NONPARAMETRIC LACK-OF-FIT TESTS IN PRESENCE OF HETEROSCEDASTIC VARIANCES

by

MOHAMMED MAHMOUD GHARAIBEH

B.S., Yarmouk University, Jordan, 1999

M.S., Yarmouk University, Jordan, 2002

AN ABSTRACT OF A DISSERTATION

submitted in partial fulfillment of the

requirements for the degree

DOCTOR OF PHILOSOPHY

Department of Statistics

College of Arts and Sciences

KANSAS STATE UNIVERSITY

Manhattan, Kansas

2014

Abstract

It is essential to test the adequacy of a specified regression model in order to have correct statistical inferences. In addition, ignoring the presence of heteroscedastic errors of regression models will lead to unreliable and misleading inferences. In this dissertation, we consider nonparametric lack-of-fit tests in presence of heteroscedastic variances. First, we consider testing the constant regression null hypothesis based on a test statistic constructed using a k-nearest neighbor augmentation. Then a lack-of-fit test of nonlinear regression null hypothesis is proposed. For both cases, the asymptotic distribution of the test statistic is derived under the null and local alternatives for the case of using fixed number of nearest neighbors. Numerical studies and real data analyses are presented to evaluate the performance of the proposed tests. Advantages of our tests compared to classical methods include: (1) The response variable can be discrete or continuous and can have variations depend on the predictor. This allows our tests to have broad applicability to data from many practical fields. (2) Using fixed number of k-nearest neighbors avoids slow convergence problem which is a common drawback of nonparametric methods that often leads to low power for moderate sample sizes. (3) We obtained the parametric standardizing rate for our test statistics, which give more power than smoothing based nonparametric methods for intermediate sample sizes. The numerical simulation studies show that our tests are powerful and have noticeably better performance than some well known tests when the data were generated from high frequency alternatives. Based on the idea of the Least Squares Cross-Validation (LSCV) procedure of Hardle and Mammen (1993), we also proposed a method to estimate the number of nearest neighbors for data augmentation that works with both continuous and discrete response variable.

NONPARAMETRIC LACK-OF-FIT TESTS IN PRESENCE OF HETEROSCEDASTIC VARIANCES

by

Mohammed Mahmoud Gharaibeh

B.S., Yarmouk University, Jordan, 1999

M.S., Yarmouk University, Jordan, 2002

A DISSERTATION

submitted in partial fulfillment of the

requirements for the degree

DOCTOR OF PHILOSOPHY

Department of Statistics

College of Arts and Sciences

KANSAS STATE UNIVERSITY

Manhattan, Kansas

2014

Approved by:

Major Professor

Haiyan Wang

Copyright

Abstract

It is essential to test the adequacy of a specified regression model in order to have correct statistical inferences. In addition, ignoring the presence of heteroscedastic errors of regression models will lead to unreliable and misleading inferences. In this dissertation, we consider nonparametric lack-of-fit tests in presence of heteroscedastic variances. First, we consider testing the constant regression null hypothesis based on a test statistic constructed using a k-nearest neighbor augmentation. Then a lack-of-fit test of nonlinear regression null hypothesis is proposed. For both cases, the asymptotic distribution of the test statistic is derived under the null and local alternatives for the case of using fixed number of nearest neighbors. Numerical studies and real data analyses are presented to evaluate the performance of the proposed tests. Advantages of our tests compared to classical methods include: (1) The response variable can be discrete or continuous and can have variations depend on the predictor. This allows our tests to have broad applicability to data from many practical fields. (2) Using fixed number of k-nearest neighbors avoids slow convergence problem which is a common drawback of nonparametric methods that often leads to low power for moderate sample sizes. (3) We obtained the parametric standardizing rate for our test statistics, which give more power than smoothing based nonparametric methods for intermediate sample sizes. The numerical simulation studies show that our tests are powerful and have noticeably better performance than some well known tests when the data were generated from high frequency alternatives. Based on the idea of the Least Squares Cross-Validation (LSCV) procedure of Hardle and Mammen (1993), we also proposed a method to estimate the number of nearest neighbors for data augmentation that works with both continuous and discrete response variable.

Table of Contents

List of Figures

List of Tables

Acknowledgments

All praise and thanks are due to Allah (God) for his blessings and making this work possible. My sincere appreciation and deepest gratitude go to my advisor, Dr. Haiyan Wang for her support, encouragement, and guidance during my research. She was always willing to share her knowledge, experience, and time with me. I learned from her a lot not only academically but also things that cannot be gained from books.

Special thanks go to my committee members: Dr. James Higgins, Dr. Gary Gadbury, Dr. Ming-Shun Chen for their support and valuable suggestions. I extend my thanks to Dr. Marianne Korten for serving as a chairperson of my committee.

My thanks go as well to the family of the Department of Statistics: faculty, staff, and students who provided support, help, and friendship throughout my studies at Kansas State University. I am also grateful to my friend and officemate Mohammed Sahtout for his help and encouragement during the journey of my study at K-State.

Last but not least, I will never find enough words to express my gratitude to my beloved father and late mother: (Mahmoud Gharaibeh and Amneh Tubeishat) for their endless prayers, patience, and outstanding support. My sincere appreciation goes to my siblings: Ola, Rana, Ziad, Amal, Reem, Areen, and Hazim for their encouragement and continuous support. Heartfelt thanks go to my wonderful wife Ola Al-Ta'ani and our children: Karam, Rose, and Arz. Their love, sacrifices and ultimate support made this work possible.

Dedication

To my beloved late mother:

Amneh Tubeishat

for all the sacrifices she made to let me achieve my goals.

To my exceptional father:

Mahmoud Gharaibeh

who have always provided me with prayer support, strength and encouragement.

To my wonderful wife:

Ola Al-Ta'ani

and our children

Karam, ***Rose*** and ***Arz***

your love inspires me and the world became a better place with you in my life.

Chapter 1

Introduction

Lack-of-fit test in regression has received a lot of attention recently. The classical lack-of-fit test with replication is given by Fisher (1922). Neill and Johnson (1984) provided a review of linear regression lack-of-fit test procedures in the case of nonreplication. Neill and Johnson (1985) proposed such a test by generalizing the pure error-lack of fit test based on a consistent estimate of the experimental error variance. Based on near replicate clusters, Neill (1988) presented a lack-of-fit test in nonlinear regression for both cases of replication and nonreplication. In all these preceding tests, random errors are assumed to have a constant variance and some assume that errors are normally distributed. Therefore, these tests are only applicable to homoscedastic regression problems. The lack-of-fit test of constant regression is a special case of testing for a nonlinear regression models.

Nonparametric lack-of-fit tests where the constant regression is assumed for the null hypothesis have been considered by many authors. The order selection test by Eubank and Hart (1992), the rank-based order selection test by Hart (2008), and the Bayes sum test by Hart (2009) are among the top few that are intuitive and easy to compute. Alternative version of the order selection test was given in Kutchibhatla and Hart (1996), which has more straightforward calculation of the p-value. A classical textbook review of extensive efforts in nonparametric lack-of-fit tests based on smoothing methods is available in Hart

(1997). Hart (2008) extended the order selection method of Eubank and Hart (1992) to rank-based test under constant variance assumption so that the test statistic is relatively insensitive to misspecification of distributional assumptions. These two order selection tests show excellent performance in low frequency alternatives. However, they may have low power in high frequency alternatives.

In a more recent paper Hart (2009), several new tests based on Laplace approximations were proposed to better handle the high frequency alternatives. In particular, one test with overall good power is the Bayes sum test with statistic of the form

$$B = \sum_{j=1}^{n} \rho_j \exp(n\hat{\phi}_j^2/(2\hat{\sigma}^2)) \text{ with } \rho_j = j^{-2}, \quad j = 1, 2, \ldots, n. \tag{1.0.1}$$

It is a modified cusum statistic with a better use of the sample Fourier coefficients $\hat{\phi}_1, \ldots, \hat{\phi}_n$ arranged in the order of increasing frequency. Hart (2009) gave two versions of critical value approximation, one based on normally generated data and the other based on bootstrap resampling of the residuals under the null hypothesis of constant regression. It is interesting to note that even though the response variable may not be from normal distribution, the normal approximation approach tends to give even higher power than the bootstrap approach. An explanation of this is that the Bayes sum test started with the canonical model that the estimators of the Fourier coefficients are normally distributed and here the sample Fourier coefficients $\hat{\phi}_j = n^{-1}\sum_{i=1}^{n} Y_i cos(\pi j X_i)$, $j = 0, \ldots, n-1$ are approximately normally distributed for large sample size. So the Bayes sum test works well for large sample size and is more powerful than the order selection test and the rank-based order selection test. For intermediate sample size, the two different approximation methods may produce very different coefficients and therefore different empirical distributions. As a result, the two versions of approximation of the Bayes sum test could produce very different results.

Beyond the aforementioned potential different results due to the two approximations of the Bayes sum test critical values, another motivation for us to write this dissertation is that the practical data may have variances vary with the covariate whereas the order selection (OS), rank-based order selection (ROS), and Bayes sum test were derived for homoscedastic regression problems. The scale parameter of the error term is assumed to be a constant in these three tests. Even in such case, different estimators of the scale parameter may be used assuming either the null or alternative hypothesis is true.

To deal with the presence of heteroscedasticity for testing the no-effect null hypothesis, Chen et al. (2001) proposed a new test statistic in addition to bootstrapping the Kuchibhatla and Hart (1996) version of the order selection test. The proposed test statistic of Chen et al. (2001) (denoted by $T_{het,n}$), has the following form

$$T_{het,n} = \max_{1 \le k \le n-1} \frac{1}{k} \sum_{j=1}^{k} \frac{\hat{\phi}_j^2}{\widehat{\mathrm{Var}}(\hat{\phi}_j^2)}, \tag{1.0.2}$$

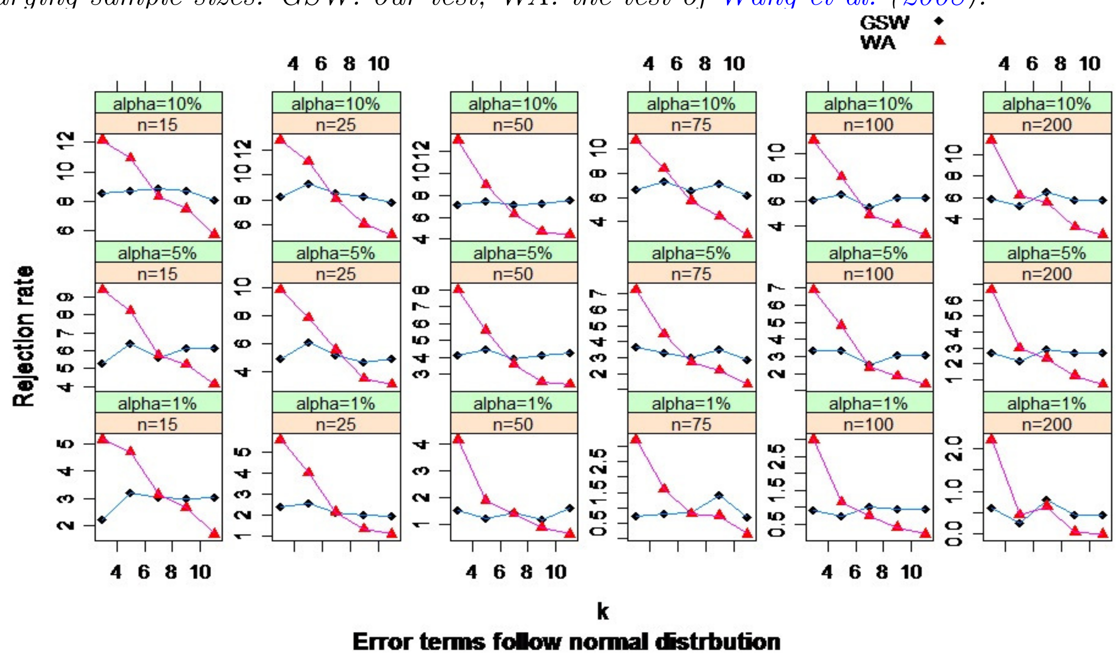
where $\hat{\phi}_1^2, ..., \hat{\phi}_k^2$ are sample Fourier coefficients and $\mathrm{Var}(\hat{\phi}_j^2) = (1/n^2) \sum_{i=1}^{n} \sigma^2(x_i) \cos^2(\pi j x_i)$ which might be estimated by $\widehat{\mathrm{Var}}(\hat{\phi}_j^2) = (1/n^2) \sum_{i=1}^{n} e_i^2 \cos^2(\pi j x_i)$. The approximate sampling distribution of the test statistics was obtained using wild bootstrap method. In the case of heteroscedasticity, Chen et al. (2001) showed that the asymptotic distribution of the Kuchibhatla and Hart (1996) version of the order selection test depends on the unknown variance function of the errors. Furthermore, they showed that the statistic $T_{het,n}$ is more robust than that of Kuchibhatla and Hart (1996) to heteroscedasticity and has better level accuracy. Chen et al. (2001) showed that wild bootstrap technique has an overall good performance in terms of level accuracy and power properties in the case of heteroscedasticity. This test was derived under the null of constant regression. In addition, our experience found that the test could have low power under high frequency alternatives.

In this dissertation, we consider a nonparametric lack-of-fit test of both constant regression and nonlinear regression models in presence of heteroscedastic variances. We construct the test statistics based on a fixed number of k-nearest neighbor augmentation defined through the ranks of the predictor variable. These tests are defined as a difference of two quadratic forms, both of which estimate a common quantity but one under the null hypothesis and the other under the alternatives. The regression function under the null hypotheses appear in one of the two quadratic forms. The asymptotic distributions of the test statistics are obtained under the null and the local alternatives for a fixed number of nearest neighbors. For data from high frequency alternatives, our tests have better power than the available tests.

The idea of using k-nearest neighbor augmentation to construct test statistic was first proposed by Wang and Akritas (2006) for analysis of covariance model, and further used in Wang et al. (2008) for a diagnostic test and in Wang et al. (2010) for a test of independence between a response variable and a covariate in presence of treatments. Wang et al. (2008) defined their test statistic for lack-of-fit test in the constant regression setting. They considered each distinct covariate value as a factor level. Then they augmented the observed data to construct what they called an artificial balanced one-way ANOVA (see section 2.1 for further description of the augmentation). This way of constructing test statistics has great potential to gain power over smoothing based methods. However, we found that the asymptotic variance of the test statistic in Wang et al. (2008) seriously underestimate the true variance for intermediate sample sizes. As a consequence, their type I error changes drastically as the number of nearest neighbors k changes regardless of the error distribution. In particular, their test has highly inflated type I error rates when k is small and becomes very conservative when k gets large. Moreover, type I error of their test depends on the sample size n. Figure 1.1 presents the relationship between the type I error and the number of nearest neighbors used in augmentation for our test and the test of Wang et al. (2008) when the error term was generated from a normal distribution. This gives the typical pattern of the type I error as a function of k with data of different sample sizes. Results for error terms generated from other distributions are presented in Section 3.2.2.

For the test of constant regression null hypothesis, we present an asymptotic variance formula for the test statistic that is very different from that in Wang et al. (2008). In the special case of homoscedastic variance, our derived asymptotic variance contains one more term (a function of k) than that in Wang et al. (2008). This explains the unstable behavior of the type I error pattern of their test. On the other hand, our test has consistent type I error rates across different sample sizes and different k values and they are very close to the nominal alpha levels (see Figure 1.1 and section 3.2.2). A discussion is also given in this dissertation to analytically explain how our test corrects the bias of the test of Wang et al. (2008).

Figure 1.1: *Relationship between type I error and the number of nearest neighbors k for data generated under Model M_0 in section 3.2.2 with error term from normal distribution for varying sample sizes. GSW: our test; WA: the test of Wang et al. (2008).*



Beyond the aforementioned test of constant regression, we also consider the test of nonlinear regression, which was not studied in Wang et al. (2008). Moreover, we give a procedure to estimate the number of nearest neighbors. Our idea extends the Least Squares Cross-Validation (LSCV) procedure of Hardle et al. (1988) in regression to the current k-nearest neighbor augmentation based on ranks. Extensive numerical studies are presented for both the test of constant regression and nonlinear regression cases. The numerical results show that our tests have encouragingly better performance in terms of type I error and power compared to the available tests.

This dissertation is organized as follows. Chapter 2 provides a review of the literature on some available methods of testing lack-of-fit in both cases of constant and nonlinear regression null hypothesis. Chapter 3 considers the nonparametric lack-of-fit test of constant regression in the presence of heteroscedastic variances. Chapter 4 presents the lack-of-fit test of a nonlinear regression model. Chapter 5 introduces the method of selecting the number of nearest neighbors. Chapter 6 provides a summary and suggested plans for future research.

In addition to the lack-of fit setting with continuous response variable and covariate, our test is also valid when the response variable is a discrete or categorical variable. Earlier work in this setting includes Hosmer and Lemesbow (1980) that gave a goodness-of-fit test for multiple logistic regression model, Brown (1982) that proposed a goodness-of-fit test for the logistic model based on score statistics, McCullagh (1986) who studied the conditional distribution of the deviance and Pearson statistics for log-linear model for Poisson data and logistic model for binomial data, Su and Wei (1991) that proposed a lack-of-fit test for the mean function in a generalized linear model based on partial sum of residuals, among others. All of the aforementioned tests assume that the data come from a particular parametric distribution. Nonparametric lack-of-fit test in a general setting without specifying a parametric distribution is still an open topic that deserves further attention.

Chapter 2

Literature Review

In this chapter, a review of some available nonparametric lack-of-fit tests is given. we discuss methods of testing constant regression hypothesis such as the order selection test of Eubank and Hart (1992), rank-based order selection test of Hart (2008), Bayes sum test of Hart (2009) and an ANOVA-type nonparametric diagnostic test for heteroscedastic regression models of Wang et al. (2008). Further, other lack-of-fit test procedures in linear and nonlinear regression are mentioned.

2.1 Order selection test

The order selection test by Eubank and Hart (1992) is one of the most intuitive methods to test the "constant regression" or "no-effect" hypothesis. In this section, a review of order selection test is given. Consider the regression model of the form

$$Y_j = r(x_j) + \epsilon_j, \quad j = 1, 2, ..., n, \tag{2.1.1}$$

where $x_j = (j - 1/2)/n, j = 1, 2, ..., n, r$ is a function that is square integrable over [0,1], and $\epsilon_1, \epsilon_2, ..., \epsilon_n$ are independent and identically distributed with finite fourth moments, $E(\epsilon_j) = 0$, and $Var(\epsilon_j) = \sigma^2$.

The goal is testing the constant regression or "no-effect" null hypothesis which can be specified as:

$$H_0 : r(x) = C \quad \text{for each} \quad x \in [0, 1], \tag{2.1.2}$$

where C is an unknown constant.

Assuming that the function r is piecewise smooth on the interval [0,1], then Fourier series might be used to represent r as the following:

$$r(x) = C + 2\sum_{j=1}^{\infty} \phi_j \cos(\pi j x), \tag{2.1.3}$$

where

$$\phi_j = \int_0^1 r(x)\cos(\pi j x)dx, \quad j = 1, 2, ... \tag{2.1.4}$$

Testing the constant regression or "no-effect" hypothesis (2.1.2) is equivalent to test:

$$H_0 : \phi_1 = \phi_2 = ... = 0. \tag{2.1.5}$$

The function r might be estimated using the following truncated series

$$\hat{r}(x; m) = \hat{C} + 2\sum_{j=1}^{m} \hat{\phi}_j \cos(\pi j x), \tag{2.1.6}$$

where $\hat{\phi}_j = 1/n \sum_{i=1}^{n} Y_i \cos(\pi j x_i)$ for $j = 1, 2, ..., n-1$, $\hat{C} = \sum_{i=1}^{n} Y_i/n$, and m is the smoothing parameter of $\hat{r}(x; m)$ which satisfies $0 \leq m < n$. It is clear that having $m = 0$ strongly supports the null hypothesis of constant regression and for $m \geq 1$ support goes for the alternative hypothesis. Define

$$T_n = \max_{0<m<n} \frac{1}{m} \sum_{j=1}^{m} \frac{2n\hat{\phi}_j^2}{\hat{\sigma}^2}, \tag{2.1.7}$$

where $\hat{\sigma}^2$ is a consistent estimator of σ^2. The order selection test rejects the null hypothesis of constant regression or "no-effect" hypothesis (2.1.2) when the statistic T_n is large. The limiting distribution of T_n is given by

$$\lim_{n\to\infty} P(T_n \leq t) = \exp\left\{-\sum_{j=1}^{\infty} \frac{P(\chi_j^2 > jt)}{j}\right\} \equiv F(t), \tag{2.1.8}$$

where t is the observed value of T_n and χ_j^2 has chi-squared distribution with j degrees of freedom. Using (2.1.8), the P-value for the observed value t is approximately $1 - F(t)$.

This test has a good power in the case of low frequency alternatives. On the other hand, it might have low power at high frequency alternatives for moderate sample sizes. This test is not valid for heteroscedastic regression problems when the error term has variance depends on the covariate.

2.2 Rank-based order selection test

Rank-based order selection test was proposed by Hart (2008). It is an extension to the order selection test of Eubank and Hart (1992). In this test, the same structure of the order selection method was applied to ranks instead of the raw data. To test the "no-effect" hypothesis in (2.1.2), define the following test statistic

$$R_n = \max_{0<m<n} \frac{1}{m} \sum_{j=1}^{m} \frac{2n\tilde{\phi}_j^2}{1/12}, \tag{2.2.1}$$

where

$$\tilde{\phi}_j = \frac{1}{n} \sum_{i=1}^{n} U_i \cos(\pi j x_i) \quad \text{for} \quad j = 1, 2, ..., n-1$$

and

$$U_i = \frac{Rank(Y_i)}{n+1}, \quad i = 1, 2, ..., n.$$

Under the null hypothesis of constant regression and the same assumptions of the order selection test without moment conditions required, the test statistic R_n has the same limiting distribution of T_n which was defined in (2.1.7). That means

$$\lim_{n\to\infty} P(R_n \le r) = \exp\left\{ -\sum_{j=1}^{\infty} \frac{P(\chi_j^2 > jr)}{j} \right\} \equiv G(r), \tag{2.2.2}$$

where r is the observed value of R_n and χ_j^2 has chi-squared distribution with j degrees of freedom.

Similar to order selection test, rank-based order selection test shows a good performance at low frequency alternatives and has low power in the case of high frequency alternatives for moderate sample sizes. Also it is only valid for homoscedastic regression problems.

2.3 Bayes sum test

Bayes sum test is one of several tests based on Laplace approximation proposed in Hart (2009). This test has an overall good power at high frequency alternatives and competitive with available tests at low frequency alternatives. A review of Bayes sum test is given in this section. Consider the model of the form

$$Y_j = \mu(x_j) + \epsilon_j, \quad j = 1, 2, ..., n+1, \tag{2.3.1}$$

where $\mu(x_j)$ is an unknown regression function, $x_j = (j - 1/2)/(n+1)$, $j = 1, 2, ..., n+1$, and $\epsilon_1, \epsilon_2, ..., \epsilon_{n+1}$ are independent and identically distributed with $N(0, 1)$. Fourier coefficients $\phi_1, \phi_2, ...$ were used to characterize the function μ. These Fourier coefficients $\phi_1, \phi_2, ..., \phi_n$ are estimated by the sample Fourier coefficients $\hat{\phi}_1, \hat{\phi}_2, ..., \hat{\phi}_n$ where

$$\hat{\phi}_j = \frac{\sqrt{2}}{(n+1)} \sum_{i=1}^{n+1} Y_i \cos(\pi j x_i), \quad j = 1, 2, ..., n.$$

To test the constant regression or "no-effect" null hypothesis $H_0 : \mu(x) = C$ where C is a constant, Hart (2009) proposed the Bayes sum statistic of the form

$$B_n = \sum_{j=1}^{n} \rho_j \exp\left(\frac{n\hat{\phi}_j^2}{2\hat{\sigma}^2}\right) \quad \text{with} \quad \rho_j = j^{-2}, \quad j = 1, 2, ..., n, \tag{2.3.2}$$

where $\hat{\sigma}^2 = \sum_{j=1}^{n} \hat{\phi}_j^2$. The test statistic B_n is a weighted sum of exponentiated squared Fourier coefficients. It was derived from Bayesian point of view based on posterior probabilities. The posterior probabilities was approximated using Laplace method and the weights $\rho_1, \rho_2, ..., \rho_n$ in (2.3.2) depend on prior probabilities. To approximate the critical value of the test statistic, two methods were given in Hart (2009). One method was done by generating data from normal distribution and the other by using bootstrap resampling from the residuals under the null hypothesis.

Bayes sum test by Hart (2009) is a useful method for lack-of-fit test that can be powerful at high frequency alternative. Furthermore, It is more powerful than order selection test of Eubank and Hart (1992) and rank-based order selection test of Hart (2008) for large sample size. However the variance of the error term is assumed to be a constant in the Bayes sum test. That means it is not applicable when data have variances varying with the covariate (i.e. heteroscedastic regression problem).

2.4 An ANOVA-type nonparametric diagnostic test for heteroscedastic regression models

In this section a discussion of an ANOVA-type nonparametric diagnostic test for heteroscedastic regression models is given. This test was proposed in Wang et al. (2008). Consider the heteroscedastic nonparametric regression model of the form

$$Y_i = m(x_i) + \sigma_i \epsilon_i \quad \text{for} \quad i = 1, 2, ..., n, \tag{2.4.1}$$

where $\sigma^2(.)$ is an unknown variance function, $m(.)$ is an unknown regression function, the errors $\epsilon_1, \epsilon_2, ..., \epsilon_n$ are independent variables with mean 0 and variance 1, and $x_1, x_2, ..., x_n$ are the design points on [0,1] satisfying

$$\int_0^{x_i} r(x)dx = \frac{i}{n} \quad \text{for} \quad i = 1, 2, ..., n,$$

where $r(x)$ is a continuous density on [0,1], . This test can be used for testing the null hypothesis of a constant regression or "no effect" hypothesis:

$$H_0 : m(x) = C \quad \text{for all} \quad x, \tag{2.4.2}$$

where C is an unknown constant. The form of the test statistic is similar to that of the classical F -statistic in analysis of variance. They constructed the test statistic based on the idea of considering each distinct covariate value as a factor level. Augmentation for the observed data have been considered to construct what is called "an artificial" balanced one-way ANOVA. This augmentation is done by considering a window W_i around each covariate value x_i that contains the k nearest covariate values.

Let

$$W_i = \left(j : |\widehat{F}(X_j) - \widehat{F}(X_i)| \leq \frac{k-1}{2n} \right),$$

where $\widehat{F}(x) = n^{-1} \sum_{j=1}^{n} I(X_j \leq x)$ denote the empirical distribution of X. To define the test statistic, consider the structure of balanced one-way ANOVA with n groups and k observations per group. Let $V_{ij}, i = 1, 2, ..., n$ and $j = 1, 2, ..., k$ denote the j^{th} observation in group i. Define

$$MST = \frac{k}{n-1} \sum_{i=1}^{n} \left(\overline{V}_{i\cdot} - \overline{V}_{\cdot\cdot} \right)^2 \quad \textit{and} \quad MSE = \frac{1}{n(k-1)} \sum_{i=1}^{n} \sum_{j=1}^{k} \left(V_{ij} - \overline{V}_{i\cdot} \right)^2. \qquad (2.4.3)$$

Consider the test statistic $MST - MSE$ but with replacing V_{ij} by Y_j , $j \in W_i$ in (2.4.3) for testing the "no effect" hypothesis. This test statistic can be written as a quadratic form $\mathbf{V}'\mathbf{A}\mathbf{V}$ where $\mathbf{V} = (Y_j, j \in W_1, ..., Y_{j'}, j' \in W_n)$ is the vector of all the observations in the artificial one-way ANOVA and

$$\mathbf{A} = \frac{nk_n - 1}{n(n-1)k_n(k_n - 1)} \bigoplus_{i=1}^{n} \mathbf{J}_{k_n} - \frac{1}{n(n-1)k_n} \mathbf{J}_{nk_n} - \frac{1}{n(k_n - 1)} \mathbf{I}_{nk_n},$$

where $\mathbf{I}_d$ is the d-dimensional identity matrix, $\mathbf{J}_d$ is a $d \times d$ matrix with all elements equal to 1, and $\bigoplus_{i=1}^{n}$ is the Kronecker sum. Under the null hypothesis and certain conditions, the quadratic form $(n/k_n)^{1/2}\mathbf{V}'\mathbf{A}\mathbf{V}$ is asymptotically equivalent to the quadratic form

$$(\frac{n}{k_n})^{1/2} (\mathbf{V} - \mathbf{C}\mathbf{1}_N)' \mathbf{A}_d (\mathbf{V} - \mathbf{C}\mathbf{1}_N),$$

which involves a block diagonal matrix $\mathbf{A}_d$ where

$$\mathbf{A}_d = \text{diag}\{\mathbf{B}_1, ..., \mathbf{B}_n\}, \quad \text{where} \quad \mathbf{B}i = \frac{1}{n(k-1)} [\mathbf{J}_k - \mathbf{I}_k]$$

and $N = nk$. This result helps to obtain the asymptotic normality for the test statistic. Under H_0 in (2.4.2) and for fixed k, the asymptotic distribution of the test statistic is

$$n^{1/2}(MST - MSE) \to N \left(0, \frac{2k(2k-1)}{3(k-1)} \tau^2 \right), \qquad (2.4.4)$$

where $\tau^2 = \int_0^1 \sigma^4(x)r(x)dx$, $r(x)$ is a positive continuous density on [0,1] and $\sigma^2(x)$ is the unknown conditional variance function of Y given $X = x$. τ^2 can be estimated by

$$\hat{\tau^2} = \frac{1}{4(n-3)} \sum_{j=2}^{n-2} R_j^2 R_{j+2}^2, \tag{2.4.5}$$

where $R_j = Y_j - Y_{j-1}, \quad j = 2, 3, ..., n$.

Consider the local alternatives $H_1 : m(x) = C + (nk)^{-1/4} g(x)$ where $g(x)$ is a Lipschitz continuous function on $[0, 1]$. For fixed k and under H_1, the asymptotic distribution of the test statistic is

$$(\frac{n}{k})^{1/2}(MST - MSE) \to N\left(\gamma^2, \frac{2(2k-1)}{3(k-1)}\tau^2\right), \tag{2.4.6}$$

where $\gamma^2 = \int_0^1 g^2(t)r(t)dt - (\int_0^1 g(t)r(t)dt)^2$.

2.5 Others

Many lack of fit tests in regression have been proposed in the literature. Some of earlier work will be mentioned in this section. The classical lack of fit test with replication is given by Fisher (1922). A review of linear regression lack of fit test procedures in the case of nonreplication is given by Neill and Johnson (1984). One such tests has been proposed by Neill and Johnson (1985). To find a useful test in the case of nonreplication, Neill and Johnson (1985) generalized the pure error-lack of fit test based on a consistent estimate of the experimental error variance. Using near replicates, this test was compared by Neill and Johnson (1989) with other available tests which is used for assessing the adequacy of a proposed linear regression model in the nonreplication case. In another paper, Neill (1988) presented a lack of fit test in nonlinear regression regardless of replication availability. Most lack of fit tests in the case of nonreplication depend on clustering techniques of the observations. One technique for choosing near replicates based on maximin power clustering criterion and implementation of this criterion are presented in Miller et al. (1998, 1999). In a recent paper, Miller and Neill (2008) proposed several tests based on different groupings of the data for detecting general lack of fit (between-cluster, within-cluster, and mixtures of the two pure types) in both cases of replication and nonreplication. All of the aforementioned tests assume that the random errors have a constant variance and some assume that errors are normally distributed. This means that these tests are only applicable in the case of homoscedastic regression problems.

Some other work has been done on lack-of-fit test include Hausman (1978), Ruud (1984), Newey(1985a; 1985b), Tauchen (1985), White (1982), White (1987), and Bierens (1990). Most of these tests are not consistent for general alternatives. Others proposed consistent nonparametric lack-of-fit test procedures using some smoothing techniques (cf Lee (1988); Yatchew (1992); Eubank and Spiegelman (1990); Hardle and Mammen (1993); Zheng (1996); Horowitz and Spokoiny (2001); Guerre and Lavergne (2005); Song and Du (2011)). Some of them are difficult to compute in addition to complicated conditions that are hard to justify. Some require estimation of the bandwidth parameter and different bandwidth parameter values may give different results. All of the aforementioned methods require the response variable to be continuous. A nonparametric lack of fit test of regression models with heteroscedastic random errors was proposed by Li (1999). However, the test of Li (1999) is not applicable in our case since Li (1999) assumes that the variance is a known function of unknown parameters. In our case the variance function is completely unknown.

Chapter 3

Nonparametric lack-of-fit test of constant regression in presence of heteroscedastic variances

3.1 Theoretical results

3.1.1 The hypotheses and test statistic

Let (X_j, Y_j), $j = 1, \ldots, N$, be a random sample of the random variables (X, Y). Let $f(x)$ and $F(x)$ denote the marginal probability density function and cumulative distribution function of X_j, respectively. Denote $\mathrm{Var}(Y_i|X_i = x) = \sigma^2(x)$ and $\varepsilon_i = Y_i - E(Y_i|X_i)$.

We would like to test whether a given function $m_0(x)$ correctly specifies the conditional mean regression function of Y given X. That is, we are testing the hypothesis:

$$H_0\text{: } E(Y|X = x) = m_0(x) \text{ , where } m_0(.) \text{ is a known function} \tag{3.1.1}$$

against:

$H_1 : E(Y|X = x) = m(x)$, which depends on x through other functions instead of $m_0(.)$.

This formulation works for both continuous and categorical response variable Y. Assume that we do not have duplicate observations for each value of X. In regression settings, the nonlinear conditional mean regression $E(Y|X)$ is often estimated through pooling observations from neighbors by one of the smoothing methods, such as loess, smoothing spline, kernel estimation, etc. For smoothing spline or kernel method, the number of observations in a window essentially needs to go to infinity as the sample size goes to infinity. K-nearest neighbor approach is a popular method for classification but the theory for fixed k is very difficult for general regression. In this research we use fixed number of k-nearest neighbor augmentation to help define a statistic for conducting lack-of-fit test. This augmentation is done for each unique value x_i of the predictor by generating a cell that contains k values of the response Y whose corresponding x values are among the k closest to x_i in rank. We consider k to be an odd number for convenience. Let c denote an index defined by the covariate value X_{j_1} where $c = j_1$ and let $\widehat{F}(x) = N^{-1}\sum_{j=1}^{N} I(X_j \leq x)$ denote the empirical distribution of X. We make the augmentation for each cell c by selecting $k-1$ pairs of observations whose covariate values are among the k closest to X_{j_1} in rank in addition to (X_{j_1}, Y_{j_1}). Let C_c denote the set of indices for the covariate values used in the augmented cell (c). Thus for any pair (X_j, Y_j) to be selected in the augmentation of the cell (c), the difference between the ranks of X_j and X_{j_1} is no more than $(k-1)/2$ if X_{j_1} is an interior point whose rank is between $(k-1)/2$ and $N-(k-1)/2$, i.e., $N|\widehat{F}(X_{j_1}) - \widehat{F}(X_j)| \leq (k-1)/2$. For X_{j_1} whose rank is less than $(k-1)/2$ or greater than $N-(k-1)/2$, the difference between the ranks of X_j and X_{j_1} is no more than $k-1$. This idea was first proposed by Wang and Akritas (2006) and further used in Wang et al. (2008) and Wang et al. (2010) for different problems. Wang et al. (2008) derived their test statistic for lack-of-fit test in the present regression setting by considering each distinct covariate value as a factor level. Then they augmented the observed data by considering a window around each x_i that contains the k_n nearest covariate values to construct what they called an artificial balanced one-way ANOVA. Similar augmentation was considered in Wang et al. (2010) when there are more than one treatment. Their results can not be applied here since the asymptotic variance calculation is ill-defined when there is no treatment factor as in our lack-of-fit setting.

Let $R_{ct}, t = 1, \ldots, k$, denote the augmented response values in cell (c) under the null hypothesis. Define $g_{Nk}(X_1, X_2) = I\left(N|\widehat{F}(X_1)-\widehat{F}(X_2)| \le \frac{k-1}{2}\right)$ to be the indicator function that the difference between the ranks of X_1 and X_2 is no more than $(k-1)/2$. Let B_N and W_N denote the average between-cell and within-cell variations defined as the following:

$$B_N = \frac{k}{N-1}\sum_{c=1}^{N}\left(\overline{R}_{c\cdot} - \overline{R}_{\cdot\cdot}\right)^2 \quad \textit{and} \quad W_N = \frac{1}{N(k-1)}\sum_{c=1}^{N}\sum_{t=1}^{k}\left(R_{ct} - \overline{R}_{c\cdot}\right)^2,$$

where $\overline{R}_{c\cdot} = k^{-1}\sum_{t=1}^{k} R_{ct}$, $\overline{R}_{\cdot\cdot} = N^{-1}\sum_{c=1}^{N}\overline{R}_{c\cdot}$. Note that B_N and W_N can be easily calculated since they resemble the mean squares statistics for an ANOVA model. The calculation is on the augmented data. In most cases in the literature, B_N/W_N is used for constructing the test statistic when B_N has fixed degrees of freedom. However, in our case, the degrees of freedom for B_N is $N-1$, which goes to infinity. Therefore, the statistic typically used in this case is $\sqrt{N}[(B_N/W_N)-1]$ (see Wang and Akritas (2011)), which involves showing that $\sqrt{N}(B_N - W_N)$ converges in distribution to normality and W_N converges in probability to a constant. With augmented data, it is complicated to show that W_N converges in probability. So we use the difference $B_N - W_N$ to construct the test statistic instead of B_N/W_N. This test statistic is similar to that proposed in Wang et al. (2008).

To express B_N and W_N in terms of the original data, we can write

$$\begin{aligned} B_N &= \frac{k}{N-1}\sum_{j_1=1}^{N}\left[\frac{1}{k}\sum_{j=1}^{N} Y_j\, g_{Nk}(X_{j1}, X_j) - \frac{1}{Nk}\sum_{j_2=1}^{N}\sum_{j=1}^{N} Y_j\, g_{Nk}(X_{j2}, X_j)\right]^2 + O_p(N^{-1}) \\ W_N &= \frac{1}{N(k-1)}\sum_{j_1=1}^{N}\sum_{j=1}^{N}\left[Y_j\, g_{Nk}(X_{j1}, X_j) - \frac{1}{k}\sum_{j_2=1}^{N} Y_{j_2}\, g_{Nk}(X_{j1}, X_{j2})\right]^2 + O_p(N^{-1}). \end{aligned}$$

In the next section, the asymptotic distribution of the test statistic will be given.

3.1.2 Asymptotic distribution of the test statistic under the null hypothesis

Asymptotic variance and distribution of our test statistic

Even though the test statistic is easy to calculate, the derivation of the asymptotic distribution is challenging since the augmented data in neighboring cells are correlated. In this section, we give the asymptotic distribution of the test statistic derived with a different strategy than that proposed in Wang et al. (2008) even though we have the same test statistic. To find the asymptotic distribution for our test statistic, we first simplify it by finding its projection. Specifically, define

$$V_{ct} = R_{ct} - E(R_{ct}|\mathbf{X}), \text{ where } \mathbf{X} = (X_1, \ldots, X_N)'. \tag{3.1.2}$$

Then we project B_N onto the space

$$\text{extended span}\{\mathbf{V}_c, c = 1, \ldots, N\}, \text{ where } \mathbf{V}_c = (V_{c1}, \cdots, V_{ck})', \tag{3.1.3}$$

of the form $\sum_{c=1}^{N} a_i g_c(\mathbf{V}_c)$, where $g_c(\mathbf{V}_c)$ is some function that is possibly nonlinear. This projection will help us to split B_N into two terms, one of which includes a summation over c and the other over c and c' for $c \neq c'$:

$$B_N = P_B(\mathbf{V}) + S_B(\mathbf{V}), \text{ where } \mathbf{V}' = (\mathbf{V}_1', \ldots, \mathbf{V}_N'),$$

and

$$P_B(\mathbf{V}) = \frac{k}{N}\sum_{c=1}^{N} \overline{V}_{c\cdot}^2, \quad S_B(\mathbf{V}) = \frac{-k}{N(N-1)}\sum_{c\neq c'}^{N} \overline{V}_{c\cdot}\overline{V}_{c'\cdot}, \tag{3.1.4}$$

where $\overline{V}_{c\cdot} = k^{-1}\sum_{t=1}^{k} V_{ct}$. Then $P_B(\mathbf{V})$ is in the space defined in (3.1.3) and $B_N - W_N = (P_B(\mathbf{V}) - W_N) + S_B(\mathbf{V}) = T_B + S_B(\mathbf{V})$, where

$$
\begin{aligned}
T_B &= \frac{1}{(k-1)N}\sum_{c=1}^{N}\sum_{t\neq t'}^{k} V_{ct}V_{ct'} = \frac{1}{(k-1)N}\sum_{c=1}^{N}\sum_{t\neq t'}^{k}(R_{ct} - E(R_{ct}|\mathbf{X}))(R_{ct'} - E(R_{ct'}|\mathbf{X})) \\
&= \frac{1}{(k-1)N}\sum_{j\neq j'}^{N}(Y_j - E(Y_j|\mathbf{X}))(Y_{j'} - E(Y_{j'}|\mathbf{X}))\sum_{c=1}^{N} I(j\in C_c)I(j'\in C_c) \\
&= \frac{1}{(k-1)N}\sum_{j\neq j'}^{N}(Y_j - E(Y_j|\mathbf{X}))(Y_{j'} - E(Y_{j'}|\mathbf{X}))K_{jj'}, \qquad (3.1.5)
\end{aligned}
$$

and

$$K_{jj'} = \sum_{c=1}^{N} I(j\in C_c)I(j'\in C_c). \qquad (3.1.6)$$

Note that the term in (3.1.5) is closely related to the expected covariance between every pair of response values with correlation induced by their dependence on $\mathbf{X}$. The $K_{jj'}$ in (3.1.6) serves as a weight function which connects the response locally with the empirical distribution function of X. The T_B term in (3.1.5) is more intuitive than $\sqrt{N}(B_N - W_N)$ to evaluate the lack-of-fit. However, T_B can not be calculated from the sample since $E(Y|X)$ is unknown. On the other hand, $\sqrt{N}(B_N - W_N)$ can be directly obtained from the sample.

We assume the following condition to obtain the result under the null hypothesis:

Assumption (A): For all x, suppose that $F(x)$ is differentiable and the fourth conditional central moments of Y_j given X_j are uniformly bounded.

The advantage of using small k instead of a large k can be seen here. Even though $S_B(\mathbf{V})$ is a quadratic form, only nearby cells have correlated observations due to the fixed number of nearest neighbor augmentation. On the other hand, when the number of nearest neighbors tends to infinity, the augmented data in a lot more cells will be correlated and therefore, $S_B(\mathbf{V})$ might diverge and the derivation of the asymptotic distribution will require unnecessarily strong conditions on the magnitude of the correlation. It is straightforward that $S_B(\mathbf{V}) = O_P(N^{-1})$ with small k. Therefore, $\sqrt{N}S_B(\mathbf{V})$ is asymptotically negligible. We state it in Lemma 3.1.1 without proof.

Lemma 3.1.1. *(Projection of B_N) Let $S_B(\mathbf{V})$ be as defined in (3.1.4). If the Assumption (A) is satisfied, then*

$$\sqrt{N} S_B(\mathbf{V}) \xrightarrow{p} 0\,, \quad as\ N \to \infty,$$

where the notation $\xrightarrow{p}$ denotes convergence in probability.

To obtain the asymptotic distribution of the test statistic under the null hypothesis, we work with

$$\sqrt{N} T_B = \frac{\sqrt{N}}{N(k-1)} \sum_{j \neq j'}^{N} (Y_j - E(Y_j|\mathbf{X}))(Y_{j'} - E(Y_{j'}|\mathbf{X}))K_{jj'} \tag{3.1.7}$$

where $K_{jj'}$ is defined in (3.1.6). We first give the large sample behavior of the variance of this term.

Theorem 3.1.2. *Under Assumption (A), $\lambda_N = Var(\sqrt{N} T_B)$ converges as $N \to \infty$ and*

$$\lim_{N \to \infty} \lambda_N = E(\lim_{N \to \infty} \delta_N),$$

where

$$\begin{aligned} \delta_N \ &= \textstyle\sum_{j<j'}^{N} \frac{4\sigma^2(X_j)\sigma^2(X_{j'})}{N(k-1)^2} \left[[k - |j'_* - j_*|]^2 + [k - |j'_* - j_*|] \right. \\ &\quad \left. -2I \left(|j'_* - j_*| \leq \tfrac{k-1}{2} \right) + O(N^{-1}) \right] I(|j'_* - j_*| \leq k-1), \end{aligned} \tag{3.1.8}$$

and j'_*, j_* are the ranks of $X_{j'}$ and X_j among the covariate values $X_1, \ldots, X_N$.

To estimate the asymptotic variance, let j_* be the rank of X_j among all covariate values. Then a consistent estimator of $\lim_{N \to \infty} \lambda_N$ is

$$\widehat{\lambda}_N = \sum_{j<j'}^{N} \frac{4\widehat{\sigma}^2(X_j)\widehat{\sigma}^2(X_{j'})}{N(k-1)^2} \left[[k - |j'_* - j_*|]^2 + [k - |j'_* - j_*|] - 2I\left(|j'_* - j_*| \leq \frac{k-1}{2} \right) \right] I(|j'_* - j_*| \leq k-1),$$

where $\widehat{\sigma}^2(X_j)$ is the sample variance based on the augmented observations for the cell determined by X_j, i.e.,

$$\widehat{\sigma}^2(X_j) = \frac{1}{k-1}\left\{\sum_{l=1}^{N} Y_l^2 g_{Nk}(X_l, X_j) - \frac{1}{k}\left(\sum_{l=1}^{N} Y_l g_{Nk}(X_l, X_j)\right)^2\right\}.$$

Note that $K_{jj'}$ are bounded counts and (3.1.7) is a clean quadratic form as defined in de Jong (1987). The Central Limit Theorem for clean quadratic forms (Proposition 3.2) in de Jong (1987) can be applied to obtain the following result. We skip the details of the proof.

Theorem 3.1.3. *Under H_0 in (3.1.1) and Assumption (A),*

$$\sqrt{N}(B_N - W_N) \xrightarrow{d} N(0, \lim_{N\to\infty} \lambda_N), \ \textit{as } N \to \infty,$$

where the notation $\xrightarrow{d}$ denotes convergence in distribution.

Comparison with the result of Wang et al. (2008)

Wang et al. (2008) expressed their test statistic as a quadratic form $\mathbf{V}'\mathbf{A}\mathbf{V}$ where $\mathbf{V}$ is the vector of all the observations in their artificial one-way ANOVA and

$$\mathbf{A} = \frac{nk_n - 1}{n(n-1)k_n(k_n-1)} \bigoplus_{i=1}^{n} \mathbf{J}_{k_n} - \frac{1}{n(n-1)k_n}\mathbf{J}_{nk_n} - \frac{1}{n(k_n-1)}\mathbf{I}_{nk_n},$$

where $\mathbf{I}_d$ is the d-dimensional identity matrix, $\mathbf{J}_d$ is a $d \times d$ matrix with all elements equal to 1, and $\bigoplus_{i=1}^{n}$ is the Kronecker sum. Then they showed that the quadratic form $(n/k_n)^{1/2}\mathbf{V}'\mathbf{A}\mathbf{V}$ is asymptotically equivalent to another quadratic form involving a block diagonal matrix. This result was used to show the asymptotic distribution of the test statistic. Their approach requires significant effort to derive the quadratic form in matrix form in addition to further difficulty to find its projection. On the other hand, Our approach is straightforward and has a much simpler form compared to Wang et al. (2008).

Wang et al. (2008) showed that the asymptotic variance of their test statistic is

$$\frac{2k(2k-1)}{3(k-1)}\tau^2, \tag{3.1.9}$$

where $\tau^2 = \int_0^1 \sigma^4(x)r(x)dx$, $r(x)$ is a positive continuous density on [0,1], and $\sigma^2(x)$ is the unknown conditional variance function of Y given $X = x$.

Note that our asymptotic variance formula for the test statistic is very different from that in Wang et al. (2008). In the special case of homoscedastic variance (i.e. $\sigma^2(x) = C$, where C is some positive constant) and under H_0, our derived asymptotic variance ($\lim_{N\to\infty} \lambda_N$) contains one more term ($2C^2(k-2)/(k-1)$) than that in Wang et al. (2008) as shown below:

$$
\begin{aligned}
\lambda_N &= E(\delta_N), \text{ where} \\
\delta_N &= \sum_{j<j'}^{N} \frac{4\sigma^2(X_j)\sigma^2(X_{j'})}{N(k-1)^2} \left[[k-|j'_*-j_*|]^2+[k-|j'_*-j_*|]\right. \\
&\quad \left. -2I\left(|j'_*-j_*|\leq \frac{k-1}{2}\right)+O(N^{-1})\right] I(|j'_*-j_*|\leq k-1) \\
&= \sum_{j<j'}^{N} \frac{4C^2}{N(k-1)^2} \left[[k-|j'_*-j_*|]^2+[k-|j'_*-j_*|]\right. \\
&\quad \left. -2I\left(|j'_*-j_*|\leq \frac{k-1}{2}\right)+O(N^{-1})\right] I(|j'_*-j_*|\leq k-1)
\end{aligned} \tag{3.1.10}
$$

If we replace the summation in (3.1.10) over the original sample index j, j' by the summation over the ranks j_*, j'_* and denoting
$M(|j'_*-j_*|) = \left[[k-|j'_*-j_*|]^2+[k-|j'_*-j_*|]-2I\left(|j'_*-j_*|\leq \frac{k-1}{2}\right)+O(N^{-1})\right] I(|j'_*-j_*|\leq k-1)$
and $m = (|j'_*-j_*|)$, we get

$$
\begin{aligned}
\delta_N^* &= \sum_{j_*<j'_*}^{N} \frac{4C^2}{N(k-1)^2} M(|j'_*-j_*|) \\
&= \frac{4C^2}{N(k-1)^2} \sum_{m=1}^{k-1} (N-m)M(m) \\
&= \frac{4C^2}{N(k-1)^2} \sum_{m=1}^{k-1} (N-m) \left[[k-m]^2+[k-m]-2I\left(m\leq \frac{k-1}{2}\right)+O(N^{-1})\right]
\end{aligned}
$$

As $N \to \infty$, we have

$$\begin{aligned}
\lim_{N\to\infty} \delta_N^* &= \frac{4C^2}{(k-1)^2}\sum_{m=1}^{k-1}\left[[k-m]^2+[k-m]-2I\left(m\leq\frac{k-1}{2}\right)\right] \\
&= \frac{2k(2k-1)}{3(k-1)}C^2 + \frac{2(k-2)}{(k-1)}C^2, \qquad (3.1.11)
\end{aligned}$$

and

$$E(\lim_{N\to\infty} \delta_N^*) = \frac{2k(2k-1)}{3(k-1)}C^2 + \frac{2(k-2)}{(k-1)}C^2. \qquad (3.1.12)$$

We can see the asymptotic variance of Wang et al. (2008) in (3.1.9) is equal to the first term in (3.1.12) for the homoscedastic case. As a result, their asymptotic variance is biased and their type I error rate depends on k (see Figures 1.1, 3.4, and 3.5). In a homoscedastic case under the alternative hypothesis, the asymptotic variance of Wang et al (2008) remains to be the same as that under H_0, which completely ignores the dependence of Y on X through the mean regression function. Our variance formula for the test statistic relies on the quadratic function $M(|j_*'-j_*|)$ of pairwise difference in ranks of the observed covariate values. In the heteroscedastic case, the expected value of $\sigma^2(X_j)\sigma^2(X_{j'})$ in our asymptotic variance formula (see (3.1.8)) is less than the τ^2 in Wang et al. (2008)'s asymptotic variance formula (3.1.9). In addition, the product is intermingled with the quadratic function $M(|j_*'-j_*|)$ of pairwise difference in ranks of the observed covariate values and therefore can not be taken outside of the expectation. Further discussions will be given in Section 3.2.2.

3.1.3 Results under local alternatives

Consider the sequence of local alternative conditional expectations $E_N(Y|X=x)$ that approach the conditional expectation of Y given X under the null hypothesis $m_0(x) = E_0(Y|X=x)$ in the order of $N^{-1/4}$. We can write the sequence of local alternative conditional expectations as

$$m(x) = E_N(Y|X=x) = m_0(x) + N^{-1/4}A(x), \qquad (3.1.13)$$

where $A(x)$ is a univariate function of x. This alternative is valid for either discrete or continuous response variable and it allows the data to have different conditional variance under the local alternatives from that under the null. For example, if $Y|X$ has a Poisson distribution with mean $m(x)$ under the alternative, then the variance is $m(x)$ instead of $m_0(x)$. Suppose (X_i, Y_i), $i = 1, \ldots, N$ are observed data under the local alternatives in (3.1.13). Let $\mathbf{Q} = \{Q_{ct};\ c = 1, \ldots, N,\ \ t = 1, \ldots, k\}$ be the augmented response values under this alternative hypothesis. Note that Q_{ct} is equal to the observed response variable whose covariate value is one of the following:

$$\begin{cases} X_{(t)} & \text{if } c < (k-1)/2 \\ X_{(c+t-(k+1)/2)} & \text{if } (k-1)/2 \le c \le N-(k-1)/2 \\ X_{(N-k+t)} & \text{if } c > N-(k-1)/2. \end{cases}$$

Then, Q_{ct} can be written as $Q_{ct} = \varepsilon_{ct} + E(Q_{ct}|\mathbf{X})$, where $E(Q_{ct}|\mathbf{X})$ includes the conditional mean under the null hypothesis and departure from the null at the rate of $N^{-1/4}$. Note that $\varepsilon_{ct} = Q_{ct} - E(Q_{ct}|\mathbf{X})$ satisfies the null hypothesis and can be viewed as the augmented data for $Z_i = Y_i - \big(m_0(X_i) + N^{-1/4}A(X_i)\big)$, whose conditional mean satisfies the null hypothesis but with $Var(Z_i|X_i)$ equal to $Var(Y_i|X_i)$ under the alternative hypothesis. As in previous section, without loss of generality, we can assume $m_0(x)$ is a constant otherwise it can be subtracted from the response variable and work with $(X_i,\ \ Y_i - m_0(X_i))$ directly.

For convenience, define A_{ct} to be the $A(x)$ function evaluated at the covariate value for augmented observation Q_{ct}. Let $\overline{A}_{c\cdot} = k^{-1}\sum_{t=1}^{k} A_{ct}$, $\overline{A}_{\cdot\cdot} = N^{-1}\sum_{c=1}^{N} \overline{A}_{c\cdot}$, $\overline{Q}_{c\cdot} = k^{-1}\sum_{t=1}^{k} Q_{ct}$, $\overline{Q}_{\cdot\cdot} = N^{-1}\sum_{c=1}^{N} \overline{Q}_{c\cdot}$, $\overline{\varepsilon}_{c\cdot} = k^{-1}\sum_{t=1}^{k} \varepsilon_{ct}$, and $\overline{\varepsilon}_{\cdot\cdot} = N^{-1}\sum_{c=1}^{N} \overline{\varepsilon}_{c\cdot}$. Denote $B_N(\mathbf{Q})$ and $W_N(\mathbf{Q})$ to be the average between-cell variations and the average within-cell variations under the local alternatives, respectively. That is,

$$
\begin{aligned}
B_N(\mathbf{Q}) &= k(N-1)^{-1}\sum_{c=1}^{N}\left(\overline{Q}_{c\cdot}-\overline{Q}_{\cdot\cdot}\right)^2\\
&= k(N-1)^{-1}\sum_{c=1}^{N}\left[(\overline{\varepsilon}_{c\cdot}-\overline{\varepsilon}_{\cdot\cdot})+N^{-1/4}\left(\overline{A}_{c\cdot}-\overline{A}_{\cdot\cdot}\right)\right]^2\\
&= k(N-1)^{-1}\sum_{c=1}^{N}\Big[(\overline{\varepsilon}_{c\cdot}-\overline{\varepsilon}_{\cdot\cdot})^2+N^{-1/2}\left(\overline{A}_{c\cdot}-\overline{A}_{\cdot\cdot}\right)^2\\
&\quad+2N^{-1/4}\left(\overline{A}_{c\cdot}-\overline{A}_{\cdot\cdot}\right)(\overline{\varepsilon}_{c\cdot}-\overline{\varepsilon}_{\cdot\cdot})\Big],
\end{aligned}
$$

and

$$
\begin{aligned}
W_N(\mathbf{Q}) &= \{N(k-1)\}^{-1}\sum_{c=1}^{N}\sum_{t=1}^{k}\left(Q_{ct}-\overline{Q}_{c\cdot}\right)^2\\
&= \{N(k-1)\}^{-1}\sum_{c=1}^{N}\sum_{t=1}^{k}\left[(\varepsilon_{ct}-\overline{\varepsilon}_{c\cdot})+N^{-1/4}\left(A_{ct}-\overline{A}_{c\cdot}\right)\right]^2\\
&= \{N(k-1)\}^{-1}\sum_{c=1}^{N}\sum_{t=1}^{k}\Big[(\varepsilon_{ct}-\overline{\varepsilon}_{c\cdot})^2+N^{-1/2}\left(A_{ct}-\overline{A}_{c\cdot}\right)^2\\
&\quad+2N^{-1/4}(\varepsilon_{ct}-\overline{\varepsilon}_{c\cdot})\left(A_{ct}-\overline{A}_{c\cdot}\right)\Big].
\end{aligned}
$$

Then the test statistic can be written as

$$
\begin{aligned}
\sqrt{N}(B_N(\mathbf{Q})-W_N(\mathbf{Q})) = \sqrt{N}\left(k(N-1)^{-1}\sum_{c=1}^{N}(\overline{\varepsilon}_{c\cdot}-\overline{\varepsilon}_{\cdot\cdot})^2-\{N(k-1)\}^{-1}\sum_{c=1}^{N}\sum_{t=1}^{k}(\varepsilon_{ct}-\overline{\varepsilon}_{c\cdot})^2\right)\\
+\quad \Delta_{N,1}+\Delta_{N,2}-\Delta_{N,3}-\Delta_{N,4}, \qquad (3.1.14)
\end{aligned}
$$

where

$$
\Delta_{N,1} = \sqrt{N}k(N-1)^{-1}\sum_{c=1}^{N}\left[N^{-1/2}\left(\overline{A}_{c\cdot}-\overline{A}_{\cdot\cdot}\right)^2\right] \qquad (3.1.15)
$$

$$
\Delta_{N,2} = \sqrt{N}k(N-1)^{-1}\sum_{c=1}^{N}\left[2N^{-1/4}\left(\overline{A}_{c\cdot}-\overline{A}_{\cdot\cdot}\right)(\overline{\varepsilon}_{c\cdot}-\overline{\varepsilon}_{\cdot\cdot})\right] \qquad (3.1.16)
$$

$$
\Delta_{N,3} = \sqrt{N}\{N(k-1)\}^{-1}\sum_{c=1}^{N}\sum_{t=1}^{k}\left[N^{-1/2}\left(A_{ct}-\overline{A}_{c\cdot}\right)^2\right] \qquad (3.1.17)
$$

and

$$\Delta_{N,4} \quad = \quad 2\sqrt{N}\{N(k-1)\}^{-1}\sum_{c=1}^{N}\sum_{t=1}^{k}(\varepsilon_{ct} - \overline{\varepsilon}_{c\cdot})\left(N^{-1/4}\left(A_{ct} - \overline{A}_{c\cdot}\right)\right). \qquad (3.1.18)$$

The following additional condition is needed for the result under the alternative hypothesis:

Assumption (B): Suppose that X_i has bounded support $[a, b]$ and $A(x)$ is locally Lipschitz continuous on $[a, b]$. That is, for each $z \in [a, b]$ there exists an $L > 0$ such that $A(x)$ is Lipschitz continuous on the neighborhood $B_L(z) = \{y \in [a, b] : |y - z| < L\}$. Further, we assume that the fourth central moments of $A(X_i)$ are uniformly bounded.

Before we give the asymptotic distribution of the test statistic under the local alternatives, we state the following results.

Lemma 3.1.4. *Under Assumptions (A) and (B) and as $N \to \infty$,*

$$(1) \qquad \Delta_{N,2} \xrightarrow{p} 0\ ,$$

$$(2) \qquad \Delta_{N,3} \xrightarrow{p} 0\ , \ \textit{and} \quad \Delta_{N,4} \xrightarrow{p} 0\ ,$$

where $\Delta_{N,2}$, $\Delta_{N,3}$ and $\Delta_{N,4}$ are defined in (3.1.16), (3.1.17) and (3.1.18), respectively.

Theorem 3.1.5. *For the sequence of local alternatives $E_N(Y|X)$ in (3.1.13) and under the Assumptions (A) and (B), the limit $\lim_{N\to\infty} \lambda_{NA}$ exists and*

$$\sqrt{N}(B_N(\mathbf{Q}) - W_N(\mathbf{Q})) \xrightarrow{d} N(k\sigma_A^2, \lim_{N\to\infty} \lambda_{NA}),$$

where λ_{NA} is defined similarly as λ_N in Theorem 3.1.2 but with $\sigma^2(X_j)$ calculated under the alternatives in (3.1.13) and

$$\sigma_A^2 = \int_{-\infty}^{\infty} A^2(x) f(x) dx - \left(\int_{-\infty}^{\infty} A(x) f(x) dx\right)^2 = Var(A(X)).$$

Note that λ_N in Theorem 3.1.2 and λ_{NA} in Theorem 3.1.5 share the same formula except that $\sigma^2(X_j) = Var(Y_j|X_j)$ in λ_{NA} needs to be calculated under the alternatives in (3.1.13).

For example, if Y given X has a Bernoulli distribution, then the conditional variance of Y given X under the local alternatives in (3.1.13) is $\sigma^2(x) = E_N(Y|X = x)(1 - E_N(Y|X = x)) = m(x)(1 - m(x))$, which is different from that under the null hypothesis $E_0(Y|X = x)(1 - E_0(Y|X = x)) = m_0(x)(1 - m_0(x))$.

In heteroscedastic regression, it is common in the literature to write $Y_i = m(X_i)+\sigma(X_i)e_i$ with e_i independent of X_i. In this formulation the entire error term $\sigma(X_i)e_i$ is uncorrelated with X_i. In the ideal case that there is no lack-of-fit, such definition is reasonable. However, when there is a lack-of-fit exist because a wrong regression function is specified, the error term still contains some systematic information of $E(Y_i|X_i)$. So it is possible that the error resulting from the specified regression function is still correlated with X_i.

3.2 Examples

3.2.1 Numerical simulation and comparisons

In this section, we present the results of a simulation study conducted to investigate the type I error and power performance of our test. The test has a parameter k to specify the number of nearest neighbors for data augmentation. The inference for our test requires the k to be a small, odd, and positive integer. We report the results for $k = 3$ and 5 and denote them as GSW3 and GSW5, respectively. This is for the user to have an idea of how the test behaves with a given k. Furthermore, we report the results of our test with k selected from 3 and 5 using our considered method that will be explained in Chapter 5 and denote it as GSW. For the GSW applied to each generated data set, the value of the k is selected using $\hat{k}$ in (5.0.1) and our test with parameter $\hat{k}$ is used to obtain the p-value.

For comparison, we also report the corresponding results for the order selection (OS) test of Eubank and Hart (1992), the rank based test (ROS) of Hart (2008), the bootstrap order selection test (BOS) of Chen et al. (2001), and the Bayes sum test of Hart (2009). As argued in Section 7.1 of Hart (1997), evenly spaced design points should be used for calculation of these four test statistics even when they are unevenly spaced. So the generated covariate values in increasing order were replaced by evenly spaced design points on (0,1) for all four tests. For the bootstrap order selection test (BOS), we apply the wild bootstrap algorithm of Chen et al. (2001) based on the residuals $Y_i - \bar{Y}, i = 1, \ldots, n$, and use the test statistic $T_{het,n}$ in (1.0.2) with 1000 bootstrap samples for each replication. For the Bayes sum test, we use the statistic in (2.3.2) that has been reported to have good power from comprehensive simulation study in Hart (2009). For approximation of the p-values of the Bayes sum test Hart (2009) gave two versions of the approximation, one assuming normality (BN) and one using the bootstrap (BB). For the BN, a random sample of the same sample size as the data was generated from the standard normal distribution and the Bayes sum test statistic was calculated from the data so generated regardless of the actual distribution of the response variable. The process was repeated 10,000 times independently and the p-value was obtained based on the empirical distribution of these 10,000 values. For the BB, 10,000 bootstrap samples were drawn from the empirical distribution of the residuals $Y_i - \overline{Y}$, $i = 1, \ldots, n$ rather than the normal distribution and the p-value approximation was carried out similarly. The scale parameter σ^2 for a given data set $Y_1, \ldots, Y_n$ in both BB and BN statistics was estimated by $\hat{\sigma}^2 = (n-2)^{-1} \sum_{i=2}^{n-1} (0.809 Y_{i-1} - 0.5 Y_i - 0.309 Y_{i+1})^2$ as was suggested in Hart (2009).

The values for the covariate X were independently generated from Uniform$(0, 1)$ while the response values were independently generated according to the following five models for $i = 1, \ldots, n$. An intermediate sample size of $n = 50$ was used in all cases.

- Model M_0: $Y_i = 10 + \epsilon_i$;
- Model M_1: $Y_i = 10\, cos(q\pi X_i) + \epsilon_i$;
- Model M_2: $Y_i = 10\, sin(q\pi X_i) + \epsilon_i$;
- Model M_3: $Y_i = e^{-2X_i} cos(q\pi X_i) + \epsilon_i$;
- Model M_4: $Y_i = 0.2 e^{-2X_i} cos(q\pi X_i) + \epsilon_i$,

where q in Models $M_1 - M_4$ represents the frequency. We first consider $q = 8$, which is a higher frequency than the simulation study reported in Hart (2009). A lower q value is considered in later section. The data for the error term ϵ_i in each model were independently generated with one of the four different types of error distribution:

- $\epsilon_i \sim \text{Uniform}(-0.1, 0.1)$;
- $\epsilon_i \sim \text{Normal}(0, 0.02^2)$;
- $\epsilon_i = V_i/30$, where V_i follows t-distribution with 5 degrees of freedom (This case is denoted as $T(5)/30$ in Table 3.1);
- $\epsilon_i = X_i \cdot e_i$ where $e_i \sim \text{Uniform}(-0.1, 0.1)$. This is a heteroscedastic regression model and denoted as $X \cdot U(-0.1, 0.1)$ in Table 3.1.

Model M_0 serves as the null model to obtain the type I error rates for all tests. For each error distribution, the data were generated from Models M_0 through M_4 with sample size $n = 50$ for 2,000 times and the rejection rate (percent of rejections) at significance level 0.05 is reported in Table 3.1.

It can be seen that the type I error estimates for all tests were below or close to the nominal level 0.05 for all models with homoscedastic errors. The Bayes sum test with Bootstrap approximation for the critical value (BB) is very conservative in these cases as the type I error rate is less than 1% (mostly zero). For the heteroscedastic regression model, the variance of the error depends on the covariate while the conditional mean of the response variable given the covariate is a constant under the Model M_0. In this case, the BB test is still conservative whereas all the other tests become liberal.

The columns M_1 to M_4 in Table 3.1 show the power comparison for the different combinations between Models M_1-M_4 and the four types of error distribution. The powers of our test with $k = 3$ (GSW3) and that with $\hat{k}$ (GSW) are very close to each other and higher than all other tests in all cases. The Bayes sum test with normal approximation for the

Table 3.1: *Rejection rate under H_0 and high frequency alternatives with sample size $n = 50$*

		Model (q=8)				
Error	Method	M_0	M_1	M_2	M_3	M_4
	BB	0.00	97.85	98.50	85.40	18.85
	BN	4.65	100	100	99.05	80.80
$U(-0.1, 0.1)$	ROS	4.05	69.60	80.95	63.55	11.10
	OS	4.45	76.25	89.91	69.20	17.00
	BOS	5.35	65.00	91.35	40.15	10.25
	GSW5	3.95	100	100	96.75	81.20
	GSW3	4.00	100	100	99.75	95.30
	GSW	5.45	100	100	99.75	94.20
	BB	0.05	98.20	98.25	87.80	80.40
	BN	4.25	100	100	99.65	99.05
$N(0, 0.02^2)$	ROS	4.10	72.95	78.85	68.30	50.40
	OS	4.25	79.45	86.85	69.15	59.25
	BOS	4.40	65.20	91.40	43.15	35.20
	GSW5	3.55	100	100	96.35	95.15
	GSW3	4.05	100	100	99.85	99.75
	GSW	4.90	100	100	99.70	99.60
	BB	0.00	98.20	98.65	85.80	45.20
	BN	4.25	100	100	99.40	92.25
$T(5)/30$	ROS	4.35	73.05	79.55	66.10	22.15
	OS	4.45	79.60	87.95	68.05	31.70
	BOS	3.50	64.95	91.45	41.35	17.95
	GSW5	2.55	100	100	96.20	89.10
	GSW3	3.00	100	100	99.90	98.20
	GSW	3.00	100	100	99.80	98.15
	BB	0.10	97.85	98.50	86.40	65.50
	BN	7.80	100	100	99.30	95.45
$X \cdot U(-0.1, 0.1)$	ROS	6.10	69.30	81.35	67.30	32.65
	OS	6.95	75.90	89.10	70.50	45.45
	BOS	5.40	65.05	91.40	42.65	24.70
	GSW5	6.50	100	100	96.95	94.55
	GSW3	7.30	100	100	99.80	99.25
	GSW	7.83	100	100	99.70	99.50

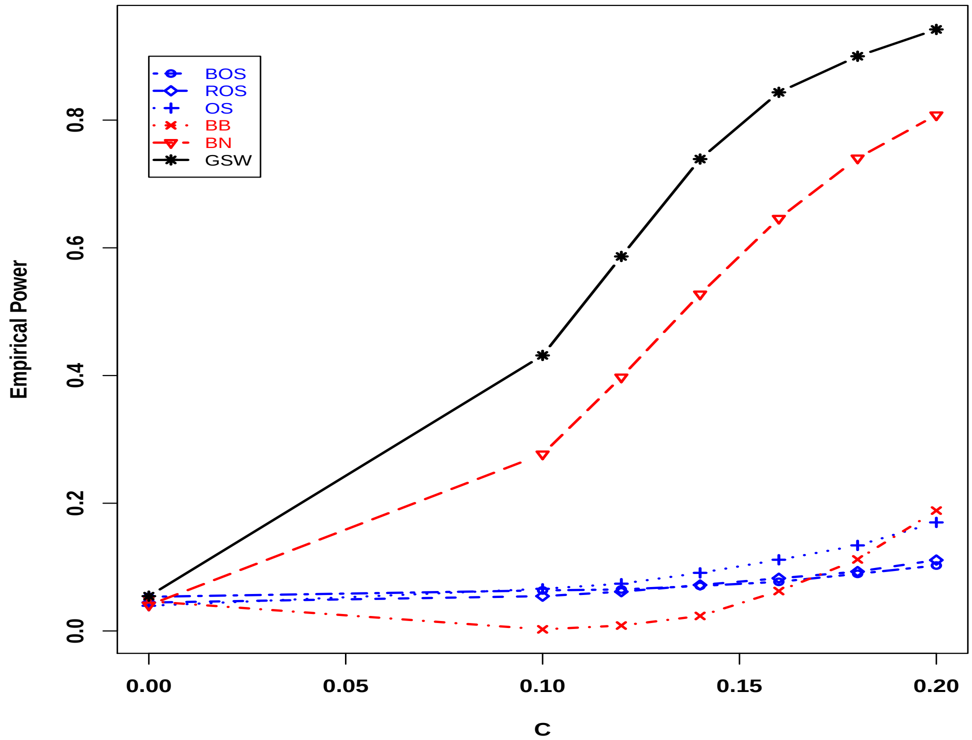
critical value (BN) has power close to our test. The order selection test (OS), the rank based test (ROS), and the bootstrap order selection test (BOS) fall far behind. The low power performance of BOS in the case of high frequency alternatives was mentioned in Chen et al. (2001) and they suggested to use smoothing squared residuals to deal with that but did not give details. For all different types of error distribution, the Bayes sum test with bootstrap approximation for the critical value (BB) has good power for Models M_1 and M_2 whereas power becomes low for Models M_3 and M_4. It is noticeable that the power of our test is 1 for Models M_1 and M_2 for all different types of error distribution and very close to 1 for Models M_3 and M_4. In addition, the power for the order selection test (OS) was slightly higher than that for the rank based test (ROS) in all cases.

Models M_3 and M_4 are similar except that Model M_4 has lower signal to noise ratio than Model M_3. With the lower signal to noise ratio, there are surprisingly big drops in the power for the four tests BB, ROS, OS and BOS. To have a closer look at the performance of all tests in even lower signal to noise ratio cases, we also considered the model $Y_i = Ce^{-2X_i}cos(8\pi X_i) + \epsilon_i$ with $C = 0.1, 0.12, 0.14, 0.16, 0.18$ and $\epsilon_i \sim \text{Uniform}(-0.1, 0.1)$. The empirical power curves are given in Figure 3.1. It is obvious that our test (GSW) has consistently higher power than the other tests.

Above discussions are for high frequency alternatives with $q = 8$ and intermediate sample size $n = 50$. When sample size increases while the frequency stays the same, the power of each test also increases. For sample size of 100 (see the columns 3 to 8 of Table 3.2), the empirical power is 1 for all four tests (Bayes sum test, order selection test, rank based test, and our test) under Models M_1-M_3. The power of BB for Model M_4 is the lowest among all methods for all error distributions. The OS and ROS have power below 1 for the uniform error case. The rest of the tests have power close to 1 for Model M_4. For the heteroscedastic error model, our test has better type I error control (5.3% and 5.5% for GSW3 and GSW5 respectively) than BN, OS, and ROS (more than 7% type I error).

To examine how the power of these tests changes with the sample size, we generated

Figure 3.1: *Power plot for data with low signal to noise ratio.*



data with model $Y_i = N^{-1/4}A(X_i) + \varepsilon_i$, where $A(X_i) = 0.3e^{-2X_i}cos(8\pi X_i) + \epsilon_i$, $\epsilon_i \sim$ Uniform$(-0.1, 0.1)$, for $N = 15, 25, 50, 75, 100, 125, 150, 175, 200, 250$. The GSW is our test with $\hat{k}$ selected from $k = 3$ and 5 based on (5.0.1). The empirical power of these tests are presented in Figure 3.2. It is obvious that the proposed (GSW) test consistently has the highest power over all the sample sizes considered.

It is worth mentioning that for lower frequency alternatives the differences among the power of the four tests will reduce. For example, when $q = 2$ and $n = 50$, the power for Models M_1-M_3 for all tests become 1. For Model M_4, the power of BB is below 1 and the rest of the tests have power close to 1 (see the last six columns of Table 3.2).

Even though the Bayes sum test (BN) showed a comparable performance to our test GSW in some cases, the running time of BN is much longer than GSW. In particular, the average running time from 10,000 runs from BEOCAT cluster machines for GSW is 0.03 second while that for the Bayes sum test is 9.7 seconds. So the GSW is more than 300 times faster than the Bayes sum test.

Figure 3.2: *Power plot for different sample sizes*

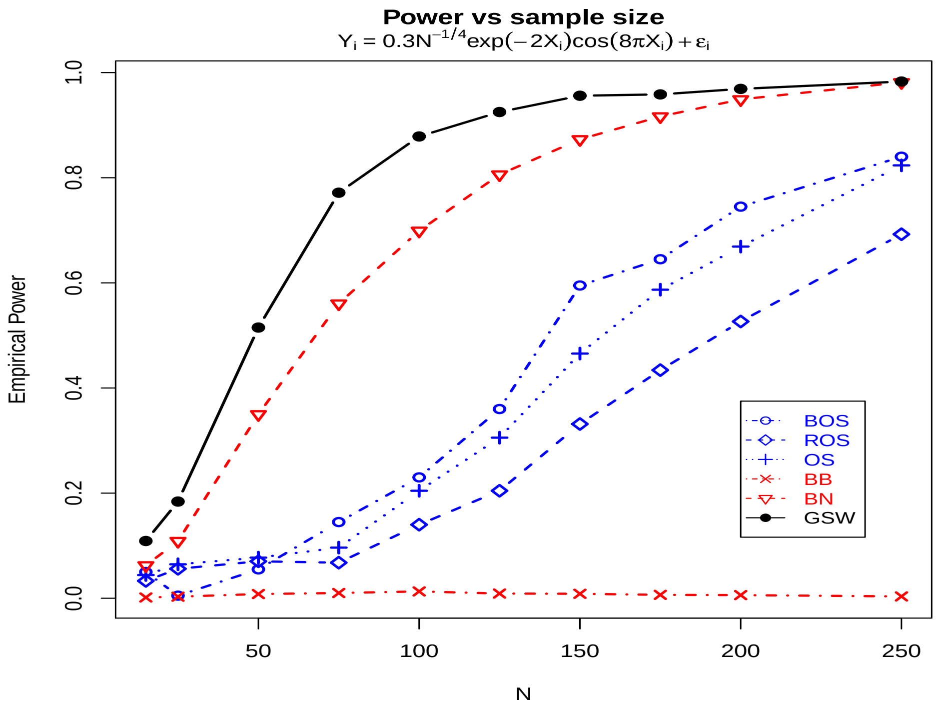


Table 3.2: *Rejection rate under H_0 and high frequency case with sample size n=100 and low frequency case with sample size n=50*

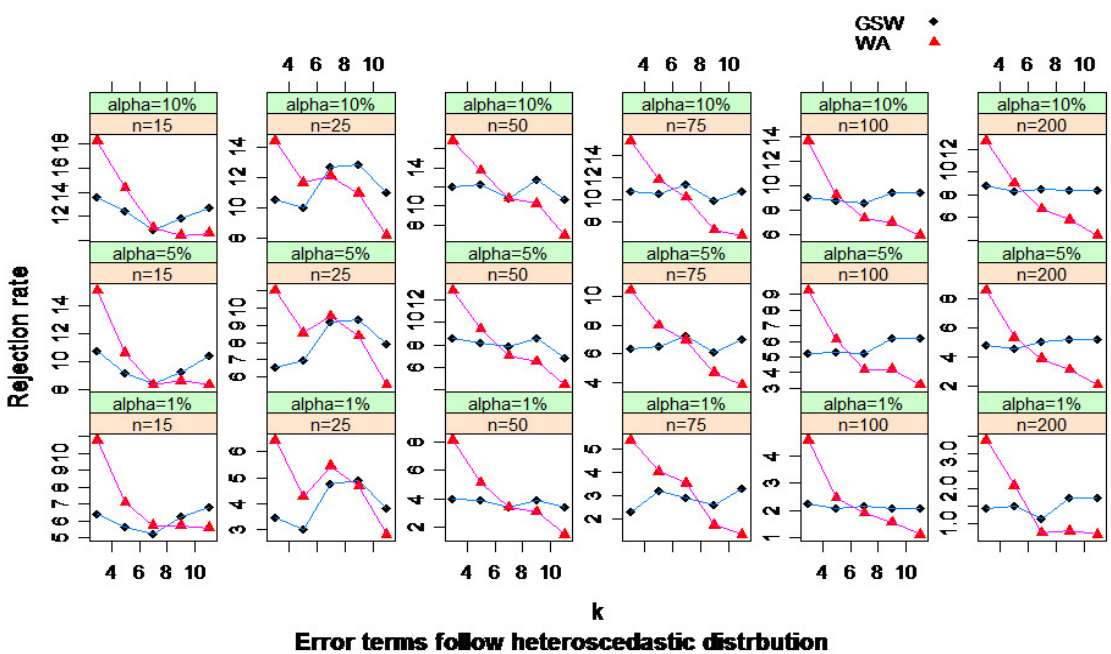
Error	Model	q=8, n=100						q=2, n=50					
		BB	BN	ROS	OS	GSW5	GSW3	BB	BN	ROS	OS	GSW5	GSW3
$U(-0.1, 0.1)$	M_0	0.0	5.2	4.5	4.7	3.5	3.8	0.0	4.6	4.0	4.4	4.0	4.0
	M_4	87.2	100	92.8	95.9	100	100	93.5	100	99.2	99.9	99.8	99.9
$N(0, 0.02^2)$	M_0	0.0	5.2	5.5	5.5	2.9	2.8	0.0	4.2	4.1	4.2	3.5	4.0
	M_4	100	100	100	100	100	100	100	100	100	100	100	100
$T(5)/30$	M_0	0.0	4.8	4.8	5.0	3.0	2.7	0.0	4.2	4.3	4.4	2.5	3.0
	M_4	98.0	100	99.9	99.4	100	100	98.4	100	100	100	100	100
$X \cdot U(-0.1, 0.1)$	M_0	0.0	7.6	7.9	7.0	5.5	5.3	0.1	7.8	6.1	7.0	6.5	7.3
	M_4	99.9	100	99.9	100	100	100	100	100	100	100	100	100

3.2.2 Numerical comparison with Wang et al. (2008)

To explain the difference between the performance of our test and that in Wang et al. (2008), we present the results of numerical studies of the type I error conducted. For each error distribution in the previous subsection (see the first column in Table 3.2), data were generated from Model M_0 with different sample sizes ($n = 15, 25, 50, 75, 100, 200$) and different values of the number of nearest neighbors ($k = 3, 5, 7, 9, 11$) for 20,000 times. The rejection rate at different significance levels ($\alpha = 0.10, 0.05, 0.01$) is given in Figures 1.1, 3.3, 3.4, and 3.5.

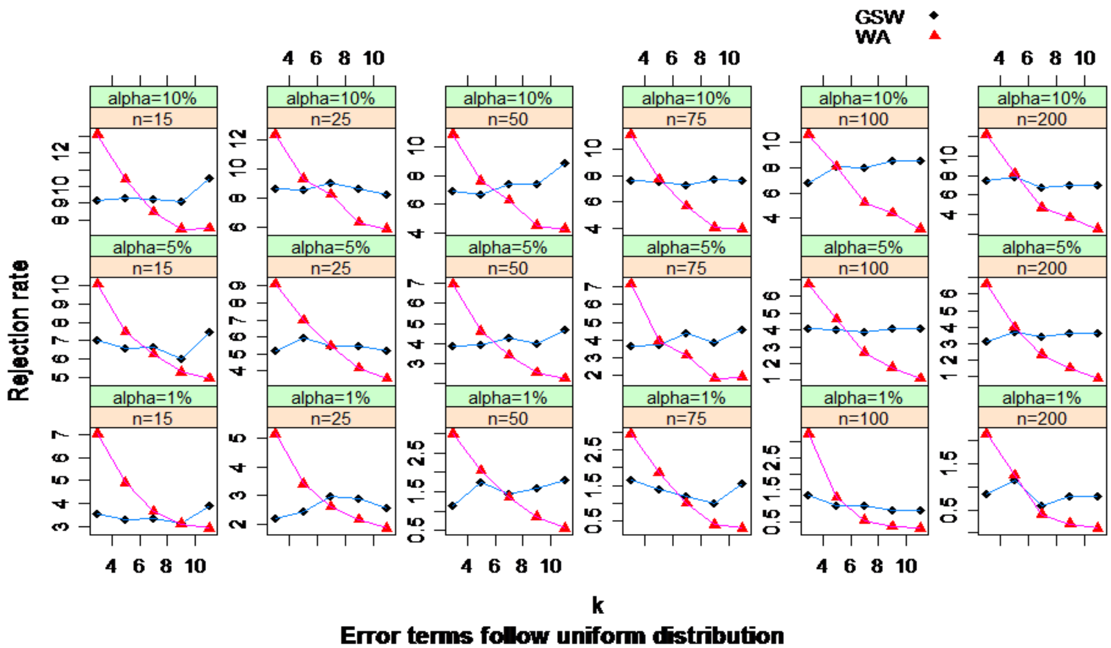
For the heteroscedastic case, both methods are liberal when $n = 15$ and $n = 25$. When the sample size n gets bigger our type I error becomes close to the nominal level but the type I error of Wang et al. (2008) still changes sharply as k varies (See Figure 3.3).

Figure 3.3: *Relationship between type I error and the number of nearest neighbors k for data generated from Model M_0 with heteroscedastic error distribution. GSW: our test; WA: the test of Wang et al. (2008).*



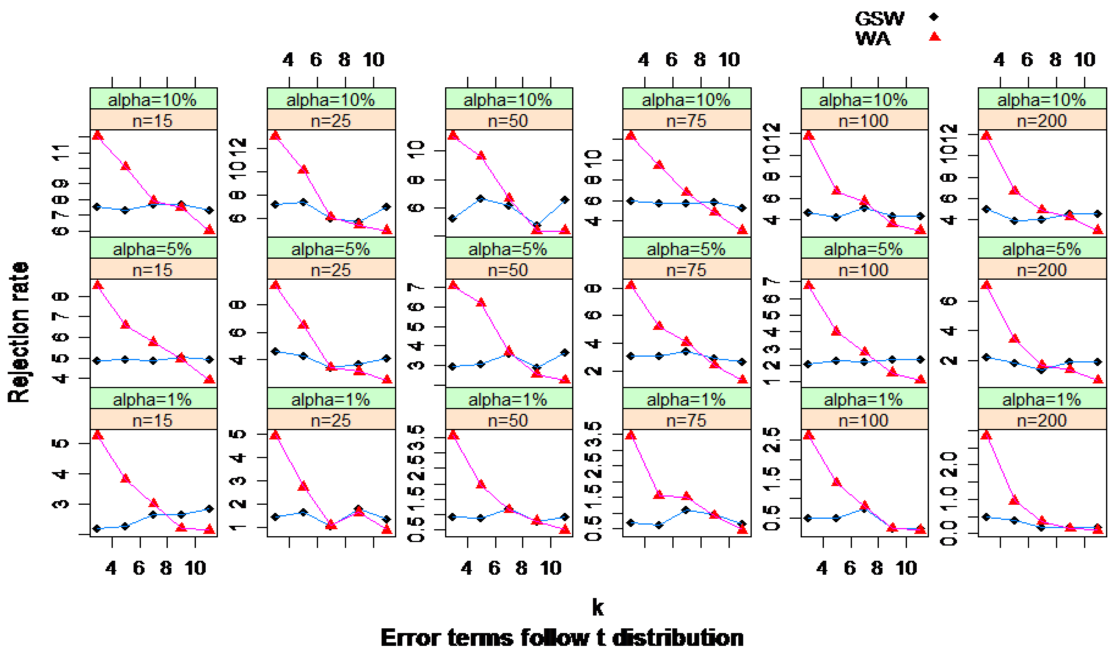
For the models with homoscedastic error distribution, it is obvious that the type I errors of our test are consistent across different sample sizes and different k values and they are very close to the nominal levels as shown in Figures 1.1, 3.4, and 3.5. On the contrary, the test of Wang et al. (2008) has unstable type I error. Their type I error changes drastically as k changes. In particular, their test is very liberal for small k and becomes very conservative when k gets large. Moreover, their type I error depends on the sample size n.

Figure 3.4: *Relationship between type I error and the number of nearest neighbors k for data generated from Model M_0 with uniform error distribution. GSW: our test; WA: the test of Wang et al. (2008).*



Even for homoscedastic case, the calculation of the test statistic depends on the covariate through the nearest neighbor augmentation. However, the asymptotic variance formula (3.1.9) of Wang et al. (2008) does not depend on X when the variance of the response variable is constant. These two facts do not agree with each other. For example, we generated data under Model M_0 ($Y_i = 10 + \epsilon_i$), Model M_2 ($Y_i = 10sin(q\pi X_i) + \epsilon_i$), and Model M_4 ($Y_i = 0.2e^{-2X_i}cos(q\pi X_i) + \epsilon_i$) where $q = 8$ and the error term ϵ_i in each model was generated from two different distributions $\epsilon_i \sim \text{Uniform}(-0.1, 0.1)$ and $\epsilon_i \sim \text{Normal}(0, 0.02^2)$.

Figure 3.5: *Relationship between type I error and the number of nearest neighbors k for data generated from Model M_0 with Student t error distribution. GSW: our test; WA: the test of Wang et al. (2008).*



The process was repeated 2,000 times using $n = 50$ and $k = 3$. The asymptotic variance of Wang et al. (2008) was calculated using formula (3.1.9). The sample variance of 2,000 test statistics was also computed. The results are reported in Table 3.3. Even though the data are under the alternative hypothesis, the asymptotic variance formula of the test statistic of Wang et al. (2008) remains the same as that under the null hypothesis (i.e. formula (3.1.9)). Empirical evidence suggests that this is not right (see Table 3.3). For models with uniform error, the asymptotic variance is 5.55×10^{-5}, whereas the sample variances from 2,000 runs are 5.4×10^{-5}, 4.97, and 4.89×10^{-4} when the data were generated with Models M_0, M_2, and M_4, respectively. They are very different from each other. This happens because their asymptotic variance is biased and does not give the true variance of their test statistic. The bias comes from the missing second term of (3.1.12). Our test statistic is identical to theirs but our calculation of the asymptotic variance depends on the covariate values.

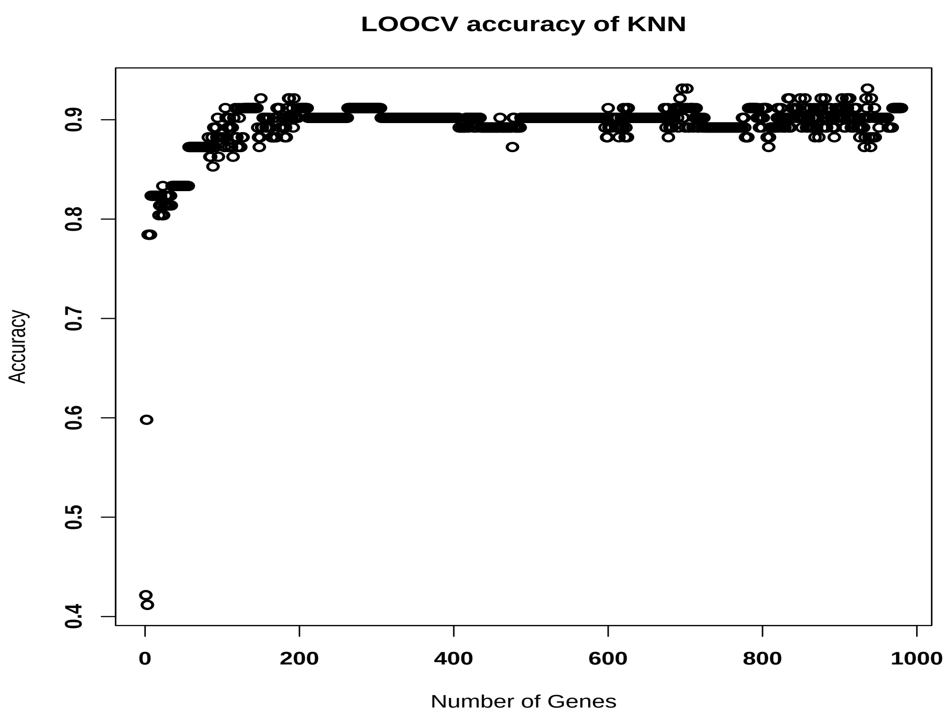
Table 3.3: *Wang et al. (2008)'s asymptotic variance and sample variance of 2,000 test statistic values*

Error	Model	Asymptotic Variance	Sample variance
	M_0		5.40×10^{-5}
$U(-0.1, 0.1)$	M_2	5.55×10^{-5}	4.97
	M_4		4.89×10^{-4}
	M_0		7.85×10^{-7}
$N(0, 0.02^2)$	M_2	8×10^{-7}	0.55
	M_4		5.19×10^{-5}

3.2.3 Application to gene expression data from patients undergoing radical prostatectomy

In this subsection, we present application of our test to gene expression data from patients undergoing radical prostatectomy in order to predict the behavior of Prostate cancer. This data set was collected between 1995 and 1997 at the Brigham and Women's Hospital from 52 tumor and 50 normal prostate samples using oligonucleotide microarrays containing probes for 12600 genes and expressed sequence tags (the data is available at http://www-genome.wi.mit.edu/MPR/prostate). The data shows heterogeneity and has a binary response variable which is the patient outcome (tumor or normal). Applying our test to the expression data from each gene, we identified 980 genes that are significantly associated with the response variable after Bonferroni correction ($p \leq 0.001/12600$). On the other hand, Singh et al. (2002) used permutation test to identify important genes. They found 456 genes whose expression values are significantly correlated with patient outcome ($p \leq 0.001$). Note that the significance declared by Singh et al. (2002) is at 0.001 level without any multiple comparison adjustment. Ours are obtained at the same significance level but with the Bonferroni control which is a very conservative method for multiple comparison adjustment. With such conservative control, we still identified more than twice of the genes than Singh et al. (2002). It is worth mentioning that our test was developed under very general assumptions that are expected to hold true for the microarry data here. These results suggest that our test is much more powerful than the permutation test of Singh et al. (2002). Furthermore, we performed k-nearest neighbor (KNN) classification on the data for the top i genes (i genes with smallest p-values, i= 1,2,....980) to predict the patient outcomes. The leave-one-out cross validation (LOOCV) was used as a validation method. The parameter k in KNN was estimated with the training part of the data in LOOCV procedure by the profile pseudolikelihood method of Holmes and Adams (2003). The leave-one-out accuracy curve with increasing number of selected top i genes is shown in Figure 3.6. We would like to comment that these genes were obtained individually. Our simple application of the test is not meant to find the best combination of genes that have the best classification accuracy. Even under such circumstances, the top genes found with our test give good LOOCV accuracy.

Figure 3.6: *The leave-one-out accuracy curve with increasing number of selected genes.*



3.3 Technical proofs

Proof of Theorem 3.1.2.

We can write $\lambda_N = \mathrm{Var}(\sqrt{N}T_B) = E(\mathrm{Var}(\sqrt{N}T_B|\mathbf{X})) + \mathrm{Var}(\sqrt{N}E(T_B|\mathbf{X}))$.

It is clear that $\mathrm{Var}(\sqrt{N}E(T_B|\mathbf{X})) = 0$ since by the definition of T_B in (3.1.5),

$$E(\sqrt{N}T_B|\mathbf{X}) = E\left(\frac{N^{-1/2}}{(k-1)}\sum_{j\neq j'}(Y_j - E(Y_j|\mathbf{X}))(Y_{j'} - E(Y_{j'}|\mathbf{X}))\middle|\mathbf{X}\right)K_{jj'} = 0 \quad a.s.$$

Therefore, we only need to consider $E(\mathrm{Var}(\sqrt{N}T_B|\mathbf{X}))$ to obtain λ_N. Let $t_{jj'} = (Y_j - E(Y_j|\mathbf{X}))(Y_{j'} - E(Y_{j'}|\mathbf{X}))K_{jj'}$. Then

$$\begin{aligned} N(k-1)^2 E(\mathrm{Var}(\sqrt{N}T_B|\mathbf{X})) &= E\left[\mathrm{Var}\left(\textstyle\sum_{j\neq j'} t_{jj'}|\mathbf{X}\right)\right] = 2E\left(\textstyle\sum_{j\neq j'} E(t^2_{jj'}|\mathbf{X})\right) \\ &= 2\textstyle\sum_{j\neq j'} E\left[\sigma^2(X_j)\sigma^2(X_{j'})K^2_{jj'}\right]. \end{aligned}$$

Let $X_{(j_*)}$ be the order statistic for X_j so that j_* is the rank of X_j among $\{X_{j_1}, j_1 = 1, \ldots, N\}$. Then

$$\begin{aligned} \lambda_N = E(\mathrm{Var}(\sqrt{N}T_B|\mathbf{X})) &= \frac{4}{N(k-1)^2}E\left\{\sum_{j<j'}\sigma^2(X_j)\sigma^2(X_{j'})E\left[K^2_{jj'}\middle| X_j, X_{j'}, j_*, j'_*\right]\right\} \\ &= \frac{4}{N(k-1)^2}E\Bigg\{\sum_{j<j'}\sigma^2(X_j)\sigma^2(X_{j'})[E^2(K_{jj'}|X_j, X_{j'}, j_*, j'_*) \\ &\qquad + \mathrm{Var}(K_{jj'}|X_j, X_{j'}, j_*, j'_*)]\Bigg\}. \end{aligned} \tag{3.3.1}$$

To find the conditional expectation, without loss of generality, assume that $X_j < X_{j'}$, so that $j_* < j'_*$. Let

$$\begin{aligned} \Lambda_{jj'} &= E\left(I(j \in C_c, j' \in C_c)|X_j, X_{j'}, j_*, j'_*\right) \\ &= P(X_j \in C_c, X_{j'} \in C_c|X_j, X_{j'}, j_*, j'_*) = \int_{X_j - L_j}^{X_j + D_j} f(x)dx I(j'_* - j_* \le k-1), \end{aligned}$$

where D_j = the upper $k/2$ spacing and L_j = the lower $(k/2 - (j'_* - j_*))$ spacing from X_j. Applying Taylor's expansion twice, we can write

$$\Lambda_{jj'} = \left\{[F(X_j + D_j) - F(X_j - L_j)] + O_p(N^{-2})\right\} I(j'_* - j_* \le k-1).$$

From the properties of spacings in Pyke (1965), we have

$$E(F(X_j+D_j)-F(X_j-L_j)|X_j,X_{j'},j_*,j'_*)=[k-(j'_*-j_*)]/(N+1)\cdot I(j'_*-j_*\le k-1).$$

Therefore, for $X_{j_1}\neq X_j$ and $X_{j_1}\neq X_{j'}$, we have

$$\begin{aligned} E(\Lambda_{jj'}|X_j,X_{j'},j_*,j'_*) \;=\; & \big\{[k-(j'_*-j_*)-2I(j'_*-j_*\le (k-1)/2)]/(N+1)+O_p(N^{-2})\big\} \\ & \times I(j'_*-j_*\le k-1); \end{aligned} \tag{3.3.2}$$

if $X_{j_1}=X_j$ (or symmetrically $X_{j_1}=X_{j'}$), then

$$\Lambda_{jj}=I(j'_*\in C_{X_{(j_*)}})=I\left(j'_*-j_*\le (k-1)/2\right). \tag{3.3.3}$$

Collecting terms from (3.3.2) and (3.3.3), we have

$$E\left(K_{jj'}|X_j,X_{j'},j_*,j'_*\right)=\big(k-(j'_*-j_*)+O_p(N^{-1})\big)\,I(j'_*-j_*\le k-1). \tag{3.3.4}$$

Now consider the conditional variance. Note that when $X_c\in\{X_j,X_{j'}\}$, the term in $K_{jj'}$ is a constant. Therefore,

$$\begin{aligned} \mathrm{Var}\left(K_{jj'}|X_j,X_{j'},j_*,j'_*\right) &= \mathrm{Var}\Big(\textstyle\sum_{c=1}^{N} I(j\in C_c)I(j'\in C_c)I(X_c\notin\{X_j,X_{j'}\})\Big|\, X_j,X_{j'},j_*,j'_*\Big) \\ &= \textstyle\sum_{c_1=1}^{N}\sum_{c_2=1}^{N}\{E[I(j\in C_{c_1})I(j'\in C_{c_1})I(j\in C_{c_2})I(j'\in C_{c_2})|X_j,X_{j'},j_*,j'_*] \\ &\quad -E[I(j\in C_{c_1})I(j'\in C_{c_1})|X_j,X_{j'},j_*,j'_*]E[I(j\in C_{c_2})I(j'\in C_{c_2})|X_j,X_{j'},j_*,j'_*]\} \\ &\quad \times I(X_{c_1}\notin\{X_j,X_{j'}\})I(X_{c_2}\notin\{X_j,X_{j'}\}) \\ &= \textstyle\sum_{c=1}^{N} E\left[I(j\in C_c)I(j'\in C_c)I(X_c\notin\{X_j,X_{j'}\})|\, X_j,X_{j'},j_*,j'_*\right] \\ &\quad - \textstyle\sum_{c=1}^{N}[E(I(j\in C_c)I(j'\in C_c)|X_j,X_{j'},j_*,j'_*)]^2 I(X_c\notin\{X_j,X_{j'}\}), \end{aligned}$$

where the last equality is due to the fact that the indicator functions involving c_1 and c_2 are conditionally independent when $c_1\neq c_2$ and neither c_1, c_2 is X_j or $X_{j'}$. Plugging (3.3.2) through (3.3.4) into the right hand side of the equation above, we obtain

$$\mathrm{Var}(K_{jj'}|X_j,X_{j'},j_*,j'_*)=\left[(k-(j'_*-j_*))-2I\left(j'_*-j_*\le\frac{k-1}{2}\right)+O_p(N^{-1})\right]I(j'_*-j_*\le k-1). \tag{3.3.5}$$

Putting (3.3.4) and (3.3.5) into (3.3.1), we have

$$\lambda_N = \sum_{j<j'}^{N} E\left\{\frac{4\sigma^2(X_j)\sigma^2(X_{j'})}{N(k-1)^2}\left[[k-(j'_*-j_*)]^2+[k-(j'_*-j_*)]-2I\left(j'_*-j_*\leq\frac{k-1}{2}\right)\right.\right.$$
$$\left.\left.+O_p(N^{-1})\right]\, I(j'_*-j_*\leq k-1)\right\}.$$

Next, we will show that the limit of λ_N exists. Note that

$$\begin{aligned}
\lambda_N &= E(\delta_N), \text{ where} \\
\delta_N &= \sum_{j<j'}^{N}\frac{4\sigma^2(X_j)\sigma^2(X_{j'})}{N(k-1)^2}\left[[k-|j'_*-j_*|]^2+[k-|j'_*-j_*|]\right. \\
&\quad \left.-2I\left(|j'_*-j_*|\leq\frac{k-1}{2}\right)+O_p(N^{-1})\right] I(|j'_*-j_*|\leq k-1).
\end{aligned}$$

It is clear that $[[k-|j'_*-j_*|]^2]\, I(|j'_*-j_*|\leq k-1)$ and $[k-|j'_*-j_*|]\, I(|j'_*-j_*|\leq k-1)$ are both at least 1, therefore, $\left[[k-|j'_*-j_*|]^2+[k-|j'_*-j_*|]-2I\left(|j'_*-j_*|\leq\frac{k-1}{2}\right)\right] I(|j'_*-j_*|\leq k-1)$ is nonnegative. Consequently, δ_N is a summation of nonnegative terms.

Under Assumption (A), the conditional variance of Y_j given X_j is uniformly bounded (i.e. there exists a constant $C>0$ such that $\sigma^2(X_j)\leq C$ for all j). We have

$$\begin{aligned}
\delta_N &= \sum_{j<j'}^{N}\frac{4\sigma^2(X_j)\sigma^2(X_{j'})}{N(k-1)^2}\left[[k-|j'_*-j_*|]^2+[k-|j'_*-j_*|]\right. \\
&\quad \left.-2I\left(|j'_*-j_*|\leq\frac{k-1}{2}\right)+O_p(N^{-1})\right] I(|j'_*-j_*|\leq k-1) \\
&\leq \sum_{j<j'}^{N}\frac{4C^2}{N(k-1)^2}\left[[k-|j'_*-j_*|]^2+[k-|j'_*-j_*|]\right. \\
&\quad \left.-2I\left(|j'_*-j_*|\leq\frac{k-1}{2}\right)+O_p(N^{-1})\right] I(|j'_*-j_*|\leq k-1) \qquad (3.3.6)
\end{aligned}$$

If we replace the summation in (3.3.6) over the original sample index j, j' by the summation over the ranks j_*, j'_* and denoting

$M(|j'_*-j_*|) = \left[[k-|j'_*-j_*|]^2+[k-|j'_*-j_*|]-2I\left(|j'_*-j_*|\leq\frac{k-1}{2}\right)+O_p(N^{-1})\right] I(|j'_*-j_*|\leq k-1)$, then we have

$$\delta_N \quad \leq \quad \sum_{j_* < j'_*}^{N} \frac{4C^2}{N(k-1)^2} M(|j'_* - j_*|). \tag{3.3.7}$$

As shown in (3.1.11), the right hand side of the inequality (3.3.7) converges to

$$\frac{2k(2k-1)}{3(k-1)} C^2 + \frac{2(k-2)}{(k-1)} C^2, \tag{3.3.8}$$

which is finite for finite C and fixed $k > 1$ (note that in our augmentation, k is a finite odd integer with minimum value of 3). Note that δ_N is the summation of nonnegative terms (with probability 1) due to the fact that $M(|j'_* - j_*|) \geq 0$. Hence the limit of δ_N exists as a result of the Comparison Test in calculus.

The convergence of $\lambda_N = E(\delta_N)$ is due to the Dominated Convergence Theorem after noticing that the expectation of (3.3.8) is finite. Applying the Dominated Convergence Theorem to λ_N, we get $\lim_{N\to\infty} \lambda_N = \lim_{N\to\infty} E(\delta_N) = E(\lim_{N\to\infty} \delta_N)$. This completes the proof.

The following lemma will be needed in the proof of Lemma 3.1.4.

Lemma 3.3.1. *For locally Lipschitz continuous function $A(x)$ on a bounded support $[a, b]$, we have*

$$A(X_i)I(i \in C_c) - A(X_j)I(j \in C_c) = O_p(N^{-1}),$$

uniformly in $i, j = 1, 2, ..., N$, for a given $c = 1, 2, ..., N$.

Sketch Proof of Lemma 3.3.1.

Recall that $f(x)$ and $F(x)$ are marginal probability density function and cumulative distribution function of X_j, respectively. Let $Y_1, Y_2, ..., Y_N$ be independent Exponential random variables with mean 1, and $U_1, U_2, ..., U_N$ be independent Uniform random variables on $(0,1)$. Without loss of generality, assume that $X_1, X_2, ..., X_N$ are ordered. Define $D_i = X_i - X_{i-1}$, for $2 \leq i \leq N$. Then from the properties of spacings on page 406 of Pyke (1965), there exists an $a_i \in [a,b]$ such that $F(a_i) \in (U_{(i-1)}, U_{(i)})$ and $D_i = (N-i+1)^{-1}\, Y_i \;\{1 - F(a_i)\}\{f(a_i)\}^{-1}$ for $2 \leq i \leq N$. For $j > i$,

$$\begin{aligned}
X_j - X_i &= D_{i+1} + D_{i+2} + ... + D_j \\
&= \sum_{l=i+1}^{j} \frac{1}{N-l+1}\, Y_l\, \frac{1-F(a_l)}{f(a_l)} \\
&\leq \sum_{l=i+1}^{j} \frac{1}{N-l+1}\, Y_l\, \frac{1-U_{(l-1)}}{f(a_l)} \\
&\leq \frac{1}{\inf_{l\in[i+1,j]} f(a_l)} \sum_{l=i+1}^{j} \frac{1}{N-l+1}\, Y_l\, (1-U_{(l-1)}) \\
&= K^* \sum_{l=i+1}^{j} \frac{1}{N-l+1}\, Y_l\, (1-U_{(l-1)}),
\end{aligned}$$

where K^* is some positive constant.

Note that the random variables Y_l and $U_{(l)}$ are independent, $1 \leq l \leq N$, and $U_{(l-1)}$ has $Beta(l-1,\ N-l+2)$ distribution. Therefore,

$$\begin{aligned}
E\left(\frac{1}{N-l+1} Y_l\, (1-U_{(l-1)})\right) &= \frac{1}{N-l+1}\, E(Y_l)\, E(1-U_{(l-1)}) \\
&= \frac{N-l+2}{(N-l+1)(N+1)} \\
&= O(N^{-1}), \qquad (3.3.9)
\end{aligned}$$

and

$$
\begin{aligned}
& Var\left(\frac{1}{N-l+1}\,Y_l\,(1-U_{(l-1)})\right) \\
= & \frac{1}{(N-l+1)^2}\,\left\{E(Y_l)^2\,E(1-U_{(l-1)})^2-\left(E(Y_l)\,E(1-U_{(l-1)})\right)^2\right\} \\
= & \frac{1}{(N-l+1)^2}\,\left\{2\left[\frac{(l-1)(N-l+2)}{(N+1)^2(N+2)}+\frac{(N-l+2)^2}{(N+1)^2}\right]-\frac{(N-l+2)^2}{(N+1)^2}\right\} \\
= & \frac{1}{(N-l+1)^2}\,\left\{\frac{2(l-1)(N-l+2)}{(N+1)^2(N+2)}+\frac{(N-l+2)^2}{(N+1)^2}\right\} \\
= & \frac{1}{(N+1)^2(N+2)}\,\left\{\frac{2(l-1)(N-l+2)}{(N-l+1)^2}+\frac{(N+2)(N-l+2)^2}{(N-l+1)^2}\right\} \\
\leq & \frac{1}{(N+1)^2(N+2)}\,\left\{\frac{2(N+2)(N-l+2)^2}{(N-l+1)^2}\right\} \qquad (3.3.10) \\
= & O(N^{-2}), \qquad (3.3.11)
\end{aligned}
$$

where the inequality in (3.3.10) is due to the fact that $2(l-1) < (N+2)(N-l+2)$. Due to (3.3.9) and (3.3.11) and by Theorem 14.4-1 in Bishop et al. (2007), we have

$$\frac{1}{N-l+1}\,Y_l\,(1-U_{(l-1)}) = O_p(N^{-1}), \quad \text{for all } l = 2, ..., N.$$

Consequently, for $X_i,\ X_j$ in the same cell,

$$X_j - X_i \leq K^* \sum_{l=i+1}^{j} \frac{1}{N-l+1}\,Y_l\,(1-U_{(l-1)}) = O_p\left(\frac{j-i}{N}\right) = O_p(N^{-1}), \qquad (3.3.12)$$

where the last equality in (3.3.12) is due to $j-i \leq 2k$ since $X_i,\ X_j$ are included in the same cell.

From the local Lipschitz continuity of $A(x)$ on $[a,\ b]$, when $N \to \infty$, the following condition is satisfied for any two $X_i,\ X_j$ inside the same cell

$$|A(X_j) - A(X_i)| \leq L^*|X_j - X_i|, \quad \text{for } i,j \in C_c, \qquad (3.3.13)$$

where L^* is a positive constant.

From (3.3.12) and (3.3.13), we have

$$|A(X_j) - A(X_i)| = O_p(N^{-1}),\ \text{for } i,j \in C_c.$$

This completes the proof.

Sketch Proof of part (1) of Lemma 3.1.4.

From (3.1.16), we have

$$\Delta_{N,2} = \sqrt{N}k(N-1)^{-1}\sum_{c=1}^{N}\left[2N^{-1/4}\left(\overline{A}_{c\cdot} - \overline{A}_{\cdot\cdot}\right)\left(\overline{\varepsilon}_{c\cdot} - \overline{\varepsilon}_{\cdot\cdot}\right)\right]$$

By Lemma 3.3.1 and Assumption (B),

$$\overline{A}_{c\cdot} = k^{-1}\sum_{t=1}^{k} A_{ct} = k^{-1}\sum_{i=1}^{N} A(X_i)I(i \in C_c) = A(X_c) + O_p(N^{-1}) \tag{3.3.14}$$

and

$$\overline{A}_{\cdot\cdot} = N^{-1}\sum_{c=1}^{N}\overline{A}_{c\cdot} = \overline{A(X)} + O_p(N^{-1}), \tag{3.3.15}$$

where $\overline{A(X)} = N^{-1}\sum_{c=1}^{N} A(X_c)$. Therefore, $\Delta_{N,2}$ can be written as

$$\Delta_{N,2} = \sqrt{N}k(N-1)^{-1}\sum_{c=1}^{N}\left[2N^{-1/4}\left(A(X_c) - \overline{A(X)}\right)\left(\overline{\varepsilon}_{c\cdot} - \overline{\varepsilon}_{\cdot\cdot}\right)\right] + o_p(1)$$

Denote $U_c = A(X_c) - E(A(X_c))$ and $\overline{U}_{\cdot} = N^{-1}\sum_{c=1}^{N} U_c$, then we can write

$$\begin{aligned}
\Delta_{N,2} &= 2kN^{\frac{-1}{4}}\left[\frac{\sqrt{N}}{(N-1)}\sum_{c=1}^{N}\left(A(X_c) - \overline{A(X)}\right)\left(\overline{\varepsilon}_{c\cdot} - \overline{\varepsilon}_{\cdot\cdot}\right)\right] + o_p(1) \\
&= 2kN^{\frac{-1}{4}}\left[\frac{\sqrt{N}}{(N-1)}\sum_{c=1}^{N}\left([A(X_c) - E(A(X_c))] - [\overline{A(X)} - E(A(X_c))]\right)\right. \\
&\quad \left.\times\left(\overline{\varepsilon}_{c\cdot} - \overline{\varepsilon}_{\cdot\cdot}\right)\right] + o_p(1) \\
&= 2kN^{\frac{-1}{4}}\left[\frac{\sqrt{N}}{(N-1)}\sum_{c=1}^{N}\left(U_c - \overline{U}_{\cdot}\right)\left(\overline{\varepsilon}_{c\cdot} - \overline{\varepsilon}_{\cdot\cdot}\right)\right] + o_p(1) \\
&= 2kN^{\frac{-1}{4}}\frac{\sqrt{N}}{(N-1)}\left[\sum_{c=1}^{N} U_c\overline{\varepsilon}_{c\cdot} - N\ \overline{U}_{\cdot}\ \overline{\varepsilon}_{\cdot\cdot}\right] + o_p(1) \\
&= 2kN^{\frac{-1}{4}}\left[\frac{\sqrt{N}}{(N-1)}\sum_{c=1}^{N} U_c\overline{\varepsilon}_{c\cdot}\right] - \frac{2kN^{\frac{1}{4}}}{(N-1)}\left[\sqrt{N}\ \overline{U}_{\cdot}\right]\left[\sqrt{N}\ \overline{\varepsilon}_{\cdot\cdot}\right] + o_p(1).
\end{aligned} \tag{3.3.16}$$

First we will show that

$$\left[\frac{\sqrt{N}}{(N-1)}\sum_{c=1}^{N} U_c\overline{\varepsilon}_{c\cdot}\right] = O_p(1) \tag{3.3.17}$$

and therefore the first term in (4.2.76) is $o_p(1)$. Note that $E(\bar{\varepsilon}_{c\cdot}|\mathbf{X}) = E(\overline{Q}_{c\cdot} - E(\overline{Q}_{c\cdot}|\mathbf{X})|\mathbf{X}) = 0$ and U_c is a function of X_c. Therefore, we have

$$E\left[\frac{\sqrt{N}}{(N-1)}\sum_{c=1}^{N} U_c\bar{\varepsilon}_{c\cdot}\right] = \frac{\sqrt{N}}{(N-1)}\sum_{c=1}^{N} E\left[U_c E(\bar{\varepsilon}_{c\cdot}|\mathbf{X})\right] = 0, \tag{3.3.18}$$

and

$$\begin{aligned}
& Var\left[\frac{\sqrt{N}}{(N-1)}\sum_{c=1}^{N} U_c\bar{\varepsilon}_{c\cdot}\right] \\
= \; & \frac{N}{(N-1)^2} E\left[\sum_{c=1}^{N} U_c\bar{\varepsilon}_{c\cdot}\right]^2 \\
= \; & \frac{N}{(N-1)^2} E\left[\sum_{c=1}^{N} U_c^2\bar{\varepsilon}_{c\cdot}^2 + \sum_{c\neq c'}^{N} U_c\bar{\varepsilon}_{c\cdot}U_{c'}\bar{\varepsilon}_{c'\cdot}\right] \\
= \; & \frac{N}{(N-1)^2}\left[\sum_{c=1}^{N} E\left(U_c^2\bar{\varepsilon}_{c\cdot}^2\right)\right] + \frac{N}{(N-1)^2}\left[\sum_{c\neq c'}^{N} E\left(U_cU_{c'}\bar{\varepsilon}_{c\cdot}\bar{\varepsilon}_{c'\cdot}\right)\right].
\end{aligned} \tag{3.3.19}$$

Denote the first term and second term in (4.2.79) as $\upsilon_{N,1}$ and $\upsilon_{N,2}$, respectively. Then

$$\begin{aligned}
\upsilon_{N,1} \; = \; & \frac{N}{(N-1)^2}\left[\sum_{c=1}^{N} E\left(U_c^2 E(\bar{\varepsilon}_{c\cdot}^2|\mathbf{X})\right)\right] \\
= \; & \frac{N}{(N-1)^2}\left[\sum_{c=1}^{N} E\left(U_c^2 E((\overline{Q}_{c\cdot} - E(\overline{Q}_{c\cdot}|\mathbf{X}))^2|\mathbf{X})\right)\right] \\
= \; & \frac{N}{(N-1)^2}\sum_{c=1}^{N} E\left\{U_c^2 E\left\{\left(\frac{1}{k}\sum_{i=1}^{N}(Y_i - E(Y_i|\mathbf{X}))I(i\in C_c)\right)^2\middle|\mathbf{X}\right\}\right\} \\
= \; & \frac{N}{k^2(N-1)^2}\sum_{c=1}^{N} E\left\{U_c^2 E\left\{\left(\sum_{i=1}^{N}(Y_i - E(Y_i|\mathbf{X}))^2 I(i\in C_c)\right.\right.\right. \\
& \left.\left.\left. + \sum_{i\neq i'}^{N}(Y_i - E(Y_i|\mathbf{X}))I(i\in C_c)(Y_{i'} - E(Y_{i'}|\mathbf{X}))I(i'\in C_c)\right)\middle|\mathbf{X}\right\}\right\} \\
= \; & \frac{N}{k^2(N-1)^2}\sum_{c=1}^{N}\sum_{i=1}^{N} E\left\{U_c^2 E((Y_i - E(Y_i|\mathbf{X}))^2\middle|\mathbf{X})I(i\in C_c)\right\} \qquad (3.3.20) \\
= \; & \frac{N}{k^2(N-1)^2}\sum_{i=1}^{N}\sum_{c=1}^{N} E\left\{U_c^2\sigma^2(X_i)I(i\in C_c)\right\}, \qquad (3.3.21)
\end{aligned}$$

where the equality in (4.2.80) is due to the fact that Y_i and $Y_{i'}$ are independent when $i \neq i'$. Similarly,

$$
\begin{aligned}
\upsilon_{N,2} &= \frac{N}{(N-1)^2}\left[\sum_{c\neq c'}^{N} E\left(U_c U_{c'} E(\bar{\varepsilon}_{c\cdot}\bar{\varepsilon}_{c'\cdot}|\mathbf{X})\right)\right] \\
&= \frac{N}{(N-1)^2}\sum_{c\neq c'}^{N} E\left\{U_c U_{c'} E\left\{\left(\frac{1}{k}\sum_{i=1}^{N}(Y_i - E(Y_i|\mathbf{X}))I(i\in C_c)\right)\right.\right. \\
&\quad \left.\left.\times\left(\frac{1}{k}\sum_{i'=1}^{N}(Y_{i'} - E(Y_{i'}|\mathbf{X}))I(i'\in C_{c'})\right)\middle|\,\mathbf{X}\right\}\right\} \\
&= \frac{N}{k^2(N-1)^2}\sum_{i=1}^{N}\sum_{c\neq c'}^{N} E\left\{U_c U_{c'} E((Y_i - E(Y_i|\mathbf{X}))^2\big|\,\mathbf{X})I(i\in C_c)I(i\in C_{c'})\right\} \\
&= \frac{N}{k^2(N-1)^2}\sum_{i=1}^{N}\sum_{c\neq c'}^{N} E\left\{U_c U_{c'}\sigma^2(X_i)I(i\in C_c\cap C_{c'})\right\}, \qquad (3.3.22)
\end{aligned}
$$

Consider individual terms under the summation in (4.2.81) and (4.2.82). By Cauchy-Schwarz inequality and Assumptions (A) and (B),

$$
\begin{aligned}
& E\left\{U_c^2\sigma^2(X_i)I(i\in C_c)\right\} \\
\leq\ & E\left\{U_c^2\sigma^2(X_i)\right\} \\
\leq\ & \left[E(U_c^4)\right]^{\frac{1}{2}}\left[E\left(\sigma^2(X_i)\right)^2\right]^{\frac{1}{2}} \\
=\ & \left[E(U_c^4)\right]^{\frac{1}{2}}\left[E\left(E((Y_i - E(Y_i|\mathbf{X}))^2\big|\,\mathbf{X})\right)^2\right]^{\frac{1}{2}} \\
\leq\ & \left[E(U_c^4)\right]^{\frac{1}{2}}\left[E\left(E((Y_i - E(Y_i|\mathbf{X}))^4\big|\,\mathbf{X})\right)^2\right]^{\frac{1}{2}} \\
<\ & \infty. \qquad (3.3.23)
\end{aligned}
$$

Similarly,

$$
\begin{aligned}
& \left| E\left\{ U_c U_{c'} \sigma^2(X_i) I(i \in C_c \cap C_{c'}) \right\} \right| \\
\leq \;& E\left\{ |U_c U_{c'}| \sigma^2(X_i) I(i \in C_c \cap C_{c'}) \right\} \\
\leq \;& E\left\{ |U_c U_{c'}| \sigma^2(X_i) \right\} \\
\leq \;& \left[E\left(U_c U_{c'} \right)^2 \right]^{\frac{1}{2}} \left[E\left(\sigma^2(X_i) \right)^2 \right]^{\frac{1}{2}} \\
= \;& \left[E\left(U_c^2 \right) \right]^{\frac{1}{2}} \left[E\left(U_{c'}^2 \right) \right]^{\frac{1}{2}} \left[E\left(E((Y_i - E(Y_i|\mathbf{X}))^2 \middle| \mathbf{X}) \right)^2 \right]^{\frac{1}{2}} \\
\leq \;& \left[E(U_c^4) \right]^{\frac{1}{2}} \left[E(U_{c'}^4) \right]^{\frac{1}{2}} \left[E\left(E((Y_i - E(Y_i|\mathbf{X}))^4 \middle| \mathbf{X}) \right)^2 \right]^{\frac{1}{2}} \\
< \;& \infty. \qquad (3.3.24)
\end{aligned}
$$

Note that X_i can only be used to augment at most $2k$ cells. That is, if the rank of X_i is r, then X_i can not be used to augment cells whose x values have ranks not in the set of positive integers $\{\max\{1, r-k\}, ..., \min\{r+k, N\}\}$. Therefore, the summation over c in (4.2.81) and that over c and c' in (4.2.82) each contains no more than $2k$ terms. As a result, the two terms $\upsilon_{N,1}$ and $\upsilon_{N,2}$ are $O(1)$ and therefore,

$$
Var\left[\frac{\sqrt{N}}{(N-1)} \sum_{c=1}^{N} U_c \bar{\varepsilon}_{c\cdot} \right] = O(1). \qquad (3.3.25)
$$

Due to (4.2.78) and (4.2.85), the proof of (4.2.77) is complete by applying Theorem 14.4-1 in Bishop et al. (2007).

Next, we will show that the second term in (4.2.76) is $o_p(1)$. The second term in (4.2.76) is

$$
\frac{-2kN^{\frac{1}{4}}}{(N-1)} \left[\sqrt{N} \ \bar{U}_{\cdot} \right] \ \left[\sqrt{N} \ \bar{\varepsilon}_{\cdot\cdot} \right].
$$

Using the same technique of the proof of (4.2.77), it can be shown that

$$
\left[\sqrt{N} \ \bar{\varepsilon}_{\cdot\cdot} \right] = O_p(1).
$$

In addition,

$$
\left[\sqrt{N} \ \bar{U}_{\cdot} \right] = O_p(1) \qquad (3.3.26)
$$

is a result of Central Limit Theorem (CLT) applied to $U_1, ..., U_N$ since they are i.i.d. due to the fact that $X_1, ..., X_N$ are i.i.d..

Consequently,

$$\Delta_{N,2} = O_p(N^{\frac{-1}{4}}) + O_p\left(\frac{N^{\frac{1}{4}}}{N-1}\right) + o_p(1) = o_p(1), \text{ as } N \to \infty.$$

This completes the proof.

Sketch Proof of part (2) of Lemma 3.1.4.

First we will show that

$$\Delta_{N,3} \xrightarrow{p} 0, \text{ as } N \to \infty. \tag{3.3.27}$$

From (3.1.17), we have

$$\begin{aligned}\Delta_{N,3} &= \sqrt{N}\{N(k-1)\}^{-1}\sum_{c=1}^{N}\sum_{t=1}^{k}\Big[N^{-1/2}\left(A_{ct}-\overline{A}_{c\cdot}\right)^2\Big] \\ &= \{N(k-1)\}^{-1}\sum_{c=1}^{N}\sum_{t=1}^{k}\Big[\left(A_{ct}-\overline{A}_{c\cdot}\right)^2\Big].\end{aligned}$$

By Lemma 3.3.1, we have $\left(A_{ct} - \overline{A}_{c\cdot}\right) = O_p(N^{-1})$. Thus,

$$\Delta_{N,3} = O_p(N^{-2}) \tag{3.3.28}$$

and therefore $\Delta_{N,3}$ is $o_p(1)$. This completes the proof of (3.3.27).

Next, we will show that $\Delta_{N,4} \xrightarrow{p} 0$. From (3.1.18), we have

$$\Delta_{N,4} = 2\sqrt{N}\{N(k-1)\}^{-1}\sum_{c=1}^{N}\sum_{t=1}^{k}(\varepsilon_{ct} - \overline{\varepsilon}_{c\cdot})\left(N^{-1/4}\left(A_{ct}-\overline{A}_{c\cdot}\right)\right).$$

By Hölder's inequality,

$$|\Delta_{N,4}| \le \left[\frac{2\sqrt{N}}{N(k-1)}\sum_{c=1}^{N}\sum_{t=1}^{k}(\varepsilon_{ct} - \overline{\varepsilon}_{c\cdot})^2\right]^{\frac{1}{2}}\left[\frac{2\sqrt{N}}{N(k-1)}\sum_{c=1}^{N}\sum_{t=1}^{k}\left(N^{-1/4}\left(A_{ct}-\overline{A}_{c\cdot}\right)\right)^2\right]^{\frac{1}{2}} \tag{3.3.29}$$

Now we will show that

$$\sum_{c=1}^{N}\sum_{t=1}^{k}(\varepsilon_{ct} - \overline{\varepsilon}_{c\cdot})^2 = O_p(N). \tag{3.3.30}$$

We can write

$$\sum_{c=1}^{N}\sum_{t=1}^{k}(\varepsilon_{ct}-\overline{\varepsilon}_{c\cdot})^2=\sum_{c=1}^{N}\sum_{t=1}^{k}\varepsilon_{ct}^2-k\sum_{c=1}^{N}\overline{\varepsilon}_{c\cdot}^2 \tag{3.3.31}$$

Note that

$$\begin{aligned}
E\left\{\sum_{c=1}^{N}\sum_{t=1}^{k}\varepsilon_{ct}^2\right\} &= E\left\{E\left(\sum_{c=1}^{N}\sum_{t=1}^{k}\varepsilon_{ct}^2|\mathbf{X}\right)\right\}\\
&= E\left\{E\left(\sum_{c=1}^{N}\sum_{i=1}^{N}\left[(Y_i-E(Y_i|\mathbf{X}))^2 I(i\in C_c)\right]|\mathbf{X}\right)\right\}\\
&= \sum_{c=1}^{N}\sum_{i=1}^{N}E\left\{E\left([Y_i-E(Y_i|\mathbf{X})]^2|\mathbf{X}\right)I(i\in C_c)\right\}\\
&= \sum_{c=1}^{N}\sum_{i=1}^{N}E\left\{\sigma^2(X_i)I(i\in C_c)\right\}=O(N),
\end{aligned} \tag{3.3.32}$$

where the last equality in (4.2.42) is due to the fact that $\sigma^2(X_i)$ is uniformly bounded by Assumption (A) and the summation over i in (4.2.42) contains only k terms.

Consider

$$\begin{aligned}
E\left\{\sum_{c=1}^{N}\sum_{t=1}^{k}\varepsilon_{ct}^{2}\right\}^{2} &= E\left\{E\left(\left[\sum_{c=1}^{N}\sum_{t=1}^{k}\varepsilon_{ct}^{2}\right]^{2}\middle|\mathbf{X}\right)\right\} \\
&= E\left\{E\left(\left[\sum_{c=1}^{N}\sum_{i=1}^{N}[Y_i - E(Y_i|\mathbf{X})]^2\, I(i\in C_c)\right]^{2}\middle|\mathbf{X}\right)\right\} \\
&= E\left\{E\left(\left[\sum_{c=1}^{N}\sum_{i=1}^{N}[Y_i - E(Y_i|\mathbf{X})]^4\, I(i\in C_c)\right]\right.\right. \\
&\quad + \left[\sum_{c=1}^{N}\sum_{i\neq i'}^{N}[Y_i - E(Y_i|\mathbf{X})]^2\, I(i\in C_c)\,[Y_{i'} - E(Y_{i'}|\mathbf{X})]^2\, I(i'\in C_c)\right] \\
&\quad + \left[\sum_{c\neq c'}^{N}\sum_{i=1}^{N}[Y_i - E(Y_i|\mathbf{X})]^2\, I(i\in C_c)\,[Y_i - E(Y_i|\mathbf{X})]^2\, I(i\in C_{c'})\right] \\
&\quad \left.\left. + \left[\sum_{c\neq c'}^{N}\sum_{i\neq i'}^{N}[Y_i - E(Y_i|\mathbf{X})]^2\, I(i\in C_c)\,[Y_{i'} - E(Y_{i'}|\mathbf{X})]^2\, I(i'\in C_{c'})\right]\middle|\mathbf{X}\right)\right\} \\
&= \sum_{c=1}^{N}\sum_{i=1}^{N} E\left\{E\left([Y_i - E(Y_i|\mathbf{X})]^4\,|\mathbf{X}\right) I(i\in C_c)\right\} && (3.3.33) \\
&\quad + \sum_{c=1}^{N}\sum_{i\neq i'}^{N} E\left\{\sigma^2(X_i)\sigma^2(X_{i'}) I(i,i'\in C_c)\right\} && (3.3.34) \\
&\quad + \sum_{c\neq c'}^{N}\sum_{i=1}^{N} E\left\{E\left([Y_i - E(Y_i|\mathbf{X})]^4\,|\mathbf{X}\right) I(i\in C_c\cap C_{c'})\right\} && (3.3.35) \\
&\quad + \sum_{c\neq c'}^{N}\sum_{i\neq i'}^{N} E\left\{\sigma^2(X_i)\sigma^2(X_{i'}) I(i\in C_c) I(i'\in C_{c'})\right\} && (3.3.36) \\
&= O(N^2), && (3.3.37)
\end{aligned}$$

where the equality in (4.2.47) is due to the fact that $\sigma^2(X_i)$ and $E\left([Y_i - E(Y_i|\mathbf{X})]^4\,|\mathbf{X}\right)$ are uniformly bounded by Assumption (A) and the summation over c in (4.2.43) and (4.2.44) and that over c and c' in (4.2.45) and (4.2.46) each contains no more than $2k$ terms.

From (4.2.42) and (4.2.47), we have

$$Var\left\{\sum_{c=1}^{N}\sum_{t=1}^{k}\varepsilon_{ct}^{2}\right\} = O(N^2). \tag{3.3.38}$$

Due to (4.2.42) and (4.2.48) and by Theorem 14.4-1 in Bishop et al. (2007), we have

$$\sum_{c=1}^{N}\sum_{t=1}^{k}\varepsilon_{ct}^2 = O_p\left(E\left\{\sum_{c=1}^{N}\sum_{t=1}^{k}\varepsilon_{ct}^2\right\}\right) + O_p\left(\left\{Var\left\{\sum_{c=1}^{N}\sum_{t=1}^{k}\varepsilon_{ct}^2\right\}\right\}^{1/2}\right) = O_p(N). \quad (3.3.39)$$

Similarly, it can be shown that the second term in (4.2.41) is $O_p(N)$ and therefore the proof of (4.2.40) is completed.

From (3.3.28), (3.3.29) and (4.2.40),

$$\begin{aligned}
|\Delta_{N,4}| &\leq \left[\frac{2\sqrt{N}}{N(k-1)}O_p(N)\right]^{\frac{1}{2}}\left[O_p(N^{-2})\right]^{\frac{1}{2}} \\
&= O_p(N^{-3/4}) = o_p(1), \text{ as } \quad N \to \infty.
\end{aligned}$$

This completes the proof.

Sketch Proof of Theorem 3.1.5.

The proof of the existence of $\lim_{N\to\infty}\lambda_{NA}$ is similar to that for $\lim_{N\to\infty}\lambda_N$ in Theorem 3.1.2. Now we will show that

$$\sqrt{N}(B_N(\mathbf{Q}) - W_N(\mathbf{Q}))\xrightarrow{d}N(k\sigma_A^2, \lim_{N\to\infty}\lambda_{NA}).$$

From (3.1.14), we have

$$\begin{aligned}
\sqrt{N}(B_N(\mathbf{Q}) - W_N(\mathbf{Q})) &= \sqrt{N}\left(k(N-1)^{-1}\sum_{c=1}^{N}(\overline{\varepsilon}_{c\cdot} - \overline{\varepsilon}_{\cdot\cdot})^2 - \{N(k-1)\}^{-1}\sum_{c=1}^{N}\sum_{t=1}^{k}(\varepsilon_{ct} - \overline{\varepsilon}_{c\cdot})^2\right) \\
&\quad +\Delta_{N,1} + \Delta_{N,2} - \Delta_{N,3} - \Delta_{N,4} \\
&= \sqrt{N}(B_N(\varepsilon) - W_N(\varepsilon)) + \Delta_{N,1} + \Delta_{N,2} - \Delta_{N,3} - \Delta_{N,4}, \qquad (3.3.40)
\end{aligned}$$

where $\Delta_{N,1}, \Delta_{N,2}, \Delta_{N,3}$, and $\Delta_{N,4}$ are defined in (3.1.15), (3.1.16), (3.1.17), and (3.1.18), respectively. The $B_N(\varepsilon)$ and $W_N(\varepsilon)$ are the average between-cell and within-cell variations for augmented observations with $Z_i = Y_i - \left(m_0(X_i) + N^{-1/4}A(X_i)\right)$ as the response. Note that the conditional mean of Z_i given $X_i = x$ satisfies the null hypothesis. But $Var(Z_i|X_i = x)$ is equal to $Var(Y_i|X_i = x)$. The result of Theorem 3.1.3 implies that

$$\sqrt{N}(B_N(\varepsilon) - W_N(\varepsilon))\xrightarrow{d}N(0, \lim_{N\to\infty}\lambda_{NA}), \qquad (3.3.41)$$

with λ_{NA} calculated with the same formula as λ_N in Theorem 3.1.2 but with $\sigma^2(X_j)$ calculated under the alternative hypothesis. By Lemma 3.1.4, we have

$$\Delta_{N,i} \xrightarrow{p} 0, \text{ as } N \to \infty, \text{ for } \quad i = 2, 3, 4. \tag{3.3.42}$$

Thus, we only need to consider $\Delta_{N,1}$ to obtain the asymptotic mean under the alternatives.

Note that $A(X_1), A(X_2), ..., A(X_N)$ are i.i.d. since $X_1, X_2, ..., X_N$ are i.i.d.. From (3.3.14) and (3.3.15), we can write $\Delta_{N,1}$ in (3.1.15) as

$$\Delta_{N,1} = \sqrt{N}k(N-1)^{-1}\sum_{c=1}^{N}\Big[N^{-\frac{1}{2}}\left(\overline{A}_{c\cdot} - \overline{A}_{\cdot\cdot}\right)^2\Big] = k(N-1)^{-1}\sum_{c=1}^{N}\Big(A(X_c) - \overline{A(X)}\Big)^2 = k\widehat{\sigma}_A^2, \tag{3.3.43}$$

where $\widehat{\sigma}_A^2$ is the sample variance of $A(X_1), A(X_2), ..., A(X_N)$. By Weak Law of Large Numbers,

$$k\widehat{\sigma}_A^2 \xrightarrow{p} k\sigma_A^2 = k\mathrm{Var}(A(X)) = k\left[\int_{-\infty}^{\infty} A^2(x)f(x)dx - \left(\int_{-\infty}^{\infty} A(x)f(x)dx\right)^2\right] \tag{3.3.44}$$

as $N \to \infty$ and k stays fixed.

From (3.3.42), (3.3.43) and (3.3.44), we have

$$\Delta_{N,1} + \Delta_{N,2} - \Delta_{N,3} - \Delta_{N,4} \xrightarrow{p} k\sigma_A^2. \tag{3.3.45}$$

From (3.3.40), (3.3.41) and (3.3.45) and by applying Slutsky's theorem, we have

$$\sqrt{N}(B_N(\mathbf{Q}) - W_N(\mathbf{Q})) \xrightarrow{d} N(k\sigma_A^2, \lim_{N\to\infty} \lambda_{NA}).$$

This completes the proof.

Chapter 4

Nonparametric lack-of-fit test of nonlinear regression in presence of heteroscedastic variances

4.1 Introduction

Even though there are plenty of studies for lack-of-fit in linear regression models (cf. Neill and Johnson (1984, 1985, 1989), Eubank and Hart (1992), Hart (2008), Miller and Neill (2008)), we found that lack-of-fit tests in nonlinear regression has not received much attention. The existing literature include for example, Neill (1988) proposed such a test based on near replicate clusters. This test is a modified version of the classical linear regression lack of fit test of Fisher (1922) and can be used in both cases of replication and nonreplication. Neill and Miller (2003) generalized the clustering based test of Christensen (1989, 1991) to the nonlinear case. Li (2005) presented a test for assessing the lack of fit of nonlinear regression models based on local linear kernel smoothers. All the preceding tests assume normality or constant variance for the random errors. Therefore, these tests are not appropriate for heteroscedastic regression problems.

However, practical data may have variances vary with the covariate i.e., the errors are heteroscedastic. In such cases, ignoring model heteroscadisity will lead to incorrect and misleading inferences. This issue is explained in the following example from the Engineering Statistics Handbook for ultrasonic reference block study. In this study, the data consist of a response variable (ultrasonic response) and a predictor variable (metal distance). The Handbook used this data to demonstrate nonlinear process modeling and the use of transformations to deal with the violation of the assumption of constant variances for the errors. The scatter plot for this data is given in Figure 4.1.

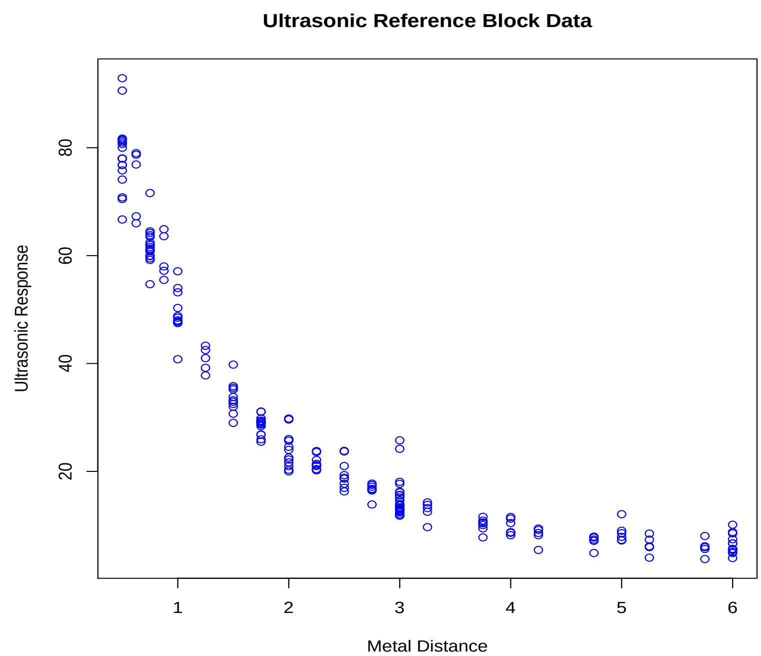


Figure 4.1: *Ultrasonic Reference Block Data*

Based on the plot and scientific and engineering knowledge, the scientists decide to fit the following theoretical model

$$y = \frac{\exp(-b_1 x)}{b_2 + b_3 x} + \epsilon, \tag{4.1.1}$$

where b_1, b_2, and b_3 are parameters to be estimated.

To check the validity of the suggested model in (4.1.1), diagnostic plots were used and these plots show that the variance of the errors is not constant (see also the residuals plot against the independent variable, Metal Distance in Figure 4.2).

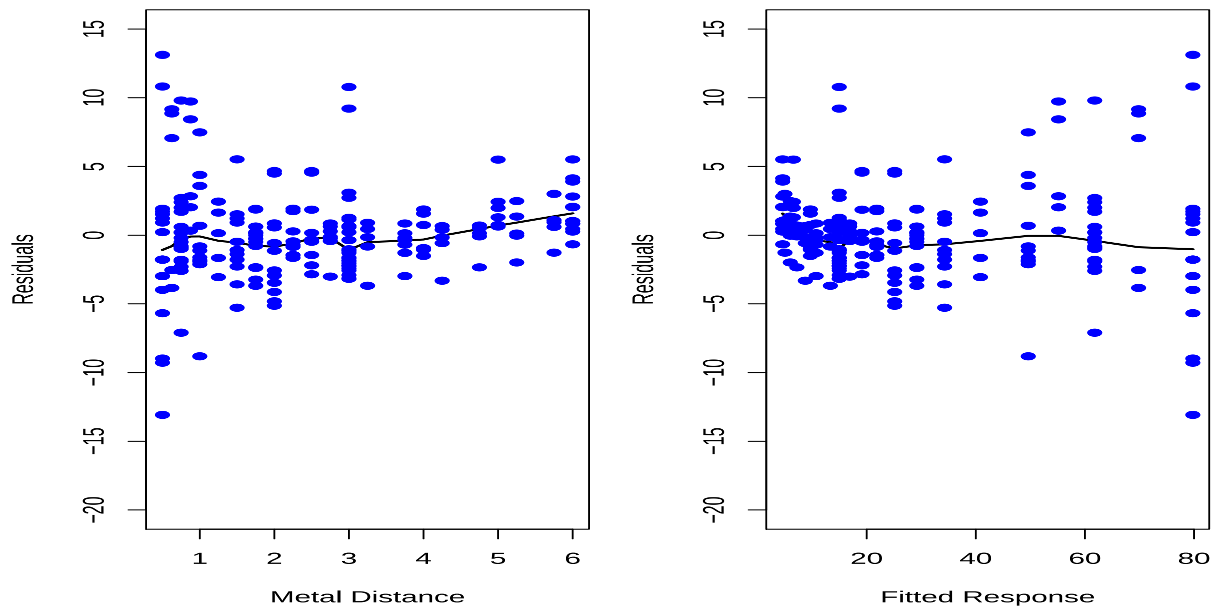


Figure 4.2: *Residuals plots from fit to untransformed data*

To deal with the violation of non constant variance, the scientists suggested to fit a model with a square root transformation for the response variable i.e.

$$y^{1/2} = \frac{\exp(-b_1 x)}{b_2 + b_3 x} + \epsilon. \tag{4.1.2}$$

The diagnostic plots show that the model (4.1.2) appear to satisfy the model assumptions better than model (4.1.1). Careful examination of the residuals plot of model (4.1.2) still (see Figure 4.3) shows some nonrandom pattern. This means that there might be lack of fit or a constant variance violation. Consequently, its important to develop a test for assessing the lack of fit for nonlinear regression models and accounting for heteroscedasticity.

In this chapter, we propose a nonparametric lack of fit test in nonlinear regression models in the presence of heteroscedastic variances. The proposed test is an extension of the lack of fit test of constant regression considered in Chapter 3 to the nonlinear regression models. Our test is valid for both continuous and discrete response variable. We constructed the test

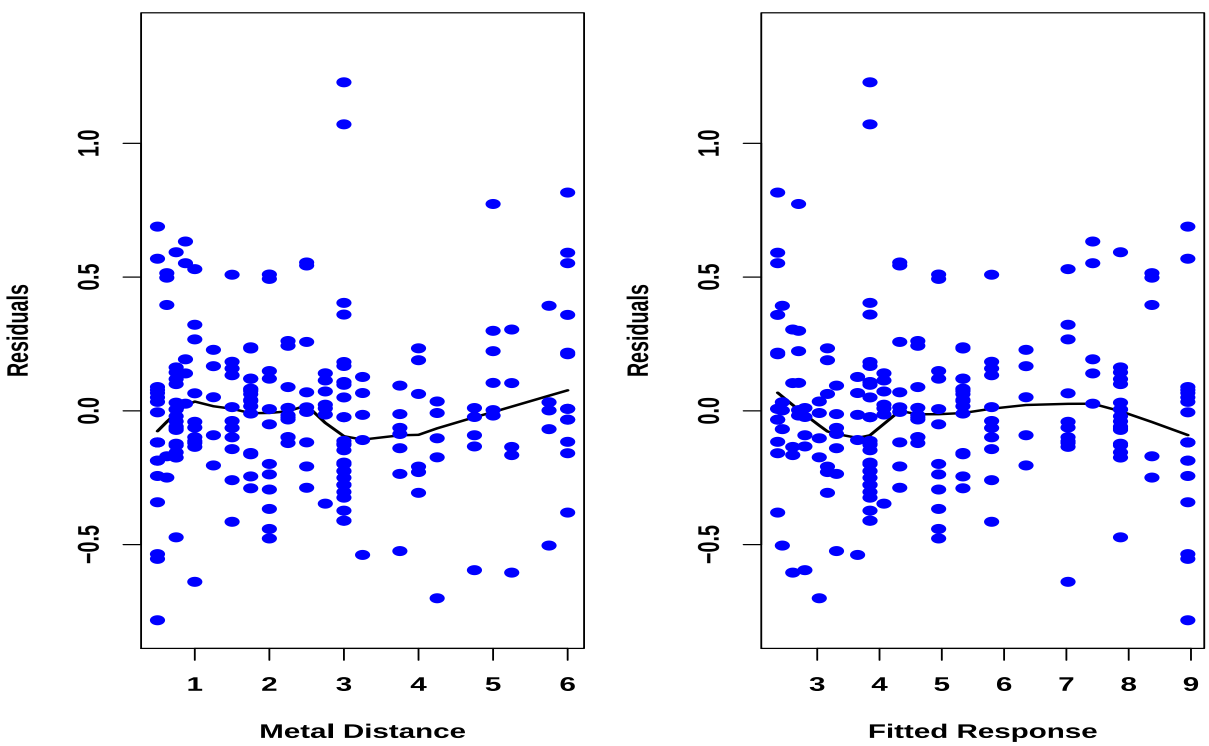


Figure 4.3: *Residuals plots from fit to transformed data*

statistic using k-nearest neighbor augmentation defined through the ranks of the predictor. This augmentation is done on the residuals from the fitted model under the null hypothesis of nonlinear regression. This idea of using k-nearest neighbor augmentation to develop test statistics was used earlier by Wang and Akritas (2006), Wang et al. (2008), and Wang et al. (2010) for different purposes.

In addition to the aforementioned lack of fit tests, Kuchibhatla and Hart (1996) proposed a new version of the order selection test of Eubank and Hart (1992). The test of Kuchibhatla and Hart (1996) was used for testing lack of fit in nonlinear regression models. In particular, consider the nonlinear regression model

$$Y_j = G(x_j; \boldsymbol{\theta}) + \epsilon_j, \quad j = 1, \ldots, N,$$

where $x_j = (j - 0.5)/N$, $Y_1, ..., Y_N$ are the observed responses, $\boldsymbol{\theta}$ is a vector of unknown parameters, the errors $\epsilon_1, ..., \epsilon_N$ are independent and identically distributed with $E(\epsilon_j) = 0$, and $Var(\epsilon_j) = \sigma^2$. Let $\hat{\boldsymbol{\theta}}$ denote the estimate of $\boldsymbol{\theta}$, then residuals can be defined as $e_j = Y_j - G(x_j; \hat{\boldsymbol{\theta}}), j = 1, ..., N$. To test the null hypothesis $H_0 : E(Y|X = x) = G(x; \boldsymbol{\theta})$, Kuchibhatla and Hart (1996) constructed a test statistic based on the residuals of the following form:

$$S_N = \max_{0<m<N} \frac{1}{m} \sum_{j=1}^{m} \frac{2N\hat{\phi}_j^2}{\hat{\sigma}^2}, \tag{4.1.3}$$

where $\hat{\sigma}^2$ is a consistent estimator of σ^2 and $\hat{\phi}_j = 1/N \sum_{i=1}^{N} e_i \cos(\pi j x_i), \quad j = 1, ..., N-1$. The test statistic S_N in (4.1.3) was also used in Hart (1997) for the same purpose. To find critical values of the test statistic S_N, Kuchibhatla and Hart (1996) and Hart (1997) suggested using large sample approximation or bootstrap algorithm. They showed that the power of the test statistic S_N converges to 1 under fixed alternatives when N goes to infinity. However, they did not give theory on the limiting distribution of the test statistic S_N in the case of testing the null hypothesis of nonlinear regression models. Kuchibhatla and Hart (1996) suggested that wild bootstrap of Hardle and Mammen (1993) might be used to handle the presence of heteroscedastic errors, which was considered in Chen et al. (2001) for testing constant regression in heteroscedastic case. In Kuchibhatla and Hart (1996) and Hart (1997), no numerical studies were reported for testing nonlinear regression null hypothesis. They only reported numerical studies for testing constant regression or linear regression null hypothesis. One drawback of the bootstrap method is the need of extensive computations which is time consuming.

For heteroscedastic nonlinear regression models, lack of fit tests have been considered by few authors. For example, Li (1999, 2003) proposed such tests based on a cosine-series smoother and a comparison of nonparametric kernel and parametric fits. However, these tests are assuming that the variance is a known function of unknown parameters which is not the case of our proposed method.

In addition to the preceding references, the literature on lack of fit test includes the following papers: Hausman (1978), Ruud (1984), Newey(1985a; 1985b), Tauchen (1985), White (1982), White (1987), and Bierens (1990). Most of these tests are not consistent for general alternatives and some of them need extensive computation. Based on smoothing techniques, consistent nonparametric lack-of-fit tests were studied by some authors (cf Lee (1988); Yatchew (1992); Eubank and Spiegelman (1990); Hardle and Mammen (1993); Zheng (1996); Horowitz and Spokoiny (2001); Guerre and Lavergne (2005); Song and Du (2011)). However, some of these tests have the drawbacks of being computationally complicated and having conditions that are hard to justify. In contrast of our proposed test, all of the proceeding methods require the response variable to be continuous.

4.2 Theoretical results

4.2.1 The hypotheses and test statistic

Consider the model

$$Y_j = G(X_j; \boldsymbol{\theta}) + \epsilon_j,$$

where G is a known function, $\boldsymbol{\theta}$ is a vector of unknown parameters $(\theta_1, ..., \theta_p)^T$ with $p < \infty$, and (X_j, Y_j), $j = 1, \ldots, N$, is a random sample of the random variables (X, Y). Let $f(x)$ and $F(x)$ denote the marginal probability density function and cumulative distribution function of X_j, respectively. Denote $\varepsilon_i^* = Y_i - E(Y_i|X_i)$.

We consider testing the hypothesis:

$$H_0\colon\ E(Y|X = x) = G(x; \boldsymbol{\theta}) \tag{4.2.1}$$

against:

$$H_1\colon\ E(Y|X = x) \neq G(x; \boldsymbol{\theta}), \tag{4.2.2}$$

in the presence of heteroscedastic variances (i.e. $\mathrm{Var}(Y_i|X_i = x) = \sigma^2(x)$). Similar to Chapter 3, fixed number of k-nearest neighbor augmentation will be used to construct a test statistic for conducting lack-of-fit test. This augmentation is done for each unique value x_i of the predictor by generating a cell that contains k values of the response Y whose corresponding x values are among the k closest to x_i in rank. We consider k to be an odd number for convenience. Let the indicator function that the difference between the ranks of X_1 and X_2 is no more than $(k-1)/2$ be defined by $g_{Nk}(X_1, X_2) = I\left(N|\widehat{F}(X_1) - \widehat{F}(X_2)| \leq \frac{k-1}{2}\right)$, where $\widehat{F}(x) = N^{-1}\sum_{j=1}^{N} I(X_j \leq x)$ denote the empirical distribution of X.

Denote

$$v(X_c; \boldsymbol{\theta}) = G(X_c; \boldsymbol{\theta}) - \overline{G}(\boldsymbol{\theta}), \text{ where } \overline{G}(\boldsymbol{\theta}) = N^{-1}\sum_{c=1}^{N} G(X_c; \boldsymbol{\theta}). \tag{4.2.3}$$

We assume the following conditions:

Assumption (C):

- (C1) For all x, suppose that $F(x)$ is differentiable and the fourth conditional central moments of Y_j given X_j are uniformly bounded.

- (C2) Assume that X_i has bounded support $\chi = [a, b]$ and the function $G(x; \boldsymbol{\theta}) : \chi \times \mathbb{R}^p \to \mathbb{R}$ is locally Lipschitz continuous with respect to its first argument x. That is, the function G is continuous and for each $(x_0; \boldsymbol{\theta}_0) \in \chi \times \mathbb{R}^p$ there are neighborhoods $U(x_0) \subseteq \chi$, $V(\boldsymbol{\theta}_0) \subseteq \mathbb{R}^p$ and a scaler $L > 0$ such that $|G(y; \mathbf{s}) - G(z; \mathbf{s})| \leq L|z - y|$ for all $z, y \in U(x_0)$ and $\mathbf{s} \in V(\boldsymbol{\theta}_0)$.

- (C3) $\frac{\partial v(x;\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}$ and $\frac{\partial^2 v(x;\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^2}$ exist.
- (C4) $E\left[\frac{\partial v(X;\boldsymbol{\theta})}{\partial \theta_m}\frac{\partial v(X;\boldsymbol{\theta})}{\partial \theta_l}\right]^2 < \infty$ and $E\left[\frac{\partial v(X;\boldsymbol{\theta})}{\partial \theta_m}\frac{\partial^2 v(X;\boldsymbol{\theta})}{\partial \theta_l \theta_u}\right]^2 < \infty$ for $m, l, u = 1, ..., p$.
- (C5) There exist $\tau_N \to \infty$, such that $\tau_N(\hat{\theta_m} - \theta_m) = O_p(1)$ for all $m = 1, ..., p$, where $\hat{\boldsymbol{\theta}} = (\hat{\theta}_1, ..., \hat{\theta}_p)^T$ is an estimate of $\boldsymbol{\theta}$.

Condition (C5) specifies that $\hat{\boldsymbol{\theta}}$ is a consistent estimator of $\boldsymbol{\theta}$ at rate τ_N. Such $\hat{\boldsymbol{\theta}}$ with different rates from nonlinear regression has been considered by various authors. For example, for homoscedastic nonlinear regression models, Jennrich (1969) derived consistency and asymptotic normality of the least squares estimator under standard sufficient conditions. In particular, he showed that $\sqrt{N}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})$ is asymptotically normally distributed. Under certain conditions imposed on the nonlinear mean regression function, the asymptotic normality of $\sqrt{\tau_N}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})$ is derived by Wu (1981), where $\tau_N \to \infty$ as $N \to \infty$. For heteroscedastic nonlinear regression models, an M-estimation and preliminary test estimation based procedures are considered by Lim (2009) and Lim et al. (2010). Under some regularity conditions, they derived the asymptotic distribution of the M-estimators and showed that $\sqrt{N}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})$ converges to normality. In all above examples, condition (C5) is satisfied.

Let B_N^* and W_N^* be defined as the following:

$$
\begin{aligned}
B_N^* &= \frac{k}{N-1}\sum_{j_1=1}^{N}\left[\frac{1}{k}\sum_{j=1}^{N} Y_j\, g_{Nk}(X_{j_1}, X_j) - \frac{1}{Nk}\sum_{j_2=1}^{N}\sum_{j=1}^{N} Y_j\, g_{Nk}(X_{j_2}, X_j) - v(x_{j_1}; \hat{\boldsymbol{\theta}})\right]^2 \\
W_N^* &= \frac{1}{N(k-1)}\sum_{j_1=1}^{N}\sum_{j=1}^{N}\left[Y_j\, g_{Nk}(X_{j_1}, X_j) - \frac{1}{k}\sum_{j_2=1}^{N} Y_{j_2}\, g_{Nk}(X_{j_1}, X_{j_2})\right]^2 .
\end{aligned}
$$

Let $e_{ct} = R_{ct} - G_{ct}(\boldsymbol{\theta})$ and $e_{ct}^* = R_{ct} - G_{ct}(\hat{\boldsymbol{\theta}})$ where $R_{ct}, t = 1, \ldots, k$, are the augmented response values in cell (c) under the null hypothesis in (4.2.1) and $G_{ct}(\boldsymbol{\theta})$ is the $G(x; \boldsymbol{\theta})$

function evaluated at the covariate value for augmented observation R_{ct}. Note that e_{ct} satisfies the null hypothesis of constant regression that we considered in Chapter 3 and can be viewed as the augmented data for $Z_i = Y_i - G(X_i;\boldsymbol{\theta})$, whose conditional mean satisfies the null hypothesis in (3.1.1). Then B_N^* and W_N^* can be expressed as the average between-cell and within-cell variations, respectively. They can be written as the following:

$$B_N^* = \frac{k}{N-1}\sum_{c=1}^{N} (\overline{e^*}_{c\cdot} - \overline{e^*}_{\cdot\cdot})^2 \quad and \quad W_N^* = \frac{1}{N(k-1)}\sum_{c=1}^{N}\sum_{t=1}^{k}(e^*{}_{ct} - \overline{e^*}_{c\cdot})^2,$$

where $\overline{e^*}_{c\cdot} = k^{-1}\sum_{t=1}^{k} e^*{}_{ct}$ and $\overline{e^*}_{\cdot\cdot} = N^{-1}\sum_{c=1}^{N}\overline{e^*}_{c\cdot}$.

We consider the test statistic $B_N^* - W_N^*$ for testing the hypothesis in (4.2.1).

4.2.2 Asymptotic distribution of the test statistic under the null hypothesis

Note that $e^*_{ct} = R_{ct} - G_{ct}(\hat{\boldsymbol{\theta}}) = R_{ct} - G_{ct}(\boldsymbol{\theta}) + G_{ct}(\boldsymbol{\theta}) - G_{ct}(\hat{\boldsymbol{\theta}}) = e_{ct} + G_{ct}(\boldsymbol{\theta}) - G_{ct}(\hat{\boldsymbol{\theta}})$. Let $\overline{G}_{c\cdot}(\boldsymbol{\theta}) = k^{-1}\sum_{t=1}^{k} G_{ct}(\boldsymbol{\theta})$ and $\overline{G}_{\cdot\cdot}(\boldsymbol{\theta}) = N^{-1}\sum_{c=1}^{N}\overline{G}_{c\cdot}(\boldsymbol{\theta})$. Then, B_N^* and W_N^* can be written as

$$\begin{aligned}
B_N^* &= \frac{k}{N-1}\sum_{c=1}^{N} (\overline{e^*}_{c\cdot} - \overline{e^*}_{\cdot\cdot})^2 \\
&= \frac{k}{N-1}\sum_{c=1}^{N} \Big(\overline{e}_{c\cdot} + \overline{G}_{c\cdot}(\boldsymbol{\theta}) - \overline{G}_{c\cdot}(\hat{\boldsymbol{\theta}}) - \overline{e}_{\cdot\cdot} - \overline{G}_{\cdot\cdot}(\boldsymbol{\theta}) + \overline{G}_{\cdot\cdot}(\hat{\boldsymbol{\theta}})\Big)^2 \\
&= \frac{k}{N-1}\sum_{c=1}^{N} \Big[(\overline{e}_{c\cdot} - \overline{e}_{\cdot\cdot})^2 + \Big(\big[\,\overline{G}_{c\cdot}(\boldsymbol{\theta}) - \overline{G}_{\cdot\cdot}(\boldsymbol{\theta})\big] - \Big[\overline{G}_{c\cdot}(\hat{\boldsymbol{\theta}}) - \overline{G}_{\cdot\cdot}(\hat{\boldsymbol{\theta}})\Big]\Big)^2 \\
&\qquad + 2\,(\overline{e}_{c\cdot} - \overline{e}_{\cdot\cdot})\Big(\big[\,\overline{G}_{c\cdot}(\boldsymbol{\theta}) - \overline{G}_{\cdot\cdot}(\boldsymbol{\theta})\big] - \Big[\overline{G}_{c\cdot}(\hat{\boldsymbol{\theta}}) - \overline{G}_{\cdot\cdot}(\hat{\boldsymbol{\theta}})\Big]\Big)\Big] \qquad (4.2.4)
\end{aligned}$$

Similarly,

$$\begin{aligned}
W_N^* &= \frac{1}{N(k-1)}\sum_{c=1}^{N}\sum_{t=1}^{k}\left({e^*}_{ct}-\overline{e^*}_{c\cdot}\right)^2 \\
&= \frac{1}{N(k-1)}\sum_{c=1}^{N}\sum_{t=1}^{k}\left[e_{ct}+G_{ct}(\boldsymbol{\theta})-G_{ct}(\hat{\boldsymbol{\theta}})-\overline{e}_{c\cdot}-\overline{G}_{c\cdot}(\boldsymbol{\theta})+\overline{G}_{c\cdot}(\hat{\boldsymbol{\theta}})\right]^2 \\
&= \frac{1}{N(k-1)}\sum_{c=1}^{N}\sum_{t=1}^{k}\Bigg[(e_{ct}-\overline{e}_{c\cdot})^2+\left(\left[\,G_{ct}(\boldsymbol{\theta})-\overline{G}_{c\cdot}(\boldsymbol{\theta})\right]-\left[G_{ct}(\hat{\boldsymbol{\theta}})-\overline{G}_{c\cdot}(\hat{\boldsymbol{\theta}})\right]\right)^2 \\
&\qquad +2\,(e_{ct}-\overline{e}_{c\cdot})\left(\left[\,G_{ct}(\boldsymbol{\theta})-\overline{G}_{c\cdot}(\boldsymbol{\theta})\right]-\left[G_{ct}(\hat{\boldsymbol{\theta}})-\overline{G}_{c\cdot}(\hat{\boldsymbol{\theta}})\right]\right)\Bigg] \qquad (4.2.5)
\end{aligned}$$

Let

$$B_N' = \frac{k}{N-1}\sum_{c=1}^{N}\left(\overline{e}_{c\cdot}-\overline{e}_{\cdot\cdot}\right)^2, \text{ and } W_N' = \frac{1}{N(k-1)}\sum_{c=1}^{N}\sum_{t=1}^{k}\left(e_{ct}-\overline{e}_{c\cdot}\right)^2, \qquad (4.2.6)$$

then the test statistic can be written as

$$\sqrt{N}(B_N^*-W_N^*) = \sqrt{N}(B_N'-W_N') + \Delta_{N,1}+\Delta_{N,2}-\Delta_{N,3}-\Delta_{N,4}, \qquad (4.2.7)$$

where

$$\begin{aligned}
\Delta_{N,1} &= \frac{k\sqrt{N}}{N-1}\sum_{c=1}^{N}\left(\left[\,\overline{G}_{c\cdot}(\boldsymbol{\theta})-\overline{G}_{\cdot\cdot}(\boldsymbol{\theta})\right]-\left[\overline{G}_{c\cdot}(\hat{\boldsymbol{\theta}})-\overline{G}_{\cdot\cdot}(\hat{\boldsymbol{\theta}})\right]\right)^2 && (4.2.8) \\
\Delta_{N,2} &= \frac{2k\sqrt{N}}{N-1}\sum_{c=1}^{N}(\overline{e}_{c\cdot}-\overline{e}_{\cdot\cdot})\left(\left[\,\overline{G}_{c\cdot}(\boldsymbol{\theta})-\overline{G}_{\cdot\cdot}(\boldsymbol{\theta})\right]-\left[\overline{G}_{c\cdot}(\hat{\boldsymbol{\theta}})-\overline{G}_{\cdot\cdot}(\hat{\boldsymbol{\theta}})\right]\right) && (4.2.9) \\
\Delta_{N,3} &= \frac{\sqrt{N}}{N(k-1)}\sum_{c=1}^{N}\sum_{t=1}^{k}\left(\left[\,G_{ct}(\boldsymbol{\theta})-\overline{G}_{c\cdot}(\boldsymbol{\theta})\right]-\left[G_{ct}(\hat{\boldsymbol{\theta}})-\overline{G}_{c\cdot}(\hat{\boldsymbol{\theta}})\right]\right)^2 && (4.2.10) \\
\Delta_{N,4} &= \frac{2\sqrt{N}}{N(k-1)}\sum_{c=1}^{N}\sum_{t=1}^{k}(e_{ct}-\overline{e}_{c\cdot})\left(\left[\,G_{ct}(\boldsymbol{\theta})-\overline{G}_{c\cdot}(\boldsymbol{\theta})\right]-\left[G_{ct}(\hat{\boldsymbol{\theta}})-\overline{G}_{c\cdot}(\hat{\boldsymbol{\theta}})\right]\right). && (4.2.11)
\end{aligned}$$

We state the following results before giving the asymptotic distribution of the test statistic.

Lemma 4.2.1. *If the Assumption (C2) is satisfied, then*

$$G(X_i;\boldsymbol{\theta})I(i\in C_c)-G(X_j;\boldsymbol{\theta})I(j\in C_c)=O_p(N^{-1}),$$

uniformly in $i,j=1,2,...,N$, *for a given* $c=1,2,...,N$.

The proof of Lemma 4.2.1 is similar to the proof of Lemma 3.3.1 in Chapter 3 and is thus skipped.

Lemma 4.2.2. *If the Assumptions (C1), (C3), and (C4) are satisfied, then*

$$Var\left[\sum_{c=1}^{N}(\overline{e}_{c\cdot}-\overline{e}_{\cdot\cdot})\,(\frac{\partial v(X_c;\boldsymbol{\theta})}{\partial\boldsymbol{\theta}})\right]=(O(N))\mathbf{J}_p \text{ as } N\to\infty,$$

where $\mathbf{J}_p$ is an $p\times p$ matrix of ones.

Proof

We can write

$$\begin{aligned}
&\mathrm{Var}\left[\sum_{c=1}^{N}(\overline{e}_{c\cdot}-\overline{e}_{\cdot\cdot})\,(\frac{\partial v(X_c;\boldsymbol{\theta})}{\partial\boldsymbol{\theta}})\right]\\
=\;& E\left\{\mathrm{Var}\left(\left[\sum_{c=1}^{N}(\overline{e}_{c\cdot}-\overline{e}_{\cdot\cdot})\,(\frac{\partial v(X_c;\boldsymbol{\theta})}{\partial\boldsymbol{\theta}})\right]\middle|\mathbf{X}\right)\right\}\\
&+\mathrm{Var}\left\{E\left(\left[\sum_{c=1}^{N}(\overline{e}_{c\cdot}-\overline{e}_{\cdot\cdot})\,(\frac{\partial v(X_c;\boldsymbol{\theta})}{\partial\boldsymbol{\theta}})\right]\middle|\mathbf{X}\right)\right\}\\
=\;& E\left\{\mathrm{Var}\left(\left[\sum_{c=1}^{N}(\overline{e}_{c\cdot}-\overline{e}_{\cdot\cdot})\,(\frac{\partial v(X_c;\boldsymbol{\theta})}{\partial\boldsymbol{\theta}})\right]\middle|\mathbf{X}\right)\right\}+\mathrm{Var}\left\{\sum_{c=1}^{N}(\frac{\partial v(X_c;\boldsymbol{\theta})}{\partial\boldsymbol{\theta}})E\left(\overline{e}_{c\cdot}-\overline{e}_{\cdot\cdot}|\mathbf{X}\right)\right\}.
\end{aligned}$$

Note that, for e_{ct} there exists some $j\in C_c$ such that $e_{ct}=(Y_j-E(Y_j|\mathbf{X}))$. Thus

$$E(e_{ct}|\mathbf{X})=E\left((Y_j-E(Y_j|\mathbf{X}))I(j\in C_c)|\mathbf{X}\right)=E\left((Y_j-E(Y_j|\mathbf{X}))|\mathbf{X}\right)I(j\in C_c)=0,$$

and

$$E\left(\overline{e}_{c\cdot}-\overline{e}_{\cdot\cdot}|\mathbf{X}\right)=0. \tag{4.2.12}$$

Therefore, by using (4.2.12) we have

$$
\begin{aligned}
&\operatorname{Var}\left[\sum_{c=1}^{N}(\overline{e}_{c\cdot}-\overline{e}_{\cdot\cdot})\,(\frac{\partial v(X_c;\boldsymbol{\theta})}{\partial \boldsymbol{\theta}})\right] \\
=\ & E\left\{\operatorname{Var}\left(\left[\sum_{c=1}^{N}(\overline{e}_{c\cdot}-\overline{e}_{\cdot\cdot})\,(\frac{\partial v(X_c;\boldsymbol{\theta})}{\partial \boldsymbol{\theta}})\right]\middle|\,\mathbf{X}\right)\right\} \\
=\ & E\left\{E\left(\left[\sum_{c=1}^{N}\sum_{c'=1}^{N}(\overline{e}_{c\cdot}-\overline{e}_{\cdot\cdot})\,(\overline{e}_{c'\cdot}-\overline{e}_{\cdot\cdot})\left(\frac{\partial v(X_c;\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right)\left(\frac{\partial v(X_{c'};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right)^T\right]\middle|\,\mathbf{X}\right)\right\} \\
=\ & \sum_{c=1}^{N}E\left\{E\left(\left[(\overline{e}_{c\cdot}-\overline{e}_{\cdot\cdot})^2\left(\frac{\partial v(X_c;\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right)\left(\frac{\partial v(X_c;\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right)^T\right]\middle|\,\mathbf{X}\right)\right\} \\
& +\sum_{c\neq c'}^{N}E\left\{E\left(\left[(\overline{e}_{c\cdot}-\overline{e}_{\cdot\cdot})\,(\overline{e}_{c'\cdot}-\overline{e}_{\cdot\cdot})\left(\frac{\partial v(X_c;\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right)\left(\frac{\partial v(X_{c'};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right)^T\right]\middle|\,\mathbf{X}\right)\right\} \\
=\ & \sum_{c=1}^{N}E\left\{\left(\frac{\partial v(X_c;\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right)\left(\frac{\partial v(X_c;\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right)^T E\left((\overline{e}_{c\cdot}-\overline{e}_{\cdot\cdot})^2\middle|\,\mathbf{X}\right)\right\} \\
& +\sum_{c\neq c'}^{N}E\left\{\left(\frac{\partial v(X_c;\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right)\left(\frac{\partial v(X_{c'};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right)^T E\left((\overline{e}_{c\cdot}-\overline{e}_{\cdot\cdot})\,(\overline{e}_{c'\cdot}-\overline{e}_{\cdot\cdot})\middle|\,\mathbf{X}\right)\right\}. \qquad (4.2.13)
\end{aligned}
$$

Next we will show that the first and second terms in (4.2.13) are $O(N)$. Denote $a_{cc'} = \left(\frac{\partial v(X_c;\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right)\left(\frac{\partial v(X_{c'};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right)^T$. Then the first and second terms in (4.2.13) can be written respectively as

$$
\begin{aligned}
& E\left\{\sum_{c=1}^{N}a_{cc}E\left((\overline{e}_{c\cdot}-\overline{e}_{\cdot\cdot})^2\middle|\,\mathbf{X}\right)\right\} \\
=\ & E\left\{\sum_{c=1}^{N}a_{cc}E\left(\overline{e}_{c\cdot}^2|\mathbf{X}\right)\right\}-2E\left\{\sum_{c=1}^{N}a_{cc}E\left(\overline{e}_{c\cdot}\overline{e}_{\cdot\cdot}|\mathbf{X}\right)\right\} \\
& +E\left\{\sum_{c=1}^{N}a_{cc}E\left(\overline{e}_{\cdot\cdot}^2|\mathbf{X}\right)\right\} \qquad (4.2.14)
\end{aligned}
$$

and

$$\begin{aligned}
& E\left\{\sum_{c\neq c'}^{N} a_{cc'}E\left(\left(\overline{e}_{c\cdot}-\overline{e}_{\cdot\cdot}\right)\left(\overline{e}_{c'\cdot}-\overline{e}_{\cdot\cdot}\right)\middle|\,\mathbf{X}\right)\right\} \\
= \ & E\left\{\sum_{c\neq c'}^{N} a_{cc'}E\left(\overline{e}_{c\cdot}\overline{e}_{c'\cdot}|\mathbf{X}\right)\right\}-2E\left\{\sum_{c\neq c'}^{N} a_{cc'}E\left(\overline{e}_{c\cdot}\overline{e}_{\cdot\cdot}|\mathbf{X}\right)\right\} \\
& +E\left\{\sum_{c\neq c'}^{N} a_{cc'}E\left(\overline{e}_{\cdot\cdot}\overline{e}_{\cdot\cdot}|\mathbf{X}\right)\right\}. \qquad (4.2.15)
\end{aligned}$$

Consider the first term in (4.2.14) and (4.2.15) and denote them by $S_{N,1}$ and $S_{N,2}$ respectively, then

$$\begin{aligned}
S_{N,1} &= E\left\{\sum_{c=1}^{N} a_{cc}E\left(\overline{e}_{c\cdot}^{2}|\mathbf{X}\right)\right\} \\
&= E\left\{\sum_{c=1}^{N} a_{cc}E\left\{\left(\frac{1}{k}\sum_{i=1}^{N}(Y_i-E(Y_i|\mathbf{X}))I(i\in C_c)\right)^2\middle|\,\mathbf{X}\right\}\right\} \\
&= \frac{1}{k^2}E\left\{\sum_{c=1}^{N} a_{cc}E\left\{\left(\sum_{i=1}^{N}(Y_i-E(Y_i|\mathbf{X}))^2 I(i\in C_c)\right.\right.\right. \\
&+ \left.\left.\left.\sum_{i\neq i'}^{N}(Y_i-E(Y_i|\mathbf{X}))I(i\in C_c)(Y_{i'}-E(Y_{i'}|\mathbf{X}))I(i'\in C_c)\right)\middle|\,\mathbf{X}\right\}\right\} \\
&= \frac{1}{k^2}\sum_{c=1}^{N}\sum_{i=1}^{N} E\left\{E((Y_i-E(Y_i|\mathbf{X}))^2\middle|\,\mathbf{X})a_{cc}I(i\in C_c)\right\} \qquad (4.2.16) \\
&= \frac{1}{k^2}\sum_{i=1}^{N}\sum_{c=1}^{N} E\left\{\sigma^2(X_i)a_{cc}I(i\in C_c)\right\}, \qquad (4.2.17)
\end{aligned}$$

and

$$\begin{aligned}
S_{N,2} &= E\left\{\sum_{c\neq c'}^{N} a_{cc'} E\left(\bar{e}_{c\cdot}\bar{e}_{c'\cdot}|\mathbf{X}\right)\right\} \\
&= E\left\{\sum_{c\neq c'}^{N} a_{cc'} E\left\{\left(\frac{1}{k}\sum_{i=1}^{N}(Y_i - E(Y_i|\mathbf{X}))I(i\in C_c)\right)\right.\right. \\
&\quad \times \left.\left.\left(\frac{1}{k}\sum_{i'=1}^{N}(Y_{i'} - E(Y_{i'}|\mathbf{X}))I(i'\in C_{c'})\right)\right|\mathbf{X}\right\}\right\} \\
&= \frac{1}{k^2}\sum_{i=1}^{N}\sum_{c\neq c'}^{N} E\left\{E(\left(Y_i - E(Y_i|\mathbf{X})\right)^2\big|\,\mathbf{X})a_{cc'}I(i\in C_c)I(i\in C_{c'})\right\} \qquad (4.2.18)\\
&= \frac{1}{k^2}\sum_{i=1}^{N}\sum_{c\neq c'}^{N} E\left\{\sigma^2(X_i)a_{cc'}I(i\in C_c\cap C_{c'})\right\}, \qquad (4.2.19)
\end{aligned}$$

where the equality in (4.2.16) and (4.2.18) is due to the fact that Y_i and $Y_{i'}$ are independent when $i \neq i'$. Note that X_i can only be used to augment at most $2k$ cells. That is, if the rank of X_i is r, then X_i can only be used to augment cells whose x values have ranks in $(r-k, r+k)$. Therefore, the summation over c in (4.2.17) and that over c and c' in (4.2.19) each contains no more than $2k$ terms. In addition, by Cauchy-Schwarz inequality,

$$\begin{aligned}
&\left|E\left\{\sigma^2(X_i)a_{cc}I(i\in C_c)\right\}\right| \\
\leq\ & E\left\{\sigma^2(X_i)|a_{cc}|I(i\in C_c)\right\} \\
\leq\ & E\left\{\sigma^2(X_i)|a_{cc}|\right\} \\
\leq\ & \left[E\left(\sigma^2(X_i)\right)^2\right]^{\frac{1}{2}}\left[E\left(\left(\frac{\partial v(X_c;\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right)\left(\frac{\partial v(X_c;\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right)^T\right)^2\right]^{\frac{1}{2}} \\
=\ & \left[E\left(\sigma^2(X_i)\right)^2\right]^{\frac{1}{2}}\left[E\left(\left(\frac{\partial v(X_c;\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right)\left[\left(\frac{\partial v(X_c;\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right)^T\left(\frac{\partial v(X_c;\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right)\right]\left(\frac{\partial v(X_c;\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right)^T\right)\right]^{\frac{1}{2}} \\
=\ & \left[E\left(\sigma^2(X_i)\right)^2\right]^{\frac{1}{2}}\left[E\left(\left[trace(a_{cc})\right]a_{cc}\right)\right]^{\frac{1}{2}}. \qquad (4.2.20)
\end{aligned}$$

Note that the elements of $E\left([trace(a_{cc})]\, a_{cc}\right)$ are

$$\begin{aligned}
& E\left(\left[\sum_{m=1}^{p}\left(\frac{\partial v(X_c;\boldsymbol{\theta})}{\partial\theta_m}\right)^2\right]\left(\frac{\partial v(X_c;\boldsymbol{\theta})}{\partial\theta_l}\right)\left(\frac{\partial v(X_c;\boldsymbol{\theta})}{\partial\theta_u}\right)\right), \text{for any integers } l,u\in[1,p] \\
= & \sum_{m=1}^{p} E\left(\left(\frac{\partial v(X_c;\boldsymbol{\theta})}{\partial\theta_m}\right)^2\left(\frac{\partial v(X_c;\boldsymbol{\theta})}{\partial\theta_l}\right)\left(\frac{\partial v(X_c;\boldsymbol{\theta})}{\partial\theta_u}\right)\right) \\
& \leq \sum_{m=1}^{p}\left[E\left(\frac{\partial v(X_c;\boldsymbol{\theta})}{\partial\theta_m}\right)^4\right]^{\frac{1}{2}}\left[E\left(\frac{\partial v(X_c;\boldsymbol{\theta})}{\partial\theta_l}\frac{\partial v(X_c;\boldsymbol{\theta})}{\partial\theta_u}\right)^2\right]^{\frac{1}{2}} \qquad (4.2.21)
\end{aligned}$$

by assumption $(C3)$, the terms in (4.2.21) are all bounded. Further, by assumption $(C1)$,

$$\sigma^2(X_i) \;=\; E\big((Y_i - E(Y_i|\mathbf{X}))^2\big|\,\mathbf{X}\big) \leq E\big((Y_i - E(Y_i|\mathbf{X}))^4\big|\,\mathbf{X}\big) < \infty.$$

Therefore, the terms in (4.2.20) are all bounded. Similarly,

$$\begin{aligned}
& \left|E\left\{\sigma^2(X_i)a_{cc'}I(i\in C_c\cap C_{c'})\right\}\right| \\
\leq\; & E\left\{\sigma^2(X_i)|a_{cc'}|I(i\in C_c\cap C_{c'})\right\} \\
\leq\; & E\left\{\sigma^2(X_i)|a_{cc'}|\right\} \\
\leq\; & \left[E\left(\sigma^2(X_i)\right)^2\right]^{\frac{1}{2}}\left[E\left(\left(\frac{\partial v(X_c;\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right)\left(\frac{\partial v(X_{c'};\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right)^T\right)^2\right]^{\frac{1}{2}} < \infty,
\end{aligned}$$

Hence, the elements of $S_{N,1}$ and $S_{N,2}$ are $O(N)$. Similarly, it can be shown that the second and third terms in (4.2.14) are $o(S_{N,1})$ and the second and third terms in (4.2.15) are $o(S_{N,2})$. This completes the proof.

Lemma 4.2.3. *Under Assumption (C) and as $N\to\infty$,*

$$\begin{aligned}
(1) \quad & \Delta_{N,1}\xrightarrow{p} 0\ , \\
(2) \quad & \Delta_{N,2}\xrightarrow{p} 0\ , \\
(3) \quad & \Delta_{N,3}\xrightarrow{p} 0\ , \\
(4) \quad & \Delta_{N,4}\xrightarrow{p} 0\ ,
\end{aligned}$$

where $\Delta_{N,i};\ i=1,2,3,4$ are defined in (4.2.8), (4.2.9), (4.2.10), and (4.2.11), respectively.

Proof of part (1) of Lemma 4.2.3

By Assumption (C2) and Lemma 4.2.1, we have

$$\overline{G}_{c\cdot}(\boldsymbol{\theta}) = \frac{1}{k}\sum_{t=1}^{k} G_{ct}(\boldsymbol{\theta}) = \frac{1}{k}\sum_{i=1}^{N} G(X_i;\boldsymbol{\theta})I(i\in C_c) = G(X_c;\boldsymbol{\theta}) + O_p(N^{-1}),$$

and

$$\overline{G}_{\cdot\cdot}(\boldsymbol{\theta}) = \frac{1}{N}\sum_{c=1}^{N} \overline{G}_{c\cdot}(\boldsymbol{\theta}) = \frac{1}{N}\sum_{c=1}^{N} G(X_c;\boldsymbol{\theta}) + O_p(N^{-1}) = \overline{G}(\boldsymbol{\theta}) + O_p(N^{-1}).$$

Therefore,

$$\overline{G}_{c\cdot}(\boldsymbol{\theta}) - \overline{G}_{\cdot\cdot}(\boldsymbol{\theta}) = G(X_c;\boldsymbol{\theta}) - \overline{G}(\boldsymbol{\theta}) + O_p(N^{-1}) = v(X_c;\boldsymbol{\theta}) + O_p(N^{-1}), \tag{4.2.22}$$

where $v(X_c;\boldsymbol{\theta})$ is defined in (4.2.3). Similarly,

$$\overline{G}_{c\cdot}(\hat{\boldsymbol{\theta}}) - \overline{G}_{\cdot\cdot}(\hat{\boldsymbol{\theta}}) = v(X_c;\hat{\boldsymbol{\theta}}) + O_p(N^{-1}). \tag{4.2.23}$$

Consequently, $\Delta_{N,1}$ in (4.2.8) can be written as

$$\begin{aligned}
\Delta_{N,1} &= \frac{k\sqrt{N}}{N-1}\sum_{c=1}^{N}\Big(\big[\,\overline{G}_{c\cdot}(\boldsymbol{\theta}) - \overline{G}_{\cdot\cdot}(\boldsymbol{\theta})\big] - \Big[\overline{G}_{c\cdot}(\hat{\boldsymbol{\theta}}) - \overline{G}_{\cdot\cdot}(\hat{\boldsymbol{\theta}})\Big]\Big)^2 \\
&= \frac{k\sqrt{N}}{N-1}\sum_{c=1}^{N}\Big\{v(X_c;\boldsymbol{\theta}) - v(X_c;\hat{\boldsymbol{\theta}}) + O_p(N^{-1})\Big\}^2 \\
&= \frac{k\sqrt{N}}{N-1}\sum_{c=1}^{N}\Big\{\Big(v(X_c;\boldsymbol{\theta}) - v(X_c;\hat{\boldsymbol{\theta}})\Big)^2 + 2\Big(v(X_c;\boldsymbol{\theta}) - v(X_c;\hat{\boldsymbol{\theta}})\Big)O_p(N^{-1}) + O_p(N^{-2})\Big\}
\end{aligned} \tag{4.2.24}$$

Using Taylor's expansion, we can write

$$v(X_c;\hat{\boldsymbol{\theta}}) = v(X_c;\boldsymbol{\theta}) + (\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})^T\frac{\partial v(X_c;\boldsymbol{\theta})}{\partial\boldsymbol{\theta}} + O_p\left((\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})^T\frac{\partial^2 v(X_c;\boldsymbol{\theta})}{\partial\boldsymbol{\theta}^2}(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})\right),$$

where

$$\frac{\partial v(X_c;\boldsymbol{\theta})}{\partial\boldsymbol{\theta}} = \begin{pmatrix} \frac{\partial v(X_c;\boldsymbol{\theta})}{\partial\theta_1} \\ \frac{\partial v(X_c;\boldsymbol{\theta})}{\partial\theta_2} \\ \vdots \\ \frac{\partial v(X_c;\boldsymbol{\theta})}{\partial\theta_p} \end{pmatrix} \text{ and } \frac{\partial^2 v(X_c;\boldsymbol{\theta})}{\partial\boldsymbol{\theta}^2} = \begin{pmatrix} \frac{\partial^2 v(X_c;\boldsymbol{\theta})}{\partial{\theta_1}^2} & \frac{\partial^2 v(X_c;\boldsymbol{\theta})}{\partial\theta_1\theta_2} & \cdots & \frac{\partial^2 v(X_c;\boldsymbol{\theta})}{\partial\theta_1\theta_p} \\ \frac{\partial^2 v(X_c;\boldsymbol{\theta})}{\partial\theta_2\theta_1} & \frac{\partial^2 v(X_c;\boldsymbol{\theta})}{\partial{\theta_2}^2} & \cdots & \frac{\partial^2 v(X_c;\boldsymbol{\theta})}{\partial\theta_2\theta_p} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 v(X_c;\boldsymbol{\theta})}{\partial\theta_p\theta_1} & \frac{\partial^2 v(X_c;\boldsymbol{\theta})}{\partial\theta_p\theta_2} & \cdots & \frac{\partial^2 v(X_c;\boldsymbol{\theta})}{\partial{\theta_p}^2} \end{pmatrix}.$$

Thus, we can write

$$\begin{aligned}
\Delta_{N,1} &= \frac{k\sqrt{N}}{N-1}\sum_{c=1}^{N}\left\{\left[(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})^T\frac{\partial v(X_c;\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}+O_p\left((\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})^T\frac{\partial^2 v(X_c;\boldsymbol{\theta})}{\partial\boldsymbol{\theta}^2}(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})\right)\right]^2\right.\\
&\left.+2\left[(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})^T\frac{\partial v(X_c;\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}+O_p\left((\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})^T\frac{\partial^2 v(X_c;\boldsymbol{\theta})}{\partial\boldsymbol{\theta}^2}(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})\right)\right]O_p(N^{-1})+O_p(N^{-2})\right\}\\
&= \frac{k\sqrt{N}}{N-1}\sum_{c=1}^{N}\left\{\left[\left((\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})^T\frac{\partial v(X_c;\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right)^2+O_p\left(\left((\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})^T\frac{\partial^2 v(X_c;\boldsymbol{\theta})}{\partial\boldsymbol{\theta}^2}(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})\right)^2\right)\right.\right.\\
&\left.+O_p\left((\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})^T\frac{\partial v(X_c;\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})^T\frac{\partial^2 v(X_c;\boldsymbol{\theta})}{\partial\boldsymbol{\theta}^2}(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})\right)\right]\\
&\left.+O_p\left(N^{-1}(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})^T\frac{\partial v(X_c;\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right)+O_p\left(N^{-1}(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})^T\frac{\partial^2 v(X_c;\boldsymbol{\theta})}{\partial\boldsymbol{\theta}^2}(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})\right)+O_p(N^{-2})\right\}\\
&= \psi_{N,\boldsymbol{\theta},1}+\psi_{N,\boldsymbol{\theta},2}+\psi_{N,\boldsymbol{\theta},3}+O_p(N^{-3/2}), \qquad (4.2.25)
\end{aligned}$$

where

$$\begin{aligned}
\psi_{N,\boldsymbol{\theta},1} &= (\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})^T k\left[\frac{\sqrt{N}}{N-1}\sum_{c=1}^{N}\frac{\partial v(X_c;\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\left(\frac{\partial v(X_c;\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right)^T\right](\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}),\\
\psi_{N,\boldsymbol{\theta},2} &= O_p\left(\frac{k\sqrt{N}}{N-1}\sum_{c=1}^{N}(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})^T\frac{\partial v(X_c;\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})^T\frac{\partial^2 v(X_c;\boldsymbol{\theta})}{\partial\boldsymbol{\theta}^2}(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})\right)\\
&\text{and}\\
\psi_{N,\boldsymbol{\theta},3} &= O_p\left(\frac{k}{N}(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})^T\left[\frac{\sqrt{N}}{N-1}\sum_{c=1}^{N}\frac{\partial v(X_c;\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right]\right).
\end{aligned}$$

Denote $(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})^T=(\Delta_1,...,\Delta_p)^T$, then

$$\begin{aligned}
\psi_{N,\boldsymbol{\theta},2} &= O_p\left(\frac{k\sqrt{N}}{N-1}\sum_{m=1}^{p}\sum_{l=1}^{p}\sum_{u=1}^{p}\Delta_m\Delta_l\Delta_u\sum_{c=1}^{N}\frac{\partial v(X_c;\boldsymbol{\theta})}{\partial\theta_m}\frac{\partial^2 v(X_c;\boldsymbol{\theta})}{\partial\theta_l\theta_u}\right)\\
&= O_p\left(k\sum_{m=1}^{p}\sum_{l=1}^{p}\sum_{u=1}^{p}\Delta_m\Delta_l\Delta_u\left[\frac{\sqrt{N}}{N-1}\sum_{c=1}^{N}\frac{\partial v(X_c;\boldsymbol{\theta})}{\partial\theta_m}\frac{\partial^2 v(X_c;\boldsymbol{\theta})}{\partial\theta_l\theta_u}\right]\right). \qquad (4.2.26)
\end{aligned}$$

Since $X_1,...,X_N$ are i.i.d., we have $(\frac{\partial v(X_1;\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}),...,(\frac{\partial v(X_N;\boldsymbol{\theta})}{\partial\boldsymbol{\theta}})$ are i.i.d., $(\frac{\partial v(X_1;\boldsymbol{\theta})}{\partial\boldsymbol{\theta}})(\frac{\partial v(X_1;\boldsymbol{\theta})}{\partial\boldsymbol{\theta}})^T,...,(\frac{\partial v(X_N;\boldsymbol{\theta})}{\partial\boldsymbol{\theta}})(\frac{\partial v(X_N;\boldsymbol{\theta})}{\partial\boldsymbol{\theta}})^T$ are iid, and for any integers $m,l,u\in[1,p]$, $\{\frac{\partial v(X_c;\boldsymbol{\theta})}{\partial\theta_m}\frac{\partial^2 v(X_c;\boldsymbol{\theta})}{\partial\theta_l\theta_u};c=1,...,N\}$ are i.i.d. as well. Therefore, under the assumptions (C3) and (C4), Central Limit Theorem (CLT) can be used to show that

$$\left[\frac{\sqrt{N}}{N-1}\sum_{c=1}^{N}\frac{\partial v(X_c;\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right] = (O_p(1))\mathbf{1}, \tag{4.2.27}$$

$$\left[\frac{\sqrt{N}}{N-1}\sum_{c=1}^{N}\frac{\partial v(X_c;\boldsymbol{\theta})}{\partial\theta_m}\frac{\partial^2 v(X_c;\boldsymbol{\theta})}{\partial\theta_l\theta_u}\right] = O_p(1), \tag{4.2.28}$$

and

$$\left[\frac{\sqrt{N}}{N-1}\sum_{c=1}^{N}\frac{\partial v(X_c;\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\left(\frac{\partial v(X_c;\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right)^T\right] = (O_p(1))\mathbf{J}_p. \tag{4.2.29}$$

Since $(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) = (O_p(\tau_N^{-1}))\mathbf{1}$ from assumption (C5) and from (4.2.27), (4.2.28) and (4.2.29), we have

$$\psi_{N,\boldsymbol{\theta},1} = O_p(\tau_N^{-2}),\ \psi_{N,\boldsymbol{\theta},2} = O_p(\tau_N^{-3}),\ \text{and}\ \psi_{N,\boldsymbol{\theta},3} = O_p(N^{-1}\tau_N^{-1}) \tag{4.2.30}$$

Putting (4.2.30) into (4.2.25), we have

$$\Delta_{N,1} = O_p(\tau_N^{-2}) + O_p(\tau_N^{-3}) + O_p(N^{-1}\tau_N^{-1}) + O_p(N^{-3/2}) = o_p(1),$$

as k stays bounded and $N\to\infty$. This completes the proof.

Proof of part (2) of Lemma 4.2.3

From (4.2.9), we have

$$\Delta_{N,2} \;=\; \frac{2k\sqrt{N}}{N-1}\sum_{c=1}^{N}(\bar{e}_{c\cdot} - \bar{e}_{\cdot\cdot})\left(\left[\,\overline{G}_{c\cdot}(\boldsymbol{\theta}) - \overline{G}_{\cdot\cdot}(\boldsymbol{\theta})\right] - \left[\overline{G}_{c\cdot}(\hat{\boldsymbol{\theta}}) - \overline{G}_{\cdot\cdot}(\hat{\boldsymbol{\theta}})\right]\right)$$

Using (4.2.22) and (4.2.22), $\Delta_{N,2}$ can be written as

$$\Delta_{N,2} \;=\; \frac{2k\sqrt{N}}{N-1}\sum_{c=1}^{N}(\bar{e}_{c\cdot} - \bar{e}_{\cdot\cdot})\left(v(X_c;\boldsymbol{\theta}) - v(X_c;\hat{\boldsymbol{\theta}}) + O_p(N^{-1})\right)$$

Using Taylor's expansion, we can write

$$v(X_c;\hat{\boldsymbol{\theta}}) = v(X_c;\boldsymbol{\theta}) + (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})^T\frac{\partial v(X_c;\boldsymbol{\theta})}{\partial\boldsymbol{\theta}} + O_p\left((\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})^T\frac{\partial^2 v(X_c;\boldsymbol{\theta})}{\partial\boldsymbol{\theta}^2}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})\right).$$

Therefore,

$$
\begin{aligned}
\Delta_{N,2} &= \frac{2k\sqrt{N}}{N-1}\sum_{c=1}^{N}(\overline{e}_{c\cdot}-\overline{e}_{\cdot\cdot})\left(-(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})^T\frac{\partial v(X_c;\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} - O_p\left((\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})^T\frac{\partial^2 v(X_c;\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^2}(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})\right)\right.\\
&\quad \left.+O_p(N^{-1})\right)\\
&= -2k(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})^T\left[\frac{\sqrt{N}}{N-1}\sum_{c=1}^{N}(\overline{e}_{c\cdot}-\overline{e}_{\cdot\cdot})\frac{\partial v(X_c;\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right]\\
&\quad -2kO_p\left((\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})^T\left[\frac{\sqrt{N}}{N-1}\sum_{c=1}^{N}(\overline{e}_{c\cdot}-\overline{e}_{\cdot\cdot})\left(\frac{\partial^2 v(X_c;\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^2}\right)\right](\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})\right)\\
&\quad +O_p\left(2kN^{-1}\left[\frac{\sqrt{N}}{N-1}\sum_{c=1}^{N}(\overline{e}_{c\cdot}-\overline{e}_{\cdot\cdot})\right]\right) \qquad (4.2.31)
\end{aligned}
$$

Next, we will show that

$$
\left[\frac{\sqrt{N}}{N-1}\sum_{c=1}^{N}(\overline{e}_{c\cdot}-\overline{e}_{\cdot\cdot})\frac{\partial v(X_c;\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right] = O_p(\mathbf{1}) \qquad (4.2.32)
$$

Since $E\left(\overline{e}_{c\cdot}-\overline{e}_{\cdot\cdot}|\mathbf{X}\right)=0$ from (4.2.12), then we have

$$
\begin{aligned}
& E\left[\frac{\sqrt{N}}{N-1}\sum_{c=1}^{N}(\overline{e}_{c\cdot}-\overline{e}_{\cdot\cdot})\frac{\partial v(X_c;\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right]\\
&= \frac{\sqrt{N}}{N-1}\sum_{c=1}^{N}E\left\{E\left[(\overline{e}_{c\cdot}-\overline{e}_{\cdot\cdot})\frac{\partial v(X_c;\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\middle|\mathbf{X}\right]\right\}\\
&= \frac{\sqrt{N}}{N-1}\sum_{c=1}^{N}E\left\{\frac{\partial v(X_c;\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}E\left(\overline{e}_{c\cdot}-\overline{e}_{\cdot\cdot}|\mathbf{X}\right)\right\} = \mathbf{0} \qquad (4.2.33)
\end{aligned}
$$

From Lemma 4.2.2, we have

$$
\mathrm{Var}\left[\frac{\sqrt{N}}{N-1}\sum_{c=1}^{N}(\overline{e}_{c\cdot}-\overline{e}_{\cdot\cdot})\frac{\partial v(X_c;\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right] = (O(1))\mathbf{J}_p \qquad (4.2.34)
$$

With (4.2.33) and (4.2.34) hold, the proof of (4.2.32) is complete if we apply Theorem 14.4-1 in Bishop et al. (2007). Similarly, it can be shown that

$$
\left[\frac{\sqrt{N}}{N-1}\sum_{c=1}^{N}(\overline{e}_{c\cdot}-\overline{e}_{\cdot\cdot})\frac{\partial^2 v(X_c;\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^2}\right] = (O_p(1))\mathbf{J}_p,, \qquad (4.2.35)
$$

and

$$\left[\frac{\sqrt{N}}{N-1}\sum_{c=1}^{N}(\overline{e}_{c\cdot}-\overline{e}_{\cdot\cdot})\right]=O_p(1). \tag{4.2.36}$$

From assumption (C5), we have

$$(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})=(O_p(\tau_N^{-1}))\mathbf{1} \tag{4.2.37}$$

Putting (4.2.32), (4.2.35), (4.2.36) and (4.2.37) into (4.2.31), we have

$$\Delta_{N,2}=O_p(\tau_N^{-1})+O_p(\tau_N^{-2})+O_p(N^{-1})=o_p(1),$$

as k stays bounded and $N\to\infty$. This completes the proof.

Proof of part (3) of Lemma 4.2.3

From (4.2.10), we have

$$\Delta_{N,3} = \frac{\sqrt{N}}{N(k-1)}\sum_{c=1}^{N}\sum_{t=1}^{k}\Big(\big[\,G_{ct}(\boldsymbol{\theta})-\overline{G}_{c\cdot}(\boldsymbol{\theta})\big]-\Big[G_{ct}(\hat{\boldsymbol{\theta}})-\overline{G}_{c\cdot}(\hat{\boldsymbol{\theta}})\Big]\Big)^2$$

By Lemma 4.2.1 and Assumption (C2),

$$G_{ct}(\boldsymbol{\theta})-\overline{G}_{c\cdot}(\boldsymbol{\theta})=O_p(N^{-1})$$

and

$$G_{ct}(\hat{\boldsymbol{\theta}})-\overline{G}_{c\cdot}(\hat{\boldsymbol{\theta}})=O_p(N^{-1}).$$

Thus,

$$\Delta_{N,3}=O_p(N^{-3/2}), \tag{4.2.38}$$

and therefore $\Delta_{N,3}$ is $o_p(1)$. This completes the proof.

Proof of part (4) of Lemma 4.2.3

From (4.2.11), we have

$$\Delta_{N,4} = \frac{2\sqrt{N}}{N(k-1)}\sum_{c=1}^{N}\sum_{t=1}^{k}(e_{ct}-\overline{e}_{c\cdot})\Big(\big[\,G_{ct}(\boldsymbol{\theta})-\overline{G}_{c\cdot}(\boldsymbol{\theta})\big]-\Big[G_{ct}(\hat{\boldsymbol{\theta}})-\overline{G}_{c\cdot}(\hat{\boldsymbol{\theta}})\Big]\Big)$$

Using Hölder's inequality and (4.2.10),

$$
\begin{aligned}
|\Delta_{N,4}| \leq & \left[\frac{2\sqrt{N}}{N(k-1)}\sum_{c=1}^{N}\sum_{t=1}^{k}(e_{ct}-\overline{e}_{c\cdot})^2\right]^{\frac{1}{2}} \\
& \times\left[\frac{2\sqrt{N}}{N(k-1)}\sum_{c=1}^{N}\sum_{t=1}^{k}\Big(\big[\,G_{ct}(\boldsymbol{\theta})-\overline{G}_{c\cdot}(\boldsymbol{\theta})\big]-\Big[G_{ct}(\hat{\boldsymbol{\theta}})-\overline{G}_{c\cdot}(\hat{\boldsymbol{\theta}})\Big]\Big)^2\right]^{\frac{1}{2}} \\
= & \left[\frac{2\sqrt{N}}{N(k-1)}\sum_{c=1}^{N}\sum_{t=1}^{k}(e_{ct}-\overline{e}_{c\cdot})^2\right]^{\frac{1}{2}}[2\ \ \Delta_{N,3}]^{\frac{1}{2}} \qquad (4.2.39)
\end{aligned}
$$

Now we will show that

$$
\frac{2\sqrt{N}}{N(k-1)}\sum_{c=1}^{N}\sum_{t=1}^{k}(e_{ct}-\overline{e}_{c\cdot})^2 = O_p(N^{\frac{1}{2}}). \qquad (4.2.40)
$$

We can write

$$
\begin{aligned}
\frac{2\sqrt{N}}{N(k-1)}\sum_{c=1}^{N}\sum_{t=1}^{k}(e_{ct}-\overline{e}_{c\cdot})^2 &= \frac{2\sqrt{N}}{N(k-1)}\sum_{c=1}^{N}\left[\sum_{t=1}^{k}e_{ct}^2-k\overline{e}_{c\cdot}^2\right] \\
&= \frac{2\sqrt{N}}{N(k-1)}\sum_{c=1}^{N}\sum_{t=1}^{k}e_{ct}^2-\frac{2k\sqrt{N}}{N(k-1)}\sum_{c=1}^{N}\overline{e}_{c\cdot}^2 \qquad (4.2.41)
\end{aligned}
$$

Note that

$$
\begin{aligned}
E\left\{\frac{2\sqrt{N}}{N(k-1)}\sum_{c=1}^{N}\sum_{t=1}^{k}e_{ct}^2\right\} &= \frac{2\sqrt{N}}{N(k-1)}E\left\{E\left(\sum_{c=1}^{N}\sum_{t=1}^{k}e_{ct}^2|\mathbf{X}\right)\right\} \\
&= \frac{2\sqrt{N}}{N(k-1)}E\left\{E\left(\sum_{c=1}^{N}\sum_{i=1}^{N}\left[(Y_i-E(Y_i|\mathbf{X}))^2I(i\in C_c)\right]|\mathbf{X}\right)\right\} \\
&= \frac{2\sqrt{N}}{N(k-1)}\sum_{c=1}^{N}\sum_{i=1}^{N}E\left\{E\left([Y_i-E(Y_i|\mathbf{X})]^2|\mathbf{X}\right)I(i\in C_c)\right\} \\
&= \frac{2\sqrt{N}}{N(k-1)}\sum_{c=1}^{N}\sum_{i=1}^{N}E\left\{\sigma^2(X_i)I(i\in C_c)\right\}=O(N^{\frac{1}{2}}), \qquad (4.2.42)
\end{aligned}
$$

where the last equality in (4.2.42) is due to the fact that $\sigma^2(X_i)$ is uniformly bounded by Assumption (C1) and the summation over i in (4.2.42) contains only k terms.

Consider

$$
\begin{aligned}
& E\left\{\frac{2\sqrt{N}}{N(k-1)}\sum_{c=1}^{N}\sum_{t=1}^{k}e_{ct}^2\right\}^2 \\
= & \frac{4N}{N^2(k-1)^2}E\left\{E\left(\left[\sum_{c=1}^{N}\sum_{t=1}^{k}e_{ct}^2\right]^2\middle|\mathbf{X}\right)\right\} \\
= & \frac{4}{N(k-1)^2}E\left\{E\left(\left[\sum_{c=1}^{N}\sum_{i=1}^{N}[Y_i - E(Y_i|\mathbf{X})]^2\, I(i\in C_c)\right]^2\middle|\mathbf{X}\right)\right\} \\
= & \frac{4}{N(k-1)^2}E\left\{E\left(\left[\sum_{c=1}^{N}\sum_{i=1}^{N}[Y_i - E(Y_i|\mathbf{X})]^4\, I(i\in C_c)\right]\right.\right. \\
& \quad + \left[\sum_{c=1}^{N}\sum_{i\neq i'}^{N}[Y_i - E(Y_i|\mathbf{X})]^2\, I(i\in C_c)\,[Y_{i'} - E(Y_{i'}|\mathbf{X})]^2\, I(i'\in C_c)\right] \\
& \quad + \left[\sum_{c\neq c'}^{N}\sum_{i=1}^{N}[Y_i - E(Y_i|\mathbf{X})]^2\, I(i\in C_c)\,[Y_i - E(Y_i|\mathbf{X})]^2\, I(i\in C_{c'})\right] \\
& \quad \left.\left. + \left[\sum_{c\neq c'}^{N}\sum_{i\neq i'}^{N}[Y_i - E(Y_i|\mathbf{X})]^2\, I(i\in C_c)\,[Y_{i'} - E(Y_{i'}|\mathbf{X})]^2\, I(i'\in C_{c'})\right]\middle|\mathbf{X}\right)\right\}
\end{aligned}
$$

$$
= \frac{4}{N(k-1)^2}\left\{\sum_{c=1}^{N}\sum_{i=1}^{N}E\left\{E\left([Y_i - E(Y_i|\mathbf{X})]^4\,|\mathbf{X}\right)I(i\in C_c)\right\}\right. \tag{4.2.43}
$$

$$
+\sum_{c=1}^{N}\sum_{i\neq i'}^{N}E\left\{\sigma^2(X_i)\sigma^2(X_{i'})I(i,i'\in C_c)\right\} \tag{4.2.44}
$$

$$
+\sum_{c\neq c'}^{N}\sum_{i=1}^{N}E\left\{E\left([Y_i - E(Y_i|\mathbf{X})]^4\,|\mathbf{X}\right)I(i\in C_c\cap C_{c'})\right\} \tag{4.2.45}
$$

$$
\left.+\sum_{c\neq c'}^{N}\sum_{i\neq i'}^{N}E\left\{\sigma^2(X_i)\sigma^2(X_{i'})I(i\in C_c)I(i'\in C_{c'})\right\}\right\} \tag{4.2.46}
$$

$$
= \frac{4}{N(k-1)^2}\left\{O(N^2)\right\} = O(N), \tag{4.2.47}
$$

where the first equality in (4.2.47) is due to the fact that $\sigma^2(X_i)$ and $E\left([Y_i - E(Y_i|\mathbf{X})]^4\,|\mathbf{X}\right)$ are uniformly bounded by Assumption (C1) and the summation over c in (4.2.43) and (4.2.44) and that over c and c' in (4.2.45) and (4.2.46) each contains no more than $2k$ terms.

From (4.2.42) and (4.2.47), we have

$$Var\left\{\frac{2\sqrt{N}}{N(k-1)}\sum_{c=1}^{N}\sum_{t=1}^{k}e_{ct}^{2}\right\}=O(N). \tag{4.2.48}$$

Due to (4.2.42) and (4.2.48) and by Theorem 14.4-1 in Bishop et al. (2007), we have

$$\frac{2\sqrt{N}}{N(k-1)}\sum_{c=1}^{N}\sum_{t=1}^{k}e_{ct}^{2}=O_p(N^{\frac{1}{2}}). \tag{4.2.49}$$

Similarly, it can be shown that the second term in (4.2.41) is $O_p(N^{\frac{1}{2}})$ and therefore the proof of (4.2.40) is completed.

From (4.2.38), (4.2.39) and (4.2.40),

$$\begin{aligned}|\Delta_{N,4}| &\leq \left[O_p(N^{\frac{1}{2}})\right]^{\frac{1}{2}}\left[O_p(N^{\frac{-3}{2}})\right]^{\frac{1}{2}}\\ &= O_p(N^{\frac{-1}{2}})=o_p(1), \text{ as } \quad N\to\infty.\end{aligned}$$

This completes the proof.

To obtain the asymptotic distribution of the test statistic $\sqrt{N}(B_N^*-W_N^*)$ in (4.2.7) under the null hypothesis, we only need to consider the first term $\sqrt{N}(B_N'-W_N')$ since the other four terms ($\Delta_{N,i};\ i=1,2,3,4$) are asymptotically negligible by Lemmas 4.2.3. Note that B_N' and W_N' are the average between-cell and within-cell variations for augmented observations with $Z_i=Y_i-G(X_i;\boldsymbol{\theta})$ as the response. Note that the conditional mean of Z_i given $X_i=x$ satisfies the null hypothesis of constant regression in (3.1.1). Therefore, the asymptotic distribution of $\sqrt{N}(B_N'-W_N')$ can be obtained by applying Theorem 3.1.3 in Chapter 3. This result is given in the following Theorem. We skip the details of the proof.

Theorem 4.2.4. *Under H_0 in (4.2.1) and Assumption (C),*

$$\sqrt{N}(B_N^*-W_N^*)\xrightarrow{d}N(0,\lim_{N\to\infty}\lambda_N) \text{ as } N\to\infty,$$

where

$$\begin{aligned}\lambda_N &= \textstyle\sum_{j<j'}^{N}E\left\{\frac{4\sigma^2(X_j)\sigma^2(X_{j'})}{N(k-1)^2}\left[[k-|j_*'-j_*|]^2+[k-|j_*'-j_*|]\right.\right.\\ &\quad \left.\left.-2I\left(|j_*'-j_*|\leq\frac{k-1}{2}\right)+O(N^{-1})\right]I(|j_*'-j_*|\leq k-1)\right\},\end{aligned} \tag{4.2.50}$$

and j_*',j_* are the ranks of $X_{j'}$ and X_j among the covariate values $\mathbf{X}=(X_1,\ldots,X_N)$..

4.2.3 Asymptotic distribution of the test statistic under local alternatives

Consider the following sequence of local alternative conditional expectations

$$m^*(x) = E_N(Y|X=x) = E_0(Y|X=x) + N^{-1/4}H(z;\boldsymbol{\gamma}), \tag{4.2.51}$$

where $E_0(Y|X=x) = G(x;\boldsymbol{\theta})$ is the conditional expectation of Y given X under the null hypothesis in (4.2.1), $H(z;\boldsymbol{\gamma})$ is a known function, z varies continuously with x, and $\boldsymbol{\gamma}$ is a vector of unknown parameters $(\gamma_1, ..., \gamma_q)$ with $q < \infty$. To express the dependence of z on x, we write $H(Z(x);\boldsymbol{\gamma})$ sometimes. In majority of situations, when it is clear, we just use the simple notation $H(z;\boldsymbol{\gamma})$. Let $Q^*_{ct};\ c = 1, \ldots, N,\ \ t = 1, \ldots, k$ be the augmented response values under the local alternatives in (4.2.51). Denote $G_{ct}(\boldsymbol{\theta})$ and $H_{ct}(\boldsymbol{\gamma})$ to be the $G(x;\boldsymbol{\theta})$ and $H(z;\boldsymbol{\gamma})$ functions evaluated at the covariate value for augmented observation Q^*_{ct}, respectively. Then, we can write Q^*_{ct} as

$$Q^*_{ct} = \varepsilon^*_{ct} + E(Q^*_{ct}|\mathbf{X}) = \varepsilon^*_{ct} + G_{ct}(\boldsymbol{\theta}) + N^{-1/4}H_{ct}(\boldsymbol{\gamma}),$$

where $\varepsilon^*_{ct} = Q^*_{ct} - E(Q^*_{ct}|\mathbf{X})$ can be viewed as the augmented data for $M_i = Y_i - G(X_i;\boldsymbol{\theta}) - N^{-1/4}H(Z_i;\boldsymbol{\gamma})$. Note that the conditional mean of M_i given $X_i = x$ satisfies the null hypothesis of constant regression in (3.1.1), but with $\mathrm{Var}(M_i|X_i)$ equals to $\mathrm{Var}(Y_i|X_i)$ under the alternative hypothesis in (4.2.51).

To define the test statistic under the local alternatives, let $r_{ct} = Q^*_{ct} - G_{ct}(\boldsymbol{\theta})$ and $r^*_{ct} = Q^*_{ct} - G_{ct}(\hat{\boldsymbol{\theta}})$. Also, denote $B^*_N(\mathbf{Q}^*)$ and $W^*_N(\mathbf{Q}^*)$ to be the average between-cell variations and the average within-cell variations under the local alternatives, respectively. Then $B^*_N(\mathbf{Q}^*)$ and $W^*_N(\mathbf{Q}^*)$ can be written as the following

$$B^*_N(\mathbf{Q}^*) = \frac{k}{N-1}\sum_{c=1}^{N} \left(\overline{r^*}_{c\cdot} - \overline{r^*}_{\cdot\cdot}\right)^2 \text{ and } W^*_N(\mathbf{Q}^*) = \frac{1}{N(k-1)}\sum_{c=1}^{N}\sum_{t=1}^{k}\left({r^*}_{ct} - \overline{r^*}_{c\cdot}\right)^2,$$

where $\overline{r^*}_{c\cdot} = k^{-1}\sum_{t=1}^{k} {r^*}_{ct}$ and $\overline{r^*}_{\cdot\cdot} = N^{-1}\sum_{c=1}^{N}\overline{r^*}_{c\cdot}$. Then the test statistic under the local alternatives is defined as $\sqrt{N}(B^*_N(\mathbf{Q}^*) - W^*_N(\mathbf{Q}^*))$. This statistic has the same form as that under the null hypothesis.

The following additional condition is needed for the result under the local alternatives:

Assumption (D): Suppose that X_i has bounded support $\chi = [a, b]$ and the function $H(z; \boldsymbol{\gamma}) : \chi \times \mathbb{R}^q \to \mathbb{R}$ is locally Lipschitz continuous with respect to its first argument. Further, assume that the fourth central moments of $H(Z_i; \boldsymbol{\gamma})$ are uniformly bounded.

Lemma 4.2.5. *If the Assumption (D) is satisfied, then*

$$H(Z_i; \boldsymbol{\theta})I(i \in C_c) - H(Z_j; \boldsymbol{\theta})I(j \in C_c) = O_p(N^{-1}),$$

uniformly in $i, j = 1, 2, ..., N$, for a given $c = 1, 2, ..., N$.

The proof of Lemma 4.2.5 is similar to the proof of Lemma 3.3.1 in Chapter 3 and is thus omitted.

In the following theorem, the asymptotic distribution of the test statistic under local alternatives is given.

Theorem 4.2.6. *Under the Assumptions (C) and (D), the limit $\lim_{N\to\infty} \lambda_{NA}$ exists and*

$$\sqrt{N}(B_N^*(\mathbf{Q}^*) - W_N^*(\mathbf{Q}^*)) \xrightarrow{d} N(k\sigma_H^2, \lim_{N\to\infty} \lambda_{NA}),$$

where λ_{NA} is defined similarly as λ_N in Theorem 4.2.4 but with $\sigma^2(X_j)$ calculated under the alternatives in (4.2.51) and

$$\sigma_H^2 = \int_{-\infty}^{\infty} H^2(Z(x); \boldsymbol{\gamma}) f(x) dx - \left(\int_{-\infty}^{\infty} H(Z(x); \boldsymbol{\gamma}) f(x) dx \right)^2 = Var(H(Z; \boldsymbol{\gamma})).$$

Proof

Note that $r_{ct}^* = Q_{ct}^* - G_{ct}(\hat{\boldsymbol{\theta}}) = \varepsilon_{ct}^* + G_{ct}(\boldsymbol{\theta}) + N^{-1/4} H_{ct}(\boldsymbol{\gamma}) - G_{ct}(\hat{\boldsymbol{\theta}})$. Let $\overline{\varepsilon^*}_{c\cdot} = k^{-1} \sum_{t=1}^k \varepsilon^*{}_{ct}$, $\overline{\varepsilon^*}_{\cdot\cdot} = N^{-1} \sum_{c=1}^N \overline{\varepsilon^*}_{c\cdot}$, $\overline{H}_{c\cdot}(\boldsymbol{\gamma}) = k^{-1} \sum_{t=1}^k H_{ct}(\boldsymbol{\gamma})$, and $\overline{H}_{\cdot\cdot}(\boldsymbol{\gamma}) = N^{-1} \sum_{c=1}^N \overline{H}_{c\cdot}(\boldsymbol{\gamma})$. Recall that $\overline{G}_{c\cdot}(\boldsymbol{\theta}) = k^{-1} \sum_{t=1}^k G_{ct}(\boldsymbol{\theta})$, and $\overline{G}_{\cdot\cdot}(\boldsymbol{\theta}) = N^{-1} \sum_{c=1}^N \overline{G}_{c\cdot}(\boldsymbol{\theta})$. Then, $B_N^*(\mathbf{Q}^*)$ and $W_N^*(\mathbf{Q}^*)$

can be written as

$$\begin{aligned}
B_N^*(\mathbf{Q}^*) &= \frac{k}{N-1}\sum_{c=1}^{N} \left(\overline{r^*}_{c\cdot} - \overline{r^*}_{\cdot\cdot}\right)^2 \\
&= \frac{k}{N-1}\sum_{c=1}^{N} \Big(\overline{\varepsilon^*}_{c\cdot} + \overline{G}_{c\cdot}(\boldsymbol{\theta}) + N^{-1/4}\overline{H}_{c\cdot}(\boldsymbol{\gamma}) - \overline{G}_{c\cdot}(\hat{\boldsymbol{\theta}}) \\
&\qquad -\overline{\varepsilon^*}_{\cdot\cdot} - \overline{G}_{\cdot\cdot}(\boldsymbol{\theta}) - N^{-1/4}\overline{H}_{\cdot\cdot}(\boldsymbol{\gamma}) + \overline{G}_{\cdot\cdot}(\hat{\boldsymbol{\theta}})\Big)^2 \\
&= \frac{k}{N-1}\Bigg[\sum_{c=1}^{N}\left(\overline{\varepsilon^*}_{c\cdot} - \overline{\varepsilon^*}_{\cdot\cdot}\right)^2 + \sum_{c=1}^{N}\Big(\left[\,\overline{G}_{c\cdot}(\boldsymbol{\theta}) - \overline{G}_{\cdot\cdot}(\boldsymbol{\theta})\right] - \Big[\overline{G}_{c\cdot}(\hat{\boldsymbol{\theta}}) - \overline{G}_{\cdot\cdot}(\hat{\boldsymbol{\theta}})\Big]\Big)^2 \\
&\quad +2\sum_{c=1}^{N}\left(\overline{\varepsilon^*}_{c\cdot} - \overline{\varepsilon^*}_{\cdot\cdot}\right)\Big(\left[\,\overline{G}_{c\cdot}(\boldsymbol{\theta}) - \overline{G}_{\cdot\cdot}(\boldsymbol{\theta})\right] - \Big[\overline{G}_{c\cdot}(\hat{\boldsymbol{\theta}}) - \overline{G}_{\cdot\cdot}(\hat{\boldsymbol{\theta}})\Big]\Big) \\
&\quad +N^{-1/2}\sum_{c=1}^{N}\left(\overline{H}_{c\cdot}(\boldsymbol{\gamma}) - \overline{H}_{\cdot\cdot}(\boldsymbol{\gamma})\right)^2 \\
&\quad +2N^{-1/4}\sum_{c=1}^{N}\left(\overline{H}_{c\cdot}(\boldsymbol{\gamma}) - \overline{H}_{\cdot\cdot}(\boldsymbol{\gamma})\right)\Big(\left[\,\overline{G}_{c\cdot}(\boldsymbol{\theta}) - \overline{G}_{\cdot\cdot}(\boldsymbol{\theta})\right] - \Big[\overline{G}_{c\cdot}(\hat{\boldsymbol{\theta}}) - \overline{G}_{\cdot\cdot}(\hat{\boldsymbol{\theta}})\Big]\Big) \\
&\quad +2N^{-1/4}\sum_{c=1}^{N}\left(\overline{\varepsilon^*}_{c\cdot} - \overline{\varepsilon^*}_{\cdot\cdot}\right)\left(\overline{H}_{c\cdot}(\boldsymbol{\gamma}) - \overline{H}_{\cdot\cdot}(\boldsymbol{\gamma})\right)\Bigg]. \qquad (4.2.52)
\end{aligned}$$

Similarly,

$$\begin{aligned}
W_N^*(\mathbf{Q}^*) &= \frac{1}{N(k-1)}\sum_{c=1}^{N}\sum_{t=1}^{k}\left(r^*_{ct} - \overline{r^*}_{c\cdot}\right)^2 \\
&= \frac{1}{N(k-1)}\Bigg[\sum_{c=1}^{N}\sum_{t=1}^{k}\left(\varepsilon^*_{ct} - \overline{\varepsilon^*}_{c\cdot}\right)^2 + \sum_{c=1}^{N}\sum_{t=1}^{k}\Big(\left[G_{ct}(\boldsymbol{\theta}) - \overline{G}_{c\cdot}(\boldsymbol{\theta})\right] - \Big[G_{ct}(\hat{\boldsymbol{\theta}}) - \overline{G}_{c\cdot}(\hat{\boldsymbol{\theta}})\Big]\Big)^2 \\
&\quad +2\sum_{c=1}^{N}\sum_{t=1}^{k}\left(\varepsilon^*_{ct} - \overline{\varepsilon^*}_{c\cdot}\right)\Big(\left[\,G_{ct}(\boldsymbol{\theta}) - \overline{G}_{c\cdot}(\boldsymbol{\theta})\right] - \Big[G_{ct}(\hat{\boldsymbol{\theta}}) - \overline{G}_{c\cdot}(\hat{\boldsymbol{\theta}})\Big]\Big) \\
&\quad +N^{-1/2}\sum_{c=1}^{N}\sum_{t=1}^{k}\left(H_{ct}(\boldsymbol{\gamma}) - \overline{H}_{c\cdot}(\boldsymbol{\gamma})\right)^2 \\
&\quad +2N^{-1/4}\sum_{c=1}^{N}\sum_{t=1}^{k}\left(H_{ct}(\boldsymbol{\gamma}) - \overline{H}_{c\cdot}(\boldsymbol{\gamma})\right)\Big(\left[G_{ct}(\boldsymbol{\theta}) - \overline{G}_{c\cdot}(\boldsymbol{\theta})\right] - \Big[G_{ct}(\hat{\boldsymbol{\theta}}) - \overline{G}_{c\cdot}(\hat{\boldsymbol{\theta}})\Big]\Big) \\
&\quad +2N^{-1/4}\sum_{c=1}^{N}\sum_{t=1}^{k}\left(\varepsilon^*_{ct} - \overline{\varepsilon^*}_{c\cdot}\right)\left(H_{ct}(\boldsymbol{\gamma}) - \overline{H}_{c\cdot}(\boldsymbol{\gamma})\right)\Bigg].
\end{aligned}$$

Then, we can write the test statistic as

$$\begin{aligned}\sqrt{N}(B_N^*(\mathbf{Q}^*) - W_N^*(\mathbf{Q}^*)) &= \sqrt{N}(B_N(\varepsilon^*) - W_N(\varepsilon^*)) + \Delta_{N,1} + \Delta_{N,2}^* - \Delta_{N,3} - \Delta_{N,4}^* \\ &+ \Delta_{N,5} + \Delta_{N,6} + \Delta_{N,7} - \Delta_{N,8} - \Delta_{N,9} - \Delta_{N,10}, \end{aligned} \tag{4.2.53}$$

where $\Delta_{N,1}, \Delta_{N,3}$ are defined in (4.2.8), (4.2.10),respectively, and

$$\Delta_{N,2}^* = \frac{2k\sqrt{N}}{N-1}\sum_{c=1}^{N}(\overline{\varepsilon^*}_{c\cdot} - \overline{\varepsilon^*}_{\cdot\cdot})\left(\left[\,\overline{G}_{c\cdot}(\boldsymbol{\theta}) - \overline{G}_{\cdot\cdot}(\boldsymbol{\theta})\right] - \left[\overline{G}_{c\cdot}(\hat{\boldsymbol{\theta}}) - \overline{G}_{\cdot\cdot}(\hat{\boldsymbol{\theta}})\right]\right) \tag{4.2.54}$$

$$\Delta_{N,4}^* = \frac{2\sqrt{N}}{N(k-1)}\sum_{c=1}^{N}\sum_{t=1}^{k}(\varepsilon^*{}_{ct} - \overline{\varepsilon^*}_{c\cdot})\left(\left[\,G_{ct}(\boldsymbol{\theta}) - \overline{G}_{c\cdot}(\boldsymbol{\theta})\right] - \left[G_{ct}(\hat{\boldsymbol{\theta}}) - \overline{G}_{c\cdot}(\hat{\boldsymbol{\theta}})\right]\right) \tag{4.2.55}$$

$$\Delta_{N,5} = \frac{k}{N-1}\sum_{c=1}^{N}\left(\overline{H}_{c\cdot}(\boldsymbol{\gamma}) - \overline{H}_{\cdot\cdot}(\boldsymbol{\gamma})\right)^2 \tag{4.2.56}$$

$$\Delta_{N,6} = \frac{2k\sqrt{N}}{N-1}\sum_{c=1}^{N}N^{-1/4}\left(\overline{H}_{c\cdot}(\boldsymbol{\gamma}) - \overline{H}_{\cdot\cdot}(\boldsymbol{\gamma})\right)\left(\left[\,\overline{G}_{c\cdot}(\boldsymbol{\theta}) - \overline{G}_{\cdot\cdot}(\boldsymbol{\theta})\right] - \left[\overline{G}_{c\cdot}(\hat{\boldsymbol{\theta}}) - \overline{G}_{\cdot\cdot}(\hat{\boldsymbol{\theta}})\right]\right) \tag{4.2.57}$$

$$\Delta_{N,7} = \frac{2k\sqrt{N}}{N-1}\sum_{c=1}^{N}(\overline{\varepsilon^*}_{c\cdot} - \overline{\varepsilon^*}_{\cdot\cdot})\,N^{-1/4}\left(\overline{H}_{c\cdot}(\boldsymbol{\gamma}) - \overline{H}_{\cdot\cdot}(\boldsymbol{\gamma})\right) \tag{4.2.58}$$

$$\Delta_{N,8} = \frac{1}{N(k-1)}\sum_{c=1}^{N}\sum_{t=1}^{k}\left(H_{ct}(\boldsymbol{\gamma}) - \overline{H}_{c\cdot}(\boldsymbol{\gamma})\right)^2 \tag{4.2.59}$$

$$\Delta_{N,9} = \frac{2\sqrt{N}}{N(k-1)}\sum_{c=1}^{N}\sum_{t=1}^{k}N^{-1/4}\left(H_{ct}(\boldsymbol{\gamma}) - \overline{H}_{c\cdot}(\boldsymbol{\gamma})\right)\left(\left[G_{ct}(\boldsymbol{\theta}) - \overline{G}_{c\cdot}(\boldsymbol{\theta})\right] - \left[G_{ct}(\hat{\boldsymbol{\theta}}) - \overline{G}_{c\cdot}(\hat{\boldsymbol{\theta}})\right]\right) \tag{4.2.60}$$

$$\Delta_{N,10} = \frac{2\sqrt{N}}{N(k-1)}\sum_{c=1}^{N}\sum_{t=1}^{k}(\varepsilon^*_{ct} - \overline{\varepsilon^*}_{c\cdot})\,N^{-1/4}\left(H_{ct}(\boldsymbol{\gamma}) - \overline{H}_{c\cdot}(\boldsymbol{\gamma})\right), \tag{4.2.61}$$

and $B_N(\varepsilon^*) = \frac{k}{N-1}\sum_{c=1}^{N}(\overline{\varepsilon^*}_{c\cdot} - \overline{\varepsilon^*}_{\cdot\cdot})^2$, $W_N(\varepsilon^*) = \frac{1}{N(k-1)}\sum_{c=1}^{N}\sum_{t=1}^{k}(\varepsilon^*_{ct} - \overline{\varepsilon^*}_{c\cdot})^2$ are the average between-cell and within-cell variations for augmented observations with $M_i = Y_i - \left(G(X_i;\boldsymbol{\theta}) + N^{-1/4}H(Z_i;\boldsymbol{\gamma})\right)$ as the response. Note that the conditional mean of M_i given $X_i = x$ satisfies the null hypothesis of constant regression in (3.1.1). But $Var(M_i|X_i)$ is equal to $Var(Y_i|X_i)$.

Therefore, the result of Theorem 3.1.3 in Chapter 3 implies that

$$\sqrt{N}(B_N(\varepsilon^*) - W_N(\varepsilon^*)) \xrightarrow{d} N(0, \lim_{N\to\infty}\lambda_{NA}), \tag{4.2.62}$$

where λ_{NA} is defined similarly as λ_N in (4.2.50) but with $\sigma^2(X_j)$ calculated under the alternatives in (4.2.51).

By parts (1) and (3) of Lemma 4.2.3,

$$\Delta_{N,i} \xrightarrow{p} 0, \text{ as } N \to \infty, \text{ for } i = 1, 3. \tag{4.2.63}$$

Also, the proof that

$$\Delta^*_{N,i} \xrightarrow{p} 0, \text{ as } N \to \infty, \text{ for } i = 2, 4., \tag{4.2.64}$$

is similar to the proof of parts (2) and (4) in Lemma 4.2.3.

In addition, we will show in Lemma 4.2.7 that

$$\Delta_{N,i} \xrightarrow{p} 0, \text{ as } N \to \infty, \text{ for } i = 6, 7, 8, 9, 10. \tag{4.2.65}$$

Thus, we only need to consider $\Delta_{N,5}$ in (4.2.53) to find the asymptotic mean of the test statistic under the local alternatives. By Lemma 4.2.5 and Assumption (D), we have

$$\overline{H}_{c\cdot}(\boldsymbol{\gamma}) = \frac{1}{k}\sum_{t=1}^{k} H_{ct}(\boldsymbol{\gamma}) = \frac{1}{k}\sum_{i=1}^{N} H(Z_i;\boldsymbol{\theta})I(i \in C_c) = H(Z(X_{(c)});\boldsymbol{\gamma}) + O_p(N^{-1}), \tag{4.2.66}$$

and

$$\overline{H}_{\cdot\cdot}(\boldsymbol{\gamma}) = \frac{1}{N}\sum_{c=1}^{N} \overline{H}_{c\cdot}(\boldsymbol{\gamma}) = \frac{1}{N}\sum_{c=1}^{N} H(Z_c;\boldsymbol{\gamma}) + O_p(N^{-1}) = \overline{H}(\boldsymbol{\gamma}) + O_p(N^{-1}), \tag{4.2.67}$$

where $\overline{H}(\boldsymbol{\gamma}) = N^{-1}\sum_{c=1}^{N} H(Z_c;\boldsymbol{\gamma})$. Therefore,

$$\begin{aligned}
\Delta_{N,5} &= \frac{k}{N-1}\sum_{c=1}^{N}\left(\overline{H}_{c\cdot}(\boldsymbol{\gamma}) - \overline{H}_{\cdot\cdot}(\boldsymbol{\gamma})\right)^2 \\
&= \frac{k}{N-1}\sum_{c=1}^{N}\left(H(Z(X_{(c)});\boldsymbol{\gamma}) - \overline{H}(\boldsymbol{\gamma}) + O_p(N^{-1})\right)^2 \\
&= \frac{k}{N-1}\sum_{c=1}^{N}\left(H(Z_c;\boldsymbol{\gamma}) - \overline{H}(\boldsymbol{\gamma})\right)^2 + O_p(N^{-2}) \qquad (4.2.68)
\end{aligned}$$

Since $X_1, X_2, ..., X_N$ are i.i.d., then $H(Z_1;\boldsymbol{\gamma}), H(Z_2;\boldsymbol{\gamma}), ..., H(Z_N;\boldsymbol{\gamma})$ are i.i.d. as well. Therefore, we can write the first term in (4.2.68) as

$$\frac{k}{N-1}\sum_{c=1}^{N}\left(H(Z_c;\boldsymbol{\gamma}) - \overline{H}(\boldsymbol{\gamma})\right)^2 = k\widehat{\sigma}^2_H, \tag{4.2.69}$$

where $\widehat{\sigma}_H^2$ is the sample variance of $H(Z_1;\boldsymbol{\gamma}), H(Z_2;\boldsymbol{\gamma}), ..., H(Z_N;\boldsymbol{\gamma})$. By the Weak Law of Large Numbers,

$$k\widehat{\sigma}_H^2 \xrightarrow{p} k\sigma_H^2 = k\mathrm{Var}(H(Z;\boldsymbol{\gamma})) = k\left[\int_{-\infty}^{\infty} H^2(Z(x);\boldsymbol{\gamma}) f(x)dx - \left(\int_{-\infty}^{\infty} H(Z(x);\boldsymbol{\gamma}) f(x)dx\right)^2\right], \quad (4.2.70)$$

as k stays fixed and $N \to \infty$.

From (4.2.68), (4.2.69), and (4.2.70), we have

$$\Delta_{N,5} \xrightarrow{p} k\sigma_H^2. \quad (4.2.71)$$

Putting (4.2.62), (4.2.63), (4.2.64), (4.2.65), and (4.2.71) in (4.2.53) and by applying Slutsky's theorem, we have

$$\sqrt{N}(B_N^*(\mathbf{Q}^*) - W_N^*(\mathbf{Q}^*)) \xrightarrow{d} N(k\sigma_H^2, \lim_{N\to\infty} \lambda_{NA}).$$

This completes the proof.

Lemma 4.2.7. *Under Assumptions (C) and (D),*

$$\Delta_{N,i} \xrightarrow{p} 0, \text{ as } N \to \infty, \text{ for } i = 6, 7, 8, 9, 10. \quad (4.2.72)$$

where $\Delta_{N,i}$, $i = 6, 7, 8, 9, 10$, are defined in (4.2.57), (4.2.58), (4.2.59), (4.2.60), and (4.2.61), respectively.

Proof of Lemma 4.2.7

First, we will show that

$$\Delta_{N,6} \xrightarrow{p} 0, \text{ as } N \to \infty. \quad (4.2.73)$$

From (4.2.57), we have

$$\Delta_{N,6} = \frac{2k\sqrt{N}}{N-1} \sum_{c=1}^{N} N^{-1/4} \left(\overline{H}_{c\cdot}(\boldsymbol{\gamma}) - \overline{H}_{\cdot\cdot}(\boldsymbol{\gamma})\right) \left(\left[\,\overline{G}_{c\cdot}(\boldsymbol{\theta}) - \overline{G}_{\cdot\cdot}(\boldsymbol{\theta})\right] - \left[\overline{G}_{c\cdot}(\hat{\boldsymbol{\theta}}) - \overline{G}_{\cdot\cdot}(\hat{\boldsymbol{\theta}})\right]\right)$$

By Hölder's inequality,

$$
\begin{aligned}
|\Delta_{N,6}| &\leq 2\left[\frac{k\sqrt{N}}{N-1}\sum_{c=1}^{N}\left(N^{-1/4}\left(H(X_{(c)};\boldsymbol{\gamma})-\overline{H}(\boldsymbol{\gamma})\right)\right)^2\right]^{\frac{1}{2}} \\
&\quad \times\left[\frac{k\sqrt{N}}{N-1}\sum_{c=1}^{N}\left(\left[\,\overline{G}_{c\cdot}(\boldsymbol{\theta})-\overline{G}_{\cdot\cdot}(\boldsymbol{\theta})\right]-\left[\overline{G}_{c\cdot}(\hat{\boldsymbol{\theta}})-\overline{G}_{\cdot\cdot}(\hat{\boldsymbol{\theta}})\right]\right)^2\right]^{\frac{1}{2}} \\
&= 2\left[\Delta_{N,5}\right]^{\frac{1}{2}}\left[\Delta_{N,1}\right]^{\frac{1}{2}}\xrightarrow{p}0, \qquad (4.2.74)
\end{aligned}
$$

where $\Delta_{N,1}$ and $\Delta_{N,5}$ are defined in (4.2.8) and (4.2.56) and the convergence in probability in (4.2.74) is due to (4.2.63) and (4.2.71). This completes the proof of (4.2.73).

Second, we will show that

$$
\Delta_{N,7}\xrightarrow{p}0,\ \text{ as } N\to\infty. \qquad (4.2.75)
$$

From (4.2.58), we have

$$
\Delta_{N,7} = \frac{2k\sqrt{N}}{N-1}\sum_{c=1}^{N}\left(\overline{\varepsilon^*}_{c\cdot}-\overline{\varepsilon^*}_{\cdot\cdot}\right)N^{-1/4}\left(\overline{H}_{c\cdot}(\boldsymbol{\gamma})-\overline{H}_{\cdot\cdot}(\boldsymbol{\gamma})\right).
$$

Using (4.2.66) and (4.2.67), we can be write

$$
\Delta_{N,7} = \sqrt{N}k(N-1)^{-1}\sum_{c=1}^{N}\left[2N^{-1/4}\left(H(Z_c;\boldsymbol{\gamma})-\overline{H}(\boldsymbol{\gamma})\right)\left(\overline{\varepsilon^*}_{c\cdot}-\overline{\varepsilon^*}_{\cdot\cdot}\right)\right]+o_p(1).
$$

Denote $U_c=H(Z_c;\boldsymbol{\gamma})-E(H(Z_c;\boldsymbol{\gamma}))$ and $\overline{U}_{\cdot}=N^{-1}\sum_{c=1}^{N}U_c$, then we can write

$$
\begin{aligned}
\Delta_{N,7} &= 2kN^{\frac{-1}{4}}\left[\frac{\sqrt{N}}{(N-1)}\sum_{c=1}^{N}\left(H(Z_c;\boldsymbol{\gamma})-\overline{H}(\boldsymbol{\gamma})\right)\left(\overline{\varepsilon^*}_{c\cdot}-\overline{\varepsilon^*}_{\cdot\cdot}\right)\right]+o_p(1) \\
&= 2kN^{\frac{-1}{4}}\left[\frac{\sqrt{N}}{(N-1)}\sum_{c=1}^{N}\left(\left[H(Z_c;\boldsymbol{\gamma})-E(H(Z_c;\boldsymbol{\gamma}))\right]-\left[\overline{H}(\boldsymbol{\gamma})-E(H(Z_c;\boldsymbol{\gamma}))\right]\right)\right. \\
&\qquad \left.\times\left(\overline{\varepsilon^*}_{c\cdot}-\overline{\varepsilon^*}_{\cdot\cdot}\right)\right]+o_p(1) \\
&= 2kN^{\frac{-1}{4}}\left[\frac{\sqrt{N}}{(N-1)}\sum_{c=1}^{N}\left(U_c-\overline{U}_{\cdot}\right)\left(\overline{\varepsilon^*}_{c\cdot}-\overline{\varepsilon^*}_{\cdot\cdot}\right)\right]+o_p(1) \\
&= 2kN^{\frac{-1}{4}}\frac{\sqrt{N}}{(N-1)}\left[\sum_{c=1}^{N}U_c\overline{\varepsilon^*}_{c\cdot}-N\ \overline{U}_{\cdot}\ \overline{\varepsilon^*}_{\cdot\cdot}\right]+o_p(1) \\
&= 2kN^{\frac{-1}{4}}\left[\frac{\sqrt{N}}{(N-1)}\sum_{c=1}^{N}U_c\overline{\varepsilon^*}_{c\cdot}\right]-\frac{2kN^{\frac{1}{4}}}{(N-1)}\left[\sqrt{N}\ \overline{U}_{\cdot}\right]\ \left[\sqrt{N}\ \overline{\varepsilon^*}_{\cdot\cdot}\right]+o_p(1). \qquad (4.2.76)
\end{aligned}
$$

Next, we will show that

$$\left[\frac{\sqrt{N}}{(N-1)}\sum_{c=1}^{N}U_c\overline{\varepsilon^*}_{c\cdot}\right] = O_p(1) \tag{4.2.77}$$

and therefore the first term in (4.2.76) is $o_p(1)$. Note that $E(\overline{\varepsilon^*}_{c\cdot}|\mathbf{X}) = E(\overline{Q^*}_{c\cdot} - E(\overline{Q^*}_{c\cdot}|\mathbf{X})|\mathbf{X}) = 0$ and U_c is a function of X_c. Therefore, we have

$$E\left[\frac{\sqrt{N}}{(N-1)}\sum_{c=1}^{N}U_c\overline{\varepsilon^*}_{c\cdot}\right] = \frac{\sqrt{N}}{(N-1)}\sum_{c=1}^{N}E\left[U_c E(\overline{\varepsilon^*}_{c\cdot}|\mathbf{X})\right] = 0, \tag{4.2.78}$$

and

$$\begin{aligned}
& Var\left[\frac{\sqrt{N}}{(N-1)}\sum_{c=1}^{N}U_c\overline{\varepsilon^*}_{c\cdot}\right] \\
= \;& \frac{N}{(N-1)^2}E\left[\sum_{c=1}^{N}U_c\overline{\varepsilon^*}_{c\cdot}\right]^2 \\
= \;& \frac{N}{(N-1)^2}E\left[\sum_{c=1}^{N}U_c^2\overline{\varepsilon^*}^2_{c\cdot} + \sum_{c\neq c'}^{N}U_c\overline{\varepsilon^*}_{c\cdot}U_{c'}\overline{\varepsilon^*}_{c'\cdot}\right] \\
= \;& \frac{N}{(N-1)^2}\left[\sum_{c=1}^{N}E\left(U_c^2\overline{\varepsilon^*}^2_{c\cdot}\right)\right] + \frac{N}{(N-1)^2}\left[\sum_{c\neq c'}^{N}E\left(U_cU_{c'}\overline{\varepsilon^*}_{c\cdot}\overline{\varepsilon^*}_{c'\cdot}\right)\right].
\end{aligned} \tag{4.2.79}$$

Denote the first term and second term in (4.2.79) as $\delta_{N,1}$ and $\delta_{N,2}$, respectively. Then

$$
\begin{aligned}
\delta_{N,1} &= \frac{N}{(N-1)^2}\left[\sum_{c=1}^{N} E\left(U_c^2 E(\overline{\varepsilon^*}_{c\cdot}^2|\mathbf{X})\right)\right] \\
&= \frac{N}{(N-1)^2}\left[\sum_{c=1}^{N} E\left(U_c^2 E((\overline{Q^*}_{c\cdot} - E(\overline{Q^*}_{c\cdot}|\mathbf{X}))^2|\mathbf{X})\right)\right] \\
&= \frac{N}{(N-1)^2}\sum_{c=1}^{N} E\left\{U_c^2 E\left\{\left(\frac{1}{k}\sum_{i=1}^{N}(Y_i - E(Y_i|\mathbf{X}))I(i\in C_c)\right)^2\middle|\mathbf{X}\right\}\right\} \\
&= \frac{N}{k^2(N-1)^2}\sum_{c=1}^{N} E\left\{U_c^2 E\left\{\left(\sum_{i=1}^{N}(Y_i - E(Y_i|\mathbf{X}))^2 I(i\in C_c)\right.\right.\right. \\
&\quad \left.\left.\left. + \sum_{i\neq i'}^{N}(Y_i - E(Y_i|\mathbf{X}))I(i\in C_c)(Y_{i'} - E(Y_{i'}|\mathbf{X}))I(i'\in C_c)\right)\middle|\mathbf{X}\right\}\right\} \\
&= \frac{N}{k^2(N-1)^2}\sum_{c=1}^{N}\sum_{i=1}^{N} E\left\{U_c^2 E((Y_i - E(Y_i|\mathbf{X}))^2\big|\,\mathbf{X})I(i\in C_c)\right\} \qquad (4.2.80) \\
&= \frac{N}{k^2(N-1)^2}\sum_{i=1}^{N}\sum_{c=1}^{N} E\left\{U_c^2\sigma^2(X_i)I(i\in C_c)\right\}, \qquad (4.2.81)
\end{aligned}
$$

where the equality in (4.2.80) is due to the fact that Y_i and $Y_{i'}$ are independent when $i \neq i'$. Similarly,

$$
\begin{aligned}
\delta_{N,2} &= \frac{N}{(N-1)^2}\left[\sum_{c\neq c'}^{N} E\left(U_c U_{c'} E(\overline{\varepsilon^*}_{c\cdot}\overline{\varepsilon^*}_{c'\cdot}|\mathbf{X})\right)\right] \\
&= \frac{N}{(N-1)^2}\sum_{c\neq c'}^{N} E\left\{U_c U_{c'} E\left\{\left(\frac{1}{k}\sum_{i=1}^{N}(Y_i - E(Y_i|\mathbf{X}))I(i\in C_c)\right)\right.\right. \\
&\quad \left.\left.\times\left(\frac{1}{k}\sum_{i'=1}^{N}(Y_{i'} - E(Y_{i'}|\mathbf{X}))I(i'\in C_{c'})\right)\middle|\mathbf{X}\right\}\right\} \\
&= \frac{N}{k^2(N-1)^2}\sum_{i=1}^{N}\sum_{c\neq c'}^{N} E\left\{U_c U_{c'} E((Y_i - E(Y_i|\mathbf{X}))^2\big|\,\mathbf{X})I(i\in C_c)I(i\in C_{c'})\right\} \\
&= \frac{N}{k^2(N-1)^2}\sum_{i=1}^{N}\sum_{c\neq c'}^{N} E\left\{U_c U_{c'}\sigma^2(X_i)I(i\in C_c\cap C_{c'})\right\}, \qquad (4.2.82)
\end{aligned}
$$

Consider individual terms under the summation in (4.2.81) and (4.2.82). By Cauchy-

Schwarz inequality and Assumptions (C) and (D),

$$\begin{aligned}
& E\left\{U_c^2\sigma^2(X_i)I(i\in C_c)\right\} \\
\leq\ & E\left\{U_c^2\sigma^2(X_i)\right\} \\
\leq\ & \left[E(U_c^4)\right]^{\frac{1}{2}}\left[E\left(\sigma^2(X_i)\right)^2\right]^{\frac{1}{2}} \\
=\ & \left[E(U_c^4)\right]^{\frac{1}{2}}\left[E\left(E((Y_i-E(Y_i|\mathbf{X}))^2\big|\,\mathbf{X})\right)^2\right]^{\frac{1}{2}} \\
\leq\ & \left[E(U_c^4)\right]^{\frac{1}{2}}\left[E\left(E((Y_i-E(Y_i|\mathbf{X}))^4\big|\,\mathbf{X})\right)^2\right]^{\frac{1}{2}} \\
<\ & \infty. \qquad (4.2.83)
\end{aligned}$$

Similarly,

$$\begin{aligned}
& \left|E\left\{U_cU_{c'}\sigma^2(X_i)I(i\in C_c\cap C_{c'})\right\}\right| \\
\leq\ & E\left\{|U_cU_{c'}|\sigma^2(X_i)I(i\in C_c\cap C_{c'})\right\} \\
\leq\ & E\left\{|U_cU_{c'}|\sigma^2(X_i)\right\} \\
\leq\ & \left[E\left(U_cU_{c'}\right)^2\right]^{\frac{1}{2}}\left[E\left(\sigma^2(X_i)\right)^2\right]^{\frac{1}{2}} \\
=\ & \left[E\left(U_c^2\right)\right]^{\frac{1}{2}}\left[E\left(U_{c'}^2\right)\right]^{\frac{1}{2}}\left[E\left(E((Y_i-E(Y_i|\mathbf{X}))^2\big|\,\mathbf{X})\right)^2\right]^{\frac{1}{2}} \\
\leq\ & \left[E(U_c^4)\right]^{\frac{1}{2}}\left[E(U_{c'}^4)\right]^{\frac{1}{2}}\left[E\left(E((Y_i-E(Y_i|\mathbf{X}))^4\big|\,\mathbf{X})\right)^2\right]^{\frac{1}{2}} \\
<\ & \infty. \qquad (4.2.84)
\end{aligned}$$

Note that X_i can only be used to augment at most $2k$ cells. That is, if the rank of X_i is r, then X_i can not be used to augment cells whose x values have ranks not in the set of positive integers $\{\max\{1,r-k\},...,\min\{r+k,N\}\}$. Therefore, the summation over c in (4.2.81) and that over c and c' in (4.2.82) each contains no more than $2k$ terms. As a result, the two terms $\delta_{N,1}$ and $\delta_{N,2}$ are $O(1)$ and therefore,

$$Var\left[\frac{\sqrt{N}}{(N-1)}\sum_{c=1}^{N}U_c\overline{\varepsilon^*}_{c\cdot}\right]=O(1). \qquad (4.2.85)$$

Due to (4.2.78) and (4.2.85), the proof of (4.2.77) is complete by applying Theorem 14.4-1 in Bishop et al. (2007).

Next, we will show that the second term in (4.2.76) is $o_p(1)$. The second term in (4.2.76) is

$$\frac{-2kN^{\frac{1}{4}}}{(N-1)}\left[\sqrt{N}\ \ \overline{U}_{\cdot}\right]\ \ \left[\sqrt{N}\ \ \overline{\varepsilon^*}_{\cdot\cdot}\right].$$

Using the same technique of the proof of (4.2.77), it can be shown that

$$\left[\sqrt{N}\ \ \overline{\varepsilon^*}_{\cdot\cdot}\right] = O_p(1).$$

In addition,

$$\left[\sqrt{N}\ \ \overline{U}_{\cdot}\right] = O_p(1) \tag{4.2.86}$$

is a result of Central Limit Theorem (CLT) applied to $U_1, ..., U_N$ since they are i.i.d. due to the fact that $X_1, ..., X_N$ are i.i.d..

Consequently,

$$\Delta_{N,7} = O_p(N^{\frac{-1}{4}}) + O_p\left(\frac{N^{\frac{1}{4}}}{N-1}\right) + o_p(1) = o_p(1), \text{ as } N \to \infty.$$

This completes the proof of (4.2.75).

Third, we will show that

$$\Delta_{N,8} \xrightarrow{p} 0, \text{ as } N \to \infty. \tag{4.2.87}$$

From (4.2.59), we have

$$\Delta_{N,8} \quad = \quad \frac{1}{N(k-1)}\sum_{c=1}^{N}\sum_{t=1}^{k}\left(H_{ct}(\boldsymbol{\gamma}) - \overline{H}_{c\cdot}(\boldsymbol{\gamma})\right)^2$$

By Lemma 4.2.5, we have $H_{ct}(\boldsymbol{\gamma}) - \overline{H}_{c\cdot}(\boldsymbol{\gamma}) = O_p(N^{-1})$. Therefore,

$$\Delta_{N,8} = O_p(N^{-2}) \tag{4.2.88}$$

and therefore $\Delta_{N,8}$ is $o_p(1)$. This completes the proof of (4.2.87).

Fourth, we will show that

$$\Delta_{N,9} \xrightarrow{p} 0, \text{ as } N \to \infty. \tag{4.2.89}$$

From (4.2.60), we have

$$\Delta_{N,9} = \frac{2\sqrt{N}}{N(k-1)} \sum_{c=1}^{N} \sum_{t=1}^{k} N^{-1/4} \left(H_{ct}(\boldsymbol{\gamma}) - \overline{H}_{c\cdot}(\boldsymbol{\gamma})\right) \left(\left[\, G_{ct}(\boldsymbol{\theta}) - \overline{G}_{c\cdot}(\boldsymbol{\theta})\right] - \left[G_{ct}(\hat{\boldsymbol{\theta}}) - \overline{G}_{c\cdot}(\hat{\boldsymbol{\theta}})\right]\right)$$

By Hölder's inequality and the definition of $\Delta_{N,1}$ in (4.2.8) and $\Delta_{N,8}$ in (4.2.59),

$$\begin{aligned} |\Delta_{N,9}| &\leq 2 \left[\frac{\sqrt{N}}{N(k-1)} \sum_{c=1}^{N} \sum_{t=1}^{k} N^{-1/2} \left(H_{ct}(\boldsymbol{\gamma}) - \overline{H}_{c\cdot}(\boldsymbol{\gamma})\right)^2\right]^{\frac{1}{2}} \\ &\quad \times \left[\frac{\sqrt{N}}{N(k-1)} \sum_{c=1}^{N} \sum_{t=1}^{k} \left(\left[\, G_{ct}(\boldsymbol{\theta}) - \overline{G}_{c\cdot}(\boldsymbol{\theta})\right] - \left[G_{ct}(\hat{\boldsymbol{\theta}}) - \overline{G}_{c\cdot}(\hat{\boldsymbol{\theta}})\right]\right)^2\right]^{\frac{1}{2}} \\ &= 2\,[\Delta_{N,8}]^{\frac{1}{2}}\,[\Delta_{N,1}]^{\frac{1}{2}} \xrightarrow{p} 0, \end{aligned} \tag{4.2.90}$$

where the convergence in probability in (4.2.90) is due to (4.2.63) and (4.2.87). This completes the proof of (4.2.89).

Finally, we will show that

$$\Delta_{N,10} \xrightarrow{p} 0, \text{ as } N \to \infty. \tag{4.2.91}$$

From (4.2.61), we have

$$\Delta_{N,10} = \frac{2\sqrt{N}}{N(k-1)} \sum_{c=1}^{N} \sum_{t=1}^{k} (\varepsilon^*_{ct} - \overline{\varepsilon^*}_{c\cdot})\, N^{-1/4} \left(H_{ct}(\boldsymbol{\gamma}) - \overline{H}_{c\cdot}(\boldsymbol{\gamma})\right)$$

Using Hölder's inequality and (4.2.59),

$$\begin{aligned} |\Delta_{N,10}| \leq &\quad \left[\frac{2\sqrt{N}}{N(k-1)} \sum_{c=1}^{N} \sum_{t=1}^{k} (\varepsilon^*_{ct} - \overline{\varepsilon^*}_{c\cdot})^2\right]^{\frac{1}{2}} \\ &\quad \times \left[\frac{2\sqrt{N}}{N(k-1)} \sum_{c=1}^{N} \sum_{t=1}^{k} N^{-1/2} \left(H_{ct}(\boldsymbol{\gamma}) - \overline{H}_{c\cdot}(\boldsymbol{\gamma})\right)^2\right]^{\frac{1}{2}} \\ &= \left[\frac{2\sqrt{N}}{N(k-1)} \sum_{c=1}^{N} \sum_{t=1}^{k} (\varepsilon^*_{ct} - \overline{\varepsilon^*}_{c\cdot})^2\right]^{\frac{1}{2}} [2\ \ \Delta_{N,8}]^{\frac{1}{2}} \end{aligned} \tag{4.2.92}$$

It can be shown that

$$\frac{2\sqrt{N}}{N(k-1)} \sum_{c=1}^{N} \sum_{t=1}^{k} (\varepsilon^*{}_{ct} - \overline{\varepsilon^*}_{c\cdot})^2 = O_p(N^{\frac{1}{2}}). \tag{4.2.93}$$

The proof of (4.2.93) is similar to that of (4.2.40).

From (4.2.88), (4.2.92), and (4.2.93), we have

$$|\Delta_{N,10}| \leq \left[O_p(N^{\frac{1}{2}})\right]^{\frac{1}{2}} \left[O_p(N^{-2})\right]^{\frac{1}{2}} = O_p(N^{-\frac{3}{4}}) = o_p(1), \quad \text{as } N \to \infty.$$

This completes the proof.

4.3 Examples

4.3.1 Numerical studies

This section will present the results of a simulation study conducted to investigate the performance of our test. Our test depends on a parameter k to determine the number of nearest neighbors for data augmentation. In Chapter 5, a discussion will be given on how to select the parameter k based on the idea of the Least Squares Cross-Validation (LSCV) procedure of Hardle et al. (1988). The regression function in this adopted procedure is estimated using k-nearest neighbors with neighbors defined through the ranks of the predictor variable. Then k is selected from a set of small odd positive integers that minimizes the leave-one-out Least Squares Cross-Validation error (see Chapter 5 for more details). For data generated under alternatives, we found that large k tends to give larger least squares error specially in the case of high frequency alternatives. For data augmentation, the smallest odd positive integer value for k is 3. Consequently in this section, the results of our test (denoted as GSW) are based on number of nearest neighbors equal to $k = 3$.

For comparison, we also report the corresponding results for the order selection test of Kuchibhatla and Hart (1996) based on the test statistic defined in (4.1.3). Two versions of critical value approximation are considered for this test, one based on bootstrap resampling procedure as recommended by Kuchibhatla and Hart (1996) and Hart (1997) (denoted as BOS), and the other based on wild bootstrap of Hardle and Mammen (1993) which was suggested by Kuchibhatla and Hart (1996) to deal with heteroscedastic nonlinear regression models and used in Chen et al. (2001) for testing constat regression with heteroscedastic errors (denoted as WBOS). In this study, we generated data from the following four models with sample size $N = 50$:

- Model M_0: $Y_i = \frac{e^{-b_1 X_i}}{b_2 + b_3 X_i} + \epsilon_i$;
- Model M_1: $Y_i = \frac{e^{-b_1 X_i}}{b_2 + b_3 X_i} + \cos(10\pi X_i) + \epsilon_i$;
- Model M_2: $Y_i = \frac{e^{-b_1 X_i}}{b_2 + b_3 X_i} + \sin(10\pi X_i) + \epsilon_i$;
- Model M_3: $Y_i = \frac{e^{-b_1 X_i}}{b_2 + b_3 X_i} + 2e^{-2X_i}\cos(10\pi X_i) + \epsilon_i$,

where the covariate values are independently generated from Uniform(0,1) and the parameters b_1, b_2, b_3 are considered to be $-5, 20, 0.6$, respectively. For each model above, the errors ϵ_i were independently generated from each of the following four distributions:

- $\epsilon_i \sim$ Uniform$(-0.8, 0.8)$ (denoted as Unif);
- $\epsilon_i \sim$ Normal$(0, 0.2)$ (denoted as Normal);
- $\epsilon_i = V_i/3$, where V_i follows t-distribution with 5 degrees of freedom (denoted as T);
- $\epsilon_i = 1.5X_i \cdot e_i$ where $e_i \sim$ Uniform$(-0.8, 0.8)$. This represents heteroscedastic errors (denoted as Heter).

For all tests, Model M_0 is considered as the null model to find the empirical type I error, while Models M_1, M_2 and M_3 are used to obtain the empirical power. For each model with each error distribution, the data were randomly generated and the GSW, BOS, and WBOS methods were applied to the data. Specifically, a nonlinear model of form M_0 was fitted with the nonlinear least squares method. The initial estimates of parameters b_1, b_2, b_3 are all set to be 0.001. Upon convergence, the residuals from the fit were obtained, which were then used in the calculation of the test statistics for all tests. To obtain p-value for GSW, we used asymptotic distribution in 4.2.4 for data generated under the null model (model

M_0) and that in 4.2.6 for data generated under the alternatives (models $M_1 - M_3$). The procedure with both data generation and application of the three tests was repeated 2,000 times.

For the generated data, BOS and WBOS tests were applied as the following:

- Sort the data according to the predictor values x_i, $i = 1, 2, ..., N$.
- Calculate nls (nonlinear least squares) fit, obtain residuals $\{e_1, e_2, ..., e_N\}$ and fitted values $\{\hat{y}_1, \hat{y}_2, ..., \hat{y}_N\}$.
- Obtain (Wild) bootstrap samples (page 11 of Kutchibhatla and Hart (1996) and page 866 of Chen et al. (2001)) from residuals $\{e_1^*, e_2^*, ..., e_N^*\}$ and calculate bootstrap observations $y_i^* = \hat{y}_i + e_i^*$,$i = 1, 2, ..., N$.
- Calculate nls fit using y_i^* and x_i ,$i = 1, 2, ..., N$, and get $\hat{e_i^*}$.
- Obtain Fourier coefficients of BOS or WBOS test using $\{\hat{e_1^*}, ..., \hat{e_N^*}\}$ and $x_i = (i - 0.5)/N$.

The order selection test statistics were then calculated for bootstrap sample if the nonlinear least squares fit can be obtained. Otherwise this bootstrap sample was discarded and a new bootstrap sample was obtained to proceed. The resampling and calculation of the bootstrap test statistic were repeated until 2,000 bootstrap test statistic values were obtained. The bootstrap p-value is the proportion of the bootstrap test statistics greater than the observed test statistic value. The resulted rejection rates are reported in Table 4.1 based on nominal levels $\alpha = 0.01$ and 0.05 for all tests. The last two columns (B BOS and B WBOS) in Tables 4.1 and 4.2 represent the average number of bootstrap resamples needed to obtain 2,000 bootstrap test statistic values for BOS and WBOS tests, respectively.

The first 4 rows of Table 4.1 show the type I error estimates for all tests with the four types of error distributions. For all tests, the type I error estimates are close to the nominal levels in all cases.

Table 4.1: *Rejection rate under H_0 and high frequency alternatives with sample size $N = 50$*

Model	Error	level 0.01			level 0.05			B BOS	B WBOS
		GSW	BOS	WBOS	GSW	BOS	WBOS		
M_0	unif	0.005	0.011	0.013	0.021	0.046	0.050	2679	2649
	normal	0.008	0.005	0.007	0.016	0.037	0.049	2702	2679
	t	0.004	0.006	0.012	0.010	0.040	0.051	2709	2699
	het5	0.008	0.005	0.016	0.023	0.041	0.067	2887	3074
M_1	unif	0.969	0.133	0.190	0.992	0.816	0.808	2442	2402
	normal	0.960	0.144	0.196	0.990	0.830	0.830	2445	2398
	t	0.957	0.164	0.237	0.988	0.878	0.869	2462	2408
	het5	0.980	0.174	0.340	0.996	0.903	0.949	2436	2363
M_2	unif	0.954	0.160	0.213	0.984	0.865	0.849	2331	2248
	normal	0.962	0.198	0.248	0.992	0.881	0.875	2340	2254
	t	0.952	0.224	0.284	0.979	0.905	0.898	2348	2260
	het5	0.981	0.276	0.377	0.997	0.937	0.959	2337	2226
M_3	unif	0.834	0.132	0.044	0.934	0.663	0.357	2262	2239
	normal	0.848	0.159	0.064	0.938	0.687	0.398	2253	2236
	t	0.874	0.177	0.052	0.946	0.726	0.394	2235	2217
	het5	0.913	0.180	0.062	0.974	0.725	0.465	2226	2206

Power comparison for the different combinations of Models $M_1 - M_3$ and the four error distributions (Unif, Normal, T, Heter) are shown in the last 12 rows of Table 4.1. It can be seen that the power of our test GSW is much higher than the other two tests in all cases. For Models M_1 and M_2 and for all different types of error distributions, the power of WBOS test is slightly higher than BOS test when $\alpha = 0.01$ and these powers become close to each other when $\alpha = 0.05$. On the contrary, BOS has significantly higher power than WBOS test for data generated under Model M_3 regardless of the error distribution and level of significance.

Models M_1, M_2, M_3 in the previous simulation represent high frequency alternatives. To investigate the power performance of the three tests (GSW, BOS, WBOS) in the case of low frequency alternatives, data were generated from the following model:

$$Y_i = \frac{e^{-b_1 X_i}}{b_2 + b_3 X_i} + \cos(2\pi X_i) + \epsilon_i, \tag{4.3.1}$$

with the four different error distributions and under the same setup used in the previous simulation. Empirical power for all tests are given in Table 4.2. Table 4.2 shows that there is not much differences between the power of the three tests in all the cases of error distribution and level of significance.

Table 4.2: *Rejection rate under low frequency alternatives in (4.3.1) with sample size* $N = 50$

Error	level 0.01			level 0.05			B BOS	B WBOS
	GSW	BOS	WBOS	GSW	BOS	WBOS		
unif	0.997	0.999	0.999	0.999	1.000	1.000	2239	2101
normal	0.994	0.999	0.999	0.999	1.000	1.000	2223	2095
t	0.982	0.996	0.997	0.989	0.999	0.999	2210	2085
het5	0.998	1.000	1.000	1.000	1.000	1.000	2175	2043

It is worth to mention that BOS and WBOS tests require a lot more bootstrap samples than the 2,000 specified because some of the bootstrap samples fail to produce successful nonlinear least squares fit (see the last two columns in Tables 4.1 and 4.2).

To have a look at the power performance of these tests with various sample sizes, we generated data from the following model

$$Y_i = \frac{e^{-b_1 X_i}}{b_2 + b_3 X_i} + e^{-2X_i} \cos(10\pi X_i) + \epsilon_i, \tag{4.3.2}$$

with $(b_1, b_2, b_3) = (-5, 20, 0.6)$ and $\epsilon_i \sim$ Uniform(-0.8,0.8) for N= 30, 50, 75, 85, 100, 115, 125, 130, 150, 175, and 200. The resulted empirical power curves of the three tests based on $\alpha = 0.01$ are shown in Figure 4.4. It can be seen that the power of our test GSW is consistently higher than the power of the other two tests (BOS and WBOS). The power of our test clearly converges to 1 faster than the BOS and WBOS as the sample size increases.

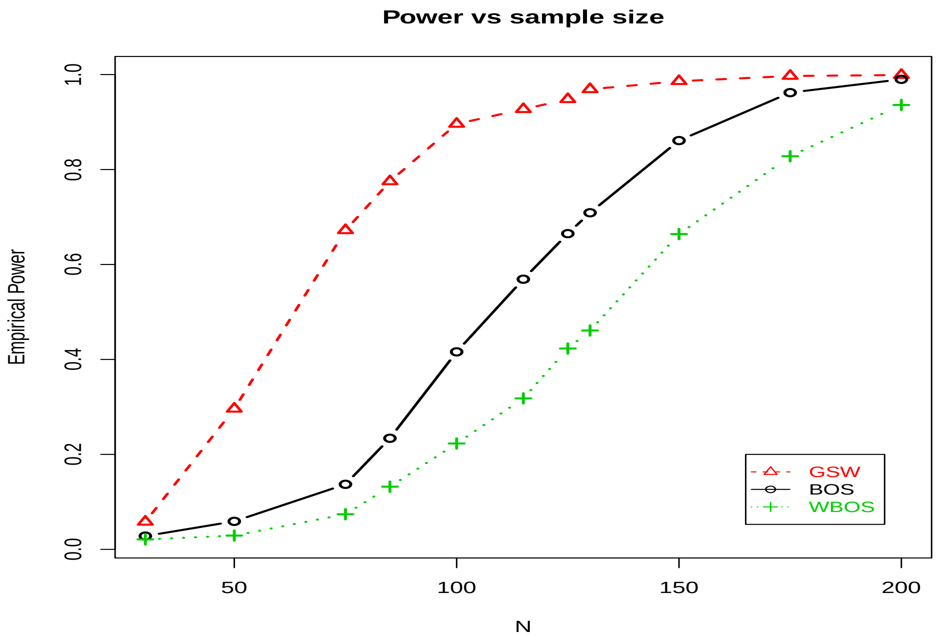


Figure 4.4: *Power plot for data generated under the model in (4.3.2) with different sample sizes*

4.3.2 Application to ultrasonic reference block data

In this section, we illustrate an application of our proposed test to the ultrasonic reference block data, which was given in Figure 4.1. These data were provided by Dan Chwirut who is a scientist at the National Institute of Standards and Technology (NIST). The data is publicly available at the Engineering Statistics Handbook. As it was mentioned in the introduction, the scientists suggested using square root transformation of the response variable to deal with the violation of non constant variance. In particular, they suggested to fit the data with the following model

$$y^{1/2} = \frac{exp(-b_1 x)}{b_2 + b_3 x} + \epsilon \tag{4.3.3}$$

The residual versus covariate and residual versus fitted value plots in Figure 4.3 still suggest some nonrandom pattern exists.

We applied our proposed test GSW to assess the lack of fit of the suggested nonlinear regression model in (4.3.3) for the ultrasonic reference block data. The order selection test of Kuchibhatla and Hart (1996) in (4.1.3) is also used for testing the adequacy of the suggested model in (4.3.3). Bootstrap and wild bootstrap are employed to obtain the critical value of the order selection test. The p-value of our proposed test GSW is 0. For the bootstrap order selection BOS and wild bootstrap order selection WBOS tests, the p-values based on 10000 resamples are 0.0214 and 0.0271, respectively. The p-values of GSW, BOS, and WBOS indicate that our proposed test GSW has more power of detecting lack of fit in such cases with the presence of heteroscedastic errors.

Chapter 5

Selection of the number of nearest neighbors

The number of nearest neighbors k in the proposed test statistics specifies the number of values augmented in each cell. In this dissertation, our theory requires that k takes a finite small odd integer. In simulations, we have found that the type I error remains close to the nominal level for different small k values and stays stable for a broad range of sample sizes and error distributions (see Figs. 1.1, 3.3, 3.4 and 3.5). Under the alternative hypothesis, different k may lead to different power for our test statistics. This chapter discusses how to select the parameter k.

Under the alternative hypothesis, our k-nearest neighbor augmentation is parallel to regression using local constant based on k-nearest neighbors. For continuous response variable, Hardle et al. (1988) suggest the Least Squares Cross-Validation (LSCV) method for smoothing parameter (bandwidth) selection in kernel regression estimation. Chen et al. (2001) recommend using the one-sided cross-validation procedure of Hart and Yi (1998) to select smoothing parameter (bandwidth) for hypothesis testing. The number of nearest neighbors k in our setting has a similar role as the smoothing parameter in kernel regression.

For categorical response variable, Holmes and Adams (2003) proposed an approach to select the parameter k in k-nearest neighbor (KNN) classification algorithm using likelihood-based inference. Choosing k in this method can be considered as a generalized linear model variable- selection problem. In particular, for multinomial data $(y_i, \mathbf{x}_i)$, $i = 1, \ldots n$, where $y_i \in \{C_0, ..., C_Q\}$ denotes the class label of the ith observation and $\mathbf{x}_i$ is a vector of p predictor variables, they considered the probability model

$$pr(y_i = C_i | y_{[-i]}, \mathbf{x}_i, k) = \frac{\exp(z_i^{(k,j)}\theta)}{\sum_{v=0}^{Q} \exp(z_i^{(k,v)}\theta)},$$

where $y_{[-i]} = \{y_1, \ldots, y_{i-1}, y_{i+1}, \ldots, y_n\}$ denotes the data with the ith observation deleted, θ is a single regression parameter and $z_i^{(k,v)}$ is the difference between the proportion of observations in class C_v and that in class C_0 within the k-nearest neighbors of $\mathbf{x}_i$, i.e.,

$$z_i^{(k,v)} = \frac{1}{k} \sum_{j \overset{k}{\sim} i} \{I(y_j = C_v) - I(y_j = C_0)\},$$

where the notation $\sum_{j \overset{k}{\sim} i}$ denotes that the summation is over the k-nearest neighbors of $\mathbf{x}_i$ in the set $\{\mathbf{x}_1, ..., \mathbf{x}_{i-1}, \mathbf{x}_{i+1}, ..., \mathbf{x}_n\}$ and the neighbors are defined based on Euclidean distance. The prediction for a new point $y_{n+1}|\mathbf{x}_{n+1}$ is given by the most common class in the k-nearest neighbors of $\mathbf{x}_{n+1}$. Afterward, the value that maximizes the profile pseudolikelihood is chosen to estimate the parameter k. However, this method is only valid when the response variable is a categorical variable and the nearest neighbor is defined using Euclidian distance.

In our case, the response variable could be continuous or categorical and our nearest neighbors are defined through ranks. So we do not recommend to use our test statistics with an estimate of k obtained with aforementioned procedures. We consider an alternative method to estimate k which uses ranks to define nearest neighbors and can be applied in both categorical and continuous response cases. Here we adopt the idea of the Least Squares Cross-Validation (LSCV) procedure of Hardle et al. (1988) to select the parameter k. Different from Hardle et al. (1988) where the regression function is estimated using kernel estimation, we consider k-nearest neighbor estimates with neighbors defined through the ranks of the predictor variable. In the case of categorical response variable, suppose we have Q classes, then we re-code the response variable to have integer values from 1 to Q. To estimate the class for the response variable, we use the majority vote (the most common value) from the k-nearest neighbors. For tied situation where there are multiple classes achieving the same highest frequency, one of them is assigned randomly to be the estimated response. In the case of continuous response variable, the regression function is estimated by the average of the k-nearest neighbors.

In leave-one-out procedure, for each $c \in \{1, \ldots, N\}$, we eliminate (X_c, Y_c) and use the rest of the observations to estimate the regression function which then is used to predict the response value Y at X_c. Here are the steps we use:

1. Find the observation in $\mathbf{X}_{[-c]} = \{\text{all } \ X_i, \text{where } \ i = 1, ..., N \ \text{ and } \ i \neq c\}$ such that the absolute difference between this observation and X_c is minimized. Denote

$$J(c) = \{\arg\min_j |X_j - X_c|, \text{ where } \ j = 1, ..., N \ \text{ and } \ j \neq c\}.$$

   Then $X_{J(c)}$ is the closest to X_c.

2. Find the k-nearest neighbors of $X_{J(c)}$ in terms of rank. We use the corresponding Y_i values such that

$$N|\widehat{F}(X_{J(c)}) - \widehat{F}(X_i)| \leq \frac{k-1}{2} \text{ for } i \neq c,$$

   to obtain the leave-one-out estimate of the regression function at X_c. That is

$$\hat{m}_{k,-c}(X_c) = \begin{cases} k^{-1} \sum_{i=1, i \neq c}^{N} Y_i \ \ I\left(N|\widehat{F}(X_{J(c)}) - \widehat{F}(X_i)| \leq \frac{k-1}{2}\right), & \text{continuous case} \\ \text{Mode of } \{Y_i : \text{all } i \neq c \text{ such that } N|\widehat{F}(X_{J(c)}) - \widehat{F}(X_i)| \leq \frac{k-1}{2}\}, & \text{categorical case,} \end{cases}$$

   where the Mode is defined as the most frequently observed value in a set of numbers. In case where the most frequently observed values are not unique, one of them is randomly selected.

3. Repeat steps 1 and 2 for $c = 1, ..., N$ to obtain all leave-one-out estimates.

Then define the leave-one-out Least Squares Cross-Validation error as

$$LSCV(k) = \frac{1}{N}\sum_{c=1}^{N}(\hat{m}_{k,-c}(X_c) - Y_c)^2$$

Finally, the number of nearest neighbors is estimated by

$$\hat{k} = \arg\min_{k\in\kappa} \quad LSCV(k), \tag{5.0.1}$$

where the set κ consists of small odd integers.

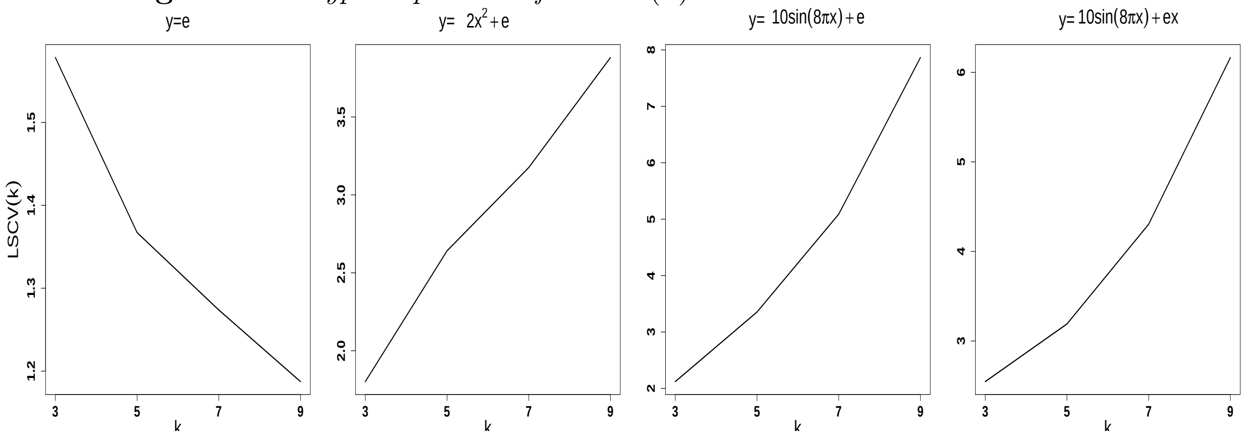
When the response variable is categorical, the estimate of k from this algorithm depends on how well the covariate values from different classes are separated and how many observations are in each class. For large class sizes, it is very possible that the resulting estimate is much greater than 10 if we leave κ unconstrained. However, our theory requires k to be a finite, positive, odd integer.

In the continuous case with k-nearest neighbor estimation, the average of a big proportion of Y values is used to approximate the response variable if a large k value is utilized. As a consequence, bigger k tends to give larger least squares error when the regression function is under the alternative hypothesis. This is especially true when the regression function has substantial curvature such as in high frequency alternatives. On the other hand, larger k tends to give smaller least squares error when the data were generated under the constant regression null hypothesis.

Figure 5.1 shows the typical pattern of $LSCV(k)$ as a function of k for $k = 3, 5, 7, 9$ when the response variable was generated as (1) $Y_i = e_i$; (2) $Y_i = 2X_i^2 + e_i$; (3) $Y_i = 10\sin(8\pi X_i) + e_i$; (4) $Y_i = 10\sin(8\pi X_i) + e_i X_i$; where e_i and X_i are i.i.d $\sim N(0, 1)$.

Regardless of categorical or continuous responses, the smallest value for k is 3 (note: $k = 1$ corresponds to the case of no data augmentation). In order to keep the least squares error minimized under the alternative hypothesis and reasonable under the null hypothesis, we recommend to let κ contain a few small integer values. For example, $\kappa = \{3, 5\}$, which is a safe choice for both moderate and large sample sizes. This choice of κ was used in the

Figure 5.1: *Typical pattern of $LSCV(k)$ versus k in continuous data.*



numerical studies of Chapter 3. The estimated $\hat{k}$ based on (5.0.1) is recommended to be used to perform the lack-of-fit tests given in Chapters 3 and 4.

Chapter 6

Summary and Future Research

6.1 Summary

In this dissertation, we studied nonparametric lack-of-fit tests in presence of heteroscedastic variances. The response variable can be discrete or continuous with unknown distribution, while the covariate is assumed to be a continuous variable. Regardless of the response variable being discrete or continuous, we formulate the hypothesis of constant regression or nonlinear regression in terms of the conditional mean of the response variable given the covariate. Assuming no replications were observed, our lack-of-fit tests first perform a data augmentation using a small number of k-nearest neighbors defined through the ranks of the predictor variable. The augmentation was done on the observed data for the constant regression null hypothesis and on the residuals from the fitted model under the null hypothesis of nonlinear regression. Then the test statistics were constructed by comparing two quadratic forms, both of which estimate a common quantity but one under the null hypothesis and the other under the local alternatives. We derived the asymptotic distribution of the test statistics under both the null and local alternative hypotheses. The theory for the test of constant regression and that for nonlinear regression were presented separately. The parametric standardizing rate is achieved for the asymptotic distribution of the proposed test statistics. As a result, the proposed tests have faster convergence rate than most of nonparametric methods. This is a consequence of fixed number of nearest neighbors augmentation. Numerical comparisons show that the proposed tests have good power to detect both low and high frequency alternatives even for moderate sample size. The tests are especially more powerful than some well known competing test procedures when data were generated under high frequency alternatives. Comparing to bootstrap or smoothing based methods, a clear advantage of the proposed tests is that the test statistics and their asymptotic distributions are easy and fast to calculate.

For the test of constant regression null hypothesis, the asymptotic distribution of the same test statistic was also given in Wang et al. (2008) but with a biased asymptotic variance. We derived the correct form of the asymptotic distribution of the test statistic under both the null hypothesis and local alternatives. The test of nonlinear regression was not as widely studied as the constant or linear regression case. The proposed test statistic in the test of nonlinear regression case is unique and is a completely new addition to the lack-of-fit literature. Since the proposed lack-of-fit tests can be applied to regression models with a discrete or continuous response variable without distributional assumptions, these tests are widely applicable to many practical data.

In addition to the inference for fixed number of nearest neighbor augmentation, this dissertation also provided a method to select the number of nearest neighbors based on the idea of the Least Squares Cross-Validation (LSCV) procedure of Hardle and Mammen (1993). We generalized the LSCV such that it works with our augmentation based on ranks of the predictor variable and can accommodate the case of discrete response variable.

Putting everything together, the results in this dissertation offer a useful tool for lack-of-fit test.

6.2 Future research

The proposed lack-of-fit tests can be simply applied to testing the equality of two regression curves when response values from both curves are available at every design point.

In particular, suppose we observe $(Y_1, Z_1), ..., (Y_N, Z_N)$ at the same design points where $Y_i = m_1(x_i) + \varepsilon_{1i}$ and $Z_i = m_2(x_i) + \varepsilon_{2i}$. To test the null hypothesis $H_0 : m_1(x_i) = m_2(x_i)$, we can define $Y_i^* = Y_i - Z_i$,$i = 1, ..., N$, then our lack-of-fit test of constant regression might be applied to the data $(Y_1^*, ..., Y_N^*)$. For future research, it might be of interest to extend our test to cover the general case when the two responses (Y_i, Z_i) are not available at every design point. This could be handled by combining our methodology in this dissertation with that in Young and Bowman (1995).

Our tests in this dissertation were developed for regression models with only one predictor variable. Extending the proposed tests to deal with the presence of more than one predictor is another issue of interest. We might use Euclidian distance or any other approach to obtain k-nearest neighbor augmentation to construct a test statistic similar to that we proposed in this dissertation.

Additionally, the test procedure developed in this dissertation can be generalized to test the fit of additive models of the form $\mathbf{Y} = \sum_{i=1}^{p} m_i(\mathbf{x}_i) + \varepsilon$ where $m_1, ..., m_p$ are unknown functions,$\mathbf{x}_1, ..., \mathbf{x}_p$ are predictor variables,$\mathbf{Y}$ is the response variable, and ε is the error term. In particular, it would be of interest to test the null hypotheses $H_0 : m_i(\mathbf{x}_i) = 0$, where $i = 1, ..., p$.

Bibliography

Bierens, H. J. (1990). A consistent conditional moment test for functional form. *Econometrica*, 58:1443–1458.

Bishop, Y. M., Fienberg, S. E., and Holland, P. W. (2007). Discrete multivariate analysis. *Springer, New York.*

Brown, C. (1982). On a goodness-of-fit test for the logistic model based on score statistics. *Communications in Statistics - Theory and Methods*, 11:1087–1105.

Chen, C.-F., Hart, J. D., and Wang, S. (2001). Bootstrapping the order selection test. *Journal of Nonparametric Statistics*, 13(6):851–882.

Christensen, R. (1989). Lack-of-fit tests based on near or exact replicates. *The Annals of Statistics*, 17(2):673–683.

Christensen, R. (1991). Small-sample characterizations of near replicate lack-of-fit tests. *Journal of the American Statistical Association*, 86(415):pp. 752–756.

de Jong, P. (1987). A central limit theorem for generalized quadratic forms. *Probability Theory*, 75:261–277.

Eubank, R. L. and Hart, J. D. (1992). Testing goodness-of-fit in regression via order selection criteria. *The Annals of Statistics*, 20:1412–1425.

Eubank, R. L. and Spiegelman, C. H. (1990). Testing the goodness of fit of a linear model via nonparametric regression techniques. *Journal of the American Statistical Association*, 85:387–392.

Fisher, R. A. (1922). The goodness of fit of regression formulae and the distribution of regression coeffcients. *Journal of the Royal Statistical Society*, 85:597–612.

Guerre, E. and Lavergne, P. (2005). Data-driven rate-optimal specification testing in regression models. *The Annals of Statistics*, 33:840–870.

Hardle, W., Hall, P., and Marron, J. S. (1988). How far are automatically chosen regression smoothing parameters from their optimum? *Journal of the American Statistical Association*, 83(401):86–95.

Hardle, W. and Mammen, E. (1993). Comparing nonparametric versus parametric regression fits. *The Annals of Statistics*, 21:1926–1947.

Hart, J. (1997). Nonparametric smoothing and lack-of-fit test. *Springer-Verlag, New York.*

Hart, J. (2008). Smoothing-inspired lack-of-fit tests based on ranks. *Beyond Parametrics in Interdisciplinary Research: Festschrift in Honor of Professor Pranab K.Sen*, 1:138–155.

Hart, J. (2009). Frequentist-Bayes lack-of-fit tests based on Laplace approximations. *Journal of Statistical Theory and Practice*, 3:681–704.

Hart, J. D. and Yi, S. (1998). One-sided cross-validation. *Journal of the American Statistical Association*, 93(442):620–631.

Hausman, J. A. (1978). Specification tests in econometrics. *Econometrica*, 46:1251–1271.

Holmes, C. C. and Adams, N. M. (2003). Likelihood inference in nearest-neighbour classification models. *Biometrika*, 90:99–112.

Horowitz, J. Z. and Spokoiny, V. G. (2001). An adaptive, rate-optimal test of a parametric mean-regression model against a nonparametric alternative. *Econometrica*, 69:599–631.

Hosmer, D. W. and Lemesbow, S. (1980). Goodness of fit tests for the multiple logistic regression model. *Communications in Statistics - Theory and Methods*, 9(10):1043–1069.

Jennrich, R. I. (1969). Asymptotic properties of non-linear least squares estimators. *The Annals of Mathematical Statistics*, 40(2):pp. 633–643.

Kuchibhatla, M. and Hart, J. D. (1996). Smoothing-based lack-of-fit tests: variations on a theme. *Journal of Nonparametric Statistics*, 7(1):1–22.

Kutchibhatla, M. and Hart, J. (1996). Smoothing-based lack-of-fit tests: variations on a theme. *Nonparametr. Statist.*, 7:1–22.

Lee, B. J. (1988). A nonparametric model specification test using a kernel regression method. ph.d. dissertation (university of wisconsin, madison, wi).

Li, C.-S. (1999). Testing lack of fit of regression models under heteroscedasticity. *The Canadian Journal of Statistics / La Revue Canadienne de Statistique*, 27(3):pp. 485–496.

Li, C.-S. (2003). A lack-of-fit test for heteroscedastic regression models via cosine-series smoothers. *Australian & New Zealand Journal of Statistics*, 45(4):477–489.

Li, C.-S. (2005). Using local linear kernel smoothers to test the lack of fit of nonlinear regression models. *Statistical Methodology*, 2(4):267 – 284.

Lim, C. (2009). Statistical theory and robust methodology for nonlinear models with application to toxicology. ph.d. dissertation (university of north carolina at chapel hill, chapel hill, nc).

Lim, C., Sen, P., and Peddada, S. (2010). Statistical inference in nonlinear regression under heteroscedasticity. *Sankhya B*, 72:202218.

McCullagh, P. (1986). The conditional distribution of goodness-of-fit statistics for discrete data. *Journal of the American Statistical Association*, 81:104–107.

Miller, F. R. and Neill, J. W. (2008). General lack of fit tests based on families of groupings. *Journal of Statistical Planning and Inference*, 138(8):2433 – 2449.

Miller, F. R., Neill, J. W., and Sherfey, B. W. (1998). Maximin clusters for near-replicate regression lack of fit tests. *The Annals of Statistics*, 26(4):pp. 1411–1433.

Miller, F. R., Neill, J. W., and Sherfey, B. W. (1999). Implementation of a maximin power clustering criterion to select near replicates for regression lack-of-fit tests. *Journal of the American Statistical Association*, 94(446):pp. 610–620.

Neill, J. and Miller, F. (2003). Limit experiments, lack of fit tests and fuzzy clusterings, technical report ii-03-1, department of statistics, kansas state univer- sity.

Neill, J. W. (1988). Testing for lack of fit in nonlinear regression. *The Annals of Statistics*, 16(2):pp. 733–740.

Neill, J. W. and Johnson, D. E. (1984). Testing for lack of fit in regression - a review. *Communications in Statistics - Theory and Methods*, 13(4):485–511.

Neill, J. W. and Johnson, D. E. (1985). Testing linear regression function adequacy without replication. *The Annals of Statistics*, 13(4):pp. 1482–1489.

Neill, J. W. and Johnson, D. E. (1989). A comparison of some lack of fit tests based on near replicates. *Communications in Statistics - Theory and Methods*, 18(10):3533–3570.

Newey, W. K. (1985a). Generalized method of moments specification testing. *Journal of Econometrics*, 29:229–256.

Newey, W. K. (1985b). Maximum likelihood specification testing and conditional moment tests. *Econometrica*, 53:1047–1070.

Pyke, R. (1965). Spacings (with discusssion). *Journal of the Royal Statistical Society. Series B (Methodological)*, 27(3):395–449.

Ruud, P. A. (1984). Tests of specification in econometrics. *Econometrics Reviews*, 3:211–242.

Singh, D., Febbo, P. G., Ross, K., Jackson, D. G., Manola, J., Ladd, C., Tamayo, P., Renshaw, A. A., D'Amico, A. V., Richie, J. P., Lander, E. S., Loda, M., Kantoff, P. W., Golub, T. R., and Sellers, W. R. (2002). Gene expression correlates of clinical prostate cancer behavior. *Cancer Cell*, 1(2):203 – 209.

Song, W. and Du, J. (2011). A note on testing the regression functions via nonparametric smoothing. *The Canadian Journal of Statistics*, 39:108–125.

Su, J. and Wei, L. J. (1991). A lack-of-fit test for the mean function in a generalized linear model. *Journal of the American Statistical Association*, 86:420–426.

Tauchen, G. (1985). Diagnostic testing and evaluation of maximum likelihod modeols. *Journal of Econometrics*, 30:415–443.

Wang, H. and Akritas, M. (2011). Asymptotically distribution free tests in heteroscedastic unbalanced high dimensional anova. *Statistica Sinica*, 21:1341–1377.

Wang, H., Tolos, S., and Wang, S. (2010). A distribution free test to detect general dependence between a response variable and a covariate in the presence of heteroscedastic treatment effects. *The Canadian Journal of Statistics*, 38(3):408–433.

Wang, L. and Akritas, M. (2006). Testing for covariate effects in the fully nonparametric analysis of covariance model. *Journal of the American Statistical Association*, 101:722–736.

Wang, L., Akritas, M. G., and Van Keilegom, I. (2008). An ANOVA-type nonparametric diagnostic test for heteroscedastic regression models. *Journal of Nonparametric Statistics*, 20(5):365–382.

White, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrics*, 50:125.

White, H. (1987). Specification testing in dynamic models. In Advances in Econometrics. Fifth World Conress. Volume I, edited by Truman F. Bewley, Cambridge: Cambridge University Press.

Wu, C.-F. (1981). Asymptotic theory of nonlinear least squares estimation. *The Annals of Statistics*, 9(3):pp. 501–513.

Yatchew, A. J. (1992). Nonparametric regression tests based on least square. *Econometric Theory*, 8:435–451.

Young, S. G. and Bowman, A. W. (1995). Non-parametric analysis of covariance. *Biometrics*, 51(3):pp. 920–931.

Zheng, J. X. (1996). A consistent test of functional form via nonparametric estimation techniques. *Journal of Econometrics*, 75:263–289.