Georgia State University
ScholarWorks @ Georgia State University

Mathematics Theses | Department of Mathematics and Statistics

12-1-2010

Data Driven Approaches to Testing Homogeneity of Intraclass Correlation Coefficients

Baohua Wu
Georgia State University

Follow this and additional works at: http://scholarworks.gsu.edu/math_theses

Recommended Citation
Wu, Baohua, "Data Driven Approaches to Testing Homogeneity of Intraclass Correlation Coefficients" (2010). *Mathematics Theses.* Paper 92.

This Thesis is brought to you for free and open access by the Department of Mathematics and Statistics at ScholarWorks @ Georgia State University. It has been accepted for inclusion in Mathematics Theses by an authorized administrator of ScholarWorks @ Georgia State University. For more information, please contact scholarworks@gsu.edu.

DATA DRIVEN APPROACHES TO TESTING HOMOGENEITY

OF INTRACLASS CORRELATION COEFFICIENTS

by

BAOHUA WU

Under the Direction of Dr. Yuanhui Xiao

ABSTRACT

The test of homogeneity for intraclass correlation coefficients has been one of the active topics in statistical research. Several chi-square tests have been proposed to test the homogeneity of intraclass correlations in the past few decades. The big concern for them is that these methods are seriously biased when sample sizes are not large. In this thesis, data driven approaches are proposed to testing the homogeneity of intraclass correlation coefficients of several populations. Through simulation study, data driven methods have been proved to be less biased and accurate than some commonly used chi-square tests.

INDEX WORDS: Intraclass correlation coefficient, Test of homogeneity, Data driven approach, Bootstrap, Chi-square test

DATA DRIVEN APPROACHES TO TESTING HOMOGENEITY OF INTRACLASS CORRELATION COEFFICIENTS

by

BAOHUA WU

A Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of

Master of Science

in the College of Arts and Sciences

Georgia State University

2010

Copyright by
Baohua Wu
2010

DATA DRIVEN APPROACHES TO TESTING HOMOGENEITY

OF INTRACLASS CORRELATION COEFFICIENTS

by

BAOHUA WU

Committee Chair	:	Yuanhui Xiao
Committee	:	Jiawei Liu
		Xu Zhang
		Yichuan Zhao

Electronic Version Approved:

Office of Graduate Studies

College of Arts and Sciences

Georgia State University

December 2010

ACKNOWLEDGEMENTS

I would like to express my gratitude to my advisor, Dr. Yuanhui Xiao, for his instruction, encouragement and constant support in my statistics study and thesis research. His patience and invaluable advice have proven of great value in bringing this project to conclusion.

I am also thankful to the other members of my committee, Dr. Yichuan Zhao, Dr. Xu Zhang and Dr. Jiawei Liu. Their suggestions, comments and additional guidance helped me complete my thesis writing. Also, I would like to thank Dr. Gengsheng Qin, Dr. Yu-sheng Hsu, Dr. Jun Han, and Dr. Yixin Fang for their excellent teaching that introduced me to this fantastic world of statistics.

Additionally, I am indebted to Dr. Gangli Wang from the Chemistry Department at Georgia State University for my Graduate Research Assistantship position in his lab to support my study. It was a real pleasure for me to work under Dr. Wang's supervision.

Also, I would like to thank my internship director and partner Mr. Alan Jordan in Abel Screening Inc. for his pleasant cooperation in our project. Through his help and guidance, I have gained valuable real world experience in statistics.

Finally, I would like to thank my wife for supporting me during my master degree study with love and patience.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS v

LIST OF TABLES viii

LIST OF FIGURES ix

1 INTRODUCTION 1

2 INTRACLASS CORRELATION AND TESTS OF HOMOGENEITY 3

2.1 The Model and Notations 3

2.2 Methods for Model Parameter Estimation 4

2.3 Several Existing Chi-square Tests 6

3 DATA DRIVEN APPROACHES TO TESTING HOMOGENEITY OF ICCs 10

3.1 Data Driven Approaches 10

3.2 Methods for Generating Null Samples 12

4 SIMULATION STUDY 15

4.1 Parameter Settings 15

4.2 Discussion of the Empirical Levels and the Estimated Powers of Tests 16

5 BIOLOGICAL DATA ANALYSIS 17

6 SUMMARY 19

REFERENCES 20

APPENDIX 23

A Tables of Empirical Levels of Tests at Various Stated Nominal Level α 23

B Figures of Estimated Powers of Tests for Testing H_0: $\rho_1 = \rho_2$ 27

C Tables of Estimated Powers of Tests for Testing H_0: $\rho_1 = \rho_2$ 35

LIST OF TABLES

3.1 Methods for generating null samples 14

5.1 P-values of tests of the homogeneity of ICCs for blood pressure 17

A.1 The empirical levels of test statistics for testing H_0: $\rho_1 = \rho_2$ (sample size $p_1 = p_2 = 25$) 23

A.2 The empirical levels of test statistics for testing H_0: $\rho_1 = \rho_2$ (sample size $p_1 = p_2 = 50$) 24

A.3 The empirical levels of test statistics for testing H_0: $\rho_1 = \rho_2 = \rho_3$ (sample size $p_1 = p_2 = p_3 = 50$) 25

A.4 The empirical levels of test statistics for testing H_0: $\rho_1 = \rho_2 = \rho_3$ (sample size $p_1 = p_2 = p_3 = 50$) 26

C.1 Estimated powers of LR test for testing H_0: $\rho_1 = \rho_2$ 35

C.2 Estimated powers of Fisher test for testing H_0: $\rho_1 = \rho_2$ 36

C.3 Estimated powers of WA test for testing H_0: $\rho_1 = \rho_2$ 37

C.4 Estimated powers of ACN test for testing H_0: $\rho_1 = \rho_2$ 38

C.5 Estimated powers of ACP test for testing H_0: $\rho_1 = \rho_2$ 39

C.6 Estimated powers of LCN test for testing H_0: $\rho_1 = \rho_2$ 40

C.7 Estimated powers of LCP test for testing H_0: $\rho_1 = \rho_2$ 41

LIST OF FIGURES

3.1 Diagram of data driven approaches to testing homogeneity of intraclass correlations 12

B.1 Estimated power curves of tests for testing $H_0 : \rho 1 = \rho 2$ under $p_1 = p_2 = 25$ and $\alpha = 0.01$ 27

B.2 Estimated power curves of tests for testing $H_0 : \rho 1 = \rho 2$ under $p_1 = p_2 = 25$ and $\alpha = 0.025$ 28

B.3 Estimated power curves of tests for testing $H_0 : \rho 1 = \rho 2$ under $p_1 = p_2 = 25$ and $\alpha = 0.05$ 29

B.4 Estimated power curves of tests for testing $H_0 : \rho 1 = \rho 2$ under $p_1 = p_2 = 25$ and $\alpha = 0.10$ 30

B.5 Estimated power curves of tests for testing $H_0 : \rho 1 = \rho 2$ under $p_1 = p_2 = 50$ and $\alpha = 0.01$ 31

B.6 Estimated power curves of tests for testing $H_0 : \rho 1 = \rho 2$ under $p_1 = p_2 = 50$ and $\alpha = 0.025$ 32

B.7 Estimated power curves of tests for testing $H_0 : \rho 1 = \rho 2$ under $p_1 = p_2 = 50$ and $\alpha = 0.05$ 33

B.8 Estimated power curves of tests for testing $H_0 : \rho 1 = \rho 2$ under $p_1 = p_2 = 50$ and $\alpha = 0.10$ 34

CHAPTER 1

INTRODUCTION

The intraclass correlation coefficient (ICC), ρ, is a measure of the degree of similarity among family/class members with respect to a specified characteristic. ICC has been widely applied in a variety of research areas. For instance, it is used in the theory of measurement errors for assessing the consistency or reproducibility of quantitative measurements made by different observers in measuring the same quantity in Shrout (1979). It also is a measure of the agreement between screening tests in clinical research, in Bland (1990), Kraemer (1975), and Kraemer (1981). In epidemiology studies, ρ is often used to measure the degree of familial resemblance with respect to biological characteristics such as blood pressure, cholesterol level, weight, height, and lung capacity, etc, in Donner (1985), Matthews (1984), Tian (2005), and Mian (1997).

It has been a common practice for researchers to collect data on the familial aggregation of a continuous outcome in samples from several populations or under different conditions. Inference issues concerning ρ from multiple samples arise naturally, in Xiao (2010), Hennekens (1980), Munoz (1986), Donner (1983), Nam (2003), Paul (1990a, 1990b), Tarone (1985), and Commenges (1994). A common problem is to test the homogeneity of ICCs of several populations. Assume there are K populations with ρ_i from the ith population ($i = 1, 2 \ldots K$), that is:

$$\begin{cases} H_0: & \rho_1 = \rho_2 = \ldots = \rho_K (= \rho), \\ H_a: & \text{not all } \rho_i\text{'s are equal.} \end{cases} \tag{1.1}$$

Many methods have been proposed for testing homogeneity of ICCs in the past research. Young (1998) and Bhandary (2000) proposed the likelihood ratio test and large sample z-test for two samples. Bhandary and Alam (2000) reported the likelihood ratio test and large sample ANOVA test for more than two samples. Donner and Bull (1983) derived the maximum likelihood estimator of ρ and constructed a likelihood ratio statistics for testing the assumption of a common ρ in two independent models. Among these methods, the most impressive methods are the chi-square tests presented by Mian and Shoukri (1997). These chi-square tests allow variable family sizes and do

not require the condition that samples taken from different populations share a common variance. Therefore, they are applicable to more general situations. Under mild conditions, the asymptotic distributions of the test statistics of these methods under H_0 are the chi-square distribution χ^2_{K-1} (where K is the number of populations under study). Thus, the null hypothesis H_0 is rejected at the significance level α (where $0 < \alpha < 1$) if the value of a test statistic exceeds the cut-off value, $\chi^2_{K-1,1-\alpha}$, the $100(1-\alpha)$ percentile of the chi-square distribution with *K-1* degrees of freedom. However, when the samples sizes are not so large (e.g. ≤ 50), $\chi^2_{K-1,1-\alpha}$ is usually lower than the actual value in general, resulting in too many false positives. Hence, these methods, though beautiful theoretically, produce too many false positives in practice. This problem was noticed by the authors themselves and verified through our simulation study in this thesis. This can be explained by the fact that asymptotic approximation may be poor when sample sizes are not large. The analytical expression for the null distributions is too complicated to be tractable, especially when the sample sizes vary. To avoid the challenge in searching for an analytical solution, we proposed several data driven methods to compute the cut-off value. Our simulation study shows that the proposed data driven methods are much less biased, and therefore are a significant improvement over the chi-square tests proposed by Mian and Shoukri (1997).

This thesis is organized as follows: Chapter 1 gives the introduction about ICC and the existing problems in previous researches on ICC; Chapter 2 describes the model and the widely used chi-square tests for testing the homogeneity of ICCs; Chapter 3 introduces our data driven methods for testing homogeneity of ICCs; Simulation study as well as the major findings are described in Chapter 4. In Chapter 5, a real biological data analysis using the proposed data driven methods is presented and compared with the results from the existing chi-square tests. A summary of this research is given in Chapter 6.

CHAPTER 2

INTRACLASS CORRELATION AND TESTS OF HOMOGENEITY

The test of homogeneity for ICCs has been one of the active topics in statistical research in the past a few decades. In this chapter, we first introduce the model as well as the estimation methods of the model parameters for testing the homogeneity of ICCs of several populations, then briefly describe the existing chi-square tests.

2.1 The Model and Notations

Assuming we are interested in the comparison of ICCs of K populations, we take a sample of p_i families from the ith population, with the family size of n_{ij} representing the number of members from the jth family. Suppose a characteristic variable, Y, is measured on each individual from each selected family, then a widely used model to study the ICCs with respect to Y is given by

$$\begin{aligned} Y_{ij,k} = \mu_i + \beta_{ij} + e_{ijk}, \quad & i = 1,\ 2,\ \ldots,\ K, \\ & j = 1,\ 2,\ \ldots,\ p_i, \\ & k = 1,\ 2,\ \ldots,\ n_{ij}, \end{aligned} \tag{2.1}$$

where μ_i is mean for the measured characteristic from the ith population, β_{ij} is the jth family effect from the ith population, and e_{ijk} is the individual effect. Here β_{ij} and e_{ijk} are assumed to be mutually independent and follow normal distribution $N(0,\ \sigma^2_{\beta,i})$ and $N(0,\ \sigma^2_{e,i})$, respectively. Let $\sigma^2_i = \sigma^2_{\beta,i} + \sigma^2_{e,i}$. The ICC, ρ_i, from the ith population is defined as

$$\rho_i = \frac{\sigma^2_{\beta,i}}{\sigma^2_i} = \frac{\sigma^2_{\beta,i}}{\sigma^2_{\beta,i} + \sigma^2_{e,i}}. \tag{2.2}$$

Under $H_0 : \rho_1 = \rho_2 = \ldots = \rho_K (= \rho)$, the ICCs are equal to one another, and the common value ρ is called the common ICC.

2.2 Methods for Model Parameter Estimation

The estimation theory has been well developed by several authors for a single population as well multiple populations. Two commonly used methods for estimating the parameters (μ_i, σ_i^2, ρ_i) in the above model are ANOVA, in Donner (1985) and Maximum Likelihood (ML), in Smith (1980), Elston (1975), and Donner (1980). For both methods, the summary statistics (n_{ij}, $\bar{Y}_{ij}$, S_{ij}^2) play the important roles, where

$$\bar{Y}_{ij} = \frac{1}{n_{ij}} \sum_{k=1}^{n_{ij}} Y_{ij,k}, \qquad S_{ij}^2 = \sum_{k=1}^{n_{ij}} (Y_{ij,k} - \bar{Y}_{ij})^2. \tag{2.3}$$

2.2.1 ANOVA Estimation

In 2.4 - 2.7 we define the notations. For the ith population with p_i families and n_{ij} members in each family, let

$$n_i = \sum_{j=1}^{p_i} n_{ij}, \qquad \bar{Y}_i = \frac{1}{n_i} \sum_{j=1}^{p_i} n_{ij} \bar{Y}_{ij}, \tag{2.4}$$

$$\lambda_i' = \frac{1}{p_i - 1} \left[n_i - \frac{1}{n_i} \sum_{j=1}^{p_i} n_{ij}^2 \right], \qquad \lambda_i'' = \sum_{j=1}^{p_i} n_{ij}^2 - \frac{2}{n_i} \sum_{j=1}^{p_i} n_{ij}^3 + \frac{1}{n_i^2} \left[\sum_{j=1}^{p_i} n_{ij}^2 \right]^2, \tag{2.5}$$

$$\mathrm{SSG}_i = \sum_{j=1}^{p_i} n_{ij} (\bar{Y}_{ij} - \bar{Y}_i)^2, \qquad \mathrm{MSG}_i = \frac{\mathrm{SSG}_i}{p_i - 1}, \tag{2.6}$$

$$\mathrm{SSW}_i = \sum_{j=1}^{p_i} \sum_{k=1}^{n_{ij}} (Y_{ij,k} - \bar{Y}_{ij})^2, \qquad \mathrm{MSW}_i = \frac{\mathrm{SSW}_i}{n_i - p_i}. \tag{2.7}$$

As in the single population case from Ruth, Dunn and Clark's previous study (2004), the ANOVA estimators of the parameters (μ_i, σ_i^2, ρ_i) for the ith population are listed as below,

respectively,

$$\tilde{\mu}_i = \bar{Y}_i, \tag{2.8}$$

$$\tilde{\sigma}_i = \frac{\mathrm{MSG}_i + (\lambda_i' - 1)\mathrm{MSW}_i}{\lambda_i'}, \tag{2.9}$$

$$\tilde{\rho}_i = \frac{\mathrm{MSG}_i - \mathrm{MSW}_i}{\mathrm{MSG}_i + (\lambda_i' - 1)\mathrm{MSW}_i}. \tag{2.10}$$

The large sample variance for the ANOVA estimator of ICC derived from Smith's results (1957) is

$$\mathrm{var}(\tilde{\rho}_i) = \frac{2(1-\rho_i)^2}{\lambda_i'^2}\left\{\frac{[1+(\lambda_i'-1)\rho_i]^2}{n_i - p_i} + \frac{(1-\rho_i)[1+(2\lambda_i'-1)\rho_i]^2}{p_i - 1} + \frac{\lambda_i''\rho_i^2}{(p_i-1)^2}\right\}. \tag{2.11}$$

Under the assumption of homogeneity of ICCs, the pooled estimate for the common ICCs ρ is given by

$$\tilde{\rho} = \frac{\sum_i^K \tilde{\rho}_i / \widetilde{\mathrm{var}}(\tilde{\rho}_i)}{\sum_i^K 1/\widetilde{\mathrm{var}}(\tilde{\rho}_i)}, \tag{2.12}$$

where $\widetilde{\mathrm{var}}(\tilde{\rho}_i)$ is an estimate of $\mathrm{var}(\tilde{\rho}_i)$ obtained by plugging-in the estimate $\tilde{\rho}_i$ of ρ_i into the Equation 2.11.

2.2.2 Maximum Likelihood Estimation (MLE)

For the ith population under the multivariate normal model, in Mian (1997), the likelihood equation with respect to ρ_i is

$$\sum_{j=1}^{p_i}\left\{\frac{n_{ij}-1}{1-\rho_i} - v_{ij}(\rho_i) - \frac{1}{\sigma_i^2}\left[\frac{S_i^2}{(1-\rho_i)^2} - u_{ij}(\rho_i)v_{ij}(\rho_i)(\bar{Y}_{ij} - \mu_i)^2\right]\right\} = 0, \tag{2.13}$$

where

$$u_{ij}(\rho) = \frac{n_{ij}}{1+(n_{ij}-1)\rho}, \qquad v_{ij}(\rho) = \frac{n_{ij}-1}{1+(n_{ij}-1)\rho}. \tag{2.14}$$

Under the assumption of homogeneity of ICCs, the maximum likelihood equation is

$$\sum_{i=1}^{K}\sum_{j=1}^{p_i}\left\{\frac{n_{ij}-1}{1-\rho}-v_{ij}(\rho)-\frac{1}{\sigma_i^2}\left[\frac{S_i^2}{(1-\rho)^2}-u_{ij}(\rho)v_{ij}(\rho)(\bar{Y}_{ij}-\mu_i)^2\right]\right\}=0. \tag{2.15}$$

Consequently the maximum likelihood estimators of μ_i and σ_i^2 are,

$$\hat{\mu}_i=\hat{\mu}_i(\rho_i)=\sum_{j=1}^{p_i}\frac{u_{ij}(\rho_i)\bar{Y}_{ij}}{\sum_{j=1}^{p_i}u_{ij}(\rho_i)}, \tag{2.16}$$

$$\hat{\sigma}_i^2=\hat{\sigma}_i^2(\rho_i)=\frac{1}{n_i}\left[\frac{\mathrm{SSW}_i}{1-\rho_i}+\sum_{j=1}^{p_i}u_{ij}(\rho_i)(\bar{Y}_{ij}-\hat{\mu}_i(\rho_i))^2\right], \tag{2.17}$$

respectively. Both estimators are functions of ρ under the null hypothesis $H_0:\rho_1=\rho_2=\ldots=\rho_K(=\rho)$ or ρ_i under H_a. Substituting the nuisance parameters μ_i and σ_i^2 in Equation 2.13 or 2.15 with the respective maximum likelihood estimators will result in a non-linear equation about the single parameter ρ_i (or ρ). Thus, the maximum likelihood estimator $\hat{\rho}$ of ρ can be computed by a safe-guard method such as the bisection method or Brent's method. The readers are referred to Mian *et al.* (1997) for the detailed description of the two estimation methods.

2.3 Several Existing Chi-square Tests

Given samples from different normal populations, we would like to test the homogeneity of ICCs of several samples with varied family sizes under the hypothesis:

$$\begin{cases} H_0: & \rho_1=\rho_2=\ldots=\rho_K(=\rho), \\ H_a: & \text{not all } \rho_i\text{'s are equal.} \end{cases} \tag{2.18}$$

In this section, we introduce three of Chi-square tests: likelihood ratio test (**LR** for short), the chi-square test derived from Fisher's variance stabilizing transformation of intraclass correlations (**Fisher** for short), in Sadanori (1989), and an *ad-hoc* chi-square test based on ANOVA estimation (**WA** for short), in Mian (1997).

2.3.1 The Likelihood Ratio Test

The log-likelihood function under the null hypothesis $H_0 : \rho_1 = \rho_2 = \ldots = \rho_K (= \rho)$ is given by (See Equation 2.15):

$$\begin{aligned} -2 \cdot lc &= C + \sum_{i=1}^{K} \left\{ n_i \ln(\sigma_i^2) + (n_i - p_i) \ln(1-\rho) + \sum_{j=1}^{p_i} \ln[1 + (n_{ij} - 1)\rho] \right. \\ &\quad \left. + \frac{1}{\sigma_i^2} \left[\frac{SSW_i}{1-\rho} + \sum_{j=1}^{p_i} w_{ij}(\rho)(\bar{Y}_{ij} - \mu_i^2) \right] \right\}. \end{aligned} \tag{2.19}$$

The log-likelihood function under the alternative hypothesis is given by (See Equation 2.13):

$$\begin{aligned} -2 \cdot l &= C + \sum_{i=1}^{K} \left\{ n_i \ln(\sigma_i^2) + +(n_i - p_i) \ln(1-\rho_i) + \sum_{j=1}^{p_i} \ln[1 + (n_{ij} - 1)\rho_i] \right. \\ &\quad \left. + \frac{1}{\sigma_i^2} \left[\frac{SSW_i}{1-\rho_i} + \sum_{j=1}^{p_i} w_{ij}(\rho_i)(\bar{Y}_{ij} - \mu_i^2) \right] \right\}. \end{aligned} \tag{2.20}$$

Suppose the maximum values of lc and l in Equations 2.19 and 2.20 are l_1 and l_0, respectively, then the likelihood ratio test statistic is given by

$$\chi_l^2 = -2(l_0 - l_1) = \sum_{i=1}^{K} \left\{ n_i \ln \left[\frac{\hat{\sigma}_i^2(\hat{\rho})}{\hat{\sigma}_i^2(\hat{\rho}_i)} \right] + (n_i - p_i) \ln \left[\frac{1-\hat{\rho}}{1-\hat{\rho}_i} \right] + \sum_{j=1}^{p_i} \ln \left[\frac{1 + (n_{ij} - 1)\rho}{1 + (n_{ij} - 1)\rho_i} \right] \right\}, \tag{2.21}$$

which is asymptotically distributed as the chi-square with *K-1* degrees of freedom under the null hypothesis H_0.

2.3.2 The Chi-square Test Derived from Fisher's z-Transformation

For a random sample of size n_{ij} from p-variate normally distributed populations, Fisher's z-transformation is given by

$$z(r_i) = \left(\frac{p-1}{2p}\right)^{1/2} \log\left[\frac{1+(p-1)r_i}{1-r_i}\right], \tag{2.22}$$

where r_i is the maximum likelihood estimator of ρ_i.

One of the advantages of Fisher's z-transformation is that the asymptotic variance of $z(r_i)$ is independent of the unknown parameters. The disadvantage is that the normal approximation could be poorer when p increases. This drawback could be overcome through approximating $z(r_i)$ by a normal distribution with the mean and variance given by

$$z(\rho_i) + n_{ij}^{-1}\frac{7-5p}{[18p(p-1)]^{1/2}} \quad \textit{and} \quad (n_{ij}-2)^{-1}, \tag{2.23}$$

respectively. See Konishi (1985). Here n_{ij} is the size from j^{th} family in the ith population. Based on the above equation, the bias could be corrected as follows:

$$z^*(r_i) = z(r_i) - n_{ij}^{-1}\frac{7-5p}{[18p(p-1)]^{1/2}}. \tag{2.24}$$

The estimate of the common mean value z(ρ) is :

$$z^* = \sum_{i=1}^{K}(n_{ij}-2)z^*(r_i) / \sum_{i=1}^{K}(n_{ij}-2). \tag{2.25}$$

To test the homogeneity of ICCs, the null hypothesis $H_0 : \rho_1 = \rho_2 = \ldots = \rho_K (= \rho)$ is equivalent to:

$$H_0 : z(r_1) = z(r_2) = \ldots = z(r_K)(= z(r)), \tag{2.26}$$

and the test statistic can be obtained as follows,

$$\sum_{i=1}^{K}(n_{ij}-2)\left[z^*(r_i)-z^*\right]^2, \tag{2.27}$$

which has the asymptotical chi-square distribution with *K-1* degrees of freedom under the null hypothesis H_0.

2.3.3 An *ad-hoc* Chi-square Test

An *ad-hoc* chi-square statistic based on ANOVA estimation of the parameters was built by Mian (1997) as follows :

$$\chi^2_{wa}=\sum_{i}^{K}\tilde{w}_i(\tilde{\rho}_i-\bar{\tilde{\rho}}_w)^2, \tag{2.28}$$

where

$$\bar{\tilde{\rho}}_w=\sum_{i}^{K}\tilde{w}_i\tilde{\rho}_i/\sum_{i}^{K}\tilde{w}_i, \tag{2.29}$$

and

$$\tilde{w}_i=\left[var(\tilde{\rho}_i)|_{\rho_i=\tilde{\rho}}\right]^{-1}. \tag{2.30}$$

The test statistic χ^2_{wa} in the Equation 2.28 is asymptotically chi-square distributed with *K-1* degrees of freedom under the null hypothesis H_0.

CHAPTER 3

DATA DRIVEN APPROACHES TO TESTING HOMOGENEITY OF ICCs

As mentioned in Chapter 1, the Chi-square tests for testing the homogeneity of ICCs of several populations are seriously biased for moderately large samples (sample sizes $\approx 25 \sim 50$ families) and produce more false positives. To reduce the number of false positives, we propose the data driven approaches to testing homogeneity of intraclass correlations in this chapter.

3.1 Data Driven Approaches

In our data driven approaches, the first step is to estimate the model parameters. The parameters μ_i, σ_i and ρ_i for the ith population may be estimated by either ANOVA or ML, which are introduced in Chapter 2. For each method, the data from different populations can be used either separately or combinedly under the assumption of homogeneity of intraclass correlations. Depending on how the data are used, a bootstrap method will be called either separate or combined. The parameters μ_i, σ_i^2, and ρ_i for the ith population may be estimated by using the data from the families of the population alone without taking into account of the data of families from other population(s). This is termed the separate estimation. When the ANOVA approach is used, the separate estimates of the parameters for the ith population are computed by Equation 2.8, 2.9, and 2.10, respectively. When ML approach is used, the separate estimates of μ_i, σ_i^2, and ρ_i are obtained by solving the Equation 2.13, 2.16 and 2.17 jointly.

The estimation of the parameters μ_i, σ_i^2, and ρ_i for the ith population may also use all the data (including data of families from other populations). This would be termed the combined estimation. For ANOVA approach, the combined estimates of μ_i, σ_i^2 remain the same, and the combined estimate of the common ρ is given by Equation 2.12. While for ML approach, the combined estimation of the parameters is made by solving Equation 2.15 and the $2K$ equations of (2.16) and (2.17) (i = 1, 2... K) jointly.

In our data driven approaches proposed in this thesis, only the combined estimation was used.

The second step is to create null samples under H_0. One direct resampling method is to simply select p_i families from the ith population at random with replacement, then use the data of the selected families to compose a bootstrap sample for the ith population. This method doesn't need to estimate the parameters, but the generated null samples may have slightly different probability distributions since the set of family sizes differs from sample to sample. For the resampling methods used in our data driven approaches, we keep the sizes of families in the original sample, but simulate data for family members. All the null samples have the same set of family sizes and distribute equally, which avoid the drawback from the direct resampling method. This resampling method is achieved by using parameter estimates (See details in Section 3.2).

In the third step, the likelihood ratio statistics are computed for both the original sample (LR_0) and each bootstrap sample (LR^*). To test the null hypothesis ($H_0 : \rho_1 = \rho_2 = \ldots = \rho_K (= \rho)$), it is critical to determine the cut-off value. If the test statistic from the original sample is larger than the cut-off value, we reject the null hypothesis H_0; otherwise, we accept H_0. In chi-square tests mentioned in Section 2.3, $\chi^2_{K-1,1-\alpha}$ was taken as the cut-off value (α is the given nominal level and K is the degree of freedom). When sample size is not large, $\chi^2_{K-1,1-\alpha}$ is lower than the actual value in general, which could bring too many false positives. However, in our proposed data driven approaches, the $100(1-\alpha)$ percentile of the test statistics ($\{TS_i^*\}_{i=1}^m$) from the bootstrap samples is selected as the cut-off value, which is certainly more accurate than $\chi^2_{K-1,1-\alpha}$ (Here, we use TS and TS^*s rather than LR_0 and LR^*s to denote the test statistics from the original sample and bootstrap samples, respectively).

For our data driven approaches, the p-value of the test is the percent of values of test statistics from the bootstrap samples which are not less than that from the original sample ($\{TS_i^*\}_{i=1}^m \geqslant TS$). Obviously, when the estimated p-value is less than or equal to α, we reject H_0.

The whole process of the data driven approaches are illustrated with components connected by the solid line in Figure 3.1.

In Figure 3.1, for each simulation run, there are m null samples generated by using bootstrap technique with the parameter settings (μ, σ, ρ, and n_{ij}). After n simulation runs, the empirical levels and estimated powers of tests are computed as the rejection rates at the stated nominal

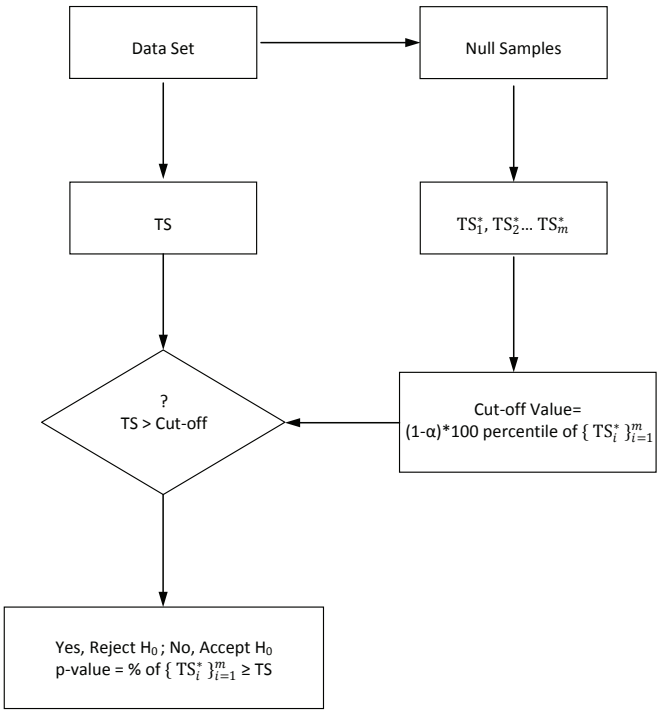

Figure 3.1: Diagram of data driven approaches to testing homogeneity of intraclass correlations level α.

3.2 Methods for Generating Null Samples

Here we describe how to generate the null samples. The covariance matrix for the measurements $Y_{ij} = (Y_{ij,1},\ Y_{ij,2},\ \ldots,\ Y_{ij,n_{ij}})^T$ of the jth family from the ith population is given by

$$\mathbf{\Sigma}_{ij} \equiv \mathbf{\Sigma}_{ij}(n_{ij},\ \rho_i,\ \sigma_i^2) = \sigma_i^2[(1-\rho_i)\mathbf{I}_{n_{ij}} + \rho_i \mathbf{J}_{n_{ij}}] = \sigma_i^2 \begin{pmatrix} 1 & \rho_i & \rho_i & \cdots & \rho_i \\ \rho_i & 1 & \rho_i & \cdots & \rho_i \\ \rho_i & \rho_i & 1 & \cdots & \rho_i \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho_i & \rho_i & \rho_i & \cdots & 1 \end{pmatrix}, \tag{3.1}$$

where $\mathbf{I}_{n_{ij}}$(use $\mathbf{I}_n$ later on) is the $n_{ij} \times n_{ij}$ identity matrix and $\mathbf{J}_{n_{ij}}(\mathbf{J}_n)$ is the $n_{ij} \times n_{ij}$ matrix whose entries are all one. To make the covariance matrix be positive definite, the condition of

$\rho_i > -1/n_{ij}$ must be met.

Let $\mathbf{A}_{ij}$ (or $\mathbf{A}$ in case of no confusion) be a $n_{ij} \times n_{ij}$ matrix such that $\mathbf{A}\mathbf{A}^T = \mathbf{\Sigma}_{ij}$. There are several options for the matrix $\mathbf{A}$. For instance, the Choleskey decomposition for positively definite matrices will give a lower triangular form of $\mathbf{A}$. Here we prefer using the matrix $\mathbf{A}$ as follows:

$$\mathbf{A} = \sigma_i[\alpha \mathbf{I}_n + \beta \mathbf{1}_n \mathbf{1}_n^T], \tag{3.2}$$

where

$$\alpha = \sqrt{1-\rho_i},$$
$$\beta = \frac{\sqrt{1+(n_{ij}-1)\rho_i} - \sqrt{1-\rho_i}}{n_{ij}} = \frac{\rho_i}{\sqrt{1-\rho_i} + \sqrt{1+(n_{ij}-1)\rho_i}}.$$

Now suppose $Z_{ij} = (Z_{ij,1},\ Z_{ij,2},\ \ldots,\ Z_{ij,n_{ij}})^T$ to be a random vector whose components $Z_{ij,k}$ are independent and distributed as the standard normal distribution $N(0,1)$. The random vector

$$Y_{ij}^* = (Y_{ij,1}^*,\ Y_{ij,2}^*,\ \ldots,\ Y_{ij,n_{ij}}^*)^T :\ Y_{ij,k}^* = \mu_i + \mathbf{A}Z_{ij},\ k = 1,\ 2,\ \ldots,\ n_{ij}, \tag{3.3}$$

would follow the model (2.1). This fact gives us the clue to generate null samples for testing H_0. Suppose ρ^*, μ_i^* and σ_i^* are the estimates of the respective parameters ρ (ρ_i), μ_i and σ_i under H_0. An approximation $\mathbf{A}^*$ to the matrix $\mathbf{A}$ is computed via Equation 3.2 by using these estimates. An approximate null sample can be generated by using the parameter estimates. The data Y_{ij}^* for each individual in a family may be simulated by generating the random vector Z_{ij} first, then computing Y_{ij}^* using Equation 3.3 with $\mathbf{A}$ replaced by $\mathbf{A}^*$. The random vector Z_{ij} can be obtained by simply simulating a series of random numbers from the standard normal distribution $N(0,1)$. This method is called the parametric method.

A non-parametric method is described as follows. Let $\mathbf{1}_{n_{ij}}$ be a $n_{ij} \times 1$ vector. The vector

$$Z_{ij} = \mathbf{A}^{-1}(Y_{ij} - \mu_i \mathbf{1}_{n_{ij}}) \tag{3.4}$$

has independent components with standard normal distribution $N(0,1)$. In practice, we may use parameter estimates to compute Z_{ij}, which are denoted by $\hat{Z}_{ij}$ if the parametric estimates

rather than true values of the parameters are used. Let $\mathbb{Z}_i$ be the collection of all components of $\hat{Z_{ij}}$ $(j = 1,\ 2,\ \ldots,\ p_i)$, then $\mathbb{Z} = \bigcup_{i=1}^{K} \mathbb{Z}_i$ is a set of random numbers which are nearly from the standard normal distribution $N(0,\ 1)$. The random vector Z_{ij} in Equation 3.3 is created by selecting n_{ij} numbers from the set $\mathbb{Z}$ at random with replacement.

Based on using either the parametric or the non-parametric resampling methods to generate the Z_{ij} (use Z in case of no confusion later on) vectors and either the ANOVA or the ML method to estimate the model parameters, there are four combined methods in total for creating null samples, which are summarized in Table 3.1.

Table 3.1: Methods for generating null samples

		Code
Parametric	ANOVA	ACP
	MLE	LCP
Non-parametric	ANOVA	ACN
	MLE	LCN

In this table, we assign a code for each data driven method. For example, ACP corresponds to the data driven approach in which we use **A**NOVA estimation, **C**ombined bootstrap technique, and **P**arametric resampling method to generate the Z vectors. LCN corresponds to the data driven approach in which we use M**L**E, **C**ombined bootstrap technique, and the **N**on-parametric resampling method to generate the Z vectors. These assigned codes are used in the text that follows.

CHAPTER 4

SIMULATION STUDY

In order to compare our data driven approaches with those existing chi-square tests, simulation study is conducted in this research. In this chapter, we present and analyze the results from the simulation study using the data driven approaches proposed in this thesis and the existing chi-square tests (LR, Fisher and WA).

4.1 Parameter Settings

In our simulation study, these model parameters are given with $\mu = 0$, $\sigma = 1$ and ρ in the range from 0.1 to 0.9 with the increment of 0.1; the number of populations is K= 2 or 3 with $p_1 = p_2 = \ldots = p_K$=25 or 50 families from each population. The family sizes (n_{ij}) are generated from the zero-truncated negative binomial distribution with the following probability mass function:

$$\text{Prob}(d) = \frac{\Gamma(m+d)}{d!\Gamma(m)} \left(\frac{P}{1+P}\right)^d [(1+P)^m - 1]^{-1}, \quad d = 1,\ 2,\ \ldots, \tag{4.1}$$

where m and P are two parameters. This distribution has good performance in the fitting of sibship size data in a variety of human populations. Here, in order to compare with the published results, we set the values $m = 2.84$ and $P = 0.93$ in our simulation study, which correspond to mean sibship size of 3.129 and a variance of 4.5201 as reported by Brass (1958). To ensure that a generated family size is at least two, an integer generated according to Equation 4.1 is increased by one. The resulting integer is truncated to 15 when it is greater than 15. The final integers thus obtained are used for each family size. Similar parameter settings have been used by Mian (1997), Tian (2005) in their simulation studies.

4.2 Discussion of the Empirical Levels and the Estimated Powers of Tests

In each simulation run, a data set as well as 5000 null samples is generated for the hypothesis testing. For each method, 5000 simulation runs have been done under each given parameter settings. In Appendix A, the empirical levels of tests at various stated nominal level α are listed in Table A.1 - A.4. The estimated power curves are shown in Figure B.1 - B.8 in Appendix B, and the corresponding values at different stated nominal levels α for each method are listed in Table C.1 - C.7 in Appendix C. Since a value of ρ beyond 0.6 rarely happens in the real world, here the estimated powers are given only for ρ changing from 0.1 to 0.6 with the increment of 0.1.

Based on the simulation study, the major findings by comparing the results from each method are summarized as follows.

1. The empirical levels from LR, Fisher and WA tests are far above the nominal level α in general; while the empirical levels from data driven methods are quite close to the stated nominal levels. Hence, LR, Fisher and WA tests are seriously biased compared to the data driven methods.

2. The empirical levels from the data driven methods have no significant difference concerning either ANOVA estimation or MLE for the model parameters, and either parametric or nonparametric method to generate the null samples.

3. When sample sizes increase from 25 to 50, the empirical levels from LR, Fisher and WA tests move toward the stated nominal level α, but are still generally worse than the results from the data driven methods.

4. The findings above are similar under either the two population case or the three population case.

CHAPTER 5

BIOLOGICAL DATA ANALYSIS

Here we use the published arterial blood pressure data collected by Miall and Oldham (1955) to illustrate our data driven methodology. The population studied is representative of the general population of a South Wales mining valley. In the past, this valley provided for the men few opportunities for occupations other than coalmining, and consequently the great majority of men in the older age groups are either miners or ex-miners. This data set was used by many researchers for different purposes, in Donner (1980) and Mian (1991). In our study, we adapted the same setting as Tian (2005). The data set is divided into 4 groups with 218 families and 1160 members in total. Each groups have sample sizes of 62, 36, 75 and 45, respectively. The first and second groups belong to female group including 98 families and 531 members, and the third and last groups belong to male group including 120 families and 629 members. The family sizes range from 2 to 12. For each family member, the data of high pressure, low pressure and measurement of arm girth are recorded. The computed p-values are listed in Table 5.1.

Table 5.1: P-values of tests of the homogeneity of ICCs for blood pressure

Method	Low Pressure			High Pressure		
	Age	Female	Male	Age	Female	Male
LR	0.2454	0.0922	0.3115	0.0094	0.0293	0.0974
Fisher	0.4234	0.1621	0.5583	0.0000	0.0187	0.0001
WA	0.2375	0.0718	0.6576	0.0000	0.0554	0.0002
ACN	0.2575	0.1017	0.3050	0.0120	0.0330	0.0919
ACP	0.2595	0.0966	0.3156	0.0114	0.0333	0.1095
LCN	0.2585	0.1060	0.3136	0.0143	0.0325	0.0943
LCP	0.2655	0.1000	0.3265	0.0129	0.0335	0.0995

As shown in Table 5.1, the data driven methods are more likely to give the similar p-values when compared to LR, Fisher and WA tests. The conclusions from the analysis of the real data set are that, the ICCs have no significant difference on testing the homogeneity for low blood pressure from the male groups or from the age groups; but for high blood pressure, the ICCs are significantly different among the female groups or the age groups. Since the p-values for testing the homogeneity of ICCs with respect to the low blood pressure from the different female groups

are quite close to 0.10, we can not arrive at a final conclusion until further research is conducted. Similar results were obtained when we test the homogeneity of the high blood pressure from the male groups.

CHAPTER 6

SUMMARY

In this thesis, we proposed several data driven approaches to testing the homogeneity of ICCs of several populations. Through simulation study, the data driven methods have been proved to be more accurate than the existing chi-square tests when testing the homogeneity of ICCs. Compared to the commonly used chi-square tests (such as LR, Fisher and WA), the data driven methods are much less biased. The empirical levels as well as the p-values of the tests from each data driven method are quite close to each other. When the samples sizes in each populations are not very large, these asymptotical chi-square tests are less precise compared to the data driven methods.

Normally, ANOVA estimation can be implemented more easily than MLE. In our simulation study, the four data driven methods are slightly different concerning bias. Based on the simulation results, ACN is recommended for use in testing homogeneity of ICCs.

REFERENCES

[1] Bhandary, M. and Alam, M.K., (2000). Test for the equality of intraclass correlation coefficients under unequal family sizes for several populations, *Communications in Statistics-Theory and Methods* **29**: 755–768.

[2] Bland, J.M. and Altman, D.G., (1990). A note on the use of the intraclass correlation-coefficient in the evaluation of agreement between 2 methods of measurement, *Computers in Biology and Medicine* **20**: 337–340.

[3] Brass, W., (1958). Simplified methods of fitting the truncated negative binomial distribtuion, *Biometrika* **45**: 59–C68.

[4] Commenges, D. and Jacquin, H., (1994). The intraclass correlation-coefficient - distribution-free definition and test, *Biometrics* **50**: 517–526.

[5] Donner, A., (1985). The analysis of intraclass correlation in multiple samples, *Annals of Human Genetics* **49**: 75–82.

[6] Donner, A. and Bull, S., (1983). Inferences concerning a common intraclass correlation-coefficient, *Biometrics* **39**: 771–775.

[7] Donner, A. and Koval, J.J., (1980). The estimation of intraclass correlation in the analysis of family data, *Biometrics* **36**: 19–25.

[8] Elston, R.C., (1975). Correlations between correlations, *Biometrika* **62**: 163–168.

[9] Hennekens, C.H. and Levine, R.S. and Rosner, B. and Klein, B.E. and Gourley, J.E. and Gelband, H. and Jesse, M.J., (1980). Aggregation of cholesterol among young families of men with premature myocardial infarction, *Journal of Chronic Diseases* **33**: 359–364.

[10] Konishi, S., (1985). Normalizing and variance stabilizing transformations for intraclass correlations, *Ann. Inst. Statist. Math.* **37**: 87–94.

[11] Kraemer, H.C., (1975). On estimation and hypothesis testing problems for correlation coefficients, *Psychometrik* **40**: 473–485.

[12] Kraemer, H.C., (1981). Extension of feldt's approach to testing homogeneity of coefficients of reliability, *Psychometrik* **46**: 41–45.

[13] Matthews, K.A. and Rosenman, R.H. and Dembroski, T.M. and Harris, E.L. and Macdougall, J.M., (1984).Familial resemblance in components of the type a behavior pattern: A reanalysis of the california type a twin study, *Psychosomatic Medicine* **46**: 512–522.

[14] Miall, W.W. and Oldham, P.D., (1955). A study of arterial blood pressure and its inheritance in a sample of the general population, *Clinical Science* **14**:459–487.

[15] Mian, I.U.H. and Shoukri, M.M., (1997). Statistical analysis of intraclass correlations from multiple samples with applications to arterial blood pressure data, *Statistics in Medicine* **16**: 1497–1514.

[16] Mian, I.U.H. and Shoukri, M.M and Tray, D.D., (1991). Maximum likelihood estimation and sibling correlations in the analysis of family data, *Journal of Statistical Planning and Inference* **27**: 125–141.

[17] Munoz, A. and Rosner B., and Carey, V., (1986). Regression analysis in the presence of heterogeneous intraclass correlations, *Biometrics* **42**: 653–658.

[18] Nam, J., (2003). Homogeneity score test for the intraclass version of the kappa statistics and sample-size determination in multiple or stratified studies, *Biometrics* **59**: 1027–1035.

[19] Paul, S.R., (1990a). Maximum likelihood estimation of intraclass correlation in the analysis of familial data: Estimating equation approach, *Biometrika* **77**: 549–555.

[20] Paul, S.R., and Barnwal, R.K., (1990b). Maximum likelihood estimation and a c(α) test for a common intraclass correlation, *The Statistician* **39**: 19–24.

[21] Sadanori, K. and Arjun, K.G., (1989). Testing the equality of several intraclass correlation coefficients, *Journal of Statistical Planning and Inference* **21**: 93–105.

[22] Shrout, P.E. and Fleiss, J.L., (1979). Intraclass correlations: Uses in assessing rater reliability, *Psychological Bulletin* **86**: 420–428.

[23] Smith, C.A.B., (1957). On the estimation of inctraclass correlation, *Annals of Human Genetics* **21**: 363–373.

[24] Smith, C.A.B., (1980). Estimating genetic correlations, *Annals of Human Genetics* **44**: 265–284.

[25] Tarone, R.E., (1985). On heterogeneity tests based on efficient scores, *Biometrika* **72**: 91–95.

[26] Tian, L., (2005). On confidence intervals of a common intraclass correlation coefficient, *Statistics in Medicine* **24**: 3311–3318.

[27] Xiao, Y. and Liu, J. and Bhandary, M., (2010). Profile likelihood based confidence intervals for common intraclass correlation coefficient, *Communications in Statistics - Simulation and Computation* **39**: 111–118.

[28] Young, D.J. and Bhandary, M., (1998). Test for equality of intraclass correlation coefficients under unequal family sizes, *Biometrics* **54**: 1363–1373.

Appendix A

Tables of Empirical Levels of Tests at Various Stated Nominal Level α

Table A.1: The empirical levels of test statistics for testing H_0: $\rho_1 = \rho_2$ (sample size $p_1 = p_2 = 25$)

ρ	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
Method	$\alpha = 0.01$								
LR	0.0162*	0.0148*	0.0138*	0.0144*	0.0096	0.0138*	0.0118	0.0098	0.0124
Fisher	0.0112	0.0156*	0.0168*	0.0160*	0.0180*	0.0206*	0.0176*	0.0236*	0.0226*
WA	0.0198*	0.0154*	0.0116	0.0126	0.0120	0.0174*	0.0264*	0.0384*	0.0456*
ACN	0.0090	0.0094	0.0096	0.0072	0.0100	0.0092	0.0096	0.0080	0.0092
ACP	0.0122	0.0118	0.0086	0.0108	0.0110	0.0118	0.0108	0.0116	0.0110
LCN	0.0120	0.0128	0.0106	0.0104	0.0110	0.0124	0.0116	0.0100	0.0100
LCP	0.0088	0.0100	0.0106	0.0120	0.0116	0.0100	0.0098	0.0104	0.0100
	$\alpha = 0.025$								
LR	0.0400*	0.0360*	0.0312	0.0296	0.0252	0.0322	0.0290	0.0246	0.0318
Fisher	0.0308	0.0324	0.0392*	0.0360*	0.0404*	0.0440*	0.0392*	0.0468*	0.0480*
WA	0.0380*	0.0332*	0.0272	0.0276	0.0284	0.0358*	0.0494*	0.0606*	0.0730*
ACN	0.0238	0.0268	0.0248	0.0228	0.0216	0.0242	0.0240	0.0190	0.0252
ACP	0.0270	0.0292	0.0242	0.0246	0.0230	0.0264	0.0236	0.0268	0.0272
LCN	0.0260	0.0258	0.0264	0.0292	0.0248	0.0270	0.0258	0.0246	0.0226
LCP	0.0256	0.0220	0.0268	0.0260	0.0290	0.0256	0.0246	0.0278	0.0248
	$\alpha = 0.05$								
LR	0.0682*	0.0654*	0.0584	0.0586	0.0542	0.0632	0.0550	0.0512	0.0582
Fisher	0.0604	0.0586	0.0684*	0.0674*	0.0724*	0.0750*	0.0724*	0.0780*	0.0876*
WA	0.0630	0.0576	0.0540	0.0496	0.0540	0.0652*	0.0806*	0.0922*	0.1030*
ACN	0.0488	0.0486	0.0488	0.0448	0.0470	0.0498	0.0474	0.0438	0.0502
ACP	0.0506	0.0534	0.0500	0.0492	0.0516	0.0548	0.0494	0.0502	0.0520
LCN	0.0514	0.0504	0.0514	0.0588	0.0452	0.0524	0.0482	0.0454	0.0450
LCP	0.0526	0.0466	0.0498	0.0470	0.0534	0.0498	0.0482	0.0512	0.0514
	$\alpha = 0.10$								
LR	0.1314*	0.1194	0.1094	0.1084	0.1114	0.1158	0.1072	0.1020	0.1106
Fisher	0.1084	0.1150	0.1242	0.1168	0.1238	0.1316*	0.1262	0.1358*	0.1492*
WA	0.1106	0.1092	0.1044	0.1012	0.1106	0.1170	0.1262	0.1490*	0.1588*
ACN	0.0932	0.0954	0.0906	0.0948	0.0964	0.0982	0.0994	0.0940	0.1066
ACP	0.1040	0.0958	0.1036	0.0984	0.0984	0.1018	0.0968	0.1014	0.0972
LCN	0.0968	0.0954	0.1030	0.1034	0.0954	0.1032	0.0966	0.0948	0.0956
LCP	0.1006	0.0964	0.1056	0.0970	0.1074	0.0976	0.0994	0.1000	0.1060

The above results are based on 5000 simulation runs with 5000 null samples generated in each simulation run.
* The empirical level is more than 30% above the stated nominal level α.

Table A.2: The empirical levels of test statistics for testing H_0: $\rho_1 = \rho_2$ (sample size $p_1 = p_2 = 50$)

ρ	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
Method	$\alpha = 0.01$								
LR	0.0162*	0.0124	0.0100	0.0076	0.0122	0.0148*	0.0112	0.0098	0.0090
Fisher	0.0122	0.0150*	0.0142*	0.0138*	0.0188*	0.0196*	0.0170*	0.0214*	0.0216*
WA	0.0112	0.0124	0.0112	0.0096	0.0068	0.0094	0.0174*	0.0224*	0.0282*
ACN	0.0094	0.0098	0.0092	0.0092	0.0094	0.0106	0.0082	0.0116	0.0082
ACP	0.0118	0.0072	0.0110	0.0102	0.0118	0.0096	0.0114	0.0122	0.0116
LCN	0.0102	0.0100	0.0086	0.0100	0.0100	0.0100	0.0078	0.0098	0.0104
LCP	0.0098	0.0102	0.0110	0.0128	0.0124	0.0126	0.0096	0.0106	0.0112
	$\alpha = 0.025$								
LR	0.0352*	0.0316	0.0258	0.0232	0.0276	0.0306	0.0250	0.0280	0.0244
Fisher	0.0290	0.0346*	0.0334*	0.0352*	0.0400*	0.0394*	0.0332*	0.0450*	0.0432*
WA	0.0234	0.0286	0.0244	0.0264	0.0222	0.0272	0.0346*	0.0388*	0.0506*
ACN	0.0222	0.0238	0.0258	0.0246	0.0248	0.0254	0.0260	0.0292	0.0260
ACP	0.0272	0.0244	0.0254	0.0282	0.0270	0.0222	0.0262	0.0248	0.0274
LCN	0.0260	0.0248	0.0228	0.0226	0.0266	0.0240	0.0232	0.0240	0.0278
LCP	0.0236	0.0236	0.0232	0.0280	0.0256	0.0246	0.0260	0.0236	0.0274
	$\alpha = 0.05$								
LR	0.0650*	0.0616	0.0500	0.0522	0.0524	0.0566	0.0484	0.0532	0.0568
Fisher	0.0560	0.0618	0.0626	0.0620	0.0716*	0.0734*	0.0688*	0.0784*	0.0758*
WA	0.0482	0.0542	0.0518	0.0494	0.0476	0.0512	0.0610	0.0692*	0.0788*
ACN	0.0490	0.0480	0.0512	0.0482	0.0468	0.0460	0.0466	0.0586	0.0522
ACP	0.0528	0.0498	0.0476	0.0486	0.0580	0.0436	0.0548	0.0478	0.0542
LCN	0.0526	0.0502	0.0482	0.0498	0.0472	0.0488	0.0504	0.0520	0.0510
LCP	0.0444	0.0478	0.0454	0.0528	0.0506	0.0518	0.0480	0.0494	0.0554
	$\alpha = 0.10$								
LR	0.1186	0.1102	0.1034	0.1058	0.1072	0.1106	0.0936	0.1076	0.1106
Fisher	0.1054	0.1084	0.1140	0.1206	0.1274	0.1300*	0.1370*	0.1396*	0.1342*
WA	0.1024	0.1008	0.1056	0.0980	0.0980	0.1004	0.1096	0.1234	0.1358*
ACN	0.1014	0.0914	0.0924	0.0960	0.0954	0.0968	0.0980	0.1084	0.1022
ACP	0.1004	0.0988	0.1038	0.0950	0.1096	0.0942	0.1050	0.0992	0.1074
LCN	0.1046	0.1010	0.0950	0.0974	0.0950	0.0958	0.1004	0.1028	0.1032
LCP	0.0952	0.0928	0.0954	0.1016	0.0952	0.1006	0.1030	0.1020	0.1052

The above results are based on 5000 simulation runs with 5000 null samples generated in each simulation run.

* The empirical level is more than 30% above the stated nominal level α.

Table A.3: The empirical levels of test statistics for testing H_0: $\rho_1 = \rho_2 = \rho_3$ (sample size $p_1 = p_2 = p_3 = 50$)

ρ	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
Method	$\alpha = 0.01$								
LR	0.0158*	0.0172*	0.0154*	0.0118	0.0150*	0.0114	0.0104	0.0120	0.0110
Fisher	0.0142*	0.0174*	0.0172*	0.0218*	0.0226*	0.0228*	0.0244*	0.0232*	0.0300*
WA	0.0198*	0.0134*	0.0096	0.0080	0.0118	0.0182*	0.0318*	0.0476*	0.0766*
ACN	0.0114	0.0116	0.0100	0.0100	0.0100	0.0108	0.0110	0.0104	0.0072
ACP	0.0104	0.0096	0.0116	0.0094	0.0116	0.0122	0.0080	0.0100	0.0078
LCN	0.0106	0.0100	0.0106	0.0070	0.0146	0.0084	0.0120	0.0090	0.0100
LCP	0.0100	0.0110	0.0076	0.0096	0.0112	0.0098	0.0106	0.0092	0.0098
	$\alpha = 0.025$								
LR	0.0364*	0.0322	0.0308	0.0322	0.0344*	0.0296	0.0306	0.0274	0.0316
Fisher	0.0324	0.0376*	0.0388*	0.0408*	0.0450*	0.0460*	0.0462*	0.0454*	0.0594*
WA	0.0414	0.0294	0.0268	0.0224	0.0280	0.0362*	0.0566*	0.0722*	0.1108*
ACN	0.0278	0.0250	0.0230	0.0206	0.0260	0.0258	0.0254	0.0246	0.0216
ACP	0.0226	0.0216	0.0260	0.0214	0.0248	0.0278	0.0224	0.0256	0.0212
LCN	0.0234	0.0272	0.0246	0.0206	0.0290	0.0254	0.0262	0.0228	0.0280
LCP	0.0272	0.0232	0.0198	0.0258	0.0272	0.0252	0.0268	0.0236	0.0254
	$\alpha = 0.05$								
LR	0.0656*	0.0604	0.0624	0.0614	0.0634	0.0598	0.0622	0.0572	0.0592
Fisher	0.0640	0.0668*	0.0680*	0.0750*	0.0786*	0.0816*	0.0820*	0.0804*	0.0916*
WA	0.0738*	0.0574	0.0526	0.0466	0.0514	0.0604	0.0864*	0.1072*	0.1490*
ACN	0.0512	0.0456	0.0506	0.0436	0.0516	0.0476	0.0496	0.0512	0.0460
ACP	0.0474	0.0460	0.0522	0.0504	0.0496	0.0492	0.0442	0.0516	0.0502
LCN	0.0502	0.0478	0.0504	0.0464	0.0562	0.0478	0.0518	0.0466	0.0516
LCP	0.0510	0.0504	0.0454	0.0524	0.0520	0.0464	0.0512	0.0530	0.0484
	$\alpha = 0.10$								
LR	0.1314*	0.1192	0.1108	0.1144	0.1156	0.1150	0.1156	0.1196	0.1134
Fisher	0.1174	0.1256	0.1230	0.1318*	0.1396*	0.1466*	0.1450*	0.1414*	0.1562*
WA	0.1304*	0.1102	0.0974	0.0958	0.1068	0.1094	0.1472*	0.1646*	0.2028*
ACN	0.0950	0.1002	0.0984	0.0952	0.1042	0.0978	0.0958	0.1026	0.0968
ACP	0.0974	0.1034	0.1068	0.1022	0.1014	0.1022	0.0900	0.1000	0.1000
LCN	0.0988	0.0944	0.1020	0.0928	0.1004	0.0964	0.1026	0.0982	0.0966
LCP	0.0954	0.1000	0.0938	0.1018	0.1040	0.0928	0.1016	0.1022	0.1014

The above results are based on 5000 simulation runs with 5000 null samples generated in each simulation run.
* The empirical level is more than 30% above the stated nominal level α.

Table A.4: The empirical levels of test statistics for testing H_0: $\rho_1 = \rho_2 = \rho_3$ (sample size $p_1 = p_2 = p_3 = 50$)

ρ	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
Method	$\alpha = 0.01$								
LR	0.0146*	0.0120	0.0096	0.0114	0.0090	0.0108	0.0086	0.0080	0.0138
Fisher	0.0134*	0.0154*	0.0174*	0.0190*	0.0176*	0.0202*	0.0192*	0.0246*	0.0266*
WA	0.0140*	0.0102	0.0124	0.0092	0.0098	0.0140*	0.0186*	0.0296*	0.0412*
ACN	0.0096	0.0090	0.0124	0.0094	0.0092	0.0104	0.0106	0.0074	0.0088
ACP	0.0084	0.0090	0.0100	0.0102	0.0096	0.0100	0.0076	0.0106	0.0090
LCN	0.0082	0.0114	0.0116	0.0110	0.0120	0.0102	0.0096	0.0092	0.0092
LCP	0.0080	0.0094	0.0092	0.0108	0.0102	0.0100	0.0124	0.0130	0.0072
	$\alpha = 0.025$								
LR	0.0314	0.0304	0.0258	0.0242	0.0256	0.0282	0.0224	0.0236	0.0302
Fisher	0.0314	0.0352*	0.0390*	0.0418*	0.0396*	0.0452*	0.0470*	0.0470*	0.0564*
WA	0.0356*	0.0260	0.0256	0.0226	0.0248	0.0320	0.0400*	0.0528*	0.0698*
ACN	0.0248	0.0248	0.0266	0.0264	0.0248	0.0256	0.0268	0.0208	0.0220
ACP	0.0236	0.0256	0.0284	0.0226	0.0248	0.0240	0.0228	0.0252	0.0242
LCN	0.0248	0.0286	0.0270	0.0262	0.0270	0.0244	0.0264	0.0262	0.0256
LCP	0.0222	0.0256	0.0252	0.0272	0.0228	0.0234	0.0258	0.0276	0.0240
	$\alpha = 0.05$								
LR	0.0628	0.0562	0.0510	0.0524	0.0520	0.0550	0.0510	0.0498	0.0554
Fisher	0.0550	0.0624	0.0692*	0.0738*	0.0692*	0.0792*	0.0802*	0.0808*	0.0922*
WA	0.0600	0.0518	0.0490	0.0486	0.0542	0.0562	0.0684*	0.0796*	0.1024*
ACN	0.0486	0.0482	0.0514	0.0520	0.0484	0.0550	0.0504	0.0482	0.0442
ACP	0.0506	0.0504	0.0510	0.0474	0.0512	0.0516	0.0482	0.0516	0.0526
LCN	0.0500	0.0534	0.0556	0.0508	0.0536	0.0506	0.0514	0.0514	0.0470
LCP	0.0482	0.0494	0.0508	0.0532	0.0474	0.0464	0.0490	0.0540	0.0462
	$\alpha = 0.10$								
LR	0.1166	0.1038	0.1020	0.1034	0.1018	0.1038	0.1022	0.1026	0.1060
Fisher	0.1088	0.1134	0.1264	0.1292	0.1304*	0.1404*	0.1442*	0.1428*	0.1588*
WA	0.1094	0.1012	0.1010	0.1034	0.1034	0.1056	0.1174	0.1332*	0.1578*
ACN	0.0988	0.0988	0.1004	0.1036	0.0966	0.1050	0.0974	0.0962	0.0976
ACP	0.1010	0.1026	0.0952	0.0972	0.0996	0.1002	0.0984	0.1066	0.1066
LCN	0.0984	0.1076	0.1002	0.0972	0.1088	0.0998	0.1004	0.1000	0.0944
LCP	0.0950	0.0960	0.1024	0.1102	0.0996	0.0940	0.0970	0.1042	0.0948

The above results are based on 5000 simulation runs with 5000 null samples generated in each simulation run.
* The empirical level is more than 30% above the stated nominal level α.

Appendix B

Figures of Estimated Powers of Tests for Testing H_0: $\rho_1 = \rho_2$

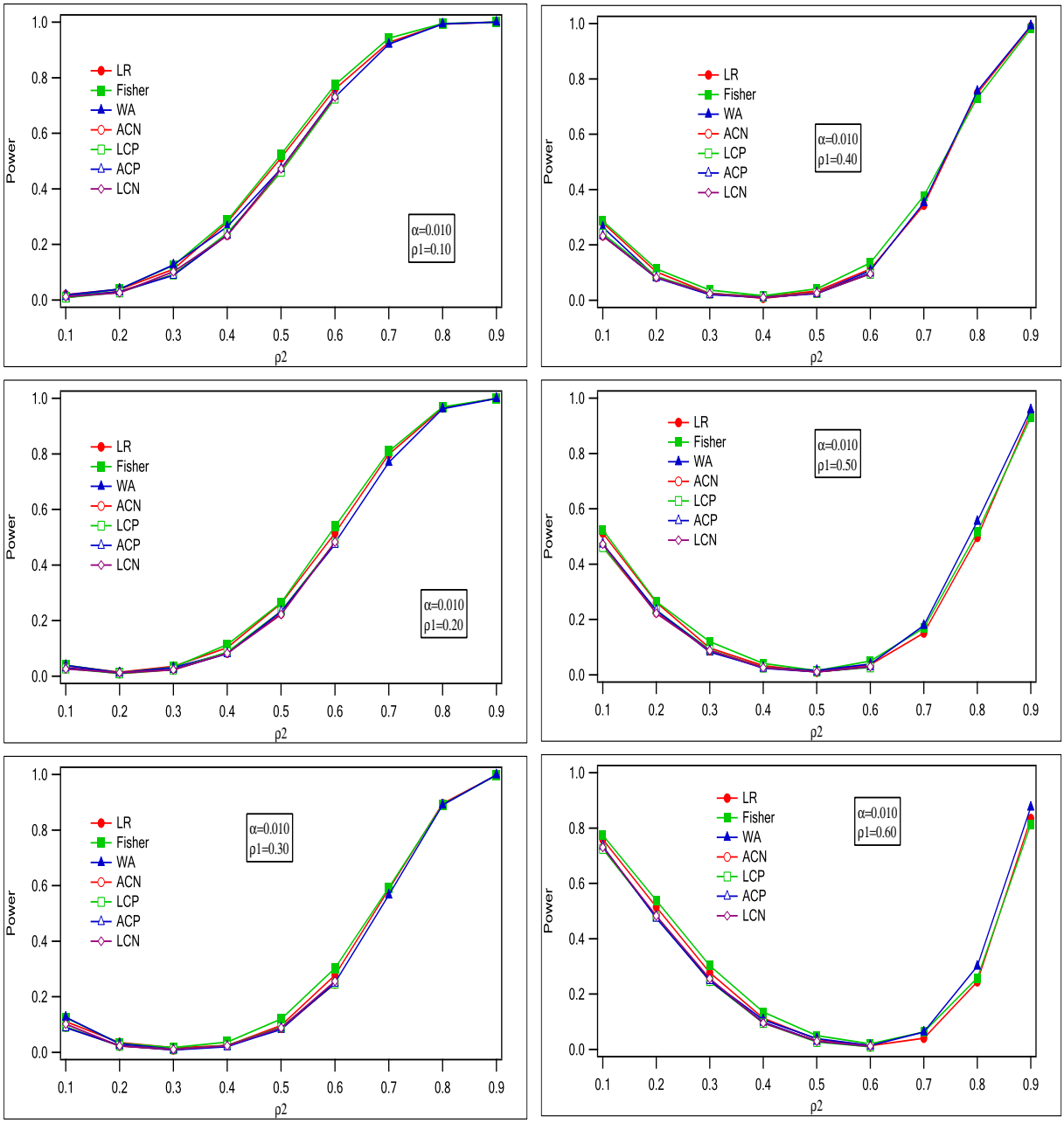

Figure B.1: Estimated power curves of tests for testing $H_0 : \rho 1 = \rho 2$ under $p_1 = p_2 = 25$ and $\alpha = 0.01$

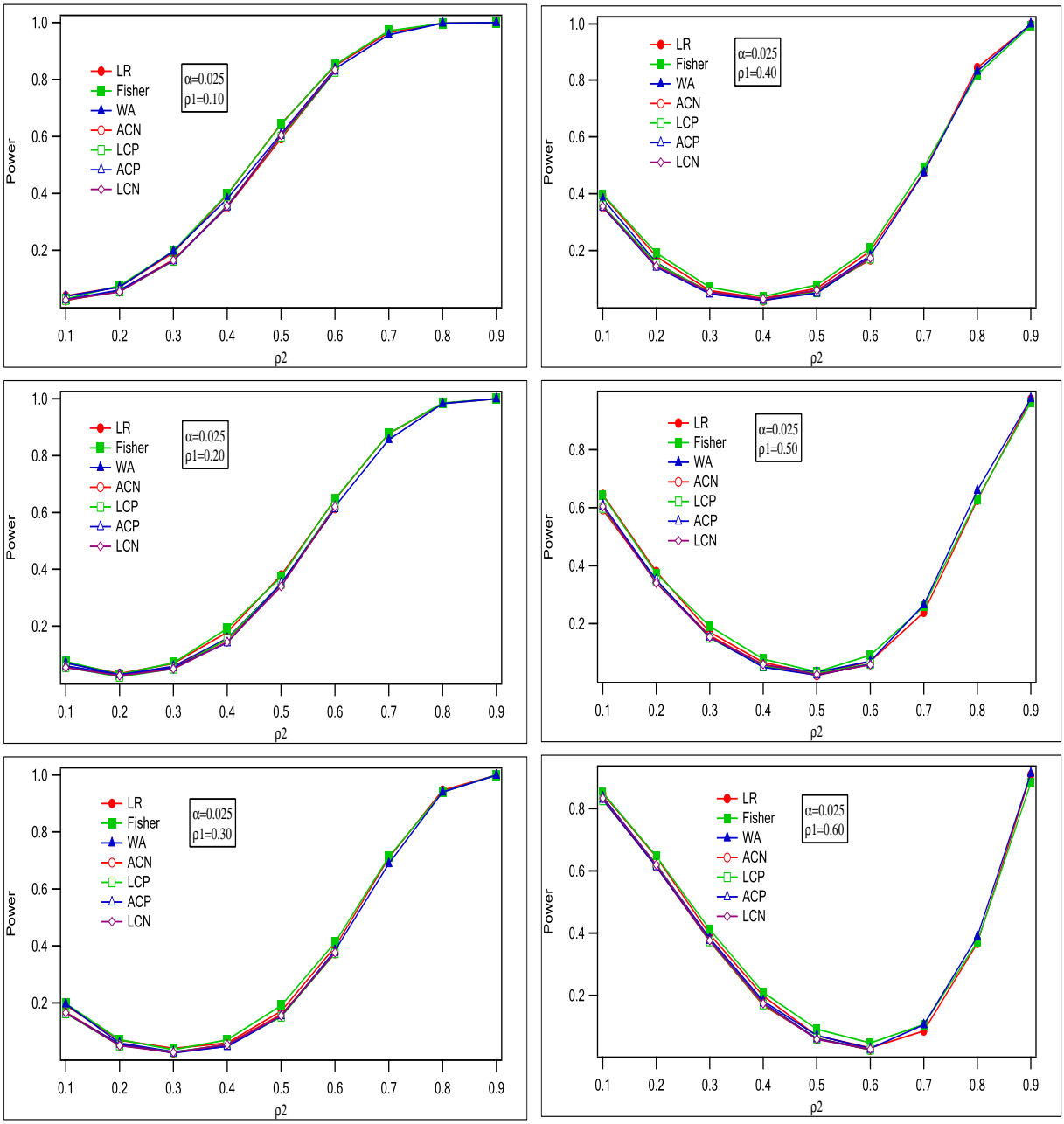

Figure B.2: Estimated power curves of tests for testing $H_0 : \rho 1 = \rho 2$ under $p_1 = p_2 = 25$ and $\alpha = 0.025$

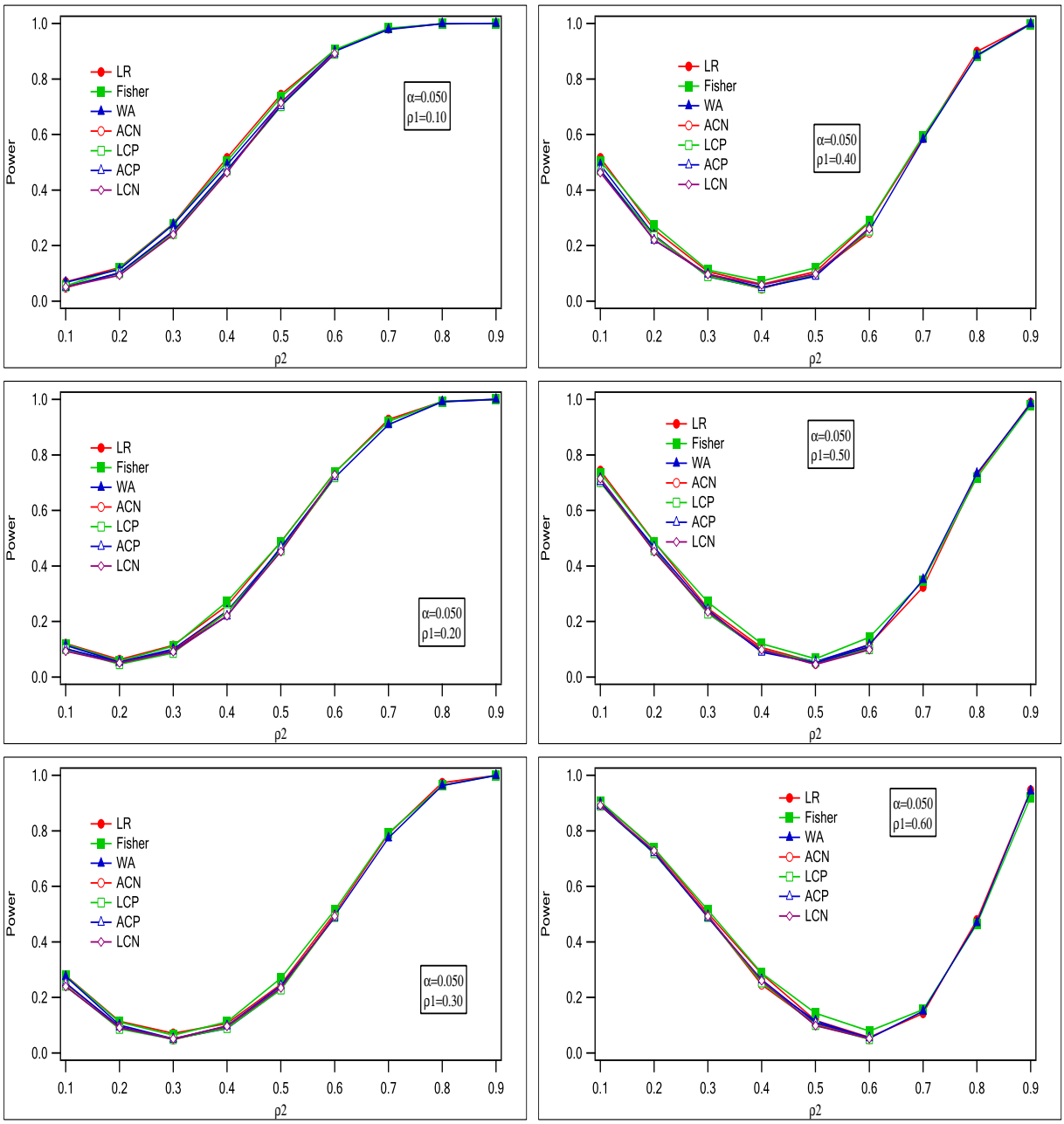

Figure B.3: Estimated power curves of tests for testing $H_0 : \rho 1 = \rho 2$ under $p_1 = p_2 = 25$ and $\alpha = 0.05$

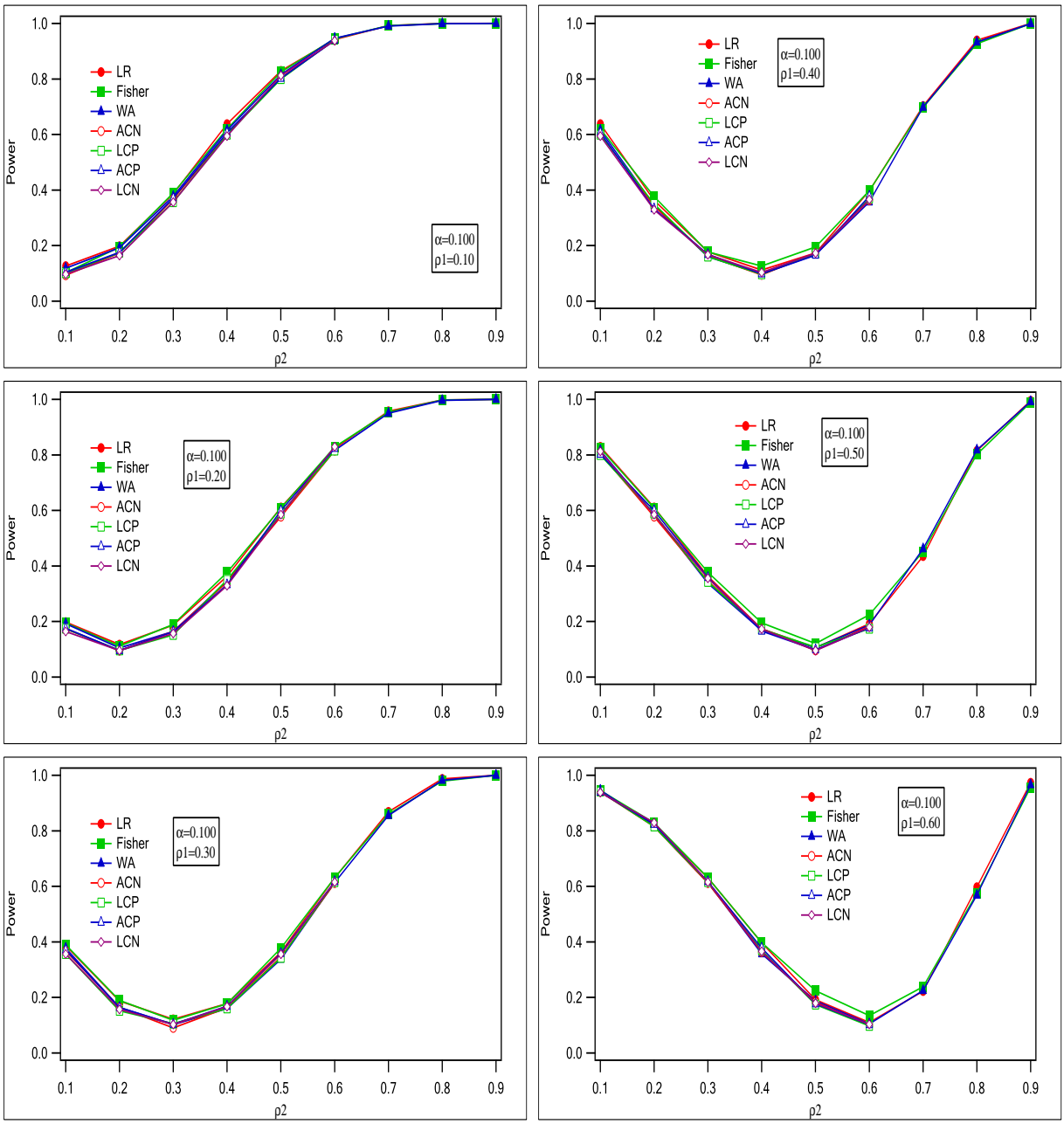

Figure B.4: Estimated power curves of tests for testing $H_0 : \rho 1 = \rho 2$ under $p_1 = p_2 = 25$ and $\alpha = 0.10$

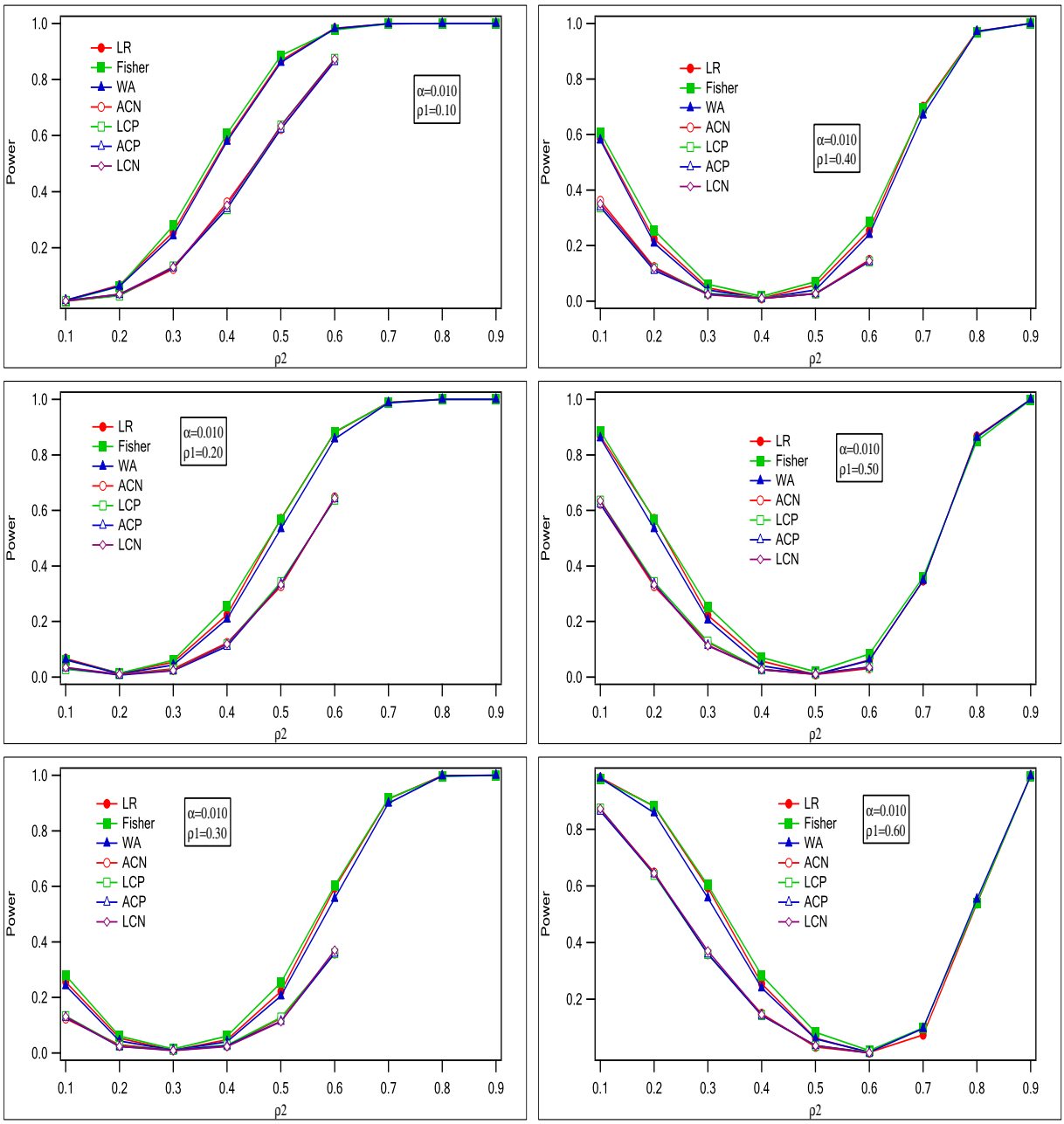

Figure B.5: Estimated power curves of tests for testing $H_0 : \rho 1 = \rho 2$ under $p_1 = p_2 = 50$ and $\alpha = 0.01$

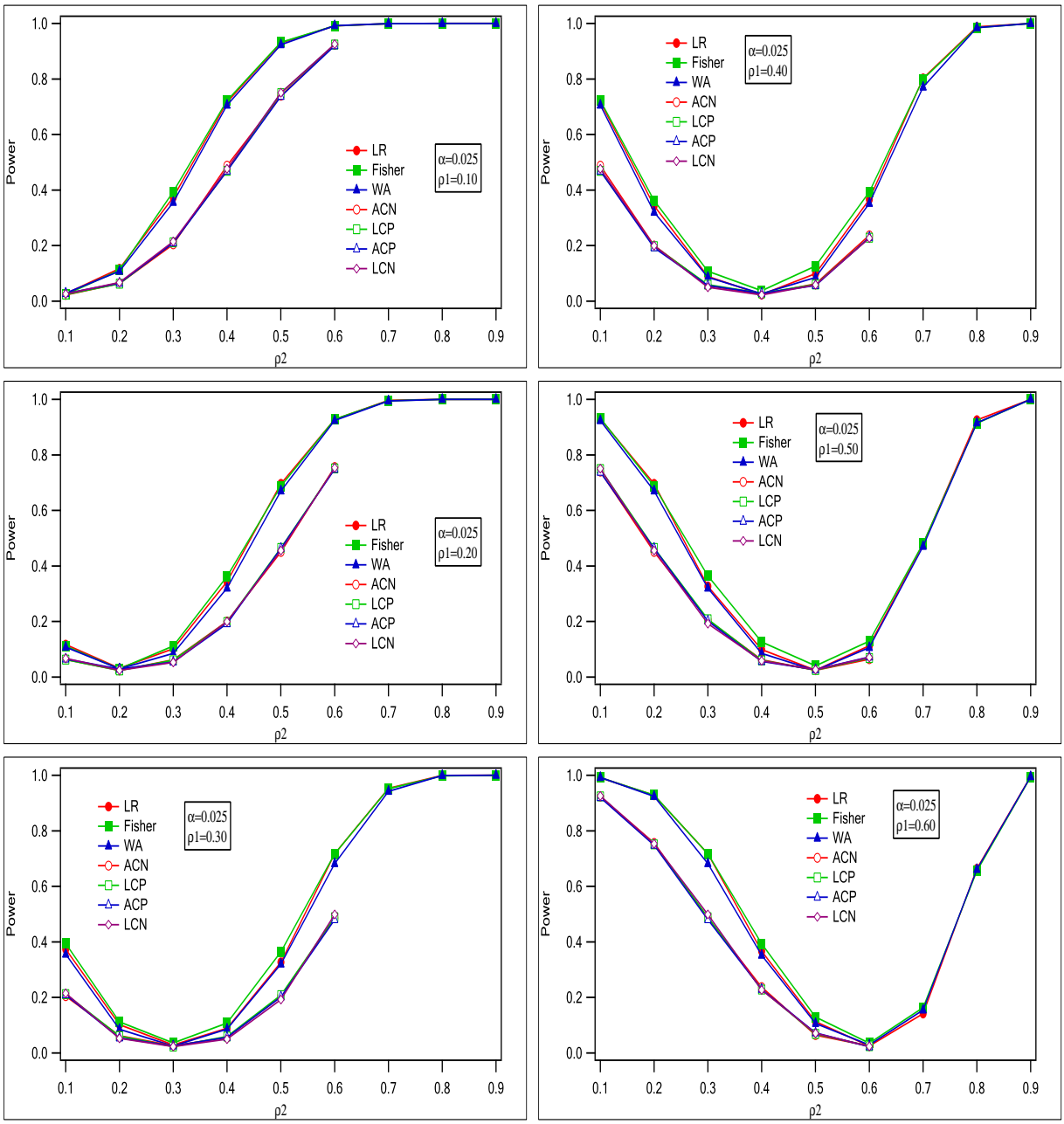

Figure B.6: Estimated power curves of tests for testing $H_0 : \rho 1 = \rho 2$ under $p_1 = p_2 = 50$ and $\alpha = 0.025$

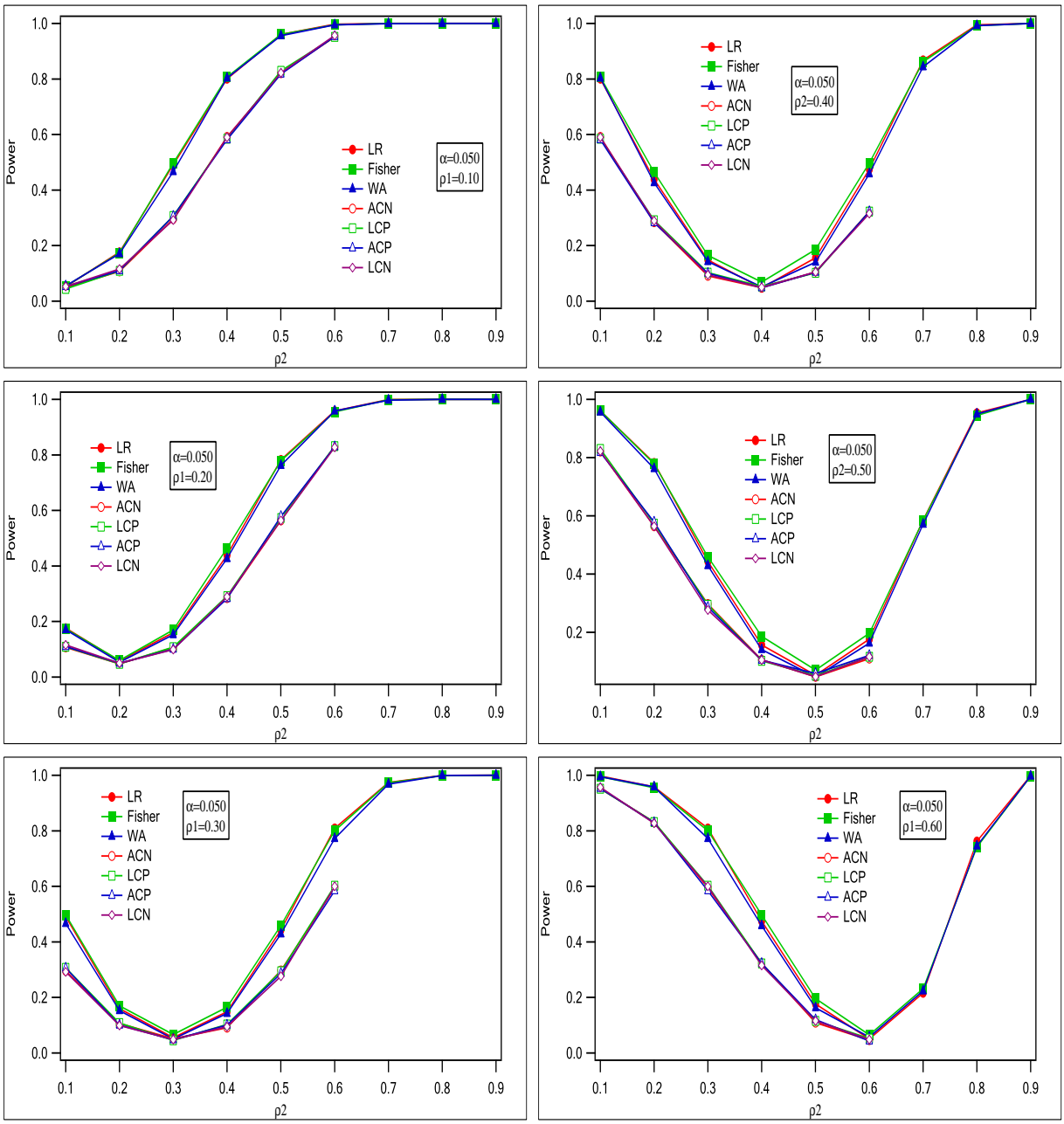

Figure B.7: Estimated power curves of tests for testing $H_0 : \rho 1 = \rho 2$ under $p_1 = p_2 = 50$ and $\alpha = 0.05$

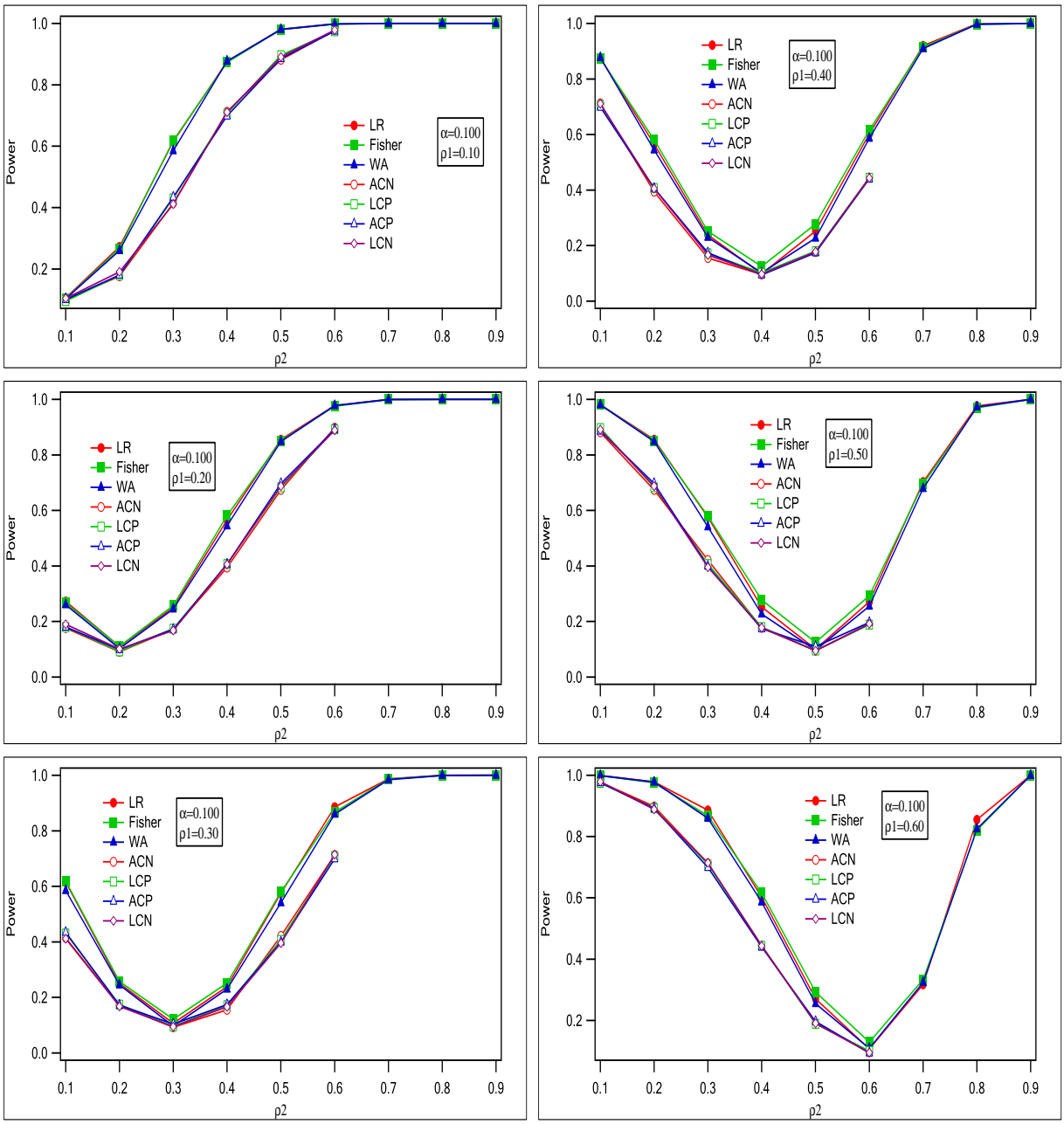

Figure B.8: Estimated power curves of tests for testing $H_0 : \rho 1 = \rho 2$ under $p_1 = p_2 = 50$ and $\alpha = 0.10$

Appendix C

Tables of Estimated Powers of Tests for Testing H_0: $\rho_1 = \rho_2$

Table C.1: Estimated powers of LR test for testing H_0: $\rho_1 = \rho_2$

	$\alpha = 0.01$, $p_1 = p_2 = 25$						$\alpha = 0.01$, $p_1 = p_2 = 50$					
ρ	0.1	0.2	0.3	0.4	0.5	0.6	0.1	0.2	0.3	0.4	0.5	0.6
0.1	**0.0192**	0.0382	0.1106	0.2796	0.5122	0.7594	**0.0122**	0.0664	0.2558	0.5864	0.8676	0.9826
0.2	0.0382	**0.0146**	0.0352	0.1026	0.2620	0.5146	0.0664	**0.0140**	0.0536	0.2244	0.5704	0.8806
0.3	0.1106	0.0352	**0.0168**	0.0246	0.0966	0.2778	0.2558	0.0536	0.0094	0.0478	0.2220	0.5934
0.4	0.2796	0.1026	0.0246	**0.0132**	0.0332	0.1128	0.5864	0.2244	0.0478	**0.0110**	0.0582	0.2546
0.5	0.5122	0.2620	0.0966	0.0332	0.0106	0.0368	0.8676	0.5704	0.2220	0.0582	0.0090	0.0614
0.6	0.7594	0.5146	0.2778	0.1128	0.0368	**0.0136**	0.9826	0.8806	0.5934	0.2546	0.0614	**0.0114**
	$\alpha = 0.025$, $p_1 = p_2 = 25$						$\alpha = 0.025$, $p_1 = p_2 = 50$					
0.1	**0.0398**	0.0718	0.1916	0.3954	0.6450	0.8482	**0.0264**	0.1170	0.3738	0.7166	0.9300	0.9914
0.2	0.0718	**0.0334**	0.0684	0.1786	0.3796	0.6448	0.1170	**0.0320**	0.1020	0.3432	0.6960	0.9274
0.3	0.1916	0.0684	**0.0406**	0.0586	0.1708	0.3960	0.3738	0.1020	**0.0286**	0.0888	0.3268	0.7176
0.4	0.3954	0.1786	0.0586	**0.0310**	0.0664	0.1976	0.7166	0.3432	0.0888	0.0238	0.0994	0.3666
0.5	0.6450	0.3796	0.1708	0.0664	**0.0266**	0.0702	0.9300	0.6960	0.3268	0.0994	0.0244	0.1128
0.6	0.8482	0.6448	0.3960	0.1976	0.0702	**0.0314**	0.9914	0.9274	0.7176	0.3666	0.1128	0.0256
	$\alpha = 0.05$, $p_1 = p_2 = 25$						$\alpha = 0.05$, $p_1 = p_2 = 50$					
0.1	**0.0696**	0.1208	0.2792	0.5162	0.7440	0.9032	**0.0522**	0.1764	0.4910	0.8006	0.9600	0.9970
0.2	0.1208	**0.0632**	0.1140	0.2572	0.4876	0.7362	0.1764	**0.0576**	0.1600	0.4390	0.7820	0.9582
0.3	0.2792	0.1140	**0.0710**	0.1064	0.2476	0.5036	0.4910	0.1600	**0.0552**	0.1482	0.4434	0.8094
0.4	0.5162	0.2572	0.1064	**0.0606**	0.1064	0.2848	0.8006	0.4390	0.1482	0.0480	0.1562	0.4752
0.5	0.7440	0.4876	0.2476	0.1064	**0.0526**	0.1170	0.9600	0.7820	0.4434	0.1562	**0.0522**	0.1762
0.6	0.9032	0.7362	0.5036	0.2848	0.1170	**0.0566**	0.9970	0.9582	0.8094	0.4752	0.1762	0.0506
	$\alpha = 0.10$, $p_1 = p_2 = 25$						$\alpha = 0.10$, $p_1 = p_2 = 50$					
0.1	**0.1262**	0.1982	0.3860	0.6374	0.8296	0.9420	**0.1054**	0.2726	0.6150	0.8754	0.9808	0.9990
0.2	0.1982	**0.1172**	0.1872	0.3632	0.6112	0.8228	0.2726	**0.1102**	0.2500	0.5652	0.8554	0.9778
0.3	0.3860	0.1872	**0.1212**	0.1780	0.3628	0.6308	0.6150	0.2500	**0.1072**	0.2388	0.5754	0.8856
0.4	0.6374	0.3632	0.1780	**0.1122**	0.1736	0.3968	0.8754	0.5652	0.2388	0.0980	0.2524	0.6024
0.5	0.8296	0.6112	0.3628	0.1736	**0.1054**	0.1914	0.9808	0.8554	0.5754	0.2524	0.1006	0.2708
0.6	0.9420	0.8228	0.6308	0.3968	0.1914	**0.1096**	0.9990	0.9778	0.8856	0.6024	0.2708	**0.1076**

The above results are based on 5000 simulation runs with 5000 null samples generated in each simulation run.
Note: The value in Bold is obviously larger than the stated nominal level α. The ratio of biased point over total is $38/48(> 0.5)$, which means LR test method is biased.

Table C.2: Estimated powers of Fisher test for testing H_0: $\rho_1 = \rho_2$

	$\alpha = 0.01$, $p_1 = p_2 = 25$						$\alpha = 0.01$, $p_1 = p_2 = 50$					
ρ	0.1	0.2	0.3	0.4	0.5	0.6	0.1	0.2	0.3	0.4	0.5	0.6
0.1	**0.0116**	0.0398	0.1242	0.2862	0.5242	0.7744	0.0104	0.0614	0.2790	0.6070	0.8856	0.9772
0.2	0.0398	**0.0124**	0.0328	0.1132	0.2650	0.5384	0.0614	**0.0136**	0.0610	0.2552	0.5676	0.8826
0.3	0.1242	0.0328	**0.0174**	0.0368	0.1202	0.3032	0.2790	0.0610	**0.0142**	0.0604	0.2532	0.6030
0.4	0.2862	0.1132	0.0368	**0.0164**	0.0410	0.1350	0.6070	0.2552	0.0604	**0.0170**	0.0704	0.2840
0.5	0.5242	0.2650	0.1202	0.0410	**0.0152**	0.0498	0.8856	0.5676	0.2532	0.0704	**0.0194**	0.0830
0.6	0.7744	0.5384	0.3032	0.1350	0.0498	**0.0204**	0.9772	0.8826	0.6030	0.2840	0.0830	**0.0180**
	$\alpha = 0.025$, $p_1 = p_2 = 25$						$\alpha = 0.025$, $p_1 = p_2 = 50$					
0.1	**0.0286**	0.0758	0.1976	0.3976	0.6422	0.8516	0.0240	0.1116	0.3944	0.7220	0.9322	0.9912
0.2	0.0758	**0.0306**	0.0706	0.1914	0.3728	0.6474	0.1116	**0.0298**	0.1122	0.3620	0.6874	0.9284
0.3	0.1976	0.0706	**0.0358**	0.0698	0.1908	0.4116	0.3944	0.1122	**0.0360**	0.1074	0.3644	0.7172
0.4	0.3976	0.1914	0.0698	**0.0376**	0.0782	0.2096	0.7220	0.3620	0.1074	**0.0378**	0.1262	0.3936
0.5	0.6422	0.3728	0.1908	0.0782	**0.0342**	0.0910	0.9322	0.6874	0.3644	0.1262	**0.0408**	0.1294
0.6	0.8516	0.6474	0.4116	0.2096	0.0910	**0.0468**	0.9912	0.9284	0.7172	0.3936	0.1294	**0.0354**
	$\alpha = 0.05$, $p_1 = p_2 = 25$						$\alpha = 0.05$, $p_1 = p_2 = 50$					
0.1	**0.0546**	0.1188	0.2774	0.5038	0.7348	0.9050	**0.0536**	0.1734	0.4974	0.8070	0.9604	0.9952
0.2	0.1188	**0.0578**	0.1110	0.2720	0.4856	0.7366	0.1734	**0.0602**	0.1698	0.4640	0.7786	0.9550
0.3	0.2774	0.1110	**0.0646**	0.1120	0.2700	0.5148	0.4974	0.1698	**0.0666**	0.1646	0.4592	0.8004
0.4	0.5038	0.2720	0.1120	**0.0722**	0.1198	0.2876	0.8070	0.4640	0.1646	**0.0688**	0.1860	0.4974
0.5	0.7348	0.4856	0.2700	0.1198	**0.0660**	0.1432	0.9604	0.7786	0.4592	0.1860	**0.0702**	0.1962
0.6	0.9050	0.7366	0.5148	0.2876	0.1432	**0.0786**	0.9952	0.9550	0.8004	0.4974	0.1962	**0.0650**
	$\alpha = 0.10$, $p_1 = p_2 = 25$						$\alpha = 0.10$, $p_1 = p_2 = 50$					
0.1	**0.1024**	0.1944	0.3894	0.6198	0.8252	0.9452	**0.1038**	0.2666	0.6186	0.8734	0.9802	0.9984
0.2	0.1944	**0.1110**	0.1892	0.3766	0.6080	0.8284	0.2666	**0.1104**	0.2574	0.5810	0.8514	0.9756
0.3	0.3894	0.1892	**0.1178**	0.1768	0.3772	0.6314	0.6186	0.2574	**0.1216**	0.2518	0.5802	0.8672
0.4	0.6198	0.3766	0.1768	**0.1268**	0.1950	0.3978	0.8734	0.5810	0.2518	**0.1250**	0.2776	0.6158
0.5	0.8252	0.6080	0.3772	0.1950	**0.1206**	0.2244	0.9802	0.8514	0.5802	0.2776	**0.1250**	0.2930
0.6	0.9452	0.8284	0.6314	0.3978	0.2244	**0.1344**	0.9984	0.9756	0.8672	0.6158	0.2930	**0.1294**

The above results are based on 5000 simulation runs with 5000 null samples generated in each simulation run.
Note: The value in Bold is obviously larger than the stated nominal level α. The ratio of biased point over total is $46/48(> 0.5)$, which means Fisher test method is biased.

Table C.3: Estimated powers of WA test for testing H_0: $\rho_1 = \rho_2$

	$\alpha = 0.01$, $p_1 = p_2 = 25$						$\alpha = 0.01$, $p_1 = p_2 = 50$					
ρ	0.1	0.2	0.3	0.4	0.5	0.6	0.1	0.2	0.3	0.4	0.5	0.6
0.1	**0.0184**	0.0392	0.1262	0.2638	0.4730	0.7304	**0.0128**	0.0626	0.2424	0.5798	0.8612	0.9812
0.2	0.0392	**0.0120**	0.0308	0.0832	0.2328	0.4804	0.0626	**0.0114**	0.0436	0.2088	0.5340	0.8578
0.3	0.1262	0.0308	**0.0118**	0.0226	0.0822	0.2508	0.2424	0.0436	0.0102	0.0406	0.2054	0.5570
0.4	0.2638	0.0832	0.0226	0.0088	0.0268	0.1072	0.5798	0.2088	0.0406	0.0104	0.0406	0.2398
0.5	0.4730	0.2328	0.0822	0.0268	**0.0158**	0.0390	0.8612	0.5340	0.2054	0.0406	0.0104	0.0594
0.6	0.7304	0.4804	0.2508	0.1072	0.0390	**0.0142**	0.9812	0.8578	0.5570	0.2398	0.0594	**0.0126**
	$\alpha = 0.025$, $p_1 = p_2 = 25$						$\alpha = 0.025$, $p_1 = p_2 = 50$					
0.1	**0.0386**	0.0708	0.1964	0.3812	0.6100	0.8374	**0.0290**	0.1076	0.3554	0.7060	0.9230	0.9922
0.2	0.0708	**0.0294**	0.0586	0.1574	0.3496	0.6208	0.1076	**0.0298**	0.0856	0.3204	0.6706	0.9244
0.3	0.1964	0.0586	**0.0286**	0.0468	0.1516	0.3832	0.3554	0.0856	0.0256	0.0860	0.3206	0.6822
0.4	0.3812	0.1574	0.0468	0.0246	0.0574	0.1820	0.7060	0.3204	0.0860	**0.0270**	0.0854	0.3522
0.5	0.6100	0.3496	0.1516	0.0574	**0.0324**	0.0704	0.9230	0.6706	0.3206	0.0854	0.0240	0.1060
0.6	0.8374	0.6208	0.3832	0.1820	0.0704	**0.0294**	0.9922	0.9244	0.6822	0.3522	0.1060	**0.0288**
	$\alpha = 0.05$, $p_1 = p_2 = 25$						$\alpha = 0.05$, $p_1 = p_2 = 50$					
0.1	**0.0678**	0.1146	0.2756	0.4912	0.7148	0.8996	**0.0548**	0.1704	0.4662	0.8032	0.9562	0.9950
0.2	0.1146	**0.0544**	0.1006	0.2382	0.4616	0.7182	0.1704	**0.0538**	0.1524	0.4266	0.7620	0.9580
0.3	0.2756	0.1006	0.0510	0.0880	0.2298	0.4948	0.4662	0.1524	0.0508	0.1424	0.4286	0.7724
0.4	0.4912	0.2382	0.0880	0.0480	0.0946	0.2538	0.8032	0.4266	0.1424	**0.0512**	0.1402	0.4586
0.5	0.7148	0.4616	0.2298	0.0946	**0.0558**	0.1158	0.9562	0.7620	0.4286	0.1402	0.0458	0.1624
0.6	0.8996	0.7182	0.4948	0.2538	0.1158	**0.0526**	0.9950	0.9580	0.7724	0.4586	0.1624	**0.0572**
	$\alpha = 0.10$, $p_1 = p_2 = 25$						$\alpha = 0.10$, $p_1 = p_2 = 50$					
0.1	**0.1188**	0.1922	0.3778	0.6154	0.8158	0.9464	**0.1030**	0.2600	0.5852	0.8768	0.9802	0.9982
0.2	0.1922	**0.1048**	0.1638	0.3446	0.5874	0.8160	0.2600	**0.1030**	0.2448	0.5448	0.8484	0.9776
0.3	0.3778	0.1638	**0.1020**	0.1600	0.3376	0.6136	0.5852	0.2448	0.0970	0.2296	0.5408	0.8598
0.4	0.6154	0.3446	0.1600	0.0986	0.1678	0.3570	0.8768	0.5448	0.2296	**0.1022**	0.2270	0.5866
0.5	0.8158	0.5874	0.3376	0.1678	**0.1060**	0.1874	0.9802	0.8484	0.5408	0.2270	0.0970	0.2550
0.6	0.9464	0.8160	0.6136	0.3570	0.1874	**0.1050**	0.9982	0.9776	0.8598	0.5866	0.2550	**0.1092**

The above results are based on 5000 simulation runs with 5000 null samples generated in each simulation run.
Note: The value in Bold is obviously larger than the stated nominal level α. The ratio of biased point over total is $34/48(> 0.5)$, which means WA test method is biased.

Table C.4: Estimated powers of ACN test for testing H_0: $\rho_1 = \rho_2$

	$\alpha = 0.01$, $p_1 = p_2 = 25$						$\alpha = 0.01$, $p_1 = p_2 = 50$					
ρ	0.1	0.2	0.3	0.4	0.5	0.6	0.1	0.2	0.3	0.4	0.5	0.6
0.1	0.0090	0.0312	0.0912	0.2330	0.4620	0.7266	0.0094	0.0302	0.1240	0.3618	0.6234	0.8706
0.2	0.0312	0.0094	0.0230	0.0830	0.2250	0.4806	0.0302	0.0098	0.0298	0.1230	0.3270	0.6474
0.3	0.0912	0.0230	0.0096	0.0250	0.0874	0.2468	0.1240	0.0298	0.0092	0.0250	0.1254	0.3578
0.4	0.2330	0.0830	0.0250	0.0072	0.0280	0.0968	0.3618	0.1230	0.0250	0.0092	0.0274	0.1488
0.5	0.4620	0.2250	0.0874	0.0280	0.0100	0.0298	0.6234	0.3270	0.1254	0.0274	0.0094	0.0310
0.6	0.7266	0.4806	0.2468	0.0968	0.0298	0.0092	0.8706	0.6474	0.3578	0.1488	0.0310	0.0106
	$\alpha = 0.025$, $p_1 = p_2 = 25$						$\alpha = 0.025$, $p_1 = p_2 = 50$					
0.1	0.0238	0.0590	0.1670	0.3516	0.5926	0.8292	0.0222	0.0634	0.2048	0.4878	0.7400	0.9236
0.2	0.0590	**0.0268**	0.0546	0.1520	0.3452	0.6138	0.0634	0.0238	0.0622	0.2004	0.4506	0.7566
0.3	0.1670	0.0546	0.0248	0.0548	0.1590	0.3730	0.2048	0.0622	0.0258	0.0558	0.2064	0.4862
0.4	0.3516	0.1520	0.0548	0.0228	0.0574	0.1676	0.4878	0.2004	0.0558	0.0246	0.0624	0.2374
0.5	0.5926	0.3452	0.1590	0.0574	0.0216	0.0616	0.7400	0.4506	0.2064	0.0624	0.0248	0.0640
0.6	0.8292	0.6138	0.3730	0.1676	0.0616	0.0242	0.9236	0.7566	0.4862	0.2374	0.0640	0.0254
	$\alpha = 0.05$, $p_1 = p_2 = 25$						$\alpha = 0.05$, $p_1 = p_2 = 50$					
0.1	0.0488	0.1018	0.2498	0.4686	0.7024	0.8922	0.0490	0.1070	0.2954	0.5916	0.8208	0.9516
0.2	0.1018	0.0486	0.0946	0.2338	0.4544	0.7200	0.1070	0.0480	0.1064	0.2842	0.5638	0.8312
0.3	0.2498	0.0946	0.0488	0.0932	0.2352	0.4930	0.2954	0.1064	**0.0512**	0.0910	0.2972	0.5946
0.4	0.4686	0.2338	0.0932	0.0448	0.1000	0.2460	0.5916	0.2842	0.0910	0.0482	0.1050	0.3232
0.5	0.7024	0.4544	0.2352	0.1000	0.0470	0.1046	0.8208	0.5638	0.2972	0.1050	0.0468	0.1092
0.6	0.8922	0.7200	0.4930	0.2460	0.1046	0.0498	0.9516	0.8312	0.5946	0.3232	0.1092	0.0460
	$\alpha = 0.10$, $p_1 = p_2 = 25$						$\alpha = 0.10$, $p_1 = p_2 = 50$					
0.1	0.0932	0.1734	0.3670	0.6018	0.8050	0.9398	**0.1014**	0.1762	0.4144	0.7124	0.8814	0.9770
0.2	0.1734	0.0954	0.1640	0.3422	0.5774	0.8156	0.1762	0.0914	0.1722	0.3944	0.6740	0.8982
0.3	0.3670	0.1640	0.0906	0.1626	0.3454	0.6116	0.4144	0.1722	0.0924	0.1558	0.4216	0.7126
0.4	0.6018	0.3422	0.1626	0.0948	0.1738	0.3640	0.7124	0.3944	0.1558	0.0960	0.1770	0.4454
0.5	0.8050	0.5774	0.3454	0.1738	0.0964	0.1832	0.8814	0.6740	0.4216	0.1770	0.0954	0.1914
0.6	0.9398	0.8156	0.6116	0.3640	0.1832	0.0982	0.9770	0.8982	0.7126	0.4454	0.1914	0.0968

The above results are based on 5000 simulation runs with 5000 null samples generated in each simulation run.
Note: The value in Bold is obviously larger than the stated nominal level α. The ratio of biased point over total is $3/48(< 0.5)$, which means ACN test method is much less biased.

Table C.5: Estimated powers of ACP test for testing H_0: $\rho_1 = \rho_2$

	$\alpha = 0.01$, $p_1 = p_2 = 25$						$\alpha = 0.01$, $p_1 = p_2 = 50$					
ρ	0.1	0.2	0.3	0.4	0.5	0.6	0.1	0.2	0.3	0.4	0.5	0.6
0.1	**0.0122**	0.0300	0.0892	0.2336	0.4732	0.7346	**0.0118**	0.0330	0.1286	0.3396	0.6234	0.8640
0.2	0.0300	**0.0118**	0.0244	0.0808	0.2326	0.4740	0.0330	0.0072	0.0232	0.1106	0.3358	0.6422
0.3	0.0892	0.0244	0.0086	0.0204	0.0876	0.2486	0.1286	0.0232	0.0110	0.0258	0.1144	0.3592
0.4	0.2336	0.0808	0.0204	0.0108	0.0240	0.0986	0.3396	0.1106	0.0258	0.0102	0.0270	0.1422
0.5	0.4732	0.2326	0.0876	0.0240	0.0110	0.0296	0.6234	0.3358	0.1144	0.0270	**0.0118**	0.0358
0.6	0.7346	0.4740	0.2486	0.0986	0.0296	**0.0118**	0.8640	0.6422	0.3592	0.1422	0.0358	0.0096
	$\alpha = 0.025$, $p_1 = p_2 = 25$						$\alpha = 0.025$, $p_1 = p_2 = 50$					
0.1	**0.0270**	0.0604	0.1636	0.3524	0.6038	0.8302	**0.0272**	0.0648	0.2096	0.4686	0.7372	0.9200
0.2	0.0604	**0.0292**	0.0546	0.1414	0.3508	0.6134	0.0648	0.0244	0.0544	0.1924	0.4636	0.7482
0.3	0.1636	0.0546	0.0242	0.0474	0.1544	0.3768	0.2096	0.0544	0.0254	0.0564	0.2014	0.4804
0.4	0.3524	0.1414	0.0474	0.0246	0.0498	0.1762	0.4686	0.1924	0.0564	**0.0282**	0.0564	0.2290
0.5	0.6038	0.3508	0.1544	0.0498	0.0230	0.0604	0.7372	0.4636	0.2014	0.0564	**0.0270**	0.0720
0.6	0.8302	0.6134	0.3768	0.1762	0.0604	**0.0264**	0.9200	0.7482	0.4804	0.2290	0.0720	0.0222
	$\alpha = 0.05$, $p_1 = p_2 = 25$						$\alpha = 0.05$, $p_1 = p_2 = 50$					
0.1	0.0506	0.1022	0.2522	0.4730	0.7042	0.8910	**0.0528**	0.1102	0.3058	0.5816	0.8176	0.9528
0.2	0.1022	**0.0534**	0.0940	0.2200	0.4682	0.7214	0.1102	0.0498	0.1002	0.2826	0.5784	0.8296
0.3	0.2522	0.0940	0.0500	0.0974	0.2424	0.4880	0.3058	0.1002	0.0476	0.1008	0.2862	0.5842
0.4	0.4730	0.2200	0.0974	0.0492	0.0898	0.2640	0.5816	0.2826	0.1008	0.0486	0.1040	0.3228
0.5	0.7042	0.4682	0.2424	0.0898	**0.0516**	0.1094	0.8176	0.5784	0.2862	0.1040	**0.0580**	0.1202
0.6	0.8910	0.7214	0.4880	0.2640	0.1094	**0.0548**	0.9528	0.8296	0.5842	0.3228	0.1202	0.0436
	$\alpha = 0.10$, $p_1 = p_2 = 25$						$\alpha = 0.10$, $p_1 = p_2 = 50$					
0.1	**0.1040**	0.1758	0.3710	0.6068	0.8026	0.9392	0.1004	0.1804	0.4344	0.6984	0.8854	0.9754
0.2	0.1758	0.0958	0.1634	0.3326	0.5994	0.8238	0.1804	0.0988	0.1720	0.4066	0.6970	0.8914
0.3	0.3710	0.1634	**0.1036**	0.1686	0.3586	0.6172	0.4344	0.1720	**0.1038**	0.1742	0.3988	0.6994
0.4	0.6068	0.3326	0.1686	0.0984	0.1656	0.3782	0.6984	0.4066	0.1742	0.0950	0.1742	0.4400
0.5	0.8026	0.5994	0.3586	0.1656	0.0984	0.1770	0.8854	0.6970	0.3988	0.1742	**0.1096**	0.1970
0.6	0.9392	0.8238	0.6172	0.3782	0.1770	**0.1018**	0.9754	0.8914	0.6994	0.4400	0.1970	0.0942

The above results are based on 5000 simulation runs with 5000 null samples generated in each simulation run.
Note: The value in Bold is obviously larger than the stated nominal level α. The ratio of biased point over total is $21/48(< 0.5)$, which means ACP test method is less biased.

Table C.6: Estimated powers of LCN test for testing H_0: $\rho_1 = \rho_2$

	$\alpha = 0.01$, $p_1 = p_2 = 25$						$\alpha = 0.01$, $p_1 = p_2 = 50$					
ρ	0.1	0.2	0.3	0.4	0.5	0.6	0.1	0.2	0.3	0.4	0.5	0.6
0.1	**0.0120**	0.0262	0.1020	0.2322	0.4720	0.7300	0.0102	0.0344	0.1308	0.3508	0.6350	0.8724
0.2	0.0262	**0.0128**	0.0230	0.0824	0.2218	0.4830	0.0344	0.0100	0.0240	0.1192	0.3328	0.6444
0.3	0.1020	0.0230	0.0106	0.0238	0.0880	0.2562	0.1308	0.0240	0.0086	0.0230	0.1126	0.3702
0.4	0.2322	0.0824	0.0238	0.0104	0.0260	0.0958	0.3508	0.1192	0.0230	0.0100	0.0270	0.1452
0.5	0.4720	0.2218	0.0880	0.0260	0.0110	0.0300	0.6350	0.3328	0.1126	0.0270	0.0100	0.0350
0.6	0.7300	0.4830	0.2562	0.0958	0.0300	**0.0124**	0.8724	0.6444	0.3702	0.1452	0.0350	0.0100
	$\alpha = 0.025$, $p_1 = p_2 = 25$						$\alpha = 0.025$, $p_1 = p_2 = 50$					
0.1	0.0260	0.0542	0.1650	0.3552	0.6040	0.8336	0.0260	0.0674	0.2146	0.4766	0.7502	0.9258
0.2	0.0542	0.0258	0.0496	0.1446	0.3394	0.6200	0.0674	0.0248	0.0528	0.1998	0.4556	0.7532
0.3	0.1650	0.0496	**0.0264**	0.0534	0.1544	0.3764	0.2146	0.0528	0.0228	0.0498	0.1920	0.4986
0.4	0.3552	0.1446	0.0534	**0.0292**	0.0598	0.1732	0.4766	0.1998	0.0498	0.0226	0.0586	0.2272
0.5	0.6040	0.3394	0.1544	0.0598	0.0248	0.0582	0.7502	0.4556	0.1920	0.0586	**0.0266**	0.0706
0.6	0.8336	0.6200	0.3764	0.1732	0.0582	**0.0270**	0.9258	0.7532	0.4986	0.2272	0.0706	0.0240
	$\alpha = 0.05$, $p_1 = p_2 = 25$						$\alpha = 0.05$, $p_1 = p_2 = 50$					
0.1	**0.0514**	0.0934	0.2392	0.4626	0.7144	0.8912	**0.0526**	0.1156	0.2924	0.5902	0.8220	0.9566
0.2	0.0934	0.0504	0.0912	0.2208	0.4522	0.7284	0.1156	0.0502	0.0988	0.2890	0.5642	0.8270
0.3	0.2392	0.0912	**0.0514**	0.0962	0.2340	0.4924	0.2924	0.0988	0.0482	0.0946	0.2762	0.6004
0.4	0.4626	0.2208	0.0962	**0.0588**	0.0980	0.2600	0.5902	0.2890	0.0946	0.0498	0.1046	0.3152
0.5	0.7144	0.4522	0.2340	0.0980	0.0452	0.0984	0.8220	0.5642	0.2762	0.1046	0.0472	0.1158
0.6	0.8912	0.7284	0.4924	0.2600	0.0984	**0.0524**	0.9566	0.8270	0.6004	0.3152	0.1158	0.0488
	$\alpha = 0.10$, $p_1 = p_2 = 25$						$\alpha = 0.10$, $p_1 = p_2 = 50$					
0.1	0.0968	0.1640	0.3566	0.5940	0.8128	0.9380	**0.1046**	0.1904	0.4112	0.7106	0.8908	0.9784
0.2	0.1640	0.0954	0.1566	0.3288	0.5850	0.8296	0.1904	0.1010	0.1674	0.4056	0.6874	0.8886
0.3	0.3566	0.1566	**0.1030**	0.1672	0.3558	0.6154	0.4112	0.1674	0.0950	0.1660	0.3954	0.7142
0.4	0.5940	0.3288	0.1672	**0.1034**	0.1730	0.3660	0.7106	0.4056	0.1660	0.0974	0.1780	0.4432
0.5	0.8128	0.5850	0.3558	0.1730	0.0954	0.1792	0.8908	0.6874	0.3954	0.1780	0.0950	0.1914
0.6	0.9380	0.8296	0.6154	0.3660	0.1792	**0.1032**	0.9784	0.8886	0.7142	0.4432	0.1914	0.0958

The above results are based on 5000 simulation runs with 5000 null samples generated in each simulation run.
Note: The value in Bold is obviously larger than the stated nominal level α. The ratio of biased point over total is $16/48(< 0.5)$, which means LCN test method is less biased.

Table C.7: Estimated powers of LCP test for testing H_0: $\rho_1 = \rho_2$

	$\alpha = 0.01$, $p_1 = p_2 = 25$						$\alpha = 0.01$, $p_1 = p_2 = 50$					
ρ	0.1	0.2	0.3	0.4	0.5	0.6	0.1	0.2	0.3	0.4	0.5	0.6
0.1	0.0088	0.0272	0.0916	0.2406	0.4610	0.7248	0.0098	0.0284	0.1334	0.3374	0.6358	0.8736
0.2	0.0272	0.0100	0.0242	0.0866	0.2340	0.4804	0.0284	0.0102	0.0262	0.1178	0.3420	0.6386
0.3	0.0916	0.0242	0.0106	0.0244	0.0904	0.2476	0.1334	0.0262	0.0110	0.0282	0.1282	0.3596
0.4	0.2406	0.0866	0.0244	**0.0120**	0.0262	0.0958	0.3374	0.1178	0.0282	**0.0128**	0.0274	0.1432
0.5	0.4610	0.2340	0.0904	0.0262	**0.0116**	0.0270	0.6358	0.3420	0.1282	0.0274	**0.0124**	0.0354
0.6	0.7248	0.4804	0.2476	0.0958	0.0270	0.0100	0.8736	0.6386	0.3596	0.1432	0.0354	**0.0126**
	$\alpha = 0.025$, $p_1 = p_2 = 25$						$\alpha = 0.025$, $p_1 = p_2 = 50$					
0.1	0.0256	0.0556	0.1632	0.3582	0.5992	0.8270	0.0236	0.0632	0.2122	0.4694	0.7488	0.9244
0.2	0.0556	0.0220	0.0504	0.1542	0.3488	0.6216	0.0632	0.0236	0.0604	0.1980	0.4652	0.7514
0.3	0.1632	0.0504	**0.0268**	0.0530	0.1512	0.3738	0.2122	0.0604	0.0232	0.0596	0.2086	0.4888
0.4	0.3582	0.1542	0.0530	0.0260	0.0526	0.1712	0.4694	0.1980	0.0596	**0.0280**	0.0602	0.2284
0.5	0.5992	0.3488	0.1512	0.0526	**0.0290**	0.0606	0.7488	0.4652	0.2086	0.0602	0.0256	0.0690
0.6	0.8270	0.6216	0.3738	0.1712	0.0606	0.0256	0.9244	0.7514	0.4888	0.2284	0.0690	0.0246
	$\alpha = 0.05$, $p_1 = p_2 = 25$						$\alpha = 0.05$, $p_1 = p_2 = 50$					
0.1	**0.0526**	0.1002	0.2424	0.4704	0.7022	0.8908	0.0444	0.1090	0.3066	0.5848	0.8296	0.9516
0.2	0.1002	0.0466	0.0870	0.2336	0.4564	0.7194	0.1090	0.0478	0.1078	0.2914	0.5726	0.8314
0.3	0.2424	0.0870	0.0498	0.0898	0.2288	0.4898	0.3066	0.1078	0.0454	0.1024	0.2940	0.6026
0.4	0.4704	0.2336	0.0898	0.0470	0.0974	0.2540	0.5848	0.2914	0.1024	**0.0528**	0.1020	0.3222
0.5	0.7022	0.4564	0.2288	0.0974	**0.0534**	0.1002	0.8296	0.5726	0.2940	0.1020	0.0506	0.1158
0.6	0.8908	0.7194	0.4898	0.2540	0.1002	0.0498	0.9516	0.8314	0.6026	0.3222	0.1158	**0.0518**
	$\alpha = 0.10$, $p_1 = p_2 = 25$						$\alpha = 0.10$, $p_1 = p_2 = 50$					
0.1	0.1006	0.1742	0.3568	0.5994	0.7998	0.9416	0.0952	0.1798	0.4304	0.7062	0.8968	0.9748
0.2	0.1742	0.0964	0.1520	0.3500	0.5886	0.8154	0.1798	0.0928	0.1740	0.4072	0.6836	0.8922
0.3	0.3568	0.1520	**0.1056**	0.1602	0.3430	0.6152	0.4304	0.1740	0.0954	0.1722	0.4070	0.7088
0.4	0.5994	0.3500	0.1602	0.0970	0.1718	0.3700	0.7062	0.4072	0.1722	**0.1016**	0.1800	0.4440
0.5	0.7998	0.5886	0.3430	0.1718	**0.1074**	0.1746	0.8968	0.6836	0.4070	0.1800	0.0952	0.1890
0.6	0.9416	0.8154	0.6152	0.3700	0.1746	0.0976	0.9748	0.8922	0.7088	0.4440	0.1890	0.1006

The above results are based on 5000 simulation runs with 5000 null samples generated in each simulation run.
Note: The value in Bold is obviously larger than the stated nominal level α. The ratio of biased point over total is $15/48(< 0.5)$, which means LCP test method is less biased.