Submitted to the Annals of Statistics
arXiv: math.PR/0000000

FUSED KERNEL-SPLINE SMOOTHING FOR REPEATEDLY MEASURED OUTCOMES IN A GENERALIZED PARTIALLY LINEAR MODEL WITH FUNCTIONAL SINGLE INDEX*

By Fei Jiang, Yanyuan Ma and Yuanjia Wang

Harvard University, University of South Carolina, and Columbia University

We propose a generalized partially linear functional single index risk score model for repeatedly measured outcomes where the index itself is a function of time. We fuse the nonparametric kernel method and regression spline method, and modify the generalized estimating equation to facilitate estimation and inference. We use local smoothing kernel to estimate the unspecified coefficient functions of time, and use B-splines to estimate the unspecified function of the single index component. The covariance structure is taken into account via a working model, which provides valid estimation and inference procedure whether or not it captures the true covariance. The estimation method is applicable to both continuous and discrete outcomes. We derive large sample properties of the estimation procedure and show different convergence rate of each component of the model. The asymptotic properties when the kernel and regression spline methods are combined in a nested fashion has not been studied prior to this work even in the independent data case.

1. Introduction. As a semiparametric regression model, single index model is a popular way to accomodate multivariate covariates while retain model flexibility. For independent outcomes, Carroll et al. (1997) introduced a generalized partially linear single index model which enriches the family of single index models by allowing an additional linear component. The goal of this paper is to develop a class of generalized partially linear single index models with functional covariate effect and explore the estimation and inference for repeatedly measured dependent outcomes.

In the longitudinal data framework, let i denote the ith individual, and k be the kth measurement, where $i = 1, \ldots, n$ and $k = 1 \ldots, M_i$. Here M_i

*This work was supported by the National Science Foundation (DMS-1206693 and DMS-1000354) and the National Institute of Neurological Disorders and Stroke (NS073671, NS082062). The authors thank the editor, associate editor and three anonymous referees for their comprehensive review which greatly improved the paper.

Keywords and phrases: B-spline, Generalized linear model, Huntington's disease, Infinite dimension, Logistic model, Semiparametric model, Single index model.

is the total number of observations available for the ith individual. Let D_{ik} be the response variable, $\mathbf{Z}_{ik}$ and $\mathbf{X}_{ik}$ be d_w and d_β dimensional covariate vectors. We assume the observations from different individuals are independent, while the responses $D_{i1}, \dots, D_{iM_i}$ assessed on the same individual at different time points are correlated but we do not attempt to model such correlation. To model the relationship between the conditional mean of the repeatedly measured outcomes D_{ik} at time T_{ik} and covariates $\mathbf{Z}_{ik}, \mathbf{X}_{ik}$, we propose a partially linear functional single index model which models the mean of D_{ik} given $\mathbf{Z}_{ik}, \mathbf{X}_{ik}$ at time T_{ik} in the form of

$$E(D_{ik} \mid \mathbf{X}_{ik}, \mathbf{Z}_{ik}, T_{ik}) = H[m\{\mathbf{w}(T_{ik})^{\mathrm{T}}\mathbf{Z}_{ik}\} + \boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}_{ik}], \tag{1}$$

where H is a known differentiable monotone link function, $\mathbf{w}(t) \in \mathrm{R}^{d_w}$ at any t, $\boldsymbol{\beta} \in \mathrm{R}^{d_\beta}$. Such model is useful when the time varying effect of $\mathbf{Z}_{ik}$ and the functional combined score effect of $\mathbf{w}(T_{ik})^{\mathrm{T}}\mathbf{Z}_{ik}$, adjusted by the covariate vector $\mathbf{X}_{ik}$, are of main interest. Note that both $\mathbf{X}_{ik}$ and $\mathbf{Z}_{ik}$ can contain components that do not vary with k, such as gender, and the ones that vary with k such as age. Here, $m(0)$ serves as the intercept term, thus $\mathbf{X}_{ik}$ does not contain the constant one. In Model (1), $\mathbf{Z}_{ik}$ includes the covariates of main research interest whose effects are usually time varying and modeled nonparametrically, and $\mathbf{X}_{ik}$ contains additional covariates of secondary scientific interest and whose effects are only modeled via a simple linear form. Here m is an unspecified smooth single index function. Further $\mathbf{w}$ is a d_w-dimensional vector of smooth functions in L_2, while $\mathbf{w}(t)$ is $\mathbf{w}$ evaluated at t, hence a d_w-dimensional vector. In addition, $\mathbf{w}(t)$ contributes to form the argument of the function m, which yields a nested nonparametric functional form. To ensure identifiability and to reflect the practial application that motivated this example, we further require $\mathbf{w}(t) > \mathbf{0}$ and $\|\mathbf{w}(t)\|_1 = 1$ $\forall t$. Here $\mathbf{w}(t) > \mathbf{0}$ means every component in $\mathbf{w}(t)$ is positive, and $\|\cdot\|_1$ denotes the vector l_1-norm, i.e. the sum of the absolute values of the components in the vector. The choice of l_1 norm incorporates the practical knowledge from our real data example described in Section 4 and is not critical. It can be modified to other norms, such as the most often used l_2 norm or sup norm in our subsequent development. We assume the observed data follow the model described above. Throughout the texts, we use subscript $_0$ to denote the true parameters. Before we proceed, we first show that

PROPOSITION 1. *Assume $m_0 \in \mathcal{M}$, where $\mathcal{M} = \{m \in C^1([0,1])$, m is one-to-one, and $m(0) = c_0\}$. Here $C^1([0,1])$ is the space of functions with continuous derivatives on $[0,1]$ and c_0 is a finite constant. Assume $\mathbf{w}_0(t) \in \mathcal{D}$, where $\mathcal{D} = \{\mathbf{w} = (w_1, \dots, w_{d_w})^{\mathrm{T}} : \|\mathbf{w}_0(t)\|_1 = 1, w_j > 0$,* *and* $w_j \in C^1([0,\tau])$ $\forall j = 1, \ldots d_w\}$. *Here* $C^1([0,\tau])$ *is the space of functions with continuous derivatives on* $[0,\tau]$ *and* τ *is a finite constant. Assume* $E(\mathbf{X}_{ik}^{\otimes 2})$ *and* $E(\mathbf{Z}_{ik}^{\otimes 2})$ *are both positive definite, where we define* $\mathbf{a}^{\otimes 2} = \mathbf{a}\mathbf{a}^{\mathrm{T}}$ *for an arbitrary vector* $\mathbf{a}$. *Then under these assumptions, the parameter set* $(\boldsymbol{\beta}_0, m_0, \mathbf{w}_0)$ *in (1) is identifiable.*

The proof of Proposition 1 is in Appendix A.1. Model (1) can be viewed as a longitudinal extension of the generalized partially linear single index risk score model introduced in Carroll et al. (1997), i.e.,

$$E(D_{ik} \mid \mathbf{X}_{ik}, \mathbf{Z}_{ik}) = H\{m(\mathbf{w}^{\mathrm{T}}\mathbf{Z}_{ik}) + \boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}_{ik}\}, \tag{2}$$

which is a popular way to increase flexibility when covariate dimension may be high. Many existing literatures explore the generalized partially linear single index model under the longitudinal settings. Jiang and Wang (2011) consider the single index function in the form of $m(\mathbf{w}^{\mathrm{T}}\mathbf{Z}_{ik}, t)$, which allows a time dependent function m, but $\mathbf{w}$ is time invariant hence it does not have the nesting structure in Model (1) to capture the time dependent effect of $\mathbf{Z}_{ik}$. Furthermore, the method does not consider the within subject correlation. Xu and Zhu (2012) adopted Model (2) as marginal model in the longitudinal data setting. Their method takes into account the within subject correlation, but, similar to Jiang and Wang (2011)'s approach, it does not allow $\mathbf{w}$ to vary with time, hence is not sufficient to describe the time varying effect of $\mathbf{Z}_{ik}$. We modify Jiang and Wang (2011) and Xu and Zhu (2012)'s models to accommodate the time dependent score effect $\mathbf{w}(t)$. In Section 4, we show that time-dependent effect is essential to improve model fit in some practical situations. In addition, we retain the virtue of Jiang and Wang (2011) and Xu and Zhu (2012)'s models by using the semiparametric functional single index model, which overcomes the curse of dimensionality, and alleviate the risk of model mis-specification (Peng and Huang, 2011).

The estimation and inference for Model (1) are challenging due to the nonparametric form of $m, \mathbf{w}$, and the complications from correlation between repeatedly measured outcomes. The estimation for single index models has been discussed extensively in both kernel and spline literatures. Carroll et al. (1997) proposed a local kernel smoothing technique to estimate the unknown function m and the finite dimensional parameters $\mathbf{w}, \boldsymbol{\beta}$ in Model (2) through iterative procedures. Later, Xia and Härdle (2006) applied a kernel-based minimum average variance estimation (MAVE) method for partially linear single index models, which was first proposed by Xia et al. (2002) for dimension reduction. When $\mathbf{Z}_{ik}$ is continuous, MAVE results in consistent estimators for the single index function m without the root-n assumption on $\mathbf{w}$ as in Carroll et al. (1997). Nevertheless, when $\mathbf{Z}_{ik}$ is discrete, the method may fail to obtain consistent estimators without prior information about $\boldsymbol{\beta}$ (Xia et al., 2002; Wang et al., 2010). Moreover, Wang and Yang (2009) showed that MAVE is unreliable for estimating single index coefficient $\mathbf{w}$ when $\mathbf{Z}_{ik}$ is unbalanced and sparse, i.e., when $\mathbf{Z}_{ik}$ is measured at different time points for each subject, and each subject may have only a few measurements.

To overcome these limitations, we apply the B-spline method to estimate the unknown function m, which is stable when the data set contains discrete or sparse $\mathbf{Z}_{ik}$. Although the B-spline method outperforms the kernel method in estimating m, problems arise if it is also used for estimating $\mathbf{w}(t)$ in our model setting. If spline approximations are used for both m and $\mathbf{w}(t)$ with k knots, then we must simultaneously solve $(d_w + 1)k$ estimating equations to get the spline coefficients associated with the spline knots, which may cause numerical instability and is computationally expensive when the parameter number increases with the sample size. To alleviate the computational burden and instability, we estimate $\mathbf{w}(t)$ by using the kernel method. At different time point t, the procedure solves $\mathbf{w}(t)$ independently and in parallel, hence it does not suffer from the numerical instability and is computationally efficient. To handle longitudinal outcomes, we use the idea from the generalized estimating equation (GEE) to combine a set of estimating equations built from the marginal model. It is worth pointing out that the GEE in its original form is only applicable when the index $\mathbf{w}$ does not change along time. In conclusion, we combine the kernel and B-spline smoothing with the GEE approach, and develop a fused kernel/B-spline procedure for estimation and inference.

The fusion of kernel and B-spline poses theoretical challenges which we address in this work. To the best of our knowledge, this is the first time kernel and spline methods are jointly implemented in a nested function setting. We study convergence properties, such as asymptotic bias and variance, for each component of the model, show that the parametric component achieves the regular root-n convergence rate, and establish the relation of the nonparametric function convergence rates to the number of B-spline basis functions and B-spline order, as well as their relation to the kernel bandwidth. These results provide guidelines for choosing the number of knots in association with spline order and bandwidth in order to optimize the performance. They also further facilitate inference, such as constructing confidence intervals and performing hypothesis testing. Although theoretical properties of kernel smoothing and spline smoothing are available separately, the properties when these two methods are combined in a nested fashion has not been studied in the literature even for the independent data case prior to this work. Because the vector functions $\mathbf{w}$ appears inside the function m, the asymptotic analysis of the spline and kernel methods are not completely separable. This requires a comprehensive analysis and integration of both methods instead of a mechanical combination of two separate techniques.

The rest of the paper is structured as the following. In Section 2, we define some notations and state assumptions in the model, introduce the fused kernel/B-spline semiparametric estimating equation, illustrate the profiling estimation procedure to obtain the estimators, and study the asymptotic properties of the resulting estimators. In Section 3, we evaluate the estimation procedure on simulated data sets. In Section 4, we apply the model and estimation procedure on the Huntington's disease data set. We conclude the paper with some discussion in Section 5. We present the technical proofs in Appendix and an online supplementary document (Jiang, Ma and Wang, 2015).

2. Estimating equations and profiling procedure. In this section, we construct estimators for $(\boldsymbol{\beta}, m, \mathbf{w})$ in Model (1). We first derive a set of estimating equations, through applying both B-spline and kernel methods. We then introduce a profiling procedure to implement the estimation. Finally, we discuss the asymptotic properties of the estimators.

Many estimation procedures have been developed for the single index risk score model. In addition to the methods describe in Section 1, for the models with uncorrelated responses, Cui, Härdle and Zhu (2011) illustrate an estimating function method based on the kernel approach for the generalized single index risk score model. Ma and Zhu (2013) discuss a doubly robust and efficient estimation procedure for the single index risk score model with high dimensional covariates. Ma and Song (2014) and Lu and Loomis (2013) propose B-spline methods for estimating the unknown regression link functions in single index risk score models. However, these methods are not adequate for the parameter estimation in our model. As shown in (1), in addition to an unknown link function m, our functional single index model contains a nonparametric function $\mathbf{w}(t)$ which is multivariate and appears inside m. Therefore, we develop a GEE type method for the parameter estimation in our model which allows to take into account the within patient correlation. In conjunction with the kernel smoothing technique and B-spline basis expansion, our fused method estimates both the coefficients as a function of time and the unspecified regression function, and simultaneously handles the complexities of repeated measurements and curses of dimensionality.

More specifically, let $\mathbf{B}_r(u) = \{B_{r1}(u), \ldots, B_{rd_\lambda}(u)\}^{\mathrm{T}}$ be the set of B-spline basis functions of order r and let $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_{d_\lambda})^{\mathrm{T}}$ be the coefficients of the B-spline approximation. Denoting $\widetilde{m}(u,\boldsymbol{\lambda}) = \mathbf{B}_r(u)^{\mathrm{T}}\boldsymbol{\lambda}$, de Boor (2001) has shown the existence of a $\boldsymbol{\lambda}_0 \in \mathrm{R}^{d_\lambda}$ so that $\widetilde{m}(u,\boldsymbol{\lambda}_0) = \mathbf{B}_r(u)^{\mathrm{T}}\boldsymbol{\lambda}_0$ converges to $m_0(u)$ uniformly on $(0,1)$ when the number of the B-spline inner knots goes to infinite (See **Fact 1** in Section S.2 in the supplementary article (Jiang, Ma and Wang, 2015)). A detailed description of the B-spline functions and the properties of their derivatives can be found in de Boor (2001).

The B-spline approximation greatly eases the parameter estimation procedure. Operationally, for a given sample size n, the problem is reduced from estimating the infinite dimensional m to estimating a finite dimensional vector $\boldsymbol{\lambda}$. Since the dimension of $\boldsymbol{\lambda}$ grows with the sample size, the estimation consistency can be achieved when the sample size goes to infinity. Let $\boldsymbol{\theta} = (\boldsymbol{\beta}^{\mathrm{T}},\boldsymbol{\lambda}^{\mathrm{T}})^{\mathrm{T}} \in \mathbb{R}^{d_\theta}$, the approximated mean function can be written as

$$H[\mathbf{B}_r\{\mathbf{w}(T_{ik})^{\mathrm{T}}\mathbf{Z}_{ik}\}^{\mathrm{T}}\boldsymbol{\lambda} + \boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}_{ik}].$$

We investigate the properties for estimating $m_0, \mathbf{w}_0, \boldsymbol{\beta}_0$ through investigating the properties of the estimators for $\boldsymbol{\lambda}_0$, $\mathbf{w}_0$ and $\boldsymbol{\beta}_0$.

2.1. *Notations.* We define some notations to present the estimation procedure. To keep the main text concise, we illustrate the specific forms of notations in the Section A.2 in Appendix. Generally, for a generic vector valued function $\mathbf{a}$ that depends on some additional parameters, we use $\widehat{\mathbf{a}}$ to denote the function with the estimated parameter values plugged in. For example, this applies to $\mathbf{S}_w, \mathbf{S}_\beta, \widehat{\mathbf{S}}_w$, $\widehat{\mathbf{S}}_\beta$ in the following text. The specific forms of $\mathbf{S}_w, \mathbf{S}_\beta, \widehat{\mathbf{S}}_w, \widehat{\mathbf{S}}_\beta$ are given in Section A.2 in Appendix.

In our profiling procedure, we estimate $\boldsymbol{\lambda}_0$ using $\widehat{\boldsymbol{\lambda}}$, considered as a functional of $\boldsymbol{\beta}_0, \mathbf{w}_0$. Then we estimate $\mathbf{w}_0$ using $\widehat{\mathbf{w}}$, considered as a function of $\boldsymbol{\beta}_0$ at different time points. Finally, we estimate $\boldsymbol{\beta}_0$ using $\widehat{\boldsymbol{\beta}}$. We further define $T_{ik}, k=1,\ldots,M_i, i=1,\ldots,n$ to be the random measurement times which are independent of $\mathbf{X}_{ik}, \mathbf{Z}_{ik}, D_{ik}$, $\mathbf{w}$ to be a function of t for $t \in [0,\tau]$, where τ is a finite constant, and $\widehat{\mathbf{w}}(\boldsymbol{\beta}), \widehat{\mathbf{w}}(\boldsymbol{\beta},t)$, considered as functions of $\boldsymbol{\beta}$, to be the estimators for $\mathbf{w}$ and $\mathbf{w}(t)$, respectively.

Let $\mathbf{Q}_\beta(\mathbf{X}_{ik}) = \mathbf{X}_{ik}$, $\mathbf{Q}_\lambda\{\mathbf{Z}_{ik};\mathbf{w}(t)\} = \mathbf{B}_r\{\mathbf{w}(t)^{\mathrm{T}}\mathbf{Z}_{ik}\}$, and $\mathbf{Q}_w\{\mathbf{Z}_{ik};\boldsymbol{\lambda},\mathbf{w}(t)\} = \mathbf{Z}_{ik}\mathbf{B}_r'\{\mathbf{w}(t)^{\mathrm{T}}\mathbf{Z}_{ik}\}^{\mathrm{T}}\boldsymbol{\lambda}$, to be the partial derivatives of $\mathbf{B}_r\{\mathbf{w}(t)^{\mathrm{T}}\mathbf{Z}_{ik}\}^{\mathrm{T}}\boldsymbol{\lambda} + \boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}_{ik}$ with respect to $\boldsymbol{\beta}$, $\boldsymbol{\lambda}$, $\mathbf{w}(t)$. In the sequel, we will frequently use $\mathbf{Q}_{\beta ik}$, $\mathbf{Q}_{\lambda ik}\{\mathbf{w}(t)\}$, $\mathbf{Q}_{wik}\{\boldsymbol{\lambda},\mathbf{w}(t)\}$ as short forms for $\mathbf{Q}_\beta(\mathbf{X}_{ik})$, $\mathbf{Q}_\lambda\{\mathbf{Z}_{ik};\mathbf{w}(t)\}$ and $\mathbf{Q}_w\{\mathbf{Z}_{ik};\boldsymbol{\lambda},\mathbf{w}(t)\}$ respectively.

In general, to simplify the notations, we use subscripts to indicate the observations, i.e. for a generic function $a(\cdot)$, we write $a_i(\cdot) \equiv a(\mathbf{O}_i;\cdot)$, where

$\mathbf{O}_i$ denotes the ith observed variables. For example we write

$$H_{ik}\{\boldsymbol{\beta}, \boldsymbol{\lambda}, \mathbf{w}(t)\} \equiv H[\mathbf{B}_r\{\mathbf{w}(t)^{\mathrm{T}}\mathbf{Z}_{ik}\}^{\mathrm{T}}\boldsymbol{\lambda} + \boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}_{ik}].$$

Further, we indicate the use of the true function instead of its B-spline approximation by replacing the argument $\boldsymbol{\lambda}$ with m, for example,

$$H_{ik}\{\boldsymbol{\beta}, m, \mathbf{w}(t)\} \equiv H[m\{\mathbf{w}(t)^{\mathrm{T}}\mathbf{Z}_{ik}\} + \boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}_{ik}].$$

We also define $\Theta(u) = dH(u)/du$ and

$$\Theta_{ik}\{\boldsymbol{\beta}, \boldsymbol{\lambda}, \mathbf{w}(t)\} = \Theta[\mathbf{B}_r\{\mathbf{w}(t)^{\mathrm{T}}\mathbf{Z}_{ik}\}^{\mathrm{T}}\boldsymbol{\lambda} + \boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}_{ik}],$$

and

$$\Theta_{ik}\{\boldsymbol{\beta}, m, \mathbf{w}(t)\} = \Theta[m\{\mathbf{w}(t)^{\mathrm{T}}\mathbf{Z}_{ik}\} + \boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}_{ik}]$$

throughout the text.

The profiling procedure has three steps. We define the details of notations used in each step and their corresponding population forms in the Section A.2 in Appendix.

2.2. *Estimation procedure via profiling.* In this section, we define the estimation procedures for m, $\mathbf{w}_0$ and $\boldsymbol{\beta}_0$ via estimating equations which are solved through a profiling procedure as we describe below. We first estimate the function m through B-splines, by treating $\mathbf{w}$ and $\boldsymbol{\beta}$ as parameters that are held fixed. This yields a set of estimating equations for the spline coefficients, as functions of $\mathbf{w}$ and $\boldsymbol{\beta}$. We then estimate the partially linear nonparametric component $\mathbf{w}(t)$ of the cognitive score profiles through local kernel smoothing, while treating $\boldsymbol{\beta}$ as fixed parameters. This further allows us to obtain a second set of estimating equations at each time point that the function $\mathbf{w}(t)$ needs to be estimated, as a function of $\boldsymbol{\beta}$. Finally, we estimate the parametric component coefficients $\boldsymbol{\beta}$ through solving its own corresponding estimating equation set. The profiling procedure achieves a certain separation by allowing us to treat only one of the three components in each of the three nested steps, hence it eases the computational complexities. Because the B-spline estimator $\widehat{\boldsymbol{\lambda}}$, kernel estimator $\widehat{\mathbf{w}}(t)$, and linear parametric estimator $\widehat{\boldsymbol{\beta}}$ have different convergence rates, such separation also facilitates analysis of the asymptotic properties, compared with a simultaneous estimation procedure.

Step 1.

We obtain $\widehat{\boldsymbol{\lambda}}(\boldsymbol{\beta}_0, \mathbf{w}_0)$ by solving

$$\sum_{i=1}^{n} \widetilde{\mathbf{Q}}_{\lambda i}\{\mathbf{w}_0(\mathbf{T}_i)\}^{\mathrm{T}}\boldsymbol{\Theta}_i\{\boldsymbol{\beta}_0, \boldsymbol{\lambda}, \mathbf{w}_0(\mathbf{T}_i)\}\boldsymbol{\Omega}_i^{-1}[\mathbf{D}_i - \mathbf{H}_i\{\boldsymbol{\beta}_0, \boldsymbol{\lambda}, \mathbf{w}_0(\mathbf{T}_i)\}] = \mathbf{0}$$

with respect to $\boldsymbol{\lambda}$, where $\boldsymbol{\Omega}_i$ is a working covariance matrix, and $\boldsymbol{\Theta}_i = \mathrm{diag}\{\Theta_{ik}\}, k = 1, \dots, M_i$ is a $M_i \times M_i$ diagonal matrix. From the first step, we obtain the B-spline coefficients to estimate the function m.

Step 2.

We obtain $\widehat{\mathbf{w}}(\boldsymbol{\beta})$ in this step. Let $\mathbf{K}_h(\mathbf{T}_i - t_0)$ be a $d_w M_i \times d_w M_i$ diagonal matrix whose kth diagonal block is $\mathrm{diag}\{K_h(T_{ik} - t_0)\}$ where $K_h(s) = h^{-1}K(s/h)$ is a Kernel function with bandwidth h.

To obtain $\widehat{\mathbf{w}}(\boldsymbol{\beta}_0, t_0)$, we solve the estimating equation

$$\begin{aligned}(3)\qquad &\sum_{i=1}^{n} \widehat{\mathbf{A}}_{wi}\{\boldsymbol{\beta}_0, \widehat{\boldsymbol{\lambda}}(\boldsymbol{\beta}_0, \mathbf{w}), \mathbf{w}(t_0)\}\widehat{\mathbf{V}}_{wi}\{\boldsymbol{\beta}_0, \widehat{\boldsymbol{\lambda}}(\boldsymbol{\beta}_0, \mathbf{w}), \mathbf{w}(t_0)\}^{-1}\\ &\times \mathbf{K}_h(\mathbf{T}_i - t_0)\widehat{\mathbf{S}}_{wi}\{\boldsymbol{\beta}_0, \widehat{\boldsymbol{\lambda}}(\boldsymbol{\beta}_0, \mathbf{w}), \mathbf{w}(t_0)\}\end{aligned}$$

with respect to $\mathbf{w}$. Recall that $\|\mathbf{w}(t_0)\|_1 = 1$. In the implementation, we parameterize $w_{d_w} = 1 - \sum_{j=1}^{d_w - 1} w_j$, and derive the score functions for the vector $(w_1, \dots, w_{d_w - 1})$. We then solve the estimating equation system which contains the $d_w - 1$ equations constructed from the score functions and the equation $\sum_{j=1}^{d_w} w_j - 1 = 0$. The roots of the estimating equation system automatically satisfy the l_1 constraint. In all our experiments, the resulting $\widehat{w}_j(t)$ are nonnegative automatically, hence we did not particularly enforce the nonnegativity as a constraint. If it is needed, one can further enforce the nonnegativity and perform a constrained optimization.

Step 3.

We obtain $\widehat{\boldsymbol{\beta}}$ by solving

$$\begin{aligned}(4)\qquad &\sum_{i=1}^{n} \widehat{\mathbf{A}}_{\beta i}[\boldsymbol{\beta}, \widehat{\boldsymbol{\lambda}}\{\boldsymbol{\beta}, \widehat{\mathbf{w}}(\boldsymbol{\beta}, \mathbf{T}_i)\}, \widehat{\mathbf{w}}(\boldsymbol{\beta})]\widehat{\mathbf{V}}_{\beta i}[\boldsymbol{\beta}, \widehat{\boldsymbol{\lambda}}\{\boldsymbol{\beta}, \widehat{\mathbf{w}}(\boldsymbol{\beta})\}, \widehat{\mathbf{w}}(\boldsymbol{\beta}, \mathbf{T}_i)]^{-1}\\ &\times \widehat{\mathbf{S}}_{\beta i}[\boldsymbol{\beta}, \widehat{\boldsymbol{\lambda}}\{\boldsymbol{\beta}, \widehat{\mathbf{w}}(\boldsymbol{\beta})\}, \widehat{\mathbf{w}}(\boldsymbol{\beta}, \mathbf{T}_i)] = \mathbf{0}.\end{aligned}$$

In above steps, we approximate $\partial\widehat{\mathbf{w}}(\boldsymbol{\beta}, \mathbf{T}_i)/\partial\boldsymbol{\beta}^{\mathrm{T}}$, $\partial\widehat{\boldsymbol{\lambda}}(\boldsymbol{\beta}, \mathbf{w})/\partial\boldsymbol{\beta}^{\mathrm{T}}$, and $\partial\widehat{\boldsymbol{\lambda}}(\boldsymbol{\beta}_0, \mathbf{w}_0)/\partial\mathbf{w}$ by the leading terms in their expansions. Their explicit forms are shown in (S.27) in the proofs of Lemma 6, (S.37) in the proofs of Lemma 11, and **Notations in Step 2** in Appendix, respectively.

2.3. *Asymptotic properties of the estimators.* The profiling estimator described in Section 2.2 is quite complex, caused by the functional nature of $\mathbf{w}(t)$, the unspecified forms of both $\mathbf{w}$ and m and their nested appearance in the model, the correlation among different observations associated with the same individual and the different numbers of observations for each individual. In addition, the fused kernel/B-spline method requires careful joint consideration of both smoothing techniques. As a consequence, the analysis to obtain the asymptotic properties of the estimator described in Section 2.2 is very challenging and involved. We first list the regularity conditions under which we perform our theoretical analysis.

(A1) The kernel function $K(\cdot)$ is non-negative, has compact support, and satisfies $\int K(s)ds$
$= 1$, $\int K(s)sds = 0$ and $\int K(s)s^2ds < \infty$, and $\int K^2(s)sds < \infty$.

(A2) The bandwidth h in the kernel smoothing satisfies $nh^2 \to \infty$ and $nh^4 \to 0$ when $n \to \infty$.

(A3) The density function of $\mathbf{w}(t)^{\mathrm{T}}\mathbf{Z}$ for each $t \in [0, \tau]$ is bounded away from 0 on $S_w(t)$ and satisfies the Lipschitz condition of order 1 on $S_w(t)$, where $\mathbf{w}$ is in a neighborhood of $\mathbf{w}_0$, and $S_w(t) = \{\mathbf{w}(t)^{\mathrm{T}}\mathbf{Z}, \mathbf{Z} \in S\}$ and S is a compact support of $\mathbf{Z}$ and $\tau < \infty$ is a finite constant. Without loss of generality, we assume $S_w(t) = [0, 1]$.

(A4) Assume $m_0 \in \{m \in C^q([0, 1])$, m is one-to-one, and $m(0) = c_0\}$. Here $C^q([0, 1])$ is the space of functions with first q continuous derivatives on $[0, 1]$. The spline order $r \geq q$. The cluster size M_i is a fixed finite number that does not diverge with the sample size, i.e. $M_i < \infty$ for all i.

(A5) Let h_p be the distance between the $(p + 1)$th and pth interior knots of the order r B-spline functions. And $h_b = \max_{r \leq p \leq N+r} h_p$. There exists $0 < c_{h_b} < \infty$, such that

$$\max_{r \leq p \leq N+r} h_{p+1} = o(N^{-1}) \text{ and } h_b / \min_{r \leq p \leq N+r} h_p < c_{h_b},$$

where N is the number of knots which satisfies $N \to \infty$ as $n \to \infty$, and $N^{-1}n(\log n)^{-1} \to \infty$ and $Nn^{-1/(2q+1)} \to \infty$. Further assuming $q > 3$ and $N^{-3}n \to \infty$.

(A6) The matrices $E(\mathbf{X}_{ik}^{\otimes 2})$, $E([\{\mathbf{X}_{ik} - E\{\mathbf{X}_{ik}|\mathbf{w}(t)^{\mathrm{T}}\mathbf{Z}_{ik}\}]^{\otimes 2})$,
$E([\mathbf{Z}_{ik} - E\{\mathbf{Z}_{ik}|\mathbf{w}(t)^{\mathrm{T}}\mathbf{Z}_{ik}\}m_0'\{\mathbf{w}(t)^{\mathrm{T}}\mathbf{Z}_{ik}\}]^{\otimes 2})$ and
$E([\{\mathbf{X}_{ik}\mathbf{Z}_{ik}^{\mathrm{T}} - E\{\mathbf{X}_{ik}\mathbf{Z}_{ik}^{\mathrm{T}}|\mathbf{w}(t)^{\mathrm{T}}\mathbf{Z}_{ik}\}m_0'\{\mathbf{w}(t)^{\mathrm{T}}\mathbf{Z}_{ik}\}]^{\otimes 2})$ are finite and positive definite for any $t \in [0, \tau]$.

The requirements $nh^4 \to 0$ in (A2) and $Nn^{-1/(2q+1)} \to \infty$ in (A5) are undersmoothing requirements on the kernel approximation and on the spline approximation respectively. They are required to ensure that the biases, $E(\widehat{\mathbf{w}}) - \mathbf{w}$ and $E(\mathbf{B}_r^{\mathrm{T}}\widehat{\boldsymbol{\lambda}}) - m_0$, are ignorable compared to other terms left in the final analysis. This kind of undersmoothing conditions are commonly required in semiparametric models.

Theorems 1–3 describe the asymptotic properties for the estimators of $\mathbf{w}_0(t)$, $\boldsymbol{\beta}_0$ and m_0, respectively.

Theorem 1. *Assume Conditions (A1)-(A6) and the identifiability conditions stated in Proposition 1 hold. Let* $\widehat{\mathbf{A}}_{wi}$, $\widehat{\mathbf{V}}_{wi}$, $\widehat{\mathbf{S}}_{wi}$, *and their population forms* $\mathbf{A}_{wi}$, $\mathbf{V}_{wi}$ *and* $\mathbf{S}_{wi}$ *be as defined in* **Notation in Step 2** *in Section A.2 in Appendix. Let* $\widehat{\mathbf{w}}(\boldsymbol{\beta}_0, t_0)$ *solve (3) and* f_T *be the probability density function of* T_{ik} *with support* $[0, \tau]$. *Define*

$$\begin{aligned}\Sigma_w &= (nh)^{-1}\{\mathbf{B}(t_0)f_T(t_0)\}^{-1}E\bigg(f_T(t_0)[\mathbf{A}_{wi}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(t_0)\}\\&\quad \times\mathbf{V}_{wi}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(t_0)\}^{-1}]\int \mathbf{K}(\mathbf{s})\mathbf{V}^*_{wi}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(t_0)\}\mathbf{K}(\mathbf{s})d\mathbf{s}\\&\quad \times[\mathbf{A}_{wi}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(t_0)\}\mathbf{V}_{wi}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(t_0)\}^{-1}]^{\mathrm{T}}\bigg)\\&\quad \{\mathbf{B}(t_0)f_T(t_0)\}^{-1}.\end{aligned}$$

Then

$$\Sigma_w^{-1/2}\{\widehat{\mathbf{w}}(\boldsymbol{\beta}_0, t_0) - \mathbf{w}_0(t_0)\} \quad \xrightarrow{d} \quad N(0, \mathbf{I}),$$

where $\mathbf{B}$ *are defined in* **Notation in Step 3** *in Section A.2 in Appendix.*

Theorem 1 establishes the large samples properties of the estimation of the multivariate weight function $\mathbf{w}_0(t)$. It shows that our method achieves the usual nonparametric convengence rate of root-nh under the conditions given.

Theorem 2. *Assume Conditions (A1)-(A6) and the identifiability conditions stated in Proposition 1 hold. Let* $\widehat{\mathbf{S}}_{\beta ik}$, $\widehat{\mathbf{A}}_{\beta ik}$, $\widehat{\mathbf{V}}_{\beta ikl}$, *and their population forms* $\mathbf{S}_{\beta ik}$, $\mathbf{A}_{\beta ik}$, $\mathbf{V}_{\beta ikl}$ *be as defined in* **Notation in Step 3** *in Section A.2 in Appendix, and* $\widehat{\mathbf{w}}(\boldsymbol{\beta})$, $\mathbf{w}(\boldsymbol{\beta})$ *be as defined in Section 2.1. Let* $\widehat{\boldsymbol{\beta}}$ *solve (4), then*

$$\begin{aligned}(5)\qquad &\sqrt{n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)\\ &= \mathbf{F}(m_0)^{-1}\bigg(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\mathbf{A}_{\beta i}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_i)\}\mathbf{V}_{\beta i}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_i)\}^{-1}\\&\quad \times\mathbf{S}_{\beta i}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_i)\} - \frac{1}{\sqrt{n}}\sum_{j=1}^{n}E(\mathbf{A}_{\beta i}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_j)\}\\&\quad \times\mathbf{V}_{\beta i}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_i)\}^{-1}\mathcal{K}(\mathbf{T}_j)|\mathbf{O}_j)\mathbf{B}(\mathbf{T}_j)^{-1}[\mathbf{A}_{wj}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_j)\}\\&\quad \times\mathbf{V}_{wj}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_j)\}^{-1}\mathbf{S}_{wj}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_j)\}]\bigg) - \mathbf{G}(m_0)\mathbf{V}^{-1}\\&\quad \times\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\widetilde{\mathbf{Q}}_{\lambda i}\{\mathbf{w}_0(\mathbf{T}_i)\}^{\mathrm{T}}\boldsymbol{\Theta}_i\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_i)\}\boldsymbol{\Omega}_i^{-1}[\mathbf{D}_i\end{aligned}$$

$$-\mathbf{H}_i\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_i)\}])\{1+o_p(1)\},$$

where $\mathcal{K}(\mathbf{T}_i) = diag\{\boldsymbol{\kappa}(T_{ik}), k = 1,\ldots,M_i\}$ *a* $d_\beta M_i \times d_w M_i$ *matrix and* $\boldsymbol{\kappa}(T_{ik})$ *is*

$$\begin{aligned}&\Bigg\{\mathbf{Q}_{\beta ik} - \boldsymbol{\delta}\{\mathbf{w}_0(T_{ik})^{\mathrm{T}}\mathbf{Z}_{ik}\} - \Bigg(\mathbf{B}(T_{ik})^{-1}E\Bigg[\mathbf{A}_{wj}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(T_{ik})\}\\ &\mathbf{V}_{wj}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(T_{ik})\}^{-1}\frac{\partial \mathbf{S}_{wj}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(T_{ik})\}}{\partial \boldsymbol{\beta}^{\mathrm{T}}} \mid \mathbf{O}_i\Bigg]\Bigg)^{\mathrm{T}}\mathbf{Z}_{ik}\\ &\times m_0'\{\mathbf{w}_0(T_{ik})^{\mathrm{T}}\mathbf{Z}_{ik}\} + \boldsymbol{\gamma}\{\mathbf{w}_0(T_{ik})^{\mathrm{T}}\mathbf{Z}_{ik}\}\Bigg\}\mathbf{Q}_{wik}\{m_0, \mathbf{w}_0(T_{ik})\}^{\mathrm{T}}\\ &\times\Theta_{ik}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(T_{ik})\}.\end{aligned}$$

$$\begin{aligned}\mathbf{F}(m_0) &= -E\Bigg\{\mathbf{A}_{\beta i}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_i)\}\mathbf{V}_{\beta i}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_i)\}^{-1}\\ &\quad\times\frac{\partial \mathbf{S}_{\beta i}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_i)\}}{\partial \boldsymbol{\beta}^{\mathrm{T}}}\Bigg\},\end{aligned}$$

and

$$\begin{aligned}\mathbf{G}(m_0) &= E[\mathbf{A}_{\beta i}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_i)\}\mathbf{V}_{\beta i}\{\boldsymbol{\beta}_0, \boldsymbol{\lambda}_0, \mathbf{w}_0(\mathbf{T}_i)\}^{-1}\mathbf{C}_i\\ &\quad\boldsymbol{\Theta}_i^*\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_i)\}\mathbf{Q}_{\lambda i}^*\{\mathbf{w}_0(\mathbf{T}_i)\}].\end{aligned}$$

Here $\mathbf{C}_i$ *is a* $d_\beta M_i \times d_\beta M_i$ *with the kth block having the form*

$$\begin{aligned}&\Bigg\{\mathbf{Q}_{\beta ik} - \boldsymbol{\delta}\{\mathbf{w}_0(T_{ik})^{\mathrm{T}}\mathbf{Z}_{ik}\} - \Bigg(\mathbf{B}(T_{ik})^{-1}E\Bigg[\mathbf{A}_{wj}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(T_{ik})\}\\ &\mathbf{V}_{wj}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(T_{ik})\}^{-1}\\ &\times\frac{\partial \mathbf{S}_{wj}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(T_{ik})\}}{\partial \boldsymbol{\beta}^{\mathrm{T}}} \mid \mathbf{O}_i\Bigg]\Bigg)^{\mathrm{T}}\mathbf{Z}_{ik}m_0'\{\mathbf{w}_0(T_{ik})^{\mathrm{T}}\mathbf{Z}_{ik}\}\\ &+\boldsymbol{\gamma}\{\mathbf{w}_0(T_{ik})^{\mathrm{T}}\mathbf{Z}_{ik}\}\Bigg\}.\end{aligned}$$

Here $\boldsymbol{\Theta}_i^*\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_i)\}$ *is a* $d_\beta M_i \times d_\beta M_i$ *matrix with the kth block being a* $d_\beta \times d_\beta$ *diagonal matrix with the element* $\Theta_{ik}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(T_{ik})\}$. *And* $\mathbf{Q}_{\lambda i}^*\{\mathbf{w}_0(\mathbf{T}_i)\}$ *is a* $d_\beta M_i \times d_\lambda$ *matrix with kth row block being a* $d_\beta \times d_\lambda$ *matrix, which is* d_β *replicates of the row vector* $\mathbf{Q}_{\lambda ik}\{\mathbf{w}_0(T_{ik})\}^{\mathrm{T}}$. $\mathbf{B}$, $\boldsymbol{\delta}$, $\boldsymbol{\gamma}$ *are functions defined in* **Notation in Step 3** *in Section A.2 in Appendix.*

Consequently, we have

$$\sqrt{n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \xrightarrow{d} N(0, \Sigma),$$

where

$$\begin{aligned}\Sigma \;=\; & \mathbf{F}(m_0)^{-1}E[([\mathbf{A}_{\beta i}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_i)\}\mathbf{V}_{\beta i}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_i)\}^{-1} \\ & \times\mathbf{S}_{\beta i}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_i)\}]^{\otimes 2}) + \{(E(\mathbf{A}_{\beta i}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_j)\} \\ & \times\mathbf{V}_{\beta i}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_i)\}^{-1}\mathcal{K}(\mathbf{T}_j)|\mathbf{O}_j)\mathbf{B}(\mathbf{T}_j)^{-1}[\mathbf{A}_{wj}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_j)\} \\ & \times\mathbf{V}_{wj}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_j)\}^{-1}\mathbf{S}_{wj}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_j)\}]))^{\otimes 2}\} \\ & +\{(\mathbf{G}(m_0)\mathbf{V}^{-1}\widetilde{\mathbf{Q}}_{\lambda i}\{\mathbf{w}_0(\mathbf{T}_i)\}^{\mathrm{T}}\boldsymbol{\Theta}_i\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_i)\}\boldsymbol{\Omega}_i^{-1}[\mathbf{D}_i \\ & -\mathbf{H}_i\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_i)\}])^{\otimes 2}\}] \\ & \times\mathbf{F}(m_0)^{-1}.\end{aligned}$$

Theorem 2 establishes the usual parametric convergence rate for $\widehat{\boldsymbol{\beta}}$, even though the estimation relies on multiple nonparametric estimates as well. The form of (5) in Theorem 2 indicates that the variance of estimating $\boldsymbol{\beta}_0$ is inflated by the estimation $\widehat{\mathbf{w}}$ as given in

$$\begin{aligned}& \frac{1}{\sqrt{n}}\sum_{j=1}^{n} E(\mathbf{A}_{\beta i}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_j)\}\mathbf{V}_{\beta i}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_i)\}^{-1}\mathcal{K}(\mathbf{T}_j)|\mathbf{O}_j) \\ & \mathbf{B}(\mathbf{T}_j)^{-1}[\mathbf{A}_{wj}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_j)\}\mathbf{V}_{wj}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_j)\}^{-1} \\ & \times\mathbf{S}_{wj}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_j)\}])\end{aligned}$$

and is also inflated by the estimation $\widehat{\lambda}$, as given in

$$\begin{aligned}& \mathbf{G}(m_0)\mathbf{V}^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\widetilde{\mathbf{Q}}_{\lambda i}\{\mathbf{w}_0(\mathbf{T}_i)\}^{\mathrm{T}}\boldsymbol{\Theta}_i\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_i)\}\boldsymbol{\Omega}_i^{-1}[\mathbf{D}_i \\ & -\mathbf{H}_i\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_i)\}].\end{aligned}$$

See Lemma 9, 11 and the proofs of Theorem 2 in the supplementary article (Jiang, Ma and Wang, 2015) for more detailed discussion.

The asymptotic normality of $\widehat{\boldsymbol{\beta}}$ established in Theorem 2 further facilitates inference on $\boldsymbol{\beta}$ such as constructing confidence intervals or performing hypothesis testing. In implementing these inference procedures, we replace the variance-covariance matrix Σ with its estimate, where we use empirical sample mean over the observed samples to replace the expectations in Theorem 2, and plug in the estimates of the corresponding parameter and function values. This is the procedure adopted in all our numerical implementation.

Theorem 3. *Assume Conditions (A1)-(A6) and the identifiability conditions stated in Proposition 1 hold. Let* $\widehat{m}\{u, \widehat{\boldsymbol{\lambda}}(\boldsymbol{\beta}, \mathbf{w})\} = \mathbf{B}_r(u)^{\mathrm{T}}\widehat{\boldsymbol{\lambda}}(\boldsymbol{\beta}, \mathbf{w})$, $\widetilde{m}\{u, \boldsymbol{\lambda}_0\} = \mathbf{B}_r(u)^{\mathrm{T}}\boldsymbol{\lambda}_0$, *where* $\widehat{\boldsymbol{\lambda}}(\boldsymbol{\beta}_0, \mathbf{w}_0)$ *solves (3) and define*

$$\sigma^2(u, \mathbf{w}_0) \;\equiv\; \frac{1}{n}\mathbf{B}_r(u)^{\mathrm{T}}E([\widetilde{\mathbf{Q}}_{\lambda i}\{\mathbf{w}_0(\mathbf{T}_i)\}^{\mathrm{T}}\boldsymbol{\Theta}_i\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_i)\}\boldsymbol{\Omega}_i^{-1}$$

$$
\begin{aligned}
&\times\boldsymbol{\Theta}_i\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_i)\}\widetilde{\mathbf{Q}}_{\lambda i}\{\mathbf{w}_0(\mathbf{T}_i)\}\}])^{-1}E([\widetilde{\mathbf{Q}}_{\lambda i}\{\mathbf{w}_0(\mathbf{T}_i)\}^{\mathrm{T}}\\
&\times\boldsymbol{\Theta}_i\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_i)\}\boldsymbol{\Omega}_i^{-1}\boldsymbol{\Omega}_i^*\boldsymbol{\Omega}_i^{-1}\boldsymbol{\Theta}_i\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_i)\}\\
&\times\widetilde{\mathbf{Q}}_{\lambda i}\{\mathbf{w}_0(\mathbf{T}_i)\}\}])E([\widetilde{\mathbf{Q}}_{\lambda i}\{\mathbf{w}_0(\mathbf{T}_i)\}^{\mathrm{T}}\boldsymbol{\Theta}_i\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_i)\}\\
&\times\boldsymbol{\Omega}_i^{-1}\boldsymbol{\Theta}_i\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_i)\}\widetilde{\mathbf{Q}}_{\lambda i}\{\mathbf{w}_0(\mathbf{T}_i)\}\}])^{-1}\mathbf{B}_r(u),
\end{aligned}
$$

where $\boldsymbol{\Omega}_i^* = E\{(\mathbf{D}_i - \mathbf{H}_i)^{\otimes 2}|\mathbf{X}_i, \mathbf{Z}_i\}$ *is the true covariance matrix, and*

$$
\begin{aligned}
\sigma_w^2 \quad \equiv \quad & \frac{1}{n}\mathbf{B}_r^{\mathrm{T}}(u)E\Bigg\{\Bigg(\mathbf{V}^{-1}E\Bigg[\sum_{k=1}^{M_i}\sum_{v=1}^{M_i}E\Bigg\{C_{ikv}\Theta_{ik}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(T_{ik})\}\\
&\times\Theta_{iv}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(T_{iv})\}\mathbf{B}_r\{\mathbf{w}_0(T_{iv})^{\mathrm{T}}\mathbf{Z}_{iv}\}m_0'(\mathbf{w}_0(T_{ik})^{\mathrm{T}}\mathbf{Z}_{ik})\mathbf{Z}_{ik}^{\mathrm{T}}\\
&\times\Bigg(\{\mathbf{B}(T_{ik})f_T(T_{ik})\}^{-1}\Bigg[\mathbf{A}_{wj}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(T_{ik})\}\mathbf{V}_{wj}\{\boldsymbol{\beta}_0, m_0,\\
&\mathbf{w}_0(T_{ik})\}^{-1}\mathbf{K}_h(\mathbf{T}_j - T_{ik})\mathbf{S}_{wj}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(T_{ik})\}\Bigg]\Bigg)\Bigg|M_i,\\
&\mathbf{O}_{\mathbf{j}}\Bigg\}\Bigg|\mathbf{O}_{\mathbf{j}}\Bigg]\Bigg)^{\otimes 2}\Bigg\}\mathbf{B}_r(u).
\end{aligned}
$$

Here $\mathbf{V}$ *is as defined in the* **Notations in Step 1** *in Section A.2 in Appendix, and* C_{ikv} *is the* (k, v)*th entry of the matrix* $\boldsymbol{\Omega}_i^{-1}$*. Then we have*

$$
\{\sigma^2(u, \mathbf{w}_0) + \sigma_w^2\}^{-1/2}\left(\widehat{m}[u, \widehat{\boldsymbol{\lambda}}\{\widehat{\boldsymbol{\beta}}, \widehat{\mathbf{w}}(\widehat{\boldsymbol{\beta}})\}] - m_0(u)\right) \quad \stackrel{d}{\to} \quad N(0,1).
$$

Further because the order of σ^2 *and* σ_λ^2 *are both* $(nh_b)^{-1}$*, together with Fact 1 in Section S.2, we have*

$$
\begin{aligned}
|\widehat{m}[u, \widehat{\boldsymbol{\lambda}}\{\widehat{\boldsymbol{\beta}}, \widehat{\mathbf{w}}(\widehat{\boldsymbol{\beta}})\}] - m_0(u)| &= O_p\{(nh_b)^{-1/2} + h_b^q\},\\
|\widehat{m}'[u, \widehat{\boldsymbol{\lambda}}\{\widehat{\boldsymbol{\beta}}, \widehat{\mathbf{w}}(\widehat{\boldsymbol{\beta}})\}] - m_0'(u)| &= O_p\{n^{-1/2}h_b^{-3/2} + h_b^{q-1}\}
\end{aligned}
$$

uniformly for $u \in (0,1)$.

Theorem 3 shows that the estimation error of $\widehat{m}[u, \widehat{\boldsymbol{\lambda}}\{\widehat{\boldsymbol{\beta}}, \widehat{\mathbf{w}}(\widehat{\boldsymbol{\beta}})\}]$ consists of two components, the approximation error of $\widehat{m}[u, \widehat{\boldsymbol{\lambda}}\{\widehat{\boldsymbol{\beta}}, \widehat{\mathbf{w}}(\widehat{\boldsymbol{\beta}})\}]$ and the approximation error of $\widetilde{m}(u, \boldsymbol{\lambda}_0)$, from their respective true functions. The errors of $\widehat{m}$ and $\widehat{m}'$ go to zero with the rates of $O_p\{(nh_b)^{-1/2}\}$ and $O_p(n^{-1/2}h_b^{-3/2})$ respectively. Under Condition (A5), $\widehat{m}$ and $\widehat{m}'$ are both consistent, and they approach the truths with the standard B-spline convergence rate. We provide an outline of the proofs for Theorems 1-3 in the supplementary article (Jiang, Ma and Wang, 2015). The proofs are highly technical and lengthy, and they require several preliminary results which we summarize as lemmas. We present and prove these lemmas in the supplementary article (Jiang, Ma and Wang, 2015).

3. Numeric evaluation via simulations. We now evaluate the finite sample performance of the proposed estimation procedure on simulated data sets. We simulate 1000 data sets from Model (1) under three settings. In Settings 1 and 2, we consider binary response and use logit link function for H, while in Setting 3, we consider continuous normal response and use an identity H function. In Setting 1, we choose m as a polynomial function with degree two. We generate $\mathbf{w}$ initially as positive linear functions on t, and then normalize the vector to have summation one. Note that the normalization function modifies the structure of $\mathbf{w}(t)$ and results in a nonlinear vector-valued function in t. Additionally, we generate $\mathbf{Z}_{ik}$ from the Poisson distribution and normalize the vectors by the sample standard deviations. Furthermore, we generate T_{ik} from the exponential distribution and the covariate $\mathbf{X}_{ik}$ from the univariate normal distribution. In Settings 2 and 3, we use the sine function for m, and generate $\mathbf{w}$ as power functions on t and then normalize the vector to have summation one. We generate covariate vector $\mathbf{X}_i$ from a three-dimensional multivariate normal distribution. In order to stabilize the computation and control numerical errors, in both settings, we transform the function $\mathbf{w}(T_{ik})^{\mathrm{T}}\mathbf{Z}_{ik}$ to $F\{\mathbf{w}(T_{ik})^{\mathrm{T}}\mathbf{Z}_{ik}\} = \boldsymbol{\Phi}\left(\left[\mathbf{w}(T_{ik})^{\mathrm{T}}\mathbf{Z}_{ik} - E\{\mathbf{w}^0(T_{ik})^{\mathrm{T}}\mathbf{Z}_{ik}\}\right] / \sqrt{\mathrm{var}\{\mathbf{w}^0(T_{ik})^{\mathrm{T}}\mathbf{Z}_{ik}\}}\right)$, where $\mathbf{w}^0$ is the initial value of $\mathbf{w}$, and $E\{\mathbf{w}^0(T_{ik})^{\mathrm{T}}\mathbf{Z}_{ik}\}$ and $\mathrm{var}\{\mathbf{w}^0(T_{ik})^{\mathrm{T}}\mathbf{Z}_{ik}\}$ are approximated by the sample mean and the sample variance. We then use B-spline to approximate $m \circ F^{-1}$ instead of m, where $\circ$ denotes composite. All other operations remain the same, and the estimation and inference of the functional single index risk score $m\{\mathbf{w}(T_{ik})^{\mathrm{T}}\mathbf{Z}_{ik}\}$, our main research interest, is carried out as described before. To recover information regarding m, one can use the Delta method to obtain the estimate and the variance of estimating m from that of estimating $m \circ F^{-1}$.

In all the implementations, we use the third order quadratic spline. We select the number of internal knots $N = \{n^{1/5}(\log n)^2/5\}$ which satisfies the Condition (A5) in Section 2.3. We choose the Gaussian kernel with bandwidth $h = n^{-2/15}h_s$, where h_s is Silverman's rule-of-thumb bandwidth (Silverman, 1986). Because $h_s = O(n^{-1/5})$, the bandwidth selection satisfies Condition (A2) in Section 2.3.

Table 1 shows the averaged point estimators of $\boldsymbol{\beta}$, the empirical standard deviations calculated from the sample variances, the averages of the estimated asymptotic standard deviation ($\Sigma^{1/2}$ in Theorem 2) over the simulated samples, and the mean squared errors (MSE) when the sample sizes are 100, 500, 800, respectively. The conclusions are similar under the three settings. To sum up, the estimation biases are consistently small across all samples sizes, the empirical standard deviations and the estimated asymptotic standard deviations are decreasing when the sample size increases. The MSE decreases as the sample size increases as well, mainly due to the declining variations. Further, the empirical standard deviation of the estimators and average of the estimated standard deviations calculated from the asymptotic results are close. In addition, the coverage probabilities of the empirical confidence intervals are close to the normal level 95%. This suggests that we can use the asymptotic properties to perform inference and can obtain sufficiently reliable results under moderate sample sizes.

Setting 1

β_0	$E(\widehat{\beta})$	$\mathrm{sd}(\widehat{\beta})$	$\widehat{\mathrm{sd}}(\widehat{\beta})$	MSE	CP
n = 100					
-0.2	-0.202	0.157	0.113	0.0247	0.957
-0.4	-0.398	0.119	0.115	0.0142	0.940
-0.6	-0.601	0.124	0.118	0.0153	0.957
n = 500					
-0.2	-0.198	0.052	0.050	0.0027	0.954
-0.4	-0.398	0.053	0.051	0.0028	0.947
-0.6	-0.601	0.056	0.053	0.0031	0.939
n = 800					
-0.2	-0.197	0.041	0.040	0.0017	0.951
-0.4	-0.398	0.041	0.040	0.0017	0.949
-0.6	-0.602	0.044	0.042	0.0019	0.946

Setting 2

	$\boldsymbol{\beta}_0$	$E(\widehat{\boldsymbol{\beta}})$	$\mathrm{sd}(\widehat{\boldsymbol{\beta}})$	$\widehat{\mathrm{sd}}(\widehat{\boldsymbol{\beta}})$	MSE	CP
n = 100						
β_1	-0.5	-0.505	0.131	0.116	0.0171	0.908
β_2	0.2	-0.200	0.122	0.112	0.0147	0.923
β_3	0.5	-0.515	0.125	0.116	0.0159	0.927
n = 500						
β_1	-0.5	-0.507	0.056	0.053	0.0032	0.946
β_2	0.2	-0.198	0.053	0.052	0.0028	0.951
β_3	0.5	-0.508	0.054	0.053	0.0031	0.944
n = 800						
β_1	-0.5	-0.504	0.043	0.042	0.0019	0.953
β_2	0.2	-0.202	0.041	0.041	0.0017	0.962
β_3	0.5	-0.505	0.043	0.042	0.0019	0.951

Setting 3

	$\boldsymbol{\beta}_0$	$E(\widehat{\boldsymbol{\beta}})$	$\mathrm{sd}(\widehat{\boldsymbol{\beta}})$	$\widehat{\mathrm{sd}}(\widehat{\boldsymbol{\beta}})$	MSE	CP
n = 100						
β_1	-0.5	-0.501	0.062	0.052	3.85e-3	0.938
β_2	0.2	-0.200	0.060	0.053	3.60e-3	0.932
β_3	0.5	-0.503	0.061	0.053	3.73e-3	0.932
n = 500						
β_1	-0.5	-0.500	0.025	0.024	6.25e-4	0.966
β_2	0.2	-0.200	0.024	0.024	5.76e-4	0.945
β_3	0.5	-0.502	0.025	0.024	6.29e-4	0.963
n = 800						
β_1	-0.5	-0.500	0.020	0.019	4.00e-4	0.949
β_2	0.2	-0.200	0.019	0.019	3.61e-4	0.949
β_3	0.5	-0.501	0.020	0.019	4.01e-4	0.952

TABLE 1

Simulation results in Setting 1, 2, 3, based on 1000 data sets. The true parameter β_0, mean (E), empirical standard deviation ($sd(\widehat{\beta})$) and average of the estimated standard deviations ($\widehat{sd}(\widehat{\beta})$) $MSE = \{sd(\widehat{\beta})\}^2 + \{E(\widehat{\beta}) - \beta\}^2$, the coverage probabilities (CP) of the 95% empirical confidnece intervals are reported.

We also examined the performances of $\widehat{\mathbf{w}}$ and $\widehat{m}$ to assess the properties of the estimated functional single index risk score. Under the first setting, because the functional single index risk score is fixed with respect to β, we only evaluate the settings with $\beta = -0.4$. To evaluate the combined score $\widehat{\mathbf{w}}(t)^{\mathrm{T}}\mathbf{Z}$ as a function of t, we fix $\mathbf{Z}$ at $\mathbf{Z}^* = (1, 2, 3, 4)$ and plot the averages of the estimated combined score $\widehat{\mathbf{w}}(t)^{\mathrm{T}}\mathbf{Z}^*$ over the 1000 simulations around the true scores $\mathbf{w}_0(t)^{\mathrm{T}}\mathbf{Z}^*$ in the upper panels of Figure 1, 2, and 3 for Setting 1, 2 and 3, respectively. Additionally, we present the 95% point wise confidence band. The results show that the estimates are close to the true function. Further, the 95% confidence band becomes narrower when the sample size increases, which indicates that the estimation variation decreases with increased sample size. Moreover, we evaluated the coverage probabilities of the empirical pointwise confidence bands of $\mathbf{w}$, by computing the coverage probabilities at a set of fixed points across t and taking their average. The average coverage probabilities for $n = 100, 500, 800$ are $0.934, 0.936, 0.939$ in Setting 1, $0.939, 0.940, 0.941$ in Setting 2, and $0.931, 0.934, 0.936$ in Setting 3, respectively. All are reasonably close to the nominal level of 95%.

To evaluate the performance of $\widehat{m}$, we plot the average of $\widehat{m}(u)$ based on the 1000 simulations, as well as the 95% point wise confidence band in the bottom panels of Figure 1, 2, and 3 for Setting 1, 2, and 3, respectively. The plots show that the estimators are close to the true functions except on the boundary when the sample size is relatively small. In addition, when the sample size increases, the confidence band becomes narrower, benefiting from the smaller estimation variation. Note that because of the additional transformation on $\mathbf{w}(t)^{\mathrm{T}}\mathbf{Z}$, it is not unexpected that the true m function does not appear to be periodic sine function on $\mathbf{w}(t)^{\mathrm{T}}\mathbf{Z}$. Moreover, we evaluate the converge probability of the empirical pointwise confidence bands of m. The average coverage probabilities are $0.943, 0.947, 0.948$ in Setting 1, $0.957, 0.960, 0.951$ in Setting 2, and $0.939, 0.947, 0.946$ in Setting 3, respectively. Again, they are all fairly close to the nominal level of 95%.

In summary, Table 1, Figures 1, 2, 3 illustrate the desirable finite sample performance of the fused kernel/B-spline combination method in estimating $\boldsymbol{\beta}, m$ and $\mathbf{w}$. In terms of parameter estimation and function estimation in the non-boundary region, the estimators show very small biases across all sample sizes, and decreasing variability as the sample size increases. The asymptotic variance and sample empirical variance in estimating $\boldsymbol{\beta}$ are close. Furthermore, the coverage probability of the empirical confidence intervals for $\boldsymbol{\beta}$ and the coverage probability of the empirical pointwise confidence bands for $\mathbf{w}$ and m are close to the nominal levels, which supports using the asymptotic results for the subsequent inferences.

4. Application. We apply the functional single index risk score model and the fused kernel/B-spline semiparametric estimation method to analyze a real data set from a Huntington's disease (HD) study. Current research in HD aims to find reliable prodromes to enable early detection of HD. The joint effect of the cognitive scores on odds of HD diagnosis is shown to change with time. In addition, the relationship between the cognitive symptoms and the log-odds of the disease diagnosis is shown to be nonlinear (Paulsen et al.,

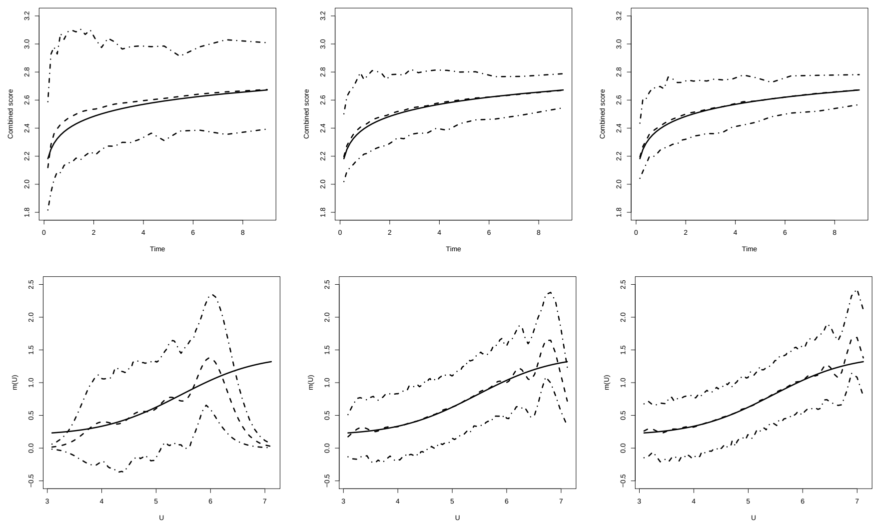

Fig 1: Estimation of $\mathbf{w}(t)^{\mathrm{T}}\mathbf{z}$ (upper) and $m(u)$ (bottom) as a function of t and u, respectively in Setting 1 with sample sizes 100 (left), 500 (middle) and 800 (right). True function (solid line), average of 1000 estimated functions (dashed lines), and 95% point wise confidence band (dash-doted lines) are provided.

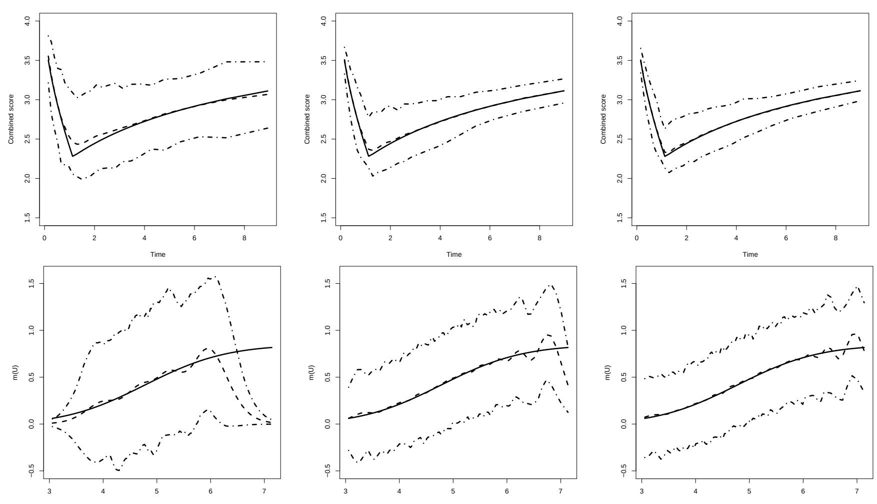

Fig 2: Estimation of $\mathbf{w}(t)^{\mathrm{T}}\mathbf{z}$ (upper) and $m(u)$ (bottom) as a function of t and u, respectively in in Setting 2 with sample sizes 100 (left), 500 (middle) and 800 (right). True function (solid line), average of 1000 estimated functions (dashed lines), and 95% point wise confidence band (dash-doted lines) are provided.

2008). Our goal is to study the nonlinear time dependent cognitive effects

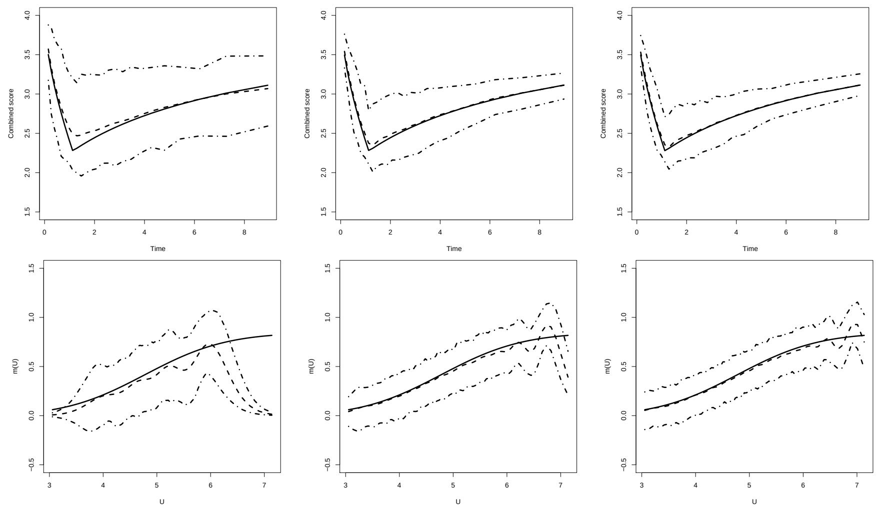

Fig 3: Estimation of $\mathbf{w}(t)^{\mathrm{T}}\mathbf{z}$ (upper) and $m(u)$ (bottom) as a function of t and u, respectively in Setting 3 with sample sizes 100 (left), 500 (middle) and 800 (right). True function (solid line), average of 1000 estimated functions (dashed lines), and 95% point wise confidence band (dash-doted lines) are provided.

so as to facilitate the early detection of HD.

Specifically, let D_{ik}, $\mathbf{Z}_{ik}$, and $\mathbf{X}_{ik}$ represent the binary disease indicator, the cognitive score vector, and the additional covariate vector for the ith individual at the jth measurement time, respectively. The cognitive scores include SDMT (Smith, 1982), stroop color, stroop word, and stroop interference tests (Stroop, 1935). They are denoted by $Z_{i1}, \ldots, Z_{i4}$, respectively. The covariates of interest are gender, education, CAP score (Zhang et al., 2011). They are denoted by $X_{i1}, \ldots, X_{i3}$, respectively. The subject's age at the visiting time serves as the time variable T_{ik}. We normalize the continuous variables to the interval $(0, 1)$ to alleviate numerical instability. Without changing notations, we transform $Z_{i1}, \ldots, Z_{i4}$, X_{i3}, T_{ik} by the normal distribution functions with means and variances estimated from the sample.

We use logit link function to model the binary outcomes, i.e., we assume

$$(6)\, H[m\{\mathbf{w}(T_{ik})^{\mathrm{T}}\mathbf{Z}_{ik}\} + \boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}_{ik}] = \frac{\exp\left[m\{\mathbf{w}(T_{ik})^{\mathrm{T}}\mathbf{Z}_{ik}\} + \boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}_{ik}\right]}{1 + \exp\left[m\{\mathbf{w}(T_{ik})^{\mathrm{T}}\mathbf{Z}_{ik}\} + \boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}_{ik}\right]}.$$

We obtain the initial estimates and working correlation matrix using the GEE method with exchangeable covariance assumption. We choose the exchangeable covariance structure because in our setting, it facilitates computation while also accounts for the longitudinal correlations. Let the working correlation coefficient matrix be $\mathbf{R}_i$, the working covariance matrix be $\widehat{\Theta}_i^{1/2}\mathbf{R}_i\widehat{\Theta}_i^{1/2}$, where $\widehat{\Theta}_i$ is $\mathbf{H}_i(1-\mathbf{H}_i)$ with estimated $\widehat{\boldsymbol{\lambda}}, \widehat{\mathbf{w}}, \widehat{\boldsymbol{\beta}}$ plugged in.

We implement the profiling procedure described in Section 2.2 in the subsequent estimation. The kernel and B-spline functions are defined in the same way as described in Section 3. We obtain the point estimators $\widehat{\boldsymbol{\beta}} = (-0.34, -0.89, 2.31)^{\mathrm{T}}$ and the asymptotic variances $\mathrm{var}(\widehat{\boldsymbol{\beta}}) = (0.0035, 0.00044, 0.011)^{\mathrm{T}}$. Consequently, the 95% asymptotic confidence intervals are $\{(-0.46, -0.23), (-0.93, -0.85), (2.09, 2.52)\}$, which demonstrate the significant effect of gender, education level, and CAP score on the disease risk. Specifically, female ($X_{i1} = 0$) tends to have higher disease risk than male ($X_{i1} = 1$). In addition, patients with lower education levels and higher CAP scores are more likely to develop Huntington's disease, which is consistent with the clinical literature (Zhang et al., 2011).

We also plot $\widehat{\mathbf{w}}(t)$ to show the variation patterns of the effect of the four cognitive scores over time. Figure 4 shows that the stroop interference score has more important effect than all the others after age 30. The 95% point wise confidence interval remains above the 0.25 level after age 27, and the stroop interference score effect largely dominates all the other effects during that period. This dominating effect indicates that the stroop inference score has the closest relationship with the onset of HD, and in turn could be used to predict HD most effectively among the four. Further, stroop color has large effect at earlier ages (before 30 or at early 30s), while the SDMT has reasonably large effect at later ages (75 or above). Moreover, stroop word have relatively small predicative effects (< 0.25) on the disease risk across all ages. The plots clearly show the time dependent nature of the cognitive score effects. More specifically, stroop color effect is decreasing over times, stroop interference effect is a concave function of time, while SDMT, stroop word effects are convex functions of time. The last three non-monotone effects reach their extreme values around the ages of 40 to 50. In summary, the results show that the stroop interference is more relevant to the disease risk than the other scores. Further, the relative magnitude of the score effects clearly change over time, which suggests the need to closely monitor specific cognitive scores for different age groups. This illustrates the importance of modeling $\mathbf{w}$ as a function of age, and the convenience of using a weighted score $\mathbf{w}(t)^{\mathrm{T}}\mathbf{Z}$ as a combined cognitive profile in practice.

The form of the function $\widehat{m}$ is shown in the left panel of Figure 5. We also plot the 95% point wise asymptotic confidence band of $\widehat{m}$ in the range of the combined scores U. The plot shows that the functional single index risk score is a decreasing function of the index. The upper confidence interval does not include 0, which shows that the functional single index risk score is significantly smaller than 0 at any age and cognitive score values in this population.

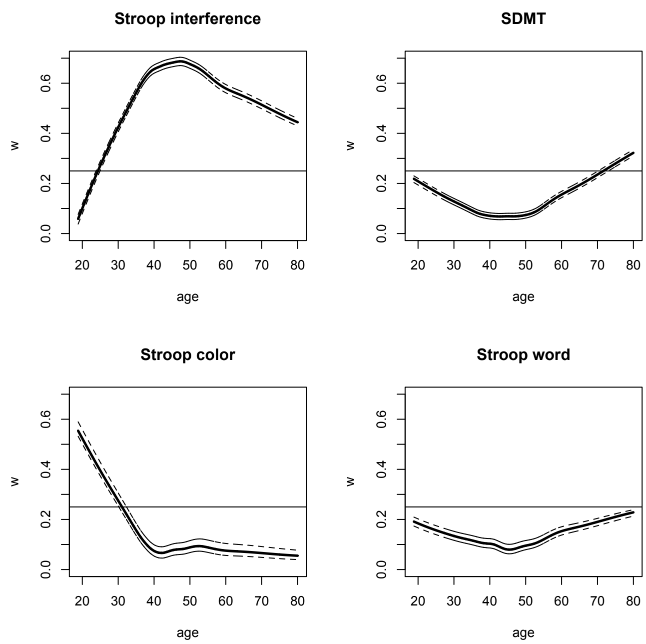

Fig 4: Estimation of the weight function $w(t)$'s and the 95% asymptotic confidence bands in Huntington's disease data. The reference line is 0.25.

In the right panel of Figure 5, we plot the disease risk (the estimated probability of $D = 1$) and the 95% point wise asymptotic confidence band, where the confidence band is based on estimated variance, calculated using the Delta method and the estimated variance of $\widehat{m}$. The results show that the disease risk decreases with the combined cognitive score value U. The 95% confidence interval does not include the 0.5 line, which shows that the disease risk in the population is smaller than 0.5 across all age and cognitive score values. Combining the two plots, Figure 5 shows that a higher value of the combined score $U = \mathbf{w}(t)^{\mathrm{T}}\mathbf{Z}$, which implies better cognitive functioning, tends to lower functional single index risk score and in turn lower the risk of HD. The effect of the functional single index cognitive risk score on HD diagnosis is approximately quadratic for a standardized score $U < 0.6$, and is approximately a constant for $U > 0.6$. The flattening of the effect reflects a ceiling effect for subjects with better cognitive performance.

Next, we perform two sensitivity analyses to justify using a more flexible generalized partially linear functional single index model as shown in (6). We compare Model (6) with two simpler models. The first one assumes the function m is linear, hence

$$(7)\, H\{\mathbf{X}_{ik}, \mathbf{Z}_{ik}; \boldsymbol{\theta}, \mathbf{w}(T_{ik})\} = \frac{\exp\{\alpha_c + \alpha_1 \mathbf{w}(T_{ik})^{\mathrm{T}}\mathbf{Z}_{ik} + \boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}_{ik}\}}{1 + \exp\{\alpha_c + \alpha_1 \mathbf{w}(T_{ik})^{\mathrm{T}}\mathbf{Z}_{ik} + \boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}_{ik}\}},$$

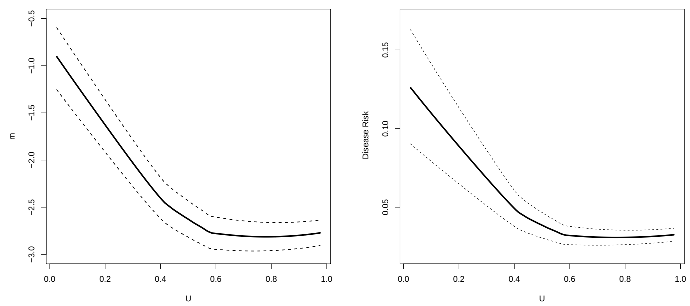

Fig 5: Function $\widehat{m}(u)$ (left) and the estimated disease risk as a function of u (right) in Huntington's disease data.

where α_c, α_1 are unknown parameters. The second one assumes the weight function $\mathbf{w}$ is time-invariant, hence

$$H(\mathbf{X}_{ik}, \mathbf{Z}_{ik}; \boldsymbol{\theta}, \mathbf{w}) = \frac{\exp\{m(\mathbf{w}^{\mathrm{T}}\mathbf{Z}_{ik}) + \boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}_{ik}\}}{1 + \exp\{m(\mathbf{w}^{\mathrm{T}}\mathbf{Z}_{ik}) + \boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}_{ik}\}}, \tag{8}$$

where $\mathbf{w}$ is an unknown parameter vector. We carried out the estimation of $\mathbf{w}(t)$ in the first model using kernel method and the estimation for m in the second model via B-spline method. We implemented 1000 5-fold cross validation analysis. We evaluated models by the mean squared predictive error (i.e., the mean squared differences between D_i and the predicted probability of $D_i = 1$ on the test set) as a function of the average of the four standardized cognitive scores $\sum_j^4 Z_j/4$, which we named the standardized score. In Figure 6, we plot the mean squared predictive error curves obtained under the proposed Model (6) and two simpler models. The results show that our original generalized partially linear model with functional single index outperforms Model (8) uniformly across the range of the standardized scores in terms of a lower mean squared error. We also plot the empirical 95% confidence intervals of the squared predictive errors under the proposed model. Compared with the simpler Model (7), our model gives significant smaller predictive errors when the standardized score is smaller than 0.36. The medians of the squared predicative errors in this range are 0.040 and 0.049 for the models (6) and (7), respectively. When the standardized score is greater than 0.5, Model (7) performs slightly, but not significantly, better than Model (6). Overall, the total mean squared error summarized by the area under the predictive error curves for models (6), (7) and (8) are respectively, 0.022, 0.028, and 0.057, which justify using the more flexible model in (6) to fit the Huntington's disease data. The results also demonstrate the potential of using our method as an exploratory tool to assess general patterns of data.

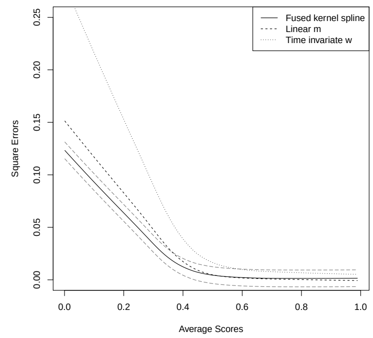

Fig 6: The mean squared predictive errors versus the standardized averaged score $\sum_{j=1}^{4} Z_{ik}/4$ in Huntington's disease data. The gray lines are the 95% confidence intervals for the fused kernel/B-spline method.

5. Conclusion and discussions. We have developed a generalized partially linear functional single index risk score model in the longitudinal data framework. We explore the relationship between the cognitive scores and the disease risk so as to predict HD diagnosis early, and in turn to intervene with the disease progression in a timely manner.

We introduce a framework of jointly using the B-spline and kernel methods in semiparametric estimation. We use B-spline to approximate the functional single index risk score function m, and use kernel smoothing technique for estimating the cognitive weight functions of time $\mathbf{w}(t)$. We integrate B-spline basis expansion, kernel smoothing and longitudinal analysis, and have proven the consistency and asymptotic normalities of the covariate coefficient estimators, the time dependent weight function estimators, and the single index risk score function estimators. The derivation relies on the assumption that the iteration procedure converges to a parameter vector value that is in a small neighborhood of the truth, which generally requires the estimating equation to have a unique zero. The unique zero property is difficult to guarantee in theory and is less likely to hold when sample size is small or moderate. To this end, empirical knowledge is usually used to select a suitable root. In our simulations, multiple roots issue did not occur and the numerical results show desirable finite sample properties of the estimators. The real data analysis yields results which are interpretable and useful in practice. In summary, the functional single index model provides rich and meaningful information regarding the association between the disease risk and the cognitive score profiles. It is of course also possible to use B-spline or kernel methods to estimate both m and $\mathbf{w}(t)$, research along this line can also be interesting.

Our method accommodates both continuous and categorical response variables as long as the link function H is continuously differentiable and has finite second derivative. One outstanding research question in these models, even in the context when the marginal model is completely parametric (for example, both m and $\mathbf{w}$ are known), is the estimation efficiency. As far as we are aware, there is no guarantee that GEE family contains the efficient estimator, and how to obtain asymptotically efficient estimator certainly worth further research.

The proposed generalized partially linear functional single index model can be used to incorporate high dimensional data, since the single index risk score is a natural method to alleviate the curse of the dimensionality. For example, the single index score could be a combination of gene expression covariates to facilitate the genetic association study. Furthermore, the generalized partially linear functional single index risk score can be used in an adaptive randomization clinical trial study to improve study efficiency. For example, we can use a single index risk score to summarize some disease related biomarkers which provide early information about the primary endpoints in adaptive trials. When a trial progresses, the information can be used to make certain intermediate decisions, such as treatment assignments among the patients, and stopping or continuation of the trial.

APPENDIX A.1: PROOF OF PROPOSITION 1

Assume there exist $m_1 \in \mathcal{M}, \mathbf{w}_1(t) \in \mathcal{D}$ and $\boldsymbol{\beta}_1 \in \mathbb{R}^{d_\beta}$, such that

$$m_1\{\mathbf{w}_1^{\mathrm{T}}(t)\mathbf{Z}\} + \boldsymbol{\beta}_1^{\mathrm{T}}\mathbf{X} = m_0\{\mathbf{w}_0^{\mathrm{T}}(t)\mathbf{Z}\} + \boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}, \tag{9}$$

where $m_0, \mathbf{w}_0(t)$ and $\boldsymbol{\beta}_0$ are the true parameter values. Taking derivative with respect to $\mathbf{Z}$ and t on both sides of the equation, we obtain

$$\begin{aligned} m_1'\{\mathbf{w}_1^{\mathrm{T}}(t)\mathbf{Z}\}\mathbf{w}_1(t) &= m_0'\{\mathbf{w}_0^{\mathrm{T}}(t)\mathbf{Z}\}\mathbf{w}_0(t), \\ m_1'\{\mathbf{w}_1^{\mathrm{T}}(t)\mathbf{Z}\}\mathbf{w}_1'(t)^{\mathrm{T}}\mathbf{Z} &= m_0'\{\mathbf{w}_0^{\mathrm{T}}(t)\mathbf{Z}\}\mathbf{w}_0'(t)^{\mathrm{T}}\mathbf{Z}. \end{aligned} \tag{10}$$

Because m_1, m_0 are one-to-one, $m_1'\{\mathbf{w}_1^{\mathrm{T}}(t)\mathbf{Z}\} = m_0'\{\mathbf{w}_0^{\mathrm{T}}(t)\mathbf{Z}\} = 0$ can hold only for a set of discrete set of $\mathbf{w}_1^{\mathrm{T}}(t)\mathbf{Z}$ and $\mathbf{w}_0^{\mathrm{T}}(t)\mathbf{Z}$ values, hence a discrete set of t values. Thus, due to the continuity of $m_1', m_0', \mathbf{w}_1$ and $\mathbf{w}_0$, (10) implies $\mathbf{w}_1'(t)^{\mathrm{T}}\mathbf{Z}/w_{1j}(t) = \mathbf{w}_0'(t)^{\mathrm{T}}\mathbf{Z}/w_{0j}(t)$ for all $j = 1, \ldots, d_w$, all $\mathbf{Z}$, and all $t \in [0, \tau]$. Thus, $\mathbf{w}_1'(t)^{\mathrm{T}}E(\mathbf{Z}^{\otimes 2})/w_{1j}(t) = \mathbf{w}_0'(t)^{\mathrm{T}}E(\mathbf{Z}^{\otimes 2})/w_{0j}(t)$. Furthermore, $E(\mathbf{Z}^{\otimes 2})$ is positive definite and in turn is invertible, it leads to $\mathbf{w}_1'(t)/w_{1j}(t) = \mathbf{w}_0'(t)/w_{0j}(t)$. In particular, we have $w_{1j}'(t)/w_{1j}(t) = w_{0j}'(t)/w_{0j}(t)$ for all $j = 1, \ldots, d_w$. This gives $w_{1j}(t) = w_{0j}(t)c_j$ for some constant c_j, or equivalently, $\mathbf{w}_1(t) = \mathbf{C}\mathbf{w}_0(t)$ where $\mathbf{C}$ is a diagonal matrix with c_j's on the diagonal. Taking derivative with respect to t, we further have $\mathbf{w}_1'(t) = \mathbf{C}\mathbf{w}_0'(t)$. Dividing $w_{1j}(t)$ on both sides, we have $\mathbf{w}_1'(t)/w_{1j}(t) = (\mathbf{C}/c_j)\mathbf{w}_0'(t)/w_{0j}(t)$. Therefore, $\mathbf{C}/c_j$ is the identity matrix. In other words, $c_j, j = 1, \dots, d_w$ are identical. Since $\|\mathbf{w}_1(t)\|_1 = \|\mathbf{w}_0(t)\|_1 = 1$ and $\mathbf{w}_1(t)$, $\mathbf{w}_0(t)$ are positive, this further implies $\mathbf{w}_1(t) = \mathbf{w}_0(t)$. Therefore, (10) reduces to $m_1'\{\mathbf{w}_0^{\mathrm{T}}(t)\mathbf{Z}\} - m_0'\{\mathbf{w}_0^{\mathrm{T}}(t)\mathbf{Z}\} = 0$. This further implies $m_1\{\mathbf{w}_0^{\mathrm{T}}(t)\mathbf{Z}\} = m_0\{\mathbf{w}_0^{\mathrm{T}}(t)\mathbf{Z}\} + C_1$ for a constant C_1. Because $m_1(0) = m_0(0) = c_0$, $C_1 = 0$, i.e. $m_1 = m_0$. (9) now leads to $\boldsymbol{\beta}_1^{\mathrm{T}}\mathbf{X} = \boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}$. The equality holds for any $\mathbf{X}$, which implies $\boldsymbol{\beta}_1^{\mathrm{T}} E(\mathbf{X}^{\otimes 2}) = \boldsymbol{\beta}_0^{\mathrm{T}} E(\mathbf{X}^{\otimes 2})$. Since $E(\mathbf{X}^{\otimes 2})$ is positive definite, and in turn is invertible, we have $\boldsymbol{\beta}_1 = \boldsymbol{\beta}_0$. Therefore, we have $\boldsymbol{\beta}_1 = \boldsymbol{\beta}_0$, $\mathbf{w}_1(t) = \mathbf{w}_0(t)$, and $m_1 = m_0$, hence the problem is identifiable. $\square$

APPENDIX A.2: NOTATION IN ESTIMATION STEP

Notation in Step 1. We define an $M_i \times d_\lambda$ matrix

$$\widetilde{\mathbf{Q}}_{\lambda i}\{\mathbf{w}(\mathbf{T}_i)\} = \begin{bmatrix} B_{r1}\{\mathbf{w}(T_{i1})^{\mathrm{T}}\mathbf{Z}_{i1}\} & \dots & B_{rd_\lambda}\{\mathbf{w}(T_{i1})^{\mathrm{T}}\mathbf{Z}_{i1}\} \\ \vdots & \vdots & \vdots \\ B_{r1}\{\mathbf{w}(T_{iM_i})^{\mathrm{T}}\mathbf{Z}_{iM_i}\} & \dots & B_{rd_\lambda}\{\mathbf{w}(T_{iM_i})^{\mathrm{T}}\mathbf{Z}_{iM_i}\} \end{bmatrix},$$

and define $\widetilde{\mathbf{Q}}_{\lambda i}\{\mathbf{w}(t_0)\}$ to be the same as $\widetilde{\mathbf{Q}}_{\lambda i}\{\mathbf{w}(\mathbf{T}_i)\}$ except we replace $T_{ik}, k = 1, \dots, M_i$ with t_0. Here and throughout the text, replacing $\mathbf{T}_i$ by t_0 means replace $T_{ik} = t_0$ for each $k, k =, 1, \dots, M_i$. Let

$$\begin{aligned} \mathbf{V}_n &= n^{-1}\sum_{i=1}^{n}[\widetilde{\mathbf{Q}}_{\lambda i}\{\mathbf{w}_0(\mathbf{T}_i)\}^{\mathrm{T}}\boldsymbol{\Theta}_i\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_i)\}\boldsymbol{\Omega}_i^{-1} \\ &\quad \times\boldsymbol{\Theta}_i\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_i)\}\widetilde{\mathbf{Q}}_{\lambda i}\{\mathbf{w}_0(\mathbf{T}_i)\}], \\ \mathbf{V} &= E([\widetilde{\mathbf{Q}}_{\lambda i}\{\mathbf{w}_0(\mathbf{T}_i)\}^{\mathrm{T}}\boldsymbol{\Theta}_i\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_i)\}\boldsymbol{\Omega}_i^{-1} \\ &\quad \times\boldsymbol{\Theta}_i\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_i)\}\widetilde{\mathbf{Q}}_{\lambda i}\{\mathbf{w}_0(\mathbf{T}_i)\}]). \end{aligned}$$

Notation in Step 2. We define $\widehat{\mathbf{S}}_{wik}\{\boldsymbol{\beta}_0, \widehat{\boldsymbol{\lambda}}(\boldsymbol{\beta}_0, \mathbf{w}), \mathbf{w}(t_0)\}$ as

$$\begin{aligned} &\left[\mathbf{Q}_{wik}\{\widehat{\boldsymbol{\lambda}}(\boldsymbol{\beta}_0, \mathbf{w}), \mathbf{w}(t_0)\} + \mathbf{Q}_{\lambda ik}\{\mathbf{w}(t_0)\}^{\mathrm{T}}\left\{\frac{\partial\widehat{\boldsymbol{\lambda}}(\boldsymbol{\beta}_0, \mathbf{w})}{\partial\mathbf{w}}\right\}\right] \\ &\times[D_{ik} - H_{ik}\{\boldsymbol{\beta}_0, \widehat{\boldsymbol{\lambda}}(\boldsymbol{\beta}_0, \mathbf{w}), \mathbf{w}(t_0)\}], \end{aligned}$$

and $\widehat{\mathbf{S}}_{wi}\{\boldsymbol{\beta}_0, \widehat{\boldsymbol{\lambda}}(\boldsymbol{\beta}_0, \mathbf{w}), \mathbf{w}(t_0)\} = [\widehat{\mathbf{S}}_{wik}\{\boldsymbol{\beta}_0, \widehat{\boldsymbol{\lambda}}(\boldsymbol{\beta}_0, \mathbf{w}), \mathbf{w}(t_0)\}^{\mathrm{T}}, k = 1, \dots, M_i]^{\mathrm{T}}$.

We now define a functional from $\mathcal{D}$ to $\mathbb{R}^{d_w}$, so that this functional evaluated at $\mathbf{w}_h$ is $\mathbf{Q}_{wik}\{\widehat{\boldsymbol{\lambda}}(\boldsymbol{\beta}_0, \mathbf{w}), \mathbf{w}(t_0)\}^{\mathrm{T}}\mathbf{w}_h(t_0)$. For notational brevity, we still use $\mathbf{Q}_{wik}\{\widehat{\boldsymbol{\lambda}}(\boldsymbol{\beta}_0, \mathbf{w}), \mathbf{w}(t_0)\}$ to denote this functional, i.e.

$$\mathbf{Q}_{wik}\{\widehat{\boldsymbol{\lambda}}(\boldsymbol{\beta}_0, \mathbf{w}), \mathbf{w}(t_0)\}(\mathbf{w}_h) \equiv \mathbf{Q}_{wik}\{\widehat{\boldsymbol{\lambda}}(\boldsymbol{\beta}_0, \mathbf{w}), \mathbf{w}(t_0)\}^{\mathrm{T}}\mathbf{w}_h(t_0).$$

Let $\widehat{\mathbf{A}}_{wi}\{\boldsymbol{\beta}_0, \widehat{\boldsymbol{\lambda}}(\boldsymbol{\beta}_0, \mathbf{w}), \mathbf{w}(t_0)\}$ be a $d_w \times d_w M_i$ matrix, with the kth size $d_w \times d_w$ column block $\widehat{\mathbf{A}}_{wik}\{\boldsymbol{\beta}_0, \widehat{\boldsymbol{\lambda}}(\boldsymbol{\beta}_0, \mathbf{w}), \mathbf{w}(t_0)\}$ being

$$\left[\mathbf{Q}_{wik}\{\widehat{\boldsymbol{\lambda}}(\boldsymbol{\beta}_0, \mathbf{w}), \mathbf{w}(t_0)\} + \mathbf{Q}_{\lambda ik}\{\mathbf{w}(t_0)\}^{\mathrm{T}} \left\{\frac{\partial \widehat{\boldsymbol{\lambda}}(\boldsymbol{\beta}_0, \mathbf{w})}{\partial \mathbf{w}}\right\}\right]^{\otimes 2}$$
$$\times \Theta_{ik}\{\boldsymbol{\beta}_0, \widehat{\boldsymbol{\lambda}}(\boldsymbol{\beta}_0, \mathbf{w}), \mathbf{w}(t_0)\}.$$

Let $\widehat{\mathbf{V}}_{wi}\{\boldsymbol{\beta}_0, \widehat{\boldsymbol{\lambda}}(\boldsymbol{\beta}_0, \mathbf{w}), \mathbf{w}(t_0)\}$ be a $d_w M_i \times d_w M_i$ matrix with the (p, q)th block $\widehat{\mathbf{V}}_{wipq}\{\boldsymbol{\beta}_0, \widehat{\boldsymbol{\lambda}}(\boldsymbol{\beta}_0, \mathbf{w}), \mathbf{w}(t_0)\}$ being

$$\left[\mathbf{Q}_{wip}\{\widehat{\boldsymbol{\lambda}}(\boldsymbol{\beta}_0, \mathbf{w}), \mathbf{w}(t_0)\} + \mathbf{Q}_{\lambda ip}\{\mathbf{w}(t_0)\}^{\mathrm{T}} \left\{\frac{\partial \widehat{\boldsymbol{\lambda}}(\boldsymbol{\beta}_0, \mathbf{w})}{\partial \mathbf{w}}\right\}\right]$$
$$\times \left[\mathbf{Q}_{wiq}\{\widehat{\boldsymbol{\lambda}}(\boldsymbol{\beta}_0, \mathbf{w}), \mathbf{w}(t_0)\} + \mathbf{Q}_{\lambda iq}\{\mathbf{w}(t_0)\}^{\mathrm{T}} \left\{\frac{\partial \widehat{\boldsymbol{\lambda}}(\boldsymbol{\beta}_0, \mathbf{w})}{\partial \mathbf{w}}\right\}\right]^{\mathrm{T}} \boldsymbol{\Omega}_{ipq},$$

where $\boldsymbol{\Omega}_{ipq}$ is the (p, q)th element of the working covariance matrix $\boldsymbol{\Omega}_i$.

We further define the population level quantities $\mathbf{S}_{wik}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(t_0)\}$ to be

$$[\mathbf{Z}_{ik} m_0'\{\mathbf{w}_0(t_0)^{\mathrm{T}} \mathbf{Z}_{ik}\} - \eta\{\mathbf{w}_0(t_0)^{\mathrm{T}} \mathbf{Z}_{ik}\}][D_{ik} - H_{ik}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(t_0)\}]$$

and $\mathbf{S}_{wi}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(t_0)\} = [\mathbf{S}_{wik}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(t_0)\}^{\mathrm{T}}, k = 1, \ldots, M_i]^{\mathrm{T}}$. Let $\mathbf{A}_{wi}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(t_0)\}$ be a $d_w \times d_w M_i$ matrix, with the kth column block $\mathbf{A}_{wik}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(t_0)\}$ being a $d_w \times d_w$ matrix

$$[\mathbf{Z}_{ik} m_0'\{\mathbf{w}_0(t_0)^{\mathrm{T}} \mathbf{Z}_{ik}\} - \eta\{\mathbf{w}_0(t_0)^{\mathrm{T}} \mathbf{Z}_{ik}\}]^{\otimes 2} \Theta_{ik}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(t_0)\}.$$

Let $\mathbf{V}_{wi}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(t_0)\}$ be a $d_w M_i \times d_w M_i$ matrix with the (p, q)th block $\mathbf{V}_{wipq}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(t_0)\}$ being

$$[\mathbf{Z}_{ip} m_0'\{\mathbf{w}_0(t_0)^{\mathrm{T}} \mathbf{Z}_{ip}\} - \eta\{\mathbf{w}_0(t_0)^{\mathrm{T}} \mathbf{Z}_{ip}\}]$$
$$\times [\mathbf{Z}_{iq} m_0'\{\mathbf{w}_0(t_0)^{\mathrm{T}} \mathbf{Z}_{iq}\} - \eta\{\mathbf{w}_0(t_0)^{\mathrm{T}} \mathbf{Z}_{iq}\}]^{\mathrm{T}} \boldsymbol{\Omega}_{ipq}.$$

Let $\mathbf{V}^*{}_{wi}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(t_0)\}$ be a $d_w M_i \times d_w M_i$ matrix. The (p, q)th block is obtained by replacing $\boldsymbol{\Omega}_{ipq}$ in $\mathbf{V}_{wipq}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(t_0)\}$ with

$$[E(D_{ip} D_{iq}) - H_{ip}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(t_0)\} H_{iq}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(t_0)\}].$$

Here η is an operator that maps functions in $C^1([0, \tau])$ to functionals from $\mathcal{D}$ to $\mathbb{R}^{d_w}$. Specifically, η minimizes

$$\sup_{\mathbf{w}_h \in \mathcal{D}} \|E([\widetilde{\mathbf{Q}}_{\mathbf{w}i}\{m_0, \mathbf{w}_h(\mathbf{T}_i)\} - \eta\{\mathbf{U}_i(\mathbf{T}_i)\}(\mathbf{w}_h)]^{\mathrm{T}} \Theta_i\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_i)\}$$

$$\times \boldsymbol{\Omega}_i^{-1}\Theta_i\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_i)\}[\widetilde{\mathbf{Q}}_{\mathbf{w}i}\{m_0, \mathbf{w}_h(\mathbf{T}_i)\} - \eta\{\mathbf{U}_i(\mathbf{T}_i)\}(\mathbf{w}_h)])\|_2$$

where

$$\begin{aligned}\widetilde{\mathbf{Q}}_{wi}\{m_0, \mathbf{w}_h(\mathbf{T}_i)\} &= [m_0'\{\mathbf{w}_0(T_{i1})^{\mathrm{T}}\mathbf{Z}_{i1}\}\mathbf{w}_h(T_{i1})^{\mathrm{T}}\mathbf{Z}_{i1}, \ldots, \\ &\quad m_0'\{\mathbf{w}_0(T_{iM_i})^{\mathrm{T}}\mathbf{Z}_{iM_i}\}\mathbf{w}_h(T_{iM_i})^{\mathrm{T}}\mathbf{Z}_{iM_i}]^{\mathrm{T}}\end{aligned}$$

and $\eta\{\mathbf{U}_i(\mathbf{T}_i)\}(\mathbf{w}_h) = [\eta\{\mathbf{w}(T_{ik})^{\mathrm{T}}\mathbf{Z}_{ik}\}(\mathbf{w}_h), k = 1, \ldots, M_i\}]^{\mathrm{T}}$ are M_i vectors. We can also write

$$\eta(\mathbf{w}_0(T_{ik})^{\mathrm{T}}\mathbf{Z}_{ik}) = E[\mathbf{Z}_{ik}m_0'\{\mathbf{w}_0(T_{ik})^{\mathrm{T}}\mathbf{Z}_{ik}\}|\mathbf{w}_0(T_{ik})^{\mathrm{T}}\mathbf{Z}_{ik}].$$

Further, we define $\widetilde{\mathbf{Q}}_{wi}\{\widehat{\boldsymbol{\lambda}}\{\widehat{\boldsymbol{\beta}}, \widehat{\mathbf{w}}(\widehat{\boldsymbol{\beta}})\}, \cdot\}$ is a $M_i \times d_w$ matrix, with row j as $\mathbf{B}_r'\{\widehat{\mathbf{w}}(\widehat{\boldsymbol{\beta}}, T_{ik})^{\mathrm{T}}\mathbf{Z}_{ik}\}\widehat{\boldsymbol{\lambda}}\{\widehat{\boldsymbol{\beta}}, \widehat{\mathbf{w}}(\widehat{\boldsymbol{\beta}})\}\mathbf{Z}_{ik}^{\mathrm{T}}$. In the estimation, we use the asymptotic form in Lemma 4 in the supplementary article in the place of $\partial\widehat{\boldsymbol{\lambda}}(\boldsymbol{\beta}_0, \mathbf{w})/\partial\mathbf{w}$ for computation.

Notation in Step 3. We define

$$\begin{aligned}&\widehat{\mathbf{S}}_{\beta ik}[\boldsymbol{\beta}, \widehat{\boldsymbol{\lambda}}\{\boldsymbol{\beta}, \widehat{\mathbf{w}}(\boldsymbol{\beta}, T_{ik})\}, \widehat{\mathbf{w}}(\boldsymbol{\beta})] \\ = &\left(\mathbf{Q}_{\beta ik} + \left[\frac{\partial\widehat{\boldsymbol{\lambda}}\{\boldsymbol{\beta}, \widehat{\mathbf{w}}(\boldsymbol{\beta})\}}{\partial\boldsymbol{\beta}^{\mathrm{T}}} + \frac{\partial\widehat{\boldsymbol{\lambda}}\{\boldsymbol{\beta}, \widehat{\mathbf{w}}(\boldsymbol{\beta})\}}{\widehat{\mathbf{w}}(\boldsymbol{\beta})}\frac{\partial\widehat{\mathbf{w}}(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}^{\mathrm{T}}}\right]^{\mathrm{T}}\mathbf{Q}_{\lambda ik}\{\widehat{\mathbf{w}}(\boldsymbol{\beta}, T_{ik})\}\right. \\ &\left. + \left\{\frac{\partial\widehat{\mathbf{w}}(\boldsymbol{\beta}, T_{ik})}{\partial\boldsymbol{\beta}^{\mathrm{T}}}\right\}^{\mathrm{T}}\mathbf{Q}_{wik}[\widehat{\boldsymbol{\lambda}}\{\boldsymbol{\beta}, \widehat{\mathbf{w}}(\boldsymbol{\beta})\}, \widehat{\mathbf{w}}(\boldsymbol{\beta}, T_{ik})]\right) \\ &\times\left(D_{ik} - H_{ik}[\boldsymbol{\beta}, \widehat{\boldsymbol{\lambda}}\{\boldsymbol{\beta}, \widehat{\mathbf{w}}(\boldsymbol{\beta})\}, \widehat{\mathbf{w}}(\boldsymbol{\beta}, T_{ik})]\right),\end{aligned}$$

and $\widehat{\mathbf{S}}_{\beta i}[\boldsymbol{\beta}, \widehat{\boldsymbol{\lambda}}\{\boldsymbol{\beta}, \widehat{\mathbf{w}}(\boldsymbol{\beta})\}, \widehat{\mathbf{w}}(\boldsymbol{\beta}, \mathbf{T}_i)] = (\widehat{\mathbf{S}}_{\beta ik}[\boldsymbol{\beta}, \widehat{\boldsymbol{\lambda}}\{\boldsymbol{\beta}, \widehat{\mathbf{w}}(\boldsymbol{\beta}, T_{ik})\}^{\mathrm{T}}, \widehat{\mathbf{w}}(\boldsymbol{\beta})], k = 1, \ldots, M_i)^{\mathrm{T}}$. Let $\widehat{\mathbf{A}}_{\beta i}[\boldsymbol{\beta}, \widehat{\boldsymbol{\lambda}}\{\boldsymbol{\beta}, \widehat{\mathbf{w}}(\boldsymbol{\beta}, \mathbf{T}_i)\}, \widehat{\mathbf{w}}(\boldsymbol{\beta})]$ be a $d_\beta \times d_\beta M_i$ matrix with the kth size $d_\beta \times d_\beta$ column block $\widehat{\mathbf{A}}_{\beta ik}[\boldsymbol{\beta}, \widehat{\boldsymbol{\lambda}}\{\boldsymbol{\beta}, \widehat{\mathbf{w}}(\boldsymbol{\beta})\}, \widehat{\mathbf{w}}(\boldsymbol{\beta}, T_{ik})]$ being

$$\begin{aligned}&\left(\mathbf{Q}_{\beta ik} + \left[\frac{\partial\widehat{\boldsymbol{\lambda}}\{\boldsymbol{\beta}, \widehat{\mathbf{w}}(\boldsymbol{\beta})\}}{\partial\boldsymbol{\beta}^{\mathrm{T}}} + \frac{\partial\widehat{\boldsymbol{\lambda}}\{\boldsymbol{\beta}, \widehat{\mathbf{w}}(\boldsymbol{\beta})\}}{\widehat{\mathbf{w}}(\boldsymbol{\beta})}\frac{\partial\widehat{\mathbf{w}}(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}^{\mathrm{T}}}\right]^{\mathrm{T}}\mathbf{Q}_{\lambda ik}\{\widehat{\mathbf{w}}(\boldsymbol{\beta}, T_{ik})\}\right. \\ &\left. + \left\{\frac{\partial\widehat{\mathbf{w}}(\boldsymbol{\beta}, T_{ik})}{\partial\boldsymbol{\beta}^{\mathrm{T}}}\right\}^{\mathrm{T}}\mathbf{Q}_{wik}[\widehat{\boldsymbol{\lambda}}\{\boldsymbol{\beta}, \widehat{\mathbf{w}}(\boldsymbol{\beta})\}, \widehat{\mathbf{w}}(\boldsymbol{\beta}, T_{ik})]\right)^{\otimes 2}\boldsymbol{\Theta}_i[\boldsymbol{\beta}, \widehat{\boldsymbol{\lambda}}\{\boldsymbol{\beta}, \widehat{\mathbf{w}}(\boldsymbol{\beta})\}, \\ &\widehat{\mathbf{w}}(\boldsymbol{\beta}, T_{ik})].\end{aligned}$$

Let $\widehat{\mathbf{V}}_{\beta i}[\boldsymbol{\beta}, \widehat{\boldsymbol{\lambda}}\{\boldsymbol{\beta}, \widehat{\mathbf{w}}(\boldsymbol{\beta})\}, \widehat{\mathbf{w}}(\boldsymbol{\beta}, \mathbf{T}_i)]^{-1}$ be a $d_\beta M_i \times d_\beta M_i$ matrix with the (p, q)th block $\widehat{\mathbf{V}}_{\beta ip}[\boldsymbol{\beta}, \widehat{\boldsymbol{\lambda}}\{\boldsymbol{\beta}, \widehat{\mathbf{w}}(\boldsymbol{\beta})\}, \widehat{\mathbf{w}}(\boldsymbol{\beta}, T_{ip})]$ being

$$\left(\mathbf{Q}_{\beta ip} + \left[\frac{\partial\widehat{\boldsymbol{\lambda}}\{\boldsymbol{\beta}, \widehat{\mathbf{w}}(\boldsymbol{\beta})\}}{\partial\boldsymbol{\beta}^{\mathrm{T}}} + \frac{\partial\widehat{\boldsymbol{\lambda}}\{\boldsymbol{\beta}, \widehat{\mathbf{w}}(\boldsymbol{\beta})\}}{\widehat{\mathbf{w}}(\boldsymbol{\beta})}\frac{\partial\widehat{\mathbf{w}}(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}^{\mathrm{T}}}\right]^{\mathrm{T}}\mathbf{Q}_{\lambda ip}\{\widehat{\mathbf{w}}(\boldsymbol{\beta}, T_{ip})\}\right.$$

$$+\left\{\frac{\partial\widehat{\mathbf{w}}(\boldsymbol{\beta},T_{ip})}{\partial\boldsymbol{\beta}^{\mathrm{T}}}\right\}^{\mathrm{T}}\mathbf{Q}_{wip}[\widehat{\boldsymbol{\lambda}}\{\boldsymbol{\beta},\widehat{\mathbf{w}}(\boldsymbol{\beta})\},\widehat{\mathbf{w}}(\boldsymbol{\beta},T_{ip})]\Bigg)\Bigg(\mathbf{Q}_{\beta iq}+\Bigg[\frac{\partial\widehat{\boldsymbol{\lambda}}\{\boldsymbol{\beta},\widehat{\mathbf{w}}(\boldsymbol{\beta})\}}{\partial\boldsymbol{\beta}^{\mathrm{T}}}$$
$$+\frac{\partial\widehat{\boldsymbol{\lambda}}\{\boldsymbol{\beta},\widehat{\mathbf{w}}(\boldsymbol{\beta})\}}{\widehat{\mathbf{w}}(\boldsymbol{\beta})}\frac{\partial\widehat{\mathbf{w}}(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}^{\mathrm{T}}}\Bigg]^{\mathrm{T}}\mathbf{Q}_{\lambda iq}\{\widehat{\mathbf{w}}(\boldsymbol{\beta},T_{iq})\}+\left\{\frac{\partial\widehat{\mathbf{w}}(\boldsymbol{\beta},T_{iq})}{\partial\boldsymbol{\beta}^{\mathrm{T}}}\right\}^{\mathrm{T}}$$
$$\times\mathbf{Q}_{wiq}[\widehat{\boldsymbol{\lambda}}\{\boldsymbol{\beta},\widehat{\mathbf{w}}(\boldsymbol{\beta})\},\widehat{\mathbf{w}}(\boldsymbol{\beta},T_{iq})]\Bigg)^{\mathrm{T}}\boldsymbol{\Omega}_{ipq}.$$

Additionally, let $\delta_u\in C^q([0,1])$ and we define $\boldsymbol{\delta}\{\mathbf{w}(T_{ik})^{\mathrm{T}}\mathbf{Z}_{ik}\}=[\delta_u\{\mathbf{w}(T_{ik})^{\mathrm{T}}\mathbf{Z}_{ik}\},\ u=1,\dots,d_\beta]\in\mathrm{R}^{d_\beta}$ which minimizes

$$\mathbf{1}_{d_\beta}^{\mathrm{T}}E([\widetilde{\mathbf{Q}}_{\beta i}-\boldsymbol{\delta}\{\mathbf{U}_i(\mathbf{T}_i)\}]^{\mathrm{T}}\Theta_i\{\boldsymbol{\beta}_0,m_0,\mathbf{w}_0(\mathbf{T}_i)\}\boldsymbol{\Omega}_i^{-1}\Theta_i\{\boldsymbol{\beta}_0,m_0,\mathbf{w}_0(\mathbf{T}_i)\}$$
$$\times[\widetilde{\mathbf{Q}}_{\beta i}-\boldsymbol{\delta}\{\mathbf{U}_i(\mathbf{T}_i)\}])\mathbf{1}_{d_\beta},$$

where $\widetilde{\mathbf{Q}}_{\beta i}=(\mathbf{X}_{i1},\dots,\mathbf{X}_{iM_i})^{\mathrm{T}}$ is a $M_i\times d_\beta$ matrix, and $\boldsymbol{\delta}\{\mathbf{U}_i(\mathbf{T}_i)\}=[\boldsymbol{\delta}\{\mathbf{w}(T_{ik})^{\mathrm{T}}\mathbf{Z}_{ik}\},k=1,\dots,M_i\}]^{\mathrm{T}}$ is a $M_i\times d_\beta$ matrix. We can also write $\boldsymbol{\delta}\{\mathbf{w}_0(T_{ik})^{\mathrm{T}}\mathbf{Z}_{ik}\}$ as $E\{\mathbf{X}|\mathbf{w}_0(T_{ik})^{\mathrm{T}}\mathbf{Z}_{ik}\}$. Further, we define

$$\begin{aligned}\mathbf{B}(t_0)\ =\ & E(\mathbf{A}_{wi}\{\boldsymbol{\beta}_0,m_0,\mathbf{w}_0(t_0)\}\mathbf{V}_{wi}\{\boldsymbol{\beta}_0,m_0,\mathbf{w}_0(t_0)\}^{-1}[\mathbf{Q}_{wi}\{m_0,\mathbf{w}_0(t_0)\}\\ & -\eta\{\mathbf{U}_i(t_0)\}]\boldsymbol{\Theta}_i^*\{\boldsymbol{\beta}_0,m_0,\mathbf{w}_0(t_0)\}\mathbf{Q}_{wi}^*\{m_0,\mathbf{w}_0(t_0)\}),\end{aligned}$$

where $\boldsymbol{\Theta}_i^*\{\boldsymbol{\beta}_0,m_0,\mathbf{w}(t_0)\}$ is a $d_wM_i\times d_wM_i$ diagonal matrix with the kth diagonal block being a $d_w\times d_w$ diagonal with the element $\Theta_{ik}\{\boldsymbol{\beta}_0,m_0,\mathbf{w}(t_0)\}$. And $\mathbf{Q}_{wi}\{m_0,\mathbf{w}(t_0)\}$ is a $d_wM_i\times d_wM_i$ diagonal matrix with the kth diagonal block being $\mathrm{diag}[\mathbf{Z}_{ik}m_0'\{\mathbf{w}(t_0)^{\mathrm{T}}\mathbf{Z}_{ik}\}]$. Moreover $\mathbf{Q}_{wi}^*\{m_0,\mathbf{w}(\mathbf{t}_0)\}$ is a $d_wM_i\times d_w$ matrix with the kth row block being a $d_w\times d_w$ matrix with d_w replications of $\mathbf{Z}_{ik}^{\mathrm{T}}m_0'\{\mathbf{w}(t_0)^{\mathrm{T}}\mathbf{Z}_{ik}\}$. And $\eta\{\mathbf{U}_i(t_0)\}=[\eta\{\mathbf{w}(t_0)\mathbf{Z}_{i1}\}^{\mathrm{T}},\dots,\eta\{\mathbf{w}(t_0)\mathbf{Z}_{iM_i}\}^{\mathrm{T}}]^{\mathrm{T}}$. Also Let $\mathbf{B}(\mathbf{T}_i)$ be the $d_wM_i\times d_wM_i$ block diagonal matrix with the kth block as $\mathbf{B}(T_{ik})$ and $\mathbf{f}_{\mathbf{T}}(\mathbf{T}_i)$ be the $d_wM_i\times d_wM_i$ block diagonal matrix with the kth block as $f_T(T_{ik})$.

Let $\gamma_u\in C^q([0,1])$ and we define $\boldsymbol{\gamma}\{\mathbf{w}(T_{ik})^{\mathrm{T}}\mathbf{Z}_{ik}\}=[\gamma_u\{\mathbf{w}(T_{ik})^{\mathrm{T}}\mathbf{Z}_{ik}\},u=1,\dots,d_\beta]\in\mathrm{R}^{d_\beta}$, which minimize

$$\mathbf{1}_\beta^{\mathrm{T}}E\Bigg[\Bigg\{\widetilde{\mathbf{Q}}_{\mathbf{w}i}\Bigg(m_0,\mathbf{B}(\mathbf{T}_i)^{-1}E\Bigg[\mathbf{A}_{wj}\{\boldsymbol{\beta}_0,m_0,\mathbf{w}_0(\mathbf{T}_i)\}\mathbf{V}_{wj}\{\boldsymbol{\beta}_0,m_0,$$
$$\mathbf{w}_0(\mathbf{T}_i)\}^{-1}\frac{\partial\mathbf{S}_{wj}\{\boldsymbol{\beta}_0,m_0,\mathbf{w}_0(\mathbf{T}_i)\}}{\partial\boldsymbol{\beta}^{\mathrm{T}}}\Bigg)\Bigg|\mathbf{O}_i\Bigg]\Bigg)-\boldsymbol{\gamma}\{\mathbf{U}_i(\mathbf{T}_i)\}\Bigg\}^{\mathrm{T}}\Theta_i\{\boldsymbol{\beta}_0,m_0,$$
$$\mathbf{w}_0(\mathbf{T}_i)\}\boldsymbol{\Omega}_i^{-1}\Theta_i\{\boldsymbol{\beta}_0,m_0,\mathbf{w}_0(\mathbf{T}_i)\}\Bigg\{\widetilde{\mathbf{Q}}_{\mathbf{w}i}\Bigg(m_0,\mathbf{B}(\mathbf{T}_i)^{-1}E\Bigg[\mathbf{A}_{wj}\{\boldsymbol{\beta}_0,m_0,$$
$$\mathbf{w}_0(\mathbf{T}_i)\}\mathbf{V}_{wj}\{\boldsymbol{\beta}_0,m_0,\mathbf{w}_0(\mathbf{T}_i)\}^{-1}\frac{\partial\mathbf{S}_{wj}\{\boldsymbol{\beta}_0,m_0,\mathbf{w}_0(\mathbf{T}_i)\}}{\partial\boldsymbol{\beta}^{\mathrm{T}}}\Bigg)\Bigg|\mathbf{O}_i\Bigg]\Bigg)$$

$$-\boldsymbol{\gamma}\{\mathbf{U}_i(\mathbf{T}_i)\}\Bigg\}\Bigg]\mathbf{1}_\beta.$$

where $\boldsymbol{\gamma}\{\mathbf{U}_i(\mathbf{T}_i)\} = [\boldsymbol{\gamma}\{\mathbf{w}(T_{ik})^{\mathrm{T}}\mathbf{Z}_{ik}\}, k = 1, \ldots, M_i\}]^{\mathrm{T}}$ is a $M_i \times d_\beta$, and

$$\widetilde{\mathbf{Q}}_{wi}\Bigg(m_0, \mathbf{B}(\mathbf{T}_i)^{-1}E\Bigg[\mathbf{A}_{wj}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_i)\}\mathbf{V}_{wj}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_i)\}^{-1} \times \frac{\partial \mathbf{S}_{wj}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_i)\}}{\partial \boldsymbol{\beta}^{\mathrm{T}}}\Bigg)\Bigg|\mathbf{O}_i\Bigg]\Bigg),$$

is a $M_i \times \boldsymbol{\beta}$ matrix with kth row as

$$\Bigg(\mathbf{B}(T_{ik})^{-1}E\Bigg[\mathbf{A}_{wj}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(T_{ik})\}\mathbf{V}_{wj}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(T_{ik})\}^{-1} \times \frac{\partial \mathbf{S}_{wj}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(T_{ik})\}}{\partial \boldsymbol{\beta}^{\mathrm{T}}} \mid \mathbf{O}_i\Bigg]\Bigg)^{\mathrm{T}}\mathbf{Z}_{ik} m_0'\{\mathbf{w}_0(T_{ik})^{\mathrm{T}}\mathbf{Z}_{ik}\}.$$

We can also write

$$\begin{aligned}\boldsymbol{\gamma}(\mathbf{w}_0(T_{ik})^{\mathrm{T}}\mathbf{Z}_{ik}) &= E\Bigg\{\Bigg(\mathbf{B}(T_{ik})^{-1}E\Bigg[\mathbf{A}_{wj}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(T_{ik})\}\mathbf{V}_{wj}\{\boldsymbol{\beta}_0, m_0, \\ &\quad \mathbf{w}_0(T_{ik})\}^{-1}\frac{\partial \mathbf{S}_{wj}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(T_{ik})\}}{\partial \boldsymbol{\beta}^{\mathrm{T}}} \mid \mathbf{O}_i\Bigg]\Bigg)^{\mathrm{T}}\mathbf{Z}_{ik} \\ &\quad \times m_0'\{\mathbf{w}_0(T_{ik})^{\mathrm{T}}\mathbf{Z}_{ik}\}\Bigg|\mathbf{w}_0(T_{ik})^{\mathrm{T}}\mathbf{Z}_{ik}\Bigg\}\end{aligned}$$

We also define the population forms $\mathbf{S}_{\beta ik}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(T_{ik})\}$ as

$$\Bigg\{\mathbf{Q}_{\beta ik} - \boldsymbol{\delta}\{\mathbf{w}_0(T_{ik})^{\mathrm{T}}\mathbf{Z}_{ik}\} - \Bigg(\mathbf{B}(T_{ik})^{-1}E\Bigg[\mathbf{A}_{wj}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(T_{ik})\} \mathbf{V}_{wj}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(T_{ik})\}^{-1}\frac{\partial \mathbf{S}_{wj}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(T_{ik})\}}{\partial \boldsymbol{\beta}^{\mathrm{T}}} \mid \mathbf{O}_i\Bigg]\Bigg)^{\mathrm{T}}\mathbf{Z}_{ik} m_0'\{\mathbf{w}_0(T_{ik})^{\mathrm{T}}\mathbf{Z}_{ik}\} + \boldsymbol{\gamma}\{\mathbf{w}_0(T_{ik})^{\mathrm{T}}\mathbf{Z}_{ik}\}\Bigg\}[D_{ik} - H_{ik}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(T_{ik})\}]$$

and $\mathbf{S}_{\beta i}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_i)\} = [\mathbf{S}_{\beta ik}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(T_{ik})\}^{\mathrm{T}}, k = 1, \ldots, M_i]^{\mathrm{T}}$. Let $\mathbf{A}_{\beta i}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_i)\}$ be a $d_\beta \times d_\beta M_i$ be the matrix with the kth block $\mathbf{A}_{\beta ik}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(T_{ik})\}$ being a $d_\beta \times d_\beta$ matrix

$$\Bigg\{\mathbf{Q}_{\beta ik} - \boldsymbol{\delta}\{\mathbf{w}_0(T_{ik})^{\mathrm{T}}\mathbf{Z}_{ik}\} - \Bigg(\mathbf{B}(T_{ik})^{-1}E\Bigg[\mathbf{A}_{wj}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(T_{ik})\} \mathbf{V}_{wj}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(T_{ik})\}^{-1}\frac{\partial \mathbf{S}_{wj}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(T_{ik})\}}{\partial \boldsymbol{\beta}^{\mathrm{T}}} \mid \mathbf{O}_i\Bigg]\Bigg)^{\mathrm{T}}\mathbf{Z}_{ik}$$

$$\times m_0'\{\mathbf{w}_0(T_{ik})^{\mathrm{T}}\mathbf{Z}_{ik}\} + \boldsymbol{\gamma}\{\mathbf{w}_0(T_{ik})^{\mathrm{T}}\mathbf{Z}_{ik}\}\Bigg\}^{\otimes 2}\Theta_{ik}[\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(T_{ik})].$$

Let $\mathbf{V}_{\beta i}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_i)\}$ be a $d_\beta M_i \times d_\beta M_i$ with the (p,q)th block $\mathbf{V}_{\beta ipq}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(T_{ip})\}$ being

$$\Bigg\{\mathbf{Q}_{\beta ip} - \boldsymbol{\delta}\{\mathbf{w}_0(T_{ip})^{\mathrm{T}}\mathbf{Z}_{ip}\} - \Bigg(\mathbf{B}(T_{ip})^{-1}E\Bigg[\mathbf{A}_{wj}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(T_{ip})\}$$
$$\mathbf{V}_{wj}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(T_{ip})\}^{-1}\frac{\partial \mathbf{S}_{wj}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(T_{ip})\}}{\partial \boldsymbol{\beta}^{\mathrm{T}}} \mid \mathbf{O}_i\Bigg]\Bigg)^{\mathrm{T}}\mathbf{Z}_{ip}$$
$$\times m_0'\{\mathbf{w}_0(T_{ip})^{\mathrm{T}}\mathbf{Z}_{ip}\}\} + \boldsymbol{\gamma}\{\mathbf{w}_0(T_{ip})^{\mathrm{T}}\mathbf{Z}_{ip}\}\Bigg\}\Bigg\{\mathbf{Q}_{\beta iq} - \boldsymbol{\delta}\{\mathbf{w}_0(T_{iq})^{\mathrm{T}}\mathbf{Z}_{iq}\}$$
$$-\Bigg(\mathbf{B}(T_{iq})^{-1}E\Bigg[\mathbf{A}_{wj}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(T_{iq})\}\mathbf{V}_{wj}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(T_{iq})\}^{-1}$$
$$\times\frac{\partial \mathbf{S}_{wj}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(T_{iq})\}}{\partial \boldsymbol{\beta}^{\mathrm{T}}} \mid \mathbf{O}_i\Bigg]\Bigg)^{\mathrm{T}}\mathbf{Z}_{iq}m_0'\{\mathbf{w}_0(T_{iq})^{\mathrm{T}}\mathbf{Z}_{iq}\}$$
$$+\boldsymbol{\gamma}\{\mathbf{w}_0(T_{iq})^{\mathrm{T}}\mathbf{Z}_{iq}\}\Bigg\}^{\mathrm{T}}\boldsymbol{\Omega}_{ipq}.$$

Let $\mathbf{V}^*_{\beta i}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_i)\}$ be a $d_\beta M_i \times d_\beta M_i$ matrix. The (p,q)th block is obtained by replacing $\boldsymbol{\Omega}_{ipq}$ in $\mathbf{V}_{\beta i}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_i)\}$ with

$$[E(D_{ip}D_{iq}) - H_{ip}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}(T_{ip})\}H_{iq}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}(T_{iq})\}].$$

SUPPLEMENTARY MATERIAL

Supplement: Supplement to "Fused Kernel-Spline Smoothing for Repeatedly Measured Outcomes in a Generalized Partially Linear Model with Functional Single Index"
(http://www.e-publications.org/ims/support/dowload). We provide the comprehensive proofs of Theorem 1, 2, 3 and additional Lemmas which support the results.

REFERENCES

Bishop, Y. M., Fienberg, S. E. and Holland, P. W. (2007). *Discrete Multivariate Analysis: Theory and Practice*. New York : Springer, c2007.

Bosq, D. (1998). *Nonparametric statistics for stochastic processes : estimation and prediction D. Bosq. Lecture notes in statistics: 110*. New York : Springer, c1998.

Carroll, R. J., Fan, J., Gijbels, I. and Wand, M. P. (1997). Generalized Partially Linear Single-Index Models. *Journal of the American Statistical Association* **92** pp. 477-489.

Cui, X., Härdle, W. K. and Zhu, L. (2011). The EFM approach for single-index models. *The Annals of Statistics* **39** 1658–1688.

de Boor, C. (2001). *A Practical Guide to Splines. Applied Mathematical Sciences* **v. 27**. Springer.

DeVore, R. A. and Lorentz, G. G. (1993). *Constructive approximation. Grundlehren der mathematischen Wissenschaften: 303.* Berlin ; New York : Springer-Verlag, c1993.

Jiang, F., Ma, Y. and Wang, Y. (2015). Supplement to "Fused Kernel-Spline Smoothing for Repeatedly Measured Outcomes in a Generalized Partially Linear Model with Functional Single Index".

Jiang, C.-R. and Wang, J.-L. (2011). Functional single index models for longitudinal data. *The Annals of Statistics* **39** 362–388.

Lu, M. and Loomis, D. (2013). Spline-based semiparametric estimation of partially linear Poisson regression with single-index models. *Journal of Nonparametric Statistics.*

Ma, S. and Song, P. X.-K. (2014). Varying Index Coefficient Models. *Journal of the American Statistical Association* 00–00.

Ma, Y. and Zhu, L. (2013). Doubly robust and efficient estimators for heteroscedastic partially linear single-index models allowing high dimensional covariates. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **75** 305–322.

Paulsen, J. S., Langbehn, D. R., Stout, J. C., Aylward, E., Ross, C. A., Nance, M., Guttman, M., Johnson, S., MacDonald, M., Beglinger, L. J., Duff, K., Kayson, E., Biglan, K., Shoulson, I., Oakes, D. and Hayden, M. (2008). Detection of Huntingtons disease decades before diagnosis: the Predict-HD study. *Journal of Neurology, Neurosurgery & Psychiatry* **79** 874-880.

Peng, H. and Huang, T. (2011). Penalized least squares for single index models. *Journal of Statistical Planning and Inference* **141** 1362 - 1379.

Silverman, B. W. (1986). *Density estimation for statistics and data analysis B.W. Silverman. Monographs on statistics and applied probability: 26.* London; New York: Chapman and Hall, 1986.

Smith, A. (1982). Symbol digits modalities test: manual. Los Angeles: Western Psychological Services.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology* **18** 643–662.

Wang, L. and Yang, L. (2009). Spline estimation of single-index models. *Statistica Sinica* **19** 765.

Wang, J.-L., Xue, L., Zhu, L. and Chong, Y. S. (2010). Estimation for a partial-linear single-index model. *The Annals of statistics* **38** 246–274.

Xia, Y. and Härdle, W. (2006). Semi-parametric estimation of partially linear single-index models. *Journal of Multivariate Analysis* **97** 1162–1184.

Xia, Y., Tong, H., Li, W. and Zhu, L.-X. (2002). An adaptive estimation of dimension reduction space. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **64** 363–410.

Xu, P. and Zhu, L. (2012). Estimation for a marginal generalized single-index longitudinal model. *Journal of Multivariate Analysis* **105** 285 - 299.

Zhang, Y., Long, J. D., Mills, J. A., Warner, J. H., Lu, W. and Paulsen, J. S. (2011). AIndexing disease progression at study entry with individuals at risk for Huntington disease. *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics* **156** 751.

655 Huntington Ave., Boston, MA, 02115. *722 West 168th S.T., Rm 205 New York, NY, 10032.*
jiang@hsph.harvard.edu *yuanjiaw@gmail.com*
1523 Greene Street Columbia, SC 29208.
yanyuanma@stat.sc.edu

Supplement to "Fused Kernel-Spline Smoothing for Repeatedly Measured Outcomes in a Generalized Partially Linear Model with Functional Single Index"

APPENDIX S.1: NOTATIONS IN THE PROOFS

We let $a \asymp b$ denote $a = O(b)$. For a vector $\boldsymbol{\xi} = \{\xi_1, \ldots, \xi_s\} \in \mathbf{R}^s$, we define $\|\boldsymbol{\xi}\|_\infty = \max_{1\le j\le s} |\xi_j|$ and $\|\boldsymbol{\xi}\|_q = (\sum_{j=1}^q |\xi_j|^q)^{1/q}$. For a $m \times s$ matrix $\mathbf{P} = (P_{ij})$, we define the norms $\|\mathbf{P}\|_\infty = \sup_{1\le i\le m} \sum_{j=1}^s |P_{ij}|$. $\|\mathbf{P}\|_q = \sup \|\mathbf{P}\boldsymbol{\xi}\|_q \|\boldsymbol{\xi}\|_q^{-1}$.

APPENDIX S.2: PROOF OF THEOREM 1

Fact 1 (de Boor, 2001): Assume $m \in C^q([0,1])$. There exists a $\boldsymbol{\lambda}_0 \in \mathrm{R}^{d_\lambda}$, such that

$$\sup_{u\in[0,1]} |m(u) - \widetilde{m}(u, \boldsymbol{\lambda}_0)| = o(h_b^q).$$

$\boldsymbol{\lambda}_0$ is the value of $\boldsymbol{\lambda}$ such that

$$E\Bigg(\widetilde{\mathbf{Q}}_{\lambda i}\{\mathbf{w}_0(\mathbf{T}_i)\}^{\mathrm{T}} \boldsymbol{\Theta}_i\{\boldsymbol{\beta}_0, \boldsymbol{\lambda}, \mathbf{w}_0(\mathbf{T}_i)\}\boldsymbol{\Omega}_i^{-1} \times [\mathbf{D}_i - \mathbf{H}_i\{\boldsymbol{\beta}_0, \boldsymbol{\lambda}, \mathbf{w}_0(\mathbf{T}_i)\}] \mid \mathbf{R}_i\Bigg) = \mathbf{0}, a.s..$$

<u>**Lemma**</u> 1. *For any $a = \{a_p, 1 \le p \le d_\lambda\}$, there exists constant $0 \le c_B \le C_B \le \infty$ such that*

(S.1)
$$c_B a^{\mathrm{T}} a h_b \le a^{\mathrm{T}} E[\mathbf{B}_r\{\mathbf{w}(T_{ik})^{\mathrm{T}}\mathbf{Z}_{ik}\}\mathbf{B}_r\{\mathbf{w}(T_{ik})^{\mathrm{T}}\mathbf{Z}_{ik}\}^{\mathrm{T}}]a \le C_B a^{\mathrm{T}} a h_b,$$

(S.2)
$$\max_{1\le p,p'\le d_\lambda} \Bigg| n^{-1}\sum_{i=1}^n\sum_{k=1}^{M_i} B_{rp}\{\mathbf{w}_0(T_{ik})^{\mathrm{T}}\mathbf{Z}_{ik}\}B_{rp'}\{\mathbf{w}_0(T_{ik})^{\mathrm{T}}\mathbf{Z}_{ik}\} - E\Bigg[\sum_{k=1}^{M_i} B_{rp}\{\mathbf{w}_0(T_{ik})^{\mathrm{T}}\mathbf{Z}_{ik}\}B_{rp'}\{\mathbf{w}_0(T_{ik})^{\mathrm{T}}\mathbf{Z}_{ik}\}\Bigg]\Bigg| = O_{a.s.}\{\sqrt{h_b n^{-1} log(n)}\},$$

and

(S.3)
$$\Bigg| n^{-1}\sum_{i=1}^n\sum_{k=1}^{M_i}\sum_{l=1}^{M_i} B_{rp}\{\mathbf{w}_0(T_{ik})^{\mathrm{T}}\mathbf{Z}_{ik}\}B_{rp'}\{\mathbf{w}_0(T_{il})^{\mathrm{T}}\mathbf{Z}_{il}\}\Bigg| = O_{a.s.}(h_b).$$

Proof: The first inequality is the direct result from Theorem 5.4.2 of the DeVore and Lorentz (1993). To prove the second result, note that

$$\Big| n^{-1}\sum_{i=1}^{n}\sum_{k=1}^{M_i} B_{rp}\{\mathbf{w}_0(T_{ik})^{\mathrm{T}}\mathbf{Z}_{ik}\}B_{rp'}\{\mathbf{w}_0(T_{ik})^{\mathrm{T}}\mathbf{Z}_{ik}\} \\ -E\Big[\sum_{k=1}^{M_i} B_{rp}\{\mathbf{w}_0(T_{ik}\}^{\mathrm{T}}\mathbf{Z}_{ik})B_{rp'}\{\mathbf{w}_0(T_{ik})^{\mathrm{T}}\mathbf{Z}_{ik}\}\Big]\Big|$$

is equal to

$$\begin{aligned} & \Big| n^{-1}\sum_{i=1}^{n}\sum_{k=1}^{M_i}\Big(B_{rp}\{\mathbf{w}_0(T_{ik})^{\mathrm{T}}\mathbf{Z}_{ik}\}B_{rp'}\{\mathbf{w}_0(T_{ik})^{\mathrm{T}}\mathbf{Z}_{ik}\} \\ & -E\Big[B_{rp}\{\mathbf{w}_0(T_{ik}\}^{\mathrm{T}}\mathbf{Z}_{ik})B_{rp'}\{\mathbf{w}_0(T_{ik})^{\mathrm{T}}\mathbf{Z}_{ik}\} \mid M_i\Big]\Big)\Big| \\ = & \Big| \frac{\sum_{i=1}^{n} M_i}{n}\frac{1}{\sum_{i=1}^{n} M_i}\sum_{i=1}^{n}\sum_{k=1}^{M_i}\Big(B_{rp}\{\mathbf{w}_0(T_{ik})^{\mathrm{T}}\mathbf{Z}_{ik}\}B_{rp'}\{\mathbf{w}_0(T_{ik})^{\mathrm{T}}\mathbf{Z}_{ik}\} \\ & -E\Big[B_{rp}\{\mathbf{w}_0(T_{ik}\}^{\mathrm{T}}\mathbf{Z}_{ik})B_{rp'}\{\mathbf{w}_0(T_{ik})^{\mathrm{T}}\mathbf{Z}_{ik}\} \mid M_i\Big]\Big)\Big| \\ = & O(1)O_{a.s.}\{\sqrt{h_b n^{-1}\log(n)}\}. \end{aligned}$$

The last equation follows directly from Bernsteins inequality (Bosq, 1998). Therefore, combining the above two results, we have

$$\Big| n^{-1}\sum_{i=1}^{n}\sum_{k=1}^{M_i} B_{rp}\{\mathbf{w}_0(T_{ik})^{\mathrm{T}}\mathbf{Z}_{ik}\}B_{rp'}\{\mathbf{w}_0(T_{ik})^{\mathrm{T}}\mathbf{Z}_{ik}\}\Big| = O_p(h_b)$$

Further, we have

$$\begin{aligned} & \Big| n^{-1}\sum_{i=1}^{n}\sum_{k=1}^{M_i}\sum_{l=1}^{M_i} B_{rp}\{\mathbf{w}_0(T_{ik})^{\mathrm{T}}\mathbf{Z}_{ik}\}B_{rp'}\{\mathbf{w}_0(T_{il})^{\mathrm{T}}\mathbf{Z}_{il}\}\Big| \\ \leq & \Big| n^{-1}\sum_{i=1}^{n}\sup_{1\leq l\leq M_i}|B_{rp'}\{\mathbf{w}_0(T_{il})^{\mathrm{T}}\mathbf{Z}_{il}\}|M_i|\sum_{k=1}^{M_i} B_{rp}\{\mathbf{w}_0(T_{ik})^{\mathrm{T}}\mathbf{Z}_{ik}\}|\Big| \\ \leq & \sup_{1\leq i\leq n} M_i\Big| n^{-1}\sum_{i=1}^{n}\sup_{1\leq l\leq M_i}|B_{rp'}\{\mathbf{w}_0(T_{il})^{\mathrm{T}}\mathbf{Z}_{il}\}||\sum_{k=1}^{M_i} B_{rp}\{\mathbf{w}_0(T_{ik})^{\mathrm{T}}\mathbf{Z}_{ik}\}|\Big| \\ = & O_{a.s.}(h_b). \end{aligned}$$

This proves the last result. □

Corollary 1. *Let C_{ik} be a random variable, if $E(C_{ik}|\mathbf{R}_i) \neq 0$, then*

$$\left\| n^{-1} \sum_{i=1}^{n} \sum_{k=1}^{M_i} C_{ik} \mathbf{B}_r \{\mathbf{w}_0(T_{ik})^{\mathrm{T}} \mathbf{Z}_{ik}\} \right\|_{\infty} = O(h_b), a.s. \tag{S.4}$$

If the $E(C_{ik}|\mathbf{R}_i) = 0$, then

$$\left\| n^{-1} \sum_{i=1}^{n} \sum_{k=1}^{M_i} C_{ik} \mathbf{B}_r \{\mathbf{w}_0(T_{ik})^{\mathrm{T}} \mathbf{Z}_{ik}\} \right\|_{\infty} = O\{\sqrt{h_b n^{-1} log(n)}\}, a.s. \tag{S.5}$$

Additionally, for a $M_i \times M_i$ bounded random matrix $\mathbf{C}_i$, and bounded positive real numbers c_a, C_a and positive random numbers c_b, c_d, C_a, C_b, C_d we have

$$c_a h_b \leq \left\| E\left[\widetilde{\mathbf{Q}}_{\lambda i}\{\mathbf{w}_0(\mathbf{T}_i)\}^{\mathrm{T}} \mathbf{C}_i \widetilde{\mathbf{Q}}_{\lambda i}\{\mathbf{w}_0(\mathbf{T}_i)\} \right] \right\|_2 \leq C_a h_b, \tag{S.6}$$

$$c_b h_b \leq \left\| n^{-1} \sum_{i=1}^{n} \widetilde{\mathbf{Q}}_{\lambda i}\{\mathbf{w}_0(\mathbf{T}_i)\}^{\mathrm{T}} \mathbf{C}_i \widetilde{\mathbf{Q}}_{\lambda i}\{\mathbf{w}_0(\mathbf{T}_i)\} \right\|_2 \leq C_b h_b, \tag{S.7}$$

and

$$c_d h_b \leq \left\| \widetilde{\mathbf{Q}}_{\lambda i}\{\mathbf{w}_0(\mathbf{T}_i)\}^{\mathrm{T}} \mathbf{C}_i \widetilde{\mathbf{Q}}_{\lambda i}\{\mathbf{w}_0(\mathbf{T}_i)\} \right\|_2 \leq C_d h_b. \tag{S.8}$$

Further

$$\left\| E\left[\widetilde{\mathbf{Q}}_{\lambda i}\{\mathbf{w}_0(\mathbf{T}_i)\}^{\mathrm{T}} \mathbf{C}_i \widetilde{\mathbf{Q}}_{\lambda i}\{\mathbf{w}_0(\mathbf{T}_i)\} \right]^{-1} \right\|_{\infty} = O(h_b^{-1}) \tag{S.9}$$

and

$$\left\| \left[n^{-1} \sum_{i=1}^{n} \widetilde{\mathbf{Q}}_{\lambda i}\{\mathbf{w}_0(\mathbf{T}_i)\}^{\mathrm{T}} \mathbf{C}_i \widetilde{\mathbf{Q}}_{\lambda i}\{\mathbf{w}_0(\mathbf{T}_i)\} \right]^{-1} \right\|_{\infty} = O_p(h_b^{-1}) \tag{S.10}$$

Moreover, for a matrix $\mathbf{C}_i$ with M_i columns, we have

$$\|\mathbf{C}_i \widetilde{\mathbf{Q}}_{\lambda i}\{\mathbf{w}_0(\mathbf{T}_i)\}\|_2 = O_p(h_b^{1/2}), \tag{S.11}$$

and

$$\begin{aligned} &\left\| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \widetilde{\mathbf{Q}}_{\lambda i}\{\mathbf{w}_0(\mathbf{T}_i)\}^{\mathrm{T}} \boldsymbol{\Theta}_i\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_i)\} \boldsymbol{\Omega}_i^{-1} \right. \\ &\left. \times [\mathbf{D}_i - \mathbf{H}_i\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_i)\}] \right\|_2 = O_p(h_b^{1/2}). \end{aligned} \tag{S.12}$$

Proof: We now show the equation (S.4). Because $B_{rp}\{\mathbf{w}_0(T_{ik})^{\mathrm{T}}\mathbf{Z}_{ik}\}$ has support of length $O(h_b)$ uniformly for all r, and $B_{rp}\{\mathbf{w}_0(T_{ik})^{\mathrm{T}}\mathbf{Z}_{ik}\}$ are bounded between 0 and 1, a direct calculation of the expectation then yields

$$E[B_{rp}\{\mathbf{w}_0(T_{ik})^{\mathrm{T}}\mathbf{Z}_{ik}\}] = O(h_b).$$

Further by the Bernstein's inequality in Bosq (1998) we have

$$\begin{aligned} & \left|n^{-1}\sum_{i=1}^{n}\sum_{k=1}^{M_i} C_{ik}B_{rp}\{\mathbf{w}_0(T_{ik})^{\mathrm{T}}\mathbf{Z}_{ik}\} - E\left[\sum_{k=1}^{M_i} C_{ik}B_{rp}\{\mathbf{w}_0(T_{ik})^{\mathrm{T}}\mathbf{Z}_{ik}\}\right]\right| \\ \asymp & \left(\sum_{i=1}^{n} E\left[n^{-1}\sum_{k=1}^{M_i} C_{ik}B_{rp}\{\mathbf{w}_0(T_{ik})^{\mathrm{T}}\mathbf{Z}_{ik}\}\right]^2 \log(n)\right)^{1/2} \\ = & O_p\{\sqrt{h_b n^{-1}\log(n)}\}. \end{aligned}$$

Here, the last equality is derived following the same line as the proof of the last result in Lemma 1. Now (S.4) follows because

$$\begin{aligned} & \left\|n^{-1}\sum_{i=1}^{n}\sum_{k=1}^{M_i} C_{ik}\mathbf{B}_r\{\mathbf{w}_0(T_{ik})^{\mathrm{T}}\mathbf{Z}_{ik}\}\right\|_\infty \\ = & \sup_{1\le p\le d_\lambda}\left|n^{-1}\sum_{i=1}^{n}\sum_{k=1}^{M_i} C_{ik}B_{rp}\{\mathbf{w}_0(T_{ik})^{\mathrm{T}}\mathbf{Z}_{ik}\}\right|, \end{aligned}$$

which is asymptotic to

$$E\left[\sum_{k=1}^{M_i} B_{rp}\{\mathbf{w}_0(T_{ik})^{\mathrm{T}}\mathbf{Z}_{ik}\}\right] = O_p(h_b).$$

To show (S.5), we simply note that when $E(C_{ik}|\mathbf{R}_{ik}) = 0$,

$$\begin{aligned} & \left|n^{-1}\sum_{i=1}^{n}\sum_{k=1}^{M_i} C_{ik}B_{rp}\{\mathbf{w}_0(T_{ik})^{\mathrm{T}}\mathbf{Z}_{ik}\}\right| \\ \asymp & \left(\sum_{i=1}^{n} E\left[n^{-1}\sum_{k=1}^{M_i} C_{ik}B_{rp}\{\mathbf{w}_0(T_{ik})^{\mathrm{T}}\mathbf{Z}_{ik}\}\right]^2 \log(n)\right)^{1/2} \\ = & O_p\{\sqrt{h_b n^{-1}\log(n)}\}. \end{aligned}$$

To show (S.6), first notes that, because of Lemma 1, we have

$$\left\|E\left[\mathbf{B}_r\{\mathbf{w}_0(T_{ik})^{\mathrm{T}}\mathbf{Z}_{ik}\}\mathbf{B}_r^{\mathrm{T}}\{\mathbf{w}_0(T_{ik})^{\mathrm{T}}\mathbf{Z}_{ik}\}\right]\right\|_2$$

$$\begin{aligned}
&\leq \quad E\left[\left\|\mathbf{B}_r\{\mathbf{w}_0(T_{ik})^{\mathrm{T}}\mathbf{Z}_{ik}\}\mathbf{B}_r^{\mathrm{T}}\{\mathbf{w}_0(T_{ik})^{\mathrm{T}}\mathbf{Z}_{ik}\}\right\|_2\right] \\
&= \quad \sup_{\|\mathbf{a}\|_2=1} E\left[\mathbf{a}^{\mathrm{T}}\mathbf{B}_r\{\mathbf{w}_0(T_{ik})^{\mathrm{T}}\mathbf{Z}_{ik}\}\mathbf{B}_r^{\mathrm{T}}\{\mathbf{w}_0(T_{ik})^{\mathrm{T}}\mathbf{Z}_{ik}\}\mathbf{a}\right] \\
&= \quad O(h_b). \qquad \text{(S.13)}
\end{aligned}$$

The second last equality holds because $\mathbf{B}_r\{\mathbf{w}_0(T_{ik})^{\mathrm{T}}\mathbf{Z}_{ik}\}\mathbf{B}_r^{\mathrm{T}}\{\mathbf{w}_0(T_{ik})^{\mathrm{T}}\mathbf{Z}_{ik}\}$ is a symmetric matrix, the 2-norm is the maximum eigenvalue of it. So by Bernsteins inequality in Bosq (1998), we have, for random variable C_{ik}

$$n^{-1}\sum_{i=1}^{n}\sum_{k=1}^{M_i}\left\|C_{ik}\mathbf{B}_r\{\mathbf{w}_0(T_{ik})^{\mathrm{T}}\mathbf{Z}_{ik}\}\mathbf{B}_r\{\mathbf{w}_0(T_{ik})^{\mathrm{T}}\mathbf{Z}_{ik}\}^{\mathrm{T}}\right\|_2 = O_p(h_b).$$

Further note that each entry of

$$\widetilde{\mathbf{Q}}_\lambda\{\mathbf{w}_0(\mathbf{T}_i)\}^{\mathrm{T}}\mathbf{C}_i\widetilde{\mathbf{Q}}_\lambda\{\mathbf{w}_0(\mathbf{T}_i)\}$$

has the form

$$\sum_{k=1}^{M_i}\sum_{l=1}^{M_i} C_{ikl}B_{rp}\{\mathbf{w}(T_{ik})^{\mathrm{T}}\mathbf{Z}_{ik}\}B_{rp'}\{\mathbf{w}(T_{il})^{\mathrm{T}}\mathbf{Z}_{il}\}.$$

Therefore,

$$\begin{aligned}
& E|\left[\left\|\widetilde{\mathbf{Q}}_\lambda\{\mathbf{w}_0(\mathbf{T}_i)\}^{\mathrm{T}}\mathbf{C}_i\widetilde{\mathbf{Q}}_\lambda\{\mathbf{w}_0(\mathbf{T}_i)\}\right\|_2\right] \\
\asymp\ & \sum_{k=1}^{M_i}\sum_{l=1}^{M_i}\sup|C_{ikl}|E\left[\left\|\mathbf{B}_r\{\mathbf{w}_0(T_{ik}^{\mathrm{T}}\mathbf{Z}_{ik})\}\mathbf{B}_r^{\mathrm{T}}\{\mathbf{w}_0(T_{il})^{\mathrm{T}}\mathbf{Z}_{il}\}\right\|_2\right] \\
=\ & \sum_{k=1}^{M_i}\sum_{l=1}^{M_i}\sup|C_{ikl}|E\Bigg(\Bigg[\sup_{\|\mathbf{a}\|_2=1}\mathbf{a}^{\mathrm{T}}\mathbf{B}_r\{\mathbf{w}_0(T_{il})^{\mathrm{T}}\mathbf{Z}_{il}\}\mathbf{B}_r^{\mathrm{T}}\{\mathbf{w}_0(T_{ik})^{\mathrm{T}}\mathbf{Z}_{ik})\} \\
& \mathbf{B}_r\{\mathbf{w}_0(T_{ik})^{\mathrm{T}}\mathbf{Z}_{ik})\}\mathbf{B}_r^{\mathrm{T}}\{\mathbf{w}_0(T_{il})^{\mathrm{T}}\mathbf{Z}_{il}\}\mathbf{a}\Bigg]^{1/2}\Bigg) \\
=\ & \sum_{k=1}^{M_i}\sum_{l=1}^{M_i}\sup|C_{ikl}|E\left[\left\|\mathbf{B}_r\{\mathbf{w}_0(T_{ik})^{\mathrm{T}}\mathbf{Z}_{ik}\}\right\|_2\left\|\mathbf{B}_r\{\mathbf{w}_0(T_{il})^{\mathrm{T}}\mathbf{Z}_{il}\}\right\|_2\right]
\end{aligned}$$

and further because

$$E\left[\left\|\mathbf{B}_r\{\mathbf{w}_0(T_{ik})^{\mathrm{T}}\mathbf{Z}_{ik}\}\right\|_2\right]E\left[\left\|\mathbf{B}_r\{\mathbf{w}_0(T_{il})^{\mathrm{T}}\mathbf{Z}_{il}\}\right\|_2\right]$$

$$
\begin{aligned}
&\leq \quad E\left[\left\|\mathbf{B}_r\{\mathbf{w}_0(T_{ik})^{\mathrm{T}}\mathbf{Z}_{ik}\}\right\|_2\left\|\mathbf{B}_r\{\mathbf{w}_0(T_{il})^{\mathrm{T}}\mathbf{Z}_{il}\}\right\|_2\right] \\
&\leq \quad E\left[\left\|\mathbf{B}_r\{\mathbf{w}_0(T_{ik})^{\mathrm{T}}\mathbf{Z}_{ik}\}\right\|_2^2\right]^{1/2} E\left[\left\|\mathbf{B}_r\{\mathbf{w}_0(T_{il})^{\mathrm{T}}\mathbf{Z}_{il}\}\right\|_2^2\right]^{1/2}
\end{aligned}
$$

and by Lemma 1,

$$
c_b h_b \leq E\left[\left\|\mathbf{B}_r\{\mathbf{w}_0(T_{ik})^{\mathrm{T}}\mathbf{Z}_{ik}\}\right\|_2^2\right] \leq C_B h_b
$$

for some positive real numbers c_b, C_B. Then equation (S.6) follows that

$$
c_A h_b \leq E| \left[\left\|\widetilde{\mathbf{Q}}_\lambda\{\mathbf{w}_0(\mathbf{T}_i)\}^{\mathrm{T}}\mathbf{C}_i\widetilde{\mathbf{Q}}_\lambda\{\mathbf{w}_0(\mathbf{T}_i)\}\right\|_2\right] \leq C_A h_b
$$

for some positive real numbers c_A, C_A. The equation (S.7) follows by the Bernsteins inequality (Bosq (1998)).

Further note that, consider

$$
\left\|\widetilde{\mathbf{Q}}_\lambda\{\mathbf{w}_0(\mathbf{T}_i)\}^{\mathrm{T}}\mathbf{C}_i\widetilde{\mathbf{Q}}_\lambda\{\mathbf{w}_0(\mathbf{T}_i)\}\right\|_2
$$

as a random variable which is nonnegative. For any give ξ we have a $M_\xi = 1/\xi < \infty$ such that

$$
Pr\left(\frac{\left\|\widetilde{\mathbf{Q}}_\lambda\{\mathbf{w}_0(\mathbf{T}_i)\}^{\mathrm{T}}\mathbf{C}_i\widetilde{\mathbf{Q}}_\lambda\{\mathbf{w}_0(\mathbf{T}_i)\}\right\|_2}{E\left[\left\|\widetilde{\mathbf{Q}}_\lambda\{\mathbf{w}_0(\mathbf{T}_i)\}^{\mathrm{T}}\mathbf{C}_i\widetilde{\mathbf{Q}}_\lambda\{\mathbf{w}_0(\mathbf{T}_i)\}\right\|_2\right]} > M_\xi\right) < \xi
$$

by the Markov inequality, i.e.,

$$
\left\|\widetilde{\mathbf{Q}}_\lambda\{\mathbf{w}_0(\mathbf{T}_i)\}^{\mathrm{T}}\mathbf{C}_i\widetilde{\mathbf{Q}}_\lambda\{\mathbf{w}_0(\mathbf{T}_i)\}\right\|_2 / E\left[\left\|\widetilde{\mathbf{Q}}_\lambda\{\mathbf{w}_0(\mathbf{T}_i)\}^{\mathrm{T}}\mathbf{C}_i\widetilde{\mathbf{Q}}_\lambda\{\mathbf{w}_0(\mathbf{T}_i)\}\right\|_2\right]
$$

is bounded in probability. The equation (S.8) follows that

$$
\begin{aligned}
& E| \left[\left\|\widetilde{\mathbf{Q}}_\lambda\{\mathbf{w}_0(\mathbf{T}_i)\}^{\mathrm{T}}\mathbf{C}_i\widetilde{\mathbf{Q}}_\lambda\{\mathbf{w}_0(\mathbf{T}_i)\}\right\|_2\right]^{-1} \\
& \times\left\|\widetilde{\mathbf{Q}}_\lambda\{\mathbf{w}_0(\mathbf{T}_i)\}^{\mathrm{T}}\mathbf{C}_i\widetilde{\mathbf{Q}}_\lambda\{\mathbf{w}_0(\mathbf{T}_i)\}\right\|_2 \\
= \quad & O_p(1).
\end{aligned}
$$

As a result,

$$E|\left[\left\|\widetilde{\mathbf{Q}}_{\lambda}\{\mathbf{w}_0(\mathbf{T}_i)\}^{\mathrm{T}}\mathbf{C}_i\widetilde{\mathbf{Q}}_{\lambda}\{\mathbf{w}_0(\mathbf{T}_i)\}\right\|_2^2\right].$$

(S.9) and (S.10) are the consequences of equation (S.6) and (S.7) and the Theorem 13.4.3 in DeVore and Lorentz (1993). The result (S.11) follows because

$$\begin{aligned}
&E\left[\left\|\mathbf{C}_i\widetilde{\mathbf{Q}}_{\lambda}\{\mathbf{w}_0(\mathbf{T}_i)\}\right\|_2^2\right] \\
= &\sup_{\|\mathbf{a}\|_2=1}\left(\mathbf{a}^{\mathrm{T}}E\left[\widetilde{\mathbf{Q}}_{\lambda}\{\mathbf{w}_0(\mathbf{T}_i)\}^{\mathrm{T}}\mathbf{C}_i^{\mathrm{T}}\mathbf{C}_i\widetilde{\mathbf{Q}}_{\lambda}\{\mathbf{w}_0(\mathbf{T}_i)\}\right]\mathbf{a}\right) \\
= &O(h_b),
\end{aligned} \tag{S.14}$$

by equation (S.6). To show (S.12), note that

$$\begin{aligned}
&\left\|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\widetilde{\mathbf{Q}}_{\lambda i}\{\mathbf{w}_0(\mathbf{T}_i)\}^{\mathrm{T}}\boldsymbol{\Theta}_i\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_i)\}\boldsymbol{\Omega}_i^{-1}[\mathbf{D}_i\right. \\
&\left.-\mathbf{H}_i\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_i)\}]\right\|_2 \\
= &\left\|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\widetilde{\mathbf{Q}}_{\lambda i}\{\mathbf{w}_0(\mathbf{T}_i)\}^{\mathrm{T}}\boldsymbol{\Theta}_i\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_i)\}\boldsymbol{\Omega}_i^{-1}[\mathbf{D}_i\right. \\
&\left.-\mathbf{H}_i\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_i)\}])^{\mathrm{T}}\right\|_2.
\end{aligned}$$

The result follows by applying the result in (S.11). □

Lemma 2. *Let $\widehat{\boldsymbol{\lambda}}(\boldsymbol{\beta}_0, \mathbf{w}_0)$ solve the equation*

$$\begin{aligned}
&\sum_{i=1}^{n}\widetilde{\mathbf{Q}}_{\lambda i}\{\mathbf{w}_0(\mathbf{T}_i)\}^{\mathrm{T}}\boldsymbol{\Theta}_i\{\boldsymbol{\beta}_0, \boldsymbol{\lambda}, \mathbf{w}_0(\mathbf{T}_i)\}\boldsymbol{\Omega}_i^{-1} \\
&\times[\mathbf{D}_i - \mathbf{H}_i\{\boldsymbol{\beta}_0, \boldsymbol{\lambda}, \mathbf{w}_0(\mathbf{T}_i)\}] = \mathbf{0},
\end{aligned} \tag{S.15}$$

where $\boldsymbol{\Omega}_i$, $\widetilde{\mathbf{Q}}_{\lambda i}\{\mathbf{w}(\mathbf{T}_i)\}$ are defined in **Notation in Step 1** *in Section 2.1. Under the Condition (A3), we have*

(S.16)

$$\begin{aligned}
&\widehat{\boldsymbol{\lambda}}(\boldsymbol{\beta}_0, \mathbf{w}_0) - \boldsymbol{\lambda}_0 \\
= &-\left(n^{-1}\sum_{i=1}^{n}[\widetilde{\mathbf{Q}}_{\lambda i}\{\mathbf{w}_0(\mathbf{T}_i)\}^{\mathrm{T}}\boldsymbol{\Theta}_i\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_i)\}\boldsymbol{\Omega}_i^{-1}\boldsymbol{\Theta}_i\{\boldsymbol{\beta}_0, m_0,\right.
\end{aligned}$$

$$\mathbf{w}_0(\mathbf{T}_i)\}\widetilde{\mathbf{Q}}_{\lambda i}\{\mathbf{w}_0(\mathbf{T}_i)\}]\Big)^{-1} n^{-1}\sum_{i=1}^{n}\widetilde{\mathbf{Q}}_{\lambda i}\{\mathbf{w}_0(\mathbf{T}_i)\}^{\mathrm{T}}\boldsymbol{\Theta}_i\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_i)\}$$
$$\boldsymbol{\Omega}_i^{-1}[\mathbf{D}_i - \mathbf{H}_i\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_i)\}]\{1 + o_p(1)\}.$$

Therefore,

$$\begin{aligned}
& var\{\widehat{\boldsymbol{\lambda}}(\boldsymbol{\beta}_0, \mathbf{w}_0) - \boldsymbol{\lambda}_0 \mid \mathbf{R}\} \\
= & \Big(\sum_{i=1}^{n}[\widetilde{\mathbf{Q}}_{\lambda i}\{\mathbf{w}_0(\mathbf{T}_i)\}^{\mathrm{T}}\boldsymbol{\Theta}_i\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_i)\}\boldsymbol{\Omega}_i^{-1}\boldsymbol{\Theta}_i\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_i)\} \\
& \times\widetilde{\mathbf{Q}}_{\lambda i}\{\mathbf{w}_0(\mathbf{T}_i)\}]\Big)^{-1}\sum_{i=1}^{n}\widetilde{\mathbf{Q}}_{\lambda i}\{\mathbf{w}_0(\mathbf{T}_i)\}^{\mathrm{T}}\boldsymbol{\Theta}_i\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_i)\} \\
& \times\boldsymbol{\Omega}_i^{-1}\boldsymbol{\Omega}_i^{*}\boldsymbol{\Omega}_i^{-1}\boldsymbol{\Theta}_i\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_i)\}\widetilde{\mathbf{Q}}_{\lambda i}\{\mathbf{w}_0(\mathbf{T}_i)\}\Big(\sum_{i=1}^{n}[\widetilde{\mathbf{Q}}_{\lambda i}\{\mathbf{w}_0(\mathbf{T}_i)\}^{\mathrm{T}} \\
& \times\boldsymbol{\Theta}_i\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_i)\}\boldsymbol{\Omega}_i^{-1}\boldsymbol{\Theta}_i\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_i)\}\widetilde{\mathbf{Q}}_{\lambda i}\{\mathbf{w}_0(\mathbf{T}_i)\}]\Big)^{-1} \\
& \times\{1 + o_p(1)\},
\end{aligned}$$

where $\mathbf{R} = \{\mathbf{R}_i, i = 1, \ldots, n\}$. *And for bounded vector* $\mathbf{a} = (a_1, \ldots, a_\lambda)^{\mathrm{T}}$

$$\mathbf{a}^{\mathrm{T}}\{\widehat{\boldsymbol{\lambda}}(\boldsymbol{\beta}_0, \mathbf{w}_0) - \boldsymbol{\lambda}_0\} = O_p\{(nh_b)^{-1/2}\}.$$

Proof:
$\widehat{\boldsymbol{\lambda}}(\boldsymbol{\beta}_0, \mathbf{w}_0)$ solves the equation

$$\sum_{i=1}^{n}\widetilde{\mathbf{Q}}_{\lambda i}\{\mathbf{w}_0(\mathbf{T}_i)\}^{\mathrm{T}}\boldsymbol{\Theta}_i\{\boldsymbol{\beta}_0, \boldsymbol{\lambda}, \mathbf{w}_0(\mathbf{T}_i)\}\boldsymbol{\Omega}_i^{-1}[\mathbf{D}_i - \mathbf{H}_i\{\boldsymbol{\beta}_0, \boldsymbol{\lambda}, \mathbf{w}_0(\mathbf{T}_i)\}] = \mathbf{0},$$

therefore, by the standard Taylor expansion, we have

(S.17)
$$\begin{aligned}
& \widehat{\boldsymbol{\lambda}}(\boldsymbol{\beta}_0, \mathbf{w}_0) - \boldsymbol{\lambda}_0 \\
= & -\Big(n^{-1}\sum_{i=1}^{n}[\widetilde{\mathbf{Q}}_{\lambda i}\{\mathbf{w}_0(\mathbf{T}_i)\}^{\mathrm{T}}\boldsymbol{\Theta}_i\{\boldsymbol{\beta}_0, \boldsymbol{\lambda}_0, \mathbf{w}_0(\mathbf{T}_i)\}\boldsymbol{\Omega}_i^{-1}\boldsymbol{\Theta}_i\{\boldsymbol{\beta}_0, \boldsymbol{\lambda}_0, \mathbf{w}_0(\mathbf{T}_i)\} \\
& \times\widetilde{\mathbf{Q}}_{\lambda i}\{\mathbf{w}_0(\mathbf{T}_i)\}]\Big)^{-1} n^{-1}\sum_{i=1}^{n}\widetilde{\mathbf{Q}}_{\lambda i}\{\mathbf{w}_0(\mathbf{T}_i)\}^{\mathrm{T}}\boldsymbol{\Theta}_i\{\boldsymbol{\beta}_0, \boldsymbol{\lambda}_0, \mathbf{w}_0(\mathbf{T}_i)\}\boldsymbol{\Omega}_i^{-1} \\
& \times[\mathbf{D}_i - \mathbf{H}_i\{\boldsymbol{\beta}_0, \boldsymbol{\lambda}_0, \mathbf{w}_0(\mathbf{T}_i)\}]\{1 + o_p(1)\} \\
= & -\Big[\Big(n^{-1}\sum_{i=1}^{n}[\widetilde{\mathbf{Q}}_{\lambda i}\{\mathbf{w}_0(\mathbf{T}_i)\}^{\mathrm{T}}\boldsymbol{\Theta}_i\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_i)\}\boldsymbol{\Omega}_i^{-1}\boldsymbol{\Theta}_i\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_i)\}
\end{aligned}$$

$$
\begin{aligned}
&\times\widetilde{\mathbf{Q}}_{\lambda i}\{\mathbf{w}_0(\mathbf{T}_i)\}]\Big)^{-1}\Big\{n^{-1}\sum_{i=1}^{n}\widetilde{\mathbf{Q}}_{\lambda i}\{\mathbf{w}_0(\mathbf{T}_i)\}^{\mathrm{T}}\boldsymbol{\Theta}_i\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_i)\}\boldsymbol{\Omega}_i^{-1} \\
&\times[\mathbf{D}_i - \mathbf{H}_i\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_i)\}] + n^{-1}\sum_{i=1}^{n}\widetilde{\mathbf{Q}}_{\lambda i}\{\mathbf{w}_0(\mathbf{T}_i)\}^{\mathrm{T}} \\
&\times\boldsymbol{\Theta}_i\{\boldsymbol{\beta}_0, \boldsymbol{\lambda}_0, \mathbf{w}_0(\mathbf{T}_i)\}\boldsymbol{\Omega}_i^{-1}[\mathbf{D}_i - \mathbf{H}_i\{\boldsymbol{\beta}_0, \boldsymbol{\lambda}_0, \mathbf{w}_0(\mathbf{T}_i)\}] \\
&\times - n^{-1}\sum_{i=1}^{n}\widetilde{\mathbf{Q}}_{\lambda i}\{\mathbf{w}_0(\mathbf{T}_i)\}^{\mathrm{T}}\boldsymbol{\Theta}_i\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_i)\}\boldsymbol{\Omega}_i^{-1} \\
&\times[\mathbf{D}_i - \mathbf{H}_i\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_i)\}]\Big\}\Big]\{1 + o_p(1)\}
\end{aligned}
$$

We further have

$$
\begin{aligned}
&\Big\|n^{-1}\sum_{i=1}^{n}\widetilde{\mathbf{Q}}_{\lambda i}\{\mathbf{w}_0(\mathbf{T}_i)\}^{\mathrm{T}}\boldsymbol{\Theta}_i\{\boldsymbol{\beta}_0, \boldsymbol{\lambda}_0, \mathbf{w}_0(\mathbf{T}_i)\}\boldsymbol{\Omega}_i^{-1}[\mathbf{D}_i \\
&-\mathbf{H}_i\{\boldsymbol{\beta}_0, \boldsymbol{\lambda}_0, \mathbf{w}_0(\mathbf{T}_i)\}] - n^{-1}\sum_{i=1}^{n}\widetilde{\mathbf{Q}}_{\lambda i}\{\mathbf{w}_0(\mathbf{T}_i)\}^{\mathrm{T}}\boldsymbol{\Theta}_i\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_i)\} \\
&\times\boldsymbol{\Omega}_i^{-1}[\mathbf{D}_i - \mathbf{H}_i\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_i)\}]\Big\|_{\infty} \\
= \;&\Big\|n^{-1}\sum_{i=1}^{n}\Big(\frac{\partial}{\partial m(\mathbf{U}_i)^{\mathrm{T}}}\widetilde{\mathbf{Q}}_{\lambda i}\{\mathbf{w}_0(\mathbf{T}_i)\}^{\mathrm{T}}\boldsymbol{\Theta}_i\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_i)\}\boldsymbol{\Omega}_i^{-1} \\
&\times[\mathbf{D}_i - \mathbf{H}_i\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_i)\}]\{\widetilde{m}(\mathbf{U}_i) - m(\mathbf{U}_i)\}\Big)\{1 + o_p(1)\}\Big]\Big\|_{\infty} \\
\leq \;&\Big\|n^{-1}\sum_{i=1}^{n}\Big(\widetilde{\mathbf{Q}}_{\lambda i}\{\mathbf{w}_0(\mathbf{T}_i)\}^{\mathrm{T}}\frac{\partial}{\partial m(\mathbf{U}_i)^{\mathrm{T}}}\boldsymbol{\Theta}_i\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_i)\}\boldsymbol{\Omega}_i^{-1} \\
&\times[\mathbf{D}_i - \mathbf{H}_i\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_i)\}]\Big)\mathbf{1}\Big\|_{\infty} \\
&\times\sup\{\widetilde{m}(u) - m_0(u)\}\{1 + o_p(1)\} \\
= \;&O_p(h_b^{q+1})
\end{aligned}
$$

The third to the last equality holds because $\widetilde{\mathbf{Q}}_{\lambda i}\{\mathbf{w}_0(\mathbf{T}_i)\}$ is not a function of $m(\mathbf{U}_i)$. The last equality holds because the matrix,

$$
\begin{aligned}
&\widetilde{\mathbf{Q}}_{\lambda i}\{\mathbf{w}_0(\mathbf{T}_i)\}^{\mathrm{T}}\frac{\partial}{\partial m(\mathbf{U}_i)^{\mathrm{T}}}\boldsymbol{\Theta}_i\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_i)\}\boldsymbol{\Omega}_i^{-1} \\
&\times[\mathbf{D}_i - \mathbf{H}_i\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_i)\}]\mathbf{1}
\end{aligned}
$$

has the form of

$$
\sum_{k=1}^{M_i}\mathbf{C}_{ik}\mathbf{B}_r\{\mathbf{w}(T_{ik})\mathbf{Z}_{ik}\}.
$$

By the equation (S.4), we have

$$\begin{aligned}&\Big\| n^{-1}\sum_{i=1}^{n}\widetilde{\mathbf{Q}}_{\lambda i}\{\mathbf{w}_0(\mathbf{T}_i)\}^{\mathrm{T}}\frac{\partial}{\partial m(\mathbf{U}_i)^{\mathrm{T}}}\boldsymbol{\Theta}_i\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_i)\}\boldsymbol{\Omega}_i^{-1}\\ &\times[\mathbf{D}_i - \mathbf{H}_i\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_i)\}]\mathbf{1}\Big\|_\infty\\ =\ & O_p(h_b).\end{aligned}$$

Further

$$\begin{aligned}&\Big\| n^{-1}\sum_{i=1}^{n}\widetilde{\mathbf{Q}}_{\lambda i}\{\mathbf{w}_0(\mathbf{T}_i)\}^{\mathrm{T}}\boldsymbol{\Theta}_i\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_i)\}\boldsymbol{\Omega}_i^{-1}[\mathbf{D}_i\\ &-\mathbf{H}_i\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_i)\}]\Big\|_\infty\\ =\ & O_p\{\sqrt{h_b n^{-1}\log(n)}\},\end{aligned}$$

and $\sqrt{h_b n^{-1}\log(n)}/h_b^{q+1} \to \infty$, so we obtain

$$\begin{aligned}&\widehat{\boldsymbol{\lambda}}(\boldsymbol{\beta}_0, \mathbf{w}_0) - \boldsymbol{\lambda}_0\\ =\ & -\Big(n^{-1}\sum_{i=1}^{n}[\widetilde{\mathbf{Q}}_{\lambda i}\{\mathbf{w}_0(\mathbf{T}_i)\}^{\mathrm{T}}\boldsymbol{\Theta}_i\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_i)\}\boldsymbol{\Omega}_i^{-1}\boldsymbol{\Theta}_i\{\boldsymbol{\beta}_0, m_0,\\ &\mathbf{w}_0(\mathbf{T}_i)\}\widetilde{\mathbf{Q}}_{\lambda i}\{\mathbf{w}_0(\mathbf{T}_i)\}]\Big)^{-1} n^{-1}\sum_{i=1}^{n}\widetilde{\mathbf{Q}}_{\lambda i}\{\mathbf{w}_0(\mathbf{T}_i)\}^{\mathrm{T}}\boldsymbol{\Theta}_i\{\boldsymbol{\beta}_0, m_0,\\ &\mathbf{w}_0(\mathbf{T}_i)\}\boldsymbol{\Omega}_i^{-1}[\mathbf{D}_i - \mathbf{H}_i\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_i)\}]\{1 + o_p(1)\}\end{aligned}$$

Additionally, because of the term $\widetilde{\mathbf{Q}}_{\lambda i}\{\mathbf{w}_0(\mathbf{T}_i)\}$ and $E(\mathbf{D}_i - \mathbf{H}_i \mid \mathbf{R}_i) = 0$, by Corollary 1, we obtain

(S.18)

$$\begin{aligned}&\Big\|\Big(n^{-1}\sum_{i=1}^{n}[\widetilde{\mathbf{Q}}_{\lambda i}\{\mathbf{w}_0(\mathbf{T}_i)\}^{\mathrm{T}}\boldsymbol{\Theta}_i\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_i)\}\boldsymbol{\Omega}_i^{-1}\boldsymbol{\Theta}_i\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_i)\}\\ &\times\widetilde{\mathbf{Q}}_{\lambda i}\{\mathbf{w}_0(\mathbf{T}_i)\}]\Big)^{-1} n^{-1}\sum_{i=1}^{n}\widetilde{\mathbf{Q}}_{\lambda i}\{\mathbf{w}_0(\mathbf{T}_i)\}^{\mathrm{T}}\boldsymbol{\Theta}_i\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_i)\}\\ &\times\boldsymbol{\Omega}_i^{-1}[\mathbf{D}_i - \mathbf{H}_i\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_i)\}]\Big\|_2 = O_p\{(nh)^{-1/2}\}.\end{aligned}$$

This proves the result. □

Lemma 3.

$$-\boldsymbol{\delta}\{\mathbf{w}_0(T_{ik})^{\mathrm{T}}\mathbf{Z}_{ik}\} - \left\{\frac{\partial\widehat{\boldsymbol{\lambda}}(\boldsymbol{\beta}_0,\mathbf{w}_0)}{\partial\boldsymbol{\beta}^{\mathrm{T}}}\right\}^{\mathrm{T}}\mathbf{Q}_{\lambda ik} = o_p(1),$$

where $\boldsymbol{\delta}$ *is defined in* **Notations in Step 3** *in Appendix, and* $\widehat{\boldsymbol{\lambda}}(\boldsymbol{\beta}_0,\mathbf{w}_0)$ *solves (S.15) in Lemma 2.*

Proof: Note that as shown in Lemma 2, $\widehat{\boldsymbol{\lambda}}(\boldsymbol{\beta},\mathbf{w}_0)$ satisfies

$$\begin{aligned}&\sum_{i=1}^{n}\widetilde{\mathbf{Q}}_{\lambda i}\{\mathbf{w}_0(\mathbf{T}_i)\}^{\mathrm{T}}\boldsymbol{\Theta}_i\{\boldsymbol{\beta},\widehat{\boldsymbol{\lambda}}(\boldsymbol{\beta},\mathbf{w}_0),\mathbf{w}_0(\mathbf{T}_i)\}\boldsymbol{\Omega}_i^{-1}\\&\times[\mathbf{D}_i - \mathbf{H}_i\{\boldsymbol{\beta},\widehat{\boldsymbol{\lambda}}(\boldsymbol{\beta},\mathbf{w}_0),\mathbf{w}_0(\mathbf{T}_i)\}] = \mathbf{0}\end{aligned}$$

for all $\boldsymbol{\beta}$. Now taking its derivative with respect to $\boldsymbol{\beta}$ on both side, we have

$$\begin{aligned}0 &= n^{-1}\sum_{i=1}^{n}\widetilde{\mathbf{Q}}_{\lambda i}\{\mathbf{w}_0(\mathbf{T}_i)\}^{\mathrm{T}}\frac{\partial\boldsymbol{\Theta}_i\{\boldsymbol{\beta},\widehat{\boldsymbol{\lambda}}(\boldsymbol{\beta},\mathbf{w}_0),\mathbf{w}_0(\mathbf{T}_i)\}}{\partial\boldsymbol{\beta}^{\mathrm{T}}}\{\mathbf{I}\otimes(\boldsymbol{\Omega}_i^{-1}\\&\quad\times[\mathbf{D}_i - \mathbf{H}_i\{\boldsymbol{\beta},\widehat{\boldsymbol{\lambda}}(\boldsymbol{\beta},\mathbf{w}_0),\mathbf{w}_0(\mathbf{T}_i)\}])\} - n^{-1}\sum_{i=1}^{n}\widetilde{\mathbf{Q}}_{\lambda i}\{\mathbf{w}_0(\mathbf{T}_i)\}^{\mathrm{T}}\\&\quad\times\boldsymbol{\Theta}_i\{\boldsymbol{\beta},\widehat{\boldsymbol{\lambda}}(\boldsymbol{\beta},\mathbf{w}_0),\mathbf{w}_0(\mathbf{T}_i)\}\boldsymbol{\Omega}_i^{-1}\frac{\partial\mathbf{H}_i\{\boldsymbol{\beta},\widehat{\boldsymbol{\lambda}}(\boldsymbol{\beta},\mathbf{w}_0),\mathbf{w}_0(\mathbf{T}_i)\}}{\partial\boldsymbol{\beta}^{\mathrm{T}}}\\&= -n^{-1}\sum_{i=1}^{n}\widetilde{\mathbf{Q}}_{\lambda i}\{\mathbf{w}_0(\mathbf{T}_i)\}^{\mathrm{T}}\boldsymbol{\Theta}_i\{\boldsymbol{\beta},\widehat{\boldsymbol{\lambda}}(\boldsymbol{\beta},\mathbf{w}_0),\mathbf{w}_0(\mathbf{T}_i)\}\boldsymbol{\Omega}_i^{-1}\\&\quad\times\frac{\partial\mathbf{H}_i\{\boldsymbol{\beta},\widehat{\boldsymbol{\lambda}}(\boldsymbol{\beta},\mathbf{w}_0),\mathbf{w}_0(\mathbf{T}_i)\}}{\partial\boldsymbol{\beta}^{\mathrm{T}}}\{1+o_p(1)\}\\&= -n^{-1}\sum_{i=1}^{n}\widetilde{\mathbf{Q}}_{\lambda i}\{\mathbf{w}_0(\mathbf{T}_i)\}^{\mathrm{T}}\boldsymbol{\Theta}_i\{\boldsymbol{\beta},\widehat{\boldsymbol{\lambda}}(\boldsymbol{\beta},\mathbf{w}_0),\mathbf{w}_0(\mathbf{T}_i)\}\boldsymbol{\Omega}_i^{-1}\\&\quad\times\boldsymbol{\Theta}_i\{\boldsymbol{\beta},\widehat{\boldsymbol{\lambda}}(\boldsymbol{\beta},\mathbf{w}_0),\mathbf{w}_0(\mathbf{T}_i)\}\left[\widetilde{\mathbf{Q}}_{\beta i} + \widetilde{\mathbf{Q}}_{\lambda i}\{\mathbf{w}_0(\mathbf{T}_i)\}\frac{\partial\widehat{\boldsymbol{\lambda}}(\boldsymbol{\beta},\mathbf{w}_0)}{\partial\boldsymbol{\beta}^{\mathrm{T}}}\right]\\&\quad\times\{1+o_p(1)\}\end{aligned}$$

where $\mathbf{I}$ is an identity matrix of dimension $d_\beta \times d_\beta$. The second equality holds because the first term

$$n^{-1}\sum_{i=1}^{n}\widetilde{\mathbf{Q}}_{\lambda i}\{\mathbf{w}_0(\mathbf{T}_i)\}^{\mathrm{T}}\frac{\partial\boldsymbol{\Theta}_i\{\boldsymbol{\beta},\widehat{\boldsymbol{\lambda}}(\boldsymbol{\beta},\mathbf{w}_0),\mathbf{w}_0(\mathbf{T}_i)\}}{\partial\boldsymbol{\beta}^{\mathrm{T}}}\{\mathbf{I}\otimes(\boldsymbol{\Omega}_i^{-1}$$

$$\times[\mathbf{D}_i - \mathbf{H}_i\{\boldsymbol{\beta}, \widehat{\boldsymbol{\lambda}}(\boldsymbol{\beta}, \mathbf{w}_0), \mathbf{w}_0(\mathbf{T}_i)\}])\}$$

contains $\mathbf{D}_i - \mathbf{H}_i$, hence it has smaller order than the second term. Because our derivation is valid for any $\boldsymbol{\beta}$, letting $\boldsymbol{\beta} = \boldsymbol{\beta}_0$, by the consistency of $\widehat{\boldsymbol{\lambda}}(\boldsymbol{\beta}_0, \mathbf{w}_0)$ we have

$$-n^{-1}\sum_{i=1}^{n} \widetilde{\mathbf{Q}}_{\lambda i}\{\mathbf{w}_0(\mathbf{T}_i)\}^{\mathrm{T}}\boldsymbol{\Theta}_i\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_i)\}\boldsymbol{\Omega}_i^{-1}\boldsymbol{\Theta}_i\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_i)\}$$
$$\times\widetilde{\mathbf{Q}}_{\beta i}\{1 + o_p(1)\} - \mathbf{V}_n \frac{\partial\widehat{\boldsymbol{\lambda}}(\boldsymbol{\beta}_0, \mathbf{w}_0)}{\partial\boldsymbol{\beta}^{\mathrm{T}}}\{1 + o_p(1)\} = 0,$$

where $\mathbf{V}_n$ is defined in **Notation in Step 1** in Section A.2. Therefore, we can write

$$\begin{aligned}
&-\frac{\partial\widehat{\boldsymbol{\lambda}}(\boldsymbol{\beta}_0, \mathbf{w}_0)}{\partial\boldsymbol{\beta}^{\mathrm{T}}} \\
&= \mathbf{V}_n^{-1}\Bigg(n^{-1}\sum_{i=1}^{n} \widetilde{\mathbf{Q}}_{\lambda i}\{\mathbf{w}_0(\mathbf{T}_i)\}^{\mathrm{T}}\boldsymbol{\Theta}_i\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_i)\}\boldsymbol{\Omega}_i^{-1} \\
&\quad \times\boldsymbol{\Theta}_i\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_i)\}\widetilde{\mathbf{Q}}_{\beta i}\Bigg)\{1 + o_p(1)\}.
\end{aligned}$$

We denote the leading term of $-\partial\widehat{\boldsymbol{\lambda}}(\boldsymbol{\beta}_0, \mathbf{w}_0)/\partial\boldsymbol{\beta}^{\mathrm{T}}$ as

(S.19)

$$\begin{aligned}
\widehat{\boldsymbol{\Delta}} &= \mathbf{V}_n^{-1}\Bigg(n^{-1}\sum_{i=1}^{n} \widetilde{\mathbf{Q}}_{\lambda i}\{\mathbf{w}_0(\mathbf{T}_i)\}^{\mathrm{T}}\boldsymbol{\Theta}_i\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_i)\}\boldsymbol{\Omega}_i^{-1} \\
&\quad \times\boldsymbol{\Theta}_i\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_i)\}\widetilde{\mathbf{Q}}_{\beta i}\Bigg).
\end{aligned}$$

Also let

$$\begin{aligned}
\widetilde{\boldsymbol{\Delta}} &= \mathbf{V}^{-1}E\Big[\widetilde{\mathbf{Q}}_{\lambda i}\{\mathbf{w}_0(\mathbf{T}_i)\}^{\mathrm{T}}\boldsymbol{\Theta}_i\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_i)\}\boldsymbol{\Omega}_i^{-1} \\
&\quad \times\boldsymbol{\Theta}_i\{\boldsymbol{\beta}_0, \boldsymbol{\lambda}_0, \mathbf{w}_0(\mathbf{T}_i)\}\widetilde{\mathbf{Q}}_{\beta i}\Big].
\end{aligned}$$

where $\mathbf{V}_n$ and $\mathbf{V}$ are defined in **Notation in Step 1**.

By Lemma 2 and in (S.2)

$$\|\mathbf{V}_n^{-1} - \mathbf{V}^{-1}\|_2 = O_p(h_b^{-2}\sqrt{n^{-1}h_b}). \tag{S.20}$$

Because the variance of

$$\widetilde{\mathbf{Q}}_{\lambda i}\{\mathbf{w}_0(\mathbf{T}_i)\}^{\mathrm{T}}\boldsymbol{\Theta}_i\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_i)\}\boldsymbol{\Omega}_i^{-1}\boldsymbol{\Theta}_i\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_i)\}\widetilde{\mathbf{Q}}_{\beta i}$$

is of the order $O_p(h_b)$ by (S.6), we have

$$\begin{aligned}
&\Bigg\|n^{-1}\Bigg(\sum_{i=1}^{n}\widetilde{\mathbf{Q}}_{\lambda i}\{\mathbf{w}_0(\mathbf{T}_i)\}^{\mathrm{T}}\boldsymbol{\Theta}_i\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_i)\}\boldsymbol{\Omega}_i^{-1}\\
&\times\boldsymbol{\Theta}_i\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_i)\}\widetilde{\mathbf{Q}}_{\beta i}\Bigg) - E\Big[\widetilde{\mathbf{Q}}_{\lambda i}\{\mathbf{w}_0(\mathbf{T}_i)\}^{\mathrm{T}}\boldsymbol{\Theta}_i\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_i)\}\\
&\times\boldsymbol{\Omega}_i^{-1}\boldsymbol{\Theta}_i\{\boldsymbol{\beta}_0, \boldsymbol{\lambda}_0, \mathbf{w}_0(\mathbf{T}_i)\}\widetilde{\mathbf{Q}}_{\beta i}\Big]\Bigg\|_2\\
= \;& O_p(n^{-1/2}h_b^{1/2})
\end{aligned}$$

Further, by the fact shown in Corollary 1 that for any bounded random matrix $\mathbf{C}_i$,

$$\begin{aligned}
&E\left[\left\|\mathbf{C}_i\widetilde{\mathbf{Q}}_{\lambda}\{\mathbf{w}_0(\mathbf{T}_i)\}\right\|_2^2\right]\\
= \;& O(h_b),
\end{aligned}$$

we have

$$\begin{aligned}
&\left\|E\left[\widetilde{\mathbf{Q}}_{\lambda i}\{\mathbf{w}_0(\mathbf{T}_i)\}^{\mathrm{T}}\boldsymbol{\Theta}_i\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_i)\}\boldsymbol{\Omega}_i^{-1}\boldsymbol{\Theta}_i\{\boldsymbol{\beta}_0, \boldsymbol{\lambda}_0, \mathbf{w}_0(\mathbf{T}_i)\}\widetilde{\mathbf{Q}}_{\beta i}\right]\right\|_2\\
= \;& O(h_b^{1/2}).
\end{aligned}$$

Combining the above results and the fact that $\|\mathbf{V}^{-1}\|_2 = O(h_b^{-1})$, we obtain

$$\begin{aligned}
&\|\widetilde{\boldsymbol{\Delta}} - \widehat{\boldsymbol{\Delta}}\|_2\\
\leq \;& \|\mathbf{V}_n^{-1} - \mathbf{V}^{-1}\|_2\Big\|E\Big[\widetilde{\mathbf{Q}}_{\lambda i}\{\mathbf{w}_0(\mathbf{T}_i)\}^{\mathrm{T}}\boldsymbol{\Theta}_i\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_i)\}\boldsymbol{\Omega}_i^{-1}\\
&\times\boldsymbol{\Theta}_i\{\boldsymbol{\beta}_0, \boldsymbol{\lambda}_0, \mathbf{w}_0(\mathbf{T}_i)\}\widetilde{\mathbf{Q}}_{\beta i}\Big]\Big\|_2 + \|\mathbf{V}\|_2\Bigg\|n^{-1}\Bigg(\sum_{i=1}^{n}\widetilde{\mathbf{Q}}_{\lambda i}\{\mathbf{w}_0(\mathbf{T}_i)\}^{\mathrm{T}}\\
&\times\boldsymbol{\Theta}_i\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_i)\}\boldsymbol{\Omega}_i^{-1}\boldsymbol{\Theta}_i\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_i)\}\widetilde{\mathbf{Q}}_{\beta i}\Bigg)\\
&-E\left[\widetilde{\mathbf{Q}}_{\lambda i}\{\mathbf{w}_0(\mathbf{T}_i)\}^{\mathrm{T}}\boldsymbol{\Theta}_i\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_i)\}\boldsymbol{\Omega}_i^{-1}\boldsymbol{\Theta}_i\{\boldsymbol{\beta}_0, \boldsymbol{\lambda}_0, \mathbf{w}_0(\mathbf{T}_i)\}\widetilde{\mathbf{Q}}_{\beta i}\right]\Bigg\|_2\\
= \;& O_p(n^{-1/2}h_b^{-1}).
\end{aligned}$$

Also

(S.21)

$$\begin{aligned}
&\Bigg\| E\Bigg(\Big[\mathbf{Q}_{\lambda ik}\{\mathbf{w}_0(T_{ik})\}^{\mathrm{T}}\widetilde{\boldsymbol{\Delta}} - \mathbf{Q}_{\lambda ik}\{\mathbf{w}_0(T_{ik})\}^{\mathrm{T}}\widehat{\boldsymbol{\Delta}}\Big] \Big[\mathbf{Q}_{\lambda ik}\{\mathbf{w}_0(T_{ik})\}^{\mathrm{T}}\widetilde{\boldsymbol{\Delta}} \\
&\times - \mathbf{Q}_{\lambda ik}\{\mathbf{w}_0(T_{ik})\}^{\mathrm{T}}\widehat{\boldsymbol{\Delta}}\Big]^{\mathrm{T}}\Bigg)\Bigg\|_2 \\
= &\Bigg\| E\Big[(\widetilde{\boldsymbol{\Delta}} - \widehat{\boldsymbol{\Delta}})^{\mathrm{T}}\mathbf{Q}_{\lambda ik}\{\mathbf{w}_0(T_{ik})\}\mathbf{Q}_{\lambda ik}\{\mathbf{w}_0(T_{ik})\}^{\mathrm{T}}(\widetilde{\boldsymbol{\Delta}} - \widehat{\boldsymbol{\Delta}})\Big]\Bigg\|_2 \\
= &\Bigg\| E\Big[(\widetilde{\boldsymbol{\Delta}} - \widehat{\boldsymbol{\Delta}})^{\mathrm{T}}\mathbf{B}_r\{\mathbf{w}_0(T_{ik})^{\mathrm{T}}\mathbf{Z}_{ik}\}\mathbf{B}_r^{\mathrm{T}}\{\mathbf{w}_0(T_{ik})^{\mathrm{T}}\mathbf{Z}_{ik}\}(\widetilde{\boldsymbol{\Delta}} - \widehat{\boldsymbol{\Delta}})\Big]\Bigg\|_2 \\
\leq &\; h_b\|\widetilde{\boldsymbol{\Delta}} - \widehat{\boldsymbol{\Delta}}\|_2^2 \\
= &\; O_p\{(nh_b)^{-1}\}.
\end{aligned}$$

by Lemma 1.

Now because $\mathbf{Q}_{\lambda ik}\{\mathbf{w}_0(T_{ik})\} = \mathbf{B}_r\{\mathbf{w}_0(T_{ik})^{\mathrm{T}}\mathbf{Z}_{ik}\}$ is a vector of B-spline bases, there exists a function $\delta_l\{\mathbf{w}_0(t_0)^{\mathrm{T}}\mathbf{Z}_{ik}\}$, $\delta_l \in L_2(0, c), c < \infty$, s.t.

$$E\Bigg(\Big[\mathbf{Q}_{\lambda ik}\{\mathbf{w}_0(T_{ik})\}^{\mathrm{T}}\widetilde{\boldsymbol{\Delta}}_l - \delta_l\{\mathbf{w}_0(T_{ik})^{\mathrm{T}}\mathbf{Z}_{ik}\}\Big]^2\Bigg) = O(h_b^{2q}),$$

by **Fact 1**, where $\widetilde{\boldsymbol{\Delta}}_l$ is the lth column of $\widetilde{\boldsymbol{\Delta}}$. Therefore we have

$$\begin{aligned}
&E\Bigg(\Big[\mathbf{Q}_{\lambda ik}\{\mathbf{w}_0(T_{ik})\}^{\mathrm{T}}\widetilde{\boldsymbol{\Delta}} - \boldsymbol{\delta}^{\mathrm{T}}\{\mathbf{w}_0(T_{ik})^{\mathrm{T}}\mathbf{Z}_{ik}\}\Big] \\
&\times \Big[\mathbf{Q}_{\lambda ik}\{\mathbf{w}_0(T_{ik})\}^{\mathrm{T}}\widetilde{\boldsymbol{\Delta}} - \boldsymbol{\delta}^{\mathrm{T}}\{\mathbf{w}_0(T_{ik})^{\mathrm{T}}\mathbf{Z}_{ik}\}\Big]^{\mathrm{T}}\Bigg) = O(h_b^{2q}).
\end{aligned}$$

where $\boldsymbol{\delta} = (\delta_1, \ldots, \delta_{d_\beta})$. Combining with (S.21), we obtain

$$\begin{aligned}
&E\Bigg(\Big[\mathbf{Q}_{\lambda ik}\{\mathbf{w}_0(T_{ik})\}^{\mathrm{T}}\widehat{\boldsymbol{\Delta}} - \boldsymbol{\delta}^{\mathrm{T}}\{\mathbf{w}_0(T_{ik})^{\mathrm{T}}\mathbf{Z}_{ik}\}\Big] \\
&\times \Big[\mathbf{Q}_{\lambda ik}\{\mathbf{w}_0(T_{ik})\}^{\mathrm{T}}\widehat{\boldsymbol{\Delta}} - \boldsymbol{\delta}^{\mathrm{T}}\{\mathbf{w}_0(T_{ik})^{\mathrm{T}}\mathbf{Z}_{ik}\}\Big]^{\mathrm{T}}\Bigg) \\
= &\; O\{(nh_b)^{-1} + h_b^{2q}\}.
\end{aligned}$$

So

$$\boldsymbol{\delta}\{\mathbf{w}_0(T_{ik})^{\mathrm{T}}\mathbf{Z}_{ik}\} - \widehat{\boldsymbol{\Delta}}^{\mathrm{T}}\mathbf{Q}_{\lambda ik} = o_p(1)$$

and in turn

$$-\boldsymbol{\delta}\{\mathbf{w}_0(T_{ik})^{\mathrm{T}}\mathbf{Z}_{ik}\} - \left\{\frac{\partial \widehat{\boldsymbol{\lambda}}(\boldsymbol{\beta}_0, \mathbf{w}_0)}{\partial \boldsymbol{\beta}^{\mathrm{T}}}\right\}^{\mathrm{T}} \mathbf{Q}_{\lambda ik} = o_p(1).$$

Now we show $\boldsymbol{\delta}$ is as defined in **Notation in Step 3**. Note that, $\widetilde{\boldsymbol{\Delta}}$ could also be obtained as

$$\begin{aligned}\widetilde{\boldsymbol{\Delta}} &= \underset{\boldsymbol{\Delta}\in\mathrm{R}^{\mathbf{d}_\lambda\times\mathbf{d}_\beta}}{\arg\min}\ \mathbf{1}_\beta^{\mathrm{T}} E\Bigg(\left[\widetilde{\mathbf{Q}}_{\beta i} - \widetilde{\mathbf{Q}}_{\lambda i}\{\mathbf{w}_0(\mathbf{T}_i)\}\boldsymbol{\Delta}\right]^{\mathrm{T}} \Theta_i\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_i)\}\boldsymbol{\Omega}_i^{-1} \\ &\quad \times\Theta_i\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_i)\}\left[\widetilde{\mathbf{Q}}_{\beta i} - \widetilde{\mathbf{Q}}_{\lambda i}\{\mathbf{w}_0(\mathbf{T}_i)\}\boldsymbol{\Delta}\right]\Bigg)\mathbf{1}_\beta,\end{aligned}$$

so we conclude that $\boldsymbol{\delta}$ minimizes

$$\begin{aligned}&\mathbf{1}_{d_\beta}^{\mathrm{T}} E\Bigg(\left[\widetilde{\mathbf{Q}}_{\beta i} - \boldsymbol{\delta}\{\mathbf{U}_i(\mathbf{T}_i)\}\right]^{\mathrm{T}} \Theta_i\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_i)\}\boldsymbol{\Omega}_i^{-1}\Theta_i\{\boldsymbol{\beta}_0, m_0, \\ &\mathbf{w}_0(\mathbf{T}_i)\}\left[\widetilde{\mathbf{Q}}_{\beta i} - \boldsymbol{\delta}\{\mathbf{U}_i(\mathbf{T}_i)\}\right]\Bigg)\mathbf{1}_{d_\beta}\end{aligned}$$

as defined in **Notation in Step 3**.

Lemma 4.

$$\begin{aligned}&\frac{\partial\widehat{\boldsymbol{\lambda}}(\boldsymbol{\beta}_0, \mathbf{w}_0)}{\partial\mathbf{w}}(\mathbf{w}_h) \\ =\ & -\Bigg(n^{-1}\sum_{i=1}^n [\widetilde{\mathbf{Q}}_{\lambda i}\{\mathbf{w}_0(\mathbf{T}_i)\}^{\mathrm{T}}\boldsymbol{\Theta}_i\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_i)\}\boldsymbol{\Omega}_i^{-1}\boldsymbol{\Theta}_i\{\boldsymbol{\beta}_0, m_0, \\ &\mathbf{w}_0(\mathbf{T}_i)\}\widetilde{\mathbf{Q}}_{\lambda i}\{\mathbf{w}_0(\mathbf{T}_i)\}]\Bigg)^{-1} n^{-1}\left[\sum_{i=1}^n \widetilde{\mathbf{Q}}_{\lambda i}\{\mathbf{w}_0(\mathbf{T}_i)\}^{\mathrm{T}}\boldsymbol{\Theta}_i\{\boldsymbol{\beta}_0, m_0,\right. \\ &\left.\mathbf{w}_0(\mathbf{T}_i)\}\boldsymbol{\Omega}_i^{-1}\boldsymbol{\Theta}_i\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_i)\}\widetilde{\mathbf{Q}}_{wi}\{m_0, \mathbf{w}_h(\mathbf{T}_i)\}\right] \\ &\times\{1 + o_p(1)\}\end{aligned}$$

where $\widetilde{\mathbf{Q}}_{wi}\{m_0, \mathbf{w}_h(\mathbf{T}_i)\}$ *is defined in* **NotationinStep1** *in Appendix.*

Proof: Note that $\mathbf{w}$ here is a function, and so $\frac{\partial\widehat{\boldsymbol{\lambda}}(\boldsymbol{\beta}_0, \mathbf{w}_0)}{\partial\mathbf{w}}$ is a functional derivative with respect to function $\mathbf{w}$, i.e. the Gâteaux derivative. Assume $\mathbf{w}_0 + \xi\mathbf{w}_h$ is a function in the space $\mathcal{D}$, and $\mathbf{w}_h$ is in the tangent space of $\mathcal{D}$, then

$$\begin{aligned}&\widehat{\boldsymbol{\lambda}}(\boldsymbol{\beta}_0, \mathbf{w}_0 + \xi\mathbf{w}_h) \\ =\ &\Bigg(n^{-1}\sum_{i=1}^n \left[\widetilde{\mathbf{Q}}_{\lambda i}\{(\mathbf{w}_0 + \xi\mathbf{w}_h)(\mathbf{T}_i)\}^{\mathrm{T}}\boldsymbol{\Theta}_i\{\boldsymbol{\beta}_0, m_0, (\mathbf{w}_0 + \xi\mathbf{w}_h)(\mathbf{T}_i)\}\boldsymbol{\Omega}_i^{-1}\right. \\ &\left.\times\boldsymbol{\Theta}_i\{\boldsymbol{\beta}_0, m_0, (\mathbf{w}_0 + \xi\mathbf{w}_h)(\mathbf{T}_i)\}\widetilde{\mathbf{Q}}_{\lambda i}\{(\mathbf{w}_0 + \xi\mathbf{w}_h)(\mathbf{T}_i)\}\right]\Bigg)^{-1}\end{aligned}$$

$$\times n^{-1}\Bigg[\sum_{i=1}^{n}\widetilde{\mathbf{Q}}_{\lambda i}\{(\mathbf{w}_0+\xi\mathbf{w}_h)(\mathbf{T}_i)\}\boldsymbol{\Theta}_i\{\boldsymbol{\beta}_0,m_0,(\mathbf{w}_0+\xi\mathbf{w}_h)(\mathbf{T}_i)\}\boldsymbol{\Omega}_i^{-1}$$
$$\times[\mathbf{D}_i-\mathbf{H}_i\{\boldsymbol{\beta}_0,m_0,(\mathbf{w}_0+\xi\mathbf{w}_h)(\mathbf{T}_i)\}]\Bigg]\{1+o_p(1)\}$$

Taking derivative of $\widehat{\boldsymbol{\lambda}}(\boldsymbol{\beta}_0,\mathbf{w}_0+\xi\mathbf{w}_h)$ with respect to ξ and evaluating at $\xi=0$, we have

$$\begin{aligned}
&\left.\frac{\partial\widehat{\boldsymbol{\lambda}}(\boldsymbol{\beta}_0,\mathbf{w}_0+\xi\mathbf{w}_h)}{\partial\xi}\right|_{\xi=0}\\
=\ &-\Bigg(n^{-1}\sum_{i=1}^{n}[\widetilde{\mathbf{Q}}_{\lambda i}\{\mathbf{w}_0(\mathbf{T}_i)\}^{\mathrm{T}}\boldsymbol{\Theta}_i\{\boldsymbol{\beta}_0,m_0,\mathbf{w}_0(\mathbf{T}_i)\}\boldsymbol{\Omega}_i^{-1}\boldsymbol{\Theta}_i\{\boldsymbol{\beta}_0,m_0,\\
&\mathbf{w}_0(\mathbf{T}_i)\}\widetilde{\mathbf{Q}}_{\lambda i}\{\mathbf{w}_0(\mathbf{T}_i)\}]\Bigg)^{-1}n^{-1}\Bigg[\sum_{i=1}^{n}\widetilde{\mathbf{Q}}_{\lambda i}\{\mathbf{w}_0(\mathbf{T}_i)\}^{\mathrm{T}}\boldsymbol{\Theta}_i\{\boldsymbol{\beta}_0,m_0,\\
&\mathbf{w}_0(\mathbf{T}_i)\}\boldsymbol{\Omega}_i^{-1}\boldsymbol{\Theta}_i\{\boldsymbol{\beta}_0,m_0,\mathbf{w}_0(\mathbf{T}_i)\}\widetilde{\mathbf{Q}}_{wi}\{m_0,\mathbf{w}_h(\mathbf{T}_i)\}\Bigg]\{1+o_p(1)\}
\end{aligned}$$

where

$$\begin{aligned}
\widetilde{\mathbf{Q}}_{wi}\{m_0,\mathbf{w}(\mathbf{T}_i)\} &= [m_0'\{\mathbf{w}_0(T_{i1})^{\mathrm{T}}\mathbf{Z}_{i1}\}\mathbf{w}_h(T_{i1})^{\mathrm{T}}\mathbf{Z}_{i1},\ldots,\\
&\quad m_0'\{\mathbf{w}_0(T_{iM_i})^{\mathrm{T}}\mathbf{Z}_{iM_i}\}\mathbf{w}_h(T_{iM_i})^{\mathrm{T}}\mathbf{Z}_{iM_i}]^{\mathrm{T}}.
\end{aligned}$$

Therefore the Gâteaux derivative is a linear function in the tangent space, and its value at $\mathbf{w}_h$ is

$$\begin{aligned}
&\frac{\partial\widehat{\boldsymbol{\lambda}}(\boldsymbol{\beta}_0,\mathbf{w}_0)}{\partial\mathbf{w}}(\mathbf{w}_h)\\
=\ &-\Bigg(n^{-1}\sum_{i=1}^{n}[\widetilde{\mathbf{Q}}_{\lambda i}\{\mathbf{w}_0(\mathbf{T}_i)\}^{\mathrm{T}}\boldsymbol{\Theta}_i\{\boldsymbol{\beta}_0,m_0,\mathbf{w}_0(\mathbf{T}_i)\}\boldsymbol{\Omega}_i^{-1}\boldsymbol{\Theta}_i\{\boldsymbol{\beta}_0,m_0,\\
&\mathbf{w}_0(\mathbf{T}_i)\}\widetilde{\mathbf{Q}}_{\lambda i}\{\mathbf{w}_0(\mathbf{T}_i)\}]\Bigg)^{-1}n^{-1}\Bigg[\sum_{i=1}^{n}\widetilde{\mathbf{Q}}_{\lambda i}\{\mathbf{w}_0(\mathbf{T}_i)\}^{\mathrm{T}}\boldsymbol{\Theta}_i\{\boldsymbol{\beta}_0,m_0,\mathbf{w}_0(\mathbf{T}_i)\}\\
&\times\boldsymbol{\Omega}_i^{-1}\boldsymbol{\Theta}_i\{\boldsymbol{\beta}_0,m_0,\mathbf{w}_0(\mathbf{T}_i)\}\widetilde{\mathbf{Q}}_{wi}\{m_0,\mathbf{w}_h(\mathbf{T}_i)\}\Bigg]\\
&\times\{1+o_p(1)\}.
\end{aligned}$$

Lemma 5.

$$-\eta(\mathbf{w}_0(T_{ik})^{\mathrm{T}}\mathbf{Z}_{ik})(\mathbf{w}_h)-\mathbf{Q}_{\lambda ik}^{\mathrm{T}}\frac{\partial\widehat{\boldsymbol{\lambda}}(\boldsymbol{\beta}_0,\mathbf{w}_0)}{\partial\mathbf{w}}(\mathbf{w}_h)=o_p(1)$$

for each $\mathbf{w}_h$ *in* E, *i.e.*

$$\sup_{\mathbf{w}_h \in \mathcal{D}} \left| -\eta(\mathbf{w}_0(T_{ik})^{\mathrm{T}} \mathbf{Z}_{ik})(\mathbf{w}_h) - \mathbf{Q}_{\lambda ik}^{\mathrm{T}} \frac{\partial \widehat{\boldsymbol{\lambda}}(\boldsymbol{\beta}_0, \mathbf{w}_0)}{\partial \mathbf{w}}(\mathbf{w}_h) \right| = o_p(1)$$

where η *is defined in* **Notations in Step 2**.

Proof: Note that as shown in Lemma 4

$$\begin{aligned}
& \frac{\partial \widehat{\boldsymbol{\lambda}}(\boldsymbol{\beta}_0, \mathbf{w}_0)}{\partial \mathbf{w}}(\mathbf{w}_h) \\
= \; & -\Bigg(n^{-1} \sum_{i=1}^{n} [\widetilde{\mathbf{Q}}_{\lambda i}\{\mathbf{w}_0(\mathbf{T}_i)\}^{\mathrm{T}} \boldsymbol{\Theta}_i\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_i)\} \boldsymbol{\Omega}_i^{-1} \boldsymbol{\Theta}_i\{\boldsymbol{\beta}_0, m_0, \\
& \mathbf{w}_0(\mathbf{T}_i)\} \widetilde{\mathbf{Q}}_{\lambda i}\{\mathbf{w}_0(\mathbf{T}_i)\}] \Bigg)^{-1} n^{-1} \Bigg[\sum_{i=1}^{n} \widetilde{\mathbf{Q}}_{\lambda i}\{\mathbf{w}_0(\mathbf{T}_i)\}^{\mathrm{T}} \boldsymbol{\Theta}_i\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_i)\} \\
& \times \boldsymbol{\Omega}_i^{-1} \boldsymbol{\Theta}_i\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_i)\} \widetilde{\mathbf{Q}}_{wi}\{m_0, \mathbf{w}_h(\mathbf{T}_i)\} \Bigg] \\
& \times \{1 + o_p(1)\}.
\end{aligned}$$

The rest of the proofs follow the same arguments as those lead to Lemma 3, except we replace

$$\begin{aligned}
& n^{-1} \sum_{i=1}^{n} \widetilde{\mathbf{Q}}_{\lambda i}\{\mathbf{w}_0(\mathbf{T}_i)\}^{\mathrm{T}} \boldsymbol{\Theta}_i\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_i)\} \boldsymbol{\Omega}_i^{-1} \\
& \times \boldsymbol{\Theta}_i\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_i)\} \widetilde{\mathbf{Q}}_{\beta i}
\end{aligned}$$

by

$$\begin{aligned}
& n^{-1} \sum_{i=1}^{n} \widetilde{\mathbf{Q}}_{\lambda i}\{\mathbf{w}_0(\mathbf{T}_i)\}^{\mathrm{T}} \boldsymbol{\Theta}_i\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_i)\} \boldsymbol{\Omega}_i^{-1} \\
& \times \boldsymbol{\Theta}_i\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_i)\} \widetilde{\mathbf{Q}}_{wi}\{m_0, \mathbf{w}_h(\mathbf{T}_i)\}.
\end{aligned}$$

□

Proof of Theorem 1: Applying a Taylor expansion, and using the fact that

$$E\{\mathbf{S}_{wi}(\boldsymbol{\beta}_0, \boldsymbol{\lambda}_0, \mathbf{w}_0) | \mathbf{z}_i, \mathbf{x}_i, \mathbf{t}_i\} = O(h_q),$$

we have

$$\frac{1}{n} \sum_{i=1}^{n} \widehat{\mathbf{A}}_{wi}\{\boldsymbol{\beta}_0, \widehat{\boldsymbol{\lambda}}(\boldsymbol{\beta}_0, \mathbf{w}_0(t_0)), \mathbf{w}_0\} \widehat{\mathbf{V}}_{wi}\{\boldsymbol{\beta}_0, \widehat{\boldsymbol{\lambda}}(\boldsymbol{\beta}_0, \mathbf{w}_0), \mathbf{w}_0(t_0)\}^{-1}$$

$$
\begin{aligned}
&\times \mathbf{K}_h(\mathbf{T}_i - t_0)\widehat{\mathbf{S}}_{wi}\{\boldsymbol{\beta}_0, \widehat{\boldsymbol{\lambda}}(\boldsymbol{\beta}_0, \mathbf{w}_0), \mathbf{w}_0(t_0)\} - \frac{1}{n}\sum_{i=1}^n \widehat{\mathbf{A}}_{wi}(\boldsymbol{\beta}_0, \boldsymbol{\lambda}_0, \mathbf{w}_0(t_0)) \\
&\times \widehat{\mathbf{V}}_{wi}(\boldsymbol{\beta}_0, \boldsymbol{\lambda}_0, \mathbf{w}_0(t_0))^{-1}\mathbf{K}_h(\mathbf{T}_i - t_0)\widehat{\mathbf{S}}_{wi}(\boldsymbol{\beta}_0, \boldsymbol{\lambda}_0, \mathbf{w}_0(t_0)) \\
= {} & \frac{1}{n}\sum_{i=1}^n \widehat{\mathbf{A}}_{wi}(\boldsymbol{\beta}_0, \boldsymbol{\lambda}_0, \mathbf{w}_0(t_0))\widehat{\mathbf{V}}_{wi}(\boldsymbol{\beta}_0, \boldsymbol{\lambda}_0, \mathbf{w}_0(t_0))^{-1}\mathbf{K}_h(\mathbf{T}_i - t_0) \\
&\times \frac{\partial}{\partial \boldsymbol{\lambda}^{\mathrm{T}}}\widehat{\mathbf{S}}_{wi}(\boldsymbol{\beta}_0, \boldsymbol{\lambda}_0, \mathbf{w}_0(t_0))\{\widehat{\boldsymbol{\lambda}}(\boldsymbol{\beta}_0, \mathbf{w}_0) - \boldsymbol{\lambda}_0\}\{1 + o_p(1)\}.
\end{aligned}
$$

The matrix

$$
\begin{aligned}
&\frac{1}{n}\sum_{i=1}^n \widehat{\mathbf{A}}_{wi}\{\boldsymbol{\beta}_0, \boldsymbol{\lambda}_0, \mathbf{w}_0(t_0)\}\widehat{\mathbf{V}}_{wi}\{\boldsymbol{\beta}_0, \boldsymbol{\lambda}_0, \mathbf{w}_0(t_0)\}^{-1} \\
&\times \mathbf{K}_h(\mathbf{T}_i - t_0)\frac{\partial}{\partial \boldsymbol{\lambda}^{\mathrm{T}}}\widehat{\mathbf{S}}_{wi}\{\boldsymbol{\beta}_0, \boldsymbol{\lambda}_0, \mathbf{w}_0(t_0)\}
\end{aligned}
$$

has the form

$$
\frac{1}{n}\sum_{i=1}^n \mathbf{C}_i \widetilde{\mathbf{Q}}_\lambda\{\mathbf{w}_0(\mathbf{T}_i)\}
$$

for a random quantity $E(\mathbf{C}_i \mid \mathbf{R}_i) \neq \mathbf{0}$, and is of order $O_p(h_b^{1/2})$ following Corollary 1. Following Lemma 2, we have

$$
\begin{aligned}
&\Bigg\| \frac{1}{n}\sum_{i=1}^n \widehat{\mathbf{A}}_{wi}\{\boldsymbol{\beta}_0, \widehat{\boldsymbol{\lambda}}(\boldsymbol{\beta}_0, \mathbf{w}_0), \mathbf{w}_0(t_0)\}\widehat{\mathbf{V}}_{wi}\{\boldsymbol{\beta}_0, \widehat{\boldsymbol{\lambda}}(\boldsymbol{\beta}_0, \mathbf{w}_0), \mathbf{w}_0(t_0)\}^{-1} \\
&\times \mathbf{K}_h(\mathbf{T}_i - t_0)\widehat{\mathbf{S}}_{wi}\{\boldsymbol{\beta}_0, \widehat{\boldsymbol{\lambda}}(\boldsymbol{\beta}_0, \mathbf{w}_0), \mathbf{w}_0(t_0)\} - \frac{1}{n}\sum_{i=1}^n \widehat{\mathbf{A}}_{wi}\{\boldsymbol{\beta}_0, \boldsymbol{\lambda}_0, \mathbf{w}_0(t_0)\} \\
&\times \widehat{\mathbf{V}}_{wi}\{\boldsymbol{\beta}_0, \boldsymbol{\lambda}_0, \mathbf{w}_0(t_0)\}^{-1}\mathbf{K}_h(\mathbf{T}_i - t_0)\widehat{\mathbf{S}}_{wi}\{\boldsymbol{\beta}_0, \boldsymbol{\lambda}_0, \mathbf{w}_0\} \Bigg\|_2 \\
= {} & O_p(h_b^{1/2})O_p\{\sqrt{(nh_b)^{-1}}\} = O_p(n^{-1/2}).
\end{aligned}
$$

Because the above matrices are of fixed dimension, each term of the above difference is of order $O_p(n^{-1/2})$.

Using the above result, we have

(S.22)

$$
\begin{aligned}
\mathbf{0} = {} & \frac{1}{n}\sum_{i=1}^n \widehat{\mathbf{A}}_{wi}[\boldsymbol{\beta}_0, \widehat{\boldsymbol{\lambda}}\{\boldsymbol{\beta}_0, \widehat{\mathbf{w}}(\boldsymbol{\beta}_0)\}, \widehat{\mathbf{w}}(\boldsymbol{\beta}_0, t_0)]\widehat{\mathbf{V}}_{wi}[\boldsymbol{\beta}_0, \widehat{\boldsymbol{\lambda}}\{\boldsymbol{\beta}_0, \widehat{\mathbf{w}}(\boldsymbol{\beta}_0)\}, \\
& \widehat{\mathbf{w}}(\boldsymbol{\beta}_0, t_0)]^{-1}\mathbf{K}_h(\mathbf{T}_i - \mathbf{t}_0)\widehat{\mathbf{S}}_{wi}[\boldsymbol{\beta}_0, \widehat{\boldsymbol{\lambda}}\{\boldsymbol{\beta}_0, \widehat{\mathbf{w}}(\boldsymbol{\beta}_0)\}, \widehat{\mathbf{w}}(\boldsymbol{\beta}_0, t_0)]
\end{aligned}
$$

$$
\begin{aligned}
&= \frac{1}{n}\sum_{i=1}^{n}\widehat{\mathbf{A}}_{wi}\{\boldsymbol{\beta}_0,\widehat{\boldsymbol{\lambda}}(\boldsymbol{\beta}_0,\mathbf{w}_0),\mathbf{w}_0(t_0)\}\widehat{\mathbf{V}}_{wi}\{\boldsymbol{\beta}_0,\widehat{\boldsymbol{\lambda}}(\boldsymbol{\beta}_0,\mathbf{w}_0),\mathbf{w}_0(t_0)\}^{-1} \\
&\quad \times\mathbf{K}_h(\mathbf{T}_i-\mathbf{t}_0)\widehat{\mathbf{S}}_{wi}\{\boldsymbol{\beta}_0,\widehat{\boldsymbol{\lambda}}(\boldsymbol{\beta}_0,\mathbf{w}_0),\mathbf{w}_0(t_0)\} \\
&\quad +n^{-1}\Bigg(\sum_{i=1}^{n}\Bigg[\frac{\partial}{\partial\mathbf{w}}\widehat{\mathbf{A}}_{wi}\{\boldsymbol{\beta}_0,\widehat{\boldsymbol{\lambda}}(\boldsymbol{\beta}_0,\mathbf{w}_0),\mathbf{w}_0(t_0)\}\widehat{\mathbf{V}}_{wi}\{\boldsymbol{\beta}_0,\widehat{\boldsymbol{\lambda}}(\boldsymbol{\beta}_0,\mathbf{w}_0), \\
&\quad \mathbf{w}_0(t_0)\}^{-1}\mathbf{K}_h(\mathbf{T}_i-\mathbf{t}_0)\widehat{\mathbf{S}}_{wi}\{\boldsymbol{\beta}_0,\widehat{\boldsymbol{\lambda}}(\boldsymbol{\beta}_0,\mathbf{w}_0),\mathbf{w}_0(t_0)\}\Bigg]+o_p(1)\Bigg) \\
&\quad \times\{\widehat{\mathbf{w}}(\boldsymbol{\beta}_0)-\mathbf{w}_0\} \\
&= \frac{1}{n}\sum_{i=1}^{n}\widehat{\mathbf{A}}_{wi}(\boldsymbol{\beta}_0,\boldsymbol{\lambda}_0,\mathbf{w}_0(t_0))\widehat{\mathbf{V}}_{wi}(\boldsymbol{\beta}_0,\boldsymbol{\lambda}_0,\mathbf{w}_0(t_0))^{-1}\mathbf{K}_h(\mathbf{T}_i-\mathbf{t}_0) \\
&\quad \times\widehat{\mathbf{S}}_{wi}(\boldsymbol{\beta}_0,\boldsymbol{\lambda}_0,\mathbf{w}_0(t_0))+O_p(n^{-1/2}) \\
&\quad +n^{-1}\Bigg(\sum_{i=1}^{n}\Bigg[\frac{\partial}{\partial\mathbf{w}}\widehat{\mathbf{A}}_{wi}\{\boldsymbol{\beta}_0,\widehat{\boldsymbol{\lambda}}(\boldsymbol{\beta}_0,\mathbf{w}_0),\mathbf{w}_0(t_0)\}\widehat{\mathbf{V}}_{wi}\{\boldsymbol{\beta}_0,\widehat{\boldsymbol{\lambda}}(\boldsymbol{\beta}_0,\mathbf{w}_0), \\
&\quad \mathbf{w}_0(t_0)\}^{-1}\mathbf{K}_h(\mathbf{T}_i-\mathbf{t}_0)\widehat{\mathbf{S}}_{wi}\{\boldsymbol{\beta}_0,\widehat{\boldsymbol{\lambda}}(\boldsymbol{\beta}_0,\mathbf{w}_0),\mathbf{w}_0(t_0)\}\Bigg]+o_p(1)\Bigg) \\
&\quad \times\{\widehat{\mathbf{w}}(\boldsymbol{\beta}_0)-\mathbf{w}_0\} \\
&= \frac{1}{n}\sum_{i=1}^{n}\mathbf{A}_{wi}(\boldsymbol{\beta}_0,m_0,\mathbf{w}_0)\mathbf{V}_{wi}(\boldsymbol{\beta}_0,m_0,\mathbf{w}_0)^{-1}\mathbf{K}_h(\mathbf{T}_i-\mathbf{t}_0) \\
&\quad \times\mathbf{S}_{wi}(\boldsymbol{\beta}_0,m_0,\mathbf{w}_0)\{1+o_p(1)\}+n^{-1}\Bigg(\sum_{i=1}^{n}\Bigg[\widehat{\mathbf{A}}_{wi}\{\boldsymbol{\beta}_0,\widehat{\boldsymbol{\lambda}}(\boldsymbol{\beta}_0,\mathbf{w}_0), \\
&\quad \mathbf{w}_0(t_0)\}\widehat{\mathbf{V}}_{wi}\{\boldsymbol{\beta}_0,\widehat{\boldsymbol{\lambda}}(\boldsymbol{\beta}_0,\mathbf{w}_0),\mathbf{w}_0(t_0)\}^{-1}\mathbf{K}_h(\mathbf{T}_i-\mathbf{t}_0) \\
&\quad \times\frac{\partial}{\partial\mathbf{w}}\widehat{\mathbf{S}}_{wi}\{\boldsymbol{\beta}_0,\widehat{\boldsymbol{\lambda}}(\boldsymbol{\beta}_0,\mathbf{w}_0),\mathbf{w}_0(t_0)\}\Bigg]+o_p(1)\Bigg)\{\widehat{\mathbf{w}}(\boldsymbol{\beta}_0)-\mathbf{w}_0\} \\
&\quad +O_p(n^{-1/2}),
\end{aligned}
$$

where the last equality follows from the result of **Fact 1** and Lemma 5, in combination with the Taylor expansion and large number theorem. On the other hand, because for each t_0,

$$
\begin{aligned}
&E\Bigg[\Bigg\{\frac{1}{n}\sum_{i=1}^{n}\mathbf{A}_{wi}\{\boldsymbol{\beta}_0,m_0,\mathbf{w}_0(t_0)\}\mathbf{V}_{wi}\{\boldsymbol{\beta}_0,m_0,\mathbf{w}_0(t_0)\}^{-1} \\
&\quad \times\mathbf{K}_h(\mathbf{T}_i-t_0)\mathbf{S}_{wi}\{\boldsymbol{\beta}_0,m_0,\mathbf{w}_0(t_0)\}\Bigg\}^{\otimes 2}\Bigg] \\
&= (nh)^{-1}E\Bigg(f_T(t_0)\left[\mathbf{A}_{wi}\{\boldsymbol{\beta}_0,m_0,\mathbf{w}_0(t_0)\}\mathbf{V}_{wi}\{\boldsymbol{\beta}_0,m_0,\mathbf{w}_0(t_0)\}^{-1}\right]
\end{aligned}
$$

$$\times \int \mathbf{K}(\mathbf{s})\mathbf{V}^*_{wi}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(t_0)\}\mathbf{K}(\mathbf{s})d\mathbf{s}\Big[\mathbf{A}_{wi}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(t_0)\}$$
$$\times \mathbf{V}_{wi}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(t_0)\}^{-1}\Big]^{\mathrm{T}}\Big)\{1+O(h^2)\},$$

we obtain

$$\begin{aligned}&\Bigg\|\frac{1}{n}\sum_{i=1}^n \mathbf{A}_{wi}(\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(t_0))\mathbf{V}_{wi}(\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(t_0))^{-1}\\ &\times \mathbf{K}_h(\mathbf{T}_i-\mathbf{t}_0)\mathbf{S}_{wi}(\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(t_0))\Bigg\|_2\\ = \;& O_p\{(nh)^{-1/2}\}.\end{aligned}$$

Thus, the $O_p(n^{-1/2})$ term in equation (S.22) does not contribute to the leading order.

Following the definition given in **Notation in Step 1** in Section 2.1, we have

$$\begin{aligned}&\frac{\partial}{\partial \mathbf{w}}\widehat{\mathbf{S}}_{wik}\{\boldsymbol{\beta}_0, \widehat{\boldsymbol{\lambda}}(\boldsymbol{\beta}_0,\mathbf{w}_0), \mathbf{w}_0(t_0)\}\\ = \;& \Bigg(\Bigg[\mathbf{Q}_{wik}\{\widehat{\boldsymbol{\lambda}}(\boldsymbol{\beta}_0,\mathbf{w}_0), \mathbf{w}_0(t_0)\} + \mathbf{Q}^{\mathrm{T}}_{\lambda ik}\{\mathbf{w}_0(t_0)\}\left\{\frac{\partial\widehat{\boldsymbol{\lambda}}(\boldsymbol{\beta}_0,\mathbf{w}_0)}{\partial \mathbf{w}}\right\}\Bigg]\\ &\times \frac{\partial}{\partial \mathbf{w}}\left[D_{ik} - H_{ik}\{\boldsymbol{\beta}_0, \widehat{\boldsymbol{\lambda}}(\boldsymbol{\beta}_0,\mathbf{w}_0), \mathbf{w}_0(t_0)\}\right]\Bigg)\{1+o_p(1)\}\\ = \;& -\Bigg(\Bigg[\mathbf{Q}_{wik}\{\widehat{\boldsymbol{\lambda}}(\boldsymbol{\beta}_0,\mathbf{w}_0), \mathbf{w}_0(t_0)\} + \mathbf{Q}^{\mathrm{T}}_{\lambda ik}\{\mathbf{w}_0(t_0)\}\left\{\frac{\partial\widehat{\boldsymbol{\lambda}}(\boldsymbol{\beta}_0,\mathbf{w}_0)}{\partial \mathbf{w}}\right\}\Bigg]\\ &\times \Bigg[\mathbf{Q}_{wik}\{\widehat{\boldsymbol{\lambda}}(\boldsymbol{\beta}_0,\mathbf{w}_0), \mathbf{w}_0(t_0)\} + \mathbf{Q}^{\mathrm{T}}_{\lambda ik}\{\mathbf{w}_0(t_0)\}\left\{\frac{\partial\widehat{\boldsymbol{\lambda}}(\boldsymbol{\beta}_0,\mathbf{w}_0)}{\partial \mathbf{w}}\right\}\Bigg]\\ &\times \Theta_{ik}\{\boldsymbol{\beta}_0, \widehat{\boldsymbol{\lambda}}(\boldsymbol{\beta}_0,\mathbf{w}_0), \mathbf{w}_0(t_0)\}\{1+o_p(1)\}\Bigg).\end{aligned}$$

Also note that plugging $\mathbf{w}_h = \widehat{\mathbf{w}}(\boldsymbol{\beta}_0) - \mathbf{w}_0$ into the results in Lemma 4, we have

(S.23)
$$\begin{aligned}&\frac{\partial\widehat{\boldsymbol{\lambda}}(\boldsymbol{\beta}_0,\mathbf{w}_0)}{\partial \mathbf{w}}\{\widehat{\mathbf{w}}(\boldsymbol{\beta}_0) - \mathbf{w}_0\}\\ = \;& -\Bigg(n^{-1}\sum_{i=1}^n \Bigg[\widetilde{\mathbf{Q}}_{\lambda i}\{\mathbf{w}_0(\mathbf{T}_i)\}^{\mathrm{T}}\boldsymbol{\Theta}_i\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_i)\}\boldsymbol{\Omega}_i^{-1}\boldsymbol{\Theta}_i\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_i)\}\end{aligned}$$

$$\times\widetilde{\mathbf{Q}}_{\lambda i}\{\mathbf{w}_0(\mathbf{T}_i)\}\Big]\Big)^{-1} n^{-1}\Big[\sum_{i=1}^{n}\widetilde{\mathbf{Q}}_{\lambda i}\{\mathbf{w}_0(\mathbf{T}_i)\}^{\mathrm{T}}\boldsymbol{\Theta}_i\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_i)\}\boldsymbol{\Omega}_i^{-1}$$
$$\times\boldsymbol{\Theta}_i\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_i)\}\widetilde{\mathbf{Q}}_{wi}\{m_0, \{\widehat{\mathbf{w}}(\boldsymbol{\beta}_0, \mathbf{T}_i) - \mathbf{w}_0(\mathbf{T}_i)\}\}\Big]\{1 + o_p(1)\}$$

which is $o_p(1)$ under the assumption that $\widehat{\mathbf{w}}(\boldsymbol{\beta}_0)$ is a consistent estimator of $\mathbf{w}_0$, and $\mathbf{Q}_{wik}\{\widehat{\boldsymbol{\lambda}}(\boldsymbol{\beta}_0, \mathbf{w}), \mathbf{w}(t_0)\}\{\widehat{\mathbf{w}}(\boldsymbol{\beta}_0) - \mathbf{w}_0\} = \mathbf{Q}_{wik}\{\widehat{\boldsymbol{\lambda}}(\boldsymbol{\beta}_0, \mathbf{w}), \mathbf{w}(t_0)\}^{\mathrm{T}}$ $\{\widehat{\mathbf{w}}(\boldsymbol{\beta}_0, t_0) - \mathbf{w}_0(t_0)\}$ as described in **Notations in Step 2**.

Thus, we obtain

$$\begin{aligned}
& n^{-1}\sum_{i=1}^{n}\widehat{\mathbf{A}}_{wi}\{\boldsymbol{\beta}_0, \boldsymbol{\lambda}_0, \mathbf{w}_0(t_0)\}\widehat{\mathbf{V}}_{wi}\{\boldsymbol{\beta}_0, \boldsymbol{\lambda}_0, \mathbf{w}_0(t_0)\}^{-1}\mathbf{K}_h(\mathbf{T}_i - \mathbf{t}_0) \\
& \times\frac{\partial}{\partial\mathbf{w}}\widehat{\mathbf{S}}_{wi}\{\boldsymbol{\beta}_0, \widehat{\boldsymbol{\lambda}}(\boldsymbol{\beta}_0, \mathbf{w}_0), \mathbf{w}_0(t_0)\}\{\widehat{\mathbf{w}}(\boldsymbol{\beta}_0) - \mathbf{w}_0\} \\
= \; & n^{-1}\sum_{i=1}^{n}\Big(\widehat{\mathbf{A}}_{wi}\{\boldsymbol{\beta}_0, \boldsymbol{\lambda}_0, \mathbf{w}_0(t_0)\}\widehat{\mathbf{V}}_{wi}\{\boldsymbol{\beta}_0, \boldsymbol{\lambda}_0, \mathbf{w}_0(t_0)\}^{-1}\mathbf{K}_h(\mathbf{T}_i - \mathbf{t}_0) \\
& \times\Big[\mathbf{Q}_{wi}\{\boldsymbol{\lambda}_0, \mathbf{w}_0(t_0)\} - \eta\{\mathbf{U}_i(t_0)\}\Big]\boldsymbol{\Theta}_i^*\{\boldsymbol{\beta}_0, \widehat{\boldsymbol{\lambda}}(\boldsymbol{\beta}_0, \mathbf{w}_0), \mathbf{w}_0(t_0)\} \\
& \times[\mathbf{Q}_{wi}^*\{\boldsymbol{\lambda}_0, \mathbf{w}_0(t_0)\}\{\widehat{\mathbf{w}}(\boldsymbol{\beta}_0, t_0) - \mathbf{w}_0(t_0)\}]\Big)\{1 + o_p(1)\} \\
= \; & E\Big(\widehat{\mathbf{A}}_{wi}\{\boldsymbol{\beta}_0, \boldsymbol{\lambda}_0, \mathbf{w}_0(t_0)\}\widehat{\mathbf{V}}_{wi}\{\boldsymbol{\beta}_0, \boldsymbol{\lambda}_0, \mathbf{w}_0(t_0)\}^{-1}\mathbf{K}_h(\mathbf{T}_i - \mathbf{t}_0) \\
& \times\Big[\mathbf{Q}_{wi}\{\boldsymbol{\lambda}_0, \mathbf{w}_0(t_0)\} - \eta\{\mathbf{U}_i(t_0)\}\Big]\boldsymbol{\Theta}_i^*\{\boldsymbol{\beta}_0, \widehat{\boldsymbol{\lambda}}(\boldsymbol{\beta}_0, \mathbf{w}_0), \mathbf{w}_0(t_0)\} \\
& \times\mathbf{Q}_{wi}^*\{\boldsymbol{\lambda}_0, \mathbf{w}_0(t_0)\}\Big)\{1 + o_p(1)\}\{\widehat{\mathbf{w}}(\boldsymbol{\beta}_0, t_0) - \mathbf{w}_0(t_0)\} \\
= \; & \int E\Big(\widehat{\mathbf{A}}_{wi}\{\boldsymbol{\beta}_0, \boldsymbol{\lambda}_0, \mathbf{w}_0(t_0)\}\widehat{\mathbf{V}}_{wi}\{\boldsymbol{\beta}_0, \boldsymbol{\lambda}_0, \mathbf{w}_0(t_0)\}^{-1}\mathbf{K}_h(\mathbf{T}_i - \mathbf{t}_0) \\
& \times\Big[\mathbf{Q}_{wi}\{\boldsymbol{\lambda}_0, \mathbf{w}_0(t_0)\} - \eta\{\mathbf{U}_i(t_0)\}\Big]\boldsymbol{\Theta}_i^*\{\boldsymbol{\beta}_0, \widehat{\boldsymbol{\lambda}}(\boldsymbol{\beta}_0, \mathbf{w}_0), \mathbf{w}_0(t_0)\} \\
& \times\mathbf{Q}_{wi}^*\{\boldsymbol{\lambda}_0, \mathbf{w}_0(t_0)\}\Big)f_{\mathbf{T}}(\mathbf{T}_i)d\mathbf{T}_i\{1 + o_p(1)\}\{\widehat{\mathbf{w}}(\boldsymbol{\beta}_0, t_0) - \mathbf{w}_0(t_0)\}
\end{aligned}$$

which suggests

(S.24)

$$n^{-1}\sum_{i=1}^{n}\widehat{\mathbf{A}}_{wi}\{\boldsymbol{\beta}_0, \boldsymbol{\lambda}_0, \mathbf{w}_0(t_0)\}\widehat{\mathbf{V}}_{wi}\{\boldsymbol{\beta}_0, \boldsymbol{\lambda}_0, \mathbf{w}_0(t_0)\}^{-1}\mathbf{K}_h(\mathbf{T}_i - \mathbf{t}_0)$$

$$
\begin{aligned}
& \times \frac{\partial}{\partial \mathbf{w}} \widehat{\mathbf{S}}_{wi}\{\boldsymbol{\beta}_0, \widehat{\boldsymbol{\lambda}}(\boldsymbol{\beta}_0, \mathbf{w}_0), \mathbf{w}_0(t_0)\}\{\widehat{\mathbf{w}}(\boldsymbol{\beta}_0) - \mathbf{w}_0\} \\
= & \int E\bigg(\widehat{\mathbf{A}}_{wi}\{\boldsymbol{\beta}_0, \boldsymbol{\lambda}_0, \mathbf{w}_0(t_0)\} \widehat{\mathbf{V}}_{wi}\{\boldsymbol{\beta}_0, \boldsymbol{\lambda}_0, \mathbf{w}_0(t_0)\}^{-1} \mathbf{K}_h(\mathbf{T}_i - \mathbf{t}_0) \\
& \times \bigg[\mathbf{Q}_{wi}\{\boldsymbol{\lambda}_0, \mathbf{w}_0(t_0)\} - \eta\{\mathbf{U}_i(t_0)\} \bigg] \boldsymbol{\Theta}_i^*\{\boldsymbol{\beta}_0, \widehat{\boldsymbol{\lambda}}(\boldsymbol{\beta}_0, \mathbf{w}_0), \mathbf{w}_0(t_0)\} \\
& \times \mathbf{Q}_{wi}^*\{\boldsymbol{\lambda}_0, \mathbf{w}_0(t_0)\} \mid t_0 + h\mathbf{s} \bigg) f_{\mathbf{T}}(t_0 + h\mathbf{s}) d\mathbf{s}\{1 + o_p(1)\} \\
& \times \{\widehat{\mathbf{w}}(\boldsymbol{\beta}_0, t_0) - \mathbf{w}_0(t_0)\} \\
= & f_T(t_0) E\bigg(\widehat{\mathbf{A}}_{wi}\{\boldsymbol{\beta}_0, \boldsymbol{\lambda}_0, \mathbf{w}_0(t_0)\} \widehat{\mathbf{V}}_{wi}\{\boldsymbol{\beta}_0, \boldsymbol{\lambda}_0, \mathbf{w}_0(t_0)\}^{-1} \bigg[\mathbf{Q}_{wi}\{\boldsymbol{\lambda}_0, \mathbf{w}_0(t_0)\} \\
& - \eta\{\mathbf{U}_i(t_0)\} \bigg] \boldsymbol{\Theta}_i^*\{\boldsymbol{\beta}_0, \widehat{\boldsymbol{\lambda}}(\boldsymbol{\beta}_0, \mathbf{w}_0), \mathbf{w}_0(t_0)\} \mathbf{Q}_{wi}^*\{\boldsymbol{\lambda}_0, \mathbf{w}_0(t_0)\} \bigg) \\
& \times \{\widehat{\mathbf{w}}(\boldsymbol{\beta}_0, t_0) - \mathbf{w}_0(t_0)\}\{1 + o_p(1) + O(h^2)\} \\
= & f_T(t_0) \mathbf{B}(t_0)\{1 + o(1)\}\{\widehat{\mathbf{w}}(\boldsymbol{\beta}_0, t_0) - \mathbf{w}_0(t_0)\},
\end{aligned}
$$

where $\mathbf{B}(t_0), \eta\{\mathbf{U}_i(t_0)\}, \mathbf{Q}_{wi}, \mathbf{Q}_{wi}^*, \boldsymbol{\Theta}_i^*$ are defined in the section **Notation in Step 3**. Plug in the result to (S.22), we obtain

$$
\begin{aligned}
& \widehat{\mathbf{w}}(\boldsymbol{\beta}_0, t_0) - \mathbf{w}_0(t_0) \\
= & -\{\mathbf{B}(t_0) f_T(t_0) + o_p(1)\}^{-1} \left[\frac{1}{n} \sum_{i=1}^n \mathbf{A}_{wi}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(t_0)\} \right. \\
& \left. \mathbf{V}_{wi}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(t_0)\}^{-1} \mathbf{K}_h(\mathbf{T}_i - t_0) \mathbf{S}_{wi}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(t_0)\} \right].
\end{aligned}
\tag{S.25}
$$

Consequently,

$$
\begin{aligned}
& \operatorname{var}\{\widehat{\mathbf{w}}(\boldsymbol{\beta}_0, t_0) - \mathbf{w}_0(t_0)\} \\
= & (nh)^{-1}\{\mathbf{B}(t_0) f_T(t_0)\}^{-1} E\bigg(f_T(t_0) \bigg[\mathbf{A}_{wi}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(t_0)\} \\
& \times \mathbf{V}_{wi}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(t_0)\}^{-1} \bigg] \int \mathbf{K}(\mathbf{s}) \mathbf{V}_{wi}^*\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(t_0)\} \mathbf{K}(\mathbf{s}) d\mathbf{s} \\
& \times \left[\mathbf{A}_{wi}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(t_0)\} \mathbf{V}_{wi}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(t_0)\}^{-1} \right]^{\mathrm{T}} \bigg) \\
& \times \{\mathbf{B}(t_0) f_T(t_0)\}^{-1} + o\{(nh)^{-1}\}.
\end{aligned}
$$

This yields the desired result. □

APPENDIX S.3: PROOF OF THEOREM ??

We define the following additional notations for proving Theorem 2. We define

$$
\mathbf{S}_{\beta ik}^{\dagger}[\boldsymbol{\beta}, \widehat{\boldsymbol{\lambda}}(\boldsymbol{\beta}_0, \mathbf{w}_0), \widehat{\mathbf{w}}(\boldsymbol{\beta}, T_{ik})]
$$

$$
\begin{aligned}
= & \Bigg\{\mathbf{Q}_{\beta ik} + \left(\frac{\partial\widehat{\boldsymbol{\lambda}}(\boldsymbol{\beta}_0,\mathbf{w}_0)}{\partial\boldsymbol{\beta}^{\mathrm{T}}} - \mathbf{Q}_{\lambda ik}\{\mathbf{w}_0(T_{ik})\}^{\mathrm{T}}\frac{\partial\widehat{\boldsymbol{\lambda}}(\boldsymbol{\beta}_0,\mathbf{w}_0)}{\partial\mathbf{w}}\mathcal{B}\right)^{\mathrm{T}} \\
& -\Bigg(\mathbf{B}(T_{ik})^{-1}E\Bigg[\mathbf{A}_{wj}\{\boldsymbol{\beta}_0,m_0,\mathbf{w}_0(T_{ik})\}\mathbf{V}_{wj}\{\boldsymbol{\beta}_0,m_0,\mathbf{w}_0(T_{ik})\}^{-1} \\
& \frac{\partial\mathbf{S}_{wj}\{\boldsymbol{\beta}_0,m_0,\mathbf{w}_0(T_{ik})\}}{\partial\boldsymbol{\beta}^{\mathrm{T}}} \mid \mathbf{O}_i\Bigg]\Bigg)^{\mathrm{T}}\mathbf{Z}_{ik}m_0'\{\mathbf{w}_0(T_{ik})^{\mathrm{T}}\mathbf{Z}_{ik}\}\Bigg\} \\
& \times\left[D_{ik} - H_{ik}\{\boldsymbol{\beta},\widehat{\boldsymbol{\lambda}}(\boldsymbol{\beta}_0,\mathbf{w}_0),\widehat{\mathbf{w}}(\boldsymbol{\beta},T_{ik})\}\right],
\end{aligned}
$$

where

$$
\begin{aligned}
\mathcal{B} = & \Bigg\{\Bigg(\mathbf{B}(T_{ik})^{-1}E\Bigg[\mathbf{A}_{wj}\{\boldsymbol{\beta}_0,m_0,\mathbf{w}_0(T_{ik})\}\mathbf{V}_{wj}\{\boldsymbol{\beta}_0,m_0,\mathbf{w}_0(T_{ik})\}^{-1} \\
& \times\frac{\partial\mathbf{S}_{wj}\{\boldsymbol{\beta}_0,m_0,\mathbf{w}_0(T_{ik})\}}{\partial\boldsymbol{\beta}^{\mathrm{T}}}\Bigg)\Bigg|\mathbf{O}_i\Bigg]\Bigg)^{\mathrm{T}}, k=1,\ldots,M_i, i=1,\ldots,n\Bigg\}^{\mathrm{T}}
\end{aligned}
$$

Let $\mathbf{A}^{\dagger}_{\beta i}[\boldsymbol{\beta}_0,\widehat{\boldsymbol{\lambda}}\{\boldsymbol{\beta}_0,\widehat{\mathbf{w}}(\boldsymbol{\beta}_0)\},\widehat{\mathbf{w}}(\boldsymbol{\beta}_0,\mathbf{T}_i)]$ be a $d_\beta\times d_\beta M_i$ matrix, with the kth $d_\beta\times d_\beta$ column block being

$$
\begin{aligned}
& \mathbf{A}^{\dagger}_{\beta ik}[\boldsymbol{\beta},\widehat{\boldsymbol{\lambda}}(\boldsymbol{\beta}_0,\mathbf{w}_0),\widehat{\mathbf{w}}(\boldsymbol{\beta},T_{ik})] \\
= & \Bigg\{\mathbf{Q}_{\beta ik} + \left(\frac{\partial\widehat{\boldsymbol{\lambda}}(\boldsymbol{\beta}_0,\mathbf{w}_0)}{\partial\boldsymbol{\beta}^{\mathrm{T}}} - \mathbf{Q}_{\lambda ik}\{\mathbf{w}_0(T_{ik})\}^{\mathrm{T}}\frac{\partial\widehat{\boldsymbol{\lambda}}(\boldsymbol{\beta}_0,\mathbf{w}_0)}{\partial\mathbf{w}}\mathcal{B}\right)^{\mathrm{T}} \\
& -\Bigg(\mathbf{B}(T_{ik})^{-1}E\Bigg[\mathbf{A}_{wj}\{\boldsymbol{\beta}_0,m_0,\mathbf{w}_0(T_{ik})\}\mathbf{V}_{wj}\{\boldsymbol{\beta}_0,m_0,\mathbf{w}_0(T_{ik})\}^{-1} \\
& \frac{\partial\mathbf{S}_{wj}\{\boldsymbol{\beta}_0,m_0,\mathbf{w}_0(T_{ik})\}}{\partial\boldsymbol{\beta}^{\mathrm{T}}} \mid \mathbf{O}_i\Bigg]\Bigg)^{\mathrm{T}}\mathbf{Z}_{ik}m_0'\{\mathbf{w}_0(T_{ik})^{\mathrm{T}}\mathbf{Z}_{ik}\}\Bigg\}^{\otimes 2} \\
& \times\boldsymbol{\Theta}_i\{\boldsymbol{\beta},\widehat{\boldsymbol{\lambda}}(\boldsymbol{\beta}_0,\mathbf{w}_0),\widehat{\mathbf{w}}(\boldsymbol{\beta},T_{ik})\}.
\end{aligned}
$$

Let $\mathbf{V}^{\dagger}_{\beta i}\{\boldsymbol{\beta}_0,\widehat{\boldsymbol{\lambda}}(\boldsymbol{\beta}_0,\mathbf{w}_0),\mathbf{w}_0(\mathbf{T}_i)\}$ be a $d_\beta M_i\times d_\beta M_i$ matrix with the (p,q)th block being

$$
\begin{aligned}
& \mathbf{V}^{\dagger}_{\beta ipq}[\boldsymbol{\beta},\widehat{\boldsymbol{\lambda}}(\boldsymbol{\beta}_0,\mathbf{w}_0),\widehat{\mathbf{w}}(\boldsymbol{\beta},T_{ip})] \\
= & \Bigg\{\mathbf{Q}_{\beta ip} + \left(\frac{\partial\widehat{\boldsymbol{\lambda}}(\boldsymbol{\beta}_0,\mathbf{w}_0)}{\partial\boldsymbol{\beta}^{\mathrm{T}}} - \mathbf{Q}_{\lambda ip}\{\mathbf{w}_0(T_{ip})\}^{\mathrm{T}}\frac{\partial\widehat{\boldsymbol{\lambda}}(\boldsymbol{\beta}_0,\mathbf{w}_0)}{\partial\mathbf{w}}\mathcal{B}\right)^{\mathrm{T}} \\
& -\Bigg(\mathbf{B}(T_{ip})^{-1}E\Bigg[\mathbf{A}_{wj}\{\boldsymbol{\beta}_0,m_0,\mathbf{w}_0(T_{ip})\}\mathbf{V}_{wj}\{\boldsymbol{\beta}_0,m_0,\mathbf{w}_0(T_{ip})\}^{-1} \\
& \frac{\partial\mathbf{S}_{wj}\{\boldsymbol{\beta}_0,m_0,\mathbf{w}_0(T_{ip})\}}{\partial\boldsymbol{\beta}^{\mathrm{T}}} \mid \mathbf{O}_i\Bigg]\Bigg)^{\mathrm{T}}\mathbf{Z}_{ip}m_0'\{\mathbf{w}_0(T_{ip})^{\mathrm{T}}\mathbf{Z}_{ip}\}\Bigg\}
\end{aligned}
$$

$$\times\Bigg\{\mathbf{Q}_{\beta iq}+\left(\frac{\partial\widehat{\boldsymbol{\lambda}}(\boldsymbol{\beta}_0,\mathbf{w}_0)}{\partial\boldsymbol{\beta}^{\mathrm{T}}}-\mathbf{Q}_{\lambda iq}\{\mathbf{w}_0(T_{iq})\}^{\mathrm{T}}\frac{\partial\widehat{\boldsymbol{\lambda}}(\boldsymbol{\beta},\mathbf{w}_0)}{\partial\mathbf{w}}\mathcal{B}\right)^{\mathrm{T}}$$
$$-\Bigg(\mathbf{B}(T_{iq})^{-1}E\Bigg[\mathbf{A}_{wj}\{\boldsymbol{\beta}_0,m_0,\mathbf{w}_0(T_{iq})\}\mathbf{V}_{wj}\{\boldsymbol{\beta}_0,m_0,\mathbf{w}_0(T_{iq})\}^{-1}$$
$$\frac{\partial\mathbf{S}_{wj}\{\boldsymbol{\beta}_0,m_0,\mathbf{w}_0(T_{iq})\}}{\partial\boldsymbol{\beta}^{\mathrm{T}}}\mid\mathbf{O}_i\Bigg]\Bigg)^{\mathrm{T}}\mathbf{Z}_{iq}m_0'\{\mathbf{w}_0(T_{iq})^{\mathrm{T}}\mathbf{Z}_{iq}\}\Bigg\}^{\mathrm{T}}\boldsymbol{\Omega}_{ipq}.$$

Note that we define $\mathbf{S}^{\dagger}_{\beta i}\{\boldsymbol{\beta},m_0,\mathbf{w}_0(\mathbf{T}_i)\}$, $\mathbf{A}^{\dagger}_{\beta i}\{\boldsymbol{\beta},m_0,\mathbf{w}_0(\mathbf{T}_i)\}$, $\mathbf{V}^{\dagger}_{\beta i}\{\boldsymbol{\beta},m_0,\mathbf{w}_0(\mathbf{T}_i)\}$ by replacing the B-spline $\mathbf{B}_r^{\mathrm{T}}\{\mathbf{w}_0(T_{ik})^{\mathrm{T}}\mathbf{Z}_{ik}\}\boldsymbol{\lambda}_0$ in $\mathbf{S}^{\dagger}_{\beta i}\{\boldsymbol{\beta},\boldsymbol{\lambda}_0,\mathbf{w}_0(\mathbf{T}_i)\}$, $\mathbf{A}^{\dagger}_{\beta i}\{\boldsymbol{\beta},\boldsymbol{\lambda}_0,\mathbf{w}_0(\mathbf{T}_i)\}$, $\mathbf{V}^{\dagger}_{\beta i}(\boldsymbol{\beta},\boldsymbol{\lambda}_0,\mathbf{w}_0(\mathbf{T}_i)\}$ with the functions

$$m\{\mathbf{U}_i(\mathbf{T}_i)\}\equiv[m\{\mathbf{w}_0(T_{ik})^{\mathrm{T}}\mathbf{Z}_{ik}\},k=1,\ldots,M_i].$$

Lemma 6. *Assume Conditions (A1)-(A5) in Section 2.3 hold. Consider* $\mathbf{w}$ *as a function of* $\boldsymbol{\beta}$*. Let* $\widehat{\mathbf{w}}(\boldsymbol{\beta})$ *and* $\mathbf{w}(\boldsymbol{\beta})$ *be as defined at the beginning of Section 2.1. Then*

(S.26)
$$\Bigg\|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\widehat{\mathbf{A}}_{\beta i}[\boldsymbol{\beta},\widehat{\boldsymbol{\lambda}}\{\boldsymbol{\beta},\widehat{\mathbf{w}}(\boldsymbol{\beta})\},\widehat{\mathbf{w}}(\boldsymbol{\beta},\mathbf{T}_i)]\widehat{\mathbf{V}}_{\beta i}[\boldsymbol{\beta},\widehat{\boldsymbol{\lambda}}\{\boldsymbol{\beta},\widehat{\mathbf{w}}(\boldsymbol{\beta})\},\widehat{\mathbf{w}}(\boldsymbol{\beta},\mathbf{T}_i)]^{-1}$$
$$\widehat{\mathbf{S}}_{\beta i}[\boldsymbol{\beta},\widehat{\boldsymbol{\lambda}}\{\boldsymbol{\beta},\widehat{\mathbf{w}}(\boldsymbol{\beta})\},\widehat{\mathbf{w}}(\boldsymbol{\beta},\mathbf{T}_i)]-\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\mathbf{A}^{\dagger}_{\beta i}[\boldsymbol{\beta},\widehat{\boldsymbol{\lambda}}\{\boldsymbol{\beta},\widehat{\mathbf{w}}(\boldsymbol{\beta})\},\widehat{\mathbf{w}}(\boldsymbol{\beta},\mathbf{T}_i)]$$
$$\mathbf{V}^{\dagger}_{\beta i}[\boldsymbol{\beta},\widehat{\boldsymbol{\lambda}}\{\boldsymbol{\beta},\widehat{\mathbf{w}}(\boldsymbol{\beta})\},\widehat{\mathbf{w}}(\boldsymbol{\beta},\mathbf{T}_i)]^{-1}\mathbf{S}^{\dagger}_{\beta i}[\boldsymbol{\beta},\widehat{\boldsymbol{\lambda}}\{\boldsymbol{\beta},\widehat{\mathbf{w}}(\boldsymbol{\beta})\},\widehat{\mathbf{w}}(\boldsymbol{\beta},\mathbf{T}_i)]\Bigg\|_{\infty}$$
$$=o_p(1).$$

Proof: We first note that following the definitions in the **Notation in Step 3** of Section 2.1, the summand of the first summation in (S.26) can be viewed as a function evaluated at

$$\mathcal{T}_1\equiv\frac{\partial\widehat{\mathbf{w}}(\boldsymbol{\beta},T_{ik})}{\partial\boldsymbol{\beta}^{\mathrm{T}}},$$

while the summand of the second summation can be viewed as the same function evaluated at a slightly different quantity

$$\mathcal{T}_2\equiv\mathbf{B}(\mathbf{T}_i)^{-1}E\Bigg[\mathbf{A}_{wj}\{\boldsymbol{\beta}_0,m_0,\mathbf{w}_0(\mathbf{T}_i)\}\mathbf{V}_{wj}\{\boldsymbol{\beta}_0,m_0,\mathbf{w}_0(\mathbf{T}_i)\}^{-1}$$
$$\times\frac{\partial\mathbf{S}_{wj}\{\boldsymbol{\beta}_0,m_0,\mathbf{w}_0(\mathbf{T}_i)\}}{\partial\boldsymbol{\beta}^{\mathrm{T}}}\mid\mathbf{O}_i\Bigg].$$

Therefore, we use a Taylor expansion to write (S.26) into the form

$$\frac{-1}{\sqrt{n}}\sum_{i=1}^{n}\mathbf{C}(\mathbf{O_i})(\mathcal{T}_1-\mathcal{T}_2)\{1+o_p(1)\}$$

where $\mathbf{C}(\mathbf{O_i})$ is the derivative of the corresponding summand evaluated at $\mathcal{T}_2$. Note that $\mathbf{C}(\mathbf{O_i})$ is a zero mean random matrix depending only on the ith observation, because it contains the term $\mathbf{D}_i-\mathbf{H}_i$ as a multiplier, which has mean zero conditioning on the covariates. Further, taking derivative of (S.25) with respect to $\boldsymbol{\beta}$, we have

(S.27)

$$\begin{aligned}
&\frac{\partial\widehat{\mathbf{w}}(\boldsymbol{\beta}_0,\mathbf{T}_i)}{\partial\boldsymbol{\beta}^{\mathrm{T}}}\\
=\;&\{\mathbf{B}(\mathbf{T}_i)\mathbf{f_T}(\mathbf{T}_i)\}^{-1}\Bigg[\frac{1}{n}\sum_{j=1}^{n}\mathbf{A}_{wj}\{\boldsymbol{\beta}_0,m_0,\mathbf{w}_0(\mathbf{T}_i)\}\mathbf{V}_{wj}\{\boldsymbol{\beta}_0,m_0,\mathbf{w}_0(\mathbf{T}_i)\}^{-1}\\
&\mathbf{K}_h(\mathbf{T}_i-\mathbf{T}_j)\frac{\partial\mathbf{S}_{wj}\{\boldsymbol{\beta}_0,m_0,\mathbf{w}_0(\mathbf{T}_i)\}}{\partial\boldsymbol{\beta}^{\mathrm{T}}}\\
&+\frac{1}{n}\sum_{j=1}^{n}\frac{\partial\mathbf{A}_{wj}\{\boldsymbol{\beta}_0,m_0,\mathbf{w}_0(\mathbf{T}_i)\}\mathbf{V}_{wj}\{\boldsymbol{\beta}_0,m_0,\mathbf{w}_0(\mathbf{T}_i)\}^{-1}}{\partial\boldsymbol{\beta}^{\mathrm{T}}}\\
&\mathbf{K}_h(\mathbf{T}_i-\mathbf{T}_j)\mathbf{S}_{wj}\{\boldsymbol{\beta}_0,m_0,\mathbf{w}_0(\mathbf{T}_i)\}\Bigg]\{1+o_p(1)\}\\
=\;&\{\mathbf{B}(\mathbf{T}_i)\mathbf{f_T}(\mathbf{T}_i)\}^{-1}\Bigg[\frac{1}{n}\sum_{j=1}^{n}\mathbf{A}_{wj}\{\boldsymbol{\beta}_0,m_0,\mathbf{w}_0(\mathbf{T}_i)\}\mathbf{V}_{wj}\{\boldsymbol{\beta}_0,m_0,\mathbf{w}_0(\mathbf{T}_i)\}^{-1}\\
&\mathbf{K}_h(\mathbf{T}_i-\mathbf{T}_j)\frac{\partial\mathbf{S}_{wj}\{\boldsymbol{\beta}_0,m_0,\mathbf{w}_0(\mathbf{T}_i)\}}{\partial\boldsymbol{\beta}^{\mathrm{T}}}\Bigg]\{1+o_p(1)\}.
\end{aligned}$$

The last equality holds because $\mathbf{S}_{wj}\{\boldsymbol{\beta}_0,m_0,\mathbf{w}_0(\mathbf{T}_i)\}$ has conditional mean 0, and therefore the second summation has smaller order than the first summation. Therefore, we have

(S.28)

$$\begin{aligned}
&\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\widehat{\mathbf{A}}_{\beta i}[\boldsymbol{\beta},\widehat{\boldsymbol{\lambda}}\{\boldsymbol{\beta},\widehat{\mathbf{w}}(\boldsymbol{\beta})\},\widehat{\mathbf{w}}(\boldsymbol{\beta},\mathbf{T}_i)]\widehat{\mathbf{V}}_{\beta i}[\boldsymbol{\beta},\widehat{\boldsymbol{\lambda}}\{\boldsymbol{\beta},\widehat{\mathbf{w}}(\boldsymbol{\beta})\},\widehat{\mathbf{w}}(\boldsymbol{\beta},\mathbf{T}_i)]^{-1}\\
&\widehat{\mathbf{S}}_{\beta i}[\boldsymbol{\beta},\widehat{\boldsymbol{\lambda}}\{\boldsymbol{\beta},\widehat{\mathbf{w}}(\boldsymbol{\beta})\},\widehat{\mathbf{w}}(\boldsymbol{\beta},\mathbf{T}_i)]-\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\mathbf{A}^{\dagger}_{\beta i}[\boldsymbol{\beta},\widehat{\boldsymbol{\lambda}}\{\boldsymbol{\beta},\widehat{\mathbf{w}}(\boldsymbol{\beta})\},\widehat{\mathbf{w}}(\boldsymbol{\beta},\mathbf{T}_i)]\\
&\mathbf{V}^{\dagger}_{\beta i}[\boldsymbol{\beta},\widehat{\boldsymbol{\lambda}}\{\boldsymbol{\beta},\widehat{\mathbf{w}}(\boldsymbol{\beta})\},\widehat{\mathbf{w}}(\boldsymbol{\beta},\mathbf{T}_i)]^{-1}\mathbf{S}^{\dagger}_{\beta i}[\boldsymbol{\beta},\widehat{\boldsymbol{\lambda}}\{\boldsymbol{\beta},\widehat{\mathbf{w}}(\boldsymbol{\beta})\},\widehat{\mathbf{w}}(\boldsymbol{\beta},\mathbf{T}_i)]
\end{aligned}$$

$$
\begin{aligned}
&= \frac{-1}{\sqrt{n}}\sum_{i=1}^{n}\mathbf{C}(\mathbf{O_i})\Bigg(\{\mathbf{B}(\mathbf{T}_i)\mathbf{f_T}(\mathbf{T}_i)\}^{-1}\Bigg[\frac{1}{n}\sum_{j=1}^{n}\mathbf{A}_{wj}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_i)\} \\
&\quad \times\mathbf{V}_{wj}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_i)\}^{-1}\mathbf{K}_h(\mathbf{T}_i-\mathbf{T}_j)\frac{\partial\mathbf{S}_{wj}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_i)\}}{\partial\boldsymbol{\beta}^{\mathrm{T}}}\Bigg] \\
&\quad \times\{1+o_p(1)\} \\
&\quad -\mathbf{B}(\mathbf{T}_i)^{-1}E\Bigg[\mathbf{A}_{wj}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_i)\}\mathbf{V}_{wj}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_i)\}^{-1} \\
&\quad \times\frac{\partial\mathbf{S}_{wj}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_i)\}}{\partial\boldsymbol{\beta}^{\mathrm{T}}}\mid\mathbf{O}_i\Bigg]\Bigg)\{1+o_p(1)\} \\
&= \frac{-1}{n^{3/2}}\sum_{i=1}^{n}\sum_{j=1}^{n}\mathbf{C}(\mathbf{O_i})\Bigg(\{\mathbf{B}(\mathbf{T}_i)\mathbf{f_T}(\mathbf{T}_i)\}^{-1}\Bigg[\mathbf{A}_{wj}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_i)\} \\
&\quad \times\mathbf{V}_{wj}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_i)\}^{-1}\mathbf{K}_h(\mathbf{T}_i-\mathbf{T}_j)\frac{\partial\mathbf{S}_{wj}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_i)\}}{\partial\boldsymbol{\beta}^{\mathrm{T}}}\Bigg] \\
&\quad -\mathbf{B}(\mathbf{T}_i)^{-1}E\Bigg[\mathbf{A}_{wj}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_i)\}\mathbf{V}_{wj}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_i)\}^{-1} \\
&\quad \frac{\partial\mathbf{S}_{wj}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_i)\}}{\partial\boldsymbol{\beta}^{\mathrm{T}}}\mid\mathbf{O}_i\Bigg]\Bigg)\{1+o_p(1)\}.
\end{aligned}
$$

Define

$$
\begin{aligned}
\mathbf{u}(\mathbf{O}_i,\mathbf{O}_j) &= \Big[\mathbf{C}(\mathbf{O_i})\{\mathbf{B}(\mathbf{T}_i)\mathbf{f_T}(\mathbf{T}_i)\}^{-1}\Big]\Big[\mathbf{A}_{wj}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_i)\} \\
&\quad \times\mathbf{V}_{wj}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_i)\}^{-1}\mathbf{K}_h(\mathbf{T}_i-\mathbf{T}_j) \\
&\quad \times\frac{\partial\mathbf{S}_{wij}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_i)\}}{\partial\boldsymbol{\beta}^{\mathrm{T}}}\Big].
\end{aligned}
$$

Then the first summation in (S.28) is

$$
\begin{aligned}
&\frac{-1}{n^{3/2}}\sum_{i=1}^{n}\mathbf{u}(\mathbf{O}_i,\mathbf{O}_j) \\
&= \frac{-1}{\sqrt{n}}\sum_{i=1}^{n}E\{\mathbf{u}(\mathbf{O}_i,\mathbf{O}_j)\mid\mathbf{O}_i\}-\frac{1}{\sqrt{n}}\sum_{j=1}^{n}E\{\mathbf{u}(\mathbf{O}_i,\mathbf{O}_j)\mid\mathbf{O}_j\} \\
&\quad +\sqrt{n}E\{\mathbf{u}(\mathbf{O}_i,\mathbf{O}_j)\}+o_p(1).
\end{aligned}
$$

Now we investigate each expectation above. It is easy to see that $E\{\mathbf{u}(\mathbf{O}_i,\mathbf{O}_j)|\mathbf{O}_j\} = \mathbf{0}, a.s.$ and $E\{\mathbf{u}(\mathbf{O}_i,\mathbf{O}_j)\} = \mathbf{0}$, because of the term $\mathbf{D}_i-\mathbf{H}_i$ in $\mathbf{C}(\mathbf{O_i})$. Further

$$E\{\mathbf{u}(\mathbf{O}_i,\mathbf{O}_j)|\mathbf{O}_i\}$$

$$
\begin{aligned}
&= E\Bigg(\left[\mathbf{C}(\mathbf{O_i})\{\mathbf{B}(\mathbf{T}_i)\mathbf{f_T}(\mathbf{T}_i)\}^{-1}\right] \mathbf{A}_{wj}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_i)\} \\
&\quad \times \mathbf{V}_{wj}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_i)\}^{-1}\mathbf{K}_h(\mathbf{T}_i - \mathbf{T}_j)\frac{\partial \mathbf{S}_{wj}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_i)\}}{\partial \boldsymbol{\beta}^{\mathrm{T}}} \mid \mathbf{O}_i \Bigg) \\
&= \left[\mathbf{C}(\mathbf{O_i})\{\mathbf{B}(\mathbf{T}_i)\mathbf{f_T}(\mathbf{T}_i)\}^{-1}\right] E\Bigg[\mathbf{A}_{wj}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_i)\}\mathbf{V}_{wj}\{\boldsymbol{\beta}_0, m_0, \\
&\quad \mathbf{w}_0(\mathbf{T}_i)\}^{-1}\mathbf{K}_h(\mathbf{T}_i - \mathbf{T}_j)\frac{\partial \mathbf{S}_{wj}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_i)\}}{\partial \boldsymbol{\beta}^{\mathrm{T}}} \mid \mathbf{O}_i\Bigg] \\
&= \mathbf{C}(\mathbf{O_i})\mathbf{B}(\mathbf{T}_i)^{-1}E\Bigg[\mathbf{A}_{wj}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_i)\}\mathbf{V}_{wj}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_i)\}^{-1} \\
&\quad \times \frac{\partial \mathbf{S}_{wj}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_i)\}}{\partial \boldsymbol{\beta}^{\mathrm{T}}} \mid \mathbf{O}_i\Bigg] + O(h^2),
\end{aligned}
$$

which differs from the second term in (S.28) by $O(h^2)$. Therefore

$$
\begin{aligned}
&\Bigg\| \frac{1}{\sqrt{n}} \sum_{i=1}^n \widehat{\mathbf{A}}_{\beta i}[\boldsymbol{\beta}, \widehat{\boldsymbol{\lambda}}\{\boldsymbol{\beta}, \widehat{\mathbf{w}}(\boldsymbol{\beta})\}, \widehat{\mathbf{w}}(\boldsymbol{\beta}, \mathbf{T}_i)]\widehat{\mathbf{V}}_{\beta i}[\boldsymbol{\beta}, \widehat{\boldsymbol{\lambda}}\{\boldsymbol{\beta}, \widehat{\mathbf{w}}(\boldsymbol{\beta})\}, \widehat{\mathbf{w}}(\boldsymbol{\beta}, \mathbf{T}_i)]^{-1} \\
&\widehat{\mathbf{S}}_{\beta i}[\boldsymbol{\beta}, \widehat{\boldsymbol{\lambda}}\{\boldsymbol{\beta}, \widehat{\mathbf{w}}(\boldsymbol{\beta})\}, \widehat{\mathbf{w}}(\boldsymbol{\beta}, \mathbf{T}_i)] - \frac{1}{\sqrt{n}} \sum_{i=1}^n \mathbf{A}^{\dagger}_{\beta i}[\boldsymbol{\beta}, \widehat{\boldsymbol{\lambda}}\{\boldsymbol{\beta}, \widehat{\mathbf{w}}(\boldsymbol{\beta})\}, \widehat{\mathbf{w}}(\boldsymbol{\beta}, \mathbf{T}_i)] \\
&\mathbf{V}^{\dagger}_{\beta i}[\boldsymbol{\beta}, \widehat{\boldsymbol{\lambda}}\{\boldsymbol{\beta}, \widehat{\mathbf{w}}(\boldsymbol{\beta})\}, \widehat{\mathbf{w}}(\boldsymbol{\beta}, \mathbf{T}_i)]^{-1}\mathbf{S}^{\dagger}_{\beta i}[\boldsymbol{\beta}, \widehat{\boldsymbol{\lambda}}\{\boldsymbol{\beta}, \widehat{\mathbf{w}}(\boldsymbol{\beta})\}, \widehat{\mathbf{w}}(\boldsymbol{\beta}, \mathbf{T}_i)]\Bigg\|_\infty \\
= &\ \frac{1}{\sqrt{n}} \sum_{i=1}^n O(h^2) + o_p(1) = o_p(1).
\end{aligned}
$$

when $nh^4 \to 0$. This yields the result in the lemma. $\square$

<u>**Lemma**</u> 7.

$$
\begin{aligned}
&\sqrt{nh_b}\mathbf{B}_r^{\mathrm{T}}(u)\left[\widehat{\boldsymbol{\lambda}}\{\widehat{\boldsymbol{\beta}}, \widehat{\mathbf{w}}(\widehat{\boldsymbol{\beta}})\} - \widehat{\boldsymbol{\lambda}}(\widehat{\boldsymbol{\beta}}, \mathbf{w}_0)\right] \\
= &\ \frac{-h_b^{1/2}}{\sqrt{n}}\mathbf{B}_r^{\mathrm{T}}(u)\mathbf{V}^{-1}\sum_{j=1}^n E\Bigg[\sum_{k=1}^{M_i}\sum_{v=1}^{M_i} E\Bigg\{C_{ikv}\Theta_{ik}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(T_{ik})\} \\
&\times \Theta_{iv}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(T_{iv})\}\mathbf{B}_r\{\mathbf{w}_0(T_{iv})^{\mathrm{T}}\mathbf{Z}_{iv}\}m_0'(\mathbf{w}_0(T_{ik})^{\mathrm{T}}\mathbf{Z}_{ik})\mathbf{Z}_{ik}^{\mathrm{T}} \\
&\times \Bigg(\{\mathbf{B}(T_{ik})f_T(T_{ik})\}^{-1}\Bigg[\mathbf{A}_{wj}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(T_{ik})\}\mathbf{V}_{wj}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(T_{ik})\}^{-1} \\
&\times \mathbf{K}_h(\mathbf{T}_j - T_{ik})\mathbf{S}_{wj}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(T_{ik})\}\Bigg]\Bigg)\Bigg| M_i, \mathbf{O_j}\Bigg\}\Bigg|\mathbf{O_j}\Bigg]\{1 + o_p(1)\} \\
= &\ O_p(1).
\end{aligned}
$$

where C_{iuv} is the u, vth entry of matrix $\boldsymbol{\Omega}^{-1}$.

Proof: Note that

$$
\begin{aligned}
\text{(S.29)}\quad & \sqrt{nh_b}\mathbf{B}_r^{\mathrm{T}}(u)\left[\widehat{\boldsymbol{\lambda}}\{\boldsymbol{\beta},\widehat{\mathbf{w}}(\boldsymbol{\beta})\}-\widehat{\boldsymbol{\lambda}}(\boldsymbol{\beta},\mathbf{w}_0)\right]\\
= & \sqrt{nh_b}\mathbf{B}_r^{\mathrm{T}}(u)\left[\frac{\partial\widehat{\boldsymbol{\lambda}}(\boldsymbol{\beta},\mathbf{w}_0)}{\partial\mathbf{w}^{\mathrm{T}}}\{\widehat{\mathbf{w}}(\boldsymbol{\beta})-\mathbf{w}_0\}\right]\{1+o_p(1)\}\\
= & -\sqrt{nh_b}\mathbf{B}_r^{\mathrm{T}}(u)\Bigg(n^{-1}\sum_{i=1}^{n}\Bigg[\widetilde{\mathbf{Q}}_{\lambda i}\{\mathbf{w}_0(\mathbf{T}_i)\}^{\mathrm{T}}\boldsymbol{\Theta}_i\{\boldsymbol{\beta},m_0,\mathbf{w}_0(\mathbf{T}_i)\}\boldsymbol{\Omega}_i^{-1}\\
& \times\boldsymbol{\Theta}_i\{\boldsymbol{\beta},m_0,\mathbf{w}_0(\mathbf{T}_i)\}\widetilde{\mathbf{Q}}_{\lambda i}\{\mathbf{w}_0(\mathbf{T}_i)\}\Bigg]\Bigg)^{-1}n^{-1}\Bigg[\sum_{i=1}^{n}\widetilde{\mathbf{Q}}_{\lambda i}\{\mathbf{w}_0(\mathbf{T}_i)\}^{\mathrm{T}}\\
& \times\boldsymbol{\Theta}_i\{\boldsymbol{\beta},m_0,\mathbf{w}_0(\mathbf{T}_i)\}\boldsymbol{\Omega}_i^{-1}\boldsymbol{\Theta}_i\{\boldsymbol{\beta},m_0,\mathbf{w}_0(\mathbf{T}_i)\}\\
& \times\widetilde{\mathbf{Q}}_{wi}\{m_0,\{\widehat{\mathbf{w}}(\boldsymbol{\beta},\mathbf{T}_i)-\mathbf{w}_0(\mathbf{T}_i)\}\}\Bigg]\\
& \times\{1+o_p(1)\}
\end{aligned}
$$

We can write out the componentwise form of (S.29) as

(S.30)

$$
\begin{aligned}
& \sqrt{nh_b}\mathbf{B}_r^{\mathrm{T}}(u)\left[\widehat{\boldsymbol{\lambda}}\{\widehat{\boldsymbol{\beta}},\widehat{\mathbf{w}}(\widehat{\boldsymbol{\beta}})\}-\widehat{\boldsymbol{\lambda}}(\widehat{\boldsymbol{\beta}},\mathbf{w}_0)\right]\\
= & \sqrt{nh_b}\mathbf{B}_r^{\mathrm{T}}(u)\mathbf{V}_n^{-1}\sum_{i=1}^{n}\sum_{k=1}^{M_i}\Bigg(\sum_{v=1}^{M_i}\frac{1}{n}C_{ikv}\Theta_{ik}\{\boldsymbol{\beta}_0,m_0,\mathbf{w}_0(T_{ik})\}\\
& \times\Theta_{iv}\{\boldsymbol{\beta}_0,m_0,\mathbf{w}_0(T_{iv})\}\mathbf{B}_r\{\mathbf{w}_0(T_{iv})^{\mathrm{T}}\mathbf{Z}_{iv}\}m_0'(\mathbf{w}_0(T_{ik})^{\mathrm{T}}\mathbf{Z}_{ik})\mathbf{Z}_{ik}^{\mathrm{T}}\Bigg)\\
& \times\Bigg(\frac{1}{n}\sum_{j=1}^{n}\{\mathbf{B}(T_{ik})f_T(T_{ik})\}^{-1}\Bigg[\mathbf{A}_{wj}\{\boldsymbol{\beta}_0,m_0,\mathbf{w}_0(T_{ik})\}\mathbf{V}_{wj}\{\boldsymbol{\beta}_0,m_0,\\
& \mathbf{w}_0(T_{ik})\}^{-1}\mathbf{K}_h(\mathbf{T}_j-T_{ik})\mathbf{S}_{wj}\{\boldsymbol{\beta}_0,m_0,\mathbf{w}_0(T_{ik})\}\Bigg]\Bigg)\{1+o_p(1)\}\\
= & h_b^{1/2}\frac{1}{n^{3/2}}\mathbf{B}_r^{\mathrm{T}}(u)\mathbf{V}_n^{-1}\sum_{j=1}^{n}\sum_{i=1}^{n}\sum_{k=1}^{M_i}\Bigg(\sum_{v=1}^{M_i}\frac{1}{n}C_{ikv}\Theta_{ik}\{\boldsymbol{\beta}_0,m_0,\mathbf{w}_0(T_{ik})\}\\
& \times\Theta_{iv}\{\boldsymbol{\beta}_0,m_0,\mathbf{w}_0(T_{iv})\}\mathbf{B}_r\{\mathbf{w}_0(T_{iv})^{\mathrm{T}}\mathbf{Z}_{iv}\}m_0'(\mathbf{w}_0(T_{ik})^{\mathrm{T}}\mathbf{Z}_{ik})\mathbf{Z}_{ik}^{\mathrm{T}}\Bigg)\\
& \times\Bigg(\{\mathbf{B}(T_{ik})f_T(T_{ik})\}^{-1}\Bigg[\mathbf{A}_{wj}\{\boldsymbol{\beta}_0,m_0,\mathbf{w}_0(T_{ik})\}\mathbf{V}_{wj}\{\boldsymbol{\beta}_0,m_0,\mathbf{w}_0(T_{ik})\}^{-1}\\
& \times\mathbf{K}_h(\mathbf{T}_j-T_{ik})\mathbf{S}_{wj}\{\boldsymbol{\beta}_0,m_0,\mathbf{w}_0(T_{ik})\}\Bigg]\Bigg)\{1+o_p(1)\}
\end{aligned}
$$

$$
\begin{aligned}
&= h_b^{1/2}\mathbf{B}_r^{\mathrm{T}}(u)\mathbf{V}_n^{-1}\Bigg[\frac{1}{\sqrt{n}}\sum_{i=1}^{n}E\{\mathbf{u}(\mathbf{O}_i,\mathbf{O}_j)\mid\mathbf{O}_i\}\\
&\quad+\frac{1}{\sqrt{n}}\sum_{j=1}^{n}E\{\mathbf{u}(\mathbf{O}_i,\mathbf{O}_j)\mid\mathbf{O}_j\}-\sqrt{n}E\{\mathbf{u}(\mathbf{O}_i,\mathbf{O}_j)\}\Bigg]\{1+o_p(1)\},
\end{aligned}
$$

where C_{iuv} is defined in the statement of this Lemma and

$$
\begin{aligned}
&\mathbf{u}(\mathbf{O}_i,\mathbf{O}_j)\\
&= \sum_{k=1}^{M_i}\Bigg(\sum_{v=1}^{M_i}\frac{1}{n}C_{ikv}\Theta_{ik}\{\boldsymbol{\beta}_0,m_0,\mathbf{w}_0(T_{ik})\}\Theta_{iv}\{\boldsymbol{\beta}_0,m_0,\mathbf{w}_0(T_{iv})\}\\
&\quad\times\mathbf{B}_r\{\mathbf{w}_0(T_{iv})^{\mathrm{T}}\mathbf{Z}_{iv}\}m_0'(\mathbf{w}_0(T_{ik})^{\mathrm{T}}\mathbf{Z}_{ik})\mathbf{Z}_{ik}^{\mathrm{T}}\Bigg)\Bigg(\{\mathbf{B}(T_{ik})\\
&\quad\times f_T(T_{ik})\}\Bigg[\mathbf{A}_{wj}\{\boldsymbol{\beta}_0,m_0,\mathbf{w}_0(T_{ik})\}\mathbf{V}_{wj}\{\boldsymbol{\beta}_0,m_0,\mathbf{w}_0(T_{ik})\}^{-1}\\
&\quad\times\mathbf{K}_h(\mathbf{T}_j-T_{ik})\mathbf{S}_{wj}\{\boldsymbol{\beta}_0,m_0,\mathbf{w}_0(T_{ik})\}\Bigg]\Bigg).
\end{aligned}
$$

We have

$$
\begin{aligned}
&E\{\mathbf{u}(\mathbf{O}_i,\mathbf{O}_j)|\mathbf{O}_i\}\\
&= \sum_{k=1}^{M_i}\Bigg(\sum_{v=1}^{M_i}\frac{1}{n}C_{ikv}\Theta_{ik}\{\boldsymbol{\beta}_0,m_0,\mathbf{w}_0(T_{ik})\}\Theta_{iv}\{\boldsymbol{\beta}_0,m_0,\mathbf{w}_0(T_{iv})\}\\
&\quad\times\mathbf{B}_r\{\mathbf{w}_0(T_{iv})^{\mathrm{T}}\mathbf{Z}_{iv}\}m_0'(\mathbf{w}_0(T_{ik})^{\mathrm{T}}\mathbf{Z}_{ik})\mathbf{Z}_{ik}^{\mathrm{T}}\Bigg)\Bigg\{\{\mathbf{B}(T_{ik})f_T(T_{ik})\}^{-1}\\
&\quad\times E\Bigg(\Bigg[\mathbf{A}_{wj}\{\boldsymbol{\beta}_0,m_0,\mathbf{w}_0(T_{ik})\}\mathbf{V}_{wj}\{\boldsymbol{\beta}_0,m_0,\mathbf{w}_0(T_{ik})\}^{-1}\\
&\quad\times\mathbf{K}_h(\mathbf{T}_j-T_{ik})\mathbf{S}_{wj}\{\boldsymbol{\beta}_0,m_0,\mathbf{w}_0(T_{ik})\}\Bigg]\Bigg|T_{ik}\Bigg)\Bigg\},
\end{aligned}
$$

and

$$
\begin{aligned}
&E\Bigg[\mathbf{A}_{wj}\{\boldsymbol{\beta}_0,m_0,\mathbf{w}_0(T_{ik})\}\mathbf{V}_{wj}\{\boldsymbol{\beta}_0,m_0,\mathbf{w}_0(T_{ik})\}^{-1}\mathbf{K}_h(\mathbf{T}_j-T_{ik})\\
&\quad\times\mathbf{S}_{wj}\{\boldsymbol{\beta}_0,m_0,\mathbf{w}_0(T_{ik})\}|T_{ik}\Bigg]\\
&= \int E[\mathbf{A}_{wj}\{\boldsymbol{\beta}_0,m_0,\mathbf{w}_0(T_{ik})\}\mathbf{V}_{wj}\{\boldsymbol{\beta}_0,m_0,\mathbf{w}_0(T_{ik})\}^{-1}\mathbf{K}_h(\mathbf{T}_j-T_{ik})\\
&\quad\times\mathbf{S}_{wj}\{\boldsymbol{\beta}_0,m_0,\mathbf{w}_0(T_{ik})\}\mid\mathbf{T}_j=\mathbf{t}]f_{\mathbf{T}}(\mathbf{t})dt
\end{aligned}
$$

$$
\begin{aligned}
&= E[\mathbf{A}_{wj}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(T_{ik})\}\mathbf{V}_{wj}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(T_{ik})\}^{-1}\mathbf{S}_{wj}\{\boldsymbol{\beta}_0, m_0, \\
&\quad \mathbf{w}_0(T_{ik})\}]f_{\mathbf{T}}(T_{ik}) + O(h^2) \\
&= O(h^2),
\end{aligned}
$$

thus we obtain

$$
\begin{aligned}
&\left\| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} E\{\mathbf{u}(\mathbf{O}_i, \mathbf{O}_j) \mid \mathbf{O}_i\} \right\|_2 \\
\leq\ & \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left\| \sum_{k=1}^{M_i} \left(\sum_{v=1}^{M_i} \frac{1}{n} C_{ikv} \Theta_{ik}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(T_{ik})\}\Theta_{iv}\{\boldsymbol{\beta}_0, m_0, \right.\right. \\
&\left.\left. \mathbf{w}_0(T_{iv})\}\mathbf{B}_r\{\mathbf{w}_0(T_{iv})^{\mathrm{T}}\mathbf{Z}_{iv}\}m_0'(\mathbf{w}_0(T_{ik})^{\mathrm{T}}\mathbf{Z}_{ik})\mathbf{Z}_{ik}^{\mathrm{T}} \right) \right\|_2 O(h^2).
\end{aligned}
$$

Because

$$
\begin{aligned}
&\left\| \sum_{k=1}^{M_i} \left(\sum_{v=1}^{M_i} \frac{1}{n} C_{ikv} \Theta_{ik}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(T_{ik})\}\Theta_{iv}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(T_{iv})\} \right.\right. \\
&\left.\left. \times\mathbf{B}_r\{\mathbf{w}_0(T_{iv})^{\mathrm{T}}\mathbf{Z}_{iv}\}m_0'(\mathbf{w}_0(T_{ik})^{\mathrm{T}}\mathbf{Z}_{ik})\mathbf{Z}_{ik}^{\mathrm{T}} \right) \right\|_2
\end{aligned}
$$

has the form $\left\| \left[\mathbf{C}_i\widetilde{\mathbf{Q}}_\lambda\{\mathbf{w}_0(\mathbf{T}_i)\}\right]^{\mathrm{T}} \right\|_2$, and from (S.11), it is of order $O_p(h_b^{1/2})$. In addition, we have $\|\mathbf{V}_n^{-1}\|_2 = O_p(h_b^{-1})$ by using (S.7). Combining these, we have

$$
\left\| \mathbf{V}_n^{-1} \frac{h_b^{1/2}}{\sqrt{n}} \sum_{i=1}^{n} E\{\mathbf{u}(\mathbf{O}_i, \mathbf{O}_j) \mid \mathbf{O}_i\} \right\|_2 = O_p(n^{1/2}h^2) = o_p(1) \tag{S.31}
$$

and

$$
\left\| \mathbf{V}_n^{-1} h_b^{1/2} \sqrt{n} E\{\mathbf{u}(\mathbf{O}_i, \mathbf{O}_j)\} \right\|_2 = O(n^{1/2}h^2) = o_p(1). \tag{S.32}
$$

due to the assumption that $nh^4 \to 0$. Further

(S.33)

$$
\begin{aligned}
&\mathbf{V}_n^{-1} h_b^{1/2} \frac{1}{\sqrt{n}} \sum_{j=1}^{n} E\{\mathbf{u}(\mathbf{O}_i, \mathbf{O}_j) \mid \mathbf{O}_j\} \\
=\ & \mathbf{V}_n^{-1} h_b^{1/2} \frac{1}{\sqrt{n}} \sum_{j=1}^{n} E\Bigg\{ \sum_{k=1}^{M_i} \Bigg(\sum_{v=1}^{M_i} \frac{1}{n} C_{ikv} \Theta_{ik}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(T_{ik})\}
\end{aligned}
$$

$$
\begin{aligned}
&\times\Theta_{iv}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(T_{iv})\}\mathbf{B}_r\{\mathbf{w}_0(T_{iv})^{\mathrm{T}}\mathbf{Z}_{iv}\}m_0'(\mathbf{w}_0(T_{ik})^{\mathrm{T}}\mathbf{Z}_{ik})\mathbf{Z}_{ik}^{\mathrm{T}}\Big)\\
&\times\Bigg(\{\mathbf{B}(T_{ik})f_T(T_{ik})\}^{-1}\Bigg[\mathbf{A}_{wj}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(T_{ik})\}\mathbf{V}_{wj}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(T_{ik})\}^{-1}\\
&\times\mathbf{K}_h(\mathbf{T}_j - T_{ik})\mathbf{S}_{wj}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(T_{ik})\}\Bigg]\Bigg)\Bigg|\mathbf{O_j}\Bigg\}\\
= \;& \mathbf{V}_n^{-1}h_b^{1/2}\frac{-1}{\sqrt{n}}\sum_{j=1}^n E\Bigg\{\sum_{k=1}^{M_i}\sum_{v=1}^{M_i} C_{ikv}\Theta_{ik}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(T_{ik})\}\\
&\times\Theta_{iv}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(T_{iv})\}\mathbf{B}_r\{\mathbf{w}_0(T_{iv})^{\mathrm{T}}\mathbf{Z}_{iv}\}m_0'(\mathbf{w}_0(T_{ik})^{\mathrm{T}}\mathbf{Z}_{ik})\mathbf{Z}_{ik}^{\mathrm{T}}\\
&\times\Bigg(\{\mathbf{B}(T_{ik})f_T(T_{ik})\}^{-1}\Bigg[\mathbf{A}_{wj}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(T_{ik})\}\mathbf{V}_{wj}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(T_{ik})\}^{-1}\\
&\times\mathbf{K}_h(\mathbf{T}_j - T_{ik})\mathbf{S}_{wj}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(T_{ik})\}\Bigg]\Bigg)\Bigg|\mathbf{O_j}\Bigg\}\\
= \;& \mathbf{V}_n^{-1}h_b^{1/2}\frac{-1}{\sqrt{n}}\sum_{j=1}^n E\Bigg[\sum_{k=1}^{M_i}\sum_{v=1}^{M_i} E\Bigg\{C_{ikv}\Theta_{ik}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(T_{ik})\}\\
&\times\Theta_{iv}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(T_{iv})\}\mathbf{B}_r\{\mathbf{w}_0(T_{iv})^{\mathrm{T}}\mathbf{Z}_{iv}\}m_0'(\mathbf{w}_0(T_{ik})^{\mathrm{T}}\mathbf{Z}_{ik})\mathbf{Z}_{ik}^{\mathrm{T}}\\
&\times\Bigg(\{\mathbf{B}(T_{ik})f_T(T_{ik})\}^{-1}\Bigg[\mathbf{A}_{wj}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(T_{ik})\}\mathbf{V}_{wj}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(T_{ik})\}^{-1}\\
&\times\mathbf{K}_h(\mathbf{T}_j - T_{ik})\mathbf{S}_{wj}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(T_{ik})\}\Bigg]\Bigg)\Bigg|M_i, \mathbf{O_j}\Bigg\}\Bigg|\mathbf{O_j}\Bigg].
\end{aligned}
$$

We consider first the order of

$$
\begin{aligned}
&\Bigg\|\frac{-1}{\sqrt{n}}\sum_{j=1}^n E\Bigg[E\Bigg\{\sum_{k=1}^{M_i}\sum_{v=1}^{M_i} C_{ikv}\Theta_{ik}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(T_{ik})\}\Theta_{iv}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(T_{iv})\}\\
&\times\mathbf{B}_r\{\mathbf{w}_0(T_{iv})^{\mathrm{T}}\mathbf{Z}_{iv}\}m_0'(\mathbf{w}_0(T_{ik})^{\mathrm{T}}\mathbf{Z}_{ik})\mathbf{Z}_{ik}^{\mathrm{T}}\Bigg(\{\mathbf{B}(T_{ik})f_T(T_{ik})\}^{-1}\\
&\times\Bigg[\mathbf{A}_{wj}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(T_{ik})\}\mathbf{V}_{wj}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(T_{ik})\}^{-1}\\
&\mathbf{K}_h(\mathbf{T}_j - T_{ik})\mathbf{S}_{wj}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(T_{ik})\}\Bigg]\Bigg)\Bigg|M_i, \mathbf{O_j}\Bigg\}\Bigg|\mathbf{O_j}\Bigg]\Bigg\|_\infty
\end{aligned}
$$

Because it is a random vector, we only have to consider the order of each element in it, i.e. the order of

(S.34)

$$
\frac{-1}{\sqrt{n}}\sum_{j=1}^n E\Bigg[E\Bigg\{\sum_{k=1}^{M_i}\sum_{v=1}^{M_i} C_{ikv}\Theta_{ik}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(T_{ik})\}\Theta_{iv}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(T_{iv})\}
$$

$$\times B_{rp}\{\mathbf{w}_0(T_{iv})^{\mathrm{T}}\mathbf{Z}_{iv}\}m_0'(\mathbf{w}_0(T_{ik})^{\mathrm{T}}\mathbf{Z}_{ik})\mathbf{Z}_{ik}^{\mathrm{T}}\Bigg(\{\mathbf{B}(T_{ik})f_T(T_{ik})\}^{-1}$$
$$\times\Bigg[\mathbf{A}_{wj}\{\boldsymbol{\beta}_0,m_0,\mathbf{w}_0(T_{ik})\}\mathbf{V}_{wj}\{\boldsymbol{\beta}_0,m_0,\mathbf{w}_0(T_{ik})\}^{-1}$$
$$\mathbf{K}_h(\mathbf{T}_j-T_{ik})\mathbf{S}_{wj}\{\boldsymbol{\beta}_0,m_0,\mathbf{w}_0(T_{ik})\}\Bigg]\Bigg)\Bigg|M_i,\mathbf{O_j}\Bigg\}\Bigg|\mathbf{O_j}\Bigg]$$

for each $1\le p\le d_\lambda$. The order of a random variable can be obtained by considering the expectation and standard deviation of the random variable (Bishop, Fienberg and Holland, 2007). Consider the the order of expectation of (S.34). Clearly, it should be less than the order of $\|\sqrt{n}E\{\mathbf{u}(\mathbf{O}_i,\mathbf{O}_j)\}\|_2$, which is of the order $O(n^{1/2}h_b^{1/2}h^2)$ by (S.32) and the fact that $\|\mathbf{V}_n^{-1}\|_2=O_p(h_b^{-1})$. Now we consider the second moment of (S.34). Note that because M_i is finite almost surely, the second moment of (S.34) has the same order as that for each k,v, i.e. the order of

$$E\Bigg(\Bigg[E\Bigg\{C_{ikv}\Theta_{ik}\{\boldsymbol{\beta}_0,m_0,\mathbf{w}_0(T_{ik})\}\Theta_{iv}\{\boldsymbol{\beta}_0,m_0,\mathbf{w}_0(T_{iv})\}$$
$$\times B_{rp}\{\mathbf{w}_0(T_{iv})^{\mathrm{T}}\mathbf{Z}_{iv}\}m_0'(\mathbf{w}_0(T_{ik})^{\mathrm{T}}\mathbf{Z}_{ik})\mathbf{Z}_{ik}^{\mathrm{T}}\Bigg(\{\mathbf{B}(T_{ik})f_T(T_{ik})\}^{-1}$$
$$\times\Bigg[\mathbf{A}_{wj}\{\boldsymbol{\beta}_0,m_0,\mathbf{w}_0(T_{ik})\}\mathbf{V}_{wj}\{\boldsymbol{\beta}_0,m_0,\mathbf{w}_0(T_{ik})\}^{-1}$$
$$\mathbf{K}_h(\mathbf{T}_j-T_{ik})\mathbf{S}_{wj}\{\boldsymbol{\beta}_0,m_0,\mathbf{w}_0(T_{ik})\}\Bigg]\Bigg)\Bigg|\mathbf{O_j}\Bigg\}\Bigg]^2\Bigg).$$

Therefore we derive the order of the above quantity as follows.

(S.35)

$$E\Bigg(\Bigg[E\Bigg\{C_{ikv}\Theta_{ik}\{\boldsymbol{\beta}_0,m_0,\mathbf{w}_0(T_{ik})\}\Theta_{iv}\{\boldsymbol{\beta}_0,m_0,\mathbf{w}_0(T_{iv})\}$$
$$\times B_{rp}\{\mathbf{w}_0(T_{iv})^{\mathrm{T}}\mathbf{Z}_{iv}\}m_0'(\mathbf{w}_0(T_{ik})^{\mathrm{T}}\mathbf{Z}_{ik})\mathbf{Z}_{ik}^{\mathrm{T}}\Bigg(\{\mathbf{B}(T_{ik})f_T(T_{ik})\}^{-1}$$
$$\times\Bigg[\mathbf{A}_{wj}\{\boldsymbol{\beta}_0,m_0,\mathbf{w}_0(T_{ik})\}\mathbf{V}_{wj}\{\boldsymbol{\beta}_0,m_0,\mathbf{w}_0(T_{ik})\}^{-1}$$
$$\mathbf{K}_h(\mathbf{T}_j-T_{ik})\mathbf{S}_{wj}\{\boldsymbol{\beta}_0,m_0,\mathbf{w}_0(T_{ik})\}\Bigg]\Bigg)\Bigg|\mathbf{O_j}\Bigg\}\Bigg]^2\Bigg)$$

$$\le\quad E\Bigg(\Bigg[\sup B_{rp}\{\mathbf{w}_0(T_{iv})^{\mathrm{T}}\mathbf{Z}_{iv}\}E\Bigg\{C_{ikv}\Theta_{ik}\{\boldsymbol{\beta}_0,m_0,\mathbf{w}_0(T_{ik})\}$$

$$
\begin{aligned}
& \times\Theta_{iv}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(T_{iv})\} m_0'(\mathbf{w}_0(T_{ik})^{\mathrm{T}}\mathbf{Z}_{ik})\mathbf{Z}_{ik}^{\mathrm{T}}\Bigg(\{\mathbf{B}(T_{ik})f_T(T_{ik})\}^{-1} \\
& \times\Bigg[\mathbf{A}_{wj}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(T_{ik})\}\mathbf{V}_{wj}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(T_{ik})\}^{-1} \\
& \mathbf{K}_h(\mathbf{T}_j - T_{ik})\mathbf{S}_{wj}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(T_{ik})\}\Bigg]\Bigg)\Bigg|\mathbf{O_j}\Bigg\}\Bigg]^2\Bigg) \\
\leq \quad & \Big(E\left[\sup B_{rp}\{\mathbf{w}_0(T_{iv})^{\mathrm{T}}\mathbf{Z}_{iv}\}\right]^2\Big)\Bigg[E\Bigg\{C_{ikv}\Theta_{ik}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(T_{ik})\} \\
& \times\Theta_{iv}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(T_{iv})\} m_0'(\mathbf{w}_0(T_{ik})^{\mathrm{T}}\mathbf{Z}_{ik})\mathbf{Z}_{ik}^{\mathrm{T}}\Bigg(\{\mathbf{B}(T_{ik})f_T(T_{ik})\}^{-1} \\
& \times\Bigg[\mathbf{A}_{wj}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(T_{ik})\}\mathbf{V}_{wj}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(T_{ik})\}^{-1} \\
& \mathbf{K}_h(\mathbf{T}_j - T_{ik})\mathbf{S}_{wj}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(T_{ik})\}\Bigg]\Bigg)\Bigg|\mathbf{O_j}\Bigg\}^2\Bigg].
\end{aligned}
$$

It is easy to see that

$$
\Big(E\left[\sup B_{rp}\{\mathbf{w}_0(T_{iv})^{\mathrm{T}}\mathbf{Z}_{iv}\}\right]^2\Big) = O(h_b).
$$

Now consider the order of the second expectation in (S.35). Let

$$
\begin{aligned}
\mathcal{J}_{ij}(T_{ik}) \quad \equiv \quad & C_{ikv}\Theta_{ik}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(T_{ik})\}\Theta_{iv}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(T_{iv})\} \\
& \times m_0'(\mathbf{w}_0(T_{ik})^{\mathrm{T}}\mathbf{Z}_{ik})\mathbf{Z}_{ik}^{\mathrm{T}}\Bigg(\{\mathbf{B}(T_{ik})\mathbf{A}_{wj}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(T_{ik})\} \\
& \times\mathbf{V}_{wj}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(T_{ik})\}^{-1}\Bigg)
\end{aligned}
$$

and $\mathcal{J}_{ijl}(T_{ik})$ be the lth column of it, then we have

$$
\begin{aligned}
& E\Bigg\{C_{ikv}\Theta_{ik}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(T_{ik})\}\Theta_{iv}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(T_{iv})\} \\
& m_0'(\mathbf{w}_0(T_{ik})^{\mathrm{T}}\mathbf{Z}_{ik})\mathbf{Z}_{ik}^{\mathrm{T}}\Bigg(\{\mathbf{B}(T_{ik})f_T(T_{ik})\}^{-1}\Bigg[\mathbf{A}_{wj}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(T_{ik})\} \\
& \times\mathbf{V}_{wj}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(T_{ik})\}^{-1}\mathbf{K}_h(\mathbf{T}_j - T_{ik})\mathbf{S}_{wj}\{\boldsymbol{\beta}_0, m_0, \\
& \mathbf{w}_0(T_{ik})\}\Bigg]\Bigg)\Bigg|\mathbf{O_j}\Bigg\} \\
= \quad & E\Bigg[\mathcal{J}_{ij}(T_{ik})f_T(T_{ik})^{-1}\mathbf{K}_h(\mathbf{T}_j - T_{ik})\mathbf{S}_{wj}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(T_{ik})\}\Bigg|\mathbf{O_j}\Bigg] \\
= \quad & E\Bigg[\sum_{l=1}^{M_j}\mathcal{J}_{ijl}(T_{ik})f_T(T_{ik})^{-1}K_h(T_{jl} - T_{ik})\mathbf{S}_{wj}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(T_{ik})\}\Bigg|\mathbf{O_j}\Bigg]
\end{aligned}
$$

$$= \sum_{l=1}^{M_j} E\left\{ \mathcal{J}_{ijl}(T_{ik}) \middle| T_{ik} = T_{jl}, \mathbf{O_j} \right\} \mathbf{S}_{wj}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(T_{jl}) + O(h^2).$$

The expectation of the square of the above quantity, i.e., the second expectation in (S.35), has the order of $O(1)$. Therefore, (S.35) has the order $O(h_b)$. As a result, under the assumption that M_i is finite almost surely, (S.34) is of order $O_p(h_b^{1/2}) + O_p(n^{1/2}h_b^{1/2}h^2) = O_p(h_b^{1/2})$. Combined with (S.33)and the fact that $\|\mathbf{V}_n^{-1}\|_\infty = h_b^{-1}$ by (S.10) we have

(S.36)

$$\begin{aligned}
& \left\| \mathbf{V}_n^{-1} h_b^{1/2} \frac{1}{\sqrt{n}} \sum_{j=1}^{n} E\{\mathbf{u}(\mathbf{O}_i, \mathbf{O}_j) \mid \mathbf{O}_j\} \right\|_\infty \\
= & \left\| \mathbf{V}_n^{-1} \frac{h_b^{1/2}}{\sqrt{n}} \sum_{j=1}^{n} E\Bigg[\sum_{k=1}^{M_i} \sum_{v=1}^{M_i} E\Bigg\{ C_{ikv}\Theta_{ik}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(T_{ik})\} \right. \\
& \times \Theta_{iv}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(T_{iv})\}\mathbf{B}_r\{\mathbf{w}_0(T_{iv})^{\mathrm{T}}\mathbf{Z}_{iv}\} m_0'(\mathbf{w}_0(T_{ik})^{\mathrm{T}}\mathbf{Z}_{ik})\mathbf{Z}_{ik}^{\mathrm{T}} \\
& \times \Bigg(\{\mathbf{B}(T_{ik}) f_T(T_{ik})\}^{-1} \Bigg[\mathbf{A}_{wj}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(T_{ik})\} \mathbf{V}_{wj}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(T_{ik})\}^{-1} \\
& \left. \times \mathbf{K}_h(\mathbf{T}_j - T_{ik}) \mathbf{S}_{wj}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(T_{ik})\} \Bigg] \Bigg) \Bigg| M_i, \mathbf{O_j} \Bigg\} \Bigg| \mathbf{O_j} \Bigg] \right\|_\infty \\
= & \; O_p(1).
\end{aligned}$$

Further, plug the (S.31), (S.32), and (S.36) into (S.30), with the fact that that $\sum_{p=1}^{d_\lambda} B_{rp}(u) = 1$, and $\|\mathbf{B}_r(u)\|_2 < \infty$, we can obtain

$$\begin{aligned}
& \left| \sqrt{nh_b}\mathbf{B}_r^{\mathrm{T}}(u) \left[\widehat{\boldsymbol{\lambda}}\{\widehat{\boldsymbol{\beta}}, \widehat{\mathbf{w}}(\widehat{\boldsymbol{\beta}})\} - \widehat{\boldsymbol{\lambda}}(\widehat{\boldsymbol{\beta}}, \mathbf{w}_0) \right] \right| \\
\asymp & \left| h_b^{1/2}\mathbf{B}_r^{\mathrm{T}}(u)\mathbf{V}_n^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} E\{\mathbf{u}(\mathbf{O}_i, \mathbf{O}_j) \mid \mathbf{O}_j\} + h_b^{1/2}\mathbf{B}_r^{\mathrm{T}}(u)\mathbf{V}_n^{-1} \right. \\
& \left. \times \frac{1}{\sqrt{n}} \sum_{j=1}^{n} E\{\mathbf{u}(\mathbf{O}_i, \mathbf{O}_j) \mid \mathbf{O}_i\} - h_b^{1/2}\mathbf{B}_r^{\mathrm{T}}(u)\mathbf{V}_n^{-1}\sqrt{n}E\{\mathbf{u}(\mathbf{O}_i, \mathbf{O}_j)\} \right| \\
\leq & \left| h_b^{1/2}\mathbf{B}_r^{\mathrm{T}}(u)\mathbf{V}_n^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} E\{\mathbf{u}(\mathbf{O}_i, \mathbf{O}_j) \mid \mathbf{O}_j\} \right| \\
& + \left| h_b^{1/2}\mathbf{B}_r^{\mathrm{T}}(u)\mathbf{V}_n^{-1} \frac{1}{\sqrt{n}} \sum_{j=1}^{n} E\{\mathbf{u}(\mathbf{O}_i, \mathbf{O}_j) \mid \mathbf{O}_i\} \right|
\end{aligned}$$

$$
\begin{aligned}
& +\left|h_b^{1/2}\mathbf{B}_r^{\mathrm{T}}(u)\mathbf{V}_n^{-1}\sqrt{n}E\{\mathbf{u}(\mathbf{O}_i,\mathbf{O}_j)\}\right| \\
\le\ & \mathbf{B}_r^{\mathrm{T}}(u)\mathbf{1}\left\|\left|h_b^{1/2}\mathbf{V}_n^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}E\{\mathbf{u}(\mathbf{O}_i,\mathbf{O}_j)\mid\mathbf{O}_j\}\right|\right\|_\infty \\
& +\|\mathbf{B}_r^{\mathrm{T}}(u)\|_2\left\|\mathbf{V}_n^{-1}h_b^{1/2}\frac{1}{\sqrt{n}}\sum_{j=1}^{n}E\{\mathbf{u}(\mathbf{O}_i,\mathbf{O}_j)\mid\mathbf{O}_i\}\right\|_2 \\
& +\|\mathbf{B}_r^{\mathrm{T}}(u)\|_2\left\|\mathbf{V}_n^{-1}h_b^{1/2}\sqrt{n}E\{\mathbf{u}(\mathbf{O}_i,\mathbf{O}_j)\}\right\|_2 \\
=\ & O_p(1).
\end{aligned}
$$

From the above derivations, the dominate term in (S.30) is

$$
h_b^{1/2}\mathbf{B}_r^{\mathrm{T}}(u)\mathbf{V}_n^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}E\{\mathbf{u}(\mathbf{O}_i,\mathbf{O}_j)\mid\mathbf{O}_j\}.
$$

Use (S.33) and the fact that

$$
\|\mathbf{V}^{-1}\|_2=\|\mathbf{V}_n^{-1}\|_2\{1+o_p(1)\}
$$

by Lemma 2, we have

$$
\begin{aligned}
& \sqrt{nh_b}\mathbf{B}_r^{\mathrm{T}}(u)\left[\widehat{\boldsymbol\lambda}\{\widehat{\boldsymbol\beta},\widehat{\mathbf{w}}(\widehat{\boldsymbol\beta})\}-\widehat{\boldsymbol\lambda}(\widehat{\boldsymbol\beta},\mathbf{w}_0)\right] \\
=\ & \frac{-h_b^{1/2}}{\sqrt{n}}\mathbf{B}_r^{\mathrm{T}}(u)\mathbf{V}_n^{-1}\sum_{j=1}^{n}E\Bigg[\sum_{k=1}^{M_i}\sum_{v=1}^{M_i}E\Bigg\{C_{ikv}\Theta_{ik}\{\boldsymbol\beta_0,m_0,\mathbf{w}_0(T_{ik})\} \\
& \times\Theta_{iv}\{\boldsymbol\beta_0,m_0,\mathbf{w}_0(T_{iv})\}\mathbf{B}_r\{\mathbf{w}_0(T_{iv})^{\mathrm{T}}\mathbf{Z}_{iv}\}m_0'(\mathbf{w}_0(T_{ik})^{\mathrm{T}}\mathbf{Z}_{ik})\mathbf{Z}_{ik}^{\mathrm{T}} \\
& \times\Bigg(\{\mathbf{B}(T_{ik})f_T(T_{ik})\}^{-1}\Bigg[\mathbf{A}_{wj}\{\boldsymbol\beta_0,m_0,\mathbf{w}_0(T_{ik})\}\mathbf{V}_{wj}\{\boldsymbol\beta_0,m_0,\mathbf{w}_0(T_{ik})\}^{-1} \\
& \times\mathbf{K}_h(\mathbf{T}_j-T_{ik})\mathbf{S}_{wj}\{\boldsymbol\beta_0,m_0,\mathbf{w}_0(T_{ik})\}\Bigg]\Bigg)\Bigg|M_i,\mathbf{O_j}\Bigg\}\Bigg|\mathbf{O_j}\Bigg]\{1+o_p(1)\} \\
=\ & \frac{-h_b^{1/2}}{\sqrt{n}}\mathbf{B}_r^{\mathrm{T}}(u)\mathbf{V}^{-1}\sum_{j=1}^{n}E\Bigg[\sum_{k=1}^{M_i}\sum_{v=1}^{M_i}E\Bigg\{C_{ikv}\Theta_{ik}\{\boldsymbol\beta_0,m_0,\mathbf{w}_0(T_{ik})\} \\
& \times\Theta_{iv}\{\boldsymbol\beta_0,m_0,\mathbf{w}_0(T_{iv})\}\mathbf{B}_r\{\mathbf{w}_0(T_{iv})^{\mathrm{T}}\mathbf{Z}_{iv}\}m_0'(\mathbf{w}_0(T_{ik})^{\mathrm{T}}\mathbf{Z}_{ik})\mathbf{Z}_{ik}^{\mathrm{T}} \\
& \times\Bigg(\{\mathbf{B}(T_{ik})f_T(T_{ik})\}^{-1}\Bigg[\mathbf{A}_{wj}\{\boldsymbol\beta_0,m_0,\mathbf{w}_0(T_{ik})\}\mathbf{V}_{wj}\{\boldsymbol\beta_0,m_0,\mathbf{w}_0(T_{ik})\}^{-1} \\
& \times\mathbf{K}_h(\mathbf{T}_j-T_{ik})\mathbf{S}_{wj}\{\boldsymbol\beta_0,m_0,\mathbf{w}_0(T_{ik})\}\Bigg]\Bigg)\Bigg|M_i,\mathbf{O_j}\Bigg\}\Bigg|\mathbf{O_j}\Bigg]\{1+o_p(1)\}.
\end{aligned}
$$

This proves the result. □

Lemma 8.

$$
\begin{aligned}
&\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\bigg(\mathbf{A}^{\dagger}_{\beta i}[\boldsymbol{\beta}_0,\widehat{\boldsymbol{\lambda}}\{\boldsymbol{\beta}_0,\widehat{\mathbf{w}}(\boldsymbol{\beta}_0)\},\widehat{\mathbf{w}}(\boldsymbol{\beta}_0,\mathbf{T}_i)]\mathbf{V}^{\dagger}_{\beta i}[\boldsymbol{\beta}_0,\widehat{\boldsymbol{\lambda}}\{\boldsymbol{\beta}_0,\widehat{\mathbf{w}}(\boldsymbol{\beta}_0)\},\\
&\widehat{\mathbf{w}}(\boldsymbol{\beta}_0,\mathbf{T}_i)]^{-1}\mathbf{S}^{\dagger}_{\beta i}[\boldsymbol{\beta}_0,\widehat{\boldsymbol{\lambda}}\{\boldsymbol{\beta}_0,\widehat{\mathbf{w}}(\boldsymbol{\beta}_0)\},\widehat{\mathbf{w}}(\boldsymbol{\beta}_0,\mathbf{T}_i)]-\mathbf{A}^{\dagger}_{\beta i}\{\boldsymbol{\beta}_0,\widehat{\boldsymbol{\lambda}}(\boldsymbol{\beta}_0,\mathbf{w}_0),\\
&\mathbf{w}_0(\mathbf{T}_i)\}\mathbf{V}^{\dagger}_{\beta i}\{\boldsymbol{\beta}_0,\widehat{\boldsymbol{\lambda}}(\boldsymbol{\beta}_0,\mathbf{w}_0),\mathbf{w}_0(\mathbf{T}_i)\}^{-1}\mathbf{S}^{\dagger}_{\beta i}\{\boldsymbol{\beta}_0,\widehat{\boldsymbol{\lambda}}(\boldsymbol{\beta}_0,\mathbf{w}_0),\mathbf{w}_0(\mathbf{T}_i)\}\bigg)\\
=\;&-\frac{1}{\sqrt{n}}\sum_{j=1}^{n}E\bigg(\mathbf{A}^{\dagger}_{\beta i}\{\boldsymbol{\beta}_0,\boldsymbol{\lambda}_0,\mathbf{w}_0(\mathbf{T}_j)\}\mathbf{V}^{\dagger}_{\beta i}\{\boldsymbol{\beta}_0,\boldsymbol{\lambda}_0,\mathbf{w}_0(\mathbf{T}_i)\}^{-1}\widehat{\mathcal{K}}(\mathbf{T}_j)\bigg|\mathbf{O}_j\bigg)\\
&\times\mathbf{B}(\mathbf{T}_j)^{-1}\bigg[\mathbf{A}_{wj}\{\boldsymbol{\beta}_0,m_0,\mathbf{w}_0(\mathbf{T}_j)\}\mathbf{V}_{wj}\{\boldsymbol{\beta}_0,m_0,\mathbf{w}_0(\mathbf{T}_j)\}^{-1}\\
&\times\mathbf{S}_{wj}\{\boldsymbol{\beta}_0,m_0,\mathbf{w}_0(\mathbf{T}_j)\}\bigg]+o_p(1),
\end{aligned}
$$

where $\mathbf{A}^{\dagger},\mathbf{V}^{\dagger},\mathbf{S}^{\dagger}$ are defined at the beginning of Section S.3,
$\widehat{\mathcal{K}}(\mathbf{T}_i)=diag\{\widehat{\boldsymbol{\kappa}}(T_{ik}),k=1,\ldots,M_i\}$ *a $d_\beta M_i\times d_w M_i$ matrix. and*

$$
\begin{aligned}
\widehat{\boldsymbol{\kappa}}(T_{ik})\;=\;&\bigg\{\mathbf{Q}_{\beta ik}+\bigg(\frac{\partial\widehat{\boldsymbol{\lambda}}(\boldsymbol{\beta}_0,\mathbf{w}_0)}{\partial\boldsymbol{\beta}^{\mathrm{T}}}-\mathbf{Q}_{\lambda ik}\{\mathbf{w}_0(T_{ik})\}^{\mathrm{T}}\frac{\partial\widehat{\boldsymbol{\lambda}}(\boldsymbol{\beta}_0,\mathbf{w}_0)}{\partial\mathbf{w}}\mathbf{B}(T_{ik})^{-1}\\
&E\bigg[\mathbf{A}_{wj}\{\boldsymbol{\beta}_0,m_0,\mathbf{w}_0(T_{ik})\}\mathbf{V}_{wj}\{\boldsymbol{\beta}_0,m_0,\mathbf{w}_0(T_{ik})\}^{-1}\\
&\times\frac{\partial\mathbf{S}_{wj}\{\boldsymbol{\beta}_0,m_0,\mathbf{w}_0(T_{ik})\}}{\partial\boldsymbol{\beta}^{\mathrm{T}}}\mid\mathbf{O}_i\bigg]\bigg)^{\mathrm{T}}\\
&-\bigg(\mathbf{B}(T_{ik})^{-1}E\bigg[\mathbf{A}_{wj}\{\boldsymbol{\beta}_0,m_0,\mathbf{w}_0(T_{ik})\}\mathbf{V}_{wj}\{\boldsymbol{\beta}_0,m_0,\mathbf{w}_0(T_{ik})\}^{-1}\\
&\frac{\partial\mathbf{S}_{wj}\{\boldsymbol{\beta}_0,m_0,\mathbf{w}_0(T_{ik})\}}{\partial\boldsymbol{\beta}^{\mathrm{T}}}\mid\mathbf{O}_i\bigg]\bigg)^{\mathrm{T}}\mathbf{Z}_{ik}m_0'\{\mathbf{w}_0(T_{ik})^{\mathrm{T}}\mathbf{Z}_{ik}\}\bigg\}\\
&\times\mathbf{Q}_{wik}\{m_0,\mathbf{w}_0(T_{ik})\}\Theta_{ik}\{\boldsymbol{\beta}_0,m_0,\mathbf{w}_0(T_{ik})\}.
\end{aligned}
$$

Proof:

$$
\begin{aligned}
&\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\bigg(\mathbf{A}^{\dagger}_{\beta i}[\boldsymbol{\beta}_0,\widehat{\boldsymbol{\lambda}}\{\boldsymbol{\beta}_0,\widehat{\mathbf{w}}(\boldsymbol{\beta}_0)\},\widehat{\mathbf{w}}(\boldsymbol{\beta}_0,\mathbf{T}_i)]\mathbf{V}^{\dagger}_{\beta i}[\boldsymbol{\beta}_0,\widehat{\boldsymbol{\lambda}}\{\boldsymbol{\beta}_0,\widehat{\mathbf{w}}(\boldsymbol{\beta}_0)\},\\
&\widehat{\mathbf{w}}(\boldsymbol{\beta}_0,\mathbf{T}_i)]^{-1}\mathbf{S}^{\dagger}_{\beta i}[\boldsymbol{\beta}_0,\widehat{\boldsymbol{\lambda}}\{\boldsymbol{\beta}_0,\widehat{\mathbf{w}}(\boldsymbol{\beta}_0)\},\widehat{\mathbf{w}}(\boldsymbol{\beta}_0,\mathbf{T}_i)]-\mathbf{A}^{\dagger}_{\beta i}\{\boldsymbol{\beta}_0,\widehat{\boldsymbol{\lambda}}(\boldsymbol{\beta}_0,\mathbf{w}_0),\\
&\mathbf{w}_0(\mathbf{T}_i)\}\mathbf{V}^{\dagger}_{\beta i}\{\boldsymbol{\beta}_0,\widehat{\boldsymbol{\lambda}}(\boldsymbol{\beta}_0,\mathbf{w}_0),\mathbf{w}_0(\mathbf{T}_i)\}^{-1}\mathbf{S}^{\dagger}_{\beta i}\{\boldsymbol{\beta}_0,\widehat{\boldsymbol{\lambda}}(\boldsymbol{\beta}_0,\mathbf{w}_0),\mathbf{w}_0(\mathbf{T}_i)\}\bigg)\\
=\;&\frac{1}{\sqrt{n}}\bigg[\sum_{i=1}^{n}\mathbf{A}^{\dagger}_{\beta i}\{\boldsymbol{\beta}_0,\widehat{\boldsymbol{\lambda}}(\boldsymbol{\beta}_0,\mathbf{w}_0),\mathbf{w}_0(\mathbf{T}_i)\}\mathbf{V}^{\dagger}_{\beta i}\{\boldsymbol{\beta}_0,\widehat{\boldsymbol{\lambda}}(\boldsymbol{\beta}_0,\mathbf{w}_0),\mathbf{w}_0(\mathbf{T}_i)\}^{-1}
\end{aligned}
$$

$$\times \frac{\partial \mathbf{S}^{\dagger}_{\beta i}\{\boldsymbol{\beta}_0, \widehat{\boldsymbol{\lambda}}(\boldsymbol{\beta}_0, \mathbf{w}_0), \mathbf{w}_0(\mathbf{T}_i)\}}{\partial \mathbf{w}}\Bigg] \{\widehat{\mathbf{w}}(\boldsymbol{\beta}_0) - \mathbf{w}_0\}\{1 + o_p(1)\}$$

Now further we have

$$\begin{aligned}
& \frac{\partial \mathbf{S}^{\dagger}_{\beta ik}[\boldsymbol{\beta}, \widehat{\boldsymbol{\lambda}}(\boldsymbol{\beta}_0, \mathbf{w}_0), \mathbf{w}_0(T_{ik})]}{\partial \mathbf{w}}\{\widehat{\mathbf{w}}(\boldsymbol{\beta}_0) - \mathbf{w}_0\} \\
= \; & \Bigg\{\mathbf{Q}_{\beta ik} + \Bigg(\frac{\partial \widehat{\boldsymbol{\lambda}}(\boldsymbol{\beta}_0, \mathbf{w}_0)}{\partial \boldsymbol{\beta}^{\mathrm{T}}} - \mathbf{Q}_{\lambda ik}\{\mathbf{w}_0(T_{ik})\}^{\mathrm{T}} \frac{\partial \widehat{\boldsymbol{\lambda}}(\boldsymbol{\beta}_0, \mathbf{w}_0)}{\partial \mathbf{w}} \mathbf{B}(T_{ik})^{-1} \\
& E\Bigg[\mathbf{A}_{wj}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(T_{ik})\}\mathbf{V}_{wj}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(T_{ik})\}^{-1} \\
& \times \frac{\partial \mathbf{S}_{wj}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(T_{ik})\}}{\partial \boldsymbol{\beta}^{\mathrm{T}}} \mid \mathbf{O}_i\Bigg]\Bigg)^{\mathrm{T}} \\
& -\Bigg(\mathbf{B}(T_{ik})^{-1} E\Bigg[\mathbf{A}_{wj}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(T_{ik})\}\mathbf{V}_{wj}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(T_{ik})\}^{-1} \\
& \frac{\partial \mathbf{S}_{wj}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(T_{ik})\}}{\partial \boldsymbol{\beta}^{\mathrm{T}}} \mid \mathbf{O}_i\Bigg]\Bigg)^{\mathrm{T}} \mathbf{Z}_{ik} m_0'\{\mathbf{w}_0(T_{ik})^{\mathrm{T}} \mathbf{Z}_{ik}\}\Bigg\} \\
& \times \frac{\partial \Big[D_{ik} - H_{ik}\{\boldsymbol{\beta}, \widehat{\boldsymbol{\lambda}}(\boldsymbol{\beta}_0, \mathbf{w}_0), \mathbf{w}_0(T_{ik})\}\Big]}{\partial \mathbf{w}} \{\widehat{\mathbf{w}}(\boldsymbol{\beta}_0) - \mathbf{w}_0\}\{1 + o_p(1)\} \\
= \; & \Bigg\{\mathbf{Q}_{\beta ik} + \Bigg(\frac{\partial \widehat{\boldsymbol{\lambda}}(\boldsymbol{\beta}_0, \mathbf{w}_0)}{\partial \boldsymbol{\beta}^{\mathrm{T}}} - \mathbf{Q}_{\lambda ik}\{\mathbf{w}_0(T_{ik})\}^{\mathrm{T}} \frac{\partial \widehat{\boldsymbol{\lambda}}(\boldsymbol{\beta}_0, \mathbf{w}_0)}{\partial \mathbf{w}} \mathbf{B}(T_{ik})^{-1} \\
& E\Bigg[\mathbf{A}_{wj}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(T_{ik})\}\mathbf{V}_{wj}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(T_{ik})\}^{-1} \\
& \times \frac{\partial \mathbf{S}_{wj}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(T_{ik})\}}{\partial \boldsymbol{\beta}^{\mathrm{T}}} \mid \mathbf{O}_i\Bigg]\Bigg)^{\mathrm{T}} \\
& -\Bigg(\mathbf{B}(T_{ik})^{-1} E\Bigg[\mathbf{A}_{wj}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(T_{ik})\}\mathbf{V}_{wj}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(T_{ik})\}^{-1} \\
& \frac{\partial \mathbf{S}_{wj}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(T_{ik})\}}{\partial \boldsymbol{\beta}^{\mathrm{T}}} \mid \mathbf{O}_i\Bigg]\Bigg)^{\mathrm{T}} \mathbf{Z}_{ik} m_0'\{\mathbf{w}_0(T_{ik})^{\mathrm{T}} \mathbf{Z}_{ik}\}\Bigg\} \\
& \times \frac{\partial \Big[D_{ik} - H_{ik}\{\boldsymbol{\beta}, \widehat{\boldsymbol{\lambda}}(\boldsymbol{\beta}_0, \mathbf{w}_0), \mathbf{w}_0(T_{ik})\}\Big]}{\partial \mathbf{w}} \{\widehat{\mathbf{w}}(\boldsymbol{\beta}_0) - \mathbf{w}_0\}\{1 + o_p(1)\} \\
= \; & \Bigg\{\mathbf{Q}_{\beta ik} + \Bigg(\frac{\partial \widehat{\boldsymbol{\lambda}}(\boldsymbol{\beta}_0, \mathbf{w}_0)}{\partial \boldsymbol{\beta}^{\mathrm{T}}} - \mathbf{Q}_{\lambda ik}\{\mathbf{w}_0(T_{ik})\}^{\mathrm{T}} \frac{\partial \widehat{\boldsymbol{\lambda}}(\boldsymbol{\beta}_0, \mathbf{w}_0)}{\partial \mathbf{w}} \mathbf{B}(T_{ik})^{-1} \\
& E\Bigg[\mathbf{A}_{wj}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(T_{ik})\}\mathbf{V}_{wj}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(T_{ik})\}^{-1}
\end{aligned}$$

$$\times \frac{\partial \mathbf{S}_{wj}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(T_{ik})\}}{\partial \boldsymbol{\beta}^{\mathrm{T}}} \mid \mathbf{O}_i\Big]\Big)^{\mathrm{T}}$$
$$-\Big(\mathbf{B}(T_{ik})^{-1} E\Big[\mathbf{A}_{wj}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(T_{ik})\}\mathbf{V}_{wj}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(T_{ik})\}^{-1}$$
$$\frac{\partial \mathbf{S}_{wj}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(T_{ik})\}}{\partial \boldsymbol{\beta}^{\mathrm{T}}} \mid \mathbf{O}_i\Big]\Big)^{\mathrm{T}} \mathbf{Z}_{ik} m_0'\{\mathbf{w}_0(T_{ik})^{\mathrm{T}}\mathbf{Z}_{ik}\}\Big\}$$
$$\times \Big[\mathbf{Q}_{wik}\{\widehat{\boldsymbol{\lambda}}(\boldsymbol{\beta}_0, \mathbf{w}_0), \mathbf{w}_0(T_{ik})\} + \mathbf{Q}_{\lambda ik}^{\mathrm{T}}\{\mathbf{w}_0(T_{ik})\}\left\{\frac{\partial \widehat{\boldsymbol{\lambda}}(\boldsymbol{\beta}_0, \mathbf{w}_0)}{\partial \mathbf{w}}\right\}\Big]$$
$$\times \Theta_{ik}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(T_{ik})\}\{\widehat{\mathbf{w}}(\boldsymbol{\beta}_0) - \mathbf{w}_0\}\{1 + o_p(1)\}.$$

Therefore, by using the same argument which lead to (S.23), we can obtain

$$\frac{\partial \mathbf{S}^{\dagger}_{\beta ik}[\boldsymbol{\beta}, \widehat{\boldsymbol{\lambda}}(\boldsymbol{\beta}_0, \mathbf{w}_0), \mathbf{w}_0(T_{ik})]}{\partial \mathbf{w}}\{\widehat{\mathbf{w}}(\boldsymbol{\beta}_0) - \mathbf{w}_0\}$$
$$= \Big\{\mathbf{Q}_{\beta ik} + \Big(\frac{\partial \widehat{\boldsymbol{\lambda}}(\boldsymbol{\beta}_0, \mathbf{w}_0)}{\partial \boldsymbol{\beta}^{\mathrm{T}}} - \mathbf{Q}_{\lambda ik}\{\mathbf{w}_0(T_{ik})\}^{\mathrm{T}} \frac{\partial \widehat{\boldsymbol{\lambda}}(\boldsymbol{\beta}_0, \mathbf{w}_0)}{\partial \mathbf{w}} \mathbf{B}(T_{ik})^{-1}$$
$$E\Big[\mathbf{A}_{wj}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(T_{ik})\}\mathbf{V}_{wj}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(T_{ik})\}^{-1}$$
$$\times \frac{\partial \mathbf{S}_{wj}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(T_{ik})\}}{\partial \boldsymbol{\beta}^{\mathrm{T}}} \mid \mathbf{O}_i\Big]\Big)^{\mathrm{T}} - \Big(\mathbf{B}(T_{ik})^{-1} E\Big[\mathbf{A}_{wj}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(T_{ik})\}$$
$$\times \mathbf{V}_{wj}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(T_{ik})\}^{-1} \frac{\partial \mathbf{S}_{wj}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(T_{ik})\}}{\partial \boldsymbol{\beta}^{\mathrm{T}}} \mid \mathbf{O}_i\Big]\Big)^{\mathrm{T}} \mathbf{Z}_{ik} m_0'\{\mathbf{w}_0(T_{ik})^{\mathrm{T}}$$
$$\times \mathbf{Z}_{ik}\}\Big\} \mathbf{Q}_{wik}\{m_0, \mathbf{w}_0(T_{ik})\}^{\mathrm{T}} \Theta_{ik}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(T_{ik})\}\{\widehat{\mathbf{w}}(\boldsymbol{\beta}_0, T_{ik}) - \mathbf{w}_0(T_{ik})\}$$
$$\times \{1 + o_p(1)\}$$

which is equal to

$$\widehat{\boldsymbol{\kappa}}(T_{ik})\{\widehat{\mathbf{w}}(\boldsymbol{\beta}_0, T_{ik}) - \mathbf{w}_0(T_{ik})\}\{1 + o_p(1)\}$$

where $\widehat{\boldsymbol{\kappa}}$ is defined in the statement of the lemma. So we have

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \Big(\mathbf{A}^{\dagger}_{\beta i}[\boldsymbol{\beta}_0, \widehat{\boldsymbol{\lambda}}\{\boldsymbol{\beta}_0, \widehat{\mathbf{w}}(\boldsymbol{\beta}_0)\}, \widehat{\mathbf{w}}(\boldsymbol{\beta}_0, \mathbf{T}_i)]\mathbf{V}^{\dagger}_{\beta i}[\boldsymbol{\beta}_0, \widehat{\boldsymbol{\lambda}}\{\boldsymbol{\beta}_0, \widehat{\mathbf{w}}(\boldsymbol{\beta}_0)\},$$
$$\widehat{\mathbf{w}}(\boldsymbol{\beta}_0, \mathbf{T}_i)]^{-1}\mathbf{S}^{\dagger}_{\beta i}[\boldsymbol{\beta}_0, \widehat{\boldsymbol{\lambda}}\{\boldsymbol{\beta}_0, \widehat{\mathbf{w}}(\boldsymbol{\beta}_0)\}, \widehat{\mathbf{w}}(\boldsymbol{\beta}_0, \mathbf{T}_i)] - \mathbf{A}^{\dagger}_{\beta i}\{\boldsymbol{\beta}_0, \widehat{\boldsymbol{\lambda}}(\boldsymbol{\beta}_0, \mathbf{w}_0),$$
$$\mathbf{w}_0(\mathbf{T}_i)\}\mathbf{V}^{\dagger}_{\beta i}\{\boldsymbol{\beta}_0, \widehat{\boldsymbol{\lambda}}(\boldsymbol{\beta}_0, \mathbf{w}_0), \mathbf{w}_0(\mathbf{T}_i)\}^{-1}\mathbf{S}^{\dagger}_{\beta i}\{\boldsymbol{\beta}_0, \widehat{\boldsymbol{\lambda}}(\boldsymbol{\beta}_0, \mathbf{w}_0), \mathbf{w}_0(\mathbf{T}_i)\}\Big)$$
$$= \frac{1}{\sqrt{n}}\Big(\sum_{i=1}^{n} \mathbf{A}^{\dagger}_{\beta i}\{\boldsymbol{\beta}_0, \widehat{\boldsymbol{\lambda}}(\boldsymbol{\beta}_0, \mathbf{w}_0), \mathbf{w}_0(\mathbf{T}_i)\}\mathbf{V}^{\dagger}_{\beta i}\{\boldsymbol{\beta}_0, \widehat{\boldsymbol{\lambda}}(\boldsymbol{\beta}_0, \mathbf{w}_0), \mathbf{w}_0(\mathbf{T}_i)\}^{-1}$$

$$
\begin{aligned}
&\times\widehat{\mathcal{K}}(\mathbf{T}_i)\{\widehat{\mathbf{w}}(\boldsymbol{\beta}_0,\mathbf{T}_i)-\mathbf{w}_0(\mathbf{T}_i)\}\Big)\{1+o_p(1)\}\\
= \quad & \frac{1}{n^{3/2}}\sum_{i=1}^n\sum_{j=1}^n\Big(\mathbf{A}_{\beta i}^{\dagger}\{\boldsymbol{\beta}_0,\widehat{\boldsymbol{\lambda}}(\boldsymbol{\beta}_0,\mathbf{w}_0),\mathbf{w}_0(\mathbf{T}_i)\}\mathbf{V}_{\beta i}^{\dagger}\{\boldsymbol{\beta}_0,\widehat{\boldsymbol{\lambda}}(\boldsymbol{\beta}_0,\mathbf{w}_0),\\
&\mathbf{w}_0(\mathbf{T}_i)\}^{-1}\widehat{\mathcal{K}}(\mathbf{T}_i)\{\mathbf{B}(\mathbf{T}_i)\mathbf{f}_{\mathbf{T}}(\mathbf{T}_i)\}^{-1}\Big[\mathbf{A}_{wj}\{\boldsymbol{\beta}_0,m_0,\mathbf{w}_0(\mathbf{T}_i)\}\mathbf{V}_{wj}\{\boldsymbol{\beta}_0,m_0,\\
&\mathbf{w}_0(\mathbf{T}_i)\}^{-1}\mathbf{K}_h(\mathbf{T}_j-\mathbf{T}_i)\mathbf{S}_{wj}\{\boldsymbol{\beta}_0,m_0,\mathbf{w}_0(\mathbf{T}_\imath)\}\Big]\Big)\{1+o_p(1)\}\\
= \quad & -\Big[\frac{1}{\sqrt{n}}\sum_{i=1}^n E\{\mathbf{u}(\mathbf{O}_i,\mathbf{O}_j)\mid\mathbf{O}_i\}+\frac{1}{\sqrt{n}}\sum_{j=1}^n E\{\mathbf{u}(\mathbf{O}_i,\mathbf{O}_j)\mid\mathbf{O}_j\}\\
&-\sqrt{n}E\{\mathbf{u}(\mathbf{O}_i,\mathbf{O}_j)\}\Big]\{1+o_p(1)\}
\end{aligned}
$$

where $\mathbf{u}(\mathbf{O}_i,\mathbf{O}_j)$ is

$$
\begin{aligned}
&\Big(\mathbf{A}_{\beta i}^{\dagger}\{\boldsymbol{\beta}_0,\widehat{\boldsymbol{\lambda}}(\boldsymbol{\beta}_0,\mathbf{w}_0),\mathbf{w}_0(\mathbf{T}_i)\}\mathbf{V}_{\beta i}^{\dagger}\{\boldsymbol{\beta}_0,\widehat{\boldsymbol{\lambda}}(\boldsymbol{\beta}_0,\mathbf{w}_0),\mathbf{w}_0(\mathbf{T}_i)\}^{-1}\widehat{\mathcal{K}}(\mathbf{T}_i)\\
&\times\{\mathbf{B}(\mathbf{T}_i)\mathbf{f}_T(\mathbf{T}_i)\}^{-1}\Big[\mathbf{A}_{wj}\{\boldsymbol{\beta}_0,m_0,\mathbf{w}_0(\mathbf{T}_i)\}\mathbf{V}_{wj}\{\boldsymbol{\beta}_0,m_0,\mathbf{w}_0(\mathbf{T}_i)\}^{-1}\\
&\mathbf{K}_h(\mathbf{T}_j-\mathbf{T}_i)\mathbf{S}_{wj}\{\boldsymbol{\beta}_0,m_0,\mathbf{w}_0(\mathbf{T}_\imath)\}\Big]\Big).
\end{aligned}
$$

Applying the standard kernel smoothing technique, we have

$$
\begin{aligned}
&\frac{1}{\sqrt{n}}\sum_{j=1}^n E\{\mathbf{u}(\mathbf{O}_i,\mathbf{O}_j)\mid\mathbf{O}_j\}\\
= \quad & \frac{1}{\sqrt{n}}\sum_{j=1}^n E\Big(\mathbf{A}_{\beta i}^{\dagger}\{\boldsymbol{\beta}_0,\widehat{\boldsymbol{\lambda}}(\boldsymbol{\beta}_0,\mathbf{w}_0),\mathbf{w}_0(\mathbf{T}_i)\}\mathbf{V}_{\beta i}^{\dagger}\{\boldsymbol{\beta}_0,\widehat{\boldsymbol{\lambda}}(\boldsymbol{\beta}_0,\mathbf{w}_0),\\
&\mathbf{w}_0(\mathbf{T}_i)\}^{-1}\widehat{\mathcal{K}}(\mathbf{T}_i)\{\mathbf{B}(\mathbf{T}_i)\mathbf{f}_{\mathbf{T}}(\mathbf{T}_i)\}^{-1}\Big[\mathbf{A}_{wj}\{\boldsymbol{\beta}_0,m_0,\mathbf{w}_0(\mathbf{T}_i)\}\\
&\times\mathbf{V}_{wj}\{\boldsymbol{\beta}_0,m_0,\mathbf{w}_0(\mathbf{T}_i)\}^{-1}\mathbf{K}_h(\mathbf{T}_j-\mathbf{T}_i)\mathbf{S}_{wj}\{\boldsymbol{\beta}_0,m_0,\mathbf{w}_0(\mathbf{T}_i)\}\Big]\Big|\mathbf{O}_j\Big)
\end{aligned}
$$

$$
\begin{aligned}
= \quad & \frac{1}{\sqrt{n}}\sum_{j=1}^n E\Big(\mathbf{A}_{\beta i}^{\dagger}\{\boldsymbol{\beta}_0,\widehat{\boldsymbol{\lambda}}(\boldsymbol{\beta}_0,\mathbf{w}_0),\mathbf{w}_0(\mathbf{T}_j)\}\mathbf{V}_{\beta i}^{\dagger}\{\boldsymbol{\beta}_0,\widehat{\boldsymbol{\lambda}}(\boldsymbol{\beta}_0,\mathbf{w}_0),\\
&\mathbf{w}_0(\mathbf{T}_i)\}^{-1}\mathbf{A}_{wj}\{\boldsymbol{\beta}_0,m_0,\mathbf{w}_0(\mathbf{T}_j)\}\mathbf{V}_{wj}\{\boldsymbol{\beta}_0,m_0,\mathbf{w}_0(\mathbf{T}_j)\}^{-1}
\end{aligned}
$$

$$\times\mathbf{S}_{wj}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_j)\}\Big] + O_p(n^{1/2}h^2),$$

and

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} E\{\mathbf{u}(\mathbf{O}_i, \mathbf{O}_j) \mid \mathbf{O}_i\} = O_p(n^{1/2}h^2), \sqrt{n}E\{\mathbf{u}(\mathbf{O}_i, \mathbf{O}_j)\} = O(n^{1/2}h^2).$$

Therefore, we have

$$\begin{aligned}
&\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\Big(\mathbf{A}^{\dagger}_{\beta i}[\boldsymbol{\beta}_0, \widehat{\boldsymbol{\lambda}}\{\boldsymbol{\beta}_0, \widehat{\mathbf{w}}(\boldsymbol{\beta}_0)\}, \widehat{\mathbf{w}}(\boldsymbol{\beta}_0, \mathbf{T}_i)]\mathbf{V}^{\dagger}_{\beta i}[\boldsymbol{\beta}_0, \widehat{\boldsymbol{\lambda}}\{\boldsymbol{\beta}_0, \widehat{\mathbf{w}}(\boldsymbol{\beta}_0)\},\\
&\widehat{\mathbf{w}}(\boldsymbol{\beta}_0, \mathbf{T}_i)]^{-1}\mathbf{S}^{\dagger}_{\beta i}[\boldsymbol{\beta}_0, \widehat{\boldsymbol{\lambda}}\{\boldsymbol{\beta}_0, \widehat{\mathbf{w}}(\boldsymbol{\beta}_0)\}, \widehat{\mathbf{w}}(\boldsymbol{\beta}_0, \mathbf{T}_i)] - \mathbf{A}^{\dagger}_{\beta i}\{\boldsymbol{\beta}_0, \widehat{\boldsymbol{\lambda}}(\boldsymbol{\beta}_0, \mathbf{w}_0),\\
&\mathbf{w}_0(\mathbf{T}_i)\}\mathbf{V}^{\dagger}_{\beta i}\{\boldsymbol{\beta}_0, \widehat{\boldsymbol{\lambda}}(\boldsymbol{\beta}_0, \mathbf{w}_0), \mathbf{w}_0(\mathbf{T}_i)\}^{-1}\mathbf{S}^{\dagger}_{\beta i}\{\boldsymbol{\beta}_0, \widehat{\boldsymbol{\lambda}}(\boldsymbol{\beta}_0, \mathbf{w}_0), \mathbf{w}_0(\mathbf{T}_i)\}\Big)\\
= \; & -\frac{1}{\sqrt{n}}\sum_{j=1}^{n} E\Big(\mathbf{A}^{\dagger}_{\beta i}\{\boldsymbol{\beta}_0, \boldsymbol{\lambda}_0, \mathbf{w}_0(\mathbf{T}_j)\}\mathbf{V}^{\dagger}_{\beta i}\{\boldsymbol{\beta}_0, \boldsymbol{\lambda}_0, \mathbf{w}_0(\mathbf{T}_i)\}^{-1}\widehat{\mathcal{K}}(\mathbf{T}_j)\Big|\mathbf{O}_j\Big)\\
&\times\mathbf{B}(\mathbf{T}_j)^{-1}\Big[\mathbf{A}_{wj}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_j)\}\mathbf{V}_{wj}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_j)\}^{-1}\\
&\times\mathbf{S}_{wj}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_j)\}\Big] + o_p(1),
\end{aligned}$$

by the assumptions that $nh^4 \to 0$. This proves the result. $\square$

Lemma 9.

$$\begin{aligned}
&\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\Big[\mathbf{A}^{\dagger}_{\beta i}\{\boldsymbol{\beta}_0, \widehat{\boldsymbol{\lambda}}(\boldsymbol{\beta}_0, \mathbf{w}_0), \mathbf{w}_0(\mathbf{T}_i)\}\mathbf{V}^{\dagger}_{\beta i}\{\boldsymbol{\beta}_0, \widehat{\boldsymbol{\lambda}}(\boldsymbol{\beta}_0, \mathbf{w}_0), \mathbf{w}_0(\mathbf{T}_i)\}^{-1}\\
&\times\mathbf{S}^{\dagger}_{\beta i}\{\boldsymbol{\beta}_0, \widehat{\boldsymbol{\lambda}}(\boldsymbol{\beta}_0, \mathbf{w}_0), \mathbf{w}_0(\mathbf{T}_i)\} - \mathbf{A}^{\dagger}_{\beta i}\{\boldsymbol{\beta}_0, \boldsymbol{\lambda}_0, \mathbf{w}_0(\mathbf{T}_i)\}\mathbf{V}^{\dagger}_{\beta i}\{\boldsymbol{\beta}_0, \boldsymbol{\lambda}_0,\\
&\mathbf{w}_0(\mathbf{T}_i)\}^{-1}\mathbf{S}^{\dagger}_{\beta i}\{\boldsymbol{\beta}_0, \boldsymbol{\lambda}_0, \mathbf{w}_0(\mathbf{T}_i)\}\Big]\\
= \; & -\frac{1}{\sqrt{n}}\mathbf{G}^{\dagger}(\boldsymbol{\lambda}_0)\mathbf{V}^{-1}\sum_{i=1}^{n}\widetilde{\mathbf{Q}}_{\lambda i}\{\mathbf{w}_0(\mathbf{T}_i)\}^{\mathrm{T}}\boldsymbol{\Theta}_i\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_i)\}\boldsymbol{\Omega}_i^{-1}\\
&\times[\mathbf{D}_i - \mathbf{H}_i\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_i)\}]\{1 + o_p(1)\}\\
= \; & O_p(1),
\end{aligned}$$

where

$$\begin{aligned}
\mathbf{G}^{\dagger}(\boldsymbol{\lambda}_0) \;=\; & E[\partial\mathbf{A}^{\dagger}_{\beta j}\{\boldsymbol{\beta}_0, \boldsymbol{\lambda}_0, \mathbf{w}_0(\mathbf{T}_j)\}\mathbf{V}^{\dagger}_{\beta j}\{\boldsymbol{\beta}_0, \boldsymbol{\lambda}_0, \mathbf{w}_0(\mathbf{T}_j)\}^{-1}\\
&\times\mathbf{S}^{\dagger}_{\beta j}\{\boldsymbol{\beta}_0, \boldsymbol{\lambda}_0, \mathbf{w}_0(\mathbf{T}_j)\}/\partial\boldsymbol{\lambda}^{\mathrm{T}}],
\end{aligned}$$

and $\mathbf{V}$ *is defined in Theorem 3.*

Proof:

$$
\begin{aligned}
& \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \Big[\mathbf{A}_{\beta i}^{\dagger}\{\boldsymbol{\beta}_0, \widehat{\boldsymbol{\lambda}}(\boldsymbol{\beta}_0, \mathbf{w}_0), \mathbf{w}_0(\mathbf{T}_i)\} \mathbf{V}_{\beta i}^{\dagger}\{\boldsymbol{\beta}_0, \widehat{\boldsymbol{\lambda}}(\boldsymbol{\beta}_0, \mathbf{w}_0), \\
& \mathbf{w}_0(\mathbf{T}_i)\}^{-1} \mathbf{S}_{\beta i}^{\dagger}\{\boldsymbol{\beta}_0, \widehat{\boldsymbol{\lambda}}(\boldsymbol{\beta}_0, \mathbf{w}_0), \mathbf{w}_0(\mathbf{T}_i)\} - \mathbf{A}_{\beta i}^{\dagger}\{\boldsymbol{\beta}_0, \boldsymbol{\lambda}_0, \mathbf{w}_0(\mathbf{T}_i)\} \\
& \times \mathbf{V}_{\beta i}^{\dagger}\{\boldsymbol{\beta}_0, \boldsymbol{\lambda}_0, \mathbf{w}_0(\mathbf{T}_i)\}^{-1} \mathbf{S}_{\beta i}^{\dagger}\{\boldsymbol{\beta}_0, \boldsymbol{\lambda}_0, \mathbf{w}_0(\mathbf{T}_i)\} \Big] \\
= \; & \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \mathbf{A}_{\beta j}^{\dagger}\{\boldsymbol{\beta}_0, \boldsymbol{\lambda}_0, \mathbf{w}_0(\mathbf{T}_j)\} \mathbf{V}_{\beta j}^{\dagger}\{\boldsymbol{\beta}_0, \boldsymbol{\lambda}_0, \mathbf{w}_0(\mathbf{T}_j)\}^{-1} \\
& \times \frac{\partial \mathbf{S}_{\beta j}^{\dagger}\{\boldsymbol{\beta}_0, \boldsymbol{\lambda}_0, \mathbf{w}_0(\mathbf{T}_j)\}}{\partial \boldsymbol{\lambda}^{\mathrm{T}}} \{\widehat{\boldsymbol{\lambda}}(\boldsymbol{\beta}_0, \mathbf{w}_0) - \boldsymbol{\lambda}_0\}\{1 + o_p(1)\} \\
= \; & -\frac{1}{\sqrt{n}} n^{-1} \sum_{j=1}^{n} \mathbf{A}_{\beta j}^{\dagger}\{\boldsymbol{\beta}_0, \boldsymbol{\lambda}_0, \mathbf{w}_0(\mathbf{T}_j)\} \mathbf{V}_{\beta j}^{\dagger}\{\boldsymbol{\beta}_0, \boldsymbol{\lambda}_0, \mathbf{w}_0(\mathbf{T}_j)\}^{-1} \\
& \times \frac{\partial \mathbf{S}_{\beta j}^{\dagger}\{\boldsymbol{\beta}_0, \boldsymbol{\lambda}_0, \mathbf{w}_0(\mathbf{T}_j)\}}{\partial \boldsymbol{\lambda}^{\mathrm{T}}} \mathbf{V}^{-1} \sum_{i=1}^{n} \widetilde{\mathbf{Q}}_{\lambda i}\{\mathbf{w}_0(\mathbf{T}_i)\}^{\mathrm{T}} \boldsymbol{\Theta}_i\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_i)\} \\
& \boldsymbol{\Omega}_i^{-1}[\mathbf{D}_i - \mathbf{H}_i\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_i)\}] \\
& \times \{1 + o_p(1)\} \\
= \; & -\frac{1}{\sqrt{n}} \Bigg(E\Big[\partial \mathbf{A}_{\beta j}^{\dagger}\{\boldsymbol{\beta}_0, \boldsymbol{\lambda}_0, \mathbf{w}_0(\mathbf{T}_j)\} \mathbf{V}_{\beta j}^{\dagger}\{\boldsymbol{\beta}_0, \boldsymbol{\lambda}_0, \mathbf{w}_0(\mathbf{T}_j)\}^{-1} \\
& \times \mathbf{S}_{\beta j}^{\dagger}\{\boldsymbol{\beta}_0, \boldsymbol{\lambda}_0, \mathbf{w}_0(\mathbf{T}_j)\} / \partial \boldsymbol{\lambda}^{\mathrm{T}} \Big] \{1 + o_p(1)\} \Bigg) \\
& \times \mathbf{V}^{-1} \sum_{i=1}^{n} \widetilde{\mathbf{Q}}_{\lambda i}\{\mathbf{w}_0(\mathbf{T}_i)\}^{\mathrm{T}} \boldsymbol{\Theta}_i\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_i)\} \boldsymbol{\Omega}_i^{-1}[\mathbf{D}_i \\
& - \mathbf{H}_i\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_i)\}]\{1 + o_p(1)\} \\
= \; & -\frac{1}{\sqrt{n}} \mathbf{G}^{\dagger}(\boldsymbol{\lambda}_0) \mathbf{V}^{-1} \sum_{i=1}^{n} \widetilde{\mathbf{Q}}_{\lambda i}\{\mathbf{w}_0(\mathbf{T}_i)\}^{\mathrm{T}} \boldsymbol{\Theta}_i\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_i)\} \boldsymbol{\Omega}_i^{-1}[\mathbf{D}_i \\
& - \mathbf{H}_i\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_i)\}]\{1 + o_p(1)\},
\end{aligned}
$$

where the second equality is due to Lemma 2. We show that the above quantity has the order of $O_p(1)$. Note that

$$
\mathbf{A}_{\beta j}^{\dagger}\{\boldsymbol{\beta}_0, \boldsymbol{\lambda}_0, \mathbf{w}_0(\mathbf{T}_j)\} \mathbf{V}_{\beta j}^{\dagger}\{\boldsymbol{\beta}_0, \boldsymbol{\lambda}_0, \mathbf{w}_0(\mathbf{T}_j)\}^{-1} \mathbf{S}_{\beta j}^{\dagger}\{\boldsymbol{\beta}_0, \boldsymbol{\lambda}_0, \mathbf{w}_0(\mathbf{T}_j)\}
$$

has the form of $\mathbf{C}_j \left[\mathbf{D}_j - \mathbf{H}_j\{\boldsymbol{\beta}_0, \boldsymbol{\lambda}_0, \mathbf{w}_0(\mathbf{T}_j)\}\right]_{d_\beta M_i}$, where $[\mathbf{D}_j - \mathbf{H}_j\{\boldsymbol{\beta}_0, \boldsymbol{\lambda}_0, \mathbf{w}_0(\mathbf{T}_j)\}]_{d_\beta M_j}$ is a $d_\beta M_j \times 1$ vector, whose the kth block is d_β replications of $D_{jk} - H_{jk}\{\boldsymbol{\beta}_0, \boldsymbol{\lambda}_0, \mathbf{w}_0(T_{jk})\}$, and we collect the terms in front into $\mathbf{C}'_j$. Thus,

$$
\begin{aligned}
(\text{S.37}) \qquad & \mathbf{G}^\dagger(\boldsymbol{\lambda}_0) \\
= \; & E[\partial \mathbf{A}^\dagger_{\beta j}\{\boldsymbol{\beta}_0, \boldsymbol{\lambda}_0, \mathbf{w}_0(\mathbf{T}_j)\}\mathbf{V}^\dagger_{\beta j}\{\boldsymbol{\beta}_0, \boldsymbol{\lambda}_0, \mathbf{w}_0(\mathbf{T}_j)\}^{-1} \\
& \times \mathbf{S}^\dagger_{\beta j}\{\boldsymbol{\beta}_0, \boldsymbol{\lambda}_0, \mathbf{w}_0(\mathbf{T}_j)\}/\partial \boldsymbol{\lambda}^{\mathrm{T}}] \\
= \; & E\left(\mathbf{C}'_j \frac{\partial \left[\mathbf{D}_j - \mathbf{H}_j\{\boldsymbol{\beta}_0, \boldsymbol{\lambda}_0, \mathbf{w}_0(\mathbf{T}_j)\}\right]_{d_\beta M_i}}{\partial \boldsymbol{\lambda}^{\mathrm{T}}}\right) \\
& + E\left(\frac{\partial \mathbf{C}'_j}{\partial \boldsymbol{\lambda}^{\mathrm{T}}} \left[\mathbf{D}_j - \mathbf{H}_j\{\boldsymbol{\beta}_0, \boldsymbol{\lambda}_0, \mathbf{w}_0(\mathbf{T}_j)\}\right]_{d_\beta M_i}\right).
\end{aligned}
$$

Consider first the second term in (S.37),

$$
\begin{aligned}
& \left\| E\left(\frac{\partial \mathbf{C}'_j}{\partial \boldsymbol{\lambda}^{\mathrm{T}}} \left[\mathbf{D}_j - \mathbf{H}_j\{\boldsymbol{\beta}_0, \boldsymbol{\lambda}_0, \mathbf{w}_0(\mathbf{T}_j)\}\right]_{d_\beta M_i}\right) \right\|_\infty \\
= \; & \left\| E\left(\frac{\partial \mathbf{C}'_j}{\partial \boldsymbol{\lambda}^{\mathrm{T}}} \left[\mathbf{D}_j - \mathbf{H}_j\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_j)\}\right]_{d_\beta M_i}\right) \right\|_\infty \\
& + \sup_{0 \le u \le 1} |\widetilde{m}(u) - m_0(u)| \\
= \; & o(h_b^q)
\end{aligned}
$$

by **Fact 1**. Now for the first term in (S.37), because

$$
\begin{aligned}
& \left\| \frac{\partial \left[\mathbf{D}_j - \mathbf{H}_j\{\boldsymbol{\beta}_0, \boldsymbol{\lambda}_0, \mathbf{w}_0(\mathbf{T}_j)\}\right]_{d_\beta M_i}}{\partial \boldsymbol{\lambda}^{\mathrm{T}}} \right\|_\infty \\
= \; & \max \sum_{p=1}^{d_\lambda} B_{rp}(\mathbf{w}_0(T_{ik})^{\mathrm{T}} \mathbf{Z}_{ik}) \\
= \; & 1
\end{aligned}
$$

the maximum row sum, we have

$$
\begin{aligned}
& \left\| E\left(\mathbf{C}'_j \frac{\partial \left[\mathbf{D}_j - \mathbf{H}_j\{\boldsymbol{\beta}_0, \boldsymbol{\lambda}_0, \mathbf{w}_0(\mathbf{T}_j)\}\right]_{d_\beta M_i}}{\partial \boldsymbol{\lambda}^{\mathrm{T}}}\right) \right\|_\infty \\
\le \; & E\left(\|\mathbf{C}'_j\|_\infty \left\| \frac{\partial \left[\mathbf{D}_j - \mathbf{H}_j\{\boldsymbol{\beta}_0, \boldsymbol{\lambda}_0, \mathbf{w}_0(\mathbf{T}_j)\}\right]_{d_\beta M_i}}{\partial \boldsymbol{\lambda}^{\mathrm{T}}} \right\|_\infty\right) \\
= \; & O(1)
\end{aligned}
$$

Therefore, we can write (S.37) as

$$
\mathbf{G}^\dagger(\boldsymbol{\lambda}_0) = E\left(\mathbf{C}'_j \frac{\partial \left[\mathbf{D}_j - \mathbf{H}_j\{\boldsymbol{\beta}_0, \boldsymbol{\lambda}_0, \mathbf{w}_0(\mathbf{T}_j)\}\right]_{d_\beta M_i}}{\partial \boldsymbol{\lambda}^{\mathrm{T}}}\right)\{1 + o_p(1)\}.
$$

Moreover, remember that

$$\widetilde{\mathbf{Q}}_\lambda\{\mathbf{w}_0(\mathbf{T}_j)\} = \boldsymbol{\Theta}_j^{-1}\frac{\partial[\mathbf{D}_j - \mathbf{H}_j\{\boldsymbol{\beta}_0, \boldsymbol{\lambda}_0, \mathbf{w}_0(\mathbf{T}_j)\}]}{\partial\boldsymbol{\lambda}^{\mathrm{T}}},$$

where $[\mathbf{D}_j - \mathbf{H}_j\{\boldsymbol{\beta}_0, \boldsymbol{\lambda}_0, \mathbf{w}_0(\mathbf{T}_j)\}]$ is a $M_i \times 1$ vector. Since d_β is finite,

$$\left\|\frac{\partial\left[\mathbf{D}_j - \mathbf{H}_j\{\boldsymbol{\beta}_0, \boldsymbol{\lambda}_0, \mathbf{w}_0(\mathbf{T}_j)\}\right]_{d_\beta M_j}}{\partial\boldsymbol{\lambda}^{\mathrm{T}}}\right\|_2 \asymp \|\widetilde{\mathbf{Q}}_\lambda\{\mathbf{w}_0(\mathbf{T}_j)\}\|_2,$$

Therefore,

$$\|\mathbf{G}^\dagger(\boldsymbol{\lambda}_0)\|_2 \le E[\|\mathbf{C}_j\widetilde{\mathbf{Q}}_\lambda\{\mathbf{w}_0(\mathbf{T}_j)\}\|_2]$$

for some bounded random matrix $\mathbf{C}_j$. Since in Corollary 1, we have shown that $E[\|\mathbf{C}_j\widetilde{\mathbf{Q}}_\lambda\{\mathbf{w}_0(\mathbf{T}_j)\}\|_2^2] = O(h_b)$, it implies $\|\mathbf{G}^\dagger(\boldsymbol{\lambda}_0)\|_2 = O(h_b^{1/2})$. Further by (S.12) and (S.6), we have

$$\begin{aligned}
&\left\|\frac{1}{\sqrt{n}}\sum_{i=1}^n \widetilde{\mathbf{Q}}_{\lambda i}\{\mathbf{w}_0(\mathbf{T}_i)\}^{\mathrm{T}}\boldsymbol{\Theta}_i\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_i)\}\boldsymbol{\Omega}_i^{-1}\right.\\
&\left.\times[\mathbf{D}_i - \mathbf{H}_i\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_i)\}]\right\|_2\\
= \;& O_p(h_b^{1/2}),
\end{aligned}$$

and $\|\mathbf{V}^{-1}\|_2 = O(h_b^{-1})$. This proves the results. □

Lemma 10.

$$-\boldsymbol{\gamma}\{\mathbf{w}_0(T_{ik})^{\mathrm{T}}\mathbf{Z}_{ik}\} - \left(\frac{\partial\widehat{\boldsymbol{\lambda}}(\boldsymbol{\beta}_0, \mathbf{w}_0)}{\partial\mathbf{w}}\mathcal{B}\right)^{\mathrm{T}}\mathbf{Q}_{\lambda ik}\{\mathbf{w}_0(T_{ik})\} = o_p(1),$$

where recall that

$$\begin{aligned}
\mathcal{B} \;=\; & \left\{\left(\mathbf{B}(T_{ik})^{-1}E\left[\mathbf{A}_{wj}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(T_{ik})\}\mathbf{V}_{wj}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(T_{ik})\}^{-1}\right.\right.\right.\\
&\left.\left.\left.\times\frac{\partial\mathbf{S}_{wj}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(T_{ik})\}}{\partial\boldsymbol{\beta}^{\mathrm{T}}}\right)\middle|\mathbf{O}_i\right]\right)^{\mathrm{T}}, k = 1, \ldots, M_i, i = 1, \ldots, n\right\}^{\mathrm{T}}
\end{aligned}$$

a $\sum_{i=1}^n d_w M_i$ *vector, and* $\boldsymbol{\gamma}$ *are defined in* **Notation in Step 3**.

Proof: Similar to (S.23) we have

$$\frac{\partial\widehat{\boldsymbol{\lambda}}(\boldsymbol{\beta}_0, \mathbf{w}_0)}{\partial\mathbf{w}}\mathcal{B}$$

$$= -\Big(n^{-1}\sum_{i=1}^{n}[\widetilde{\mathbf{Q}}_{\lambda i}\{\mathbf{w}_0(\mathbf{T}_i)\}^{\mathrm{T}}\boldsymbol{\Theta}_i\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_i)\}\boldsymbol{\Omega}_i^{-1}\boldsymbol{\Theta}_i\{\boldsymbol{\beta}_0, m_0,$$

$$\mathbf{w}_0(\mathbf{T}_i)\}\widetilde{\mathbf{Q}}_{\lambda i}\{\mathbf{w}_0(\mathbf{T}_i)\}]\Big)^{-1} n^{-1}\sum_{i=1}^{n}\widetilde{\mathbf{Q}}_{\lambda i}\{\mathbf{w}_0(\mathbf{T}_i)\}^{\mathrm{T}}\boldsymbol{\Theta}_i\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_i)\}$$

$$\boldsymbol{\Omega}_i^{-1}\boldsymbol{\Theta}_i\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_i)\}\widetilde{\mathbf{Q}}_{wi}\Big(m_0, \mathbf{B}(\mathbf{T}_i)^{-1}E\Big[\mathbf{A}_{wj}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_i)\}$$

$$\times\mathbf{V}_{wj}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_i)\}^{-1}\frac{\partial\mathbf{S}_{wj}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_i)\}}{\partial\boldsymbol{\beta}^{\mathrm{T}}}\Big)\Big|\mathbf{O}_i\Big]\Big)$$

$$\times\{1+o_p(1)\}$$

where recall that

$$\widetilde{\mathbf{Q}}_{wi}\Big(m_0, \mathbf{B}(\mathbf{T}_i)^{-1}E\Big[\mathbf{A}_{wj}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_i)\}\mathbf{V}_{wj}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_i)\}^{-1}$$

$$\times\frac{\partial\mathbf{S}_{wj}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_i)\}}{\partial\boldsymbol{\beta}^{\mathrm{T}}}\Big)\Big|\mathbf{O}_i\Big]\Big)$$

is a $M_i \times \boldsymbol{\beta}$ matrix, with the k th row as

$$m_0'\{\mathbf{w}_0(T_{ik})\}\mathbf{Z}_{ik}^{\mathrm{T}}\mathbf{B}(T_{ik})^{-1}E\Big[\mathbf{A}_{wj}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(T_{ik})\}\mathbf{V}_{wj}\{\boldsymbol{\beta}_0, m_0,$$

$$\mathbf{w}_0(T_{ik})\}^{-1}\frac{\partial\mathbf{S}_{wj}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(T_{ik})\}}{\partial\boldsymbol{\beta}^{\mathrm{T}}}\Big)\Big|\mathbf{O}_i\Big].$$

The result is proved by using the same argument as those lead to Lemma 3. □

<u>Lemma</u> 11.

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\mathbf{A}_{\beta i}^{\dagger}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_i)\}\mathbf{V}_{\beta i}^{\dagger}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_i)\}^{-1}$$

$$\times\mathbf{S}_{\beta i}^{\dagger}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_i)\}$$

$$= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\mathbf{A}_{\beta i}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_i)\}\mathbf{V}_{\beta i}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_i)\}^{-1}$$

$$\times\mathbf{S}_{\beta i}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_i)\}\{1+o_p(1)\},$$

$$E\Big(\mathbf{A}_{\beta i}^{\dagger}\{\boldsymbol{\beta}_0, \boldsymbol{\lambda}_0, \mathbf{w}_0(\mathbf{T}_j)\}\mathbf{V}_{\beta i}^{\dagger}\{\boldsymbol{\beta}_0, \boldsymbol{\lambda}_0, \mathbf{w}_0(\mathbf{T}_i)\}^{-1}\widehat{\mathcal{K}}(\mathbf{T}_j)\Big|\mathbf{O}_j\Big)\mathbf{B}(\mathbf{T}_j)$$

$$\times\Big[\mathbf{A}_{wj}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_j)\}\mathbf{V}_{wj}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_j)\}^{-1}$$

$$
\begin{aligned}
&\times \mathbf{S}_{wj}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_j)\}\Big] \\
= \;& E\bigg(\mathbf{A}_{\beta i}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_j)\}\mathbf{V}_{\beta i}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_i)\}^{-1}\mathcal{K}(\mathbf{T}_j)\bigg|\mathbf{O}_j\bigg)\mathbf{B}(\mathbf{T}_j) \\
&\times \Big[\mathbf{A}_{wj}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_j)\}\mathbf{V}_{wj}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_j)\}^{-1} \\
&\times \mathbf{S}_{wj}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_j)\}\Big]\{1 + o_p(1)\}, \\
&\mathbf{G}^{\dagger}(m) = \mathbf{G}(m_0)\{1 + o_p(1)\}, \;\; \textit{and } \mathbf{F}^{\dagger}(m) = \mathbf{F}(m_0)\{1 + o_p(1)\}.
\end{aligned}
$$

where $\mathcal{K}$ is defined in Theorem 2.

$$
\begin{aligned}
\mathbf{G}^{\dagger}(m) \;\; = \;\; & E\Big[\mathbf{A}^{\dagger}_{\beta i}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_i)\}\mathbf{V}^{\dagger}_{\beta i}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_i)\}^{-1}\mathbf{C}^{\dagger}_i \\
& \times \boldsymbol{\Theta}^*_i\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_i)\}\mathbf{Q}^*_{\lambda i}\{\mathbf{w}_0(\mathbf{T}_i)\}\Big].
\end{aligned}
$$

where $\mathbf{C}^{\dagger}_i$ is a $d_\beta M_i \times d_\beta M_i$ with the kth block has the form of

$$
\begin{aligned}
&\bigg\{\mathbf{Q}_{\beta ik} + \bigg\{\frac{\partial\widehat{\boldsymbol{\lambda}}(\boldsymbol{\beta}_0, \mathbf{w}_0)}{\partial\boldsymbol{\beta}^{\mathrm{T}}} + \frac{\partial\widehat{\boldsymbol{\lambda}}\{\boldsymbol{\beta}, \widehat{\mathbf{w}}(\boldsymbol{\beta})\}}{\widehat{\mathbf{w}}(\boldsymbol{\beta})}\mathcal{B}\bigg\}^{\mathrm{T}}\mathbf{Q}_{\lambda ik} - \bigg(\mathbf{B}(T_{ik})^{-1} \\
&\times E\Big[\mathbf{A}_{wj}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_j)\}\mathbf{V}_{wj}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_j)\}^{-1} \\
&\times \frac{\partial\mathbf{S}_{wj}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_j)\}}{\partial\boldsymbol{\beta}^{\mathrm{T}}} \mid \mathbf{O}_i\Big]\bigg)^{\mathrm{T}}\mathbf{Z}_{ik}m_0'\{\mathbf{w}_0(T_{ik})^{\mathrm{T}}\mathbf{Z}_{ik}\}\bigg\},
\end{aligned}
$$

$\boldsymbol{\Theta}^*_i\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_i)\}$, $\mathbf{Q}^*_{\lambda i}\{\mathbf{w}_0(\mathbf{T}_i)\}$ *are defined in Theorem 2, $\mathcal{B}$ is defined in* **Notations in Step 3** .

$$
\begin{aligned}
& \mathbf{F}^{\dagger}(m) \\
= \;\; & -E\bigg\{\mathbf{A}^{\dagger}_{\beta i}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_i)\}\mathbf{V}^{\dagger}_{\beta i}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_i)\}^{-1} \\
& \times \frac{\partial\mathbf{S}^{\dagger}_{\beta i}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_i)\}}{\partial\boldsymbol{\beta}^{\mathrm{T}}}\bigg\},
\end{aligned}
$$

and $\mathbf{G}(m_0)$, $\mathbf{F}(m_0)$ *are defined in Theorem 2.*

Proof: The result follows by using Taylor expansion combine with Lemma 3, 5, and 10. □

Proof of Theorem 2: From a Taylor expansion and 8, it is easy to see that

$$
\sqrt{n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)
$$

$$
\begin{aligned}
&= -\Bigg(\frac{1}{n}\sum_{i=1}^{n}\widehat{\mathbf{A}}_{\beta i}[\boldsymbol{\beta}_0,\widehat{\boldsymbol{\lambda}}\{\boldsymbol{\beta}_0,\widehat{\mathbf{w}}(\boldsymbol{\beta}_0)\},\widehat{\mathbf{w}}(\boldsymbol{\beta}_0,\mathbf{T}_i)]\widehat{\mathbf{V}}_{\beta i}[\boldsymbol{\beta}_0,\widehat{\boldsymbol{\lambda}}\{\boldsymbol{\beta}_0,\widehat{\mathbf{w}}(\boldsymbol{\beta}_0)\},\\
&\quad \widehat{\mathbf{w}}(\boldsymbol{\beta}_0,\mathbf{T}_i)]^{-1}\frac{\partial\widehat{\mathbf{S}}_{\beta i}[\boldsymbol{\beta}_0,\widehat{\boldsymbol{\lambda}}\{\boldsymbol{\beta}_0,\widehat{\mathbf{w}}(\boldsymbol{\beta}_0)\},\widehat{\mathbf{w}}(\boldsymbol{\beta},\mathbf{T}_i)]}{\partial\boldsymbol{\beta}^{\mathrm{T}}}\Bigg)^{-1}\\
&\quad \times\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\widehat{\mathbf{A}}_{\beta i}[\boldsymbol{\beta}_0,\widehat{\boldsymbol{\lambda}}\{\boldsymbol{\beta}_0,\widehat{\mathbf{w}}(\boldsymbol{\beta}_0)\},\widehat{\mathbf{w}}(\boldsymbol{\beta}_0,\mathbf{T}_i)]\widehat{\mathbf{V}}_{\beta i}[\boldsymbol{\beta}_0,\widehat{\boldsymbol{\lambda}}\{\boldsymbol{\beta}_0,\widehat{\mathbf{w}}(\boldsymbol{\beta}_0)\},\\
&\quad \widehat{\mathbf{w}}(\boldsymbol{\beta}_0,\mathbf{T}_i)]^{-1}\widehat{\mathbf{S}}_{\beta i}[\boldsymbol{\beta}_0,\widehat{\boldsymbol{\lambda}}\{\boldsymbol{\beta}_0,\widehat{\mathbf{w}}(\boldsymbol{\beta}_0)\},\widehat{\mathbf{w}}(\boldsymbol{\beta}_0,\mathbf{T}_i)]\{1+o_p(1)\}\\
&= -\Bigg(\frac{1}{n}\sum_{i=1}^{n}\mathbf{A}^{\dagger}_{\beta i}[\boldsymbol{\beta}_0,\widehat{\boldsymbol{\lambda}}\{\boldsymbol{\beta}_0,\widehat{\mathbf{w}}(\boldsymbol{\beta}_0)\},\widehat{\mathbf{w}}(\boldsymbol{\beta}_0,\mathbf{T}_i)]\mathbf{V}^{\dagger}_{\beta i}[\boldsymbol{\beta}_0,\widehat{\boldsymbol{\lambda}}\{\boldsymbol{\beta}_0,\widehat{\mathbf{w}}(\boldsymbol{\beta}_0)\},\\
&\quad \widehat{\mathbf{w}}(\boldsymbol{\beta}_0,\mathbf{T}_i)]^{-1}\frac{\partial\mathbf{S}^{\dagger}_{\beta i}[\boldsymbol{\beta}_0,\widehat{\boldsymbol{\lambda}}\{\boldsymbol{\beta}_0,\widehat{\mathbf{w}}(\boldsymbol{\beta}_0)\},\widehat{\mathbf{w}}(\boldsymbol{\beta},\mathbf{T}_i)]}{\partial\boldsymbol{\beta}^{\mathrm{T}}}+o_p(1)\Bigg)^{-1}\\
&\quad \times\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\mathbf{A}^{\dagger}_{\beta i}[\boldsymbol{\beta}_0,\widehat{\boldsymbol{\lambda}}\{\boldsymbol{\beta}_0,\widehat{\mathbf{w}}(\boldsymbol{\beta}_0)\},\widehat{\mathbf{w}}(\boldsymbol{\beta}_0,\mathbf{T}_i)]\mathbf{V}^{\dagger}_{\beta i}[\boldsymbol{\beta}_0,\widehat{\boldsymbol{\lambda}}\{\boldsymbol{\beta}_0,\widehat{\mathbf{w}}(\boldsymbol{\beta}_0)\},\\
&\quad \widehat{\mathbf{w}}(\boldsymbol{\beta}_0,\mathbf{T}_i)]^{-1}\mathbf{S}^{\dagger}_{\beta i}[\boldsymbol{\beta}_0,\widehat{\boldsymbol{\lambda}}\{\boldsymbol{\beta}_0,\widehat{\mathbf{w}}(\boldsymbol{\beta}_0)\},\widehat{\mathbf{w}}(\boldsymbol{\beta}_0,\mathbf{T}_i)]\{1+o_p(1)\}
\end{aligned}
$$

$$
\begin{aligned}
&= [\mathbf{F}^{\dagger}\{\widehat{\boldsymbol{\lambda}}\{\boldsymbol{\beta}_0,\widehat{\mathbf{w}}(\boldsymbol{\beta}_0)\}+o_p(1)]^{-1}\\
&\quad\times\Bigg(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\mathbf{A}^{\dagger}_{\beta i}\{\boldsymbol{\beta}_0,\widehat{\boldsymbol{\lambda}}(\boldsymbol{\beta}_0,\mathbf{w}_0),\mathbf{w}_0(\mathbf{T}_i)\}\mathbf{V}^{\dagger}_{\beta i}\{\boldsymbol{\beta}_0,\widehat{\boldsymbol{\lambda}}(\boldsymbol{\beta}_0,\mathbf{w}_0),\mathbf{w}_0(\mathbf{T}_i)\}^{-1}\\
&\quad\mathbf{S}^{\dagger}_{\beta i}\{\boldsymbol{\beta}_0,\widehat{\boldsymbol{\lambda}}(\boldsymbol{\beta}_0,\mathbf{w}_0),\mathbf{w}_0(\mathbf{T}_i)\}-\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\sum_{j=1}^{n}E\Bigg(\mathbf{A}^{\dagger}_{\beta i}\{\boldsymbol{\beta}_0,m_0,\mathbf{w}_0(\mathbf{T}_j)\}\\
&\quad\times\mathbf{V}^{\dagger}_{\beta i}\{\boldsymbol{\beta}_0,m_0,\mathbf{w}_0(\mathbf{T}_i)\}^{-1}\widehat{\mathcal{K}}(\mathbf{T}_j)\Bigg|\mathbf{O}_j\Bigg)\mathbf{B}(\mathbf{T}_j)^{-1}\Bigg[\mathbf{A}_{wj}\{\boldsymbol{\beta}_0,m_0,\mathbf{w}_0(\mathbf{T}_j)\}\\
&\quad\times\mathbf{V}_{wj}\{\boldsymbol{\beta}_0,m_0,\mathbf{w}_0(\mathbf{T}_j)\}^{-1}\mathbf{S}_{wj}\{\boldsymbol{\beta}_0,m_0,\mathbf{w}_0(\mathbf{T}_j)\}\Bigg]\Bigg)\\
&= \{\mathbf{F}^{\dagger}(\boldsymbol{\lambda}_0)+o_p(1)\}^{-1}\Bigg(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\mathbf{A}^{\dagger}_{\beta i}\{\boldsymbol{\beta}_0,\boldsymbol{\lambda}_0,\mathbf{w}_0(\mathbf{T}_i)\}\mathbf{V}^{\dagger}_{\beta i}\{\boldsymbol{\beta}_0,\boldsymbol{\lambda}_0,\\
&\quad\mathbf{w}_0(\mathbf{T}_i)\}^{-1}\mathbf{S}^{\dagger}_{\beta i}\{\boldsymbol{\beta}_0,\boldsymbol{\lambda}_0,\mathbf{w}_0(\mathbf{T}_i)\}-\frac{1}{\sqrt{n}}\sum_{j=1}^{n}E\Bigg(\mathbf{A}^{\dagger}_{\beta i}\{\boldsymbol{\beta}_0,m_0,\mathbf{w}_0(\mathbf{T}_j)\}\\
&\quad\times\mathbf{V}^{\dagger}_{\beta i}\{\boldsymbol{\beta}_0,m_0,\mathbf{w}_0(\mathbf{T}_i)\}^{-1}\widehat{\mathcal{K}}(\mathbf{T}_j)\Bigg|\mathbf{O}_j\Bigg)\mathbf{B}(\mathbf{T}_j)^{-1}\Bigg[\mathbf{A}_{wj}\{\boldsymbol{\beta}_0,m_0,\mathbf{w}_0(\mathbf{T}_j)\}\\
&\quad\times\mathbf{V}_{wj}\{\boldsymbol{\beta}_0,m_0,\mathbf{w}_0(\mathbf{T}_j)\}^{-1}\mathbf{S}_{wj}\{\boldsymbol{\beta}_0,m_0,\mathbf{w}_0(\mathbf{T}_j)\}\Bigg]\Bigg)-\mathbf{G}^{\dagger}(\boldsymbol{\lambda}_0)\mathbf{V}^{-1}
\end{aligned}
$$

$$\times \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \widetilde{\mathbf{Q}}_{\lambda i}\{\mathbf{w}_0(\mathbf{T}_i)\}^{\mathrm{T}} \boldsymbol{\Theta}_i\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_i)\} \boldsymbol{\Omega}_i^{-1} [\mathbf{D}_i - \mathbf{H}_i\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_i)\}] \Bigg).$$

So we can further write

$$\begin{aligned}
&\sqrt{n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \\
= \;& \{\mathbf{F}^{\dagger}(m) + o_p(1)\}^{-1} \Bigg(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \mathbf{A}_{\beta i}^{\dagger}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_i)\} \mathbf{V}_{\beta i}^{\dagger}\{\boldsymbol{\beta}_0, m_0, \\
& \mathbf{w}_0(\mathbf{T}_i)\}^{-1} \mathbf{S}_{\beta i}^{\dagger}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_i)\} - \frac{1}{\sqrt{n}} \sum_{j=1}^{n} E\Bigg(\mathbf{A}_{\beta i}^{\dagger}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_j)\} \\
& \mathbf{V}_{\beta i}^{\dagger}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_i)\}^{-1} \widehat{\mathcal{K}}(\mathbf{T}_j) \Bigg| \mathbf{O}_j \Bigg) \mathbf{B}(\mathbf{T}_j)^{-1} \Bigg[\mathbf{A}_{wj}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_j)\} \\
& \mathbf{V}_{wj}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_j)\}^{-1} \mathbf{S}_{wj}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_j)\} \Bigg] \Bigg) - \mathbf{G}^{\dagger}(m) \mathbf{V}^{-1} \\
& \times \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \widetilde{\mathbf{Q}}_{\lambda i}\{\mathbf{w}_0(\mathbf{T}_i)\}^{\mathrm{T}} \boldsymbol{\Theta}_i\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_i)\} \boldsymbol{\Omega}_i^{-1} [\mathbf{D}_i \\
& - \mathbf{H}_i\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_i)\}] \Bigg) \\
= \;& \{\mathbf{F}(m_0) + o_p(1)\}^{-1} \Bigg(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \mathbf{A}_{\beta i}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_i)\} \mathbf{V}_{\beta i}\{\boldsymbol{\beta}_0, m_0, \\
& \mathbf{w}_0(\mathbf{T}_i)\}^{-1} \mathbf{S}_{\beta i}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_i)\} - \frac{1}{\sqrt{n}} \sum_{j=1}^{n} E\Bigg(\mathbf{A}_{\beta i}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_j)\} \\
& \times \mathbf{V}_{\beta i}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_i)\}^{-1} \mathcal{K}(\mathbf{T}_j) \Bigg| \mathbf{O}_j \Bigg) \mathbf{B}(\mathbf{T}_j)^{-1} \Bigg[\mathbf{A}_{wj}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_j)\} \\
& \times \mathbf{V}_{wj}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_j)\}^{-1} \mathbf{S}_{wj}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_j)\} \Bigg] \Bigg) - \mathbf{G}(m_0) \mathbf{V}^{-1} \\
& \times \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \widetilde{\mathbf{Q}}_{\lambda i}\{\mathbf{w}_0(\mathbf{T}_i)\}^{\mathrm{T}} \boldsymbol{\Theta}_i\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(\mathbf{T}_i)\} \boldsymbol{\Omega}_i^{-1} [\mathbf{D}_i - \mathbf{H}_i\{\boldsymbol{\beta}_0, m_0, \\
& \mathbf{w}_0(\mathbf{T}_i)\}] \Bigg),
\end{aligned}$$

where $\widehat{\mathcal{K}}$ is defined in Lemma 8, $\mathbf{G}^{\dagger}$ is defined in Lemma 9, $\mathbf{F}^{\dagger}$ is defined in Lemma 11. $\mathcal{K}$, $\mathbf{G}$, and $\mathbf{F}$ are defined in Theorem 2. The second equality is because of Lemma 6. The third equality is the result of Lemma 8. The fourth equality is the result of Lemma 9. The fifth equality is the result of **Fact 1**. The last equality the result of Lemma 11. This proves the results. □

Proof of Theorem 3: From the Lemma 2 we have

$$
\begin{aligned}
& \mathrm{var}[\mathbf{B}_r(u)^{\mathrm{T}}\{\widehat{\boldsymbol{\lambda}}(\boldsymbol{\beta}_0,\mathbf{w}_0)-\boldsymbol{\lambda}_0\}\mid \mathbf{R}] \\
= \ & \mathbf{B}_r(u)^{\mathrm{T}}\bigg(\sum_{i=1}^{n}[\widetilde{\mathbf{Q}}_{\lambda i}\{\mathbf{w}_0(\mathbf{T}_i)\}^{\mathrm{T}}\boldsymbol{\Theta}_i\{\boldsymbol{\beta}_0,m_0,\mathbf{w}_0(\mathbf{T}_i)\}\boldsymbol{\Omega}_i^{-1} \\
& \times\boldsymbol{\Theta}_i\{\boldsymbol{\beta}_0,m_0,\mathbf{w}_0(\mathbf{T}_i)\}\widetilde{\mathbf{Q}}_{\lambda i}\{\mathbf{w}_0(\mathbf{T}_i)\}]\bigg)^{-1}\sum_{i=1}^{n}\widetilde{\mathbf{Q}}_{\lambda i}\{\mathbf{w}_0(\mathbf{T}_i)\}^{\mathrm{T}} \\
& \times\boldsymbol{\Theta}_i\{\boldsymbol{\beta}_0,m_0,\mathbf{w}_0(\mathbf{T}_i)\}\boldsymbol{\Omega}_i^{-1}\boldsymbol{\Omega}_i^{*}\boldsymbol{\Omega}_i^{-1}\boldsymbol{\Theta}_i\{\boldsymbol{\beta}_0,m_0,\mathbf{w}_0(\mathbf{T}_i)\}\widetilde{\mathbf{Q}}_{\lambda i}\{\mathbf{w}_0(\mathbf{T}_i)\} \\
& \times\bigg(\sum_{i=1}^{n}[\widetilde{\mathbf{Q}}_{\lambda i}\{\mathbf{w}_0(\mathbf{T}_i)\}^{\mathrm{T}}\boldsymbol{\Theta}_i\{\boldsymbol{\beta}_0,m_0,\mathbf{w}_0(\mathbf{T}_i)\}\boldsymbol{\Omega}_i^{-1}\boldsymbol{\Theta}_i\{\boldsymbol{\beta}_0,m_0,\mathbf{w}_0(\mathbf{T}_i)\} \\
& \times\widetilde{\mathbf{Q}}_{\lambda i}\{\mathbf{w}_0(\mathbf{T}_i)\}]\bigg)^{-1}\mathbf{B}_r(u)\{1+o_p(1)\} \\
= \ & \mathbf{B}_r(u)^{\mathrm{T}}n^{-1}\mathbf{V}_n^{-1}n^{-1}\sum_{i=1}^{n}\widetilde{\mathbf{Q}}_{\lambda i}\{\mathbf{w}_0(\mathbf{T}_i)\}^{\mathrm{T}}\boldsymbol{\Theta}_i\{\boldsymbol{\beta}_0,m_0,\mathbf{w}_0(\mathbf{T}_i)\} \\
& \times\boldsymbol{\Omega}_i^{-1}\boldsymbol{\Omega}_i^{*}\boldsymbol{\Omega}_i^{-1}\boldsymbol{\Theta}_i\{\boldsymbol{\beta}_0,m_0,\mathbf{w}_0(\mathbf{T}_i)\}\widetilde{\mathbf{Q}}_{\lambda i}\{\mathbf{w}_0(\mathbf{T}_i)\}\mathbf{V}_n^{-1}\mathbf{B}_r(u) \\
= \ & \widehat{\sigma}^2(u,\mathbf{w}_0).
\end{aligned}
$$

Further, because

$$\|\mathbf{V}_n^{-1}-\mathbf{V}^{-1}\|_2 = O_p(h_b^{-2}\sqrt{n^{-1}h_b})$$

as shown in (S.20). And

$$
\begin{aligned}
& \bigg\|n^{-1}\sum_{i=1}^{n}\widetilde{\mathbf{Q}}_{\lambda i}\{\mathbf{w}_0(\mathbf{T}_i)\}^{\mathrm{T}}\boldsymbol{\Theta}_i\{\boldsymbol{\beta}_0,m_0,\mathbf{w}_0(\mathbf{T}_i)\}\boldsymbol{\Omega}_i^{-1}\boldsymbol{\Omega}_i^{*}\boldsymbol{\Omega}_i^{-1} \\
& \times\boldsymbol{\Theta}_i\{\boldsymbol{\beta}_0,m_0,\mathbf{w}_0(\mathbf{T}_i)\}\widetilde{\mathbf{Q}}_{\lambda i}\{\mathbf{w}_0(\mathbf{T}_i)\} - E[\widetilde{\mathbf{Q}}_{\lambda i}\{\mathbf{w}_0(\mathbf{T}_i)\}^{\mathrm{T}} \\
& \times\boldsymbol{\Theta}_i\{\boldsymbol{\beta}_0,m_0,\mathbf{w}_0(\mathbf{T}_i)\}\boldsymbol{\Omega}_i^{-1}\boldsymbol{\Omega}_i^{*}\boldsymbol{\Omega}_i^{-1}\boldsymbol{\Theta}_i\{\boldsymbol{\beta}_0,m_0,\mathbf{w}_0(\mathbf{T}_i)\}\widetilde{\mathbf{Q}}_{\lambda i}\{\mathbf{w}_0(\mathbf{T}_i)\}]\bigg\|_2
\end{aligned}
$$

has the same order as the square root of the 2-norm of its variance, which has at most the same order as

$$
\begin{aligned}
& \bigg\|n^{-1}E\bigg[\widetilde{\mathbf{Q}}_{\lambda i}\{\mathbf{w}_0(\mathbf{T}_i)\}^{\mathrm{T}}\boldsymbol{\Theta}_i\{\boldsymbol{\beta}_0,m_0,\mathbf{w}_0(\mathbf{T}_i)\}\boldsymbol{\Omega}_i^{-1}\boldsymbol{\Omega}_i^{*} \\
& \boldsymbol{\Omega}_i^{-1}\boldsymbol{\Theta}_i\{\boldsymbol{\beta}_0,m_0,\mathbf{w}_0(\mathbf{T}_i)\}\bigg]^{\otimes 2}\bigg\|_2^{1/2}
\end{aligned}
$$

because all the entries of $\widetilde{\mathbf{Q}}_{\lambda i}\{\mathbf{w}_0(\mathbf{T}_i)\}$ are less than 1. Therefore by (S.14), we have

$$\Big\| n^{-1}\sum_{i=1}^{n}\widetilde{\mathbf{Q}}_{\lambda i}\{\mathbf{w}_0(\mathbf{T}_i)\}^{\mathrm{T}}\boldsymbol{\Theta}_i\{\boldsymbol{\beta}_0,m_0,\mathbf{w}_0(\mathbf{T}_i)\}\boldsymbol{\Omega}_i^{-1}\boldsymbol{\Omega}_i^{*}\boldsymbol{\Omega}_i^{-1} \\ \times\boldsymbol{\Theta}_i\{\boldsymbol{\beta}_0,m_0,\mathbf{w}_0(\mathbf{T}_i)\}\widetilde{\mathbf{Q}}_{\lambda i}\{\mathbf{w}_0(\mathbf{T}_i)\} - E[\widetilde{\mathbf{Q}}_{\lambda i}\{\mathbf{w}_0(\mathbf{T}_i)\}^{\mathrm{T}} \\ \times\boldsymbol{\Theta}_i\{\boldsymbol{\beta}_0,m_0,\mathbf{w}_0(\mathbf{T}_i)\}\boldsymbol{\Omega}_i^{-1}\boldsymbol{\Omega}_i^{*}\boldsymbol{\Omega}_i^{-1}\boldsymbol{\Theta}_i\{\boldsymbol{\beta}_0,m_0,\mathbf{w}_0(\mathbf{T}_i)\} \\ \times\widetilde{\mathbf{Q}}_{\lambda i}\{\mathbf{w}_0(\mathbf{T}_i)\}]\Big\|_2 \le O_p(n^{-1/2}h_b^{1/2}).$$

Combine above results with the fact that

$$\Big\| E[\widetilde{\mathbf{Q}}_{\lambda i}\{\mathbf{w}_0(\mathbf{T}_i)\}^{\mathrm{T}}\boldsymbol{\Theta}_i\{\boldsymbol{\beta}_0,m_0,\mathbf{w}_0(\mathbf{T}_i)\}\boldsymbol{\Omega}_i^{-1}\boldsymbol{\Omega}_i^{*}\boldsymbol{\Omega}_i^{-1} \\ \times\boldsymbol{\Theta}_i\{\boldsymbol{\beta}_0,m_0,\mathbf{w}_0(\mathbf{T}_i)\}\widetilde{\mathbf{Q}}_{\lambda i}\{\mathbf{w}_0(\mathbf{T}_i)\}]\Big\|_2 = O(h_b),$$

we have

$$\begin{aligned} & |\widehat{\sigma}^2(u,\mathbf{w}_0)-\sigma^2(u,\mathbf{w}_0)| \\ = \;& n^{-1}\|\mathbf{B}_r(u)\|_2^2\Bigg(\Bigg\|E\Bigg[\widetilde{\mathbf{Q}}_{\lambda i}\{\mathbf{w}_0(\mathbf{T}_i)\}^{\mathrm{T}}\boldsymbol{\Theta}_i\{\boldsymbol{\beta}_0,m_0,\mathbf{w}_0(\mathbf{T}_i)\}\boldsymbol{\Omega}_i^{-1}\boldsymbol{\Omega}_i^{*}\boldsymbol{\Omega}_i^{-1} \\ & \times\boldsymbol{\Theta}_i\{\boldsymbol{\beta}_0,m_0,\mathbf{w}_0(\mathbf{T}_i)\}\widetilde{\mathbf{Q}}_{\lambda i}\{\mathbf{w}_0(\mathbf{T}_i)\}\Bigg]\Bigg\|_2\|\mathbf{V}_n^{-1}-\mathbf{V}^{-1}\|_2^2 \\ & +\|\mathbf{V}^{-1}\|_2^2\Bigg\|n^{-1}\sum_{i=1}^{n}\widetilde{\mathbf{Q}}_{\lambda i}\{\mathbf{w}_0(\mathbf{T}_i)\}^{\mathrm{T}}\boldsymbol{\Theta}_i\{\boldsymbol{\beta}_0,m_0,\mathbf{w}_0(\mathbf{T}_i)\}\boldsymbol{\Omega}_i^{-1}\boldsymbol{\Omega}_i^{*} \\ & \boldsymbol{\Omega}_i^{-1}\boldsymbol{\Theta}_i\{\boldsymbol{\beta}_0,m_0,\mathbf{w}_0(\mathbf{T}_i)\}\widetilde{\mathbf{Q}}_{\lambda i}\{\mathbf{w}_0(\mathbf{T}_i)\} - E\Bigg[\widetilde{\mathbf{Q}}_{\lambda i}\{\mathbf{w}_0(\mathbf{T}_i)\}^{\mathrm{T}} \\ & \times\boldsymbol{\Theta}_i\{\boldsymbol{\beta}_0,m_0,\mathbf{w}_0(\mathbf{T}_i)\}\boldsymbol{\Omega}_i^{-1}\boldsymbol{\Omega}_i^{*}\boldsymbol{\Omega}_i^{-1}\boldsymbol{\Theta}_i\{\boldsymbol{\beta}_0,m_0,\mathbf{w}_0(\mathbf{T}_i)\} \\ & \times\widetilde{\mathbf{Q}}_{\lambda i}\{\mathbf{w}_0(\mathbf{T}_i)\}\Bigg]\Bigg\|_2\Bigg) \\ = \;& n^{-1}\|\mathbf{B}_r(u)\|_2^2\{O_p(h_bh_b^{-4}n^{-1}h_b)+O_p(h_b^{-2}n^{-1/2}h_b^{1/2})\} \\ = \;& \|\mathbf{B}_r(u)\|_2^2\{O_p(h_b^{-2}n^{-2})+O_p(h_b^{-3/2}n^{-3/2})\} \end{aligned}$$

which is $o_p\{(nh_b)^{-1}\}$. Moreover, by Lemma 1, we have, with probability 1,

$$c_{\widehat{\sigma}}(nh_b)^{-1} \le \inf_{u\in[0,1]}\widehat{\sigma}^2(u,\mathbf{w}_0) \le \sup_{u\in[0,1]}\widehat{\sigma}^2(u,\mathbf{w}_0) \le C_{\widehat{\sigma}}(nh_b)^{-1}$$

so

$$c_{\sigma}(nh_b)^{-1} \le \inf_{u\in[0,1]}\sigma^2(u,\mathbf{w}_0) \le \sup_{u\in[0,1]}\sigma^2(u,\mathbf{w}_0) \le C_{\sigma}(nh_b)^{-1}$$

where $0 < c_{\widehat{\sigma}} < C_{\widehat{\sigma}} < \infty$, and $0 < c_\sigma < C_\sigma < \infty$.

As a result, for $u \in [0,1]$,

$$\widehat{\sigma}(u, \mathbf{w}_0)^{-1}[\widehat{m}\{u, \widehat{\boldsymbol{\lambda}}(\boldsymbol{\beta}_0, \mathbf{w}_0)\} - \widetilde{m}(u, \boldsymbol{\lambda}_0)] \xrightarrow{d} N(0,1),$$

or

$$[\sigma(u, \mathbf{w}_0)\{1 + o_p(1)\}]^{-1}[\widehat{m}\{u, \widehat{\boldsymbol{\lambda}}(\boldsymbol{\beta}_0, \mathbf{w}_0)\} - \widetilde{m}(u, \boldsymbol{\lambda}_0)] \xrightarrow{d} N(0,1).$$

By Fact 1,

$$\sup_{u\in[0,1]} |\widetilde{m}(u, \boldsymbol{\lambda}_0) - m_0(u)| = o(h_b^q),$$

along with the boundedness of $\widehat{\sigma}(u, \mathbf{w}_0)$ and Slutsky's rule, we obtain that

$$\sigma(u, \mathbf{w}_0)^{-1}[\widehat{m}\{u, \widehat{\boldsymbol{\lambda}}(\boldsymbol{\beta}_0, \mathbf{w}_0)\} - m_0(u)] \quad \xrightarrow{d} \quad N(0,1).$$

We have

$$|\widehat{m}\{u, \widehat{\boldsymbol{\lambda}}(\boldsymbol{\beta}_0, \mathbf{w}_0)\} - \widetilde{m}(u, \boldsymbol{\lambda}_0)| = O_p\{(nh_b)^{-1/2}\}.$$

To obtain the final results, by Lemma 7, we have

$$\left| \sqrt{nh_b}\mathbf{B}_r^{\mathrm{T}}(u) \left[\widehat{\boldsymbol{\lambda}}\{\widehat{\boldsymbol{\beta}}, \widehat{\mathbf{w}}(\widehat{\boldsymbol{\beta}})\} - \widehat{\boldsymbol{\lambda}}(\widehat{\boldsymbol{\beta}}, \mathbf{w}_0)\right] \right| = O_p(1).$$

Because $\sqrt{n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) = O_p(1)$,

$$\left| \sqrt{nh_b}\mathbf{B}_r^{\mathrm{T}}(u) \left[\widehat{\boldsymbol{\lambda}}(\widehat{\boldsymbol{\beta}}, \mathbf{w}_0) - \widehat{\boldsymbol{\lambda}}(\boldsymbol{\beta}_0, \mathbf{w}_0)\right] \right| = O_p(h_b^{1/2}).$$

Further,

$$\begin{aligned}
& \mathrm{var}\left(\mathbf{B}_r^{\mathrm{T}}(u)\left[\widehat{\boldsymbol{\lambda}}\{\widehat{\boldsymbol{\beta}}, \widehat{\mathbf{w}}(\widehat{\boldsymbol{\beta}})\} - \widehat{\boldsymbol{\lambda}}(\widehat{\boldsymbol{\beta}}, \mathbf{w}_0)\right]\right) \\
\to\ & \mathbf{B}_r^{\mathrm{T}}(u)\frac{1}{n}E\Bigg\{\Bigg(\mathbf{V}^{-1}E\Bigg[\sum_{k=1}^{M_i}\sum_{v=1}^{M_i}E\Bigg\{C_{ikv}\Theta_{ik}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(T_{ik})\} \\
& \times\Theta_{iv}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(T_{iv})\}\mathbf{B}_r\{\mathbf{w}_0(T_{iv})^{\mathrm{T}}\mathbf{Z}_{iv}\}m_0'(\mathbf{w}_0(T_{ik})^{\mathrm{T}}\mathbf{Z}_{ik})\mathbf{Z}_{ik}^{\mathrm{T}} \\
& \times\Bigg(\{\mathbf{B}(T_{ik})f_T(T_{ik})\}^{-1}\Bigg[\mathbf{A}_{wj}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(T_{ik})\}\mathbf{V}_{wj}\{\boldsymbol{\beta}_0, m_0, \\
& \times\mathbf{w}_0(T_{ik})\}^{-1}\mathbf{K}_h(\mathbf{T}_j - T_{ik}) \\
& \times\mathbf{S}_{wj}\{\boldsymbol{\beta}_0, m_0, \mathbf{w}_0(T_{ik})\}\Bigg]\Bigg)\Bigg|M_i, \mathbf{O_j}\Bigg\}\Bigg|\mathbf{O_j}\Bigg]\Bigg)^{\otimes 2}\Bigg\}\mathbf{B}_r(u) \\
=\ & \sigma_w^2.
\end{aligned}$$

Therefore,

$$\{\sigma^2(u, \mathbf{w}_0) + \sigma_w^2\}^{-1/2}\left(\widehat{m}[u, \widehat{\boldsymbol{\lambda}}\{\widehat{\boldsymbol{\beta}}, \widehat{\mathbf{w}}(\widehat{\boldsymbol{\beta}})\}] - m_0(u)\right) \quad \xrightarrow{d} \quad N(0,1).$$

Together with Fact 1, we have

$$\begin{aligned}
|\widehat{m}[u, \widehat{\boldsymbol{\lambda}}\{\widehat{\boldsymbol{\beta}}, \widehat{\mathbf{w}}(\widehat{\boldsymbol{\beta}})\}] - m_0(u)| &= O_p\{(nh_b)^{-1/2} + h_b^q\}, \\
|\widehat{m}'[u, \widehat{\boldsymbol{\lambda}}\{\widehat{\boldsymbol{\beta}}, \widehat{\mathbf{w}}(\widehat{\boldsymbol{\beta}})\}] - m_0'(u)| &= O_p\{n^{-1/2}h_b^{-3/2} + h_b^{q-1}\},
\end{aligned}$$

uniformly for $u \in (0,1)$. □

APPENDIX S.4: CONVERGENCY.

We plot the values of $\widehat{\boldsymbol{\beta}}$ in the last few iterations to show the convergence of the estimation in Section 4.

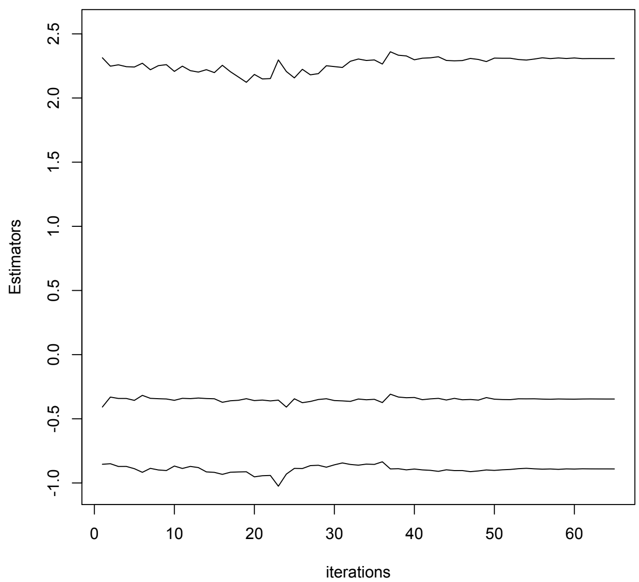

Fig S.1: The values in the last few iterations.

REFERENCES

Bishop, Y. M., Fienberg, S. E. and Holland, P. W. (2007). *Discrete Multivariate Analysis: Theory and Practice*. New York : Springer, c2007.

Bosq, D. (1998). *Nonparametric statistics for stochastic processes : estimation and prediction D. Bosq. Lecture notes in statistics: 110*. New York : Springer, c1998.

de Boor, C. (2001). *A Practical Guide to Splines. Applied Mathematical Sciences* **v. 27**. Springer.

DeVore, R. A. and Lorentz, G. G. (1993). *Constructive approximation. Grundlehren der mathematischen Wissenschaften: 303*. Berlin ; New York : Springer-Verlag, c1993.