A PERMUTATION TEST FOR PARTIAL REGRESSION COEFFICIENTS ON FIRST-ORDER AUTOCORRELATION

Pradthana Minsan

A Dissertation Submitted in Partial
Fulfillment of the Requirements for the Degree of
Doctor of Philosophy (Statistics)
School of Applied Statistics
National Institute of Development Administration
2010

A PERMUTATION TEST FOR PARTIAL REGRESSION COEFFICIENTS ON FIRST-ORDER AUTOCORRELATION

Pradthana Minsan

Graduate School of Applied Statistics

Associate Professor............Pachitjanut Siripanich............ Major Advisor

(Pachitjanut Siripanich, Ph.D)

The Examining Committee Approved This Dissertation Submitted in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy (Statistics).

Associate Professor............Jirawan Jitthavech............Committee Chairperson

(Jirawan Jitthavech, Ph.D)

Associate Professor............Pachitjanut Siripanich............Committee

(Pachitjanut Siripanich, Ph.D)

Associate Professor..Committee

(Samruam Chongcharoen, Ph.D)

Associate Professor............Montip Tiensuwan............Committee

(Montip Tiensuwan, Ph.D)

Instructor............Lersan Bosuwan............Dean

(Lersan Bosuwan, Ph.D.)

November 2010

ABSTRACT

Title of Dissertation A Permutation Test for Partial Regression Coefficients on First-Order Autocorrelation

Author Mrs. Pradthana Minsan

Degree Doctor of Philosophy (Statistics)

Year 2010

This dissertation proposes a permutation test (F_P) and a permutation procedure for testing on partial regression coefficients from a multiple linear regression with first-order autocorrelation where the distribution of the error terms is not necessarily normal. The proposed permutation procedure can be directly conducted in the test without having to fit back to the model, which is not the same procedure as in previous permutation tests, and a proposed permutation test is considered based on a random permutation test.

In addition, the asymptotic analysis of the proposed test can be obtained when errors are i.i.d. with mean zero and finite variance. The asymptotic distribution of F_P, called the asymptotic chi-squared test, can be used to perform a significance test of partial regression coefficients.

It was found that, for a small sample size (T=12), the proposed permutation method has the same type I error rate as the partial F-statistic and is not significantly different from the significance level α, and has a higher power when compare with the other methods in the case where autocorrelation approached ± 1. However, with a moderate sample size (T=16, 20), the asymptotic chi-squared test is preferred (in terms of type I error and power of the test).

ACKNOWLEDGEMENTS

I am greatly appreciative of my dissertation advisor, Associate Professor Pachitjanut Siripanich, for her guidance, supervision, encouragement and invaluable comments during the completion of this dissertation. I would also like to gratefully acknowledge all of the Dissertation and Examining committee, Associate Professor Jirawan Jitthavech, Associate Professor Samruam Chongcharoen and Associate Professor Montip Tiensuwan, for their helpful comments and suggestions.

I would like to express my gratitude to Professor Prachoom Suwattee, Associate Professor Vichit Lorchirachoonkul, Assistant Professor Arthur L. Dryver , Associate Professor Virool Boonyasombat and Asssistant Professor Lily Insrisawang for their invaluable lectures, imparting their knowledge and experiences, which really help to guide me throughout this study.

This dissertation could not have been completed without the help and support of several people and my organization, Chiang Mai Rajabhat University. Thanks are also dedicated to those who awarded me a scholarship under the Staff Development Project of the Office of the Higher Education Commission. I am very grateful to Dr. John McMorris for his kindness towards me by editing my English, which has made the manuscript more readable.

I am sure that I have forgotten some names of contributors to this study (not least all my friends). I would like to express my deep thanks to all persons, if they have been mentioned above or not, for their energy and help, spirit, patience, and co-operation.

Finally, I am greatly indebted to my husband, my parents, and my brothers and sister, for their patience and encouragement throughout the period of this accomplishment.

Pradthana Minsan

November 2010

TABLE OF CONTENTS

	Page
ABSTRACT	iii
ACKNOWLEDGEMENTS	iv
TABLE OF CONTENTS	v
LIST OF TABLES	vii
LIST OF FIGURES	viii
CHAPTER 1 INTRODUCTION	1
1.1 Background	1
1.2 Objectives of the Study	3
1.3 Scope of the Study	4
1.4 Definitions	4
CHAPTER 2 LITERATURE REVIEW	6
2.1 Permutation Tests	6
2.2 A Permutation Test for a Multiple Linear Regression Model	8
CHAPTER 3 A PROPOSED PEMUTATION TEST	16
3.1 Error Transformation	16
3.2 A Permutation Test for Multiple Linear Regression with First-order Autocorrelation	19
3.3 The Proposed Permutation Method on the AR(1) Model	20
3.4 Procedure of proposed permutation Test (PPT)	22
3.5 Asymptotic Theory	23
CHAPTER 4 SIMULATION STUDY	39
4.1 Simulation method	40
4.2 Results of the Simulation Studies	41

CHAPTER 5 SUMMARY AND FUTURE RESEARCH 52

5.1 Summary 52

5.2 Results of the Study 53

5.3 Recommendations for Future Research 56

BIBLIOGRAPHY 58

APPENDICES 62

APPENDIX A Properties of the Error Terms and Transformation of Error to I.I.D. 63

APPENDIX B Quadratic Form and Its Properties 68

APPENDIX C The Ordinary Least Squares Estimator 74

APPENDIX D The Residuals under the Reduced Model 78

APPENDIX E The F_{ref} Statistic 81

APPENDIX F Rejection Rates of the Tests 84

APPENDIX G Permutation Simulation Flowchart 93

BIOGRAPHY 96

LIST OF TABLES

Tables **Page**

F.1 Empirical Type I Error Rates of the Five Tests at $\alpha = 0.01^*$ when $\boldsymbol{\varepsilon} \sim N(0,1)$ and $(\theta_1, \theta_2, \theta_3) = (1, 1, 1)$ 85

F.2 Empirical Type I Error Rates of the Five Tests at $\alpha = 0.01^*$ when $\boldsymbol{\varepsilon} \sim U(-2,2)$ and $(\theta_1, \theta_2, \theta_3) = (1, 1, 1)$ 86

F.3 Empirical Type I Error Rates of the Five Tests at $\alpha = 0.05^*$ when $\boldsymbol{\varepsilon} \sim N(0,1)$ and $(\theta_1, \theta_2, \theta_3) = (1, 1, 1)$ 87

F.4 Empirical Type I Error Rates of the Five Tests at $\alpha = 0.05^*$ when $\boldsymbol{\varepsilon} \sim U(-2,2)$ and $(\theta_1, \theta_2, \theta_3) = (1, 1, 1)$ 88

F.5 Empirical Powers of the Five Tests at $\alpha = 0.01^*$ when $\boldsymbol{\varepsilon} \sim N(0,1)$, $(\theta_1, \theta_2, \theta_3) = (1, 1, 1)$ and $(\beta_1, \beta_2) = (1, 1)$ 89

F.6 Empirical Power rates of five tests at $\alpha = 0.01^*$ when $\boldsymbol{\varepsilon} \sim U(-2,2)$, $(\theta_1, \theta_2, \theta_3) = (1, 1, 1)$ and $(\beta_1, \beta_2) = (1, 1)$. 90

F.7 Empirical Powers of the Five Tests at $\alpha = 0.05^*$ when $\boldsymbol{\varepsilon} \sim N(0,1)$, $(\theta_1, \theta_2, \theta_3) = (1, 1, 1)$ and $(\beta_1, \beta_2) = (1, 1)$ 91

F.8 Empirical Powers of the Five Tests at $\alpha = 0.05^*$ when $\boldsymbol{\varepsilon} \sim U(-2,2)$, $(\theta_1, \theta_2, \theta_3) = (1, 1, 1)$ and $(\beta_1, \beta_2) = (1, 1)$ 92

LIST OF FIGURES

Figures **Page**

4.1 Empirical Type I Errors of the five Tests with increasing Autocorrelation (ρ) at $\alpha = 0.01$ for different Sample Sizes (T) where Errors are Standard Normal and $(\theta_0, \theta_1, \theta_2, \theta_3) = (1,1,1,1)$ 43

4.2 Empirical Type I Errors of the five Tests with increasing Autocorrelation (ρ) at $\alpha = 0.01$ for different Sample Sizes (T) where Errors are Uniform (-2,2) and $(\theta_0, \theta_1, \theta_2, \theta_3) = (1,1,1,1)$ 44

4.3 Empirical Type I Error of the five Tests with increasing Autocorrelation (ρ) at $\alpha = 0.05$ for different Sample Sizes (T) where Errors are Standard Normal and $(\theta_0, \theta_1, \theta_2, \theta_3) = (1,1,1,1)$ 45

4.4 Empirical Type I Errors of the five Tests with increasing Autocorrelation (ρ) at $\alpha = 0.05$ for different Sample Sizes (T) where Errors are Uniform (-2,2) and $(\theta_0, \theta_1, \theta_2, \theta_3) = (1,1,1,1)$ 46

4.5 Empirical Power of the five Tests with increasing Autocorrelation (ρ) at $\alpha = 0.01$ for different Sample Sizes (T) where Errors are Standard Normal, $(\theta_0, \theta_1, \theta_2, \theta_3) = (1,1,1,1)$ and $(\beta_1, \beta_2,) = (1,1)$ 48

4.6 Empirical Powers of the five tests with increasing Autocorrelation (ρ) at $\alpha = 0.01$ for different Sample Sizes (T) where Errors are Uniform (-2,2), $(\theta_0, \theta_1, \theta_2, \theta_3) = (1,1,1,1)$ and $(\beta_1, \beta_2,) = (1,1)$ 49

4.7 Empirical Power of the five tests with increasing Autocorrelation (ρ) at $\alpha = 0.05$ for different Sample Sizes (T) where Errors are Standard Normal, $(\theta_0, \theta_1, \theta_2, \theta_3) = (1,1,1,1)$ and $(\beta_1, \beta_2,) = (1,1)$ 50

4.8 Empirical Power of the five tests with increasing Autocorrelation (ρ) at $\alpha = 0.05$ for different Sample Sizes (T) where Errors are Uniform (-2,2), $(\theta_0, \theta_1, \theta_2, \theta_3) = (1,1,1,1)$ and $(\beta_1, \beta_2,) = (1,1)$. 51

CHAPTER 1

INTRODUCTION

1.1 Background

In an ordinary regression model, the random error terms are assumed to be uncorrelated random variables and, may be, distriubeted as normal. However, in some situations, many regression applications involve time series data where the assumption of uncorrelated or independent error terms is not applicable, i.e. they are frequently correlated positively over time which is called autocorrelated or serially correlated (Kutner et al., 2005: 481). It may occur when one or several key variables are omitted from model, caused by lacking of complete knowledge of the problem and/or the data is not sufficiently precise. The most popular form of autocorrelation is known as the first-order autoregressive model and starts with a linear multiple regression model, such as

$$y_t = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + \ldots + \beta_k X_{kt} + u_t, \qquad \text{for } t = 1, 2, \ldots, T. \qquad (1.1)$$

In the first-order autoregressive model, AR(1), an additional assumption is that u_t's are not i.i.d. for $t = 1, 2, \ldots, T$, but

$$u_t = \rho u_{t-1} + \varepsilon_t,$$

where ρ is called the autocorrelation coefficient, $|\rho| < 1$ and for ε_t's are independent identically distributed random errors.

In practice, a normality assumption on ε_t, as required by traditional statistical methods, is not tenable especially when n is small. As a consequence, the use of ordinary least squares procedure is no longer efficient and the usual t and F tests cannot be suitably applied. Nevertheless, the presence of autocorrelation does not affect other desired properties of an OLS estimator, i.e. the estimator of regression coefficients are still unbiased, but the minimum variance property is no longer satisfied and it may be quite inefficient (Kutner et al., 2005: 481). Anderson (2001: 626-639) suggested the use of permuation test as an alternative approach for situation that does not rely on traditional underlying assumptions.

The permutation test, one of the resampling methods for linear models, was invented in the work of Fisher (1935) and Pitman (1937a: 119-130, 1937b: 225-232, 1937c: 322-325). It had not been of as much interest as the traditional normal-theory tests until the middle of the last decade when computers became very powerful. In the case of a linear model, a permutation test can be applied when the normality assumption for the error terms in the model is not preserved.

Different contexts and applications of permutation tests for linear models have been studied and developed by many authors. Brown and Maritz (1982: 318-331) modified the least squares estimation equation in simple linear regression leading to an exact distribution-free inference about the slope. Freedman and Lane (1983: 292-298), Oja (1987: 91-100), Welch (1990: 693-698), ter Braak (1992: 79-85), Kennedy (1995: 85-94), Manly (1991: 250-261) and Huh and Jhun (2001: 2023-2032) proposed different methods of permutation to test the significance of one or more regression coefficients of multiple linear regression model. Furthermore, Cade (2005: 1049-1054) applied a permutation test for quantile regression, Ptitsyn, Zvonic and Gimble (2006: 1-9) used a permutation test for periodicity in short time series data, and Jung, Jhun and Song (2006: 47-62) developed an exact method of random permutations when testing both interaction and main effects in a two-way ANOVA model. Subsequently, comparison between some proposed permutation methods were done via simulation by many authors. For example, Anderson and Legendre (1999: 271-303) compared type I errors and the power of permutation methods proposed by Freedman and Lane (FL), Manly (M), Kennedy (K), ter Braak (TB) for a test of significance of a single partial regression coefficient in a multiple regression model, the normal theory t test was also included in the study. Anderson and Robinson (2001: 75-88) compared the distributions of a test statistic under the various permutation methods. They found that FL seems to be better in the sense that it is most consistent and reliable. In addition, FL has asymptotic correlation equal to 1 with an exact test on partial regression coefficient. Thus, FL is used as a "control" for comparison to the permutation test proposed in this dissertation.

A major assumption needed for permutation tests is exchangeability, i.e., under null hypothesis, the joint distribution of the observations is invariant under permutations of the subscripts (Good, 2002: 243-247). Note that observations may be exchangeable even though they are not i.i.d., but they must have the same probability of any particular joint outcome regardless of the order in which the observations are considered (Good, 2000 quoted in Lehmann, 1986: 231). Therefore, the autoregressive regression models have to be transformed so that the error terms are written in the form of i.i.d. random variables. Consequently, exchangeable errors are attained no matter what kind of distribution they have (Anderson, 2001: 626-639).

The motivation behind this dissertation is to develop a permutation test for testing partial regression coefficients with first-order autocorrelation where the distribution of the errors term, ε_t, is not necessary normal. Firstly, a time series regression model is transformed so that the errors are i.i.d. and so exchangeability can be applied. After this, a permutation procedure is also proposed for testing.

1.2 Objectives of the Study

This dissertation focuses on an analysis of a permutation test for multiple linear regression with first-order autocorrelation without outliers. The error terms ε_t's are assumed to be i.i.d. with zero mean and constant variance σ^2, but the distribution that is otherwise general is not necessarily normal. The objectives of the study are as follows:

1. To propose a permutation test and a permutation procedure for testing partial regression coefficients for multiple linear regression with first-order autocorrelation.
2. To investigate an asymptotic distribution of the proposed permutation test in 1.
3. To compare the empirical type I error rates and the empirical power of the

proposed permutation test in 1. to a partial F-test from some previously reported permutation tests and the asymptotic test obtained in 2.

1.3 Scope of the Study

1. This study considers the multiple linear regression model in (1.1), where the errors, ε_t, are assumed to be i.i.d. with zero mean and constant variance σ^2 but not normally distributed and no outliers exist.

2. A random permutation test is considered, that is, some randomly elected but not all permutation of subscripts involve.

3. Approximate or asymptotic distribution of the proposed permutation test statistic under the null and alternative hypotheses will be investigated.

4. Estimated level of significance and of power of the proposed permutation test will be compared to those of the partial F-test and some previous works on permutation test.

1.4 Definitions

The following definitions are used in this dissertation:

1. Permutation Matrix. A permutation matrix $\mathbf{S}$ is a square matrix whose columns are be obtained by permuting (rearranging) the columns of an identity matrix. That is, if $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_T$ represent the first, second, …, T$^{\text{th}}$ columns of $\mathbf{I}_T$, then a $T \times T$ permutation matrix is

$$\mathbf{S} = \left(\mathbf{u}_{\pi_1}, \mathbf{u}_{\pi_2}, \ldots, \mathbf{u}_{\pi_T}\right),$$

where $\left(\pi_1, \pi_2, \ldots, \pi_T\right)$ represents a permutation of the positive integers $1, 2, \ldots, T$. (Harville, 1997: 86).

2. Exchangeable. A collection of random variables $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_T$ is said to be exchangeable if, for every permutation $\left(\pi_1, \pi_2, \ldots, \pi_T\right)$ of the integers $\left(1, 2, \ldots, T\right)$,

$$\left(\varepsilon_1, \ldots, \varepsilon_T\right) \stackrel{d}{=} \left(\varepsilon_{\pi 1}, \ldots, \varepsilon_{\pi T}\right),$$

where $\stackrel{d}{=}$ denotes "has the same distribution as" (Randles and Wolfe, 1979: 15)

It was proved by Randles and Wolfe (1979: Theorem 1.3.5, 15), that is if $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_T$ are a collection of i.i.d. random variables, then they are exchangeable. Additionally, Good (2002: 243-247) said that the joint distribution of a set of normally distributed random variables whose covariance matrix is such that all diagonal elements have the same value σ^2 and all the off-diagonal elements have the same value ρ^2 , is invariant under permutations of the variable subscripts.

3. Exact Test. A test is said to be exact if the probability of making a type I error is exactly α for each and every one of the possibilities that make up the hypothesis.

CHAPTER 2

LITERATURE REVIEW

First, this chapter begins with the definition of permutation tests and follows by a review of the relevant features of the test.

2.1 Permutation Tests

Permutation tests, or randomization tests as they are sometimes called, were described by Fisher (1935) and studied theoretically by Pittman (1937a: 119-130, 1937b: 225-232, 1937c: 322-325). Although these terms can be used interchangeably, the difference between permutation tests and randomization tests was mentioned. Hinkelmann and Kempthorne (1994) pointed out the fundamental fact that permutation tests are based on random sampling while the physical act of randomization has been used as the primary justification for randomization tests. Edgington (1995: 337) discussed randomization tests as special type of permutation tests and the production of randomization test reference sets by permuting data, but the rationale is different. The distinction between the assumptions, the rationale underlying the tests and the conceptual gap between both of them were discussed throughout the work of Kempthorne (1955: 946-967, 1969: 231-248).

Although, the first concept of a randomization test was introduced by Fisher (1935), likewise Pitman (1937a: 119-130, 1937b: 225-232, 1937c: 322-325) primarily studied the theory and predicated on the exchangeability of the error distribution. If the interpretation follows Pitman’s studies, then it can be extended to the circumstances of multiple regression not to the data themselves, but to the permutation of the unobservable errors. Conversely, Kempthorne’s work emphasized the analysis of variance which focused on the idea of how the data were randomized or the permuting of subjects to certain positions. The use and theory of permutation or randomization tests have been argued, that is, the idea originated by Fisher and extended by Pitman, and are a cause for theoretical discussion. One of the discussions is without the physical act of randomization and it is difficult to view even the analysis of covariance in terms of a randomization test. Because of the relations of covariance, which can be used in many ways to permute data and can yield different significance tests, so the different theories have implications.

Principally, a permutation test is based on the observed test statistic compared to the possible values of the test statistic given some conditioning sigma-field. Thus, the test statistic is assumed to take on a finite set of values which are equally likely under the null hypothesis. Accordingly, the null hypothesis is rejected if the observed test statistic is too large or too small (or both) when compared to the possible values of the test statistic.

Hoeffding (1952: 169-192) summarized the general idea of the permutation test, as follows:

Let $\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3, \ldots$ be a sequence of p-dimensional vectors defined on some measurable space (Ω, A). Let $\{P_\theta : \theta \in \Theta_1\}$ be a family of probability measures on (Ω, A) and $\mathbf{X}_i$ distributed according to P_θ. The hypothesis we wish to test is

$$H_0 : \theta \in \Theta_0 \quad \text{against} \quad H_1 : \theta \in \Theta_1 .$$

Let V be a test statistic of interest, i.e.,

$$V = g(\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n).$$

Supposed F is some conditioning σ-field and allow any permutation $S(1), S(2), \ldots, S(n)$ of the first n positive integers $(1, 2, \ldots, n)$, then, the permutated data $(\mathbf{X}_{S(1)}, \ldots, \mathbf{X}_{S(n)})$ is a random sample of size n drawn without replacement from $(\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n)$.

As a result, $V^* = g(\mathbf{X}_{S(1)}, \ldots, \mathbf{X}_{S(n)})$ is a permuted statistic.

Let

$$V^{*(1)} \le V^{*(2)} \le \ldots \le V^{*(n!)} ,$$

be the order of all n! possible distinct values of the permuted statistic, including the statistic V, and

$$P_{\theta}\left\{V = V^{*(j)} \mid F\right\} = \frac{1}{n!}, \qquad \text{for all } j = 1, 2, \ldots, n! \text{ and for all } \theta \in \Theta_0,$$

then a test of size α is obtained where the criteria for rejecting the null hypothesis is $V \geq V^{*(k)}$, where $k = n! - [n!\alpha]$ and $[n!\alpha]$ denotes the largest integer less than or equal to $n!\alpha$.

In practice, since the number of permutations n! is usually very large, Vadiveloo (1983: 1581-1596) suggested the sampling permutations test of size q when q was an integer-value represented the numbers of permutation test statistic chosen at random without replacement from n! possible permutatios of V. In addition, for given α, q should be chosen to be large enough so that

$$\frac{k}{q+1} = \alpha .$$

2.2 A Permutation Test for a Multiple Linear Regression Model

Consider the following linear equation for multiple regression;

$$\mathbf{y} = \mathbf{X}\boldsymbol{\tau} + \boldsymbol{\varepsilon} = \mathbf{X}_1\boldsymbol{\theta} + \mathbf{X}_2\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \tag{2.1}$$

where $\mathbf{y}$ is the $n \times 1$ response variable. $\mathbf{X} = (\mathbf{X}_1 \ \ \mathbf{X}_2)$ is an $n \times p$ matrix of predictors of rank p, where $\mathbf{X}_1$ is an $n \times q$ of rank q and $\mathbf{X}_2$ is an $n \times (p-q)$ of rank $p-q$. $\boldsymbol{\tau} = (\boldsymbol{\theta} \ \ \boldsymbol{\beta})'$ is a $p \times 1$ vector of regression coefficient of $\mathbf{y}$, where $\boldsymbol{\theta}$ is a $q \times 1$ vector and $\boldsymbol{\beta}$ is a $(p-q) \times 1$ vector of $\mathbf{X}_1$ and $\mathbf{X}_2$, respectively, and $\boldsymbol{\varepsilon}$ is an $n \times 1$ i.i.d. random error term. From (2.1), normal equations written in matrix form, are the following.

$$\begin{pmatrix} \mathbf{X}_1'\mathbf{X}_1 & \mathbf{X}_1'\mathbf{X}_2 \\ \mathbf{X}_2'\mathbf{X}_1 & \mathbf{X}_2'\mathbf{X}_2 \end{pmatrix}\begin{pmatrix} \boldsymbol{\theta} \\ \boldsymbol{\beta} \end{pmatrix} = \begin{pmatrix} \mathbf{X}_1'\mathbf{y} \\ \mathbf{X}_2'\mathbf{y} \end{pmatrix}. \tag{2.2}$$

Solving the first normal equation for $\boldsymbol{\theta}$ gives $\widehat{\boldsymbol{\theta}} = \left(\mathbf{X}_1'\mathbf{X}_1\right)^{-1}\mathbf{X}_1'\left(\mathbf{y} - \mathbf{X}_2\boldsymbol{\beta}\right)$. When this is absorbed into the second normal equation,

$$\mathbf{X}_2'\left[\mathbf{I} - \mathbf{X}_1\left(\mathbf{X}_1'\mathbf{X}_1\right)^{-1}\mathbf{X}_1'\right]\mathbf{X}_2\boldsymbol{\beta} = \mathbf{X}_2'\left[\mathbf{I} - \mathbf{X}_1'\left(\mathbf{X}_1'\mathbf{X}_1'\right)^{-1}\mathbf{X}_1\right]\mathbf{y},$$

or

$$\mathbf{X}_2'\mathbf{M}_{\mathbf{x}_1}\mathbf{X}_2\boldsymbol{\beta} = \mathbf{X}_2'\mathbf{M}_{\mathbf{x}_1}\mathbf{y}, \tag{2.3}$$

where $\mathbf{M}_{\mathbf{X}_1} = \mathbf{I} - \mathbf{X}_1\left(\mathbf{X}_1'\mathbf{X}_1'\right)^{-1}\mathbf{X}_1'$.

Note that, if $\mathbf{X}_2$ and $\mathbf{X}_1$ are orthogonal, then $\mathbf{M}_{\mathbf{x}_1}\mathbf{X}_1 = 0$ and $\mathbf{M}_{\mathbf{x}_1}\mathbf{X}_2 = \mathbf{X}_2$. The ordinary least squares estimators of the regression coefficient are obtained as follows:

$$\begin{aligned} \hat{\boldsymbol{\beta}}_{\mathrm{OLS}} &= \left(\mathbf{X}_2'\mathbf{M}_{\mathbf{X}_1}\mathbf{X}_2\right)^{-1}\mathbf{X}_2'\mathbf{M}_{\mathbf{X}_1}\mathbf{y}, \\ &= \boldsymbol{\beta} + \left(\mathbf{X}_2'\mathbf{M}_{\mathbf{x}_1}\mathbf{X}_2\right)^{-1}\mathbf{X}_2'\mathbf{M}_{\mathbf{x}_1}\boldsymbol{\varepsilon}. \end{aligned} \tag{2.4}$$

The null hypothesis that we wish to test is $\mathrm{H}_0 : \boldsymbol{\beta} = \mathbf{0}$ (i.e. that the partial regression coefficient for $\mathbf{X}_2$ is not significantly different from zero). Let $\mathrm{F}(\mathbf{y}, \mathbf{X}_1, \mathbf{X}_2)$ be the conventional F statistic used to test this hypothesis: F is defined algebraically in the following equation:

$$\mathrm{F} = \frac{\mathbf{y}'\left[\mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}' - \mathbf{X}_1\left(\mathbf{X}_1'\mathbf{X}_1\right)^{-1}\mathbf{X}_1'\right]\mathbf{y} \Big/ (\mathrm{p} - \mathrm{q})}{\mathbf{y}'\left[\mathbf{I} - \mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\right]\mathbf{y} \Big/ (\mathrm{n} - \mathrm{p})}. \tag{2.5}$$

The degrees of freedom for F are $p-q$ in the numerator and $n-p$ in the denominator. For the nuisance parameter, Lafleur (1999: 6) expressed the sum of squares for the adjusted hypothesis, $R(\boldsymbol{\beta} \mid \boldsymbol{\theta})$, as

$$R(\mathbf{X}_2|\mathbf{X}_1) = (\mathbf{X}_2\boldsymbol{\beta} + \boldsymbol{\varepsilon})' \mathbf{M}_{\mathbf{x}_1} \mathbf{X}_2 (\mathbf{X}_2' \mathbf{M}_{\mathbf{x}_1} \mathbf{X}_2)^{-1} \mathbf{X}_2' \mathbf{M}_{\mathbf{x}_1} (\mathbf{X}_2\boldsymbol{\beta} + \boldsymbol{\varepsilon}). \tag{2.6}$$

A vector of data can be shuffled by multiplying it by a square shuffling matrix $\mathbf{S}$, as described by Harville (1997: 86), so that each column contains zeros save for a single 1 and each column has its 1 in a different row than the 1's appearing in all others columns.

The special properties of a permutation matrix $\mathbf{S}$ are that $\mathbf{S}'\mathbf{S} = \mathbf{I}$ and

$$\begin{aligned}
\mathbf{M}_{\mathbf{S}'\mathbf{x}_1} &= \mathbf{I} - \mathbf{S}'\mathbf{X}_1\left[(\mathbf{S}'\mathbf{X}_1)'\mathbf{S}'\mathbf{X}_1\right]^{-1}(\mathbf{S}'\mathbf{X}_1)', \\
&= \mathbf{S}'\mathbf{S} - \mathbf{S}'\mathbf{X}_1\left[\mathbf{X}_1'\mathbf{X}_1\right]^{-1}(\mathbf{S}'\mathbf{X}_1)', \\
&= \mathbf{S}'\left(\mathbf{I} - \mathbf{X}_1\left[\mathbf{X}_1'\mathbf{X}_1\right]^{-1}\mathbf{X}_1'\right)\mathbf{S}, \\
&= \mathbf{S}'\mathbf{M}_{\mathbf{x}_1}\mathbf{S}.
\end{aligned}$$

Kennedy (1995: 85-94) explored six shuffling methods of conducting a randomization test which were used to permute various parts of the model.

Additionally, under the null hypothesis, $\mathrm{H}_0 : \boldsymbol{\beta} = \mathbf{0}$, the unknown parameter is solved for each shuffling methods. Five of the shuffling methods, as outlined in Kennedy, are discussed below.

2.2.1 Shuffling on $\mathbf{X}_2$. The F statistic for testing $\boldsymbol{\beta} = \mathbf{0}$ is calculated and then compared to the F statistics produced by shuffling $\mathbf{X}_2$. The formula for the shuffled $\boldsymbol{\beta}$ estimate is given by

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}_2'\mathbf{S}'\mathbf{M}_{\mathbf{x}_1}\mathbf{S}\mathbf{X}_2)^{-1}\mathbf{X}_2'\mathbf{S}'\mathbf{M}_{\mathbf{x}_1}(\mathbf{X}_2\boldsymbol{\beta} + \boldsymbol{\varepsilon}), \tag{2.7}$$

and

$$R(\mathbf{X}_2|\mathbf{X}_1) = (\mathbf{X}_2\boldsymbol{\beta} + \boldsymbol{\varepsilon})' \mathbf{M}_{\mathbf{x}_1}\mathbf{S}\mathbf{X}_2\left(\mathbf{X}_2'\mathbf{S}'\mathbf{M}_{\mathbf{x}_1}\mathbf{S}\mathbf{X}_2\right)^{-1}\mathbf{X}_2'\mathbf{S}'\mathbf{M}_{\mathbf{x}_1}(\mathbf{X}_2\boldsymbol{\beta} + \boldsymbol{\varepsilon}).$$

Note that when $\mathbf{S} = \mathbf{I}$ this estimate is reduced to the same value as the estimate $\hat{\boldsymbol{\beta}}_{\mathrm{OLS}}$ of the model with no permutations. Kennedy suggested that this method violates the ancillarity principle, meaning that the natural correlation structure between $\mathbf{X}_1$ and $\mathbf{X}_2$ is split by using this permutation plan.

2.2.2 Shuffling on $\mathbf{y}$. In this method, contributed by Manly (1991: 250-261), the F statistic for testing $\boldsymbol{\beta} = \mathbf{0}$ is calculated and then compared to F statistics produced by shuffling on $\mathbf{y}$. The parameter estimate is given by

$$\begin{aligned}\hat{\boldsymbol{\beta}} &= \left(\mathbf{X}_2'\mathbf{M}_{\mathbf{x}_1}\mathbf{X}_2\right)^{-1}\mathbf{X}_2'\mathbf{M}_{\mathbf{x}_1}\mathbf{S}'\mathbf{y}, \\ &= \left(\mathbf{X}_2'\mathbf{M}_{\mathbf{x}_1}\mathbf{X}_2\right)^{-1}\mathbf{X}_2'\mathbf{M}_{\mathbf{x}_1}\mathbf{S}'(\mathbf{X}_1\boldsymbol{\theta} + \mathbf{X}_2\boldsymbol{\beta} + \boldsymbol{\varepsilon}),\end{aligned} \tag{2.8}$$

and

$$R(\mathbf{X}_2|\mathbf{X}_1) = (\mathbf{X}_1\boldsymbol{\theta} + \mathbf{X}_2\boldsymbol{\beta} + \boldsymbol{\varepsilon})'\mathbf{S}\mathbf{M}_{\mathbf{x}_1}\mathbf{X}_2\left(\mathbf{X}_2'\mathbf{M}_{\mathbf{x}_1}\mathbf{X}_2\right)^{-1}\mathbf{X}_2'\mathbf{M}_{\mathbf{x}_1}\mathbf{S}'(\mathbf{X}_1\boldsymbol{\theta} + \mathbf{X}_2\boldsymbol{\beta} + \boldsymbol{\varepsilon}).$$

Kennedy, noted that this method is unsuitable because it does not recognize that, under the null hypothesis $\boldsymbol{\theta} \neq \mathbf{0}$, part of $\mathbf{y}$ must stay paired with $\mathbf{X}_1$. Moreover, Kennedy and Cade (1996: 923-936), and Anderson and Legendre (1999: 271-303), have shown that the presence of an outlier in the covariable destabilizes this test, often leading to an inflated type I error. In addition, Anderson (2001: 626-639) said shuffling on $\mathbf{y}$ is not a method that should generally be used for tests of partial correlation and if the sample size is quite small (i.e., $\mathrm{n} < 10$), this method avoids having to calculate residuals as estimates of errors.

2.2.3 Residualize $\mathbf{y}$. In order to eliminate the nuisance parameter, $\boldsymbol{\theta}$, and avoid the confounding role played by $\mathbf{X}_1$, Levin and Robbins (1983: 247-267) and Gail, Tan and Piantadosi (1988: 57-64), suggested doing this by residualizing $\mathbf{y}$ for $\mathbf{X}_1$ and then treating the residualized $\mathbf{y}$ as the dependent variable in a regression on $\mathbf{X}_2$. The F statistic for testing $\boldsymbol{\beta}=\mathbf{0}$ is calculated and then compared to F statistics produced by shuffling the residualized $\mathbf{y}$. Under the terms of a permutation analysis,

$$\begin{aligned}\hat{\boldsymbol{\beta}} &= \left(\mathbf{X}_2'\mathbf{X}_2\right)^{-1}\mathbf{X}_2'\mathbf{S}'\mathbf{M}_{\mathbf{x}1}\mathbf{y}\,,\\ &= \left(\mathbf{X}_2'\mathbf{X}_2\right)^{-1}\mathbf{X}_2'\mathbf{S}'\mathbf{M}_{\mathbf{x}1}\left(\mathbf{X}_1\boldsymbol{\theta}+\mathbf{X}_2\boldsymbol{\beta}+\boldsymbol{\varepsilon}\right),\end{aligned} \tag{2.9}$$

and

$$R\left(\mathbf{X}_2|\mathbf{X}_1\right)=\left(\mathbf{X}_1\boldsymbol{\theta}+\mathbf{X}_2\boldsymbol{\beta}+\boldsymbol{\varepsilon}\right)'\mathbf{S}\mathbf{X}_2\left(\mathbf{X}_2'\mathbf{X}_2\right)^{-1}\mathbf{X}_2'\mathbf{S}'\left(\mathbf{X}_1\boldsymbol{\theta}+\mathbf{X}_2\boldsymbol{\beta}+\boldsymbol{\varepsilon}\right).$$

Kennedy found fault with this shuffle plan because type I errors are too low when $\mathbf{X}_1$ and $\mathbf{X}_2$ are not orthogonal. Additionally, Gail et al., (1988: 57-64) assumed that $\mathbf{X}_2$ is assigned independently of $\mathbf{X}_1$ which would cause near-orthogonality, but allowed that desirable properties are only found when $\mathbf{X}_1$ and $\mathbf{X}_2$ are orthogonal.

2.2.4 Residualize $\mathbf{y}$ and $\mathbf{X}_2$. An alternative method suggested by Kennedy first residualizes both $\mathbf{y}$ and $\mathbf{X}_2$ for $\mathbf{X}_1$ and then permutes these residualized values together. This estimate then becomes

$$\begin{aligned}\hat{\boldsymbol{\beta}} &= \left(\mathbf{X}_2'\mathbf{M}_{\mathbf{x}1}\mathbf{X}_2\right)^{-1}\mathbf{X}_2'\mathbf{M}_{\mathbf{x}1}\mathbf{S}'\mathbf{M}_{\mathbf{x}1}\mathbf{y}\,,\\ &= \left(\mathbf{X}_2'\mathbf{M}_{\mathbf{x}1}\mathbf{X}_2\right)^{-1}\mathbf{X}_2'\mathbf{M}_{\mathbf{x}1}\mathbf{S}'\mathbf{M}_{\mathbf{x}1}\left(\mathbf{X}_2\boldsymbol{\beta}+\boldsymbol{\varepsilon}\right),\end{aligned} \tag{2.10}$$

and

$$R\left(\mathbf{X}_2|\mathbf{X}_1\right)=\left(\mathbf{X}_2\boldsymbol{\beta}+\boldsymbol{\varepsilon}\right)'\mathbf{M}_{\mathbf{x}1}\mathbf{S}\mathbf{X}_2\mathbf{M}_{\mathbf{x}1}\left(\mathbf{X}_2'\mathbf{M}_{\mathbf{x}1}\mathbf{X}_2\right)^{-1}\mathbf{M}_{\mathbf{x}1}\mathbf{X}_2'\mathbf{S}'\mathbf{M}_{\mathbf{x}1}\left(\mathbf{X}_2\boldsymbol{\beta}+\boldsymbol{\varepsilon}\right).$$

Note that, if $\mathbf{S}=\mathbf{I}$, then this becomes $\hat{\boldsymbol{\beta}}_{\mathrm{OLS}}$.

2.2.5 Shuffling on Residuals. In this shuffling method, first $\mathbf{y}$ is regressed on $\mathbf{X}_1$, then these $\mathbf{y}$ values are residualized with respect to $\mathbf{X}_1$, followed by the residualized values being permuted, as under shuffle $\mathbf{y}$. These values are then added to the predicted $\mathbf{y}$ to form new $\mathbf{y}$ and regressed on $\mathbf{X}_1$ and $\mathbf{X}_2$. The parameter estimate can be written as

$$\begin{aligned}\hat{\boldsymbol{\beta}} &= \left(\mathbf{X}_2'\mathbf{M}_{\mathbf{x}_1}\mathbf{X}_2\right)^{-1}\mathbf{X}_2'\mathbf{M}_{\mathbf{x}_1}\mathbf{S}'\mathbf{M}_{\mathbf{x}_1}\mathbf{y}, \\ &= \left(\mathbf{X}_2'\mathbf{M}_{\mathbf{x}_1}\mathbf{X}_2\right)^{-1}\mathbf{X}_2'\mathbf{M}_{\mathbf{x}_1}\mathbf{S}'\mathbf{M}_{\mathbf{x}_1}\left(\mathbf{X}_1\boldsymbol{\theta}+\mathbf{X}_2\boldsymbol{\beta}+\boldsymbol{\varepsilon}\right),\end{aligned} \tag{2.11}$$

and

$$R\left(\mathbf{X}_2|\mathbf{X}_1\right)=\left(\mathbf{X}_1\boldsymbol{\theta}+\mathbf{X}_2\boldsymbol{\beta}+\boldsymbol{\varepsilon}\right)'\mathbf{M}_{\mathbf{x}_1}\mathbf{S}\mathbf{M}_{\mathbf{x}_1}\mathbf{X}_2\left(\mathbf{X}_2'\mathbf{M}_{\mathbf{x}_1}\mathbf{X}_2\right)^{-1}\mathbf{X}_2'\mathbf{M}_{\mathbf{x}_1}\mathbf{S}'\mathbf{M}_{\mathbf{x}_1}\left(\mathbf{X}_1\boldsymbol{\theta}+\mathbf{X}_2\boldsymbol{\beta}+\boldsymbol{\varepsilon}\right).$$

For a permutation test in this situation, if the null hypothesis was true, the model would be:

$$\mathbf{y}=\mathbf{X}_1\boldsymbol{\theta}+\boldsymbol{\varepsilon}', \tag{2.12}$$

where $\boldsymbol{\theta}$ is the simple regression coefficient of $\mathbf{y}$ versus $\mathbf{X}_1$. Under the assumption of i.i.d. errors, the exchangeable units under the null hypothesis are error $\boldsymbol{\varepsilon}'$ which should be shuffled for the exact permutation test. However, the parameter $\boldsymbol{\theta}$ is not known and so for the test of partial regression, no exact permutation test is possible (Anderson and Robinson, 2001: 75-88).

The shuffling on residuals method was suggested by, e.g. Freedman and Lane (1983: 292-298), which has been referred to as **permutation under the reduced model**. The rationale for Freedman and Lane's method is that, under the assumption of i.i.d. errors, the exchangeable units under $\mathrm{H}_0:\boldsymbol{\beta}=\mathbf{0}$ are $\mathbf{y}-\mathbf{X}_1\boldsymbol{\theta}$. Although $\boldsymbol{\theta}$ is unknown, it can be estimated by the regression coefficient $\hat{\boldsymbol{\theta}}_{\mathrm{Red}}$. That is, the residuals of this regression $\mathbf{R}_{\mathbf{y}|\mathbf{X}_1}=\mathbf{y}-\mathbf{X}_1\boldsymbol{\theta}$ approximate the error on $\boldsymbol{\varepsilon}'$ that are exchangeable under the null hypothesis (Anderson, 2001: 626-639). Subsequently, for each of n! possible re-orderings, the test statistic under permutation is calculated from the permuted residuals $\mathbf{R}_{\mathbf{y}|\mathbf{X}_1}$. However, $\mathbf{X}_1$ and $\mathbf{X}_2$ are not permuted but remain in their original order. The p-value is calculated as the proportion of values tested under permutations that equal or exceed the value of the test statistic obtained from for the original data.

In addition, ter Braak (1992: 79-85) proposed the permutation of residuals of the full model, rather than residuals of the reduced model as in Freedman and Lane. The idea of ter Braak's method is, if the null hypothesis $\mathrm{H}_0 : \boldsymbol{\beta} = \mathbf{0}$ is true, then the distribution of errors of the full model ($\boldsymbol{\varepsilon}$ in eq. 2.1) should be like the distribution of errors under the null hypothesis ($\boldsymbol{\varepsilon}'$ in eq. 2.13) (Anderson, 2001: 626-639). The values of $\boldsymbol{\theta}$ and $\boldsymbol{\beta}$ are unknown, but the least-squares estimates of these can be obtained as $\hat{\boldsymbol{\theta}}$ and $\hat{\boldsymbol{\beta}}$, respectively. After this, these residuals estimating the error $(\boldsymbol{\varepsilon})$ are calculated as $\mathbf{R}_{\mathbf{y}|\mathbf{X}_1\mathbf{X}_2} = \mathbf{y} - \mathbf{X}_1\boldsymbol{\theta} - \mathbf{X}_2\boldsymbol{\beta}$. These residuals are permuted and, for each of the n! re-orderings, a value of the test statistic is calculated using the permuted residual $\mathbf{R}_{\mathbf{y}|\mathbf{X}_1\mathbf{X}_2}$. As with Freedman and Lane's method, $\mathbf{X}_1$ and $\mathbf{X}_2$ are not permuted but remain in their original order.

Note that these existing methods have been compared theoretically by Anderson and Robinson (2001: 75-88) and in empirical simulations by Anderson and Legendre (1999: 271-303).

Basically, the residuals from a multiple regression model are not independent and variance is not constant. Subsequently, Huh and Jhun (2001: 2023-2032) suggested an alternative to simply randomizing the residuals for testing purposes by proposing the transformation of the residuals to uncorrelated variables. After randomization it is possible to transform back to find the randomized data for the original regression equation, although this is not necessary because testing can be done in terms of the transformed model. This method has been compared by Zocchi and Manly (2006) using simulation with an alternative to using an F-distribution, randomizing observations, randomizing the residual under the full model, and randomizing the residuals under the reduced model. They found that with one extreme data point in the term of the **X** variable and extremely nonnormal regression errors, the Huh and Jhun method gave results similar to those from the other methods, and there was no evidence of it giving better control of size or more power than the others. Given the complexity of the calculations for the Huh and Jhun method, Zocchi and Manly suggested that there is no reason to use this method.

CHAPTER 3

A PROPOSED PERMUTATION TEST

Because of the limitations of the parametric procedures for multiple linear regression with first-order autocorrelation and the advantages of a permutation test, this dissertation proposes a permutation test for testing partial regression coefficients for multiple linear regression with first-order autocorrelation where the distribution of error term is not necessarily normal. In order to propose such a test, it must be established that the observations are exchangeable under the null hypothesis. Therefore, the Prais-Winsten transformation was applied so that errors are i.i.d and hence exchangeability is then obtained. In addition, a permutation method that can be directly conducted to the test without fitting back to the model is proposed. The asymptotic distribution of the proposed test statistic is considered as well.

3.1 Error Transformation

Consider a linear regression model:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\tau} + \mathbf{u}, \tag{3.1}$$

$$u_t = \rho\, u_{t-1} + \varepsilon_t, \qquad \text{for } t = 1,2,\ldots,T,$$

where $\mathbf{y}$ is a $T \times 1$ column vector,

$\mathbf{X}$ is a $T \times p$ matrix of rank p,

$\boldsymbol{\tau}$ is a $p \times 1$ vector of unknown parameter and $\boldsymbol{\tau} = (\boldsymbol{\theta}, \boldsymbol{\beta})'$.

The model (3.1) is called first-order autoregressive if and only if $\mathbf{u}$ is a vector of serial correlated random variables where u_t's when

$$u_t = \rho\, u_{t-1} + \varepsilon_t, \tag{3.2}$$

when ρ is called an autocorrelation coefficient and $|\rho| < 1$, and

$\boldsymbol{\varepsilon}_t$'s are i.i.d. random error with mean zero and variance σ^2.

Exchangeability of observations is a key condition for applying permutation tests. Good (2002: 243-247) gave a definition of exchangeability as "the observations are exchangeable if, under the (null) hypothesis, the joint distribution of the observations is invariant under permutations of the subscripts", as well as "the observations are exchangeable if they are independent and, identically distributed (i.i.d.), or if they are jointly normal with identical covariance. Certainly, elements of $\mathbf{u}$ are not exchangeable. First of all, consider model (3.1), where $\mathbf{u}$ is a vector of errors that are serially correlated. To be able to apply a permutation test, the autocorrelation errors of the data must be appropriately transformed before the permutation testing procedure can be performed (Kennedy, 1995: 85-94). Therefore, model (3.1) must be rewritten in terms of $\boldsymbol{\varepsilon}$, which is a vector of i.i.d. elements as in the following equation (See Appendix A2)

$$\tilde{\mathbf{y}} = \tilde{\mathbf{X}}\ \boldsymbol{\tau} + \boldsymbol{\varepsilon}\ . \tag{3.3}$$

Now that the transformation on the vector $\mathbf{y}$ of dimention $T \times 1$ in equation (3.1) yields in vector $\tilde{\mathbf{y}}$ of dimention $(T-1)\times 1$ in equation (3.3). In small sample, it has been documented that keeping the first observation or omitting it can make a substantial difference in regression results, whereas the loss of one observation in large samples tends to be inconsequential (Jaggia and Kelly-Hawke, 2005). Accordingly, the Prais-Winsten transformation procedure preserves the first observation and equation (3.3) becomes

$$\begin{bmatrix} (1-\rho^2)^{1/2} y_1 \\ y_2 - \rho y_1 \\ y_3 - \rho y_2 \\ \vdots \\ y_T - \rho y_{T-1} \end{bmatrix} = \begin{bmatrix} (1-\rho^2)^{1/2} & (1-\rho^2)^{1/2} X_{11} & \cdots & (1-\rho^2)^{1/2} X_{p1} \\ 1-\rho & X_{12} - \rho X_{11} & \cdots & X_{p2} - \rho X_{p1} \\ 1-\rho & X_{13} - \rho X_{12} & \cdots & X_{p3} - \rho X_{p2} \\ \vdots & \vdots & \ddots & \vdots \\ 1-\rho & X_{1T} - \rho X_{1,T\text{-}1} & \cdots & X_{pT} - \rho X_{p,T\text{-}1} \end{bmatrix} \begin{bmatrix} \tau_0 \\ \tau_1 \\ \tau_2 \\ \vdots \\ \tau_p \end{bmatrix} + \begin{bmatrix} (1-\rho^2)^{1/2} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \vdots \\ \varepsilon_T \end{bmatrix},$$

or

$$\mathbf{y}^* = \mathbf{X}^*\boldsymbol{\tau} + \boldsymbol{\varepsilon}.$$

This is the same as equation (A14) except the first row of $\mathbf{y}^*$ and $\mathbf{X}^*$ are added. In this dissertation, testing of partial regression coefficients is considered. That is, $\mathbf{X}^*$ is partitioned into two submatrixes as $\mathbf{X}^* = \left(\mathbf{X}_1^*, \mathbf{X}_2^*\right)$ where the dimensions of $\mathbf{X}_1^*$ and $\mathbf{X}_2^*$ are $\mathrm{T} \times \mathrm{q}$ and $\mathrm{T} \times \left(\mathrm{p} - \mathrm{q}\right)$, respectively. Consequently, $\boldsymbol{\tau}$ is partitioned accordingly so that

$$\mathbf{y}^* = \mathbf{X}^*\boldsymbol{\tau} + \boldsymbol{\varepsilon} = \mathbf{X}_1^*\boldsymbol{\theta} + \mathbf{X}_2^*\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \tag{3.4}$$

The hypothesis of interest is

$$\mathrm{H}_0 : \boldsymbol{\beta} = \mathbf{0} \qquad \text{vs.} \qquad \mathrm{H}_1 : \boldsymbol{\beta} \neq \mathbf{0}. \tag{3.5}$$

Rejecting H_0 means there is no significant effect of variables in matrix $\mathbf{X}_2^*$ on $\mathbf{y}^*$, and hence $\mathbf{X}_2^*$ should be omitted from the model. So under normality assumption for testing the hypothesis (3.5), is

$$\mathrm{F} = \frac{\mathbf{y}^{*\prime}\left[\mathbf{X}^*\left(\mathbf{X}^{*\prime}\mathbf{X}^*\right)^{-1}\mathbf{X}^{*\prime} - \mathbf{X}_1^*\left(\mathbf{X}_1^{*\prime}\mathbf{X}_1^*\right)^{-1}\mathbf{X}_1^{*\prime}\right]\mathbf{y}^* \Big/ \left(\mathrm{p} - \mathrm{q}\right)}{\mathbf{y}^{*\prime}\left[\mathbf{I} - \mathbf{X}^*\left(\mathbf{X}^{*\prime}\mathbf{X}^*\right)^{-1}\mathbf{X}^{*\prime}\right]\mathbf{y}^* \Big/ \left(\mathrm{T} - \mathrm{p}\right)}. \tag{3.6}$$

where $\hat{\boldsymbol{\beta}}$ is the OLS estimator of the partial regression coefficient $\boldsymbol{\beta}$ and is difined as

$$\hat{\boldsymbol{\beta}} = \left(\mathbf{X}_2^{*\prime}\mathbf{M}\mathbf{X}_2^*\right)^{-1}\left(\mathbf{M}\mathbf{X}_2^*\right)^{\prime}\mathbf{y}^*, \tag{3.7}$$

and $\mathbf{M} = \mathbf{I}_\mathrm{T} - \mathbf{X}_1^*\left(\mathbf{X}_1^{*\prime}\mathbf{X}_1^*\right)^{-1}\mathbf{X}_1^{*\prime}$.

3.2 A Permutation Test for Multiple Linear Regression with First-order Autocorrelation

From equation (3.4), the error term is transformed to i.i.d. and is exchangeable so that a permutation test can be applied. Next, a test of a partial time series regression is developed and proposed. A permuation test statistic, denoted by $\mathrm{F_{ref}}$, can be developed from the partial F statistic defined in equation (3.6) by omitting the number of the degree of freedoms. We can obtain

$$\mathrm{F_{ref}} = \frac{\hat{\boldsymbol{\beta}}'\left(\mathbf{X}_2^{*'}\mathbf{M}\,\mathbf{X}_2^{*}\right)\hat{\boldsymbol{\beta}}}{\mathbf{y}^{*'}\,\mathbf{H}\,\mathbf{y}^{*}}, \tag{3.8}$$

and $\mathbf{H} = \mathbf{I}_\mathrm{T} - \mathbf{X}^*\left(\mathbf{X}^{*'}\mathbf{X}^*\right)^{-1}\mathbf{X}^{*'}$ (See Appendix E). Note that the degrees of freedom are not included in the numerator and denominator because they are multiplicative constants trugh the permutation, thus, there is no effect on the test, neither on power nor type I error (Tantawanich, 2005: 16). In fact, the partial F statistic is a multiple of $\mathrm{F_{ref}}$. When the null hypothesis $\mathrm{H_0}: \boldsymbol{\beta} = \mathbf{0}$ is true, the model (3.4) becomes

$$\mathbf{y}^* = \mathbf{X}_1^*\boldsymbol{\theta} + \boldsymbol{\varepsilon}, \tag{3.9}$$

where $\boldsymbol{\theta}$ is the vector of regression coefficients of $\mathbf{y}^*$ on $\mathbf{X}_1^*$. Therefore, under the assumption of i.i.d. errors, the exchangeable units under $\mathrm{H_0}$ are $\mathbf{y}^* - \mathbf{X}_1^*\boldsymbol{\theta}$.

3.3 The Proposed Permutation Method on the AR(1) Model

This proposed method is based on permuting the residuals under the reduced model in the same manner as the Freedman and Lane (1983: 292-298) method. However, this test can be directly conducted without fitting back to the model which is not the same as, e.g., Freedman and Lane (1983: 292-298), ter Braak (1992: 79-85) and Manly (1991: 250-261).

Notation π is assumed to be a uniformly-chosen random permutation of $\{1, 2, ..., T\}$. Equal weight $1/T!$ are assigned to each of the possible T! true permutations. Used as a superscript, π denotes that the superscripted variable itself is permuted, whereas a subscript π denotes that the subscripted variable is derived from permuted and non- permuted variables.

Recall that, under the null hypothesis, the reduced model is $\mathbf{y}^* = \mathbf{X}_1^*\boldsymbol{\theta} + \boldsymbol{\varepsilon}$ and $\hat{\boldsymbol{\theta}}_{\text{Red}} = \left(\mathbf{X}_1^{*\prime}\mathbf{X}_1^*\right)^{-1}\mathbf{X}_1^{*\prime}\mathbf{y}^*$ is an estimate of the unknown vector of parameter $\boldsymbol{\theta}$ (See Appendix C). The residual of this regression approximates the error $\boldsymbol{\varepsilon}$ that is exchangeable under the null hypothesis and is given by

$$\begin{aligned}
\mathbf{R}_{\mathbf{y}^*|\mathbf{X}_1^*} &= \mathbf{y}^* - \mathbf{X}_1^*\hat{\boldsymbol{\theta}}_{\text{Red}}, \\
&= \mathbf{y}^* - \mathbf{X}_1^*\left[\left(\mathbf{X}_1^{*\prime}\mathbf{X}_1^*\right)^{-1}\mathbf{X}_1^{*\prime}\mathbf{y}^*\right], \\
&= \mathbf{M}\mathbf{y}^*, \qquad (3.10)
\end{aligned}$$

where $\mathbf{M} = \mathbf{I}_T - \mathbf{X}_1^*\left(\mathbf{X}_1^{*\prime}\mathbf{X}_1^*\right)^{-1}\mathbf{X}_1^{*\prime}$.

According to Kennedy (1995: 946-967), a vector of data can be shuffled by multiplying it by a permutation matrix $\mathbf{S}$. Note that the special properties of this matrix is $\mathbf{S}'\mathbf{S} = \mathbf{I}$. Therefore, formally, the new $\mathbf{y}_\pi^* = \mathbf{X}_1^*\left(\mathbf{X}_1^{*\prime}\mathbf{X}_1^*\right)^{-1}\mathbf{X}_1^*\mathbf{y}^* + \mathbf{SMy}^*$.

Hence an estimator $\hat{\boldsymbol{\beta}}$ of $\boldsymbol{\beta}$ in (3.7) under permuted procedure becomes

$$\hat{\boldsymbol{\beta}}_{\pi(P)} = \left(\mathbf{X}_2^{*\prime}\mathbf{M}\mathbf{X}_2^{*}\right)^{-1}\left(\mathbf{M}\mathbf{X}_2^{*}\right)'\left[\mathbf{X}_1^{*}\left(\mathbf{X}_1^{*\prime}\mathbf{X}_1^{*}\right)^{-1}\mathbf{X}_1^{*}\mathbf{y}^{*} + \mathbf{S}\mathbf{M}\mathbf{y}^{*}\right],$$

$$= \left(\mathbf{X}_2^{*\prime}\mathbf{M}\mathbf{X}_2^{*}\right)^{-1}\left(\mathbf{M}\mathbf{X}_2^{*}\right)'\mathbf{S}\mathbf{R}_{\mathbf{y}^{*}|\mathbf{X}_1^{*}}, \tag{3.11}$$

where $\mathbf{M}\mathbf{X}_1^{*} = 0$ (See Appendix B).

After that, the test can be carried out without fitting back to the model which differs from the permutation test based on Freedman and Lane's test. Hence, the proposed test statistic under permutation is

$$F_P = \frac{\hat{\boldsymbol{\beta}}'_{\pi(P)}\left(\mathbf{X}_2^{*\prime}\,\mathbf{M}\,\mathbf{X}_2^{*}\right)\hat{\boldsymbol{\beta}}_{\pi(P)}}{\mathbf{R}^{\pi\prime}_{\mathbf{y}^{*}|\mathbf{X}_1^{*}}\,\mathbf{H}\,\mathbf{R}^{\pi}_{\mathbf{y}^{*}|\mathbf{X}_1^{*}}}. \tag{3.12}$$

For n randomly selected permutations $\{n < T!\}$, the values of the test statistic F_P are calculated via coresponing permuted residual $\mathbf{R}_{\mathbf{y}_\pi^{*}|\mathbf{X}_1^{*}}$. Thus, the significance of H_0 can be obtained as $p-\text{value}=P\left[F_P \geq F_{ref}\right]$. Since, under the null hypothesis, $\boldsymbol{\varepsilon}$ is exchangeable and, under a few weak additional conditions, the residuals $\mathbf{R}_{\mathbf{y}^{*}|\mathbf{X}_1^{*}}$ are asymptotically exchangeable (Schmoyer, 1994: 1507-1516), then F_{ref} has the same distribution as F_P for any permutation on $\mathbf{R}_{\mathbf{y}^{*}|\mathbf{X}_1^{*}}$.

3.4 Procedure for the Proposed Permutation Test (PPT)

The proposed permutation procedure proceeds as follows:

1. Regress $\mathbf{y}^*$ on $\mathbf{X}_1^*$ and $\mathbf{X}_2^*$ and obtain an estimate $\hat{\boldsymbol{\beta}}$ of $\boldsymbol{\beta}$, then calculate

reference value $F_{ref} = \dfrac{\hat{\boldsymbol{\beta}}'\left(\mathbf{X}_2^{*\prime}\,\mathbf{M}\,\mathbf{X}_2^*\right)\hat{\boldsymbol{\beta}}}{\mathbf{y}^{*\prime}\,\mathbf{H}\,\mathbf{y}^*}$

2. Regress $\mathbf{y}^*$ on $\mathbf{X}_1^*$ alone according to the model in Eq. (3.9), and obtain an

estimate $\hat{\boldsymbol{\theta}}_{Red}$ of $\boldsymbol{\theta}$ and residual $\mathbf{R}_{\mathbf{y}^*|\mathbf{X}_1^*}$

3. Randomly permute the residual $\mathbf{R}_{\mathbf{y}^*|\mathbf{X}_1^*}$ to produce a permuted vector $\mathbf{R}^{\pi}_{\mathbf{y}^*|\mathbf{X}_1^*}$

4. Calculate $\hat{\boldsymbol{\beta}}_{\pi(P)} = \left(\mathbf{X}_2^{*\prime}\mathbf{M}\mathbf{X}_2^*\right)^{-1}\left(\mathbf{M}\mathbf{X}_2^*\right)'\mathbf{R}^{\pi}_{\mathbf{y}^*|\mathbf{X}_1^*}$ and $F_P = \dfrac{\hat{\boldsymbol{\beta}}'_{\pi(P)}\left(\mathbf{X}_2^{*\prime}\,\mathbf{M}\,\mathbf{X}_2^*\right)\hat{\boldsymbol{\beta}}_{\pi(P)}}{\mathbf{R}^{\pi\prime}_{\mathbf{y}^*|\mathbf{X}_1^*}\,\mathbf{H}\,\mathbf{R}^{\pi}_{\mathbf{y}^*|\mathbf{X}_1^*}}$

5. Repeat steps 3) and 4) n times where $n < T!$, so that n values of F_P are obtained

6. The values of the statistic F_{ref} and n values of the permuted statistic F_P are ranked from high to low

7. Calculate an estimated significance level $\hat{p} = \dfrac{k}{n+1}$, where k is the rank of F_{ref} among the n values of F_P

8. Reject the null hypothesis if $\hat{p}$ is less than or equal to the significance level α

3.5 Asymptotic Theory

In this section, we begin with the basic assumptions and definitions which are essential for examining the asymptotic distribution of $\mathrm{F_P}$. Using these, the asymptotic distribution of $\mathrm{F_P}$ can be described.

Assumption 3.1 $\boldsymbol{\varepsilon} = (\varepsilon_1, \varepsilon_2, ..., \varepsilon_T)'$ is an vector of independent and identically distributed random errors ε_t, $t = 1, 2, ..., T$ with $E(\varepsilon_t) = 0$, $0 < E(\varepsilon_t^2) = \sigma^2 < \infty$, $E(\varepsilon_s \varepsilon_t) = 0 \ \forall s \neq t$, and $E(\varepsilon_t^4) = \delta^4 < \infty$.

Assumption 3.2 The columns of $\mathbf{X}_1^*$ are linearly independent of the columns of $\mathbf{X}_2^*$**.**

Definition 3.1 The sequence of random vector $\mathbf{y}_t^*$ is bounded in probability denoted by $\mathbf{y}_t^* = O_p(1)$ if, for any $\varepsilon > 0$, there exists a constant K and a value T_0 such that

$$\mathrm{P}\left(\left|\mathbf{y}_t^*\right| \geq K\right) < \varepsilon, \qquad \text{for all } T > T_0. \tag{3.13}$$

See Lehmann (1999: 53-66).

Definition 3.2 The matrix $\mathbf{A}$ is called a **positive definite matrix** if $\mathbf{x}'\mathbf{A}\mathbf{x} > \mathbf{0}$ for all non-null vectors $\mathbf{x} \in \mathcal{R}^T$ and $\mathbf{x}'\mathbf{A}\mathbf{x} = \mathbf{0}$ only when $\mathbf{x}$ is the null vector, i.e. when $\mathbf{x} = \mathbf{0}$.

3.5.1 Asymptotic Distribution of the Permuted Statistic $\mathbf{F_P}$ under the Null Hypothesis

In this section, lemmas involved the asymptotic distribution under the null hypothesis are stated and proved.

Lemma 3.1 Let $\mathbf{S}$ be a $\mathrm{T} \times \mathrm{T}$ permutation matrix. If $\mathbf{A}$ is a $\mathrm{T} \times \mathrm{T}$ positive definite matrix, then

1. $\mathbf{S}'\mathbf{AS}$ is positive definite,
2. $tr(\mathbf{S}'\mathbf{AS}) = tr(\mathbf{A})$,
3. $\lambda_{\mathrm{T}}(\mathbf{S}'\mathbf{AS}) = \lambda_{\mathrm{T}}(\mathbf{A})$, where $\lambda_{\mathrm{T}}(\mathbf{A})$ denotes the largest eigenvalue of $\mathbf{A}$.

Proof

1. Since $\mathbf{A}$ is positive definite, then $\mathbf{v}'\mathbf{Av} > \mathbf{0}$, for all $\mathbf{v} \in \mathcal{R}^{\mathcal{T}}$and $\mathbf{v}'\mathbf{Av} = \mathbf{0}$ when $\mathbf{v} = \mathbf{0}$. Suppose that $\mathbf{x} \in \mathcal{R}^{\mathcal{T}}$, then

$$\mathbf{x}'(\mathbf{S}'\mathbf{AS})\mathbf{x} = (\mathbf{Sx})\mathbf{A}(\mathbf{Sx}),$$

only when $\mathbf{Sx} = 0$. A permutation matrix $\mathbf{S}$ is an elementary matrix when its calculate are permuted columns of the identity matrix. Tharefore is $\mathbf{S}$ invertible. Since $\mathbf{S}$ is invertible, so $\mathbf{x} = 0$ and $\mathbf{x}'(\mathbf{S}'\mathbf{AS})\mathbf{x} = 0$, hence the proof is completes. ■

2. From Harville (1997: 87), a permutation matrixs is an orthogonal matrixs and so $tr(\mathbf{S}'\mathbf{AS}) = tr(\mathbf{ASS}') = tr(\mathbf{A})$, Rencher and Schaalje (2007: 45). ■

3. Suppose $\lambda = \lambda(\mathbf{S}'\mathbf{AS})$ is an eigenvalue of a matrix $\mathbf{S}'\mathbf{AS}$. Therefore, if $\mathbf{x}$ is an eigenvetor of of a matrix $\mathbf{S}'\mathbf{AS}$, $(\mathbf{S}'\mathbf{AS})\mathbf{x} = \lambda\mathbf{x}$. Since $\mathbf{S}'\mathbf{S} = \mathbf{I}$, for $\mathbf{S}$ is orthogonal, so $\mathbf{ASx} = \lambda\mathbf{Sx}$. That is, $\lambda = \lambda(\mathbf{A})$ is eigenvalue of $\mathbf{A}$ and $\mathbf{Sx}$ is igenvector of $\mathbf{A}$. This means that matrix $\mathbf{S}'\mathbf{AS}$ and $\mathbf{A}$ have the same eigenvalues. Hence the largest eigenvalue λ_{T} are concident. ■

Lemma 3.2 Let $\mathbf{C} = \{c_{ij}\}$ be a $\mathrm{T} \times \mathrm{T}$ positive definite matrix such that $tr(\mathbf{C}) \le c < \infty$, where c is a constant, then $\mathbf{R}'_{\mathbf{y}^*}\big|_{\mathbf{x}_1^*} \mathbf{CR}_{\mathbf{y}^*}\big|_{\mathbf{x}_1^*}$ is bounded in probability, i.e. $\mathbf{R}'_{\mathbf{y}^*}\big|_{\mathbf{x}_1^*} \mathbf{CR}_{\mathbf{y}^*}\big|_{\mathbf{x}_1^*} = O_P(1)$.

Proof Since $\mathbf{C}$ is positive definite, so $\left|\mathbf{R}'_{y^*|x_1^*}\mathbf{C}\mathbf{R}_{y^*|x_1^*}\right| = \left(\mathbf{R}'_{y^*|x_1^*}\mathbf{C}\mathbf{R}_{y^*|x_1^*}\right)$.

In addition, using definition 2.4.2 from Ravishanker and Dey (2002: 51),

$$\begin{aligned}
\mathrm{E}\left|\mathbf{R}'_{y^*|x_1^*}\mathbf{C}\mathbf{R}_{y^*|x_1^*}\right| &= \mathrm{E}\left(\mathbf{R}'_{y^*|x_1^*}\mathbf{C}\mathbf{R}_{y^*|x_1^*}\right), \\
&= \mathrm{E}\left(\sum_{i=1}^{T}\sum_{j=1}^{T} c_{ij}\,\mathrm{R}_i\mathrm{R}_j\right), \\
&= \sum_{i=1}^{T} c_{ii}\mathrm{E}\left(\mathrm{R}_i^2\right), \\
&= tr(\mathbf{C})\sigma^2\mathbf{M} \quad \le c\sigma^2\mathbf{M}, \text{ (See Appendix D2).}
\end{aligned}$$

It can now be shown that, for any $\varepsilon > 0$, there exists a constant λ such that $P\left(\left|\mathbf{R}'_{y^*|x_1^*}\mathbf{C}\mathbf{R}_{y^*|x_1^*}\right| \ge \lambda\right) < \varepsilon$.

Let $\varepsilon > 0$ and choose $\lambda = \dfrac{2c\sigma^2\mathbf{M}}{\varepsilon}$ then, by using Markov's inequality,

$$P\left(\left|\mathbf{R}'_{y^*|x_1^*}\mathbf{C}\mathbf{R}_{y^*|x_1^*}\right| \ge \lambda\right) < \frac{\mathrm{E}\left(\left|\mathbf{R}'_{y^*|x_1^*}\mathbf{C}\mathbf{R}_{y^*|x_1^*}\right|\right)}{\lambda} \le \frac{c\sigma^2\mathbf{M}\varepsilon}{2c\sigma^2\mathbf{M}} = \frac{\varepsilon}{2} < \varepsilon,$$

and hence $\mathbf{R}'_{y^*|x_1^*}\mathbf{C}\mathbf{R}_{y^*|x_1^*} = O_P(1)$ as defined in definition 3.1. ■

Lemma 3.3 Suppose the assumption 3.2 holds, then the following quadratic forms are all bounded in probability:

1. $\mathbf{R}'_{y^*|x_1^*}\mathbf{H}_{x^*}\mathbf{R}_{y^*|x_1^*}$

2. $\mathbf{R}'_{y^*|x_1^*}\mathbf{W}\mathbf{R}_{y^*|x_1^*}$

3. $\mathbf{R}'_{y^*|x_1^*}\mathbf{S}'\mathbf{H}_{x^*}\,\mathbf{S}\,\mathbf{R}_{y^*|x_1^*}$

4. $\mathbf{R}'_{y^*|x_1^*}\,\mathbf{S}'\,\mathbf{W}\,\mathbf{S}\,\mathbf{R}_{y^*|x_1^*}$

where $\mathbf{H}_{\mathbf{X}^*} = \mathbf{X}^*\left(\mathbf{X}^{*\prime}\mathbf{X}^*\right)^{-1}\mathbf{X}^{*\prime}$.

Proof

From appendix B1, $\mathbf{H}_{\mathbf{X}^*}$ is a positive definite which gives $tr(\mathbf{H}_{\mathbf{X}^*}) = \mathrm{p}$ and $\mathbf{H}_{\mathbf{X}^*}\mathbf{H}_{\mathbf{X}^*} = \mathbf{H}_{\mathbf{X}^*}$. In addition, from appendix B5, $\mathbf{W}$ is a positive definite which gives $tr(\mathbf{W}) = \mathrm{p} - \mathrm{q}$. Therefore:

1. and 2. The proof follows directly from the result of lemma 3.2. ■

3. Since $\mathbf{H}_{\mathbf{X}^*}$ is positive definite and by lemma 3.1, then $\mathbf{S}'\,\mathbf{H}_{\mathbf{X}^*}\mathbf{S}$ is positive definite and $tr\left(\mathbf{S}'\,\mathbf{H}_{\mathbf{X}^*}\,\mathbf{S}\right) = tr\left(\mathbf{H}_{\mathbf{X}^*}\right) = \mathrm{p} < \infty$.

Hence, $\mathbf{R}'_{y^*|x_1^*}\,\mathbf{S}'\,\mathbf{H}_{\mathbf{X}^*}\,\mathbf{S}\,\mathbf{R}_{y^*|x_1^*}$ is bounded in probability. ■

4. Since $\mathbf{W}$ is positive definite and by lemma 3.1, then $\mathbf{S}'\mathbf{W}\mathbf{S}$ is positive definite and $tr(\mathbf{S}'\mathbf{W}\mathbf{S}) = tr(\mathbf{W}) = \mathrm{p} - \mathrm{q} < \infty$.

Hence, $\mathbf{R}'_{y^*|x_1^*}\,\mathbf{S}'\,\mathbf{W}\,\mathbf{S}\,\mathbf{R}_{y^*|x_1^*}$ is bounded in probability. ■

Lemma 3.4 Suppose assumption 3.1 and 3.2 hold, then:

1. $\mathrm{T}^{-1}\left(\mathbf{R}'_{y^*|x_1^*}\,\mathbf{R}_{y^*|x_1^*}\right) \xrightarrow{P} \sigma^2\mathbf{M}$

2. $\mathrm{T}^{-1}\left(\mathbf{R}'_{y^*|x_1^*}\,\mathbf{S}'\,\mathbf{H}\,\mathbf{S}\,\mathbf{R}_{y^*|x_1^*}\right) \xrightarrow{P} \sigma^2\mathbf{M}$

Proof

1. Want to prove that

$$\lim_{T\to\infty} P\left(\left|T^{-1}\left(\mathbf{R}'_{y^*|x_1^*}\mathbf{R}_{y^*|x_1^*}\right)-\sigma^2\mathbf{M}\right|>\varepsilon\right)\le 0 \ , \ \forall\varepsilon>0\ . \tag{3.14}$$

According to appendix D2, the mean of variance of $T^{-1}\mathbf{R}'_{y^*|x_1^*}\mathbf{R}_{y^*|x_1^*}$ are

$$E\left(T^{-1}\mathbf{R}'_{y^*|x_1^*}\mathbf{R}_{y^*|x_1^*}\right)=\sigma^2\mathbf{M} \text{ and } \mathrm{Var}\left(T^{-1}\mathbf{R}'_{y^*|x_1^*}\mathbf{R}_{y^*|x_1^*}\right)=2T^{-2}\sigma^4(T-q).$$

Applying the Chebyshev inequality we get

$$0\le P\left(\left|T^{-1}\left(\mathbf{R}'_{y^*|x_1^*}\mathbf{R}_{y^*|x_1^*}\right)-\sigma^2\mathbf{M}\right|>\varepsilon\right)\le\frac{1}{\varepsilon^2}\mathrm{Var}\left(T^{-1}\mathbf{R}'_{y^*|x_1^*}\mathbf{R}_{y^*|x_1^*}\right)$$

$$=\frac{2\sigma^4(T-q)}{T^2\varepsilon^2},\ \forall\varepsilon>0\ .$$

As $T\to\infty$. The right hand side to zero and hence the proof is completes. ■

2. Since $\mathbf{H}=\mathbf{I}_T-\mathbf{H}_{X^*}$, then

$$T^{-1}\left(\mathbf{R}'_{y^*|x_1^*}\mathbf{S}'\mathbf{H}\,\mathbf{S}\,\mathbf{R}_{y^*|x_1^*}\right)=T^{-1}\left[\mathbf{R}'_{y^*|x_1^*}\mathbf{S}'\left(\mathbf{I}_T-\mathbf{H}_{X^*}\right)\mathbf{S}\mathbf{R}_{y^*|x_1^*}\right],$$

$$=T^{-1}\mathbf{R}'_{y^*|x_1^*}\mathbf{R}_{y^*|x_1^*}-T^{-1}\mathbf{R}'_{y^*|x_1^*}\mathbf{S}'\,\mathbf{H}_{X^*}\,\mathbf{S}\,\mathbf{R}_{y^*|x_1^*}\ . \tag{3.15}$$

By lemma 3.3, $\mathbf{R}'_{y^*|x_1^*}\mathbf{S}'\,\mathbf{H}_{X^*}\,\mathbf{S}\,\mathbf{R}_{y^*|x_1^*}$ is bounded in probability and the property of permutation matrix $\mathbf{S}$ is such that $\mathbf{S}'\mathbf{S}=\mathbf{I}$, then $T^{-1}\mathbf{R}'_{y^*|x_1^*}\mathbf{S}'\,\mathbf{H}_{X^*}\,\mathbf{S}\,\mathbf{R}_{y^*|x_1^*}\xrightarrow{P}0$.

In addition, recall the result from (1); as $T\to\infty$, $T^{-1}\left(\mathbf{R}'_{y^*|x_1^*}\mathbf{R}_{y^*|x_1^*}\right)\xrightarrow{P}\sigma^2\mathbf{M}$.

This completes the proof. ■

For the following results, want to find the distribution of the permuted statistic F_P. From (3.11), the permuted statistic F_P in (3.12) can be expressed as

$$F_P = \frac{\hat{\boldsymbol{\beta}}'_{\pi(P)}\left(\mathbf{X}_2^{*\prime}\mathbf{M}\,\mathbf{X}_2^*\right)\hat{\boldsymbol{\beta}}_{\pi(P)}}{\mathbf{R}^{\pi\prime}_{\mathbf{y}^*|\mathbf{x}_1^*}\,\mathbf{H}\,\mathbf{R}^{\pi}_{\mathbf{y}^*|\mathbf{x}_1^*}} \; ,$$

$$= \frac{\left[\left(\mathbf{X}_2^{*\prime}\mathbf{M}\mathbf{X}_2^*\right)^{-1}\left(\mathbf{M}\mathbf{X}_2^*\right)'\mathbf{S}\mathbf{R}_{\mathbf{y}^*|\mathbf{x}_1^*}\right]'\left(\mathbf{X}_2^{*\prime}\mathbf{M}\,\mathbf{X}_2^*\right)\left[\left(\mathbf{X}_2^{*\prime}\mathbf{M}\mathbf{X}_2^*\right)^{-1}\left(\mathbf{M}\mathbf{X}_2^*\right)'\mathbf{S}\mathbf{R}_{\mathbf{y}^*|\mathbf{x}_1^*}\right]}{\mathbf{R}'_{\mathbf{y}^*|\mathbf{x}_1^*}\,\mathbf{S}'\,\mathbf{H}\,\mathbf{S}\,\mathbf{R}_{\mathbf{y}^*|\mathbf{x}_1^*}},$$

$$= \frac{\mathbf{R}'_{\mathbf{y}^*|\mathbf{x}_1^*}\,\mathbf{S}'\,\mathbf{M}\,\mathbf{X}_2^*\left(\mathbf{X}_2^{*\prime}\mathbf{M}\,\mathbf{X}_2^*\right)^{-1}\mathbf{X}_2^{*\prime}\mathbf{M}\,\mathbf{S}\mathbf{R}_{\mathbf{y}^*|\mathbf{x}_1^*}}{\mathbf{R}'_{\mathbf{y}^*|\mathbf{x}_1^*}\,\mathbf{S}'\,\mathbf{H}\,\mathbf{S}\,\mathbf{R}_{\mathbf{y}^*|\mathbf{x}_1^*}},$$

$$= \frac{\mathbf{R}'_{\mathbf{y}^*|\mathbf{x}_1^*}\,\mathbf{S}'\,\mathbf{W}\,\mathbf{S}\mathbf{R}_{\mathbf{y}^*|\mathbf{x}_1^*}}{\mathbf{R}'_{\mathbf{y}^*|\mathbf{x}_1^*}\,\mathbf{S}'\,\mathbf{H}\,\mathbf{S}\,\mathbf{R}_{\mathbf{y}^*|\mathbf{x}_1^*}} \; , \tag{3.16}$$

where $\mathbf{W} = \mathbf{M}\mathbf{X}_2^*\left(\mathbf{X}_2^{*\prime}\mathbf{M}\mathbf{X}_2^*\right)^{-1}\mathbf{X}_2^{*\prime}\mathbf{M}$.

Next, we want to find the distribution of the quadratic form $\mathbf{R}'_{\mathbf{y}^*|\mathbf{x}_1^*}\,\mathbf{S}'\,\mathbf{W}\,\mathbf{S}\,\mathbf{R}_{\mathbf{y}^*|\mathbf{x}_1^*}$. It can be shown that $\mathbf{A} = \mathbf{S}'\mathbf{W}\mathbf{S}$ is symmetric idempotent matrix of rank $(p-q) = d$ (see lemma 3.3), so there exists a $T \times T$ orthogonal matrix $\mathbf{P}$ such that

$$\mathbf{P}'\mathbf{A}\mathbf{P} = \begin{bmatrix} \mathbf{I}_d & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}, \tag{3.17}$$

(Ravishanker and Dey, 2002: 172).

Define the $T \times 1$ random vector $\mathbf{Z}$ by $\mathbf{Z} = \mathbf{P}'\mathbf{R}$. Suppose that we partition $\mathbf{P}$ as $\mathbf{P} = (\mathbf{P}_1 \quad \mathbf{P}_2)$, where $\mathbf{P}_1{}'\mathbf{P}_1 = \mathbf{I}_d$, and create partition $\mathbf{Z}' = \left(\mathbf{Z}_1{}' \quad \mathbf{Z}_2{}'\right)$, where $\mathbf{Z}_1$ is a $d \times 1$ vector, then

$$\mathbf{Z}_1 = \mathbf{P}_1{}'\mathbf{R} = \begin{bmatrix} p_{11} & \cdots & p_{1d} \\ \vdots & \ddots & \vdots \\ p_{T1} & \cdots & p_{Td} \end{bmatrix}' \begin{bmatrix} R_1 \\ \vdots \\ R_T \end{bmatrix},$$

$$= \begin{bmatrix} p_{11}R_1 + p_{21}R_2 + \ldots + p_{T1}R_T \\ p_{12}R_1 + p_{22}R_2 + \ldots + p_{T2}R_T \\ \vdots \\ p_{1d}R_1 + p_{2d}R_2 + \ldots + p_{Td}R_T \end{bmatrix},$$

$$= \left(\sum_{i=1}^{T} p_{i1}R_i \quad \sum_{i=1}^{T} p_{i2}R_i \quad \cdots \quad \sum_{i=1}^{T} p_{id}R_i \right)'. \tag{3.18}$$

Hence,

$$\mathbf{R}'_{y^*}\big|_{x_1^*} \mathbf{A}\, \mathbf{R}_{y^*}\big|_{x_1^*} = (\mathbf{PZ})'\mathbf{A}(\mathbf{PZ}) = \mathbf{Z}'\mathbf{P}'\mathbf{APZ},$$

$$= \left(\mathbf{Z}_1{}' \quad \mathbf{Z}_2{}'\right) \begin{pmatrix} \mathbf{I}_d & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \begin{pmatrix} \mathbf{Z}_1 \\ \mathbf{Z}_2 \end{pmatrix},$$

$$= \mathbf{Z}_1{}'\mathbf{Z}_1 \quad = \sum_{j=1}^{d} \left(\sum_{i=1}^{T} p_{ij}R_i \right)^2,$$

$$= \sum_{j=1}^{d} v_{\bullet j}^2 \qquad \text{where } v_{\bullet j} = \sum_{i=1}^{T} p_{ij}R_i \quad . \tag{3.19}$$

Result 3.1 Let $\bar{v}_{\bullet j} = \frac{1}{T}\sum_{i=1}^{T}(p_{ij}R_i)$, so $E(\bar{v}_{\bullet j}) = 0$ and $\mathrm{Var}(\bar{v}_{\bullet j}) = \sigma^2\mathbf{M}/T^2 < \infty$.

Proof Since $\mathbf{P} = (p_{ij})$ is orthogonal so that $v_{\bullet j}$ are i.i.d.,

then

$$\mathrm{E}(\bar{v}_{\bullet j}) = \frac{1}{\mathrm{T}}\sum_{i=1}^{\mathrm{T}} \mathrm{p_{ij}E(R_i)} = 0,$$

$$\mathrm{Var}(\bar{v}_{\bullet j}) = \frac{1}{\mathrm{T}^2}\sum_{i=1}^{\mathrm{T}} \mathrm{p_{ij}^2 Var(R_i)} = \frac{\sigma^2 \mathbf{M}}{\mathrm{T}^2}, \text{ where } \sum_{i=1}^{\mathrm{T}} \mathrm{p_{ij}^2} = 1 \text{ ,from } \mathbf{P}_1'\mathbf{P}_1 = \mathbf{I}_\mathbf{d}. \ ■$$

Lemma 3.5 Let $Q_j = \dfrac{\bar{v}_{\bullet j} - \mathrm{E}(\bar{v}_{\bullet j})}{\sqrt{\mathrm{Var}(\bar{v}_{\bullet j})}}$, for $j = 1, 2, ..., \mathrm{d}$, be i.i.d , and let Φ be the distribution function of $Z \sim N(0,1)$, then

$$Q_j \xrightarrow{L} \Phi(z) \ .$$

Proof In the following proof, we set out with

$$Q_j = \frac{\bar{v}_{\bullet j} - \mathrm{E}(\bar{v}_{\bullet j})}{\sqrt{\mathrm{Var}(\bar{v}_{\bullet j})}} = \frac{\left(\sum_{i=1}^{\mathrm{T}} \mathrm{p_{ij}R_i} \Big/ \mathrm{T}\right)}{\left(\sigma^2 \mathbf{M}/\mathrm{T}^2\right)^{1/2}} = \sigma^{-1}\mathbf{M}^{-1/2}\sum_{i=1}^{\mathrm{T}} \mathrm{p_{ij}R_i} \ . \tag{3.20}$$

Subsequently,

$\mathrm{E}(Q_j) = 0$ and

$$\mathrm{Var}(Q_j) = \sigma^{-2}\mathbf{M}^{-1}\sum_{i=1}^{\mathrm{T}} \mathrm{p_{ij}^2 Var(R_i)} = 1, \text{ where } \sum_{i=1}^{\mathrm{T}} \mathrm{p_{ij}^2} = 1 \text{ , from } \mathbf{P}_1'\mathbf{P}_1 = \mathbf{I}_\mathbf{d},$$

which is the result following from theorem 2.4.1, the classical central limit theorem (CLT), from Lehmann (1999). Hence $Q_j \xrightarrow{L} \Phi(z)$. ■

Lemma 3.6 Let $Q_j = \dfrac{\bar{v}_{\bullet \mathrm{j}} - \mathrm{E}(\bar{v}_{\bullet \mathrm{j}})}{\sqrt{\mathrm{Var}(\bar{v}_{\bullet \mathrm{j}})}}$, for $j = 1, 2, ..., \mathrm{d}$, be i.i.d., then;

1. $Q_j^2 \xrightarrow{\mathrm{L}} \chi^2_{(1)}$

2. $\sum_{\mathrm{j}=1}^{\mathrm{d}} Q_{\mathrm{j}}^2 \xrightarrow{\mathrm{L}} \chi^2_{(d)}$

Proof

1. From lemma 3.5, we have $Q_j^2 = \sigma^{-2}\mathbf{M}^{-1}\left(\sum_{i=1}^{T} p_{ij}R_i\right)^2 = \sigma^{-2}\mathbf{M}^{-1}v_{\bullet j}^2$.

It follows directly from lemma 5.3.2 in Casella and Berger (2002) that $Q_j^2 \xrightarrow{L} \chi^2_{(1)}$.

2. Similarly, we can also write $\sum_{j=1}^{d} Q_j^2 = \sum_{j=1}^{d} \frac{\left(\sum_{i=1}^{T} p_{ij}R_i\right)^2}{\sigma^2\mathbf{M}}$. It follows directly from lemma 5.3.2 in Casella and Berger (2002) that $\sum_{j=1r}^{d} Q_j^2 \xrightarrow{L} \chi^2_{(d)}$. ■

The following theorem is the main result of the proof of the asymptotic distribution of F_P under the null hypothesis.

Theorem 3.1 Consider a multiple linear regression model with first-order autoregression, $\mathbf{y}^* = \mathbf{X}_1^*\boldsymbol{\theta} + \mathbf{X}_2^*\boldsymbol{\beta} + \boldsymbol{\varepsilon}$; $\mathbf{y}^*$ and $\mathbf{X}^*$ are given in (3.4), where the errors $\boldsymbol{\varepsilon}$ are assumed to be i.i.d. $(0, \sigma^2)$. Under the null hypothesis, $H_0 : \boldsymbol{\beta} = \mathbf{0}$,

$$T^{-1}F_P \xrightarrow{L} \chi^2_{(p-q)}, \tag{3.21}$$

where the permuted statistic F_P is given in (3.16), $p-q$ is the total number of partial regression coefficients to be tested and T is the sample size. Subsequently, $T^{-1}F_P$ converges in distribution to a central chi-squared distribution with $p-q$ degrees of freedom.

Proof From lemma 3.5, we have

$$\sigma^{-1}\mathbf{M}^{-1/2}\sum_{i=1}^{T} p_{ij}R_i \xrightarrow{L} \Phi(z),$$

where Φ is the distribution function of $Z \sim N(0,1)$. (3.22)

Similarly, from lemma 3.6,

$$\sigma^{-2}\mathbf{M}^{-1}\left(\sum_{i=1}^{T} p_{ij}R_i\right)^2 \xrightarrow{L} \chi^2_{(1)}. \tag{3.23}$$

This implies that

$$\sum_{j=1}^{p-q}\left(\sum_{i=1}^{T} p_{ij}R_i\right)^2 \xrightarrow{L} \sigma^2\mathbf{M}\,\chi^2_{(p-q)}. \tag{3.24}$$

Since, from lemma 3.4(2), $T^{-1}\left(\mathbf{R}'_{y^*}\big|_{x_1^*}\mathbf{S}'\,\mathbf{H}\,\mathbf{S}\,\mathbf{R}_{y^*}\big|_{x_1^*}\right) \xrightarrow{P} \sigma^2\mathbf{M}$, by using Slutsky's theorem, we get

$$\frac{\sum_{j=1}^{p-q}\left(\sum_{i=1}^{T} p_{ij}R_i\right)^2}{T^{-1}\left(\mathbf{R}'_{y^*}\big|_{x_1^*}\mathbf{S}'\,\mathbf{H}\,\mathbf{S}\,\mathbf{R}_{y^*}\big|_{x_1^*}\right)} \xrightarrow{L} \chi^2_{(p-q)}. \tag{3.25}$$

From equation (3.19), $\mathbf{R}'_{y^*}\big|_{x_1^*}\mathbf{S}'\,\mathbf{W}\,\mathbf{S}\,\mathbf{R}_{y^*}\big|_{x_1^*} = \sum_{j=1}^{p-q} v^2_{\bullet j}$, where $v_{\bullet j} = \sum_{i=1}^{T} p_{ij}R_i$, then, from (3.25), this can be rewritten as

$$T^{-1}F_p \xrightarrow{L} \chi^2_{(p-q)}, \tag{3.26}$$

which completes the proof. ■

3.5.2 Asymptotic Distribution of Permuted Statistic F_P under the Alternative Hypothesis

Lemmas and results, necessary for finding the asymptotic distribution of the test statistic are now stated.

Result 3.2 From assumption 3.1, the error vector $\boldsymbol{\varepsilon}$ is independent and identically distributed with $E(\varepsilon_t)=0$, $0<E(\varepsilon_t^2)=\sigma^2<\infty$, for $t=1,2,\dots,T$. If the null hypothesis is false, $H_1:\boldsymbol{\beta}\neq\mathbf{0}$, then $E(\mathbf{y}^*)=\mathbf{X}_1^*\boldsymbol{\theta}+\mathbf{X}_2^*\boldsymbol{\beta}$ and $\mathrm{Var}(\mathbf{y}^*)=\sigma^2\mathbf{I}$.

Proof $E(\mathbf{y}^*)=E(\mathbf{X}_1^*\boldsymbol{\theta}+\mathbf{X}_2^*\boldsymbol{\beta}+\boldsymbol{\varepsilon})=\mathbf{X}_1^*\boldsymbol{\theta}+\mathbf{X}_2^*\boldsymbol{\beta}$ and

$$\mathrm{Var}(\mathbf{y}^*)=\mathrm{Var}(\mathbf{X}_1^*\boldsymbol{\theta}+\mathbf{X}_2^*\boldsymbol{\beta}+\boldsymbol{\varepsilon})=\sigma^2\mathbf{I}. \qquad \blacksquare$$

Result 3.3 Under the alternative hypothesis, $H_1:\boldsymbol{\beta}\neq\mathbf{0}$, and result 3.2 , then:

1. $E\left(\mathbf{R}_{\mathbf{y}^*|\mathbf{x}_1^*}\right)=\mathbf{M}\mathbf{X}_2^*\boldsymbol{\beta}$

2. $\mathrm{Var}\left(\mathbf{R}_{\mathbf{y}^*|\mathbf{x}_1^*}\right)=\sigma^2\mathbf{M}$

Proof

1. $E\left(\mathbf{R}_{\mathbf{y}^*|\mathbf{x}_1^*}\right)=E(\mathbf{M}\mathbf{y}^*)$,

$$=\mathbf{M}\,E(\mathbf{X}_1^*\boldsymbol{\theta}+\mathbf{X}_2^*\boldsymbol{\beta}+\boldsymbol{\varepsilon})=\mathbf{M}\mathbf{X}_2^*\boldsymbol{\beta}. \qquad \blacksquare$$

2. $\mathrm{Var}\left(\mathbf{R}_{\mathbf{y}^*|\mathbf{x}_1^*}\right)=\mathbf{M}'\,\mathrm{Var}(\mathbf{y}^*)\mathbf{M}$,

$$= \mathbf{M}' \, \mathrm{Var}\left(\mathbf{X}_1^{*}\boldsymbol{\theta} + \mathbf{X}_2^{*}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \right)\mathbf{M} \;\; = \mathbf{M}'\left(\sigma^2\mathbf{I}\right)\mathbf{M} \;\; = \sigma^2\mathbf{M}. \qquad \blacksquare$$

Using the following results, we want to find a distribution of quadratic form, $\mathbf{R}'_{y^*|\mathbf{x}_1^*}\mathbf{S}'\,\mathbf{W}\,\mathbf{S}\,\mathbf{R}_{y^*|\mathbf{x}_1^*}$. Since $\mathbf{A} = \mathbf{S}'\mathbf{W}\mathbf{S}$ is a symmetric and idempotent matrix of rank $(p-q) = d$ (see lemma 3.3), then there exists a $T \times T$ orthogonal matrix $\mathbf{P}$ such that

$$\mathbf{P}'\mathbf{A}\mathbf{P} = \begin{bmatrix} \mathbf{I}_d & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}, \tag{3.27}$$

(see Ravishanker and Dey, 2002: 172).

Define the $T \times 1$ random vector $\mathbf{Z}$ as $\mathbf{Z} = \mathbf{P}'\,\mathbf{R}_{y^*|\mathbf{x}_1^*}$. Suppose we partition $\mathbf{P}$ as $\mathbf{P} = \left(\mathbf{P}_1 \quad \mathbf{P}_2\right)$, where $\mathbf{P}_1'\mathbf{P}_1 = \mathbf{I}_d$, and partition $\mathbf{Z}' = \left(\mathbf{Z}_1' \quad \mathbf{Z}_2'\right)$, where $\mathbf{Z}_1$ is a $d \times 1$ vector, then

$$\begin{aligned}
\mathbf{Z}_1 = \mathbf{P}_1'\mathbf{R}_{y^*|\mathbf{x}_1^*} &= \begin{bmatrix} p_{11} & \cdots & p_{1d} \\ \vdots & \ddots & \vdots \\ p_{T1} & \cdots & p_{Td} \end{bmatrix}' \begin{bmatrix} R_1 \\ \vdots \\ R_T \end{bmatrix}, \\
&= \left[\mathbf{p}_1'\mathbf{R} \quad \mathbf{p}_2'\mathbf{R} \quad \ldots \quad \mathbf{p}_d'\mathbf{R}\right]', \\
&= \left[\sum_{i=1}^{T} p_{i1}R_i \quad \sum_{i=1}^{T} p_{i2}R_i \quad \cdots \quad \sum_{i=1}^{T} p_{id}R_i\right]',
\end{aligned} \tag{3.28}$$

where $\mathbf{p}_j' = \left[p_{1j} \quad p_{2j} \quad \ldots \quad p_{Tj}\right]$ is the jth row of $\mathbf{P}_1'$, for $j = 1, 2, \ldots, d$.

Hence,

$$\begin{aligned}
\mathbf{R}'_{y^*|\mathbf{x}_1^*}\,\mathbf{A}\,\mathbf{R}_{y^*|\mathbf{x}_1^*} &= \left(\mathbf{P}\mathbf{Z}\right)'\mathbf{A}\left(\mathbf{P}\mathbf{Z}\right) = \mathbf{Z}'\,\mathbf{P}'\,\mathbf{A}\,\mathbf{P}\,\mathbf{Z}, \\
&= \left(\mathbf{Z}_1' \quad \mathbf{Z}_2'\right)\begin{pmatrix} \mathbf{I}_d & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}\begin{pmatrix} \mathbf{Z}_1 \\ \mathbf{Z}_2 \end{pmatrix},
\end{aligned}$$

$$= \mathbf{Z}_1{}'\mathbf{Z}_1 \quad = \sum_{j=1}^{d}\left(\sum_{i=1}^{T} p_{ij}R_i\right)^2 ,$$

$$= \sum_{j=1}^{d}\left(\mathbf{p}_j^{'}\mathbf{R}\right)^2 , \quad \text{where } \mathbf{p}_j^{'}\mathbf{R} = \sum_{i=1}^{T} p_{ij}R_i \ . \tag{3.29}$$

Result 3.4 Let $\bar{v}_{\bullet j} = T^{-1}\mathbf{p}_j^{'}\mathbf{R}$, so $E(\bar{v}_{\bullet j}) = \frac{1}{T}\mathbf{p}_j^{'}(\mathbf{M}\mathbf{X}_2^*\boldsymbol{\beta})$ and $Var(\bar{v}_{\bullet j}) = T^{-1}\sigma^2\mathbf{M} < \infty$.

Proof Since $\mathbf{P} = (p_{ij})$ is orthogonal so that $v_{\bullet j}$'s are i.i.d., and by result 3.3, then

$$E(\bar{v}_{\bullet j}) = \frac{1}{T}E(\mathbf{p}_j^{'}\mathbf{R}) = \frac{1}{T}\mathbf{p}_j^{'}(\mathbf{M}\mathbf{X}_2^*\boldsymbol{\beta}) \text{ and}$$

$$Var(\bar{v}_{\bullet j}) = \frac{1}{T^2}Var(\mathbf{p}_j^{'}\mathbf{R}) = \frac{1}{T^2}\mathbf{p}_j^{'}(\sigma^2\mathbf{M})\mathbf{p}_j = T^{-2}\sigma^2\mathbf{M},$$

where $\sum_{i=1}^{T} p_{ij}^2 = 1$, from $\mathbf{P}_1{}'\mathbf{P}_1 = \mathbf{I}_d$. ■

Lemma 3.7 Let $Q_j = \frac{\bar{v}_{\bullet j} - E(\bar{v}_{\bullet j})}{\sqrt{Var(\bar{v}_{\bullet j})}}$, for $j = 1, 2, ..., d$, be i.i.d, and let Φ be the distribution function of $Z \sim N(0,1)$, then

$$Q_j \xrightarrow{L} \Phi(z) \ .$$

Proof Obviously. ■

Lemma 3.8 $\mathbf{AM}$ is idempotent where $\mathbf{A} = \mathbf{S'WS}$ and $\mathbf{M}$ be symmetric and positive definite.

Proof Since $\mathbf{A} = \mathbf{S'WS}$ is symmetric (see appendix B4), $\mathbf{M}$ is symmetric and positive definite (see appendix B4), it follows directly from lemma 9 of Searle (1971: 37), that $\mathbf{AM}$ is idempotent. ■

Theorem 3.2 Consider a multiple linear regression model with first-order autoregression model, $\mathbf{y}^* = \mathbf{X}_1^*\boldsymbol{\theta} + \mathbf{X}_2^*\boldsymbol{\beta} + \boldsymbol{\varepsilon}$; $\mathbf{y}^*$ and $\mathbf{X}^*$ are given in (3.4), where the errors $\boldsymbol{\varepsilon}$ are assumed to be i.i.d.$\left(0, \sigma^2\right)$, then, under the alternative hypothesis, $\mathrm{H}_1 : \boldsymbol{\beta} \neq \mathbf{0}$,

$$\mathrm{T}^{-1}\mathrm{F}_{\mathrm{P}} \xrightarrow{L} \chi^2_{\left(p-q,\, \lambda^*\right)}, \tag{3.30}$$

where the permuted statistic F_{P} is given in (3.16) and $\mathrm{p}-\mathrm{q}$ is the total number of partial regression coefficients to be tested, T is the sample size and noncentrality parameter $\lambda^* = \frac{1}{2\sigma^2}\left(\mathbf{MX}_2^*\boldsymbol{\beta}\right)' \mathbf{S'WS}\left(\mathbf{MX}_2^*\boldsymbol{\beta}\right)$. That is to say, $\mathrm{T}^{-1}\mathrm{F}_{\mathrm{P}}$ converges in distribution to a noncentral chi-squared distribution with $\mathrm{p}-\mathrm{q}$ degrees and a noncentrality parameter λ^*.

Proof

From lemma 3.7, we obtain

$$Q_j = \frac{\mathrm{T}^{-1}\mathbf{p}_j'\mathbf{R} - \mathrm{T}^{-1}\mathbf{p}_j'\left(\mathbf{MX}_2^*\boldsymbol{\beta}\right)}{\left(\mathrm{T}^{-2}\sigma^2\mathbf{M}\right)^{1/2}} \xrightarrow{\mathrm{L}} Z \sim N(0,1), \text{ for } j = 1, 2, \ldots, \mathrm{p}-\mathrm{q}. \tag{3.31}$$

In addition,

$$\mathbf{p}_j'\mathbf{R} \xrightarrow{L} Z_j \sim N\left(\mathbf{p}_j'\left(\mathbf{M}\mathbf{X}_2^*\boldsymbol{\beta}\right), \sigma^2\mathbf{M}\right). \tag{3.32}$$

This implies that,

$$\sum_{j=1}^{p-q}\left(\mathbf{p}_j'\mathbf{R}\right)^2 \xrightarrow{L} \sigma^2\mathbf{M}\sum_{j=1}^{p-q} Z_j^2 \sim \sigma^2\mathbf{M}\,\chi^2_{\left(p-q,\lambda^*\right)}, \tag{3.33}$$

where

$$\lambda^* = \frac{1}{2}\left(\mathbf{M}\mathbf{X}_2^*\boldsymbol{\beta}\right)'\mathbf{A}\left(\mathbf{M}\mathbf{X}_2^*\boldsymbol{\beta}\right).$$

Following directly from lemma 3.8, and using theorem 2 from Searle (1971: 57), $\mathbf{AM}$ is idempotent.

Hence, 3.29 and 3.33 implies that

$$\mathbf{R}'_{y^*}\big|_{x_1^*}\mathbf{S}'\,\mathbf{W}\,\mathbf{S}\,\mathbf{R}_{y^*}\big|_{x_1^*} \xrightarrow{L} \sigma^2\mathbf{M}\,\chi^2_{\left(p-q,\lambda^*\right)}. \tag{3.34}$$

Since from lemma 3.4(2), $T^{-1}\left(\mathbf{R}'_{y^*}\big|_{x_1^*}\mathbf{S}'\,\mathbf{H}\,\mathbf{S}\,\mathbf{R}_{y^*}\big|_{x_1^*}\right) \xrightarrow{P} \sigma^2\mathbf{M}$, and by using Slutsky's theorem, we get

$$\frac{\mathbf{R}'_{y^*}\big|_{x_1^*}\mathbf{S}'\,\mathbf{W}\,\mathbf{S}\,\mathbf{R}_{y^*}\big|_{x_1^*}}{\mathbf{R}'_{y^*}\big|_{x_1^*}\mathbf{S}'\,\mathbf{H}\,\mathbf{S}\,\mathbf{R}_{y^*}\big|_{x_1^*}} \xrightarrow{L} \chi^2_{\left(p-q,\lambda^*\right)}. \tag{3.35}$$

Hence, $T^{-1}F_P$ converges in distribution to a noncentral chi-squared distribution with p-q degrees of freedom and a noncentrality parameter λ^*, which completes the proof. ■

As a consequence of Theorem 3.1 and 3.2, an asymptotic chi-squared significance test for testing $H_0 : \boldsymbol{\beta} = \mathbf{0}$ can be performed. Therefore, for a test of size α, the null hypothesis is rejected if $F_P \geq T^{-1}\chi^2_{(\alpha, p-q)}$, where $\chi^2_{(\alpha, p-q)}$ is the value of the $(1-\alpha)100^{th}$ percentile of chi-squared with p-q degrees of freedom.

CHAPTER 4

SIMULATION STUDY

In this chapter, numerical simulations designed to explore the performance of the procedures proposed in chapter 3 for a multiple linear regression model with first-order autoregression which is transformed so that exchangeability property is hold. Consider the multiple linear regression model that is transformed from the model:

$$\mathbf{y}^* = \mathbf{X}_1^*\boldsymbol{\theta} + \mathbf{X}_2^*\boldsymbol{\beta} + \boldsymbol{\varepsilon} \ , \tag{4.1}$$

where

$$\mathbf{y}^* = \begin{bmatrix} (1-\rho^2)^{1/2} y_1 \\ y_2 - \rho y_1 \\ y_3 - \rho y_2 \\ \vdots \\ y_T - \rho y_{T-1} \end{bmatrix}, \mathbf{X}_1^* = \begin{bmatrix} (1-\rho^2)^{1/2} & (1-\rho^2)^{1/2} X_{11} & (1-\rho^2)^{1/2} X_{21} & (1-\rho^2)^{1/2} X_{31} \\ 1-\rho & X_{12} - \rho X_{11} & X_{22} - \rho X_{21} & X_{32} - \rho X_{31} \\ 1-\rho & X_{13} - \rho X_{12} & X_{23} - \rho X_{22} & X_{33} - \rho X_{32} \\ \vdots & \vdots & \vdots & \vdots \\ 1-\rho & X_{1T} - \rho X_{1,T-1} & X_{2T} - \rho X_{2,T-1} & X_{3t} - \rho X_{3,T-1} \end{bmatrix},$$

$$\mathbf{X}_2^* = \begin{bmatrix} (1-\rho^2)^{1/2} X_{41} & (1-\rho^2)^{1/2} X_{51} \\ X_{42} - \rho X_{41} & X_{52} - \rho X_{51} \\ X_{43} - \rho X_{42} & X_{53} - \rho X_{52} \\ \vdots & \vdots \\ X_{4T} - \rho X_{4,T-1} & X_{5T} - \rho X_{5,T-1} \end{bmatrix}, \quad \boldsymbol{\varepsilon} = \begin{bmatrix} (1-\rho^2)^{1/2} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \vdots \\ \varepsilon_T \end{bmatrix},$$

$\boldsymbol{\theta} = (\theta_0, \theta_1, \theta_2, \theta_3)$ and $\boldsymbol{\beta} = (\beta_1, \beta_2)$.

Note that $\boldsymbol{\varepsilon}$ is a vector of i.i.d. errors (see appendix A).

In testing the hypothesis

$$H_0 : \boldsymbol{\beta} = 0 \qquad \text{vs.} \qquad H_1 : \boldsymbol{\beta} \neq 0, \tag{4.2}$$

the methods compared are the proposed permutation method (PPT), the permutation method based on Freedman and Lane (1983: 292-298) (FPT) (see section 2.2.5) and the permutation method based on Manly (1991: 250-261) (MPT) (see section 2.2.2). The Partial F-test (PFT) and asymptotic chi-squared test (ACT) are also considered in the simulations.

4.1 Simulation Method

In the study, data were generated under the following choices of entities.

1. The sample size: $T = \{12, 16, 20\}$
2. The autocorrelation coefficient: $\rho = \{\pm 0.25, \pm 0.5, \pm 0.75\}$
3. The distribution of random errors: $(\boldsymbol{\varepsilon}) : \{\text{Normal}(0,1), \text{Uniform}(-2,2)\}$

The regression coefficient $\boldsymbol{\theta} = (\theta_0, \theta_1, \theta_2, \theta_3) = (1,1,1,1)$, meanwhile, $\boldsymbol{\beta} = (\beta_1, \beta_2)$ are set at $(0,0)$ and $(1,1)$ under the null and alternative hypotheses, repectively. Most economic time series generally exhibit positive autocorrelation because most of them either move upwards or downwards over extended time periods and do not exhibit a constant up-and-down movement (Gujarati, 2003: 449). However, the autocorrelation can be positive as well as negative so, in this simulation study, both positive and negative autocorrelation values are included.

Significance levels of $\alpha = 0.01$ and $\alpha = 0.05$ are considered and each of the five tests (PPT, FPT, MPT, ACT and PFT) is carried out on 10,000 data sets simulated with all possible combinations of simulated attribute distributions of $\boldsymbol{\varepsilon}$, regression coefficients ρ and sample sizes T.

The empirical type I error of each test in each situation is the rejection rate in 2,000 data sets under the null hypothesis and the empirical power of the test is the rejection rate in 2,000 data sets under the alternative hypothesis.

According to Vadiveloo (1983: 1581-1596), the number of permutations in each random permutation test at α level of significance ($\alpha = 0.01$ and $\alpha = 0.05$) is chosen to be q = 999 times.

Computer programs for simulations and calculations were written and compiled using MathLab (the Mathwork Release 14 with Service Pack1 product family).

Furthermore, in each situation, testing whether the empirical type I error $\hat{\rho}$ is statistically different assigned value of $\alpha = \rho_0$ and is done by using Z-test as follow. That is,

$$Z = \frac{\hat{p} - p}{\sqrt{p(1-p)/2000}} \tag{4.3}$$

is a statistic for testing $H_0 : p = p_0$ against $H_1 : p \neq p_0$, where Z is asymptotically normally distributed under H_0. Recall that the number of replication is 2,000. In particular, two values of $p_0 = \alpha$ were tested, i.e., $p_0 = 0.01$ and $p_0 = 0.05$.

4.2 Results of the Simulation Studies

4.2.1 Type I error

Figures 4.1 - 4.4 and tables F.1 - F.4 in appendix F present the empirical type I error rates obtained for the various permutation methods (PPT, FPT and MPT), as well as for ACT and PFT.

The simulations show that, for data generated under standard normal or uniform (-2, 2) errors, the MPT method of permutation results in inflated type I errors for all cases, i.e., type I errors are significantly far away from $\alpha = 0.01$ and $\alpha = 0.05$ in all situations. Hence, the graphs of MPT are not appeared in figures 4.1 - 4.4 but their values are shown in table F.1 – F.4. Although the errors are transformed to i.i.d and the assumption on exchangeability for a permutation test are valid, the MPT method, shuffing on $\mathbf{y}^*$, is obviously not suitable for autoregressive data.

At significance level $\alpha = 0.01$, for standard normal errors and a small sample size (T = 12), the results show that the PPT method had a smaller type I error rate than the ACT method for all value of ρ. In fact, the difference of the empirical type I error rate of PPT, FPT and PFT methods are significantly invisible for every value of ρ when T is small.

Conversely, as the sample size T increase, the empirical type I error rate of ACT method decrease and for T=16, 20 is smaller than the others three methods for all value of ρ but the different are not significant. Conclusions are similar when the distribution of error is uniform (-2, 2).

Additionally, for uniform (-2, 2) errors, the PPT, FPT and PFT methods had approximately the same empirical type I error rate. It seems that PPT's performance, in terms of empirical type I error, is more likely ACT's performance as T increases, though, the empirical type I error of all four methods are very slightly different. Findings are similar for uniform (-2, 2) errors where the difference are smaller in all situations. However, for increases in sample size T, the ACT mothod yields smaller type I error rate for all values of autocorrelation same as uniform (-2, 2) errors.

In the case of significance level $\alpha = 0.05$, for standard normal errors, it maybe reported that there are no significant differences among the PPT, FPT, ACT and PFT methods when the sample size is small. However, with uniform errors, the PPT method had type I errors more often lower than the others, especially as the autocorrelation coefficient approached ± 0.75.

Furthermore, the results of the simulation show that the type I error rates of all of the methods were higher when the autocorrelation coefficient approached ± 0.25.

Note that, since the MPT method of permutation result in inflated type I errors (see appendix F, tables F.1 - F.4) for all situations, the empirical type I error rates aren't shown in Fig. 4.1- 4.4

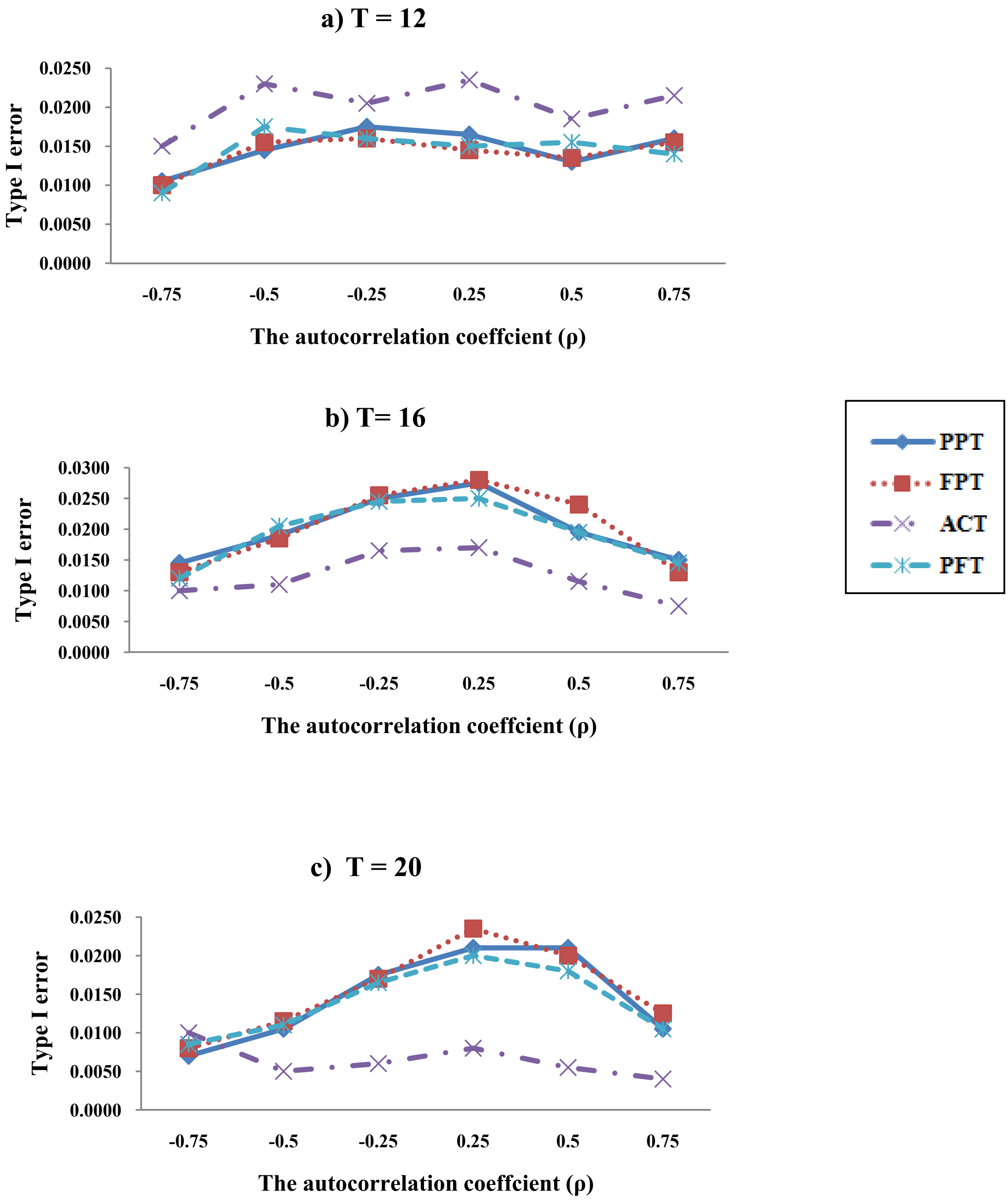


Figure 4.1 Empirical Type I Errors of the five Tests with increasing Autocorrelation (ρ) at $\alpha = 0.01$ for different Sample Sizes (T) where Errors are Standard Normal and $(\theta_0, \theta_1, \theta_2, \theta_3) = (1,1,1,1)$

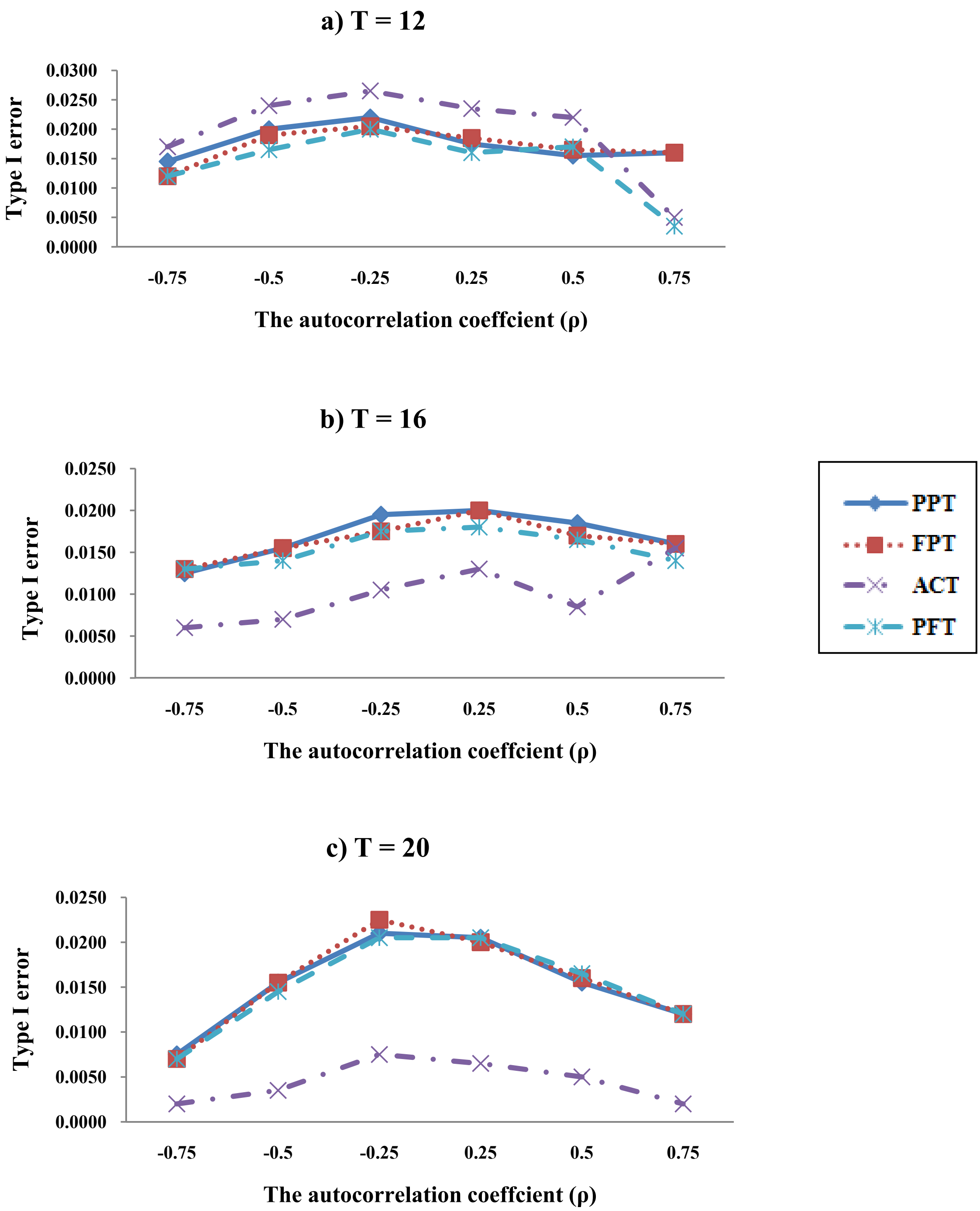


Figure 4.2 Empirical Type I Errors of the five Tests with increasing Autocorrelation (ρ) at $\alpha = 0.01$ for different Sample Sizes (T) where Errors are Uniform$(-2, 2)$ and $(\theta_0, \theta_1, \theta_2, \theta_3) = (1,1,1,1)$

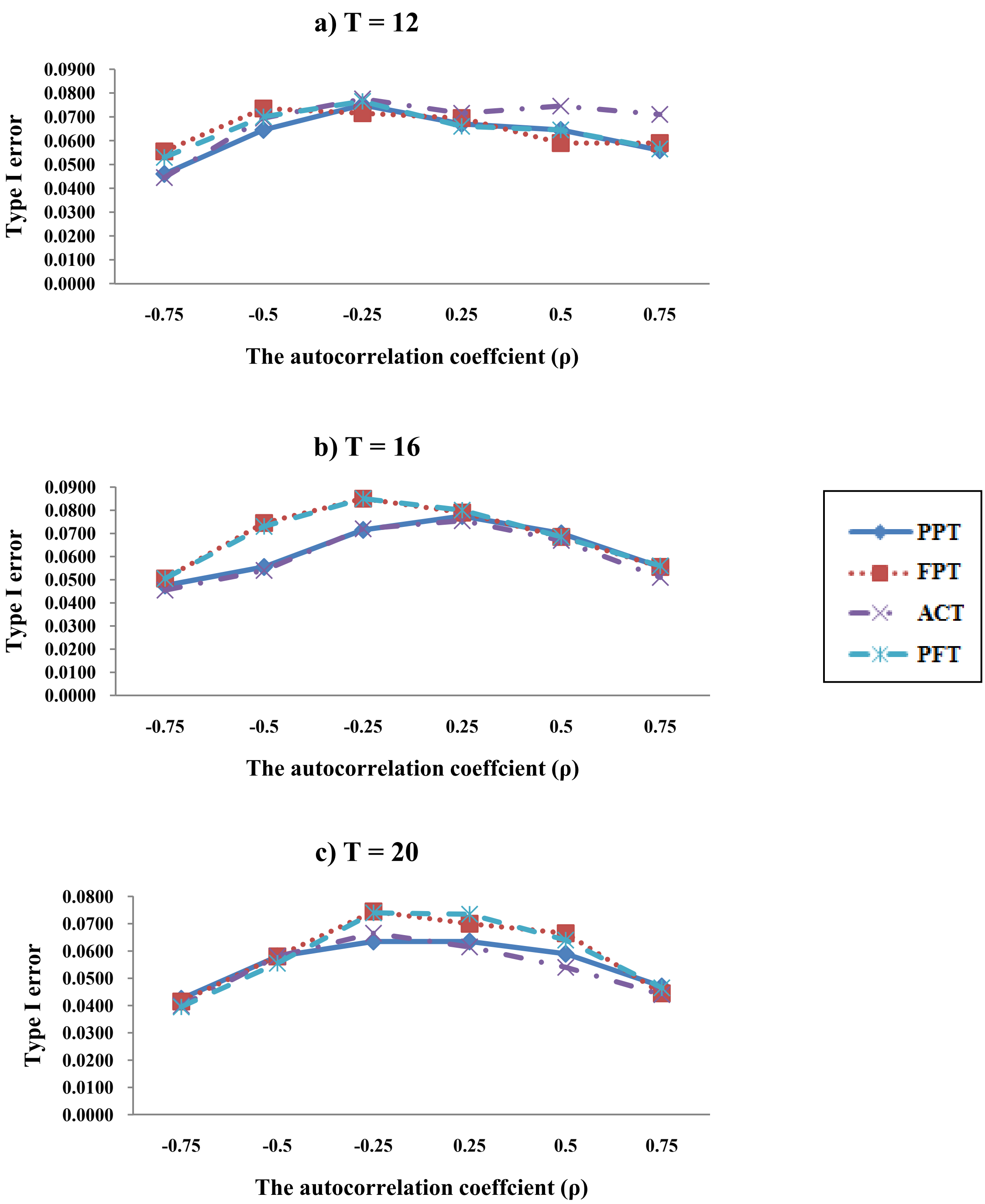


Figure 4.3 Empirical Type I Error of the five Tests with increasing Autocorrelation (ρ) at $\alpha = 0.05$ for different Sample Sizes (T) where Errors are Standard Normal and $(\theta_0, \theta_1, \theta_2, \theta_3) = (1,1,1,1)$

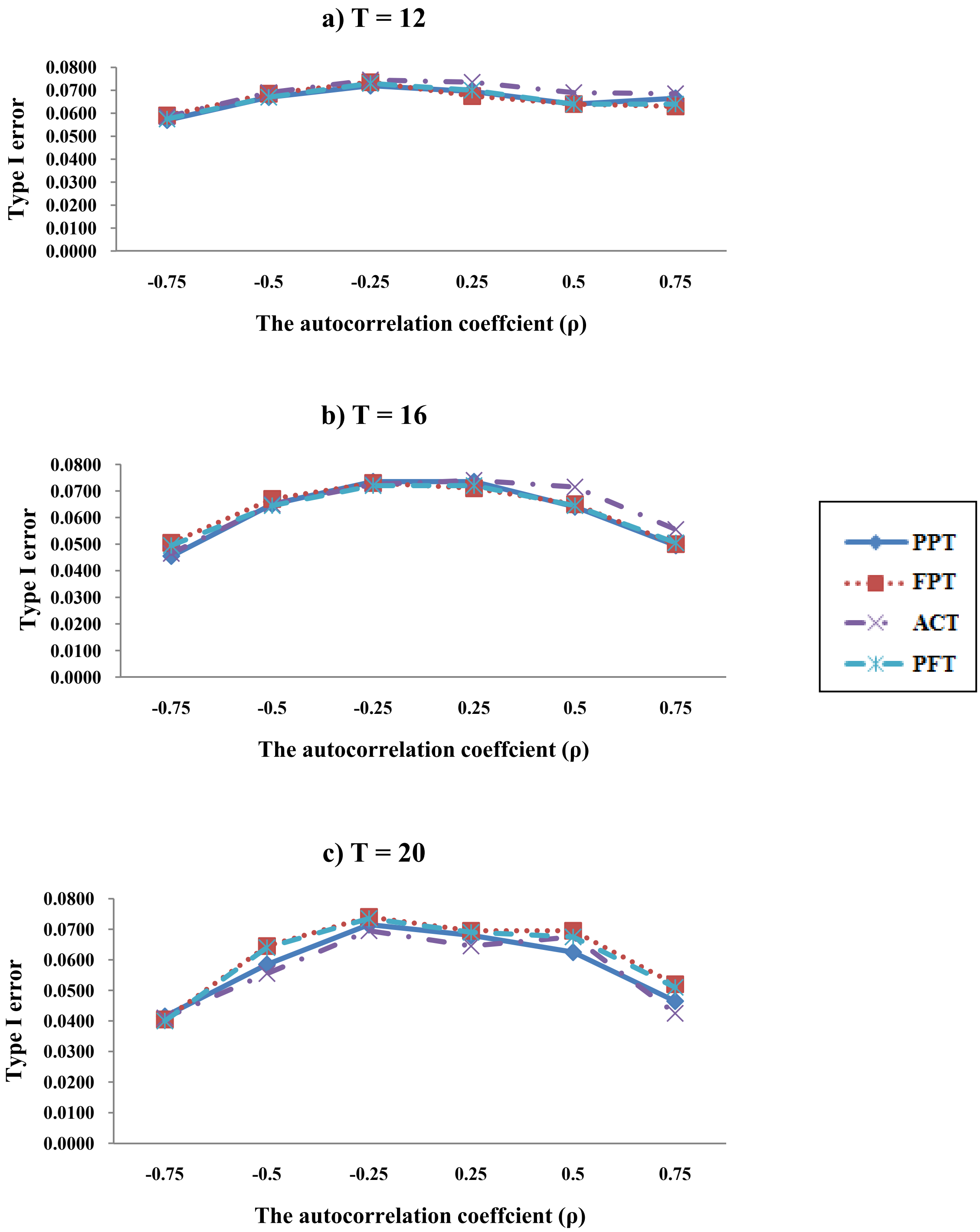


Figure 4.4 Empirical Type I Errors of the five Tests with increasing Autocorrelation (ρ) at $\alpha = 0.05$ for different Sample Sizes (T) where Errors are Uniform (-2,2) and $(\theta_0, \theta_1, \theta_2, \theta_3) = (1,1,1,1)$

4.2.2 Power of the Test

Figures 4.5 - 4.8 and tables F.5 – F.8 in appendix F present the powers of the test obtained from five permutation methods namely PPT, FPT, MPT, ACT and PFT. Again, the graph of the power of MPT method were not included in figures 4.5 4.8 since their values are too high to be fitted in the graphs due to the high values of empirical type I error rate, compare to the others methods.

As might be expected, for all methods (PPT, FPT, ACT and PFT), there is increase in power with an increase in sample size.

For data generated with normal or non normal errors, there are no significant differences in power among the permutation methods, including the Partial F test. However, for small sample size, the ACT method has a higher power than the others at significance level 0.01 for all situations. When the sample size increases, the ACT method had slightly smaller than the others due to the smaller value of the empirical type I error rate which and approach to the decrease value α.

When $\alpha = 0.05$, there are very slightly difference of power between methods of testing the hypothesis. Moreover, when T is large (T=20), the power of all testing methods approaches 1 for all situations.

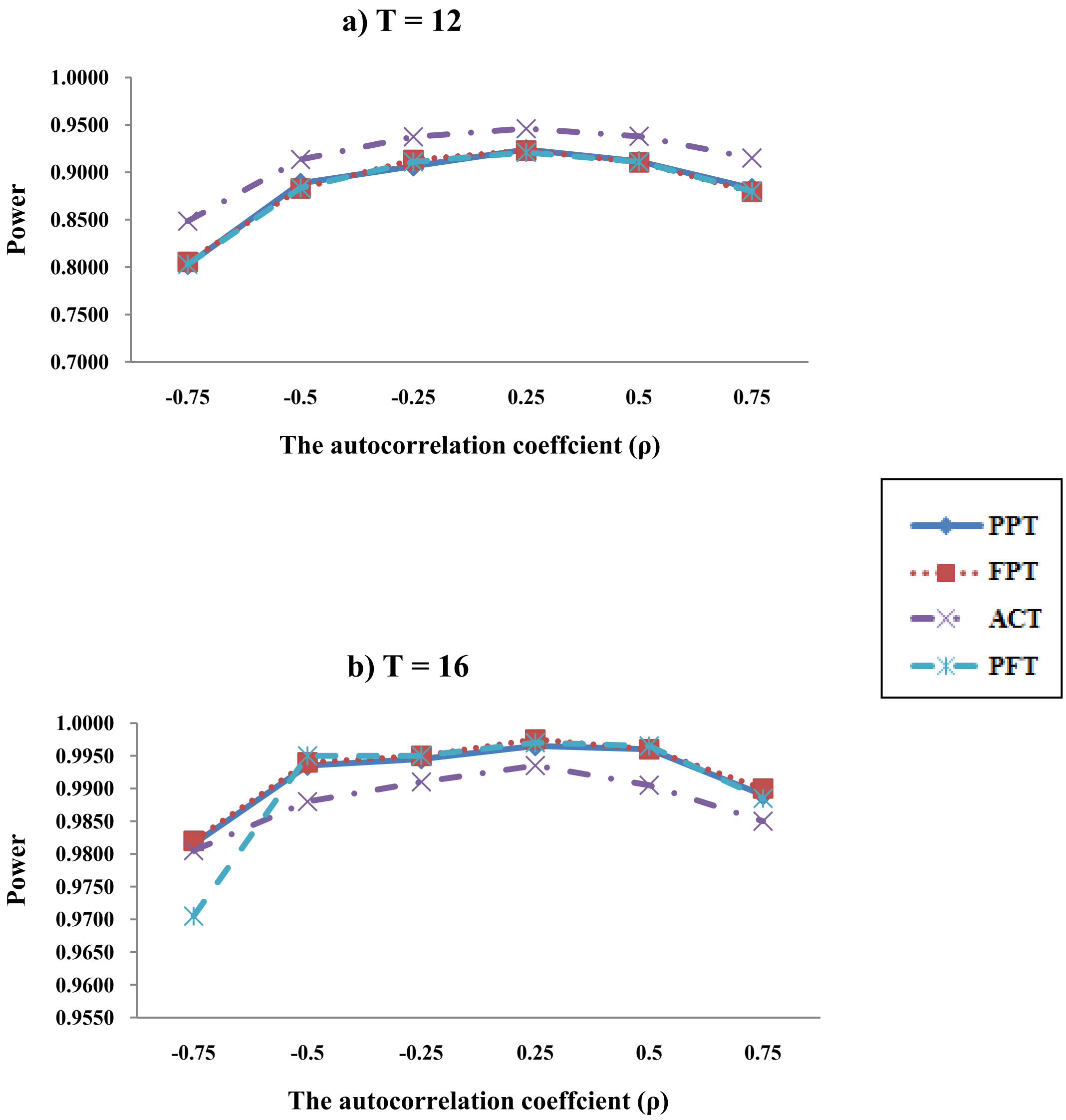


Figure 4.5 Empirical Power of the five Tests with increasing Autocorrelation (ρ) at $\alpha = 0.01$ for different Sample Sizes (T) where Errors are Standard Normal, $(\theta_0, \theta_1, \theta_2, \theta_3) = (1,1,1,1)$ and $(\beta_1, \beta_2,) = (1,1)$

Note: The empirical powers of all five Tests at T = 20 are larger than 0.99.

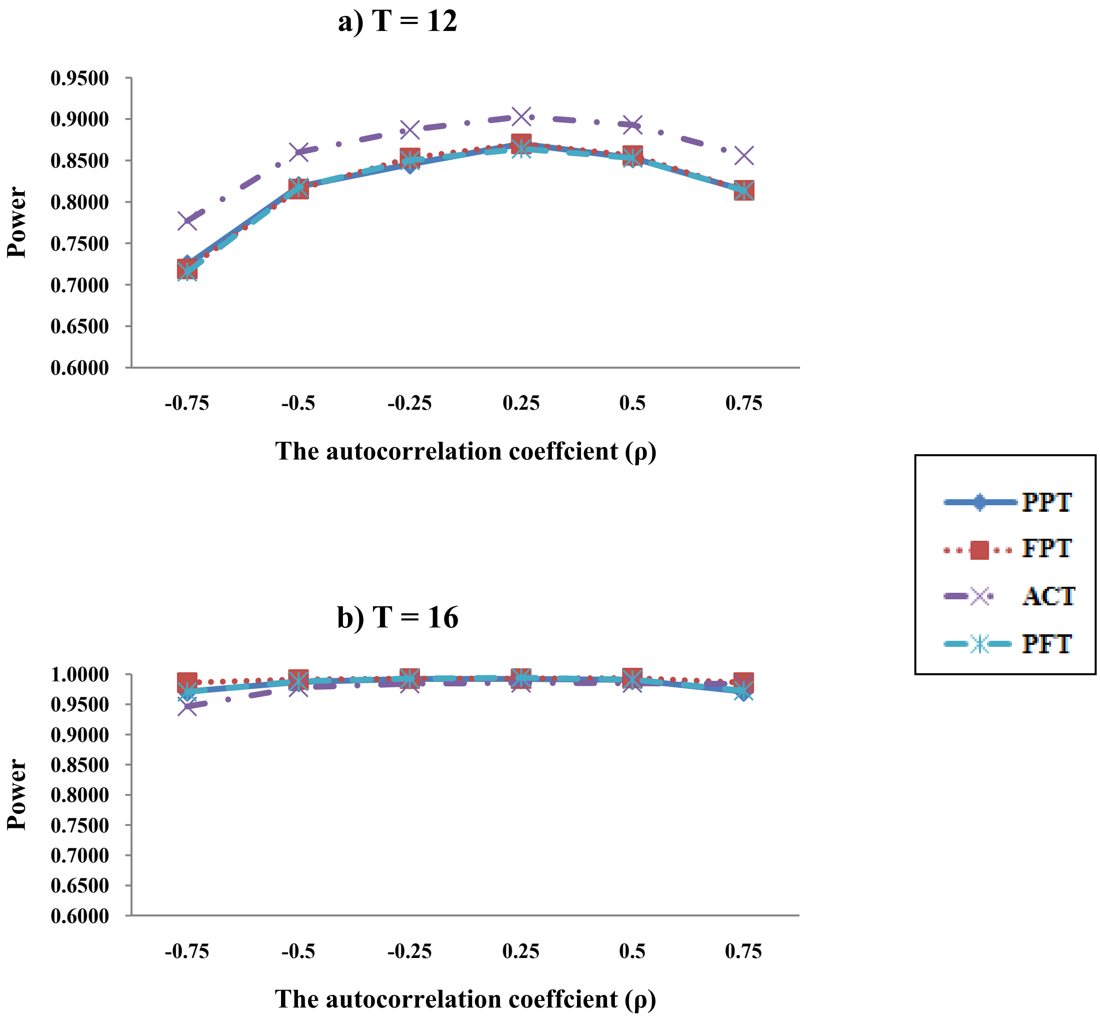


Figure 4.6 Empirical Powers of the five tests with increasing Autocorrelation (ρ) at $\alpha = 0.01$ for different Sample Sizes (T) where Errors are Uniform (-2,2), $(\theta_0, \theta_1, \theta_2, \theta_3) = (1,1,1,1)$ and $(\beta_1, \beta_2,) = (1,1)$

Note: The empirical powers of all five Tests at T = 20 are larger than 0.99.

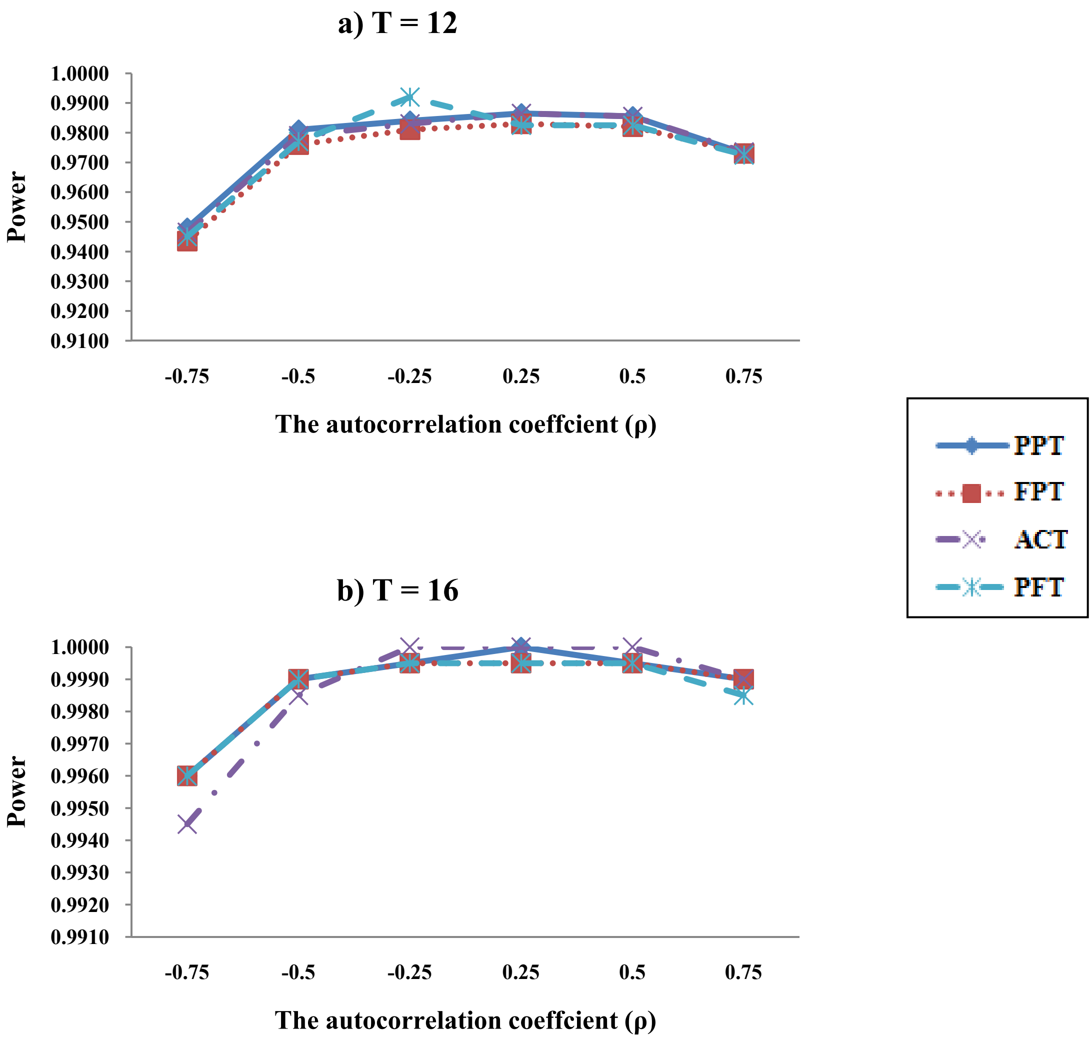


Figure 4.7 Empirical Power of the five tests with increasing Autocorrelation (ρ) at $\alpha = 0.05$ for different Sample Sizes (T) where Errors are Standard Normal, $(\theta_0, \theta_1, \theta_2, \theta_3) = (1,1,1,1)$ and $(\beta_1, \beta_2,) = (1,1)$

Note: The empirical powers of all five Tests at T = 20 are larger than 0.99.

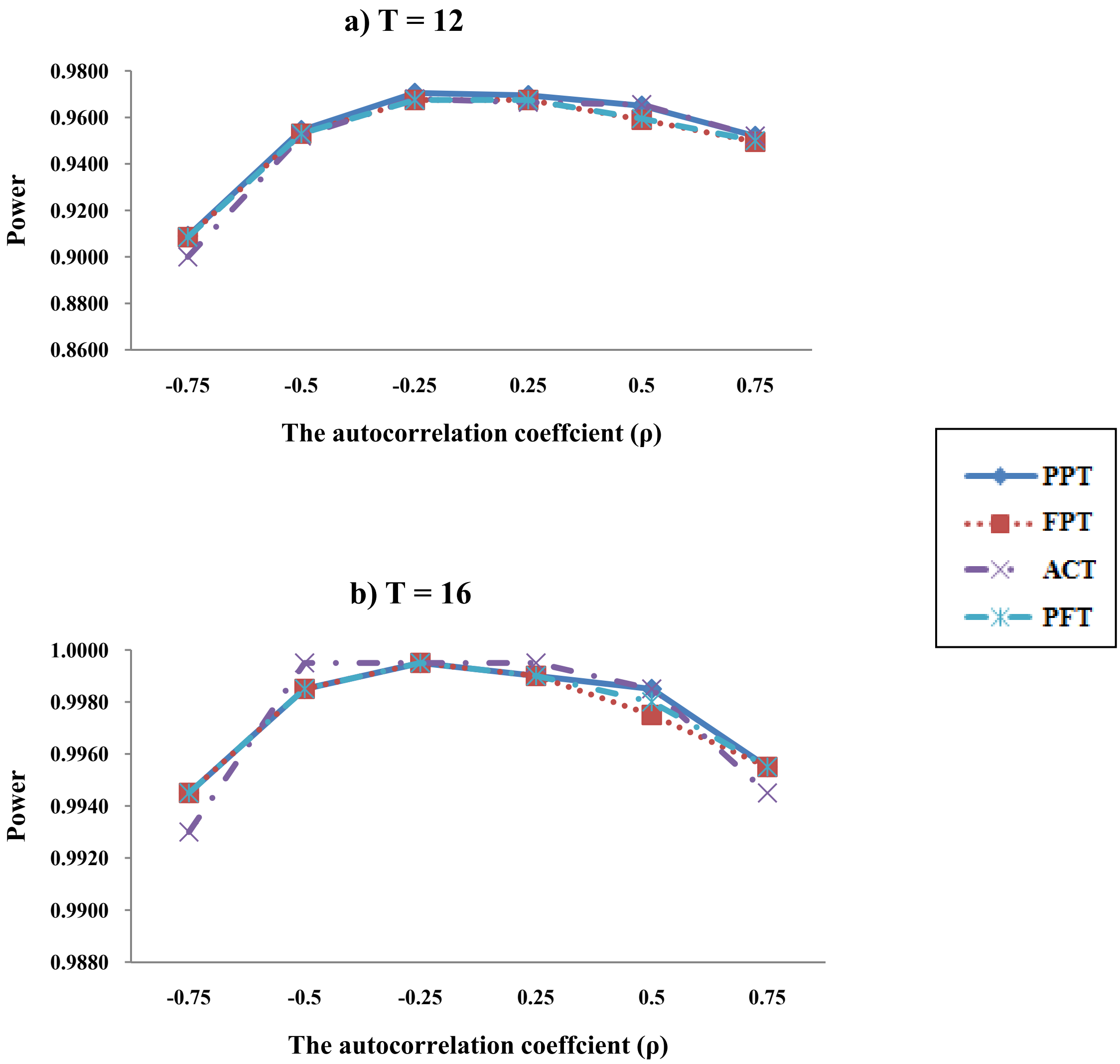


Figure 4.8 Empirical Power of the five tests with increasing Autocorrelation (ρ) at $\alpha = 0.05$ for different Sample Sizes (T) where Errors are Uniform (-2,2), $(\theta_0, \theta_1, \theta_2, \theta_3) = (1,1,1,1)$ and $(\beta_1, \beta_2,) = (1,1)$.

Note: The empirical powers of all five Tests at T = 20 are larger than 0.99.

CHAPTER 5

SUMMARY AND FUTURE RESEARCH

5.1 Summary

5.1.1 The Problem

In some situations, many regression applications involve time series data where the assumption of uncorrelated or independent error terms is not suitable, i.e., it is more likely that the errors are serially correlated. Consider a linear regression first-order autoregressive model as follows;

$$\mathbf{y} = \mathbf{X}\boldsymbol{\tau} + \mathbf{u}$$

$$u_t = \rho\, u_{t-1} + \varepsilon_t, \qquad \text{where } \varepsilon_t \text{ is a } T \times 1 \text{ vector of i.i.d.} \left(0, \sigma^2\right).$$

However, even if the error term is autocorrelated, the OLS parameter estimates will remain unbiased but the variance of the error term may be seriously underestimated.

Since the usual t and F tests are not generally reliable, an alternative to this traditional approach is to use a permutation test. Moreover, in practice, permutation procedures have increased steadily with computing power and it can now easily be employed in many situations without concerning for computing difficulties.

The aims of this dissertation are to propose a permutation test of partial regression coefficients for multiple linear regression with first-order autocorrelation where the distribution of the error terms are not necessary normal and the proposed permutation test statistic and others can be directly obtained without fitting back to the model, which is not the same as with previously reported permutation tests, e.g. Freedman and Lane, Manly, etc.

The asymptotic distribution of the proposed test is also considered. The type I error rate and power of the proposed method along with the other permutation methods are studied using numerical simulation and compared to the results of the classic parametric F-test.

5.1.2 Methodology

In this study, for a small sample size T, a permutation test statistic for testing partial regression coefficients for multiple linear regression with first-order autocorrelation without outliers is proposed. In order that the errors are exchangeable, the Prais-Winsten transformation (a procedure for determining errors free from serial correlation) is used, so that the model becomes $\mathbf{y}^* = \mathbf{X}^*\boldsymbol{\tau} + \boldsymbol{\varepsilon} = \mathbf{X}_1^*\boldsymbol{\theta} + \mathbf{X}_2^*\boldsymbol{\beta} + \boldsymbol{\varepsilon}$.

A proposed permutation test (PPT) based on a random permutation test is considered. The hypotheses of interest are $H_0 : \boldsymbol{\beta} = \mathbf{0}$ vs. $H_1 : \boldsymbol{\beta} \neq \mathbf{0}$. The properties of the proposed permutation test (PPT) that were investigated are asymptotic distribution under the null and alternative hypotheses. The level of significance and the power of PPT were compared to the permutation methods based on Freedman and Lane (1983: 292-298) (FPT) and Manly (1991: 250-261) (MPT), along with the Partial F-test (PFT) and the Asymptotic Chi-squared test (ACT). This comparison was carried out using a simulation study.

5.2 Results of the Study

5.2.1 The Proposed Permutation Test Statistic

The basic idea behind permutation methods is to generate a reference distribution by recalculating a statistic for many permutations of the data (Ernst, 2004: 676-685). Therefore, a permutation test was developed and proposed as follows:

$$F_{ref} = \frac{\hat{\boldsymbol{\beta}}'\left(\mathbf{X}_2^{*\prime}\,\mathbf{M}\,\mathbf{X}_2^*\right)\hat{\boldsymbol{\beta}}}{\mathbf{y}^{*\prime}\,\mathbf{H}\,\mathbf{y}^*}\,, \tag{5.1}$$

where $\mathbf{H} = \mathbf{I}_T - \mathbf{X}^*\left(\mathbf{X}^{*'}\mathbf{X}^*\right)^{-1}\mathbf{X}^{*'}$.

The idea of the proposed permutation procedure is, if the null hypothesis $H_0 : \boldsymbol{\beta} = \mathbf{0}$ is true, then $\mathbf{y}^* = \mathbf{X}_1^*\boldsymbol{\theta} + \boldsymbol{\varepsilon}$, where $\boldsymbol{\theta}$ is the simple regression coefficient of $\mathbf{y}^*$ versus $\mathbf{X}_1^*$. The residual under the reduced model is given by

$$\mathbf{R}_{\mathbf{y}^*|\mathbf{X}_1^*} = \mathbf{M}\mathbf{y}^*, \qquad \text{where } \mathbf{M} = \mathbf{I}_T - \mathbf{X}_1^*\left(\mathbf{X}_1^{*'}\mathbf{X}_1^*\right)^{-1}\mathbf{X}_1^{*'}. \tag{5.2}$$

Next, this residual is randomly permuted to produce a permuted vector $\mathbf{R}^{\pi}_{\mathbf{y}^*|\mathbf{X}_1^*}$ for $T!$ times. The vector of the permuted estimator of $\boldsymbol{\beta}$ is

$$\hat{\boldsymbol{\beta}}_{\pi(P)} = \left(\mathbf{X}_2^{*'}\mathbf{M}\mathbf{X}_2^*\right)^{-1}\left(\mathbf{M}\mathbf{X}_2^*\right)'\mathbf{R}^{\pi}_{\mathbf{y}^*|\mathbf{X}_1^*} \tag{5.3}$$

and the permuted statistic F_P is obtained by,

$$F_P = \frac{\hat{\boldsymbol{\beta}}'_{\pi(P)}\left(\mathbf{X}_2^{*'}\mathbf{M}\,\mathbf{X}_2^*\right)\hat{\boldsymbol{\beta}}_{\pi(P)}}{\mathbf{R}^{\pi\,'}_{\mathbf{y}^*|\mathbf{X}_1^*}\,\mathbf{H}\,\mathbf{R}^{\pi}_{\mathbf{y}^*|\mathbf{X}_1^*}}. \tag{5.4}$$

Therefore, n out of $T!$ values of F_P are randomly selected. For instant, set n = 999. The criteria for rejecting the hypothesis involves the propotion of the number of F_p's that is greater than or equal to F_{ref}. In other words, supposes n of F_p's and F_{ref} are ordered from largest to smallest, then rank them from 1 to (n+1) and F_{ref} is at k^{th} rank. The empirical p-value is equal to $\frac{k}{n+1}$. Hence, for a test of size α, the null hypothesis is rejected when $\frac{k}{n+1} \le \alpha$.

The rationale behind the permutation on the residual under the reduced model $\mathbf{R}_{\mathbf{y}^*|\mathbf{x}_1^*}$ can be found in the empirical studies of Anderson and Legendre (1999: 271-303), showing that the Freedman and Lane (1983: 292-298) method that permutated the residual under the model generally gives the best results (in terms of type I error or power), and the theoretical results of Anderson and Robinson (2001: 75-88) confirm this.

5.2.2 Asymptotic Distribution of the Proposed Permutation Test

Under the assumption that errors $\boldsymbol{\varepsilon}$ are i.i.d. $\left(0, \sigma^2\right)$, the asymptotic distribution of F_P converges in distribution as follows:

1. If the null hypothesis $H_0 : \boldsymbol{\beta} = \mathbf{0}$ is true, an asymptotic distribution of $T^{-1}F_P$ converges in distribution to $\chi^2_{(p-q)}$, where T is the sample size and $p-q$ is the total number of partial regression coefficients to be tested

2. If the null hypothesis $H_0 : \boldsymbol{\beta} = \mathbf{0}$ is false, an asymptotic distribution of $T^{-1}F_P$ converges in distribution to $\chi^2_{\left(p-q,\, \lambda^*\right)}$, where T is the sample size, $p-q$ is the total number of partial regression coefficients to be tested and $\lambda^* = \frac{1}{2\sigma^2}\left(\mathbf{M}\mathbf{X}_2^*\boldsymbol{\beta}\right)' \mathbf{S}'\mathbf{W}\mathbf{S}\left(\mathbf{M}\mathbf{X}_2^*\boldsymbol{\beta}\right)$ is a noncentrality parameter.

An alternative significance test of $H_0 : \boldsymbol{\beta} = \mathbf{0}$ is proposed based on the asymptotic distribution of F_P. Hence, for a test of size α, the null hypothesis is rejected if $F_P \geq T^{-1}\chi^2_{(\alpha,p-q)}$, where $\chi^2_{(\alpha,p-q)}$ is the value of the Chi-squared percentile with p-q degrees of freedom.

5.2.3 Results of the Simulation Study

A comparison of the five tests (i.e. the proposed permutation method (PPT), the permutation method based on Freedman and Lane (1983: 292-298) (FPT), the permutation method based on Manly (1991: 250-261) (MPT), the Partial F-test (PFT) and the Asymptotic Chi-squared test (ACT)) was made. Empirical type I errors and the power of the test were compared by using simulations with regard to four factors: the sample size (T), the autocorrelation coefficient (ρ), the size of the covariable's parameter $(\boldsymbol{\theta})$ and the distribution of random errors $(\boldsymbol{\varepsilon})$.

The results of the simulation study indicate that, for both data generated with standard normal and uniform errors distributions, the MPT method of permutation resulted in inflated type I errors for data from either distribution and virtually any combination of test factors. Therefore, it can be said that the MPT method is not suitable for autoregressive data, although the errors had been transformed to i.i.d following the assumption of exchangeability in a permutation test.

However, the proposed permutation method (PPT), with errors generated from both standard normal and uniform distributions, has the same type I error rates as the PFT method and performed better than the others (in terms of type I error rates) in cases where autocorrelation approached ± 1 with a small sample size (T=12). Additionally, for both error distributions, the PPT method has a slightly higher power than the others.

With moderate sample sizes (T=16, 20), the ACT method performed better than the others for both error distributions. Although the power of the ACT method was not significantly different from the others, the type I error rate is smaller.

5.3 Recommendations for Future Research

In this dissertation considers a multiple regression model with first-order autoregression. The distribution of random errors, $\boldsymbol{\varepsilon}$, are symmetrical distribution that is standard normal and uniform (-2, 2) , so it would have been interesting to study non- symmetrical distribution. Furthermore, transformation of the errors needs to be applied so that the error terms become i.i.d. and hence exchangeability holds and permutaitos on these terms are valid. Many researchers have proposed methods of transformation; the Prais-Winsten transformation was chosen for this study. Though, the autocorrelation coefficient ρ is not a parameter of the main interest, it is one of the key parameter that effects our result. Recall that $y_t^* = y_t - \rho y_{t-1}$, so ρ should be known or, if not, estimation on ρ is needed to be done. There are many authers studied on estimation of ρ, e.g. Cochrane-Orcutt method (Cochrane and Orcutt, 1949). In this study, ρ is fixed and, yet, we do not consider the efficient of ρ. So it would have been interesting to find a more suitable estimator of ρ before applying the permutation procedure.

BIBLIOGRAPHY

Anderson, M. J. 2001. Permutation Tests for Univariate or Multivariate Analysis of Variance and Regression. **Canadian Journal of Fisheries and Aquatic Science**. 58 (3): 626 – 639.

Anderson, M.J. and Legendre, P. 1999. An Empirical Comparison of Permutation Methods for Tests of Partial Regression Coefficients in a Linear Model. **Journal of Statistical Computation and Simulation.** 62: 271– 303.

Anderson, M.J. and Robinson, J. 2001. Permutation Tests for Linear Models. **Australian New Zealand Journal of Statistics.** 43 (1): 75–88.

Brown, B. M. and Maritz, J. S. 1982. Distribution Free Methods in Regression. **The Australian Journal of Statistics.** 24 (3): 318-331.

Cade, B. S. 2005. Linear Models: Permutation Methods. In **Encyclopedia of Statistics in Behavioral Science**. Vol. 2. Chichester: Wiley. Pp. 1049-1054.

Casella, G. and Berger, R. L. 2002. **Statistical Inference**. 2nd ed. Australia: Thomson Learning.

Cochrane, D. and Orcutt, G. H. 1949. Application of Least Squares Regression to Relationships Containing Autocorrelated Error Terms. **Journal of the American Statistical Association**. 44: 32–61.

Edgington, E. S. 1995. **Randomization Tests**. 3rd ed. New York: Marcel Dekker.

Ernst, M. D. 2004. Permutation Methods: A Basis for Exact Inference. **Statistical Science.** 19 (4): 676-685.

Fisher, R. A. 1935. **Design of Experiments.** Edinburgh: Oliver and Boyd.

Freedman, D. and Lane, D. 1983. A Nonstochastic Interpretation of Reported Significance Levels. **Journal of Business and Economic Statistics**. 1 (4): 292–298.

Gail, M. H.; Tan, W. Y. and Piantadosi, S. 1988. Tests for No Treatment Effect in Randomized Clinical Trials. **Biometrika**. 75 (1): 57-64.

Good, P.I. 1999. **Resampling Methods A Practical Guide to Data Analysis**. Boston: Birkhäuser. Good, P.I. 2002. Extensions of the Concept of Exchangeability and Their Applications. **Journal of Modern Applied Statistical Methods.** 1 (2): 243-247.

Gujarati, D. 2003. **Basic Econometrics**. 4th ed. Boston: McGraw-Hill.

Harville, D. A. 1997. **Matrix Algebra From a Statistician's Perspective**. New York: Springer-Verlag.

Hinkelman, K. and Kempthorne, O. 1994. **Design and Analysis of Experiments.** New York: Wiley.

Hoeffding, W. 1952. The Large-sample Power of Tests Based on Permutations of the Observations. **Annals of Mathematical Statistics.** 23 (2) : 169–192.

Huh, M. H. and Jhun, M. 2001. Random Permutation Testing in Multiple Linear Regression. **Communications in Statistics - Theory and Methods**. 30 (10): 2023–2032.

Jaggia, S and Kelly-Hawke, A. 2005. Practical Considerations when Estimating in the Presence of Autocorrelation. Retrieved May, 15, 2010 from http://papers.ssrn.com/sol3/papers.cfm?abstract_id=805525.

Jung, B. C.; Jhun, M. and Song, S. H. 2006. A New Random Permutation Test in ANOVA Models. **Statistical Papers.** 48 : 47 - 62.

Kempthorne, O. 1955. The Randomization Theory of Experimental Inference. **Journal of the American Statistical Association**. 50 (271): 946–967.

Kempthorne, O. and Doerfler, T. E. 1969. The Behaviour of Some Significance Tests under Experimental Randomization. **Biometrika.** 56 (2): 231–248.

Kennedy, P.E. 1995. Randomization Tests in Econometrics. **Journal of Business and Economic Statistics.** 25 (1): 85-94.

Kennedy, P.E. and Cade, B. S. 1996. Randomization Tests for Multiple Regression. **Journal of Statistical Computation and Simulation.** 25: 923-936.

Kutner, M. H.; Nachtsheim, C. J.; Neter, J. and Li, W. 2005. **Applied Linear Statistical Models.** 5th ed. Chicago,IL: McGraw-Hill/Irwin.

Lafleur, B. 1999. **Application of Permutation Methods to the Generalized Linear Model.** Doctoral dissertation, University of Colorado.

Lehmann, E. L. 1999. **Elements of Large Sample Theory**. New York: Springer Verlag.

Levin, B. and Robbins, H. 1983. Urn Models for Regression Analysis, with Applications to Employment Discrimination Studies. **Statistical Inference in Litigation.** 46 (4): 247-267.

Manly, B.F.J. 1991. **Randomization, and Monte Carlo Methods in Biology.** 1st ed. London: Chapman and Hall.

Manly, B.F.J. 1997. **Randomization, Bootstrap and Monte Carlo Methods in Biology.** 2nd ed. London: Chapman and Hall.

Manly, B.F.J. 2007. **Randomization, Bootstrap and Monte Carlo Methods in Biology.** London: Chapman and Hall.

Oja, H. 1987. On Permutation Test in Multiple Regression and Analysis of Covariance Problems. **Australian Journal of Statistics**. 29 (1): 91-100.

Pitman, E. J. G. 1937a. Significance Tests Which May Be Applied to Samples from Any Populations. **Journal of the Royal Statistical Society, Series B**. 4 (1): 119–130.

Pitman, E. J. G. 1937b. Significance Tests Which May Be Applied to Samples from Any Populations II. The Correlation Coefficient Test. **Journal of the Royal Statistical Society, Series B**. 4 (2): 225–232.

Pitman, E. J. G. 1937c. Significance Tests Which May Be Applied to Samples from Any Populations III. The Analysis of Variance Test. **Biometrika.** 29: 322-325.

Ptitsyn, A. A; Zvonic, S. and Gimble, J. M. 2006. Permutation Test for Periodicity in Short Time Series Data. **BMC Bioinformatics.** 7: 1-9.

Randles, R. H. and Wolfe, D. A. 1979. **Introduction to the Theory of Nonparametric Statistics.** New York: Wiley.

Ravishanker, N. and Dey, D. K. 2002. **A First Course in Linear Model Theory**. New York: Chapman and Hall/CRC.

Rencher, A. C. and Schaalje, G. B. 2007. **Linear Models in Statistics**. 2nd ed. New Jersey: Wiley.

Schmoyer, R. L. 1994. Permutation Tests for Correlation in Regression Errors. **Journal of the American Statistical Association.** 89 (428): 1507-1516.

Searle, S. R. 1971. **Linear Models.** New York: Wiley.

Seber, G. A. F. 1977. **Linear Regression Analysis.** Toronto: Wiler.

Tantawanich, S. 2005**. Permutation Test of Partial Regression Coefficients**. Doctoral dissertation, National Institute of Development Administration.

ter Braak, C.J.F. 1992. Permutation Versus Bootstrap Significance Tests in Multiple Regression and ANOVA. In **Bootstrapping and Related Techniques.** K.H. Jockel, G. Rothe and W. Sendler, eds. New York: Springer-Verlag. Pp. 79-85.

Vadeviloo, J. 1983. On the Theory of Modified Randomization Tests for Nonparametric Hypotheses. **Communications in Statistics - Theory and Methods.** 12 (14): 1581-1596.

Welch, W. J. 1990. Construction of Permutation Tests. **Journal of the American Statistical Association.** 85 (411): 693-698.

Zocchi, S. and Manly, B. F. J. 2006. Generating Different Data Sets for Linear Regression Models with the Same estimates. Retrieved from February, 22, 2010 from http://www.ime.usp.br/~abe/ICOTS7/ Proceedings /PDFs/ContributedPapers/C108.pdf.

APPENDICES

APPENDIX A

PROPERTIES OF THE ERROR TERMS AND TRANSFORMATION OF ERRORS TO I.I.D.

(A1) Consider errors with first-order autocorrelation,

$$u_t = \rho\, u_{t\text{-}1} + \varepsilon_t\,, \tag{A1}$$

where ρ is the autocorrelation coefficient such that $|\rho| < 1$, and

ε_t are the white noise error terms which are assumed to have the following properties:

1. $E(\varepsilon_t) = 0$, $\forall t$ (zero mean assumption)
2. $E(\varepsilon_t^2) = \sigma^2$, $\forall t$ (constant variance assumption)
3. $E(\varepsilon_s \varepsilon_t) = 0$, $\forall s \neq t$ (uncorrelated of errors assumption)
4. ε_t are i.i.d

By repeated substitution, we get

$$u_t = \varepsilon_t + \rho\varepsilon_{t-1} + \rho^2\varepsilon_{t-2} + \ldots \quad = \sum_{r=0}^{\infty} \rho^r \varepsilon_{t-r}\,. \tag{A2}$$

The expectation and variance of u_t are respectively,

$$E[u_t] = 0\,. \tag{A3}$$

$$\mathrm{Var}[u_t] = \sum_{r=0}^{\infty} \rho^{2r} \mathrm{Var}[\varepsilon_{t-r}] = \sum_{r=0}^{\infty} \rho^{2r}\sigma^2 = \frac{\sigma^2}{1-\rho^2} = \sigma_u^2 \text{ , say.} \tag{A4}$$

The covariance of $u_t, u_{t\text{-}1}$ is

$$\mathrm{Cov}(u_t, u_{t-1}) = E[u_t u_{t-1}] \quad = E[u_{t-1}(\rho u_{t-1} + \varepsilon_t)] \quad = \rho V[u_{t-1}] \quad = \rho\left(\frac{\sigma^2}{1-\rho^2}\right). \tag{A5}$$

In the same manner, we get

$$E[u_t u_{t+s}] = \begin{cases} \dfrac{\rho^s \sigma^2}{1-\rho^2}, & s \neq t \\ \dfrac{\sigma^2}{1-\rho^2}, & s = t \end{cases} \qquad \text{where } s, t = 1, 2, \ldots$$

Therefore, we can represent this as

$$\mathrm{Var}[u] = \frac{\sigma^2}{1-\rho^2}\begin{bmatrix} 1 & \rho & \rho^2 & \rho^3 & \cdots & \rho^{T-1} \\ \rho & 1 & \rho & \rho^2 & \cdots & \rho^{T-2} \\ \rho^2 & \rho & 1 & \rho & \cdots & \rho^{T-3} \\ \vdots & \vdots & \vdots & \vdots & & \vdots \\ \rho^{T-1} & \rho^{T-2} & \rho^{T-3} & \rho^{T-4} & \cdots & 1 \end{bmatrix}_{T\times T} , \qquad \text{(A6)}$$

$$= \sigma^2\Phi , \qquad \text{(A7)}$$

where matrix $\Phi = \Phi(\rho) = \left(\dfrac{1}{1-\rho^2}\right)V$ and $V = \left(\rho^{|i-j|}\right)_{T\times T}$. (A8)

Inaddition the inverse of Φ is given by

$$\Phi^{-1} = \begin{bmatrix} 1 & -\rho & 0 & \cdots & 0 & 0 \\ -\rho & 1+\rho^2 & -\rho & \cdots & 0 & 0 \\ 0 & -\rho & 1+\rho^2 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1+\rho^2 & -\rho \\ 0 & 0 & 0 & \cdots & -\rho & 1 \end{bmatrix}_{T\times T} . \qquad \text{(A9)}$$

(A.2) Consider the multiple linear regression with first-order autocorrelation

$$y_t = \beta_0 + \beta_1 X_{1t} + \ldots + \beta_p X_{pt} + u_t, \quad \text{(A10)}$$

$$u_t = \rho u_{t-1} + \varepsilon_t \quad \text{(A11)}$$

for $t = 1, 2, \ldots, T$.

Substitute u_t in (A11) into equation (A10), we have

$$y_t = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + \ldots + \beta_p X_{pt} + \rho u_{t-1} + \varepsilon_t. \quad \text{(A12)}$$

At time $t-1$, multiply equation (A12) by ρ to get

$$\rho y_{t\text{-}1} = \rho\beta_0 + \rho\beta_1 X_{1,t\text{-}1} + \ldots + \rho\beta_p X_{p,t\text{-}1} + \rho^2 u_{t-2} + \rho\varepsilon_{t\text{-}1}. \quad \text{(A13)}$$

Subtract equation (A13) from (A12) and obtain the following;

$$\begin{aligned} y - \rho y_{t\text{-}1} &= \beta_0(1-\rho) + \beta_1(X_{1t} - \rho X_{1,\,t\text{-}1}) + \ldots + \beta_p(X_{pt} - \rho X_{p,\,t\text{-}1}) \\ &\quad + \rho u_{t-1} + \varepsilon_t - \rho(\rho u_{t-2} + \varepsilon_{t\text{-}1}), \\ &= \beta_0(1-\rho) + \beta_1(X_{1t} - \rho X_{1,\,t\text{-}1}) + \ldots + \beta_p(X_{pt} - \rho X_{p,\,t\text{-}1}) + u_t - \rho u_{t-1}, \\ &= \beta_0(1-\rho) + \beta_1(X_{1t} - \rho X_{1,\,t\text{-}1}) + \ldots + \beta_p(X_{pt} - \rho X_{p,\,t\text{-}1}) + \varepsilon_t, \end{aligned}$$

for $t = 2,\ 3,\ \ldots,\ T$.

Thus, it can be written in a matrix form as

$$\begin{bmatrix} y_2 - \rho y_1 \\ y_3 - \rho y_2 \\ \vdots \\ y_T - \rho y_{T-1} \end{bmatrix} = \begin{bmatrix} 1-\rho & X_{12} - \rho X_{11} & \cdots & X_{p2} - \rho X_{p1} \\ 1-\rho & X_{13} - \rho X_{12} & \cdots & X_{p3} - \rho X_{p2} \\ \vdots & \vdots & \ddots & \vdots \\ 1-\rho & X_{1T} - \rho X_{1,\,T-1} & \cdots & X_{pT} - \rho X_{p,\,T-1} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_T \end{bmatrix}, \quad \text{(A14)}$$

or

$$\tilde{\mathbf{y}} = \tilde{\mathbf{X}}\ \boldsymbol{\beta} + \boldsymbol{\varepsilon}, \tag{A15}$$

where the element of $\tilde{\mathbf{y}}$ on (t-1)th row is $\tilde{y} = (y_t - \rho y_{t\text{-}1})$, all element of $\tilde{\mathbf{X}}$ on the 1st column are $1-\rho$ and (i-1)th column $(i+1, t)^{th}$ is $\tilde{X}_{it} = X_{it} - \rho X_{i,\,t\text{-}1}$ for $i = 3, 4, \ldots, p$ and $t = 2, 3, \ldots, T$.

APPENDIX B

QUADRATIC FORM AND ITS PROPERTIES

Here, the properties of some quadratic forms and matrices of quadratic forms are proven.

Given that $\mathbf{X}^*$ is a $\mathrm{T} \times \mathrm{p}$ constant matrix of rank q for which $\mathbf{X}^* = (\mathbf{X}_1^*, \mathbf{X}_2^*)$, where $\mathbf{X}_1^*$ is a $\mathrm{T} \times \mathrm{q}$ matrix of rank q and $\mathbf{X}_2^*$ is an $\mathrm{T} \times (\mathrm{p}-\mathrm{q})$ matrix of rank $\mathrm{p}-\mathrm{q}$, then

(B1) Let $\mathbf{H}_{\mathbf{X}^*} = \mathbf{X}^*\left(\mathbf{X}^{*\prime}\mathbf{X}^*\right)^{-1}\mathbf{X}^{*\prime} = \{h_{ij}\}$ be a $\mathrm{T} \times \mathrm{T}$ matrix. Then

1. $\mathbf{H}_{\mathbf{X}^*}$ is symmetric and idempotent
2. rank $\mathbf{H}_{\mathbf{X}^*} = tr(\mathbf{H}_{\mathbf{X}^*}) = \mathrm{p}$
3. $\mathbf{H}_{\mathbf{X}^*}\mathbf{X}_1^* = \mathbf{X}_1^*$ and $\mathbf{H}_{\mathbf{X}^*}\mathbf{X}_2^* = \mathbf{X}_2^*$
4. $\mathbf{H}_{\mathbf{X}^*}$ is positive definite
5. $1/\mathrm{T} \le h_{ii} \le 1$, for $i = 1, 2, \ldots, \mathrm{T}$, and $-.5 \le h_{ij} \le .5$, for all $j \ne i$

Proof

1. Verification is straightforward as follow

$$\mathbf{H}'_{\mathbf{X}^*} = \left[\mathbf{X}^*\left(\mathbf{X}^{*\prime}\mathbf{X}^*\right)^{-1}\mathbf{X}^{*\prime}\right]' = \mathbf{X}^*\left(\mathbf{X}^{*\prime}\mathbf{X}^*\right)^{-1}\mathbf{X}^{*\prime} = \mathbf{H}_{\mathbf{X}^*} \; ;$$

$$\mathbf{H}_{\mathbf{X}^*}\mathbf{H}_{\mathbf{X}^*} = \mathbf{X}^*\left(\mathbf{X}^{*\prime}\mathbf{X}^*\right)^{-1}\mathbf{X}^{*\prime}\mathbf{X}^*\left(\mathbf{X}^{*\prime}\mathbf{X}^*\right)^{-1}\mathbf{X}^{*\prime} = \mathbf{H}_{\mathbf{X}^*} \, .$$ ■

2. Since $\mathbf{H}_{\mathbf{X}^*}$ is symmetric and idempotent, then

$$\begin{aligned}
\text{rank } \mathbf{H}_{\mathbf{X}^*} &= trace(\mathbf{H}_{\mathbf{X}^*}) \\
&= tr\left[\mathbf{X}^*\left(\mathbf{X}^{*\prime}\mathbf{X}^*\right)^{-1}\mathbf{X}^{*\prime}\right] = tr\left[\mathbf{X}^{*\prime}\mathbf{X}^*\left(\mathbf{X}^{*\prime}\mathbf{X}^*\right)^{-1}\right], \\
&= tr\left[\mathbf{I}_{\mathrm{p}}\right] = \mathrm{p}\,.
\end{aligned}$$

Note: Use Result 2.3.9 from Ravishanker and Dey (2002: 50). ■

3. Recall that $\mathbf{H}_{\mathbf{X}^*} = \mathbf{X}^*\left(\mathbf{X}^{*\prime}\mathbf{X}^*\right)^{-1}\mathbf{X}^{*\prime}$ and $\mathbf{X}^* = (\mathbf{X}_1^*, \mathbf{X}_2^*)$

Therefore,

$$\begin{aligned}\mathbf{H}_{\mathbf{X}^*}\mathbf{X}^* &= \mathbf{H}_{\mathbf{X}^*}\left(\mathbf{X}_1^*, \mathbf{X}_2^*\right)\\ &= \left(\mathbf{H}_{\mathbf{X}^*}\mathbf{X}_1^*, \mathbf{H}_{\mathbf{X}^*}\mathbf{X}_2^*\right)\end{aligned}$$

and
$$\begin{aligned}\mathbf{H}_{\mathbf{X}^*}\mathbf{X}^* &= \mathbf{X}^*\left(\mathbf{X}^{*\prime}\mathbf{X}^*\right)^{-1}\mathbf{X}^{*\prime}\mathbf{X}^*\\ &= \mathbf{X}^*.\end{aligned}$$

Hence $\mathbf{H}_{\mathbf{X}^*}\mathbf{X}_1^* = \mathbf{X}_1^*$ and $\mathbf{H}_{\mathbf{X}^*}\mathbf{X}_2^* = \mathbf{X}_2^*$. ■

4. It can be provent that $\mathbf{v}'\mathbf{H}_{\mathbf{X}^*}\mathbf{v} > 0$ for all $\mathbf{v} \in \mathcal{R}^T$ except $\mathbf{v} = 0$.

Suppose $\mathbf{H}_{\mathbf{X}^*} = \mathbf{P}'\mathbf{P}$ for a nonsingular $\mathbf{P}$, then

$$\mathbf{v}'\mathbf{H}_{\mathbf{X}^*}\mathbf{v} = \mathbf{v}'\mathbf{P}'\mathbf{P}\mathbf{v} = (\mathbf{P}\mathbf{v})'(\mathbf{P}\mathbf{v}),$$

which is the sum of squares of the elements of $\mathbf{P}\mathbf{v}$. Hence, $\mathbf{v}'\mathbf{H}_{\mathbf{X}^*}\mathbf{v} > 0$ for all $\mathbf{P}\mathbf{v} \neq 0$, and $\mathbf{v}'\mathbf{H}_{\mathbf{X}^*}\mathbf{v} = 0$ for all $\mathbf{P}\mathbf{v} = 0$. However, $\mathbf{P}\mathbf{v} = 0$ only when $\mathbf{v} = 0$ because $\mathbf{P}^{-1}$ exists. Since $\mathbf{v}'\mathbf{H}_{\mathbf{X}^*}\mathbf{v} > 0$ for all $\mathbf{v} \neq 0$, and $\mathbf{v}'\mathbf{H}_{\mathbf{X}^*}\mathbf{v} = 0$ only for $\mathbf{v} = 0$, then $\mathbf{H}_{\mathbf{X}^*}$ is positive definite. ■

5. The proof follows directly from theorem 9.2(i) (Rencher and Schaalje: 2007: 231). ■

(B2) Let $\mathbf{H} = \mathbf{I}_{\mathrm{T}} - \mathbf{X}^{*}\left(\mathbf{X}^{*\prime}\mathbf{X}^{*}\right)^{-1}\mathbf{X}^{*\prime} = \mathbf{I}_{\mathrm{T}} - \mathbf{H}_{\mathbf{X}^{*}}$ be a $\mathrm{T} \times \mathrm{T}$ matrix, then:

1. $\mathbf{H}$ is symmetric and idempotent
2. Rank $\mathbf{H} = tr(\mathbf{H}) = \mathrm{T} - \mathrm{p}$
3. $\mathbf{HX}^{*} = 0$
4. $\mathbf{HX}_{1}^{*} = 0$ and $\mathbf{HX}_{2}^{*} = 0$.

Proof

1. From result B1, $\mathbf{H}_{\mathbf{X}^{*}}$ is symmetric and idempotent, then

$$\mathbf{H}' = \left[\mathbf{I}_{\mathrm{T}} - \mathbf{H}_{\mathbf{X}^{*}}\right]' = \mathbf{I}_{\mathrm{T}} - \mathbf{H}'_{\mathbf{X}^{*}} = \mathbf{I}_{\mathrm{T}} - \mathbf{H}_{\mathbf{X}^{*}} = \mathbf{H}$$

and

$$\mathbf{HH} = \left[\mathbf{I}_{\mathrm{T}} - \mathbf{X}^{*}\left(\mathbf{X}^{*\prime}\mathbf{X}^{*}\right)^{-1}\mathbf{X}^{*\prime}\right]\left[\mathbf{I}_{\mathrm{T}} - \mathbf{X}^{*}\left(\mathbf{X}^{*\prime}\mathbf{X}^{*}\right)^{-1}\mathbf{X}^{*\prime}\right],$$

$$= \mathbf{I}_{\mathrm{T}} - \mathbf{X}^{*}\left(\mathbf{X}^{*\prime}\mathbf{X}^{*}\right)^{-1}\mathbf{X}^{*\prime} - \mathbf{X}^{*}\left(\mathbf{X}^{*\prime}\mathbf{X}^{*}\right)^{-1}\mathbf{X}^{*\prime} + \mathbf{X}^{*}\left(\mathbf{X}^{*\prime}\mathbf{X}^{*}\right)^{-1}\mathbf{X}^{*\prime},$$

$$= \mathbf{I}_{\mathrm{T}} - \mathbf{H}_{\mathbf{X}^{*}} = \mathbf{H}. \qquad \blacksquare$$

2. Since $\mathbf{H}$ is symmetric and idempotent, then

$$\text{rank}\,\mathbf{H} = \text{rank}\left[\mathbf{I}_{\mathrm{T}} - \mathbf{H}_{\mathbf{X}^{*}}\right],$$

$$= tr\left[\mathbf{I}_{\mathrm{T}} - \mathbf{H}_{\mathbf{X}^{*}}\right],$$

$$= \mathrm{T} - tr\left[\mathbf{H}_{\mathbf{X}^{*}}\right] = \mathrm{T} - \mathrm{p}. \qquad \blacksquare$$

3. $\mathbf{HX}^{*} = \left[\mathbf{I}_{\mathrm{T}} - \mathbf{X}^{*}\left(\mathbf{X}^{*\prime}\mathbf{X}^{*}\right)^{-1}\mathbf{X}^{*\prime}\right]\mathbf{X}^{*} = \mathbf{X}^{*} - \mathbf{X}^{*} = 0.$ $\blacksquare$

4. We can write $\mathbf{HX}^{*} = \mathbf{H}(\mathbf{X}_1^{*}, \mathbf{X}_2^{*}) = (\mathbf{HX}_1^{*}, \mathbf{HX}_2^{*})$ and, using result (3), it follows that $\mathbf{HX}^{*} = 0$, then $\mathbf{HX}_1^{*} = 0$ and $\mathbf{HX}_2^{*} = 0$. ■

(B3) Let $\mathbf{M}_{\mathbf{X}_1^{*}} = \mathbf{X}_1^{*}\left(\mathbf{X}_1^{*\prime}\mathbf{X}_1^{*}\right)^{-1}\mathbf{X}_1^{*\prime}$ be a $\mathrm{T}\times\mathrm{T}$ matrix, then:

1. $\mathbf{M}_{\mathbf{X}_1^{*}}$ is symmetric and idempotent
2. rank $\mathbf{M}_{\mathbf{X}_1^{*}} = tr\left(\mathbf{M}_{\mathbf{X}_1^{*}}\right) = \mathrm{q}$
3. $\mathbf{M}_{\mathbf{X}_1^{*}}\mathbf{X}_1^{*} = \mathbf{X}_1^{*}$

Proof

The proof is omitted since it can be done in the same manner as (B1). ■

(B4) Let $\mathbf{M} = \mathbf{I}_{\mathrm{T}} - \mathbf{X}_1^{*}\left(\mathbf{X}_1^{*\prime}\mathbf{X}_1^{*}\right)^{-1}\mathbf{X}_1^{*\prime} = \mathbf{I}_{\mathrm{T}} - \mathbf{M}_{\mathbf{X}_1^{*}}$ be a $\mathrm{T}\times\mathrm{T}$ matrix, then:

1. $\mathbf{M}$ is symmetric and idempotent
2. rank $\mathbf{M} = tr(\mathbf{M}) = \mathrm{T} - \mathrm{q}$
3. $\mathbf{M}$ is positive definite
4. $\mathbf{MX}_1^{*} = 0$

Proof

1. and 2. They can be proved in the same manner as having done in (B3). ■

3. It can be proven that $\mathbf{v}'\mathbf{Mv} > 0$ for all $\mathbf{v} \in \mathcal{R}^{\mathcal{T}}$ except $\mathbf{v} = 0$.

Suppose $\mathbf{M} = \mathbf{P}'\mathbf{P}$ for a nonsingular $\mathbf{P}$, then

$$\mathbf{v}'\mathbf{Mv} = \mathbf{v}'\mathbf{P}'\mathbf{Pv} = (\mathbf{Pv})'(\mathbf{Pv}),$$

which is the sum of squares of the elements of $\mathbf{Pv}$. Hence, $\mathbf{v'Mv} > 0$ for all $\mathbf{Pv} \neq 0$, and $\mathbf{v'Mv} = 0$ for all $\mathbf{Pv} = 0$, but $\mathbf{Pv} = 0$ only when $\mathbf{v} = 0$ because $\mathbf{P}^{-1}$ exists. Since $\mathbf{v'Mv} > 0$ for all $\mathbf{v} \neq 0$, and $\mathbf{v'Mv} = 0$ only for $\mathbf{v} = 0$, then $\mathbf{M}$ is positive definite. ■

4. $\mathbf{MX}_1^* = \left[\mathbf{I}_T - \mathbf{X}_1^*\left(\mathbf{X}_1^{*\prime}\mathbf{X}_1^*\right)^{-1}\mathbf{X}_1^{*\prime}\right]\mathbf{X}_1^* = \mathbf{X}_1^* - \mathbf{X}_1^* = 0$. ■

(B5) Let $\mathbf{W} = \mathbf{MX}_2^*\left(\mathbf{X}_2^{*\prime}\mathbf{MX}_2^*\right)^{-1}\mathbf{X}_2^{*\prime}\mathbf{M}$ be a $T \times T$ matrix since:

1. $\mathbf{W}$ is symmetric and idempotent
2. rank $\mathbf{W} = tr(\mathbf{W}) = p - q$
3. $\mathbf{W}$ is positive definite
4. $\mathbf{WX}_1^* = 0$

Proof

See (B1) and (B3). ■

APPENDIX C

THE ORDINARY LEAST SQUARES ESTIMATOR

Consider a multiple linear regression model, which is transformed from the first-order autoregression model. And it is free from serial correlation as follow;

$$\mathbf{y}^* = \mathbf{X}^* \boldsymbol{\tau} + \boldsymbol{\varepsilon} = \mathbf{X}_1^* \boldsymbol{\theta} + \mathbf{X}_2^* \boldsymbol{\beta} + \boldsymbol{\varepsilon}, \tag{C1}$$

where $\boldsymbol{\varepsilon}$ is a $\mathrm{T} \times 1$ vector of i.i.d. random vector with mean 0 and variance σ^2.

(C1) The OLS estimators of the regression coefficient of $\boldsymbol{\tau}$, $\boldsymbol{\theta}$, $\boldsymbol{\beta}$ and $\boldsymbol{\theta}_{\mathrm{Red}}$ are:

1. $\hat{\boldsymbol{\tau}} = \left(\mathbf{X}^{*\prime} \mathbf{X}^*\right)^{-1} \mathbf{X}^{*\prime} \mathbf{y}^*$
2. $\hat{\boldsymbol{\theta}} = \left(\mathbf{X}_1^{*\prime} \mathbf{X}_1^*\right)^{-1} \mathbf{X}_1^{*\prime} \left(\mathbf{y}^* - \mathbf{X}_2^* \hat{\boldsymbol{\beta}}\right)$
3. $\hat{\boldsymbol{\beta}} = \left(\mathbf{X}_2^{*\prime} \mathbf{M} \mathbf{X}_2^*\right)^{-1} \left(\mathbf{M} \mathbf{X}_2^*\right)' \mathbf{y}^*$
4. $\hat{\boldsymbol{\theta}}_{\mathrm{Red}} = \left(\mathbf{X}_1^{*\prime} \mathbf{X}_1^*\right)^{-1} \mathbf{X}_1^{*\prime} \mathbf{y}$

Proof

1. Given $\mathbf{y}^* = \mathbf{X}^* \boldsymbol{\tau} + \boldsymbol{\varepsilon}$, then the residual $\mathbf{e} = \mathbf{y}^* - \hat{\mathbf{y}}^* = \mathbf{y}^* - \mathbf{X}^* \hat{\boldsymbol{\tau}}$,

$$\mathbf{e}'\mathbf{e} = \left(\mathbf{y}^* - \mathbf{X}^* \hat{\boldsymbol{\tau}}\right)' \left(\mathbf{y}^* - \mathbf{X}^* \hat{\boldsymbol{\tau}}\right),$$

$$= \mathbf{y}^{*\prime} \mathbf{y}^* - \mathbf{y}^{*\prime} \mathbf{X}^* \hat{\boldsymbol{\tau}} - \hat{\boldsymbol{\tau}}' \mathbf{X}^{*\prime} \mathbf{y}^* + \hat{\boldsymbol{\tau}}' \mathbf{X}^{*\prime} \hat{\boldsymbol{\tau}}.$$

Assume that $\partial \mathbf{e}'\mathbf{e} / \partial \hat{\tau} = 0$ is solved. Thus, we have

$$-2\mathbf{X}^{*\prime} \mathbf{y}^* + 2\mathbf{X}^{*\prime} \mathbf{X}^* \hat{\boldsymbol{\tau}} = 0. \tag{C2}$$

From (C2), the OLS estimator of $\boldsymbol{\tau}$ is

$$\hat{\boldsymbol{\tau}} = \left(\mathbf{X}^{*\prime}\mathbf{X}^{*}\right)^{-1}\mathbf{X}^{*\prime}\mathbf{y}^{*}. \tag{C3}$$

(C2) Given $\mathbf{y}^{*} = \mathbf{X}_1^{*}\boldsymbol{\theta} + \mathbf{X}_2^{*}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, then the residual $\mathbf{e} = \mathbf{y}^{*} - \hat{\mathbf{y}}^{*} = \mathbf{y}^{*} - \mathbf{X}_1^{*}\hat{\boldsymbol{\theta}} + \mathbf{X}_2^{*}\hat{\boldsymbol{\beta}}$,

$$\begin{aligned}
\mathbf{e}'\mathbf{e} &= \left(\mathbf{y}^{*} - \mathbf{X}_1^{*}\hat{\boldsymbol{\theta}} + \mathbf{X}_2^{*}\hat{\boldsymbol{\beta}}\right)'\left(\mathbf{y}^{*} - \mathbf{X}_1^{*}\hat{\boldsymbol{\theta}} + \mathbf{X}_2^{*}\hat{\boldsymbol{\beta}}\right), \\
&= \mathbf{y}^{*\prime}\mathbf{y}^{*} - \mathbf{y}^{*\prime}\mathbf{X}_1^{*}\hat{\boldsymbol{\theta}} - \mathbf{y}^{*\prime}\mathbf{X}_2^{*}\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\theta}}'\mathbf{X}_1^{*\prime}\mathbf{y}^{*} + \hat{\boldsymbol{\theta}}'\mathbf{X}_1^{*\prime}\mathbf{X}_1^{*}\hat{\boldsymbol{\theta}} + \hat{\boldsymbol{\theta}}'\mathbf{X}_1^{*\prime}\mathbf{X}_2^{*}\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}'\mathbf{X}_2^{*\prime}\mathbf{y}^{*} \\
&\quad + \hat{\boldsymbol{\beta}}'\mathbf{X}_2^{*\prime}\mathbf{X}_1^{*}\hat{\boldsymbol{\theta}} + \hat{\boldsymbol{\beta}}'\mathbf{X}_2^{*\prime}\mathbf{X}_2^{*}\hat{\boldsymbol{\beta}}, \\
&= \mathbf{y}^{*\prime}\mathbf{y}^{*} - 2\hat{\boldsymbol{\theta}}'\mathbf{X}_1^{*\prime}\mathbf{y}^{*} - 2\hat{\boldsymbol{\beta}}'\mathbf{X}_2^{*\prime}\mathbf{y}^{*} + 2\hat{\boldsymbol{\beta}}'\mathbf{X}_2^{*\prime}\mathbf{X}_1^{*}\hat{\boldsymbol{\theta}} + \hat{\boldsymbol{\theta}}'\mathbf{X}_1^{*\prime}\mathbf{X}_1^{*}\hat{\boldsymbol{\theta}} + \hat{\boldsymbol{\beta}}'\mathbf{X}_2^{*\prime}\mathbf{X}_2^{*}\hat{\boldsymbol{\beta}}.
\end{aligned}$$

So as to find $\hat{\boldsymbol{\theta}}$ and $\hat{\boldsymbol{\beta}}$, we set $\partial\mathbf{e}'\mathbf{e}/\partial\hat{\boldsymbol{\theta}} = 0$ and $\partial\mathbf{e}'\mathbf{e}/\partial\hat{\boldsymbol{\beta}} = 0$. Thus, the following equations are obtained:

$$-2\mathbf{X}_1^{*\prime}\mathbf{y}^{*} + 2\mathbf{X}_1^{*\prime}\mathbf{X}_2^{*}\hat{\boldsymbol{\beta}} + 2\mathbf{X}_1^{*\prime}\mathbf{X}_1^{*}\hat{\boldsymbol{\theta}} = \mathbf{0} \tag{C4}$$

$$-2\mathbf{X}_2^{*\prime}\mathbf{y}^{*} + 2\mathbf{X}_2^{*\prime}\mathbf{X}_1^{*}\hat{\boldsymbol{\theta}} + 2\mathbf{X}_2^{*\prime}\mathbf{X}_2^{*}\hat{\boldsymbol{\beta}} = \mathbf{0} \tag{C5}$$

From (C4), the OLS estimator of $\boldsymbol{\theta}$ is

$$\hat{\boldsymbol{\theta}} = \left(\mathbf{X}_1^{*\prime}\mathbf{X}_1^{*}\right)^{-1}\mathbf{X}_1^{*\prime}\left(\mathbf{y}^{*} - \mathbf{X}_2^{*}\hat{\boldsymbol{\beta}}\right). \tag{C6}$$

(C3) Substituting (C6) into (C5) leads to

$$\mathbf{X}_2^{*\prime}\mathbf{X}_2^{*}\hat{\boldsymbol{\beta}} = \mathbf{X}_2^{*\prime}\mathbf{y}^{*} - \mathbf{X}_2^{*\prime}\mathbf{X}_1^{*}\left(\mathbf{X}_1^{*\prime}\mathbf{X}_1^{*}\right)^{-1}\mathbf{X}_1^{*\prime}\left(\mathbf{y}^{*} - \mathbf{X}_2^{*}\hat{\boldsymbol{\beta}}\right)$$

$$\mathbf{X}_2^{*'}\left[\mathbf{I}_T-\mathbf{X}_1^*\left(\mathbf{X}_1^{*'}\mathbf{X}_1^*\right)^{-1}\mathbf{X}_1^{*'}\right]\mathbf{X}_2^*\hat{\boldsymbol{\beta}}=\mathbf{X}_2^{*'}\left[\mathbf{I}_T-\mathbf{X}_1^*\left(\mathbf{X}_1^{*'}\mathbf{X}_1^*\right)^{-1}\mathbf{X}_1^{*'}\right]\mathbf{y}^*$$

or

$$\mathbf{X}_2^{*'}\mathbf{M}\mathbf{X}_2^*\hat{\boldsymbol{\beta}}=\mathbf{X}_2^{*'}\mathbf{M}\mathbf{y}^* \ , \tag{C7}$$

where $\mathbf{M}=\mathbf{I}_T-\mathbf{X}_1^*\left(\mathbf{X}_1^{*'}\mathbf{X}_1^*\right)^{-1}\mathbf{X}_1^{*'}$.

Since $\left(\mathbf{X}_2^{*'}\mathbf{M}\mathbf{X}_2^*\right)$ is nonsingular (Seber, 1977: 65), then the OLS estimator of $\boldsymbol{\beta}$ is

$$\hat{\boldsymbol{\beta}}=\left(\mathbf{X}_2^{*'}\mathbf{M}\mathbf{X}_2^*\right)^{-1}\mathbf{X}_2^{*'}\mathbf{M}\mathbf{y}^* \ . \tag{C8}$$

(C4) Given that $\mathbf{y}^*=\mathbf{X}_1^*\boldsymbol{\theta}+\boldsymbol{\varepsilon}^*$, then the residual $\mathbf{e}^*=\mathbf{y}^*-\hat{\mathbf{y}}^*=\mathbf{y}^*-\mathbf{X}_1^*\hat{\boldsymbol{\theta}}_{\text{Red}}$,

$$\begin{aligned}\mathbf{e}^{*'}\mathbf{e}^* &= \left(\mathbf{y}^*-\mathbf{X}_1^*\hat{\boldsymbol{\theta}}_{\text{Red}}\right)'\left(\mathbf{y}^*-\mathbf{X}_1^*\hat{\boldsymbol{\theta}}_{\text{Red}}\right), \\ &= \mathbf{y}^{*'}\mathbf{y}^*-\mathbf{y}^{*'}\mathbf{X}_1^*\hat{\boldsymbol{\theta}}_{\text{Red}}-\hat{\boldsymbol{\theta}}'_{\text{Red}}\mathbf{X}_1^{*'}\mathbf{y}^*+\hat{\boldsymbol{\theta}}'_{\text{Red}}\mathbf{X}_1^{*'}\mathbf{X}_1^*\hat{\boldsymbol{\theta}}'_{\text{Red}} \ .\end{aligned}$$

In order to find $\hat{\boldsymbol{\theta}}_{\text{Red}}$, the normal equation $\partial\mathbf{e}'\mathbf{e}/\hat{\boldsymbol{\theta}}_{\text{Red}}$ is solved. Thus, we have

$$-2\mathbf{X}_1^{*'}\mathbf{y}^*+2\mathbf{X}_1^{*'}\mathbf{X}_1^*\hat{\boldsymbol{\theta}}_{\text{Red}}=0 \ . \tag{C9}$$

From (C9), the OLS estimator of $\boldsymbol{\theta}_{\text{Red}}$ is

$$\hat{\boldsymbol{\theta}}_{\text{Red}}=\left(\mathbf{X}_1^{*'}\mathbf{X}_1^*\right)^{-1}\mathbf{X}_1^{*'}\mathbf{y} \ . \tag{C10}$$

APPENDIX D

THE RESIDUALS UNDER THE REDUCED MODEL

Under the null hypothesis, the reduced model is $\mathbf{y}^* = \mathbf{X}_1^*\boldsymbol{\theta} + \boldsymbol{\varepsilon}^*$ and $\hat{\boldsymbol{\theta}}_{\text{Red}} = \left(\mathbf{X}_1^{*\prime}\mathbf{X}_1^*\right)^{-1}\mathbf{X}_1^{*\prime}\mathbf{y}$ is an estimate of the unknown parameter $\boldsymbol{\theta}$. The residual of this regression approximates the error $\boldsymbol{\varepsilon}$ that are exchangeable under the null hypothesis and is given by

$$\begin{aligned}
\mathbf{R}_{\mathbf{y}^*|\mathbf{X}_1^*} &= \mathbf{y}^* - \mathbf{X}_1^*\hat{\boldsymbol{\theta}}_{\text{Red}}, \\
&= \mathbf{y}^* - \mathbf{X}_1^*\left[\left(\mathbf{X}_1^{*\prime}\mathbf{X}_1^*\right)^{-1}\mathbf{X}_1^{*\prime}\mathbf{y}^*\right], \\
&= \mathbf{M}\mathbf{y}^*,
\end{aligned}$$

where $\mathbf{M} = \mathbf{I}_T - \mathbf{X}_1^*\left(\mathbf{X}_1^{*\prime}\mathbf{X}_1^*\right)^{-1}\mathbf{X}_1^{*\prime}$. ■

(D1) From assumption 3.1, the error vector $\boldsymbol{\varepsilon}$ is independent and identically distributed with $E(\varepsilon_t) = 0$, $0 < E\left(\varepsilon_t^2\right) = \sigma^2 < \infty$, for $t = 1, 2, ..., T$. Under the null hypothesis, $H_0 : \boldsymbol{\beta} = \mathbf{0}$, $E\left(\mathbf{y}^*\right) = \mathbf{X}_1^*\boldsymbol{\theta}$ and $\text{Var}\left(\mathbf{y}^*\right) = \sigma^2\mathbf{I}$.

Proof

$$E\left(\mathbf{y}^*\right) = E\left(\mathbf{X}_1^*\boldsymbol{\theta} + \boldsymbol{\varepsilon}\right) = \mathbf{X}_1^*\boldsymbol{\theta} \quad \text{and}$$

$$\text{Var}\left(\mathbf{y}^*\right) = \text{Var}\left(\mathbf{X}_1^*\boldsymbol{\theta} + \boldsymbol{\varepsilon}\right) = \sigma^2\mathbf{I}. \quad ■$$

(D2) Since $\mathbf{R}_{\mathbf{y}^*|\mathbf{X}_1^*} = \mathbf{M}\mathbf{y}^* = \left[\mathbf{I}_T - \mathbf{X}_1^*\left(\mathbf{X}_1^{*\prime}\mathbf{X}_1^*\right)^{-1}\mathbf{X}_1^{*\prime}\right]\mathbf{y}^*$ and using result **D1**, then

1. $E\left(\mathbf{R}_{\mathbf{y}^*|\mathbf{X}_1^*}\right) = 0$

2. $\text{Var}\left(\mathbf{R}_{\mathbf{y}^*|\mathbf{X}_1^*}\right) = \sigma^2\mathbf{M}$

3. $E\left(T^{-1}\mathbf{R}'_{\mathbf{y}^*|\mathbf{x}_1^*}\mathbf{R}_{\mathbf{y}^*|\mathbf{x}_1^*}\right) = \sigma^2\mathbf{M}$

4. $\mathrm{Var}\left(T^{-1}\mathbf{R}'_{\mathbf{y}^*|\mathbf{x}_1^*}\mathbf{R}_{\mathbf{y}^*|\mathbf{x}_1^*}\right) = 2T^{-2}\sigma^4(T - q)$

Proof

1. $$E\left(\mathbf{R}_{\mathbf{y}^*|\mathbf{x}_1^*}\right) = E\left(\mathbf{M}\mathbf{y}^*\right),$$

$$= \left[\mathbf{I}_T - \mathbf{X}_1^*\left(\mathbf{X}_1^{*\prime}\mathbf{X}_1^*\right)^{-1}\mathbf{X}_1^{*\prime}\right]E\left(\mathbf{y}^*\right),$$

$$= \left[\mathbf{I}_T - \mathbf{X}_1^*\left(\mathbf{X}_1^{*\prime}\mathbf{X}_1^*\right)^{-1}\mathbf{X}_1^{*\prime}\right]\mathbf{X}_1^*\boldsymbol{\theta} = 0 .$$ ■

2. $$\mathrm{Var}\left(\mathbf{R}_{\mathbf{y}^*|\mathbf{x}_1^*}\right) = \mathbf{M}'\,\mathrm{Var}\left(\mathbf{y}^*\right)\mathbf{M} = \mathbf{M}'\left(\sigma^2\mathbf{I}\right)\mathbf{M},$$

$$= \sigma^2\mathbf{M}.$$ ■

3. $$E\left(T^{-1}\mathbf{R}'_{\mathbf{y}^*|\mathbf{x}_1^*}\mathbf{R}_{\mathbf{y}^*|\mathbf{x}_1^*}\right) = T^{-1}E\left(\sum_{t=1}^{T}R_t^2\right),$$

$$= T^{-1}\sum_{t}^{T}E\left(R_t^2\right) = \sigma^2\mathbf{M} .$$ ■

4. $$\mathrm{Var}\left(T^{-1}\mathbf{R}'_{\mathbf{y}^*|\mathbf{x}_1^*}\mathbf{R}_{\mathbf{y}^*|\mathbf{x}_1^*}\right) = T^{-2}\mathrm{Var}\left(\mathbf{y}^{*\prime}\mathbf{M}\mathbf{y}^*\right),$$

$$= T^{-2}\left\{2tr\left(\mathbf{M}\sigma^2\mathbf{I}\right)^2 + 4\left(\mathbf{X}_1^*\boldsymbol{\theta}\right)'\mathbf{M}\sigma^2\mathbf{I}\left(\mathbf{X}_1^*\boldsymbol{\theta}\right)\right\},$$

$$= T^{-2}\left\{2\sigma^4(T - q)\right\}$$ (Rencher and Schaalje, 2007: 109). ■

Appendix E

THE F_{ref} STATISTIC

Recall that, using (C2) and (C9) from appendix C, we obtain

$$\hat{\boldsymbol{\tau}}'\mathbf{X}^{*'}\mathbf{y}^* - \hat{\boldsymbol{\theta}}_{\text{Red}}\mathbf{X}_1^{*'}\mathbf{y}^* = \hat{\boldsymbol{\tau}}'\mathbf{X}^{*'}\mathbf{X}^*\hat{\boldsymbol{\tau}} - \hat{\boldsymbol{\theta}}'_{\text{Red}}\mathbf{X}_1^{*'}\mathbf{X}_1^*\hat{\boldsymbol{\theta}}_{\text{Red}},$$

$$= \left(\mathbf{X}^*\hat{\boldsymbol{\tau}}\right)'\mathbf{X}^*\hat{\boldsymbol{\tau}} - \hat{\boldsymbol{\theta}}'_{\text{Red}}\mathbf{X}_1^{*'}\mathbf{X}_1^*\hat{\boldsymbol{\theta}}_{\text{Red}}. \qquad \text{(E1)}$$

However, $\hat{\boldsymbol{\theta}}_{\text{Red}} = \left(\mathbf{X}_1^{*'}\mathbf{X}_1^*\right)^{-1}\mathbf{X}_1^{*'}\mathbf{y}$ can be rewritten as

$$\hat{\boldsymbol{\theta}}_{\text{Red}} = \left(\mathbf{X}_1^{*'}\mathbf{X}_1^*\right)^{-1}\mathbf{X}_1^{*'}\mathbf{y},$$

$$= \left(\mathbf{X}_1^{*'}\mathbf{X}_1^*\right)^{-1}\mathbf{X}_1^{*'}\left(\mathbf{y}^* - \mathbf{X}_2^*\hat{\boldsymbol{\beta}}\right) + \left(\mathbf{X}_1^{*'}\mathbf{X}_1^*\right)^{-1}\mathbf{X}_1^{*'}\mathbf{X}_2^*\hat{\boldsymbol{\beta}},$$

$$= \hat{\boldsymbol{\theta}} + \left(\mathbf{X}_1^{*'}\mathbf{X}_1^*\right)^{-1}\mathbf{X}_1^{*'}\mathbf{X}_2^*\hat{\boldsymbol{\beta}}. \qquad \text{(E2)}$$

From (E2) and $\mathbf{X}^*\hat{\boldsymbol{\tau}} = \mathbf{X}_1^*\hat{\boldsymbol{\theta}} + \mathbf{X}_2^*\hat{\boldsymbol{\beta}}$, then (E1) becomes

$$\hat{\boldsymbol{\tau}}'\mathbf{X}^{*'}\mathbf{y}^* - \hat{\boldsymbol{\theta}}_{\text{Red}}\mathbf{X}_1^{*'}\mathbf{y}$$

$$= \left(\mathbf{X}_1^*\hat{\boldsymbol{\theta}} + \mathbf{X}_2^*\hat{\boldsymbol{\beta}}\right)'\left(\mathbf{X}_1^*\hat{\boldsymbol{\theta}} + \mathbf{X}_2^*\hat{\boldsymbol{\beta}}\right) - \left(\hat{\boldsymbol{\theta}} + \left(\mathbf{X}_1^{*'}\mathbf{X}_1^*\right)^{-1}\mathbf{X}_1^{*'}\mathbf{X}_2^*\hat{\boldsymbol{\beta}}\right)'\mathbf{X}_1^{*'}\mathbf{X}_1^*\left(\hat{\boldsymbol{\theta}} + \left(\mathbf{X}_1^{*'}\mathbf{X}_1^*\right)^{-1}\mathbf{X}_1^{*'}\mathbf{X}_2^*\hat{\boldsymbol{\beta}}\right),$$

$$= \hat{\boldsymbol{\theta}}'\mathbf{X}_1^{*'}\mathbf{X}_1^*\hat{\boldsymbol{\theta}} + \hat{\boldsymbol{\theta}}'\mathbf{X}_1^{*'}\mathbf{X}_2^*\hat{\boldsymbol{\beta}} + \hat{\boldsymbol{\beta}}'\mathbf{X}_2^{*'}\mathbf{X}_1^*\hat{\boldsymbol{\theta}} + \hat{\boldsymbol{\beta}}'\mathbf{X}_2^{*'}\mathbf{X}_2^*\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\theta}}'\mathbf{X}_1^{*'}\mathbf{X}_1^*\hat{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}}'\mathbf{X}_1^{*'}\mathbf{X}_1^*\left(\mathbf{X}_1^{*'}\mathbf{X}_1^*\right)^{-1}\mathbf{X}_1^{*'}\mathbf{X}_2^*\hat{\boldsymbol{\beta}}$$

$$-\hat{\boldsymbol{\beta}}'\mathbf{X}_2^{*'}\mathbf{X}_1^*\left(\mathbf{X}_1^{*'}\mathbf{X}_1^*\right)^{-1}\mathbf{X}_1^{*'}\mathbf{X}_1^*\hat{\boldsymbol{\theta}} - \hat{\boldsymbol{\beta}}'\mathbf{X}_2^{*'}\mathbf{X}_1^*\left(\mathbf{X}_1^{*'}\mathbf{X}_1^*\right)^{-1}\mathbf{X}_1^{*'}\mathbf{X}_1^*\left(\mathbf{X}_1^{*'}\mathbf{X}_1^*\right)^{-1}\mathbf{X}_1^{*'}\mathbf{X}_2^*\hat{\boldsymbol{\beta}}$$

$$= \hat{\boldsymbol{\beta}}'\mathbf{X}_2^{*'}\mathbf{X}_2^*\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}'\mathbf{X}_2^{*'}\mathbf{X}_1^*\left(\mathbf{X}_1^{*'}\mathbf{X}_1^*\right)^{-1}\mathbf{X}_1^{*'}\mathbf{X}_2^*\hat{\boldsymbol{\beta}},$$

$$= \hat{\boldsymbol{\beta}}'\mathbf{X}_2^{*'}\left[\mathbf{I}_{\text{T}} - \mathbf{X}_1^*\left(\mathbf{X}_1^{*'}\mathbf{X}_1^*\right)^{-1}\mathbf{X}_1^{*'}\right]\mathbf{X}_2^*\hat{\boldsymbol{\beta}},$$

$$= \hat{\boldsymbol{\beta}}'\mathbf{X}_2^{*'}\mathbf{M}\mathbf{X}_2^*\hat{\boldsymbol{\beta}}, \qquad \text{where } \mathbf{M} = \mathbf{I}_{\text{T}} - \mathbf{X}_1^*\left(\mathbf{X}_1^{*'}\mathbf{X}_1^*\right)^{-1}\mathbf{X}_1^{*'}. \qquad \text{(E3)}$$

Consider the formula of the partial F statistic in (3.5);

$$F = \frac{\mathbf{y}^{*'}\left[\mathbf{X}^{*}\left(\mathbf{X}^{*'}\mathbf{X}^{*}\right)^{-1}\mathbf{X}^{*'} - \mathbf{X}_1^{*}\left(\mathbf{X}_1^{*'}\mathbf{X}_1^{*}\right)^{-1}\mathbf{X}_1^{*'}\right]\mathbf{y}^{*} \Big/ (p-q)}{\mathbf{y}^{*'}\left[\mathbf{I} - \mathbf{X}^{*}\left(\mathbf{X}^{*'}\mathbf{X}^{*}\right)^{-1}\mathbf{X}^{*'}\right]\mathbf{y}^{*} \Big/ (T-p)},$$

$$= \frac{\left(\hat{\boldsymbol{\tau}}'\mathbf{X}^{*'}\mathbf{y}^{*} - \hat{\boldsymbol{\theta}}_{\text{Red}}\mathbf{X}_1^{*'}\mathbf{y}^{*}\right)' \Big/ (p-q)}{\mathbf{y}^{*'}\mathbf{H}\,\mathbf{y}^{*} \Big/ (T-p)},$$

$$= \frac{\hat{\boldsymbol{\beta}}'\mathbf{X}_2^{*'}\mathbf{M}\mathbf{X}_2^{*}\hat{\boldsymbol{\beta}} \Big/ (p-q)}{\mathbf{y}^{*'}\mathbf{H}\,\mathbf{y}^{*} \Big/ (T-p)}, \tag{E4}$$

where $\mathbf{H} = \mathbf{I}_T - \mathbf{X}^{*}\left(\mathbf{X}^{*'}\mathbf{X}^{*}\right)^{-1}\mathbf{X}^{*'}$.

(E4) can be rewritten in a form that omits the number of partial regression coefficients and the degrees of freedom of the partial F statistic to produce F_{ref};

$$F_{ref} = \frac{\hat{\boldsymbol{\beta}}'\left(\mathbf{X}_2^{*'}\mathbf{M}\,\mathbf{X}_2^{*}\right)\hat{\boldsymbol{\beta}}}{\mathbf{y}'\mathbf{H}\,\mathbf{y}}. \tag{E5}$$

APPENDIX F

REJECTION RATES OF THE TESTS

Table F.1 Empirical Type I Error Rates of the Five Tests at $\alpha = 0.01^{*}$ when $\boldsymbol{\varepsilon} \sim N(0,1)$ and $(\theta_1, \theta_2, \theta_3) = (1,1,1)$

T	Test	ρ					
		-0.75	**-0.5**	**-0.25**	**0.25**	**0.5**	**0.75**
12	PPT	0.0105	0.0145	0.0175	0.0165	0.0130	0.0160
	FPT	0.0160	0.0155	0.0160	0.0145	0.0135	0.0155
	MPT	0.0810	0.0760	0.0810	0.0720	0.0655	0.0505
	ACT	0.0250	0.0230	0.0205	0.0235	0.0185	0.0215
	PFT	0.0160	0.0175	0.0160	0.0150	0.0155	0.0140
16	PPT	0.0145	0.0190	0.0250	0.0275	0.0195	0.0150
	FPT	0.0130	0.0185	0.0255	0.0280	0.0240	0.0130
	MPT	0.0670	0.0835	0.0930	0.0775	0.0740	0.0540
	ACT	0.0100	0.0110	0.0165	0.0170	0.0115	0.0075
	PFT	0.0185	0.0250	0.0245	0.0250	0.0195	0.0145
20	PPT	0.0070	0.0105	0.0175	0.0210	0.0210	0.0105
	FPT	0.0080	0.0115	0.0170	0.0235	0.0200	0.0125
	MPT	0.0675	0.0745	0.0705	0.0705	0.0600	0.0605
	ACT	0.0100	0.0050	0.0060	0.0080	0.0055	0.0040
	PFT	0.0085	0.0110	0.0165	0.0200	0.0180	0.0105

* *PPT, FPT and MPT are based on 2,000 Monte Carlo replications.*

Table F.2 Empirical Type I Error Rates of the Five Tests at $\alpha = 0.01^{*}$ when $\boldsymbol{\varepsilon} \sim U(-2,2)$ and $(\theta_1, \theta_2, \theta_3) = (1,1,1)$

T	Test	ρ					
		-0.75	**-0.5**	**-0.25**	**0.25**	**0.5**	**0.75**
	PPT	0.0145	0.0200	0.0220	0.0175	0.0155	0.0160
	FPT	0.0120	0.01900	0.0205	0.0175	0.0165	0.0160
12	MPT	0.0590	0.0715	0.0805	0.0710	0.0690	0.0655
	ACT	0.0170	0.0240	0.0265	0.0235	0.0220	0.0050
	PFT	0.0255	0.0165	0.0200	0.0160	0.0170	0.0035
	PPT	0.0125	0.0155	0.0195	0.0200	0.0185	0.0160
	FPT	0.0130	0.0155	0.0175	0.0200	0.017	0.0160
16	MPT	0.0665	0.0755	0.0770	0.0760	0.0655	0.0655
	ACT	0.0060	0.0070	0.0105	0.0130	0.0085	0.0205
	PFT	0.0130	0.0035	0.0175	0.0180	0.0165	0.0140
	PPT	0.0075	0.0155	0.0210	0.0205	0.0155	0.0120
	FPT	0.0070	0.0155	0.0105	0.0200	0.0160	0.0120
20	MPT	0.0665	0.0715	0.0665	0.0560	0.0570	0.0520
	ACT	0.0035	0.0040	0.0075	0.0065	0.0050	0.0020
	PFT	0.0070	0.0145	0.0205	0.0205	0.0165	0.0120

* *PPT, FPT and MPT are based on 2,000 Monte Carlo replications.*

Table F.3 Empirical Type I Error Rates of the Five Tests at $\alpha = 0.05^*$ when $\boldsymbol{\varepsilon} \sim N(0,1)$ and $(\theta_1, \theta_2, \theta_3) = (1,1,1)$

T	Test	ρ					
		-0.75	**-0.5**	**-0.25**	**0.25**	**0.5**	**0.75**
12	PPT	0.0460	0.0645	0.0750	0.0670	0.0645	0.0560
	FPT	0.0555	0.0735	0.0715	0.0695	0.0640	0.0590
	MPT	0.1170	0.1340	0.1380	0.1330	0.1205	0.1045
	ACT	0.0445	0.0695	0.0775	0.0715	0.0745	0.0710
	PFT	0.0530	0.0700	0.0765	0.0660	0.0645	0.0565
16	PPT	0.0475	0.0555	0.0715	0.0775	0.0700	0.0555
	FPT	0.0505	0.0745	0.0850	0.0790	0.0685	0.0555
	MPT	0.1015	0.1320	0.1310	0.1275	0.1205	0.0980
	ACT	0.0455	0.0540	0.0720	0.0755	0.0670	0.0510
	PFT	0.0505	0.0730	0.0850	0.0800	0.0685	0.0560
20	PPT	0.0425	0.0580	0.0635	0.0635	0.0590	0.0470
	FPT	0.0415	0.0580	0.0745	0.0700	0.0665	0.0445
	MPT	0.1025	0.1155	0.1230	0.1105	0.1095	0.1050
	ACT	0.0400	0.0580	0.0665	0.0615	0.0540	0.0440
	PFT	0.0395	0.0555	0.0740	0.0735	0.0640	0.0465

* *PPT, FPT and MPT are based on 2,000 Monte Carlo replications.*

Table F.4 Empirical Type I Error Rates of the Five Tests at $\alpha = 0.05^{*}$ when $\boldsymbol{\varepsilon} \sim U(-2,2)$ and $(\theta_1, \theta_2, \theta_3) = (1,1,1)$

T	Test	ρ					
		-0.75	**-0.5**	**-0.25**	**0.25**	**0.5**	**0.75**
12	PPT	0.0570	0.0670	0.0720	0.0695	0.0640	0.0665
	FPT	0.0590	0.0685	0.0735	0.0675	0.0640	0.0630
	MPT	0.1140	0.1370	0.1360	0.1265	0.1190	0.1075
	ACT	0.0585	0.0690	0.0745	0.0735	0.0690	0.0685
	PFT	0.0575	0.0670	0.0730	0.0700	0.0640	0.0640
16	PPT	0.0455	0.0650	0.0735	0.0735	0.0640	0.0495
	FPT	0.0505	0.0670	0.0730	0.0710	0.0650	0.0500
	MPT	0.1055	0.1260	0.1320	0.1255	0.1120	0.0985
	ACT	0.0465	0.0650	0.0725	0.0740	0.0715	0.0555
	PFT	0.0495	0.0645	0.0720	0.0720	0.0645	0.0505
20	PPT	0.0415	0.0585	0.0715	0.0680	0.0625	0.0465
	FPT	0.0405	0.0645	0.0740	0.0695	0.0695	0.0520
	MPT	0.1010	0.1190	0.1180	0.1100	0.1095	0.1010
	ACT	0.0410	0.0555	0.0695	0.0645	0.0675	0.0425
	PFT	0.0400	0.0640	0.0735	0.0690	0.0675	0.0510

** PPT, FPT and MPT are based on 2,000 Monte Carlo replications.*

. **Table F.5** Empirical Powers of the Five Tests at $\alpha = 0.01^{*}$ when $\boldsymbol{\varepsilon} \sim N(0,1)$, $(\theta_1, \theta_2, \theta_3) = (1,1,1)$ and $(\beta_1, \beta_2) = (1,1)$

T	**Test**	**ρ**					
		-0.75	**-0.5**	**-0.25**	**0.25**	**0.5**	**0.75**
12	PPT	0.8025	0.8885	0.9065	0.9240	0.9115	0.8830
	FPT	0.8055	0.9940	0.9130	0.9230	0.9105	0.8795
	MPT	0.9170	0.9390	0.9510	0.9495	0.9470	0.9400
	ACT	0.8485	0.9135	0.9375	0.9460	0.9380	0.9150
	PFT	0.8030	0.8835	0.9110	0.9210	0.9110	0.9885
16	PPT	0.9815	0.9935	0.9945	0.9965	0.9960	0.9890
	FPT	0.9820	0.9940	0.9980	0.9975	0.9960	0.9900
	MPT	0.9910	0.9970	0.9950	0.9970	0.9980	0.9940
	ACT	0.9805	0.9880	0.9910	0.9935	0.9905	0.9850
	PFT	0.9705	0.9950	0.9950	0.9970	0.9965	0.9885
20	PPT	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	FPT	0.9995	1.0000	1.0000	1.0000	1.0000	1.0000
	MPT	0.9995	0.9995	1.0000	1.0000	1.0000	1.0000
	ACT	0.9975	1.0000	1.0000	1.0000	0.9995	0.9985
	PFT	0.9995	1.0000	1.0000	1.0000	1.0000	1.0000

** PPT, FPT and MPT are based on 2,000 Monte Carlo replications.*

Table F.6 Empirical Power rates of five tests at $\alpha = 0.01^{*}$ when $\boldsymbol{\varepsilon} \sim U(-2,2)$, $(\theta_1,\theta_2,\theta_3) = (1,1,1)$ and $(\beta_1,\beta_2) = (1,1)$.

T	Test	ρ					
		-0.75	**-0.5**	**-0.25**	**0.25**	**0.5**	**0.75**
12	PPT	0.7240	0.8185	0.8455	0.8700	0.8530	0.8140
	FPT	0.7190	0.8155	0.8530	0.8700	0.8560	0.8140
	MPT	0.8750	0.9090	0.9190	0.9240	0.9220	0.9150
	ACT	0.7770	0.8600	0.8870	0.9030	0.8930	0.8560
	PFT	0.7155	0.8175	0.8500	0.8640	0.8530	0.8135
16	PPT	0.9710	0.9870	0.9920	0.9920	0.9905	0.9710
	FPT	0.9680	0.9880	0.9925	0.9925	0.9915	0.9680
	MPT	0.9860	0.9910	0.9925	0.9935	0.9935	0.9860
	ACT	0.9465	0.9780	0.9840	0.9855	0.9850	0.9835
	PFT	0.9705	0.9880	0.9925	0.9940	0.9910	0.9725
20	PPT	0.9960	0.9990	0.9995	0.9980	0.9975	0.9975
	FPT	0.9945	0.9995	0.9995	0.9990	0.9990	0.9990
	MPT	0.9985	0.9980	0.9995	0.9995	0.9990	0.9990
	ACT	0.9890	0.9940	0.9960	0.9960	0.9955	0.9955
	PFT	0.9950	0.9985	1.0000	0.9985	0.9980	0.9980

* *PPT, FPT and MPT are based on 2,000 Monte Carlo replications.*

Table F.7 Empirical Powers of the Five Tests at $\alpha = 0.05^{*}$ when $\boldsymbol{\varepsilon} \sim N(0,1)$, $(\theta_1, \theta_2, \theta_3) = (1, 1, 1)$ and $(\beta_1, \beta_2) = (1, 1)$

T	Test	ρ					
		-0.75	**-0.5**	**-0.25**	**0.25**	**0.5**	**0.75**
12	PPT	0.9480	0.9810	0.9840	0.9865	0.9855	0.9730
	FPT	0.9435	0.9760	0.9810	0.9830	0.9820	0.9730
	MPT	0.9355	0.9540	0.9660	0.9680	0.9695	0.9585
	ACT	0.9465	0.9790	0.9830	0.9865	0.9855	0.9735
	PFT	0.9450	0.9770	0.9920	0.9825	0.9825	0.9725
16	PPT	0.9960	0.9990	0.9995	0.9999	0.9995	0.9990
	FPT	0.9960	0.9990	0.9995	0.9995	0.9995	0.9990
	MPT	0.9940	0.9980	0.9985	0.9995	0.9995	0.9960
	ACT	0.9945	0.9985	0.9999	0.9999	0.9999	0.9999
	PFT	0.9960	0.9990	0.9995	0.9995	0.9995	0.9985
20	PPT	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	FPT	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	MPT	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	ACT	0.9975	0.9990	0.9995	1.0000	1.0000	1.0000
	PFT	0.9995	1.0000	1.0000	1.0000	1.0000	1.0000

* *PPT, FPT and MPT are based on 2,000 Monte Carlo replications.*

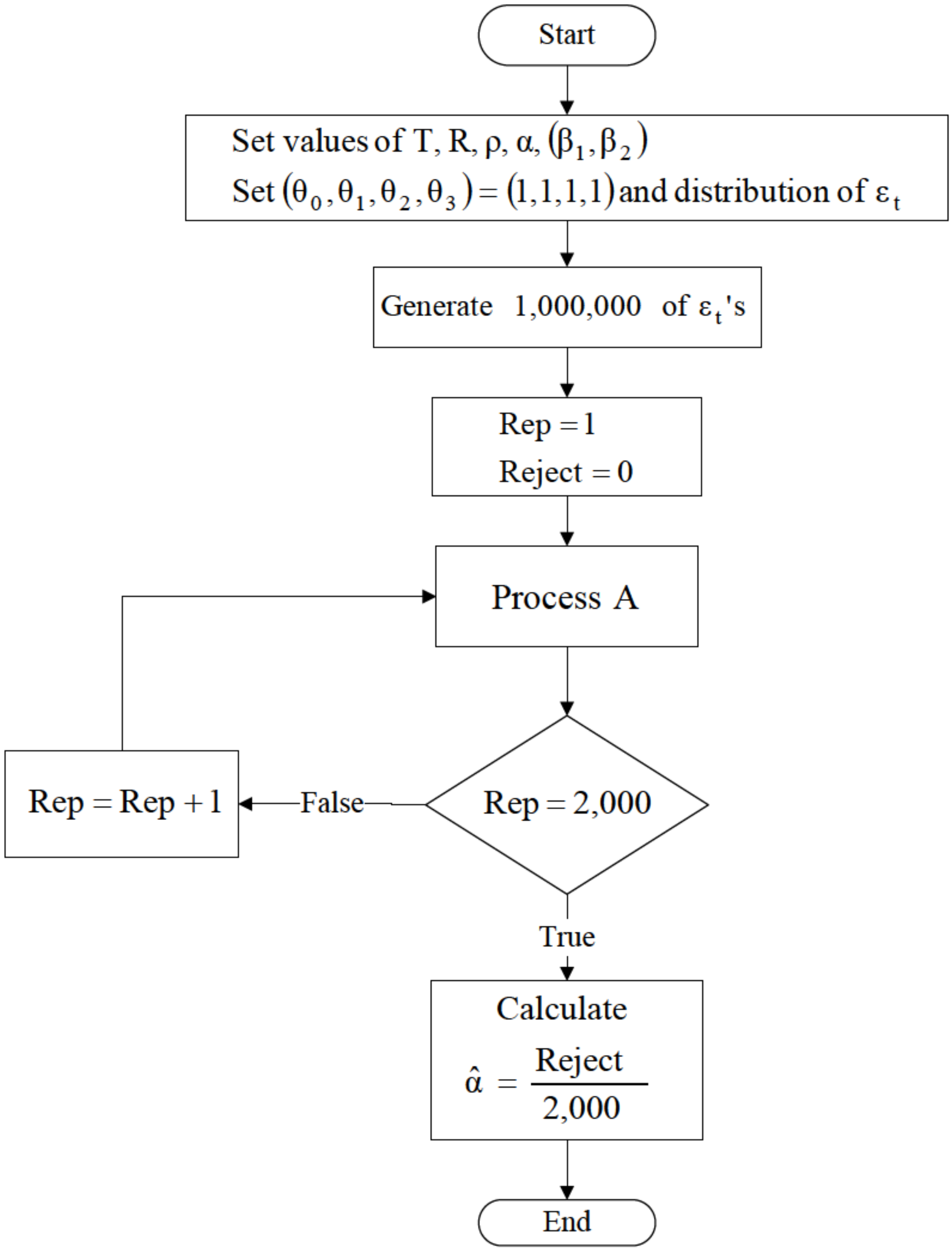
Table F.8 Empirical Powers of the Five Tests at $\alpha = 0.05^{*}$ when $\boldsymbol{\varepsilon} \sim U(-2,2)$, $(\theta_1, \theta_2, \theta_3) = (1,1,1)$ and $(\beta_1, \beta_2) = (1,1)$

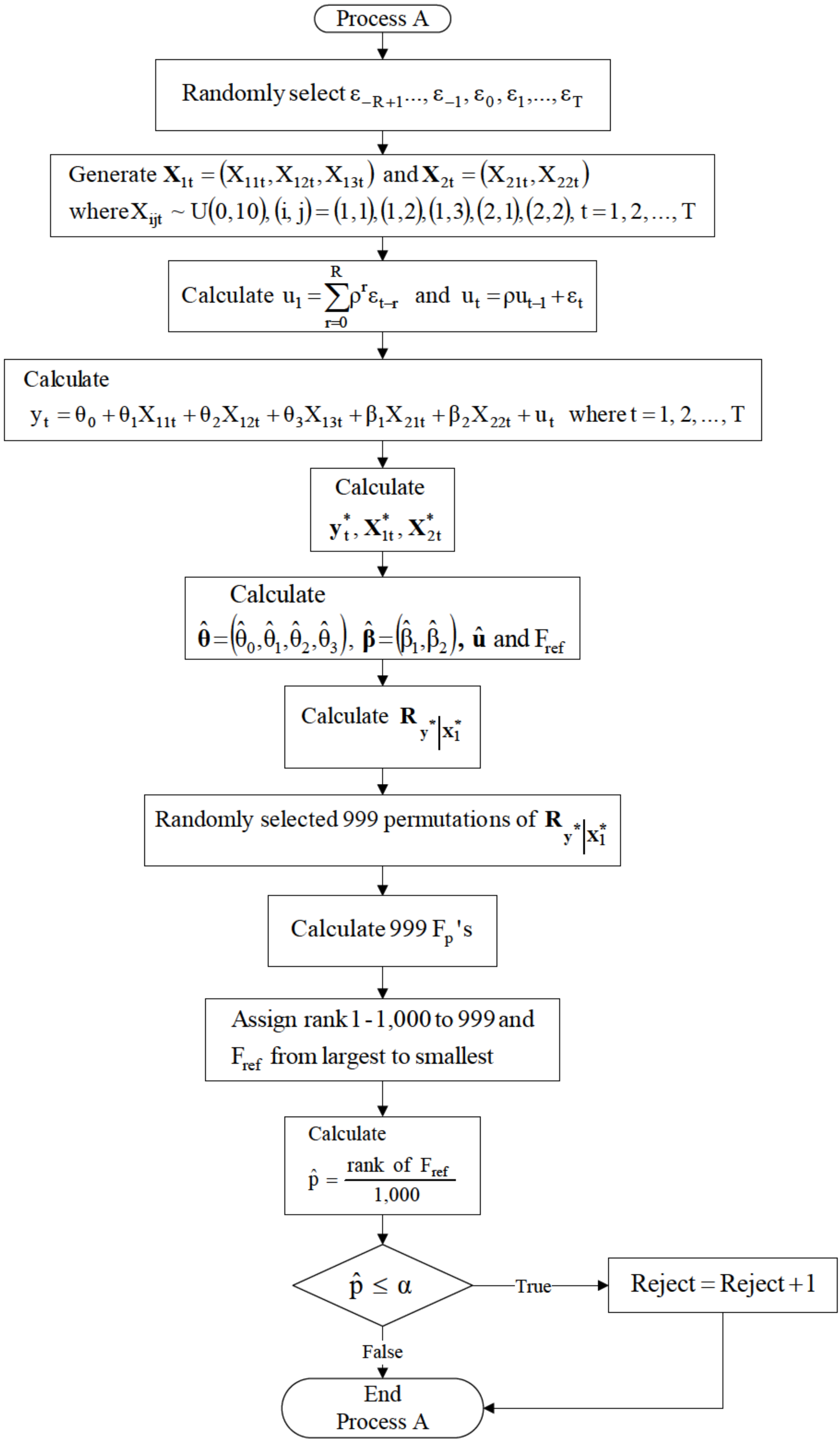
		ρ					
T	**Tests**	**-075**	**-0.5**	**-0.25**	**0.25**	**0.5**	**0.75**
	PPT	0.9090	0.9545	0.9750	0.9695	0.9650	0.9520
	FPT	0.9085	0.9530	0.9455	0.9675	0.9590	0.9495
12	MPT	0.9030	0.9325	0.9675	0.9600	0.9565	0.9450
	ACT	0.900	0.9520	0.9680	0.9665	0.9655	0.9520
	PFT	0.9085	0.9530	0.9675	0.9675	0.9595	0.9500
	PPT	0.9945	0.9958	0.9995	0.9990	0.9985	0.9955
	FPT	0.9945	0.9985	0.9995	0.9990	0.9975	0.9955
16	MPT	0.9890	0.9930	0.9950	0.9960	0.9965	0.9930
	ACT	0.9930	0.9995	0.9995	0.9995	0.9985	0.9945
	PFT	0.9945	0.9985	0.9995	0.9990	0.9980	0.9955
	PPT	0.9985	1.0000	1.0000	1.0000	1.0000	1.0000
	FPT	0.9985	1.0000	1.0000	1.0000	1.0000	1.0000
20	MPT	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	ACT	0.9885	1.0000	1.0000	1.0000	0.9960	0.9930
	PFT	0.9950	1.0000	1.0000	1.0000	0.9985	0.9970

** PPT, FPT and MPT are based on 2,000 Monte Carlo replications.*

APPENDIX G

PERMUTAION SIMULATION FLOWCHART


Start
Set values of T, R, ρ, α, (β1, β2)
Set (θ0, θ1, θ2, θ3) = (1,1,1,1) and distribution of εt
Generate 1,000,000 of εt's
Rep = 1
Reject = 0
Process A
Rep = 2,000
False
Rep = Rep + 1
True
Calculate
α̂ = Reject / 2,000
End


Process A
Randomly select $\varepsilon_{-R+1}, \ldots, \varepsilon_{-1}, \varepsilon_0, \varepsilon_1, \ldots, \varepsilon_T$
Generate $\mathbf{X}_{1t} = (X_{11t}, X_{12t}, X_{13t})$ and $\mathbf{X}_{2t} = (X_{21t}, X_{22t})$
where $X_{ijt} \sim U(0,10), (i, j) = (1,1), (1,2), (1,3), (2,1), (2,2), t = 1, 2, \ldots, T$
Calculate $u_1 = \sum_{r=0}^{R} \rho^r \varepsilon_{t-r}$ and $u_t = \rho u_{t-1} + \varepsilon_t$
Calculate
$y_t = \theta_0 + \theta_1 X_{11t} + \theta_2 X_{12t} + \theta_3 X_{13t} + \beta_1 X_{21t} + \beta_2 X_{22t} + u_t$ where $t = 1, 2, \ldots, T$
Calculate
$\mathbf{y}_t^*, \mathbf{X}_{1t}^*, \mathbf{X}_{2t}^*$
Calculate
$\hat{\boldsymbol{\theta}} = (\hat{\theta}_0, \hat{\theta}_1, \hat{\theta}_2, \hat{\theta}_3)$, $\hat{\boldsymbol{\beta}} = (\hat{\beta}_1, \hat{\beta}_2)$, $\hat{\mathbf{u}}$ and F_{ref}
Calculate $\mathbf{R}_{y^*|x_1^*}$
Randomly selected 999 permutations of $\mathbf{R}_{y^*|x_1^*}$
Calculate 999 F_p's
Assign rank 1 - 1,000 to 999 and
F_{ref} from largest to smallest
Calculate
$\hat{p} = \frac{\text{rank of } F_{ref}}{1,000}$
$\hat{p} \le \alpha$
True
Reject = Reject + 1
False
End
Process A

BIOGRAPHY

NAME	Mrs Pradthana Minsan
ACADEMIC BACKGROUND	B.Sc.(Statistics) Chiang Mai University, Thailand, 1998. M.S.(Applied Statistics), Chaing Mai University, Thailand, 2000.
PRESENT POSITION	Lecturer, Department of Mathematics and Statistics, Faculty of Science and Technology, Chiangmai Rajabhat University, Thailand.
EXPERIENCE	
2000-present	Department of Mathematics and Statistics, Faculty of Science and Technology, Chiangmai Rajabhat University, Chiang Mai, Thailand.
Publications	1) Minsan, P. and Siripanich, P. 2010. Random Permutation Test on Partial Regression Coefficient of AR(1). In **Proceedings of The 2010 International Conference on Statistical Analysis of Complex Data at Yunnan University, Kunming, China**, July 1-3, 2010.

2) Minsan, P. and Siripanich, P. 2010. Permutation Test for Partial Regression Coefficient on First-Order Autocorrelation Random Permutation Test on Partial regression. In **Proceedings of The Conference Applied Statistics 2010 at Ribno (Bled), Slovenia**, September 19-22, 2010.