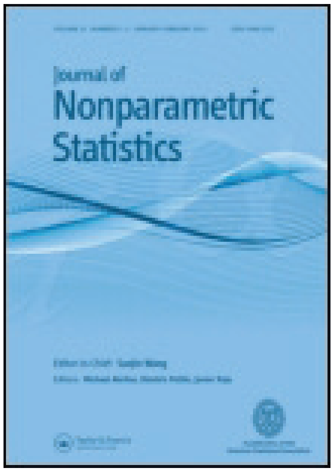

Journal of Nonparametric Statistics
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/gnst20

Inference from heteroscedastic functional data

Haiyan Wang [a] & Michael G. Akritas [b]
[a] Department of Statistics , Kansas State University , 101 Dickens Hall, Manhattan, KS, 66503, USA
[b] Department of Statistics , Pennsylvania State University , 316 Thomas Building, University Park, PA, 16802, USA
Published online: 27 Jan 2010.

Journal of Nonparametric Statistics
Vol. 22, No. 2, February 2010, 149–168

Inference from heteroscedastic functional data

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

[a]*Department of Statistics, Kansas State University, 101 Dickens Hall, Manhattan, KS 66503, USA;* [b]*Department of Statistics, Pennsylvania State University, 316 Thomas Building, University Park, PA 16802, USA*

(*Received 8 July 2008; final version received 6 July 2009*)

Technological advancements have produced an abundance of data sets in which a large number of repeated measurements are observed within a subject or stratum. Many of these data sets are based on a small number of subjects rendering most existing inferential methods unsuitable. This paper develops test procedures based on a novel model for nested heteroscedastic high-dimensional data which we propose. The novelty of the model rests on the fact that the random effects are assumed to be neither uncorrelated nor normal. The model is nonparametric in the sense that it leaves the covariance structure unspecified and applies to both discrete and continuous data. The test procedures developed are useful for evaluating the effects of time as well as their interactions with the crossed factors on the stratum. The asymptotic theory of the test statistics is driven by a large number of measurements per subject and the assumption of nonstationary α-mixing on the error term. Simulation studies and real applications show that the proposed tests are more powerful in detecting effects compared with benchmark methods in data with very limited number of replications.

Keywords: functional data analysis; nonparametric inference; hypothesis testing; high dimensional multivariate analysis

AMS Subject Classification: 62G10; 62G20; 62H15

1. Introduction

The profusion of clustered high-dimensional data, made possible by recent technological advances in many disciplines, raises new statistical challenges. Such data often contain a large number of repeated observations per cluster over time or space, while the number of clusters or subjects may be very limited due to multiple concerns such as cost. For example, fast functional magnetic resonance imaging data contain measurements from the brain recorded at a time scale of seconds to help diagnose and determine how the normal, diseased or injured brain functions, as well as for assessing the potential risks of invasive treatments of the brain. Other examples arise in the context of identifying the disease status using signal intensities from thousands of signal intensities of proteins, lipids, fatty acids or other metabolites. These data are referred to as functional data. A common characteristic of such data is high dimensionality with spatial or temporal correlations while only a small amount of replications is available.

*Corresponding author. Email: hwang@ksu.edu

ISSN 1048-5252 print/ISSN 1029-0311 online

DOI: 10.1080/10485250903171621
http://www.informaworld.com

In this paper, we deal with hypothesis-testing problems to evaluate the effect of functional data and their interactions with several baseline or time-independent factors (treatment, diseased *vs.* healthy groups, sex, age, dose level, etc.).

Existing methods were mainly focused on models developed under certain parametric assumptions on the underlying distribution or covariance structure. Inferences for these models were mainly obtained under the setting of a large number of clusters with small number of within-cluster measurements (often referred to as longitudinal data) (Davidian and Giltinan 1995; Hand and Crowder 1996; Brunner, Domhof, and Langer 2002; Diggle, Heagerty, Liang, and Zeger 2002; and the references therein). The major obstacle to the successful application of such methods to functional data is the estimation of within-cluster correlations. Because the sample covariance matrix is not invertible when the dimensionality is larger than the sample sizes (Lin and Perlman 1985; Johnstone 2001), it is a common practice to pick a stationary covariance matrix from a parametric family with few parameters provided by popular software packages. Such estimators of the covariance matrix could have considerable bias when the selected structure is far from the truth. An alternative solution is to model functional data as aggregates of statistically independent components. For example, Fan, Fan, and Lv (2008) considered a multi-factor model to reduce dimensionality and to estimate the covariance matrix. Another example is that stationary series can be transformed to approximately independent Gaussian series after Fourier transformation (Brockwell and Davis 1991, Chapter 10). Fan and Lin (1998) is an example of testing for the simple effect of treatment in a high-dimensional fixed effects ANOVA model with completely balanced designs and generalising naturally to mixed model with stationary correlation after Fourier transformation. However, the crucial assumption of stationarity is often not satisfied in practice. Bayesian covariance estimators have been recently considered to improve small sample performance (Schäfer and Strimmer 2005; Zhu and Hero 2007).

To analyse correlated categorical data, the generalised estimating equation (GEE) approach is among the most popular methods. This approach assumes distributions belonging to the exponential family and relies on a generalised linear model and the use of a sandwich estimator of variance and a working correlation matrix to describe dependencies within clusters. Under a correct model for the marginal mean, this method results in consistent regression parameters and their variances even if the assumed working correlation matrix is misspecified. However, use of the sandwich estimator for inference requires a large number of clusters, and otherwise may give inflated type I error rates (Lin and Wei 1989; Emrich and Piedmonte 1992; Gunsolley, Getchell, and Chinchilli 1995; Fay, Graubard, Freedman, and Midthune 1998). A few authors considered bias-corrected covariance estimates to improve GEE coverage probability when the number of unknown parameters is smaller than the number of clusters (Fay and Graubard 2001; Kauermann and Carroll 2001; Mancl and DeRouen 2001). Even when the number of clusters is large and the number of parameters is small, the sandwich estimator can be inefficient compared with model-based classical variance estimates (Kauermann and Carroll 2001). In fact, as shown in (Kauermann and Carroll 2001), the sandwich estimator in univariate Poisson log-linear regression has efficiency decreasing to 0 as the background event rate increases. In addition to the aforementioned efficiency problem, the link function and variance function in the generalised linear model have to follow specific forms depending on the underlying distribution. In practice, however, such a variance function is often inadequate to describe the variations among the data and there are often over- and under-dispersion problems.

In this paper, we propose a nonparametric model for nested heteroscedastic high-dimensional data. Unlike the usual nested mixed-effects model, the random components (effects and error) are not assumed to be normal and the random effects are not assumed to be uncorrelated. Thus, the model pertains to both continuous and discrete numerical data. In the context of this model, we propose new test statistics to test for effects that need to be described by a large number of parameters (e.g. main time effects and interactions between time and treatment). The asymptotic distribution of the test statistics is given under nonstationary α-mixing condition for the error terms, which cover most common time-series models. The test statistics are in the form of a difference of two quadratic forms and have a limiting normal distribution. The asymptotic theory requires novel techniques that rely on the number of within-cluster measurements tending to infinity under an unknown general covariance structure. The number of subjects can be small or large and the convergence rate of the test statistics does not depend on the number of subjects being large or small. This is achieved through the special construction of the test statistics in which the number of subjects appear in the numerators and denominators in the same order.

Simulation studies confirm that our methods are effective under general conditions. The present inference procedures can be used in conjunction with existing descriptive methods such as principal component analysis and dynamic time warping; see (Ramsay and Silverman 1997; Ramsay and Li 1998; Ke and Wang 2001) or smoothing spline mixed-effects model for data from normal distribution with constant variance for both the error terms and random effects (Brumback and Rice 1998).

The rest of the paper is organised as follows. Section 2 describes the model and hypotheses. Section 3 is devoted to results about the test statistics and their asymptotic distributions under the null hypotheses and local alternatives. Simulation results and real applications are presented in Section 5 followed by concluding remarks. The proofs are given in Appendix.

2. A nonparametric model for nested high-dimensional data

Consider subjects nested within a total of a factor levels such that each subject is measured at b time points. Thus, a randomly selected subject $T_{(i)}$ nested within level i generates a time series

$$\mathbf{X}_{iT_{(i)}} = (X_{i1T_{(i)}}, \ldots, X_{ibT_{(i)}})', \quad i = 1, \ldots, a.$$

Let j index the time points and set

$$\mu_{ijT_{(i)}} = E(X_{ijT_{(i)}} | T_{(i)}), \quad \mu = (ab)^{-1} \sum_i \sum_j E(\mu_{ijT_{(i)}}). \tag{1}$$

The random means $\mu_{ijT_{(i)}}$ can be decomposed to yield the fixed main and interaction effects (α_i, β_j, γ_{ij}) and a nested subject-specific intercept ($B_{k(i)}$):

$$\alpha_i = b^{-1} \sum_j E(\mu_{ijT_{(i)}}) - \mu, \quad B_{iT_{(i)}} = b^{-1} \sum_j \mu_{ijT_{(i)}} - \mu - \alpha_i,$$
$$\beta_j = a^{-1} \sum_i E(\mu_{ijT_{(i)}}) - \mu, \quad \gamma_{ij} = E(\mu_{ijT_{(i)}}) - \mu - \alpha_i - \beta_j. \tag{2}$$

Note that these quantities do not add up to $\mu_{ijT_{(i)}}$, and the remainder term $D_{ijT_{(i)}} = \mu_{ijT_{(i)}} - \mu - \alpha_i - \beta_j - B_{iT_{(i)}} - \gamma_{ij}$ reflects a random interaction between the subject and time that is not identifiable unless there are replicated measurements from the same subject at identical time points. For such a reason, we denote a general error term that is not independent:

$$u_{ijT_{(i)}} = X_{ijT_{(i)}} - \mu - \alpha_i - \beta_j - B_{iT_{(i)}} - \gamma_{ij}.$$

From its definition, it is easily established that

$$\sum_i \alpha_i = \sum_j \beta_j = \sum_i \gamma_{ij} = \sum_j \gamma_{ij} = E(B_{iT_{(i)}}) = E(u_{ijT_{(i)}}) = 0. \tag{3}$$

This set of effects is unique, subject to conditions (3). Note also that the random subject effects $B_{iT_{(i)}}$ and the error $u_{ijT_{(i)}}$ are generally not uncorrelated without making further assumptions.

Suppose now that we have n_i randomly selected subjects, $T_{k(i)},\ k = 1, \ldots, n_i$, nested in level i. For simplicity, set

$$\mathbf{X}_{ik} = \mathbf{X}_{iT_{k(i)}}, \quad X_{ijk} = X_{ijT_{k(i)}}, \quad B_{k(i)} = B_{iT_{k(i)}}, \quad u_{ijk} = u_{ijT_{k(i)}}.$$

Using the above simplified notation and combining Equations (1) and (2), we have

$$X_{ijk} = \mu + \alpha_i + \beta_j + B_{k(i)} + \gamma_{ij} + u_{ijk}. \tag{4}$$

We assume that the time series $\mathbf{X}_{ik}$ are independent for different k and i, and that $X_{ijk} \sim F_{ij}$, some unknown distribution function that depends on i and j.

Though Equation (4) reminds us of Gaussian models, we do not assume normality or homoscedasticity of the random effects and error terms. In fact, the model is completely nonparametric and applies to both continuous and discrete numerical data. Its advantage over the generalised linear model is that there is no need to specify a link function and the variance does not need to be restricted to be a function of the mean.

Under the above nonparametric model, we develop test procedures for the hypotheses

$$H_0(\beta): \text{ all } \beta_j = 0, \quad H_0(\gamma): \text{ all } \gamma_{ij} = 0. \tag{5}$$

The test procedures to be developed apply to data sets having b large and a small. The sample sizes n_i are allowed to be small. To allow for asymptotics, we assume that the time series $u_{ijk},\ j = 1, \ldots, b$, satisfy an α-mixing condition. That is, assume for some sequence $\alpha_m \to 0$,

$$|P(A \cap C) - P(A)P(C)| \le \alpha_m$$

holds for all $A \in \sigma(u_{i1k}, \ldots, u_{i\ell k})$, $C \in \sigma(u_{i,\ell+m,k}, u_{i,\ell+m+1,k}, \ldots)$, and all i, k, where $\sigma(\cdot)$ denotes the σ-field generated by the indicated random variables. The α-mixing condition basically requires the correlation between observations on the same time series to decay as the time lag m increases (Billingsley 1995, p. 365). We will make the assumption that the decay rate is $\alpha_m = O(m^{-5})$, which is weaker than the commonly used exponential decay rate of an autoregressive (AR) covariance structure. Many common time-series models have been shown to satisfy this condition. In particular, both ARCH processes and additive AR processes with exogenous variables are α-mixing under some mild conditions (Masry and Tjostheim 1995, 1997). Both stationary and nonstationary time series are allowed in this paper.

The following two lemmas generalise the central limit theorem for stationary α-mixing processes in (Billingsley 1995) to nonstationary α-mixing processes. Their proofs are given in Appendix. They will be used extensively in establishing the asymptotic distribution of the test statistics.

LEMMA 2.1 *Suppose that $X_1, X_2, \ldots$ is α-mixing with $\alpha_m = O(m^{-5})$, and such that $E(X_i) = 0$, $\limsup_i E(X_i^{16}) < \infty$. Set $S_b = \sum_{j=1}^{b} X_j$. Then the following results hold*:

(a) *There exists K, such that*

$$\sum_{j<j'}^{b} E(X_j X_{j'}) \le Kb, \quad \sum_{j_1<j_2<j_3}^{b} E(X_{j_1}^2 X_{j_2} X_{j_3}) \le Kb^2, \quad \sum_{j_1<j_2<j_3<j_4}^{b} E(X_{j_1} X_{j_2} X_{j_3} X_{j_4}) \le Kb^2. \tag{6}$$

(b) *If $\lim_{b\to\infty} \mathrm{Var}(S_b)/b$ exists and is greater than 0, $S_b/\sqrt{\mathrm{Var}(S_b)} \xrightarrow{d} N(0, 1)$.*

LEMMA 2.2 *Let $\boldsymbol{X}_k = (X_{k1}, X_{k2}, \ldots,)'$, $k = 1, \ldots, L$, be a sequence of independent vectors satisfying the following condition: for each k, $\boldsymbol{X}_k$ is α-mixing with $\alpha_m = O(m^{-5})$ and $E(X_{ki}) = 0$,*

$\limsup_i E(X_{ki}^{16}) < \infty$. *Denote* $\bar{X}_{\cdot j} = L^{-1}\sum_{i=1}^{L} X_{ij}$. *Then there exists* K, *such that*

$$\sum_{j<j'}^{b} E(\bar{X}_{\cdot j}\bar{X}_{\cdot j'}) \le \frac{Kb}{L}, \quad \sum_{j_1<j_2<j_3}^{b} E(\bar{X}_{\cdot j_1}^2 \bar{X}_{\cdot j_2}\bar{X}_{\cdot j_3}) \le \frac{Kb^2}{L^2}, \quad \sum_{j_1<j_2<j_3<j_4}^{b} E(\bar{X}_{\cdot j_1}\bar{X}_{\cdot j_2}\bar{X}_{\cdot j_3}\bar{X}_{\cdot j_4}) \le \frac{Kb^2}{L^2}.$$

Further, define $S_{b2} = \sum_{j=1}^{b}\bar{X}_{\cdot j}$. *Then we have* $S_{b2}/\sqrt{Var(S_{b2})} \xrightarrow{d} N(0,1)$ *no matter* L *stays bounded or goes to infinity as* $b \to \infty$, *provided that* $LVar(S_{b2})/b$ *has nonzero limit.*

Throughout the rest of the paper, the following notation will be used:

$$n = \sum_{i=1}^{a} n_i, \quad N = nb, \quad \bar{X}_{ij.} = \frac{1}{n_i}\sum_{k=1}^{n_i} X_{ijk}, \quad \tilde{X}_{i..} = \bar{X}_{i..} = \frac{1}{bn_i}\sum_{j=1}^{b}\sum_{k=1}^{n_i} X_{ijk},$$

$$\tilde{X}_{...} = \frac{1}{ab}\sum_{i=1}^{a}\sum_{j=1}^{b}\frac{1}{n_i}\sum_{k=1}^{n_i} X_{ijk}, \quad \tilde{X}_{.j.} = \frac{1}{a}\sum_{i=1}^{a}\frac{1}{n_i}\sum_{k=1}^{n_i} X_{ijk}, \quad \bar{X}_{i.k} = \frac{1}{b}\sum_{j=1}^{b} X_{ijk}.$$

3. Test statistics and their asymptotic distributions

3.1. *The test statistics*

ANOVA-type statistics are commonly used in the homoscedastic case. Under heteroscedasticity, however, the two mean sum of squares do not have the same expectation. To remedy this, we introduce suitable variations of these sum of squares. In particular, these adjusted mean (AM) squares are defined as

$$\mathrm{AM}\beta = \frac{1}{b-1}\sum_{i=1}^{a}\sum_{j=1}^{b}(\tilde{X}_{.j.} - \tilde{X}_{...})^2, \tag{7}$$

$$\mathrm{AM}\gamma = \frac{1}{(a-1)(b-1)}\sum_{i=1}^{a}\sum_{j=1}^{b}(\bar{X}_{ij.} - \tilde{X}_{i..} - \tilde{X}_{.j.} + \tilde{X}_{...})^2, \tag{8}$$

$$\mathrm{AM}E = \frac{1}{a(b-1)}\sum_{i=1}^{a}\sum_{j=1}^{b}\frac{1}{n_i(n_i-1)}\sum_{k=1}^{n_i}(X_{ijk} - \bar{X}_{ij.} - \bar{X}_{i.k} + \tilde{X}_{i..})^2. \tag{9}$$

Proposition 3.1 *Let* $\sigma_{ijj'} = \mathrm{Cov}(u_{ijk}, u_{ij'k})$ *and* $\sigma_{ijj} = \sigma_{ij}^2 = \mathrm{Var}(u_{ijk})$. *Then*,

$$E(\mathrm{AM}E) = \frac{1}{a(b-1)}\sum_{i,j}\frac{\sigma_{ij}^2}{n_i} - \frac{1}{ab(b-1)}\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{j'=1}^{b}\frac{\sigma_{ijj'}}{n_i},$$

$E(\mathrm{AM}\beta) = E(\mathrm{AM}E)$, *under* $H_0(\beta)$, *and* $E(\mathrm{AM}\gamma) = E(\mathrm{AM}E)$, *under* $H_0(\gamma)$.

The proof of this proposition is straightforward. Set $F_\beta = \mathrm{AM}\beta/\mathrm{AM}E$, $F_\gamma = \mathrm{AM}\gamma/\mathrm{AM}E$. Motivated by Proposition 3.1, we propose to use $\sqrt{b}(F_\beta - 1)$ and $\sqrt{b}(F_\gamma - 1)$ as test statistics for $H_0(\beta)$, and $H_0(\gamma)$, respectively.

3.2. *Results under the null hypotheses*

The asymptotic theory of the test statistics is given first. The proof, given in Appendix, uses a generalised version of Hájek's projection method. After obtaining the projection and showing its asymptotic equivalence to the projected statistic, the asymptotic distribution of the projection is obtained through the central limit theorems for nonstationary α-mixing series given in Lemmas 2.1 and 2.2.

THEOREM 3.2 *Assume that for each i and k, u_{ijk}, $j = 1, 2, \ldots$ is α-mixing with $\alpha_m = O(m^{-5})$. In addition, assume that* $\limsup_j E[u_{ijk}^{32}] < \infty$. *Let $n(a) = \min_i\{n_i\}$. Then as $b \to \infty$ while a remains bounded, the limits of*

$$\zeta_1 = \frac{2}{a^2 b}\sum_{j=1}^{b}\sum_{j'=1}^{b}\sum_{i=1}^{a}\frac{\sigma_{ijj'}^2}{n_i(n_i-1)}, \quad \zeta_2 = \frac{2}{a^2 b}\sum_{j=1}^{b}\sum_{j'=1}^{b}\sum_{i\neq i'}^{a}\frac{\sigma_{ijj'}\sigma_{i'jj'}}{n_i n_{i'}} \tag{10}$$

exist regardless of whether the n_i are bounded or go to ∞ as $b \to \infty$. Moreover, with $\sigma^2 = \lim_{b\to\infty} E(\mathrm{AM}E)$ and $\sigma_*^2 = \lim_{b\to\infty} E(n(a)\mathrm{AM}E)$,

(1) *for $n_i \geq 2$ bounded,*
- *under $H_0(\beta)$, $\sqrt{b}(F_\beta - 1) \xrightarrow{d} N(0, \tau_\beta^2/\sigma^4)$, where $\tau_\beta^2 = \lim_{b\to\infty}(\zeta_1 + \zeta_2)$;*
- *under $H_0(\gamma)$, $\sqrt{b}(F_\gamma - 1) \xrightarrow{d} N(0, \tau_\gamma^2/\sigma^4)$, where $\tau_\gamma^2 = \lim_{b\to\infty}(\zeta_1 + \zeta_2/(a-1)^2)$;*

(2) *if $n_i \to \infty$ as $b \to \infty$, under the additional assumption $\max_i\{n_i\}/n(a) = O(1)$, we have*
- *under $H_0(\beta)$, $\sqrt{b}(F_\beta - 1) \xrightarrow{d} N(0, \tau_{\beta*}^2/\sigma_*^4)$, where $\tau_{\beta*}^2 = \lim_{b\to\infty} n^2(a)(\zeta_1 + \zeta_2)$;*
- *under $H_0(\gamma)$, $\sqrt{b}(F_\gamma - 1) \xrightarrow{d} N(0, \tau_{\gamma*}^2/\sigma_*^4)$, where $\tau_{\gamma*}^2 = \lim_{b\to\infty} n^2(a)(\zeta_1 + \zeta_2/(a-1)^2)$.*

Remark Under the extra (often unrealistic) assumption of stationarity, the assumption $\limsup_j E[u_{ijk}^{32}] < \infty$ becomes $\limsup_j E[u_{ijk}^{16}] < \infty$.

To estimate the limits of ζ_1 and ζ_2 which enter the expressions for the asymptotic variance, we find that it is not efficient to use all pairwise sample correlations in the estimation. Instead, the idea of thresholding to reduce noise can be applied to obtain a consistent estimator of ζ_1 and ζ_2. The following proposition states the result.

PROPOSITION 3.3 *For each $j = 1, \ldots, b$, let $C_u(j,h) = [\min\{b, j + b^h\}]$, $C_l(j,h) = [\max\{0, j - b^h\}]$, for some $0 < h < 1$, where $[x]$ denotes the largest integer less than or equal to x. Let $Y_{ijk} = X_{ijk} - \bar{X}_{i.k}$ and set*

$$\widehat{\zeta_1} = \frac{2}{a^2 b}\sum_{j=1}^{b}\sum_{j'=C_l(j,h)}^{C_u(j,h)}\sum_{i=1}^{a}\frac{\widehat{\sigma_{ijj'}^2}}{n_i(n_i-1)}, \quad \widehat{\zeta_2} = \frac{2}{a^2 b}\sum_{j=1}^{b}\sum_{j'=C_l(j,h)}^{C_u(j,h)}\sum_{i\neq i'}^{a}\frac{\widehat{\sigma}_{ijj'}\widehat{\sigma}_{i'jj'}}{n_i n_{i'}},$$

$$\widehat{\sigma}^2 = \frac{1}{a(b-1)}\sum_{i,j}\frac{\widehat{\sigma}_{ijj}}{n_i} - \frac{1}{ab(b-1)}\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{j'=C_l(j,h)}^{C_u(j,h)}\frac{\widehat{\sigma}_{ijj'}}{n_i}, \qquad \widehat{\sigma}_*^2 = n(a)\widehat{\sigma}^2,$$

where

$$\widehat{\sigma}_{ijj'} = \sum_{k=1}^{n_i}\frac{(Y_{ijk} - \bar{Y}_{ij.})(Y_{ij'k} - \bar{Y}_{ij'.})}{n_i - 1},$$

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

Then under the assumptions of Theorem 3.2, as $b \to \infty$, $n(a)^2(\widehat{\zeta_1} - \zeta_1) \xrightarrow{p} 0$ and $n(a)^2(\widehat{\zeta_2} - \zeta_2) \xrightarrow{p} 0$. If $b \to \infty$ and n_i stay bounded, then $\widehat{\sigma}^2$ is a consistent estimator of σ^2; if both b and n_i go to ∞, $\widehat{\sigma}_*^2$ is a consistent estimator of σ_*^2.

3.3. *Asymptotic results under local alternatives*

In this section, we give the asymptotic distribution of the test statistics under local alternatives when the number of within-cluster observations is large.

THEOREM 3.4 *Let $X_{ijk} = \mu + \alpha_i + \beta_{jb} + \gamma_{ij} + B_{k(i)} + u_{ijk}$ be the observations with the time effect given by the sequence of local alternatives $\beta_{jb} = b^{-1/4}\beta_j$. Define $\eta_1 = \lim_{b\to\infty} a\sqrt{b}/(b-1)\sum_{j=1}^{b}\beta_{jb}^2$, $\eta_1^* = \lim_{n(a),b\to\infty} n(a)\eta_1$, and let σ^2, σ_*^2, ζ_1 and ζ_2 be as given in Theorem* 3.2. *Then under the assumptions of Theorem* 3.2,

(1) *for $n_i \geq 2$ bounded, $\sqrt{b}(F_\beta - 1) \xrightarrow{d} N(\eta_1/\sigma^2, \tau_\beta^2/\sigma^4)$, where $\tau_\beta^2 = \lim_{b\to\infty}(\zeta_1 + \zeta_2)$;*
(2) *if $n_i \to \infty$ as $b \to \infty$, under the additional assumption $\max_i\{n_i\}/n(a) = O(1)$, we have $\sqrt{b}(F_\beta - 1) \xrightarrow{d} N(\eta_1^*/\sigma_*^2, \tau_{\beta*}^2/\sigma_*^4)$, where $\tau_{\beta*}^2 = \lim_{b\to\infty} n^2(a)(\zeta_1 + \zeta_2)$.*

THEOREM 3.5 *Let $X_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij,b} + B_{k(i)} + u_{ijk}$ be the observations with the interaction effect given by the sequence of local alternatives $\gamma_{ij,b} = b^{-1/4}\gamma_{ij}$. Define $\eta_2 = \lim_{b\to\infty}\sqrt{b}(a-1)^{-1}(b-1)^{-1}\sum_{i=1}^{a}\sum_{j=1}^{b}\gamma_{ij,b}^2$, $\eta_2^* = \lim_{n(a),b\to\infty} n(a)\eta_2$, and let σ^2, σ_*^2, ζ_1 and ζ_2 be as given in Theorem* 3.2. *Then under the assumption of Theorem* 3.2,

(1) *for $n_i \geq 2$ bounded, we have $\sqrt{b}(F_\gamma - 1) \xrightarrow{d} N(\eta_2/\sigma^2, \tau_\gamma^2/\sigma^4)$, where $\tau_\gamma^2 = \lim_{b\to\infty}(\zeta_1 + \zeta_2/(a-1)^2)$;*
(2) *if $n_i \to \infty$ as $b \to \infty$, under the additional assumption $\max_i\{n_i\}/n(a) = O(1)$, we have $\sqrt{b}(F_\gamma - 1) \xrightarrow{d} N(\eta_2^*/\sigma_*^2, \tau_{\gamma*}^2/\sigma_*^4)$, where $\tau_{\gamma*}^2 = \lim_{b\to\infty} n^2(a)(\zeta_1 + \zeta_2/(a-1)^2)$.*

4. Diagnostics about α-mixing condition

The α-mixing condition is assumed for $\mathbf{u}_{ik}$ as a stochastic process. Let $\widehat{\sigma}_{ijj'}$ be as defined in Proposition 3.3. Similar to the proof of Proposition 3.3, it can be shown that under the condition of Theorem 3.2, $\widehat{\sigma}_{ijj'}$ is a consistent estimator of $\sigma_{ijj'}$ as $b \to \infty$. Then the correlation matrix per group can be estimated by $\widehat{\text{corr}}_{ijj'} = \widehat{\sigma}_{ijj'}/\sqrt{\widehat{\sigma}_{ijj}\widehat{\sigma}_{ij'j'}}$. Note that $\widehat{\text{corr}}_{ijj'}$ is the sample correlation based on $Y_{ijk} = X_{ijk} - \bar{X}_{i.k}$, $k = 1, \ldots, n_i$. For small n_i, the estimator has a large variance. In addition, as $b \to \infty$, the dimension of the correlation matrix tends to ∞. Therefore, it becomes impractical to check whether the α-mixing condition holds for each j as $|j - j'| \to \infty$. Instead, we can combine the estimated correlations for each value of $|j - j'|$ in some way. Autocorrelation function is a choice. Empirical semi-variogram function provides an alternative to examine the lagged correlation (Diggle and Ribeiro 2007). A polynomial fit of the semi-variogram can be obtained (Pinheiro and Bates 2000) and compared with the required convergence rate in the α-mixing condition. This is illustrated in the analysis of soil water content data in Section 5.1.

5. Numerical studies

5.1. *Analysis of soil water content data*

Inappropriate soil water content can cause stress for plants and therefore lead to agricultural yield loss (Venuprasad, Lafitte, and Atlin 2007; Skinner 2008). An experiment was conducted at Kansas State University to compare the soil water content at a depth of 10 cm following random precipitation for no-till soil and chisel soil. Three plots were randomly assigned to each treatment and the soil water content from each plot was collected daily for 158 days. The top panel of

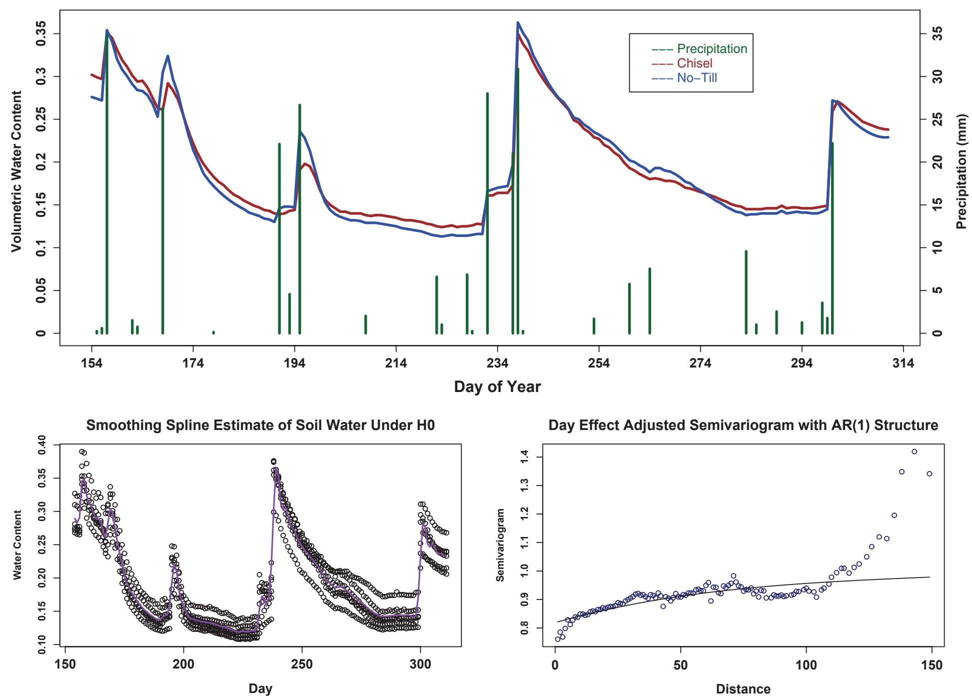

Figure 1. Top panel is the average soil water content taken over time for both treatments with three plots per treatment. Lower left panel gives the scatter plot and the smoothing spline estimate of the mean soil water content when no treatment effect exists. Lower right panel is the estimated semivariogram after the day effect from the smoothing spline fit is removed. The semi-variogram increases as the time distance increases indicates that the lag-m correlation decreases as m increases. An estimated variogram from AR(1) model is also plotted (black curve).

Figure 1 shows the average soil water content for the three plots over time at soil depth 10 cm for the two treatments. The scatter plot in the lower left panel of Figure 1 shows that the variations are different at different time points.

Note that the proposed test requires at least four replicates to obtain the estimator for ζ_1. If we assume $\sigma^2_{1jj'} \neq \sigma^2_{2jj'}$, the number of replicates is only three and we cannot use Proposition 3.3. We cannot perform any formal test to evaluate whether the two 158-dimensional covariance matrices are the same based on three replicates from each population. However, the soil water content under the two treatments follows nearly identical patterns, and the sample mean and variance from the two treatments are very close at each time point (mean-squared difference for the sample averages and variances at each time point between the two treatments is 1.2×10^{-4} and 5.48×10^{-7}, respectively). This leads us to assume that the within-plot covariance matrices are the same for the two treatments. Hence, we pulled the replicates from both treatments at each time point to obtain a common estimator of the squared covariances. That is, the estimate $\widehat{\sigma^2_{1jj'}} = \widehat{\sigma^2_{2jj'}}$ was obtained with $n = 6$ replications in Proposition 3.3:

$$\widehat{\sigma^2_{1jj'}} = \widehat{\sigma^2_{2jj'}} = \sum_{l_1 \neq l_2 \neq l_3 \neq l_4}^{n} \frac{(Y_{l_1 j} - Y_{l_2 j})(Y_{l_1 j'} - Y_{l_2 j'})(Y_{l_3 j} - Y_{l_4 j})(Y_{l_3 j'} - Y_{l_4 j'})}{4n(n-1)(n-2)(n-3)}.$$

The empirical semi-variogram of the water content data, given in the lower right panel of Figure 1, is obtained using the R package nlme after the day effect fitted with smoothing spline is removed. This plot shows an increasing pattern, which indicates that the lagged correlation is a decreasing function of the distance in time (since the semi-variogram is a decreasing function

Table 1. P-values for the main effect of time and treatment by time interaction effect for the soil water content data.

Effect	NPorg	lmeRan	lmeRanAR1	lmeRanGaus	glsAR1	geeInd	geequasi
Time	0	1.00	0.962	0.995	0.971	1.00	1.00
Trt:Time	0.882	1.00	0.965	0.990	0.998	1.00	1.00

Note: NPorg: the proposed test; lmeRan: 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: GLS with AR(1) serial correlation; geeInd: GEE with Gaussian family using independent working correlation; geequasi: GEE with quasi-likelihood using independent working correlation. For the proposed test, the samples from both treatments are pooled together in estimation of the asymptotic variances. Trt:time refers to treatment × time interaction.

of the lagged correlations). However, the whole profile does not resemble known semi-variogram models such as exponential, Gaussian, linear, rational quadratic or spherical correlation structure (Pinheiro and Bates 2000, p. 233). The black line is a fitted semi-variogram from an exponentially decaying model. Compared with the fitted value, the empirical semi-variogram increases more dramatically at both ends.

We apply the proposed test and some benchmark methods on the data set. Specifically, the linear mixed-effect (lme) model and generalised least-square (gls) methods with various covariance structures are applied to the data (R version 2.6.2 using packages *nlme* for lme model and gls method). GEE approach using package *gee* with independent working correlation using robust variance estimator and Gaussian family or quasi-likelihood is also considered. For other covariance structures, GEE fails to converge for this data set.

The p-values for all tests mentioned above for no main effect of time as well as the interaction effect are reported in Table 1. None of the tests give significant results for the interaction effect. The proposed test yields highly significant p-value for the main time effect while the other tests have p-values close to 1 though it can be seen clearly from the top panel of Figure 1 that the water content is very different over time. For the GEE method applied to this data set, the results can only be produced with the independence working correlation using either the Gaussian family or quasi-likelihood approach.

5.2. *Simulation study*

Here we use simulated data, which have features similar to those of the soil water content data set, to compare the performance of the proposed and the traditional tests used in Section 5.1 as the water content of the two soil treatments becomes increasingly different (in terms of treatment–time interaction and in terms of main time effects). All results reported in this section are based on 3000 runs. Throughout this section, the covariance function to be used for the simulation is given by

$$\sigma_{ijj'} = \sigma_{ij}\sigma_{ij'}\mathrm{e}^{-|j-j'|/m}, \quad \text{for } i = 1, \ldots, a, \quad j, j' = 1, \ldots, m. \tag{11}$$

This correlation structure is also used in (Wu and Chiang 2000; Diggle and Ribeiro 2007) to generate continuous data from the Gaussian process, though they considered the constant variance case. With this data-generation scheme, the lagged correlation reduces in exponential rate as the time lag increases. However, the correlation also depends on the total number of time points m. Therefore, the lagged correlation is at least $\mathrm{e}^{-(m-1)/m} \to \mathrm{e}^{-1} = 0.368$. That is, the reducing speed of the lagged correlation is faster than the α-mixing condition, but the lagged correlation does not reduce to zero as the time lag increases to ∞. Instead, we can think of the 0.368 as the correlation incurred due to a hidden random subject effect, though it is not specifically written in the model. Four replications are generated for each treatment.

For the null hypothesis of no interaction effect, we generated the data using multivariate normal (MVN) distribution

$$\mathrm{MVN}_b(\boldsymbol{\mu}_{01}, \Sigma), \tag{12}$$

where $\boldsymbol{\mu}_{01}$ is a b-dimensional vector of the pooled smoothing spline estimate of the soil water content from both treatments at all time points and Σ has (j, j') element given by Equation (11) with σ_{ij} being the pooled sample standard deviation at the jth time point from the soil water content data.

The proposed test (NPorg) and the likelihood-based test using lme model with a random intercept plus an AR(1) error structure (lmeRanAR1) have acceptable sizes (type I error estimate at level 0.01 is 0.003 for NPorg and 0 for lmeRanAR1). The lme with Gaussian serial correlation (lmeRanGaus) is slightly inflated as the type I error estimate at the 0.01 level is 0.023. The rest of the tests have type I error at least 0.063 at the 0.01 level. The inflation of type I error for GEE agrees with earlier findings by other authors when the cluster sizes are small. These type I error estimates at level 0.01 are presented in the top panel of Table 2 when θ value is labelled as $H01$.

For the power comparison when there is a treatment by time interaction effect, we generated the data in each treatment with MVN distributions:

$$\mathrm{MVN}(\boldsymbol{\mu}_{\mathrm{NT}}, \Sigma_1) \quad \text{and} \quad \mathrm{MVN}(\theta\boldsymbol{\mu}_{\mathrm{CH}}, \Sigma_2), \tag{13}$$

where $\boldsymbol{\mu}_{\mathrm{NT}}$ and $\boldsymbol{\mu}_{\mathrm{CH}}$ are the vectors of sample mean soil water content from the No-till soil and chisel soil, respectively. The value of θ ranges from 1 to 1.7. The elements of the covariance matrices Σ_i, $i = 1, 2$, are given by Equation (11) with σ_{ij}, $i = 1, 2$, being the sample standard deviation from the soil water content data at time point j in corresponding no-till soil and chisel soil, respectively (see Table 3 for the five-number summary of the mean and standard deviation over time).

The estimated power for testing of no treatment by time interaction effect at significance level 0.01 is given in the top panel of Table 2. The proposed method labelled as NPorg has superior power compared with the traditional methods, even though the traditional methods have inflated

Table 2. Proportion of rejections at level 0.01 for data with covariance structure given in Equation (11).

Effect	θ	NPOrg	lmeRan	lmeRanAR1	lmeRanGaus	glsAR1	geeInd	geequasi
Trt:time	$H01$	0.003	0.076	0	0.023	0.063	0.070	0.070
	1	0.049	0.110	0	0.044	0.126	0.086	0.086
	1.2	0.141	0.120	0	0.054	0.232	0.122	0.122
	1.3	0.349	0.161	0	0.079	0.298	0.164	0.164
	1.4	0.630	0.181	0	0.091	0.368	0.201	0.201
	1.5	0.842	0.211	0	0.106	0.437	0.256	0.256
	1.6	0.937	0.239	0	0.114	0.500	0.298	0.298
	1.7	0.983	0.266	0	0.131	0.581	0.346	0.346
Time	$H02$	0.0003	0.704	0.008	0.625	0.772	0.077	0.077
	$H01$	1.000	0.328	0	0.189	0.379	0.269	0.269
	1	1.000	0.427	0.0003	0.262	0.614	0.259	0.259
	1.2	1.000	0.541	0.0003	0.331	0.772	0.380	0.380
	1.3	1.000	0.594	0	0.373	0.822	0.435	0.435
	1.4	1.000	0.656	0	0.406	0.863	0.506	0.506
	1.5	1.000	0.710	0	0.440	0.912	0.569	0.569
	1.6	1.000	0.735	0	0.462	0.936	0.618	0.618
	1.7	1.000	0.782	0	0.507	0.958	0.683	0.683

Note: The top panel is for testing no treatment by time interaction effect and the bottom panel is for testing no main Time effect. $H01$ refers the data generation setting under H_0: no treatment by time interaction effect and $H02$ refers to the data generation setting under H_0: no time effect. The label of each test is same as in Table 1.

Table 3. The sample mean and sample standard deviation over time for the two treatments in water content data.

	Mean over time		Standard deviation over time	
	CH	NT	CH	NT
Mininum	0.124	0.113	0.014	0.001
First quantile	0.146	0.140	0.020	0.005
Median	0.168	0.171	0.027	0.007
Mean	0.195	0.193	0.027	0.009
Third quantile	0.244	0.240	0.031	0.013
Maximum	0.351	0.364	0.048	0.031

type I error estimates. As the power of the proposed method approaches 1, the power of all the traditional methods compared is still below 0.4. Among the traditional methods, gls method with AR(1) structure performs better than others in terms of power. GEE with Gaussian family assumption produced exactly the same result as the GEE with quasi-likelihood. This is expected as the data were generated from MVN distribution. The lme with a random intercept plus an AR(1) correlation structure does not have any power for any values of θ considered.

For the test of no main time effect, we generated the data in both treatments with MVN distributions. Under the null hypothesis, the data were generated from MVN$(\boldsymbol{\mu}_{0i} = \bar{x}_{i..}\mathbf{1}_b, \Sigma_{0i})$, $i = 1, 2$, where $\bar{x}_{1..} = 0.1926$ and $\bar{x}_{2..} = 0.1946$ are the average soil water content over all three plots at all time points from no-till soil and chisel soil, respectively; the elements of Σ_{0i} are given by Equation (11) with $\sigma_{1j} = 0.06614$ and $\sigma_{2j} = 0.06645$ being the sample standard deviation calculated using all observations from the no-till soil and chisel soil, respectively. The type I error estimate at level 0.01 is 0.0003 for the proposed test, 0.703 for lme with a random intercept, 0.0083 for lme with a random intercept plus an AR(1) error, 0.625 for lme with a random intercept plus a Gaussian serial correlation, 0.772 for gls with an AR(1) error, 0.077 for GEE with independent working correlation with Gaussian family or quasi-likelihood. They are given in the bottom panel of Table 2 when the θ value is labelled as $H02$. The lme with a random intercept plus an AR(1) model has accurate type I error estimate and the proposed test has an acceptable type I error estimate. All the other tests have wrong sizes. Note that in this data generation setting, the data are nearly homoscedastic. The liberality of the lme tests except lmeRanAR1 is mainly due to misspecification of the covariance structure and the four replications are too few to give a reliable estimate for a 158×158 covariance matrix. The GEE tests were not as liberal as the lme tests.

For the power estimate, we generate data using exactly the same model as given in Equation (13). The data are heteroscedastic since the variances of the observations at different time points are different. The proportion of rejections at level 0.01 is given in the bottom panel of Table 2. Note that the case with θ labelled as $H01$ generated from Equation (12) is now under the alternative hypothesis in the test of no main time effect. We can see that the proposed test outperforms the traditional methods for all values of θ in consideration. The lack of power behaviour under the alternatives is due to a combination of misspecification of the covariance structure, heteroscedastic variances over time and small replications. Same as the test for the interaction effect, the likelihood-based test of no main time effect for the lme model assuming a random intercept plus an AR(1) correlation does not have power for any θ value considered even though this test has the most accurate type I error estimate.

5.3. *Additional simulation study for data with other covariance structures*

We also conducted a simulation study to evaluate the performance of the proposed test when the data were generated with compound symmetry or unstructured covariance matrices. The data generation scheme is similar to the descriptions in Section 5.2, except for the correlation matrix. In the compound symmetry case, we let all the correlations be to 0.8 for observations from the same subject. For the unstructured case, we generated the correlation matrix by $\mathbf{e}_b\mathbf{e}_b' + \text{diag}_2(\mathbf{1}_b - \text{diag}_1(\mathbf{e}_b\mathbf{e}_b'))$, where $\mathbf{e}_b$ is a b-dimensional vector whose elements are iid from uniform (0, 1), $\mathbf{1}_b$ is a vector of ones of dimension b, diag_1 is a function to extract the diagonal elements of a matrix to form a vector, and diag_2 is a function to turn a vector into a diagonal matrix with diagonal elements given by the vector.

The proportion of rejections at level 0.01 for the compound symmetry case is given in the top panel of Table 4. The proposed test has similar performance as given in previous section. We remark that the model with the constant correlation is equivalent to the case where there is a random intercept with constant variance and the error terms are uncorrelated. As uncorrelated error is a special case of the α-mixing assumption, this is an extreme case of our proposed model. The data generated for testing no treatment and time interaction effect are heteroscedastic since the variances are different at different time points. Consequently, none of the lme tests have any power to detect the deviations from the null hypothesis. The glsAR1 test has acceptable type I error but only medium power. The GEE tests show good power but the type I errors remain inflated. Note that the data generation under the null hypothesis of no main time effect is almost perfect for the lme model with only a random intercept. As a result, all the lme tests and gls test have acceptable type I error estimates. Under the alternatives, lme with a random intercept (lmeRan) and glsAR1 both have good performance but not as powerful as the proposed test. This happens because the data under the alternatives are heteroscedastic and the lmeRan and glsAR1 were derived under the homoscedastic assumption.

For the unstructured covariance, the result is presented in the bottom panel of Table 4. It is interesting to see that the results follow a pattern similar to those for the compound symmetry case except that the proposed test is now slightly liberal under the null hypothesis with a type I error estimate of 0.031 for the test of no interaction effect and 0.026 of no main time effect at the 0.01 level. The other tests either have no power at all or less power with even worse type I error estimates.

Table 4. Proportion of rejections at 0.01 level when the data have compound symmetry (CS) or unstructured (UN) covariance matrix.

Cov	Effect	θ	NPOrg	lmeRan	lmeRanAR1	lmeRanGaus	glsAR1	geeInd	geeQuasi
CS	Trt:time	$H01$	0.0003	0	0	0	0.004	0.062	0.062
		1.2	0.170	0	0	0	0.082	0.540	0.540
		1.4	0.609	0.001	0	0	0.196	0.897	0.897
		1.6	0.937	0.004	0	0	0.378	0.993	0.993
		1.8	0.994	0.041	0	0.0003	0.572	1.000	1.000
	Time	$H02$	0	0.002	0.001	0.001	0.014	0.079	0.079
		$H01$	1.000	0.001	0	0	0.222	1.000	1.000
		1.2	1.000	0.686	0	0.055	0.749	1.000	1.000
		1.4	1.000	0.966	0	0.238	0.885	1.000	1.000
		1.6	1.000	0.999	0	0.568	0.971	1.000	1.000
		1.8	1.000	1.000	0	0.805	0.990	1.000	1.000
UN	Trt:time	$H01$	0.031	0	0	0	0.002	0.067	0.067
		1.2	0.301	0	0	0	0.019	0.468	0.468
		1.4	0.903	0	0	0	0.061	0.846	0.846
		1.6	0.999	0.008	0	0	0.132	0.979	0.979
	Time	$H02$	0.026	0.049	0.053	0.048	0.050	0.086	0.086
		$H01$	1.000	0.052	0	0.002	0.071	0.952	0.952
		1.2	1.000	0.578	0	0.155	0.374	0.993	0.993
		1.4	1.000	0.881	0	0.396	0.544	0.999	0.999
		1.6	1.000	0.980	0	0.650	0.704	1.000	1.000

Note: The legends are same as those in Table 2.

6. Summary

In this paper, we developed some inference for testing hypotheses in high-dimensional functional data. We began with a general model setup and proposed nonparametric test statistics for the test of no main effect of time and no interaction effects between the factor and time when there are heteroscedastic variances and unstructured within-subject correlations. The asymptotic distribution of the test statistics was given when the number of within-cluster measurements approaches infinity and the number of clusters can be small or large. In the case of a small number of clusters, the test statistics gain power from the asymptotics offered by a large number of repeated measurements. This is achieved through an extended Hájek projection method and a new central limit theorem on nonstationary α-mixing series shown in Lemma 2.1. That is, we turn the difficulty of high dimensionality into power for our advantage. The test statistics hold generally under the assumption of α-mixing condition for the the error term. Real application and simulation studies show that the proposed tests have superior power compared with the benchmark methods. As nonstationary data are allowed in our model, we anticipate that the results in this paper have broad applications in genomics as well as other high throughput screenings.

Acknowledgements

We thank the Editor, Associate Editor and two anonymous referees for their valuable constructive suggestions, which significantly helped improve the structure and presentation of this paper. We also thank Prof. James Higgins, Meghan Buckley and Prof. Gerard Kluitenberg at Kansas State University for sharing the soil water content data with us for illustration of the analysis. We are grateful to Prof. Stephen B. Vardeman from Iowa State University and Prof. Raymond J. Carroll from Texas A&M University for their helpful comments that led to essential improvement of the earlier versions of the manuscript.

References

Billingsley, P. (1995), *Probability and Measure* (3rd ed.), NewYork: Wiley.

Brockwell, P., and Davis, R. (1991), *Time Series: Theory and Methods* (2nd ed.), NewYork: Springer.

Brumback, A.B., and Rice, J. (1998), 'Smoothing Spline Models for the Analysis of Nested and Crossed Samples of Curves', *Journal of the American Statistical Association*, 93, 961–976.

Brunner, E., Domhof, S., and Langer, F. (2002), *Nonparametric Analysis of Longitudinal Data in Factorial Experiments*, NewYork: Wiley.

Davidian, M., and Giltinan, D.M. (1995), *Nonlinear Models for Repeated Measurement Data*, NewYork: Chapman and Hall.

Diggle, P., Heagerty, P., Liang, K.-Y., and Zeger, S. (2002), *Analysis of Longitudinal Data*, Oxford: Oxford University Press.

Diggle, P.J., and Ribeiro, P.J. (2007), *Model-Based Geostatistics*, New-York: Springer.

Emrich, L.J., and Piedmonte, M.R. (1992), 'On Some Small Sample Properties of Generalized Estimating Equation Estimates for Multivariate dichotomous Outcomes', *Journal of Statistical Computation and Simulation*, 41, 19–29.

Fan, J., and Lin, S. (1998), 'Test of Significance When Data are Curves', *Journal of the American Statistical Association*, 93, 1007–1021.

Fan, J., Fan, Y., and Lv, J. (2008), 'High Dimensional Covariance Matrix Estimation Using a Factor Model', *Journal of Econometrics*, 147, 186–197. *http://arxiv.org/abs/math.ST/0701124v1*.

Fay, M.P., and Graubard, B.I. (2001), 'Small-sample Adjustments for Wald-type Tests Using Sandwich Estimators', *Biometrics*, 57, 1198–1206.

Fay, M.P., Graubard, B.I., Freedman, L.S., and Midthune, D.N. (1998), 'Conditional Logistic Regression with Sandwich Estimators: Application to a Meta-analysis', *Biometrics*, 54, 195–208.

Gunsolley, J.C., Getchell, C., and Chinchilli, V.M. (1995), 'Small Sample Characteristics of Generalized Estimating Equations', *Communications in Statistics – Simulation and Computation*, 24, 869–878.

Hand, D., and Crowder, M. (1996), *Practical Longitudinal Data Analysis*, New-York: Chapman and Hall.

Johnstone, I.M. (2001), 'On the Distribution of the Largest Eigenvalue in Principal Components Analysis', *Annals of Statistics*, 29, 295–327.

Kauermann, G., and Carroll, R.J. (2001), 'A Note on the Efficiency of Sandwich Covariance Matrix Estimation', *Journal of the American Statistical Association*, 96, 1387–1396.

Ke, C., and Wang, Y. (2001), 'Semiparametric and Nonlinear Mixed-effects Models and their Applications', *Journal of the American Statistical Association*, 96, 1272–1298.

Lin, S.P., and Perlman, M.D. (1985), 'A Monte Carlo Comparison of Four Estimators of a Covariance Matrix', in *Multivariate Analysis* 6, eds. P.R. Krishnaiah, North-Holland, Amsterdam, pp. 411–429.
Lin, D.Y., and Wei, L.J. (1989), 'The Robust inference for the Cox Proportional hazards Model', *Journal of the American Statistical Association*, 84, 1074–1078.
Mancl, L.A., and DeRouen, T.A. (2001), 'A Covariance Estimator for Gee with Improved Small-sample Properties', *Biometrics*, 57, 126–134.
Masry, E., and Tjostheim, D. (1995), 'Nonparametric Estimation and Identification of Nonlinear Arch Time Series: Strong Convergence and Asymptotic Normality', *Econometric Theory*, 11, 258–289.
Masry, E., and Tjostheim, D. (1997), 'Additive Nonlinear ARX Time Series and Projection Estimates', *Econometric Theory*, 13, 214–252.
Pinheiro, J.C., and Bates, D. (2000), *Mixed-Effects Models in S and S-PLUS*, New-York: Springer.
Ramsay, J., and Li, X. (1998), 'Curve Registration', *Journal of the Royal Statistical Society. Series B*, 60, 351–363.
Ramsay, J., and Silverman, B. (1997), *Functional Data Analysis*, New-York: Springer.
Schäfer, J., and Strimmer, K. (2005), 'A Shrinkage Approach to Large-scale Covariance Matrix Estimation and Implications for Functional Genomics', *Statistical Applications in Genetics and Molecular Biology*, 4, 1–32.
Skinner, R.H. (2008), 'Yield, Root Growth, and Soil Water Content in Drought-stressed Pasture Mixtures Containing Chicory', *Crop Science*, 48, 380–388.
Venuprasad, R., Lafitte, H.R., and Atlin, G.N. (2007), 'Response to Direct Selection for Grain Yield Under Drought Stress in Rice', *Crop Science*, 47, 285–293.
Wang, H. (2004), 'Testing in Multifactor Heteroscedastic ANOVA and Repeated Measures Designs with Large Number of Levels', Ph.D. thesis, Pennsulvania State University.
Wu, C.O., and Chiang, C.T. (2000), 'Kernel Smoothing on Varying Coefficient Models with Longitudinal Dependent Variable', *Statistica Sinica*, 10, 433–456.
Zhu, D., and Hero, A. (2007), 'Bayesian Hierarchical Model for Large-scale Covariance Matrix Estimation', *Journal of Computational Biology*, 14, 1311–1326.

Appendix. Proofs

In this section, we give sketch of technical proofs for the lemmas and theorems. All omitted details can be found in Wang (2004).

Proof of Lemma 2.1 (a) By the given condition $\limsup_i E(X_i^{16}) < \infty$ and Lemma 3 of (Billingsley 1995, p. 377), there exists K_1, such that

$$\begin{aligned}
\sum_{j<j'}^{b} E(X_j X_{j'}) &= \sum_{j'=2}^{b}\sum_{j=1}^{j'-1} E(X_j X_{j'}) \\
&\le \sum_{j'=2}^{b}\sum_{j=1}^{j'-1} 8[1+E(X_j^4)+E(X_{j'}^4)]\alpha_{j'-j}^{1/2} \le K_1 \sum_{j'=2}^{b}\sum_{j=1}^{j'-1}\alpha_{j'-j}^{1/2} \\
&= K_1\sum_{m=1}^{b-1}(b-m)\alpha_m^{1/2} \le 2K_1 b\sum_{m=1}^{\infty}\alpha_m^{1/2} \le Kb,
\end{aligned}$$

since $\sum_{j_1=1}^{\infty}\alpha_{j_1}^{1/2} = O(\sum_{m=1}^{\infty} m^{-5/2}) < \infty$.

$$\begin{aligned}
\sum_{j_1<j_2<j_3}^{b} E(X_{j_1}^2 X_{j_2} X_{j_3}) &\le \sum_{j_1<j_2<j_3}^{b} 8[1+E(X_{j_1}^2)^4 + E(X_{j_2}X_{j_3})^4]\alpha_{j_2-j_1}^{1/2} \\
&\le K_1 \sum_{j_1<j_2<j_3}^{b}\alpha_{j_2-j_1}^{1/2} \le K_1 b\sum_{j_1<j_2}^{b}\alpha_{j_2-j_1}^{1/2} \le Kb^2.
\end{aligned}$$

For $j_1 < j_2 < j_3 < j_4$, by α-mixing condition, there exists K_2, such that

$$\begin{aligned}
|E(X_{j_1}X_{j_2}X_{j_3}X_{j_4})| &\le 8[1+E(X_{j_1}X_{j_2})^4 + E(X_{j_3}X_{j_4})^4]\alpha_{j_3-j_2}^{1/2} \\
&\le 8[1+[E(X_{j_1}^8)]^{1/2}[E(X_{j_2}^8)]^{1/2} + [E(X_{j_3}^8)]^{1/2}[E(X_{j_4}^8)]^{1/2}]\alpha_{j_3-j_2}^{1/2} \\
&\le K_2\alpha_{j_3-j_2}^{1/2}.
\end{aligned}$$

Also,

$$|E(X_{j_1}X_{j_2}X_{j_3}X_{j_4})| \le 8[1 + E(X_{j_1}^4) + E(X_{j_2}X_{j_3}X_{j_4})^4]\alpha_{j_2-j_1}^{1/2} \le K_2\alpha_{j_2-j_1}^{1/2}.$$

So

$$|E(X_{j_1}X_{j_2}X_{j_3}X_{j_4})| \le K_2 \min\left(\alpha_{j_2-j_1}^{1/2}, \alpha_{j_3-j_2}^{1/2}\right),$$

and

$$\begin{aligned}
\sum_{j_1<j_2<j_3<j_4}^{b} E(X_{j_1}X_{j_2}X_{j_3}X_{j_4}) &\le K_2 \sum_{j_1<j_2<j_3<j_4}^{b} \min\left(\alpha_{j_2-j_1}^{1/2}, \alpha_{j_3-j_2}^{1/2}\right)\\
&\le K_2 b \sum_{j_1<j_2<j_3}^{b} \min\left(\alpha_{j_2-j_1}^{1/2}, \alpha_{j_3-j_2}^{1/2}\right)\\
&\le K_2 b \sum_{j_3=3}^{b} \sum_{t\ge1, m\ge1, t+m<j_3} \min\left(\alpha_m^{1/2}, \alpha_t^{1/2}\right) \qquad \text{(A1)}\\
&\le 2K_2 b \sum_{j_3=3}^{b} \sum_{m\le t, t+m<j_3} \alpha_t^{1/2} = 2K_2 b \sum_{j_3=3}^{b}\sum_{t=1}^{j_3} t\alpha_t^{1/2}\\
&\le 2K_2 b^2 \sum_{t=1}^{\infty} t\alpha_t^{1/2} \le Kb^2, \qquad \text{(A2)}
\end{aligned}$$

where Equation (A1) is because $\alpha_m^{1/2} \ge \alpha_t^{1/2}$ when $m \le t$, and Equation (A2) is because $\sum_{t=1}^{\infty} t\alpha_t^{1/2} < \infty$.

(b) Note that by Equation (6), we have

$$\begin{aligned}
E(S_b^2) &= \sum_{j=1}^{b}\sum_{j'=1}^{b} E(X_j X_{j'})\\
&= \sum_{j=1}^{b} E(X_j^2) + 2\sum_{j<j'}^{b} E(X_j X_{j'}) \le Kb.
\end{aligned}$$

Thus,

$$E(\text{Var}(S_b/\sqrt{b})) = b^{-1}E(S_b^2) = O(1). \qquad \text{(A3)}$$

To verify the asymptotic distribution, we split the sum $X_1 + \cdots + X_b$ into alternate blocks of length B_b (the big blocks) and L_b (the small blocks). Namely, let

$$U_{bj} = X_{(j-1)(B_\text{b}+L_\text{b})+1} + \cdots + X_{(j-1)(B_\text{b}+L_\text{b})+B_\text{b}}, \quad 1 \le j \le r_\text{b}, \qquad \text{(A4)}$$

where r_b is the largest integer j for which $(j-1)(B_\text{b}+L_\text{b}) + B_\text{b} < b$. Further, let

$$V_{bj} = X_{(j-1)(B_\text{b}+L_\text{b})+B_\text{b}+1} + \cdots + X_{j(B_\text{b}+L_\text{b})}, \quad 1 \le j < r_\text{b}, \qquad \text{(A5)}$$

$$V_{br_\text{b}} = X_{(r_\text{b}-1)(B_\text{b}+L_\text{b})+B_\text{b}+1} + \cdots + X_\text{b}. \qquad \text{(A6)}$$

Then $S_b = \sum_{j=1}^{r_\text{b}} U_{bj} + \sum_{j=1}^{r_\text{b}} V_{bj}$, and the technique will be to choose the L_b small enough that $\sum_{j=1}^{r_\text{b}} V_{bj}$ is small in comparison with $\sum_{j=1}^{r_\text{b}} U_{bj}$ but large enough that the U_{bj} are nearly independent, so that Lyapounov's theorem can be adapted to prove $\sum_{j=1}^{r_\text{b}} U_{bj}/\sqrt{\text{Var}(S_b)}$ asymptotically normal.

Take $B_\text{b} = [b^{3/4}]$ and $L_\text{b} = [b^{1/4}]$. If r_b is the largest integer j such that $(j-1)(B_\text{b}+L_\text{b}) + B_\text{b} < b$, then

$$B_\text{b} \sim b^{3/4}, \quad L_\text{b} \sim b^{1/4}, \quad r_\text{b} \sim b^{1/4}. \qquad \text{(A7)}$$

Applying Equation (A3) on V_{bj}, we have $E(V_{bj})^2 \leq KL_{\rm b}$, and so (see Wang (2004) for details)

$$\frac{1}{\sqrt{\mathrm{Var}(S_b)}}\sum_{j=1}^{r_{\rm b}-1} V_{bj} = o_p(1), \quad \frac{V_{br_{\rm b}}}{\sqrt{\mathrm{Var}(S_b)}} = o_p(1).$$

It suffices to show that $1/\sqrt{\mathrm{Var}(S_b)}\sum_{j=1}^{r_{\rm b}} U_{bj} \xrightarrow{d} N(0,1)$. Let U'_{bj}, $1 \leq j \leq r_{\rm b}$, be independent random variables having the same distribution as U_{bj}. Consider the difference in the characteristic function of $\sum_{j=1}^{r_{\rm b}} U_{bj}/\sqrt{\mathrm{Var}(S_b)}$ and of $\sum_{j=1}^{r_{\rm b}} U'_{bj}/\sqrt{\mathrm{Var}(S_b)}$.

By adding and subtracting terms and noticing that U'_{bj} has the same characteristic function as U_{bj}, we can write the difference in the characteristic functions as

$$\begin{aligned}
&E\left(\mathrm{e}^{it\sum_{j=1}^{r_{\rm b}} U_{bj}/\sqrt{\mathrm{Var}(S_b)}}\right) - E\left(\mathrm{e}^{it\sum_{j=1}^{r_{\rm b}} U'_{bj}/\sqrt{\mathrm{Var}(S_b)}}\right)\\
&\quad = E\left(\mathrm{e}^{it\sum_{j=1}^{r_{\rm b}} U_{bj}/\sqrt{\mathrm{Var}(S_b)}}\right) - E\left(\mathrm{e}^{it\sum_{j=2}^{r_{\rm b}} U_{bj}/\sqrt{\mathrm{Var}(S_b)}}\right)E\left(\mathrm{e}^{itU_{b1}/\sqrt{\mathrm{Var}(S_b)}}\right)\\
&\qquad + \left[E\left(\mathrm{e}^{\sum_{j=2}^{r_{\rm b}} itU_{bj}/\sqrt{\mathrm{Var}(S_b)}}\right) - E\left(\mathrm{e}^{\sum_{j=3}^{r_{\rm b}} itU_{bj}/\sqrt{\mathrm{Var}(S_b)}}\right)E\left(\mathrm{e}^{itU_{b2}/\sqrt{\mathrm{Var}(S_b)}}\right)\right]E\left(\mathrm{e}^{itU_{b1}/\sqrt{\mathrm{Var}(S_b)}}\right)\\
&\qquad + \cdots\\
&\qquad + \left[E\left(\mathrm{e}^{it(U_{b,r_{\rm b}}+U_{b,r_{\rm b}-1})/\sqrt{\mathrm{Var}(S_b)}}\right) - E\left(\mathrm{e}^{itU_{b,r_{\rm b}}/\sqrt{\mathrm{Var}(S_b)}}\right)E\left(\mathrm{e}^{itU_{b,r_{\rm b}-1}/\sqrt{\mathrm{Var}(S_b)}}\right)\right]\prod_{j=1}^{r_{\rm b}-2} E\left(\mathrm{e}^{itU_{bj}/\sqrt{\mathrm{Var}(S_b)}}\right).
\end{aligned} \tag{A8}$$

Using the α-mixing condition and applying Lemma 2 on page 376 of (Billingsley 1995), the absolute value of the real and imaginary parts of Equation (A8) is bounded by $8\alpha_{L_{\rm b}}$. Similarly, the absolute value of the real and imaginary parts of all other rows are bounded by $16\alpha_{L_{\rm b}}$ since

$$\prod_{j=1}^{k} E(\mathrm{e}^{itU_{bj}/\sqrt{\mathrm{Var}(S_b)}}) = E\left[\cos\left(\sum_{j=1}^{k}\frac{tU'_{bj}}{\sqrt{\mathrm{Var}(S_b)}}\right) + \mathrm{i}\sin\left(\sum_{j=1}^{k}\frac{tU'_{bj}}{\sqrt{\mathrm{Var}(S_b)}}\right)\right]$$

for all k. Therefore, both the real and imaginary parts of the difference of the two characteristic functions differ by at most $\sum_{j=1}^{r_{\rm b}} \alpha_{L_{\rm b}} = O(b^{-1})$. So $\sum_{j=1}^{r_{\rm b}} U_{bj}/\sqrt{\mathrm{Var}(S_b)}$ has the same asymptotic distribution as $\sum_{j=1}^{r_{\rm b}} U'_{bj}/\sqrt{\mathrm{Var}(S_b)}$.

It remains to show that $\sum_{j=1}^{r_{\rm b}} U'_{bj}/\sqrt{\mathrm{Var}(S_b)}$ has an asymptotic distribution $N(0,1)$.

We will verify Lyapounov's condition for $\delta = 2$. Applying Equation (A3) on U_{bj}, we have

$$\mathrm{Var}\left(\sum_{j=1}^{r_{\rm b}}\frac{U'_{bj}}{\sqrt{\mathrm{Var}(S_b)}}\right) = \sum_{j=1}^{r_{\rm b}}\frac{E(U^2_{bj})}{\mathrm{Var}(S_b)} \leq \sum_{j=1}^{r_{\rm b}}\frac{KB_{\rm b}}{\mathrm{Var}(S_b)} = O(1).$$

Lyapounov's condition will be satisfied if $L(U,b) = \sum_{j=1}^{r_{\rm b}} E(U'_{bj}/\sqrt{\mathrm{Var}(S_b)})^4 \to 0$. This is true because

$$\begin{aligned}
L(U,b) &= \sum_{j=1}^{r_{\rm b}}\frac{E(U'_{bj})^4}{(\mathrm{Var}(S_b))^2}\\
&\leq \sum_{j=1}^{r_{\rm b}}\frac{KB^2_{\rm b}}{(\mathrm{Var}(S_b))^2} = O(b^{-1/4}) \to 0,
\end{aligned}$$

which used inequality $E(S_m^4) \leq K_3 m^2$, for some finite K_3 independent of m (See Wang (2004) for proof). Thus we complete the proof of this lemma. ■

Proof of Lemma 2.2 The inequalities and part (1) of the lemma are easily shown by applying the results of Lemma 2.1 together with the independence among $\mathbf{X}_k$, $k = 1, \ldots, L$. Here we only show part (2) of the lemma. By Lemma 2.1,

$$E\left(LS_{b2}^2\right) = L^{-1} E\left[\sum_{k_1,k_2}^{L} \sum_{j_1,j_2}^{b} X_{k_1 j_1} X_{k_2 j_2}\right]$$
$$= L^{-1} \sum_{k=1}^{L} \sum_{j_1,j_2}^{b} E(X_{kj_1} X_{kj_2}) = O(b).$$

Lyapounov's condition can be shown since the following is true due to Equation (6):

$$l(b) = \sum_{k=1}^{L} E\left(\sqrt{\mathrm{Var}(S_{b2})}\frac{1}{L}\sum_{j=1}^{b} X_{kj}\right)^4$$
$$= \sum_{k=1}^{L} \frac{1}{L^4[E(S_{b2}^2)]^2} E\left(\sum_{j=1}^{b} X_{kj}\right)^4 = O(L^{-1}) \to 0.$$

So the proof is complete. ■

For the rest of this section, we let X_{ijk}, $i = 1, \ldots, a$, $j = 1, \ldots, b$, $k = 1, \ldots, n_i$ be as given in Theorem 3.2, and denote $\mathbf{u} = (u_{111}, \ldots, u_{11n_1}, \ldots, u_{121}, \ldots, u_{12n_1}, u_{ab1}, \ldots, u_{abn_a})'$ to be the vecotor of random interaction plus the error terms.

Lemma A.1 *Assume that* $\limsup_j E(u_{ijk}^{32}) < \infty$. *Then with* AM$E$ *defined in Equation* (9), *as* $b \to \infty$ *regardless of whether* n_i *are fixed or not*,

(a) $n(a)\sqrt{b}[\mathrm{AM}E - P_{\mathrm{AM}E}(\boldsymbol{u})] = o_p(1)$, *where* $P_{\mathrm{AM}E}$ *is defined as*

$$P_{\mathrm{AM}E}(\boldsymbol{u}) = \frac{1}{ab}\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n_i}\frac{u_{ijk}^2}{n_i^2} - \frac{1}{ab}\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k\neq k'}^{n_i}\frac{u_{ijk}u_{ijk'}}{n_i^2(n_i-1)}; \tag{A9}$$

(b) $n(a)\mathrm{AM}E \xrightarrow{p} \sigma_*^2$, *where* $\sigma_*^2 = \lim_{b\to\infty}(ab)^{-1}\sum_{i,j}\sigma_{ij}^2 n(a)/n_i$, *provided that the limit exists.*

Proof Note that $\mathrm{AM}E = P_{\mathrm{AM}E}(\mathbf{u}) + D_1(\mathbf{u}) + D_2(\mathbf{u})$, where

$$D_1(\mathbf{u}) = \frac{-1}{ab(b-1)}\sum_{i=1}^{a}\sum_{k=1}^{n_i}\sum_{j\neq j'}^{b}\frac{u_{ijk}u_{ij'k}}{n_i(n_i-1)}, \quad D_2(\mathbf{u}) = \frac{1}{ab(b-1)}\sum_{i=1}^{a}\sum_{j\neq j'}^{b}\frac{\bar{u}_{ij\cdot}\bar{u}_{ij'\cdot}}{n_i-1}. \tag{A10}$$

For part (a), we need to show that $n(a)D_l(\mathbf{u}) = o_p(b^{-1/2})$, $l = 1, 2$. Indeed,

$$E(n^2(a)D_1^2(\mathbf{u})) = \frac{n^2(a)}{a^2b^2(b-1)^2}\sum_{i=1}^{a}\sum_{k=1}^{n_i}\sum_{j\neq j'}^{b}\sum_{j_1\neq j_2}^{b}\frac{E(u_{ijk}u_{ij'k}u_{ij_1k}u_{ij_2k})}{n_i(n_i-1)n_i(n_i-1)}$$
$$+ \frac{n^2(a)}{a^2b^2(b-1)^2}\sum_{i=1}^{a}\sum_{i_1=1}^{a}\sum_{k\neq k_1}^{n_i}\sum_{j\neq j'}^{b}\sum_{j_1\neq j_2}^{b}\frac{E(u_{ijk}u_{ij'k})E(u_{i_1j_1k_1}u_{i_1j_2k_1})}{n_i(n_i-1)n_{i_1}(n_{i_1}-1)}.$$

By Lemma 2.1, $\sum_{j\neq j'}^{b} E(u_{ijk}u_{ij'k}) = O(b)$ and $\sum_{j\neq j'}^{b}\sum_{j_1\neq j_2}^{b} E(u_{ijk}u_{ij'k}u_{ij_1k}u_{ij_2k}) = O(b^2)$. So $E(n^2(a)D_1^2(\mathbf{u})) = O(b^{-2})$. Similarly, it can be shown that $E(n^2(a)D_2^2(\mathbf{u})) = O(b^{-2})$. This shows parts (a). Using (a), part (b) will follow from $n(a)P_{\mathrm{AM}E}(\mathbf{u}) \xrightarrow{p} \sigma_*^2$, or from

$$n(a)P_{\mathrm{AM}E} - \sum_{i,j}\frac{n(a)\sigma_{ij}^2}{abn_i} = \frac{n(a)}{ab}\sum_{i,j,k}\frac{u_{ijk}^2-\sigma_{ij}^2}{n_i^2} - \frac{n(a)}{ab}\sum_{i,j}^{a}\sum_{k\neq k'}^{n_i}\frac{u_{ijk}u_{ijk'}}{n_i^2(n_i-1)} = o_p(1).$$

See Wang (2004) for detailed proof of this. Thus we have completed the proof of part (b) of the Lemma for both cases. ■

LEMMA A.2 *Assume* $\limsup_j E(u_{ijk}^{32}) < \infty$. *Let* AM$\beta$ *and* AMγ *be defined in Equations* (7) *and* (8), *respectively. Then as* $b \to \infty$ *while a remains fixed, regardless of whether* n_i *stay fixed or* $n(a) \to \infty$,

(a) *under* $H_0(\beta)$, $n(a)\sqrt{b}(\mathrm{AM}\beta - P_\beta(\boldsymbol{u})) \xrightarrow{p} 0$, *where* $P_\beta(\boldsymbol{u})$ *is defined in Equation* (A11);
(b) *under* $H_0(\gamma)$, $n(a)\sqrt{b}(\mathrm{AM}\gamma - P_\gamma(\boldsymbol{u})) \xrightarrow{p} 0$, *where* $P_\gamma(\boldsymbol{u})$ *is defined in Equation* (A11).

Proof Note that $\mathrm{AM}\beta - P_\beta(\mathbf{u}) = -D_3(\mathbf{u})$, $\quad \mathrm{AM}\gamma - P_\gamma(\mathbf{u}) = \frac{D_3(\mathbf{u})-D_4(\mathbf{u})}{a-1}$, where

$$D_3(\mathbf{u}) = \frac{a}{b(b-1)}\sum_{j\neq j'}^{b} \tilde{u}_{.j.}\tilde{u}_{.j'.}, \quad D_4(\mathbf{u}) = \frac{1}{b(b-1)}\sum_{i=1}^{a}\sum_{j\neq j'}^{b} \bar{u}_{ij.}\bar{u}_{ij'.}.$$

The proof of $n(a)\sqrt{b}D_4(\mathbf{u}) = o_p(1)$ is similar to that of $n(a)\sqrt{b}D_2(\mathbf{u}) = o_p(1)$, by showing that $E(n^2(a)D_4^2(\mathbf{u})) = O(1/b^2)$ as in the proof of Lemma A1. So we only need to show that $n(a)\sqrt{b}D_3(\mathbf{u}) = o_p(1)$. Write $D_3(\mathbf{u}) = D_{31}(\mathbf{u}) + D_{32}(\mathbf{u}) + D_{33}(\mathbf{u})$, where $D_{31}(\mathbf{u}) = a^{-1}b^{-1}(b-1)^{-1}\sum_{i\neq i'}^{a}\sum_{j\neq j'}^{b}\bar{u}_{ij.}\bar{u}_{i'j'.}$,

$$D_{32}(\mathbf{u}) = \frac{1}{ab(b-1)}\sum_{i=1}^{a}\sum_{j\neq j'}^{b}\sum_{k\neq k'}^{n_i}\frac{u_{ijk}u_{ij'k'}}{n_i^2}, \quad D_{33}(\mathbf{u}) = \frac{1}{ab(b-1)}\sum_{i=1}^{a}\sum_{j\neq j'}^{b}\sum_{k=1}^{n_i}\frac{u_{ijk}u_{ij'k}}{n_i^2}.$$

That $n(a)\sqrt{b}D_{33}(\mathbf{u}) = o_p(1)$ follows by the same argument as $n(a)\sqrt{b}D_1(\mathbf{u}) = o_p(1)$, which is given in the proof of Lemma A1. By independence of the error terms from different subjects or groups,

$$E(n^2(a)bD_{32}^2(\mathbf{u})) = \frac{4n^2(a)}{a^2b(b-1)^2}\sum_{i=1}^{a}\sum_{j\neq j'}^{b}\sum_{k\neq k'}^{n_i}\sum_{j_1\neq j_2}^{b}\frac{E(u_{ijk}u_{ij_1k})E(u_{ij'k'}u_{ij_2k'})}{n_i^4} = O(b^{-1}),$$

where the last step used Lemma 2.1. That $n(a)\sqrt{b}D_{31}(\mathbf{u}) = o_p(1)$ can be shown similarly, and this completes the proof. ■

Proof of Theorem 3.2 By Lemma A1(b), we only need to consider the asymptotic distribution of $\sqrt{b}n(a)(\mathrm{AM}\beta - \mathrm{AM}E)$ and $\sqrt{b}n(a)(\mathrm{AM}\gamma - \mathrm{AM}E)$ under $H_0(\beta)$ and $H_0(\gamma)$, respectively. They will be obtained by the projection method. Namely, AMβ, AMγ and AME will be projected onto the class of random variables of the form $\sum_{j=1}^{b} g_j(\mathbf{u}_j)$, where $\mathbf{u}_j = (u_{1j1}, \ldots, u_{1jn_1}, \ldots, u_{ajn_a})'$ and g_j are measurable with $Eg_j^2(\mathbf{u}_j) < \infty$. The projection of AM$E$ is defined in Equation (A9) while those of AMβ under $H_0(\beta)$ and AMγ under $H_0(\gamma)$ are given by

$$P_\beta(\mathbf{u}) = \frac{a}{b}\sum_{j=1}^{b}\tilde{u}_{.j.}^2, \quad P_\gamma(\mathbf{u}) = \frac{1}{(a-1)b}\sum_{i=1}^{a}\sum_{j=1}^{b}\bar{u}_{ij.}^2 - \frac{a}{b(a-1)}\sum_{j=1}^{b}\tilde{u}_{.j.}^2, \tag{A11}$$

respectively. We note that the projections are only formal, in the sense that they are computed ignoring the dependence of the data. However, Lemmas A1 and A2 imply that $n(a)\sqrt{b}(\mathrm{AM}\beta - \mathrm{AM}E)$ and $n(a)\sqrt{b}(\mathrm{AM}\gamma - \mathrm{AM}E)$ have same asymptotic distribution as their projections $n(a)\sqrt{b}(P_\beta(\mathbf{u}) - P_{\mathrm{AM}E})$ and $n(a)\sqrt{b}(P_\gamma(\mathbf{u}) - P_{\mathrm{AM}E})$, under $H_0(\beta)$ and $H_0(\gamma)$, respectively. Letting

$$P_1 = \frac{1}{ab}\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k\neq k'}^{n_i}\frac{u_{ijk}u_{ijk'}}{n_i(n_i-1)}, \quad P_2 = \frac{1}{ab}\sum_{i\neq i'}^{a}\sum_{j=1}^{b}\bar{u}_{ij.}\bar{u}_{i'j.}, \tag{A12}$$

we note that $P_\beta(\mathbf{u}) - P_{\mathrm{AM}E} = P_1 + P_2$, $P_\gamma(\mathbf{u}) - P_{\mathrm{AM}E} = P_1 - P_2/(a-1)$. We will only derive the asymptotic distribution of $n(a)\sqrt{b}(P_\beta(\mathbf{u}) - P_{\mathrm{AM}E})$. That of $n(a)\sqrt{b}(P_\gamma(\mathbf{u}) - P_{\mathrm{AM}E})$ can be derived similarly. Write $n(a)\sqrt{b}(P_\beta(\mathbf{u}) - P_{\mathrm{AM}E}) = b^{-1/2}\sum_{j=1}^{b} W_j$, where

$$W_j = \frac{n(a)}{a}\sum_{i=1}^{a}\sum_{k\neq k'}^{n_i}\frac{u_{ijk}u_{ijk'}}{n_i(n_i-1)} + \frac{n(a)}{a}\sum_{i\neq i'}^{a}\bar{u}_{ij.}\bar{u}_{i'j.}.$$

We will apply Lemma 2.1 to prove the asymptotic normality. First note that $E(W_j) = 0$.

Further, we can show (see Wang (2004) for details) that

$$E(W_j^{16}) \leq \frac{n^{16}(a)2^{15}}{a}\left[\sum_{i=1}^{a}\frac{E\left(\sum_{k=1}^{n_i}u_{ijk}\right)^{32}}{n_i^{16}(n_i-1)^{16}} + a^2\sum_{i\neq i'}^{a}E(\bar{u}_{ij.})^{16}E(\bar{u}_{i'j.})^{16}\right] \tag{A13}$$

and

$$E\left(\sum_{k=1}^{n_i}u_{ijk}\right)^{32} \leq A_{32}n_i^{16}E(u_{ij1}^{32}); \quad E(\bar{u}_{ij.})^{16} \leq \frac{A_{16}}{n_i^8}E(u_{ij1}^{16}), \tag{A14}$$

where A_{32} and A_{16} are some finite-positive constants. Relations (A13) and (A14) together with the assumption $\limsup_j E(u_{ijk}^{32}) < \infty$ imply that $\limsup_j E(W_j^{16}) < \infty$ and thus Lemma 2.1 can be applied to yield the asymptotic normality of $n(a)\sqrt{b}(P_\beta(\mathbf{u}) - P_{\mathrm{AM}E})$ with asymptotic variance $\tau_{\beta*}^2 = \lim E(\sum_{j=1}^{b} W_j)^2/b = \mathrm{Var}(\sum_{j=1}^{b} W_j)/b$. ■

Proof of Proposition 3.3 Let $\hat{\zeta}_1(\mathbf{u})$, $\hat{\zeta}_2(\mathbf{u})$ be as defined in Proposition 3.3 but calculated based on the vector $\mathbf{u}$ instead of $\mathbf{Y}$. We will first show that $\hat{\zeta}_1(\mathbf{u}) - \zeta_1$ and $\hat{\zeta}_2(\mathbf{u}) - \zeta_2$ converge to zero in probability.

In this proof, we will use the inequality $|\sigma_{ijj'}| \leq K\alpha_{|j-j'|}^{1/2}$ (Billingsley 1995). Since $E(\widehat{\sigma_{ijj'}^2}) = \sigma_{ijj'}^2$, we have

$$\begin{aligned}|\zeta_1 - E(\hat{\zeta}_1(\mathbf{u}))| &\leq \frac{2}{a^2 b}\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{|j'-j|>b^h}\frac{K^2\alpha_{|j'-j|}}{n_i(n_i-1)} \\ &\leq \frac{2}{a^2}\sum_{i=1}^{a}\sum_{j_2>b^h}^{b}\frac{K^2\alpha_{j_2}}{n_i(n_i-1)} \to 0.\end{aligned}$$

Thus, to show $\hat{\zeta}_1(\mathbf{u}) - \zeta_1 \xrightarrow{p} 0$, it remains to show that $\mathrm{Var}(\hat{\zeta}_1(\mathbf{u})) \to 0$. We have

$$\mathrm{Var}(\hat{\zeta}_1(\mathbf{u})) = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{|j'-j|<b^h}\sum_{j_1=1}^{b}\sum_{|j_1'-j_1|<b^h}\frac{4[E(\widehat{\sigma_{ijj'}^2}(\mathbf{u})\widehat{\sigma_{ij_1j_1'}^2}(\mathbf{u})) - \sigma_{ijj'}^2\sigma_{ij_1j_1'}^2]}{a^4b^2n_i^2(n_i-1)^2}.$$

Through some calculation, it can be shown that

$$\begin{aligned}E(\widehat{\sigma_{ijj'}^2}(\mathbf{u})\widehat{\sigma_{ij_1j_1'}^2}(\mathbf{u})) - \sigma_{ijj'}^2\sigma_{ij_1j_1'}^2 &= \frac{(n_i-4)(n_i-5)\sigma_{ijj'}\sigma_{ij_1j_1'}}{(n_i-1)(n_i-2)(n_i-3)}[g_{ijj'j_1j_1'} - \sigma_{ijj'}\sigma_{ij_1j_1'}] \\ &\quad + \frac{[9g_{ijj'j_1j_1'} + 33\sigma_{ijj'}\sigma_{ij_1j_1'}][g_{ijj'j_1j_1'} - \sigma_{ijj'}\sigma_{ij_1j_1'}]}{n_i(n_i-1)(n_i-2)(n_i-3)},\end{aligned}$$

where $g_{ijj'j_1j_1'} = E[u_{ijk}u_{ij'k}u_{ij_1k}u_{ij_1'k}]$. Thus, it suffices to show

$$\sum_{j,j_1}^{b}\sum_{|j'-j|<b^h}\sum_{|j_1'-j_1|<b^h}\sigma_{ijj'}\sigma_{ij_1j_1'}[g_{ijj'j_1j_1'} - \sigma_{ijj'}\sigma_{ij_1j_1'}] = o(b^2), \tag{A15}$$

$$\sum_{j,j_1}^{b}\sum_{|j'-j|<b^h}\sum_{|j_1'-j_1|<b^h}[9g_{ijj'j_1j_1'} + 33\sigma_{ijj'}\sigma_{ij_1j_1'}][g_{ijj'j_1j_1'} - \sigma_{ijj'}\sigma_{ij_1j_1'}] = o(b^2). \tag{A16}$$

Relations (A15) and (A16) can be shown by considering seperately the summation when all four indices are different, when three indices are different, when two indices are different and when all indices are same. In each of the cases, the relations follow by considering certain bounds. For details, we refer to Wang (2004).

Next, we will show $\hat{\zeta}_2(\mathbf{u}) \xrightarrow{p} \zeta_2$. It is clear that $E(\hat{\zeta}_2(\mathbf{u})) = \zeta_2$. Write

$$\begin{aligned}\mathrm{Var}(\hat{\zeta}_2(\mathbf{u})) &= \sum_{i\neq i'}^{a}\sum_{i_1\neq i_1'}^{a}\sum_{j=1}^{b}\sum_{|j'-j|<b^h}\sum_{j_1=1}^{b} 4(a^4b^2 n_i n_{i'} n_{i_1} n_{i_1'})^{-1}\\ &\quad\times\sum_{|j_1'-j_1|<b^h}[E(\hat{\sigma}_{ijj'}(\mathbf{u})\hat{\sigma}_{i'jj'}(\mathbf{u})\hat{\sigma}_{i_1j_1j_1'}(\mathbf{u})\hat{\sigma}_{i_1'j_1j_1'}(\mathbf{u})) - \sigma_{ijj'}\sigma_{i'jj'}\sigma_{i_1j_1j_1'}\sigma_{i_1'j_1j_1'}]\\ &= \frac{8}{a^4b^2}\sum_{j=1}^{b}\sum_{|j'-j|<b^h}\sum_{j_1=1}^{b}\sum_{|j_1'-j_1|<b^h}\left\{\sum_{i\neq i'\neq i_1}^{a}\frac{[E(\hat{\sigma}_{ijj'}(\mathbf{u})\hat{\sigma}_{ij_1j_1'}(\mathbf{u})) - \sigma_{ijj'}\sigma_{ij_1j_1'}]}{\sigma_{i'jj'}^{-1}\sigma_{i_1j_1j_1'}^{-1}n_i^2 n_{i'}n_{i_1}}\right.\\ &\quad\left.+\sum_{i\neq i'}^{a}\frac{[E(\hat{\sigma}_{ijj'}(\mathbf{u})\hat{\sigma}_{ij_1j_1'}(\mathbf{u}))E(\hat{\sigma}_{i'jj'}(\mathbf{u})\hat{\sigma}_{i'j_1j_1'}(\mathbf{u})) - \sigma_{ijj'}\sigma_{i'jj'}\sigma_{ij_1j_1'}\sigma_{i'j_1j_1'}]}{n_i^2n_{i'}^2}\right\}. \qquad \text{(A17)}\end{aligned}$$

Straightforward calculations yield

$$\begin{aligned}E(\hat{\sigma}_{ijj'}(\mathbf{u})\hat{\sigma}_{ij_1j_1'}(\mathbf{u})) - \sigma_{ijj'}\sigma_{ij_1j_1'} &= \frac{1}{(n_i-1)^2}\left[\sum_{k_1,k_2}^{n_i}E(u_{ijk_1}u_{ij'k_1}u_{ij_1k_2}u_{ij_1'k_2})\right.\\ &\quad-\frac{1}{n_i}\sum_{k_1,k_2,k_3}^{n_i}E(u_{ijk_1}u_{ij'k_1}u_{ij_1k_2}u_{ij_1'k_3}) - \sum_{k_1,k_2,k_3}^{n_i}\frac{E(u_{ijk_1}u_{ij'k_2}u_{ij_1k_3}u_{ij_1'k_3})}{n_i}\\ &\quad\left.+\sum_{k_1,k_2,k_3,k_4}^{n_i}\frac{E(u_{ijk_1}u_{ij'k_2}u_{ij_1k_3}u_{ij_1'k_4})}{n_i^2}\right] - \sigma_{ijj'}\sigma_{ij_1j_1'}\\ &= \frac{1}{n_i}\left[E\left(u_{ijk_1}u_{ij'k_1}u_{ij_1k_1}u_{ij_1'k_1}\right) - \sigma_{ijj'}\sigma_{ij_1j_1'}\right] + \frac{\sigma_{ijj_1}\sigma_{ij'j_1'} + \sigma_{ijj_1'}\sigma_{ij'j_1}}{n_i(n_i-1)},\end{aligned}$$

and

$$\begin{aligned}&E(\hat{\sigma}_{ijj'}(\mathbf{u})\hat{\sigma}_{ij_1j_1'}(\mathbf{u}))E(\hat{\sigma}_{i'jj'}\hat{\sigma}_{i'j_1j_1'}) - \sigma_{ijj'}\sigma_{i'jj'}\sigma_{ij_1j_1'}\sigma_{i'j_1j_1'}\\ &\quad= [E(\hat{\sigma}_{ijj'}(\mathbf{u})\hat{\sigma}_{ij_1j_1'})(\mathbf{u}) - \sigma_{ijj'}\sigma_{ij_1j_1'}][E(\hat{\sigma}_{i'jj'}(\mathbf{u})\hat{\sigma}_{i'j_1j_1'}(\mathbf{u})) - \sigma_{i'jj'}\sigma_{i'j_1j_1'}] + [E(\hat{\sigma}_{ijj'}(\mathbf{u})\\ &\qquad\times\hat{\sigma}_{ij_1j_1'}(\mathbf{u})) - \sigma_{ijj'}\sigma_{ij_1j_1'}]\sigma_{i'jj'}\sigma_{i'j_1j_1'} + [E(\hat{\sigma}_{i'jj'}(\mathbf{u})\hat{\sigma}_{i'j_1j_1'}(\mathbf{u})) - \sigma_{i'jj'}\sigma_{i'j_1j_1'}]\sigma_{ijj'}\sigma_{ij_1j_1'}.\end{aligned}$$

Each of the terms on the right-hand side of Equation (A17) can be shown to be $o(1)$ by considering the bounds similar to those for Equations (A15) and (A16). Thus $\hat{\zeta}_2(\mathbf{u}) - \zeta_2 \to 0$

The proof will be complete by showing that $\hat{\zeta}_1(\mathbf{Y}) - \hat{\zeta}_1(\mathbf{u}) \to 0$ and $\hat{\zeta}_2(\mathbf{Y}) - \hat{\zeta}_2(\mathbf{u}) \to 0$ in probability. These follow similar arguments as the first part of the proof. The details are omitted. ■

Proof of Theorems 3.4 and 3.5 We only need to consider the additional terms existing in the test statistics under corresponding local alternatives. Note that

$$\begin{aligned}\mathrm{AM}\beta &= \frac{1}{b-1}\sum_{i=1}^{a}\sum_{j=1}^{b}(\tilde{X}_{.j.} - \tilde{X}_{...})^2 = \frac{1}{b-1}\sum_{i=1}^{a}\sum_{j=1}^{b}(\beta_{jb} + \tilde{u}_{.j.} - \tilde{u}_{...})^2\\ &= \frac{1}{b-1}\sum_{i=1}^{a}\sum_{j=1}^{b}\beta_{jb}^2 + \frac{1}{b-1}\sum_{i=1}^{a}\sum_{j=1}^{b}(\tilde{u}_{.j.} - \tilde{u}_{...})^2 + O_p(b^{-1/2}b^{-1/4}).\end{aligned}$$

Similarly, when the observation are given by $X_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij,b} + B_{k(i)} + u_{ijk}$,

$$\mathrm{AM}\gamma = \frac{1}{(a-1)(b-1)}\sum_{i=1}^{a}\sum_{j=1}^{b}\{(\tilde{u}_{ij.} - \tilde{u}_{i..} - \tilde{u}_{.j.} + \tilde{u}_{...})^2 + \gamma_{ij,b}^2\} + O_p(b^{-3/4}).$$

The rest of the proof follows from the results under corresponding null hypotheses. ■