Testing for Covariate Effects in the Fully Nonparametric Analysis of Covariance Model

Lan WANG and Michael G. AKRITAS

Traditional inference questions in the analysis of covariance mainly focus on comparing different factor levels by adjusting for the continuous covariates, which are believed to also exert a significant effect on the outcome variable. Testing hypotheses about the covariate effects, although of substantial interest in many applications, has received relatively limited study in the semiparametric/nonparametric setting. In the context of the fully nonparametric analysis of covariance model of Akritas et al., we propose methods to test for covariate main effects and covariate–factor interaction effects. The idea underlying the proposed procedures is that covariates can be thought of as factors with many levels. The test statistics are closely related to some recent developments in the asymptotic theory for analysis of variance when the number of factor levels is large. The limiting normal distributions are established under the null hypotheses and local alternatives by asymptotically approximating a new class of quadratic forms. The test statistics bear similar forms to the classical F-test statistics and thus are convenient for computation. We demonstrate the methods and their properties on simulated and real data.

KEY WORDS: Covariate effects; Fully nonparametric model; Heteroscedasticity; Interaction effects; Nearest-neighborhood windows; Nonparametric hypotheses.

1. INTRODUCTION

In many scientific fields, researchers frequently face the problem of analyzing a response variable at the presence of both factors (categorical explanatory variables) and covariates (continuous explanatory variables). Traditional inference focuses mainly on covariate-adjusted factor effects. Testing for covariates effects, although of substantial interest in itself, has received relatively limited attention, especially in a semiparametric/nonparametric setting. In this article we discuss formulating and testing appropriate hypotheses for covariate effects in a completely nonparametric fashion.

The classical analysis of covariance model imposes a set of stringent assumptions including linearity, parallelism, homoscedasticity, and normality. In this framework, testing the significance of covariate effects is equivalent to checking whether the regression coefficients are zero, and testing for the presence of factor–covariate interaction effects amounts to investigating whether the regression lines have the same slope. Due to its simplicity, the classical approach remains the most popular tool in practice.

Early work to relax the restrictive assumptions of the classical ANCOVA model involves the use of rank-based methods. This literature is very extensive, but it is fair to state that although successful in some respects, the rank-based methods are not totally satisfactory. For example, the test of Quade (1967) can be applied only to test for covariate-adjusted treatment effects; it avoids assuming normality and linearity but requires that the covariate have the same distribution across all treatment levels. The approach of McKean and Schrader (1980) does not need to assume normality but is based on parametric modeling of the covariate effect. Another line of research yields interesting alternative models, including the partial linear model (Speckman 1988; Hastie and Tibshirani 1990), the varying-coefficient model (Hastie and Tibshirani 1993), and the model of Sen (1996). But these approaches still require certain parametric assumptions and have not considered the type of hypotheses that we discuss in this article.

Closer to the spirit of the present article is recent work on testing the equality of two or more mean regression functions (see Härdle and Marron 1990; Hall and Hart 1990; King, Hart, and Wehrly 1991; Delgado 1993; Kulasekera 1995; Hall, Huber, and Speckman 1997; Dette and Munk 1998; Dette and Neumeyer 2001; Koul and Schick 2003, among others). The null hypothesis considered by these authors amounts to the hypothesis of no factor-simple effects (i.e., no factor main effects and no factor–covariate interaction effects). These tests are developed under the assumption of a location-scale model; as our simulation (Sec. 4) suggests, this is not a benign assumption. Moreover, it is not immediately clear how the hypotheses of no covariate effects or no factor–covariate interaction effects can be directly formulated and tested with their approaches. One exception is the work of Young and Bowman (1995), who considered testing the parallelism of several mean regression functions. They used a partial linear model, which provides one way to formulate the hypothesis of no factor–covariate interaction effects. They did not consider any hypothesis related to main covariate effects; for their method to work, the random errors need to have homoscedastic normal distributions. Finally, it also is not clear how the aforementioned methods can readily accommodate more than one factor.

To overcome the aforementioned difficulties, this article extends the state of art of nonparametric analysis of covariance in a number of ways. The new methodology applies to testing all common hypotheses, allows for the presence of several factors, includes both continuous and discrete ordinal response variable, permits heteroscedastic errors, and does not require the covariate to be equally spaced or to have the same distribution at different treatment level combinations. The main vehicle for achieving this extension is the nonparametric model of Akritas, Arnold, and Du (2000), which generalizes beyond the common location-scale model.

To introduce the fully nonparametric model, let (X_{ij}, Y_{ij}), $i = 1, \dots, k$, $j = 1, \dots, n_i$, denote the observable variables. More

Lan Wang is Assistant Professor, School of Statistics, University of Minnesota, Minneapolis, MN 55455 (E-mail: *lan@stat.umn.edu*). Michael G. Akritas is Professor, Department of Statistics, Pennsylvania State University, University Park, PA 16802 (E-mail: *mga@stat.psu.edu*). The authors thank the associate editor, an anonymous referee, and the editor for comments that significantly improved the article. This work was supported in part by National Science Foundation grant SES-0318200.

Journal of the American Statistical Association
June 2006, Vol. 101, No. 474, Theory and Methods
DOI 10.1198/016214505000001276

specifically, i enumerates the cells (or factor-level combinations) and j denotes the observations within each cell. This notation is typical for one-way analysis of covariance and also can be used for higher-way designs, with the understanding that the single index i enumerates all factor-level combinations; X_{ij} may also represent a vector of covariates if needed. More discussion on higher-way designs and multiple covariates is given in Section 6. The fully nonparametric model of Akritas et al. (2000) specifies only that

$$Y_{ij}|X_{ij} = x \sim F_{ix}; \tag{1}$$

that is, given $X_{ij} = x$, the distribution of Y_{ij} depends on i and x. Thus (1) does not specify how this conditional distribution changes with the cell or covariate value. In this sense it is completely nonparametric (also nonlinear and nonadditive). Choose a distribution function $G(x)$ and let

$$\overline{F}_{i\cdot}^{G}(y) = \int F_{ix}(y)\, dG(x) \qquad \text{and} \qquad \overline{F}_{\cdot x}(y) = \frac{1}{k}\sum_{i=1}^{k} F_{ix}(y).$$

If the X_{ij}'s are random, then G can be taken as their overall distribution function. Therefore, if the covariate has the same distribution in all groups, then $\overline{F}_{i\cdot}^{G}(y)$ is the marginal distribution function of Y_{ij}. The hypotheses of interest in model (1) for the one-way design are as follows:

$$\text{No main group effect, or } \overline{F}_{i\cdot}^{G} \text{ does not depend on } i; \tag{2}$$

$$\text{No main covariate effect, or } \overline{F}_{\cdot x}(y) \text{ does not depend on } x; \tag{3}$$

$$\text{No group-covariate interaction, or } F_{ix}(y) = \overline{F}_{i\cdot}^{G}(y) + K_x(y); \tag{4}$$

$$\text{No simple group effect, or } F_{ix}(y) \text{ does not depend on } i; \tag{5}$$

and

$$\text{No simple covariate effect, or } F_{ix}(y) \text{ does not depend on } x, \tag{6}$$

where in (4) $K_x(y)$ is some function independent of i. Effects for these hypotheses are defined in terms of an ANOVA-type decomposition,

$$F_{ix}(y) = M(y) + A_i(y) + D_x(y) + C_{ix}(y), \tag{7}$$

where, under the side conditions $\sum_{i=1}^{k} A_i(y) = 0$ for all y, $\int D_x(y)\, dG(x) = 0$ for all y, $\sum_{i=1}^{k} C_{ix}(y) = 0$ for all x and y, and $\int C_{ix}(y)\, dG(x) = 0$ for all i and y, the effects are given by $M(y) = k^{-1}\sum_{i=1}^{k} \overline{F}_{i\cdot}^{G}(y)$, $A_i(y) = \overline{F}_{i\cdot}^{G}(y) - M(y)$, $D_x(y) = \overline{F}_{\cdot x}(y) - M(y)$, and $C_{ix}(y) = F_{ix}(y) - M(y) - A_i(y) - D_x(y)$. Borrowing terminology from the classical ANCOVA, A_i is the covariate-adjusted main effect of cell i, D_x is the main effect of the covariate value x, and C_{ix} is the interaction effect between cell i and covariate value x. The nonparametric hypotheses simply state that the corresponding nonparametric effects are zero. In particular, the hypotheses (3) and (4) can be restated as

$$H_0(D): D_x(y) = 0 \quad \text{for all } x \text{ and all } y$$

and

$$H_0(C): C_{ix}(y) = 0 \quad \text{for all } i,\ \text{all } x,\ \text{and all } y.$$

An important property of the nonparametric hypotheses is that they are invariant to monotone transformations of the response. This property is particularly desirable for ordinal response variables because they can be coded using different scales by different researchers.

The use of mid-rank procedures for testing (2) was considered in Akritas et al. (2000). In this article we develop procedures for testing the hypotheses (3) and (4). Procedures for testing (5) and (6) can be constructed similarly. Note that (4) and (5) generalize the hypotheses of parallelism and equality of regression lines, which were considered by Young and Bowman (1995). Development of the present tests requires new types of statistics and different techniques from those for testing (2). The basic idea for constructing the test procedures is to treat the continuous covariate as a factor with many levels and borrow suitable test statistics from the ANOVA. The methods are therefore motivated by recent advances in the asymptotic theory for heteroscedastic ANOVA with a large number of factor levels (Akritas and Papadatos 2004; Wang and Akritas 2002). The technical derivation involves the asymptotic approximation of a new class of quadratic forms (Sec. 3.2). Some new asymptotic tools are developed for this purpose.

In the sequel, we use $\mathrm{N}(\mu, \sigma^2)$ to represent a normal distribution with mean μ and variance σ^2 and let $a \wedge b = \min(a, b)$ and $a \vee b = \max(a, b)$. We let $\mathbf{1}_d$ denote the $d \times 1$ column vector of 1's, $\mathbf{J}_d = \mathbf{1}_d\mathbf{1}_d'$, $\mathbf{I}_d$ represent the d-dimensional identity matrix, and $\oplus$ and $\otimes$ denote the operations of Kronecker sum and Kronecker product.

The rest of the article is organized as follows. In the next section we introduce the test statistics. In Section 3 we present the asymptotic techniques and main theoretical properties, and in Section 4 we report results from several simulation studies. In Section 5 we apply the tests to two real datasets: the White Spanish Onion data and the North Carolina Rain data. We provide some generalizations and discussions in Section 6 and give proofs of the main theorems and some auxiliary technical results in two appendixes.

2. TEST STATISTICS

We adopt the convention that in each category i, the observed pairs (X_{ij}, Y_{ij}), $i = 1, \dots, k$, $j = 1, \dots, n_i$, are enumerated in ascending order of the covariate values; thus $X_{i1} < X_{i2} < \cdots < X_{in_i}$ holds for each i. Also, we let $X_1 < X_2 < \cdots < X_N$ denote the pooled covariate values X_{ij}'s in ascending order, where $N = \sum_{i=1}^{k} n_i$ is the total number of observations.

In the context of the nonparametric ANCOVA model (1), there is no conceptual distinction between the ANCOVA and ANOVA designs, because the effect of the covariate is not modeled. The idea, then, for constructing the test statistics is to treat the ANCOVA design as an ANOVA design where one factor, corresponding to the covariate, has levels $X_1, X_2, \dots, X_N$. Thus recent advances in testing hypotheses in ANOVA with one factor having many levels become relevant for the present testing problems. Such methods are not directly applicable, however, because of the sparseness issue (lack of replications in the cells) caused by the continuity of the covariate. Our proposal for dealing with the sparseness in the ANOVA design is to use the fact that the covariate factor is actually a regressor and to use smoothness assumptions.

Basically, a smoothness assumption stipulates that the conditional distribution of the response changes smoothly with the covariate value. Different versions of smoothness assumptions are ubiquitous in nonparametric regression. They are used for enhancing or augmenting the available information about the conditional distribution of the response at a given x-value by also considering the responses at covariate values that are "close" to x or belong in a local "window" around x. In our case we need to augment the available information for each F_{ix}, where $i = 1, \dots, k$ and x takes any of the covariate values $X_1, X_2, \dots, X_N$. Because the present hypothesis testing objective is different than that of nonparametric estimation, the window size n is required to increase with N at a rate only faster than $\log N$ [see Assumption A1(b)]. This allows the use of fairly small local window sizes even for large N, and simulations suggest that the actual choice of the local window size does not seem to have a large influence on the performance of the testing procedures.

Augmenting the information that is available for F_{ix} means, in terms of the ANOVA design, augmenting the information available in the corresponding cell (i, x). In particular, each cell (i, X_r) is augmented to include the n responses from category i whose corresponding covariate values X_{ij} are nearest to X_r in the sense that

$$|\widehat{G}_i(X_{ij}) - \widehat{G}_i(X_r)| \le \frac{n-1}{2n_i},$$

where $\widehat{G}_i(x) = n_i^{-1} \sum_{j=1}^{n_i} I(X_{ij} \le x)$ is the empirical distribution function of the covariate in group i. We use the notation W_{ir} to denote either the set of responses or the set of indices of the observations in the augmented cell (i, X_r); the meaning should be clear from the context. To distinguish the observations in the augmented ANOVA design from those of the original observations, the tth observation in cell (i, X_r) is denoted by Z_{irt}. Specifically, $Z_{irt} = Y_{ij}$ if and only if $\sum_{l=1}^{n_i} I(X_{il} \le x_{ij}, Y_{il} \in W_{ir}) = t$.

Now let F_D and F_C denote the classical F statistics in balanced two-way ANOVA for testing the null hypotheses of no main column effects and no row–column interaction effects, evaluated on the augmented data. Then F_D and F_C have the expressions

$$\begin{aligned} F_D &= \frac{MST_D}{MSE} \\ &= \frac{kn(N-1)^{-1} \sum_{r=1}^N (\overline{Z}_{\cdot r \cdot} - \overline{Z}_{\dots})^2}{(Nk(n-1))^{-1} \sum_{i=1}^k \sum_{r=1}^N \sum_{t=1}^n (Z_{irt} - \overline{Z}_{ir\cdot})^2} \end{aligned}$$

and

$$\begin{aligned} F_C &= \frac{MST_C}{MSE} \\ &= \frac{n\{(N-1)(k-1)\}^{-1}}{(Nk(n-1))^{-1}} \\ &\quad \times \frac{\sum_{i=1}^k \sum_{r=1}^N (\overline{Z}_{ir\cdot} - \overline{Z}_{\cdot r \cdot} - \overline{Z}_{i\cdot\cdot} + \overline{Z}_{\dots})^2}{\sum_{i=1}^k \sum_{r=1}^N \sum_{t=1}^n (Z_{irt} - \overline{Z}_{ir\cdot})^2}. \end{aligned}$$

Unlike in the classical ANOVA model, $E(Z_{irt} - \overline{Z}_{ir\cdot}) = 0$ is no longer true. Similarly, $E(\overline{Z}_{\cdot r \cdot} - \overline{Z}_{\dots}) \ne 0$ under $H_0(D)$ and $E(\overline{Z}_{ir\cdot} - \overline{Z}_{\cdot r\cdot} - \overline{Z}_{i\cdot\cdot} + \overline{Z}_{\dots}) \ne 0$ under $H_0(C)$. The reason why F_D and F_C are suitable for testing the nonparametric hypotheses (3) and (4) is that the aforementioned expectations tend to 0 under their respective null hypotheses and smoothness assumptions; see Lemma B.2 in Appendix B.

For the ANOVA with independent observations and a large number of factor levels, asymptotic distributions are obtained by considering $N^{1/2}(F_D - 1)$ and $N^{1/2}(F_C - 1)$ (see Wang and Akritas 2002). But these results do not apply here, due to the dependence of the observations in the augmented layout. In Section 3.2 a new set of asymptotic approximation techniques is developed that is suitable for our purpose. Due to the local smoothing, the parametric convergence rate $N^{1/2}$ is impossible to achieve. The next section reveals that an appropriate scaling constant should be of order $N^{1/2}n^{-1/2}$. Because under weak regularity conditions, the MSE tends to a constant in probability, by Slutsky's theorem, it suffices to study the asymptotic distribution of

$$N^{1/2}n^{-1/2}T_D = N^{1/2}n^{-1/2}(MST_D - MSE) \tag{8}$$

and

$$N^{1/2}n^{-1/2}T_C = N^{1/2}n^{-1/2}(MST_C - MSE), \tag{9}$$

which we use as the test statistics for $H_0(D)$ and $H_0(C)$, from now on.

3. ASYMPTOTIC DISTRIBUTIONS OF THE TEST STATISTICS

3.1 Conventions and Assumptions

The test statistics in (8) and (9) can be expressed as $T_D = \mathbf{Z}'\mathbf{A}_D\mathbf{Z}$ and $T_C = \mathbf{Z}'\mathbf{A}_C\mathbf{Z}$, where $\mathbf{Z} = (Z_{111}, Z_{112}, \dots, Z_{kNn})'$ is the vector of all observations in the augmented design and

$$\begin{aligned} \mathbf{A}_D = {} & \frac{1}{(N-1)kn} \bigoplus_{r=1}^N \mathbf{J}_{kn} - \frac{1}{N(N-1)kn} \mathbf{J}_{Nkn} \\ & - \frac{1}{Nk(n-1)} \mathbf{I}_{Nkn} + \frac{1}{Nk(n-1)n} \bigoplus_{r=1}^N \bigoplus_{i=1}^k \mathbf{J}_n \end{aligned}$$

and

$$\begin{aligned} \mathbf{A}_C = {} & \frac{1}{(N-1)(k-1)n} \left(\bigoplus_{r=1}^N \bigoplus_{i=1}^k \mathbf{1}_n \right) \\ & \times \left(\mathbf{I}_{Nk} - \mathbf{J}_N \otimes \left(\bigoplus_{i=1}^k \frac{1}{N} \right) - \bigoplus_{r=1}^N \frac{1}{k} \mathbf{J}_k + \frac{1}{Nk} \mathbf{J}_{Nk} \right) \\ & \times \left(\bigoplus_{r=1}^N \bigoplus_{i=1}^k \mathbf{1}_n' \right) - \frac{1}{Nk(n-1)} \left(\mathbf{I}_{Nkn} - \frac{1}{n} \bigoplus_{r=1}^N \bigoplus_{i=1}^k \mathbf{J}_n \right). \end{aligned}$$

For later reference, we also define two other quadratic forms, $\mathbf{Z}'\mathbf{B}_D\mathbf{Z}$ and $\mathbf{Z}'\mathbf{B}_C\mathbf{Z}$, where

$$\mathbf{B}_D = \frac{1}{Nkn} \bigoplus_{r=1}^N \mathbf{J}_{kn} - \frac{1}{Nk(n-1)} \mathbf{I}_{Nkn} + \frac{1}{Nk(n-1)n} \bigoplus_{i=1}^k \bigoplus_{r=1}^N \mathbf{J}_n$$

is block diagonal with block size $kn \times kn$ and

$$\mathbf{Z}'\mathbf{B}_C\mathbf{Z} = \frac{1}{Nk(n-1)}\sum_{i=1}^{k}\sum_{r=1}^{N}[\mathbf{Z}_{ir}'\mathbf{J}_n\mathbf{Z}_{ir} - \mathbf{Z}_{ir}'\mathbf{Z}_{ir}] - \frac{1}{Nk(k-1)n}\sum_{i_1\neq i_2}^{k}\sum_{r=1}^{N}\mathbf{Z}_{i_1r}'\mathbf{J}_n\mathbf{Z}_{i_2r},$$

where $\mathbf{Z}_{ir} = (Z_{ir1}, \dots, Z_{irn})'$ is a vector of all observations in the augmented cell (i, X_r). Moreover, we set $K(u) = I(|u| \le 1)$, $G(x)$ for the overall distribution function of X, G_i for the distribution function of X in the ith group, and $h_i = (n-1)/(2n_i)$, and define

$$K_{i,G}(X_1, X_2) = K\left(\frac{G_i(X_1) - G_i(X_2)}{h_i}\right). \tag{10}$$

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

Assumption A1. (a) For each i, $n_i/N \to \lambda_i \in (0,1)$ as $n_i \to \infty$.

(b) $\ln(Nn^{-1}) = o(n)$, $\ln(Nn^{-1})/\ln\ln N \to \infty$, $N^{-3}n^5 = o(1)$ as $N \to \infty$.

Assumption A2. (a) The covariate X_i is a continuous random variable, $i = 1, \dots, k$.

(b) The X_i's have common bounded support S, $i = 1, \dots, k$.

(c) The density g_i of X_i is bounded away from 0 and ∞ on S and is differentiable, $i = 1, \dots, k$.

(d) $\int y\, dF_i(y|x)$ are uniformly Lipschitz continuous of order 1 in the following sense: There exists some positive constant c such that

$$\left|\int y\, dF_i(y|x_1) - \int y\, dF_i(y|x_2)\right| \le c|x_1 - x_2| \quad \text{for all } i, x_1, x_2.$$

Assumption A3. $E(Y_{ij}^4|X_{ij} = x)$ are uniformly bounded in i, j, x.

The restriction that Assumption A1(b) puts on n is very weak; for example, n is allowed to go to infinity at the rate $\log N$. This assumption also implies that $n = o(N^{3/5})$. We find that a typical order of $N^{1/2}$ often works well in practice. Our simulation results indicate that the nonparametric test is usually valid for a wide range of choices of n. Assumptions A2 and A3 impose weak smoothness and moment constraints.

3.2 A New Class of Quadratic Forms and Asymptotic Approximations

The test statistics that we are dealing with are quadratic forms of the type $\mathbf{Z}'\mathbf{M}\mathbf{Z}$, where $\mathbf{M}$ is a symmetric matrix and $\mathbf{Z}$ is a random column vector whose dimension goes to infinity. When the coordinates of $\mathbf{Z}$ are independent and homoscedastic random variables, the asymptotic distribution of this type of quadratic form was studied by Boos and Brownie (1995), Jiang (1996), and Akritas and Arnold (2000); when the coordinates of $\mathbf{Z}$ are independent but heteroscedastic, it was investigated by Akritas and Papadatos (2004), who applied a novel variant of Hájek's projection method. In our setting, besides the fact that $\mathbf{Z}$ is nonnormal and heteroscedastic, its distribution depends on a vector of covariates, and its coordinates are correlated due to the local augmentation. To our knowledge, no asymptotic method in the literature is directly applicable.

Our approach for obtaining the asymptotic distribution of this class of quadratic forms consists of four steps. First, it can be assumed without loss of generality that the observations are conditionally centered under the null hypothesis; this is done in Proposition 1. Second, the test statistics are approximated by simpler quadratic forms obtained by an application of Akritas and Papadatos's (2004) variant of Hájek's projection method; this is done in Proposition 2 (although the projection calculations, which are quite tedious, are not shown). Third, the simpler quadratic forms are further approximated by the clean generalized quadratic forms (de Jong 1987), which involve only independent variables; this is done in Proposition 3. Finally, the conditions for establishing the asymptotic normality of the clean generalized quadratic forms (proposition 3.2 of de Jong 1987) are verified, which proves the asymptotic normality of the proposed test statistics; the last step is presented in the next section. The new techniques developed in this article are of independent interest and are likely to be useful for asymptotic analysis of other similar statistics.

Proposition 1. Assume that Assumptions A1 and A2 hold. Then the following results apply:

(a) Under $H_0(D)$, $N^{1/2}n^{-1/2}[\mathbf{Z}'\mathbf{A}_D\mathbf{Z} - (\mathbf{Z} - E(\mathbf{Z}|X))' \times \mathbf{A}_D(\mathbf{Z} - E(\mathbf{Z}|X))] \to 0$ in probability.

(b) Under $H_0(C)$, $N^{1/2}n^{-1/2}[\mathbf{Z}'\mathbf{A}_C\mathbf{Z} - (\mathbf{Z} - E(\mathbf{Z}|X))' \times \mathbf{A}_C(\mathbf{Z} - E(\mathbf{Z}|X))] \to 0$ in probability.

Proposition 2. Assume that Assumptions A1–A3 hold. Then the following results apply:

(a) Under $H_0(D)$, $N^{1/2}n^{-1/2}(\mathbf{Z}'\mathbf{A}_D\mathbf{Z} - \mathbf{Z}'\mathbf{B}_D\mathbf{Z}) \to 0$ in probability.

(b) Under $H_0(C)$, $N^{1/2}n^{-1/2}(\mathbf{Z}'\mathbf{A}_C\mathbf{Z} - \mathbf{Z}'\mathbf{B}_C\mathbf{Z}) \to 0$ in probability.

Proposition 3. Assume that Assumptions A1–A3 hold. Then the following results apply:

(a) Under $H_0(D)$, $N^{1/2}n^{-1/2}(\mathbf{Z}'\mathbf{B}_D\mathbf{Z} - T) \to 0$ in probability, where $T = T_1 + T_2$, with

$$T_1 = \frac{1}{k(n-1)}\sum_{i=1}^{k}\sum_{l_1\neq l_2}^{n_i} Y_{il_1}Y_{il_2} \times \int K_{i,G}(X_{il_1}, x)K_{i,G}(X_{il_2}, x)\, dG(x)$$

and

$$T_2 = \frac{1}{kn}\sum_{i_1\neq i_2}^{k}\sum_{l_1=1}^{n_{i_1}}\sum_{l_2=1}^{n_{i_2}} Y_{i_1l_1}Y_{i_2l_2} \times \int K_{i_1,G}(X_{i_1l_1}, x)K_{i_2,G}(X_{i_2l_2}, x)\, dG(x).$$

(b) Under $H_0(C)$, $N^{1/2}n^{-1/2}(\mathbf{Z}'\mathbf{B}_C\mathbf{Z} - \widetilde{T}) \to 0$ in probability, where $\widetilde{T} = T_1 - (k-1)^{-1}T_2$, for T_1 and T_2 defined earlier.

3.3 Asymptotic Distribution Under the Null

The two theorems that follow provide the asymptotic distributions of the test statistics under the null hypotheses of no nonparametric main covariate effects and no nonparametric factor–covariate interaction effects.

Theorem 1. Assume that Assumptions A1–A3 hold. Then, under $H_0(D)$,

$$N^{1/2}n^{-1/2}T_D \to \mathrm{N}\left(0, \frac{4}{3k^2}(\xi^4+\eta^4)\right),$$

where, setting $\sigma_i^2(x) = \operatorname{var}(Y_{ij}|X_{ij}=x)$, $\rho_i(t) = \lambda_i g_i(t)$, $g(t) = \sum_{i=1}^k \rho_i(t)$, and $\tau_{i_1,i_2}(t) = 3(\rho_{i1}(t)\wedge\rho_{i2}(t)) - (\rho_{i1}(t)\wedge\rho_{i2}(t))^2/(\rho_{i1}(t)\vee\rho_{i2}(t))$,

$$\xi^4 = \sum_{i=1}^k \int \sigma_i^4(t)\frac{g^2(t)}{\rho_i(t)}\,dt \qquad \text{and}$$

$$\eta^4 = \sum_{i_1<i_2} \int \sigma_{i_1}^2(t)\sigma_{i_2}^2(t)\frac{g^2(t)}{\rho_{i1}(t)\rho_{i2}(t)}\tau_{i_1,i_2}(t)\,dt.$$

Theorem 2. Assume that Assumptions A1–A3 hold. Then, under $H_0(C)$,

$$N^{1/2}n^{-1/2}T_C \to \mathrm{N}\left(0, \frac{4}{3k^2}\xi^4 + \frac{4}{3k^2(k-1)^2}\eta^4\right),$$

where ξ^4 and η^4 are defined as in Theorem 1.

To implement the tests, we need to estimate the asymptotic variances. In simulations and data analysis, we use the following simple and consistent estimators for ξ^4 and η^4:

$$\widehat{\xi}^4 = \frac{3}{2Nn(n-1)^2}\sum_{i=1}^k\sum_{l_1\neq l_2}^{n_i}\widehat{\sigma}_i^2(x_{il_1})\widehat{\sigma}_i^2(x_{il_2}) \times \left[\sum_{r=1}^N I(x_{il_1}\in W_{ir})I(x_{il_2}\in W_{ir})\right]^2$$

and

$$\widehat{\eta}^4 = \frac{3}{2Nn^3}\sum_{i_1\neq i_2}^{k}\sum_{l_1=1}^{n_{i_1}}\sum_{l_2=1}^{n_{i_2}}\widehat{\sigma}_{i_1}^2(x_{i_1l_1})\widehat{\sigma}_{i_2}^2(x_{i_2l_2}) \times \left[\sum_{r=1}^N I(x_{i_1l_1}\in W_{i_1r})I(x_{i_2l_2}\in W_{i_2r})\right]^2,$$

where, for $\widehat{\sigma}_i^2(x_{il_1})$ we take the sample variance of all observations in the window centered around x_{il_1} and similarly for the other conditional variances. One advantage of these estimators is their computational convenience because they avoid directly estimating the density functions of the covariates.

3.4 Asymptotic Distribution Under Local Alternatives

To investigate the power properties of the tests, we consider the sequence of local alternatives

$$F_{N,ix}(y) = F_{ix}(y) + (Nn)^{-1/4}R_{ix}(y), \tag{11}$$

where $F_{ix}(y)$, $i=1,\dots,k$, satisfy the null hypothesis of interest and $R_{ix}(y)$, $i=1,\dots,k$, are such that $\int y\,dR_{ix}(y)$ and $\int y^2\,dR_{ix}(y)$ are uniformly bounded for all i and x. It is also assumed that for some positive constant C, $|R_{ix_1}(y) - R_{ix_2}(y)| \le C|x_1-x_2|$, for all i, y. Define

$$\theta_D = \int\left[\frac{1}{k}\sum_{i=1}^k\int y\,dR_{ix}(y)\right]^2 dG(x) - \left[\frac{1}{k}\sum_{i=1}^k\int y\,dR_{ix}(y)\,dG(x)\right]^2$$

and

$$\theta_C = \sum_{i=1}^k\int\left[\int y\,dR_{ix}(y) - \frac{1}{k}\sum_{j=1}^k\int y\,dR_{jx}(y) - \int y\,dR_{it}(y)\,dG(t) + \frac{1}{k}\sum_{j=1}^k\int y\,dR_{jt}(y)\,dG(t)\right]^2 dG(x).$$

The next lemma plays a central role in the derivation of the asymptotic distributions under the local alternatives. Let $E_N(\cdot|X)$ denote expectation under the local conditional distribution functions specified by (11); then, instead of Proposition 1, we have the following result.

Lemma 1. Assume that Assumptions A1 and A2 hold and that the sequence of conditional distributions $F_{N,ix}(y)$ satisfies (11). Then the following results apply:

(a) If $F_{ix}(y)$, $i=1,\dots,k$, satisfy $H_0(D)$, then

$$N^{1/2}n^{-1/2}\big[\mathbf{Z}'\mathbf{A}_D\mathbf{Z} - (\mathbf{Z}-E_N(\mathbf{Z}|X))'\mathbf{A}_D(\mathbf{Z}-E_N(\mathbf{Z}|X))\big] \to k\theta_D$$

in probability.

(b) If $F_{ix}(y)$, $i=1,\dots,k$, satisfy $H_0(C)$, then

$$N^{1/2}n^{-1/2}\Big[\mathbf{Z}'\mathbf{A}_C\mathbf{Z} - (\mathbf{Z}-E_N(\mathbf{Z}|X))'\mathbf{A}_C(\mathbf{Z}-E_N(\mathbf{Z}|X))\Big] \to \frac{\theta_C}{k-1}$$

in probability.

The asymptotic distributions of the test statistics under the foregoing local alternative sequences are given by the following two theorems.

Theorem 3. Let ξ^4 and η^4 be as defined in Theorem 1. Then, under Assumptions A1–A3, for the sequence of local alternatives $F_{N,ix}(y)$ in (11) with $F_{ix}(y)$ satisfying $H_0(D)$,

$$N^{1/2}n^{-1/2}T_D \to \mathrm{N}\left(k\theta_D, \frac{4(\xi^4+\eta^4)}{3k^2}\right).$$

Theorem 4. Let ξ^4 and η^4 be as defined in Theorem 1. Then, under Assumptions A1–A3, for the sequence of local alternatives $F_{N,ix}(y)$ in (11) with $F_{ix}(y)$ satisfying $H_0(C)$,

$$N^{1/2}n^{-1/2}T_C \to \mathrm{N}\left(\frac{\theta_C}{k-1}, \frac{4\xi^4}{3k^2} + \frac{4\eta^4}{3k^2(k-1)^2}\right).$$

4. SIMULATIONS

We carried out Monte Carlo simulations to compare the performance of the proposed nonparametric test (NP test) with that of some common alternatives. We focus the comparisons on testing the null hypothesis of no interaction effects. All simulation results use a nominal level of .05 and are based on 500 runs. Covariate values are generated separately for each group from the uniform distribution on (0, 1) unless specified otherwise. In this section the ϵ_{ij}'s represent independent standard normal random variables.

For a continuous response variable, the NP test is compared with the semiparametric test (YB test) of Young and Bowman, the classical F test (CF test), and the drop test of McKean and Schrader (1980) (see also Hettmansperger and McKean 1998, sec. 3.6). We note that the term "drop test" was used by Terpstra and McKean (2004), who recently developed R software for the implementation of this test. For a dichotomous response variable, the NP test is compared with the deviance test in logistic regression model.

We first present results from three small simulations that demonstrate that the three alternative tests for the case of a continuous response variable are sensitive to departures from the assumptions that underly their asymptotic distribution theory. All three simulations use two groups and one continuous covariate. To investigate the effect of heteroscedasticity, we take $n_1 = 60$ and $n_2 = 40$ and generate data according to $Y_{1j} = .1X_{1j}^2\epsilon_{1j}$ and $Y_{2j} = .1X_{2j}^2\epsilon_{2j}$. The observed type I error for the NP test is .076 ($n = 7$), .069 ($n = 9$), and .057 ($n = 11$), for the YB test it is .117, for the CF test it is .131, and for the drop test it is .264. To investigate the effect of covariate distribution, we take $n_1 = n_2 = 30$ and generate the covariate of the first group from uniform distribution on $(10, 20)$ and the covariate of the second group from $8 + 14 * Beta(3, 3)$ restricted on $(10, 20)$; the response variables are generated by $Y_{ij} = 1 + .3X_{ij} + \epsilon_{ij}$. The observed type I error for the NP test is .058 ($n = 5$), .042 ($n = 7$), .034 ($n = 9$), that for the YB test is .328, that for the CF test is .052, and that for the drop test is .054. Finally, to investigate the effect of departures from the location-scale model, we take $n_1 = 60$ and $n_2 = 40$, and generate data according to a mixture distribution, $Y_{1j}|X_{1j} = x \sim F_{1x}(y) = .5\Phi((y-2)/.1) + .5\Phi((4x-y)/.1)$, $Y_{2j}|X_{2j} = x \sim F_{2x}(y) = .5\Phi(y/.1) + .5\Phi((4x-y)/.1)$, where $\Phi(\cdot)$ is the cumulative distribution function of the standard normal distribution; thus the conditional means for groups 1 and 2 are $1 + 2x$ and $2x$. The observed type I error for the NP test is .078, .064, .056 for $n = 7, 9, 11$, that for the YB test is .182, that for the CF test is .144, and that for the drop test is .556.

As expected, the achieved level of the NP test is reasonable for a wide variety of data-generating mechanisms. Because of the sensitivity of the other tests to violations of certain assumptions, Examples 1 and 2 use homoscedastic normal errors.

Example 1. Here the data are generated according to

$$Y_{1j} = .1\epsilon_{1j} \qquad \text{and} \qquad Y_{2j} = \theta X_{2j} + .1\epsilon_{2j} \tag{12}$$

for $\theta = 0, .1, .2, .3$, where $\theta = 0$ corresponds to the null hypothesis. The results are summarized in Table 1.

Table 1. Estimated Powers of the NP Test, YB Test, CF Test, CO Test, and Drop Test for the Linear Alternative (12) at Level $\alpha = .05$, Based on 500 Simulations

		θ	0	.1	.2	.3
n_i	Test	n	level	power	power	power
30	NP	3	.090	.150	.395	.651
		5	.076	.145	.413	.694
		7	.062	.141	.427	.707
	YB		.054	.088	.201	.415
	CF		.057	.201	.578	.881
	Drop		.048	.158	.540	.860
50	NP	5	.062	.172	.529	.872
		7	.052	.175	.564	.908
		9	.051	.185	.592	.915
	YB		.053	.107	.304	.658
	CF		.033	.280	.807	.982
	Drop		.036	.294	.784	.984

In this example the assumptions of the classical ANCOVA are satisfied, and, as expected, the CF test outperforms the other three. The drop test is slightly less powerful than the CF test, whereas the NP test outperforms the YB test. The reduced power of the NP and YB tests for linear alternatives is the price that they must pay for guarding against omnibus alternatives, as is demonstrated in the next example.

Example 2. Here the data are generated according to

$$\begin{aligned} Y_{1j} &= .1\epsilon_{1j} \qquad \text{and} \\ Y_{2j} &= \theta(X_{2j}^2 - X_{2j} + .15) + .1\epsilon_{2j} \end{aligned} \tag{13}$$

for $\theta = 0, .5, 1.0, 1.5$, where $\theta = 0$ corresponds to the null hypothesis. The results are summarized in Table 2. The superiority of the NP test is obvious, whereas the powers of the CF test and the drop test are close to the specified nominal level and the YB test has power in between.

Example 3. We consider data with a binary response variable. The CF test, drop test, and YB test are inappropriate for such data. The logistic regression model is commonly used for handling this type of data. It assumes that the binary response has a Bernoulli distribution and that the covariate has a linear effect on the logit of the success probability. A test for interaction effects can be obtained by comparing the deviance under

Table 2. Estimated Powers of the NP Test, YB Test, CF Test, CO Test, and Drop Test for the Quadratic Alternative (13) at Level $\alpha = .05$, Based on 500 Simulations

		θ	0	.5	1.0	1.5
n_i	Test	n	level	power	power	power
30	NP	3	.089	.198	.498	.812
		5	.060	.162	.438	.813
		7	.042	.129	.397	.742
	YB		.043	.107	.304	.646
	CF		.068	.053	.078	.099
	Drop		.040	.042	.084	.074
50	NP	5	.070	.239	.670	.964
		7	.053	.236	.687	.964
		9	.042	.213	.666	.961
	YB		.058	.164	.467	.866
	CF		.043	.053	.059	.078
	Drop		.034	.038	.066	.088

Table 3. Estimated Powers of the NP Test and the Deviance Test of the Logistic Regression Model for (14) at Level $\alpha = .05$, Based on 500 Simulations

n_i	Test	θ n	0 level	1 power	2 power	3 power	4 power
30	NP	3	.060	.106	.306	.486	.590
		5	.026	.074	.266	.458	.596
		7	.016	.062	.206	.380	.518
	Logistic		.068	.084	.058	.084	.060
50	NP	5	.038	.094	.468	.746	.874
		7	.030	.096	.490	.800	.890
		9	.028	.078	.482	.788	.910
	Logistic		.068	.058	.064	.052	.064

the full model and under the reduced model. To make a comparison with the NP test, we simulate an experiment with three different groups ($k = 3$),

$$Y_{1j} \sim \text{Bernoulli}\left(\frac{\exp(\theta\cos(2\pi X_{1j}))}{1+\exp(\theta\cos(2\pi X_{1j}))}\right) \tag{14}$$

for $\theta = 0, 1, 2, 3, 4$; Y_{2j} and Y_{3j} are independent Bernoulli random variables with success probability .5. The results from the NP test and the deviance test based on logistic regression model are displayed in Table 3. The NP test is very powerful, whereas the deviance test has power close to its nominal level.

5. EMPIRICAL STUDIES

5.1 Case Study 1: White Spanish Onion Data

Many studies in agricultural science investigate the relationship between yield and planting density for different crops. This dataset, taken from Ratkowsky (1983), consists of 84 measurements on white Spanish onions from two different locations in south Australia, Purnong Landing and Virginia.

Of interest is to investigate whether density (plants/m^2) has a significant effect on yield (g/plant) and if the yield–density relationship is the same for the two locations. Following Young and Bowman (1995), who analyzed this dataset for interaction effects, we took off the two obvious outliers from the Virginia group corresponding to the smallest and greatest densities. Figure 1(a) depicts the relation between the yield on the natural log scale and the density, where the circles represent the Purnong Landing group and the crosses represent the Virginia group. The local linear smoothers imposed on the scatter plot use Gaussian kernel and bandwidth selected by the direct plug-in method (Ruppert, Sheather, and Wand 1995).

For the covariate effects, the NP test (for a wide range of window sizes), the CF test, and the drop test all give p values of 0, indicating a significant density effect on yield.

For the interaction effects, the CF test gives a p value of .0363, the YB test gives high p values for a wide range of smoothing parameters, the NP test gives p values of .7100, .6187, and .3247 for window size 7, 9, 11, and the drop test gives a p value of .1051. Moreover, Ratkowsky (1983) using an approximate F test within Holliday's (1960) nonlinear yield-density model, obtained a p value of .2523.

A careful exploration of the data helps explain why the CF test suggests significant interaction effects when the other four tests do not. Although the normality assumption seems reasonable, the homoscedasticity assumption appears to be violated. This is suggested by plots of the standardized residuals from the classical parametric model and of the estimated (by the method of Ruppert, Wand, Holst, and Hössjer 1997) conditional variance functions displayed in Figure 1. The simulations reported before Example 1 of Section 4 point out the possible liberal behavior of the classical F test under such model violations. The standardized residual plot in the Figure 1(c) also reveals that the observation $(21.25, 5.4606)$ in the Virginia group has a standardized residual of 2.745. When this data point is removed, the p value of the CF test jumps to .0857, suggesting its unstable behavior. We remark that with this data point removed, the other tests continue to indicate nonsignificant interaction effects, whereas all p values for testing for no covariate effects continue to be near 0.

All other tests but the fully nonparametric test also can be sensitive to the assumption of homoscedasticity. Investigation of the extent of such sensitivity of various tests in a controlled experimental setting is a worthwhile topic for future study. Our experience indicates that the divergence of the parametric and nonparametric tests usually suggests some kind of violation of the underlying parametric assumptions. Although the fully nonparametric test provides certain protection against model misspecification, we recommend that it is always good practice for a data analyst to validate any underlying model assumption before making inference.

5.2 Case Study 2: North Carolina Rain Data

The rain data were obtained as part of the National Atmospheric Deposition Program, which monitors wet atmospheric deposition at different locations across the United States. As an example, we consider data on the pH level of the precipitation in two North Carolina cities, Coweeta and Lewiston, and analyze it for time effects and location–time interaction effects. There are 120 monthly readings from January 1988 to December 1997. A scatterplot is given in Figure 2, with dots denoting observations from Coweeta and crosses denoting those from Lewiston. Local linear smoothers are imposed on this plot, with the solid line for Coweeta data and the dashed line for Lewiston data. Although clear seasonal oscillation is seen, a residual analysis does not suggest correlation over time.

For the interaction effects, the CF test gives a p value of .9876, the drop test gives a p value of .9492, the YB test gives p values close to 1 for a large range of smoothing parameters, and the NP test gives p values of .9135, .9561, .9615, .9659, and .9606 for local window sizes 7, 9, 11, 13, and 15. Thus all methods indicate that the interaction effects are statistically negligible.

For the covariate effects, the CF test gives a p value of .0579, the drop test gives a p value of .1229, and the NP test gives p values of 0, .0003, .0108, .0135, and .0065 for local window sizes 7, 9, 11, 13, and 15. The failure of the CF and drop tests to indicate a significant covariate effect is probably due to the fact that in this example the effect is not linear.

6. GENERALIZATIONS AND DISCUSSIONS

We have provided a thorough treatment of testing covariate effects and covariate–factor interaction effects for a single factor and a single continuous covariate. In this section we briefly consider data analysis in designs with more than one factor

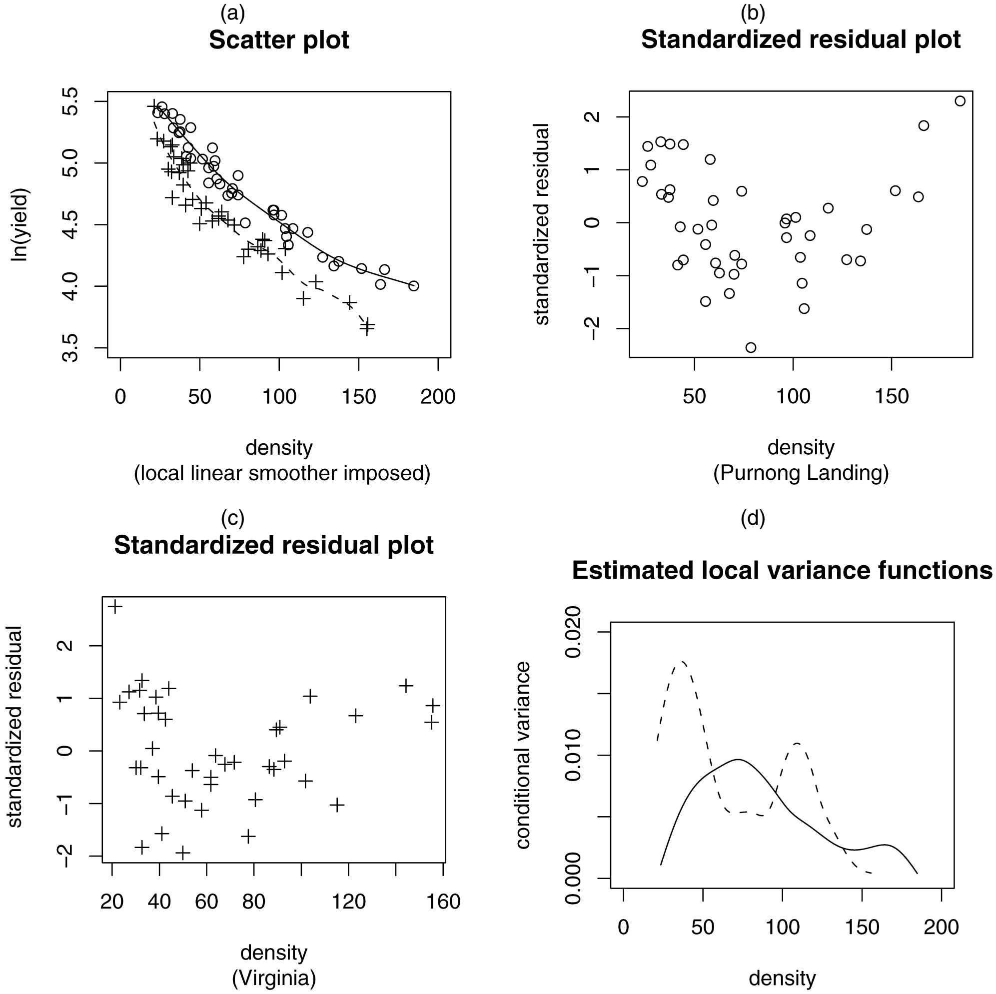

Figure 1. Plots of the White Spanish Onion Data. (a) A scatterplot where the circles denote the observations from Purnong Landing and the crosses denote the observations from Virginia, local linear smoothers are imposed. (b) The standardized residuals from the classical parametric model for the data of Purnong Landing. (c) The standardized residuals from the classical parametric model for the data of Virginia. (d) The local linear estimators of the two conditional variance functions, where the solid line represents Purnong Landing and the dashed line represents Virginia.

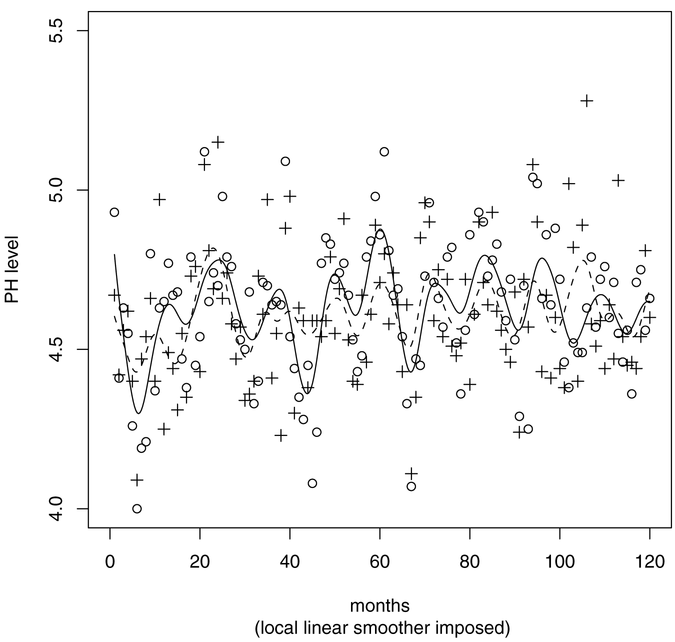

Figure 2. Scatterplot of North Carolina Rain Data.

and/or multiple covariates. We also discuss a new test statistic that can incorporate general weighting scheme and the challenges for finding optimal weights.

6.1 Two-Way and Higher-Way ANCOVA

The case of more than one factor is readily incorporated in the nonparametric model (1) by imposing a structure on the index i. To illustrate the ideas, we consider a two-way ANCOVA model, where the row factor has a levels and the column factor has b levels. The nonparametric model assumes only that conditional on the covariate value $X_{ijk} = x$, the responses Y_{ijk} are independent with conditional distribution function

$$Y_{ijk}|X_{ijk} = x \sim F_{ijx}.$$

The decomposition (7) can be similarly written as

$$F_{ijx}(y) = M(y) + A_i(y) + B_j(y) + (AB)_{ij}(y) + D_x(y) + C_{ijx}(y),$$

where $M(y) = (ab)^{-1}\sum_{i=1}^{a}\sum_{j=1}^{b}\int F_{ijx}(y)\,dG(x)$, $A_i(y) = b^{-1}\sum_{j=1}^{b}\int F_{ijx}(y)\,dG(x) - M(y)$, $B_j(y) = a^{-1}\sum_{i=1}^{a}\int F_{ijx}(y) \times dG(x) - M(y)$, $D_x(y) = (ab)^{-1}\sum_{i=1}^{a}\sum_{j=1}^{b} F_{ijx}(y) - M(y)$,

$(AB)_{ij}(y) = \int F_{ijx}(y)\,dG(x) - A_i(y) - B_j(y) + M(y)$, and $C_{ijx}(y) = F_{ijx}(y) - M(y) - A_i(y) - B_j(y) - (AB)_{ij}(y) - D_x(y)$. The last term, $C_{ijx}(y)$, includes all first- and second-order interaction terms and can be further decomposed as

$$C_{ijx}(y) = (AD)_{ix}(y) + (BD)_{jx}(y) + (ABD)_{ijx}(y),$$

where $(AD)_{ix}(y) = b^{-1}\sum_{j=1}^{b} F_{ijx}(y) - A_i(y) - D_x(y) + M(y)$, $(BD)_{jx}(y) = a^{-1}\sum_{i=1}^{a} F_{ijx}(y) - B_j(y) - D_x(y) + M(y)$, and $(ABD)_{ijx}(y) = C_{ijx}(y) - (AD)_{ix}(y) - (BD)_{jx}(y)$.

In the foregoing two-way model, testing the null hypothesis of no main covariate effects [$H_0 : D_x(y) = 0$ for all x and y] will be exactly the same as in the one-way case, because we can pretend that the ab treatment combinations come from a single factor with ab levels. This is also true for testing the null hypothesis of no interaction effects between the covariate and the two factors [$H_0 : C_{ijx}(y) = 0$ for all i, j, x, and y]. Testing for no interaction effects between the covariate and a particular factor is not included in the previous development, but a test procedure can be constructed using the method described in Section 2. More specifically, to test for no interaction effects between the covariate and the row factor A [i.e., $H_0(AD) : (AD)_{ix}(y) = 0$ for all $i = 1, \dots, a$ and all x and y], one formulates a hypothetical three-way ANOVA with a factor of a levels (those of row factor A), a factor of b levels (those of column factor B), and a factor of N levels (corresponding to the covariate). This three-way design is then augmented to n observations per cell. If we denote the pseudo-observations by Z_{ijrt}, then the test statistic for testing $H_0(AD)$ is $T_{AD} = MSAD - MSE$, where

$$MSAD = \frac{bn\sum_{i=1}^{a}\sum_{r=1}^{N}(\overline{Z}_{i\cdot r\cdot} - \overline{Z}_{i\cdots} - \overline{Z}_{\cdot\cdot r\cdot} + \overline{Z}_{\cdots\cdot})^2}{(a-1)(N-1)}$$

and

$$MSE = \frac{\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{r=1}^{N}\sum_{t=1}^{n}(Z_{ijrt} - \overline{Z}_{ijr\cdot})^2}{abN(n-1)},$$

where $\overline{Z}_{ijr\cdot} = n^{-1}\sum_{t=1}^{n} Z_{ijrt}$, $\overline{Z}_{i\cdot r\cdot} = b^{-1}\sum_{j=1}^{b}\overline{Z}_{ijr\cdot}$, $\overline{Z}_{i\cdots} = (bN)^{-1}\sum_{j=1}^{b}\sum_{r=1}^{N}\overline{Z}_{ijr\cdot}$, $\overline{Z}_{\cdot\cdot r\cdot} = (ab)^{-1}\sum_{i=1}^{a}\sum_{j=1}^{b}\overline{Z}_{ijr\cdot}$, and $\overline{Z}_{\cdots\cdot} = (abN)^{-1}\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{r=1}^{N}\overline{Z}_{ijr\cdot}$. Following the same arguments as in the proof of Theorem 1, with slightly more complex notation, we can prove the asymptotic normality of T_{AD}. Under suitable regularity assumptions, we can show that when $H_0(AD)$ holds,

$$N^{1/2}n^{-1/2}T_{AD} \to \mathrm{N}\left(0, \frac{4}{3a^2b^2}(\xi^4 + \eta^4 + \phi^4)\right), \qquad (15)$$

where, setting $\sigma_{ij}^2(x) = \mathrm{var}(Y_{ijk}|X_{ijk} = x)$, $\lambda_{ij} = \lim_{N\to\infty} N^{-1} \times n_{ij}$, $\rho_{ij}(t) = \lambda_{ij}g_{ij}(t)$, $g(t) = \sum_{i=1}^{a}\sum_{j=1}^{b}\rho_{ij}(t)$, and $\tau_{i_1j_1,i_2j_2}(t) = 3(\rho_{i_1j_1}(t) \wedge \rho_{i_2j_2}(t)) - (\rho_{i_1j_1}(t) \wedge \rho_{i_2j_2}(t))^2/(\rho_{i_1j_1}(t) \vee \rho_{i_2j_2}(t))$,

$$\xi^4 = \sum_{i=1}^{a}\sum_{j=1}^{b}\int \sigma_{ij}^4(t)\frac{g^2(t)}{\rho_{ij}(t)}\,dt,$$

$$\eta^4 = \sum_{i=1}^{a}\sum_{j_1<j_2}^{b}\int \sigma_{ij_1}^2(t)\sigma_{ij_2}^2(t)\frac{g^2(t)}{\rho_{ij_1}(t)\rho_{ij_2}(t)}\tau_{ij_1,ij_2}(t)\,dt,$$

Table 4. Estimated Powers of the NP Test and CF Test for the Two-Way ANCOVA (16) at Level $\alpha = .05$, Based on 500 Simulations

Test	θ / n	0 / level	.5 / power	1.0 / power	1.5 / power
NP	5	.057	.257	.810	.987
	7	.040	.183	.787	.983
	9	.033	.160	.700	.970
CF		.040	.070	.073	.057

and

$$\phi^4 = \sum_{i_1<i_2}^{a}\sum_{j_1=1}^{b}\sum_{j_2=1}^{b}\int \sigma_{i_1j_1}^2(t)\sigma_{i_2j_2}^2(t)\frac{g^2(t)}{\rho_{i_1j_1}(t)\rho_{i_2j_2}(t)}\tau_{i_1j_1,i_2j_2}(t)\,dt.$$

These can be estimated similarly as described in Section 3.3.

We remark that the result in (15) covers testing for interaction effects between the covariate and a factor in multifactor ANCOVA designs. To see this, let factor A denote the factor whose interaction with the covariate is to be tested, and let factor B having as levels the factor-level combinations of all the other factors.

To illustrate the performance of the foregoing test, Table 4 reports the results of a small simulation study done using a two-way ANCOVA design in which both the row and column factors have two levels. Covariate values for each of the four cells are generated independently from a uniform distribution on (0, 1), and the responses are generated as

$$\begin{aligned} &Y_{11k} = .1\epsilon_{11k}, \qquad Y_{12k} = .1\epsilon_{12k}, \\ &Y_{21k} = \theta\cos(2\pi X_{21k}) + .1\epsilon_{21k}, \qquad \text{and} \\ &Y_{22k} = \theta\cos(2\pi X_{22k}) + .1\epsilon_{22k}, \end{aligned} \qquad (16)$$

for $k = 1, \dots, 30$ and $\theta = 0, .5, 1.0, 1.5$, where the ϵ_{ijk}'s are independent standard normal random variables. It is seen that the NP test exhibits satisfactory behavior and is more powerful than the CF test.

6.2 Multiple Covariates

Although the foregoing discussion is restricted to scalar covariate, the method of constructing test statistics can nonetheless be generalized to the multivariate covariate case. The ordering of a single covariate is straightforward. In practice, the ordering of a multivariate vector is often based on either certain distance measure, such as Euclidean distance based on standardized coordinates and Mahalanobis distance, or some associated score, such as sample score of variation. With appropriate ordering, a nearest-neighborhood window can be formed.

In the presence of multiple covariates, testing main covariate effects or factor–covariate interaction effects may proceed similarly if the covariate vector can be treated as a whole. Probably the simplest way to create nearest-neighborhood windows is to use the so-called "statistically equivalent blocks" or "Gassaman rule" (see, e.g., Gassaman 1970; Lugosi and Nobel 1996), where many small blocks are created such that they all contain equal numbers of observations. Each block may be considered a hypothetical cell; the test statistic and the asymptotic theory remain the same. It is also possible that a single covariate calls for special attention; for example, we may wish to test whether a particular covariate is significant. In this case this particular covariate should be separated from the others and treated as a factor with many levels. Test statistics are readily available by borrowing those from analysis of variance, the asymptotic tools developed in this article can be applied to derive their asymptotic distributions.

6.3 General Weights

In each local window, all of the observations are weighted equally. It is possible to design a similar but different test statistic such that a more general weighting scheme can be incorporated. Each hypothetical observation Z_{irt} can be assigned a weight that depends on the distance between its associated covariate X_{il} and the center of the corresponding local window, and these weighted observations are used in the test statistics. More specifically, we define

$$Z^*_{irt} = Z_{irt} K\left(\frac{\widehat{G}_i(X_{il}) - \widehat{G}_i(X_r)}{h_i}\right),$$

where $K(\cdot)$ represents a general kernel function. To test for the null hypotheses $H_0(D)$ and $H_0(C)$, we define the test statistics as in (8) and (9) but with Z_{irt} replaced by Z^*_{irt}. Then the test statistics considered in Section 2 are simply special cases with $K(u) = I(|u| \le 1)$. The techniques developed in this article can be generalized to analyze the new test statistics.

It is well known that in estimation theory, the Epanechnikov kernel is optimal for kernel smoothing in the sense that it minimizes the asymptotic mean integrated squared error. But this optimality does not automatically carry over to the present setting of hypothesis testing, where the power is the performance criterion. For any given alternative, the power function usually depends on the true unknown distribution of the data. The optimal weighting derived by maximizing this power function, even if mathematically tractable, is of limited practical use because it requires knowledge of both the alternative and the unknown data distribution. The fact that the alternative space considered in this article is an infinite-dimensional functional space makes the problem of optimal testing even more challenging. However, some general results might still be possible, as suggested by some recent work. For a one-sample problem of testing for no covariate effects, Eubank (2000) derived an asymptotically efficient smoothing lack-of-fit test that is most powerful over a class of alternatives. In a related problem of testing for the superiority among two regression curves, Koul and Schick (2003) showed that for any given local alternative, it is possible to find weights such that the test achieves the upper bound on the asymptotic power of all asymptotic level-α tests. We expect that these will stimulate similar future developments in the nonparametric analysis of covariance problem.

APPENDIX A: SKETCH OF PROOFS

The sketch of proofs is for the test statistic T_D. The asymptotic results for T_C can be proven similarly. For notational convenience, we take n to be odd.

Proof of Proposition 1

Let $Q = N^{1/2}n^{-1/2}[\mathbf{Z}'\mathbf{A}_D\mathbf{Z} - (\mathbf{Z} - E(\mathbf{Z}|X))'\mathbf{A}_D(\mathbf{Z} - E(\mathbf{Z}|X))]$, we show that $E(Q^2) \to 0$. After some algebraic manipulation,

$$E(Q^2|X) = Nn^{-1}\big[4E(\mathbf{Z}|X)'\mathbf{A}_D \operatorname{cov}(\mathbf{Z}|X)\mathbf{A}_D E(\mathbf{Z}|X) + \big(E(\mathbf{Z}|X)'\mathbf{A}_D E(\mathbf{Z}|X)\big)^2\big].$$

The square root of the second term in the foregoing sum is equal to

$$\begin{aligned} N^{1/2}n^{-1/2}&\Bigg[\frac{kn}{N-1}\sum_{r=1}^{N}\big(E(\overline{Z}_{\cdot r \cdot}|X) - E(\overline{Z}_{\cdots}|X)\big)^2 \\ &- \frac{1}{Nk(n-1)}\sum_{i=1}^{k}\sum_{r=1}^{N}\sum_{t=1}^{n}\big(E(Z_{irt}|X) - E(\overline{Z}_{ir\cdot}|X)\big)^2\Bigg] \\ &= N^{1/2}n^{-1/2}O(N^{-1}n)O(N)O(N^{-2}n^2) \\ &\quad - N^{1/2}n^{-1/2}O(N^{-1}n^{-1})O(Nn)O(N^{-2}n^2) \\ &= O\big(N^{-3/2}n^{5/2}\big), \end{aligned}$$

where the first equation is obtained by Lemma B.2 in Appendix B. We now evaluate the order of the first term of $E(Q^2|X)$. First, note that

$$\begin{aligned} \mathbf{A}_D E(\mathbf{Z}|X) &= \left[\frac{1}{(N-1)kn}\bigoplus_{r=1}^{N}\mathbf{J}_{kn} - \frac{1}{N(N-1)kn}\mathbf{J}_{Nkn}\right]E(\mathbf{Z}|X) \\ &\quad - \left[\frac{1}{Nk(n-1)}\mathbf{I}_{Nkn} - \frac{1}{Nk(n-1)n}\bigoplus_{r=1}^{N}\bigoplus_{i=1}^{k}\mathbf{J}_n\right]E(\mathbf{Z}|X), \end{aligned}$$

because the coordinate of the first $Nkn \times 1$ vector is of the form $(N-1)^{-1}(E(\overline{Z}_{\cdot r \cdot}|X) - E(\overline{Z}_{\cdots}|X))$, and that of the second vector is of the form $\{Nk(n-1)\}^{-1}(E(Z_{irt}|X) - E(\overline{Z}_{ir\cdot}|X))$, the order of $\mathbf{A}_D E(\mathbf{Z}|X)$ is $O(N^{-2}n)\mathbf{1}_{Nkn} + O(N^{-2})\mathbf{1}_{Nkn} = O(N^{-2}n)\mathbf{1}_{Nkn}$. The order of the first term of $E(Q^2|X)$ is $O(N^{-3}n)\mathbf{1}'_{Nkn}\operatorname{cov}(Z|X)\mathbf{1}_{Nkn} = O(N^{-2}n^3)$ by Lemma B.3 in Appendix B. Thus $E(Q^2|X) = O(N^{-2} \times n^3) + O(N^{-3}n^5) = O(N^{-3}n^5)$. By dominated convergence theorem (DCT), $E(Q^2) \to 0$.

Proof of Proposition 2

Denote $Q = N^{1/2}n^{-1/2}(\mathbf{Z}'\mathbf{B}_D\mathbf{Z} - \mathbf{Z}'\mathbf{A}_D\mathbf{Z})$, then

$$Q = \frac{1}{N^{1/2}(N-1)kn^{3/2}}\sum_{i_1=1}^{k}\sum_{i_2=1}^{k}\sum_{r_1 \ne r_2}^{N}\sum_{t_1=1}^{n}\sum_{t_2=1}^{n} Z_{i_1 r_1 t_1}Z_{i_2 r_2 t_2}.$$

We have

$$\begin{aligned} E(Q^2|X) &= O(N^{-3}n^{-3})\sum_{i=1}^{k}\sum_{r_1 \ne r_2}^{N}\sum_{r_3 \ne r_4}^{N}\sum_{l_1,l_2,l_3,l_4}^{n_i} E\big(y_{il_1}y_{il_2}y_{il_3}y_{il_4}|X\big) \\ &\quad \times I\big(l_m \in W_{ir_m}, m = 1,2,3,4\big) \\ &\quad + O(N^{-3}n^{-3}) \\ &\quad \times \sum_{i_1 \ne i_2}^{k}\sum_{r_1 \ne r_2}^{N}\sum_{r_3 \ne r_4}^{N}\sum_{l_1,l_2=1}^{n_{i_1}}\sum_{l_3,l_4=1}^{n_{i_2}} E\big(y_{i_1 l_1}y_{i_1 l_2}y_{i_2 l_3}y_{i_2 l_4}|X\big) \\ &\quad \times I\big(l_1 \in W_{i_1 r_1}, l_2 \in W_{i_1 r_2}, l_3 \in W_{i_2 r_3}, l_4 \in W_{i_2 r_4}\big). \end{aligned}$$

Note that by Lemma B.3 in Appendix B, the first term is of order $O(N^{-3}n^{-3})O(N^2n^4) = O(N^{-1}n)$; the order of the second term is also $(N^{-3}n^{-3})O(N^2n^4) = O(N^{-1}n)$. Thus $E(Q^2) \to 0$ by the DCT and $Q \xrightarrow{p} 0$.

By Proposition 2, it suffices to work with $N^{1/2}n^{-1/2}\mathbf{Z}'\mathbf{B}_D\mathbf{Z}$,

$$\mathbf{Z}'\mathbf{B}_D\mathbf{Z} = \frac{1}{Nk(n-1)}\sum_{i=1}^{k}\sum_{r=1}^{N}\sum_{t_1\neq t_2}^{n} Z_{irt_1}Z_{irt_2} + \frac{1}{Nkn}\sum_{i_1\neq i_2}^{k}\sum_{r=1}^{N}\sum_{t_1=1}^{n}\sum_{t_2=1}^{n} Z_{i_1rt_1}Z_{i_2rt_2} = D_1 + D_2,$$

where the definitions of D_i, $i=1,2$, are clear from the context. We can write

$$D_1 = \frac{1}{Nk(n-1)}\sum_{i=1}^{k}\sum_{r=1}^{N}\sum_{l_1\neq l_2}^{n_i} Y_{il_1}Y_{il_2}K_{i,\widehat{G}}(X_{il_1},X_r)K_{i,\widehat{G}}(X_{il_2},X_r)$$

and

$$D_2 = \frac{1}{Nkn}\sum_{i_1\neq i_2}^{k}\sum_{r=1}^{N}\sum_{l_1=1}^{n_{i_1}}\sum_{l_2=1}^{n_{i_2}} Y_{i_1l_1}Y_{i_2l_2} \times K_{i_1,\widehat{G}}(X_{i_1l_1},X_r)K_{i_2,\widehat{G}}(X_{i_2l_2},X_r).$$

Proof of Proposition 3

Let $T^* = D_1^* + D_2^*$, D_i^* is defined as D_i $(i=1,2)$ with the empirical distribution function $\widehat{G}_i$ replaced by the true distribution function G_i. We prove the proposition by verifying the following two expressions:

$$N^{1/2}n^{-1/2}(\mathbf{Z}'\mathbf{B}_D\mathbf{Z} - T^*) \xrightarrow{p} 0 \tag{A.1}$$

and

$$N^{1/2}n^{-1/2}(T^* - T) \xrightarrow{p} 0. \tag{A.2}$$

To prove (A.1), first note that $N^{1/2}n^{-1/2}(D_1 - D_1^*) = D_{11} + D_{12}$, where

$$D_{11} = \frac{N^{1/2}n^{-1/2}}{Nk(n-1)}\sum_{i=1}^{k}\sum_{r=1}^{N}\sum_{l_1\neq l_2}^{n_i} Y_{il_1}Y_{il_2} \times K_{i,\widehat{G}}(X_{il_2},X_r)\big[K_{i,\widehat{G}}(X_{il_1},X_r) - K_{i,G}(X_{il_1},X_r)\big]$$

and

$$D_{12} = \frac{N^{1/2}n^{-1/2}}{Nk(n-1)}\sum_{i=1}^{k}\sum_{r=1}^{N}\sum_{l_1\neq l_2}^{n_i} Y_{il_1}Y_{il_2} \times K_{i,G}(X_{il_1},X_r)\big[K_{i,\widehat{G}}(X_{il_2},X_r) - K_{i,G}(X_{il_2},X_r)\big].$$

It is obvious that $E(D_{11}) = 0$. After some algebra,

$$\begin{aligned} E(D_{11}^2|X) &= \frac{1}{Nk^2(n-1)^2n}\sum_{i=1}^{k}\sum_{r_1=1}^{N}\sum_{r_2=1}^{N}\sum_{l_1\neq l_2}^{n_i}\sum_{l_3\neq l_4}^{n_i} E(y_{il_1}y_{il_2}y_{il_3}y_{il_4}|X) \\ &\quad\times \big[K_{i,\widehat{G}}(X_{il_1},X_{r_1}) - K_{i,G}(X_{il_1},X_{r_1})\big]K_{i,\widehat{G}}(X_{il_2},X_{r_1}) \\ &\quad\times \big[K_{i,\widehat{G}}(X_{il_3},X_{r_2}) - K_{i,G}(X_{il_3},X_{r_2})\big]K_{i,\widehat{G}}(X_{il_4},X_{r_2}) \\ &= E_1 + E_2. \end{aligned}$$

The expectation is zero unless $l_1 = l_3$ and $l_2 = l_4$, or $l_1 = l_4$ and $l_2 = l_3$, represent the sum pertaining to the former case by E_1 and that pertaining to the latter case by E_2,

$$\begin{aligned} E_1 &= \frac{1}{Nk^2(n-1)^2n}\sum_{i=1}^{k}\sum_{r_1=1}^{N}\sum_{r_2=1}^{N}\sum_{l_1\neq l_2}^{n_i} E(y_{il_1}^2y_{il_2}^2|X) \\ &\quad\times \big[K_{i,\widehat{G}}(X_{il_1},X_{r_1}) - K_{i,G}(X_{il_1},X_{r_1})\big]K_{i,\widehat{G}}(X_{il_2},X_{r_1}) \\ &\quad\times \big[K_{i,\widehat{G}}(X_{il_1},X_{r_2}) - K_{i,G}(X_{il_1},X_{r_2})\big]K_{i,\widehat{G}}(X_{il_2},X_{r_2}). \end{aligned}$$

By Lemma B.4, we have

$$\begin{aligned} |E_1| &\le [Nk^2(n-1)^2n]^{-1}O(Nn)O\big(n^{1/2}(\log(Nn^{-1}))^{1/2}\big)^2 \\ &= O\big(n^{-1}\log(Nn^{-1})\big). \end{aligned}$$

Similarly, the order of $|E_2|$ is bounded by $O(n^{-1}\log(Nn^{-1}))$. Thus $E(D_{11}^2) \to 0$ by dominated convergence theorem and $D_{11} \xrightarrow{p} 0$. Similarly, we can show that $D_{12} \xrightarrow{p} 0$. We can prove $N^{1/2}n^{-1/2}(D_2 - D_2^*) = o_p(1)$ exactly the same way. Therefore, we verify (A.1). To prove (A.2), write $N^{1/2}n^{-1/2}(T^* - T) = A_1 + A_2$, where

$$A_1 = \frac{N^{1/2}n^{-1/2}}{k(n-1)}\sum_{i=1}^{k}\sum_{l_1\neq l_2}^{n_i} y_{il_1}y_{il_2} \times \int K_{i,G}(x_{il_1},x)K_{i,G}(x_{il_2},x)\,d\big(\widehat{G}(x) - G(x)\big)$$

and

$$A_2 = \frac{N^{1/2}n^{-1/2}}{kn}\sum_{i_1\neq i_2}^{k}\sum_{l_1=1}^{n_{i_1}}\sum_{l_2=1}^{n_{i_2}} y_{i_1l_1}y_{i_2l_2} \times \int K_{i_1,G}(x_{i_1l_1},x)K_{i_2,G}(x_{i_2l_2},x)\,d\big(\widehat{G}(x) - G(x)\big).$$

It is obvious that $E(A_1) = 0$,

$$E(A_1^2|X) = \frac{2N}{k^2n(n-1)^2}\sum_{i=1}^{k}\sum_{l_1\neq l_2}^{n_i}\sigma_i^2(X_{il_1})\sigma_i^2(X_{il_2}) \times \left[\int K_{i,G}(x_{il_1},x)K_{i,G}(x_{il_2},x)\,d\big(\widehat{G}(x) - G(x)\big)\right]^2.$$

The integrand will be 0 unless $G_i(x_{il_m}) - h_i \le G_i(x) \le G_i(x_{il_m}) + h_i$, $m = 1,2$; thus if the intersection of $[G_i(x_{il_1}) - h_i, G_i(x_{il_1}) + h_i]$ and $[G_i(x_{il_2}) - h_i, G_i(x_{il_2}) + h_i]$ is not empty, then we must have $|x_{il_1} - x_{il_2}| \le ch_i$ for some positive constant c, then

$$\begin{aligned} &\int K_{i,G}(x_{il_1},x)K_{i,G}(x_{il_2},x)\,d\big(\widehat{G}(x) - G(x)\big) \\ &\le |\widehat{G}(b) - G(b) - \widehat{G}(a) - G(a)| \quad (\text{where } |a-b| \le ch) \\ &\le KN^{-1}n^{1/2}(\log(Nn^{-1}))^{1/2} \end{aligned}$$

for some positive constant K almost surely for N large enough by (B.2) in Appendix B. By the uniform boundedness of $\sigma_i^2(\cdot)$,

$$\begin{aligned} E(A_1^2|X) &\le C_2\frac{N}{n^3}Nn\big(N^{-1}n^{1/2}(\log(Nn^{-1}))^{1/2}\big)^2 \\ &= C_2n^{-1}\log(Nn^{-1}), \end{aligned}$$

for some positive constant C_2. Thus $E(A_1^2) \to 0$ by DCT, so $A_1 \xrightarrow{p} 0$. Similarly, we can show that $A_2 \xrightarrow{p} 0$.

Proof of Theorem 1

Because T_1 and T_2 are uncorrelated, the asymptotic variance of $N^{1/2}n^{-1/2}T$ can be obtained from Lemma B.6. To prove the asymptotic normality, we will make use of proposition 3.2 of de Jong (1987). For ease of presentation, we restrict ourselves to the case with $k = 2$. The case with $k \ge 3$ follows by the same argument with an additional amount of algebra and notation. Let $\mathbf{U} = (\mathbf{U}_1,\ldots,\mathbf{U}_{n_1},\mathbf{U}_{n_1+1},\ldots,\mathbf{U}_{n_1+n_2})'$, with $\mathbf{U}_i = (X_{1i},Y_{1i})$ if $1 \le i \le n_1$, and $\mathbf{U}_i = (X_{2,i-n_1},$

$Y_{2,i-n_1})$ if $n_1+1\le i\le n_1+n_2$. Thus the components of $\mathbf{U}$ are independent. Let $W^*_{ii}=0$, $i=1,\dots,n_1+n_2$, and for $i\neq j$, let W^*_{ij} be

$$\frac{N^{1/2}Y_{1i}Y_{1j}}{k(n-1)n^{1/2}}\int K_{1,G}(X_{1i},x)K_{1,G}(X_{1j},x)g(x)\,dx \quad \text{if } 1\le i,j\le n_1,$$

$$\frac{N^{1/2}Y_{1i}Y_{2,j-n_1}}{k(n-1)n^{1/2}}\int K_{1,G}(X_{1i},x)K_{2,G}\big(X_{2,j-n_1},x\big)g(x)\,dx \quad \text{if } 1\le i\le n_1<j\le n_1+n_2,$$

$$\frac{N^{1/2}Y_{2,i-n_1}Y_{1j}}{k(n-1)n^{1/2}}\int K_{2,G}\big(X_{2,i-n_1},x\big)K_{1,G}(X_{1j},x)g(x)\,dx \quad \text{if } 1\le j\le n_1<i\le n_1+n_2,$$

and

$$\frac{N^{1/2}Y_{2,i-n_1}Y_{2,j-n_1}}{k(n-1)n^{1/2}}\int K_{2,G}\big(X_{2,i-n_1},x\big)K_{2,G}\big(X_{2,j-n_1},x\big)g(x)\,dx \quad \text{if } n_1+1\le i,j\le n_1+n_2.$$

Then $N^{1/2}n^{-1/2}T=\sum_{1\le i\le n_1+n_2}\sum_{1\le j\le n_1+n_2}W^*_{ij}$ is a generalized quadratic form (de Jong 1987). It is easy to check that $W^*_{ij}=W^*_{ji}$; let $W_{ij}=W^*_{ij}+W^*_{ji}$, then $N^{1/2}n^{-1/2}T=\sum_{1\le i<j\le n_1+n_2}W_{ij}$. To apply proposition 3.2 of de Jong (1987), following his notations, we need to check the following five conditions: (1) $N^{1/2}n^{-1/2}T$ is clean (see de Jong 1987 for the definition); (2) $\mathrm{var}(N^{1/2}n^{-1/2}T)\to\tau^2$, some positive constant; (3) G_I is of smaller order than $\mathrm{var}(N^{1/2}n^{-1/2}T)$; (4) G_{II} is of smaller order than $\mathrm{var}(N^{1/2}n^{-1/2}T)$; and (5) G_{IV} is of smaller order than $\mathrm{var}(N^{1/2}n^{-1/2}T)$, where

$$G_I=\sum_{1\le i<j\le n_1+n_2}E(W^4_{ij}),$$

$$G_{II}=\sum_{1\le i<j<k\le n_1+n_2}\big\{E(W^2_{ij}W^2_{ik})+E(W^2_{ji}W^2_{jk})+E(W^2_{ki}W^2_{kj})\big\},$$

and

$$G_{IV}=\sum_{1\le i<j<k<l\le n_1+n_2}\big\{E(W_{ij}W_{ik}W_{lj}W_{lk})+E(W_{ij}W_{il}W_{kj}W_{kl})+E(W_{ik}W_{il}W_{jk}W_{jl})\big\}.$$

Because $E(W_{ij}|\mathbf{U}_i)=E(W_{ij}|\mathbf{U}_j)=0$, condition (1) is satisfied by the definition of de Jong (1987). We have proved condition (2). It remains to show that $G_I=o(1)$, $G_{II}=o(1)$, and $G_{IV}=o(1)$. We first show $G_I=o(1)$:

$$G_I=\sum_{1\le i<j\le n_1}E(W^4_{ij})+\sum_{n_1+1\le i<j\le n_1+n_2}E(W^4_{ij})+\sum_{1\le i\le n_1<j\le n_1+n_2}E(W^4_{ij}).$$

Look at the case $1\le i<j\le n_1$,

$$W_{ij}=\frac{2N^{1/2}}{k(n-1)n^{1/2}}Y_{1i}Y_{1j}\int K_{1,G}(X_{1i},x)K_{1,G}(X_{1j},x)g(x)\,dx.$$

We have

$$E(W^4_{ij}|\mathbf{X})=\frac{16N^2}{k^4(n-1)^4n^2}E(Y^4_{1i}|\mathbf{X})E(Y^4_{1j}|\mathbf{X})\times\left[\int K_{1,G}(X_{1i},x)K_{1,G}(X_{1j},x)g(x)\,dx\right]^4.$$

Consider the integral. Let $y=\frac{G_1(X_{1i})-G_1(x)}{h_1}$; then it equals

$$\int K(y)K\left(\frac{G_1(X_{1i})-G_1(X_{1j})}{h_1}-y\right)\times\frac{-h_1g(G_1^{-1}(G_1(X_{1i})-yh_1))\,dy}{g_1(G_1^{-1}(G_1(X_{1i})-yh_1))}=O(h_1).$$

Thus $E(W^4_{ij}|\mathbf{X})=O(N^2n^{-6})O(h_1^4)=O(N^{-2}n^{-2})$. The boundedness of $\mathbf{X}$ yields $E(W^4_{ij})=O(N^{-2}n^{-2})$ for the case where $1\le i<j\le n_1$. Similarly, we may prove that $E(W^4_{ij})=O(N^{-2}n^{-2})$ for the other two cases. Notice that the order of number of indices satisfying $1\le i<j\le n_1+n_2$ is actually $O(Nn)$ for $E(W^4_{ij})$ to be nonzero; thus $G_I=O(Nn)O(N^{-2}n^{-2})=O(N^{-1}n^{-1})=o(1)$. Applying Hölder's inequality, we can similarly show that $G_{II}=o(1)$ and $G_{IV}=o(1)$.

Proof of Lemma 1

As before, let $\mathbf{Z}$ be the vector denoting observations in the hypothetical ANOVA constructed from (X_{ij},Y_{ij}), and let $\mathbf{g}$ represent another vector denoting observations in a hypothetical two-way ANOVA constructed from $(X_{ij},(Nn)^{-1/4}\int y\,dR_{iX_{ij}}(y))$, $i=1,\dots,k$, $j=1,\dots,n_i$. The lemma is proven if we can show the following three steps:

$$N^{1/2}n^{-1/2}\big[\mathbf{Z}'\mathbf{A}_D\mathbf{Z}-\big(\mathbf{Z}-E_N(\mathbf{Z}|X)+\mathbf{g}\big)'\times\mathbf{A}_D\big(\mathbf{Z}-E_N(\mathbf{Z}|X)+\mathbf{g}\big)\big]\xrightarrow{p}0, \tag{A.3}$$

$$N^{1/2}n^{-1/2}\big[\big(\mathbf{Z}-E_N(\mathbf{Z}|X)\big)'\mathbf{A}_D\big(\mathbf{Z}-E_N(\mathbf{Z}|X)\big)-\big(\mathbf{Z}-E_N(\mathbf{Z}|X)+g\big)'\mathbf{A}_D\big(\mathbf{Z}-E_N(\mathbf{Z}|X)+\mathbf{g}\big)+\mathbf{g}'\mathbf{A}_D\mathbf{g}\big]\xrightarrow{p}0, \tag{A.4}$$

and

$$N^{1/2}n^{-1/2}\mathbf{g}'\mathbf{A}_D\mathbf{g}\xrightarrow{p}k\theta_D. \tag{A.5}$$

To prove (A.3), notice that

$$\mathbf{Z}'\mathbf{A}_D\mathbf{Z}-\big(\mathbf{Z}-E_N(\mathbf{Z}|X)+\mathbf{g}\big)'\mathbf{A}_D\big(\mathbf{Z}-E_N(\mathbf{Z}|X)+\mathbf{g}\big)=2\mathbf{Z}'\mathbf{A}_D\big(E_N(\mathbf{Z}|X)-\mathbf{g}\big)-\big(E_N(\mathbf{Z}|X)-\mathbf{g}\big)'\mathbf{A}_D\big(E_N(\mathbf{Z}|X)-\mathbf{g}\big).$$

Similar to the proof of Proposition 1, we can show that $N^{1/2}n^{-1/2}\times(E_N(\mathbf{Z}|X)-\mathbf{g})'\mathbf{A}_D(E_N(\mathbf{Z}|X)-\mathbf{g})=O_p(n^{5/2}N^{3/2})=o_p(1)$ and $N^{1/2}n^{-1/2}\mathbf{Z}'\mathbf{A}_D(E_N(\mathbf{Z}|X)-\mathbf{g})=o_p(1)$. The left side of (A.4) is equal to $-2N^{1/2}n^{-1/2}(\mathbf{Z}-E_N(\mathbf{Z}|X))'\mathbf{A}_D\mathbf{g}$, which can be shown to be $o_p(1)$ as in the proof of Proposition 1. As for (A.5), let g_{irt} be the elements of $\mathbf{g}$, then

$$\begin{aligned}
&N^{1/2}n^{-1/2}\mathbf{g}'\mathbf{A}_D\mathbf{g}\\
&=N^{1/2}n^{-1/2}\left[\frac{kn}{N-1}\sum_{r=1}^N(\bar g_{\cdot r\cdot}-\bar g_{\dots})^2-\frac{1}{Nk(n-1)}\sum_{i=1}^k\sum_{r=1}^N\sum_{t=1}^n(g_{irt}-\bar g_{ir\cdot})^2\right]\\
&=N^{1/2}n^{-1/2}\frac{kn}{N-1}(Nn)^{-1/2}\times\sum_{r=1}^N\left[\frac{1}{k}\sum_{i=1}^k\int y\,dR_{ix_r}(y)-\frac{1}{Nk}\sum_{r'=1}^N\sum_{i=1}^k\int y\,dR_{ix_{r'}}(y)\right]^2\\
&\quad+O_p\big(n^{3/2}N^{-3/2}\big)\\
&\xrightarrow{p}k\theta_D.
\end{aligned}$$

Proof of Theorem 3

Similar to the proof of Theorem 1, we can show that

$$N^{1/2}n^{-1/2}\big(\mathbf{Z}-E_N(\mathbf{Z}|X)\big)'\mathbf{A}_D\big(\mathbf{Z}-E_N(\mathbf{Z}|X)\big) \to \mathrm{N}\bigg(0, \frac{4}{3k^2}(\xi_1^4+\eta_1^4)\bigg),$$

where ξ_1^4 and η_1^4 are defined as in ξ and η but with $\sigma_i^2(t)$ replaced by $\sigma_{N,i}^2(t)$, the conditional variance function calculated from the conditional distribution function (11). It is straightforward to show that $\sigma_{N,i}^2(t)=\sigma_i^2(t)+O(N^{-1/2})$. The proof is finished by combining the results of Theorem 1 and Lemma 1.

APPENDIX B: SOME AUXILIARY RESULTS

Lemma B.1. Let $h(\cdot)$ be any Lipschitz continuous function. Then for any cell (i,r) in this hypothetical two-way ANOVA, there exists some $C>0$, for N large enough, such that

$$\left|\frac{1}{n}\sum_{j=1}^{n_i}h(X_{ij})I(j\in W_{ir})-h(X_{im})\right|\le CnN^{-1},$$

$$\forall X_{im}\in W_{ir}, 1\le i\le k, 1\le r\le N.$$

Proof. The proof is rather straightforward and thus is omitted.

Lemma B.2. Assume Assumptions A1 and A2. Then under $H_0(D)$, for N large enough, there exists some $C>0$ such that for N large enough, $\forall\, i,r,t$,

$$\big|E(Z_{irt}-\overline{Z}_{ir\cdot}|X)\big|\le CnN^{-1}, \qquad \big|E(\overline{Z}_{\cdot r\cdot}-\overline{Z}_{\cdots}|X)\big|\le CnN^{-1}.$$

Proof. Let x_{ij} be the covariate corresponding to Z_{irt} in the hypothetical two-way ANOVA (which is the response $Y_{ij}=y_{ij}$ in the truly observed ANCOVA model, remember the dual notations). Apply Lemma B.1 [let $h(x)=\int y\,dF_{ix}(y)$], for N large enough, there exists some $C>0$ such that $\forall\, i,r,t$,

$$\begin{aligned}\big|E(Z_{irt}-\overline{Z}_{ir\cdot}|X)\big| &= \left|\int y\,d\left[F_{ix_{ij}}(y)-\frac{1}{n}\sum_{l=1}^{n_i}F_i(y|x_{il})I(l\in W_{ir})\right]\right|\\ &\le CnN^{-1}.\end{aligned} \tag{B.1}$$

Relationship (B.1) ensures that $\forall\, i,r$, $|\int y\,d\overline{F}_{ir\cdot}(y)-\int y\,dF_{ix_r}(y)|\le CnN^{-1}$, where x_r is the center of window W_{ir}, $\overline{F}_{ir\cdot}(y)=\frac{1}{n}\times\sum_{l=1}^{n_i}F_{ix_{il}}(y)I(l\in W_{ir})$. Under the null hypothesis, $\sum_{i=1}^{k}F_{ix}(y)=M(y)$, $\forall x, \forall y$; therefore,

$$\begin{aligned}&\big|E(\overline{Z}_{\cdot r\cdot}|X)-E(\overline{Z}_{\cdots}|X)\big|\\ &\quad=\left|\int y\,d\left[\frac{1}{k}\sum_{i=1}^{k}\overline{F}_{ir\cdot}(y)-\frac{1}{Nk}\sum_{i=1}^{k}\sum_{r=1}^{N}\overline{F}_{ir\cdot}(y)\right]\right|\\ &\quad\le 2CnN^{-1}.\end{aligned}$$

Lemma B.3. $\forall X_1,X_2$, if $|\widehat{G}_i(X_1)-\widehat{G}_i(X_2)|=O_p(h_i)$, where $h_i=\frac{n-1}{2n_i}$. Then there exists some $C>0$, for N large enough,

$$|\widehat{G}(X_1)-\widehat{G}(X_2)|\le Ch_i,$$

where $\widehat{G}_i$ is the empirical distribution of the covariate in the ith group and $\widehat{G}$ is the overall empirical distribution of the covariate.

Proof. By the main result stated by Pyke (1965, sec. 2.1), we have $|X_1-X_2|=O_p(h_i)$; thus we immediately have $|G_i(X_1)-G_i(X_2)|=O_p(h_i)$. Applying the mean value theorem to $G_i(X_1)-G_i(X_2)$ and $G(X_1)-G(X_2)$ and combining the results, we have, for some X_1^*, $X_1^{**}\in[X_1,X_2]$,

$$|G(X_1)-G(X_2)|=\frac{g(X_1^*)}{g_i(X_1^{**})}|G_i(X_1)-G_i(X_2)|\le c_1|G_i(X_1)-G_i(X_2)|,$$

where the positive constant c_1 can be taken independently of X_1 and X_2 by the boundedness assumption on the density. Therefore, we have $|G(X_1)-G(X_2)|=O_p(h_i)$. By theorem 2.11 of Stute (1982), we have that for N large enough, for some positive constant c_2,

$$\begin{aligned}&|\widehat{G}(X_1)-\widehat{G}(X_2)|\\ &\quad\le|G(X_1)-G(X_2)|+\big|\widehat{G}(X_1)-\widehat{G}(X_2)-G(X_1)+G(X_2)\big|\\ &\quad=O_p(h_i)+c_2N^{-1}n^{1/2}(\log(Nn^{-1}))^{1/2}\\ &\quad=O_p(h_i).\end{aligned}$$

Remark. An immediate consequence of this lemma is that for any observation in the one-way ANCOVA, it belongs to at most $O(n)$ cells in the hypothetical two-way ANOVA.

Lemma B.4. For N large enough, there exists some $C>0$ such that $\forall\, i, X$,

$$\sum_{r=1}^{N}|K_{i,\widehat{G}}(X,X_r)-K_{i,G}(X,X_r)|\le Cn^{1/2}(\log(n_in^{-1}))^{1/2}$$

in probability, where X_r, $r=1,\dots,N$, are the centers of cells in the hypothetical two-way ANOVA.

Proof. By theorem 2.11 of Stute (1982), there exists $K_1>0$, such that for N large enough,

$$\begin{aligned}&\big|\widehat{G}_i(X)-\widehat{G}_i(X_r)-G_i(X)+G_i(X_r)\big|\\ &\qquad\le K_1n_i^{-1}n^{1/2}(\log(n_in^{-1}))^{1/2}\quad\text{a.s.,}\end{aligned} \tag{B.2}$$

$K_{i,\widehat{G}}(X,X_r)-K_{i,G}(X,X_r)\ne 0$ only if (1) $|G_i(X)-G_i(X_r)|\le h_i$ but $|\widehat{G}_i(X)-\widehat{G}_i(X_r)|>h_i$ or (2) $|G_i(X)-G_i(X_r)|>h_i$ but $|\widehat{G}_i(X)-\widehat{G}_i(X_r)|\le h_i$. In case (1) we thus have, for N large enough,

$$h_i-K_1n_i^{-1}n^{1/2}(\log(n_in^{-1}))^{1/2}<|G_i(X)-G_i(X_r)|\le h_i. \tag{B.3}$$

After performing second-order Taylor expansion for $G_i(x)-G_i(x_r)$ and $G(x)-G(x_r)$, then combining the results, we have

$$G(x)-G(x_r)=\frac{g(x_r)}{g_i(x_r)}\big(G_i(x)-G_i(x_r)\big)+\left[\frac{g'(x_r^{**})}{2}-\frac{g_i'(x_r^*)g(x_r)}{2g_i(x_r)}\right](x-x_r)^2,$$

where x_r^* and x_r^{**} are between x and x_r. If $|G_i(x)-G_i(x_r)|\le h_i$, then $|x-x_r|\le\sup_p|(G_i^{-1})'(p)|\cdot|G_i(x)-G_i(x_r)|=c\frac{n}{2n_i}$ for some constant c, and we have

$$|G(X)-G(X_r)|=c_1|G_i(X_{ij})-G_i(X_r)|+O(n^2n_i^{-2})$$

for some positive constant c_1. Combined with (B.3), we thus have

$$\begin{aligned}&c_1h_i-c_1K_1n_i^{-1}n^{1/2}(\log(n_in^{-1}))^{1/2}+O(n^2n_i^{-2})\\ &\quad<|G(X)-G(X_r)|\\ &\quad\le c_1h_i+O(n^2n_i^{-2}).\end{aligned}$$

Applying theorem 2.11 of Stute (1982) again, we have

$$\begin{aligned}&c_1h_i-c_2K_1n_i^{-1}n^{1/2}(\log(n_in^{-1}))^{1/2}+O(n^2n_i^{-2})\\ &\quad<|\widehat{G}(X)-\widehat{G}(X_r)|\\ &\quad\le c_1h_i+c_3K_1n_i^{-1}n^{1/2}(\log(n_in^{-1}))^{1/2}+O(n^2n_i^{-2})\end{aligned} \tag{B.4}$$

for some positive constants c_2 and c_3. We have similar inequalities for case (2). As a result,

$$\sum_{r=1}^{N} |K_{i,\widehat{G}}(X, X_r) - K_{i,G}(X, X_r)| \le NO\big(n_i^{-1} n^{1/2} (\log(n_i n^{-1}))^{1/2}\big)$$
$$\le c^* n^{1/2} (\log(n_i n^{-1}))^{1/2}$$

for some positive constant c^*.

Lemma B.5. For any positive constant C, we have

$$\int \left[\int K(y)K(z - Cy)\,dy\right]^2 dz = \begin{cases} 8\left(1 - \dfrac{C}{3}\right), & 0 < C \le 1 \\ 8\left(\dfrac{1}{C} - \dfrac{1}{3C^2}\right), & C > 1. \end{cases}$$

Proof. Direct integration.

Lemma B.6. Let ξ^4 and η^4 be defined as in Theorem 1. Then

$$\operatorname{var}\big(N^{1/2} n^{-1/2} T_1\big) \to \frac{4\xi^4}{3k^2} \qquad \text{and} \qquad \operatorname{var}\big(N^{1/2} n^{-1/2} T_2\big) \to \frac{4\eta^4}{3k^2}.$$

Proof.

$$T_1 = \frac{1}{k(n-1)} \sum_{i=1}^{k} \sum_{l_1 \ne l_2}^{n_i} y_{il_1} y_{il_2} \int K_{i,G}\big(X_{il_1}, x\big) K_{i,G}\big(X_{il_2}, x\big)\,dG(x).$$

It is obvious that $E(T_1) = 0$,

$$E(Nn^{-1} T_1^2 | X) = \frac{2N}{k^2 (n-1)^2 n} \sum_{i=1}^{k} \sum_{l_1 \ne l_2}^{n_i} \sigma_i^2\big(X_{il_1}\big) \sigma_i^2\big(X_{il_2}\big)$$
$$\times \left[\int K_{i,G}\big(X_{il_1}, x\big) K_{i,G}\big(X_{il_2}, x\big)\,dG(x)\right]^2.$$

Thus

$$\operatorname{var}\big(N^{1/2} n^{-1/2} T_1\big)$$
$$= \frac{2N}{k^2 (n-1)^2 n} \sum_{i=1}^{k} n_i (n_i - 1) \int\int \sigma_i^2(u) \sigma_i^2(v)$$
$$\times \left[\int K_{i,G}(u, x) K_{i,G}(v, x)\,dG(x)\right]^2 g_i(u) g_i(v)\,du\,dv.$$

Fix u and v, and let $y = \frac{G_i(u) - G_i(x)}{h_i}$; then $x = G_i^{-1}(G_i(u) - y h_i)$, $dx = -h_i \frac{dy}{g_i(G_i^{-1}(G_i(u) - y h_i))}$, and

$$\operatorname{var}\big(N^{1/2} n^{-1/2} T_1\big) = \frac{2N}{k^2 (n-1)^2 n} \sum_{i=1}^{k} n_i (n_i - 1) \int\int \sigma_i^2(u) \sigma_i^2(v)$$
$$\times \left[\int K(y) K\left(\frac{G_i(u) - G_i(v)}{h_i} - y\right)\right.$$
$$\left.\times \frac{-h_i g(G_i^{-1}(G_i(u) - y h_i))}{g_i(G_i^{-1}(G_i(u) - y h_i))}\,dy\right]^2$$
$$\times g_i(u) g_i(v)\,du\,dv$$
$$= \frac{2N}{k^2 (n-1)^2 n} \sum_{i=1}^{k} n_i (n_i - 1) h_i^2 \int\int \sigma_i^2(u) \sigma_i^2(v)$$
$$\times \left[\int K(y) K\left(\frac{G_i(u) - G_i(v)}{h_i} - y\right) dy\right]^2$$
$$\times \frac{g^2(u) g_i(v)}{g_i(u)}\,du\,dv + o(1).$$

Fix v, and let $z = \frac{G_i(u) - G_i(v)}{h_i}$; then $u = G_i^{-1}(G_i(v) + z h_i)$, $du = h_i \frac{dz}{g_i(G_i^{-1}(G_i(v) + z h_i))}$, and

$$\operatorname{var}\big(N^{1/2} n^{-1/2} T_1\big) = \frac{2N}{k^2 (n-1)^2 n} \sum_{i=1}^{k} n_i (n_i - 1) h_i^3 \int\int \sigma_i^4(v)$$
$$\times \left[\int K(y) K(z - y)\,dy\right]^2 \frac{g^2(v)}{g_i(v)}\,dz\,dv + o(1).$$

Apply Lemma B.5 to prove the first part of the lemma. The second part can be verified similarly.

Lemma B.7. If Assumptions A1 and A2 hold, then, under the local alternative sequence (11), there exists some $C > 0$ such that, for N large enough, $\forall\, r, i, t$,

$$\big|E_N(Z_{rit} - \overline{Z}_{ri\cdot} | X)\big| \le CnN^{-1},$$

$$\Bigg|E_N(\overline{Z}_{r\cdot\cdot} - \overline{Z}_{\cdot\cdot\cdot} | X)$$
$$- (Nn)^{-1/4} \left[\frac{1}{k} \sum_{i=1}^{k} \int y\,dR_{ix_r}(y) - \frac{1}{Nk} \sum_{r=1}^{N} \sum_{i=1}^{k} \int y\,dR_{ix_r}(y)\right]\Bigg|$$
$$\le CnN^{-1}.$$

Proof. This follows by inspecting the proof of Lemma B.2 and noting that now,

$$E_N(Z_{rit} | X) = \int y\,dF_{ix_{ij}}(y) + (Nn)^{-1/4} \int y\,dR_{ix_{ij}}(y).$$

[Received October 2004. Revised August 2005.]

REFERENCES

Akritas, M. G., and Arnold, S. (2000), "Asymptotics for ANOVA When the Number of Levels Is Large," *Journal of the American Statistical Association*, 95, 212–226.

Akritas, M. G., Arnold, S., and Du, Y. (2000), "Nonparametric Models and Methods for Nonlinear Analysis of Covariance," *Biometrika*, 87, 507–526.

Akritas, M. G., and Papadatos, N. (2004), "Heteroscedastic One-Way ANOVA and Lack-of-Fit Test," *Journal of the American Statistical Association*, 99, 368–382.

Arnold, S. (1980), "Asymptotic Validity of F Tests for the Ordinary Linear Model and the Multiple Correlation Model," *Journal of the American Statistical Association*, 75, 890–894.

Boos, D. D., and Brownie, C. (1995), "ANOVA and Rank Tests When the Number of Treatments Is Large," *Statistics & Probability Letters*, 23, 183–191.

de Jong, P. (1987), "A Central Limit Theorem for Generalized Quadratic Forms," *Probability Theory and Related Fields*, 75, 261–277.

Delgado, M. A. (1993), "Testing the Equality of Nonparametric Regression Curves," *Statistics & Probability Letters*, 17, 199–204.

Dette, H., and Munk, A. (1998), "Nonparametric Comparison of Several Regression Functions: Exact and Asymptotic Theory," *The Annals of Statistics*, 26, 2339–2368.

Dette, H., and Neumeyer, N. (2001), "Nonparametric Analysis of Covariance," *The Annals of Statistics*, 29, 1361–1400.

Eubank, R. L. (2000), "Testing for No Effect by Cosine Series Methods," *Scandinavian Journal of Statistics*, 27, 747–763.

Gassaman, M. P. (1970), "A Consistent Nonparametric Multivariate Density Estimator Based on Statistical Equivalent Blocks," *The Annals of Mathematical Statistics*, 41, 1344–1346.

Hall, P., and Hart, J. D. (1990), "Bootstrap Tests for Difference Between Means in Nonparametric Regression," *Journal of the American Statistical Association*, 85, 1039–1049.

Hall, P., Huber, C., and Speckman, P. L. (1997), "Covariate-Matched One-Sided Tests for the Difference Between Functional Means," *Journal of the American Statistical Association*, 92, 1074–1083.

Härdle, W., and Marron, J. S. (1990), "Semiparametric Comparison of Regression Curves," *The Annals of Statistics*, 18, 63–89.

Hastie, T. J., and Tibshirani, R. J. (1990), *Generalized Additive Models*, New York: Chapman & Hall.

——— (1993), "Varying-Coefficient Models" (with discussion), *Journal of the Royal Statistical Society*, Ser. B, 55, 757–796.

Hettmansperger, T. P., and McKean, J. W. (1998), *Robust Nonparametric Statistical Methods*, London: Arnold.

Holliday, R. (1960), "Plant Population and Crop Yield," *Field Crop Abstracts*, 13, 159–167, 247–254.

Jiang, J. (1996), "REML Estimation: Asymptotic Behavior and Related Topics," *The Annals of Statistics*, 24, 255–286.

King, E. C., Hart, J. D., and Wehrly, T. E. (1991), "Testing the Equality of Regression Curves Using Linear Smoothers," *Statistics & Probability Letters*, 12, 239–247.

Koul, H. L., and Schick, A. (2003), "Testing for Superiority Among Two Regression Curves," *Journal of Statistical Planning and Inference*, 117, 15–33.

Kulasekera, K. B. (1995), "Comparison of Regression Curves Using Quasi-Residuals," *Journal of the American Statistical Association*, 90, 1085–1093.

Lugosi, G., and Nobel, A. (1996), "Consistency of Data-Driven Histogram Methods for Density Estimation and Classification," *The Annals of Statistics*, 24, 687–706.

McKean, J. W., and Schrader, R. M. (1980), "The Geometry of Robust Procedures in Linear Models," *Journal of the Royal Statistical Society*, Ser. B, 42, 366–371.

Pyke, R. (1965), "Spacings" (with discussion), *Journal of the Royal Statistical Society*, Ser. B, 27, 395–449.

Quade, D. (1967), "Rank Analysis of Covariance," *Journal of the American Statistical Association*, 62, 1187–1200.

Ratkowsky, D. A. (1983), *Nonlinear Regression Modelling: A Unified Practical Approach*, New York: Marcel Dekker.

Ruppert, D., Sheather, S. J., and Wand, M. P. (1995), "An Effective Bandwidth Selector for Local Least Squares Regression," *Journal of the American Statistical Association*, 90, 1257–1270.

Ruppert, D., Wand, M. P., Holst, U., and Hossjer, O. (1997), "Local Polynomial Variance Function Estimation," *Technometrics*, 39, 262–273.

Sen, P. (1996), "Regression Rank Scores Estimations in ANOCOVA," *The Annals of Statistics*, 24, 1586–1601.

Speckman, P. (1988), "Kernel Smoothing in Partial Linear Models," *Journal of the Royal Statistical Society*, Ser. B, 50, 413–436.

Stute, W. (1982), "The Oscillation Behavior of Empirical Process," *The Annals of Probability*, 10, 86–107.

Terpstra, J., and McKean, J. (2004), "Rank-Based Analyses of Linear Models Using R," technical report, Western Michigan University, Dept. of Statistics.

Wang, L., and Akritas, M. G. (2002), "Two-Way Heteroscedastic ANOVA When the Number of Levels Is Large," *Statistica Sinica*, to appear.

Young, S., and Bowman, A. (1995), "Non-Parametric Analysis of Covariance," *Biometrics*, 51, 920–931.