Journal of Multivariate Analysis 101 (2010) 1791–1805

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

Rank test for heteroscedastic functional data

Haiyan Wang [a,*], Michael G. Akritas [b]

[a] Department of Statistics, Kansas State University, Manhattan, KS, United States
[b] Department of Statistics, Pennsylvania State University, United States

ARTICLE INFO

Article history:
Received 23 April 2009
Available online 31 March 2010

AMS subject classifications:
primary 62G10
62G35

Keywords:
Repeated measures
Nonparametric inference
Hypothesis testing
High-dimensional multivariate analysis

ABSTRACT

In this paper, we consider (mid-)rank based inferences for testing hypotheses in a fully nonparametric marginal model for heteroscedastic functional data that contain a large number of within subject measurements from possibly only a limited number of subjects. The effects of several crossed factors and their interactions with time are considered. The results are obtained by establishing asymptotic equivalence between the rank statistics and their asymptotic rank transforms. The inference holds under the assumption of α-mixing without moment assumptions. As a result, the proposed tests are applicable to data from heavy-tailed or skewed distributions, including both continuous and ordered categorical responses. Simulation results and a real application confirm that the (mid-) rank procedures provide both robustness and increased power over the methods based on original observations for non-normally distributed data.

© 2010 Elsevier Inc. All rights reserved.

1. Introduction

Functional data in which a large number of within cluster measurements are collected temporally or spatially are becoming more common as technology advances. An example is the tiling microarrays which use probes with partially overlapping sequences to cover the entire genome. Another example comes from fast functional magnetic resonance imaging data (fMRI) which contain measurements from the brain recorded at a time scale of seconds. We are interested in the effect of several baseline or time-independent factors and their interactions with time. For the fMRI example, such effects can help us to diagnose and determine how a normal, diseased or injured brain functions and to assess the potential risks of invasive treatments of the brain.

Extensive studies have been done for functional data from various perspectives such as exploratory analysis by fixed or mixed effects functional ANOVA models [1], smoothing spline models [2,3], and varying coefficient models [4–7]. A Gaussian distribution was typically assumed for the models mentioned above or a large number of clusters or subjects were available to estimate the inverse of the within cluster covariance matrix. Without these assumptions, the inference faces major challenges such as nonconsistency due to a large number of nuisance parameters, difficulties to estimate large unknown heteroscedastic covariance matrices using a limited number of subjects, as well as loss of power due to high dimensionality. Since we would like to impose no restrictions on the distributions of data and the asymptotic setting of a large number of within cluster measurements requires very different techniques from those assuming a large number of independent clusters, we skip lengthy discussions and refer the reader to [8] for a discussion of some references.

To incorporate both numerical and ordinal functional data, Wang and Akritas [8] and Wang et al. [9] proposed nonparametric test statistics based on original observations for evaluating the significance of fixed effects related to time and treatment in heteroscedastic functional data, respectively. The test statistics in [9] are mainly quadratic forms with a limiting χ^2 distribution and those in [8] are based on a difference of two quadratic forms with a limiting normal distribution.

* Corresponding author.
E-mail addresses: hwang@ksu.edu (H. Wang), mga@stat.psu.edu (M.G. Akritas).

0047-259X/$ – see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmva.2010.03.012

The asymptotic distributions of their test statistics were both obtained under nonclassical asymptotics in the sense that the number of within cluster measurements is large without requiring the number of subjects being large. Strong moment conditions were assumed for validity of their inference. It is well known that inferences requiring higher order moment conditions perform poorly for highly skewed or heavy-tailed data. Particularly, the moments do not exist for Cauchy distribution.

Rank based inferences have the advantage of not requiring moment conditions or distributional assumptions. Rank tests were available for some multivariate factorial designs. For example, Thompson [10,11] studied nonparametric tests based on overall ranking of data from balanced one-factor design assuming continuous distributions. Akritas and Arnold [12] and Akritas and Brunner [13] gave nonparametric tests using overall rankings of data from multivariate repeated measures designs allowing both discrete and continuous data. Munzel and Brunner [14,15] proposed an approach using separate rankings for different variables. The asymptotic results in these papers are under the classical asymptotic setting where the sample size per treatment tends to infinity and the number of unknown effects remains fixed. For other rank results under the classical setting, see [16,17] and the references therein. Recent papers by Harrar and Bathke [18] and Bathke and Harrar [19] proposed several tests based on separate rankings of data when there are a large number of fixed effects and a small number of correlated measurements and sample sizes. However, there are no rank results available for functional data in the literature.

In this paper, we develop robust (mid-)rank based inference for hypothesis testing in functional data. We will formulate the nonparametric main effect of treatment, time, and their interactions through the underlying unknown distributions. The test statistics are constructed using rank transforms of those based on original observations. We will show that the rank statistics are asymptotically equivalent to their counterpart based on asymptotic rank transforms defined through the average cumulative distribution function under the asymptotic setting that the number of within cluster measurements is large while the number of clusters per treatment may be small. The response variable in this paper can be measured on a continuous or discrete ordinal scale since we work with the unknown distribution functions through (mid-)ranks. We do not restrict the allowed pattern of heteroscedasticity and no specified parametric model is needed to describe the change in the response distribution from one value to another. With the general setup and least assumptions, our theory applies to a wide range of data including those from skewed or heavy-tailed distributions that are commonly seen in data from high-throughput studies. Simulations and real data analysis provide additional supportive evidence to the proposed tests.

The rest of the paper is organized as follows. Section 2 describes the nonparametric model and hypotheses. In Section 3, we state the main theoretical results about the test statistics and their asymptotic distributions. Analysis of real data and simulation studies are presented in Section 4 followed by concluding remarks. The main technical arguments are given in Appendix A and proofs of the supporting lemmas are given in Appendix B.

2. The nonparametric model and hypotheses

Consider subjects nested within a total of a factor levels such that each subject is measured at b time points $t_1, \dots, t_b$. The kth subject in factor level i generates a time series $\mathbf{X}_{ik} = (X_{i1k}, \dots, X_{ibk})'$, $i = 1, \dots, a$, $k = 1, \dots, n_i$. The a groups can be factor level combinations of several factors and the individual effect can be recovered through contrast matrices. The total number of subjects $n = \sum_{i=1}^{a} n_i$ can be large or small. Suppose X_{ijk} have marginal distribution $F_{ij}(x)$ for all $k = 1, \dots n_i$ for some unknown $F_{ij}(\cdot)$.

We consider rank procedures that test hypothesis related to the distribution functions by decomposing F_{ij} in a way similar to the decomposition of the mean μ_{ij} as in the parametric ANOVA model. They were first introduced by Akritas and Arnold [12] as purely nonparametric hypotheses.

$$F_{ij}(x) = M(x) + A_i(x) + B_j(x) + C_{ij}(x), \tag{2.1}$$

where $\sum_{i=1}^{a} A_i(x) = \sum_{j=1}^{b} B_j(x) = \sum_{i=1}^{a} C_{ij}(x) = \sum_{j=1}^{b} C_{ij}(x) = 0$, $\forall x$. The functions A_i, B_j, C_{ij} in (2.1) are the fully nonparametric effects and the nonparametric hypotheses specify that the corresponding effects are zero. Specifically we denote

$$H_0(B):\ \text{all } B_j = 0, \quad H_0(C):\ \text{all } C_{ij} = 0, \quad H_0(D):\ \text{all } D_{ij} = A_i + C_{ij} = 0,$$
$$\text{and} \quad H_0(A):\ \text{all } A_i = 0, \quad \text{or, more generally } \widetilde{H}_0(A): \mathbf{C}\overline{\mathbf{F}} = \mathbf{0},\ \text{where } \overline{\mathbf{F}} = (\overline{F}_{1.}, \dots, \overline{F}_{a.})',$$

where $\mathbf{C}$ is a contrast matrix with full row rank, and $\overline{F}_{i.} = b^{-1} \sum_{j=1}^{b} F_{ij}(x)$. The null hypothesis $H_0(C)$ means that the marginal distribution is a mixture of two components with equal mixture probability, one depending on the row factor, and the other one depending on the column factor. The other hypotheses can be similarly interpreted. The fully nonparametric hypotheses are stronger than the usual parametric hypotheses in the sense that the former one implies but are not implied by the latter.

To achieve weak convergence, it is necessary to control the correlation among observations. We assume that the time series corresponding to different subjects are independent and each time series satisfies an α-mixing condition, i.e., assume that for some sequence $\alpha_m \to 0$ as $m \to \infty$, $|P(A \cap B) - P(A)P(B)| \le \alpha_m$ holds for all $A \in \sigma(X_{i1k}, \dots, X_{i\ell k})$, $B \in \sigma(X_{i,\ell+m,k}, X_{i,\ell+m+1,k}, \dots)$, and all i, k, where $\sigma(\cdot)$ denotes the σ-field generated by the random variables. The α-mixing condition basically requires the correlation between observations from the same subject to decay as the time lag m increases. Billingsley [20] gave a central limit theory for strictly stationary α-mixing process. A few others considered nonstationary weak dependent processes (cf. [21] and the references therein). Herrndorf [21] gave conditions for weak convergence for α-mixing short dependence processes allowing the mixing coefficient to approach 0 with a rate faster than $O(m^{-1})$. However, additional assumption $E(S_b^2/b) \to \sigma^2$, for some σ^2, was assumed in [21], where $S_b = \sum_{j=1}^b \epsilon_j$ is the partial sum of an α-mixing process $\{\epsilon_j, j = 1, \ldots, \}$ (for ease of discussion on CLT, we focus on one process and drop the index i for treatment and k for subject temporarily). This condition is not easy to verify without making further assumptions.

The decay rate $\alpha_m = O(m^{-5})$ was assumed in [8] for residuals conditional on the random intercept and a CLT was given for nonstationary α-mixing process under this rate without assuming the order of $E(S_b^2/b)$. It is possible to relax the rate for α_m to include long range dependence processes but additional assumptions need to be imposed. For example, the result of [22] gave a CLT allowing $\alpha_m > m^{-1}$. The CLT stated that $S_b/\sqrt{kh(p_b)}$ is asymptotically normally distributed, where $h(b)$ is the order of $E(S_b^2)$ and p_b satisfies four conditions: (1) $k(p_b + q_b) = b$; (2) $p_b, q_b, k \to \infty$ and $q_b/p_b \to 0$ as $b \to \infty$; (3) $h(q_b) = o(h(b)/k^3)$; (4) $k \le [-\log \alpha_{q_b}]^{1/2}$.

With these constraints, the order of $kh(p_b)$ is between b and b^2 (not including the end points). Therefore, if this result is applied to the test statistics based on original observations, the standardizing rate for the test statistics would then be $b/\sqrt{kh(p_b)} = O(b^\beta)$ for some $\beta < 1/2$. That is, allowing slower rate for the mixing coefficient will lead to slower convergence rate for the test statistics. [23] contain a few other results on long dependence processes though most of them are for stationary or Gaussian processes with the exception of some references studied linear processes (a linear combination of iid random variables with zero mean and variance one). To allow heteroscedastic data and the parametric standardizing rate, we follow the decay rate assumption $\alpha_m = O(m^{-5})$ for nonstationary processes in this paper. Should slower rate for the mixing coefficient is required and the convergence rate for the test statistics is not a concern, we refer to the references above and [9] to extend the results. For simplicity, we impose the α-mixing condition directly on the observations and only give a remark about how to proceed if the lagged correlations do not tend to zero as this is straightforward.

3. Rank tests

In this section we give the (mid-)rank based test statistics and their asymptotic distribution for the hypotheses stated in Section 2.

(Mid-)Ranks can be expressed as a special transformation of the original observations. Specifically, let $H(x) = N^{-1} \sum_{i=1}^a \sum_{j=1}^b n_i F_{ij}(x)$ be the average distribution function, $c(x, y) = [I(x < y) + I(x \le y)]/2$, and $\widehat{H}(x) = N^{-1} \sum_{i=1}^a \sum_{j=1}^b n_i \widehat{F}_{ij}(x)$, $\widehat{F}_{ij}(x) = n_i^{-1} \sum_{k=1}^{n_i} c(X_{ijk}, x)$, be its empirical version of the average distribution function. Then $R_{ijk} = 1/2 + N\widehat{H}(X_{ijk})$ is the (mid-)rank of X_{ijk} among all observations. $\widehat{H}(X_{ijk})$ and $H(X_{ijk})$ are referred as *rank transforms* and *asymptotic rank transforms* (ART), respectively, as they relate the observations to their overall (mid-)ranks directly or asymptotically. The concept of ART was first introduced in [24]. Rank transforms are not independent even when the original observations are independent. On the contrary, the ART may inherit some good properties from the original data, such as correlation structure. (Mid-)rank based tests typically involve showing the asymptotic equivalence between quadratic forms based on rank transforms and their counterparts based on ART's. This has been commonly adopted by many authors (cf. [24,12–15,18,19] among others). The first five references show such equivalence under the large sample size setting, and the latter two references show such equivalence under the small sample size but a large number of fixed effects setting. In this article, we establish such equivalence under the functional data setting, i.e., a large number of within cluster measurements but with possibly a small number of clusters.

For any transformation Y_{ijk} of X_{ijk}, the following notations are used throughout the paper:

$$N = nb, \qquad \widetilde{n} = \min_{1 \le i \le a}\{n_i\}, \qquad \overline{Y}_{ij.} = \frac{1}{n_i}\sum_{k=1}^{n_i} Y_{ijk}, \qquad \widetilde{Y}_{i..} = \overline{Y}_{i..} = \frac{1}{b}\sum_{j=1}^{b} \overline{Y}_{ij.}, \qquad \overline{Y}_{...} = \frac{1}{N}\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n_i} Y_{ijk},$$

$$\overline{Y}_{i.k} = \frac{1}{b}\sum_{j=1}^{b} Y_{ijk}, \qquad \widetilde{Y}_{...} = \frac{1}{ab}\sum_{i=1}^{a}\sum_{j=1}^{b} \overline{Y}_{ij.}, \qquad \widetilde{Y}_{.j.} = \frac{1}{a}\sum_{i=1}^{a} \overline{Y}_{ij.}, \qquad \overline{Y}_{.j.} = \frac{1}{n}\sum_{i=1}^{a}\sum_{k=1}^{n_i} Y_{ijk}.$$

Let $ASB_R, ASC_R, ASD_R, ASE_R$ and $ASE_{D,R}$ be the variations of the mean squares defined in (mid-)ranks as follows:

$$ASB_R = a\sum_{j=1}^{b} \frac{\left(\widetilde{R}_{.j.} - \widetilde{R}_{...}\right)^2}{b-1}, \qquad ASC_R = \sum_{i,j} \frac{(\overline{R}_{ij.} - \widetilde{R}_{i..} - \widetilde{R}_{.j.} + \widetilde{R}_{...})^2}{(a-1)(b-1)}, \tag{3.1}$$

$$ASE_R = \sum_{i,j}\sum_{k=1}^{n_i} \frac{\left(R_{ijk} - \overline{R}_{ij.} - \overline{R}_{i.k} + \widetilde{R}_{i..}\right)^2}{a(b-1)n_i(n_i-1)}, \tag{3.2}$$

$$ASD_R = \sum_{i,j} \frac{\left(\overline{R}_{ij.} - \widetilde{R}_{.j.}\right)^2}{(a-1)b}, \qquad ASE_{D,R} = \frac{1}{ab}\sum_{i,j,k} \frac{(R_{ijk} - \overline{R}_{ij.})^2}{n_i(n_i-1)}. \tag{3.3}$$

The statistics given above are the rank transform of the corresponding statistics based on the original observations in [8,9]. These mean squares are different from those used in traditional mixed effects models in that the un-weighted means are used and the summation over the subjects is removed from the sums of squares. In addition, the corresponding mean squares for errors are adjusted to match the expectations under the heteroscedastic setting. Though bearing exactly the same form as those based on original observations, however, the components under the summations are always correlated due to the rankings used.

Define $F_{R,B} = ASB_R/ASE_R$, $F_{R,C} = ASC_R/ASE_R$ and $F_{R,D} = ASD_R/ASE_{D,R}$. When there is only one treatment, i.e., $a = 1$, model (2.1) is reduced to $F_j(x) = M(x) + B_j(x)$. In this case, the only interest would be to evaluate the time effect of the single treatment using $F_{R,B}$ that is still well defined through ASB_R and ASE_R.

Set $Y_{ijk} = H(X_{ijk})$, $\widetilde{\sigma}_{ijj'} = \text{cov}(Y_{ijk}, Y_{ij'k})$, and $\widetilde{\sigma}_{ijj} = \widetilde{\sigma}_{ij}^2 = \text{Var}(Y_{ijk})$. Let $\widetilde{\sigma}^2 = \lim_{b\to\infty} E(ASE_R)$, and $\widetilde{\sigma}_*^2 = \lim_{b\to\infty} E(\widetilde{n}ASE_R)$, where

$$E(ASE_R) = \frac{1}{a(b-1)}\sum_{i,j}\frac{\widetilde{\sigma}_{ij}^2}{n_i} - \frac{1}{ab(b-1)}\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{j'=1}^{b}\frac{\widetilde{\sigma}_{ijj'}}{n_i}.$$

To obtain the asymptotic distribution of the rank tests, we need to establish the equivalence between the rank statistics and their asymptotic rank transforms. Note that the latter are the test statistics evaluated at $H(X_{ijk})$, for all i, j, k. Since $H(\cdot)$ is the average cumulative distribution function, it is bounded by one uniformly and therefore the strong moment assumption on the original observations are automatically satisfied for $H(X_{ijk})$. Further, the α-mixing property on $\{X_{ijk}, j = 1, \ldots\}$ is carried over to the process $\{H(X_{ijk}), j = 1, \ldots\}$ by Theorem 5.2 of [25] about Borel functions of independent α-mixing processes. We state it as a lemma here for convenience.

As multiple α-mixing sequences are involved in the lemma, the dependence coefficients α_m will be denoted $\alpha(\mathbf{X}, m)$ for a given α-mixing sequence $\mathbf{X} = \{X_j, j \in \mathcal{Z}\}$, where $\mathcal{Z}$ is some index set.

Lemma 3.1. *Suppose that for each $i = 1, 2, 3, \ldots$, $\mathbf{X}^{(i)} = \{X_j^{(i)}, j \in \mathcal{Z}\}$ is a (not necessarily stationary) sequence of α-mixing random variables. Suppose these sequences $\mathbf{X}^{(i)}$, $i = 1, 2, 3, \ldots$ are independent of each other. Suppose that for each $j \in Z$, $h_j : \mathcal{R} \times \mathcal{R} \times \mathcal{R} \times \cdots \longrightarrow \mathcal{R}$ is a Borel function. Define the sequence $\mathbf{U} = \{U_j, j \in \mathcal{Z}\}$ of random variables by $U_j = h_j(X_j^{(1)}, X_j^{(2)}, X_j^{(3)}, \ldots)$, $j \in \mathcal{Z}$. Then for each $m \geq 1$, $\alpha(\mathbf{U}, m) \leq \sum_{i=1}^{\infty}\alpha(\mathbf{X}^{(i)}, m)$.*

Applying this lemma to $H(x)$, a Borel function with a single argument, we know that $\{H(X_{ijk}), j = 1, \ldots\}$ is an α-mixing process with the same dependence coefficient α_m as $\{X_{ijk}, j = 1, \ldots\}$, for each i, k. Then the results based on the original observations can be applied to the asymptotic rank transforms. The next theorem gives the asymptotic distributions of the test statistics.

Theorem 3.2. *Assume for each subject in each group, $\{X_{ijk}, j = 1, 2, \ldots, \}$ is α-mixing with $\alpha_m = O\left(m^{-5}\right)$, $i = 1, \ldots, a$, $k = 1, \ldots, n_i$. Then as $b \to \infty$ while a remains bounded, the limits of*

$$\widetilde{\zeta}_1 = \frac{2}{a^2 b}\sum_{j=1}^{b}\sum_{j'=1}^{b}\sum_{i=1}^{a}\frac{\widetilde{\sigma}_{ijj'}^2}{n_i(n_i-1)} \quad \text{and} \quad \widetilde{\zeta}_2 = \frac{2}{a^2 b}\sum_{j=1}^{b}\sum_{j'=1}^{b}\sum_{i\neq i'}^{a}\frac{\widetilde{\sigma}_{ijj'}\widetilde{\sigma}_{i'jj'}}{n_i n_{i'}} \tag{3.4}$$

exist regardless of whether the n_i stay bounded or go to ∞. Further,

(1) *if $n_i \geq 2$ are bounded, then*

$$\text{under } H_0(B),\quad \sqrt{b}(F_{R,B}-1) \xrightarrow{d} N\left(0, \widetilde{\tau}_B^2/\widetilde{\sigma}^4\right);$$

$$\text{under } H_0(C),\quad \sqrt{b}(F_{R,C}-1) \xrightarrow{d} N\left(0, \widetilde{\tau}_C^2/\widetilde{\sigma}^4\right);$$

$$\text{under } H_0(D),\quad \sqrt{b}(F_{R,D}-1) \xrightarrow{d} N\left(0, \widetilde{\tau}_C^2/\widetilde{\sigma}^4\right);$$

where $\widetilde{\tau}_B^2 = \lim_{b\to\infty}(\widetilde{\zeta}_1 + \widetilde{\zeta}_2)$ and $\widetilde{\tau}_C^2 = \lim_{b\to\infty}\left(\widetilde{\zeta}_1 + \widetilde{\zeta}_2/(a-1)^2\right)$.

(2) *if $n_i \to \infty$ as $b \to \infty$, assume $\max_i\{n_i\}/\widetilde{n} = O(1)$. Then*

$$\text{under } H_0(B),\quad \sqrt{b}(F_{R,B}-1) \xrightarrow{d} N\left(0, \widetilde{\tau}_{B*}^2/\widetilde{\sigma}_*^4\right);$$

$$\text{under } H_0(C),\quad \sqrt{b}(F_{R,C}-1) \xrightarrow{d} N\left(0, \widetilde{\tau}_{C*}^2/\widetilde{\sigma}_*^4\right);$$

$$\text{under } H_0(D),\quad \sqrt{b}(F_{R,D}-1) \xrightarrow{d} N\left(0, \widetilde{\tau}_{C*}^2/\widetilde{\sigma}_*^4\right);$$

where $\widetilde{\tau}_{B*}^2 = \lim_{b\to\infty}\widetilde{n}^2(\widetilde{\zeta}_1 + \widetilde{\zeta}_2)$ and $\widetilde{\tau}_{C*}^2 = \lim_{b\to\infty}\widetilde{n}^2\left(\widetilde{\zeta}_1 + \widetilde{\zeta}_2/(a-1)^2\right)$.

Remark 3.1. The convergence rate of the test statistics in Theorem 3.2 depends on the number of within subject measurements b but not on the number of subjects n_i. This is achieved since the n_i in the numerator and denominator of both the test statistics and the asymptotic variances have the same order.

In the proposition below, we give consistent estimators for $\tilde{\zeta}_1$ and $\tilde{\zeta}_2$.

Proposition 3.3. *Let $\tilde{\zeta}_1$ and $\tilde{\zeta}_2$ be as given in (3.4). Define $\overline{C}(j,h) = [\min\{b, j+b^h\}]$ and $\underline{C}(j,h) = [\max\{0, j-b^h\}]$ for some $0 < h < 1$, where $[x]$ denotes the largest integer less than or equal to x. Set*

$$\widehat{\zeta_1} = \frac{2}{a^2 b}\sum_{j=1}^{b}\sum_{j'=\underline{C}(j,h)}^{\overline{C}(j,h)}\sum_{i=1}^{a}\frac{\widehat{\psi}_{ijj'}}{n_i(n_i-1)}, \qquad \widehat{\zeta_2} = \frac{2}{a^2 b}\sum_{j=1}^{b}\sum_{j'=\underline{C}(j,h)}^{\overline{C}(j,h)}\sum_{i\neq i'}^{a}\frac{\widehat{\sigma}_{ijj'}\widehat{\sigma}_{i'jj'}}{n_i n_{i'}},$$

where

$$\widehat{\psi}_{ijj'} = \sum_{k_1\neq k_2\neq k_3\neq k_4}^{n_i}\frac{(R_{ijk_1}-R_{ijk_2})(R_{ij'k_1}-R_{ij'k_2})(R_{ijk_3}-R_{ijk_4})(R_{ij'k_3}-R_{ij'k_4})}{4n_i(n_i-1)(n_i-2)(n_i-3)},$$

and

$$\widehat{\sigma}_{ijj'} = \sum_{k=1}^{n_i}\frac{(R_{ijk}-\overline{R}_{ij.})(R_{ij'k}-\overline{R}_{ij'.})}{n_i-1}.$$

Then as $b \to \infty$, regardless of whether n_i stay bounded or go to ∞, $\widehat{\zeta_1}/N^4 - \tilde{\zeta}_1 \xrightarrow{p} 0$ and $\widehat{\zeta_2}/N^4 - \tilde{\zeta}_2 \xrightarrow{p} 0$, provided that $h < 1/2$.

Remark 3.2. The diagnostics for the α-mixing condition is discussed in [8]. If the lagged correlations do not go to zero as the lagged distance increases to ∞, then a hidden random intercept effect may be contributing to the covariances. In such case, the results for testing $H_0(B)$, $H_0(C)$ in Theorem 3.2 and Proposition 3.3 still hold if the α-mixing condition is assumed on the residuals defined as the observations minus their conditional mean given the random intercept. Correspondingly, the estimators $\widehat{\psi}_{ijj'}$ and $\widehat{\sigma}_{ijj'}$ also need to be adjusted by replacing R_{ijk} by $R_{ijk} - \overline{R}_{i.k}$. The test for $H_0(D)$ no longer holds in this case.

For testing of no group effect, the hypothesis involves only a small number of parameters but we need to handle a large number of nuisance parameters from the unknown correlations. The result is stated below.

Theorem 3.4. *Let $\mathbf{W}_R = (\overline{R}_{1..}, \ldots, \overline{R}_{a..})'$. Assume $\{X_{ijk}, j = 1, 2, \ldots,\}$ is α-mixing with $\alpha_m = O\left(m^{-5}\right)$ for all i, k. Let*

$$\widehat{\eta}_{Ri} = \frac{n}{bn_i(n_i-1)}\sum_{j=1}^{b}\sum_{j'=\underline{C}(j,h)}^{\overline{C}(j,h)}\sum_{k=1}^{n_i}\left(R_{ijk}-\overline{R}_{ij.}\right)\left(R_{ij'k}-\overline{R}_{ij'.}\right), \quad i = 1, \ldots, a,$$

where $\underline{C}(j,h)$, $\overline{C}(j,h)$ are given in Proposition 3.3. Then under $\widetilde{H}_0(A)$, for a contrast matrix $\mathbf{C}_a$ with full row rank r,

$$N\mathbf{W}_R'\mathbf{C}_a'\left[\mathbf{C}_a\mathrm{diag}(\hat{\eta}_{R1}, \ldots, \hat{\eta}_{Ra})\mathbf{C}_a'\right]^{-1}\mathbf{C}_a\mathbf{W}_R \xrightarrow{d} \chi_r^2, \text{ as } b \to \infty,$$

regardless of whether the n_i remain bounded or tend to ∞, provided that $\max_i\{n_i\}/\widetilde{n} = O(1)$ and $h < 1/2$.

The proofs of Theorems 3.2 and 3.4 and Proposition 3.3 are given in Appendix A together with some supporting lemmas.

Remark 3.3. The use of $\underline{C}(j,h)$ and $\overline{C}(j,h)$ in Proposition 3.3 and Theorem 3.4 are important in that it not only is necessary for the asymptotic convergence but also controls the amount of noises included in the test statistics. Without such control, the estimators may accumulate too much noises.

4. Numerical study

In [8,9], their tests based on original observations (NPorg) showed superior performance in terms of both type I error and power in simulations when compared with traditional tests including linear mixed models and generalized least square methods with various covariance structures. These traditional tests failed to maintain type I error rate or exhibit low power under local alternatives. In this section, we first examine how b^h affects the behavior of the test statistics and then compare the performance of the rank test with NPorg and linear mixed effects model with a random intercept on skewed, heavy-tailed and normally distributed data.

Note that the estimator for $\widetilde{\zeta}_1$ requires the number of subjects to be at least four if we use unbiased estimator $\widehat{\psi}_{ijj'}$ for $\sigma^2_{ijj'}$. Throughout this section when the number of subjects per treatment is less than four, we use a Jackknife bias corrected estimator of $\sigma^2_{ijj'}$. The following notations are used in Sections 4.1 and 4.2: $N(\mu, \sigma)$ denote normal distribution with mean μ and standard deviation σ; lognormal(μ, σ) to denote lognormal distribution whose log transform follows normal distribution with mean μ and standard deviation σ; Cauchy(μ, σ) to denote Cauchy distribution with location parameter μ and scale parameter σ.

4.1. Influence of b^h on performance of the rank test

First we examine the influence of b^h on the test statistics in simulations. This term exists in both the test statistic for the main effect of treatment and the estimators for the asymptotic variances in the other tests. For $i = 1, 2, j = 1, \ldots, b$,

Table 1
Influence of b^h on type I error estimates of the proposed rank tests at level 0.05. The number of treatments is $a = 2$, each of size $n = 4$, and the number of time points is b.

b^h	Effect	$b = 50$			$b = 80$		
		Normal	Lognormal	Cauchy	Normal	Lognormal	Cauchy
3	Treatment (A)	0.0620	0.0490	0.0560	0.0545	0.0490	0.0540
	Time (B)	0.0565	0.0440	0.0505	0.0495	0.0500	0.0445
	Interaction (C)	0.0490	0.0500	0.0585	0.0505	0.0460	0.0500
	Simple treatment effect (D)	0.0520	0.0510	0.0655	0.0520	0.0490	0.0505
4	Treatment (A)	0.0630	0.0500	0.0540	0.0530	0.0500	0.0525
	Time (B)	0.0535	0.0395	0.0475	0.0455	0.0440	0.0375
	Interaction (C)	0.0440	0.0455	0.0530	0.0460	0.0430	0.0445
	Simple treatment effect (D)	0.0465	0.0445	0.0615	0.0470	0.0440	0.0460

Table 2
Influence of b^h on type I error estimates of the proposed rank tests for *unbalanced design* at level 0.05. The number of treatments is $a = 2$, with $n_1 = 4$, $n_2 = 7$, and the number of time points is $b = 30$.

	$b^h = 4$			$b^h = 5$			$b^h = 6$		
Time (B)	0.0540	0.0560	0.0530	0.0490	0.0545	0.0505	0.0460	0.0510	0.0490
Interaction (C)	0.0530	0.0535	0.0525	0.0490	0.0500	0.0505	0.0475	0.0485	0.0480
Simple treatment effect (D)	0.0595	0.0590	0.0655	0.0580	0.0565	0.0615	0.0540	0.0545	0.0575

$k = 1, 2, 3, 4$, we generated data from model

$$X_{ijk} = d_{ijk} + \xi_{ijk} \tag{4.1}$$

$$d_{ijk} = 0.4d_{i,j-1,k} + \epsilon_{ijk} \tag{4.2}$$

where $d_{i1k} = 0$, and $\xi_{ijk}, \epsilon_{ijk}$ are iid from normal, lognormal, or Cauchy distribution with parameter $\mu = 0$ and $\sigma = 1$. When $b = 80$, for b^h being integers from 3 to 7, the type I error estimates at level $\alpha = 0.05$ for the rank test of no treatment effect is 0.0545, 0.0530, 0.0530, 0.0520, 0.0530 under the normal model, 0.049, 0.050, 0.051, 0.053, 0.0545 under the lognormal model, and 0.0575, 0.054, 0.0525, 0.0525, 0.0520, 0.0555 under the Cauchy model. Stable results are also observed for the tests on the effect of time, treatment by time interaction, and simple treatment effect (see Table 1). In another simulation using unbalanced design with $n_1 = 4$, $n_2 = 7$ when $b = 30$, we also obtained acceptable type I error with $b^h = 4, 5, 6$ for each effect (see Table 2). So the test statistics are stable over a range of values for b^h.

Note that b^h is the size of the overlapping window to capture the correlations between nearby observations. Therefore, too small values of b^h, such as 1 or 2, may not be enough to give reliable estimation of the asymptotic variance. On the other hand, it is required that $h < 1/2$ in Proposition 3.3 and Theorem 3.4. In practice, we recommend to take b^h to be an integer that is reasonably large such that it is smaller than $b^{1/2}$. We set $b^h = 4$ for the rest of the simulation study.

4.2. *Comparison of performance in simulated data*

4.2.1. *Performance on data satisfying α-mixing condition*

In this section, we compare the rank test with those based on original observations (NPorg) in [8,9], and linear mixed effects model with a random intercept on skewed data from lognormal, heavy-tailed data from Cauchy distribution, and data from normal distribution. For interaction or main time effect, NPorg refers to the test statistics in [8]. For main or simple treatment effect, NPorg refers to those in [9]. The data were generated with the same model as in Section 4.1. All results reported in this subsection are based on 2000 runs.

The type I error estimates for testing no main effect of treatment, time, treatment effect over time, and no simple treatment effect at $\alpha = 0.05$ are reported in Table 3 for designs with the number of time points $b = 20, 30, 50$, and 80. Both the *lme* and NPorg perform poorly under the lognormal and Cauchy distributions. On the other hand, the rank test has good type I error estimates in all situations. In a separate simulation where data were generated from model (4.1) using an unbalanced design with the number of subjects $n_1 = 4$ and $n_2 = 7$ and the error term ζ_{ijk} having scale parameter i, we obtained similar results as reported above.

Results from power simulations are summarized in Fig. 1. The simulations are under alternatives to the setting of Table 3 for testing for no treatment effect with $b = 50$ (top three panels), and for testing for no interaction effect with $b = 20$ (lower three panels). Specifically, the observations were generated as in (4.1) but with d_{ijk} from model $d_{ijk} = (-1)^i\tau + 0.4d_{i,j-1,k} + \epsilon_{ijk}$ for the test of no main treatment effect, and $d_{ijk} = (-1)^i\tau f(j-1) + 0.4d_{i,j-1,k} + \epsilon_{ijk}$ for the test of no interaction effect for various τ values (see Fig. 1). The function $f(\cdot)$ takes form $f(x) = \log(x)/2$ for the normal and $f(x) = x$ for lognormal and Cauchy distributions.

Table 3
Comparison of type I error estimates at level 0.05. The number of treatments is $a = 2$, each of size $n = 4$, and the number of time points is b.

b	Effect	Normal			Lognormal			Cauchy		
		lme	NPorg	Rank	lme	NPorg	Rank	lme	NPorg	Rank
20	Treatment (A)	0.050	0.069	0.067	0.028	0.054	0.071	0.021	0.031	0.069
	Time (B)	0.054	0.048	0.050	0.029	0.002	0.050	0.021	0.003	0.059
	Interaction (C)	0.055	0.054	0.060	0.019	0.001	0.060	0.020	0.003	0.052
	Simple treatment (D)	0.061	0.058	0.063	0.019	0.001	0.063	0.020	0.004	0.063
30	Treatment (A)	0.050	0.060	0.061	0.034	0.047	0.057	0.019	0.030	0.064
	Time (B)	0.065	0.055	0.061	0.027	0.002	0.061	0.021	0.002	0.057
	Interaction (C)	0.057	0.050	0.053	0.027	0.001	0.053	0.021	0.002	0.044
	Simple treatment (D)	0.059	0.052	0.053	0.029	0.001	0.053	0.021	0.002	0.055
50	Treatment (A)	0.040	0.060	0.057	0.045	0.044	0.071	0.019	0.028	0.056
	Time (B)	0.059	0.044	0.054	0.029	0.001	0.040	0.025	0.001	0.048
	Interaction (C)	0.058	0.040	0.044	0.021	0.001	0.046	0.017	0.001	0.053
	Simple treatment (D)	0.060	0.047	0.047	0.021	0.001	0.045	0.019	0.001	0.061
80	Treatment (A)	0.052	0.056	0.053	0.021	0.027	0.050	0.025	0.028	0.052
	Time (B)	0.064	0.044	0.045	0.027	0.001	0.044	0.020	0.000	0.038
	Interaction (C)	0.053	0.039	0.046	0.017	0.001	0.043	0.017	0.000	0.045
	Simple treatment (D)	0.055	0.042	0.047	0.017	0.001	0.044	0.019	0.001	0.046

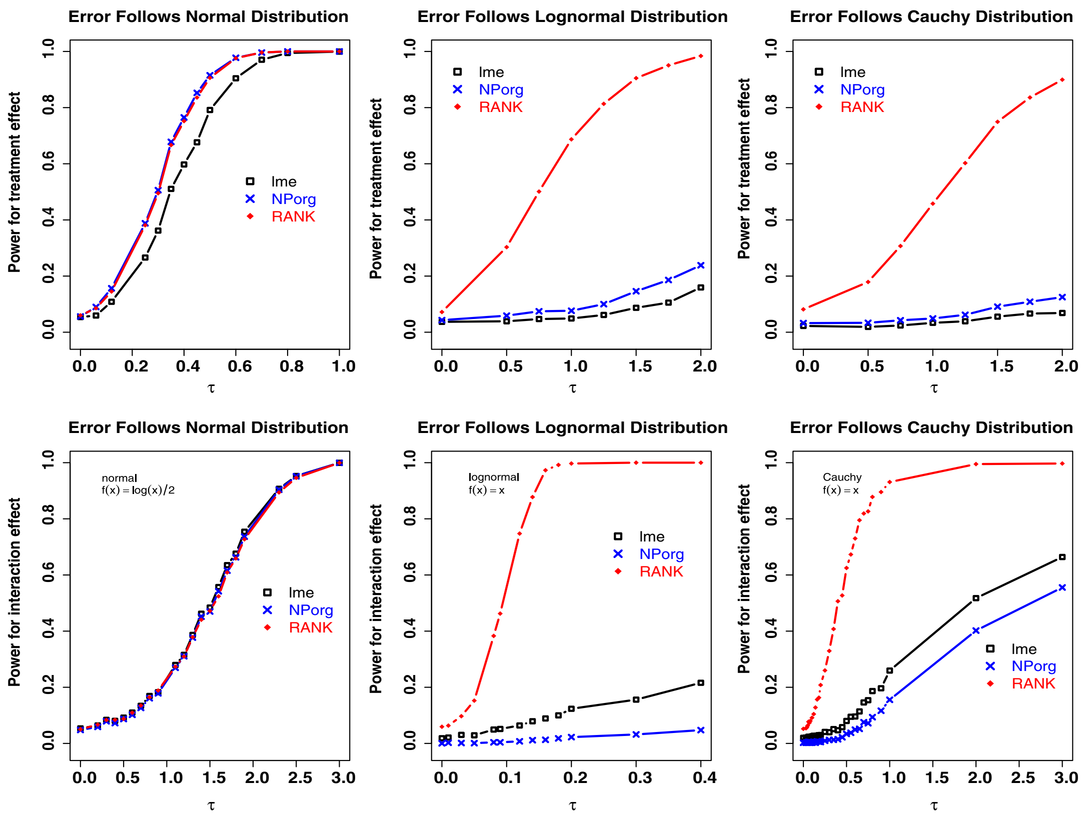

Fig. 1. Estimated power under alternatives. The top three panels are for treatment effect when $d_{ijk} = (-1)^i \tau + 0.4 d_{i,j-1,k} + \epsilon_{ijk}$ and the lower three panels are for interaction effect when $d_{ijk} = (-1)^i \tau f(j-1) + 0.4 d_{i,j-1,k} + \epsilon_{ijk}$. In both cases, $X_{ijk} = d_{ijk} + \xi_{ijk}$ with ϵ_{ijk} and ξ_{ijk} follow i.i.d. normal, lognormal or Cauchy distribution. The function $f(x) = \log(x)/2$ is for the normal and $f(x) = x$ is for lognormal and Cauchy distributions.

For the test of no main treatment effect (top panels of Fig. 1), *lme* is not as powerful as NPorg and the rank test. The rank test clearly outperforms NPorg and *lme*. NPorg has almost no power under lognormality and the Cauchy distribution. Under normality, the rank test and NPorg have comparable power.

For the test of no interaction effect (lower panels), all three tests have almost identical power under normality whereas the rank test significantly outperforms NPorg and *lme* under lognormality and the Cauchy distribution. The *lme* has better power than NPorg for the given form of alternatives under lognormality and the Cauchy distribution.

The poor performance of NPorg under lognormality and Cauchy distribution is because the theoretical argument of NPorg requires at least 16th central moment to be finite. This is not satisfied for Cauchy distribution and the sample (co)variance based on small replications is a poor estimator for the (co)variance in lognormal distribution. The rank test, however, does not rely on any moment assumptions.

4.2.2. Performance on unstructured data

The α-mixing condition is general enough to allow many different covariance structures such as autoregressive (AR) models, some autoregressive moving average models (ARMA), Gaussian serial correlation, as well as other unnamed covariance structures as long as the mixing coefficients satisfy the required polynomial decay rate. Such decay rate makes it possible to have the central limit theorem on nonstationary dependent processes. Even though this rate could be relaxed to include long range dependent process, the asymptotic variance estimators will employ different forms than those presented in this manuscript (cf. [9]). For such reason, we advise the reader to take special care if the data do not satisfy the α-mixing condition with the specified mixing coefficient. Particularly, if the condition is violated, the test statistics in this manuscript may be used but with bootstrap to determine significance. In the rest of this subsection, we illustrate with a simulation study for data generated with unstructured covariance matrix upon the request of a reviewer.

Note that unstructured covariance matrix does not satisfy the α-mixing constraint required in this manuscript. To describe data generation, let $\mathrm{diag}(\cdot)$ be the function that extracts the diagonal elements of a matrix to form a vector. Let $\mathrm{Diag}(c)$ be the function that uses the values in vector c to form a diagonal matrix. Denote $\mathbf{1}_b$ to be a b-dimensional column vector of ones and $I(i=1)$ to be an indicator function for $i=1$. We generated data Y_{ijk} as below (let e^A be an elementwise exponential transform of a vector A) :

$$\begin{aligned}&\mathbf{Y}_{ik} = e^{\mathbf{L}\mathbf{X}_{ik}}, \text{ where } \mathbf{X}_{ik} = (X_{i1k}, \ldots, X_{ibk})' \text{ has independent element}\\ &X_{ijk} \sim N(\tau(I(i=1) - I(i=2)), c_j^2), \text{ and } c_j, j = 1, \ldots, b \text{ are iid from } N(1, 0.2^2),\end{aligned} \tag{4.3}$$

and $\mathbf{L}$ is the lower half triangular matrix from Cholesky decomposition of an unstructured correlation matrix $\Sigma = \left\{\boldsymbol{\rho}_b\boldsymbol{\rho}_b' + \mathrm{Diag}(\mathbf{1}_b - \mathrm{diag}(\boldsymbol{\rho}_b\boldsymbol{\rho}_b'))\right\}$, with $\boldsymbol{\rho}_b$ being a b-dimensional column vector with elements generated from uniform (0, 1) independently. In this data generation setting, the marginal distribution of the data is lognormal distribution. The simulation is done with $a=2$, $b=50$, and $n_1=n_2=4$. When $\tau=0$, the null hypothesis of no main treatment effect and no interaction effect are satisfied. When $\tau\neq 0$, both main effect of treatment and the interaction effects of treatment over time exist.

We compare the proposed rank test with some benchmark methods: linear mixed effects model with various covariance structures (R version 2.7.2 in linux using package *nlme* with command *lme*), generalized least squares method (using package *nlme* with command *gls*), and generalized estimating equation (GEE) approach (using package *gee*) with independent working correlation using robust variance estimator and Gaussian family or quasi-likelihood. For other covariance structures, GEE fails to converge.

The proportions of rejections at 0.01 level for no treatment by time interactions and no main treatment effects based on 3000 runs are reported in Tables 5 and 6 respectively. The parameters of interaction effects lie in a high-dimensional space while those for the main treatment effects lie in a low-dimensional space. The estimated level corresponds to the case with $\tau=0$ that is under the null hypothesis. It can be seen that the tests of interaction effect become conservative under the null with the exception of generalized estimating equations approach being liberal. Under the alternative where there are interaction effects, the rank test has the best power. Note that if the dimension b is small but n_i's are large, then both linear mixed effects model and generalized estimating equations allow to specify unstructured covariance structure. However, in the small n_i, large b setting, this option is not possible for these classical methods.

For the test of no main treatment effect, none of the tests considered have acceptable type I error under this data generation setting (see first column of Table 6). Due to this reason, we can not use proportion of p-values that are less than the true level 0.01 to estimate the power. Instead, we find the rejection region using percentiles of all p-values when the data were generated under the null hypothesis. Then we use the proportion of p-values for data generated under the alternatives falling inside the rejection region to give bootstrap estimate of the power. The power estimates reported in Table 6 are the bootstrap power. It can be seen that the likelihood based tests from linear mixed effects models barely have any power to detect deviations from the null hypothesis.

In summary, the simulation study in this subsection suggests that the test statistics with asymptotic normal distribution still maintain acceptable type I error with good power when the covariance matrix is randomly generated with no structure. However, the test of no treatment effect based on asymptotic chi-square distribution may fail to maintain the type I error as is also the case for all other tests compared in this subsection. Correspondingly, bootstrap estimate of the significance with the proposed test statistic could be used to give a powerful test for the treatment effect.

4.3. Application to Hessian fly data

In this section, we apply the proposed rank test to a data set collected for studying the effect of pheromone and color in attracting Hessian fly in Kansas State University. The experiment involves two factors: treatment with 3 levels (blank, pheromone CP4, or pheromone CP5) and color with 2 levels (yellow or white). Four traps are used as replications for each treatment and color combination. A fixed amount of the same pheromone compounds was added to each trap that uses

Table 4
p-values for testing the effects in Hessian fly data.

Effects	NPorg	Rank	Effects from GEE	GEE
Treatment	1.11×10^{-16}	0	Treatment	0.787
Color	0.0135	0.0027	Color	0.783
Treatment:color	2.78×10^{-8}	0	Treatment:color	0.393
Time	0	0	Time	0.441
Treatment:time	0.115	5.3×10^{-8}	Treatment:time	0.877
CP4.vs.CP5	0.846	0.29	Color:time	0.662
Pheromone.vs.Blank	3.83×10^{-10}	0	Treatment:color:time	0.885

Table 5
Proportion of rejections at 0.01 level for no treatment and time interaction effect based on 3000 runs. The data Y_{ijk} were generated from lognormal distribution as in (4.3) with unstructured covariance matrix. Table legend: NPorg: the Wang and Akritas [8] test based on original observations; Rank: the proposed rank test; lmeRan: linear mixed effects model (lme) with a random intercept; lmeRanAR1: lme with a random intercept plus an AR(1) serial correlation; lmeRanGaus: lme with a random intercept plus Gaussian serial correlation; glsAR1: generalized least squares test with AR(1) serial correlation; geeInd: GEE with Gaussian family using independent working correlation; geeQuasi: GEE with Quasi-likelihood using independent working correlation.

	Estimated	Power estimate					
	level	$\tau = 0.8$	$\tau = 1$	$\tau = 1.25$	$\tau = 1.5$	$\tau = 2.5$	$\tau = 3$
NPOrg	0.000	0.039	0.089	0.167	0.241	0.441	0.465
Rank	0.002	0.603	0.729	0.816	0.871	0.971	0.983
lmeRan	0.007	0.413	0.516	0.608	0.689	0.862	0.888
lmeRanAR1	0.004	0.371	0.470	0.570	0.655	0.840	0.866
lmeRanGaus	0.004	0.377	0.475	0.574	0.658	0.838	0.866
glsAR1	0.001	0.327	0.434	0.522	0.610	0.801	0.838
geeInd	0.056	0.526	0.590	0.648	0.690	0.736	0.731
geequasi	0.056	0.526	0.590	0.648	0.690	0.736	0.731

Table 6
Proportion of rejections at 0.01 level for no treatment effect based on 3000 runs. The data Y_{ijk} were generated from lognormal distribution as in (4.3) with unstructured covariance matrix. Table legend is same as those in Table 5.

	Estimated	Bootstrap	Bootstrap power estimate					
	level	level	$\tau = 0.8$	$\tau = 1$	$\tau = 1.25$	$\tau = 1.5$	$\tau = 2.5$	$\tau = 3$
NPOrg	0.207	0.010	0.749	0.744	0.718	0.679	0.539	0.494
Rank	0.238	0.010	0.993	0.998	1.000	1.000	1.000	1.000
lmeRan	0.139	0.011	0.135	0.115	0.081	0.049	0.016	0.009
lmeRanAR1	0.114	0.012	0.161	0.146	0.098	0.061	0.022	0.012
lmeRanGaus	0.118	0.012	0.161	0.140	0.093	0.057	0.020	0.010
glsAR1	0.083	0.011	0.187	0.166	0.115	0.077	0.020	0.018
geeInd	0.047	0.012	0.359	0.367	0.322	0.286	0.200	0.193
geequasi	0.047	0.012	0.359	0.367	0.322	0.286	0.200	0.193

pheromone every two weeks. The number of flies caught in each trap was recorded every other day leading to 36 repeated measurements per trap. Of interest to the entomologist is the pheromone effect compared to blank with no pheromone, color effect, difference between CP4 and CP5, and interactions among the factors. Yellow traps with CP4 caught a total of 742 flies while white traps with CP4 only caught 488 flies; yellow traps with CP5 caught 899 flies but white traps with CP5 only caught 267 flies; blank yellow and white traps caught 80 and 23 flies respectively. These summary statistics suggested a significant color effect and pheromone effect. In addition, a boxplot (not shown) suggested that many of the counts are zero over time and the variations of the counts differ over time and over different treatment and color combinations.

To apply the rank tests, the treatment and color combinations are treated as the levels of a single factor and the individual effect of color or pheromone is obtained through contrast matrices. The *p*-values are listed in Table 4 along with the *p*-values from NPorg and GEE in the same table for comparison. GEE does not have power to detect any significant effects for this data set. This happens because GEE requires a large number of subjects and a small number of within subject measurements for its validity. This data set, on the contrary, has only 4 traps with a large number of within trap measurements.

For the effect of color, the rank test yields *p*-value 0.0027 while the NPorg test gives *p*-value of 0.0135. The *p*-value for treatment by time interaction effect is 5.3×10^{-8} for the rank test and 0.115 for the test of NPorg. For the rest of the effects, both the rank test and NPorg yield highly significant results except that CP4 is not significantly different from CP5 in either test. Taken into account the summary statistics, we see that the rank test provides stronger evidence for the color effect and treatment by time interaction effect than the tests based on original observations.

5. Summary

In this paper, we developed rank test for high-dimensional heteroscedastic functional data. The theory obtained under a nonstationary α-mixing condition is valid for both small and large number of clusters with large unknown correlation matrices. In the case of only a small number of clusters, the test statistics gain power from a large number of repeated measurements. The test statistic for the hypothesis of no main effect of treatment has asymptotically a Chi-square distribution. Other test statistics that deal with hypotheses involving a large number of parameters have asymptotically normal distributions. In a real application and simulation studies, the proposed rank test significantly outperforms those based on original observations for skewed or heavy-tailed data as in lognormal or Cauchy case.

Acknowledgments

We thank Dr. Jeffrey R. Woodward, Erik Echegaray and Dr. Ming-Shun Chen for the Hessian fly data. We are grateful to Dr. Stephen B. Vardeman at Iowa State University, Drs. Raymond J. Carroll, Suojin Wang at Texas A&M University, Ming-Shun Chen, and the anonymous referees for their helpful comments that led to significant improvement of the manuscript.

Appendix A. Some major technical arguments

In this section, we state some lemmas and give the derivation of the theorems. Detailed proofs of the lemmas are given in Appendix B. Denote $u_{ijk} = X_{ijk} - \mu_{ij}$ and $p_{ij} = E(H(X_{ijk}))$.

Lemma A.1. *Let $\bar{\mathbf{F}} = (\bar{F}_{1.}, \dots, \bar{F}_{a.})'$ and $\bar{\bar{\mathbf{F}}} = (\bar{\bar{F}}_{1.}, \dots, \bar{\bar{F}}_{a.})'$. Then as $b \to \infty$, regardless of whether the n_i go to infinity or remain bounded, $\sqrt{N}\int(\widehat{H} - H)\, d(\bar{\bar{\mathbf{F}}} - \bar{\mathbf{F}}) \xrightarrow{p} \mathbf{0}$.*

Above relationship also holds for $\widetilde{n} \to \infty$ and b stays bounded. Here the asymptotics is under the setting of $b \to \infty$.

Lemma A.2. *Let $\mathbf{R}^c = (R_{111} - \mu_{R,11}, R_{112} - \mu_{R,11}, \dots, R_{11n_1} - \mu_{R,11}, R_{121} - \mu_{R,12}, \dots, R_{abn_a} - \mu_{R,ab})$, where $\mu_{R,ij} = NE(H(X_{ijk})) + 1/2$, and let $P_{ASE}(\cdot)$ be the function defined as*

$$P_{ASE}(\mathbf{u}) = \frac{1}{ab}\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n_i}\frac{u_{ijk}^2}{n_i^2} - \frac{1}{ab}\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k\neq k'}^{n_i}\frac{u_{ijk}u_{ijk'}}{n_i^2(n_i-1)}. \tag{A.1}$$

Then

(a) $\widetilde{n}\sqrt{b}[ASE_R - P_{ASE}(\mathbf{R}^c)]/N^2 = o_p(1)$.

(b) $\widetilde{n}ASE_R/N^2 \xrightarrow{p} \widetilde{\sigma}_*^2$, *where $\widetilde{\sigma}_*^2 = \lim_{b\to\infty}(ab)^{-1}\sum_{i,j,k}\widetilde{n}\widetilde{\sigma}_{ij}^2/n_i^2$ provided that the limit exists.*

Lemma A.3. *Let $P_B(\cdot)$, $P_C(\cdot)$ be the functions defined as in*

$$P_B(\mathbf{u}) = \frac{a}{b}\sum_{j=1}^{b}\widetilde{u}_{.j.}^2, \qquad P_C(\mathbf{u}) = \frac{1}{(a-1)b}\sum_{i=1}^{a}\sum_{j=1}^{b}\bar{u}_{ij.}^2 - \frac{a}{b(a-1)}\sum_{j=1}^{b}\widetilde{u}_{.j.}^2, \tag{A.2}$$

with u_{ijk} replaced by $R_{ijk} - \mu_{R,ij}$, where $\mu_{R,ij} = 1/2 + Np_{ij}$. and $\mathbf{R}^c$ is given in Lemma A.2. Then as $b \to \infty$ while a remains bounded,

(a) *under $H_0(B)$, $\widetilde{n}\sqrt{b}(ASB_R - P_B(\mathbf{R}^c))/N^2 \xrightarrow{p} 0$;*

(b) *under $H_0(C)$, $\widetilde{n}\sqrt{b}(ASC_R - P_C(\mathbf{R}^c))/N^2 \xrightarrow{p} 0$.*

Proof of Theorem 3.2. First we show that the limit of $\widetilde{\zeta}_1$ and $\widetilde{\zeta}_2$ exists. By Lemma 3.1, we know that $\{Y_{ijk} = H(X_{ijk}), j = 1, \dots\}$ is bounded by one and is an α-mixing process with the same decay rate as $\{X_{ijk}, j = 1, \dots\}$. Therefore $|\widetilde{\sigma}_{ijj'}| = |\text{cov}(Y_{ijk}, Y_{ij'k})| \le \alpha_{|j-j'|}$ (see Lemma 2 on page 365 of [20]). We have for all b,

$$\widetilde{\zeta}_1 = \frac{2}{a^2b}\sum_{j=1}^{b}\sum_{j'=1}^{b}\sum_{i=1}^{a}\frac{\widetilde{\sigma}_{ijj'}^2}{n_i(n_i-1)} \le \frac{2}{ab}\sum_{j=1}^{b}\sum_{j'=1}^{b}\alpha_{|j-j'|}^2 = \frac{4}{ab}\sum_{m=0}^{b-1}(b-m)\alpha_m^2 \le \frac{4}{a}\sum_{m=0}^{b-1}\alpha_m^2 < \infty.$$

Similarly,

$$|\widetilde{\zeta}_2| \le \frac{2}{a^2b}\sum_{j=1}^{b}\sum_{j'=1}^{b}\sum_{i\neq i'}^{a}\frac{|\widetilde{\sigma}_{ijj'}\widetilde{\sigma}_{i'jj'}|}{n_in_{i'}} \le \frac{2}{b}\sum_{j=1}^{b}\sum_{j'=1}^{b}\alpha_{|j-j'|}^2 < \infty, \quad \text{as } b \to \infty.$$

Absolute convergence of $\widetilde{\zeta}_1$ and $\widetilde{\zeta}_2$ implies that their limits exist.

Let $P_{ASE}(\mathbf{R}^c)$, $P_B(\mathbf{R}^c)$ and $P_C(\mathbf{R}^c)$ be the functions defined by $P_{ASE}(\cdot)$ in (A.1), $P_B(\cdot)$ and $P_C(\cdot)$ in (A.2), respectively, with argument $\mathbf{R}^c$ given in Lemma A.2. By Lemmas A.2 and A.3, we only need to consider the asymptotic distribution of the projections $\widetilde{n}\sqrt{b}(P_B(\mathbf{R}^c) - P_{ASE}(\mathbf{R}^c))/N^2$ and $\widetilde{n}\sqrt{b}(P_C(\mathbf{R}^c) - P_{ASE}(\mathbf{R}^c))/N^2$ when n_i go to ∞ as $b \to \infty$, under $H_0(\widetilde{B})$ and

$H_0(\widetilde{C})$, respectively. Let $P_1(\cdot), P_2(\cdot)$ be the functions defined as P_1 and P_2 in

$$P_1 = \frac{1}{ab}\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k\neq k'}^{n_i}\frac{u_{ijk}u_{ijk'}}{n_i(n_i-1)}, \qquad P_2 = \frac{1}{ab}\sum_{i\neq i'}^{a}\sum_{j=1}^{b}\overline{u}_{ij.}\overline{u}_{i'j.}, \tag{A.3}$$

respectively, but with u_{ijk} replaced by $R_{ijk} - \mu_{R,ij}$, and let $P_1(\mathbf{Y})$, $P_2(\mathbf{Y})$ be similarly defined but with $Y_{ijk} - p_{ij}$ replacing the u_{ijk}. Then $P_B(\mathbf{R}^c) - P_{ASE}(\mathbf{R}^c) = P_1(\mathbf{R}^c) + P_2(\mathbf{R}^c)$ and $P_C(\mathbf{R}^c) - P_{ASE}(\mathbf{R}^c) = P_1(\mathbf{R}^c) - P_2(\mathbf{R}^c)/(a-1)$. We also have $ASD_R - ASE_{D,R} = P_1(\mathbf{R}^c) - P_2(\mathbf{R}^c)/(a-1)$. If we can show the following

$$\widetilde{n}\sqrt{b}\left(\frac{P_1(\mathbf{R}^c)}{N^2} - P_1(\mathbf{Y}^c)\right) = o_p(1) \quad\text{and}\quad \widetilde{n}\sqrt{b}\left(\frac{P_2(\mathbf{R}^c)}{N^2} - P_2(\mathbf{Y}^c)\right) = o_p(1), \tag{A.4}$$

then the result would follow from that of Theorem 3.2 of [8]. Write

$$\frac{P_1(\mathbf{R}^c)}{N^2} - P_1(\mathbf{Y}^c) = \frac{D_5 + 2D_6}{ab}, \qquad \frac{P_2(\mathbf{R}^c)}{N^2} - P_2(\mathbf{Y}^c) = \frac{D_7 + 2D_8}{ab},$$

where

$$D_5 = \sum_{i,j}\sum_{k\neq k'}^{n_i}\frac{(Z_{ijk}-Y_{ijk})(Z_{ijk'}-Y_{ijk'})}{n_i(n_i-1)}, \qquad D_6 = \sum_{i,j}\sum_{k\neq k'}^{n_i}\frac{(Z_{ijk}-Y_{ijk})(Y_{ijk'}-p_{ij})}{n_i(n_i-1)},$$

$$D_7 = \sum_{i\neq i'}^{a}\sum_{j=1}^{b}(\overline{Z}_{ij.}-\overline{Y}_{ij.})(\overline{Z}_{i'j.}-\overline{Y}_{i'j.}), \qquad D_8 = \sum_{i\neq i'}^{a}\sum_{j=1}^{b}(\overline{Z}_{ij.}-\overline{Y}_{ij.})(\overline{Y}_{i'j.}-p_{i'j}).$$

It can be shown that $\widetilde{n}D_5/(a\sqrt{b}), \widetilde{n}/(a\sqrt{b})D_6, \widetilde{n}/(a\sqrt{b})D_7, \widetilde{n}/(a\sqrt{b})D_8$ are all $o_p(1)$, which completes the proof. Here we will only show that $\widetilde{n}/(a\sqrt{b})D_6 = o_p(1)$.

$$E\left(\frac{n^2(a)}{a^2 b}D_6^2\right) = \frac{4n^2(a)}{a^2bN^2}\sum_{i,j}\sum_{k\neq k'}^{n_i}\sum_{i_2,j_2}\sum_{k_2\neq k_2'}^{n_i}\sum_{i_1,j_1,k_1}\sum_{i_3,j_3,k_3}E\left\{\frac{(Y_{ijk'}-p_{ij})(Y_{i_2j_2k_2'}-p_{i_2j_2})}{(n_i-1)(n_{i_2}-1)}\right.$$
$$\left.\frac{1}{n_{i_2}n_i}[c(X_{i_1j_1k_1},X_{ijk}) - F_{i_1j_1}(X_{ijk})][c(X_{i_3j_3k_3},X_{i_2j_2k_2}) - F_{i_3j_3}(X_{i_2j_2k_2})]\right\}.$$

Note that the expectation under the summation is zero if the number of different elements in set $\{k, k', k_2, k_2', k_1, k_3\}$ is five or six. When the number of different elements in set $\{j, j_2, j_1, j_3\}$ is at most two, the summation is $O(b^{-1})$. For $j \neq j_2 \neq j_1 \neq j_3$, the expectation is not zero only if

(a) All four terms under the expectation are correlated. In this situation, we must have $i = i_1 = i_2 = i_3$ and $k' = k_1 = k_2' = k_3$. A representative term in above summation is

$$\sum_{j<j_2<j_1<j_3}E\left\{\frac{(Y_{ijk'}-p_{ij})(Y_{ij_2k'}-p_{ij_2})[c(X_{ij_1k'},X_{ijk})-F_{ij_1}(X_{ijk})]}{(n_i-1)(n_i-1)}\frac{1}{n_in_i}[c(X_{ij_3k'},X_{ij_2k'})-F_{ij_3}(X_{ij_2k'})]\right\}$$
$$\leq \sum_{j<j_2<j_1<j_3}^{b} 64\min(\alpha_{j_2-j}^{1/2},\alpha_{j_1-j_2}^{1/2})\frac{1}{(n_i-1)^2n_i^2} = O\left(\frac{b^2}{(n_i-1)^2n_i^2}\right);$$

or

(b) the four terms under the expectation form two independent groups with two correlated terms in each group. The proof in this situation is similar to (B.6). When the number of different elements in set $\{j, j_2, j_1, j_3\}$ is three, the summation is $O(b^{-1})$ and the proof is similar to that when $j \neq j_2 \neq j_1 \neq j_3$. Thus $\widetilde{n}/(a\sqrt{b})D_6 = o_p(1)$. □

Proof of Proposition 3.3. We show only $\widehat{\widetilde{\zeta}}_1/N^4 - \widetilde{\zeta}_1 \xrightarrow{p} 0$. The other one is similar and is omitted. To make symbols easier, we write $\widehat{\zeta}_1(\mathbf{Z})$ and $\widehat{\zeta}_1(\mathbf{Y})$ as the statistics when the X_{ijk} in $\widehat{\zeta}_1$ is replaced by Z_{ijk} and Y_{ijk} respectively. Similar notations for $\widehat{\sigma}_{ijj'}(\mathbf{Z})$ and $\widehat{\sigma}_{ijj'}(\mathbf{Y})$ will be used. Note that $\widehat{\widetilde{\zeta}}_1 = N^4\widehat{\zeta}_1(\mathbf{Z})$. Apply Proposition 3.3 in [8] on Y_{ijk}, we have $\widehat{\zeta}_1(\mathbf{Y}) - \widetilde{\zeta}_1 \xrightarrow{p} 0$. So it remains to show that $\widehat{\zeta}_1(\mathbf{Z}) - \widehat{\zeta}_1(\mathbf{Y}) = o_p(1)$. Write each of the difference $Z_{ijk_1} - Z_{ijk_2}$ as $Z_{ijk_1} - Z_{ijk_2} - (Y_{ijk_1} - Y_{ijk_2}) + (Y_{ijk_1} - Y_{ijk_2})$, and note that Z_{ijk}, Y_{ijk} are uniformly bounded by 1, we have

$$\widehat{\sigma}_{ijj'}(\mathbf{Z}) - \widehat{\sigma}_{ijj'}(\mathbf{Y}) \leq \sum_{k_1\neq k_2\neq k_3\neq k_4}^{n_i}\frac{4[Z_{ijk_1}-Z_{ijk_2}-(Y_{ijk_1}-Y_{ijk_2})][Z_{ij'k_1}-Z_{ij'k_2}-(Y_{ij'k_1}-Y_{ij'k_2})]}{n_i(n_i-1)(n_i-2)(n_i-3)}$$
$$+ \sum_{k_1\neq k_2\neq k_3\neq k_4}^{n_i}\frac{6[Z_{ijk_1}-Z_{ijk_2}-(Y_{ijk_1}-Y_{ijk_2})](Y_{ij'k_1}-Y_{ij'k_2})}{n_i(n_i-1)(n_i-2)(n_i-3)}$$

$$+\sum_{k_1\neq k_2\neq k_3\neq k_4}^{n_i}\frac{6[Z_{ij'k_1}-Z_{ij'k_2}-(Y_{ij'k_1}-Y_{ij'k_2})](Y_{ijk_1}-Y_{ijk_2})}{n_i(n_i-1)(n_i-2)(n_i-3)}$$
$$=O_p(N^{-1/2}).$$

So $\widehat{\zeta}_1(\mathbf{Z})-\widehat{\zeta}_1(\mathbf{Y})=O_p\left(b^{-1}N^{-1/2}\sum_{j=1}^{b}\sum_{j'=\underline{C}(j,h)}^{\overline{C}(j,h)}\right)=O_p(b^h/\sqrt{N})=o_p(1)$. □

Proof of Theorem 3.4. By Slutsky's Theorem, we only need to show

$$\frac{1}{\sqrt{N}}\mathbf{C}_a\mathbf{W}_R\xrightarrow{d}N(\mathbf{0}_r,\mathbf{C}_a\,\mathrm{Diag}(\eta_{Y1},\ldots,\eta_{Ya})\mathbf{C}_a'),\qquad \widehat{\eta}_{Ri}/N^2-\eta_{Yi}=o_p(1),\tag{A.5}$$

where $\mathbf{0}_r$ is a vector of zeros and $\eta_{Yi}=\lim_{b\to\infty}N\mathrm{Var}\left(\overline{Y}_{i..}-E(\overline{Y}_{i..})\right)$. Let $\widehat{\eta}_{Yi}$ be similarly defined as $\widehat{\eta}_{Ri}$ with R replaced by Y. To show the asymptotic normality in (A.5), note that $\int\widehat{H}\,\mathrm{d}\overline{\overline{F}}_{i.}=\frac{1}{b}\sum_{j=1}^{b}\frac{1}{n_i}\sum_{k=1}^{n_i}\widehat{H}(X_{ijk})=N^{-1}\left(\overline{R}_{i..}-\frac{1}{2}\right)$. So $\mathbf{W}_R=N\int\widehat{H}\,\mathrm{d}\overline{\overline{\mathbf{F}}}+\frac{1}{2}\cdot\mathbf{1}_a$, where $\mathbf{1}_a$ is an a dimensional vector of one's. Since $\mathbf{C}_a$ is a contrast matrix. So under $\widetilde{H}(A)$,

$$\frac{1}{\sqrt{N}}\mathbf{C}_a\mathbf{W}_R=\sqrt{N}\mathbf{C}_a\int\widehat{H}\,d\overline{\overline{\mathbf{F}}}=\sqrt{N}\mathbf{C}_a\int\widehat{H}\,\mathrm{d}(\overline{\overline{\mathbf{F}}}-\overline{\mathbf{F}}).$$

It suffices to find the asymptotic distribution of $\sqrt{N}\int\widehat{H}\,\mathrm{d}(\overline{\overline{\mathbf{F}}}-\overline{\mathbf{F}})$. Note that

$$\sqrt{N}\int H\,\mathrm{d}(\overline{\overline{\mathbf{F}}}-\overline{\mathbf{F}})=\sqrt{N}\left(\overline{Y}_{1..}-E(\overline{Y}_{1..}),\ldots,\overline{Y}_{a..}-E(\overline{Y}_{a..})\right)'=\sqrt{N}[\mathbf{W}_Y-E(\mathbf{W}_Y)],$$

where $\mathbf{W}_Y=(\overline{Y}_{1..},\ldots,\overline{Y}_{a..})'$.

Apply the proof for Theorem 3.2 of [9] on Y_{ijk}, we get $\sqrt{N}(\overline{Y}_{i..}-E(\overline{Y}_{i..}))\xrightarrow{d}N(0,\eta_{Yi})$. Note that $\eta_{Yi}<\infty$ is guaranteed by the α-mixing condition. By independence of the observations from different group, we have $\sqrt{N}[\mathbf{W}_Y-E(\mathbf{W}_Y)]\xrightarrow{d}N_a(\mathbf{0},\mathbf{V}_Y)$ as $a\to\infty, c\to\infty$, where $\mathbf{V}_Y=\mathrm{Diag}(\eta_{Y1},\ldots,\eta_{Ya})$. Apply Continuous Mapping Theorem, we have $\sqrt{N}\mathbf{C}_a\mathbf{W}_Y\xrightarrow{d}N_{a-1}(\mathbf{0},\mathbf{C}_a\mathbf{V}_Y\mathbf{C}_a')$. Therefore, the proof of the theorem is complete since $\sqrt{N}\int(\widehat{H}-H)\,\mathrm{d}(\overline{\overline{\mathbf{F}}}-\overline{\mathbf{F}})\xrightarrow{p}\mathbf{0}$ by Lemma A.1.

To show the second equation in (A.5), we only need to show $\widehat{\eta}_{Ri}/N^2-\widehat{\eta}_{Yi}=o_p(1)$ because $\widehat{\eta}_{Yi}$ is a consistent estimator for η_{Yi}.

$$\begin{aligned}\frac{bn_i(n_i-1)}{n}\left(\widehat{\eta}_{Ri}/N^2-\widehat{\eta}_{Yi}\right)=&\sum_{j=1}^{b}\sum_{j'=-b^h}^{b^h}\sum_{k=1}^{n_i}\left[Z_{ijk}-\overline{Z}_{ij.}-Y_{ijk}+\overline{Y}_{ij.}\right]\left[Z_{i(j+j')k}-\overline{Z}_{i(j+j').}-Y_{i(j+j')k}+\overline{Y}_{i(j+j').}\right]\\&+\sum_{j=1}^{b}\sum_{j'=-b^h}^{b^h}\sum_{k=1}^{n_i}\left[Z_{ijk}-\overline{Z}_{ij.}-Y_{ijk}+\overline{Y}_{ij.}\right]\left[Y_{i(j+j')k}+\overline{Y}_{i(j+j').}\right]\\&+\sum_{j=1}^{b}\sum_{j'=-b^h}^{b^h}\sum_{k=1}^{n_i}\left[Z_{i(j+j')k}-\overline{Z}_{i(j+j').}-Y_{i(j+j')k}+\overline{Y}_{i(j+j').}\right]\left[Y_{ijk}+\overline{Y}_{ij.}\right].\end{aligned}$$

The first term is $O(2n_ib^{\frac{1}{4}}/n)$ and the second and the third terms are both $O(\sqrt{b}n_i/\sqrt{n})$. Thus $\widehat{\eta}_{Ri}/N^2-\widehat{\eta}_{Yi}=O_p\left(\sqrt{n}(n_i-1)^{-1}b^{-1/4}\right)=o_p(1)$. □

Appendix B. Proof of Lemmas

Proof of Lemma A.1. The component of the vector $\sqrt{N}\int(\widehat{H}-H)\,d(\hat{\mathbf{F}}-\mathbf{F})$ is

$$\begin{aligned}\sqrt{N}\int(\widehat{H}-H)\,\mathrm{d}(\widehat{F}_{ij}-F_{ij})&=\frac{1}{\sqrt{N}}\sum_{i_1,j_1}n_{i_1}\int(\widehat{F}_{i_1j_1}-F_{i_1j_1})\,\mathrm{d}(\widehat{F}_{ij}-F_{ij})\\&=\frac{1}{\sqrt{N}}\sum_{i_1,j_1}\left\{\frac{n_{i_1}}{n_i}\sum_{k=1}^{n_i}(\widehat{F}_{i_1j_1}(X_{ijk})-F_{i_1j_1}(X_{ijk}))-\sum_{k_1=1}^{n_{i_1}}\left[1-F_{ij}(X_{i_1j_1k_1})-\int F_{i_1j_1}\,\mathrm{d}F_{ij}\right]\right\}\\&=\frac{1}{\sqrt{N}}\sum_{j_1=1}^{b}nh(\mathbf{X}_{j_1},\mathbf{X}_{ij}),\end{aligned}$$

where

$$h(\mathbf{X}_{j_1}, \mathbf{X}_{ij}) = \frac{1}{nn_i} \sum_{i_1,k_1} \sum_{k=1}^{n_i} \left\{ c(X_{i_1j_1k_1}, X_{ijk}) - F_{i_1j_1}(X_{ijk}) - \left[1 - F_{ij}(X_{i_1j_1k_1}) - \int F_{i_1j_1}\, \mathrm{d}F_{ij} \right] \right\},$$

$\mathbf{X}_{j_1} = (X_{1j_11}, \ldots, X_{aj_1n_a})'$, and $\mathbf{X}_{ij} = (X_{ij1}, \ldots, X_{ijn_i})'$. Note that $h(\mathbf{X}_{j_1}, \mathbf{X}_{ij})$ is uniformly bounded by 4 and satisfies

$$Eh(\mathbf{X}_{j_1}, \mathbf{X}_{ij}) = E[h(\mathbf{X}_{j_1}, \mathbf{X}_{ij})|\mathbf{X}_{j_1}] = E[h(\mathbf{X}_{j_1}, \mathbf{X}_{ij})|\mathbf{X}_{ij}] = 0. \tag{B.1}$$

The component of the vector $\sqrt{N} \int (\widehat{H} - H)\, d(\widehat{\overline{\mathbf{F}}} - \overline{\mathbf{F}})$ is

$$\frac{\sqrt{N}}{b} \sum_{j=1}^{b} \int (\widehat{H} - H)\, \mathrm{d}(\widehat{F}_{ij} - F_{ij}) = \frac{\sqrt{N}}{bN} \sum_{j=1}^{b} \sum_{j_1=1}^{b} nh(\mathbf{X}_{j_1}, \mathbf{X}_{ij})$$

and

$$\begin{aligned} E\left(\frac{\sqrt{N}}{b} \sum_{j=1}^{b} \int (\widehat{H} - H)\, \mathrm{d}(\widehat{F}_{ij} - F_{ij}) \right)^2 &= \frac{n^2}{b^2N} \sum_{j=1}^{b} \sum_{j_1=1}^{b} \sum_{j_2=1}^{b} \sum_{j_3=1}^{b} E\left[h(\mathbf{X}_{j_1}, \mathbf{X}_{ij}) h(\mathbf{X}_{j_2}, \mathbf{X}_{ij_3}) \right] \\ &= \frac{n^2}{b^2N} \sum_{j=1}^{b} \sum_{j_3=1}^{b} \left[2 \sum_{j_1<j_2}^{b} E\left[h(\mathbf{X}_{j_1}, \mathbf{X}_{ij}) h(\mathbf{X}_{j_2}, \mathbf{X}_{ij_3}) \right] + \sum_{j_1=1}^{b} E\left[h(\mathbf{X}_{j_1}, \mathbf{X}_{ij}) h(\mathbf{X}_{j_1}, \mathbf{X}_{ij_3}) \right] \right] \\ &= u_{D1} + u_{D2} + u_{D3} + u_{D4} + u_{D5}, \end{aligned}$$

where

$$u_{D1} = \frac{2n^2}{b^2N} \sum_{j_1<j_2}^{b} \sum_{j<j_3}^{b} E\left[h(\mathbf{X}_{j_1}, \mathbf{X}_{ij}) h(\mathbf{X}_{j_2}, \mathbf{X}_{ij_3}) \right],$$

$$u_{D2} = \frac{2n^2}{b^2N} \sum_{j_1<j_2}^{b} \sum_{j>j_3}^{b} E\left[h(\mathbf{X}_{j_1}, \mathbf{X}_{ij}) h(\mathbf{X}_{j_2}, \mathbf{X}_{ij_3}) \right],$$

$$u_{D3} = \frac{2n^2}{b^2N} \sum_{j_1<j_2}^{b} \sum_{j=1}^{b} E\left[h(\mathbf{X}_{j_1}, \mathbf{X}_{ij}) h(\mathbf{X}_{j_2}, \mathbf{X}_{ij}) \right],$$

$$u_{D4} = \frac{2n^2}{b^2N} \sum_{j_1=1}^{b} \sum_{j<j_3}^{b} E\left[h(\mathbf{X}_{j_1}, \mathbf{X}_{ij}) h(\mathbf{X}_{j_1}, \mathbf{X}_{ij_3}) \right], \qquad u_{D5} = \frac{n^2}{b^2N} \sum_{j=1}^{b} \sum_{j_1=1}^{b} E[h^2(\mathbf{X}_{j_1}, \mathbf{X}_{ij})].$$

Obviously, $u_{D5} = O(n/(b\widetilde{n})) = o(1)$, since $E(h(\cdot,\cdot)^2) = O(n_i^{-1})$ which can be verified by examining the expected value in further detail. Note that

$$u_{D1} = \frac{2n^2}{b^2N} \sum_{j<j_3}^{b} \sum_{j_1<j_2}^{b} E\left[E[h(\mathbf{X}_{j_1}, \mathbf{X}_{ij}) h(\mathbf{X}_{j_2}, \mathbf{X}_{ij_3}) | (\mathbf{X}_{j_1}, \mathbf{X}_{ij})] \right] = 0.$$

Similarly $u_{D2} = 0$ and

$$\begin{aligned} |u_{D3}| &\le \frac{2n^2}{b^2N} \sum_{j_1<j_2}^{b} \sum_{j=1}^{b} \left| E\left[h(\mathbf{X}_{j_1}, \mathbf{X}_{ij}) h(\mathbf{X}_{j_2}, \mathbf{X}_{ij}) \right] \right| \\ &= \frac{2n^2}{b^2N} \sum_{j_1<j_2}^{b} \sum_{j=1}^{b} \left| E\left(E\left[h(\mathbf{X}_{j_1}, \mathbf{X}_{ij}) h(\mathbf{X}_{j_2}, \mathbf{X}_{ij}) | \mathbf{X}_{ij} \right] \right) \right| \\ &= \frac{2n^2}{b^2N} \sum_{j_1<j_2}^{b} \sum_{j=1}^{b} I(j = j_1 \text{ or } j = j_2) \left| E\left(E\left[h(\mathbf{X}_{j_1}, \mathbf{X}_{ij}) h(\mathbf{X}_{j_2}, \mathbf{X}_{ij}) | \mathbf{X}_{ij} \right] \right) \right| = 0, \end{aligned}$$

where the last equality is due to (B.1). The proof of $|u_{D4}| = o(1)$ is similar to that of $|u_{D3}| = o(1)$. Hence $\frac{\sqrt{N}}{b} \sum_{j=1}^{b} \int (\widehat{H} - H)\, \mathrm{d}(\widehat{F}_{ij} - F_{ij}) = o_p(1)$ and the proof is complete. □

Proof of Lemma A.2. We will use the decomposition $ASE_R = P_{ASE}(\mathbf{R}^c) + D_1(\mathbf{R}^c) + D_2(\mathbf{R}^c)$, where $P_{ASE}(\mathbf{R})$, $D_1(\mathbf{R})$ and $D_2(\mathbf{R})$ are similarly defined as P_{ASE}, and

$$D_1(\mathbf{u}) = -\frac{1}{ab(b-1)}\sum_{i=1}^{a}\sum_{k=1}^{n_i}\sum_{j\neq j'}^{b}\frac{u_{ijk}u_{ij'k}}{n_i(n_i-1)}, \qquad D_2(\mathbf{u}) = \frac{1}{ab(b-1)}\sum_{i=1}^{a}\sum_{j\neq j'}^{b}\frac{\bar{u}_{ij.}\bar{u}_{ij'.}}{n_i-1}, \tag{B.2}$$

with u_{ijk} replaced by $R_{ijk}-\mu_{R,ij}$. Define $\mathbf{Y}^c = (Y_{111}-p_{11}, \ldots, Y_{11n_1}-p_{11}, Y_{121}-p_{12}, \ldots, Y_{abn_a}-p_{ab})$, where $p_{ij} = E(H(X_{ijk})) = E(Y_{ijk})$. By Lemma A.1 in [8], the proof will be done if we show that in both cases,

$$\widetilde{n}\sqrt{b}(D_1(\mathbf{R}^c)/N^2 - D_1(\mathbf{Y}^c)) = o_p(1), \qquad \widetilde{n}\sqrt{b}(D_2(\mathbf{R}^c)/N^2 - D_2(\mathbf{Y}^c)) = o_p(1), \tag{B.3}$$

$$\widetilde{n}\sqrt{b}(P_{ASE}(\mathbf{R}^c)/N^2 - P_{ASE}(\mathbf{Y}^c)) = o_p(1). \tag{B.4}$$

We will first show (B.3). Let $Z_{ijk} = \widehat{H}(X_{ijk})$, and write $D_1(\mathbf{R}^c)/N^2 - D_1(\mathbf{Y}^c) = D_{11} + D_{12}$, $D_2(\mathbf{R}^c)/N^2 - D_2(\mathbf{Y}^c) = D_{21} + D_{22}$, where

$$D_{11} = -\sum_{i,k}\sum_{j\neq j'}^{b}\frac{(Z_{ijk}-Y_{ijk})(Z_{ij'k}-Y_{ij'k})}{ab(b-1)n_i(n_i-1)}, \qquad D_{12} = -2\sum_{i,k}\sum_{j\neq j'}^{b}\frac{(Z_{ijk}-Y_{ijk})(Y_{ij'k}-p_{ij'})}{ab(b-1)n_i(n_i-1)},$$

$$D_{21} = -\sum_{i,k,k'}\sum_{j\neq j'}^{b}\frac{(Z_{ijk}-Y_{ijk})(Z_{ij'k'}-Y_{ij'k'})}{ab(b-1)n_i^2(n_i-1)}, \qquad D_{22} = -2\sum_{i,k,k'}\sum_{j\neq j'}^{b}\frac{(Z_{ijk}-Y_{ijk})(Y_{ij'k'}-p_{ij'})}{ab(b-1)n_i^2(n_i-1)}.$$

It is not hard to show that D_{11} and D_{21} are $O_p(N^{-1}\widetilde{n}^{-1})$. The proofs for D_{12} and D_{22} are similar and we will give that for D_{22}. Write

$$D_{22} = 2\sum_{i=1}^{a}\sum_{k=1}^{n_i}\sum_{k'=1}^{n_i}\sum_{j\neq j'}^{b}\sum_{i_1=1}^{a}\sum_{j_1=1}^{b}\sum_{k_1=1}^{n_{i_1}}\frac{\left[c(X_{i_1j_1k_1}, X_{ijk}) - F_{i_1j_1}(X_{ijk})\right]\left(Y_{ij'k'}-p_{ij'}\right)}{ab(b-1)Nn_i^2(n_i-1)},$$

$$E(D_{22}^2) = \frac{4}{a^2b^2(b-1)^2N^2}\sum_{i=1}^{a}\sum_{k=1}^{n_i}\sum_{k'=1}^{n_i}\sum_{j\neq j'}^{b}\sum_{i_1=1}^{a}\sum_{j_1=1}^{b}\sum_{k_1=1}^{n_{i_1}}\sum_{i_2=1}^{a}\sum_{k_2=1}^{n_{i_2}}\sum_{k_2'=1}^{n_{i_2}}\sum_{j_2\neq j_3}^{b}\sum_{i_4=1}^{a}\sum_{j_4=1}^{b}\sum_{k_4=1}^{n_{i_4}} E\left\{\frac{\left(Y_{ij'k'}-p_{ij'}\right)}{n_i^2(n_i-1)}\right.$$
$$\left.\frac{(Y_{i_2j_3k_2'}-p_{i_2j_3})\left[c(X_{i_1j_1k_1}, X_{ijk}) - F_{i_1j_1}(X_{ijk})\right]\left[c(X_{i_4j_4k_4}, X_{i_2j_2k_2}) - F_{i_4j_4}(X_{i_2j_2k_2})\right]}{n_{i_2}^2(n_{i_2}-1)}\right\}.$$

Note that $E(c(X_{i_1j_1k_1}, X_{ijk}) - F_{i_1j_1}(X_{ijk})|X_{ijk}) = 0$, so the expectation under the summation is zero if the number of different elements in set $\{k, k', k_1, k_2, k_2', k_4\}$ is five or six. If the number of different elements in set $\{j, j', j_1, j_2, j_3, j_4\}$ is four or less, the summation is of order $O(b^{-2}\widetilde{n}^{-2})$. When all elements in the set $\{j, j', j_1, j_2, j_3, j_4\}$ are different, without loss of generality, we can consider a representative case in which $j < j' < j_1 < j_2 < j_3 < j_4$, $k_4 = k$ and $k' = k_2'$ (Note: the expectation under the summation is not zero only when the number of different elements in $\{k, k', k_1, k_2, k_2', k_4\}$ is four or less). In this case,

$$\sum_{j'<j_1<j_3<j_4}^{b} E\left\{\left(Y_{ij'k'}-p_{ij'}\right)\left[c(X_{i_1j_1k_1}, X_{ijk}) - F_{i_1j_1}(X_{ijk})\right](Y_{i_2j_3k_2'}-p_{i_2j_3})\left[c(X_{i_4j_4k_4}, X_{i_2j_2k_2}) - F_{i_4j_4}(X_{i_2j_2k_2})\right]\right\}$$
$$= \sum_{j'<j_1<j_3<j_4}^{b} E\left\{\left[c(X_{i_1j_1k_1}, X_{ijk}) - F_{i_1j_1}(X_{ijk})\right]\left[c(X_{i_4j_4k}, X_{i_2j_2k_2}) - F_{i_4j_4}(X_{i_2j_2k_2})\right]\right.$$
$$\left.\times\ \left(Y_{ij'k'}-p_{ij'}\right)\left(Y_{i_2j_3k'}-p_{i_2j_3}\right)\right\} \quad (\text{ note that } k_4 = k \text{ and } k' = k_2'\ ). \tag{B.5}$$

If $k' \notin \{k, k_1, k_2\}$,

$$\text{(B.5)} \leq \sum_{j_1<j_4}^{b}\left|E\left[c(X_{i_1j_1k_1}, X_{ijk}) - F_{i_1j_1}(X_{ijk})\right]\left[c(X_{i_4j_4k}, X_{i_2j_2k_2}) - F_{i_4j_4}(X_{i_2j_2k_2})\right]\right|\sum_{j'<j_3}\left|E\left(Y_{ij'k'}-p_{ij'}\right)\left(Y_{i_2j_3k'}-p_{i_2j_3}\right)\right|$$
$$\leq \sum_{j_1<j_4}^{b} 4\times 2\times 2\alpha_{j_4-j_1}^{1/2} I(k_1 = k_2)\sum_{j'<j_3}^{b} 4\times 2\times 2\alpha_{j_3-j'}^{1/2} = O(b^2 I(k_1 = k_2)). \tag{B.6}$$

If $k' \in \{k, k_1, k_2\}$, the total number of different elements in set $\{k, k', k_1, k_2, k_2', k_4\}$ is three or less, we have the following 3 cases:

- All four terms under the expectation are correlated, which can be dealt with similarly as $E(D_{12}^2)$.
- The four terms under the expectation form two independent groups with two terms in each group correlated. This situation can be handled similarly as (B.6).

- The four terms under the expectation form two independent groups and one of the groups contains three correlated terms, or the four terms form three or four independent groups. In this situation, the expectation is zero.

Therefore, $E(D_{22}^2) = O(b^{-2}n^{-3}(a))$ and so $\sqrt{b}\widetilde{n}D_{22} = o_p(1)$.
To show (B.4), write $P_{ASE}(\mathbf{R}^c)/N^2 - P_{ASE}(\mathbf{Y}^c) = P_{RY1} - P_{RY2}$, where

$$\begin{aligned} P_{RY1} &= \frac{1}{ab}\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n_i}\frac{(Z_{ijk}-p_{ij})^2-(Y_{ijk}-p_{ij})^2}{n_i^2} \\ &= \frac{1}{ab}\sum_{i,j,k}\frac{(Z_{ijk}-Y_{ijk})^2}{n_i^2}+\frac{2}{ab}\sum_{i,j,k}\frac{(Z_{ijk}-Y_{ijk})(Y_{ijk}-p_{ij})}{n_i^2}, \\ P_{RY2} &= \frac{1}{ab}\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k\neq k'}^{n_i}\frac{(Z_{ijk}-p_{ij})(Z_{ijk'}-p_{ij})-(Y_{ijk}-p_{ij})(Y_{ijk'}-p_{ij})}{n_i^2(n_i-1)} \\ &= \sum_{i,j}\sum_{k\neq k'}^{n_i}\frac{(Z_{ijk}-Y_{ijk})(Z_{ijk'}-Y_{ijk'})}{abn_i^2(n_i-1)}+2\sum_{i,j}\sum_{k\neq k'}^{n_i}\frac{(Z_{ijk}-Y_{ijk})(Y_{ijk'}-p_{ij})}{abn_i^2(n_i-1)}. \end{aligned}$$

The first summations in both P_{RY1} and P_{RY2} are $O_p(N^{-1}\widetilde{n}^{-1})$. The second summations in both P_{RY1} and P_{RY2} are $o_p(b^{-1}\widetilde{n}^{-3})$ and the proof is similar to that of $\sqrt{b}\widetilde{n}D_{22} = o_p(1)$ and is omitted. Thus the proof of this lemma is completed. □

Proof of Lemma A.3. Note that $ASB_R - P_B(\mathbf{R}^c) = -D_3(\mathbf{R}^c)$, $ASC_R - P_C(\mathbf{R}^c) = (D_3(\mathbf{R}^c) - D_4(\mathbf{R}^c))/(a-1)$, where $D_3(\mathbf{R}^c)$ and $D_4(\mathbf{R}^c)$ are similarly defined as $D_3(\mathbf{u})$ and $D_4(\mathbf{u})$ in the proof of Lemma A.2 in [8] with $\mathbf{u}$ replaced by $\mathbf{R}^c$. Note that the expression of $D_3(\mathbf{R}^c)$ and $D_4(\mathbf{R}^c)$ is very close to $D_2(\mathbf{R}^c)$. When $\widetilde{n} \to \infty$ as $b \to \infty$, the proof of $\widetilde{n}\sqrt{b}(D_4(\mathbf{R}^c)/N^2 - D_4(\mathbf{Y}^c)) = o_p(1)$ follows that of (B.3) (see the proof of Lemma A.2). Due to independence of the observations in different groups, the proof of $\widetilde{n}\sqrt{b}(D_3(\mathbf{R}^c)/N^2 - D_3(\mathbf{Y}^c)) = o_p(1)$ is not much different from that of (B.3). When n_i are bounded, treat $\widetilde{n}$ as a bounded number in above argument. Then we complete the proof of the lemma. □

References

[1] J. Ramsay, B. Silverman, Functional Data Analysis, Springer, 1997.
[2] B. Brumback, J. Rice, Smoothing spline models for the analysis of nested and crossed samples of curves, J. Amer. Statist. Assoc. 93 (1998) 961–976.
[3] Y. Wang, Mixed-effects smoothing spline anova, J. Roy. Statist. Soc. Ser. B 60 (1998) 159–174.
[4] D.R. Hoover, J.A. Rice, C. Wu, L. Yang, Nonparametric smoothing estimates of time-varing coefficient models with longitudinal data, Biometrika 85 (4) (1998) 809–822.
[5] J. Fan, J. Zhang, Two-step estimation of funtional linear models with application to longitudinal data, J. Roy. Statist. Soc. Ser. B 62 (2000) 303–322.
[6] Z. Cai, J. Fan, Q. Yao, Functional-coefficient regression models for nonlinear time series, J. Amer. Statist. Assoc. 95 (2000) 941–956.
[7] C.-T. Chiang, J. Rice, C. Wu, Smoothing spline estimation for varying coefficient models with repeatedly measured dependent variables, J. Amer. Statist. Assoc. 96 (2001) 605–619.
[8] H. Wang, M. Akritas, Inference from heteroscedastic functional data, J. Nonparametr. Stat. 22 (2) (2010) 149–168, doi:10.1080/10485250903171621.
[9] H. Wang, J. Higgins, D. Blasi, Distribution-free tests for no effect of treatment in heteroscedastic functional data under both weak and long range dependence, Statist. Probab. Lett. 80 (2010) 390–402, doi:10.1016/j.spl.2009.11.016.
[10] G. Thompson, Asymptotic distribution of rank statistics under dependencies with multivariate application, J. Multivariate Anal. 33 (1990) 183–211.
[11] G. Thompson, A unified approach to rank tests for multivariate and repeated measures designs, J. Amer. Statist. Assoc. 86 (1991) 410–419.
[12] M. Akritas, S. Arnold, Fully nonparametric hypotheses for factorial designs i: multivariate repeated measures designs, J. Amer. Statist. Assoc. 89 (1994) 336–343.
[13] M. Akritas, E. Brunner, A unified approach to rank tests for mixed models, J. Statist. Plann. Inference 61 (1997) 249–277.
[14] U. Munzel, E. Brunner, Nonparametric methods in multivariate factorial designs, J. Statist. Plann. Inference 88 (2000) 117–132.
[15] U. Munzel, E. Brunner, Nonparametric tests in the unbalanced multivariate one-way design, Biom. J. 42 (2000) 837–854.
[16] E. Brunner, M. Denker, Rank statistics under dependent observations and applications to factorial designs, J. Statist. Plann. Inference 42 (1994) 353–378.
[17] E. Brunner, S. Domhof, F. Langer, Nonparametric analysis of longitudinal data in factorial experiments, 2002.
[18] S. Harrar, A. Bathke, Nonparametric methods for unbalanced multivariate data and many factor levels, J. Multivariate Anal. 99 (2008) 1635–1664.
[19] A. Bathke, S. Harrar, Nonparametric methods in multivariate factorial designs for large number of factor levels, J. Statist. Plann. Inference 138 (2008) 588–610.
[20] P. Billingsley, Probability and Measure, Third edition, Wiley, New-York, 1995.
[21] N. Herrndorf, A functional central limit theorem for weakly dependent sequences of random variables, Ann. Probab. 12 (1984) 141–153.
[22] M. Rosenblatt, A central limit theorem and a strong mixing condition, Proc. Natl. Acad. Sci. USA 42 (1956) 43–47.
[23] P. Doukhan, G. Oppenheim, M.S. Taqqu, Theory and Applications of Long-Range Dependence, Birkhauser Boston, c/o Springer-Verlag, New York, 2003.
[24] M.G. Akritas, The rank transform method in some two-factor designs, J. Amer. Statist. Assoc. 85 (1990) 73–78.
[25] R. Bradley, Basic properties of strong mixing conditions. A survey and some open questions, Probab. Surv. 2 (2005) 1549–5787.