Probability in the Engineering and Informational Sciences (2024), **38:**1 82–99
doi:10.1017/S0269964822000511

CAMBRIDGE
UNIVERSITY PRESS

RESEARCH ARTICLE

Incorporating covariate into mean and covariance function estimation of functional data under a general weighing scheme

Xingyu Yan[1], Hao Wang[2], Hong Sun[3] and Peng Zhao[4]

[1] School of Mathematics and Statistics, Jiangsu Normal University, Xuzhou 221116, China
[2] School of Mathematics and Statistics, Anhui Normal University, Wuhu 241000, China
[3] School of Mathematics and Statistics, Lanzhou University, Lanzhou 730000, China
[4] School of Mathematics and Statistics and RIMS, Jiangsu Provincial Key Laboratory of Educational Big Data Science and Engineering, Jiangsu Normal University, Xuzhou 221116, China. E-mail: zhaop@jsnu.edu.cn.

Keywords: Functional data, Local linear smoothing, Uniform convergence, Weighing schemes

Abstract
This paper develops the estimation method of mean and covariance functions of functional data with additional covariate information. With the strength of both local linear smoothing modeling and general weighing scheme, we are able to explicitly characterize the mean and covariance functions with incorporating covariate for irregularly spaced and sparsely observed longitudinal data, as typically encountered in engineering technology or biomedical studies, as well as for functional data which are densely measured. Theoretically, we establish the uniform convergence rates of the estimators in the general weighing scheme. Monte Carlo simulation is conducted to investigate the finite-sample performance of the proposed approach. Two applications including the children growth data and white matter tract dataset obtained from Alzheimer's Disease Neuroimaging Initiative study are also provided.

1. Introduction

The estimation of mean and covariance functions plays a fundamental role in the analysis of functional data. How to appropriately model these functions is fascinating but challenging and has drawn much attention from statisticians in the past several decades. Influential works in this area include, but are not limited to, Rice and Silverman [19], James *et al.* [9], Yao *et al.* [23], Li and Hsing [12], Peng and Paul [17], Cai and Yuan [2], Ogden and Greene [16], Chen and Müller [4], Xiao *et al.* [22], Zhou *et al.* [26], Meister [15] and the references therein. Well-known monographs by Ramsay and Silverman [18], Ferraty and Vieu [7] and Kokoszka and Reimherr [11] provided comprehensive discussions on the methods and applications.

Nowadays, an important question is how to estimate mean and covariance function with available covariate information. Such covariate information is commonly encountered in biomedical studies and informational sciences, which requires us to use the additional covariates to model the trajectory realistically. They have been receiving increasing attention recently [3,20,21]. Specifically, Chiou and Wang [6] discussed the influence of covariates on a sample of response curves through a semiparametric model under the framework of dense functional data. Jiang and Wang [10] described a general approach incorporating a covariate effect to model the mean and covariance function for sparse longitudinal data. Liebl [14] considered inference problem for the mean and covariance functions of covariate adjusted functional data. Zhang *et al.* [25] proposed a new functional regression model with covariate-dependent mean and covariance structures to analyze the Avon Longitudinal Study of Parents and Children datasets.

However, the aforementioned works only address the sparse or dense functional data. Little is known on how to incorporate covariate information in modeling for a general type of functional data, so our goal in this paper is to provide a general weighing approach to incorporate the covariate information that is applicable to both dense and sparse functional data. For the sake of simplicity and convenience, throughout this article, we consider the case of a one-dimensional covariate Z_i for $i = 1, \ldots, n$. Let Y_{ij} be the jth observation of the random function $X_i(T_{ij}, Z_i)$, made at a discrete time points $T_{ij} \in [0, 1]$ with a covariate $Z_i \in [0, 1]$ and independent identically distributed measurement errors ε_{ij} for $j = 1, \ldots, m_i$. Thus, the observed data are often written as

$$Y_{ij} = X_i(T_{ij}, Z_i) + \varepsilon_{ij} \quad \text{for } i = 1, \ldots, n, \tag{1}$$

where the sampling locations T_{ij} and Z_i are independently drawn from a distribution of random variables T and Z with density function $f(\cdot)$ and $g(\cdot)$ on bounded support $[0, 1]$, respectively. One is generally interested in estimating mean function $E\{X_i(t, z)\} = \mu(t, z)$ and covariance function $\text{cov}\{X_i(t, z), X_i(s, z)\} = K(t, s, z)$ based on the observation of Y_{ij} for $j = 1, \ldots, m_i, i = 1, \ldots, n$.

To broaden the applicability of the aforementioned model, we propose to estimate the mean and covariance functions not only by allowing the functions to depend on the additional scalar covariate but also in the framework of the general weighted local linear smoothing. We further carefully demonstrate the uniform convergence rate for the proposed estimators. The derived convergence rates of mean and covariance functions provide an essential theoretical result for the future research, such as functional principal component analysis and functional regression issues.

The rest of the article is organized as follows. We would introduce the proposed estimation procedure in Section 2.1 and present theoretical results in Section 2.2. Regularity conditions and technical proofs are delegated to the Appendix. Simulation studies are conducted to verify the theoretical results. The approach is applied to analyze the growth curves of children dataset and produce meaningful and interesting results. Both are shown in Section 3. The concluding remarks are given in Section 4.

2. Methodology

2.1. Estimation procedure

This section describes the method of estimation of mean and covariance functions. To obtain mean function $\mu(t, z)$ and covariance function $K(s, t, z)$, we apply the weighted local linear smoothing method [24]. Specifically, the weight ω_i is attached to each observation for the ith subject such that $\sum_{i=1}^{n} m_i \omega_i = 1$, we define the weighted local linear smoother for $\mu(t, z)$ by minimizing

$$\begin{aligned} \widehat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} \sum_{i=1}^{n} \omega_i \sum_{j=1}^{m_i} K_{h_{\mu t}}(T_{ij} - t) K_{h_{\mu z}}(Z_i - z) \\ \times \{Y_{ij} - \beta_0 - \beta_1(T_{ij} - t) - \beta_2(Z_i - z)\}^2, \end{aligned} \tag{2}$$

with respect to $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2)^{\mathrm{T}}$. The estimate of $\mu(t, z)$ is then $\widehat{\mu}(t, z) = \widehat{\beta}_0$. Once the $\widehat{\mu}(\cdot, \cdot)$ is obtained, we are then ready to estimate the covariance function $G(t, s, z)$. Let $G_{ijk} = \{Y_{ij} - \widehat{\mu}(T_{ij}, Z_i)\}\{Y_{ik} - \widehat{\mu}(T_{ik}, Z_i)\}$ be the input data, and the weight ν_i is attached to each G_{ijk} for the ith subject such that $\sum_{i=1}^{n} m_i(m_i - 1)\nu_i = 1$. The weighted local linear smoother for the covariance function $G(t, s, z)$ is $\widehat{G}(t, s, z) = \widehat{\beta}_0$ where $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2, \beta_3)^{\mathrm{T}}$,

$$\begin{aligned} \widehat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} \sum_{i=1}^{n} \nu_i \sum_{1 \le j \ne k \le m_i} K_{h_{Gt}}(T_{ij} - t) K_{h_{Gt}}(T_{ik} - s) K_{h_{Gz}}(Z_i - z) \\ \times \{G_{ijk} - \beta_0 - \beta_1(T_{ij} - t) - \beta_2(T_{ik} - s) + \beta_3(Z_i - z)\}^2. \end{aligned} \tag{3}$$

https://doi.org/10.1017/S0269964822000511 Published online by Cambridge University Press

The bandwidths $h_{\mu t}$ and $h_{\mu z}$ for the estimated mean function are chosen via the leave-one-curve-out cross validation and the bandwidths h_{Gt} and h_{Gz} of the covariance function estimators are chosen via a 10-fold cross-validation procedure to save computing time throughout this article. More details about choices of ω_i and ν_i of (2) and (3) can be found in Huang *et al.* [8] and Zhang and Wang [24]. In general, when the observation is sparse, assignment $\omega_i = 1/\sum_{i=1}^n m_i$ along with $\nu_i = 1/\sum_{i=1}^n m_i(m_i-1)$ leads to the same weight to each observation scheme (OBS); often in the dense observation, assignment $\omega_i = 1/(nm_i)$ along with $\nu_i = 1/nm_i(m_i-1)$ leads to the same weight to each subject scheme (SUBJ). Although OBS and SUBJ are the most commonly used schemes, for some cases, Zhang and Wang [24] suggested using a mixture of the OBS and SUBJ schemes (MIX). That is, $\omega_i = \alpha_1/(\sum_{i=1}^n m_i) + (1-\alpha_1)/(nm_i)$ and $\nu_i = \alpha_2/\{\sum_{i=1}^n m_i(m_i-1)\} + (1-\alpha_2)/\{nm_i(m_i-1)\}$ for some $0 \le \alpha_1, \alpha_2 \le 1$. Moreover, they showed that the MIX scheme was likely to achieve a better performance. In our simulation part, we will see that this weighing scheme is also likely to be better than OBS and SUBJ with incorporating covariate information.

2.2. Asymptotic results

In this section, we establish the uniform convergence rates for the estimates. Our results cover the case of sparse designs, where the number of design points m_i is bounded, and the case of dense designs, where $m_i \to \infty$. The additional assumptions and proof of these results are relegated to Appendix. Based on the general weighing scheme, we provide the uniform convergence rates of mean and covariance function estimators in Theorems 1 and 2, respectively.

Theorem 1. *Under conditions (A1)–(A5) and (B1)–(B4) in the Appendix,*

$$\sup_{t,z\in[0,1]} |\widehat{\mu}(t,z) - \mu(t,z)| = O(h_{\mu t}^2 + h_{\mu z}^2 + h_{\mu t}h_{\mu z} + \delta_{n1}) \quad a.s.,$$

where

$$\delta_{n1} = \left[\log(n)\left\{\frac{\sum_{i=1}^n m_i\omega_i^2}{h_{\mu t}h_{\mu z}} + \frac{\sum_{i=1}^n m_i(m_i-1)\omega_i^2}{h_{\mu z}}\right\}\right]^{1/2}.$$

Theorem 2. *Under conditions (A1)–(A6) and (C1)–(C4) in the Appendix,*

$$\sup_{s,t,z\in[0,1]} |\widehat{G}(t,s,z) - G(t,s,z)| = O(h_{\mu t}^2 + h_{\mu z}^2 + h_{\mu t}h_{\mu z} + \delta_{n1} + h_{Gt}^2 + h_{Gz}^2 + h_{Gt}h_{Gz} + \delta_{n2}) \quad a.s.,$$

where

$$\delta_{n2} = \left[\log(n)\left\{\frac{\sum_{i=1}^n m_i(m_i-1)\nu_i^2}{h_{Gt}^2 h_{Gz}} + \frac{\sum_{i=1}^n m_i(m_i-1)(m_i-2)\nu_i^2}{h_{Gt}h_{Gz}} + \frac{\sum_{i=1}^n m_i(m_i-1)(m_i-2)(m_i-3)\nu_i^2}{h_{Gz}}\right\}\right]^{1/2}.$$

The results in Theorems 1 and 2 are natural extensions of Theorems 5.1 and 5.2 of Zhang and Wang [24] for the case of additional covariate information available, respectively. Similar conclusion for the OBS scheme was established by Chen and Müller [4].

https://doi.org/10.1017/S0269964822000511 Published online by Cambridge University Press

3. Numerical studies

3.1. *Simulation results*

To assess which weighing scheme is better in the case where covariate information is incorporated, we evaluate the finite-sample numerical performances of the most commonly used OBS and SUBJ schemes, together with the weighing schemes based on their mixture as defined in Section 2. The simulation is set as follows: for each subject i, the covariate-adjusted true trajectories are generated as $Y_i(t) = \mu(t,z) + \sum_{k=1}^{2} \xi_{ik}\psi_k(t,z) + \epsilon_i(t)$, where the covariate z is generated from $U(0,1)$, and mean function is $\mu(t,z) = (1+z)(t+\sin(t))$. The principal component scores are $\xi_{ik} \sim \mathcal{N}\{0, \lambda_k(z)\}$ for $k = 1,2$ and the eigenfunctions are $\psi_1(t,z) = \sqrt{2}\sin(\pi(t+z/2))$ and $\psi_2(t,z) = -\sqrt{2}\cos(\pi(t+z/2))$, respectively. We set $\{\lambda_1(z), \lambda_2(z)\} = (z/9, z/36)$ and $\epsilon_i(t) \sim \mathcal{N}(0, 0.2)$.

For more generality in practice, we provide the simulation studies under different number of observation points setting, which include sparse, dense and mixture of them shown as follows:

Case 1: m_i is sampled with equal probability from $\{2, \ldots, 6\}$.
Case 2: m_i is sampled from a discrete uniform distribution on the interval $[n/6, n/3]$.
Case 3: $m_i = n/4$ with probability $1/2$, and follows Case 1 with probability $1/2$.

We compare the performance of three different weighing scheme using the corresponding mean integrated square errors (MISE) for the $\widehat{\mu}$ defined as below

$$\mathrm{MISE}(\widehat{\mu}, h_\mu) = \frac{1}{n}\sum_{i=1}^{n} \int \{\widehat{\mu}(t, z_i) - \mu(t, z_i)\}^2 \, dt.$$

To evaluate the covariance estimators, we assume that the mean function μ is known so that the covariance estimation is disentangled from the mean estimation and define the MISE of $\widehat{G}$ as follows:

$$\mathrm{MISE}(\widehat{G}, h_G) = \frac{1}{n}\sum_{i=1}^{n} \iint \{\widehat{G}(s, t, z_i) - G(s, t, z_i)\}^2 \, ds \, dt.$$

In each case, the sample size is 50, 100 and 200, respectively. We present the mean and standard deviation (in parentheses) of the MISE over 200 repetitions for each case. Hereafter, estimators with the subscript "obs," "subj" and "mix" represent the OBS, SUBJ and MIX estimators, respectively.

In Table 1, similar to the case where covariate information is not incorporated, the OBS scheme outperforms the SUBJ scheme in Case 1 and Case 3, while the SUBJ scheme is superior in Case 2. Moreover, as expected, the performance improves as the number of subjects increases. With respect to MISE of covariance function from Table 2, we have similar results to their counterparts in Table 1, which conforms with the case of ignoring the covariate information.

As shown in Tables 1 and 2, both $\widehat{\mu}_{\mathrm{mix}}$ and $\widehat{G}_{\mathrm{mix}}$ performed equally well or better than the counterparts of OBS and SUBJ estimators. This provides the numerical evidence for the benefit of using a mixture of the OBS and SUBJ schemes as discussed in Section 2.

3.2. *Application to children growth data*

We apply the proposed methodology to analyze a real medical study dataset about the growth curves of children, which is publicly available on https://content.sph.harvard.edu/fitzmaur/ala/fev1.txt. This dataset was presented by Brunekreef *et al.* [1] and was commonly used as an example for a longitudinal study designed to characterize lung growth as measured by changes in pulmonary function in children and adolescents.

The dataset consists of 300 children, with a minimum of 1 and a maximum of 12 observations over time. The following four variables are included: age (ranging from 6 to 18 years old), height, FEV_1 (forced expiratory volume in one second) and initial height, which is obtained from a randomly selected

https://doi.org/10.1017/S0269964822000511 Published online by Cambridge University Press

Table 1. *MISE of the estimator of mean function, where $\widehat{\mu}_{\rm obs}$ is the OBS scheme, $\widehat{\mu}_{\rm subj}$ refers to SUBJ scheme, $\widehat{\mu}_{\rm mix}$ denotes the mixture of the above two weighing scheme ($\alpha_1 = 1/2$).*

n	Method	Case 1	Case 2	Case 3
50	$\widehat{\mu}_{\rm obs}$	0.0563 (0.0185)	0.0450 (0.0163)	0.0495 (0.0164)
	$\widehat{\mu}_{\rm subj}$	0.0567 (0.0191)	0.0450 (0.0161)	0.0519 (0.0178)
	$\widehat{\mu}_{\rm mix}$	0.0555 (0.0186)	0.0449 (0.0162)	0.0495 (0.0166)
100	$\widehat{\mu}_{\rm obs}$	0.0465 (0.0080)	0.0392 (0.0085)	0.0489 (0.0124)
	$\widehat{\mu}_{\rm subj}$	0.0475 (0.0092)	0.0387 (0.0081)	0.0419 (0.0104)
	$\widehat{\mu}_{\rm mix}$	0.0460 (0.0084)	0.0388 (0.0082)	0.0426 (0.0105)
200	$\widehat{\mu}_{\rm obs}$	0.0376 (0.0124)	0.0325 (0.0112)	0.0320 (0.0108)
	$\widehat{\mu}_{\rm subj}$	0.0384 (0.0127)	0.0323 (0.0110)	0.0373 (0.0079)
	$\widehat{\mu}_{\rm mix}$	0.0371 (0.0123)	0.0323 (0.0110)	0.0366 (0.0093)

The corresponding standard errors are in parentheses.

Table 2. *MISE of the estimator of covariance function, where $\widehat{G}_{\rm obs}$ is the OBS scheme, $\widehat{G}_{\rm subj}$ refers to SUBJ scheme, $\widehat{G}_{\rm mix}$ denotes the mixture of the above two weighing scheme ($\alpha_2 = 1/2$).*

n	Method	Case 1	Case 2	Case 3
50	$\widehat{G}_{\rm obs}$	0.0316 (0.0081)	0.0225 (0.0074)	0.0279 (0.0063)
	$\widehat{G}_{\rm subj}$	0.0318 (0.0085)	0.0224 (0.0070)	0.0296 (0.0075)
	$\widehat{G}_{\rm mix}$	0.0304 (0.0079)	0.0223 (0.0072)	0.0276 (0.0063)
100	$\widehat{G}_{\rm obs}$	0.0229 (0.0053)	0.0186 (0.0052)	0.0212 (0.0042)
	$\widehat{G}_{\rm subj}$	0.0236 (0.0059)	0.0183 (0.0045)	0.0215 (0.0063)
	$\widehat{G}_{\rm mix}$	0.0221 (0.0051)	0.0183 (0.0048)	0.0195 (0.0042)
200	$\widehat{G}_{\rm obs}$	0.0182 (0.0041)	0.0167 (0.0037)	0.0175 (0.0036)
	$\widehat{G}_{\rm subj}$	0.0195 (0.0041)	0.0163 (0.0034)	0.0186 (0.0034)
	$\widehat{G}_{\rm mix}$	0.0179 (0.0036)	0.0164 (0.0035)	0.0167 (0.0025)

The corresponding standard errors are in parentheses.

subset of the participants living in Topeka, Kansas in USA. What we are interested in is to explore the influence of children's initial height on the shape changes of the mean functions of height's curve. Therefore, the adjusted covariate is initial height and we exclude the children whose observations are less than two times, such that the sample size $n = 258$. Since the range of children's age is sparse (the maximum is 12 observations), we adopt OBS weighing scheme to model the height curve.

To compare the developed approach with the approach that does not incorporate the initial height information, we also display the fitted curves by using the conventional OBS weighing scheme estimation procedure [24]. We randomly select four estimated mean height curves shown in Figure 1. The overall trend of the height curve as shown is incremental with a decreasing rate of growth, which is consistent with common knowledge. Specifically, the left top plot shows that our estimation procedure is closer to the observed height's curve compared with the approach without the initial height adjusted. According to other plots, we have similar results. This demonstrates that the data support the simple covariate-adjusted approach when modeling curves. The differences between the two estimated methods may somewhat claim that the proposed method is more efficient numerically than conventional approaches without incorporating covariate information. Such an extension in this paper could be valuable for doctors who study how the children's height changes with their age when covariate information is available.

https://doi.org/10.1017/S0269964822000511 Published online by Cambridge University Press

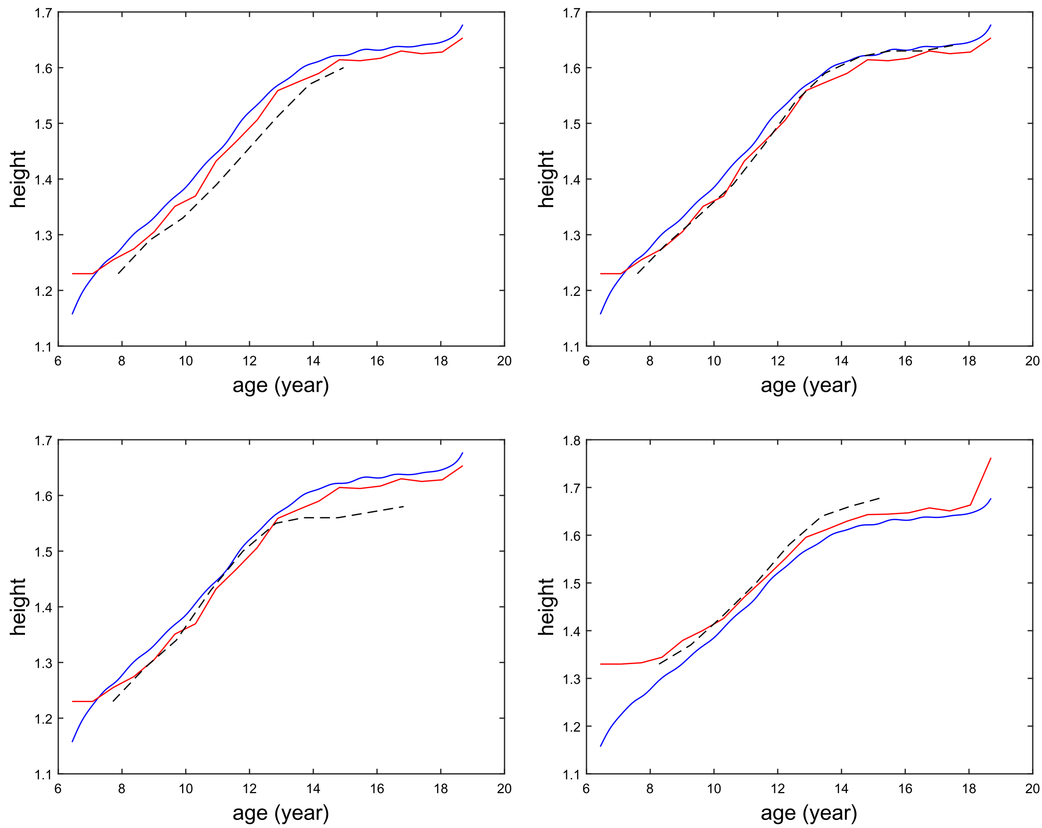

Figure 1. *Plots of an estimated mean function of children's height. The black dashed lines are the observed height's curve. The blue solid lines are the estimated mean function of height by Zhang and Wang [24] and the red solid lines denote the curve estimated by our estimation procedure.*

3.3. Application to diffusion tensor imaging study

The proposed method is also applied to a real data set of diffusion tensor imaging (DTI). In diffusion data analysis, fractional anisotropy (FA), which quantifies the directional strength of white matter tract structure, at a particular location, is one of the most used measures and has been widely applied to statistical analyses in many imaging studies, see, such as Zhu *et al.* [27], Li *et al.* [13] and reference therein. The data can be downloaded from the ADNI publicly available database (https://adni.loni.usc.edu/) or http://www.nitrc.org/projects/fadtts/.

We are interested in delineating the trend of the variability of these functional FA with a set of available covariates of interest, such as age. In particular, we use 64 healthy infants from the neonatal project on early brain development. The gestational ages of these infants range from 262 to 433 days, and their mean gestational age is 298 days with standard deviation 17.6 days. The dataset was preprocessed by Zhu *et al.* [27]. Finally, we fit model (1) to the FA values from 64 subjects at 75 grid points along the genu tract of the corpus callosum, in which Z_i = Age. Since the data are dense, SUBJ is adopted.

Figure 2 presents the three randomly selected estimated FA trajectories with age information adjusted, as well as the traditional non-information adjusted estimation curves. It also shows that the proposed method is more efficient. Moreover, the results confirm that neonatal microstructural development of FA correlates with age. Such findings are consistent with those of previous works. This shows that the proposed estimate is more reasonable in describing the true characteristic of the mean function.

https://doi.org/10.1017/S0269964822000511 Published online by Cambridge University Press

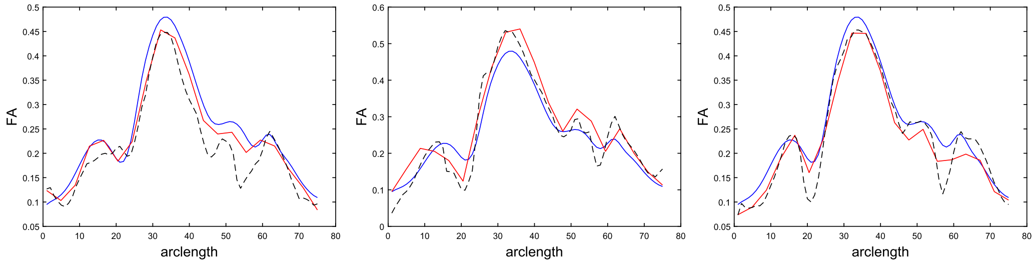

Figure 2. *Plots of the estimated mean function of functional FA. The black dashed lines are the observed FA values. The blue solid lines are the estimated mean function of FA without age adjusted and the red solid lines denote the trajectory estimated by our estimation procedure.*

4. Discussion

In this paper, we focused on covariate adjusted local linear smoothers when estimating the mean and covariance functions. This is an important extension for the study of functional mean-covariance model, because we incorporate the available covariate information that may improve the accuracy of estimation. Moreover, we are particularly interested in the framework of general weighing scheme which incorporates the commonly used OBS and SUBJ schemes. The carefully derived convergence rates in the framework of general weighing scheme expanded the results that ignore the covariate information. Numerical performances of OBS and SUBJ schemes are also systematically compared.

It is also of great interest to establish the asymptotic distribution and optimal convergence rate of $\widehat{\mu}(t, z)$ and $\widehat{G}(t, z)$ under the general weighing framework, which we left for future work. Furthermore, the general weighing framework may be used in functional data regression, classification, clustering, etc., and hence the theoretical results here could be extended to those cases as well. This will also be pursued in future work.

Acknowledgments. We thank two reviewers, an associate editor and the editor, for their most helpful comments that lead to substantial improvements in the paper. The work is supported by National Natural Science Foundation of China (project number: 12101270, 11871252, 11971034). Data used in the preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in the analysis or writing of this report. A complete listing of ADNI investigators can be found at: http://adni.loni.usc.edu/wpcontent/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf.

Conflict of interest. The authors have not disclosed any competing interests.

References

[1] Brunekreef, B., Janssen, N.A., de Hartog, J., Harssema, H., Knape, M., & van Vliet, P. (1997). Air pollution from truck traffic and lung function in children living near motorways. *Epidemiology* 8(3): 298–303.
[2] Cai, T.T. & Yuan, M. (2011). Optimal estimation of the mean function based on discretely sampled functional data: Phase transition. *The Annals of Statistics* 39(5): 2330–2355.
[3] Cardot, H. & Sarda, P. (2008). Varying-coefficient functional linear regression models. *Communications in Statistics -– Theory and Methods* 37(20): 3186–3203.
[4] Chen, K.H. & Müller, H.-G. (2012). Modeling repeated functional observations. *Journal of the American Statistical Association* 107(500): 1599–1609.
[5] Chen, K.H., Delicado, P., & Müller, H.-G. (2017). Modelling function-valued stochastic processes, with applications to fertility dynamics. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 79(1): 177–196.
[6] Chiou J.-M., Müller, H.-G., & Wang, J.-L. (2003). Functional quasi-likelihood regression models with smooth random effects. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 65(2): 405–423.
[7] Ferraty, F. & Vieu, P. (2006). *Nonparametric functional data analysis: Theory and practice*. New York: Springer.
[8] Huang, J.H., Wu, C.O., & Zhou, L. (2002). Varying-coefficient models and basis function approximations for the analysis of repeated measurements. *Biometrika* 89(1): 111–128.
[9] James, G.M., Hastie, T.J., & Sugar, C.A. (2000). Principal component models for sparse functional data. *Biometrika* 87(3): 587–602.

[10] Jiang, C.-R. & Wang, J.-L. (2010). Covariate adjusted functional principal components analysis for longitudinal data. *The Annals of Statistics* 38(2): 1194–1226.
[11] Kokoszka, P. & Reimherr, M. (2017). *Introduction to functional data analysis*. New York: Chapman and Hall/CRC.
[12] Li, Y.H. & Hsing, T. (2010). Uniform convergence rates for nonparametric regression and principal component analysis in functional/longitudinal data. *The Annals of Statistics* 38(6): 3321–3351.
[13] Li, J., Huang, C., Hongtu, Z., & Initiative, A.D.N. (2017). A functional varying-coefficient single-index model for functional response data. *Journal of the American Statistical Association* 112(519): 1169–1181.
[14] Liebl, D. (2019). Inference for sparse and dense functional data with covariate adjustments. *Journal of Multivariate Analysis* 170: 315–335.
[15] Meister, A. (2016). Optimal classification and nonparametric regression for functional data. *Bernoulli* 22(3): 1729–1744.
[16] Ogden, R.T. & Greene, E. (2010). Wavelet modeling of functional random effects with application to human vision data. *Journal of Statistical Planning and Inference* 140(12): 3797–3808.
[17] Peng, J. & Paul, D. (2009). A geometric approach to maximum likelihood estimation of the functional principal components from sparse longitudinal data. *Journal of Computational and Graphical Statistics* 18(4): 995–1015.
[18] Ramsay, J.O. & Silverman, B.W. (2005). *Functional data analysis*. New York: Springer.
[19] Rice, J.A. & Silverman, B.W. (1991). Estimating the mean and covariance structure nonparametrically when the data are curves. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 53(1): 233–243.
[20] Stallrich, J., Islam, M.N., Staicu, A.-M., Crouch, D., Pan, L., & Huang, H. (2020). Optimal EMG placement for a robotic prosthesis controller with sequential, adaptive functional estimation (SAFE). *The Annals of Applied Statistics* 14(3): 1164–1181.
[21] Wu, Y.C., Fan, J.Q., & Müller, H.-G. (2010). Varying-coefficient functional linear regression. *Bernoulli* 16(3): 730–758.
[22] Xiao, L., Li, Y.X., Ruppert, D. (2013). Fast bivariate p-splines: the sandwich smoother. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 75(3): 577–599.
[23] Yao, F., Müller, H.-G., & Wang, J.-L. (2005). Functional data analysis for sparse longitudinal data. *Journal of the American Statistical Association* 100(470): 577–590.
[24] Zhang, X.K. & Wang, J.-L. (2016). From sparse to dense functional data and beyond. *The Annals of Statistics* 44(5): 2281–2321.
[25] Zhang, C., Lin, H., Liu, L., Liu, J., & Li, Y. (2022). Functional data analysis with covariate-dependent mean and covariance structures. *Biometrics*. doi: doi.org/10.1111/biom.13744.
[26] Zhou, L., Lin, H.Z., & Liang, H. (2018). Efficient estimation of the nonparametric mean and covariance functions for longitudinal and sparse functional data. *Journal of the American Statistical Association* 113(524): 1550–1564.
[27] Zhu, H.T., Li, R.Z., & Kong, L.L. (2012). Multivariate varying coefficient model for functional responses. *Annals of Statistics* 40(5): 2634–2666.

Appendix

Conditions

Similar to the conditions in Chen and Müller [4], Chen *et al.* [5] and Zhang and Wang [24], the following conditions are imposed to facilitate the proofs.

Conditions for kernel function and time points

(A1) $K(\cdot)$ is a symmetric probability density function on $[-1, 1]$ and

$$\sigma_K^2 = \int u^2 K(u)\, du < \infty, \quad \|K\|^2 = \int K(u)^2\, du < \infty$$

(A2) $K(\cdot)$ is Lipschitz continuous and there exists $0 < L < \infty$ such that

$$|K(u) - K(v)| \le L|u - v|, \quad \text{for any } u, v \in [0, 1].$$

This implies $K(\cdot) \le M_K$ for a constant M_K.
(A3) $\{T_{ij} : i = 1, \ldots, n, j = 1, \ldots, m_i\}$, are i.i.d. copies of a random variable T defined on [0,1]. The density $f(\cdot)$ of T is bounded from below and above: $0 < m_f \le \min_{t\in[0,1]} f(t) \le \max_{t\in[0,1]} f(t) \le M_f < \infty$.
(A4) $f^{(2)}(\cdot)$ and $g^{(2)}(\cdot)$, the second derivatives of $f(\cdot)$ and $g(\cdot)$, are bounded, respectively.

https://doi.org/10.1017/S0269964822000511 Published online by Cambridge University Press

(A5) $\partial^2\mu(t,z)/\partial t$ and $\partial^2\mu(t,z)/\partial z$, the second derivatives of $\mu(t,z)$ with respect to t and z, are bounded on $[0,1]^2$, respectively.
(A6) $\partial^2 G(s,t,z)/\partial s^2$, $\partial^2 G(s,t,z)/\partial t^2$, $\partial^2 G(s,t,z)/\partial z^2$, $\partial^2 G(s,t,z)/\partial s\partial t$, $\partial^2 G(s,t,z)/\partial s\partial z$, and $\partial^2 G(s,t,z)/\partial t\partial z$ are bounded on $[0,1]^3$.

Conditions for mean function estimation

(B1) $h_{\mu t}\to 0$. For some finite ρ_μ, $h_{\mu t}/h_{\mu z}\to\rho_\mu$.
(B2) $\log(n)\sum_{i=1}^n m_i\omega_i^2/h_{\mu t}h_{\mu z}\to 0$, and $\log(n)\sum_{i=1}^n m_i(m_i-1)\omega_i^2/h_{\mu z}\to 0$.
(B3) For some $\alpha>2$, $E\{\sup_{t,z\in[0,1]}|U(t,z)|^\alpha\}<\infty$, $E|\varepsilon|^\alpha<\infty$ and

$$n\left\{\sum_{i=1}^n m_i\omega_i^2 h_{\mu t}h_{\mu z}+\sum_{i=1}^n m_i(m_i-1)\omega_i^2 h_{\mu t}^2 h_{\mu z}\right\}\left\{\frac{\log(n)}{n}\right\}^{2/\alpha-1}\to\infty.$$

(B4) $\sup_n(n\max_i m_i\omega_i)\le B<\infty$.

Conditions for covariance function estimation

(C1) $h_{Gt}\to 0$ and $h_{Gt}/h_{Gz}\to\rho_G$ for some $0<\rho_G<\infty$.
(C2) $\log(n)\sum_{i=1}^n m_i(m_i-1)\nu_i^2/h_{Gt}^2 h_{Gz}\to 0$, $\log(n)\sum_{i=1}^n m_i(m_i-1)(m_i-2)\nu_i^2/h_{Gt}h_{Gz}\to 0$, $\log(n)\sum_{i=1}^n m_i(m_i-1)(m_i-2)(m_i-3)\nu_i^2/h_{Gz}\to 0$.
(C3) For some $\beta>2$, $E\{\sup_{t,z\in[0,1]}|U(t,z)|^{2\beta}\}<\infty$, $E|\varepsilon|^{2\beta}<\infty$, and

$$\begin{aligned}n\Bigg\{&\sum_{i=1}^n m_i(m_i-1)\nu_i^2 h_{Gt}^2 h_{Gz}+\sum_{i=1}^n m_i(m_i-1)(m_i-2)\nu_i^2 h_{Gt}^3 h_{Gz}\\&+\sum_{i=1}^n m_i(m_i-1)(m_i-2)(m_i-3)\nu_i^2 h_{Gt}^4 h_{Gz}\Bigg\}\left\{\frac{\log(n)}{n}\right\}^{2/\beta-1}\to\infty.\end{aligned}$$

(C4) $\sup_n(n\max_i m_i(m_i-1)\nu_i)\le B'<\infty$.

Conditions (A1)–(A6) are standard in the literature of functional data analysis (FDA) and nonparametric smoothing such as Yao *et al.* [23]. Conditions (B1)–(B4) are used to guarantee the consistency of the estimator of the mean function. Their counterparts for the OBS and SUBJ schemes are commonly used in FDA, similar versions can refer to Chen and Müller [4]. Likewise, the Conditions (C1)–(C4) are used to establish the consistency of the estimated covariance function. Condition (C1) is a mild condition that requires the bandwidth h_{Gt} and h_{Gz} to converge to zero at the same rate. Conditions (C2) and (C3) impose restrictions on m_i, n and $h_{Gt}(h_{Gz})$ in theory. Condition (C4) allows m_i is different for each subject i but not too irregular. More detailed demonstration can refer to Zhang and Wang [24].

Auxiliary results and proofs

In this subsection, we provide detailed proofs for the estimators in this paper. Below, for any square matrix $\mathbf{A}$, $|\mathbf{A}|$ denotes the determinant. We denote

$$L(t,z)=\sum_{i=1}^n\omega_i\sum_{j=1}^{m_i}K\left(\frac{T_{ij}-t}{h_{\mu t}}\right)K\left(\frac{Z_i-z}{h_{\mu z}}\right)U_{ij}^+,$$

where U_{ij}^+ is the positive part of U_{ij} and $U_{ij}:=U(T_{ij},Z_i)=Y_{ij}-\mu(T_{ij},Z_i)$.

The following lemma is used to prove Theorem 1.

https://doi.org/10.1017/S0269964822000511 Published online by Cambridge University Press

Lemma A.1. *Under the assumptions for Theorem 1,*

$$\sup_{t,z\in[0,1]} |L(t,z) - EL(t,z)| = O(a_n) \quad a.s.,$$

where

$$a_n = \left[\log(n)\left\{\sum_{i=1}^{n} m_i\omega_i^2 h_{\mu t}h_{\mu z} + \sum_{i=1}^{n} m_i(m_i-1)\omega_i^2 h_{\mu t}^2 h_{\mu z}\right\}\right]^{1/2}.$$

Proof. For the sake of simplicity, we denote $A_n = a_n\{n/\log(n)\}$. By (B2), we can choose ρ such that $n^{\rho}h_{\mu t}a_n \to \infty$ (i.e., $n^{\rho}h_{\mu z}a_n \to \infty$ by (B1)) and let $\chi(\rho)$ be an equally sized mesh on $[0,1]^2$ with grid $n^{-\rho}$ by $n^{-\rho}$. Observe that

$$\begin{aligned}
\sup_{t,z\in[0,1]} |L(t,z) - EL(t,z)| \le{}& \sup_{t,z\in\chi(\rho)} \sup_{\substack{t,s,z,z'\in[0,1]\\ |t-s|,|z-z'|\le n^{-\rho}}} |L(t,z) - L(s,z')| \\
&+ \sup_{t,z\in\chi(\rho)} |L(t,z) - EL(t,z)| \\
&+ \sup_{t,z\in\chi(\rho)} \sup_{\substack{t,s,z,z'\in[0,1]\\ |t-s|,|z-z'|\le n^{-\rho}}} |EL(t,z) - EL(s,z')| \\
\equiv{}& Q_1 + Q_2 + Q_3.
\end{aligned}$$

We first consider Q_1 and Q_3. For all t and z, by (A2) and Hölder inequality, it follows that

$$\begin{aligned}
Q_1 ={}& \sup_{|t-s|,|z-z'|\le n^{-\rho}} |L(t,z) - L(s,z')| \\
\le{}& \sup_{|t-s|,|z-z'|\le n^{-\rho}} \left|\sum_{i=1}^{n}\omega_i\sum_{j=1}^{m_i} U_{ij}^{+}K\left(\frac{Z_i-z}{h_{\mu z}}\right)\left\{K\left(\frac{T_{ij}-t}{h_{\mu t}}\right) - K\left(\frac{T_{ij}-s}{h_{\mu t}}\right)\right\}\right| \\
&+ \sup_{|t-s|,|z-z'|\le n^{-\rho}} \left|\sum_{i=1}^{n}\omega_i\sum_{j=1}^{m_i} U_{ij}^{+}K\left(\frac{T_{ij}-s}{h_{\mu t}}\right)\left\{K\left(\frac{Z_i-z}{h_{\mu z}}\right) - K\left(\frac{Z_i-z'}{h_{\mu z}}\right)\right\}\right| \\
\le{}& \left(\sum_{i=1}^{n}\omega_i\sum_{j=1}^{m_i} U_{ij}^{+}\right) M_K L n^{-\gamma}(1/h_{\mu t} + 1/h_{\mu z}) \\
\le{}& \left(\sum_{i=1}^{n}\omega_i\sum_{j=1}^{n} 1^{\alpha/(\alpha-1)}\right)^{(\alpha-1)/\alpha}\left(\sum_{i=1}^{n}\omega_i\sum_{j=1}^{n} U_{ij}^{\alpha}\right)^{1/\alpha} M_K L n^{-\gamma}(1/h_{\mu t} + 1/h_{\mu z}) \\
\le{}& \left\{\sum_{i=1}^{n} m_i\omega_i \sup_{t,z\in[0,1]} |U_i(t,z)|^{\alpha}\right\}^{1/\alpha} M_K L n^{-\gamma}(1/h_{\mu t} + 1/h_{\mu z}).
\end{aligned}$$

Combing the Conditions (B3) and (B4), and the strong law of large numbers, we have

$$\begin{aligned}
\left\{\sum_{i=1}^{n} m_i\omega_i \sup_{t,z\in[0,1]} |U_i(t,z)|^{\alpha}\right\}^{1/\alpha} \le{}& \left(n\max_i m_i\omega_i\right)\frac{1}{n}\sum_{i=1}^{n}\sup_{t,z\in[0,1]} |U_i(t,z)|^{\alpha} \\
\le{}& B\frac{1}{n}\sum_{i=1}^{n}\sup_{t,z\in[0,1]} |U_i(t,z)|^{\alpha}
\end{aligned}$$

https://doi.org/10.1017/S0269964822000511 Published online by Cambridge University Press

$$\to BE \sup_{t,z\in[0,1]} |U_i(t,z)|^\alpha$$
$$< \infty, \quad \text{a.s..}$$

$n^\rho h_{\mu t} a_n \to \infty$ and $n^\rho h_{\mu z} a_n \to \infty$ lead to $n^{-\rho}/h_{\mu t} = o(a_n)$ and $n^{-\rho}/h_{\mu z} = o(a_n)$, respectively, which implies that $Q_1 = o(a_n)$, a.s.. The term Q_3 can be dealt with similarly and we skip the details here. We now consider the Q_2. Observe that

$$Q_2 \le \sup_{t,z\in\chi(\rho)} |L(t,z)^* - EL(t,z)^*|$$
$$+ \sup_{t,z\in\chi(\rho)} \sum_{i=1}^{n} \omega_i \sum_{j=1}^{m_i} K\left(\frac{T_{ij}-t}{h_{\mu t}}\right) K\left(\frac{Z_i - z}{h_{\mu z}}\right) U_{ij}^{+} I(U_{ij}^{+} > A_n)$$
$$+ \sup_{t,z\in\chi(\rho)} \sum_{i=1}^{n} \omega_i \sum_{j=1}^{m_i} E\left\{K\left(\frac{T_{ij}-t}{h_{\mu t}}\right) K\left(\frac{Z_i - z}{h_{\mu z}}\right) U_{ij}^{+} I(U_{ij}^{+} > A_n)\right\},$$

where $L(t,z)^*$ is the truncation of $L(t,z)$, that is

$$L(t,z)^* = \sum_{i=1}^{n} \omega_i \sum_{j=1}^{m_i} K\left(\frac{T_{ij}-t}{h_{\mu t}}\right) K\left(\frac{Z_i - z}{h_{\mu z}}\right) U_{ij}^{+} I(U_{ij}^{+} \le A_n),$$

where $I(\cdot)$ is the indicator function. Combing the Conditions (A2), (B3)–(B4) and $A_n = a_n\{n/\log(n)\}$, for every $t, z \in \chi(\rho)$, we have

$$\sum_{i=1}^{n} \omega_i \sum_{j=1}^{m_i} K\left(\frac{T_{ij}-t}{h_{\mu t}}\right) K\left(\frac{Z_i - z}{h_{\mu z}}\right) U_{ij}^{+} I(U_{ij}^{+} > A_n)$$
$$\le M_K A_n^{1-\alpha} \sum_{i=1}^{n} \omega_i \sum_{j=1}^{m_i} |U_{ij}|^\alpha$$
$$\le BM_K A_n^{1-\alpha} \left\{ n^{-1} \sum_{i=1}^{n} \sup_{t,z\in[0,1]} |U_i(t,z)|^\alpha \right\}$$
$$= o(a_n) \quad \text{a.s.,}$$

where $o(\cdot)$a.s. is uniform in $t, z \in \chi(\rho)$. Similarly,

$$\sup_{t,z\in\chi(\rho)} \sum_{i=1}^{n} \omega_i \sum_{j=1}^{m_i} E\left\{K\left(\frac{T_{ij}-t}{h_{\mu t}}\right) K\left(\frac{Z_i - z}{h_{\mu z}}\right) U_{ij}^{+} I(U_{ij}^{+} > A_n)\right\} = o(a_n), \quad \text{a.s.}$$

Note that $L(t,z)^* - EL(t,z)^* = \sum_{i=1}^{n}(V_i - EV_i)$, where

$$V_i = \omega_i \sum_{j=1}^{m_i} K((T_{ij}-t)/h_{\mu t}) K((Z_i - z)/h_{\mu z}) U_{ij}^{+} I(U_{ij}^{+} \le A_n).$$

https://doi.org/10.1017/S0269964822000511 Published online by Cambridge University Press

It is easy to check that, for some constant $M_U > 0$,

$$\begin{aligned}
\operatorname{var} V_i &\le EV_i^2 \\
&\le \sum_{i=1}^n m_i \omega_i^2 E\left\{K\left(\frac{T_{ij}-t}{h_{\mu t}}\right)K\left(\frac{Z_i-z}{h_{\mu z}}\right)U_{ij}^+\right\}^2 \\
&\quad + \sum_{i=1}^n m_i(m_i-1)\omega_i^2 E\left\{K\left(\frac{T_{ij}-t}{h_{\mu t}}\right)K\left(\frac{T_{ij'}-t}{h_{\mu t}}\right)K^2\left(\frac{Z_i-z}{h_{\mu z}}\right)U_{ij}^+U_{ij'}^+\right\} \\
&\le M_U\left\{\sum_{i=1}^n m_i\omega_i^2 h_{\mu t}h_{\mu z} + \sum_{i=1}^n m_i(m_i-1)\omega_i^2 h_{\mu t}^2 h_{\mu z}\right\}
\end{aligned}$$

and $|V_i - EV_i| \le 2m_i\omega_i M_K^2 A_n \le 2BM_K^2 A_n/n$ implied by (B4). By Bernstein inequality, for any $M > 0$,

$$\begin{aligned}
&P\left(\sup_{t,z\in\chi(\rho)} |L(t,z)^* - EL(t,z)^*| > Ma_n\right) \\
&= P\left(\bigcup_{t,z\in\chi(\rho)} |L(t,z)^* - EL(t,z)^*| > Ma_n\right) \\
&\le n^\rho P\left(\left|\sum_{i=1}^n (V_i - EV_i)\right| > Ma_n\right) \\
&\le 2n^\rho \exp\left(-M^2a_n^2/2 \Bigg/ \left[M_U\left\{\sum_{i=1}^n m_i\omega_i^2(h_{\mu t}h_{\mu z} + h_{\mu t}^3 h_{\mu z} + h_{\mu t}h_{\mu z}^3)\right.\right.\right. \\
&\quad \left.\left.\left. + \sum_{i=1}^n m_i(m_i-1)\omega_i^2 h_{\mu t}^2 h_{\mu z}\right\} + 2BM_K^2 A_n M a_n/3n\right]\right) \\
&= 2n^\rho \exp\left(-\frac{M^2}{2M_U/\log n + 4BM_K^2 M/3\log n}\right) \\
&= 2n^{\rho - M^*},
\end{aligned}$$

where $M^* = M^2/(2M_U + 4BM_k^2 M/3)$. So $P(\sup_{t,z\in\chi(\rho)} |L(t,z)^* - EL(t,z)^*| > Ma_n)$ is summable in n if we select M enough such that $M^* - \rho > 1$. By Borel–Cantelli's lemma, $\sup_{t,z\in\chi(\rho)} |L(t,z)^* - EL(t,z)^*| = O(a_n)$ a.s.. This completes the proof. □

Proof of Theorem 1. Denote

$$S_{pq} = \sum_{i=1}^n \omega_i \sum_{j=1}^{m_i} K_{h_{\mu t}}(T_{ij}-t)K_{h_{\mu z}}(Z_i - z)\left(\frac{T_{ij}-t}{h_{\mu t}}\right)^p\left(\frac{Z_i-z}{h_{\mu z}}\right)^q$$

and

$$R_{pq} = \sum_{i=1}^n \omega_i \sum_{j=1}^{m_i} K_{h_{\mu t}}(T_{ij}-t)K_{h_{\mu z}}(Z_i - z)\left(\frac{T_{ij}-t}{h_{\mu t}}\right)^p\left(\frac{Z_i-z}{h_{\mu z}}\right)^q Y_{ij},$$

https://doi.org/10.1017/S0269964822000511 Published online by Cambridge University Press

for $p, q = 0, \ldots, 2$. It is easy to check that

$$\begin{aligned}
&\widehat{\mu}(t,z) - \mu(t,z) \\
&= (S_{20}S_{02} - S_{11}^2)\left\{R_{00} - \mu(t,z)S_{00} - h_{\mu t}\frac{\partial \mu}{\partial t}(t,z)S_{10} - h_{\mu z}\frac{\partial \mu}{\partial z}(t,z)S_{01}\right\} \\
&\quad \Big/ S_{20}S_{02} - S_{11}^2)S_{00} - (S_{10}S_{02} - S_{01}S_{11})S_{10} + (S_{10}S_{11} - S_{01}S_{20})S_{01}\} \\
&\quad - (S_{10}S_{02} - S_{01}S_{11})\left\{R_{10} - \mu(t,z)S_{10} - h_{\mu t}\frac{\partial \mu}{\partial t}(t,z)S_{20} - h_{\mu z}\frac{\partial \mu}{\partial z}(t,z)S_{11}\right\} \\
&\quad \Big/ \{(S_{20}S_{02} - S_{11}^2)S_{00} - (S_{10}S_{02} - S_{01}S_{11})S_{10} + (S_{10}S_{11} - S_{01}S_{20})S_{01}\} \\
&\quad + (S_{10}S_{11} - S_{01}S_{20})\left\{R_{01} - \mu(t,z)S_{01} - h_{\mu t}\frac{\partial \mu}{\partial t}(t,z)S_{11} - h_{\mu z}\frac{\partial \mu}{\partial z}(t,z)S_{02}\right\} \\
&\quad \Big/ \{(S_{20}S_{02} - S_{11}^2)S_{00} - (S_{10}S_{02} - S_{01}S_{11})S_{10} + (S_{10}S_{11} - S_{01}S_{20})S_{01}\},
\end{aligned} \tag{A.1}$$

where

$$\begin{aligned}
&R_{00} - \mu(t,z)S_{00} - h_{\mu t}\frac{\partial \mu}{\partial t}(t,z)S_{10} - h_{\mu z}\frac{\partial \mu}{\partial z}(t,z)S_{01} \\
&= \sum_{i=1}^{n}\omega_i\sum_{j=1}^{m_i}K_{h_{\mu t}}(T_{ij}-t)K_{h_{\mu z}}(Z_i-z) \\
&\quad \times\left[\delta_{ij} + \mu(T_{ij},Z_i) - \mu(t,z) - (T_{ij}-t)\frac{\partial \mu}{\partial t}(t,z) - (Z_i-z)\frac{\partial \mu}{\partial z}(t,z)\right] \\
&= \sum_{i=1}^{n}\omega_i\sum_{j=1}^{m_i}K_{h_{\mu t}}(T_{ij}-t)K_{h_{\mu z}}(Z_i-z)\delta_{ij} + O(h_{\mu t}^2) + O(h_{\mu z}^2) + O(h_{\mu t}h_{\mu z}) \quad \text{a.s..}
\end{aligned}$$

By Lemma A.1, $\sup_{t,z\in[0,1]}|\sum_{i=1}^{n}\omega_i\sum_{j=1}^{m_i}K_{h_{\mu t}}(T_{ij}-t)K_{h_{\mu z}}(Z_i-z)\delta_{ij}| = O(a_n/h_{\mu t}h_{\mu z})$, which yields that

$$\begin{aligned}
&R_{00} - \mu(t,z)S_{00} - h_{\mu t}\frac{\partial \mu}{\partial t}(t,z)S_{10} - h_{\mu z}\frac{\partial \mu}{\partial z}(t,z)S_{01} \\
&= O(h_{\mu t}^2 + h_{\mu z}^2 + h_{\mu t}h_{\mu z} + a_n/h_{\mu t}h_{\mu z}) \quad \text{a.s..}
\end{aligned}$$

We also note that $ES_{20} = f(t)f(z)\sigma_K^2 + O(h_{\mu t} + h_{\mu z})$, $ES_{02} = f(t)f(z)\sigma_K^2 + O(h_{\mu t} + h_{\mu z})$ and $ES_{11} = O(h_{\mu t}h_{\mu z})$, which imply that $S_{20}S_{02} - S_{11}^2$ is positive and bounded away from 0 on $[0,1]^2$a.s. and it is easy to see that $\sup_{t,z\in[0,1]}|S_{20}S_{02} - S_{11}^2| = O_p(1)$, $\sup_{t,z\in[0,1]}|S_{10}S_{02} - S_{01}S_{11}| = O_p(1)$ and $\sup_{t,z\in[0,1]}|S_{10}S_{11} - S_{01}S_{20}t| = O_p(1)$. Therefore, the order of the first term of (A.1) is

$$\begin{aligned}
&(S_{20}S_{02} - S_{11}^2)\left\{R_{00} - \mu(t,z)S_{00} - h_{\mu t}\frac{\partial \mu}{\partial t}(t,z)S_{10} - h_{\mu z}\frac{\partial \mu}{\partial z}(t,z)S_{01}\right\} \\
&\quad \Big/ \{(S_{20}S_{02} - S_{11}^2)S_{00} - (S_{10}S_{02} - S_{01}S_{11})S_{10} + (S_{10}S_{11} - S_{01}S_{20})S_{01}\} \\
&= O(h_{\mu t}^2 + h_{\mu z}^2 + h_{\mu t}h_{\mu z} + a_n/h_{\mu t}h_{\mu z}) \quad \text{a.s..}
\end{aligned}$$

The same rate can also be similarly seen to hold for the other two terms of (A.1). This completes the proof. □

In the following, we present the convergence rate of covariance function. Firstly, let

$$L(t,s,z)=\sum_{i=1}^{n} v_i \sum_{1\le j\neq k\le m_i} K\left(\frac{T_{ij}-t}{h_{Gt}}\right)K\left(\frac{T_{ik}-s}{h_{Gt}}\right)K\left(\frac{Z_i-z}{h_{Gz}}\right)U^{+}_{ijk},$$

where U^{+}_{ijk} is the positive part of $U_{ij}U_{ik}$.

Lemma A.2. *Under the assumptions for Theorem 2,*

$$\sup_{s,t,z\in[0,1]} |L(t,s,z)-EL(t,s,z)| = O(b_n), \quad a.s.$$

where

$$b_n=\left[\log(n)\left\{\sum_{i=1}^{n} m_i(m_i-1)v_i^2 h_{Gt}^2 h_{Gz}+\sum_{i=1}^{n} m_i(m_i-1)(m_i-2)v_i^2 h_{Gt}^3 h_{Gz}\right.\right.$$
$$\left.\left.+\sum_{i=1}^{n} m_i(m_i-1)(m_i-2)(m_i-3)v_i^2 h_{Gt}^4 h_{Gz}\right\}\right]^{1/2}.$$

Proof. Denotes $B_n=b_n\{n/\log(n)\}$. Using (C2), we can also choose $\gamma>0$ such that $n^{\gamma}h_{Gt}b_n\to\infty$ (and $n^{\gamma}h_{Gz}b_n\to\infty$ by (C1)). Let $\chi(\gamma)$ be a three-dimensional grid on $[0,1]^3$ with grid size $n^{-\gamma}\times n^{-\gamma}\times n^{-\gamma}$. Therefore, by (C3) and (C4), we have

$$\sup_{t,s,z\in[0,1]} |L(t,s,z)-EL(t,s,z)| \le \sup_{t,s,z\in\chi(\gamma)} |L(t,s,z)-EL(t,s,z)|+Q_1+Q_2 \tag{A.2}$$

where

$$Q_1=\sup_{|t-t'|,|s-s'|,|z-z'|\le n^{-\gamma}} |L(t,s,z)-L(t',s',z')|,$$

and

$$Q_2=\sup_{|t-t'|,|s-s'|,|z-z'|\le n^{-\gamma}} |EL(t,s,z)-EL(t',s',z')|.$$

We note that

$$\begin{aligned}
Q_1 \le{}& \sup_{|t-t'|,|s-s'|,|z-z'|\le n^{-\gamma}} \sum_{i=1}^{n} v_i \sum_{j\neq k}^{m_i} K\left(\frac{T_{ik}-s}{h_{Gt}}\right)K\left(\frac{Z_i-z}{h_{Gz}}\right)\\
&\times U^{+}_{ijk}\left|K\left(\frac{T_{ij}-t}{h_{Gt}}\right)-K\left(\frac{T_{ij}-t'}{h_{Gt}}\right)\right|\\
&+\sup_{|t-t'|,|s-s'|,|z-z'|\le n^{-\gamma}} \sum_{i=1}^{n} v_i \sum_{j\neq k}^{m_i} K\left(\frac{T_{ij}-t'}{h_{Gt}}\right)K\left(\frac{Z_i-z}{h_{Gz}}\right)\\
&\times U^{+}_{ijk}\left|K\left(\frac{T_{ij}-s}{h_{Gt}}\right)-K\left(\frac{T_{ij}-s'}{h_{Gt}}\right)\right|\\
&+\sup_{|t-t'|,|s-s'|,|z-z'|\le n^{-\gamma}} \sum_{i=1}^{n} v_i \sum_{j\neq k}^{m_i} K\left(\frac{T_{ij}-t'}{h_{Gt}}\right)K\left(\frac{T_{ik}-s'}{h_{Gt}}\right)\\
&\times U^{+}_{ijk}\left|K\left(\frac{T_{ij}-z}{h_{Gz}}\right)-K\left(\frac{T_{ij}-z'}{h_{Gt}}\right)\right|\\
\equiv{}& A_{n1}+A_{n2}+A_{n3}.
\end{aligned}$$

https://doi.org/10.1017/S0269964822000511 Published online by Cambridge University Press

Now, by (A2), (C4), Hölder inequality and strong law of large numbers,

$$\begin{aligned}
A_{n1} &\le \sup_{|t-t'|,|s-s'|,|z-z'|\le n^{-\gamma}} M_K^2 \sum_{i=1}^{n} \nu_i \sum_{j\neq k}^{m_i} U_{ijk}^{+} L n^{-\gamma}/h_{Gt} \\
&\le \left\{ \sum_{i=1}^{n} \nu_i \sum_{j\neq k}^{m_i} (U_{ijk}^{+})^{\beta} \right\}^{1/\beta} M_K^2 L n^{-\gamma}/h_{Gt} \\
&\le \left\{ \sum_{i=1}^{n} m_i(m_i-1)\nu_i \sup_{t,z\in[0,1]} |U_i(t,z)|^{2\beta} \right\}^{1/\beta} L n^{-\gamma}/h_{Gt} \\
&= o(b_n) \quad \text{a.s..}
\end{aligned}$$

The similar demonstration can be obtained for A_{n2}, A_{n3} and Q_2. So, $Q_1, Q_2 = o(b_n)$a.s.. Let the truncated $L(s,t,z)$ be

$$L(s,t,z)^* = \sum_{i=1}^{n} \nu_i \sum_{1\le j\neq k\le m_i} K\left(\frac{T_{ij}-t}{h_{Gt}}\right) K\left(\frac{T_{ik}-s}{h_{Gt}}\right) K\left(\frac{Z_i-z}{h_{Gz}}\right) U_{ijk}^{+} I(U_{ijk}^{+} \le B_n),$$

where $I(\cdot)$ is the indicator function. Then,

$$\sup_{t,s,z\in\chi(\gamma)} |L(t,s,z) - EL(t,s,z)| \le \sup_{t,s,z\in\chi(\gamma)} |L(t,s,z)^* - EL(t,s,z)^*| + Q_1^* + Q_2^*, \tag{A.3}$$

where

$$\begin{aligned}
Q_1^* &= \sup_{t,s,z\in\chi(\gamma)} \sum_{i=1}^{n} \nu_i \sum_{1\le j\neq k\le m_i} K\left(\frac{T_{ij}-t}{h_{Gt}}\right) K\left(\frac{T_{ik}-s}{h_{Gt}}\right) K\left(\frac{Z_i-z}{h_{Gz}}\right) U_{ijk}^{+} I(U_{ijk}^{+} > B_n), \\
Q_2^* &= \sup_{t,s,z\in\chi(\gamma)} \sum_{i=1}^{n} \nu_i \sum_{1\le j\neq k\le m_i} E\left\{ K\left(\frac{T_{ij}-t}{h_{Gt}}\right) K\left(\frac{T_{ik}-s}{h_{Gt}}\right) K\left(\frac{Z_i-z}{h_{Gz}}\right) U_{ijk}^{+} I(U_{ijk}^{+} > B_n) \right\}.
\end{aligned}$$

Combing the Conditions (A2), (C3)–(C4) and $B_n = b_n\{n/\log(n)\}$, it follows that $Q_1^*, Q_2^* = o(b_n), a.s..$ Now, we rewrite $L(s,t,z)^* - EL(s,t,z)^* = \sum_{i=1}^{n}(V_i - EV_i)$, where

$$V_i = \nu_i \sum_{j\neq k} K\left(\frac{T_{ij}-t}{h_{Gt}}\right) K\left(\frac{T_{ik}-s}{h_{Gt}}\right) K\left(\frac{Z_i-z}{h_{Gz}}\right) U_{ijk}^{+} I(U_{ijk}^{+} \le B_n).$$

Notice that

$$\begin{aligned}
E(V_i - EV_i)^2 &\le EV_i^2 \\
&\le m_i(m_i-1) E\left\{ K\left(\frac{T_{ij}-t}{h_{Gt}}\right) K\left(\frac{T_{ik}-s}{h_{Gt}}\right) K\left(\frac{Z_i-z}{h_{Gz}}\right) U_{ijk}^{+} \right\}^2
\end{aligned}$$

https://doi.org/10.1017/S0269964822000511 Published online by Cambridge University Press

$$
\begin{aligned}
&+ m_i(m_i-1)(m_i-2)E\left\{K\left(\frac{T_{ij}-t}{h_{Gt}}\right)K\left(\frac{T_{ij'}-t}{h_{Gt}}\right)\right.\\
&\left.\times K^2\left(\frac{T_{ik}-s}{h_{Gt}}\right)K^2\left(\frac{Z_i-z}{h_{Gz}}\right)U^+_{ijk}U^+_{ij'k}\right\}\\
&+ m_i(m_i-1)(m_i-2)(m_i-3)E\left\{K\left(\frac{T_{ij}-t}{h_{Gt}}\right)K\left(\frac{T_{ik}-s}{h_{Gt}}\right)\right.\\
&\left.\times K\left(\frac{T_{ij'}-t}{h_{Gt}}\right)K\left(\frac{T_{ik'}-s}{h_{Gt}}\right)K^2\left(\frac{Z_i-z}{h_{Gz}}\right)U^+_{ijk}U^+_{ij'k'}\right\}\\
\le{}& M_U\left\{\sum_{i=1}^n m_i(m_i-1)\nu_i^2h_{Gt}^2h_{Gz}+\sum_{i=1}^n m_i(m_i-1)(m_i-2)\nu_i^2h_{Gt}^3h_{Gz}\right.\\
&\left.+\sum_{i=1}^n m_i(m_i-1)(m_i-2)(m_i-3)\nu_i^2h_{Gt}^4h_{Gz}\right\},
\end{aligned}
$$

for some constant $M_U > 0$ and $|V_i - EV_i| \le 2M_K^3 m_i(m_i-1)\nu_i B_n \le 2BM_K^3 B_n/n$. Similar to the proof of Lemma A.1, by Bernstein inequality, we have

$$
\sup_{t,s,z\in\chi(\gamma)} |L(s,t,z)^* - EL(s,t,z)^*| = O(b_n) \quad \text{a.s..}
$$

Together with (A.2) and (A.3), the proof is completed. □

Now, we give the proof of Theorem 2.

Proof of Theorem 2. We denote that

$$
\begin{aligned}
S_{pq\ell} = \sum_{i=1}^n \sum_{1\le j\ne k\le m_i} & K_{h_{Gt}}(T_{ij}-t)K_{h_{Gt}}(T_{ik}-s)K_{h_{Gz}}(Z_i-z)\\
&\times\left(\frac{T_{ij}-t}{h_{Gt}}\right)^p\left(\frac{T_{ij}-s}{h_{Gt}}\right)^q\left(\frac{Z_i-z}{h_{Gz}}\right)^\ell,
\end{aligned}
$$

and

$$
\begin{aligned}
R_{pq\ell} = \sum_{i=1}^n \sum_{1\le j\ne k\le m_i} & K_{h_{Gt}}(T_{ij}-t)K_{h_{Gt}}(T_{ik}-s)K_{h_{Gz}}(Z_i-z)\\
&\times\left(\frac{T_{ij}-t}{h_{Gt}}\right)^p\left(\frac{T_{ij}-s}{h_{Gt}}\right)^q\left(\frac{Z_i-z}{h_{Gz}}\right)^\ell C_{ijk},
\end{aligned}
$$

where $C_{ijk} = \{Y_{ij} - \widehat{\mu}(T_{ij}, Z_i)\}\{Y_{ik} - \widehat{\mu}(T_{ik}, Z_i)\}$, $p, q, \ell = 0, 1, 2$. Let

$$
\mathbf{S} = \begin{pmatrix} S_{000} & S_{100} & S_{010} & S_{001}\\ S_{100} & S_{200} & S_{110} & S_{101}\\ S_{010} & S_{110} & S_{020} & S_{011}\\ S_{001} & S_{101} & S_{011} & S_{002}\end{pmatrix}, \quad \mathbf{R} = \begin{pmatrix} R_{000}\\ R_{100}\\ R_{010}\\ R_{001}\end{pmatrix},
$$

https://doi.org/10.1017/S0269964822000511 Published online by Cambridge University Press

and note that the algebraic cofactors of $\mathbf{S}$ are

$$\mathbf{S}^{*}=\begin{pmatrix} C_{11} & C_{21} & C_{31} & C_{41} \\ C_{12} & C_{22} & C_{32} & C_{42} \\ C_{13} & C_{23} & C_{33} & C_{43} \\ C_{14} & C_{24} & C_{34} & C_{44} \end{pmatrix},$$

where C_{jk} are the cofactors of the (j,k) element of $\mathbf{S}$. Then,

$$\widehat{G}(t,s,z) = \frac{(C_{11}R_{000} + C_{21}R_{100} + C_{31}R_{010} + C_{41}R_{001})}{|\mathbf{S}|},$$
$$G(t,s,z) = \frac{C_{11}S_{000} + C_{21}S_{100} + C_{31}S_{010} + C_{41}S_{001}}{|\mathbf{S}|}G.$$

By Cramer's rule,

$$\begin{aligned}\widehat{G}(t,s,z) - G(t,s,z) = |\mathbf{S}|^{-1}\Big\{&(C_{11}R_{000} + C_{21}R_{100} + C_{31}R_{010} + C_{41}R_{001})\\ &- (C_{11}S_{000} + C_{21}S_{100} + C_{31}S_{010} + C_{41}S_{001})G\\ &- (C_{11}S_{010} + C_{21}S_{110} + C_{31}S_{010} + C_{41}S_{011})\frac{\partial G}{\partial s}h_{Gt}\\ &- (C_{11}S_{001} + C_{21}S_{101} + C_{31}S_{011} + C_{41}S_{002})\frac{\partial G}{\partial z}h_{Gz}\\ &-(C_{11}S_{100} + C_{21}S_{200} + C_{31}S_{110} + C_{41}S_{101})\frac{\partial G}{\partial t}h_{Gt}\Big\}\\ = \sum_{k=1}^{4}\mathscr{C}_k,\end{aligned}$$

where

$$\mathscr{C}_1 = |\mathbf{S}|^{-1}C_{11}\left(R_{000} - GS_{000} - h_{Gt}\frac{\partial G}{\partial s}S_{010} - h_{Gz}\frac{\partial G}{\partial z}S_{001} - h_{Gt}\frac{\partial G}{\partial t}S_{100}\right),$$
$$\mathscr{C}_2 = |\mathbf{S}|^{-1}C_{21}\left(R_{100} - GS_{100} - h_{Gt}\frac{\partial G}{\partial s}S_{110} - h_{Gz}\frac{\partial G}{\partial z}S_{101} - h_{Gt}\frac{\partial G}{\partial t}S_{200}\right),$$
$$\mathscr{C}_3 = |\mathbf{S}|^{-1}C_{31}\left(R_{010} - GS_{010} - h_{Gt}\frac{\partial G}{\partial s}S_{010} - h_{Gz}\frac{\partial G}{\partial z}S_{011} - h_{Gt}\frac{\partial G}{\partial t}S_{110}\right),$$

and

$$\mathscr{C}_4 = |\mathbf{S}|^{-1}C_{41}\left(R_{001} - GS_{001} - h_{Gt}\frac{\partial G}{\partial s}S_{011} - h_{Gz}\frac{\partial G}{\partial z}S_{002} - h_{Gt}\frac{\partial G}{\partial t}S_{101}\right).$$

We will focus on $\mathscr{C}_1$. The other three terms are of the same order and can be dealt with similarly. Specifically,

$$\begin{aligned}\mathscr{C}_1 = |\mathbf{S}|^{-1}C_{11}\Big(&R_{000} - G(s,t,z)S_{000} - h_{Gt}\frac{\partial G(s,t,z)}{\partial s}S_{010}\\ &-h_{Gz}\frac{\partial G(s,t,z)}{\partial z}S_{001} - h_{Gt}\frac{\partial G(s,t,z)}{\partial t}S_{100}\Big)\end{aligned}$$

https://doi.org/10.1017/S0269964822000511 Published online by Cambridge University Press

$$
\begin{aligned}
&= |\mathbf{S}|^{-1}C_{11}\left[\sum_{i=1}^{n}\sum_{1\le j\ne k\le m_i} K_{h_{Gt}}(T_{ij}-t)K_{h_{Gt}}(T_{ik}-s)K_{h_{Gz}}(Z_i-z)\right.\\
&\quad\times\left\{C_{ijk}-G(s,t,z)-\frac{\partial G(s,t,z)}{\partial s}(T_{ik}-s)\right.\\
&\quad\left.\left.-\frac{\partial G(s,t,z)}{\partial z}(Z_i-z)-\frac{\partial G(s,t,z)}{\partial t}(T_{ij}-t)\right\}\right]\\
&= |\mathbf{S}|^{-1}C_{11}\left[\sum_{i=1}^{n}\sum_{1\le j\ne k\le m_i} K_{h_{Gt}}(T_{ij}-t)K_{h_{Gt}}(T_{ik}-s)K_{h_{Gz}}(Z_i-z)\right.\\
&\quad\left.\times\{C_{ijk}-G(T_{ik},T_{ij},Z_i)+h_{Gt}^2+h_{Gz}^2+h_{Gt}h_{Gz}\}\right],
\end{aligned}
$$

where

$$
\begin{aligned}
C_{ijk} &= \delta_{ij}\delta_{ik}+\delta_{ij}\{\mu(T_{ik},Z_i)-\widehat{\mu}(T_{ik},Z_i)\}+\delta_{ik}\{\mu(T_{ij},Z_i)-\widehat{\mu}(T_{ij},Z_i)\}\\
&\quad+\{\widehat{\mu}(T_{ij},Z_i)-\mu(T_{ij},Z_i)\}\{\widehat{\mu}(T_{ik},Z_i)-\mu(T_{ik},Z_i)\},
\end{aligned}
$$

and $\delta_{ij}=Y_{ij}-\mu(T_{ij},Z_i)$. By Lemma A.1, for all T_{ij} and T_{ik},

$$
\widehat{\mu}(T_{ij},Z_i)-\mu(T_{ij},Z_i),\widehat{\mu}(T_{ik},Z_i)-\mu(T_{ik},Z_i)=O(h_{\mu t}^2+h_{\mu z}^2+h_{\mu t}h_{\mu z}+\delta_{n1}) \quad \text{a.s.}
$$

Similar to Lemma 2 in Zhang and Wang [24], the $S_{pq\ell}$ converges almost surely to their respective means in supremum norm and are thus bounded almost surely for $p, q, \ell = 0, 1, 2$, so that C_{p1} is bounded almost surely for $p = 1, 2, 3, 4$ Then $|\mathbf{S}|^{-1}$ is bounded away from 0 by the almost sure supremum convergence of $S_{pq\ell}$ and Slutsky's theorem. Then using Lemma A.2 and Theorem 5.2 in Zhang and Wang [24], we can obtain that

$$
\begin{aligned}
&\sup_{s,t,z\in[0,1]}|\widehat{G}(t,s,z)-G(t,s,z)|\\
&= O(h_{\mu t}^2+h_{\mu z}^2+h_{\mu t}h_{\mu z}+\delta_{n1}+h_{Gt}^2+h_{Gz}^2+h_{Gt}h_{Gz}+\delta_{n2}) \quad \text{a.s..}
\end{aligned}
$$

This completes the proof. □

Cite this article: Yan X, Wang H, Sun H and Zhao P (2024). Incorporating covariate into mean and covariance function estimation of functional data under a general weighing scheme. *Probability in the Engineering and Informational Sciences* **38**, 82–99. https://doi.org/10.1017/S0269964822000511