Varying-Coefficient Additive Models for Functional Data

XIAOKE ZHANG and JANE-LING WANG

Department of Statistics, University of California, Davis,
Davis, California 95616 U.S.A.
xkzhang@ucdavis.edu   janelwang@ucdavis.edu

October 30, 2013

Abstract

Both varying-coefficient and additive models have been studied extensively in the literature as extensions to linear models. They have also been extended to functional response data. However, existing extensions are still not sufficiently flexible to reflect the functional feature of the responses. In this paper, we extend both varying-coefficient and additive models to a much more flexible "varying-coefficient additive model" and propose a simple algorithm to estimate the nonparametric additive and varying-coefficient components of this model. We establish the L^2 rate of convergence for each component function and demonstrate the applicability of the new model through traffic data.

Key Words: B-splines, functional data, varying-coefficient models, additive structure.

1 INTRODUCTION

Varying-coefficient models are widely used in longitudinal data analysis due to their simplicity, flexibility and interpretability. They provide a natural extension of the linear regression model to a nonparametric setting and can easily incorporate multiple covariates. Let $W(t)$ be a smooth random response function, $\mathbf{Z} = (Z_1, \ldots, Z_d)^\top$ be a d-dimensional vector of covariates, and m be the regression function $m(t, \mathbf{z}) = E\{W(t) \mid \mathbf{Z} = \mathbf{z}\}$. The varying-coefficient model proposed by Hoover et al. (1998) takes the form:

$$m(t, \mathbf{z}) = \beta_0(t) + \sum_{k=1}^{d} \beta_k(t) z_k. \tag{1}$$

Model (1) assumes that at each time t, the relation between the covariate vector $\mathbf{Z}$ and the response $W(t)$ is linear but allows the regression coefficients to vary across time. However, the linear assumption may not hold or should at least be verified. This motivates us to consider a natural extension of model (1), the *varying-coefficient additive model*:

$$m(t, \mathbf{z}) = \beta_0(t) + \sum_{k=1}^{d} \beta_k(t)\phi_k(z_k), \tag{2}$$

where the relation between Z and W might be nonlinear and unknown. Here the $\beta_k, k = 1, \ldots, d$, are varying-coefficient functions and $\phi_k, k = 1, \ldots, d$, the additive component functions.

An appealing feature of this model is that at any fixed time point t model (2) is additive with respect to $\mathbf{z}$. Hence model (2) is also an extension of the conventional additive model (Stone (1985)), which avoids the curse of dimensionality through the additive structure. Model (2) is thus an extension of both the varying-coefficient and additive models. A key advantage of the model is the multiplicative form $\beta_k(\cdot)\phi_k(\cdot)$, which separates the joint influence of covariates and time and provides easy interpretations. It also facilitates model checking for either the varying-coefficient model or additive model.

Thanks to the dense design of the time points for functional data, we are able to separate the estimation of the additive component functions from that of the varying-coefficient functions. A two-step estimation procedure is proposed in Section 2. The proposed approach involves fitting an additive model for independent data and then the fitting of a varying-coefficient model. The errors induced in estimating the additive component functions ϕ_k complicate the asymptotic theory for the coefficient function as errors are induced on the covariates in the regression models and need to be handled carefully.

Algorithms to estimate component functions for additive models include ordinary backfitting (Buja et al. (1989)), marginal integration (Linton & Nielsen (1995)), smooth backfitting (Mammen et al. (1999)) and regression splines (Stone (1985), Wang & Yang (2007)). For varying-coefficient models, the unknown coefficient functions can be estimated by smoothing splines (Brumback & Rice (1998), Hoover et al. (1998)), local polynomial smoothing (Hoover et al. (1998), Fan & Zhang (1999)) and polynomial splines (Huang et al. (2002), Huang et al. (2004)). Due to computational efficiency and stability, we employed the B-spline approach to fit both the additive component and coefficient functions.

The rest of the paper proceeds as follows. The model and a two-step estimation approach are introduced in Section 2. Section 3 gives the asymptotic properties of the functional estimators. Section 4 and 5, respectively, present a simulation study and a real data analysis. Conclusions and discussions are in Section 6.

2 METHODOLOGY

The varying-coefficient additive model (2) requires some identifiability conditions. Assume for simplicity that the random function $W(t)$ from a subject is an L^2 stochastic process on the interval $[0, 1]$, and its covariate $\mathbf{Z} = (Z_1, \ldots, Z_d)^\top$, also has the property that $Z_k \in [0, 1]$, for all $k = 1, \ldots, d$.

PROPOSITION 1. *The varying-coefficient additive model (2) is identifiable under the constraints*

$$E\{\phi_k(Z_k)\} = 0 \quad \text{and} \quad \int_0^1 \beta_k(t)dt = 1, \quad k = 1, \ldots, d. \tag{3}$$

Then, $\beta_0(t) = E\{W(t)\}$, *the overall mean function.*

2.1 Estimation

Under the above identifiability conditions, an alternative form of model (2) is:

$$W(t \mid \mathbf{Z} = \mathbf{z}) = \beta_0(t) + \sum_{k=1}^{d} \beta_k(t)\phi_k(z_k) + U(t), \tag{4}$$

where U is the stochastic component of the process W, which is independent of $\mathbf{Z}$, and $E\{U(t)\} = 0$, $E\{\phi_k(Z_k)\} = 0$, and $\int \beta_k(t)dt = 1$ for $k = 1, \cdots, d$.

Due to the dense design at the time points of the functional data $W(t \mid \mathbf{z})$, we are able to separate the estimation of the additive component functions from that of the coefficient functions. To see this, integrate both sides of (4) with respect to t, we get

$$\int W(t \mid \mathbf{z})dt = \tilde{\beta}_0 + \sum_{k=1}^{d} \phi_k(z_k) + \tilde{U}, \tag{5}$$

where $\tilde{\beta}_0 = \int \beta_0(t)dt$ and $\tilde{U} = \int U(t)dt$.

This leads to an additive model in $\mathbf{Z}$, so the additive component functions ϕ_k can be estimated through standard additive model approaches for independent data. Once the ϕ_k are estimated, we can then employ any one of the several approaches for varying-coefficient models to estimate the β_k. We choose the B-spline smoother to estimate both ϕ_k and β_k and develop asymptotic theory for both estimators. A caveat is that the left-hand side of (5) involves integrating $\mathbf{W}$ over t but we do not observe the entire process $\mathbf{W}$, so there is an approximation error in the integration that needs to be handled properly. This will be addressed technically later in the proofs.

Suppose that we have a sample of n independent subjects, i.e., $(W_i, \mathbf{Z}_i)$'s are independently and identically distributed copies of $(W, \mathbf{Z})$. The process W_i is not observed, but instead, measurements are made densely at time points T_{ij}, $j = 1, \ldots, N_i$, for subject i, and the response at T_{ij} is contaminated with a random error e_{ij}. These errors, across time and subjects, are assumed to be independent and identical copies of e, where $E(e) = 0$ and $\text{var}(e) = \sigma^2$. Thus we observe $\{(Y_{ij}, T_{ij}, \mathbf{Z}_i), i = 1, \ldots, n; j = 1, \ldots, N_i\}$, where $Y_{ij} = W_i(T_{ij}) + e_{ij}$. This means the observations can be written as

$$Y_{ij} = \beta_0(T_{ij}) + \sum_{k=1}^{d} \beta_k(T_{ij})\phi_k(Z_{ik}) + U_{ij} + e_{ij}, \tag{6}$$

where $U_{ij} = U_i(T_{ij})$. For brevity of notation, hereafter we abbreviate $\delta_{ij} = U_{ij} + e_{ij}$.

Step I

Following (5), the first step is to construct an additive model with additive component functions ϕ_k. For this, we sort the data within each subject in ascending order of time and denote by $T^*_{i1} \leq T^*_{i2} \leq \cdots \leq T^*_{i,N_i}$ the ordered T_{ij} with Y^*_{ij} and δ^*_{ij} as concomitants, i.e. $Y^*_{ij} = Y_{ik}$ when $T_{ik} = T^*_{ij}$.

Applying the trapezoidal rule of integration, we define the integrated response for the ith subject by

$$\tilde{Y}_i = \frac{1}{2} \sum_{j=1}^{N_i - 1} (Y^*_{ij} + Y^*_{i,j+1})(T^*_{i,j+1} - T^*_{ij}) + Y^*_{i1} T^*_{i1} + Y^*_{i,N_i}(1 - T^*_{i,N_i}). \tag{7}$$

Then we fit an additive model on $\{(\tilde{Y}_i, \mathbf{Z}_i), i = 1, \ldots, n\}$ to estimate the additive component functions $\phi_k, k = 1, \ldots, d$, as follows:

$$\tilde{Y}_i = \tilde{\beta}_0 + \sum_{k=1}^{d} \phi_k(Z_{ik}) + \eta_i, \tag{8}$$

where η_i is defined in Eq. (16) in Appendix B.

To approximate ϕ_k by B-splines, we use basis functions of order $p_{k,A} \geq 1$ and $K_{k,A}$ interior knots, where the subscript A suggests that this is for the additive component function. Denote the basis functions by $\{\psi_{kl}, l = 1, \ldots, J_{k,A}\}$, where $J_{k,A} = K_{k,A} + p_{k,A}$ is the total number of bases to fit ϕ_k. For each $k = 1, \ldots, d$, we approximate $\phi_k(z) \approx \sum_{l=1}^{J_{k,A}} f_{kl}\psi_{kl}(z)$ by suitable spline coefficients $\mathbf{f}_k = \{f_{kl}, l = 1, \ldots, J_{k,A}\}^\top$. Let $\mathbf{f} = (\mathbf{f}_1, \ldots, \mathbf{f}_d)^\top$. The initial estimate for $\mathbf{f}$ is

$$(\breve{\beta}_0, \hat{\mathbf{f}}) = \underset{\tilde{\beta}_0, \mathbf{f}}{\operatorname{argmin}} \sum_{i=1}^{n} \left\{ \tilde{Y}_i - \tilde{\beta}_0 - \sum_{k=1}^{d} \sum_{l=1}^{J_{k,A}} f_{kl}\psi_{kl}(Z_{ik}) \right\}^2, \tag{9}$$

which is then modified to satisfy the empirical version of the constraints (3), i.e., $n^{-1} \sum_{i=1}^{n} \hat{\phi}_k(Z_{ik}) = 0$. This leads to

$$\hat{\phi}_k(z) = \sum_{l=1}^{J_{k,A}} \hat{f}_{kl}\psi_{kl}(z) - \frac{1}{n} \sum_{i=1}^{n} \sum_{l=1}^{J_{k,A}} \hat{f}_{kl}\psi_{kl}(Z_{ik}), \quad k = 1, \ldots, d. \tag{10}$$

Step II

Once we have obtained the estimates $\hat{\phi}_k, k = 1, \ldots, d$, we can estimate $\beta_k, k = 0, \ldots, d$, by plugging the estimates $\hat{\phi}_k$ into (6), and this leads to

$$Y_{ij} = \beta_0(T_{ij}) + \sum_{k=1}^{d} \beta_k(T_{ij})\hat{\phi}_k(Z_{ik}) + \epsilon_{ij}, \quad \text{where} \quad \epsilon_{ij} = \delta_{ij} - \sum_{k=1}^{d} \beta_k(T_{ij})\{\hat{\phi}_k(Z_{ik}) - \phi_k(Z_{ik})\}. \tag{11}$$

To estimate β_k, we use basis functions of order $p_{k,C} \geq 1$ and $K_{k,C}$ interior knots to approximate them, where the subscript C suggests that this is for the varying-coefficient function. Denote the basis functions by $\{\theta_{kl}, l = 1, \ldots, J_{k,C}\}$, where $J_{k,C} = K_{k,C} + p_{k,C}$. Then for each $k = 0, \ldots, d$, β_k is approximated by $\sum_{l=1}^{J_{k,C}} g_{kl}\theta_{kl}(t)$ with suitable spline coefficients $\mathbf{g}_k = \{g_{kl}, l = 1, \ldots, J_{k,C}\}^\top$. Let $\mathbf{g} = (\mathbf{g}_0, \mathbf{g}_1, \ldots, \mathbf{g}_d)^\top$, the estimate of $\mathbf{g}$ is

$$\hat{\mathbf{g}} = \underset{\mathbf{g}}{\text{argmin}} \sum_{i=1}^{n} \sum_{j=1}^{N_i} \left[Y_{ij} - \sum_{l=1}^{J_{0,C}} g_{0,l}\theta_{0,l}(T_{ij}) - \sum_{k=1}^{d} \left\{ \hat{\phi}_k(Z_{ik}) \sum_{l=1}^{J_{k,C}} g_{kl}\theta_{kl}(T_{ij}) \right\} \right]^2. \tag{12}$$

Hence, the estimates of the overall mean function β_0 is

$$\hat{\beta}_0(t) = \sum_{l=1}^{J_{0,C}} \hat{g}_{0,l}\theta_{0,l}(t), \tag{13}$$

and the estimates of the coefficient functions $\beta_k, k = 1, \ldots, d$, subject to the constraints in (3), are

$$\hat{\beta}_k(t) = c_k^{-1} \sum_{l=1}^{J_{k,C}} \hat{g}_{kl}\theta_{kl}(t), \quad \text{where} \quad c_k = \int \left\{ \sum_{l=1}^{J_{k,C}} \hat{g}_{kl}\theta_{kl}(t) \right\} dt, \quad k = 1, \ldots, d. \tag{14}$$

2.2 Remarks

The estimation of the additive component functions ϕ_k hinges on a reliable estimate of $\int \beta_k(t)dt$, which should be close to 1. This is guaranteed by the dense design of the time points. Details are provided in Appendix B.

In Step I, after intra-subject integration as in (7), the subsequent procedures, including (8), (9) and (10), are similar to fitting an additive model for independent data. In Step II, Eq. (11) can be considered as a varying-coefficient model with vector covariates $\hat{\phi}_k(Z_k)$, so Eq. (12) is similar to the estimation procedure in Huang et al. (2002). However, the estimation errors in $\hat{\phi}_k(Z_k)$ need to be tracked as in the proof of Theorem 2 presented in Appendix B.

The two-step approach can be iterated to update the estimates, $\hat{\phi}_k$ and $\hat{\beta}_k$. However, numerical studies based on simulations not presented in the paper showed no improvements to iterate the estimates.

3 ASYMPTOTIC PROPERTIES

In this section we establish the L^2 rates of convergence for $\hat{\phi}_k, k = 1, \ldots, d$ and $\hat{\beta}_k, k = 0, \ldots, d$. For any two sequences $a(n), b(n) > 0$ we introduce the notation $a(n) \preceq b(n)$ to mean that $\limsup_{n\to\infty} a(n)/b(n) < \infty$, and $a(n) \asymp b(n)$ to mean that $0 < \limsup_{n\to\infty} a(n)/b(n) < \infty$. The L^2 norm of a univariate function $\eta(\cdot) \in [0, 1]$ is denoted by $\|\eta\|_2 = \{\int_0^1 \eta(s)^2 ds\}^{1/2}$.

3.1 Component functions

Let $G_{n,k}$ be the linear space spanned by $\{\psi_{kl}, l = 1, \ldots, J_{k,A}\}$. Denote

$$K_A = \max_{1\le k\le d} K_{k,A} \quad \text{and} \quad \rho_A = \sum_{k=1}^{d} \inf_{\mu_k \in G_{n,k}} \sup_{z_k\in[0,1]} |\phi_k(z_k) - \mu_k(z_k)|.$$

THEOREM 1. *Under Assumptions 1–6, 8–9 in Appendix A, for any $k = 1, \ldots, d$,*

$$\|\hat{\phi}_k(z_k) - \phi_k(z_k)\|_2^2 = O_p\left\{\frac{K_A}{n} + \frac{1}{n^2}(\sum_{i=1}^{n} N_i^{-2})^2 + \rho_A^2\right\}.$$

Remark 1. Compared to the convergence for one-dimensional nonparametric smoothers, Theorem 1 has one additional term $n^{-2}(\sum_{i=1}^{n} N_i^{-2})^2$. This term comes from the intra-subject integration in Step I, which reflects the approximation error of the trapezoid rule to the theoretical integral. The approximation becomes negligible when the relative magnitude of N_i to n is sufficiently large. In particular, if $\min_{1\le i\le n} N_i \ge n^{1/4}$, this additional term is dominated by K_A/n and the rate of convergence in Theorem 1 is exactly the same as a one-dimensional nonparametric smoother. This phenomenon indicates that for dense functional data, the estimated ϕ_k attain the same rate as those whose true β_k are known.

Remark 2. We can obtain more accurate rates of convergence if specific smoothness conditions are provided. For example, if for all $k = 1, \ldots, d$, ϕ_k has bounded second derivatives and $G_{n,k}$ is a cubic spline space, that is, $r_k = 2$ in Assumption 6 and $p_{k,A} = 4$, then by Theorem 6.27 of Schumaker (2007) and Assumption 9,

$$\rho_A \preceq \sum_{k=1}^{d} K_{k,A}^{-2} \asymp K_A^{-2}.$$

If we let $K_A \asymp n^{1/5}$ and $\min_{1\le i\le n} N_i \ge n^{1/4}$, we have the optimal rate of convergence $n^{2/5}$ as in Stone (1985).

3.2 Coefficient functions

Let $N_s = \sum_{i=1}^{n} N_i$, $\bar{N} = N_s/n$, $N_{max} = \max_{1\le i\le n} N_i$, and $H_{n,k}$ be the linear space spanned by $\{\theta_{kl}, l = 1, \ldots, J_{k,C}\}$. Denote

$$K_C = \max_{0\le k\le d} K_{k,C} \quad \text{and} \quad \rho_C = \sum_{k=0}^{d} \inf_{\nu_k\in H_{n,k}} \sup_{t\in[0,1]} |\beta_k(t) - \nu_k(t)|.$$

THEOREM 2. *Under Assumptions 1–10 in Appendix A, for any $k = 0, \ldots, d$, if $\lim_{n\to\infty} \rho_A = 0$,*

$$\|\hat{\beta}_k(t) - \beta_k(t)\|_2^2 = O_p\left[\frac{K_C}{N_s} + \frac{\sum_{i=1}^{n} N_i^2}{N_s^2} + \frac{N_{max}}{\bar{N}}\rho_C^2 + \frac{K_C N_{max} + (N_{max})^2 + \bar{N}N_{max}}{\bar{N}^2}\left\{\frac{K_A}{n} + \frac{1}{n^2}(\sum_{i=1}^{n} N_i^{-2})^2 + \rho_A^2\right\}\right]$$

Remark 3. Compared to the results for varying-coefficient models, the rate of convergence in Theorem 2 has an additional term $\bar{N}^{-2}\{K_C N_{max} + (N_{max})^2 + \bar{N}N_{max}\}\{n^{-1}K_A + n^{-2}(\sum_{i=1}^{n} N_i^{-2})^2 + \rho_A^2\}$, and the term $\bar{N}^{-1}N_{max}\rho_C^2$ replaces the bias ρ_C^2 in varying-coefficient models. They both reflect the effects of replacing the "true" covariate in a varying-coefficient model by an estimate of it, and underscore the strong impact of error-in-variable in regression models.

Remark 4. More accurate rate of convergence can be obtained with more specific conditions on the smoothness of the varying-coefficient functions β_k and the additive component functions ϕ_k, and on the number of observations N_i. For example, if for all $k = 0, \ldots, d, \beta_k$ has bounded second derivatives and $H_{n,k}$ is a cubic spline space, i.e., $s_k = 2$ in Assumption 4 and $p_{k,C} = 4$, then by Theorem 6.27 of Schumaker (2007) and Assumption 10,

$$\rho_C \preceq \sum_{k=0}^{d} K_{k,C}^{-2} \asymp K_C^{-2}.$$

Additionally, suppose that for $k = 1, \ldots, d$, ϕ_k has bounded second derivatives and $G_{n,k}$ is a cubic spline space. If we let $N_{max} \asymp \min_{1\le i\le n} N_i \ge n^{1/4}$, $K_C \asymp n^{1/4}$ and $K_A \asymp n^{1/5}$, then we can achieve the rate $n^{2/5}$ for $\hat{\beta}_k$, which is the comparable rate for one-dimensional nonparametric smoother based on independently and identically distributed data. In the special case when the additive component functions falls in a known and finite dimensional spline spaces, i.e. $\phi_k \in G_{n,k}$ with $K_{k,A}$ known for all k, then $\rho_A = 0$ and we can obtain root-n rate for $\hat{\beta}_k$ if we let $N_{max} \asymp \min_{1\le i\le n} N_i \ge n^{1/4}$, $K_C \asymp n^{1/4}$ and $K_A \asymp 1$.

4 SIMULATION

We generated $Q = 500$ samples with $n = 100$ subjects in each sample. The number of measurements per subject is $N_i = 40$ and the time points $T_{ij}, j = 1, \ldots, N_i$ are equidistant on $[0, 1]$. We set $d = 2$ and the two covariates $(Z_{i1}, Z_{i2})^\top, i = 1, \ldots, n$ are independently and identically distributed copies generated from the MATLAB function copularnd using a Gaussian copula with linear correlation parameter 0.6. Thus, marginally, Z_{i1} and Z_{i2} both have a uniform distribution on $[0, 1]$.

The true functions were chosen as:

$$\begin{aligned} &\phi_1(z_1) = \sin(2\pi z_1), \qquad \phi_2(z_2) = 4z_2^3 - 1; \\ &\beta_0(t) = 1.5\sin\{3\pi(t + 0.5)\} + 4t^3, \quad \beta_1(t) = 3(1 - t)^2, \quad \beta_2(t) = 4t^3. \end{aligned}$$

It is obvious that ϕ_1, ϕ_2, β_1 and β_2 satisfy the constraints (3). We generated $U_{ij} = \sum_{l=1}^{4} A_{il}\gamma_l(T_{ij})$, where A_{il} are independent from $N(0, \lambda_l)$ with $\lambda_l = 1/(l + 1)^2, l = 1, \ldots, 4$, and

$$\begin{aligned} \gamma_1(t) &= 2^{1/2}\cos(2\pi t), \quad \gamma_2(t) = 2^{1/2}\sin(2\pi t), \\ \gamma_3(t) &= 2^{1/2}\cos(4\pi t), \quad \gamma_4(t) = 2^{1/2}\sin(4\pi t). \end{aligned}$$

The random errors e_{ij} are independently drawn from $N(0, 0.01)$. Thus the observed response was obtained by

$$Y_{ij} = \beta_0(T_{ij}) + \sum_{k=1}^{d} \beta_k(T_{ij})\phi_k(Z_{ik}) + U_{ij} + e_{ij}.$$

The evaluation criterion is the mean integrated squared error (MISE) based on $Q = 500$ estimates $\{\hat{\phi}_k(\cdot)^{[q]}, k = 1, \ldots, d, q = 1, \ldots, Q\}$ and $\{\hat{\beta}_k(\cdot)^{[q]}, k = 0, \ldots, d, q = 1, \ldots, Q\}$ from the simulated samples:

$$\text{MISE}(\phi_k) = Q^{-1} \sum_{q=1}^{Q} \int \{\hat{\phi}_k(z_k)^{[q]} - \phi_k(z_k)\}^2 dz_k, \quad \text{MISE}(\beta_k) = Q^{-1} \sum_{q=1}^{Q} \int \{\hat{\beta}_k(t)^{[q]} - \beta_k(t)\}^2 dt.$$

We further define summary measures $\text{MISE}(\phi) = \sum_{k=1}^{d} \text{MISE}(\phi_k)$, $\text{MISE}(\beta) = \sum_{k=0}^{d} \text{MISE}(\beta_k)$, and $\text{MISE}(\phi, \beta) = \text{MISE}(\phi) + \text{MISE}(\beta)$.

AIC or BIC was used to select the number of knots for $\beta_k, k = 0, \ldots, d$ and $\phi_k, k = 1, \ldots, d$. A summary of the selected knots is shown in Table 1 together with the optimal number of knots in the simulation setting. The corresponding MISE values are presented in Table 2, where the preferred choice is to use BIC to select the knots for both ϕ_k's and β_k's. This leads to MISE values comparable to the optimal ones, hence the proposed procedure seems to work well.

5 DATA APPLICATION

The methodology in Section 2 was implemented to a traffic dataset from the Freeway Performance Measurement System (PeMS, `http://pems.eecs.berkeley.edu`). The dataset is available at the University of California, Irvine machine learning repository. It was collected from a loop sensor at an on-ramp for the 101 North freeway in Los Angeles located near Dodger Stadium, which is the home stadium of the Los Angeles Dodgers baseball team. The sensor would detect the traffic volume after a Dodgers game. The period of data we used spanned from April 2005 to October 2005. Measurements were taken every five minutes and the total number of cars in this five-minutes interval were recorded as one measurement. On a day when the Dodgers had a home game, additional information is available, including the time when the game started and ended, game attendance and final scores of the two teams.

Different from Ihler et al. (2006) whose objective was to predict the occurrence of a baseball game at Dodgers stadium, we focused on game days and investigated how car counts around the end of the game are related to game attendance and score difference (home score minus away score). The motivation of our study is to check whether people are likely to leave before the game ends when the attendance is large or when the home team falls behind. We regarded the end time of each game as the onset time 0 and focused on the car counts between 30 minutes before the end ($t = -30$) and 120 minutes after the game (time $t = 120$). The car counts on the 78 game days during April 2005 to October 2005 are shown in Fig. 1, which reveals that 30 minutes before the game ends, the number of cars gradually increases and reaches the peak around 20 minutes after the game and then decreases throughout afterwards.

We fitted this data with the varying-coefficient additive model. For both the additive component and coefficient functions, we used cubic splines and equidistant interior knots, and the number of knots were selected via the BIC method based on our experience in the simulation study. The results were $(K_{1,A}, K_{2,A}) = (1, 2)$ and $(K_{0,C}, K_{1,C}, K_{2,C}) = (2, 5, 1)$. The estimated additive component and coefficient functions are shown in Fig. 2 and Fig. 3 together with their 95% simultaneous confidence bands. The simultaneous confidence bands for the additive component functions were constructed through the Scheffé's method based on the estimated spline coefficients, $\hat{\mathbf{f}}$ in equation (9). This is a simple and practical way to approximate the simultaneous confidence bands because only the subject specific integrated responses, which are independent, were used to estimate the additive component functions. As for the coefficient functions, we used the bootstrap method to construct the simultaneous confidence band, because the coefficient functions were estimated from all Y_{ij}, which are not independent, so the method is not applicable for such dependent data.

Fig. 2 and Fig. 3 reveal that at the 5% level, both functions $\hat{\phi}_1$ and $\hat{\beta}_1$ for attendance are significantly different, respectively, from 0 and the constant function 1 but $\hat{\phi}_2$ and $\hat{\beta}_2$ corresponding to the score difference are not significant. Therefore, game attendance can be considered a significant covariate for car traffic counts and the traditional additive model may not fit this data well. The additive component function became positive when the attendance exceeded approximately $45,000$. This and $\hat{\beta}_1$ in Fig. 3 suggest two additional smaller peaks, one slightly before time 0 and the other around time 50, in addition to the primary peak traffic around time 20 depicted by the mean curve in Fig. 1. This seems reasonable as in order to avoid the primary traffic peak, some people may leave shortly before the game ends, while others who took longer time to leave (could be by choice to avoid the peak traffic) created the peak at time 50.

For attendance less than $45,000$, the additive component function is negative and there is no additional peak traffic near the end of the game. This could be explained as most people would like to stay until the end of the game, if the attendance is relatively small. An interesting finding is that the primary peak size at time 20 is offset by a trough in $\hat{\beta}_1$ in Fig. 3. This is reasonable and underscores the advantage of the proposed varying-coefficient additive model.

6 DISCUSSIONS

The varying-coefficient additive model proposed in this paper alleviates the curse of dimensionality, maintains modeling flexibility, and captures the time dynamic characteristics of functional data. Densely sampled functional responses facilitate separate estimation of the additive component functions and the varying-coefficient functions by respectively fitting an additive model and a varying-coefficient model. If the data is sufficiently dense in the sense that $\max_{1\le i\le n} N_i \asymp \min_{1\le i\le n} N_i \ge n^{1/4}$, we can achieve the optimal L^2 rate of convergence for one-dimensional nonparametric smoothing for both the additive component function and the coefficient function estimates. The estimators perform satisfactorily in the simulation study.

The varying-coefficient additive model (2) can also be used for model checking for two submodels: if each $\phi_k(z_k)$ is a linear function of z_k, model (2) becomes the varying-coefficient model (1); if all $\beta_k(t) = 1, k = 1, \ldots, d$, then model (2) becomes the additive model (5). For the analysis of traffic data, we demonstrate how simultaneous confidence band can be obtained by bootstrapping and it suggests a lack of fit of the traditional additive model. Model checking for the varying-coefficient model is more complicated and will be a future project.

ACKNOWLEDGEMENTS

The research of Jane-Ling Wang was supported by NSF grant DMS-09-06813.

APPENDIX A: Assumptions

We first list the assumptions needed to prove the asymptotic properties.

Assumption 1. $\{T_{ij}, i = 1, \ldots, n, j = 1, \ldots, N_i\}$ are independent and identical copies of T whose density function f_T is bounded from below and above, i.e., $m_T \le f_T(t) \le M_T$, for all $t \in [0, 1]$.

Assumption 2. The joint density of $\mathbf{Z}$, $p(\mathbf{z})$ is bounded below and above on its domain $[0, 1]^d$:

$$m_p \le \inf_{\mathbf{z}\in[0,1]^d} p(\mathbf{z}) \le \sup_{\mathbf{z}\in[0,1]^d} p(\mathbf{z}) \le M_p.$$

Assumption 3. $N_i \to \infty$ as $n \to \infty$ for all $i = 1, \ldots, n$.

Assumption 4. $\beta_k(t)$ has bounded s_kth derivatives, where $s_k \ge 2$ for all $k = 0, \ldots, d$.

Assumption 5. $\sup_{s,t\in[0,1]} |\mathrm{cov}\{U(s), U(t)\}| \le G < \infty$.

Assumption 6. ϕ_k has bounded r_kth derivatives, for $k = 1, \ldots, d$., where $r_k \ge 0$.

Assumption 7. Let $\Phi(\mathbf{z}) = (1, \phi_1(z_1), \ldots, \phi_d(z_d))^\top$. The eigenvalues of $\int \Phi(\mathbf{z})\Phi(\mathbf{z})^\top p(\mathbf{z})d\mathbf{z}$ are bounded away from 0 and ∞.

Assumption 8. The knots for $\phi_k, k = 1, \ldots, d$

$$0 = \tau_{k,1-p_{k,A}} = \cdots = \tau_{k,0} < \tau_{k,1} < \cdots < \tau_{k,K_{k,A}} < \tau_{k,K_{k,A}+1} = \cdots = \tau_{k,K_{k,A}+p_{k,A}} = 1,$$

and the knots for $\beta_k, k = 0, \ldots, d$

$$0 = \zeta_{k,1-p_{k,C}} = \cdots = \zeta_{k,0} < \zeta_{k,1} < \cdots < \zeta_{k,K_{k,C}} < \zeta_{k,K_{k,C}+1} = \cdots = \zeta_{k,K_{k,C}+p_{k,C}} = 1,$$

have bounded mesh ratio:

$$\limsup_{n\to\infty} \max_{1\le k\le d} \frac{\max_{1\le l\le K_{k,A}+1}\{\tau_{k,l} - \tau_{k,l-1}\}}{\min_{1\le l\le K_{k,A}+1}\{\tau_{k,l} - \tau_{k,l-1}\}} < \infty, \quad \limsup_{n\to\infty} \max_{0\le k\le d} \frac{\max_{1\le l\le K_{k,C}+1}\{\zeta_{k,l} - \zeta_{k,l-1}\}}{\min_{1\le l\le K_{k,C}+1}\{\zeta_{k,l} - \zeta_{k,l-1}\}} < \infty.$$

Assumption 9. $\limsup_{n\to\infty}(K_A / \min_{1\le k\le d} K_{k,A}) < \infty$ and $K_A \log(K_A)/n \to 0$ as $n \to \infty$.

Assumption 10. $\limsup_{n\to\infty}(K_C / \min_{0\le k\le d} K_{k,C}) < \infty$ and $n\,(N_{max})^2 K_C \log(K_C)/N_s^2 \to 0$ as $n \to \infty$.

Assumptions 1 and 2 are standard assumptions for the measurement schedule and covariates in the literature of varying-coefficient models. The postulated dense designs and the smoothness assumption with $s_k = 2$ in Assumption 3 and 4 are used in Lemma 1 below to control the precision in the approximation of the trapezoid integral in (7). Higher order of smoothness coupled with higher order splines lead to better orders for the bias term in the spline method. Assumptions 5 and 6 on the covariance and additive component functions are fairly standard. Assumption 7, which is used in the proof of Theorem 2, is very similar to Assumption 2 in Huang et al. (2002) and Assumption (C2) in Huang et al. (2004). Assumption 8 is standard for spline regression methods. Assumptions 9 and 10 are similar to Assumption (C5) and the assumptions used in Lemma A.2 of Huang et al. (2004).

APPENDIX B: Proofs

Identifiability

Proof of Proposition 1. Since $\beta_0(t) = E\{W(t)\}$, it suffices to prove that β_k and ϕ_k are identifiable from (3), for $k = 1, \ldots, d$. Suppose that there exist $\{\beta_k, \phi_k, k = 1, \ldots, d\}$ and $\{\check{\beta}_k, \check{\phi}_k, k = 1, \ldots, d\}$ such that

$$\sum_{k=1}^{d} \beta_k(t)\phi_k(z_k) = \sum_{k=1}^{d} \check{\beta}_k(t)\check{\phi}_k(z_k), \tag{15}$$

for all t and $\mathbf{z} = (z_1, \ldots, z_d)^\top$. To identify β_j and ϕ_j, multiply both sides of (15) by the joint density of $\mathbf{z}_{-j}$, which represents all z_k but z_j. Next, integrate both sides with respect to $\mathbf{z}_{-j}$ and we have

$$\beta_j(t)\phi_j(z_j) = \check{\beta}_j(t)\check{\phi}_j(z_j).$$

Integrating with respect to t then leads to $\phi_j(\cdot) = \check{\phi}_j(\cdot)$ and thus $\beta_j(\cdot) = \check{\beta}_j(\cdot)$. □

Errors from numerical integration

We can rewrite $\tilde{Y}_i$ in (7) as

$$\tilde{Y}_i = \tilde{\beta}_0 + \sum_{k=1}^{d} \phi_k(Z_{ik}) + \eta_i, \quad \text{where} \quad \eta_i = \tilde{\delta}_i - \Delta_0^i - \sum_{k=1}^{d} \Delta_k^i \phi_k(Z_{ik}),$$
$$\tilde{\delta}_i = \frac{1}{2}\sum_{j=1}^{N_i-1}(\delta_{ij}^* + \delta_{i,j+1}^*)(T_{i,j+1}^* - T_{ij}^*) + \delta_{i1}^* T_{i1}^* + \delta_{i,N_i}^*(1 - T_{i,N_i}^*), \tag{16}$$

and Δ_k^i, for any $k = 0, \ldots, d$, is the error when using the trapezoid rule to approximate the integral $\int_0^1 \beta_k(t)dt$ for subject i:

$$\Delta_k^i = \int_0^1 \beta_k(t)dt - \frac{1}{2}\sum_{j=1}^{N_i-1}\{\beta_k(T_{ij}^*) + \beta_k(T_{i,j+1}^*)\}(T_{i,j+1}^* - T_{ij}^*) - \beta_k(T_{i1}^*)T_{i1}^* - \beta_k(T_{i,N_i}^*)(1 - T_{i,N_i}^*). \tag{17}$$

The following lemma gives the properties of Δ_k^i which is used to prove Theorem 1.

LEMMA 1. *Under Assumptions 1 and 4, there exits a constant* $0 < B_1 < \infty$ *such that for any* $i = 1, \ldots, n$ *and* $k = 0, \ldots, d$,

$$E(|\Delta_k^i|) \leq B_1\{(N_i + 1)(N_i + 2)\}^{-1}, \quad E(|\Delta_k^i|^2) \leq B_1.$$

Proof. We can decompose Δ_k^i in (17) into $\Delta_k^i = I_k^i + II_k^i + III_k^i$ where

$$I_k^i = \int_0^{T_{i1}^*} \{\beta_k(t) - \beta_k(T_{i1}^*)\}\, dt; \qquad II_k^i = \int_{T_{i,N_i}^*}^1 \left\{\beta_k(t) - \beta_k(T_{i,N_i}^*)\right\} dt;$$
$$III_k^i = \int_{T_{i1}^*}^{T_{i,N_i}^*} \beta_k(t)dt - \frac{1}{2}\sum_{j=1}^{N_i-1}\{\beta_k(T_{ij}^*) + \beta_k(T_{i,j+1}^*)\}(T_{i,j+1}^* - T_{ij}^*).$$

The proof of this lemma uses the following three properties related to I_k^i, II_k^i and III_k^i:

Property 1. $E\{(T_{i1}^*)^2\} \leq 2\{m_T^2(N_i + 1)(N_i + 2)\}^{-1}$ and $E\{(1 - T_{i,N_i}^*)^2\} \leq 2\{m_T^2(N_i + 1)(N_i + 2)\}^{-1}$.

Property 2. There exists $C_1 > 0$ such that for all $k = 0, \ldots, d$, $E(|I_k^i|) \leq C_1\{(N_i + 1)(N_i + 2)\}^{-1}$, $E(|II_k^i|) \leq C_1\{(N_i + 1)(N_i + 2)\}^{-1}$ and $E(|III_k^i|) \leq C_1(N_i - 1)\{(N_i + 1)(N_i + 2)(N_i + 3)\}^{-1}$.

Property 3. $E(|I_k^i|^2)$, $E(|II_k^i|^2)$ and $E(|III_k^i|^2)$ are bounded by a constant $0 < C_2 < \infty$.

The proofs of the three properties are as follows:

Proof of Property 1: Note that $T_{i1}^* \geq 0$ and

$$E\{(T_{i1}^*)^2\} = 2\int_0^1 t\{\text{pr}\,(T_{i1}^* > t)\}dt = 2\int_0^1 t\,\{1 - F(t)\}^{N_i}\, dt \leq \frac{2}{m_T}\int_0^1 t\{1 - F(t)\}^{N_i} dF(t)$$
$$= \frac{2}{m_T(N_i + 1)}\int_0^1 \{1 - F(t)\}^{N_i+1}dt \leq \frac{2}{m_T^2(N_i + 1)(N_i + 2)}.$$

Similarly $E\{(1 - T_{i,N_i}^*)^2\} \leq 2\{m_T^2(N_i + 1)(N_i + 2)\}^{-1}$.

Proof of Property 2: By Assumption 4, there exists $\tilde{C}_1 > 0$ such that the derivative $|\beta_k(t)^{(1)}| \leq \tilde{C}_1$ for $t \in [0, 1]$ and all $k = 0, \ldots, d$. Therefore $E(|I_k^i|) \leq \tilde{C}_1 E\{(T_{i1}^*)^2\} \leq 2\tilde{C}_1\{m_T^2(N_i + 1)(N_i + 2)\}^{-1}$ and $E(|II_k^i|) \leq \tilde{C}_1 E\{(1 - T_{i,N_i}^*)^2\} \leq 2\tilde{C}_1\{m_T^2(N_i + 1)(N_i + 2)\}^{-1}$. By Dragomir et al. (2000), $E(|III_k^i|) \leq \tilde{C}_2\sum_{j=1}^{N_i-1} E(h_{ij}^3)/12$, where $\sup_{t\in[0,1]} |\beta_k^{(2)}(t)| \leq \tilde{C}_2$ by Assumption 4 and $h_{ij} = T_{i,j+1}^* - T_{ij}^*$. By Corollary 1 of Jones & Balakrishnan (2002), $E(h_{ij}^3) \leq 6\{m_T^3(N_i + 1)(N_i + 2)(N_i + 3)\}^{-1}$, so $E(|III_k^i|) \leq \tilde{C}_2(N_i - 1)\{2m_T^3(N_i + 1)(N_i + 2)(N_i + 3)\}^{-1}$. Let $C_1 = \max\{2\tilde{C}_1/m_T^2, \tilde{C}_2/2m_T^3\}$.

Proof of Property 3: $E(|I_k^i|^2) \le \tilde{C}_1^2 E\{(T_{i1}^*)^4\} \le \tilde{C}_1^2$. Similarly $E(|II_k^i|^2) \le \tilde{C}_1^2$.

$$E(|III_k^i|^2) \le \tilde{C}_2^2 E\{(\sum_{j=1}^{N_i-1} h_{ij})^2\}/12^2 \le \tilde{C}_2^2/12^2.$$

Let $C_2 = \max\{\tilde{C}_1^2, \tilde{C}_2^2/12^2\}$.

The lemma now follows by letting $B_1 = \max\{3C_1, 9C_2\}$. □

Random error $\tilde{\delta}_i$ in Eq. (16)

LEMMA 2. *Under Assumption 5, we have, for any $i = 1, \ldots, n$,*

$$E(\tilde{\delta}_i) = 0, \qquad E(\tilde{\delta}_i^2) \le G + \sigma^2.$$

Proof. Note $\delta_{ij} = U_i(T_{ij}) + e_{ij}$. It is obvious that $E(\tilde{\delta}_i) = 0$ since $E(\delta_{ij}) = 0$. For $E(\tilde{\delta}_i^2)$, note that

$$\begin{aligned}\tilde{\delta}_i &= \frac{1}{2}\sum_{j=1}^{N_i-1}(\delta_{ij}^* + \delta_{i,j+1}^*)(T_{i,j+1}^* - T_{ij}^*) + \delta_{i1}^* T_{i1}^* + \delta_{i,N_i}^*(1 - T_{i,N_i}^*) \\ &= \frac{1}{2}\delta_{i1}^*(T_{i1}^* + T_{i2}^*) + \delta_{i,N_i}^*(1 - \frac{1}{2}T_{i,N_i}^* - \frac{1}{2}T_{i,N_i-1}^*) + \frac{1}{2}\sum_{j=2}^{N_i-1}\delta_{ij}^*(h_{ij} + h_{i,j-1}),\end{aligned}$$

where $h_{ij} = T_{i,j+1}^* - T_{ij}^*$. Since $E\{(\delta_{ij}^*)^2 \mid \{T_{ij}^*, j = 1, \ldots, N_i\}\} \le G + \sigma^2$ and $E(\delta_{ij}^*\delta_{il}^* \mid \{T_{ij}^*, j = 1, \ldots, N_i\}) \le G + \sigma^2$ by Cauchy–Schwartz inequality, we have

$$\begin{aligned}E\left(\tilde{\delta}_i^2 \mid \{T_{ij}^*, j = 1, \ldots, N_i\}\right) &\le (G + \sigma^2)\left\{\frac{1}{2}(T_{i1}^* + T_{i2}^*) + (1 - \frac{1}{2}T_{i,N_i}^* - \frac{1}{2}T_{i,N_i-1}^*) + \frac{1}{2}\sum_{j=2}^{N_i-1}(h_{ij} + h_{i,j-1})\right\}^2 \\ &= G + \sigma^2.\end{aligned}$$

Thus $E(\tilde{\delta}_i^2) \le G + \sigma^2$. □

Theorem 1

We first introduce the notations needed to prove Theorem 1. For each $k = 1, \ldots, d$, let $\Psi_k(\cdot) = (\psi_{k,1}(\cdot), \ldots, \psi_{k,J_{k,A}}(\cdot))^\top$ and $\psi_{kl} = J_{k,A}^{1/2} B_{kl}$, where $\{B_{kl}, l = 1, \ldots, J_{k,A}\}$ are B-splines defined in de Boor (2001). The properties of B-splines imply that $\psi_{kl}(\cdot) \ge 0$, $\sum_{l=1}^{J_{k,A}} \psi_{kl}(z) = J_{k,A}^{1/2}$, $\int \psi_{kl}(z)dz \le M_1 J_{k,A}^{-1/2}$, and for any vector $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_{J_{k,A}})^\top$,

$$M_2\|\boldsymbol{\alpha}\|^2 \leq \int \left(\sum_{l=1}^{J_{k,A}} \alpha_l \psi_{kl}(z) \right)^2 dz \leq M_3\|\boldsymbol{\alpha}\|^2 \tag{18}$$

where M_1, M_2, M_3 are positive constants and $\|\boldsymbol{\alpha}\|$ represents the Euclidean norm of a vector $\boldsymbol{\alpha}$.

Let $w(\mathbf{z}) = \tilde{\beta}_0 + \sum_{k=1}^{d} \phi_k(z_k)$ be the true regression function in equation (5) and $\hat{w}(\mathbf{z}) = \check{\beta}_0 + \sum_{k=1}^{d} \hat{\phi}_k(z_k)$, where $\check{\beta}_0$ and $\hat{\phi}_k$ are the spline estimator obtained from (9) and (10). Denote $\hat{\boldsymbol{\alpha}} = (\check{\beta}_0, \hat{\mathbf{f}})^\top$, $V(\mathbf{Z}) = (1, \Psi_1(Z_1)^\top, \ldots, \Psi_d(Z_d)^\top)^\top$, and $V_i = V(\mathbf{Z}_i)$, then $\hat{\boldsymbol{\alpha}} = (\sum_{i=1}^{n} V_i V_i^\top)^{-1}(\sum_{i=1}^{n} V_i \tilde{Y}_i)$ and $\hat{\phi}_k(z_k) = \check{\phi}_k(z_k) - n^{-1} \sum_{i=1}^{n} \check{\phi}_k(Z_{ik})$ where $\check{\phi}_k(z_k) = \Psi_k(z_k)^\top \hat{\mathbf{f}}_k$. Also define $\tilde{\boldsymbol{\alpha}} = (\beta_0^*, \tilde{\mathbf{f}})^\top = (\sum_{i=1}^{n} V_i V_i^\top)^{-1}(\sum_{i=1}^{n} V_i w_i)$, where $w_i = w(\mathbf{Z}_i)$ and $\tilde{\phi}_k(z_k) = \Psi_k(z_k)^\top \tilde{\mathbf{f}}_k$.

We next provide a lemma which will be repeatedly used in the subsequent proof.

LEMMA 3. *Under the assumptions in Theorem 1, both $n^{-1} \sum_{i=1}^{n} V_i V_i^\top$ and $n^{-1} \sum_{i=1}^{n} \Psi_k(Z_{ik})\Psi_k(Z_{ik})^\top$ have eigenvalues bounded away from 0 and ∞, with probability tending to one as $n \to \infty$.*

Proof. The proof can follow similar arguments in Lemma A.2 of Huang et al. (2004) and thus is omitted. □

Proof of Theorem 1. By Cauchy–Schwartz inequality,

$$\begin{aligned}\|\hat{\phi}_k(z_k) - \phi_k(z_k)\|^2 \leq 5\|\check{\phi}_k(z_k) - \tilde{\phi}_k(z_k)\|^2 + \frac{5}{n} \sum_{i=1}^{n} \{\check{\phi}_k(Z_{ik}) - \tilde{\phi}_k(Z_{ik})\}^2 \\ +5\|\tilde{\phi}_k(z_k) - \phi_k(z_k)\|^2 + \frac{5}{n} \sum_{i=1}^{n} \{\tilde{\phi}_k(Z_{ik}) - \phi_k(Z_{ik})\}^2 + 5\left|\frac{1}{n} \sum_{i=1}^{n} \phi_k(Z_{ik})\right|^2.\end{aligned}$$

Since $E\{\phi_k(Z_{ik})\} = 0$, $|n^{-1} \sum_{i=1}^{n} \phi_k(Z_{ik})|^2 = O_p(n^{-1})$. It suffices to handle the approximation error terms $\|\tilde{\phi}_k(z_k) - \phi_k(z_k)\|^2$ and $n^{-1} \sum_{i=1}^{n} \{\tilde{\phi}_k(Z_{ik}) - \phi_k(Z_{ik})\}^2$, and the stochastic error terms $\|\check{\phi}_k(z_k) - \tilde{\phi}_k(z_k)\|^2$ and $n^{-1} \sum_{i=1}^{n} \{\check{\phi}_k(Z_{ik}) - \tilde{\phi}_k(Z_{ik})\}^2$.

Approximation errors: We show below the following rate of approximation errors:

$$\|\tilde{\phi}_k(z_k) - \phi_k(z_k)\|^2 = O_p\left(\rho_A^2\right), \quad \text{and} \quad \frac{1}{n} \sum_{i=1}^{n} \{\tilde{\phi}_k(Z_{ik}) - \phi_k(Z_{ik})\}^2 = O_p\left(\rho_A^2\right). \tag{19}$$

By the definition of ρ_A, we can find $\boldsymbol{\alpha}^* = (\tilde{\beta}_0, \mathbf{f}^*)^\top$ and $\phi_k(z_k)^* = \Psi_k(z_k)^\top \mathbf{f}_k^*$ such that $\sup_{z_k \in [0,1]} |\phi_k(z_k)^* - \phi_k(z_k)| = O(\rho_A)$. Thus $\|\phi_k^* - \phi_k\|^2 = O(\rho_A^2)$ and $n^{-1} \sum_{i=1}^{n} \{\phi_k(Z_{ik})^* - \phi_k(Z_{ik})\}^2 = O(\rho_A^2)$. Now it suffices to prove $\|\tilde{\phi}_k(z_k) - \phi_k(z_k)^*\|^2 = O_p(\rho_A^2)$ and $n^{-1} \sum_{i=1}^{n} \{\tilde{\phi}_k(Z_{ik}) - \phi_k(Z_{ik})^*\}^2 = O_p\left(\rho_A^2\right)$.

By Assumption 2 and (18), $\|\tilde{\phi}_k(z_k) - \phi_k(z_k)^*\|^2 = \|\Psi_k(z_k)^\top(\tilde{\mathbf{f}}_k - \mathbf{f}_k^*)\|^2 \asymp \|\tilde{\mathbf{f}}_k - \mathbf{f}_k^*\|^2 \le \|\tilde{\boldsymbol{\alpha}} - \boldsymbol{\alpha}^*\|^2$. By Lemma 3, $n^{-1}\sum_{i=1}^n\{\tilde{\phi}_k(Z_{ik}) - \phi_k(Z_{ik})^*\}^2 = (\tilde{\mathbf{f}}_k - \mathbf{f}_k^*)^\top\{n^{-1}\sum_{i=1}^n \Psi_k(Z_{ik})\Psi_k(Z_{ik})^\top\}(\tilde{\mathbf{f}}_k - \mathbf{f}_k^*) \asymp \|\tilde{\mathbf{f}}_k - \mathbf{f}_k^*\|^2 \le \|\tilde{\boldsymbol{\alpha}} - \boldsymbol{\alpha}^*\|^2$. Therefore we only need to show $\|\tilde{\boldsymbol{\alpha}} - \boldsymbol{\alpha}^*\|^2 = O_p(\rho_A^2)$.

Since $n^{-1}\sum_{i=1}^n V_i(w_i - V_i^\top\tilde{\boldsymbol{\alpha}}) = \mathbf{0}$, Lemma 3 implies, with probability approaching one,

$$\|\tilde{\boldsymbol{\alpha}} - \boldsymbol{\alpha}^*\|^2 \asymp \frac{1}{n}\sum_{i=1}^n\{V_i^\top(\tilde{\boldsymbol{\alpha}} - \boldsymbol{\alpha}^*)\}^2 \le \frac{1}{n}\sum_{i=1}^n(w_i - V_i^\top\boldsymbol{\alpha}^*)^2 = \frac{1}{n}\sum_{i=1}^n\left[\sum_{k=1}^d\{\phi_k(Z_{ik}) - \phi_k(Z_{ik})^*\}\right]^2 = O(\rho_A^2).$$

Therefore, (19) is proved.

Stochastic error: We next show the following rate of stochastic errors:

$$\begin{aligned} \|\check{\phi}_k(z_k) - \tilde{\phi}_k(z_k)\|^2 &= O_p\left\{\frac{K_A}{n} + \frac{1}{n^2}(\sum_{i=1}^n N_i^{-2})^2\right\}, \\ \frac{1}{n}\sum_{i=1}^n\{\check{\phi}_k(Z_{ik}) - \tilde{\phi}_k(Z_{ik})\}^2 &= O_p\left\{\frac{K_A}{n} + \frac{1}{n^2}(\sum_{i=1}^n N_i^{-2})^2\right\}. \end{aligned} \tag{20}$$

By (18) and Lemma 3, $\|\check{\phi}_k(z_k) - \tilde{\phi}_k(z_k)\|^2 \le \|\hat{\boldsymbol{\alpha}} - \tilde{\boldsymbol{\alpha}}\|^2$ and $n^{-1}\sum_{i=1}^n\{\check{\phi}_k(Z_{ik}) - \tilde{\phi}_k(Z_{ik})\}^2 \le \|\hat{\boldsymbol{\alpha}} - \tilde{\boldsymbol{\alpha}}\|^2$, so it suffices to handle $\|\hat{\boldsymbol{\alpha}} - \tilde{\boldsymbol{\alpha}}\|^2$. Since $\|\hat{\boldsymbol{\alpha}} - \tilde{\boldsymbol{\alpha}}\|^2 = \|(\sum_{i=1}^n V_iV_i^\top)^{-1}(\sum_{i=1}^n V_i\eta_i)\|^2 \asymp \|n^{-1}\sum_{i=1}^n V_i\eta_i\|^2$ by Lemma 3, we only need to focus on $\|n^{-1}\sum_{i=1}^n V_i\eta_i\|^2$.

The rest of the proof can mostly follow standard procedure in tracking the order for the variance of the stochastic error term for spline methods, except that η_i in (8) does not have zero conditional mean given the value of $\mathbf{Z}_i$ due to the integration error associated with $\tilde{Y}_i$. So we need to track $E(\eta_i \mid \mathbf{Z}_i)$ carefully. From the expression of η_i in (16),

$$\eta_i = \tilde{\delta}_i - \Delta_0^i - \sum_{k=1}^d \Delta_k^i\phi_k(Z_{ik}).$$

By Assumption 6, Lemmas 1 and 2, there exists $0 < B_2 < \infty$ such that, for all $i = 1, \ldots, n$,

$$\left|E(\eta_i \mid \mathbf{Z}_i)\right| \le B_2N_i^{-2}, \quad E(\eta_i^2 \mid \mathbf{Z}_i) \le B_2.$$

By (18) and Assumption 2, $EV_i^\top V_i = 1 + \sum_{k=1}^d\sum_{l=1}^{J_{k,A}} E\psi_{kl}^2(Z_{ik}) = O(K_A)$ and $EV_i^\top V_j = 1 +$

$\sum_{k=1}^{d}\sum_{l=1}^{J_{k,A}} E\psi_{kl}(Z_{ik})E\psi_{kl}(Z_{jk}) = O(1)$. Therefore,

$$E\left(\left\|\frac{1}{n}\sum_{i=1}^{n} V_i\eta_i\right\|^2\right) = \frac{1}{n^2}\left[\sum_{i=1}^{n} E\left\{V_i^\top V_i E(\eta_i^2 \mid \mathbf{Z}_i)\right\} + \sum_{i\neq j} E\left\{V_i^\top V_j E(\eta_i \mid \mathbf{Z}_i)E(\eta_j \mid \mathbf{Z}_j)\right\}\right]$$
$$\leq \frac{1}{n^2}\left\{B_2\sum_{i=1}^{n} E(V_i^\top V_i) + B_2^2\sum_{i\neq j}\frac{1}{N_i^2}\frac{1}{N_j^2}E(V_i^\top V_j)\right\} = O\left\{\frac{K_A}{n} + \frac{1}{n^2}\left(\sum_{i=1}^{n}\frac{1}{N_i^2}\right)^2\right\},$$

and thus

$$\|\hat{\boldsymbol{\alpha}} - \tilde{\boldsymbol{\alpha}}\|^2 = O_p\left\{\frac{K_A}{n} + \frac{1}{n^2}\left(\sum_{i=1}^{n}\frac{1}{N_i^2}\right)^2\right\},$$

which completes the proof for (20).

Therefore, Theorem 1 holds by (19) and (20). □

Theorem 2

We first introduce the notations to prove Theorem 2. Let $\theta_{kl} = J_{k,C}^{1/2}B_{kl}$, where $\{B_{kl}, l = 1, \ldots, J_{k,C}\}$ are B-splines defined in de Boor (2001). Similar to (18), $\theta_{kl}(\cdot) \geq 0$, $\sum_{l=1}^{J_{k,C}}\theta_{kl}(t) = J_{k,C}^{1/2}$, $\int\theta_{kl}(t)dt \leq M_4 J_{k,C}^{-1/2}$, and for any vector $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_{J_{k,C}})^\top$,

$$M_5\|\boldsymbol{\gamma}\|^2 \leq \int\left(\sum_{l=1}^{J_{k,C}}\gamma_l\theta_{kl}(t)\right)^2 dt \leq M_6\|\boldsymbol{\gamma}\|^2, \tag{21}$$

where M_4, M_5, M_6 are positive constants and $\|\boldsymbol{\gamma}\|$ represents the Euclidean norm of a vector $\boldsymbol{\gamma}$.

Define $\Theta_k(t) = \{\theta_{k,1}(t), \ldots, \theta_{k,J_{k,C}}(t)\}^\top$ for all $k = 0, \ldots, d$ and also define

$$\Theta(t) = \begin{pmatrix} \Theta_0(t)^\top & 0 & \cdots & 0 \\ & \ddots & \ddots & \\ 0 & \cdots & 0 & \Theta_d(t)^\top \end{pmatrix},$$

$\Phi(\mathbf{z}) = \{1, \phi_1(z_1), \ldots, \phi_d(z_d)\}^\top$, $\hat{\Phi}(\mathbf{z}) = \{1, \hat{\phi}_1(z_1), \ldots, \hat{\phi}_d(z_d)\}^\top$, $D_{ij} = \left\{\Phi(\mathbf{Z}_i)^\top\Theta(T_{ij})\right\}^\top$, $\tilde{D}_{ij} = \left\{\hat{\Phi}(\mathbf{Z}_i)^\top\Theta(T_{ij})\right\}^\top$, $\mathbf{D}_i = (D_{i1}, \ldots, D_{i,N_i})^\top$, $\tilde{\mathbf{D}}_i = (\tilde{D}_{i1}, \ldots, \tilde{D}_{i,N_i})^\top$, and $\mathbf{Y}_i = (Y_{i1}, \ldots, Y_{i,N_i})^\top$. Then $\hat{\mathbf{g}}$ in (12) can be expressed as

$$\hat{\mathbf{g}} = \left(\frac{1}{N_s}\sum_{i=1}^{n}\tilde{\mathbf{D}}_i^\top\tilde{\mathbf{D}}_i\right)^{-1}\left(\frac{1}{N_s}\sum_{i=1}^{n}\tilde{\mathbf{D}}_i^\top\mathbf{Y}_i\right).$$

Let $m(t,\mathbf{z}) = \beta_0(t) + \sum_{k=1}^{d}\beta_k(t)\phi_k(z_k)$, $\mathbf{m}_i = \{m(T_{i1},\mathbf{Z}_i), \ldots, m(T_{i,N_i},\mathbf{Z}_i)\}^\top$, $\tilde{m}(t,\mathbf{z}) = \beta_0(t) + \sum_{k=1}^{d}\beta_k(t)\hat{\phi}_k(z_k)$, and $\tilde{\mathbf{m}}_i = \{\tilde{m}(T_{i1},\mathbf{Z}_i), \ldots, \tilde{m}(T_{i,N_i},\mathbf{Z}_i)\}^\top$. Define $\tilde{\mathbf{g}} = (N_s^{-1}\sum_{i=1}^{n}\tilde{\mathbf{D}}_i^\top\tilde{\mathbf{D}}_i)^{-1}(N_s^{-1}\sum_{i=1}^{n}\tilde{\mathbf{D}}_i^\top\mathbf{m}_i)$,

$\check{\beta}_k(t) = \Theta_k(t)^\top \hat{\mathbf{g}}_k$ and $\tilde{\beta}_k(t) = \Theta_k(t)^\top \tilde{\mathbf{g}}_k$ for $k = 0, \ldots, d$. Thus $\hat{\beta}_k(t) = \check{\beta}_k(t)/\int \check{\beta}_k(t)dt$.

The next lemma will be repeatedly used in the subsequent proof.

LEMMA 4. *Under the assumptions in Theorem 2, the eigenvalues of $N_s^{-1}\sum_{i=1}^n \tilde{\mathbf{D}}_i^\top \tilde{\mathbf{D}}_i$ are bounded away from 0 and ∞, with probability tending to one as $n \to \infty$.*

Proof. By Theorem 1 and the fact $\lim_{n\to\infty}\rho_A = 0$, $\|\hat{\phi}_k(z_k) - \phi_k(z_k)\|^2 = o_p(1)$ and thus by Assumption 7, the eigenvalues of $\int \hat{\Phi}(\mathbf{z})\hat{\Phi}(\mathbf{z})^\top p(\mathbf{z})d\mathbf{z}$ are bounded away from 0 and ∞ with probability approaching one as $n \to \infty$. The rest of the proof can follow similar arguments in Lemma A.2 of Huang et al. (2004) and thus is omitted. □

Proof of Theorem 2. By $\int \beta_k(t)dt = 1$ in (3) and Cauchy–Schwartz inequality,

$$\|\hat{\beta}_k(t) - \beta_k(t)\|^2 = \left\|\frac{\check{\beta}_k(t)}{\int \check{\beta}_k(t)dt} - \beta_k(t)\right\|^2 \le 2\|\check{\beta}_k(t) - \beta_k(t)\|^2 + 2\left\|\frac{\check{\beta}_k(t)}{\int \check{\beta}_k(t)dt} - \check{\beta}_k(t)\right\|^2$$
$$\le 2\|\check{\beta}_k(t) - \beta_k(t)\|^2 + 4\|\check{\beta}_k(t) - \beta_k(t)\|^2\left|\frac{\int\left\{\check{\beta}_k(t) - \beta_k(t)\right\}dt}{\int\left\{\check{\beta}_k(t) - \beta_k(t)\right\}dt + 1}\right|^2 + 4\|\beta_k(t)\|^2\left|\frac{\int\left\{\check{\beta}_k(t) - \beta_k(t)\right\}dt}{\int\left\{\check{\beta}_k(t) - \beta_k(t)\right\}dt + 1}\right|^2,$$

it thus suffices to prove the convergence of $\|\check{\beta}_k(t) - \beta_k(t)\|^2$. Moreover, since $\|\check{\beta}_k(t) - \beta_k(t)\|^2 \le 2\|\check{\beta}_k(t) - \tilde{\beta}_k(t)\|^2 + 2\|\tilde{\beta}_k(t) - \beta_k(t)\|^2$, we only need to handle the approximation error term $\|\tilde{\beta}_k(t) - \beta_k(t)\|^2$ and the stochastic error term $\|\check{\beta}_k(t) - \tilde{\beta}_k(t)\|^2$.

Approximation error: There are two sources of approximation error, one attributed to the spline approximation of the coefficient function β_k and the other induced by estimating the additive component function ϕ_k in (11). The combined approximation error is shown below to have the following rate:

$$\|\tilde{\beta}_k(t) - \beta_k(t)\|^2 = O_p\left[\frac{N_{max}}{\bar{N}}\rho_C^2 + \frac{N_{max}}{\bar{N}}\left\{\frac{K_A}{n} + \frac{1}{n^2}(\sum_{i=1}^n N_i^{-2})^2 + \rho_A^2\right\}\right]. \tag{22}$$

By the definition of ρ_C, there exists $\mathbf{g}^*$ and $\beta_k(t)^* = \Theta_k(t)^\top \mathbf{g}_k^*$ such that $\sup_{t\in[0,1]}|\beta_k(t)^* - \beta_k(t)| = O(\rho_C)$ for $k = 0, \ldots, d$. Therefore $\|\beta_k(t)^* - \beta_k(t)\|^2 = O_p(\rho_C^2)$ and it suffices to prove $\|\tilde{\beta}_k(t) - \beta_k(t)^*\|^2 = O_p[\bar{N}^{-1}(N_{max})\{\rho_C^2 + n^{-1}K_A + n^{-2}(\sum_{i=1}^n N_i^{-2})^2 + \rho_A^2\}]$. By (21), $\|\tilde{\beta}_k(t) - \beta_k(t)^*\|^2 = \|\Theta_k(t)^\top(\tilde{\mathbf{g}}_k - \mathbf{g}_k^*)\|^2 \asymp \|\tilde{\mathbf{g}}_k - \mathbf{g}_k^*\|^2 \le \|\tilde{\mathbf{g}} - \mathbf{g}^*\|^2$, so we only need to show $\|\tilde{\mathbf{g}} - \mathbf{g}^*\|^2 = O_p[\bar{N}^{-1}(N_{max})\{\rho_C^2 + n^{-1}K_A + n^{-2}(\sum_{i=1}^n N_i^{-2})^2 + \rho_A^2\}]$.

Since $N_s^{-1}\sum_{i=1}^n \tilde{D}_i^\top(\mathbf{m}_i - \tilde{D}_i\tilde{\mathbf{g}}) = \mathbf{0}$, Lemma 4 implies, with probability approaching one

$$\|\tilde{\mathbf{g}} - \mathbf{g}^*\|^2 \asymp \frac{1}{N_s}\sum_{i=1}^n \|\tilde{D}_i^\top(\tilde{\mathbf{g}} - \mathbf{g}^*)\|^2 \le \frac{1}{N_s}\sum_{i=1}^n \|\mathbf{m}_i - \tilde{D}_i\mathbf{g}^*\|^2 \le \frac{2}{N_s}\sum_{i=1}^n \|\mathbf{m}_i - \tilde{\mathbf{m}}_i\|^2 + \frac{2}{N_s}\sum_{i=1}^n \|\tilde{\mathbf{m}}_i - \tilde{D}_i\mathbf{g}^*\|^2.$$

Now it suffices to focus on $N_s^{-1}\sum_{i=1}^n \|\mathbf{m}_i - \tilde{\mathbf{m}}_i\|^2$ and $N_s^{-1}\sum_{i=1}^n \|\tilde{\mathbf{m}}_i - \tilde{D}_i\mathbf{g}^*\|^2$.

Note that by (19) and (20),

$$\frac{1}{n}\sum_{i=1}^n\{\hat{\phi}_k(Z_{ik}) - \phi_k(Z_{ik})\}^2 = O_p\left\{\frac{K_A}{n} + \frac{1}{n^2}(\sum_{i=1}^n N_i^{-2})^2 + \rho_A^2\right\} = o_p(1), \quad k = 1,\dots,d. \tag{23}$$

Thus by Assumption 4 and Cauchy–Schwartz inequality,

$$\begin{aligned}
\frac{1}{N_s}\sum_{i=1}^n \|\mathbf{m}_i - \tilde{\mathbf{m}}_i\|^2 &= \frac{1}{N_s}\sum_{i=1}^n\sum_{j=1}^{N_i}\left[\sum_{k=1}^d \beta_k(T_{ij})\{\hat{\phi}_k(Z_{ik}) - \phi_k(Z_{ik})\}\right]^2 \\
&\le \frac{1}{N_s}\sum_{i=1}^n\sum_{j=1}^{N_i}\sum_{k=1}^d\{\hat{\phi}_k(Z_{ik}) - \phi_k(Z_{ik})\}^2 = \frac{1}{N_s}\sum_{k=1}^d\sum_{i=1}^n\left[N_i\{\hat{\phi}_k(Z_{ik}) - \phi_k(Z_{ik})\}^2\right] \\
&\le \frac{N_{max}}{N_s}\sum_{k=1}^d\sum_{i=1}^n\{\hat{\phi}_k(Z_{ik}) - \phi_k(Z_{ik})\}^2 = O_p\left[\frac{N_{max}}{\bar{N}}\left\{\frac{K_A}{n} + \frac{1}{n^2}(\sum_{i=1}^n N_i^{-2})^2 + \rho_A^2\right\}\right].
\end{aligned}$$

Similarly by (23) and Cauchy–Schwartz inequality,

$$\begin{aligned}
\frac{1}{N_s}\sum_{i=1}^n \|\tilde{\mathbf{m}}_i - \tilde{D}_i\mathbf{g}^*\|^2 &= \frac{1}{N_s}\sum_{i=1}^n\sum_{j=1}^{N_i}\left[\{\beta_0(T_{ij}) - \beta_0(T_{ij})^*\} + \sum_{k=1}^d\hat{\phi}_k(Z_{ik})\{\beta_k(T_{ij}) - \beta_k(T_{ij})^*\}\right]^2 \\
&\le \frac{1}{N_s}\sum_{i=1}^n\sum_{j=1}^{N_i}\left[\{\beta_0(T_{ij}) - \beta_0(T_{ij})^*\}^2 + \sum_{k=1}^d\hat{\phi}_k(Z_{ik})^2\{\beta_k(T_{ij}) - \beta_k(T_{ij})^*\}^2\right] \le \rho_C^2 + \frac{\rho_C^2}{N_s}\sum_{k=1}^d\sum_{i=1}^n\{N_i\hat{\phi}_k(Z_{ik})^2\} \\
&\le \rho_C^2 + \frac{\rho_C^2 N_{max}}{N_s}\sum_{k=1}^d\left[\sum_{i=1}^n\{\hat{\phi}_k(Z_{ik}) - \phi_k(Z_{ik})\}^2 + \sum_{i=1}^n\phi_k(Z_{ik})^2\right] = O_p\left(\frac{N_{max}}{\bar{N}}\rho_C^2\right).
\end{aligned}$$

Thus (22) is proved.

Stochastic error: The stochastic error here is attributed to the usual stochastic error in spline smoothing plus the error induced by $\hat{\phi}_k$ in estimating β_k in (11). We show below that the rate of the combined stochastic error is:

$$\|\check{\beta}_k(t) - \tilde{\beta}_k(t)\|^2 = O_p\left[\frac{K_C}{N_s} + \frac{\sum_{i=1}^n N_i^2}{N_s^2} + \frac{K_C N_{max} + (N_{max})^2}{\bar{N}^2}\left\{\frac{K_A}{n} + \frac{1}{n^2}(\sum_{i=1}^n N_i^{-2})^2 + \rho_A^2\right\}\right]. \tag{24}$$

By (21), Lemma 4, and Cauchy–Schwartz inequality, with probability approaching one we have

$$
\begin{aligned}
\|\breve{\beta}_k(t)-\tilde{\beta}_k(t)\|^2 &\asymp \|\hat{\mathbf{g}}_k-\tilde{\mathbf{g}}_k\|^2 \le \|\hat{\mathbf{g}}-\tilde{\mathbf{g}}\|^2 = \|(\frac{1}{N_s}\sum_{i=1}^n \tilde{\mathbf{D}}_i^\top \tilde{\mathbf{D}}_i)^{-1}\frac{1}{N_s}\sum_{i=1}^n \tilde{\mathbf{D}}_i^\top \boldsymbol{\delta}_i\|^2 \\
&\asymp \|\frac{1}{N_s}\sum_{i=1}^n \tilde{\mathbf{D}}_i^\top \boldsymbol{\delta}_i\|^2 \le 2\|\frac{1}{N_s}\sum_{i=1}^n \mathbf{D}_i^\top \boldsymbol{\delta}_i\|^2 + 2\|\frac{1}{N_s}\sum_{i=1}^n (\tilde{\mathbf{D}}_i-\mathbf{D}_i)^\top \boldsymbol{\delta}_i\|^2.
\end{aligned}
$$

Now we only need to focus on $\|N_s^{-1}\sum_{i=1}^n \mathbf{D}_i^\top \boldsymbol{\delta}_i\|^2$ and $\|N_s^{-1}\sum_{i=1}^n (\tilde{\mathbf{D}}_i-\mathbf{D}_i)^\top \boldsymbol{\delta}_i\|^2$.

By Assumption 1, 5, 6 and (21), for any $k=0,\ldots,d$, $E[\sum_{l=1}^{J_{k,C}}\{\sum_{j=1}^{N_i}\theta_{kl}(T_{ij})\delta_{ij}\}^2] = O(K_C N_i + N_i^2) = O(K_C N_{max} + N_{max}^2)$. Therefore

$$
\begin{aligned}
E\left(\|\frac{1}{N_s}\sum_{i=1}^n \mathbf{D}_i^\top \boldsymbol{\delta}_i\|^2\right) &= \frac{1}{N_s^2}\sum_{i=1}^n \left(E\left[\sum_{l=1}^{J_{0,C}}\left\{\sum_{j=1}^{N_i}\theta_{0,l}(T_{ij})\delta_{ij}\right\}^2\right]\right. \\
&\quad \left. + \sum_{k=1}^d E\{\phi_k(Z_{ik})^2\} E\left[\sum_{l=1}^{J_{k,C}}\left\{\sum_{j=1}^{N_i}\theta_{kl}(T_{ij})\delta_{ij}\right\}^2\right]\right) = O\left(\frac{K_C}{N_s}+\frac{\sum_{i=1}^n N_i^2}{N_s^2}\right),
\end{aligned}
$$

and thus by Markov inequality

$$
\|\frac{1}{N_s}\sum_{i=1}^n \mathbf{D}_i^\top \boldsymbol{\delta}_i\|^2 = O_p\left(\frac{K_C}{N_s}+\frac{\sum_{i=1}^n N_i^2}{N_s^2}\right).
$$

Moreover by Cauchy–Schwartz inequality,

$$
\begin{aligned}
\|\frac{1}{N_s}\sum_{i=1}^n (\tilde{\mathbf{D}}_i-\mathbf{D}_i)^\top \boldsymbol{\delta}_i\|^2 &= \frac{1}{N_s^2}\sum_{k=1}^d\sum_{l=1}^{J_{k,C}}\left[\sum_{i=1}^n\sum_{j=1}^{N_i}\{\hat{\phi}_k(Z_{ik})-\phi_k(Z_{ik})\}\theta_{kl}(T_{ij})\delta_{ij}\right]^2 \\
&\le \frac{n}{N_s^2}\sum_{k=1}^d\sum_{l=1}^{J_{k,C}}\sum_{i=1}^n\left[\{\hat{\phi}_k(Z_{ik})-\phi_k(Z_{ik})\}^2\{\sum_{j=1}^{N_i}\theta_{kl}(T_{ij})\delta_{ij}\}^2\right] \\
&= \frac{n}{N_s^2}\sum_{k=1}^d\sum_{i=1}^n\left(\{\hat{\phi}_k(Z_{ik})-\phi_k(Z_{ik})\}^2\left[\sum_{l=1}^{J_{k,C}}\{\sum_{j=1}^{N_i}\theta_{kl}(T_{ij})\delta_{ij}\}^2\right]\right) \\
&\asymp \frac{n(K_C N_{max}+N_{max}^2)}{N_s^2}\sum_{k=1}^d\sum_{i=1}^n\{\hat{\phi}_k(Z_{ik})-\phi_k(Z_{ik})\}^2 \\
&= O_p\left[\frac{K_C N_{max}+(N_{max})^2}{\bar{N}^2}\left\{\frac{K_A}{n}+\frac{1}{n^2}(\sum_{i=1}^n N_i^{-2})^2+\rho_A^2\right\}\right].
\end{aligned}
$$

Now the proof for (24) is complete. Therefore Theorem 2 holds by (22) and (24). □

REFERENCES

BRUMBACK, B. & RICE, J. (1998). Smoothing spline models for the analysis of nested and crossed samples of curves. *Journal of the American Statistical Association* , 961–976.

BUJA, A., HASTIE, T. & TIBSHIRANI, R. (1989). Linear smoothers and additive models. *The Annals of Statistics* **17**, 453–555.

DE BOOR, C. (2001). *A Practical Guide to Splines*, vol. 27. Springer Verlag.

DRAGOMIR, S. S., CERONE, P. & SOFO, A. (2000). Some remarks on the trapezoid rule in numerical integration. *Indian Journal of Pure and Applid Mathematics* **31**, 475–494.

FAN, J. & ZHANG, W. (1999). Statistical estimation in varying coefficient models. *The annals of Statistics* **27**, 1491–1518.

HOOVER, D., RICE, J., WU, C. & YANG, L. (1998). Nonparametric smoothing estimates of time-varying coefficient models with longitudinal data. *Biometrika* **85**, 809–822.

HUANG, J., WU, C. & ZHOU, L. (2002). Varying-coefficient models and basis function approximations for the analysis of repeated measurements. *Biometrika* **89**, 111–128.

HUANG, J., WU, C. & ZHOU, L. (2004). Polynomial spline estimation and inference for varying coefficient models with longitudinal data. *Statistica Sinica* **14**, 763–788.

IHLER, A., HUTCHINS, J. & SMYTH, P. (2006). Adaptive event detection with time-varying poisson processes. In *Proceedings of the 12th ACM SIGKDD international conference on knowledge discovery and data mining*. ACM.

JONES, M. & BALAKRISHNAN, N. (2002). How are moments and moments of spacings related to distribution functions? *Journal of statistical planning and inference* **103**, 377–390.

LINTON, O. B. & NIELSEN, J. P. (1995). A kernel method of estimating structured nonparametric regression based on marginal integration. *Biometrika* **82**, 93–100.

MAMMEN, E., LINTON, O. B. & NIELSEN, J. P. (1999). The existence and asymptotic properties of a backfitting projection algorithm under weak conditions. *The Annals of Statistics* **27**, 1443–1490.

SCHUMAKER, L. (2007). *Spline Functions: Basic Theory*. Cambridge University Press.

STONE, C. (1985). Additive regression and other nonparametric models. *The Annals of Statistics* , 689–705.

WANG, L. & YANG, L. (2007). Spline-backfitted kernel smoothing of nonlinear additive autoregression model. *The Annals of Statistics* **35**, 2474–2503.

Table 1: Basic summary statistics of number of knots selected by AIC and BIC. The mean, median and standard deviation of the selected knots are shown. The optimal number of knots resulted from the simulation is shown in the last column.

ϕ	AIC			BIC			OPTIMAL
	mean	med	std	mean	med	std	
$K_{1,A}$	5.14	5	1.17	4.13	4	0.86	5
$K_{2,A}$	2.42	1	2.11	1.11	1	0.45	2

β	AIC			BIC			AIC			BIC			OPTIMAL
	mean	med	std	mean	med	std	mean	med	std	mean	med	std	
$K_{0,C}$	2.38	2	1.00	2	2	0	2.38	2	1.00	2	2	0	2
$K_{1,C}$	1.63	1	0.78	1.08	1	0.28	1.63	1	0.78	1.08	1	0.28	1
$K_{2,C}$	1.63	1	0.78	1.08	1	0.30	1.63	1	0.78	1.08	1	0.30	1

Table 2: MISE values of additive component function estimates and coefficient function estimates with the number of knots selected by AIC and BIC. The MISE corresponding to the optimal number of knots is shown in the last column.

Component functions (10^{-2})					
	AIC		BIC		OPTIMAL
MISE(ϕ_1)	0.5176		0.5174		0.5171
MISE(ϕ_2)	1.2545		1.2532		1.2528
MISE(ϕ)=MISE(ϕ_1)+MISE(ϕ_2)	1.7721		1.7711		1.7699
Coefficient functions (10^{-2})					
	AIC	BIC	AIC	BIC	OPTIMAL
MISE(β_0)	4.1412	4.1252	4.1412	4.1252	4.1323
MISE(β_1)	1.0632	0.9954	1.0625	0.9949	0.9734
MISE(β_2)	0.4499	0.4236	0.4498	0.4235	0.4148
MISE(β)=MISE(β_0)+MISE(β_1)+MISE(β_2)	5.6529	5.5427	5.6520	5.5421	5.5205
MISE(ϕ,β)=MISE(ϕ)+ MISE(β)	7.4250	7.3148	7.4231	7.3132	7.2904

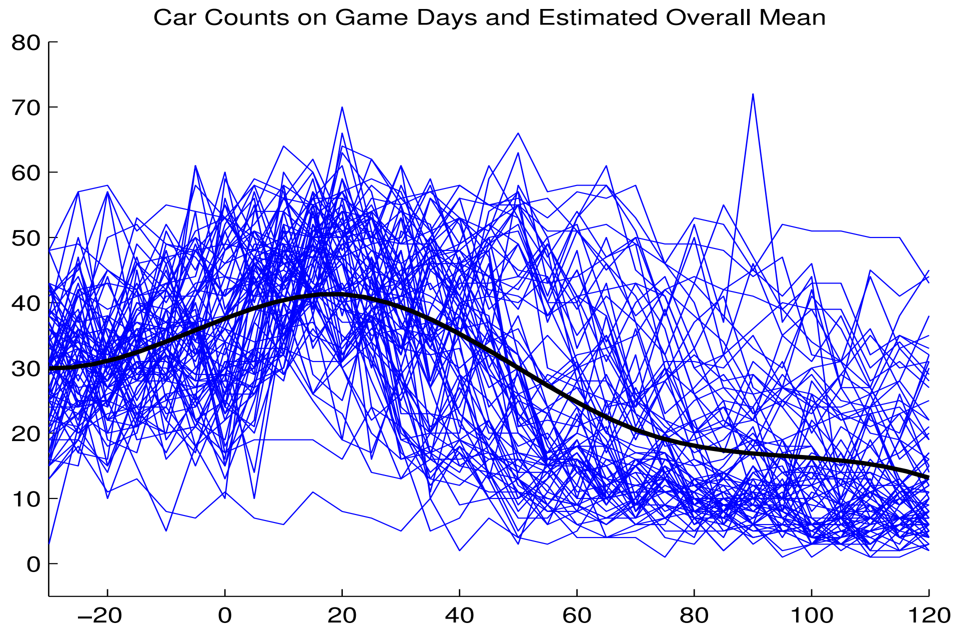

Fig. 1: Car counts on 78 game days over 30 minutes before the end of a game until two hours after the game. The bold line is the estimated $\beta_0(t)$.

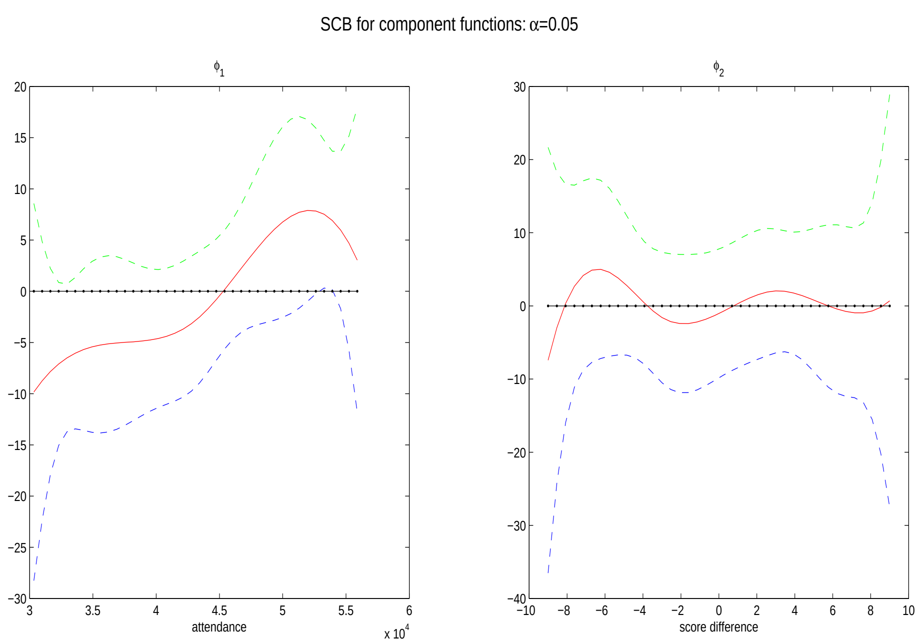

Fig. 2: Fitted additive component functions and 95% simultaneous confidence bands using Scheffé's method. In each panel, the solid line corresponds to the function estimate and the dashed lines represent the upper and lower bounds. A reference dash-and-dot line is also given for constant 0.

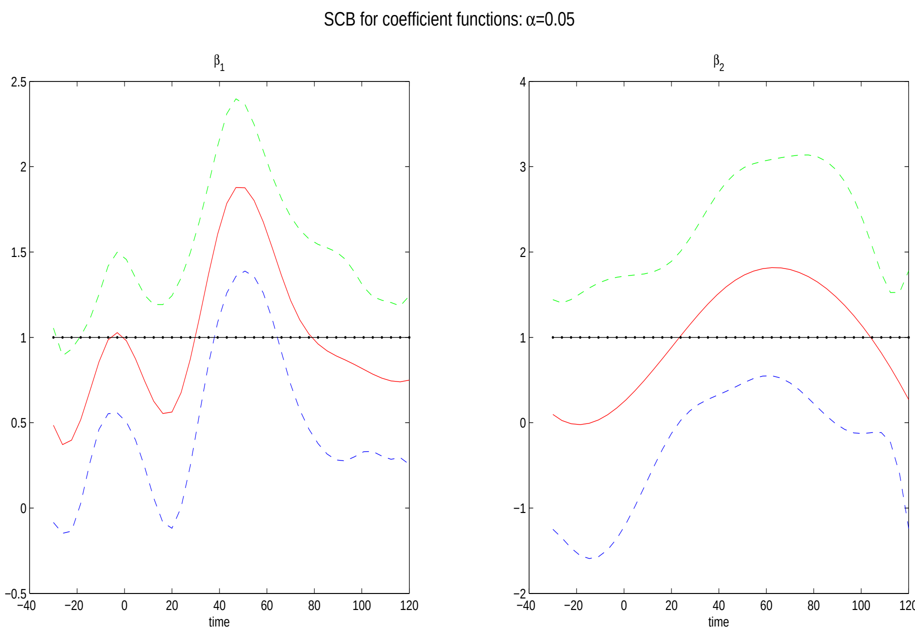

Fig. 3: Fitted coefficient functions and 95% bootstrap simultaneous confidence bands. In each panel, the solid line corresponds to the function estimate and the dashed lines represent the upper and lower bounds. A reference dash-and-dot line is also given for constant 1.