ANOVA FOR LONGITUDINAL DATA WITH MISSING VALUES

By Song Xi Chen∗, Ping-Shou Zhong∗

Iowa State University and Peking University, Iowa State University†

We carry out an ANOVA analysis to compare multiple treatment effects for longitudinal studies with missing values. The treatment effects are modeled semiparametrically via a partially linear regression which is flexible in quantifying the time effects of treatments. The empirical likelihood is employed to formulate nonparametric ANOVA tests for treatment effects with respect to covariates and the nonparametric time-effect functions. The proposed tests can be readily modified for ANOVA tests when the underlying regression model for the longitudinal study is either parametric or nonparametric. The asymptotic distributions of the ANOVA test statistics are derived. A bootstrap procedure is proposed to improve the ANOVA test for the time-effect functions. We analyze an HIV-CD4 data set and compare the effects of four treatments.

1. Introduction. Randomized clinical trials and observational studies are often used to evaluate treatment effects. While the treatment versus control studies are popular, multi-treatment comparisons beyond two samples are commonly practiced in clinical trials and observational studies. In addition to evaluate overall treatment effects, investigators are also interested in intra-individual changes over time by collecting repeated measurements on each individual over time. Although most longitudinal studies are desired to have all subjects measured at the same set of time points, such “balanced” data may not be available in practice due to missing values. Missing values arise when scheduled measurements are not made, which make the data “unbalanced”. There is a good body of literature on parametric, semiparametric and semiparametric estimation for longitudinal data with or without missing values. This includes Liang and Zeger (1986), Laird and Ware (1982), Wu (1998, 2000), Fitzmaurice et al. (2004) for methods developed for longitudinal data without missing values; and Little and Rubin (2002), Little (1995), Laird (2004), Robins, Rotnitzky and Zhao (1995) for missing values.

The aim of this paper is to develop ANOVA tests for multi-treatment comparisons in longitudinal studies with or without missing values. Suppose that at time \( t \), corresponding to \( k \) treatments there are \( k \) mutually independent samples:

\[
\{(Y_{1i}(t), X_{1i}(t)) \}_{i=1}^{n_1}, \ldots, \{(Y_{ki}(t), X_{ki}(t)) \}_{i=1}^{n_k}
\]

where the response variable \( Y_{ji}(t) \) and the covariate \( X_{ji}(t) \) are supposed to be measured at time points \( t = t_{j11}, \ldots, t_{j1T_j} \). Here \( T_j \) is the fixed number of scheduled observations for the \( j \)-th treatment. However, \( \{Y_{ji}(t), X_{ji}(t)\} \) may not be observed at some times, resulting in missing values in either the response \( Y_{ji}(t) \) or the covariates \( X_{ji}(t) \).

∗The authors acknowledge support from NSF grants SES-0518904, DMS-0604563 and DMS-0714978.
Keywords and phrases: Analysis of Variance; Empirical likelihood; Kernel smoothing; Missing at random; Semiparametric model; Treatment effects.
We consider a semiparametric partially linear regression to model the treatment effects

\[(1.1) \quad Y_{ji}(t) = X^T_{ji}(t)\beta_{j0} + g_{j0}(t) + \varepsilon_{ji}(t), \quad j = 1, 2, \cdots, k\]

where \(\beta_{j0}\) are \(p\)-dimensional parameters representing covariate effects and \(g_{j0}(t)\) are unknown smooth functions representing the time effect, and \(\{\varepsilon_{ji}(t)\}\) are residuals time series. Such a semiparametric model has been used for longitudinal data analysis; see Zeger and Diggle (1994), Zhang, Lin, Raz and Sowers (1998), Lin and Ying (2001), Wang, Carroll and Lin (2005), Wu et al. (1998) and Wu and Chiang (2000) proposed estimation and confidence regions for a related semiparametric varying coefficient regression models.

Despite the substantial amount of works on estimation for longitudinal data with or without missing values, analysis of variance for longitudinal data with or without missing values have attracted much less attention. Among a few exceptions are Forcina (1992) who proposed an ANOVA test in a fully parametric setting; and Scheike and Zhang (1998) who considered ANOVA tests in a fully nonparametric setting with two treatments.

In this paper, we propose ANOVA tests for the semiparametric model (1.1) to test for differences among the \(\beta_{j0}\)'s and the baseline functions \(g_{j0}\)'s respectively. The ANOVA statistics are formulated based on the empirical likelihood (Owen, 1988 and 2001), which can be viewed as a nonparametric counterpart of the conventional parametric likelihood. We propose two empirical likelihood ANOVA tests: one for the equivalence of the treatment effects with respect to the covariate; and the other for the equivalence of the time effect functions \(g_{j0}(\cdot)\)'s. These produce, as special cases, an ANOVA test for parametric regression in the absence of the baseline time effect functions, and an ANOVA test for nonparametric regression in the absence of the parametric part in (1.1). We show that, as in the conventional parametric likelihood formulation of the ANOVA test as in Forcina (1992), the empirical likelihood ANOVA statistic is limiting chi-square distributed for testing the parametric covariate effects even in the presence of the missing values. For testing the nonparametric time effects, a bootstrap calibration that respects the longitudinal nature and missing values is proposed to obtain the critical values for the ANOVA test. We applied the proposed ANOVA tests to analyze an HIV-CD4 data set that consists of four treatments, and found significant differences among the treatments in both the covariates and the baseline time effect functions.

Empirical likelihood (EL) introduced by Owen (1988) is a computer-intensive statistical method which can facilitate either nonparametric or semiparametric inference. Despite its not requiring a fully parametric model, the empirical likelihood enjoys some nice properties of a conventional likelihood, like Wilks’ theorem (Owen 1990, Qin and Lawless 1994, Fan and Zhang 2004) and Bartlett correction (DiCiccio, Hall and Romano 1991; Chen and Cui 2006). For missing values, Wang et al. (2002, 2004) considered empirical likelihood inference with kernel regression imputation for missing responses, and Liang and Qin (2008) treated estimation for the partially linear models with missing covariates. Cao and Van Keilegom (2006) employed the empirical likelihood to formulate a two samples test for the equivalence of two probability densities. For longitudinal data, Xue and Zhu (2007a, 2007b) proposed a bias correction method so that the empirical likelihood statistic is asymptotically pivotal for the nonparametric part in a one sample partially linear model; You, Chen and Zhou (2007) applied the blocking technique in their formulation for a semiparametric model, and Huang, Qin and Follman (2008) considered estimation.

The paper is organized as below. In Section 2, we describe the models and the missing value mechanism. Section 3 outlines the ANOVA test for comparing treatment effects due to
the covariates; whereas that for the nonparametric time effects is given in Section 4 with a theoretical justification. The bootstrap calibration to the ANOVA test on the nonparametric part is outlined in Section 5. Section 6 reports simulation results and Section 7 analyze the HIV-CD4 data. All technical details are given in the Appendix.

2. Models, Hypotheses and Missing Values. For the i-th individual of the j-th treatment, the measurements taken at time \( t_{jim} \) follow a partially linear model

\[
Y_{ji}(t_{jim}) = X_{ji}(t_{jim})'\beta_j + g_j(t_{jim}) + \varepsilon_{ji}(t_{jim}),
\]

for \( j = 1, \ldots, k, \ i = 1, \ldots, n_j, \ m = 1, \ldots, T_j \). Here \( \beta_j \) are unknown \( p \)-dimensional parameters and \( g_j(t) \) are unknown functions representing the time effects of the treatments. The time points \( \{ t_{jim} \}_{m=1}^{T_j} \) are known design points. For each individual, the residuals \( \{ \varepsilon_{ji}(t) \} \) satisfy \( \text{Var}(\varepsilon_{ji}(t)|X_{ji}(t)) = \sigma_j^2(t) \) and

\[
\text{Cov}(\varepsilon_{ji}(t), \varepsilon_{ji}(s)|X_{ji}(t), X_{ji}(s)) = \rho_j(s,t)\sigma_j(t)\sigma_j(s)
\]

where \( \rho_j(s, t) \) is the conditional correlation coefficient between two residuals at two different times. And the residual time series \( \{ \varepsilon_{ji}(t) \} \) from different subjects and different treatments are independent. Without loss of generality, we assume \( t, s \in [0, 1] \).

As commonly exercised in the partially linear model (Speckman 1988; Linton 1995), there is a secondary model for the covariate \( X_{ji} \):

\[
X_{ji}(t_{jim}) = h_j(t_{jim}) + u_{jim}, \ j = 1, 2, \ldots, k, \ i = 1, \ldots, n_j, \ m = 1, \ldots, T_j,
\]

where \( h_j(\cdot) \) are \( p \)-dimensional smooth functions with continuous second derivatives, the residual \( u_{jim} = (u_{jim1}^T, \ldots, u_{jimp}^T)^T \) satisfy \( \text{E}(u_{jim}) = 0 \) and \( u_{jl} \) and \( u_{jk} \) are independent for \( l \neq k \). For the purpose of identifying \( \beta_j \) and \( g_j(t) \), the covariance matrix of \( u_{jim} \) is assumed to be finite and positive definite (Härdle, Liang and Gao 2000).

We are interested in testing two hypotheses. One is on the treatment effects with respect to the covariates:

\[
H_{0a} : \beta_{10} = \beta_{20} = \ldots = \beta_{k0} \quad \text{vs} \quad H_{1a} : \beta_{10} \neq \beta_{j0} \quad \text{for some} \quad i \neq j;
\]

and the other is regarding the time effect functions:

\[
H_{0b} : g_{10}(\cdot) = \ldots = g_{k0}(\cdot) \quad \text{vs} \quad H_{1b} : g_{10}(\cdot) \neq g_{j0}(\cdot) \quad \text{for some} \quad i \neq j.
\]

For the ease of notation, we write \( \{Y_{jim}, X_{jim}\} \) to denote \( \{Y_{ji}(t_{jim}), X_{ji}(t_{jim})\} \), and let \( X_{ji} = \{X_{ji1}, \ldots, X_{jiT_j}\} \) and \( Y_{ji} = \{Y_{ji1}, \ldots, Y_{jiT_j}\} \) be the complete time series of covariates and responses of the \( (j, i) \)-th subject (the \( i \)-th subject in the \( j \)-th treatment), and \( \tilde{Y}_{jit,d} = \{Y_{ji(t-d)}, \ldots, Y_{ji(t-1)}\} \) and \( \tilde{X}_{jit,d} = \{X_{ji(t-d)}, \ldots, X_{ji(t-1)}\} \) be the past \( d \) observations at time \( t \) for a positive integer \( d \). For \( t < d \), we set \( d = t - 1 \).

Define the missing value indicator \( \delta_{jit} = 1 \) if \( (X_{jit}, Y_{jit}) \) is observed and \( \delta_{jit} = 0 \) if \( (X_{jit}, Y_{jit}) \) is missing. Here, we assume \( X_{jit} \) and \( Y_{jit} \) are either both observed or both missing, and \( \delta_{jit} = 1 \), namely the first observation for each subject is always observed. This simultaneous missingness of \( X_{jit} \) and \( Y_{jit} \) is for the ease of mathematical exposition. Our method can be extended to the case where the missingness of \( X_{jit} \) and \( Y_{jit} \) is not simultaneous.
We assume the missingness of \((X_{jit}, Y_{jit})\) at \(t\) is missing at random (MAR) (Rubin 1976) given its immediate past \(d\) complete observations. Let \(\delta_{jit,d} = \prod_{t=1}^{d} \delta_{ji(t-t)}\), the MAR means that for each \(j = 1, \ldots, k\),

\[
P(\delta_{jit} = 1|\delta_{jit,d} = 1, X_{jit}, Y_{jit}) = P(\delta_{jit} = 1|\delta_{jit,d} = 1, \tilde{X}_{jit,d}, \tilde{Y}_{jit,d}) = p_j(\tilde{X}_{jit,d}, \tilde{Y}_{jit,d}; \theta_{j0}).
\]

The form of the missing propensity \(p_j\) is known up to a parameter \(\theta_{j0}\), whose dimension depends on \(j\) and \(d\). The monotone missingness in the sense that \(\delta_{jit} = 0\) if \(\delta_{ji(t-1)} = 0\), considered in Robins et al (1995), is a special case with \(d \geq T_j\). Some guidelines on how to choose models for the missing propensity are given in Section 7 in the context of the empirical study. The robustness of the ANOVA tests with respect to the missing propensity model are discussed in Sections 3 and 4.

The parameters \(\theta_{j0}\) can be estimated by maximizing the binary likelihood

\[
(2.3) \quad \mathcal{L}_{ij}(\theta_j) = \prod_{i=1}^{n_j} \prod_{t=1}^{T_j} [p_j(\tilde{X}_{jit,d}, \tilde{Y}_{jit,d}; \theta_j)^{\delta_{jit}} \{1 - p_j(\tilde{X}_{jit,d}, \tilde{Y}_{jit,d}; \theta_j)\}^{(1-\delta_{jit})}]^{\delta_{jit,d}}.
\]

Under some regular conditions, the binary maximum likelihood estimator \(\hat{\theta}_j\) is \(\sqrt{n}\)-consistent estimator for \(\theta_{j0}\); see Chen et al (2008) for results on a related situation.

3. ANOVA Test for Covariate Effects. We consider testing for \(H_{0j} : \beta_{1j0} = \beta_{2j0} = \ldots = \beta_{kj0}\) with respect to the covariates. Let \(\pi_{jim}(\theta_j) = \prod_{m=d}^{n_j} p_j(\tilde{X}_{jil,d}, \tilde{Y}_{jil,d}; \theta_j)\) be the overall missing propensity for the \((j,i)-th\) subject up to time \(t_{jim}\). To remove the nonparametric part in (2.1), we first estimate the nonparametric function \(g_{j0}(t)\). If \(\beta_{j0}\) were known, \(g_{j0}(t)\) would be estimated by

\[
(3.1) \quad \hat{g}_j(t; \beta_{j0}) = \sum_{i=1}^{n_j} \sum_{m=1}^{T_j} w_{jim,h}(t)(Y_{jim} - X_{jim}^r(\beta_{j0})),
\]

where

\[
(3.2) \quad w_{jim,h}(t) = \frac{(\delta_{jim}/\pi_{jim}(\hat{\theta}_j))K_{h_j}(t_{jim} - t)}{\sum_{s=1}^{n_j} \sum_{t=1}^{T_j} (\delta_{jsl}/\pi_{jsl}(\hat{\theta}_j))K_{h_j}(t_{jsl} - t)}
\]

is a kernel weight that itself has been inversely weighted by the propensity \(\pi_{jim}(\hat{\theta}_j)\) to correct for selection bias due to the missing values. In (3.2), \(K\) is a univariate kernel function which is a symmetric probability density, \(K_{h_j}(t) = K(t/h_j)/h_j\) and \(h_j\) is a smoothing bandwidth. The conventional kernel estimation of \(g_{j0}(t)\) without weighting by \(\pi_{jsl}(\hat{\theta}_j)\) may not be consistent if the missingness depends on the responses \(Y_{jil}\), which may happen for missing covariates.

Center each \(\{X_{jim}, Y_{jim}\}\) by

\[
(3.3) \quad \tilde{X}_{jim} = X_{jim} - \sum_{i_1=1}^{n_j} \sum_{m_1=1}^{T_j} w_{jim_1,h_j(t_{jim})}X_{jim_1m_1}
\]

\[
(3.4) \quad \tilde{Y}_{jim} = Y_{jim} - \sum_{i_1=1}^{n_j} \sum_{m_1=1}^{T_j} w_{jim_1,h_j(t_{jim})}Y_{jim_1m_1}
\]
as is commonly exercised in the partially linear regression (Härdle, Liang and Gao 2000). Then, an estimating function for the \((j, i)\)-th subject is

\[
Z_{ji}(\beta_j) = \sum_{m=1}^{T_j} \frac{\delta_{jim}}{\tau_{jim}(\hat{\beta}_j)} X_{jim}(Y_{jim} - \hat{X}_{jim} \beta_j).
\]

At the true parameter \(\beta_j^0\), \(E\{Z_{ji}(\beta_j^0)\} = o(1)\). Although it is not exactly zero, \(Z_{ji}(\beta_j^0)\) can still be used as an approximate zero mean estimating function to formulate an empirical likelihood for \(\beta_j\) as follows.

Let \(\{p_{ji}\}_{i=1}^{n_j}\) be non-negative weights allocated to \(\{(X_{ji}^\tau, Y_{ji})\}_{i=1}^{n_j}\). Then the empirical likelihood for \(\beta_j\) based on measurements from the \(j\)-th treatment is

\[
L_{n_j}(\beta_j) = \max \left\{ \prod_{i=1}^{n_j} p_{ji} \right\},
\]

subject to \(\sum_{i=1}^{n_j} p_{ji} = 1\) and \(\sum_{i=1}^{n_j} p_{ji} Z_{ji}(\beta_j) = 0\).

By introducing a Lagrange multiplier \(\lambda_j\) to solve the above optimization problem and following the standard derivation in empirical likelihood (Owen, 2001), it can be shown that

\[
L_{n_j}(\beta_j) = \prod_{i=1}^{n_j} \left\{ \frac{1}{n_j} \frac{1}{1 + \lambda_j Z_{ji}(\beta_j)} \right\},
\]

where \(\lambda_j\) satisfies

\[
\sum_{i=1}^{n_j} \frac{Z_{ji}(\beta_j)}{1 + \lambda_j^2 Z_{ji}(\beta_j)} = 0.
\]

The maximum of \(L_{n_j}(\beta_j)\) is \(\prod_{i=1}^{n_j} \frac{1}{n_j}\), achieved at \(\beta_j = \hat{\beta}_j\) and \(\lambda_j = 0\), where \(\hat{\beta}_j\) is the solution of \(\sum_{i=1}^{n_j} Z_{ji}(\hat{\beta}_j) = 0\), which can be solved by the Newton Raphson method.

Let \(n = \sum_{i=1}^{k} n_j, n_j/n \to \rho_j\) for some non-zero \(\rho_j\) as \(n \to \infty\) such that \(\sum_{j=1}^{k} \rho_j = 1\). As the \(k\) samples are independent, the joint empirical likelihood for \((\beta_1, \beta_2, \ldots, \beta_k)\) is

\[
L_n(\beta_1, \beta_2, \ldots, \beta_k) = \prod_{j=1}^{k} L_{n_j}(\beta_j).
\]

The log likelihood ratio test statistic for \(H_{\text{0a}}\) is

\[
\ell_n : = -2 \max_{j} \log L_n(\beta, \beta, \ldots, \beta) + \sum_{j=1}^{k} n_j \log n_j
\]

\[
= 2 \min_{\beta} \sum_{j=1}^{k} \sum_{i=1}^{n_j} \log \{1 + \lambda_j^2 Z_{ji}(\beta)\}.
\]

Using a Taylor expansion and the Lagrange method to carry out the minimization in (3.8) (See Appendix), the optimal solution to \(\beta\) is

\[
(3.9) \quad \left( \sum_{j=1}^{k} \Omega_x B_j^{-1} \Omega_x \right)^{-1} \left( \sum_{j=1}^{k} \Omega_x B_j^{-1} \Omega_x y_j \right) + o_p(1),
\]
where \( B_j = \lim_{n_j \to \infty} (n_jT_j)^{-1} \sum_{i=1}^{n_j} T_{ij}^{-1} E\{Z_{ji}(\beta_0)Z_{ji}(\beta_0)^\top \} \),

\[
\Omega_{x_j} = \frac{1}{\sqrt{n_jT_j}} \sum_{i=1}^{n_j} \sum_{m=1}^{T_j} E\left\{ \frac{\delta_{jim}(\hat{\theta}_j)}{\pi_{jim}(\hat{\theta}_j)} \tilde{X}_{jim} \tilde{X}_{jim}^\top \right\}
\]

and

\[
\Omega_{x_jy_j} = \frac{1}{\sqrt{n_jT_j}} \sum_{i=1}^{n_j} \sum_{m=1}^{T_j} \frac{\delta_{jim}}{\pi_{jim}(\hat{\theta}_j)} \tilde{X}_{jim} \tilde{Y}_{jim}.
\]

The ANOVA test statistic (3.8) can be viewed as a nonparametric counterpart of the conventional parametric likelihood ratio ANOVA test statistic, for instance that considered in Forcina (1992). Indeed, like its parametric counterpart, the Wilks’ theorem is maintained for \( \ell_n \).

**Theorem 1.** If the conditions (A1-A4) given in the Appendix hold, then under \( H_{0a} \),

\[ \ell_n \xrightarrow{d} \chi^2_{(k-1)p} \quad \text{as} \quad n \to \infty. \]

The theorem suggests an empirical likelihood ANOVA test that rejects \( H_{0a} \) if \( \ell_n > \chi^2_{(k-1)p,\alpha} \) where \( \alpha \) is the level of significance and \( \chi^2_{(k-1)p,\alpha} \) is the upper \( \alpha \) quantile of \( \chi^2_{(k-1)p} \) distribution.

We next evaluate the power of the empirical likelihood ANOVA test under a series of local alternative hypotheses:

\[ H_{1a}: \beta_{j0} = \beta_{10} + c_n n_j^{-1/2} \quad \text{for} \quad 2 \leq j \leq k \]

where \( \{c_n\} \) is a sequence of bounded constants and \( n_j \) such that \( n = \rho_j^{-1} n_j \) as we defined as above. Define \( \Delta_j = (\beta_{10}^- - \beta_{10}^+, \beta_{10}^- - \beta_{20}^+, \ldots, \beta_{10}^- - \beta_{k0}^+) \), \( D_{ij} = \Omega_{xx}^{-1}\Omega_{xy} - \Omega_{xx}^{-1}\Omega_{xy} \) for \( 2 \leq j \leq k \) and \( D = (D_{12}, D_{13}, \ldots, D_{kk})^\top \). Let \( \Sigma_D = \text{Var}(D) \) and \( \gamma^2 = \Delta_j^\top \Sigma_D^{-1} \Delta_j \). Theorem 2 gives the asymptotic distribution of \( \ell_n \) under the local alternative hypothesis.

**Theorem 2.** Suppose the conditions (A1-A4) in the Appendix hold, and under \( H_{1a} \),

\[ \ell_n \xrightarrow{d} \chi^2_{(k-1)p}(\gamma^2) \quad \text{as} \quad n \to \infty. \]

It can be shown that

\[ \Sigma_D = \Omega_{xx}^{-1}B_1\Omega_{xx}^{-1}1_{(k-1)} \otimes 1_{(k-1)} + \text{diag}\{\Omega_{xx}^{-1}B_2\Omega_{xx}^{-1}, \ldots, \Omega_{xx}^{-1}B_k\Omega_{xx}^{-1}\}. \]

As each \( \Omega_{xx}^{-1} \) is \( O(n^{1/2}) \), the non-central component \( \gamma^2 \) is non-zero and bounded. The power of the \( \alpha \)-level empirical likelihood ANOVA test is

\[ \beta(\gamma) = \Pr(\chi^2_{(k-1)p}(\gamma^2) > \chi^2_{(k-1)p,\alpha}) \]

This indicates that test is able to detect local departures of size \( O(n^{-1/2}) \) from \( H_{0a} \) which is the best rate we can achieve under the local alternative set-up for finite dimensional parameters. This is achieved despite the fact that nonparametric kernel estimation is involved in the formulation, which has a slower rate of convergence than \( \sqrt{n} \), as the centering in (3.3) and (3.4) essentially eliminate the effects of the nonparametric estimation.

**Remark 1.** When there is no missing values, namely all \( \delta_{jim} = 1 \), we will assign all \( \pi_{jim}(\hat{\theta}_j) = 1 \) and there is no need to estimate each \( \theta_j \). In this case, Theorems 1 and 2 remain valid as we are concerning with testing. It is a different matter for estimation as estimation efficiency with missing values will be less than that without missing values.
Remark 2. The above ANOVA test is robust against mis-specifying the missing propensity $p_j(\cdot; \theta_0)$ provided the missingness does not depend on the responses $Y_{j(it,d)}$. This is because despite the mis-specification, the mean of $Z_{ji}(\beta)$ is still approximately zero and the empirical likelihood formulation remains valid, as well as Theorems 1 and 2. However, if the missingness depends on the responses and if the model is mis-specified, Theorems 1 and 2 will be affected.

Remark 3. The empirical likelihood test can be readily modified for ANOVA testing on pure parametric regressions with some parametric time effects $g_{j0}(t; \eta_j)$ with parameters $\eta_j$. We may formulate the empirical likelihood for $\{g_{j0}(t; \eta_j)\}$ as the estimating function for the $(j,i)$-th subject. The ANOVA test can be formulated following the same procedures from (3.6) to (3.8), and both Theorems 1 and 2 remaining valid after updating $p$ with $p + q$ where $q$ is the dimension of parameter $\eta_j$.

In our formulation for the ANOVA test here and in the next section, we rely on the Nadaraya-Watson type kernel estimator. The local linear kernel estimator may be employed as the boundary bias may seem to be an issue. However, as we are interested in ANOVA tests instead of estimation, the boundary bias does not have any leading order effects. Nevertheless, the local linear kernel smoothing can be used without affecting the main results of the paper.

4. ANOVA Test for Time Effects. In this section, we consider the ANOVA test for the nonparametric part

$$H_{0b} : g_{10}(\cdot) = \ldots = g_{kb}(\cdot).$$

We will formulate an empirical likelihood for $g_{j0}(t)$ at each $t$, which then lead to an overall likelihood ratio for $H_{0b}$. We need an estimator of $g_{j0}(t)$ that is less biased than the one in (3.1). Plugging-in the estimator $\hat{\beta}_j$ to (3.1), we have

$$g_j(t) = \sum_{i=1}^{n_j} \sum_{m=1}^{T_j} w_{jim,h_j}(t)(Y_{jim} - X_{jim}^\tau \hat{\beta}_j).$$

It follows that, for any $t \in [0,1]$,

$$\tilde{g}_j(t) - g_{j0}(t) = \sum_{i=1}^{n_j} \sum_{m=1}^{T_j} w_{jim,h_j}(t) \left\{ \varepsilon_{ji}(t_{jim}) + X_{jim}^\tau (\beta_j - \hat{\beta}_j) + g_{j0}(t_{jim}) - g_{j0}(t) \right\}.$$

However, there is a bias of order $h_j^2$ in the kernel estimation since

$$\sum_{i=1}^{n_j} \sum_{m=1}^{T_j} w_{jim,h_j}(t) \{g_{j0}(t_{jim}) - g_{j0}(t)\} = \frac{1}{2} \int z^2 K(z) dz g_{j0}''(t) h_j^2 + o_p(h_j^2).$$

If we formulated the empirical likelihood based on $\tilde{g}_j(t)$, the bias will contribute to the asymptotic distribution of the ANOVA test statistic. To avoid that, we use the bias-correction method proposed in Xue and Zhu (2007a) so that the estimator of $g_{j0}$ is

$$\hat{g}_j(t) = \sum_{i=1}^{n_j} \sum_{m=1}^{T_j} w_{jim,h_j}(t) \{Y_{jim} - X_{jim}^\tau \hat{\beta}_j - (\tilde{g}_j(t_{jim}) - \tilde{g}_j(t))\}.$$
Based on this modified estimator \( \hat{g}_j(t) \), we define the auxiliary variable

\[
R_{ji}(g_j(t)) = \sum_{m=1}^{T_j} \frac{\delta_{jm}}{\pi_{jm}(\theta_j)} K \left( \frac{t_{jm} - t}{h_j} \right) \left\{ Y_{jm} - X_{jm} \beta_j - g_j(t) - (\hat{g}_j(t_{jm}) - \hat{g}(t)) \right\}
\]

for empirical likelihood formulation. At true function \( g_{j0}(t) \), \( E[R_{ji}(g_{j0}(t))] = o(1) \).

Using a similar procedure to \( L_{nj}(\beta_j) \) as given in (3.6) and (3.7), the empirical likelihood for \( g_{j0}(t) \) is

\[
L_{nj}(g_{j0}(t)) = \max \left\{ \prod_{i=1}^{n_j} p_{ji} \right\}
\]

subject to \( \sum_{i=1}^{n_j} p_{ji} = 1 \) and \( \sum_{i=1}^{n_j} p_{ji}R_{ji}(g_j(t)) = 0 \). The latter is obtained in a similar fashion as we obtain (3.6) by introducing Langrange multipliers. It can be shown that

\[
L_{nj}(g_{j0}(t)) = \prod_{i=1}^{n_j} \left\{ \frac{1}{n_j} \frac{1}{1 + \eta_j(t)R_{ji}(g_{j0}(t))} \right\},
\]

where \( \eta_j(t) \) is a Lagrange multiplier that satisfies

\[
\sum_{i=1}^{n_j} \frac{R_{ji}(g_{j0}(t))}{1 + \eta_j(t)R_{ji}(g_{j0}(t))} = 0.
\]

The log empirical likelihood ratio for \( g_{10}(t) = \ldots = g_{k0}(t) := g(t) \), say, is

\[
\mathcal{L}_n(t) = 2 \min_{g(t)} \sum_{j=1}^{k} \sum_{i=1}^{n_j} \log(1 + \eta_j(t)R_{ji}(g(t)));
\]

which is analogues of \( \ell_n \) in (3.8).

Let \( v_j(t, h_j) = \sum_{i=1}^{n_j} R_{ji}^2(g(t)) \) and \( d_j(t, h_j) = \sum_{i=1}^{n_j} \sum_{m=1}^{T_j} \frac{\delta_{jm}}{\pi_{jm}(\theta_j)} K \left( \frac{t_{jm} - t}{h_j} \right) \), where for notation simplification, we have suppressed \( n_j \) in the arguments of these two functions. Then by a Taylor expansion,

\[
\mathcal{L}_n(t) = \sum_{j=1}^{k} v_j^{-1}(t, h_j) \left[ \sum_{i=1}^{n_j} R_{ji}\{0\} - d_j(t, h_j) \right] \left( \sum_{s=1}^{k} v_s^{-1}(t, h_s) d_s^2(t, h_s) \right)^{-1} \times \sum_{s=1}^{k} v_s^{-1}(t, h_s) d_s(t, h_s) \sum_{i=1}^{n_s} R_{si}\{0\}^2 + \tilde{O}_p \left( (n_jh_j)^{-1} \log^2 n_j \right).
\]

It may be shown that the leading term of \( \mathcal{L}_n(t) \) is

\[
\left\{ \sum_{j=1}^{k} \frac{2}{\zeta_j} \right\}^{-1} \left\{ \hat{g}_1(t) - \hat{g}_2(t) \right\}^2
\]

with \( \zeta_j = v_j(t, h_j)/d_j^2(t, h_j) \) for \( k = 2 \); If \( k = 3 \), the leading term of \( \mathcal{L}_n(t) \) will be

\[
\left( \hat{g}_1(t) - \hat{g}_2(t) \right) \left( \hat{g}_1(t) - \hat{g}_3(t) \right) H_n \left( \hat{g}_1(t) - \hat{g}_3(t) \right)
\]
where
\[
H_n = \left\{ \sum_{j=1}^{3} \zeta_j^{-1} \right\}^{-1} \times \left( \begin{array}{ccc}
\zeta_2^{-1}(\zeta_1^{-1} + \zeta_3^{-1}) & \zeta_2^{-1}\zeta_3^{-1} \\
\zeta_2^{-1}\zeta_3^{-1} & \zeta_3^{-1}(\zeta_1^{-1} + \zeta_2^{-1}) 
\end{array} \right).
\]

An expression for a general \( k \) is available which shows that the leading order term of the ANOVA statistic \( \mathcal{L}_n(t) \) is a studentized \( L_2 \) distance between \( \hat{g}_1(t) \) and the other \( \hat{g}_j(t) \) (\( j \neq 1 \)). This means that \( \mathcal{L}_n(t) \) is able to test for equivalence of \( \{g_{j0}(t)\}_{j=1}^{k} \) at any \( t \in [0,1] \). In summary, the leading order term of the \( \mathcal{L}_n(t) \) is a studentized version of the distance
\[
(\hat{g}_1(t) - \hat{g}_2(t), \hat{g}_1(t) - \hat{g}_3(t), \ldots, \hat{g}_1(t) - \hat{g}_k(t)),
\]

namely between \( \hat{g}_1(t) \) and the other \( \hat{g}_j(t) \) (\( j \neq 1 \)). This motivates us to propose using
\[
(4.8) \quad T_n = \int_0^1 \mathcal{L}_n(t) \varpi(t) \, dt
\]

to test for the equivalence of \( \{g_{j0}(\cdot)\}_{j=1}^{k} \), where \( \varpi(t) \) is a probability weight function over \([0,1]\).

We consider a sequence of local alternative hypotheses:
\[
(4.9) \quad g_{j0}(t) = g_{10}(t) + C_{jn} \Delta_{jn}(t),
\]

where \( C_{jn} = (n_jT_j)^{-1/2}h_j^{1/4} \) for \( j = 2, \ldots, k \) and \( \{\Delta_{jn}(t)\}_{n \geq 1} \) is a sequence of uniformly bounded functions.

To define the asymptotic distribution of \( T_n \), we need some notations. We assume without loss of generality that for each \( h_j \) and \( T_j \), \( j = 1, \ldots, k \), there exist fixed finite positive constants \( \alpha_j \) and \( b_j \) such that \( \alpha_jT_j = T \) and \( b_jh_j = h \) for some \( T \) and \( h \) as \( h \to 0 \). Effectively, \( T \) is the smallest common multiple of \( T_1, \ldots, T_k \). Let \( K^{(2)}_c(t) = \int K(w)K(t - cw) \, dt \) and \( K^{(4)}_c(0) = \int K^{(2)}_c(w\sqrt{c})K^{(2)}_c(w\sqrt{c}) \, dw \). For \( c = 1 \), we resort to the standard notations of \( K^{(2)}(t) \) and \( K^{(4)}(0) \) for \( K^{(2)}_1(t) \) and \( K^{(4)}_1(0) \), respectively. For each treatment \( j \), let \( f_j \) be the super-population density of the design points \( \{t_{jm}\} \). Let \( \rho_j = \rho_j^{-1/2} \alpha_j \),
\[
W_j(t) = \frac{f_j(t) / \{a_jb_j\sigma_{\hat{g}_j}^2\}}{\sum_{l=1}^{k} f_l(t) / \{a_lb_l\sigma_{\hat{g}_l}^2\}}
\]

and \( V_j(t) = K^{(2)}(0)\sigma_{\hat{g}_j}^2f_j(t) \) where \( \sigma_{\hat{g}_j}^2 = \frac{1}{n_jT_j} \sum_{i=1}^{T_j} E\left\{ \frac{\hat{g}_{jm}^2}{\sigma_{\hat{g}_j}^2} \right\} \). Furthermore, we define
\[
\Lambda(t) = \sum_{j=1}^{k} b_j^{-1}K^{(4)}(0)\left(1 - W_j(t)\right)^2 + \sum_{j \neq j_1} K^{(4)}_{b_j/b_{j_1}}(0)W_j(t)W_{j_1}(t)
\]
and
\[
\mu_1 = \int_0^1 \left[ \sum_{j=1}^{k} b_j^{-1}V_j^{-1}(t)f_j^2(t)\Delta_{n_j}(t) - \left( \sum_{s=1}^{k} b_s^{-1}V_s^{-1}(t)f_s^2(t)\Delta_{n_s}(t) \right)^2 \right] \varpi(t) \, dt.
\]

**Theorem 3.** Assume conditions (A1-A4) in the Appendix hold, and \( h = O(n^{-1/5}) \), then under (4.9),
\[
h^{-1/2}(T_n - \mu_0) \overset{d}{\to} N(0, \sigma_0^2),
\]

where \( \mu_0 = (k - 1) + h^{1/2}\mu_1 \) and \( \sigma_0^2 = 2K^{(2)}(0)^{-2} \int_0^1 \Lambda(t)\varpi^2(t) \, dt \).
We note that under $H_{0b}$: $g_{10}(\cdot) = \ldots = g_{k0}(\cdot)$, $\Delta_{j0}(t) = 0$ which yields $\mu_1 = 0$ and

$$h^{-1/2}\{T_n - (k - 1)\} \xrightarrow{d} N(0, \sigma^2_0).$$

This may lead to an asymptotic test at a nominal significance level $\alpha$ that rejects $H_{0b}$ if

$$T_n \geq h^{1/2}\sigma_0 z_\alpha + (k - 1),$$

where $z_\alpha$ is the upper $\alpha$ quantile of $N(0, 1)$ and $\sigma_0$ is a consistent estimator of $\sigma_0$. The asymptotic power of the test under the local alternatives is $1 - \Phi(z_\alpha - \frac{\mu_1}{\sigma_0})$, where $\Phi(\cdot)$ is the standard normal distribution function. This indicates that the test is powerful in differentiating null hypothesis and its local alternative at the convergence rate $O(n_j^{-1/2}h_j^{-1/4})$ for $C_{jn}$. The rate is the best we could attain when a single bandwidth is used; see Härdle and Mammen (1993).

If all the $h_j$ ($j = 1, \ldots, k$) are the same, the asymptotic variance $\sigma_0^2 = 2(k - 1)K(0)^{-2} \times K(4)(0) \int_0^1 \varphi^2(t)dt$, which means that the test statistic under $H_{0b}$ is asymptotic pivotal under null hypothesis. However, when the bandwidths are not the same, which is the most likely case as different treatments may require different amount of smoothness in the estimation of $g_{j0}(\cdot)$, the asymptotical pivotalness of $T_n$ is no longer available, and estimation of $\sigma_0^2$ is needed for conducting the asymptotic test in (4.10). We will propose a test based on bootstrap calibration to the distribution of $T_n$ in the next section.

**Remark 4.** Similar to Remarks 1 and 2 made on the ANOVA tests for the covariate effects, the proposed ANOVA test for the nonparametric baseline functions (Theorem 3) remains valid in the absence of missing values and/or if the missing propensity is misspecified as long as the responses do not contribute to the missingness.

**Remark 5.** We note that the proposed test is not affected by the within-subject dependent structure (the longitudinal aspect) due to the fact that the formulation of the empirical likelihood is made for each subject as shown in the construction of $R_{ji}\{g_j(t)\}$ as the nonparametric functions can be well separated from the covariate effects in the semiparametric model. Again this would be changed if we are interested in estimation as the correlation structure in the longitudinal data will affect the efficiency of the estimation. However, the test will be dependent on the choice of the weight function $\varphi(\cdot)$, and $\{\alpha_j\}$, $\{\rho_j\}$ and $\{b_j\}$, the relative ratios among $\{T_j\}$, $\{n_j\}$ and $\{h_j\}$, as would be normally the case for other nonparametric goodness of fit tests.

**Remark 6.** The ANOVA test statistics for the time effects for the semiparametric model can be modified for ANOVA test for purely nonparametric regression by simply setting $\hat{\beta}_j = 0$ in the formulation of the test statistic $L_n(t)$. In this case, the model (2.1) takes the form

$$Y_{ji}(t) = g_j(X_{ji}(t), t) + \varepsilon_{ji}(t),$$

where $g_j(\cdot)$ is the unknown nonparametric function of $X_{ji}(t)$ and $t$. The proposed ANOVA test can be viewed as generalization of the tests considered in Mund and Detter (1998), Pardo-Fernández, Van Keilegom and González-Manteiga (2007) and Wang, Akritas and Van Keilegom (2008) by considering both the longitudinal and missing aspects.

**5. Bootstrap Calibration.** To avoid direct estimation of $\sigma_0^2$ in Theorem 3 and to speed up the convergence of $T_n$, we resort to the bootstrap. While the wild bootstrap (Wu 1986, Liu 1988 and Härdle and Mammen 1993) originally proposed for parametric regression
without missing values has been modified by Shao and Sitter (1996) to take into account missing values, we extend it further to suit the longitudinal feature.

Let \( \tilde{\eta}_j \) and \( \tilde{\eta}_n \) be the sets of the time points with full and missing observations, respectively. According to model (2.2), we impute a missing \( X_{ji}(t) \) from \( \hat{X}_{ji}(t) \), \( t \in \tilde{\eta}_j \), so that for any \( t \in \tilde{\eta}_n \)

\[
\hat{X}_{ji}(t) = \sum_{i=1}^{n_j} \sum_{m=1}^{T_j} w_{jim,h_j}(t) X_{jim},
\]

where \( w_{jim,h_j}(t) \) is the kernel weight defined in (3.2).

To mimic the heteroscedastic and correlation structure in the longitudinal data, we estimate the covariance matrix for each subject in each treatment. Let

\[
\hat{\varepsilon}_{jim} = Y_{jim} - X_{jim}^* \hat{\beta}_j - \hat{g}_j(t_{jim}).
\]

An estimator of \( \sigma_j^2(t) \), the variance of \( \varepsilon_{ji}(t) \), is \( \hat{\sigma}_j^2(t) = \sum_{i=1}^{n_j} \sum_{m=1}^{T_j} w_{jim,h_j}(t) \hat{\varepsilon}_{jim}^2 \) and an estimator of \( \rho_j(s,t) \), the correlation coefficient between \( \varepsilon_{ji}(t) \) and \( \varepsilon_{ji}(s) \) for \( s \neq t \), is

\[
\hat{\rho}_j(s,t) = \frac{n_j \sum_{m=1}^{T_j} H_{jim,m'}(s,t) \hat{\varepsilon}_{jim}(s) \hat{\varepsilon}_{jim'}(t)}{\sum_{m=1}^{T_j} H_{jim,m'}(s,t) \hat{\varepsilon}_{jim}(s) \hat{\varepsilon}_{jim'}(t)}.
\]

where \( \hat{\varepsilon}_{jim} = \varepsilon_{jim}/\hat{\sigma}_j(t_{jim}) \),

\[
H_{jim,m'}(s,t) = \frac{\delta_{jim} \delta_{jim'} K_{h_j} (s - t_{jim}) K_{h_j} (t - t_{jim'}) / \pi_{jim,m'}(\hat{\theta}_j)}{\sum_{i=1}^{n_j} \sum_{m=1}^{T_j} \delta_{jim} \delta_{jim'} K_{h_j} (s - t_{jim}) K_{h_j} (t - t_{jim'}) / \pi_{jim,m'}(\hat{\theta}_j)},
\]

and \( \pi_{jim,m'}(\hat{\theta}_j) = \pi_{jim}(\hat{\theta}_j) \pi_{jim'}(\hat{\theta}_j) \) if \( |m - m'| > d \); \( \pi_{jim,m'}(\hat{\theta}_j) = \pi_{jim}(\hat{\theta}_j) \) if \( |m - m'| \leq d \) where \( m_0 = \max(m, m') \). Here \( h_j \) is a smoothing bandwidth which may be different from the bandwidth \( h_{j,i} \) for calculating the test statistics \( T_n \) (Fan, Huang and Li 2007). Then, the covariance \( \Sigma_{ji} \) of \( \varepsilon_{ji} = (\varepsilon_{j1}, \ldots, \varepsilon_{jiT})' \) is estimated by \( \hat{\Sigma}_{ji} \) which has \( \hat{\sigma}_j^2(t_{jim}) \) as its \( m \)-th diagonal element and \( \hat{\rho}_j(t_{jik},t_{jil}) \) as its \((k,l)-th\) element for \( k \neq l \).

Let \( Y_{ji}, \delta_{ji}, t_{ji} \) be the vector of random variables of subject \( i \) in the \( j \)-th treatment, \( X_{ji} = (X_{ji}(t_{j1}), \ldots, X_{ji}(t_{jT}))' \) and \( g_{j0}(t_{sl}) = (g_{j0}(t_{s1}), \ldots, g_{j0}(t_{sT}))' \), where \( s \) may be different from \( j \). Let \( X_{ji}^s = \{X_{ji}^s, X_{ji}^m \} \), where \( X_{ji}^m \) contains observed \( X_{ji}(t) \) for \( t_j \in \tilde{\eta}_j \) and \( X_{ji}^m \) collects the imputed \( X_{ji}(t) \) for \( t \in \tilde{\eta}_n \) according to (5.1).

The proposed bootstrap procedure consists of the following steps:

**Step 1.** Generate a bootstrap re-sample \( \{Y_{ji}^*, X_{ji}^s, \delta_{ji}^*, t_{ji} \} \) for the \((j,i)-th\) subject by

\[
Y_{ji}^* = X_{ji}^s \hat{\beta}_j + \hat{g}_j(t_{ji}) + \hat{\Sigma}_{ji} e_{ji}^*,
\]

where \( e_{ji}^* \)'s are i.i.d. random vectors simulated from a distribution satisfying \( E(e_{ji}^*) = 0 \) and \( \text{Var}(e_{ji}^*) = I_{T_j} \), \( \delta_{ji}^* \sim \text{Bernoulli}(\pi_{jim}(\hat{\theta}_j)) \) where \( \hat{\theta}_j \) is estimated based on the original sample as given in (2.3). Here, \( \hat{g}_j(t_{ji}) \) is used as the common nonparametric time effect to mimic the null hypothesis \( H_{0b} \).

**Step 2.** For each treatment \( j \), we re-estimate \( \beta_j, \theta_j \) and \( g_{j}(t) \) based on the re-sample \( \{Y_{ji}^*, X_{ji}^s, \delta_{ji}^*, t_{ji} \} \) and denote them as \( \hat{\beta}_j^*, \hat{\theta}_j^* \) and \( \hat{g}_j^*(t) \). The bootstrap version of \( R_{ji}(g_{j}(t)) \) is

\[
R_{ji}^*(\hat{g}_j(t)) = \frac{T_j}{\sum_{m=1}^{T_j} \pi_{jim}(\hat{\theta}_j)} K \left( \frac{t_{jim} - t}{h_j} \right) \left[ Y_{ji}^* - X_{ji}^s \hat{\beta}_j^* - \hat{g}_j(t) - \{\hat{g}^*_j(t_{jim}) - \hat{g}_j(t)\} \right]
\]

for any \( t \in \tilde{\eta}_n \).
and use it to substitute $R_{ji}\{g_j(t)\}$ in the formulation of $L_n(t)$, we obtain $L_n^*(t)$ and then $T_n^* = \int L_n^*(t)\varpi(t)dt$.

**Step 3.** Repeat the above two steps $B$ times for a large integer $B$ and obtain $B$ bootstrap values $(T_{nb})_{b=1}^B$. Let $t_\alpha$ be the $1 - \alpha$ quantile of $(T_{nb})_{b=1}^B$, which is a bootstrap estimate of the $1 - \alpha$ quantile of $T_n$. Then, we reject the null hypothesis $H_0$ if $T_n > t_\alpha$.

The following theorem justifies the bootstrap procedure.

**Theorem 4.** Assume conditions (A1-A4) in the Appendix hold and $h = O(n^{-1/3})$. Let $X_n$ denote the original sample and $h, \sigma_0^2$ be defined as in Theorem 3. The conditional distribution of $h^{-1/2}(T_n - \mu_0)$ given $X_n$ converges to $N(0, \sigma_0^2)$ almost surely, namely,

$$h^{-1/2}(T_n - (k - 1))|X_n \overset{d}{\longrightarrow} N(0, \sigma_0^2) \quad a.s.$$

**6. Simulation Results.** In this section, we report results from simulation studies which were designed to confirm the proposed ANOVA tests proposed in the previous sections. We simulated data from the following three-treatment model:

$$Y_{jim} = X_{jim}\beta_j + g_j(t_{jim}) + \epsilon_{jim} \quad \text{and} \quad X_{jim} = 2 - 1.5t_{jim} + u_{jim},$$

where $\epsilon_{jim} = e_{ji} + \nu_{jim}$, $u_{jim} \sim N(0, \sigma_{\epsilon_jim}^2)$, $e_{ji} \sim N(0, \sigma_{\epsilon_jim}^2)$ and $\nu_{jim} \sim N(0, \sigma_{\nu_jim}^2)$ for $j = \{1, 2, 3\}$, $i = 1, \cdots, n_j$ and $m = 1, \cdots, T_j$. This structure used to generate $\{\epsilon_{jim}\}_{m=1}^{T_j}$ ensures dependence among the repeated measurements $\{Y_{jim}\}$ for each subject $i$. The correlation between $Y_{jim}$ and $Y_{jil}$ for any $m \neq l$ is $\sigma_{\epsilon_jim}/(\sigma_{\epsilon_jim}^2 + \sigma_{\epsilon_jil}^2)$. The time points $\{t_{jim}\}_{m=1}^{T_j}$ were obtained by first independently generating uniform $[0, 1]$ random variables and then sorted in the ascending order. We set the number of repeated measures $T_j$ to be the same, say $T_j$, for all three treatments; and chose $T = 5$ and 10 respectively. The standard deviation parameters in (6.1) were $\sigma_{\epsilon_1} = 0.5, \sigma_{\epsilon_2} = 0.5, \sigma_{\epsilon_3} = 0.2$ for the first treatment, $\sigma_{\epsilon_2} = 0.5, \sigma_{\epsilon_3} = 0.2$ for the second and $\sigma_{\epsilon_3} = 0.6, \sigma_{h_3} = 0.6, \sigma_{c_3} = 0.3$ for the third.

The parameters and the time effects for the three treatments were

**Treatment 1:** $\beta_1 = 2, \quad g_1(t) = 2\sin(2\pi t)$;

**Treatment 2:** $\beta_2 = 2 + D_{2n}, \quad g_2(t) = 2\sin(2\pi t) - \Delta_{2n}(t)$;

**Treatment 3:** $\beta_3 = 2 + D_{3n}, \quad g_3(t) = 2\sin(2\pi t) - \Delta_{3n}(t)$.

We designated different values of $D_{2n}, D_{3n}, \Delta_{2n}(t)$, and $\Delta_{3n}(t)$ in the evaluation of the size and the power, whose details will be reported shortly.

We considered two missing data mechanisms. In the first mechanism (I), the missing propensity was

$$logit\{P(\delta_{jim} = 1|\delta_{jim,m-1} = 1, X_{ji}, Y_{ji})\} = \theta_j X_{ji(m-1)} \quad \text{for} \quad m > 1,$$

which is not dependent on the response $Y$, with $\theta_1 = 3, \theta_2 = 2$ and $\theta_3 = 2$. In the second mechanism (II),

$$logit\{P(\delta_{jim} = 1|\delta_{jim,m-1} = 1, X_{ji}, Y_{ji})\} = \begin{cases} \theta_{j1} X_{ji(m-1)} + \theta_{j2} \{Y_{ji(m-1)} - Y_{ji(m-2)}\}, & \text{if} \ m > 2, \\ \theta_{j1} X_{ji(m-1)}, & \text{if} \ m = 2; \end{cases}$$

which is influenced by both covariate and response, with $\theta_1 = (\theta_{11}, \theta_{12})^T = (2, -1)^T, \theta_2 = (\theta_{21}, \theta_{22})^T = (2, -1.5)^T$ and $\theta_3 = (\theta_{31}, \theta_{32})^T = (2, -1.5)^T$. In both mechanisms, the first observation ($m = 1$) for each subject was always observed as we have assumed earlier.
We used the Epanechnikov kernel $K(u) = 0.75(1-u^2)_+$ throughout the simulation where $(\cdot)_+$ stands for the positive part of a function. The bandwidths were chosen by the ‘leave-one-subject’ out cross-validation. Specifically, we chose the bandwidth $h_j$ that minimized the cross-validation score functions

$$
\sum_{i=1}^{n_j} \sum_{m=1}^{T_j} \frac{\delta_{jim}}{\pi_{jim}(\theta_j)} (Y_{jim} - X_{jim}^T \hat{\beta}_j^{(-i)} - \hat{g}_j^{(-i)}(t_{jim}))^2,
$$

where $\hat{\beta}_j^{(-i)}$ and $\hat{g}_j^{(-i)}(t_{jim})$ were the corresponding estimates without using observations of the $i$-th subject. We fixed the number of simulations to be 500.

The average missing percentages based on 500 simulations for the missing mechanism I were 8%, 15% and 17% for Treatments 1-3 respectively when $T = 5$, and were 16%, 28% and 31% when $T = 10$. In the missing mechanism II, the average missing percentages were 10%, 8% and 15% for $T = 5$, and 23%, 20% and 36% for $T = 10$, respectively.

For the ANOVA test for $H_{0a} : \beta_{j0} = \beta_{20} = \beta_{30}$ with respect to the covariate effects, three values of $D_{2n}$ and $D_{3n}$: 0, 0.1 and 0.2, were used respectively, while fixing $\Delta_{2n}(t) = 0$ and $\Delta_{3n}(t) = 0$. Table 1 summarizes the empirical size and power of the proposed EL ANOVA test with 5% nominal significant level for $H_{0a}$ for 9 combinations of $(D_{2n}, D_{3n})$, where the sizes corresponding to $D_{2n} = 0$ and $D_{3n} = 0$. We observed that the size of the ANOVA tests improved as the sample sizes and the observational length $T$ increased, and the overall level of size were close to the nominal 5%. This is quite re-assuring considering the ANOVA test is based on the asymptotic chi-square distribution. We also observed that the power of the test increased as sample sizes and $T$ were increased, and as the distance among the three $\beta_{j0}$ was increased. For example, when $D_{2n} = 0.0$ and $D_{3n} = 0.2$, the $L_2$ distance was $\sqrt{0.2^2 + 0.2^2} = 0.283$, which is larger than $\sqrt{0.1^2 + 0.1^2 + 0.2^2} = 0.245$ for $D_{2n} = 0.1$ and $D_{3n} = 0.2$. This explains why the ANOVA test was more powerful for $D_{2n} = 0.0$ and $D_{3n} = 0.2$ than $D_{2n} = 0.1$ and $D_{3n} = 0.2$. At the same time, we see similar power performance between the two missing mechanisms.

We then evaluate the power and size of the proposed ANOVA test regarding the nonparametric components. To study the local power of the test, we set $\Delta_{3n}(t) = U_n \sin(2\pi t)$ and $\Delta_{3n}(t) = 2 \sin(2\pi t) - 2 \sin(2\pi (t + V_n))$, and fixed $D_{2n} = 0$ and $D_{3n} = 0.2$. Here $U_n$ and $V_n$ were designed to adjust the amplitude and phase of the sine function. The same kernel and bandwidths chosen by the cross-validation as outlined earlier in the parametric ANOVA test were used in the test for the nonparametric time effects. We calculated the test statistic $T_n$ with $w(t)$ being the kernel density estimate based on all the time points in all treatments. We applied the wild bootstrap proposed in Section 5 with $B = 100$ to obtain $T_{0.05}$, the bootstrap estimator of the 5% critical value. The simulation results of the nonparametric ANOVA test for the time effects are given in Table 2. The sizes of the nonparametric ANOVA test were obtained when $U_n = 0$ and $V_n = 0$, which were quite close to the nominal 5%. We observe that the power of the test increased when the distance among $g_1(\cdot)$, $g_2(\cdot)$ and $g_3(\cdot)$ were becoming larger, and when the sample size or repeated measurement $T$ were increased. We noticed that the power was more sensitive to change in $V_n$, the initial phase of the sine function, than $U_n$.

We then compared the proposed tests with a test proposed by Scheike and Zhang (1998). Scheike and Zhang’s test was comparing two treatments for the nonparametric regression model (4.11) for longitudinal data without missing values. Their test was based on a cumulative statistic

$$
T(z) = \int_0^z (\hat{g}_1(t) - \hat{g}_2(t)) dt,
$$
Table 1
Empirical size and power of the ANOVA test for $H_{0a}: \beta_{10} = \beta_{20} = \beta_{30}$.

<table>
<thead>
<tr>
<th>Sample Size</th>
<th>Missingness</th>
<th>Missingness</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n_1$</td>
<td>$n_2$</td>
<td>$n_3$</td>
</tr>
<tr>
<td>60</td>
<td>65</td>
<td>55</td>
</tr>
<tr>
<td>60</td>
<td>65</td>
<td>55</td>
</tr>
<tr>
<td>60</td>
<td>65</td>
<td>55</td>
</tr>
<tr>
<td>60</td>
<td>65</td>
<td>55</td>
</tr>
<tr>
<td>60</td>
<td>65</td>
<td>55</td>
</tr>
<tr>
<td>60</td>
<td>65</td>
<td>55</td>
</tr>
<tr>
<td>60</td>
<td>65</td>
<td>55</td>
</tr>
<tr>
<td>60</td>
<td>65</td>
<td>55</td>
</tr>
<tr>
<td>60</td>
<td>65</td>
<td>55</td>
</tr>
<tr>
<td>100</td>
<td>110</td>
<td>105</td>
</tr>
<tr>
<td>100</td>
<td>110</td>
<td>105</td>
</tr>
<tr>
<td>100</td>
<td>110</td>
<td>105</td>
</tr>
<tr>
<td>100</td>
<td>110</td>
<td>105</td>
</tr>
<tr>
<td>100</td>
<td>110</td>
<td>105</td>
</tr>
<tr>
<td>100</td>
<td>110</td>
<td>105</td>
</tr>
<tr>
<td>100</td>
<td>110</td>
<td>105</td>
</tr>
<tr>
<td>100</td>
<td>110</td>
<td>105</td>
</tr>
<tr>
<td>100</td>
<td>110</td>
<td>105</td>
</tr>
<tr>
<td>325</td>
<td>324</td>
<td>330</td>
</tr>
<tr>
<td>325</td>
<td>324</td>
<td>330</td>
</tr>
<tr>
<td>325</td>
<td>324</td>
<td>330</td>
</tr>
<tr>
<td>325</td>
<td>324</td>
<td>330</td>
</tr>
<tr>
<td>325</td>
<td>324</td>
<td>330</td>
</tr>
<tr>
<td>325</td>
<td>324</td>
<td>330</td>
</tr>
<tr>
<td>325</td>
<td>324</td>
<td>330</td>
</tr>
<tr>
<td>325</td>
<td>324</td>
<td>330</td>
</tr>
<tr>
<td>325</td>
<td>324</td>
<td>330</td>
</tr>
</tbody>
</table>

Table 2
Empirical size and power of ANOVA test for $H_{0b}: g_1(\cdot) = g_2(\cdot) = g_3(\cdot)$ with $\Delta_{2n}(t) = U_n \sin(2\pi t)$ and $\Delta_{3n}(t) = 2\sin(2\pi t) - 2\sin(2\pi (t + V_n))$.

<table>
<thead>
<tr>
<th>Sample Size</th>
<th>Missingness</th>
<th>Missingness</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n_1$</td>
<td>$n_2$</td>
<td>$n_3$</td>
</tr>
<tr>
<td>60</td>
<td>65</td>
<td>55</td>
</tr>
<tr>
<td>60</td>
<td>65</td>
<td>55</td>
</tr>
<tr>
<td>60</td>
<td>65</td>
<td>55</td>
</tr>
<tr>
<td>60</td>
<td>65</td>
<td>55</td>
</tr>
<tr>
<td>60</td>
<td>65</td>
<td>55</td>
</tr>
<tr>
<td>100</td>
<td>110</td>
<td>105</td>
</tr>
<tr>
<td>100</td>
<td>110</td>
<td>105</td>
</tr>
<tr>
<td>100</td>
<td>110</td>
<td>105</td>
</tr>
<tr>
<td>100</td>
<td>110</td>
<td>105</td>
</tr>
<tr>
<td>100</td>
<td>110</td>
<td>105</td>
</tr>
</tbody>
</table>

where $a, z$ are in a common time interval $[0, 1]$. They showed that $\sqrt{n_1 + n_2} T(z)$ converges to a Gaussian Martingale with mean 0 and variance function $\rho_1^{-1} h_1(z) + \rho_2^{-1} h_2(z)$, where $h_j(z) = \int_z^a \sigma_j^2(y) f_j^{-1}(y) dy$. Hence, the test statistic $T(1 - a)/\sqrt{\text{Var}(T(1 - a))}$ is used for two group time-effect functions comparison.

To make the proposed test and the test of Scheike and Zhang (1998) comparable, we conducted simulation in a set-up that mimics the setting of model (6.1) but with only the
first two treatments, no missing values and only the nonparametric part in the regression by setting $\beta_j = 0$. Specifically, we test for $H_0: g_1(\cdot) = g_2(\cdot)$ vs $H_1: g_1(\cdot) = g_2(\cdot) + \Delta_{2n}(\cdot)$ for three cases of the alternative shift function $\Delta_{2n}(\cdot)$ functions which are spelt out in Table 3 and set $\alpha = 0$ in the test of Scheike and Zhang. The simulation results are summarized in Table 3. We found that in the first two cases (I and II) of the alternative shift function $\Delta_{2n}$, the test of Scheike and Zhang had little power. It was only in the third case (III), the test started to pick up some power although it was still not as powerful as the proposed test.

### Table 3

<table>
<thead>
<tr>
<th>Sample Size</th>
<th>$n_1$</th>
<th>$n_2$</th>
<th>$n_3$</th>
<th>$U_n$</th>
<th>Tests</th>
<th>Tests</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>CZ</td>
<td>SZ</td>
</tr>
<tr>
<td>60</td>
<td>65</td>
<td>55</td>
<td></td>
<td></td>
<td>$T$</td>
<td>$T$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.00(size) 5</td>
<td>0.060</td>
<td>0.032</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.30</td>
<td>0.736</td>
<td>0.046</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.50</td>
<td>1.000</td>
<td>0.048</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Case II: $\Delta_{2n}(t) = 2\sin(2\pi t) - 2\sin(2\pi(t + U_n))$</td>
<td>0.05</td>
<td>1.000</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.10</td>
<td>1.000</td>
<td>0.024</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Case III: $\Delta_{2n}(t) = -U_n$</td>
<td>0.10</td>
<td>0.196</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.20</td>
<td>0.562</td>
<td>0.514</td>
</tr>
<tr>
<td>100</td>
<td>110</td>
<td>105</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.00(size) 5</td>
<td>0.056</td>
<td>0.028</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.30</td>
<td>0.982</td>
<td>0.038</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.50</td>
<td>1.000</td>
<td>0.054</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Case II: $\Delta_{2n}(t) = 2\sin(2\pi t) - 2\sin(2\pi(t + U_n))$</td>
<td>0.05</td>
<td>1.000</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.10</td>
<td>1.000</td>
<td>0.026</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Case III: $\Delta_{2n}(t) = -U_n$</td>
<td>0.10</td>
<td>0.290</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.20</td>
<td>0.780</td>
<td>0.774</td>
</tr>
</tbody>
</table>

#### 7. Analyzing the HIV-CD4 Data.

In this section, we analyzed a longitudinal data set from AIDS Clinical Trial Group 193A Study (Henry et al. 1998), which was a randomized, double-blind study of HIV-AIDS patients with advanced immune suppression. The study was carried out in 1993 with 1309 patients who were randomized to four treatments with regard to HIV-1 reverse transcriptase inhibitors. Patients were randomly assigned to one of four daily treatment regimes: 600mg of zidovudine alternating monthly with 400mg didanosine (Treatment I); 600mg of zidovudine plus 2.25mg of zalcitabine (Treatment II); 600mg of zidovudine plus 400mg of didanosine (Treatment III); or 600mg of zidovudine plus 400mg of didanosine plus 400mg of nevirapine (Treatment VI). The four treatments had 325, 324, 330 and 330 patients respectively.

The aim of our analysis was to compare the effects of age (Age), baseline CD4 counts (PreCD4), and gender (Gender) on $Y = \log(\text{CD4 counts} + 1)$. The semiparametric model regression is, for $j = 1, 2, 3$ and 4,

$$Y_{ji}(t) = \beta_{j1}\text{Age}_i(t) + \beta_{j2}\text{PreCD4}_i + \beta_{j3}\text{Gender}_i + g_j(t) + \varepsilon_{ji}(t),$$

with the intercepts absorbed in the nonparametric $g_j(\cdot)$ functions, and $\beta_j = (\beta_{j1}, \beta_{j2}, \beta_{j3})^\top$ is the regression coefficients to the covariates (Age(t), PreCD4, Gender).
To make \( g_{j}(t) \) more interpretable, we centralized \( \text{Age}(t) \) and \( \text{PreCD4} \) so that their sample means in each treatment were 0, respectively. As a result, \( g_{j}(t) \) can be interpreted as the baseline evolution of \( Y \) for a female (Gender=0) with the average PreCD4 counts and the average age in Treatment \( j \). This kind of normalization is used in Wu and Chiang (2000) in their analyzes for another CD4 data set. Our objectives were to detect any difference in the treatments with respect to (i) the covariates; and (ii) the nonparametric baseline functions.

Measurements of CD4 counts were scheduled at the start time 1 and at a 8-week intervals during the follow-up. However, the data were unbalanced due to variations from the planned measurement time and missing values resulted from skipped visits and dropouts. The number of CD4 measurements for patients during the first 40 weeks of follow-up varied from 1 to 9, with a median of 4. There were 5036 complete measurements of CD4, and 2826 scheduled measurements were missing. Hence, considering missing values is very important in this analysis.

7.1. Monotone Missingness. We considered three logistic regression models for the missing propensities and used the AIC and BIC criteria to select the one that was the mostly supported by data. The first model (M1) was a logistic regression model for \( p_{j}(\hat{X}_{jit,3}, \hat{Y}_{jit,3}; \theta_{j0}) \) that effectively depends on \( X_{jit} \) (the PreCD4) and \( (Y_{ji(t-1)}, Y_{ji(t-2)}, Y_{ji(t-3)}) \) if \( t \geq 3 \). For \( t < 3 \), it relies on all \( Y_{jit} \) observed before \( t \). In the second model (M2), we replace the \( X_{jit} \) in the first model with an intercept. In the third model (M3), we added to the second logistic model with covariates representing the square of \( Y_{ji(t-1)} \) and the interactions between \( Y_{ji(t-1)} \) and \( Y_{ji(t-2)} \). The difference of AIC and BIC values among these models for four treatment groups is given in Table 4. Under the BIC criterion, M2 was the best model for all four treatments. For Treatments II and III, M3 had smaller AIC values than M2, but the differences were very small. For Treatments I and VI, M2 had smaller AIC than M3. As the AIC tends to select more explanatory variables, we chose M2 as the model for the parametric missing propensity.

| Table 4 | Difference in the AIC and BIC scores among three models (M1-M3) |
|---------|-----------------|-----------------|-----------------|-----------------|-----------------|
|         | Treatment I     | Treatment II    | Treatment III   | Treatment VI    |
| Models  | AIC  | BIC  | AIC  | BIC  | AIC  | BIC  | AIC  | BIC  | AIC  | BIC  |
| M2-M3   | -2.47| -11.47| 0.93 | -8.12| 0.30 | -8.75| -3.15| -12.27|      |      |

Table 5 reports the parameter estimates \( \hat{\beta}_{j} \) of \( \beta_{j} \) based on the estimating function \( Z_{jit}(\beta_{j}) \) given in Section 3. It contains the standard errors of the estimates, which were obtained from the length of the EL confidence intervals based on the marginal empirical likelihood ratio for each \( \beta_{j} \) as proposed in Chen and Hall (1994). In getting these estimates, we use the ‘leave-one-subject’ cross-validation (Rice and Silverman 1991) to select the smoothing bandwidths \( \{h_{j}\}_{j=1}^{4} \) for the four treatments, which were 12.90, 7.61, 8.27 and 16.20 respectively. We see that the estimates of the coefficients for the Age\( (t) \) and PreCD4 were similar among all four treatments with comparable standard errors, respectively. In particular, the estimates of the Age coefficients endured large variations while the estimates of the PreCD4 coefficients were quite accurate. However, estimates of the Gender coefficients were largely different among the treatments.

We then formally tested \( H_{0a} : \beta_{1} = \beta_{2} = \beta_{3} = \beta_{4} \). The empirical likelihood ratio statistic \( \ell_{n} \) was 20.0486, which was larger than \( \chi^{2}_{2,0.95} = 16.9190 \), which produced a p-value of 0.0176.
This led to rejection of $H_{0a}$ at a significant level 5%. The parameter estimates reported in Table 5 suggested similar covariate effects between Treatments I and II, and between Treatments III and IV, respectively; but different effects between the first two treatments and the last two treatments. To verify this suggestion, we carry out formal ANOVA test for pair-wise equality among the $\beta_j$’s as well as for equality of any three $\beta_j$’s. The p-values of these ANOVA test are reported in Table 6. Indeed, the difference between the first two treatments and between the last two treatments were insignificant. There were significant differences due to the covariates between the first two dual therapy treatments (I and II) and the triple therapy Treatment IV. These differences, in light of all the p-values, was the main cause for the rejection of $H_{0a}$.

We then tested for the nonparametric baseline time effects. The kernel estimates $\hat{g}_j(t)$ are displayed in Figure 1, which shows that Treatments I and II and Treatments III and IV had similar baselines evolution overtime, respectively. However, a big difference existed between the first two treatments and the last two treatments. Treatment IV decreased more slowly than that of the other three treatments, which seemed to be the most effective in slowing down the decline of CD4. We also found that during the first 16 weeks the CD4 counts decrease slowly and then the decline became faster after 16 weeks for Treatments I, II and III.

The $p$-value for testing $H_{0b}: g_1(\cdot) = g_2(\cdot) = g_3(\cdot) = g_4(\cdot)$ is shown in Table 7. The entries were based on 500 bootstrapped resamples according to the procedure introduced in Section 4. The statistics $T_n$ for testing $H_{0b}: g_1(\cdot) = g_2(\cdot) = g_3(\cdot) = g_4(\cdot)$ was 3965.00, where we take $\varpi(t) = 1$ over the range of $t$. The $p$-value of the test was 0.004. Thus, there existed

### Table 5

<table>
<thead>
<tr>
<th>Coefficients</th>
<th>Treatment I</th>
<th>Treatment II</th>
<th>Treatment III</th>
<th>Treatment VI</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta_1$</td>
<td>0.0063(0.0039)</td>
<td>0.0050(0.0040)</td>
<td>0.0047(0.0058)</td>
<td>0.0056(0.0046)</td>
</tr>
<tr>
<td>$\beta_2$</td>
<td>0.7308(0.0462)</td>
<td>0.7724(0.0378)</td>
<td>0.7587(0.0523)</td>
<td>0.8431(0.0425)</td>
</tr>
<tr>
<td>$\beta_3$</td>
<td>0.1009(0.0925)</td>
<td>0.1045(0.0920)</td>
<td>-0.3300(0.1510)</td>
<td>-0.3055(0.1136)</td>
</tr>
<tr>
<td>$\beta_4$</td>
<td>0.9172</td>
<td>0.0847</td>
<td>0.0070</td>
<td>0.0686</td>
</tr>
<tr>
<td>$\beta_5$</td>
<td>0.0168</td>
<td>0.1857</td>
<td>0.0192</td>
<td>0.0320</td>
</tr>
<tr>
<td>$\beta_6$</td>
<td>0.3938</td>
<td>0.0070</td>
<td>0.0192</td>
<td>0.0716</td>
</tr>
</tbody>
</table>

### Table 6

<table>
<thead>
<tr>
<th>$H_{0a}$</th>
<th>$p$-value</th>
<th>$H_{0a}$</th>
<th>$p$-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta_1 = \beta_2$</td>
<td>0.9172</td>
<td>$\beta_1 = \beta_2 = \beta_3$</td>
<td>0.1857</td>
</tr>
<tr>
<td>$\beta_1 = \beta_3$</td>
<td>0.0847</td>
<td>$\beta_1 = \beta_2 = \beta_4$</td>
<td>0.0200</td>
</tr>
<tr>
<td>$\beta_1 = \beta_4$</td>
<td>0.0070</td>
<td>$\beta_1 = \beta_3 = \beta_4$</td>
<td>0.0192</td>
</tr>
<tr>
<td>$\beta_2 = \beta_3$</td>
<td>0.0686</td>
<td>$\beta_2 = \beta_3 = \beta_4$</td>
<td>0.0320</td>
</tr>
<tr>
<td>$\beta_2 = \beta_4$</td>
<td>0.0168</td>
<td>$\beta_1 = \beta_2 = \beta_3 = \beta_4$</td>
<td>0.0176</td>
</tr>
<tr>
<td>$\beta_3 = \beta_4$</td>
<td>0.3938</td>
<td>$\beta_1 = \beta_2 = \beta_3 = \beta_4$</td>
<td>0.0176</td>
</tr>
</tbody>
</table>

We then tested for the nonparametric baseline time effects. The kernel estimates $\hat{g}_j(t)$ are displayed in Figure 1, which shows that Treatments I and II and Treatments III and IV had similar baselines evolution overtime, respectively. However, a big difference existed between the first two treatments and the last two treatments. Treatment IV decreased more slowly than that of the other three treatments, which seemed to be the most effective in slowing down the decline of CD4. We also found that during the first 16 weeks the CD4 counts decrease slowly and then the decline became faster after 16 weeks for Treatments I, II and III.

Figure 1: (a) The raw data excluding missing values plots with the estimates of $g_j(t)$ ($j = 1, 2, 3, 4$). (b) The estimates of $g_j(t)$ in the same plot: Treatment I (solid line), Treatment II (short dashed line), Treatment III (dashed and dotted line) and Treatment IV (long dashed line).
significant difference in the baseline time effects $g_j(\cdot)$’s among Treatments I-IV. At the same time, we also calculate the test statistics $T_n$ for testing $g_1(\cdot) = g_2(\cdot)$ and $g_3(\cdot) = g_4(\cdot)$. The statistics values were 19.26 and 26.22, with p-values 0.894 and 0.860, respectively. From Table 7, we see the p-value is much bigger than 0.05. We conclude that treatment I and II has similar baseline time effects, but they are significantly distinct from the baseline time effects of treatment III and IV, respectively. P-values of testing other combinations on equalities of $g_1(\cdot), g_2(\cdot), g_3(\cdot)$ and $g_4(\cdot)$ are also reported in Table 7.

### Table 7

<table>
<thead>
<tr>
<th>$H_{06}$</th>
<th>p-value</th>
<th>$H_{06}$</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$g_1(\cdot) = g_2(\cdot)$</td>
<td>0.894</td>
<td>$g_1(\cdot) = g_2(\cdot) = g_3(\cdot)$</td>
<td>0.046</td>
</tr>
<tr>
<td>$g_1(\cdot) = g_3(\cdot)$</td>
<td>0.018</td>
<td>$g_1(\cdot) = g_2(\cdot) = g_4(\cdot)$</td>
<td>0.010</td>
</tr>
<tr>
<td>$g_2(\cdot) = g_4(\cdot)$</td>
<td>0.004</td>
<td>$g_1(\cdot) = g_3(\cdot) = g_4(\cdot)$</td>
<td>0.000</td>
</tr>
<tr>
<td>$g_3(\cdot) = g_4(\cdot)$</td>
<td>0.020</td>
<td>$g_1(\cdot) = g_3(\cdot) = g_4(\cdot)$</td>
<td>0.014</td>
</tr>
<tr>
<td>$g_4(\cdot) = g_4(\cdot)$</td>
<td>0.006</td>
<td>$g_1(\cdot) = g_2(\cdot) = g_3(\cdot)$</td>
<td>0.004</td>
</tr>
<tr>
<td>$g_4(\cdot) = g_4(\cdot)$</td>
<td>0.860</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

#### 7.2. Not-monotone Missingness.
We analyzed the model the data without assuming missing as monotone for the missing values in this subsection. Instead of monotone assumption, we assume the missing propensity depends on the past $d(t)$ observations for a given time $t$ as we described at Section 2. Recall that from Section 2, if we assume small $d$ for the missing propensity function, more data could be used for analysis than monotone assumption. We presented the results with $d = 1, 2, 3$ in this subsection.

For $d = 1$, three logistic models were used to model the missing propensity functions. In the first model (M1) we include intercept, PreCD4, and $Y_{j(t−1)}$ as covariates. In the second model (M2), only intercept and $Y_{j(t−1)}$ are included. In the third model, we used a nonlinear model with intercept, $Y_{j(t−1)}$, $Y_{j(t−1)}^2$, and PreCD4 as covariates. As we did in previous monotone case, AIC and BIC values are given in the following Table 7. We observed that model M1 had the smallest AIC at four Treatments among M1-M3. M1 also had the smaller BIC values than M3, for Treatment II-IV, M2 had slightly smaller BIC values. So, overall we would choose M1 to model the missing propensity. For $d = 2$ and $d = 3$, we chose the missing propensity function in a similar way, but we do not report the AIC and BIC values here for saving space.

### Table 7: Difference in the AIC and BIC scores among three models (M1-M3)

<table>
<thead>
<tr>
<th>Models</th>
<th>Treatment I</th>
<th>Treatment II</th>
<th>Treatment III</th>
<th>Treatment VI</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>AIC</td>
<td>BIC</td>
<td>AIC</td>
<td>BIC</td>
</tr>
<tr>
<td>M1-M2</td>
<td>-7.885</td>
<td>-2.992</td>
<td>-3.870</td>
<td>1.039</td>
</tr>
</tbody>
</table>

Table 8 reported the parameter estimates and their corresponding standard errors. The estimates for the coefficient of PrdCD4 are very much similar for $d = 1, 2, 3$, but the estimates for the coefficient of Age(t) and Gender seems more variable among $d = 1, 2, 3$. Nevertheless, all of the estimates at one $d$ value are in the 95% confidence interval of the estimates at another $d$ value. For example, the 95% confidence interval for PreCD4 in Treatment I with $d = 1$ is (0.6812, 0.8368) and the corresponding estimates with $d = 2, 3$
are in this confidence interval. Basically, we may say the estimates at \( d = 1, 2, 3 \) are not significantly different.

Next, we summarized the ANOVA test results on \( \beta \)s with \( d = 1, 2, 3 \) at Table 9. The p-values are consistent in the sense that the order of the p-values at different \( d \) values were almost the same. For instance, the test for \( \beta_2 = \beta_4 \) always had the smallest p-value among all the p-values with same \( d \). We observed that when \( d = 2, 3 \) the tests for \( \beta_1 = \beta_4 \) and \( \beta_2 = \beta_4 \) had p-values less than 0.05. The tests between \( \beta_1, \beta_2 \) and \( \beta_4 \) had smaller p-values than the other tests. All the test results showed the similarity treatment effects due to covariates among Treatments I-III (dual therapy treatments) and difference with Treatment IV (triple therapy treatments).

Finally, Table 10 illustrate the ANOVA test for the nonparametric baseline time effect functions. The p-values are obtain from the bootstrap calibration test we introduced in Section 5. Each p-value were based on 500 times resampling. The bandwidth selection method and the weight function \( w(t) \) are the same with the monotone case. We found that the p-values when \( d = 3 \) are quite similar to the monotone case.

This data set has been analyzed by Fitzmaurice, Laird and Ware (2004) using a random effects model that applied the Restricted Maximum Likelihood (REML) method. They
conducted a two sample comparison test via parameters in the model for the difference between the dual therapy (Treatment I-III) versus triple therapy (Treatment VI) without considering the missing values. More specifications, they denote Group = 1 if subject in the triple therapy treatment and Group = 0 if subject in the dual therapy treatment, and the linear mixed effect was

$$E(Y|b) = \beta_1 + \beta_2 t + \beta_3 (t - 16)_+ + \beta_4 \text{Group} \times t + \beta_5 \text{Group} \times (t - 16)_+ + b_1 + b_2 t + b_3 (t - 16)_+,$$

where $b = (b_1, b_2, b_3)$ are random effects. They tested $H_0 : \beta_4 = \beta_5 = 0$. This is equivalent to test the null hypothesis of no treatment group difference in the changes in log CD4 counts between therapy and dual treatments. Both Wald test and likelihood ratio test rejected the null hypothesis, indicating the difference between dual and triple therapy in the change of log CD4 counts. This result is consistent with the result we illustrated in Table 6 and 10.

8. Appendix: Proofs. The following assumptions are made in the paper:

A1. Let $S(\theta_j)$ be the score function of the partial likelihood $L_{b_j}(\theta_j)$ for a q-dimensional parameter $\theta_j$ defined in (2.3), and $\theta_{j0}$ is in the interior of compact $\Theta_j$. We assume $E\{S(\theta_j)\} \neq 0$ if $\theta_j \neq \theta_{j0}$, $\text{Var}(S(\theta_{j0}))$ is finite and positive definite, and $E(\frac{\partial S(\theta_{j0})}{\partial \theta_{j0}})$ exists and is invertible. The missing propensity $\pi_{ijm}(\theta_{j0}) > b_0 > 0$ for all $j, i, m$.

A2. (i) The kernel function $K$ is a symmetric probability density which is differentiable of Lipschitz order 1 on its support $[-1,1]$. The bandwidths satisfy $n_j h_j^2 / \log^2 n_j \to \infty$, $n_j^{1/2} h_j^4 \to 0$ and $h_j \to 0$ as $n_j \to \infty$.

(ii) For each design point $j$, $(j = 1, \cdots, k)$, the design points $\{t_{ijm}\}$ are thought to be independent and identically distributed from a super-population with density $f_j(t)$. There exist constants $b_l$ and $b_u$ such that $0 < b_l \leq \sup_{t \in S} f_j(t) \leq b_u < \infty$.

(iii) For each $h_j$ and $T_j$, $j = 1, \cdots, k$, there exist finite positive constants $\alpha_j$, $b_j$ and $T$ such that $\alpha_j T_j = T$ and $b_j h_j = h$ for some $h$ as $h \to 0$. Let $n = \sum_{i=1}^k n_j, n_j/n \to \rho_j$ for some non-zero $\rho_j$ as $n \to \infty$ such that $\sum_{i=1}^k \rho_j = 1$.
A3. The residuals \( \{ \varepsilon_{ji} \} \) and \( \{ u_{ji} \} \) are independent of each other and each of \( \{ \varepsilon_{ji} \} \) and \( \{ u_{ji} \} \) are mutually independent among different \( j \) or \( i \), respectively; \( \text{max}_{1\leq i\leq n_j} E|\varepsilon_{jim}|^{4+r} < \infty \), \( \text{max}_{1\leq i\leq n_j} \| u_{jim} \| = o_p(\sqrt[n_j]{\log n_j})^{-1} \) for some \( r > 0 \); \( \{ u_{jim} \} \)
satisfy, for each \( j \)
\[
\lim_{n_j \to \infty} \sum_{i=1}^{n_j} \sum_{m=1}^{T_j} E\{ \frac{\delta_{jim}}{\pi_{jim}(\theta_{jm})} u_{jim}^r u_{jim} \} = \Sigma_u > 0.
\]

A4. The functions \( g_{j0}(t) \) and \( h_j(t) \) are, respectively, 1-dimensional and \( p \)-dimensional smooth functions with continuously second derivatives on \( S = [0, 1] \).

Remark: Condition A1 are the regular conditions for the consistency of the binary MLE for the parameters in the missing propensity. Condition A2(i) are the usual conditions for different samples. The positive definite of \( \Sigma_u \) and \( \Sigma_u \) is a mild assumption similar to the ones in Müller (1987). Condition A2(ii) is a common assumption similar to the ones in Möller (1987). Condition A2(iii) is a mild assumptions on the the relationship between bandwidths and sample sizes among different samples. The positive definite of \( \Sigma_u \) in Condition A3 is used for model identification (Härde, Liang and Gao, 2000).

Derivation of (3.9): To appreciate this, we note from (3.7) that via standard derivations in empirical likelihood (Owen, 1990) that \( \| \lambda_j \| = O_p(n_j^{-1/2}) \), and
\[
\lambda_j = (\sum_{i=1}^{n_j} Z_{ji}(\beta) Z_{ji}(\beta)^T)^{-1} \sum_{i=1}^{n_j} Z_{ji}(\beta) + o_p(n_j^{-1/2}), \quad j = 1, 2, \ldots, k.
\]

Then we can write
\[
\ell_n = 2 \min_{\beta} \left\{ \sum_{j=1}^{k} \left( \sum_{i=1}^{n_j} Z_{ji}^T(\beta) \left( \sum_{i=1}^{n_j} Z_{ji}(\beta) Z_{ji}(\beta)^T \right)^{-1} \sum_{i=1}^{n_j} Z_{ji}(\beta) \right) + o_p \left( \min_{j} n_j^{-1/2} \right) \right\}
\]
(4.1)
\[
\left( \sum_{j=1}^{k} \frac{1}{n_j T_j} \left( \sum_{i=1}^{n_j} Z_{ji}^T(\beta) B_j^{-1} \sum_{i=1}^{n_j} Z_{ji}(\beta) \right) + o_p \left( \min_{j} n_j^{-1/2} \right) \right)
\]
where \( B_j := \lim_{n_j \to \infty} \frac{1}{n_j T_j} \sum_{i=1}^{n_j} E\{ Z_{ji}(\beta_0) Z_{ji}(\beta_0)^T \} \), which is not related with \( \beta \) for any \( \beta = \beta_0 + \Delta_jn \) and \( \Delta_jn = O(n_j^{-1/2}) \).

Using the Lagrange method to carry out the minimizations in (A.1), we want to minimize
\[
Q = \frac{1}{2} \sum_{j=1}^{k} \frac{1}{n_j T_j} \left( \sum_{i=1}^{n_j} Z_{ji}^T(\beta_j) B_j^{-1} \sum_{i=1}^{n_j} Z_{ji}(\beta_j) \right) \sum_{j=2}^{k} \eta_j (\beta_1 - \beta_j),
\]
where \( \eta_1, \ldots, \eta_k \) are lagrange multipliers. Then
\[
\frac{\partial Q}{\partial \beta_1} = \frac{1}{n_1 T_1} \sum_{i=1}^{n_1} Z_{i1}^T(\beta_1) B_1^{-1} \sum_{i=1}^{n_1} T_1 \frac{\delta_{1im}}{\pi_{1im}(\theta)} \bar{X}_{1im} \bar{X}_{1im} - \sum_{j=2}^{k} \eta_j,
\]
and
\[
\frac{\partial Q}{\partial \beta_j} = \frac{1}{n_j T_j} \sum_{i=1}^{n_j} Z_{ji}^T(\beta_j) B_j^{-1} \sum_{i=1}^{n_j} T_j \frac{\delta_{jim}}{\pi_{jim}(\theta)} \bar{X}_{jim} \bar{X}_{jim} + \eta_j, \quad j = 2, \ldots, k,
\]
Setting \( \beta_1 = \beta_2 = \cdots = \beta_k = \beta \), then the minima \( \beta \) satisfies

\[
A.2 \quad \sum_{j=1}^{k} \frac{1}{\sqrt{n_j}T_j} \Omega_{x_j} B_j^{-1} \sum_{i=1}^{n_j} Z_{ji}(\beta) = o_p(1).
\]

Inverting \((A.2)\) for \( \beta \), we have

\[
\beta = \left( \sum_{j=1}^{k} \Omega_{x_j} B_j^{-1} \Omega_{y_j} \right)^{-1} \left( \sum_{j=1}^{k} \Omega_{x_j} B_j^{-1} \Omega_{x_j y_j} \right) + o_p(1).
\]

**LEMMA 1.** Suppose \((e_{i1}, \ldots, e_{iT})_{i=1}^{T} \) is a sequence of \(T\)-dimensional independent random vectors and \( T \) is a fixed finite number, and \( \max_{1 \leq i \leq n} E(|e_{im}|^4) < \infty \) for some \( \delta > 1 \) and all \( m \). Let \( \{a_{jim}, 1 \leq i, j \leq n, 1 \leq m \leq T \} \) be a collection of real numbers such that \( \max_{1 \leq j \leq n} \sum_{i=1}^{n} \sum_{m=1}^{T} |a_{jim}| < \infty \). Let \( d_n = \max_{1 \leq i, j \leq n, 1 \leq m \leq T} |a_{jim}| \), then

\[
\max_{1 \leq j \leq n} \left| \sum_{i=1}^{n} \sum_{m=1}^{T} a_{jim} e_{im} \right| = O\{n^{1/\delta} d_n (\log n)^{1/2} \} \quad a.s.
\]

**PROOF.** This can be proved in a similar way as Lemma 1 of Shi and Lau (2000).

**LEMMA 2.** Under assumptions A1, A2(i), A3 and A4, we have

(i) if \( \tilde{g}(t) = g(t) - \sum_{i_1=1}^{n_j} \sum_{m_1=1}^{T_j} w_{jim_1} A(t_{jim_1} g(t_{jim_1} m_1), \) then \( \tilde{g}(t) = O_p(h^2 + \frac{1}{\sqrt{n_j}h}) \);

(ii) for \( B_j \) defined after \((A.1)\) and \( \beta = \beta_{0j} + \Delta_{jn} \), where \( \beta_{0j} \) is the true value of \( \beta_j \) and \( \Delta_{jn} = O(n^{-1/2}) \),

\[
B_j = \lim_{n_j \to \infty} \frac{1}{n_j} \sum_{i=1}^{n_j} \sum_{m=1}^{T_j} E\{\nu_{jim} \nu_{jim1} \varepsilon_{jim} \varepsilon_{jim1} u_{jim} u_{jim1}^{T} \},
\]

where \( \nu_{jim} = \delta_{jim} \pi_{jim}^{-1}(\theta_{j0}) \);

(iii) for any \( 1 \leq l \neq g \leq k \), under the hypothesis: \( \beta_{0j} = \beta_{0g} \),

\[
\{(\Omega_{x_l}^{-1} B_l \Omega_{x_l}^{-1}) + (\Omega_{x_g}^{-1} B_g \Omega_{x_g}^{-1})\}^{-1/2}(\Omega_{x_l}^{-1} \Omega_{x_l t} - \Omega_{x_g}^{-1} \Omega_{x_g t}) \overset{d}{\to} N(0, I_p).
\]

**PROOF.** We only give the proofs for (i) and (iii) as that for (ii) is straightforward. For convenience we define \( \tilde{A}(t_{jim}) = \sum_{i_1=1}^{n_j} \sum_{m_1=1}^{T_j} w_{jim_1} A(t_{jim_1} m_1) \) for a generic function \( A \). The result in (i) will be true if \( \text{Bias}\{\tilde{g}(t)\} = O(h^2) \) and \( \text{Var}\{\tilde{g}(t)\} = O\{(n_j h)^{-1}\} \). Note that

\[
\tilde{g}(t) = \frac{\tilde{\varphi}(t)}{f(t)} := \frac{1}{n_j} \sum_{i_1=1}^{n_j} \sum_{m_1=1}^{T_j} (\delta_{jim} \pi_{jim}(\hat{\theta}_j)) K_h(t_{jim} - t) g(t_{jim}).
\]

Following a standard procedure in nonparametric regression, it can be shown that \( E\{\tilde{g}(t)\} = \mu_{\varphi}(t)/\mu_f(t) + O(h^2) \) and

\[
\text{Var}\{\tilde{g}(t)\} = \text{Var}(\varphi)/\mu_f^2(t) + \mu_{\varphi}^2(t) \text{Var}(\hat{f})/\mu_f^4(t) - 2 \mu_{\varphi}(t) \text{Cov}(\hat{f}, \varphi)/\mu_f^3(t)
\]
where $\mu_p(t) = E\{\hat{\phi}(t)\}$ and $\mu_f(t) = E\{\hat{f}(t)\}$. Now we can write
\[
\frac{\delta_{jm}}{\pi_{jm}(\theta_j)} = \left\{ \delta_{jm} \pi^{-1}_{jm}(\theta_j) \right\} \left\{ 1 - \frac{\pi'_{jm}(\theta_j)}{\pi_{jm}(\theta_j)}(\hat{\theta}_j - \theta_j) + o_p\left(\frac{1}{n_j}\right) \right\}.
\]
By the MAR assumption, we have $\mu_p(t) = \frac{1}{n_j T_j} \sum_{i=1}^{n_j} \sum_{m=1}^{T_j} K_h(t_{jm} - t)g(t_{jm})\{1 + o(1)\}$ and $\mu_f(t) = \frac{1}{n_j T_j} \sum_{i=1}^{n_j} \sum_{m=1}^{T_j} K_h(t_{jm} - t)\{1 + o(1)\}$. Since $\{t_{jm}\}$ satisfy Condition A2(ii), a Taylor expansion can be used to show that $\text{Bias}\{\hat{g}(t)\} = O(h^2)$.

For $\text{Var}(\hat{\phi})$, we can get the following expansion,
\[
\text{Var}(\hat{\phi}) = \frac{1}{(n_j T_j)^2} \sum_{i=1}^{n_j} \left( \frac{T_j}{\pi_{jm}(\theta_j)} \sum_{m=1}^{T_j} K_h(t_{jm} - t)g(t_{jm}) - \mu_{f,jm} \right)^2 + \frac{1}{(n_j T_j)^2} \sum_{i=1}^{n_j} \sum_{m=1}^{T_j} E\left[ \left( \frac{\delta_{jm}}{\pi_{jm}(\theta_j)} \sum_{m=1}^{T_j} K_h(t_{jm} - t)g(t_{jm}) - \mu_{f,jm} \right) \times \left( \frac{\delta_{jm}}{\pi_{jm}(\theta_j)} \sum_{m=1}^{T_j} K_h(t_{jm} - t)g(t_{jm}) - \mu_{f,jm} \right) \right],
\]
where $\mu_{f,jm} = E\{\frac{\delta_{jm}}{\pi_{jm}(\theta_j)} K_h(t_{jm} - t)g(t_{jm})\}$. The first term is obviously $O\{n_j h^{-1}\}$, since $T_j$ is fixed and finite. The second term equals
\[
\frac{1}{(n_j T_j)^2} \sum_{i=1}^{n_j} \sum_{m=1}^{T_j} E\left\{ \left( \frac{\delta_{jm}}{\pi_{jm}(\theta_j)} K_h(t_{jm} - t)g(t_{jm}) - (1 - \frac{\pi'_{jm}(\theta_j)}{\pi_{jm}(\theta_j)}(\hat{\theta}_j - \theta_j)) - \mu_{f,jm} \right) \times \left( \frac{\delta_{jm}}{\pi_{jm}(\theta_j)} K_h(t_{jm} - t)g(t_{jm}) - (1 - \frac{\pi'_{jm}(\theta_j)}{\pi_{jm}(\theta_j)}(\hat{\theta}_j - \theta_j)) - \mu_{f,jm} \right) \right\},
\]
Therefore, $\text{Var}(\hat{\phi}) = O\{n_j h^{-1}\}$. In a similar way, we can prove that $\text{Var}(\hat{f})$ and $\text{Cov}(\hat{f}, \hat{\phi})$ are also $O\{n_j h^{-1}\}$. Therefore, we have $\text{Var}\{\hat{g}(t)\} = O\{(n_j h)^{-1}\}$.

We now prove (iii). Since we know that
\[
\Omega_{x_j}^{-1} \Omega_{x_j,y_j} = \frac{1}{\sqrt{n_j T_j}} \sum_{i=1}^{n_j} \sum_{m=1}^{T_j} \frac{\delta_{jm}}{\pi_{jm}(\theta_j)}(\hat{X}_{jm} \check{Y}_{jm} - \check{X}_{jm} \check{Y}_{jm})
\]
\[
= \frac{1}{\sqrt{n_j T_j}} \sum_{i=1}^{n_j} \sum_{m=1}^{T_j} \frac{\delta_{jm}}{\pi_{jm}(\theta_j)}(\hat{X}_{jm} \check{Y}_{jm} - \check{X}_{jm} \beta_{jm} + \beta_{jm})
\]
\[
= \frac{1}{\sqrt{n_j T_j}} \sum_{i=1}^{n_j} \sum_{m=1}^{T_j} \frac{\delta_{jm}}{\pi_{jm}(\theta_j)}(\hat{X}_{jm} \check{Y}_{jm} - \check{X}_{jm} \beta_{jm}) + \beta_{jm}
\]
\[
= \frac{1}{\sqrt{n_j T_j}} \sum_{i=1}^{n_j} \sum_{m=1}^{T_j} \frac{\delta_{jm}}{\pi_{jm}(\theta_j)}(\hat{X}_{jm} \check{Y}_{jm} - \check{X}_{jm} \beta_{jm}) + \beta_{jm},
\]
and because samples $l$ and $g$ are mutually independent, we need to show that, for $j = l, g$,
\[
(\Omega_{x_j}^{-1} B_j \Omega_{x_j}^{-1})^{-1/2} \Omega_{x_j}^{-1} \Omega_{x_j,y_j} \overset{d}{\rightarrow} N(0, I_p),
\]
which is equivalent to show that
\[
\frac{1}{\sqrt{n_j T_j}} \sum_{i=1}^{n_j} \sum_{m=1}^{T_j} (\hat{Z}_{jm}(\beta_{jm}) \overset{d}{\rightarrow} N(0, B_j).
Recall that $\tilde{Y}_{jim} = \tilde{X}_{jim} \beta_j + \tilde{g}_j(t_{jim}) + \tilde{\varepsilon}_{jim}$, where $\tilde{g}_j(t_{jim}) = g_j(t_{jim}) - \bar{g}_j(t_{jim})$, $\tilde{\varepsilon}_{jim} = \varepsilon_{jim} - \bar{\varepsilon}_{jim}$ and $\mathcal{A}(t_{jim}) = \sum_{i=1}^{n_j} \sum_{m=1}^{T_j} w_{jim1,h}(t_{jim}) A(t_{jim})$. Then, it follows that

$$
\sum_{i=1}^{n_j} Z_{ji}(\beta_j) = \sum_{i=1}^{n_j} \sum_{m=1}^{T_j} \delta_{jim} \tilde{X}_{jim} \{\tilde{g}_j(t_{jim}) + \tilde{\varepsilon}_{jim}\}
$$

$$
= \sum_{i=1}^{n_j} \sum_{m=1}^{T_j} \left\{ \delta_{jim} \pi_{jim}^{-1}(\theta_0) \right\} \left\{ 1 - \frac{\pi_{jim}(\theta_0)}{\pi_{jim}(\theta_0)} (\tilde{\theta}_j - \theta_0) + o_p(n_j^{-1/2}) \right\} \tilde{X}_{jim} \{\tilde{g}_j(t_{jim}) + \tilde{\varepsilon}_{jim}\}
$$

$$
= \sum_{i=1}^{n_j} \sum_{m=1}^{T_j} \left\{ \delta_{jim} \pi_{jim}^{-1}(\theta_0) \right\} \tilde{X}_{jim} \{\tilde{g}_j(t_{jim}) + \tilde{\varepsilon}_{jim}\} \{1 + o_p(1)\}.
$$

The last equality is true, since $\tilde{\theta}_j - \theta_0 = O_p(n_j^{-1/2})$. At the same time, we can decompose

$$
\tilde{X}_{jim} \{\tilde{g}_j(t_{jim}) + \tilde{\varepsilon}_{jim}\} = \{\tilde{h}(t_{jim}) + u_{jim} - \bar{u}(t_{jim})\} \{\tilde{g}_j(t_{jim}) + \varepsilon_{jim} - \bar{\varepsilon}_{jim}\}
$$

$$
= u_{jim} \varepsilon_{jim} + \{\tilde{h}(t_{jim}) - \bar{u}(t_{jim})\} \varepsilon_{jim} + \{\tilde{g}_j(t_{jim}) - \bar{g}_j(t_{jim}) - \bar{\varepsilon}_{jim}\} u_{jim}
$$

$$
+ \{\tilde{h}(t_{jim}) - \bar{u}(t_{jim})\}(\tilde{g}_j(t_{jim}) - \bar{\varepsilon}_{jim})
$$

$$
:= I_1 + I_2 + I_3, \text{ say.}
$$

From assumptions in A2(i) and the facts that max$_{1 \leq i, i' \leq n_j, 1 \leq m, m' \leq T_j} w_{jim1,m'1,h}(t_{jim}) = O(\{n_j, h_j\}^{-1})$ and $\sum_{i=1}^{n_j} \sum_{m=1}^{h_j} w_{jim1,m1,h}(t_{jim}) = 1$, we have, by applying Lemma 1

$$(A.3) \quad \max_{1 \leq i \leq n_j} \|\tilde{h}(t_{jim}) - \bar{u}(t_{jim})\| = o(1) \text{ a.s.,} \quad \max_{1 \leq i \leq n_j} |\tilde{g}_j(t_{jim}) - \bar{\varepsilon}_{jim}| = o(1) \text{ a.s.,}
$$

$$
\max_{1 \leq i \leq n_j} \|\tilde{h}(t_{jim}) - \bar{u}(t_{jim})\| (\tilde{g}_j(t_{jim}) - \bar{\varepsilon}_{jim}) = o(n_j^{-1/2}) \text{ a.s.}
$$

Therefore,

$$
\frac{1}{\sqrt{n_j T_j}} \sum_{i=1}^{n_j} Z_{ji}(\beta_j) = \frac{1}{\sqrt{n_j T_j}} \sum_{i=1}^{n_j} \sum_{m=1}^{T_j} \left\{ \delta_{jim} \pi_{jim}^{-1}(\theta_0) \right\} (I_1 + I_2 + I_3)
$$

$$
:= J_1 + J_2 + J_3, \text{ say.}
$$

It is easy to see $J_3 = o_p(1)$, and from (A.3),

$$
|J_2| \leq o_p(1) \times \left\|\frac{1}{\sqrt{n_j T_j}} \sum_{i=1}^{n_j} \sum_{m=1}^{T_j} \left\{ \delta_{jim} \pi_{jim}^{-1}(\theta_0) \right\} u_{jim} \right\|
$$

$$
+ o_p(1) \times \left\|\frac{1}{\sqrt{n_j T_j}} \sum_{i=1}^{n_j} \sum_{m=1}^{T_j} \left\{ \delta_{jim} \pi_{jim}^{-1}(\theta_0) \right\} \varepsilon_{jim} \right\| = o_p(1).
$$

We note that Var$(J_1) = B_J$ and $J_1$ is a sum of independent random variables. Therefore, we will complete our proof by verifying the Linderberg-Feller condition for $\alpha'J_1$, for any $\alpha \in \mathbb{R}^p$. Let $\nu_{jim} = \delta_{jim} \pi_{jim}^{-1}(\theta_0)$. Then for any $\epsilon > 0$, let

$$
L_n := \sum_{i=1}^{n_j} \text{Var} \left\{ \sum_{m=1}^{T_j} \alpha' u_{jim} \nu_{jim} \varepsilon_{jim} \right\} = O(n_j),
$$
\[ \Lambda_n(\epsilon) = \frac{1}{L_n} \sum_{i=1}^{n_j} E \left[ I \left( \sum_{m=1}^{T_j} \alpha' u_{jim} v_{jim} \varepsilon_{jim} \geq \epsilon \sqrt{L_n} \right) \right] \]

\[ \leq \frac{1}{L_n} \sum_{i=1}^{n_j} \left( E \left( I \left( \sum_{m=1}^{T_j} \alpha' u_{jim} v_{jim} \varepsilon_{jim} \geq \epsilon \sqrt{L_n} \right) \right) \right)^{2+r} \left( E \left| \sum_{m=1}^{T_j} \alpha' u_{jim} v_{jim} \varepsilon_{jim} \right|^{4+r} \right)^{\frac{2}{4+r}} \]

\[ \leq \frac{1}{L_n} \sum_{i=1}^{n_j} \sum_{m=1}^{T_j} \alpha' u_{jim} v_{jim} \varepsilon_{jim} \left| \varepsilon_{jim} \right|^{4+r} \log((4+r)(n_j)) \rightarrow 0, \]

where \( C \) is a finite positive constant. This completes the proof of (iii).

**Lemma 3.** If the conditions A1, A2, A3 and A4 hold, then under null hypothesis \( H_{0a} \), i.e. \( \beta_{10} = \beta_{20} \), \( \ell_n \to \chi^2_p \).

**Proof.** Let \( S_1 := \sum_{j=1}^{k} \Omega x_j B_j^{-1} \Omega x_j \) and \( S_2 := \sum_{j=1}^{k} \Omega x_j B_j^{-1} \Omega x_j y_j \). In this Lemma, \( k = 2 \). Then,

\[ \ell_n = (\Omega^{r}_{x_1 y_1} - S_2^r S_1^{-1} \Omega x_1) B_1^{-1} (\Omega x_{1 y_1} - \Omega x_1 S_1^{-1} S_2) \]

\[ + (\Omega^{r}_{x_2 y_2} - S_2^r S_1^{-1} \Omega x_2) B_2^{-1} (\Omega x_{2 y_2} - \Omega x_2 S_1^{-1} S_2) + o_p(1) \]

\[ = (\Omega^{r}_{x_1 y_1} - S_2^r S_1^{-1} \Omega x_1) B_1^{-1} \Omega x_{1 y_1} S_1^{-1} (\Omega x_1 \Omega x_{1 y_1} - S_2) \]

\[ + (\Omega^{r}_{x_2 y_2} - S_2^r S_1^{-1} \Omega x_2) B_2^{-1} \Omega x_{2 y_2} S_1^{-1} (\Omega x_2 \Omega x_{2 y_2} - S_2) + o_p(1) \]

It is easy to show that

\[ S_1 (\Omega x_1^{-1} \Omega x_{1 y_1} - S_2) = \Omega x_2 B_2^{-1} \Omega x_{2 y_2} (\Omega x_2^{-1} \Omega x_{1 y_1} - \Omega x_2^{-1} \Omega x_{2 y_2}), \]

\[ S_1 (\Omega x_2^{-1} \Omega x_{2 y_2} - S_2) = \Omega x_1 B_1^{-1} \Omega x_{1 y_1} (\Omega x_1^{-1} \Omega x_{2 y_2} - \Omega x_1^{-1} \Omega x_{1 y_1}). \]

Then

\[ \ell_n = (\Omega^{r}_{x_1 y_1} - S_2^r S_1^{-1} \Omega x_1) V (\Omega x_{1 y_1} - \Omega x_{2 y_2}) + o_p(1), \]

where

\[ V = (\Omega x_2 B_2^{-1} \Omega x_2) S_1^{-1} (\Omega x_2 B_1^{-1} \Omega x_1) S_1^{-1} (\Omega x_2 B_2^{-1} \Omega x_2) \]

\[ + (\Omega x_1 B_1^{-1} \Omega x_1) S_1^{-1} (\Omega x_2 B_2^{-1} \Omega x_2) S_1^{-1} (\Omega x_1 B_1^{-1} \Omega x_1) \]

\[ := P_1 + P_2, \text{say.} \]

We note that

\[ P_1 = (\Omega x_2 B_2^{-1} \Omega x_2) S_1^{-1} (\Omega x_2 B_1^{-1} \Omega x_1) - (\Omega x_2 B_2^{-1} \Omega x_2) S_1^{-1} (\Omega x_1 B_1^{-1} \Omega x_1) S_1^{-1} (\Omega x_2 B_1^{-1} \Omega x_1) \]

and

\[ P_2 = (\Omega x_2 B_2^{-1} \Omega x_2) S_1^{-1} (\Omega x_1 B_1^{-1} \Omega x_1) - (\Omega x_2 B_2^{-1} \Omega x_2) S_1^{-1} (\Omega x_2 B_2^{-1} \Omega x_2) S_1^{-1} (\Omega x_1 B_1^{-1} \Omega x_1). \]
It follows that \( V = (\Omega_x B^{-1}_2 \Omega_x) S_1^{-1} (\Omega_x B^{-1}_1 \Omega_x) \). Thus, to prove the theorem, we just need to show that

\[
(\Omega^{-1}_x x_1 y_1 - \Omega^{-1}_x x_2 y_2)^\tau \left( (\Omega^{-1}_x B_1 \Omega^{-1}_x) + (\Omega^{-1}_x B_2 \Omega^{-1}_x) \right)^{-1} (\Omega^{-1}_x x_1 y_1 - \Omega^{-1}_x x_2 y_2) \xrightarrow{d} \chi^2_p,
\]

which is true as Lemma 2(iii) implies

\[
\left( (\Omega^{-1}_x B_1 \Omega^{-1}_x) + (\Omega^{-1}_x B_2 \Omega^{-1}_x) \right)^{-1/2} (\Omega^{-1}_x x_1 y_1 - \Omega^{-1}_x x_2 y_2) \xrightarrow{d} N(0, I_p).
\]

This completes the proof of Lemma 3.

**Proof of Theorem 1:** Let \( S_1 := \sum_{j=1}^{k} \Omega_x B^{-1}_j \Omega_x \) and \( S_2 := \sum_{j=1}^{k} \Omega_x B^{-1}_j \Omega_x y_j \). From (3.8),

\[
\ell_n = \sum_{j=1}^{k} (\Omega_{x,j} y_j - S_2 S_1^{-1} \Omega_{x,j}) B^{-1}_j (\Omega_{x,j} y_j - \Omega_{x,j} S_1^{-1} S_2) + o_p(1)
\]

\[
= \sum_{j=1}^{k} (\Omega_{x,j} y_j - S_2 S_1^{-1} \Omega_{x,j}) B^{-1}_j \Omega_{x,j} B^{-1}_j S_1^{-1} (S_1 \Omega^{-1}_x \Omega_x y_j - S_2) + o_p(1).
\]

It can be shown that

\[
\ell_n = \begin{pmatrix}
\Omega^{-1}_x x_1 y_1 - \Omega^{-1}_x x_2 y_2 \\
\Omega^{-1}_x x_1 y_1 - \Omega^{-1}_x x_3 y_3 \\
\vdots \\
\Omega^{-1}_x x_1 y_1 - \Omega^{-1}_x x_k y_k
\end{pmatrix}^\tau \Sigma_0 \begin{pmatrix}
\Omega^{-1}_x x_1 y_1 - \Omega^{-1}_x x_2 y_2 \\
\Omega^{-1}_x x_1 y_1 - \Omega^{-1}_x x_3 y_3 \\
\vdots \\
\Omega^{-1}_x x_1 y_1 - \Omega^{-1}_x x_k y_k
\end{pmatrix} + o_p(1),
\]

where \( \Sigma_0 \) is a \((k-1)p \times (k-1)p\) matrix with \((j-1)\)-th \((j = 2, \ldots, k)\) diagonal matrix component as \((\Omega_x B^{-1}_j \Omega_x) - (\Omega_x B^{-1}_j \Omega_x) S_1^{-1} (\Omega_x B^{-1}_j \Omega_x)\) and \((p-1, q-1)\)-th \((p,q = 2, \ldots, k)\) matrix component is \(\Omega_x B^{-1}_p \Omega_x y_q - S_1^{-1} (\Omega_x B^{-1}_p \Omega_x y_q)\).

To make the derivation easily presentable, we only present the detail proof for \( k = 3 \), as the general case can be done similarly except more tedious. From (A.4), we have

\[
\ell_n = \begin{pmatrix}
\Omega^{-1}_x x_1 y_1 - \Omega^{-1}_x x_2 y_2 \\
\Omega^{-1}_x x_1 y_1 - \Omega^{-1}_x x_3 y_3 \\
\vdots \\
\Omega^{-1}_x x_1 y_1 - \Omega^{-1}_x x_k y_k
\end{pmatrix}^\tau \Sigma_0 \begin{pmatrix}
\Omega^{-1}_x x_1 y_1 - \Omega^{-1}_x x_2 y_2 \\
\Omega^{-1}_x x_1 y_1 - \Omega^{-1}_x x_3 y_3 \\
\vdots \\
\Omega^{-1}_x x_1 y_1 - \Omega^{-1}_x x_k y_k
\end{pmatrix} + o_p(1),
\]
where $\Sigma_0 = \begin{pmatrix} A & C \\ C^T & B \end{pmatrix}$,

$$A = V_{121} + V_{321} + V_{323} + V_{123} + V_{232} + V_{212},$$

$$= (\Omega_{x_2} B_2^{-1} \Omega_{x_2}) S_1^{-1} (\Omega_{x_1} B_1^{-1} \Omega_{x_1}) - V_{221} + (\Omega_{x_2} B_2^{-1} \Omega_{x_2}) S_1^{-1} (\Omega_{x_3} B_3^{-1} \Omega_{x_3}) - V_{223} + V_{212}$$

$$= (\Omega_{x_2} B_2^{-1} \Omega_{x_2}) S_1^{-1} (\Omega_{x_1} B_1^{-1} \Omega_{x_1}) + (\Omega_{x_2} B_2^{-1} \Omega_{x_2}) S_1^{-1} (\Omega_{x_3} B_3^{-1} \Omega_{x_3})$$

$$+ (\Omega_{x_2} B_2^{-1} \Omega_{x_2}) S_1^{-1} (S_1 - \Omega_{x_2} B_2^{-1} \Omega_{x_2}) S_1^{-1} (\Omega_{x_2} B_2^{-1} \Omega_{x_2})$$

$$- (\Omega_{x_2} B_2^{-1} \Omega_{x_2}) S_1^{-1} (\Omega_{x_2} B_2^{-1} \Omega_{x_2}) S_1^{-1} (S_1 - \Omega_{x_2} B_2^{-1} \Omega_{x_2})$$

$$= (\Omega_{x_2} B_2^{-1} \Omega_{x_2}) S_1^{-1} (\Omega_{x_1} B_1^{-1} \Omega_{x_1}) + (\Omega_{x_2} B_2^{-1} \Omega_{x_2}) S_1^{-1} (\Omega_{x_3} B_3^{-1} \Omega_{x_3})$$

$$+ (\Omega_{x_2} B_2^{-1} \Omega_{x_2}) S_1^{-1} (\Omega_{x_2} B_2^{-1} \Omega_{x_2}) - (\Omega_{x_2} B_2^{-1} \Omega_{x_2}) S_1^{-1} (\Omega_{x_2} B_2^{-1} \Omega_{x_2}) - V_{222} + V_{222}$$

$$= (\Omega_{x_2} B_2^{-1} \Omega_{x_2}) - (\Omega_{x_2} B_2^{-1} \Omega_{x_2}) S_1^{-1} (\Omega_{x_2} B_2^{-1} \Omega_{x_2}).$$

$$B = V_{313} + V_{323} + V_{131} + V_{232} + V_{123} + V_{231}$$

$$= \Omega_{x_3} B_3^{-1} \Omega_{x_3} - (\Omega_{x_3} B_3^{-1} \Omega_{x_3}) S_1^{-1} (\Omega_{x_3} B_3^{-1} \Omega_{x_3})$$

$$C = V_{213} - V_{323} - V_{123} - V_{232} - V_{231}$$

$$= -(\Omega_{x_2} B_2^{-1} \Omega_{x_2}) S_1^{-1} (\Omega_{x_3} B_3^{-1} \Omega_{x_3}).$$

From the proof of Lemma 2, we know that

$$\Sigma_1 = \text{Var} \left( \begin{pmatrix} \Omega_{x_1}^{-1} \Omega_{x_1 y_1} & \Omega_{x_2}^{-1} \Omega_{x_2 y_2} \\ \Omega_{x_1}^{-1} \Omega_{x_1 y_1} & \Omega_{x_3}^{-1} \Omega_{x_3 y_3} \end{pmatrix} \right)$$

$$= \begin{pmatrix} \Omega_{x_1}^{-1} B_1 \Omega_{x_1}^{-1} + \Omega_{x_2}^{-1} B_2 \Omega_{x_2}^{-1} \\ \Omega_{x_1}^{-1} B_1 \Omega_{x_1}^{-1} + \Omega_{x_3}^{-1} B_3 \Omega_{x_3}^{-1} \end{pmatrix} \Omega_{x_1}^{-1} B_1 \Omega_{x_1}^{-1} + \Omega_{x_3}^{-1} B_3 \Omega_{x_3}^{-1}.$$

As $\Sigma_0 = \Sigma_1^{-1}$, from (A.5) $\ell_n \overset{d}{\to} \chi^2_{2p}$. This completes the proof.

**Proof of Theorem 2:** We note that $\Sigma_D^{-1/2} D \overset{d}{\to} N_{(k-1)p}(\gamma, I_{(k-1)p})$, where $D$ and $\Sigma_D$ are defined before Theorem 2 and (3.10) respectively, and $\gamma = \Sigma_D^{-1/2} D$. From (A.5), $\ell_n = D^T \Sigma_D^{-1} D + o_p(1)$, therefore $\ell_n \to \chi^2_{(k-1)p}(\gamma^2)$, which completes the proof of the theorem.

Next we give the proof of $\sup_{t \in [0,1]} |\eta_j(t)| = O_p\{n_j h_j^{-1/2} \log n_j\}$, which requires the following corollary from Lemma 1.

**Corollary 1.** Suppose $e_1, \ldots, e_n$ are independent random variables with $E(e_i) = 0$, $(i = 1, \ldots, n)$ and $\max_{1 \leq i \leq n} E(|e_i|^\delta) < \infty$ for some $\delta > 2$. Let $G_i(t)$ be smooth functions of $t$ on $[0,1]$ and is Lipschitz continuous of order $1$, $\max_{1 \leq i \leq n} \sup_{t \in S} |G_i(t)| = O(d_n)$ and $\sup_{t \in S} \sum_{i=1}^n |G_i(t)| < \infty$ as $n \to \infty$, then

$$\sup_{t \in S} \left| \sum_{i=1}^n G_i(t) e_i \right| = O\{\max(n^{1/3}d_n, n^{1/2}) \log n\} \quad a.s.$$
and for some $j$,
\[
\sup_{t \in S} \left| \sum_{i=1}^{n} G_i(t)e_i \right| = \sup_{t \in S} \left| \sum_{i=1}^{n} G_i(t)e_i - \sum_{i=1}^{n} G_i(t_j)e_i + \sum_{i=1}^{n} G_i(t_j)e_i \right| 
\leq C \left| n^{-1} \tau_n \sum_{i=1}^{n} e_i \right| + \max_{1 \leq j \leq n} \left| \sum_{i=1}^{n} G_i(t_j)e_i \right|.
\]

By Lemma 1, we know $\max_{1 \leq j \leq n} \left| \sum_{i=1}^{n} G_i(t_j)e_i \right| = O(\tau_n)$ a.s. We also need to show $|n^{-1} \tau_n \sum_{i=1}^{n} e_i|$ is $O(\tau_n)$ a.s. Let $e_i = e_i I\{|e_i| \leq i^{1/\delta}\}$. Then
\[
|n^{-1} \tau_n \sum_{i=1}^{n} e_i| \leq |n^{-1} \tau_n \sum_{i=1}^{n} (e_i - e_i')| + |n^{-1} \tau_n \sum_{i=1}^{n} (e_i' - E(e_i'))| + |n^{-1} \tau_n \sum_{i=1}^{n} E(e_i' - e_i)|
\leq K_1 + K_2 + K_3, \quad \text{say}
\]

For $K_1$, since $\max_{1 \leq i \leq n} E(|e_i|^{1/\delta}) < \infty$, $\sum_{i=1}^{\infty} P(|e_i| > i^{1/\delta}) < \infty$. According to Borel-Cantelli lemma, $\sum_{i=1}^{n} e_i I\{|e_i| > i^{1/\delta}\} < \infty$ a.s. Thus, $K_1 = o(\tau_n)$ a.s.. For $K_3$, we note that, for $\delta > 2$,
\[
|n^{-1} \tau_n \sum_{i=1}^{n} E(e_i' - e_i)| \leq n^{-1} \tau_n \max_{1 \leq t \leq n} E|e_i|^{1/\delta} \sum_{i=1}^{n} E(i^{-(\delta-1)/\delta} I\{|e_i| > i\})
\leq n^{-1} \tau_n \sum_{i=1}^{n} i^{-(\delta-1)/\delta} \leq \tau_n \sum_{i=1}^{n} i^{-(2\delta-1)/\delta} = O(\tau_n).
\]

Let $M_n = 2n^{1/\delta}$. Then $P\{|e_i' - E(e_i')| \leq 2M_n\} = 1$ for each $i \leq n$. Applying Bernstein inequality, we get
\[
P\left[ n^{-1} \tau_n \sum_{i=1}^{n} (e_i' - E(e_i')) > \tau_n \right] \leq 2 \exp \left\{ - \frac{n^2}{8C^2 n^{(\delta+2)/\delta} + \frac{4C}{\delta} n^{(\delta+1)/\delta}} \right\} \leq 2 \exp(-C_1 n^{1-\frac{2}{\delta}}),
\]
where $C_1 = \frac{3}{2C^2+4C} < \infty$. As $\sum_{i=1}^{\infty} \exp(-C_1 n^{1-\frac{2}{\delta}}) < \infty$ and apply Borel-Cantelli lemma, we have $K_2 = o(\tau_n)$ a.s.. In summary, $|n^{-1} \tau_n \sum_{i=1}^{n} e_i| = O(\tau_n)$ a.s.. This completes the proof.

**Lemma 4.** Under Assumptions A1, A2, A3 and A4, and suppose $h_j = O(n^{-1/5})$, then
\[
\sup_{t_0 \in [0,1]} |\eta_j(t)| = O_p\{(n_j h_j)^{-1/2} \log n_j\}.
\]

**Proof.** Generalizing Owen(1990), we need to show
\[
(A.6) \quad \sup_{t \in [0,1]} \left| \frac{1}{n_j h_j T} \sum_{i=1}^{n_j} R_{ji}(g_{j0}(t)) \right| = O_p((n_j h_j)^{-1/2} \log n_j),
\]
\[
(A.7) \quad \max_{1 \leq i \leq n_j} \sup_{t \in [0,1]} |R_{ji}(g_{j0}(t))| = O_p((n_j h_j)^{1/2} \log^{-1} n_j),
\]
\[
(A.8) \quad P\left\{ \frac{1}{n_j h_j T} \sum_{i=1}^{n_j} R_{ji}^2(g_{j0}(t)) \geq d_0 \right\} = 1 \quad \text{for a positive} \quad d_0 > 0.
\]
Thus (plugging in the leading term of $L$ we know $\beta$ (2003), which we omit the details here.

Under the local alternative $G_n \doteq \theta_j - \beta_j$: Let $\theta_j := \theta_j \{n_j \to \infty\}$ and $\beta_j := \beta_j \{n_j \to \infty\}$.

Then for all $i, m$, $\max_{t \in S} |(n_j h_j T_j)^{-\frac{1}{2}} S_{j,i,m}| = O_p\{\sqrt{h_j}\}$. Applying Corollary 1, $\sup_{t \in [0,1]} |(n_j h_j T_j)^{-\frac{1}{2}} S_{j1}(t)| = O_p\{\log n_j\}$.

Thus (A.6) is proved. For (A.7) and (A.8), the proofs are similar to Lemma 1 in Chen et al. (2003), which we omit the details here.

**Proof of Theorem 3:** Let $v_j(t, h_j) = \sum_{i=1}^{n_j} R_{ji}^2 \{g(t)\}$ and $d_j(t, h_j) = \sum_{i=1}^{n_j} \sum_{m=1}^{T_j} \frac{\delta_{jim}}{\pi_{jim}(\theta_j)} K \left( \frac{t_{jim} - t}{h_j} \right)$.

To simplify notation, we sometimes hide the arguments of $v_j(t, h_j)$ and $d_j(t, h_j)$. After plugging in the leading term of $L_n(t)$ into $T_n$, we have the leading term of $T_n$, which is $\int \sum_{j=1}^k v_j^{-1} \left[ \sum_{i=1}^{n_j} R_{ji} \{0\} - d_j \left( \sum_{s=1}^k v_s^{-1} d_s^2 \right)^{-1} \sum_{s=1}^k v_s^{-1} d_s \sum_{i=1}^{n_s} R_{si} \{0\} \right]^2 \varpi(t) dt$.

Under the local alternative $g_{s0}(t) = g_{s1}(t) + C_{ns} \Delta_{ns}(t)$ for $s = 2, \ldots, k$, the test statistic $T_n$ can be written as $T_n = \int_0^1 \sum_{j=1}^k v_j^{-1} \left\{ B_{n}^2(t) + A_n^2(t) + 2 A_n(t) B_n(t) \right\} \varpi(t) dt + o_p(1)$ (A.9) := T_{n1} + T_{n2} + T_{n3} + o_p(1),
where $A_n(t) = d_j \left\{ C_{nj} \Delta_{nj}(t) - \left( \sum_{s=1}^{k} v_s^{-1} \sigma_s^2 \right)^{-1} \sum_{s=1}^{k} v_s^{-1} \sigma_s^2 C_{ns} \Delta_{ns}(t) \right\}$ and

$$B_n(t) = \sum_{i=1}^{n_j} R_{ji} \{ g_{j0}(t) \} - d_j \left( \sum_{s=1}^{k} v_s^{-1} \sigma_s^2 \right)^{-1} \sum_{s=1}^{k} v_s^{-1} \sigma_s^2 \sum_{i=1}^{n_i} R_{si} \{ g_{s0}(t) \}.$$  

Define $\sigma_{\epsilon_j}^2 = \frac{1}{n_j h_j T_j} \sum_{i=1}^{n_j} \sum_{m=1}^{T_j} E \left\{ \frac{\epsilon_j^2}{\sigma_j(t)} \right\}$, $R(K) = \int K^2(t) dt$ and $V_j(t) = R(K) \sigma_{\epsilon_j}^2 f_j(t)$. We first show that $(n_j T_j h_j)^{-1} v_j(t, h_j) \overset{p}{\to} V_j(t)$. According to the definition of $v_j(t, h_j)$, $R_{ji} \{ g(t) \}$ and $g(t) = g_{j0}(t) + O((n_j h_j)^{-1/2})$, we get

$$\frac{1}{n_j h_j T_j} \sum_{i=1}^{n_j} R_{ji}^2 \{ g(t) \} = \frac{1}{n_j h_j T_j} \sum_{i=1}^{n_j} \left( \sum_{m=1}^{T_j} K \left( \frac{t_{jim} - t}{h_j} \right) \nu_{jim} \right)^2$$

$$+ \frac{1}{n_j h_j T_j} \sum_{i=1}^{n_j} \left( \sum_{m=1}^{T_j} K \left( \frac{t_{jim} - t}{h_j} \right) \nu_{jim} \right)^2$$

$$+ \frac{1}{n_j h_j T_j} \sum_{i=1}^{n_j} \left( \sum_{m=1}^{T_j} K \left( \frac{t_{jim} - t}{h_j} \right) \nu_{jim} X_{jim} (\beta_{j0} - \beta_j) \right)^2 + o_p(1)$$

$$:= A_1(t) + A_2(t) + A_3(t) + o_p(1).$$

It is easy to see that $A_3(t) = O_p(n_j^{-1})$, since $\beta_{j0} - \beta_j = O_p(n^{-1/2})$. For $A_2(t)$, we note that the kernel $K(t)$ has support on $[-1, 1]$ and is Lipschitz continuous from assumption A2(i). Then a Taylor expansion yields

$$A_2(t) = \frac{1}{n_j h_j T_j} \sum_{i=1}^{n_j} \nu_{jim} K \left( \frac{t_{jim} - t}{h_j} \right) (g_{j0}(t) - \hat{g}_j(t))(t - t_{jim}) + O_p(h_j^2)$$

$$\leq \frac{1}{n_j h_j T_j} \sum_{i=1}^{n_j} \nu_{jim} K \left( \frac{t_{jim} - t}{h_j} \right)^2 |g_{j0}(t) - \hat{g}_j(t)|^2 h_j^2 + o_p(h_j^2) = o_p(h_j^2),$$

since $g_{j0}(t) - \hat{g}_j(t) = o_p(1)$. Note that $A_1(t)$ can be written as

$$A_1(t) = \frac{1}{n_j h_j T_j} \sum_{i=1}^{n_j} \nu_{jim} K \left( \frac{t_{jim} - t}{h_j} \right) K \left( \frac{t_{jim1} - t}{h_j} \right) \nu_{jim1} \nu_{jim} \nu_{jim1} \nu_{jim}$$

$$= \frac{1}{n_j h_j T_j} \sum_{i=1}^{n_j} \nu_{jim} K \left( \frac{t_{jim} - t}{h_j} \right) K \left( \frac{t_{jim1}}{h_j} \right) \nu_{jim} \nu_{jim1}$$

$$+ \frac{1}{n_j h_j T_j} \sum_{i=1}^{n_j} \nu_{jim} K \left( \frac{t_{jim} - t}{h_j} \right) K \left( \frac{t_{jim1} - t}{h_j} \right) \nu_{jim} \nu_{jim1}$$

$$:= A_{11}(t) + A_{12}(t).$$

Then

$$E \left\{ A_{12}(t) \right\} = \frac{T_j - 1}{h_j} \int_0^1 \int_0^1 K \left( \frac{x - t}{h_j} \right) K \left( \frac{y - t}{h_j} \right) \rho_j(x, y) \epsilon_j(x) \epsilon_j(y) f_j(x) f_j(y) dx dy$$

$$= h_j (T_j - 1) \sigma_{\epsilon_j}^2 f_j^2(t) \{ 1 + o(1) \} = O(h_j),$$
which is the case since $T_j$ is finite. Note here, when $m \neq m_1$, $A_{12}(t)$ is similar to the kernel estimator for a bivariate function. Whereas in two dimensional kernel estimator are divided by $n_j T_j h_j^2$. However, the denominator is $n_j T_j h_j$ in $A_{12}(t)$, so this term become smaller order term relative to $A_{11}(t)$. From assumptions A2(ii) and A3, we know $A_{11}(t) \overset{P}{\rightarrow} V_j(t)$.

Let us first consider the first term in $T_n$ given in (A.9),

$$T_{n1} = \int_0^1 \sum_{j=1}^k v_j^{-1} B_n^2(t) \varpi(t) dt$$

$$= \int_0^1 \sum_{j=1}^k \left( 1 - \left[ \sum_{s=1}^k v_s^{-1} d_s^2 \right]^{v_j^{-1} d_j^2} \left[ \sum_{i=1}^{n_j} R_{ji}(g_{j0}(t)) \right] \right)^2 \varpi(t) dt$$

$$- \int_0^1 \sum_{j=1}^k \left[ \sum_{s=1}^k v_s^{-1} d_s^2 \right]^{-v_j^{-1} d_j^2} \left[ \sum_{i=1}^{n_j} R_{ji}(g_{j0}(t)) \right] \left[ \sum_{i=1}^{n_j} R_{ji}(g_{j0}(t)) \right] \varpi(t) dt$$

(A.10) \quad := T_{n1}^{(1)} - T_{n1}^{(2)} \quad , \quad \text{say},

Let

$$S_{j1}^2(t) = \frac{1}{n_j h_j T_j} \sum_{i=1}^{n_j} \sum_{m=1}^{T_j} \sum_{j=1}^{n_j} \delta_{jim} \delta_{jim} \varepsilon_{jim} \varepsilon_{jim} \left( \frac{t_{jim} - t_{jim}}{h_j} \right) K \left( \frac{t_{jim} - t_{jim}}{h_j} \right)$$

and

$$S_{j2}^2(t) = \frac{2}{n_j h_j T_j} \sum_{i=1}^{n_j} \sum_{m=1}^{T_j} \delta_{jim} \delta_{jim} \varepsilon_{jim} \varepsilon_{jim} \left( \frac{t_{jim} - t_{jim}}{h_j} \right) K \left( \frac{t_{jim} - t_{jim}}{h_j} \right)$$

We observe that

$$T_{n1}^{(1)} = \int_0^1 \sum_{j=1}^k \{1 - W_j(t)\} V_j^{-1}(t) S_{j1}^2(t) \varpi(t) dt$$

$$+ \int_0^1 \sum_{j=1}^k \{1 - W_j(t)\} V_j^{-1}(t) S_{j2}^2(t) \varpi(t) dt \{1 + o_p(1)\}$$

$$:= (T_{n1}^{(11)} + T_{n1}^{(12)}) \{1 + o_p(1)\}$$

Since $E\{S_{j1}^2(t)\} = V_j(t) + O(h), E(T_{n1}^{(1)}) = \sum_{j=1}^k \int_0^1 \{1 - W_j(t)\} \varpi(t) dt = k - 1 + O(h)$ and the variance of $T_{n1}^{(11)}$ is $O(n^{-1/2}) = o(h)$, under the condition $h = O(n^{-1/5})$. Thus

(A.11) \quad $h^{-1/2} \{T_{n1}^{(11)} - (k - 1)\} \overset{P}{\rightarrow} 0$.

Define $\xi_{ji}(t) = \frac{1}{\sqrt{h_j T_j}} \sum_{m=1}^{T_j} K \left( \frac{t_{jim} - t_{jim}}{h_j} \right) \varepsilon_{jim}$, then we have

(A.12) \quad $T_{n1}^{(12)} = \sum_{j=1}^k \sum_{j=1}^{n_j} v_j^{-1} \int_0^1 \{1 - W_j(t)\} V_j^{-1}(t) \xi_{ji}(t) \xi_{ji}(t) \varpi(t) dt + o_p(h^{1/2})$.

and

(A.13) \quad $T_{n1}^{(2)} = \sum_{j=1}^{n_j} \sum_{i=1}^{n_j} \sum_{i=1}^{n_j} (n_j n_{j1})^{-1/2} \int_0^1 \left( \frac{W_j(t) V_j(t)}{V_j(t) V_j(t)} \right)^{1/2} \xi_{ji}(t) \xi_{ji}(t) \varpi(t) dt + o_p(h^{1/2})$. 
Let $N = \sum_{j=1}^{k} n_j$. We stack $\xi_{ji}$ $(j = 1, \ldots, k, i = 1, \ldots, n_j)$ to form a sequence $\phi_s$, $s = 1, \ldots, N$. Let $G_j$ be the collection of the subscripts of $\phi$ whose corresponding $\xi$ are in Treatment $j$. Define
\begin{equation}
C_{ps}(t) = \frac{1}{n(p, s)} \left\{ \sum_{j=1}^{k} I(p \in G_j, s \in G_j) V_j^{-1}(t) - \sum_{j=1}^{k} \sum_{l=1}^{k} \frac{W_j(t)W_l(t)}{V_j(t)V_l(t)} I(p \in G_j, s \in G_l) \right\},
\end{equation}

where $n(p, s) = \sum_{j=1}^{k} \sum_{l=1}^{k} (n_jn_l)^{1/2} I(p \in G_j, s \in G_l)$ and $I(p \in G_j, s \in G_l)$ is the usual indicator function. Using these notations, we may write
\begin{equation}
U_N := T_{n_1}^{(12)} - T_{n_1}^{(2)} = 2 \sum_{p=1}^{N} \sum_{s<p} \psi(\phi_p, \phi_s),
\end{equation}

where $\psi(\phi_p, \phi_s) = \int_{0}^{1} C_{ps}(t)\phi_p(t)\phi_s(t)\varphi(t)dt$. Then (A.15) is a quadratic form with kernel $\psi(\phi_p, \phi_s)$. Let $\sigma_{ps}^2 = \text{Var}(\psi(\phi_p, \phi_s))$. Using results for generalized quadratic form with independent but not identically distributed random variables (de Jong, 1987) if
\begin{equation}
\{\text{Var}(U_N)\}^{-1} \max_{1 \leq p \leq N} \sum_{s=1}^{N} \sigma_{ps}^2 \to 0 \quad \text{and}
\end{equation}
\begin{equation}
\{\text{Var}(U_N)\}^{-2} EU_N^4 \to 3,
\end{equation}

then (A.15) is asymptotically normally distributed with mean 0 and variance
\begin{equation}
\text{Var}(U_N) = \text{Var}(T_{n_1}^{(12)}) + \text{Var}(T_{n_1}^{(2)}) - 2\text{Cov}(T_{n_1}^{(12)}, T_{n_1}^{(2)}).
\end{equation}

Let us first derive Var($U_N$). We note that $\text{Var}(T_{n_1}^{(12)}) = \sum_{j=1}^{k} \sum_{m_1} \sigma_{1,jm_1}^2$ where
\begin{align*}
\sigma_{1,jm_1}^2 &= E_i E_{i1} \left\{ \int_{0}^{1} \int_{0}^{1} \left\{ 1 - W_j(t) \right\} \left\{ 1 - W_j(u) \right\} \frac{\xi_{ji}(t)\xi_{ji}(t)}{V_j(t)V_j(u)} \xi_{ji}(u)\xi_{ji}(u) \varphi(t)\varphi(u) dt du \right\} \\
&= \frac{1}{T_j^2} \sum_{m_1} \frac{1 - W_j(t_{jm1})}{V_j^2(t_{jm1})} \sigma_{ejj_1}^2(t_{jm1}) \sigma_{ejj_1}^2(t_{jm1}) \varphi^2(t_{jm1}) \\
&\quad \times \left(K^{(2)}(t_{jm1} - t_{jm1}) \right)^2 \left\{ 1 + o(1) \right\},
\end{align*}

where $\sigma_{ejj_1}^2(t_{jm1}) = E_i\{z_{jm1}^2/\pi_{jm1}(\theta_{j0})\}$. Since $\{t_{jm1}\}$ are fixed design points generated from a density $f_j(t)$, via a Taylor expansion and by Assumption A2(ii),
\begin{equation}
\text{Var}(T_{n_1}^{(12)}) = 2hR(K)^{-2} K_1^{(4)}(0) \sum_{j=1}^{k} b_j^{-1} \int_{0}^{1} (1 - W_j(t))^2 \varphi^2(t) dt \{ 1 + o(1) \}.
\end{equation}

Similar to our derivation for the variance of $T_{n_1}^{(12)}$, it may be shown that
\begin{equation}
\text{Var}(T_{n_1}^{(2)}) = 2hR(K)^{-2} \sum_{j \neq j_1} K_{b_j/b_{j_1}}^{(4)}(0) (b_j b_{j_1})^{-1/2} \int_{0}^{1} W_j(t) W_{j_1}(t) \varphi^2(t) dt \{ 1 + o(1) \}.
\end{equation}
From (A.18), we also need to calculate the covariance between $\mathcal{T}_{n_1}^{(12)}$ and $\mathcal{T}_{n_1}^{(2)}$. Using the same method for calculating variance for $\mathcal{T}_{n_1}^{(12)}$ and $\mathcal{T}_{n_1}^{(2)}$, we may show that

$$\text{(A.21)} \quad \text{Cov}(\mathcal{T}_{n_1}^{(12)}, \mathcal{T}_{n_1}^{(2)}) = O(h^2),$$

In summary of (A.19), (A.20) and (A.21),

$$\text{(A.22)} \quad \text{Var}(U_N) := h\sigma_0^2 = 2hK^{(2)}(0)^{-2} \int_0^1 \Lambda(t) \varpi^2(t) dt \{1 + o(1)\},$$

where $\Lambda(t)$ is defined just before Theorem 3.

Next we need to establish the conditions (A.16) and (A.17). For (A.16), we have

\[
\{\text{Var}(U_N)\}^{-1} \max_{1 \leq p < N} \sum_{s=1}^N \sigma_{ps}^2 = (h\sigma_0^2)^{-1} \max_{1 \leq i \leq n_j} \left\{ \frac{1}{n_j} \sum_{i=1}^{n_j} \sigma_{1,i;i_1}^2 + \sum_{j_1=1}^k \frac{1}{n_j n_{j_1}} \sum_{i=1}^{n_{j_1}} \sigma_{2,j_1,i_1}^2 \right\} 
\leq \left( h\sigma_0^2 \right)^{-1} \left[ \max_{1 \leq j_1 \leq j} \left\{ \frac{1}{n_j} \sum_{i=1}^{n_j} \sigma_{1,j;i_1}^2 \right\} + \max_{1 \leq j \leq n_j} \left\{ \sum_{j_1=1}^k \frac{1}{n_j n_{j_1}} \sum_{i=1}^{n_{j_1}} \sigma_{2,j_1;i_1}^2 \right\} \right].
\]

From conditions (A2) and (A3),

\[
\max_{1 \leq j \leq n_j} \frac{1}{n_j} \sum_{i=1}^{n_j} \sigma_{1,j;i_1}^2 = \max_{1 \leq j \leq n_j} \frac{1}{n_j T_j} \sum_{m} \left\{ \frac{1 - W_j(t_{jim})}{V_j^2(t_{jim})} \sigma_{\varepsilon_1,j_1}(t_{jim}) \varpi^2(t_{jim}) \right\} 
\times \left\{ \frac{1}{n_j T_j} \sum_{i=1}^{n_j} \sum_{m=1}^{T_j} \sigma_{\varepsilon_1,j_1}(t_{jim}) \left( K^{(2)}(t_{jim} - t_{jim}) \right)^2 \right\} 
\leq \max_{1 \leq j \leq n_j} \left\{ \frac{1}{n_j T_j} \sum_{m} \left\{ \frac{1 - W_j(t_{jim})}{V_j^2(t_{jim})} \sigma_{\varepsilon_1,j_1}(t_{jim}) \sigma_{\varepsilon_1,j_1}(t_{jim}) \varpi^2(t_{jim}) \right\} 
\times \left\{ R(K)^{-2} K_1^{(4)}(0) \right\} \right\} h_j = O(n^{-1}h).
\]

And similarly, \[ \max_{1 \leq j \leq n_j} \left\{ \sum_{j_1=1}^k \frac{1}{n_j n_{j_1}} \sum_{i=1}^{n_{j_1}} \sigma_{2,j_1;i_1}^2 \right\} = O(n^{-1}h). \] These imply (A.16).

It is remain to check (A.17). By (A.15), we have

\[
E(U_N^4) = E(T_{n_1}^{(12)}T_{n_1}^{(2)})^4 - 4E(T_{n_1}^{(12)})^4 T_{n_1}^{(2)} + 6E(T_{n_1}^{(12)})^2 (T_{n_1}^{(2)})^2 + 4E(T_{n_1}^{(12)})^2 (T_{n_1}^{(2)})^3 + E(T_{n_1}^{(2)})^4.
\]

(A.23)
Therefore, (A.24) is $O(n^{-2})$, hence is negligible; and the second term on the right hand side converges to $3\{\text{Var}(T_{n1}^{(12)})\}^2$. Similarly, we can show that $E(T_{n1}^{(12)})^4 \to 3\{\text{Var}(T_{n1}^{(12)})\}^2$ and $6E\{(T_{n1}^{(12)})^2(T_{n1}^{(22)})^2\} \to 6\text{Var}(T_{n1}^{(12)})\text{Var}(T_{n1}^{(22)})$. From (A.23),

$$
\lim_{n \to \infty} \{\text{Var}(U_N)\}^{-2}E(U_N^4) = \lim_{n \to \infty} 3\{\text{Var}(U_N)\}^{-2}\{\text{Var}(T_{n1}^{(12)}) + \text{Var}(T_{n1}^{(22)})\}^2 = 3.
$$

Therefore, (A.17) is verified and then we have the asymptotic normality of $U_N$.

In summary of (A.11), (A.15) and (A.22),

(A.25) $$h^{-1/2}\{T_{n1} - (k - 1)\} \xrightarrow{d} N(0, \sigma_0^2).$$

Let us consider $T_{n2} = \int_0^1 \sum_{j=1}^k v_j^{-1} A_n^2(t)\varpi(t)dt$. Recall the definition of $A_n(t)$ in (A.9). From Assumption A2(iii) that there exist finite number $a_j$ and $b_j$ such that $C_{jn} = a_j^{-1/2}b_j^{-1/4}(nT)^{-1/2}h^{-1/4} n_j T_j h_j = (a_jb_j)^{-1}nT$. Thus, $h^{-1/2}(T_{n2} - \mu_1) = \omega_p(1)$ and where

$$
\mu_1 = \int_0^1 \left[ \sum_{j=1}^k b_j^{-1} V_j^{-1}(t) f_j^2(t)\Delta_{n_j}^2(t) - \left( \sum_{s=1}^k b_s^{-1} V_s^{-1}(t) f_s^2(t)\Delta_{n_s}^2(t) \right)^2 \right] \varpi(t)dt.
$$

It remains to consider $T_{n3} = 2 \int_0^1 \sum_{j=1}^k v_j^{-1} A_n(t)B_n(t)\varpi(t)dt$. Using the expression of $A_n(t)$ and $B_n(t)$, we can decompose $T_{n3}$ as

$$
T_{n3} = 2 \int_0^1 \sum_{j=1}^k v_j^{-1} d_j C_{nj} \Delta_{n_j}(t) \sum_{i=1}^{n_j} R_{ji}\{g_{j0}(t)\} \varpi(t)dt
$$

$$
- 2 \int_0^1 \left( \sum_{j=1}^k v_j^{-1} d_j^2 \right)^{-1} \left( \sum_{j=1}^k v_j^{-1} d_j^2 C_{nj} \Delta_{n_j}(t) \right) \left( \sum_{s=1}^k v_s^{-1} d_s \sum_{i=1}^{n_s} R_{si}\{g_{s0}(t)\} \right) \varpi(t)dt
$$

$$
:= T_{n3}^{(1)} - T_{n3}^{(2)}, \quad \text{say}.
$$

We know that

$$
T_{n3}^{(1)} = 2 \int_0^1 \sum_{j=1}^k V_j(t)^{-1} f_j(t) C_{nj} \Delta_{n_j}(t) \sum_{i=1}^{n_j} R_{ji}\{g_{j0}(t)\} \varpi(t)dt = 2 \sum_{j=1}^k h_j^{3/4} T_{n3}^{(1)} \{1 + \omega_p(1)\}
$$
\[ T^{(1j)}_{n3} = (n_j T_j)^{-1/2} \sum_{i=1}^{n_j} \sum_{m=1}^{T_j} \nu_{jim} \varepsilon_{jim} \int_0^1 V_j(t)^{-1} f_j(t) \Delta_{nj}(t) K_h(t - t_{jim}) \pi(t) dt. \]

Thus we can say \( T^{(1)}_{n3} \) is \( O_p(h^{3/4}) \) if we can show \( T^{(1j)}_{n3} = O_p(1) \). It is sufficient to show that \( \text{Var}(T^{(1j)}_{n3}) = O(1) \). Indeed, after some algebra, we get
\[
\text{Var}(T^{(1j)}_{n3}) = R(K)^{-2} \int_0^1 \sigma_{ij}(y) \Delta_\pi^2(y) dy \int K^{(2)}(z) dz \{1 + o(1)\} = O(1).
\]

Therefore \( T^{(1)}_{n3} \) is \( O_p(h^{3/4}) \). The second term of \( T^{(2)}_{n3} \) can be written in a similar form as \( T^{(1)}_{n3} \), which is also \( O_p(h^{3/4}) \). Thus \( T_{n3} = O_p(h^{3/4}) \).

In summary of these and (A.25),
\[ h^{-1/2}(T_n - (k - 1) - \mu) \Rightarrow N(0, \sigma_0^2). \]
Thus the proof is completed.

**Proof of Theorem 4**: We want to establish the bootstrap version of Theorem 3. To avoid repetition, we only outline some important steps in proving this theorem.

We use \( v^*_j(t, h_j) \) and \( d^*(t, h_j) \) to denote the bootstrap counterparts of \( v_j(t, h_j) \) and \( d(t, h_j) \) respectively. Let \( O_p^*(1) \) and \( O_p^*(1) \) be the stochastic order with respect to the conditional probability measure given the original samples.

We want to show first that
\[
(A.26) \quad (n_j h_j T_j)^{-1} v^*_j(t, h_j) - V^*(t) = o_p^*(1), \quad \text{as} \quad n_j \to \infty.
\]

where \( V^*(t) = R(K) \sigma_2T_j f_j(t) \). This can be seen from the following decomposition,

\[
\frac{1}{n_j h_j T_j} \sum_{j=1}^{n_j} \sum_{i=1}^{T_j} R^2_{ji} \{ \hat{g}_i(t) \} = \frac{1}{n_j h_j T_j} \sum_{j=1}^{n_j} \sum_{m=1}^{T_j} \nu^*_{jim} K \left( t_{jm} - t \right) \hat{\varepsilon}_{jim}^2 + \frac{1}{n_j h_j T_j} \sum_{j=1}^{n_j} \sum_{m=1}^{T_j} \nu^*_{jim} K \left( t_{jm} - t \right) X_{jm}^2 \left( \beta_j - \hat{\beta}_j \right) + o_p^*(1) := A_1^* + A_2^* + A_3^* + o_p^*(1),
\]

where \( \nu^*_{jim} = \frac{\pi^*_{ijm}}{\pi_{ijm}(\theta_j)} = \frac{\pi^*_{ijm}}{\pi_{ijm}(\theta_j)} \left( 1 - \pi_{ijm}(\theta_j) - \pi_{ijm}(\theta_j) \right) \).

Then we can apply \( \pi_{ijm}(\theta_j) - \pi_{ijm}(\theta_j) = O_p(n^{-1/2}) \), \( \hat{\beta}_j - \beta_j = O_p(n^{-1/2}) \) and \( \hat{\beta}_j (t) - \beta_j (t) = o_p^*(1) \) to \( A_j + A_3 \). By the similar procedure as we derive expression for \( v_j(t, h_j) \) in the proof of Theorem 3, we can get (A.26).

Corresponding to the leading term of \( T_n \), the leading term of \( T_n^* \) is

\[
\int_0^1 \sum_{j=1}^k v_{s1}^{*1} \left[ \sum_{i=1}^{n_j} R^2_{ji} \{ \hat{g}_i(t) \} - d_j^{-1} \left( \sum_{s=1}^k R^2_{s1} \{ \hat{g}_s(t) \} \right) \right] \varpi(t) dt + \int_0^1 \sum_{j=1}^k \{1 - W_j^*(t)\} V_j^{*1} S_j^2(t) \varpi(t) dt + \int_0^1 \sum_{j=1}^k \{1 - W_j^*(t)\} V_j^{*1} S_j^2(t) \varpi(t) dt
\]

\[
:= B_1^* + B_2^*.
\]
where \( W_j^*(t) = \frac{f_j(t)/(\sigma_j^2 \rho_j^2)}{\sum_{l=1}^n f_l(t)/(\sigma_l^2 \rho_l^2)} \), \( S_j^2(t) \) and \( S_j^{2*}(t) \) are the bootstrap version of \( S_j^2(t) \) and \( S_j^{2*}(t) \) defined in the proof of Theorem 3. Then, using a similar approach to the one used in establishing the asymptotic normality of \( \mathcal{T}_{n1} \) in (A.10) in the proof of Theorem 3. We may show that

\[
H^{-1/2} \{ B_1^* - (k - 1) \} = o_p(1) \quad \text{and} \quad H^{-1/2} B_2^* \xrightarrow{d} N(0, \sigma_0^2) \quad \text{a.s.}
\]

Hence, Theorem 4 is established.

References.


ANOVA FOR LONGITUDINAL DATA


[34] RUBIN, B. (1976), Inference and Missing Data (with discussion), Biometrika, 63, 581-592.


