X}_j^\top \boldsymbol{M}_0 \boldsymbol{X}_j)\right]^2 \\
&= \left[\boldsymbol{\epsilon}_j^\top \left[\boldsymbol{M}_0 \boldsymbol{X}_j \boldsymbol{X}_j^\top \boldsymbol{M}_0 - \frac{\mathbb{I}_{IT}}{IT - I - K - 1}\text{trace}(\boldsymbol{X}_j^\top \boldsymbol{M}_0 \boldsymbol{X}_j)\right]\boldsymbol{\epsilon}_j\right]^2 \\
&= \left[\text{trace}\left[\boldsymbol{\epsilon}_j^\top \left[\boldsymbol{M}_0 \boldsymbol{X}_j \boldsymbol{X}_j^\top \boldsymbol{M}_0 - \frac{\mathbb{I}_{IT}}{IT - I - K - 1}\text{trace}(\boldsymbol{X}_j^\top \boldsymbol{M}_0 \boldsymbol{X}_j)\right]\boldsymbol{\epsilon}_j\right]\right]^2 \\
&= \left[\text{trace}\left[\boldsymbol{\epsilon}_j \boldsymbol{\epsilon}_j^\top \left[\boldsymbol{M}_0 \boldsymbol{X}_j \boldsymbol{X}_j^\top \boldsymbol{M}_0 - \frac{\mathbb{I}_{IT}}{IT - I - K - 1}\text{trace}(\boldsymbol{X}_j^\top \boldsymbol{M}_0 \boldsymbol{X}_j)\right]\right]\right]^2.
\end{aligned}$$

Define the quantities

$$\begin{aligned}
\boldsymbol{Z}_{j1} &= \boldsymbol{\epsilon}_j \boldsymbol{\epsilon}_j^\top \\
\boldsymbol{Z}_{j2} &= \boldsymbol{M}_0 \boldsymbol{X}_j \boldsymbol{X}_j^\top \boldsymbol{M}_0 - \frac{\mathbb{I}_{IT}}{IT - I - K - 1}\text{trace}(\boldsymbol{X}_j^\top \boldsymbol{M}_0 \boldsymbol{X}_j).
\end{aligned}$$

Therefore, $z_j = (\text{trace}[\boldsymbol{Z}_{j1}\boldsymbol{Z}_{j2}])^2$.

Using $\boldsymbol{Z}_{j1}$ and $\boldsymbol{Z}_{j2}$ are symmetric matrix, Theorem 2.41(3), and Assumption 3,

$$\begin{aligned}
E[z_j^2] &\leq E\left[[\text{trace}(\boldsymbol{Z}_{j1}^\top \boldsymbol{Z}_{j1})]^2[\text{trace}(\boldsymbol{Z}_{j2}^\top \boldsymbol{Z}_{j2})]^2\right] \\
&= E\left[E\left([\text{trace}(\boldsymbol{Z}_{j1}^\top \boldsymbol{Z}_{j1})]^2[\text{trace}(\boldsymbol{Z}_{j2}^\top \boldsymbol{Z}_{j2})]^2\right)|\boldsymbol{X}_j\right] \\
&= E\left[E\left([\text{trace}(\boldsymbol{Z}_{j1}^\top \boldsymbol{Z}_{j1})]^2\right)|\boldsymbol{X}_j\right] E[\text{trace}(\boldsymbol{Z}_{j2}^\top \boldsymbol{Z}_{j2})]^2 \\
&\leq E[(\boldsymbol{\epsilon}_j^\top \boldsymbol{\epsilon}_j)^4|\boldsymbol{X}_j] E[\text{trace}[(\boldsymbol{X}_j^\top \boldsymbol{M}_0 \boldsymbol{X}_j)^4]] \\
&\leq J_1 J_2 < \infty.
\end{aligned}$$

Then

$$\frac{1}{J^2}\sum_{j=1}^{J} E(z_j^2) \to 0 \quad \text{as} \quad J \to \infty,$$

and we can apply Chebychev's Weak Law of Large Numbers to obtain

$$\begin{aligned}
\tilde{\sigma}_{clm}^2 &= \frac{1}{J}\sum_{j=1}^{J}\left[\tilde{\boldsymbol{\epsilon}}_j^\top \boldsymbol{M}_0 \boldsymbol{X}_j \boldsymbol{X}_j^\top \boldsymbol{M}_0 \tilde{\boldsymbol{\epsilon}}_j - \tilde{\sigma}_j^2 \text{trace}(\boldsymbol{X}_j^\top \boldsymbol{M}_0 \boldsymbol{X}_j)\right]^2 \\
&\xrightarrow{p} \lim_{J\to\infty}\frac{1}{J}\sum_{j=1}^{J} E\left[\boldsymbol{\epsilon}_j^\top \boldsymbol{M}_0 \boldsymbol{X}_j \boldsymbol{X}_j^\top \boldsymbol{M}_0 \boldsymbol{\epsilon}_j - \frac{\boldsymbol{\epsilon}_j^\top \boldsymbol{\epsilon}_j}{IT - I - K - 1}\text{trace}(\boldsymbol{X}_j^\top \boldsymbol{M}_0 \boldsymbol{X}_j)\right]^2 \\
&= \sigma_{clm}^2.
\end{aligned}$$

□

CHAPTER IV
NUMERICAL RESULT

In this chapter, we show an empirical study to numerically test the performance of our proposed test statistic by considering the size of the proposed hypothesis. The setting of the simulation is as follows.
Consider the model under null hypothesis

$$y_{ijt} = \alpha_i + \gamma_j + \boldsymbol{x}_{ijt}^\top \boldsymbol{\beta} + \epsilon_{ijt}, \tag{4.1}$$

where α_i is drawn from the $N(0,1)$ distribution and γ_j is drawn from the $N(0, 0.001)$ distribution. In addition, we construct

$$\begin{aligned}
x_{ijt1} &= (\boldsymbol{\iota}_T \otimes (\boldsymbol{\iota}_J \otimes \mathbb{I}_I))\alpha_i + (\boldsymbol{\iota}_T \otimes (\mathbb{I}_J \otimes \boldsymbol{\iota}_I))\gamma_j + w_{ijt1} \quad \text{with} \quad w_{ijt1} \sim N(1,1), \\
x_{ijt2} &= (\boldsymbol{\iota}_T \otimes (\boldsymbol{\iota}_J \otimes \mathbb{I}_I))\alpha_i + (\boldsymbol{\iota}_T \otimes (\mathbb{I}_J \otimes \boldsymbol{\iota}_I))\gamma_j + w_{ijt2} \quad \text{with} \quad w_{ijt2} \sim N(2,1), \\
x_{ijt3} &= (\boldsymbol{\iota}_T \otimes (\boldsymbol{\iota}_J \otimes \mathbb{I}_I))\alpha_i + (\boldsymbol{\iota}_T \otimes (\mathbb{I}_J \otimes \boldsymbol{\iota}_I))\gamma_j + w_{ijt3} \quad \text{with} \quad w_{ijt3} \sim N(1.5,1), \\
x_{ijt4} &= (\boldsymbol{\iota}_T \otimes (\boldsymbol{\iota}_J \otimes \mathbb{I}_I))\alpha_i + (\boldsymbol{\iota}_T \otimes (\mathbb{I}_J \otimes \boldsymbol{\iota}_I))\gamma_j + w_{ijt4} \quad \text{with} \quad w_{ijt4} \sim N(1.75,1), \\
x_{ijt5} &= (\boldsymbol{\iota}_T \otimes (\boldsymbol{\iota}_J \otimes \mathbb{I}_I))\alpha_i + (\boldsymbol{\iota}_T \otimes (\mathbb{I}_J \otimes \boldsymbol{\iota}_I))\gamma_j + w_{ijt5} \quad \text{with} \quad w_{ijt5} \sim N(1.25,1),
\end{aligned}$$

$\boldsymbol{x}_{ijt}$ is a 5×1 covariate vector and simulate the slope coefficient vector $\boldsymbol{\beta}$ of size 5×1 from

$$\begin{aligned}
\beta^{(1)} &\sim N(1,1), \\
\beta^{(2)} &\sim N(1.5,1), \\
\beta^{(3)} &\sim N(2,1), \\
\beta^{(4)} &\sim N(1.75,1), \\
\beta^{(5)} &\sim N(2.5,1).
\end{aligned}$$

From the above construction, we can see that $\alpha_i,\ \gamma_j$, and $\boldsymbol{x}_{ijt}$ are correlated. The variance of the error terms are given by $\mathrm{E}(\epsilon_{ijt}^2) = \sigma_\epsilon^2$ with $\sigma_\epsilon^2 \sim U(0, 0.01)$. Sample sizes are $I = 5, 7, 9,\ J = 25, 50, 75, 100$, and $T = 7$.

We study the size of a hypothesis testing by letting $\boldsymbol{\beta}_{ij} = \boldsymbol{\beta}$ for all i, j. For each experiment, we perform 5000 replications. We compare the different significant levels such as $\alpha = 0.1,\ 0.05,\ 0.01, 0.0001$. In our simulation, we examine the accuracy of the limiting distribution of CLM by generating model error ϵ_{ijt} in (4.1) from three different distributions such as a normal distribution, a t-distribution with five degrees of freedom and a centered chi-square distribution with four degrees of freedom. Tables 4.1 - 4.3 show the size of CLM in percentage and Figures 4.1a - 4.9d show the range of statistics CLM.

From Tables 4.1 - 4.3, we can see that size is generally conservative for all error distributions and all tests with the exception when level of factor 1 is greater than the level of time index, the test is oversized at significant level 0.1 but decreases very quickly as significant level 0.0001. In particular, for the case of chi-square distribution, for significant level 0.05, size is 8.24% for $I = 5, J = 25$, and $T = 7$, but decreases to 5.32% when $J = 100$. Moreover, for all tests and all error distributions, the statistics behave well, in particular when the significant level is small. For example, for normal errors, when $I = 9, J = 100$, and $T = 7$, the size of significant level 0.1, test is 54.08%, but decreases to 0.02% when significant level 0.0001. The simulation results confirm our theorem.

			$\alpha = 0.1$	$\alpha = 0.05$	$\alpha = 0.01$	$\alpha = 0.0001$
I = 5	T=7	J= 25	12.42	7.6	2.74	0
		J= 50	11.08	5.98	2.00	0.08
		J= 75	10.14	4.94	1.00	0.02
		J= 100	10.74	5.36	1.26	0.10
I = 7	T = 7	J= 25	10.18	5.12	1.02	0
		J= 50	10.3	5.10	1.16	0.04
		J= 75	9.9	4.66	0.84	0.02
		J= 100	10.28	5.26	1.20	0
I = 9	T = 7	J= 25	16.52	6.86	0.68	0
		J= 50	29.14	15.12	2.28	0
		J= 75	45.00	28.52	6.38	0.02
		J= 100	54.08	37.70	11.10	0.02

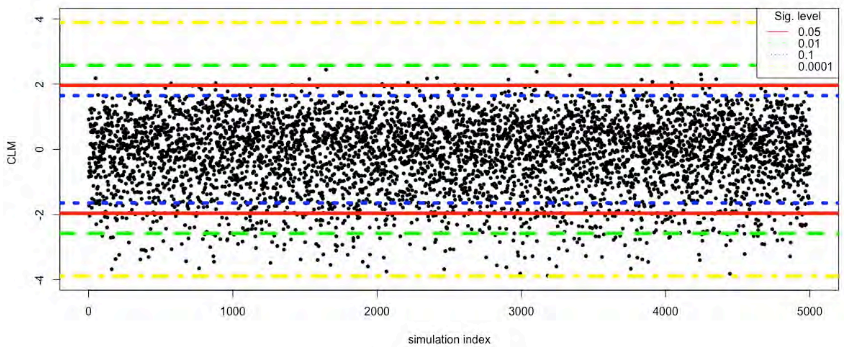
Table 4.1: Size of CLM (in percentage) when error terms are generated from a normal distribution

			$\alpha = 0.1$	$\alpha = 0.05$	$\alpha = 0.01$	$\alpha = 0.0001$
I = 5	T=7	J= 25	13.28	7.52	2.52	0.08
		J= 50	11.10	6.00	1.64	0.14
		J= 75	10.82	5.30	1.42	0.10
		J= 100	11.18	5.76	1.24	0.08
I = 7	T = 7	J= 25	10.3	4.98	1.10	0.02
		J= 50	9.64	4.54	1.10	0
		J= 75	10.66	5.14	0.94	0.02
		J= 100	9.92	4.86	0.88	0
I = 9	T = 7	J= 25	14.54	6.20	0.66	0
		J= 50	28.36	14.28	2.06	0
		J= 75	41.52	26.68	6.28	0
		J= 100	50.98	35.54	10.28	0.02

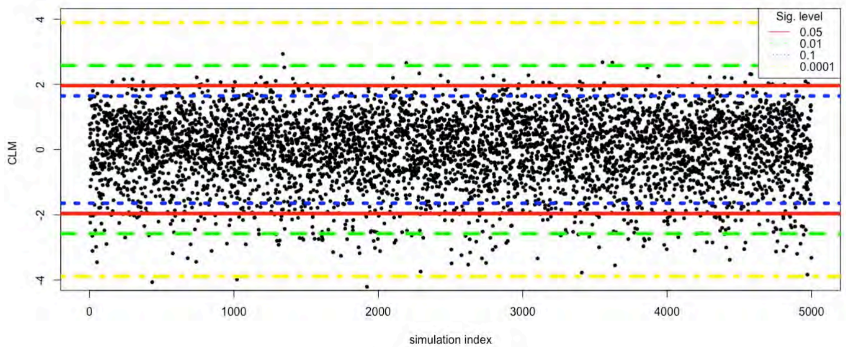
Table 4.2: Size of CLM (in percentage) when error terms are generated from a t-distribution

			$\alpha = 0.1$	$\alpha = 0.05$	$\alpha = 0.01$	$\alpha = 0.0001$
I = 5	T=7	J= 25	13.04	8.24	3.00	0.04
		J= 50	10.42	5.82	2.02	0.04
		J= 75	11.24	5.76	1.44	0.10
		J= 100	10.92	5.32	1.22	0.06
I = 7	T = 7	J= 25	10.94	5.24	1.1	0
		J= 50	10.82	5.48	1.32	0.04
		J= 75	10.54	4.82	1.02	0
		J= 100	10.16	4.80	0.84	0
I = 9	T = 7	J= 25	15.24	6.26	0.72	0
		J= 50	27.14	14.10	2.22	0
		J= 75	42.44	26.54	6.12	0
		J= 100	53.40	36.52	10.92	0.02

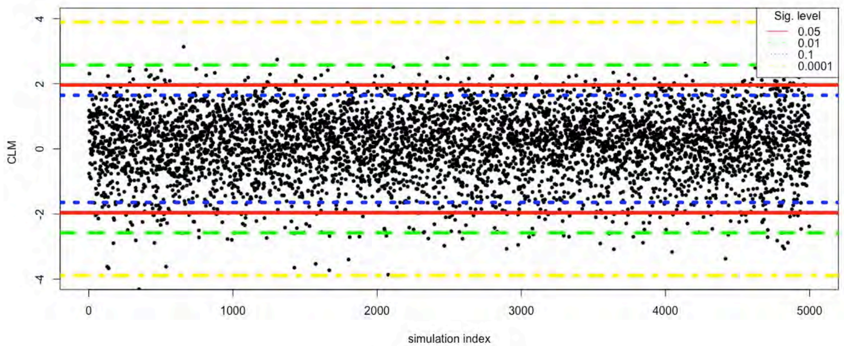
Table 4.3: Size of CLM (in percentage) when error terms are generated from a chi-square distribution



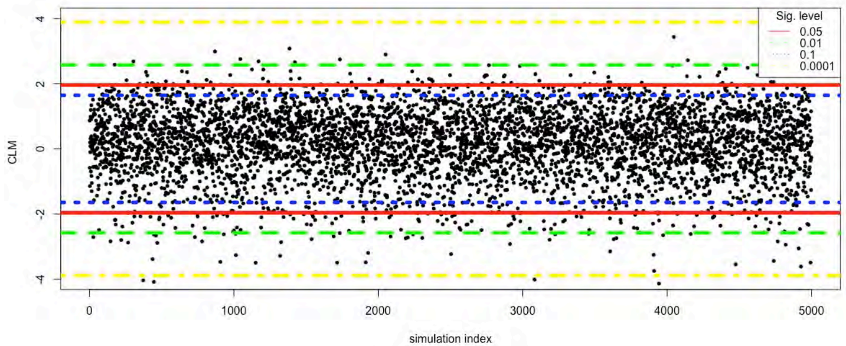
(a) I = 5, J=25, T=7



(b) I = 5, J=50, T=7

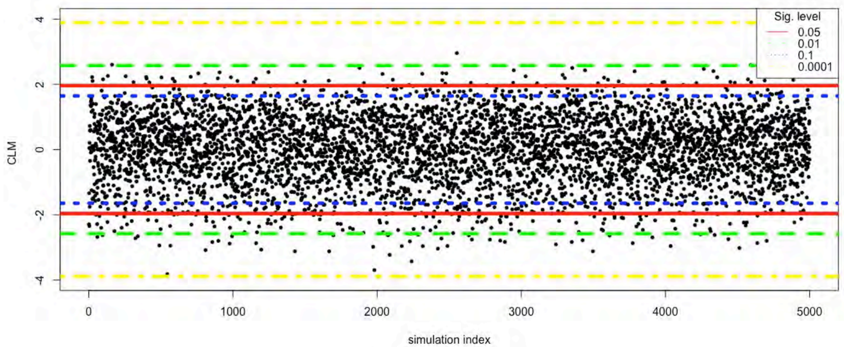


(c) I = 5, J=75, T=7

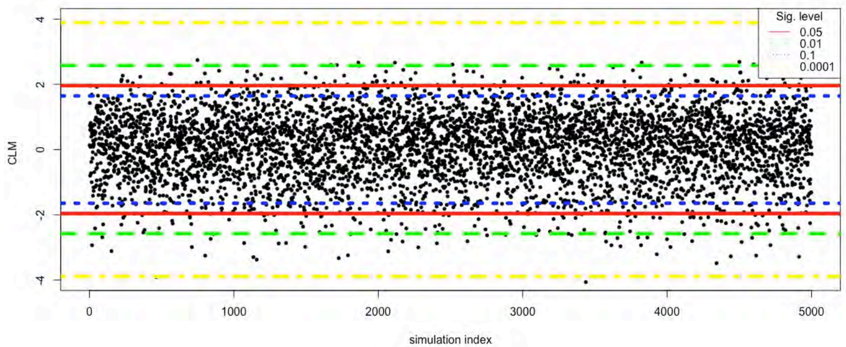


(d) I = 5, J=100, T=7

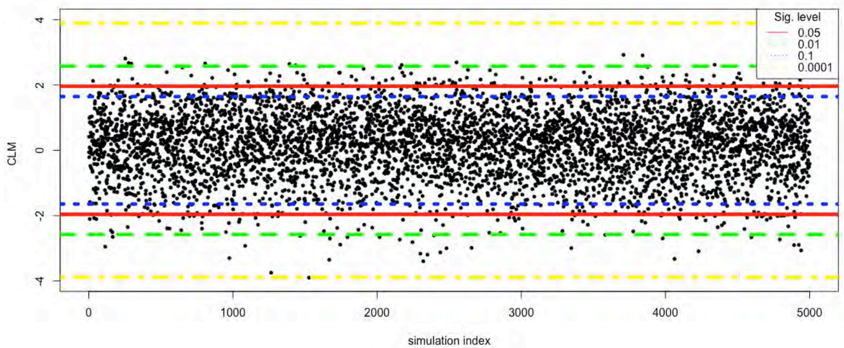
Figure 4.1: Ranges of statistics with respect to standard normal value when error distributed by a normal distribution.



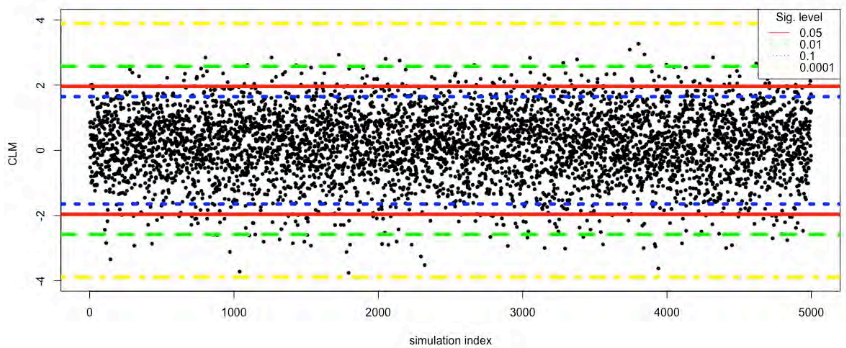
(a) I = 7, J=25, T=7



(b) I = 7, J=50, T=7

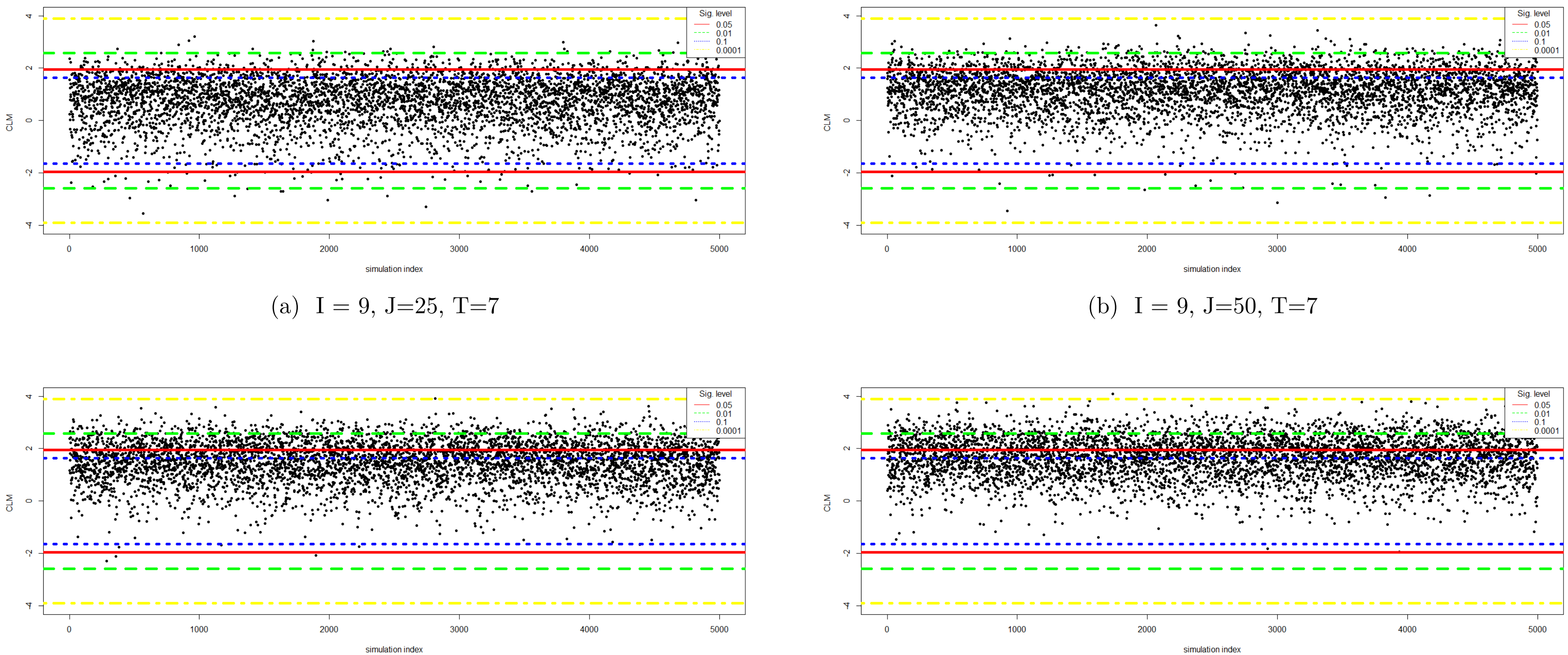


(c) I = 7, J=75, T=7



(d) I = 7, J=100, T=7

Figure 4.2: Ranges of statistics with respect to standard normal value when error distributed by a normal distribution.



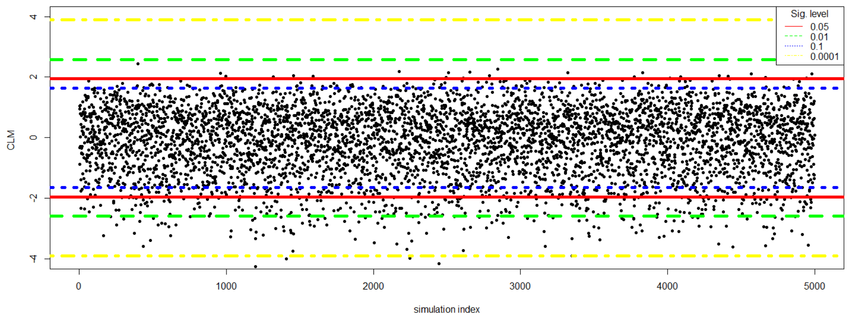
(a) I = 9, J=25, T=7

(b) I = 9, J=50, T=7

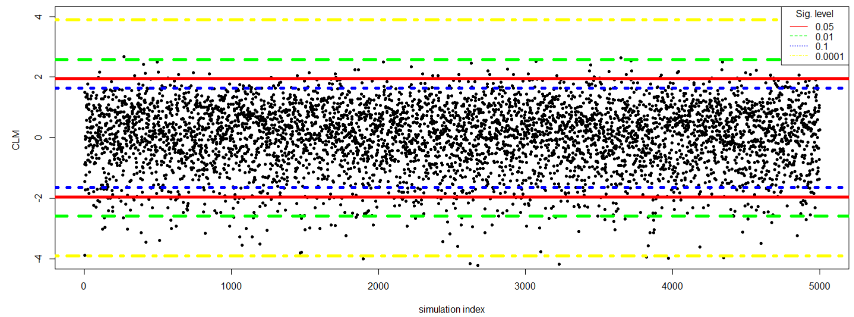
(c) I = 9, J=75, T=7

(d) I = 9, J=100, T=7

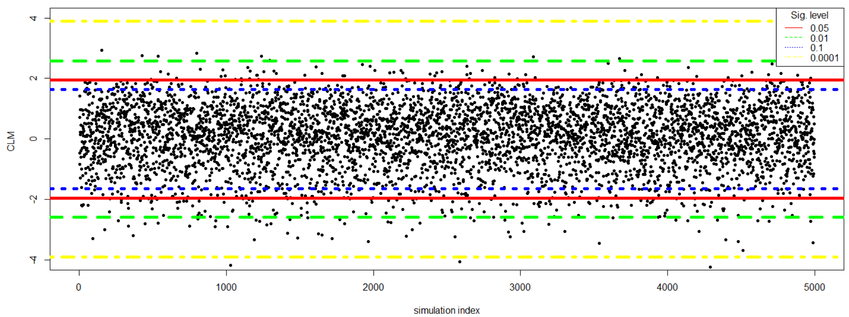
Figure 4.3: Ranges of statistics with respect to standard normal value when error distributed by a normal distribution.



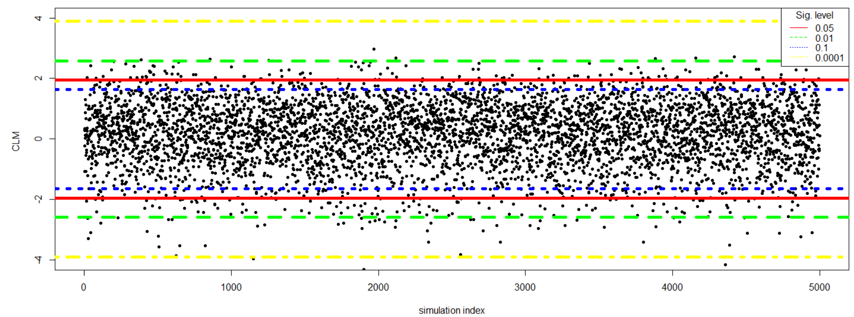
(a) I = 5, J=25, T=7



(b) I = 5, J=50, T=7

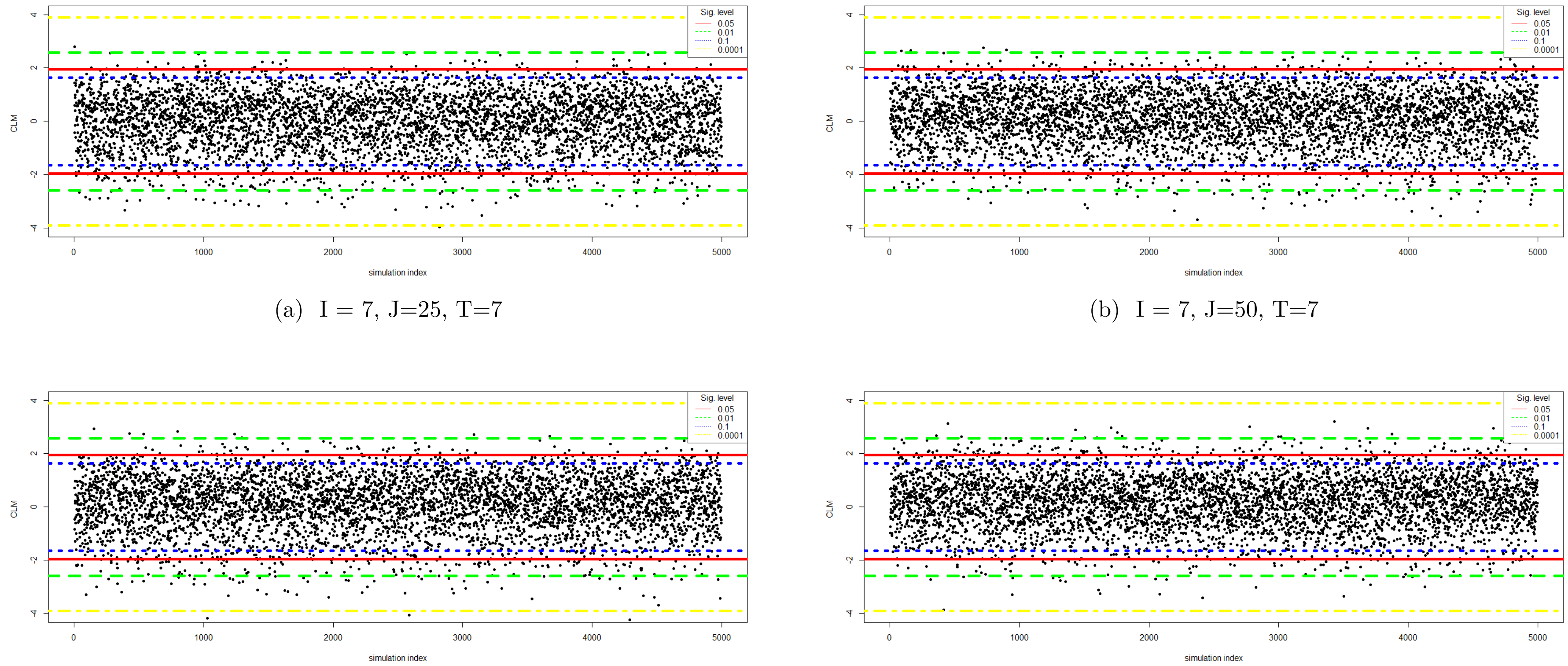


(c) I = 5, J=75, T=7



(d) I = 5, J=100, T=7

Figure 4.4: Ranges of statistics with respect to standard normal value when error distributed by a t-distribution with five degrees of freedom.



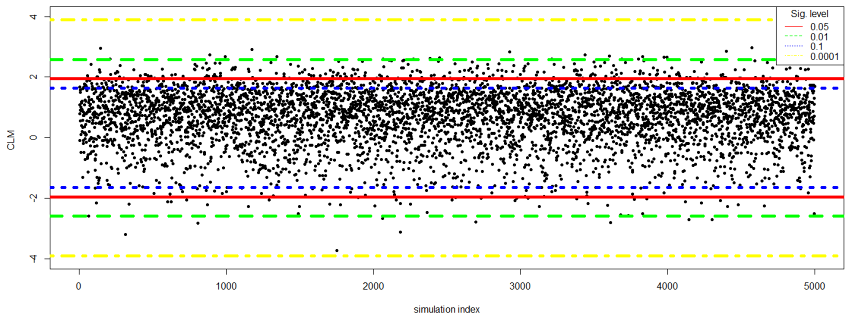
(a) I = 7, J=25, T=7

(b) I = 7, J=50, T=7

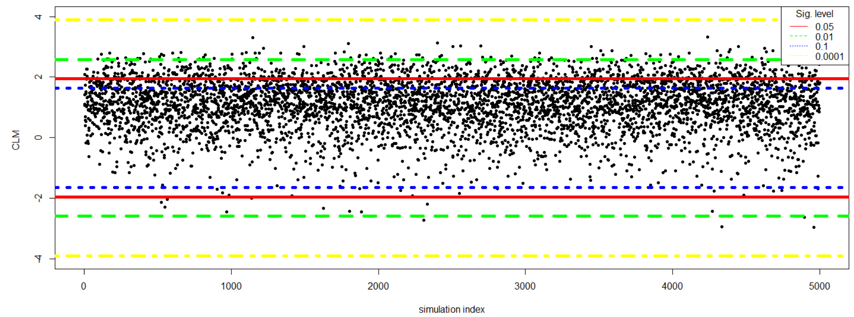
(c) I = 7, J=75, T=7

(d) I = 7, J=100, T=7

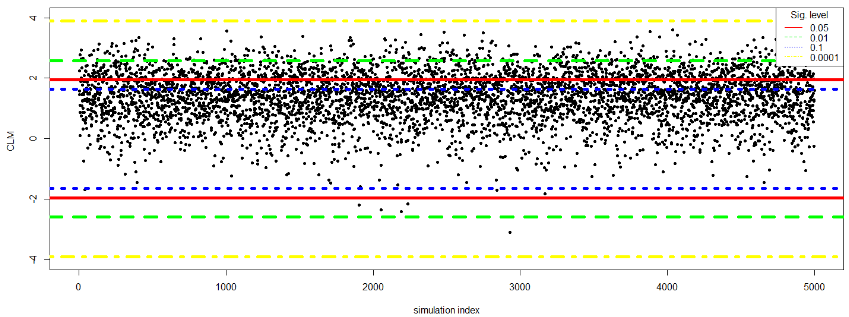
Figure 4.5: Ranges of statistics with respect to standard normal value when error distributed by a *t*-distribution with five degrees of freedom.



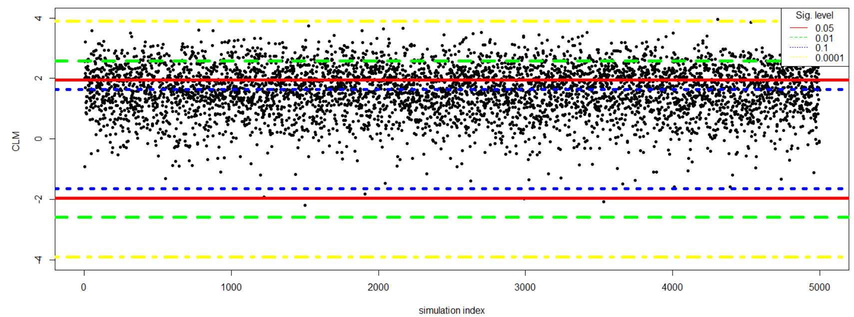
(a) I = 9, J=25, T=7



(b) I = 9, J=50, T=7

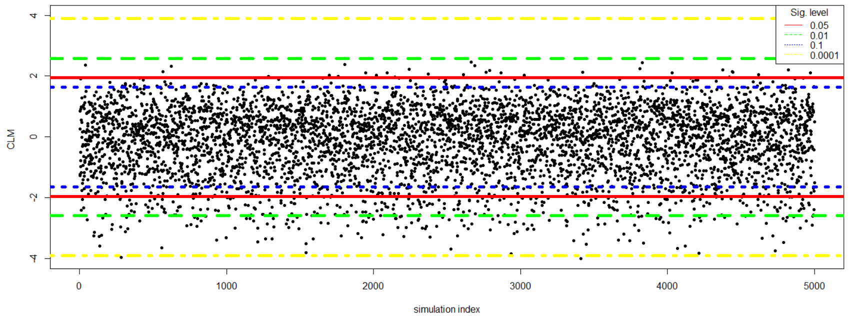


(c) I = 9, J=75, T=7

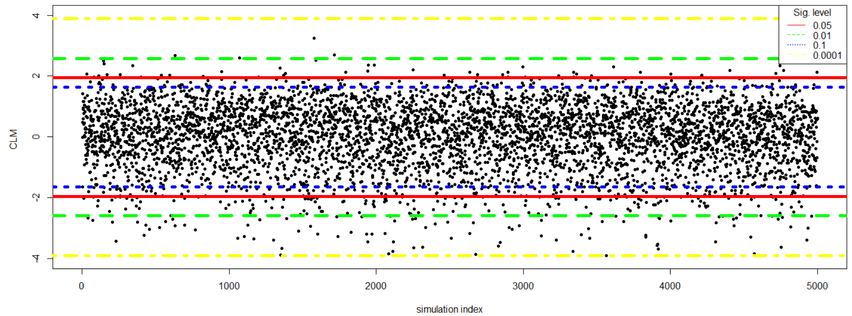


(d) I = 9, J=100, T=7

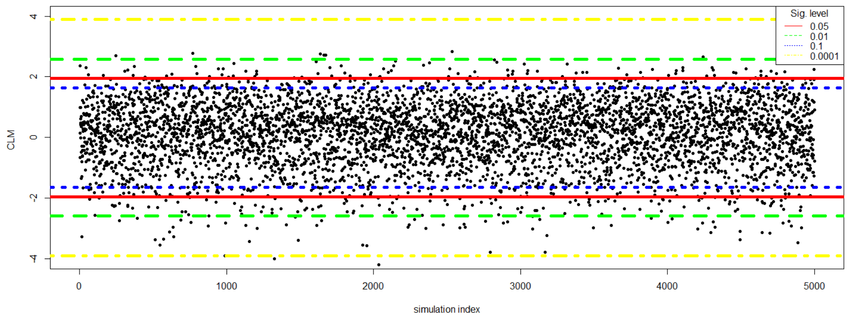
Figure 4.6: Ranges of statistics with respect to standard normal value when error distributed by a t-distribution with five degrees of freedom.



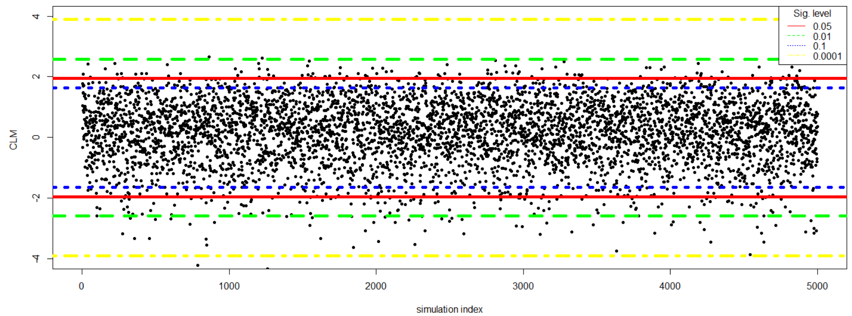
(a) I = 5, J=25, T=7



(b) I = 5, J=50, T=7

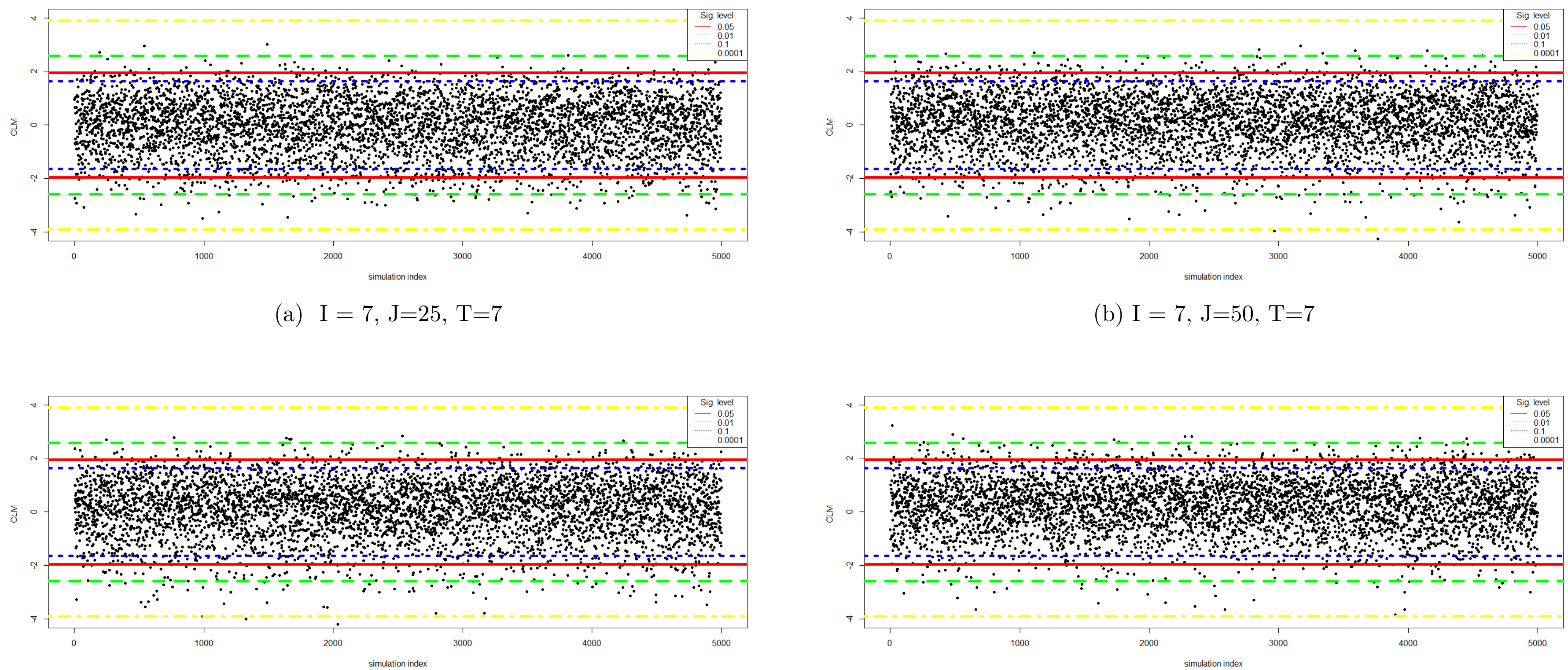


(c) I = 5, J=75, T=7



(d) I = 5, J=100, T=7

Figure 4.7: Ranges of statistics with respect to standard normal value when error distributed by a centered chi-square distribution with four degrees of freedom.



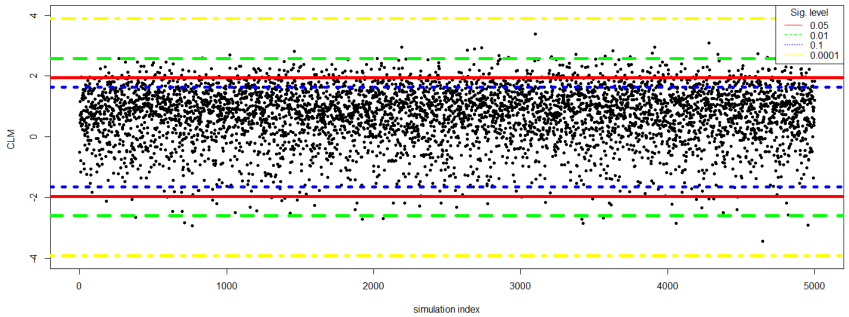
(a) I = 7, J=25, T=7

(b) I = 7, J=50, T=7

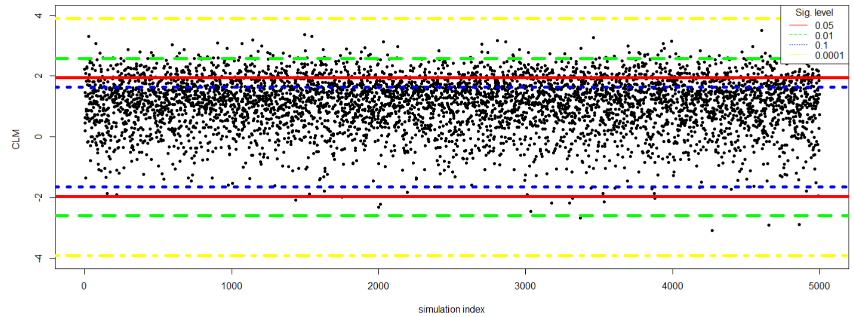
(c) I = 7, J=75, T=7

(d) I = 7, J=100, T=7

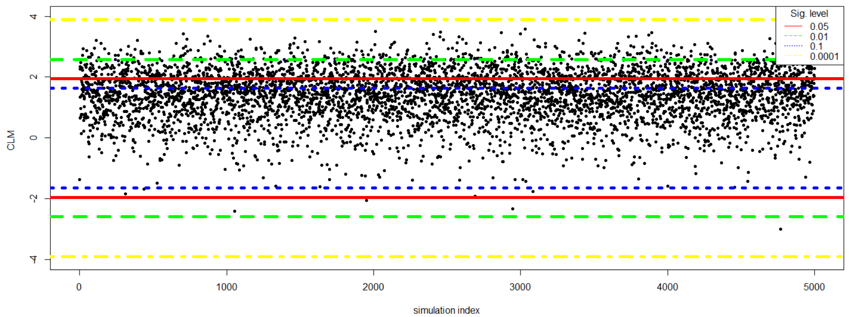
Figure 4.8: Ranges of statistics with respect to standard normal value when error distributed by a centered chi-square distribution with four degrees of freedom.



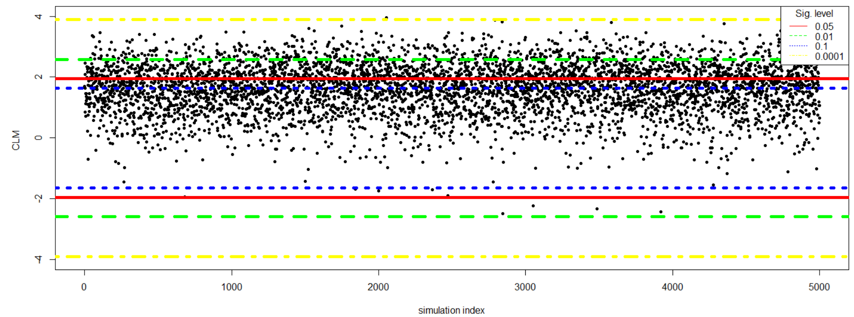
(a) I = 9, J=25, T=7



(b) I = 9, J=50, T=7



(c) I = 9, J=75, T=7



(d) I = 9, J=100, T=7

Figure 4.9: Ranges of statistics with respect to standard normal value when error distributed by a centered chi-square distribution with four degrees of freedom.

CHAPTER V
CONCLUSIONS

In this thesis, we have constructed a test for slope heterogeneity in three-dimensional panel data model based on a conditional Lagrange multiplier test. The advantage of our test is more available and important in applications for multi-dimensional panel data. From the simulation results in Tables 4.1 - 4.3, the size of CLM do not continuously decrease when $J \to \infty$. When $J = 100$, it may be not large enough to draw conclusions how effective the approach will be. Therefore, we added experiments with $J = 200$ when $I = 5, T = 7$ for all error distributions. The results of this experiment as follows. When error terms are generated from a normal distribution, a t-distribution and a chi-squared distribution for significant level 0.05, the size (in percentage) is 5.76%, 5.60%, and 5.78%, respectively. In conclusions, the size of CLM is generally conservative when $J \to \infty$ for all error distributions and all tests with the exception when level of factor 1 is greater than the level of time index. Moreover, our test does not require large I and T. Even though, we do not have a real data applications in this thesis. We hope that the test may be applicable to some applications in the future work, for example, a real data application could be a two-way ANOVA with time series data available for applications in the timeline of thesis.

REFERENCES

[1] Aitchison, J. and Silvey, S. D. (1958). Maximum-likelihood Estimation of Parameters Subject to Restraints. *Annual of Mathematical Statistics*, **29**, 813–828.

[2] Aitchison, J. and Silvey, S. D. (1960). Maximum-likelihood Estimation Procedures and Associated Tests of Significance. *Journal of the Royal Statistical Soceity*, Series B, **22**, 154–171.

[3] Anellano, M. (2003). *Panel Data Econometric.* Oxford. Oxford University Press.

[4] Baltagi, B. H. *Econometric Analysis of Panel Data.* Chichester. John Wiley.

[5] Baltagi, B. H. and Li, Q. (1990). A Lagrange Multiplier Test for the Error Components Model with Incomplete Panels. *Econometric Review*, **9**, 103–107.

[6] Breusch, T. and Pagan, A. (1980). The Lagrange Multiplier Test and Its Application in Econometrics. *Review of Economic Studies*, **47**, 239–253.

[7] Byron, R. P. (1970). The Restricted Aitken Estimation of Sets of Demand Relations. *Econometrica*, **38**, 816–830.

[8] Casella, G. and Berger, R. L. (2001). *Statistical Inference.* California: Duxbury Press.

[9] Chamberlain, G. (1980). Analysis of Covariance with Qualitative Data. *Review of Economic Studies*, **47**, 225–238.

[10] Daren, B. H. (1989). Consistency for Least Squares Regression Estimators with Infinite Variance Data. *Journal of Statistical Planning and Inference*, **23**, 163–179.

[11] Facker, P. L. *Notes on Matrix Calculus.* Available from http:// www4.ncsu.edu/ pfackler/MatCalc.pdf, Sept. 2005.

[12] Greene, W. H. (2011). *Econometric Analysis.* New Jersey: Pearson.

[13] Hsiao, C. (2003). *Analysis of Panel Data.* New York: Cambridge University Press.

[14] Inoue, A. and Solon, G. (2006). A Portmanteau Test for Serially Correlated Errors in Fixed Effects Models. *Econometric Theory*, **22**, 835–851.

[15] Jean-Claude, B. and Ariane, T. (2004). Bilateral Donors' Aid Allocation Decisions–a Three-dimensional Panel Analysis. *International Review of Economics and Finance*, **13**, 253–274.

[16] Juhl, T. and Lugovskyy, O. (2014). A Test for Slope Heterogeneity in Fixed Effects Models. *Econometric Reviews* **33**(8), 906–935.

[17] Magnus, J. and Neudecker, H. (1999). *Matrix Differential Calculus with Applications in Statistics and Econometrics*. Chichester: John Wiley.

[18] Matyas, L. (1997). Proper Econometric Specification of the Gravity Model. *The World Economy*, **20**, 363–369.

[19] Matyas, L. and Balazsi, L. (2012). The Estimation of Multi-dimensional Fixed Effects Panel Data Models [working paper]. Department of Economics, Central European University.

[20] Pesaran, M, H. and R. Smith. (2008). Testing Slope Homogeneity in Large Panels. *Journal of Econometics*, **142**, 50–93.

[21] Pesaran, M. H., Yamagata, T. (2008). Testing Slope Homogeneity in Large Panels. *Journal of Econometrics* **142**, 50–93.

[22] Rao, C. (1970). Estimation of Heteroscedastic Variances in Linear Models. *Journal of the American Statistical Association*, **65**, 161–172.

[23] Rencher, A. C. and Bruce, S. G. (2007). *Linear Models in Statistics*. New Jersey: John Wiley.

[24] Sakata, S. Testing Parameter Constancy Across Many Groups. Discussion paper, University of British Columbia.

[25] Smyth, G. K. and Verbyla, A. P. (1996). A Conditional Likelihood Approach to Residual Maximum Likelihood Estimation in Generalized Linear Models. *Journal of the Royal Statistical Society*, **58**, 565–572.

[26] Swamy, P. (1970). Efficient Inference in a Random Coefficient Regression Model. *Econometrica*, **38**, 311–323.

[27] Wand, M. P. (2002). Vector Differential Calculus in Statistics. *The American Statistician*, **56**, 1–8.

[28] White, H. (2005). A Heteroskedasticity-consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity. *Econometica*, **48**, 817–838.

[29] White, H. (2001). *Asymptotic Theory for Econometricians*. New York, NY: Academic Press.

[30] Yair, E. and Liran, E. (2004). Determinants of International Tourism : a Three-dimensional Panel Data Analysis. *Applied Economics*, **36**, 1315–1327.

[31] Yamazaki, D. and Kurozumi, E. (2014). Testing for Parameter Constancy in the Time Series Direction in Panel Data Models. *Journal of Statistical Computation and Simulation*.

VITA

Name	:	Miss Sasitorn Kanokhilunyakorn
Date of Birth	:	10 November 1991
Place of Birth	:	Uttaradit, Thailand
Education	:	B.Sc. (Mathematics),
		Chulalongkorn University, 2013

Publication

- Kanokhilunyakorn, S. and Suntornchost, J. (2017). Hypothesis Testing for Slope Heterogeneity for Three Dimensional Panel Data Model. *Proceeding of AMM 2017*, Thailand, 2-4 June 2017. STA-02-1–STA-02-14.