The Pennsylvania State University

The Graduate School

The Eberly College of Science

HETEROSCEDASTIC UNBALANCED NESTED DESIGNS AND FULLY NONPARAMETRIC ANALYSIS OF COVARIANCE

A Dissertation in

Statistics

by

Shu-Min Liao

© 2009 Shu-Min Liao

Submitted in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

August 2009

The dissertation of Shu-Min Liao was reviewed and approved* by the following:

Michael G. Akritas
Professor of Statistics
Dissertation Advisor, Chair of Committee

Steven F. Arnold
Professor of Statistics

Francesca Chiaromonte
Associate Professor of Statistics and Public Health Sciences

Vernon M. Chinchilli
Distinguished Professor and Chair of Public Health Sciences, and Professor of Statistics

Bruce G. Lindsay
Willaman Professor of Statistics
Head of the Department of Statistics

*Signatures are on file in the Graduate School.

Abstract

Analysis of variance is a corner stone of statistical applications. The classical asymptotic results were built either under the normality and homoscedasticity assumptions, or on cases when the numbers of factor levels are all fixed. However, the past decade has witnessed the generation of large data sets which involve a multitude of factor levels while the number of replications per factor combination is very small. The asymptotic theory is considerably more complicated when testing against those high-dimensional alternatives.

In the first part of this thesis, we consider the problem of testing for the sub-class effect in the unbalanced two-fold nested models with a large number of sub-classes. It is shown that the classical F-statistic is very sensitive to departures from homoscedasticity, even in balanced designs. We propose new testing procedures to accommodate heteroscedasticity, and the asymptotic distributions of the proposed test statistics, both under the null and local alternative hypotheses, are established. Simulation studies examine the finite sample performance of the proposed statistics and the competing classical F-test. Two real data sets are analyzed and ramifications of these results to the hypothesis of no covariate effect in the analysis of covariance are discussed, which leads to a more sophisticated approach described in the second part of the thesis. Testing for the class effect is also investigated.

In the second part of this thesis, we introduce a new approach for testing the covariate effect in the context of the fully nonparametric ANCOVA model which capitalizes on the connection to the testing problems in nested designs. The basic idea behind the proposed method is to think of each distinct covariate value as a level of a sub-class nested in each group/class. A projection-based tool is developed to obtain a new class of quadratic forms, whose asymptotic behavior is then studied to establish the limiting distributions of the proposed test statistic under the null hypothesis and local alternatives. Simulation studies show that this new method, compared with existing alternatives, has better power properties and achieves the nominal level under violations of the classical assumptions. Three data sets are analyzed, and asymptotic results concerning testing for the covariate-adjusted group effect are also included.

Table of Contents

List of Figures

List of Tables

Acknowledgments

First of all, I would like to express my deep and hearty gratitude to my dissertation advisor, Dr. Michael G. Akritas, for his extraordinary insights and encouraging guidance. Without his enthusiasm, direction, and motivation, it would be impossible for me to develop such an understanding of the subject and further complete this dissertation. He has been a great academic *father* who always shares his invaluable knowledge and experiences generously.

In addition, my warm and sincere thanks are given to the other committee members, Dr. Steven F. Arnold, Dr. Francesca Chiaromonte, and Dr. Vernon M. Chinchilli. Their consistent assistance and valuable comments have surely had a remarkable influence on my research work and entire career. I am also grateful to Dr. James L. Rosenberger for introducing the Pennsylvania State University to me and being such a good mentor in all these years; to Dr. Bruce G. Lindsay, my initial academic advisor, for his continuous support and wise/honest suggestions from the very beginning to now; to Dr. Thomas P. Hettmansperger for impassioning me to be a statistician; to Dr. David R. Hunter for teaching me "Asymptotic Tools"; to Dr. Bing Li for teaching me "Statistical Inference"; to Dr. Runze Li for teaching me "Probability Theory"; and to Mr. Robert F. Heckard for sharing with me his experiences and wisdom in statistical education. The financial support of the Department of Statistics at the Pennsylvania State University is gratefully acknowledged, too.

Many of my peers and friends have played important roles in my life at Penn State. I am especially thankful to Trent Gaugler and Min Hee Kim for being my academic *brother* and *sister*; to Yeojin Chung, Kion Kim, Jianping Sun, Mina Yoo, Hyang Min Lee, Juyoun Lee, Min Kyung Kim, Lu Zhang, Yijia Feng and Heather Ma for their friendship and help when my family was not around me; and to Yuexiao Dong and Bo Kai for being my "battle companions" and fighting against a bad job market together in the past year. To all the others who assisted me in any respect during the completion of this thesis, I offer my best regards and blessings.

Among all, I am greatly in debt to my husband, Daeyoung Kim, for bringing so much happiness to me and for being such a wonderful partner in life. His understanding, lasting encouragement and mental support are surely the main power resources which push me forward in these years. I also truly appreciate my parents-in-law, Hyung Seok Kim and Young Soon Baek, for loving me as their own and for praying for me wholeheartedly, and consider myself very lucky because of my brother, Chung-Ching Liao.

Lastly, I would like to dedicate this dissertation to my parents, Hui-Sung Liao and Tsui-Feng Tsai, whose love, sacrifices and unconditional support make this dissertation and all good in my life possible.

Chapter 1

Introduction

1.1 Heteroscedastic Unbalanced Two-fold Nested Model

The classical ANOVA model assumes that the error terms are i.i.d. normal, in which case F-statistics have certain optimality properties (cf. Arnold (1981), Chapter 7). Arnold (1980) showed that the classical F-test is robust to the normality assumption if the sample sizes are large while the number of factor levels or groups is small. The past decade has witnessed the generation of large data sets, involving a multitude of factor levels, in several areas of scientific investigation. For example, in agricultural trials it is not uncommon to see a large number of treatments with a small number of replications per treatment. Another application arises in certain type of microarray data in which the nested factor corresponds to a large number of genes. As a consequence, testing in designs with a large number of factor levels has attracted considerable attention.

The asymptotic theory of inference is considerably more complicated when the number of parameters increases with the sample sizes. The seminal paper by Neyman and Scott (1948) highlights these difficulties. See also Andersen (1970), Portnoy (1985), Fan and Lin (1998), Simons and Yao (1999), Li et al. (2003) and Hall et al. (2005) for some representative publications. Li et al. (2003) distinguish two types of frameworks for the development of asymptotic theory for high-dimensional data: the Neyman-Scott framework, where the sample sizes remain fixed while the number of parameters tend to infinity, and the rectangular array framework, where both sample sizes and number of parameters tend to infinity. The asymptotic theory in the present paper falls under the Neyman-Scott framework.

Testing in factorial designs with a large number of factor levels appears to have been initiated by Brownie and Boos (1994) who, however, used a specialized technique applicable only to a few designs with independent data. More general approaches for finding the asymptotic distribution of F-statistics, which are of the form $F = MST/MSE$, were developed in Akritas and Arnold (2000), Bathke (2002), and Akritas and Papadatos (2004). Wang and Akritas (2006) applied the Akritas and Papadatos (2004) approach to two-way designs, Gupta et al. (2006) consider designs with multivariate data, while Wang and Akritas (2004) and Bathke and Harrar (2008) consider methods based on ranks. When the degrees of freedom of both the numerator and the denominator of F-statistics tend to infinity, inference is based on the asymptotic distribution of $F - 1$ (with some scaling that depends on the number of factor levels). In all cases known to the authors, the asymptotic distribution of $F - 1$ is normal and the test procedure rejects the null hypothesis at level α when $F - 1$ is larger than the $100(1 - \alpha)$th percentile of its limiting distribution. This is because under the alternative $E(MST) > E(MSE)$.

In this thesis we study the two-fold nested design. The motivating application comes from the Mussel Watch Project of the National Oceanic and Atmospheric Administration (NOAA), which monitors chemical and biological contaminant trends in sediment and bivalve tissue collected from hundreds of EDAs (Estuarine Drainage Areas) in the West Coast, the East Coast (North, Middle and South Atlantic), the Gulf of Mexico, and the Great Lakes. Since each coastal region has its own EDAs, results of crossed designs are not appropriate for studying differences among the different EDAs. In this data set, the number of EDAs within each coastal region is relatively large, ranging from 30-60, while the cell sizes within each sub-class is small. While normality and homoscedasticity are difficult to ascertain with small sample sizes, Figure 2.1 suggests that these assumptions are violated. Thus, there is need for an asymptotic theory that accommodates these features.

It is known that the classical, normality-based F-test is sensitive to departures from the homoscedasticity assumption, especially when the design is unbalanced. For example, based on $10,000$ simulated replications from an unbalanced, between-classes heteroscedastic nested design (i.e. homoscedasticity within each class, but not between classes), the classical F-test achieved α level of 0.57 at nominal $\alpha = 0.05$, as reported in Table3.2. This simulation study used $r = 5$ classes, $(c_1,\ c_2,\ c_3,\ c_4,\ c_5) = (2,\ 4,\ 5,\ 6,\ 8)$ sub-classes and $(2, 2)$, $(1, 2, 2, 2)$, $(1, 5, 1, 3, 1)$, $(12, 8, 13, 12, 10, 11)$, $(5, 2, 1, 3, 3, 1, 2, 2)$ cell sample sizes in each sub-class. For the same setting, the proposed unweighted (between-classes) heteroscedastic test procedure (based on Theorem 2.2.2) achieved an α level of 0.115. For the same setting but with larger number of sub-classes ($(c_1,\ c_2,\ c_3,\ c_4,\ c_5) = (50,\ 75,\ 100,\ 125,\ 150)$), the proposed procedure achieved an α level of 0.067, while the F-test rejected 100% of the time. In fact, for the nested design we consider, even under homoscedasticity, the classical F-test is not asymptotically valid in the unbalanced design if the cell sizes are small, unless the model is normal. More details are discussed in Chapters 2–4.

1.2 Fully Nonparametric Analysis of Covariance and Fully Nonparametric Hypotheses

For a k-group (one-way) analysis of covariance (ANCOVA) model, let (X_{ij}, Y_{ij}) denote the paired covariate and the response variables for the j-th observation in the i-th group, $i = 1, \cdots, k$, $j = 1, \cdots, n_i$. The classical ANCOVA model specifies that, conditionally on $X_{ij} = x$,

$$Y_{ij} = \mu + \alpha_i + \delta_i\ x + e_{ij}, \tag{1.2.1}$$

where e_{ij} are independent and identically distributed normal errors. Without the normality assumption, model (1.2.1) is known as the semiparametric ANCOVA.

In this thesis, we consider the nonparametric ANCOVA model of Akritas et al. (2000), which assumes only that, conditionally on $X_{ij} = x$, the distribution of Y_{ij} depends on i and x:

$$Y_{ij}|X_{ij} = x \sim F_{ix}. \tag{1.2.2}$$

As pointed out in their paper, this model does not place any assumption on normality, homoscedasticity, linearity and additivity, so they called it “completely nonparametric” or “fully nonparametric” as used in Akritas and Arnold (1994) when they first introduced this concept. Also note that this model allows ordinal categorical covariates and its model interpretation is scale-free.

Next, choose distribution functions $G_i(x),\ i = 1, \cdots, k$, and define

$$\bar{F}_{i\cdot}^{G_i}(y) = \int F_{ix}(y)\ dG_i(x). \tag{1.2.3}$$

If X_{ij}'s are random, one can think of G_i as the distribution function of X_{ij} in group i, and $\bar{F}_{i\cdot}^{G_i}$ as the marginal distribution function of Y_{ij} in the same group i. Note that the definition in (1.2.3) uses the individual $G_i(x)$, instead of the overall $G(x)$, so it allows the covariate to have different distributions and different supports for different groups.

Under this setting, there are two hypotheses of interest:

$$\text{No covariate-adjusted group effect} \iff H_0:\ \bar{F}_{i\cdot}^{G_i} \text{ does not depend on } i; \tag{1.2.4}$$

$$\text{No covariate effect} \iff H_0:\ F_{ix} \text{ does not depend on } x. \tag{1.2.5}$$

Note that our null hypothesis of no covariate-adjusted group effect, as defined in (1.2.4), is different from the one used by Akritas et al. (2000):

$$H_0:\ F_{i\cdot}(y) \text{ does not depend on } i, \quad \text{where } F_{i\cdot}(y) = \int F_{ix}(y)\ dG(x). \tag{1.2.6}$$

The distribution function used in $F_{i\cdot}(y)$ is the overall $G(x)$. Compare with $F_{i\cdot}(y)$, we believe that $\bar{F}_{i\cdot}^{G_i}(y)$ has at least two advantages:

1. $\bar{F}_{i\cdot}^{G_i}(y)$ is a more natural way to define the average effect of the covariate variable on the response variable, because it allows the covariate X_{ij} to have different ranges in different groups, while $F_{i\cdot}(y)$ forces the covariate to have the same range, which may cause the issue of missing values in applications.

2. In many cases, it makes more sense to compare $\bar{F}_{i\cdot}^{G_i}(y)$, instead of $F_{i\cdot}(y)$ when

testing the covariate-adjusted group effect. Take a simple medical study as an example. Suppose that the response is the reduction in blood pressure, while the covariate is the baseline measurement. Suppose that the group effect of interest is *ethnicity*. Since different races represent different *populations*, it might not be reasonable to assume that the covariate variables for different ethnicity groups come from a common distribution $G(x)$. As a consequence, the definition of $F_{i\cdot}(y)$ itself may be questionable under those cases and testing hypotheses based on it may also make no sense.

Conditionally on $X_{ij} = x$, we can further decompose F_{ix} as follows:

$$F_{ix}(y) = M(y) + A_i(y) + D_i(y;x),\ \ i = 1, \cdots, k, \tag{1.2.7}$$

where

$$\sum_{i=1}^{k}\sum_{j=1}^{n_i} A_i(y) = \sum_{i=1}^{k} n_i A_i(y) = 0, \forall y; \qquad \int D_i(y;x)\ dG_i(x) = 0, \forall i,\ \forall y.$$

Thus, letting $N = \sum_i n_i$,

$$\begin{aligned} M(y) &= \frac{1}{N}\sum_{i=1}^{k} n_i \bar{F}_{i\cdot}^{G_i}; \\ A_i(y) &= \bar{F}_{i\cdot}^{G_i} - M(y); \\ D_i(y;x) &= F_{ix}(y) - A_i(y) - M(y). \end{aligned}$$

In this unique decomposition, the functions $M(y)$, $A_i(y)$ and $D_i(y;x)$ can be thought of as the overall effect, the covariate-adjusted group effect of group i, and the covariate effect with the value x, respectively. The null hypothesis (1.2.4) of no covariate-adjusted group effect can then be equivalently rewritten as

$$H_0(A):\ \ A_i(y) = 0 \quad \text{for all } i \text{ and all } y, \tag{1.2.8}$$

while the null hypothesis (1.2.5) of no covariate effect can be restated as

$$H_0(D): \; D_i(y;x) = 0 \quad \text{for all } i, \text{ all } x \text{ and all } y. \tag{1.2.9}$$

Since these hypotheses are clearly invariant under monotone transformations of the response and do not depend on any modeling assumptions, they are "fully nonparametric" as well. For the importance of test procedures being invariant under monotone transformations, see Patel and Hoel (1973), Akritas and Arnold (1994), Akritas et al. (1997) and references therein.

1.3 Connections between Two Models

The idea for constructing test statistics for the null hypotheses in the fully nonparametric ANCOVA model, (1.2.8) and (1.2.9), is inspired by the similarity of model (1.2.7) to the model for the two-fold nested design, with the group variable corresponding to the class factor and the covariate variable corresponding to the sub-class factor. What makes the connection between these two models feasible is that in the context of the fully nonparametric ANCOVA model, the covariate effect is in fact not modeled. In spite of this conceptual similarity, however, the classical asymptotic test procedures in the two-fold nested model were driven by a large number of replications on fixed numbers of levels of the class and sub-class factors, while the classical ANCOVA built its standard asymptotics with a large number of observations per group, which makes the number of *levels* of the 'covariate factor' tend to infinity. Therefore, the asymptotic results in the classical two-fold nested model are not directly applicable to the classical ANCOVA model.

To construct a link between these two models using the conceptual similarity stated above, we first consider the asymptotic test procedures using the Neyman-Scott framework according to which the number of sub-class levels tends to infinity with the number of replications being at least two, but otherwise allowed (but not required) to remain fixed. Note that, however, there is typically only one observation per covariate value, if assuming no ties on the covariate. One simple way to solve this 'sparseness' issue, due to the continuity of the covariate, is to *discretise* the covariate factor. More specifically, one can simply partition the paired observations $(X_{ij},\ Y_{ij})$ in the same group i into n_i/w non-overlapping *windows*, according to their sorted covariate values, so that there are only a small number w of observations per window. Those artificially-created non-overlapping small windows serve as the sub-classes in the two-fold nested model setting, and the asymptotic results derived for the two-fold nested model with a large number of sub-classes (and a small number of observations in each sub-class) can then apply directly. We call this "non-overlapping windows approach", and a simple application of this approach can be found in Section 2.4.2.

Another way to remedy this sparseness issue, or the issue of lack of replications on each of the covariate values, is to consider a window W_{ij} around each X_{ij} consisting of the w nearest covariate values from group i. That is, we utilize smoothness assumptions to augment the observed data in ANCOVA to construct a large number of *overlapping* local windows in order to form an artificial two-fold nested model with each sub-class having w replications. Under the assumption that the conditional distribution F_{ix} of the response at a given covariate value x changes smoothly with x, simultaneously taking into account the responses having covariate values close to the given x-value can magnify the information available for F_{ix}, and hence enlarge the power of the corresponding test procedures. This approach is called "overlapping windows approach" in this thesis, and more details about how to implement this approach can be found in Chapter 5. Since the overlapping windows in this artificial two-fold nested model have common observations with other windows close by, the asymptotic results obtained in Chapters 2–4 do not apply. A new set of asymptotic approximation techniques using the projection principle is then introduced in Section 5.2 to accommodate such kind of augmented dependence in our design. It can be shown that the "overlapping windows approach" is indeed more powerful than the naive "non-overlapping windows approach" (see Section 5.3.2 for some numerical evidence).

1.4 Thesis Outline

The rest of this thesis is organized as follows. The first part comprises three chapters. In Chapter 2, we consider the problem of testing for the sub-class effect in the unbalanced two-fold nested model, when the number of of sub-classes is large while the number of classes and the number of observations per sub-class remain fixed. The designs under homoscedasticity and under heteroscedasticity are all investigated. Appropriate test procedures are developed for different designs, and the asymptotic distributions of the proposed test statistics, both under the null hypothesis and local alternatives, are established. Simulation studies examine the finite sample performance of the proposed test procedures and the competing classical F-test. Two real data sets are analyzed: one is from a project monitoring the chemical contaminants in the coastal areas, and the other is used to illustrate ramifications of these results to the hypothesis of no covariate effect in the nonparametric analysis of covariance.

In Chapter 3, the extension of the results in Chapter 2 to the case when the number of classes and the number of sub-classes are both large is introduced. We derive the asymptotic theories of the proposed test procedures both under the null hypothesis and local alternatives, and examine their performances using some simulation studies. An application of this methodology for testing the lack-of-fit in regression is also discussed.

In Chapter 4, testing for the class effect in the two-fold nested model with a large number of sub-classes is considered. Asymptotic results are presented for the hypotheses with arbitrary weights on sub-classes. Both homoscedastic designs and heteroscedastic designs are included.

In the second part of this thesis, which consists of two chapters, we investigate fully nonparametric analysis of covariance, as introduced in Section 1.2. In Chapter 5, we establish a new methodology to test for the covariate effect by utilizing the connection to testing in the nested models. Although the proposed test statistic has a form similar to those for testing the sub-class effect in nested models when the number of sub-classes is large, the asymptotic derivations of its limiting distributions under the null hypotheses and local alternatives involves a different class of quadratic forms, and hence needs a new asymptotic tool based on the projection principle which we introduce. Simulation studies are performed to demonstrate the proposed method and compare its properties with existing alternatives. Three real data sets are analyzed.

In Chapter 6, we further consider testing for the covariate-adjusted group effect in the fully nonparametric analysis of covariance model. Asymptotic results for the proposed test statistic are given and proved.

Finally, we complete this thesis with a brief summary of our work and some possible future research topics in Chapter **??**.

Chapter 2

Testing for the Sub-class Effect in Two-fold Nested Model when the number of sub-classes is large

The purpose of the present chapter is to provide valid test procedures for the sub-class effect which can perform well in unbalanced and/or heteroscedastic designs when the number of sub-classes is large. The proposed test statistics are of the general form $MST - MSE$, but the MSE is chosen so that, under the null hypothesis, $E(MSE) = E(MST)$. Note that this last relation does not hold under heteroscedasticity for the classical definition MSE. The basic asymptotic technique we apply is based on finding the joint limiting distribution of (MST, MSE) through a suitable representation by a simpler, asymptotically equivalent, random vector.

The rest of this chapter is organized as follows. Section 2.1 describes the statistical model for the unbalanced heteroscedastic two-fold nested design, and reviews the classical F-test procedure for the hypothesis of no sub-class effect. In Section 2.2.1 we present the asymptotic theory for the classical F-statistic in the homoscedastic case. In Section 2.2.2 we propose two test statistics (one weighted and one unweighted) for the between-classes heteroscedastic model, and present their asymptotic distributions. In Section 2.2.3 we propose an unweighted test statistic for the model with general heteroscedasticity, and present its asymptotic theory. Simulation results are presented in Section 2.3, while the lead concentration data set from the Mussel Watch Project is analyzed in Section 2.4.1. A ramification of these results for analysis of covariance is illustrated in Section 2.4.2 using the Acid Rain data from the National Atmospheric Deposition Program (NADP). Finally, proofs of the results presented in Section 2.2 are provided in Section 2.5.

2.1 The Statistical Model and the Test Statistic

In the *general* unbalanced two-fold nested model, we observe

$$Y_{ijk} = \mu_{ij} + \sigma_{ij} \cdot e_{ijk}, \quad i = 1, \cdots, r; \quad j = 1, \cdots, c_i; \quad k = 1, \cdots, n_{ij}, \qquad (2.1.1)$$

where the μ_{ij} and σ_{ij} are bounded and e_{ijk} are independent with

$$E(e_{ijk}) = 0, \quad Var(e_{ijk}) = 1. \qquad (2.1.2)$$

Note that the general model (2.1.1), (2.1.2) does not assume that the errors e_{ijk} are normally, or even identically, distributed. Thus, ordinal discrete data are included in this formulation. Let

$$C = \sum_{i=1}^{r} c_i, \quad n_{i\cdot} = \sum_{j=1}^{c_i} n_{ij}, \quad N_C = \sum_{i=1}^{r} \sum_{j=1}^{c_i} n_{ij} = \sum_{i=1}^{r} n_{i\cdot}.$$

The means μ_{ij} are typically decomposed as

$$\mu_{ij} = \mu + \alpha_i + \delta_{ij}, \qquad (2.1.3)$$

where we assume that

$$\sum_{i=1}^{r} n_{i\cdot} \alpha_i = 0 \text{ and } \sum_{j=1}^{c_i} n_{ij} \delta_{ij} = 0, \forall i.$$

In this chapter, we are mainly interested in testing H_0: $\delta_{ij} = 0$ (no sub-class effect). Let

$$MS\delta = \frac{\sum_{i=1}^{r}\sum_{j=1}^{c_i} n_{ij}(\bar{Y}_{ij\cdot} - \bar{Y}_{i\cdot\cdot})^2}{C - r}, \tag{2.1.4}$$

$$MSE = \frac{\sum_{i=1}^{r}\sum_{j=1}^{c_i}\sum_{k=1}^{n_{ij}}(Y_{ijk} - \bar{Y}_{ij\cdot})^2}{N_C - C}, \tag{2.1.5}$$

where $\bar{Y}_{ij\cdot}$ and $\bar{Y}_{i\cdot\cdot}$ are the corresponding unweighted means of Y_{ijk} within each sub-class and within each class, i.e.

$$\bar{Y}_{ij\cdot} = \frac{1}{n_{ij}}\sum_{k=1}^{n_{ij}} Y_{ijk}, \quad \bar{Y}_{i\cdot\cdot} = \frac{1}{n_{i\cdot}}\sum_{j=1}^{c_i}\sum_{k=1}^{n_{ij}} Y_{ijk} = \frac{1}{n_{i\cdot}}\sum_{j=1}^{c_i} n_{ij}\bar{Y}_{ij\cdot}.$$

Then, the usual F-test statistic for testing H_0: $\delta_{ij} = 0$ is

$$F_C^{\delta} = \frac{MS\delta}{MSE}. \tag{2.1.6}$$

Under the normal homoscedastic model, i.e. if e_{ijk} are assumed to be iid $N(0,1)$ and all $\sigma_{ij} = \sigma$, we have that

$$F_C^{\delta} \sim F_{C-r,\, N_C - C}, \text{ under } H_0 : \delta_{ij} = 0. \tag{2.1.7}$$

In what follows we examine the robustness of this procedure to departures from the assumptions of normality and homoscedasticity as the number of sub-classes gets large.

In all that follows we will use the notation

$$\bar{n}_{ic_i} = \frac{1}{c_i}\sum_{j=1}^{c_i} n_{ij} = \frac{n_{i\cdot}}{c_i}, \quad \underline{n}_{ic_i} = \frac{1}{c_i}\sum_{j=1}^{c_i}\frac{1}{n_{ij}}.$$

All results, except those of Section 5, are derived under the following conditions on the sample sizes: There exist numbers $\lambda_i \in (0,1)$, $\bar{n}_i > 1$, and $\underline{n}_i \in (0,\infty)$ such that as $\min(c_i) \to \infty$,

$$\sqrt{C}\left(\frac{c_i}{C} - \lambda_i\right) \to 0, \quad \sqrt{c_i}\,(\bar{n}_{ic_i} - \bar{n}_i) \to 0, \quad \underline{n}_{ic_i} \longrightarrow \underline{n}_i. \tag{2.1.8}$$

Finally, we define

$$\bar{n} = \sum_{i=1}^{r} \lambda_i \bar{n}_i. \tag{2.1.9}$$

2.2 Main Results

2.2.1 Homoscedastic Designs

In this subsection we consider the unbalanced two-fold nested design with homoscedastic errors and derive the asymptotic distribution of F_C^δ, defined in (2.1.6). As a corollary of Theorem 2.2.1 below, we obtain that the usual, normal-based, F-test procedure is not robust to departures from the normality assumption even under homoscedasticity.

Theorem 2.2.1. *Consider the model and assumptions given in (2.1.1) with $\sigma_{ij} = \sigma$, (2.1.2), (2.1.8) and the decomposition of the means given in (2.1.3). In addition assume that*

$$E(e_{ijk}^3) = 0, \quad E(e_{ijk}^4) = \kappa_i, \textit{ and } E|e_{ijk}|^{4+2\epsilon} < \infty \textit{ for some } \epsilon > 0.$$

Then, under alternatives δ_{ij} which satisfy

$$\sqrt{c_i}\left(\frac{1}{c_i}\sum_{j=1}^{c_i} n_{ij}\frac{\delta_{ij}^2}{\sigma^2} - \theta_i\right) \to 0, \quad \textit{for some numbers } \theta_i \in (0,\infty),$$

as $\min(c_i) \to \infty$ while r, n_{ij} stay fixed,

$$\sqrt{C}\left(F_C^\delta - (1+\theta)\right) \xrightarrow{d} N\left(0, \Sigma_s\right), \tag{2.2.1}$$

where λ_i, $\bar{n}_i$, $\underline{n}$, and $\bar{n}$ are as given in (2.1.8) and (2.1.9), $\theta = \sum_{i=1}^r \lambda_i\theta_i$, and

$$\begin{aligned}\Sigma_s &= 2 + 4\theta + \frac{2(1+\theta)^2}{\bar{n}-1} \\ &\quad + \sum_{i=1}^{r}\left[(\kappa_i - 3)\lambda_i \frac{(2\theta+\bar{n})(\bar{n}\underline{n}_i - 1) + (2\theta+1)(\bar{n}_i - \bar{n}) + \theta^2(\bar{n}_i + \underline{n}_i - 2)}{(\bar{n}-1)^2}\right].\end{aligned}$$

Under the null hypothesis $H_0 : \delta_{ij} = 0$, which results in $\theta = 0$, we then have

$$\sqrt{C}\left(F_C^\delta - 1\right) \xrightarrow{d} N\left(0,\, 2 + \frac{2}{\bar{n}-1} + \sum_{i=1}^{r}\left[\frac{(\kappa_i - 3)\lambda_i(\bar{n}^2\underline{n}_i - 2\bar{n} + \bar{n}_i)}{(\bar{n}-1)^2}\right]\right). \qquad (2.2.2)$$

Corollary 2.2.1. *Under the model and assumptions of Theorem 2.2.1, the classical, normality-based, F-test procedure for the hypothesis $H_0 : \delta_{ij} = 0$, shown in (2.1.7), is* not *asymptotically valid when the model is not normal, unless $n_{ij} = n$.*

It can be shown that if normality holds, the test procedure implied by Theorem 2.2.1 is asymptotically equivalent to the classical F-test procedure under $H_0 : \delta_{ij} = 0$.

2.2.2 Between-classes Heteroscedastic Designs

Two possible statistics

In this subsection we consider the heteroscedastic unbalanced two-fold nested design, but assume we have *between-classes heteroscedasticity*, i.e. $\sigma_{ij} = \sigma_i$ in the relation (2.1.1). It can be shown that if the design is unbalanced, then, under heteroscedasticity, it is no longer true that $E(MSE) = E(MS\delta)$ under the null hypothesis $H_0 : \delta_{ij} = 0$. Thus it is clear that the usual F-test procedure is not valid even under normality. In this section, we will first introduce two possible test statistics, one *unweighted* and one *weighted*.

The unweighted statistic simply replaces MSE with

$$MSE^* = \frac{1}{C-r}\sum_{i=1}^{r}\frac{c_i - 1}{n_{i\cdot} - c_i}\sum_{j=1}^{c_i}\sum_{k=1}^{n_{ij}}\left(Y_{ijk} - \bar{Y}_{ij\cdot}\right)^2.$$

It is easily seen that MSE^* satisfies $E(MSE^*) = E(MS\delta)$ under the null hypothesis. Thus, the unweighted statistic is

$$F_C^* - 1 = \frac{MS\delta}{MSE^*} - 1. \qquad (2.2.3)$$

It is easy to verify that, in the balanced case, $F_C^* = F_C^\delta$, where F_C^δ is the classical F-statistic given in (2.1.6).

The weighted statistic will be derived from a totally different angle. Assume first that the σ_i are known and consider the mean-error decomposition of the weighted random variables $Y'_{ijk} = Y_{ijk}/\sigma_i$:

$$Y'_{ijk} = \frac{Y_{ijk}}{\sigma_i} \triangleq \mu'_{ij} + e_{ijk}, \text{ where } \mu'_{ij} = \frac{\mu}{\sigma_i} + \frac{\alpha_i}{\sigma_i} + \frac{\delta_{ij}}{\sigma_i}. \tag{2.2.4}$$

The new means μ'_{ij} are then further decomposed (uniquely) as

$$\mu'_{ij} = \mu' + \alpha'_i + \delta'_{ij}, \text{ where } \sum_{i=1}^{r} n_{i\cdot}\alpha'_i = 0 \text{ and } \sum_{j=1}^{c_i} n_{ij}\delta'_{ij} = 0, \forall i, \tag{2.2.5}$$

where

$$\begin{aligned} \mu' &= \frac{1}{N_C}\sum_{i=1}^{r}\sum_{j=1}^{c_i} n_{ij}\mu'_{ij} = \frac{1}{N_C}\sum_{i=1}^{r} n_{i\cdot}\left(\frac{\mu+\alpha_i}{\sigma_i}\right), \\ \alpha'_i &= \sum_{j=1}^{c_i}\frac{n_{ij}}{n_{i\cdot}}\mu'_{ij} - \mu' = \frac{\mu+\alpha_i}{\sigma_i} - \mu', \\ \delta'_{ij} &= \mu'_{ij} - (\mu' + \alpha'_i) = \frac{\delta_{ij}}{\sigma_i}. \end{aligned}$$

Note that the original null hypothesis $H_0 : \delta_{ij} = 0, \ \forall i, \ j$, is equivalent to the corresponding hypothesis for the model for the weighted random variables Y'_{ijk}, namely, $H_0 : \delta'_{ij} = 0, \ \forall i, \ j$. Let $MS\delta'$, MSE', and F'_C be as defined in (2.1.4)–(2.1.6) but with Y'_{ijk} replacing Y_{ijk}, that is,

$$MS\delta' = \frac{\sum_i \sum_j n_{ij}(\bar{Y}'_{ij\cdot} - \bar{Y}'_{i\cdot\cdot})^2}{C-r} = \frac{1}{C-r}\sum_{i=1}^{r}\sum_{j=1}^{c_i} n_{ij}\left(\frac{\bar{Y}_{ij\cdot} - \bar{Y}_{i\cdot\cdot}}{\sigma_i}\right)^2, \tag{2.2.6}$$

$$MSE' = \frac{\sum_i \sum_j \sum_k (Y'_{ijk} - \bar{Y}'_{ij\cdot})^2}{N_C - C} = \frac{1}{N_C - C}\sum_{i=1}^{r}\sum_{j=1}^{c_i}\sum_{k=1}^{n_{ij}}\left(\frac{Y_{ijk} - \bar{Y}_{ij\cdot}}{\sigma_i}\right)^2, \tag{2.2.7}$$

$$F'_C = \frac{MS\delta'}{MSE'}. \tag{2.2.8}$$

Of course, the σ_i are not known and thus they need to be estimated. Let

$$S_{ij}^2 = \frac{1}{n_{ij}-1}\sum_{k=1}^{n_{ij}}(Y_{ijk}-\bar{Y}_{ij\cdot})^2, \quad S_i^2 = \frac{1}{n_{i\cdot}-c_i}\sum_j (n_{ij}-1)S_{ij}^2, \tag{2.2.9}$$

be the usual estimators of σ_{ij}^2 and σ_i^2. Note that if σ_i is replaced by S_i in MSE', the expression in (2.2.7) is identically equal to one. The proposed weighted test statistic is thus

$$\widehat{F'_C} - 1, \quad \text{where} \quad \widehat{F'_C} = \frac{1}{C-r}\sum_{i=1}^{r}\frac{1}{S_i^2}\sum_{j=1}^{c_i} n_{ij}\left(\bar{Y}_{ij\cdot}-\bar{Y}_{i\cdot\cdot}\right)^2. \tag{2.2.10}$$

Asymptotic Theory for the Unweighted Statistic

The asymptotic distribution of the unweighted statistic, $F_C^* - 1$, defined in (2.2.3), is given by Theorem 2.2.2. As a corollary to this theorem we obtain that, under heteroscedasticity, the classical F-test procedure is not valid in the balanced case (where $F_C^* = F_C^\delta$) even under normality.

Theorem 2.2.2. *Consider the model and assumptions given in Theorem 2.2.1, except that the variances $\sigma_{ij} = \sigma_i$ are allowed to vary among classes. Then, under alternatives δ_{ij} which satisfy*

$$\sqrt{c_i}\left(\frac{1}{c_i}\sum_{j=1}^{c_i} n_{ij}\frac{\delta_{ij}^2}{\sigma_i^2} - \theta_i\right) \to 0, \quad \textit{for some numbers } \theta_i \in (0,\infty),$$

as $\min(c_i) \to \infty$ *while* $r,\ n_{ij}$ *stay fixed,*

$$\sqrt{C}\left(F_C^* - (1+\theta^*)\right) \xrightarrow{d} N\left(0, \Sigma_s^*\right), \tag{2.2.11}$$

where λ_i, $\bar{n}_i$, $\underline{n}_i$ *and* $\bar{n}$ *are given in (2.1.8), and*

$$\theta^* = \frac{\theta^\sigma}{\beta}, \quad \textit{where} \quad \theta^\sigma = \sum_{i=1}^{r}\sigma_i^2\lambda_i\theta_i \quad \textit{and} \quad \beta = \sum_{i=1}^{r}\sigma_i^2\lambda_i, \quad \textit{and}$$

$$\Sigma_s^* = \sum_{i=1}^{r}\frac{\lambda_i\sigma_i^4}{\beta^2}\left\{2 + 4\theta_i + \frac{2(1+\theta^*)^2}{\bar{n}_i - 1}\right.$$

$$+\frac{\kappa_i - 3}{(\bar{n}_i - 1)^2}\left[(\bar{n}_i + 2\theta^*)(\bar{n}_i\underline{n}_i - 1) + \theta^{*2}(\bar{n}_i + \underline{n}_i - 2)\right]\Bigg\}.$$

Under the null hypothesis $H_0 : \delta_{ij} = 0$, which results in $\theta^* = 0$, we then have

$$\sqrt{C}\,(F_C^* - 1) \xrightarrow{d} N\left(0, \sum_{i=1}^{r} \frac{\lambda_i \sigma_i^4}{\beta^2}\left[2 + \frac{2}{\bar{n}_i - 1} + \frac{(\kappa_i - 3)\bar{n}_i(\bar{n}_i\underline{n}_i - 1)}{(\bar{n}_i - 1)^2}\right]\right). \quad (2.2.12)$$

Corollary 2.2.2. *Under the model and assumptions of Theorem 2.2.2, if the design is balanced (i.e. $c_i = c$ and $n_{ij} = n$), then the unweighted test statistic F_C^* is equal to the classical F-test statistic F_C^δ, and as $c \to \infty$*

$$\sqrt{C}\,(F_C^* - 1) \xrightarrow{d} N\left(0, \sum_{i=1}^{r} \frac{\lambda_i \sigma_i^4}{\beta^2}\left[2 + \frac{2}{n - 1}\right]\right), \quad (2.2.13)$$

under the null hypothesis $H_0 : \delta_{ij} = 0$. Thus the classical F-test procedure based on F_C^δ, is not *asymptotically valid even when the design is balanced and normality holds.*

Asymptotic Theory for the Weighted Statistic

The asymptotic distribution of the weighted statistic, $\widehat{F_C'} - 1$, as defined in (2.2.10), is given by the following theorem.

Theorem 2.2.3. *Consider the model and assumptions given in Theorem 2.2.1, except that the variances $\sigma_{ij} = \sigma_i$ are allowed to vary among classes. Then, under alternatives δ_{ij} which satisfy*

$$\sqrt{c_i}\left(\frac{1}{c_i}\sum_{j=1}^{c_i} n_{ij}\frac{\delta_{ij}^2}{\sigma_i^2} - \theta_i\right) \to 0, \quad \textit{for some numbers } \theta_i \in (0, \infty),$$

as $\min(c_i) \to \infty$ *while r, n_{ij} stay fixed,*

$$\sqrt{C}\left(\widehat{F_C'} - (1 + \theta)\right) \xrightarrow{d} N\left(0, \hat{\Sigma}_s\right), \quad (2.2.14)$$

where λ_i, $\bar{n}_i$, *and* $\underline{n}$ *are as given in (2.1.8),* $\theta = \sum_i \lambda_i \theta_i$ *and*

$$\hat{\Sigma}_s = 2 + 4\theta + \sum_{i=1}^{r} \left[\frac{2\lambda_i}{\bar{n}_i - 1}(1+\theta_i)^2 (\kappa_i - 3)\lambda_i \frac{(2\theta_i + \bar{n}_i)(\bar{n}_i \underline{n}_i - 1) + \theta_i^2(\bar{n}_i + \underline{n}_i - 2)}{(\bar{n}_i - 1)^2} \right].$$

Under the null hypothesis $H_0 : \delta_{ij} = 0$*, which results in* $\theta = 0$*, we then have*

$$\sqrt{C}\left(\widehat{F'_C} - 1\right) \xrightarrow{d} N\left(0, \sum_{i=1}^{r} \lambda_i \left[2 + \frac{2}{\bar{n}_i - 1} + \frac{(\kappa_i - 3)\bar{n}_i(\bar{n}_i \underline{n}_i - 1)}{(\bar{n} - 1)^2}\right]\right). \quad (2.2.15)$$

Corollary 2.2.3. *Consider the model and assumptions of Theorem 2.2.3. If the design is balanced (*$c_i = c$ *and* $n_{ij} = n$*), then under the null hypothesis* $H_0 : \delta_{ij} = 0$*,*

$$\sqrt{C}\left(\widehat{F'_C} - 1\right) \xrightarrow{d} N\left(0,\, 2 + \frac{2}{n-1}\right), \quad \textit{as } c \to \infty.$$

2.2.3 General Heteroscedastic Designs

In this subsection we consider the general unbalanced heteroscedastic two-fold nested model, as defined in (2.1.1). As remarked in the context of between-classes heteroscedasticity, the relation $E(MSE) = E(MS\delta)$ is no longer true if the design is unbalanced. Moreover, the usual F-test procedure is not valid even in the balanced case under normality.

In the previous subsection, we introduced two statistics, unweighted and weighted. Conceptually, we should be able to extend both ideas to the present model which allows general heteroscedasticity. However, the weighted statistic is very unstable when estimation of the σ_{ij} is based on small sample sizes. In fact, the asymptotic theory of the weighted test statistic in this case requires the sample sizes to also tend to infinity. Therefore, we will only consider the unweighted statistic in this subsection.

The idea of the unweighted statistic is to replace MSE by a different linear combination of the cell sample variances in order to match the expected value of $MS\delta$ under the null hypothesis. This achieved by replacing MSE by MSE^{**}, defined as

$$MSE^{**} = \frac{1}{C - r} \sum_{i=1}^{r} \sum_{j=1}^{c_i} \left(1 - \frac{n_{ij}}{n_{i\cdot}}\right) S_{ij}^2, \quad (2.2.16)$$

where S_{ij}^2 is as given in (2.2.9). The unweighted statistic for the general heteroscedastic case would be then defined as

$$F_C^{**} - 1 = \frac{MS\delta}{MSE^{**}} - 1. \tag{2.2.17}$$

It is easy to verify that, in the balanced case, $F_C^{**} = F_C^* = F_C^\delta$, where F_C^δ is the classical F-statistic given in (2.1.6) and F_C^* is the unweighted statistic under between-classes heteroscedastic designs. The asymptotic distribution of the unweighted statistic $F_C^{**} - 1$ is given by the following theorem.

Theorem 2.2.4. *Consider the model and assumptions given in (2.1.1), (2.1.2), and the decomposition of the means given in (2.1.3). In addition, assume that there exist* κ_{ij}, λ_i, a_{1i}, a_{2i}, b_{1i}, b_{2i} *and* b_{3i} *such that, as* $\min(c_i) \to \infty$,

$$E(e_{ijk}^3) = 0, \quad E(e_{ijk}^4) = \kappa_{ij}, \text{ and } E|e_{ijk}|^{4+2\epsilon} < \infty \text{ for some } \epsilon > 0;$$

$$\sqrt{C}\left(\frac{c_i}{C} - \lambda_i\right) \to 0, \quad \sqrt{c_i}\left(\frac{1}{c_i}\sum_{j=1}^{c_i}\sigma_{ij}^2 - a_{1i}\right) \to 0, \quad \frac{1}{n_{i\cdot}}\sum_{j=1}^{c_i} n_{ij}\sigma_{ij}^2 \longrightarrow a_{2i}, \tag{2.2.18}$$

$$\frac{1}{c_i}\sum_{j=1}^{c_i}\sigma_{ij}^4 \longrightarrow b_{1i}, \quad \frac{1}{c_i}\sum_{j=1}^{c_i}\frac{\sigma_{ij}^4}{n_{ij}-1} \longrightarrow b_{2i}, \quad \frac{1}{c_i}\sum_{j=1}^{c_i}\frac{\sigma_{ij}^4(\kappa_{ij}-3)}{n_{ij}} \longrightarrow b_{3i}.$$

Then, under alternatives δ_{ij} *which satisfy, as* $\min(c_i) \to \infty$ *while* r, n_{ij} *stay fixed,*

$$\sqrt{c_i}\left(\frac{1}{c_i}\sum_{j=1}^{c_i} n_{ij}\delta_{ij}^2 - \theta_{1i}\right) \to 0, \quad \frac{1}{c_i}\sum_{j=1}^{c_i} n_{ij}\delta_{ij}^2\sigma_{ij}^2 \longrightarrow \theta_{2i},$$

for some numbers $\theta_{1i} \in (0, \infty)$ *and* $\theta_{2i} \in (0, \infty)$, *we have*

$$\sqrt{C}\left(F_C^{**} - (1 + \theta^{**})\right) \xrightarrow{d} N\left(0, \frac{1}{a_1^2}\left[2(b_1 + b_2) + 4(\theta_2 + b_2\theta^{**}) + (2b_2 + b_3)\theta^{**2}\right]\right),$$

where

$$\theta^{**} = \frac{\theta_1}{a_1}, \quad \theta_1 = \sum_{i=1}^r \lambda_i\theta_{1i}, \quad a_1 = \sum_{i=1}^r \lambda_i a_{1i}, \tag{2.2.19}$$

$$b_1 = \sum_{i=1}^{r} \lambda_i b_{1i}, \quad b_2 = \sum_{i=1}^{r} \lambda_i b_{2i}, \quad b_3 = \sum_{i=1}^{r} \lambda_i b_{3i}, \quad \theta_2 = \sum_{i=1}^{r} \lambda_i \theta_{2i}.$$

Under the null hypothesis $H_0 : \delta_{ij} = 0$, which results in $\theta^{**} = 0$, we then have

$$\sqrt{C}\,(F_C^{**} - 1) \xrightarrow{d} N\left(0,\ \frac{2b_1 + 2b_2}{a_1^2}\right). \qquad (2.2.20)$$

2.3 Simulation Studies

In this section, simulations are used to compare the achieved sizes and/or powers of several test procedures. Let *CF* denote the classical F-test procedure, shown in (2.1.7), and *HOM*, *UW*, *WT*, *HET* denote the test procedures implied by the asymptotic results of (2.2.2), (2.2.12), (2.2.15) and (3.1.3), respectively. The procedure *CF* is compared with *HOM* for homoscedastic designs (Section 2.3.1), and with both of *UW* and *WT* for between-classes heteroscedastic designs (Section 2.3.2). In Section 2.3.3 the *HET* procedure is compared to *UW* and *HOM* for both homoscedastic and heteroscedastic designs.

For all simulations except those of Section 2.3.3, the number of classes used in all simulations is five ($r = 5$). The different combinations of numbers of sub-classes studied here, with the average $\bar{c}$ in each case, are:

- $\bar{c} = 5 \Leftrightarrow (c_1,\ c_2,\ c_3,\ c_4,\ c_5) = (2,\ 4,\ 5,\ 6,\ 8)$;
- $\bar{c} = 30 \Leftrightarrow (c_1,\ c_2,\ c_3,\ c_4,\ c_5) = (15,\ 23,\ 30,\ 37,\ 45)$;
- $\bar{c} = 100 \Leftrightarrow (c_1,\ c_2,\ c_3,\ c_4,\ c_5) = (50,\ 75,\ 100,\ 125,\ 150)$;
- $\bar{c} = 500 \Leftrightarrow (c_1,\ c_2,\ c_3,\ c_4,\ c_5) = (250,\ 375,\ 500,\ 625,\ 750)$.

The number of observations in each sub-class (n_{ij}) is generated by truncated Poisson distributions. More specifically, $n_{ij} = Z_{ij} + \upsilon_i \times I(Z_{ij} = 0)$, where I is an indicator function and $Z_{ij} \sim Poisson(\upsilon_i),\ i = 1, \cdots, 5;\ j = 1, \cdots, c_i$. The value of υ_i used in our simulations is $(\upsilon_1,\ \upsilon_2,\ \upsilon_3,\ \upsilon_4, \upsilon_5)' = (2,\ 2,\ 2,\ 12,\ 2)$. The values of the other parameters in the decomposition (2.1.3) are as follows: $\mu = 0$, $(\alpha_1, \alpha_2, \alpha_3, \alpha_4)' = (-3, -2, -1, 2)'$

and α_5 is chosen so that $\sum_i n_{i\cdot}\alpha_i = 0$. After generating the n_{ij} and fixing all parameters, we randomly generate errors e_{ijk} from one of the following five distributions: (i) **Normal**: the standard normal; (ii) **Exponen**: the exponential distribution with $\lambda = 1$; (iii) **LogNorm**: the log-normal distribution whose logarithm has mean 0 and standard deviation 1; (iv) **Mixture**: the mixture distribution defined as $U_1 \cdot X_1 + (1 - U_1) \cdot Y_1$, where $U_1 \sim$ Bernoulli($p = 0.9$), $X_1 \sim N(-1, 1)$ and $Y_1 \sim N(9, 1)$; and (v) **Multi-d**: when $r = 1,\ 2,\ 3,\ 4$, generate e_{ijk} from *Normal*, *Exponen*, *LogNorm*, and *Mixture* as described above, respectively. When $r = 5$, generate e_{ijk} from another mixture distribution defined as $U_2 \cdot X_2 + (1 - U_2) \cdot Y_2$, where $U_2 \sim$ Bernoulli($p = 0.5$), $X_2 \sim N(-3, .5)$ and $Y_2 \sim N(3, .5)$. All e_{ijk} are standardized to have mean 0 and standard deviation 1. As for the variances, we use $\sigma_{ij} = \sigma = 1,\ \forall i,\ j$ for homoscedastic designs, use $(\sigma_{ij}) = (\sigma_i) = (\sigma_1,\ \sigma_2,\ \sigma_3,\ \sigma_4,\ \sigma_5) = (1,\ 1,\ 5,\ 1,\ 1),\ \forall j$ for between-classes heteroscedastic designs, and $\sigma_{ij} = 4 \cdot I(i = 3) + 5 \cdot I(j < 0.3\ c_i) + (j/c_i),\ \forall i,\ j$, where $I(\cdot)$ is an indicator function, for general heteroscedastic designs.

2.3.1 Simulations under Homoscedastic designs

We first compare the achieved sizes of two procedures, *CF* and *HOM*, under homoscedastic designs. The first procedure, *CF*, based on the classical normality-based F-test theorem, rejects at level α if

$$F_C^{\delta} > F_{C-r,N_C-C}^{\alpha}, \tag{2.3.1}$$

where F_C^{δ} is defined in (2.1.6) and F_{C-r,N_C-C}^{α} is the $(1-\alpha)$100th percentile of the F_{C-r,N_C-C} distribution. The second procedure, *HOM*, using the asymptotic null distribution shown in (2.2.2), rejects at level α if

$$\sqrt{C}(F_C^{\delta} - 1) > \sqrt{2 + \frac{2}{\hat{\bar{n}} - 1} + \sum_{i=1}^{r}\left[\frac{(\hat{\kappa}_i - 3)\hat{\lambda}_i(\hat{\bar{n}}^2\underline{\hat{n}}_i - 2\hat{\bar{n}} + \hat{\bar{n}}_i)}{(\hat{\bar{n}} - 1)^2}\right]}\, Z_{\alpha}, \tag{2.3.2}$$

where F_C^{δ} is as before and Z_{α} is the $(1-\alpha)$100th percentile of the standard normal distribution. In addition, $\hat{\lambda}_i$, $\hat{\bar{n}}_i$, $\underline{\hat{n}}_i$, $\hat{\bar{n}}$, and $\hat{\kappa}_i$ are the empirical versions of λ_i, $\bar{n}_i$, $\underline{n}_i$, $\bar{n}$, and κ_i, namely

$$\hat{\lambda}_i = \frac{c_i}{C}, \quad \hat{\bar{n}}_i = \frac{1}{c_i}\sum_{j=1}^{c_i} n_{ij}, \quad \hat{\underline{n}}_i = \frac{1}{c_i}\sum_{j=1}^{c_i}\frac{1}{n_{ij}}, \quad \hat{\bar{n}} = \sum_{i=1}^{r}\hat{\lambda}_i\hat{\bar{n}}_i, \tag{2.3.3}$$

$$\hat{\kappa}_i = \frac{\hat{\nu}_i}{(MSE)^2}, \quad \text{where} \quad \hat{\nu}_i = \frac{1}{n_{i\cdot}}\sum_j\sum_k (Y_{ijk} - \bar{Y}_{i\cdot\cdot})^4. \tag{2.3.4}$$

It can be easily verified that, under the null hypothesis, $\hat{\kappa}_i \xrightarrow{P} \kappa_i$, as $\min(c_i) \to \infty$. The simulated sizes, based on 10,000 simulation runs, are shown in Table 3.1.

Table 2.1. Achieved α-levels over 10,000 simulation runs under homoscedastic designs at nominal $\alpha = 0.05$.

	$\overline{c} = 5$		$\overline{c} = 30$		$\overline{c} = 100$		$\overline{c} = 500$	
	CF	HOM	CF	HOM	CF	HOM	CF	HOM
Normal	0.0489	0.0839	0.0516	0.0604	0.0525	0.0581	0.0502	0.0518
Exponen	0.0943	0.0869	0.1044	0.0554	0.1110	0.0488	0.1151	0.0465
LogNorm	0.1392	0.0826	0.1917	0.0534	0.2314	0.0507	0.2679	0.0470
Mixture	0.0953	0.0868	0.0919	0.0576	0.0980	0.0561	0.1021	0.0535
Multi-d	0.0818	0.0875	0.0806	0.0573	0.0742	0.0526	0.0839	0.0521

The results in Table 3.1 confirm the conclusions stated in Corollary 2.2.1. Thus, the classical *CF* is liberal in this unbalanced design for all non-normal distributions, with the achieved α-level increasing with the number of sub-classes. Its performance is worse for the log-normal distribution. On the other hand, the proposed *HOM* procedure performed well for all distributions, though somewhat liberal in the case of small number of sub-classes.

2.3.2 Simulations under between-classes Heteroscedastic designs

Here we compare the achieved sizes of *CF*, *UW* and *WT*. The first procedure, *CF*, is as shown in (2.3.1). The second procedure, *UW*, using the asymptotic null distribution shown in (2.2.12), rejects at level α if

$$\sqrt{C}(F_C^* - 1) > \sqrt{\sum_{i=1}^{r}\frac{\hat{\lambda}_i\hat{\sigma}_i^4}{\hat{\beta}^2}\left[2 + \frac{2}{\hat{\bar{n}}_i - 1} + \frac{(\tilde{\kappa}_i - 3)\hat{\bar{n}}_i(\hat{\bar{n}}_i\hat{\underline{n}}_i - 1)}{(\hat{\bar{n}}_i - 1)^2}\right]} Z_\alpha, \tag{2.3.5}$$

where F_C^* is defined in (2.2.3), while the third procedure, WT, using the asymptotic null distribution shown in (2.2.15), rejects at level α if

$$\sqrt{C}(\widehat{F_C'} - 1) > \sqrt{2 + \frac{2}{\hat{\bar{n}} - 1} + \sum_{i=1}^{r} \left[\frac{(\tilde{\kappa}_i - 3)\hat{\lambda}_i(\hat{\bar{n}}^2 \underline{\hat{n}}_i - 2\hat{\bar{n}} + \hat{\bar{n}}_i)}{(\hat{\bar{n}} - 1)^2} \right]} Z_\alpha, \tag{2.3.6}$$

where $\widehat{F_C'}$ is defined in (2.2.10). The empirical quantities $\hat{\lambda}_i$, $\hat{\bar{n}}_i$, $\underline{\hat{n}}_i$, and $\hat{\bar{n}}$ are as defined in (2.3.3). Moreover, $\hat{\beta}$, $\hat{\sigma}_i^4$, and $\tilde{\kappa}_i$ above, are as follows:

$$\hat{\beta} = \sum_{i=1}^{r} \hat{\lambda}_i \hat{\sigma}_i^2, \quad \text{where } \hat{\sigma}_i^2 = S_i^2; \quad \hat{\sigma}_i^4 = (\hat{\sigma}_i^2)^2; \quad \text{and } \tilde{\kappa}_i = \frac{\hat{\nu}_i}{\hat{\sigma}_i^4}, \tag{2.3.7}$$

where $\hat{\nu}_i$ is defined in (2.3.4). Again, it can be easily verified that, as $\min(c_i) \to \infty$, $\tilde{\kappa}_i$ converges in probability to κ_i under the null hypothesis. The corresponding simulated sizes under heteroscedastic designs, based on 10,000 runs, are shown in Table 3.2.

Table 2.2. Achieved α-levels over 10,000 simulation runs under between-classes heteroscedastic designs with smaller average cell sizes at nominal $\alpha = 0.05$.

	$\bar{c} = 5$			$\bar{c} = 30$			$\bar{c} = 100$			$\bar{c} = 500$		
	CF	UW	WT	CF	UW	WT	CF	UW	WT	CF	UW	WT
Normal	.57	.115	.299	.99	.079	.120	1.0	.067	.077	1.0	.059	.062
Exponen	.47	.155	.333	.97	.089	.152	1.0	.072	.084	1.0	.057	.067
LogNorm	.39	.154	.350	.89	.105	.193	.99	.083	.118	1.0	.074	.094
Mixture	.37	.159	.351	.95	.082	.170	1.0	.068	.085	1.0	.054	.060
Multi-d	.48	.156	.332	.97	.089	.138	1.0	.067	.080	1.0	.063	.065

Table 3.2 makes it clear that the traditional *CF* procedure is quite inappropriate under between-classes heteroscedastic designs. More specifically, when $\bar{c}$ is large enough, regardless of the underlying distribution, the *CF* procedure rejects the null hypothesis almost all the times under the null hypothesis. When comparing the two proposed procedures, UW and WT, we can see that procedure WT appears to be more liberal than procedure UW, but becomes less so when $\bar{c}$ increases. The most likely explanation for this is the small-sample instability of the variance estimators that are used to standardize the observations in the WT statistic. This explanation is confirmed by the results in Table 2.3, where the only difference is that the sample sizes were generated from truncated

Poisson distributions with larger mean values: $(\upsilon_1,\ \upsilon_2,\ \upsilon_3,\ \upsilon_4, \upsilon_5)' = (5,\ 5,\ 5,\ 12,\ 5))$.

Table 2.3. Achieved α-levels over 10,000 simulation runs under between-classes Heteroscedasticity with larger average cell sizes at nominal $\alpha = 0.05$.

	$\bar{c} = 5$			$\bar{c} = 30$			$\bar{c} = 100$			$\bar{c} = 500$		
	CF	UW	WT	CF	UW	WT	CF	UW	WT	CF	UW	WT
Normal	.25	.105	.183	.55	.072	.080	.91	.063	.059	1.0	.055	.058
Exponen	.23	.097	.175	.52	.072	.082	.91	.068	.068	1.0	.056	.057
LogNorm	.21	.083	.156	.46	.060	.076	.85	.070	.071	.99	.057	.062
Mixture	.21	.079	.157	.52	.073	.081	.90	.065	.065	1.0	.059	.059
Multi-d	.23	.094	.183	.52	.072	.078	.90	.068	.065	1.0	.059	.061

With the same simulation settings used for Table 2.3, but taking $\bar{c} = 100$, Table 2.4 compares the achieved powers of the two proposed test statistics UW and WT, under the alternatives $\delta_{ij} = t \times (2j/c_i - 1)$, for $t = 0.6,\ 0.8,\ 1.0,\ 1.2$ and $i = 1, \cdots, 5$, $j = 1, \cdots, c_i - 1$. For each i, δ_{ic_i} is chosen so that $\sum_j n_{ij}\delta_{ij} = 0$.

Table 2.4. Powers over 10,000 simulation runs under between-classes heteroscedastic designs with larger average cell sizes at nominal $\alpha = 0.05$ ($\bar{c} = 100$).

	t=0.6		t=0.8		t=1.0		t=1.2	
	UW	WT	UW	WT	UW	WT	UW	WT
Normal	.3245	1.0	.5676	1.0	.9725	1.0	.9999	1.0
Exponen	.2828	1.0	.5002	1.0	.9465	1.0	.9989	1.0
LogNorm	.9027	1.0	.9670	1.0	.9945	1.0	.9985	1.0
Mixture	.2813	1.0	.5090	1.0	.9576	1.0	.9997	1.0
Multi-d	.2855	1.0	.5006	1.0	.9476	1.0	.9985	1.0

As expected, Table 2.4 reveals that the WT procedure is much more powerful in detecting the sub-class effect than the unweighted procedure UW. Note that with the larger cell sample sizes used in Table 2.3, WT is only mildly liberal. Thus, the power advantage of WT does not come at the expense of level accuracy.

More discussion and comparisons between the procedures UW and WT are provided in Appendix A.

2.3.3 Simulations under general Heteroscedastic designs

The simulations in the previous subsection demonstrate that the classical *CF* procedure is very liberal under between-classes heteroscedasticity. Simulations under general heteroscedasticity, not shown here, reveal similar behavior. Thus, the tables in this subsection exclude the *CF* procedure.

In Table 2.5 we compare the achieved α-levels of *UW* and *HET*, under general heteroscedasticity. The former procedure is described in (2.3.5), while the latter uses the statistic F_C^{**} given in (2.2.17) and its asymptotic null distribution shown in (3.1.3). Thus, the *HET* procedure rejects at level α if

$$\sqrt{C}(F_C^{**} - 1) > \sqrt{\frac{2\hat{b}_1 + 2\hat{b}_2}{\hat{a}_1^2}} Z_\alpha, \qquad (2.3.8)$$

where $\hat{a}_1$, $\hat{b}_1$ and $\hat{b}_2$ are consistent estimators of a_1, b_1 and b_2. Note that consistent estimation of b_1 and b_2 needs unbiased estimation of each σ_{ij}^4. For such unbiased estimation we use the U-statistics with the kernel $(Y_{ij1} - Y_{ij2})^2/2 \times (Y_{ij3} - Y_{ij4})^2/2$. As a consequence, the application of procedure *HET* requires $n_{ij} \geq 4$, although Theorem 3.1.2 requires only $n_{ij} \geq 2$. Because of this constraint, we generate n_{ij} using $n_{ij} = Z_{ij} \times I(Z_{ij} \geq 4) + \upsilon_i \times I(Z_{ij} < 4)$, where I is an indicator function and $Z_{ij} \sim Poisson(\upsilon_i)$ with $(\upsilon_1,\ \upsilon_2,\ \upsilon_3,\ \upsilon_4, \upsilon_5)' = (5,\ 5,\ 5,\ 12,\ 5)$, $i = 1, \cdots, 5;\ j = 1, \cdots, c_i$.

Table 2.5. Achieved α-levels over $10,000$ simulation runs at nominal $\alpha = 0.05$ under general Heteroscedasticity.

	$\bar{c} = 5$		$\bar{c} = 30$		$\bar{c} = 100$		$\bar{c} = 500$	
	UW	HET	UW	HET	UW	HET	UW	HET
Normal	.1089	.1043	.0985	.0837	.1479	.0701	.1560	.0575
Exponen	.1120	.1012	.0896	.0717	.1333	.0597	.1545	.0530
LogNorm	.1064	.1052	.0738	.0568	.1092	.0491	.1225	.0461
Mixture	.1192	.1018	.0852	.0567	.1378	.0554	.1507	.0508
Multi-d	.1093	.0968	.0910	.0718	.1366	.0635	.1576	.0550

From Table 2.5 we see that both procedures are liberal when the average number of sub-classes is 5, but *HET* becomes less so as $\bar{c}$ increases. On the other hand, *UW* becomes more liberal as $\overline{c}$ increases, a behavior which is expected in view of the fact that it is not designed to allow the present type of heteroscedasticity.

The next two tables perform a more detailed comparison of the procedures *UW* and *HET* under the setting of between-classes heteroscedasticity when both are asymptotically valid. Table 2.6 suggests that the achieved α-levels of the two procedures are comparably close to the nominal level, with *HET* slightly less liberal for non-normal distributions. Table 2.7 shows the achieved powers of these two procedures, only for the case of $\overline{c} = 100$, under the alternatives used in Table 2.4. Again the procedures have comparable power with *HET* being slightly more powerful for non-normal distributions.

Table 2.6. Achieved α-levels over 10,000 simulation runs under between-classes heteroscedasticity at nominal $\alpha = 0.05$.

	$\overline{c} = 5$		$\overline{c} = 30$		$\overline{c} = 100$		$\overline{c} = 500$	
	UW	HET	UW	HET	UW	HET	UW	HET
Normal	.1077	.1194	.0701	.0771	.0606	.0612	.0513	.0539
Exponen	.0907	.1017	.0720	.0698	.0628	.0560	.0566	.0519
LogNorm	.0757	.0871	.0655	.0529	.0651	.0525	.0624	.0440
Mixture	.0798	.0851	.0750	.0633	.0650	.0536	.0584	.0494
Multi-d	.0921	.1041	.0711	.0655	.0626	.0567	.0560	.0505

Table 2.7. Powers over 10,000 simulation runs under between-classes heteroscedastic designs at nominal $\alpha = 0.05$ ($\overline{c} = 100$).

	t=0.6		t=0.8		t=1.0		t=1.2	
	UW	HET	UW	HET	UW	HET	UW	HET
Normal	.3353	.3410	.6249	.6214	.9204	.9173	1.000	.9998
Exponen	.3308	.3361	.6210	.6436	.9083	.9259	.9997	.9997
LogNorm	.9750	.9986	.9941	1.000	.9986	1.000	.9996	1.000
Mixture	.3289	.3218	.6115	.6417	.9135	.9350	.9998	1.000
Multi-d	.3325	.3347	.6243	.6549	.9101	.9304	.9994	.9997

The final two tables perform a more detailed comparison of the procedures *HOM*, *UW* and *HET* under homoscedasticity when all three are asymptotically valid. The results reported in Table 2.8 suggest that the achieved α-levels of the three procedures are comparably close to the nominal level (the results for $\overline{c} = 500$ are very close to those for $\bar{c} = 100$, so they are omitted). Table 2.9 compares the achieved powers of these three procedures, only for the case of $\bar{c} = 100$, under alternatives $\delta_{ij} = t \times (2j/c_i - 1)$, for $t = 0.20,\ 0.25,\ 0.35$ and $i = 1, \cdots, 5,\ j = 1, \cdots, c_i - 1$. For each i, δ_{ic_i} is chosen so that $\sum_j n_{ij}\delta_{ij} = 0$. Note that the cell sizes used here are larger than those used in Table 3.1, as required for the applicability of *HET*. The results suggest that, even though procedure *HET* estimates more parameters, this does not compromise its power.

Table 2.8. Achieved α-levels over 10,000 simulation runs under Homoscedasticity at nominal $\alpha = 0.05$.

	$\bar{c} = 5$			$\bar{c} = 30$			$\bar{c} = 100$		
	HOM	UW	HET	HOM	UW	HET	HOM	UW	HET
Normal	.092	.087	.099	.058	.057	.058	.053	.053	.053
Exponen	.088	.075	.083	.061	.058	.056	.056	.053	.050
LogNorm	.078	.057	.072	.062	.053	.044	.057	.053	.041
Mixture	.094	.080	.080	.064	.064	.058	.058	.059	.052
Multi-d	.085	.076	.084	.066	.062	.062	.059	.058	.056

Table 2.9. Powers over 10,000 simulation runs under Homoscedasticity at nominal $\alpha = 0.05$ ($\bar{c} = 100$).

	$t = 0.20$			$t = 0.25$			$t = 0.35$		
	HOM	UW	HET	HOM	UW	HET	HOM	UW	HET
Normal	.436	.429	.427	.735	.725	.724	.993	.992	.992
Exponen	.411	.418	.426	.706	.716	.732	.992	.993	.994
LogNorm	.977	.989	1.00	.995	.998	1.00	.998	1.00	1.00
Mixture	.417	.421	.428	.711	.714	.727	.992	.992	.995
Multi-d	.419	.420	.422	.716	.720	.725	.992	.992	.993

2.4 Data Analyses: Two Empirical Studies

2.4.1 Application to the Mussel Watch Project Data

One real-world application for our methodology can be found through the National Oceanic and Atmospheric Administration's National Status and Trends Program. In 1986, this division undertook a very large scale project to monitor the levels of numerous chemical contaminants and organic chemical constituents in marine sediment and bivalve (mollusk) tissue samples. This project, dubbed the Mussel Watch Project, is still ongoing and there are no apparent plans to discontinue it in the near future. There are currently over 300 coastal sites at which sediment and bivalve samples are collected and analyzed for the project. Each site is categorized as being within a certain Estuarine Drainage Area (EDA). See O'Connor (1998) for more details on this project. For our data analysis, we chose to analyze the Lead concentrations from years 1998 to 2005. We chose to analyze concentrations of Lead in tissue samples, specifically in the *Crassostrea virginica*, or American Oysters, from two different regions: Middle and South Atlantic, and the Gulf of Mexico. Due to the fact that nested in each region there are many EDAs, it is natural to consider regions as classes, and EDAs as sub-classes in our analysis. The main interest of our study is the sub-class effect. The boxplots of the lead concentration levels at each EDA, shown in Figure 2.1, suggest heteroscedasticity among different EDAs in the same region (general heteroscedasticity). Thus, the procedure *HET* seems to be the appropriate one for analyzing this data set. However, the results of application of the other procedures mentioned in this chapter (i.e. *CF*, *HOM*, *UW*, and *WT*) are also included for comparison purposes. Because the *HET* procedure requires at least 4 observations within each sub-class, we remove four EDAs with less than four observations from our data, resulting in 58 EDAs in total. (Another approach would be to impute values, but this will be pursued elsewhere.)

Application of five procedures, *CF*, *HOM*, *UW*, *WT* and *HET*, on this data set yields p-values of 0.1076, 0.3136, 0.2008, 0.0005 and 0.0246, respectively, for the hypothesis of no EDA effect. Note that only procedures *HET* and *WT* detect the effect of EDA at $\alpha = 0.05$. A closer examination of the data reveals that the largest sample variance estimate from EDA 'G120x' in the Gulf of Mexico region is 69.21, while the second largest one is only 2.45. This high variance of the data in EDA 'G120x' in fact results from a few outliers in a site named 'CBPP', as shown in Figure 2.1. After four data points from 'CBPP' are removed (see the changed boxplots in Figure 2.2), the sample variance estimate of the EDA 'G120x' becomes 0.0921 and heteroscedasticity is not so pronounced. With the outliers removed, the p-values of five procedures are all very close

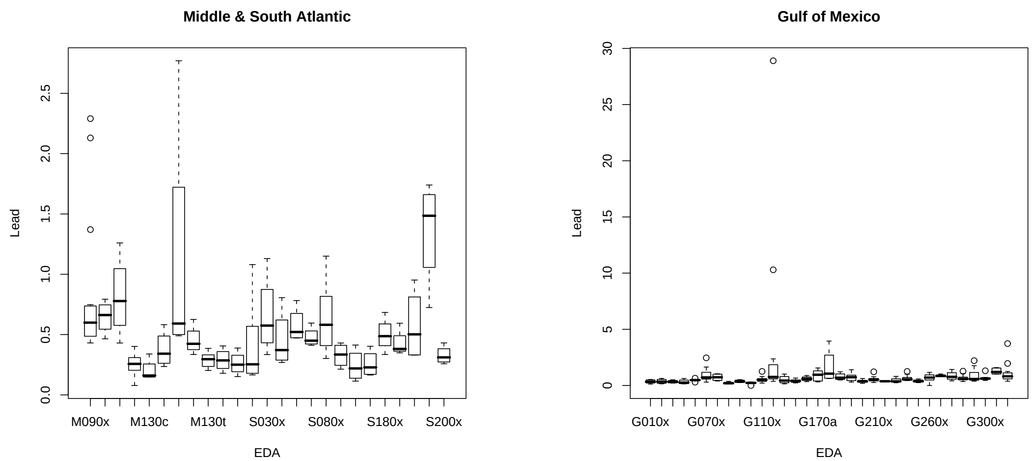

Figure 2.1. Mussel Watch Project. The boxplots of the Lead concentration levels in American Oysters at EDAs nested in Middle and South Atlantic (left) and in the Gulf of Mexico (right).

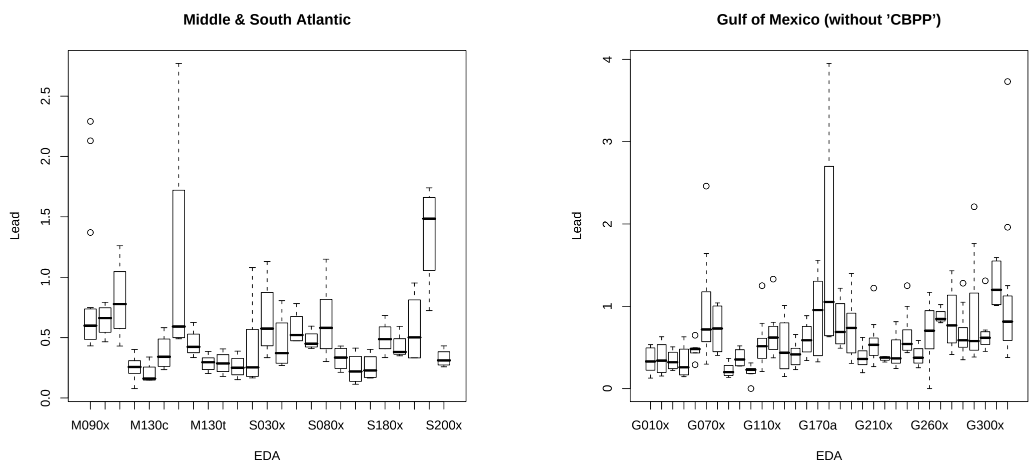

Figure 2.2. Mussel Watch Project. The boxplots of the Lead concentration levels in American Oysters at EDAs nested in Middle and South Atlantic (left) and in the Gulf of Mexico (right), after four observations at 'CBPP' removed.

to zero (less than 10^{-12}). This dramatic change confirms the instability of procedures *CF*, *HOM* and *UW* under general heteroscedasticity.

2.4.2 Application to the NADP Data: Ramification for ANCOVA

Another real-world application for our methodology can be found through the National Atmospheric Deposition Program (2009), which monitors of geographical and temporal long-term trends on the chemistry of precipitation. Starting from only 22 stations in 1978, NAPD has grown as a nationwide network of over 250 sites for which precipitation samples are collected and analyzed in the Central Analytical Laboratory (CAL) weekly. For our data analysis, we chose to analyze the pH level (reported as the negative log of hydrogen ion concentration) of precipitation samples as measured in the CAL from the first week of January 2003 to the last week of December 2007. We consider comparing the data in two North Carolina towns, Lewiston and Coweeta, and are interested in the effect of Time. Among the total 233 weeks in this period, we notice that there are several weeks in which data were missing at one or both locations. After removing those missing values, there are 180 weeks of data for each of these two towns, i.e. $n_1 = n_2 = 180$, although this balancedness is simply a coincidence. A further examination reveals that the missing data in fact happen at different time points for the two locations.

This data set can be analyzed as a simple one-way ANCOVA model with locations as groups and time as the covariate. Here, in order to directly apply the asymptotic theory for the two-fold nested model, we think of two locations as two *classes* and form artificial *sub-classes* by dividing the observations in the same class into non-overlapping 'windows' of a fixed size 5. More specifically, the first time sub-class consists of observations from weeks 1–5, the second time sub-class consists of observations from weeks 6–10, and so on. Since there are 180 observations in each of two locations, this division results in $180/5 = 36$ sub-classes each class, i.e. $c_1 = c_2 = 36$. We call this simple ramification for the analysis of covariance, as outlined above, the *non-overlapping windows approach.* The boxplots of the pH levels at each of these 36 times for the two locations are shown in Figure 2.3 (the left panel for Lewiston, the right one for Coweeta). A simple time series analysis does not indicate meaningful correlation over time (see Appendix B), so it appears reasonable to implement our methodology in this study.

For the sub-class effect of Time, the five procedures mentioned in this chapter, *CF*, *HOM*, *UW*, *WT* and *HET*, give p-values of 0.0929, 0.0757, 0.0760, 0.0773 and 0.0508, respectively. Given that the assumption of homoscedasticity is clearly violated as shown in Figure 2.3, it is not surprising that the *HET* procedure is the only one which finds the effect of Time (marginally) significant at $\alpha = 0.05$. Although this simple ramification of

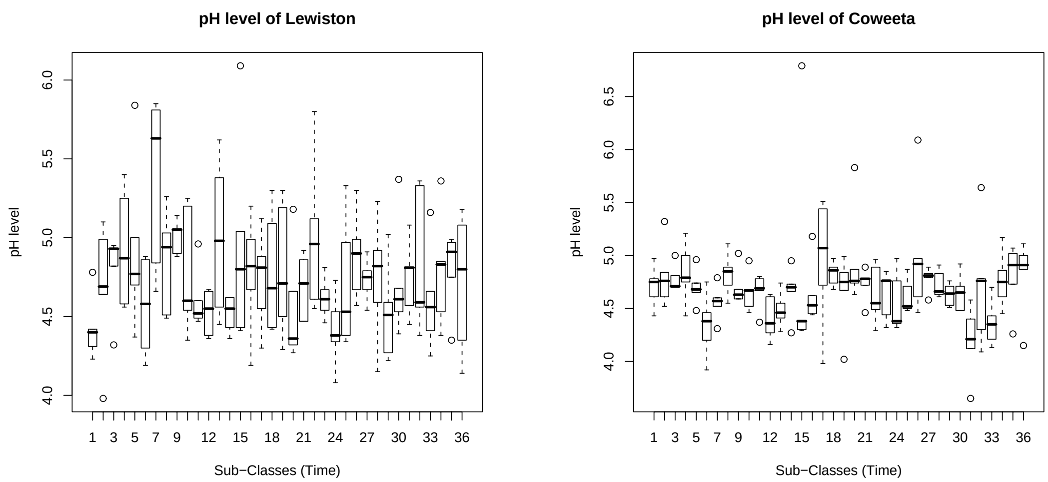

Figure 2.3. NADP Rain Data. The boxplots of the pH levels of precipitation at different Times from January 2003 to December 2007 in two towns in North Carolina: Lewiston (left) and Coweeta (right).

the present methodology for the nested model seems to work for analysis of covariance as well, we look for a more powerful tool. A further analysis shows that if we apply the *HET* procedure on the data from Year 1998 to Year 2007 (i.e. use the weekly data from 10 years, instead of just 5 years), the p-value for the null hypothesis of no time effect is as small as 0.000189, implying that a larger number of 'windows' may result in a better power in testing. A more advanced approach for analysis of covariance, called the *overlapping windows approach*, is proposed in Chapter 5.

2.5 Proofs

In this appendix we will use the following notations:

$$\mathsf{U}_c \approx \mathsf{V}_c \quad \Leftrightarrow \quad \sqrt{c}(\mathsf{U}_c - \mathsf{V}_c) \xrightarrow{P} 0, \quad \text{as } c \to \infty,$$

$$\mathsf{a}_c \approx \mathsf{b}_c \quad \Leftrightarrow \quad \sqrt{c}(\mathsf{a}_c - \mathsf{b}_c) \to 0, \quad \text{as } c \to \infty,$$

where U_c and V_c are two sequences of random vectors, while a_c and b_c are two sequences of constant vectors.

Proof of Theorem 2.2.1

Define

$$U_{ij}^{\delta} = n_{ij}(\bar{e}_{ij\cdot} + \frac{\delta_{ij}}{\sigma})^2, \quad \bar{U}_{ic}^{\delta} = \frac{1}{c_i}\sum_{j=1}^{c_i} U_{ij}^{\delta}, \quad \bar{U}_{C}^{\delta} = \frac{1}{C}\sum_{i=1}^{r} \sigma^2 c_i \bar{U}_{ic}^{\delta},$$

$$W_{ij} = \frac{\sum_{k=1}^{n_{ij}}(e_{ijk} - \bar{e}_{ij\cdot})^2}{\bar{n}-1}, \quad \bar{W}_{ic} = \frac{1}{c_i}\sum_{j=1}^{c_i} W_{ij}, \quad \bar{W}_{C} = \frac{1}{C}\sum_{i=1}^{r} \sigma^2 c_i \bar{W}_{ic}, \tag{2.5.1}$$

$$\mathsf{V}_{ij}^{\delta} = \begin{pmatrix} U_{ij}^{\delta} \\ W_{ij} \end{pmatrix}, \quad \bar{\mathsf{V}}_{ic}^{\delta} = \begin{pmatrix} \bar{U}_{ic}^{\delta} \\ \bar{W}_{ic} \end{pmatrix}, \quad \bar{\mathsf{V}}_{C}^{\delta} = \begin{pmatrix} \bar{U}_{C}^{\delta} \\ \bar{W}_{C} \end{pmatrix}.$$

Note that

$$\begin{aligned} \bar{U}_{C}^{\delta} &= MS\delta + \left[\frac{1}{C-r}\sum_{i=1}^{r}\sigma^2 n_{i\cdot}\bar{e}_{i\cdot\cdot}^2 - \frac{r}{C-r}\bar{U}_{C}^{\delta}\right], \quad \text{and} \\ \bar{W}_{C} &= MSE + \left[\sum_{i=1}^{r}\left(\frac{c_i}{C} - \frac{c_i(\bar{n}-1)}{N_C - C}\right)\sigma^2 \bar{W}_{ic}\right]. \end{aligned}$$

It can be easily verified that, as $\min(c_i) \to \infty$ and $r,\ n_{ij}$ remain fixed,

$$\sqrt{C}\ \frac{1}{C-r}\sum_{i=1}^{r} n_{i\cdot}\bar{e}_{i\cdot\cdot}^2 \xrightarrow{P} 0, \quad \sqrt{C}\ \frac{r}{C-r}\bar{U}_{C}^{\delta} \xrightarrow{P} 0, \quad \text{and} \tag{2.5.2}$$

$$\sqrt{C}\ \sum_{i=1}^{r}\left(\frac{c_i}{C} - \frac{c_i(\bar{n}-1)}{N_C - C}\right)\sigma^2 \bar{W}_{ic} = \left(1 - \frac{C(\bar{n}-1)}{N_C - C}\right)\sqrt{C}\bar{W}_C \xrightarrow{P} 0.$$

Combining the above we have that, as $\min(c_i) \to \infty$ and $r,\ n_{ij}$ remain fixed,

$$\bar{\mathsf{V}}_{C}^{\delta} \approx \mathsf{M}_{C}^{\delta} \equiv \begin{pmatrix} MS\delta \\ MSE \end{pmatrix}. \tag{2.5.3}$$

Hence, the asymptotic joint distribution of $MS\delta$ and MSE is the same as the asymptotic joint distribution of $\bar{U}_{C}^{\delta}$ and $\bar{W}_{C}$.

It can be shown that, under normality, U_{ij}^{δ} and W_{ij} are independent, and

$$U_{ij}^{\delta} \sim \chi_1^2\left(\frac{n_{ij}\delta_{ij}^2}{\sigma^2}\right), \quad (\bar{n}-1)W_{ij} \sim \chi_{n_{ij}-1}^2.$$

Using known results regarding the mean and covariance of quadratic forms (cf. Theorem 1 in Akritas and Arnold (2000)) and the facts that $E(\chi_a^2(a\gamma)) = a(1+\gamma)$, $Var(\chi_a^2(a\gamma)) = a(2+4\gamma)$, we obtain

$$E(\mathsf{V}_{ij}^{\delta}) = \begin{pmatrix} 1+\frac{n_{ij}\delta_{ij}^2}{\sigma^2} \\ \frac{n_{ij}-1}{\bar{n}-1} \end{pmatrix},$$

$$Cov(\mathsf{V}_{ij}^{\delta}) = \begin{pmatrix} 2+4\frac{n_{ij}\delta_{ij}^2}{\sigma^2} & 0 \\ 0 & \frac{2(n_{ij}-1)}{(\bar{n}-1)^2} \end{pmatrix} + \frac{\kappa_i - 3}{n_{ij}} \begin{pmatrix} 1 & \frac{n_{ij}-1}{\bar{n}-1} \\ \frac{n_{ij}-1}{\bar{n}-1} & \left(\frac{n_{ij}-1}{\bar{n}-1}\right)^2 \end{pmatrix}.$$

Let $\theta_{ic_i}^{\delta} = \frac{1}{c_i}\sum_{j=1}^{c_i} n_{ij}\frac{\delta_{ij}^2}{\sigma^2}$. Then, for each class i, as $c_i \to \infty$,

$$E(\bar{\mathsf{V}}_{ic}^{\delta}) = \begin{pmatrix} 1+\theta_{ic_i}^{\delta} \\ \frac{\bar{n}_{ic_i}-1}{\bar{n}-1} \end{pmatrix} \approx \begin{pmatrix} 1+\theta_i \\ \frac{\bar{n}_i-1}{\bar{n}-1} \end{pmatrix} \triangleq \boldsymbol{\mu}_i, \quad \text{and} \tag{2.5.4}$$

$$\begin{aligned} & c_i \cdot Cov(\bar{\mathsf{V}}_{ic}^{\delta}) \\ &= \begin{pmatrix} 2+4\theta_{ic}^{\delta} & 0 \\ 0 & \frac{2}{\bar{n}-1}\frac{\bar{n}_{ic_i}-1}{\bar{n}-1} \end{pmatrix} + \frac{\kappa_i-3}{(\bar{n}-1)^2} \begin{pmatrix} (\bar{n}-1)^2\underline{n}_{ic_i} & (\bar{n}-1)(1-\underline{n}_{ic_i}) \\ (\bar{n}-1)(1-\underline{n}_{ic_i}) & \bar{n}_{ic_i}+\underline{n}_{ic_i}-2 \end{pmatrix} \\ &\longrightarrow \begin{pmatrix} 2+4\theta_i & 0 \\ 0 & \frac{2(\bar{n}_i-1)}{(\bar{n}-1)^2} \end{pmatrix} + \frac{\kappa_i-3}{(\bar{n}-1)^2} \begin{pmatrix} (\bar{n}-1)^2\underline{n}_i & (\bar{n}-1)(1-\underline{n}_i) \\ (\bar{n}-1)(1-\underline{n}_i) & \bar{n}_i+\underline{n}_i-2 \end{pmatrix} \triangleq \Sigma_i. \end{aligned}$$

Under the assumption that $E|e_{ijk}|^{4+2\epsilon} < \infty$ for some $\epsilon > 0$, Lindeberg-Feller's theorem together with Cramér-Wold's theorem yield

$$\sqrt{c_i}(\bar{\mathsf{V}}_{ic}^{\delta} - \boldsymbol{\mu}_i) \xrightarrow{d} N_2(\mathsf{0}, \Sigma_i).$$

Using the independence among $\bar{\mathsf{V}}_{ic}^{\delta}$ and the assumption (2.1.8), one can be show that

$$\sqrt{C}(\bar{\mathsf{V}}_{C}^{\delta}-\boldsymbol{\mu}) \xrightarrow{d} N_2(\mathsf{0},\sigma^4\sum_{i=1}^{r}\lambda_i\Sigma_i), \quad \text{where } \boldsymbol{\mu}=\sigma^2\begin{pmatrix}1+\theta\\ 1\end{pmatrix}. \tag{2.5.5}$$

By the asymptotic equivalence between $\bar{\mathsf{V}}_{C}^{\delta}$ and M_{C}^{δ} shown in (2.5.3), we then have

$$\sqrt{C}(\mathsf{M}_{C}^{\delta}-\boldsymbol{\mu}) \xrightarrow{d} N_2(\mathsf{0},\sigma^4\sum_{i=1}^{r}\lambda_i\Sigma_i), \text{ as } \min{(c_i)}\to\infty.$$

Note that if $\mathsf{s}'=(1,-(1+\theta))/\sigma^2$, $\sqrt{C}\,\mathsf{s}'(\mathsf{M}_{C}^{\delta}-\boldsymbol{\mu})=\sqrt{C}[MS\delta-(1+\theta)MSE]/\sigma^2$ which, by Slutsky's theorem, is asymptotically equivalent to $\sqrt{C}\left(F_{C}^{\delta}-(1+\theta)\right)$. Thus, by the Δ-method, as $\min{(c_i)}\to\infty$,

$$\sqrt{C}\left(F_{C}^{\delta}-(1+\theta)\right) \xrightarrow{d} N\left(0,\sigma^4\sum_{i=1}^{r}\lambda_i\mathsf{s}'\Sigma_i\mathsf{s}\right)=N\left(\mathsf{0},\,\Sigma_s\right),$$

where Σ_s is as defined in Theorem 2.2.1.

Proof of Corollary 2.2.1

It can be easily verified that for C large enough, the approximate distribution of the classical F-test under $H_0:\delta_{ij}=0$, and under the normality assumption is:

$$\sqrt{C}\left(F_{C}^{\delta}-1\right)\dot{\sim} N\left(0,2\left(1+\frac{C}{N_C-C}\right)\right), \tag{2.5.6}$$

where $\dot{\sim}$ means "approximately distributed". The relation (2.5.6) is obviously not equivalent to (2.2.2), unless $n_{ij}=n$ for all i and j so that

$$\bar{n}_{ic_i}=n=\bar{n}_i, \quad \underline{n}_{ic_i}=\frac{1}{n}=\underline{n}_i \;\longrightarrow\; \bar{n}_i\underline{n}_i-1=0,$$

and hence both of (2.2.2) and (2.5.6) would become

$$\sqrt{C}\left(F_{C}^{\delta}-1\right)\xrightarrow{d} N\left(0,2\left(1+\frac{1}{n-1}\right)\right). \tag{2.5.7}$$

Proof of Theorem 2.2.2

Define new quantities U_{ij}^{δ}, $\bar{U}_C^{\delta}$, $\bar{W}_C$ to be as the corresponding quantities in (2.5.1) but with σ_i^2 replacing σ^2, and the new quantity W_{ij} to be as the corresponding quantity in (2.5.1) but with $\bar{n}_i$ replacing $\bar{n}$. Finally, let $\bar{U}_{ic}^{\delta}$, $\bar{W}_{ic}$, V_{ij}^{δ}, $\bar{\mathsf{V}}_{ic}^{\delta}$, and $\bar{\mathsf{V}}_C^{\delta}$ be as defined in (2.5.1) but using the above new quantities. It can then be shown that $\bar{U}_C^{\delta}$, $\bar{W}_C$ are related to $MS\delta$, MSE^* via

$$\begin{aligned}
\bar{U}_C^{\delta} &= MS\delta + \left[\frac{1}{C-r}\sum_{i=1}^{r}\sigma_i^2 n_{i\cdot}\bar{e}_{i\cdot\cdot}^2 - \frac{r}{C-r}\bar{U}_C^{\delta}\right], \\
\bar{W}_C &= MSE^* + \left[\sum_{i=1}^{r}\left(\frac{c_i}{C} - \frac{(c_i-1)c_i(\bar{n}_i-1)}{(C-r)(n_{i\cdot}-c_i)}\right)\sigma_i^2\bar{W}_{ic}\right].
\end{aligned}$$

Using (2.5.2), and the fact that, as $\min(c_i)\to\infty$ and $r,\ n_{ij}$ remain fixed,

$$\sqrt{C}\ \sum_{i=1}^{r}\left(\frac{c_i}{C} - \frac{(c_i-1)c_i(\bar{n}_i-1)}{(C-r)(n_{i\cdot}-c_i)}\right)\sigma_i^2\bar{W}_{ic} \xrightarrow{P} 0, \tag{2.5.8}$$

we have that, as $\min(c_i)\to\infty$ and $r,\ n_{ij}$ remain fixed,

$$\bar{\mathsf{V}}_C^{\delta} \approx \mathsf{M}_C^* \equiv \begin{pmatrix} MS\delta \\ MSE^* \end{pmatrix}. \tag{2.5.9}$$

Following the same derivation in the proof of Theorem 2.2.1, one can easily get

$$\sqrt{c_i}(\bar{\mathsf{V}}_{ic}^{\delta} - \boldsymbol{\mu}_i^*) \xrightarrow{d} N_2(\mathsf{0},\Sigma_i^*), \quad \text{where } \boldsymbol{\mu}_i^* \text{ and } \Sigma_i^* \text{ are defined by} \tag{2.5.10}$$

$$E(\bar{\mathsf{V}}_{ic}^{\delta}) = \begin{pmatrix} 1+\theta_{ic_i}^{\delta} \\ \frac{\bar{n}_{ic_i}-1}{\bar{n}_i-1} \end{pmatrix} \approx \begin{pmatrix} 1+\theta_i \\ 1 \end{pmatrix} \triangleq \boldsymbol{\mu}_i^*, \quad \text{where } \theta_{ic_i}^{\delta} = \frac{1}{c_i}\sum_{j=1}^{c_i} n_{ij}\frac{\delta_{ij}^2}{\sigma_i^2}, \quad \text{and}$$

$$c_i \cdot Cov(\bar{\mathsf{V}}_{ic}^{\delta})$$

$$= \begin{pmatrix} 2+4\theta_{ic}^{\delta} & 0 \\ 0 & \frac{2}{\bar{n}_i-1}\frac{\bar{n}_{ic_i}-1}{\bar{n}_i-1} \end{pmatrix} + \frac{\kappa_i-3}{(\bar{n}_i-1)^2}\begin{pmatrix} (\bar{n}_i-1)^2\underline{n}_{ic_i} & (\bar{n}_i-1)(1-\underline{n}_{ic_i}) \\ (\bar{n}_i-1)(1-\underline{n}_{ic_i}) & \bar{n}_{ic_i}+\underline{n}_{ic_i}-2 \end{pmatrix}$$

$$\longrightarrow \begin{pmatrix} 2+4\theta_i & 0 \\ 0 & \frac{2}{\bar{n}_i-1} \end{pmatrix} + \frac{\kappa_i-3}{(\bar{n}_i-1)^2}\begin{pmatrix} (\bar{n}_i-1)^2\underline{n}_i & (\bar{n}_i-1)(1-\underline{n}_i) \\ (\bar{n}_i-1)(1-\underline{n}_i) & \bar{n}_i+\underline{n}_i-2 \end{pmatrix} \triangleq \Sigma_i^*.$$

By the independence among $\bar{\mathsf{V}}_{ic}^{\delta}$ and the assumption (2.1.8), it can be shown that

$$\sqrt{C}(\bar{\mathsf{V}}_C^{\delta} - \boldsymbol{\mu}^*) \xrightarrow{d} N_2(\mathsf{0}, \sum_{i=1}^{r} \sigma_i^4 \lambda_i \Sigma_i^*), \quad \text{where } \boldsymbol{\mu}^* = \begin{pmatrix} \beta+\theta^{\sigma} \\ \beta \end{pmatrix},$$

where β and θ^{σ} are as defined in Theorem 2.2.2. Because $\bar{\mathsf{V}}_C^{\delta}$ and M_C^* are asymptotically equivalent, as shown in (2.5.9), we then have

$$\sqrt{C}(\mathsf{M}_C^* - \boldsymbol{\mu}^*) \xrightarrow{d} N_2(\mathsf{0}, \sum_{i=1}^{r} \sigma_i^4 \lambda_i \Sigma_i^*), \text{ as } \min(c_i) \to \infty.$$

Finally, by the Δ-method with $\mathsf{s}^{*\prime} = (1, -(1+\theta^*))/\beta$, where $\theta^* = \theta^{\sigma}/\beta$, it can be easily verified that, as $\min(c_i) \to \infty$,

$$\sqrt{C}\left(F_C^* - (1+\theta^*)\right) \xrightarrow{d} N\left(0, \sum_{i=1}^{r} \sigma_i^4 \lambda_i \mathsf{s}^{*\prime}\Sigma_i^*\mathsf{s}^*\right) = N\left(0,\, \Sigma_s^*\right),$$

where Σ_s^* is as defined in Theorem 2.2.2.

Proof of Corollary 2.2.2

The fact that, when the design is balanced, the unweighted statistic F_C^* equals the classical F-statistic is clear. Next, (2.2.13) follows directly from Theorem 2.2.2. Finally, the fact that the classical F-test procedure is not valid follows by comparing (2.5.7) and (2.2.13).

Proof of Theorem 2.2.3

Let U_{ij}^{δ}, $\bar{U}_{ic}^{\delta}$, W_{ij}, $\bar{W}_{ic}$, and V_{ij}^{δ} be as defined in Proof of Theorem 2.2.2. It follows easily from the definition (2.2.10) and the assumption (2.1.8) that as $\min(c_i) \to \infty$,

$$\widehat{F_C'} = \sum_{i=1}^{r} \frac{c_i}{C-r} \times \frac{\bar{U}_{ic}^{\delta} + o_p(c_i^{-.5})}{\bar{W}_{ic} + o_p(c_i^{-.5})} \approx \sum_{i=1}^{r} \frac{c_i}{C-r} \times \frac{\bar{U}_{ic}^{\delta}}{\bar{W}_{ic}}.$$

In addition, (2.5.10) tells us that

$$\sqrt{c_i}\left(\begin{pmatrix} \bar{U}_{ic}^{\delta} \\ \bar{W}_{ic} \end{pmatrix} - \begin{pmatrix} 1+\theta_i \\ 1 \end{pmatrix}\right) \xrightarrow{d} N_2\left(\mathsf{0},\, \Sigma_i^*\right), \text{ where } \Sigma_i^* \text{ is as defined there.}$$

Thus, by the Δ-method with $\mathsf{s_i}' = (1, -(1+\theta_i))$, the assumption (2.1.8), and the independence among $\bar{U}_{ic}^{\delta}$ and $\bar{W}_{ic}$, we obtain (2.2.14).

Proof of Corollary 2.2.3

The proof follows directly from Theorem 2.2.2.

Proof of Theorem 3.1.2

Define $\mathsf{V}_{ij}^{\delta} = (U_{ij}^{\delta},\ W_{ij})'$, $\bar{\mathsf{V}}_{ic}^{\delta} = (\bar{U}_{ic}^{\delta},\ \bar{W}_{ic})'$, and $\bar{\mathsf{V}}_{C}^{\delta} = (\bar{U}_{C}^{\delta},\ \bar{W}_{C})'$, where

$$\begin{aligned} U_{ij}^{\delta} &= \sigma_{ij}^2 n_{ij}(\bar{e}_{ij\cdot} + \frac{\delta_{ij}}{\sigma_{ij}})^2, \quad \bar{U}_{ic}^{\delta} = \frac{1}{c_i}\sum_{j=1}^{c_i} U_{ij}^{\delta}, \quad \bar{U}_{C}^{\delta} = \frac{1}{C}\sum_{i=1}^{r} c_i \bar{U}_{ic}^{\delta}, \\ W_{ij} &= \frac{\sigma_{ij}^2}{n_{ij}-1}\sum_{k=1}^{n_{ij}} (e_{ijk} - \bar{e}_{ij\cdot})^2, \quad \bar{W}_{ic} = \frac{1}{c_i}\sum_{j=1}^{c_i} W_{ij}, \quad \bar{W}_{C} = \frac{1}{C}\sum_{i=1}^{r} c_i \bar{W}_{ic}. \end{aligned}$$

Note that

$$\bar{U}_{C}^{\delta} = MS\delta + \left[\frac{1}{C-r}\sum_{i=1}^{r}\frac{1}{n_{i\cdot}}\left(\sum_{j=1}^{c_i}\sigma_{ij} n_{ij}\bar{e}_{ij\cdot}\right)^2 - \frac{r}{C-r}\bar{U}_{C}^{\delta}\right], \text{ and}$$

$$\bar{W}_C = MSE^{**} - \frac{r}{C-r}\sum_{i=1}^{r}\frac{1}{C}\sum_{j=1}^{c_i} S_{ij}^2 + \frac{1}{C-r}\sum_{i=1}^{r}\frac{1}{n_{i\cdot}}\sum_{j=1}^{c_i} n_{ij}S_{ij}^2.$$

Under the assumptions in (3.1.3), it can be easily verified that, as $\min(c_i) \to \infty$,

$$\sqrt{C}\ \frac{1}{C-r}\sum_{i=1}^{r}\frac{1}{n_{i\cdot}}\left(\sum_{j=1}^{c_i}\sigma_{ij}n_{ij}\bar{e}_{ij\cdot}\right)^2 \xrightarrow{P} 0, \quad \sqrt{C}\ \frac{r}{C-r}\bar{U}_C^{\delta} \xrightarrow{P} 0,$$

$$\sqrt{C}\frac{r}{C-r}\sum_{i=1}^{r}\frac{1}{C}\sum_{j=1}^{c_i} S_{ij}^2 \xrightarrow{P} 0, \quad \sqrt{C}\ \frac{1}{C-r}\sum_{i=1}^{r}\frac{1}{n_{i\cdot}}\sum_{j=1}^{c_i} n_{ij}S_{ij}^2 \xrightarrow{P} 0.$$

Combining the above we have that, as $\min(c_i) \to \infty$ and $r,\ n_{ij}$ remain fixed,

$$\bar{\mathsf{V}}_C^{\delta} \approx \mathsf{M}_C^{**} \equiv \begin{pmatrix} MS\delta \\ MSE^{**} \end{pmatrix}. \tag{2.5.11}$$

Following the same derivation in the proof of Theorem 2.2.1, one can easily get the asymptotic distribution of $\bar{\mathsf{V}}_{ic}^{\delta}$ as

$$\sqrt{c_i}(\bar{\mathsf{V}}_{ic}^{\delta} - \boldsymbol{\mu}_i^{**}) \xrightarrow{d} N_2(\mathbf{0}, \Sigma_i^{**}),$$

where $\boldsymbol{\mu}_i^{**}$ and Σ_i^{**} are

$$E(\bar{\mathsf{V}}_{ic}^{\delta}) = \begin{pmatrix} \frac{1}{c_i}\sum_j \sigma_{ij}^2 + \frac{1}{c_i}\sum_j n_{ij}\delta_{ij}^2 \\ \frac{1}{c_i}\sum_j \sigma_{ij}^2 \end{pmatrix} \approx \begin{pmatrix} a_{1i} + \theta_{1i} \\ a_{1i} \end{pmatrix} \triangleq \boldsymbol{\mu}_i^{**}, \quad \text{and}$$

$$\begin{aligned} & c_i \cdot Cov(\bar{\mathsf{V}}_{ic}^{\delta}) \\ &= \begin{pmatrix} 2\frac{1}{c_i}\sum_j \sigma_{ij}^4 + 4\frac{1}{c_i}\sum_j n_{ij}\sigma_{ij}^2\delta_{ij}^2 & 0 \\ 0 & 2\frac{1}{c_i}\sum_j \frac{\sigma_{ij}^4}{n_{ij}-1} \end{pmatrix} + \frac{1}{c_i}\sum_j \frac{\sigma_{ij}^4(\kappa_{ij}-3)}{n_{ij}} \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \\ &\longrightarrow \begin{pmatrix} 2b_{1i} + 4\theta_{2i} & 0 \\ 0 & 2b_{2i} \end{pmatrix} + b_{3i}\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \triangleq \Sigma_i^{**}. \end{aligned}$$

By the independence among $\bar{\mathsf{V}}_{ic}^{\delta}$, the assumptions in (3.1.3) and the asymptotic equivalence shown in (2.5.11), we then have

$$\sqrt{C}\left(\mathsf{M}_C^{**} - \begin{pmatrix} a_1 + \theta_1 \\ a_1 \end{pmatrix}\right) \xrightarrow{d} N_2\left(\mathsf{0}, \sum_{i=1}^{r} \lambda_i \Sigma_i^{**}\right),$$

where a_1 and θ_1 are as defied in the theorem above. Finally, using the Δ-method with $\mathsf{s}^{**\prime} = (1, -(1+\theta^{**}))/a_1$, $\theta^{**} = \theta_1/a_1$, one can easily get the limiting distribution of F_C^{**} as shown in (2.2.19) and complete the proof.

Chapter 3

Testing for the Sub-classes Effect in Two-fold Nested Model with large numbers of sub-classes and classes

The purpose of this chapter is to extend the results in Chapter 2 to cases having a large number of classes (r) and a large number of subclasses (c_i), while the cell sizes remain small. It is organized as follows. The first section that follows gives the asymptotic theories for the proposed test statistics in the two-fold nested model, both under homoscedasticity and under heteroscedasticity (more specifically, under *general* heteroscedasticity, as described in the previous chapter). In Section 3.2, simulations are used to compare the corresponding procedures with the classical F test procedures. In Section 3.3, we demonstrate an interesting application of our methodology for testing the lack-of-fit in regression, which is in fact the core of this chapter.

3.1 Main Results

The proofs of the asymptotic theorems presented in this section are basically similar to those in Section 2.5 (with some slightly more tedious calculations) and hence omitted.

3.1.1 Homoscedastic Designs

Under homoscedastic assumption (i.e. $\sigma_{ij} = \sigma$), we have the following theorem giving the asymptotic distributions for the test statistic F_C^δ, as defined in (2.1.7), when the numbers of classes and sub-classes (i.e. r and all c_i) are both large.

Theorem 3.1.1. *Consider the model and assumptions given in (2.1.1), (2.1.2), and the decomposition of the means given in (2.1.3). In addition, we assume that there exist* κ, $\overline{n} > 1$, *and* $\underline{n} \in (0, \infty)$ *such that as* $r \to \infty$ *and* $\min(c_i) \to \infty$,

$$E(e_{ijk}^3) = 0, \quad E(e_{ijk}^4) = \kappa, \text{ and } E|e_{ijk}|^{4+2\epsilon} < \infty \text{ for some } \epsilon > 0;$$

$$\sqrt{C}\left(\frac{N_C}{C} - \overline{n}\right) \longrightarrow 0, \quad \frac{1}{C}\sum_{i=1}^{r}\sum_{j=1}^{c_i}\frac{1}{n_{ij}} \longrightarrow \underline{n} < \infty, \tag{3.1.1}$$

Then, under alternatives δ_{ij} *which satisfy*

$$\sqrt{C}\left(\frac{1}{C}\sum_{i=1}^{r}\sum_{j=1}^{c_i} n_{ij}\frac{\delta_{ij}^2}{\sigma^2} - \theta\right) \longrightarrow 0,$$

as $r \to \infty$, $\min(c_i) \to \infty$ *while* n_{ij} *stay fixed, we have*

$$\sqrt{C}\left(F_C^\delta - (1+\theta)\right) \xrightarrow{d} N(0, \Sigma_s),$$

where

$$\Sigma_s = 2 + 4\theta + 2\frac{(1+\theta)^2}{\overline{n}-1} + (\kappa - 3)\frac{(\overline{n}+2\theta)(\overline{n}\underline{n} - 1) + \theta^2(\overline{n} + \underline{n} - 2)}{(\overline{n}-1)^2}.$$

Under the null hypothesis $H_0 : \delta_{ij} = 0$, *which results in* $\theta = 0$, *we then have*

$$\sqrt{C}\left(F_C^\delta - 1\right) \xrightarrow{d} N\left(0,\, 2 + \frac{2}{\overline{n}-1}\frac{(\kappa-3)(\overline{n}^2\underline{n} - \overline{n})}{(\overline{n}-1)^2}\right). \tag{3.1.2}$$

3.1.2 Heteroscedastic Designs

Consider the unbalanced two-fold nested model with (general) heteroscedasticity, as described in (2.1.1). The following theorem giving the asymptotic distributions for the proposed statistic F_C^{**}, defined in (2.2.17), when the numbers of classes and sub-classes are both large.

Theorem 3.1.2. *Consider the model and assumptions given in (2.1.1), (2.1.2), and the decomposition of the means given in (2.1.3). In addition, we assume that there exist κ_{ij}, a_1, b_1, b_2 and b_3 such that as $r \to \infty$ and $\min(c_i) \to \infty$,*

$$E(e_{ijk}^3) = 0, \quad E(e_{ijk}^4) = \kappa_{ij}, \ \textit{and} \ \ E|e_{ijk}|^{4+2\epsilon} < \infty \ \ \textit{for some} \ \epsilon > 0;$$
$$\sqrt{C}\left(\frac{1}{C}\sum_{i=1}^{r}\sum_{j=1}^{c_i}\sigma_{ij}^2 - a_1\right) \longrightarrow 0, \quad \frac{1}{C}\sum_{i=1}^{r}\sum_{j=1}^{c_i}\sigma_{ij}^4 \longrightarrow b_1,$$
$$\frac{1}{C}\sum_{i=1}^{r}\sum_{j=1}^{c_i}\frac{\sigma_{ij}^4}{n_{ij}-1} \longrightarrow b_2, \quad \frac{1}{C}\sum_{i=1}^{r}\sum_{j=1}^{c_i}\frac{\sigma_{ij}^4(\kappa_{ij}-3)}{n_{ij}} \longrightarrow b_3.$$

Then, under alternatives δ_{ij} which satisfy

$$\sqrt{C}\left(\frac{1}{C}\sum_{i=1}^{r}\sum_{j=1}^{c_i} n_{ij}\delta_{ij}^2 - \theta_1\right) \longrightarrow 0, \quad \frac{1}{C}\sum_{i=1}^{r}\sum_{j=1}^{c_i} n_{ij}\delta_{ij}^2\sigma_{ij}^2 \longrightarrow \theta_2,$$

as $r \to \infty$, $\min(c_i) \to \infty$ while n_{ij} stay fixed, we have, with $\theta^* = \theta_1/a_1$,

$$\sqrt{C}\,(F_C^{**} - (1+\theta^*)) \xrightarrow{d} N\left(0, \ \frac{1}{a_1^2}\left[2(b_1+b_2) + 4(\theta_2 + b_2\theta^*) + (2b_2+b_3)\theta^{*2}\right]\right).$$

Under the null hypothesis $H_0 : \delta_{ij} = 0$, which results in $\theta^* = 0$, we then have

$$\sqrt{C}\,(F_C^{**} - 1) \xrightarrow{d} N\left(0, \ \frac{2b_1 + 2b_2}{a_1^2}\right). \tag{3.1.3}$$

3.2 Simulation Studies

In this section, we use simulations to compare the achieved α-levels of several test procedures when the number of classes, r, equal to 5, 30, and 60. For each r, we study different combinations of numbers of sub-classes with the average $\bar{c}$, defined as $C/r = \sum_i c_i/r$, equal to 5, 30, and 100 respectively. More specifically, for $i = 1, \cdots, r-1$, we use $c_i = round(r\ \bar{c}\ \lambda_i)$, where $\lambda_i = 0.02 + (1 - 0.02\ r) \times \frac{2\ i}{r\ (r+1)}$, and c_r is so chosen that

$\sum_{i=1}^{r} c_i = r \cdot \bar{c}$. Then the number of observation in each sub-class (n_{ij}) is generated by truncated Poisson distributions: $n_{ij} = Z_{ij} \times [1 - I(Z_{ij} \leq 3)] + \nu_i \times I(Z_{ij} \leq 3)$, where I is an indication function and $Z_{ij} \sim Poisson(\nu_i),\ i = 1, \cdots, r;\ j = 1, \cdots, c_i$. The values of ν_i used in our simulations are $\nu_i = 2 + 10 \times I(i = 4,\ 14, \cdots, 54)$ for cases under homoscedasticity, while use $\nu_i = 5 + 7 \times I(i = 4,\ 14, \cdots, 54)$ for cases under heteroscedasticity. As for the values of parameters in the decomposition (2.1.3), we use $\mu = 0$, $\alpha_i = -(r/2) + i - 1$ for $i = 1, \cdots, r-1$, and α_r so chosen that $\sum_i n_{i\cdot}\alpha_i = 0$. In addition, $\sigma_{ij} = \sigma = 1$ are used in the cases under homoscedastic designs, while $\sigma_{ij} = 5 \times I(i = 3,\ 13, \cdots, 53) + (j/c_i),\ \forall i = 1, \cdots, r;\ j = 1, \cdots, c_i$, are used when the designs are heteroscedastic.

For each case, four distributions of e_{ijk} are studies: (i) **Normal** : the standard normal; (ii) **Exponen**: the exponential distribution with $\lambda = 1$; (iii) **LogNorm**: the log-normal distribution whose logarithm has mean 0 and standard deviation 1; and (iv) **Mixture**: the mixture distribution defined as $U_1 \cdot X_1 + (1 - U_1) \cdot Y_1$, where $U_1 \sim$ Bernoulli($p = 0.9$), $X_1 \sim N(-1, 1)$ and $Y_1 \sim N(9, 1)$.

3.2.1 Simulations under Homoscedastic Designs

First, simulations are used to compare two test procedures: the classical F-test procedure, shown in (2.1.7), and the proposed test procedure of (3.1.2). Let *CF* and *HOM* denote them respectively. For the latter one, procedure *HOM*, the empirical versions of $\overline{n}$, $\underline{n}$ and κ, denoted as $\hat{\overline{n}}$, $\hat{\underline{n}}$ and $\hat{\kappa}$, are needed. More specifically,

$$\hat{\overline{n}} = \frac{1}{C}\sum_{i=1}^{r}\sum_{j=1}^{c_i} n_{ij}, \quad \hat{\underline{n}} = \frac{1}{C}\sum_{i=1}^{r}\sum_{j=1}^{c_i} \frac{1}{n_{ij}}, \quad \hat{\kappa} = \frac{\sum_{i=1}^{r}\sum_{j=1}^{c_i}\sum_{k=1}^{n_{ij}}(Y_{ijk} - \bar{Y}_{i\cdot\cdot})^4}{N_C \cdot (MSE)^2}.$$

The achieved simulated sizes, based on 10,000 simulation runs, are shown in Table 3.1.

From Table 3.1, we can see that the classical *CF* procedure is liberal in the unbalanced design at all non-normal distributions, especially in the case of the log-normal distribution. On the other hand, the proposed *HOM* procedure performs well with all distributions with the achieved α-levels approaching the nominal one as the number of classes (r) and the average number of sub-classes ($\bar{c}$) are both large.

	$r = 5, \bar{c} = 5$		$r = 30, \bar{c} = 5$		$r = 60, \bar{c} = 5$	
	CF	HOM	CF	HOM	CF	HOM
Normal	0.0500	0.0883	0.0480	0.0768	0.0488	0.0745
Exponen	0.0853	0.0978	0.1203	0.0983	0.1136	0.0921
LogNorm	0.1267	0.0998	0.2084	0.1135	0.2167	0.0997
Mixture	0.0827	0.0901	0.1062	0.0873	0.0938	0.0827
	$r = 5, \bar{c} = 30$		$r = 30, \bar{c} = 30$		$r = 60, \bar{c} = 30$	
	CF	HOM	CF	HOM	CF	HOM
Normal	0.0446	0.0541	0.0500	0.0556	0.0463	0.0519
Exponen	0.1064	0.0690	0.1200	0.0641	0.1206	0.0630
LogNorm	0.1890	0.0681	0.2425	0.0683	0.2715	0.0705
Mixture	0.0911	0.0583	0.0941	0.0563	0.1035	0.0573
	$r = 5, \bar{c} = 100$		$r = 30, \bar{c} = 100$		$r = 60, \bar{c} = 100$	
	CF	HOM	CF	HOM	CF	HOM
Normal	0.0477	0.0520	0.0511	0.0542	0.0504	0.0517
Exponen	0.1120	0.0588	0.1195	0.0555	0.1166	0.0527
LogNorm	0.2344	0.0673	0.2813	0.0659	0.2886	0.0637
Mixture	0.0949	0.0531	0.1005	0.0546	0.0972	0.0512

Table 3.1. Achieved α-levels over $10,000$ simulation runs under homoscedastic and unbalanced design at nominal $\alpha = 0.05$.

3.2.2 Simulations under Heteroscedastic Designs

In this section, we compare three test procedures: the classical F-test procedure, the proposed test procedure of (3.1.2), and the proposed test procedure of (3.1.3). Let *CF* and *HOM* denote the first two as before, while let *HET* denote the last one. Note that the accuracy of procedure *HET* requires consistent estimation of b_1 and b_2, which needs unbiased estimation of σ_{ij}^4 for all $i,\ j$. For such unbiased estimation we use the U-statistics with the kernel $(Y_{ij1} - Y_{ij2})^2/2 \times (Y_{ij3} - Y_{ij4})^2/2$. As a consequence, the application of procedure *HET* requires $n_{ij} \geq 4$, although Theorem 3.1.2 requires only $n_{ij} \geq 2$. The achieved simulated sizes, based on 10,000 simulation runs, are shown in Table 3.2.

As expected, Table 3.2 reveals that the proposed procedure *HET*, although somewhat liberal as r and $\bar{c}$ are small, outperforms both of *CF* and *HOM* procedures in all simulations.

	$r=5,\bar{c}=5$			$r=30,\bar{c}=5$			$r=60,\bar{c}=5$		
	CF	HOM	HET	CF	HOM	HET	CF	HOM	HET
Normal	0.322	0.348	0.120	0.321	0.329	0.115	0.252	0.259	0.096
Exponen	0.313	0.329	0.102	0.301	0.289	0.096	0.244	0.230	0.087
LogNorm	0.284	0.282	0.087	0.285	0.239	0.088	0.242	0.190	0.075
Mixture	0.280	0.281	0.084	0.288	0.272	0.090	0.247	0.233	0.077
	$r=5,\bar{c}=30$			$r=30,\bar{c}=30$			$r=60,\bar{c}=30$		
	CF	HOM	HET	CF	HOM	HET	CF	HOM	HET
Normal	0.613	0.603	0.075	0.421	0.402	0.066	0.536	0.518	0.062
Exponen	0.626	0.570	0.072	0.430	0.367	0.065	0.538	0.473	0.057
LogNorm	0.605	0.437	0.057	0.417	0.246	0.052	0.533	0.332	0.050
Mixture	0.594	0.537	0.060	0.420	0.361	0.053	0.533	0.478	0.056
	$r=5,\bar{c}=100$			$r=30,\bar{c}=100$			$r=60,\bar{c}=100$		
	CF	HOM	HET	CF	HOM	HET	CF	HOM	HET
Normal	0.912	0.906	0.063	0.610	0.586	0.057	0.714	0.691	0.058
Exponen	0.908	0.875	0.058	0.609	0.536	0.054	0.720	0.651	0.052
LogNorm	0.905	0.745	0.050	0.586	0.347	0.043	0.682	0.421	0.045
Mixture	0.917	0.889	0.055	0.603	0.543	0.049	0.704	0.652	0.052

Table 3.2. Achieved α-levels over $10,000$ simulation runs under heteroscedastic and unbalanced design at nominal $\alpha = 0.05$.

3.3 Application: Ramification for Lack-of-Fit testing in Regression

Consider the following heteroscedastic nonparametric regression model with two covariates: conditioning on $X_{1i} = x_{1i}$ and $X_{2i} = x_{2i}$,

$$Y_i = m(x_{1i}, x_{2i}) + \sigma(x_{1i}, x_{2i})\, \epsilon_{Yi}, \quad i = 1, \cdots, N,$$

where $m(\cdot,\cdot)$ and $\sigma(\cdot,\cdot)$ are unknown functions. The test of interest is to check whether the second covariate, X_2, should be included in the model. More specifically, the null hypothesis of interest is $H_0:\ m(x_1, x_2) = m(x_1),\ \forall x_1,\ \forall x_2$.

Consider the following decomposition:

$$m(x_1, x_2) \quad = \quad \mu + \alpha(x_1) + \delta(x_1, x_2)$$

$$\mu = E[m(X_1, X_2)]$$

$$\alpha(x_1) = E[m(X_1, X_2)|X_1 = x_1] - \mu$$

$$\delta(x_1, x_2) = m(x_1, x_2) - E[m(X_1, X_2)|X_1 = x_1]$$

Then, testing

$$H_0: \ \delta(x_1, x_2) = 0, \forall x_1, \ \forall x_2 \quad \Leftrightarrow \quad H_0: \ m(x_1, x_2) = C(x_1), \ \forall x_1, \ \forall x_2.$$

The idea for constructing the test statistic in this regression setting is to discretise the first covariate, X_1, into r "windows" or classes, and then further discretise the second covariate, X_2, nested in class i, into c_i "sub-windows" or sub-classes so that there are n_{ij} observations in each of those sub-classes. Here, we consider the simplest case: when the random errors are homoscedastic and normal (i.e. $\sigma(\cdot, \cdot) = 1$ and $\epsilon_Y \sim N(0, \sigma_Y^2)$) and the design is balanced (i.e. $c_i = c$, and $n_{ij} = n$). We first independently generate X_{1i} and X_{3i} from i.i.d Unif(0,1), $i = 1, \cdots, N$, where $N = r \times c \times n = 30 \times 30 \times 3$. Let $X_2 = 2 \cdot X_1 + sin(4\ \pi\ X_3) + \epsilon_X$, where ϵ_X are i.i.d. $N(0,\ \sigma_X^2 = 1)$. Then, generate the response Y using $Y = 2 \cdot X_1 + \beta_2 \cdot X_2 + \epsilon_Y$, where the random errors ϵ_Y are i.i.d. standard normal. In Table below, we compare our test for the homoscedastic cases (denoted as *HOM*) with the usual sequential test on the regression model $Y = \beta_1 X_1 + \beta_3 X_3 + \epsilon$ with the null hypothesis $H_0: \ \beta_3 = 0$ (denoted as *REG*).

β_2	0.00	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90	1.00
HOM	.053	.143	.669	.995	1.00	1.00	1.00	1.00	1.00	1.00	1.00
REG	.042	.404	.811	.899	.060	.112	1.00	.992	1.00	1.00	.246

Table 3.3. Achieved Powers/α-levels over $1,000$ simulation runs at nominal $\alpha = 0.05$.

Two phenomenon are observed from Table 3.3. First, both procedures *HOM* and *REG* achieve reasonable α-levels (i.e. $\beta_2 = 0$). Second, The power of *HOM* increases when the value of β_2 increases, and achieves the power of 1.00 as $\beta_2 \geq 0.40$; however, the performance of *REG* seems very unstable when $\beta_2 \neq 0$, suggesting its lack of power to detect the effect of the additional covariate under alternatives.

Chapter 4

Asymptotic Theorems in Testing for the Class Effect in Two-fold Nested Model when the number of sub-classes is large

4.1 The Hypotheses with arbitrary Weights

Unlike the crossed models, the nested models under homoscedasticity are *orthogonal* designs for a particular set of weights. The classical two-fold model (2.1.1) with $\sigma_{ij} = \sigma$ and the mean decomposition as specified in (2.1.3) is an example:

$$\mu_{ij} = \mu + \alpha_i + \delta_{ij}, \quad \text{assuming} \quad \sum_{i=1}^{r} n_{i\cdot}\alpha_i = 0 \text{ and } \sum_{j=1}^{c_i} n_{ij}\delta_{ij} = 0, \forall i,$$

where μ, α_i, and δ_{ij} are $\mu = \bar{\mu}_{\cdot\cdot}$, $\quad \alpha_i = \bar{\mu}_{i\cdot} - \bar{\mu}_{\cdot\cdot}$, $\quad \delta_{ij} = \mu_{ij} - \bar{\mu}_{i\cdot}$,

$$\bar{\mu}_{i\cdot} = \frac{1}{n_{i\cdot}} \sum_{j=1}^{c_i} n_{ij}\mu_{ij}, \quad \bar{\mu}_{\cdot\cdot} = \frac{1}{N_C} \sum_{i=1}^{r} n_{i\cdot}\bar{\mu}_{i\cdot}.$$

In this model, the null hypothesis of no sub-class effect is stated as "μ_{ij} does not depend on j", or equivalently $H_0^\delta: \ \delta_{ij} = 0$, as shown in Section 2.1, while the null hypothesis of no class effect is often specified as "$\bar{\mu}_{i\cdot}$ does not depend on i", or equivalently $H_0^\alpha: \alpha_i = 0 \Leftrightarrow \bar{\mu}_{i\cdot} = \bar{\mu}_{\cdot\cdot}$. The classical F test statistic and the corresponding procedure under normality and homoscedasticity for $H_0: \ \delta_{ij} = 0$ are given in (2.1.6) and (2.1.7). As for the class effect, the classical F test statistic for testing $H_0: \ \alpha_i = 0$ is

$$F_C^\alpha = \frac{MS\alpha}{MSE}, \quad \text{where} \quad MS\alpha = \frac{\sum_{i=1}^r n_{i\cdot}(\bar{Y}_{i\cdot\cdot} - \bar{Y}_{\cdot\cdot\cdot})^2}{r-1}, \tag{4.1.1}$$

and MSE is as defined in (2.1.5). The $\bar{Y}_{\cdot\cdot\cdot}$ and $\bar{Y}_{i\cdot\cdot}$ in the above definition (4.1.1) are the unweighted sample overall means and the unweighted sample class means of Y_{ijk}, i.e.

$$\bar{Y}_{\cdot\cdot\cdot} = \frac{1}{N_C}\sum_{i=1}^r n_{i\cdot}\bar{Y}_{i\cdot\cdot}, \quad \bar{Y}_{i\cdot\cdot} = \frac{1}{n_{i\cdot}}\sum_{j=1}^{c_i} n_{ij}\bar{Y}_{ij\cdot}, \quad \bar{Y}_{ij\cdot} = \frac{1}{n_{ij}}\sum_{k=1}^{n_{ij}} Y_{ijk}.$$

Then, under normality and homoscedasticity,

$$F_C^\alpha \sim F_{r-1,\, N_C - C}, \text{ under } H_0: \ \alpha_i = 0. \tag{4.1.2}$$

It can be shown that the hypothesis of no sub-class effect is unaffected by the weights in the definition; however, the hypothesis of no class effect is actually testing the equality of $\bar{\mu}_{i\cdot}$, which is equivalently testing the *weighted averages* of the sub-class means weighted by the *sample sizes*. This might not be a sensible way to test for the class effect, since two researchers investigating the same class effects are in fact testing different hypotheses if the sample sizes used in their studies are not the same. As a consequence, a better way of testing for the class effect in the nested models might be to choose the appropriate weights w_{ij} for the sub-classes, based on their relative importance within a class, and then decide the sampling scheme accordingly. More details and examples can be found in Arnold (1981). This chapter is established for providing the suitable test procedures for such two-fold nested models with arbitrary weights.

We define the average performance of each class, based on the chosen weights w_{ij} on sub-classes, by

$$\bar{\mu}_{i\cdot}^{w} = \frac{\sum_{j=1}^{c_i} w_{ij}\mu_{ij}}{\sum_{j=1}^{c_i} w_{ij}} = \frac{1}{w_{i\cdot}}\sum_{j=1}^{c_i} w_{ij}\mu_{ij}, \quad \text{where } w_{i\cdot} = \sum_{j=1}^{c_i} w_{ij}, \tag{4.1.3}$$

and test the equality of the weighted means $\bar{\mu}_{i\cdot}^{w}$,

$$H_0^w : \bar{\mu}_{i\cdot}^{w} \text{ does not depend on } i \quad \Leftrightarrow \quad \bar{\mu}_{i\cdot}^{w} = \bar{\mu}_{\cdot\cdot}^{w}, \tag{4.1.4}$$

where

$$\bar{\mu}_{\cdot\cdot}^{w} = \frac{1}{Q_\cdot}\sum_{i=1}^{r} Q_i\bar{\mu}_{i\cdot}^{w}, \quad \frac{1}{Q_i} = \frac{1}{w_{i\cdot}^2}\sum_{j=1}^{c_i}\frac{w_{ij}^2}{n_{ij}}, \quad Q_\cdot = \sum_{i=1}^{r} Q_i,$$

as our null hypothesis of no class effect.

Further define

$$A_i = \frac{1}{w_{i\cdot}}\sum_{j=1}^{c_i} w_{ij}\bar{Y}_{ij\cdot}, \quad \bar{A} = \frac{1}{Q_\cdot}\sum_{i=1}^{r} Q_i A_i, \tag{4.1.5}$$

where $\bar{Y}_{ij\cdot}$ are the unweighted means of Y_{ijk} within each sub-class, as defined before. The asymptotic results based on A_i and $\bar{A}$ for testing H_0^w in the nested models, both under homoscedasticity and under heteroscedasticity, are given in the next section, while the details of these theoretical derivations can be found in Section 4.3.

4.2 Main Results

4.2.1 Homoscedastic Designs

In this subsection we consider the unbalanced two-fold nested models with arbitrary weights under homoscedasticity, and derive the asymptotic results for testing for the class effect both under the null hypothesis H_0^w as specified in (4.2.3) and under some alternatives. As a corollary of Theorem 4.2.1 below, we obtain that the usual, normal-based, F-test procedure for the class effect is robust to departures from the normality assumption if the number of sub-classes is large, no matter whether the model is unbalanced.

Lemma 4.2.1. *Consider the model and assumptions given in (2.1.1) with* $\sigma_{ij} = \sigma$*, (2.1.2), and (2.1.8). In addition assume that there exists some positive* $\epsilon > 0$ *such that*

$$\limsup \frac{1}{n_{i\cdot}} \sum_{j=1}^{c_i} \sum_{k=1}^{n_{ij}} E|e_{ijk}|^{2+\epsilon} < \infty, \ \forall i.$$

Then, under the null hypothesis $H_0^w : \bar{\mu}_{i\cdot}^w = \bar{\mu}_{\cdot\cdot}^w$ *with arbitrary* w_{ij}*, as specified in (4.1.4), we have*

$$\frac{1}{\sigma^2} \sum_{i=1}^{r} Q_i \left(A_i - \bar{A}\right)^2 \xrightarrow{d} \chi^2_{(r-1)}, \quad as \ \min(c_i) \to \infty, \tag{4.2.1}$$

where A_i *and* $\bar{A}$ *are as defined in (4.1.5).*

In addition, under alternatives $\bar{\mu}_{i\cdot}^w$ *satisfying*

$$\frac{1}{\sigma^2} \sum_{i=1}^{r} Q_i \left(\bar{\mu}_{i\cdot}^w - \bar{\mu}_{\cdot\cdot}^w\right)^2 \to \eta, \quad for\ some\ \eta \in (0, \infty), \tag{4.2.2}$$

as $\min(c_i) \to \infty$ *while* $r,\ n_{ij}$ *stay fixed,*

$$\frac{1}{\sigma^2} \sum_{i=1}^{r} Q_i \left(A_i - \bar{A}\right)^2 \xrightarrow{d} \chi^2_{(r-1)}(\eta). \tag{4.2.3}$$

Theorem 4.2.1. *Consider the model and assumptions given in Lemma 4.2.1. Provided that* $C^{-2} \sum_j \sum_k E\left(e_{ijk}^4\right) \to 0,\ \forall i,$ *as* $\min(c_i) \to \infty$*, under the null hypothesis* H_0^w *of no class effect, we have*

$$\frac{\sum_{i=1}^{r} Q_i \left(A_i - \bar{A}\right)^2}{MSE} \xrightarrow{d} \chi^2_{(r-1)}, \quad as \ \min(c_i) \to \infty,$$

where MSE *as defined in (2.1.5).*

Corollary 4.2.1. *Under the model and assumptions of Theorem 4.2.1, the classical, normality-based,* F*-test procedure for the hypothesis* $H_0^\alpha : \alpha_i = 0$*, shown in (4.1.2), is asymptotically* valid *even when the model is not normal.*

4.2.2 Heteroscedastic Designs

Lemma 4.2.2. *Consider the model and assumptions given in (2.1.1), (2.1.2), and (2.1.8), and let* $\bar{\mu}_{i\cdot}$, $\bar{\mu}_{\cdot\cdot}$, Q_i, $Q_\cdot$, A_i *and* $\bar{A}$ *be as defined in (4.1.3), (4.1.4) and (4.1.5) with arbitrary weights* w_{ij}. *Assume that there exist some positive* $\epsilon > 0$ *and* $\sigma^2_{A_i} \in (0, \infty),\ i = 1, \cdots, k$, *such that*

$$\limsup \frac{1}{n_{i\cdot}} \sum_{j=1}^{c_i} \sum_{k=1}^{n_{ij}} E|e_{ijk}|^{2+\epsilon} < \infty,\ \forall i.$$

and

$$Var\left(\sqrt{Q_\cdot} A_i\right) = \frac{Q_\cdot}{w_{i\cdot}^2} \sum_{j=1}^{c_i} \frac{w_{ij}^2}{n_{ij}} \sigma^2_{ij} \longrightarrow \sigma^2_{A_i}, \quad as \ \ \min{(c_i)} \to \infty. \tag{4.2.4}$$

Define the contrast matrix $\mathbf{H}$ *as*

$$\mathbf{H} = \begin{pmatrix} 1 & -1 & 0 & \cdots & 0 \\ 1 & 0 & -1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & 0 & \cdots & -1 \end{pmatrix}_{(r-1)\times r} = \left(1'_{r-1} \mid -\mathbf{I}_{r-1}\right).$$

Let $\mathbf{A} = (A_1, \cdots, A_r)'$, $\boldsymbol{\mu}_A = (\bar{\mu}^w_{1\cdot}, \cdots, \bar{\mu}^w_{r\cdot})'$, *and* $\mathbf{V}_A = diag(\sigma^2_{A_1}, \cdots, \sigma^2_{A_r})$. *Then, under the null hypothesis* $H_0^w: \ \bar{\mu}^w_{i\cdot} = \bar{\mu}^w_{\cdot\cdot}$, *which results in* $\mathbf{H}\boldsymbol{\mu}_A \overset{H_0^w}{=} \mathbf{H}\bar{\mu}^w_{\cdot\cdot}1_r = 0_{r-1}$, *we have*

$$Q_\cdot \left(\mathbf{HA}\right)' \left(\mathbf{H}\mathbf{V}_A\mathbf{H}'\right)^{-1} \left(\mathbf{HA}\right) \xrightarrow{d} \chi^2_{(r-1)}, \quad as \ \ \min{(c_i)} \to \infty. \tag{4.2.5}$$

Under alternatives $\bar{\mu}^w_{i\cdot}$ *satisfying*

$$Q_\cdot \left(\mathbf{H}\boldsymbol{\mu}_A\right)' \left(\mathbf{H}\mathbf{V}_A\mathbf{H}'\right)^{-1} \left(\mathbf{H}\boldsymbol{\mu}_A\right) \to \eta, \quad for\ some\ \eta \in (0, \infty), \tag{4.2.6}$$

as $\min(c_i) \to \infty$ *while* r, n_{ij} *stay fixed, we obtain*

$$Q_{\cdot}\left(\mathbf{HA}\right)'\left(\mathbf{HV}_A\mathbf{H}'\right)^{-1}\left(\mathbf{HA}\right) \xrightarrow{d} \chi^2_{(r-1)}(\eta). \tag{4.2.7}$$

Theorem 4.2.2. *Consider the model and assumptions given in Lemma 4.2.1. Define* $\widehat{\mathbf{V}}_A = diag(\hat{\sigma}^2_{A_1}, \cdots, \hat{\sigma}^2_{A_r})$ *where*

$$\hat{\sigma}^2_{A_i} = \frac{Q_{\cdot}}{w^2_{i\cdot}}\sum_{j=1}^{c_i}\frac{w^2_{ij}}{n_{ij}}\hat{\sigma}^2_{ij}, \quad and \quad \hat{\sigma}^2_{ij} = \frac{1}{n_{ij}-1}\sum_{k=1}^{n_{ij}}\left(Y_{ijk} - \bar{Y}_{ij\cdot}\right)^2. \tag{4.2.8}$$

Provided that $E\left(e^4_{ijk}\right)$ *are all bounded away from* 0 *and* ∞*, under the null hypothesis* H^w_0 *of no class effect, we have*

$$Q_{\cdot}\left(\mathbf{HA}\right)'\left(\mathbf{H}\widehat{\mathbf{V}}_A\mathbf{H}'\right)^{-1}\left(\mathbf{HA}\right) \xrightarrow{d} \chi^2_{(r-1)}, \quad as \ \min(c_i) \to \infty.$$

Using the Sherman-Marrison-Woodbury Theorem to obtain the inverse of $\mathbf{HV}_A\mathbf{H}'$, along with some algebra calculation (see Section 4.3), one can get

$$\begin{aligned} & Q_{\cdot}\left(\mathbf{HA}\right)'\left(\mathbf{HV}_A\mathbf{H}'\right)^{-1}\left(\mathbf{HA}\right) \\ = \ & Q_{\cdot}\mathbf{A}'\left[\mathbf{V}_A^{-1} - \frac{1}{\mathbf{1}_r'\mathbf{V}_A^{-1}\mathbf{1}_r}\mathbf{V}_A^{-1}\mathbf{J}_r\mathbf{V}_A^{-1}\right]\mathbf{A} \\ = \ & Q_{\cdot}\left(\sum_{i=1}^{r}\frac{A_i^2}{\sigma^2_{A_i}}\right) - \frac{Q_{\cdot}}{\left(\sum_{i=1}^{r}\frac{1}{\sigma^2_{A_i}}\right)}\left(\sum_{i=1}^{r}\frac{A_i}{\sigma^2_{A_i}}\right)^2 \end{aligned} \tag{4.2.9}$$

where $\mathbf{J}_r = \mathbf{1}_r\mathbf{1}_r'$ is a $r \times r$ matrix with all elements equal to one.

This relation is the main tool used to prove the asymptotic equivalence under Homoscedasticity between the test procedures based on Theorem 4.2.1 and Theorem 4.2.2, as stated in the corollary below.

Corollary 4.2.2. *Consider the model and assumptions of Theorem 4.2.2. Under Homoscedasticity, i.e.* $\sigma_{ij} = \sigma$, $Q_{\cdot}\left(\mathbf{HA}\right)'\left(\mathbf{HV}_A\mathbf{H}'\right)^{-1}\left(\mathbf{HA}\right)$ *in Lemma 4.2.2 is asymptotically equivalent to* $\sum_{i=1}^{r} Q_i\left(A_i - \bar{A}\right)^2/\sigma^2$ *in Lemma 4.2.1.*

In addition, $Q_{\cdot}\left(\mathbf{HA}\right)'\left(\mathbf{H}\widehat{\mathbf{V}}_A\mathbf{H}'\right)^{-1}\left(\mathbf{HA}\right)/(r-1)$ *in Theorem 4.2.2 is asymptotically equivalent to the classical F test statistic* F_C^{α}*, as defined in (4.1.2), under homoscedasticity.*

4.3 Proofs

Proof of Lemma 4.2.1

Recall that

$$A_i = \frac{1}{w_{i\cdot}}\sum_{j=1}^{c_i} w_{ij}\bar{Y}_{ij\cdot} = \frac{1}{w_{i\cdot}}\sum_{j=1}^{c_i}\sum_{k=1}^{n_{ij}}\frac{w_{ij}}{n_{ij}}Y_{ijk}$$

and note that

$$\begin{aligned} E(A_i) &= \frac{1}{w_{i\cdot}}\sum_{j=1}^{c_i} w_{ij}\mu_{ij} = \bar{\mu}_{i\cdot}^{w}, \\ Var(A_i) &= \frac{1}{w_{i\cdot}^2}\sum_{j=1}^{c_i} w_{ij}^2\, Var(\mu_{ij}+\sigma\cdot\bar{e}_{ij\cdot}) \ = \ \frac{\sigma^2}{Q_i}. \end{aligned}$$

Let

$$m = \liminf\left(\frac{w_{ij}}{n_{ij}}\right), \quad M = \limsup\left(\frac{w_{ij}}{n_{ij}}\right).$$

To prove the asymptotic normality of A_i, or more specifically, for the limiting distribution of $\sum_j\sum_k w_{ij}(Y_ijk-\mu_{ij})/n_{ij}$, we check the Lyapounov condition: $\exists\epsilon\in(0,\infty)$,

$$\begin{aligned} Ł_C^{\epsilon} &= \frac{1}{\left(\sqrt{\sum_{j=1}^{c_i}\sum_{k=1}^{n_{ij}}\frac{w_{ij}^2}{n_{ij}^2}\sigma^2}\right)^{2+\epsilon}}\sum_{j=1}^{c_i}\sum_{k=1}^{n_{ij}} E\left|\frac{w_{ij}}{n_{ij}}\left(Y_{ijk}-\mu_{ij}\right)\right|^{2+\epsilon} \\ &\leq \frac{M^{2+\epsilon}\cdot\sigma^{2+\epsilon}}{(\sigma m)^{2+\epsilon}(n_{i\cdot})^{\epsilon/2}}\cdot\frac{1}{n_{i\cdot}}\sum_{j=1}^{c_i}\sum_{k=1}^{n_{ij}} E\left|e_{ijk}\right|^{2+\epsilon} \longrightarrow 0, \quad \text{as } \min\left(c_i\right)\to\infty. \end{aligned}$$

Since the Lyapounov condition holds, we know that, as $\min(n_i) \to \infty$,

$$\frac{A_i - \bar{\mu}^w_{i\cdot}}{\sigma/\sqrt{Q_i}} = \frac{\sum_j \sum_k \frac{w_{ij}}{n_{ij}} (Y_{ijk} - \mu_{ij})}{\sqrt{\sum_j \sum_k Var\left[\cdot\frac{w_{ij}}{n_{ij}} (Y_{ijk} - \mu_{ij})\right]}} \xrightarrow{d} N(0,1).$$

Note that

$$\bar{A} = \frac{1}{Q.} \sum_{i=1}^{r} Q_i A_i \implies E(\bar{A}) = \bar{\mu}^w_{\cdot\cdot} \quad \text{and} \quad Var(\bar{A}) = \frac{\sigma^2}{Q.},$$

and hence as $\min(c_i) \to \infty$,

$$\frac{\sqrt{Q.}}{\sigma}\left(\bar{A} - \bar{\mu}^w_{\cdot\cdot}\right) \xrightarrow{d} N(0,1) \implies \frac{Q.}{\sigma^2}\left(\bar{A} - \bar{\mu}^w_{\cdot\cdot}\right)^2 \xrightarrow{d} \chi^2_{(1)}.$$

In addition, note that

$$\frac{\sqrt{Q_i}}{\sigma}(A_i - \bar{\mu}^w_{\cdot\cdot}) = \frac{\sqrt{Q_i}}{\sigma}(A_i - \bar{\mu}^w_{i\cdot}) + \frac{\sqrt{Q_i}}{\sigma}(\bar{\mu}^w_{i\cdot} - \bar{\mu}^w_{\cdot\cdot}),$$

suggesting that

$$\sum_{i=1}^{r} \frac{Q_i}{\sigma^2}(A_i - \bar{\mu}^w_{\cdot\cdot})^2 \quad \xrightarrow{d} \quad \chi^2_{(r)}(\eta),$$

where η is as defined in (4.2.2).

Then, (4.2.3) follows the fact that $\bar{A}$ and $\sum_i Q_i(A_i - \bar{A})^2$ are independent, and

$$\sum_i Q_i (A_i - \bar{\mu}^w_{\cdot\cdot})^2 = \sum_i Q_i \left(A_i - \bar{A}\right)^2 + \sum_i Q_i \left(\bar{A} - \bar{\mu}^w_{\cdot\cdot}\right)^2, \text{ as } \sum_i Q_i \left(A_i - \bar{A}\right) = 0.$$

The null distribution (4.2.1) is a straightforward result from $H^w_0 : \bar{\mu}^w_{i\cdot} = \bar{\mu}^w_{\cdot\cdot}, \forall i \Rightarrow \eta = 0$.

Proof of Theorem 4.2.1

By WLLN, $MSE \xrightarrow{P} \sigma^2$. The theorem then follows the Slutsky's Theorem.

Proof of Corollary 4.2.1

Take the weights $w_{ij} = n_{ij}$. Then, $w_{i\cdot} = n_{i\cdot} = Q_i$, $Q_\cdot = N_C$, $\bar{\mu}^w_{i\cdot} = \bar{\mu}_{i\cdot}$, $\bar{\mu}^w_{\cdot\cdot} = \bar{\mu}_{\cdot\cdot}$, $A_i = \bar{Y}_{i\cdot\cdot}$ and $\bar{A} = \bar{Y}_{\cdot\cdot\cdot}$. Hence, testing $H^w_0: \ \bar{\mu}^w_{i\cdot} = \bar{\mu}^w_{\cdot\cdot}, \ \forall i$ is clearly equivalent to testing $H_0: \ \alpha_i = \bar{\mu}_{i\cdot} - \bar{\mu}_{\cdot\cdot} = 0, \forall i$, and

$$\sum_i Q_i \left(A_i - \bar{A}\right)^2 = \sum_i n_{i\cdot} \left(\bar{Y}_{i\cdot\cdot} - \bar{Y}_{\cdot\cdot\cdot}\right)^2 = (r-1)MS\alpha.$$

Then, this corollary follows from Theorem 4.2.1.

Proof of Lemma 4.2.2

Let

$$m^\sigma = \liminf \left(\frac{w_{ij}}{n_{ij}}\sigma_{ij}\right), \quad M^\sigma = \limsup \left(\frac{w_{ij}}{n_{ij}}\sigma_{ij}\right),$$

and note that

$$E(A_i) = \bar{\mu}^w_{i\cdot}, \quad Var(A_i) = \frac{1}{w^2_{i\cdot}} \sum_{j=1}^{c_i} \frac{w^2_{ij}}{n_{ij}} \sigma^2_{ij} \ \Rightarrow \ \sqrt{Q_\cdot} Var(A_i) \longrightarrow \sigma^2_{A_i},$$

where $\sigma^2_{A_i}$ are as defined in (4.2.4).

Check the Lyapounov condition: $\exists \epsilon \in (0, \infty)$,

$$\begin{aligned} Ł^\epsilon_C &= \frac{1}{\left(\sqrt{\sum_{j=1}^{c_i}\sum_{k=1}^{n_{ij}} \frac{w^2_{ij}}{n^2_{ij}}\sigma^2_{ij}}\right)^{2+\epsilon}} \sum_{j=1}^{c_i}\sum_{k=1}^{n_{ij}} E\left|\frac{w_{ij}}{n_{ij}}\left(Y_{ijk} - \mu_{ij}\right)\right|^{2+\epsilon} \\ &\leq \frac{(M^\sigma)^{2+\epsilon}}{(m^\sigma)^{2+\epsilon}(n_{i\cdot})^{\epsilon/2}} \cdot \frac{1}{n_{i\cdot}} \sum_{j=1}^{c_i}\sum_{k=1}^{n_{ij}} E\left|e_{ijk}\right|^{2+\epsilon} \longrightarrow 0, \quad \text{as } \min(c_i) \to \infty, \end{aligned}$$

and hence

$$\sqrt{Q_\cdot}\left(A_i - \bar{\mu}^w_{i\cdot}\right) \xrightarrow{d} N(0, \sigma^2_{A_i}), \ \forall i \quad \Rightarrow \quad \sqrt{Q_\cdot}\left(\mathbf{A} - \boldsymbol{\mu}_A\right) \xrightarrow{d} N_r(\mathbf{0}_r, \mathbf{V}_A).$$

Under $H_0^w: \ \bar{\mu}_{i\cdot}^w = \bar{\mu}_{\cdot\cdot}^w, \ \forall i \ \Rightarrow \ \mathbf{H}\boldsymbol{\mu}_A \overset{H_0^w}{=} \mathbf{H}\bar{\mu}_{\cdot\cdot}^w \mathbf{1}_r = \mathbf{0}_{r-1}$, we have

$$\sqrt{Q}.\mathbf{HA} \xrightarrow{d} N_{r-1}(\mathbf{0}_{r-1}, \mathbf{HV}_A\mathbf{H}') \tag{4.3.1}$$

$$\Longrightarrow \quad Q.\left(\mathbf{HA}\right)'\left(\mathbf{HV}_A\mathbf{H}'\right)^{-1}\left(\mathbf{HA}\right) \xrightarrow{d} \chi^2_{(r-1)}, \text{ as } \min\left(c_i\right) \to \infty.$$

On the other hand, under alternatives $\bar{\mu}_{i\cdot}^w$ satisfying (4.2.6), in which η is defined,

$$Q.\left(\mathbf{HA}\right)'\left(\mathbf{HV}_A\mathbf{H}'\right)^{-1}\left(\mathbf{HA}\right) \xrightarrow{d} \chi^2_{(r-1)}(\eta).$$

Proof of Theorem 4.2.2

Since $\hat{\sigma}_{ij}^2$ are unbiased estimators of σ_{ij} and $E(e_{ijk}^4)$ are bounded, by WLLN,

$$\hat{\sigma}_{A_i}^2 \xrightarrow{P} \sigma_{A_i}^2, \quad \text{as } \min\left(c_i\right) \to \infty, \ \forall i = 1, \cdots, r.$$

As r is fixed/finite, we further have $\hat{\sigma}_{A_i}^2 - \sigma_{A_i}^2 \xrightarrow{P} 0$, uniformly in i. Since $\mathbf{V}_A$ and $\widehat{\mathbf{V}}_A$ are $r \times r$ diagonal matrices with elements $\sigma_{A_i}^2$ and $\hat{\sigma}_{A_i}^2$, $\widehat{\mathbf{V}}_A - \mathbf{V}_A \xrightarrow{P} 0$ follows.

In addition, because for any $r \times r$ squared matrix $\mathbf{C}$, the elements of $\mathbf{HCH}'$ are continuous functions of the elements of $\mathbf{C}$, and the dimension r is finite, by CMT (continuous mapping theorem), we get

$$\mathbf{H}\widehat{\mathbf{V}}_A\mathbf{H}' - \mathbf{HV}_A\mathbf{H}' \xrightarrow{P} 0, \quad \text{as } \min\left(c_i\right) \to \infty.$$

Similarly, by the fact that the elements of any inverse matrix with finite dimensions are continuous functions of the elements of the original matrix, CMT further tells us that

$$\left(\mathbf{H}\widehat{\mathbf{V}}_A\mathbf{H}'\right)^{-1} - \left(\mathbf{HV}_A\mathbf{H}'\right)^{-1} \xrightarrow{P} 0, \quad \text{as } \min\left(c_i\right) \to \infty.$$

Under the null hypothesis H_0^w, by (4.3.1) and the Slutsky's theorem,

$$\left(\sqrt{Q}.\mathbf{HA}\right)'\left[\left(\mathbf{H}\widehat{\mathbf{V}}_A\mathbf{H}'\right)^{-1} - \left(\mathbf{HV}_A\mathbf{H}'\right)^{-1}\right]\left(\sqrt{Q}.\mathbf{HA}\right) \xrightarrow{P} 0, \quad \text{as } \min\left(c_i\right) \to \infty,$$

which completes the proof.

Calculation of Relation (4.2.9)

Let $D = diag(\sigma^2_{A_2}, \cdots, \sigma^2_{A_r})$. Then,

$$\begin{aligned}
\mathbf{H}\mathbf{V}_A\mathbf{H}' &= (\mathbf{1}_{r-1} \mid -\mathbf{I}_{r-1})_{(r-1)\times r} \begin{pmatrix} \sigma^2_{A_1} & \mathbf{0}'_{r-1} \\ \mathbf{0}_{r-1} & D \end{pmatrix} \begin{pmatrix} \mathbf{1}'_{r-1} \\ -\mathbf{I}_{r-1} \end{pmatrix}_{r\times(r-1)} \\
&= (\mathbf{1}_{r-1} \mid -\mathbf{I}_{r-1})_{(r-1)\times r} \begin{pmatrix} \sigma^2_{A_1}\mathbf{1}'_{r-1} \\ -D \end{pmatrix} \\
&= \sigma^2_{A_1}\mathbf{1}_{r-1}\mathbf{1}'_{r-1} + D.
\end{aligned}$$

Recall the Sherman-Marrison-Woodbury Theorem:

$$\left(\Gamma_{p\times p} + \theta \mathbf{x}_{p\times 1}\mathbf{x}'_{1\times p}\right)^{-1} = \Gamma^{-1} - \frac{\theta}{1+\theta\mathbf{x}'\Gamma^{-1}\mathbf{x}}\Gamma^{-1}\mathbf{x}\mathbf{x}'\Gamma^{-1}.$$

Hence,

$$\begin{aligned}
\left(\mathbf{H}\mathbf{V}_A\mathbf{H}'\right)^{-1} &= D^{-1} - \frac{\sigma^2_{A_1}}{1+\sigma^2_{A_1}\mathbf{1}'_{r-1}D^{-1}\mathbf{1}_{r-1}} D^{-1}\mathbf{1}_{r-1}\mathbf{1}'_{r-1}D^{-1} \\
\Longrightarrow \mathbf{H}'\left(\mathbf{H}\mathbf{V}_A\mathbf{H}'\right)^{-1}\mathbf{H} &= \mathbf{H}'D^{-1}\mathbf{H} - \frac{1}{\left(\sum_{i=1}^r \frac{1}{\sigma^2_{A_i}}\right)}\mathbf{H}'D^{-1}\mathbf{1}_{r-1}\mathbf{1}'_{r-1}D^{-1}\mathbf{H}.
\end{aligned}$$

Because

$$\begin{aligned}
\mathbf{H}'D^{-1}\mathbf{H} &= \begin{pmatrix} \mathbf{1}'_{r-1} \\ -\mathbf{I}_{r-1} \end{pmatrix}_{r\times(r-1)} D^{-1} \ (\mathbf{1}_{r-1} \mid \mathbf{I}_{r-1})_{(r-1)\times r} \\
&= \begin{pmatrix} 1/\sigma^2_{A_2} & \cdots & 1/\sigma^2_{A_r} \\ & -D^{-1} & \end{pmatrix}_{r\times(r-1)} (\mathbf{1}_{r-1} \mid \mathbf{I}_{r-1})_{(r-1)\times r}
\end{aligned}$$

$$= \begin{pmatrix} \zeta & -\frac{1}{\sigma^2_{A_2}} & \cdots & -\frac{1}{\sigma^2_{A_r}} \\ -\frac{1}{\sigma^2_{A_2}} & -\frac{1}{\sigma^2_{A_2}} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ -\frac{1}{\sigma^2_{A_r}} & 0 & \cdots & -\frac{1}{\sigma^2_{A_r}} \end{pmatrix}_{r\times r}, \text{ where } \zeta = \left(\sum_{i=2}^{r} \frac{1}{\sigma^2_{A_i}}\right),$$

and

$$\begin{aligned} \mathbf{H}'D^{-1}\mathbf{1}_{r-1}\mathbf{1}'_{r-1}D^{-1}\mathbf{H} &= \begin{pmatrix} \zeta \\ -\frac{1}{\sigma^2_{A_2}} \\ \vdots \\ -\frac{1}{\sigma^2_{A_r}} \end{pmatrix}_{r\times 1} \begin{pmatrix} \zeta, & -\frac{1}{\sigma^2_{A_2}}, & \cdots, & -\frac{1}{\sigma^2_{A_r}} \end{pmatrix}_{1\times r} \\ &= \begin{pmatrix} \zeta^2 & -\frac{1}{\sigma^2_{A_2}}\cdot\zeta & \cdots & -\frac{1}{\sigma^2_{A_r}}\cdot\zeta \\ -\frac{1}{\sigma^2_{A_2}}\cdot\zeta & \frac{1}{\left(\sigma^2_{A_2}\right)^2} & \cdots & \frac{1}{\sigma^2_{A_2}}\frac{1}{\sigma^2_{A_r}} \\ \vdots & \vdots & \ddots & \vdots \\ -\frac{1}{\sigma^2_{A_r}}\cdot\zeta & \frac{1}{\sigma^2_{A_2}}\frac{1}{\sigma^2_{A_r}} & \cdots & \frac{1}{\left(\sigma^2_{A_r}\right)^2} \end{pmatrix}_{r\times r}, \end{aligned}$$

we obtain

$$\begin{aligned} &\left(\frac{1}{\sum_{i=1}^{r}\sigma^2_{A_i}}\right)\mathbf{H}'\left(\mathbf{H}\mathbf{V}_A\mathbf{H}'\right)^{-1}\mathbf{H} \\ &= \begin{pmatrix} \frac{1}{\sigma^2_{A_1}}\left(\sum_{i\neq 1}\frac{1}{\sigma^2_{A_i}}\right) & -\frac{1}{\sigma^2_{A_1}\sigma^2_{A_2}} & \cdots & -\frac{1}{\sigma^2_{A_1}\sigma^2_{A_r}} \\ -\frac{1}{\sigma^2_{A_2}\sigma^2_{A_1}} & \frac{1}{\sigma^2_{A_2}}\left(\sum_{i\neq 2}\frac{1}{\sigma^2_{A_i}}\right) & \cdots & -\frac{1}{\sigma^2_{A_2}\sigma^2_{A_r}} \\ \vdots & \vdots & \ddots & \vdots \\ -\frac{1}{\sigma^2_{A_r}\sigma^2_{A_1}} & -\frac{1}{\sigma^2_{A_r}\sigma^2_{A_2}} & \cdots & \frac{1}{\sigma^2_{A_r}}\left(\sum_{i\neq r}\frac{1}{\sigma^2_{A_i}}\right) \end{pmatrix}_{r\times r}. \end{aligned}$$

On the other hand,

$$\left(\mathbf{1}'_r\mathbf{V}_A^{-1}\mathbf{1}_r\right)\mathbf{V}_A^{-1} - \mathbf{V}_A^{-1}\mathbf{1}_r\mathbf{1}'_r\mathbf{V}_A^{-1}$$

$$= \left(\sum_{i=1}^{r} v \frac{1}{\sigma_{A_i}^2}\right) \begin{pmatrix} 1/\sigma_{A_1}^2 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 1/\sigma_{A_r}^2 \end{pmatrix} - \begin{pmatrix} 1/\sigma_{A_1}^2 \\ \vdots \\ 1/\sigma_{A_r}^2 \end{pmatrix} \begin{pmatrix} 1/\sigma_{A_1}^2 & \cdots & 1/\sigma_{A_r}^2 \end{pmatrix}$$

$$= (m_{ij})_{r\times r}, \quad \text{where } m_{ij} = \begin{cases} \frac{1}{\sigma_{A_i}^2}\left(\sum_{l\neq i} \frac{1}{\sigma_{A_l}^2}\right), & \text{if } i = j; \\ -\frac{1}{\sigma_{A_i}^2 \cdot \sigma_{A_j}^2}, & \text{if } i \neq j; \end{cases}$$

which completes the proof of Relation (4.2.9).

REMARK Let $\mathbf{W} = \mathbf{V}_A^{-1} - \mathbf{V}_A^{-1}\mathbf{J}_r\mathbf{V}_A^{-1}/\mathbf{1}_r^{'}\mathbf{V}_A^{-1}\mathbf{1}_r$, which is a $r \times r$ square matrix of rank $(r-1)$. It can be shown that (1) $\mathbf{W1}_r = \mathbf{0}_r$ and (2) $\mathbf{WV}_A\mathbf{W} = \mathbf{W}$, i.e. $\mathbf{V}_A$ is a generalized inverse of $\mathbf{W}$ (see Lemma A.1. of Akritas et al. (1995)).

Hence, Relation (4.2.9) may be further extended to

$$\begin{aligned} Q_{\cdot}\left(\mathbf{HA}\right)^{'}\left(\mathbf{HV}_A\mathbf{H}^{'}\right)^{-1}\left(\mathbf{HA}\right) &= Q_{\cdot}\mathbf{A}^{'}\mathbf{WA} \\ &= Q_{\cdot}\left(\mathbf{WA}\right)^{'}\left(\mathbf{WV}_A\mathbf{W}^{'}\right)^{-}\left(\mathbf{WA}\right). \end{aligned}$$

Note that under $H_0^w: \ \bar{\mu}_{i\cdot}^w = \bar{\mu}_{\cdot\cdot}^w, \ \forall i$, not only $\mathbf{HA} = \mathbf{0}_{r-1}$ but also $\mathbf{WA} = \mathbf{0}_r$.

Proof of Corollary 4.2.2

When $\sigma_{ij} = \sigma$, (4.2.4) tells us that

$$\frac{Q_{\cdot}}{Q_i}\sigma^2 \longrightarrow \sigma_{A_i}^2, \quad \text{as} \quad \min(c_i) \to \infty.$$

Using (4.2.9), it is clear that $Q_{\cdot}\left(\mathbf{HA}\right)^{'}\left(\mathbf{HV}_A\mathbf{H}^{'}\right)^{-1}\left(\mathbf{HA}\right)$ is approximately equal to

$$Q_{\cdot}\left(\sum_{i=1}^{r}\frac{A_i^2}{Q_{\cdot}\sigma^2/Q_i}\right) - \frac{Q_{\cdot}}{\left(\sum_{i=1}^{r}\frac{Q_i}{Q_{\cdot}\sigma^2}\right)}\left(\sum_{i=1}^{r}\frac{A_i}{Q_{\cdot}\sigma^2/Q_i}\right)^2 = \frac{1}{\sigma^2}\sum_{i=1}^{r}Q_i\left(A_i - \bar{A}\right)^2.$$

The asymptotical equivalence between $Q_{\cdot}\left(\mathbf{HA}\right)^{'}\left(\mathbf{HV}_A\mathbf{H}^{'}\right)^{-1}\left(\mathbf{HA}\right)/(r-1)$ and F_C^{α} follows similarly.

Chapter 5

Testing for the Covariate Effect in the Fully Nonparametric ANCOVA

Consider the fully nonparametric ANCOVA model with the observed pairs $(X_{ij}, Y_{ij}),\ i = 1, \cdots, k,\ j = 1, \cdots, n_i$, as introduced in the beginning of this thesis (Section 1.2 in particular), which specifies only that

$$Y_{ij}|X_{ij} = x \sim F_{ix}(y) = M(y) + A_i(y) + D_i(y;x),\ i = 1, \cdots, k,$$

where the decomposition of F_{ix} is defined in (1.2.7). In the present chapter, we develop a test procedure for testing the null hypothesis of no covariate effect in the fully nonparametric fashion:

$$H_0(D):\ D_i(y;x) = 0 \quad \Leftrightarrow \quad F_{ix}(y) = F_i^0(y), \quad \text{for all } i, \text{ all } x \text{ and all } y. \qquad (5.0.1)$$

The basic idea used to construct the test procedure is to think of the continuous covariate variable as a factor with many levels and utilize suitable test statistics from the heteroscedastic unbalanced two-fold nested model. As outlined in Section 1.3, there are two possible approaches: *non-overlapping windows approach* and *overlapping windows approach*. The former one applies directly the asymptotic results from Chapter 2 and has been demonstrated on a real data set in Section 2.4.2. On the other hand, the technical derivation of the *overlapping windows approach* involves some complications resulting from the augmented dependence in our design. New asymptotic tools based on the projection principle are hence developed in this chapter to accommodate those challenges.

The rest of this chapter is organized as follows. Section 5.1 introduces how to implement the overlapping windows approach and the corresponding test statistic. In Section 5.2 we present the asymptotic techniques and the main theoretical results, while in Section 5.3 we summarize numerical results from several simulation studies. Three real data sets are analyzes in Section 5.4: the Low Birth Weight data, the Ethanol data, and the Acid Rain data from the National Atmospheric Deposition Program (NADP). Finally, we provide proofs of the main theorems and some technical details in Section 5.5.

5.1 The Test Statistic

To implement the overlapping windows approach, we first enumerate the observed pairs $(X_{ij}, Y_{ij}),\ i = 1, \cdots, k;\ j = 1, \cdots, n_i$, such that $X_{i1} < X_{i2} < \cdots < X_{in_i}$ (i.e. assuming no ties) for each i. Thus, each ordered covariate value corresponds to a level of the nested factor in the artificial two-fold nested design. The observations at each such level (i, X_{ir}) are augmented by including in it the responses corresponding to the w covariate values X_{ij} that are nearest to X_{ir} in the sense that

$$|\hat{G}_i(X_{ij}) - \hat{G}_i(X_{ir})| \leq \frac{w-1}{2n_i}, \quad \text{where } \hat{G}_i(x) = \frac{1}{n_i}\sum_{j=1}^{n_i} I(X_{ij} \leq x).$$

Note that $\hat{G}_i(x)$ is the empirical distribution function of the covariate in group i. For simplicity, we only consider w to be odd in this chapter.

Let Z_{irt} denote the t^{th} observation at level (i, X_{ir}) of subclass factor group i in this artificial two-fold nested design. In particular, letting

$$W_{ir} = \left\{1 \leq j \leq n_i : \ |\hat{G}_i(X_{ij}) - \hat{G}_i(X_{ir})| \leq \frac{w-1}{2n_i}\right\},$$

we have

$$Z_{irt} = Y_{ij} \quad \text{iff} \qquad \sum_{l=1}^{n_i} I(X_{il} \le x_{ij};\ l \in W_{ir}) = t, \tag{5.1.1}$$
$$i = 1, \cdots, k;\ r = 1, \cdots, n_i;\ t = 1, \cdots, w_{ir}.$$

Note that, for each $i = 1, \ldots, k$, the number of observations in levels (i, X_{ir}), $r = 1, \ldots, (w-1)/2$ and $r = n_i - (w+1)/2 + 1, \ldots, n_i$ is less than w. More specifically,

$$\begin{cases} w_{ir} = \frac{w-1}{2} + r, & if \;\; 1 \le r \le \frac{w-1}{2}; \\ w_{ir} = w, & if \;\; \frac{w-1}{2} + 1 \le r \le n_i - \frac{w-1}{2}; \\ w_{ir} = \frac{w-1}{2} + 1 + n_i - r, & if \;\; n_i - \frac{w-1}{2} + 1 \le r \le n_i. \end{cases} \tag{5.1.2}$$

It can be shown that

$$\begin{aligned} w_{i\cdot} &= \sum_{r=1}^{n_i} w_{ir} \\ &= \left(\sum_{r=1}^{\frac{w-1}{2}} \frac{w-1}{2} + r \right) + \left(\sum_{r=\frac{w-1}{2}+1}^{n_i - \frac{w-1}{2}} w \right) + \left(\sum_{r=n_i-\frac{w-1}{2}+1} n_i \frac{w-1}{2} + 1 + n_i - r \right) \\ &= w(n_i - w + 1) + 2\left[\frac{w-1}{2} \cdot \frac{w-1}{2} + \frac{1}{2} \cdot \frac{w-1}{2} \cdot \frac{w+1}{2} \right] \\ &= w(n_i - w + 1) + \frac{(w-1)(3w-1)}{4} \\ \Rightarrow \quad w_{i\cdot} &= n_i w - \frac{w^2-1}{4} \\ \Rightarrow \quad w_{\cdot\cdot} &= \sum_{i=1}^{k} w_{i\cdot} = Nw - \frac{k(w^2-1)}{4}, \end{aligned}$$

where $N = \sum_{i=1}^{k} n_i$.

Consider a heteroscedastic unbalanced two-fold nested design with k classes/groups and n_i sub-classes nested in class i, $i = 1, \cdots, k$. Let $V_{ir1}, \cdots, V_{irw_{ir}}$ denote the w_{ir}

observations of sub-class r in class i, $i = i, \cdots, k;\ r = 1, \cdots, n_i$. Define

$$\begin{aligned} F_D &= \frac{MST_D}{MSE_D}, \quad \text{where} \quad MST_D = \frac{1}{N-k} \sum_{i=1}^{k} \sum_{r=1}^{n_i} w_{ir} \left(\bar{V}_{ir\cdot} - \bar{V}_{i\cdot\cdot} \right)^2; \\ MSE_D &= \frac{1}{N-k} \sum_{i=1}^{k} \sum_{r=1}^{n_i} \left(1 - \frac{w_{ir}}{w_{i\cdot}} \right) \frac{1}{w_{ir}-1} \sum_{t=1}^{w_{ir}} \left(V_{irt} - \bar{V}_{ir\cdot} \right)^2, \end{aligned} \tag{5.1.3}$$

where

$$\bar{V}_{ir\cdot} = \frac{1}{w_{ir}} \sum_{t=1}^{w_{ir}} V_{irt}, \quad \bar{V}_{i\cdot\cdot} = \frac{1}{w_{i\cdot}} \sum_{r=1}^{n_i} w_{ir} \bar{V}_{ir\cdot}.$$

We replace V_{irt} in (5.1.3) with $Z_{irt},\ i = 1, \cdots, k;\ r = 1, \cdots, n_i;\ t = 1, \cdots, w_{ir}$, and have

$$\begin{aligned} F_D &= \frac{MST_D}{MSE_D}, \quad \text{where} \quad MST_D = \frac{1}{N-k} \sum_{i=1}^{k} \sum_{r=1}^{n_i} w_{ir} \left(\bar{Z}_{ir\cdot} - \bar{Z}_{i\cdot\cdot} \right)^2, \\ MSE_D &= \frac{1}{N-k} \sum_{i=1}^{k} \sum_{r=1}^{n_i} \left(1 - \frac{w_{ir}}{w_{i\cdot}} \right) \frac{1}{w_{ir}-1} \sum_{t=1}^{w_{ir}} \left(Z_{irt} - \bar{Z}_{ir\cdot} \right)^2, \\ \bar{Z}_{ir\cdot} &= \frac{1}{w_{ir}} \sum_{t=1}^{w_{ir}} Z_{irt}, \quad \bar{Z}_{i\cdot\cdot} = \frac{1}{w_{i\cdot}} \sum_{r=1}^{n_i} w_{ir} \bar{Z}_{ir\cdot}. \end{aligned} \tag{5.1.4}$$

Define the test statistic for the null hypothesis of no covariate effect (5.0.1) as

$$T_D = MST_D - MSE_D. \tag{5.1.5}$$

In this chapter we study the asymptotic distributions of T_D under both the null and the local alternative hypotheses. Although we do not use F_D in (5.1.4) as our test statistic here, one can easily prove that the MSE_D converges to some constant in probability as $\min(n_i) \to \infty$ under weak regularity conditions, and apply the Slutsky's theorem to obtain the asymptotic distributions of $(F_D - 1)$ if desired.

5.2 Main Results

5.2.1 Some notations and Assumptions

Let $\mathbf{Z}_{ir} = (Z_{ir_1}, Z_{ir2}, \cdots, Z_{irw_{ir}})'$ be $w_{ir} \times 1$ vectors, $\mathbf{Z}_i = (\mathbf{Z}_{i1}^{'}, \mathbf{Z}_{i2}^{'}, \cdots, \mathbf{Z}_{in_i}^{'})'$ be $w_{i\cdot} \times 1$ vectors, and $\mathbf{Z} = (\mathbf{Z}_1^{'}, \mathbf{Z}_2^{'}, \cdots, \mathbf{Z}_k^{'})'$ be a $w_{\cdot\cdot} \times 1$ vector containing all observations in the augmented design. In addition, let $\mathbf{1}_d$ denote the $d \times 1$ vector of 1's, $\mathbf{J}_d = \mathbf{1}_d \mathbf{1}_d^{'}$, and $\mathbf{I}_d$ be the d-dimensional identity matrix. $\oplus$ denotes the operation of Kronecker sum.

It can be shown that $MSE_D = \mathbf{Z}'\mathbf{T}_1\mathbf{Z}$ (as in Section 5.5.1) with

$$\mathbf{T}_1 = \frac{1}{N-k} \bigoplus_{i=1}^{k} \bigoplus_{r=1}^{n_i} \left(1 - \frac{w_{ir}}{w_{i\cdot}}\right) \frac{1}{w_{ir}-1} \left[\mathbf{I}_{w_{ir}} - \frac{1}{w_{ir}} \mathbf{J}_{w_{ir}}\right], \tag{5.2.1}$$

while $MST_D = \mathbf{Z}'\mathbf{T}_2\mathbf{Z}$ with

$$\mathbf{T}_2 = \frac{1}{N-k} \bigoplus_{i=1}^{k} \left[\bigoplus_{r=1}^{n_i} \frac{1}{w_{ir}} \mathbf{J}_{w_{ir}} - \frac{1}{w_{i\cdot}} \mathbf{J}_{w_{i\cdot}}\right]. \tag{5.2.2}$$

Then, the test statistic $T_D = MST_D - MSE_D = \mathbf{Z}'(\mathbf{T}_2 - \mathbf{T}_1)\mathbf{Z} = \mathbf{Z}'\mathbf{A}\mathbf{Z}$, where

$$\begin{aligned}
\mathbf{A} = \mathbf{T}_2 - \mathbf{T}_1 &= \frac{1}{N-k} \bigoplus_{i=1}^{k} \bigoplus_{r=1}^{n_i} \left[\frac{1}{w_{ir}} + \left(1 - \frac{w_{ir}}{w_{i\cdot}}\right) \frac{1}{w_{ir}-1} \cdot \frac{1}{w_{ir}}\right] \mathbf{J}_{w_{ir}} \\
&- \frac{1}{N-k} \bigoplus_{i=1}^{k} \frac{1}{w_{i\cdot}} \mathbf{J}_{w_{i\cdot}} - \frac{1}{N-k} \bigoplus_{i=1}^{k} \bigoplus_{r=1}^{n_i} \left(1 - \frac{w_{ir}}{w_{i\cdot}}\right) \frac{1}{w_{ir}-1} \mathbf{I}_{w_{ir}} \\
&= diag\{\mathbf{A}_1, \ \cdots, \mathbf{A}_k\}, \\
\text{with} \quad \mathbf{A}_i &= \bigoplus_{r=1}^{n_i} (\alpha_{ir}\mathbf{J}_{w_{ir}} - \gamma_{ir}\mathbf{I}_{w_{ir}}) - \beta_i \mathbf{J}_{w_{i\cdot}} \\
&= \begin{pmatrix} \mathbf{B}_{i1} & -\beta_i \mathbf{1}_{w_{i1}} \mathbf{1}_{w_{i2}}^{'} & \cdots & -\beta_i \mathbf{1}_{w_{i1}} \mathbf{1}_{w_{in_i}}^{'} \\ -\beta_i \mathbf{1}_{w_{i2}} \mathbf{1}_{w_{i1}}^{'} & \mathbf{B}_{i2} & & \vdots \\ \vdots & & \ddots & \vdots \\ -\beta_i \mathbf{1}_{w_{in_i}} \mathbf{1}_{w_{i1}}^{'} & \cdots & \cdots & \mathbf{B}_{in_i} \end{pmatrix}_{w_{i\cdot} \times w_{i\cdot}} \\
\alpha_{ir} &= \frac{1}{(N-k)(w_{ir}-1)} \left(1 - \frac{1}{w_{i\cdot}}\right); \quad \beta_i = \frac{1}{(N-k) \cdot w_{i\cdot}};
\end{aligned}$$

$$\gamma_{ir} = \frac{1}{(N-k)(w_{ir}-1)}\left(1-\frac{w_{ir}}{w_{i\cdot}}\right);$$

$$\mathbf{B}_{ir} = (b_{ir,gh})_{w_{ir}\times w_{ir}}, \quad b_{ir,gh} = \begin{cases} 0, & \text{if } g = h; \\ \alpha_{ir} - \beta_i = \gamma_{ir}, & \text{if } g \neq h. \end{cases}$$

The following notations will be frequently used in this chapter:

$$\mathbf{Z}^* = \mathbf{Z} - E(\mathbf{Z}|\mathbf{X}); \quad Y_{ij}^* = Y_{ij} - E(Y_{ij}|X_{ij}); \quad \sigma_i^2(x) = Var(Y_{ij} \mid X_{ij} = x).$$

To study the asymptotic distributions of T_D, we need to further define two quadratic forms, $\mathbf{Z}^{*\prime}\mathbf{A}_D\mathbf{Z}^*$ and $\mathbf{Z}^{*\prime}\mathbf{A}_D^*\mathbf{Z}^*$, where

$$\mathbf{A}_D = \bigoplus_{i=1}^{k}\bigoplus_{r=1}^{n_i}\mathbf{B}_{ir} \quad \text{and} \quad \mathbf{A}_D^* = \bigoplus_{i=1}^{k}\bigoplus_{r=1}^{n_i}\mathbf{B}_{ir}^*, \tag{5.2.3}$$

with the $\mathbf{B}_{ir}^*$ defined as $\mathbf{B}_{ir}$, with γ_{ir} replaced by

$$\kappa_i = \frac{1}{(N-k)(w-1)}\left(1-\frac{1}{n_i}\right). \tag{5.2.4}$$

These two quadratic forms will serve as approximations of the test statistic $T_D = \mathbf{Z}'\mathbf{A}\mathbf{Z}$ in our asymptotic derivations. $\mathbf{Z}^{*\prime}\mathbf{A}_D\mathbf{Z}^*$ is obtained by an application of the projection method, as introduced in Akritas and Papadatos (2004), while $\mathbf{Z}^{*\prime}\mathbf{A}_D^*\mathbf{Z}^*$ is designed to simplify our calculations in the proofs. (The calculation of the projection matrix $\mathbf{A}_D$ is quite tedious and hence omitted here; see Section 5.5.1 for details.)

The asymptotic distributions of T_D are derived under the following assumptions:

Assumption A1. (1) $\forall i,\ \exists\lambda_i \in (0,1)$, such that $n_i/N \to \lambda_i$, as $\underline{n} = \min(n_i) \to \infty$;
(2) $w \geq 2$ and $\underline{n}^{-1+a}w \to 0$, for all $0 < a < 1$;

Assumption A2. $E(Y_{ij}^4 \mid X_{ij} = x)$ are uniformly bounded in $i,\ x$;

Assumption A3. The covariate X is a continuous random variable with bounded support S_i, c.d.f G_i, and p.d.f g_i, $i = 1, \cdots, k$. The density g_i is bounded away from 0 on S_i.

5.2.2 Asymptotic Distribution under the Null

To obtain the asymptotic distribution of $N^{1/2}w^{-1/2}T_D$ under $H_0(D)$, we first show in Lemma 5.2.1 that observations are (conditionally) centered under the null hypothesis. In addition, Lemma 5.2.2 shows that this centered quadratic form can be approximated by another (centered) quadratic form based on the projection matrix $\mathbf{A}_D$, while Lemma 5.2.3 tells us that this projection quadratic form can be further approximated by a simpler form based on $\mathbf{A}_D^*$ as defined in (5.2.3). The asymptotic variance of this simpler quadratic form is then given in Lemma 5.2.4, while Theorem 5.2.1 provides the asymptotic distribution of the proposed test statistic under $H_0(D)$ (see Section 5.5.2 for all corresponding proofs).

Lemma 5.2.1. *Under $H_0(D)$,*

$$\mathbf{Z}'\mathbf{A}\mathbf{Z} = \mathbf{Z}^{*\prime}\mathbf{A}\mathbf{Z}^*.$$

Lemma 5.2.2. *Under $H_0(D)$ and Assumptions A1–A2,*

$$N^{1/2}w^{-1/2}\left[\mathbf{Z}^{*\prime}\mathbf{A}\mathbf{Z}^* - \mathbf{Z}^{*\prime}\mathbf{A}_D\mathbf{Z}^*\right] \xrightarrow{P} 0, \quad \text{as } \min(n_i) \to \infty.$$

Lemma 5.2.3. *Under $H_0(D)$ and Assumptions A1–A2,*

$$N^{1/2}w^{-1/2}\left[\mathbf{Z}^{*\prime}\mathbf{A}_D\mathbf{Z}^* - \mathbf{Z}^{*\prime}\mathbf{A}_D^*\mathbf{Z}^*\right] \xrightarrow{P} 0, \quad \text{as } \min(n_i) \to \infty.$$

Lemma 5.2.4. *Under $H_0(D)$ and Assumption A1,*
provided that $\sigma_i^2(x) \overset{H_0(D)}{=} \sigma_i^2 > 0, \ \forall x, \ \left(\sigma_i^4 = (\sigma_i^2)^2\right)$

$$Var(N^{1/2}w^{-1/2}\mathbf{Z}^{*\prime}\mathbf{A}_D^*\mathbf{Z}^*) = \frac{2(2w-1)}{3(w-1)}\sum_{i=1}^{k}\lambda_i\sigma_i^4 + o(1), \quad \text{as } \min(n_i) \to \infty.$$

Theorem 5.2.1. [Asymptotic Null Distribution]
Under $H_0(D)$ as defined in (1.2.9) and Assumptions A1–A2,
provided that $\sigma_i^2(x) \overset{H_0(D)}{=} \sigma_i^2 > 0, \ \forall x, \ \left(\sigma_i^4 = (\sigma_i^2)^2\right)$

(1) *if w is fixed,*

$$N^{1/2}w^{-1/2}T_D \xrightarrow{d} N\left(0,\ \frac{2(2w-1)}{3(w-1)}\sum_{i=1}^{k}\lambda_i\sigma_i^4\right),\ \ as\ \min(n_i)\to\infty;$$

(2) *if* $w\to\infty$ *as* $\min(n_i)\to\infty$ *(while Assumption A1 holds),*

$$N^{1/2}w^{-1/2}T_D \xrightarrow{d} N\left(0,\ \frac{4}{3}\sum_{i=1}^{k}\lambda_i\sigma_i^4\right),\ \ as\ \min(n_i)\to\infty.$$

5.2.3 Asymptotic Distribution under Local Alternatives

The asymptotic power is investigated by considering the local alternative sequence:

$$H_a(D):\ \ D_i(y;x)=(n_i\cdot w)^{-1/4}C_i(y;x), \tag{5.2.5}$$

where $C_i(y;x),\ i=1,\cdots,k$, are chosen so that $\int ydC_i(y;x)$ are uniformly bounded and uniformly Lipschitz continuous for all i and x. Note that (5.2.5) implies that the alternatives need to approach the null at the rate of $(\inf_i n_i w)^{-1/4}$ to ensure nontrivial power. Define

$$\theta_D=\sum_{i=1}^{k}\sqrt{\lambda_i}\left\{\int\left[\int ydC_i(y;x)\right]^2 dG_i(x)-\left[\int\int ydC_i(y;x)dG_i(x)\right]^2\right\}. \tag{5.2.6}$$

Recall that

$$\begin{aligned} N^{1/2}w^{-1/2}T_D &= N^{1/2}w^{-1/2}(\mathbf{Z}'\mathbf{A}\mathbf{Z}-\mathbf{Z}^{*\prime}\mathbf{A}\mathbf{Z}^*)+N^{1/2}w^{-1/2}(\mathbf{Z}^{*\prime}\mathbf{A}\mathbf{Z}^*-\mathbf{Z}^{*\prime}\mathbf{A}_D\mathbf{Z}^*) \\ &\quad +N^{1/2}w^{-1/2}(\mathbf{Z}^{*\prime}\mathbf{A}_D\mathbf{Z}^*-\mathbf{Z}^{*\prime}\mathbf{A}_D^*\mathbf{Z}^*)+N^{1/2}w^{-1/2}\mathbf{Z}^{*\prime}\mathbf{A}_D^*\mathbf{Z}^*. \end{aligned}$$

In this section we first show that, under the local alternative sequence as defined in (5.2.5), the first term converges in probability to θ_D as $\min(n_i)\to\infty$ in Lemma 5.2.5, while the second and third terms remain negligible as shown in Lemmas 5.2.6 and 5.2.7. Next, the asymptotic variance of the simpler quadratic form is given in Lemma 5.2.8. Finally, Theorem 5.2.2 provides the asymptotic distribution of the proposed test statistic under $H_a(D)$ (see Section 5.5.2 for proofs).

Lemma 5.2.5. *Under $H_a(D)$ and Assumptions A1–A3, as* $\min(n_i) \to \infty$,

$$N^{1/2}w^{-1/2}\left[\mathbf{Z}'\mathbf{A}\mathbf{Z}' - \mathbf{Z}^{*\prime}\mathbf{A}\mathbf{Z}^*\right] \xrightarrow{P} \theta_D,$$

where θ_D is as defined in (5.2.6).

Lemma 5.2.6. *Under $H_a(D)$ and Assumptions A1–A3,*

$$N^{1/2}w^{-1/2}\left[\mathbf{Z}^{*\prime}\mathbf{A}\mathbf{Z}^* - \mathbf{Z}^{*\prime}\mathbf{A}_D\mathbf{Z}^*\right] \xrightarrow{P} 0, \quad \text{as } \min(n_i) \to \infty.$$

Lemma 5.2.7. *Under $H_a(D)$ and Assumptions A1–A3,*

$$N^{1/2}w^{-1/2}\left[\mathbf{Z}^{*\prime}\mathbf{A}_D\mathbf{Z}^* - \mathbf{Z}^{*\prime}\mathbf{A}_D^*\mathbf{Z}^*\right] \xrightarrow{P} 0, \quad \text{as } \min(n_i) \to \infty.$$

Lemma 5.2.8. *Under $H_a(D)$ and Assumptions A1–A3, provided that $\sigma_i^2(x)$ is positive and uniformly Lipschitz continuous in x and i,* $\left(\sigma_i^4(x) = [\sigma_i^2(x)]^2\right)$

$$Var(N^{1/2}w^{-1/2}\mathbf{Z}^{*\prime}\mathbf{A}_D^*\mathbf{Z}^*) = \frac{2(2w-1)}{3(w-1)}\sum_{i=1}^{k}\lambda_i E[\sigma_i^4(X)] + o(1), \quad \text{as } \min(n_i) \to \infty.$$

Theorem 5.2.2. [Asymptotic Distribution under Local Alternatives]
Under $H_a(D)$ as defined in (5.2.5) and Assumptions A1–A3, provided that $\sigma_i^2(x)$ is positive and uniformly Lipschitz continuous in x and i,

(1) *if w is fixed, as* $\min(n_i) \to \infty$,

$$N^{1/2}w^{-1/2}T_D \xrightarrow{d} N\left(\theta_D,\ \frac{2(2w-1)}{3(w-1)}\sum_{i=1}^{k}\lambda_i E[\sigma_i^4(X)]\right);$$

(2) *if $w \to \infty$ as $\min(n_i) \to \infty$ (while Assumption A1 holds),*

$$N^{1/2}w^{-1/2}T_D \xrightarrow{d} N\left(\theta_D,\ \frac{4}{3}\sum_{i=1}^{k}\lambda_i E[\sigma_i^4(X)]\right),$$

where θ_D is as defined in (5.2.6).

5.3 Simulation Studies

In this section, simulations are used to compare the achieved sizes and/or powers of several test procedures for testing the null hypothesis of no (simple) covariate effect. Let *NP(w)* denote the proposed nonparametric test using overlapping windows of size w. By Theorem 5.2.1, *NP(w)* rejects at level α if

$$\left(\frac{N}{w}\right)^{1/2} T_D > \sqrt{\frac{2(2w-1)}{3(w-1)} \sum_{i=1}^{k} \hat{\lambda}_i \hat{\sigma}_i^4}\, Z_\alpha,$$

where Z_α is the $(1-\alpha)$100th percentile of the standard normal distribution and $\hat{\lambda}_i$ is the empirical versions of λ_i, namely $\hat{\lambda}_i = n_i/N$. The $\hat{\sigma}_i^4$ can be any consistent estimator for σ_i^4, and the one we use in the simulations studies shown in this section is

$$\hat{\sigma}_i^4 \;=\; \frac{1}{n_i}\sum_{r=1}^{n_i} \hat{\sigma}_i^4\left(X_{ir}\right),$$

where $\hat{\sigma}_i^4(X_{ir})$ is the U-statistic of the observations in window W_{ir} (i.e. the window centered around X_{ir}) with the kernel $h(z_1, z_2, z_3, z_4) = (z_1 - z_2)^2 \times (z_3 - z_4)^2/4$. Another possible simple estimator of σ_i^4 is

$$\hat{\sigma}_i^{4*} \;=\; \frac{1}{4(n_i-3)} \sum_{j=1}^{n_i-3} \left(Y_{i,j} - Y_{i,j+1}\right)^2 \times \left(Y_{i,j+2} - Y_{i,j+3}\right)^2, \qquad (5.3.1)$$

which can be thought of as a modified Rice's estimator (1984). From simulations (not shown here), the performance of *NP(w)* using $\hat{\sigma}_i^4$ seems slightly better than the one using $\hat{\sigma}_i^{4*}$, so we only include the results using the former. In addition, for simplicity, we only use windows of size 5 to compute the estimator $\hat{\sigma}_i^4$ in our simulations, regardless of the window size used in computing the test statistic T_D.

All simulations are demonstrated using two groups ($k = 2$) and one continuous covariate whose values are randomly generated from the standard uniform distribution $U(0,1)$. The proposed *NP(w)* test is compared with two alternatives: the classical F test (denoted as *CF*) and the rank-based test of McKean and Sheather (1991) (denoted as *Drop*). Both of these tests can be directly implemented in R (2009). For more details about the Drop test in R, see Terpstra and McKean (2005).

5.3.1 Simulations under Heteroscedasticity and Non-Normality

We first compare the achieved α-levels of these three test procedures when homoscedasticity does not hold. The data for our two groups are generated from

$$Y_{1j} = \theta \cdot e_{1j} \qquad \text{and} \qquad Y_{2j} = 2 + e_{2j},$$

where $e_{ij} \sim N(0,1)$, $i = 1,\ 2$, for $\theta = 1,\ 3,\ 5,\ 10$. Note that, when $\theta = 1$, the homoscedastic assumption in fact holds. The achieved α-levels, based on $10,000$ simulation runs, at nominal level of .05 are shown in Table 5.1.

Table 5.1. Achieved α-levels over $10,000$ simulation runs under homoscedasticity ($\theta = 1$) and under heteroscedasticity ($\theta \neq 1$) at nominal $\alpha = 0.05$.

$(n_1,\ n_2)$	θ	CF	Drop	NP(5)	NP(7)	NP(9)
(20, 30)	1	.0508	.0515	.0457	.0324	.0243
	3	.1085	.1408	.0664	.0476	.0345
	5	.1237	.2088	.0702	.0506	.0367
	10	.1308	.3212	.0730	.0544	.0407
(40, 60)	1	.0462	.0464	.0482	.0402	.0340
	3	.1007	.1358	.0692	.0546	.0481
	5	.1159	.2158	.0681	.0570	.0480
	10	.1258	.3352	.0728	.0609	.0526
(150, 200)	1	.0519	.0498	.0508	.0453	.0404
	3	.0901	.1265	.0645	.0605	.0563
	5	.1019	.2019	.0616	.0564	.0532
	10	.1083	.3219	.0639	.0610	.0572

Table 5.1 makes it clear that both the *CF* test and the *Drop* test are too liberal under heteroscedasticity, although the *CF* test appears to become less so when the number of observations, n_1 and n_2, increase. On the other hand, the proposed procedure *NP(w)* performs reasonably well both under homoscedasticity and under heteroscedasticity, though it is somewhat conservative in the case of homoscedasticity, especially when the number of observations are small.

In addition, we investigate the effect of normality by considering data generated from some non-normal distributions. More specifically, we generate Y_{ij} from the exponential distribution with rate equal to either .2 or 1 (denoted as *exp(0.2)* and *exp(1.0)*), and from the the log-normal distribution whose logarithm has mean 0 and standard deviation equal to either .2 or 1 (denoted as *lnorm(0.2)* and *lnorm(1.0)*) . The achieved α-levels, based on $10,000$ simulation runs, at nominal level of .05 are shown in Table 5.2.

Table 5.2. Achieved α-levels over $10,000$ simulation runs under non-normality at nominal $\alpha = 0.05$.

$(n_1,\ n_2)$	Y_{1j}	Y_{2j}	CF	Drop	NP(5)	NP(7)	NP(9)
(20, 30)	exp(0.2)	exp(1.0)	.1186	.2138	.0557	.0399	.0275
	exp(1.0)	exp(1.0)	.0502	.0461	.0445	.0318	.0227
	lnorm(0.2)	lnorm(1.0)	.0528	.0922	.0516	.0407	.0302
	lnorm(1.0)	lnorm(1.0)	.0541	.0451	.0358	.0232	.0168
(40, 60)	exp(0.2)	exp(1.0)	.1137	.2247	.0550	.0446	.0362
	exp(1.0)	exp(1.0)	.0495	.0445	.0421	.0325	.0285
	lnorm(0.2)	lnorm(1.0)	.0533	.0987	.0469	.0376	.0344
	lnorm(1.0)	lnorm(1.0)	.0504	.0449	.0356	.0271	.0216
(150, 200)	exp(0.2)	exp(1.0)	.0978	.2069	.0553	.0483	.0466
	exp(1.0)	exp(1.0)	.0518	.0484	.0434	.0381	.0360
	lnorm(0.2)	lnorm(1.0)	.0637	.1131	.0463	.0410	.0388
	lnorm(1.0)	lnorm(1.0)	.0517	.0456	.0362	.0309	.0293

Although it is well-known that the classical F test is robust to the departure from the normality assumption when the number of observation goes to infinity (see Arnold (1980)), Table 5.2 reveals something very interesting. When the underlying distributions in the two groups are the same, even though non-normal, the classical *CF* test and the rank-based *Drop* test do perform well. However, if the observations in the two groups are generated from different distributions, like *exp(0.2) vs exp(1.0)* or *lnorm(0.2) vs lnorm(1.0)*, these two test procedures can become much too liberal, as seen in Table 5.2. This seems to suggest that the insensitivity of the *CF* test and the *Drop* test to the normality assumption only holds within groups, not between groups. As for the proposed *NP(w)* test, its achieved α-levels are all reasonable under various data-generating mechanisms, though a comparative conservativeness is still observed with smaller w.

5.3.2 Simulations under Linearity and Non-Linearity

Due to the sensitivity of the *CF* and *Drop* procedures to departures from certain model assumptions (see previous subsection), we only use homoscedastic normal errors in the present subsection.

When the assumptions of the classical ANCOVA model all hold (i.e. under homoscedasticity, normality and linearity), the *CF* test unsurprisingly outperforms the other test procedures, as shown in Table 5.3, where the data are generated from

$$Y_{1j} = .2e_{1j} \qquad \text{and} \qquad Y_{2j} = \theta \cdot X_{2j} + .2e_{2j}, \tag{5.3.2}$$

with $e_{ij} \sim N(0,1)$, $i = 1,\ 2$, for $\theta = 0,\ 0.1,\ 0.3,\ 0.5$. Note that when $\theta = 0$, data in the two groups are generated under the null hypothesis.

Table 5.3. Powers over 10,000 simulation runs under linear alternatives at nominal $\alpha = 0.05$.

$(n_1,\ n_2)$	θ	CF	Drop	NP(5)	NP(7)	NP(9)	HOM(5)
(20, 30)	0.0 (level)	.0521	.0502	.0464	.0347	.0260	.0964
	0.1 (power)	.0952	.0923	.0748	.0611	.0497	.1323
	0.3 (power)	.5197	.4775	.3187	.3198	.3109	.3849
	0.5 (power)	.9192	.8939	.7473	.7697	.7724	.7877
(40, 60)	0.0 (level)	.0525	.0519	.0503	.0409	.0353	.0854
	0.1 (power)	.1470	.1370	.0768	.0739	.0705	.1171
	0.3 (power)	.8342	.8083	.4821	.5308	.5609	.4850
	0.5 (power)	.9988	.9979	.9489	.9674	.9757	.9419
(150, 200)	0.0 (level)	.0513	.0501	.0521	.0479	.0459	.0670
	0.1 (power)	.4260	.4060	.1124	.1210	.1284	.1228
	0.3 (power)	.9999	.9997	.8757	.9279	.9527	.8097
	0.5 (power)	1.000	1.000	1.000	1.000	1.000	1.000

To investigate the sensitivity of these test procedures to departures from the linearity assumption, we further generate data from two non-linear alternatives:

$$Y_{1j} = .2e_{1j} \qquad \text{versus} \qquad Y_{2j} = \theta \cdot (X_{2j}^2 - X_{2j}) + .2e_{2j}; \tag{5.3.3}$$

$$Y_{1j} = .2e_{1j} \qquad \text{versus} \qquad Y_{2j} = \theta \cdot cos(2\ \pi\ X_{2j}) + .2e_{2j}; \tag{5.3.4}$$

where $e_{ij} \sim N(0,1)$, $i = 1,\ 2$, and θ are as specified in the tables. As in the previous simulation study, $\theta = 0$ gives us the achieved α-levels. The achieved α-levels and powers, based on $10,000$ simulation runs, at nominal level of .05 are shown in Table 5.4 and Table 5.5, respectively.

Table 5.4. Powers over $10,000$ simulation runs under quadratic alternatives as specified in (5.3.3) at nominal $\alpha = 0.05$.

$(n_1,\ n_2)$	θ	CF	Drop	NP(5)	NP(7)	NP(9)	HOM(5)
(20, 30)	0.0 (level)	.0471	.0475	.0467	.0351	.0234	.0980
	0.5 (power)	.0503	.0488	.0753	.0584	.0428	.1385
	1.0 (power)	.0550	.0536	.1951	.1690	.1350	.2626
	1.5 (power)	.0633	.0582	.4247	.3957	.3389	.4888
(40, 60)	0.0 (level)	.0508	.0506	.0527	.0439	.0364	.0837
	0.5 (power)	.0517	.0536	.1060	.0999	.0936	.1419
	1.0 (power)	.0564	.0528	.3387	.3610	.3671	.3693
	1.5 (power)	.0605	.0567	.7322	.7625	.7706	.7253
(150, 200)	0.0 (level)	.0506	.0490	.0541	.0495	.0459	.0738
	0.5 (power)	.0520	.0532	.1685	.1900	.2050	.1651
	1.0 (power)	.0572	.0581	.7188	.7920	.8337	.6323
	1.5 (power)	.0641	.0651	.9917	.9953	.9979	.9769

Table 5.5. Powers over $10,000$ simulation runs under cosine alternatives as specified in (5.3.4) at nominal $\alpha = 0.05$.

$(n_1,\ n_2)$	θ	CF	Drop	NP(5)	NP(7)	NP(9)	HOM(5)
(20, 30)	0.0 (level)	.0483	.0458	.0498	.0346	.0253	.1010
	0.1 (power)	.0514	.0488	.1969	.1797	.1459	.2639
	0.2 (power)	.0514	.0515	.7023	.6952	.6513	.7281
	0.3 (power)	.0558	.0613	.9700	.9693	.9588	.9643
(40, 60)	0.0 (level)	.0497	.0500	.0505	.0411	.0350	.0829
	0.1 (power)	.0495	.0501	.3160	.3422	.3544	.3357
	0.2 (power)	.0513	.0512	.9392	.9600	.9666	.9200
	0.3 (power)	.0550	.0592	.9999	.9999	.9999	.9995
(150, 200)	0.0 (level)	.0515	.0482	.0506	.0478	.0422	.0692
	0.1 (power)	.0506	.0519	.6572	.7382	.7935	.5693
	0.2 (power)	.0513	.0505	1.000	1.000	1.000	1.000
	0.3 (power)	.0564	.0628	1.000	1.000	1.000	1.000

From Table 5.4 and Table 5.5, one can easily notice the superiority of the proposed

NP(w) test under non-linear alternatives, regardless of the local window size w used. Tables 5.4 and 5.5 also spotlight the lack of power in detecting non-linear covariate effects when using the *CF* test and the *Drop* test; their achieved powers are very close to the specified nominal level .05.

Also note that we include the simulation results using the *non-overlapping windows approach*, which directly utilizes the asymptotic theorem from the two-fold nested model. Since the errors used are homoscedastic, only the results of the *HOM(5)* procedure are shown here, where the number 5 means that the non-overlapping windows used have size 5 (i.e. the size of each 'sub-class' is 5); see Section 2.4.2 for a description of the *non-overlapping windows approach*. Unsurprisingly, *HOM(5)* procedure performs too liberally in all simulations shown here, due to the fact that the corresponding numbers of 'sub-classes' are not large enough to make the asymptotic theorem work. For example, for the case with ($n_1 = 40$, $n_2 = 60$), the corresponding numbers of sub-classes are as small as $c_1 = 8$ and $c_2 = 12$. Even though the achieved α-levels of the *HOM* procedure do decrease to about .07 (at the nominal level .05) in cases with $n_1 = 150$ and $n_2 = 200$, Tables 5.3 – 5.5 show that its achieved powers are not as promising as those of the *NP* procedure, based on the *overlapping windows approach*. This indicates that the newly-proposed *overlapping windows approach* does outperform the naive *non-overlapping windows approach* in analysis of covariance when testing for the covariate effect.

5.4 Data Analyses: Three Empirical Studies

5.4.1 Example using Low Birth Weight Data

In 1986, the Baystate Medical Center in Springfield, Massachusetts, collected data from 189 females, 59 of which had low birth weight babies (weighing less than 2500 grams) while the other 130 of which had babies with normal birth weights. The main objective of this study was to identify influential factors which would result in low birth weights, and among all, two variables of interest were race (96 Whites, 26 Blacks, and 67 Others) and weight of the mother in pounds at her last menstrual period (LWT). See Hosmer

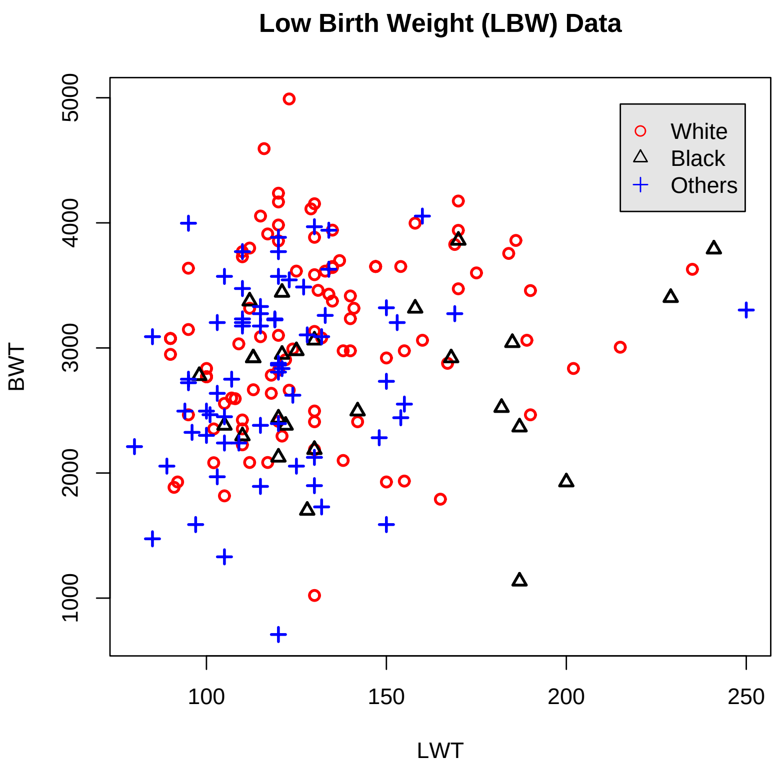

Figure 5.1. Scatterplot of Low Birth Weight Data.

and Lemeshow (2000) for more details on this data set. Here we would like to investigate whether LWT has a significant effect on babies' birth weights in grams (BWT) when the factor RACE is being considered. Figure 5.1 is the scatter plot of BWT versus LWT, where the circles represent the White group, the triangles represent the Black group, and the crosses represent the Others group.

Due to recording purposes, the LWT values were rounded to integers in the data, resulting in some ties on the covariate values in this study. To implement our methodology on this data set, we use the uniform random generator on (0.0001, 0.01) to add a small random quantity on the LWT values, and then sorted the observed pairs of (original LWT, BWT) for each of the three race groups separately, according to the modified LWT values. The average p-values of *NP(5)*, *NP(7)*, and *NP(9)*, over 100 repeat runs, for the covariate effect are .2572, .3271, and .3065, respectively, indicating an insignifi-

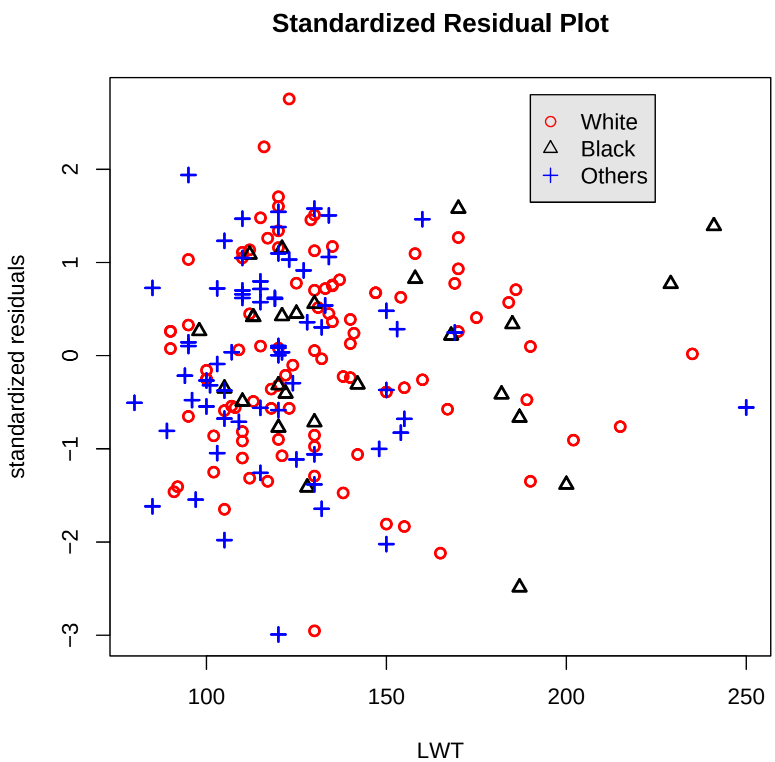

Figure 5.2. Low Birth Weight Data: Standardized Residual Plot from the classical ANCOVA model.

cant LWT effect on BWT. However, the *CF* test gives a p-value of .0578 and the *Drop* test gives a p-value of .0382, suggesting otherwise.

A closer examination of the data reveals a serious violation of the homoscedasticity assumption in this study, which can be seen from Figure 5.2, the standardized residual plot from the classical ANCOVA model. This is also suggested by Rice's estimators, as defined in (5.3.1), since $\hat{\sigma}^{4*}_{white} = 2.05\times10^{11}$, $\hat{\sigma}^{4*}_{black} = 1.14\times10^{11}$, and $\hat{\sigma}^{4*}_{others} = 1.18\times10^{11}$. As shown by simulations reported in Table 5.1, the *CF* test and the *Drop* test tend to perform more liberally than they should be when the model is not homoscedastic, so the small p-values they give in the study of low birth weight are in fact not surprising.

5.4.2 Example using Ethanol Data

As another example, we analyze a data set which has been studied previously by Kulasekera (1995). The data set consists of 88 observations from an experimental study in which Ethanol fuel was used in a single-cylinder engine. Three variables were recorded: the concentration of nitrogen oxides (NOx, in μg/J), the compression ratio (CR), and the equivalence ratio (ER), a measure of the richness of the air and fuel mixture. Since NOx is one of the major air pollutants, the main goal of the original study was to understand how the concentration of NOx depends on various settings of the compression ratio and the equivalence ratio. For more details, see also Cleveland (1993). Following Kulasekera (1995), we categorize the compression ratio as either *Low*: $CR < 10$ or *High*: $CR >= 10$, which is then utilized to divide the observed pairs (ER, NOx) into two groups. As a consequence, the *Low* group has 39 observations, while the *High* group has 49 observations. Figure 5.3 shows the relation between the equivalence ratio (ER) and the concentration of NOx, where the circles represent the *Low* group and the asterisks represent the *High* group. Local linear Gaussian kernel regression estimates are fitted for the two groups separately, with bandwidths selected by the direct plug-in methodology of Ruppert et al. (1995). The dashed line represents Group *Low*, while the solid line represents Group *High*.

Note that the local kernel regression curves show a high nonlinearity for both groups. Application of the two test procedures *CF* and *Drop* on this data set yields p-values of .6373 and .6285 respectively, for the hypothesis of no ER effect. On the other hand, the proposed *NP* test yields very small p-values (less than 10^{-12}) for a wide range of window sizes, suggesting that the effect of the equivalence ratio is in fact significant. Clearly, the nonlinearity of the ER effect causes the failure of the *CF* and *Drop* tests in detecting a significant covariate effect in this study. This result echoes our findings in the previous section 5.3.2.

5.4.3 Example using NADP Data

The third real-world application for our methodology can be found through the National Atmospheric Deposition Program (2009), which monitors geographical and temporal

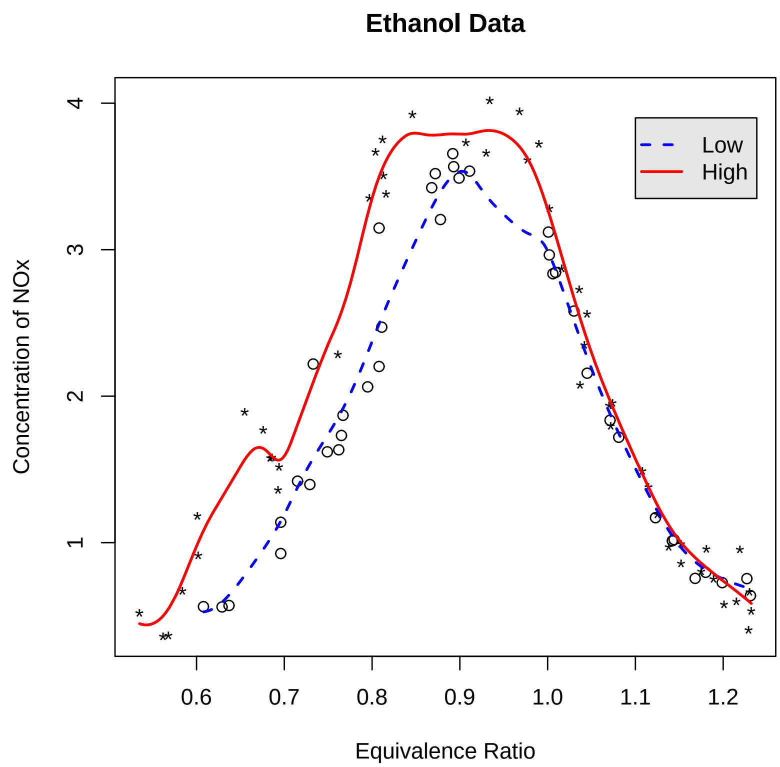

Figure 5.3. Scatterplot of Ethanol Data.

long-term trends on the chemistry of precipitation. Starting from only 22 stations in 1978, NAPD has grown as a nationwide network of over 250 sites at which precipitation samples are collected and analyzed in the Central Analytical Laboratory (CAL) weekly. For our data analysis, we chose to analyze the pH level (reported as the negative log of hydrogen ion concentration) of precipitation samples as measured in the CAL from the first week of January 2003 to the last week of January 2007. We consider comparing the data in two North Carolina towns, Lewiston and Coweeta, and are interested in the covariate effect of Time. The data, along with local linear kernel regression estimates, are shown in Figure 5.4. The circles and the dash line are for Lewiston data, while the asterisks and the solid line are for Coweeta data. Since a simple time series analysis does not indicate meaningful correlation over time (see Appendix B), it appears reasonable to implement our methodology in this study.

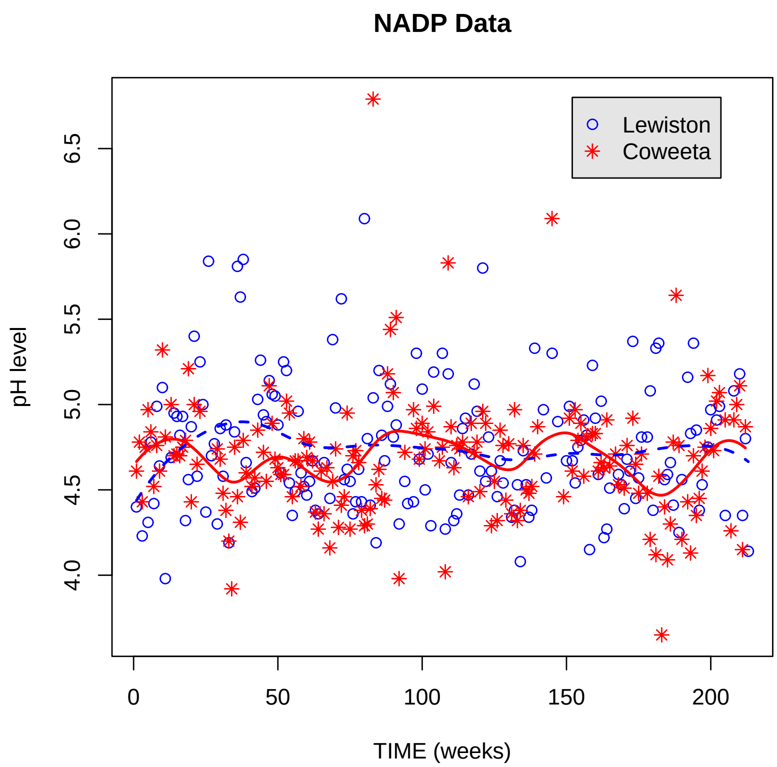

Figure 5.4. Scatterplot of NADP Data.

It should be pointed out that there are several weeks in which data were missing at one or both locations. One interesting feature in this data set is that there are 180 weeks of data in both Lewiston and Coweeta among the total 233 weeks, although this balancedness of the design is simply a coincidence. A further examination reveals that the missing data in fact happen at different time points in the two locations. This fact does not cause any difficulties to the implementation of our methodology. However, it does matter for some procedures which are applicable to studies like NADP, although they were designed for different type of hypotheses. For instance, the bootstrap test of Hall and Hart (1990) for the hypothesis of no location effect can be conducted using only those weeks in which there are no missing data at both locations, meaning that it can only use less than 150 weeks of data in this study. That is a loss of $30/180 = 1/6$ of the data. This constraint could be a crucial drawback in practice, especially when the number of observations in each group is smaller moderate.

For the covariate effect of time in our analysis, the *CF* test yields a p-value of .5737, the *Drop* test yields a p-value of .7527, and the *NP(w)* test yields p-values of .0217, .0239, and .0221, for $w = 5,\ 7,\ 9$, respectively. As seen in the previous two examples, the choice of the local window size w does not seem to affect the testing result of the proposed procedure *NP*. In addition, the failure of the *CF* and *Drop* tests in detecting the time effect in this example again confirms their lack of power when the effect of interest is not linear.

Recall that the application of the *non-overlapping windows approach* on this data set is demonstrated in Section 2.4.2 and the *HET* procedure gives a p-value of 0.0508 for the effect of time. The small p-values of the *NP(w)* procedures, compared with that of the *HET* procedure, confirms the superiority of the *overlapping windows approach* over the *non-overlapping windows approach* for detecting the effect of the covariate in analysis of covariance.

5.5 Technical Details

5.5.1 Some Calculations

Calculation of T_1 and T_2

Firstly, for T_1 in Equation 5.2.1,

$$\because \quad \sum_{t=1}^{w_{ir}} \left(Z_{irt} - \bar{Z}_{ir\cdot}\right)^2 = \sum_t Z_{irt}^2 - w_{ir}\bar{Z}_{ir\cdot}^2 = \sum_t Z_{irt}^2 - \frac{1}{w_{ir}}(\sum_t Z_{irt})^2$$
$$= \mathbf{Z}_{ir}^{'}\mathbf{I}_{w_{ir}}\mathbf{Z}_{ir} - \frac{1}{w_{ir}}\mathbf{Z}_{ir}^{'}\mathbf{1}_{w_{ir}}\mathbf{1}_{w_{ir}}^{'}\mathbf{Z}_{ir}$$

$$MSE_D = \frac{1}{N-k}\sum_{i=1}^{k}\sum_{r=1}^{n_i}\left(1-\frac{w_{ir}}{w_{i\cdot}}\right)\frac{1}{w_{ir}-1}\sum_{t=1}^{w_{ir}}\left(Z_{irt}-\bar{Z}_{ir\cdot}\right)^2$$
$$= \frac{1}{N-k}\sum_{i=1}^{k}\sum_{r=1}^{n_i}\left(1-\frac{w_{ir}}{w_{i\cdot}}\right)\frac{1}{w_{ir}-1}\mathbf{Z}_{ir}^{'}\left[\mathbf{I}_{w_{ir}}-\frac{1}{w_{ir}}\mathbf{J}_{w_{ir}}\right]\mathbf{Z}_{ir}$$

$$= \mathbf{Z}' \left\{ \frac{1}{N-k} \bigoplus_{i=1}^{k} \bigoplus_{r=1}^{n_i} \left(1 - \frac{w_{ir}}{w_{i\cdot}}\right) \frac{1}{w_{ir}-1} \left[\mathbf{I}_{w_{ir}} - \frac{1}{w_{ir}} \mathbf{J}_{w_{ir}} \right] \right\} \mathbf{Z}'.$$

Secondly, for T_2 in Equation 5.2.2,

$$\begin{aligned}
\because \quad \sum_{r=1}^{n_i} w_{ir} \left(\bar{Z}_{ir\cdot} - \bar{Z}_{i\cdot\cdot}\right)^2 &= \sum_r w_{ir} \bar{Z}_{ir\cdot}^2 - w_{i\cdot} \bar{Z}_{i\cdot\cdot}^2 \\
&= \sum_r w_{ir} \left(\frac{1}{w_{ir}} \sum_t Z_{irt} \right)^2 - w_{i\cdot} \left(\frac{1}{w_{i\cdot}} \sum_r \sum_t Z_{irt} \right)^2 \\
&= \sum_r \frac{1}{w_{ir}} \mathbf{Z}_{ir}' \mathbf{J}_{w_{ir}} \mathbf{Z}_{ir} - \frac{1}{w_{i\cdot}} \mathbf{Z}_i' \mathbf{J}_{w_{i\cdot}} \mathbf{Z}_i \\
&= \mathbf{Z}_i' \left(\bigoplus_r \frac{1}{w_{ir}} \mathbf{J}_{w_{ir}} - \frac{1}{w_{i\cdot}} \mathbf{J}_{w_{i\cdot}} \right) \mathbf{Z}_i
\end{aligned}$$

$$\begin{aligned}
MST_D &= \frac{1}{N-k} \sum_{i=1}^{k} \sum_{r=1}^{n_i} w_{ir} \left(\bar{Z}_{ir\cdot} - \bar{Z}_{i\cdot\cdot}\right)^2 \\
&= \mathbf{Z}' \left\{ \frac{1}{N-k} \bigoplus_i \left(\bigoplus_r \frac{1}{w_{ir}} \mathbf{J}_{w_{ir}} - \frac{1}{w_{i\cdot}} \mathbf{J}_{w_{i\cdot}} \right) \right\} \mathbf{Z}
\end{aligned}$$

Calculation of the projection matrix $\mathbf{A}_D$

Let $\mathbf{U}_{ir} = (U_{ir1}, \cdots, U_{irw_{ir}})'$ be *independent* random vectors with independent components; WLOG, assume that $E(U_{irt}) = 0$.

Let $\mathbf{U}' = (\mathbf{U}_1', \cdots, \mathbf{U}_k') = ((\mathbf{U}_{11}', \cdots, \mathbf{U}_{1n_1}'), \cdots, (\mathbf{U}_{k1}', \cdots, \mathbf{U}_{kn_k}'))$. Then,

$$\mathbf{U}'\mathbf{A}\mathbf{U} = \sum_{i=1}^{k} \mathbf{U}_i' \mathbf{A}_i \mathbf{U}_i$$

and the projection (Hájek's projection) of $\mathbf{U}'\mathbf{A}\mathbf{U}$ is defined as

$$\sum_i \sum_r E(\mathbf{U}'\mathbf{A}\mathbf{U} | \mathbf{U}_{ir}) - (N-k) E(\mathbf{U}'\mathbf{A}\mathbf{U}).$$

Note that

$$\mathbf{U}_i^{'}\mathbf{A}_i\mathbf{U}_i = (\mathbf{U}_{i1}^{'},\cdots,\mathbf{U}_{in_i}^{'})\begin{pmatrix} \mathbf{B}_{i1} & -\beta_i\mathbf{1}_{w_{i1}}\mathbf{1}_{w_{i2}}^{'} & \cdots & -\beta_i\mathbf{1}_{w_{i1}}\mathbf{1}_{w_{in_i}}^{'} \\ -\beta_i\mathbf{1}_{w_{i2}}\mathbf{1}_{w_{i1}}^{'} & \mathbf{B}_{i2} & & \vdots \\ \vdots & & \ddots & \vdots \\ -\beta_i\mathbf{1}_{w_{in_i}}\mathbf{1}_{w_{i1}}^{'} & \cdots & \cdots & \mathbf{B}_{in_i} \end{pmatrix}\begin{pmatrix} \mathbf{U}_{i1} \\ \vdots \\ \vdots \\ \mathbf{U}_{in_i}\end{pmatrix}$$

$$= \left(\mathbf{U}_{i1}^{'}\mathbf{B}_{i1} - \beta_i\sum_{l\neq 1}\mathbf{U}_{il}^{'}\mathbf{1}_{w_{il}}\mathbf{1}_{w_{i1}}^{'},\cdots,\mathbf{U}_{in_i}^{'}\mathbf{B}_{in_i} - \beta_i\sum_{l\neq n_i}\mathbf{U}_{il}^{'}\mathbf{1}_{w_{il}}\mathbf{1}_{w_{in_i}}^{'}\right)\begin{pmatrix} \mathbf{U}_{i1} \\ \vdots \\ \vdots \\ \mathbf{U}_{in_i}\end{pmatrix}$$

$$= \sum_{r=1}^{n_i}\mathbf{U}_{ir}^{'}\mathbf{B}_{ir}\mathbf{U}_{ir} - \beta_i\sum_{r=1}^{n_i}\left(\sum_{l\neq r}\mathbf{U}_{il}^{'}\mathbf{1}_{w_{il}}\right)\mathbf{1}_{w_{ir}}^{'}\mathbf{U}_{ir}$$

$$= \sum_{r=1}^{n_i}\mathbf{U}_{ir}^{'}\mathbf{B}_{ir}\mathbf{U}_{ir} - \beta_i\left(\sum_{r_1=1}^{n_i}\sum_{t_1=1}^{w_{ir_1}}U_{ir_1t_1}\right)\left(\sum_{r_2\neq r_1}^{n_i}\sum_{t_2=1}^{w_{ir_2}}U_{ir_2t_2}\right)$$

$$= \sum_{r=1}^{n_i}\mathbf{U}_{ir}^{'}\mathbf{B}_{ir}\mathbf{U}_{ir} - \beta_i\sum_{r_1=1}^{n_i}\sum_{r_2\neq r_1}^{n_i}\sum_{t_1=1}^{w_{ir_1}}\sum_{t_2=1}^{w_{ir_2}}U_{ir_1t_1}U_{ir_2t_2}$$

So,

$$\mathbf{U}^{'}\mathbf{A}\mathbf{U} = \sum_{i=1}^{k}\mathbf{U}_i^{'}\mathbf{A}_i\mathbf{U}_i = \sum_{i=1}^{k}\left\{\sum_{r=1}^{n_i}\mathbf{U}_{ir}^{'}\mathbf{B}_{ir}\mathbf{U}_{ir} - \beta_i\sum_{r_1=1}^{n_i}\sum_{r_2\neq r_1}^{n_i}\sum_{t_1=1}^{w_{ir_1}}\sum_{t_2=1}^{w_{ir_2}}U_{ir_1t_1}U_{ir_2t_2}\right\}.$$

Because

$$E(U_{ir_1t_1}U_{ir_2t_2}|\mathbf{U}_{il}) = 0, \quad \forall l,\ \forall r_1\neq r_2,$$

we have

$$E(\mathbf{U}_i^{'}\mathbf{A}_i\mathbf{U}_i|\mathbf{U}_{il}) = E\left(\sum_{r=1}^{n_i}\mathbf{U}_{ir}^{'}\mathbf{B}_{ir}\mathbf{U}_{ir} - \beta_i\sum_{r_1=1}^{n_i}\sum_{r_2\neq r_1}^{n_i}\sum_{t_1=1}^{w_{ir_1}}\sum_{t_2=1}^{w_{ir_2}}U_{ir_1t_1}U_{ir_2t_2}\mid\mathbf{U}_{il}\right)$$

$$= \sum_{r=1}^{n_i}E(\mathbf{U}_{ir}^{'}\mathbf{B}_{ir}\mathbf{U}_{ir}|\mathbf{U}_{il})$$

$$= \mathbf{U}_{il}^{'}\mathbf{B}_{il}\mathbf{U}_{il} + \sum_{r\neq l} E(\mathbf{U}_{ir}^{'}\mathbf{B}_{ir}\mathbf{U}_{ir})$$

Note that

$$\begin{aligned}
E(\mathbf{U}_{ir}^{'}\mathbf{B}_{ir}\mathbf{U}_{ir}) &= E\left\{(U_{ir1}, U_{ir2}, \cdots, U_{irw_{ir}})\begin{pmatrix} 0 & \gamma_{ir} & \cdots & \gamma_{ir} \\ \gamma_{ir} & 0 & \cdots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_{ir} & \gamma_{ir} & \cdots & 0 \end{pmatrix}\begin{pmatrix} U_{ir1} \\ U_{ir2} \\ \vdots \\ U_{irw_{ir}} \end{pmatrix}\right\} \\
&= \sum_{t=1}^{w_{ir}} E\left(U_{irt}^2 \times 0\right) + \gamma_{ir}\ E\left(\sum_{t_1}\sum_{t_2\neq t_1} U_{irt_1}U_{irt_2}\right) \\
\Longrightarrow E(\mathbf{U}_{ir}^{'}\mathbf{B}_{ir}\mathbf{U}_{ir}) &= 0,\ \forall i,\ \forall r.
\end{aligned}$$

Therefore,

$$E(\mathbf{U}_i^{'}\mathbf{A}_i\mathbf{U}_i|\mathbf{U}_{il}) = \mathbf{U}_{il}^{'}\mathbf{B}_{il}\mathbf{U}_{il}$$

In addition,

$$\begin{aligned}
\because\ E(\mathbf{U}_i^{'}\mathbf{A}_i\mathbf{U}_i) &= E(E(\mathbf{U}_i^{'}\mathbf{A}_i\mathbf{U}_i|\mathbf{U}_{il})) = E(\mathbf{U}_{il}^{'}\mathbf{B}_{il}\mathbf{U}_{il}) = 0 \\
\therefore\ E(\mathbf{U}^{'}\mathbf{A}\mathbf{U}|\mathbf{U}_{il}) &= E\left(\sum_{i_1=1}^{k}\mathbf{U}_{i_1}^{'}\mathbf{A}_{i_1}\mathbf{U}_{i_1}|\mathbf{U}_{il}\right) = E\left(\mathbf{U}_i^{'}\mathbf{A}_i\mathbf{U}_i|\mathbf{U}_{il}\right) + E\left(\sum_{i_1\neq i}\mathbf{U}_{i_1}^{'}\mathbf{A}_{i_1}\mathbf{U}_{i_1}|\mathbf{U}_{il}\right) \\
&= \mathbf{U}_{il}^{'}\mathbf{B}_{il}\mathbf{U}_{il} + E\left(\sum_{i_1\neq i}\mathbf{U}_{i_1}^{'}\mathbf{A}_{i_1}\mathbf{U}_{i_1}\right) \\
&= \mathbf{U}_{il}^{'}\mathbf{B}_{il}\mathbf{U}_{il} \\
\Longrightarrow E(\mathbf{U}^{'}\mathbf{A}\mathbf{U}) &= E[E(\mathbf{U}^{'}\mathbf{A}\mathbf{U}|\mathbf{U}_{il})] = E(\mathbf{U}_{il}^{'}\mathbf{B}_{il}\mathbf{U}_{il}) = 0.
\end{aligned}$$

Therefore, the projection (Hájek's projection) of $\mathbf{U}^{'}\mathbf{A}\mathbf{U}$ is defined as

$$\sum_i\sum_r E(\mathbf{U}^{'}\mathbf{A}\mathbf{U}|\mathbf{U}_{ir}) - (N-k)E(\mathbf{U}^{'}\mathbf{A}\mathbf{U}) = \sum_i\sum_r \mathbf{U}_{ir}^{'}\mathbf{B}_{ir}\mathbf{U}_{ir}$$

$$\Longrightarrow \quad \therefore\ \mathbf{A}_D = \bigoplus_{i=1}^{k}\bigoplus_{r=1}^{n_i}\mathbf{B}_{ir}.$$

5.5.2 Proofs of Lemmas and Theorems

Here are some more notations used in the following proofs. Let $\mathbf{Z}^0 = \mathbf{Z} - E_{F_i^0}(\mathbf{Z}|\mathbf{X})$ where

$$E_{F_i^0}(\mathbf{Z}|\mathbf{X}) = E(\mathbf{Z}|\mathbf{X};\ \text{under } H_0(D) \text{ as defined in (1.2.9)}).$$

In addition, let $\mathbf{A}^* = \mathbf{A} - \mathbf{A}_D$, where $\mathbf{A}_D$ is as define in (5.2.3).

Proof of Lemma 5.2.1

Write $F_{ix}(y)$ as $F_i(y;x)$. Consider $Z_{irt} = Y_{ij}$ with the covariate $X_{ij} = x_{ij}$; in other words, $\sum_{l=1}^{n_i} I(X_{il} \le x_{ij};\, l \in W_{ir}) = t$. Although in general

$$\begin{aligned}
E[Z_{irt} - \bar{Z}_{ir\cdot} \mid \mathbf{X} = \mathbf{x}] &= E\left[Y_{ij} - \frac{1}{w_{ir}} \sum_{l=1}^{n_i} Y_{il} I(l \in W_{ir}) \mid \mathbf{X} = \mathbf{x}\right] \\
&= \int y dF_i(y; x_{ij}) - \frac{1}{w_{ir}} \sum_{l=1}^{n_i} \int y dF_i(y; x_{il}) \cdot I(l \in W_{ir}) \\
&\neq 0,
\end{aligned}$$

under $H_0(D):\ F_i(y;x) = F_i^0(y),\ \forall x,$

$$\begin{aligned}
E[Z_{irt} - \bar{Z}_{ir\cdot} \mid \mathbf{X} = \mathbf{x}] &\overset{H_0(D)}{=} \int y dF_i^0(y) - \frac{1}{w_{ir}} \sum_{l=1}^{n_i} \int y dF_i^0(y) \cdot I(l \in W_{ir}) \\
&\overset{H_0(D)}{=} 0,\ \forall \mathbf{x},\ \forall i,\ \forall r,\ \forall t.
\end{aligned}$$

Similarly,

$$\begin{aligned}
& E[\bar{Z}_{ir\cdot} - \bar{Z}_{i\cdot\cdot} \mid \mathbf{X} = \mathbf{x}] \\
= \quad & \frac{1}{w_{ir}} \sum_{j=1}^{n_i} \int y\ dF_i(y; x_{ij}) \cdot I(j \in W_{ir}) - \frac{1}{w_{i\cdot}} \sum_{r=1}^{n_i} \sum_{j=1}^{n_i} \int y\ dF_i(y; x_{ij}) \cdot I(j \in W_{ir}) \\
= \quad & \int y\ d\left[\frac{1}{w_{ir}} \sum_{j=1}^{n_i} F_i(y; x_{ij}) \cdot I(j \in W_{ir}) - \frac{1}{w_{i\cdot}} \sum_{r=1}^{n_i} \sum_{j=1}^{n_i} F_i(y; x_{ij}) \cdot I(j \in W_{ir})\right]
\end{aligned}$$

$$\overset{H_0(D)}{=} \int y \, dF_i^0(y) \left[\frac{1}{w_{ir}} \sum_{j=1}^{n_i} I(j \in W_{ir}) - \frac{1}{w_{i\cdot}} \sum_{r=1}^{n_i} \sum_{j=1}^{n_i} I(j \in W_{ir}) \right]$$

$$\overset{H_0(D)}{=} 0, \ \forall \mathbf{x}, \ \forall i, \ \forall r.$$

Recall that $\mathbf{A} = \mathbf{T}_2 - \mathbf{T}_1$ with $\mathbf{T}_1$ and $\mathbf{T}_2$ as defined in (5.2.1) and (5.2.2). It can be shown that

$$\begin{aligned} \mathbf{T}_1\mathbf{Z} &= \frac{1}{N-k} \left[\bigoplus_{i=1}^{k} \bigoplus_{r=1}^{n_i} \left(1 - \frac{w_{ir}}{w_{i\cdot}} \right) \frac{1}{w_{ir}-1} \left(\mathbf{I}_{w_{ir}} - \frac{1}{w_{ir}} \mathbf{J}_{w_{ir}} \right) \right] \mathbf{Z} \\ &= \frac{1}{N-k} \begin{pmatrix} (1 - \frac{w_{11}}{w_{1\cdot}}) \ \frac{1}{w_{11}-1} \ (\mathbf{Z}_{11} - \bar{Z}_{11\cdot} \ \mathbf{1}_{w_{11}}) \\ \vdots \\ (1 - \frac{w_{1n_1}}{w_{1\cdot}}) \ \frac{1}{w_{1n_1}-1} \ (\mathbf{Z}_{1n_1} - \bar{Z}_{1n_1\cdot} \ \mathbf{1}_{w_{1n_1}}) \\ (1 - \frac{w_{21}}{w_{2\cdot}}) \ \frac{1}{w_{21}-1} \ (\mathbf{Z}_{21} - \bar{Z}_{21\cdot} \ \mathbf{1}_{w_{21}}) \\ \vdots \\ (1 - \frac{w_{kn_k}}{w_{k\cdot}}) \ \frac{1}{w_{kn_k}-1} \ (\mathbf{Z}_{kn_k} - \bar{Z}_{kn_k\cdot} \ \mathbf{1}_{w_{kn_k}}) \end{pmatrix}_{w_{\cdot\cdot} \times 1} \\ \Longrightarrow \quad & \mathbf{T}_1 E_{F_i^0}(\mathbf{Z} \mid \mathbf{X}) = \mathbf{0}_{w_{\cdot\cdot} \times 1}. \end{aligned}$$

and

$$\begin{aligned} \mathbf{T}_2\mathbf{Z} &= \frac{1}{N-l} \left(\bigoplus_{i=1}^{k} \bigoplus_{r=1}^{n_i} \frac{1}{w_{ir}} \mathbf{J}_{w_{ir}} \right) \mathbf{Z} - \frac{1}{N-k} \left(\bigoplus_{i=1}^{k} \frac{1}{w_{i\cdot}} \ \mathbf{J}_{w_{i\cdot}} \right) \mathbf{Z} \\ &= \frac{1}{N-k} \begin{pmatrix} (\bar{Z}_{11\cdot} - \bar{Z}_{1\cdot\cdot})\mathbf{1}_{w_{11}} \\ \vdots \\ (\bar{Z}_{1n_1\cdot} - \bar{Z}_{1\cdot\cdot})\mathbf{1}_{w_{1n_1}} \\ (\bar{Z}_{21\cdot} - \bar{Z}_{2\cdot\cdot})\mathbf{1}_{w_{21}} \\ \vdots \\ (\bar{Z}_{kn_k\cdot} - \bar{Z}_{k\cdot\cdot})\mathbf{1}_{w_{kn_k}} \end{pmatrix}_{w_{\cdot\cdot} \times 1} \\ \Longrightarrow \quad & \mathbf{T}_2 E_{F_i^0}(\mathbf{Z} \mid \mathbf{X}) = \mathbf{0}_{w_{\cdot\cdot} \times 1}. \end{aligned}$$

Hence,

$$\mathbf{A}\ E(\mathbf{Z} \mid \mathbf{X}) \overset{H_0(D)}{=} (\mathbf{T}_2 - \mathbf{T}_1)E_{F_i^0}(\mathbf{Z} \mid \mathbf{X}) = \mathbf{0}_{w..\times 1}. \tag{5.5.1}$$

As a consequence,

$$\begin{aligned}\mathbf{Z}^{*\prime}\mathbf{A}\mathbf{Z}^{*} &= (\mathbf{Z} - E(\mathbf{Z} \mid \mathbf{X}))'\,\mathbf{A}\,(\mathbf{Z} - E(\mathbf{Z} \mid \mathbf{X}))\\ &= \mathbf{Z}'\mathbf{A}\mathbf{Z} - E(\mathbf{Z} \mid \mathbf{X})'\mathbf{A}\mathbf{Z} - (\mathbf{Z} - E(\mathbf{Z} \mid \mathbf{X}))'\,\mathbf{A}E(\mathbf{Z} \mid \mathbf{X})\\ &\overset{H_0(D)}{=} \mathbf{Z}'\mathbf{A}\mathbf{Z},\end{aligned}$$

which completes the proof.

Proof of Lemma 5.2.5

First note that

$$\begin{aligned}\mathbf{Z}^{*\prime}\mathbf{A}\mathbf{Z}^{*} &= [\mathbf{Z} - E(\mathbf{Z}|\mathbf{X})]'\mathbf{A}[\mathbf{Z} - E(\mathbf{Z}|\mathbf{X})]\\ &= \mathbf{Z}'\mathbf{A}\mathbf{Z} - 2E(\mathbf{Z}|\mathbf{X})'\mathbf{A}[\mathbf{Z} - E(\mathbf{Z}|\mathbf{X})] - E(\mathbf{Z}|\mathbf{X})'\mathbf{A}E(\mathbf{Z}|\mathbf{X}).\end{aligned}$$

From (5.5.1), we learned that $\mathbf{A}E_{F_i^0}(\mathbf{Z}|\mathbf{X}) = \mathbf{0}_{w..\times 1}$, so

$$\begin{aligned}&\mathbf{Z}'\mathbf{A}\mathbf{Z} - \mathbf{Z}^{*\prime}\mathbf{A}\mathbf{Z}^{*} = 2\ E(\mathbf{Z}|\mathbf{X})'\mathbf{A}\mathbf{Z}^{*} + E(\mathbf{Z}|\mathbf{X})'\mathbf{A}E(\mathbf{Z}|\mathbf{X})\\ =\ & 2\left[E(\mathbf{Z}|\mathbf{X}) - E_{F_i^0}(\mathbf{Z}|\mathbf{X})\right]'\mathbf{A}\mathbf{Z}^{*} + \left[E(\mathbf{Z}|\mathbf{X}) - E_{F_i^0}(\mathbf{Z}|\mathbf{X})\right]'\mathbf{A}\left[E(\mathbf{Z}|\mathbf{X}) - E_{F_i^0}(\mathbf{Z}|\mathbf{X})\right]\\ =\ & 2\ d(\mathbf{X})'\mathbf{A}E(\mathbf{Z}|\mathbf{X}) + d(\mathbf{X})'\mathbf{A}d(\mathbf{X}),\end{aligned}$$

where

$$d(\mathbf{X}) \equiv E(\mathbf{Z}|\mathbf{X}) - E_{F_i^0}(\mathbf{Z}|\mathbf{X}).$$

Let d_{irt} be the elements of $d(\mathbf{X})$, i.e. $d_{irt} = \int y dD_i(y; X_{irt})$, and

$$d_{ij} = \int y dD_i(y; X_{ij}), \quad \text{and} \quad c_{ij} = \int y dC_i(y; X_{ij}). \tag{5.5.2}$$

Then, it is clear that $d_{ij} = (n_i w)^{-1/4} c_{ij},\ i = 1, \cdots, k;\ j = 1, \cdots, n_i$.

To prove that $N^{1/2} w^{-1/2} (\mathbf{Z}'\mathbf{A}\mathbf{Z} - \mathbf{Z}^{*\prime}\mathbf{A}\mathbf{Z}^*) \to \theta_D$ in probability, it's sufficient to show that (1) $N^{1/2} w^{-1/2} d(\mathbf{X})'\mathbf{A} d(\mathbf{X}) = \theta_D + o_p(1)$ and (2) $N^{1/2} w^{-1/2} d(\mathbf{X})'\mathbf{A}\mathbf{Z}^* = o_p(1)$.

We first prove (1). Note that

$$\begin{aligned}
\mathbf{d}(\mathbf{X})'\mathbf{A}\mathbf{d}(\mathbf{X}) &= \frac{1}{N-k} \sum_{i=1}^{k} \sum_{r=1}^{n_i} w_{ir} \left(\bar{d}_{ir\cdot} - \bar{d}_{i\cdot\cdot} \right)^2 \\
&\quad - \frac{1}{N-k} \sum_{i=1}^{k} \sum_{r=1}^{n_i} \left(1 - \frac{w_{ir}}{w_{i\cdot}} \right) \frac{1}{w_{ir} - 1} \sum_{t=1}^{w_{ir}} \left(d_{irt} - \bar{d}_{ir\cdot} \right)^2 \\
&= \frac{1}{N-k} \sum_{i=1}^{k} \sum_{r}^{n_i} w_{ir} \bar{d}_{ir\cdot}^2 - \frac{1}{N-k} \sum_{i=1}^{k} w_{i\cdot} \bar{d}_{i\cdot\cdot}^2 \\
&\quad - \frac{1}{N-k} \sum_{i=1}^{k} \sum_{r=1}^{n_i} \left(1 - \frac{w_{ir}}{w_{i\cdot}} \right) \frac{1}{w_{ir} - 1} \left\{ \sum_{t=1}^{w_{ir}} d_{irt}^2 - w_{ir} \bar{d}_{ir\cdot}^2 \right\}
\end{aligned}$$

Since $\int y dC_i(y; x)$ is uniformly Lipschitz continuous (by Assumption 3), Lemma 5.5.1 tells us that for all n_i large enough,

$$\begin{aligned}
\bar{d}_{ir\cdot} &= \frac{1}{w_{ir}} \sum_{t=1}^{w_{ir}} d_{irt} = \frac{1}{w_{ir}} \sum_{j=1}^{n_i} d_{ij} \cdot I(j \in W_{ir}) \\
&= (n_i w)^{-\frac{1}{4}} \frac{1}{w_{ir}} \sum_{j=1}^{n_i} \int y dC_i(y; X_{ij}) \cdot I(j \in W_{ir}) \\
&= (n_i w)^{-\frac{1}{4}} \left[\int y dC_i(y; X_i r) + O\left(\frac{w}{n_i} + n_i^{-1+\delta} \right) \right],\ \text{uniformly a.s.} \\
&= (n_i w)^{-\frac{1}{4}} c_{ir} + O\left(n_i^{-\frac{5}{4}} w^{\frac{3}{4}} + n_i^{-\frac{5}{4}+\delta} w^{-\frac{1}{4}} \right),\ \text{u.a.s.}
\end{aligned}$$

for all $\delta > 0$.

As a consequence, we have firstly,

$$\begin{aligned}
\bar{d}_{ir\cdot}^2 &= (n_i w)^{-\frac{1}{2}} c_{ir}^2 + O\left[(n_i w)^{-\frac{1}{4}} \left(n_i^{-\frac{5}{4}} w^{\frac{3}{4}} + n_i^{-\frac{5}{4}+\delta} w^{-\frac{1}{4}} \right) \right],\ \text{u.a.s.} \\
&= (n_i w)^{-\frac{1}{2}} c_{ir}^2 + O\left(n_i^{-\frac{3}{2}} w^{\frac{1}{2}} + n_i^{-\frac{3}{2}+\delta} w^{-\frac{1}{2}} \right),\ \text{u.a.s.;}
\end{aligned}$$

secondly,

$$\begin{aligned}
\bar{d}_{i\cdot\cdot} &= \frac{1}{w_{i\cdot}}\sum_{r=1}^{n_i} w_{ir}\bar{d}_{ir\cdot} \\
&= \frac{(n_i w)^{-\frac{1}{4}}}{w_{i\cdot}}\sum_{r=1}^{n_i} w_{ir}c_{ir} + O\left(n_i^{-\frac{5}{4}}w^{\frac{3}{4}} + n_i^{-\frac{5}{4}+\delta}w^{-\frac{1}{4}}\right), \text{ u.a.s.} \\
\Longrightarrow \bar{d}_{i\cdot\cdot}^2 &= (n_i w)^{-\frac{1}{2}}\left(\frac{1}{w_{i\cdot}}\sum_{r=1}^{n_i} w_{ir}c_{ir}\right)^2 + O\left(n_i^{-\frac{3}{2}}w^{\frac{1}{2}} + n_i^{-\frac{3}{2}+\delta}w^{-\frac{1}{2}}\right), \text{ u.a.s.;}
\end{aligned}$$

and thirdly,

$$\sum_{t=1}^{w_{ir}} d_{irt}^2 = \sum_{j=1}^{n_i} d_{ij}^2 \times I(j \in W_{ir}) = (n_i w)^{-\frac{1}{2}}\sum_{j=1}^{n_i} c_{ij}^2 \times I(j \in W_{ir})$$

$$\begin{aligned}
\Longrightarrow & \sum_{r=1}^{n_i}\left(1 - \frac{w_{ir}}{w_{i\cdot}}\right)\frac{1}{w_{ir}-1}\left[\sum_{t=1}^{w_{ir}} d_{irt}^2 - w_{ir}\bar{d}_{ir\cdot}^2\right] \\
= \quad & O\left(\frac{1}{w}\right)\sum_{r=1}^{n_i}\Bigg[(n_i w)^{-\frac{1}{2}}\sum_{j=1}^{n_i} c_{ij}^2 \times I(j \in W_{ir}) \\
& \qquad -w_{ir}\left\{(n_i w)^{-\frac{1}{2}}c_{ir}^2 + O\left(n_i^{-\frac{3}{2}}w^{\frac{1}{2}} + n_i^{-\frac{3}{2}+\delta}w^{-\frac{1}{2}}\right)\right\}\Bigg], \ u.a.s. \\
= \quad & O\left(\frac{1}{w}\right)\Bigg[(n_i w)^{-\frac{1}{2}}\sum_{j=1}^{n_i} w_{ij}c_{ij}^2 - (n_i w)^{-\frac{1}{2}}\sum_{j=1}^{n_i} w_{ij}c_{ij}^2 \\
& \qquad +O\left(n_i w\left(n_i^{-\frac{3}{2}}w^{\frac{1}{2}} + n_i^{-\frac{3}{2}+\delta}w^{-\frac{1}{2}}\right)\right)\Bigg], \ u.a.s. \\
= \quad & O\left(n_i^{-.5}w^{.5} + n_i^{-.5+\delta}w^{-.5}\right), \ u.a.s
\end{aligned}$$

Therefore,

$$\begin{aligned}
& N^{1/2}w^{-1/2}d(\mathbf{X})'\mathbf{A}d(\mathbf{X}) \\
= \quad & \frac{N^{.5}w^{-.5}}{N-k}\sum_i\sum_r w_{ir}\left[(n_i w)^{-\frac{1}{2}}c_{ir}^2 + O\left(n_i^{-\frac{3}{2}}w^{\frac{1}{2}} + n_i^{-\frac{3}{2}+\delta}w^{-\frac{1}{2}}\right)\right] \\
& -\frac{N^{.5}w^{-.5}}{N-k}\sum_i w_{i\cdot}\left[(n_i w)^{-\frac{1}{2}}\left(\frac{1}{w_{i\cdot}}\sum_{r=1}^{n_i} w_{ir}c_{ir}\right)^2 + O\left(n_i^{-\frac{3}{2}}w^{\frac{1}{2}} + n_i^{-\frac{3}{2}+\delta}w^{-\frac{1}{2}}\right)\right]
\end{aligned}$$

$$
\begin{aligned}
& -\frac{N^{.5}w^{-.5}}{N-k}\sum_i O\left(n_i^{-.5}w^{.5}+n_i^{-.5+\delta}w^{-.5}\right), \ u.a.s. \\
= \quad & \frac{N}{N-k}\sum_{i=1}^{k}\sqrt{\frac{n_i}{N}}\frac{w_{i\cdot}}{n_i w}\left[\left(\frac{1}{w_{i\cdot}}\sum_r w_{ir}c_{ir}^2\right)-\left(\frac{1}{w_{i\cdot}}\sum_r w_{ir}c_{ir}\right)^2\right] \\
& +O\left[N^{-.5}w^{-.5}\left(n_i^{-\frac{3}{2}}w^{\frac{1}{2}}+n_i^{-\frac{3}{2}+\delta}w^{-\frac{1}{2}}\right)\right] \\
& +O\left[N^{-.5}w^{-.5}\left(n_i^{-.5}w^{.5}+n_i^{-.5+\delta}w^{-.5}\right)\right], \ u.a.s. \\
= \quad & O\left(N^{-1}w+N^{-1+\delta}\right), \ u.a.s. \\
= \quad & o(1)
\end{aligned}
$$

By WLLN,

$$
\begin{aligned}
\frac{1}{w_{i\cdot}}\sum_{r=1}^{n_i} w_{ir}c_{ir} \quad & = \quad \frac{1}{w_{i\cdot}}\sum_{r=1}^{n_i} w_{ir}\int y dC_i(y;X_{ir}) \xrightarrow{P} \int\int y dC_i(y;x)dG_i(x) \\
\frac{1}{w_{i\cdot}}\sum_{r=1}^{n_i} w_{ir}c_{ir}^2 \quad & \xrightarrow{P} \quad \int\left[\int y dC_i(y;x)\right]^2 dG_i(x)
\end{aligned}
$$

Therefore,

$$
\left(\frac{N}{w}\right)^{1/2} d(\mathbf{X})'\mathbf{A}d(\mathbf{X}) \xrightarrow{P} \theta_D, \ \text{as } \min(n_i)\to\infty.
$$

We now prove (2). Since $\mathbf{A}=\mathbf{A}_D+\mathbf{A}^*$, it is equivalent to prove that

$$
R_1 = N^{1/2}w^{-1/2}d(\mathbf{X})'\mathbf{A}_D\mathbf{Z}^* = o_p(1), \tag{5.5.3}
$$

$$
\text{and} \quad R_2 = N^{1/2}w^{-1/2}d(\mathbf{X})'\mathbf{A}^*\mathbf{Z}^* = o_p(1). \tag{5.5.4}
$$

To prove (5.5.3), first note that

$$
\begin{aligned}
d(\mathbf{X})'\mathbf{A}_D\mathbf{Z}^* \quad & = \quad \sum_{i=1}^{k}\sum_{r=1}^{n_i}\mathbf{d}_{ir}'\mathbf{B}_{ir}\mathbf{Z}_{ir}^* = \sum_{i=1}^{k}\sum_{r=1}^{n_i}\gamma_{ir}\sum_{t_1=1}^{w} d_{irt_1}\sum_{t_2\neq t_1} Z_{irt_2}^* \\
& = \quad \sum_{i=1}^{k}\sum_{r=1}^{n_i}\gamma_{ir}\sum_{j_1\neq j_2} d_{ij_1}Y_{ij_2}^* I(j_1,j_2\in W_{ir}).
\end{aligned}
$$

Because

$$E\left(R_1|\mathbf{X}\right) = \left(\frac{N}{w}\right)^{1/2} \sum_{i=1}^{k}\sum_{r=1}^{n_i} \gamma_{ir} \sum_{j_1 \neq j_2} d_{ij_1} E(Y^*_{ij_2}|\mathbf{X}) I(j_1, j_2 \in W_{ir}) = 0,$$

and

$$\begin{aligned}
& E\left(R_1^2|\mathbf{X}\right) = \frac{N}{w} E\left[\left(d(\mathbf{X})'\mathbf{A}_D\mathbf{Z}^*\right)^2 |\mathbf{X}\right] \\
= \; & \frac{N}{w} \sum_{i_1}\sum_{i_2}\sum_{r_1}\sum_{r_2} \gamma_{i_1r_1}\gamma_{i_2r_2} \sum_{j_1\neq j_2}\sum_{l_1 \neq l_2} d_{i_1j_1} d_{i_2l_1} E(Y^*_{i_1j_2} Y^*_{i_2l_2}|\mathbf{X}) \\
& \quad \times I(j_1, j_2 \in W_{i_1r_1}) \times I(l_1, l_2 \in W_{i_2r_2}) \\
= \; & \sum_{i=1}^{k}\sum_{r_1=1}^{n_i}\sum_{r_2=1}^{n_i} O\left(N^{-2}w^{-2}\right) \sum_{m=1}^{n_i}\sum_{j\neq m}\sum_{l \neq m} d_{ij} d_{il} E(Y^{*2}_{im}|\mathbf{X}) \times I(j, m \in W_{ir_1}) \times I(l, m \in W_{ir_2}) \\
& \left(\because \; E(Y^*_{i_1j_2}Y^*_{i_2l_2}|\mathbf{X}) = 0 \text{ unless } i_1 = i_2 \text{ and } j_2 = l_2 = m; \; \gamma_{ir} = O(N^{-1}w^{-1})\right) \\
= \; & O\left(N^{-1}w^{-3}\right) \sum_{i=1}^{k} 4 \sum_{m=1}^{n_i - 1}\sum_{s_1=1}^{w_{im}-1}\sum_{s_2=1}^{w_{im}-1} d_{i,m+s_1} d_{i,m+s_2} E(Y^{*2}_{im}|\mathbf{X}) \\
& \quad \times \sum_{r_1=1}^{n_i} I(m, m+s_1 \in W_{ir_1}) \cdot \gamma_{ir_1} \times \sum_{r_2=1}^{n_i} I(m, m+s_2 \in W_{ir_2}) \cdot \gamma_{ir_2}.
\end{aligned}$$

Since $d_{ij} = O((n_i w)^{-1/4})$, uniformly a.s.; by Assumption 2, $E(Y^{*2}_{ij}|\mathbf{X}) = O(1)$, u.a.s.; and $\forall m = 1, \cdots, n_i - 1$,

$$\sum_{r=1}^{n_i} I(m, m+s \in W_{ir}) \leq w - s = O(w), \; \forall s = 1, \cdots, w-1,$$

we have

$$E\left(R_1^2|\mathbf{X}\right) = O\left(N^{-1}w^{-3}\right) \cdot O\left(n_i \cdot w^2 \cdot (n_i w)^{-1/2} \cdot w^2\right) = O(N^{-1/2}w^{1/2}), \; u.a.s.$$

Then, by DCT, $E(R_1) = E[E(R_1|\mathbf{X})] = 0$ and $E(R_1^2) = E[E(R_1^2|\mathbf{X})] = o(1)$, which completes the proof of $R_1 = o_p(1)$.

Similarly, to prove (5.5.4), note that

$$\begin{aligned} d(\mathbf{X})'\mathbf{A}^*\mathbf{Z}^* &= -\sum_{i=1}^{k}\beta_i \sum_{r_1\neq r_2}^{n_i}\sum_{t_1=1}^{w_{ir_1}}\sum_{t_2=1}^{w_{ir_2}} d_{ir_1t_1}Z^*_{ir_2t_2} \\ &= -\sum_{i=1}^{k}\beta_i \sum_{r_1\neq r_2}\sum_{j_1=1}^{n_i}\sum_{j_2=1}^{n_i} d_{ij_1}Y^*_{ij_2}\times I(j_1\in W_{ir_1})\times I(j_2\in W_{ir_2}). \end{aligned}$$

Since $E(R_2|\mathbf{X})=0$ and

$$E(R_2^2|\mathbf{X}) = \frac{N}{w}E\left[\left(d(\mathbf{X})'\mathbf{A}^*\mathbf{Z}^*\right)^2|\mathbf{X}\right]$$

$$\begin{aligned} &= \frac{N}{w}\sum_{i_1}\sum_{i_2}\beta_{i_1}\beta_{i_2}\sum_{r_1\neq r_2}\sum_{r_3\neq r_4}\sum_{j_1,\ j_2}\sum_{l_1,\ l_2} d_{i_1j_1}d_{i_2l_1}E(Y^*_{i_1j_2}Y^*_{i_2l_2}|\mathbf{X}) \\ &\qquad \times I(j_1\in W_{i_1r_1})\times I(j_2\in W_{i_1r_2})\times I(l_1\in W_{i_2r_3})\times I(l_2\in W_{i_2r_4}) \\ &= \frac{N}{w}\sum_{i=1}^{k}\beta_i^2\sum_{r_1\neq r_2}\sum_{r_3\neq r_4}\sum_{j=1}^{n_i}\sum_{l=1}^{n_i}\sum_{m=1}^{n_i} d_{ij}d_{il}E(Y^{*2}_{im}|\mathbf{X}) \\ &\qquad \times I(j\in W_{ir_1})\times I(l\in W_{ir_3})\times I(m\in W_{ir_2}\cap W_{ir_4}) \\ &\leq \frac{N}{w}\sum_{i=1}^{k}\beta_i^2\left(\sum_j d_{ij}\right)^2\cdot\left(\sum_m E(Y^{*2}_{im}|\mathbf{X})\right)\times O(w^4) \\ &= Nw^{-1}O((N^{-2}w^{-1})^2\cdot(n_i^{3/4}w^{-1/4})^2\cdot n_i\cdot w^4) = O(N^{-1/2}w^{1/2}),\ \ u.a.s. \end{aligned}$$

$$\left(\because \beta_i = O(N^{-2}w^{-1});\ \sum_j d_{ij} = O(n_i\cdot(n_iw)^{-1/4}),\ \ u.a.s.;\ E(Y^{*2}_{im}|\mathbf{X}) = O(1),\ \ u.a.s.\right)$$

by DCT, $E(R_2) = E[E(R_2|\mathbf{X})] = 0$ and $E(R_2^2) = E[E(R_2^2|\mathbf{X})] = o(1)$, so $R_2 = o_p(1)$, which completes the proof.

Proof of Lemmas 5.2.2 and 5.2.6

Since $\mathbf{A}^* = \mathbf{A} - \mathbf{A}_D$,

$$\mathbf{Z}^{*\prime}\mathbf{A}^*\mathbf{Z}^* = -\sum_{i=1}^{k}\beta_i\cdot\left(\sum_{r_1=1}^{n_i}\sum_{r_2\neq r_1}^{n_i}\sum_{t_1=1}^{w_{ir_1}}\sum_{t_2=1}^{w_{ir_2}} Z^*_{ir_1t_1}Z^*_{ir_2t_2}\right)$$

$$= \quad -\sum_{i=1}^{k} \beta_i \left(\sum_{r_1=1}^{n_i} \sum_{r_2 \neq r_1}^{n_i} \sum_{j_1=1}^{n_i} \sum_{j_2=1}^{n_i} Y_{ij_1}^* Y_{ij_2}^* I(j_1 \in W_{ir_1}) I(j_2 \in W_{ir_2}) \right)$$

Let $Q_1 = N^{1/2} w^{-1/2} \mathbf{Z}^{*\prime} \mathbf{A}^* \mathbf{Z}^*$. To show $Q_1 = o_p(1)$, first note that, conditionally on $\mathbf{X}$, Y_{ij_1} and Yij_2 are independent if $j_1 \neq j_2$, and hence $E(Y_{ij_1}^* Y_{ij_2}^* | \mathbf{X}) \neq 0$ only if $j_1 = j_2$. This tells us that

$$\begin{aligned}
E(Q_1|\mathbf{X}) &= -\left(\frac{N}{w}\right)^{\frac{1}{2}} \sum_{i=1}^{k} \beta_i \sum_{j=1}^{n_i} E\left(Y_{ij}^{*2} | \mathbf{X}\right) \sum_{r_1=1}^{n_i} \sum_{r_2 \neq r_1}^{n_i} I(j \in W_{ir_1} \cap W_{ir_2}) \\
&= N^{1/2} w^{-1/2} \cdot O(N^{-2} w^{-1}) \cdot O(n_i \cdot w^2), \text{ uniformly a.s.} \\
&\quad \left(\because \beta_i = O(N^{-2} w^{-1});\ E(Y_{ij}^{*2}|\mathbf{X}) = O(1),\ u.a.s., \text{by A2;}\right) \\
&\quad \left(\because \sum_{r_1 \neq r_2} I(j \in W_{ir_1} \cap W_{ir_2}) = w_{ij}(w_{ij} - 1) \leq w^2.\right) \\
&= O(N^{-1/2} w^{1/2}),\ u.a.s.
\end{aligned}$$

In addition,

$$\begin{aligned}
E(Q_1^2|\mathbf{X}) &= E\left[(N^{1/2} w^{-1/2} \mathbf{Z}^{*\prime} \mathbf{A}^* \mathbf{Z}^*)^2 | \mathbf{X}\right] \\
&= N w^{-1} \sum_{i=1}^{k} \sum_{r_1 \neq r_2}^{n_i} \sum_{r_3 \neq r_4}^{n_i} \sum_{j_1,\ j_2=1}^{n_i} \sum_{j_3,\ j_4=1}^{n_i} \beta_i^2 E\left(Y_{ij_1}^* Y_{ij_2}^* Y_{ij_3}^* Y_{ij_4}^* | \mathbf{X}\right) \\
&\qquad \times I(j_1 \in W_{ir_1},\ j_2 \in W_{ir_2},\ j_3 \in W_{ir_3},\ j_4 \in W_{ir_4}) \\
&+ N w^{-1} \sum_{i_1 \neq i_2}^{k} \sum_{r_1 \neq r_2}^{n_{i_1}} \sum_{r_3 \neq r_4}^{n_{i_2}} \sum_{j_1,\ j_2=1}^{n_{i_1}} \sum_{j_3,\ j_4=1}^{n_{i_2}} \beta_{i_1} \beta_{i_2} E\left(Y_{i_1 j_1}^* Y_{i_1 j_2}^* | \mathbf{X}\right)\ E\left(Y_{i_2 j_3}^* Y_{i_2 j_4}^* | \mathbf{X}\right) \\
&\qquad \times I(j_1 \in W_{i_1 r_1},\ j_2 \in W_{i_1 r_2},\ j_3 \in W_{i_2 r_3},\ j_4 \in W_{i_2 r_4}),
\end{aligned}$$

and because (1) $E(Y_{ij_1}^* Y_{ij_2}^* Y_{ij_3}^* Y_{ij_4}^* | \mathbf{X}) \neq 0$ only if $j_1 = j_2 = j_3 = j_4$ or $(j_1,\ j_2,\ j_3,\ j_4)$ are two pairs of equal indices; (2) $E(Y_{i_1 j_1}^* Y_{i_1 j_2}^* | \mathbf{X}) \cdot E(Y_{i_2 j_3}^* Y_{i_2 j_4}^* | \mathbf{X}) \neq 0$ only if $j_1 = j_2$ and $j_3 = j_4$; (3) $\sum_{r=1}^{n_i} I(j \in W_{ir}) = w_{ij} \leq w,\ \forall i,\ j$, we know

$$E(Q_1^2|\mathbf{X}) \quad \leq \quad N w^{-1} \sum_{i=1}^{k} \beta_i^2 \left[\sum_{l=1}^{n_i} E(Y_{il}^{*4}|\mathbf{X}) + \frac{4!}{2!\,2!} \sum_{l_1 < l_2}^{n_i} E(Y_{il_1}^{*2}|\mathbf{X}) E(Y_{il_2}^{*2}|\mathbf{X}) \right] \cdot w^4$$

$$+ Nw^{-1}\sum_{i_1\neq i_2}^{k}\beta_{i_1}\beta_{i_2}\left[\sum_{j_1=j_2}^{n_{i_1}}\sum_{j_3=j_4}^{n_{i_2}}E(Y^{*2}_{i_1j_1}|\mathbf{X})E(Y^{*2}_{i_2j_3}|\mathbf{X})\right]\cdot w^4$$

$$= Nw^{-1}O(N^{-4}w^{-2})\{[O(n_i)+O(n_i^2)]+O(n_i^2)\}\cdot w^4,\ u.a.s.\,(\text{by A2})$$

$$= O(N^{-1}w),\ u.a.s.$$

Then, by DCT, $E(Q_1) = E[E(Q_1|\mathbf{X})] = o(1)$ and $E(Q_1^2) = E[E(Q_1^2|\mathbf{X})] = o(1)$, which completes the proof.

Proof of Lemmas 5.2.3 and 5.2.7

First note that

$$\mathbf{Z}^{*\prime}\left(\mathbf{A}_D - \mathbf{A}_D^*\right)\mathbf{Z}^* = \sum_{i=1}^{k}\sum_{r=1}^{n_i}(\gamma_{ir}-\kappa_i)\sum_{j_1=1}^{n_i}\sum_{j_2\neq j_1}^{n_i}Y^*_{ij_1}Y^*_{ij_2}I(j_1,j_2\in W_{ir}),\ \text{where}$$

$$\gamma_{ir} = \frac{1}{(N-k)(w_{ir}-1)}\left(1-\frac{w_{ir}}{w_{i\cdot}}\right),\quad \text{and}\quad \kappa_i = \frac{1}{(N-k)(w-1)}\left(1-\frac{1}{n_i}\right).$$

Let $Q_2 = N^{1/2}w^{-1/2}\mathbf{Z}^{*\prime}\left(\mathbf{A}_D - \mathbf{A}_D^*\right)\mathbf{Z}^*$. since $E(Y^*_{ij_1}Y^*_{ij_2}\times I(j_1,j_2\in W_{ir})|\mathbf{X}),\ \forall j_1\neq j_2$, it's clear that $E(Q_2|\mathbf{X}) = 0$. In addition,

$$E\left\{Q_2^2 \mid \mathbf{X}\right\}$$

$$= \frac{N}{w}\sum_{i=1}^{k}\sum_{r_1=1}^{n_i}\sum_{r_2=1}^{n_i}(\gamma_{ir_1}-\kappa_i)(\gamma_{ir_2}-\kappa_i)\sum_{j_1\neq j_2}\sum_{l_1\neq l_2}E\left(Y^*_{ij_1}Y^*_{ij_2}Y^*_{il_1}Y^*_{il_2}\mid\mathbf{X}\right)$$
$$\times I(j_1,j_2\in W_{ir_1})\times I(l_1,l_2\in W_{ir_2})$$

$$= \frac{N}{w}\sum_{i=1}^{k}\sum_{r_1=1}^{n_i}\sum_{r_2=1}^{n_i}(\gamma_{ir_1}-\kappa_i)(\gamma_{ir_2}-\kappa_i)\sum_{j_1\neq j_2}E\left(Y^{*2}_{ij_1}Y^{*2}_{ij_2}\mid\mathbf{X}\right)\times I(j_1,j_2\in W_{ir_1}\cap W_{ir_2})$$

$$= \frac{N}{w}\sum_{i=1}^{k}\sum_{r_1=1}^{n_i}\sum_{r_2=1}^{n_i}(\gamma_{ir_1}-\kappa_i)(\gamma_{ir_2}-\kappa_i)\sum_{j_1=1}^{n_i-1}\sum_{j_2>j_1}O(1)\times I(j_1,j_2\in W_{ir_1}\cap W_{ir_2}),\ u.a.s.$$
$$\left(\because\ \text{By A2},\ E\left(Y^{*2}_{ij_1}Y^{*2}_{ij_2}\mid\mathbf{X}\right) = O(1),\ \text{uniformly a.s.},\forall j_1\neq j_2.\right)$$

$$= \frac{N}{w}\sum_{i=1}^{k}\sum_{j=1}^{n_i-1}\sum_{s=1}^{w_{ij}-1}\sum_{r_1,\ r_2=\max(1,\ j+s-\frac{w-1}{2})}^{\min(j+\frac{w-1}{2},\ n_i)}(\gamma_{ir_1}-\kappa_i)(\gamma_{ir_2}-\kappa_i)\times O(1),\ u.a.s.$$

Recall that $w_{i\cdot} = n_i w - (w^2-1)/4$. It can be easily verified that, if $1+(w-1)/2 \le r \le n_i - (w-1)/2$ ($\because w_{ir} = w$),

$$\begin{aligned}\gamma_{ir} - \kappa_i &= \frac{1}{(N-k)(w-1)}\left(1 - \frac{w}{w_{i\cdot}} - 1 + \frac{1}{n_i}\right) \\ &= \frac{1}{(N-k)(w-1)} \times \frac{n_i w - \frac{w^2-1}{4} - n_i w}{n_i w_{i\cdot}} \\ &= O\left(\frac{w^2}{NwNNw}\right) = O(N^{-3});\end{aligned}$$

otherwise,

$$\begin{aligned}\gamma_{ir} - \kappa_i &\le \frac{1}{(N-k)(w_{ir}-1)}\left(1 - \frac{w_{ir}}{w_{i\cdot}} - 1 + \frac{1}{n_i}\right) = \frac{w_{i\cdot} - n_i w_{ir}}{(N-k)(w_{ir}-1)n_i w_{i\cdot}} \\ &\le \frac{n_i(w - w_{ir})}{(N-k)(w_{ir}-1)n_i w_{i\cdot}} = O\left(\frac{w}{Nwn_i w}\right) \\ &= O(N^{-2}w^{-1}).\end{aligned}$$

Hence,

$$\begin{aligned}& E\left(Q_2^2 \mid \mathbf{X}\right) \\ =\ & \frac{N}{w}\sum_{i=1}^{k}\sum_{j=w}^{n_i-w}\sum_{s=1}^{w-1}\sum_{r_1,\ r_2=j+s-\frac{w-1}{2}}^{j+\frac{w-1}{2}}\left[O(N^{-3})\right]^2 \\ +\ & \frac{N}{w}\sum_{i=1}^{k}\left(\sum_{j=1}^{w-1} + \sum_{j=n_i-w+1}^{n_i-1}\right)\sum_{s=1}^{w_{ij}-1}\sum_{r_1,\ r_2=\max(1,\ j+s-\frac{w-1}{2})}^{\min(j+\frac{w-1}{2},\ n_i)}\left[O(N^{-2}w^{-1})\right]^2,\ u.a.s. \\ =\ & O(N^{-4}w^2) + O(N^{-3}w) = O(N^{-3}w),\ u.a.s.\end{aligned}$$

By DCT, $E(Q_2) = E[E(Q_2|\mathbf{X})] = 0$ and $E(Q_2^2) = E[E(Q_2^2|\mathbf{X})] = o(1)$, so $Q_2 = o_p(1)$.

Proof of Lemmas 5.2.4 and 5.2.8

First note that $E(Y_{ij}^* \mid \mathbf{X}) = 0 \ \Rightarrow\ E[\mathbf{Z}^{*\prime}\mathbf{A}_D^*\mathbf{Z}^* \mid \mathbf{X}] = 0$, and

$$\sigma_i^2(X_{ij}) = Var(Y_{ij}|X_{ij}) = E\{[Y_{ij} - E(Y_{ij}|X_{ij})]^2|X_{ij}\} = E(Y_{ij}^{*2}|X_{ij})$$

$$= \quad Var(Y^*_{ij}|X_{ij})$$

In addition,

$$\begin{aligned}
& E[(\mathbf{Z}^{*\prime}\mathbf{A}^*_D\mathbf{Z}^*)^2|\mathbf{X}] \\
= \; & \sum_{i=1}^{k}\kappa_i^2\sum_{r_1=1}^{n_i}\sum_{r_2=1}^{n_i}\sum_{j_1\neq j_2}^{n_i}\sum_{l_1\neq l_2}^{n_i} E(Y^*_{ij_1}Y^*_{ij_2}Y^*_{il_1}Y^*_{il_2}|\mathbf{X})I(j_1,j_2\in W_{ir_1})I(l_1,l_2\in W_{ir_2}) \\
= \; & \sum_{i=1}^{k}\kappa_i^2\cdot 4\sum_{r_1=1}^{n_i}\sum_{r_2=1}^{n_i}\sum_{j_1=l_1<j_2=l_2} E(Y^{*2}_{ij_1}Y^{*2}_{ij_2}|\mathbf{X})\times I(j_1,j_2\in W_{ir_1}\cap W_{ir_2}) \\
= \; & \sum_{i=1}^{k}\kappa_i^2\cdot 4\sum_{r_1=1}^{n_i}\sum_{r_2=1}^{n_i}\sum_{j_1<j_2}\sigma_i^2(X_{ij_1})\cdot\sigma_i^2(X_{ij_2})\times I(j_1,j_2\in W_{ir_1}\cap W_{ir_2}).
\end{aligned}$$

From the proof of Lemma 5.5.1, we learn that for all n_i large enough,

$$X^U_{ir} - X^L_{ir} = O\left(n_i^{-1+\delta} + \frac{w}{n_i}\right), \text{ uniformly a.s..}$$

Since, by Assumption, $\sigma_i^2(x)$ are uniformly Lipschitz continuous in x, we know that $\forall j_1,\ j_2 \in W_{ir}$, there exists a positive constant C such that, for n_i large enough,

$$\begin{aligned}
|\sigma_i^2(X_{ij_1}) - \sigma_i^2(X_{ij_1})| \quad & \leq \quad C|X_{ij_1} - X_{ij_2}| \\
& \leq \quad C(X^U_{ir} - X^L_{ir}) = O\left(n_i^{-1+\delta} + \frac{w}{n_i}\right),\ u.a.s.,\ \forall\delta > 0,
\end{aligned}$$

implying that

$$\sigma_i^2(X_{ij_2}) = \sigma_i^2(X_{ij_1}) + O\left(n_i^{-1+\delta} + \frac{w}{n_i}\right),\ u.a.s.$$

So,

$$\begin{aligned}
& E[(\mathbf{Z}^{*\prime}\mathbf{A}^*_D\mathbf{Z}^*)^2|\mathbf{X}] \\
= \; & \sum_{i=1}^{k}\kappa_i^2\cdot 4\sum_{r_1=1}^{n_i}\sum_{r_2=1}^{n_i}\sum_{j_1<j_2}\sigma_i^2(X_{ij_1})\left[\sigma_i^2(X_{ij_1}) + O\left(n_i^{-1+\delta} + \frac{w}{n_i}\right)\right] \\
& \times I(j_1,j_2\in W_{ir_1}\cap W_{ir_2}),\ u.a.s.
\end{aligned}$$

$$= \quad 4 \sum_{i=1}^{k} \kappa_i^2 \sum_{j=1}^{n_i-1} \left[\sigma_i^4(X_{ij}) + O\left(n_i^{-1+\delta} + \frac{w}{n_i} \right) \right]$$

$$\times \sum_{s=1}^{w_{ij}-1} \sum_{r_1=1}^{n_i} \sum_{r_2=1}^{n_i} I(j,\ j+s \in W_{ir_1} \cap W_{ir_2}),\ u.a.s.$$

Note that, if $1 + \frac{w-1}{2} \le j \le n_i - \frac{w-1}{2} \ \Rightarrow\ w_{ij} = w$,

$$\sum_{s=1}^{w-1} \sum_{r_1=1}^{n_i} \sum_{r_2=1}^{n_i} I(j, j+s \in W_{ir_1} \cap W_{ir_2}) \quad = \quad \sum_{s=1}^{w-1} \sum_{r_1, r_2 = j+s-\frac{w-1}{2}}^{j+\frac{w-1}{2}} I(j, j+s \in W_{ir_1} \cap W_{ir_2})$$

$$= \quad \sum_{s=1}^{w-1} (w-s)^2 \quad = \quad \frac{w(w-1)(2w-1)}{6};$$

otherwise, $\sum_s \sum_{r_1} \sum_{r_2} I(j, j+s \in W_{ir_1} \cap W_{ir_2}) = O(w^3)$.

This tells us that

$$E[(\mathbf{Z}^{*\prime} \mathbf{A}_D^* \mathbf{Z}^*)^2 | \mathbf{X}]$$

$$= \quad 4 \sum_{i=1}^{k} \kappa_i^2 \sum_{j=1+\frac{w-1}{2}}^{n_i - \frac{w-1}{2}} \left[\sigma_i^4(X_{ij}) + O\left(n_i^{-1+\delta} + \frac{w}{n_i} \right) \right] \times \frac{w(w-1)(2w-1)}{6}$$

$$+ \sum_{i=1}^{k} O\left(N^{-2} w^{-2} \right) \cdot \left(\sum_{j=1}^{\frac{w-1}{2}} + \sum_{j=n_i-\frac{w-1}{2}+1}^{n_i-1} \right) \cdot O\left(w^3 \right),\ u.a.s.$$

$$= \quad \sum_{i=1}^{k} \frac{4}{(N-k)^2 (w-1)^2} \left(1 - \frac{1}{n_i} \right)^2 \frac{w(w-1)(2w-1)}{6} \sum_{j=1+\frac{w-1}{2}}^{n_i-\frac{w-1}{2}} \sigma_i^4(X_{ij})$$

$$+ O\left(N^{-2} w^{-2} n_i \cdot w^3 \cdot \left(n_i^{-1+\delta} + \frac{w}{n_i} \right) \right) + O\left(N^{-2} w^2 \right),\ u.a.s.$$

$$= \quad \frac{w}{N-k} \cdot \frac{2(2W-1)}{3(w-1)} \sum_{i=1}^{k} \left(1 - \frac{1}{n_i} \right)^2 \cdot \frac{n_i}{N-k} \left[\frac{1}{n_i} \sum_{j=1+\frac{w-1}{2}}^{n_i-\frac{w-1}{2}} \sigma_i^4(X_{ij}) \right]$$

$$+ O\left(N^{-2+\delta} w + N^{-2} w^2 \right),\ u.a.s.$$

By WLLN,

$$\frac{1}{n_i}\sum_{j=1+\frac{w-1}{2}}^{n_i-\frac{w-1}{2}} \sigma_i^4(X_{ij}) = E[\sigma_i^4(X)] + o_p(1).$$

Therefore,

$$\begin{aligned} E\left[\left(N^{1/2}w^{-1/2}\mathbf{Z}^{*\prime}\mathbf{A}_D\mathbf{Z}^*\right)^2 \mid \mathbf{X}\right] &= \frac{2(2w-1)}{3(w-1)}\sum_{i=1}^{k}\lambda_i E[\sigma_i^4(X)] + o_p(1) \\ &\qquad + O\left(N^{-1+\delta} + N^{-1}w\right),\ u.a.s. \end{aligned}$$

By DCT,

$$\begin{aligned} & Var(N^{1/2}w^{-1/2}\mathbf{Z}^{*\prime}\mathbf{A}_D\mathbf{Z}^*) \\ = \;& Var[E(N^{1/2}w^{-1/2}\mathbf{Z}^{*\prime}\mathbf{A}_D\mathbf{Z}^*|\mathbf{X})] + E[Var(N^{1/2}w^{-1/2}\mathbf{Z}^{*\prime}\mathbf{A}_D\mathbf{Z}^*|\mathbf{X})] \\ = \;& E\{E[(N^{1/2}w^{-1/2}\mathbf{Z}^{*\prime}\mathbf{A}_D\mathbf{Z}^*)^2|\mathbf{X}]\} = \frac{2(2w-1)}{3(w-1)}\sum_{i=1}^{k}\lambda_i E[\sigma_i^4(X)] + o(1), \end{aligned}$$

which completes the proof.

Proof of Theorems 5.2.1 and 5.2.2

The following is in fact the proof of Theorem 5.2.2, but the proof of Theorem 5.2.1 should be similar (and in act easier).

Following Lemmas 5.2.5–5.2.8, it remains to verify the asymptotic normality of $N^{1/2}w^{-1/2}\mathbf{Z}^{*\prime}\mathbf{A}_D^*\mathbf{Z}^*$ under $H_a(D)$. Write $N^{1/2}w^{-1/2}\mathbf{Z}^{*\prime}\mathbf{A}_D^*\mathbf{Z}^*$ as follows:

$$\begin{aligned} & \left(\frac{N}{w}\right)^{1/2}\mathbf{Z}^{*\prime}\mathbf{A}_D^*\mathbf{Z}^* \\ = \;& \left(\frac{N}{w}\right)^{1/2}\sum_{i=1}^{k}\frac{1}{(N-k)(w-1)}\left(1-\frac{1}{n_i}\right)\sum_{r=1}^{n_i}\sum_{j_1}^{n_i}\sum_{j_2\neq j_1} Y_{ij_1}^* Y_{ij_2}^* I(j_1, j_2 \in W_{ir}) \\ = \;& \sqrt{\frac{N}{N-k}}\sum_{i=1}^{k}\left(1-\frac{1}{n_i}\right)\sqrt{\frac{n_i}{N-k}}\left[\frac{1}{\sqrt{n_i w}}\sum_{r=1}^{n_i} T_{i,r}\right] \end{aligned}$$

$$\text{where } T_{i,r} = \frac{1}{w-1} \sum_{j_1 \neq j_2} Y^*_{ij_1} Y^*_{ij_2} I(j_1, j_2 \in W_{ir}).$$

We now first prove the asymptotic normality of $N^{1/2}w^{-1/2}\mathbf{Z}^{*\prime}\mathbf{A}^*_D\mathbf{Z}^*$, conditionally on $\mathbf{X}$. Due to the independence among different groups, it's sufficient to prove the asymptotic normality of $(n_i w)^{-1/2}\sum_{r=1}^{n_i} T_{i,r}$, $\forall i$, conditionally on $\mathbf{X}$.

Define $t_{n_i} = \lfloor n_i^{2/3} \rfloor$, $s_{n_i} = \lfloor n_i/(t_{n_i}+w) \rfloor$, $l_{n_i} = s_{n_i}(t_{n_i}+w)$, where $\lfloor a_n \rfloor$ denotes the largest integer which is not greater than a_n. Further define, $\forall r = 1, \cdots, s_{n_i}$,

$$\begin{aligned} U_{ir} &= T_{i,(r-1)(t_{n_i}+w)+1} + \cdots + T_{i,r(t_{n_i}+w)-w}, \\ V_{ir} &= T_{i,r(t_{n_i}+w)-w+1} + \cdots + T_{i,r(t_{n_i}+w)}. \end{aligned}$$

Note that there are t_{n_i} terms in each U_{ir} while only w terms in each V_{ir}. Hence, for all n_i large enough, conditionally on $\mathbf{X}$, $\{U_{ir}; r = 1, \cdots, s_{n_i}\}$ are independent; so do $\{V_{ir}\}$. Now, decompose $(n_i w)^{-1/2}\sum_{r=1}^{n_i} T_{i,r}$ as three parts:

$$(n_i w)^{-1/2} \sum_{r=1}^{n_i} T_{i,r} \equiv Q_{i1} + Q_{i2} + Q_{i3},$$

where

$$Q_{i1} = (n_i w)^{-1/2} \sum_{r=1}^{s_{n_i}} U_{ir}, \quad Q_{i2} = (n_i w)^{-1/2} \sum_{r=1}^{s_{n_i}} V_{ir}, \quad Q_{i3} = (n_i w)^{-1/2} \sum_{r=l_{n_i}+1}^{n_i} T_{i,r}.$$

Our plan here is to show that, conditionally on $\mathbf{X}$, both of Q_{i2} and Q_{i3} are $o_p(1)$, and Q_{i1} converges in distribution to normal as $\min(n_i) \to \infty$. Then, Slusky's theorem completes the proof.

First note that

$$\begin{aligned} & E\left(T_{i,r} \mid \mathbf{X}\right) = \frac{1}{w-1} \sum_{j_1 \neq j_2} E\left(Y^*_{ij_1} Y^*_{ij_2} \mid \mathbf{X}\right) \cdot I(j_1,\ j_2 \in W_{ir}) = 0,\ \forall i,\ \forall r \\ & \Longrightarrow E(Q_{i1}|\mathbf{X}) = E(Q_{i2}|\mathbf{X}) = E(Q_{i3}|\mathbf{X}) = 0. \end{aligned}$$

Next, consider $U_{i1} = \sum_{r=1}^{t_{n_i}} T_{i,r} \;\Rightarrow\; E(U_{i1}|\mathbf{X}) = 0$ and

$$\begin{aligned}
E(U_{i1}^2|\mathbf{X}) &= O(w^{-2}) \sum_{r_1=1}^{t_{n_i}} \sum_{r_2=1}^{t_{n_i}} \sum_{j_1 \neq j_2} \sum_{l_1 \neq l_2} E\left(Y_{ij_1}^* Y_{ij_2}^* Y_{il_1}^* Y_{il_2}^* |\mathbf{X}\right) \\
&\quad \times I(j_1, j_2 \in W_{ir_1}) \times I(l_1, l_2 \in W_{ir_2}) \\
&= O(w^{-2}) \sum_{r_1=1}^{t_{n_i}} \sum_{r_2=1}^{t_{n_i}} 4 \sum_{j_1<j_2} E(Y_{ij_1}^{*2} Y_{ij_2}^{*2}|\mathbf{X}) \times I(j_1,\ j_2 \in W_{ir_1} \cap W_{ir_2}) \\
&= O(w^{-2}) \sum_{j=1}^{t_{n_i}^*} \sum_{s=1}^{w_{ij}-1} E(Y_{ij}^{*2} Y_{i,j+s}^{*2}|\mathbf{X}) \sum_{r_1=1}^{t_{n_i}} \sum_{r_2=1}^{t_{n_i}} I(j,\ j+s \in W_{ir_1} \cap W_{ir_2}) \\
&\quad \text{where } t_{n_i}^* = t_{n_i} + (w-1)/2 - 1 \\
&= O(w^{-2} \cdot t_{n_i} w \cdot w^2) \times O(1),\ u.a.s \\
&\left(\because E(Y_{ij}^*|\mathbf{X}) = O(1),\ u.a.s.,\ \text{by A2};\ \sum_{r_1} \sum_{r_2} I(j,\ j+s \in W_{ir_1} \cap W_{ir_2}) = O(w^2)\right) \\
&= O(t_{n_i} w),\ u.a.s.
\end{aligned}$$

Similarly, one can show that $\forall r = 1, \cdots, s_{n_i}$,

$$E(U_{ir}^2|\mathbf{X}) = O(t_{n_i} \cdot w) = O(n_i^{2/3} \cdot w),\ u.a.s.; \tag{5.5.5}$$

$$E(V_{ir}^2|\mathbf{X}) = O(w^2),\ u.a.s.; \tag{5.5.6}$$

$$E\left[\left(\sum_{r=l_{n_i}+1}^{n_i} T_{i,r}\right)^2 |\mathbf{X}\right] \leq O((t_{n_i} + w) \cdot w) = O(n_i^{2/3} \cdot w),\ u.a.s. \tag{5.5.7}$$

Given that $E(Q_{i2}|\mathbf{X}) = E(Q_{i3}|\mathbf{X}) = 0$, (5.5.6) and (5.5.7) tell us that

$$\begin{aligned}
Var(Q_{i2}|\mathbf{X}) &= (n_i w)^{-1} \sum_{r=1}^{s_{n_i}} Var(V_{ir}|\mathbf{X}) = (n_i w)^{-1} \sum_{r=1}^{s_{n_i}} E(V_{ir}^2|\mathbf{X}) \\
&= (n_i w)^{-1} O\left(s_{n_i} \cdot w^2\right),\ u.a.s. \ = \ o(1),\ u.a.s.
\end{aligned}$$

and

$$\begin{aligned} Var(Q_{i3}|\mathbf{X}) &= E(Q_{i3}^2|\mathbf{X}) = (n_i w)^{-1} E\left[\left(\sum_{r=l_{n_i}+1}^{n_i} T_{i,r}\right)^2 |\mathbf{X}\right] \\ &= (n_i w)^{-1} \times O\left((t_{n_i}+w)\cdot w\right),\ u.a.s.\ =\ o(1),\ u.a.s. \end{aligned}$$

Therefore, it's clear that, conditionally on $\mathbf{X}$, both Q_{i2} and Q_{i3} are $o_p(1)$, uniformly a.s.

Now, we verify the asymptotic normality of Q_{i1}, conditionally on $\mathbf{X}$. Since conditionally on $\mathbf{X}$, $U_{i1}, \cdots, U_{in_i}$ are independent, it is sufficient to check the Lyapounov's condition:

$$\mathrm{L}_n^{\delta=2} = \frac{\sum_{r=1}^{s_{n_i}} E(U_{ir}^4|\mathbf{X})}{\left[\sum_{r=1}^{s_{n_i}} E(U_{ir}^2|\mathbf{X})\right]^2}.$$

Note that

$$\begin{aligned} & E(U_{i1}^4|\mathbf{X}) \\ = \ & O(w^{-4}) \sum_{r_1}\sum_{r_2}\sum_{r_3}\sum_{r_4}\sum_{a_1\neq a_2}\sum_{b_1\neq b_2}\sum_{c_1\neq c_3}\sum_{d_1\neq d_2} E\left(Y_{ia_1}^* Y_{ia_2}^* Y_{ib_1}^* Y_{ib_2}^* Y_{ic_1}^* Y_{ic_2}^* Y_{id_1}^* Y_{id_2}^*|\mathbf{X}\right) \\ & \times I(a_1, a_2 \in W_{ir_1}) \cdot I(b_1, b_2 \in W_{ir_2}) \cdot I(c_1, c_2 \in W_{ir_3}) \cdot I(d_1, d_2 \in W_{ir_4}). \end{aligned}$$

Since the nonzero expected terms in the above equation must be one of the following forms: $E(Y_{ia}^{*4} Y_{ib}^{*4}|\mathbf{X})$, $E(Y_{ia}^{4*} Y_{ib}^{2*} Y_{ic}^{2*}|\mathbf{X})$, $E(Y_{ia}^{3*} Y_{ib}^{3*} Y_{ic}^{2*}|\mathbf{X})$, or $E(Y_{ia}^{2*} Y_{ib}^{2*} Y_{ic}^{2*} Y_{id}^{2*}|\mathbf{X})$, where $a \neq b \neq c \neq d$, we have

$$\begin{aligned} & E(U_{i1}^4|\mathbf{X}) \\ = \ & O(w^{-4}) \sum_{a=1}^{t_{n_i}^*}\sum_{s=1}^{w_{ia}-1} E(Y_{ia}^{*4} Y_{i,a+s}^{*4}|\mathbf{X}) \times O(w^4) \\ & + O(w^{-4}) \sum_{a=1}^{t_{n_i}^*}\sum_{s_2=1}^{w_{ia}-1}\sum_{s_1<s_2} E(Y_{ia}^{*4} Y_{i,a+s_1}^{*2} Y_{i,a+s_2}^{*2}|\mathbf{X}) \times O(w^4) \\ & + O(w^{-4}) \sum_{a=1}^{t_{n_i}^*}\sum_{s_1\neq s_2}^{w_{ia}-1} E(Y_{ia}^{*2} Y_{i,a+s_1}^{*4} Y_{i,a+s_2}^{*2}|\mathbf{X}) \times O(w^4) \end{aligned}$$

$$+ O(w^{-4}) \sum_{a=1}^{t_{n_i}^*} \sum_{s_1 \neq s_2}^{w_{ia}-1} E(Y_{ia}^{*3} Y_{i,a+s_1}^{*3} Y_{i,a+s_2}^{*2} | \mathbf{X}) \times O(w^4)$$

$$+ O(w^{-4}) \sum_{a=1}^{t_{n_i}^*} \sum_{s_1 \neq s_2}^{w_{ia}-1} E(Y_{ia}^{*2} Y_{i,a+s_1}^{*3} Y_{i,a+s_2}^{*3} | \mathbf{X}) \times O(w^4)$$

$$+ O(w^{-4}) \left[\sum_{a=1}^{t_{n_i}^*} \sum_{s_1=1}^{w_{ia}-1} E(Y_{ia}^{*2} Y_{i,a+s_1}^{*2} | \mathbf{X}) \right] \left[\sum_{b \neq a}^{t_{n_i}^*} \sum_{s_2=1}^{w_{ib}-1} E(Y_{ib}^{*2} Y_{i,b+s_2}^{*2} | \mathbf{X}) \right] \times O(w^4)$$

where $t_{n_i}^* = t_{n_i} + (w-1)/(2) - 1$

$$= O(t_{n_i} \cdot w) + O(t_{n_i} \cdot w^2) + O(t_{n_i} \cdot w^2) + O(t_{n_i}^2 \cdot w^2), \ u.a.s.$$

$$= O(t_{n_i}^2 \cdot w^2), \ u.a.s.$$

Similarly,

$$E(U_{ir}^4 | \mathbf{X}) = O(t_{n_i}^2 \cdot w^2), \ u.a.s., \ \forall r = 1, \cdots, s_{n_i}.$$

Combining with (5.5.5), we then have

$$Ł_n^{\delta=2} = \frac{\sum_{r=1}^{s_{n_i}} E(U_{ir}^4 | \mathbf{X})}{\left[\sum_{r=1}^{s_{n_i}} E(U_{ir}^2 | \mathbf{X})\right]^2} = \frac{O(s_{n_i} \cdot t_{n_i}^2 w^2)}{[O(s_{n_i} \cdot t_{n_i} w)]^2}, \ u.a.s.$$

$$= O(s_{n_i}^{-1}) = o(1), \ u.a.s.$$

This completes the proof of the asymptotic conditional normality of $N^{1/2} w^{-1/2} \mathbf{Z}^{*\prime} \mathbf{A}_D^* \mathbf{Z}^*$. In other words, conditionally on $\mathbf{X}$,

$$\frac{N^{1/2} w^{-1/2} \mathbf{Z}^{*\prime} \mathbf{A}_D^* \mathbf{Z}^*}{Var\left[N^{1/2} w^{-1/2} \mathbf{Z}^{*\prime} \mathbf{A}_D^* \mathbf{Z}^* | \mathbf{X}\right]} \xrightarrow{d} N(0,1).$$

By the proof of Lemma 5.2.8, we know that, as $\min(n_i) \to \infty$,

$$Var\left(N^{1/2} w^{-1/2} \mathbf{Z}^{*\prime} \mathbf{A}_D^* \mathbf{Z}^* | \mathbf{X}\right) \longrightarrow \eta, \ a.s.$$

for some positive constant $\eta = \dfrac{2(2w-1)}{3(w-1)} \sum_{i=1}^{k} \lambda_i E[\sigma_i^4(X)]$.

By the Slutsky's Theorem that, conditionally on $\mathbf{X}$,

$$N^{1/2} w^{-1/2} \mathbf{Z}^{*\prime} \mathbf{A}_D^* \mathbf{Z}^* \stackrel{d}{\longrightarrow} N(0, \eta).$$

Since the limiting distribution is the same for all $\mathbf{X}$, this weak convergence also holds unconditionally (see Lemma 5.5.2), which completes the proof.

5.5.3 Some Auxiliary Results

Lemma 5.5.1. *Under Assumption 3, for any Lipschitz continuous function $h_i(x)$ on S_i and n_i large enough,*

$$\frac{1}{w_{ir}} \sum_{j=1}^{n_i} h_i(X_{ij}) I(j \in W_{ir}) - h_i(X_{il}) = O\left(n_i^{-1+\delta} + \frac{w}{n_i}\right), \quad \textit{uniformly a.s.,}$$

$$\forall \delta > 0,\ \forall l \in W_{ir},\ 1 \leq i \leq k,\ 1 \leq r \leq n_i.$$

Proof. Let $X_{ir}^U = \max(X_{ij},\ j \in W_{ir})$ and $X_{ir}^L = \min(X_{ij};\ j \in W_{ir})$ for an arbitrary window W_{ir}. For any $l \in W_{ir}$,

$$\left| \frac{1}{w_{ir}} \sum_{j=1}^{n_i} h_i(X_{ij}) I(j \in W_{ir}) - h_i(X_{il}) \right| \leq \frac{1}{w_{ir}} \sum_{j=1}^{n_i} |h_i(X_{ij}) - h_i(X_{il})| \cdot I(j \in W_{ir})$$

$$\leq \frac{1}{w_{ir}} \sum_{j=1}^{n_i} K_i \cdot |X_{ij} - X_{il}| \cdot I(j \in W_{ir}), \text{ for some constant } K_i > 0$$

($\because h_i$ are Lipschitz continuous functions.)

$$\leq \frac{1}{w_{ir}} \sum_{j=1}^{n_i} K_i \cdot |X_{ir}^U - X_{ir}^L| \cdot I(j \in W_{ir}) = K_i \cdot (X_{ir}^U - X_{ir}^L).$$

Recall that

$$\hat{G}_i(X_{ir}^U) - \hat{G}_i(X_{ir}^L) = \frac{w_{ir} - 1}{n_i}.$$

By Smirnov's LIL (Laws of the Iterated Logarithm), we have ($log_2 = log\,log$)

$$\overline{\lim}_{n_i \to \infty} \sqrt{\frac{n_i}{\log_2 n_i}} \sup_x \left| \hat{G}_i(x) - G_i(x) \right| = \frac{1}{\sqrt{2}}, \ a.s.$$

$$\Rightarrow \quad \overline{\lim}_{n_i \to \infty} \sqrt{\frac{n_i}{\log_2 n_i}} \ \left| [\hat{G}_i(X_{ir}^U) - G_i(X_{ir}^U)] - [\hat{G}_i(X_{ir}^L) - G_i(X_{ir}^L)] \right| \leq \sqrt{2}, \ a.s.$$

$$\Rightarrow \quad \overline{\lim}_{n_i \to \infty} \sqrt{\frac{n_i}{\log_2 n_i}} \ \left| G_i(X_{ir}^U) - G_i(X_{ir}^L) - \frac{w_{ir} - 1}{n_i} \right| \leq \sqrt{2}, \ a.s.$$

Since by MVT (the mean value theorem),

$$|G_i(X_{ir}^U) - G_i(X_{ir}^L)| = g_i(\widetilde{X}_i r) \cdot |X_{ir}^U - X_{ir}^L|, \text{ for some } X_{ir}^U \leq \widetilde{X}_{ir} \leq X_{ir}^L,$$

we know that, for all n_i large enough,

$$\left| g_i(\widetilde{X}_i r) \cdot [X_{ir}^U - X_{ir}^L] - \frac{w_{ir} - 1}{n_i} \right| \leq \sqrt{\frac{2 \ log_2 n_i}{n_i}}, \ a.s.$$

$$\Rightarrow \quad |X_{ir}^U - X_{ir}^L| \leq \frac{1}{g_i(\widetilde{X}_{ir})} \left(\sqrt{\frac{2 \ log_2 n_i}{n_i}} + \frac{w_{ir} - 1}{n_i} \right), \ a.s.$$

$$\Rightarrow \quad |X_{ir}^U - X_{ir}^L| \leq M \left(\sqrt{\frac{2 \ log_2 n_i}{n_i}} + \frac{w - 1}{n_i} \right), \ a.s., \text{ for some } M > 0.$$

$$(\because g_i \text{ are bounded away from } 0, by Assumption A3.)$$

Let

$$a_{n_i, \ 1} = M \left(\sqrt{\frac{2 \ log_2 n_i}{n_i}} + \frac{w - 1}{n_i} \right) = O\left(n_i^{-.5 + \delta_1} + \frac{w}{n_i} \right), \ \forall \delta_1 > 0.$$

Then, by Theorem 2.11 of Stute (1982), for all n_i large enough,

$$\sqrt{n_i} \left| G_i(X_{ir}^U) - G_i(X_{ir}^L) - \frac{w_{ir} - 1}{n_i} \right| \leq \sqrt{2 \ a_{n_i, \ 1} \log a_{n_i, \ 1}^{-1}}, \ a.s.$$

$$\Rightarrow \quad \left| g_i(\widetilde{X}_{ir}) \cdot (X_{ir}^U - X_{ir}^L) - \frac{w_{ir} - 1}{n_i} \right| \leq \sqrt{2 \ \frac{a_{n_i, \ 1}}{n_i} \log a_{n_i, \ 1}^{-1}}, \ a.s. \quad (by MVT)$$

$$\Rightarrow \quad \left| X_{ir}^U - X_{ir}^L \right| \leq M \left(\sqrt{2 \ \frac{a_{n_i, \ 1}}{n_i} \log a_{n_i, \ 1}^{-1}} + \frac{w - 1}{n_i} \right), \ a.s. \quad (by A3).$$

Let

$$a_{n_i, \ 2} = M \left(\sqrt{2 \ \frac{a_{n_i, \ 1}}{n_i} \log a_{n_i, \ 1}^{-1}} + \frac{w - 1}{n_i} \right) = O\left(n_i^{-.75 + \delta_2} + \frac{w}{n_i} \right), \ \forall \delta_2 > 0.$$

Then, by Theorem 2.11 of Stute (1982) again, for all n_i large enough,

$$\sqrt{n_i}\left|G_i(X_{ir}^U) - G_i(X_{ir}^L) - \frac{w_{ir}-1}{n_i}\right| \le \sqrt{2\ a_{n_i,\,2}\log a_{n_i,\,2}^{-1}},\ \ a.s.$$

$$\Rightarrow\quad \left|X_{ir}^U - X_{ir}^L\right| \le M\left(\sqrt{2\ \frac{a_{n_i,\,2}}{n_i}\log a_{n_i,\,2}^{-1}} + \frac{w-1}{n_i}\right),\ \ a.s.,$$

where

$$M\left(\sqrt{2\ \frac{a_{n_i,\,2}}{n_i}\log a_{n_i,\,2}^{-1}} + \frac{w-1}{n_i}\right) = O\left(n_i^{-\frac{7}{8}+\delta_3} + \frac{w}{n_i}\right),\ \ \forall\delta_3 > 0.$$

Repeatedly using Stute (1982), we can then obtain that, for n_i large enough,

$$X_{ir}^U - X_{ir}^L = O\left(n_i^{-1+\delta} + \frac{w}{n_i}\right),\ \text{uniformly a.s.,} \tag{5.5.8}$$

which completes the proof.

Lemma 5.5.2. *Suppose that, conditionally on* $\mathbf{U}_n = (U_1, \cdots, U_n)'$,

$$\widehat{\theta}_n\left(\mathbf{V}_n = (V_1, \cdots, V_n)'\right) \xrightarrow{d} Z, \quad \text{as } n \to \infty, \tag{5.5.9}$$

where the distribution of Z *is continuous and does not depend on* $\mathbf{X}$*. If* $U_i,\ i = 1, \cdots, n$*, are from a continuous random variable with bounded support* S*, then the above weak convergence holds unconditionally.*

Proof. Since the limiting distribution is continuous, (5.5.9) implies that

$$P\left(\widehat{\theta}_n \le t \mid \mathbf{U}_n = \mathbf{u}\right) \longrightarrow F_Z(t), \quad \text{for all } \mathbf{u},$$

where $F_Z(t) = P(Z \le t)$, which follows that the convergence is uniformly in $\mathbf{u}$, i.e.

$$\sup_{\mathbf{u}} |P\left(\widehat{\theta}_n \le t \mid \mathbf{U}_n = \mathbf{u}\right) - F_Z(t)| \longrightarrow 0, \quad \text{as } n \to \infty.$$

This implies that

$$\forall\epsilon > 0,\ P\left[|P\left(\widehat{\theta}_n \le t \mid \mathbf{U}_n\right) - F_Z(t)| > \epsilon\right] \to 0, \quad \text{as } n \to \infty$$

$$\Longleftrightarrow \quad P\left(\widehat{\theta}_n \leq t \mid \mathbf{U}_n\right) \stackrel{P}{\longrightarrow} F_Z(t), \text{ as } n \to \infty.$$

This convergence i probability tells use that, for every subsequence $\{k_1,\ k_2,\cdots\} \subseteq \{1,\cdots,n\}$, there exists a further subsequence $\{m_1,\ m_2,\cdots,m_j\} \subseteq \{k_1,\ k_2,\cdots\}$ such that

$$P\left(\widehat{\theta}_n \leq t \mid \mathbf{U}_{m_j}\right) \stackrel{a.s.}{\longrightarrow} F_Z(t), \quad \text{as } j \to \infty.$$

Then, by the Lebesgue Dominated Convergence Theorem (DCT),

$$E\left\{P\left(\widehat{\theta}_n \leq t \mid \mathbf{U}_{m_j}\right)\right\} \longrightarrow F_Z(t), \quad \text{as } j \to \infty.$$

Since the above equation holds for every subsequence of $\{1,\cdots,n\}$, we have

$$P\left(\widehat{\theta}_n \leq t\right) = E\left\{P\left(\widehat{\theta}_n \leq t \mid \mathbf{U}_n\right)\right\} \longrightarrow F_Z(t), \quad \text{as } n \to \infty,$$

which completes the proof.

Chapter 6

Asymptotics in Testing for the Group effect in the Fully Nonparametric ANCOVA

Recall the fully nonparametric ANCOVA model as defined in Section 1.2, which assumes only that

$$Y_{ij}|X_{ij} = x \sim F_{ix}(y) = M(y) + A_i(y) + D_i(y;x),\ \ i = 1, \cdots, k,$$

where $(X_{ij},\ Y_{ij})$ are paired observable variables while the decomposition of F_{ix} is defined in (1.2.7). Also recall that, for any given distribution functions $G_i(x),\ i = 1, \cdots, k$, we define

$$\bar{F}_{i\cdot}^{G_i}(y) = \int F_{ix}(y)\ dG_i(x),$$

and it is assumed that $\int D_i(y;x)\ dG_i(x) = 0,\ \forall i,\ \forall y.$ The null hypothesis of no covariate-adjusted group effect in the fully nonparametric fashion is then specified as:

$$H_0(A):\ A_i(y) = 0 \quad \Leftrightarrow \quad \bar{F}_{i\cdot}^{G_i}(y) = F^0(y),\ \text{for all } i \text{ and all } y. \tag{6.0.1}$$

As introduced in the previous chapter, there are two possible approaches to utilize the similarity between the nested model and the analysis of covariate: *non-overlapping windows approach* and *overlapping windows approach*. For testing (6.0.1), the *non-overlapping windows approach* can apply directly the asymptotic results from Chapter 4, while the present chapter is constructed to tackle the challenges coming with the augmented dependence in the design of the overlapping windows approach, and to provide an appropriate test procedure accordingly. We first introduce the test statistic in the next section, and then provide its asymptotic distributions, both under the null and local alternatives, in Section 6.2. All proofs can be found in Section 6.3.

6.1 The Test Statistic

For a k-group ANCOVA model with overlapping windows of a fixed size w (assuming to be *odd* for simplicity), consider the notations in the hypothetical two-fold nested design as introduced in Section 5.1: letting $(X_{ij},\ Y_{ij}),\ i = 1, \cdots, k;\ j = 1, \cdots, n_i$, be the *enumerated* pairs with $X_{i1} < X_{i2} < \cdots < X_{in_i}$ for each i,

$$Z_{irt} = Y_{ij} \quad \text{iff} \qquad \sum_{l=1}^{n_i} I(X_{il} \le x_{ij};\ l \in W_{ir}) = t,$$
$$i = 1, \cdots, k;\ r = 1, \cdots, n_i;\ t = 1, \cdots, w_{ir},$$

where w_{ir} are as shown in (5.1.2) and

$$W_{ir} = \left\{ 1 \le j \le n_i :\ |\widehat{G}_i(X_{ij}) - \widehat{G}_i(X_{ir})| \le \frac{w-1}{2n_i} \right\}, \quad \widehat{G}_i(x) = \frac{1}{n_i} \sum_{j=1}^{n_i} I(X_{ij} \le x).$$

Note that W_{ir} are in fact functions of X_{ij}'s. Also recall that $w_{i\cdot} = \sum_r w_{ir} = n_i w - (w^2 - 1)/4$ and $w_{\cdot\cdot} = \sum_i w_{i\cdot} = Nw - k(w^2 - 1)/4$.

Define

$$\bar{Z}_{ir\cdot} = \frac{1}{w_{ir}} \sum_{t=1}^{w_{ir}} Z_{irt} \qquad \text{and} \qquad \bar{Z}_{i\cdot\cdot} = \frac{1}{w_{i\cdot}} \sum_{r=1}^{n_i} w_{ir} \bar{Z}_{ir\cdot}.$$

For testing the null hypothesis of no covariate-adjusted group effect (6.0.1), we consider a test statistic based on the random vector $\mathbf{U} = (U_1, \cdots, U_k)'$ where $U_i \equiv \bar{Z}_{i\cdot\cdot},\ i = 1, \cdots, k$. More specifically, since the aforementioned hypothesis (6.0.1) is equivalent to

$$H_0(\mathbf{H}):\ \mathbf{H}\,\mathbf{F} = \mathbf{0}_{k-1}, \quad \text{where } \mathbf{F} = (\bar{F}_{1\cdot}^{G_1}, \cdots, \bar{F}_{k\cdot}^{G_k})',$$

and the contrast matrix $\mathbf{H}$ is defined as

$$\mathbf{H} = \begin{pmatrix} 1 & -1 & 0 & \cdots & 0 \\ 1 & 0 & -1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & 0 & \cdots & -1 \end{pmatrix}_{(k-1)\times k} = \left(\mathbf{1}_{k-1}' \mid -\mathbf{I}_{k-1}\right), \tag{6.1.1}$$

the test statistic we consider in this chapter for testing (6.0.1) is of the form

$$N\,(\mathbf{H}\,\mathbf{U})'\left(\mathbf{H}\,\widehat{\boldsymbol{\Sigma}}\,\mathbf{H}'\right)^{-1}(\mathbf{H}\,\mathbf{U})$$

where $\mathbf{H}$ is the contrast matrix defined above, and $\widehat{\boldsymbol{\Sigma}}$ is a suitable estimator of the asymptotic covariance matrix of $\sqrt{N}\mathbf{U}$. As shown in the next section, under the null hypothesis $H_0(A)$, this test statistic converges in distribution to a χ^2 distribution with degree of freedom $(k-1)$, as $\min(n_i) \to \infty$. Its asymptotic distributions under suitable local alternatives are also included in the next section.

6.2 The Main Results

6.2.1 Assumptions and Lemmas

Here are two notations which are frequently used in this chapter:

$$\mu_i(x) \equiv E(Y_{ij} \mid X_{ij} = x); \quad \sigma_i^2(x) \equiv Var(Y_{ij} \mid X_{ij} = x),$$

and all assumptions required for Lemmas and Theorems presented in this section are as follows:

Assumption S1. $\forall i,\ \exists \lambda_i \in (0,1)$, such that $n_i/N \to \lambda_i$, as $\underline{n} = \min(n_i) \to \infty$;

Assumption S2. The covariate X is a continuous random variable with bounded support S_i, c.d.f G_i, and p.d.f g_i, $i = 1, \cdots, k$. The density g_i is bounded away from 0 on S_i uniformly in i;

Assumption S3. $E(Y_{ij}^4 \mid X_{ij} = x)$ are uniformly bounded in i and x;

Assumption S4. $\sigma_i^2(x)$ are bounded away from 0 and ∞ uniformly in i and x.

Assumption S5. $\mu_i(x)$ and $\sigma_i^2(x)$ are uniformly Lipschitz continuous in x.

The following four lemmas which are basic vehicles for deriving the asymptotic distributions of the test statistic under the null and the local alternatives.

Lemma 6.2.1. *Under Assumptions S1–S4, for all* $i = 1, \cdots, k$, *we have that, conditionally on* $\mathbf{X}$,

$$\sqrt{N}\left(U_i - E(U_i \mid \mathbf{X})\right) \stackrel{d}{\longrightarrow} N\left(0, a_i^2\right), \quad as \ \min(n_i) \to \infty,$$

where $U_i = \bar{Z}_{i\cdot\cdot}$ *and*

$$a_i^2 = \frac{1}{\lambda_i} \int \sigma_i^2(x)\, dG_i(x).$$

Lemma 6.2.2. *Under Assumptions S1, S3–S4, for all* $i = 1, \cdots, k$,

$$\sqrt{N}\left(E(U_i \mid \mathbf{X}) - \int\int y\, dF_{ix}(y)\, d\widehat{G}_i(x)\right) \longrightarrow 0, \ a.s., \quad as \ \min(n_i) \to \infty.$$

Lemma 6.2.3. *Under Assumptions S1–S2, for all* $i = 1, \cdots, k$,

$$\sqrt{N}\int\int y\, dF_{ix}(y)\, d\left(\widehat{G}_i(x) - G_i(x)\right) \stackrel{d}{\longrightarrow} N\left(0, b_i^2\right), \quad as \ \min(n_i) \to \infty,$$

where

$$b_i^2 = \frac{1}{\lambda_i}\left\{\int\left[\int y\, dF_{ix}(y)\right]^2 dG_i(x) - \left[\int\int y\, dF_{ix}(y)\, dG_i(x)\right]^2\right\}.$$

Lemma 6.2.4. *Under Assumptions S1–S5,*

$$\widehat{a}_i^2 = \frac{1}{\widehat{\lambda}_i} \cdot \frac{1}{n_i} \sum_{j=1}^{n_i} \widehat{\sigma}_i^2 \left(X_{ij}\right), \quad \textit{where}$$

$$\widehat{\sigma}_i^2 \left(X_{ij}\right) = \frac{1}{w_{ij}-1} \sum_{l=1}^{n_i} Y_{il}^2 \times I(l \in W_{ij}) - \frac{w_{ij}}{w_{ij}-1} \left[\frac{1}{w_{ij}} \sum_{l=1}^{n_i} Y_{il} \times I(l \in W_{ij}) \right]^2 ;$$

$$\widehat{b}_i^2 = \frac{1}{\widehat{\lambda}_i} \left\{ \frac{1}{n_i} \sum_{j=1}^{n_i} \left(C_2^{w_{ij}}\right)^{-1} \sum_{l_1 \neq l_2} \frac{Y_{il_1}\, Y_{il_2}}{2} \times I(l_i,\ l_2 \in W_{ij}) - \left[\frac{1}{n_i} \sum_{j=1}^{n_i} \widehat{\mu}_i \left(X_{ij}\right) \right]^2 \right\},$$

$$\textit{where} \quad \widehat{\mu}_i \left(X_{ij}\right) = \frac{1}{w_{ij}} \sum_{l=1}^{n_i} Y_{il} \times I(l \in W_{ij});$$

are strongly consistent estimators of a_i^2 *and* b_i^2*, as defined in Lemma 6.2.1 and in Lemma 6.2.3, respectively.*

6.2.2 Asymptotic Distribution under the Null

Let

$$\mathbf{U} = (U_1, \cdots, U_k)' = (\bar{Z}_{1\cdot\cdot}, \cdots, \bar{Z}_{k\cdot\cdot})', \tag{6.2.1}$$

$$\mathbf{\Sigma} = diag\left(a_1^2 + b_1^2, \cdots, \cdots, a_k^2 + b_k^2\right), \tag{6.2.2}$$

where a_i^2 and b_i^2, $i = 1, \cdots, k$, are as defined in Lemma 6.2.1 and in Lemma 6.2.3.

The following Lemma gives the asymptotic result under the null hypothesis of no covariate-adjusted group effect when the covariance elements a_i^2 and b_i^2, $i = 1, \cdots, k$, are known, while Theorem 6.2.1 that follows states the asymptotic distribution of the proposed test statistic under $H_0(A)$.

Lemma 6.2.5. *Under* $H_0(A)$ *as defined in (6.0.1) and Assumptions S1–S4,*

$$N \left(\mathbf{H}\,\mathbf{U}\right)' \left(\mathbf{H}\,\mathbf{\Sigma}\,\mathbf{H}'\right)^{-1} \left(\mathbf{H}\,\mathbf{U}\right) \xrightarrow{d} \chi_{k-1}^2, \quad \textit{as } \min\left(n_i\right) \to \infty,$$

where $\mathbf{H}$ *is the contrast matrix as defined in (6.1.1) while* $\mathbf{U}$ *and* $\mathbf{\Sigma}$ *are as defined in (6.2.1) and (6.2.2).*

Theorem 6.2.1. *Under $H_0(A)$ as defined in (6.0.1) and Assumptions S1–S5,*

$$N\left(\mathbf{H}\,\mathbf{U}\right)'\left(\mathbf{H}\,\widehat{\mathbf{\Sigma}}\,\mathbf{H}'\right)^{-1}\left(\mathbf{H}\,\mathbf{U}\right)\xrightarrow{d}\chi^2_{k-1},\quad as\ \min\left(n_i\right)\to\infty,$$

where $\mathbf{H}$ is the $(k-1)\times k$ contrast matrix as defined in (6.1.1), $\mathbf{U}$ is as defined in (6.2.1), and $\widehat{\mathbf{\Sigma}} = diag\left(\widehat{a}_1^2+\widehat{b}_1^2,\cdots,\cdots,\widehat{a}_k^2+\widehat{b}_k^2\right)$, where the $\widehat{a}_i^2$ and $\widehat{b}_i^2$, $i=1,\cdots,k$, are as defined in Lemma 6.2.4.

6.2.3 Asymptotic Distribution under Local Alternatives

The asymptotic power property is investigated by considering the local alternative sequence:

$$H_a(A):\ A_i(y)=(n_i)^{-1/2}B_i(y), \tag{6.2.3}$$

where $B_i(y)$, $i=1,\cdots,k$, are so chosen that $\int y dB_i(y)$ are uniformly bounded for all i. Note that (6.2.3) implies that the alternatives need to approach the null at the rate of $(\inf_i n_i)^{-1/2}$ to ensure nontrivial power. Define

$$\zeta=\left(\sum_{i=1}^{k}\frac{\lambda_i\int y\ dB_i(y)}{a_i^2+b_i^2}\right)-\frac{1}{\sum_{i=1}^{k}1/\left(a_i^2+b_i^2\right)}\left(\sum_{i=1}^{k}\frac{\sqrt{\lambda_i}\int y\ dB_i(y)}{a_i^2+b_i^2}\right)^2. \tag{6.2.4}$$

Then, Theorem 6.2.2 below provides the asymptotic distribution of the proposed test statistic under the local alternatives $H_a(A)$.

Theorem 6.2.2. *Under $H_a(A)$ as defined in (6.2.3) and Assumptions S1–S5,*

$$N\left(\mathbf{H}\,\mathbf{U}\right)'\left(\mathbf{H}\,\widehat{\mathbf{\Sigma}}\,\mathbf{H}'\right)^{-1}\left(\mathbf{H}\,\mathbf{U}\right)\xrightarrow{d}\chi^2_{k-1}\left(\zeta\right),\quad as\ \min\left(n_i\right)\to\infty,$$

where $\mathbf{H}$, $\mathbf{U}$, and $\widehat{\mathbf{\Sigma}}$ are as defined in Theorem 6.2.1, while ζ is as defined in (6.2.4).

6.3 Technical Details

Recall that X_{ij} are *ordered* covariate variable. Letting $X_{ij}^{(o)}$ be the *original* covariate variable, i.e. the one without ordering, there is one fact which is repeated used in our proofs: For any well-defined function $f(x)$, it should be clear that $\sum_{j=1}^{n_i} f(X_{ij})$ has the same distribution as $\sum_{j=1}^{n_i} f(X_{ij}^{(o)})$, since the only difference between the original $\{X_{ij}^{(o)},\ j = 1, \cdots, n_i\}$ and the ordered $\{X_{ij},\ j = 1, \cdots, n_i\}$ are the permutations. Note that we take advantage of the $X_{ij}^{(o)}$'s independent and identically distributed feature (within the group i) to significantly simplify the proofs in this section. Similarly, let $(X_{ij}^{(o)}, Y_{ij}^{(o)}),\ j = 1, \cdots, n_i$, denote the original un-ordered paired data, which are i.i.d.

Let

$$\boldsymbol{\mu}_G = \left(\int\int y\, dF_{1x}(y)\, dG_1(x), \cdots, \int\int y\, dF_{kx}(y)\, dG_k(x) \right)'. \tag{6.3.1}$$

6.3.1 Proofs of Lemmas and Theorems

Proof of Lemma 6.2.1

First note that

$$U_i \;=\; \bar{Z}_{i\cdot\cdot} = \frac{1}{w_{i\cdot}} \sum_{r=1}^{n_i} w_{ir} \bar{Z}_{ir\cdot} = \frac{1}{w_{i\cdot}} \sum_{r=1}^{n_i} \sum_{t=1}^{w_{ir}} Z_{irt} = \frac{1}{w_{i\cdot}} \sum_{r=1}^{n_i} \sum_{j=1}^{n_i} Y_{ij} \times I(j \in W_{ir}),$$

which, conditionally on $\mathbf{X}$,

$$U_i = \frac{1}{w_{i\cdot}} \sum_{j=1}^{n_i} w_{ij} \cdot Y_{ij}.$$

So,

$$E\left(U_i \mid \mathbf{X}\right) = \frac{1}{w_{i\cdot}} \sum_{j=1}^{n_i} w_{ij}\, \mu_i\left(X_{ij}\right), \quad Var\left(U_i \mid \mathbf{X}\right) = \frac{1}{w_{i\cdot}^2} \sum_{j=1}^{n_i} w_{ij}^2\, \sigma_i^2\left(X_{ij}\right).$$

Also note that under Assumptions S3 and S4, both of $\mu_i(X_{ij})$ and $\sigma_i^2(X_{ij})$ are $O(1)$, uniformly almost surely.

For the limiting distribution of $\sum_j w_{ij}\left[Y_{ij} - E(Y_{ij} \mid X_{ij})\right]$, check the Lyapounov condition : conditionally on $\mathbf{X}$,

$$\begin{aligned} Ł_{n_i}^{\delta=2} &= \frac{\sum_{j=1}^{n_i} E\left\{\left|w_{ij}\left[Y_{ij} - E\left(Y_{ij} \mid X_{ij}\right)\right]\right|^4 \mid \mathbf{X}\right\}}{\left(\sqrt{\sum_{j=1}^{n_i} w_{ij}^2 \, Var\left(Y_{ij} \mid X_{ij}\right)}\right)^4} \\ &= \frac{n_i^{-1}}{\left(\frac{1}{n_i}\sum_{j=1}^{n_i} w_{ij}^2 \sigma_i^2\left(X_{ij}\right)\right)^2} \cdot \frac{1}{n_i}\sum_{j=1}^{n_i} w_{ij}^4 \, E\left(Y_{ij}^* \mid X_{ij}\right), \\ &\qquad \text{where } Y_{ij}^* = Y_{ij} - E\left(Y_{ij} \mid X_{ij}\right), \\ &= o(1), \;\; u.a.s., \;\; \text{as } \min\left(n_i\right) \to \infty. \end{aligned}$$

Since the Lyapounov condition holds, we have that, conditionally on $\mathbf{X}$,

$$\frac{\sum_{j=1}^{n_i} w_{ij} \cdot Y_{ij}^*}{\sqrt{\sum_{j=1}^{n_i} w_{ij}^2 \cdot \sigma_i^2(X_{ij})}} = \frac{\sqrt{N}\left[U_i - E\left(U_i \mid \mathbf{X}\right)\right]}{\sqrt{Var\left(\sqrt{N}\, U_i \mid \mathbf{X}\right)}} \xrightarrow{d} N(0,1), \quad \text{as } \min\left(n_i\right) \to \infty.$$

Because by Lemma 6.3.1,

$$Var\left(\sqrt{N} U_i \mid \mathbf{X}\right) \longrightarrow a_i^2, \; a.s., \quad \text{as } \min\left(n_i\right) \to \infty,$$

for some positive constant a_i^2 are as defined in Lemma 6.2.1, we know by the Slutsky's Theorem that, conditionally on $\mathbf{X}$,

$$\sqrt{N}\left(U_i - E(U_i \mid \mathbf{X})\right) \xrightarrow{d} N\left(0, a_i^2\right), \quad \text{as } \min\left(n_i\right) \to \infty,$$

which completes the proof.

Proof of Lemma 6.2.2

First note that

$$E(U_i \mid \mathbf{X}) - \int\int y \, dF_{ix}(y)\, d\widehat{G}_i(x) = \frac{1}{w_{i\cdot}}\sum_{j=1}^{n_i} w_{ij} E\left(Y_{ij} \mid X_{ij}\right) - \frac{1}{n_i}\sum_{j=1}^{n_i} E\left(Y_{ij} \mid X_{ij}\right).$$

Since

$$\frac{w_{ij}}{w_{i\cdot}} - \frac{1}{n_i} = \frac{n_i \cdot w_{ij} - w_{i\cdot}}{n_i \cdot w_{i\cdot}} = \frac{n_i\ w - \left(n_i\ w - \frac{w^2-1}{4}\right)}{n_i\ w_{i\cdot}} = \frac{\left(w^2-1\right)/4}{n_i\left(n_i\ w - \frac{w^2-1}{4}\right)},$$

if $\frac{w-1}{2} + 1 \leq j \leq n_i - \frac{w-1}{2}$; otherwise, it's $O(n_i^{-1})$. In addition, note that $E\left(Y_{ij} \mid X_{ij}\right) = O(1)$, u.a.s. by Assumptions S3 and S4. Therefore,

$$\begin{aligned} & E(U_i \mid \mathbf{X}) - \int\int y\, dF_{ix}(y)\, d\widehat{G}_i(x) \\ = & \left(\sum_{j=1}^{\frac{w-1}{2}} O\left(n_i^{-1}\right) + \sum_{j=\frac{w-1}{2}+1}^{n_i-\frac{w-1}{2}} O\left(n_i^{-2}\right) + \sum_{j=n_i-\frac{w-1}{2}+1}^{n_i} O\left(n_i^{-1}\right)\right) \times O(1),\ u.a.s. \\ = & O(N^{-1}),\ u.a.s., \end{aligned}$$

which completes the proof.

Proof of Lemma 6.2.3

First note that

$$\begin{aligned} \int\int y\, dF_{ix}(y)\, d\widehat{G}_i(x) &= \int E\left[Y_{ij} \mid X_{ij} = x\right] d\widehat{G}_i(x) = \frac{1}{n_i}\sum_{j=1}^{n_i} \mu_i\left(X_{ij}\right); \\ \int\int y\, dF_{ix}(y)\, dG_i(x) &= \int E\left[Y_{ij} \mid X_{ij} = x\right] dG_i(x) = E\left[\mu_i\left(X_{ij}\right)\right]. \end{aligned}$$

Let $(X_{ij}^{(o)}, Y_{ij}^{(o)}),\ j = 1, \cdots, n_i$, denote the original un-ordered paired data, and let $\mu_i\left(X_{ij}^{(o)}\right) = E\left(Y_{ij}^{(o)} \mid X_{ij}^{(o)}\right)$ and $\sigma_i^2\left(X_{ij}^{(o)}\right) = Var\left(Y_{ij}^{(o)} \mid X_{ij}^{(o)}\right)$. Note that

$$\frac{1}{n_i}\sum_{j=1}^{n_i} \mu_i\left(X_{ij}^{(o)}\right) \stackrel{d}{=} \frac{1}{n_i}\sum_{j=1}^{n_i} \mu_i\left(X_{ij}\right) = \int\int y\, dF_{ix}(y)\, d\widehat{G}_i(x).$$

In addition, by Lemma 13.1 of van der Vaart (2000),

$$E\left[\frac{1}{n_i}\sum_{j=1}^{n_i} \mu_i\left(X_{ij}^{(o)}\right)\right] = E\left\{E\left[\frac{1}{n_i}\sum_{j=1}^{n_i} \mu_i\left(X_{ij}^{(o)}\right) \mid \mathbf{R}_{n_i}\right]\right\},$$

where $\mathbf{R}_{n_i}$ are the vector of *ranks* of $X_{ij}^{(o)},\ j = 1, \cdots, n_i$;

$$= \quad E\left\{E\left[\frac{1}{n_i}\sum_{j=1}^{n_i}\mu_i\left(X_{ij}\right)\right]\right\} \;=\; E\left[\frac{1}{n_i}\sum_{j=1}^{n_i}\mu_i\left(X_{ij}\right)\right].$$

Similarly,

$$Var\left[\frac{1}{n_i}\sum_{j=1}^{n_i}\mu_i\left(X_{ij}^{(o)}\right)\right] \;=\; Var\left[\frac{1}{n_i}\sum_{j=1}^{n_i}\mu_i\left(X_{ij}\right)\right].$$

Since $X_{ij}^{(o)}$ are i.i.d., by CLT,

$$\frac{\sqrt{N}\left\{\frac{1}{n_i}\sum_{j=1}^{n_i}\left[\mu_i\left(X_{ij}^{(o)}\right) - E\left(\mu_i\left(X_{ij}^{(o)}\right)\right)\right]\right\}}{Var\left(\frac{\sqrt{N}}{n_i}\sum_{j=1}^{n_i}\mu_i\left(X_{ij}^{(o)}\right)\right)} \xrightarrow{d} N(0,1), \quad \text{as } \min\left(n_i\right)\to\infty,$$

while

$$\begin{aligned} E\left[\frac{1}{n_i}\sum_{j=1}^{n_i}\mu_i\left(X_{ij}^{(o)}\right)\right] &= E\left[\frac{1}{n_i}\sum_{j=1}^{n_i}\mu_i\left(X_{ij}\right)\right] = \int\int y\, dF_{ix}(y)\, dG_i(x), \text{ and} \\ Var\left[\frac{\sqrt{N}}{n_i}\sum_{j=1}^{n_i}\mu_i\left(X_{ij}^{(o)}\right)\right] &= \frac{N}{n_i}Var\left[E\left(Y_{ij}^{(o)} \mid X_{ij}^{(o)}\right)\right] = \frac{N}{n_i}\,\lambda_i\cdot b_i^2, \end{aligned}$$

where b_i^2 is as defined in Lemma 6.2.3. This completes the proof.

Proof of Lemma 6.2.4

Firstly,

$$\begin{aligned} &\widehat{\mu}_i\left(X_{ij}\right) = \frac{1}{w_{ij}}\sum_{l=1}^{n_i} Y_{il}\times I(l\in W_{ij}) \\ \Longrightarrow\quad &\frac{1}{n_i}\sum_{j=1}^{n_i}\left[\widehat{\mu}_i\left(X_{ij}\right) - \mu_i\left(X_{ij}\right)\right] = \frac{1}{n_i}\sum_{j=1}^{n_i}\frac{1}{w_{ij}}\sum_{l=1}^{n_i}\left[Y_{il} - \mu_i\left(X_{ij}\right)\right]\times I(l\in W_{ij}) \\ &= \frac{1}{n_i}\sum_{j=1}^{n_i}\frac{1}{w_{ij}}\sum_{l=1}^{n_i}\left[Y_{il} - \mu_i\left(X_{il}\right)\right]\times I(l\in W_{ij}) \end{aligned}$$

$$+\frac{1}{n_i}\sum_{j=1}^{n_i}\frac{1}{w_{ij}}\sum_{l=1}^{n_i}\left[\mu_i\left(X_{il}\right)-\mu_i\left(X_{ij}\right)\right]\times I(l\in W_{ij})$$
$$\equiv\ M_{1i}+M_{2i}.$$

Note that

$$M_{1i}=\frac{1}{n_i}\sum_{l=1}^{n_i}\left[Y_{il}-\mu_i\left(X_{il}\right)\right]\sum_{j=1}^{n_i}\frac{1}{w_{ij}}I(l\in W_{ij}).$$

It can be shown that for any l,

$$\frac{\frac{w-1}{2}+1}{w}\ <\ \sum_{j=1}^{n_i}\frac{1}{w_{ij}}I(l\in W_{ij})\ =\ \sum_{j=\max\left(1,l-\frac{w-1}{2}\right)}^{\min\left(n_i,l+\frac{w-1}{2}\right)}\frac{1}{w_{ij}}\ <\ \frac{w}{\frac{w-1}{2}+1},\tag{6.3.2}$$

and

$$\begin{aligned}
E\left[Y_{il}-\mu_i\left(X_{il}\right)\right] &= E\left[Y_{il}-E\left(Y_{il}\mid X_{il}\right)\right]=0,\\
Var\left[Y_{il}-\mu_i\left(X_{il}\right)\right] &= E\left[Var\left(Y_{il}-\mu_i\left(X_{il}\right)\mid X_{il}\right)\right]+Var\left[E\left(Y_{il}-\mu_i\left(X_{il}\right)\mid X_{il}\right)\right],\\
&\qquad \text{where } E\left(Y_{il}-\mu_i\left(X_{il}\right)\mid X_{il}\right)=0\\
&= E\left[\sigma_i^2\left(X_{il}\right)\right]\ =\ O(1),\quad \text{uniformly.}
\end{aligned}$$

Since $\sum_{l=1}^{n_i}\left[Y_{il}-\mu_i\left(X_{il}\right)\right]\overset{d}{=}\sum_{l=1}^{n_i}\left[Y_{il}^{(o)}-\mu_i\left(X_{il}^{(o)}\right)\right]$ and the original $(X_{ij}^{(o)},\ Y_{ij}^{(o)})$ are i.i.d., by SLLN,

$$M_{1i}\longrightarrow 0,\ a.s.,\quad \text{as } n_i\to\infty.$$

On the other hand, under Assumption S5, we know by Lemma 5.5.1 that, for n_i large enough, $\forall\delta>0$,

$$\frac{1}{w_{ij}}\sum_{l=1}^{n_i}\mu_i(X_{il})I(l\in W_{ij})-\mu_i(X_{ij})=O\left(n_i^{-1+\delta}+\frac{w}{n_i}\right),\quad \text{uniformly a.s., } \forall j,$$

so

$$M_{2i} = \frac{1}{n_i} \sum_{j=1}^{n_i} O\left(n_i^{-1+\delta} + \frac{w}{n_i}\right) \longrightarrow 0,\ a.s., \quad \text{as } n_i \to \infty,$$

which completes the proof of

$$\frac{1}{n_i} \sum_{j=1}^{n_i} \left[\widehat{\mu}_i\left(X_{ij}\right) - \mu_i\left(X_{ij}\right)\right] \longrightarrow 0,\ a.s., \quad \text{as } n_i \to \infty.$$

Moreover, under Assumptions S2–S4, by SLLN,

$$\frac{1}{n_i} \sum_{j=1}^{n_i} \mu_i\left(X_{ij}\right) - E\left[\mu_i\left(X_{ij}\right)\right] \longrightarrow 0,\ a.s., \quad \text{as } n_i \to \infty.$$

Therefore,

$$\frac{1}{n_i} \sum_{j=1}^{n_i} \widehat{\mu}_i\left(X_{ij}\right) \ = \ E\left[\mu_i\left(X_{ij}\right)\right] + o(1),\ a.s. \tag{6.3.3}$$

Secondly,

$$\begin{aligned}
\widehat{\sigma}_i^2\left(X_{ij}\right) &= \frac{1}{w_{ij}-1} \sum_{l=1}^{n_i} Y_{il}^2 \times I(l \in W_{ij}) - \frac{w_{ij}}{w_{ij}-1} \left[\widehat{\mu}_i\left(X_{ij}\right)\right]^2,\ \text{where} \\
\left[\widehat{\mu}_i\left(X_{ij}\right)\right]^2 &= \left[\frac{1}{w_{ij}} \sum_{l=1}^{n_i} Y_{il} \times I(l \in W_{ij})\right]^2 \\
&= \frac{1}{w_{ij}} \left\{\frac{1}{w_{ij}} \sum_{l=1}^{n_i} Y_{il}^2 \times I(l \in W_{ij})\right\} + \frac{1}{w_{ij}^2} \sum_{l_1 \neq l_2} Y_{il_1}\ Y_{il_2} \times I(l_1,\ l_2 \in W_{ij}) \\
\Longrightarrow \widehat{\sigma}_i^2\left(X_{ij}\right) &= \left(\frac{w_{ij}}{w_{ij}-1} - \frac{w_{ij}}{w_{ij}-1} \times \frac{1}{w_{ij}}\right) \frac{1}{w_{ij}} \sum_{l=1}^{n_i} Y_{il}^2 \times I(l \in W_{ij}) \\
&\quad - \frac{w_{ij}}{w_{ij}-1} \times \frac{1}{w_{ij}^2} \sum_{l_1 \neq l_2} Y_{il_1}\ Y_{il_2} \times I(l_1,\ l_2 \in W_{ij}) \\
&= \frac{1}{w_{ij}} \sum_{l=1}^{n_i} Y_{il}^2 \times I(l \in W_{ij}) - \frac{1}{w_{ij}(w_{ij}-1)} \sum_{l_1 \neq l_2} Y_{il_1}\ Y_{il_2} \times I(l_1,\ l_2 \in W_{ij})
\end{aligned}$$

To prove that

$$\frac{1}{n_i}\sum_{j=1}^{n_i}\left[\widehat{\sigma}_i\left(X_{ij}\right)-\sigma_i\left(X_{ij}\right)\right]\longrightarrow 0,\ a.s.,\quad \text{as } n_i\to\infty, \tag{6.3.4}$$

it suffices to show that

$$Q_{1i}\equiv\frac{1}{n_i}\sum_{j=1}^{n_i}\left[\frac{1}{w_{ij}}\sum_{l=1}^{n_i}Y_{il}^2\times I(l\in W_{ij})-E\left(Y_{ij}^2\mid X_{ij}\right)\right]=o(1),\ a.s.$$

and

$$Q_{2i}\equiv\frac{1}{n_i}\sum_{j=1}^{n_i}\left[\frac{1}{w_{ij}(w_{ij}-1)}\sum_{l_1\neq l_2}Y_{il_1}\ Y_{il_2}\times I(l_1,\ l_2\in W_{ij})-\left(E\left(Y_{ij}\mid X_{ij}\right)\right)^2\right]=o(1),\ a.s.$$

Further write Q_{1i} as

$$\begin{aligned}
Q_{1i} &= \frac{1}{n_i}\sum_{j=1}^{n_i}\frac{1}{w_{ij}}\sum_{l=1}^{n_i}\left[Y_{il}^2-E\left(Y_{il}^2\mid X_{il}\right)\right]\times I(l\in W_{ij})\\
&\quad+\frac{1}{n_i}\sum_{j=1}^{n_i}\frac{1}{w_{ij}}\sum_{l=1}^{n_i}\left[E\left(Y_{il}^2\mid X_{il}\right)-E\left(Y_{ij}^2\mid X_{ij}\right)\right]\times I(l\in W_{ij})\\
&\equiv T_{1i}+T_{2i}.
\end{aligned}$$

$T_{1i}=o(1),\ a.s.$, because by (6.3.2),

$$\sum_{j=1}^{n_i}\frac{1}{w_{ij}}I(l\in W_{ij})=O(1),\ \text{uniformly in } l, \tag{6.3.5}$$

and by SLLN,

$$\frac{1}{n_i}\sum_{l=1}^{n_i}\left[Y_{il}^2-E\left(Y_{il}^2\mid X_{il}\right)\right]\ \longrightarrow\ 0,\quad \text{as } n_i\to\infty.$$

In addition, by Lemma 5.5.1 and under Assumption S5, for n_i large enough,

$$\frac{1}{w_{ij}}\sum_{l=1}^{n_i}E\left(Y_{il}^2\mid X_{il}\right)\times I(l\in W_{ij})-E\left(Y_{ij}^2\mid X_{ij}\right)=O\left(n_i^{-1+\delta}+\frac{w}{n_i}\right),\quad \text{uniformly a.s.,}$$

$\forall \delta > 0$. Hence,

$$T_{2i} = \frac{1}{n_i} \sum_{j=1}^{n_i} O\left(n_i^{-1+\delta} + \frac{w}{n_i}\right) \longrightarrow 0,\ a.s., \quad \text{as } n_i \to \infty,$$

which completes the proof of $Q_{1i} = o(1),\ a.s.$

Similarly, one can further decompose Q_{2i} as

$$\begin{aligned} Q_{2i} &= \frac{1}{n_i} \sum_{j=1}^{n_i} \frac{1}{w_{ij}(w_{ij}-1)} \sum_{l_1 \neq l_2} \left[Y_{il_1}\ Y_{il_2} - E\left(Y_{il_1} Y_{il_2} \mid X_{il_1}, X_{il_2}\right)\right] \times I(l_1, l_2 \in W_{ij}) \\ &+ \frac{1}{n_i} \sum_{j=1}^{n_i} \frac{1}{w_{ij}(w_{ij}-1)} \sum_{l_1 \neq l_2} \left[E\left(Y_{il_1} Y_{il_2} \mid X_{il_1}, X_{il_2}\right) - \left(E\left(Y_{ij} \mid X_{ij}\right)\right)^2\right] I(l_1, l_2 \in W_{ij}) \\ &\equiv T_{3i} + T_{4i}. \end{aligned}$$

We prove $T_{3i} = o(1),\ a.s.$ in Lemma 6.3.2. For T_{4i}, note that under Assumption S5, Equation (5.5.8) tells us that, for n_i large enough, there exists some constant $K > 0$ such that, $\forall l \in W_{ij}$ and $\forall \delta > 0$,

$$\begin{aligned} \left|E\left(Y_{il} \mid X_{il}\right) - E\left(Y_{ij} \mid X_{ij}\right)\right| &\leq K \cdot \left|X_{il} - X_{ij}\right| \\ &\leq K \cdot \left(X_{ij}^U - X_{ij}^L\right) = O\left(n_i^{-1+\delta} + \frac{w}{n_i}\right),\ u.a.s. \end{aligned}$$

where $X_{ij}^U = \max\left(X_{il},\ l \in W_{ij}\right)$ and $X_{ij}^L = \min\left(X_{il},\ l \in W_{ij}\right)$. Hence, $\forall l_1,\ l_2 \in W_{ij}$,

$$\begin{aligned} E\left(Y_{il_1} \mid X_{il_2}\right) E\left(Y_{il_2} \mid X_{il_2}\right) &= \left[E\left(Y_{ij} \mid X_{ij}\right) + O\left(n_i^{-1+\delta} + \frac{w}{n_i}\right)\right]^2 \\ &= \left[E\left(Y_{ij} \mid X_{ij}\right)\right]^2 + o(1),\ u.a.s. \end{aligned}$$

which leads to

$$\begin{aligned} T_{4i} &= \frac{1}{n_i} \sum_{j=1}^{n_i} o(1) \times \frac{1}{w_{ij}(w_{ij}-1)} \sum_{l_1 \neq l_2} I(l_1, l_2 \in W_{ij}) = \frac{1}{n_i} \sum_{j=1}^{n_i} o(1) \\ &\longrightarrow 0,\ a.s., \quad \text{as } n_i \to \infty. \end{aligned}$$

This also completes the proof of (6.3.4).

Furthermore, under Assumptions S2–S4, by SLLN,

$$\frac{1}{n_i}\sum_{j=1}^{n_i}\sigma_i^2\left(X_{ij}\right)-E\left[\sigma_i^2\left(X_{ij}\right)\right]\xrightarrow{a.s.}0,\quad \text{as } n_i\to\infty.$$

Therefore,

$$\widehat{\lambda}_i\ \widehat{a}_i^2\ =\ \int\sigma_i^2(x)\ dG_i(x)+o(1)=\lambda_i\ a_i^2+o(1),\ a.s.$$

which completes the proof of the asymptotic consistency of $\widehat{a}_i^2$ to a_i^2.

To prove the asymptotic consistency of $\widehat{b}_i^2$ to b_i^2, first note that by (6.3.3) and CMT,

$$\left[\frac{1}{n_i}\sum_{j=1}^{n_i}\widehat{\mu}_i\left(X_{ij}\right)\right]^2\ \longrightarrow\ \left[\int\int y\ dF_{ix}(y)\ dG_i(x)\right]^2,\ \text{ as }\min\left(n_i\right)\to\infty.$$

In addition, using the proved fact that $Q_{2i}=o(1),\ a.s.$ and SLLN, one can easily get that, under Assumptions S2–S5, as $\min\left(n_i\right)\to\infty$,

$$\frac{1}{n_i}\sum_{j=1}^{n_i}\sum_{j=1}^{n_i}\left(C_2^{w_{ij}}\right)^{-1}\sum_{l_1\neq l_2}\frac{Y_{il_1}Y_{il_2}}{2}\times I(l_1,\ l_2\in W_{ij})\longrightarrow\int\left[\int y\ dF_{ix}(y)\right]^2\ dG_i(x),$$

which completes the proof.

Proof of Lemma 6.2.5

First by Lemma 6.2.2,

$$\begin{aligned}&\sqrt{N}\left(U_i-\int\int y\,dF_{ix}(y)\ dG_i(x)\right)\\ =\ &\sqrt{N}\left[U_i-E\left(U_i\mid\mathbf{X}\right)\right]+\sqrt{N}\int\int y\ dF_{ix}(y)\ d\left[\widehat{G}_i(x)-G_i(x)\right]+O(N^{-1/2}),\ u.a.s.\end{aligned}$$

Since conditionally on $\mathbf{X}$, $\left[U_i-E\left(U_i\mid\mathbf{X}\right)\right]$ and $\int\int y\ dF_{ix}(y)\ d\left[\widehat{G}_i(x)-G_i(x)\right]$ are independent, by Lemmas 6.2.1 and 6.2.3, we know: conditionally on $\mathbf{X}$,

$$\sqrt{N}\left(U_i-\int\int y\,dF_{ix}(y)\ dG_i(x)\right)\xrightarrow{d}N(0,\ a_i^2+b_i^2),\ \text{ as }\min\left(n_i\right)\to\infty.$$

Since the limiting distribution is the same for all $\mathbf{X}$, this weak convergence also holds unconditionally (by Lemma 5.5.2). Hence,

$$\sqrt{N}\,(\mathbf{U} - \boldsymbol{\mu}_G) \xrightarrow{d} N_k(\mathbf{0}, \boldsymbol{\Sigma}), \quad \text{as } \min(n_i) \to \infty,$$

where $\mathbf{U}$, $\boldsymbol{\mu}_G$, and $\boldsymbol{\Sigma}$ are as defined in (6.2.1), (6.3.1), and (6.2.2), respectively.

Recall that

$$F_{ix}(y) = M(y) + A_i(y) + D_i(y;x), \quad \text{where } \int D_i(y;x)\ dG_i(x) = 0,\ \forall i, \forall y.$$

As a consequence,

$$\begin{aligned}\int\int y\ dF_{ix}(y)\ dG_i(x) &= \int y\ d\left[\int F_{ix}(y)\ dG_i(x)\right] \\ &= \int y\ d\left[\int (M(y) + A_i(y) + D_i(y;x))\ \ dG_i(x)\right] \\ &= \int y\ dM(y) + \int y\ dA_i(y).\end{aligned}$$

Under $H_0(A)$ as defined in (6.0.1),

$$\int\int y\ dF_{ix}(y)\ dG_i(x) \overset{H_0(A)}{=} \int y\ dM(y) \Longrightarrow \mathbf{H}\boldsymbol{\mu}_G \overset{H_0(A)}{=} \mathbf{H}\left(\int y\ dM(y)\right)\mathbf{1}_k = \mathbf{0}_{k-1}.$$

Hence,

$$\begin{aligned}&\sqrt{N}\,\mathbf{HU} \xrightarrow{d} N_{k-1}\left(\mathbf{0},\ \mathbf{H\Sigma H'}\right), \quad \text{under } H_0(A); &(6.3.6)\\ \Longrightarrow\quad &N\,(\mathbf{HU})'\left(\mathbf{H\Sigma H'}\right)^{-1}(\mathbf{HU}) \xrightarrow{d} \chi^2_{k-1}, \quad \text{under } H_0(A),\end{aligned}$$

which completes the proof.

Proof of Theorem 6.2.1

First note that $\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma} \xrightarrow{P} 0$, as $\min(n_i) \to \infty$, since they are $k \times k$ diagonal matrices with diagonal elements $\widehat{a}_i^2 + \widehat{b}_i^2$ and $a_i^2 + b_i^2$ respectively, while $\widehat{a}_i^2 + \widehat{b}_i^2$ are consistent estimators of $a_i^2 + b_i^2$ by Lemma 6.2.4. In addition, note that for any $k \times k$ squared matrix $\mathbf{C}$, the elements of $\mathbf{HCH}'$ are continuous functions of the elements of $\mathbf{C}$, by the Continuous Mapping Theorem (CMT), we get

$$\mathbf{H}\widehat{\mathbf{\Sigma}}\mathbf{H}' - \mathbf{H\Sigma H}' \xrightarrow{P} 0, \quad \text{as } \min(n_i) \to \infty.$$

By applying the CMT again along with the fact that the elements of any inverse matrix with finite dimensions are continuous functions of the elements of the original matrix, we further get

$$\left(\mathbf{H}\widehat{\mathbf{\Sigma}}\mathbf{H}'\right)^{-1} - \left(\mathbf{H\Sigma H}'\right)^{-1} \xrightarrow{P} 0, \quad \text{as } \min(n_i) \to \infty.$$

Then, by the Slutsky's Theorem and (6.3.6), we have that, under $H_0(A)$,

$$\left(\sqrt{N}\mathbf{HU}\right)' \left[\left(\mathbf{H}\widehat{\mathbf{\Sigma}}\mathbf{H}'\right)^{-1} - \left(\mathbf{H\Sigma H}'\right)^{-1}\right] \left(\sqrt{N}\mathbf{HU}\right) \xrightarrow{P} 0, \quad \text{as } \min(n_i) \to \infty,$$

which completes the proof.

Proof of Theorem 6.2.2

First recall that

$$\begin{aligned} \boldsymbol{\mu}_G &= \left(\int\int y \; dF_{1x}(y) \; dG_1(x), \cdots, \int\int y \; dF_{kx}(y) \; dG_k(x)\right)' \\ &= \left(\int y \; dM(y) + \int y \; dA_1(y), \cdots, \int y \; dM(y) + \int y \; dA_k(y)\right)'. \end{aligned}$$

Define

$$\boldsymbol{\mu}_A \equiv \left(\int y \; dA_1(y), \cdots, \int y \; dA_k(y)\right)' \equiv (\mu_{A_1}, \cdots, \mu_{A_k})',$$

where $\mu_{A_i} = \int y \; dA_i(y), \; i = 1, \cdots, k$. It is clear that $\mathbf{H}\boldsymbol{\mu}_G = \mathbf{H}\boldsymbol{\mu}_A$, where $\mathbf{H}$ is the contrast matrix as defined in (6.1.1). In addition, under the local alternatives $H_a(A)$ as defined in (6.2.3), $\mu_{A_i} = (n_i)^{-1/2} \int y \; dB_i(y) \equiv (n_i)^{-1/2}\mu_{B_i}, \; i = 1, \cdots, k$.

The non-centrality of the asymptotic χ^2 distribution is then decided by

$$
\begin{aligned}
& N\left(\mathbf{H}\boldsymbol{\mu}_G\right)'\left(\mathbf{H}\boldsymbol{\Sigma}\mathbf{H}'\right)^{-1}\left(\mathbf{H}\boldsymbol{\mu}_G\right) = N\left(\mathbf{H}\boldsymbol{\mu}_A\right)'\left(\mathbf{H}\boldsymbol{\Sigma}\mathbf{H}'\right)^{-1}\left(\mathbf{H}\boldsymbol{\mu}_A\right) \\
= \ & N\left(\sum_{i=1}^{k}\frac{\mu_{A_i}^2}{c_i^2}\right) - \frac{N}{\left(\sum_{i=1}^{k} 1/c_i^2\right)}\left(\sum_{i=1}^{k}\frac{\mu_{A_i}}{c_i^2}\right)^2, \quad \text{where } c_i^2 = a_i^2 + b_i^2, \qquad (6.3.7) \\
= \ & \left(\sum_{i=1}^{k}\frac{\frac{N}{n_i}\mu_{B_i}^2}{c_i^2}\right) - \frac{1}{\left(\sum_{i=1}^{k} 1/c_i^2\right)}\left(\sum_{i=1}^{k}\frac{\left(\frac{N}{n_i}\right)^{1/2}\mu_{B_i}}{c_i^2}\right)^2 \longrightarrow \zeta, \quad \text{as } \min\left(n_i\right)\to\infty,
\end{aligned}
$$

where ζ is as defined in (6.2.4). For the calculation of (6.3.7), see Section 4.3.

Finally, we complete the proof using the fact that $\widehat{a}_i^2 + \widehat{b}_i^2$ are consistent estimators of $a_i^2 + b_i^2$ (see Lemma 6.2.4) and the arguments stated in the proof of Theorem 6.2.1.

6.3.2 Some Auxiliary Results

Lemma 6.3.1. *Under Assumptions S1, S2 and S4,*

$$
Var\left(\sqrt{N}U_i \mid \mathbf{X}\right) \longrightarrow a_i^2, \ a.s., \quad as \ \min\left(n_i\right)\to\infty,
$$

where a_i^2 are as defined in Lemma 6.2.1.

Proof. Let $\Delta_i = Var\left(\sqrt{N}U_i \mid \mathbf{X}\right)$ and write Δ_i as

$$
\begin{aligned}
\Delta_i &= \frac{N}{w_{i\cdot}^2}\sum_{j=1}^{n_i} Var\left[Y_{ij}\sum_{r=1}^{n_i} I(j\in W_{ir}) \mid \mathbf{X}\right] \\
&= \frac{N}{n_i}\frac{1}{n_i}\left(\sum_{j=1}^{\frac{w-1}{2}} + \sum_{j=\frac{w-1}{2}+1}^{n_i-\frac{w-1}{2}} + \sum_{j=n_i-\frac{w-1}{2}+1}^{n_i}\right)\frac{w_{ij}^2}{(w_{i\cdot}/n_i)^2}\sigma_i^2\left(X_{ij}\right) \\
&= \frac{N}{n_i}\left\{\frac{w^2}{\left(w-\frac{w^2-1}{4n_i}\right)^2}\cdot\frac{1}{n_i}\sum_{j=1}^{n_i}\sigma_i^2\left(X_{ij}\right) + \frac{1}{n_i}\sum_{j=1}^{\frac{w-1}{2}}\frac{\left(\frac{w-1}{2}+j\right)^2 - w^2}{\left(w-\frac{w^2-1}{4n_i}\right)^2}\sigma_i^2\left(X_{ij}\right)\right. \\
&\qquad \left. + \frac{1}{n_i}\sum_{j=n_i-\frac{w-1}{2}+1}^{n_i}\frac{\left(n_i+\frac{w-1}{2}+1-j\right)^2 - w^2}{\left(w-\frac{w^2-1}{4n_i}\right)^2}\sigma_i^2\left(X_{ij}\right)\right\}
\end{aligned}
$$

$$\equiv \quad \frac{N}{n_i}\left\{\Lambda_1+\Lambda_2+\Lambda_3\right\}.$$

Note that

$$\frac{1}{n_i}\sum_{j=1}^{n_i}\sigma_i^2\left(X_{ij}\right)=\frac{1}{n_i}\sum_{j=1}^{n_i}\sigma_i^2\left(X_{ij}^{(o)}\right), \quad \text{for all } \mathbf{X}.$$

Since the original covariate $X_{ij}^{(o)}$ are i.i.d. and $\sigma_i^2(x)$ are uniformly bounded in x (by Assumptions S2 and S4), $\sigma_i^2\left(X_{ij}^{(o)}\right)$, $j=1,\cdots,n_i$, are i.i.d. with finite means. Therefore, by SLLN,

$$\Lambda_1 \longrightarrow \int \sigma_i^2(x)\,dG_i(x), \ a.s., \quad \text{as } \min\left(n_i\right)\to\infty.$$

In addition, by Assumption S4, it is clear that $\Lambda_2=O\left(n_i^{-1}\right)$ and $\Lambda_3=O\left(n_i^{-1}\right)$, u.a.s. Then the remaining part of the proof is completed by Assumption S1.

Lemma 6.3.2. *Under Assumptions in Lemma 6.2.4,*

$$\begin{aligned} T_{3i} &= \frac{1}{n_i}\sum_{j=1}^{n_i}\frac{1}{w_{ij}(w_{ij}-1)}\sum_{l_1\neq l_2}\left[Y_{il_1}\ Y_{il_2}-E\left(Y_{il_1}Y_{il_2}\mid X_{il_1},X_{il_2}\right)\right]\times I(l_1,l_2\in W_{ij}) \\ &\longrightarrow 0, \ a.s., \quad \text{as } n_i\to\infty. \end{aligned}$$

Proof. To simplify the notations, write $E(Y_{ij}\mid X_{ij})$ as μ_{ij} in this proof, and let $Y_{ij}^*=Y_{ij}-\mu_{ij}$, which have zero means: $E(Y_{ij}^*)=0$. Note that

$$\begin{aligned} Y_{il_1}^*Y_{il_2}^* &= \left[Y_{il_1}-\mu_{il_1}\right]\left[Y_{il_2}-\mu_{il_2}\right] \\ &= Y_{il_1}Y_{il_2}-Y_{il_1}\mu_{il_2}-\mu_{il_1}Y_{il_2}+\mu_{il_1}\mu_{il_2} \\ \therefore\ Y_{il_1}Y_{il_2}-\mu_{il_1}\mu_{il_2} &= Y_{il_1}^*Y_{il_2}^*+Y_{il_1}^*\mu_{il_2}+\mu_{il_1}Y_{il_2}^*. \end{aligned}$$

Hence,

$$T_{3i}=R_{1n_i}+R_{2n_i}+R_{3n_i},$$

where

$$R_{1n_i} = \frac{1}{n_i}\sum_{j=1}^{n_i}\frac{1}{w_{ij}(w_{ij}-1)}\sum_{l_1\neq l_2} Y^*_{il_1}Y^*_{il_2}\times I(l_1,l_2\in W_{ij});$$
$$R_{2n_i} = \frac{1}{n_i}\sum_{j=1}^{n_i}\frac{1}{w_{ij}(w_{ij}-1)}\sum_{l_1\neq l_2} Y^*_{il_1}\mu_{il_2}\times I(l_1,l_2\in W_{ij});$$
$$R_{3n_i} = \frac{1}{n_i}\sum_{j=1}^{n_i}\frac{1}{w_{ij}(w_{ij}-1)}\sum_{l_1\neq l_2} \mu_{il_1}Y^*_{il_2}\times I(l_1,l_2\in W_{ij}).$$

To prove $T_{3i} = o(1),\ a.s.$, it is sufficient to prove that all of R_{1n_i}, R_{2n_i} and R_{3n_i} are $o(1),\ a.s.$ It is easy to prove the latter two. For example, $R_{2n_i} = o(1),\ a.s.$ because $\mu_{il_2} = O(1)$, uniformly in l_2, and hence $\exists K_2 > 0$ such that

$$\begin{aligned} R_{2n_i} &= \frac{1}{n_i}\sum_{l_1=1}^{n_i} Y^*_{il_1}\left\{\sum_{|l_2-l_1|<w}\mu_{il_2}\sum_{j=1}^{n_i}\frac{1}{w_{ij}(w_{ij}-1)}I(l_1,l_2\in W_{ij})\right\} \\ &\leq K_2\cdot\frac{1}{n_i}\sum_{l_1=1}^{n_i} Y^*_{il_1} \longrightarrow 0,\ a.s.,\ \text{ by SLLN.} \end{aligned}$$

Similarly, $R_{3n_i} = o(1),\ a.s.$

Now we prove $R_{1n_i} = o(1),\ a.s.$ It suffices to show that $\forall\epsilon > 0$,

$$\sum_{n_i=1}^{\infty} P\left(|R_{1n_i}| > \epsilon\right) < \infty. \tag{6.3.8}$$

Let

$$S_i = n_i\cdot R_{1n_i} = O(1)\sum_{j=1}^{n_i}\sum_{l_1\neq l_2} Y^*_{il_1}Y^*_{il_2}I(l_1,l_2\in W_{ij}),\quad \text{as } \frac{1}{w_{ij}(w_{ij}-1)} \leq \frac{4}{w^2-1},$$

where the window size w is fixed. Then,

$$\begin{aligned} & E(S_i^4|\mathbf{X}) \\ = \ & O(1)\sum_{j_1}\sum_{j_2}\sum_{j_3}\sum_{j_4}\sum_{a_1\neq a_2}\sum_{b_1\neq b_2}\sum_{c_1\neq c_3}\sum_{d_1\neq d_2} E\left(Y^*_{ia_1}Y^*_{ia_2}Y^*_{ib_1}Y^*_{ib_2}Y^*_{ic_1}Y^*_{ic_2}Y^*_{id_1}Y^*_{id_2}|\mathbf{X}\right) \end{aligned}$$

$$
\begin{aligned}
& \times I(a_1, a_2 \in W_{ij_1}) \cdot I(b_1, b_2 \in W_{ij_2}) \cdot I(c_1, c_2 \in W_{ij_3}) \cdot I(d_1, d_2 \in W_{ij_4}) \\
= \quad & O(1) \sum_{a=1}^{n_i-1} \sum_{s=1}^{w_{ia}-1} E(Y_{ia}^{*4} Y_{i,a+s}^{*4} | \mathbf{X}) \times O(w^4) \\
& + O(1) \sum_{a=1}^{n_i-1} \sum_{s_2=1}^{w_{ia}-1} \sum_{s_1<s_2} E(Y_{ia}^{*4} Y_{i,a+s_1}^{*2} Y_{i,a+s_2}^{*2} | \mathbf{X}) \times O(w^4) \\
& + O(1) \sum_{a=1}^{n_i-1} \sum_{s_1 \neq s_2}^{w_{ia}-1} E(Y_{ia}^{*2} Y_{i,a+s_1}^{*4} Y_{i,a+s_2}^{*2} | \mathbf{X}) \times O(w^4) \\
& + O(1) \sum_{a=1}^{n_i-1} \sum_{s_1 \neq s_2}^{w_{ia}-1} E(Y_{ia}^{*3} Y_{i,a+s_1}^{*3} Y_{i,a+s_2}^{*2} | \mathbf{X}) \times O(w^4) \\
& + O(1) \sum_{a=1}^{n_i-1} \sum_{s_1 \neq s_2}^{w_{ia}-1} E(Y_{ia}^{*2} Y_{i,a+s_1}^{*3} Y_{i,a+s_2}^{*3} | \mathbf{X}) \times O(w^4) \\
& + O(1) \left[\sum_{a=1}^{n_i-1} \sum_{s_1=1}^{w_{ia}-1} E(Y_{ia}^{*2} Y_{i,a+s_1}^{*2} | \mathbf{X}) \right] \left[\sum_{b \neq a}^{n_i-1} \sum_{s_2=1}^{w_{ib}-1} E(Y_{ib}^{*2} Y_{i,b+s_2}^{*2} | \mathbf{X}) \right] \times O(w^4) \\
= \quad & O(n_i^2), \; u.a.s.
\end{aligned}
$$

Therefore, $E\left(S_i^4\right) = O\left(n_i^2\right)$.

Then, by Markov's Inequality with order 4, $\forall \epsilon > 0$,

$$
P\left(|R_{1n_i}| > \epsilon\right) \;\leq\; \frac{E\left(R_{1n_i}^4\right)}{\epsilon^4} = \frac{E\left(S_i^4\right)}{n_i^4 \cdot \epsilon^4} \;\leq\; \frac{K_1}{n_i^2} + o\left(n_i^{-2}\right),
$$

for some positive constant K_1 which does not depend on n_i.

Since $\sum_{n_i=1}^{\infty} n_i^{-2} = \pi^2/6 < \infty$, the condition (6.3.8) holds. Therefore, $R_{1n_i} = o(1), \; a.s.$, which completes the proof.

Chapter 7

Summary and Future Work

7.1 Summary

In the context of a nonparametric model for the unbalanced heteroscedastic two-fold nested design, we considered the problem of testing for the sub-class effect. We have established, via theoretical derivations and numerical evidence, that, when the number of sub-classes is large, the classical F-test (CF) procedure is very sensitive to departures from homoscedasticity regardless of whether the model is balanced or unbalanced. Even under homoscedasticity, it is still not asymptotically valid in unbalanced designs with non-normal errors. For this reason, we developed procedures which are asymptotically valid under heteroscedasticity.

We distinguished between what we call general heteroscedasticity and between-classes heteroscedasticity. For the latter case we develop two test procedures, one based on unweighted (UW) and one on weighted observations (WT). The UW procedure is extended also to the case of general heteroscedasticity (HET). Our simulations indicate that the HET procedure is very competitive against the CF and the UW procedures in cases where the last two are valid. Thus, we recommend the procedure HET for general applicability provided $n_{ij} \geq 4$ in all sub-classes. The procedure WT is preferable to HET when the between-classes heteroscedasticity assumption appears tenable and there is either a very large number of sub-classes, or large cell sizes. The procedure CF is preferable to HET when the assumptions of normality and homoscedasticity appear tenable.

A connection is made between testing for no sub-class effect in the nested model, and testing for no covariate effect in nonparametric ANCOVA. We call this naive approach the *non-overlapping windows approach* for analysis of covariance, which leads to the more sophisticated nonparametric ANCOVA approach, called the *overlapping windows approach*, described in the second half of the thesis. Testing for the class effect in the two-fold nested model is also investigated.

In the second half of the thesis, we propose a new method to test for the covariate effect in the context of the fully nonparametric ANCOVA model by capitalizing on the connection, alluded to in the previous paragraph, to testing in nested designs. The basic idea behind the proposed method can be briefly outlined as three steps:

- treat the levels of the categorical group variable as the classes in the two-fold nested model;
- consider each distinct covariate value as a *sub-class* nested in each group/class;
- take a small "window" around each distinct covariate value which consists of the w nearest covariate values nested in the same group to artificially create duplicates within each sub-class.

The key advantage of this idea is allowing the covariate to behave differently and to have possibly different ranges in each group. Of course the aforementioned asymptotic results for the two-fold nested model cannot be directly applied here, due to the dependence of the observations resulting from the overlapping windows. Some new asymptotic tools, based on an application of the projection principle, are thus developed to obtain a new class of quadratic forms, whose asymptotical approximation is then utilized to establish the limiting distributions of the proposed test statistic under the null hypotheses and local alternatives. Our simulations and real data analyses confirm that the proposed test procedure, compared with other existing methodologies, is very powerful and has unique flexibility. Testing for the group effect in the fully nonparametric ANCOVA model is also investigated.

7.2 Future Research

In the near future, we will investigate, among others, the following topics.

Lack-of-Fit Test in ANCOVA

When parametric assumptions can be validated, the usual parametric analysis of covariance is more powerful than nonparametric ANCOVA. A general parametric ANCOVA model is of the form $Y_{ij} = \mu + \alpha_i + g_i(x_{ij}, \theta) + \sigma_i(x_{ij})\,\epsilon_{ij}$, where $g_i(x, \theta)$ are known functions depending on the unknown parameter θ. A simple, yet common, version of this model uses $g_i(x, \theta) = \beta\, x$, while more complicated models allow not only β to differ among groups, but also g_i to be of completely different form in different groups. To test the feasibility of such a model, I plan to propose the statistic for testing for no covariate effect in ANCOVA (see previous paragraph), applied on the residuals $Y_{ij} - g_i(x_{ij}, \hat{\theta})$. The dependence of the residuals should pose some new methodological challenges which I plan to handle.

Generalizations of Fully Nonparametric ANCOVA

As a nature extension, I will focus on the development of *higher-way* ANCOVA model (i.e. more than one factors) with *multiple covariates*. The main challenge may rest on the construction of nearest-neighborhood windows among covariates, which requires appropriate ordering of a multivariate vector. This might be achieved via the incorporation of data depth measurements or the application of some multivariate clustering techniques (such as K-means).

Extension to Designs with Dependent Observations

The nested design considered in my thesis assumes independent observations. In many applications it is reasonable to assume random sub-class effects. Successful application of our methodology to designs with dependent data will also lead to lack-of-fit testing for certain stochastic regression models. One example I plan to consider is lack-of-fit testing in nonlinear time series models such as the nonlinear autoregressive model

$$Y_i = \mu(Y_{i-1}) + \sigma(Y_{i-1})\,\epsilon_i,$$

which includes as special cases the AR $Y_i = \gamma\, Y_{i-1} + \epsilon_i$, ARCH $Y_i = \left(\alpha_0^2 + \alpha_1^2\, Y_{i-1}^2\right)^{1/2}\,\epsilon_i$, EAR $Y_i = \left[u + v\exp\left(-w\, Y_{i-1}^2\right)\right] Y_{i-1} + \epsilon_i$ and TAR $Y_i = a\max\left(Y_i; 0\right) + b\min\left(Y_i; 0\right) + \epsilon_i$ models. These tools can then be utilized to select the most appropriate model for studying diverse issues such as global warming/environmental issues and problems in financial econometrics.

More Statistical Learning: Classification and Clustering

In 2005 I proposed an innovative classification methodology in my master thesis, called *test-based classification (TBC)*, which applies to any dimensional, high or low, data setting provided a suitable k-sample test exists and uses p-values to quantify the similarity of a new observation to each of the training data sets. Because the performance of this classification method relies heavily on the efficiency of the test procedure used, my subsequent efforts have been concentrated in developing test procedures involving high-dimensional alternatives. In particular, I have been working on test procedures in the nested models and in the fully nonparametric ANCOVA models in this thesis. As a natural next step, I plan to incorporate the asymptotic results of testing in this thesis with the TBC rule, and apply the combined methodologies to different classification challenges in bioinformatics, bio-medical research, statistical genetic/genomics studies, and other data mining problems. I expect that these will simulate similar developments in the cluster analysis as well.

Appendix A

Some comments on two procedures: UW and WT

To understand why UW couldn't perform as well as WT under the local alternatives, we try more simulations (not shown here) under difference settings and found that the UW procedure is comparatively sensitive to the departure of homoscedasticity under the local alternatives. More specifically, if the value of σ_3 decreases, given that the original σ_i values used in Table 2.4 are $(\sigma_1,\ \sigma_2,\ \sigma_3,\ \sigma_4,\ \sigma_5) = (1,\ 1,\ 5,\ 1,\ 1)$, the resulting powers of the UW procedure could change dramatically, especially when the value of σ_3 is large. Figure A.1 shows the achieved powers, over 10,000 simulation runs at each different value of σ_3, of the UW and WT procedures, under the normality and $C = 500$, from which one can see that the smaller the value of σ_3, the better powers of UW, while the achieved powers of the WT procedure remains stably high at any values of σ_3. Also note that the difference in the achieved powers between UW and WT becomes less distinguishable as the values of δ_{ij} increases.

Two more simulation studies are done to investigate the performances of procedures UW and WT under homoscedastic designs and under balanced designs. For the former one, we repeat the simulations done in Table 3.2 except let all $\sigma_i = 1$, to compare procedures UW and WT with procedures HOM and CF under homoscedasticity. The corresponding results are shown in Table A.1. Comparing Table A.1 with Table 3.1,

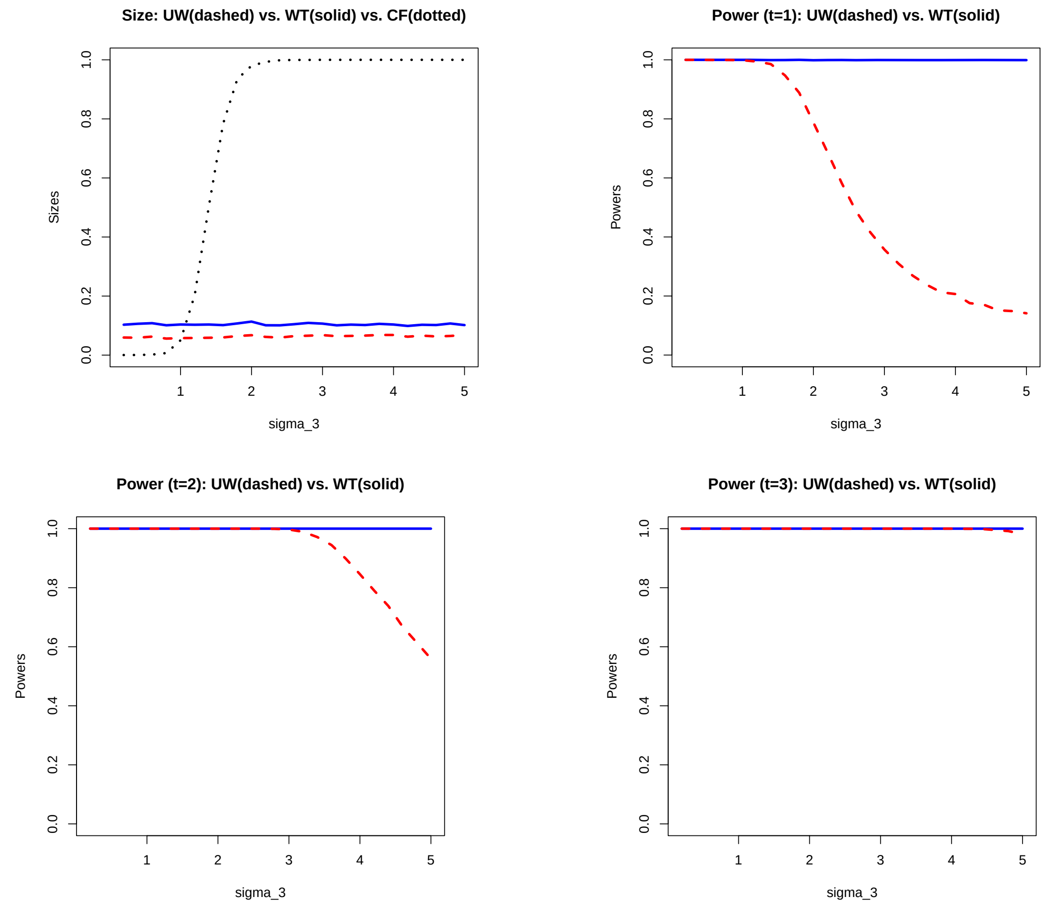

Figure A.1. Sizes and Powers under Normality and between-classes Heteroscedasticity (unbalanced design with $r = 5$ and $C = 500$; $\alpha = 0.05$)

one can easily see that procedure *HOM* still outperform both of *UW* and *WT*, although the differences among them become smaller as C increases. Also note that all three procedures (*HOM*, *UW*, and *WT*) perform better than procedure *CF* in the non-normal cases when C is large enough. Although the theory tells us that the *WT* procedure is asymptotically equivalent to *HOM* under homoscedasticity (see Remark before Corollary 2.2.3), as addressed in the previous paragraph, this asymptotic equivalence highly depends on the accuracy of sample variance estimates, and hence needs either larger numbers of sub-classes or the larger cell sizes to achieve. For instance, if using $(\upsilon_1,\ \upsilon_2,\ \upsilon_3,\ \upsilon_4, \upsilon_5)' = (5,\ 5,\ 5,\ 12,\ 5)$ to generate n_{ij}, our simulations (not shown here) indicates that, three procedures (*HOM*, *UW* and *WT*) could in fact perform very

$\alpha = 0.05$	$\overline{c} = 5$ UW	$\overline{c} = 5$ WT	$\overline{c} = 30$ UW	$\overline{c} = 30$ WT	$\overline{c} = 100$ UW	$\overline{c} = 100$ WT	$\overline{c} = 500$ UW	$\overline{c} = 500$ WT
Normal	0.0902	0.2536	0.0632	0.1408	0.0583	0.1029	0.0499	0.0827
Exponen	0.0988	0.2716	0.0713	0.1362	0.0611	0.0959	0.0564	0.0713
LogNorm	0.0946	0.2624	0.0791	0.1423	0.0726	0.1008	0.0658	0.0679
Mixture	0.0993	0.2649	0.0700	0.1466	0.0546	0.0913	0.0519	0.0698
Multi-d	0.0871	0.2438	0.0662	0.1315	0.0618	0.0975	0.0557	0.0732

Table A.1. Sizes over $10,000$ simulation runs under Homoscedasticity ($r = 5$, unbalanced design)

similarly under the null when C is large.

For the case under balanced designs, we redo the simulation as in Table 3.2 except letting all $c_i = c \equiv C/r$ and all $n_{ij} = 3$ while the number of classes remains the same as before ($r = 5$). The goal of this simulation study is to see whether the balance of the design would affect the performances of three procedures: *UW*, *WT*, and *CF*. The achieved sizes are shown in Table A.2. Recall that under homoscedasticity, if the design

	c=5			c=30		
$\alpha = 0.05$	CF	UW	WT	CF	UW	WT
Normal	0.1763	0.1293	0.2224	0.1878	0.0810	0.0987
Exponen	0.1421	0.1090	0.1956	0.1807	0.0845	0.0991
LogNorm	0.1094	0.0861	0.1679	0.1509	0.0715	0.0954
Mixture	0.1017	0.0783	0.1546	0.1723	0.0806	0.1067
Multi-d	0.1467	0.1107	0.2041	0.1781	0.0847	0.1026
	c =100			c=500		
$\alpha = 0.05$	CF	UW	WT	CF	UW	WT
Normal	0.1931	0.0710	0.0751	0.2029	0.0604	0.0607
Exponen	0.1903	0.0684	0.0720	0.1950	0.0615	0.0660
LogNorm	0.1550	0.0664	0.0787	0.1731	0.0626	0.0622
Mixture	0.1858	0.0733	0.0866	0.1938	0.0616	0.0626
Multi-d	0.1938	0.0712	0.0787	0.1945	0.0602	0.0566

Table A.2. Sizes over $10,000$ simulation runs under between-classes heteroscedastic designs ($r = 5$, balanced design with $c_i = c \equiv C/r$ and $n_{ij} = 3$).

is balanced, the classical F-test procedure is in fact asymptotically valid even when the model is not normal (see Corollary 2.2.1). However, Table A.2 tells us that this asymptotical validity of the classical F-test procedure no longer holds under between-classes heteroscedastic designs, even when the design is balanced and the model is normal. This confirms our findings in Corollary 2.2.2. In addition, comparing Table A.2 with Table 3.2, once can easily see how much the balance of design could improve the performances of procedures CF and WT, especially for the CF procedure, while its influence on the UW procedure is comparatively limited.

To sum, based on simulations done above, we would in general recommend the unweighted test statistic and the corresponding UW procedure for the nested model under heteroscedastic and unbalanced designs, while when C is large enough and the cell sizes are not too small, procedure UW would be a good choice as well.

Appendix B

A simple Time Series Analysis on NADP Data

In Section 2.4.2 and Section 5.4.3, we analyze the rain data from NADP (National Atmospheric Deposition Program). The response variable of interest is the pH levels of precipitation in two towns, Lewiston and Coweeta, while the effect of interest is Time. To ensure the implementations of our methodologies on this data set are valid, a simple time series analysis is performed to check the correlations of observations over time in two locations. The easiest way to complete this mission is to check the plots of the autocorrelation functions (ACF) for two locations, as shown in Figure B.1. The top ACF plot is for Lewiston, while the bottom one is for Coweeta. The confidence limits in the plots assume an MA(k-1), the moving average model of order $k-1$, input for lag k, instead of a white noise input.

As seen in Figure B.1, the absolute values of all autocorrelations at different lags in two locations are less than .20 and all are not significant at level .05, except one. This only exception is the lag 20 autocorrelation of Coweeta. Given that its value is as small as $-.2256$ while the value of the lag is as large as 20, it does not seem meaningful to take this small possible correlation into consideration in the analysis. We hence ignore it in our empirical studies.

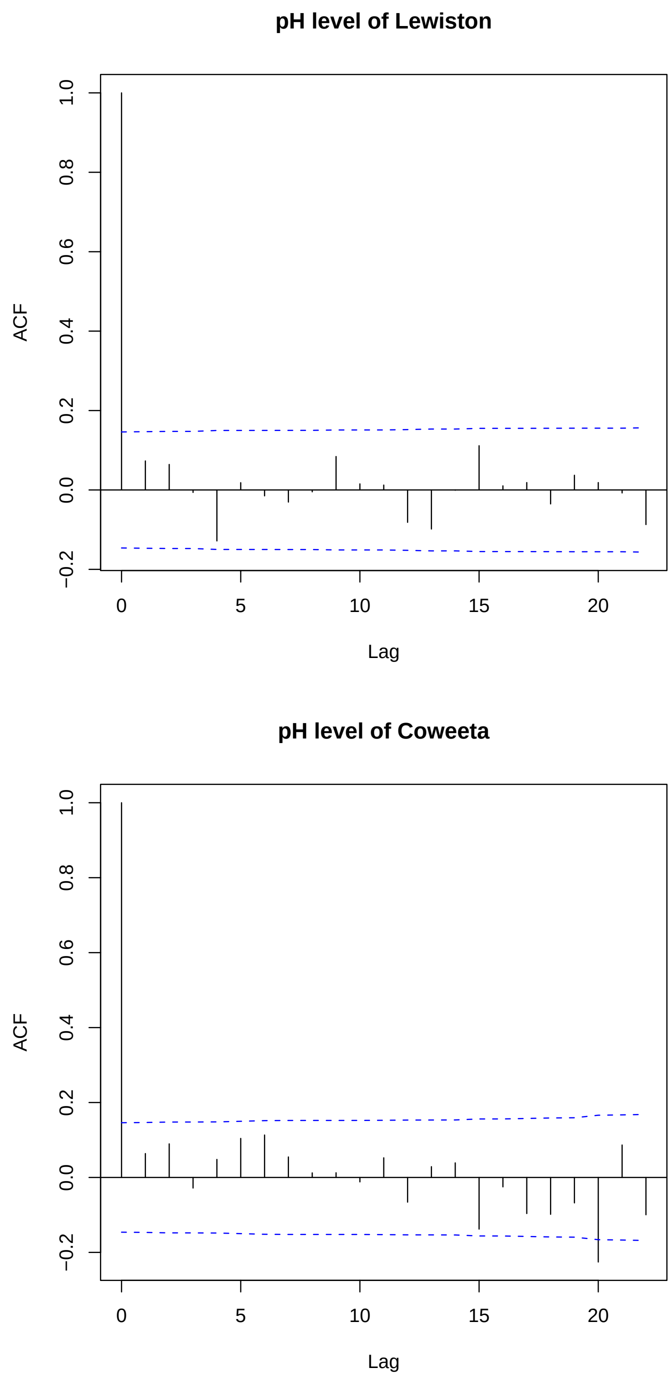

Figure B.1. NADP Data. ACF plots of pH levels in two locations: Lewiston (left) and Coweeta (right).

Bibliography

Akritas, M. G. and Arnold, S. F. (1994), "Fully Nonparametric Hypotheses for Factorial Designs I: Multivariate Repeated Measures Designs," *Journal of the American Statistical Association*, 89, 336–343.

Akritas, M. G. and Arnold, S. F. (2000), "Asymptotics for Analysis of Variance when the Number of Levels is Large," *Journal of the American Statistical Association*, 95, 212–226.

Akritas, M. G., Arnold, S. F., and Brunner, E. (1995), "A Unified Approach to Rank Tests for Mixed Models," Tech. rep., Department of Statistics, The Pennsylvania State University.

— (1997), "Nonparametric Hypotheses and Rank Statistics for Unbalanced Factorial Designs," *Journal of the American Statistical Association*, 92, 259–265.

Akritas, M. G., Arnold, S. F., and Du, Y. (2000), "Nonparametric Models and Methods for Nonlinear Analysis of Covariance," *Biometrika*, 87, 507–526.

Akritas, M. G. and Papadatos, N. (2004), "Heteroscedastic One-Way ANOVA and Lack-of-Fit tests," *Journal of the American Statistical Association*, 99, 368–382.

Andersen, E. B. (1970), "On Fisher's lower bound to asymptotic variances in case of infinitely many nuisance parameters," *Skandinavisk Aktuarietidskrift*, 53, 78–85.

Arnold, S. F. (1980), "Asymptotic Validity of F Tests for the Ordinary Linear Model and the Multiple Correlation Model," *Journal of the American Statistical Association*, 75, 890–894.

— (1981), *The Theory of Linear Models and Multivariate Analysis*, New York: Wiley.

Bathke, A. (2002), "ANOVA for a Large Number of Treatments," *Mathematical Methods of Statistics*, 11, 118–132.

Bathke, A. and Harrar, S. (2008), "Nonparametric methods in multivariate factorial designs for large number of factor levels," *Journal of Multivariate Analysis*, 38, 588–610.

Brownie, C. and Boos, D. D. (1994), “Type I Error Robustness of ANOVA and ANOVA on Ranks when the Number of Treatments is Large,” *Biometrics*, 50, 542–549.

Cleveland, W. S. (1993), *Visualizing Data*, Summit, NJ: Hobart Press.

Fan, J. and Lin, S.-K. (1998), “Test of significance when data are curves,” *Journal of the American Statistical Association*, 93, 1007–1021.

Gupta, A., Harrar, S., and Fujikoshi, Y. (2006), “Asymptotics for testing hypothesis in some multivariate variance components model under non-normality,” *Journal of Multivariate Analysis*, 97, 148–178.

Hall, P. and Hart, J. D. (1990), “Bootstrap Tests for Difference Between Means in Nonparametric Regression,” *Journal of the American Statistical Association*, 85, 1039–1049.

Hall, P., Marron, J. S., and Neeman, A. (2005), “Geometric representation of high dimension, low sample size data,” *Journals of the Royal Statistical Society, Series B*, 67, 427–444.

Hosmer, D. W. and Lemeshow, S. (2000), *Applied Logistic Regression*, New York: Wiley, 2nd ed.

Kulasekera, K. B. (1995), “Testing the Equality of Regression Curves Using Quasi-Residuals,” *Journal of the American Statistical Association*, 90, 1085–1093.

Li, H., Lindsay, B., and Waterman, R. (2003), “Efficiency of projected score methods in rectangular array asymptotics,” *Journals of the Royal Statistical Society, Series B*, 65, 191–208.

McKean, J. W. and Sheather, S. J. (1991), “Small sample properties of robust analyses of linear models based on R-estimates: A survey,” *Institute for Mathematics and Its Applications*, 34, 1–+.

National Atmospheric Deposition Program (NRSP-3). (2009), *Illinois State Water Survey*, NADP Program Office, 2204 Griffith Dr., Champaign, IL 61820.

Neyman, J. and Scott, E. (1948), “Consistent estimates based on partially consistent observations,” *Econometrica*, 16, 1–32.

O’Connor, T. (1998), “Chemical Contaminants in Oysters and Mussels,” Tech. rep., Nationl Oceanic and Atmospheric Administration (NOAA), Silver Spring, MD.

Patel, K. M. and Hoel, D. G. (1973), “A Nonparametric Test for Interaction in Factorial Experiments,” *Journal of the American Statistical Association*, 68, 615–620.

Portnoy, S. (1985), “Asymptotic behavior of M-estimators of p regression parameters when p^2/n is large: II. Normal approximation,” *The Annals of Statistics*, 13, 1403–1417.

R Development Core Team (2009), *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, ISBN 3-900051-07-0.

Rice, J. (1984), "Bandwidth choice for nonparametric regression," *The Annals of Statistics*, 12, 1215–1230.

Ruppert, D., Sheather, S. J., and Wand, M. P. (1995), "An effective bandwidth selector for local least squares regression," *Journal of the American Statistical Association*, 90, 1257–1270.

Simons, G. and Yao, Y. (1999), "Asymptotics when the number of parameters tends to infinity in the Bradley-Terry model for paired comparisons," *The Annals of Statistics*, 27, 1041–1060.

Stute, W. (1982), "The Oscillation Behavior of Empirical Processes," *The Annals of Probability*, 10, 86–107.

Terpstra, J. and McKean, J. (2005), "Rank-based analyses of linear models using R," *Journal of Statistical Software*, 14.

van der Vaart, A. W. (2000), *Asymptotic Statistics*, New York: Cambridge University Press.

Wang, H. and Akritas, M. G. (2004), "Rank Tests for ANOVA with Large Nuumber of Factor Levels," *Journal of Nonparametric Statistics*, 16, 563–590.

Wang, L. and Akritas, M. G. (2006), "Two-way Heteroscedastic ANOVA when the Number of Levels is Larg," *Statistica Sinica*, 16, 1387–1408.

Vita

Shu-Min Liao

Shu-Min Liao was born in Gangshan Township, Kaohsiung County, Taiwan (R.O.C) on March 21st, 1978. She is the daughter of Hui-Sung Liao and Tsui-Feng Liao Tsai. She graduated *Summa Cum Laude* with a Bachelors of Science degree in in Finance from National Chengchi University, Taipei, Taiwan, in May 2001. In the fall of 2003, she enrolled in the graduate program in Statistics at The Pennsylvania State University, and received her Master degree in Science in December 2005. Her research interests include nonparametric theory and methodology, testing against high dimensional alternatives, statistical learning and data mining, experimental designs, linear model, survival analysis, and computational statistics.